CartIQ - A Customer Segment Analysis

An Internship Project Report Submitted
In partial fulfillment of the requirements for the successful completion of the Internship at Tavant Technologies

Submitted by
SATTI HARISH REDDY
Employee ID: INT0223

Introduction

CartIQ is a data-driven customer segmentation system developed to analyze purchasing behavior using RFM (Recency, Frequency, Monetary) modeling. In today's competitive market, understanding customer value is critical for targeted marketing, retention, and revenue growth. This project leverages historical transaction data to segment customers into meaningful groups—such as VIPs, Loyal, At-Risk, and Lost—using KMeans clustering after normalizing RFM scores. Built during an internship at Tavant Technologies, CartIQ demonstrates practical applications of data preprocessing, behavioral analytics, and machine learning to drive customer-centric business insights.

ABSTRACT

The increasing need for personalized marketing strategies in modern business environments has made customer segmentation a critical component of data analytics. CartIQ is a customer segmentation system developed during an internship at Tavant Technologies that leverages RFM (Recency, Frequency, Monetary) analysis and KMeans clustering to classify customers based on their purchasing behavior. The project involves end-to-end data preprocessing, normalization, and application of machine learning techniques to identify and segment customers into value-based groups such as VIPs, Loyal, At-Risk, and Lost. This segmentation allows businesses to understand customer lifetime value and take data-driven actions to retain high-value customers and re-engage those at risk. The solution is built using Python and scalable data processing tools, demonstrating practical applications of behavioral analytics in real-world business scenarios.

Existing System

- Relies on manual or rule-based segmentation with limited criteria like last purchase.
- Lacks automation and behavioral analysis (e.g., no RFM or clustering).
- Results in ineffective targeting and poor customer retention strategies.

Proposed System

- Uses **RFM modeling** to quantify customer behavior.
- Applies **KMeans clustering** to segment customers into VIP, Loyal, At-Risk, and Lost.
- Enables **automated, data-driven segmentation** with high accuracy.
- Supports **personalized marketing** and retention strategies through actionable insights.

Hardware Requirements

- Processor: Intel i5 or above (or equivalent AMD)
- RAM: Minimum 8 GB (16 GB recommended)
- Storage: At least 512 MB free disk space
- System Type: 64-bit Operating System
- Display: HD display (for data visualization)

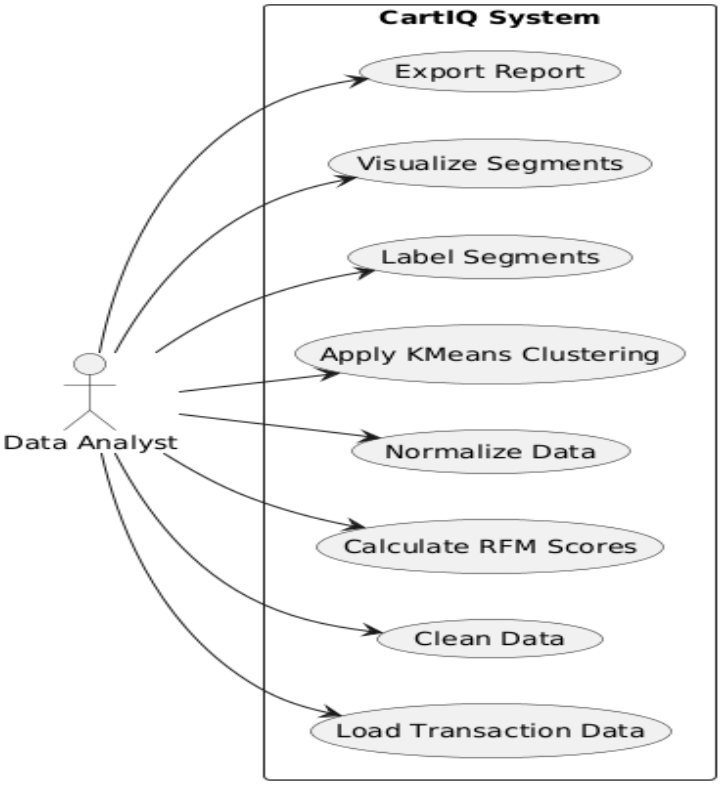
Software Requirements

    - Programing Language: python
    - Development environment: Jupyter Notebook / Google Colab
- Libraries:
    - pandas, NumPy (data manipulation)
    - matplotlib, seaborn (visualization)
    - scikit-learn (clustering & normalization)
- **Excel / Google Sheets**: For data viewing and exporting

Dataset :

Online Retail II - UCI Machine Learning Repository

UML Diagrams



The diagram illustrates the interaction between the **Data Analyst** and the core functionalities of the **CartIQ system**. It highlights major tasks such as loading and cleaning data, calculating RFM scores, applying clustering, labeling segments, visualizing results, and exporting the final report.