

We will get started at 1:00pm PT

In the meantime...

1. Are you signed into the course Slack team? https://bit.ly/IntuitDataScience_SDG3
2. Have you created an account on GA's Github for Enterprise? <https://git.generalassembly.ly/intuit-cohort6>
3. Have you posted your Github username to the **#classroom** channel on Slack?
4. Fill out this quick survey (posted in slack too!) <https://forms.gle/KLd9txu3RzMjUD4o8>



GENERAL ASSEMBLY

Intuit Data Science Program Orientation

April 20, 2020

Overview

1:00 - 1:50 pm PT: Installfest

- GA Introduction
- Instructional Team Introductions
- Managing Your Digital Classroom
- Technology Check
- Course Management

Break: 1:50 - 2:00 PM PT

2:00 - 4:00 pm: Course kickoff:

- Intuit Introduction
- Icebreaker
- Intro to Data Science
- Break (10 min)
- Data Science Workflow



Meet General Assembly

GA Team Introduction

Jenny Morris



**Senior Customer
Success Manager**

New York City

Michelle Wengrover



**Customer Success
Manager**

New York City



We empower people to pursue work they **love** through education in technology, business, and design.





30+
Global Campuses

70,000+
Course Alumni

500,000+
Workshop
Attendees

20,000 +
Expert Instructors

Meet Your Instructors

Instructional Team



Greg Swanson
GA Instructor



Brahm Rhodes
IA



John Hazard
IA

Managing your Digital Classroom

Become Familiar with the Remote Learning Environment

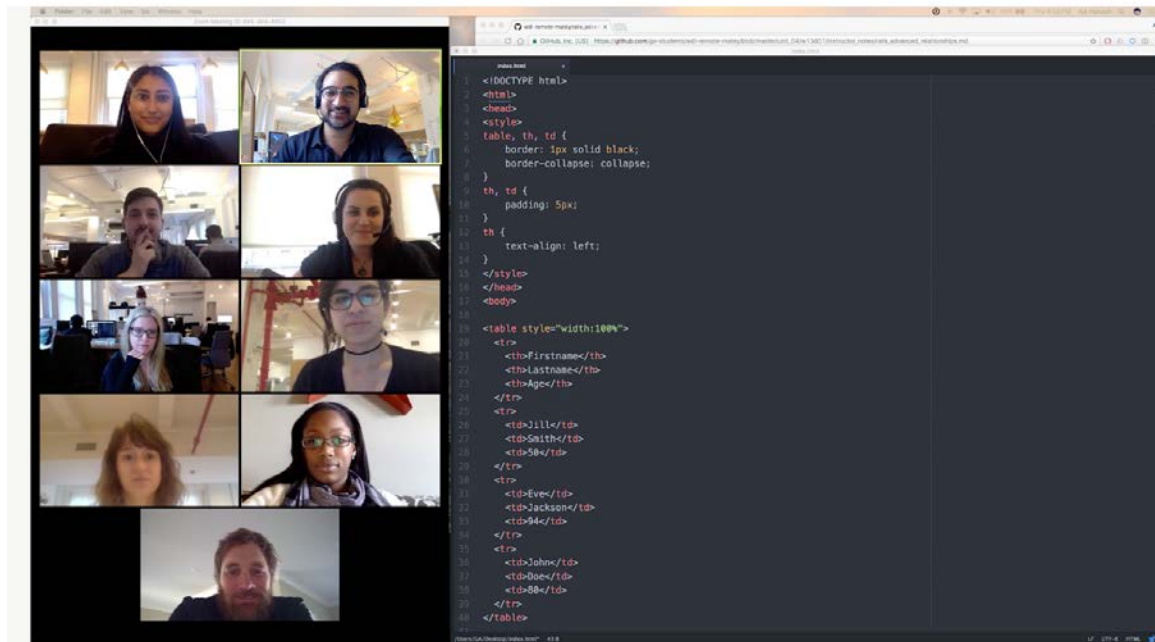
At GA, we leverage different Platforms when teaching online to make the experience as interactive & engaging as possible:

- Zoom for video conferencing
- Slack for in-class communication & checks for understanding.

Because online classes are hosted live & are interactive, getting used to managing those platforms on a single-screen is the first step to ensuring your success in learning online. Watch the video linked on our next slide for how to set up your screen with a dual or single monitor set-up!

Online Norms & Best Practices for Learning

Become Familiar with the Remote Learning Environment



Zoom Norms & Best Practices

Zoom is a video-conferencing platform. It's similar to Google Hangouts or Skype, but it provides your instructional team with robust options such as:

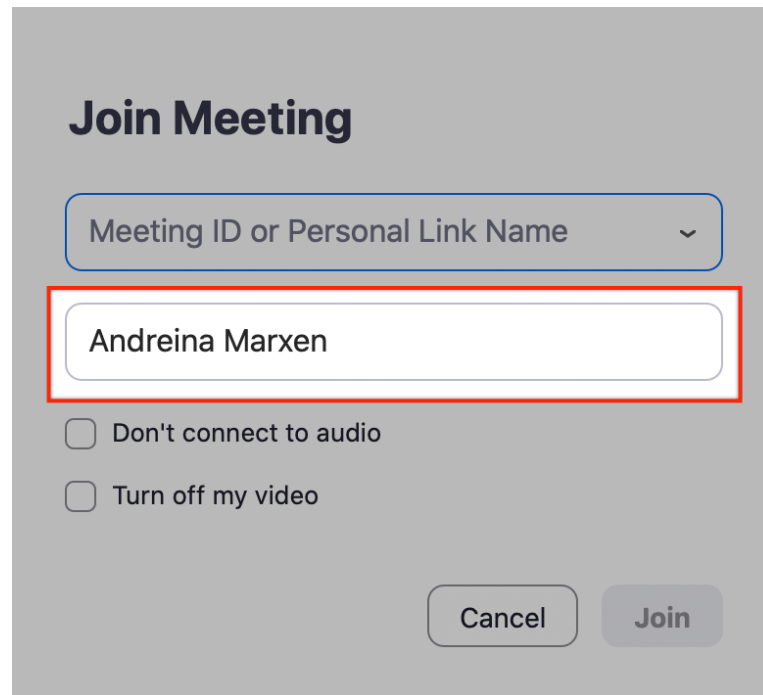
- **Screen-sharing** so that you can follow along with their slides & demonstrations.
- **Breakout Rooms** so that you can work in small groups on exercises & projects.
- **Annotations** for whiteboarding exercises that allows your entire class to participate.
- **Recording sessions** to the cloud for you to review important lessons as needed.

Online Norms & Best Practices for Learning

Zoom Norms & Best Practices

In order to join a Zoom call, you'll need to:

- [Sign-up](#) for an account.
- [Test your audio & video](#) ahead of time.
- Set your display name to your First & Last Name.
- Make sure you have a strong internet connection - ethernet cords are strongly recommended!



