IE6400 Foundations Data Analytics Engineering
Fall Semester 2023

Group Projects

Project 2:

Topic: Customer Segmentation using RFM Analysis

Introduction:
In this project assignment, you will work with the eCommerce dataset provided (https://www.kaggle.com/datasets/carrie1/ecommerce-data) to create a Customer Segmentation model using the RFM (Recency, Frequency, Monetary) analysis method. RFM segmentation is a powerful technique used by businesses to group customers based on their recent purchasing behavior, purchase frequency, and monetary value, enabling more targeted marketing and customer engagement strategies.

Objective: Your objective is to perform RFM analysis on the dataset and segment the customers into distinct groups based on their RFM scores. These segments will provide valuable insights for marketing and customer retention strategies.

Tasks:

1. **Data Preprocessing:**
    - Import the dataset and perform necessary data preprocessing steps, including data cleaning, handling missing values, and converting data types if needed.
2. **RFM Calculation:**
    - Calculate the RFM metrics for each customer:
    - Recency (R): How recently a customer made a purchase. Calculate the number of days since the customer's last purchase.
    - Frequency (F): How often a customer makes a purchase. Calculate the total number of orders for each customer.
    - Monetary (M): The total monetary value of a customer's purchases. Calculate the sum of the total price for each customer.
3. **RFM Segmentation:**
    - Assign RFM scores to each customer based on their quartiles (or custom-defined bins). You can use quartiles (1 to 4) or custom scores (e.g., 1 to 5) for each RFM metric.
    - Combine the RFM scores to create a single RFM score for each customer.
4. **Customer Segmentation:**
    - Use clustering techniques (e.g., K-Means clustering) to segment customers based on their RFM scores.
    - Experiment with different numbers of clusters to find the optimal number that provides meaningful segments.
5. **Segment Profiling:**

    - Analyze and profile each customer segment. Describe the characteristics of customers in each segment, including their RFM scores and any other relevant attributes.
6. **Marketing Recommendations:**
    - Provide actionable marketing recommendations for each customer segment. How can the business tailor its marketing strategies for each group to improve customer retention and maximize revenue?
7. **Visualization:**
    - Create visualizations (e.g., bar charts, scatter plots, or heat maps) to illustrate the RFM distribution and the clusters formed.
8. **Documentation:**
    - Prepare a detailed report or presentation summarizing your findings, methodology, and recommendations.
    - Include code documentation and comments for clarity.

Find the solutions to these questions:

1. **Data Overview**
    - What is the size of the dataset in terms of the number of rows and columns?
    - Can you provide a brief description of each column in the dataset?
    - What is the time period covered by this dataset?
2. **Customer Analysis**
    - How many unique customers are there in the dataset?
    - What is the distribution of the number of orders per customer?
    - Can you identify the top 5 customers who have made the most purchases by order count?
3. **Product Analysis**
    - What are the top 10 most frequently purchased products?
    - What is the average price of products in the dataset?
    - Can you find out which product category generates the highest revenue?
4. **Time Analysis**
    - Is there a specific day of the week or time of day when most orders are placed?
    - What is the average order processing time?
    - Are there any seasonal trends in the dataset?
5. **Geographical Analysis**
    - Can you determine the top 5 countries with the highest number of orders?
    - Is there a correlation between the country of the customer and the average order value?
6. **Payment Analysis**
    - What are the most common payment methods used by customers?
    - Is there a relationship between the payment method and the order amount?
7. **Customer Behavior**
    - How long, on average, do customers remain active (between their first and last purchase)?
    - Are there any customer segments based on their purchase behavior?
8. **Returns and Refunds**

    - What is the percentage of orders that have experienced returns or refunds?
    - Is there a correlation between the product category and the likelihood of returns?
9. **Profitability Analysis**
    - Can you calculate the total profit generated by the company during the dataset's time period?
    - What are the top 5 products with the highest profit margins?
10. **Customer Satisfaction**
    - Is there any data available on customer feedback or ratings for products or services?
    - Can you analyze the sentiment or feedback trends, if available?

The student would explore various aspects of the dataset, from customer behavior to sales trends and profitability. You can use data visualization techniques and statistical analysis to derive insights and answers to these questions.

Deliverables:

1. Jupyter Notebook or Python script containing your code.
2. A report summarizing your analysis, insights, and recommendations.

Evaluation Criteria:

- Accuracy and correctness of the RFM calculations and clustering.
- Quality of customer segments and their descriptions.
- Clarity and depth of analysis.
- Practicality and effectiveness of marketing recommendations.
- Presentation and documentation quality.

Submission Deadline: December 1st, 2023

Note: You are encouraged to explore additional techniques and tools beyond the basic RFM analysis if you feel it would enhance your project. Good luck with your Customer Segmentation project!