Enhancing Customer Segmentation in Virtual Shopping through RFM Analysis

Nilam N. Parmar[1], Dr. Sweta S. Panchal[2]

[1]Research Scholar, Computer Science and Engineering, Dr. Subhash University, Gujarat, India
[2]Associate Professor, Electronics and Communication, Dr. Subhash University, Gujarat, India

---***---

Abstract - *Customer segmentation in virtual shopping is enhanced through RFM analysis, focusing on the recency, frequency, and monetary value of purchases. RFM analysis allows for a deeper understanding of customer behaviors and preferences, enabling businesses to tailor their marketing strategies effectively. By categorizing customers based on their past purchasing patterns, businesses can optimize their approaches for different customer segments. Analyzing the recency, frequency, and monetary value of purchases provides valuable insights that can guide targeted marketing efforts. This study emphasizes the importance of utilizing RFM analysis in virtual shopping environments to improve customer segmentation and enhance overall marketing strategies. By leveraging RFM analysis, businesses can better understand their customers' needs and preferences, leading to more personalized and effective marketing campaigns.*

Key Words: Customer segmentation, Virtual shopping, Recency, Frequency and Monetary Value, RFM Analysis

1. INTRODUCTION

Marketers are aware that customers have diverse needs and desires. To effectively identify and understand different customer groups, companies have employed various segmentation criteria and techniques, enabling them to offer tailored products and services that meet these varying needs. Segmentation also allows companies to create profitable segments and respond strategically to these segments based on their competitive strengths. However, many marketers struggle with accurately identifying the right customer segments for organizing marketing campaigns [1].

To formulate a marketing strategy, segmentation is employed to cluster customers based on their loyalty criteria. Segmenting the customer base is one of the initial steps in developing a business model. This process involves dividing a customer base into uniform subgroups, each considered a distinct marketing audience. Customer segmentation helps quantify customer value, enabling businesses to identify high-revenue clients and those who generate less revenue [2].

In data segmentation, customers are grouped into sets of individuals with distinct similarities. Relevant attributes for customer segmentation include gender, age, lifestyle, location, purchase behavior, and income behavior. These attributes are primarily categorized based on historical purchasing behavior, which can lead to specific outcomes, such as increased sales and higher profits for the company. [3].

Clustering involves dividing or grouping customers based on their interactions with the company, whether direct or indirect. Customer data can include metrics such as time spent on social media platforms, transaction data, or time spent on specific posts. This paper focuses on the transaction data of customers from a UK online retail ecommerce platform. Although the dataset contains numerous attributes, selecting the most relevant ones is crucial for optimal results. To address this challenge, many data scientists prefer using the K-means algorithm, an unsupervised learning method, in conjunction with the RFM model. RFM stands for recency, frequency, and monetary value of a customer. The collected data can then serve as a foundation for customer segmentation [4].

The aim of this paper is to identify the type of customer (super customers, intermediate customers, base customers) and determine their value so that companies can discern which customer classes generate substantial revenue and which do not. This information will help companies develop new market strategies to improve their revenue growth [4].

2. LITERATURE SURVEY

Recency, frequency, and monetary (RFM) analysis is an effective method for market segmentation and behavioral analysis. The main advantage of the RFM model is its ability to provide a detailed behavioral analysis of customers, grouping them into homogeneous clusters. Additionally, it helps develop a marketing plan tailored to each specific market segment. RFM analysis enhances market segmentation by examining when customers made purchases (recency), how often they made purchases (frequency), and how much money they spent (monetary). Customers who have bought most recently, most frequently, and have spent the most money are more likely to respond to future promotions [5].

The strength of the RFM model lies in its use of several observable and objective variables, all of which are derived from each customer's past order history. These variables are categorized according to three independent criteria: recency, frequency, and monetary value. Recency refers to the time interval between a customer's last purchase and a specified reference point, with a shorter interval indicating a higher likelihood of repeat purchases. Frequency measures the number of transactions a customer has made within a particular timeframe, while monetary value represents the amount of money spent during that period. Traditionally, the RFM model is applied by sorting customer data according to each RFM variable and dividing them into quintiles. Segmentation starts by ranking customers based on recency, followed by frequency, and then monetary value [5].

