

Python For Data Science

Manaranjan Pradhan

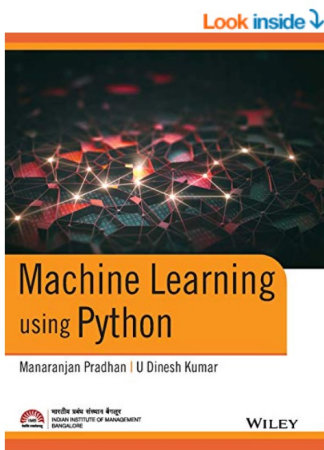
Know Your Instructor

- Has over 20+ years of industry experience
- Consulting & Training on Big Data, Machine Learning, Deep Learning & MLOps
- Have trained more than 1000+ persons on Machine Learning
- CISCO, HP, Fidelity, Goldman Sachs, TESCO, Accenture, Software AG etc.
- Visiting faculty for IIM Bangalore and ISB Hyderabad



Manaranjan Pradhan

<https://www.linkedin.com/in/manaranjanpradhan>



Machine Learning using Python [Print Replica] Kindle Edition

by [Manaranjan Pradhan](#) (Author), [U Dinesh Kumar](#) (Author) | Format: Kindle Edition

★★★★☆ 363 ratings

[See all formats and editions](#)

Kindle Edition

₹386.40

Paperback

₹494.00 [prime](#)

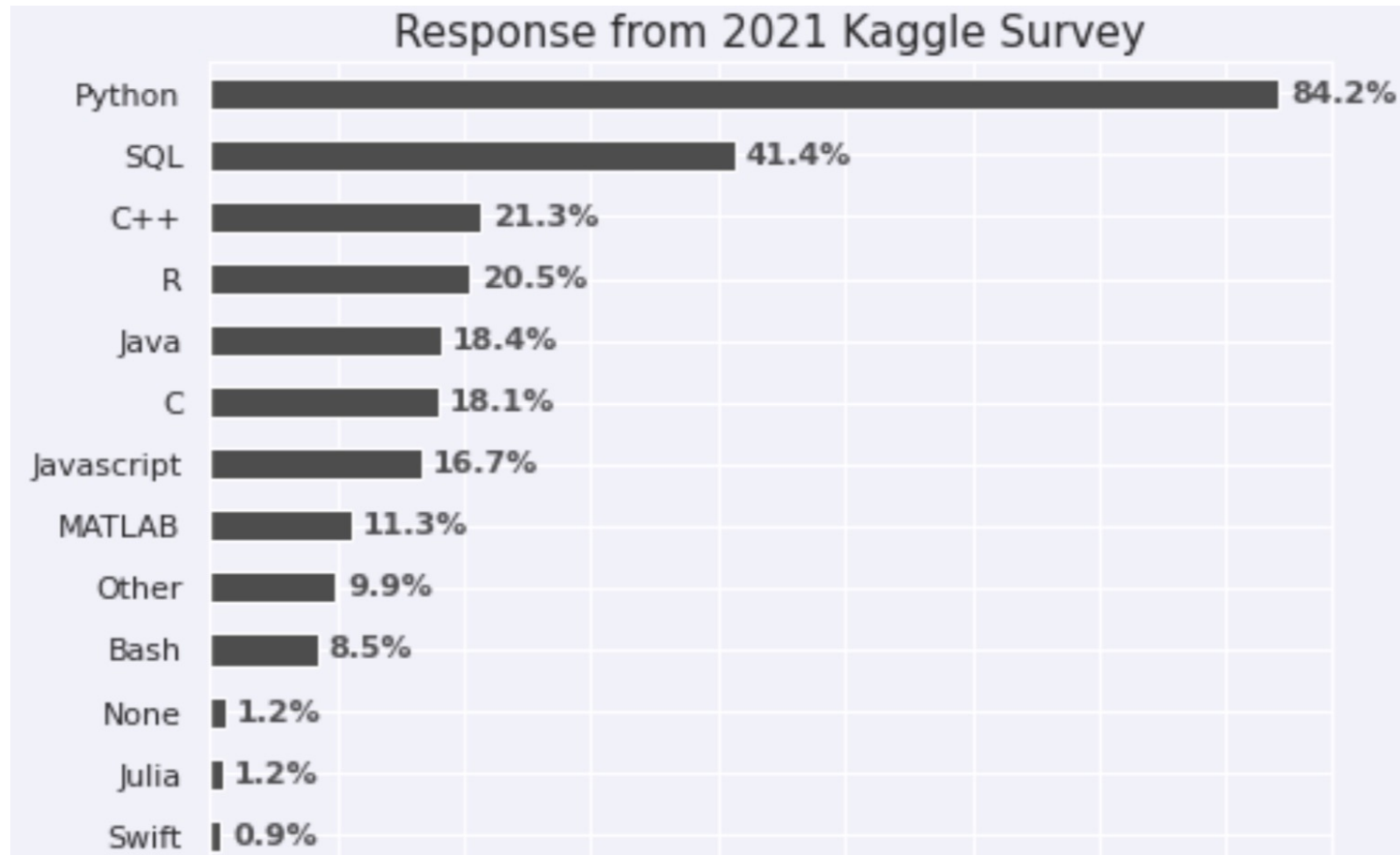
[Read with Our Free App](#)

