Problem Statement

A retail company aims to understand customer purchasing behavior and proactively identify churn-risk customers to improve retention, engagement, and revenue optimization. The objective is to analyze demographic and behavioral purchase patterns, uncover key drivers of customer loyalty, and develop predictive signals that enable targeted marketing and customer retention strategies.

Deliverables

The project delivers an end-to-end analytics and machine learning solution spanning descriptive analysis, predictive modeling, and business insight delivery.

Data Preparation & Feature Engineering (Python)
Cleaned and transformed customer-level retail data, engineered behavioral engagement features, and constructed a multi-factor churn definition based on purchase frequency, purchase history, subscription status, and spending patterns.

Exploratory & SQL Analytics (PostgreSQL)
Performed structured analysis of customer segments, purchasing trends, subscription behavior, and revenue drivers to uncover patterns in engagement and retention risk.

Predictive Modeling (scikit-learn)
Developed and evaluated Logistic Regression and Random Forest churn prediction models using pipeline-based preprocessing (encoding, imputation, leakage control), achieving ROC-AUC 0.91.

Visualization & Business Insights (Power BI)
Designed interactive dashboards highlighting customer segments, spending behavior, and churn-risk indicators to support data-driven marketing and retention strategies.

End-to-End Data Workflow (GitHub)
Integrated Python, SQL, and visualization artifacts into a structured repository demonstrating the complete lifecycle from raw data analysis to predictive insight delivery.