

RFM ANALYSIS IN R

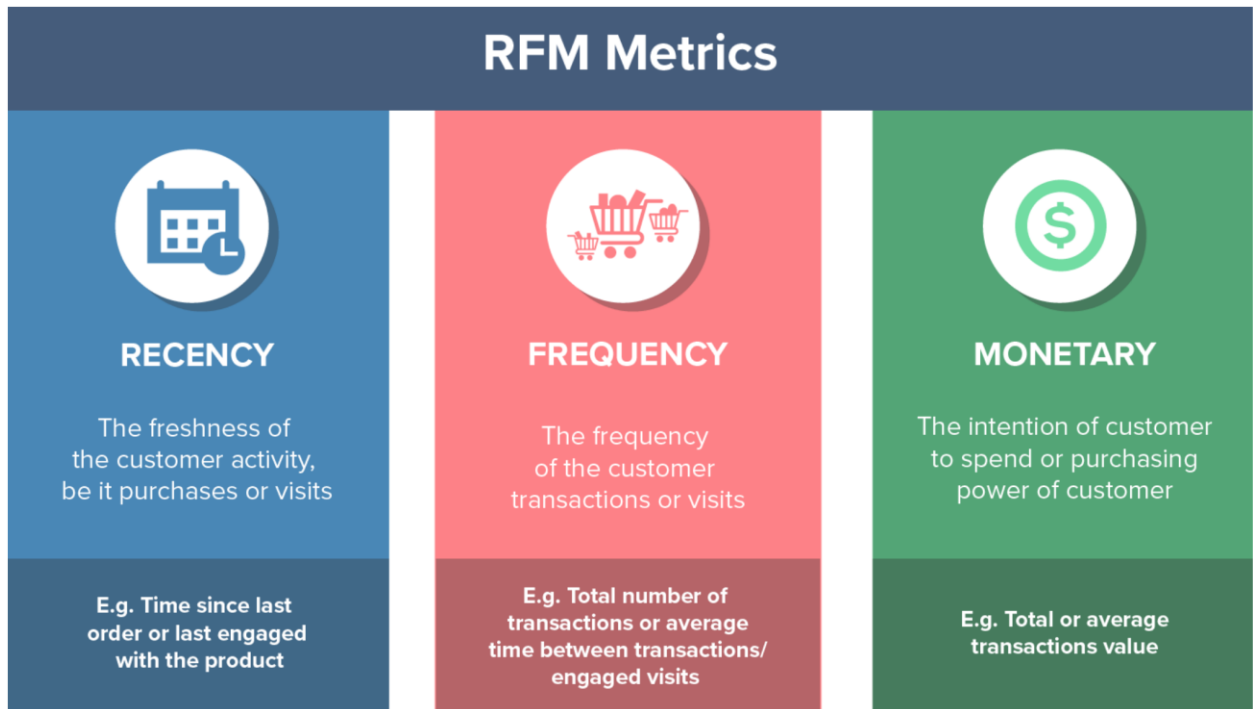
Recency, Frequency, and Monetary



PRIYANK MEHAR

1.RFM ANALYSIS

RFM stands for Recency, Frequency, and Monetary value, each corresponding to some key customer trait. These RFM metrics are important indicators of a customer's behavior because frequency and monetary value affects a customer's lifetime value, and recency affects retention, a measure of engagement.



1.1 RFM Table

Key element of the RFM dataset

```
> head(RFM_SRC %>% select(ClientID, NoofVisit, InActivePeriod, TotalPurchase))
  ClientID NoofVisit InActivePeriod TotalPurchase
1    55010         5          115         837.10
2     8803         5           39        1838.70
3    73411         7           85         679.80
4    21980        16          200        3735.97
5    53497         5           39         675.32
6    26226         7           39        2853.09
```

1.2 RFM Score

For computing RFM score we have to go with the below steps which explain the process:

