Customer Segmentation and Behavioural Analysis Using RFM Analysis

Project Overview:

In today's competitive market landscape, understanding customer behavior is vital for making informed, data-driven business decisions. This project leverages RFM analysis (Recency, Frequency, Monetary) to segment customers based on their purchasing behavior and value contribution. Additionally, time series analysis and clustering techniques are applied to uncover patterns in engagement trends and customer lifecycle stages.

To ensure the accuracy and consistency of the analysis, the data was cleaned and pre-processed using Python. While the data wrangling steps are not detailed in this report for the sake of clarity and focus, the full code and methodology are available via an accessible link.

What is RFM Analysis?

RFM Analysis is a widely used technique for segmenting customers based on their historical purchasing behavior. It focuses on three key metrics:

- Recency (R): The number of days since the customer's last purchase.
- Frequency (F): The total number of transactions made by the customer over the analysis period.
- Monetary (M): The total amount spent by the customer during the analysis period.

This model is particularly effective for businesses with repeat purchases, such as in retail or e-commerce, and helps in identifying high-value customers, inactive users, and those at risk of churning.

About the Dataset

The dataset used in this analysis comes from an e-commerce store and consists of customer transactions. It includes the following fields:

- InvoiceNo: Unique invoice number (with a "C" prefix indicating cancelled orders)

- StockCode: Unique product identifier
- Description: Description of the product
- Quantity: Quantity of items purchased per transaction
- InvoiceDate: Date and time of the transaction
- UnitPrice: Price per product unit
- CustomerID: Unique identifier for the customer
- Country: Country where the transaction took place

Methodology

- **Data Cleaning & Preparation:**
  The raw transaction data was cleaned and transformed using Python libraries such as pandas and NumPy. This included handling missing values, removing cancelled orders, and correcting data types to ensure consistency.
- **RFM Metric Calculation:**
  - Recency: Calculated as the number of days between a customer's most recent purchase and a reference date.
  - Frequency: Count of unique transactions (invoices) per customer.
  - Monetary: Sum of the total value spent by each customer.
- **RFM Scoring:**
  Each customer is scored on a scale (e.g., 1–5) for each RFM metric, with higher scores indicating more desirable customer behavior. Composite RFM scores are then used for segmentation.
- **Customer Segmentation:**
  Based on RFM scores, customers are grouped into segments such as:
  - Champions
  - Loyal Customers
  - Potential Loyalists
  - At Risk
  - Lost Customers

    - New Customers
    - Hibernating Customers
- **Visualization & Interpretation:**
  Data visualizations were created using Python libraries (matplotlib, seaborn) to explore the distribution of customer segments and provide intuitive insights into customer behavior.

Deliverables

- **Comprehensive RFM Analysis**: Cleaned and structured customer transaction data with Recency, Frequency, and Monetary scores assigned to each customer, enabling effective segmentation.
- **Strategic Business Insights**: Clear visualizations and actionable recommendations tailored to each customer segment, supporting data-driven decision-making.
- **Interactive Dashboard**: A user-friendly, internal dashboard for exploring customer segments and monitoring engagement trends in real time.