13 New from ₹483.00

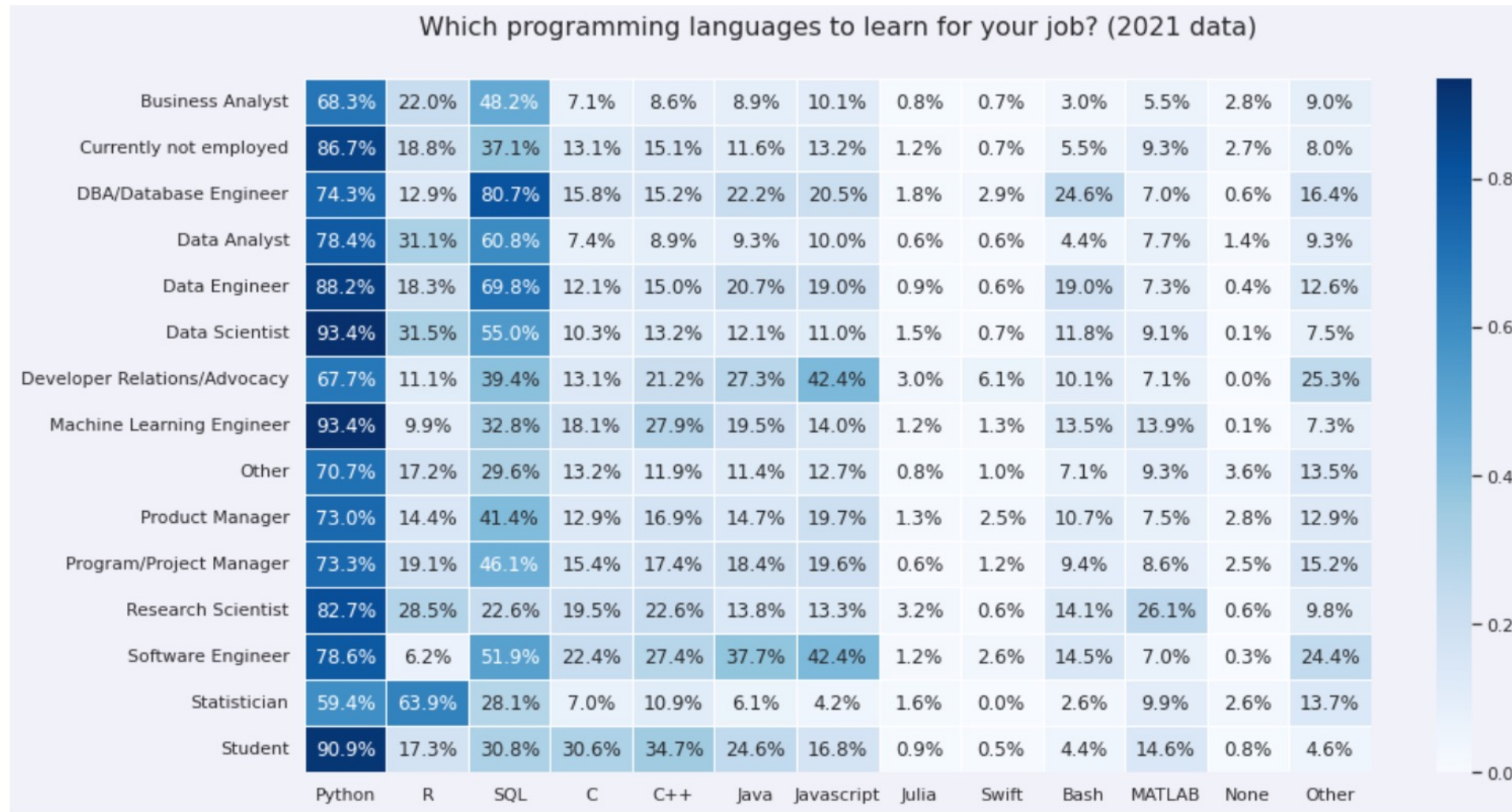
This book is written to provide a strong foundation in Machine Learning using Python libraries by providing real-life case studies and examples. It covers topics such as Foundations of Machine Learning, Introduction to Python, Descriptive Analytics and Predictive Analytics. Advanced Machine Learning concepts such as decision tree learning, random forest, boosting, recommender systems, and text analytics are covered. The book takes a balanced approach between theoretical understanding and

