

Executive Summary

E-Commerce Data Analysis using Python & SQL

This project focuses on a comprehensive E-commerce data analysis combining SQL-based data extraction with Python-driven analytics and visualizations to uncover meaningful insights related to customer behavior, geographic trends, order patterns, payment preferences, product performance, seller contribution, and revenue dynamics. The analysis transforms raw transactional data into decision-ready business intelligence, supported by well-structured charts and statistical measures.

1. Customer & Geographic Distribution Insights

The customer analysis shows that customer distribution is highly skewed geographically rather than evenly spread across all regions.

- A small subset of states contributes the majority of customers, accounting for well over 60–70% of the total customer base, while the remaining states collectively represent a much smaller share.
 - City-level analysis further reinforces this trend, where top cities contribute a disproportionately high percentage of total orders.
 - The average number of products per order varies by city, indicating that certain regions demonstrate higher basket sizes, suggesting stronger purchasing power or higher engagement.
-

2. Order Volume & Seasonal Trends

Time-based analysis of orders reveals clear seasonal and temporal patterns.

- Monthly order trends show that peak months contribute a significantly higher percentage of total annual orders, while off-season months contribute comparatively less.
 - Using moving average analysis, short-term fluctuations are smoothed to expose long-term demand trends, making it easier to distinguish between temporary spikes and sustainable growth.
 - Order distribution indicates that a majority of orders are concentrated within specific time windows, highlighting predictable demand cycles.
-

3. Payment Behavior Analysis

Payment analysis highlights strong customer preference patterns:

- Installment-based payments account for a dominant percentage of total transactions, indicating that flexible payment options significantly influence purchasing decisions.
 - Single-payment transactions form a smaller portion, suggesting that customers are more comfortable making higher-value purchases when installments are available.
-

4. Product Category & Revenue Contribution

Revenue analysis by product category reveals a Pareto-like distribution:

- Top product categories contribute a majority (over 70–80%) of total revenue, while the remaining categories generate comparatively smaller revenue shares.
 - Some categories show high order volume but lower revenue contribution, indicating lower average selling prices.
 - Other categories generate high revenue with fewer orders, reflecting premium pricing or higher-margin products.
-

5. Seller Performance & Revenue Concentration

Seller-level analysis shows strong revenue concentration:

- A small percentage of sellers generate a large share of total platform revenue, highlighting reliance on top-performing sellers.
 - Using window functions (DENSE_RANK), sellers are ranked based on revenue contribution, clearly separating top-tier sellers from long-tail contributors.
 - Lower-ranked sellers collectively contribute a smaller percentage but represent growth potential if performance is improved.
-

6. Statistical Relationship Analysis (Price vs Demand)

Correlation analysis between order count and average price provides quantitative insights into customer price sensitivity:

- The correlation coefficient indicates the direction and strength of the relationship between pricing and demand.

- Results suggest that price changes influence purchasing behavior, though the relationship is not perfectly linear, implying the presence of other influencing factors such as brand value, category type, and discounts.
-

Overall Business Value Delivered

- ✓ Converted raw transactional data into actionable insights
- ✓ Identified high-value customers, regions, products, and sellers
- ✓ Highlighted seasonal demand patterns and payment preferences
- ✓ Applied statistical analysis for evidence-based decision-making
- ✓ Demonstrated end-to-end analytical workflow using SQL + Python