Research Paper

EPRA International Journal of Economic and Business Review-Peer Reviewed Journal
Volume - 12, Issue - 7, July 2024 | e-ISSN: 2347 - 9671| p- ISSN: 2349 – 0187

SJIF Impact Factor (2024): 8.808 || ISI Value: 1.433 || Journal DOI URL: https://doi.org/10.36713/epra2012

CUSTOMER SEGMENTATION USING RFM ANALYSIS

Bharghav Madhiraju[1], Sukesh Reddy[2], Dr. G Sasikala[3]

[1]*R22DG007,Department of Computer Science and Applications, REVA University,Bengaluru, India*
[2]*R22DG075,Department of Computer Science and Applications, REVA University,Bengaluru, India*
[3]*Head of Department,Department of Computer Science and Applications, REVA University,Bengaluru, India*

ABSTRACT **DOI No: 10.36713/epra17685** **Article DOI: https://doi.org/10.36713/epra17685**

This research presents a comprehensive approach to customer segmentation using Recency, Frequency, and Monetary (RFM) analysis, combining statistical insights, data visualization, and machine learning techniques. The study utilizes a real-world dataset obtained from a retail environment, aiming to categorize customers based on their recent purchasing behavior, visit frequency, and monetary contributions to the store. The code begins with data preparation and exploration, ensuring data integrity by addressing issues such as negative quantities and missing customer identifiers. Following this, the Recency, Frequency, and Monetary metrics are computed, providing a holistic view of customer engagement and spending patterns. Visualizations, including violin plots, histograms, and box plots, are employed to intuitively convey the distribution of these metrics. The research then delves into the quantile-based segmentation of customers, allowing for a more granular classification. Quantiles are calculated to divide customers into four segments for each RFM metric. The resulting quantile labels are applied to the dataset, enabling the creation of a compound RFM quantile that combines recency, frequency, and monetary information. This combined quantile facilitates the definition of distinct customer segments. To further enhance the interpretability of customer segments, the study introduces a set of rules for labeling customers based on their RFM quantiles. These rules yield segments such as "Best Customer," "Loyal Customer," "Big Spender," "Dead Beats," and "Lost Customer." The resulting customer segmentation is presented visually through histograms and a pie chart, providing a clear and concise representation of the distribution of customers across different segments. Moreover, the research integrates machine learning models, including XGBClassifier, and CatBoostClassifier, to explore the potential of automating the segmentation process and predicting customer segments based on historical data. However, the machine learning aspect is introduced with commented-out sections, leaving room for further exploration and experimentation.

In conclusion, this research contributes a comprehensive and detailed code implementation for RFM-based customer segmentation. The integration of visualization techniques aids in the interpretation of customer behavior, while the inclusion of machine learning models opens avenues for predictive analytics in customer segmentation. The presented approach provides valuable insights for businesses seeking to tailor marketing and customer relationship strategies based on individualized customer segments.

KEYWORDS— ***Customer Segmentation, Frequency, Monetary Value, Recency, RFM Analysis.***

I. INTRODUCTION

RFM analysis is a type of marketing procedure that is usually used to group the customers and quantitively rank the customers based on the recency, frequency, and monetary values of their transactions. The analysis is usually performed to identify the customers along with the targeted marketing campaigns. Recency, Frequency and Monetary are the three quantifiable factors in the RFM Model. The Recency is generally classified based on what is the time elapsed since the last transaction and activity that happened with the brand. Here the activity refers to the interaction or transaction taken place from the customer with the brand. Next factor or metric in the RFM analysis is Frequency, the frequency refers to how frequently does the customer interact with the brand during a particular time of duration. The customers who are frequently engaged with the brand are treated as more loyal, than the customers who rarely interact or communicate with the brand. Monetary values are another metric in RFM analysis that reflects on the money spent by the customer on the brand in the particular duration of time. The customers who usually spend more amount are segmented under the big spender category.

Usually, the RFM analysis help the marketers for targeting the specific customer clusters with the communications much relevant for generating much higher rate of customer lifetime value and higher rate of response and added with increased loyalty. Based on the recent purchases, transaction frequency and total spending is the prime agenda behind the RFM analysis. The RFM analysis usually categorize the customers into different segments and more targeted and effective marketing strategies were tailored to their specific spending patterns. This analysis on the Recency, Frequency and Monetary values allows the implementation of effective marketing strategies in the businesses.

