📊 E-Commerce Sales & Customer Behavior Analysis

Executive Summary

This project presents an end-to-end analysis of an e-commerce dataset with the objective of extracting actionable business insights related to sales performance, customer behavior, payment patterns, seller contribution, and retention trends. The analysis was conducted using **MySQL** for data extraction and transformation, followed by **Python (Pandas, NumPy, Matplotlib, Seaborn)** for exploratory data analysis and visualization.

Key findings indicate that revenue is highly concentrated within a small number of product categories and sellers (Pareto effect), installment-based payments dominate transaction behavior (~99.99%), and customer demand is weakly correlated with price. Seasonal patterns in order volume and geographically concentrated customer bases further highlight opportunities for targeted marketing, logistics optimization, and retention strategies.

---

1. Business Problem Statement

E-commerce platforms operate in highly competitive environments where data-driven decision-making is critical. The core business challenges addressed in this project include: - Identifying high-revenue product categories and sellers - Understanding customer purchasing and payment behavior - Detecting seasonal and regional sales trends - Evaluating customer retention and loyalty patterns

Objectives: 1. Analyze overall sales and order trends 2. Identify top-performing product categories and sellers 3. Examine customer geographic distribution 4. Study payment methods and purchasing behavior 5. Derive insights to support business strategy

---

2. Dataset Overview

Data Source

Relational e-commerce database consisting of multiple normalized tables.

Tables Used

- `customers`
- `orders`
- `order_items`
- `products`
- `payments`
- `sellers`

Scope

- Customers distributed across thousands of cities and multiple states
- Orders spanning multiple years
- Wide range of product categories and sellers

3. Tools & Technologies

Category	Tools	MySQL
Database	Python	Pandas,
Programming Language	NumPy	Matplotlib,
Libraries	Seaborn	Jupyter
Visualization	Notebook	
Environment		

4. Data Extraction & Cleaning

- Complex SQL joins across multiple tables
- Aggregations using SUM, COUNT, AVG
- Use of window functions such as RANK, DENSE_RANK,LAG
- Handling duplicate records and missing values
- Conversion of SQL query outputs into Pandas DataFrames

---

5. Exploratory Data Analysis & Insights

5.1 Customer Geographic Distribution

Customer distribution analysis revealed a strong concentration in a few states, particularly São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG).

Insight: Logistics and marketing operations should prioritize these regions to maximize efficiency and return on investment.

5.2 Order Volume Trends

Order volume exhibits noticeable seasonality, with peaks during specific months and a decline toward the later part of the year.

Insight: Promotional campaigns should be strategically aligned with high-demand periods.

5.3 Sales by Product Category (Pareto Analysis)

A small subset of product categories contributes disproportionately to total revenue, demonstrating a classic Pareto (80/20) effect.

Insight: Inventory planning and supplier negotiations should focus on high-performing categories.

5.4 Payment Behavior Analysis

Analysis shows that approximately **99.99% of transactions are paid via installments**.

Business Implication: Installment payment options are a critical conversion factor and should be maintained or enhanced.

5.5 Price vs Purchase Frequency

Correlation analysis between product price and purchase frequency yielded a value of approximately **–0.10**, indicating a very weak negative correlation.

Insight: Customer demand is driven more by product utility and category relevance than by price alone.

5.6 Seller Performance Analysis

Revenue contribution among sellers is highly skewed, with top sellers accounting for a significant portion of overall sales.

Actionable Insight: Seller incentive and retention programs should focus on top-performing sellers to stabilize revenue streams.

5.7 Customer Purchase Behavior

Average number of products per order varies significantly across cities, purchasing power and consumer behavior.

Use Case: City-level bundling and upselling strategies can be developed.

5.8 Customer Retention Analysis

Retention was defined as customers making repeat purchases within a six-month period. A substantial fraction of customers were identified as one-time buyers.

Insight: Early engagement strategies are essential to improve customer lifetime value.

5.9 High-Value Customer Identification

Using yearly aggregation and ranking functions, the top three customers per year were identified based on total spending.

Business Value: Enables targeted VIP programs and personalized offers.

---

6. Advanced Analytics Techniques Used

- SQL Window Functions
- Moving Averages for trend smoothing
- Year-over-Year (YoY) analysis
- Ranking and cumulative metrics

These techniques enhance analytical depth and demonstrate industry-level data analysis proficiency.

---

7. Key Business Recommendations

1. Allocate marketing budgets toward top-performing product categories
2. Preserve and optimize installment-based payment systems
3. Implement seller reward programs for high contributors
4. Introduce regional marketing strategies for high-basket cities
5. Launch early-stage customer retention campaigns
6. Use seasonality insights for demand forecasting

---

8. Limitations & Future Scope

Limitations

- Lack of customer demographic and review data
- Nodirectinformationon marketing campaigns

Future Enhancements

- RFM-based customer segmentation
- Predictive sales forecasting models
- Churn prediction using machine learning
- Interactive dashboards using Tableau or Power BI

---

9. Conclusion

This project demonstrates a comprehensive analytical workflow from raw data extraction to business insight generation. The methodologies and insights presented can directly support strategic decision- making in an e-commerce environment and reflect industry-ready data analytics skills.

4