

CUSTOMER SEGMENTATION BASED ON BUYING BEHAVIOUR

SHUVADEEP MAITY, TANUKA CHATTARAJ, RIDDHIMAN SYED

Contents

1. Abstract
2. Introduction
3. Problem statement
4. Objective
5. Data Overview
6. Methods
7. Implementation
8. Outcomes
9. Conclusion

Abstract - Customer segmentation through the RFM model offers an in-depth insight into customer purchasing patterns. This segmentation method categorizes customers into specific groups based on recency, frequency, and monetary value, enabling retailers to customize their marketing strategies and elevate customer satisfaction. The ultimate result is improved customer retention and loyalty.

Customer segmentation via RFM analysis stands as a widely embraced approach for customer stratification and targeting. RFM, denoting Recency, Frequency, and Monetary Value, comprises three pivotal metrics employed to classify customers according to their transactional behaviours.

Introduction - In the ever-evolving landscape of business and marketing, understanding and effectively engaging with customers has become more crucial than ever before. One powerful strategy that has gained significant traction is Customer Segmentation based on RFM analysis. RFM, which stands for Recency, Frequency, and Monetary, offers a comprehensive framework for dissecting a customer database to identify distinct customer segments with unique behaviours and preferences. By harnessing the power of RFM analysis, businesses can tailor their marketing efforts and strategies to target diverse customer segments more effectively. This approach not only enhances customer engagement but also optimizes resource allocation, leading to improved customer satisfaction, loyalty, and ultimately, sustainable growth. In this exploration, we will delve into the concept of Customer Segmentation through RFM analysis, uncovering its key principles and practical applications in the quest to better understand and serve our valued customers.

Problem statement - In the retail industry, there exists a significant challenge in identifying the primary revenue drivers and evaluating their respective significance in influencing the top-line. This lack of comprehensive insight into the factors impacting revenue generation poses a potential hindrance to effective strategic planning and well-informed decision-making within the retail landscape. To thrive in this competitive environment, it is crucial for retailers to successfully understand and cater to diverse customer segments, fostering tailored interactions and building lasting customer loyalty.

However, conventional methods of customer segmentation, including geographic, demographic, and psychographic approaches, often prove inadequate in achieving this objective. These methods carry the risk

of overlooking latent high-value customer segments and may struggle to develop precise, targeted strategies aimed at transitioning customers into more financially rewarding segments.

Objective - The primary objective of this study is customer segmentation using the RFM (Recency, Frequency, and Monetary) model in improving marketing strategies for retailers. Specifically, this research aims to:

- ✓ Categories the customers based on their transactional behaviour.
- ✓ Analysis customize RFM model to created 16 segments of customer.
- ✓ Segregated the customers in active and inactive by their purchase recency.

Data Overview - The data that has been used to demonstrate is 15 months of transactional data of a particular retail store in the United States. The data has been anonymized to protect the privacy of the customers.

Columns	Description	Type	Type/Units	
CUST_ID	An unique multi digit number identifying each customer	Nominal	Identifier	
VISIT_ID	An unique multi digit number identifying each purchase by the customer (Invoice Number)	Nominal	Identifier	
VISIT_DT	The date of the transaction by the customer	Date	Date	
VISIT_TM	The time of the transaction by the customer	Time	Time	
SALES	The value of the product purchased in \$	Continuous	US Dollars(\$)	
VOLUME	The quantity of the product in units	Discrete Count	Units	

Methods - Customer segmentation using RFM analysis is a widely employed technique in order to segregate and target customers. RFM stands for Recency, Frequency, and Monetary Value, which are three essential metrics used to categorize customers based on their transactional behaviour. Each of these metrics provides valuable insights into a customer's engagement and value to the business. Before RFM is implemented, the customers are segregated according to the frequency of their visits.

Step 1: The 'VISIT_DT' variable was split into ranges of weeks, months, quarters & years.

For weeks, we got 65 'VISIT_DT' variables(dimensions).

For months, we got 15 'VISIT_DT' variables(dimensions).

