Hands-on: Data Analysis and Machine Learning with Python

Course Outline

Part 1: Introduction to Python

- Introduction to Python and Jupyter Notebooks
- Variable Types
  - Numbers
  - Strings
  - Lists
  - Tuples
  - Dictionaries
  - Booleans
- Control Flow
  - If-Else
  - While Loops
  - For Loops
  - List Comprehensions
- Functions
  - Built-in functions
  - Importing python packages/modules and using their functions
  - Writing functions
  - Lambda Functions
- Simple Data I/O
  - Reading and writing to text files

Part 2: Data Analysis with Python

- Introduction to Jupyter
- What a Jupyter notebook is
  - How to create and edit a Jupyter notebook
  - How Jupyter notebooks can be effectively used in the workplace
- Loading Data into a Pandas DataFrame
  - Loading data from text files (csv, tab delimited, whitespace delimited)
  - Loading data from databases
  - Tips and tricks for loading data
- Describing Data
  - Describe a datasets size, shape, and variable types
- Selecting Data from a DataFrame
  - Select data by position
  - Select data by conditions
- Handling Dates in a DataFrame
  - Convert dates in data to datetime objects
  - Calculate time deltas
  - Convert time zones
- Manipulate Data DataFrame
  - Update values, delete values, drop rows and columns
- Analyzing Groups of Data
  - Easily sort data into groups and perform group wise analyses and comparisons

- Custom Analysis (apply your own functions to DataFrames)
    - Go beyond the built in analysis functions and write your own formulas and algorithms
- Merging DataFrames
    - Combining dataframes
- Basic Data Visualization
    - Create line plots, histograms, and bar plots
    - Format plot labels and axes

Part 3: Developing with Python

- Introduction to Visual Studio Code for Python
    - Explore the basics of Visual Studo Code and learn how you can use it
- Write scripts that accept command line arguments for reusability and automation
- Write python modules for organized and reusable code
- Learn how to write and install your own python packages
    - Learn how these can be co-developed and used across your team for increased efficiency and productivity
- Testing your code with pytest
    - Learn why tests are important and how to easily implement them in python
- Effectively incorporating logging in you code
    - Learn how to set up logging and handle different levels of logging messages (error, info, debug, etc...)

Part 4: Machine Learning with Python

- Introduction to Jupyter
- Collecting and Describing Data
- Exploring Data
- Training Regression and Classification Models
    - Logistic Regression
    - Decision Trees
    - Linear Regression
    - Neural Networks
    - K-Nearest Neighbors
- Using Cross Validation To Tune Hyper parameters
    - What Cross Validation Is and Why It Is Important
- Preventing Overfitting
    - Strategies to prevent overfitting
    - What regularization is and how to use it.
- Testing Models
    - How to test your models and what to consider when aligning tests with real world business problems.

Part 5: Workshop

- Workshop scenario and data will be identified during pre-course planning