

CUSTOMER SEGMENTATION

[USING RFM ANALYSIS]



BALAMURUGESH NAGARAJ

JUNIOR DATA ANALYST

**Customer Segmentation using RFM Analysis**

**1. Introduction**

Customer segmentation is a crucial process in business analytics, helping companies understand their customer base and tailor marketing strategies accordingly. This project applies **RFM Analysis** (Recency, Frequency, Monetary) and **K-Means Clustering** to segment customers based on purchasing behavior.

**2. Dataset Overview**

The dataset consists of transaction records, including:

* InvoiceNo: Unique transaction identifier
* CustomerID: Unique customer identifier
* InvoiceDate: Date of transaction
* Quantity: Number of units purchased
* UnitPrice: Price per unit

**3. Data Preprocessing**

**Description:** Data preprocessing involves handling missing values, creating necessary features, and ensuring data consistency.

import pandas as pd

from sklearn.preprocessing import MinMaxScaler

# Load dataset

df = pd.read\_csv("online-retail-dataset.csv")

# Display basic info

print(df.info())

# Check for missing values

print(df.isnull().sum())

# Fill missing values in 'Description'

df['Description'] = df['Description'].fillna('')

# Drop rows where 'CustomerID' is missing

df = df.dropna(subset=['CustomerID'])

# Convert 'CustomerID' to integer

df['CustomerID'] = df['CustomerID'].astype(int)

# Create 'TotalPrice' column

df['TotalPrice'] = df['Quantity'] \* df['UnitPrice']

# Convert 'InvoiceDate' to datetime

df['InvoiceDate'] = pd.to\_datetime(df['InvoiceDate'])

import pandas as pd

from sklearn.preprocessing import MinMaxScaler

# Load dataset

df = pd.read\_csv("online-retail-dataset.csv")

# Display basic info

print(df.info())

# Check for missing values

print(df.isnull().sum())

# Fill missing values in 'Description'

df['Description'] = df['Description'].fillna('')

# Drop rows where 'CustomerID' is missing

df = df.dropna(subset=['CustomerID'])

# Convert 'CustomerID' to integer

df['CustomerID'] = df['CustomerID'].astype(int)

# Create 'TotalPrice' column

df['TotalPrice'] = df['Quantity'] \* df['UnitPrice']

# Convert 'InvoiceDate' to datetime

df['InvoiceDate'] = pd.to\_datetime(df['InvoiceDate'])

**4. RFM Analysis**

**Description:** RFM (Recency, Frequency, and Monetary) analysis helps categorize customers based on their purchasing behavior.

# Define the most recent date in dataset

latest\_date = df['InvoiceDate'].max()

# Recency: Calculate the number of days since last purchase

recency\_df = df.groupby('CustomerID')['InvoiceDate'].max().reset\_index()

recency\_df['Recency'] = (latest\_date - recency\_df['InvoiceDate']).dt.days

recency\_df.drop(columns=['InvoiceDate'], inplace=True)

# Frequency: Count the number of unique purchases per customer

frequency\_df = df.groupby('CustomerID')['InvoiceNo'].nunique().reset\_index()

frequency\_df.rename(columns={'InvoiceNo': 'Frequency'}, inplace=True)

# Monetary: Sum total spending per customer

monetary\_df = df.groupby('CustomerID')['TotalPrice'].sum().reset\_index()

# Merge RFM Data

rfm\_df = recency\_df.merge(frequency\_df, on='CustomerID').merge(monetary\_df, on='CustomerID')

**5. Data Normalization**

**Description:** To ensure equal weighting of features, the RFM values are scaled using MinMaxScaler.

# Select features and scale the data

rfm\_data = rfm\_df[['Recency', 'Frequency', 'TotalPrice']]

scaler = MinMaxScaler()

rfm\_scaled = scaler.fit\_transform(rfm\_data)

# Convert to DataFrame

rfm\_scaled\_df = pd.DataFrame(rfm\_scaled, columns=['Recency', 'Frequency', 'TotalPrice'])

**6. K-Means Clustering**

**Description:** K-Means clustering is used to segment customers into distinct groups. The Elbow Method helps determine the optimal number of clusters.

