RFM CUstomer segmentation

REPORT

QIVOS

Chatzimoschou angeliki

june 2017

# TABLE OF CONTENTS

[TABLE OF CONTENTS 2](#_Toc486162765)

[PART I – DATA REPRESANTATION AND EXPLORATION 3](#_Toc486162766)

[PART II – DATA PRE-PROCESSING AND TRANFORMATION 4](#_Toc486162767)

[PART III – descriptive VisualiZations 5](#_Toc486162768)

[PART IV – RFM SEGMENTATION 10](#_Toc486162769)

# PART I – DATA REPRESANTATION AND EXPLORATION

In this project, we are assigned to deploy RFM customer segmentation on the “rfm” dataset provided by Qivos. This dataset summarizes a heterogeneous set of features about historical transactions with various field columns collected in a period of 360 days.

Specifically, this dataset contains 4 attributes (with numerical, categorical and date values) describing different aspects of each customer’s purchase, from a total of 830.853 purchases. It is composed by 4 independent variables, such as: transaction date, client Id, Invoice id and amount spent per purchase.

The goal of RFM Analysis is to segment customers based on their historical buying actions. Afterwards, we rank them based on each individual RFM factor, and finally pull all the factors together to create RFM segments for targeted marketing.

RFM stands for:

* **Recency** represents the “freshness” of customer activity. This attribute is the number of days since last order or interaction.
* **Frequency** captures how often the customer buys. This attribute is the total number of orders in the last year or during the whole customer’s lifetime.
* **Monetary**indicates how much the customer is spending. This attribute is the sum of all orders’ paid amount.

# PART II – DATA PRE-PROCESSING AND TRANFORMATION

To accomplish our analysis, we used the following tools:

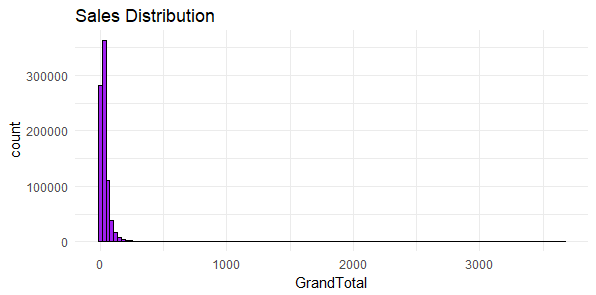
* R Studio version 1.0.136
* M.R.O version 3.3.3

For the stage of preprocessing, we deployed a few techniques using the tool R Studio and Microsoft R Open programming language. More specifically, we performed data pre-processing techniques in order to examine at a deeper level the dataset. After checking for missing values, we realized that our dataset didn’t contain any, so we simply transformed the “Timestamp” feature at a proper “Date” form.

In addition, we decided to create some additional variables to help us create more insightful visualizations at a monthly basis. At first, we created the feature called “ym”, which keeps only the month and year for each transaction. Then, “Total Orders”, “Total Sales” and “Average Sales” summarized by month too.

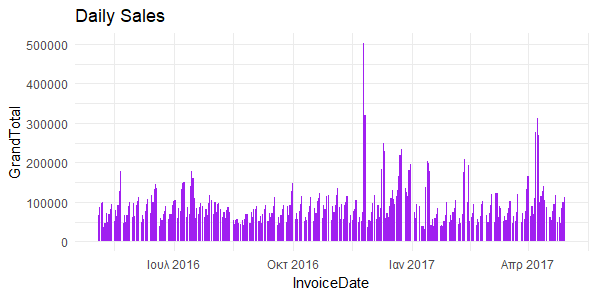
# PART III – descriptive VisualiZations

At first, by deploying a histogram on Sales data, we observe that this variable is highly right –skewed, which may lead to wrong segmentation since our data are of different scale.   
However, we won’t perform any suitable normalization technique (ie. Log-transformation) in this project and will stick to the typical steps as the final segmentation procedure is already given to us and we won’t create our own segments.