Join Meeting

Meeting ID or Personal Link Name ▼

Andreina Marxen

☐ Don't connect to audio

☐ Turn off my video

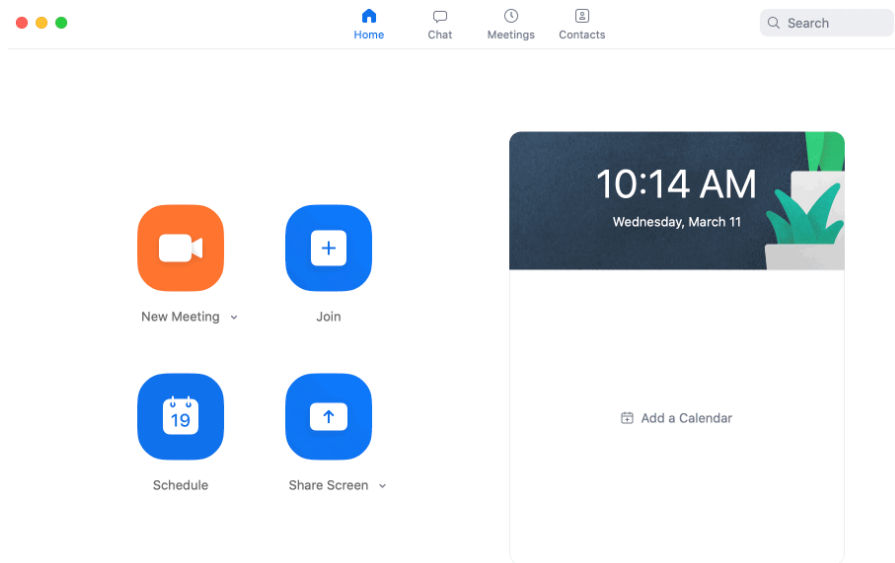
Cancel Join

Online Norms & Best Practices for Learning

Zoom Norms & Best Practices

On the day of the meeting:

- Join the zoom call a few minutes before the start of class.
- You can join the meeting by
 - clicking on the zoom link or
 - by typing in the zoom Meeting ID in Join Meeting on your desktop app.
- Set your yourself to **mute!**
- Make sure your **video is turned on** .



Zoom Norms & Best Practices

Be Present

- Class time should be a distraction free zone. Please set boundaries in your household by letting everyone know that you will be unavailable for those active class hours. This way you can direct your full attention to class!
- You should be on camera for the entirety of the class, unless you need to step away briefly. If you do need to step away, let your instructional team know via direct message on Slack.

Be Mindful

- You should be on mute during lecture & when you are not speaking. This will allow for no background noise to impede learning.
- Make sure you're in a well-lit space with as few distractions in the background as possible.

Be Participatory

- Since most of the time you will be on mute, slack will come in handy whenever you want to share your thoughts but we also encourage you to come off mute and speak out loud!

Slack Norms & Best Practices

Slack is a messaging platform that allows for easy, organized communication between teams. Instructors leverage slack throughout the course in order to:

- Share the day's learning objectives, in-class resources, slides, and exercise prompts.
- Take attendance & make announcements to the whole class.
- Check for understanding as they cover key concepts.
- Gauge how much longer students need on group exercises.

Before Day 1, it's important that you become familiar with [finding slack channels](#), using [emoji responses](#) & [replying via thread](#)!

Slack Norms & Best Practices

Slack will be your main point of communication with each other and the instructional team.

Before Day 1, please do the following:

- Make sure you download slack on your phones & desktops.
- Accept the invitation to the Slack Workspace for your course. Once you're in please add yourselves to the class channel.
- Please keep an eye on the slack channel for important announcements by the instructional team.

Tech Check

Data Science happens in quite a few languages



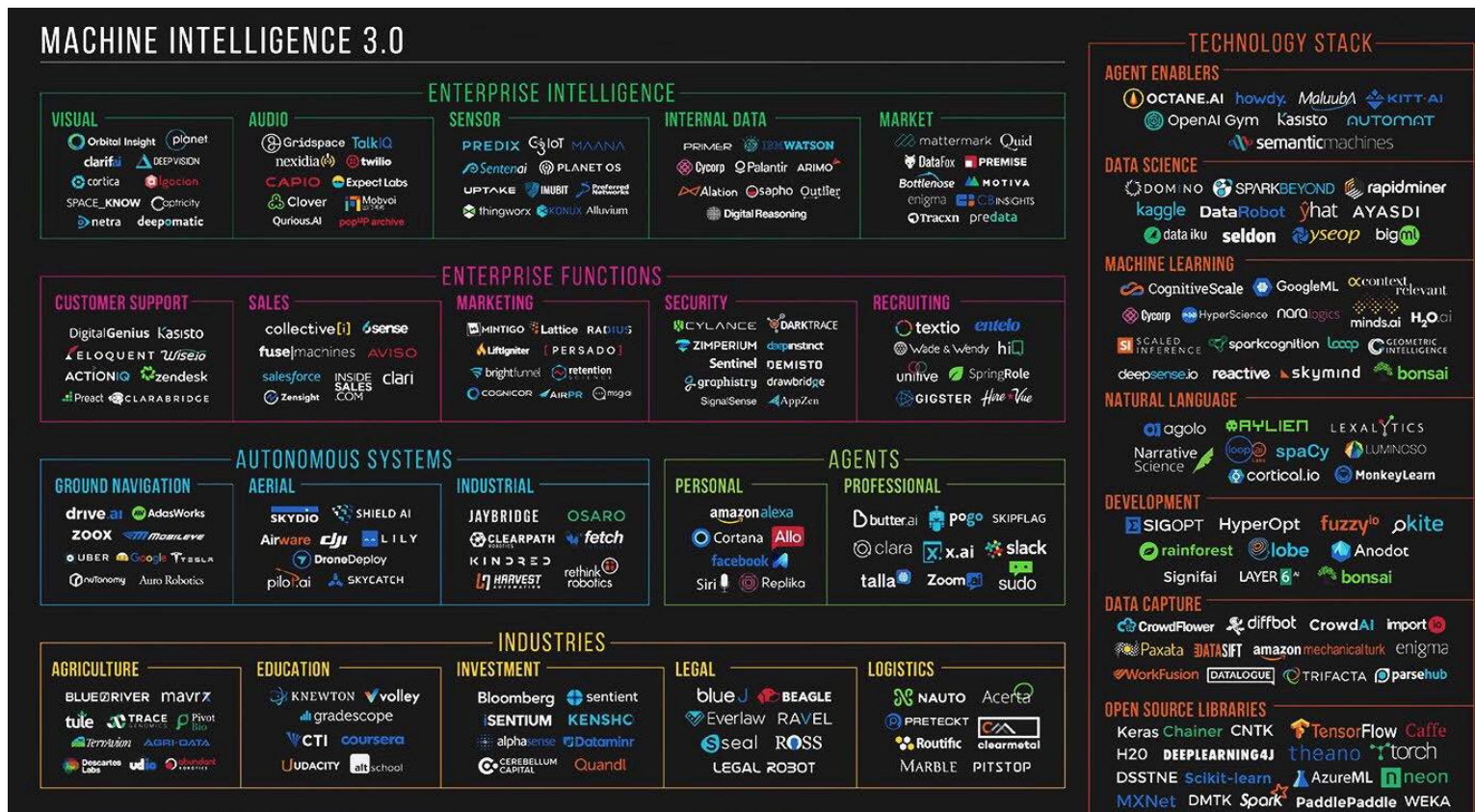
- Simple Object Oriented Language
- Bread and butter of many Data Scientists
- Considered a great “learn to code” language
- Widest access to machine learning algorithms
- Common and powerful Interpretive language
- Great for statistics, visuals and modeling
- Many packages available across a number of industries



- Higher level Object Oriented Language
- Incredibly powerful and robust
- Wide use cases to utilize enterprise level computing and data streaming platforms like Apache Spark and Kafka



Switch to tools and the list is extensive!