[Read more](#)

Why Python?

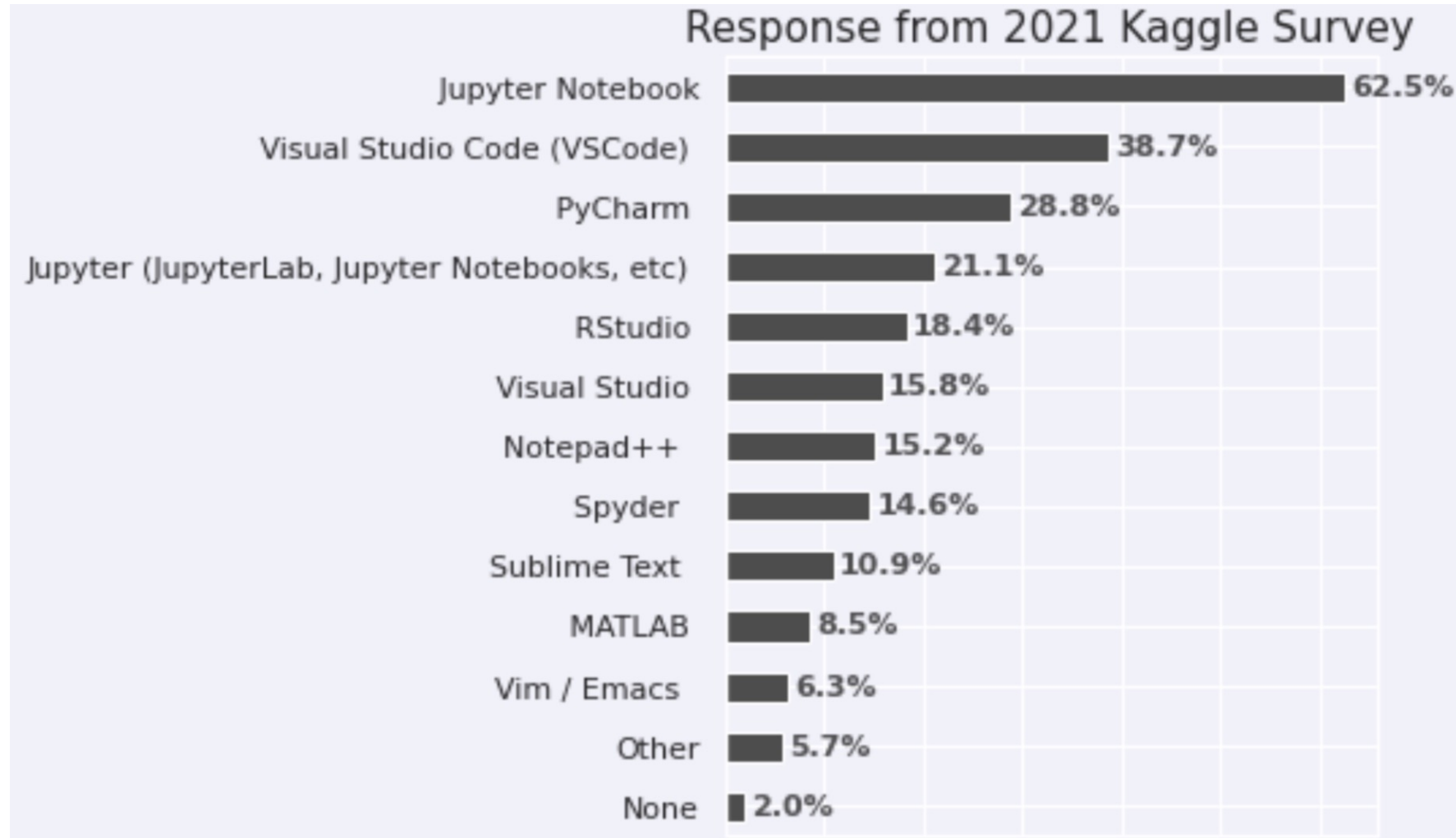


Who is using Python?



