RFM Analysis for Retail Shop Data Using K-Means Clustering

Introduction:

In this project, I performed an RFM (Recency, Frequency, Monetary) analysis to segment customers of a retail shop based on their purchasing behavior. I used the K-means clustering algorithm to identify distinct customer segments, which can then be leveraged for targeted marketing, personalized promotions, and overall customer relationship management.

What is RFM Analysis?

RFM analysis is a marketing technique used to analyze and segment customers based on three key factors:

Recency (R): How recently a customer made a purchase.

Frequency (F): How often a customer makes a purchase.

Monetary (M): How much money a customer spends.

These three dimensions help in understanding customer behavior, predicting future purchasing patterns, and targeting the right customer segments effectively.

Dataset Overview

The dataset used in this analysis contains transaction data from a retail shop. It includes the following features:

Customer ID: A unique identifier for each customer.

Transaction Date: The date when the customer made a purchase.

Transaction Amount: The total amount spent by the customer in that transaction.

Steps Involved in the Analysis

Data Preprocessing:

Cleaned the dataset to handle missing values.

Converted the transaction date into a datetime object to calculate the recency.

RFM Metrics Calculation:

Recency: Calculated as the number of days since the last purchase.

Frequency: The total number of purchases made by each customer.

Monetary: The total amount spent by each customer.

Data Normalization:

Since K-means clustering is sensitive to the scale of features, I standardized the RFM values to ensure they all contribute equally to the clustering process.

K-Means Clustering:

Applied the K-means clustering algorithm to group customers into distinct segments based on their RFM values.

Used the Elbow Method to determine the optimal number of clusters.

Customer Segmentation:

Interpreted the resulting clusters to understand the characteristics of each segment.

Labeled segments based on customer behavior, e.g., "High-Value Customers", "At-Risk Customers", etc.

Visualization:

Created visualizations like scatter plots and cluster heatmaps to present the results.

Key Insights

High-Value Customers: These customers have made recent purchases, purchase frequently, and spend a lot of money. They are the most valuable segment and should be the target for loyalty programs and personalized marketing.

At-Risk Customers: These customers have not purchased recently but spent a considerable amount in the past. Targeting them with re-engagement campaigns could bring them back.

New Customers: This segment includes customers who have made only one or a few recent purchases. Offering discounts or promotions could increase their frequency of purchases.

Conclusion:

RFM analysis with K-means clustering is a powerful technique for customer segmentation. By identifying distinct customer groups, businesses can tailor their marketing strategies to maximize customer retention, sales, and overall satisfaction. This analysis can be further enhanced by incorporating additional features like customer demographics and purchase categories.

Repo Structure:

Customer\_Segmentation-RFM\_Analysis.ipynb: Jupyter notebook containing the code for data preprocessing, RFM metric calculation, clustering, and visualization.

**OnlineRetail.csv** File containing the raw dataset.

Installation:

To run the notebook, clone this repository and install the necessary libraries:

git clone https://github.com/SifatSaeed/Customer\_Segmentation-RFM\_Analysis.git

cd Customer\_Segmentation-RFM\_Analysis

Requirements:

Python 3.x, Pandas , NumPy, Matplotlib, Scikit-learn, Seaborn

This post is structured to explain the methodology and results of your RFM analysis, while also making it easy for others to follow and reproduce your work. You can modify the content depending on your personal findings and any additional tools or methods you used.