Kicking the tires

To keep everything simple, consistent and effective – we use four primary tools in the classroom. Let's get used to all of them.



zoom

 slack

 git

 ANACONDA®



Course Code Collaboration



GitHub

Git is a programming language for version-control and collaboration. In our class we'll be using git repositories (folders) to:

- Course lessons & projects
- Collaborate on your homework & capstone projects

Let's make sure everyone has git installed properly.

- Open Terminal (mac) or gitbash (PC) then run the below command

```
git --version
```



Anaconda – One distribution to keep us on the same page



Single distribution including the following in one install:

- **Conda** – package manager with virtual environment support
- **Python** – latest version of the Python interpreter
- **Standard libraries** – standardized versions of common Python libraries
- **ipython** – improved interactive Python shell
- **Spyder** – Python Interactive Development Environment
- **Jupyter** – “Notebook” environment predominantly for Data related purposes

Open Terminal (mac) or Anaconda Prompt (PC) then run the below command

```
conda list anaconda
```



Jupyter – Popular Notebook environment

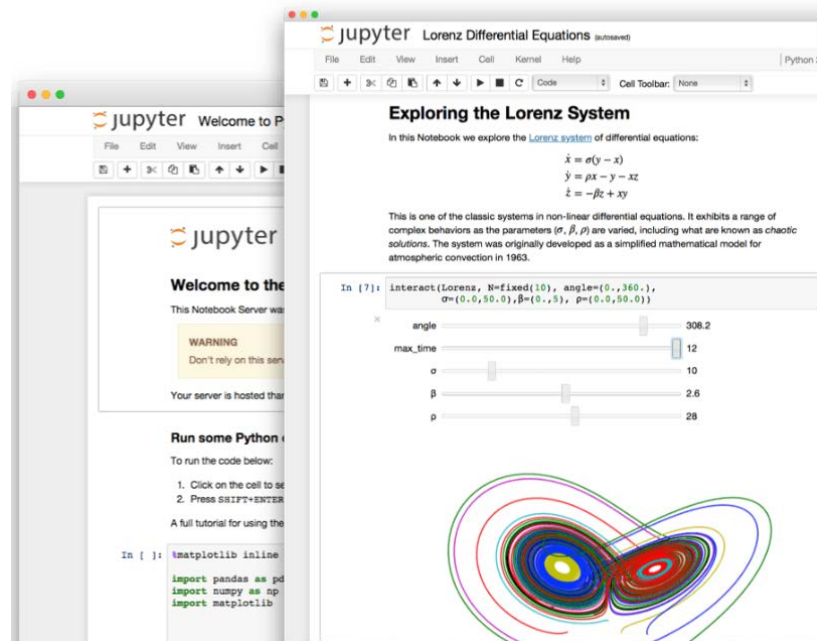


Popular open-source application allowing you to create and share documents that contain live code, equations, visualizations and narrative text.

Notebooks are heavily used in data-related fields like statistics, ETL/ELT, analytics, visualization and machine learning

Let's open jupyter together. In terminal (mac) or Anaconda Prompt (PC) type

jupyter notebook



Course Management

Classroom Expectations



- Be on time to class
 - We take attendance
 - You will be marked as late after 15 minutes
- Follow the attendance and absence protocol
 - Do not miss more than 3 classes
 - Let an IA know in advance if you need to miss class
 - Set aside time to review lesson materials, and attend office hours with questions
- Outside of class time
 - ~5-7 hours per week is a good guide



The Program



Every Monday from 4/20 - 10/13

1:00 pm - 4:00 PM PT

Capstone presentations on 10/13

Your class Slack:

- Go to the class Slack:
[Intuit Data Science Classroom](#)

Being on camera on Zoom is necessary for attendance!



To join class:

1. Sign on to Slack.
2. Go to the class Slack:

[Intuit Data Science Classroom](#)

3. Click on the Zoom link posted in the meeting invite 5-10 minutes before the start of class.

Graduation Requirements

- Attendance
 - Missing more than the 3 full classes will put you at risk of not graduating
- Projects
 - Meet expectations on at least 80% of project milestones
- Engagement
 - Class participation, engagement and giving feedback



Data Science Schedule: San Diego (Cohort 6)

Step	Learning Component	Format	Dates
1	Data Science Level 1 Assessment (DS1)	Online, timed	Jan 15 - Jan 23
2	Data Science self-paced pre-work	Online, self-paced	Jan 29 - Apr 17
3	Data Science 60-hour training program	Online	Apr 21 - Sep 29

- Class meets once a week on Tuesday for 3 hours
- Office hours: 12:30-1:00 PM PT before class, additional office hours to be scheduled with IAs
- Last day of lessons: Sep 1
- Capstone prep between Sep 8 - Sep 22; capstone presentations on Sep 29
- Holidays on May 26



Curriculum Overview

Unit	Title	Topics / Skills	Sample Project
1	Fundamentals	Git, Command Line, Python, Jupyter Notebook	Python Coding Challenges -- reverse an inputted string
2	Working with Data	Pandas, NumPy, SciPy, Descriptive/Inferential Statistics	Explore popular movies on IMDB; explore retail sales data
3	Data Science Modeling	Linear Regression, Classification	Predict home prices in Boston; classify healthcare results
4	Data Science Applications	Unsupervised Learning, Principal Component Analysis, NLP, Time Series, Neural Networks	Capstone
Capstone	Guided project working sessions, Capstone presentations		



Instructor Support

Instructor Support:

- In-class
- Slack
- Optional office hours

Optional Office Hours with IAs:

- Office hour: Mondays 12:30 - 1:00 pm PT (Brahm/John)
- Remote office hours
 - Remote office hours need to be booked. Link to book a timeslot will be posted in Slack and GH. You will receive a Zoom video conferencing link at your scheduled time.

Break

Intuit Introduction

Intuit Introduction



Oleg Kislyuk
Manager, Learning & Development

Ice Breaker

Meet Your Instructors



Steven Longstreet

Hilton - Senior Director Data & Analytics Strategy
GA - Lead Data Science Instructor (Enterprise), Data
Science Standards Board Member, Distinguished Faculty

Steven is a lead Data Science Instructor for General Assembly as well as a Board Member of the General Assembly Data Science Standards Board. In his primary job at Hilton, he leads the Strategy, Communications, Change Management and Production Readiness teams for Hilton Data and Analytics. Previous roles at Hilton include the development of new Hotel brands, rebuilding Hilton's signature benefit (Team Member Travel) and building data driven approaches to Innovation & Product. Before joining Hilton Steven started his career with technology startups then served as a consultant for a range of both government agencies and Fortune 100 companies.