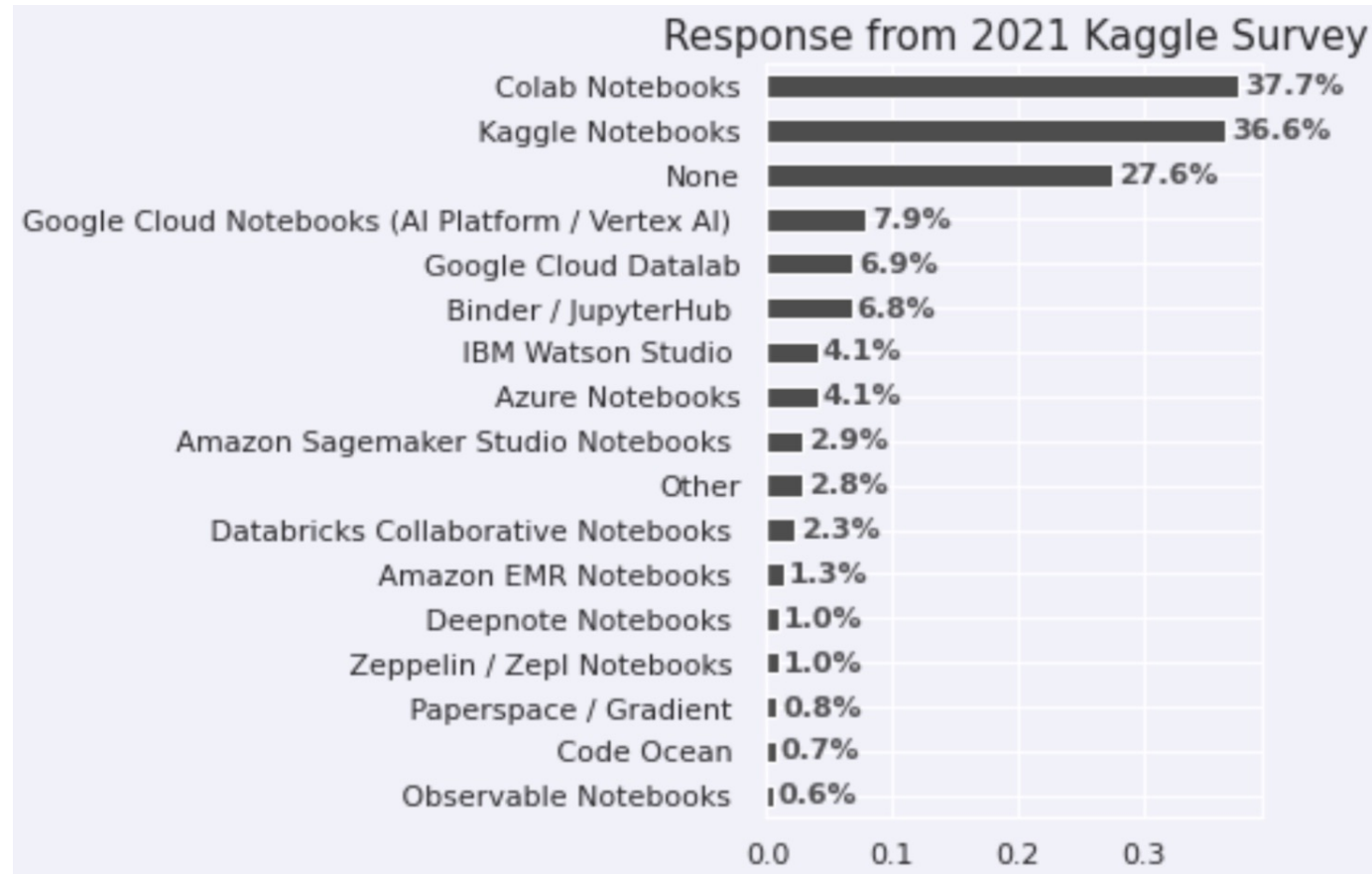
Source: <https://www.kaggle.com/code/lynnxy/a-deep-dive-into-the-kaggle-survey-from-2017-2021>

