

Python for Data Science

Instructor: Michelle Miller

Goal: The purpose of this course will be to give the students an introduction in working with Python as a data analysis tool. The world of data science has exploded exponentially in recent years and the need for understanding how to use code to analyze data is an imperative skill for students to learn as they develop their careers. In this course they will walk away with a working knowledge of Python. Further, they will be able to load data into Python and write code to analyze said data. This includes being given some of the mathematical tools to work with data and time permitting the algorithms used to do more advanced analysis. They will be able to be given a data set and being thinking about how to frame their own questions of how to analyze said data.

Meetings: Sessions will be twice a week on Mondays and Fridays at 10am EST for approximately 60 minutes depending on the amount of material needed to be covered. This can include writing code tailored to sessions, working through students' code, answering questions, preparing material to hone in on specific topics or answer specific questions, handling hardware, etc.

Structure: The course will mostly consist of sessions via Zoom. The course will begin by presenting the first couple of weeks reviewing the basics of python. This will include data types, data structures, simple algorithms, etc. It is important the students feel more comfortable with Python before they are able to do much computation with it. Assignments will be given based on the relevance of the lectures to challenge the students' understanding of lecture. They will have weekly assignments and based on how motivated the students are, I will incorporate more nontrivial problems. Though the overall pace of this course will be set by the pace of the students, this course is anticipated to take anywhere from 5-6 weeks.

Material:

- Key Texts:
 - [Python Data Science Handbook:](#)
 - <https://jakevdp.github.io/PythonDataScienceHandbook/>
 - [Fundamentals of Python:](#)
 - This is the same textbook used by the intro python course at Harvard for undergraduates. If we manage to get through the first 11 chapters of this book, they will have the equivalent of a first semester course in programming
- We will be following the materials from these books as guidance for the course material.

I will be providing lessons through a mix of powerpoints and jupyter notebooks (python code meant to show the output of the code in line – often used as a teaching tool). The students will learn the fundamentals of Python as well as how to use Python for basic data analysis. This will build off the lessons done during the IGNITE program in which they will learn how to run some hypothesis tests in Python. They will learn more advanced statistical techniques using Python packages (numpy, sci-kit learn, pandas) and how to visualize it. Time permitting and depending on how challenging they find this work, they will be introduced to some introductory machine learning topics.

- Topic 01 – Review of fundamentals in Python

- Topic 02 – Working with Data in Python
- Topic 03 - Plotting in Python/Visualization
- Topic 04 - Working with NumPy Arrays
- Topic 05 – Functions and Classes in Python
- Topic 06 – Basic Statistical Analysis (Statistical Moments)
- Topic 07 – Curve Fitting data (Spline Fitting) ; Regression
- Topic 08 – Probability Theory and Data Analysis
- Topic 09 – (Time permitting) Naïve Bayes; Support Vector Machine; and/or K-means clustering and other Machine learning topics

Topics of lessons are subject to change based on the background of the students. If students prove the earlier lessons are too basic, topics in lesson 9 can easily be expanded into more lessons to allow for deeper challenges.