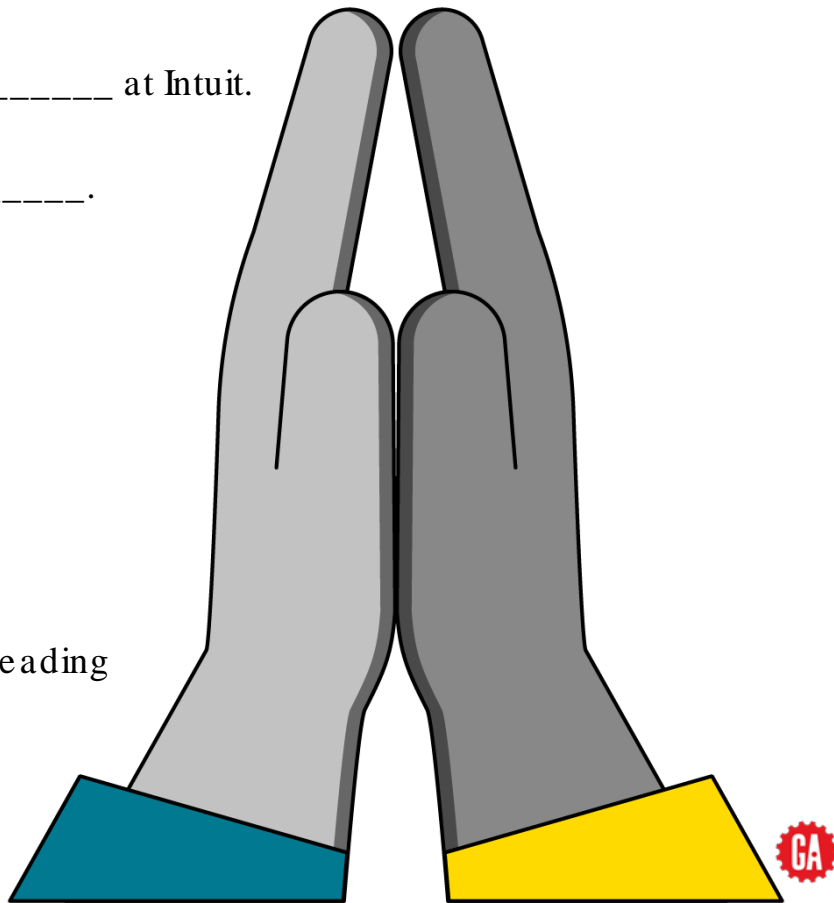
Let's Get to Know You!



1. My name is _____ and I am a _____ at Intuit.
2. I'm learning data science because _____.
3. A fun fact about me is _____.

Some ideas...(pick one)

- Your most used emoji
- Your guilty pleasure
- The worst job you've ever had
- An app you can't live without
- What you're currently binge-watching or reading

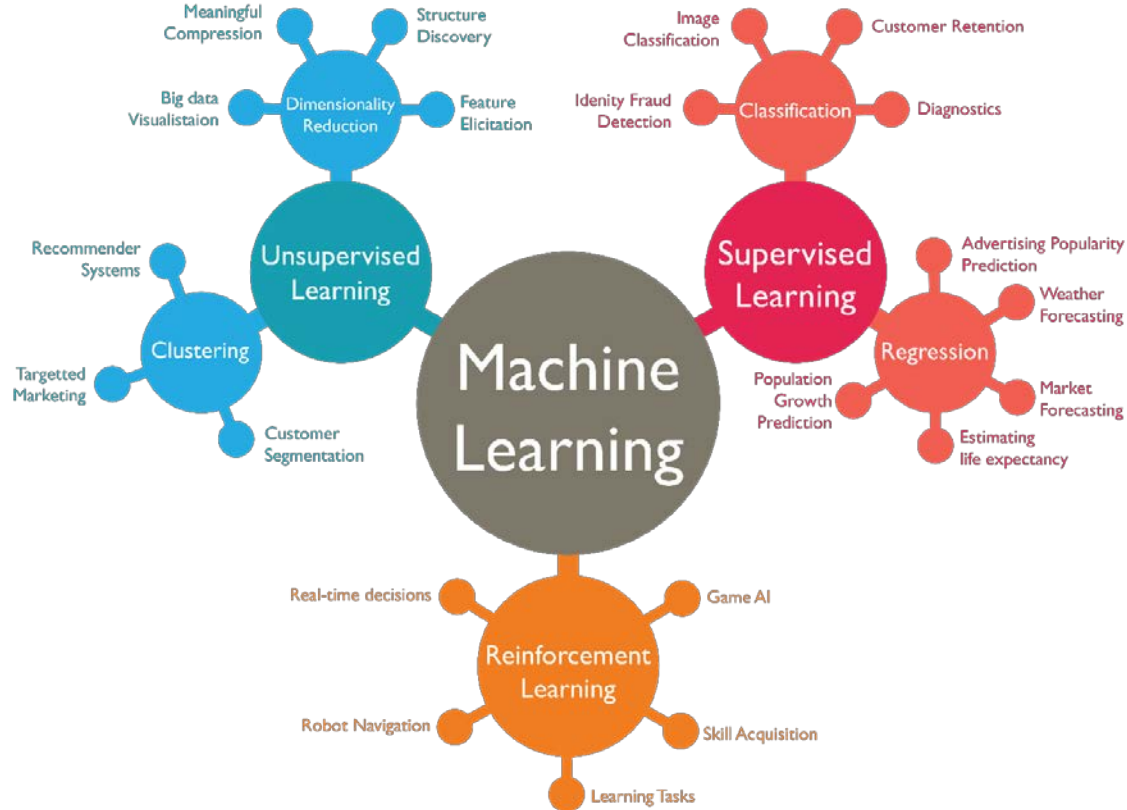


INTRODUCTION

WHAT IS DATA SCIENCE?



Data Science applications are vast!



DATA SCIENCE IN REAL WORLD

In Slack, list 3 products or services you think utilize data science.



EYES IN THE SKY HELP FARMERS ON THE GROUND

Aerial imaging companies using high-resolution cameras and artificial intelligence data analysis are giving the agricultural community a high-tech boost. NYTimes, Lauren Smiley, Sept. 8, 2019



What is Data Science?

