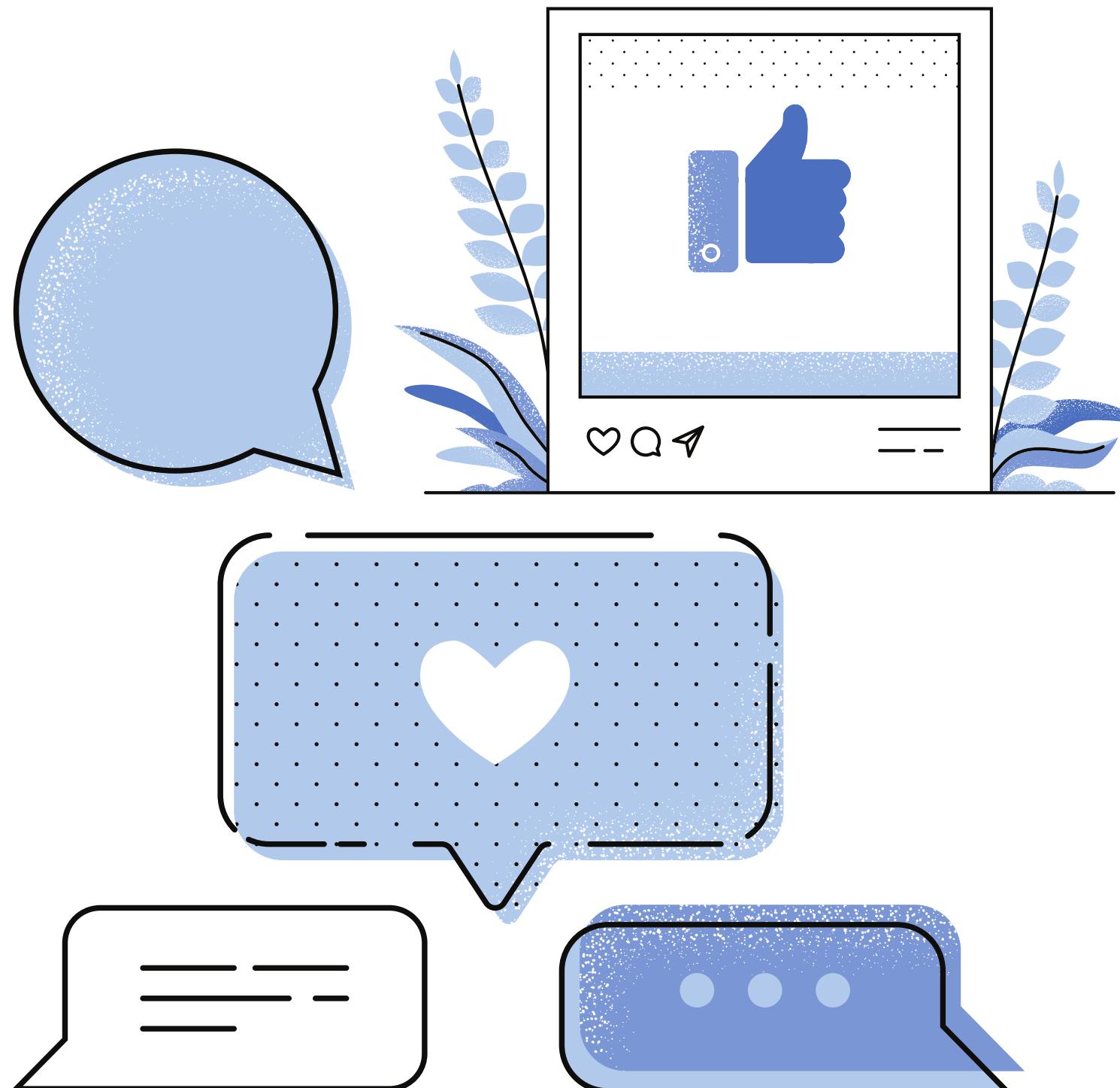
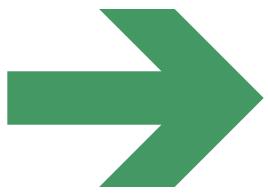


Customer Segmentation & Mini ML Model

Unlocking Customer Insights with RFM Analysis and Decision Tree





Project Overview

In this project, we aim to segment customers based on RFM (Recency, Frequency, Monetary) analysis and build a simple machine learning model to predict customer segments. This allows businesses to better understand customer behavior and develop targeted marketing strategies.

