

## A Heuristic Approach to Predictive Modeling... RFM Analysis



Ashutosh Kar

Mihir Ghosh

Nitish

Raghuveer Kodali

Sangitha Ajith

Sri Valli

Purba H. Rao

**Myra School of Business, Mysore, Karnataka.**

## Abstract

This paper discusses the basic postulates of Recency, Frequency & Monetary (RFM) analysis, a heuristic modeling approach, used in Predictive Analytics, to segment a target market into preferred segments and not so preferred segments. The preferred segments are characterized by their high response rate or high willingness to purchase as opposed to other segments which are not as preferred.

The paper also exemplifies these concepts with the help of a case study in the tele-communication sector where a company uses an existing data base to arrive at RFM categorization as well as identifies the profile of customers in the preferred segments.

## Introduction

Predictive modeling, the way it is understood in the Business Analytics context, is a way of predicting consumer behavior by analyzing a database either existing in the company concerned or on a database created with the help of an empirical survey. Essentially, a modeling approach, predictive modeling helps the company to identify profiles of consumers who would be more likely to purchase a product or a service which the company might be offering to a specified and defined target market. Applications of predictive modeling can be seen over different industries and in different managerial functions.

For instance, for an entrepreneur offering a new product in a specified target market, predictive modeling can help in understanding the consumer needs and preferences with respect to the attributes defining the product. For a service oriented company it can help to determine the profile of the most preferred segment and predict the percentage of customers who may actually purchase a new service being offered. For a credit card company or an organization offering loans of any kind, predictive modeling may evolve guidelines as to what kind of consumer profile would merit a preferred treatment and to whom loans may be extended with a softer level of interest.

Once the predictive modeling context is well understood and the objective in terms of what phenomenon is to be predicted has been clearly stated, the approach would define measurable variables for each item of the in the situation ... the predictor variables... as well as the variable to be predicted ... the dependent or target variable.

Thereafter the predictive modeling uses either:

(a) Analytical approach like Logistics regression, Linear Regression Analysis, Factor and Cluster Analysis, Conjoint analysis, or

(b) Heuristic approach... like RFM analysis, or

(c) Data mining approach, which combines Heuristic and Statistical approach such as Classification Trees.

This paper proposes to discuss the basic concepts of the Heuristic Approach of RFM Analysis and provide an example of RFM Analysis applied on the database of a company operating in the telecommunication field in India.

## Predictive Modeling and RFM analysis.

In strategic decision making companies often strive to determine who are the most valuable customers whom they would give special privileges to, invest to build up long term relations with, say in a CRM scenario, or target offers for mail orders, catalogue buying or any kind of direct marketing initiatives. The objective, in most of such situations, is to find out who the most likely buyers are, who makes purchases most frequently, who spend the most and who have the greater probability of coming back for repurchase. In many such initiatives, RFM analysis, recency-frequency-monetary analysis, helps identify consumer segments and customer profiles having such characteristics.

‘The fundamental premise underlying RFM analysis is that customers who have purchased recently - , have made more purchases and have made larger purchases are more likely to respond to your offering than other customers who have purchased less recently, less often and in smaller amounts.’

[Charlotte Mason, 2003, University of North Carolina].

The analysis helps an organization to focus on a smaller section of the target population which again follows another managerial premise, Pareto Principle that 80 % of the business comes from 20 % of the customers.

In the past 30 years, direct mailing marketers for non-profit organizations have used an informal RFM analysis to target their mailings to customers most likely to make donations. The reasoning behind RFM was simple: people who donated once were more likely to donate again. Currently, with the availability of CRM software and the use of e-mail marketing, RFM analysis has become an even more important tool. Using RFM analysis, customers are assigned a ranking number of 1,2,3,4, or 5 (with 5 being highest) for each RFM parameter. The three scores together are referred to as an RFM composite score. The database is sorted to determine which customers have been the best customers in the past, with a composite score “111” being ideal. Of course, in some organizations marketers consider 5 to be the most preferred RFM parameter, in which case ‘555’ would be the most preferred customer.