For quarters, we got 5 'VISIT_DT' variables(dimensions).

For years, we got 2 'VISIT_DT' variables(dimensions).

A pivot table was generated for all the customers(rows) and the splits, with the count of the visits for each customer as values in the pivot. Any values before the first visit of the customer were replaced with Nan.

	2022-01	2022-02	2022-03	2022-04	2022-05	2022-06	2022-07	2022-08	2022-09	2022-10	2022-11	2022-12	2023-01	2023-02	2023-03
CUST_ID															
25005600	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	0.0	0.0	0
25005623	NaN	NaN	NaN	NaN	NaN	2.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
25005686	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0
25005692	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0
25005707	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	0.0	0.0	0.0	0.0	0

Then the 25 percentile values were calculated for the count of visits for each customer. This translates to the fact that, the customer visited more than the Q1 value(25 percentile) [if the Q1 value is 2 for weekly data for a customer, that means 75% of the weeks counted for the given case he has come more than twice in a week] for 75% of the time period(either weekly, daily, monthly, or quarterly). This threshold helps in the segregation.

Step 2: The customers were then segregated into the below mentioned categories based on their frequency of visits:

- **Daily:** Customer who have visited more than once a week, are segregated as daily customers. The left-over customers are then selected for weekly calculation.
- **Weekly:** Customer who have visited at least twice or more in a month, are segregated as weekly customers. The left-over customers are then selected for monthly calculation.
- **Monthly:** Customers who have visited at least twice or more in a quarter, are segregated as monthly customers. The left-over customers are then selected for quarterly calculation.
- **Quarterly:** Customers who have visited more than thrice or more in a year, are segregated. The left-over customers are unsegmented customer due to the infrequency of their visits.

	CUST_ID	Perc_25
7	25005746	2.0
16	25005909	3.5
27	25006296	6.0
49	25006517	3.5
62	25006641	2.5
...
13954	364506563	2.0
13959	364506580	4.0
13989	364506677	3.0
14014	364506755	3.0
14023	364506785	4.0

1815 rows × 2 columns

Step 3: The RFM analysis was then performed

Recency (R): Recency refers to the time elapsed since a customer's last purchase (Number of inactive days since last purchase). Customers who have made recent purchases are considered more engaged and likely to make repeat purchases in the near future. The figure is always in days.

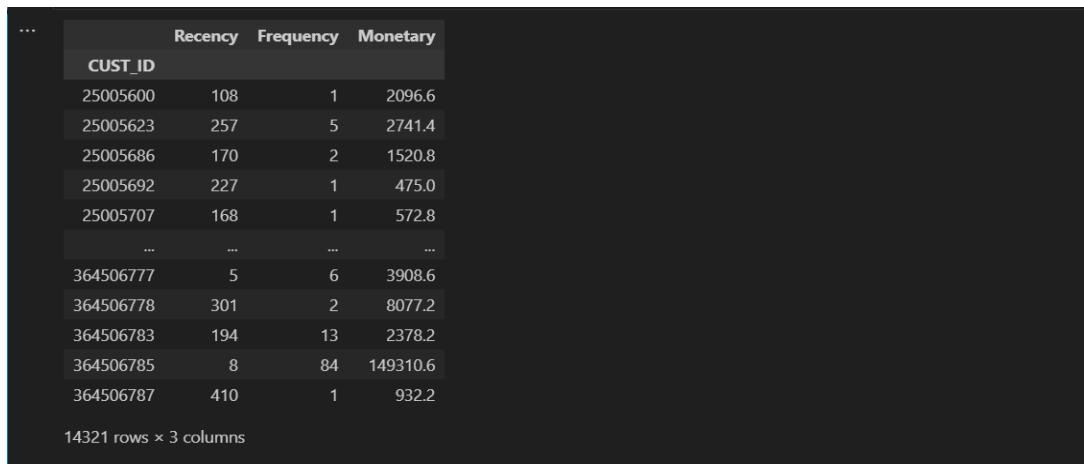
For example, if the last purchase of the customer was on 1st of December, and we are calculating on 31st December, then the recency in days would be 30 days.

