FACTOR HAIR REVISED

Karthik Paranthaman

UNIVERSITY OF TEXAS

1. PROJECT OVERVIEW

The main aim of this project is to explore the data set provided by the client "Factor Hair Revised" to find business insights and improve services. The R-Studio Platform is used to analyze the dataset using descriptive statistical tools such as Mean, Median, IQR, Standard Deviation, Variances, Linear Regression and Principal Component Analysis and Factor Analysis. The outputs are visualized using varies graphical methodologies to interpret with an aid of plots like Histogram, Bar Plot, Box Plot, Density Plot, and etc.

2. ASSUMPTIONS

I feel the rating is very subjective to customer's opinion and the mindset he or she was at during data collection period. Also, there was another study shows that most people doesn't understand the method of rating due to the fact of it's just number from customer point of view.
It was assumed that all information provided by the customers are accurate. The error caused during data entry is been identified using Sanity Check methodologies.

3. EXPLORATORY DATA ANALYSIS

In this Chapter we will be discussing about the data exploration approaches to find business insights. The exploratory process consists of the following stages.

1. Environment Setup and Data Import
2. Variable Identification
3. Univariate Analysis
4. Bi-Variate Analysis
5. Missing Value Identification
6. Variable Transformation
7. Feature Creation & Exploration

3.1 Environment Setup and Data Import

3.1.1 WORKING DIRECTORY

The working directory has been set. The dataset is called using the name "Customer_Feedback" in R-Studio (Refer to APPENDIX -1 – for R Code).

3.1.2 LIBRARY PACKAGES

The library packages used in this project are listed as follows:

1. readr
2. ggplot2
3. rpivotTable
4. dplyr
5. esquisse

3.2 Missing Value Identification

The Customer_Feedback dataset contains 13 Variables and 100 observations.
Sanity check was done to make sure there are no missing values in the dataset by using the following code:

anyNA(Customer_Feedback)*# Outputs if any Missing values were spotted*

In addition to that "head" & "tail" commands are used to call top & bottom rows to identify for any formatting issue in the dataset given. Furthermore, output from "summary (Customer_Feedback)" command also approves the Sanity Check from missing values and data classes.

3.3 UNIVARIATE ANALYSIS

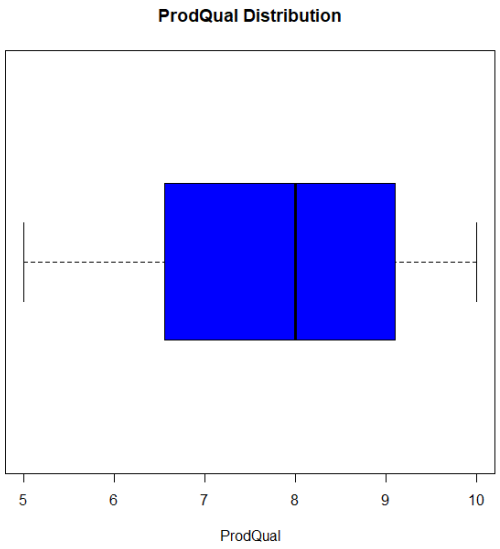
5 Number Summary analysis is done for each variable and the output is given below.

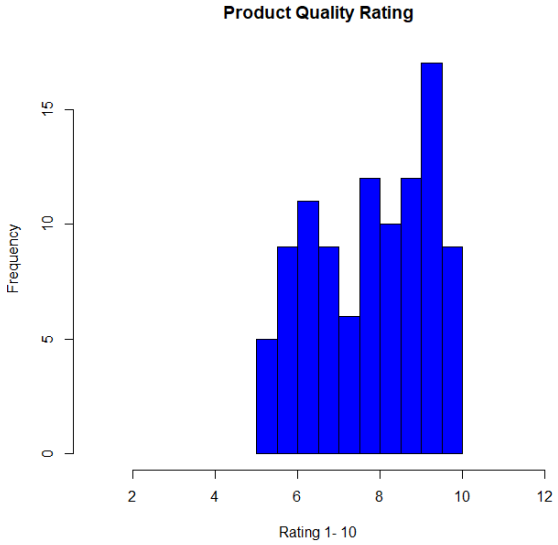
```
    ProdQual          Ecom          TechSup         CompRes
 Min.   : 5.000   Min.   :2.200   Min.   :1.300   Min.   :2.600
 1st Qu.: 6.575   1st Qu.:3.275   1st Qu.:4.250   1st Qu.:4.600
 Median : 8.000   Median :3.600   Median :5.400   Median :5.450
 Mean   : 7.810   Mean   :3.672   Mean   :5.365   Mean   :5.442
 3rd Qu.: 9.100   3rd Qu.:3.925   3rd Qu.:6.625   3rd Qu.:6.325
 Max.   :10.000   Max.   :5.700   Max.   :8.500   Max.   :7.800

  Advertising       ProdLine      SalesFImage      ComPricing
 Min.   :1.900   Min.   :2.300   Min.   :2.900   Min.   :3.700
 1st Qu.:3.175   1st Qu.:4.700   1st Qu.:4.500   1st Qu.:5.875
 Median :4.000   Median :5.750   Median :4.900   Median :7.100
 Mean   :4.010   Mean   :5.805   Mean   :5.123   Mean   :6.974
 3rd Qu.:4.800   3rd Qu.:6.800   3rd Qu.:5.800   3rd Qu.:8.400
 Max.   :6.500   Max.   :8.400   Max.   :8.200   Max.   :9.900

   WartyClaim      OrdBilling       DelSpeed      Satisfaction
 Min.   :4.100   Min.   :2.000   Min.   :1.600   Min.   :4.700
 1st Qu.:5.400   1st Qu.:3.700   1st Qu.:3.400   1st Qu.:6.000
 Median :6.100   Median :4.400   Median :3.900   Median :7.050
 Mean   :6.043   Mean   :4.278   Mean   :3.886   Mean   :6.918
 3rd Qu.:6.600   3rd Qu.:4.800   3rd Qu.:4.425   3rd Qu.:7.625
 Max.   :8.100   Max.   :6.700   Max.   :5.500   Max.   :9.900
```

An interesting observation we can highlight in the 5 Number Summary is that all attributes the mean and median of ratings are almost equal. The chapter below discusses the univariate analysis of each attribute.

PRODUCT QUALITY

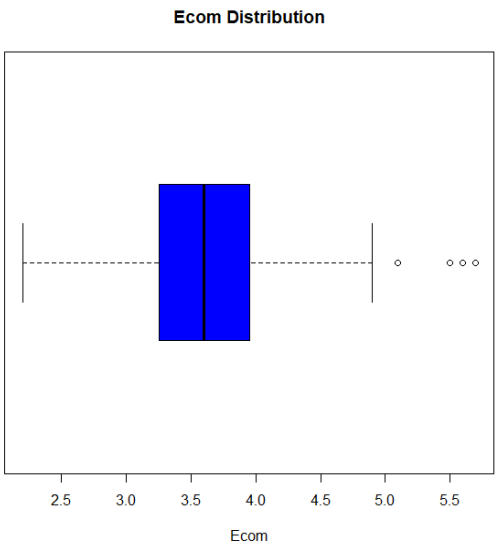
The range of rating values for the quality of the product 5 to 10. There are some customers feedback have reached a top ratings. However, mean is 7.8 rating, 25% of the customers have given ratings of 3.9 and 75% of the ratings are below 6.575. This rating indicates that popularity of product quality is slightly above average which require a serious improvement due to the fact quality of a product considered to be a key element of any successful business.

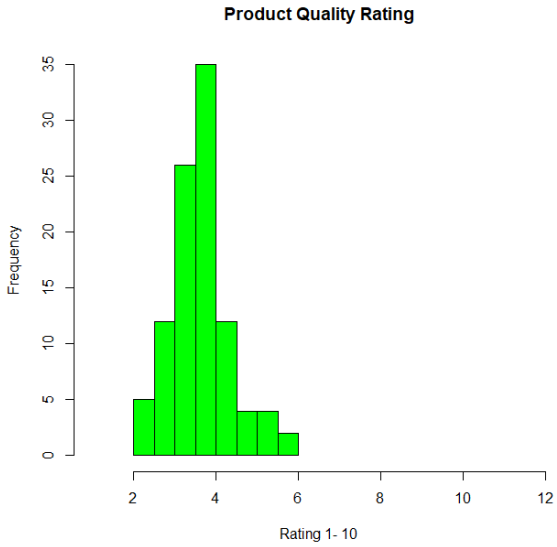




Ecommerce

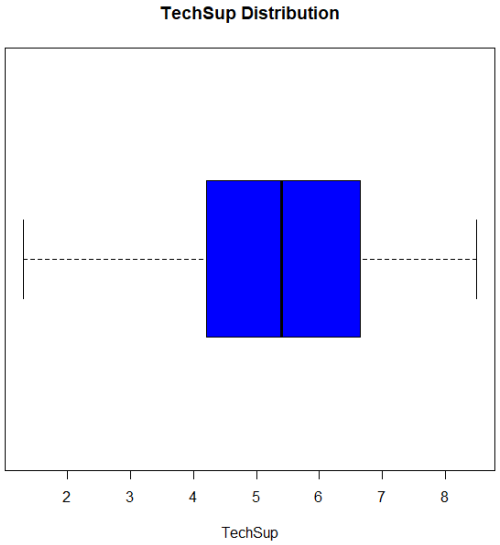
Range of ratings given for Ecom service lies in between 2.2 to 5.7. The mean of Ecom rating is about 3.672. the quantile analysis describes the 75% of the people have given ratings below 3.925 and 25% of the people have given 3.275 ratings. Box plot and R command for outliers analysis shows there are 7 outliers. That is 2.2, 5.5, 5.7, 5.1, 5.1, 5.1, and 5.5. This service must be inspected thoroughly in order to improve on the customer ratings while taking business decisions.

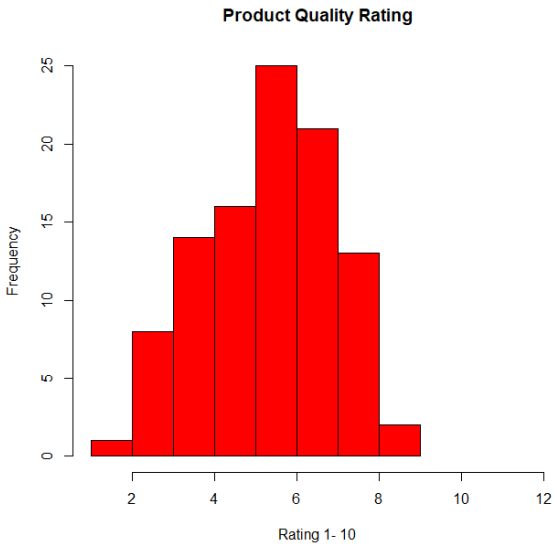




Technical Support

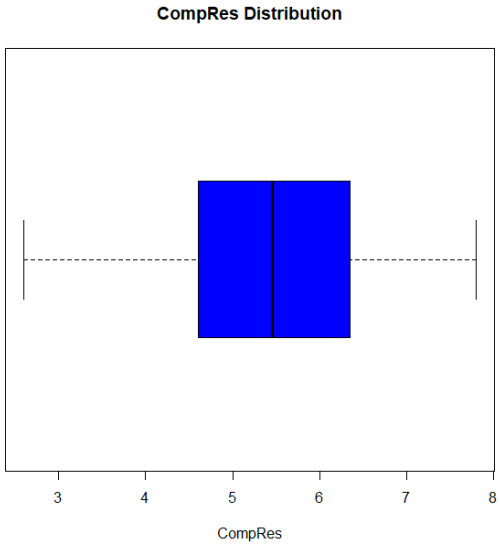
The ratings range is 1.3 to 8.5. The average rating for technical support is about 5.365. At least, 75% of the customers have found the Technical support to be 6.625 while 25% have found the technical support not very helpful and given a rating below 4.25.