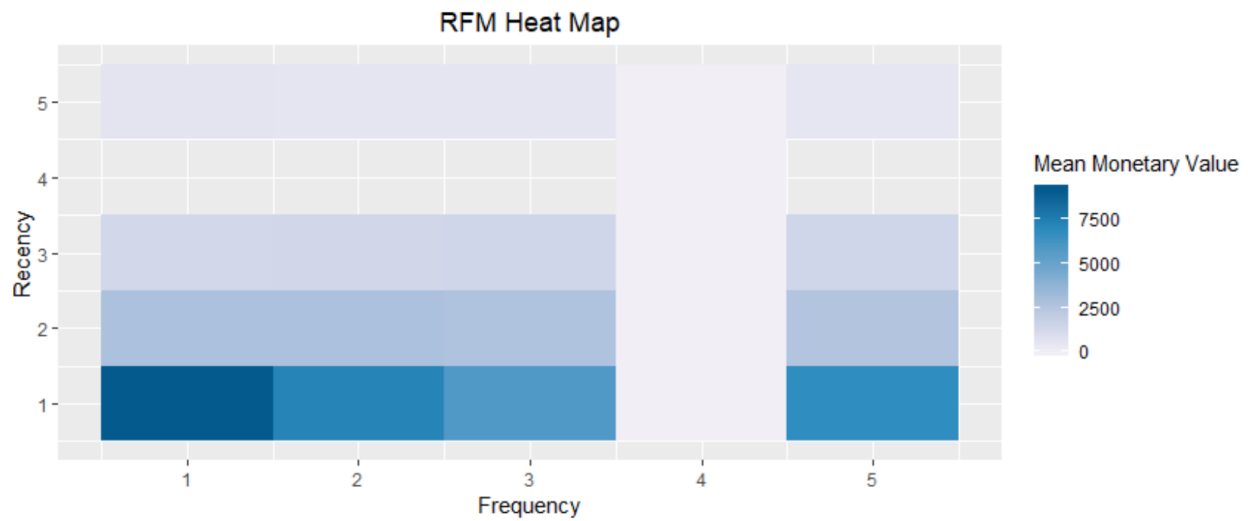
- A recency score is assigned to each customer based on date of most recent purchase. The score is generated by binning the recency values into a number of categories (default is 5). For example, if you use four categories, the customers with the most recent purchase dates receive a recency ranking of 4, and those with purchase dates in the distant past receive a recency ranking of 1.
- A frequency ranking is assigned in a similar way. Customers with high purchase frequency are assigned a higher score (4 or 5) and those with lowest frequency are assigned a score 1.
- Monetary score is assigned on the basis of the total revenue generated by the customer in the period under consideration for the analysis. Customers with highest revenue/order amount are assigned a higher score while those with lowest revenue are assigned a score of 1.
- A fourth score, RFM score is generated which is simply the three individual scores concatenated into a single value.

In our code we are using “ rfm ” library to calculate the score. And we have got the score for our dataset as follows:

	customer_id	recency_days	transaction_count	amount	recency_score	frequency_score	monetary_score	rfm_score
1	55010	5	115	837.10	2	2	3	223
2	8803	5	39	1838.70	2	1	4	214
3	73411	7	85	679.80	1	2	3	123
4	21980	16	200	3735.97	1	3	5	135
5	53497	5	39	675.32	2	1	3	213
6	26226	7	39	2853.09	1	1	4	114