([http://searchdatamanagement.techtargget.com/sDefinition/0,290660,sid91\\_gci751219,00.html](http://searchdatamanagement.techtargget.com/sDefinition/0,290660,sid91_gci751219,00.html))

There are many justifications as to why RFM analysis works. Customers who bought most recently from an organization, are more likely to respond to the next promotion than those whose last purchase has been way back in the past. This is a universal marketing phenomenon and has been observed in many industries such as insurance, banks, cataloging, retail, travel, etc. In a similar manner, customers who have purchased frequently are more likely to respond than the less frequent ones. Also customers who are big spenders often exhibit much higher response rates than small spenders.

### **Applying RFM Analysis to Indian Telecom Sector**

Telecommunication has been recognized around the world to be an important tool for socio-economic development of a nation and plays a central role in the growth and modernization of various other sectors of the economy. Telecom growth is directly correlated to the GDP of developing countries.

After a period of rapid growth, Indian telecom operators face challenges now, as growth of voice revenues has slowed down, while the data growth has not been meaningful enough to make a substantial impact on the sector revenues yet. Hyper-competition, large investments in spectrum and network along with declining tariffs have led to significant pressures on operator profitability. A number of regulatory issues have also acted as roadblocks to growth.

Telecom sector in India is characterized by vast subscriber base, intense competition among the mobile operators and amongst the lowest tariffs in the world. Domestic subscriber base has made it one of the fastest growing industries. 40mn subscribers were added during the year 2012-13. The sector has experienced a rapid growth in the past decade owing to 1) increased penetration of mobile network, declining cost of mobiles, aggressive marketing and branding by telecom operators and a growing population with a need to be connected.

An interesting trend is, while the wireless subscribers have kept pace with the increase in overall subscribers, there has actually been a decline in the number of wireline subscribers over the years.

Telecom sector has experienced a rapid growth in the past decade owing to 1) increased penetration of mobile network, declining cost of mobiles, aggressive marketing and branding by telecom operators and a growing population with a need to be connected.

A case to support this point is the fact that there were 391.76mn subscribers as of March 2009. This number stands at 931.95 million as of March 2014 - an increase of about 600 million. Hence, on an average, a net 10 million subscribers were added each month which has put Indian telecom companies amongst the largest in the world - next only to China in terms of subscribers.

Sector regulator Telecom Regulatory Authority of India (TRAI) reports that the current teledensity is ~75% as of Mar 2014. This means that there is still a huge scope for increasing the number of subscribers. In other words, approximately ~300 million subscribers are still out of the gambit of network operators. The rural teledensity stands at 43.96%. Considering the fact that

most of India's population resides in rural areas, the rural market provides huge opportunities for expansion. The active subscriber penetration (measured on VLR) is 64% ~ 791 million subscribers. This leaves a vast scope of subscriber addition and hence increase in voice revenues.

## HOW CAN ANALYTICS HELP

Analytics is increasingly being used as a tool to solve complex organizational problems – leading to better decisions. These are the decisions which were once taken solely by gut instincts.

The success of any company in the Telecom industry currently depends on two broad factors –

- Ability to add new subscribers (both data and voice)
- Ability to retain existing subscribers (Since, Mobile Number Portability is now available by all operators)

This paper focuses on the second part, which is on the indicators which would help the company minimize the tendency of subscribers to switch from their service to others. One of the major indicators of this tendency is measured by **Port-in Port-out ratio** (it is also commonly referred to as Churn).

The number of subscribers switching to a given provider from others is referred to as Port-in. Port-out indicates the number of subscribers switching to a different provider from the given company. A port-in port-out ratio of less than 1, hence, is good for the company because more subscribers are coming in than out. If the ratio is greater than 1, it is considered bad for the company.

**There are two ways of making this ratio healthy – increase the number of port-ins or decrease the number of port-outs.** In order to do so, the company would need to strategically connect with the individual subscriber base – the better their needs are taken care of, the less likely it is that they will switch to other provider.

Keeping the above fact in mind, a telecom provider usually comes up with a number of plans to woo the existing customers. This comes with a catch though. It is almost impossible to roll out tailor made plans for every subscriber – such would be too expensive. At the same time, covering the entire subscriber base with a few plans would not go down well with specific consumers whose needs might be different.