Let's start with discussing key terms:

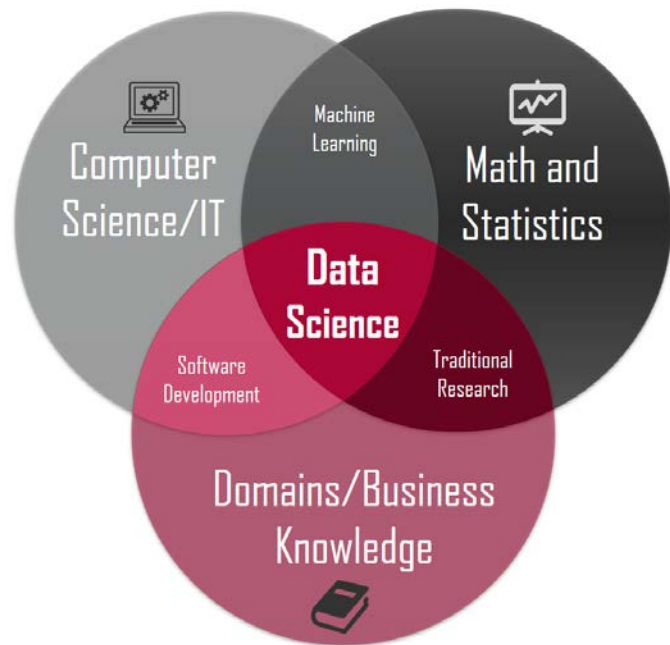
- Statistics
- Machine Learning
- Artificial Intelligence



What is Data Science?

At a high level...

- Application of machine learning to solve real world problems
- Capable of creating immense value in everyday life



“

**One problem: trying to agree on a
definition for data science.**

Board for AI & Data Science Standards

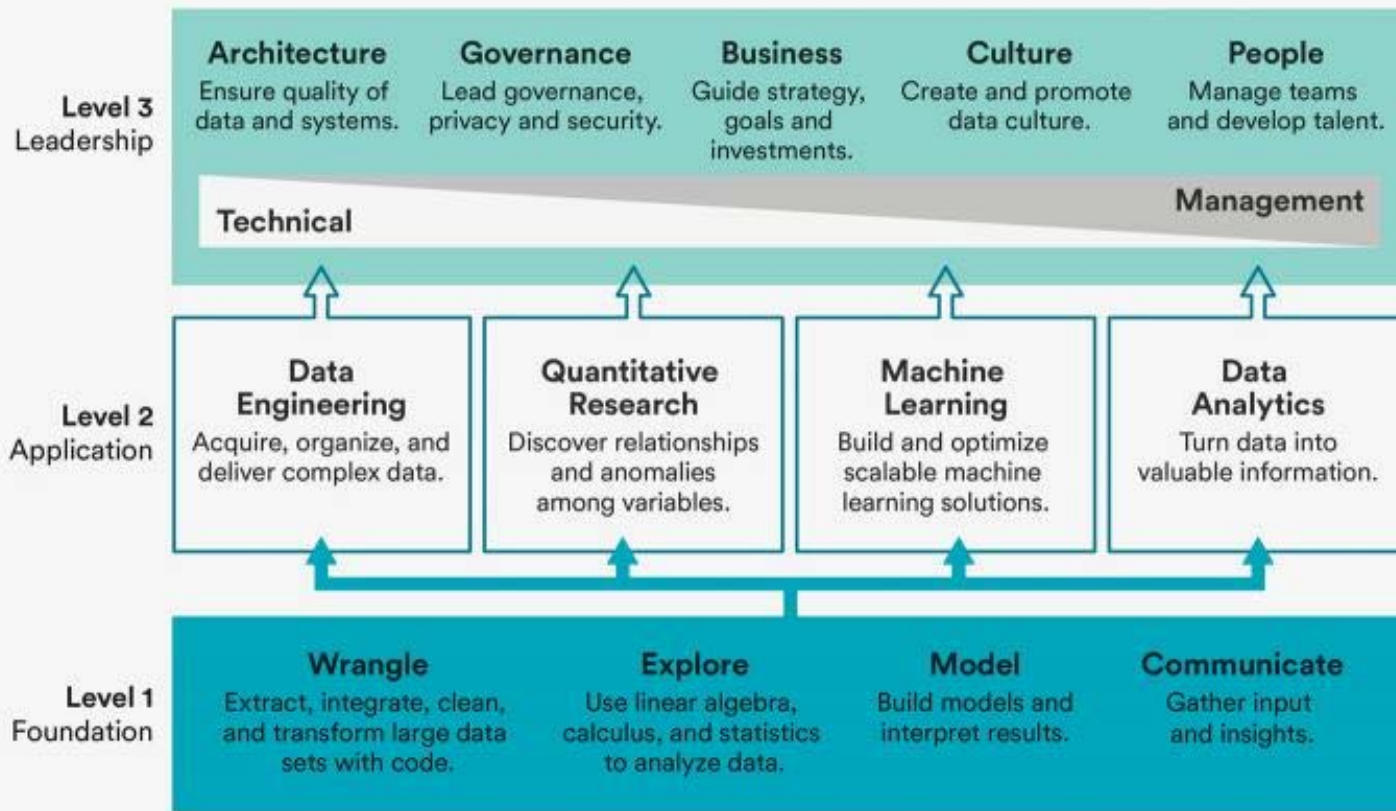
Data science is the practice of

- acquiring, organizing and delivering complex data,
- discovering relationships and anomalies among variables,
- building and deploying machine learning models,
- and synthesizing data to influence decision making.

Machine Learning: A model or equation that changes and improves its outputs over time



The Data Science Career Framework





Why do we care about Data Science?

Interested in applying data science to my current role

47%

Interested in becoming a data scientist

33%

Overall industry trend towards data science & AI

28%

Feeling limited by current tools

25%

I am a data scientist/not applicable

18%

I work with data scientists

14%

Launching a startup or ML/AI project

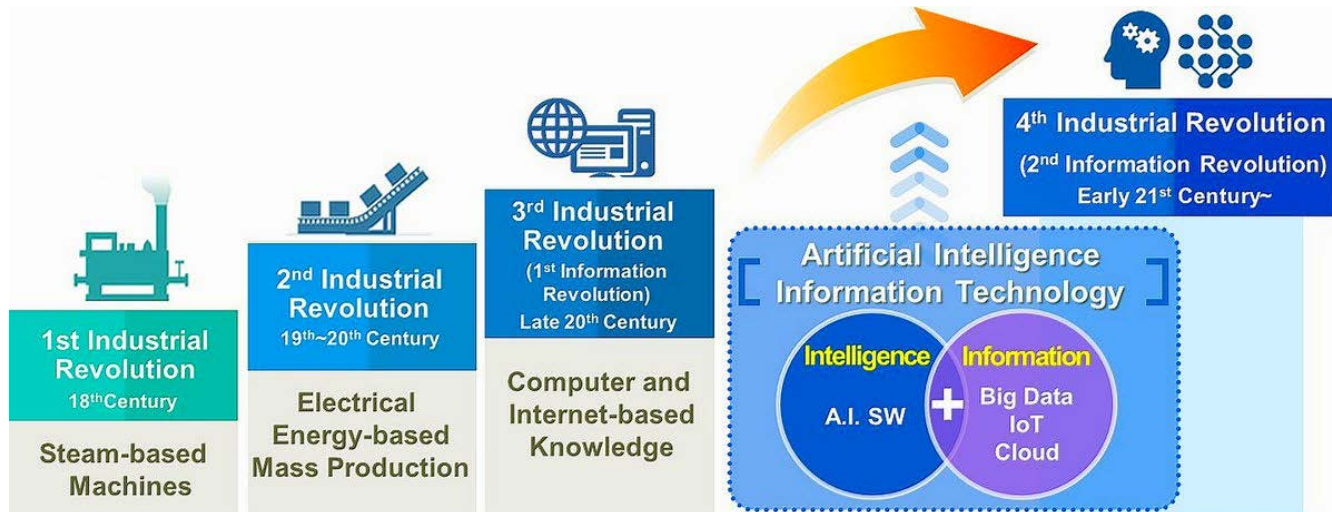
8%

2019 ANACONDA STATE OF DATA SCIENCE

WHY ARE YOU
LEARNING ABOUT
DATA SCIENCE?