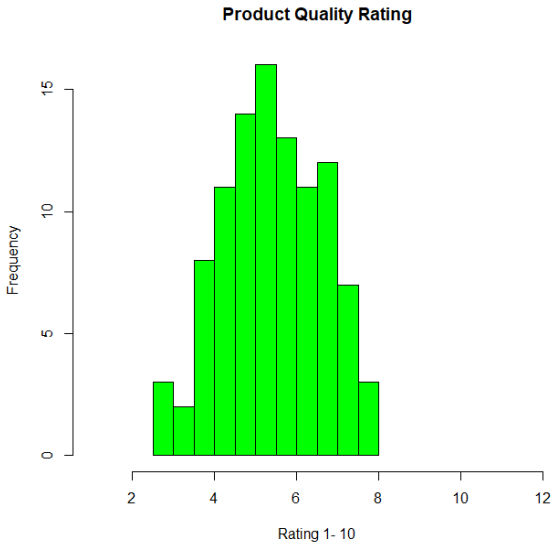




CompRes

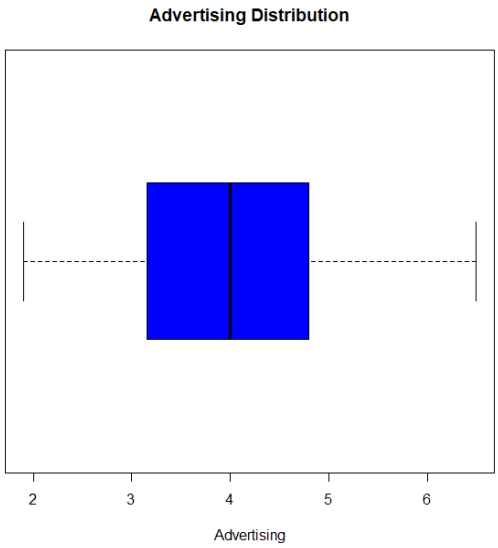
The CompRes rating range is from 2.6 to 7.8. Average rating for this particular attribute is about 5.442. 25% of the people have shown an enthusiasm of giving a rating of 4.6 and below while 75% of people rated below 6.325. There are no outliers which effects the decision.

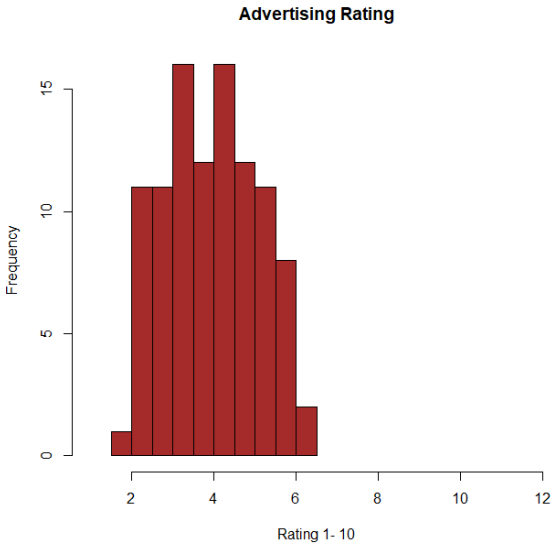




ADVERTISING

Advertising is an important tool in any given business. This would be the first point of interaction with the customer. Management must take a serious note on improving this attributes due to the power rating lies between 1.9 a very low level to 6.5 just on the average border. The average is about 4.01 rating. 25% people see the advertisement of the company to be not very attractive and rated below 3.175. 50% of the people have rated between 3.175 and 4.8. Management must come up with better strategy to attract more customer by improving the advertising tools.

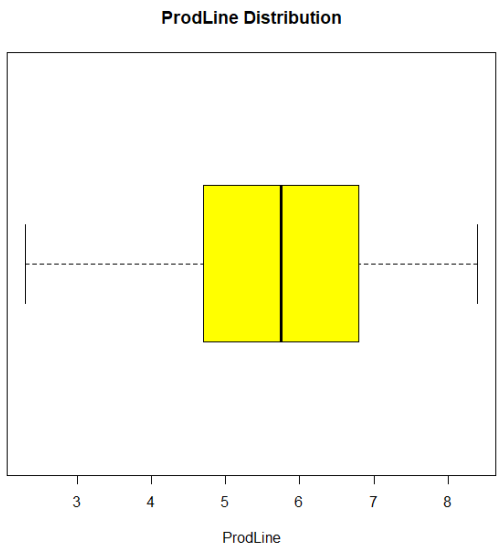


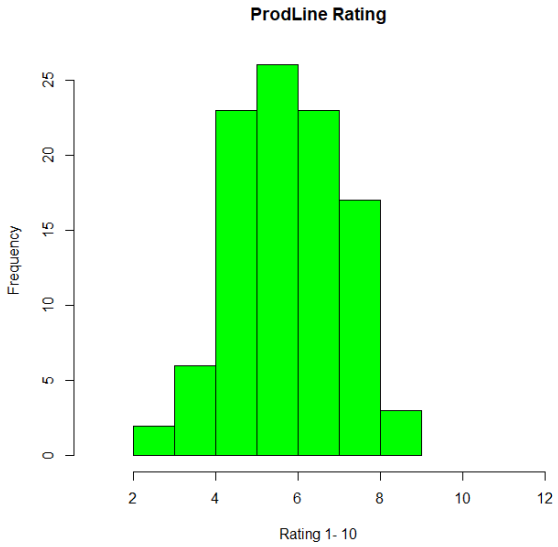


Product Line

The product line rating is from 2.3 to 8.4. The average rating for product Line is about 5.805.

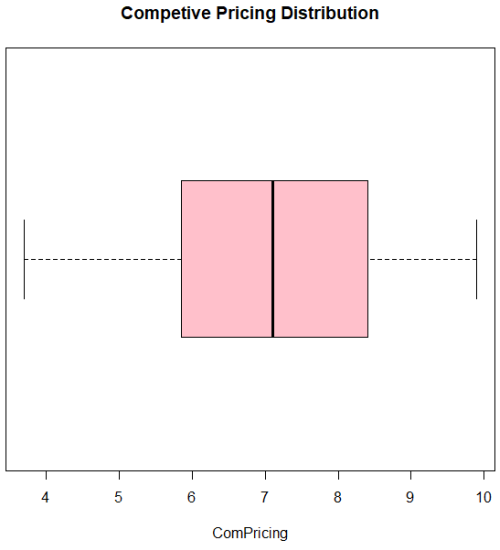
75% of the rating value is below 6.8 and 25% of the ratings were given below 4.7. No outliers in the rating distribution.

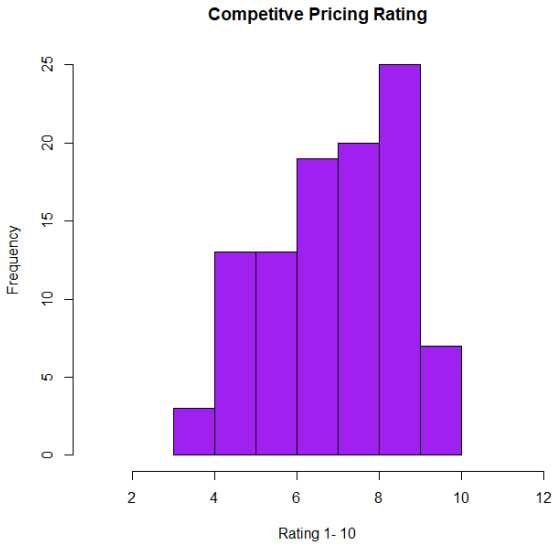




Competitve Pricing

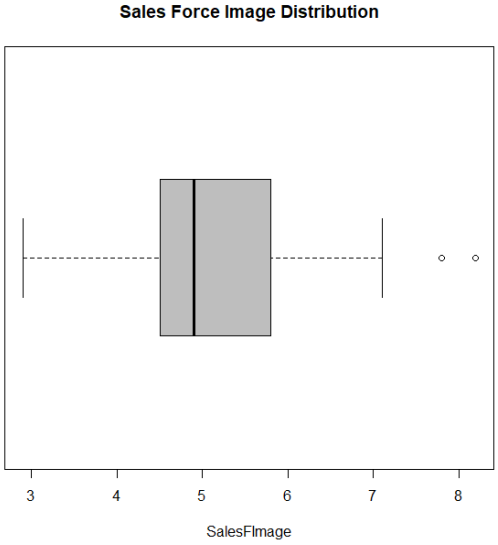
The Competitve pricing range is from 3.7 to 9.9. The first quanitle is 5.875 and the 3rd quantile is 8.4. Therfore the 50% of the population think the product pricing is reasonable which is a very positive feedback. The boxplot diagram illustrates a very balanced rating distribution on the competetive price variable too.

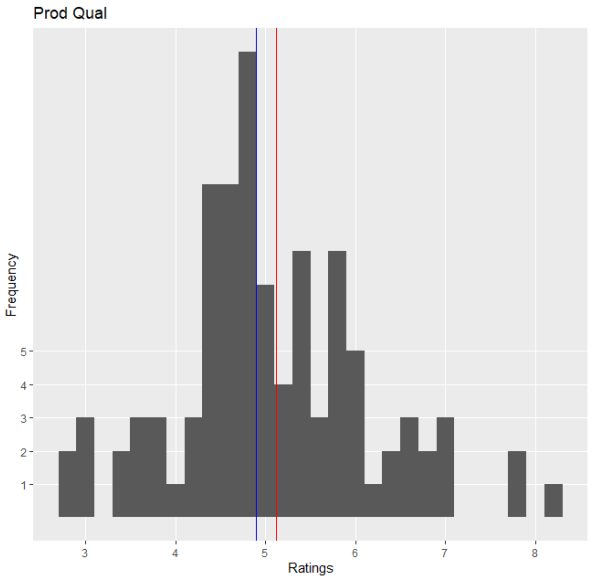




Sales Force Image

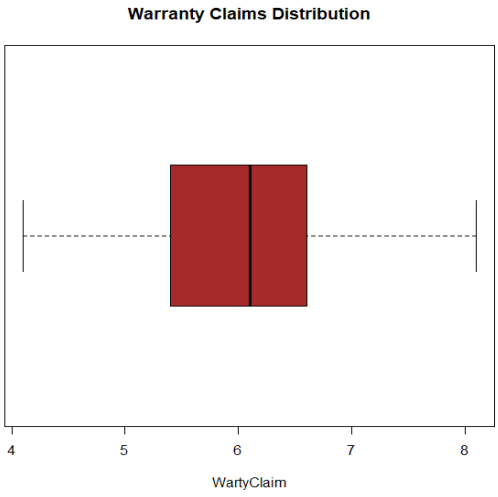
The range of Sales Force Image starts from 2.9 to 8.2. The mean of the Sales Force Image rating is about 5.123. 75% of the ratings are given below 5.8 rating which is just around the average ratings. This indication tells the management to give extra attention in training the sales force. There are 3 outliers could be identifeid in the rating distribution.(i.e: 7.8, 7.8, 8.2). From the histogram we can clearly say that distribution is negatively skewed.

