Customer Segmentation & Churn Risk Dashboard

Tech Stack: Python, MySQL, Power BI, DAX, RFM Segmentation

Dataset Used: Brazilian E-Commerce Public Dataset – Olist (Kaggle)

# Business Objective

- Identify high-value, loyal, and at-risk customers using RFM segmentation.  
- Analyze customer behavior by region, payment methods, and product categories.  
- Help marketing teams design targeted campaigns to retain valuable customers and re-engage churned ones.

# Business Questions Answered

1. Who are the most valuable customers?  
2. Which customers are at risk of churning?  
3. What is the distribution of customers by recency, frequency, and monetary value?  
4. Which states or cities generate the most revenue?  
5. What is the average customer lifecycle and purchasing behavior?

# Project Workflow

## 1. Data Preparation (Python + Excel)

- Loaded 5 CSV files: orders.csv, customers.csv, order\_items.csv, payments.csv, products.csv.  
- Removed duplicates, fixed column formats, and ensured proper datetime parsing.  
- Final cleaned columns: customer\_unique\_id, order\_id, order\_purchase\_timestamp, order\_delivered\_customer\_date, price, payment\_value, customer\_state, customer\_city.

## 2. Data Import to MySQL

- Imported cleaned CSVs into MySQL.  
- Created indexes and foreign key relationships for joins.

## 3. RFM Table Creation (SQL)

- Calculated Recency, Frequency, and Monetary values per customer.  
- Created SQL View rfm\_segmented with segment logic for recency, frequency, monetary.

## 4. Churn Risk Tagging

- Applied logic: Low Risk = Recent+Frequent+High Spender.

- Medium Risk = Warm+Occasional+Mid Spender.

- High Risk = Inactive+Rare+Low Spender.

## 5. Export Final Data to CSV

- Exported rfm\_segmented to CSV for use in Power BI.

# Dashboard Design (Power BI)

## Page 1: Executive Overview

- Filters: Date, State, City, RFM Segment  
- KPIs: Total Customers, Total Revenue, Avg Recency, Frequency, Monetary, Churn %, Active %, CLV.  
- Visuals: RFM Segment Pie, Customer Value Bar, Churn Risk Bar, Top Revenue by State.

## Page 2: Customer & Regional Analysis

- KPIs: Top Cities by Frequency, Top States by Revenue, Avg Order Value  
- Graphs: Churn Heatmap by State, Frequency vs Monetary, Monthly Revenue Trend, Payment Methods

# DAX Measures Used

Total Revenue = SUM(final\_data[monetary])  
Total Customers = DISTINCTCOUNT(final\_data[customer\_unique\_id])  
Avg Recency = AVERAGE(final\_data[recency])  
Avg Frequency = AVERAGE(final\_data[frequency])  
Avg Monetary = AVERAGE(final\_data[monetary])  
Active Customers = CALCULATE([Total Customers], final\_data[churn\_risk] = "Low Risk")  
Churned Customers = CALCULATE([Total Customers], final\_data[churn\_risk] = "High Risk")  
Churned % = DIVIDE([Churned Customers], [Total Customers], 0)  
CLV = [Avg Frequency] \* [Avg Monetary]

# Key Insights

- 1,100+ high-value customers contributed most of the revenue.  
- 89K+ customers showed low spending or infrequent activity.  
- SP, RJ, MG were top-performing states.  
- Churn rate around 30%, suggesting need for re-engagement campaigns.

# Conclusion

This end-to-end project demonstrates customer segmentation and retention analysis using SQL and Power BI. It simulates a real-world CRM problem—helping business stakeholders identify high-value customers, reduce churn, and optimize regional marketing strategies.