# Identification of Customer Clusters using RFM Model: A Case of Diverse Purchaser Classification

Publishing India Group

*International Journal of Business Analytics and Intelligence*

Article can be accessed online at http://www.publishingindia.com

# Identification of Customer Clusters using RFM Model: A Case of Diverse Purchaser Classification

**Riktesh Srivastava***

**Abstract**

Competitive world today stresses of having virtuous marketing strategies to appeal new customers while holding longstanding customers. Organisations use instruments to embrace both types of customers, thereby, probing better return on investments and ensuing increasing revenues. The notion of "customer clustering" is used by organisations to categorise diverse fragments of customers and offer them with varied services. The present study takes the four fragments of customers, viz., active, warm, cold, and inactive and does added exploration of these fragments. It was found that these fragments are not enough for defining marketing strategies and need further analysis. The paper magnifies the fragment using RFM analysis then performing clustering on the values obtained from this analysis. This analysis spawns the pertinent rules for each customer segment obtained after clustering.

**JEL Codes:** G31, G32

**Keywords:** RFM, Customer Value Pyramid (CVP), Customer Clusters, Clustering without Classification, Clustering with Classification

## Introduction

RFM model is an apparatus of clustering customers into 3-dimensions, specifically, recency (R), frequency (F), and monetary value (M). In added arguments, RFM model helps to determine the top 20% of customers, who bring in 80% of revenue. In RFM model, recency (R) is defined as the intermission from the time when the latest consumption happens to the present, frequency (F) is the number of consumption within a certain period, and monetary (M) is the amount of money spent within a certain period. An earlier study showed that customers with bigger R, F, and M values are more likely to make a new transaction (Wu& Lin, 2002).

In order to group customers and perform analysis, a customer segmentation model-Customer pyramid model is used (Curry & Curry, 2000). Allowance of customer pyramid to model group customers by the revenue they generate is shown in Fig.1 (http://mnama.blogspot.ae).

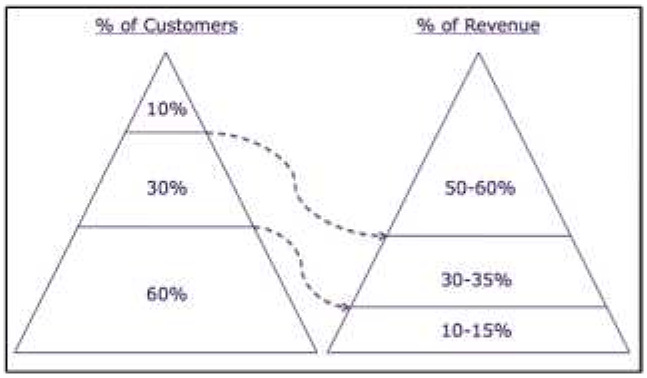


**Fig. 1: % of Customers v/s % of Revenue**

As stated in Fig. 1, the uppermost 10% of customers epitomizes amid 50-60% of revenue, next 30% embodies 30-35% of revenue. The bottom 60% of customers has awfully low value, and gives less than 15% of total revenue. These three stages of the customer value pyramid can be divided as active, warm, and cold. Added elaboration of the pyramid into 4 dimensions comprises the following four customer types– active, warm, cold, and inactive (https://lawsonhembree.wordpress.com).

Both the studies (http://mnama.blogspot.ae, https://lawsonhembree.wordpress.com) suggest that the customer exhibiting high RFM score should normally conduct more transactions and result in higher revenue. RFM analysis (Im, & Park, 1999; Madeira, 2002) is

* Associate Professor, Information Systems, Skyline University College, Sharjah, UAE. Email: riktesh.srivastava@gmail.com

used to further enhance the customer value pyramid into different subsections. As mentioned by Cheng & Chen (2009), there are two opinions on the importance of R, F and M values, while the three parameters are considered equally important in Miglautsch (2000). They are unequally weighted due to the characteristics of industry in Tsai and Chiu (2004). 96 data objects with 1659 observations collected for data analysis in the study adopted the weighted characteristics of R, F and M and further classified the customer clusters into 8 segments.