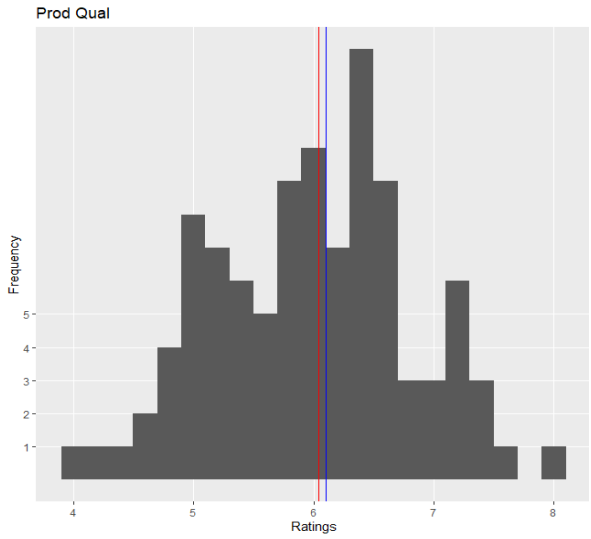




Warranty Claim

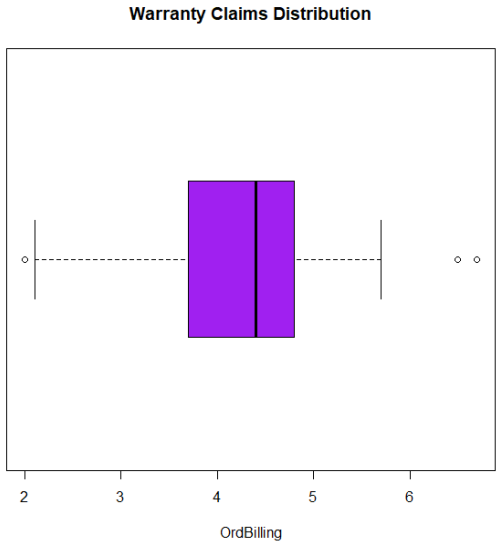
The rang of rating is 4.1 to 8.1 for Warranty and Claim service.There are no outliers to be identified. 75% of the rating is below 6.6 and the 25% have rated 5.4 below. The histogram show couple of peaks at the centre. The mean and median are almost equal, however, there are some flat bars could be observed.

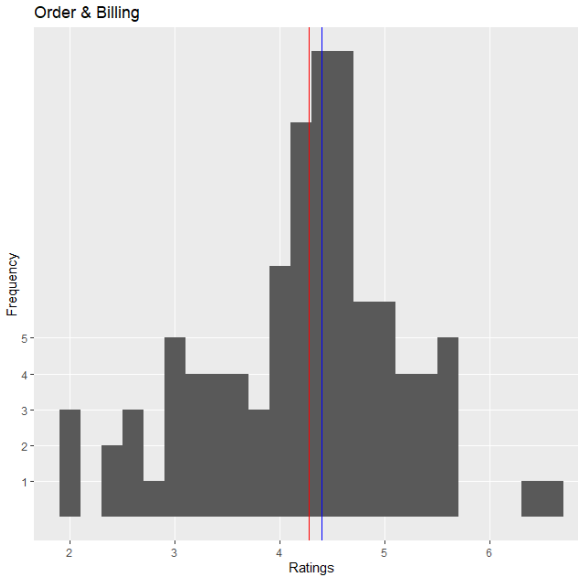




Order & Billing

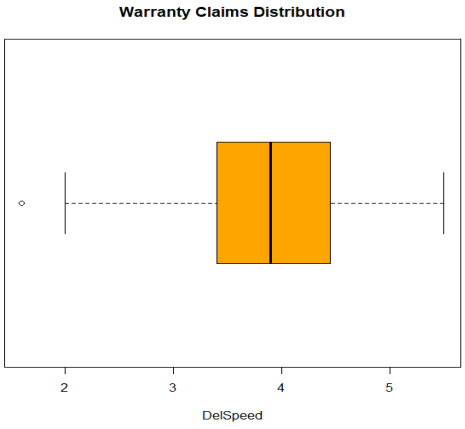
The order & billing rating range is from 2.0 to 6.7. The caluclated mean rating for this attribute is around the value 4.3. Evidently box plot shows the biased mean towards the 3rd quantile. 75% of the people have given their rating below 4.8. 25% have given rating of below 3.7.The upper and lower outliers are 6.7,6.5 and 2,2 respectively. The histogram is left skewed.

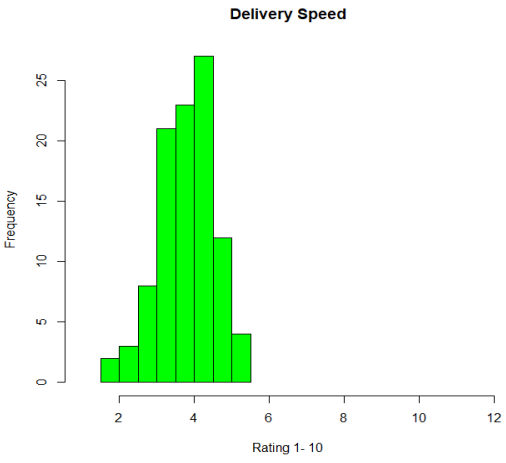




Delivery Speed

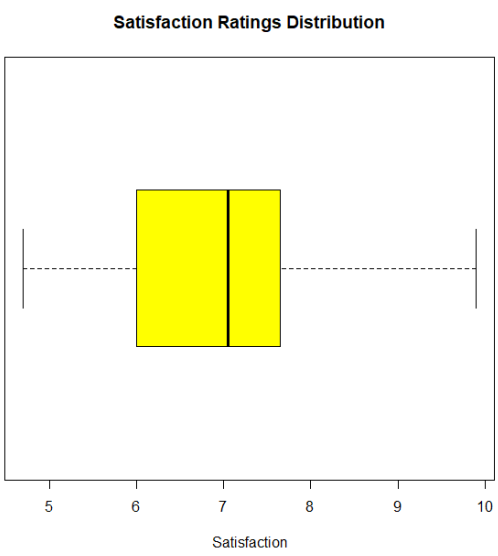
The range for delivery speed is 1.6 to 5.5. The box plot indicates an evidence of having one outlier. The average rating is about 3.9. 50% of the cutomers from the given population has rated the delivery speed to be 3.4 to 4.425. This outcome of the analysis must be taken into consideration and the delivery methods must be redesigned in such a way to gain customer trust. The outlier 1.83 can be seen.



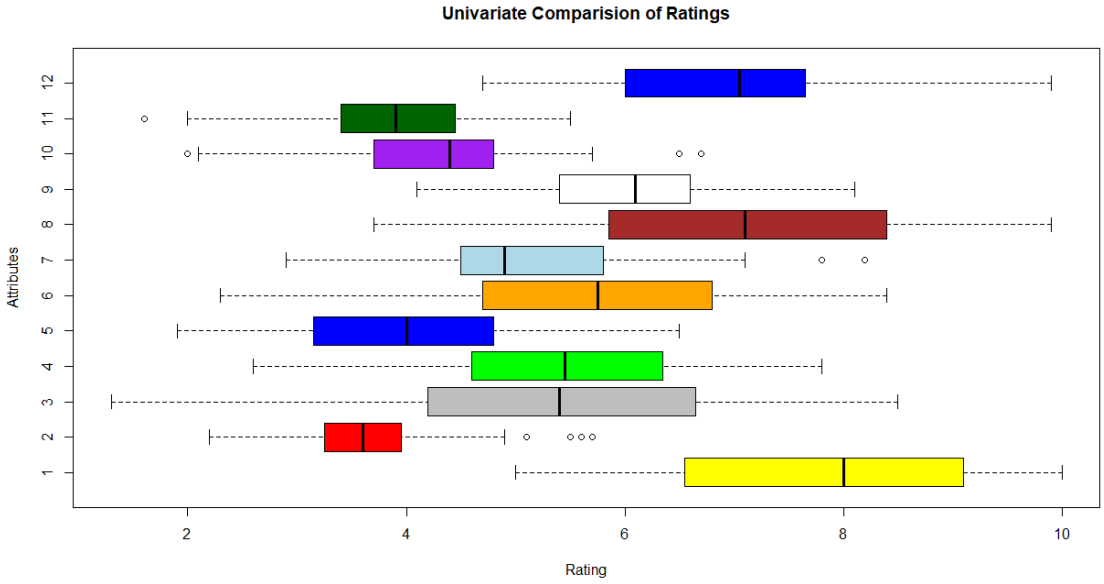


Satisfaction

The rating given by the customers for the satisfactory of the customer service is from 4.7 to 9.9.It can be said that the mean rating for the satisfaction attribute is 6.9 which is above average where average considered to be rating value 5. The 75% of the population have rated below the 7.6 and 25% have rated below 6. The histogram shows multiple peaks at the centre with equal heights which does hold any nature of normal distribution.

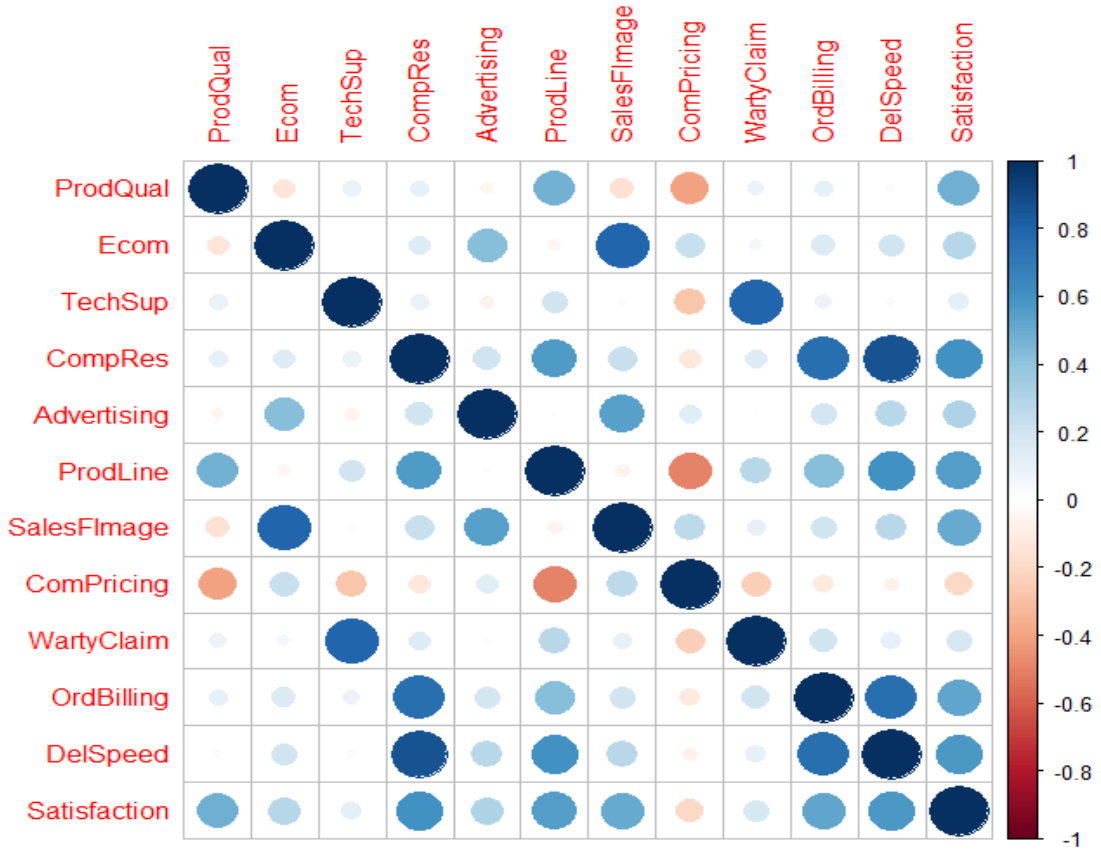


3.4 MULTIVARIATE ANALYSIS



3.5 MULTI-COLINEARITY ANALYSIS

As Multivariate analysis Corelation test been done among all variables. As shown in the bleow diagram and the table it can be seen that Circles with Dark Blue are highly positively correlated and red means high negative correlation. This approach will help to make sense out of the all variables. We can see that Competivie Pricing of the product is not very much correlated with any of the other 11 attributes. Furthermore, "ProdQual","Advertising" and "ProdLine" attributes also shares a similar uncorrelation with the other attributes.



		ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed	Satisfaction
2	ProdQual	1	-0.1372	0.0956	0.1064	-0.0535	0.4775	-0.1518	-0.4013	0.0883	0.1043	0.0277	0.4863
3	Ecom	-0.137163217	1.0000	0.0009	0.1402	0.4299	-0.0527	0.7915	0.2295	0.0519	0.1561	0.1916	0.2827
4	TechSup	0.095600454	0.0009	1.0000	0.0967	-0.0629	0.1926	0.0170	-0.2708	0.7972	0.0801	0.0254	0.1126
5	CompRes	0.106370001	0.1402	0.0967	1.0000	0.1969	0.5614	0.2298	-0.1280	0.1404	0.7569	0.8651	0.6033
6	Advertising	-0.053473134	0.4299	-0.0629	0.1969	1.0000	-0.0116	0.5422	0.1342	0.0108	0.1842	0.2759	0.3047
7	ProdLine	0.477493413	-0.0527	0.1926	0.5614	-0.0116	1.0000	-0.0613	-0.4949	0.2731	0.4244	0.6019	0.5505
8	SalesFImage	-0.151812874	0.7915	0.0170	0.2298	0.5422	-0.0613	1.0000	0.2646	0.1075	0.1951	0.2716	0.5002
9	ComPricing	-0.401281884	0.2295	-0.2708	-0.1280	0.1342	-0.4949	0.2646	1.0000	-0.2450	-0.1146	-0.0729	-0.2083
10	WartyClaim	0.088312306	0.0519	0.7972	0.1404	0.0108	0.2731	0.1075	-0.2450	1.0000	0.1971	0.1094	0.1775
11	OrdBilling	0.104303074	0.1561	0.0801	0.7569	0.1842	0.4244	0.1951	-0.1146	0.1971	1.0000	0.7510	0.5217
12	DelSpeed	0.027718003	0.1916	0.0254	0.8651	0.2759	0.6019	0.2716	-0.0729	0.1094	0.7510	1.0000	0.5770
13	Satisfaction	0.486324998	0.2827	0.1126	0.6033	0.3047	0.5505	0.5002	-0.2083	0.1775	0.5217	0.5770	1.0000

MULTI-COLINEARITY BETWEEN SATISFACTIONA & OTHER ATTRIBUTES.

I have run t and thehe Varriance Inflation Factor using a "vif" function in R and the result is tabulated below. Since the range of rating is very small, I have made an assumption of considering the values which is more than 3.Therfore all the variables which has ore than 3 vif values assumed to be well correlated. For those attribute Simple Linear Regression technique will be applied. Detail SLM analysis will be discussed in upcoming Chapters.

Attributes are:- CompRes,ProdLine.SalesFImage,WartyClaim & DelSpeed

ATTRIBUTES	VARRIANCE INFLATION FACTOR
Prod_Qual	1.635797
Ecom	2.756694
TechSup	2.976796
CompRes	4.730448
Advertising	1.508933
ProdLine	3.488185
SalesFImage	3.439420
ComPricing	1.635000
WartyClaim	3.198337
OrdBilling	2.902999
DelSpeed	6.516014

3.5 SIMPLE LINEAR REGRESSION ANALYSIS

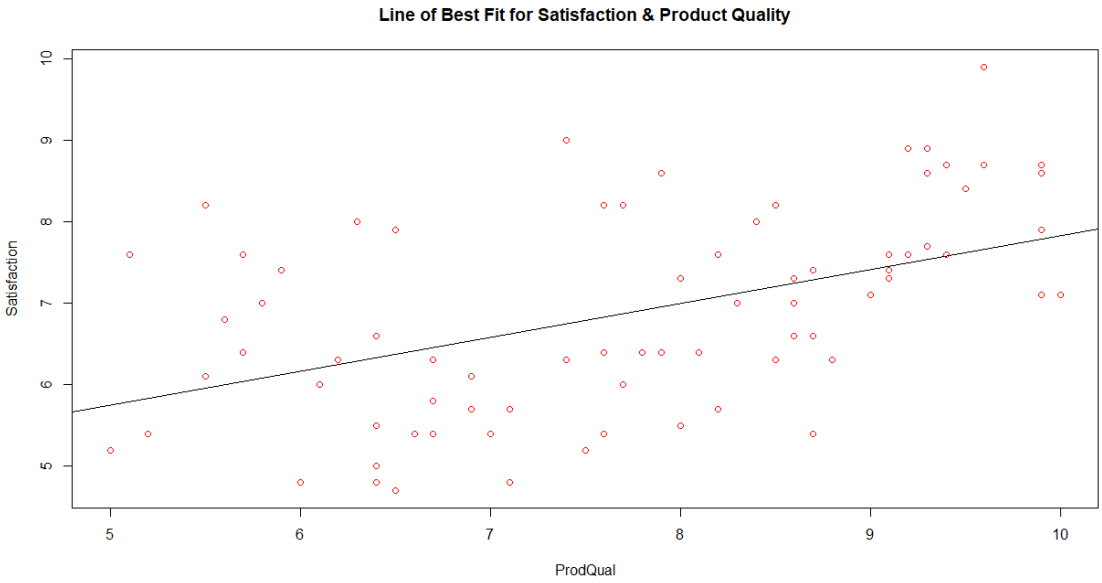
The linear regression model is defined on 4 important assumptions, often referred to as (***LINE***)
Assumption 1: The regression model is ***linear*** in the parameters i.e. $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$,
Assumption 2: The observations Y_i (or the error terms ϵ_i) are ***independent***
Assumption 3: The error variables ϵ_i are ***normally*** distributed.
Assumption 4: The errors have no bias (that is, $E(\epsilon_i)=0$) and they are *homoscedastic*, that is, they have ***equal variance***.

Dependent Variable = Satisfaction

Independent Variable = Other 11 attributes.

Satisfaction & Product Quality

The below plot illustrates that Satisfaction and product quality attributes are 49% correlated and the line of best fit drawn using ordinary least square method shown below shows the intensity of the correlation nature graphically. It can be said that the relationship is non-linear and fitted line does not follow the pattern observed at all points. So this graphical representation is not sufficient enough to come to a conclusion about the colinearity.



Inorder to investigate further on the relationship line linear regression model analysis is analysed. It can be seen that R^2 gives a pretty low score of 0.23. It means linear model we just fit in the data is explaining a mere 23% of the variance observed in the data.

About the residual error of this particular combinations of attributes can be interpreted as for any prediction made based on the Product Quality will show a deviation on the Satisfaction Rating approximately 1.047. Well, the Residual Standard Error for Satisfaction is 1.047 while the mean rating for satisfaction variable is about 6.92. We can assume that percentage error for any given point is around 85% which is a very high square

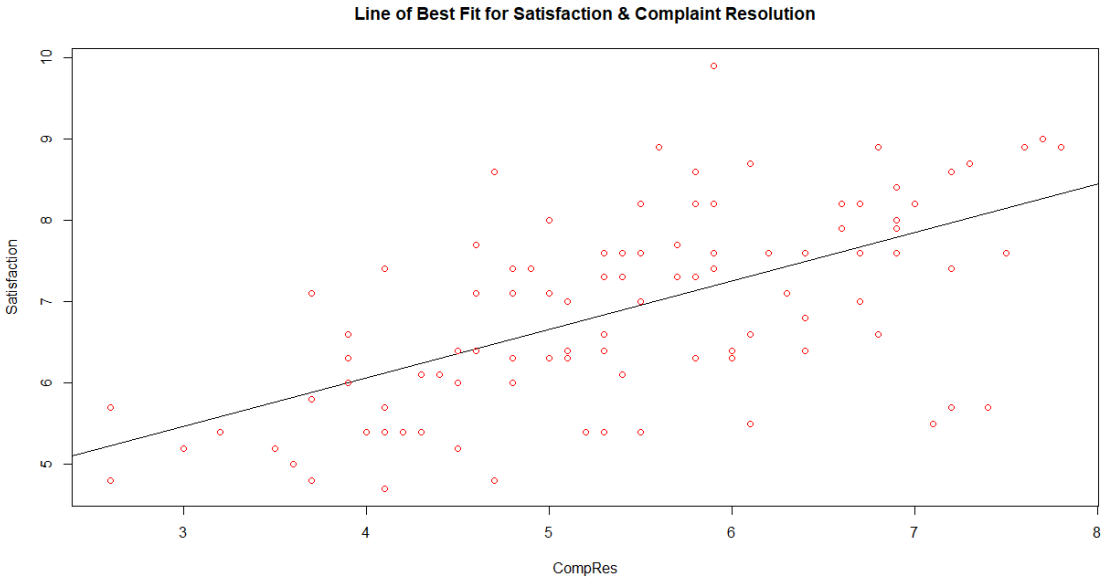
Satisfaction_Rating = 3.67 + 0.41512Product_Qual_Rating.

```
 Call:
lm(formula = Satisfaction ~ ProdQual, data = Customer_Rating)

Residuals:
     Min       1Q   Median       3Q      Max
-1.88746 -0.72711 -0.01577  0.85641  2.25220
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.67593    0.59765   6.151 1.68e-08 ***
ProdQual     0.41512    0.07534   5.510 2.90e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.047 on 98 degrees of freedom
Multiple R-squared:  0.2365,    Adjusted R-squared:  0.2287
F-statistic: 30.36 on 1 and 98 DF,  p-value: 2.901e-07
```

Satisfaction & Complaint Resolution

The below plot illustrates that Satisfaction and Complaint Resolution attributes are 60% correlated and the line of best fit drawn using ordinary least square method shown below shows the intensity of the correlation nature graphically. It can be said that the relationship is non-linear and fitted line does not follow the pattern observed at all points. So this graphical representation is not sufficient enough to come to a conclusion about the colinearity.



Inorder to investigate further on the relationship line linear regression model analysis is analysed. It can be seen that R^2 gives a pretty low score of 0.37. It means linear model we just fit in the data is explaining a mere 37% of the variance observed in the data.

