Contents

Welcome to Programming for Data Science

Welcome to the course manual for CSC310 at URI.

This website will contain the syllabus, class notes and other reference material for the class.

Course Calendar with zoom links on BrightSpace

> **Tip**
>
> subscribe to that calendar in your favorite calendar application

Syllabus

Welcome to CSC/DSP310: Programming For Data Science.

In this syllabus you will find an overview of the course, information about your instructor, course policies, restatements of URI policies, reminders of relevant resources, and a schedule for the course.

About

About this course

Data science exists at the intersection of computer science, statistics, and machine learning. That means writing programs to access and manipulate data so that it becomes available for analysis using statistical and machine learning techniques is at the core of data science. Data scientists use their data and analytical ability to find and interpret rich data sources; manage large amounts of data despite hardware, software, and bandwidth constraints; merge data sources; ensure consistency of datasets; create visualizations to aid in understanding data; build mathematical models using the data; and present and communicate the data insights/findings.

This course provides a survey of data science. Topics include data driven programming in Python; data sets, file formats and meta-data; descriptive statistics, data visualization, and foundations of predictive data modeling and machine learning; accessing web data and databases; distributed data management. You will work on weekly substantial programming problems such as accessing data in database and visualize it or build machine learning models of a given data set.

Basic programming skills (CSC201 or CSC211) are a prerequisite to this course. This course is a prerequisite course to machine learning, where you learn how machine learning algorithms work. In this course, we will start with a very fast review of basic programming ideas, since you've already done that before. We will learn how to *use* machine learning algorithms to do data science, but not how to *build* machine learning algorithms, we'll use packages that implement the algorithms for us.

About this semester

This semester is a lot of new things for all of us. This course will be completely online all semester, so we will get to use a single instructional format all semester, including when all campus activities move remote after Thanksgiving. I recognize that those last two weeks of the semester may change your obligations with siblings, parents, work, etc. In light of that, we will cover all of the most important topics and you will have the opportunity to achieve all of the course learning outcomes before Thanksgiving. The material in the last two weeks of the semester will be more advanced, likely interesting and definitely useful material, but if your ability to participate in class is less at that time, it will not hurt your grade.

About this syllabus

This syllabus is a *living* document and accessible from BrightSpace, as a pdf for download directly online at rhodyprog4ds.github.io/BrownFall20/syllabus. If you choose to download a copy of it, note that it is only a copy. You can get notification of changes from GitHub by "watching" the repository. You can view the date of changes and exactly what

changes were made on the Github commits page.

Creating an issue on the repository is also a good way to ask questions about anything in the course it will prompt additions and expand the FAQ section.

About your instructor

Name: Dr. Sarah Brown Office hours: TBA via zoom, link in BrightSpace

Dr. Brown is a new Assistant Professor of Computer Science, who does research on how social context changes machine learning. Dr. Brown earned a PhD in Electrical Engineering from Northeastern University, completed a postdoctoral fellowship at University of California Berkeley, and worked as a postdoctoral research associate at Brown University before joining URI. At Brown University, Dr. Brown taught the Data and Society course for the Master's in Data Science Program.

The best way to contact me is e-mail or by dropping into my office hours. Please include `[CSC310]` or `[DSP310]` in the subject line of your email along with the topic of your message. This is important, because your messages are important, but I also get a lot of e-mail. Consider these a cheat code to my inbox: I have setup a filter that will flag your e-mail if you use one of those in the subject to ensure that I see it. I rarely check e-mail between 6pm and 9am, on weekends or holidays. You might see me post or send things during these hours, but I will not reliably see emails that arrive during those hours.

> **Note**
>
> Whether you use CSC or DSP does not matter.

Tools and Resources

We will use a variety of tools to conduct class and to facilitate your programming. You will need a computer with Linux, MacOS, or Windows. It is unlikely that a tablet or Chromebook will be able to do all of the things required in this course.

All of the tools below are either: - paid for by URI or - freely available online.

BrightSpace

This will be the central location from which you can access all other materials. Any links that are for private discussion among those enrolled in the course will be available only from our course Brightspace site. This is also where your grades will appear.

Zoom

This is where we will meet for synchronous class sessions. You will find the link to class zoom sessions on Brightspace.

URI provides all faculty, staff, and students with a paid Zoom account. It *can* run in your browser or on a mobile device, but you will be able to participate in class best if you download the Zoom client on your computer. Please log in and configure your account. Please add a photo of yourself to your account so that we can still see your likeness in some form when your camera is off. You may also wish to use a virtual background and you are welcome to do so.

Class will be interactive, so if you cannot be in a quiet place at class time, headphones with a built in microphone are strongly recommended.

For help, you can access the instructions provided by IT.

> **Important**
>
> TL;DR [1]
>
> - check Brightspace
> - Install Zoom
> - Setup your URI Zoom Account
> - Log in to Prismia Chat
> - Make a GitHub Account
> - Install Python
> - Install Git

Prismia chat

Our class link for Prismia chat is available on Brightspace. We will use this for chatting and in-class understanding checks.

On Prismia, all students see the instructor's messages, but only the Instructor and TA see student responses.

Course Manual

The course manual will have content including the class policies, scheduling, class notes, assignment information, and additional resources. This will be linked from Brightspace and available publicly online at rhodyprog4ds.github.io/BrownFall20/. Links to the course reference text and code documentation will also be included

here in the assignments and class notes.

GitHub Classroom

You will need a GitHub Account. If you do not already have one, please create one by the first day of class. There will be a link to our class GitHub Classroom on Brightspace.

Programming Environment

This a programming course, so you will need a programming environment. In order to complete assignments you need the items listed in the requirements list. The easiest way to meet these requirements is to follow the recommendations below. I will provide instruction assuming that you have followed the recommendations.

Requirements:

- Python with scientific computing packages (numpy, scipy, jupyter, pandas, etc)
- Git
- A web browser compatible with Jupyter Notebooks

Recommendation:

- Install python via Anaconda
- if you use Windows, install Git with GitBash (video instructions).
- if you use MacOS, install Git with the Xcode Command Line Tools. On Mavericks (10.9) or above you can do this by trying to run git from the Terminal the very first time.`git --version`

Optional:

- Text Editor: you may want a text editor outside of the Jupyter environment. Jupyter can edit markdown files (that you'll need for your portfolio), in browser, but it is more common to use a text editor like Atom or Sublime for this purpose.

> **Note**
>
> all Git instructions will be given as instructions for the command line interface and GitHub specific instructions via the web interface. You may choose to use GitHub desktop or built in IDE tools, but the instructional team may not be able to help.

> **Note**
>
> I use atom, but I decided to use it by downloading both Atom and Sublime and trying different things in each for a week. I liked Atom better after that and I've stuck with it since. I used Atom to write all of the content in this syllabus.

Textbook

The text for this class is a reference book and iwll not be a source of assignments. It will be a helpful reference and you may be directed there for answers to questions or alternate explanations ot topics.

Python for Data Science is available free online:

---

[1] Too long; didn't read.

Grading

This section of the syllabus describes the principles and mechanics of the grading for the course. This course will be graded on a basis of a set of *skills* (described in detail the next section of the syllabus). This is in contrast to more common grading on a basis of points earned through assignments.

Principles of Grading

Learning happens through practice and feedback. My goal as a teacher is for you to learn. The grading in this course is based on your learning of the material, rather than your completion of the activities that are assigned.

This course is designed to encourage you to work steadily at learning the material and demonstrating your new knowledge. There are no single points of failure, where you lose points that cannot be recovered. Also, you cannot cram anything one time and then forget it. The material will build and you have to demonstrate that you retained things.

- Earning a C in this class is intended to be easier than typical grading. I expect everyone to get at least a C.
- Earning a B in this class is intended to be very accessible, you can make a lot of mistakes along the way as you learn, as long as you learn by the end.
- Earning an A in this class will be challenging, but is possible even with making mistakes while you learn.

Grading this way also is more amenable to the fact that there are correct and incorrect ways to do things, but there is not always a single correct answer to a realistic data science problem. Your work will be assessed on whether or not it demonstrates your learning of the targeted skills. You will also receive feedback on how to improve.

How it works

There are 15 skills that you will be graded on in this course. While learning these skills, you will work through a progression of learning. Your grade will be based on earning 45 achievements that are organized into 15 skill groups with 3 levels for each.

These map onto letter grades roughly as follows:

- If you achieve level 1 in all of the skills, you will earn at least a C in the course.
- To earn a B, you must earn all of the level 1 and level 2 achievements.
- To earn an A, you must earn all of the achievements.

You will have at least three opportunities to earn every level 2 achievement. You will have at least two opportunities to earn every level 3 achievement. You will have three *types* of opportunities to demonstrate your current skill level: participation, assignments, and a portfolio.

Each level of achievement corresponds to a phase in your learning of the skill:

- To earn level 1 achievements, you will need to demonstrate basic awareness of the required concepts and know approximately what to do, but you may need specific instructions of which things to do or to look up examples to modify every step of the way. You can earn level 1 achievements in class, assignments, or portfolio submissions.
- To earn level 2 achievements you will need to demonstrate understanding of the concepts and the ability to apply them with instruction after earning the level 1 achievement for that skill. You can earn level 2 achievements in assignments or portfolio submissions.
- To earn level 3 achievements you will be required to consistently execute each skill and demonstrate deep understanding of the course material, after achieving level 2 in that skill. You can earn level 3 achievements only through your portfolio submissions.

Participation

While attending synchronous class sessions, there will be understanding checks and in class exercises. Completing in class exercises and correctly answering questions in class can earn level 1 achievements. In class questions will be administered through the classroom chat platform Prismia.chat; these records will be used to update your skill progression. You can also earn level 1 achievements from adding annotation to a section of the class notes.

Assignments

For your learning to progress and earn level 2 achievements, you must practice with the skills outside of class time.

Assignments will each evaluate certain skills. After your assignment is reviewed, you will get qualitative feedback on your work, and an assessment of your demonstration of the targeted skills.

Warning

If you will skip an assignment, please accept the GitHub assignment and then close the Feedback pull request with a comment. This way we can make sure that you have support you need.

Portfolio Checks

To earn level 3 achievements, you will build a portfolio consisting of reflections, challenge problems, and longer analyses over the course of the semester. You will submit your portfolio for review 4 times. The first two will cover the skills taught up until 1 week before the submission deadline.

The third and fourth portfolio checks will cover all of the skills. The fourth will be due during finals. This means that, if you have achieved mastery of all of the skills by the 3rd portfolio check, you do not need to submit the fourth one.

Portfolio prompts will be given throughout the class, some will be strucutred questions, others may be questions that arise in class, for which there is not time to answer.

TLDR

You *could* earn a C through in class participation alone, if you make nearly zero mistakes. To earn a B, you must complete assignments and participate in class. To earn an A you must participate, complete assignments, and build a portfolio.

Detailed mechanics

On Brightspace there are 45 Grade items that you will get a 0 or a 1 grade for. These will be revealed, so that you can view them as you have an opportunity to demonstrate each one. The table below shows the minimum number of skills at each level to earn each letter grade.

letter grade	Level 3	Level 2	Level 1
A	15	15	15
A-	10	15	15
B+	5	15	15
B	0	15	15
B-	0	10	15
C+	0	5	15
C	0	0	15
C-	0	0	10
D+	0	0	5
D	0	0	3

For example, if you achieve level 2 on all of the skills and level 3 on 7 skills, that will be a B+.

If you achieve level 3 on 14 of the skills, but only level 1 on one of the skills, that will be a B-, because the minimum number of level 2 achievements for a B is 15. In this scenario the total number of achievements is 14 at level 3, 14 at level 2 and 15 at level 3, because you have to earn achievements within a skill in sequence.

> **Note**
>
> In this example, you will have also achieved level 1 on all of the skills, because it is a prerequisite to level 2.

The letter grade can be computed as follows

```
def compute_grade(num_level1,num_level2,num_level3):
    '''
    Computes a grade for CSC/DSP310 from numbers of achievements at each level

    Parameters:
    ------------
    num_level1 : int
      number of level 1 achievements earned
    num_level2 : int
      number of level 2 achievements earned
    num_level3 : int
      number of level 3 achievements earned

    Returns:
    --------
    letter_grade : string
      letter grade with modifier (+/-)
    '''
    if num_level1 == 15:
        if num_level2 == 15:
            if num_level3 == 15:
                grade = 'A'
            elif num_level3 >= 10:
                grade = 'A-'
            elif num_level3 >=5:
                grade = 'B+'
            else:
                grade = 'B'
        elif num_level2 >=10:
            grade = 'B-'
        elif num_level2 >=5:
            grade = 'C+'
        else:
            grade = 'C'
    elif num_level1 >= 10:
        grade = 'C-'
    elif num_level1 >= 5:
        grade = 'D+'
    elif num_level1 >=3:
        grade = 'D'
    else:
        grade = 'F'


    return grade
```

```
compute_grade(15,15,15)
```

```
'A'
```

```
compute_grade(14,14,14)
```

```
'C-'
```

```
assert compute_grade(14,14,14) == 'C-'
```

```
assert compute_grade(15,15,15) == 'A'
```

```
assert compute_grade(15,15,11) == 'A-'
```

Late work

No late work will be graded. Every skill will be assessed through more than one assignment, so missing assignments occasionally *may* not hurt your grade. If you do not submit any assignments that cover a given skill, you may earn the level 2 achievement in that skill through a portfolio check, but you will not be able to earn the level 3 achievement in that skill.

Note

You may visit office hours to discuss assignments that you did not complete on time to get feedback and check your own understanding, but they will not count toward skill demonstration.

Examples

Important

The following will make more sense after you read the next section of the syllabus and see the skills rubric sections.

If you always attend and get everything correct, you will earn and A and you won't need to submit the 4th portfolio check or assignment 13.

Getting A Without Perfection

In this example the student made several mistakes, but still earned an A. This is the advantage to this grading scheme. For the `python`, `process`, and `classification` skills, the level 1 achievements were earned on assignments, not in class. For the `process` and `classification` skills, the level 2 achievements were not earned on assignments, only on portfolio checks, but they were earned on the first portfolio of those skills, so the level 3 achievements were earned on the second portfolio check for that skill. This student's fourth portfolio only demonstrated two skills: `optimize` and `unstructured`. It included only 1 analysis, a text analysis with optimizing the parameters of the model. Assignments 4 and 7 were both submitted, but didn't earn any achievements, the student got feedback though, that they were able to apply in later assignments to earn the achievements. The student missed class week 6 and chose to not submit assignment 6 and use week 7 to catch up. The student had too much work in another class and chose to skip assignment 8. The student tried assignment 12, but didn't finish it on time, so it was not graded, but the student visited office hours to understand and be sure to earn the level 2 `unstructured` achievement on assignment 13.

Getting a B with minimal work

In this example, the student earned all level 1 achievements in class and all level 2 on assignments. This student was content with getting a B and chose to not submit a portfolio.

Getting a B while having trouble

In this example, the student struggled to understand in class and on assignments. Assignments were submitted that showed some understanding, but all had some serious mistakes, so only level 1 achievements were earned from assignments. The student wanted to get a B and worked hard to get the level 2 achievements on the portfolio checks.

Learning Objective, Schedule, and Rubric

Learning Outcomes

There are five learning outcomes for this course.

1. (process) Describe the process of data science, define each phase, and identify standard tools
2. (data) Access and combine data in multiple formats for analysis
3. (exploratory) Perform exploratory data analyses including descriptive statistics and visualization
4. (modeling) Select models for data by applying and evaluating mutiple models to a single dataset
5. (communicate) Communicate solutions to problems with data in common industry formats

We will build your skill in the `process` and `communicate` outcomes over the whole semester. The middle three skills will correspond roughly to the content taught for each of the first three portfolio checks.

Schedule

The course will meet MWF 1-1:50pm on Zoom. Every class will include participatory live coding (instructor types, students follow along)) instruction and small exercises for you to progress toward level 1 achievements of the new skills introduced in class that day.

Programming assignments that will be due each week Tuesday by 11:59pm. *until week 5 they were due Sundays*

Note

On the BrightSpace calendar page you can get a feed link to add to the calendar of your choice by clicking on the subscribe (star) button on the top right of the page. Class is for 1 hour there because of Brightspace/zoom integration limitations, but that calendar includes the zoom link.

week	topics	skills
1	[admin, python review]	process
2	Loading data, Python review	[access, prepare, summarize]
3	Exploratory Data Analysis	[summarize, visualize]
4	Data Cleaning	[prepare, summarize, visualize]
5	Databases, Merging DataFrames	[access, construct, summarize]
6	Modeling, Naive Bayes, classification performance metrics	[classification, evaluate]
7	decision trees, cross validation	[classification, evaluate]
8	Regression	[regression, evaluate]
9	Clustering	[clustering, evaluate]
10	SVM, parameter tuning	[optimize, tools]
11	KNN, Model comparison	[compare, tools]
12	Text Analysis	[unstructured]
13	Topic Modeling	[unstructured, tools]
14	Deep Learning	[tools, compare]

Skill Rubric

The skill rubric describes how your participation, assignments, and portfolios will be assessed to earn each achievement. The keyword for each skill is a short name that will be used to refer to skills throughout the course materials; the full description of the skill is in this table.

keyword	skill	Level 1	Level 2	Level 3
python	pythonic code writing	python code that mostly runs, occasional pep8 adherance	python code that reliably runs, frequent pep8 adherance	reliable, efficient, pythonic code that consistently adheres to pep8
process	describe data science as a process	Identify basic components of data science	Describe and define each stage of the data science process	Compare different ways that data science can facilitate decision making
access	access data in multiple formats	load data from at least one format; identify the most common data formats	Load data for processing from the most common formats; Compare and constrast most common formats	access data from both common and uncommon formats and identify best practices for formats in different contexts
construct	construct datasets from multiple sources	identify what should happen to merge datasets or when they can be merged	apply basic merges	merge data that is not automatically aligned
summarize	Summarize and describe data	Describe the shape and structure of a dataset in basic terms	compute summary statndard statistics of a whole dataset and grouped data	Compute and interpret various summary statistics of subsets of data
visualize	Visualize data	identify plot types, generate basic plots from pandas	generate multiple plot types with complete labeling with pandas and seaborn	generate complex plots with pandas and plotting libraries and customize with matplotlib or additional parameters
prepare	prepare data for analysis	identify if data is or is not ready for analysis, potential problems with data	apply data reshaping, cleaning, and filtering as directed	apply data reshaping, cleaning, and filtering manipulations reliably and correctly by assessing data as received
classification	Apply classification	identify and describe what classification is, apply pre-fit classification models	fit preselected classification model to a dataset	fit and apply classification models and select appropriate classification models for different contexts
regression	Apply Regression	identify what data that can be used for regression looks like	can fit linear regression models	can fit and explain regrularized or nonlinear regression
clustering	Clustering	describe what clustering is	apply basic clustering	apply multiple clustering techniques, and interpret results
evaluate	Evaluate model performance	Explain basic performance metrics for different data science tasks	Apply basic model evaluation metrics to a held out test set	Evaluate a model with multiple metrics and cross validation
optimize	Optimize model parameters	Identify when model parameters need to be optimized	Manually optimize basic model parameters such as model order	Select optimal parameters based of mutiple quanttiateve criteria and automate parameter tuning
compare	compare models	Qualitatively compare model classes	Compare model classes in specific terms and fit models in terms of traditional model performance metrics	Evaluate tradeoffs between different model comparison types

	skill	Level 1	Level 2	Level 3
keyword				
unstructured	model unstructured data	Identify options for representing text data and use them once data is tranformed	Apply at least one representation to transform unstructured data for model fitting or summarizing	apply multiple representations and compare and contrast them for different end results
workflow	use industry standard data science tools and workflows to solve data science problems	Solve well strucutred problems with a single tool pipeline	Solve semi-strucutred, completely specified problems, apply common structure to learn new features of standard tools	Scope, choose an appropriate tool pipeline and solve data science problems, describe strengths and weakensses of common tools

Assignments and Skills

Using the keywords from the table above, this table shows which assignments you will be able to demonstrate which skills and the total number of assignments that assess each skill. This is the number of opportunities you have to earn Level 2 and still preserve 2 chances to earn Level 3 for each skill.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	# Assignment
keyword														
python	1	1	1	1	0	0	0	0	0	0	0	0	0	
process	1	1	0	0	0	0	0	0	0	0	0	0	0	
access	0	1	1	1	0	0	0	0	0	0	0	0	0	
construct	0	0	0	0	1	1	0	0	0	0	0	0	0	
summarize	0	0	1	1	1	1	1	1	1	1	1	1	1	1
visualize	0	0	1	1	0	1	1	1	1	1	1	1	1	1
prepare	0	0	0	1	1	0	0	0	0	0	0	0	0	
classification	0	0	0	0	0	1	1	0	0	1	0	0	0	
regression	0	0	0	0	0	0	0	1	0	0	1	0	0	
clustering	0	0	0	0	0	0	0	0	1	0	1	0	0	
evaluate	0	0	0	0	0	0	0	0	0	1	1	0	0	
optimize	0	0	0	0	0	0	0	0	0	1	1	0	0	
compare	0	0	0	0	0	0	0	0	0	0	1	0	1	
unstructured	0	0	0	0	0	0	0	0	0	0	0	1	1	
workflow	0	0	0	0	0	0	0	0	0	1	1	1	1	

Portfolios and Skills

The objective of your portfolio submissions is to earn Level 3 achievements. The following table shows what Level 3 looks like for each skill and identifies which portfolio submissions you can earn that Level 3 in that skill.

	Level 3	P1	P2	P3	P4
keyword					
python	reliable, efficient, pythonic code that consistently adheres to pep8	1	1	0	0
process	Compare different ways that data science can facilitate decision making	0	1	1	0
access	access data from both common and uncommon formats and identify best practices for formats in different contexts	1	1	0	0
construct	merge data that is not automatically aligned	1	1	0	0
summarize	Compute and interpret various summary statistics of subsets of data	1	1	0	0
visualize	generate complex plots with pandas and plotting libraries and customize with matplotlib or additional parameters	1	1	0	0
prepare	apply data reshaping, cleaning, and filtering manipulations reliably and correctly by assessing data as received	1	1	0	0
classification	fit and apply classification models and select appropriate classification models for different contexts	0	1	1	0
regression	can fit and explain regrularized or nonlinear regression	0	1	1	0
clustering	apply multiple clustering techniques, and interpret results	0	1	1	0
evaluate	Evaluate a model with multiple metrics and cross validation	0	1	1	0
optimize	Select optimal parameters based of mutiple quanttiateve criteria and automate parameter tuning	0	0	1	1
compare	Evaluate tradeoffs between different model comparison types	0	0	1	1
unstructured	apply multiple representations and compare and contrast them for different end results	0	0	1	1
workflow	Scope, choose an appropriate tool pipeline and solve data science problems, describe strengths and weakensses of common tools	0	0	1	1

Support

Academic Enhancement Center

Located in Roosevelt Hall, the AEC offers free face to face and web-based services to undergraduate students seeking academic support. Peer tutoring is available for STEM-related courses through drop-in centers and small group tutoring. The Writing Center offers peer tutoring focused on supporting undergraduate writers at any stage of a writing assignment. The UCS160 course and academic skills consultations offer students strategies and activities aimed at improving their studying and test-taking skills. Complete details about each of these programs, up-to-date schedules, contact information and self-service study resources are all available on the AEC website.

- **STEM Tutoring** helps students navigate 100 and 200 level math, chemistry, physics, biology, and other select STEM courses. The STEM Tutoring program offers free online and limited in-person peer-tutoring this fall. Undergraduates in introductory STEM courses have a variety of small group times to choose from and can select occasional or weekly appointments. Appointments and locations will be visible in the TutorTrac system on September 14th, 2020. The TutorTrac application is available through URI Microsoft 365 single sign-on and by visiting aec.uri.edu. More detailed information and instructions can be found on the AEC tutoring page.
- **Academic Skills Development** resources helps students plan work, manage time, and study more effectively. In Fall 2020, all Academic Skills and Strategies programming are offered both online and in-person. UCS160: Success in Higher Education is a one-credit course on developing a more effective approach to studying. Academic Consultations are 30-minute, 1 to 1 appointments that students can schedule on Starfish with Dr. David Hayes to address individual academic issues. Study Your Way to Success is a self-guided web portal connecting students to tips and strategies on studying and time management related topics. For more information on these programs, visit the Academic Skills Page or contact Dr. Hayes directly at davidhayes@uri.edu.
- The **Undergraduate Writing Center** provides free writing support to students in any class, at any stage of the writing process: from understanding an assignment and brainstorming ideas, to developing, organizing, and revising a draft. Fall 2020 services are offered through two online options: 1) real-time synchronous appointments with a peer consultant (25- and 50-minute slots, available Sunday - Friday), and 2) written asynchronous

consultations with a 24-hour turn-around response time (available Monday - Friday). Synchronous appointments are video-based, with audio, chat, document-sharing, and live captioning capabilities, to meet a range of accessibility needs. View the synchronous and asynchronous schedules and book online, visit uri.mywconline.com.

Policies

Anti-Bias Statement

We respect the rights and dignity of each individual and group. We reject prejudice and intolerance, and we work to understand differences. We believe that equity and inclusion are critical components for campus community members to thrive. If you are a target or a witness of a bias incident, you are encouraged to submit a report to the URI Bias Response Team at www.uri.edu/brt. There you will also find people and resources to help.

Disability Services for Students Statement

Your access in this course is important. Please send me your Disability Services for Students (DSS) accommodation letter early in the semester so that we have adequate time to discuss and arrange your approved academic accommodations. If you have not yet established services through DSS, please contact them to engage in a confidential conversation about the process for requesting reasonable accommodations in the classroom. DSS can be reached by calling: 401-874-2098, visiting: web.uri.edu/disability, or emailing: dss@etal.uri.edu. They are available to meet with students enrolled in Kingston as well as Providence courses.

Academic Honesty

Students are expected to be honest in all academic work. A student's name on any written work, quiz or exam shall be regarded as assurance that the work is the result of the student's own independent thought and study. Work should be stated in the student's own words, properly attributed to its source. Students have an obligation to know how to quote, paraphrase, summarize, cite and reference the work of others with integrity. The following are examples of academic dishonesty.

- Using material, directly or paraphrasing, from published sources (print or electronic) without appropriate citation
- Claiming disproportionate credit for work not done independently
- Unauthorized possession or access to exams
- Unauthorized communication during exams
- Unauthorized use of another's work or preparing work for another student
- Taking an exam for another student
- Altering or attempting to alter grades
- The use of notes or electronic devices to gain an unauthorized advantage during exams
- Fabricating or falsifying facts, data or references
- Facilitating or aiding another's academic dishonesty
- Submitting the same paper for more than one course without prior approval from the instructors

URI COVID-19 Statement

The University is committed to delivering its educational mission while protecting the health and safety of our students. At this uncertain time, those concerns include minimizing the potential spread of COVID-19 within our community. While the university has worked this summer to create a healthy learning environment for all, it is up to all of us to ensure our campus stays that way.

As members of the URI community, students are required to comply with standards of conduct and take precautions to keep themselves and others safe. Students are required to comply with Rhode Island state laws, including the Rhode Island Executive Orders related to health and safety, ordinances, regulations, and guidance adopted by the University as it relates to public health crises, such as COVID-19.

An addendum on policies and guidelines concerning your obligations during this crisis has recently been integrated into the Student Handbook. These obligations include:

- Wearing of face masks by all community members when on a URI campus in the presence of others
- Maintaining physical distancing of at least six feet at all times
- Following state rules on the number of individuals allowed in a group gathering

- Completing a daily health self-assessment also available through the Rhody Connect app before coming to campus
- Submitting to COVID-19 testing as the University monitors the health of our community
- Following the University's quarantine and isolation requirements

If you answer yes to any of the questions on the daily health assessment, do not go to campus. YOU MUST STAY HOME/IN YOUR ROOM and notify URI Health Services via phone at 401-874-2246 immediately.

If you are already on campus and start to feel ill, you need to remove yourself from the public and notify URI Health Services via phone immediately at 401-874-2246 and go home/back to your room and self-isolate while you await direction from Health Services.

If you are unable to attend class, please notify me at brownsarahm@uri.edu or through the medium we have established for the class. We will work together to ensure that course instruction and work is completed for the semester.

Class Notes

Class notes will get posted here day by day

- 2020-09-11: Jupyter Notebook tour, conditionals, functions
- 2020-09-14: Iterables, Pandas
- 2020-09-16: Pandas loading and exploring
- 2020-09-18: Pandas, Functions as Object, Dictionaries
- 2020-09-21: Exploratory Data Analysis, Split, apply, Combine
- 2020-09-23: Visualization for EDA
- 2020-09-25: Viz & Cleaning
- 2020-09-28: Preparing data for analysis
- 2020-09-30: Reading complex data and Ray Summit
- 2020-10-02: Cleaning Data
- 2020-10-05: Merging DataFrames
- 2020-10-07: Merging & Databases
- 2020-10-12: Intro to ML
- 2020-10-14: Naive Bayes Classifier
- 2020-10-16: Classification Metrics
- 2020-10-19: Mid-Semester Feedback & Naive Bayes Decision rules
- 2020-10-21: Decision Trees
- 2020-10-23: Decision Trees & Cross validation
- 2020-10-26: Regression
- 2020-10-28: More Regression & More Evaluation
- 2020-10-30: Interpretting Regression Metrics
- 2020-11-02: Clustering
- 2020-11-04: Evaluating Clustering
- 2020-11-06: More Clustering Models
- 2020-11-09: Optimizing Clustering Models
- 2020-11-13: SVM & Grid Search
- 2020-11-16: Model Selection
- 2020-11-18: Learning and Validation Curves
- 2020-11-20: Confidence Intervals
- 2020-11-23: Intro to NLP
- 2020-11-25: More NLP
- 2020-11-20: More text representations

Class 2: intro to notebooks and python

Agenda:

1. review Data Science
2. `jupyter notebook`
3. `python`: conditionals and functions

Jupyter Notebooks

To launch a Jupyter notebook, in your Anaconda prompt on Windows or terminal on Linux or Mac:

```
cd dir/you/want/to/work/in
jupyter notebook
```

A Jupyter notebook has two modes. When you first open, it is in command mode. The border is blue in command mode.

When you press a key in command mode it works like a shortcut. For example `p` shows the command search menu.

If you press `enter` (or `return`) or click on the cell it changes to edit mode. The border is green in edit mode

Type code or markdown into boxes called cells. There are two type of cells that we will used: code and markdown. You can change that in command mode with `y` for code and `m` for markdown or on the cell type menu at the top of the notebook.

You can treat markdown cells like plain text, or use special formatting. Here is a [markdown cheatsheet]

Code cells can run like a calculator. If there is a value returned by the last line of a cell, it will be displayed.

```
4+5
```

```
9
```

For example, when we assign, python “returns” `None` so there is no output from this cell

```
a = 9
```

But this one does display the value of the cell

```
a
```

```
9
```

```
b = 4
b
```

```
4
```

```
a
```

```
9
```

> **Note**
>
> Here in class we changed the value of a above and noted that the new value is shows here, but not in the previous cell that had ouput the value of a

Getting Help in Python and Jupyter

The standard way to get help in the help function

```
help(print)
```

```
Help on built-in function print in module builtins:

print(...)
    print(value, ..., sep=' ', end='\n', file=sys.stdout, flush=False)

    Prints the values to a stream, or to sys.stdout by default.
    Optional keyword arguments:
    file:  a file-like object (stream); defaults to the current sys.stdout.
    sep:   string inserted between values, default a space.
    end:   string appended after the last value, default a newline.
    flush: whether to forcibly flush the stream.
```

Note that these are green in the jupyter notebook because they're python reserved words.

There are two special ways to get help in Jupyter, one dynamically while you're working and one that stays displayed for a while

Python comments are indicated by the # symbol.

```
print() # shift +tab to view help
```

That will look like this:

Press tab twice for the longer version.

A Question mark puts it in a popup window that stays until you close it

```
print?
```

```
In [3]: print?
```

```
Docstring:
print(value, ..., sep=' ', end='\n', file=sys.stdout, flush=False)

Prints the values to a stream, or to sys.stdout by default.
Optional keyword arguments:
file:  a file-like object (stream); defaults to the current sys.stdout.
sep:   string inserted between values, default a space.
end:   string appended after the last value, default a newline.
flush: whether to forcibly flush the stream.
Type:      builtin_function_or_method
```

This means you can then use the displayed help to remember how to call the function

```
print(a,b, 'hello',sep='-')
```

```
9-4-hello
```

```
a
b
```

```
4
```

If Statements

:::{warning} if the *concept* of an `if` in programming is new to you, you should talk to Dr. Brown. Basic programming is a prerequisite to this course, we're *reviewing* basic ideas but only at a level of detail to serve as a reminder. :::

The synta

```
if a >b:
    print('greater')
```

```
greater
```

```
if b > a:
    print('greater')
```

> **Tip**
>
> this is updated to include things that were skipped in class and discussed after the breakouts

You can check the contents of a string with the `in` keyword

```
name = 'sarah'
if 'a' in name:
    print(name, 'has an a')
```

```
sarah has an a
```

> **Tip**
>
> `in` works for all lists, we'll learn more about that next week

if we copy and change the name we get no output

```
name = 'Beibhinn'
if 'a' in name:
    print(name, 'has an a')
```

Functions

> **Tip**
>
> this is also updated to include things that were skipped in class and discussed after the breakouts

How to write functions in python:

- the `def` keyword starts a function definition
- then the function name
- then the parameters it accepts in `()`
- end that line with a `:`
- the body of the function is spaced over one tab, but Jupyter will do it automatically for you. if it doesn't you might have forgotten the `:`

```
def greeting(name):
    '''
    a function that greets the person name by printing

    parameters
    ----------
    name: string
      a name to be greeeted

    Returns
    -------
    nothing
    '''
    print('hello', name)
```

Warning

this is actually not a great function, because the printing is only a *side effect*. It's better to return the output of the function

```
greeting('sarah')
```

```
hello sarah
```

A better version of that function might be:

```
def greeting(name):
    '''
    a function that greets the person name by printing

    parameters
    ----------
    name: string
      a name to be greeeted

    Returns
    -------
    nothing
    '''
    return 'hello ' + name
```

Tip

you can append strings with `+`

Try it Yourself!

Write a function that checks if a string has a space in it and returns "please rename" if there is a space.

Remember a docstring. Call your function a couple of times to confirm it works.

Unhide the cell below to see the answer.

This is what some calls of the function look lik

```
check_string("my data.csv")
```

```
'please rename'
```

If there's no string we see no output

```
check_string("my_data.csv")
```

What does python actually return?

```
NoneType
```

Further Reading

- How Ipython Works
- Ipython Overiver
- Jupyter Notebooks Technical Overview
- Python If Statements
- Python Functions

Class 3: Welcome to Week 2

This week we will:

- clarify how this grading really works
- learn about accessing data
- use accessing data as motivation to review more python

Grading and Assignment 1

- Solution function posted.
- note: not a sum
- read the rubric
- Brightspace will show grades as they're earned
- In class, respond on prismia
- Portfolio
    - will start posting prompts The docstring functions like a property of the function object. so it has to be inside.

Iterables

Python has a general data type for objects that are designed to facilitate repetition of some sort, they're called `iterable`s

We've already seen one. Strings are `Iterable`s

```
name = 'sarah'
```

which means we can index them

```
name[3]
```

```
'a'
```

> **Tip**
>
> remember python indexes from 0

Indexing with a negative number counts from the end

```
name[-1]
```

```
'h'
```

Loops in python have similar syntax to the `if` and functions we saw last week:

```
for char in name:
    print(char*3)
```

```
sss
aaa
rrr
aaa
hhh
```

some notes:

- `char` is called the loop variable
- `name` is called the collection- this can be any iterable type object in python
- `print(char*3)` is called the loop body
- python lets us use mathematical operations on strings

Lists and List Comprehensions

We make a list with square brackets

```
names = ['sarah', 'Jose', 'Cam', 'Bri']
```

we can also build lists by folding a loop into the list construction

```
['hello' + n for n in names]
```

```
['hellosarah', 'helloJose', 'helloCam', 'helloBri']
```

this is called a list comprehension

```
greetings = ['hello ' + n for n in names]
```

```
greetings[0]
```

```
'hello sarah'
```

Dictionaries

Dictionaries are a useful datatype in python. It is denoted by `{}` and contains `key: value` pairs separated by commas.

```
gh_names = {'brownsarahm':'Sarah Brown',
            'briannakathrynm1' : 'Brianna MacDonald',
            'jdion62':'Jacob Dion'}
gh_names
```

```
{'brownsarahm': 'Sarah Brown',
 'briannakathrynm1': 'Brianna MacDonald',
 'jdion62': 'Jacob Dion'}
```

You can think of it like a list of the values with a named index.

```
gh_names['jdion62']
```

```
'Jacob Dion'
```

we can iterate over both the key and the value by using the `items` method on a dictionary. That makes another iterable object that can be used as a loop collection. It functions as a set of pairs now, so we get two loop variables:

```
for key, value in gh_names.items():
    print(value, "'s username is ", key)
```

```
Sarah Brown 's username is  brownsarahm
Brianna MacDonald 's username is  briannakathrynm1
Jacob Dion 's username is  jdion62
```

If we iterate over the dictionary without that method, we get the keys.

```
for val in gh_names:
    print(val)
```

```
brownsarahm
briannakathrynm1
jdion62
```

Libraries

To use libraries in python we import them

We will use `pandas` a lot in this class. It's the Python Data Analysis Library.

```
import pandas
```

Once we import we can use the functions, datatypes, and values a library provides by using a `.` after the name. In a notebook, pressing tab will show you the options.

```
pandas.read_csv()
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-14-374a1a6f9f7e> in <module>
----> 1 pandas.read_csv()

TypeError: read_csv() missing 1 required positional argument: 'filepath_or_buffer'
```

We can also use an alias to give a library a nickname to make it easier to use. `pd` is the standard alias for `pandas`

```
import pandas as pd
```

We can read in from a local path or a url. Let's read in the course map page of our course website.

```
pd.read_html('https://rhodyprog4ds.github.io/BrownFall20/syllabus/course_map.html')
```

Tip

note that `import` is a keyword and that in a Jupyter notebook, we can import anywhere and then the library can be used in any cell that is *run* after the import cell is *run*. It's good practice to put them at the top and make your notebook runnable in sequence, but Jupyter won't force you to.

Tip

For example if you don't remember what kind of read functions there are in pandas, type `pandas.read` and then press tab to see options.

```
[   Unnamed: 0_level_0                                          topics  \
                  week                                Unnamed: 1_level_1
 0                   1                            [admin, python review]
 1                   2                        Loading data, Python review
 2                   3                          Exploratory Data Analysis
 3                   4                                       Data Cleaning
 4                   5                      Databases, Merging DataFrames
 5                   6  Modeling, Naive Bayes, classification performa...
 6                   7                   decision trees, cross validation
 7                   8                                         Regression
 8                   9                                         Clustering
 9                  10                             SVM, parameter tuning
 10                 11                             KNN, Model comparison
 11                 12                                     Text Analysis
 12                 13                                    Topic Modeling
 13                 14                                     Deep Learning

                            skills
                  Unnamed: 2_level_1
 0                          process
 1      [access, prepare, summarize]
 2            [summarize, visualize]
 3   [prepare, summarize, visualize]
 4    [access, construct, summarize]
 5        [classification, evaluate]
 6        [classification, evaluate]
 7            [regression, evaluate]
 8            [clustering, evaluate]
 9                 [optimize, tools]
 10                 [compare, tools]
 11                   [unstructured]
 12            [unstructured, tools]
 13                 [tools, compare]  ,
    Unnamed: 0_level_0                                           skill  \
                keyword                                Unnamed: 1_level_1
 0               python                            pythonic code writing
 1              process                 describe data science as a process
 2               access                    access data in multiple formats
 3            construct           construct datasets from multiple sources
 4            summarize                       Summarize and describe data
 5            visualize                                   Visualize data
 6              prepare                         prepare data for analysis
 7      classification                              Apply classification
```

```
8          regression                                  Apply Regression
9          clustering                                        Clustering
10           evaluate                        Evaluate model performance
11           optimize                         Optimize model parameters
12            compare                                    compare models
13       unstructured                            model unstructured data
14           workflow  use industry standard data science tools and w...

                                                Level 1  \
                                     Unnamed: 2_level_1
0   python code that mostly runs, occasional pep8 ...
1           Identify basic components of data science
2   load data from at least one format; identify t...
3   identify what should happen to merge datasets ...
4   Describe the shape and structure of a dataset ...
5   identify plot types, generate basic plots from...
6   identify if data is or is not ready for analys...
7   identify and describe what classification is, ...
8   identify what data that can be used for regres...
9                        describe what clustering is
10  Explain basic performance metrics for differen...
11  Identify when model parameters need to be opti...
12               Qualitatively compare model classes
13  Identify options for representing text data an...
14  Solve well strucutred problems with a single t...

                                                Level 2  \
                                     Unnamed: 3_level_1
0   python code that reliably runs, frequent pep8 ...
1   Describe and define each stage of the data sci...
2   Load data for processing from the most common ...
3                                  apply basic merges
4   compute summary statndard statistics of a whol...
5   generate multiple plot types with complete lab...
6   apply data reshaping, cleaning, and filtering ...
7   fit preselected classification model to a dataset
8                    can fit linear regression models
9                                apply basic clustering
10  Apply basic model evaluation metrics to a held...
11  Manually optimize basic model parameters such ...
12  Compare model classes in specific terms and fi...
13  Apply at least one representation to transform...
14  Solve semi-strucutred, completely specified pr...

                                                Level 3
                                     Unnamed: 4_level_1
0   reliable, efficient, pythonic code that consis...
1   Compare different ways that data science can f...
2   access data from both common and uncommon form...
3        merge data that is not automatically aligned
4   Compute and interpret various summary statisti...
5   generate complex plots with pandas and plottin...
6   apply data reshaping, cleaning, and filtering ...
7   fit and apply classification models and select...
8   can fit and explain regrularized or nonlinear ...
9   apply multiple clustering techniques, and inte...
10  Evaluate a model with multiple metrics and cro...
11  Select optimal parameters based of mutiple qua...
12  Evaluate tradeoffs between different model com...
13  apply multiple representations and compare and...
14  Scope, choose an appropriate tool pipeline and...  ,
   Unnamed: 0_level_0                A1                A2  \
              keyword Unnamed: 1_level_1 Unnamed: 2_level_1
0             python                  1                  1
1            process                  1                  1
2             access                  0                  1
3          construct                  0                  0
4          summarize                  0                  0
5          visualize                  0                  0
6            prepare                  0                  0
7     classification                  0                  0
8         regression                  0                  0
9         clustering                  0                  0
10          evaluate                  0                  0
11          optimize                  0                  0
12           compare                  0                  0
13      unstructured                  0                  0
14          workflow                  0                  0

                   A3                A4                A5  \
   Unnamed: 3_level_1 Unnamed: 4_level_1 Unnamed: 5_level_1
0                   1                  1                  0
1                   0                  0                  0
2                   1                  1                  0
3                   0                  0                  1
4                   1                  1                  1
5                   1                  1                  0
```

```
6                  0                  1                  1
7                  0                  0                  0
8                  0                  0                  0
9                  0                  0                  0
10                 0                  0                  0
11                 0                  0                  0
12                 0                  0                  0
13                 0                  0                  0
14                 0                  0                  0

                   A6                 A7                 A8  \
   Unnamed: 6_level_1 Unnamed: 7_level_1 Unnamed: 8_level_1
0                  0                  0                  0
1                  0                  0                  0
2                  0                  0                  0
3                  1                  0                  0
4                  1                  1                  1
5                  1                  1                  1
6                  0                  0                  0
7                  1                  1                  0
8                  0                  0                  1
9                  0                  0                  0
10                 0                  0                  0
11                 0                  0                  0
12                 0                  0                  0
13                 0                  0                  0
14                 0                  0                  0

                   A9                 A10                 A11  \
   Unnamed: 9_level_1 Unnamed: 10_level_1 Unnamed: 11_level_1
0                  0                   0                   0
1                  0                   0                   0
2                  0                   0                   0
3                  0                   0                   0
4                  1                   1                   1
5                  1                   1                   1
6                  0                   0                   0
7                  0                   1                   0
8                  0                   0                   1
9                  1                   0                   1
10                 0                   1                   1
11                 0                   1                   1
12                 0                   0                   1
13                 0                   0                   0
14                 0                   1                   1

                   A12                 A13       # Assignments
   Unnamed: 12_level_1 Unnamed: 13_level_1 Unnamed: 14_level_1
0                   0                   0                   4
1                   0                   0                   2
2                   0                   0                   3
3                   0                   0                   2
4                   1                   1                  11
5                   1                   1                  10
6                   0                   0                   2
7                   0                   0                   3
8                   0                   0                   2
9                   0                   0                   2
10                  0                   0                   2
11                  0                   0                   2
12                  0                   1                   2
13                  1                   1                   2
14                  1                   1                   4  ,
   Unnamed: 0_level_0                                          Level 3  \
              keyword                                Unnamed: 1_level_1
0             python  reliable, efficient, pythonic code that consis...
1            process  Compare different ways that data science can f...
2             access  access data from both common and uncommon form...
3          construct       merge data that is not automatically aligned
4          summarize  Compute and interpret various summary statisti...
5          visualize  generate complex plots with pandas and plottin...
6            prepare  apply data reshaping, cleaning, and filtering ...
7     classification  fit and apply classification models and select...
8         regression  can fit and explain regrularized or nonlinear ...
9         clustering  apply multiple clustering techniques, and inte...
10          evaluate  Evaluate a model with multiple metrics and cro...
11          optimize  Select optimal parameters based of mutiple qua...
12           compare  Evaluate tradeoffs between different model com...
13      unstructured  apply multiple representations and compare and...
14          workflow  Scope, choose an appropriate tool pipeline and...

                   P1                 P2                 P3                 P4
   Unnamed: 2_level_1 Unnamed: 3_level_1 Unnamed: 4_level_1 Unnamed: 5_level_1
0                  1                  1                  0                  0
1                  0                  1                  1                  0
2                  1                  1                  0                  0
3                  1                  1                  0                  0
```

```
4                   1                   1                   0                   0
5                   1                   1                   0                   0
6                   1                   1                   0                   0
7                   0                   1                   1                   0
8                   0                   1                   1                   0
9                   0                   1                   1                   0
10                  0                   1                   1                   0
11                  0                   0                   1                   1
12                  0                   0                   1                   1
13                  0                   0                   1                   1
14                  0                   0                   1                   1  ]
```

This makes a `list` of `pandas.DataFrame` objects. We can check that with the following

> ⚠ **Warning**
>
> This cell was added after class, but the explanation was given in class

```
type(pd.read_html('https://rhodyprog4ds.github.io/BrownFall20/syllabus/course_map.html'
))
```

```
list
```

To work with it though, we should save to a variable, then we can index into that list.

```
df_list =
pd.read_html('https://rhodyprog4ds.github.io/BrownFall20/syllabus/course_map.html')
df_list[0]
```

	Unnamed: 0_level_0	topics	skills
	week	Unnamed: 1_level_1	Unnamed: 2_level_1
0	1	[admin, python review]	process
1	2	Loading data, Python review	[access, prepare, summarize]
2	3	Exploratory Data Analysis	[summarize, visualize]
3	4	Data Cleaning	[prepare, summarize, visualize]
4	5	Databases, Merging DataFrames	[access, construct, summarize]
5	6	Modeling, Naive Bayes, classification performa...	[classification, evaluate]
6	7	decision trees, cross validation	[classification, evaluate]
7	8	Regression	[regression, evaluate]
8	9	Clustering	[clustering, evaluate]
9	10	SVM, parameter tuning	[optimize, tools]
10	11	KNN, Model comparison	[compare, tools]
11	12	Text Analysis	[unstructured]
12	13	Topic Modeling	[unstructured, tools]
13	14	Deep Learning	[tools, compare]

When you display `DataFrames` in jupyter, they get nice formatting.

Review & Further Reading

- strings are iterable, more specifically sequences
- `for` loops, looping techniques and list comprehensions
- dictionaries
- imported `pandas` and read data from a website

If you've made it this far, let me know how you found these notes.

Class 4: Pandas

Today we will:

Remember, Programming is a Practice

- if you're curious about something try it
- you don't need me to give you answers about how code works, the interpreter will tell you
- if you don't remember details, remember you can get help from Jupyter

with a `?` after the function name withouth `()`

```
print?
```

or using the `tab` key inside the `()` for a function

```
print()
```

or from the core python, with the `help` fucntion

```
help(print)
```

```
Help on built-in function print in module builtins:

print(...)
    print(value, ..., sep=' ', end='\n', file=sys.stdout, flush=False)

    Prints the values to a stream, or to sys.stdout by default.
    Optional keyword arguments:
    file:  a file-like object (stream); defaults to the current sys.stdout.
    sep:   string inserted between values, default a space.
    end:   string appended after the last value, default a newline.
    flush: whether to forcibly flush the stream.
```

Data in Pandas

We can import `pandas` again as before

```
import pandas as pd
```

and we can read in data.

```
pd.read_csv('https://raw.githubusercontent.com/brownsarahm/python-socialsci-files/master/data/SAFI_clean.csv')
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	roo
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	
2	3	God	2016-11-17T00:00:00Z	10	15	burntbricks	
3	4	God	2016-11-17T00:00:00Z	7	6	burntbricks	

to be able to use this, we need to save it to a variable.

```
safi_df = pd.read_csv('https://raw.githubusercontent.com/brownsarahm/python-socialsci-files/master/data/SAFI_clean.csv')
```

This is an excerpt from the SAFI dataset.

Another important thing to do is to check datatypes, this is how we know what things we can do with a variable.

```
type(safi_df)
```

```
pandas.core.frame.DataFrame
```

An important thing to check is the size of the dataset.

```
safi_df.shape
```

```
(131, 14)
```

Tip

we can also see the size when we print the whole dataset as in the first time we read the data in.

Recall that you can also tab complete

```
safi_df.shape
```

```
(131, 14)
```

To see the first 5 rows of the dataset, use the `head()` function

```
safi_df.head()
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1
2	3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1
3	4	God	2016-11-17T00:00:00Z	7	6	burntbricks	1
4	5	God	2016-11-17T00:00:00Z	7	40	burntbricks	1

We can call this function with a value to change how many rows are returned

```
safi_df.head(3)
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1
2	3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1

To know how this works, we can view the documentation for the function

```
help(safi_df.head)
```

Warning

this was changed from using ? for help in class so that the help is desplayed in the rendered website, but the pop up was fine in real time

```
Help on method head in module pandas.core.generic:

head(n: 'int' = 5) -> 'FrameOrSeries' method of pandas.core.frame.DataFrame instance
    Return the first `n` rows.

    This function returns the first `n` rows for the object based
    on position. It is useful for quickly testing if your object
    has the right type of data in it.

    For negative values of `n`, this function returns all rows except
    the last `n` rows, equivalent to ``df[:-n]``.

    Parameters
    ----------
    n : int, default 5
        Number of rows to select.

    Returns
    -------
    same type as caller
        The first `n` rows of the caller object.

    See Also
    --------
    DataFrame.tail: Returns the last `n` rows.

    Examples
    --------
    >>> df = pd.DataFrame({'animal': ['alligator', 'bee', 'falcon', 'lion',
    ...                    'monkey', 'parrot', 'shark', 'whale', 'zebra']})
    >>> df
          animal
    0  alligator
    1        bee
    2     falcon
    3       lion
    4     monkey
    5     parrot
    6      shark
    7      whale
    8      zebra

    Viewing the first 5 lines

    >>> df.head()
          animal
    0  alligator
    1        bee
    2     falcon
    3       lion
    4     monkey

    Viewing the first `n` lines (three in this case)

    >>> df.head(3)
          animal
    0  alligator
    1        bee
    2     falcon

    For negative values of `n`

    >>> df.head(-3)
          animal
    0  alligator
    1        bee
    2     falcon
    3       lion
    4     monkey
    5     parrot
```

Since it says `n =5` we know that the default value of the parameter `n` is 5. When a function has a default value, we can call the function without a value.

To view the last few lines, we use `tail`

```
safi_df.tail()
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	roo
126	126	Ruaca	2017-05-18T00:00:00Z	3	7	burntbricks	
127	193	Ruaca	2017-06-04T00:00:00Z	7	10	cement	
128	194	Ruaca	2017-06-04T00:00:00Z	4	5	muddaub	
129	199	Chirodzo	2017-06-04T00:00:00Z	7	17	burntbricks	
130	200	Chirodzo	2017-06-04T00:00:00Z	8	20	burntbricks	

We can also get an `Index` for the columns of the DataFrame.

```
safi_df.columns
```

```
Index(['key_ID', 'village', 'interview_date', 'no_membrs', 'years_liv',
       'respondent_wall_type', 'rooms', 'memb_assoc', 'affect_conflicts',
       'liv_count', 'items_owned', 'no_meals', 'months_lack_food',
       'instanceID'],
      dtype='object')
```

`Index` is another data type that is defined by `pandas`.

an `Index` variable is iterable so we can index into it

Try it Yourself

How would you view the name of the 3rd column?

First the correct answer:

```
'interview_date'
```

Now some misconceptions:

```
safi_df['interview_date']
```

```
0      2016-11-17T00:00:00Z
1      2016-11-17T00:00:00Z
2      2016-11-17T00:00:00Z
3      2016-11-17T00:00:00Z
4      2016-11-17T00:00:00Z
                ...
126    2017-05-18T00:00:00Z
127    2017-06-04T00:00:00Z
128    2017-06-04T00:00:00Z
129    2017-06-04T00:00:00Z
130    2017-06-04T00:00:00Z
Name: interview_date, Length: 131, dtype: object
```

Indexing with the column name) will return the *values* in the column

```
safi_df.columns(2)
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-17-bd02c7e8a4a6> in <module>
----> 1 safi_df.columns(2)

TypeError: 'Index' object is not callable
```

Using `()` returns an error, because `columns` is an *attribute* which is referenced as is with no `()`. We get a type error because functions in python are objects of type `callable` and properties are values not functions.

```
pd.DataFrame.columns[2]
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-18-40e277f3074e> in <module>
----> 1 pd.DataFrame.columns[2]

TypeError: 'pandas._libs.properties.AxisProperty' object is not subscriptable
```

This doesn't work because `columns` is an attribute of an object of type `pandas.DataFrame` and `pd.DataFrame.columns` is not an object.

We can see what the type of pd.DataFrame is with the `type` function.

```
type(pd.DataFrame)
```

```
type
```

Knowing about types is helpful for the individual columns of a dataset as well.

```
safi_df.dtypes
```

```
key_ID                   int64
village                 object
interview_date          object
no_membrs                int64
years_liv                int64
respondent_wall_type    object
rooms                    int64
memb_assoc              object
affect_conflicts        object
liv_count                int64
items_owned             object
no_meals                 int64
months_lack_food        object
instanceID              object
dtype: object
```

Note that it uses `int64` and `object` as the types.

```
safi_df.head(2)
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1

We might want to look at what villages were included in the data.

```
pd.unique(safi_df['village'])
```

```
array(['God', 'Chirodzo', 'Ruaca'], dtype=object)
```

We can also get count of the number of of each value

```
safi_df['village'].value_counts()
```

```
Ruaca       49
God         43
Chirodzo    39
Name: village, dtype: int64
```

Try it Yourself!

how many surveyed farms have all type `mauddaub`?

Review and Further reading

- reading data with pandas
- Python built in functions and in particular the `type` function
- Pandas DataFrames
- `value_counts`

If you've made it this far, let me know how you found these notes.

Class 5: Accessing Data, continued

Today's agenda:

- warm up/ review
- announcements
- working with dataframes
- the power of functions as objects
- (maybe) exploratory data analysis

Try it out –>

Read the tables off of the syllabus course map page with `read_html` and make a list of the shapes of all of the tables on the page. Save the output to a variable and paste the *value* of that variable as your answer to the question.

```
import pandas as pd
[df.shape for df in
pd.read_html('https://rhodyprog4ds.github.io/BrownFall20/syllabus/course_map.html')]
```

```
[(14, 3), (15, 5), (15, 15), (15, 6)]
```

Announcements

- annotated notes are up
- beginning portfolio prompts and instructions are up
- Assignment due Sunday,
- office hours will remain Fridays
- TA office hours posted.

More Pandas

We'll go back to the SAFI dataset from Wednesday.

```
safi_df = pd.read_csv('https://raw.githubusercontent.com/brownsarahm/python-socialsci-files/master/data/SAFI_clean.csv')
```

We downloaded the data into memory, but we can also write it to disk.

```
safi_df.to_csv('safi_clean.csv')
```

It will go to the same folder as the notebook, but we can also use a relative path. If we make a `data` folder in the folder where we've saved the notebook, we can write the file there instead.

```
safi_df.to_csv('data/safi_clean.csv')
```

Now we can read it in using the same path

```
safi_df2= pd.read_csv('data/safi_clean.csv')
```

Note that now it has an extra column

```
safi_df2.head(2)
```

	Unnamed: 0	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_ty
0	0	1	God	2016-11-17T00:00:00Z	3	4	mudda
1	1	1	God	2016-11-17T00:00:00Z	7	9	mudda

```
safi_df.head(2)
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1

We can prevent this by writing it out with the `index` parameter set to `False`

```
safi_df.to_csv('data/safi_clean.csv',index=False)
```

Now when we read it in, there's no extra column.

```
safi_df3 = pd.read_csv('data/safi_clean.csv')
safi_df3.head(3)
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1
2	3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1

Recall, we indexed a column with the name in square brackets

Tip

In class we used the Jupyter GUI to create a new folder. You could also use your computer's default file management tool (Windows Explorer, Mac Finder, etc). Here, since the notebooks have to run completely automatically for this website, we use a ipython bash magic cell to make the folder. Jupyter notebooks use an ipython kernel.

Tip

False must be with a capital letter to be a boolean variable in python, as with True. You'll know if you did it right in your jupyter notebook, if the word terms bold and green.

```
safi_df['village']
```

```
0           God
1           God
2           God
3           God
4           God
         ...
126       Ruaca
127       Ruaca
128       Ruaca
129    Chirodzo
130    Chirodzo
Name: village, Length: 131, dtype: object
```

To index rows, we can use `loc`

```
safi_df.loc[3]
```

```
key_ID                                                                  4
village                                                               God
interview_date                                          2016-11-17T00:00:00Z
no_membrs                                                               7
years_liv                                                               6
respondent_wall_type                                            burntbricks
rooms                                                                   1
memb_assoc                                                            NaN
affect_conflicts                                                      NaN
liv_count                                                               2
items_owned             bicycle;radio;cow_plough;solar_panel;mobile_phone
no_meals                                                                2
months_lack_food                                          Sept;Oct;Nov;Dec
instanceID                      uuid:148d1105-778a-4755-aa71-281eadd4a973
Name: 3, dtype: object
```

To select a range, use `:`

```
safi_df.loc[3:5]
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
3	4	God	2016-11-17T00:00:00Z	7	6	burntbricks	1
4	5	God	2016-11-17T00:00:00Z	7	40	burntbricks	1
5	6	God	2016-11-17T00:00:00Z	3	3	muddaub	1

You only have to have a number on one side of the colon, it will go from the beginnig up to that number like this:

```
safi_df.loc[:4]
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1
2	3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1
3	4	God	2016-11-17T00:00:00Z	7	6	burntbricks	1
4	5	God	2016-11-17T00:00:00Z	7	40	burntbricks	1

With two `::` we can also set an increment

```
safi_df.loc[::5]
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	roo
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	
5	6	God	2016-11-17T00:00:00Z	3	3	muddaub	
10	11	God	2016-11-21T00:00:00Z	6	20	sunbricks	
15	16	God	2016-11-24T00:00:00Z	6	47	muddaub	
20	21	God	2016-11-21T00:00:00Z	8	20	burntbricks	
25	26	Ruaca	2016-11-21T00:00:00Z	3	20	burntbricks	
30	31	Ruaca	2016-11-21T00:00:00Z	3	2	muddaub	
35	36	Chirodzo	2016-11-17T00:00:00Z	6	23	sunbricks	
40	41	God	2016-11-17T00:00:00Z	7	22	muddaub	
45	46	Chirodzo	2016-11-17T00:00:00Z	10	42	burntbricks	
50	51	Chirodzo	2016-11-16T00:00:00Z	5	30	muddaub	
55	56	Chirodzo	2016-11-16T00:00:00Z	12	23	burntbricks	
60	61	Chirodzo	2016-11-16T00:00:00Z	10	14	muddaub	
65	66	Chirodzo	2016-11-16T00:00:00Z	10	37	burntbricks	
70	71	Ruaca	2016-11-18T00:00:00Z	6	14	burntbricks	
75	155	God	2016-11-24T00:00:00Z	4	4	burntbricks	
80	182	God	2016-11-25T00:00:00Z	7	21	muddaub	

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	roo
85	197	God	2016-11-28T00:00:00Z	5	19	burntbricks	
90	73	Ruaca	2017-04-26T00:00:00Z	7	9	burntbricks	
95	101	God	2017-04-27T00:00:00Z	3	4	muddaub	
100	104	Ruaca	2017-04-28T00:00:00Z	14	52	sunbricks	
105	113	Ruaca	2017-05-03T00:00:00Z	11	26	burntbricks	
110	108	God	2017-05-11T00:00:00Z	15	22	burntbricks	
115	150	Ruaca	2017-05-18T00:00:00Z	7	8	muddaub	
120	167	Ruaca	2017-06-03T00:00:00Z	8	24	muddaub	
125	192	Chirodzo	2017-06-03T00:00:00Z	9	20	burntbricks	
130	200	Chirodzo	2017-06-04T00:00:00Z	8	20	burntbricks	

These can be combined to index a subset at an increment.

We can index columns in two ways, as we did on Wednesday

```
safi_df['village'].head(2)
```

```
0    God
1    God
Name: village, dtype: object
```

Or using a `.`

```
safi_df.village.head(2)
```

```
0    God
1    God
Name: village, dtype: object
```

We can select multiple columns, using a `list` of column names. We can define the list inline.

```
safi_df[['village','no_membrs','years_liv']].head(2)
```

	village	no_membrs	years_liv
0	God	3	4
1	God	7	9

or in a separate variable

```
columns_of_interest = ['village','no_membrs','years_liv']
safi_df[columns_of_interest].head(2)
```

	village	no_membrs	years_liv
0	God	3	4
1	God	7	9

Functions are objects

```
syllabus_df_list =
pd.read_html('https://rhodyprog4ds.github.io/BrownFall20/syllabus/course_map.html')
```

And we can put them in a dictionary. `lambda` functions are special functions defined in a single line.

```
greetingl = lambda name: 'hello ' + name
greetingl('sarah')
```

```
'hello sarah'
```

is the same as

```
def greetingf(name):
    return 'hello ' + name
greetingf('sarah')
```

```
'hello sarah'
```

So, we can define a function in a dictionary like this:

```
view_rows = {0: lambda df: print(df.head()),
             1: lambda df: print(df.tail())}
```

Tip

this is how to do the equivalent of a switch or case in other languages in python

The `len` function works on all iterables

```
for df in syllabus_df_list:
    num_row = len(df)
    view_rows[num_row%2](df)
```

```
  Unnamed: 0_level_0                        topics  \
                week            Unnamed: 1_level_1
0                  1         [admin, python review]
1                  2    Loading data, Python review
2                  3      Exploratory Data Analysis
3                  4                 Data Cleaning
4                  5  Databases, Merging DataFrames

                           skills
                Unnamed: 2_level_1
0                         process
1     [access, prepare, summarize]
2           [summarize, visualize]
3  [prepare, summarize, visualize]
4   [access, construct, summarize]
   Unnamed: 0_level_0                                          skill  \
              keyword                                  Unnamed: 1_level_1
10           evaluate                         Evaluate model performance
11           optimize                          Optimize model parameters
12            compare                                     compare models
13       unstructured                            model unstructured data
14           workflow  use industry standard data science tools and w...

                                              Level 1  \
                                  Unnamed: 2_level_1
10  Explain basic performance metrics for differen...
11  Identify when model parameters need to be opti...
12                Qualitatively compare model classes
13  Identify options for representing text data an...
14  Solve well strucutred problems with a single t...
```

```
                                          Level 2  \
                                 Unnamed: 3_level_1
10  Apply basic model evaluation metrics to a held...
11  Manually optimize basic model parameters such ...
12  Compare model classes in specific terms and fi...
13  Apply at least one representation to transform...
14  Solve semi-strucutred, completely specified pr...

                                          Level 3
                                 Unnamed: 4_level_1
10  Evaluate a model with multiple metrics and cro...
11  Select optimal parameters based of mutiple qua...
12  Evaluate tradeoffs between different model com...
13  apply multiple representations and compare and...
14  Scope, choose an appropriate tool pipeline and...
   Unnamed: 0_level_0                A1                 A2  \
              keyword Unnamed: 1_level_1 Unnamed: 2_level_1
10           evaluate                 0                  0
11           optimize                 0                  0
12            compare                 0                  0
13       unstructured                 0                  0
14           workflow                 0                  0

                   A3                 A4                 A5  \
   Unnamed: 3_level_1 Unnamed: 4_level_1 Unnamed: 5_level_1
10                  0                  0                  0
11                  0                  0                  0
12                  0                  0                  0
13                  0                  0                  0
14                  0                  0                  0

                   A6                 A7                 A8  \
   Unnamed: 6_level_1 Unnamed: 7_level_1 Unnamed: 8_level_1
10                  0                  0                  0
11                  0                  0                  0
12                  0                  0                  0
13                  0                  0                  0
14                  0                  0                  0

                   A9                 A10                 A11  \
   Unnamed: 9_level_1 Unnamed: 10_level_1 Unnamed: 11_level_1
10                  0                   1                   1
11                  0                   1                   1
12                  0                   0                   1
13                  0                   0                   0
14                  0                   1                   1

                   A12                 A13       # Assignments
   Unnamed: 12_level_1 Unnamed: 13_level_1 Unnamed: 14_level_1
10                   0                   0                   2
11                   0                   0                   2
12                   0                   1                   2
13                   1                   1                   2
14                   1                   1                   4
   Unnamed: 0_level_0                                          Level 3  \
              keyword                                  Unnamed: 1_level_1
10           evaluate  Evaluate a model with multiple metrics and cro...
11           optimize  Select optimal parameters based of mutiple qua...
12            compare  Evaluate tradeoffs between different model com...
13       unstructured  apply multiple representations and compare and...
14           workflow  Scope, choose an appropriate tool pipeline and...

                   P1                 P2                 P3                 P4
   Unnamed: 2_level_1 Unnamed: 3_level_1 Unnamed: 4_level_1 Unnamed: 5_level_1
10                  0                  1                  1                  0
11                  0                  0                  1                  1
12                  0                  0                  1                  1
13                  0                  0                  1                  1
14                  0                  0                  1                  1
```

The beginning of Exploratory Data Analysis

Pandas will give us descriptive statistics

```
safi_df.describe()
```

	key_ID	no_membrs	years_liv	rooms	liv_count	no_meals
count	131.000000	131.00000	131.000000	131.000000	131.000000	131.000000
mean	85.473282	7.19084	23.053435	1.740458	2.366412	2.603053
std	63.151628	3.17227	16.913041	1.092547	1.082775	0.491143
min	1.000000	2.00000	1.000000	1.000000	1.000000	2.000000
25%	32.500000	5.00000	12.000000	1.000000	1.000000	2.000000
50%	66.000000	7.00000	20.000000	1.000000	2.000000	3.000000
75%	138.000000	9.00000	27.500000	2.000000	3.000000	3.000000
max	202.000000	19.00000	96.000000	8.000000	5.000000	3.000000

The statistics of the `key_ID` column don't make a lot of sense. We can avoid that by making it the index

```
safi_df.head()
```

	key_ID	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms
0	1	God	2016-11-17T00:00:00Z	3	4	muddaub	1
1	1	God	2016-11-17T00:00:00Z	7	9	muddaub	1
2	3	God	2016-11-17T00:00:00Z	10	15	burntbricks	1
3	4	God	2016-11-17T00:00:00Z	7	6	burntbricks	1
4	5	God	2016-11-17T00:00:00Z	7	40	burntbricks	1

the `inplace` parameter of a `pandas` functions applies the operation to the DataFrame in memory, but then the function returns nothing, but if we display after that, we see that now the `key_ID` column is now the index.

```
safi_df.set_index('key_ID',inplace=True)
safi_df.head(2)
```

	village	interview_date	no_membrs	years_liv	respondent_wall_type	rooms	me
key_ID							
1	God	2016-11-17T00:00:00Z	3	4	muddaub	1	
1	God	2016-11-17T00:00:00Z	7	9	muddaub	1	

and if we describe again, we see it doesn't compute on that column

```
safi_df.describe()
```

	no_membrs	years_liv	rooms	liv_count	no_meals
count	131.00000	131.000000	131.000000	131.000000	131.000000
mean	7.19084	23.053435	1.740458	2.366412	2.603053
std	3.17227	16.913041	1.092547	1.082775	0.491143
min	2.00000	1.000000	1.000000	1.000000	2.000000
25%	5.00000	12.000000	1.000000	1.000000	2.000000
50%	7.00000	20.000000	1.000000	2.000000	3.000000
75%	9.00000	27.500000	2.000000	3.000000	3.000000
max	19.00000	96.000000	8.000000	5.000000	3.000000

We can also call any of those on one column or one statistic.

```
safi_df['rooms'].mean()
```

```
1.7404580152671756
```

Pandas also has some built in plotting functions.

```
safi_df.plot.scatter('no_membrs','rooms')
```

```
<AxesSubplot:xlabel='no_membrs', ylabel='rooms'>
```

After Class Questions

How can we clean data? what other topics will we cover in this class?

How does the syntax on the question from prismia today work?

Where I can go to find a list of all the syntax.

More Practice

These additional questions are for if you want more practice with things we've done this week, before class next week.

Which of the following is a dictionary?

What type is the shape of a pandas DataFrame?

What does indexing with -1 do?

Further Reading

If you've made it this far, let me know how you found these notes.