**One solution** might be to come up with a plan, say, '**Pay-Per-Use**' plan – and in order to do so, a broad survey of consumers is required – most of their usage details are already with the company. **It is their preferences which need to be mapped with their eagerness to take up a new plan from the same provider instead of switching to another service provider.** This new service was called 'dataplan' by the telecommunication company.

Thereafter the company wished to use Analytics to help identify, who among the existing subscribers are most willing to take up the new plans. It is with this end in mind, that the data has been collected, data base was constructed and analyzed. The database elements have been discussed in the next section.

### **RFM Analysis on Telecom Company database.**

To apply RFM the telecommunication company considered a database to be analyzed before the launch of a new service (data plan). They have included various variables for the research. Some of them are taken from the company's own data base such as recency, frequency and amount spent. Other variables are taken by calling to a sample of their existing customers such as if they have children, broadband connection etc.

The company basically wants to know if the new service would attract the customers or not for its 'Pay-Per-Use' plan.

We have conducted **RFM analysis** followed by **one way ANOVA** test to come with a profile of customers to whom company should target.

**Variables Included and their Explanation:**

The following is an explanation of the variables –

1. CustomerID – A list of auto-generated customer ids with the telecom company.
2. ZipCode – The zip code of the customer filled in the survey. Although the VLR (Variable lookup register) might give a different pin code in the same state.
3. DaysLastRecharge – This is the number of days since the customer's last recharge for any services provided by the company. The lower this value, the better for the company.
4. Amount – This is the amount of money spent by the consumer in the last one full year on all services. The higher this value, the better for the company.
5. RechargeFrequency – This is the number of times, in a year, the consumer has recharged the cellular phone.
6. DataUsage – This is the amount of data (in GBs) used by the subscriber in last 1 year. This data field has been categorized into 5 parts and named it dataUsagecategory:

1	2	3	4	5
> 11 GB	(9.1 – 11) GB	(7.1 – 9) GB	(5.1 – 7) GB	< 5 GB

7. HasChild – If the consumer has child/children or not.
8. HasBroadband - If the consumer has broadband connection

***Applying RFM Approach to the Telecom database:***

***Creating Recency Quintile***

From the datafield 'DaysLastRecharge', which gave the number of days since the last recharge, the Recency quintiles were created and labeled so that the Recency category '1' referred to the 20 % of all customers who had the smallest measures on 'DaysLastRecharge'. The Recency category '5' referred to the 20 % of all customers who had the largest measures on 'DaysLastRecharge'

***Creating Frequency Quintile***

From the datafield RechargeFrequency, the number of times, in a year, the consumer has recharged the cellular phone, Frequency Quintiles were created and labeled so that the Frequency category '1' referred to the 20 % of all customers who had the largest measures on

‘RechargeFrequency’. The Frequency category ‘5’ referred to the 20 % of all customers who had the smallest measures on ‘RechargeFrequency’.

### *Creating Monetary Quintile*

From the datafield, ‘Amount’, which is the amount of money spent by the consumer in the last one full year on all services, the ‘Monetary’ quintiles were created and labeled so that the Monetary category ‘1’ referred to the 20 % of all customers who had the largest measures on ‘Amount’. The Monetary category ‘5’ referred to the 20 % of all customers who had the smallest measures on ‘Amount’.

### *Creating Composite Score and Composite Quintile*

Having created the three components of RFM Analysis, R for Recency, F for Frequency and M for Monetary, the combined 3-digit composite score was created as R for first digit, F for next digit and M for the last (third) digit. The Combined score thus labeled implies smaller the score the more preferred the customer is. Once again, quintiles were constructed on Composite Scores with category ‘1’ for the 20 % smallest scores for 20 % most preferred customers.

The composite scores on one hand identified the most preferred customers in the database. On the other hand they identified the profiles of these most preferred customers.