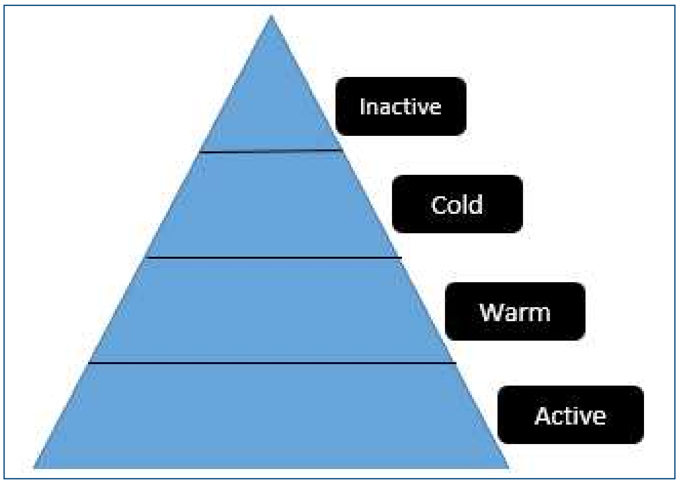


**Fig. 2: Four Dimensions of Customer Value Pyramid**

The complete paper is organized in 5 sections. Second section exemplifies the data analysis "without further classifications", labeled as "Clustering without classification", for R,F and M for all 96 data objects. Third section does the investigation "with further classifications", named as "Clustering with classification". Fourth section conducts the revenue analysis of 2015, 2014 and 2013 and gauges the customer clustering of 8 segments and revenue generated. Fifth section accomplishes the paper with recommendations and interpretations.

## Clustering without Classification

In clustering without classification, the customer value pyramid is divided in 4 layers, namely, Active, Warm, Cold and Inactive respectively. The output obtained in portrayed in Fig. 3.

There are two important observations from CVP:

1. No customer is inactive.
2. Majority of customers, 84.4% fall under "Active Cluster", which appears to be a worthy signal for the organisations.

The R, F and M accompanied for the 1659 observations are quantified in Figs.4a, 4b, and 4c, respectively.

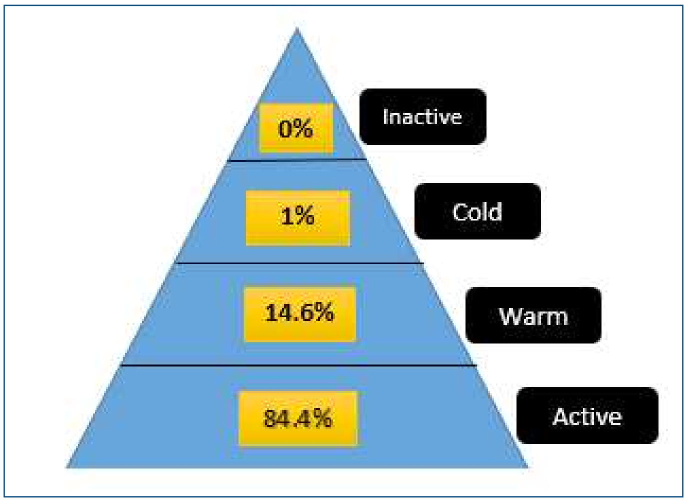


**Fig. 3: Clustering Without Classification**

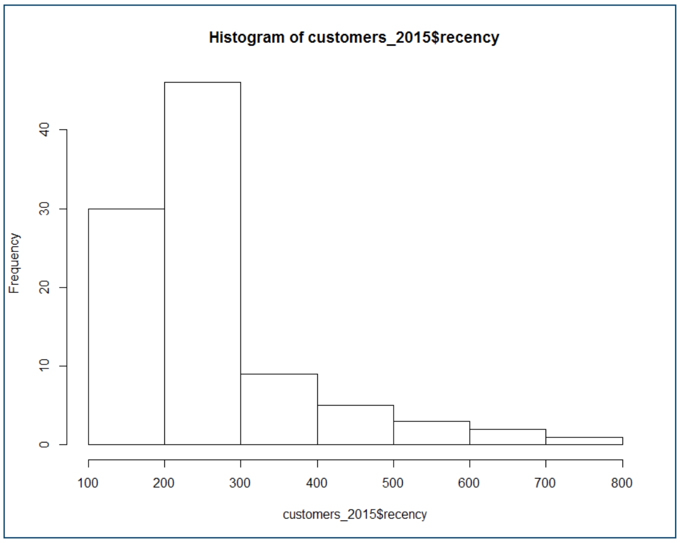


**Fig. 4a: Recency for Clustering without Classification**

The key observations from Figs. 4a, 4b and 4c are as under:

- Recency between two shopping space is between 100-300 days.
- The frequency of customers is quite high, and falls between 0-25 times.

