



Pandas Foundation

Online Course by DataRockie



Pandas Foundation

- Python Language
- 30+ Video Full HD
- Beginner Friendly
- Just 2 Hours Quick and easy to learn!



Course Outline

- Introduction

- Python Review

variable

types

list

dictionary

function

- Basic Pandas

`read_csv()`

add/ drop
columns

inspect
dataframe

summarise
dataframe

- Intermediate Pandas

select
columns

filter
rows

merge
dataframe

transformation



Your Friend (Instructor)



{

"name": "Toy",

"page": "DataRockie",

"skills": ["Economics", "Marketing", "Data Science"],

"hobbies": ["Reading", "Learning", "Sharing"],

"Tools": ["Google Sheets", "SQL", "R",
"Python", "Data Studio"]

}



What exactly is DataFrame?

	id	name	gpa	C	city
0	1	Toy	3.42	good	Tokyo
1	2	Joe	3.50	good	London
2	3	Ann	4.00	good	Bangkok
3	4	Mary	2.58	ok	Bangkok
4	5	David	2.98	ok	Seoul



Python library
designed to work
with dataframe



Like Excel Table

Columns

Rows

	id	name	gpa	C	city
0	1	Toy	3.42	good	Tokyo
1	2	Joe	3.50	good	London
2	3	Ann	4.00	good	Bangkok
3	4	Mary	2.58	ok	Bangkok
4	5	David	2.98	ok	Seoul

Value



Easy to Learn

```
import pandas as pd  
import numpy as np
```

import pandas

```
raw_data = {  
    "id": [1,2,3,4,5],  
    "name": ["Toy", "Joe", "Ann", "Mary", "David"]  
}
```

```
df = pd.DataFrame(  
    raw_data  
)
```

df

create dataframe

	id	name
0	1	Toy
1	2	Joe
2	3	Ann
3	4	Mary
4	5	David




Basic Python (easy!)

- Variable
- Type
- List/ Dictionary
- Function/ Method
- OOP (not required, but good to know)



Google Colab

<https://colab.research.google.com/>

 Welcome To Colaboratory

File Edit View Insert Runtime Tools Help

Table of contents

Getting started

Data science

Machine learning

More Resources

Featured examples

Section

+ Code + Text Copy to Drive

• easy snaring

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Getting started

The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

```
[ ] 1 seconds_in_a_day = 24 * 60 * 60
    2 seconds_in_a_day

86400
```

To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut "Command/Ctrl+Enter". To edit the code, just click the cell and start editing.

Variables that you define in one cell can later be used in other cells:

```
[ ] 1 seconds_in_a_week = 7 * seconds_in_a_day
    2 seconds_in_a_week

604800
```

Colab notebooks allow you to combine **executable code** and **rich text** in a single document, along with **images**, **HTML**, **LaTeX** and more. When you create your own Colab notebooks, they are stored in your Google Drive account. You can easily share your Colab notebooks with co-workers or friends, allowing them to comment on your notebooks or even edit them. To learn more, see [Overview of Colab](#). To create a new Colab notebook you can use the File menu above, or use the following link: [create a new Colab notebook](#).

Colab notebooks are Jupyter notebooks that are hosted by Colab. To learn more about the Jupyter project, see [jupyter.org](#).



DataLore Notebook

```
[28] 0.1s
import pandas as pd
import numpy as np

[31]
raw_data = {
    "id": [1,2,3,4,5],
    "name": ["Toy", "Joe", "Ann", "Mary", "David"]
}

df = pd.DataFrame(
    raw_data
)

df
```

	id	name
0	1	Toy
1	2	Joe
2	3	Ann
3	4	Mary
4	5	David

```
[32]
df['gpa'] = [3.42, 3.50, 4.00, 2.58, 2.98]
```

<https://datalore.jetbrains.com/>



Free, easy to use,
lots of functions



Share Your Work

The screenshot shows the DataLore IDE interface. The main workspace displays a notebook with the following code:

```
[28] import pandas as pd
import numpy as np

[31] ▶ 0.3s
raw_data = {
    "id": [1,2,3,4,5],
    "name": ["Toy", "Joe", "Ann", "Mary", "David"]
}

df = pd.DataFrame(
    raw_data
)

df
```

Below the code, a table view of the DataFrame is shown:

	id	name
0	1	Toy
1	2	Joe
2	3	Ann
3	4	Mary
4	5	David

At the bottom, the status bar indicates: Reactive mode, Python 3.8.12, Calculated: 0, In process: 0, Errors: 0, Running, Machine: t2.medium, CPU: 0%, FreeMem: 2222MB.

On the right, a sidebar titled "Share" is open, showing options to share the notebook:

- Get link for this notebook**: Any user with the link can view the notebook. A dropdown menu shows "View access link" and a URL: <https://datalore.jetbrains.com/notebook/>. A "Copy link" button is present.
- Share this notebook with someone by email**: Invite people by email and collaborate with them live. An "Add and manage people" button is present.
- Update published copy**: Use the link below to open the static copy of this notebook and manage its visibility. A URL is shown: <https://datalore.jetbrains.com/view/notebook/41>. An "Update" button is present and circled with a hand-drawn arrow.