II. LITERATURE REVIEW

In [1] customer segmentation has been discussed, where it states that customer segmentation is the method of targeting the customers for increasing the revenue of the company. With the help of streamlit a web application has been developed for the segmentation purpose. The real time paper has been used for customer segmentation. The data consists of various variables like age of the customers etc. The main reason behind the application is to complete the segmentation of the customer automatically. Python and Machine Learning approach has been used to obtain the result. To obtain the clusters the Elbow method has been used. The obtained variables are fitting into the K-Means model. The details and the mentalities of the various customers are achieved using the plot, which will be useful for increasing their sales and improving their products as well.

In [2] it has been stated that in the customer relationship management customer value analysis is key work. For the analyzing the customer value Recency, Frequency and Frequency (RFM) model is used and for clustering the similar type of customers K- means clustering is used. With the obtained results various the customers have been classified into various categories. A crawlers program using python has been compiled to extract the data from an enterprise’s information management system. Based on the purchase frequency and transaction amount the customers are basically classified in to four categories as C1, C2, C3 and C4. Some advance algorithms can be used for obtaining the more accuracy.

In [3] states that the problem of High Utility Itemset Mining (HUIM) has been studied in the recent times. The HUIM algorithms are only used for revealing the profitable itemset from the transaction databases, but item sets are generalized. The datamining techniques are useful only for reflecting the sales trend of the customers and these cannot be used for creating the marketing strategies. Due to the limitations of HUIM analysis on the customer behavior maintenance of the specific customers became a difficult task. Here RFM miner is proposed for finding the RFM patterns which are very recent, profitable, and frequent for the transactional databases. For calculating the upper bounds like transaction-weighted utilization, subtree and local utility, the algorithm relies on the array-bin structure in the linear time and space. Even in small projected database the RFM miner searches for the extension items in the itemset. Four real-life and synthetic datasets have been used and it has been concluded that the RFM- Miner is performing very well in terms of runtime as well memory consumption. In the case of dense datasets, the RFM Miner had achieved the better performance than the state-of-the-art.

In [4] the RFM analysis has been performed on the official Indian elections results of the year 2019. The data consists of the past 30 years (i.e. 1989- 2019). In this analysis the RFM model has been used along with the K- Means Clustering and XG Boost for the clustering the score of the various candidates who are participating in the general Lok Sabha elections. It is concluded that the machine learning models can be used for analyzing the candidate’s party loyalty and ability to secure the votes.

In [5] The key function of the marketing department in the industries is to find out the potential users for the specific products. With the use of Machine Learning and Text Mining a new method has been proposed to find the method for finding the high value ranked audience for the specific brand. For building the non-target and target dataset has been collected for the 10 Twitter accounts. Five data resampling has been assessed using five datasets to resolve the program of imbalance. Bagging and Boosting algorithms of the ensemble learning are used for building the classifier. SMOTE outperforms has been the outperforming resampling method and from the classifier models the AdaBoosting is the outperformer.

In [6] The RFM analysis is to analyze the customer behavior any of the organization. Here the analysis has been performed using the RFM and SQL fuzzy. By using the general marketing strategies, the customers who are falling around the border clusters cannot be clustered. The fuzzy database is obtained by creating the fuzzy SQL database, where the data can be updated after the retrieval of the data. To address the disadvantages of the conventional RFM the proposed fuzzy database can be implemented.

In [7] With the behavior of the products and interaction with the products by the customers, the proposed research work can be used for executing the new marketing strategies and marketing methodologies for the profit. In the present work the focus is on the challenges faced during the analysis of the consumer buying patterns and interaction of the customer with the products. A study has been conducted on the various aspects such as purchasing patterns of the customers, obtaining the customer lifetime value, customer trajectory determination which helps us in the attractive shelf of the customer and for the determination of the high value customer. With the help of the Apriori algorithm, the association rules have been developed. The machine learning models that are used for the segmentation of the customer are K-Mean, Arima model and Agglomerative. The customer value standardization is calculated using the BG/NMD model and gamma-gamma model. Along with these algorithms the image processing algorithms, video analysis algorithms and customer route tracking were also performed by using different colors. The CCTV was used for tracking the patterns of the purchase and the interaction of the customers with the products.

In [8] At the Amazon.com, the revenue of 35% is specifically has been obtained by the Operation Analysis and Recommendation System. These systems play the crucial role for the increase in the average order value by the recommendations during the online shopping. With the help of the intelligent predictions, big data and data mining this research focuses on the recommendation system for the prediction of the product that a person is most likely to purchase based on the customers browsing data and shopping history. The main agenda of the work is to build a full stack recommendation engine by using Apriori algorithm, Python and RFM analysis.