Organizations often collect a vast amount of customer data, but without structured analysis, the full potential of these data remains untapped. RFM segmentation offers a proven methodology to identify key customer groups based on purchasing behaviors.

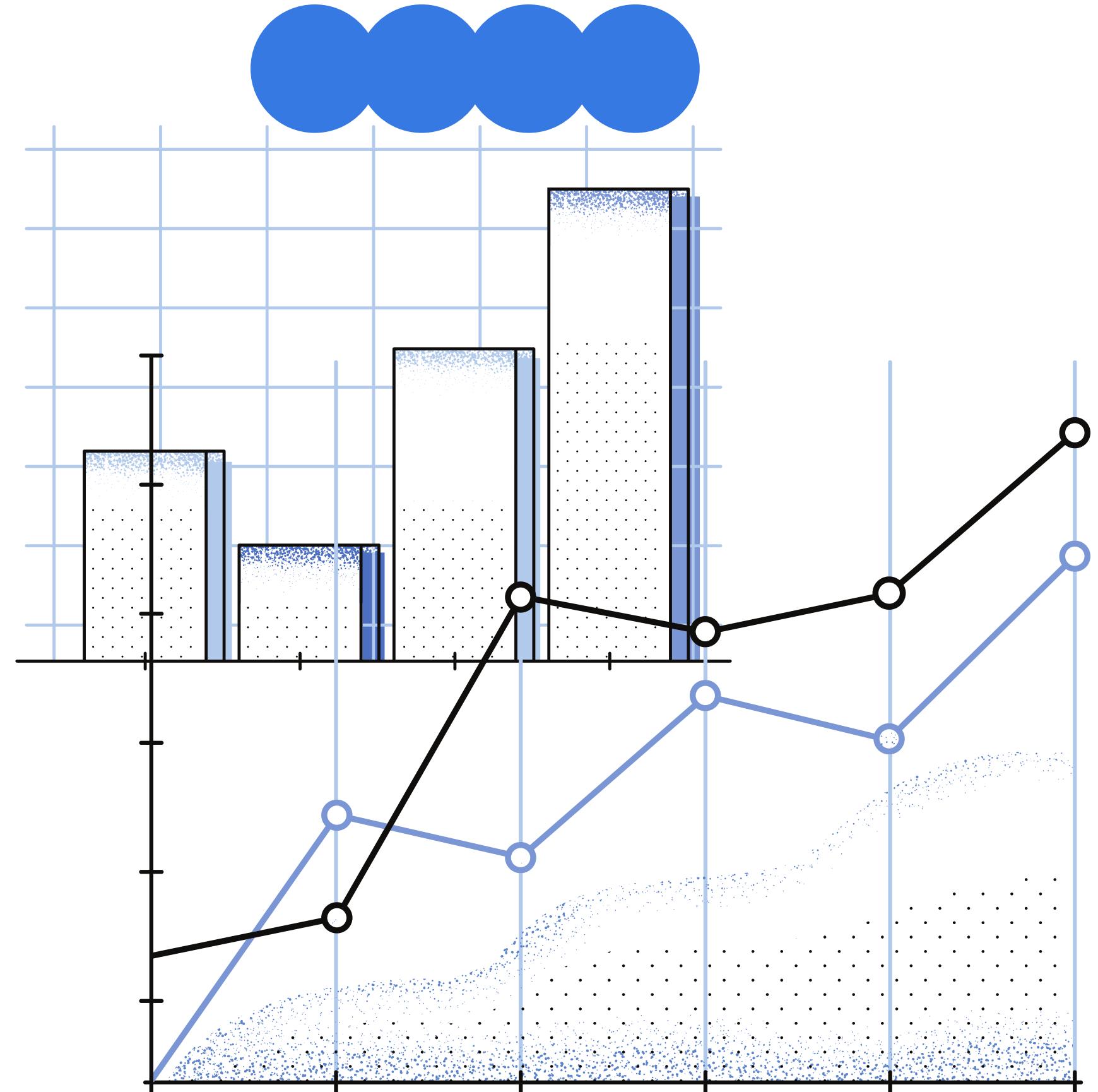
By combining RFM with a Decision Tree model, we can not only classify existing customers but also predict the segment of new or future customers, enhancing marketing effectiveness and customer retention.

Objectives

- Perform RFM Analysis to cluster customers into meaningful groups.
- Develop a Mini Machine Learning Model (Decision Tree) to classify customer segments.
- Build a dashboard to visualize customer segments and key behaviors.
- Provide actionable insights to support data-driven marketing strategies.