Class 6: Exploratory Data Analysis

Warmup

```
topics = ['what is data science', 'jupyter', 'conditional','functions', 'lists',
'dictionaries','pandas' ]
topics
```

```
['what is data science',
 'jupyter',
 'conditional',
 'functions',
 'lists',
 'dictionaries',
 'pandas']
```

What happens when we index with `-1`?

```
topics[-1]
```

```
'pandas'
```

We get the last value.

Recall last class we used `:` to index `DataFrames` with `.loc`, we can do that with lists too:

```
topics[-2:]
```

```
['dictionaries', 'pandas']
```

Announcements

- next assignment will go up today
- try practingin throughout the week
- Check Brightspace for TA office hours

Why Exploratory Data Analysis (EDA) before cleaning?

Typically a Data process looks like one of these figures:

Data cleaning, which some of you asked about on Friday, would typically, *happen* before EDA, but it's hard to know what good data looks like, until you're used to manipulating it. Also, it's hard to check if data is good (and your cleaning worked) without some EDA. So, I'm choosing to *teach* EDA first. We'll do data cleaning next, with these tools from EDA available as options for checking and determining what cleaning to do.

EDA

```
# a version of the SAFI data with more columns
data_url = 'https://raw.githubusercontent.com/brownsarahm/python-socialsci-files/master/data/SAFI_full_shortname.csv'
```

As usual, we import `pandas`

```
import pandas as pd
```

and pull in the data.

```
safi_df = pd.read_csv(data_url)
safi_df.head(2)
```

	key_id	interview_date	quest_no	start	end	province	distric
0	1	17 November 2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Manic
1	2	17 November 2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Manic

2 rows × 65 columns

Note

In class, we used `%load http://drsmb.co/310` to pull in the url for the data. Load is an ipython (ther kernel thta Jupyter uses) magic function. It allows us to read in data from another place. I've set that short url to link to the download url for this hackmd as markdown(plain text). That way I can paste things and you can pull them directly into your notebook. We'll use it like a shared copy-paste.

Recall on Friday, we can use `describe` to see several common descriptive statistics

```
safi_df.describe()
```

	key_id	quest_no	years_farm	note2	no_membrs	members_count	years
count	131.000000	131.000000	131.000000	0.0	131.00000	131.00000	131.000
mean	66.000000	85.473282	15.832061	NaN	7.19084	7.19084	23.053
std	37.960506	63.151628	10.903883	NaN	3.17227	3.17227	16.913
min	1.000000	1.000000	1.000000	NaN	2.00000	2.00000	1.000
25%	33.500000	32.500000	8.000000	NaN	5.00000	5.00000	12.000
50%	66.000000	66.000000	15.000000	NaN	7.00000	7.00000	20.000
75%	98.500000	138.000000	20.500000	NaN	9.00000	9.00000	27.500
max	131.000000	202.000000	60.000000	NaN	19.00000	19.00000	96.000

8 rows × 25 columns

Then we remember that this includes `key_id` as a variable tht we don't actually want to treat as a variable, it's an index. We can change that with `set_index` and we can do it in memory (without assignment) using the `inplace` keyword.

```
safi_df.set_index('key_id',inplace=True)
safi_df.describe()
```

	quest_no	years_farm	note2	no_membrs	members_count	years_liv	respon
count	131.000000	131.000000	0.0	131.00000	131.00000	131.000000	
mean	85.473282	15.832061	NaN	7.19084	7.19084	23.053435	
std	63.151628	10.903883	NaN	3.17227	3.17227	16.913041	
min	1.000000	1.000000	NaN	2.00000	2.00000	1.000000	
25%	32.500000	8.000000	NaN	5.00000	5.00000	12.000000	
50%	66.000000	15.000000	NaN	7.00000	7.00000	20.000000	
75%	138.000000	20.500000	NaN	9.00000	9.00000	27.500000	
max	202.000000	60.000000	NaN	19.00000	19.00000	96.000000	

8 rows × 24 columns

Try it yourself!

You can use settings on `pd.read_csv` to set the index when the data is read in instead of doing this after the fact

We can also use the descriptive statistics on a single variable.

```
safi_df['years_farm'].describe()
```

```
count    131.000000
mean      15.832061
std       10.903883
min        1.000000
25%        8.000000
50%       15.000000
75%       20.500000
max       60.000000
Name: years_farm, dtype: float64
```

Note however that this is not well formatted, that's because it's a `Series` instead of a DataFrame

```
type(safi_df['years_farm'].describe())
```

```
pandas.core.series.Series
```

We can use `reset_index()` to make it back into a dataframe

```
safi_df['years_farm'].describe().reset_index()
```

	index	years_farm
0	count	131.000000
1	mean	15.832061
2	std	10.903883
3	min	1.000000
4	25%	8.000000
5	50%	15.000000
6	75%	20.500000
7	max	60.000000

And we can drop the added index:

```
safi_df['years_farm'].describe().reset_index().set_index('index')
```

	years_farm
index	
count	131.000000
mean	15.832061
std	10.903883
min	1.000000
25%	8.000000
50%	15.000000
75%	20.500000
max	60.000000

Note

See that we can chain operations together. This is a helpful feature, but to be used with care. Pep8, the style conventions for python, recommends no more than 79 characters per line. A summary and another summary

We can also use individual summary statistics on `DataFrame` or `Series` objects

```
safi_df['no_membrs'].max()
```

```
19
```

Tip

one column of a `pd.DataFrame` is a `pd.Series`

Split - Apply - Combine

A powerful tool in pandas (and data analysis in general) is the ability to apply functions to subsets of the data.

Note

In class, we used the load magic again here to get this url. Then split the cell, changed the second one to markdown with the code for the image

For example, we can get descriptive statistics per village

```
safi_df.groupby('village').describe()
```

	quest_no								years_farm	
	count	mean	std	min	25%	50%	75%	max	count	mean
village										
Chirodzo	39.0	62.487179	44.261705	8.0	44.5	55.0	64.5	200.0	39.0	17.28205
God	43.0	81.720930	77.863839	1.0	14.5	40.0	168.5	202.0	43.0	14.88372
Ruaca	49.0	107.061224	55.024013	23.0	72.0	113.0	152.0	194.0	49.0	15.51020

3 rows × 192 columns

We can also rearrange this into a more usable format than this very wide format:

```
safi_df.groupby('village').describe().unstack().reset_index()
```

	level_0	level_1	village	0
0	quest_no	count	Chirodzo	39.000000
1	quest_no	count	God	43.000000
2	quest_no	count	Ruaca	49.000000
3	quest_no	mean	Chirodzo	62.487179
4	quest_no	mean	God	81.720930
...	...	...	...	...
571	gps_Accuracy	75%	God	13.500000
572	gps_Accuracy	75%	Ruaca	12.000000
573	gps_Accuracy	max	Chirodzo	39.000000
574	gps_Accuracy	max	God	30.000000
575	gps_Accuracy	max	Ruaca	2099.999000

576 rows × 4 columns

This, however, gives some funny variable names, to fix that, we first save it to a variable.

```
village_summary_df = safi_df.groupby('village').describe().unstack().reset_index()
```

Now we can use the rename function

```
help(village_summary_df.rename)
```

Tip

To include an image in a notebook, use:

```
![alt text here]
(url/or/path/to/image
)
```

in a markdown cell

Tip

Rename is also one thing you might do in cleaning a dataset.

```
Help on method rename in module pandas.core.frame:
```

```
Help on method rename in module pandas.core.frame:

rename(mapper=None, index=None, columns=None, axis=None, copy=True, inplace=False,
level=None, errors='ignore') method of pandas.core.frame.DataFrame instance
    Alter axes labels.

    Function / dict values must be unique (1-to-1). Labels not contained in
    a dict / Series will be left as-is. Extra labels listed don't throw an
    error.

    See the :ref:`user guide <basics.rename>` for more.

    Parameters
    ----------
    mapper : dict-like or function
        Dict-like or function transformations to apply to
        that axis' values. Use either ``mapper`` and ``axis`` to
        specify the axis to target with ``mapper``, or ``index`` and
        ``columns``.
    index : dict-like or function
        Alternative to specifying axis (``mapper, axis=0``
        is equivalent to ``index=mapper``).
    columns : dict-like or function
        Alternative to specifying axis (``mapper, axis=1``
        is equivalent to ``columns=mapper``).
    axis : {0 or 'index', 1 or 'columns'}, default 0
        Axis to target with ``mapper``. Can be either the axis name
        ('index', 'columns') or number (0, 1). The default is 'index'.
    copy : bool, default True
        Also copy underlying data.
    inplace : bool, default False
        Whether to return a new DataFrame. If True then value of copy is
        ignored.
    level : int or level name, default None
        In case of a MultiIndex, only rename labels in the specified
        level.
    errors : {'ignore', 'raise'}, default 'ignore'
        If 'raise', raise a `KeyError` when a dict-like `mapper`, `index`,
        or `columns` contains labels that are not present in the Index
        being transformed.
        If 'ignore', existing keys will be renamed and extra keys will be
        ignored.

    Returns
    -------
    DataFrame or None
        DataFrame with the renamed axis labels or None if ``inplace=True``.

    Raises
    ------
    KeyError
        If any of the labels is not found in the selected axis and
        "errors='raise'".

    See Also
    --------
    DataFrame.rename_axis : Set the name of the axis.

    Examples
    --------
    ``DataFrame.rename`` supports two calling conventions

    * ``(index=index_mapper, columns=columns_mapper, ...)``
    * ``(mapper, axis={'index', 'columns'}, ...)``

    We *highly* recommend using keyword arguments to clarify your
    intent.

    Rename columns using a mapping:

    >>> df = pd.DataFrame({"A": [1, 2, 3], "B": [4, 5, 6]})
    >>> df.rename(columns={"A": "a", "B": "c"})
       a  c
    0  1  4
    1  2  5
    2  3  6

    Rename index using a mapping:

    >>> df.rename(index={0: "x", 1: "y", 2: "z"})
       A  B
    x  1  4
    y  2  5
    z  3  6

    Cast index labels to a different type:

    >>> df.index
```

```
RangeIndex(start=0, stop=3, step=1)
>>> df.rename(index=str).index
Index(['0', '1', '2'], dtype='object')

>>> df.rename(columns={"A": "a", "B": "b", "C": "c"}, errors="raise")
Traceback (most recent call last):
KeyError: ['C'] not found in axis

Using axis-style parameters:

>>> df.rename(str.lower, axis='columns')
   a  b
0  1  4
1  2  5
2  3  6

>>> df.rename({1: 2, 2: 4}, axis='index')
   A  B
0  1  4
2  2  5
4  3  6
```

Rename takes a dict-like or function that maps from what's there to what to change it to. We can either provide the mapper and an axis or pass the mapper as the columns parameter. The columns parameter makes it more human readable, and explicit what we're doing.

```
renaming_dict = {'level_0':'variable',
                 'level_1':'statistic',
                 0:'value'}
village_summary_df.rename(columns= renaming_dict,inplace=True)
```

`inplace` again. See questions at the bottom for more on what it does

Now we have a much better summary table. This is in what's called tidy format. Each colum

```
village_summary_df.head()
```

	variable	statistic	village	value
0	quest_no	count	Chirodzo	39.000000
1	quest_no	count	God	43.000000
2	quest_no	count	Ruaca	49.000000
3	quest_no	mean	Chirodzo	62.487179
4	quest_no	mean	God	81.720930

How could we use this to compute the average across villages for each statistic of each variable?

> **Try it yourself!**
>
> What would it look like to do this from the first result of `describe()` on the `groupby`, instead of with the `unstack` and `reset_index`?

We *do* need to `groupby 'statistic'` and use `mean`, but that's not enough:

```
village_summary_df.groupby('statistic').mean()
```

	value
statistic	
25%	44.807522
50%	48.428030
75%	54.062308
count	30.388889
max	105.468983
mean	49.868824
min	13.756913
std	24.588967

This averages across all of the variables, too. So instead we need to groupby two variables.

```
village_summary_df.groupby(['statistic','variable']).mean()
```

statistic	variable	value
25%	buildings_in_compound	1.000000
	gps_Accuracy	8.000000
	gps_Altitude	691.833333
	gps_Latitude	-19.112221
	gps_Longitude	33.480944
...	...	...
std	respondent_wall_type_other	NaN
	rooms	1.082409
	years_farm	10.832121
	years_liv	16.549343
	yes_group_count	1.074600

192 rows × 1 columns

```
safi_df.columns
```

```
Index(['interview_date', 'quest_no', 'start', 'end', 'province', 'district',
       'ward', 'village', 'years_farm', 'agr_assoc', 'note2', 'no_membrs',
       'members_count', 'remittance_money', 'years_liv', 'parents_liv',
       'sp_parents_liv', 'grand_liv', 'sp_grand_liv', 'respondent_roof_type',
       'respondent_wall_type', 'respondent_wall_type_other',
       'respondent_floor_type', 'window_type', 'buildings_in_compound',
       'rooms', 'other_buildings', 'no_plots', 'plots_count', 'water_use',
       'no_group_count', 'yes_group_count', 'no_enough_water',
       'months_no_water', 'period_use', 'exper_other', 'other_meth',
       'res_change', 'memb_assoc', 'resp_assoc', 'fees_water',
       'affect_conflicts', 'note', 'need_money', 'money_source',
       'money_source_other', 'crops_contr', 'emply_lab', 'du_labour',
       'liv_owned', 'liv_owned_other', 'liv_count', 'poultry',
       'du_look_aftr_cows', 'items_owned', 'items_owned_other', 'no_meals',
       'months_lack_food', 'no_food_mitigation', 'gps_Latitude',
       'gps_Longitude', 'gps_Altitude', 'gps_Accuracy', 'instanceID'],
      dtype='object')
```

What is the most common combination of `respondent_wall_type` and `respondent_floor_type`?

```
safi_df.groupby(['respondent_wall_type','respondent_floor_type'])
['instanceID'].count().reset_index()
```

	respondent_wall_type	respondent_floor_type	instanceID
0	burntbricks	cement	30
1	burntbricks	earth	37
2	cement	cement	1
3	muddaub	cement	3
4	muddaub	earth	43
5	sunbricks	cement	4
6	sunbricks	earth	13

We can read this table to see the result. Or we might want it in a different way

```
safi_df.groupby(['respondent_wall_type','respondent_floor_type'])
['instanceID'].count().unstack()
```

respondent_floor_type	cement	earth
respondent_wall_type		
burntbricks	30.0	37.0
cement	1.0	NaN
muddaub	3.0	43.0
sunbricks	4.0	13.0

Questions After Class

What does the inplace parameter do?

We used inplace for the first time today right after we read in the data. Let's make a new DataFrame and compare what happens with and without.

```
safi_df_noinplace = pd.read_csv(data_url)
safi_df_inplace = pd.read_csv(data_url)
```

```
type(safi_df_noinplace.set_index('key_id'))
```

```
pandas.core.frame.DataFrame
```

```
safi_df_noinplace.head()
```

	key_id	interview_date	quest_no	start	end	province	distri
0	1	17 November 2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Manic
1	2	17 November 2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Manic
2	3	17 November 2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	Manica	Manic
3	4	17 November 2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	Manica	Manic
4	5	17 November 2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	Manica	Manic

5 rows × 65 columns

```
type(safi_df_inplace.set_index('key_id',inplace=True))
```

```
NoneType
```

```
safi_df_inplace.head()
```

Tip

I'm surprised I've remembered it repeatedly in class. I usually forget `inplace` don't get the output I want and then go back and add it. That's a normal part of programming.

	interview_date	quest_no	start	end	province	district
key_id						
1	17 November 2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Manica
2	17 November 2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Manica
3	17 November 2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	Manica	Manica
4	17 November 2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	Manica	Manica
5	17 November 2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	Manica	Manica

5 rows × 64 columns

Without `inplace`, `set_index` returns a dataframe, with the index changed, with `inplace` it returns `None`. But, without `inplace`, we don't see the desired effect when we do our next step, but with `inplace` the index is changed.

Without the inplace, to save the changes you need to use an assignment. The following is equivalent to using `inplace`.

```
safi_df_noinplace= safi_df_noinplace.set_index('key_id'))
safi_df_noinplace.head()
```

```
  File "<ipython-input-30-1a84c65392b3>", line 1
    safi_df_noinplace= safi_df_noinplace.set_index('key_id'))
                                                            ^
SyntaxError: invalid syntax
```

Why is a `for` loop slower than a `pandas` operation?

TL;DR: for this example, it's actually possible to relatively easily write a faster loop, but for other operations, it's unlikely.

Basically, the advantage of using the library functions is that someone has already put a lot of though into the optimal way to implement things and it leverages core features of the data structure. Here is a blog post about iterating and applying functions in pandas. It covers topics we haven't yet seen, but will.
The `pandas` implementation is also fewer lines and can scale for large datasets easily.

> **Note**
>
> For example using modin that allows you to change 1 line of code (`import pandas as pd` to `import modin.pandas as pd`) to leverage parallelized computation. Base python loops won't do that.

To test, we can do an experiment. Let's take the last thing we covered in class today, finding the most common combination of floor and wall types.

First we'll look at two solutions and verify that they're the same.

First, how to do it with `pandas`

```
wall_floor_df = safi_df.groupby(['respondent_wall_type','respondent_floor_type'])
['instanceID'].count()
wall_floor_df
```

```
respondent_wall_type  respondent_floor_type
burntbricks           cement                  30
                      earth                   37
cement                cement                   1
muddaub               cement                   3
                      earth                   43
sunbricks             cement                   4
                      earth                   13
Name: instanceID, dtype: int64
```

We can read the answer off of that table, but let's get it programmatically:

```
wall_floor_df.idxmax()
```

```
('muddaub', 'earth')
```

Next, how to do it with a for loop (if you have a better for loop, make a PR so share it).

```
wall_floor_dict = {}

for wall,floor  in
zip(safi_df['respondent_wall_type'],safi_df['respondent_floor_type']):
    wf_key = '_'.join([wall,floor])
    if wf_key in wall_floor_dict.keys():
        wall_floor_dict[wf_key] +=1
    else:
        wall_floor_dict[wf_key] =1

wall_floor_dict
```

```
{'muddaub_earth': 43,
 'burntbricks_cement': 30,
 'burntbricks_earth': 37,
 'sunbricks_earth': 13,
 'muddaub_cement': 3,
 'sunbricks_cement': 4,
 'cement_cement': 1}
```

Again, we can read from the dict, but lets find it

```
max(wall_floor_dict, key=wall_floor_dict.get).split('_')
```

```
['muddaub', 'earth']
```

Now we can use a special feature of Jupyter notebooks to time them and see which is faster, called the `%timeit` magic. Displaying visual things is always slow, so we'll take the line that would display out of both options.

Now timing the pandas way, with a human deciding

```
%%timeit -o

wall_floor_df = safi_df.groupby(['respondent_wall_type','respondent_floor_type'])
['instanceID'].count().reset_index()
```

```
2.31 ms ± 60.8 µs per loop (mean ± std. dev. of 7 runs, 100 loops each)
```

```
<TimeitResult : 2.31 ms ± 60.8 µs per loop (mean ± std. dev. of 7 runs, 100 loops
each)>
```

```
t_pandas = _
```

And capture the time by using the _ it gets the std out from the previous cell.

For loop with required interpretation

```
%%timeit -o
wall_floor_dict = {}

for wall,floor  in
zip(safi_df['respondent_wall_type'],safi_df['respondent_floor_type']):
    wf_key = '_'.join([wall,floor])
    if wf_key in wall_floor_dict.keys():
        wall_floor_dict[wf_key] +=1
    else:
        wall_floor_dict[wf_key] =1

max(wall_floor_dict, key=wall_floor_dict.get).split('_')
```

Tip

This timing can be good when you have a large dataset, you can load a small part and and compare two ways of doing something you want to do with that. Then use only the faster on the whole dataset.

More detail on timing and profiling from the text book

```
87.1 µs ± 801 ns per loop (mean ± std. dev. of 7 runs, 10000 loops each)
```

```
<TimeitResult : 87.1 µs ± 801 ns per loop (mean ± std. dev. of 7 runs, 10000 loops
each)>
```

```
t_loop = _
```

On my local computer it was about 12x faster with that loop. Here, we'll calculate it for the GitHub servers that host the course manual.

```
t_pandas.average/t_loop.average
```

```
26.57731776507862
```

For a more a complex question (say we wanted to know a statistic more complicated than the count) the loop gets harder (and probably slower), but the pandas operation stays about the same.

If you've made it this far, let me know how you found these notes.

Tip

You can also try it on your computer and see how it compares. Or try other loops/ways of doing it in pandas.

Class 7: Visualization for EDA

Announcements

Syllabus updated

1. rubric for summarize and visualize are slightly changed
2. Please accept assignments if you plan to not complete for any reason

Assignment updated to clarify continuous and categorical variables

Loading Data

Importing the libraries for today. We'll continue plotting with pandas and we'll use `seaborn` as well. Seaborn provides higher level plotting functions and better formatting.

```
import pandas as pd
import seaborn as sns
```

The alias for `seaborn` is `sns` the result of an inside joke among the developers in reference so Samuel Norman Seaborn on The West Wing, per stackexchange

Loading the data as usual.

```
data_url = 'https://raw.githubusercontent.com/brownsarahm/python-socialsci-
files/master/data/SAFI_full_shortname.csv'
```

We know that the `key_id` column should be used as an index, not as data, so we'll use the `index_col` parameter t do that from the beginning.

```
safi_df = pd.read_csv(data_url,index_col='key_id')
safi_df.head()
```

key_id	interview_date	quest_no	start	end	province	district
1	17 November 2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Manica
2	17 November 2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Manica
3	17 November 2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	Manica	Manica
4	17 November 2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	Manica	Manica
5	17 November 2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	Manica	Manica

5 rows × 64 columns

We can make scatter plots as we saw Friday.

```
safi_df.plot.scatter('gps_Latitude','gps_Longitude')
```

```
<AxesSubplot:xlabel='gps_Latitude', ylabel='gps_Longitude'>
```

With seaborn, however, we can control it more, changing the color of the points based on a column of the data.

```
sns.relplot(x= 'gps_Latitude',y='gps_Longitude',
            data=safi_df, hue='village')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f05c06aea50>
```

We can also plot a single variable to see the quantiles (the box is 25%-50%) and see if there are outliers (the points outside the box).

```
safi_df.boxplot('years_farm')
```

```
<AxesSubplot:>
```

We can do some more conditiioning, even with only pandas. Using the `by` parameter will do a `groupby` operation first and then make the plot.

```
safi_df.boxplot('years_farm', by='village')
```

```
<AxesSubplot:title={'center':'years_farm'}, xlabel='village'>
```

We can also make the figure larger

```
safi_df.boxplot('years_farm', by='village',figsize=(12,8))
```

```
<AxesSubplot:title={'center':'years_farm'}, xlabel='village'>
```

We can see how a single variable is distributed in more detail with the [seaborn displot](#).

```
sns.displot(data=safi_df, x='years_farm',col='village')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f05c0326c50>
```

```
sns.displot(data=safi_df, x='years_farm',col='village',kind='kde')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f05c032d150>
```

Updating Seaborn

the `displot` is new in seaborn 0.11, on your terminal (mac, linux) or anaconda prompt(Windows):

```
pip install update seaborn
```

Then restart your notebook's kernel and re-run all cells

or, in a notebook, you can update with

```
!pip install seaborn==0.11.0
```

This is set to not run on the served site

Regression plots

We can also plot with some calculations done for us already.

```
sns.lmplot(x='years_farm',y='years_liv', data=safi_df,
           hue='village')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f05bde67e50>
```

And we can make grids of plots with the `row` and `col` parameters. We can turn off the regression lines with the `fit_reg` parameter.

```
sns.lmplot(x='years_farm',y='years_liv', data=safi_df,
           hue='village',  col='respondent_wall_type',row='respondent_floor_type',
           fit_reg=False)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f05c017bfd0>
```

```
safi_df.groupby('poultry')['members_count'].mean()
```

```
poultry
no     5.894737
yes    7.720430
Name: members_count, dtype: float64
```

Questions after class

More practice

Test out the parameters of the plotting functions to see what they do.

Further Reading

- Pandas Plotting
- Seaborn Gallery
- Seaborn Tutorial

If you've made it this far, let me know how you found these notes.

Class 8: Visualization and Starting to Clean Data

Announcements

1. closing notebooks
    - recall that they don't stop when you close the tabs
    - you need to stop it at the terminal it launched from
    - with: ctr + c as it says when you first launch a notebook
2. restart and rerun notebooks
    - notebooks are a continuous REPL as long as you have it open
    - if you change code, you could have, for example a variable that is no longer defined in the notebook as written, but that still exists in memory, so code that depends on it will still run, for now, but not after you restart next (eg if we run it while grading)
    - to check, restart and rerun your notebook
3. say hello on zoom for attendance

Setup

First we `import` packages and load data as normal.

```
import pandas as pd
import seaborn as sns

data_url = 'https://raw.githubusercontent.com/brownsarahm/python-socialsci-files/master/data/SAFI_full_shortname.csv'
sns.set(font_scale=1.25)
```

I've added one new command here, Seaborn's []`set` function](https://seaborn.pydata.org/generated/seaborn.set.html). It sets a bunch of theme aspects for plotting, here I used it to increase the font size.

Warmup Activity

How recreate this plot?

First, load the data

```
safi_df = pd.read_csv(data_url)
```

This is a `displot`.

```
sns.displot(data=safi_df, x='members_count',row='parents_liv',
            col = 'sp_parents_liv',hue='village',kind='kde')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f710ba0c690>
```

This allows comparisons of how the village, respondent's parents living (`parents_liv`), and the spouse's (`sp_parents_liv`) influence the number of members of the household (`members_count`). The `kind='kde'` parameter uses the underlying `kdeplot` to apply kernel density estimation

We can understand better what this does, by comparing what happens when we take it out.

```
sns.displot(data=safi_df, x='members_count',row='parents_liv',
            col = 'sp_parents_liv',hue='village')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f70cab6fdd0>
```

Tip

This is a general strategy. Testing out parameters with different values is a valuable way to learn what they do, and to build intuition about what they do. The documentation also has demos of a lot of values.

What about the rest of the columns?

We've used this SAFI dataset a lot, but we've only used a few of the many columns. Let's look at all of them.

```
safi_df.columns
```

```
Index(['key_id', 'interview_date', 'quest_no', 'start', 'end', 'province',
       'district', 'ward', 'village', 'years_farm', 'agr_assoc', 'note2',
       'no_membrs', 'members_count', 'remittance_money', 'years_liv',
       'parents_liv', 'sp_parents_liv', 'grand_liv', 'sp_grand_liv',
       'respondent_roof_type', 'respondent_wall_type',
       'respondent_wall_type_other', 'respondent_floor_type', 'window_type',
       'buildings_in_compound', 'rooms', 'other_buildings', 'no_plots',
       'plots_count', 'water_use', 'no_group_count', 'yes_group_count',
       'no_enough_water', 'months_no_water', 'period_use', 'exper_other',
       'other_meth', 'res_change', 'memb_assoc', 'resp_assoc', 'fees_water',
       'affect_conflicts', 'note', 'need_money', 'money_source',
       'money_source_other', 'crops_contr', 'emply_lab', 'du_labour',
       'liv_owned', 'liv_owned_other', 'liv_count', 'poultry',
       'du_look_aftr_cows', 'items_owned', 'items_owned_other', 'no_meals',
       'months_lack_food', 'no_food_mitigation', 'gps_Latitude',
       'gps_Longitude', 'gps_Altitude', 'gps_Accuracy', 'instanceID'],
      dtype='object')
```

We can look at the first few row to recall what sort of data are available in each column.

```
safi_df.head()
```

	key_id	interview_date	quest_no	start	end	province	distric
0	1	17 November 2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Manic
1	2	17 November 2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Manic
2	3	17 November 2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	Manica	Manic
3	4	17 November 2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	Manica	Manic
4	5	17 November 2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	Manica	Manic

5 rows × 65 columns

It might be interesting to use the `items_owned` column to compare find out if there are differences between farms based on what items they have.

Maybe we'd want to see how many farms own each item, maybe we want to know how many farms own bicycle. Let's try `value_counts` out like we used before.

```
safi_df['items_owned'].value_counts()
```

```
['radio']
4
['mobile_phone']
4
['bicycle' ;  'radio' ;  'cow_plough' ;  'solar_panel' ;  'table' ;  'mobile_phone']
3
['bicycle' ;  'radio' ;  'mobile_phone']
3
['bicycle' ;  'radio' ;  'cow_plough' ;  'solar_panel' ;  'solar_torch' ;
'mobile_phone']                                          3

..
['radio' ;  'cow_plough' ;  'table' ;  'mobile_phone']
1
['motorcyle' ;  'television' ;  'radio' ;  'cow_plough' ;  'solar_panel' ;
'mobile_phone']                                         1
['radio' ;  'table' ;  'sofa_set' ;  'mobile_phone']
1
['cow_cart' ;  'motorcyle' ;  'bicycle' ;  'television' ;  'radio' ;  'cow_plough' ;
'solar_panel' ;  'solar_torch' ;  'table']    1
['motorcyle' ;  'bicycle' ;  'radio' ;  'table' ;  'mobile_phone']
1
Name: items_owned, Length: 95, dtype: int64
```

The problem is this gives the count of each *list* of items instead of each item. What we'd want instead is one column for each item, where the values in the column are 1 when the farm owns that item and 0 when they don't. This representation is sometimes called 1 hot encoding and other times (including in pandas) it's called dummy variables.

Using Dummy Variables

Let's try this function out, first on the `village` column.

```
pd.get_dummies(data= safi_df,columns=['village'])
```

	key_id	interview_date	quest_no	start	end	province	dis
0	1	17 November 2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Ma
1	2	17 November 2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Ma
2	3	17 November 2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	Manica	Ma
3	4	17 November 2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	Manica	Ma
4	5	17 November 2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	Manica	Ma
...	...	...	...	...	...	...	
126	127	18 May 2017	126	2017-05-18T04:13:37.000Z	2017-05-18T04:35:47.000Z	Manica	Ma
127	128	04 June 2017	193	2017-06-04T09:36:20.000Z	2017-06-04T10:13:32.000Z	Manica	Ma
128	129	04 June 2017	194	2017-06-04T10:13:36.000Z	2017-06-04T10:32:06.000Z	Manica	Ma
129	130	04 June 2017	199	2017-06-04T10:33:55.000Z	2017-06-04T10:52:22.000Z	Manica	Ma
130	131	04 June 2017	200	2017-06-04T10:52:46.000Z	2017-06-04T11:08:13.000Z	Manica	Ma

131 rows × 67 columns

Instead of scrolling, we can isolate the new columns we just created.

```
safi_df_villages = pd.get_dummies(data= safi_df,columns=['village'])
keep_cols = [col for col in safi_df_villages.columns if 'village_' in col]
safi_df_villages[keep_cols]
```

	village_Chirodzo	village_God	village_Ruaca
0	0	1	0
1	0	1	0
2	0	1	0
3	0	1	0
4	0	1	0
...	...	...	...
126	0	0	1
127	0	0	1
128	0	0	1
129	1	0	0
130	1	0	0

131 rows × 3 columns

Note that now we have 3 columns, one for each village. This column was already usable, but it was well formatted an useful to illustrate what `get_dummies` does.

> **Note**
>
> try this out and compare the two, test out `get_dummies` on another column to be sure you know what it does.

Now, we can try it on `items_owned`. _this time we'll filter the columns for display and inspection right away

```
safi_df_items = pd.get_dummies(data= safi_df,columns=['items_owned'])
keep_cols = [col for col in safi_df_items.columns if 'village_' in col]
safi_df_items[keep_cols]
```

0

1

2

3

4

...

126

127

128

129

130

131 rows × 0 columns

This still isn't quite what we want, because each value in the `items_owned` column *looks* like a list but it's actually a string, we can check that with the `type` function.

```
type(safi_df['items_owned'][0])
```

```
str
```

Cleaning One Cell of Data

So, let's clean the value from one cell of that column and see what else is going on. First, we save it to a variable.

```
farm1_items = safi_df['items_owned'][0]
```

Let's look at it first.

```
farm1_items
```

```
"['bicycle' ;  'television' ;  'solar_panel' ;  'table']"
```

We could try first casting it to a list.

```
list(farm1_items)
```

```
['[',
 "'",
 'b',
 'i',
 'c',
 'y',
 'c',
 'l',
 'e',
 "'",
 ' ',
 ';',
 ' ',
 ' ',
 "'",
 't',
 'e',
 'l',
 'e',
 'v',
 'i',
 's',
 'i',
 'o',
 'n',
 "'",
 ' ',
 ';',
 ' ',
 ' ',
 "'",
 's',
 'o',
 'l',
 'a',
 'r',
 '_',
 'p',
 'a',
 'n',
 'e',
 'l',
 "'",
 ' ',
 ';',
 ' ',
 ' ',
 "'",
 't',
 'a',
 'b',
 'l',
 'e',
 "'",
 ']']
```

That doesn't quite do it though, let's try separating it with the split method at the ';'.

```
farm1_items.split(';')
```

```
["['bicycle ', "  'television ', "  'solar_panel ', "  'table']"]
```

This still has some extra characters, in it though. We should remove those, probably before we do the split so that we don't end up with empty items in the list.

```
farm1_items.replace('[','').replace(']','').replace("'","")
```

```
'bicycle ;  television ;  solar_panel ;  table'
```

That is now ready to split.

```
farm1_items.replace('[','').replace(']','').replace("'","").split(';')
```

```
['bicycle ', '  television ', '  solar_panel ', '  table']
```

There are some empty spaces lest. Python has a string function `strip` that removes leading and trailing (on the ends) whitespace (spaces, tabs, new lines), we have to apply it to each individual item of that list above. We can use a list comprehension for this.

```
[i.strip() for i in
farm1_items.replace('[','').replace(']','').replace("'","").split(';')]
```

```
['bicycle', 'television', 'solar_panel', 'table']
```

Applying this to the rest of the data

We can put that in a function so that we can reuse it.

```
def separate_items(row):
    '''
    clean the items owned column of one row of the safi_df
    '''

    return [i.strip() for i in
row['items_owned'].replace('[','').replace(']','').replace("'","").split(';')]
```

Pandas provides us special function to apply a function to a dataframe along a given axis.

```
safi_df.apply(separate_items,axis=1)
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-20-26c2d810f773> in <module>
----> 1 safi_df.apply(separate_items,axis=1)

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/frame.py in apply(self, func, axis, raw, result_type, args,
**kwds)
   7766             kwds=kwds,
   7767         )
-> 7768         return op.get_result()
   7769
   7770     def applymap(self, func, na_action: Optional[str] = None) -> DataFrame:

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/apply.py in get_result(self)
    183             return self.apply_raw()
    184
--> 185         return self.apply_standard()
    186
    187     def apply_empty_result(self):

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/apply.py in apply_standard(self)
    274
    275     def apply_standard(self):
--> 276         results, res_index = self.apply_series_generator()
    277
    278         # wrap results

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/apply.py in apply_series_generator(self)
    288             for i, v in enumerate(series_gen):
    289                 # ignore SettingWithCopy here in case the user mutates
--> 290                 results[i] = self.f(v)
    291                 if isinstance(results[i], ABCSeries):
    292                     # If we have a view on v, we need to make a copy because

<ipython-input-19-f7c175567037> in separate_items(row)
      4     '''
      5
----> 6     return [i.strip() for i in
row['items_owned'].replace('[','').replace(']','').replace("'","").split(';')]
      7

AttributeError: 'float' object has no attribute 'replace'
```

This error means that there are some elements of that column that is a float instead of a string. We should check what that might be. We can check with a list comprehension and look at the ones that are `float`

```
[item_list for item_list in safi_df['items_owned'] if type(item_list)==float]
```

```
[nan, nan, nan, nan, nan, nan, nan, nan, nan, nan]
```

Since the floats are `nan`, we know that our function more or less is the right thing to do, but we need to modify our function to accommodate those values. We can replace them with an empty list.

```
def separate_items(row):
    '''
    '''
    if type(row['items_owned'])==str:
        return [i.strip() for i in
row['items_owned'].replace('[','').replace(']','').replace("'","").split(';')]
    else:
        return []
```

> **Tip**
>
> If this were a regular analysis and not a tutorial, we would probably just edit the cell above, but for the purpose of making this tutorial more readable, we'll copy the function from above, into another cell, edit that cell, and overwrite in python memory by leaving it the same function name.
>
> If instead we wanted to be able to compare the two functions, we'd give the second one a different name.

We can try this again:

```
safi_df['items_owned_clean'] = safi_df.apply(separate_items,axis=1)
```

it runs successfully and we can look at the output.

```
safi_df[['items_owned_clean']].head()
```

	items_owned_clean
0	[bicycle, television, solar_panel, table]
1	[cow_cart, bicycle, radio, cow_plough, solar_p...
2	[solar_torch]
3	[bicycle, radio, cow_plough, solar_panel, mobi...
4	[motorcyle, radio, cow_plough, mobile_phone]

This looks all good and we'll pick up from here on Monday.

Class 9: Preparing Data For Analysis

- Say hello in the zoom chat
- join Prismia

Checking types is an important part of cleaning data, we need to figure out what is wrong with data before we can fix it.

Remember that data cleaning is a lot of exploration and iteration in the data. Let's review a few data types we've seen.

Warmup: type review

Lists are enclosed by square brackets(`[]`), they're ordered and iterable.

```
type([char for char in 'abcde'])
```

```
list
```

Dictionaries are comprised of `key:value` pairs and enclosed in curly brackets (`{}`) they're iterable and indexable by the keys.

```
type({char:i for i, char in enumerate('abcde')})
```

```
dict
```

This is a tuple, this data type is used to tie together items, but not usually to iterate over. It's used for pairs of things often or groups to iterate over multiple things at once. For example with the zip function.

```
type(('a','b','c','d','e'))
```

```
tuple
```

Splitting a string makes a list.

```
type('a b c d e'.split(' '))
```

```
list
```

Loading Data

Loading the data and packages as normal.

```
# %load http://drsmb.co/310
import pandas as pd
import seaborn as sns
```

```
safi_url = 'https://raw.githubusercontent.com/brownsarahm/python-socialsci-files/master/data/SAFI_full_shortname.csv'
safi_df = pd.read_csv(safi_url)
```

Exploring Types

On Friday we figured out that the following function would clean one cell of the `safi_df['items_owned']` column.

```
def separate_items(row):
    '''
    '''

    return [i.strip() for i in row['items_owned'].replace('[','').replace(']','').replace("'","").split(';')]
```

But when we applied it, we got an error.

```
safi_df.apply(separate_items,axis=1)
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-8-26c2d810f773> in <module>
----> 1 safi_df.apply(separate_items,axis=1)

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/frame.py in apply(self, func, axis, raw, result_type, args,
**kwds)
   7766             kwds=kwds,
   7767         )
-> 7768         return op.get_result()
   7769
   7770     def applymap(self, func, na_action: Optional[str] = None) -> DataFrame:

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/apply.py in get_result(self)
    183             return self.apply_raw()
    184
--> 185         return self.apply_standard()
    186
    187     def apply_empty_result(self):

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/apply.py in apply_standard(self)
    274
    275     def apply_standard(self):
--> 276         results, res_index = self.apply_series_generator()
    277
    278         # wrap results

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/core/apply.py in apply_series_generator(self)
    288             for i, v in enumerate(series_gen):
    289                 # ignore SettingWithCopy here in case the user mutates
--> 290                 results[i] = self.f(v)
    291                 if isinstance(results[i], ABCSeries):
    292                     # If we have a view on v, we need to make a copy because

<ipython-input-7-1663c693d570> in separate_items(row)
      3     '''
      4
----> 5     return [i.strip() for i in
row['items_owned'].replace('[','').replace(']','').replace("'","").split(';')]

AttributeError: 'float' object has no attribute 'replace'
```

Let's figure out why. First we'll look at the data again to remember what we expected.

```
safi_df['items_owned']
```

```
0      ['bicycle' ;  'television' ;  'solar_panel' ; ...
1      ['cow_cart' ;  'bicycle' ;  'radio' ;  'cow_pl...
2                                       ['solar_torch']
3      ['bicycle' ;  'radio' ;  'cow_plough' ;  'sola...
4      ['motorcyle' ;  'radio' ;  'cow_plough' ;  'mo...
                              ...
126            ['motorcyle' ;  'radio' ;  'solar_panel']
127    ['car' ;  'lorry' ;  'television' ;  'radio' ;...
128    ['radio' ;  'solar_panel' ;  'solar_torch' ;  ...
129    ['cow_cart' ;  'lorry' ;  'motorcyle' ;  'comp...
130    ['radio' ;  'cow_plough' ;  'solar_panel' ;  '...
Name: items_owned, Length: 131, dtype: object
```

The error says that some item is a float, let's look at the value of only the ones that are `float`

```
[item_list for item_list in  safi_df['items_owned'] if type(item_list)==float]
```

```
[nan, nan, nan, nan, nan, nan, nan, nan, nan, nan]
```

Now we can modify the cleaning function

```
def separate_items(row):
    '''
    '''
    if type(row['items_owned'])==str:
        return [i.strip() for i in
row['items_owned'].replace('[','').replace(']','').replace("'","").split(';')]
    else:
        return []
```

and apply it again, to see if it works now.

```
safi_df.apply(separate_items,axis=1)
```

```
0              [bicycle, television, solar_panel, table]
1      [cow_cart, bicycle, radio, cow_plough, solar_p...
2                                          [solar_torch]
3      [bicycle, radio, cow_plough, solar_panel, mobi...
4           [motorcyle, radio, cow_plough, mobile_phone]
                                ...
126                      [motorcyle, radio, solar_panel]
127    [car, lorry, television, radio, sterio, cow_pl...
128      [radio, solar_panel, solar_torch, mobile_phone]
129    [cow_cart, lorry, motorcyle, computer, televis...
130    [radio, cow_plough, solar_panel, solar_torch, ...
Length: 131, dtype: object
```

Dummies From Lists

```
safi_df['items_owned'] = safi_df.apply(separate_items,axis=1)
```

```
pd.get_dummies(safi_df['items_owned'].apply(pd.Series).stack())
```

		bicycle	car	computer	cow_cart	cow_plough	electricity	fridge	lorry	mobile
0	**0**	1	0	0	0	0	0	0	0	
	1	0	0	0	0	0	0	0	0	
	2	0	0	0	0	0	0	0	0	
	3	0	0	0	0	0	0	0	0	
1	**0**	0	0	0	1	0	0	0	0	
...	**...**	...	...	...	...	...	...	...	...	
130	**1**	0	0	0	0	1	0	0	0	
	2	0	0	0	0	0	0	0	0	
	3	0	0	0	0	0	0	0	0	
	4	0	0	0	0	0	0	0	0	
	5	0	0	0	0	0	0	0	0	

621 rows × 17 columns

```
safi_df['items_owned'].shape
```

```
(131,)
```

```
pd.get_dummies(safi_df['items_owned'].apply(pd.Series).stack()).index
```

```
MultiIndex([(  0,  0),
            (  0,  1),
            (  0,  2),
            (  0,  3),
            (  1,  0),
            (  1,  1),
            (  1,  2),
            (  1,  3),
            (  1,  4),
            (  1,  5),
            ...
            (129,  8),
            (129,  9),
            (129, 10),
            (129, 11),
            (130,  0),
            (130,  1),
            (130,  2),
            (130,  3),
            (130,  4),
            (130,  5)],
           length=621)
```

```
pd.get_dummies(safi_df['items_owned'].apply(pd.Series).stack()).sum(level=0)
```

	bicycle	car	computer	cow_cart	cow_plough	electricity	fridge	lorry	mobile_phe
0	1	0	0	0	0	0	0	0	
1	1	0	0	1	1	0	0	0	
2	0	0	0	0	0	0	0	0	
3	1	0	0	0	1	0	0	0	
4	0	0	0	0	1	0	0	0	
...	...	...	...	...	...	...	...	...	
126	0	0	0	0	0	0	0	0	
127	0	1	0	0	1	1	1	1	
128	0	0	0	0	0	0	0	0	
129	0	0	1	1	1	1	0	1	
130	0	0	0	0	1	0	0	0	

121 rows × 17 columns

```
safi_df['items_owned'].head()
```

```
0            [bicycle, television, solar_panel, table]
1    [cow_cart, bicycle, radio, cow_plough, solar_p...
2                                        [solar_torch]
3    [bicycle, radio, cow_plough, solar_panel, mobi...
4         [motorcyle, radio, cow_plough, mobile_phone]
Name: items_owned, dtype: object
```

```
safi_item_df =
pd.get_dummies(safi_df['items_owned'].apply(pd.Series).stack()).sum(level=0)
```

Pandas concat

```
safi_df = pd.concat([safi_df, safi_item_df],axis=1)
```

```
safi_df.shape
```

```
(131, 82)
```

```
safi_df.head()
```

	key_id	interview_date	quest_no	start	end	province	distric
0	1	17 November 2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	Manica	Manic
1	2	17 November 2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	Manica	Manic
2	3	17 November 2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	Manica	Manic
3	4	17 November 2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	Manica	Manic
4	5	17 November 2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	Manica	Manic

5 rows × 82 columns

Working with Dummy Variables

```
safi_df['bicycle']
```

```
0      1.0
1      1.0
2      0.0
3      1.0
4      0.0
      ...
126    0.0
127    0.0
128    0.0
129    0.0
130    0.0
Name: bicycle, Length: 131, dtype: float64
```

```
sum(safi_df['bicycle'])
```

```
nan
```

```
safi_df['bicycle'].sum()
```

```
60.0
```

```
safi_df['bicycle'].count()
```

```
121
```

```
safi_df['bicycle'].shape
```

```
(131,)
```

```
pd.concat([safi_df, safi_item_df],axis=0).shape
```

```
(252, 82)
```

Questions after class

What does the `get_dummies` function do?

To illustrate, let's first make a small dataframe to look at it

```
ex_data = [['a',2,'x'],['b',5,'o'],['a',4,'o'],['c',2,'x'],['b',3,'x']]
ex_df = pd.DataFrame(data =ex_data, columns=['char','num','symbol'])
ex_df
```

	char	num	symbol
0	a	2	x
1	b	5	o
2	a	4	o
3	c	2	x
4	b	3	x

When we apply `get_dummies` to one column (a `pd.Series`) it makes a column for each value of that column. First we look at that column to focus on what it looks like. Note that there are 3 different values.

```
ex_df['char']
```

```
0    a
1    b
2    a
3    c
4    b
Name: char, dtype: object
```

Now conver it to dummy variables, now we have 3 columns. For each row there is a 1 in the column corresponding to the value and zeros elsewhere.

```
pd.get_dummies(ex_df['char'])
```

	a	b	c
0	1	0	0
1	0	1	0
2	1	0	0
3	0	0	1
4	0	1	0

We can confirm by summing across the rows, it's all ones.

```
pd.get_dummies(ex_df['char']).sum(axis=1)
```

```
0    1
1    1
2    1
3    1
4    1
dtype: int64
```

If we sum down the columns, we get the count of each value

```
pd.get_dummies(ex_df['char']).sum(axis=0)
```

```
a    2
b    2
c    1
dtype: int64
```

it's the same thing as `value_counts()`

```
ex_df['char'].value_counts()
```

```
a    2
b    2
c    1
Name: char, dtype: int64
```

We can also apply it one column, but append it to the whole dataset this way.

```
pd.get_dummies(ex_df,columns=['char'])
```

	num	symbol	char_a	char_b	char_c
0	2	x	1	0	0
1	5	o	0	1	0
2	4	o	1	0	0
3	2	x	0	0	1
4	3	x	0	1	0

This way, it prepends the column values with the original column name. This way concatenates on it's own and we don't have to do it separately.

When we apply `get_dummies` on a whole DataFrame, without indicating a column it converts all of the columns of pandas type `object`

```
ex_df.dtypes
```

```
char      object
num        int64
symbol    object
dtype: object
```

```
pd.get_dummies(ex_df,)
```

	num	char_a	char_b	char_c	symbol_o	symbol_x
0	2	1	0	0	0	1
1	5	0	1	0	1	0
2	4	1	0	0	1	0
3	2	0	0	1	0	1
4	3	0	1	0	0	1

Class 10: Cleaning review and Ray Summit Keynotes

- Say hello on zoom chat
- join prismia
- sign up so you can watch Ray Summit talks by Pandas and Scikit learn

```
import pandas as pd
```

```
# %load http://drsmb.co/310
data_url = 'https://github.com/rhodyprog4ds/inclass-data/raw/main/ca_dds_summary.xlsx'
```

Let's look at the data

```
pd.read_excel(data_url)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-3-cb7bb7f4d96e> in <module>
----> 1 pd.read_excel(data_url)

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/util/_decorators.py in wrapper(*args, **kwargs)
    297                 )
    298                 warnings.warn(msg, FutureWarning, stacklevel=stacklevel)
--> 299             return func(*args, **kwargs)
    300
    301         return wrapper

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in read_excel(io, sheet_name, header, names,
index_col, usecols, squeeze, dtype, engine, converters, true_values, false_values,
skiprows, nrows, na_values, keep_default_na, na_filter, verbose, parse_dates,
date_parser, thousands, comment, skipfooter, convert_float, mangle_dupe_cols,
storage_options)
    334     if not isinstance(io, ExcelFile):
    335         should_close = True
--> 336         io = ExcelFile(io, storage_options=storage_options, engine=engine)
    337     elif engine and engine != io.engine:
    338         raise ValueError(

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in __init__(self, path_or_buffer, engine,
storage_options)
   1101             if ext != "xls" and xlrd_version >= "2":
   1102                 raise ValueError(
-> 1103                     f"Your version of xlrd is {xlrd_version}. In xlrd >= 2.0,
"
   1104                     f"only the xls format is supported. Install openpyxl
instead."
   1105                 )

ValueError: Your version of xlrd is 2.0.1. In xlrd >= 2.0, only the xls format is
supported. Install openpyxl instead.
```

We can read multiple rows in as the header

```
pd.read_excel(data_url,header=list(range(4)))
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-4-aeb280e6386d> in <module>
----> 1 pd.read_excel(data_url,header=list(range(4)))

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/util/_decorators.py in wrapper(*args, **kwargs)
    297                 )
    298                 warnings.warn(msg, FutureWarning, stacklevel=stacklevel)
--> 299             return func(*args, **kwargs)
    300
    301         return wrapper

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in read_excel(io, sheet_name, header, names,
index_col, usecols, squeeze, dtype, engine, converters, true_values, false_values,
skiprows, nrows, na_values, keep_default_na, na_filter, verbose, parse_dates,
date_parser, thousands, comment, skipfooter, convert_float, mangle_dupe_cols,
storage_options)
    334     if not isinstance(io, ExcelFile):
    335         should_close = True
--> 336         io = ExcelFile(io, storage_options=storage_options, engine=engine)
    337     elif engine and engine != io.engine:
    338         raise ValueError(

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in __init__(self, path_or_buffer, engine,
storage_options)
   1101             if ext != "xls" and xlrd_version >= "2":
   1102                 raise ValueError(
-> 1103                     f"Your version of xlrd is {xlrd_version}. In xlrd >= 2.0,
"
   1104                     f"only the xls format is supported. Install openpyxl
instead."
   1105                 )

ValueError: Your version of xlrd is 2.0.1. In xlrd >= 2.0, only the xls format is
supported. Install openpyxl instead.
```

Looks good, let's save this to a DataFrame

```
ca_dds_df = pd.read_excel(data_url,header=list(range(4)))
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-5-627b4cbdff1b> in <module>
----> 1 ca_dds_df = pd.read_excel(data_url,header=list(range(4)))

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/util/_decorators.py in wrapper(*args, **kwargs)
    297                 )
    298                 warnings.warn(msg, FutureWarning, stacklevel=stacklevel)
--> 299             return func(*args, **kwargs)
    300
    301         return wrapper

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in read_excel(io, sheet_name, header, names,
index_col, usecols, squeeze, dtype, engine, converters, true_values, false_values,
skiprows, nrows, na_values, keep_default_na, na_filter, verbose, parse_dates,
date_parser, thousands, comment, skipfooter, convert_float, mangle_dupe_cols,
storage_options)
    334     if not isinstance(io, ExcelFile):
    335         should_close = True
--> 336         io = ExcelFile(io, storage_options=storage_options, engine=engine)
    337     elif engine and engine != io.engine:
    338         raise ValueError(

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in __init__(self, path_or_buffer, engine,
storage_options)
   1101             if ext != "xls" and xlrd_version >= "2":
   1102                 raise ValueError(
-> 1103                     f"Your version of xlrd is {xlrd_version}. In xlrd >= 2.0,
"
   1104                     f"only the xls format is supported. Install openpyxl
instead."
   1105                 )

ValueError: Your version of xlrd is 2.0.1. In xlrd >= 2.0, only the xls format is
supported. Install openpyxl instead.
```

```
ca_dds_df.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-6-8d811b295c4b> in <module>
----> 1 ca_dds_df.head()

NameError: name 'ca_dds_df' is not defined
```

```
ca_dds_df.columns
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-7-2f51e10b6c51> in <module>
----> 1 ca_dds_df.columns

NameError: name 'ca_dds_df' is not defined
```

Ray Summit Notes

contribute things you learned here

Pandas, by Wes

- Pandas was designed to do data science on your laptop
- It's designed to be coupled tightly to numpy, which is why it's not very fast, especially with strings
-

Scikit Learn

- Data science for the many not the mighty
- Machine learning for all

Class 11: Cleaning Data

```
import pandas as pd
```

```
# %load http://drsmb.co/310
data_url = 'https://github.com/rhodyprog4ds/inclass-data/raw/main/ca_dds_summary.xlsx'
```

Recall from Wednesday

```
ca_dds_df = pd.read_excel(data_url, header=list(range(3)))
ca_dds_df
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-3-4542499f505a> in <module>
----> 1 ca_dds_df = pd.read_excel(data_url, header=list(range(3)))
      2 ca_dds_df

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/util/_decorators.py in wrapper(*args, **kwargs)
    297                 )
    298                 warnings.warn(msg, FutureWarning, stacklevel=stacklevel)
--> 299             return func(*args, **kwargs)
    300
    301         return wrapper

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in read_excel(io, sheet_name, header, names,
index_col, usecols, squeeze, dtype, engine, converters, true_values, false_values,
skiprows, nrows, na_values, keep_default_na, na_filter, verbose, parse_dates,
date_parser, thousands, comment, skipfooter, convert_float, mangle_dupe_cols,
storage_options)
    334     if not isinstance(io, ExcelFile):
    335         should_close = True
--> 336         io = ExcelFile(io, storage_options=storage_options, engine=engine)
    337     elif engine and engine != io.engine:
    338         raise ValueError(

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/pandas/io/excel/_base.py in __init__(self, path_or_buffer, engine,
storage_options)
   1101             if ext != "xls" and xlrd_version >= "2":
   1102                 raise ValueError(
-> 1103                     f"Your version of xlrd is {xlrd_version}. In xlrd >= 2.0,
"
   1104                     f"only the xls format is supported. Install openpyxl
instead."
   1105                 )

ValueError: Your version of xlrd is 2.0.1. In xlrd >= 2.0, only the xls format is
supported. Install openpyxl instead.
```

```
print(ca_dds_df)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-4-e640822e617f> in <module>
----> 1 print(ca_dds_df)

NameError: name 'ca_dds_df' is not defined
```

First we can look at the index

```
ca_dds_df.index
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-5-f24e6493175a> in <module>
----> 1 ca_dds_df.index

NameError: name 'ca_dds_df' is not defined
```

then the first's name

```
ca_dds_df.columns[0]
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-6-6dcb8c4e6570> in <module>
----> 1 ca_dds_df.columns[0]

NameError: name 'ca_dds_df' is not defined
```

For this we want to se that as the index

```
ca_dds_df.set_index(ca_dds_df.columns[0],inplace=True)

ca_dds_df
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-7-894c600386fd> in <module>
----> 1 ca_dds_df.set_index(ca_dds_df.columns[0],inplace=True)
      2
      3 ca_dds_df

NameError: name 'ca_dds_df' is not defined
```

We can get rid of rows that are all `nan`, that have no data in them, using `dropna` with `how='all'`

```
ca_dds_df.dropna(how='all',inplace=True)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-8-b9f52e336c27> in <module>
----> 1 ca_dds_df.dropna(how='all',inplace=True)

NameError: name 'ca_dds_df' is not defined
```

```
ca_dds_df.index
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-9-f24e6493175a> in <module>
----> 1 ca_dds_df.index

NameError: name 'ca_dds_df' is not defined
```

We want to still rename this, we we'll fill that in manually

```
ca_dds_df.index.rename('Age Cohort',inplace=True)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-10-fcfa8204b94f> in <module>
----> 1 ca_dds_df.index.rename('Age Cohort',inplace=True)

NameError: name 'ca_dds_df' is not defined
```

Now we want to unstack,

```
ca_dds_df.unstack(level= [0,1])
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-11-06d6e6545889> in <module>
----> 1 ca_dds_df.unstack(level= [0,1])

NameError: name 'ca_dds_df' is not defined
```

but this doesn't quite get what we want, instead we can

We can rotate (transpose `.T`)this, then unstack and rotate it back to get the outcome we want here.

```
ca_dds_df.T.unstack(level= [0,1]).T.reset_index()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-12-c8bbbfe0591e> in <module>
----> 1 ca_dds_df.T.unstack(level= [0,1]).T.reset_index()

NameError: name 'ca_dds_df' is not defined
```

This looks good, so we'll save it to a new DataFrame

```
ca_dds_clean = ca_dds_df.T.unstack(level= [0,1]).T.reset_index()
ca_dds_clean.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-13-eff451103ff6> in <module>
----> 1 ca_dds_clean = ca_dds_df.T.unstack(level= [0,1]).T.reset_index()
      2 ca_dds_clean.head()

NameError: name 'ca_dds_df' is not defined
```

We want to rename the two with nondescript names, but the index is not mutable

```
ca_dds_clean.columns[1] = 'Race'
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-14-f771ccd0e865> in <module>
----> 1 ca_dds_clean.columns[1] = 'Race'

NameError: name 'ca_dds_clean' is not defined
```

```
type(ca_dds_clean.columns)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-15-0372e536c94d> in <module>
----> 1 type(ca_dds_clean.columns)

NameError: name 'ca_dds_clean' is not defined
```

Instead we can use rename with a dictionary, this way is good also because it's explicit and readable.

```
ca_dds_clean.rename(columns= {'level_1':'Race', 'level_2':'Gender'},inplace=True)
ca_dds_clean.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-16-e447bbfeee5e> in <module>
----> 1 ca_dds_clean.rename(columns= {'level_1':'Race',
'level_2':'Gender'},inplace=True)
      2 ca_dds_clean.head()

NameError: name 'ca_dds_clean' is not defined
```

```
ca_dds_clean['count'][0]
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-17-f579f06c67b3> in <module>
----> 1 ca_dds_clean['count'][0]

NameError: name 'ca_dds_clean' is not defined
```

```
ca_dds_clean.Gender[0]
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-18-2e64fd505c34> in <module>
----> 1 ca_dds_clean.Gender[0]

NameError: name 'ca_dds_clean' is not defined
```

```
ca_dds_clean.Gender[0].lower()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-19-693531faa940> in <module>
----> 1 ca_dds_clean.Gender[0].lower()

NameError: name 'ca_dds_clean' is not defined
```

```
clean_cols = {c:c.lower().replace(' ','_') for c in ca_dds_clean.columns}
ca_dds_clean.rename(columns=clean_cols,inplace=True)
ca_dds_clean.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-20-bbc80f90b452> in <module>
----> 1 clean_cols = {c:c.lower().replace(' ','_') for c in ca_dds_clean.columns}
      2 ca_dds_clean.rename(columns=clean_cols,inplace=True)
      3 ca_dds_clean.head()

NameError: name 'ca_dds_clean' is not defined
```

```
ca_dds_clean.fillna('*')
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-21-3bf797621552> in <module>
----> 1 ca_dds_clean.fillna('*')

NameError: name 'ca_dds_clean' is not defined
```

Questions after class

What happens when we use dropna?

We can see the [documentation]() it drops the whole row or column.

```
import numpy as np #
test_df = pd.DataFrame(data = [[1,3,4,5],[2 ,6, np.na]])
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-22-482bb49100aa> in <module>
      1 import numpy as np #
----> 2 test_df = pd.DataFrame(data = [[1,3,4,5],[2 ,6, np.na]])

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-packages/numpy/__init__.py
in __getattr__(attr)
    302
    303             raise AttributeError("module {!r} has no attribute "
--> 304                                  "{!r}".format(__name__, attr))
    305
    306         def __dir__():

AttributeError: module 'numpy' has no attribute 'na'
```

Class 12: Constructing Datasets from Multiple Sources

```
import pandas as pd
```

Introduction

Sometimes we use data from multiple sources. For example data may be provided across a bunch of tables and we want to put them together. Or we may want to build our own dataset to answer the questions that we want to ask. For example we have a dataset of drivers who have been pulled over by the police, and need to compare distribution of them between day and night times. If we only have time of being pulled over without mentioning day/night, we can use another dataset

including time of sunset for every day and build our own dataset out of these two dataset to find the answer. Later we will also learn about getting data out of databses with some small database operations through SQLite and some python libararies.

To use relative paths as in class (`data/2018-games.csv`) instead of a full url `https://raw.githubusercontent.com/rhodyprog4ds/inclass-data/main/2018-games.csv`, download data from [this GitHub repo] by clicking on the green code button and choosing .zip. Then unzip the data and save it in a data folder in the same folder as the notebook. For the class notes, the urls make it so that the notebook can run without having to store the data in another place.

```
games_df18 = pd.read_csv('https://raw.githubusercontent.com/rhodyprog4ds/inclass-data/main/2018-games.csv')
games_df19 = pd.read_csv('https://raw.githubusercontent.com/rhodyprog4ds/inclass-data/main/2019-games.csv')
```

Stacking DataFrames

Let's look at the first couple of rows to see how the configuration of data is.

```
games_df18.head(2)
```

	Unnamed: 0	GAME_DATE_EST	GAME_ID	GAME_STATUS_TEXT	HOME_TEAM_ID	VISI
0	16196	2019-06-13	41800406	Final	1610612744	
1	16197	2019-06-10	41800405	Final	1610612761	

2 rows × 22 columns

```
games_df19.head(2)
```

	Unnamed: 0	GAME_DATE_EST	GAME_ID	GAME_STATUS_TEXT	HOME_TEAM_ID	VISI
0	0	2020-03-01	21900895	Final	1610612766	
1	1	2020-03-01	21900896	Final	1610612750	

2 rows × 22 columns

As we can see, both datasets have the same columns, and just for two different years.

```
games_df18.shape
```

```
(1378, 22)
```

```
games_df19.shape
```

```
(965, 22)
```

Let's concatenate two dataframes to make one dataframe out of them.

```
games_df = pd.concat([games_df18,games_df19])
games_df.shape
```

```
(2343, 22)
```

As we can see rows are added up but we have the same numbe of rcolums.

```
games_df.columns
```

```
Index(['Unnamed: 0', 'GAME_DATE_EST', 'GAME_ID', 'GAME_STATUS_TEXT',
       'HOME_TEAM_ID', 'VISITOR_TEAM_ID', 'SEASON', 'TEAM_ID_home', 'PTS_home',
       'FG_PCT_home', 'FT_PCT_home', 'FG3_PCT_home', 'AST_home', 'REB_home',
       'TEAM_ID_away', 'PTS_away', 'FG_PCT_away', 'FT_PCT_away',
       'FG3_PCT_away', 'AST_away', 'REB_away', 'HOME_TEAM_WINS'],
      dtype='object')
```

Let's drop the column we do not need.

```
games_df.drop(columns= 'Unnamed: 0',inplace=True,)
games_df.head(2)
```

	GAME_DATE_EST	GAME_ID	GAME_STATUS_TEXT	HOME_TEAM_ID	VISITOR_TEAM_I
0	2019-06-13	41800406	Final	1610612744	16106127
1	2019-06-10	41800405	Final	1610612761	16106127

2 rows × 21 columns

Merging Data Frames

Now we read another dataset which some of its columns are the same as dataframe "games_df" and some are different.

```
teams_df = pd.read_csv('https://raw.githubusercontent.com/rhodyprog4ds/inclass-data/main/teams.csv')
teams_df.head(2)
```

	LEAGUE_ID	TEAM_ID	MIN_YEAR	MAX_YEAR	ABBREVIATION	NICKNAME	YEAR
0	0	1610612737	1949	2019	ATL	Hawks	
1	0	1610612738	1946	2019	BOS	Celtics	

We use left_on='TEAM_ID' and right_on = 'HOME_TEAM_ID' to match two dataframes.

```
merge1_df = pd.merge(teams_df,games_df,left_on='TEAM_ID', right_on = 'HOME_TEAM_ID')
merge1_df.head(2)
```

	LEAGUE_ID	TEAM_ID	MIN_YEAR	MAX_YEAR	ABBREVIATION	NICKNAME	YEAR
0	0	1610612737	1949	2019	ATL	Hawks	
1	0	1610612737	1949	2019	ATL	Hawks	

2 rows × 35 columns

We want information for each game and append the team info onto that. So, lets try another settings in our mergging.

```
merge1_df.shape
```

```
(2343, 35)
```

```
merge2_df = pd.merge(teams_df, games_df,left_on='TEAM_ID', right_on = 'HOME_TEAM_ID', how='outer')
```

```
merge2_df.head(2)
```

	LEAGUE_ID	TEAM_ID	MIN_YEAR	MAX_YEAR	ABBREVIATION	NICKNAME	YEAR
0	0	1610612737	1949	2019	ATL	Hawks	
1	0	1610612737	1949	2019	ATL	Hawks	

2 rows × 35 columns

```
merge2_df.shape
```

```
(2343, 35)
```

We can group by “ARENA” and then look at the “mean” statistics.

```
merge1_df.groupby('ARENA').mean()
```

	LEAGUE_ID	TEAM_ID	MIN_YEAR	MAX_YEAR	YEARFOUNDED
ARENA					
AT&T Center	0.0	1.610613e+09	1976.000000	2019.0	1976.000000
American Airlines Center	0.0	1.610613e+09	1980.000000	2019.0	1980.000000
AmericanAirlines Arena	0.0	1.610613e+09	1988.000000	2019.0	1988.000000
Amway Center	0.0	1.610613e+09	1989.000000	2019.0	1989.000000
Bankers Life Fieldhouse	0.0	1.610613e+09	1976.000000	2019.0	1976.000000
Barclays Center	0.0	1.610613e+09	1976.000000	2019.0	1976.000000
Capital One Arena	0.0	1.610613e+09	1961.000000	2019.0	1961.000000
Chase Center	0.0	1.610613e+09	1946.000000	2019.0	1946.000000
Chesapeake Energy Arena	0.0	1.610613e+09	1967.000000	2019.0	1967.000000
FedExForum	0.0	1.610613e+09	1995.000000	2019.0	1995.000000
Fiserv Forum	0.0	1.610613e+09	1968.000000	2019.0	1968.000000
Golden 1 Center	0.0	1.610613e+09	1948.000000	2019.0	1948.000000
Little Caesars Arena	0.0	1.610613e+09	1948.000000	2019.0	1948.000000
Madison Square Garden	0.0	1.610613e+09	1946.000000	2019.0	1946.000000
Moda Center	0.0	1.610613e+09	1970.000000	2019.0	1970.000000
Pepsi Center	0.0	1.610613e+09	1976.000000	2019.0	1976.000000
Quicken Loans Arena	0.0	1.610613e+09	1970.000000	2019.0	1970.000000
Scotiabank Arena	0.0	1.610613e+09	1995.000000	2019.0	1995.000000
Smoothie King Center	0.0	1.610613e+09	2002.000000	2019.0	2002.000000
Spectrum Center	0.0	1.610613e+09	1988.000000	2019.0	1988.000000
Staples Center	0.0	1.610613e+09	1959.210191	2019.0	1959.210191
State Farm Arena	0.0	1.610613e+09	1949.000000	2019.0	1949.000000
TD Garden	0.0	1.610613e+09	1946.000000	2019.0	1946.000000
Talking Stick Resort Arena	0.0	1.610613e+09	1968.000000	2019.0	1968.000000
Target Center	0.0	1.610613e+09	1989.000000	2019.0	1989.000000
Toyota Center	0.0	1.610613e+09	1967.000000	2019.0	1967.000000
United Center	0.0	1.610613e+09	1966.000000	2019.0	1966.000000
Vivint Smart Home Arena	0.0	1.610613e+09	1974.000000	2019.0	1974.000000
Wells Fargo Center	0.0	1.610613e+09	1949.000000	2019.0	1949.000000

29 rows × 25 columns

How to combine the same data for different outcomes

We can combine datasets in different ways to learn different things about the data. Let's now read info about the players.

```
players18 = pd.read_csv('https://raw.githubusercontent.com/rhodyprog4ds/inclass-data/main/2018-players.csv')
players19 = pd.read_csv('https://raw.githubusercontent.com/rhodyprog4ds/inclass-data/main/2019-players.csv')
players18.head()
```

	Unnamed: 0	PLAYER_NAME	TEAM_ID	PLAYER_ID	SEASON
0	626	Kawhi Leonard	1610612761	202695	2018
1	627	Pascal Siakam	1610612761	1627783	2018
2	628	Marc Gasol	1610612761	201188	2018
3	629	Danny Green	1610612761	201980	2018
4	630	Kyle Lowry	1610612761	200768	2018

First let's look at the shape of each of them to have a reference for what happens when we try different merges.

```
players18.shape, players19.shape
```

```
((748, 5), (626, 5))
```

One thing we might want to do is to put all of the information into one long DataFrame. We can do this with `concat`.

```
pd.concat([players18,players19])
```

	Unnamed: 0	PLAYER_NAME	TEAM_ID	PLAYER_ID	SEASON
0	626	Kawhi Leonard	1610612761	202695	2018
1	627	Pascal Siakam	1610612761	1627783	2018
2	628	Marc Gasol	1610612761	201188	2018
3	629	Danny Green	1610612761	201980	2018
4	630	Kyle Lowry	1610612761	200768	2018
...	...	...	...	...	...
621	621	Anthony Bennett	1610612745	203461	2019
622	622	Ray Spalding	1610612737	1629034	2019
623	623	Devyn Marble	1610612744	203906	2019
624	624	Hassani Gravett	1610612753	1629755	2019
625	625	JaKeenan Gant	1610612754	1629721	2019

1374 rows × 5 columns

This allows us to see all the players for each year, we could do groupby and count to see how many players played each year for example.

