Introduction to Python for Exploratory Data Analytics (EDA)

In the past few years, Python has emerged as a choice tool in the professional world, from basic data analysis and model building to Big Data Analysis and generating complex machine learning (ML) algorithms. This course offers a foundational understanding and hands-on expertise with Python in data analytics/visualization. Beginning with the basic knowledge of variables to store/retrieve/calculate values to data cleanup techniques, the course aims to introduce the core Python programming tools (functions and loops) for complex data analysis and visualization.

Totally based on hands-on exercises with Python, this course would introduce methods for:

− Data cleaning and preprocessing (e.g. handling missing values, formatting, normalizing, binning)
− Data manipulation using dataframes, summarizing data, and understanding data distribution, creating new information
− Performing exploratory data analysis with real-word datasets
− Performing data visualization and presentation

In the process, the course would introduce various data libraries such as

• Pandas
• Numpy
• Seaborn
• Matplotlib

Expertise with above impart market-based soft-skill development ideal for aspiring data professionals for the industry as well as academia.

The course would operate on a 'continuous assessment' mode with 4-5 hands-on exercises (65-70 pts in total), and a final exercise (~30% of the total). The exercises would be open book/notes, and either group assignments or individualistic, based on circumstances. Reading materials would be provided during lectures/labs as necessary.

PREREQUISITES: NO PRIOR EXPERIENCE required with Python; Interest in data science and quantitative analysis; willingness to write codes.