Welcome to the 4th Industrial Revolution!

Big data, immense advances in processing power, machine learning and Data Science are at the forefront of an explosive new chapter in human history



Source : WorldBank.org

shared via @pradeeprao_



facebook

The Seattle Times

WHAT YOU DON'T KNOW ABOUT HOW FACEBOOK USES YOUR DATA

The social-media giant also collects so-called biometric facial data without users' explicit "opt-in" consent, and helps video-game companies target "high-value players" who are likely to spend on in-app purchases. Seattle Times, Natasha Singer, April 11, 2018

WHO USES DATA SCIENCE?

The Switch

Red light, green light: New Audis will predict the time until that stoplight turns green [Article](#)



Innovations

Netflix reveals what images hook viewers on new shows [Article](#)



WHO USES DATA SCIENCE?

NETFLIX



amazon.com



Google



Hilton

Bloomberg



Leveraging Data Science

What problem are we trying to solve?

TESLA'S SELF DRIVING TRUCK

Here's some food for thought:

- 4,000 truckers die every year
- The American Trucking Associations reports the annual driver turnover for large truckload carriers reached a whopping 90 percent this year, and it projects a 50,000-driver shortage by the end of 2017.
- Labor is the most expensive component of truck transport
- The US is experiencing a freight rail crisis plagued by bottle necks and unprecedented delays – and we aren't investing heavily in new rail lines



Empowering Truck Drivers vs Replacing them

To be realistic there's not a lot of skill to driving a truck on the interstate. What can our machine learning algorithms do for us?

- Keep the truck between the lines
- Adjust for and avoid other vehicles
- Read and adjust for road signs

Where does it have problems? Off highway and maneuvering for loading docks.

What if...one operator could drive 5 trucks?



SO MUCH OF "AI" IS JUST FIGURING OUT WAYS TO OFFLOAD WORK ONTO RANDOM STRANGERS.

American Truck Simulator

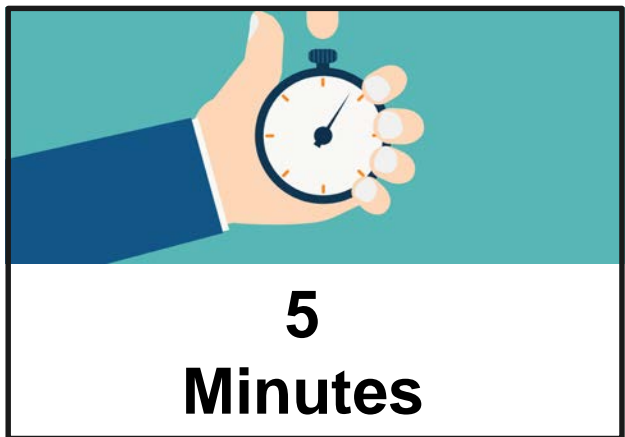
Two things to know:

- This is both a real and popular game!
- People make incredible rigs to simulate driving trucks

How is this any different than piloting a drone?



What problem do you want to solve?



It's early – but what piqued your interest in this course? The best projects center around personal interest. Let's dig into this in small groups of 3-4 people

Your group will discuss & report back the following on slack:

- What you find interesting about Data Science?
- Types of problems you'd like to solve with Data Science



INTRODUCTION

Machine Learning

MACHINE LEARNING TERMINOLOGY

Supervised learning (a.k.a., “predictive modeling”):

- *Classification and regression*
- Predicts an outcome based on input data
 - **Example:** Predicts whether an email is spam or ham.
- Attempts to generalize
- Requires past data on the element we want to predict (the target)

Unsupervised learning:

- *Clustering and dimensionality reduction*
- Extracts structure from data
 - **Example:** Segmenting grocery store shoppers into “clusters” that exhibit similar behaviors.
- Attempts to represent
- Does not require past data on the element we want to predict.

