Customer Segmentation Project Documentation

# Project Overview

This project focuses on Customer Segmentation using the RFM (Recency, Frequency, Monetary) model.

We use unsupervised machine learning (KMeans) to group customers based on their purchasing behavior.

This helps businesses:

- Tailor marketing strategies

- Improve customer retention

- Personalize offers and communication

# Data Preprocessing

Steps performed on the raw e-commerce dataset:

1. Removed missing values — especially in the CustomerID field.

2. Filtered out canceled orders — identified by InvoiceNo starting with 'C'.

3. Calculated TotalPrice — computed as Quantity × UnitPrice for each order.

# RFM Feature Engineering

For each unique CustomerID, we calculated:

- Recency: Days since the customer's last purchase (from a reference date).

- Frequency: Number of unique purchases made by the customer.

- Monetary: Total amount spent by the customer.

# Feature Scaling

Applied StandardScaler to normalize Recency, Frequency, and Monetary columns.

Scaling ensures fair contribution to the clustering algorithm.

# KMeans Clustering

1. Elbow Method: Ran KMeans for cluster counts 1–10 to determine the optimal number.

2. Optimal Clusters: Chose n\_clusters = 4 based on the elbow curve.

3. Fit the Model: Applied KMeans on scaled RFM data.

4. Cluster Labels: Assigned each customer to a cluster.

5. Evaluation: Used Silhouette Score to assess cluster quality.

# Visualizations

Used Seaborn's pairplot to compare clusters across RFM dimensions.

This helps visualize how customer segments differ.

# Cluster Interpretation

A possible interpretation of the 4 clusters:

- Cluster 0: High-value, loyal customers.

- Cluster 1: At-risk customers.

- Cluster 2: New or one-time buyers.

- Cluster 3: Low-value or inactive customers.

# Outputs Delivered

- Clustered Dataset with CustomerID, Recency, Frequency, Monetary, and Cluster columns.

- Visuals: Pairplots of RFM distribution.

- Silhouette Score as performance metric.

# Conclusion

With this model, businesses can:

- Retain high-value customers

- Win back at-risk customers

- Engage new buyers

- Improve marketing ROI through segmentation