The RFM model is highly effective in building customer segmentation models. However, its simplicity can limit its power, as improvements often require manual adjustments. Additionally, RFM models may struggle to adapt to changes in the business environment, necessitating ad hoc managerial decisions. This paper identifies optimal definitions for R, F, and M to create a dynamic RFM model. By incorporating K-Means clustering, we propose an R+FM model that dynamically builds customer segmentation models [6].

The researcher employs the RFM (Recency, Frequency, and Monetary) model, a widely used approach for segmenting customers based on their last visit time, visit frequency, and revenue generated for the company. One of the primary reasons for persisting with the RFM model is its simplicity and ease of implementation within companies. Moreover, RFM is readily comprehensible to managers and marketing decision-makers. [6].

3. METHODOLOGY

There are multiple distinct steps in the procedure. Additionally, the dataset's available data will be preprocessed using the Knowledge Discovery technique as part of the first phase. After that, the important data will be extracted for training using a machine learning method.

After that, the Min-Max normalization method is utilized. Finding the probability that a score will fall into the dataset's normal distribution is its main goal. Clustering is used in this procedure to examine unlabeled data and find clusters. This stage involves grouping the data according to how similar they are to one another in terms of a particular attribute.

Finally, Silhouette Analysis is applied to assess how close the data points are to one another in each cluster. This makes it easier to create a graphical depiction that accurately represents the classification outcomes produced by the algorithm.

3.1 Data Description

The dataset utilized comprises transactional information from an online retail business based in the UK, aimed at clustering analysis. It consists of 525,461 records with 8 distinct attributes:

Invoice: Each transaction is assigned a unique integral number, representing the invoice number. If a transaction is cancelled, it is marked with a 'c'.

StockCode: A unique nominal number assigned to each product, serving as its product code.

Description: This attribute denotes the name of the product, categorized as nominal.

Quantity: Numeric data representing the quantity of each product involved in a transaction.

InvoiceDate: Indicates the date and time of each transaction, recorded as numeric data.

Price: Numeric data representing the unit price of a product, denoting the product price per unit.

Customer ID: Each customer is assigned a unique integral number, serving as their customer ID.

Country: Denotes the country name of the customer involved in the transaction.

This dataset offers a comprehensive view of the online retail transactions, facilitating insights into customer behavior, product popularity, and geographic distribution.

Table -1: Feature Analysis

Data columns (total 8 columns):				
#	**Column**	**Non-Null**	**Count**	**Dtype**
0	Invoice	370929	non-null	object
1	StockCode	370929	non-null	object
2	Description	370929	non-null	object
3	Quantity	370929	non-null	int64
4	InvoiceDate	370929	non-null	datetime64[ns]
5	Price	370929	non-null	float64
6	CustomerID	370929	non-null	object
7	Country	370929	non-null	object
8	Amount	370929	non-null	float64

```
Summary..
Number of invoices:  17612
Number of products bought:  17612
Number of customers: 3969
Percentage of customers NA:  0.0 %
Average quantity of product purchased by a customer:  1121.0
Average revenue generated per customer:  1868.17
Average product quantity sold per transaction:  12.0
Average revenue generated per transaction:  19.99
```

Fig -1: Data Analysis of each Attribute

The study of Table 1 and Fig. 1 above shows that the dataset contains a great deal of rich and diverse information, offering many chances for extraction and research. Owing to its extensive feature set, which includes transaction history, product details, and consumer demographics, the dataset presents a wide range of analytical opportunities. It makes it possible to do thorough analyses of consumer behavior, product popularity, long-term purchasing trends, and relationships between different factors. The depth of the dataset also makes it easier to divide up the consumer base into discrete groups, which allows for more specialized marketing and customized tactics.

