Task 2 – Exploratory Data Analysis (EDA)

Introduction:
Exploratory Data Analysis (EDA) is the process of analyzing datasets to summarize their main characteristics, often using visual methods. It helps in understanding the data's structure, identifying patterns, spotting anomalies, and checking assumptions before applying any machine learning model. EDA is a crucial step in the data analytics process because it provides deep insights into data and helps in better decision-making. **Objective:**
The main objective of this task is to perform EDA on the collected dataset to understand its behavior, identify missing values, detect outliers, and visualize relationships among variables. EDA helps to convert raw data into meaningful information through summary statistics and visual exploration. **Steps Performed:**
1. Imported required Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn.
2. Loaded the dataset using Pandas.
3. Checked dataset shape, column names, and data types.
4. Handled missing values and removed duplicate records.
5. Used summary statistics to understand central tendency and dispersion.
6. Detected outliers using boxplots and handled them appropriately.
7. Visualized relationships between variables using graphs and charts.
Data Visualization:
Visualization plays a vital role in EDA. Various types of graphs were used such as:
- Histogram: To show data distribution.
- Box Plot: To identify outliers.
- Bar Chart: To compare categorical values.
- Heatmap: To show correlation between variables.
- Scatter Plot: To analyze relationships between two numerical variables.
These visualizations helped in understanding the patterns and hidden relationships within the dataset. **Insights:**
From the analysis and visualizations, meaningful insights were drawn such as the presence of missing or skewed data, the relationship between dependent and independent variables, and overall data trends. The analysis provided a clear picture of which factors influence the target variable and how they interact. **Conclusion:**
This task enhanced my understanding of how to explore and analyze datasets using Python. I learned how to use visualization tools to find meaningful insights and prepare data for further modeling or prediction. Exploratory Data Analysis is an essential step for any data-driven project as it ensures data quality and improves the reliability of results.