Which development tool to use?



Source: <https://www.kaggle.com/code/lynnxy/a-deep-dive-into-the-kaggle-survey-from-2017-2021>

On Cloud?



Which libraries in Python?

Efficient storage of arrays and matrices. Backbone of all scientific calculations and algorithms.

NumPy 

Statistical Analysis

 statsmodels

Plotting and visualization

 seaborn

matplotlib 

 **pandas**

Fast, powerful, flexible and easy to use open source data analysis and manipulation tool

 scikit
learn

Machine learning library.
Collection of ML algorithms.

What tools should we use?



The Jupyter Notebook is a **web-based interactive computing platform**.

<https://www.jupyter.org/>



GitHub

The Jupyter Notebook is a **web-based interactive computing platform**.

<https://www.github.com/>

Where to start?



Most popular open-source
Python distribution platform

Anaconda Distribution

Download 

For MacOS

Python 3.9 • 64-Bit Graphical Installer • 688 MB

Get Additional Installers



<https://www.anaconda.com/products/distribution>



Goole Colaboratory is a hosted Jupyter
notebook environment that is free to
use and requires no setup.

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

On my computer?



Data science technology for a better world.

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. Start working with thousands of open-source packages and libraries today.

Download 

For MacOS

Python 3.9 • 64-Bit Graphical Installer • 688 MB

Get Additional Installers



<https://www.anaconda.com/>

Python Primer for Machine Learning

Language Features

- Variables
- Conditional Statements
- Control Flows
- Functions

Collections

- List
- Tuple
- Set
- Dictionary

Functional Programming

- Lambda
- List Comprehension
- map()

Where to find the resources for this class?

- https://github.com/manaranjanp/MU_PythonForDS

The screenshot shows the GitHub repository page for `manaranjanp / MU_PythonForDS`. The repository is public and has 1 branch (main) and 0 tags. The commit history shows three commits: 'Add files via upload' (1b5ce29, now), 'Add files via upload' (now), and 'Initial commit' (11 minutes ago). The README.md file is visible, containing the title 'MU_PythonForDS' and the description 'Python for Data Science Course for Masters Union Batch'.

