

Project Proposal

Adithya Murali, Aditya Sindhavad, Ammar Mustafa, John Hwang, Manasa Maganti,
Sameer Khan, Shashank Rao, Varsha Manju Jayakumar

Summary of the Dataset and Its Relevance

The dataset for this project is of Olist, a Brazilian e-commerce platform and is sourced from Kaggle. The dataset consists of multiple interconnected tables covering customer demographics, order details, product categories, seller information, payment methods, and customer reviews.

This dataset is particularly relevant because it provides comprehensive information on order trends, revenue generation, and customer behaviors. By analyzing this dataset, we can uncover patterns in order volumes, peak sales times, and revenue generation trends, which can help businesses optimize their sales strategies.

Problem Statement

The goal of this project is to analyze Olist's sales data to determine patterns in order volume and revenue generation. Specifically, we aim to answer the following questions:

- What is the distribution of orders across different days and times?
- When are the peak sales periods (days and hours)?
- What are the trends in revenue generation across different periods?
- How do different payment methods impact revenue and order volumes?
- Are there any seasonality patterns or specific time frames where sales perform better?

By addressing these questions, businesses using Olist can optimize their sales strategies, promotional campaigns, and inventory management.

Intended Audience and Story Framing

The primary audience for this analysis includes stakeholders and board members at Olist. The analysis will be conducted using Python notebooks for data exploration and insights generation, and the findings will be presented in a structured presentation format. This approach ensures that stakeholders can understand and act upon key insights effectively through a combination of interactive analysis and visual storytelling.

Initial Ideas on Visualization Types and Tools

For this project, we plan to use the following visualization techniques:

- **Line Charts:** To examine order volume trends over time and identify seasonal peaks.

- **Bar Charts:** To compare sales distribution across different time periods, such as days of the week and hours of the day.
- **Pie Charts:** To display the proportion of different payment methods used by customers.
- **Tree Plot:** To visualize the distribution of orders across different product categories, allowing for quick identification of the most and least popular segments.
- **Geo-Spatial Plot:** To analyze the relationship between state population, average income, and total purchase value, providing insights into regional purchasing behavior and market potential.

The tools considered for this project include:

- Python (Pandas, Matplotlib, Seaborn, Plotly): For data analysis and visualization.
- Excel/Google Sheets: For preliminary data exploration and reporting.

Next Steps

- Data Preprocessing: Cleaning and merging datasets for analysis.
- Exploratory Data Analysis (EDA): Identifying trends, anomalies, and key insights.
- Visualization Development: Creating charts and graphs for key findings.
- Interpretation & Recommendations: Providing actionable insights based on the analysis.

By completing this project, we aim to provide meaningful insights that can help Olist sellers maximize their revenue and optimize their business strategies effectively.