3.2 Model Description

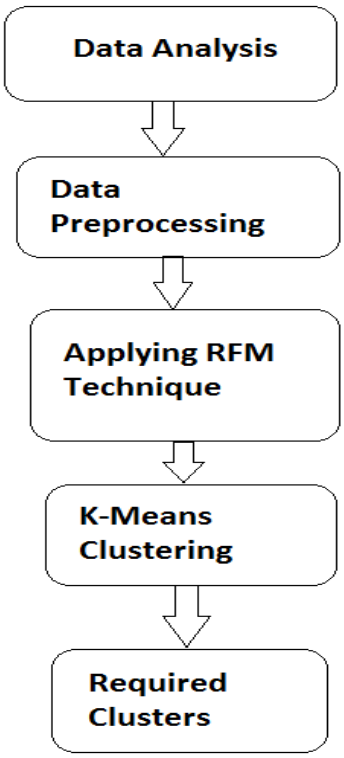

Fig -2: Cluster Analysis Model

The model consists of the following phases to acquire the cluster, which is nothing more than customer segmentation, as illustrated in Fig.-2 Here, the entire dataset was first obtained and loaded, and then all preprocessing or cleaning of the data was completed by applying methods such as scaling, addressing missing values, and removing duplicate values.

Using the RFM model to calculate each customer's Recency, Frequency, and Monetary score; finally, using the K-means method to place the client in the appropriate cluster.

3.3 Data Preprocessing

A CSV file containing the dataset discussed earlier was imported to create a new data frame. Through a series of preprocessing steps, the data was cleaned and aggregated as per the specified requirements. The original dataset, comprising approximately 525,461 tuples, was refined to a retail sales dataset containing 370,929 tuples.

During preprocessing, NA values were removed, and a new attribute named Revenue was introduced, calculated by multiplying the quantity by the unit price. Additionally, two additional datasets were created through aggregation: customer aggregate data and invoice aggregate data.

Several steps were taken to clean the data further: invoices unrelated to customers were discarded, along with duplicate entries, fully cancelled transactions, and entries with missing values in the Customer ID column. These measures ensured the integrity and quality of the dataset for subsequent analysis.

3.4 RFM Technique

RFM (Recency, Frequency, Monetary) analysis serves as a robust marketing tool for customer segmentation, drawing insights from customer behavior. This model categorizes customers based on their transactional patterns, focusing on how recently, how often, and how much they have purchased. By leveraging RFM analysis, businesses can identify clusters of customers who may be at risk of discontinuing their relationship with the company.

1. **Recency**: This metric gauges the freshness of customer activities by measuring the time elapsed since their last transaction.

2. **Frequency**: Frequency assesses the rate of customer transactions, often quantified as the total number of recorded transactions.

3. **Monetary**: Monetary value encapsulates the total amount spent by each customer, indicating their overall transactional value.

The RFM factors shed light on key insights:

1. Recent purchases indicate heightened responsiveness to promotional efforts from customers.

2. Higher frequency in purchases typically reflects greater customer engagement and satisfaction.

3. Monetary value helps differentiate between customers who are heavy spenders and those who make lower-value purchases.

Employing RFM analysis allows businesses to tailor their marketing strategies more effectively, targeting specific customer segments with tailored approaches based on their transactional behaviors.[1]

3.5 RFM Implementation

The process of RFM calculation involves several stages:

1. Recency Calculation: Recency is determined for each client by subtracting their most recent transaction date from the most recent date found in the entire dataset, which in this case is December 11, 2010. This reference date is used to calculate the recency value for each customer.

2. Frequency Calculation: Frequency is calculated by summing up the total number of purchases made by each individual customer, providing insights into how often they engage in transactions.

3. Monetary Computation: Monetary value is computed by considering the quantity of each product purchased and its corresponding unit price. By summing up the totals of all transactions made by each customer, their overall expenditure is determined.

These steps enable businesses to gain a comprehensive understanding of their customers' behavior, allowing for more targeted marketing strategies and personalized approaches to customer engagement.

Table-2. Result for Recency, Frequency and Monetary Calculation

	Customer ID	Recency	Frequency	Monetary
0	12346.0	160	11	214.24
1	12608.0	35	1	64.90
2	12745.0	117	2	89.52
3	12746.0	171	1	15.00
4	12747.0	0	16	834.11
...	...	...	...	...
3964	18283.0	13	6	13.94
3965	18284.0	62	1	50.00
3966	18285.0	291	1	29.90
3967	18286.0	107	2	35.00
3968	18287.0	13	4	80.40

From Table 2 above Customers are given values for the Recency, Frequency, and Monetary parameters in the first stages of RFM analysis. The customer list is then divided into tiers according to three dimensions (R, F, and M), which serve as the foundation for specialized advertising and marketing efforts. With the help of this smart methodology, firms are able to successfully customize their plans, boosting corporate impact and consumer engagement.