In [9] The customer analysis has been performed for on the combined experimental data provided by the BilliBilli and the K-Means is used for the customer cluster analysis. The model that is proposed is useful for the customer value evaluation in the video-on-demand service platform. Combining the characters, a customer value evaluation model and customer value evaluation RFMD (Recency, Frequency, Monitory and Duration) model is suggested. At last, it is concluded that the RFMD model gives the good consumer value judgment. The same can be used for the building the accurate recommendation systems and for developing high quality-customers since the RFMD model has the very good customer value judgment.

In [10] Recency, Frequency and Monetary (RFM) model and classification model like logistic regression, neural network, decision tree and ensemble model have been used for the purpose of increasing the predictive accuracy. Neural network has shown the 2x accuracy more than the RFM model from 42.9% to 78.2%. In case of requirement of strong explanation, the decision tree can be used with sacrifice of 2% loss in accuracy. When the predictive model has been used there is the improvement in sampling and reduction in the computing power along with positive response rate targeting. The study should have focused on the false positive transaction so that the model will go more forward.

III. METHODOLOGY

In this section we are going to see the detailed understanding on the dataset, source of the dataset, RFM model and the implementation of the segmentation of customers using the Recency, Frequency and Monetary metrics.

A. Dataset Description

The dataset considered for the present RFM analysis is Superstore dataset, which is obtained from the Kaggle. The dataset considered consists of 8 columns with the 541909 entries. Here is the brief description about the various attributes present in the dataset. Let us have a look into the attribute:

- **Invoice No**: This is the attribute where the Invoice number is the unique number assigned to each sale or transaction performed. Generally, this is useful for customers and businesses for tracking and managing their respective orders.
- **Stock code**: In the E- Commerce dataset considered, the stock code acts as the unique identifier or the unique code that is assigned to the products that are present in the inventory.
- **Description**: In the column named description in the present dataset the description about the products is given.
- **Quantity**: The quantity attribute in the dataset gives the with what quantity the product has been sold in the superstore.
- **Invoice Date**: Here in the E-Commerce dataset, the invoice date is the date on which the sale of the invoice has been issued for the particular transaction. This attribute is an important one because with this attribute we can obtain the tracking of sale occurred at the time of generation of corresponding invoice.
- **Unit price**: The term unit price refers to the price of the single product sold in the store. For calculating the total cost of the quantity of the items this attribute is used. The product of unit price and quantity usually gives the total cost of the product.
- **Customer ID**: This attribute is the unique identifier that is usually used for distinguishing the customers from each other. The customer id usually used for associating the orders, transactions, and other activities for associating the specific customers.
- **Country:** The attribute country in the dataset is generally used for tracking the transactions and sales taking place in the region. This is also used for managing the inventory across the various locations, getting the knowledge about the patterns in the regional sales. This attribute is also useful for getting the data about the marketing efforts in various regions.

B. RFM Model

RFM model is the one of the most famous models used for the segmenting the customers based on their Recency, Frequency and Monetary values. For performing customer segmentation, the data is required from the E-Commerce domain. The reason behind considering the E-Commerce domain is because it consists of the attributes based on the transactions and sales. In the Fig. 1 the architecture of the RFM model.

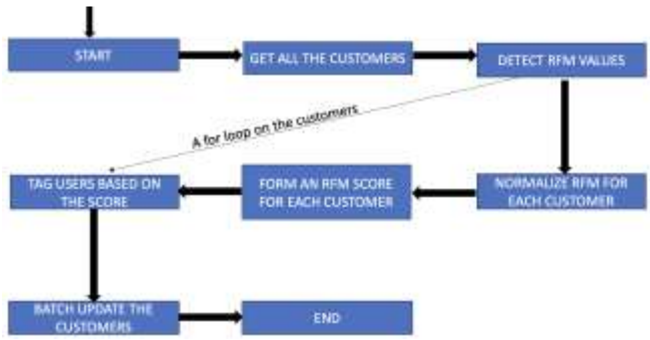


Fig. 1: Architecture of RFM Segmentation

Coming to the working of the working of the model considered for the customer segmentation the model starts working with the collecting the customers data from the E-Commerce domain. The preprocessing of the customer data collected is the second that should be performed. In the preprocessing of the data the cleaning of the noise present in the collected data takes place. After the preprocessing of the data the Recency, Frequency and Monetary values are identified for each individual customer. Next step in the segmentation part is to normalize the RFM score for each customer. Here normalizing the score refers to assigning the RFM values between 0 and 1. This normalization is optional.

