

E-Commerce Marketplace Analytics Using SQL

Final Project Summary

Project Overview

This project analyzes the Brazilian Olist e-commerce marketplace dataset using SQL to uncover insights about customer behavior, revenue distribution, logistics performance, and seller activity.

The goal is to simulate real-world marketplace analytics using a structured relational database and business-driven SQL queries.

The project includes data cleaning, relational modeling, and 50 analytical SQL queries designed to answer practical business questions.

Dataset Overview

Dataset source: Brazilian Olist E-Commerce Dataset (Kaggle)

The dataset represents a multi-vendor online marketplace and simulates real transactional behavior.

Tables included in the project:

- customers
- orders
- order_items
- products
- sellers
- payments
- reviews
- geolocation
- category_translation

The dataset was cleaned in Python and imported into MySQL for structured SQL analysis. This environment replicates a realistic analytics warehouse setup.

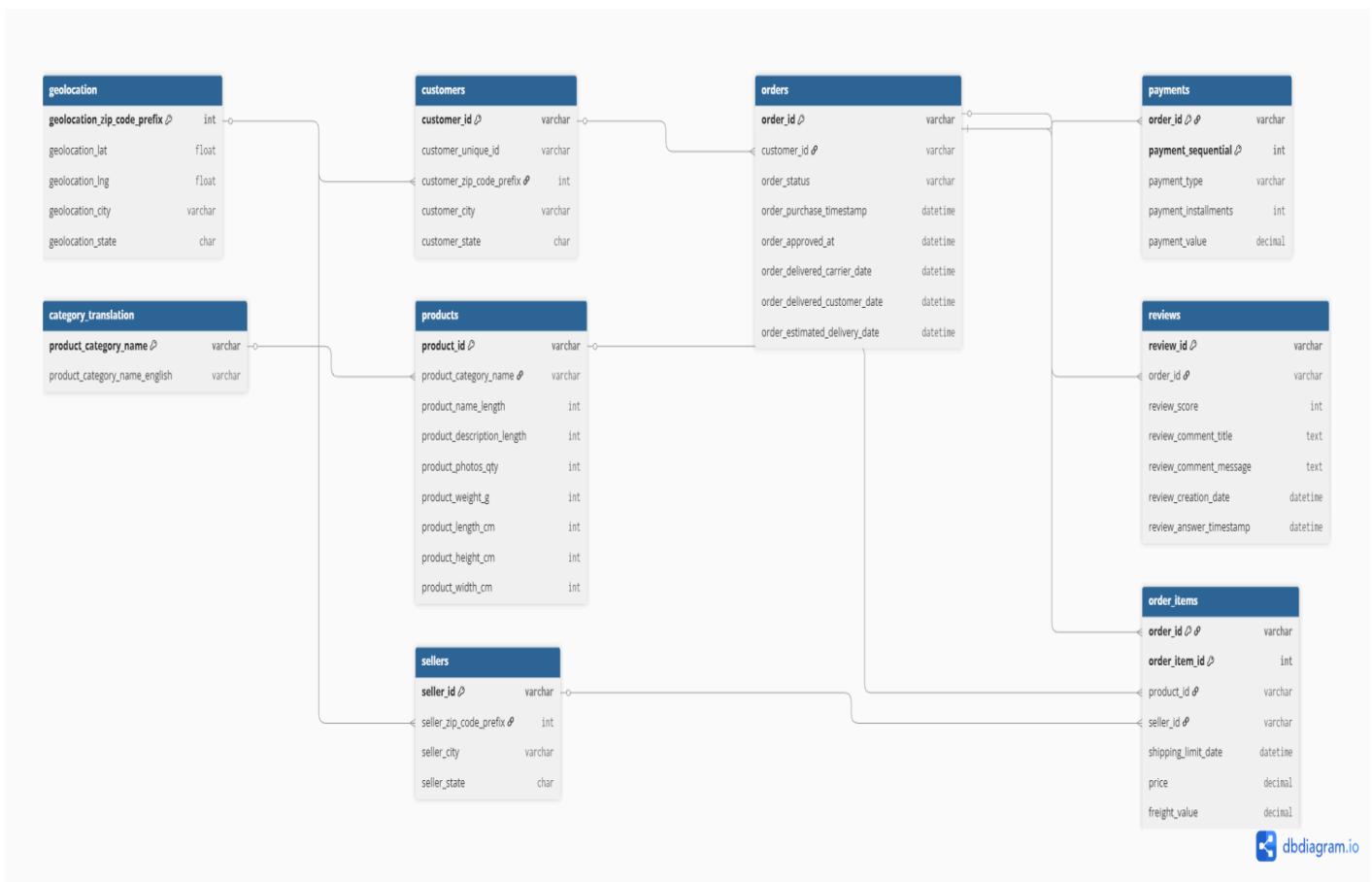


Figure 1 - Database ER Diagram

Database Structure

The relational schema follows a normalized marketplace design.

Core transaction flow:

customers → orders → order_items

Supporting tables extend analytics capability:

payments — transaction behavior
 reviews — customer satisfaction
 sellers — marketplace supply
 products — catalog metadata
 geolocation — regional analysis

Foreign keys maintain integrity and enable multi-table joins required for business analytics.

Key Business Questions

The SQL analysis focuses on answering practical marketplace questions:

- Who are the highest spending customers?
- Which cities generate the most revenue? • How efficient is delivery performance?
- Which sellers drive marketplace growth?
- How does delivery speed impact customer ratings?
- Is revenue concentrated among a small group of buyers?
- Are customers repeating purchases?

These questions simulate real analyst decision support.

Major Insights

- Insight 1 — Revenue concentration
A small group of customers generates a large portion of total revenue.
- Insight 2 — Geographic dominance
Major cities contribute disproportionately to platform revenue.
- Insight 3 — Seller concentration
Top sellers handle a significant share of marketplace activity.
- Insight 4 — Delivery impact
Slower deliveries correlate with lower review scores.
- Insight 5 — Retention gap
Most customers purchase only once, indicating low retention.
- Insight 6 — Logistics variation
Delivery performance varies widely across regions.

Skills Demonstrated

- Complex SQL joins
- Aggregations and GROUP BY
- Subqueries
- Window functions
- Business KPI analysis
- Relational database modeling
- Data cleaning and preparation

- Structured analytical reasoning

Tools Used

- MySQL
- SQL
- Python (Jupyter Notebook)

Conclusion

This project demonstrates practical SQL analytics applied to a real marketplace dataset.

It showcases the ability to model relational data, answer business questions, and translate SQL results into actionable insights.

The work reflects readiness for entry-level data analyst and SQL analytics roles.