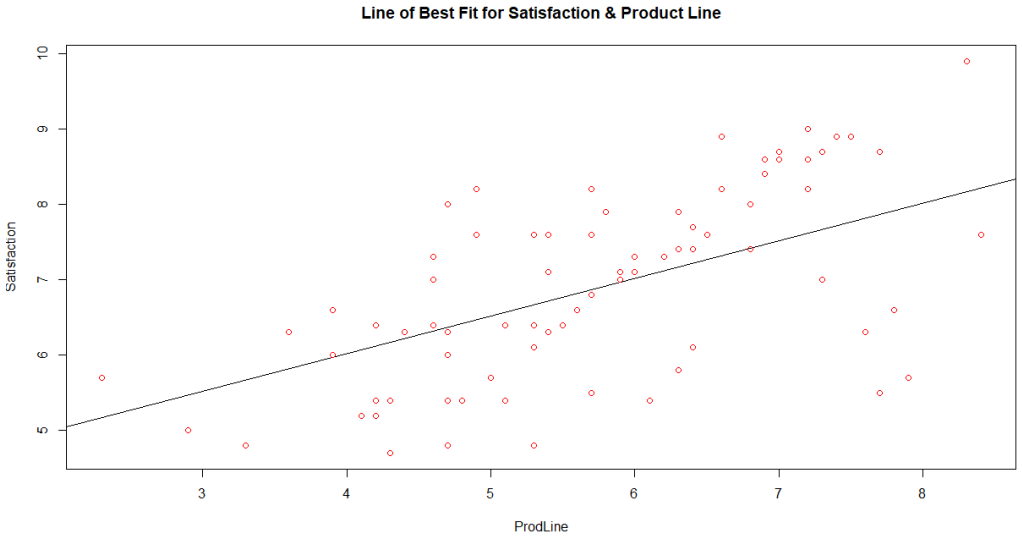
About the residual error of this particular combinations of attributes can be interpreted as for any prediction made based on the Complaint Resolution will show a deviation on the Satisfaction

Rating approximately 0.9554. Well, the Residual Standard Error for Satisfaction is 0.9554 while the mean rating for satisfaction variable is about 6.92. We can assume that percentage error for any given point is around 85% which is a very high square

Satisfaction_Rating = 3.68 + 0.59499ComplaintResolution_Rating.

```
Call:
lm(formula = Satisfaction ~ CompRes, data = Customer_Rating)

Residuals:
     Min       1Q   Median       3Q      Max
-2.40450 -0.66164  0.04499  0.63037  2.70949

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.68005    0.44285   8.310 5.51e-13 ***
CompRes      0.59499    0.07946   7.488 3.09e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9554 on 98 degrees of freedom
Multiple R-squared:  0.3639,    Adjusted R-squared:  0.3574
F-statistic: 56.07 on 1 and 98 DF,  p-value: 3.085e-11
```

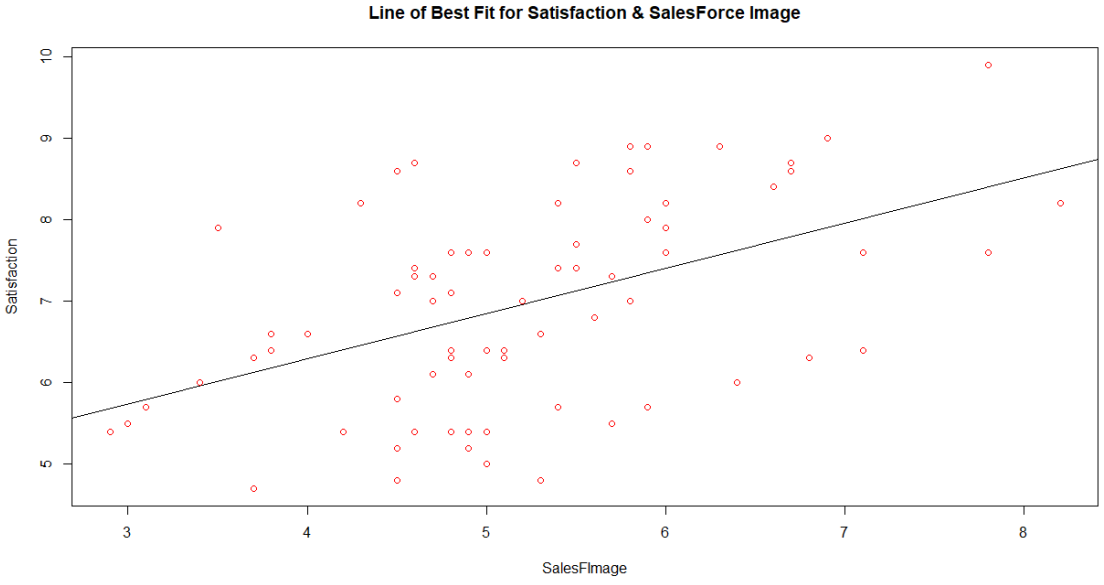
Satisfaction & Product Line



```
Call:
lm(formula = Satisfaction ~ ProdLine, data = Customer_Rating)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3634 -0.7795  0.1097  0.7604  1.7373

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.02203    0.45471   8.845 3.87e-14 ***
ProdLine     0.49887    0.07641   6.529 2.95e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1 on 98 degrees of freedom
Multiple R-squared:  0.3031,    Adjusted R-squared:  0.296
F-statistic: 42.62 on 1 and 98 DF,  p-value: 2.953e-09
```

Satisfaction_Rating = 4.02 + 0.49887ProductLine_Rating.

Satisfaction & Product Line



```
Call:
lm(formula = Satisfaction ~ SalesFImage, data = Customer_Rating)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2164 -0.5884  0.1838  0.6922  2.0728

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.06983    0.50874   8.000 2.54e-12 ***
SalesFImage  0.55596    0.09722   5.719 1.16e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.037 on 98 degrees of freedom
Multiple R-squared:  0.2502,    Adjusted R-squared:  0.2426
F-statistic:  32.7 on 1 and 98 DF,  p-value: 1.164e-07
```

Satisfaction_Rating = 4.069 + 0.55596 SalesFImage_Rating.

3.6 PCA /FA ANALYSIS

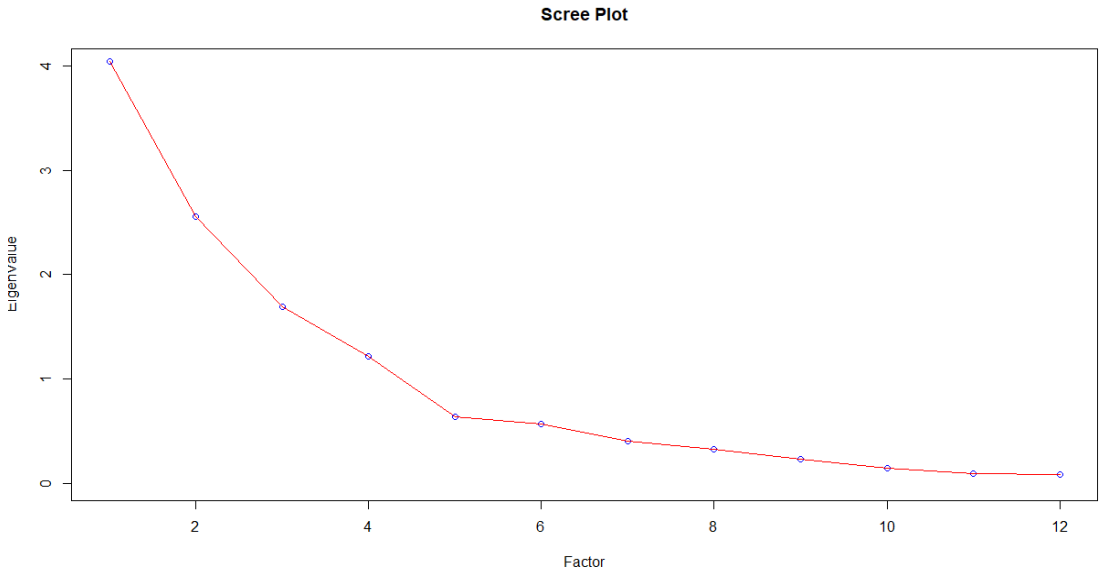
The eigen value coputation is being done and the results are given below.

4.04285991, 2.55292440, 1.69222417, 1.21754639, 0.63596293, 0.56853132, 0.40282774,

0.32448016, 0.23613948, 0.4422355, 0.09913845 and 0.08314143

This eigen vector is nothing but the principal component. Same eigen values can be obtained using square of each attributes eigen vectors.

Scree plot was drawn using the obtained 12 eigen values.Ac



According to the Kaizer Rule number of factors been selected for all eigen values which are more than 1. In this case we can obtain **4 Factors.**

Next by applying the principal component analysis conducted without rotating the matrix and the output is displayed below.

```
Call: principal(r = Customer_Rating, nfactors = 4, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
                PC1    PC2    PC3    PC4    h2     u2  com
ProdQual      0.319 -0.500 -0.096  0.678 0.820 0.1798 2.37
Ecom          0.334  0.704  0.308  0.217 0.749 0.2512 2.08
TechSup       0.252 -0.381  0.802 -0.198 0.890 0.1098 1.80
CompRes       0.850  0.002 -0.256 -0.309 0.883 0.1172 1.46
Advertising   0.363  0.571  0.117  0.227 0.523 0.4768 2.16
ProdLine      0.709 -0.476 -0.145  0.110 0.763 0.2372 1.91
SalesFImage   0.438  0.743  0.313  0.220 0.890 0.1095 2.24
ComPricing   -0.271  0.667 -0.067 -0.266 0.594 0.4058 1.69
WartyClaim    0.352 -0.321  0.788 -0.209 0.891 0.1088 1.92
OrdBilling    0.780  0.014 -0.202 -0.338 0.764 0.2360 1.52
DelSpeed      0.849  0.087 -0.284 -0.320 0.911 0.0886 1.55
Satisfaction  0.830  0.038 -0.037  0.365 0.826 0.1739 1.38

                       PC1   PC2   PC3   PC4
SS loadings          4.043 2.553 1.692 1.218
Proportion Var       0.337 0.213 0.141 0.101
Cumulative Var       0.337 0.550 0.691 0.792
Proportion Explained 0.425 0.269 0.178 0.128
Cumulative Proportion 0.425 0.694 0.872 1.000

Mean item complexity =  1.8
Test of the hypothesis that 4 components are sufficient.
```

