**Week 6:  
  
Tasks & Resources :-**

“Welcome to the Week 6 of PrepInsta’s Data Analytics Internship program.

This task enables you to work in fields like data profiling, missing data handling, and visual exploration to uncover insights and patterns within the data.

**Project 6:- Python Exploratory Data Analysis**

**Objective:**

Engage in exploratory data analysis (EDA) by analyzing a real-world dataset. Your goal is to perform tasks such as data profiling, missing data handling, and visual exploration to uncover insights and patterns within the data.

**Steps to Perform:**

* **Dataset Selection**:

Choose a real-world dataset for analysis. It could be related to any field of interest—healthcare, finance, social sciences, etc.

Ensure that the dataset is comprehensive and has enough complexity to allow for meaningful exploration.

* **Data Profiling**:

Conduct data profiling to gain an initial understanding of the dataset’s structure.

Examine the data types, unique values, and basic statistics of each column.

* **Missing Data Handling:**

Identify and handle missing data in the dataset. Employ strategies such as imputation or removal of missing values based on the nature of the data.

* **Data Cleaning:**

Clean the dataset by addressing any inconsistencies, errors, or outliers.

Document the steps taken for data cleaning and the rationale behind each decision.

* **Descriptive Statistics:**

Calculate and present descriptive statistics for key variables in the dataset.

Utilize statistical measures to describe the central tendency, dispersion, and shape of the data.

* **Visual Exploration:**

Create visualizations to explore relationships and patterns within the data.

Utilise charts, graphs, and other visual tools to represent the dataset’s characteristics effectively.

* **Correlation Analysis:**

Perform correlation analysis to identify relationships between variables.

Interpret the correlation coefficients and understand the implications for the dataset.

* **Insights and Patterns:**

Extract meaningful insights and patterns from your analysis.

Formulate hypotheses or conclusions based on the observed trends in the data.

* **Documentation:**

Document your entire EDA process, including the dataset selection, data profiling, cleaning steps, and the rationale behind your analytical choices.

Create a report summarising your findings and insights.

**Pre-requisites**

* Basic understanding of data analysis concepts.
* Proficiency in using tools like Python, Jupyter Notebooks, or any other preferred data analysis tool.
* Knowledge of descriptive statistics and data visualization techniques.

Note:- In case you want to revise the pre-reqs, just head over to the resources section for a quick brush up.

**What you need to do?**

**Good Practices:**

* Ensure reproducibility by clearly documenting and commenting on your code.
* Experiment with different visualization styles to effectively communicate your findings.
* Collaborate with peers or mentors to gather diverse perspectives on the dataset.
* Consider the ethical implications of your analysis, especially if the dataset contains sensitive information.

This task is designed to assess your ability to explore and understand real-world datasets. It emphasizes the importance of thorough data profiling, handling missing data, and extracting meaningful insights through visual exploration. Approach the task with curiosity and a critical analytical mindset. Best of luck!