Project Brief: E-Commerce Customer Retention Analysis

Overview

This independent project explores customer retention within Brazil's leading e-commerce marketplace, using real-world data from Olist. With increasing competition and the growing importance of personalization and customer lifetime value (CLV), businesses must better understand customer behaviour to drive retention.

Context

As a data analyst, the objective was to use SQL (PostgreSQL) to model, clean, and transform eight core datasets into a series of analytical views. These views were designed to answer key business questions related to purchasing patterns, delivery performance, customer churn, and satisfaction. Power BI was used to visualize these insights in a clear and strategic format.

Objectives

- Identify factors influencing customer retention and churn
- Analyse delivery trends, payment behaviour, and product category performance
- Segment high-value customers using cohort and RFM analysis
- Build visual dashboards that support data-driven decision-making

Tools Used

- PostgreSQL (pgAdmin 4) – Data prep, transformation, CTEs, window functions
- Power BI – Interactive dashboards for business insights
- Python – Pre-SQL cleaning: null handling, formatting, timestamp parsing
- ERD Tools – Data structure mapping for relational clarity

Business Questions Answered

- Which regions have the highest-value and most loyal customers?
- How does delivery performance affect customer satisfaction and retention?
- How do payment methods and average basket size vary across customer segments?
- Which product categories perform best by revenue and satisfaction?
- Where is customer churn highest — and what factors contribute to it?

Approach

- Cleaned and structured raw CSV data using Python
- Loaded datasets into PostgreSQL and created 19 tailored views
- Imported SQL views into Power BI using direct connectors
- Built a data model and created visualizations including heatmaps, order funnels, retention curves, and category dashboards

Deliverables

- Reusable SQL scripts with documented logic
- Power BI file with interactive dashboards
- ERD and schema documentation
- Business summary with insights and recommendations

Outcome- The project revealed clear retention trends, regional insights, and the impact of delivery delays on satisfaction. By combining SQL modelling and data storytelling through Power BI, the project delivers a strategic view of customer behaviour — helping businesses improve decision-making across marketing, logistics, and operations.