

Slash Data Analysis Task

Objective:

Analyze the Amazon sales report dataset attached in the mail to extract meaningful insights, preprocess the data, create visualizations using Python libraries (matplotlib and seaborn), build predictive models, and develop a dashboard for comprehensive data presentation.

Detailed Task Breakdown

Step 1: Exploratory Data Analysis (EDA)

1. Data Inspection:

- Load the dataset and inspect the first few rows to understand its structure.
- Check the data types of each column and identify any potential issues.

2. Summary Statistics:

- Generate summary statistics for numerical and categorical variables.
- Visualize the distribution of key features to identify trends and patterns.

Step 2: Data Preprocessing

1. Handling Missing Values:

- Identify columns with missing values and decide on appropriate strategies to handle them (e.g., imputation, removal).

2. Data Type Conversion:

- Convert relevant columns to appropriate data types (e.g., converting **Date** column to datetime format).

3. Outlier Detection and Treatment:

- Identify and treat outliers in numerical columns to ensure data quality.

Step 3: Data Visualization

1. Using Matplotlib and Seaborn:

- Create visualizations to understand data distributions and relationships.
- Examples include histograms, bar plots, line plots, and heatmaps.

2. Visual Analysis:

- Visualize sales trends over time (e.g., monthly sales trends).
- Identify top-selling products and categories using bar plots.
- Analyze regional sales distributions using geographical visualizations.

Step 4: Predictive Modeling

1. Building Predictive Models:

- Develop models to predict the order status (**Shipped**, **Canceled**, etc.).
- Use classification algorithms such as logistic regression, decision trees, or random forests.

2. Model Evaluation:

- Evaluate the models using appropriate metrics (e.g., accuracy, precision, recall).
- Perform cross-validation to ensure model robustness.

Step 5: Dashboard Development

1. Dashboard Design:

- Create an interactive dashboard to present key insights and visualizations.
- Ensure the dashboard is user-friendly and provides actionable insights at a glance.

2. Tools:

- Use Python libraries like Dash, Plotly, or Streamlit to build the dashboard.
- Integrate visualizations created using matplotlib and seaborn into the dashboard.

Deliverables

1. Cleaned and Processed Dataset:

- Submit the final version of the cleaned dataset used for analysis and modeling.

2. Exploratory Data Analysis Report:

- Provide a report summarizing the findings from the EDA phase, including key statistics and visualizations.

3. Predictive Models:

- Submit the predictive models along with their evaluation metrics and a brief explanation of the modeling process.

4. Interactive Dashboard:

- Develop and share an interactive dashboard that presents the key insights and visualizations from the analysis.

5. Documentation and Presentation:

- Document the entire analysis process, including data cleaning, EDA, visualization, and modeling.
- Prepare a presentation summarizing key findings, insights, and recommendations.

Evaluation Criteria

- **Data Cleaning and Preparation:** Effectiveness of data preprocessing and handling of missing values.
- **Exploratory Data Analysis:** Depth and quality of initial data analysis and insights derived.
- **Visualization Skills:** Clarity, relevance, and effectiveness of visualizations created using matplotlib and seaborn.
- **Predictive Modeling:** Accuracy and interpretability of the predictive models.
- **Dashboard Design:** Usability, interactivity, and completeness of the developed dashboard.
- **Communication:** Quality of documentation and presentation.

Submission Form:

<https://forms.gle/JrzNJA9vUkWp7X6A8>

Deadline:

Sunday 23 June at midnight