Interpreting the above output it can be seen that amalgamation effect is influencing our interpretation. In order to overcome this particular effect I will implying VarimaxOrthogonal Rotation to interpret futher.

```
C:/Users/user/Desktop/Data Science/FUNDAMENTAL OF BUSINESS STATISTICS/PRINCIPAL COMPONENT - PROJECT/
Standardized loadings (pattern matrix) based upon correlation matrix
                 RC1     RC2     RC4     RC3     h2      u2    com
ProdQual      -0.0065 -0.0259  0.9048 -0.0287 0.8202 0.17978 1.004
Ecom           0.0506  0.8547 -0.1155  0.0488 0.7488 0.25119 1.050
TechSup        0.0182 -0.0186  0.0955  0.9383 0.8902 0.10976 1.022
CompRes        0.9246  0.1205  0.1058  0.0472 0.8828 0.11717 1.066
Advertising    0.1399  0.7062 -0.0107 -0.0699 0.5232 0.47675 1.099
ProdLine       0.5882 -0.0991  0.6176  0.1602 0.7628 0.23717 2.188
SalesFImage    0.1311  0.9275 -0.0952  0.0630 0.8905 0.10952 1.071
ComPricing    -0.0883  0.2841 -0.6576 -0.2707 0.5942 0.40583 1.775
WartyClaim     0.1089  0.0563  0.0937  0.9314 0.8912 0.10876 1.055
OrdBilling     0.8620  0.1091  0.0467  0.0831 0.7640 0.23603 1.057
DelSpeed       0.9375  0.1723  0.0532 -0.0026 0.9114 0.08856 1.074
Satisfaction   0.5223  0.4788  0.5678  0.0397 0.8261 0.17393 2.955

                       RC1    RC2    RC4    RC3
SS loadings          3.1549 2.4695 2.0115 1.8696
Proportion Var       0.2629 0.2058 0.1676 0.1558
Cumulative Var       0.2629 0.4687 0.6363 0.7921
Proportion Explained 0.3319 0.2598 0.2116 0.1967
Cumulative Proportion 0.3319 0.5917 0.8033 1.0000
```

	RC1	RC2	RC3	RC4
Prod_Qual	-0.0065	-0.0259	-0.0287	0.9048
Ecom	0.0506	0.8547	0.0488	-0.1155
TechSup	0.182	0.0186	0.9383	0.0955
CompRes	0.9246	0.1205	0.0472	0.1058
Advertising	0.1399	0.7062	-0.0699	-0.0107
ProdLine	0.5882	-0.0991	0.1602	0.6176
SalesFImage	0.1311	0.9275	0.0630	-0.0952
ComPricing	-0.0883	0.2841	-0.2707	-0.6576
WartyClaim	0.1089	0.0563	0.9314	0.0937
OrdBilling	0.8620	0.1091	0.0831	0.0467
DelSpeed	0.9375	0.1723	-0.0026	0.0532
Satisfaction	0.5223	0.4788	0.0397	0.5678

According to the above loadig factor we could ammalgmate the 12 attributes into the 4 different catergories as highligted

LOADINGS	VARIABLES
PRE-SALE SERVICE	OrdBilling
	DelSpeed
BRAND REPUTATION	Ecom
	Advertising
	SalesFImage
POST SALES SERVICE	TechSup
	WartyClaim
PRODUCT QUALITY	Prod_Qual
	CompRes
	ProdLine
	ComPricing
	Satisfaction

3.7 MULTIPLE LINEAR REGRESSION

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.43005 -0.31165  0.07621  0.37190  0.90120

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.66961    0.81233  -0.824  0.41199
ProdQual     0.37137    0.05177   7.173 2.18e-10 ***
Ecom        -0.44056    0.13396  -3.289  0.00145 **
TechSup      0.03299    0.06372   0.518  0.60591
CompRes      0.16703    0.10173   1.642  0.10416
Advertising -0.02602    0.06161  -0.422  0.67382
ProdLine     0.14034    0.08025   1.749  0.08384 .
SalesFImage  0.80611    0.09775   8.247 1.45e-12 ***
ComPricing  -0.03853    0.04677  -0.824  0.41235
WartyClaim  -0.10298    0.12330  -0.835  0.40587
OrdBilling   0.14635    0.10367   1.412  0.16160
DelSpeed     0.16570    0.19644   0.844  0.40124
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