- Maximum amount spend by customers falls in range of $900-$1000.

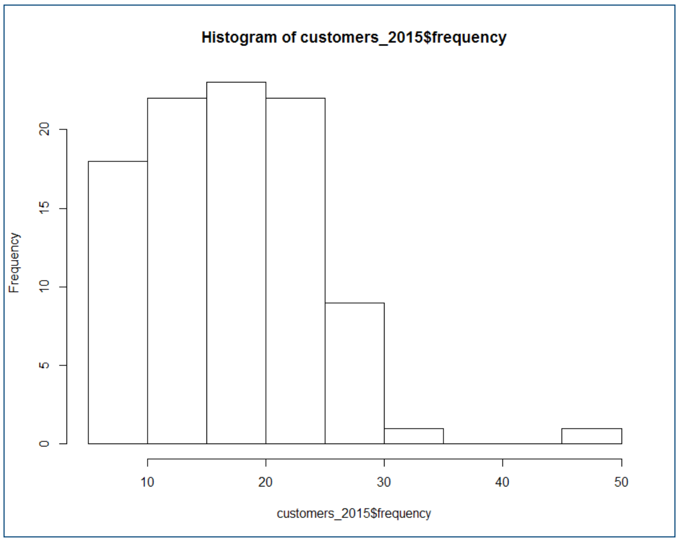


**Fig. 4b: Frequency for Clustering without Classification**

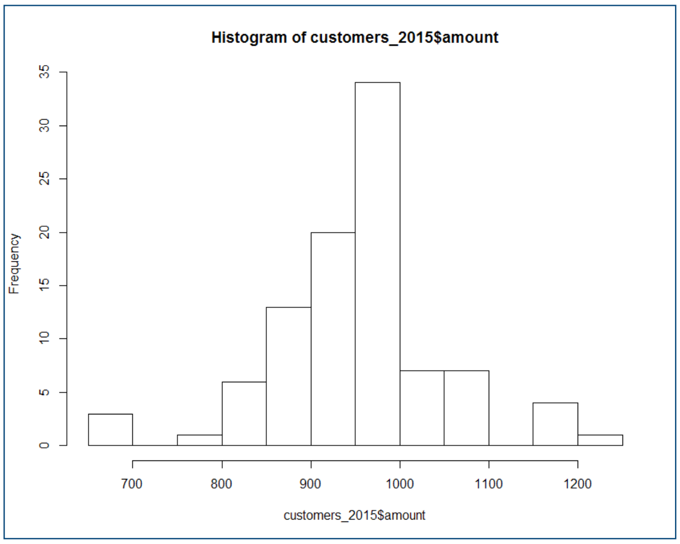


**Fig. 4c: Monetary Value for Clustering without Classification**

RFM analysis conducted for 4-dimension customer is shown in Fig. 5.

```
  Group.1  recency first_purchase frequency   amount
1  active 221.8873       975.8380  17.41975 951.5221
2    cold 756.8750       908.8750  21.00000 980.7143
3    warm 466.3036       954.2321  16.21429 950.8530
>
```

**Fig. 5: RFM Analysis for Clustering without Classification**

The observations are:

- The shopping space between active customers is 221.8873, however, the average amount spend by them is least at $951.5221.
- The shopping space between cold customers is 756.8750, however, the average amount spend by them is maximum at $951.5221.

## Clustering with Classification

Clustering with classification stretches the improved representation of the different types of customers in “active” and “warm” section, being two most vital categories of customers. The active type of customers is divided into 3 subsections –Active High, Active Low, and New Active, where New Active is the customer whose first purchase is within 365 days. Active High and Active Low are the classifications for the Monetary value (M) more than or less than 100 respectively. The warm type of customers is also alienated into 3 subsections – Active Warm, Active Warm and New Warm, where New Warm is the customer whose first purchase is within 365 days. Active Warm and Active Warm are the classifications for the Monetary value (M) more than or less than 100 correspondingly.

These added classifications are stated on CVP in Fig.6.