Then the customers are tagged based on their respective RFM score obtained. Once the RFM scores of the customers the clusters are created based the various categories. There are 3 basic clusters in which the clusters are created. The 3 clusters created are Best Customers, Middle level Customers, Bad Customers. Basically, the Best Customers are the group of customers who are most frequent visitors and key revenue generators. The Mid-Level customers are the one who generates the less revenue compared to the Best Customers. Bad Customers are the one who creates the least revenue and sales compared to all the other type of customers. The customers who come new to the stores are the key people whom the brand can convert into the best customers. Once the clustering is done for the customer as per the requirements then the segmentation of the customers ends. Then based on these segments further strategic steps can be initiated for the increase in the sales and profits of the brand.

Here are the formulas that are used to obtained the Recency, Frequency and Monetary values in the RFM Analysis:
Recency = Recent- Last Date of Transaction. (1)
Frequency = Total number of unique customers/Total number of transactions (or visits). (2)
Monetary = Sum of (Quantity*Unit Price). (3)

C. Code Description:
Once we are done with the architecture and the equation, we have used in the segmentation of the customers using RFM analysis, we are going to discuss how we implemented the same using the programming. The programming language used in the present task is Python with the version 3.10.12. Google Colab was used to perform the coding for the customer segmentation.

Importing the required libraries and data:
We started the implementation with import and installing the libraries. Let us have a look on the libraries used in the code. First, we have import the os, this module usually provides a way to interact with the operating system. Usually, this module helps in performing the tasks such as manipulation, process management and environment variable access. After that NumPy and Pandas were imported for the mathematical operations and data manipulation. Here the pandas library is used above the NumPy. Matplotlib and Seaborn is used for the purpose of visualization the plots and graphs. Using these libraries, we can create the static interactive and animated visualizations in Python. Sklearn is the abbreviation of Scikit-learn is one of the most popular library used for the machine learning. This tool is used for performing data analysis and modelling using various machine learning algorithms. The we imported the XGBClassifier from and xgboost and CatBoostClassifier from the catboost. These are the libraries that are used for the building a strong prediction models with the combination of typical decision trees and multiple weak models. These two models are used for the customer segmentation for performing the feature engineering, model training, marketing strategies and predictive segmentation. Then the warnings module is used for getting alert about the deprecated features present in the code. This is also used for filtering the warning in the code.

Data cleaning and exploration:
After importing the libraries and modules required for the segmentation, then we performed the data cleaning and explored the data. Here exploring the code meant to getting the statistical information about the data used. In this segment, we obtained the shape of the data. In the shape of the data, we get the count of rows and columns present inside the dataset. After getting the shape of the data we are displaying the top five entries of the dataset using the head() function. Then using the info() we obtain the columns and non-null values present in the dataset along with the datatypes present in the dataset. After using the function, the datatypes present in the dataset are float64, int64. Then the memory used by the dataset is 169.5 KB. Then the describe function is used for obtaining the count, mean, standard deviation, minimum and maximum values present inside the data taken.

RFM Analysis
Now comes the key task that is performing the analysis on the Recency, Frequency and Monetary values based on the formulas discussed earlier. The recency, frequency and monetary values is calculated in python by using the below mentioned formulae:
Recency:
```
recency_df['recency'] = recency_df['RecencyDate'].apply(
    lambda x: (recent_purchase_date - x).days)
```
Frequency:
```
frequency_series = data['column_name'].value_counts()
```
Monetary:
```
df['Total'] = df['UnitPrice'] * df['Quantity']
```

Once the Recency, Frequency and Monetary values are obtained of each customer then the combination of the all the scores of recency, frequency and monetary values takes place. The quantile score calculation of each metric of RFM becomes the next task that must be performed for all the three metrics. Once the quantile score of each metric is being calculated then all three individual scores of the metrics must combined as the separate column. Based on the quantile score of the recency, frequency, and monetary metrics the segmentation of the customer takes place. Here is the reference for the segmenting the customer based on the quantile score of the customers. If all the quantile scores are 444 then the customer is known to be the best customer, because the customer visits the store regularly, the customer also frequently purchases the product at the same time the customer does generate the good revenue at the same time. If the frequency score is 4 for the customer than the customer, then the customer is segmented under the category of Loyal Customer. When the monetary value of the customer is 4 then the customers are segmented under the category of Big spender, as the customer generates the revenue for the company. But if the all the score of value 1, then the customer is segmented under the category of deadbeats. These are segmented as the deadbeats because they do not have a good recency, frequency, and monetary score. The quantiles are calculated in the RFM analysis to determine the time gap between the most recent purchase date and purchase date of the previous transaction.

