Case Study: Olist E-Commerce Analysis (Reviews, Revenue & Order Performance)

Introduction

Olist is a Brazilian e-commerce platform that connects small businesses to major marketplaces. The dataset contains customer, order, product, and review information, making it ideal for analyzing business performance, customer experience, and operational efficiency.

Business Question:
👉 *"How do delivery delays and product mix affect review scores, repeat purchases, and revenue?"*

Key KPIs Defined:

- **Revenue** – Total sales (sum of item prices).
- **Orders** – Total and delivered orders.
- **Average Review Score** – Customer satisfaction metric.
- **90-Day Repeat Rate** – Customers who purchased again within 90 days.
- **SLA Breach Rate** – % of orders delivered after the shipping limit date.
- **Returns/Cancellations** – Orders not fulfilled.

Data Preparation

- **Source:** Olist e-commerce public dataset.
- **Database:** MySQL (tables: orders, order_items, order_reviews, order_payments, products).
- **Transformations:**
  - Removed incomplete/missing geo data.
  - Standardized date formats (order_purchase_timestamp, review_creation_date).
  - Handled NULL values (e.g., missing review scores).
  - Exported KPI queries into **CSV files** for Power BI.

Analysis & Queries

- **Revenue by Month** → Growth trends and seasonality.
- **Revenue by Product Category** → Identified top-selling categories.
- **Average Review Score** → Overall customer satisfaction.
- **Order Status Distribution** → Delivered vs canceled/unavailable orders.
- **SLA Breach Rate** → % of late deliveries impacting reviews.

- **Returns/Cancellations** → Proportion of lost orders.

(SQL scripts for each KPI are included in the project folder under /queries)

Power BI Dashboard

The dashboard was built using exported CSVs from SQL queries.

Visuals Included:

- **Revenue by Month (Line Chart)** – Trend of sales growth.
- **Revenue by Product Category (Bar Chart)** – Category contribution.
- **Average Review Score (Card)** – Highlight of customer satisfaction.
- **Total Orders (Card)** – Count of orders placed.
- **Delivered Orders (Card)** – Successful deliveries.
- **Cancellations/Unavailable (Card)** – Orders not fulfilled.

Insights

Revenue Growth:

i. Consistent increase in revenue month-over-month.
ii. Certain categories (e.g., electronics, furniture) dominate sales.

Customer Reviews:

i. Average review score ~4.0 shows good customer experience.
ii. Late deliveries correlate with lower review scores.

Operational Efficiency:

i. SLA breach rate around ~8% indicates logistics inefficiency.
ii. 1.24% of total orders were canceled or unavailable.

