

# **Spring 2021 BDAA DataMastery Python Syllabus**

## **Course Description**

Training in the foundational tools necessary for data analysis in Python, such as data manipulation, and data visualization. Students will be equipped with the knowledge to use software tools such as Pandas, Matplotlib, and Seaborn, to aid in any data analytics related tasks. The course emphasizes the use of hands on learning through live workshops with code along exercises to aid in retention in material. The course is structured to target people from all different backgrounds, and with a wide variety of computer science experience, from beginner to expert.

## **Course Learning Outcomes**

By the end of this course, students should successfully be able to:

- Understand what data science is, and the overall workflow in a project
- Use Pandas for data manipulation and data cleaning
- Explore data with data visualization techniques using matplotlib and seaborn
- Be able to take a dataset, frame a question, and use the given tools to tell a story with data

## **Course Technology**

- Computer: current Mac (OS X) or PC (Windows 10+) with high speed internet connection
- Webcam: built-in or external webcam, fully installed
- Microphone: built-in laptop or tablet mic or external microphone

## **Necessary Software**

- This course will be using Google Colaboratory as the Python Environment for workshops.
- Many of you may already have a gmail account, but in case you don't make sure this is done

## Course Delivery

Workshops will be held every Thursday on Zoom, from 7:30pm-8:30pm. The workshops will be recorded, and uploaded onto the Canvas page for any asynchronous learners. Further resources to dive deeper into topics covered in a workshop will be posted as a separate module, such as books and free courses. After 2 weeks of instruction, there will be a short, brief carmen quiz to check for understanding. Note: This is not something to stress over, rather it is just to make sure there is a general understanding of the processes and ideas that were covered in the workshop. These will usually be small coding snippets on a small dataset.

## Course Content

### **Intro to the Data Science Workflow + Intro to Data Manipulation in Python**

- What is Data Science? What are its applications?
- What is the Data Science Workflow?
  - How does Data Manipulation fit?
- Why do we do Data Manipulation and Data Cleaning?
- Introduction to Pandas for Data Manipulation
- Live Coding Exercises
  - How to inspect a data frame
  - How to check properties of columns
  - Calculating summary statistics
  - Subsetting Data
  - Renaming Columns

### **Intro to Data Visualization in Python**

- How does Data Visualization fit in the workflow?
- Why is Data Visualization Important?
- Exploratory Data Analysis with Matplotlib
- Live Coding
  - Histogram
  - Line Plots
  - Scatter Plots
  - Barcharts

## **Intermediate Data Manipulation in Python**

- Advanced subsetting methods
  - Loc, iloc
  - Query
  - Goupbys
  - Aggregation functions
  - List Comprehensions
  - Iterating over rows
- DataFrame operations
  - Apply functions
  - Lambda functions
  - Data Reshaping
  - Data Imputation

## **Intermediate Data Visualization in Python**

- Introduction to Seaborn for Data Visualization
  - Stacked Barcharts
  - Faceted Plots
  - Scatter Plots with lines
  - Time Series Graphs
  - Colors and visual aesthetics

## **Advanced Data Manipulation in Python**

- Joining Data Frames in Pandas
- Methods for joining Data Frames
  - Joining on a key
  - .merge
  - Inner Joins
  - Left Joins
  - Right Joins
  - Outer Joins
- Concatenating DataFrames
  - Horizontally
  - Vertically