Data Visualization
Once the code is ready then comes the visualization, the key reason for the visualizing the values is that the images depict the values more clearly and the clarity will be more when

compared to numerical values. Let us now have a glance on the various plots that has been plotted in the analysis.

The 2X2 grid of subplots of histograms were plotted for identifying the recency, frequency, and monetary values. For plotting the 3 various graphs, 3 different colors have been used. The 3 colors used is Red, Blue, and Yellow. The fig.2 gives us the graphical representation of the RFM values through the histograms.

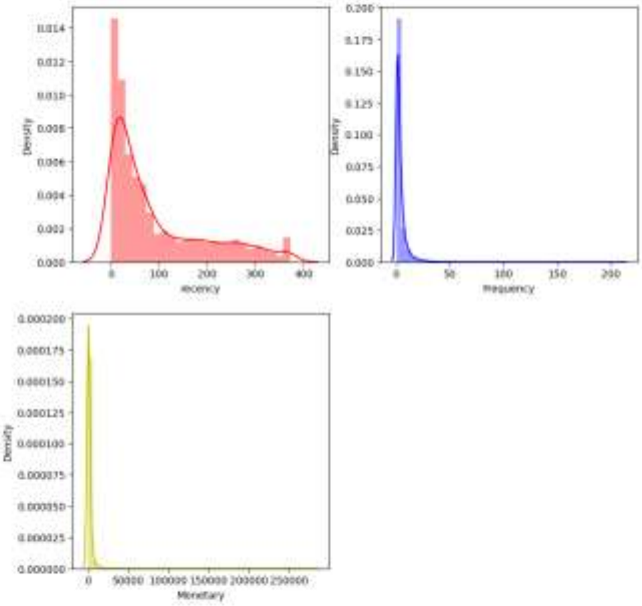


Fig. 3 Graphical representation of RFM values

The next plot used in the analysis is box plot for determining the frequency using the Plotly Express library. Representing the IQR (interquartile range), the bottom and the top edges represent the Q1 and Q3. Here the inside line represents the median (Q2). Fig. 4 depicts the box plot for the frequency value of the RFM analysis.

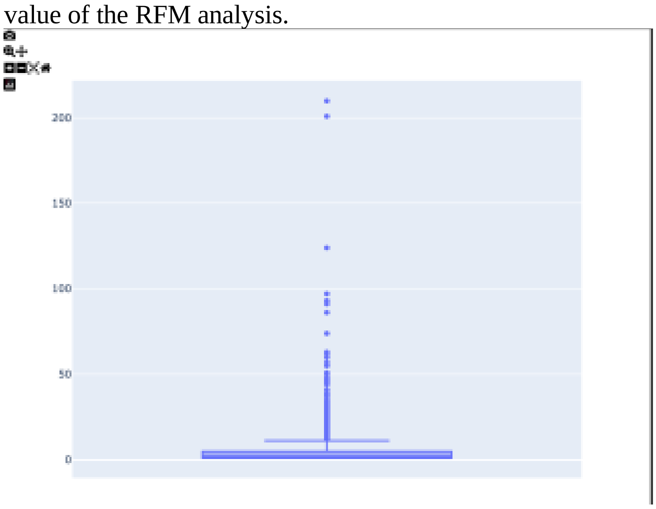


Fig. 4 Frequency values of the RFM analysis

The next box plot is for getting visual summary of the monetary values across the customers. The central line inside the box plot represents the 50^{th} percentile of the Monetary values. The fig. 5 illustrates the monetary values of all the customers.

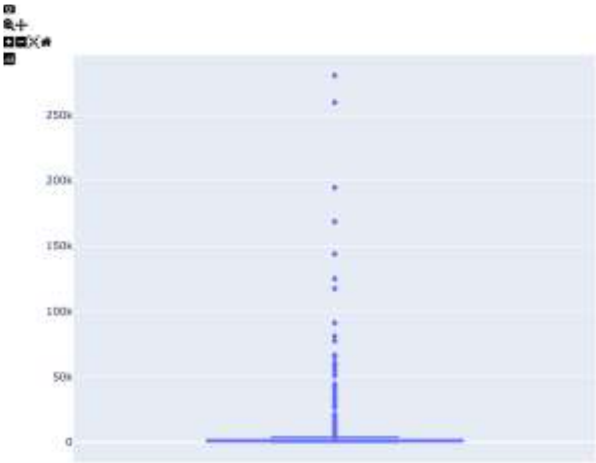


Fig.5 Monetary values of the RFM analysis

The 2X2 grid of histograms with using the Seaborn library for visualizing the quantiles of the Recency, Frequency and Monetary metrics. In this visualization, each x-axis of the subplot is representing the quantile values, and the y-axis gives the frequency of in each quantile. Fig.6 gives the visual representation of the quantiles across the customers data.