**

*Figure 1 – Sales Distribution Histogram*

Here we perform a daily representation of our Sales data and observe that there are days with really low Sales value compared to the general Daily Sales trend, such as: 16-04-2017 (17.70€), 15-08-2016 (19.70€), 26-06-2016 (37.91€), 24-07-2016 (74.10€) and 14-08-2016 (87.02€). On the opposite side, the top 5 Days with the highest Sales amount to collect for the company are:

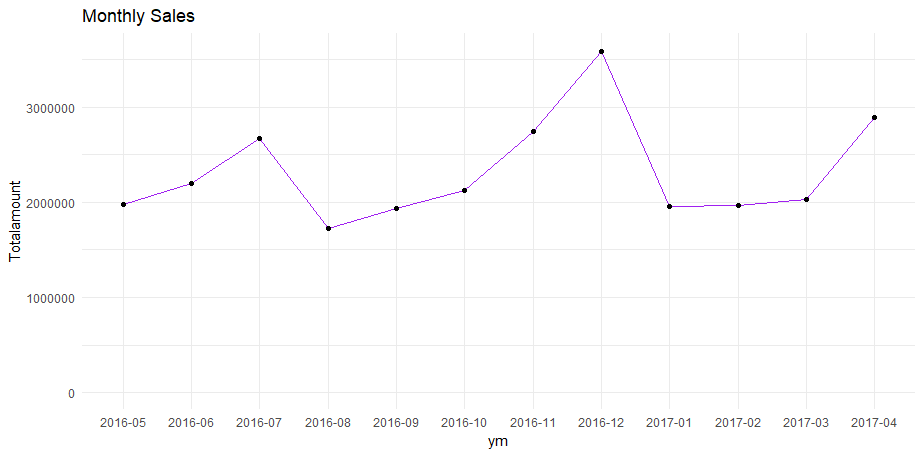
**

*Figure 2 – Daily Sales*

|  |  |
| --- | --- |
| DATE | AMOUNT |
| 25 - 11 – 2016 | 503.738€ |
| 26 – 11 – 2016 | 320.388,66€ |
| 08 – 04 – 2017 | 313.423,98€ |
| 07 – 04 – 2017 | 277.215,99€ |
| 09 – 04 – 2017 | 269.479,54€ |

*Table 1 – TOP 5 Days with highest Sales*

Observing the Total Sales per month, we may conclude that there is an obvious fluctuation during the examined time period (May 2016 – April 2017) with the lowest values pointing on August 2016 and January 2017. Depending on the product, August is the month that the majority of people take their summer vacation and probably this is why sales hit their lowest point then and also January comes after Christmas holidays that people have spent already much money and thus tend to spend less afterwards.

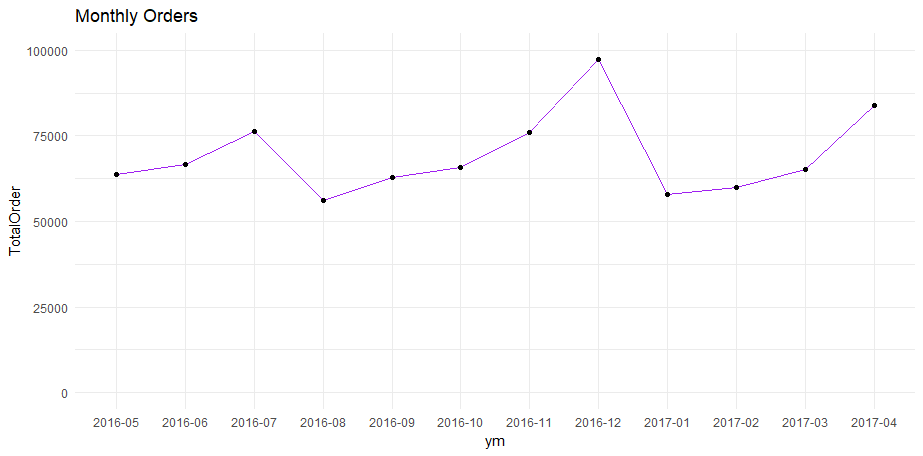
**

*Figure 3 – Total Sales per month*

|  |  |
| --- | --- |
| MONTH | TOTAL AMOUNT |
| 12 – 2016 | 3.585.089€ |
| 04 – 2017 | 2.888.814€ |
| 11 – 2016 | 2.747.909€ |
| 07 – 2016 | 2.676.350€ |
| 06 – 2016 | 2.196.376€ |