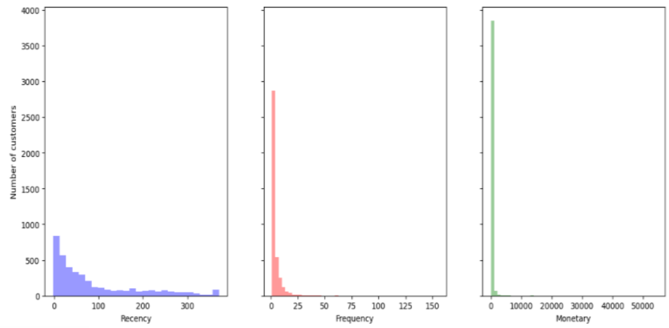

Chart -1: RFM Visualization

Chart-1 graph assists businesses in modifying their marketing strategies and product offerings by displaying the distribution of customers across different recency, frequency and monetary divisions. Higher bars indicate a larger proportion of customers falling into that demographic, whereas shorter bars indicate fewer people. through examining the distribution of the bars in the areas of monetary, frequency, and recency.

4. CONCLUSION

The findings of this study can be mapped out and potential clients identified using a decision support system in the lending industry. By employing the RFM approach for item segmentation and clustering by item category, we may increase the recommendation accuracy and better represent the attributes of the items. Consequently, we are able to suggest the k-means clustering of item category based on RFM personalized recommendation system.

This study's primary goal was to use the RFM model to separate the customer IDs from the 525461 transaction data sets.

5. FUTURE WORK

Future work could focus on refining the decision support system by integrating advanced machine learning techniques to enhance the accuracy of item segmentation and clustering. Additionally, exploring the implementation of real-time data analysis and incorporating customer feedback could improve the personalization and effectiveness of the recommendation system. Expanding the scope to include more diverse datasets and testing the model in different industry contexts would provide valuable insights and broaden its applicability.

Furthermore, investigating the integration of other customer behavior metrics beyond RFM could offer a more holistic approach to customer segmentation and recommendation accuracy.

REFERENCES

[1] Ms Sarika Rathi and Prof. Vijay Karwande, "Review Paper on Customer Segmentation Approach Using RFM and K-Means Clustering Technique," © 2022 IJCRT,Volume 10, Issue 12 December 2022 ISSN: 2320-2882.

[2] Ali Khumaidi, Herry Wahyono, "RFM-AR Model for Customer Segmentation using K-Means Algorithm" E3S Web of Conferences 465, 02005 (2023) ICIMECE 2023, DOI:10.1051/e3sconf/202346502005

[3] P. Anitha, Malini M. Patil "RFM model for customer purchase behavior using K-Means algorithm" https://doi.org/10.1016/j.jksuci.2019.12.011 1319-1578/ 2019 The Authors. Published by Elsevier B.V. on behalf of King Saud University.

[4] Rahul Shirole, Laxmiputra Salokhe "Customer Segmentation using RFM Model and K-Means Clustering" International Journal of Scientific Research in Science and Technology Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com) doi : https://doi.org/10.32628/IJSRST2183118.

[5] Rachid AIT DAOUD, Belaid BOUIKHALENE, Abdellah AMINE "Combining RFM Model and Clustering Techniques for Customer Value Analysis of a Company selling online" 978-1-5090-0478-2/15/$31.00 ©2015 IEEE.

[6] Nikita Bagul, Prerana Berad, Prof. Priya Surana "Retail Customer Churn Analysis using RFM Model and K-Means Clustering" International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org ISSN: 2278-0181 IJERTV10IS030170 (This work is licensed under a Creative Commons Attribution 4.0 International License.) Published by : www.ijert.org Vol. 10 Issue 03, March-2021.