Expected Outcomes

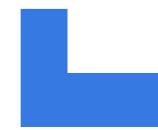
- Clear segmentation of customers into 4 groups based on RFM scores.
- A mini Decision Tree model that predicts customer segments with high accuracy.
- An interactive dashboard to explore customer behaviors by segment.
- Practical recommendations for marketing and customer relationship strategies.



DATA SET

SOURCE: Online Retail - UCI Machine Learning Repository

Link: <https://archive.ics.uci.edu/dataset/352/online+retail>



Dataset Information

This is a transactional data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

Data Preparation

- Filtered only completed transactions (excluded canceled orders)
- Removed missing CustomerID entries
- Selected only records with positive Quantity and UnitPrice

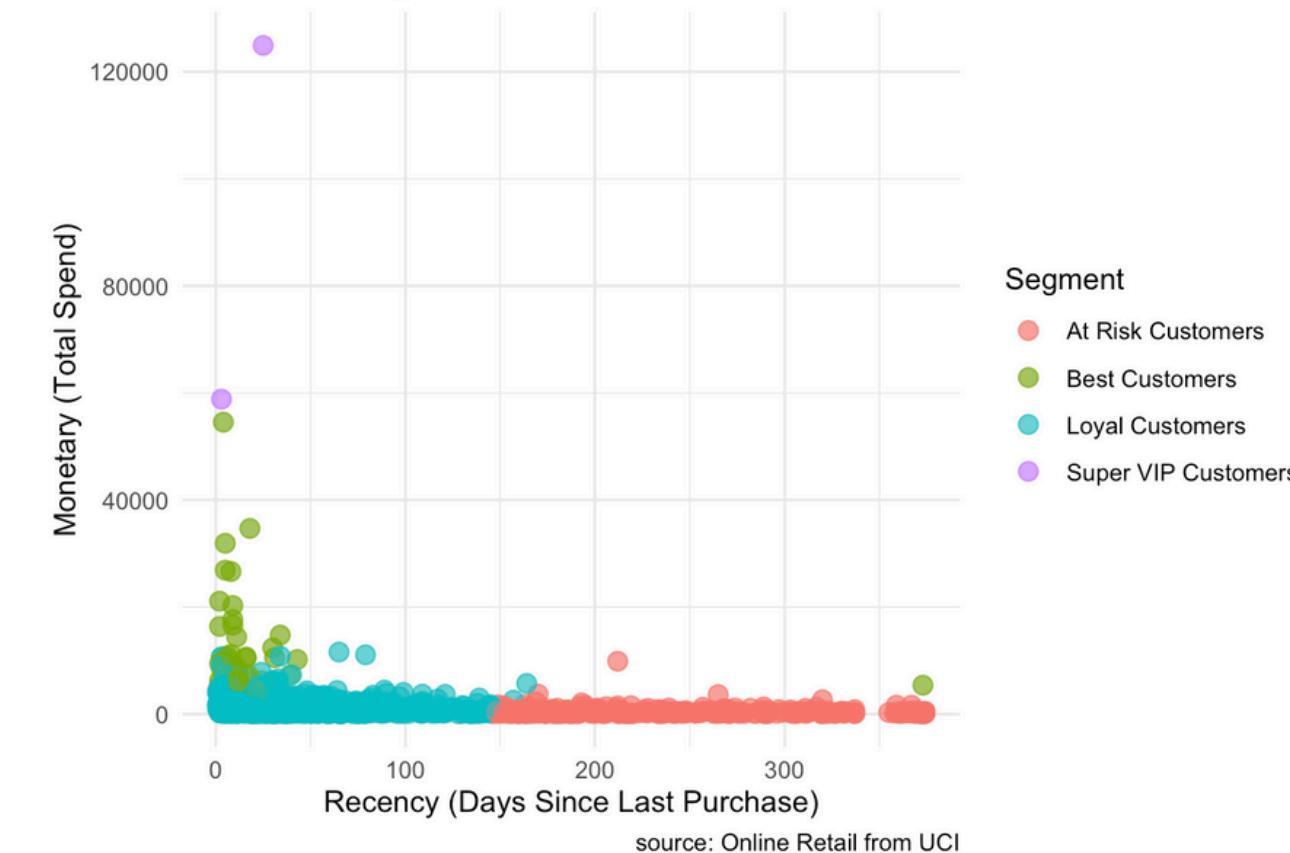
Fields Used for Analysis

- CustomerID : Customer identifier
- InvoiceDate : Transaction date
- Quantity : Number of items purchased
- UnitPrice : Price per item
- InvoiceNo : Invoice number
- Country : Country of transaction