We can check that this is the size we expected.

```
pd.concat([players18,players19]).shape
```

```
(1374, 5)
```

If we use the default merge settings we get an empty result because the two DataFrames have the same columns so pandas tries to merge `on` all of the columns, but there are no rows that have the same value in all of the columns, so there's nothing left.

```
pd.merge(players18,players19,)
```

Unnamed: 0	PLAYER_NAME	TEAM_ID	PLAYER_ID	SEASON

If we merge with on='PLAYER_ID', it only requires that one column to be the same to match rows from the two DataFrames together. With the default value for `how` or explicitly setting `how='inner'` we get the info of players who played both seasons.

```
pd.merge(players18,players19,on='PLAYER_ID', how='inner')
```

	Unnamed: 0_x	PLAYER_NAME_x	TEAM_ID_x	PLAYER_ID	SEASON_x	Unnamed: 0_y	PLA
0	626	Kawhi Leonard	1610612761	202695	2018	299	
1	627	Pascal Siakam	1610612761	1627783	2018	275	
2	628	Marc Gasol	1610612761	201188	2018	276	
3	1157	Marc Gasol	1610612763	201188	2018	276	
4	629	Danny Green	1610612761	201980	2018	202	
...	...	...	...	...	...	...	
533	1339	Abdul Gaddy	1610612760	203583	2018	561	
534	1342	Andre Roberson	1610612760	203460	2018	566	A
535	1348	Norvel Pelle	1610612755	203658	2018	24	
536	1350	Denzel Valentine	1610612741	1627756	2018	58	D
537	1353	C.J. Wilcox	1610612754	203912	2018	573	

538 rows × 9 columns

When we use `outer` we get one row for each player who played in either season or both seasons. From this we can for example see who changed teams, who are the rookies in 2019 and who retired or was unsigned in 2019.

```
pd.merge(players18,players19,on='PLAYER_ID', how='outer')
```

	Unnamed: 0_x	PLAYER_NAME_x	TEAM_ID_x	PLAYER_ID	SEASON_x	Unnamed: 0_y	PI
0	626.0	Kawhi Leonard	1.610613e+09	202695	2018.0	299.0	
1	627.0	Pascal Siakam	1.610613e+09	1627783	2018.0	275.0	
2	628.0	Marc Gasol	1.610613e+09	201188	2018.0	276.0	
3	1157.0	Marc Gasol	1.610613e+09	201188	2018.0	276.0	
4	629.0	Danny Green	1.610613e+09	201980	2018.0	202.0	
...	...	...	...	...	...	...	
922	NaN	NaN	NaN	1629097	NaN	619.0	
923	NaN	NaN	NaN	203461	NaN	621.0	
924	NaN	NaN	NaN	203906	NaN	623.0	
925	NaN	NaN	NaN	1629755	NaN	624.0	
926	NaN	NaN	NaN	1629721	NaN	625.0	

927 rows × 9 columns

Using `left` gives us the `'PLAYER_ID'` that are in the left(`players18`) DataFrame, including those that are in both DataFrame `right` would give players in the `players19` DataFrame or both DataFrames. With this result, we can see who retired, but not the 2019 rookies.

```
pd.merge(players18,players19,on='PLAYER_ID', how='left')
```

	Unnamed: 0_x	PLAYER_NAME_x	TEAM_ID_x	PLAYER_ID	SEASON_x	Unnamed: 0_y	PLA
0	626	Kawhi Leonard	1610612761	202695	2018	299.0	
1	627	Pascal Siakam	1610612761	1627783	2018	275.0	
2	628	Marc Gasol	1610612761	201188	2018	276.0	
3	629	Danny Green	1610612761	201980	2018	202.0	
4	630	Kyle Lowry	1610612761	200768	2018	451.0	
...	...	...	...	...	...	...	
749	1369	Tyrius Walker	1610612752	1629246	2018	NaN	
750	1370	Marcus Lee	1610612748	1629159	2018	NaN	
751	1371	Trey Lewis	1610612762	1629163	2018	NaN	
752	1372	Emanuel Terry	1610612743	1629150	2018	NaN	
753	1373	Justin Bibbs	1610612738	1629167	2018	NaN	

754 rows × 9 columns

Try it yourself

Try different merges and inspect them:

- how many rows & columns?
- Where are NaN values inserted?
- What rows from the original datasets are not included?
- describe each type of merge in your own words

Split a DataFrame into separate data frames by subsetting the columns and indexing the rows with `loc`, then use concat to put it back together. Programmatically check that it's back together correctly.

Class 13: Data from multiple sources and Databases

Welcome:

1. Say hello on zoom
2. Download data from Mondays' notes if you don't have it already
3. log onto Prismia.chat

```
import pandas as pd
```

Merging review

```
df18_players = pd.read_csv('data/2018-players.csv')
df19_players = pd.read_csv('data/2019-players.csv')
```

```
df18_players.shape, df19_players.shape
```

```
((748, 5), (626, 5))
```

```
pd.merge(df18_players,df19_players,how='inner',on='PLAYER_ID')
```

	Unnamed: 0_x	PLAYER_NAME_x	TEAM_ID_x	PLAYER_ID	SEASON_x	Unnamed: 0_y	PLA
0	626	Kawhi Leonard	1610612761	202695	2018	299	
1	627	Pascal Siakam	1610612761	1627783	2018	275	
2	628	Marc Gasol	1610612761	201188	2018	276	
3	1157	Marc Gasol	1610612763	201188	2018	276	
4	629	Danny Green	1610612761	201980	2018	202	
...	...	...	...	...	...	...	
533	1339	Abdul Gaddy	1610612760	203583	2018	561	
534	1342	Andre Roberson	1610612760	203460	2018	566	A
535	1348	Norvel Pelle	1610612755	203658	2018	24	
536	1350	Denzel Valentine	1610612741	1627756	2018	58	D
537	1353	C.J. Wilcox	1610612754	203912	2018	573	

538 rows × 9 columns

```
pd.merge(df18_players,df19_players,how='outer',on='PLAYER_ID')
```

	Unnamed: 0_x	PLAYER_NAME_x	TEAM_ID_x	PLAYER_ID	SEASON_x	Unnamed: 0_y	PI
0	626.0	Kawhi Leonard	1.610613e+09	202695	2018.0	299.0	
1	627.0	Pascal Siakam	1.610613e+09	1627783	2018.0	275.0	
2	628.0	Marc Gasol	1.610613e+09	201188	2018.0	276.0	
3	1157.0	Marc Gasol	1.610613e+09	201188	2018.0	276.0	
4	629.0	Danny Green	1.610613e+09	201980	2018.0	202.0	
...	...	...	...	...	...	...	
922	NaN	NaN	NaN	1629097	NaN	619.0	
923	NaN	NaN	NaN	203461	NaN	621.0	
924	NaN	NaN	NaN	203906	NaN	623.0	
925	NaN	NaN	NaN	1629755	NaN	624.0	
926	NaN	NaN	NaN	1629721	NaN	625.0	

927 rows × 9 columns

```
pd.merge(df18_players,df19_players,how='left',on='PLAYER_ID')
```

	Unnamed: 0_x	PLAYER_NAME_x	TEAM_ID_x	PLAYER_ID	SEASON_x	Unnamed: 0_y	PLA
0	626	Kawhi Leonard	1610612761	202695	2018	299.0	
1	627	Pascal Siakam	1610612761	1627783	2018	275.0	
2	628	Marc Gasol	1610612761	201188	2018	276.0	
3	629	Danny Green	1610612761	201980	2018	202.0	
4	630	Kyle Lowry	1610612761	200768	2018	451.0	
...	...	...	...	...	...	...	
749	1369	Tyrius Walker	1610612752	1629246	2018	NaN	
750	1370	Marcus Lee	1610612748	1629159	2018	NaN	
751	1371	Trey Lewis	1610612762	1629163	2018	NaN	
752	1372	Emanuel Terry	1610612743	1629150	2018	NaN	
753	1373	Justin Bibbs	1610612738	1629167	2018	NaN	

754 rows × 9 columns

```
df18_games = pd.read_csv('data/2018-games.csv')
df19_games = pd.read_csv('data/2019-games.csv')
```

```
games_df = pd.concat([df18_games,df19_games])
games_df.head()
```

	Unnamed: 0	GAME_DATE_EST	GAME_ID	GAME_STATUS_TEXT	HOME_TEAM_ID	VISI
0	16196	2019-06-13	41800406	Final	1610612744	
1	16197	2019-06-10	41800405	Final	1610612761	
2	16198	2019-06-07	41800404	Final	1610612744	
3	16199	2019-06-05	41800403	Final	1610612744	
4	16200	2019-06-02	41800402	Final	1610612761	

5 rows × 22 columns

Data

In order to work with large dataset, one option is to use a computer that has more memory, such as an HPC system. An easier option is to load data from a database instead of loading all data into memory. One way of working with databases is to use a Python library named `sqlite3` which allows us to create a connection between the database and Python so that we can pull data into Python to work with it.

> ⚠️ **Warning**
>
> To work with a database it needs to be downloaded, we can't load it via url like we did with the `.csv` files that we've been using, because when we need to make a connection to the database that can dynamically interact with the database.

Today, we'll work with a database version of the same SAFI dataset we've been working with. Download it and put it in a `data` folder that's in the same folder as your notebook.

First we'll import the needed library, this one doesn't have a standard alias, so we'll use it by referring to the whole name.

```
import sqlite3
```

First we'll connect to the database

```
con = sqlite3.connect('data/SQL_SAFI.sqlite')
```

Next we establish a cursor with the connection and then run a query.

```
cur = con.cursor()
cur.execute("SELECT * FROM Farms")
```

```
<sqlite3.Cursor at 0x7f4994c69180>
```

Here, "SELECT * FROM Farms" is a query we made into the cursor. It does not return the data, it just says what data to get ready, this is another opportunity to get the data in smaller chunks, we can choose in the next step to bring all or only some of the data from the query's result into Python.

This query says to get all (`*`) the data from the `Farms` table.

Now, we can use this cursor to pull all of the results from that query into our notebook.

```
rows = cur.fetchall()
rows
```

```
[(1,
  'Moz',
  '17/11/2016',
  1,
  '2017-03-23T09:49:57.000Z',
  '2017-04-02T17:29:08.000Z',
  'Manica',
  'Manica',
  'Bandula',
```

```
 'God',
 11,
 'no',
 3,
 3,
 'no',
 4,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'television', 'solar_panel', 'table']",
 2,
 "['Jan']",
 "['na', 'rely_less_food', 'reduce_meals', 'day_night_hungry']",
 -19.11225943,
 33.48345609,
 698.0,
 14.0,
 'uuid:ec241f2c-0609-46ed-b5e8-fe575f6cefef'),
(2,
 'Moz',
 '17/11/2016',
 1,
 '2017-04-02T09:48:16.000Z',
 '2017-04-02T17:26:19.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 2,
 'yes',
 7,
 7,
 'no',
 9,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug', 'Sept']",
 2.0,
```

```
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'once',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch',
'table', 'mobile_phone']",
  2,
  "['Jan', 'Sept', 'Oct', 'Nov', 'Dec']",
  "['na', 'reduce_meals', 'restrict_adults', 'borrow_food', 'seek_government']",
  -19.11247712,
  33.48341568,
  690.0,
  19.0,
  'uuid:099de9c9-3e5e-427b-8452-26250e840d6e'),
 (3,
  'Moz',
  '17/11/2016',
  3,
  '2017-04-02T14:35:26.000Z',
  '2017-04-02T17:26:53.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  40,
  'no',
  10,
  10,
  'no',
  15,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  1,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['solar_torch']",
  2,
  "['Jan', 'Feb', 'Mar', 'Oct', 'Nov', 'Dec']",
  "['na', 'restrict_adults', 'lab_ex_food']",
  -19.1121076,
  33.48344998,
  674.0,
  13.0,
```

```
 'uuid:193d7daf-9582-409b-bf09-027dd36f9007'),
(4,
 'Moz',
 '17/11/2016',
 4,
 '2017-04-02T14:55:18.000Z',
 '2017-04-02T17:27:16.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 6,
 'no',
 7,
 7,
 'no',
 6,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['oxen', 'cows']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
 2,
 "['Sept', 'Oct', 'Nov', 'Dec']",
 "['na', 'reduce_meals', 'restrict_adults', 'lab_ex_food']",
 -19.11222901,
 33.48342395,
 679.0,
 5.0,
 'uuid:148d1105-778a-4755-aa71-281eadd4a973'),
(5,
 'Moz',
 '17/11/2016',
 5,
 '2017-04-02T15:10:35.000Z',
 '2017-04-02T17:27:35.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 18,
 'no',
 7,
 7,
 'no',
 40,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
```

```
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats', 'poultry']",
 None,
 4,
 'yes',
 'no',
 "['motorcyle', 'radio', 'cow_plough', 'mobile_phone']",
 2,
 "['Aug', 'Sept', 'Oct', 'Nov']",
 "['na', 'go_forest', 'migrate']",
 -19.11221722,
 33.48342524,
 689.0,
 10.0,
 'uuid:2c867811-9696-4966-9866-f35c3e97d02d'),
(6,
 'Moz',
 '17/11/2016',
 6,
 '2017-04-02T15:27:25.000Z',
 '2017-04-02T17:28:02.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 3,
 'no',
 3,
 3,
 'no',
 3,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
```

```
 'no',
 None,
 2,
 "['Aug', 'Sept', 'Oct']",
 "['borrow_food', 'lab_ex_food', 'seek_government']",
 -19.112195899999996,
 33.48339187,
 692.0,
 12.0,
 'uuid:daa56c91-c8e3-44c3-a663-af6a49a2ca70'),
(7,
 'Moz',
 '17/11/2016',
 7,
 '2017-04-02T15:38:01.000Z',
 '2017-04-02T17:28:19.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 20,
 'no',
 6,
 6,
 'no',
 38,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'yes',
 4,
 4,
 'yes',
 None,
 4.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen']",
 None,
 1,
 'no',
 'no',
 "['motorcyle', 'cow_plough']",
 3,
 "['Nov']",
 "['lab_ex_food']",
 -19.11221904,
 33.48336498,
 709.0,
 11.0,
 'uuid:ae20a58d-56f4-43d7-bafa-e7963d850844'),
(8,
 'Moz',
 '16/11/2016',
 8,
 '2017-04-02T15:59:52.000Z',
 '2017-04-02T17:28:39.000Z',
 'Manica',
 'Manica',
 'Manica',
 'Chirodzo',
 16,
 'yes',
 12,
 12,
 'no',
 70,
```

```
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  3,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct']",
  10.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'goats']",
  None,
  2,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'television', 'radio', 'cow_plough', 'solar_panel',
'solar_torch', 'table', 'fridge']",
  2,
  "['Jan']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'borrow_food']",
  -19.11215963,
  33.48341914,
  700.0,
  9.0,
  'uuid:d6cee930-7be1-4fd9-88c0-82a08f90fb5a'),
 (9,
  'Moz',
  '16/11/2016',
  9,
  '2017-04-02T16:23:36.000Z',
  '2017-04-02T16:42:08.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  16,
  'no',
  8,
  8,
  'no',
  6,
  'yes',
  'no',
  'yes',
  'no',
  'grass',
  'burntbricks',
  'earth',
  'no',
  2,
  1,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Oct', 'Nov']",
  6.0,
  'yes',
  'no',
  None,
  'no',
  None,
```

```
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['television', 'solar_panel', 'solar_torch']",
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'restrict_adults',
'lab_ex_food']",
  -19.11221518,
  33.48343695,
  701.0,
  11.0,
  'uuid:846103d2-b1db-4055-b502-9cd510bb7b37'),
 (10,
  'Moz',
  '16/12/2016',
  10,
  '2017-04-02T17:03:28.000Z',
  '2017-04-02T17:25:11.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  22,
  'no',
  12,
  12,
  'no',
  23,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  5,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  22.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_panel', 'solar_torch', 'table']",
  3,
  "['Jan', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_portion', 'restrict_adults', 'lab_ex_food']",
  -19.1122147,
  33.48339436,
  710.0,
  14.0,
  'uuid:8f4e49bc-da81-4356-ae34-e0d794a23721'),
 (11,
  'Moz',
  '21/11/2016',
```

```
 11,
 '2017-04-03T03:16:15.000Z',
 '2017-04-03T03:31:10.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 6,
 'no',
 6,
 6,
 'no',
 20,
 'yes',
 'yes',
 'no',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['oxen', 'cows']",
 None,
 2,
 'no',
 'no',
 "['radio', 'cow_plough']",
 2,
 "['Oct', 'Nov']",
 "['rely_less_food', 'lab_ex_food']",
 -19.11219126,
 33.48345933,
 707.0,
 10.0,
 'uuid:d29b44e3-3348-4afc-aa4d-9eb34c89d483'),
(12,
 'Moz',
 '21/11/2016',
 12,
 '2017-04-03T03:31:13.000Z',
 '2017-04-03T03:58:34.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 20,
 'no',
 7,
 7,
 'yes',
 20,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 2,
 3,
 'no',
 2,
 2,
```

```
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov']",
 20.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows']",
 None,
 2,
 'yes',
 'no',
 "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'table']",
 3,
 "['Sept', 'Oct']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.11229465,
 33.48341504,
 696.0,
 13.0,
 'uuid:e6ee6269-b467-4e37-91fc-5e9eaf934557'),
(13,
 'Moz',
 '21/11/2016',
 13,
 '2017-04-03T03:58:43.000Z',
 '2017-04-03T04:19:36.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 7,
 'yes',
 6,
 6,
 'no',
 8,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'yes',
 "['Oct']",
 7.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['cows', 'goats', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'mobile_phone']",
 2,
 "['Sept', 'Oct', 'Nov']",
```

```
 "['lab_ex_food']",
 -19.11236935,
 33.48355635,
 706.0,
 15.0,
 'uuid:6c00c145-ee3b-409c-8c02-2c8d743b6918'),
(14,
 'Moz',
 '21/11/2016',
 14,
 '2017-04-03T04:19:57.000Z',
 '2017-04-03T04:50:05.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 20,
 'yes',
 10,
 10,
 'no',
 20,
 'yes',
 'yes',
 'no',
 'yes',
 'grass',
 'burntbricks',
 'earth',
 'no',
 3,
 3,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'yes',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'table', 'mobile_phone']",
 3,
 "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
 "['reduce_meals', 'lab_ex_food']",
 -19.11222089,
 33.4834388,
 698.0,
 11.0,
 'uuid:9b21467f-1116-4340-a3b1-1ab64f13c87d'),
(15,
 'Moz',
 '21/11/2016',
 15,
 '2017-04-03T05:12:17.000Z',
 '2017-04-03T05:28:44.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 30,
 'yes',
 5,
 5,
 'no',
 30,
 'yes',
 'yes',
 'yes',
 'yes',
```

```
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 30.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'table']",
 2,
 "['Jan', 'Feb', 'Mar', 'Apr', 'May', 'June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
 "['na', 'rely_less_food', 'limit_portion', 'reduce_meals', 'lab_ex_food']",
 -19.1120793,
 33.48339657,
 715.0,
 9.0,
 'uuid:a837e545-ff86-4a1c-a1a5-6186804b985f'),
(16,
 'Moz',
 '24/11/2016',
 16,
 '2017-04-03T05:29:24.000Z',
 '2017-04-03T05:40:53.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 24,
 'yes',
 6,
 6,
 'no',
 47,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'yes',
 2,
 1,
 'yes',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
```

```
 'yes',
 "['oxen', 'cows', 'goats', 'poultry']",
 None,
 4,
 'no',
 'no',
 "['radio', 'cow_plough', 'solar_panel', 'solar_torch']",
 3,
 "['Jan', 'Feb']",
 "['lab_ex_food']",
 -19.11210678,
 33.48344397,
 709.0,
 9.0,
 'uuid:d17db52f-4b87-4768-b534-ea8f9704c565'),
(17,
 'Moz',
 '21/11/2016',
 17,
 '2017-04-03T05:41:42.000Z',
 '2017-04-03T05:57:57.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 10,
 'yes',
 8,
 8,
 'no',
 20,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['goats']",
 None,
 1,
 'no',
 'yes',
 "['mobile_phone']",
 2,
 "['Nov', 'Dec']",
 "['lab_ex_food']",
 -19.11218314,
 33.48346325,
 710.0,
 10.0,
 'uuid:4707f3dc-df18-4348-9c2c-eec651e89b6b'),
(18,
 'Moz',
 '21/11/2016',
 18,
 '2017-04-03T12:27:04.000Z',
 '2017-04-03T12:39:48.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 6,
```

```
 'no',
 4,
 4,
 'no',
 20,
 'yes',
 'yes',
 'no',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['goats', 'pigs', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'mobile_phone']",
 2,
 "['Oct', 'Nov']",
 "['reduce_meals', 'lab_ex_food']",
 -19.11133515,
 33.47630848,
 685.0,
 17.0,
 'uuid:7ffe7bd1-a15c-420c-a137-e1f006c317a3'),
(19,
 'Moz',
 '21/11/2016',
 19,
 '2017-04-03T12:40:14.000Z',
 '2017-04-03T12:53:29.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 20,
 'yes',
 9,
 9,
 'no',
 23,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'yes',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'yes',
 "['oxen', 'goats']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_portion', 'lab_ex_food', 'seek_government']",
 -19.11142642,
 33.47640837,
 716.0,
 30.0,
 'uuid:e32f2dc0-0d05-42fb-8e21-605757ddf07d'),
(20,
 'Moz',
 '21/11/2016',
 20,
 '2017-04-03T14:04:50.000Z',
 '2017-04-03T14:20:04.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 24,
 'yes',
 6,
 6,
 'no',
 1,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'burntbricks',
 'earth',
 'yes',
 1,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['cows']",
 None,
 1,
 'yes',
 'yes',
 "['bicycle', 'cow_plough', 'solar_torch']",
 2,
 "['Oct', 'Nov']",
 "['rely_less_food', 'lab_ex_food']",
 -19.11147317,
 33.47619213,
 700.0,
 27.0,
 'uuid:d1005274-bf52-4e79-8380-3350dd7c2bac'),
(21,
 'Moz',
```

```
 '21/11/2016',
 21,
 '2017-04-03T14:24:58.000Z',
 '2017-04-03T14:44:39.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 20,
 'yes',
 8,
 8,
 'no',
 20,
 'no',
 'yes',
 'no',
 'yes',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 1,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'yes',
 "['Oct']",
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 None,
 2,
 "['Jan', 'Feb', 'Mar', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'restrict_adults', 'day_night_hungry', 'lab_ex_food']",
 -19.11155014,
 33.47623134,
 707.0,
 20.0,
 'uuid:6570a7d0-6a0b-452c-aa2e-922500e35749'),
(22,
 'Moz',
 '21/11/2016',
 22,
 '2017-04-03T16:28:52.000Z',
 '2017-04-03T16:40:47.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 14,
 'no',
 4,
 4,
 'no',
 20,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
```

```
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['goats']",
  None,
  1,
  'yes',
  'no',
  "['radio']",
  2,
  "['Jan', 'Feb', 'Mar', 'Apr', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'reduce_meals', 'go_forest', 'lab_ex_food']",
  -19.11212691,
  33.48350764,
  722.0,
  9.0,
  'uuid:a51c3006-8847-46ff-9d4e-d29919b8ecf9'),
 (23,
  'Moz',
  '21/11/2016',
  23,
  '2017-04-03T16:41:04.000Z',
  '2017-04-03T17:04:28.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  12,
  'yes',
  10,
  10,
  'no',
  20,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  4,
  'yes',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'bicycle', 'television', 'radio', 'cow_plough', 'solar_panel',
'electricity', 'mobile_phone']",
```

```
 3,
 "['none']",
 "['na']",
 -19.11219352,
 33.4833833,
 699.0,
 10.0,
 'uuid:58b37b6d-d6cd-4414-8790-b9c68bca98de'),
(24,
 'Moz',
 '21/11/2016',
 24,
 '2017-04-03T17:19:49.000Z',
 '2017-04-03T17:43:01.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 8,
 'no',
 6,
 6,
 'no',
 4,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 2,
 2,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Sept', 'Oct']",
 8.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'no',
 'no',
 "['radio', 'table', 'sofa_set', 'mobile_phone']",
 2,
 "['Nov', 'Dec']",
 "['lab_ex_food']",
 -19.11218803,
 33.48343475,
 745.0,
 11.0,
 'uuid:661457d3-7e61-45e8-a238-7415e7548f82'),
(25,
 'Moz',
 '21/11/2016',
 25,
 '2017-04-04T04:01:58.000Z',
 '2017-04-04T04:29:47.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 10,
 'yes',
 11,
 11,
 'no',
 6,
 'no',
 'no',
```

```
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  3,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept']",
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'yes',
  "['business', 'non_farm_prod']",
  None,
  'yes',
  'no',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'television', 'radio', 'cow_plough', 'solar_panel',
'solar_torch', 'table', 'sofa_set', 'mobile_phone']",
  2,
  "['Jan', 'Feb', 'Oct']",
  "['na']",
  -19.11223416,
  33.4834841,
  698.0,
  11.0,
  'uuid:45ed84c4-114e-4df0-9f5d-c800806c2bee'),
 (26,
  'Moz',
  '21/11/2016',
  26,
  '2017-04-04T04:30:19.000Z',
  '2017-04-04T04:44:19.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  2,
  'yes',
  3,
  3,
  'no',
  20,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  2,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Oct', 'Nov']",
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
```

```
 None,
 None,
 'yes',
 'yes',
 "['oxen', 'cows']",
 None,
 2,
 'yes',
 'no',
 "['radio', 'cow_plough', 'table', 'mobile_phone']",
 2,
 "['none']",
 "['na']",
 -19.11230411,
 33.48354341,
 706.0,
 12.0,
 'uuid:1c54ee24-22c4-4ee9-b1ad-42d483c08e2e'),
(27,
 'Moz',
 '21/11/2016',
 27,
 '2017-04-05T04:59:42.000Z',
 '2017-04-05T05:14:45.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 36,
 'no',
 7,
 7,
 'no',
 36,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 3,
 2,
 'yes',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.043000699999997,
 33.40508367,
 679.0,
 14.0,
 'uuid:3197cded-1fdc-4c0c-9b10-cfcc0bf49c4d'),
(28,
 'Moz',
 '21/11/2016',
 28,
 '2017-04-05T05:14:49.000Z',
 '2017-04-05T05:36:18.000Z',
 'Manica',
 'Manica',
```

```
 'Bandula',
 'Ruaca',
 2,
 'no',
 2,
 2,
 'no',
 2,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov']",
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 3,
 "['Aug', 'Sept', 'Oct']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'go_forest', 'lab_ex_food']",
 -19.04290893,
 33.40506932,
 721.0,
 7.0,
 'uuid:1de53318-a8cf-4736-99b1-8239f8822473'),
(29,
 'Moz',
 '21/11/2016',
 29,
 '2017-04-05T05:37:30.000Z',
 '2017-04-05T06:05:44.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 10,
 'yes',
 7,
 7,
 'no',
 10,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
```

```
  4.0,
  'yes',
  'yes',
  "['less_work']",
  'yes',
  'yes',
  'no',
  'frequently',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'radio', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb']",
  "['rely_less_food', 'limit_variety']",
  -19.04302413,
  33.40507276,
  657.0,
  6.0,
  'uuid:adcd7463-8943-4c67-b25f-f72311409476'),
 (30,
  'Moz',
  '21/11/2016',
  30,
  '2017-04-05T06:05:58.000Z',
  '2017-04-05T06:20:39.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  22,
  'yes',
  7,
  7,
  'yes',
  22,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  2,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  2,
  "['Jan', 'Feb']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'restrict_adults',
'lab_ex_food']",
  -19.04300478,
  33.40505449,
  669.0,
```

```
 5.0,
 'uuid:59341ead-92be-45a9-8545-6edf9f94fdc6'),
(31,
 'Moz',
 '21/11/2016',
 31,
 '2017-04-05T06:21:20.000Z',
 '2017-04-05T06:38:26.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 15,
 'yes',
 3,
 3,
 'no',
 2,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 7,
 1,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 3,
 "['none']",
 "['na']",
 -19.04302176,
 33.40509382,
 704.0,
 4.0,
 'uuid:cb06eb49-dd39-4150-8bbe-a599e074afe8'),
(32,
 'Moz',
 '21/11/2016',
 32,
 '2017-04-05T06:38:55.000Z',
 '2017-04-05T08:05:32.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 53,
 'yes',
 19,
 19,
 'yes',
 69,
 'yes',
 'yes',
 'yes',
 'no',
 'mabatipitched',
 'muddaub',
 'earth',
 'no',
```

```
 8,
 2,
 'yes',
 8,
 8,
 'yes',
 None,
 8.0,
 'yes',
 "['Sept', 'Oct']",
 45.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'more_once',
 'yes',
 "['farming']",
 None,
 'yes',
 'yes',
 "['oxen', 'cows', 'goats', 'pigs', 'poultry']",
 None,
 5,
 'yes',
 'no',
 "['cow_cart', 'motorcyle', 'radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
 2,
 "['none']",
 "['na']",
 -19.04411399,
 33.40390565,
 703.0,
 8.0,
 'uuid:25597af3-cd79-449c-a48a-fb9aea6c48bf'),
(33,
 'Moz',
 '21/11/2016',
 33,
 '2017-04-05T08:08:19.000Z',
 '2017-04-05T08:25:48.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 20,
 'yes',
 8,
 8,
 'no',
 34,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'cement',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'poultry']",
 None,
 2,
```

```
  'yes',
  'no',
  "['cow_cart', 'lorry', 'motorcyle', 'sterio', 'cow_plough', 'solar_panel',
'mobile_phone']",
  2,
  "['none']",
  "['na']",
  -19.04414887,
  33.40383602,
  695.0,
  5.0,
  'uuid:0fbd2df1-2640-4550-9fbd-7317feaa4758'),
 (34,
  'Moz',
  '17/11/2016',
  34,
  '2017-04-05T16:00:47.000Z',
  '2017-04-05T16:21:59.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  18,
  'yes',
  8,
  8,
  'no',
  18,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  2,
  3,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Oct', 'Nov']",
  2.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['television', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'table',
'mobile_phone']",
  2,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'restrict_adults',
'lab_ex_food']",
  -19.11219065,
  33.48341559,
  706.0,
  11.0,
  'uuid:14c78c45-a7cc-4b2a-b765-17c82b43feb4'),
 (35,
  'Moz',
  '17/11/2016',
  35,
  '2017-04-05T16:22:13.000Z',
  '2017-04-05T16:50:25.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  45,
  'yes',
```

```
  5,
  5,
  'no',
  45,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  3,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  20.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'cow_plough']",
  3,
  "['Jan', 'Sept', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'restrict_adults', 'go_forest']",
  -19.11211362,
  33.48342515,
  733.0,
  11.0,
  'uuid:ff7496e7-984a-47d3-a8a1-13618b5683ce'),
 (36,
  'Moz',
  '17/11/2016',
  36,
  '2017-04-05T16:50:48.000Z',
  '2017-04-05T17:10:53.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  23,
  'yes',
  6,
  6,
  'no',
  23,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'sunbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  23.0,
  'yes',
  'no',
```

```
  None,
  'yes',
  'no',
  'no',
  'once',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['oxen', 'cows', 'pigs']",
  None,
  3,
  'no',
  'no',
  "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.11218058,
  33.4833843,
  710.0,
  10.0,
  'uuid:c90eade0-1148-4a12-8c0e-6387a36f45b1'),
 (37,
  'Moz',
  '17/11/2016',
  37,
  '2017-04-05T17:17:48.000Z',
  '2017-04-05T17:26:51.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  8,
  'no',
  3,
  3,
  'no',
  8,
  'yes',
  'yes',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  3,
  1,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'television', 'radio', 'cow_plough', 'solar_panel', 'solar_torch',
'mobile_phone']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'lab_ex_food']",
  -19.11217951,
  33.48339525,
  711.0,
  12.0,
  'uuid:408c6c93-d723-45ef-8dee-1b1bd3fe20cd'),
 (38,
```

```
 'Moz',
 '17/11/2016',
 38,
 '2017-04-05T17:28:12.000Z',
 '2017-04-05T17:50:57.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 19,
 'yes',
 10,
 10,
 'yes',
 19,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct']",
 9.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'yes',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'table', 'mobile_phone']",
 3,
 "['Nov']",
 "['limit_variety', 'lab_ex_food']",
 -19.11222939,
 33.48337467,
 696.0,
 9.0,
 'uuid:81309594-ff58-4dc1-83a7-72af5952ee08'),
(39,
 'Moz',
 '17/11/2016',
 39,
 '2017-04-06T08:31:17.000Z',
 '2017-04-06T08:44:47.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 22,
 'yes',
 6,
 6,
 'no',
 22,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
```

```
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen']",
 None,
 1,
 'yes',
 'no',
 None,
 3,
 "['Nov']",
 "['lab_ex_food']",
 -19.043361800000003,
 33.404667100000005,
 0.0,
 20.0,
 'uuid:c0fb6310-55af-4831-ae3d-2729556c3285'),
(40,
 'Moz',
 '17/11/2016',
 40,
 '2017-04-06T08:44:51.000Z',
 '2017-04-06T09:03:47.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 23,
 'yes',
 9,
 9,
 'yes',
 23,
 'no',
 'yes',
 'yes',
 'yes',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct']",
 23.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'table', 'mobile_phone']",
```

```
 3,
 "['Sept', 'Oct', 'Nov']",
 "['lab_ex_food']",
 -19.043361800000003,
 33.404667100000005,
 0.0,
 22.112,
 'uuid:c0b34854-eede-4e81-b183-ef58a45bfc34'),
(41,
 'Moz',
 '17/11/2016',
 41,
 '2017-04-06T09:03:50.000Z',
 '2017-04-06T09:14:05.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 22,
 'yes',
 7,
 7,
 'no',
 22,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['motorcyle', 'bicycle', 'radio', 'cow_plough', 'table']",
 3,
 "['Oct', 'Nov']",
 "['limit_portion', 'reduce_meals']",
 -19.04339398,
 33.40485363,
 679.0,
 13.0,
 'uuid:b3ba34d8-eea1-453d-bc73-c141bcbbc5e5'),
(42,
 'Moz',
 '17/11/2016',
 42,
 '2017-04-06T09:14:22.000Z',
 '2017-04-06T09:30:54.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 8,
 'yes',
 8,
 8,
 'no',
 8,
 'no',
 'no',
```

```
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'yes',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept']",
 5.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['cows', 'goats', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.04346041,
 33.40484617,
 699.0,
 10.0,
 'uuid:e3a1dd8a-1bda-428c-a014-2b527f11ae64'),
(43,
 'Moz',
 '17/11/2016',
 43,
 '2017-04-06T09:31:56.000Z',
 '2017-04-06T09:53:53.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 3,
 'no',
 7,
 7,
 'no',
 29,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
```

```
 None,
 'yes',
 'no',
 "['oxen', 'cows']",
 None,
 2,
 'no',
 'no',
 "['cow_plough', 'mobile_phone']",
 2,
 "['Jan', 'Feb', 'Oct', 'Nov', 'Dec']",
 "['lab_ex_food']",
 -19.04303063,
 33.40472726,
 605.0,
 30.0,
 'uuid:b4dff49f-ef27-40e5-a9d1-acf287b47358'),
(44,
 'Moz',
 '17/11/2016',
 44,
 '2017-04-06T14:44:32.000Z',
 '2017-04-06T14:53:01.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 3,
 'no',
 2,
 2,
 'no',
 6,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['oxen', 'cows', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['radio', 'solar_torch']",
 2,
 "['Jan', 'Dec']",
 "['borrow_food']",
 -19.04315107,
 33.40458039,
 716.0,
 11.0,
 'uuid:f9fadf44-d040-4fca-86c1-2835f79c4952'),
(45,
 'Moz',
 '17/11/2016',
 45,
 '2017-04-06T14:53:04.000Z',
 '2017-04-06T15:11:57.000Z',
 'Manica',
 'Manica',
 'Bandula',
```