Frequency (F): Frequency denotes the number of distinct days a customer made a purchase within a specific time period. Customers with a higher frequency of purchases are often more loyal and hold significant potential for upselling and cross-selling.

For example, if the customer has made 5 purchases days in a span of 1 year, then the frequency would be 5 for a particular time period, and if the customer purchases 2 time or more in a single day its also be count as a 1 single purchase day.

Monetary Value (M): Monetary Value represents the total amount of money a customer has spent on purchases. Customers with higher monetary value are valuable to the business as they contribute more significantly to the overall revenue.

For example, if the customer has spent \$1000 in a span of 1 year, then the monetary value would be \$1000.



The screenshot displays a data table with the following structure:

	Recency	Frequency	Monetary
CUST_ID			
25005600	108	1	2096.6
25005623	257	5	2741.4
25005686	170	2	1520.8
25005692	227	1	475.0
25005707	168	1	572.8
...
364506777	5	6	3908.6
364506778	301	2	8077.2
364506783	194	13	2378.2
364506785	8	84	149310.6
364506787	410	1	932.2

14321 rows × 3 columns

Once these metrics were obtained, we segregated the customers in active and inactive customers, in the following manner:

Daily Active Customers: Daily Customers who have made a purchase in the last 30 days are considered active customers. Left over customers are classified inactive.

Weekly Active Customers: Weekly Customers who have made a purchase in the last 60 days (8 weeks) are considered active customers. Left over customers are classified inactive.

Monthly Active Customers: Monthly Customers who have made a purchase in the last 90 days (3 months) are considered active customers. Left over customers are classified inactive.

Quarterly Active Customers: Quarterly Customers who have made a purchase in the last 180 days (2 quarters) are considered active customers. Left over customers are classified inactive.

By combining these metrics with segregated active customers, RFM analysis creates different segments of customers with distinct characteristics and behaviours as mentioned below, which can be used to tailor marketing strategies and drive customer engagement.

Step 4: For each of the active customer segments obtained above two independent metrics were calculated:

Median of Frequency (F): The median of the frequency of the customers in that segment.

Median of Monetary Value (M): The median of the monetary value of the customers in that segment.

After which each segment was further divided into 4 sub segments based on the median of the frequency and monetary value and a set of conditions:

High-Frequency High-Value (HFHM): - These are the most valuable customers who make frequent purchases and spend more money than the median monetary values.

High-Frequency Low-Value (HFLM): - These are customers who make more frequent purchases than the median but spend less money than the median monetary value.

Low-Frequency High-Value (LFHM): - These are customers who make fewer purchases but spend more money than the median monetary value.

Low-Frequency Low-Value (LFLM): - These are customers who make fewer purchases and spend less money than the median values.

HFHM	HFLM
Median Recency 5.0 Median Frequency 115.0 Median Monetary 97962.2 count 655.0 Name: HFHM, dtype: float64	Median Recency 5.0 Median Frequency 90.0 Median Monetary 41045.0 count 234.0 Name: HFLM, dtype: float64
LFHM	LFLM
Median Recency 7.0 Median Frequency 57.0 Median Monetary 77594.7 count 248.0 Name: LFHM, dtype: float64	Median Recency 8.0 Median Frequency 50.0 Median Monetary 29604.2 count 669.0 Name: LFLM, dtype: float64

By understanding the distinct needs and behaviours of each segment, retailers can optimize customer retention, increase customer lifetime value, and drive sustainable growth in the highly competitive retail landscape.

