

Brazilian Olist Store Data Analysis – Project Documentation

1. Overview

This project analyzes the Brazilian Olist Store Dataset to understand e-commerce performance across customers, sellers, orders, payments, and products using Python, SQL, and Power BI. The final result is an interactive dashboard showing insights such as customer count, seller count, revenue, orders, and delivery performance.

2. Objectives

- Clean and preprocess the raw dataset using Python.
- Build relational data models using SQL.
- Perform exploratory data analysis (EDA).
- Create KPI dashboards and visual insights in Power BI.
- Identify trends that affect customer satisfaction and store performance.

3. Tools & Technologies

- **Python** (Pandas, NumPy, Matplotlib)
- **SQL** (Data modeling, joins, relationships)
- **Power BI** (Dashboard creation)
- **GitHub** (Version control & documentation)

4. Dataset Description

The Olist dataset contains:

- Customers data
- Orders and order items
- Sellers and products
- Payments & delivery status
- Reviews and satisfaction scores

5. Data Cleaning Using Python

Performed using Pandas:

- Handling missing values
- Converting date formats
- Removing duplicates
- Normalizing categorical fields
- Outlier detection for delivery time & payments

6. Data Modeling in SQL

- Created relational schemas
- Built relationships using primary and foreign keys
- Performed joins to prepare fact and dimension tables
- Generated analytical views for Power BI loading

7. Power BI Dashboard Highlights

The dashboard includes:

- Total customers & sellers

- Total revenue
- Order status distribution
- Delivery performance
- Payment method insights
- Customer review trends

8. Key Insights

- Delivery delays affect customer review scores.
- Most revenue is generated from specific high-volume categories.
- Credit card is the dominant payment method.
- Sellers with more product variety show higher order frequency.

9. Repository Structure

- `/python/` – Data cleaning scripts
- `/sql/` – SQL data modeling files
- `/powerbi/` – PBIX dashboard
- `/docs/` – Documentation

10. Conclusion

This project demonstrates a complete data workflow: cleaning, modeling, analyzing, and visualizing. The resulting dashboard helps businesses understand customer behavior, optimize operations, and improve decision-making.