APPENDIX

```
#Name : Karthik

#Project: Project 3 Product Service Management.


#-------------------------------------------------------------------------------


setwd("C:/Users/user/Desktop/Data Science/FUNDAMENTAL OF BUSINESS STATISTICS/PRINCIPAL COMPONENT - PROJECT/")

getwd()


Customer_Feedback = read.csv("Factor-Hair-Revised.csv")


#-------------------------------------------------------------------------------


#Installing Library PackProdQual to Read & Write .csv files

install.packProdQuals("readr")

library(readr)

#Installing Library PackProdQual to work with plots

install.packProdQuals("ggplot2")

library(ggplot2)

#Calling Library PackProdQuals to work with Table

library(rpivotTable)

#Calling Library PackProdQuals to Work with data manipulation

library(dplyr)

#Calling Library PackProdQuals to Work with Correlation

library(corrplot)

#Calling Library For Correlation Graph Analysis

library(PerformanceAnalytics)

#Calling Library For Exporting to .csv
```

```
library(rio)
#Calling Library to do Varriance Inflation Factor Analysis
library("car")

library(nFactors)
library(psych)
#------------------------------------------------------------------------------
#SANITY CHECK
#------------------------------------------------------------------------------

str(Customer_Feedback)
summary(Customer_Feedback)

#Dimentions
dim(Customer_Feedback)
## MISSING VALUE IDENTIFICATION
anyNA(Customer_Feedback)      # Outputs if any Missing values were spotted
sum(is.na(Customer_Feedback)) # Number of Missing Values

# ANALYSING FORMATTING ISSUES
head(Customer_Feedback,10)
tail(Customer_Feedback,10)



# Removing Unwanted Coloumns

Customer_Rating = -select(Customer_Feedback,-ID)
Customer_Rating = Customer_Rating*-1
View(Customer_Rating)
```

```
summary(Customer_Rating)

#UNIVARIATE ANALYSIS
#Product Quality
names(Customer_Rating)
range(Customer_Rating$ProdQual)

boxplot(Customer_Rating$ProdQual,horizontal = TRUE,col="Blue",
      main = "ProdQual Distribution", xlab= "ProdQual")

# OUTLIER CALCULATION
mean(Customer_Rating$ProdQual)
median(Customer_Rating$ProdQual)
ProdQual_IQR = IQR(Customer_Rating$ProdQual)
ProdQual_IQR

#OUTLIERS - Q3
Q3_ProdQual = quantile(Customer_Rating$ProdQual,0.75)
Q3_ProdQual = unname(Q3)
Q3_ProdQual

Q1_ProdQual = unname(quantile(Customer_Rating$ProdQual,0.5))
Q1_ProdQual

#OUTLIERS - Q1
Q1_ProdQual = unname(quantile(Customer_Rating$ProdQual,0.25))
Q1_ProdQual
Max_ProdQual_IQR = Q3_ProdQual+1.5*ProdQual_IQR # Positive Outlier Calculation
Max_ProdQual_IQR
```

```
Min_ProdQual_IQR = Q1_ProdQual-1.5*ProdQual_IQR # Negative Outlier Calculation
Min_ProdQual_IQR
# No Of OUTLIERS
sum(Customer_Rating$ProdQual > Max_ProdQual_IQR)
sum(Customer_Rating$ProdQual < Min_ProdQual_IQR)

# Print all Outliers
Customer_Rating$ProdQual[Customer_Rating$ProdQual > Max_ProdQual_IQR]
Customer_Rating$ProdQual[Customer_Rating$ProdQual < Min_ProdQual_IQR]
# Histogram

hist(Customer_Rating$ProdQual,col="Blue",xlim=c(1,12),
    main="Product Quality Rating",xlab="Rating 1- 10")


qplot(Customer_Rating$ProdQual,data=Customer_Rating,geom ="histogram",binwidth=1)+
  labs(title = "Prod Qual")+labs(x="Ratings")+labs(y="Frequency")+
  scale_y_continuous( breaks=c(1:20),minor_breaks=NULL)+
  scale_x_continuous( breaks=c(1:10),minor_breaks=NULL)+
  geom_vline(xintercept = mean(Customer_Rating$ProdQual),show.legend=TRUE,col="red")+
  geom_vline(yintercept=median(Customer_Rating$ProdQual),show.legend=TRUE,col="blue")
#=============================================================================
qplot(Customer_Rating$ProdQual,data=Customer_Rating,geom ="histogram",binwidth=0.2)+
  labs(title = "Prod Qual")+labs(x="Ratings")+labs(y="Frequency")+
  scale_y_continuous( breaks=c(1:5),minor_breaks=NULL)+
  scale_x_continuous( breaks=c(1:10),minor_breaks=NULL)+
  geom_vline(xintercept = mean(Customer_Rating$ProdQual),show.legend=TRUE,col="red")+
  geom_vline(xintercept=median(Customer_Rating$ProdQual),show.legend=TRUE,col="blue")
#=============================================================================
```

```
#UNIVARIATE of Ecom
range(Customer_Rating$Ecom)

boxplot(Customer_Rating$Ecom,horizontal = TRUE,col="Blue",
      main = "Ecom Distribution", xlab= "Ecom")
# OUTLIER CALCULATION
mean(Customer_Rating$Ecom)
median(Customer_Rating$Ecom)
Ecom_IQR = IQR(Customer_Rating$Ecom)
Ecom_IQR
#OUTLIERS - Q3
Q3_Ecom = quantile(Customer_Rating$Ecom,0.75)
Q3_Ecom = unname(Q3_Ecom)
Q3_Ecom
#OUTLIERS - Q1
Q1_Ecom = unname(quantile(Customer_Rating$Ecom,0.25))
Q1_Ecom
Max_Ecom_IQR = Q3_Ecom+1.5*Ecom_IQR # Positive Outlier Calculation
Max_Ecom_IQR
Min_Ecom_IQR = Q1-1.5*Ecom_IQR # Negative Outlier Calculation
Min_Ecom_IQR
# No Of OUTLIERS
sum(Customer_Rating$Ecom > Max_Ecom_IQR)
sum(Customer_Rating$Ecom < Min_Ecom_IQR)

# Print all Outliers
Customer_Rating$Ecom[Customer_Rating$Ecom > Max_Ecom_IQR]
Customer_Rating$Ecom[Customer_Rating$Ecom < Min_Ecom_IQR]
```

```
# Histogram

hist(Customer_Rating$Ecom,col="Green",xlim=c(1,12),
    main="Product Quality Rating",xlab="Rating 1- 10")

#=============================================================================

#UNIVARIATE of TechSup
names(Customer_Rating)
range(Customer_Rating$TechSup)

boxplot(Customer_Rating$TechSup,horizontal = TRUE,col="Blue",
      main = "TechSup Distribution", xlab= "TechSup")
# OUTLIER CALCULATION
mean(Customer_Rating$TechSup)
median(Customer_Rating$TechSup)
TechSup_IQR = IQR(Customer_Rating$TechSup)
TechSup_IQR
#OUTLIERS - Q3
Q3_TechSup = quantile(Customer_Rating$TechSup,0.75)
Q3_TechSup = unname(Q3_TechSup)
Q3_TechSup
#OUTLIERS - Q1
Q1_TechSup = unname(quantile(Customer_Rating$TechSup,0.25))
Q1_TechSup
Max_TechSup_IQR = Q3_TechSup+1.5*TechSup_IQR # Positive Outlier Calculation
Max_TechSup_IQR
Min_TechSup_IQR = Q1-1.5*TechSup_IQR # Negative Outlier Calculation
Min_TechSup_IQR
```

```
# No Of OUTLIERS
sum(Customer_Rating$TechSup > Max_TechSup_IQR)
sum(Customer_Rating$TechSup < Min_TechSup_IQR)

# Print all Outliers
Customer_Rating$TechSup[Customer_Rating$TechSup > Max_TechSup_IQR]
Customer_Rating$TechSup[Customer_Rating$TechSup < Min_TechSup_IQR]
# Histogram

hist(Customer_Rating$TechSup,col="red",xlim=c(1,12),
    main="Product Quality Rating",xlab="Rating 1- 10")
names(Customer_Rating)

#================================================================================

#UNIVARIATE of CompRes
names(Customer_Rating)
range(Customer_Rating$CompRes)

boxplot(Customer_Rating$CompRes,horizontal = TRUE,col="Blue",
       main = "CompRes Distribution", xlab= "CompRes")
# OUTLIER CALCULATION
mean(Customer_Rating$CompRes)
median(Customer_Rating$CompRes)
CompRes_IQR = IQR(Customer_Rating$CompRes)
CompRes_IQR

#OUTLIERS - Q3
Q3_CompRes = quantile(Customer_Rating$CompRes,0.75)
```

```
Q3_CompRes = unname(Q3_CompRes)
Q3_CompRes
#OUTLIERS - Q1
Q1_CompRes = unname(quantile(Customer_Rating$CompRes,0.25))
Q1_CompRes
Max_CompRes_IQR = Q3_CompRes+1.5*CompRes_IQR # Positive Outlier Calculation
Max_CompRes_IQR
Min_CompRes_IQR = Q1-1.5*CompRes_IQR # Negative Outlier Calculation
Min_CompRes_IQR
# No Of OUTLIERS
sum(Customer_Rating$CompRes > Max_CompRes_IQR)
sum(Customer_Rating$CompRes < Min_CompRes_IQR)

# Print all Outliers
Customer_Rating$CompRes[Customer_Rating$CompRes > Max_CompRes_IQR]
Customer_Rating$CompRes[Customer_Rating$CompRes < Min_CompRes_IQR]
# Histogram

hist(Customer_Rating$CompRes,col="Green",xlim=c(1,12),
    main="Product Quality Rating",xlab="Rating 1- 10")

#=============================================================================

#UNIVARIATE of Advertising
names(Customer_Rating)
range(Customer_Rating$Advertising)

boxplot(Customer_Rating$Advertising,horizontal = TRUE,col="Blue",
       main = "Advertising Distribution", xlab= "Advertising")
```

```
# OUTLIER CALCULATION
mean(Customer_Rating$Advertising)
median(Customer_Rating$Advertising)
Advertising_IQR = IQR(Customer_Rating$Advertising)
Advertising_IQR
#OUTLIERS - Q3
Q3_Advertising = quantile(Customer_Rating$Advertising,0.75)
Q3_Advertising = unname(Q3_Advertising)
Q3_Advertising
#OUTLIERS - Q1
Q1_Advertising = unname(quantile(Customer_Rating$Advertising,0.25))
Q1_Advertising
Max_Advertising_IQR = Q3_Advertising+1.5*Advertising_IQR # Positive Outlier Calculation
Max_Advertising_IQR
Min_Advertising_IQR = Q1-1.5*Advertising_IQR # Negative Outlier Calculation
Min_Advertising_IQR
# No Of OUTLIERS
sum(Customer_Rating$Advertising > Max_Advertising_IQR)
sum(Customer_Rating$Advertising < Min_Advertising_IQR)

# Print all Outliers
Customer_Rating$Advertising[Customer_Rating$Advertising > Max_Advertising_IQR]
Customer_Rating$Advertising[Customer_Rating$Advertising < Min_Advertising_IQR]

# Histogram

hist(Customer_Rating$Advertising,col="brown",xlim=c(1,12),
    main="Advertising Rating",xlab="Rating 1- 10")
#================================================================================
```

```
#UNIVARIATE of ProdLine
names(Customer_Rating)
range(Customer_Rating$ProdLine)

boxplot(Customer_Rating$ProdLine,horizontal = TRUE,col="Yellow",
      main = "ProdLine Distribution", xlab= "ProdLine")
# OUTLIER CALCULATION
mean(Customer_Rating$ProdLine)
median(Customer_Rating$ProdLine)
ProdLine_IQR = IQR(Customer_Rating$ProdLine)
ProdLine_IQR
#OUTLIERS - Q3
Q3_ProdLine = quantile(Customer_Rating$ProdLine,0.75)
Q3_ProdLine = unname(Q3_ProdLine)
Q3_ProdLine
#OUTLIERS - Q1
Q1_ProdLine = unname(quantile(Customer_Rating$ProdLine,0.25))
Q1_ProdLine
Max_ProdLine_IQR = Q3_ProdLine+1.5*ProdLine_IQR # Positive Outlier Calculation
Max_ProdLine_IQR
Min_ProdLine_IQR = Q1-1.5*ProdLine_IQR # Negative Outlier Calculation
Min_ProdLine_IQR
# No Of OUTLIERS
sum(Customer_Rating$ProdLine > Max_ProdLine_IQR)
sum(Customer_Rating$ProdLine < Min_ProdLine_IQR)

# Print all Outliers
Customer_Rating$ProdLine[Customer_Rating$ProdLine > Max_ProdLine_IQR]
```

```
Customer_Rating$ProdLine[Customer_Rating$ProdLine < Min_ProdLine_IQR]
# Histogram

hist(Customer_Rating$ProdLine,col="Green",xlim=c(1,12),
    main="ProdLine Rating",xlab="Rating 1- 10")

#================================================================================

#UNIVARIATE of ComPricing
names(Customer_Rating)
range(Customer_Rating$ComPricing)

boxplot(Customer_Rating$ComPricing,horizontal = TRUE,col="pink",
       main = "Competive Pricing Distribution", xlab= "ComPricing")
# OUTLIER CALCULATION
mean(Customer_Rating$ComPricing)
median(Customer_Rating$ComPricing)
ComPricing_IQR = IQR(Customer_Rating$ComPricing)
ComPricing_IQR
#OUTLIERS - Q3
Q3_ComPricing = quantile(Customer_Rating$ComPricing,0.75)
Q3_ComPricing = unname(Q3_ComPricing)
Q3_ComPricing
#OUTLIERS - Q1
Q1_ComPricing = unname(quantile(Customer_Rating$ComPricing,0.25))
Q1_ComPricing
Max_ComPricing_IQR = Q3_ComPricing+1.5*ComPricing_IQR # Positive Outlier Calculation
Max_ComPricing_IQR
Min_ComPricing_IQR = Q1-1.5*ComPricing_IQR # Negative Outlier Calculation
```

```
Min_ComPricing_IQR
# No Of OUTLIERS
sum(Customer_Rating$ComPricing > Max_ComPricing_IQR)
sum(Customer_Rating$ComPricing < Min_ComPricing_IQR)

# Print all Outliers
Customer_Rating$ComPricing[Customer_Rating$ComPricing > Max_ComPricing_IQR]
Customer_Rating$ComPricing[Customer_Rating$ComPricing < Min_ComPricing_IQR]
# Histogram

hist(Customer_Rating$ComPricing,col="purple",xlim=c(1,12),
    main="Competitve Pricing Rating",xlab="Rating 1- 10")

#================================================================================

#UNIVARIATE of SalesFImage
names(Customer_Rating)
range(Customer_Rating$SalesFImage)

boxplot(Customer_Rating$SalesFImage,horizontal = TRUE,col="grey",
      main = "Sales Force Image Distribution", xlab= "SalesFImage")
# OUTLIER CALCULATION
mean(Customer_Rating$SalesFImage)