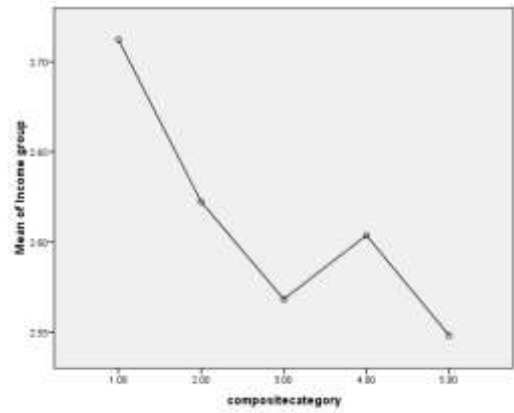
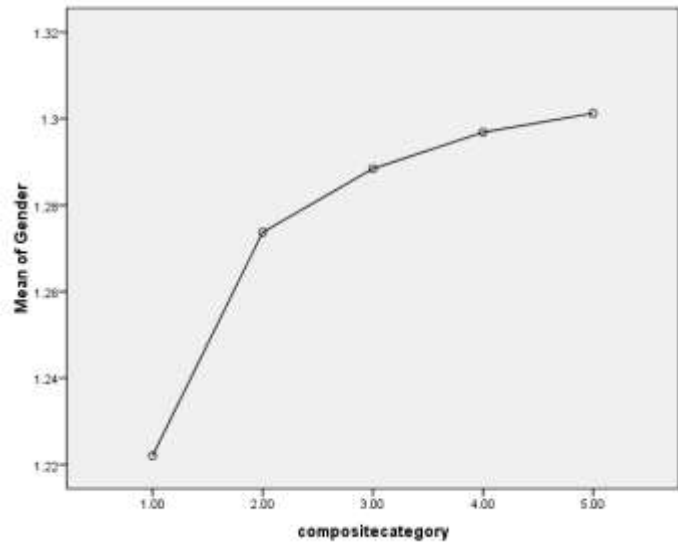
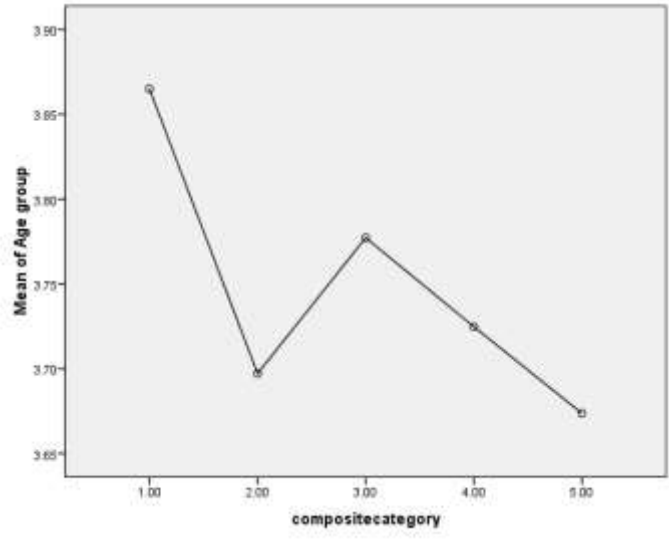
To arrive at these profiles One-way ANOVA was conducted with Composite score quintiles as the Factor Variable.