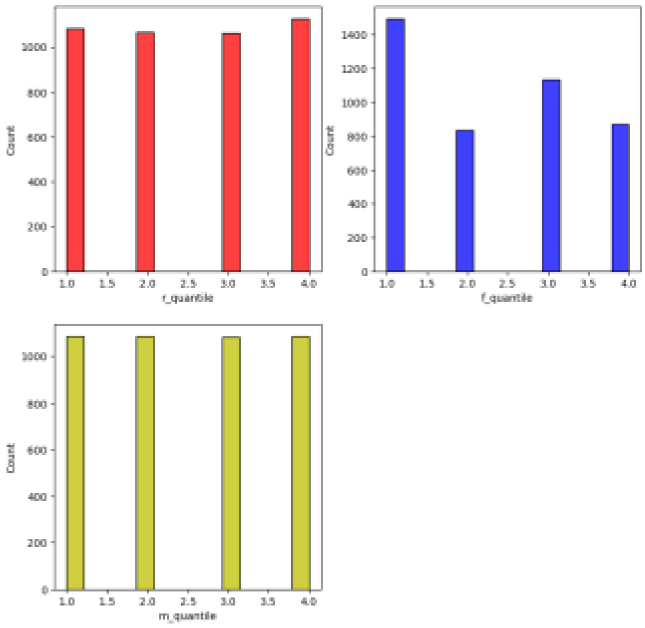


Fig. 7 Quantile values across the dataset

IV. REPORT AND DISCUSSION

At the beginning the understanding of the RFM analysis takes place, where the metrics used in the RFM analysis was known. Then customer segmentation concept was also included in this. Then literature review on the RFM analysis and customer segmentation was done. Various papers were referred from different conferences and journals for knowing the state-of-art in the field of customer segmentation.

Then the coding part was executed with the importing libraries required for performing the RFM analysis for segmentation using the customer segmentation. The libraries used for this were catboost, xgboost for performing the predictions. For the purpose of visualization, the seaborn and plotly were setup, and warnings were set up for ignoring the warnings. Then the data loading was performed, then the exploratory data analysis was performed on the data for getting the clearer idea. Then the data cleaning was done for clearing the missing values in the CustomerID column and null values were dropped from the column. Then the all the values in the Quantity column that are less than 0 were removed.

Then the Recency, Frequency, and Monetary values are calculated. Once the RFM values are obtained then they are combined into the single DataFrame. Then the data visualization is performed for plotting the recency, frequency, and monetary values. The histograms and boxplots were used for plotting these values. Then the calculations of the quantiles are performed and visualized. Once the quantile calculation is done then the same is visualized. In the below mentioned tables, Table 1 describes the recency score, Table 2 describes the frequency score, monetary score has been displayed in the Table 3. Table 4 is where we can find the combined score of recency, frequency, and monetary scores.

TABLE1. RECENCY SCORES

Customer ID	*Recency*
12346	325
12347	1
12348	74
12349	18
12350	309

In the Table.1 we can observe the recency scores of the 5 customers. It is clearly known that the customer with the high recency score is the one who visits the store most frequently. Hence the customer with the CustomerID 12347 has the highest recency and customer with CustomerID 12346 has the least recency rate among the considered 5 customers.

TABLE 2. FREQUENCY SCORES

Customer ID	*Frequency*
12346	1
12347	7
12348	4
12349	1
12350	1

In the Table.2 we can observe the frequency scores of the 5 customers. It is clearly known that the customer with the high frequency score is the one who most unique invoices among the other customers. Hence the customer with the CustomerID 12347 has the highest frequency and customer with CustomerID 12346, 12349, 12350 has the least frequency rate among the considered 5 customers.

Table 3. Monetary Scores

Customer ID	*Monetary*
12346	77183
12347	1
12348	74
12349	18
12350	309

In the Table.3 we can observe the monetary values of the 5 customers. It is clearly known that the customer with the high Monetary values is the one who spends the most among the other customers. Hence the customer with the CustomerID 12346 has the highest Monetary score and customer with CustomerID 12347 has the least Monetary rate among the considered 5 customers.

TABLE 4. RFM SCORES

Customer ID	*Recency*	*Frequency*	*Monetary*
12346	325	1	77183
12347	1	7	1
12348	74	4	74
12349	18	1	18
12350	309	1	309

In the Table 4 all the values of recency, frequency and monetary values are combined into the single table for better understanding about the values obtained for the same metrics.

