Project 2 – Hair Factor Case Study

Assignment Report

- By Samrat Mallik

Table of Contents

1.Project Objective

The objective of this assignment to explore the datasets “Factor-Hair-Revised.csv” in R and generate insights about the datasets. The exploration procedure will consist of the following:

- Importing the datasets in R
- Understanding the structure of the datasets
- Graphical exploration
- Statistical evaluation
- Generate meaningful insights from the datasets

We also want to the answer the following questions with regard to the datasets:

Perform exploratory data analysis on the dataset. Showcase some charts, graphs. Check for outliers and missing values.

Is there evidence of multicollinearity ? Showcase your analysis

Perform simple linear regression for the dependent variable with every independent variable

Perform PCA/Factor analysis by extracting 4 factors. Interpret the output and name the Factors

Perform Multiple linear regression with customer satisfaction as dependent variables and the four factors as independent variables. Comment on the Model output and validity. Your remarks should make it meaningful for everybody

2.Assumptions

- We assume that both the datasets are normally distribued.
- Linear relationship
- Multivariate normality
- No or little multicollinearity
- No auto-correlation
- Homoscedasticity

Project 2 - Factor Hair

Samrat Mallik

8/2/2019

3.Exploratory Data Analysis

3.1.Environment Setup and Data Import

3.1.1.Installing Necessary Packages and Invoking LIbraries

```
library(ggplot2)

## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang

library(ggcorrplot)

## Warning: package 'ggcorrplot' was built under R version 3.6.1

library(ellipse)

## Warning: package 'ellipse' was built under R version 3.6.1

##
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':
##
##     pairs

library(RColorBrewer)
library(nFactors)

## Warning: package 'nFactors' was built under R version 3.6.1

## Loading required package: MASS

## Loading required package: psych

##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

## Loading required package: boot

##
## Attaching package: 'boot'

## The following object is masked from 'package:psych':
##
##     logit

## Loading required package: lattice

##
## Attaching package: 'lattice'

## The following object is masked from 'package:boot':
##
##     melanoma

##
## Attaching package: 'nFactors'

## The following object is masked from 'package:lattice':
##
##     parallel

library(psych)
library(lattice)
```

3.1.2.Setting Up Working Directory

```
setwd("C:/Users/Sam/Documents/R/Directories")
getwd()

## [1] "C:/Users/Sam/Documents/R/Directories"
```

3.1.3.Importing and Reading the Dataset

```
data = read.csv("Factor-Hair-Revised.csv")
View(data)
data = data[,-1]
View(data)
```

Since the first column is simply the index for the number of rows we remove it for ease of operation.

3.2.Variable Identification

```r
dim(data)

## [1] 100  12

summary(data)

##     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  

str(data)

## 'data.frame':    100 obs. of  12 variables:
##  $ ProdQual    : num  8.5 8.2 9.2 6.4 9 6.5 6.9 6.2 5.8 6.4 ...
##  $ Ecom        : num  3.9 2.7 3.4 3.3 3.4 2.8 3.7 3.3 3.6 4.5 ...
##  $ TechSup     : num  2.5 5.1 5.6 7 5.2 3.1 5 3.9 5.1 5.1 ...
##  $ CompRes     : num  5.9 7.2 5.6 3.7 4.6 4.1 2.6 4.8 6.7 6.1 ...
##  $ Advertising : num  4.8 3.4 5.4 4.7 2.2 4 2.1 4.6 3.7 4.7 ...
##  $ ProdLine    : num  4.9 7.9 7.4 4.7 6 4.3 2.3 3.6 5.9 5.7 ...
##  $ SalesFImage : num  6 3.1 5.8 4.5 4.5 3.7 5.4 5.1 5.8 5.7 ...
##  $ ComPricing  : num  6.8 5.3 4.5 8.8 6.8 8.5 8.9 6.9 9.3 8.4 ...
##  $ WartyClaim  : num  4.7 5.5 6.2 7 6.1 5.1 4.8 5.4 5.9 5.4 ...
##  $ OrdBilling  : num  5 3.9 5.4 4.3 4.5 3.6 2.1 4.3 4.4 4.1 ...
##  $ DelSpeed    : num  3.7 4.9 4.5 3 3.5 3.3 2 3.7 4.6 4.4 ...
##  $ Satisfaction: num  8.2 5.7 8.9 4.8 7.1 4.7 5.7 6.3 7 5.5 ...

names(data)

##  [1] "ProdQual"     "Ecom"         "TechSup"      "CompRes"     
##  [5] "Advertising"  "ProdLine"     "SalesFImage"  "ComPricing"  
##  [9] "WartyClaim"   "OrdBilling"   "DelSpeed"     "Satisfaction"
```

```r
head(data,10)
```