manaranjanp / MU_PythonForDS

Type to search

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

MU_PythonForDS Public Pin Unwatch 1

main 1 branch 0 tags Go to file Add file <> Code

manaranjanp Add files via upload 1b5ce29 now 3 commits

IntroToPython Add files via upload now

README.md Initial commit 11 minutes ago

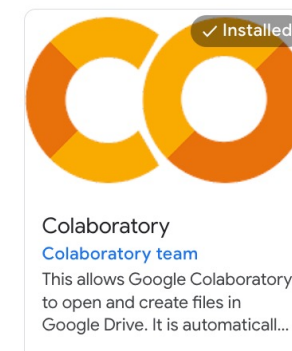
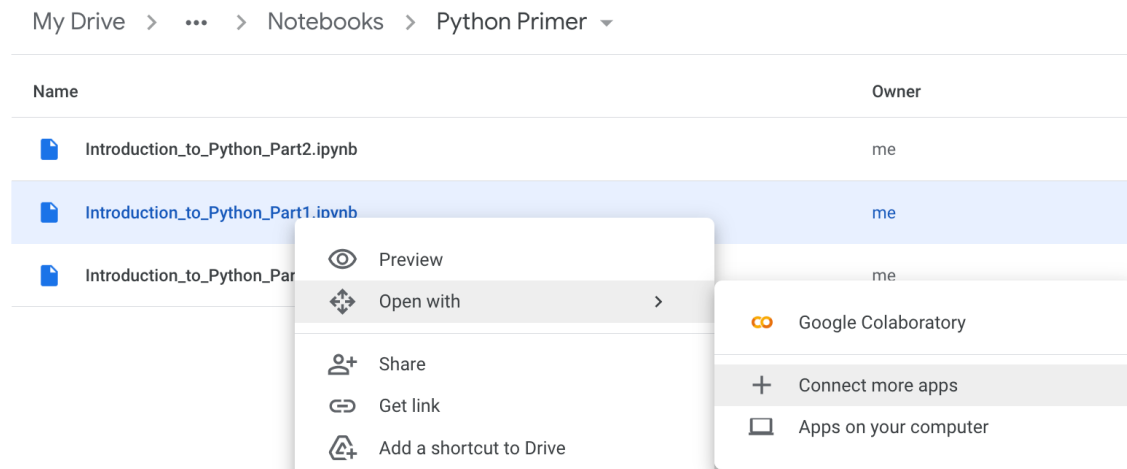
README.md