Feature Engineering

- Created an RFM table (Recency, Frequency, Monetary) for customer segmentation
- Developed the "PredictedCluster" feature based on Machine Learning models

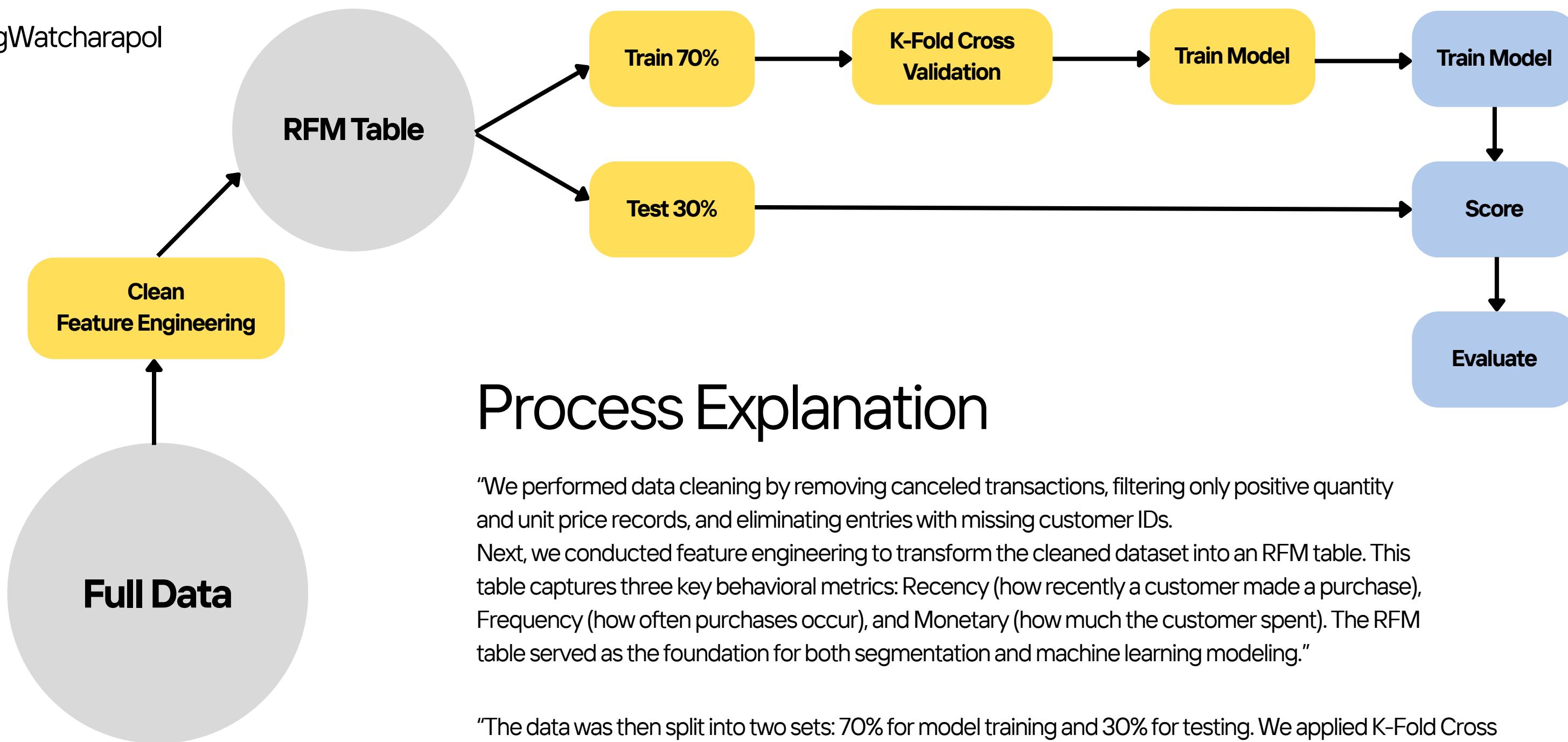
Customer Segmentation Scatter Plot



Data Set Clean : [online_retail_clean.csv](#)

RFM Table : [rfm_dashboard.xlsx](#)

Dashboard : [RFM Dashboard](#)