Implementation –

1. Installation-

Following code is to install package.

pip install businessmodels

```
pip install businessmodels

Collecting businessmodels
  Downloading businessmodels-0.0.7-py3-none-any.whl (6.4 kB)
Collecting business-models-initial (from businessmodels)
  Downloading business_models_initial-0.0.2-py3-none-any.whl (4.0 kB)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from business-models-initial->businessmodels) (1.5.3)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from business-models-initial->businessmodels) (1.11.3)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->business-models-initial->businessmodels) (2.8.1)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->business-models-initial->businessmodels) (2020.1)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas->business-models-initial->businessmodels) (1.21.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas->business-models-initial->businessmodels) (1.16.0)
Installing collected packages: business-models-initial, businessmodels
Successfully installed business-models-initial-0.0.2 businessmodels-0.0.7
```

Import the 'businessmodels' library and its 'customersegmentation' module

```
from businessmodels import customersegmentation
```

Import the 'pandas' library as 'pd'

```
import pandas as pd
```

```
from businessmodels import customersegmentation
import pandas as pd
```

2. Data Load-

Read a CSV file from Google Drive or any type of storage it in the 'df' DataFrame

```
df = pd.read_csv('/content/drive/MyDrive/store_1586_q1_2022_q1_2023.csv')
```

```
# Display the first few rows of the DataFrame
```

```
df.head()
```

3.Function Call -

```
# Perform customer segmentation using the 'businessmodels' library
```

```
# 'customersegmentation.cust_seg' is a function in the library
```

```
# It takes the 'df' DataFrame and a date (Ex.- '2023-04-02') as arguments
```

```
Segments = customersegmentation.cust_seg(df, '2023-04-02')
```

```
# Extract all segmentation results into variables 'customer_rfm', 'active_customer_segments', 'inactive_customers', and 'segments_summary'.
```

For Daily Customers,

```
Customer_rfm,active_customer_segments, inactive_customers, segments_summary = Segments.Daily()
```

```
# Print or use the resulting data as needed
```

```
print("\nCustomer RFM:")
```

```
print(Customer_rfm)
```

```
print("\nCustomer Segments:")
```

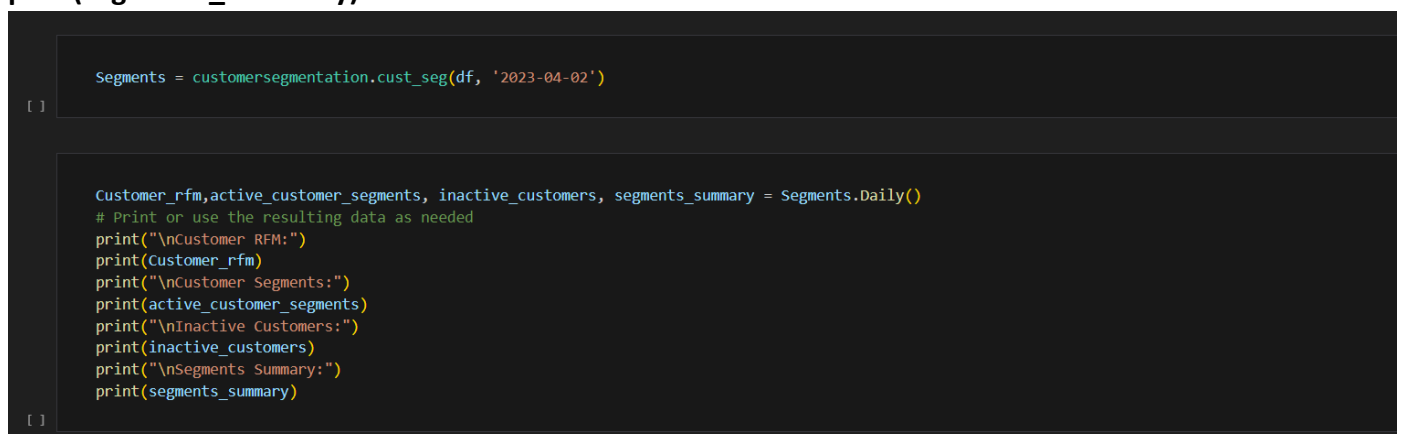
```
print(active_customer_segments)
```

```
print("\nInactive Customers:")
```

```
print(inactive_customers)
```

```
print("\nSegments Summary:")
```

```
print(segments_summary)
```



```
[ ]

Segments = customersegmentation.cust_seg(df, '2023-04-02')

[ ]

Customer_rfm,active_customer_segments, inactive_customers, segments_summary = Segments.Daily()
# Print or use the resulting data as needed
print("\nCustomer RFM:")
print(Customer_rfm)
print("\nCustomer Segments:")
print(active_customer_segments)
print("\nInactive Customers:")
print(inactive_customers)
print("\nSegments Summary:")
print(segments_summary)

[ ]
```