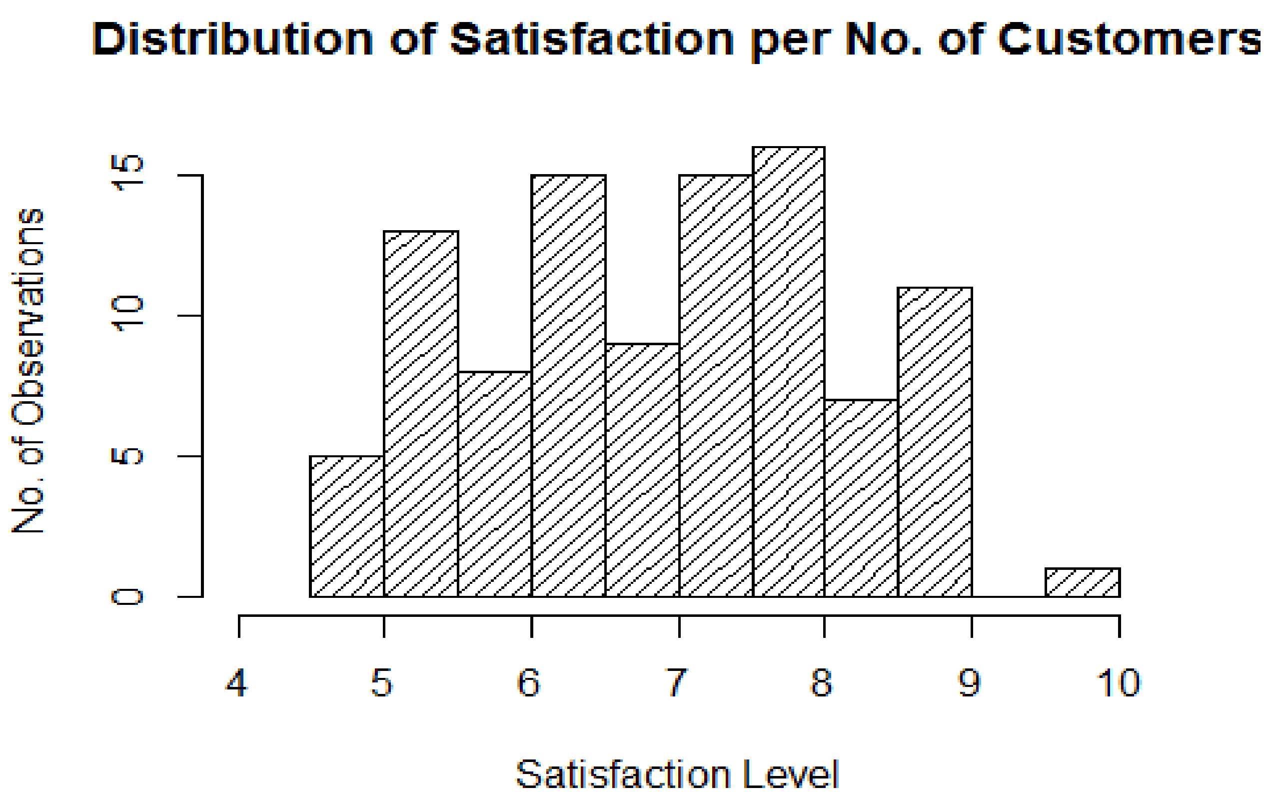
```
##    ProdQual Ecom TechSup CompRes Advertising ProdLine SalesFImage
## 1       8.5  3.9     2.5     5.9         4.8      4.9         6.0
## 2       8.2  2.7     5.1     7.2         3.4      7.9         3.1
## 3       9.2  3.4     5.6     5.6         5.4      7.4         5.8
## 4       6.4  3.3     7.0     3.7         4.7      4.7         4.5
## 5       9.0  3.4     5.2     4.6         2.2      6.0         4.5
## 6       6.5  2.8     3.1     4.1         4.0      4.3         3.7
## 7       6.9  3.7     5.0     2.6         2.1      2.3         5.4
## 8       6.2  3.3     3.9     4.8         4.6      3.6         5.1
## 9       5.8  3.6     5.1     6.7         3.7      5.9         5.8
## 10      6.4  4.5     5.1     6.1         4.7      5.7         5.7
##    ComPricing WartyClaim OrdBilling DelSpeed Satisfaction
## 1         6.8        4.7        5.0      3.7          8.2
## 2         5.3        5.5        3.9      4.9          5.7
## 3         4.5        6.2        5.4      4.5          8.9
## 4         8.8        7.0        4.3      3.0          4.8
## 5         6.8        6.1        4.5      3.5          7.1
## 6         8.5        5.1        3.6      3.3          4.7
## 7         8.9        4.8        2.1      2.0          5.7
## 8         6.9        5.4        4.3      3.7          6.3
## 9         9.3        5.9        4.4      4.6          7.0
## 10        8.4        5.4        4.1      4.4          5.5
```

Inferences: The dataset contains 100 rows and 12 columns. All the columns contain numeric data. The columns are named “ProdQual”, “Ecom”, “TechSup”, “CompRes”, “Advertising”, “ProdLine”, “SalesFImage”, “ComPricing”, “WartyClaim”, “OrdBilling”, “DelSpeed” and “Satisfaction” respectively. Here the last column “Satisfaction” is the dependant variable and the remaining are the independant variables. The numeric values in the dataset range from 1.3 - 10.

3.3.Univariate Analysis

We try to get a better understanding of the dataset by implementing some graphs/charts for easy visualisation of the data.

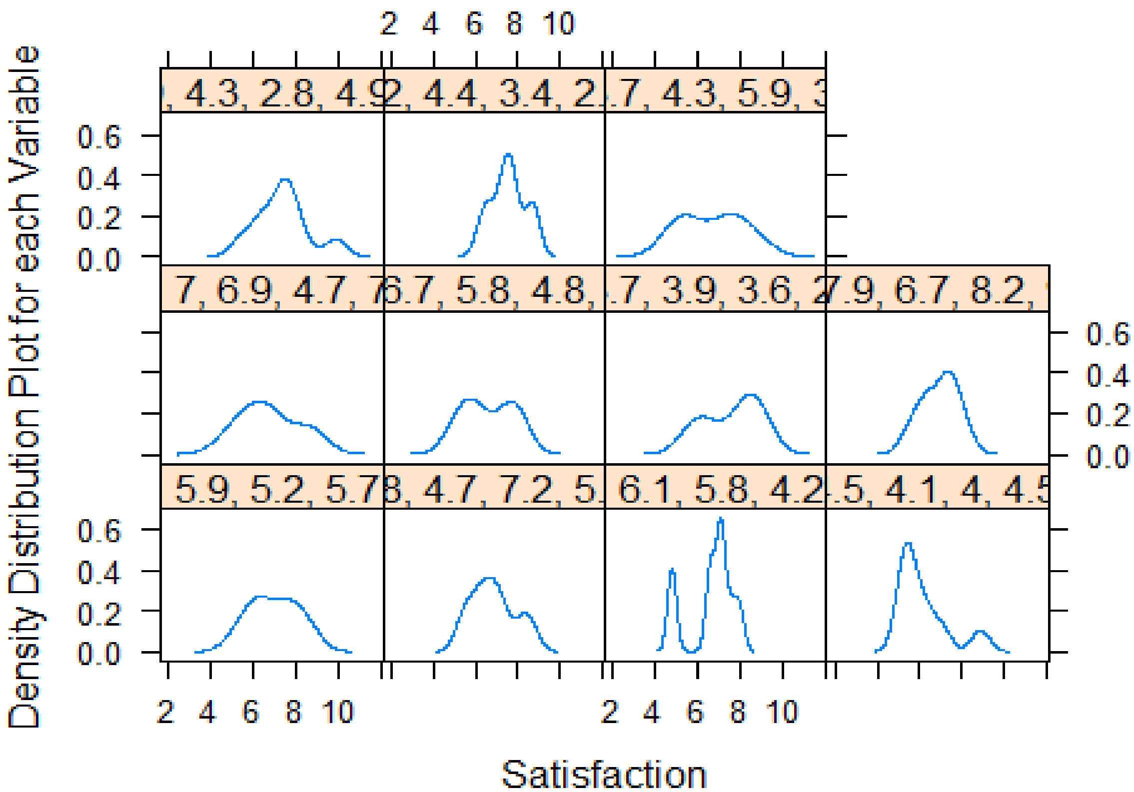
```
attach(data)
hist(Satisfaction, density = 20, col = "black", xlim = range(4:10), plot =
TRUE, main = "Distribution of Satisfaction per No. of Customers",
      xlab = "Satisfaction Level",
      ylab = "No. of Observations")
```