Process

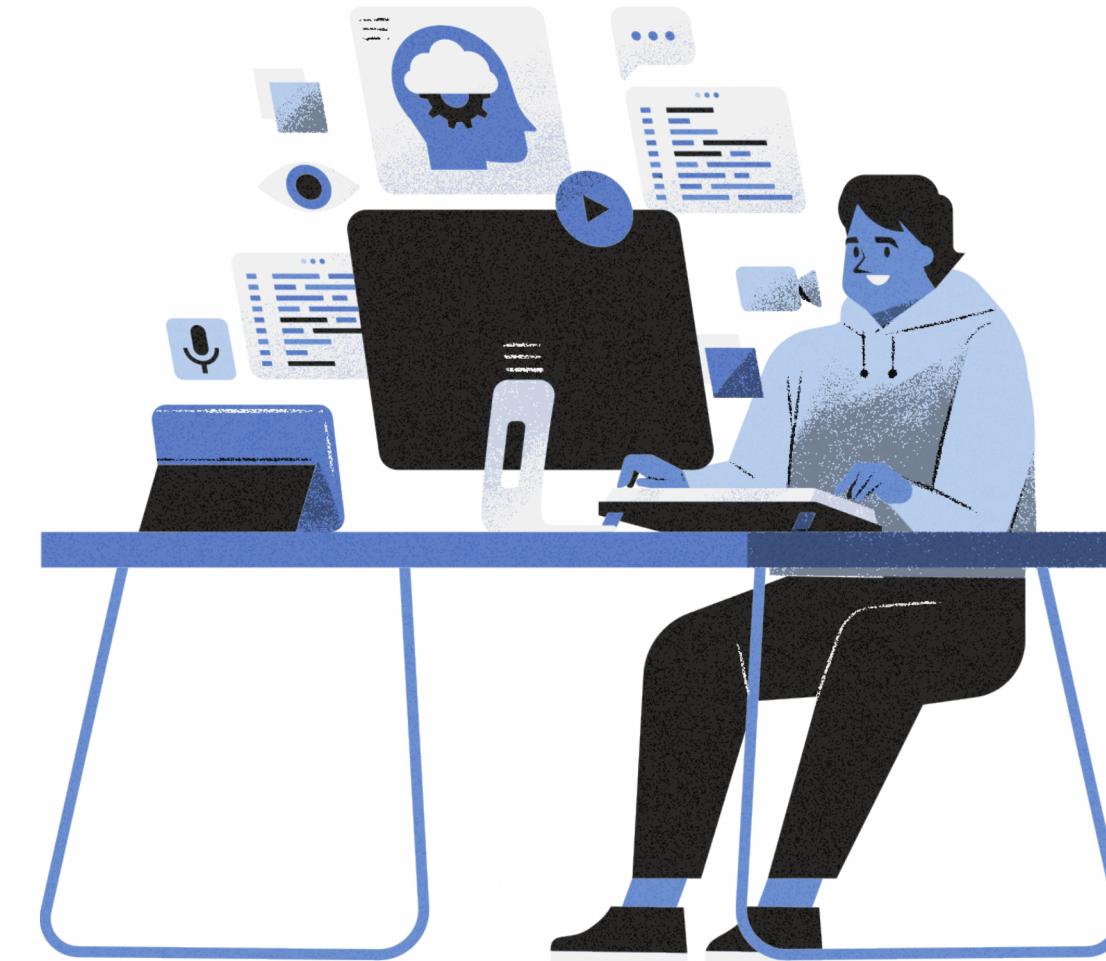
Process Explanation

"We performed data cleaning by removing canceled transactions, filtering only positive quantity and unit price records, and eliminating entries with missing customer IDs.

Next, we conducted feature engineering to transform the cleaned dataset into an RFM table. This table captures three key behavioral metrics: Recency (how recently a customer made a purchase), Frequency (how often purchases occur), and Monetary (how much the customer spent). The RFM table served as the foundation for both segmentation and machine learning modeling."

"The data was then split into two sets: 70% for model training and 30% for testing. We applied K-Fold Cross Validation on the training set to train three different machine learning models — K-Nearest Neighbors (KNN), Decision Tree, and Random Forest — ensuring robust and generalized model performance. After model training, we used the test set to predict customer segments and evaluated the models based on accuracy and stability. This process allowed us to identify the best-performing model for customer classification, supporting strategic insights and future marketing actions."

Machine Learning Models



What is a Machine Learning Model ?

A Machine Learning (ML) model is an algorithm trained to recognize patterns within data and make predictions or decisions without explicit programming. In this project, ML models are used to classify customers into predefined segments based on their behavioral features (Recency, Frequency, Monetary).

