Olist E-Commerce Analysis Report

1. Business Problem

The goal of this analysis was to explore and understand sales, customer behavior, and product performance in the Brazilian e-commerce market (Olist dataset). The insights help stakeholders answer:

- Which products and categories drive the most revenue?
- Who are the most valuable customers and how can we retain them?
- What delivery issues impact satisfaction?
- How can marketing and operations be optimized?

2. Dataset Description

- **Dataset:** Brazilian E-Commerce Public Dataset by Olist
- **Source:** Kaggle
- **Size:** ~100k orders across multiple CSV files
- **Time Period:** 2016 – 2018
- **Main Tables:** Orders, Order Items, Customers, Products, Payments, Reviews, Geolocation

3. Methodology

3.1 Data Cleaning (Python)

- Removed duplicates, handled missing values, corrected date formats, merged datasets.

3.2 Exploratory Data Analysis (EDA)

- Analyzed sales trends, top categories, delivery times, customer distribution.

3.3 Customer Segmentation

- Applied **RFM (Recency, Frequency, Monetary)** analysis to classify customers into VIP, Loyal, At-Risk, and New.

3.4 Visualization

- Built dashboards in **Power BI** and Python (Matplotlib, Seaborn).

4. Key Findings

- Sales peak in **November** (Black Friday surge).
- **Health & Beauty**, **Bed/Bath/Table**, and **Sports & Leisure** top categories.
- **8% late deliveries** often linked to low reviews.
- **15% of customers** generate **50% of revenue**.
- Installment payments dominate large purchases.

5. Business Insights

- Focus on **VIP customers** via loyalty programs.

- Improve **delivery logistics** to reduce late deliveries.
- Increase **marketing spend** for Black Friday.
- Promote **high-margin categories** more aggressively.
- Encourage **upfront payments** to improve liquidity.

6. Limitations

- Dataset limited to Olist marketplace (Brazil) only.
- Minimal customer demographic data.
- Missing data for some deliveries and reviews.

7. Future Work

- Add demographic & marketing campaign data.
- Build predictive sales/churn models.
- Compare performance with industry benchmarks.