EYouth X DEPI Tech Challenge

Submitted by:
EYouth

Brazilian E-Commerce Data Challenge

Challenge Description:

This competition focuses on analyzing the Brazilian E-Commerce Public Dataset by Olist, which includes detailed information about orders, products, customers, and payments. Participants will follow a structured approach using various tools such as Excel or SQL, Python, and Power BI to extract valuable insights into customer behavior and e-commerce performance.

Implementation in Steps:

A) Data Preparation:

> Load and merge the datasets on order_id.
> Create a unified dataset with relevant columns:

- Order Date (order_purchase_timestamp)
- Payment Details (payment_value, payment_type)
- Order Status (order_status)
- Product Prices (price, freight_value)

B) Reconciliation Criteria:

1. Sales Performance:

- Total Revenue = price + freight_value from successfully delivered orders.
- Expected Revenue = Total of all approved orders.
- Canceled Orders = Orders with order_status == "canceled".
- Late Deliveries = Orders where order_delivered_customer_date > order_estimated_delivery_date.

2. Payment Accuracy

C) Monthly Financial Dashboard Requirements:

1. Financial Overview:
- Total Revenue (Delivered Orders)
- Total Payments Received
- Expected Revenue (Approved Orders)
- % Revenue Reconciliation (Actual vs. Expected)

2. Order Breakdown:
- Total Orders by Month
- % of Canceled, Pending, and Delivered Orders
- Average Order Value

3. Delivery Insights:
- Count of Late Deliveries by Month
- Average Delay in Days

4. Interactive Features:
- Filters by Order Status, Payment Type, and Date Range.
- Drill-Down to Order-Level Details.

D) Bonus Analytical Tasks:

1. Profitability Analysis:
- Net Profit = price - freight_value (if available).
- Gross Margin % = (Net Profit / Total Revenue) * 100

2. Revenue Forecasting:
- Use previous monthly revenue trends to predict future sales and revenue growth

Supporting Material:

- Dataset:[https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/data?select=olist_orders_dataset.csv]

Time Frame:

- Duration: 1 Week
- Deadline: Friday, 28/2/2025

Student Deliverables:

- **Cleaned Dataset** – A well-processed dataset with missing values handled, errors corrected, and data properly structured for analysis.Source code and implementation files.
- **Python Analysis Code** – A Jupyter Notebook or Python script showcasing data exploration, statistical analysis, and key insights using libraries like Pandas, Matplotlib, and Seaborn.Any additional requirements specified in the challenge.
- **Power BI Dashboard** – An interactive visualization report presenting key findings with charts, graphs, and insights.

Concise Report – A structured document summarizing:

- Data cleaning and preprocessing steps.
- Key analytical findings and trends.
- Visualizations and interpretations.
- Business or strategic insights derived from the data.

Evaluation Criteria:

1. Data Cleaning and Preprocessing (20%)

- **Completeness**: All missing values are handled appropriately.
- **Accuracy**: Errors in the dataset are corrected.
- **Structure**: Data is properly structured and formatted for analysis.
- **Documentation**: Clear documentation of the cleaning process is provided.

2. Data Analysis (30%)

- **Depth of Analysis**: Use of advanced analytical techniques and libraries (e.g., Pandas, Matplotlib, Seaborn).
- **Insightfulness**: Identification of meaningful patterns, trends, and correlations.
- **Relevance**: Analysis is relevant to the e-commerce context and provides actionable insights.
- **Code Quality**: Clean, well-commented, and efficient code.

3. Visualization and Reporting (30%)

- **Clarity**: Visualizations are clear, well-labeled, and easy to understand.
- **Interactivity**: Power BI dashboard is interactive and user-friendly.
- **Insight Presentation**: Key findings are effectively communicated through visualizations.
- **Aesthetics**: Visual appeal and professional presentation of the dashboard and report.

4. Report Quality (20%)

- **Structure**: The report is well-organized and follows a logical flow.
- **Content**: Comprehensive coverage of data cleaning, analysis, and findings.
- **Interpretation**: Clear interpretation of results and their implications.
- **Business Insights**: Practical business or strategic insights derived from the data.