For Weekly Customers,

```
Customer_rfm ,active_customer_segments, inactive_customers, segments_summary = Segments.  
Weekly ()
```

```
# Print or use the resulting data as needed
```

```

print("\nCustomer RFM:")
print(Customer_rfm)
print("\nCustomer Segments:")
print(active_customer_segments)
print("\nInactive Customers:")
print(inactive_customers)
print("\nSegments Summary:")
print(segments_summary)

For Monthly Customers,

Customer_rfm ,active_customer_segments, inactive_customers, segments_summary = Segments.
Monthly ()

# Print or use the resulting data as needed
print("\nCustomer RFM:")
print(Customer_rfm)
print("\nCustomer Segments:")
print(active_customer_segments)
print("\nInactive Customers:")
print(inactive_customers)
print("\nSegments Summary:")
print(segments_summary)

For Quarterly Customers,

Customer_rfm ,active_customer_segments, inactive_customers, segments_summary = Segments.
Quarterly ()

# Print or use the resulting data as needed
print("\nCustomer RFM:")
print(Customer_rfm)
print("\nCustomer Segments:")
print(active_customer_segments)
print("\nInactive Customers:")
print(inactive_customers)
print("\nSegments Summary:")
print(segments_summary)

```

Outcomes - The customers were segmented according to their time frequency of visit into daily, weekly, monthly and quarterly customers. The RFM analysis was performed on each of these segments resulting in the following active customer segments for our 14321 customers;

Segment	Active Count	Segment Total	Percentage Count (Active)	Percentage (Total - Segmented)	Percentage (Total Customers)
Daily Active Customers	295	296	99.7%	5.7%	2.0%
Weekly Active Customers	1806	1815	99.5%	34.9%	12.7%
Monthly Active Customers	2222	2370	93.8%	42.9%	15.5%
Quarterly Active Customers	847	957	88.5%	16.4%	5.9%
---	---	---	---		
Total Segmented	5170	14321	36.1%		
Unsegmented Customers	8883	14321	62.0%		

- This shows that about 36% of customers are actually active, with the biggest chunk of active customers being the monthly customers at 43% followed by the weekly customers at 35%.
- The daily and quarterly customers are the least active at 6% and 16% respectively.
- 62% of the customers are unsegmented.

Daily Customers Summary :

Customer_segments	Median Recency	Median Frequency	Median Monetary \$	count
HFHM	2.0	359.0	2,68,5823.4	89
HFLM	2.0	321.0	1,25,026.6	51
LFHM	3.0	208.0	2,19,582.2	51
LFLM	3.0	204.0	1,21,226.0	90

Weekly Customer Summary :

Customer_segments	Median Recency	Median Frequency	Median Monetary	count
HFHM	5.0	115.0	97,980.8	655
HFLM	5.0	90.0	41,156.0	234
LFHM	7.0	57.0	77,882.5	248
LFLM	8.0	50.0	29,583.2	669

Monthly Customer Summary :

Customer_segments	Median Recency	Median Frequency	Median Monetary	count
HFHM	13.0	34.0	28,742.6	831
HFLM	14.0	28.0	11,715.6	263
LFHM	19.0	18.0	22,133.0	280
LFLM	23.0	13.0	6,449.3	848

Quaterly Customer Summary :

Customer_segments	Median Recency	Median Frequency	Median Monetary	count
HFHM	27.5	18.0	15,876.9	256
HFLM	26.0	16.0	5,852.6	105
LFHM	23.0	11.0	12,236.6	167
LFLM	23.0	10.0	4,774.6	319

Conclusion - The customer segmentation approach showed us that the most valuable customers are those who are daily in nature, which is contrary to popular belief. This insight was used to suggest that the business should focus on retaining and nurturing the all-segments customers through promotions.