'Chirodzo',
25,
'yes',
9,
9,
'no',
7,
'no',
'no',
'no',
'no',
'grass',
'muddaub',
'earth',
'no',
2,
1,
'no',
3,
3,
'yes',
None,
3.0,
'yes',
"['Sept', 'Oct']",
20.0,
'yes',
'no',
None,
'no',
None,
'no',
'never',
'yes',
"['farming', 'non_farm_prod']",
None,
'yes',
'no',
"['oxen', 'cows', 'goats', 'poultry']",
None,
4,
'no',
'no',
"['motorcyle', 'bicycle', 'television', 'radio', 'cow_plough', 'solar_panel',
'solar_torch', 'table', 'mobile_phone']",
3,
"['none']",
"['na']",
-19.04312371,
33.40466493,
703.0,
28.0,
'uuid:e3554d22-35b1-4fb9-b386-dd5866ad5792'),
(46,
'Moz',
'17/11/2016',
46,
'2017-04-06T15:19:41.000Z',
'2017-04-06T15:45:32.000Z',
'Manica',
'Manica',
'Bandula',
'Chirodzo',
21,
'yes',
10,
10,
'no',
42,
'yes',
'yes',
'yes',
'yes',
'mabatisloping',
'burntbricks',
'cement',
'no',
3,
2,
'no',
5,
5,
'yes',
None,
5.0,
'yes',
"['Aug', 'Sept', 'Oct']",

```
  21.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'once',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'poultry']",
  None,
  2,
  'no',
  'no',
  "['motorcyle', 'computer', 'television', 'sterio', 'solar_panel', 'solar_torch',
'table', 'mobile_phone']",
  2,
  "['Sept', 'Oct', 'Nov']",
  "['lab_ex_food']",
  -19.04307032,
  33.40458421,
  703.0,
  5.0,
  'uuid:35f297e0-aa5d-4149-9b7b-4965004cfc37'),
 (47,
  'Moz',
  '17/11/2016',
  47,
  '2017-04-07T14:05:25.000Z',
  '2017-04-07T14:19:45.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  2,
  'yes',
  2,
  2,
  'no',
  2,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  2.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'once',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['solar_torch', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.11226093,
  33.48339791,
  689.0,
```

```
 5.0,
 'uuid:2d0b1936-4f82-4ec3-a3b5-7c3c8cd6cc2b'),
(48,
 'Moz',
 '16/11/2016',
 48,
 '2017-04-07T14:19:49.000Z',
 '2017-04-07T14:40:23.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 48,
 'yes',
 7,
 7,
 'no',
 58,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['radio']",
 3,
 "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
 "['limit_variety']",
 -19.11222978,
 33.48353345,
 689.0,
 12.0,
 'uuid:e180899c-7614-49eb-a97c-40ed013a38a2'),
(49,
 'Moz',
 '16/11/2016',
 49,
 '2017-04-07T14:43:09.000Z',
 '2017-04-07T14:55:51.000Z',
 'Manica',
 'Manica',
 'Bandula',
 '49',
 12,
 'no',
 6,
 6,
 'no',
 26,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'yes',
```

```
  2,
  2,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['goats', 'poultry']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'table',
'mobile_phone']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['reduce_meals', 'go_forest', 'lab_ex_food']",
  -19.11228265,
  33.48334844,
  694.0,
  12.0,
  'uuid:2303ebc1-2b3c-475a-8916-b322ebf18440'),
 (50,
  'Moz',
  '16/11/2016',
  50,
  '2017-04-07T14:56:01.000Z',
  '2017-04-07T15:26:23.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  6,
  'yes',
  6,
  6,
  'no',
  7,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  1,
  1,
  'yes',
  None,
  1.0,
  'yes',
  "['Sept', 'Oct']",
  1.0,
  'no',
  'no',
  None,
  'yes',
  'no',
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['poultry']",
  None,
```

```
 1,
 'yes',
 'no',
 "['solar_torch']",
 2,
 "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
 "['borrow_food', 'go_forest', 'lab_ex_food']",
 -19.11220496,
 33.48344521,
 718.0,
 12.0,
 'uuid:4267c33c-53a7-46d9-8bd6-b96f58a4f92c'),
(51,
 'Moz',
 '16/11/2016',
 51,
 '2017-04-07T15:27:45.000Z',
 '2017-04-07T15:39:10.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 11,
 'yes',
 5,
 5,
 'no',
 30,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'cement',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['poultry']",
 None,
 1,
 'no',
 'no',
 "['radio']",
 3,
 "['Oct', 'Nov']",
 "['go_forest', 'lab_ex_food']",
 -19.11221446,
 33.48338443,
 709.0,
 12.0,
 'uuid:18ac8e77-bdaf-47ab-85a2-e4c947c9d3ce'),
(52,
 'Moz',
 '16/11/2016',
 52,
 '2017-04-08T04:44:09.000Z',
 '2017-04-08T05:02:58.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 15,
 'yes',
 11,
 11,
```

```
  'no',
  15,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  3,
  'yes',
  1,
  1,
  'yes',
  None,
  1.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  15.0,
  'no',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'poultry']",
  None,
  3,
  'yes',
  'no',
  "['motorcyle', 'television', 'radio', 'cow_plough', 'solar_panel',
'mobile_phone']",
  3,
  "['Aug', 'Sept', 'Oct', 'Nov']",
  "['limit_variety', 'reduce_meals']",
  -19.11223637,
  33.48335287,
  694.0,
  10.0,
  'uuid:6db55cb4-a853-4000-9555-757b7fae2bcf'),
 (53,
  'Moz',
  '16/11/2016',
  21,
  '2017-04-08T05:03:08.000Z',
  '2017-04-08T05:33:51.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  16,
  'yes',
  8,
  8,
  'yes',
  16,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  3,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept', 'Oct']",
  16.0,
  'yes',
  'no',
  None,
  'yes',
```

```
 'no',
 'no',
 'frequently',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'goats']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'mobile_phone']",
 2,
 "['Nov']",
 "['rely_less_food', 'limit_portion']",
 -19.11216571,
 33.48343865,
 687.0,
 10.0,
 'uuid:cc7f75c5-d13e-43f3-97e5-4f4c03cb4b12'),
(54,
 'Moz',
 '16/11/2016',
 54,
 '2017-04-08T05:36:55.000Z',
 '2017-04-08T05:52:15.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 10,
 'no',
 7,
 7,
 'yes',
 15,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 2,
 "['Sept', 'Oct', 'Nov']",
 "['rely_less_food', 'lab_ex_food']",
 -19.11220247,
 33.48335527,
 681.0,
 9.0,
 'uuid:273ab27f-9be3-4f3b-83c9-d3e1592de919'),
(55,
 'Moz',
 '16/11/2016',
 55,
```

```
 '2017-04-08T05:52:32.000Z',
 '2017-04-08T06:05:41.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 23,
 'yes',
 9,
 9,
 'no',
 23,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 2,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['television', 'cow_plough', 'mobile_phone']",
 2,
 "['Oct', 'Nov']",
 "['lab_ex_food']",
 -19.11228974,
 33.48333393,
 702.0,
 11.0,
 'uuid:883c0433-9891-4121-bc63-744f082c1fa0'),
(56,
 'Moz',
 '16/11/2016',
 56,
 '2017-04-08T06:05:59.000Z',
 '2017-04-08T06:26:12.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 23,
 'yes',
 12,
 12,
 'no',
 23,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 4,
 2,
 'no',
 2,
 2,
 'yes',
```

```
 None,
 2.0,
 'yes',
 "['Sept', 'Oct']",
 23.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'goats']",
 None,
 2,
 'yes',
 'no',
 "['motorcyle', 'bicycle', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11217264,
 33.48341837,
 689.0,
 10.0,
 'uuid:973c4ac6-f887-48e7-aeaf-4476f2cfab76'),
(57,
 'Moz',
 '16/11/2016',
 57,
 '2017-04-08T06:26:22.000Z',
 '2017-04-08T06:39:40.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 20,
 'yes',
 4,
 4,
 'no',
 27,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'burntbricks',
 'earth',
 'no',
 4,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['radio']",
 2,
 "['none']",
 "['na']",
```

```
  -19.11227947,
  33.48338576,
  695.0,
  10.0,
  'uuid:a7184e55-0615-492d-9835-8f44f3b03a71'),
 (58,
  'Moz',
  '16/11/2016',
  58,
  '2017-04-08T08:25:49.000Z',
  '2017-04-08T08:48:51.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  35,
  'yes',
  11,
  11,
  'no',
  45,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  5,
  3,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  35.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'no',
  'no',
  "['motorcyle', 'bicycle', 'television', 'radio', 'cow_plough', 'solar_panel',
'mobile_phone']",
  2,
  "['none']",
  "['na']",
  -19.1121758,
  33.48332076,
  723.0,
  11.0,
  'uuid:a7a3451f-cd0d-4027-82d9-8dcd1234fcca'),
 (59,
  'Moz',
  '16/11/2016',
  59,
  '2017-04-08T08:52:05.000Z',
  '2017-04-08T09:02:34.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  60,
  'no',
  2,
  2,
  'no',
  60,
  'yes',
  'yes',
  'yes',
  'yes',
```

```
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 3,
 'yes',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'poultry']",
 None,
 3,
 'no',
 'no',
 None,
 2,
 "['none']",
 "['na']",
 -19.112339499999997,
 33.48333251,
 683.0,
 13.0,
 'uuid:1936db62-5732-45dc-98ff-9b3ac7a22518'),
(60,
 'Moz',
 '16/11/2016',
 60,
 '2017-04-08T09:03:01.000Z',
 '2017-04-08T09:20:18.000Z',
 'Manica',
 'Bandula',
 'Bandula',
 'Chirodzo',
 12,
 'yes',
 8,
 8,
 'no',
 15,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'burntbricks',
 'earth',
 'no',
 3,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 12.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
```

```
  'yes',
  "['oxen', 'cows', 'goats', 'poultry']",
  None,
  4,
  'yes',
  'no',
  "['cow_plough']",
  2,
  "['none']",
  "['na']",
  -19.11225763,
  33.48341208,
  694.0,
  11.0,
  'uuid:85465caf-23e4-4283-bb72-a0ef30e30176'),
 (61,
  'Moz',
  '16/11/2016',
  61,
  '2017-04-08T10:47:11.000Z',
  '2017-04-08T11:14:09.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  14,
  'yes',
  10,
  10,
  'no',
  14,
  'no',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  4,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Aug', 'Sept', 'Oct']",
  13.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'no',
  'more_once',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_panel', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults']",
  -19.11218035,
  33.48341635,
  712.0,
  13.0,
  'uuid:2401cf50-8859-44d9-bd14-1bf9128766f2'),
 (62,
  'Moz',
  '16/11/2016',
  62,
  '2017-04-08T13:27:58.000Z',
  '2017-04-08T13:41:21.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
```

```
  5,
  'no',
  5,
  5,
  'no',
  5,
  'yes',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  3,
  1,
  'no',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  3,
  "['Aug', 'Sept', 'Oct', 'Nov']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'restrict_adults', 'borrow_food', 'go_forest', 'lab_ex_food']",
  -19.11216869,
  33.48339699,
  719.0,
  18.0,
  'uuid:c6597ecc-cc2a-4c35-a6dc-e62c71b345d6'),
 (63,
  'Moz',
  '16/11/2016',
  63,
  '2017-04-08T13:41:39.000Z',
  '2017-04-08T13:52:07.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  1,
  'yes',
  4,
  4,
  'no',
  10,
  'yes',
  'yes',
  'no',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'yes',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
```

```
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  None,
  3,
  "['Jan', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults']",
  -19.11220024,
  33.483390299999996,
  702.0,
  25.0,
  'uuid:86ed4328-7688-462f-aac7-d6518414526a'),
 (64,
  'Moz',
  '16/11/2016',
  64,
  '2017-04-08T13:52:30.000Z',
  '2017-04-08T14:02:24.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  1,
  'no',
  6,
  6,
  'no',
  1,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'no',
  "['goats']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'solar_torch', 'table', 'sofa_set', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'lab_ex_food']",
  -19.1121962,
  33.48339576,
  704.0,
  9.0,
```

```
 'uuid:28cfd718-bf62-4d90-8100-55fafbe45d06'),
(65,
 'Moz',
 '16/11/2016',
 65,
 '2017-04-08T14:02:49.000Z',
 '2017-04-08T14:15:45.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 20,
 'no',
 8,
 8,
 'no',
 20,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 2,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['motorcyle', 'radio', 'cow_plough', 'table']",
 3,
 "['Jan', 'Feb', 'Mar']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.11221474,
 33.48341209,
 706.0,
 39.0,
 'uuid:143f7478-0126-4fbc-86e0-5d324339206b'),
(66,
 'Moz',
 '16/11/2016',
 66,
 '2017-04-08T21:09:38.000Z',
 '2017-04-08T21:33:45.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 16,
 'yes',
 10,
 10,
 'no',
 37,
 'yes',
 'no',
 'no',
 'no',
 'mabatipitched',
 'burntbricks',
 'cement',
 'yes',
 5,
```

```
  3,
  'yes',
  1,
  1,
  'yes',
  None,
  1.0,
  'yes',
  "['Aug', 'Sept', 'Oct']",
  10.0,
  'no',
  'no',
  None,
  'yes',
  'no',
  'no',
  'frequently',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats', 'donkeys']",
  None,
  4,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_panel', 'solar_torch', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.112172199999996,
  33.48339142,
  702.0,
  8.0,
  'uuid:a457eab8-971b-4417-a971-2e55b8702816'),
 (67,
  'Moz',
  '16/11/2016',
  67,
  '2017-04-08T21:34:23.000Z',
  '2017-04-08T21:49:02.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  9,
  'yes',
  5,
  5,
  'no',
  31,
  'yes',
  'no',
  'yes',
  'no',
  'mabatipitched',
  'burntbricks',
  'cement',
  'yes',
  1,
  2,
  'no',
  1,
  1,
  'yes',
  None,
  1.0,
  'yes',
  "['Oct', 'Nov']",
  9.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'more_once',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['oxen', 'cows', 'goats', 'donkeys']",
  None,
  4,
```

```
 'yes',
 'no',
 "['motorcyle', 'radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11209398,
 33.48338805,
 717.0,
 11.0,
 'uuid:6c15d667-2860-47e3-a5e7-7f679271e419'),
(68,
 'Moz',
 '16/11/2016',
 68,
 '2017-04-08T21:49:40.000Z',
 '2017-04-09T22:06:57.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 21,
 'yes',
 8,
 8,
 'no',
 52,
 'yes',
 'yes',
 'no',
 'no',
 'mabatipitched',
 'burntbricks',
 'earth',
 'no',
 3,
 3,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Nov', 'Dec']",
 21.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['motorcyle', 'television', 'sterio', 'solar_panel', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11214985,
 33.4834782,
 710.0,
 10.0,
 'uuid:ef04b3eb-b47d-412e-9b09-4f5e08fc66f9'),
(69,
 'Moz',
 '16/11/2016',
 69,
 '2017-04-09T22:08:07.000Z',
 '2017-04-09T22:21:08.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 12,
 'yes',
 4,
 4,
 'no',
```

```
 12,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['Nov', 'Dec']",
 12.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_torch', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11216693,
 33.48340465,
 708.0,
 8.0,
 'uuid:f86933a5-12b8-4427-b821-43c5b039401d'),
(70,
 'Moz',
 '16/11/2016',
 70,
 '2017-04-09T22:21:23.000Z',
 '2017-04-09T22:40:57.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 20,
 'yes',
 8,
 8,
 'no',
 25,
 'no',
 'yes',
 'no',
 'no',
 'mabatipitched',
 'burntbricks',
 'earth',
 'no',
 2,
 2,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug', 'Sept']",
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
```

```
  'more_once',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats', 'poultry']",
  None,
  4,
  'no',
  'no',
  "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
  2,
  "['none']",
  "['na']",
  -19.11217415,
  33.48341471,
  707.0,
  8.0,
  'uuid:1feb0108-4599-4bf9-8a07-1f5e66a50a0a'),
 (71,
  'Moz',
  '18/11/2016',
  71,
  '2017-04-09T15:00:19.000Z',
  '2017-04-09T15:19:22.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  12,
  'yes',
  6,
  6,
  'no',
  14,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'burntbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  1.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'goats', 'poultry']",
  None,
  3,
  'no',
  'no',
  "['radio', 'cow_plough', 'mobile_phone']",
  2,
  "['Aug', 'Sept', 'Oct', 'Nov']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'restrict_adults', 'borrow_food', 'lab_ex_food']",
  -19.11220603,
  33.48332601,
  696.0,
  4.0,
  'uuid:761f9c49-ec93-4932-ba4c-cc7b78dfcef1'),
 (72,
  'Moz',
  '16/11/2016',
  127,
  '2017-04-09T05:16:06.000Z',
```

```
 '2017-04-09T05:27:41.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Chirodzo',
 10,
 'yes',
 4,
 4,
 'no',
 18,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 3,
 8,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['pigs']",
 None,
 1,
 'no',
 'no',
 "['mobile_phone']",
 2,
 "['Aug', 'Sept', 'Oct']",
 "['borrow_food', 'lab_ex_food']",
 -19.11221396,
 33.48341359,
 676.0,
 8.0,
 'uuid:f6d04b41-b539-4e00-868a-0f62b427587d'),
(73,
 'Moz',
 '23/11/2016',
 133,
 '2017-04-09T05:27:46.000Z',
 '2017-04-09T05:43:51.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 2,
 'no',
 5,
 5,
 'yes',
 25,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'yes',
 1,
 1,
 'yes',
 None,
```

```
  1.0,
  'yes',
  "['Nov']",
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats', 'sheep', 'poultry']",
  None,
  5,
  'yes',
  'no',
  "['cow_cart', 'car', 'lorry', 'motorcyle', 'bicycle', 'television', 'sterio',
'cow_plough', 'solar_panel', 'solar_torch', 'electricity', 'table', 'sofa_set',
'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Oct', 'Nov']",
  "['na']",
  -19.1041288,
  33.4777622,
  0.0,
  2099.999,
  'uuid:429d279a-a519-4dcc-9f64-4673b0fd5d53'),
 (74,
  'Moz',
  '24/11/2016',
  152,
  '2017-04-09T05:47:31.000Z',
  '2017-04-09T06:16:11.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  15,
  'yes',
  10,
  10,
  'no',
  16,
  'yes',
  'no',
  'yes',
  'no',
  'grass',
  'burntbricks',
  'cement',
  'no',
  3,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  16.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'once',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'yes',
  "['motorcyle', 'bicycle', 'radio', 'sterio', 'cow_plough', 'solar_panel',
'mobile_phone']",
  3,
```

```
  "['none']",
  "['na']",
  -19.1121034,
  33.48344669,
  702.0,
  11.0,
  'uuid:59738c17-1cda-49ee-a563-acd76f6bc487'),
 (75,
  'Moz',
  '24/11/2016',
  153,
  '2017-04-09T06:16:49.000Z',
  '2017-04-09T06:28:48.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  41,
  'no',
  5,
  5,
  'no',
  41,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'burntbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['goats']",
  None,
  1,
  'yes',
  'no',
  None,
  2,
  "['Oct', 'Nov']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals', 'no_food',
'day_night_hungry']",
  -19.11223572,
  33.48345393,
  691.0,
  12.0,
  'uuid:7e7961ca-fa1c-4567-9bfa-a02f876e4e03'),
 (76,
  'Moz',
  '24/11/2016',
  155,
  '2017-04-09T06:35:16.000Z',
  '2017-04-09T06:48:01.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  5,
  'no',
  4,
  4,
  'no',
  4,
  'no',
  'no',
```

```
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['electricity']",
 2,
 "['Jan', 'Sept', 'Oct', 'Nov', 'Dec']",
 "['limit_variety', 'reduce_meals', 'borrow_food', 'go_forest', 'lab_ex_food']",
 -19.11220708,
 33.48342364,
 713.0,
 13.0,
 'uuid:77b3021b-a9d6-4276-aaeb-5bfcfd413852'),
(77,
 'Moz',
 '25/11/2016',
 178,
 '2017-04-09T06:54:49.000Z',
 '2017-04-09T07:14:16.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 20,
 'yes',
 5,
 5,
 'no',
 79,
 'yes',
 'yes',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Oct', 'Nov']",
 20.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'no',
 'frequently',
 'no',
 None,
```

```
 None,
 'yes',
 'yes',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'no',
 'no',
 "['radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.1123581,
 33.48355586,
 714.0,
 50.0,
 'uuid:2186e2ec-f65a-47cc-9bc1-a0f36dd9591c'),
(78,
 'Moz',
 '25/11/2016',
 177,
 '2017-04-09T07:59:49.000Z',
 '2017-04-09T08:22:34.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 11,
 'yes',
 10,
 10,
 'no',
 13,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Oct', 'Nov']",
 11.0,
 'no',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'yes',
 'no',
 "['oxen', 'cows']",
 None,
 2,
 'yes',
 'no',
 "['motorcyle', 'television', 'cow_plough', 'solar_panel', 'mobile_phone']",
 3,
 "['Nov']",
 "['na']",
 -19.112053699999997,
 33.48349493,
 713.0,
 9.0,
 'uuid:87998c33-c8d2-49ec-9dae-c123735957ec'),
(79,
 'Moz',
 '25/11/2016',
 180,
 '2017-04-09T08:23:05.000Z',
 '2017-04-09T08:42:02.000Z',
 'Manica',
 'Manica',
 'Bandula',
```

```
 'Ruaca',
 4,
 'no',
 7,
 7,
 'no',
 50,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'yes',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov']",
 4.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['cow_plough', 'solar_panel']",
 3,
 "['Oct', 'Nov']",
 "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
 -19.11204921,
 33.48346659,
 701.0,
 4.0,
 'uuid:ece89122-ea99-4378-b67e-a170127ec4e6'),
(80,
 'Moz',
 '25/11/2016',
 181,
 '2017-04-09T08:43:08.000Z',
 '2017-04-09T09:07:48.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 20,
 'yes',
 11,
 11,
 'no',
 25,
 'yes',
 'no',
 'yes',
 'no',
 'mabatipitched',
 'sunbricks',
 'earth',
 'yes',
 4,
 2,
 'yes',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['Sept', 'Oct']",
 23.0,
```

```
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_panel', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.11207501,
  33.48351086,
  701.0,
  5.0,
  'uuid:bf373763-dca5-4906-901b-d1bacb4f0286'),
 (81,
  'Moz',
  '25/11/2016',
  182,
  '2017-04-09T09:08:04.000Z',
  '2017-04-09T09:34:12.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  20,
  'no',
  7,
  7,
  'no',
  21,
  'no',
  'no',
  'no',
  'no',
  'mabatipitched',
  'muddaub',
  'earth',
  'no',
  2,
  3,
  'yes',
  1,
  1,
  'yes',
  None,
  1.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  20.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['solar_panel']",
  3,
  "['Jan', 'Feb', 'Nov', 'Dec']",
  "['na']",
  -19.11204151,
  33.48345447,
  694.0,
  4.0,
```

```
 'uuid:394033e8-a6e2-4e39-bfac-458753a1ed78'),
(82,
 'Moz',
 '28/11/2016',
 186,
 '2017-04-09T15:20:26.000Z',
 '2017-04-09T15:46:14.000Z',
 'Manica',
 'Manica',
 'Manica',
 'God',
 24,
 'no',
 7,
 7,
 'no',
 24,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['Sept', 'Oct']",
 21.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows']",
 None,
 2,
 'no',
 'no',
 "['cow_plough', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11232028,
 33.48346266,
 690.0,
 10.0,
 'uuid:268bfd97-991c-473f-bd51-bc80676c65c6'),
(83,
 'Moz',
 '28/11/2016',
 187,
 '2017-04-09T15:48:14.000Z',
 '2017-04-09T16:12:46.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 1,
 'yes',
 5,
 5,
 'no',
 43,
 'yes',
 'no',
 'no',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
```

```
  2,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  24.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats', 'poultry']",
  None,
  4,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_panel', 'solar_torch', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.1122345,
  33.48349248,
  691.0,
  13.0,
  'uuid:0a42c9ee-a840-4dda-8123-15c1bede5dfc'),
 (84,
  'Moz',
  '28/11/2016',
  195,
  '2017-04-09T16:13:19.000Z',
  '2017-04-09T16:35:24.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  7,
  'yes',
  5,
  5,
  'no',
  48,
  'yes',
  'yes',
  'no',
  'no',
  'grass',
  'burntbricks',
  'earth',
  'no',
  3,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'poultry']",
  None,
  3,
```

```
 'no',
 'no',
 "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'solar_torch']",
 2,
 "['Sept', 'Oct', 'Nov']",
 "['lab_ex_food']",
 -19.11220559,
 33.48342104,
 706.0,
 9.0,
 'uuid:2c132929-9c8f-450a-81ff-367360ce2c19'),
(85,
 'Moz',
 '28/11/2016',
 196,
 '2017-04-09T18:00:41.000Z',
 '2017-04-09T18:31:40.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 21,
 'yes',
 7,
 7,
 'no',
 49,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Oct', 'Nov']",
 21.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['radio', 'cow_plough', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11220357,
 33.48340018,
 694.0,
 10.0,
 'uuid:44e427d1-a448-4bf2-b529-7d67b2266c06'),
(86,
 'Moz',
 '28/11/2016',
 197,
 '2017-04-09T18:32:09.000Z',
 '2017-04-09T19:14:52.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 11,
 'no',
 5,
 5,
 'yes',
```

```
  19,
  'yes',
  'yes',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov']",
  11.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'more_once',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'yes',
  "['bicycle', 'television', 'radio', 'cow_plough', 'solar_torch', 'table',
'mobile_phone']",
  2,
  "['Nov']",
  "['rely_less_food', 'limit_variety', 'reduce_meals']",
  -19.11218716,
  33.4833828,
  714.0,
  11.0,
  'uuid:85c99fd2-775f-40c9-8654-68223f59d091'),
 (87,
  'Moz',
  '28/11/2016',
  198,
  '2017-04-09T19:15:21.000Z',
  '2017-04-09T19:27:56.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  11,
  'no',
  3,
  3,
  'no',
  49,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'burntbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Jan', 'Dec']",
  11.0,
  'yes',
  'no',
  None,
  'no',
  None,
```

```
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 3,
 "['Nov']",
 "['reduce_meals', 'lab_ex_food']",
 -19.11212338,
 33.48338774,
 716.0,
 7.0,
 'uuid:28c64954-739c-444c-a6e0-355878e471c8'),
(88,
 'Moz',
 '21/11/2016',
 201,
 '2017-04-09T19:31:47.000Z',
 '2017-04-09T19:45:38.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 6,
 'yes',
 4,
 4,
 'no',
 6,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['oxen', 'cows']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_torch', 'mobile_phone']",
 2,
 "['Oct', 'Nov', 'Dec']",
 "['borrow_food', 'lab_ex_food']",
 -19.11217226,
 33.48338504,
 685.0,
 10.0,
 'uuid:9e79a31c-3ea5-44f0-80f9-a32db49422e3'),
(89,
 'Moz',
 '17/11/2016',
 202,
 '2017-04-09T19:48:09.000Z',
```

```
  '2017-04-09T20:08:15.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  12,
  'yes',
  12,
  12,
  'no',
  12,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  4,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
  9.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'table',
'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'lab_ex_food']",
  -19.11218863,
  33.48340466,
  705.0,
  5.0,
  'uuid:06d39051-38ef-4757-b68b-3327b1f16b9d'),
 (90,
  'Moz',
  '26/04/2017',
  72,
  '2017-04-26T15:46:24.000Z',
  '2017-04-26T16:13:33.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  24,
  'yes',
  6,
  6,
  'no',
  24,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  3,
  3,
```

```
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov']",
  4.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['bicycle', 'radio', 'cow_plough']",
  2,
  "['Jan', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'lab_ex_food']",
  -19.11221489,
  33.48347358,
  716.0,
  5.0,
  'uuid:c4a2c982-244e-45a5-aa4b-71fa53f99e18'),
 (91,
  'Moz',
  '26/04/2017',
  73,
  '2017-04-26T16:13:50.000Z',
  '2017-04-26T16:45:01.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  6,
  'yes',
  7,
  7,
  'no',
  9,
  'yes',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  2,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  4.0,
  'yes',
  'yes',
  "['cheaper', 'less_work', 'train_outside_org']",
  'yes',
  'no',
  'no',
  'more_once',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_panel', 'table', 'mobile_phone']",
```

```
 3,
 "['Jan', 'Sept', 'Oct']",
 "['rely_less_food', 'limit_variety']",
 -19.11219747,
 33.48341488,
 714.0,
 4.0,
 'uuid:ac3da862-9e6c-4962-94b6-f4c31624f207'),
(92,
 'Moz',
 '26/04/2017',
 76,
 '2017-04-26T16:45:28.000Z',
 '2017-04-26T17:26:21.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 22,
 'yes',
 17,
 17,
 'no',
 48,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatipitched',
 'burntbricks',
 'cement',
 'no',
 5,
 2,
 'no',
 6,
 6,
 'yes',
 None,
 6.0,
 'yes',
 "['Sept', 'Oct']",
 22.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats', 'poultry']",
 None,
 4,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11220193,
 33.48341762,
 702.0,
 9.0,
 'uuid:4178a296-903a-4a8e-9cfa-0cd6143476e8'),
(93,
 'Moz',
 '27/04/2017',
 83,
 '2017-04-27T12:27:31.000Z',
 '2017-04-27T12:56:42.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 17,
 'yes',
 5,
 5,
 'no',
 22,
 'no',
 'no',
```

```
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 17.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows']",
 None,
 2,
 'no',
 'no',
 "['radio', 'cow_plough', 'solar_torch']",
 2,
 "['Aug', 'Sept', 'Oct']",
 "['borrow_food', 'lab_ex_food']",
 -19.11231638,
 33.48349754,
 693.0,
 33.0,
 'uuid:a1e9df00-c8ae-411c-931c-c7df898c68d0'),
(94,
 'Moz',
 '27/04/2017',
 85,
 '2017-04-27T12:58:02.000Z',
 '2017-04-27T13:23:06.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 12,
 'yes',
 7,
 7,
 'no',
 40,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 3,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Sept', 'Oct']",
 12.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
```

```
  None,
  'yes',
  'yes',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['radio', 'cow_plough']",
  2,
  "['Oct', 'Nov']",
  "['na']",
  -19.11235114,
  33.48358508,
  697.0,
  10.0,
  'uuid:4d0f472b-f8ae-4026-87c9-6b5be14b0a70'),
 (95,
  'Moz',
  '27/04/2017',
  89,
  '2017-04-27T16:11:23.000Z',
  '2017-04-27T16:41:41.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  10,
  'no',
  5,
  5,
  'no',
  10,
  'yes',
  'yes',
  'no',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  2,
  2,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct']",
  10.0,
  'no',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'no',
  'no',
  "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'table',
'mobile_phone']",
  3,
  "['Oct', 'Nov']",
  "['rely_less_food', 'limit_variety', 'go_forest', 'lab_ex_food']",
  -19.11231667,
  33.48348329,
  710.0,
  13.0,
  'uuid:b3b309c6-f234-4830-8b30-87d26a17ee1d'),
 (96,
  'Moz',
  '27/04/2017',
  101,
  '2017-04-27T16:42:02.000Z',
  '2017-04-27T18:11:54.000Z',
  'Manica',
  'Manica',
```

```
  'Bandula',
  'God',
  10,
  'no',
  3,
  3,
  'no',
  4,
  'yes',
  'yes',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['goats']",
  None,
  1,
  'no',
  'yes',
  "['bicycle', 'solar_torch']",
  3,
  "['Sept', 'Oct', 'Nov']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'borrow_food', 'lab_ex_food']",
  -19.11223501,
  33.48341767,
  702.0,
  11.0,
  'uuid:3c174acd-e431-4523-9ad6-eb14cddca805'),
 (97,
  'Moz',
  '27/04/2017',
  103,
  '2017-04-27T17:38:53.000Z',
  '2017-04-27T18:09:45.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  50,
  'no',
  6,
  6,
  'no',
  96,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'sunbricks',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
```

```
  "['Aug', 'Sept', 'Oct', 'Nov']",
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats', 'pigs', 'poultry']",
  None,
  5,
  'yes',
  'no',
  "['cow_cart', 'cow_plough', 'solar_panel', 'sofa_set', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['go_forest']",
  -19.11219049,
  33.48337002,
  712.0,
  9.0,
  'uuid:e9d79844-ef14-493b-bbd6-d13691cc660e'),
 (98,
  'Moz',
  '28/04/2017',
  102,
  '2017-04-28T06:27:07.000Z',
  '2017-04-28T06:52:04.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  15,
  'yes',
  12,
  12,
  'no',
  15,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  3,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Aug', 'Sept', 'Oct']",
  2.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'no',
  'frequently',
  'no',
  None,
  None,
  'no',
  'yes',
  "['cows', 'goats']",
  None,
  2,
  'no',
  'no',
  "['cow_plough', 'table', 'sofa_set', 'mobile_phone']",
  3,
  "['Jan', 'Feb']",
  "['lab_ex_food']",
  -19.04413333,
  33.40388996,
  691.0,
```

```
 11.0,
 'uuid:76206b0b-af74-4344-b24f-81e839f0d7b0'),
(99,
 'Moz',
 '28/04/2017',
 78,
 '2017-04-28T07:09:39.000Z',
 '2017-04-28T07:31:38.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 20,
 'no',
 6,
 6,
 'no',
 48,
 'yes',
 'no',
 'yes',
 'no',
 'mabatipitched',
 'burntbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept']",
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows']",
 None,
 2,
 'no',
 'no',
 "['cow_plough']",
 2,
 "['Aug', 'Sept', 'Oct']",
 "['na']",
 -19.04402591,
 33.40398184,
 723.0,
 11.0,
 'uuid:da3fa7cc-5ce9-44fd-9a78-b8982b607515'),
(100,
 'Moz',
 '28/04/2017',
 80,
 '2017-04-28T09:01:47.000Z',
 '2017-04-28T09:25:51.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 12,
 'no',
 5,
 5,
 'no',
 12,
 'no',
 'no',
 'no',
 'no',
 'mabatipitched',
 'muddaub',
 'earth',
 'no',
```

```
 1,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 12.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch']",
 3,
 "['none']",
 "['na']",
 -19.04415001,
 33.40390046,
 689.0,
 6.0,
 'uuid:a85df6df-0336-46fa-a9f4-522bf6f8b438'),
(101,
 'Moz',
 '28/04/2017',
 104,
 '2017-04-28T14:25:13.000Z',
 '2017-04-28T15:18:10.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 35,
 'no',
 14,
 14,
 'no',
 52,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'sunbricks',
 'cement',
 'no',
 4,
 1,
 'yes',
 4,
 4,
 'yes',
 None,
 4.0,
 'yes',
 "['Aug', 'Sept']",
 16.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats', 'pigs']",
 None,
 4,
```

```
 'yes',
 'no',
 "['cow_cart', 'bicycle', 'cow_plough']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['go_forest', 'lab_ex_food']",
 -19.11225606,
 33.4833734,
 711.0,
 9.0,
 'uuid:bb2bb365-7d7d-4fe9-9353-b21269676119'),
(102,
 'Moz',
 '28/04/2017',
 105,
 '2017-04-28T15:32:38.000Z',
 '2017-04-28T15:58:10.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 20,
 'yes',
 6,
 6,
 'no',
 40,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov']",
 1.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'no',
 'frequently',
 'no',
 None,
 None,
 'no',
 'yes',
 "['cows', 'poultry']",
 None,
 2,
 'yes',
 'yes',
 "['motorcyle', 'radio', 'cow_plough', 'solar_panel', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['lab_ex_food']",
 -19.11221203,
 33.48350526,
 735.0,
 11.0,
 'uuid:af0904ee-4fdb-4090-973f-599c81ddf022'),
(103,
 'Moz',
 '30/04/2017',
 106,
 '2017-04-30T05:51:18.000Z',
 '2017-04-30T06:47:01.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 22,
 'yes',
 15,
 15,
 'no',
```

```
  22,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'sunbricks',
  'cement',
  'yes',
  4,
  5,
  'yes',
  5,
  5,
  'yes',
  None,
  5.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov']",
  21.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'radio', 'sterio', 'cow_plough',
'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
  3,
  "['Oct', 'Nov', 'Dec']",
  "['lab_ex_food', 'seek_government']",
  -19.11220965,
  33.48340866,
  709.0,
  5.0,
  'uuid:468797c1-4a65-4f35-9c83-e28ce46972a2'),
 (104,
  'Moz',
  '03/05/2017',
  109,
  '2017-05-03T13:14:43.000Z',
  '2017-05-03T13:37:53.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  12,
  'yes',
  4,
  4,
  'no',
  12,
  'yes',
  'no',
  'yes',
  'no',
  'grass',
  'sunbricks',
  'earth',
  'no',
  1,
  1,
  'yes',
  3,
  3,
  'no',
  3.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
```

```
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'table']",
 3,
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 "['rely_less_food', 'limit_portion', 'reduce_meals']",
 -19.11222402,
 33.48338036,
 699.0,
 10.0,
 'uuid:602cd3f6-4a97-49c6-80e3-bcfd5c78dfa4'),
(105,
 'Moz',
 '03/05/2017',
 110,
 '2017-05-03T13:37:57.000Z',
 '2017-05-03T13:58:06.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 22,
 'no',
 6,
 6,
 'no',
 22,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'sunbricks',
 'cement',
 'no',
 2,
 3,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov']",
 22.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'table', 'mobile_phone']",
 2,
 "['none']",
 "['na']",
 -19.11219756,
 33.48339714,
 704.0,
 9.0,
 'uuid:e7c51ac4-24e4-475e-88e7-f85e896945e3'),
(106,
 'Moz',
 '03/05/2017',
 113,
 '2017-05-03T14:00:13.000Z',
```

```
  '2017-05-03T14:27:03.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  16,
  'no',
  11,
  11,
  'yes',
  26,
  'yes',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  3,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct']",
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats', 'poultry']",
  None,
  4,
  'no',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'radio', 'cow_plough', 'solar_panel',
'solar_torch', 'table', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.11222682,
  33.48340708,
  706.0,
  12.0,
  'uuid:01210861-aba1-4268-98d0-0260e05f5155'),
 (107,
  'Moz',
  '04/05/2017',
  118,
  '2017-05-04T10:26:35.000Z',
  '2017-05-04T10:46:35.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  6,
  'no',
  5,
  5,
  'no',
  25,
  'yes',
  'yes',
  'yes',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  1,
  1,
  'no',
```

```
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['radio', 'solar_torch', 'mobile_phone']",
  3,
  "['Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food',
'lab_ex_food']",
  -19.11220573,
  33.48344178,
  713.0,
  9.0,
  'uuid:77335b2e-8812-4a35-b1e5-ca9ab626dfea'),
 (108,
  'Moz',
  '04/05/2017',
  125,
  '2017-05-04T10:47:05.000Z',
  '2017-05-04T11:16:05.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  7,
  'yes',
  5,
  5,
  'no',
  14,
  'no',
  'yes',
  'no',
  'yes',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov']",
  3.0,
  'yes',
  'yes',
  "['less_work']",
  'no',
  None,
  'no',
  'more_once',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['oxen', 'cows']",
  None,
  2,
  'no',
  'yes',
  "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'mobile_phone']",
  3,
  "['Jan', 'Sept', 'Oct', 'Nov', 'Dec']",
```

```
 "['borrow_food', 'lab_ex_food']",
 -19.11218005,
 33.4834101,
 682.0,
 11.0,
 'uuid:02b05c68-302e-4e7a-b229-81cb1377fd29'),
(109,
 'Moz',
 '04/05/2017',
 119,
 '2017-05-04T11:16:57.000Z',
 '2017-05-04T11:38:38.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca-Nhamuenda',
 14,
 'no',
 3,
 3,
 'yes',
 14,
 'yes',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 4,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['oxen', 'cows', 'goats', 'poultry']",
 None,
 4,
 'no',
 'no',
 "['bicycle', 'cow_plough', 'solar_panel', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11225658,
 33.48334839,
 706.0,
 5.0,
 'uuid:fa201fce-4e94-44b8-b435-c558c2e1ed55'),
(110,
 'Moz',
 '11/05/2017',
 115,
 '2017-05-11T05:24:25.000Z',
 '2017-05-11T05:41:56.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca - Nhamuenda',
 16,
 'no',
 4,
 4,
 'no',
 16,
 'yes',
 'yes',
 'no',
 'no',
```

```
  'mabatisloping',
  'sunbricks',
  'cement',
  'no',
  3,
  2,
  'yes',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.1114691,
  33.4761047,
  0.0,
  20.0,
  'uuid:628fe23d-188f-43e4-a203-a4bf3257d461'),
 (111,
  'Moz',
  '11/05/2017',
  108,
  '2017-05-11T05:42:08.000Z',
  '2017-05-11T06:08:58.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'God',
  22,
  'no',
  15,
  15,
  'no',
  22,
  'no',
  'yes',
  'no',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  3,
  2,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  16.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
```

```
  'no',
  'no',
  "['oxen', 'cows', 'goats', 'poultry']",
  None,
  4,
  'yes',
  'no',
  "['cow_cart', 'bicycle', 'radio', 'cow_plough', 'solar_panel', 'table',
'mobile_phone']",
  3,
  "['Aug', 'Sept', 'Oct', 'Nov']",
  "['rely_less_food', 'limit_portion', 'reduce_meals']",
  -19.1114691,
  33.4761047,
  0.0,
  20.0,
  'uuid:e4f4d6ba-e698-45a5-947f-ba6da88cc22b'),
 (112,
  'Moz',
  '11/05/2017',
  116,
  '2017-05-11T06:09:56.000Z',
  '2017-05-11T06:22:19.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca-Nhamuenda',
  21,
  'yes',
  5,
  5,
  'no',
  25,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'burntbricks',
  'cement',
  'no',
  2,
  3,
  'yes',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'television', 'radio', 'cow_plough', 'solar_panel',
'solar_torch', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
  -19.1114691,
  33.4761047,
  0.0,
  20.0,
  'uuid:cfee6297-2c0e-4f8a-94cc-9aaee0bd64cb'),
 (113,
  'Moz',
  '11/05/2017',
  117,
  '2017-05-11T06:28:02.000Z',
  '2017-05-11T06:55:35.000Z',
  'Manica',
  'Manica',
```

```
  'Bandula',
  'Ruaca-Nhamuenda',
  1,
  'no',
  10,
  10,
  'no',
  28,
  'yes',
  'yes',
  'yes',
  'no',
  'grass',
  'muddaub',
  'cement',
  'no',
  1,
  4,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'television', 'radio', 'solar_panel', 'solar_torch', 'table',
'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'borrow_food', 'lab_ex_food']",
  -19.1114691,
  33.4761047,
  0.0,
  20.0,
  'uuid:3fe626b3-c794-48e1-a80f-5bfe440c507b'),
 (114,
  'Moz',
  '18/05/2017',
  144,
  '2017-05-18T04:36:23.000Z',
  '2017-05-18T05:02:38.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  5,
  'no',
  7,
  7,
  'no',
  5,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  3,
  4,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
```

```
  'yes',
  "['Sept', 'Oct', 'Nov']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'yes',
  'frequently',
  'yes',
  "['farming']",
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats', 'sheep']",
  None,
  4,
  'yes',
  'no',
  "['cow_cart', 'television', 'radio', 'cow_plough', 'solar_panel', 'solar_torch',
'table', 'mobile_phone']",
  2,
  "['none']",
  "['na']",
  -19.11218801,
  33.48342233,
  710.0,
  5.0,
  'uuid:0670cef6-d233-4852-89d8-36955261b0a3'),
 (115,
  'Moz',
  '18/05/2017',
  143,
  '2017-05-18T05:55:04.000Z',
  '2017-05-18T06:37:10.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  24,
  'yes',
  10,
  10,
  'no',
  24,
  'yes',
  'no',
  'no',
  'no',
  'grass',
  'burntbricks',
  'earth',
  'no',
  3,
  2,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  12.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'frequently',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'television', 'radio', 'cow_plough', 'solar_torch',
'table', 'mobile_phone']",
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety']",
```

```
 -19.1124845,
 33.4763322,
 0.0,
 1911.0,
 'uuid:9a096a12-b335-468c-b3cc-1191180d62de'),
(116,
 'Moz',
 '18/05/2017',
 150,
 '2017-05-18T10:37:37.000Z',
 '2017-05-18T10:56:00.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 8,
 'no',
 7,
 7,
 'no',
 8,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Apr', 'May', 'June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
 8.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['cows']",
 None,
 1,
 'no',
 'yes',
 "['mobile_phone']",
 3,
 "['Sept', 'Oct', 'Nov']",
 "['reduce_meals', 'lab_ex_food']",
 -19.11147739,
 33.47618369,
 709.0,
 17.0,
 'uuid:92613d0d-e7b1-4d62-8ea4-451d7cd0a982'),
(117,
 'Moz',
 '18/05/2017',
 159,
 '2017-05-18T10:56:16.000Z',
 '2017-05-18T11:07:35.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 17,
 'yes',
 4,
 4,
 'no',
 24,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
```

```
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['radio', 'solar_panel', 'solar_torch']",
 3,
 "['Sept', 'Oct', 'Nov']",
 "['lab_ex_food']",
 -19.11149498,
 33.47617506,
 712.0,
 28.0,
 'uuid:37577f91-d665-443e-8d70-b914954cef4b'),
(118,
 'Moz',
 '03/06/2017',
 160,
 '2017-06-03T05:08:49.000Z',
 '2017-06-03T05:32:19.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'God',
 3,
 'yes',
 7,
 7,
 'no',
 13,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Jan', 'Dec']",
 3.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'no',
 'frequently',
 'no',
 None,
 None,
 'yes',
 'yes',
```

```
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['cow_cart', 'cow_plough', 'solar_torch', 'mobile_phone']",
  2,
  "['Nov']",
  "['lab_ex_food']",
  -19.11218456,
  33.4834495,
  711.0,
  12.0,
  'uuid:f22831ec-6bc3-4b73-9197-4b01e01abb66'),
 (119,
  'Moz',
  '03/06/2017',
  165,
  '2017-06-03T05:32:33.000Z',
  '2017-06-03T05:51:49.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  14,
  'no',
  9,
  9,
  'no',
  14,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'burntbricks',
  'earth',
  'no',
  1,
  1,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['cow_cart', 'motorcyle', 'bicycle', 'television', 'radio', 'cow_plough',
'solar_torch', 'electricity', 'table', 'sofa_set', 'mobile_phone', 'fridge']",
  3,
  "['none']",
  "['na']",
  -19.11217437,
  33.48346513,
  708.0,
  9.0,
  'uuid:62f3f7af-f0f3-4f88-b9e0-acf8baa49ae4'),
 (120,
  'Moz',
  '03/06/2017',
  166,
  '2017-06-03T05:53:28.000Z',
  '2017-06-03T06:25:06.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  16,
```

```
 'no',
 11,
 11,
 'no',
 16,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_torch', 'mobile_phone']",
 2,
 "['Feb', 'Mar']",
 "['go_forest', 'lab_ex_food']",
 -19.113858899999997,
 33.4826653,
 0.0,
 1799.999,
 'uuid:40aac732-94df-496c-97ba-5b67f59bcc7a'),
(121,
 'Moz',
 '03/06/2017',
 167,
 '2017-06-03T06:25:09.000Z',
 '2017-06-03T06:45:06.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 16,
 'no',
 8,
 8,
 'no',
 24,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 5,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
 16.0,
 'yes',
 'no',
```

```
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'yes',
  "['motorcyle', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'table',
'mobile_phone']",
  2,
  "['Jan', 'Nov', 'Dec']",
  "['lab_ex_food']",
  -19.114988699999998,
  33.4882685,
  0.0,
  2000.0,
  'uuid:a9d1a013-043b-475d-a71b-77ed80abe970'),
 (122,
  'Moz',
  '03/06/2017',
  174,
  '2017-06-03T06:50:47.000Z',
  '2017-06-03T07:20:21.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  13,
  'no',
  12,
  12,
  'no',
  25,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'burntbricks',
  'cement',
  'no',
  2,
  2,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct', 'Nov', 'Dec']",
  13.0,
  'yes',
  'yes',
  "['cheaper']",
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['car', 'lorry', 'motorcyle', 'radio', 'sterio', 'cow_plough', 'solar_panel',
'solar_torch', 'table', 'sofa_set', 'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['lab_ex_food']",
  -19.11216751,
  33.48340539,
  703.0,
  3.0,
  'uuid:43ec6132-478c-4f87-878d-fb3c0c4d0c74'),
```

```
 (123,
  'Moz',
  '03/06/2017',
  175,
  '2017-06-03T07:21:48.000Z',
  '2017-06-03T07:51:29.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  11,
  'no',
  7,
  7,
  'no',
  36,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Oct', 'Nov']",
  11.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['oxen', 'cows', 'goats', 'poultry']",
  None,
  4,
  'yes',
  'yes',
  "['motorcyle', 'bicycle', 'radio', 'sterio', 'cow_plough', 'solar_panel', 'table',
'mobile_phone']",
  2,
  "['Jan', 'Oct', 'Nov', 'Dec']",
  "['na']",
  -19.11217963,
  33.48336019,
  705.0,
  5.0,
  'uuid:64fc743e-8176-40f6-8ae4-36ae97fac1d9'),
 (124,
  'Moz',
  '03/06/2017',
  189,
  '2017-06-03T15:23:01.000Z',
  '2017-06-03T15:53:10.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  10,
  'yes',
  15,
  15,
  'no',
  16,
  'no',
  'yes',
  'no',
  'yes',
  'mabatisloping',
  'sunbricks',
  'earth',
  'no',
  2,
```

```
  1,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
  7.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'yes',
  'no',
  "['motorcyle', 'radio', 'sterio', 'cow_plough', 'solar_panel', 'table',
'mobile_phone']",
  3,
  "['Nov']",
  "['na']",
  -19.11216796,
  33.48341529,
  714.0,
  4.0,
  'uuid:c17e374c-280b-4e78-bf21-74a7c1c73492'),
 (125,
  'Moz',
  '03/06/2017',
  191,
  '2017-06-03T15:54:03.000Z',
  '2017-06-03T16:17:26.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  3,
  'no',
  10,
  10,
  'no',
  5,
  'yes',
  'no',
  'yes',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  4,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  3.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['cows']",
  None,
  1,
```

```
  'yes',
  'no',
  "['radio', 'cow_plough', 'solar_panel', 'solar_torch', 'mobile_phone']",
  2,
  "['Oct', 'Nov', 'Dec']",
  "['go_forest', 'lab_ex_food']",
  -19.11219207,
  33.48338051,
  709.0,
  9.0,
  'uuid:dad53aff-b520-4015-a9e3-f5fdf9168fe1'),
 (126,
  'Moz',
  '03/06/2017',
  192,
  '2017-06-03T16:17:55.000Z',
  '2017-06-03T17:16:39.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  15,
  'yes',
  9,
  9,
  'no',
  20,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'burntbricks',
  'cement',
  'yes',
  4,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Nov']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'once',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'television', 'radio', 'sterio', 'solar_panel', 'solar_torch',
'table', 'mobile_phone']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['borrow_food']",
  -19.11215404,
  33.48335905,
  705.0,
  4.0,
  'uuid:f94409a6-e461-4e4c-a6fb-0072d3d58b00'),
 (127,
  'Moz',
  '18/05/2017',
  126,
  '2017-05-18T04:13:37.000Z',
  '2017-05-18T04:35:47.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  5,
  'yes',
  3,
  3,
```

```
 'no',
 7,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 4.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['oxen', 'cows', 'goats']",
 None,
 3,
 'yes',
 'no',
 "['motorcyle', 'radio', 'solar_panel']",
 3,
 "['Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'lab_ex_food']",
 -19.11219355,
 33.48337856,
 700.0,
 7.0,
 'uuid:69caea81-a4e5-4e8d-83cd-9c18d8e8d965'),
(128,
 'Moz',
 '04/06/2017',
 193,
 '2017-06-04T09:36:20.000Z',
 '2017-06-04T10:13:32.000Z',
 'Manica',
 'Manica',
 'Bandula',
 'Ruaca',
 10,
 'no',
 7,
 7,
 'no',
 10,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'cement',
 'cement',
 'yes',
 3,
 3,
 'yes',
 4,
 4,
 'yes',
 None,
 4.0,
 'yes',
 "['Oct', 'Nov', 'Dec']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
```

```
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'yes',
  "['oxen', 'cows', 'goats']",
  None,
  3,
  'no',
  'no',
  "['car', 'lorry', 'television', 'radio', 'sterio', 'cow_plough', 'solar_torch',
'electricity', 'table', 'sofa_set', 'mobile_phone', 'fridge']",
  3,
  "['none']",
  "['na']",
  -19.11215668,
  33.48339039,
  720.0,
  9.0,
  'uuid:5ccc2e5a-ea90-48b5-8542-69400d5334df'),
 (129,
  'Moz',
  '04/06/2017',
  194,
  '2017-06-04T10:13:36.000Z',
  '2017-06-04T10:32:06.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Ruaca',
  5,
  'no',
  4,
  4,
  'no',
  5,
  'no',
  'no',
  'yes',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Aug', 'Sept', 'Oct']",
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'more_once',
  'no',
  None,
  None,
  'no',
  'yes',
  "['poultry']",
  None,
  1,
  'no',
  'no',
  "['radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
  3,
  "['Sept', 'Oct', 'Nov']",
  "['lab_ex_food']",
  -19.11227133,
  33.48347111,
  719.0,
  10.0,
  'uuid:95c11a30-d44f-40c4-8ea8-ec34fca6bbbf'),
 (130,
  'Moz',
  '04/06/2017',
  199,
```

```
  '2017-06-04T10:33:55.000Z',
  '2017-06-04T10:52:22.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  17,
  'yes',
  7,
  7,
  'no',
  17,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  6.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'no',
  'more_once',
  'yes',
  "['farming', 'business']",
  None,
  'yes',
  'no',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['cow_cart', 'lorry', 'motorcyle', 'computer', 'television', 'radio', 'sterio',
'cow_plough', 'solar_panel', 'solar_torch', 'electricity', 'mobile_phone']",
  3,
  "['Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'lab_ex_food']",
  -19.11227751,
  33.48338952,
  711.0,
  5.0,
  'uuid:ffc83162-ff24-4a87-8709-eff17abc0b3b'),
 (131,
  'Moz',
  '04/06/2017',
  200,
  '2017-06-04T10:52:46.000Z',
  '2017-06-04T11:08:13.000Z',
  'Manica',
  'Manica',
  'Bandula',
  'Chirodzo',
  20,
  'yes',
  8,
  8,
  'no',
  20,
  'no',
  'yes',
  'no',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  2,
  'yes',
  2,
```

```
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['oxen', 'goats', 'pigs']",
 None,
 3,
 'yes',
 'no',
 "['radio', 'cow_plough', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
 3,
 "['Oct', 'Nov']",
 "['rely_less_food', 'restrict_adults', 'borrow_food']",
 -19.11218302,
 33.48337175,
 681.0,
 20.0,
 'uuid:aa77a0d7-7142-41c8-b494-483a5b68d8a7'),
(132,
 'Moz',
 '11/04/2017',
 2,
 '2017-04-11T14:13:11.000Z',
 '2017-04-13T06:46:37.000Z',
 'Nampula',
 'Nampula',
 'Namikopo',
 'Namikopo',
 2,
 'no',
 6,
 6,
 'yes',
 4,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 2,
```

```
 "['Feb', 'Mar', 'Apr']",
 "['rely_less_food', 'limit_portion', 'borrow_food', 'no_food', 'lab_ex_food']",
 -19.11215974,
 33.48339748,
 719.0,
 5.0,
 'uuid:159cb921-0c77-4e1e-aedc-5eba3e9102a3'),
(133,
 'Moz',
 '12/04/2017',
 4,
 '2017-04-12T07:44:51.000Z',
 '2017-04-13T06:45:59.000Z',
 'Nampula',
 'Nampula',
 'Namikopo',
 'Namikopo',
 12,
 'no',
 2,
 2,
 'no',
 65,
 'no',
 'no',
 'no',
 'yes',
 'mabatisloping',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'yes',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 2,
 "['Jan', 'Feb', 'Mar']",
 "['limit_variety']",
 -19.04475097,
 33.40409627,
 739.0,
 49.0,
 'uuid:7a594aa8-231d-4086-b072-d9b3c2ccfb42'),
(134,
 'Moz',
 '17/06/2017',
 1,
 '2017-06-17T06:02:27.000Z',
 '2017-06-17T06:50:50.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 7,
 'yes',
 5,
 5,
 'yes',
 9,
 'yes',
 'yes',
 'yes',
```

```
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 3,
 3,
 'yes',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 7.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['goats', 'pigs']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
 2,
 "['none']",
 "['na']",
 -19.38398448,
 34.36420606,
 23.0,
 4.0,
 'uuid:cf33dea7-7fdb-43f9-8b9f-c5d92d63a1c7'),
(135,
 'Moz',
 '17/06/2017',
 4,
 '2017-06-17T06:53:26.000Z',
 '2017-06-17T07:28:07.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 9,
 'yes',
 7,
 7,
 'no',
 27,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 2.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'no',
 'never',
 'no',
 None,
 None,
```

```
 'no',
 'yes',
 "['sheep']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Oct', 'Nov', 'Dec']",
 "['limit_variety', 'lab_ex_food']",
 -19.38389493,
 34.36418707,
 32.0,
 11.0,
 'uuid:d4934aad-96e2-43c7-8d82-1c14bbd20e0d'),
(136,
 'Moz',
 '17/06/2017',
 9,
 '2017-06-17T07:32:09.000Z',
 '2017-06-17T07:57:24.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 22,
 'yes',
 2,
 2,
 'no',
 22,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['May', 'June', 'July', 'Aug', 'Sept', 'Oct']",
 22.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.38410401,
 34.36431072,
 -4.0,
 30.0,
 'uuid:50e494bc-e0bf-4910-92a8-a7646a93256d'),
(137,
 'Moz',
 '17/06/2017',
 10,
 '2017-06-17T07:57:47.000Z',
 '2017-06-17T08:29:56.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
```

```
  20,
  'yes',
  8,
  8,
  'no',
  20,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  3,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
  20.0,
  'no',
  'no',
  None,
  'yes',
  'no',
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['pigs', 'poultry']",
  None,
  2,
  'no',
  'no',
  "['bicycle', 'radio', 'sterio', 'solar_torch', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food', 'go_forest',
'lab_ex_food']",
  -19.38402658,
  34.36435587,
  -3.0,
  10.0,
  'uuid:242648cb-a33c-44a2-a25d-1b48aaad0236'),
 (138,
  'Moz',
  '17/06/2017',
  3,
  '2017-06-17T08:33:09.000Z',
  '2017-06-17T08:51:21.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato',
  20,
  'no',
  4,
  4,
  'no',
  4,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'burntbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  3.0,
```

```
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['goats']",
 None,
 1,
 'yes',
 'yes',
 "['radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'restrict_adults', 'lab_ex_food']",
 -19.38403877,
 34.36428872,
 3.0,
 12.0,
 'uuid:7de16ed6-a6f8-45ec-ac34-b1b3f5cc09bd'),
(139,
 'Moz',
 '17/06/2017',
 25,
 '2017-06-17T08:51:24.000Z',
 '2017-06-23T14:42:33.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 45,
 'no',
 5,
 5,
 'no',
 65,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['Oct', 'Nov']",
 4.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['solar_panel']",
 3,
 "['Sept', 'Oct', 'Nov']",
 "['limit_portion', 'reduce_meals', 'lab_ex_food']",
 -19.38397881,
 34.36433516,
 28.0,
 8.0,
 'uuid:7ef28793-6419-4763-85a3-6f0b96694b37'),
```

```
(140,
 'Moz',
 '22/06/2017',
 2,
 '2017-06-22T07:32:51.000Z',
 '2017-06-22T08:15:21.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 25,
 'no',
 5,
 5,
 'yes',
 25,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 8.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 2,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'borrow_food', 'lab_ex_food']",
 -19.29882071,
 34.38481958,
 -8.0,
 13.0,
 'uuid:08e76e63-8959-4264-be56-4fff09944483'),
(141,
 'Moz',
 '22/06/2017',
 6,
 '2017-06-22T08:29:23.000Z',
 '2017-06-22T09:05:15.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 20,
 'no',
 8,
 8,
 'no',
 32,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
```

```
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'yes',
 "['farming']",
 None,
 'yes',
 'no',
 "['goats', 'pigs']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_torch']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'borrow_food', 'lab_ex_food']",
 -19.29897236,
 34.38478469,
 16.0,
 5.0,
 'uuid:aaee5983-ca3e-4554-ac9d-9afb13c15a29'),
(142,
 'Moz',
 '22/06/2017',
 7,
 '2017-06-22T09:06:11.000Z',
 '2017-06-22T09:37:32.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 7,
 'yes',
 6,
 6,
 'yes',
 11,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 5,
 5,
 'yes',
 None,
 5.0,
 'no',
 None,
 6.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'yes',
```

```
 "['bicycle', 'radio', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['limit_variety', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.29892635,
 34.38484077,
 32.0,
 5.0,
 'uuid:d659f3b6-b330-4a4f-be95-b6e49ade3278'),
(143,
 'Moz',
 '22/06/2017',
 8,
 '2017-06-22T09:39:40.000Z',
 '2017-06-22T09:56:02.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 7,
 'no',
 5,
 5,
 'no',
 7,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 7.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['sterio', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'borrow_food', 'lab_ex_food']",
 -19.29895979,
 34.38482849,
 43.0,
 5.0,
 'uuid:4694511b-602c-4dc5-8602-14a867c71c08'),
(144,
 'Moz',
 '22/06/2017',
 11,
 '2017-06-22T10:02:35.000Z',
 '2017-06-22T11:42:16.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 12,
 'no',
 8,
 8,
 'yes',
 12,
 'no',
```

```
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Aug', 'Sept', 'Oct']",
  12.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['pigs', 'poultry']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'radio', 'sterio', 'mobile_phone']",
  2,
  "['Jan', 'Feb']",
  "['rely_less_food', 'limit_variety', 'borrow_food', 'day_night_hungry',
'lab_ex_food']",
  -19.40661027,
  34.43743139,
  17.0,
  6.0,
  'uuid:c959b0d3-7ebb-4a10-8cdf-d5badb3be296'),
 (145,
  'Moz',
  '22/06/2017',
  12,
  '2017-06-22T11:48:38.000Z',
  '2017-06-22T12:18:31.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato',
  43,
  'yes',
  6,
  6,
  'no',
  43,
  'no',
  'yes',
  'no',
  'yes',
  'grass',
  'sunbricks',
  'earth',
  'no',
  1,
  3,
  'yes',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
```

```
 None,
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'limit_variety', 'borrow_food', 'go_forest']",
 -19.40658067,
 34.4374628,
 -5.0,
 5.0,
 'uuid:bc54082a-baef-4635-ac72-58e5011a0aa7'),
(146,
 'Moz',
 '23/06/2017',
 13,
 '2017-06-23T06:57:40.000Z',
 '2017-06-23T07:31:36.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 23,
 'yes',
 3,
 3,
 'no',
 23,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'cement',
 'no',
 2,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 23.0,
 'yes',
 'yes',
 "['cheaper']",
 'yes',
 'yes',
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['poultry', 'none']",
 None,
 2,
 'yes',
 'no',
 "['television', 'tractor', 'solar_panel', 'mobile_phone']",
 2,
 "['Jan', 'Feb']",
 "['limit_variety']",
 -19.287941699999998,
 34.20577802,
 62.0,
 4.0,
 'uuid:9b7d8a74-1586-49fb-9783-cd8752442ac5'),
(147,
 'Moz',
 '23/06/2017',
 14,
 '2017-06-23T07:34:58.000Z',
 '2017-06-23T07:58:36.000Z',
 'Sofala',
```

```
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 42,
 'yes',
 7,
 7,
 'no',
 42,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug']",
 22.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['motorcyle', 'radio', 'sterio', 'solar_panel', 'mobile_phone']",
 2,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'go_forest']",
 -19.290154,
 34.20626915,
 60.0,
 5.0,
 'uuid:4c96ef7f-e5f8-429a-9593-93cb4663b0cf'),
(148,
 'Moz',
 '23/06/2017',
 15,
 '2017-06-23T08:01:44.000Z',
 '2017-06-23T08:34:34.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 42,
 'yes',
 7,
 7,
 'no',
 42,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
```

```
 "['June', 'July', 'Aug', 'Sept', 'Oct']",
 42.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_panel', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'limit_portion', 'lab_ex_food']",
 -19.27281975,
 34.20381567,
 75.0,
 5.0,
 'uuid:d59a5dfa-65a0-45c6-b2f2-43c0e563674a'),
(149,
 'Moz',
 '23/06/2017',
 16,
 '2017-06-23T08:48:54.000Z',
 '2017-06-23T09:19:03.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 25,
 'yes',
 5,
 5,
 'no',
 25,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 2,
 4,
 'yes',
 4,
 4,
 'no',
 4.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.30333993,
 34.25722303,
 45.0,
```

```
 5.0,
 'uuid:64589d7b-4120-4fbd-a27e-425518a7bc6a'),
(150,
 'Moz',
 '23/06/2017',
 5,
 '2017-06-23T09:23:41.000Z',
 '2017-06-24T11:26:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 18,
 'no',
 5,
 5,
 'no',
 22,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 4.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'lab_ex_food']",
 -19.35829038,
 34.29253063,
 3.0,
 5.0,
 'uuid:fc0d05dd-6da5-4c2c-8682-9b8ee61486d9'),
(151,
 'Moz',
 '23/06/2017',
 17,
 '2017-06-23T09:50:58.000Z',
 '2017-06-23T10:10:02.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 11,
 'yes',
 5,
 5,
 'no',
 11,
 'no',
 'yes',
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
```

```
 1,
 2,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['radio']",
 3,
 "['Sept', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.373838600000003,
 34.253811600000006,
 19.0,
 5.0,
 'uuid:30781fce-6e6b-404b-80aa-d3db1b99094c'),
(152,
 'Moz',
 '23/06/2017',
 18,
 '2017-06-23T10:12:02.000Z',
 '2017-06-23T11:19:22.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 42,
 'yes',
 6,
 6,
 'yes',
 45,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 2,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 21.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
```

```
 'yes',
 'no',
 "['bicycle', 'sterio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'lab_ex_food']",
 -19.36809049,
 34.26296784,
 46.0,
 5.0,
 'uuid:51a4408e-bc77-4c47-9c43-fee305a8ead2'),
(153,
 'Moz',
 '23/06/2017',
 19,
 '2017-06-23T10:36:58.000Z',
 '2017-06-23T10:55:30.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 45,
 'yes',
 2,
 2,
 'no',
 45,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 3,
 "['Jan', 'Feb', 'Mar', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals', 'no_food']",
 -19.32893291,
 34.31076777,
 24.0,
 10.0,
 'uuid:dc799843-449c-42a9-9abb-7d62a982ea8f'),
(154,
 'Moz',
 '23/06/2017',
 20,
 '2017-06-23T10:55:59.000Z',
 '2017-06-23T11:16:19.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 49,
 'yes',
 5,
 5,
 'no',
```

```
 45,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'yes',
 1,
 1,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.28611445,
 34.22582387,
 43.0,
 4.0,
 'uuid:983add41-b0c8-4e9b-aa1e-85debbb240f2'),
(155,
 'Moz',
 '24/06/2017',
 21,
 '2017-06-24T09:14:05.000Z',
 '2017-06-24T09:33:14.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 8,
 'no',
 6,
 6,
 'no',
 8,
 'no',
 'yes',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 3,
 'yes',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['television', 'radio', 'mobile_phone']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.111524199999998,
 33.4760439,
 0.0,
 22.318,
 'uuid:416fdd6d-f0bb-413d-aa40-602ea2c7c146'),
(156,
 'Moz',
 '24/06/2017',
 22,
 '2017-06-24T09:33:20.000Z',
 '2017-06-24T09:45:32.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 7,
 'no',
 5,
 5,
 'no',
 7,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 4.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['solar_panel']",
 3,
 "['Jan', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.111524199999998,
 33.4760439,
 0.0,
 21.953000000000003,
 'uuid:cf678826-6a9e-4ce2-ada1-569d2c6939c1'),
(157,
 'Moz',
 '24/06/2017',
 23,
 '2017-06-24T10:02:09.000Z',
 '2017-06-24T10:24:01.000Z',
```

```
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 50,
 'yes',
 3,
 3,
 'yes',
 50,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 30.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 3,
 "['Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.1115117,
 33.4760548,
 0.0,
 21.976999999999997,
 'uuid:32f32401-eb07-4c11-9b2a-5ab765ecf614'),
(158,
 'Moz',
 '24/06/2017',
 24,
 '2017-06-24T10:24:04.000Z',
 '2017-06-24T11:05:56.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 3,
 'no',
 5,
 5,
 'no',
 3,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
```

```
  'no',
  None,
  3.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  None,
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'lab_ex_food']",
  -19.11154348,
  33.4761323,
  705.0,
  54.0,
  'uuid:da757d90-8245-42e7-bc00-4fe1e352eb7b'),
 (159,
  'Moz',
  '17/06/2017',
  26,
  '2017-06-17T09:17:09.000Z',
  '2017-06-22T09:59:04.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato',
  12,
  'no',
  5,
  5,
  'no',
  35,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'sunbricks',
  'earth',
  'no',
  1,
  2,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'table']",
  3,
  "['Oct', 'Nov']",
  "['reduce_meals', 'lab_ex_food']",
  -19.38400777,
```

```
 34.36429742,
 5.0,
 5.0,
 'uuid:b4d6edf2-de19-4e7a-8e12-cf8c7239c0e0'),
(160,
 'Moz',
 '24/06/2017',
 27,
 '2017-06-24T11:06:28.000Z',
 '2017-06-24T11:25:22.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 20,
 'no',
 10,
 10,
 'no',
 10,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 2,
 "['Oct', 'Nov']",
 "['reduce_meals', 'lab_ex_food']",
 -19.1114983,
 33.47606917,
 707.0,
 66.0,
 'uuid:06d21f6d-62ab-484d-80a6-a5ef3e67043a'),
(161,
 'Moz',
 '24/06/2017',
 28,
 '2017-06-24T16:30:56.000Z',
 '2017-06-24T17:47:34.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 45,
 'no',
 4,
 4,
 'no',
 40,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
```

```
 'earth',
 'no',
 2,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 2,
 "['none']",
 "['reduce_meals', 'day_night_hungry', 'seek_government']",
 -19.11219076,
 33.48338349,
 704.0,
 9.0,
 'uuid:c5f16ba4-bb4f-4878-9e10-e76803df9cd5'),
(162,
 'Moz',
 '25/06/2017',
 29,
 '2017-06-25T06:39:20.000Z',
 '2017-06-25T07:09:00.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 50,
 'no',
 6,
 6,
 'no',
 45,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'yes',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
```

```
 None,
 1,
 'no',
 'no',
 None,
 3,
 "['Aug', 'Sept', 'Oct']",
 "['limit_portion', 'reduce_meals', 'lab_ex_food']",
 -19.11203235,
 33.48332918,
 768.0,
 15.0,
 'uuid:58b0e9c3-6fff-4f22-9ff0-670911e2d97c'),
(163,
 'Moz',
 '25/06/2017',
 30,
 '2017-06-25T07:09:03.000Z',
 '2017-06-25T07:28:19.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 10,
 'no',
 8,
 8,
 'no',
 10,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 3,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle']",
 2,
 "['Sept', 'Oct', 'Nov']",
 "['reduce_meals', 'lab_ex_food']",
 -19.1121875,
 33.4834042,
 712.0,
 8.0,
 'uuid:102746ea-822f-4d36-972b-b9054522b35e'),
(164,
 'Moz',
 '25/06/2017',
 31,
 '2017-06-25T07:29:15.000Z',
 '2017-06-25T07:46:38.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 31,
 'no',
 8,
```

```
 8,
 'no',
 1,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'yes',
 1,
 1,
 'yes',
 None,
 1.0,
 'no',
 None,
 1.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['bicycle']",
 3,
 "['Aug', 'Sept', 'Oct']",
 "['rely_less_food', 'limit_portion', 'reduce_meals']",
 -19.11230668,
 33.48348781,
 698.0,
 9.0,
 'uuid:68f7f590-c5d6-460a-a3fa-0715eed8412c'),
(165,
 'Moz',
 '25/06/2017',
 32,
 '2017-06-25T07:46:55.000Z',
 '2017-06-25T08:08:01.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 46,
 'no',
 7,
 7,
 'no',
 46,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 4,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 46.0,
 'yes',
 'no',
 None,
 'no',
```

```
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['bicycle']",
 3,
 "['Jan']",
 "['rely_less_food', 'limit_portion', 'reduce_meals']",
 -19.11216152,
 33.48332066,
 719.0,
 11.0,
 'uuid:4cab3d08-26cf-4b90-8fdb-78b0f8b9f250'),
(166,
 'Moz',
 '25/06/2017',
 33,
 '2017-06-25T08:28:37.000Z',
 '2017-06-25T08:40:18.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 35,
 'no',
 3,
 3,
 'no',
 30,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 35.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 3,
 "['none']",
 "['limit_variety', 'limit_portion', 'lab_ex_food']",
 -19.11221439,
 33.48344455,
 689.0,
 7.0,
 'uuid:b9007988-8d49-444f-946f-f889a3ef4bd9'),
(167,
 'Moz',
 '25/06/2017',
 34,
```

```
 '2017-06-25T08:40:21.000Z',
 '2017-06-25T08:53:50.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 30,
 'no',
 4,
 4,
 'no',
 40,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 3,
 2,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'no',
 None,
 30.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle']",
 3,
 "['Aug', 'Sept', 'Oct']",
 "['reduce_meals', 'lab_ex_food']",
 -19.11204515,
 33.4832909,
 809.0,
 11.0,
 'uuid:f97ce99d-69d3-476a-804f-f53f23c69c8a'),
(168,
 'Moz',
 '25/06/2017',
 35,
 '2017-06-25T08:54:26.000Z',
 '2017-06-25T09:39:18.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 5,
 'no',
 5,
 5,
 'yes',
 5,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'yes',
 2,
 2,
 'yes',
```

```
 None,
 2.0,
 'no',
 None,
 5.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['bicycle', 'solar_panel', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['limit_variety', 'reduce_meals']",
 -19.11227976,
 33.48338034,
 718.0,
 10.0,
 'uuid:a5ebfca2-2c3f-4288-9fb7-d0b02096e714'),
(169,
 'Moz',
 '25/06/2017',
 36,
 '2017-06-25T13:09:17.000Z',
 '2017-06-25T13:31:12.000Z',
 'Sofala',
 'Nhamatanda',
 'Tica',
 'Nhansato',
 10,
 'no',
 1,
 1,
 'no',
 36,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'no',
 None,
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 3,
 "['Jan', 'Dec']",
 "['limit_portion', 'reduce_meals']",
```

```
 -19.11220854,
 33.48338486,
 698.0,
 7.0,
 'uuid:ec19828c-e352-4c18-b661-b11d44f21d4e'),
(170,
 'Moz',
 '25/06/2017',
 37,
 '2017-06-25T13:38:58.000Z',
 '2017-06-25T13:55:54.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 6,
 'no',
 6,
 6,
 'no',
 6,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 3,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 6.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle']",
 3,
 "['Sept', 'Oct', 'Nov']",
 "['limit_portion', 'reduce_meals', 'lab_ex_food']",
 -19.11212614,
 33.48333008,
 708.0,
 10.0,
 'uuid:b47f8116-439b-4318-9f50-a398a1d6d480'),
(171,
 'Moz',
 '25/06/2017',
 38,
 '2017-06-25T13:57:19.000Z',
 '2017-06-25T14:30:50.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 10,
 'no',
 9,
 9,
 'no',
 8,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
```

```
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['radio', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'reduce_meals', 'go_forest', 'lab_ex_food']",
 -19.11221763,
 33.48336608,
 703.0,
 43.0,
 'uuid:ff3ec25c-b3bd-47d8-95ee-b6d7a2b22a40'),
(172,
 'Moz',
 '25/06/2017',
 39,
 '2017-06-25T14:31:12.000Z',
 '2017-06-25T15:01:07.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 23,
 'no',
 9,
 9,
 'no',
 24,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 5.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
```

```
 "['goats', 'pigs']",
 None,
 2,
 'yes',
 'yes',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'go_forest']",
 -19.11222159,
 33.48342054,
 709.0,
 5.0,
 'uuid:5e162232-71f3-4a33-95d0-329c5ed88f30'),
(173,
 'Moz',
 '25/06/2017',
 40,
 '2017-06-25T15:01:19.000Z',
 '2017-06-25T15:18:31.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 25,
 'no',
 10,
 10,
 'no',
 10,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 7.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'tractor', 'solar_panel', 'mobile_phone']",
 3,
 "['Jan', 'Oct', 'Nov', 'Dec']",
 "['reduce_meals', 'lab_ex_food']",
 -19.11220082,
 33.48341163,
 709.0,
 9.0,
 'uuid:0f1bde99-7e0f-4d67-98fa-1f6a90ee03a6'),
(174,
 'Moz',
 '26/06/2017',
 41,
 '2017-06-26T05:34:45.000Z',
 '2017-06-26T05:52:38.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 30,
 'no',
```

```
 3,
 3,
 'yes',
 3,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 7.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'mobile_phone']",
 2,
 "['Jan', 'Feb']",
 "['rely_less_food', 'limit_portion', 'reduce_meals', 'lab_ex_food']",
 -19.11213629,
 33.48334281,
 718.0,
 9.0,
 'uuid:62059296-cf7a-463b-a8a0-e37e8a980b09'),
(175,
 'Moz',
 '26/06/2017',
 42,
 '2017-06-26T05:53:11.000Z',
 '2017-06-26T06:14:06.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 41,
 'no',
 7,
 7,
 'no',
 41,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'yes',
 1,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct']",
 41.0,
 'yes',
 'no',
 None,
```

```
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['pigs', 'poultry']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'solar_panel', 'mobile_phone']",
  2,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'reduce_meals', 'borrow_food', 'seek_government']",
  -19.11214616,
  33.48331151,
  711.0,
  6.0,
  'uuid:8dc14381-f7e4-4ce7-8b2b-c7159cd005ed'),
 (176,
  'Moz',
  '26/06/2017',
  43,
  '2017-06-26T06:14:11.000Z',
  '2017-06-26T06:30:39.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato',
  35,
  'no',
  4,
  4,
  'no',
  45,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['July', 'Aug', 'Sept', 'Oct']",
  30.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats', 'poultry']",
  None,
  2,
  'yes',
  'no',
  "['mobile_phone']",
  2,
  "['Jan', 'Feb']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food', 'go_forest',
'lab_ex_food', 'seek_government']",
  -19.11223376,
  33.48340679,
  699.0,
  8.0,
  'uuid:71fb6807-fe27-4920-9bae-41d883c48cf9'),
 (177,
  'Moz',
```

```
  '26/06/2017',
  44,
  '2017-06-26T06:37:19.000Z',
  '2017-06-26T06:52:48.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato',
  50,
  'no',
  6,
  6,
  'no',
  50,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['May', 'June', 'July', 'Aug', 'Sept', 'Oct']",
  32.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats', 'pigs', 'poultry']",
  None,
  3,
  'yes',
  'no',
  None,
  2,
  "['Jan', 'Feb', 'Mar', 'Apr', 'Dec']",
  "['rely_less_food', 'reduce_meals', 'borrow_food', 'go_forest',
'seek_government']",
  -19.11204336,
  33.48324378,
  735.0,
  6.0,
  'uuid:ee3561a8-2a79-4c2d-b7da-84f34e007002'),
 (178,
  'Moz',
  '26/06/2017',
  45,
  '2017-06-26T06:53:31.000Z',
  '2017-06-26T07:08:35.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato',
  50,
  'no',
  3,
  3,
  'no',
  50,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  2,
  'no',
```

```
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
 50.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'borrow_food', 'seek_government']",
 -19.11219614,
 33.4834441,
 722.0,
 7.0,
 'uuid:f376b8a8-80cd-4ab8-a3ad-641bfeaa0d50'),
(179,
 'Moz',
 '26/06/2017',
 46,
 '2017-06-26T07:08:49.000Z',
 '2017-06-26T07:22:05.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 38,
 'no',
 1,
 1,
 'no',
 38,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['May', 'June', 'July', 'Aug', 'Sept', 'Oct']",
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
```

```
 3,
 "['Jan', 'Feb', 'Dec']",
 "['limit_portion', 'borrow_food', 'go_forest', 'seek_government']",
 -19.11232025,
 33.48343542,
 678.0,
 9.0,
 'uuid:7520ec70-9270-4504-b593-5375a480bacf'),
(180,
 'Moz',
 '26/06/2017',
 47,
 '2017-06-26T07:23:18.000Z',
 '2017-06-26T07:39:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato -Castanheira',
 10,
 'no',
 10,
 10,
 'no',
 10,
 'yes',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 4,
 2,
 'yes',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 2.0,
 'yes',
 'yes',
 "['cheaper']",
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Sept', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_portion']",
 -19.1035586,
 33.4778296,
 0.0,
 2124.0,
 'uuid:dadf7328-e428-4d1b-8161-20bc9c8a5ac7'),
(181,
 'Moz',
 '26/06/2017',
 48,
 '2017-06-26T08:51:17.000Z',
 '2017-06-26T09:07:47.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato-Castanheira',
 37,
 'no',
 9,
 9,
 'no',
 37,
 'yes',
 'yes',
```

```
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'table']",
  3,
  "['Jan', 'Feb', 'Sept', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'day_night_hungry',
'lab_ex_food']",
  -19.11218834,
  33.48334860000001,
  708.0,
  11.0,
  'uuid:15978a1d-9275-4af7-af0a-95e85ab9766d'),
 (182,
  'Moz',
  '26/06/2017',
  49,
  '2017-06-26T09:08:35.000Z',
  '2017-06-26T09:22:34.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato-Castanheira',
  21,
  'yes',
  8,
  8,
  'yes',
  21,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'sunbricks',
  'earth',
  'no',
  2,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
```

```
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Sept', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'borrow_food']",
  -19.11218347,
  33.48337383,
  725.0,
  9.0,
  'uuid:dc8a263a-cddd-4292-bb81-c32991993de2'),
 (183,
  'Moz',
  '26/06/2017',
  50,
  '2017-06-26T09:22:55.000Z',
  '2017-06-26T09:38:45.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato-Castanheira',
  24,
  'yes',
  9,
  9,
  'yes',
  40,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  4.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'solar_torch', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food',
'lab_ex_food']",
  -19.11219658,
  33.48342006,
  720.0,
  7.0,
  'uuid:6b65122d-291b-4c0b-80de-3106ae1448cf'),
 (184,
  'Moz',
  '27/06/2017',
  51,
  '2017-06-27T08:13:58.000Z',
  '2017-06-27T17:02:23.000Z',
```

```
 'Sofala',
 '51',
 'Lamego',
 'Massequece',
 5,
 'yes',
 5,
 5,
 'no',
 9,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio']",
 2,
 "['Sept', 'Oct', 'Nov', 'Dec']",
 "['reduce_meals', 'lab_ex_food']",
 -19.11134046,
 33.47604298,
 717.0,
 23.0,
 'uuid:57130433-3503-4353-8016-95370469e57b'),
(185,
 'Moz',
 '27/06/2017',
 52,
 '2017-06-27T08:41:23.000Z',
 '2017-06-27T09:05:40.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 11,
 'no',
 9,
 9,
 'no',
 34,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
```

```
 'yes',
 "['Aug', 'Sept']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'solar_panel', 'mobile_phone']",
 2,
 "['Jan', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
 "['reduce_meals', 'restrict_adults']",
 -19.11153271,
 33.47620975,
 712.0,
 29.0,
 'uuid:4a9fd7e8-fb7f-4b7e-9718-f6e59059509a'),
(186,
 'Moz',
 '27/06/2017',
 53,
 '2017-06-27T09:06:35.000Z',
 '2017-06-27T09:33:44.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 10,
 'no',
 6,
 6,
 'no',
 17,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['radio', 'solar_panel']",
 3,
 "['Sept', 'Oct', 'Nov']",
 "['rely_less_food', 'limit_portion', 'lab_ex_food']",
 -19.11152561,
 33.47614903,
```

```
  722.0,
  42.0,
  'uuid:8adb7af9-3216-43bc-892f-325f4473732a'),
 (187,
  'Moz',
  '27/06/2017',
  58,
  '2017-06-27T13:02:39.000Z',
  '2017-06-27T13:23:12.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  18,
  'no',
  9,
  9,
  'no',
  18,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'sunbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  18.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  None,
  2,
  "['Jan']",
  "['rely_less_food', 'limit_portion', 'borrow_food', 'go_forest',
'seek_government']",
  -19.1115207,
  33.476051399999996,
  0.0,
  20.0,
  'uuid:846846c3-2eac-4412-81c1-52bc3c225ab0'),
 (188,
  'Moz',
  '27/06/2017',
  59,
  '2017-06-27T13:28:33.000Z',
  '2017-06-27T13:50:51.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  15,
  'no',
  11,
  11,
  'no',
  14,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
```

```
 'earth',
 'no',
 3,
 2,
 'yes',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'television', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
 2,
 "['Jan', 'Feb', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food']",
 -19.1115207,
 33.476051399999996,
 0.0,
 20.0,
 'uuid:ba051d36-fcc3-469c-b8f3-a2938272d8ba'),
(189,
 'Moz',
 '27/06/2017',
 60,
 '2017-06-27T13:54:05.000Z',
 '2017-06-27T14:07:31.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 30,
 'no',
 3,
 3,
 'yes',
 20,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
```

```
 None,
 1,
 'yes',
 'no',
 "['bicycle']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'lab_ex_food']",
 -19.1115207,
 33.476051399999996,
 0.0,
 20.0,
 'uuid:4e945d34-f420-4b0b-bd81-97b397da29a9'),
(190,
 'Moz',
 '27/06/2017',
 54,
 '2017-06-27T15:39:19.000Z',
 '2017-06-27T15:55:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 22,
 'no',
 4,
 4,
 'no',
 22,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 22.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 None,
 2,
 "['Jan', 'Dec']",
 "['rely_less_food', 'borrow_food']",
 -19.11214136,
 33.48343551,
 701.0,
 4.0,
 'uuid:6491aacb-c3b7-472f-b00f-1a905c173a2c'),
(191,
 'Moz',
 '27/06/2017',
 55,
 '2017-06-27T15:56:26.000Z',
 '2017-06-27T16:14:52.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequese',
 42,
 'yes',
 4,
```

```
 4,
 'yes',
 45,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 42.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio']",
 3,
 "['none']",
 "['na']",
 -19.11223034,
 33.48340883,
 697.0,
 16.0,
 'uuid:7e3e956f-c59f-4e7e-be35-cba5d80a76a0'),
(192,
 'Moz',
 '27/06/2017',
 56,
 '2017-06-27T16:15:37.000Z',
 '2017-06-27T16:39:42.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 57,
 'yes',
 5,
 5,
 'yes',
 16,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 3,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 10.0,
 'yes',
 'no',
 None,
 'yes',
```

```
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_torch', 'table', 'mobile_phone']",
 3,
 "['Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'lab_ex_food']",
 -19.11221906,
 33.4833719,
 730.0,
 10.0,
 'uuid:45f70b5c-89c8-4f98-a43e-66125e2f290d'),
(193,
 'Moz',
 '27/06/2017',
 57,
 '2017-06-27T16:40:01.000Z',
 '2017-06-27T16:57:22.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 17,
 'no',
 6,
 6,
 'no',
 17,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 2,
 "['Jan', 'Nov', 'Dec']",
 "['borrow_food', 'lab_ex_food']",
 -19.11223961,
 33.48326329,
 708.0,
 24.0,
 'uuid:9c3021f6-c520-45f9-afed-e4508b9a758a'),
(194,
 'Moz',
 '28/06/2017',
 61,
```

```
 '2017-06-28T04:14:11.000Z',
 '2017-06-28T04:29:48.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 6,
 'yes',
 8,
 8,
 'no',
 6,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 4,
 1,
 'yes',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['pigs', 'poultry']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_torch', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_portion']",
 -19.112245,
 33.48324839,
 711.0,
 7.0,
 'uuid:4535b659-4a1b-4ffc-bdf0-14115e93090b'),
(195,
 'Moz',
 '28/06/2017',
 62,
 '2017-06-28T04:29:54.000Z',
 '2017-06-28T04:44:02.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 10,
 'no',
 6,
 6,
 'no',
 23,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 3,
 'no',
 3,
 3,
 'yes',
```

```
 None,
 3.0,
 'yes',
 "['May', 'June', 'July', 'Aug', 'Sept', 'Oct']",
 25.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['table']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'go_forest']",
 -19.11194214,
 33.48341973,
 713.0,
 11.0,
 'uuid:e6eaf4a6-610c-4f9c-90f1-0ea90ca73513'),
(196,
 'Moz',
 '28/06/2017',
 63,
 '2017-06-28T05:55:53.000Z',
 '2017-06-28T06:18:47.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 15,
 'no',
 5,
 5,
 'no',
 15,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
 5.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['poultry']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food', 'go_forest',
```