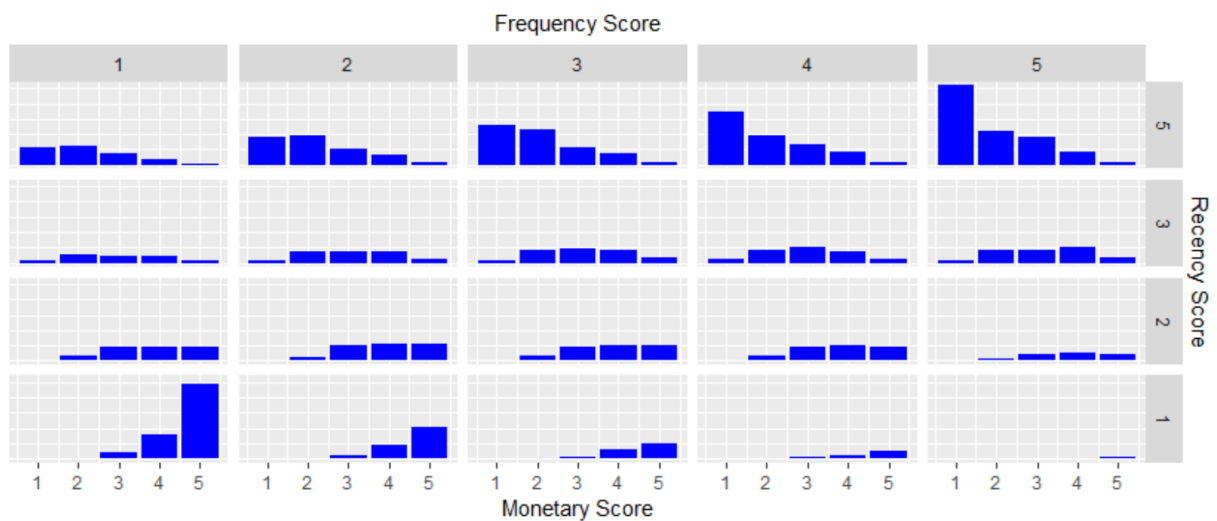
1.3 RFM Heat Map

The heat map shows the average monetary value for different categories of recency and frequency scores. Higher scores of frequency and recency are characterized by higher average monetary value as indicated by the darker areas in the heatmap.



1.4 RFM Bar Chart

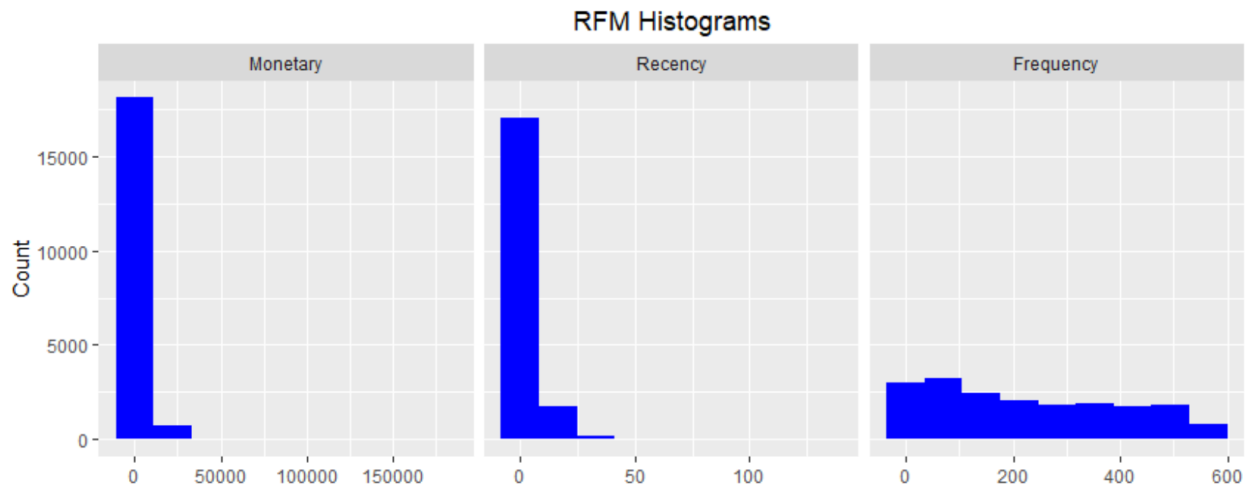
Below bar chart shows the distribution of monetary scores for the different combinations of frequency and recency scores.



