

A Guided Tour to Data Science using Python

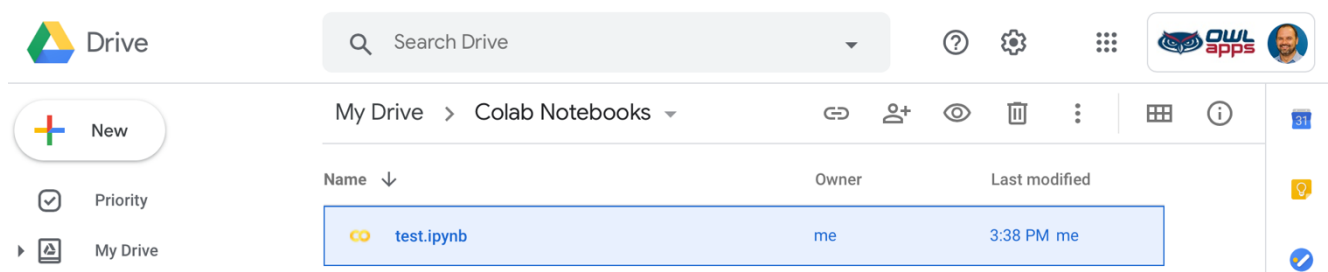
© Oge Marques, PhD – 2019

Introduction

- This document guides you through several tutorials, papers, websites, videos, books, and resources related to Data Science using Python.
- It assumes no prior exposure to Machine Learning, Data Science, or Python.
- It is structured as a step-by-step guide. It is best that you follow it in the intended sequence.

Part 1 – Setting up Google Colab(oratory)

1. Log on to your Google account using your FAU credentials.
2. Go to <https://colab.research.google.com/notebooks/welcome.ipynb> and watch the (3-min) welcome video.
3. (OPTIONAL) Spend a minute or two playing with the two code cells under the "Getting Started" portion of the welcome notebook.
4. Download a test notebook (**test.ipynb**, available on Canvas) to your computer.
5. Upload the test notebook (**test.ipynb**) from your computer and run it.
6. Locate your Colab Notebooks folder in your Google Drive and confirm that the test notebook is there. You should see a screen like this:



Part 2 - Learning the basics of Jupyter notebooks

1. (OPTIONAL) Learn more about IPython (Jupyter's predecessor) at <https://ipython.org/>
2. (OPTIONAL) Bookmark the documentation page at: <https://jupyter.org/documentation>
3. Follow the steps in the notebook at https://colab.research.google.com/notebooks/basic_features_overview.ipynb
4. Read Chapter 1 of "textbook 1" (VanderPlas) and use the companion notebook (on Colab: <https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/01.00-IPython-Beyond-Normal-Python.ipynb>) to practice the main concepts, commands, etc.

Part 3 - Building a Python environment

There are good tutorials, videos, and explanations online on how to set up a Python environment for your machine. Your choice of editor, IDE, and environment is ultimately yours, of course, but I recommend Anaconda (<https://www.anaconda.com/>) as a Python distribution environment (and Conda <https://conda.io/docs/index.html> as a package manager), due (among other things) to their popularity in the data science community.

1. Watch the video "Which Python Package Manager Should You Use?" <https://youtu.be/3J02sec99RM>
2. If you choose to use Anaconda (as I recommend), follow the steps at <https://www.anaconda.com/distribution/> to download, install, and configure Anaconda in your machine.
3. (OPTIONAL) For the IDE, you may want to consider PyCharm (<https://www.jetbrains.com/pycharm/features/>) – **free** for FAU students, see <https://www.jetbrains.com/student/> for details – or any other editor/IDE you may prefer.
4. If you want to install and learn how to use Jupyter Notebooks in your own computer, follow the steps at: <http://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html>

Part 4 – Learning (or refreshing your knowledge of) the basics of Python

[Skip this part if you have already taken Python programming classes and/or used Python in other courses, projects, and work-related tasks.]

1. Go through the Jupyter notebooks (available at: <https://github.com/jakevdp/WhirlwindTourOfPython/>) for the book: "A Whirlwind Tour of Python," by Jake VanderPlas (available at: <https://www.oreilly.com/learning/a-whirlwind-tour-of-python>). It is a great summary and works both as refresher as well as "quick reference".
2. (OPTIONAL) Take the Google's Python class (FREE): <https://developers.google.com/edu/python/>
3. (OPTIONAL) Take the "Python Programming: A Concise Introduction" course (FREE, with the "no certificate" option) at Coursera: <https://www.coursera.org/learn/python-programming-introduction>
4. (OPTIONAL) Take the "Introduction to Computing using Python" course (FREE, with the "no certificate" option) at edX: <https://www.edx.org/course/introduction-computing-using-python-gtx-cs1301x>
5. (OPTIONAL) Take the Udacity "Introduction to Python" (FREE) course: <https://www.udacity.com/course/introduction-to-python--ud1110>
6. (OPTIONAL) Follow the LearnPython.org interactive Python tutorials: <https://www.learnpython.org/>

Part 5 - Learning Numpy

1. Read Chapter 2 of "textbook 1" (VanderPlas) and use the companion notebook (on Colab: <https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/02.00-Introduction-to-NumPy.ipynb>) to practice the main concepts, commands, etc.
2. (OPTIONAL) Bookmark this page for quick access to official NumPy documentation: <https://docs.scipy.org/doc/numpy/index.html>
3. (OPTIONAL) Take the DataCamp "Intro to Python for Data Science" course (FREE): <https://www.datacamp.com/courses/intro-to-python-for-data-science>

Part 6 - Learning Pandas

1. Read Chapter 3 of "textbook 1" (VanderPlas) and use the companion notebook (on Colab: <https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/03.00-Introduction-to-Pandas.ipynb>) to practice the main concepts, commands, etc.

Part 7 - Learning Matplotlib

1. Read Chapter 4 of "textbook 1" (VanderPlas) and use the companion notebook (on Colab: <https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/03.00-Introduction-to-Pandas.ipynb>) to practice the main concepts, commands, etc.
2. (OPTIONAL) Take the DataCamp "Intermediate Python for Data Science" course (FREE): <https://www.datacamp.com/courses/intermediate-python-for-data-science>
3. (OPTIONAL) Take the "Introduction to Computing using Python" course (FREE, with the "no certificate" option) at edX: <https://www.edx.org/course/introduction-python-data-science-microsoft-dat208x-8>
4. (OPTIONAL) Take the Udacity "Intro to Data Analysis" (FREE) course: <https://www.udacity.com/course/intro-to-data-analysis--ud170>
5. (OPTIONAL) Take the (FREE) Udemy course "Deep Learning Prerequisites: the Numpy Stack in Python": <https://www.udemy.com/deep-learning-prerequisites-the-numpy-stack-in-python/>
6. Work on **Assignment 1: The Python data science stack**