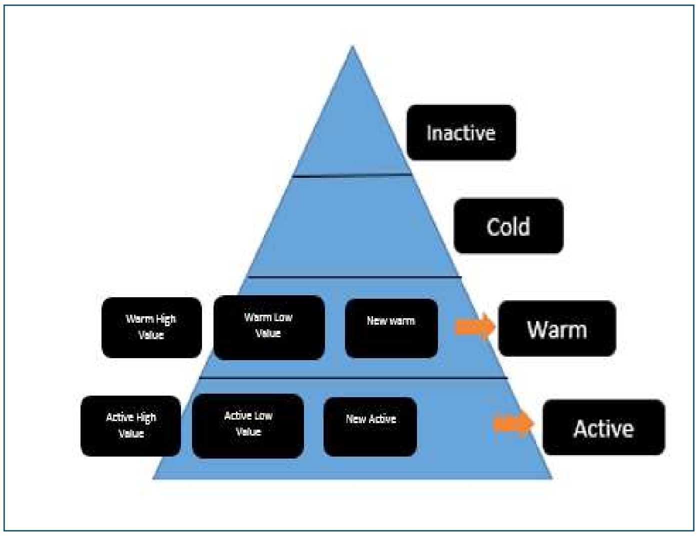


**Fig.6: Cluster with Classification**

Based on these classifications, trials were again piloted and effects witnessed are declared in Figs.7a, 7b and 7c, respectively.

```
> table(customers_2013$segment)

        inactive              cold   warm high value    warm low value          new warm active high value  active low value
               0                 0                 0                 0                91                 0                 0
      new active
               0
> aggregate(x = customers_2013[, 2:5], by = list(customers_2013$segment), mean)
   Group.1  recency first_purchase frequency   amount
1 new warm 375.4245       632.4684  8.648352 951.4591
>
```

**Fig. 7a: Outcome of Year 2013 for Cluster with Classification**

```
> table(customers_2014$segment)

        inactive              cold   warm high value    warm low value          new warm active high value  active low value
               0                 0                 0                 0                 7                84                 0
      new active
               5
> pie(table(customers_2014$segment), col = rainbow(24))
> aggregate(x = customers_2014[, 2:5], by = list(customers_2014$segment), mean)
            Group.1   recency first_purchase frequency    amount
1          new warm 379.58929       610.0179  8.571429  845.6222
2 active high value  64.30357       634.3393 13.440476  960.1452
3        new active  23.87500       143.2750 16.400000 1025.9247
>
```

**Fig. 7b: Outcome of Year 2014 for Cluster with Classification**

```
> table(customers_2015$segment)

        inactive              cold   warm high value    warm low value          new warm active high value  active low value
               0                 1                13                 0                 1                81                 0
      new active
               0
> aggregate(x = customers_2015[, 2:5], by = list(customers_2015$segment), mean)
            Group.1  recency first_purchase frequency   amount
1              cold 756.8750       908.8750  21.00000 980.7143
2   warm high value 469.1827       979.0288  15.76923 950.8312
3          new warm 428.8750       631.8750  22.00000 951.1364
4 active high value 221.8873       975.8380  17.41975 951.5221
```

**Fig. 7c: Outcome of Year 2015 for Cluster with Classification**

These outcomes give the improved perspective of the comprehensive cluster breakdown for three year interval. The foremost results are:

- There is only one category of customer – new warm – in year 2013, indicating no new customer was acquired during this period.
- New customers were acquired in year 2014 and they resulted in maximum monetary value.
- Many customers were "cold" in the year 2015, indicating they are not repeated customers.

## Revenue Analysis: Segmenting Database Retrospectively

The segment does retrospective analysis of the customers in 2013 and 2014, and their status in 2015, as shown in Fig. 8 and estimates the revenue created by these sets of customers in 2015. The analysis aids the organisation to outline certain set of policies for customer clusters.

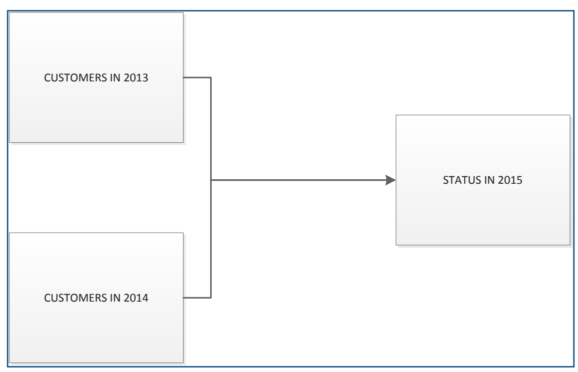


**Fig. 8: Retrospective Analysis of customer status in 2015**

The revenue generated by these sets of customers in 2015 is shown in Fig. 9.

