

# Marketing Segmentation on Sales Data

---

Keval Patel

[Kp14@myscc.ca](mailto:Kp14@myscc.ca)

# Table of Contents

<b>Market Segmentation .....</b>	<b>3</b>
<b>Data Overview and Cleaning .....</b>	<b>3</b>
<b>Exploratory data analysis .....</b>	<b>4</b>
Plot 1 (Box plot Distribution) .....	4
Plot 2 (Countries by Sales) .....	4
Plot 3 (Annual Revenue) .....	4
Plot 4 (Monthly Revenue) .....	4
Plot 5 (Monthly Sales Growth Rate).....	5
Plot 6 (Average Sales per Order).....	5
<b>RFM Analysis.....</b>	<b>5</b>
RFM Distribution.....	5
RFM boxplot .....	6
Clean_RFM boxplot.....	6
Clean_RFM Distributions .....	6
Elbow plots for RFM and Clean_RFM.....	6
Frequency vs Monetary RFM Uncleaned data .....	6
Monetary vs Recency RFM Uncleaned data .....	6
Frequency vs Monetary RFM_Clean .....	7
Monetary and Recency for RFM_Clean.....	7
RFM plot in 3-D.....	7
<b>Conclusion .....</b>	<b>7</b>
<b>References .....</b>	<b>7</b>

## Market Segmentation

Market Segmentation is a method of separating a market of Customers into groups based on different features. The groups were based on customers who response in the same way to the market and who share a similar interest.

In terms of my dataset, we can find customers who have a similar interest or who respond the same to the market by doing RFM analysis and focus on a particular group of clusters to create a better profit or improve their business.

## Data Overview and Cleaning

For Data Overview, I first import the necessary packages and then import data. As part of importing data, I didn't apply 'index\_col=0' because if I do that I will unstructured index or I have to reset index afterward. After importing data, I check for the datatype, null values, and shape of data. In next step, I remove unnecessary columns 'Unnamed: 0', 'PHONE', 'ADDRESSLINE1', 'ADDRESSLINE2', 'STATE', 'POSTALCODE', 'TERRITORY' because these all are not helpful in analysis or marketing segmentation and some has null values.

# Exploratory data analysis

## Plot 1 (Box plot Distribution)

Plot 1 is based on distributions of all numeric columns of data in box plot and by box plot, we can easily see outliers, skewness of data, variance in data.

- In data, we have some outliers in 'SALES' and 'QUANTITYORDERED'.
- We have some skewed in data such as 'PRICEEACH', 'YEAR\_ID', 'ORDERLINENUMBER' etc.
- We have some high variance data such as 'MSRP', 'PRICEEACH', 'ORDERLINENUMBER' etc.

## Plot 2 (Countries by Sales)

In Plot 2, I describe countries by sales so, by seeing this data we can say that most people who buy products are from the USA and so on we can describe more.

By seeing this graph, we can say that where we want to increase our business, or we need to find why this particular country has a lower Sale.

## Plot 3 (Annual Revenue)

In Plot 3, I describe Annual Revenue which describes that total amount of money made per year and as we can see in the graph, Annual Revenue of 2004 is more than 2003 and 2005 but we should know that we don't have data of whole 2005 so, by this, we can't say that in 2004 we have highest Annual Revenue.

## Plot 4 (Monthly Revenue)

In Plot 4, I describe Monthly Revenue which describe that total amount of money made per month in particular month and as we can see in graph, in November we have highest revenue the reason behind this could be festival such as Black Friday, thanksgiving etc. we can also understand by seeing this each year sale is increasing then last year so, every year sales is increasing. In 2005, we have limited data so, we can see sale up to May.

### Plot 5 (Monthly Sales Growth Rate)

In Plot 5, we can see Monthly Sales Growth Rate, in which sales below 0.0 that means the company is in loss and more than 0.0 means company is in profit and by calculating we can also say that how much percentage profit company made, and we can also compare it with last month. So, for example, in May-2005 company made 75% more profit than last month.

### Plot 6 (Average Sales per Order)

In Plot 6, I describe Average Sales per order, in which we can see that which month has more average sale than others when we compare with other months and by seeing this, I can say that April- 2005 has more average sale than any other months.

## RFM Analysis

RFM analysis is a marketing strategy used to determine quantitatively which customers are the best ones by examining how recently a customer has purchased (recency), how often they purchase (frequency), and how much the customer spends (monetary).

### RFM Distribution

Recency histogram shows that customers order date are extremely high and recent between day 0 and approximately day 50. their purchase recency (order date) tends to decline over time and increase after around 150 days and again decrease after 200 days.

Frequency histogram shows that customers re-order or purchase are extremely frequent between 0 to approximately 25 days.

Monetary histogram shows that most of the customers spend/purchase more between 0 - 100,000.

## RFM boxplot

In this plot we can see the outliers in RFM data. So, we need to remove that outliers.

## Clean\_RFM boxplot

In this plot I removed outliers and see the distributions.

## Clean\_RFM Distributions

Recency histogram shows that customers order date are extremely high and recent between day 0 and approximately day 50. their purchase recency (order date) tends to decline over time and increase after around 150 days and again decrease after 250 days.

Frequency histogram shows that customers re-order or purchase are extremely frequent between around 22 to 27 days.

Monetary histogram shows that most of the customers spend/purchase more between around \$75000 - \$85000.

## Elbow plots for RFM and Clean\_RFM

By seeing Elbow plot, we can define that hoe many clusters we need to choose and here we choose number of clusters is 2 because after 2 clusters variance is constantly decreeing but Elbow method just give the idea about clusters. it's just suggestion not a perfect answer.

## Frequency vs Monetary RFM Uncleaned data

In this plot, we can see that how frequent customers are buying products and with range of cost. In this data we have 2 outliers which is second cluster.

## Monetary vs Recency RFM Uncleaned data

In this plot, we can see that how recent customers are buying product with range of cost. In this data we have 2 outliers which is second cluster.

## Frequency vs Monetary RFM\_Clean

In this plot, we can see that how frequent customers are buying products and with range of cost and here, we have 2 defined clusters.

## Monetary and Recency for RFM\_Clean

In this plot, we can see that how recent customers are buying product with range of cost and here, we have 2 defined clusters.

## RFM plot in 3-D

In this plot, we can see that we can see that how frequent customers are buying products and with range of cost and how recent customers are buying product and we can see clusters in better way and also identified customers in the clusters. By this technique, we can target specific customers based on cluster.

## Conclusion

Based on the results, I provide recommendations on which cluster should the company target for the next marketing campaign. Always keep the size of the cluster in mind when recommending it.

By seeing 3D plot, we can say that company should target cluster 2 for next marketing campaign because when we see the values of cluster 2, we can see that customers are not beneficial as cluster1 because In cluster2 customers are buying products which are costly and they buy products very recently but not frequently such as Customer-1 buys Product-1, Product-2, and Product-3 within two days but he only buy products once in a year. On the other hand, Customers who belong to cluster1 they spend less money on products, but they buy products very frequently and recently.

## References

Dataset: <https://www.kaggle.com/kyanyoga/sample-sales-data>

Market Segmentation: <https://trackmaven.com/marketing-dictionary/market-segmentation/>

RFM Analysis: <https://searchdatamanagement.techtarget.com/definition/RFM-analysis>