COURSE - 3 PYTHON PROGRAMMING FOR DATA ANALYSIS

Lab Instructions

Objective:

Your task is to perform a comprehensive Exploratory Data Analysis (EDA) on the provided dataset. The goal is to uncover patterns, spot anomalies, test hypotheses, and check assumptions with the help of statistical summaries and graphical representations.

Tools:

- Python (Colab Notebook)
- Libraries: Pandas, NumPy, Matplotlib, Seaborn

Dataset:

1. **Walmart Dataset**

    1.1. **Order ID**: Identifier for the order (string)
    1.2. **Order Date**: The date when the order was placed (**datetime64**)
    1.3. **Ship Date**: The date when the order was shipped (**datetime64**)
    1.4. **EmailID**: Email address of the customer (string)
    1.5. **Geography**: Geographical information about the order, likely including country, city, and state (string)
    1.6. **Category**: The category of the product (string)
    1.7. **Product Name**: Name of the product (string)
    1.8. **Sales**: Sales amount (float64)
    1.9. **Quantity**: Quantity of the product ordered (int64)
    1.10. **Profit**: Profit amount (float64)

LAB Instructions:

1. **Load Dataset:**
    - Import the dataset using Pandas.

- Examine basic information from the dataset using pandas functions. (Informations like number of rows and columns, types of data [numerical, categorical], three quartiles, nas, etc.)

2. **Data Cleaning:**
   - Check and Handle missing values: Decide whether to fill or drop them.
   - Identify and resolve duplicate data entries.
3. **Descriptive Statistics:**
   - Analyze measures like mean, median, mode, range, variance, and standard deviation for numerical data.
4. **Data Visualization:**
   - Use charts and plots (like histograms, scatter plots, box plots, bar charts, pie charts) to visualize data distributions and relationships.
5. **Identifying Relationships:**
   - Explore correlations between different variables.
6. **Anomaly Detection:**
   - Identify and investigate any outliers or unusual data points
7. **Data Discovery:**
   - **Trend Analysis**:
     i. Analyze the sales and profit trends over the years. Are there any noticeable patterns or seasonal variations?
     ii. Determine the product category that has shown the most growth in terms of sales over the years.
   - **Customer Analysis**:
     i. Identify the top 5 customers based on the number of orders placed and total sales generated. What insights can you derive about the buying behavior of these customers?
     ii. Analyze the repeat purchase behavior by calculating the average time between orders for each customer.