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

# Find optimal K using the Elbow Method

wcss = []

for k in range(1, 11):

kmeans = KMeans(n\_clusters=k, random\_state=42, n\_init=10)

kmeans.fit(rfm\_scaled)

wcss.append(kmeans.inertia\_)

# Plot Elbow Curve

plt.figure(figsize=(8, 5))

plt.plot(range(1, 11), wcss, marker='o', linestyle='--')

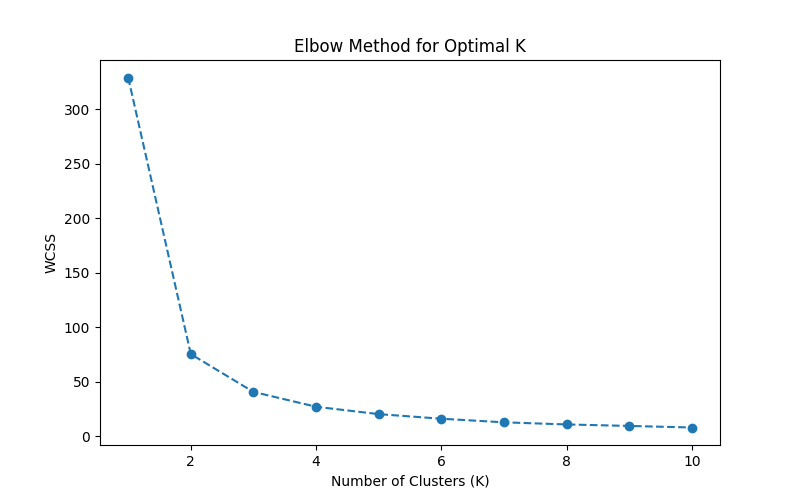
plt.xlabel("Number of Clusters (K)")

plt.ylabel("WCSS")

plt.title("Elbow Method for Optimal K")

plt.show()

**Figure 1: Elbow Method Graph**



# Apply K-Means clustering

optimal\_k = 3

kmeans = KMeans(n\_clusters=optimal\_k, random\_state=42, n\_init=10)

rfm\_scaled\_df['Cluster'] = kmeans.fit\_predict(rfm\_scaled)

**7. Clustering Results & Analysis**

**Description:** After clustering, we analyze the number of customers in each segment and their average RFM scores.

# Count customers in each cluster

print(rfm\_scaled\_df['Cluster'].value\_counts())

# Calculate cluster-wise mean values

rfm\_cluster\_analysis = rfm\_scaled\_df.groupby("Cluster").mean()

print(rfm\_cluster\_analysis)

# Cluster-wise RFM Mean Values

rfm\_cluster\_analysis.plot(kind="bar", figsize=(10, 5))

plt.title("Customer Segment Profiles")

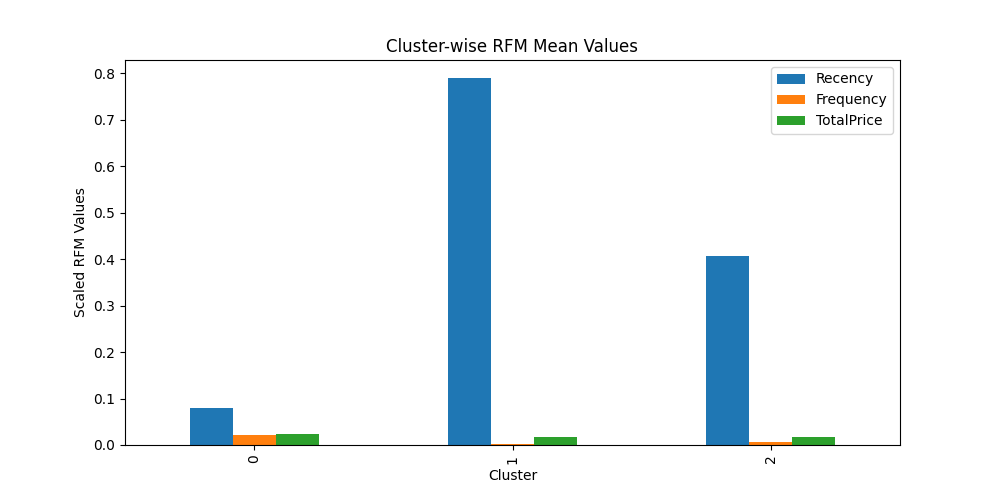
plt.ylabel("Scaled RFM Values")