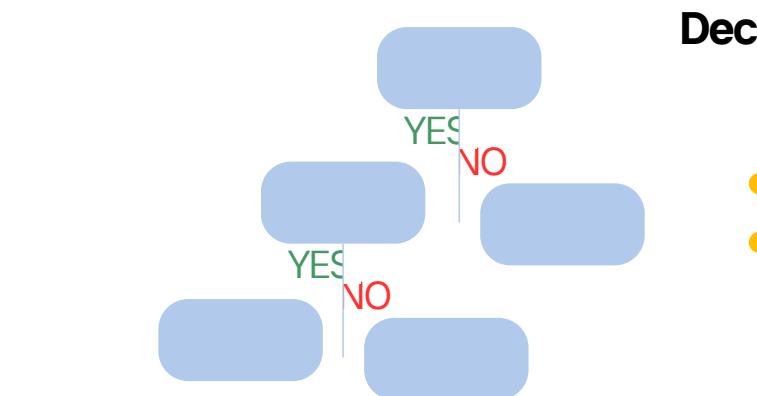
Why Use Machine Learning ?

Traditional segmentation methods group customers based on historical data but lack predictive power. Machine Learning enables automatic, scalable customer classification, improves targeting strategies, and adapts to changing customer behaviors in real time.

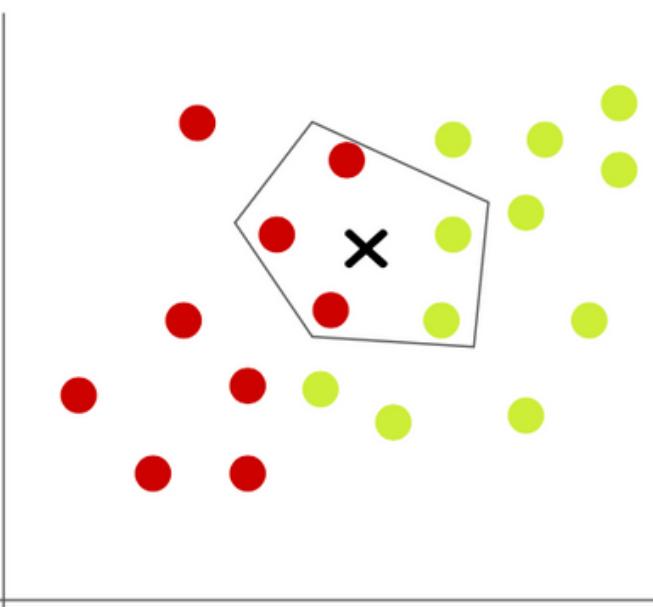
Machine Learning Models Used

- **K-Nearest Neighbors (KNN)** – Classifies based on nearby customers.
 - **Decision Tree** – Classifies using yes/no decision rules.
 - **Random Forest** – Combines many trees for better accuracy.
- Each model brings unique strengths in customer classification and prediction accuracy.

Applied ML Models



Decision Tree



K-Nearest Neighbors (KNN)



Random Forest

We used Supervised Learning because customer segments were labeled based on RFM scores.

This approach allows accurate prediction of new customer groups.



Decision Tree

Decision Tree works like a flow of yes-or-no questions, where each answer leads to a new branch until a final group is reached. It's like following paths on a tree based on simple decisions.



K-Nearest Neighbors (KNN)

KNN works like asking, "What are my closest neighbors doing?" It looks at nearby customers with similar behaviors and groups a new customer based on majority voting from these neighbors.



Random Forest

Random Forest is like asking the same set of questions to many different trees, then combining their answers to make a final decision. It gives more accurate and stable results than using just one tree.

ML Model Comparison

We compared three supervised learning models — K-Nearest Neighbors (KNN), Decision Tree, and Random Forest — for customer segmentation.

Although Random Forest achieved the highest accuracy, Decision Tree was selected for its interpretability and practical insights.