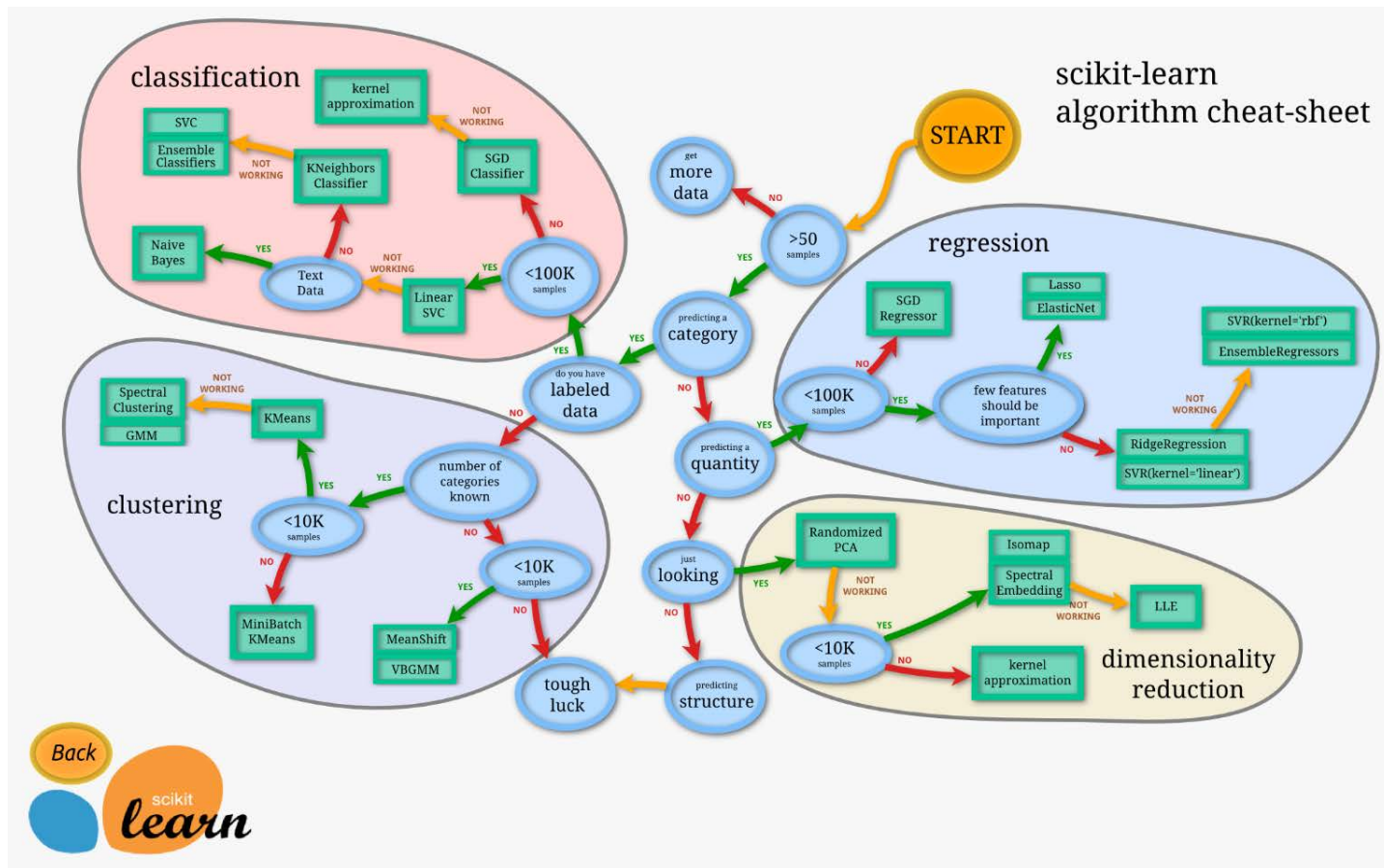


USING MACHINE LEARNING

- Oftentimes, we may combine both types of machine learning in a project to reduce the cost of data collection by learning a better representation. This is referred to as transfer learning.
- Unsupervised learning tends to present more difficult problems because its goals are amorphous.
- Supervised learning has goals that are almost too clear and can lead people into the trap of optimizing metrics without considering business value.



Slack Exercise – Let's explore a few areas of Data Science we'll be learning in class!



SUPERVISED LEARNING

Most frequent type of work that data scientists do and will be the main focus of this course.
How does supervised learning work?

1. We train a **machine learning model** (more on that shortly) using **labeled data** (the "response" label from earlier).
2. We make predictions on **new data** for which the response is unknown.

The primary goal of supervised learning is to build a model that “generalizes” —i.e., accurately predicts the **future** rather than the **past**!

CLASSIFICATION VS. REGRESSION

There are two categories of supervised learning:

- **Regression**

- The outcome we are trying to predict is a continuous value.
- Can you think of anything we would want to predict like this?

- **Classification**

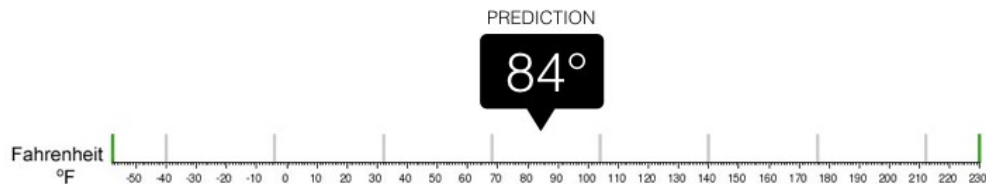
- The outcome we are trying to predict is categorical (i.e., it comes in one of a set number of classes).
- Can you think of anything that we would want to predict like this?

The type of supervised learning problem has nothing to do with the features; only the response matters!



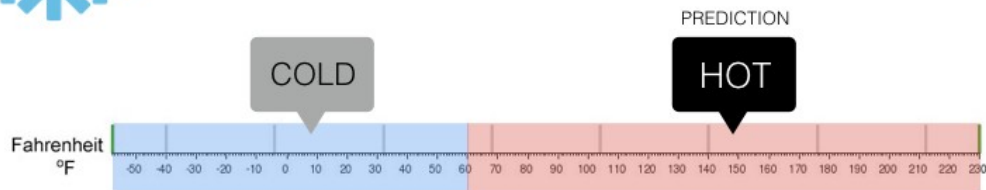
Regression

What is the temperature going to be tomorrow?



Classification

Will it be Cold or Hot tomorrow?



UNSUPERVISED LEARNING

Common Types of Unsupervised Learning

- **Clustering:** Groups “similar” data points together.
- **Dimensionality reduction :** Reduce the dimensionality of a data set by extracting features that capture most of the variance in the data.

Types of Customers at a Bar

- **Observations:** Customers.
- **Features:** Drink purchases, people they interact with, etc.
- **Response or target variable:** There isn't one —instead, we group similar customers together.

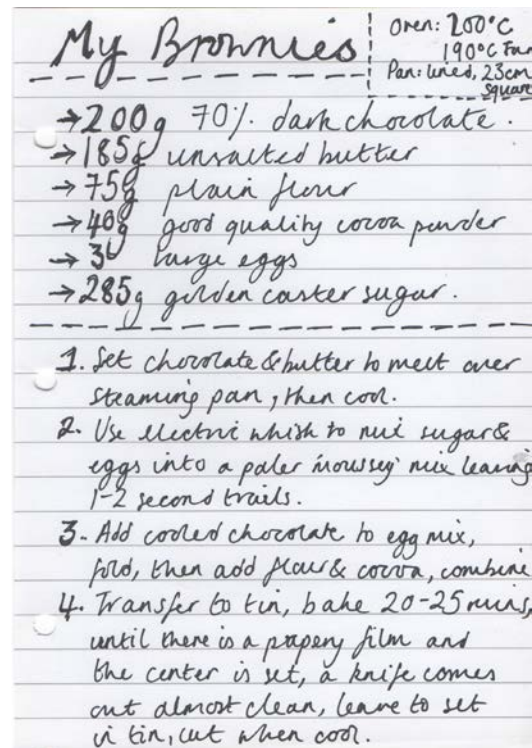


Understanding algorithms

As defined: a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Think of it like a recipe. A Step-by-step process needed to get a successful outcome.

- There are many ready to use
- You may need to modify it
- Often you'll want to build your own from scratch



Breakout Group Exercise!



ASKING GOOD QUESTIONS



Asking a Good Question

Why do we need a good question?

“A problem well stated is half solved.”—Charles Kettering

Even though all data science projects have different general flows, they start in the same place: with a problem. From this problem statement arises questions; questions we will ask the data in order to gain more information so we can attempt to find a solution to that problem.

Components of a Good Question

1. Sets you up for success as you begin analysis
2. Establishes the basis for reproducibility
3. Enables collaboration through clear goals



Overview of the SMART Goals Framework

1. **Specific:** The data set and key variables are clearly defined
2. **Measurable:** The type of analysis and major assumptions are articulated
3. **Attainable:** The question you are asking is feasible for your data set and not likely to be biased
4. **Reproducible:** Another person (or future you) can read and understand exactly how your analysis is performed
5. **Time-bound:** You clearly state the time period and population to which this analysis pertains



What Are Some Common Questions Asked in Data Science?

Machine learning more or less asks the following questions:

- Does X predict Y? (Where X is a set of data and y is an outcome.)
- Are there any distinct groups in our data?
- What are the key components of our data?
- Is one of our observations “weird”?

From the business the ask is often:

- What is the likelihood that a customer will buy this product?
- Is this a good or bad review?
- How much demand will there be for my service tomorrow?
- Is this the cheapest way to deliver my goods?
- Is there a better way to segment my marketing strategies?
- What groups of products are customers purchasing together?
- Can we automate this simple yes/no decision?