'lab_ex_food', 'seek_government']",
  -19.06138693,
  33.39081845,
  698.0,
  10.0,
  'uuid:1541f85b-7219-4401-8efd-7e984b5c2b99'),
 (197,
  'Moz',
  '28/06/2017',
  64,
  '2017-06-28T07:02:34.000Z',
  '2017-06-28T07:16:52.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  15,
  'no',
  4,
  4,
  'no',
  15,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  4,
  4,
  'yes',
  None,
  4.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  None,
  2,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
  -19.04336707,
  33.40438933,
  715.0,
  52.0,
  'uuid:1e4b10f6-57d0-4309-aa0d-b22135ddde18'),
 (198,
  'Moz',
  '28/06/2017',
  65,
  '2017-06-28T07:19:02.000Z',
  '2017-06-28T07:31:20.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  25,
  'no',
  3,
  3,
  'no',
  50,
  'no',
  'no',
  'no',
  'no',

```
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct']",
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['goats', 'pigs', 'poultry']",
 None,
 3,
 'no',
 'no',
 "['bicycle', 'solar_panel', 'mobile_phone']",
 2,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'go_forest', 'lab_ex_food']",
 -19.04260643,
 33.40413019,
 767.0,
 55.0,
 'uuid:596816b5-a163-4d1e-9ba4-d0a27944d4c8'),
(199,
 'Moz',
 '28/06/2017',
 66,
 '2017-06-28T10:53:22.000Z',
 '2017-06-28T11:20:37.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 45,
 'yes',
 3,
 3,
 'yes',
 45,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'yes',
 1,
 1,
 'yes',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'yes',
 'never',
 'yes',
 "['support']",
 None,
 'yes',
```

```
  'no',
  "['goats', 'pigs']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.1014914,
  33.427975,
  628.0,
  4.0,
  'uuid:27d46183-f616-4933-a5dc-4d3171bc1767'),
 (200,
  'Moz',
  '28/06/2017',
  67,
  '2017-06-28T13:55:22.000Z',
  '2017-06-28T14:09:46.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  10,
  'no',
  6,
  6,
  'yes',
  30,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  3.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'borrow_food',
'lab_ex_food']",
  -19.1115207,
  33.476051399999996,
  0.0,
  20.0,
  'uuid:62829412-1df0-4858-99f9-ae88b6aedfc3'),
 (201,
  'Moz',
  '28/06/2017',
  68,
  '2017-06-28T14:11:03.000Z',
  '2017-06-28T14:25:01.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
```

```
 18,
 'yes',
 3,
 3,
 'no',
 19,
 'yes',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'yes',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 12.0,
 'yes',
 'yes',
 "['cheaper', 'obtn_more_water', 'less_work']",
 'yes',
 'no',
 'yes',
 'never',
 'yes',
 "['support']",
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.1115207,
 33.476051399999996,
 0.0,
 20.0,
 'uuid:461cf089-a400-4407-9c1b-1b93762fba39'),
(202,
 'Moz',
 '28/06/2017',
 69,
 '2017-06-28T14:29:43.000Z',
 '2017-06-28T14:45:10.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Lamego',
 19,
 'no',
 2,
 2,
 'no',
 19,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'yes',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 19.0,
 'yes',
```

```
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['radio']",
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
  -19.111524199999998,
  33.4760439,
  0.0,
  21.473000000000003,
  'uuid:85359846-2b98-413b-9671-6501980fa4f1'),
 (203,
  'Moz',
  '28/06/2017',
  6,
  '2017-06-28T14:54:17.000Z',
  '2017-06-28T15:11:10.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  3,
  'yes',
  6,
  6,
  'yes',
  18,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'sunbricks',
  'earth',
  'no',
  1,
  1,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  3.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['table']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'lab_ex_food']",
  -19.1115207,
  33.476051399999996,
  0.0,
  20.9,
  'uuid:8c1ca96b-8154-4cf4-8b77-e78efc724f2c'),
```

```
 (204,
  'Moz',
  '29/06/2017',
  71,
  '2017-06-29T03:51:01.000Z',
  '2017-06-29T04:01:27.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  13,
  'no',
  4,
  4,
  'no',
  13,
  'yes',
  'yes',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  3.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['goats', 'poultry']",
  None,
  2,
  'yes',
  'no',
  "['bicycle']",
  3,
  "['Jan', 'Feb', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'borrow_food', 'no_food', 'day_night_hungry', 'lab_ex_food']",
  -19.11173102,
  33.48353913,
  741.0,
  10.0,
  'uuid:4402581d-0b04-441d-95f2-d6978b65b610'),
 (205,
  'Moz',
  '29/06/2017',
  72,
  '2017-06-29T04:02:00.000Z',
  '2017-06-29T04:42:38.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  30,
  'no',
  7,
  7,
  'yes',
  30,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
```

```
 2,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_torch']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11197377,
 33.48346522,
 726.0,
 23.0,
 'uuid:495eb0f0-d931-4fc4-bc5c-f37e1aee3f74'),
(206,
 'Moz',
 '29/06/2017',
 73,
 '2017-06-29T13:27:31.000Z',
 '2017-06-29T13:46:02.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 20,
 'no',
 7,
 7,
 'no',
 20,
 'no',
 'yes',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 1.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
```

```
 'no',
 "['bicycle', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.11132209,
 33.47602087,
 745.0,
 49.0,
 'uuid:7018c11e-946c-4e93-8c45-5f9ba741dabf'),
(207,
 'Moz',
 '29/06/2017',
 74,
 '2017-06-29T13:46:16.000Z',
 '2017-06-29T14:11:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 13,
 'no',
 5,
 5,
 'no',
 13,
 'yes',
 'yes',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 13.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['bicycle', 'radio', 'mobile_phone']",
 3,
 "['Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11156919,
 33.4762273,
 805.0,
 30.0,
 'uuid:79d40e59-3cf8-4980-abc7-a508b6308f68'),
(208,
 'Moz',
 '29/06/2017',
 75,
 '2017-06-29T14:13:31.000Z',
 '2017-06-29T14:23:10.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 4,
 'no',
 4,
 4,
 'no',
 4,
```

```
 'no',
 'yes',
 'no',
 'no',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'electricity', 'table']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food']",
 -19.1114463,
 33.475955600000006,
 0.0,
 72.9,
 'uuid:775f4335-4372-4943-9f28-5c9c3614eea1'),
(209,
 'Moz',
 '29/06/2017',
 76,
 '2017-06-29T16:40:25.000Z',
 '2017-06-29T16:50:28.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 28,
 'no',
 4,
 4,
 'no',
 28,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 13.0,
 'no',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
```

```
 'no',
 None,
 None,
 'no',
 'no',
 "['goats', 'pigs', 'poultry']",
 None,
 3,
 'no',
 'no',
 "['bicycle', 'television', 'radio', 'electricity', 'table']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11225279,
 33.48349952,
 698.0,
 33.0,
 'uuid:fbebd4ea-8b68-4e7a-ac58-b8af66d2d386'),
(210,
 'Moz',
 '29/06/2017',
 77,
 '2017-06-29T16:51:51.000Z',
 '2017-06-29T17:06:35.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 14,
 'no',
 3,
 3,
 'no',
 14,
 'no',
 'yes',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 4.0,
 'yes',
 'yes',
 "['previous_unavailable']",
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['radio', 'solar_torch', 'table']",
 3,
 "['none']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11218703,
 33.4834991,
 725.0,
 7.0,
 'uuid:4721d9e5-00ff-47bf-8321-a457a85bdbe0'),
(211,
 'Moz',
 '29/06/2017',
 78,
 '2017-06-29T17:06:38.000Z',
 '2017-06-29T17:18:47.000Z',
 'Sofala',
```

```
  'Nhamatanda',
  'Lamego',
  'Massequece',
  12,
  'no',
  6,
  6,
  'no',
  12,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['goats', 'pigs']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'electricity', 'table',
'mobile_phone']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals']",
  -19.11217481,
  33.48347033,
  706.0,
  5.0,
  'uuid:b5200aab-a013-4c9b-9d87-90bcfe3c9d58'),
 (212,
  'Moz',
  '29/06/2017',
  79,
  '2017-06-29T17:19:21.000Z',
  '2017-06-29T17:34:24.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  15,
  'yes',
  5,
  5,
  'no',
  20,
  'yes',
  'yes',
  'no',
  'no',
  'mabatisloping',
  'cement',
  'cement',
  'yes',
  1,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
```

```
  'no',
  None,
  6.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'yes',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'television', 'radio', 'sterio', 'solar_panel', 'electricity',
'table', 'sofa_set', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar']",
  "['rely_less_food', 'limit_variety']",
  -19.11211864,
  33.48337579,
  716.0,
  14.0,
  'uuid:997e97ae-ca9a-431d-8222-79da30c93b4b'),
 (213,
  'Moz',
  '29/06/2017',
  80,
  '2017-06-29T17:34:36.000Z',
  '2017-06-29T17:45:39.000Z',
  'Sofala',
  'Nhamantanda',
  'Lamego',
  'Massequece',
  5,
  'no',
  2,
  2,
  'no',
  22,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct']",
  2.0,
  'yes',
  'no',
  None,
  'yes',
  'no',
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food',
'lab_ex_food']",
```

```
  -19.11216772,
  33.48345188,
  692.0,
  10.0,
  'uuid:a67bc75f-c70e-490e-9bae-5454e7bad76f'),
 (214,
  'Moz',
  '30/06/2017',
  81,
  '2017-06-30T05:32:47.000Z',
  '2017-06-30T05:50:23.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nhansato',
  20,
  'no',
  9,
  9,
  'no',
  25,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct']",
  12.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['poultry']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'radio', 'solar_torch', 'mobile_phone']",
  2,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'restrict_adults', 'borrow_food']",
  -19.04362725,
  33.40450004,
  674.0,
  9.0,
  'uuid:65064927-d9ad-4a31-b95c-8e38d0c5d599'),
 (215,
  'Moz',
  '30/06/2017',
  83,
  '2017-06-30T13:10:52.000Z',
  '2017-06-30T13:23:29.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  19,
  'no',
  3,
  3,
  'no',
  12,
  'yes',
  'yes',
  'no',
  'no',
```

```
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  None,
  3,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'day_night_hungry',
'go_forest']",
  -19.11159001,
  33.47559586,
  735.0,
  139.0,
  'uuid:bc5bd0b1-83a6-4d91-b04c-5e48d19bffc8'),
 (216,
  'Moz',
  '30/06/2017',
  84,
  '2017-06-30T13:25:24.000Z',
  '2017-06-30T13:36:57.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  6,
  'no',
  6,
  6,
  'no',
  15,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct']",
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
```

```
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['solar_torch']",
  2,
  "['Jan', 'Feb', 'Mar', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'restrict_adults', 'borrow_food', 'lab_ex_food', 'seek_government']",
  -19.1115207,
  33.476051399999996,
  0.0,
  20.0,
  'uuid:887b0ef4-ccfc-40d2-b807-71d3218627ce'),
 (217,
  'Moz',
  '30/06/2017',
  82,
  '2017-06-30T12:53:02.000Z',
  '2017-06-30T13:04:36.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  30,
  'no',
  6,
  6,
  'no',
  10,
  'yes',
  'yes',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'radio', 'solar_torch', 'mobile_phone']",
  2,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'restrict_adults', 'lab_ex_food']",
  -19.11167546,
  33.47622988,
  734.0,
  26.0,
  'uuid:d79aa0e1-d38b-4aee-a8bb-712c7cf79877'),
 (218,
  'Moz',
  '30/06/2017',
  85,
  '2017-06-30T15:22:08.000Z',
  '2017-06-30T15:42:35.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
```

```
  'Massequece',
  10,
  'no',
  6,
  6,
  'no',
  7,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct']",
  7.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'television', 'solar_panel', 'sofa_set', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals', 'go_forest',
'lab_ex_food']",
  -19.11155625,
  33.45908041,
  706.0,
  14.0,
  'uuid:47518791-d202-4ea3-8285-748a6432d6eb'),
 (219,
  'Moz',
  '30/06/2017',
  86,
  '2017-06-30T15:43:34.000Z',
  '2017-06-30T16:13:25.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  20,
  'no',
  6,
  6,
  'no',
  25,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['May', 'June', 'July', 'Aug', 'Sept', 'Oct']",
```

```
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['poultry']",
  None,
  1,
  'no',
  'no',
  "['radio', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'restrict_adults', 'borrow_food', 'go_forest', 'lab_ex_food']",
  -19.1113203,
  33.45912408,
  704.0,
  7.0,
  'uuid:ebe34e84-e09c-4b5d-aa80-ec490289e8cf'),
 (220,
  'Moz',
  '30/06/2017',
  87,
  '2017-06-30T16:14:41.000Z',
  '2017-06-30T16:31:19.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  25,
  'no',
  5,
  5,
  'no',
  35,
  'yes',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept']",
  15.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats', 'pigs', 'poultry']",
  None,
  3,
  'yes',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety']",
  -19.11123552,
  33.45907741,
  705.0,
```

```
 11.0,
 'uuid:686b184f-c2fc-4964-ac62-a08465846eac'),
(221,
 'Moz',
 '30/06/2017',
 88,
 '2017-06-30T16:32:30.000Z',
 '2017-06-30T16:44:09.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 20,
 'no',
 4,
 4,
 'no',
 30,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 15.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['poultry']",
 None,
 1,
 'no',
 'yes',
 "['bicycle', 'radio', 'solar_panel', 'mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11132791,
 33.45916222,
 702.0,
 10.0,
 'uuid:93295999-d643-42d0-b96d-0028f98b4952'),
(222,
 'Moz',
 '30/06/2017',
 89,
 '2017-06-30T16:44:18.000Z',
 '2017-06-30T17:07:47.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 20,
 'yes',
 10,
 10,
 'no',
 20,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
```

```
  2,
  3,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  20.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'yes',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['poultry']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'television', 'radio', 'sterio', 'solar_panel', 'solar_torch',
'electricity', 'table', 'sofa_set', 'mobile_phone']",
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'borrow_food']",
  -19.11131064,
  33.45913834,
  716.0,
  22.0,
  'uuid:7eae29c9-e402-4e77-b960-2b5bc09bb5ea'),
 (223,
  'Moz',
  '30/06/2017',
  90,
  '2017-06-30T17:10:24.000Z',
  '2017-06-30T17:22:47.000Z',
  'Sofala',
  'Nhamatanda',
  'Lomego',
  'Massequece',
  40,
  'no',
  6,
  6,
  'yes',
  40,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
```

```
 1,
 'yes',
 'no',
 "['solar_torch']",
 2,
 "['Jan', 'Feb', 'Sept', 'Oct', 'Nov', 'Dec']",
 "['go_forest']",
 -19.11131805,
 33.45906793,
 704.0,
 8.0,
 'uuid:88f0276d-bda9-4ce7-9394-eceba1bbcac5'),
(224,
 'Moz',
 '01/07/2017',
 92,
 '2017-07-01T17:19:27.000Z',
 '2017-07-01T17:37:44.000Z',
 'Sofala',
 'Nhamatanda',
 '91',
 'Massequece',
 45,
 'yes',
 6,
 6,
 'yes',
 70,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 2,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'yes',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['motorcyle', 'radio', 'solar_panel', 'solar_torch', 'table']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'borrow_food']",
 -19.11227186,
 33.4834346,
 690.0,
 10.0,
 'uuid:09452e0d-22d5-4f45-8698-9249c614bab9'),
(225,
 'Moz',
 '01/07/2017',
 92,
 '2017-07-01T17:38:11.000Z',
 '2017-07-01T17:50:39.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 5,
 'no',
 4,
 4,
```

```
 'no',
 5,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_torch']",
 2,
 "['Jan', 'Feb', 'Mar', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'go_forest', 'lab_ex_food']",
 -19.11225355,
 33.483395,
 697.0,
 5.0,
 'uuid:c8fd2fdb-df73-461a-9c2c-38f288c65e1b'),
(226,
 'Moz',
 '01/07/2017',
 93,
 '2017-07-01T17:51:10.000Z',
 '2017-07-01T18:05:14.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 16,
 'no',
 10,
 10,
 'no',
 15,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'sterio', 'solar_torch']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.11222287,
 33.48341093,
 689.0,
 13.0,
 'uuid:3dfe8fd8-659b-437d-b235-b25084f99002'),
(227,
 'Moz',
 '01/07/2017',
 94,
 '2017-07-01T18:07:22.000Z',
 '2017-07-01T18:20:55.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 5,
 'no',
 4,
 4,
 'no',
 5,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 5.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['pigs']",
 None,
 1,
 'yes',
 'no',
 "['computer', 'solar_torch']",
 2,
 "['Jan', 'Feb', 'Mar', 'Sept', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'borrow_food', 'lab_ex_food']",
 -19.11219416,
 33.48343164,
 684.0,
 5.0,
 'uuid:d2452e59-1a21-45da-a184-325828070df1'),
(228,
 'Moz',
 '02/07/2017',
 95,
 '2017-07-02T07:23:51.000Z',
```

```
 '2017-07-02T07:39:41.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 20,
 'no',
 5,
 5,
 'no',
 20,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.11223405,
 33.4833483,
 723.0,
 11.0,
 'uuid:a10d89ad-8ade-4ab5-9b15-6ed66709074a'),
(229,
 'Moz',
 '02/07/2017',
 96,
 '2017-07-02T07:44:52.000Z',
 '2017-07-02T07:56:08.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhamatanda',
 4,
 'no',
 4,
 4,
 'no',
 4,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
```

```
 2.0,
 'no',
 None,
 4.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['radio']",
 2,
 "['Jan', 'Feb', 'Mar', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
 -19.112238100000003,
 33.48342821,
 713.0,
 10.0,
 'uuid:8d7261cc-7c9b-4578-9106-88b807f35362'),
(230,
 'Moz',
 '02/07/2017',
 97,
 '2017-07-02T07:56:56.000Z',
 '2017-07-02T08:10:50.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 17,
 'no',
 6,
 6,
 'no',
 17,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 17.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['pigs']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
 -19.1121819,
```

```
 33.48340442,
 714.0,
 6.0,
 'uuid:0791ec0f-4afd-409c-8972-c8e403001bcf'),
(231,
 'Moz',
 '02/07/2017',
 98,
 '2017-07-02T08:11:09.000Z',
 '2017-07-02T08:23:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 13,
 'no',
 3,
 3,
 'no',
 10,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 4,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['lorry', 'solar_panel', 'mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['limit_portion', 'lab_ex_food']",
 -19.11219328,
 33.48336924,
 709.0,
 13.0,
 'uuid:6d23af1e-09b6-4bfa-8b7b-4367113fee51'),
(232,
 'Moz',
 '02/07/2017',
 99,
 '2017-07-02T08:24:09.000Z',
 '2017-07-02T08:39:12.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 25,
 'no',
 4,
 4,
 'no',
 40,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
```

```
 'cement',
 'no',
 3,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 25.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'yes',
 "['farming']",
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'table', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11218694,
 33.483386100000004,
 716.0,
 11.0,
 'uuid:18c5d990-c1a1-42e2-a4da-ecf940cefdc3'),
(233,
 'Moz',
 '02/07/2017',
 100,
 '2017-07-02T08:39:15.000Z',
 '2017-07-02T08:50:55.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 15,
 'no',
 3,
 3,
 'no',
 37,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['none']",
```

```
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.11225776,
  33.48338638,
  701.0,
  6.0,
  'uuid:bb10b080-d919-46b8-838b-a3bed2609788'),
 (234,
  'Moz',
  '04/07/2017',
  101,
  '2017-07-04T05:20:53.000Z',
  '2017-07-04T05:37:16.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  5,
  'no',
  8,
  8,
  'no',
  40,
  'yes',
  'no',
  'yes',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  6,
  'no',
  4,
  4,
  'yes',
  None,
  4.0,
  'yes',
  "['Aug', 'Sept']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'television', 'electricity', 'table', 'sofa_set', 'mobile_phone',
'fridge']",
  3,
  "['none']",
  "['na']",
  -14.71751016,
  34.35387906,
  1281.0,
  9.0,
  'uuid:33247b5c-964b-4422-b385-71d1e635b4c9'),
 (235,
  'Moz',
  '04/07/2017',
  102,
  '2017-07-04T05:39:31.000Z',
  '2017-07-04T05:48:27.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  1,
  'yes',
```

```
 4,
 4,
 'no',
 30,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['radio', 'table']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food']",
 -14.71758448,
 34.35371339,
 1273.0,
 17.0,
 'uuid:0cd22fe9-76d3-467c-b202-0f75d17707c1'),
(236,
 'Moz',
 '05/07/2017',
 103,
 '2017-07-05T05:47:50.000Z',
 '2017-07-05T06:07:58.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 9,
 'yes',
 4,
 4,
 'no',
 27,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 2,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
```

```
 'yes',
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'electricity', 'mobile_phone', 'fridge']",
 3,
 "['none']",
 "['na']",
 -14.71758322,
 34.35406703,
 1270.0,
 16.0,
 'uuid:36b8d9fc-29b8-41c9-bbac-1d6598be651b'),
(237,
 'Moz',
 '05/07/2017',
 104,
 '2017-07-05T06:34:39.000Z',
 '2017-07-07T06:45:05.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 3,
 'no',
 8,
 8,
 'no',
 3,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['goats', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['bicycle', 'radio']",
 2,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food']",
 -14.721824800000002,
 34.3585566,
 0.0,
 2185.0,
 'uuid:a7547669-7cf8-4982-ab61-b742f71337ac'),
(238,
 'Moz',
 '05/07/2017',
```

```
 105,
 '2017-07-05T11:29:24.000Z',
 '2017-07-05T11:42:08.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 50,
 'no',
 4,
 4,
 'no',
 50,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Sept', 'Oct']",
 15.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['bicycle', 'solar_torch', 'table']",
 2,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food']",
 -14.7214103,
 34.3582717,
 0.0,
 2200.0,
 'uuid:c800d6e8-aaad-4e54-bce3-78b12ee456a9'),
(239,
 'Moz',
 '06/07/2017',
 106,
 '2017-07-06T06:30:20.000Z',
 '2017-07-06T07:00:02.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 20,
 'yes',
 7,
 7,
 'no',
 30,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 2,
 3,
 'no',
 1,
 1,
```

```
 'yes',
 None,
 1.0,
 'yes',
 "['Aug', 'Sept']",
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Dec']",
 "['limit_variety', 'reduce_meals']",
 -14.72334659,
 34.34416543,
 1272.0,
 29.0,
 'uuid:919b4115-86e6-4b12-8a45-0b7d85b52de0'),
(240,
 'Moz',
 '06/07/2017',
 107,
 '2017-07-06T07:03:16.000Z',
 '2017-07-06T07:59:31.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 30,
 'yes',
 6,
 6,
 'no',
 30,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 2,
 3,
 'yes',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['goats', 'poultry']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'electricity', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Oct', 'Nov', 'Dec']",
```

```
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -14.72315773,
 34.34406807,
 1290.0,
 42.0,
 'uuid:b5b9ab62-24c9-4132-aafe-758ffdcc869b'),
(241,
 'Moz',
 '07/07/2017',
 108,
 '2017-07-07T05:21:40.000Z',
 '2017-07-07T05:39:43.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 6,
 'no',
 6,
 6,
 'no',
 6,
 'yes',
 'yes',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 2,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 6.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['motorcyle', 'bicycle', 'television', 'radio', 'table', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -14.71749213,
 34.35370834,
 1302.0,
 15.0,
 'uuid:c2d63f59-2626-4b8b-ac6a-16fe4af1f8ea'),
(242,
 'Moz',
 '07/07/2017',
 109,
 '2017-07-07T05:41:50.000Z',
 '2017-07-07T05:53:47.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 20,
 'no',
 5,
 5,
 'no',
 20,
 'no',
 'no',
 'no',
 'no',
```

```
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  13.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['poultry']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'lab_ex_food']",
  -14.71761138,
  34.35387854,
  1240.0,
  11.0,
  'uuid:ea053298-8556-4577-8218-78d4d4767e34'),
 (243,
  'Moz',
  '07/07/2017',
  110,
  '2017-07-07T06:27:14.000Z',
  '2017-07-07T06:44:51.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  35,
  'no',
  3,
  3,
  'no',
  35,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  20.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
```

```
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'solar_panel', 'table', 'mobile_phone']",
 2,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -14.72343526,
 34.34431177,
 1260.0,
 24.0,
 'uuid:b0e0740f-22a6-42e3-918e-e5a7bff24a52'),
(244,
 'Moz',
 '09/07/2017',
 120,
 '2017-07-09T17:08:35.000Z',
 '2017-07-09T17:29:19.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 12,
 'no',
 10,
 10,
 'no',
 33,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -18.95176687,
 33.26177387,
 647.0,
 7.0,
 'uuid:897f340d-1ba0-47fe-9dfc-0a76a39af7ed'),
(245,
 'Moz',
 '10/07/2017',
 121,
 '2017-07-10T05:14:53.000Z',
 '2017-07-10T05:33:26.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
```

```
 12,
 'no',
 6,
 6,
 'no',
 5,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'more_once',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar', 'Apr']",
 "['rely_less_food', 'borrow_food']",
 -18.95173891,
 33.26171114,
 664.0,
 5.0,
 'uuid:45a11e51-009c-41da-8bf2-a8b66d6a66a2'),
(246,
 'Moz',
 '10/07/2017',
 122,
 '2017-07-10T05:33:29.000Z',
 '2017-07-10T05:53:07.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 5,
 'no',
 5,
 5,
 'no',
 3,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 5.0,
 'yes',
```

```
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'reduce_meals']",
 -18.95175155,
 33.261753399999996,
 667.0,
 5.0,
 'uuid:c16129e9-c69c-4c02-8bba-52c7e42bed3b'),
(247,
 'Moz',
 '10/07/2017',
 123,
 '2017-07-10T05:53:13.000Z',
 '2017-07-10T06:16:53.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 25,
 'no',
 7,
 7,
 'no',
 15,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 3,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food']",
 -18.95171584,
 33.26170882,
 655.0,
 5.0,
 'uuid:837e0768-3c25-4aeb-a35b-85a41404f2f0'),
(248,
```

```
  'Moz',
  '10/07/2017',
  124,
  '2017-07-10T06:27:59.000Z',
  '2017-07-10T06:46:10.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  15,
  'no',
  11,
  11,
  'no',
  9,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  4,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'television', 'radio', 'solar_panel', 'solar_torch', 'electricity',
'table', 'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Aug', 'Sept']",
  "['rely_less_food']",
  -18.95175349,
  33.26169067,
  656.0,
  11.0,
  'uuid:0ce65b2d-6061-4b55-a674-47148b54f696'),
 (249,
  'Moz',
  '10/07/2017',
  125,
  '2017-07-10T06:47:12.000Z',
  '2017-07-10T07:19:29.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  27,
  'no',
  6,
  6,
  'no',
  27,
  'yes',
  'no',
  'yes',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  3,
```

```
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  27.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'television', 'radio', 'solar_panel', 'solar_torch',
'electricity', 'table', 'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['reduce_meals', 'borrow_food']",
  -18.95177568,
  33.26174639,
  649.0,
  15.0,
  'uuid:c9d22767-dba3-4d7f-9f37-4a07da81e80a'),
 (250,
  'Moz',
  '10/07/2017',
  126,
  '2017-07-10T07:39:31.000Z',
  '2017-07-10T07:53:41.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  5,
  'no',
  7,
  7,
  'no',
  5,
  'yes',
  'no',
  'yes',
  'no',
  'mabatisloping',
  'cement',
  'cement',
  'yes',
  1,
  3,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'no',
```

```
  'no',
  "['bicycle', 'television', 'radio', 'solar_torch', 'electricity', 'table',
'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'reduce_meals']",
  -18.95177745,
  33.26172438,
  651.0,
  56.0,
  'uuid:afc26fe7-a10e-4263-88ee-0edc411823f2'),
 (251,
  'Moz',
  '10/07/2017',
  127,
  '2017-07-10T07:54:26.000Z',
  '2017-07-10T08:18:10.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  16,
  'no',
  9,
  9,
  'no',
  3,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  1,
  1,
  'yes',
  None,
  1.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov']",
  16.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats']",
  None,
  1,
  'yes',
  'yes',
  "['bicycle', 'table', 'mobile_phone']",
  3,
  "['Jan']",
  "['rely_less_food', 'lab_ex_food']",
  -18.95188039,
  33.26170161,
  619.0,
  8.0,
  'uuid:5d3b42f7-8f16-4d11-bef6-083443da5eb9'),
 (252,
  'Moz',
  '10/07/2017',
  129,
  '2017-07-10T11:41:35.000Z',
  '2017-07-10T11:59:07.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  20,
  'no',
  8,
  8,
  'no',
```

```
 20,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 1,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -18.95175855,
 33.26184303,
 623.0,
 11.0,
 'uuid:71d03142-40af-4227-9659-453e0cbaccda'),
(253,
 'Moz',
 '10/07/2017',
 128,
 '2017-07-10T11:59:11.000Z',
 '2017-07-10T12:22:32.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 10,
 'no',
 15,
 15,
 'no',
 20,
 'no',
 'yes',
 'no',
 'yes',
 'mabatipitched',
 'cement',
 'tiles',
 'yes',
 2,
 3,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept']",
 1.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
```

```
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['car', 'lorry', 'motorcyle', 'television', 'electricity', 'table', 'sofa_set',
'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety']",
  -18.95180249,
  33.26176401,
  646.0,
  20.0,
  'uuid:49996ba1-b219-44be-8854-ec1fa2a1cd26'),
 (254,
  'Moz',
  '10/07/2017',
  130,
  '2017-07-10T12:28:31.000Z',
  '2017-07-10T12:52:44.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  7,
  'no',
  3,
  3,
  'yes',
  7,
  'yes',
  'no',
  'yes',
  'no',
  'mabatisloping',
  'cement',
  'cement',
  'yes',
  1,
  1,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats', 'poultry']",
  None,
  2,
  'no',
  'no',
  "['car', 'television', 'radio', 'sterio', 'electricity', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar']",
  "['rely_less_food', 'limit_variety', 'go_forest']",
  -18.95178118,
  33.26170729,
  634.0,
  17.0,
  'uuid:8c9af539-52ad-4ea0-be86-e9d215d97690'),
 (255,
  'Moz',
  '07/07/2017',
  111,
  '2017-07-07T07:05:51.000Z',
```

```
  '2017-07-07T07:20:46.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  40,
  'no',
  2,
  2,
  'yes',
  40,
  'no',
  'yes',
  'no',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept', 'Oct']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['poultry']",
  None,
  1,
  'no',
  'no',
  "['solar_torch']",
  3,
  "['Jan', 'Feb', 'Mar', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'borrow_food']",
  -14.72354118,
  34.34408304,
  1254.0,
  20.0,
  'uuid:b4616110-9691-484f-8b20-a4bbc02ff742'),
 (256,
  'Moz',
  '07/07/2017',
  112,
  '2017-07-07T16:57:16.000Z',
  '2017-07-07T17:17:43.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Massequece',
  12,
  'no',
  6,
  6,
  'no',
  15,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
```

```
 None,
 2.0,
 'no',
 None,
 8.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['television', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'go_forest', 'lab_ex_food']",
 -14.7256636,
 34.3610469,
 0.0,
 2099.999,
 'uuid:e039c1ca-2102-40d6-ab16-f42d11aa3429'),
(257,
 'Moz',
 '07/07/2017',
 114,
 '2017-07-07T17:32:39.000Z',
 '2017-07-07T17:48:46.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 6,
 'no',
 5,
 5,
 'no',
 6,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'yes',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 6.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_torch']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'borrow_food']",
```

```
 -14.71746301,
 34.35380822,
 1249.0,
 18.0,
 'uuid:1bed1b67-f55f-4393-afc4-6b791e0154b9'),
(258,
 'Moz',
 '07/07/2017',
 115,
 '2017-07-07T17:50:15.000Z',
 '2017-07-07T18:06:30.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 40,
 'no',
 4,
 4,
 'no',
 40,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 35.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['radio', 'solar_panel', 'electricity', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['reduce_meals', 'go_forest']",
 -14.71745811,
 34.35378321,
 1276.0,
 10.0,
 'uuid:a279f8ba-3519-4751-b363-72da36617f3b'),
(259,
 'Moz',
 '09/07/2017',
 116,
 '2017-07-09T14:52:36.000Z',
 '2017-07-09T15:06:41.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 20,
 'no',
 6,
 6,
 'no',
 19,
 'no',
 'no',
 'no',
 'no',
 'grass',
```

```
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 13.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['solar_panel']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['reduce_meals', 'borrow_food']",
 -18.95161101,
 33.26158652,
 651.0,
 12.0,
 'uuid:83b2469d-b9e2-47ea-a8a1-9cb638f4b0a8'),
(260,
 'Moz',
 '09/07/2017',
 117,
 '2017-07-09T15:12:30.000Z',
 '2017-07-09T15:31:38.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 30,
 'no',
 10,
 10,
 'yes',
 30,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 25.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
```

```
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_torch']",
 2,
 "['Jan', 'Dec']",
 "['reduce_meals', 'borrow_food', 'go_forest', 'lab_ex_food']",
 -18.95407352,
 33.26594692,
 653.0,
 4.0,
 'uuid:d0db831a-9910-47ee-b4be-3801e667d2b0'),
(261,
 'Moz',
 '09/07/2017',
 118,
 '2017-07-09T15:31:48.000Z',
 '2017-07-09T15:48:30.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 15,
 'no',
 8,
 8,
 'no',
 60,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 3,
 2,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 2,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'borrow_food']",
 -18.95406826,
 33.26594114,
 651.0,
 5.0,
 'uuid:f94c27c3-4a9c-4fd8-83d7-3112e666dc8e'),
(262,
 'Moz',
 '09/07/2017',
 119,
 '2017-07-09T16:21:06.000Z',
 '2017-07-09T17:07:51.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nhansato',
 16,
 'no',
```

```
 3,
 3,
 'no',
 30,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'no',
 None,
 6.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['poultry']",
 None,
 1,
 'no',
 'no',
 "['radio', 'solar_panel']",
 3,
 "['none']",
 "['na']",
 -18.95182097,
 33.2618461,
 658.0,
 15.0,
 'uuid:a6b32653-0202-4453-a1e5-89bde5a271fd'),
(263,
 'Moz',
 '13/07/2017',
 131,
 '2017-07-13T16:12:12.000Z',
 '2017-07-13T16:31:00.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 23,
 'no',
 6,
 6,
 'no',
 20,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 2,
 1,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct']",
 2.0,
 'yes',
 'no',
 None,
```

```
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'borrow_food', 'go_forest']",
  -19.11220694,
  33.48337946,
  703.0,
  19.0,
  'uuid:10e16c6c-db8e-4034-b324-ac8cea072b5c'),
 (264,
  'Moz',
  '13/07/2017',
  132,
  '2017-07-13T16:31:13.000Z',
  '2017-07-13T16:44:15.000Z',
  'Sofala',
  'Madangua',
  'Lamego',
  'Madangua',
  11,
  'no',
  8,
  8,
  'no',
  11,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  3,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug']",
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals', 'go_forest',
'lab_ex_food']",
  -19.11228821,
  33.48340581,
  685.0,
  32.0,
  'uuid:4b6078b4-7e3b-45bd-9ad6-2553e7c8a829'),
 (265,
  'Moz',
```

```
  '13/07/2017',
  133,
  '2017-07-13T16:44:46.000Z',
  '2017-07-14T04:57:46.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  5,
  'no',
  3,
  3,
  'no',
  26,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'sunbricks',
  'cement',
  'no',
  1,
  4,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['July', 'Aug', 'Sept', 'Oct']",
  5.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['television', 'electricity', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'restrict_adults',
'lab_ex_food']",
  -19.11217152,
  33.48342965,
  710.0,
  4.0,
  'uuid:ddfadef3-45c2-4679-b1cd-7b949cb4c9f2'),
 (266,
  'Moz',
  '13/07/2017',
  134,
  '2017-07-13T16:57:18.000Z',
  '2017-07-14T04:58:05.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  2,
  'no',
  9,
  9,
  'no',
  2,
  'yes',
  'no',
  'yes',
  'no',
  'grass',
  'sunbricks',
  'earth',
  'no',
  2,
  1,
  'no',
```

```
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['July', 'Aug', 'Sept']",
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['goats']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'television', 'radio', 'sterio', 'electricity', 'table',
'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals']",
  -19.11220491,
  33.48340178,
  706.0,
  21.0,
  'uuid:23544c71-a6e8-406b-8687-f44afa9a8bc9'),
 (267,
  'Moz',
  '13/07/2017',
  135,
  '2017-07-13T17:23:51.000Z',
  '2017-07-14T04:57:21.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  17,
  'no',
  2,
  2,
  'no',
  17,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct']",
  17.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['poultry']",
  None,
  1,
  'no',
  'no',
```

```
 None,
 3,
 "['Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'seek_government']",
 -19.11219406,
 33.48344529,
 673.0,
 30.0,
 'uuid:ee598272-5610-4172-81ef-60e65a6e4165'),
(268,
 'Moz',
 '14/07/2017',
 136,
 '2017-07-14T04:40:36.000Z',
 '2017-07-14T04:53:52.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 3,
 'no',
 4,
 4,
 'no',
 5,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct']",
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_torch', 'table']",
 3,
 "['Jan', 'Feb', 'Mar']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.11210546,
 33.48331551,
 713.0,
 14.0,
 'uuid:64daed29-5de4-4b95-a338-a83e2514f8fa'),
(269,
 'Moz',
 '14/07/2017',
 137,
 '2017-07-14T04:58:19.000Z',
 '2017-07-14T05:12:58.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 13,
 'no',
 6,
 6,
 'no',
 13,
 'no',
```

```
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct']",
 13.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'solar_torch', 'table']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
 -19.11238719,
 33.48328691,
 689.0,
 12.0,
 'uuid:8c0cb1b7-c5ee-4f29-bb6a-fbea72d5bc79'),
(270,
 'Moz',
 '14/07/2017',
 138,
 '2017-07-14T14:50:39.000Z',
 '2017-07-14T15:01:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 10,
 'no',
 5,
 5,
 'no',
 10,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct']",
 7.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
```

```
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'table']",
  3,
  "['Jan', 'Feb', 'Mar']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'borrow_food', 'lab_ex_food']",
  -19.11637936,
  33.47159471,
  703.0,
  5.0,
  'uuid:f2c47d3a-946b-4102-8f98-daa6e3b1ff2d'),
 (271,
  'Moz',
  '14/07/2017',
  139,
  '2017-07-14T15:02:07.000Z',
  '2017-07-14T15:11:27.000Z',
  'Sofala',
  'Nhamatandda',
  'Lamego',
  'Madangua',
  10,
  'no',
  5,
  5,
  'no',
  10,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'sunbricks',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct']",
  10.0,
  'yes',
  'yes',
  "['cheaper']",
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle']",
  3,
  "['Jan', 'Feb', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'lab_ex_food']",
  -19.11642689,
  33.47149112,
  694.0,
  4.0,
  'uuid:1d25b110-699d-4651-8740-33d60489881c'),
 (272,
  'Moz',
  '14/07/2017',
  140,
  '2017-07-14T15:11:53.000Z',
  '2017-07-14T15:28:01.000Z',
```

```
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  8,
  'no',
  9,
  9,
  'no',
  25,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  2,
  3,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov']",
  8.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'table']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'borrow_food', 'no_food', 'lab_ex_food']",
  -19.11637826,
  33.4715523,
  692.0,
  5.0,
  'uuid:bfd1149b-54d8-4d57-ac8a-bf9d0c1ab265'),
 (273,
  'Moz',
  '16/07/2017',
  141,
  '2017-07-16T08:04:48.000Z',
  '2017-07-16T08:20:31.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  5,
  'no',
  4,
  4,
  'no',
  45,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'sunbricks',
  'earth',
  'no',
  1,
  2,
  'no',
  3,
  3,
  'yes',
  None,
```

```
  3.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
  1.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'yes',
  "['farming']",
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  None,
  3,
  "['Jan', 'Feb', 'Mar']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'restrict_adults', 'borrow_food', 'no_food', 'lab_ex_food']",
  -19.11206303,
  33.48352159,
  733.0,
  5.0,
  'uuid:01d672d7-2854-4d2f-a392-e46cdef18e71'),
 (274,
  'Moz',
  '16/07/2017',
  142,
  '2017-07-16T08:20:37.000Z',
  '2017-07-16T08:39:49.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  25,
  'no',
  10,
  10,
  'no',
  15,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  3,
  'yes',
  2,
  2,
  'yes',
  None,
  2.0,
  'yes',
  "['June', 'July', 'Aug', 'Sept', 'Oct']",
  15.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['poultry']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'television', 'radio', 'table', 'fridge']",
  3,
  "['none']",
  "['na']",
```

```
 -19.11210615,
 33.48352978,
 702.0,
 4.0,
 'uuid:eb343026-25b8-423e-848d-ff9c5b3b3f4e'),
(275,
 'Moz',
 '16/07/2017',
 143,
 '2017-07-16T08:43:19.000Z',
 '2017-07-16T08:59:44.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 10,
 'no',
 5,
 5,
 'no',
 40,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'no',
 'no',
 "['radio', 'table', 'mobile_phone']",
 3,
 "['Jan']",
 "['rely_less_food', 'reduce_meals', 'restrict_adults', 'borrow_food']",
 -19.11199287,
 33.48352391,
 736.0,
 4.0,
 'uuid:1d1106c3-dff0-4bc7-9aca-33c8d326bd47'),
(276,
 'Moz',
 '16/07/2017',
 144,
 '2017-07-16T09:00:30.000Z',
 '2017-07-16T09:18:41.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 20,
 'no',
 5,
 5,
 'no',
 51,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
```

```
 'sunbricks',
 'cement',
 'no',
 2,
 2,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['June', 'July', 'Aug', 'Sept', 'Oct', 'Nov']",
 19.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'solar_panel', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11208087,
 33.48350786,
 686.0,
 4.0,
 'uuid:69dd39e1-8ecc-4b6b-9fb9-3c616081aa5b'),
(277,
 'Moz',
 '16/07/2017',
 145,
 '2017-07-16T09:18:45.000Z',
 '2017-07-16T09:36:16.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 15,
 'no',
 10,
 10,
 'no',
 40,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'sunbricks',
 'cement',
 'no',
 2,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 6.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
```

```
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'television', 'radio', 'table', 'sofa_set']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'reduce_meals', 'borrow_food']",
 -19.11207978,
 33.48348548,
 701.0,
 4.0,
 'uuid:bed12ca0-b067-4d00-ac0a-30dc2bd1c017'),
(278,
 'Moz',
 '16/07/2017',
 148,
 '2017-07-16T09:41:24.000Z',
 '2017-07-16T09:54:46.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 5,
 'no',
 3,
 3,
 'no',
 22,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 5.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle']",
 3,
 "['Jan', 'Feb', 'Mar']",
 "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
 -19.11208063,
 33.48351109,
 704.0,
 4.0,
 'uuid:aa4da4ea-e72a-427a-8798-60c03ccdbef7'),
(279,
 'Moz',
 '16/07/2017',
 149,
 '2017-07-16T14:13:44.000Z',
 '2017-07-16T14:43:13.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 50,
 'yes',
```

```
 8,
 8,
 'no',
 49,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 40.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle']",
 2,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.11227751,
 33.48340728,
 699.0,
 5.0,
 'uuid:4a671885-24bf-416d-87ee-fd760a9ab134'),
(280,
 'Moz',
 '16/07/2017',
 147,
 '2017-07-16T15:01:10.000Z',
 '2017-07-16T15:16:24.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 1,
 'no',
 7,
 7,
 'no',
 60,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 2,
 3,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 2.0,
 'yes',
 'no',
 None,
```

```
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11223296,
 33.48343545,
 696.0,
 10.0,
 'uuid:1e1795cb-f470-4597-bb80-bda16d399fb6'),
(281,
 'Moz',
 '16/07/2017',
 146,
 '2017-07-16T15:16:45.000Z',
 '2017-07-16T15:33:24.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 2,
 'no',
 8,
 8,
 'no',
 31,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'mobile_phone']",
 3,
 "['Sept', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
 -19.11224829,
 33.48345129,
 705.0,
 12.0,
 'uuid:1b29fefa-d169-47f6-8c22-dc2191e5702b'),
(282,
 'Moz',
 '16/07/2017',
```

```
 150,
 '2017-07-16T15:39:56.000Z',
 '2017-07-16T16:03:11.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 17,
 'yes',
 5,
 5,
 'no',
 20,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 3,
 1,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
 17.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'table', 'mobile_phone']",
 2,
 "['Jan', 'Feb', 'Sept', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food']",
 -19.1122128,
 33.48341528,
 714.0,
 10.0,
 'uuid:3c22eb00-e327-4ffd-b9b1-368b4962f4bb'),
(283,
 'Moz',
 '17/07/2017',
 151,
 '2017-07-17T06:04:11.000Z',
 '2017-07-17T06:24:29.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 36,
 'yes',
 5,
 5,
 'yes',
 20,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 3,
 1,
 'no',
 4,
 4,
```

```
  'yes',
  None,
  4.0,
  'yes',
  "['July', 'Aug', 'Sept', 'Oct', 'Nov']",
  30.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'reduce_meals', 'go_forest', 'lab_ex_food']",
  -19.11200034,
  33.48319807,
  743.0,
  44.0,
  'uuid:37f32d47-ae52-4a25-9a34-cc005309d2cb'),
 (284,
  'Moz',
  '17/07/2017',
  152,
  '2017-07-17T14:40:59.000Z',
  '2017-07-17T15:08:49.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  6,
  'yes',
  8,
  8,
  'no',
  37,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  2,
  3,
  'no',
  1,
  1,
  'yes',
  None,
  1.0,
  'yes',
  "['Sept', 'Oct']",
  3.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'television', 'radio', 'electricity', 'table',
'sofa_set', 'mobile_phone', 'fridge']",
  3,
```

```
  "['Oct']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'lab_ex_food']",
  -19.11223544,
  33.48346775,
  715.0,
  10.0,
  'uuid:182f85e1-d63d-45d4-b196-dfd384200a79'),
 (285,
  'Moz',
  '17/07/2017',
  153,
  '2017-07-17T15:09:01.000Z',
  '2017-07-17T15:26:21.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  23,
  'yes',
  3,
  3,
  'yes',
  23,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  13.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['goats', 'poultry']",
  None,
  2,
  'no',
  'no',
  "['radio', 'table']",
  3,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
  -19.11218811,
  33.48341404,
  709.0,
  12.0,
  'uuid:0abd7c8e-7d97-4718-a2f0-1bb9cbade583'),
 (286,
  'Moz',
  '17/07/2017',
  154,
  '2017-07-17T15:26:56.000Z',
  '2017-07-17T15:44:54.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  23,
  'yes',
  6,
  6,
  'no',
  23,
  'no',
  'no',
```

```
 'no',
 'no',
 'mabatisloping',
 'sunbricks',
 'earth',
 'yes',
 3,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 23.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['sheep', 'poultry']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'solar_panel', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
 -19.11220926,
 33.48346351,
 710.0,
 13.0,
 'uuid:b257e20e-3aec-469d-bfe0-73e89938ce0b'),
(287,
 'Moz',
 '17/07/2017',
 155,
 '2017-07-17T15:45:17.000Z',
 '2017-07-17T15:57:42.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 2,
 'yes',
 5,
 5,
 'no',
 28,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov']",
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
```

```
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  None,
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'borrow_food', 'lab_ex_food']",
  -19.11216523,
  33.48345336,
  708.0,
  11.0,
  'uuid:c81c353f-3c8c-4766-bc08-e86d84dadfd9'),
 (288,
  'Moz',
  '18/07/2017',
  156,
  '2017-07-18T07:42:06.000Z',
  '2017-07-18T07:57:52.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  6,
  'yes',
  5,
  5,
  'no',
  30,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  2,
  2,
  'no',
  4,
  4,
  'yes',
  None,
  4.0,
  'no',
  None,
  6.0,
  'no',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'table']",
  3,
  "['Jan', 'Feb', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'lab_ex_food']",
  -19.04349814,
  33.40419468,
  686.0,
  11.0,
  'uuid:1d66cfa7-7cb7-4cd7-bf1c-4fa06ddd177e'),
 (289,
  'Moz',
  '18/07/2017',
  157,
  '2017-07-18T07:58:46.000Z',
  '2017-07-18T08:07:16.000Z',
  'Sofala',
  'Nhamatanda',
```

```
 'Lamego',
 'Madangua',
 21,
 'yes',
 1,
 1,
 'no',
 21,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 21.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_torch', 'table']",
 3,
 "['Jan', 'Feb', 'Mar', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals']",
 -19.04342373,
 33.40441698,
 641.0,
 33.0,
 'uuid:773c37d3-9469-40d4-b02a-68e00d44f1d2'),
(290,
 'Moz',
 '18/07/2017',
 158,
 '2017-07-18T11:49:12.000Z',
 '2017-07-18T12:08:56.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 2,
 'yes',
 7,
 7,
 'yes',
 1,
 'yes',
 'yes',
 'no',
 'yes',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'no',
 None,
```

```
  2.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals']",
  -19.11202245,
  33.48328227,
  709.0,
  40.0,
  'uuid:095f0ebd-4c28-4b89-bcc1-4c9837777d20'),
 (291,
  'Moz',
  '18/07/2017',
  159,
  '2017-07-18T12:08:59.000Z',
  '2017-07-18T12:28:22.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  25,
  'yes',
  8,
  8,
  'yes',
  45,
  'yes',
  'no',
  'no',
  'yes',
  'mabatisloping',
  'muddaub',
  'earth',
  'no',
  2,
  3,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  25.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'mobile_phone']",
  2,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'borrow_food']",
  -19.11215367,
  33.4833973,
  707.0,
```

```
 13.0,
 'uuid:8d426c93-f27f-4c97-bff3-836af81ef182'),
(292,
 'Moz',
 '18/07/2017',
 160,
 '2017-07-18T12:29:03.000Z',
 '2017-07-18T12:41:56.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 1,
 'no',
 4,
 4,
 'no',
 7,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['radio', 'fridge']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'reduce_meals']",
 -19.11223556,
 33.48339619,
 699.0,
 19.0,
 'uuid:36629d68-8d69-45eb-a2b6-bd9f7043b0c8'),
(293,
 'Moz',
 '19/07/2017',
 161,
 '2017-07-19T09:32:27.000Z',
 '2017-07-19T09:47:12.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 6,
 'yes',
 9,
 9,
 'no',
 49,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
```

```
 3,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 6.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['goats']",
 None,
 1,
 'yes',
 'no',
 "['radio', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.11224265,
 33.48344019,
 689.0,
 9.0,
 'uuid:beeea76e-ac18-4779-a64f-8a497069b9ca'),
(294,
 'Moz',
 '19/07/2017',
 162,
 '2017-07-19T09:47:15.000Z',
 '2017-07-19T10:01:21.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 16,
 'no',
 8,
 8,
 'no',
 13,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'yes',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 13.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
```

```
  'no',
  'yes',
  "['mobile_phone']",
  3,
  "['Jan', 'Feb', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'lab_ex_food']",
  -19.11218757,
  33.48340916,
  758.0,
  8.0,
  'uuid:02488447-c91a-4de1-bffe-c5aefed5502c'),
 (295,
  'Moz',
  '19/07/2017',
  163,
  '2017-07-19T10:01:34.000Z',
  '2017-07-19T10:19:54.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  12,
  'no',
  9,
  9,
  'no',
  12,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'muddaub',
  'earth',
  'no',
  1,
  2,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  12.0,
  'yes',
  'yes',
  "['obtn_more_water', 'less_work']",
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'radio', 'table']",
  3,
  "['Jan', 'Sept', 'Oct', 'Nov', 'Dec']",
  "['rely_less_food', 'reduce_meals']",
  -19.11213241,
  33.48338908,
  700.0,
  9.0,
  'uuid:46d85123-b73a-4549-a108-3dcd1390c0ad'),
 (296,
  'Moz',
  '19/07/2017',
  164,
  '2017-07-19T10:21:19.000Z',
  '2017-07-19T10:35:32.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  19,
  'no',
  4,
  4,
```

```
 'no',
 19,
 'yes',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 19.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['solar_torch', 'table', 'mobile_phone']",
 3,
 "['none']",
 "['na']",
 -19.11224278,
 33.48341985,
 702.0,
 5.0,
 'uuid:fa4ff645-8c7e-4fa9-b7fe-e3b198b3107e'),
(297,
 'Moz',
 '19/07/2017',
 165,
 '2017-07-19T10:35:48.000Z',
 '2017-07-19T10:49:44.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 2,
 'no',
 6,
 6,
 'no',
 4,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 3,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['limit_variety', 'reduce_meals']",
 -19.11224829,
 33.4834212,
 702.0,
 9.0,
 'uuid:b261846a-6231-4a7b-9ccd-aa863b33a8e7'),
(298,
 'Moz',
 '19/07/2017',
 166,
 '2017-07-19T10:56:38.000Z',
 '2017-07-19T11:10:18.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 1,
 'no',
 6,
 6,
 'no',
 2,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'solar_torch', 'electricity', 'mobile_phone']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11215659,
 33.48337539,
 705.0,
 10.0,
 'uuid:c6be9fea-59c2-471f-a80b-f2c65c2786be'),
(299,
 'Moz',
 '19/07/2017',
 167,
 '2017-07-19T11:10:31.000Z',
```

```
 '2017-07-19T11:21:27.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 4,
 'no',
 5,
 5,
 'no',
 4,
 'no',
 'yes',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 4.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'solar_torch', 'table']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11213466,
 33.48341985,
 697.0,
 11.0,
 'uuid:8de7d8da-1e63-4a90-8895-3c5033e41ba0'),
(300,
 'Moz',
 '19/07/2017',
 168,
 '2017-07-19T11:21:30.000Z',
 '2017-07-19T11:32:48.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 1,
 'no',
 4,
 4,
 'no',
 1,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'yes',
 None,
```

```
  1.0,
  'no',
  None,
  1.0,
  'yes',
  'yes',
  "['previous_unavailable']",
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  None,
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals', 'no_food',
'lab_ex_food']",
  -19.11210894,
  33.48327438,
  699.0,
  13.0,
  'uuid:a9e71668-712e-43f4-9007-7c7359a7ebfb'),
 (301,
  'Moz',
  '19/07/2017',
  169,
  '2017-07-19T11:32:51.000Z',
  '2017-07-19T11:44:07.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  11,
  'no',
  7,
  7,
  'yes',
  11,
  'no',
  'yes',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  2,
  2,
  'no',
  1,
  1,
  'no',
  1.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'radio', 'solar_torch', 'electricity', 'table',
'mobile_phone']",
  3,
  "['Jan', 'Dec']",
```

```
 "['rely_less_food', 'limit_variety', 'reduce_meals']",
 -19.11220623,
 33.48351684,
 708.0,
 11.0,
 'uuid:1b2cef54-227f-41c4-8254-1dbed17e9f78'),
(302,
 'Moz',
 '19/07/2017',
 170,
 '2017-07-19T11:44:10.000Z',
 '2017-07-19T11:55:32.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 3,
 'no',
 5,
 5,
 'no',
 3,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 1,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.11222346,
 33.48340619,
 701.0,
 10.0,
 'uuid:34daa5f8-ab78-4b35-9f1d-47245acdc6ca'),
(303,
 'Moz',
 '19/07/2017',
 171,
 '2017-07-19T15:47:30.000Z',
 '2017-07-19T16:02:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 10,
 'no',
 7,
 7,
 'no',
 10,
 'no',
 'no',
 'no',
 'no',
```

```
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  3,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  10.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'television', 'radio', 'solar_panel', 'solar_torch', 'electricity',
'table', 'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_portion', 'borrow_food']",
  -19.11222518,
  33.48350478,
  704.0,
  5.0,
  'uuid:cb483f76-7902-4106-b552-ad5fc2e2ec77'),
 (304,
  'Moz',
  '19/07/2017',
  172,
  '2017-07-19T16:03:50.000Z',
  '2017-07-19T16:29:04.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  30,
  'no',
  5,
  5,
  'yes',
  20,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  2,
  'no',
  4,
  4,
  'yes',
  None,
  4.0,
  'no',
  None,
  20.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
```

```
 'no',
 'yes',
 "['goats', 'pigs', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'solar_torch']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'lab_ex_food']",
 -19.11219731,
 33.48345445,
 684.0,
 15.0,
 'uuid:cbd0b040-2b8a-4bc7-8661-6c50d301a168'),
(305,
 'Moz',
 '19/07/2017',
 173,
 '2017-07-19T16:29:16.000Z',
 '2017-07-19T16:41:43.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 23,
 'no',
 6,
 6,
 'no',
 23,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 2,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 20.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_panel']",
 3,
 "['Jan', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
 -19.11221544,
 33.483467499999996,
 697.0,
 10.0,
 'uuid:ccbfaec0-dd21-40d0-8c5b-dba74c593efd'),
(306,
 'Moz',
 '19/07/2017',
 174,
 '2017-07-19T16:43:33.000Z',
 '2017-07-19T16:54:47.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
```

```
 15,
 'no',
 4,
 4,
 'yes',
 13,
 'yes',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'mobile_phone']",
 3,
 "['Jan', 'Oct', 'Nov', 'Dec']",
 "['limit_variety', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.11218126,
 33.48349666,
 702.0,
 7.0,
 'uuid:402ead38-98d8-4ec4-8aac-9be71d797947'),
(307,
 'Moz',
 '19/07/2017',
 175,
 '2017-07-19T16:54:50.000Z',
 '2017-07-19T17:11:48.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 19,
 'no',
 9,
 9,
 'no',
 17,
 'yes',
 'no',
 'yes',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
 3,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'no',
 None,
 17.0,
 'yes',
```

```
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['motorcyle', 'radio', 'solar_torch']",
  3,
  "['Jan', 'Feb', 'Mar']",
  "['rely_less_food', 'reduce_meals', 'lab_ex_food']",
  -19.11218621,
  33.48344957,
  702.0,
  9.0,
  'uuid:63b9fe42-e366-476b-a6b5-6cfc0032381f'),
 (308,
  'Moz',
  '20/07/2017',
  176,
  '2017-07-20T04:41:56.000Z',
  '2017-07-20T04:58:07.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  20,
  'no',
  10,
  10,
  'no',
  30,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  1,
  3,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  20.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'solar_panel', 'table']",
  2,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'reduce_meals', 'restrict_adults', 'borrow_food',
'lab_ex_food']",
  -19.11201461,
  33.48341571,
  747.0,
  11.0,
  'uuid:dba91fea-bc88-464e-9e6e-763fafd82a29'),
```

```
(309,
 'Moz',
 '20/07/2017',
 177,
 '2017-07-20T04:58:14.000Z',
 '2017-07-20T05:09:57.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 3,
 'no',
 5,
 5,
 'no',
 17,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 3.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['goats', 'pigs']",
 None,
 2,
 'yes',
 'no',
 "['television', 'radio', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.11215612,
 33.48341474,
 716.0,
 10.0,
 'uuid:170b87e6-46c3-491a-a31c-9c8be40be862'),
(310,
 'Moz',
 '20/07/2017',
 178,
 '2017-07-20T05:11:55.000Z',
 '2017-07-20T05:28:02.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 45,
 'no',
 10,
 10,
 'no',
 40,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
 1,
```

```
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 40.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Nov', 'Dec']",
 "['rely_less_food', 'borrow_food', 'go_forest']",
 -19.11217484,
 33.48339463,
 717.0,
 5.0,
 'uuid:a5247389-a739-41e2-82da-4cd6a9903757'),
(311,
 'Moz',
 '24/07/2017',
 179,
 '2017-07-24T04:51:43.000Z',
 '2017-07-24T16:58:19.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 12,
 'no',
 6,
 6,
 'no',
 15,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'sunbricks',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['goats', 'pigs', 'poultry']",
 None,
 3,
 'yes',
 'no',
```

```
  "['motorcyle', 'radio', 'solar_torch', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'borrow_food', 'go_forest', 'lab_ex_food']",
  -19.33407781,
  34.31642189,
  -3.0,
  3.0,
  'uuid:b53acc44-ab9a-474e-943d-d81c32e0304f'),
 (312,
  'Moz',
  '24/07/2017',
  180,
  '2017-07-24T05:14:10.000Z',
  '2017-07-24T16:58:42.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  10,
  'no',
  5,
  5,
  'no',
  5,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  3,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  3.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'no',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['radio', 'sterio', 'electricity', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion']",
  -19.33409091,
  34.31629507,
  36.0,
  5.0,
  'uuid:e1a0e35e-b93f-447f-b64c-36434883a4e2'),
 (313,
  'Moz',
  '24/07/2017',
  181,
  '2017-07-24T09:45:53.000Z',
  '2017-07-24T10:06:54.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  7,
  'no',
  8,
  8,
  'no',
  7,
```

```
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 2,
 2,
 'no',
 3,
 3,
 'no',
 3.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['oxen']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['rely_less_food', 'limit_variety']",
 -19.4064424,
 34.43747345,
 -19.0,
 5.0,
 'uuid:786eff8c-0020-4603-8dce-d72fc6739003'),
(314,
 'Moz',
 '24/07/2017',
 182,
 '2017-07-24T10:13:41.000Z',
 '2017-07-24T16:59:00.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 5,
 'no',
 5,
 5,
 'no',
 5,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 3,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'sterio', 'cow_plough', 'solar_panel', 'sofa_set']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'borrow_food', 'day_night_hungry', 'lab_ex_food']",
  -19.40654201,
  34.43746759,
  -17.0,
  10.0,
  'uuid:fb7248ba-5ace-46d1-9e2d-e6dd5e23e468'),
 (315,
  'Moz',
  '24/07/2017',
  183,
  '2017-07-24T10:26:04.000Z',
  '2017-07-24T10:39:59.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  3,
  'no',
  5,
  5,
  'yes',
  3,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  3,
  'yes',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  3.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'television', 'radio', 'sterio', 'table', 'sofa_set', 'mobile_phone',
'fridge']",
  3,
  "['none']",
  "['na']",
  -19.40647934,
  34.43747881,
  -22.0,
  13.0,
  'uuid:0f19c5dc-7222-4e1b-b000-a326d1f95dad'),
 (316,
  'Moz',
  '24/07/2017',
  184,
  '2017-07-24T12:54:14.000Z',
```

```
  '2017-07-24T13:09:10.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Lamego - Madangua',
  10,
  'no',
  8,
  8,
  'no',
  10,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  1,
  2,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  1.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['poultry']",
  None,
  1,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'television', 'radio', 'sterio', 'solar_panel',
'electricity', 'table', 'mobile_phone']",
  3,
  "['Jan', 'Feb']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'go_forest']",
  -19.40667636,
  34.43746277,
  0.0,
  5.0,
  'uuid:707bdb52-ad0f-43d5-8fd8-ad6cdce122b9'),
 (317,
  'Moz',
  '24/07/2017',
  185,
  '2017-07-24T13:09:36.000Z',
  '2017-07-24T13:21:14.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  35,
  'yes',
  4,
  4,
  'yes',
  4,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  3,
  'no',
  4,
  4,
  'yes',
```

```
 None,
 4.0,
 'yes',
 "['Dec']",
 35.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Dec']",
 "['go_forest', 'lab_ex_food']",
 -19.406550699999997,
 34.43743141,
 14.0,
 12.0,
 'uuid:cb1086cf-50e9-47e8-8ae9-eab03238b3b4'),
(318,
 'Moz',
 '24/07/2017',
 186,
 '2017-07-24T15:05:20.000Z',
 '2017-07-24T15:32:11.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 7,
 'no',
 6,
 6,
 'no',
 10,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'yes',
 2,
 1,
 'yes',
 5,
 5,
 'yes',
 None,
 5.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 7.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['goats', 'pigs', 'poultry']",
 None,
 3,
 'no',
 'no',
 "['bicycle', 'table', 'mobile_phone']",
 2,
 "['Jan', 'Feb']",
 "['rely_less_food', 'limit_variety', 'go_forest']",
```

```
 -19.27284544,
 34.20390833,
 76.0,
 5.0,
 'uuid:877b4334-b938-4b5a-acdf-7d856d9f48d1'),
(319,
 'Moz',
 '24/07/2017',
 187,
 '2017-07-24T16:33:57.000Z',
 '2017-07-24T16:45:20.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 5,
 'no',
 6,
 6,
 'no',
 5,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 3,
 'yes',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'yes',
 'no',
 "['goats']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'table']",
 2,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_variety', 'borrow_food', 'go_forest']",
 -19.27286151,
 34.20393386,
 54.0,
 5.0,
 'uuid:a73ff65b-9b63-4936-b0dc-425ee1460d54'),
(320,
 'Moz',
 '24/07/2017',
 188,
 '2017-07-24T16:45:39.000Z',
 '2017-07-24T16:57:09.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 1,
 'no',
 3,
 3,
 'no',
 5,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
```

```
 'burntbricks',
 'cement',
 'yes',
 1,
 2,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'radio', 'electricity', 'mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food']",
 -19.27285846,
 34.20390732,
 64.0,
 3.0,
 'uuid:f25842db-6aa6-46ba-a105-ceda9f423957'),
(321,
 'Moz',
 '07/07/2017',
 113,
 '2017-07-07T17:18:13.000Z',
 '2017-07-24T17:00:52.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Massequece',
 55,
 'no',
 1,
 1,
 'yes',
 60,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'yes',
 1,
 1,
 'no',
 1,
 1,
 'yes',
 None,
 1.0,
 'no',
 None,
 2.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
```

```
 "['none']",
 None,
 1,
 'no',
 'no',
 None,
 2,
 "['Jan', 'Dec']",
 "['limit_portion', 'reduce_meals', 'borrow_food', 'seek_government']",
 -14.7174565,
 34.35393313,
 1261.0,
 72.0,
 'uuid:dbcf8573-60c8-46f0-818d-048965b81683'),
(322,
 'Moz',
 '25/07/2017',
 190,
 '2017-07-25T10:53:07.000Z',
 '2017-07-25T11:10:34.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 1,
 'no',
 6,
 6,
 'no',
 27,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'yes',
 2,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 1.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['television', 'electricity', 'table', 'mobile_phone']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['reduce_meals', 'borrow_food']",
 -19.41217768,
 34.44026935,
 22.0,
 11.0,
 'uuid:3c85d8d3-22f9-4208-a40a-ec7a35081d85'),
(323,
 'Moz',
 '01/12/2016',
 1,
 '2017-08-08T11:54:54.000Z',
 '2017-08-08T13:59:46.000Z',
 'Manica',
 'Vanduzi',
 'Vanduzi',
 'Belas',
 17,
 'yes',
```

```
 2,
 2,
 'yes',
 17,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
 1,
 'no',
 5,
 5,
 'yes',
 None,
 5.0,
 'yes',
 "['Sept', 'Oct', 'Nov', 'Dec']",
 3.0,
 'yes',
 'yes',
 "['avail_money']",
 'yes',
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['oxen', 'cows', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'radio', 'cow_plough', 'solar_torch', 'mobile_phone']",
 2,
 "['Jan']",
 "['seek_government']",
 -19.11734948,
 33.47660169,
 699.0,
 5.0,
 'uuid:fa93a7dd-5b9a-4663-9302-76d20096ba52'),
(324,
 'Moz',
 '01/12/2016',
 2,
 '2017-08-09T11:07:27.000Z',
 '2017-08-09T11:53:48.000Z',
 'Manica',
 'Vanduzi',
 'Vanduzi',
 'Belas',
 9,
 'yes',
 6,
 6,
 'no',
 9,
 'yes',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 3,
 4,
 'no',
 4,
 4,
 'yes',
 None,
 4.0,
 'yes',
 "['Aug', 'Sept', 'Oct']",
 3.0,
 'yes',
 'yes',
 "['train_outside_org']",
```

```
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['cows', 'goats', 'pigs', 'poultry']",
 None,
 4,
 'yes',
 'no',
 "['bicycle', 'radio', 'sterio', 'cow_plough', 'solar_panel', 'solar_torch']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'go_forest', 'seek_government']",
 -19.11739121,
 33.47659249,
 716.0,
 5.0,
 'uuid:126ed6ff-d601-47c0-8b5b-8aa39d7b94f5'),
(325,
 'Moz',
 '01/12/2016',
 3,
 '2017-08-09T11:55:38.000Z',
 '2017-08-09T12:43:36.000Z',
 'Manica',
 'Vanduzi',
 'Vanduzi',
 'Belas',
 24,
 'yes',
 5,
 5,
 'yes',
 21,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'yes',
 "['Aug', 'Sept', 'Oct', 'Nov', 'Dec']",
 3.0,
 'yes',
 'yes',
 "['train_outside_org']",
 'yes',
 'no',
 'no',
 'frequently',
 'no',
 None,
 None,
 'no',
 'no',
 "['goats', 'pigs', 'poultry']",
 None,
 3,
 'yes',
 'no',
 "['bicycle', 'television', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Sept', 'Oct', 'Nov', 'Dec']",
 "['reduce_meals', 'borrow_food', 'seek_government']",
 -19.11738955,
 33.47659911,
 707.0,
 5.0,
 'uuid:3c853174-090b-45be-ac26-254b083ed156'),
(326,
 'Moz',
 '01/12/2016',
```

```
  4,
  '2017-08-09T12:43:54.000Z',
  '2017-08-09T14:03:41.000Z',
  'Manica',
  'Vanduzi',
  'Vanduzi',
  'Belas',
  13,
  'yes',
  12,
  12,
  'no',
  15,
  'yes',
  'yes',
  'yes',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  4,
  2,
  'no',
  6,
  6,
  'yes',
  None,
  6.0,
  'yes',
  "['Sept', 'Oct', 'Nov']",
  1.0,
  'yes',
  'yes',
  "['train_outside_org']",
  'no',
  None,
  'no',
  'frequently',
  'no',
  None,
  None,
  'no',
  'no',
  "['cows', 'goats', 'poultry']",
  None,
  3,
  'yes',
  'no',
  "['bicycle', 'radio', 'sterio', 'cow_plough', 'solar_torch', 'table',
'mobile_phone']",
  3,
  "['Oct', 'Nov', 'Dec']",
  "['limit_variety']",
  -19.11737137,
  33.47655563,
  702.0,
  5.0,
  'uuid:dcbb8590-15ee-493a-9896-2062a901fcfe'),
 (327,
  'Moz',
  '01/12/2016',
  5,
  '2017-08-10T09:21:21.000Z',
  '2017-08-10T11:35:48.000Z',
  'Manica',
  'Vanduzi',
  'Vanduzi',
  'Belas',
  7,
  'yes',
  10,
  10,
  'no',
  28,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'sunbricks',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
```

```
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct']",
 7.0,
 'yes',
 'yes',
 "['obtn_more_water', 'less_work']",
 'yes',
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['cows']",
 None,
 1,
 'no',
 'no',
 "['bicycle', 'table', 'mobile_phone']",
 2,
 "['Aug', 'Sept', 'Oct']",
 "['limit_variety']",
 -19.11732142,
 33.4765294,
 714.0,
 11.0,
 'uuid:62decb21-69a2-4a25-9610-7460aab3c9ac'),
(328,
 'Moz',
 '01/12/2016',
 6,
 '2017-08-10T11:38:30.000Z',
 '2017-08-10T12:44:31.000Z',
 'Manica',
 'Vanduzi',
 'Vanduzi',
 'Belas',
 2,
 'yes',
 3,
 3,
 'no',
 15,
 'yes',
 'yes',
 'no',
 'yes',
 'mabatisloping',
 'burntbricks',
 'earth',
 'no',
 1,
 2,
 'yes',
 1,
 1,
 'yes',
 None,
 1.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 2.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'sterio', 'solar_torch', 'table', 'mobile_phone']",
 2,
```

```
 "['Sept', 'Oct', 'Nov', 'Dec']",
 "['limit_variety']",
 -19.117406600000002,
 33.4765576,
 0.0,
 21.6,
 'uuid:4b03f943-3ce7-4379-992a-9f65f1b7b05a'),
(329,
 'Moz',
 '01/12/2016',
 7,
 '2017-08-10T12:52:37.000Z',
 '2017-08-10T13:38:49.000Z',
 'Manica',
 'Vanduzi',
 'Vanduzi',
 'Belas',
 7,
 'yes',
 5,
 5,
 'no',
 32,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'sunbricks',
 'earth',
 'no',
 2,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct', 'Nov']",
 7.0,
 'yes',
 'no',
 None,
 'yes',
 'no',
 'yes',
 'never',
 'no',
 None,
 None,
 'no',
 'yes',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['sterio', 'solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Aug', 'Sept', 'Oct', 'Nov']",
 "['lab_ex_food']",
 -19.117406600000002,
 33.4765576,
 0.0,
 20.1,
 'uuid:afd25082-1857-442c-9296-b911b27751d5'),
(330,
 'Moz',
 '01/12/2016',
 8,
 '2017-08-11T09:03:05.000Z',
 '2017-08-11T11:16:35.000Z',
 'Manica',
 'Vanduzi',
 'Vanduzi',
 'Belas',
 5,
 'yes',
 5,
 5,
 'no',
 25,
 'yes',
 'no',
 'yes',
```

```
  'no',
  'mabatisloping',
  'burntbricks',
  'earth',
  'no',
  2,
  2,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Aug', 'Sept', 'Oct', 'Nov']",
  2.0,
  'yes',
  'yes',
  "['obtn_more_water', 'less_work']",
  'yes',
  'no',
  'yes',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['computer', 'radio', 'sterio', 'cow_plough', 'solar_panel', 'table',
'mobile_phone']",
  3,
  "['Oct', 'Nov', 'Dec']",
  "['limit_portion', 'reduce_meals']",
  -19.11739911,
  33.47657075,
  713.0,
  13.0,
  'uuid:a56f105a-6f25-48a1-9a24-a32211dcfe31'),
 (331,
  'Moz',
  '01/12/2016',
  9,
  '2017-08-11T11:28:04.000Z',
  '2017-08-11T12:53:59.000Z',
  'Manica',
  'Va',
  'Vanduzi',
  'Belas',
  16,
  'yes',
  8,
  8,
  'no',
  39,
  'yes',
  'yes',
  'no',
  'no',
  'mabatipitched',
  'sunbricks',
  'cement',
  'yes',
  2,
  3,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  1.0,
  'yes',
  'yes',
  "['less_work']",
  'yes',
  'no',
  'no',
  'never',
  'yes',
  "['farming']",
```

```
  None,
  'yes',
  'no',
  "['oxen', 'goats', 'poultry']",
  None,
  3,
  'yes',
  'no',
  "['television', 'radio', 'cow_plough', 'solar_panel', 'solar_torch', 'table',
'mobile_phone']",
  3,
  "['none']",
  "['na']",
  -19.11739116,
  33.47658977,
  718.0,
  10.0,
  'uuid:009fc260-87cd-4acb-b695-24bff2c44cb6'),
 (332,
  'Moz',
  '21/08/2017',
  191,
  '2017-08-15T05:18:46.000Z',
  '2017-08-21T06:10:54.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  15,
  'yes',
  4,
  4,
  'no',
  50,
  'no',
  'yes',
  'no',
  'yes',
  'grass',
  'sunbricks',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  15.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['cows']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'mobile_phone']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'reduce_meals', 'borrow_food']",
  -19.1115104,
  33.4759125,
  713.4000244,
  21.813000000000002,
  'uuid:61f243ba-786e-4acb-9fb7-777e339622c7'),
 (333,
  'Moz',
  '21/08/2017',
  192,
  '2017-08-21T06:19:05.000Z',
  '2017-08-21T06:50:50.000Z',
  'Sofala',
  'Nhamatanda',
```

```
 'Lamego',
 'Madangua',
 17,
 'yes',
 12,
 12,
 'no',
 18,
 'no',
 'no',
 'no',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 5,
 3,
 'no',
 3,
 3,
 'yes',
 None,
 3.0,
 'no',
 None,
 17.0,
 'yes',
 'no',
 None,
 'yes',
 'yes',
 'yes',
 'never',
 'yes',
 "['farming', 'wages', 'business']",
 None,
 'yes',
 'yes',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'electricity', 'mobile_phone']",
 3,
 "['Jan', 'Oct', 'Nov', 'Dec']",
 "['rely_less_food', 'limit_variety', 'lab_ex_food']",
 -19.11147603,
 33.47614013,
 699.0,
 25.0,
 'uuid:be47bcee-3b40-492a-9219-6b19d7e53be4'),
(334,
 'Moz',
 '21/08/2017',
 193,
 '2017-08-21T10:31:06.000Z',
 '2017-08-21T10:50:24.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 60,
 'yes',
 6,
 6,
 'no',
 51,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'cement',
 'cement',
 'yes',
 2,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
```

```
  6.0,
  'yes',
  'no',
  None,
  'yes',
  'yes',
  'yes',
  'never',
  'no',
  None,
  None,
  'yes',
  'no',
  "['none']",
  None,
  1,
  'no',
  'no',
  "['bicycle', 'television', 'radio', 'electricity', 'table', 'sofa_set',
'mobile_phone', 'fridge']",
  3,
  "['Jan', 'Feb', 'Nov', 'Dec']",
  "['rely_less_food', 'reduce_meals']",
  -19.40793648,
  33.29214318,
  607.0,
  17.0,
  'uuid:76d38063-777a-4cec-8fad-caba34499145'),
 (335,
  'Moz',
  '21/08/2017',
  194,
  '2017-08-21T10:50:30.000Z',
  '2017-08-21T11:13:21.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  8,
  'yes',
  8,
  8,
  'no',
  13,
  'no',
  'no',
  'no',
  'no',
  'mabatisloping',
  'burntbricks',
  'cement',
  'yes',
  1,
  4,
  'no',
  2,
  2,
  'yes',
  None,
  2.0,
  'no',
  None,
  8.0,
  'yes',
  'yes',
  "['avail_money']",
  'yes',
  'yes',
  'yes',
  'frequently',
  'no',
  None,
  None,
  'yes',
  'no',
  "['oxen', 'cows']",
  None,
  2,
  'yes',
  'no',
  "['motorcyle', 'bicycle', 'television', 'electricity', 'table', 'mobile_phone',
'fridge']",
  3,
  "['Jan', 'Feb', 'Mar']",
  "['na']",
  -19.40803975,
  33.29180602,
```

```
 598.0,
 13.0,
 'uuid:f6a8cfeb-419b-4852-ad03-e2bd75a1e52e'),
(336,
 'Moz',
 '21/08/2017',
 195,
 '2017-08-21T11:14:03.000Z',
 '2017-08-21T11:28:32.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 15,
 'yes',
 3,
 3,
 'no',
 40,
 'yes',
 'no',
 'yes',
 'no',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 15.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'yes',
 'yes',
 "['bicycle', 'radio', 'sofa_set']",
 3,
 "['Jan', 'Nov', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'lab_ex_food']",
 -19.40794986,
 33.29208865,
 588.0,
 7.0,
 'uuid:02d6de35-3dc6-4729-b82d-ad59c94ed789'),
(337,
 'Moz',
 '21/08/2017',
 196,
 '2017-08-21T13:07:57.000Z',
 '2017-08-21T13:22:51.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 5,
 'yes',
 7,
 7,
 'yes',
 40,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'sunbricks',
 'earth',
```

```
 'no',
 2,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'yes',
 "['Sept', 'Oct']",
 5.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['goats', 'poultry']",
 None,
 2,
 'yes',
 'no',
 "['bicycle', 'radio', 'mobile_phone']",
 3,
 "['Jan', 'Feb']",
 "['rely_less_food', 'limit_portion', 'reduce_meals']",
 -19.11140353,
 33.47609595,
 730.0,
 28.0,
 'uuid:30777507-6d70-4837-9bd9-a2b4587f3f39'),
(338,
 'Moz',
 '21/08/2017',
 197,
 '2017-08-21T13:44:42.000Z',
 '2017-08-21T14:01:45.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 30,
 'yes',
 9,
 9,
 'yes',
 30,
 'no',
 'yes',
 'no',
 'yes',
 'grass',
 'sunbricks',
 'earth',
 'no',
 3,
 2,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 30.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['goats', 'poultry']",
 None,
```

```
  2,
  'yes',
  'no',
  "['bicycle', 'solar_panel']",
  3,
  "['Jan', 'Dec']",
  "['rely_less_food', 'reduce_meals', 'restrict_adults', 'borrow_food']",
  -19.11147023,
  33.47612499,
  705.0,
  39.0,
  'uuid:74c5d3b9-775d-4aa6-80b2-931c169198cb'),
 (339,
  'Moz',
  '21/08/2017',
  199,
  '2017-08-21T14:38:13.000Z',
  '2017-08-21T15:06:45.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  1,
  'yes',
  9,
  9,
  'no',
  17,
  'no',
  'yes',
  'no',
  'yes',
  'mabatisloping',
  'burntbricks',
  'cement',
  'no',
  4,
  3,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'yes',
  "['Sept', 'Oct']",
  1.0,
  'yes',
  'no',
  None,
  'no',
  None,
  'no',
  'never',
  'no',
  None,
  None,
  'yes',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'solar_panel', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_portion', 'reduce_meals', 'borrow_food',
'lab_ex_food']",
  -19.1115104,
  33.4759125,
  713.4000244,
  21.836,
  'uuid:27fc9dcb-aef9-4068-8464-1466fccdbe97'),
 (340,
  'Moz',
  '21/08/2017',
  198,
  '2017-08-22T04:18:31.000Z',
  '2017-08-23T06:22:40.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Nguinea',
  18,
  'no',
  6,
```

```
  6,
  'no',
  18,
  'no',
  'yes',
  'no',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  2,
  1,
  'no',
  2,
  2,
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['none']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'radio', 'sterio', 'solar_panel', 'mobile_phone']",
  3,
  "['Jan', 'Nov', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'restrict_adults',
'lab_ex_food']",
  -19.11218866,
  33.48341368,
  735.0,
  5.0,
  'uuid:f4e489af-a8b2-4dec-9b59-3baf390206bf'),
 (341,
  'Moz',
  '21/08/2017',
  200,
  '2017-08-22T04:38:27.000Z',
  '2017-08-22T08:27:40.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Madangua',
  40,
  'yes',
  3,
  3,
  'yes',
  40,
  'no',
  'no',
  'no',
  'no',
  'grass',
  'sunbricks',
  'earth',
  'no',
  2,
  1,
  'no',
  3,
  3,
  'yes',
  None,
  3.0,
  'no',
  None,
  6.0,
  'yes',
  'no',
  None,
```

```
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'yes',
 "['none']",
 None,
 1,
 'no',
 'no',
 "['solar_panel', 'solar_torch', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar']",
 "['rely_less_food', 'limit_portion', 'lab_ex_food']",
 -19.1126712,
 33.4764327,
 0.0,
 1902.0,
 'uuid:bd3d281a-6782-4d9a-bd69-d5736bc30df9'),
(342,
 'Moz',
 '22/08/2017',
 201,
 '2017-08-22T09:26:55.000Z',
 '2017-08-22T09:51:58.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Madangua',
 10,
 'yes',
 10,
 10,
 'no',
 40,
 'yes',
 'yes',
 'yes',
 'yes',
 'mabatisloping',
 'burntbricks',
 'cement',
 'no',
 1,
 3,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 10.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'table', 'mobile_phone']",
 2,
 "['Jan', 'Oct', 'Dec']",
 "['rely_less_food', 'reduce_meals']",
 -19.1114042,
 33.4759344,
 0.0,
 23.514,
 'uuid:d2b25f5f-4359-42ef-99bd-ffb59cfc8407'),
(343,
 'Moz',
 '22/08/2017',
```

```
 202,
 '2017-08-22T09:52:02.000Z',
 '2017-08-22T10:07:46.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nguinea',
 18,
 'no',
 2,
 2,
 'no',
 25,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'yes',
 None,
 2.0,
 'no',
 None,
 16.0,
 'yes',
 'no',
 None,
 'no',
 None,
 'no',
 'never',
 'no',
 None,
 None,
 'yes',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['bicycle', 'radio', 'solar_panel', 'mobile_phone']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals']",
 -19.1114087,
 33.4759125,
 0.0,
 22.566999999999997,
 'uuid:46ce8db9-5c01-48af-8b77-0d1f0d5fc258'),
(344,
 'Moz',
 '22/08/2017',
 203,
 '2017-08-22T10:10:14.000Z',
 '2017-08-22T10:24:23.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Nguinea',
 19,
 'no',
 10,
 10,
 'no',
 22,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 3,
 2,
 'no',
 2,
 2,
```

```
  'no',
  2.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'no',
  "['goats', 'poultry']",
  None,
  2,
  'yes',
  'no',
  "['bicycle', 'solar_panel', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Dec']",
  "['rely_less_food', 'limit_variety', 'limit_portion', 'reduce_meals',
'go_forest']",
  -19.11138608,
  33.47606571,
  733.0,
  118.0,
  'uuid:2abe07c7-4efa-4d0e-8e19-8712bd987845'),
 (345,
  'Moz',
  '22/08/2017',
  204,
  '2017-08-22T10:24:38.000Z',
  '2017-08-22T10:34:43.000Z',
  'Sofala',
  'Nhamatanda',
  'Lamego',
  'Ndedja',
  35,
  'no',
  2,
  2,
  'no',
  15,
  'no',
  'yes',
  'no',
  'yes',
  'grass',
  'burntbricks',
  'cement',
  'yes',
  2,
  2,
  'yes',
  3,
  3,
  'no',
  3.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'no',
  "['goats']",
  None,
  1,
  'no',
  'no',
  "['radio', 'mobile_phone']",
  3,
```

```
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'reduce_meals']",
 -19.1113884,
 33.4759125,
 0.0,
 23.504,
 'uuid:925708f3-a2d9-4331-a5d4-856efd184dc8'),
(346,
 'Moz',
 '22/08/2017',
 207,
 '2017-08-22T10:35:17.000Z',
 '2017-08-22T10:41:35.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Ndedja',
 30,
 'no',
 1,
 1,
 'no',
 55,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 1,
 1,
 'no',
 1.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['radio']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_variety', 'borrow_food']",
 -19.11098351,
 33.47564293,
 722.0,
 117.0,
 'uuid:abdba24f-7bca-4d79-8a01-273547da52f5'),
(347,
 'Moz',
 '22/08/2017',
 209,
 '2017-08-22T10:42:15.000Z',
 '2017-08-23T06:21:54.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Ndedja',
 35,
 'no',
 1,
 1,
 'yes',
 35,
 'no',
 'no',
 'no',
```

```
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 1,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'no',
 "['none']",
 None,
 1,
 'yes',
 'no',
 None,
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'reduce_meals', 'borrow_food', 'go_forest', 'lab_ex_food']",
 -19.1113139,
 33.476521999999996,
 0.0,
 82.5,
 'uuid:ef587b92-5dc4-47b3-b631-a3b191649066'),
(348,
 'Moz',
 '22/08/2017',
 205,
 '2017-08-23T05:08:13.000Z',
 '2017-08-23T05:27:06.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Ndedja',
 20,
 'no',
 8,
 8,
 'no',
 20,
 'no',
 'no',
 'no',
 'no',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 5,
 'no',
 4,
 4,
 'no',
 4.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
 'no',
 'no',
 "['pigs', 'poultry']",
 None,
 2,
 'no',
 'no',
 "['bicycle', 'radio', 'solar_torch', 'mobile_phone']",
 3,
 "['none']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'go_forest', 'lab_ex_food']",
 -19.1113139,
 33.476521999999996,
 0.0,
 72.9,
 'uuid:2eb827ea-c94d-4a55-bd64-59bc67459c17'),
(349,
 'Moz',
 '22/08/2017',
 206,
 '2017-08-23T05:39:40.000Z',
 '2017-08-23T05:53:54.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Ndedja',
 35,
 'no',
 5,
 5,
 'no',
 68,
 'yes',
 'yes',
 'yes',
 'yes',
 'grass',
 'muddaub',
 'earth',
 'no',
 1,
 2,
 'no',
 2,
 2,
 'no',
 2.0,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 'no',
 'yes',
 "['poultry']",
 None,
 1,
 'yes',
 'no',
 "['radio']",
 3,
 "['Jan', 'Feb', 'Mar', 'Dec']",
 "['rely_less_food', 'limit_variety', 'reduce_meals', 'go_forest', 'lab_ex_food']",
 -19.1114023,
 33.4758475,
 0.0,
 45.6,
 'uuid:d8dc7800-eb7d-4b8f-84be-4dcddeb55b18'),
(350,
 'Moz',
 '22/08/2017',
 208,
 '2017-08-23T05:53:59.000Z',
 '2017-08-23T06:04:39.000Z',
 'Sofala',
 'Nhamatanda',
 'Lamego',
 'Ndedja',
```

```
  10,
  'no',
  7,
  7,
  'no',
  10,
  'yes',
  'yes',
  'yes',
  'yes',
  'grass',
  'muddaub',
  'earth',
  'no',
  1,
  2,
  'no',
  3,
  3,
  'no',
  3.0,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  None,
  'no',
  'yes',
  "['poultry']",
  None,
  1,
  'yes',
  'no',
  "['bicycle', 'solar_panel', 'mobile_phone']",
  3,
  "['Jan', 'Feb', 'Mar', 'Dec']",
  "['rely_less_food', 'limit_variety', 'reduce_meals', 'day_night_hungry',
'lab_ex_food']",
  -19.11192854,
  33.476780600000005,
  696.0,
  72.0,
  'uuid:02cde43c-adcf-4b1d-8ef7-bd665ccab756')]
```

We can see it returns the data without much structure

```
type(rows)
```

```
list
```

and examine the type of one element of it as well.

```
type(rows[0])
```

```
tuple
```

we can also combine use of this library (to access databases) with `Pandas` functions to pull data from databases directly into a datagrame.

```
df = pd.read_sql_query("SELECT * FROM Farms",con)

df.head()
```

	Id	Country	A01_interview_date	A03_quest_no	A04_start	A05_end	A
0	1	Moz	17/11/2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	
1	2	Moz	17/11/2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	
2	3	Moz	17/11/2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	
3	4	Moz	17/11/2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	
4	5	Moz	17/11/2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	

5 rows × 61 columns

This is the familiar SAFI data that we've been working with

More Specific Queries

There are several ways that we can make our queries return only subsets of the data, this is where the database can start to provide real advantages, we can do sophisticated subsetting of our dataset in the database, which is optimized for performance at exactly these tasks, instead of in Python which is a general purpose programming language.