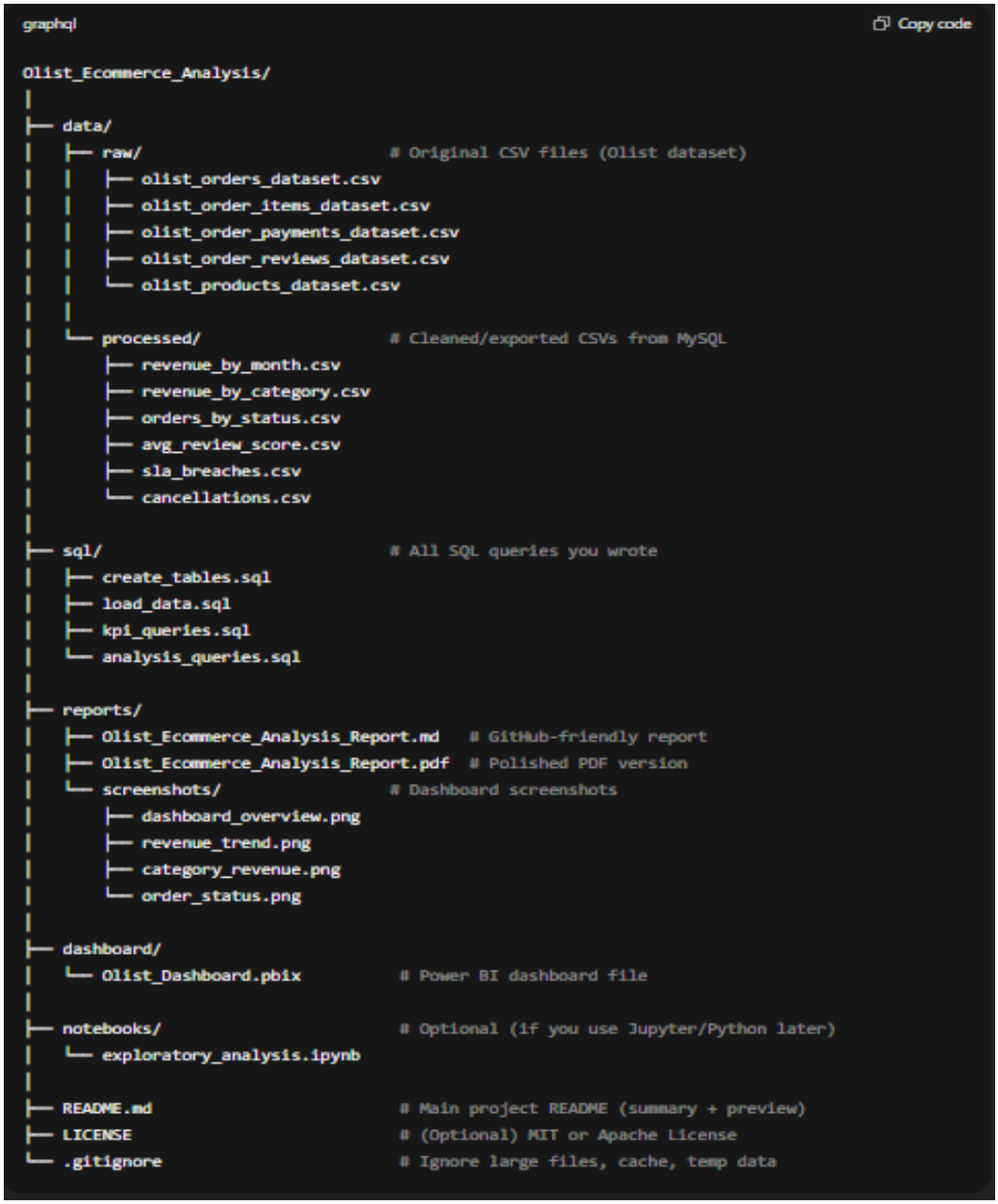
Customer Retention:

i. Repeat purchase rate was **very low** in dataset (0%).
ii. Indicates customer loyalty challenges, possibly due to delivery issues.

Conclusion

- **Delivery performance** (on-time shipping) is strongly tied to **review scores**.
- **Product mix** affects revenue concentration — dependency on a few categories poses risk.
- **Customer loyalty is weak** → Olist must invest in faster delivery, post-purchase engagement, and targeted promotions.

Project Structure



```graphql
Olist_Ecommerce_Analysis/
|
├── data/
|   ├── raw/                        # Original CSV files (Olist dataset)
|   |   ├── olist_orders_dataset.csv
|   |   ├── olist_order_items_dataset.csv
|   |   ├── olist_order_payments_dataset.csv
|   |   ├── olist_order_reviews_dataset.csv
|   |   └── olist_products_dataset.csv
|   |
|   └── processed/                  # Cleaned/exported CSVs from MySQL
|       ├── revenue_by_month.csv
|       ├── revenue_by_category.csv
|       ├── orders_by_status.csv
|       ├── avg_review_score.csv
|       ├── sla_breaches.csv
|       └── cancellations.csv
|
├── sql/                            # All SQL queries you wrote
|   ├── create_tables.sql
|   ├── load_data.sql
|   ├── kpi_queries.sql
|   └── analysis_queries.sql
|
├── reports/
|   ├── Olist_Ecommerce_Analysis_Report.md   # GitHub-friendly report
|   ├── Olist_Ecommerce_Analysis_Report.pdf  # Polished PDF version
|   └── screenshots/                # Dashboard screenshots
|       ├── dashboard_overview.png
|       ├── revenue_trend.png
|       ├── category_revenue.png
|       └── order_status.png
|
├── dashboard/
|   └── Olist_Dashboard.pbix         # Power BI dashboard file
|
├── notebooks/                      # Optional (if you use Jupyter/Python later)
|   └── exploratory_analysis.ipynb
|
├── README.md                       # Main project README (summary + preview)
├── LICENSE                         # (Optional) MIT or Apache License
└── .gitignore                      # Ignore large files, cache, temp data
```

Future Work

- Incorporate customer demographic segmentation.
- Predict **churn probability** using ML models.
- Optimize **delivery routes & SLA management**.