INTRODUCTION

The Data Science Workflow

OVERVIEW OF THE DATA SCIENCE WORKFLOW

- A methodology for doing Data Science
- Similar to the scientific method
- Helps produce *reliable* and *reproducible* results
 - *Reliable*: Accurate findings
 - *Reproducible*: Others can follow your steps and get the same results

Throughout this course and for our projects, we'll be following a general workflow

OVERVIEW OF THE DATA SCIENCE WORKFLOW

Steps in the Data Science Workflow

- **Frame:** Develop a hypothesis-driven approach to your analysis.
- **Prepare:** Select, import, explore, and clean your data
- **Analyze:** Structure, visualize, and complete your analysis
- **Interpret:** Derive recommendations and business decisions from your data
- **Communicate:** Present (edited) insights from your data to different audiences



A Quick Note about Workflows

This is our workflow. It's not necessarily your workflow or the workflow for a specific project.

- Some projects may not require every step
- These steps are iterative; it's normal to go back and repeat certain steps a few times in a row
- The process is cyclical; after completing the process, you may restart it on new findings

Key Components:

- Reliable
- Reproducible
- (Explainable)



Application: Data Science Workflow Through Ames Data

What is Ames Data?

DESCRIPTIVE ABSTRACT: Data set contains information from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010.

Why are you getting Ames Data?

For many of our lessons, projects and activities we'll be providing you access to various datasets to explore, discover and build analysis upon. It provides a common reference point for learning

Ideal Data vs. Available Data

Oftentimes, we'll start by identifying the *ideal data* we would want for a project, but as data is acquired, we realize the limitations of the data and must decide if it is enough to find a reasonable conclusion.



Application: Data Science Workflow Through Ames Data

Frame

- Identify the business/product objectives.
- Identify and hypothesize goals and criteria for success.
- Create a set of questions to help you identify the correct data set.

We work for a real estate company interested in using data science to determine the best properties to buy and resell. Specifically, your company would like to identify the characteristics of residential houses that estimate their sale price and the cost-effectiveness of doing renovations.

Application: Data Science Workflow Through Ames Data

Identify the Business/Product Objectives

The customer tells us their business goals are to accurately predict prices for houses (so that they can sell them for as large a profit as possible) and to identify which kinds of features in the housing market would be more likely to lead to foreclosure and other abnormal sales (which could represent more profitable sales for the company).

Identify and Hypothesize Goals and Criteria for Success

Ultimately, the customer wants us to:

- Deliver a presentation to the real estate team.
- Write a business report discussing results, procedures used, and rationales.
- Build an API that provides estimated returns.

Create a Set of Questions to Help You Identify the Correct Data Set

- Can you think of questions that would help this customer deliver on their business goals?
- What sort of features or columns would you want to see in the data?



Understand Dataset



EXERCISE

Take a moment to look through the [data description](#).
How closely does the set match the ideal data that you envisioned? Would it be sufficient for our purposes?
What limitations does it have?

What are some questions we should ask during the acquisition process?

What are some questions we should ask when checking the data for quality?

Moving to Prepare Step

Often, we are given secondary data, or data that were collected previously. In these cases, we have to learn as much as possible about our data using tools like data dictionaries and source documentation to determine how the set was gathered.

Here's an example of a data dictionary:

Variable	Description	Type of Variable
Square Footage	Floating Point	Continuous
Street Type	1 - Gravel, 2 - Paved	Categorical
Neighborhood	String, e.g., 'Tenderloin'	Categorical
Number of Bedrooms	Integer	Discrete

Common considerations when preparing our data include:

- Ensuring data is clearly defined and structured
- Check and clean data formatting as needed

Common considerations for cleaning include:

- Most data will not come perfectly clean and ready to use. Cleaning data is normally the most time-consuming task a data scientist faces.
- As you can see, the "Prepare" phase of the data science workflow encompasses several steps: the act of reviewing, indexing, and cleaning your data. This normally consumes a great deal of time!

Moving to Analyze

Analyze

As an example of basic statistics, Data scientists often check the mean, standard deviation, or specific frequency counts of their data. Statistics that we might expect for the earlier housing variables include:

Variable	Mean or Frequency (%)
Square Footage	2201.3
Street Type - Gravel	8%
Street Type - Paved	92%
Number of Bedrooms	1.8

What sort of questions do these types of statistics allow us to answer? Why would we do this?

Continuing with Analyze

Creating a Predictive Model

We generate predictive models based on the SMART goal we decided upon earlier. Typically, our interest is in predicting or guessing some sort of value we might be interested in (such as the housing price for a house given some set of fixed characteristics).

*What are some other business goals we can support as data scientists for this realty company?
What are some values we would like to guess?

What do you think are the steps for model building?

We'll be spending much of our time in this course on data analysis and predictive modeling.



Onto Interpret!

Develop Recommendations and Decisions

- Now that you have a model, what are some things you should check?
- Now that you have a model, can you convert your model's finding into a conclusion or next step for your employer?



Finally Communicate

Share the Results of Your Analysis

Presentations are a critical part of your analysis. It doesn't matter how brilliant your model is or how illuminating your findings are — without effective communication, your work will not be used.

The most basic form of a data science presentation should include a simple sentence that describes your results:

"Customers from large companies had twice (CI 1.9, 2.1) the odds for placing another order with Planet Express compared to customers from small companies."

Data science presentations can also be far more complex and exciting, like some of the research presented by Nate Silver's FiveThirtyEight blog.

Communicate Considerations

When crafting a presentation, always consider your audience and make sure to practice your presentation beforehand. Consider the types of questions people might ask or — better yet — test your presentation on a few people and pay attention to their response. Clarify and refine your presentation accordingly.

Make sure to consider your needs and goals, as well as those of your audience. A presentation created for your fellow data scientists will be vastly different than a presentation intended for executives trying to make a business decision.



A Note About Iteration

Iteration is an important part of *every step* in the data science workflow. At any given point in the process, you may find yourself repeating or going back and redoing steps in order to better understand your data, clarify your model, and refine your presentation.

What are some things you may want to redo or iterate over after presenting your findings?



Summary

Crafting good questions is key.

- Without a thoughtful, targeted, and SMART question, it can be difficult to create an effective model.

Use the data science workflow to iteratively develop solutions.

- Frame: Develop a hypothesis-driven approach to your analysis.
- Prepare: Select, import, explore, and clean your data.
- Analyze: Structure, visualize, and complete your analysis.
- Interpret: Derive recommendations and business decisions from your data.
- Communicate: Present (edited) insights from your data to different audiences. 3) Informed by your past work, continue to refine your findings and models.

While the data science workflow may appear to be linear, we consistently return to past steps to implement new findings



CONCLUSION

Check to see if you can answer the following questions easily:

- What is data science?
- What is the data science workflow?
- What is the difference between supervised and unsupervised learning?
- What is the difference between regression and classification?
- What is an algorithm?



WELCOME TO DATA SCIENCE

Q&A

EXIT TICKET