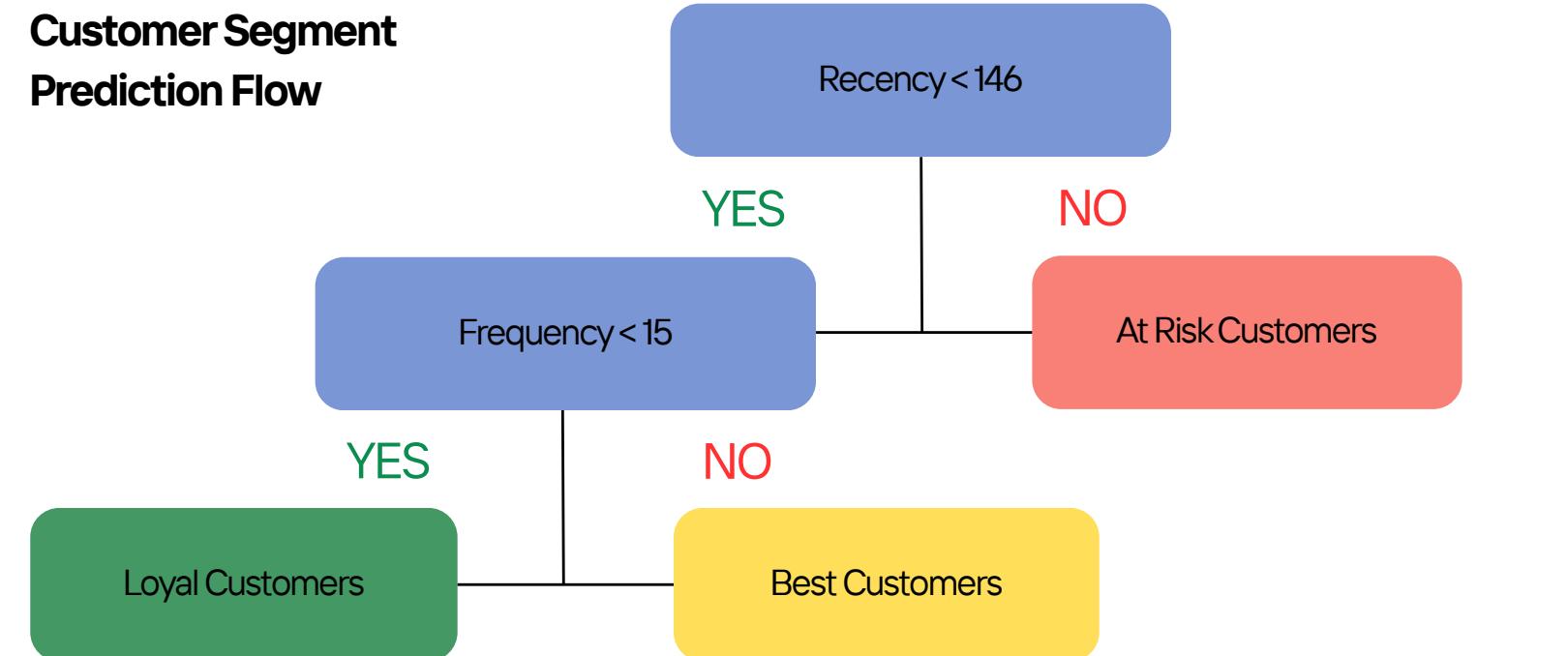
Why We Selected Decision Tree ?

Although Random Forest had the highest accuracy, the difference was minor.

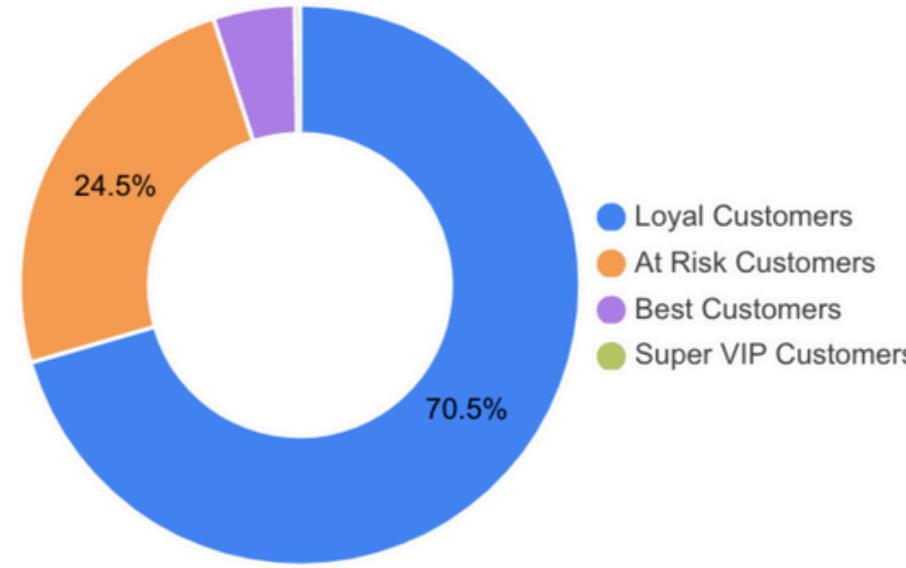
Decision Tree was chosen because it allows clear understanding of how customers are classified through simple decision rules, making it highly actionable for business users.