Inferences: We conduct univariate analysis on the dependant variable “Satisfaction”. We can deduce from the histogram that the overall customer satisfaction ratings range from 4.7 to 9.9. Largest number of observations lie in between 6.1 to 7.9 and very few observations are recorded above the 9 mark.

3.4.Bivariate Analysis

```
densityplot(~Satisfaction|data[1:11], ylab = "Density Distribution Plot for each Variable", type = "percent")
```



Inferences: The density plot gives us the kernel density estimates of Satisfcation. Here the 11 independant variables are treated as functions of “Satisfcation” and corresponding estimates are given by the plot. We see that “Ecom”,“OrdBilling” and “DelSpeed” have a narrower distribution compared to the other variables signifying lower variations.

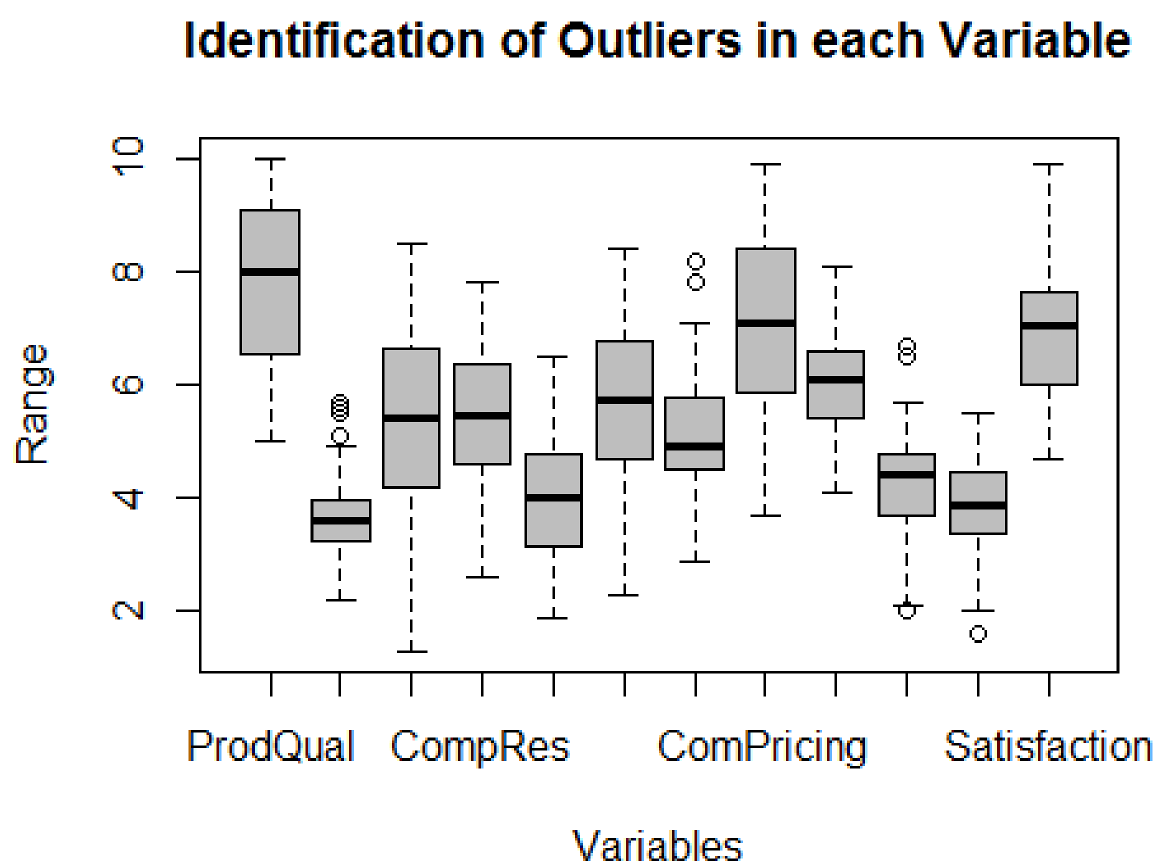
3.5.Missing Value Identification

```
sum(is.na(data))

## [1] 0
```

Inferences: There are no missing values in the dataset.