*Table 2 – TOP 5 Months with highest Total Sales*

Additionaly, the Total Orders trend line per month appears to follow a similar fluctuation with the one of Total Sales with the only difference that on July 2016 there were spotted more Orders than on November 2016.

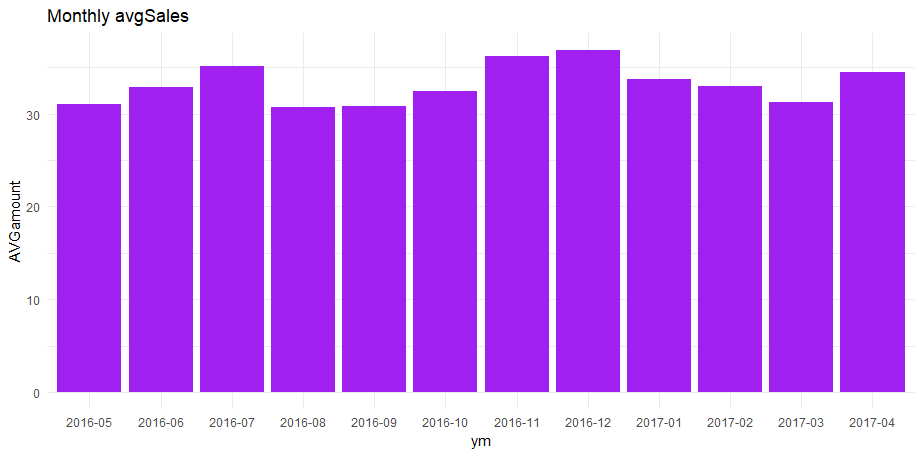
**

*Figure 4 – Total Orders per month*

|  |  |
| --- | --- |
| MONTH | TOTAL ORDERS |
| 12 – 2016 | 97.201 |
| 04 – 2017 | 83.855 |
| 07 – 2016 | 76.209 |
| 11 – 2016 | 75.831 |
| 06 – 2016 | 66.772 |

*Table 3 – TOP 5 Months with highest Total Orders*

Checking on the average Sales though, apart from the main characteristics mentioned on our previous charts, the months with the highest average sales are different than the ones with the highest total sales. This may have happened, after having a lot of orders with lower value amount at a specific month.

