DATA ANALYSIS AND DATA SCIENCE WITH PYTHON

TASK - 2

Exploratory Data Analysis (EDA)

Objective

Perform an in-depth exploratory data analysis (EDA) on a dataset to identify trends, patterns, anomalies, and factors influencing performance.

Project 1: General EDA

Steps to Follow

1. **Dataset Selection**
    - Choose a dataset like "Global Superstore" containing columns such as Sales, Profit, Region, and Product Categories.
2. **Tasks to Perform**
    - **Clean Data:**
        - Handle missing values by filling them with appropriate measures (mean, median, or placeholders) or by removing affected rows/columns.
        - Remove duplicates to ensure the dataset's integrity.
        - Detect and handle outliers using statistical techniques (e.g., IQR or Z-scores).
    - **Statistical Analysis:**
        - Use measures like mean, median, standard deviation, and variance to understand the data distribution.
        - Compute correlations between variables to study relationships.
    - **Data Visualization:**
        - Use histograms to explore distributions of numerical data.
        - Use boxplots to identify outliers in continuous variables.
        - Use heatmaps to visualize correlations and relationships between features.

3. **Deliverables**

    - A cleaned dataset free from missing values, duplicates, and outliers.
    - A summary report highlighting trends, patterns, and anomalies.
    - Visualizations: Histograms, boxplots, heatmaps, and other relevant graphs.

Project 2: Sales Performance Analysis

Objective

Analyze sales data to identify trends, relationships, and factors affecting sales performance.

Steps to Follow

1. **Dataset Selection**

    - **Dataset Name:** `sales_data.csv`
    - **Columns:**
        - Product, Region, Sales, Profit, Discount, Category, Date
2. **Tasks to Perform**

    - **Load and Explore the Dataset:**

        - Use libraries like Pandas and NumPy to load and inspect the dataset (shape, missing values, data types).
    - **Data Cleaning:**

        - Remove duplicates using `drop_duplicates()`.
        - Fill missing values using appropriate strategies like the mean or median.
        - Convert the Date column to a `datetime` object for trend analysis.
    - **Exploratory Data Analysis:**

        - Plot time series graphs to observe trends in Sales over time.
        - Use scatter plots to study the relationship between Profit and Discount.
        - Visualize sales distribution by Region and Category using bar plots or pie charts.
    - **Predictive Modeling:**

        - Train a Linear Regression Model to predict Sales using Profit and Discount as features.

Main Flow Services and Technologies Pvt. Ltd.
Contact Us. +91 9389641586, +91 97736 99074
Email-Add. contact.mainflow@gmail.com
www.mainflow.in

- Evaluate model performance using metrics like R^2 score and Mean Squared Error (MSE).

Deliverables

1. **Visualizations:**
    - Sales trends over time (time series plot).
    - Scatter plot showing Profit vs. Discount.
    - Bar or pie charts showing Sales by Region and Category.
2. **Predictive Model:**
    - A Linear Regression Model capable of predicting Sales based on key variables.
3. **Insights and Recommendations:**
    - Provide actionable insights on improving sales (e.g., optimal discount rates, top-performing regions, or categories).

Expected Outcomes

- Develop the ability to clean and analyze real-world datasets.
- Gain insights into the factors driving sales performance.
- Build simple predictive models to support business decisions.
- Present findings with effective visualizations and actionable recommendations.

Deadline Compliance

- **Restriction**: **Submit the project within 7 days** from the start date.
- **Reason**: Meeting deadlines is crucial in the real-world software development environment. This restriction helps students practice **time management** and **task prioritization**. In professional settings, tight deadlines are often the norm, and learning to meet them without compromising quality is an essential skill.
- **Learning Outcome**: Students will learn to manage their time effectively, complete projects under pressure, and **deliver results on time**, which are all important skills in the workplace.

Main Flow Services and Technologies Pvt. Ltd.
Contact Us. +91 9389641586, +91 97736 99074
Email-Add. contact.mainflow@gmail.com
www.mainflow.in