3.6.Outlier Identification

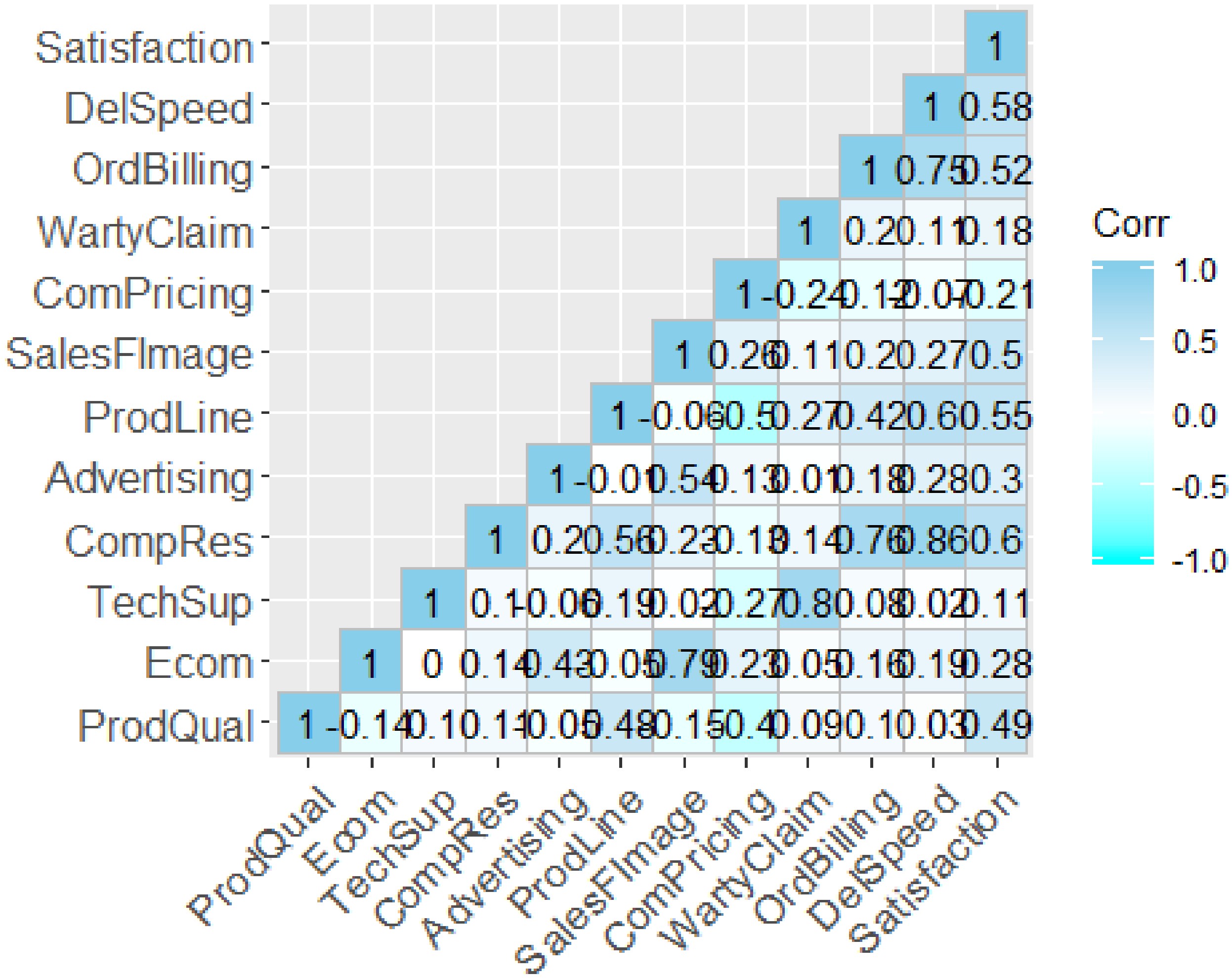
```
boxplot(data[1:12], col = "grey", main = "Identification of Outliers in each
Variable",
        xlab = "Variables",
        ylab = "Range")
```



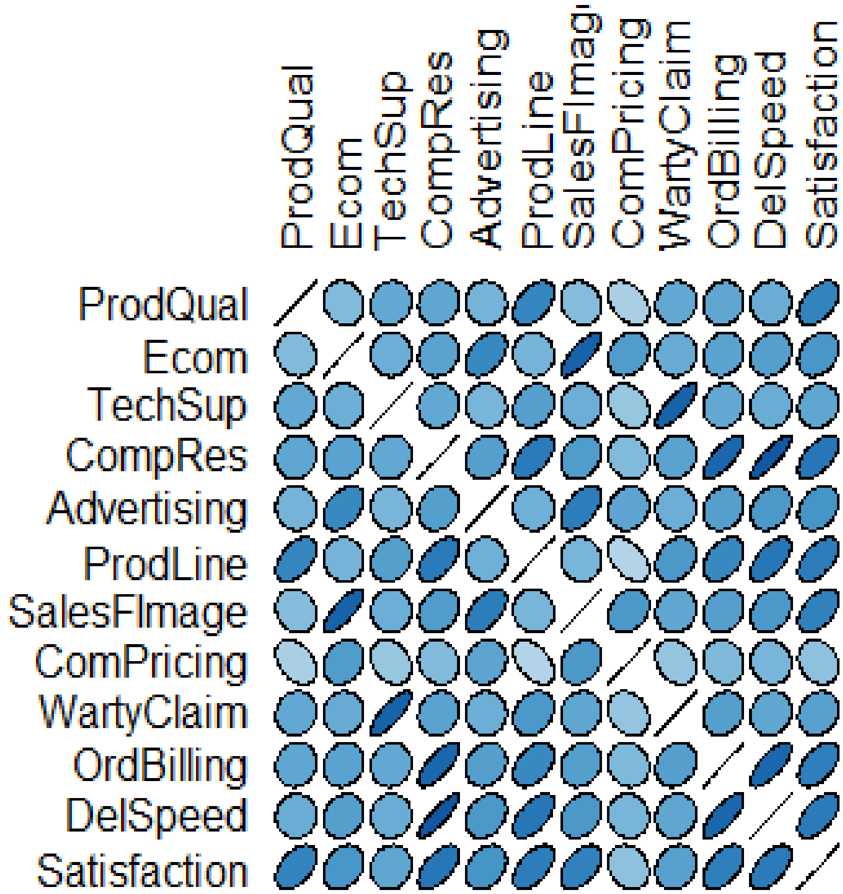
Infernces: The boxplot gives us a clear image of all possible outliers in the dataset. We can infer that “Ecom” has 4 outliers, “SalesFImage” has 2 outliers, “OrdBilling” has 3 outliers and “DelSpeed” has 1 outlier.