The quantile is used of segmenting the customers into four segments based their recency, frequency, and monetary values. Once the quantile scores are obtained then the customer segmentation is performed. And the customers are categorized as best customers, loyal customers, big Spenders, deadbeats, and lost Customers. Table 5 and Table 6 discuss about the final RFM quantile and Final segmentation of the customers.

TABLE 5. RFM QUANTILES

Customer ID	*RFM_ Quantile*
12346	114
12347	444
12348	234
12349	314
12350	112

From Table 5 we get the RFM Quantile scores of the considered 5 customers. Based on these quantile values the segmentation of customers as the mentioned as per the provided categories. Usually, these quantiles are used for engagment of the various customer group. For the low recency, customers some of the targeting the customers with the low recency rate. In the Table 6 the customers are classified based the quantile score obtained in Table 5. Here basically the customers are segmented in to 4 categories. If the RFM score is 444 then the customer is treated as the Best Customer due to their high recency, frequency, and the monetary score. The customer with the quantile score as 144 is known to be the loyal customers since their monetary value and frequency is very high. The customers who have the high monetary values and even if their recency and frequency is not that high then they are segmented as the big spenders.

TABLE 5. COMBINED RFM SCORE AND SEGMENTATION

Customer ID	*Recency*	*Frequency*	*Monetary*	*RFM Quantiles*	*Customer Segmentation*
12346	325	1	77183	114	Big Spender
12347	1	7	1	444	Best Customer
12348	74	4	74	234	Big spender
12349	18	1	18	314	Big Spender
12350	309	1	309	112	None

In the Fig. 8 the histogram has been plotted using the seaborn library function in python for the visualizing the customers distribution across the various Recency, Frequency quantiles. The visualization performed here will be useful in the distribution of the customers at different segments, helping businesses in understanding their composition of their customer base.

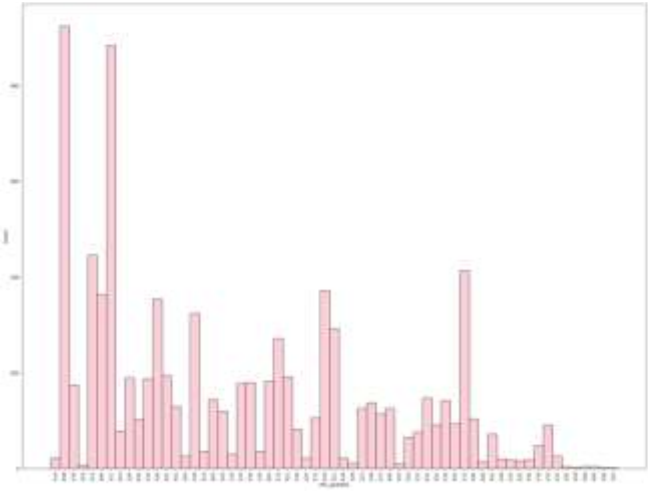
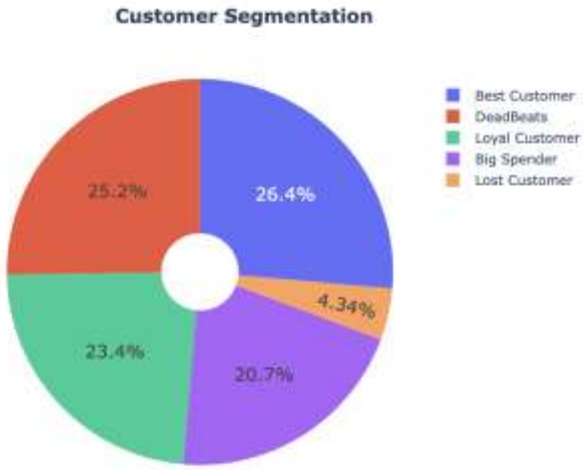


Fig.8 Customers distributions across RFM quantiles

The pie chart is generated using the Plotly for visualizing the customer distributions that are obtained from the RFM Analysis. This figure gives the customer segmentations across the segments obtained through the RFM analysis. In the below mentioned pie chart we can see that the Best Customer are of 26.4% with the count of 462. Then the deadbeats are at the 25.2% with the count of 442. Then the loyal customers are with 23.4% with the total count of 410. Then comes the best customer who covers the 20.7% the count is 373. Then comes layer with the selection of lost customer with the 4.34 % with the count of 76. The below mentioned Table 7 is the count of segmented customers across the various categories.