Model	Accuracy	Characteristics
• K-Nearest Neighbors (KNN)	95.4%	• Simple and intuitive. No training phase needed. Slower when handling large datasets.
• Decision Tree	98.6%	• Easy to interpret and visualize. Risk of overfitting with complex data.
• Random Forest	99.6%	• High accuracy and robustness. Harder to interpret compared to a single tree.

Customer Segment Prediction Flow



Customer Segment Distribution



The majority of customers (70.5%) are Loyal Customers, followed by At Risk Customers (24.5%). Best Customers and Super VIP Customers together form a small but high-value segment that should be carefully nurtured.

Segment	Characteristics	Suggested Focus
Loyal Customers (70.5%)	Frequently purchase, moderate spending, steady relationship.	Reward loyalty, offer exclusive promotions to maintain engagement.
At Risk Customers (24.5%)	Previously active but now showing declining or no recent activity.	Re-engagement campaigns with special offers to win them back.
Best Customers	Highly recent and frequent purchasers, significant spenders.	Personalize communication, provide VIP treatment, early access offers.
Super VIP Customers	Extremely high-value, high-frequency, and high-loyalty customers.	Ultra-exclusive perks, personalized services, special recognition programs.

Customer Insights

Our RFM segmentation and Decision Tree classification revealed four distinct customer groups. Each segment shows unique behaviors that can inform marketing strategies.

Understanding these customer segments enables us to allocate marketing resources effectively and maximize customer lifetime value.

Campaign Recommendations

Based on customer segmentation insights, we propose the following targeted campaigns to maximize engagement, retention, and revenue.



Loyal Customers

- Loyalty rewards program
- Early access to new products
- Exclusive membership benefits

At Risk Customers

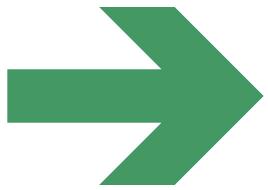
- Re-engagement emails with personalized discounts
- Limited-time comeback offers
- Customer feedback requests to regain trust

Best Customers

- Personalized VIP treatment
- Birthday and anniversary rewards
- Private sales invitations

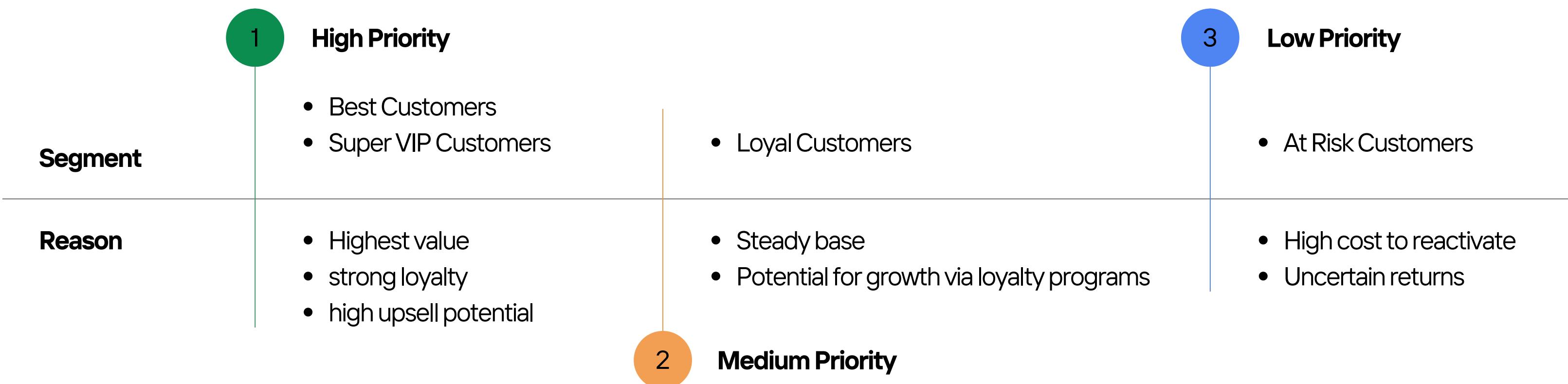
Super VIP Customers

- Ultra-exclusive event invitations
- Personal account manager services
- First priority access to limited-edition products



Focus Area Summary

To maximize marketing effectiveness, we propose prioritizing customer segments based on value contribution and potential for engagement.



Focus first on nurturing high-value customers while maintaining loyalty to ensure sustainable growth.

Thank You

- Linkedin : [Watcharapol Charoensuk](#)
- frongbook@hotmail.com
- www.frongwatcharapol.com