Strucutred Query Langauge (SQL) is a powerful way of working with databases. We'll use it only through pandas going forward so that we can also use the nice visualizations from the notebook, but there are many other ways to use SQL.

For example we might only be interested in the location of the farms that were surveyed, so we can choose only those columns, by putting the column names in place of the * in the query.

```
location_df = pd.read_sql_query("SELECT Country, A06_Province, A07_district, A08_ward,
A09_village FROM Farms",con)
location_df.head()
```

	Country	A06_province	A07_district	A08_ward	A09_village
0	Moz	Manica	Manica	Bandula	God
1	Moz	Manica	Manica	Bandula	God
2	Moz	Manica	Manica	Bandula	God
3	Moz	Manica	Manica	Bandula	God
4	Moz	Manica	Manica	Bandula	God

We can limit for the number of rows we get by using "LIMIT" keyword in the query.

```
pd.read_sql_query("SELECT * FROM Farms LIMIT 10",con)
```

	Id	Country	A01_interview_date	A03_quest_no	A04_start	A05_end	A
0	1	Moz	17/11/2016	1	2017-03-23T09:49:57.000Z	2017-04-02T17:29:08.000Z	
1	2	Moz	17/11/2016	1	2017-04-02T09:48:16.000Z	2017-04-02T17:26:19.000Z	
2	3	Moz	17/11/2016	3	2017-04-02T14:35:26.000Z	2017-04-02T17:26:53.000Z	
3	4	Moz	17/11/2016	4	2017-04-02T14:55:18.000Z	2017-04-02T17:27:16.000Z	
4	5	Moz	17/11/2016	5	2017-04-02T15:10:35.000Z	2017-04-02T17:27:35.000Z	
5	6	Moz	17/11/2016	6	2017-04-02T15:27:25.000Z	2017-04-02T17:28:02.000Z	
6	7	Moz	17/11/2016	7	2017-04-02T15:38:01.000Z	2017-04-02T17:28:19.000Z	
7	8	Moz	16/11/2016	8	2017-04-02T15:59:52.000Z	2017-04-02T17:28:39.000Z	
8	9	Moz	16/11/2016	9	2017-04-02T16:23:36.000Z	2017-04-02T16:42:08.000Z	
9	10	Moz	16/12/2016	10	2017-04-02T17:03:28.000Z	2017-04-02T17:25:11.000Z	

10 rows × 61 columns

Try it Yourself!

How would you return only the first 5 rows of the columns about food?

We can use a “where” keyword to filter and only choose a subset of the rows conditionally, based on *values* of the data. For example, we can look only at the farms that have living parents.

```
parents_df = pd.read_sql_query("SELECT Id, B17_parents_liv FROM Farms WHERE
B17_parents_liv = 'yes'",con)
parents_df.head()
```

	Id	B17_parents_liv
0	2	yes
1	5	yes
2	7	yes
3	8	yes
4	9	yes

Try it yourself

1. Write a queries to read from the database, the Id column and two subsets of columns to 2 separate dataframes. Then use pandas to merge the dataframes into one.

2.

Class 15: Intro to ML & Modeling

handwritten notes from class

What is a Model?

A model is a simplified representation of some part of the world. A famous quote about models is:

> All models are wrong, but some are useful –George Box[^wiki]

You might have seen models in chemistry class, for example about an atom:

read more in theModel Based Machine Learning Book

Fig. 1 Brighterorange / CC BY-SA

An atom doesn't actually *look* like this, but this is a *useful* representation to help people learn how they function.

In machine learning, we use models, that are generally *statistical* models.

> A statistical model is a mathematical model that embodies a set of statistical assumptions concerning the generation of sample data (and similar data from a larger population). A statistical model represents, often in considerably idealized form, the data-generating process
>
> —wikipedia

Types of Models in Machine Learning

Starting from a dataset, we first make an additional designation about how we will use the different variables (columns). We will call most of them the *features*, which we denote mathematically with $\mathbf{X}$ and we'll choose one to be the *target* or *labels*, denoted by $\mathbf{y}$.

The core assumption for just about all machine learning is that there exists some function f so that for the ith sample

$$ y_i = f(\mathbf{x}_i) $$

Then with different additional assumptions we get different types of machine learning:

- if both features ($\mathbf{X}$) and target ($\mathbf{y}$) are observed (contained in our dataset) it's **supervised learning** code
- if only the features ($\mathbf{X}$) are observed, it's **unsupervised learning** code

There are more types, that we won't cover much in class but will mention here for completeness:

- if we have many samples for the features ($\mathbf{X}$) but only labels for some it's **semi-supervised learning**
- if we have only features ($\mathbf{X}$) and an oracle (a person or some other labeler) that we can ask for labels from with a budget, so we can't get all of them, it's *active learning)*
- reinforcement learning involves taking actions and getting rewards.

Tip

a common convention in mathematical notation is to denote matrices (tables) by capitalized bold characters , vectors (lists) by lowercase bold characters and scalars (single values) by lowercase regular characters.

For example:

- all of the features would be: $\mathbf{X}$
- one variable for all samples or all variables for one sample would be $\mathbf{x}$
- one variable for one sample would be x

Supervised Learning

we'll focus on supervised learning first. we can take that same core assumption and use it with additional information about our target variable to determine learning **task** we are working to do.

\[y_i = f(\mathbf{x}_i) \]

- if \(y_i\) are discrete (eg flower species) we are doing **classification**
- if \(y_i\) are continuous (eg height) we are doing **regression**

Machine Learning Pipeline

To do machine learning we start with **training data** which we put as input to the **learning algorithm**. A learning algorithm might be a generic optimization procedure or a specialized procedure for a specific model. The learning algorithm outputs a trained **model** or the parameters of the model. When we deploy a model we pair the **fit model** with a **prediction algorithm** or **decision** algorithm to evaluate a new sample in the world.

In experimenting and design, we need **testing data** to evaluate how well our learning algorithm understood the world. We need to use previously unseen data, because if we don't we can't tell if the prediction algorithm is using a rule that the learning algorithm produced or just looking up from a lookup table the result. This can be thought of like the difference between memorization and understanding.

When the model does well on the training data, but not on test data, we say that it does not generalize well.

Try it yourself

1. List different machine learning applications you've interacted with and try to figure out what the features would be, what the target would be, and then what type of learning it is

Glossary

> **Tip**
>
> This week we've learned a lot of new terms. Contribute definition below to form a glossary.

Term	Definition
Model	a mathematical representation of assumptions about the world

Class 16: Naive Bayes Classification

To learn a classifier, we need labeled data (features and target)

We split our data twice:

- sample-wise: test and train
- variable-wise: features and target

Naive Bayes with Sci-kit Learn

We will use a new package today, Scikit-Learn. Its package name for importing is `sklearn` but we don't import it with an alias, in general. It's a large module and we most often import just the parts we need.

To do that we use a new Python keyword `from`. We can identify a package and then import a submodule or a package and submodule with `.` and then import specific functions or classes.

> **Tip**
>
> Recall when a word turns green & bold in a notebook, it's a python keyword, or reserved word.

```
import pandas as pd
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
```

We can tell from this code that `test_train_split` is probably a function because it's in lowercase and `sklearn` follows PEP 8 the Python Style Guide pretty strictly. We can also check with type

```
type(train_test_split)
```

```
function
```

We can tell `GaussianNB` is probably a class because it's in CapWords, also known as camel case.

Again we can check.

```
type(GaussianNB)
```

```
abc.ABCMeta
```

That's an abstract base class.

Today we'll work with the iris dataset, which has been used for demonstrating statistical analyses since 1936. It contains 4 measurements of flowers from 3 different species.

```
iris_df = sns.load_dataset('iris')
```

As usual, we look at the structure.

```
iris_df.head()
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Next we examine the data litle further to consider what the structure is like for classification purposes.

```
sns.pairplot(data= iris_df, hue='species')
```

```
<seaborn.axisgrid.PairGrid at 0x7f003d1fee50>
```

In order for classification to work, we're looking to see that the groups of samples of different classes (here, species) do not overlap too much. We're also looking at the shape of the groups of samples of each class.

Naive Bayes

Naive Bayes assumes that the features are uncorrelated (Naive) and that we can pick the most probable class. It can assume different distributions of the features conditioned on the class, though. We'll use Gaussian Naive Bayes. Gaussian distributed samples in the plot above would be roughly ovals with more points at the center and uncorrelated Gaussian data would be distributed in circles, not ovals. We have some ovals and some overlap, but enough we can expect this classifier to work pretty well.

First we instantiate the classifier object with the constructor method.

```
gnb = GaussianNB()
```

We just made it exist so far, nothing more, so we can check its type.

```
type(gnb)
```

```
sklearn.naive_bayes.GaussianNB
```

We can also use the `get_params` method to see what it looks like.

```
gnb.get_params()
```

```
{'priors': None, 'var_smoothing': 1e-09}
```

Training a Model

Before we trian the model, we have to get our data out of the DataFrame, because our `gnb.fit` method takes arrays. We can use the `values` attribute

```
iris_df.values
```

```
array([[5.1, 3.5, 1.4, 0.2, 'setosa'],
       [4.9, 3.0, 1.4, 0.2, 'setosa'],
       [4.7, 3.2, 1.3, 0.2, 'setosa'],
       [4.6, 3.1, 1.5, 0.2, 'setosa'],
       [5.0, 3.6, 1.4, 0.2, 'setosa'],
       [5.4, 3.9, 1.7, 0.4, 'setosa'],
       [4.6, 3.4, 1.4, 0.3, 'setosa'],
       [5.0, 3.4, 1.5, 0.2, 'setosa'],
       [4.4, 2.9, 1.4, 0.2, 'setosa'],
       [4.9, 3.1, 1.5, 0.1, 'setosa'],
       [5.4, 3.7, 1.5, 0.2, 'setosa'],
       [4.8, 3.4, 1.6, 0.2, 'setosa'],
       [4.8, 3.0, 1.4, 0.1, 'setosa'],
       [4.3, 3.0, 1.1, 0.1, 'setosa'],
       [5.8, 4.0, 1.2, 0.2, 'setosa'],
       [5.7, 4.4, 1.5, 0.4, 'setosa'],
       [5.4, 3.9, 1.3, 0.4, 'setosa'],
       [5.1, 3.5, 1.4, 0.3, 'setosa'],
       [5.7, 3.8, 1.7, 0.3, 'setosa'],
       [5.1, 3.8, 1.5, 0.3, 'setosa'],
       [5.4, 3.4, 1.7, 0.2, 'setosa'],
       [5.1, 3.7, 1.5, 0.4, 'setosa'],
       [4.6, 3.6, 1.0, 0.2, 'setosa'],
       [5.1, 3.3, 1.7, 0.5, 'setosa'],
       [4.8, 3.4, 1.9, 0.2, 'setosa'],
       [5.0, 3.0, 1.6, 0.2, 'setosa'],
       [5.0, 3.4, 1.6, 0.4, 'setosa'],
       [5.2, 3.5, 1.5, 0.2, 'setosa'],
       [5.2, 3.4, 1.4, 0.2, 'setosa'],
       [4.7, 3.2, 1.6, 0.2, 'setosa'],
       [4.8, 3.1, 1.6, 0.2, 'setosa'],
       [5.4, 3.4, 1.5, 0.4, 'setosa'],
       [5.2, 4.1, 1.5, 0.1, 'setosa'],
       [5.5, 4.2, 1.4, 0.2, 'setosa'],
       [4.9, 3.1, 1.5, 0.2, 'setosa'],
       [5.0, 3.2, 1.2, 0.2, 'setosa'],
       [5.5, 3.5, 1.3, 0.2, 'setosa'],
```

```
[4.9, 3.6, 1.4, 0.1, 'setosa'],
[4.4, 3.0, 1.3, 0.2, 'setosa'],
[5.1, 3.4, 1.5, 0.2, 'setosa'],
[5.0, 3.5, 1.3, 0.3, 'setosa'],
[4.5, 2.3, 1.3, 0.3, 'setosa'],
[4.4, 3.2, 1.3, 0.2, 'setosa'],
[5.0, 3.5, 1.6, 0.6, 'setosa'],
[5.1, 3.8, 1.9, 0.4, 'setosa'],
[4.8, 3.0, 1.4, 0.3, 'setosa'],
[5.1, 3.8, 1.6, 0.2, 'setosa'],
[4.6, 3.2, 1.4, 0.2, 'setosa'],
[5.3, 3.7, 1.5, 0.2, 'setosa'],
[5.0, 3.3, 1.4, 0.2, 'setosa'],
[7.0, 3.2, 4.7, 1.4, 'versicolor'],
[6.4, 3.2, 4.5, 1.5, 'versicolor'],
[6.9, 3.1, 4.9, 1.5, 'versicolor'],
[5.5, 2.3, 4.0, 1.3, 'versicolor'],
[6.5, 2.8, 4.6, 1.5, 'versicolor'],
[5.7, 2.8, 4.5, 1.3, 'versicolor'],
[6.3, 3.3, 4.7, 1.6, 'versicolor'],
[4.9, 2.4, 3.3, 1.0, 'versicolor'],
[6.6, 2.9, 4.6, 1.3, 'versicolor'],
[5.2, 2.7, 3.9, 1.4, 'versicolor'],
[5.0, 2.0, 3.5, 1.0, 'versicolor'],
[5.9, 3.0, 4.2, 1.5, 'versicolor'],
[6.0, 2.2, 4.0, 1.0, 'versicolor'],
[6.1, 2.9, 4.7, 1.4, 'versicolor'],
[5.6, 2.9, 3.6, 1.3, 'versicolor'],
[6.7, 3.1, 4.4, 1.4, 'versicolor'],
[5.6, 3.0, 4.5, 1.5, 'versicolor'],
[5.8, 2.7, 4.1, 1.0, 'versicolor'],
[6.2, 2.2, 4.5, 1.5, 'versicolor'],
[5.6, 2.5, 3.9, 1.1, 'versicolor'],
[5.9, 3.2, 4.8, 1.8, 'versicolor'],
[6.1, 2.8, 4.0, 1.3, 'versicolor'],
[6.3, 2.5, 4.9, 1.5, 'versicolor'],
[6.1, 2.8, 4.7, 1.2, 'versicolor'],
[6.4, 2.9, 4.3, 1.3, 'versicolor'],
[6.6, 3.0, 4.4, 1.4, 'versicolor'],
[6.8, 2.8, 4.8, 1.4, 'versicolor'],
[6.7, 3.0, 5.0, 1.7, 'versicolor'],
[6.0, 2.9, 4.5, 1.5, 'versicolor'],
[5.7, 2.6, 3.5, 1.0, 'versicolor'],
[5.5, 2.4, 3.8, 1.1, 'versicolor'],
[5.5, 2.4, 3.7, 1.0, 'versicolor'],
[5.8, 2.7, 3.9, 1.2, 'versicolor'],
[6.0, 2.7, 5.1, 1.6, 'versicolor'],
[5.4, 3.0, 4.5, 1.5, 'versicolor'],
[6.0, 3.4, 4.5, 1.6, 'versicolor'],
[6.7, 3.1, 4.7, 1.5, 'versicolor'],
[6.3, 2.3, 4.4, 1.3, 'versicolor'],
[5.6, 3.0, 4.1, 1.3, 'versicolor'],
[5.5, 2.5, 4.0, 1.3, 'versicolor'],
[5.5, 2.6, 4.4, 1.2, 'versicolor'],
[6.1, 3.0, 4.6, 1.4, 'versicolor'],
[5.8, 2.6, 4.0, 1.2, 'versicolor'],
[5.0, 2.3, 3.3, 1.0, 'versicolor'],
[5.6, 2.7, 4.2, 1.3, 'versicolor'],
[5.7, 3.0, 4.2, 1.2, 'versicolor'],
[5.7, 2.9, 4.2, 1.3, 'versicolor'],
[6.2, 2.9, 4.3, 1.3, 'versicolor'],
[5.1, 2.5, 3.0, 1.1, 'versicolor'],
[5.7, 2.8, 4.1, 1.3, 'versicolor'],
[6.3, 3.3, 6.0, 2.5, 'virginica'],
[5.8, 2.7, 5.1, 1.9, 'virginica'],
[7.1, 3.0, 5.9, 2.1, 'virginica'],
[6.3, 2.9, 5.6, 1.8, 'virginica'],
[6.5, 3.0, 5.8, 2.2, 'virginica'],
[7.6, 3.0, 6.6, 2.1, 'virginica'],
[4.9, 2.5, 4.5, 1.7, 'virginica'],
[7.3, 2.9, 6.3, 1.8, 'virginica'],
[6.7, 2.5, 5.8, 1.8, 'virginica'],
[7.2, 3.6, 6.1, 2.5, 'virginica'],
[6.5, 3.2, 5.1, 2.0, 'virginica'],
[6.4, 2.7, 5.3, 1.9, 'virginica'],
[6.8, 3.0, 5.5, 2.1, 'virginica'],
[5.7, 2.5, 5.0, 2.0, 'virginica'],
[5.8, 2.8, 5.1, 2.4, 'virginica'],
[6.4, 3.2, 5.3, 2.3, 'virginica'],
[6.5, 3.0, 5.5, 1.8, 'virginica'],
[7.7, 3.8, 6.7, 2.2, 'virginica'],
[7.7, 2.6, 6.9, 2.3, 'virginica'],
[6.0, 2.2, 5.0, 1.5, 'virginica'],
[6.9, 3.2, 5.7, 2.3, 'virginica'],
[5.6, 2.8, 4.9, 2.0, 'virginica'],
[7.7, 2.8, 6.7, 2.0, 'virginica'],
[6.3, 2.7, 4.9, 1.8, 'virginica'],
```

```
       [6.7, 3.3, 5.7, 2.1, 'virginica'],
       [7.2, 3.2, 6.0, 1.8, 'virginica'],
       [6.2, 2.8, 4.8, 1.8, 'virginica'],
       [6.1, 3.0, 4.9, 1.8, 'virginica'],
       [6.4, 2.8, 5.6, 2.1, 'virginica'],
       [7.2, 3.0, 5.8, 1.6, 'virginica'],
       [7.4, 2.8, 6.1, 1.9, 'virginica'],
       [7.9, 3.8, 6.4, 2.0, 'virginica'],
       [6.4, 2.8, 5.6, 2.2, 'virginica'],
       [6.3, 2.8, 5.1, 1.5, 'virginica'],
       [6.1, 2.6, 5.6, 1.4, 'virginica'],
       [7.7, 3.0, 6.1, 2.3, 'virginica'],
       [6.3, 3.4, 5.6, 2.4, 'virginica'],
       [6.4, 3.1, 5.5, 1.8, 'virginica'],
       [6.0, 3.0, 4.8, 1.8, 'virginica'],
       [6.9, 3.1, 5.4, 2.1, 'virginica'],
       [6.7, 3.1, 5.6, 2.4, 'virginica'],
       [6.9, 3.1, 5.1, 2.3, 'virginica'],
       [5.8, 2.7, 5.1, 1.9, 'virginica'],
       [6.8, 3.2, 5.9, 2.3, 'virginica'],
       [6.7, 3.3, 5.7, 2.5, 'virginica'],
       [6.7, 3.0, 5.2, 2.3, 'virginica'],
       [6.3, 2.5, 5.0, 1.9, 'virginica'],
       [6.5, 3.0, 5.2, 2.0, 'virginica'],
       [6.2, 3.4, 5.4, 2.3, 'virginica'],
       [5.9, 3.0, 5.1, 1.8, 'virginica']], dtype=object)
```

Then we create test an train splits of our data. we'll to equal parts (`test_size = .5`) and set the random state so that we call get the same result

```
X_train, X_test,  y_train, y_test = train_test_split(iris_df.values[:,:4],
                                                     iris_df.values[:,-1],
                                                    test_size = .5,
                                                    random_state =0)
```

try it yourself!

1. rerun the test `train_test_split` without random state, to see how it's different
2. change the `test_size` to different sizes and see what happens.

Now we use the *training* data to fit our model.

```
gnb.fit(X_train, y_train)
```

```
GaussianNB()
```

Now we can predict using the *test* data's features only.

```
y_pred = gnb.predict(X_test)
y_pred
```

```
array(['virginica', 'versicolor', 'setosa', 'virginica', 'setosa',
       'virginica', 'setosa', 'versicolor', 'versicolor', 'versicolor',
       'versicolor', 'versicolor', 'versicolor', 'versicolor',
       'versicolor', 'setosa', 'versicolor', 'versicolor', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'virginica', 'setosa', 'setosa', 'versicolor', 'versicolor',
       'setosa', 'virginica', 'versicolor', 'setosa', 'virginica',
       'virginica', 'versicolor', 'setosa', 'versicolor', 'versicolor',
       'versicolor', 'virginica', 'setosa', 'virginica', 'setosa',
       'setosa', 'versicolor', 'virginica', 'virginica', 'versicolor',
       'virginica', 'versicolor', 'virginica', 'versicolor', 'versicolor',
       'virginica', 'versicolor', 'versicolor', 'virginica', 'versicolor',
       'virginica', 'versicolor', 'setosa', 'virginica', 'versicolor',
       'versicolor', 'versicolor', 'versicolor', 'virginica', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'versicolor'], dtype='<U10')
```

We can compare this to the y_test to see how well our classifier works.

```
y_test
```

```
array(['virginica', 'versicolor', 'setosa', 'virginica', 'setosa',
       'virginica', 'setosa', 'versicolor', 'versicolor', 'versicolor',
       'virginica', 'versicolor', 'versicolor', 'versicolor',
       'versicolor', 'setosa', 'versicolor', 'versicolor', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'virginica', 'setosa', 'setosa', 'versicolor', 'versicolor',
       'setosa', 'virginica', 'versicolor', 'setosa', 'virginica',
       'virginica', 'versicolor', 'setosa', 'versicolor', 'versicolor',
       'versicolor', 'virginica', 'setosa', 'virginica', 'setosa',
       'setosa', 'versicolor', 'virginica', 'virginica', 'virginica',
       'virginica', 'versicolor', 'virginica', 'versicolor', 'versicolor',
       'virginica', 'virginica', 'virginica', 'virginica', 'versicolor',
       'virginica', 'versicolor', 'setosa', 'virginica', 'versicolor',
       'versicolor', 'versicolor', 'versicolor', 'virginica', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'versicolor'], dtype=object)
```

We can simply use base python to get the number correct with a boolean and sum since `False` turns to 0 and `True` to 1.

```
sum(y_pred == y_test)
```

```
71
```

and compare that with the total

```
len(y_pred)
```

```
75
```

or compute accuracy

```
sum(y_pred == y_test)/len(y_pred)
```

```
0.9466666666666667
```

Questions after class

I'm curious about other classifiers

Great! We'll see more Friday, next week, and next month. You're also encouraged to try out others and experiment with sci-kit learn.

How could this be related to NLP?

Naive Bayes isn't a good classifier for NLP most of the time because word are not independent, but for very coarse distinctions it can be applied. The *Gaussian* naive Bayes we used today doesn't apply to NLP without a lot of transformation because in general, words are more like categorical (or binary/dummies) variables than continuous valued.

We'll work with text data toward the end of the semester.

How we got the values from the DataFrame

A DataFrame is an object and one of the attributes is `values`, we accessed that directly. However, there is a new, safer way through the `to_numpy` method.

Class 17: Evaluating Classification and Midsemester Feedback

1. share your favorite rainy day activity (or just say hi) in the zoom chat for attendance
2. log onto prismia

Naive Bayes Review

Main assumptions:

- classification assumes that features will separate the gorups
- NB: conditionally independent

```
# %load http://drsmb.co/310
import pandas as pd
import seaborn as sns
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
```

```
iris = sns.load_dataset("iris")
iris.head()
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

```
X_train, X_test, y_train, y_test = train_test_split(iris.values[:,:4],
                                                    iris.species.values,
                                                    test_size=0.5, random_state=0)
```

```
gnb = GaussianNB()
y_pred = gnb.fit(X_train, y_train).predict(X_test)
```

```
y_pred
```

```
array(['virginica', 'versicolor', 'setosa', 'virginica', 'setosa',
       'virginica', 'setosa', 'versicolor', 'versicolor', 'versicolor',
       'versicolor', 'versicolor', 'versicolor', 'versicolor',
       'versicolor', 'setosa', 'versicolor', 'versicolor', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'virginica', 'setosa', 'setosa', 'versicolor', 'versicolor',
       'setosa', 'virginica', 'versicolor', 'setosa', 'virginica',
       'virginica', 'versicolor', 'setosa', 'versicolor', 'versicolor',
       'versicolor', 'virginica', 'setosa', 'virginica', 'setosa',
       'setosa', 'versicolor', 'virginica', 'virginica', 'versicolor',
       'virginica', 'versicolor', 'virginica', 'versicolor', 'versicolor',
       'virginica', 'versicolor', 'versicolor', 'virginica', 'versicolor',
       'virginica', 'versicolor', 'setosa', 'virginica', 'versicolor',
       'versicolor', 'versicolor', 'versicolor', 'virginica', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'versicolor'], dtype='<U10')
```

```
y_test
```

```
array(['virginica', 'versicolor', 'setosa', 'virginica', 'setosa',
       'virginica', 'setosa', 'versicolor', 'versicolor', 'versicolor',
       'virginica', 'versicolor', 'versicolor', 'versicolor',
       'versicolor', 'setosa', 'versicolor', 'versicolor', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'virginica', 'setosa', 'setosa', 'versicolor', 'versicolor',
       'setosa', 'virginica', 'versicolor', 'setosa', 'virginica',
       'virginica', 'versicolor', 'setosa', 'versicolor', 'versicolor',
       'versicolor', 'virginica', 'setosa', 'virginica', 'setosa',
       'setosa', 'versicolor', 'virginica', 'virginica', 'virginica',
       'virginica', 'versicolor', 'virginica', 'versicolor', 'versicolor',
       'virginica', 'virginica', 'virginica', 'virginica', 'versicolor',
       'virginica', 'versicolor', 'setosa', 'virginica', 'versicolor',
       'versicolor', 'versicolor', 'versicolor', 'virginica', 'setosa',
       'setosa', 'virginica', 'versicolor', 'setosa', 'setosa',
       'versicolor'], dtype=object)
```

```
sum(y_pred == y_test)
```

```
71
```

```
len(y_pred)
```

```
75
```

Evaluating Classification

We can use the `score` method to compute accuracy by providing both features and target for the test set.

```
gnb.score(X_test, y_test)
```

```
0.9466666666666667
```

Which is the same as the manual way we did before

```
sum(y_pred == y_test)/len(y_pred)
```

```
0.9466666666666667
```

Scikit learn also provides a whole metrics module. We'll look at two functions from it today.

```
from sklearn.metrics import confusion_matrix, classification_report
```

First a Confusion matrix, which is the basic table of how the decisions were made. It counts how many from each true class were predicted to be in each class. My favorite refernce on these is the table on the wikipedia article.

```
confusion_matrix(y_test,y_pred,)
```

```
array([[21,  0,  0],
       [ 0, 30,  0],
       [ 0,  4, 20]])
```

We can see this makes sense from looking at the data.

```
sns.pairplot(data =iris, hue='species')
```

```
<seaborn.axisgrid.PairGrid at 0x7f3233337210>
```

We can also get a formatted report that uses the confusion matrix to compute multiple metrics

```
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

      setosa       1.00      1.00      1.00        21
  versicolor       0.88      1.00      0.94        30
   virginica       1.00      0.83      0.91        24

    accuracy                           0.95        75
   macro avg       0.96      0.94      0.95        75
weighted avg       0.95      0.95      0.95        75
```

Inspecting a classifier

We can serialize any python object to inspect it with `__dict__`

```
gnb.__dict__
```

```
{'priors': None,
 'var_smoothing': 1e-09,
 'n_features_in_': 4,
 'epsilon_': 3.6399040000000003e-09,
 'classes_': array(['setosa', 'versicolor', 'virginica'], dtype='<U10'),
 'theta_': array([[4.97586207, 3.35862069, 1.44827586, 0.23448276],
        [5.935     , 2.71      , 4.185     , 1.3       ],
        [6.77692308, 3.09230769, 5.73461538, 2.10769231]]),
 'sigma_': array([[0.10321047, 0.13208086, 0.01629013, 0.00846612],
        [0.256275  , 0.0829    , 0.255275  , 0.046     ],
        [0.38869823, 0.10147929, 0.31303255, 0.04763314]]),
 'class_count_': array([29., 20., 26.]),
 'class_prior_': array([0.38666667, 0.26666667, 0.34666667])}
```

It includes attributes `theta_` and `sigma_` these characterize the mean and the variance of each class. Naive Bayes is a generative classifier, since we're using the Guassian NB, it means it assumes that the data are Gaussian in each class. For a generative classifier, we can draw synthetic data from the model it learned. We can use this to decide if what it learned makes sense by comparing it qualitatively (or statistically) to the real data.

```
# %load http://drsmb.co/310
gnb_df = pd.DataFrame(np.concatenate([np.random.multivariate_normal(mu,
sig*np.eye(4),20)
                                      for mu, sig in zip(gnb.theta_,gnb.sigma_)]))
gnb_df['species'] = [ci for cl in [[c]*20 for c in gnb.classes_] for ci in cl]
sns.pairplot(data =gnb_df, hue='species')
```

```
<seaborn.axisgrid.PairGrid at 0x7f322e750910>
```

Feedback

Please complete this feedback survey to provide input on how the semester is going for you so far.

More practice

1. Try breaking down all the steps that are in the last cell one by one to understand them.
2. Read the documentation for naive bayes, how does it use the model parameters to make a prediction?

Class 18: Mid Semester Checkin, Git, & How GNB makes decisions

- Share your favorite thing about fall (or just say hi) in the zoom chat for attendance
- log onto Prismia
- accept Assignment 6
- install git or GitBash if needed

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
sns.set_theme(font_scale = 2)
```

Feedback Review

Thank you for the feedback. You can add more on Friday's notes if you did not give any yet.

```
feedback_df = pd.read_csv('data/feedback_structured.csv')
feedback_df.head()
```

	How much do you think you've learned so far this semester?	How much of the material that's been taught do you feel you understand?	How do you think the achievements you've earned so far align with your understanding?	Rank the following as what you feel the grading (when you do/not earn achievements) so far actually reflects about your performance in the class [How well I follow instructions]	Rank the following as what you feel the grading (when you do/not earn achievements) so far actually reflects about your performance in the class [What I understand about the material]	Rank th following a what you fe the gradir (when yo do/not ea achievement so far actual reflects abo you performanc in the clas [How muc effort I p int assignment
0	4	3	I think they reflect my understanding well	Reflected moderately in the grading	Reflected strongly in the grading	Reflecte perfectly in th gradir
1	4	4	I think the achievements underestimate what I ...	Reflected strongly in the grading	Reflected strongly in the grading	Reflecte moderately the gradir
2	3	3	I think they reflect my understanding well	Reflected moderately in the grading	Reflected moderately in the grading	Reflecte moderately the gradir
3	4	3	I think they reflect my understanding well	Reflected moderately in the grading	Reflected moderately in the grading	Reflecte moderately the gradir
4	3	3	I think they reflect my understanding well	Reflected strongly in the grading	Reflected moderately in the grading	Reflecte strongly in th gradir

```
el_meaning = {1: 'much less',
 2: 'a little less',
 3: 'about right',
 4: 'a little more ',
 5: 'much more'}

expected_learning = feedback_df[feedback_df.columns[0]]
el_counts,_ = np.histogram(expected_learning,bins = [i+.5 for i in range(6)])
el_df = pd.DataFrame(data = el_counts,index = el_meaning.values(),columns=
[expected_learning._name],)
# el_df.reset_index(inplace=True)
# el_df = el_df.rename(columns= {'index':'amount'}).set_index('amount')
el_df.rename_axis(index='amount relative to expectations',inplace=True)
el_df.plot.bar(legend=False,title=expected_learning._name,);
```

```
u_meaning = {0:'None', 1:'A little',3:'Half', 4:'Almost all',5:'All'}
u_col = feedback_df[feedback_df.columns[1]]
u_counts,_ = np.histogram(u_col,bins = [i+.5 for i in range(6)])
understanding_df = pd.DataFrame(data = u_counts,index = u_meaning.values(),columns=
[u_col._name],)

understanding_df.rename_axis(index='amount of material',inplace=True)
understanding_df.plot.bar(legend=False,title=u_col._name,);
```

```
align = [c_name for c_name in feedback_df.columns if 'align' in c_name][0]
align_df = feedback_df[align].copy()
align_counts = align_df.value_counts()
align_kw = {'well':'well aligned','more':'I know more','less':'I know less'}
# # kw_replace =
align_counts.rename(index={s: [align_kw[kw] for kw in align_kw.keys() if kw in s][0]
for s in align_counts.index},inplace=True)
align_counts.plot.bar(title=align_counts._name);
```

```
reflect_cols = [c_name for c_name in feedback_df.columns if 'reflects' in c_name]
reflect_df = feedback_df[reflect_cols].copy()
reflect_df.rename(columns = lambda c: c.split('[')[1].replace(']',''),inplace=True)
reflect_df = reflect_df.melt(var_name='factor',value_name='grading reflection')
reflect_df['grading reflection'] = reflect_df['grading reflection'].str.replace(' in
the grading','')
reflect_df['grading reflection'] = reflect_df['grading
reflection'].str.replace('Reflected','')

factor_kw = ['follow instructions','understand','effort']
# kw_replace =
reflect_df.replace({s: [kw for kw in factor_kw if kw in s][0] for s in
reflect_df['factor']},inplace=True)
g = sns.displot(data = reflect_df, x='grading reflection',col='factor', hue = 'grading
reflection',)
g.set(xticklabels=[]);
```

```
fair_cols = [c_name for c_name in feedback_df.columns if 'fair' in c_name]
fair_df = feedback_df[fair_cols].copy()
fair_df.rename(columns = lambda c: c.split('[')[1].replace(']',''),inplace=True)
fair_df = fair_df.melt(var_name='factor',value_name='fairness')
fair_df['fairness'] = fair_df['fairness'].str.replace(' in the grading','')
# reflect_df['grading reflection'] = reflect_df['grading
reflection'].str.replace('Reflected','')

factor_kw = ['follow instructions','understand','effort']
# # kw_replace =
fair_df.replace({s: [kw for kw in factor_kw if kw in s][0] for s in
fair_df['factor']},inplace=True)
g = sns.displot(data = reflect_df, x='grading reflection',col='factor', hue = 'grading
reflection',)
g.set(xticklabels=[]);
# fair_df
```

```
activity_col = [cname for cname in feedback_df.columns if 'outside' in cname][0]
activity_counts =
pd.get_dummies(feedback_df[activity_col].str.split(',').apply(pd.Series).stack()).sum()
activity_counts = activity_counts.rename_axis(index='activites outside of class')
# activity_counts.plot.bar()
ac_clean = activity_counts.reset_index().rename(columns={0:'students'})
sns.barplot(data=ac_clean,y='activites outside of class',x='students',);
```

Changes based on your feedback for the rest of the semester

- Assignments posted by Wed, due Tuesday (after this one, but I'm giving class time for it)
- Assignments will have a bit more detail or advice (this one is very structured because of the short time, but not all will be this closed).
- Office hours will stay the same, but deadlines moved
- Will collect descriptions of any issued with Prismia.chat to pass on to the developers

Assignment

- Accept
- Git demo
- Read & make notes of a plan (what do you need to do? what methods will you use?)
- Discuss your plan in a breakout room

Questions:

1. Errors means misclassifications, samples where the classifier's prediction is not correct.

More on Naive Bayes

From last week:

```
# %load http://drsmb.co/310
import pandas as pd
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
```

```
iris = sns.load_dataset("iris")

X_train, X_test, y_train, y_test = train_test_split(iris.values[:,:4],
                                                    iris.values[:,-1],
                                                    test_size=0.5, random_state=5)


gnb = GaussianNB()
y_pred = gnb.fit(X_train, y_train).predict(X_test)
```

Remember we can check accuracy with the score method

```
gnb.score(X_test,y_test)
```

```
0.9466666666666667
```

We can inspect the trained classifier

```
gnb.__dict__
```

```
{'priors': None,
 'var_smoothing': 1e-09,
 'n_features_in_': 4,
 'epsilon_': 3.3870506666666674e-09,
 'classes_': array(['setosa', 'versicolor', 'virginica'], dtype='<U10'),
 'theta_': array([[4.98571429, 3.40714286, 1.47142857, 0.23214286],
        [5.96190476, 2.76666667, 4.34285714, 1.31904762],
        [6.61923077, 2.95384615, 5.59230769, 2.01153846]]),
 'sigma_': array([[0.15551021, 0.17709184, 0.03275511, 0.00718113],
        [0.23188209, 0.0984127 , 0.15482994, 0.03392291],
        [0.48232249, 0.1071006 , 0.34686391, 0.06948225]]),
 'class_count_': array([28., 21., 26.]),
 'class_prior_': array([0.37333333, 0.28      , 0.34666667])}
```

Gaussian Niave Bayes learns the mean and variance for each feature in each class. We can use that to generate synthetic data.

```
df = pd.DataFrame(np.concatenate([np.random.multivariate_normal(mu, sig*np.eye(4),20)
                                  for mu, sig in zip(gnb.theta_,gnb.sigma_)]))
df['species'] = [ci for cl in [[c]*20 for c in gnb.classes_] for ci in cl]
sns.pairplot(data =df, hue='species')
```

```
<seaborn.axisgrid.PairGrid at 0x7fbb4c0d68d0>
```

It decides for it's predictions by computing probabilities and predicting the highest probability one. We can have it return these probabilities directly.

```
gnb.predict_proba(X_test)`
```

```
  File "<ipython-input-14-e6a42f6fa178>", line 1
    gnb.predict_proba(X_test)`
                             ^
SyntaxError: invalid syntax
```

We can use these to see how confident it is on each point. First, we'll get the probability of the predicted class (the max in each row of the matrix).

```
pob_ypred = np.max(gnb.predict_proba(X_test),axis=1)
pob_ypred
```

```
array([0.9999498 , 0.70485761, 0.99999988, 1.        , 1.        ,
       0.99999528, 1.        , 0.88532403, 1.        , 0.98995685,
       0.98928212, 0.69406879, 0.99999999, 0.99999997, 1.        ,
       1.        , 0.90927501, 0.99999997, 1.        , 1.        ,
       0.99972898, 0.99999839, 1.        , 0.92449508, 0.99970908,
       0.99999997, 0.99857943, 0.99965594, 0.98844145, 0.99999983,
       1.        , 0.99993213, 0.9999939 , 1.        , 0.80885454,
       1.        , 1.        , 1.        , 1.        , 0.99999993,
       0.98915134, 0.99998516, 1.        , 1.        , 0.95513579,
       0.85016072, 0.99525954, 0.99999988, 1.        , 1.        ,
       0.98735446, 1.        , 1.        , 1.        , 1.        ,
       0.99906987, 0.67288225, 0.99999514, 0.9999897 , 0.96679497,
       0.99999729, 0.99986416, 0.99999998, 0.55849652, 1.        ,
       0.98840745, 1.        , 0.94307153, 0.99999991, 0.99990409,
       1.        , 0.99977424, 0.92205013, 0.99937796, 0.99994699])
```

Now, we'll put the test data into a `DataFrame` so we can use high level plotting functions from `pandas` and `seaborn`. We could use the `numpy` data structures that scikit learn uses directly with `matplotlib` plotting and `numpy` statistics, but this is a little more compact and readable.

```
iris_test_df = pd.DataFrame(X_test,columns = iris.columns[:4])
iris_test_df['species'] = y_test
iris_test_df['species_pred'] = y_pred
iris_test_df['pob_ypred'] = pob_ypred
iris_test_df.head()
```

	sepal_length	sepal_width	petal_length	petal_width	species	species_pred	pob_ypr
0	5.8	2.7	3.9	1.2	versicolor	versicolor	0.9999
1	6.1	2.6	5.6	1.4	virginica	virginica	0.7048
2	5.8	2.8	5.1	2.4	virginica	virginica	1.0000
3	4.4	3.2	1.3	0.2	setosa	setosa	1.0000
4	7.2	3.6	6.1	2.5	virginica	virginica	1.0000

First we'll look at the one pairwise view of the data that was the most visually separate, `'petal_width'` vs `'petal_length'`

```
iris_test_df.plot.scatter(x='petal_width', y = 'petal_length',c='pob_ypred')
```

```
<AxesSubplot:xlabel='petal_width', ylabel='petal_length'>
```

Notice that the points near the boundary are the lowest probability. We can inspect further by adding a `'correct'` column with boolean values

```
iris_test_df['correct'] = iris_test_df['species'] == iris_test_df['species_pred']
```

Now we can use EDA (exploratory data analysis) to further examine.

```
iris_test_df.groupby('correct')['pob_ypred'].describe()
```

	count	mean	std	min	25%	50%	75%	max
correct								
False	4.0	0.770240	0.176779	0.558497	0.660176	0.789696	0.899761	0.943072
True	71.0	0.980814	0.059866	0.672882	0.999224	1.000000	1.000000	1.000000

The incorrectly (`False`) predicted samples were much lower confidence and higher variance in confidence. This suggests that the classifier is a good fit for the data, though it's imperfect on the test set, and that we should trust how it's working.

Try it yourself

1. Can any of these things help with Assignment 6?
2. Review the datasets you've used for prior analyses, can they be used for any classification?
3. What kinds of data would be needed to train a classifier for some ML systems that you interact with?

Class 19: Decision Trees

1. log onto Prismia
2. say hello in the zoom chat

```
# %load http://drsmb.co/310
import pandas as pd
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.datasets import load_iris
from sklearn import tree
```

First we'll review the datasets from the assignment briefly. More details on them will be on the solutions repo.

```
a6_data = 'https://raw.githubusercontent.com/rhodyprog4ds/06-naive-
bayes/main/data/dataset'
data_urls = [a6_data + str(i) +'.csv' for i in range(1,7)]

# read in only the columns with actual data
[sns.pairplot(data =pd.read_csv(url,usecols=lambda c: not('Unnamed' in c)),
hue='char') for url in data_urls]
```

```
[<seaborn.axisgrid.PairGrid at 0x7f5ff881f550>,
 <seaborn.axisgrid.PairGrid at 0x7f5ff84f5150>,
 <seaborn.axisgrid.PairGrid at 0x7f5fecbb35d0>,
 <seaborn.axisgrid.PairGrid at 0x7f5febbbfc90>,
 <seaborn.axisgrid.PairGrid at 0x7f5feba02410>,
 <seaborn.axisgrid.PairGrid at 0x7f5feabeb890>]
```

x0
x1
x2
x3
x4
char
A
B
C

x0
x1
char
a
c
b
d

Now we're looking at dataset 6 more deeply.

```
df6= pd.read_csv(data_urls[-1],usecols=[1,2,3])
df6.head()
```

	x0	x1	char
0	6.14	2.10	B
1	2.22	2.39	A
2	2.27	5.44	B
3	1.03	3.19	A
4	2.25	1.71	A

This one Naive Bayes does poorly on, even though the classes are separable because each class is not defined by a single region. Since each class has two noncontinuous resions.

```
dt = DecisionTreeClassifier()
X_train, X_test, y_train,  y_test = train_test_split(df6.values[:,:2],df6.values[:,2])
```

```
y_train
```

```
array(['B', 'B', 'B', 'A', 'B', 'A', 'B', 'A', 'A', 'B', 'B', 'A', 'A',
       'A', 'B', 'A', 'B', 'B', 'A', 'B', 'B', 'B', 'B', 'B', 'A', 'B',
       'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'A', 'A', 'B', 'A', 'B',
       'B', 'B', 'B', 'A', 'B', 'A', 'B', 'A', 'B', 'B', 'B', 'A', 'A',
       'B', 'A', 'B', 'A', 'B', 'B', 'B', 'B', 'B', 'A', 'A', 'A', 'A',
       'B', 'A', 'A', 'B', 'B', 'B', 'A', 'A', 'A', 'B', 'A', 'A', 'B',
       'A', 'B', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'B',
       'B', 'B', 'A', 'A', 'B', 'B', 'B', 'A', 'A', 'A', 'A', 'A', 'A',
       'A', 'B', 'A', 'B', 'A', 'A', 'B', 'B', 'A', 'B', 'B', 'A', 'A',
       'B', 'B', 'A', 'A', 'A', 'B', 'A', 'B', 'B', 'A', 'B', 'B', 'B',
       'A', 'B', 'B', 'B', 'B', 'B', 'A', 'A', 'A', 'A', 'A', 'A', 'A',
       'A', 'B', 'B', 'B', 'A', 'A', 'A'], dtype=object)
```

```
dt.fit(X_train,y_train)
```

```
DecisionTreeClassifier()
```

```
tree.plot_tree(dt)
```

```
[Text(96.00882352941177, 201.90857142857143, 'X[1] <= 1.58\ngini = 0.5\nsamples =
150\nvalue = [73, 77]'),
 Text(39.38823529411765, 170.84571428571428, 'X[0] <= 3.83\ngini = 0.32\nsamples =
15\nvalue = [3, 12]'),
 Text(19.694117647058825, 139.78285714285715, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(59.082352941176474, 139.78285714285715, 'gini = 0.0\nsamples = 12\nvalue = [0,
12]'),
 Text(152.62941176470588, 170.84571428571428, 'X[1] <= 1.94\ngini = 0.499\nsamples =
135\nvalue = [70, 65]'),
 Text(98.47058823529412, 139.78285714285715, 'X[0] <= 4.275\ngini = 0.337\nsamples =
14\nvalue = [11, 3]'),
 Text(78.7764705882353, 108.72, 'gini = 0.0\nsamples = 11\nvalue = [11, 0]'),
 Text(118.16470588235295, 108.72, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
 Text(206.78823529411767, 139.78285714285715, 'X[0] <= 1.86\ngini = 0.5\nsamples =
121\nvalue = [59, 62]'),
 Text(157.5529411764706, 108.72, 'X[1] <= 3.98\ngini = 0.417\nsamples = 27\nvalue =
[8, 19]'),
 Text(137.85882352941178, 77.65714285714284, 'gini = 0.0\nsamples = 8\nvalue = [8,
0]'),
 Text(177.24705882352941, 77.65714285714284, 'gini = 0.0\nsamples = 19\nvalue = [0,
19]'),
 Text(256.02352941176474, 108.72, 'X[1] <= 2.955\ngini = 0.496\nsamples = 94\nvalue =
[51, 43]'),
 Text(216.63529411764708, 77.65714285714284, 'X[0] <= 4.01\ngini = 0.464\nsamples =
41\nvalue = [15, 26]'),
 Text(196.94117647058823, 46.59428571428572, 'gini = 0.0\nsamples = 15\nvalue = [15,
0]'),
 Text(236.3294117647059, 46.59428571428572, 'gini = 0.0\nsamples = 26\nvalue = [0,
26]'),
 Text(295.4117647058824, 77.65714285714284, 'X[0] <= 4.095\ngini = 0.436\nsamples =
53\nvalue = [36, 17]'),
 Text(275.71764705882356, 46.59428571428572, 'X[1] <= 4.065\ngini = 0.188\nsamples =
19\nvalue = [2, 17]'),
 Text(256.02352941176474, 15.531428571428563, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(295.4117647058824, 15.531428571428563, 'gini = 0.0\nsamples = 17\nvalue = [0,
17]'),
 Text(315.1058823529412, 46.59428571428572, 'gini = 0.0\nsamples = 34\nvalue = [34,
0]')]
```

```
print(tree.export_text(dt))
```

```
|--- feature_1 <= 1.58
|   |--- feature_0 <= 3.83
|   |   |--- class: A
|   |--- feature_0 >  3.83
|   |   |--- class: B
|--- feature_1 >  1.58
|   |--- feature_1 <= 1.94
|   |   |--- feature_0 <= 4.27
|   |   |   |--- class: A
|   |   |--- feature_0 >  4.27
|   |   |   |--- class: B
|   |--- feature_1 >  1.94
|   |   |--- feature_0 <= 1.86
|   |   |   |--- feature_1 <= 3.98
|   |   |   |   |--- class: A
|   |   |   |--- feature_1 >  3.98
|   |   |   |   |--- class: B
|   |   |--- feature_0 >  1.86
|   |   |   |--- feature_1 <= 2.96
|   |   |   |   |--- feature_0 <= 4.01
|   |   |   |   |   |--- class: A
|   |   |   |   |--- feature_0 >  4.01
|   |   |   |   |   |--- class: B
|   |   |   |--- feature_1 >  2.96
|   |   |   |   |--- feature_0 <= 4.10
|   |   |   |   |   |--- feature_1 <= 4.07
|   |   |   |   |   |   |--- class: A
|   |   |   |   |   |--- feature_1 >  4.07
|   |   |   |   |   |   |--- class: B
|   |   |   |   |--- feature_0 >  4.10
|   |   |   |   |   |--- class: A
```

```
dt.score(X_test,y_test)
```

```
1.0
```

```
dt.get_depth()
```

```
6
```

```
dt2 = DecisionTreeClassifier(max_depth=2)
```

```
dt2.fit(X_train, y_train,)
```

```
DecisionTreeClassifier(max_depth=2)
```

```
dt2.score(X_test,y_test)
```

```
0.5
```

```
print(tree.export_text(dt2))
```

```
|--- feature_1 <= 1.58
|   |--- feature_0 <= 3.83
|   |   |--- class: A
|   |--- feature_0 >  3.83
|   |   |--- class: B
|--- feature_1 >  1.58
|   |--- feature_1 <= 1.94
|   |   |--- class: A
|   |--- feature_1 >  1.94
|   |   |--- class: B
```

```
dt2.score(X_train,y_train)
```

```
0.5866666666666667
```

Class 20: Decision Trees and Cross Validation

1. Share your favorite beverage (or say hi) in the zoom chat
2. log onto prismia
3. Accept assignment 7

Assignment 7

Make a plan with a group:

- what methods do you need to use in part 1?
- try to outline with psuedocode what you'll do for part 2 & 3

Share any questions you have.

Followup:

1. assignment clarified to require 3 values for the parameter in part 2
2. more tips on finding data sets added to assignment text

Complexity of Decision Trees

```
# %load http://drsmb.co/310
import pandas as pd
import seaborn as sns
import numpy as np
from sklearn import tree
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
d6_url = 'https://raw.githubusercontent.com/rhodyprog4ds/06-naive-
bayes/main/data/dataset6.csv'
```

```
df6= pd.read_csv(d6_url,usecols=[1,2,3])
df6.head()
```

	x0	x1	char
0	6.14	2.10	B
1	2.22	2.39	A
2	2.27	5.44	B
3	1.03	3.19	A
4	2.25	1.71	A

```
X_train, X_test, y_train,  y_test = train_test_split(df6.values[:,:2],df6.values[:,2],
                                                     train_size=.8)
```

Note

`df6.values` is a numpy array, which is a good datastructure for storing matrices of data. We can index into numpy arrays using `[rows, columns]`. Here, `df6.values[:,:2]` we take all the rows (`:`) and the columns up to, but not including index 2 for the features (X) `:2` and use columns at index 2 for the target(y).

```
dt = tree.DecisionTreeClassifier(min_samples_leaf = 10)
dt.fit(X_train,y_train)
```

```
DecisionTreeClassifier(min_samples_leaf=10)
```

```
print(tree.export_text(dt))
```

```
|--- feature_0 <= 5.88
|   |--- feature_1 <= 3.98
|   |   |--- feature_0 <= 4.07
|   |   |   |--- class: A
|   |   |--- feature_0 >  4.07
|   |   |   |--- class: B
|   |--- feature_1 >  3.98
|   |   |--- feature_0 <= 4.09
|   |   |   |--- class: B
|   |   |--- feature_0 >  4.09
|   |   |   |--- class: A
|--- feature_0 >  5.88
|   |--- feature_1 <= 3.89
|   |   |--- class: B
|   |--- feature_1 >  3.89
|   |   |--- class: A
```

```
dt2 = tree.DecisionTreeClassifier(min_samples_leaf = 50)
dt2.fit(X_train,y_train)
```

```
DecisionTreeClassifier(min_samples_leaf=50)
```

```
print(tree.export_text(dt2))
```

```
|--- feature_0 <= 5.88
|   |--- feature_1 <= 3.98
|   |   |--- class: A
|   |--- feature_1 >  3.98
|   |   |--- class: B
|--- feature_0 >  5.88
|   |--- class: B
```

```
dt2.score(X_test,y_test)
```

```
0.6
```

```
dt.score(X_test,y_test)
```

```
1.0
```

```
df6.shape
```

```
(200, 3)
```

Training, Test set size and Cross Validation

```
dt3 = tree.DecisionTreeClassifier()
dt3.fit(df6.values[:-1,:2],df6.values[:-1,2],)
```

```
DecisionTreeClassifier()
```

```
print(tree.export_text(dt3))
```

```
|--- feature_0 <= 5.88
|   |--- feature_1 <= 5.33
|   |   |--- feature_0 <= 4.07
|   |   |   |--- feature_1 <= 4.00
|   |   |   |   |--- class: A
|   |   |   |--- feature_1 >  4.00
|   |   |   |   |--- class: B
|   |   |--- feature_0 >  4.07
|   |   |   |--- feature_1 <= 3.91
|   |   |   |   |--- class: B
|   |   |   |--- feature_1 >  3.91
|   |   |   |   |--- class: A
|   |--- feature_1 >  5.33
|   |   |--- feature_0 <= 4.09
|   |   |   |--- class: B
|   |   |--- feature_0 >  4.09
|   |   |   |--- class: A
|--- feature_0 >  5.88
|   |--- feature_1 <= 3.89
|   |   |--- class: B
|   |--- feature_1 >  3.89
|   |   |--- class: A
```

```
dt4 = tree.DecisionTreeClassifier(max_depth=2)
cv_scores = cross_val_score(dt4,df6.values[:,:2],df6.values[:,2],cv=100 )
cv_scores
```

```
/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/sklearn/model_selection/_split.py:668: UserWarning: The least populated
class in y has only 99 members, which is less than n_splits=100.
  % (min_groups, self.n_splits)), UserWarning)
```

```
array([1. , 1. , 0.5, 0.5, 1. , 0.5, 1. , 0.5, 0.5, 1. , 1. , 0.5, 0.5,
       1. , 1. , 0.5, 1. , 1. , 0.5, 1. , 1. , 0.5, 0.5, 1. , 0. , 1. ,
       1. , 1. , 0.5, 1. , 0.5, 0.5, 0.5, 0.5, 1. , 0.5, 1. , 1. , 0.5,
       0.5, 1. , 0.5, 0.5, 0.5, 1. , 0.5, 0.5, 1. , 1. , 0.5, 1. , 1. ,
       1. , 1. , 1. , 0.5, 1. , 1. , 1. , 1. , 1. , 0. , 1. , 0.5, 0.5,
       1. , 0. , 1. , 0.5, 1. , 0.5, 0. , 1. , 1. , 1. , 1. , 0.5, 0.5,
       0.5, 1. , 1. , 1. , 1. , 0. , 1. , 1. , 1. , 1. , 1. , 0.5, 1. ,
       1. , 1. , 0.5, 1. , 1. , 1. , 0.5, 0. , 0.5])
```

```
np.mean(cv_scores)
```

```
0.755
```

Class 21: Regression

1. Log onto prismia
2. Share a topic you're most interested in applying data science to in the zoom chat

Regression Introduction

What is the difference in data that's well suited for regression vs classification

- [ ] regression is better for more features
- [ ] regression can work with categorical features
- [x] regression uses a continuous target variable

Explanation: the difference is that for regression we can use continuous target variables. Either classification or regression can work with high dimensional data (dimension refers to the number of features). Either can also work with a mixture of categorical or continuous valued features.

```
import numpy as np
import seaborn as sns
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score
import pandas as pd
sns.set_theme(font_scale=2)
```

Today we will work with the [diabetes dataset]() and we'll use the `sklearn.datasets` module to load it. When we load data this way, it gets loaded as a `numpy` array

```
diabetes_X, diabetes_y = datasets.load_diabetes(return_X_y = True)
```

Since it's not a `pandas.DataFrame` we don't have the head method, but we can index using square brackets.

```
diabetes_X[:5]
```

```
array([[ 0.03807591,  0.05068012,  0.06169621,  0.02187235, -0.0442235 ,
        -0.03482076, -0.04340085, -0.00259226,  0.01990842, -0.01764613],
       [-0.00188202, -0.04464164, -0.05147406, -0.02632783, -0.00844872,
        -0.01916334,  0.07441156, -0.03949338, -0.06832974, -0.09220405],
       [ 0.08529891,  0.05068012,  0.04445121, -0.00567061, -0.04559945,
        -0.03419447, -0.03235593, -0.00259226,  0.00286377, -0.02593034],
       [-0.08906294, -0.04464164, -0.01159501, -0.03665645,  0.01219057,
         0.02499059, -0.03603757,  0.03430886,  0.02269202, -0.00936191],
       [ 0.00538306, -0.04464164, -0.03638469,  0.02187235,  0.00393485,
         0.01559614,  0.00814208, -0.00259226, -0.03199144, -0.04664087]])
```

and the same for the labels.

```
diabetes_y[:5]
```

```
array([151.,  75., 141., 206., 135.])
```

Since this data is not in any sorted order, we can split into test and train using indexing instead of the `test_train_split` function. This way we all get the same split, without setting the `ranomd_seed` parameter. In most cases, it's best to use the function, but it's good to know different ways to do things.

We'll start by using only one feature, the one in column 8.

```
diabetes_X_train = diabetes_X[:-20,8]
diabetes_X_test = diabetes_X[-20:,8]
diabetes_y_train = diabetes_y[:-20]
diabetes_y_test = diabetes_y[-20:]
```

Now we can instantiate the object

```
regr = linear_model.LinearRegression()
```

All `sklearn` estimators have the same methods, and all take any specialized parameters in the constructor. We've used the default values here, but this is an important design feature of scikit learn, because it makes their pipeline infrastructure and functions for cross validation work.

```
regr.fit(diabetes_X_train,diabetes_y_train)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-7-7563ebc7151a> in <module>
----> 1 regr.fit(diabetes_X_train,diabetes_y_train)

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/sklearn/linear_model/_base.py in fit(self, X, y, sample_weight)
    517
    518         X, y = self._validate_data(X, y, accept_sparse=accept_sparse,
--> 519                                    y_numeric=True, multi_output=True)
    520
    521         if sample_weight is not None:

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-packages/sklearn/base.py in
_validate_data(self, X, y, reset, validate_separately, **check_params)
    431                 y = check_array(y, **check_y_params)
    432             else:
--> 433                 X, y = check_X_y(X, y, **check_params)
    434             out = X, y
    435

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/sklearn/utils/validation.py in inner_f(*args, **kwargs)
     61             extra_args = len(args) - len(all_args)
     62             if extra_args <= 0:
---> 63                 return f(*args, **kwargs)
     64
     65             # extra_args > 0

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/sklearn/utils/validation.py in check_X_y(X, y, accept_sparse,
accept_large_sparse, dtype, order, copy, force_all_finite, ensure_2d, allow_nd,
multi_output, ensure_min_samples, ensure_min_features, y_numeric, estimator)
    876                     ensure_min_samples=ensure_min_samples,
    877                     ensure_min_features=ensure_min_features,
--> 878                     estimator=estimator)
    879     if multi_output:
    880         y = check_array(y, accept_sparse='csr', force_all_finite=True,

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/sklearn/utils/validation.py in inner_f(*args, **kwargs)
     61             extra_args = len(args) - len(all_args)
     62             if extra_args <= 0:
---> 63                 return f(*args, **kwargs)
     64
     65             # extra_args > 0

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/sklearn/utils/validation.py in check_array(array, accept_sparse,
accept_large_sparse, dtype, order, copy, force_all_finite, ensure_2d, allow_nd,
ensure_min_samples, ensure_min_features, estimator)
    696                     "Reshape your data either using array.reshape(-1, 1) if "
    697                     "your data has a single feature or array.reshape(1, -1) "
--> 698                     "if it contains a single sample.".format(array))
    699
    700         # make sure we actually converted to numeric:
```

```
ValueError: Expected 2D array, got 1D array instead:
array=[ 0.01990842 -0.06832974  0.00286377  0.02269202 -0.03199144 -0.04118039
 -0.06291295 -0.03581673 -0.01495648  0.06773633 -0.06291295 -0.09643322
 -0.03075121  0.03839325 -0.03199144  0.03605579  0.05228     0.02736771
 -0.01811827 -0.00894402 -0.01190068 -0.07212845 -0.0611766   0.13359898
 -0.02595242  0.01919903 -0.0425721  -0.01599827 -0.00060925  0.0594238
 -0.02712865 -0.03712835  0.00027149 -0.01811827 -0.0594727   0.02131085
  0.054724    0.01703713  0.07142403  0.01919903  0.01255315 -0.04986847
 -0.00991896 -0.0425721   0.03243323 -0.01495648 -0.0079794  -0.01811827
 -0.0594727   0.03546194 -0.02139368 -0.00239668 -0.01811827 -0.03324879
 -0.04118039 -0.06648815  0.03365681 -0.05615757 -0.00060925  0.03839325
 -0.05140054  0.0366458  -0.08238148 -0.03075121 -0.01919705  0.0423449
  0.00027149 -0.01190068 -0.01290794 -0.05140054 -0.06291295  0.02671426
  0.08449528 -0.00514531  0.00371174  0.00620932 -0.03075121 -0.07408887
  0.04289569 -0.06832974  0.02539313 -0.00608025 -0.07020931 -0.01599827
 -0.06832974 -0.04118039 -0.09643322  0.04289569 -0.03452372 -0.02479119
 -0.07212845  0.03723201  0.01255315 -0.03324879 -0.04118039  0.0011438
  0.06078775  0.00286377 -0.00608025  0.0117839   0.03723201 -0.03075121
 -0.00514531 -0.01090444 -0.0611766  -0.01090444 -0.0594727  -0.05295879
  0.04560081 -0.02595242 -0.12609739 -0.02364456 -0.02364456  0.04666077
  0.09924023  0.08379677  0.02200405  0.07912108  0.02061233  0.02736771
 -0.03980959  0.01990842  0.03243323  0.01556684 -0.01599827  0.08058546
 -0.09393565  0.00455189 -0.03075121  0.03896837  0.02269202 -0.04687948
 -0.01290794 -0.07212845 -0.02712865  0.05078151 -0.06648815  0.01703713
  0.02801651  0.054724   -0.03324879  0.04506617  0.04560081 -0.04118039
  0.00620932 -0.01599827  0.01482271  0.02605609 -0.03075121  0.06123791
 -0.01919705 -0.00894402  0.02671426 -0.03452372  0.00864028  0.03781448
 -0.02139368  0.01022564 -0.03075121  0.0423449  -0.03581673  0.07573759
 -0.03075121 -0.02364456  0.01776348 -0.05615757 -0.07814091  0.07763279
  0.06898221  0.13359898 -0.09643322 -0.05140054  0.07341008 -0.01811827
  0.01703713 -0.02712865 -0.00330371  0.07222365 -0.04836172  0.01407245
  0.01630495 -0.02595242  0.01556684  0.01556684  0.03723201  0.06731722
  0.04718617 -0.08682899 -0.00149859 -0.02028875  0.02337484  0.01022564
 -0.02952762  0.03953988  0.00200784  0.00783714 -0.03324879  0.01331597
 -0.07408887  0.08989869 -0.02832024 -0.05295879  0.08449528  0.02539313
  0.00200784  0.06345592  0.00943641  0.04613233 -0.00149859 -0.01190068
 -0.03324879 -0.02251217 -0.00060925 -0.02251217 -0.03980959  0.0366458
  0.04560081  0.03365681  0.00702686  0.0011438  -0.10164355  0.0011438
  0.0011438  -0.01811827 -0.03581673  0.06301662 -0.03980959 -0.03845911
 -0.02712865 -0.02364456  0.05027649  0.00943641  0.02405258 -0.02595242
  0.06432823 -0.00514531  0.04067226 -0.04836172  0.0110081   0.06604821
  0.03365681 -0.02952762 -0.07814091 -0.0439854  -0.0439854  -0.04986847
  0.04506617 -0.01090444  0.06389312  0.0702113   0.07573759  0.08449528
 -0.00421986  0.03486419  0.12005338 -0.00330371  0.02801651 -0.00514531
 -0.03712835 -0.0439854  -0.08023654 -0.01290794  0.03486419 -0.07212845
 -0.0594727   0.01482271 -0.08682899  0.02337484  0.02930041 -0.01190068
  0.00864028 -0.01811827 -0.01811827  0.10635428  0.01556684  0.0110081
  0.05757286 -0.02139368  0.0011438  -0.01599827  0.01630495  0.02405258
 -0.00894402 -0.06832974  0.00286377  0.05988072 -0.01919705  0.03243323
 -0.07020931  0.06123791  0.02993565  0.05803913 -0.06291295 -0.05454415
 -0.07408887  0.01630495 -0.05615757 -0.06468302 -0.03075121 -0.02595242
  0.04289569 -0.02139368  0.03119299  0.07380215 -0.02251217  0.01776348
 -0.03980959  0.02337484 -0.0611766   0.02671426 -0.01811827  0.01255315
 -0.03845911  0.05520504 -0.01190068 -0.01090444  0.07496834  0.02472532
  0.00943641  0.02866072  0.03304707  0.09864637  0.13339573  0.12901941
  0.06604821  0.08449528  0.02993565 -0.00060925 -0.00149859 -0.0170521
 -0.02364456 -0.04118039  0.05710419  0.02930041 -0.06291295 -0.05140054
  0.08094791  0.00538437 -0.02251217  0.04344317 -0.03581673 -0.02479119
 -0.02364456 -0.05780007  0.00702686 -0.00060925  0.06257518  0.00371174
 -0.01090444 -0.0611766   0.0702113  -0.05780007 -0.02028875  0.13237265
  0.07419254 -0.03324879 -0.02832024  0.04976866 -0.08238148  0.01482271
  0.01331597 -0.07212845  0.00371174  0.01919903  0.01482271  0.03119299
  0.07102158  0.03365681  0.08553312 -0.00514531 -0.0611766   0.11934399
 -0.01090444 -0.0425721  -0.00060925  0.02801651  0.04613233  0.00943641
  0.02269202 -0.04687948 -0.05140054 -0.01290794  0.10413761 -0.01811827
  0.02405258 -0.02595242 -0.03075121 -0.10436482  0.02671426 -0.04118039
  0.10329226 -0.08913686 -0.04118039 -0.03980959  0.13008061  0.03181522
 -0.09643322  0.03605579  0.04506617  0.03953988 -0.03452372 -0.0425721
  0.01556684  0.06168585 -0.02712865  0.0366458  -0.05615757 -0.04836172
  0.05803913  0.03056649  0.06604821 -0.05140054  0.00620932 -0.03581673
  0.054724    0.01482271 -0.02952762 -0.04687948 -0.02139368 -0.07408887
 -0.03324879  0.04976866].
Reshape your data either using array.reshape(-1, 1) if your data has a single feature
or array.reshape(1, -1) if it contains a single sample.
```

we get an error because the training data (`diabetes_X_train`) is only one dimensional. Scikit learn's `fit` methods require a 2d-array even if fitting with only one feature.