Share your work,
create portfolio



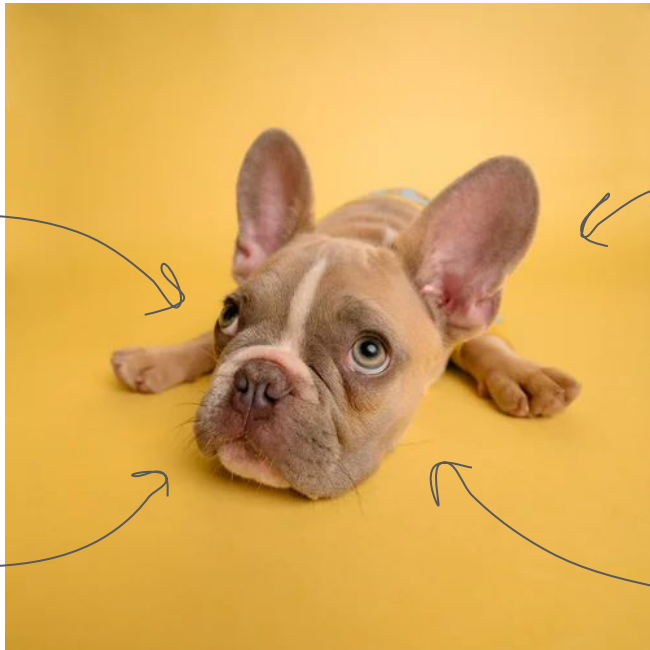
OOP

Name

Age

Color

Breed



sitting(), barking(), running()



JSON

JavaScript Object Notation

```
File Edit Selection View Go Run Terminal Help
data.json - Visual Studio Code

data.json x
C: > Users > eskim > Desktop > {} data.json > ...
1  {
2    "ebook": ["Getting started with Python", "Introduction to R", "SQL for Beginners"],
3    "language": ["python", "r", "sql"],
4    "amazonRating": [4.89, 4.88, 4.75]
5  }
```



Mean Imputation

	species	island	bill_length_mm	bill_depth_mm	flipper_length_...	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
3	Adelie	Torgersen	nan	nan	nan	nan	nan
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE

5 rows × 7 columns

Replace nan with
column mean



numpy

Numerical Python <https://numpy.org/>

Quantum
Computing



QuTiP
PyQuil
Qiskit

Statistical
Computing



Pandas
statsmodels
Xarray
Seaborn

Signal Processing



SciPy
PyWavelets
python-control

Image
Processing



Scikit-image
OpenCV
Mahotas

Graphs and
Networks



NetworkX
graph-tool
igraph
PyGSP

Astronomy
Processes



AstroPy
SunPy
SpacePy

Cognitive Psychology



PsychoPy

Bioinformatics



BioPython
Scikit-Bio
PyEnsembl
ETE

Bayesian
Inference



PyStan
PyMC3
ArviZ
emcee

Mathematical
Analysis



SciPy
SymPy
cvxpy
FEniCS

Chemistry



Cantera
MDAnalysis
RDKit

Geoscience



Pangeo
Simpeg
ObsPy
Fatiando a Terra

Geographic
Processing



Shapely
GeoPandas
Folium

Architecture &
Engineering



COMPAS
City Energy Analyst
Sverchok



IF else using numpy

```
import numpy as np
```

```
np.where( score > 80, True, False )
```

 a condition



Merge DataFrames

left

key	name	age
1	toy	25
2	joe	28
3	jane	30
4	anna	22

right

key	city	zip
1	Bangkok	1001
2	London	2504
3	Seoul	2094
4	Tokyo	9802

result

key	name	age	city	zip
1	toy	25	Bangkok	1001
2	joe	28	London	2504
3	jane	30	Seoul	2094
4	anna	22	Tokyo	9802

```
result = pd.merge(left, right, on="key")
```

https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html



Course Review

- Python Language
- Pandas Basics
- Pandas Intermediate (Data Analyst)
- Next - Final Project





Final Project

1. Go to my datalore template

<https://datalore.jetbrains.com/view/notebook/3h5IZr0I3bZAKLkLB3omTf>

2. Upload csv file [sample-store.csv]



3. Write pandas to answer all questions in this notebook

4. Share your work online

The screenshot shows the Datalore web interface. The top bar includes the DataCamp logo, navigation links (File, Tools, Kernel, View, Run, Help), the course title 'Course Pandas Foundation - Final Project (Template)', a 'Share' button, and user avatars. The left sidebar contains icons for file explorer, attachments, and a list of notebooks, with the last one marked with a red 'x'. The main area displays a notebook titled 'Final Project - Analyzing Sales Data'. Below the title, it lists 'Date: 29 December 2021', 'Author: Kasidis Satangmongkol (Toy DataRockie)', and 'Course: Pandas Foundation'. A handwritten note with an arrow points to the author name, saying 'Change date and author name to yours'. At the bottom, a code cell [22] contains the following code:

```
# TODO - import data
import pandas as pd
```



Final Project

TODO 01 - how many columns, rows in this dataset

TODO 02 - is there any missing values?, if there is, which column? how many nan values?

TODO 03 - your friend ask for 'California' data, filter it and export csv for him

TODO 04 - your friend ask for all order data in 'California' and 'Texas' in 2017 (Look at Order Date), send him csv file

TODO 05 - how much total sales, average sales, and standard deviation of sales your company make in 2017

TODO 06 - which Segment has the highest profit in 2018

TODO 07 - which top 5 States have the least total sales between 15 April 2019 - 31 December 2019

TODO 08 - what is the proportion of total sales (%) in West + Central in 2019 e.g. 25%

TODO 09 - find top 10 popular products in terms of number of orders vs. total sales during 2019-2020

TODO 10 - plot at least 2 plots, any plot you think interesting :)



Pandas Foundation

Thank you and happy learning!