1.5 RFM Histogram

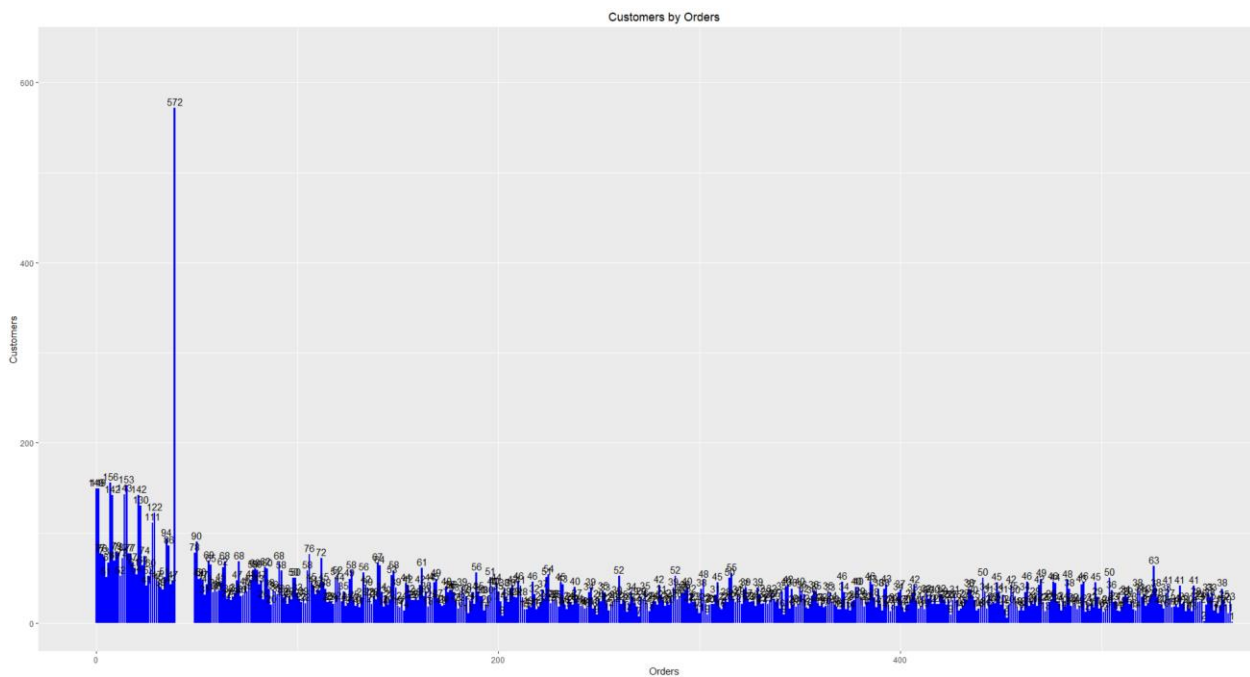
Histogram helps us here to examine the relative distribution of -

- monetary value (total revenue generated by each customer)
- recency days (days since the most recent visit for each customer)
- frequency (transaction count for each customer)



1.6 RFM Customers by Orders

This helps us to visualize the distribution of customers across orders.

