

DSL 810 (Data Driven Design)

Instructor: Dr Jay Dhariwal, Dept of Design, IIT Delhi

This handout describes how to get started with *Python for Data Science* and the resources available to learn further on your own.

We have already introduced you to learning Data Science in MATLAB at length. We have also given you video tutorial for R. We wanted to introduce you to Python for Data Science. Some of you already know Python for Data Science but for others, we would like them to know about it. After this, you are free to choose the data science platform of your choice.

Introduction to Google Colab

We would use Google Colaboratory (Colab) for working with examples in Python for Data Science. It is similar to MATLAB online's Live script editor which we use to write code, output and formatted text together in an executable notebook. In general, the concepts remain the same while the software can be chosen based on own preference. Those of you, who have followed MATLAB so far, can relate the concepts from MATLAB to Python for Data Science to help make it easier to pick it up.

For those who have worked with Python for Data Science, Colaboratory is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud. With Colaboratory, we can write and execute code, save and share analyses, and access powerful computing resources, all for free from your browser.

To begin, please go to <https://colab.research.google.com/notebooks/intro.ipynb>

Go through this Colab notebook to get started.

Python, Numpy, Pandas, Matplotlib basics

Python for Data Science assumes that you know the Python language. You can revise the basics of Python and the data science specific modules from <https://www.w3schools.com/python/default.asp>.

Another online resource for Python, Numpy and Pandas is <https://pythonexamples.org/>

Python for Data Science MOOC

We can select from the freely available MOOCs on *Python for Data Science* which we could give a crash course on and the students could refer to that MOOC for details.

For instance, this [link](#) gives a link of free AI/ML/DS MOOCs on the internet.

This [MOOC \(Python for Data Science!\) on edx](#) from UCSD seems to be well aligned to our course objectives. We would be introducing this MOOC to you.

We were also curious to know if what all we covered in the course so far was aligned with the best practices from around the world.

We can also go through Linear Regression example for the Life Expectancy dataset used earlier and the marks dataset for Logistic Regression.

How to know which machine learning model to use and when?

https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html