3.7.Checking for Multicollinearity

```
corr.matrix = round(cor(data),3)
corr.matrix

##               ProdQual   Ecom TechSup CompRes Advertising ProdLine
## ProdQual         1.000 -0.137   0.096   0.106      -0.053    0.477
## Ecom            -0.137  1.000   0.001   0.140       0.430   -0.053
## TechSup          0.096  0.001   1.000   0.097      -0.063    0.193
## CompRes          0.106  0.140   0.097   1.000       0.197    0.561
## Advertising     -0.053  0.430  -0.063   0.197       1.000   -0.012
## ProdLine         0.477 -0.053   0.193   0.561      -0.012    1.000
## SalesFImage     -0.152  0.792   0.017   0.230       0.542   -0.061
## ComPricing      -0.401  0.229  -0.271  -0.128       0.134   -0.495
## WartyClaim       0.088  0.052   0.797   0.140       0.011    0.273
## OrdBilling       0.104  0.156   0.080   0.757       0.184    0.424
## DelSpeed         0.028  0.192   0.025   0.865       0.276    0.602
## Satisfaction     0.486  0.283   0.113   0.603       0.305    0.551
##               SalesFImage ComPricing WartyClaim OrdBilling DelSpeed
## ProdQual           -0.152     -0.401      0.088      0.104    0.028
## Ecom                0.792      0.229      0.052      0.156    0.192
## TechSup             0.017     -0.271      0.797      0.080    0.025
## CompRes             0.230     -0.128      0.140      0.757    0.865
## Advertising         0.542      0.134      0.011      0.184    0.276
## ProdLine           -0.061     -0.495      0.273      0.424    0.602
## SalesFImage         1.000      0.265      0.107      0.195    0.272
## ComPricing          0.265      1.000     -0.245     -0.115   -0.073
## WartyClaim          0.107     -0.245      1.000      0.197    0.109
## OrdBilling          0.195     -0.115      0.197      1.000    0.751
## DelSpeed            0.272     -0.073      0.109      0.751    1.000
## Satisfaction        0.500     -0.208      0.178      0.522    0.577
##               Satisfaction
## ProdQual             0.486
## Ecom                 0.283
## TechSup              0.113
## CompRes              0.603
## Advertising          0.305
## ProdLine             0.551
## SalesFImage          0.500
## ComPricing          -0.208
## WartyClaim           0.178
## OrdBilling           0.522
## DelSpeed             0.577
## Satisfaction         1.000

ggcorrplot(corr.matrix, type = "lower", ggtheme = ggplot2::theme_gray,
show.legend = TRUE, show.diag = TRUE, colors = c("cyan","white","sky blue"),
lab = TRUE)
```




```r
my_colors <- brewer.pal(7, "Blues")
my_colors = colorRampPalette(my_colors)(100)

plotcorr(corr.matrix , col=my_colors[corr.matrix*50+50] , mar=c(1,1,1,1), )
```



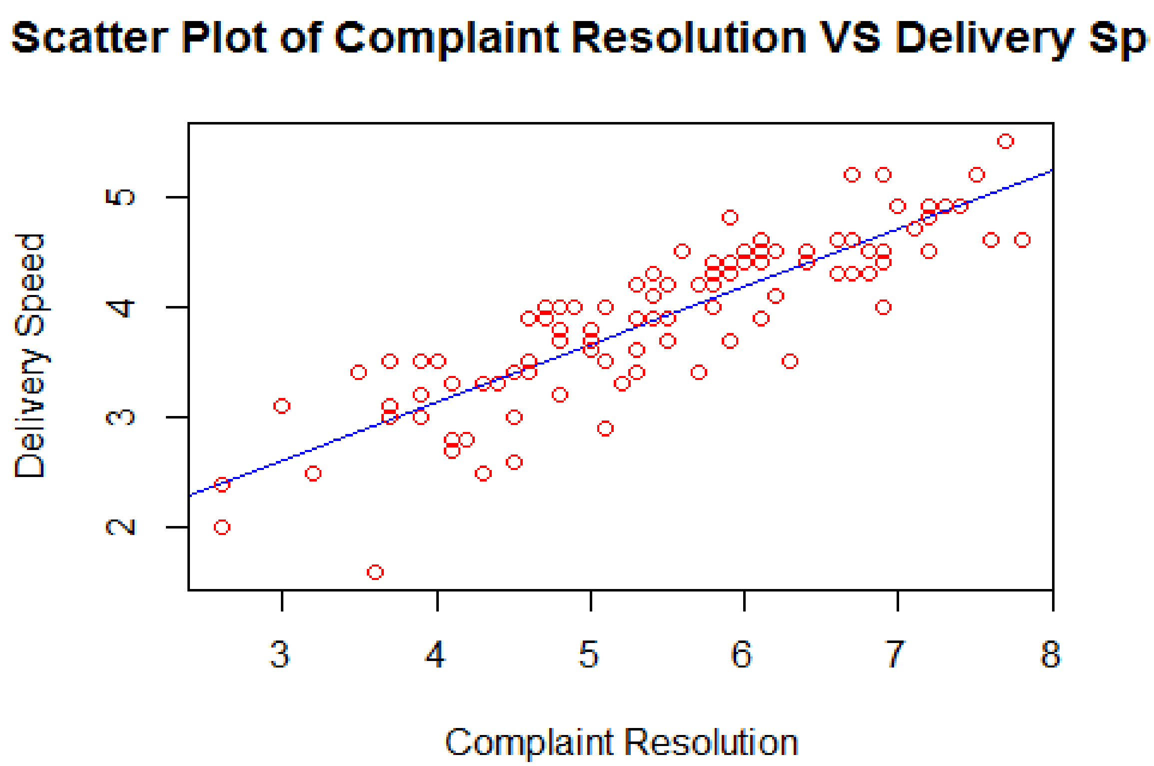
Inferences: We extract the correlation matrix of the dataset. But, since the data is quite large, we fail to effeciently deduce any conclusions from it. So, we plot the correlation matrix for visualisation purpose. We notice a high correlation between "CompRes" and "OrdBilling"&"DelSpeed", "TechSup" and "WartyClaims", "Ecom" and "SalesFImage" and moderately high correlation between "CompRes" and "Satisfaction" and "ProdLine" and "DelSpeed". Hence, we can see that the dependant variable is not highly correlated with all the indepandant variables. Furthermore, some of the independant variables are highly correlated with one another. Hence, there is evidence suggesting multicollinearity.