### **Results from One-way ANOVA.**

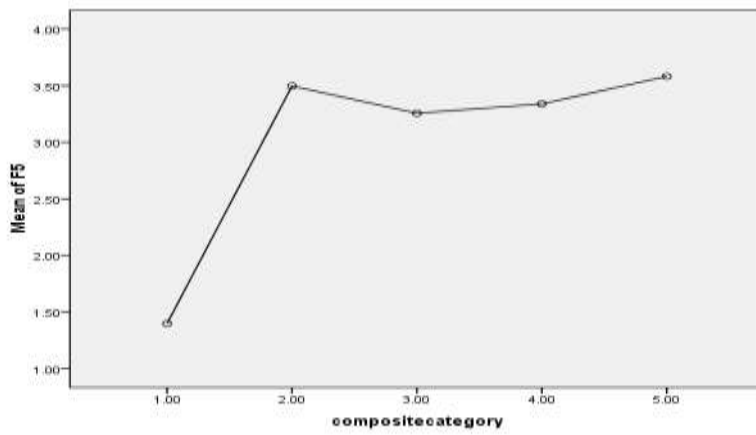
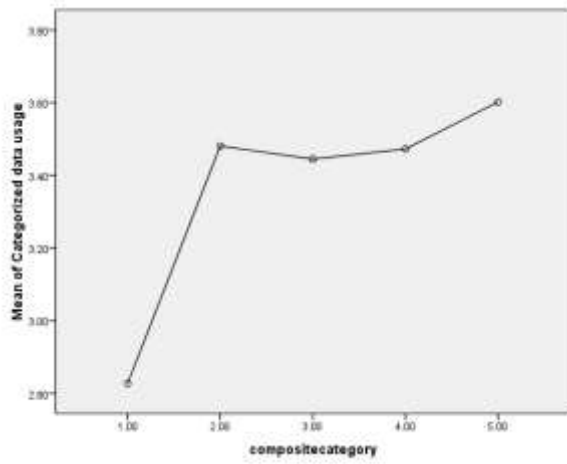
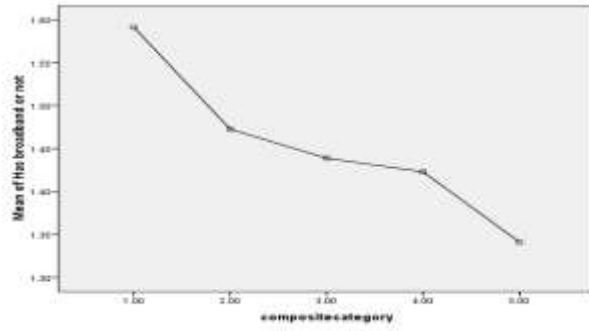
Using composite score category as factor the following variables emerged as significant judging from the significance of the F-statistic, at the 5% level of significance:

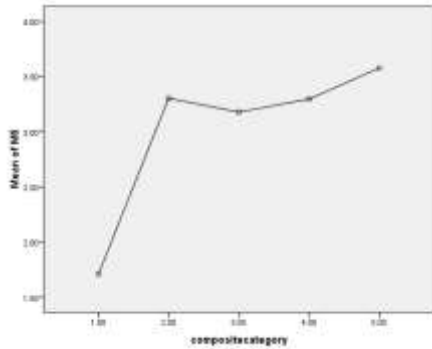
<b>Variable</b>	<b>F-statistic</b>	<b>Significance</b>
Age group	4.635	.001
Gender	2.883	.021
Income group	2.529	.039
Has broad band	18.212	.000
R5=Recency	22674.315	.000
F5= Frequency	462.154	.000
M5=Monetary	195.5	.000
Data usage category	138.515	.000

The Means plot for some of the above significant variables are as follows:









From the above analysis *the profile of most preferred customers* emerge as:

Higher Age group  
 Predominantly male  
 Higher income group  
 People who have broad band  
 People who are in lower data usage category  
 People who have recharged recently  
 People who recharge frequently.  
 People who spent large amount of money in recharging

## Conclusion

Using a Heuristic approach in Predictive Analytics such as RFM Analysis, as explained above, many organizations who have large volumes of transactions, can identify a customer segment in their existing database who belong to the most preferred category.

The analysis can be further extrapolated to identify the preferred segment, best 20%, next best segment and finally the least preferred segment in the primary target market.

## References

Rao, P, 2013, Business Analytics an Application Focus , PHI, New Delhi.

Rao, P., Ray, S. and Kumar, P. 2013. Business Analytics, A Perspective, International Journal of Business Analytics & Intell8igence, Publishing India Group, New Delhi.

Reichheld, F. F., (1996), The Loyalty Effect, Harvard Business School Press, U.S.

Recency Frequency Monetary Analysis : RFM Analysis.

[http://cmason.myweb.uga.edu/Course\\_Roadmap/\\_RFM\\_Analysis/Note\\_-\\_RFM\\_Analysis.pdf](http://cmason.myweb.uga.edu/Course_Roadmap/_RFM_Analysis/Note_-_RFM_Analysis.pdf)