Graphical representations of the revenue generated are also depicted in Figs.10a, 10b and 10c.

| | | |
|---|---|---|
| Group.1 x<br>4 active high value 4558.951<br>1 cold NA<br>2 warm high value NA<br>3 new warm NA | Group.1 x<br>3 new active 8556.000<br>2 active high value 3662.381<br>1 new warm 2693.571 | Group.1 x<br>1 new warm 3587.857 |
| Revenue generated by Customers acquired in 2015 and revenue in 2015 | Revenue generated by Customers acquired in 2014 and revenue in 2015 | Revenue generated by Customers acquired in 2013 and revenue in 2015 |

**Fig. 9: Revenue Status in 2015**

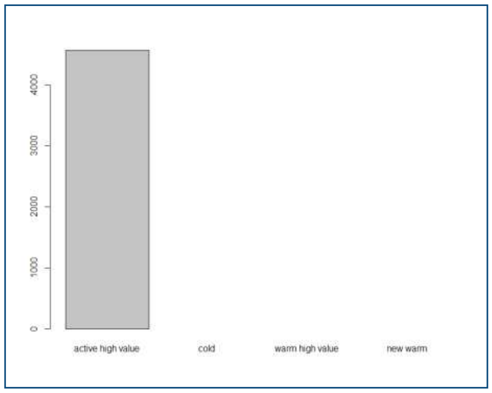


**Fig.10a: Revenue Generated by Customers Acquired in 2015 and Revenue in 2015**

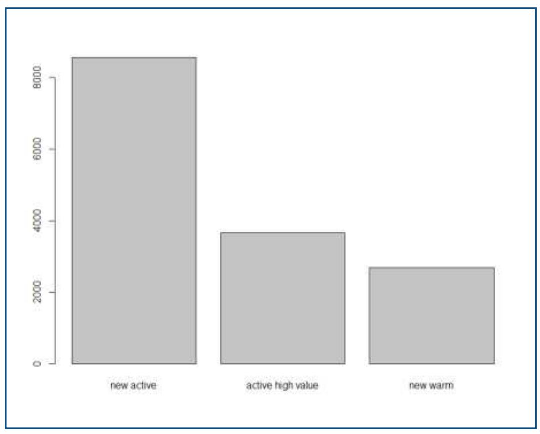


**Fig.10b: Revenue Generated by Customers Acquired in 2014 and Revenue in 2015**


**Fig.10c: Revenue Generated by Customers Acquired in 2013 and Revenue in 2015**

## Recommendations and Conclusion

The study divulges an acumen of the customer analysis and portrays a vital results for the organisations. R programming language is used to scrutinise the data objects and stretches an improved appreciative of the 1659 observations for 96 data objects (customers). These observations are reasonably keys to mount the strategies for customer acquisition. The result exposes that only 1.06% of active customers of 2013 and 2014 get transformed in 2015, which is frightening position for the organisation. Also, only 0.397% of warm category customers of 2013 and 2014 get transformed in 2015. This stipulates that 98.5% target is attained in 2015, and does not assure continual purchase in subsequent years. The company must outline the strategies for customer retaining, which might include announcing loyalty offers, sale deal, or superior deal for these customers.

## References


Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications,* 4176-4184

Curry, J., & Curry, A., (2000). *The customer marketing method: How to implement and profit from customer relationship management,* 7-15.

http://mnama.blogspot.ae/2010_11_01_archive.html(Online-checked on 18/12/2015)

https://lawsonhembree.wordpress.com/2011/03/15/marketing-alchemy/(Online-checked on 18/12/2015)

Im, K., & Park, S. (1999). *A study on analyzing characteristics of target customers from refined sales data*. APIEMS.

Madeira S. A. (2002). *Comparison of target selection methods in direct marketing.*MSc Thesis, Technical University of Lisbon.

Miglautsch, J. (2000). Thoughts on RFM scoring. *Journal of Database Marketing,* 67-72.

Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology.*Expert Systems with Applications.*

Wu, J., & Lin, Z. (2002). Research on customer segmentation model by clustering. *In Proceedings of the 7th ACM ICEC International conference on electronic commerce.*