TABLE 5. SEGMENTED

Segmentation	*Count of customer*
Best Customer	462
Dead beats	443
Loyal Customers	410
Big Spender	314



V. CONCLUSION AND FUTURE ENHANCEMENT

In the conclusion of the RFM analysis, the analysis performed on the Superstore store dataset for the segmenting the customers of the superstore into the various categories based on the recency, frequency, and monetary values of the individual customer. The customer segmentation is performed for understanding the sales patterns and the other aspect of the business. Once the segmentation is over then the results are used for making the new strategies.

Regarding the future enhancement of the present customer segmentation using RFM analysis is to get the count and percentage of people, who are loyal, best, and deadbeat for the company. With the inclusion of certain features here are some of the suggestions. Interactive dashboards can be implemented with the tools like plotly Dash in python or Streamlit. This will be very useful for getting the customer segments interactively and insights will be interactive. The data loading process can be automated, so that the data obtained from database or an external API. The present insights can be integrated with CRM systems in which actionable insights will be provided along with the customer relationship strategies.

REFERENCES

1. *N. Gankidi, S. Gundu, M. v. Ahmed, T. Tanzeela, C. R. Prasad and S. Yalabaka, "Customer Segmentation Using Machine Learning," 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-5, doi: 10.1109/CONIT55038.2022.9848389.*
2. *M. Liu, Y. Du and X. Xu, "Customer Value Analysis Based on Python Crawler," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 4345-4350, doi: 10.1109/CCDC.2019.8832805..*
3. *S. Wan, J. Deng, W. Gan, J. Chen and P. S. Yu, "Fast Mining RFM Patterns for Behavioral Analytics," 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA), Shenzhen, China, 2022, pp. 1-10, doi: 10.1109/DSAA54385.2022.10032434..*
4. *J. B. Mavinakunte Nagesh, "Generating Political Scores Using RFM Model and Cluster Prediction by XGBoost," 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2019, pp. 1333-1336, doi: 10.1109/CSCI49370.2019.00249..*

5. Y. Yang and M. O. Shafiq, "Identifying High Value Users in Twitter Based on Text Mining Approaches," 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), Rotorua, New Zealand, 2019, pp. 634-641, doi: 10.1109/TrustCom/BigDataSE.2019.00090..

A. Parkhan and M. R. A. Purnomo, "Industrial Marketing Strategy Formulation Based on RFM-Data Analytics and Fuzzy Database," 2023 International Conference On Cyber Management And Engineering (CyMaEn), Bangkok, Thailand, 2023, pp. 463-468, doi: 10.1109/CyMaEn57228.2023.10051008..

A. M. P. Fernando, A. M. T. T. Adhikari, W. H. A. T. K. Wijesekara, T. V. T. I. Vithanage, A. Gamage and T. Jayalath, "Planning Marketing Strategies in Small-Scale Business Using Data Analysis," 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 2022, pp. 1-6, doi: 10.1109/IISEC56263.2022.9998200.

6. O. Syrotkina, S. Bhatta and K. P. Jacob, "Recency-Frequency-Monetary Analysis and Recommendation System using Apriori Algorithm on E-Commerce Sales Data," 2023 13th International Conference on Advanced Computer Information Technologies (ACIT), Wrocław, Poland, 2023, pp. 321-326, doi: 10.1109/ACIT58437.2023.10275513.

7. J. Sun and Z. Meng, "Research on Customer Value Identification of Video-on-demand services Based on RFM Improved Model," 2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Nanjing, China, 2022, pp. 464-467, doi: 10.1109/CEI57409.2022.9950144.

8. T. Lim, "RFM and Classification Predictive Modelling to Improve Response Prediction Rate," 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), Novi Sad, Serbia, 2020, pp. 333-337, doi: 10.1109/ZINC50678.2020.9161800.

9. S. H. Shihab, S. Afroge and S. Z. Mishu, "RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-4, doi: 10.1109/ECACE.2019.8679376.

10. J. Tri Wibowo and M. Suryanegara, "Segmenting the Subscribers of An Indonesian 4G Service Operator Using RFM Method," 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Purwokerto, Indonesia, 2021, pp. 280-283, doi: 10.1109/COMNETSAT53002.2021.9530772.

11. Y. Feng, X. Wang and L. Li, "The Application Research of Customer Segmentation Model in Bank Financial Marketing," 2019 2nd International Conference on Safety Produce Informatization (IICSPI), Chongqing, China, 2019, pp. 564-569, doi: 10.1109/IICSPI48186.2019.9095900.