median(Customer_Rating$SalesFImage)
SalesFImage_IQR = IQR(Customer_Rating$SalesFImage)
SalesFImage_IQR
#OUTLIERS - Q3
Q3_SalesFImage = quantile(Customer_Rating$SalesFImage,0.75)
Q3_SalesFImage = unname(Q3_SalesFImage)
```

```
Q3_SalesFImage
#OUTLIERS - Q1
Q1_SalesFImage = unname(quantile(Customer_Rating$SalesFImage,0.25))
Q1_SalesFImage
Max_SalesFImage_IQR = Q3_SalesFImage+1.5*SalesFImage_IQR # Positive Outlier Calculation
Max_SalesFImage_IQR
Min_SalesFImage_IQR = Q1-1.5*SalesFImage_IQR # Negative Outlier Calculation
Min_SalesFImage_IQR
# No Of OUTLIERS
sum(Customer_Rating$SalesFImage > Max_SalesFImage_IQR)
sum(Customer_Rating$SalesFImage < Min_SalesFImage_IQR)

# Print all Outliers
Customer_Rating$SalesFImage[Customer_Rating$SalesFImage > Max_SalesFImage_IQR]
Customer_Rating$SalesFImage[Customer_Rating$SalesFImage < Min_SalesFImage_IQR]
# Histogram

hist(Customer_Rating$SalesFImage,col="Green",xlim=c(1,12),
    main="Sales Force Image Rating",xlab="Rating 1- 10")

qplot(Customer_Rating$SalesFImage,data=Customer_Rating,geom ="histogram",binwidth=0.2)+
  labs(title = "Prod Qual")+labs(x="Ratings")+labs(y="Frequency")+
  scale_y_continuous(breaks=c(1:5),minor_breaks=NULL)+
  scale_x_continuous(breaks=c(1:10),minor_breaks=NULL)+
  geom_vline(xintercept = mean(Customer_Rating$SalesFImage),show.legend=TRUE,col="red")+
  geom_vline(xintercept=median(Customer_Rating$SalesFImage),show.legend=TRUE,col="blue")
#=============================================================================

#UNIVARIATE of WartyClaim
```

```
names(Customer_Rating)
range(Customer_Rating$WartyClaim)

boxplot(Customer_Rating$WartyClaim,horizontal = TRUE,col="brown",
      main = "Warranty Claims Distribution", xlab= "WartyClaim")
# OUTLIER CALCULATION
mean(Customer_Rating$WartyClaim)
median(Customer_Rating$WartyClaim)
WartyClaim_IQR = IQR(Customer_Rating$WartyClaim)
WartyClaim_IQR
#OUTLIERS - Q3
Q3_WartyClaim = quantile(Customer_Rating$WartyClaim,0.75)
Q3_WartyClaim = unname(Q3_WartyClaim)
Q3_WartyClaim
#OUTLIERS - Q1
Q1_WartyClaim = unname(quantile(Customer_Rating$WartyClaim,0.25))
Q1_WartyClaim
Max_WartyClaim_IQR = Q3_WartyClaim+1.5*WartyClaim_IQR # Positive Outlier Calculation
Max_WartyClaim_IQR
Min_WartyClaim_IQR = Q1-1.5*WartyClaim_IQR # Negative Outlier Calculation
Min_WartyClaim_IQR
# No Of OUTLIERS
sum(Customer_Rating$WartyClaim > Max_WartyClaim_IQR)
sum(Customer_Rating$WartyClaim < Min_WartyClaim_IQR)

# Print all Outliers
Customer_Rating$WartyClaim[Customer_Rating$WartyClaim > Max_WartyClaim_IQR]
Customer_Rating$WartyClaim[Customer_Rating$WartyClaim < Min_WartyClaim_IQR]
# Histogram
```

```
hist(Customer_Rating$WartyClaim,col="Green",xlim=c(1,12),
    main="Warranty Claims Rating",xlab="Rating 1- 10")


qplot(Customer_Rating$WartyClaim,data=Customer_Rating,geom ="histogram",binwidth=0.2)+
 labs(title = "Prod Qual")+labs(x="Ratings")+labs(y="Frequency")+
 scale_y_continuous(breaks=c(1:5),minor_breaks=NULL)+
 scale_x_continuous(breaks=c(1:10),minor_breaks=30)+
 geom_vline(xintercept = mean(Customer_Rating$WartyClaim),show.legend=TRUE,col="red")+
 geom_vline(xintercept=median(Customer_Rating$WartyClaim),show.legend=TRUE,col="blue")


#=============================================================================

#UNIVARIATE of OrdBilling
names(Customer_Rating)
range(Customer_Rating$OrdBilling)

boxplot(Customer_Rating$OrdBilling,horizontal = TRUE,col="Purple",
      main = "Warranty Claims Distribution", xlab= "OrdBilling")
# OUTLIER CALCULATION
mean(Customer_Rating$OrdBilling)
median(Customer_Rating$OrdBilling)
OrdBilling_IQR = IQR(Customer_Rating$OrdBilling)
OrdBilling_IQR
#OUTLIERS - Q3
Q3_OrdBilling = quantile(Customer_Rating$OrdBilling,0.75)
Q3_OrdBilling = unname(Q3_OrdBilling)
```

```
Q3_OrdBilling
#OUTLIERs - Q1
Q1_OrdBilling = unname(quantile(Customer_Rating$OrdBilling,0.25))
Q1_OrdBilling
Max_OrdBilling_IQR = Q3_OrdBilling+1.5*OrdBilling_IQR # Positive Outlier Calculation
Max_OrdBilling_IQR
Min_OrdBilling_IQR = Q1_OrdBilling-1.5*OrdBilling_IQR # Negative Outlier Calculation
Min_OrdBilling_IQR
# No Of OUTLIERS
sum(Customer_Rating$OrdBilling > Max_OrdBilling_IQR)
sum(Customer_Rating$OrdBilling < Min_OrdBilling_IQR)

# Print all Outliers
Customer_Rating$OrdBilling[Customer_Rating$OrdBilling > Max_OrdBilling_IQR]
Customer_Rating$OrdBilling[Customer_Rating$OrdBilling < Min_OrdBilling_IQR]
# Histogram

hist(Customer_Rating$OrdBilling,col="Green",xlim=c(1,12),
    main="Warranty Claims Rating",xlab="Rating 1- 10")

qplot(Customer_Rating$OrdBilling,data=Customer_Rating,geom ="histogram",binwidth=0.2)+
  labs(title = "Order & Billing")+labs(x="Ratings")+labs(y="Frequency")+
  scale_y_continuous(breaks=c(1:5),minor_breaks=NULL)+
  scale_x_continuous(breaks=c(1:10),minor_breaks=30)+
  geom_vline(xintercept = mean(Customer_Rating$OrdBilling),show.legend=TRUE,col="red")+
  geom_vline(xintercept=median(Customer_Rating$OrdBilling),show.legend=TRUE,col="blue")


#================================================================================
```

```
#UNIVARIATE of DelSpeed
names(Customer_Rating)
range(Customer_Rating$DelSpeed)

boxplot(Customer_Rating$DelSpeed,horizontal = TRUE,col="Orange",
      main = "Warranty Claims Distribution", xlab= "DelSpeed")
# OUTLIER CALCULATION
mean(Customer_Rating$DelSpeed)
median(Customer_Rating$DelSpeed)
DelSpeed_IQR = IQR(Customer_Rating$DelSpeed)
DelSpeed_IQR
#OUTLIERS - Q3
Q3_DelSpeed = quantile(Customer_Rating$DelSpeed,0.75)
Q3_DelSpeed = unname(Q3_DelSpeed)
Q3_DelSpeed
#OUTLIERs - Q1
Q1_DelSpeed = unname(quantile(Customer_Rating$DelSpeed,0.25))
Q1_DelSpeed
Max_DelSpeed_IQR = Q3_DelSpeed+1.5*DelSpeed_IQR # Positive Outlier Calculation
Max_DelSpeed_IQR
Min_DelSpeed_IQR = Q1_DelSpeed-1.5*DelSpeed_IQR # Negative Outlier Calculation
Min_DelSpeed_IQR
# No Of OUTLIERS
sum(Customer_Rating$DelSpeed > Max_DelSpeed_IQR)
sum(Customer_Rating$DelSpeed < Min_DelSpeed_IQR)

# Print all Outliers
Customer_Rating$DelSpeed[Customer_Rating$DelSpeed > Max_DelSpeed_IQR]
```

```
Customer_Rating$DelSpeed[Customer_Rating$DelSpeed < Min_DelSpeed_IQR]
# Histogram

hist(Customer_Rating$DelSpeed,col="Green",xlim=c(1,12),
    main="Delivery Speed",xlab="Rating 1- 10")

#============================================================================

#UNIVARIATE of Satisfaction
names(Customer_Rating)
range(Customer_Rating$Satisfaction)

boxplot(Customer_Rating$Satisfaction,horizontal = TRUE,col="Yellow",
      main = "Satisfaction Ratings Distribution", xlab= "Satisfaction")
# OUTLIER CALCULATION
mean(Customer_Rating$Satisfaction)
median(Customer_Rating$Satisfaction)
Satisfaction_IQR = IQR(Customer_Rating$Satisfaction)
Satisfaction_IQR
#OUTLIERS - Q3
Q3_Satisfaction = quantile(Customer_Rating$Satisfaction,0.75)
Q3_Satisfaction = unname(Q3_Satisfaction)
Q3_Satisfaction
#OUTLIERs - Q1
Q1_Satisfaction = unname(quantile(Customer_Rating$Satisfaction,0.25))
Q1_Satisfaction
Max_Satisfaction_IQR = Q3_Satisfaction+1.5*Satisfaction_IQR # Positive Outlier Calculation
Max_Satisfaction_IQR
Min_Satisfaction_IQR = Q1_Satisfaction-1.5*Satisfaction_IQR # Negative Outlier Calculation
```

```
Min_Satisfaction_IQR
# No Of OUTLIERS
sum(Customer_Rating$Satisfaction > Max_Satisfaction_IQR)
sum(Customer_Rating$Satisfaction < Min_Satisfaction_IQR)

# Print all Outliers
Customer_Rating$Satisfaction[Customer_Rating$Satisfaction > Max_Satisfaction_IQR]
Customer_Rating$Satisfaction[Customer_Rating$Satisfaction < Min_Satisfaction_IQR]
# Histogram

hist(Customer_Rating$Satisfaction,col="Green",xlim=c(1,12),
    main="Satisfaction Rating",xlab="Rating 1- 10")
rm(Satisfaction.c)

names(Customer_Rating)
attach (Customer_Rating)

boxplot(ProdQual,Ecom,TechSup,CompRes,Advertising,ProdLine,SalesFImage,
       ComPricing,WartyClaim,OrdBilling,DelSpeed,Satisfaction,
       horizontal = TRUE,

       col=c("Yellow","Red","Grey","Green","Blue",
            "Orange","Light Blue","Brown","White",
            "purple","Dark Green","Blue"),

       main = "Univariate Comparision of Ratings",
       xlab= "Rating",ylab="Attributes",
      # names=c("1","2","3","4","5","6","7","8","9","10","11","12"))
       names=c("ProdQual","Ecom","TechSup","CompRes",
```

```
          "Advertising","ProdLine","SalesFImage",
           "ComPricing","WartyClaim","OrdBilling",
           "DelSpeed","Satisfaction"))
names(Customer_Rating)

#-----------------------------------------------------------------------
#MULTI-COLINEARITY
names(Customer_Rating)
Correlation_Test=cor(Customer_Rating[1:12])
table(Correlation_Test)
Correlation_Test= as.data.frame(Correlation_Test)
View(Correlation_Test)

install.packages("xlsx")
library(xlsx)
write.xlsx(Correlation_Test, "c:/Correlation_Test.xlsx")


# Pairwise correlation
install.packages("PerformanceAnalytics")
library(PerformanceAnalytics)
#(1) Correlation: Alcohol and Fixed Acidity
chart.Correlation(cbind(ProdQual,Satisfaction))

install.packages("PerformanceAnalytics")
#---------------------
# EXPORTING TO .CSV
library(rio)
Correlation_Test= as.data.frame(Correlation_Test)
```

```
export(Correlation_Test, "Correlation.csv")
#----------------------
corrplot(cor(Customer_Rating[,1:12]))
SLM= lm(Satisfaction~ProdQual+Ecom+TechSup+CompRes+Advertising+ProdLine+SalesFImage+
        ComPricing+WartyClaim+OrdBilling+DelSpeed,,data=Customer_Rating)



vif(SLM)

#-------------------------------------------------------------------------------------
#SIMPLE LINEAR REGRESSION BETWEEN DEPENDENT VARIABLE & EVERY INDEPENDENT VARIABLE

#ProdQual
attach(Customer_Rating)
cor(Satisfaction,ProdQual)
plot(ProdQual,Satisfaction,col="red",
    main="Line of Best Fit for Satisfaction & Product Quality",
    abline(lm(Satisfaction~ProdQual,col="Blue")))
SLM_ProdQual = lm(Satisfaction~ProdQual,data=Customer_Rating)
summary(SLM_ProdQual)



#CompRes
cor(Satisfaction,CompRes)
plot(CompRes,Satisfaction,col="red",
    main="Line of Best Fit for Satisfaction & Complaint Resolution",
    abline(lm(Satisfaction~CompRes,col="Blue")))
SLM_CompRes = lm(Satisfaction~CompRes,data=Customer_Rating)
summary(SLM_CompRes)
```

```
#ProdLine
cor(Satisfaction,ProdLine)
plot(ProdLine,Satisfaction,col="red",
    main="Line of Best Fit for Satisfaction & Product Line",
    abline(lm(Satisfaction~ProdLine,col="Blue")))
SLM_ProdLine = lm(Satisfaction~ProdLine,data=Customer_Rating)
summary(SLM_ProdLine)
#SalesFImage,
cor(Satisfaction,SalesFImage)
plot(SalesFImage,Satisfaction,col="red",
    main="Line of Best Fit for Satisfaction & SalesForce Image",
    abline(lm(Satisfaction~SalesFImage,col="Blue")))
SLM_SalesFImage = lm(Satisfaction~SalesFImage,data=Customer_Rating)
summary(SLM_SalesFImage)

#WartyClaim

#DelSpeed

#---------------------------------------------------------------------------------
#PCA / FA
install.packages("nFactors")



ev=eigen(cor(Customer_Rating)) # Obtaining the Eigen Values
print(ev,digits=5)
ev
EigenValue=ev$values
EigenValue
```

```
Factor = c(1,2,3,4,5,6,7,8,9,10,11,12)
Scree = data.frame(Factor,EigenValue)
plot(Scree,main="Scree Plot",col="Blue",ylim=c(0,4))
lines(Scree,col="Red")

Unrotate=principal(Customer_Rating,nfactors=4,rotate="none")
print(Unrotate,digits=3)

UnrotatedProfile = plot(Unrotate,row.names(Unrotae$loadings))

Rotate=principal(Customer_Rating,nfactors=4,rotate="varimax")
print(Rotate,digits=4)
Rotate$scores
#-------------------------------------------------------------------------------
#MULTI REGRESSION ANALYSIS
Model=lm(Satisfaction~ProdQual+Ecom+TechSup+CompRes+
                Advertising+ProdLine+SalesFImage+
                ComPricing+WartyClaim+OrdBilling+
                DelSpeed+Satisfaction,data=Customer_Rating)
summary(Model)
```