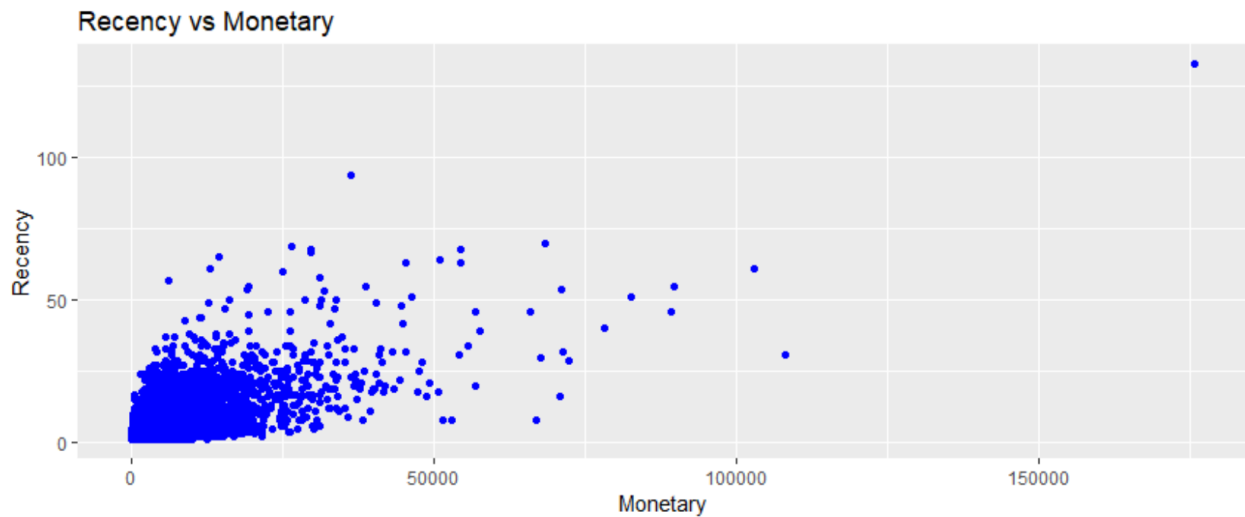


1.7 Scatter Plots

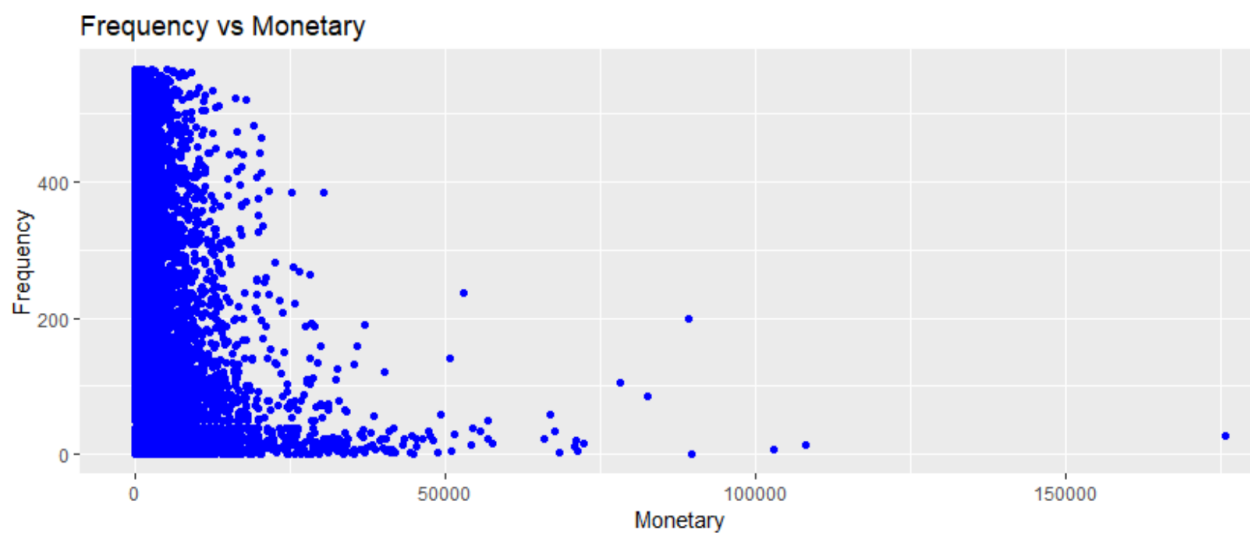
The best client are those who:

