Milestone 1: Order Pattern Analysis & Data Integration

Dataset Overview

The Olist Brazilian E-commerce dataset (2016–2018) consisting of 9 relational CSV files was integrated into a master dataset. The integration started with the orders table as the base and merged order items, products, category translations, customers, and sellers. The final cleaned dataset includes delivered orders between January 2017 and August 2018.

Data Cleaning Steps

- Filtered only delivered orders - Removed missing delivery dates - Removed negative prices and freight values - Filled missing product categories as 'unknown' - Converted all date columns to datetime format - Removed duplicate records

Order Volume Trends

Monthly order volume showed consistent growth across the period, with noticeable peaks during holiday seasons (especially November and December). Month-over-month growth rates indicate increasing demand with seasonal fluctuations.

Product Category Insights

Top product categories by volume and revenue were identified. Some categories led in order volume but not necessarily in revenue, indicating price differences across categories.

Geographic & Seasonality Insights

Orders were concentrated in specific Brazilian states, with São Paulo leading in volume. Cross-state deliveries represented a significant portion of orders. Seasonality analysis confirmed increased activity during year-end periods.

Key Business Observations

- Demand is growing steadily with strong seasonal spikes - Certain states dominate both supply and demand - Cross-state logistics likely impact delivery times - Inventory planning must account for holiday season surges