4.Statistical Data Analysis

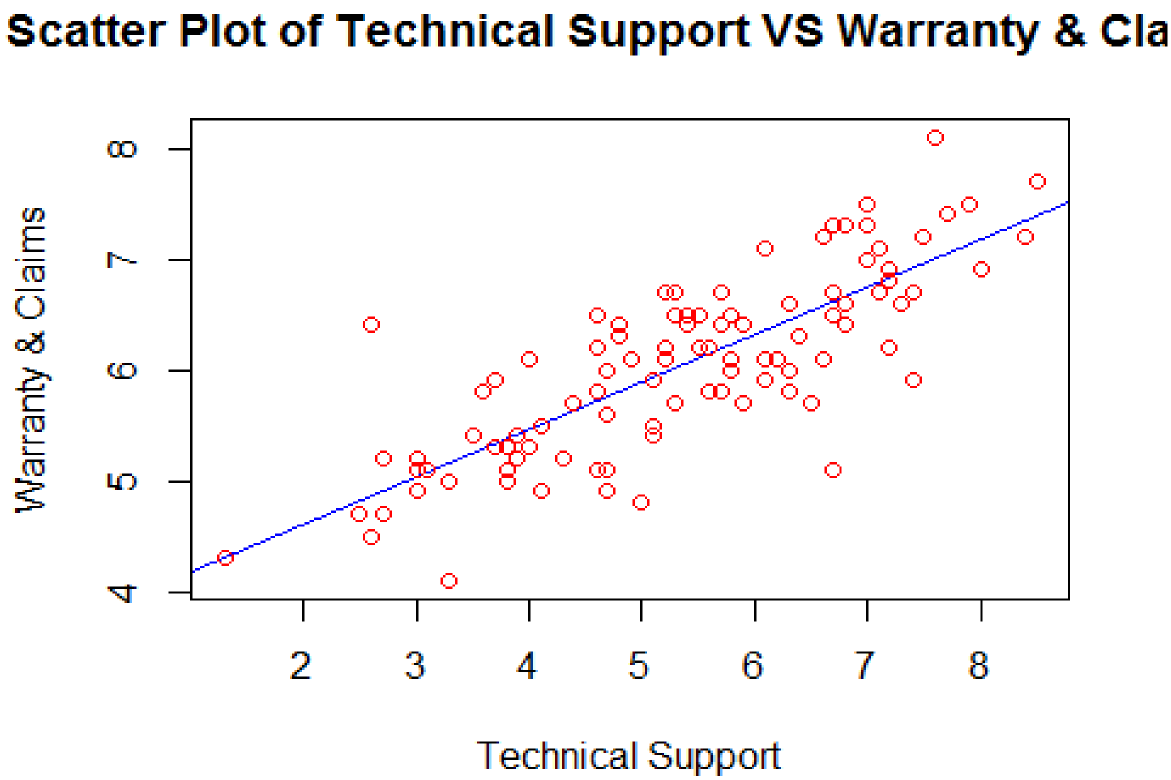
4.1.Simple Linear Regression

We draw a few scatterplots to get a better understanding of the highly correlated variables.

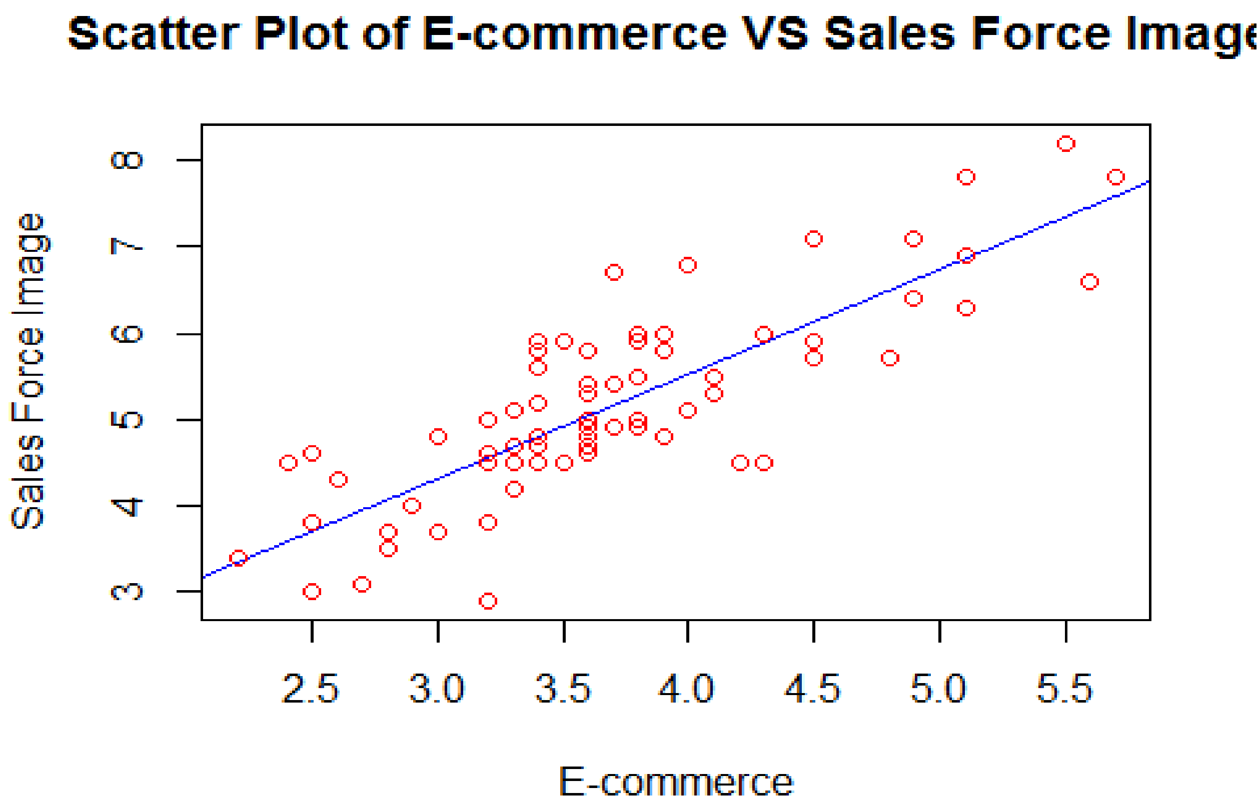
```
scatter1 = plot(CompRes,DelSpeed, col = "Red", abline(lm(DelSpeed~CompRes), col = "Blue"), main = "Scatter Plot of Complaint Resolution VS Delivery Speed",
                xlab = "Complaint Resolution",
                ylab = "Delivery Speed")
```