- bought most recently
- most often
- and spend the most

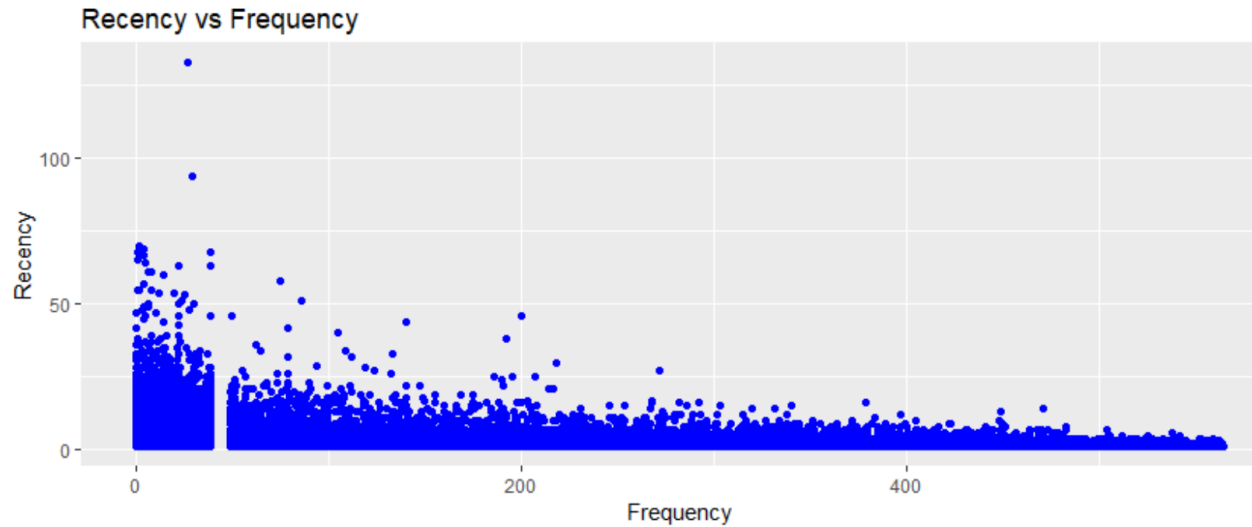
So let's examine various relationship between Recency, Frequency and Monetary:



Client who visited more recently generated more revenue compared to those who visited in the distant past. The clients who visited in the recent past are more likely to return compared to those who visited long time ago as most of those would be lost clients.



As the frequency of visits increases, not necessarily the revenue also increases. Clients who visit more frequently are your champion, loyal or potential most probably drive higher revenue.



Clients with low frequency visited in the distant past while those with high frequency have visited in the recent past. Again, the clients who visited in the recent past are more likely to return compared to those who visited long time ago

1.8 Segmented Customer Data

Let us classify our clients based on the individual recency, frequency and monetary scores. We can use the segmented data to identify

- best customers
- loyal customers
- at risk customers
- and lost customers

In general below score card helps to identify customer score

Segment	Description	Classification	R	F	M
Champions	Bought recently, buy often and spend the most	Top	4 - 5	4 - 5	4 - 5
Loyal Customers	Spend good money. Responsive to promotions	Loyal	2 - 5	3 - 5	3 - 5
Potential Loyalist	Recent customers, spent good amount, bought more than once	Potential	3 - 5	1 - 3	1 - 3
New Customers	Bought more recently, but not often		4 - 5	<= 1	<= 1
Promising	Recent shoppers, but haven't spent much		3 - 4	<= 1	<= 1
Need Attention	Above average recency, frequency & monetary values	Need Attention	2 - 3	2 - 3	2 - 3
About To Sleep	Below average recency, frequency & monetary values		2 - 3	<= 2	<= 2
At Risk	Spent big money, purchased often but long time ago	High Spent	<= 2	2 - 5	2 - 5
Can't Lose Them	Made big purchases and often, but long time ago		<= 1	4 - 5	4 - 5
Hibernating	Low spenders, low frequency, purchased long time ago	Lost	1 - 2	1 - 2	1 - 2
Lost	Lowest recency, frequency & monetary scores		<= 2	<= 2	<= 2

Once we have classified a customer into a particular segment, we can take appropriate action to increase his/her lifetime value.

So we have created a dataframe correspond to the above table to run our analysis on top of the data we have and segment customers based on that.

1.9 Segment Size

The distribution of customers across the segments. Ideally, we should have very few or no customer in segments such as At Risk or Needs Attention.

	Segment	Count
	<fct>	<int>
1	Loyal	3506
2	Potential	3200
3	Need Attention	1380
4	High Spent	925
5	Top	511
6	Lost	96

But the analysis shows that we have quite considerable number of customer in the “Need Attention” segment, by various promotion offers/ discount and other schemes.

1.10 RFM Median

In this section will analyze the median recency, frequency and monetary value across segments to ensure that the logic used for customer classification is sound and practical.