plt.xlabel("Cluster")

plt.legend(["Recency", "Frequency", "TotalPrice"])

plt.show()

**Figure 2: Customer Segment Profiles**

****

# Cluster distribution

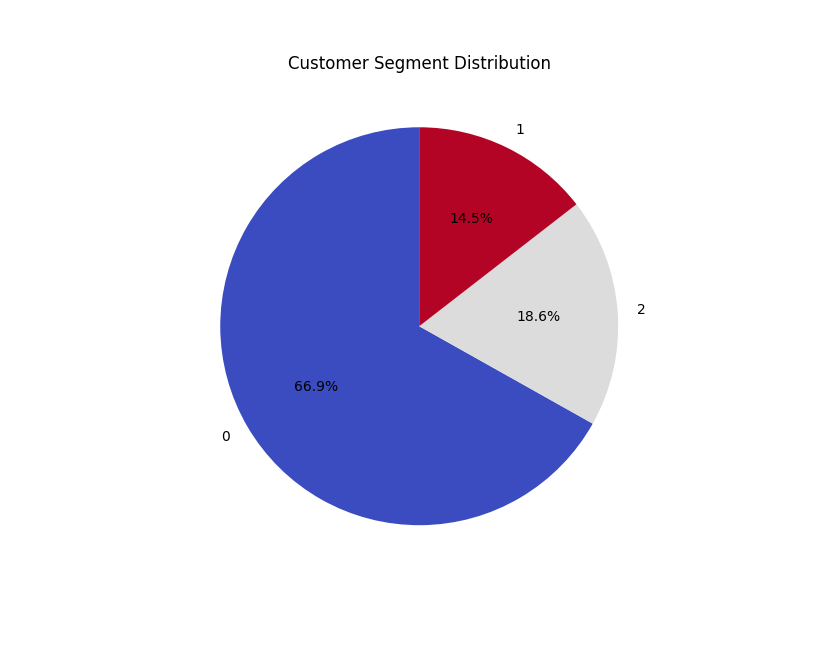
rfm\_scaled\_df['Cluster'].value\_counts().plot(kind='pie', autopct='%1.1f%%', figsize=(7, 7), startangle=90, cmap='coolwarm')

plt.title("Customer Segment Distribution")

plt.ylabel("")

plt.show()

**Figure 3: Customer Segment Distribution**



**8. Business Strategies Based on Customer Segments**

**Description:** Based on the RFM analysis, different strategies can be applied to maximize customer engagement and revenue.

**Cluster 0: High-Value Customers**

* Offer loyalty programs and exclusive discounts.
* Personalized recommendations to maintain engagement.
* Premium support services to enhance customer experience.

**Cluster 1: Low-Value Customers**

* Send re-engagement emails with targeted promotions.
* Offer small-value coupons to encourage purchases.
* Identify reasons for low spending and address them.

**Cluster 2: Mid-Value Customers**

* Provide incentives for larger purchases.
* Implement referral programs to expand customer base.
* Encourage repeat purchases through follow-up emails.

**9. Conclusion**

**Description:** This segmentation model enables businesses to make data-driven decisions and improve marketing efforts. Understanding customer behavior through RFM analysis helps in enhancing customer retention and maximizing revenue.

**Figure 4: Final Summary of Customer Segments**