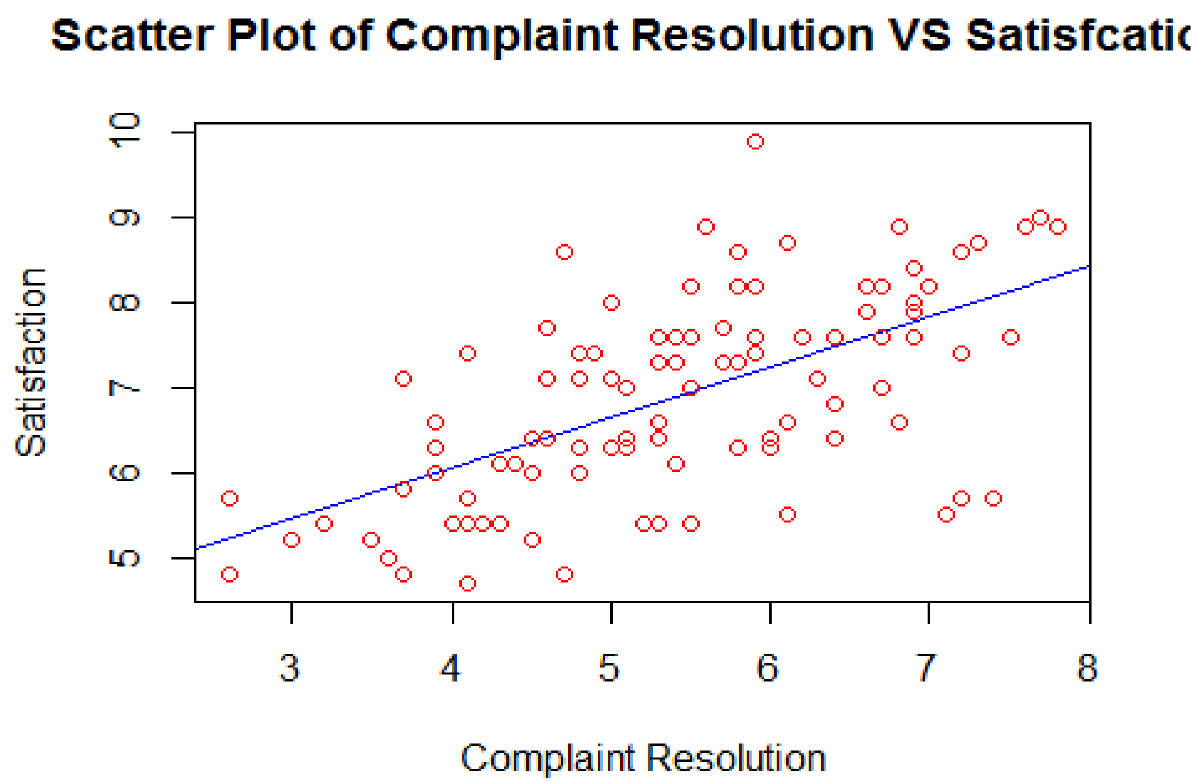

```
scatter2 = plot(TechSup, WartyClaim, col = "Red",
abline(lm(WartyClaim~TechSup), col = "Blue"), main = "Scatter Plot of Technical Support VS Warranty & Claims",
                xlab = "Technical Support",
                ylab = "Warranty & Claims")
```




```
scatter3 = plot(Ecom, SalesFImage, col = "Red", abline(lm(SalesFImage~Ecom),
col = "Blue"), main = "Scatter Plot of E-commerce VS Sales Force Image ",
                xlab = "E-commerce",
                ylab = "Sales Force Image")
```




```
scatter4 = plot(CompRes, Satisfaction, col = "Red",
abline(lm(Satisfaction~CompRes), col = "Blue"), main = "Scatter Plot of
Complaint Resolution VS Satisfcation",
                xlab = "Complaint Resolution",
                ylab = "Satisfaction")
```



