AICTE EY-GDS Internship December'23
Self-Paced Study Material for Internship:
Topic: Exploratory Data Analytics (EDA)

Introduction to EDA:

- Exploratory Data Analysis (EDA) involves systematically analyzing and visualizing data to discover patterns, anomalies, and insights, playing a crucial role in understanding the underlying structure of the data.

Importing and loading data:

- Data can be imported into Python using various formats such as CSV, Excel, or SQL, providing a foundation for EDA and subsequent analysis.

Data cleaning and preprocessing:

- Cleaning and preprocessing steps, including handling missing values, outliers, and inconsistencies, are essential for ensuring the accuracy and reliability of the data.

Descriptive statistics:

- Descriptive statistics, encompassing measures of central tendency and dispersion, offer a summary of the main characteristics of the dataset.

Data visualization:

- Visualizations like histograms, boxplots, and scatter plots provide a powerful means to explore data distributions, relationships, and outliers, enhancing the interpretability of the dataset.

Identifying patterns and relationships:

- EDA enables the identification of patterns and relationships within the data, helping to uncover hidden insights and guide subsequent analysis.

Univariate and bivariate analysis:

- Univariate analysis focuses on individual variables, while bivariate analysis explores relationships between pairs of variables, offering a comprehensive understanding of the dataset's structure.

Feature engineering:

- Feature engineering involves creating new features from existing data, enhancing the dataset with additional information to improve the performance of machine learning models.

Hypothesis generation:

- EDA findings often lead to hypothesis generation, fostering a deeper understanding of the data and guiding further research questions or analytical approaches.

Resources:

1. DataCamp's Exploratory Data Analysis with Python: https://www.datacamp.com/courses/introduction-to-data-science-in-python
2. Pandas Tutorial for Data Analysis: https://pandas.pydata.org/docs/getting_started/intro_tutorials/
3. Seaborn Data Visualization Tutorial: https://seaborn.pydata.org/tutorial.html
4. Hands-On Exploratory Data Analysis with Python: https://www.oreilly.com/library/view/python-for-data/9781449323592/
5. Exploratory Data Analysis for Real-World Projects: https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2