```
diabetes_X_train.shape
```

```
(422,)
```

we can use `np.newaxis` to fix this problem.

```
diabetes_X_train = diabetes_X[:-20,np.newaxis,8]
diabetes_X_test = diabetes_X[-20:,np.newaxis,8]
```

and check the shape

```
diabetes_X_train.shape
```

```
(422, 1)
```

It's the same shape, in a mathematical sense, but in terms of the underlying data structure this one has a second dimension (it's just only 1 long) in that dimension. This is tricky, but an important thing to know how to fix.

Now we can fit our model

```
regr.fit(diabetes_X_train,diabetes_y_train)
```

```
LinearRegression()
```

and examine the coefficient

```
regr.coef_
```

```
array([900.39171612])
```

and make predictions with the model

```
diabetes_y_pred = regr.predict(diabetes_X_test)
```

we can evaluate with `mean_squared_error` (see below for more detail), which is as it's named. It computes the error in each prediction (`y_pred- y_true`), squares it, then averages them all together.

```
mean_squared_error(diabetes_y_test,diabetes_y_pred)
```

```
2923.342534244987
```

We can also use the \(r^2\) score, which is more normalized, it's best value is 1.

```
r2_score(diabetes_y_test,diabetes_y_pred)
```

```
0.3948984231023219
```

To get an even better idea, we can plot. We'll use `matplotlib` to plot, since our data is not in a DataFrame. `seaborn` and `pandas` use `matplotlib` under the hood, so it will have a lot of familiar methods, but it takes a lot more work to build complicated figures with `matplotlib`. The alias `plt` is used by convention for the `pyplot` module of `matplotlib`

```
import matplotlib.pyplot as plt
```

Then we can scatter plot the true data and make a line of the predictions in blue.

```
plt.scatter(diabetes_X_test,diabetes_y_test, color='black')
plt.plot(diabetes_X_test,diabetes_y_pred, color='blue')
```

```
[<matplotlib.lines.Line2D at 0x7f022e23db90>]
```

We can modify this plot to highlight the errors, too.

```
plt.scatter(diabetes_X_test, diabetes_y_test,  color='black')
plt.plot(diabetes_X_test, diabetes_y_pred, color='blue', linewidth=3)

[plt.plot([x,x],[yp,yt], color='red', linewidth=3)
                 for x, yp, yt in zip(diabetes_X_test,
diabetes_y_pred,diabetes_y_test)];
```

Tip

the semicolon here prevents it from printing out the axis information for each element of the list comprehension

Try it yourself

Fit another regression model to this data that uses all of the feautures instead of only one

```
diabetes_X_train2 = diabetes_X[:-20]
diabetes_X_test2 = diabetes_X[-20:]
diabetes_y_train2 = diabetes_y[:-20]
diabetes_y_test2 = diabetes_y[-20:]

regr2 = linear_model.LinearRegression()
regr2.fit(diabetes_X_train2,diabetes_y_train2)
diabetes_y_pred2 = regr2.predict(diabetes_X_test2)
r2_score(diabetes_y_test2,diabetes_y_pred2)
```

```
0.5850753022690574
```

This does better on \(r^2\), let's also check the mse.

```
mean_squared_error(diabetes_y_test2,diabetes_y_pred2)
```

```
2004.5676026898211
```

More practice

1. try out `regr2.score` what does scikite learn use for the score for `linear_model.LinearRegression` models?
2. repeat the above using `test_train_split`, can you make the test size 20 samples again?
3. try another single feature model using a different feature than we did above. Is your new feature a better or worse predictor? Make plots to compare and try to build intuition for the performance metrics geometrically.

Questions After class

What is considered a good r2 value?

This will depend on context in some sense. You can read about it in the `r2_score` docs and the corresponding [wikipedia page]

Here's a quick visual of what different r2 score values lookl like. I simulated data by ranomly picking 10 points, then made the "predicted" y values by picking a slope of 3 and computing `3*x`. Then I simulated various levels of noise, by sampling noise and multiplying the same noise vector by different scales and adding all of those to a data frame with the column name the r score for if that column of target values was the truth.

Then I added some columns of y values that were with different slopes and different functions of x. These all have the small amount of noise.

I used the pandas `melt` method to restructure the DataFrame so that I could use `FaceGrid` with `col` and `col_wrap` to show all of the results.

```
x = 10*np.random.random(20)
y_pred = 3*x
ex_df = pd.DataFrame(data = x,columns = ['x'])
ex_df['y_pred'] = y_pred
n_levels = range(1,18,2)
noise = (np.random.random(20)-.5)*2
for n in n_levels:
    y_true = y_pred + n* noise
    ex_df['r2 = '+ str(np.round(r2_score(y_pred,y_true),3))] = y_true

f_x_list = [2*x,3.5*x,.5*x**2, .03*x**3, 10*np.sin(x)+x*3,3*np.log(x**2)]
for fx in f_x_list:
    y_true = fx + noise
    ex_df['r2 = '+ str(np.round(r2_score(y_pred,y_true),3))] = y_true

xy_df = ex_df.melt(id_vars=['x','y_pred'],var_name='rscore',value_name='y')
# sns.lmplot(x='x',y='y', data = xy_df,col='rscore',col_wrap=3,)
g = sns.FacetGrid(data = xy_df,col='rscore',col_wrap=3,aspect=1.5,height=3)
g.map(plt.plot, 'x','y_pred',color='k')
g.map(sns.scatterplot, "x", "y",)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f022c1808d0>
```

The `r2_score` alone doesn't tell the whole picture. We'll also often want to analyze the errors in greater detail. We'll look at how to do that to investigate how the model really fits more later this week.

Can some of the methods applied in regression can be used in the same format in classification?

the `sklearn` estimator objects all have the same methods, you can see details on the API on the [sklearn developer page](#).

Why couldn't we use `train_test_split` for this model?

We absolutely can. Splitting this way was to show *another* way to split the data, and point out under which conditions this way is ok.

I am still confused about machine learning

Try reading the notes from the [introduction class](#).

The scikit-learn [choose an estimator](#) graph can also be helpful.

What is the difference between test and train data?

We split data, sample-wise, into two pieces, in order to evaluate what we are doing.

There is nothing inherently different between test data and train data. In fact, by applying the model we learned with the training data to predict on the test data we are assuming that they are very similar.

We split our data so that we can see how well we expect our trained model to work on new data, since the goal is to send our learned model into the world and use it for something else. For example, a model fit on today's data, about diabetes, might be used to predict which patients are going to have the most severe disease in a year based on their stats today and then enroll them in a program to help them learn to better manage their diabetes.

The test set is meant to simulate future data that we've never seen before, but it's usually data *we* have seen, but our learning algorithm did not use to fit the model.

What does cross validation mean?

Cross validation will come up again later, but basically cross validation is a more elaborate evaluation. It repeatedly splits the data into test and train sets, fits the model, scores the model, and then returns all fo the results.

What does `mean_squared_error` do?

It is a popular performance metric for regression. To understand it, let's look back at what regression is doing. Regression is tyring to predict a continuous valued quantity, so unlike in classification where we can just check if it matches or not, to compute performance, it makes sense consider the *size* of the erorr, beacuse it's unlikely to get *exactly* the right value for any prediction.

So, we might be tempted to compute error like this:

```
diabetes_y_test2-diabetes_y_pred2
```

```
array([ 35.38153092, -64.43979328, -61.88665147,  40.46462721,
       -44.80054784, -64.06954875,  50.87762239,  -6.47935157,
        65.9398948 , -58.30503555, -45.36632793,  10.80168716,
       -83.25046751, -56.3332925 ,  -4.54458691, -16.03798088,
         1.42860298,   8.43395013,   8.9653683 ,   4.39664326])
```

this, however is a whole vector of quantities, so we would need to take the average. Taking the average here though, will be very small, because some are positive and some are negative and they will cancel one another out.

```
np.mean(diabetes_y_test2-diabetes_y_pred2)
```

```
-13.941182850978748
```

In this case, it turns out to even be negative.

Instead, we can square each error, to make them all postive. We *could* also take the absolute value, but in a lot of calculations, we want to take the derivative, so suqaring instead of absolute value became more popular.

```
(diabetes_y_test2-diabetes_y_pred2)**2
```

```
array([1.25185273e+03, 4.15248696e+03, 3.82995763e+03, 1.63738606e+03,
       2.00708909e+03, 4.10490708e+03, 2.58853246e+03, 4.19819967e+01,
       4.34806973e+03, 3.39947717e+03, 2.05810371e+03, 1.16676446e+02,
       6.93064034e+03, 3.17343984e+03, 2.06532702e+01, 2.57216831e+02,
       2.04090647e+00, 7.11315149e+01, 8.03778287e+01, 1.93304719e+01])
```

now we can take the mean of these to get a total average error.

```
mse_manual = np.mean((diabetes_y_test2-diabetes_y_pred2)**2)
mse_manual
```

```
2004.5676026898211
```

we can confirm this gets the same answer as the sklearn function

```
mse_sk = mean_squared_error(diabetes_y_test2,diabetes_y_pred2)
mse_sk
```

```
2004.5676026898211
```

We can even assert that they're the same, they're always going to be exactly the same

```
assert mse_manual == mse_sk
```

Class 22: More Regression, More Evaluation and LASSO

1. log prismia chat
2. say hello in the zoom chat

```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split

# Load the diabetes dataset
diabetes_X, diabetes_y = datasets.load_diabetes(return_X_y=True)
```

Questions after class Monday

Some good questions were asked on the form Monday. Check the notes for insight on the `r2_score` and `mean_square_error`.

Review of Test Train splits

```
X_train,X_test, y_train,y_test = train_test_split(diabetes_X, diabetes_y ,
                                                  test_size=20,random_state=0)
```

What metric does the `score` method of `LinearRegression` use?

```
regr = linear_model.LinearRegression()
regr.fit(X_train, y_train)
```

```
LinearRegression()
```

```
regr.score(X_test,y_test)
```

```
0.5195208400616668
```

```
y_pred = regr.predict(X_test)
r2_score(y_test,y_pred)
```

```
0.5195208400616668
```

```
mean_squared_error(y_test,y_pred)
```

```
2850.3176917525775
```

Digging in deeper to the Linear Regression model

Linear regression fitting involvles learning the coefficients

```
regr.coef_
```

```
array([ -32.3074285 , -257.44432972,  513.31945939,  338.46656647,
       -766.86983748,  455.85416891,   92.55795582,  184.75163454,
        734.92318647,   82.7231425 ])
```

and an intercept

```
regr.intercept_
```

```
152.39189054201842
```

The linear regression model is

\[y = wx + b\]
wehere it stores \(w\) in the `coef_` attribute and `b` as the `intercept_`

We can check how this works by multiplying one sample by the coefficients, to get a vector

```
X_test[0]*regr.coef_
```

```
array([ -0.64334474, -13.0473092 ,  53.80033982,  23.71721361,
        27.58260658, -12.16168908,  -2.31326921,  -0.47892464,
         2.72784249,   3.33733048])
```

then taking the sum and adding the intercept

```
np.sum(X_test[0]*regr.coef_) + regr.intercept_
```

```
234.9126866563482
```

and then comparing to the predictions.

```
y_pred[0]
```

```
234.91268665634823
```

These are not exactly the same due to float rounding erros, but they're very close.

We can also check for the whole test set

```
errors = np.sum(X_test*regr.coef_,axis=1) + regr.intercept_ -y_pred
errors
```

```
array([-2.84217094e-14,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        1.42108547e-14,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00, -1.42108547e-14, -2.84217094e-14,  0.00000000e+00])
```

Tip

by default `np.sum` sums across both axes so it gets a single value. we want to only sum on 1 so that we get t

and confirm these are very small

```
np.max(errors)
```

```
1.4210854715202004e-14
```

It's doing linear regression as we expected.

Changing the complexity in Linear regression

It can be important to also know what the model is really doing and see how different features are used or not. Maybe, for example all of the features are expensive to measure, but for testing we measured a lot of them. We might want to simultaneously learn which features we actually need *and* the linear model.

LASSO can do that for us it's objective is like linear regression, but it adds an extra term. This term forces some of the learned coefficients to be 0. Mulitplying by 0 gives 0, so that's like throwing away that feature. The term is called the \(1-norm\), the details of the math are not important here, just the idea that it can reduce the number of features used as well.

\[||y-wx||_2^2 + \alpha ||w||_1 \]

```
lasso = linear_model.Lasso()
lasso.fit(X_train, y_train)
lasso.score(X_test,y_test)
```

```
0.4224899448032938
```

It uses many fewer

```
lasso.coef_
```

```
array([  0.        ,  -0.        , 358.27498703,   9.7141813 ,
         0.        ,   0.        ,  -0.        ,   0.        ,
       309.50796119,   0.        ])
```

It has a parameter alpha that must be >0 that we can use to control how many features it uses. When `alpha = 0` it's the same as linear regression, but the regular linear regression estimator uses a more numerically stable algorithm for the `fit` method

```
lasso2 = linear_model.Lasso(alpha=.5)
lasso2.fit(X_train, y_train)
lasso2.score(X_test,y_test)
```

```
0.5272242529276977
```

We see tha fewer are 0 now.

```
lasso2.coef_
```

```
array([  0.        ,  -0.        , 466.09039819, 140.20195776,
        -0.        ,  -0.        , -61.96474668,   0.        ,
       405.95829094,   0.        ])
```

```
regr.score(X_train, y_train)
```

```
0.5170933599105026
```

```
lasso2.score(X_train, y_train)
```

```
0.4516267981295532
```

```
lasso_cv = cross_val_score(lasso2, diabetes_X, diabetes_y)
lasso_cv
```

```
array([0.3712605 , 0.45675648, 0.45066569, 0.42134051, 0.47736463])
```

```
regr_cv = cross_val_score(regr,diabetes_X, diabetes_y )
regr_cv
```

```
array([0.42955643, 0.52259828, 0.4826784 , 0.42650827, 0.55024923])
```

```
np.mean(lasso_cv), np.mean(regr_cv)
```

```
(0.43547756049731523, 0.48231812211149394)
```

```
cross_val_score(regr,diabetes_X, diabetes_y ,cv=10)
```

```
array([0.55614411, 0.23056092, 0.35357777, 0.62190498, 0.26587602,
       0.61819338, 0.41815916, 0.43515232, 0.43436983, 0.68568514])
```

We can do cross validation for all three of these models and look at both the mean and the standard deviation. The mean tells us how well on average each model does on different subsets of the data. The standard deviation is a measure of *spread* of the scores. For intuition, another measure of spread is `max(cv_scores) - min(cv_scores)`.

```
regr_cv = cross_val_score(regr,diabetes_X, diabetes_y ,cv=10)
np.mean(regr_cv), np.std(regr_cv)
```

```
(0.461962361958337, 0.14698789185375888)
```

```
lasso_cv = cross_val_score(lasso,diabetes_X, diabetes_y ,cv=10)
np.mean(lasso_cv), np.std(lasso_cv)
```

```
(0.3211351084864853, 0.09122225780232662)
```

```
lasso2_cv = cross_val_score(lasso2,diabetes_X, diabetes_y ,cv=10)
np.mean(lasso2_cv), np.std(lasso2_cv)
```

```
(0.41808972445133297, 0.11927596370400496)
```

Tip

To keep things concise, we can put two values on one line. This creates a `tuple` and prints out both values since it's on the last line of the cell.

Questions after class

What does the `cv` parameter in `cross_val_score` do?

First, let's look at the help.

```
help(cross_val_score)
```

```
Help on function cross_val_score in module sklearn.model_selection._validation:

cross_val_score(estimator, X, y=None, *, groups=None, scoring=None, cv=None,
n_jobs=None, verbose=0, fit_params=None, pre_dispatch='2*n_jobs', error_score=nan)
    Evaluate a score by cross-validation

    Read more in the :ref:`User Guide <cross_validation>`.

    Parameters
    ----------
    estimator : estimator object implementing 'fit'
        The object to use to fit the data.

    X : array-like of shape (n_samples, n_features)
        The data to fit. Can be for example a list, or an array.

    y : array-like of shape (n_samples,) or (n_samples, n_outputs),
default=None
        The target variable to try to predict in the case of
        supervised learning.

    groups : array-like of shape (n_samples,), default=None
        Group labels for the samples used while splitting the dataset into
        train/test set. Only used in conjunction with a "Group" :term:`cv`
        instance (e.g., :class:`GroupKFold`).

    scoring : str or callable, default=None
        A str (see model evaluation documentation) or
        a scorer callable object / function with signature
        ``scorer(estimator, X, y)`` which should return only
        a single value.

        Similar to :func:`cross_validate`
        but only a single metric is permitted.

        If None, the estimator's default scorer (if available) is used.

    cv : int, cross-validation generator or an iterable, default=None
        Determines the cross-validation splitting strategy.
        Possible inputs for cv are:

        - None, to use the default 5-fold cross validation,
        - int, to specify the number of folds in a `(Stratified)KFold`,
        - :term:`CV splitter`,
        - An iterable yielding (train, test) splits as arrays of indices.
```

```
    For int/None inputs, if the estimator is a classifier and ``y`` is
    either binary or multiclass, :class:`StratifiedKFold` is used. In all
    other cases, :class:`KFold` is used. These splitters are instantiated
    with `shuffle=False` so the splits will be the same across calls.

    Refer :ref:`User Guide <cross_validation>` for the various
    cross-validation strategies that can be used here.

    .. versionchanged:: 0.22
        ``cv`` default value if None changed from 3-fold to 5-fold.

n_jobs : int, default=None
    Number of jobs to run in parallel. Training the estimator and computing
    the score are parallelized over the cross-validation splits.
    ``None`` means 1 unless in a :obj:`joblib.parallel_backend` context.
    ``-1`` means using all processors. See :term:`Glossary <n_jobs>`
    for more details.

verbose : int, default=0
    The verbosity level.

fit_params : dict, default=None
    Parameters to pass to the fit method of the estimator.

pre_dispatch : int or str, default='2*n_jobs'
    Controls the number of jobs that get dispatched during parallel
    execution. Reducing this number can be useful to avoid an
    explosion of memory consumption when more jobs get dispatched
    than CPUs can process. This parameter can be:

        - None, in which case all the jobs are immediately
          created and spawned. Use this for lightweight and
          fast-running jobs, to avoid delays due to on-demand
          spawning of the jobs

        - An int, giving the exact number of total jobs that are
          spawned

        - A str, giving an expression as a function of n_jobs,
          as in '2*n_jobs'

error_score : 'raise' or numeric, default=np.nan
    Value to assign to the score if an error occurs in estimator fitting.
    If set to 'raise', the error is raised.
    If a numeric value is given, FitFailedWarning is raised.

    .. versionadded:: 0.20

Returns
-------
scores : ndarray of float of shape=(len(list(cv)),)
    Array of scores of the estimator for each run of the cross validation.

Examples
--------
>>> from sklearn import datasets, linear_model
>>> from sklearn.model_selection import cross_val_score
>>> diabetes = datasets.load_diabetes()
>>> X = diabetes.data[:150]
>>> y = diabetes.target[:150]
>>> lasso = linear_model.Lasso()
>>> print(cross_val_score(lasso, X, y, cv=3))
[0.33150734 0.08022311 0.03531764]

See Also
---------
cross_validate : To run cross-validation on multiple metrics and also to
    return train scores, fit times and score times.

cross_val_predict : Get predictions from each split of cross-validation for
    diagnostic purposes.

sklearn.metrics.make_scorer : Make a scorer from a performance metric or
    loss function.
```

It does different things, depending on what type of value we pass it. We passed it an `int` (`10`), so what it did was split the data into 10 groups and then trains on 9/10 of those parts and tests on the last one. Then it iterates over those folds.

For classification, it can take into consideration the target value (classes) and an additionally specified group in the data. A good visualization of what it does is shown in the sklearn docs.

It uses StratifiedKfold for classification, but since we're using regression it will use `KFold`. `test_train_split` uses `ShuffleSplit` by default, let's load that too to see what it does.

> **Warning**
>
> The key in the following is to get the *concepts* not all of the details in how I evaluate and visualize. I could have made figures separately to explain the concept, but I like to show that Python is self contained.

```
from sklearn.model_selection import KFold, ShuffleSplit
```

```
kf = KFold(n_splits = 10)
```

When we use the `split` method it gives us a generator.

```
kf.split(diabetes_X, diabetes_y)
```

```
<generator object _BaseKFold.split at 0x7fd98e8c5550>
```

We can use this in a loop to get the list of indices that will be used to get the test and train data for each fold. To visualize what this is doing, see below.

```
N_samples = len(diabetes_y)
kf_tt_df = pd.DataFrame(index=list(range(N_samples)))
i = 1
for train_idx, test_idx in kf.split(diabetes_X, diabetes_y):
    kf_tt_df['split ' + str(i)] = ['unused']*N_samples
    kf_tt_df['split ' + str(i)][train_idx] = 'Train'
    kf_tt_df['split ' + str(i)][test_idx] = 'Test'
    i +=1
```

How would you use those indices to get a out actual test and train data?

We can count how many times ‘Test’ and ‘Train’ appear

```
count_test = lambda part: len([v for v in part if v=='Test'])
count_train = lambda part: len([v for v in part if v=='Train'])
```

When we apply this along `axis=1` we to check that each sample is used in exactly 1 test set how may times each sample is used

```
sum(kf_tt_df.apply(count_test,axis = 1) ==1)
```

```
442
```

and exactly 9 training sets

```
sum(kf_tt_df.apply(count_test,axis = 1) ==9)
```

```
0
```

the describe helps ensure that all fo the values are exa

We can also visualize:

```
cmap = sns.color_palette("tab10",10)
g = sns.heatmap(kf_tt_df.replace({'Test':1,'Train':0}),cmap=cmap[7:9],cbar_kws=
{'ticks':[.25,.75]},linewidths=0,
    linecolor='gray')
colorbar = g.collections[0].colorbar
colorbar.set_ticklabels(['Train','Test'])
```

> **Tip**
>
> `sns.heatmap` doesn’t work on strings, so we can replace them for the plotting

Note that unlike `test_train_split` this does not always randomize and shuffle the data before splitting.

If we apply those `lambda` functions along `axis=0`, we can see the size of each test set

```
kf_tt_df.apply(count_test,axis = 0)
```

```
split 1     45
split 2     45
split 3     44
split 4     44
split 5     44
split 6     44
split 7     44
split 8     44
split 9     44
split 10    44
dtype: int64
```

and training set:

```
kf_tt_df.apply(count_train,axis = 0)
```

```
split 1     397
split 2     397
split 3     398
split 4     398
split 5     398
split 6     398
split 7     398
split 8     398
split 9     398
split 10    398
dtype: int64
```

We can verify that these splits are the same size as what `test_train_split` does using the right settings. 10-fold splits the data into 10 parts and tests on 1, so that makes a test size of 1/10=.1, so we can use the `train_test_split` and check the length.

```
X_train2,X_test2, y_train2,y_test2 = train_test_split(diabetes_X, diabetes_y ,
                                                      test_size=.1,random_state=0)

[len(split) for split in [X_train2,X_test2,]]
```

Under the hood `train_test_split` uses `ShuffleSplit` We can do a similar experiment as above to see what `ShuffleSplit` does.

```
skf = ShuffleSplit(10)
N_samples = len(diabetes_y)
ss_tt_df = pd.DataFrame(index=list(range(N_samples)))
i = 1
for train_idx, test_idx in skf.split(diabetes_X, diabetes_y):
    ss_tt_df['split ' + str(i)] = ['unused']*N_samples
    ss_tt_df['split ' + str(i)][train_idx] = 'Train'
    ss_tt_df['split ' + str(i)][test_idx] = 'Test'
    i +=1

ss_tt_df
```

	split 1	split 2	split 3	split 4	split 5	split 6	split 7	split 8	split 9	split 10
0	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
1	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
2	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
3	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
4	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
...	...	...	...	...	...	...	...	...	...	...
437	Train	Train	Train	Train	Test	Train	Train	Train	Test	Train
438	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
439	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
440	Train	Train	Train	Train	Train	Train	Train	Train	Train	Train
441	Train	Train	Train	Train	Test	Train	Train	Train	Train	Test

442 rows × 10 columns

And plot

```
cmap = sns.color_palette("tab10",10)
g = sns.heatmap(ss_tt_df.replace({'Test':1,'Train':0}),cmap=cmap[7:9],cbar_kws=
{'ticks':[.25,.75]},linewidths=0,
    linecolor='gray')
colorbar = g.collections[0].colorbar
colorbar.set_ticklabels(['Train','Test'])
```

Now, we see the samples in each training set (gray along the columns) are a random subset of all of all of the samples.

And check the usage of each sample

```
sum(ss_tt_df.apply(count_test,axis = 1) ==1)
```

```
172
```

and exactly 9 training sets

```
sum(ss_tt_df.apply(count_test,axis = 1) ==9)
```

```
0
```

And the size of the splits

```
ss_tt_df.apply(count_test,axis = 0)
```

```
split 1     45
split 2     45
split 3     45
split 4     45
split 5     45
split 6     45
split 7     45
split 8     45
split 9     45
split 10    45
dtype: int64
```

and training set:

```
ss_tt_df.apply(count_train,axis = 0)
```

```
split 1     397
split 2     397
split 3     397
split 4     397
split 5     397
split 6     397
split 7     397
split 8     397
split 9     397
split 10    397
dtype: int64
```

Again the same sizes

Class 23: Interpretting Regression Evaluations

1. Snow in October. Share your thoughts in the zoom chat
2. Log onto prismia

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
```

```
X, y = datasets.load_boston(return_X_y= True)
```

```
X[:5]
```

```
array([[6.3200e-03, 1.8000e+01, 2.3100e+00, 0.0000e+00, 5.3800e-01,
        6.5750e+00, 6.5200e+01, 4.0900e+00, 1.0000e+00, 2.9600e+02,
        1.5300e+01, 3.9690e+02, 4.9800e+00],
       [2.7310e-02, 0.0000e+00, 7.0700e+00, 0.0000e+00, 4.6900e-01,
        6.4210e+00, 7.8900e+01, 4.9671e+00, 2.0000e+00, 2.4200e+02,
        1.7800e+01, 3.9690e+02, 9.1400e+00],
       [2.7290e-02, 0.0000e+00, 7.0700e+00, 0.0000e+00, 4.6900e-01,
        7.1850e+00, 6.1100e+01, 4.9671e+00, 2.0000e+00, 2.4200e+02,
        1.7800e+01, 3.9283e+02, 4.0300e+00],
       [3.2370e-02, 0.0000e+00, 2.1800e+00, 0.0000e+00, 4.5800e-01,
        6.9980e+00, 4.5800e+01, 6.0622e+00, 3.0000e+00, 2.2200e+02,
        1.8700e+01, 3.9463e+02, 2.9400e+00],
       [6.9050e-02, 0.0000e+00, 2.1800e+00, 0.0000e+00, 4.5800e-01,
        7.1470e+00, 5.4200e+01, 6.0622e+00, 3.0000e+00, 2.2200e+02,
        1.8700e+01, 3.9690e+02, 5.3300e+00]])
```

```
y[:5]
```

```
array([24. , 21.6, 34.7, 33.4, 36.2])
```

```
X_train, X_test, y_train, y_teest = train_test_split(X,y)
```

```
regr = linear_model.LinearRegression()
regr.fit(X_train, y_train)
```

```
LinearRegression()
```

```
regr.score(X_test,y_teest)
```

```
0.7076649858990918
```

```
y_pred = regr.predict(X_test)
```

```
# %load http://drsmb.co/310
col_name = ['ZN ',
'INDUS ',
'CHAS ',
'NOX ',
'RM ',
'AGE ',
'DIS ',
'RAD ',
'TAX ',
'PTRATIO ',
'B ',
'LSTAT ',
'MEDV ']
```

```
test_df = pd.DataFrame(data=X_test, columns = col_name)
test_df.head()
```

	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LS
0	0.03871	52.5	5.32	0.0	0.405	6.209	31.3	7.3172	6.0	293.0	16.6	39
1	73.53410	0.0	18.10	0.0	0.679	5.957	100.0	1.8026	24.0	666.0	20.2	1
2	0.04301	80.0	1.91	0.0	0.413	5.663	21.9	10.5857	4.0	334.0	22.0	38
3	15.28800	0.0	18.10	0.0	0.671	6.649	93.3	1.3449	24.0	666.0	20.2	36
4	5.87205	0.0	18.10	0.0	0.693	6.405	96.0	1.6768	24.0	666.0	20.2	39

```
test_df['y_test'] = y_teest
test_df['y_pred'] = y_pred
```

```
error = lambda r: r['y_test'] - r['y_pred']
test_df['pred_error'] = test_df.apply(error,axis=1)
```

```
test_df['errror'] = y_teest - y_pred
```

```
test_df.head()
```

	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LS
0	0.03871	52.5	5.32	0.0	0.405	6.209	31.3	7.3172	6.0	293.0	16.6	39
1	73.53410	0.0	18.10	0.0	0.679	5.957	100.0	1.8026	24.0	666.0	20.2	1
2	0.04301	80.0	1.91	0.0	0.413	5.663	21.9	10.5857	4.0	334.0	22.0	38
3	15.28800	0.0	18.10	0.0	0.671	6.649	93.3	1.3449	24.0	666.0	20.2	36
4	5.87205	0.0	18.10	0.0	0.693	6.405	96.0	1.6768	24.0	666.0	20.2	39

```
sns.displot(data=test_df, x='errror', kind = 'kde')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f13cafe5090>
```

Class 24: Clustering

1. Log onto prismia
2. Say hello on Zoom

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from sklearn import datasets
from sklearn.cluster import KMeans
import string
import pandas as pd
```

How Does Kmeans work?

We'll start with the data from dataset5 in assignment 6.

```
C = 4
N = 200
classes = list(string.ascii_uppercase[:C])
mu = {c: i for c, i in zip(classes,[[2,2, 2,2], [6,6,6,6], [2,6,6,2],[6,2,2,6]])}
sigma = {c: i*.5 for c, i in zip(classes,np.random.random(4))}
sigma

target5 = np.random.choice(classes,N)
data5 = [np.random.multivariate_normal(mu[c],.25*np.eye(C)) for c in target5]
df5 = pd.DataFrame(data = data5,columns = ['x' + str(i) for i in range(C)]).round(2)
rand_target = np.random.choice(classes,N)
obs_target = [np.random.choice([t,r],p=[.85,.15]) for t,r in zip(target5,rand_target)]
df5['char_noisy'] = obs_target
df5['char'] = target5

sns.pairplot(data =df5, hue='char')
```

```
<seaborn.axisgrid.PairGrid at 0x7fd973659d50>
```

We'll work with just two columns and the true labels

```
data_cols = ['x0','x1']
df = df5[data_cols]
df.head()
```

	x0	x1
0	4.87	5.68
1	6.04	6.02
2	1.85	6.36
3	6.45	6.14
4	1.35	2.17

We'll also setup some things for making our plots how we'd like in advance

```
def mu_to_df(mu,i):
    mu_df = pd.DataFrame(mu,columns=data_cols)
    mu_df['iteration'] = str(i)
    mu_df['class'] = ['M'+str(i) for i in range(K)]
    mu_df['type'] = 'mu'
    return mu_df

cmap_pt = sns.color_palette('tab20',8)[1::2]
cmap_mu = sns.color_palette('tab20',8)[0::2]
```

Now, we'll start our clustering. First we set how many clusters we ant and then we'll try out randomly assigning the samples and see where that puts the means

```
K = 4

df['0'] = np.random.choice(K,size=N)
sfig = sns.scatterplot(data =df,x='x0',y='x1',hue='0',palette=cmap_pt,legend=False)

mu = df.groupby('0')[data_cols].mean().values
mu_df = mu_to_df(mu,0)

# sns.scatterplot(data
=mu_df,x='x0',y='x1',hue='class',palette=cmap_mu,ax=sfig,legend=False)
sfig.get_figure().savefig('kmeans00.png')
#     sfig = sns.scatterplot(data =df,x='x0',y='x1',hue='1'
sns.scatterplot(data =mu_df,x='x0',y='x1',hue='class',palette=cmap_mu,ax=sfig)
```

```
/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  This is separate from the ipykernel package so we can avoid doing imports until
```

```
<AxesSubplot:xlabel='x0', ylabel='x1'>
```

This puts the means all really close to the center, so we'll try a different initialization for those: randomly sampling 4 points

```
mu = df[data_cols].sample(n=K).values
mu
```

```
array([[6.5 , 2.57],
       [7.42, 6.08],
       [2.06, 1.61],
       [6.04, 4.75]])
```

Now we can compute the distance from each point to each of the four means and use the closest one as an assignment. We'll look at the assignments again.

```
i = 1
df[str(i)] = pd.concat([((df[data_cols]-mu_i)**2).sum(axis=1) for mu_i in
mu],axis=1).idxmin(axis=1)


sfig = sns.scatterplot(data =df,x='x0',y='x1',hue='1',palette=cmap_pt,legend=False)
# plt.plot(mu[:,0],mu[:,1],marker='s',linewidth=0)
mu_df = mu_to_df(mu,i)
sns.scatterplot(data
=mu_df,x='x0',y='x1',hue='class',palette=cmap_mu,ax=sfig,legend=False)
sfig.get_figure().savefig('kmeans01.png')
```

```
/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

Now, we cna use those assignments, to compute new means:

```
mu = df.groupby('1')[data_cols].mean().values
# sfig = sns.scatterplot(data =df,x='x0',y='x1',hue='1')
# plt.plot(mu[:,0],mu[:,1],marker='s',linewidth=0)
fig = plt.figure()
mu_df = mu_to_df(mu,i)
sfig = sns.scatterplot(data =df,x='x0',y='x1',hue='1',palette=cmap_pt,legend=False)
sns.scatterplot(data
=mu_df,x='x0',y='x1',hue='class',palette=cmap_mu,ax=sfig,legend=False)
sfig.get_figure().savefig('kmeans02.png')
```

Now we can set up a while loop to continue the iterations

```
mu_list = [mu_to_df(mu,i)]
cur_old = str(i-1)
cur_new = str(i)
while sum(df[cur_old] !=df[cur_new]) >0:
    cur_old = cur_new
    i +=1
    cur_new = str(i)
    #     update the assigments and plot with the associated means
    df[cur_new] = pd.concat([((df[data_cols]-mu_i)**2).sum(axis=1) for mu_i in
mu],axis=1).idxmin(axis=1)
    fig = plt.figure()
    sfig = sns.scatterplot(data
=df,x='x0',y='x1',hue=cur_new,palette=cmap_pt,legend=False)
    sns.scatterplot(data
=mu_df,x='x0',y='x1',hue='class',palette=cmap_mu,ax=sfig,legend=False)
    file_num = str(i*2 -1).zfill(2)
    sfig.get_figure().savefig('kmeans' +file_num + '.png')


    #     update the means and plot with current generating assignments
    mu = df.groupby(cur_new)[data_cols].mean().values
    mu_df = mu_to_df(mu,i)
    mu_list.append(mu_df)

    fig = plt.figure()
    sfig = sns.scatterplot(data
=df,x='x0',y='x1',hue=cur_new,palette=cmap_pt,legend=False)
    sns.scatterplot(data
=mu_df,x='x0',y='x1',hue='class',palette=cmap_mu,ax=sfig,legend=False)
#     plt.plot(mu[:,0],mu[:,1],marker='s',linewidth=0)

    file_num = str(i*2).zfill(2)
    sfig.get_figure().savefig('kmeans' +file_num + '.png')

n_iter = i
```

```
/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/ipykernel_launcher.py:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  if __name__ == '__main__':
```

These plots can be saved and merged into a gif with, for example [imagemagick]() to create the gifs above. Since we've saved them in a single dataframe, we can also look at that

```
df.head()
```

	x0	x1	0	1	2	3	4
0	4.87	5.68	1	3	3	1	1
1	6.04	6.02	0	3	1	1	1
2	1.85	6.36	1	3	3	3	3
3	6.45	6.14	2	1	1	1	1
4	1.35	2.17	2	2	2	2	2

We can also manipuate the dataframe to make plotting them all together easier.

```
df_vis = df.melt(id_vars = ['x0','x1'], var_name ='iteration',value_name='class')
df_vis.replace({'class':{i:c for i,c in
enumerate(string.ascii_uppercase[:C])}},inplace=True)

df_vis['type'] = 'pt'
df_mu_vis = pd.concat([pd.concat(mu_list),df_vis])
cmap = sns.color_palette('tab20',8)
n_iter = i

sns.relplot(data=df_mu_vis,x='x0',y='x1',hue='class',col='iteration',
            col_wrap=3,hue_order = ['M0','A','M1','B','M2','C','M3','D'],
           palette = cmap,size='type',col_order=[str(i) for i in range(n_iter+1)])
```

```
<seaborn.axisgrid.FacetGrid at 0x7fd96c512150>
```

Here are a few different runs of the algorithm on the same dtaset.

k means

k means

k means

k means

k means

KMeans with Sklearn

```
iris_X, iris_y = datasets.load_iris(return_X_y=True)
```

```
km3 = KMeans(n_clusters=3)
```

```
iris_cluster3 = km3.fit_predict(iris_X)
```

```
plt.scatter(iris_X[:,0],iris_X[:,1],c=iris_cluster3)
```

```
<matplotlib.collections.PathCollection at 0x7fd965ec8bd0>
```

```
iris_cluster3 = km3.fit_predict(iris_X)
plt.scatter(iris_X[:,0],iris_X[:,1],c=iris_cluster3)
```

```
<matplotlib.collections.PathCollection at 0x7fd965d80b90>
```

Class 25: Evaluating Clustering

1. Log onto Prismia
2. (forgot to use this on monday, honoring post-halloween) share your favorite candy in the zoom chat (or just say hi)

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from sklearn import datasets
from sklearn import cluster
from sklearn import metrics
```

```
iris_X, iris_y = datasets.load_iris(return_X_y=True)
```

```
km3 = cluster.KMeans(n_clusters=3)
```

```
km3.fit(iris_X)
```

```
KMeans(n_clusters=3)
```

```
km3.cluster_centers_
```

```
array([[5.9016129 , 2.7483871 , 4.39354839, 1.43387097],
       [5.006     , 3.428     , 1.462     , 0.246     ],
       [6.85      , 3.07368421, 5.74210526, 2.07105263]])
```

```
plt.scatter(iris_X[:,0],iris_X[:,1],c=km3.labels_)
```

```
<matplotlib.collections.PathCollection at 0x7f44a3ab1590>
```

```
km2 = cluster.KMeans(n_clusters=2)
km2.fit(iris_X)
plt.scatter(iris_X[:,0],iris_X[:,1],c=km2.labels_)
```

```
<matplotlib.collections.PathCollection at 0x7f44a39cf250>
```

```
km4 = cluster.KMeans(n_clusters=4)
km4.fit(iris_X)
plt.scatter(iris_X[:,0],iris_X[:,1],c=km4.labels_)
```

```
<matplotlib.collections.PathCollection at 0x7f44a39cff90>
```

\[s = \frac{b-a}{max(a,b)}\]

a: The mean distance between a sample and all other points in the same class.

b: The mean distance between a sample and all other points in the next nearest cluster.

```
metrics.silhouette_score(iris_X, km3.labels_)
```

```
0.5528190123564091
```

```
metrics.silhouette_score(iris_X, km2.labels_)
```

```
0.681046169211746
```

```
metrics.silhouette_score(iris_X, km4.labels_)
```

```
0.49745518901737446
```

```
km3a = cluster.KMeans(n_clusters=3)
km3a.fit(iris_X)
plt.scatter(iris_X[:,0],iris_X[:,1],c=km3.labels_)
```

```
<matplotlib.collections.PathCollection at 0x7f44a38db790>
```

```
metrics.adjusted_rand_score(iris_y,km3.labels_)
```

```
0.7302382722834697
```

```
metrics.adjusted_rand_score(iris_y,km2.labels_)
```

```
0.5399218294207123
```

```
metrics.adjusted_rand_score(iris_y,km4.labels_)
```

```
0.6460787233460947
```

```
metrics.adjusted_mutual_info_score(iris_y,km3.labels_)
```

```
0.7551191675800482
```

```
metrics.adjusted_mutual_info_score(iris_y,km2.labels_)
```

```
0.6538380713762781
```

```
metrics.adjusted_mutual_info_score(iris_y,km4.labels_)
```

```
0.7151742230795862
```

Class 26: More Clustering Models

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from sklearn import datasets
from sklearn import cluster
from sklearn import metrics
```

```
iris_X, iris_y = datasets.load_iris(return_X_y = True)
```

```
msc = cluster.MeanShift()
```

```
msc.fit(iris_X)
```

```
MeanShift()
```

```
plt.scatter(iris_X[:,0],iris_X[:,1], c= msc.labels_)
```

```
<matplotlib.collections.PathCollection at 0x7f4f1d1d2110>
```

```
sns.pairplot(iris_X, hue= msc.labels_)
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-6-7e1512556191> in <module>
----> 1 sns.pairplot(iris_X, hue= msc.labels_)

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/seaborn/_decorators.py in inner_f(*args, **kwargs)
     44             )
     45         kwargs.update({k: arg for k, arg in zip(sig.parameters, args)})
---> 46         return f(**kwargs)
     47     return inner_f
     48

/opt/hostedtoolcache/Python/3.7.10/x64/lib/python3.7/site-
packages/seaborn/axisgrid.py in pairplot(data, hue, hue_order, palette, vars, x_vars,
y_vars, kind, diag_kind, markers, height, aspect, corner, dropna, plot_kws, diag_kws,
grid_kws, size)
   1972         raise TypeError(
   1973             "'data' must be pandas DataFrame object, not:
{typefound}".format(
-> 1974                 typefound=type(data)))
   1975
   1976     plot_kws = {} if plot_kws is None else plot_kws.copy()

TypeError: 'data' must be pandas DataFrame object, not: <class 'numpy.ndarray'>
```

```
spc = cluster.SpectralClustering(n_clusters=3)
```

```
spc.fit(iris_X)
```

```
SpectralClustering(n_clusters=3)
```

```
plt.scatter(iris_X[:,0],iris_X[:,1], c=spc.labels_)
```

```
<matplotlib.collections.PathCollection at 0x7f4f11874990>
```

```
km3 = cluster.KMeans(n_clusters=3)
km3.fit(iris_X)
plt.scatter(iris_X[:,0],iris_X[:,1], c=km3.labels_)
```

```
<matplotlib.collections.PathCollection at 0x7f4f11807b50>
```

```
from sklearn.neighbors import NearestNeighbors
```

```
nbrs = NearestNeighbors()
```

```
nbrs.fit(iris_X)
```

```
NearestNeighbors()
```

```
dists, indx = nbrs.kneighbors(iris_X)
```

```
dists
```

```
array([[0.        , 0.1       , 0.14142136, 0.14142136, 0.14142136],
       [0.        , 0.14142136, 0.14142136, 0.14142136, 0.17320508],
       [0.        , 0.14142136, 0.24494897, 0.26457513, 0.26457513],
       [0.        , 0.14142136, 0.17320508, 0.2236068 , 0.24494897],
       [0.        , 0.14142136, 0.14142136, 0.17320508, 0.17320508],
       [0.        , 0.33166248, 0.34641016, 0.36055513, 0.37416574],
       [0.        , 0.2236068 , 0.26457513, 0.3       , 0.31622777],
       [0.        , 0.1       , 0.14142136, 0.17320508, 0.2       ],
       [0.        , 0.14142136, 0.3       , 0.31622777, 0.34641016],
       [0.        , 0.1       , 0.17320508, 0.17320508, 0.17320508],
       [0.        , 0.1       , 0.28284271, 0.3       , 0.33166248],
       [0.        , 0.2236068 , 0.2236068 , 0.28284271, 0.3       ],
       [0.        , 0.14142136, 0.17320508, 0.2       , 0.2       ],
       [0.        , 0.24494897, 0.31622777, 0.34641016, 0.47958315],
       [0.        , 0.41231056, 0.46904158, 0.54772256, 0.55677644],
       [0.        , 0.36055513, 0.54772256, 0.6164414 , 0.6164414 ],
       [0.        , 0.34641016, 0.36055513, 0.38729833, 0.38729833],
       [0.        , 0.1       , 0.14142136, 0.17320508, 0.17320508],
       [0.        , 0.33166248, 0.38729833, 0.46904158, 0.50990195],
       [0.        , 0.14142136, 0.14142136, 0.24494897, 0.26457513],
       [0.        , 0.28284271, 0.3       , 0.36055513, 0.36055513],
       [0.        , 0.14142136, 0.24494897, 0.24494897, 0.26457513],
       [0.        , 0.45825757, 0.50990195, 0.50990195, 0.51961524],
       [0.        , 0.2       , 0.26457513, 0.37416574, 0.38729833],
       [0.        , 0.3       , 0.37416574, 0.41231056, 0.42426407],
       [0.        , 0.17320508, 0.2       , 0.2236068 , 0.2236068 ],
       [0.        , 0.2       , 0.2236068 , 0.2236068 , 0.24494897],
       [0.        , 0.14142136, 0.14142136, 0.14142136, 0.17320508],
       [0.        , 0.14142136, 0.14142136, 0.14142136, 0.17320508],
       [0.        , 0.14142136, 0.17320508, 0.2236068 , 0.2236068 ],
       [0.        , 0.14142136, 0.14142136, 0.17320508, 0.2236068 ],
       [0.        , 0.28284271, 0.3       , 0.3       , 0.31622777],
       [0.        , 0.34641016, 0.34641016, 0.37416574, 0.42426407],
       [0.        , 0.34641016, 0.36055513, 0.38729833, 0.41231056],
       [0.        , 0.1       , 0.14142136, 0.14142136, 0.17320508],
       [0.        , 0.2236068 , 0.3       , 0.31622777, 0.33166248],
       [0.        , 0.3       , 0.31622777, 0.33166248, 0.34641016],
       [0.        , 0.14142136, 0.24494897, 0.26457513, 0.26457513],
       [0.        , 0.14142136, 0.2       , 0.24494897, 0.3       ],
       [0.        , 0.1       , 0.14142136, 0.14142136, 0.14142136],
       [0.        , 0.14142136, 0.17320508, 0.17320508, 0.24494897],
       [0.        , 0.6244998 , 0.71414284, 0.76811457, 0.78102497],
       [0.        , 0.2       , 0.2236068 , 0.3       , 0.3       ],
       [0.        , 0.2236068 , 0.26457513, 0.31622777, 0.37416574],
       [0.        , 0.36055513, 0.37416574, 0.41231056, 0.41231056],
       [0.        , 0.14142136, 0.2       , 0.2       , 0.24494897],
       [0.        , 0.14142136, 0.24494897, 0.24494897, 0.3       ],
       [0.        , 0.14142136, 0.14142136, 0.2236068 , 0.2236068 ],
       [0.        , 0.1       , 0.2236068 , 0.24494897, 0.24494897],
       [0.        , 0.14142136, 0.17320508, 0.2236068 , 0.2236068 ],
```

```
[0.        , 0.26457513, 0.33166248, 0.43588989, 0.45825757],
[0.        , 0.26457513, 0.31622777, 0.34641016, 0.37416574],
[0.        , 0.26457513, 0.28284271, 0.31622777, 0.34641016],
[0.        , 0.2       , 0.3       , 0.31622777, 0.43588989],
[0.        , 0.24494897, 0.31622777, 0.37416574, 0.37416574],
[0.        , 0.3       , 0.31622777, 0.31622777, 0.33166248],
[0.        , 0.26457513, 0.37416574, 0.42426407, 0.45825757],
[0.        , 0.14142136, 0.38729833, 0.45825757, 0.72111026],
[0.        , 0.24494897, 0.24494897, 0.31622777, 0.31622777],
[0.        , 0.38729833, 0.50990195, 0.51961524, 0.52915026],
[0.        , 0.36055513, 0.45825757, 0.67082039, 0.71414284],
[0.        , 0.3       , 0.33166248, 0.36055513, 0.36055513],
[0.        , 0.48989795, 0.51961524, 0.54772256, 0.58309519],
[0.        , 0.14142136, 0.2236068 , 0.24494897, 0.42426407],
[0.        , 0.42426407, 0.4472136 , 0.50990195, 0.51961524],
[0.        , 0.14142136, 0.31622777, 0.31622777, 0.34641016],
[0.        , 0.2       , 0.3       , 0.38729833, 0.41231056],
[0.        , 0.24494897, 0.28284271, 0.33166248, 0.36055513],
[0.        , 0.26457513, 0.50990195, 0.53851648, 0.678233  ],
[0.        , 0.17320508, 0.24494897, 0.26457513, 0.26457513],
[0.        , 0.2236068 , 0.3       , 0.36055513, 0.42426407],
[0.        , 0.33166248, 0.34641016, 0.37416574, 0.4       ],
[0.        , 0.36055513, 0.36055513, 0.41231056, 0.42426407],
[0.        , 0.2236068 , 0.3       , 0.38729833, 0.43588989],
[0.        , 0.2       , 0.26457513, 0.36055513, 0.38729833],
[0.        , 0.14142136, 0.24494897, 0.26457513, 0.31622777],
[0.        , 0.31622777, 0.34641016, 0.34641016, 0.37416574],
[0.        , 0.31622777, 0.37416574, 0.41231056, 0.42426407],
[0.        , 0.2       , 0.24494897, 0.33166248, 0.34641016],
[0.        , 0.34641016, 0.42426407, 0.43588989, 0.4472136 ],
[0.        , 0.14142136, 0.17320508, 0.3       , 0.3       ],
[0.        , 0.14142136, 0.26457513, 0.34641016, 0.43588989],
[0.        , 0.14142136, 0.26457513, 0.28284271, 0.3       ],
[0.        , 0.33166248, 0.36055513, 0.36055513, 0.37416574],
[0.        , 0.2       , 0.41231056, 0.47958315, 0.48989795],
[0.        , 0.37416574, 0.42426407, 0.45825757, 0.46904158],
[0.        , 0.28284271, 0.31622777, 0.31622777, 0.33166248],
[0.        , 0.26457513, 0.57445626, 0.59160798, 0.60827625],
[0.        , 0.17320508, 0.17320508, 0.2236068 , 0.31622777],
[0.        , 0.2       , 0.24494897, 0.3       , 0.3       ],
[0.        , 0.26457513, 0.31622777, 0.42426407, 0.42426407],
[0.        , 0.14142136, 0.2       , 0.3       , 0.34641016],
[0.        , 0.14142136, 0.24494897, 0.26457513, 0.26457513],
[0.        , 0.14142136, 0.36055513, 0.38729833, 0.64807407],
[0.        , 0.17320508, 0.2236068 , 0.26457513, 0.3       ],
[0.        , 0.14142136, 0.17320508, 0.24494897, 0.33166248],
[0.        , 0.14142136, 0.14142136, 0.17320508, 0.2236068 ],
[0.        , 0.2       , 0.33166248, 0.34641016, 0.34641016],
[0.        , 0.38729833, 0.38729833, 0.72111026, 0.79372539],
[0.        , 0.14142136, 0.17320508, 0.2236068 , 0.24494897],
[0.        , 0.42426407, 0.5       , 0.50990195, 0.55677644],
[0.        , 0.        , 0.26457513, 0.31622777, 0.33166248],
[0.        , 0.38729833, 0.4       , 0.41231056, 0.45825757],
[0.        , 0.24494897, 0.24494897, 0.33166248, 0.38729833],
[0.        , 0.3       , 0.31622777, 0.36055513, 0.38729833],
[0.        , 0.26457513, 0.52915026, 0.54772256, 0.54772256],
[0.        , 0.73484692, 0.76157731, 0.79372539, 0.87749644],
[0.        , 0.26457513, 0.43588989, 0.52915026, 0.54772256],
[0.        , 0.55677644, 0.6       , 0.6164414 , 0.6164414 ],
[0.        , 0.63245553, 0.67082039, 0.70710678, 0.75498344],
[0.        , 0.2236068 , 0.37416574, 0.42426407, 0.42426407],
[0.        , 0.34641016, 0.37416574, 0.37416574, 0.38729833],
[0.        , 0.17320508, 0.34641016, 0.36055513, 0.37416574],
[0.        , 0.26457513, 0.26457513, 0.33166248, 0.51961524],
[0.        , 0.48989795, 0.50990195, 0.50990195, 0.51961524],
[0.        , 0.3       , 0.37416574, 0.37416574, 0.38729833],
[0.        , 0.14142136, 0.24494897, 0.36055513, 0.38729833],
[0.        , 0.41231056, 0.81853528, 0.86023253, 1.00498756],
[0.        , 0.41231056, 0.54772256, 0.89442719, 0.92736185],
[0.        , 0.43588989, 0.51961524, 0.53851648, 0.58309519],
[0.        , 0.2236068 , 0.26457513, 0.3       , 0.3       ],
[0.        , 0.31622777, 0.31622777, 0.33166248, 0.45825757],
[0.        , 0.26457513, 0.41231056, 0.60827625, 0.678233  ],
[0.        , 0.17320508, 0.24494897, 0.36055513, 0.36055513],
[0.        , 0.3       , 0.31622777, 0.37416574, 0.37416574],
[0.        , 0.34641016, 0.38729833, 0.43588989, 0.46904158],
[0.        , 0.17320508, 0.24494897, 0.28284271, 0.38729833],
[0.        , 0.14142136, 0.24494897, 0.28284271, 0.3       ],
[0.        , 0.1       , 0.31622777, 0.33166248, 0.37416574],
[0.        , 0.34641016, 0.50990195, 0.51961524, 0.55677644],
[0.        , 0.26457513, 0.45825757, 0.46904158, 0.50990195],
[0.        , 0.41231056, 0.88317609, 0.92736185, 0.93273791],
[0.        , 0.1       , 0.3       , 0.42426407, 0.43588989],
[0.        , 0.33166248, 0.36055513, 0.37416574, 0.43588989],
[0.        , 0.53851648, 0.55677644, 0.58309519, 0.66332496],
[0.        , 0.53851648, 0.54772256, 0.66332496, 0.678233  ],
[0.        , 0.24494897, 0.38729833, 0.42426407, 0.43588989],
```

```
       [0.        , 0.14142136, 0.24494897, 0.38729833, 0.43588989],
       [0.        , 0.14142136, 0.2236068 , 0.28284271, 0.31622777],
       [0.        , 0.17320508, 0.36055513, 0.36055513, 0.37416574],
       [0.        , 0.24494897, 0.26457513, 0.34641016, 0.34641016],
       [0.        , 0.24494897, 0.36055513, 0.46904158, 0.50990195],
       [0.        , 0.        , 0.26457513, 0.31622777, 0.33166248],
       [0.        , 0.2236068 , 0.31622777, 0.31622777, 0.34641016],
       [0.        , 0.24494897, 0.3       , 0.31622777, 0.4       ],
       [0.        , 0.24494897, 0.36055513, 0.36055513, 0.37416574],
       [0.        , 0.24494897, 0.37416574, 0.38729833, 0.41231056],
       [0.        , 0.2236068 , 0.34641016, 0.36055513, 0.36055513],
       [0.        , 0.24494897, 0.3       , 0.55677644, 0.6164414 ],
       [0.        , 0.28284271, 0.31622777, 0.33166248, 0.33166248]])
```

```
indx
```

```
array([[  0,  17,   4,  39,  27],
       [  1,  34,  45,  12,   9],
       [  2,  47,   3,  12,   6],
       [  3,  47,  29,  30,   2],
       [  4,  37,   0,  17,  40],
       [  5,  18,  10,  48,  44],
       [  6,  47,   2,  11,  42],
       [  7,  39,  49,   0,  17],
       [  8,  38,   3,  42,  13],
       [  9,  34,   1,  30,  12],
       [ 10,  48,  27,  36,  19],
       [ 11,  29,   7,  26,  24],
       [ 12,   1,   9,  45,  34],
       [ 13,  38,  42,   8,  47],
       [ 14,  33,  16,  15,  18],
       [ 15,  33,  14,   5,  16],
       [ 16,  10,  48,  33,  19],
       [ 17,   0,  40,   4,  39],
       [ 18,   5,  10,  48,  20],
       [ 19,  21,  46,  48,   4],
       [ 20,  31,  27,  28,  10],
       [ 21,  19,  46,  17,   4],
       [ 22,   6,   2,  37,  40],
       [ 23,  26,  43,  39,   7],
       [ 24,  11,  29,  26,  30],
       [ 25,  34,   9,   1,  30],
       [ 26,  23,  43,   7,  39],
       [ 27,  28,   0,  39,  17],
       [ 28,  27,   0,  39,  17],
       [ 29,  30,   3,  11,  47],
       [ 30,  29,  34,   9,  25],
       [ 31,  20,  28,  27,  36],
       [ 32,  46,  33,  19,  48],
       [ 33,  32,  15,  16,  14],
       [ 34,   9,   1,  30,  25],
       [ 35,  49,   1,   2,  40],
       [ 36,  10,  31,  28,  48],
       [ 37,   4,   0,  40,   7],
       [ 38,   8,  42,  13,   3],
       [ 39,   7,   0,  28,  27],
       [ 40,  17,   0,   4,   7],
       [ 41,   8,  38,  45,  13],
       [ 42,  38,  47,   3,   2],
       [ 43,  26,  23,  21,  17],
       [ 44,  46,   5,  21,  19],
       [ 45,   1,  12,  34,  30],
       [ 46,  19,  21,  48,   4],
       [ 47,   3,   2,  42,   6],
       [ 48,  10,  27,  19,  46],
       [ 49,   7,  39,  35,   0],
       [ 50,  52,  86,  65,  76],
       [ 51,  56,  75,  65,  86],
       [ 52,  50,  86,  77,  76],
       [ 53,  89,  80,  69,  81],
       [ 54,  58,  75,  76,  86],
       [ 55,  66,  90,  96,  94],
       [ 56,  51,  85,  91, 127],
       [ 57,  93,  98,  60,  81],
       [ 58,  75,  54,  65,  76],
       [ 59,  89,  94,  53,  80],
       [ 60,  93,  57,  81,  80],
       [ 61,  96,  78,  95,  99],
       [ 62,  92,  69,  67,  80],
       [ 63,  91,  73,  78,  97],
       [ 64,  82,  79,  88,  99],
       [ 65,  75,  58,  86,  51],
       [ 66,  84,  55,  96,  78],
       [ 67,  92,  82,  99,  69],
       [ 68,  87,  72, 119,  54],
```

```
       [ 69,  80,  89,  81,  92],
       [ 70, 138, 127, 149,  85],
       [ 71,  97,  82,  92,  61],
       [ 72, 133, 123, 146,  83],
       [ 73,  63,  91,  78,  97],
       [ 74,  97,  75,  58,  54],
       [ 75,  65,  58,  74,  54],
       [ 76,  58,  86,  52,  54],
       [ 77,  52,  86, 147, 110],
       [ 78,  91,  63,  61,  97],
       [ 79,  81,  80,  69,  64],
       [ 80,  81,  69,  89,  53],
       [ 81,  80,  69,  79,  89],
       [ 82,  92,  99,  67,  69],
       [ 83, 133, 101, 142, 149],
       [ 84,  66,  55,  96,  88],
       [ 85,  56,  70,  51,  91],
       [ 86,  52,  65,  58,  50],
       [ 87,  68,  72,  62,  54],
       [ 88,  96,  95,  99,  94],
       [ 89,  53,  69,  80,  94],
       [ 90,  94,  55,  96,  89],
       [ 91,  63,  78,  73,  97],
       [ 92,  82,  67,  99,  69],
       [ 93,  57,  60,  98,  81],
       [ 94,  99,  96,  90,  89],
       [ 95,  96,  88,  99,  94],
       [ 96,  95,  99,  88,  94],
       [ 97,  74,  71,  91,  78],
       [ 98,  57,  93,  60,  79],
       [ 99,  96,  94,  88,  95],
       [100, 136, 144, 104, 143],
       [101, 142, 113, 121, 149],
       [102, 125, 120, 143, 130],
       [103, 116, 137, 128, 111],
       [104, 132, 128, 140, 124],
       [105, 122, 107, 135, 118],
       [106,  84,  59,  90,  89],
       [107, 130, 125, 105, 102],
       [108, 128, 103, 132, 116],
       [109, 143, 120, 144, 102],
       [110, 147, 115,  77, 145],
       [111, 147, 128, 146, 103],
       [112, 139, 140, 120, 145],
       [113, 101, 142, 121, 114],
       [114, 121, 101, 142, 113],
       [115, 148, 145, 110, 147],
       [116, 137, 103, 147, 111],
       [117, 131, 105, 109, 135],
       [118, 122, 105, 135, 107],
       [119,  72,  83,  68, 146],
       [120, 143, 140, 124, 144],
       [121, 101, 142, 113, 149],
       [122, 105, 118, 107, 130],
       [123, 126, 146, 127,  72],
       [124, 120, 143, 140, 112],
       [125, 129, 102, 107, 130],
       [126, 123, 127, 138, 146],
       [127, 138, 126, 149,  70],
       [128, 132, 104, 103, 111],
       [129, 125, 130, 102, 107],
       [130, 107, 102, 125, 129],
       [131, 117, 105, 135, 109],
       [132, 128, 104, 103, 111],
       [133,  83,  72, 123, 126],
       [134, 103,  83, 133, 111],
       [135, 130, 105, 102, 107],
       [136, 148, 115, 100, 144],
       [137, 116, 103, 147, 128],
       [138, 127,  70, 126, 149],
       [139, 112, 145, 141, 120],
       [140, 144, 120, 112, 143],
       [141, 145, 139, 112, 110],
       [101, 142, 113, 121, 149],
       [143, 120, 124, 144, 140],
       [144, 140, 120, 143, 124],
       [145, 141, 147, 139, 112],
       [146, 123, 111, 126,  72],
       [147, 110, 111, 116, 145],
       [148, 136, 115, 110, 147],
       [149, 127, 138, 101, 142]])
```

Class 27: Model Optimization- Choosing K

1. On the zoom chat, say hello
2. Log onto Prismia

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd
from sklearn import datasets
from sklearn import cluster
from sklearn import metrics

url ='https://raw.githubusercontent.com/rhodyprog4ds/06-naive-
bayes/main/data/dataset2.csv'
```

```
df = pd.read_csv(url)
```

```
X = df.values[:,:-1]
```

```
score = []

for k in range(2,10):
    km = cluster.KMeans(n_clusters=k)
    km.fit(X)
    score.append(metrics.silhouette_score(X,km.labels_))
```

```
score
```

```
[0.6151956537579685,
 0.5689398715407455,
 0.5388658222509434,
 0.5156765142705673,
 0.4948750100931285,
 0.47805666819267567,
 0.45856852844475937,
 0.4437858458538978]
```

```
plt.plot(range(2,10), score)
```

```
[<matplotlib.lines.Line2D at 0x7fdea0059310>]
```

```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7fde9ffc1c10>
```

```
iris_X, _ = datasets.load_iris(return_X_y = True)
```

```
score_iris = []

for k in range(2,10):
    km = cluster.KMeans(n_clusters=k)
    km.fit(iris_X)
    score_iris.append(metrics.silhouette_score(iris_X,km.labels_))
```

```
plt.plot(range(2,10), score_iris)
```

```
[<matplotlib.lines.Line2D at 0x7fde9ee967d0>]
```

```
silohette_iris = []
ch_iris = []
db_iris = []

for k in range(2,10):
    km = cluster.KMeans(n_clusters=k)
    km.fit(iris_X)
    silohette_iris.append(metrics.silhouette_score(iris_X,km.labels_))
    ch_iris.append(metrics.calinski_harabasz_score(iris_X,km.labels_))
    db_iris.append(metrics.davies_bouldin_score(iris_X,km.labels_))
```

```
ch_iris = np.asarray(ch_iris)/np.max(ch_iris)
plt.plot(range(2,10), silohette_iris)
plt.plot(range(2,10), ch_iris)
plt.plot(range(2,10), db_iris)
```

```
[<matplotlib.lines.Line2D at 0x7fde9caaa9d0>]
```

Class 28: SVM & Model Optimization

1. Use 1 word to say how your portfolio check 2 is going in the zoom chat
2. Log onto prismia & share any final questions you have about the portfolio

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from sklearn import datasets
from sklearn import cluster
from sklearn import svm, datasets
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import train_test_split
```

```
iris_X, iris_y = datasets.load_iris(return_X_y= True)
```

```
iris_X_train, iris_X_test, iris_y_train, iris_y_test = train_test_split(iris_X,iris_y)
```

```
svm_clf = svm.SVC(kernel='linear')
```

```
svm_clf.fit(iris_X_train,iris_y_train)
```

```
SVC(kernel='linear')
```

```
svm_clf.score(iris_X_test, iris_y_test)
```

```
1.0
```

```
param_grid = {'kernel':['linear','rbf'], 'C':[.5, 1, 10]}
svm_opt = GridSearchCV(svm_clf,param_grid,)
```

```
svm_opt.fit(iris_X, iris_y)
```

```
GridSearchCV(estimator=SVC(kernel='linear'),
             param_grid={'C': [0.5, 1, 10], 'kernel': ['linear', 'rbf']})
```

```
type(svm_opt.best_estimator_)
```

```
sklearn.svm._classes.SVC
```

```
svm_opt.best_params_
```

```
{'C': 0.5, 'kernel': 'linear'}
```

```
svm_opt.cv_results_
```

```
{'mean_fit_time': array([0.00061193, 0.00076184, 0.00062265, 0.00072966, 0.00060325,
        0.00064535]),
 'std_fit_time': array([1.05270116e-04, 7.82869614e-05, 4.61122982e-05, 4.40765221e-
05,
        2.30151865e-05, 3.56278736e-05]),
 'mean_score_time': array([0.00028019, 0.00037417, 0.00029044, 0.00034089,
0.00027933,
        0.00030575]),
 'std_score_time': array([5.55222356e-05, 6.42569874e-05, 2.11074649e-05,
2.79559992e-05,
        3.98430475e-05, 2.54760488e-05]),
 'param_C': masked_array(data=[0.5, 0.5, 1, 1, 10, 10],
              mask=[False, False, False, False, False, False],
        fill_value='?',
             dtype=object),
 'param_kernel': masked_array(data=['linear', 'rbf', 'linear', 'rbf', 'linear',
'rbf'],
              mask=[False, False, False, False, False, False],
        fill_value='?',
             dtype=object),
 'params': [{'C': 0.5, 'kernel': 'linear'},
  {'C': 0.5, 'kernel': 'rbf'},
  {'C': 1, 'kernel': 'linear'},
  {'C': 1, 'kernel': 'rbf'},
  {'C': 10, 'kernel': 'linear'},
  {'C': 10, 'kernel': 'rbf'}],
 'split0_test_score': array([0.96666667, 0.93333333, 0.96666667, 0.96666667, 1.
,
        0.96666667]),
 'split1_test_score': array([1.        , 0.96666667, 1.        , 0.96666667, 1.
,
        1.        ]),
 'split2_test_score': array([1.        , 0.96666667, 0.96666667, 0.96666667, 0.9
,
        0.96666667]),
 'split3_test_score': array([0.96666667, 0.93333333, 0.96666667, 0.93333333,
0.96666667,
        0.96666667]),
 'split4_test_score': array([1., 1., 1., 1., 1., 1.]),
 'mean_test_score': array([0.98666667, 0.96      , 0.98      , 0.96666667,
0.97333333,
        0.98      ]),
 'std_test_score': array([0.01632993, 0.02494438, 0.01632993, 0.02108185, 0.03887301,
        0.01632993]),
 'rank_test_score': array([1, 6, 2, 5, 4, 2], dtype=int32)}
```

```
import pandas as pd
```

```
pd.DataFrame(svm_opt.cv_results_)
```

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_C	param_kerne
0	0.000612	0.000105	0.000280	0.000056	0.5	linea
1	0.000762	0.000078	0.000374	0.000064	0.5	rb
2	0.000623	0.000046	0.000290	0.000021	1	linea
3	0.000730	0.000044	0.000341	0.000028	1	rb
4	0.000603	0.000023	0.000279	0.000040	10	linea
5	0.000645	0.000036	0.000306	0.000025	10	rb

```
svm_opt.best_estimator_.predict(iris_X_test)
```

```
array([2, 0, 0, 2, 2, 2, 0, 0, 1, 0, 2, 2, 2, 2, 1, 2, 0, 1, 2, 1, 0, 2,
       0, 0, 2, 2, 0, 0, 2, 2, 1, 0, 0, 2, 0, 2, 0, 1])
```

```
from sklearn import tree
```