**

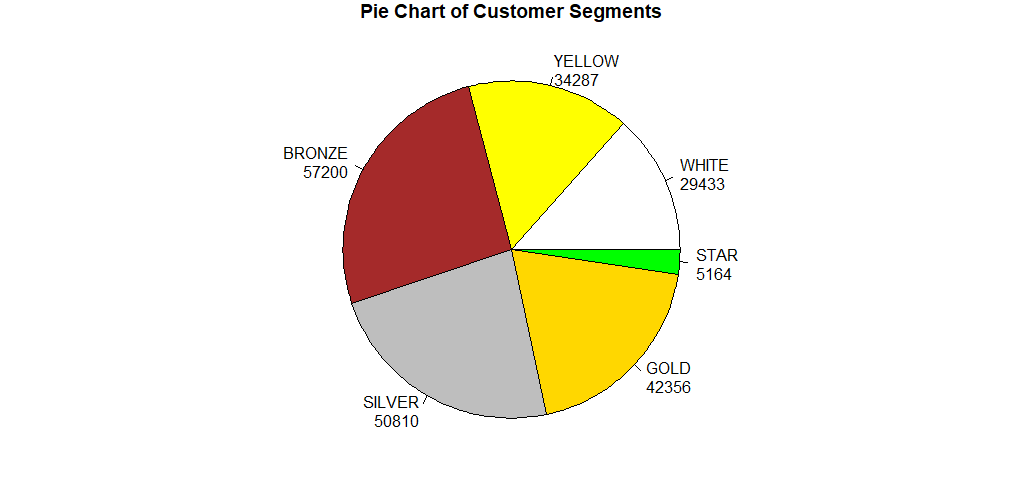
*Figure 5 – Average Sales per month*

|  |  |
| --- | --- |
| MONTH | Average SALES |
| 12 – 2016 | 36,88€ |
| 11 – 2016 | 36,24€ |
| 07 – 2016 | 35,12€ |
| 04 – 2017 | 34,45€ |
| 01 – 2017 | 33,71€ |

*Table 4 – TOP 5 Months with highest Average Sales*

# PART IV – RFM SEGMENTATION

After implementing the adequate procedure to get the initial RFM customer segments, we also calculated the final weighted RFM score for each group. Since we deploy our own weights, there is no need for any further normalization on the R - F - M values.

*****Figure 6 – RFM groups*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RFM GROUP | No. of CUSTOMERS | TOTAL AMOUNT | TOTAL ORDERS | AvgBASKET |
| STAR | 5.164 | 1.593.598,4€ | 37.335 | 308,60€ |
| GOLD | 42.356 | 15.635.767,2€ | 404.536 | 369,15€ |
| SILVER | 50.810 | 6.105.191,5€ | 189.932 | 120,16€ |
| BRONZE | 57.200 | 3.116.634,1€ | 116.948 | 54,49€ |
| YELLOW | 34.287 | 812.653,6€ | 47.614 | 23,70€ |
| WHITE | 29.433 | 548.034€ | 34.488 | 18,62€ |

*Table 6 – RFM Customer Groups*

Now that all customers are divided into RFM segments, we have identified a number of different groups that need to be approached by very different marketing strategies. Here are the customer groups using this technique:

Cluster **STAR** contains the best customers who haven’t bought that recently, but on the other hand they seem to buy often and tend to spend a lot on each purchase. It’s likely that they will continue to do so. Since they already like the company so much, consider marketing to them without price incentives to preserve profit margin. It is proposed to tell these customers about new products, how to connect on social networks, and any loyalty programs or social media incentives the company runs.

Cluster **GOLD** contains good customers, who have bought recently, are frequent clients and spend a considerably good amount of money on the products offered. We may treat them similarly to STAR cluster.

Cluster **SILVER** contains the kind of customers who recently bought from our shop, are slightly frequent and spend medium to low amount of money on their orders. Since their frequency rate is medium, this is the kind of customers we want to convert into loyal, regular customers that love our products and brand. A good suggestion would be to offer them a set of discounts periodically.

Cluster **BRONZE** contains lost customers (churners), who used to buy frequently from us, and at one point they used to spend a lot with us, but they’ve stopped. Now it’s time to win them back. They might be lost to a competitor, they might not have need of our products anymore, or they might have had a bad customer service experience. Despite this fact, they were an extremely valuable customer that should be approached again with extreme care.

Cluster **YELLOW** contains the worst customers, in other words the ones who spend very little, have bought very few times, and last ordered quite a while ago. They are unlikely to be worth much time, so it is suggested to put them in our general house list and consider a re-opt-in campaign.

Cluster **WHITE** contains the almost lost and some freshly new customers. For the almost lost clients, it has been some time since their last purchase and the newly comers have just showed up. These customers might need more aggressive discounts so that we can win them back before it’s too late, since it is much less expensive to maintain them compared to attracting new ones. For the newly-freshers, it is proposed to welcome them and thank them for making a first purchase and follow it up with unique offers to lure them back again. Consider branding the email with a special note from the CEO and include a survey to ask about their experience.

Many times segments can be combined to receive an offer. For example, you may combine “Lost Customers” and “Almost Lost” for a win-back campaign. “Going silent” or creating “contact fatigue” have negative consequences and should be managed appropriately.