MU_PythonForDS

Python for Data Science Course for Masters Union Batch

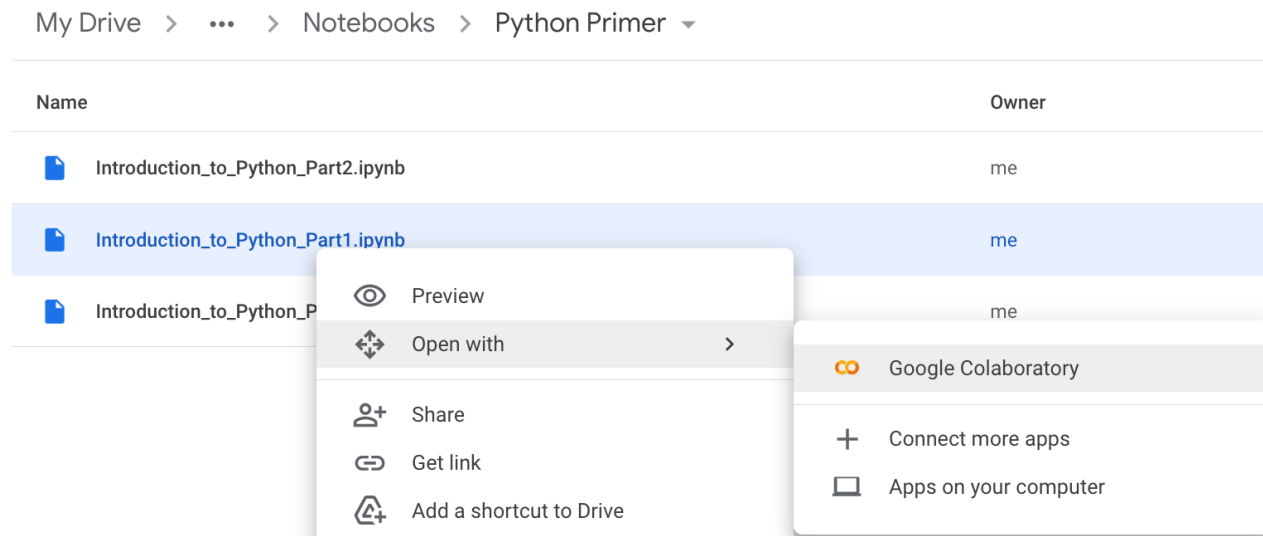
Notebook on Google Colab: Step 1

- Upload Notebooks to Google Drive
- Click on the Notebook
 - Select “**+Connect more apps**”
 - Search for Colaboratory
 - Install



Notebook on Google Colab: Step 2

- Upload Notebooks to Google Drive
- Click on the Notebook
 - Select **“Google Colaboratory”**



Notebooks on Local

- Start Jupyter notebook
 - Option 1: Enter “**jupyter notebook**” from command prompt
 - Anaconda prompt on Windows

```
(trainingnew) → Work jupyter notebook
[I 17:12:20.428 NotebookApp] [jupyter_nbextensions_configurator] enabled 0.4.1
[I 17:12:20.555 NotebookApp] JupyterLab extension loaded from /opt/anaconda3/envs/trainingnew/lib/python3.8/site-packages/jupyterlab
ab
[I 17:12:20.555 NotebookApp] JupyterLab application directory is /opt/anaconda3/envs/trainingnew/share/jupyter/lab
[I 17:12:20.557 NotebookApp] Serving notebooks from local directory: /Users/manaranjan/Documents/Work
[I 17:12:20.557 NotebookApp] The Jupyter Notebook is running at:
[I 17:12:20.557 NotebookApp] http://localhost:8888/?token=205b81da57906b8a4e3be3f624d2422fb5b1af0f3e3e9136
[I 17:12:20.558 NotebookApp] or http://127.0.0.1:8888/?token=205b81da57906b8a4e3be3f624d2422fb5b1af0f3e3e9136
[I 17:12:20.558 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 17:12:20.563 NotebookApp]

To access the notebook, open this file in a browser:
    file:///Users/manaranjan/Library/Jupyter/runtime/nbserver-17038-open.html
Or copy and paste one of these URLs:
    http://localhost:8888/?token=205b81da57906b8a4e3be3f624d2422fb5b1af0f3e3e9136
    or http://127.0.0.1:8888/?token=205b81da57906b8a4e3be3f624d2422fb5b1af0f3e3e9136
```

- Option 2: Start Anaconda Navigator -> Click on Jupyter Notebook

Notebooks on Local



Quit

Logout

Files

Running

Clusters

Nbextensions

Select items to perform actions on them.

Upload

New ▾



0 ▾ / ML Mentorship / Class / PythonPrimer

	Name	
<input type="checkbox"/>	..	
<input type="checkbox"/>	Introduction_to_Python_Part1.ipynb	
<input type="checkbox"/>	Introduction_to_Python_Part2 Answers.ipynb	
<input type="checkbox"/>	Introduction_to_Python_Part2.ipynb	

Notebook:

- Apache Toree - PySpark
- Apache Toree - SQL
- Apache Toree - Scala
- Apache Toree - SparkR
- Python 3**
- Spark 2.4.0

To run an existing notebook click on it.

Create a new notebook

Course Outline

Session #	Topic	Sub-topics
1	Python overview	Overview of Python platform and libraries
		Core Language Features
		Variable and Loops
		Using Jupyter Notebook and Google Colab
2	Advanced Language Features	Functions
		Collections
		Reading and Writing files
3	Basic Statistical Analysis	Dealing with real world data
		Basic Data Analysis
		Stock price Analysis
4	Exploratory Data Analysis using Python - 1	Data Preparation like grouping, filtering, joining, sorting
		Univariate Analysis
		Histogram, KDE Plot
		Box Plot
5	Exploratory Data Analysis using Python - 2	Bivariate Analysis
		Scatter Plot
		Correlation

Evaluation

- In Class Participation
- One Group Assignment using real world dataset