Find the optimal criterion, max_depth and min_samples_leaf for a decisoin tree on the iris data

```
dt = tree.DecisionTreeClassifier()
params_dt = {'criterion':['gini','entropy'],'max_depth':[2,3,4],
             'min_samples_leaf':list(range(2,20,2))}
dt_opt = GridSearchCV(dt,params_dt)
dt_opt.fit(iris_X,iris_y)
```

```
GridSearchCV(estimator=DecisionTreeClassifier(),
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_depth': [2, 3, 4],
                         'min_samples_leaf': [2, 4, 6, 8, 10, 12, 14, 16, 18]})
```

```
dt_opt.best_params_
```

```
{'criterion': 'gini', 'max_depth': 3, 'min_samples_leaf': 2}
```

Class 29: Choosing a Model

1. log onto prismia
2. share your favorite restaraunt on/near campus in the zoom chat

Portfolio PR

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd
from sklearn import datasets
from sklearn import cluster
from sklearn import svm
from sklearn import tree
from sklearn import model_selection
```

```
iris_X, iris_y = datasets.load_iris(return_X_y= True)
iris_X_train, iris_X_test, iris_y_train, iris_y_test =
model_selection.train_test_split(iris_X,iris_y)
dt = tree.DecisionTreeClassifier()
params_dt = {'criterion':['gini','entropy'],'max_depth':[2,3,4],
             'min_samples_leaf':list(range(2,20,2))}
dt_opt = model_selection.GridSearchCV(dt,params_dt)
dt_opt.fit(iris_X_train,iris_y_train)
```

```
GridSearchCV(estimator=DecisionTreeClassifier(),
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_depth': [2, 3, 4],
                         'min_samples_leaf': [2, 4, 6, 8, 10, 12, 14, 16, 18]})
```

```
dt_opt.predict(iris_X_test)
```

```
array([1, 0, 0, 1, 2, 2, 0, 2, 0, 2, 0, 0, 2, 1, 2, 0, 2, 2, 2, 2, 0, 2,
       2, 0, 2, 1, 0, 1, 1, 0, 0, 2, 1, 2, 2, 1, 2, 2])
```

```
dt_opt.best_estimator_.predict(iris_X_test)
```

```
array([1, 0, 0, 1, 2, 2, 0, 2, 0, 2, 0, 0, 2, 1, 2, 0, 2, 2, 2, 2, 0, 2,
       2, 0, 2, 1, 0, 1, 1, 0, 0, 2, 1, 2, 2, 1, 2, 2])
```

```
pd.DataFrame(dt_opt.cv_results_)
```

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_criterion	parar
0	0.000561	0.000138	0.000276	0.000041	gini	
1	0.000443	0.000023	0.000238	0.000013	gini	
2	0.000484	0.000022	0.000266	0.000011	gini	
3	0.000447	0.000027	0.000239	0.000015	gini	
4	0.000489	0.000022	0.000255	0.000008	gini	
5	0.000473	0.000031	0.000244	0.000019	gini	
6	0.000460	0.000031	0.000250	0.000013	gini	
7	0.000447	0.000031	0.000249	0.000019	gini	
8	0.000477	0.000021	0.000266	0.000032	gini	
9	0.000475	0.000038	0.000248	0.000021	gini	
10	0.000502	0.000014	0.000261	0.000023	gini	
11	0.000470	0.000040	0.000243	0.000024	gini	
12	0.000438	0.000026	0.000234	0.000016	gini	
13	0.000474	0.000046	0.000252	0.000026	gini	

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_criterion	parar
14	0.000499	0.000018	0.000263	0.000010	gini	
15	0.000500	0.000020	0.000258	0.000008	gini	
16	0.000495	0.000023	0.000259	0.000009	gini	
17	0.000481	0.000022	0.000249	0.000015	gini	
18	0.000486	0.000025	0.000246	0.000021	gini	
19	0.000499	0.000019	0.000257	0.000008	gini	
20	0.000481	0.000033	0.000250	0.000025	gini	
21	0.000476	0.000029	0.000251	0.000017	gini	
22	0.000480	0.000043	0.000248	0.000020	gini	
23	0.000473	0.000036	0.000247	0.000023	gini	
24	0.000497	0.000020	0.000255	0.000016	gini	
25	0.000425	0.000049	0.000234	0.000027	gini	
26	0.000433	0.000034	0.000225	0.000008	gini	
27	0.000472	0.000049	0.000247	0.000022	entropy	
28	0.000465	0.000039	0.000248	0.000017	entropy	

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_criterion	parar
29	0.000451	0.000057	0.000229	0.000030	entropy	
30	0.000438	0.000037	0.000226	0.000025	entropy	
31	0.000412	0.000060	0.000224	0.000040	entropy	
32	0.000480	0.000034	0.000240	0.000025	entropy	
33	0.000465	0.000035	0.000246	0.000023	entropy	
34	0.000470	0.000032	0.000236	0.000024	entropy	
35	0.000443	0.000062	0.000224	0.000025	entropy	
36	0.000450	0.000054	0.000237	0.000034	entropy	
37	0.000466	0.000049	0.000223	0.000029	entropy	
38	0.000419	0.000030	0.000211	0.000014	entropy	
39	0.000478	0.000057	0.000237	0.000036	entropy	
40	0.000463	0.000032	0.000237	0.000017	entropy	
41	0.000427	0.000049	0.000225	0.000033	entropy	
42	0.000453	0.000041	0.000223	0.000026	entropy	
43	0.000422	0.000049	0.000224	0.000032	entropy	
44	0.000393	0.000039	0.000203	0.000021	entropy	
45	0.000459	0.000069	0.000225	0.000041	entropy	

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_criterion	parar
46	0.000407	0.000035	0.000200	0.000014	entropy	
47	0.000489	0.000054	0.000256	0.000019	entropy	
48	0.000438	0.000053	0.000226	0.000027	entropy	
49	0.000413	0.000072	0.000209	0.000032	entropy	
50	0.000400	0.000035	0.000218	0.000026	entropy	
51	0.000445	0.000060	0.000241	0.000029	entropy	
52	0.000440	0.000031	0.000230	0.000024	entropy	
53	0.000470	0.000039	0.000255	0.000019	entropy	

```
df = pd.DataFrame(dt_opt.cv_results_)
df.plot('mean_score_time','mean_test_score')
```

```
<AxesSubplot:xlabel='mean_score_time'>
```

```
%load http://drsmb.co
```

```
param_grid = {'kernel':['linear','rbf'], 'C':[.5, 1, 10]}
svm_clf = svm.SVC(kernel='linear')
svm_opt =model_selection.GridSearchCV(svm_clf,param_grid,)
svm_opt.fit(iris_X_train, iris_y_train)
```

```
GridSearchCV(estimator=SVC(kernel='linear'),
             param_grid={'C': [0.5, 1, 10], 'kernel': ['linear', 'rbf']})
```

```
df_svm = pd.DataFrame(svm_opt.cv_results_)
df_svm.plot.scatter('mean_score_time','mean_test_score')
```

```
<AxesSubplot:xlabel='mean_score_time', ylabel='mean_test_score'>
```

```
df_svm.plot.scatter('mean_score_time','std_test_score')
```

```
<AxesSubplot:xlabel='mean_score_time', ylabel='std_test_score'>
```

Class 30: Learning Curves, Validation Curves

1. Join prismia
2. say hello in the zoom chat

```
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd
from sklearn import datasets
from sklearn import cluster
from sklearn import svm
from sklearn import tree
from sklearn import model_selection
```

```
iris_X , iris_y = datasets.load_iris(return_X_y= True)
iris_X_train, iris_X_test, iris_y_train, iris_y_test =
model_selection.train_test_split(
            iris_X , iris_y,test_size =.2)
```

```
param_grid = {'kernel':['linear','rbf'], 'C':[.5, 1, 10]}
svm_clf = svm.SVC(kernel='linear')
svm_opt =model_selection.GridSearchCV(svm_clf,param_grid,)
svm_opt.fit(iris_X_train, iris_y_train)
```

```
GridSearchCV(estimator=SVC(kernel='linear'),
             param_grid={'C': [0.5, 1, 10], 'kernel': ['linear', 'rbf']})
```

```
df_svm = pd.DataFrame(svm_opt.cv_results_)
df_svm.sort_values(by='mean_score_time',inplace=True)
plt.errorbar(df_svm['mean_score_time'],df_svm['mean_test_score'],
df_svm['std_test_score'])
```

```
<ErrorbarContainer object of 3 artists>
```

```
model_selection.learning_curve(svm_opt.best_estimator_,iris_X_train, iris_y_train,
                               train_sizes= [.4,.5,.6,.8])
```

```
(array([38, 48, 57, 76]),
 array([[0.97368421, 1.        , 1.        , 1.        , 1.        ],
        [0.97916667, 0.97916667, 1.        , 1.        , 1.        ],
        [0.98245614, 0.98245614, 0.98245614, 1.        , 1.        ],
        [0.98684211, 0.98684211, 0.97368421, 0.97368421, 0.97368421]]),
 array([[1.        , 0.95833333, 0.91666667, 0.95833333, 1.        ],
        [1.        , 0.95833333, 0.91666667, 0.91666667, 0.95833333],
        [1.        , 1.        , 0.91666667, 0.91666667, 0.95833333],
        [1.        , 1.        , 0.91666667, 0.95833333, 0.95833333]]))
```

```
# %load http://drsmb.co/310
def lc_plot(train_sizes, train_scores, valid_scores):
    ts_len = len(train_sizes)
    ts = np.reshape(train_sizes,(ts_len,1))

    cols = ['size']
    cols.extend(['split_' + str(i) for i in range(5)])

    df = pd.DataFrame(np.block([[ts,train_scores],[ts,valid_scores]]),columns = cols)
    df['type'] = ['train']*ts_len + ['validation']*ts_len

    df_m = df.melt(id_vars=['size','type'],value_name='score',var_name='split')
    g = sns.lmplot(data=df_m,x='size', y='score',hue='type',)
    return g
```

```
train_sizes, train_scores, valid_scores =
model_selection.learning_curve(svm_opt.best_estimator_,iris_X_train, iris_y_train,
                               train_sizes= [.4,.5,.6,.8])
lc_plot(train_sizes, train_scores, valid_scores)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f8f3ce22c90>
```

```
train_scores, valid_scores =
model_selection.validation_curve(tree.DecisionTreeClassifier(),iris_X_train,
iris_y_train,
                         param_name='max_depth', param_range=list(range(1,6)))
```

```
param_values = list(range(1,6))
```

```
lc_plot(param_values, train_scores, valid_scores)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f8f3cca3b90>
```

```
digits_X, digits_y = datasets.load_digits(return_X_y = True)
```

```
digits_X.shape
```

```
(1797, 64)
```

Fit an SVM, with rbf kernel, examine the learning curve and the validation curve for paramters gamma

```
param_range = np.logspace(-6, -1, 5)
train_scores, valid_scores = model_selection.validation_curve(svm.SVC(),
                                                              digits_X, digits_y,
                        param_name='gamma', param_range=param_range)
lc_plot(param_range, train_scores, valid_scores)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f8f3cbed290>
```

```
train_sizes, train_scores, valid_scores = model_selection.learning_curve(svm.SVC(),
                                                              digits_X, digits_y)
```

Class 31: Confidence Intervals

1. respond in zoom chat: iced vs. hot coffee (or tea, or beverages in general). one all year? seasonal?
2. log onto prismia chat
3. respond on prismia with any questions you have about the course material so far:

- what are you confused about?
- wht do you want to know about about?
- what thing keeps tripping you up on assignments?

Admin

Confidence intervals

```
# %load http://drsmb.co/310
# %load http://drsmb.co/310
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd
from sklearn import datasets
from sklearn import cluster
from sklearn import svm
from sklearn import tree
from sklearn import model_selection
```

```
iris_X , iris_y = datasets.load_iris(return_X_y= True)
iris_X_train, iris_X_test, iris_y_train, iris_y_test =
model_selection.train_test_split(
            iris_X , iris_y,test_size =.2, random_state=0)

param_grid = {'kernel':['linear','rbf'], 'C':[.5, 1, 10]}
svm_clf = svm.SVC(kernel='linear')
svm_opt =model_selection.GridSearchCV(svm_clf,param_grid,)
svm_opt.fit(iris_X_train, iris_y_train)

df_svm = pd.DataFrame(svm_opt.cv_results_)
df_svm.sort_values(by='mean_score_time',inplace=True)
plt.errorbar(df_svm['mean_score_time'],df_svm['mean_test_score'],
df_svm['std_test_score'])
```

```
<ErrorbarContainer object of 3 artists>
```

```
iris_X.shape
```

```
(150, 4)
```

training

```
150*.8
```

```
120.0
```

```
iris_X_train.shape
```

```
(120, 4)
```

Cross validation tests

```
120*.2
```

```
24.0
```

```
iris_X_test
```

```
array([[5.8, 2.8, 5.1, 2.4],
       [6. , 2.2, 4. , 1. ],
       [5.5, 4.2, 1.4, 0.2],
       [7.3, 2.9, 6.3, 1.8],
       [5. , 3.4, 1.5, 0.2],
       [6.3, 3.3, 6. , 2.5],
       [5. , 3.5, 1.3, 0.3],
       [6.7, 3.1, 4.7, 1.5],
       [6.8, 2.8, 4.8, 1.4],
       [6.1, 2.8, 4. , 1.3],
       [6.1, 2.6, 5.6, 1.4],
       [6.4, 3.2, 4.5, 1.5],
       [6.1, 2.8, 4.7, 1.2],
       [6.5, 2.8, 4.6, 1.5],
       [6.1, 2.9, 4.7, 1.4],
       [4.9, 3.6, 1.4, 0.1],
       [6. , 2.9, 4.5, 1.5],
       [5.5, 2.6, 4.4, 1.2],
       [4.8, 3. , 1.4, 0.3],
       [5.4, 3.9, 1.3, 0.4],
       [5.6, 2.8, 4.9, 2. ],
       [5.6, 3. , 4.5, 1.5],
       [4.8, 3.4, 1.9, 0.2],
       [4.4, 2.9, 1.4, 0.2],
       [6.2, 2.8, 4.8, 1.8],
       [4.6, 3.6, 1. , 0.2],
       [5.1, 3.8, 1.9, 0.4],
       [6.2, 2.9, 4.3, 1.3],
       [5. , 2.3, 3.3, 1. ],
       [5. , 3.4, 1.6, 0.4]])
```

```
svm_opt.score(iris_X_test,iris_y_test)
```

```
1.0
```

```
# %load http://drsmb.co/310
def classification_confint(acc, n):
    '''
    Compute the 95% confidence interval for a classification problem.
      acc -- classification accuracy
      n   -- number of observations used to compute the accuracy
    Returns a tuple (lb,ub)
    '''
    interval = 1.96*np.sqrt(acc*(1-acc)/n)
    lb = max(0, acc - interval)
    ub = min(1.0, acc + interval)
    return (lb,ub)
```

```
classification_confint(svm_opt.score(iris_X_test,iris_y_test),len(iris_y_test))
```

```
(1.0, 1.0)
```

```
classification_confint(.9999,len(iris_y_test))
```

```
(0.9963217248848085, 1.0)
```

```
classification_confint(.85,len(iris_y_test))
```

```
(0.722223632858028, 0.9777763671419719)
```

```
classification_confint(.93,len(iris_y_test))
```

```
(0.8386968127609995, 1.0)
```

```
classification_confint(.85,50)
```

```
(0.7510248516040516, 0.9489751483959483)
```

```
classification_confint(.93,50)
```

```
(0.8592768552735387, 1.0)
```

```
df_svm
```

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_C	param_kerne
5	0.000566	0.000049	0.000252	0.000019	10	rb
4	0.000640	0.000034	0.000290	0.000014	10	linea
2	0.000668	0.000023	0.000318	0.000018	1	linea
0	0.000746	0.000053	0.000320	0.000031	0.5	linea
3	0.000738	0.000040	0.000332	0.000022	1	rb
1	0.000827	0.000016	0.000372	0.000015	0.5	rb

Try it yourself

How many samples would it take for accuracies of 85% and 93% to be statistically significantly different for 95% confidence interval?

Class 32: Intro to NLP

1. say hello on zoom
2. share a sentence on the doc linked on prismia

```
import numpy as np
```

Reviewing Confidence Intervals

```
# %load http://drsmb.co/310
def classification_confint(acc, n):
    '''
    Compute the 95% confidence interval for a classification problem.
      acc -- classification accuracy
      n   -- number of observations used to compute the accuracy
    Returns a tuple (lb,ub)
    '''
    interval = 1.96*np.sqrt(acc*(1-acc)/n)
    lb = max(0, acc - interval)
    ub = min(1.0, acc + interval)
    return (lb,ub)
```

If you trained to classifiers on the same data and evaluated on 50 test samples, to get accuracies of 78% and 90% is the difference significant?

To check, we compute the confidence interval for each.

```
classification_confint(.78,50)
```

```
(0.6651767828355258, 0.8948232171644742)
```

```
classification_confint(.9,50)
```

```
(0.816844242532462, 0.983155757467538)
```

Then we check to see if the intervals overlap. They do, so these are not significantly different.

This means that while those seem meaningfully different, with 50 samples, 78% vs 50% is not statistically significantly different. This means that we can't formally guarantee that the two classifiers have reliablly different perforamnce.

If we had more samples, it could be, for example, for 200 samples we see that they are different.

```
N =200
classification_confint(.9,N)
```

```
(0.8584221212662311, 0.941577878733769)
```

```
classification_confint(.78,N)
```

```
(0.722588391417763, 0.8374116085822371)
```

Natural Language Processing

The first thing we need to do to be able to do to model text is transform to a numerical representation. We can't use any of the models we've seen so far, or other models, on non numerical data.

terms:

- document: unit of text we're analyzing (one sample)
- token: sequence of characters in some particular document that are grouped together as a useful semantic unit for processing (basically a word)
- stop words: no meaning, we don't need them (like a, the, an,). Note that this is context dependent. [more info]

Representation

vector or bag of words, implemented by the `CountVectorizer`

Some sample text:

```
# %load http://drsmb.co/310
text = {
'Demeus Alves':'Hope everybody is staying safe',
'Ryan Booth':'The power is out where I live, might be forced to leave soon',
'Brianna MacDonald':'Rainy days',
'Jair Delgado':'Can not wait for lunch... hungry',
'Shawn Vincent':'I am excited for Thanksgiving',
'Jacob Afonso':'Short weeks are the best!',
'Ryan Buquicchio':'The sentence is sentence. (Best sentence ever)',
'Nick McCaffery':'Very windy today',
'David Perrone':'this is a sentence',
'Masoud':'It is rainy here. What about there?',
'Rony Lopes':'I get to relax later this week',
'Patrick Dowd':'It is cold out today',
'Ruifang Kuang':'Happy Thanksgiving!',
}
```

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import euclidean_distances
```

Let's try it on one:

```
s1 = text['Demeus Alves']
s1
```

```
'Hope everybody is staying safe'
```

first we initalize the object

```
counts = CountVectorizer()
```

Then we can fit and transform at once, this will build the representation and return the input represented that way.

```
counts.fit_transform([s1])
```

```
<1x5 sparse matrix of type '<class 'numpy.int64'>'
        with 5 stored elements in Compressed Sparse Row format>
```

It tells us the size and that it's a "sparse matrix" but that doesnt' display much more To see more we can cast it to a regular array

```
counts.fit_transform([s1]).toarray()
```

```
array([[1, 1, 1, 1, 1]])
```

This doesn't tell us much because this is all ones.

Or look at the "vocabulary" also called the "dictionary" for the whole representation

```
counts.vocabulary_
```

```
{'hope': 1, 'everybody': 0, 'is': 2, 'staying': 4, 'safe': 3}
```

We can instead apply to the whole dataset.

```
counts.fit_transform(text.values())
```

```
<13x48 sparse matrix of type '<class 'numpy.int64'>'
        with 65 stored elements in Compressed Sparse Row format>
```

Now there are more rows (samples/documents) and more columns (words in vocabulary)

```
counts.vocabulary_
```

```
{'hope': 16,
 'everybody': 9,
 'is': 18,
 'staying': 34,
 'safe': 30,
 'the': 36,
 'power': 27,
 'out': 26,
 'where': 46,
 'live': 22,
 'might': 24,
 'be': 3,
 'forced': 12,
 'to': 39,
 'leave': 21,
 'soon': 33,
 'rainy': 28,
 'days': 7,
 'can': 5,
 'not': 25,
 'wait': 42,
 'for': 11,
 'lunch': 23,
 'hungry': 17,
 'am': 1,
 'excited': 10,
 'thanksgiving': 35,
 'short': 32,
 'weeks': 44,
 'are': 2,
 'best': 4,
 'sentence': 31,
 'ever': 8,
 'very': 41,
 'windy': 47,
 'today': 40,
 'this': 38,
 'it': 19,
 'here': 15,
 'what': 45,
 'about': 0,
 'there': 37,
 'get': 13,
 'relax': 29,
 'later': 20,
 'week': 43,
 'cold': 6,
 'happy': 14}
```

We can save the transformed data to a variable

```
mat = counts.fit_transform(text.values()).toarray()
mat
```

```
array([[0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0],
       [0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1,
        1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0,
        0, 0, 1, 0],
       [0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0],
       [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
        0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
        0, 0, 0, 0],
       [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0],
       [0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
        1, 0, 0, 0],
       [0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
        0, 0, 0, 1],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
        0, 0, 0, 0],
       [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
        0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
        0, 1, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0,
        0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1,
        0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
        0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
        0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
        0, 0, 0, 0]])
```

To make it easier to read, we can use a dataFrame

```
import pandas as pd
```

The index is the keys of the dictionary of the sentences. The columns are the words from the vocabulary. The `get_feature_names` method will return them as a sorted list instead of a dictionary with numbers.

```
text_df = pd.DataFrame(data=mat, index = text.keys(),
columns=counts.get_feature_names() )
text_df
```

	about	am	are	be	best	can	cold	days	ever	everybody	...	this	to
Demeus Alves	0	0	0	0	0	0	0	0	0	1	...	0	0
Ryan Booth	0	0	0	1	0	0	0	0	0	0	...	0	1
Brianna MacDonald	0	0	0	0	0	0	0	1	0	0	...	0	0
Jair Delgado	0	0	0	0	0	1	0	0	0	0	...	0	0
Shawn Vincent	0	1	0	0	0	0	0	0	0	0	...	0	0
Jacob Afonso	0	0	1	0	1	0	0	0	0	0	...	0	0
Ryan Buquicchio	0	0	0	0	1	0	0	0	1	0	...	0	0
Nick McCaffery	0	0	0	0	0	0	0	0	0	0	...	0	0
David Perrone	0	0	0	0	0	0	0	0	0	0	...	1	0
Masoud	1	0	0	0	0	0	0	0	0	0	...	0	0
Rony Lopes	0	0	0	0	0	0	0	0	0	0	...	1	1
Patrick Dowd	0	0	0	0	0	0	1	0	0	0	...	0	0
Ruifang Kuang	0	0	0	0	0	0	0	0	0	0	...	0	0

13 rows × 48 columns

To compute the distances we use the `euclidean_distances` function. To make this easy to read, we will put this in a dataframe as well.

```
dist_df = pd.DataFrame(data = euclidean_distances(text_df),
                       index=  text.keys(), columns= text.keys())
dist_df
```

	Demeus Alves	Ryan Booth	Brianna MacDonald	Jair Delgado	Shawn Vincent	Jacob Afonso	Ryan Buquicchio
Demeus Alves	0.000000	3.872983	2.645751	3.316625	3.000000	3.162278	4.000000
Ryan Booth	3.872983	0.000000	3.741657	4.242641	4.000000	3.872983	4.582576
Brianna MacDonald	2.645751	3.741657	0.000000	2.828427	2.449490	2.645751	3.872983
Jair Delgado	3.316625	4.242641	2.828427	0.000000	2.828427	3.316625	4.358899
Shawn Vincent	3.000000	4.000000	2.449490	2.828427	0.000000	3.000000	4.123106
Jacob Afonso	3.162278	3.872983	2.645751	3.316625	3.000000	0.000000	3.741657
Ryan Buquicchio	4.000000	4.582576	3.872983	4.358899	4.123106	3.741657	0.000000
Nick McCaffery	2.828427	3.872983	2.236068	3.000000	2.645751	2.828427	4.000000
David Perrone	2.449490	3.605551	2.236068	3.000000	2.645751	2.828427	2.828427
Masoud	3.162278	4.123106	2.645751	3.605551	3.316625	3.464102	4.242641
Rony Lopes	3.316625	4.000000	2.828427	3.464102	3.162278	3.316625	4.358899
Patrick Dowd	2.828427	3.605551	2.645751	3.316625	3.000000	3.162278	4.000000
Ruifang Kuang	2.645751	3.741657	2.000000	2.828427	2.000000	2.645751	3.872983

How can we find who's sentence was most similar to Masoud's?

We can select his column and take the min.

```
dist_df['Masoud'].min()
```

```
0.0
```

But this will return zero, because it's the distance to the same sentence, so we can drop that row of the column

```
dist_df['Masoud'].drop('Masoud')
```

```
Demeus Alves         3.162278
Ryan Booth           4.123106
Brianna MacDonald    2.645751
Jair Delgado         3.605551
Shawn Vincent        3.316625
Jacob Afonso         3.464102
Ryan Buquicchio      4.242641
Nick McCaffery       3.162278
David Perrone        2.828427
Rony Lopes           3.605551
Patrick Dowd         2.828427
Ruifang Kuang        3.000000
Name: Masoud, dtype: float64
```

Then min gives us the the value that's the minumum.

```
dist_df['Masoud'].drop('Masoud').min()
```

```
2.6457513110645907
```

We can use idx min instead.

```
dist_df['Masoud'].drop('Masoud').idxmin()
```

```
'Brianna MacDonald'
```

Try it yourself

1. Which two people wrote the most similar sentences?
2. Using the feature space defined by the text above, what would the following sentence be as a vector?

- "Thanksgiving is a short week"?
- "Rainy, windy days are cold"

1. What word was used the most in the whole set of sentences?

Class 33: Tools, Workflow & more NLP

Review

sklearn estimators have: fit, predict and score methods

split, train and test are the *steps* we do, but not the names of methods of objects in `sklearn`

matplotlib vs seaborn:

We've learned two plotting libraries. Why would you use seaborn over matplotlib?

- seaborn works better when the data is in a dataframe while matplotlib can work better with data in array form
- Matplotlib is more basic plotting, whereas seaborn has less syntax but more plotting customization and themes
- Seaborn uses simpler syntax and easier to make complex intuitive plots with short commands

- seaborn is a higher level library that uses matplotlib “under the hood”. It’s basically an easier way to graph, but directly using matplotlib can give us more freedom, but forces you to make all of the decisions.
- seaborn helps make “good” types of plots, both common and easy for people, on average, to read. Matplotlib, lets you do anything you want, even things that are likely to be confusing
- seaborn can show some statistical calculations while plotting
- seaborn makes categories easier to show, with for example the `hue`, `row`, and `column` parameters

Vector representation review

Given this Vocabulary:

[‘and’, ‘are’, ‘cat’, ‘cats’, ‘dogs’, ‘pets’, ‘popular’, ‘videos’]

represent: Cats and dogs are pets

Since we have the vocabulary, we can go word by word in the vocabulary and [1,1,0,1,1,1,0,0]

build the vocabulary and transform with “Cats and dogs are pets” [1,1,1,1,1]

Classification with text

```
# %load http://drsmb.co/310
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import euclidean_distances
from sklearn import datasets
import pandas as pd
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
```

```
ng_X,ng_y = datasets.fetch_20newsgroups(categories =['comp.graphics','sci.crypt'],
                                        return_X_y = True)
```

```
ng_X[0], ng_y[0]
```

("From: robert@cpuserver.acsc.com (Robert Grant)\nSubject: Virtual Reality for X on the CHEAP!\nOrganization: USCACSC, Los Angeles\nLines: 187\nDistribution: world\nReply-To: robert@cpuserver.acsc.com (Robert Grant)\nNNTP-Posting-Host: cpuserver.acsc.com\n\nHi everyone,\n\nI thought that some people may be interested in my VR\nsoftware on these groups:\n\n*******Announcing the release of Multiverse-1.0.2*******\n\nMultiverse is a multi-user, non-immersive, X-Windows based Virtual Reality\nsystem, primarily focused on entertainment/research.\n\nFeatures:\n\n Client-Server based model, using Berkeley Sockets.\n No limit to the number of users (apart from performance).\n Generic clients.\n Customizable servers.\n Hierachical Objects (allowing attachment of cameras and light sources).\n Multiple light sources (ambient, point and spot).\n Objects can have extension code, to handle unique functionality, easily\n   attached.\n\nFunctionality:\n\n Client:\n The client is built around a 'fast' render loop. Basically it changes things\n when told to by the server and then renders an image from the user's\n viewpoint. It also provides the server with information about the user's\n actions - which can then be communicated to other clients and therefore to\n other users.\n\n The client is designed to be generic - in other words you don't need to\n develop a new client when you want to enter a new world. This means that\n resources can be spent on enhancing the client software rather than adapting\n it. The adaptations, as will be explained in a moment, occur in the servers.\n\n This release of the client software supports the following functionality:\n\n   o Hierarchical Objects (with associated addressing)\n\n   o Multiple Light Sources and Types (Ambient, Point and Spot)\n\n   o User Interface Panels\n\n   o Colour Polygonal Rendering with Phong Shading (optional wireframe for\n\tfaster frame rates)\n\n   o Mouse and Keyboard Input\n\n (Some people may be disappointed that this software doesn't support the\n PowerGlove as an input device - this is not because it can't, but because\n I don't have one! This will, however, be one of the first enhancements!)\n\n Server(s):\n This is where customization can take place. The following basic support is\n provided in this release for potential world server developers:\n\n   o Transparent Client Management\n\n   o Client Message Handling\n\n This may not sound like much, but it takes away the headache of\naccepting and\n terminating clients and receiving messages from them - the\napplication writer\n can work with the assumption that things are happening locally.\n\n Things get more interesting in the object extension functionality. This is\n what is provided to allow you to animate your objects:\n\n   o Server Selectable Extension Installation:\n   What this means is that you can decide which objects have extended\n   functionality in your world. Basically you call the extension\n   initialisers you want.\n\n   o Event Handler Registration:\n When you develop extensions for an object you basically write callback\n functions for the events that you want the object to respond to.\n   (Current events supported: INIT, MOVE, CHANGE, COLLIDE & TERMINATE)\n\n   o Collision Detection Registration:\n   If you want your object to respond to collision events just provide\n   some basic information to the collision detection management software.\n   Your callback will be activated when a collision occurs.\n\n   This software is kept separate from the worldServer applications because\n   the application developer wants to build a library of extended objects\n from which to choose.\n\n   The following is all you need to make a World Server application:\n\n   o Provide an initWorld function:\n   This is where you choose what object extensions will be supported, plus\n   any initialization you want to do.\n\n   o Provide a positionObject function:\n   This is where you determine where to place a new client.\n\n   o Provide an installWorldObjects function:\n   This is where you load the world (.wld) file for a new client.\n\n o Provide a getWorldType function:\n   This is where you tell a new client what persona they should have.\n\n   o Provide an animateWorld function:\n   This is where you can go wild! At a minimum you should let the objects\n   move (by calling a move function) and let the server sleep for a bit\n   (to avoid outrunning the clients).\n\n   That's all there is to it! And to prove it here are the line counts for the\n   three world servers I've provided:\n\n   generic - 81 lines\n   dactyl - 270 lines (more complicated collision detection due to the\n   stairs! Will probably be improved with future\n versions)\n   dogfight - 72 lines\n\nLocation:\n\n This software is located at the following site:\n ftp.u.washington.edu\n\n Directory:\n pub/virtual-worlds\n\n File:\n multiverse-1.0.2.tar.Z\n\nFutures:\n\n Client:\n\n   o Texture mapping.\n\n   o More realistic rendering: i.e. Z-Buffering (or similar), Gouraud shading\n\n   o HMD support.\n\n   o Etc, etc....\n\n Server:\n\n   o Physical Modelling (gravity, friction etc).\n\n   o Enhanced Object Management/Interaction\n\n   o Etc, etc....\n\n Both:\n\n   o Improved Comms!!!\n\nI hope this provides people with a good understanding of the Multiverse\nsoftware,\nunfortunately it comes with practically zero documentation, and I'm not sure\nwhether that will ever be able to be rectified! :-(\n\nI hope people enjoy this software and that it is useful in our explorations of\nthe Virtual Universe - I've certainly found fascinating developing it, and I\nwould *LOVE* to add support for the PowerGlove...and an HMD :-)!!\n\nFinally one major disclaimer:\n\nThis is totally amateur code. By that I mean there is no support for this code\nother than what I, out the kindness of my heart, or you, out of pure\ndesperation, provide. I cannot be held responsible for anything good or bad\nthat may happen through the use of this code - USE IT AT YOUR OWN RISK!\n\nDisclaimer over!\n\nOf course if you love it, I would like to here from you. And anyone with\nPOSITIVE contributions/criticisms is also encouraged to contact me. Anyone who\nhates it: > /dev/null!\n\n*****\n*********\nAnd if anyone wants to let me do this for a living: you know where to\nwrite :-)!\n*****\n*****\n\nThanks,\n\nRobert.\n\nrobert@acsc.com\n^^^^^^^^^^^^^^^\n",
 0)

```
counts = CountVectorizer()
ng_vec = counts.fit_transform(ng_X)
```

```
ng_vec
```

```
<1179x24257 sparse matrix of type '<class 'numpy.int64'>'
        with 188291 stored elements in Compressed Sparse Row format>
```

```
ng_vec[:1].toarray()
```

```
array([[0, 0, 0, ..., 0, 0, 0]])
```

```
clf = MultinomialNB()
```

```
ng_vec_train, ng_vec_test, ng_y_train, ng_y_test = train_test_split(ng_vec,ng_y)
```

```
clf.fit(ng_vec_train,ng_y_train).score(ng_vec_test,ng_y_test)
```

```
0.9932203389830508
```

Class : More Representations of Text

[[1],[1],[1],[1],[1]]

```
# %load http://drsmb.co/310
from sklearn.feature_extraction import text
from sklearn.metrics.pairwise import euclidean_distances
from sklearn import datasets
import pandas as pd
from sklearn.naive_bayes import MultinomialNB
import numpy as np
```

```
ng_X,ng_y = datasets.fetch_20newsgroups(categories =['comp.graphics','sci.crypt'],
                                        return_X_y = True)
```

```
count_vec = text.CountVectorizer()
```

```
ng_X_vec = count_vec.fit_transform(ng_X)
```

```
ng_X_vec.toarray()[:3]
```

```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [1, 1, 0, ..., 0, 0, 0]])
```

```
ng_X_vec.shape
```

```
(1179, 24257)
```

```
np.max(ng_X_vec)
```

```
549
```

```
tfidf = text.TfidfTransformer()
```

```
ng_X_tfidf = tfidf.fit_transform(ng_X_vec)
```

```
ng_X_tfidf.toarray()[:3]
```

```
array([[0.        , 0.        , 0.        , ..., 0.        , 0.        ,
        0.        ],
       [0.        , 0.        , 0.        , ..., 0.        , 0.        ,
        0.        ],
       [0.05125476, 0.0576842 , 0.        , ..., 0.        , 0.        ,
        0.        ]])
```

```
ng_X_tfidf.shape
```

```
(1179, 24257)
```

```
np.max(ng_X_tfidf)
```

```
0.938424962381379
```

Does this representation improve our classification for this task?

```
counts_bigram = text.CountVectorizer(ngram_range = (2,2))
counts_bigram.fit_transform(ng_X)
```

```
<1179x149885 sparse matrix of type '<class 'numpy.int64'>'
        with 313557 stored elements in Compressed Sparse Row format>
```

```
counts_bigram = text.CountVectorizer(ngram_range = (2,2),stop_words = 'english')
counts_bigram.fit_transform(ng_X)
```

```
<1179x125898 sparse matrix of type '<class 'numpy.int64'>'
        with 192349 stored elements in Compressed Sparse Row format>
```

Assignments

All assignments are due on Sunday at 11:59pm, via github unless otherwise noted.

Assignment TOC:

- Assignment 1 Due September 13
- Assignment 2 Due September 20
- Assignment 3 Due September 29
- Assignment 4 Due October 4
- Assignment 5 Due October 11
- Assignment 6 Due October 20
- Assignment 7 Due October 27
- Assignment 8 Due November 3
- Assignment 9 Due November 10
- Assignment 10 Due November 17
- Assignment 11 Due November 24
- Assignment 12 Due December 1

Assignment 1: Portfolio Setup, Data Science, and Python

Due: 2020-09-13

Objective & Evaluation

This assignment is an opportunity to earn level 2 achievements for the `process` and `python` and confirm that you have all of your tools setup, including your portfolio.

To Do

Your task is to:

1. Install required software
2. Setup your portfolio, by accepting the assignment and following the instructions in the README file on your repository.
3. Add your own definition of data science to the introduction of your portfolio, in `about/index.md`
4. Add a Jupyter notebook called `grading.ipynb` to the `about` folder and write a function that computes a grade for this course, with the following docstring. Include:

- a Markdown cell with a heading
- your function called `compute_grade`
- three calls to your function that verify it returns the correct value for different number of badges that produce at three different letter grades.

1. Uncomment the line `#  - file: about/grading` in your `_toc.yml` file.

> **Note**
>
> If you get stuck on any of this after accepting the assignment and creating a repository, you can create an issue on your repository, describing what you're stuck on and tag us with `@rhodyprog4ds/fall20instructors`.
>
> To do this click Issues at the top, the green "New Issue" button and then type away.

```
'''
Computes a grade for CSC/DSP310 from numbers of achievements at each level

Parameters:
------------
num_level1 : int
  number of level 1 achievements earned
num_level2 : int
  number of level 2 achievements earned
num_level3 : int
  number of level 3 achievements earned

Returns:
--------
letter_grade : string
  letter grade with possible modifier (+/-)
'''
```

Here are some sample tests you could run to confirm that your function works correctly:

```
assert compute_grade(15,15,15) == 'A'

assert compute_grade(15,15,13) == 'A-'

assert compute_grade(15,14,14) == 'B-'

assert compute_grade(14,14,14) == 'C-'

assert compute_grade(4,3,1) == 'D'

assert compute_grade(15,15,6) ==='B+'
```

> **Warning**
>
> your function can have a different name than `compute_grade`, but make sure it's your function name, with those parameter values in your tests.

> **Note**
>
> when the value of the expression after `assert` is `True`, it will look like nothing happened. `assert` is used for testing

Submission Instructions

Create a Jupyter Notebook with your function and Add the notebook to your portfolio by uploading it to your repository, or adding to the folder off line and committing and pushing the changes.

View the `gh-pages` branch to see your compiled submission, as `portfolio.pdf` or by viewing your website.

There will be a pull request on your repository that is made by GitHub classroom, request a review from the team `rhodyprog4ds/Fall20instructors`.

Solutions

One solution is added to the Detailed Mechanics part of the Grading section of the syllabus.

Assignment 2: Practicing Python and Accessing Data

Due: 2020-09-20

Objective & Evaluation

This assignment is an opportunity to earn level 1 or 2 achievements in `python`, `process` and `access` and begin working toward level 1 in `summarize`.

Accept the assignment on GitHub Classroom. It contains a notebook with some template structure (and will set you up for grading). The template will also convert notebooks that are added to markdown, which makes reading on GitHub for easier grading. If you want to incorporate feedback you receive back into a notebook file, Jupytext can do that.

To work with this notebook you can either:

- download the repository as .zip from the green code button, unzip, and re-upload, OR
- clone the repository with git and the push your changes. See Git/GitHub help on cloning, committing, and pushing, for example this tutorial on git to learn more about git.

Accessing Data with Python and pandas

(for python and access)

Find 3 datasets of interest to you that are provided in different file formats. Choose datasets that are not too big, so that they do not take more than a few second to load. At least one dataset, must have non numerical (eg string or boolean) data in at least 1 column. Complete a dictionary for each with the url, a name, and what function should be used to load the data into a `pandas.DataFrame`.

Use a list of those dictionaries to iterate over the datasets and build a table that describes them, with the following columns `['name','source','num_rows', 'num_columns','source_file_name']`. The source column should be the url where you loaded the data from or the source if you downloaded it from a website first The `source_file_name` should be the part of the url after the last `/`, you should extract this programmatically. Display that summary table as a dataframe and save it as a csv, named `dataset_summary.csv`.

> **Tip**

> **Tip**
>
> Urls are strings. The `string` class in python has a lot of helpful methods for manipulating strings, like `split`.

> **Note**
>
> If you download the datasets (or find them as .zip and need to) you can use the local path instead of the url, but include a markdown cell with links to where you got your data from.

For one dataset (must include nonnumerical data):

- display the heading with the last seven rows
- make and display a new data frame with only the non numerical columns
- was the format that the data was provided in a good format? why or why not?

> **Tip**

> **Tip**
>
> You can create a `pandas DataFrame` using the constructor and you can build lists (or lists of lists) using the `append` method

For a second dataset:

- display the heading and the first three rows
- display the datatype for each column
- Are there any variables where pandas may have read in the data as a datatype that's not what you expect (eg a numerical column mistaken for strings)?

For the third dataset:

- display the first 5 even rows of the data for three columns of your choice

For any dataset:

- try reading it in with the wrong `read_` function. If you had done this by accident, how could you tell?

Data Science Process

(for the `process` skill)

Make a list of a data science pipeline and denote which types of programming might be helpful at each staged. Include this in a markdown cell in the same notebook with your analysis.

> **Tip**
>
> Remember that this will be graded based on the rubric and that it should reflect your understanding, not be simply copied from a source. Also, always cite your sources, informal linking is ok.
>
> To make a link in markdown
>
> ```
> [text to display](http://url.com/of/the/site/your/are/linking/)
> ```

Assignment 3: Exploratory Data Analysis

Due: 2020-09-27

Objective & Evaluation

This assignment is an opportunity to earn level 1 or 2 achievements in `summarize`, `visualize`, or `access`. You can earn level 2 in `python`.

Accept the assignment on GitHub Classroom. The template will convert notebooks that are added to markdown, which makes reading on GitHub for easier grading. It will sync between .ipynb and .md style notebooks stored in your repository.

This week I encourage you to try working with git, but if you're not comfortable with that you can work via upload again.

Exploratory Data Analysis

This week your goal is to do a small exploratory data analysis for two datasets of your choice. One dataset must include at least two continuous valued variables and at least one categorical variable(d1). One dataset must include at least two categorical variables and at least one continuous valued variable(d2).

> **Tip**
>
> A continuous valued variable is one that can take on infinitely many values. Examples include temperatue, length, age, etc. A categorical value is one that can only take on one of a subset of specific values. These require a bit more context to define often. For example color could be recorded in a basically continuous way as RGB values in HEX, but color of the stoplight will only be red, green, or yellow. Town of birth in a dataset of celebrities would not be a good categorical variable, there might not be many duplicates to do analysis with, but metro area of current residence for celebrities would be because we'd see mostly major cities and could do analyses.

Use a separate notebook for each dataset, name them `dataset_01.ipynb` and `dataset_02.ipynb`.

For each dataset:

1. Include a markdown header with a title for your analysis
2. Load the data to a notebook as a `DataFrame` from url.
3. Explore the dataset in a notebook enough to describe its structure
    - shape
    - columns
    - variable types
4. Write a short description of what the data contains and what it could be used for
5. Complete an exploratory analysis with statistics and plots. Your analysis should include markdown cells describing the results you see, not only what you did. Your analysis should be structured to follow the steps below, for the corresponding dataset.

For d1:

1. Display all of the summary statistics for a subset of 5 of your choice or all variables if there are fewer than 5 numerical values
2. Display all of summary statistics grouped by a categorical variable
3. For two continuous variables make a scatter plot and color the points by a categorical variable
4. Pose one question for this dataset that can be answered with summary statistics, compute a statistic and plot that help answer that exploratory question.

For d2:

1. Display two individual summary statistics for one variable
2. Group the data by two categorical variables and display a table of one summary statistic
3. Use a seaborn plotting function with the `col` parameter or a `FacetGrid` to make a plot that shows something informative about this data, using both categorical variables and at least one numerical value. Describe what this tells you about the data.
4. Produce one additional plot of a different plot type that shows something about this data.

> **Tip**
>
> if you use a local copy of the dataset you can load from a relative path instead of a url. Please include the data in a data folder that's in the folder where your notebook is so that the path is something like: `data/best-dataset-ever.csv` and upload the data to your assignment repository as well.

In total, in each of your notebooks you will have:

- a loaded dataset
- a basic description
- at least two summary statistic calculations
- at least two plots

> **Tip**
>
> Recall how we used a single function to see many statistics in class, use that for dataset 1. For dataset 2, you can make your own combination of summary statistics with `pandas.DataFrame.agg`

Assignment 4: Preparing Data for Analysis

Due: 2020-10-04

Objective & Evaluation

You can earn prepare level 2 by cleaning two datasets as outlined in the `cs_degrees.ipynb` and `travel_time.ipynb` notebooks.

To earn access level 2 you must clean a third dataset, following the `airlines.ipynb` or `safi_full.ipynb` notebooks. Note that this is the SAFI dataset we've been working with in class, but a messier version of it. You know what some of the fixes are because they're in the notes and some of what it will look like when it's done.

To earn python level 2, you must complete `safi_full.ipynb`. Specifically, you must successfully use list or dictionary comprehensions and conditional statements, beyond those provided as hints.

To earn level 2 for summarize and visualize, include additional analyses after cleaning the datasets. You'll need at least two types of plots and two different ways of using summary statistics and to interpret the results.

Instructions

For this assignment, your assignment is to clean datasets in notebooks and include narrative description of how you're making decisions about the data cleaning. At the end of each cleaning, save the cleaned dataset to the data folder with an informative file name.

When you accept the assignment, there will be template notebooks in the repository. These have significant hints to guide your efforts.

> **Tip**
>
> You can download any of the files and open them with another program if you need to by pasting the url in your browser. You'll need to display relevant parts in a notebook for grading, but if looking another way helps, you can try that.

There is an issue type with a todo list, try that out to plan and track questions you have.

You can work with this assignment in 2 ways:

Git Workflow

1. Click the green code button and copy the url.
2. clone the repository as below where th eurl is the one you copied from GitHub. In your terminal on Linux or Mac or on the GitBash on Windows (install instructions on tools section of syllabus)

```
cd path/where/you/want/a/new/folder/for/this/assignment
git clone http://github.com/rhodyprog4ds/04-prepare-username
```

1. then launch your notebook from the newly created folder.

on Linux or Mac, in the same terminal (remember you can use tab complete)

```
cd 04-prepare-username
jupyter notebook
```

or on windows, in your anaconda prompt

```
cd path/where/you/want/a/new/folder/for/this/assignment/04-prepare-username
jupyter notebook
```

1. work on your assignment, by opening the notebooks there and editing them.

2. commit your changes when you want to save a point in your progress. Either you have part workign and want to save that, or you want feedback, or you're done. On whichever terminal you used for the `git clone` command above

```
git add .
git commit -m 'description of current status of project'
```

1. push your changes when you want to share them on GitHub, (eg need help, want feedback or complete)

```
git push
```

1. if you will keep working after pushing, first pull down the .md conversion that was added by GitHub actions. If you don't you'll have to merge, it should be fine, but ask if you're not sure.

```
git pull
```

Download and upload workflow

1. Click the green code button
2. choose the download via zip option
3. unzip
4. launch a notebook in the folder where you unzipped
5. Upload only the files you changed into the repository to replace their previous versions with the add file via upload button on GitHub. Put the notebooks at the top level and the data files in the data folder.

Tips and hints

- Pandas can convert string to dates to be able to work with them more flexibly
- With python date objects you can get the day of the week from a date
- Pandas `read_excel` has:
  - a `header` parameter can take a list as its value
  - two parameters that allow you to skip rows when you read in the data
- `list(range(7))` check out what this line does
- Pandas drop can work a lot of different ways

The goal of the plot for the CS degrees dataset is:

Assignment 5: Constructing Datasets and Using Databases

Due: 2020-10-11

there will be no tolerance of late submissions on this assignment because it will be graded Monday, so that you can use that feedback to finish up your portfolio.

Accept the assignment

Instructions

this will earn level 2 for construct

Your goal is to build and prepare two ready to analyze datasets. You will submit a notebook that describes how you built and prepared each dataset.

- Each finished dataset must be produced from two or more DataFrames.
- At least one must come from an sqlite database, either by merging results from multiple queries or multiple tables.
- You should use at least two different merges and one concatenate

Your completed datasets should have:

- column names that are well formatted (only lowercase letters, numbers and _)
- an added column that is derived from one or more other columns (string operation or calculation)

For each dataset, pose one question that could not be directly answered from the input data files as provided and demonstrate how to answer it with the dataset you built. This could be something that can be answered with using only `shape` of the merged data, but if you need summarize and visualize level 2 achievements, you should use more statistics and plots.

Your notebooks must be in the top level of the repository, not in a subfolder.

Additional Achievements:

if you already earned prior achievements you can ignore the following

To earn level 2 for prepare, one of your analyses must use datasets with missing values and one must be provided as excel files with merged columns (for example from NCES). You may use one dataset with both merged columns and missing data or one of each. You must also use datasets that have column names that need repair.
For the merged column data, either before or after merging, you must additionally:

- create separate separate tables for original and aggregate values (eg percentages or sums that can be recovered from the other columns)
- unstack all levels of the data to create a single level index over the columns

To earn level 2 for summarize and/or visualize, include additional analyses after building the datasets. You'll need at least two types of plots for visualize and/or two different ways of using summary statistics for summarize and to interpret the results for either.

Assignment 6: Naive Bayes

Due: 2020-10-20

Remember, even if you aren't sure how to do every part of the assignment, submitting pseudocode, a list of your questions indicating where you got stuck, or a partial solution will get you personalized feedback. You might even earn level 1 achievements for the partial effort.

For Classification Level 2:

Acccept the Assignment

Create one notebook where you load examine each of the provided datasets for for suitability to use with Gaussian Naive Bayes. Label the sections `Dataset #`. Use exploratory data analysis (visualisation and statistics) in pandas and Gaussian Naive Bayes from scikit learn to answer the following for each dataset:

1. Do you expect Gaussian Naive Bayes to work well on this dataset, why or why not? (think about the assumptions of naive bayes and classification in general) *explanation is essential here, because you can actually use the classifier to check*
2. How well does a Gaussian Naive Bayes classifier work on this dataset? Do you think a different classifier might work better or do you think this data cannot be predicted any better than this?
3. How does the actual performance compare to your prediction? If it performs much better or much worse than you expected, what might you use to figure out why? (*you do not have to figure out why your predictions were not correct, just list tools you've learned in class that might help you figure that out*)

This assignment will be easiest if you take advantage of the template via git:

Git Workflow

1. Click the green code button and copy the url.
2. clone the repository as below where th eurl is the one you copied from GitHub. In your terminal on Linux or Mac or on the GitBash on Windows ([install instructions on tools section of syllabus](#))

```
cd path/where/you/want/a/new/folder/for/this/assignment
git clone http://github.com/rhodyprog4ds/06-classification-username
```

1. then launch your notebook from the newly created folder.

on Linux or Mac, in the same terminal (remember you can use tab complete)

```
cd 06-classification
jupyter notebook
```

or on windows, in your Anaconda prompt

```
cd path/where/you/want/a/new/folder/for/this/assignment/06-classification-username
jupyter notebook
```

1. work on your assignment, by opening the notebooks there and editing them.
2. commit your changes when you want to save a point in your progress. Either you have part workign and want to save that, or you want feedback, or you're done. On whichever terminal you used for the `git clone` command above

```
git add .
git commit -m 'description of your changes since las commit'
```

1. push your changes when you want to share them on GitHub, (eg need help, want feedback or complete)

```
git push
```

1. if you will keep working after pushing, first pull down the .md conversion that was added by GitHub actions. If you don't you'll have to merge, it should be fine, but ask if you're not sure.

```
git pull
```

For construct level 2

Build your own dataset for classification by merging data from separate sources that you're interested in.

Summarize and Viz level 2

Show some extra summary stats and plots

Assignment 7: Decision Trees

Due: 2020-10-27

Decision Trees

[Accept this repo for submission](#)

Choose a datasets that is well suited for classification and that has all numerical features. (eg the Wisconsin Breast Cancer data from UCI)

Part 1: DT Basics

1. Include a basic description of the data(what the features are)

2. Write your own description of what the classification task is and why a decision tree is a reasonable model to try for this data.
3. Include one summary visualization of the data.
4. Fit a decision tree with the default parameters on 50% of the data
5. Test it on 50% held out data and generate a classification report
6. Inspect the model by visualizing and interpreting the results

- Does this model make sense?
- Are there any leaves that are vey small?
- Is this an interpretable number of levels?

1. Repeat with the entropy criterion. Does using the entropy criterion make a big difference or small difference in the overall classifier?

> **Tip**
>
> See the documentation for the sklearn `DecisionTreeClassifier`. One of the parameters is called `criterion`

Part 2: DT parameters

Do an experiment to see how `max_depth`, `min_values_split`, or `min_values_leaf` impacts the model.

1. Choose one of these and say explain why and how you hypothesize it will impact the performance
2. Use the model you fit above and EDA to choose minimum and maximum values for your parameter. Choose a total of 3 values for the parameter.
3. Retrain the model for each value of the parameter
4. Test and use at least 3 metrics to describe the performance, compiling your results into a DataFrame
5. Plot and interpret your results

you should use a loop for this part

Part 3: Test and Train Sizes

Do an experiment to compare test set size vs performance:

1. Train a decision tree on 20%, 30%, … , 80% of the data, using one of the training parameter combinations you tried above and explain why you chose the one you chose.
2. Save the results of both test and train accuracy for each size training data in a DataFrame with columns ['train_pct','n_train_samples','n_test_samples','train_acc','test_acc']
3. Plot the accuracies vs training percentage.
4. Explain these results. What is the best test/train split. Why?

you should use a loop for this part

Evaluation

> **Tip**
>
> Start thinking about level 3 for evaluate by trying this experiment with cross validation.

This assignment's focus is Classification and Evaluation:

- part 1 will show that you understand classification and can apply it for level 2
- parts 2 & 3 are opportunities to earn level 2 for evaluation.
- If you don't successfully complete the whole assignment, pseudocode or partial answers to the questions could earn you level 1 for either skill.

Additionally:

- Using functions, loops, and conditionals appropriately while completing those tasks could earn level 2 for python if needed.
- You can earn level 2 for viz, summarize, prepare, or construct by choosing a messy dataset and/or including extra exploratory data analysis.

FAQ

How do I find a good dataset?

Look for a dataset with numerical features and a categorical target variable.

If you look at the UCI website you can search for datasets for Classification and numerical and look through what those search filters give you some options. You might want to choose a dataset with less than about 10 attributes to make your decision tree readable. To make the training fast, try to find a dataset with 1000 samples or less.

Assignment 8: Linear Regression

Due: 2020-11-03

Linear Regression

Find a dataset suitable for regression. We recommend a dataset from the UCI repository (see below)

Fit a linear regression model, measure the fit with two metrics, and make a plot that helps visualize the result.

Examine the coefficients or the residuals to try to interpret the results and explain what this regression model is doing.

Try fitting the model only on one feature. Justify your choice of feature based on the results above. Plot this result.

Accept the assignment to create your submission repository

Part 2: Test Train Splits

if you successfully completed this experiment in assignment 7, you don't need to repeat it, but it might be interesting to try

Do an experiment to compare test set size vs performance:

1. Train a the linear regression model on 20%, 30%, … , 80% of the data.
2. Save the results of both test and train accuracy for each size training data in a DataFrame with columns ['train_pct','n_train_samples','n_test_samples','train_acc','test_acc']
3. Plot the accuracies vs training percentage.
4. Explain these results. What is the best test/train split. Why?

Part 3: Other models

Try fitting LASSO and explaining what it does. Does LASSO make better predictions on your data?

Do you think a model more complex than linear would be better? How could you tell?

Grading

Include description of what you're doing, why you're doing it, and what the results mean at each step of your analysis so that we can tell that you understand.

For regression level 2, complete part 1.

For evaluate level 2, complete part 2.

If you're curious, try a more complex regression model as in part 3. This will get you an early start on level 3 for regression.

If you don't successfully complete the whole assignment, pseudocode or partial answers to the questions could earn you level 1 for either skill.

FAQ

How do I find a good dataset?

Look for a dataset with numerical features and a categorical target variable.

If you look at the UCI website you can search for datasets for Classification and numerical and look through what those search filters give you some options. You might want to choose a dataset with less than about 20 attributes. To make the training fast, try to find a dataset with 1000 samples or less, or use read functions to use only a small chunk of the data.

Remember to read about the dataset and note what you're predicting.

Assignment 9

Due: 2020-11-10

assignment submission

1. Apply Kmeans to the dataset you used for assignment 7.
2. Evaluate how well clustering worked on the data:

- using clustering metrics and
- using visualization.
- the ground truth labels

1. Include a discussion of your results that addresses the following:

- describes what the clustering means
- what the metrics show
- Does this clustering work better or worse than expected based on the classification performance (if you didn't complete assignment 7, also apply a classifier)

Evaluation and Grading

The goal of this is for clustering level 2.

You may also earn evaluation level 2. To do this, use at least two metrics, explain them thoroughly, and use train-test splits.

You can earn visualize and summarize by adding and interpretting visualizations and summary statistics.

As always, psuedo code or an outline can earn level 1

Assignment 10: Optimizing Models

Due: 2020-11-17

Choose a dataset with a modeling task that you are interested in. Choose a plausible model for the dataset and optimize the parameters of the model.

Explain and interpret your results. Keep your notebook focused (don't include things that don't advance the story or go on tangents) and complete (enough detail for someone to learn from your notebook and for us to know you understand)

Classification or regression will be easier to use for this assignment and next week's assignment. Next week, you'll continue with the same dataset and compare the model you optimized this week to another model.

submit via github

This is an opportunity to earn any modeling (classification, regression, or clustering) skill, evaluation, and optimize. If there are older skills that you would like evaluated at this point include a note in a markdown cell at the top of your notebook.

Assignment 11: Model Comparison

Due: 2020-11-24

submission

Choose a dataset, it can be appropriate for classification, regression, or clustering. Fit at least two models for the same task and choose the appropriate metrics to compare the fit. Determine which model fits the data better and write a brief report. Summarize your findings with plots and tables as appropriate.

Explain and interpret your results. Keep your notebook focused (don't include things that don't advance the story or go on tangents) and complete (enough detail for someone to learn from your notebook and for us to know you understand).

This is an opportunity to earn any modeling (classification, regression, or clustering) skill, evaluation, and optimize. If there are older skills that you would like evaluated at this point include a note in a markdown cell at the top of your notebook.

Assignment 12: Fake News

Due: 2020-12-01

Submission Template

For this exercise you will build a classifier that can distinguish real news from fake news. A training set for this is available here: https://raw.githubusercontent.com/lutzhamel/fake-news/master/data/fake_or_real_news.csv

The fields you are interested in are 'text' and 'label' with the obvious interpretations. The data set contains a large number of articles (takes a long time to train), you can downsample this to approximately 1,000 articles in order to speed up training and evaluation (hint: use shuffle).

1. How accurately can you predict real vs fake news from the text?
2. Are titles of real or fake news more similar to one another based on euclidean distance? *for this question, describe what you would need to do to answer it and answer it*

For `unstructured` and `workflow` plan out solutions, determine what tools you'll need and answer the two questions. You can earn unstructured by representing the text for analysis, to earn workflow, you'll need to answer both questions.

Portfolio

This section of the site has a set of portfolio prompts and this page has instructions for portfolio submissions.

Starting in week 3 it is recommended that you spend some time each week working on items for your portfolio, that way when it's time to submit you only have a little bit to add before submission.

The portfolio is your only chance to earn Level 3 achievements, however, if you have not earned a level 2 for any of the skills in a given check, you could earn level 2 then instead. The prompts provide a starting point, but remember that to earn achievements, you'll be evaluated by the rubric. You can see the full rubric for all portfolios in the syllabus. Your portfolio is also an opportunity to be creative, explore things, and answer your own questions that we haven't answered in class to dig deeper on the topics we're covering. Use the feedback you get on assignments to inspire your portfolio.

Each submission should include an introduction and a number of 'chapters'. The grade will be based on both that you demonstrate skills through your chapters that are inspired by the prompts and that your summary demonstrates that you *know* you learned the skills. See the formatting tips for advice on how to structure files.

The third submission will be graded on the following criteria and due on December 4:

	Level 3
keyword	
process	Compare different ways that data science can facilitate decision making
classification	fit and apply classification models and select appropriate classification models for different contexts
regression	can fit and explain regrularized or nonlinear regression
clustering	apply multiple clustering techniques, and interpret results
evaluate	Evaluate a model with multiple metrics and cross validation
optimize	Select optimal parameters based of mutiple quanttiateve criteria and automate parameter tuning
compare	Evaluate tradeoffs between different model comparison types
unstructured	apply multiple representations and compare and contrast them for different end results
workflow	Scope, choose an appropriate tool pipeline and solve data science problems, describe strengths and weakensses of common tools

On each chapter(for a file) of your portfolio, you should identify which skills by their keyword, you are applying.

You can view a (fake) example in this repository as a pdf or as a rendered website

Upcoming Checks

Portfolio 4

For the fourth submission, due December 19, you may earn level 1 for *any skill* an unlimited number, as long as your submission is clear and concise. I recommend that you make a plan early (by updating your `subimssion_4_intro.md` file or making an issue on your repo) and ask for feedback in office hours.

It will be graded primarily on the following criteria:

	Level 3
keyword	
optimize	Select optimal parameters based of mutiple quanttiateve criteria and automate parameter tuning
compare	Evaluate tradeoffs between different model comparison types
unstructured	apply multiple representations and compare and contrast them for different end results
workflow	Scope, choose an appropriate tool pipeline and solve data science problems, describe strengths and weakensses of common tools

You may also earn level 2 for any of those skills and *one additional skill*. For a single skill of your choice that you have attempted at least twice (assignments and portfolios) you may get an additional attempt in check 4. For this case, you must link to your two previous attempts and describe how you approached working on understanding this skill.

Linking in markdown uses `[]` for display text and `()` for the url like:

```
[previous attempt here](https://github.com/rhodyprog4ds/portfolio-brownsarahm/blob/main/check1/loading_data.md)
```

that would render like: [previous attempt here]

You can also use relative paths as in the [example intro file].

Past Checks

Portfolio 2

The second submission will be graded on the following criteria and due on November 13:

	Level 3
keyword	
python	reliable, efficient, pythonic code that consistently adheres to pep8
process	Compare different ways that data science can facilitate decision making
access	access data from both common and uncommon formats and identify best practices for formats in different contexts
construct	merge data that is not automatically aligned
summarize	Compute and interpret various summary statistics of subsets of data
visualize	generate complex plots with pandas and plotting libraries and customize with matplotlib or additional parameters
prepare	apply data reshaping, cleaning, and filtering manipulations reliably and correctly by assessing data as received
classification	fit and apply classification models and select appropriate classification models for different contexts
regression	can fit and explain regrularized or nonlinear regression
clustering	apply multiple clustering techniques, and interpret results
evaluate	Evaluate a model with multiple metrics and cross validation

For this portfolio check I encourage you to dig in on either data from one domain and think about how clustering, regression, and classification apply in this domain OR to inspect these more carefully in a performance prospective, taking a more CS-basics approach.

Formatting Tips

Your portfolio is a jupyter book. This means a few things:

- it uses myst markdown
- it will run and compile Jupyter notebooks

This page will cover a few basic tips.

Organization

The summary of for the `part` or whole submission, should match the skills to the chapters. Which prompt you're addressing is not important, the prompts are a *starting point* not the end goal of your portfolio.

Data Files

Also note that for your portfolio to build, you will have to:

- include the data files in the repository and use a relative path OR
- load via url

using a full local path **will not work** and will render your portfolio unreadable.

Structure of plain markdown

Use a heading like this:

```
# Heading of page
```

in the file and it will appear in the sidebar.

File Naming

It is best practice to name files without spaces. Each `chapter` or file should have a descriptive file name (`with_no_spaces`) and descriptive title for it.

Syncing markdown and ipynb files

To sync feedback received to your runnable notebook files, change the related GitHub Actions file: `.github/workflows/` In the step named convert that looks like:

```
- name: convert
  run: |
      jupytext */*.ipynb --to myst
```

change it to:

```
- name: convert
  run: |
      jupytext --set-formats ipynb,md */*.ipynb  # Turn .ipynb into a paired ipynb/py notebook
      jupytext --sync */*.ipynb                  # Update whichever of .ipynb/notebook.md is
outdated
```

This means if you accept suggestion commits from the the `.md` file, the action will upate your `.ipynb` file. If you update your `.ipynb file` the action will update the .md file.

Adding annotations with formatting or margin notes

You can either install jupytext and convert locally or upload /push a notebook to your repository and let GitHub convert. Then edit the .md file with a text editor of your choice. You can run by uploading if you don't have jupytext installed, or locally if you have installed jupytext or jupyterbook.

In your .md file use backticks to mark special content blocks

```
```{note}
Here is a note!
```
```

```
```{warning}
Here is a warning!
```
```

```
```{tip}
Here is a tip!
```
```

```
```{margin}
Here is a margin note!
```
```

For a complete list of options, see the `sphinx-book-theme` documentation.

Links

Markdown syntax for links

```
[text to show](path/or/url)
```

Configurations

Things like the menus and links at the top are controlled as settings, in `_config.yml`. The following are some things that you might change in your configuration file.

Show errors and continue

To show errors and continue running the rest, add the following to your configuration file:

```
# Execution settings
execute:
  allow_errors              : true
```

Using additional packages

You'll have to add any additional packages you use (beyond pandas and seaborn) to the `requirements.txt` file in your portfolio.

Reflective Prompts

These prompts are more reflective to help demonstrate your understanding of skills. These are more writing than new coding.

Correcting a Prior Assignment

Python	Process	Access	Summarize	Visualize	Prepare	Construct	Classification	Clustering	Regression	Evaluation
Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 1 Eligible Skills

Choose an assignment that you did not achieve the target level for. Write a blog style notebook analysis that corrects what you could have done better, what you learned, and addresses the misconception if applicable.

Data Science Pipeline

Like the day 1 activity, find two different sources that describe the data science pipeline or lifecycle. Write a blog style post that discusses their differences and hypothesizes about why they may be different? Are they for different audiences? Is one domain specific. How do they emphasize different modeling tasks?

Podcast

Python	Process	Access	Summarize	Visualize	Prepare	Construct	Classification	Clustering	Regression	Evaluation
No	Yes	No	No	No	No	No	No	No	No	No

Table 2 Eligible Skills

Watch an episode of a high quality[1] podcast and write a blog style summary and review of the episode. Highlight what you learned and how it relates to topics covered in class.

Approved Podcasts:

- Pod of Asclepius, Fall Series: The Philosophy of Data Science

What Can Data Science do?

Python	Process	Access	Summarize	Visualize	Prepare	Construct	Classification	Clustering	Regression	Evaluation
No	Yes	No	No	No	No	No	No	No	No	No

Table 3 Eligible Skills

Write a blog post describing different ways

---

[1] approved by Dr. Brown by creating a pull request to add it to the list on this page that is successfully merged. To create a PR, use the suggest an edit button at the top of this page.

Analysis Prompts

Loading Data

Python	Process	Access	Summarize	Visualize	Prepare	Construct
No	No	Yes	No	No	No	No

Table 4 Eligible Skills

Look at all of the different ways pandas can load data. Consider some of the questions below and add a notebook that's styled like a report that answers a few of them with text and code.

- Which seem like good idea? are any dangerous?
- Which seem more or less common?
- Can you compare them on speed? Is it ever worth transforming a dataset before loading?
- How much can you repair a dataset using the parameters of the load functions?

CheatSheet

Python	Process	Access	Summarize	Visualize	Prepare	Construct	Classification	Clustering	Regression	Evaluation
No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5 Eligible Skills

Make a cheatsheet with examples of the several different parameter settings for common operations for one topic.

This cheatsheet is an example, it's too broad, but it's the same idea. Yours should

Deeper Analysis

Python	Process	Access	Summarize	Visualize	Prepare	Construct	Classification	Clustering	Regression	Evaluation
Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 6 Eligible Skills

For one of the assignments, if there was something you were curious about. Try it out and investigate how to answer it. Vary parameters and document your investigation.

New Analysis

Python	Process	Access	Summarize	Visualize	Prepare	Construct	Classification	Clustering	Regression	Evaluation
Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 7 Eligible Skills

For one a topic of interest, clean, explore and model the data. Work with messy data or data provided in multiple files to earn prepare and construct achievements or use clean data to earn only EDA and modeling achievements.

Modeling Experiments

Python	Process	Access	Summarize	Visualize	Prepare	Construct	Classification	Clustering	Regression	Evaluation
No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 8 Eligible Skills

FAQ

This section will grow as questions are asked and new content is introduced to the site. You can submit questions:

- via e-mail to Dr. Brown (brownsarahm) or Beibhinn (beibhinn)
- via Prismia.chat during class
- by creating an issue

Syllabus FAQ

How much does assignment x, class participation, or a portfolio check weigh in my grade?

Can I submit this assignment late if …?

GitHub FAQ

My portfolio won't compile

Help! I accidentally merged the Feedback Pull Request before my assignment was graded

Common Debugging Issues

Key Error

Resources

This section will compile resources for the course over time into various sections:

- Data sources of data to use for assignments
- Programming info on the tools and libraries we're using in class, background info, etc
- Reference example tips
- Tips general, semi-related information
- Cheatsheet patterns and checkpoints

General Tips and Resources

This section is for materials that are not specific to this course, but are likely useful. They are not generally required readings or installs, but are options or advice I provide frequently.

on email

- how to e-mail professors

References on Python

- Course Text

Data Sources

- Kaggle
- Google Dataset Search
- UCI Data Repository
- Json Datasets
- Databases

If you have others please share by creating a pull request or issue on this repo (from the GitHub logo at the top right).

By Professor Sarah M Brown
© Copyright 2021.