We perform simple linear regression of the dependant variable "Satisfcation" with all the independant variables. null Hypothesis(Ho) = all Betas are equal alternative Hypothesis(Ha) = atleast one alternative Beta exists.

```
model1 = lm(Satisfaction~ProdQual)
summary(model1)

##
## Call:
## lm(formula = Satisfaction ~ ProdQual)
##
## 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

anova(model1)

## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq F value    Pr(>F)
## ProdQual   1  33.26  33.260  30.358 2.901e-07 ***
## Residuals 98 107.37   1.096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model2 = lm(Satisfaction~Ecom)
summary(model2)

##
## Call:
## lm(formula = Satisfaction ~ Ecom)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.37200 -0.78971  0.04959  0.68085  2.34580
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.1516     0.6161   8.361 4.28e-13 ***
## Ecom          0.4811     0.1649   2.918  0.00437 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.149 on 98 degrees of freedom
## Multiple R-squared:  0.07994,    Adjusted R-squared:  0.07056
## F-statistic: 8.515 on 1 and 98 DF,  p-value: 0.004368

anova(model2)

## Analysis of Variance Table
##
## Response: Satisfaction
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## Ecom       1  11.242 11.2424  8.5153 0.004368 **
## Residuals 98 129.385  1.3203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model3 = lm(Satisfaction~TechSup)
summary(model3)

##
## Call:
```

```
## lm(formula = Satisfaction ~ TechSup)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.26136 -0.93297  0.04302  0.82501  2.85617
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.44757    0.43592  14.791   <2e-16 ***
## TechSup      0.08768    0.07817   1.122    0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.19 on 98 degrees of freedom
## Multiple R-squared:  0.01268,    Adjusted R-squared:  0.002603
## F-statistic: 1.258 on 1 and 98 DF,  p-value: 0.2647
```

```
anova(model3)
```

```
## Analysis of Variance Table
##
## Response: Satisfaction
##           Df  Sum Sq Mean Sq F value Pr(>F)
## TechSup    1   1.783  1.7829  1.2584 0.2647
## Residuals 98 138.845  1.4168
```

```
model4 = lm(Satisfaction~CompRes)
summary(model4)
```

```
##
## Call:
## lm(formula = Satisfaction ~ CompRes)
##
## 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
```

```
anova(model4)
```

```
## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq F value    Pr(>F)
## CompRes    1 51.178  51.178   56.07 3.085e-11 ***
## Residuals 98 89.450   0.913
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model5 = lm(Satisfaction~Advertising)
summary(model5)
```

```
##
## Call:
## lm(formula = Satisfaction ~ Advertising)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.34033 -0.92755  0.05577  0.79773  2.53412
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.6259     0.4237  13.279  < 2e-16 ***
## Advertising   0.3222     0.1018   3.167  0.00206 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.141 on 98 degrees of freedom
## Multiple R-squared:  0.09282,    Adjusted R-squared:  0.08357
## F-statistic: 10.03 on 1 and 98 DF,  p-value: 0.002056
```

```
anova(model5)
```

```
## Analysis of Variance Table
##
## Response: Satisfaction
##             Df  Sum Sq Mean Sq F value   Pr(>F)
## Advertising  1  13.054 13.0535  10.027 0.002056 **
## Residuals   98 127.574  1.3018
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model6 = lm(Satisfaction~ProdLine)
summary(model6)
```

```
##
## Call:
## lm(formula = Satisfaction ~ ProdLine)
##
## 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

anova(model6)

## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq F value    Pr(>F)
## ProdLine   1 42.624  42.624  42.623 2.953e-09 ***
## Residuals 98 98.003   1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model7 = lm(Satisfaction~SalesFImage)
summary(model7)

##
## Call:
## lm(formula = Satisfaction ~ SalesFImage)
##
## 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

anova(model7)

## Analysis of Variance Table
##
## Response: Satisfaction
```

```
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## SalesFImage  1  35.186  35.186  32.703 1.164e-07 ***
## Residuals   98 105.442   1.076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model8 = lm(Satisfaction~ComPricing)
summary(model8)
```

```
##
## Call:
## lm(formula = Satisfaction ~ ComPricing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9728 -0.9915 -0.1156  0.9111  2.5845
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.03856    0.54427  14.769   <2e-16 ***
## ComPricing  -0.16068    0.07621  -2.108   0.0376 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.172 on 98 degrees of freedom
## Multiple R-squared:  0.04339,    Adjusted R-squared:  0.03363
## F-statistic: 4.445 on 1 and 98 DF,  p-value: 0.03756
```

```
anova(model8)
```

```
## Analysis of Variance Table
##
## Response: Satisfaction
##             Df  Sum Sq Mean Sq F value  Pr(>F)
## ComPricing  1   6.101  6.1014  4.4448 0.03756 *
## Residuals  98 134.526  1.3727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model9 = lm(Satisfaction~WartyClaim)
summary(model9)
```

```
##
## Call:
## lm(formula = Satisfaction ~ WartyClaim)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36504 -0.90202  0.03019  0.90763  2.88985
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   5.3581     0.8813   6.079 2.32e-08 ***
## WartyClaim    0.2581     0.1445   1.786   0.0772 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.179 on 98 degrees of freedom
## Multiple R-squared:  0.03152,    Adjusted R-squared:  0.02164 
## F-statistic:  3.19 on 1 and 98 DF,  p-value: 0.0772
```

```
anova(model9)
```

```
## Analysis of Variance Table
## 
## Response: Satisfaction
##            Df  Sum Sq Mean Sq F value Pr(>F)  
## WartyClaim  1   4.433  4.4329  3.1897 0.0772 .
## Residuals  98 136.195  1.3897                 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model10 = lm(Satisfaction~OrdBilling)
summary(model10)
```

```
## 
## Call:
## lm(formula = Satisfaction ~ OrdBilling)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -2.4005 -0.7071 -0.0344  0.7340  2.9673 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   4.0541     0.4840   8.377 3.96e-13 ***
## OrdBilling    0.6695     0.1106   6.054 2.60e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.022 on 98 degrees of freedom
## Multiple R-squared:  0.2722, Adjusted R-squared:  0.2648 
## F-statistic: 36.65 on 1 and 98 DF,  p-value: 2.602e-08
```

```
anova(model10)
```

```
## Analysis of Variance Table
## 
## Response: Satisfaction
##            Df  Sum Sq Mean Sq F value    Pr(>F)    
## OrdBilling  1  38.279  38.279  36.653 2.602e-08 ***
## Residuals  98 102.348   1.044                      
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model11 = lm(Satisfaction~DelSpeed)
summary(model11)
```

```
##
## Call:
## lm(formula = Satisfaction ~ DelSpeed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.22475 -0.54846  0.08796  0.54462  2.59432
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.2791     0.5294   6.194 1.38e-08 ***
## DelSpeed      0.9364     0.1339   6.994 3.30e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9783 on 98 degrees of freedom
## Multiple R-squared:  0.333,  Adjusted R-squared:  0.3262
## F-statistic: 48.92 on 1 and 98 DF,  p-value: 3.3e-10
```

```
anova(model11)
```

```
## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq F value  Pr(>F)
## DelSpeed   1 46.826  46.826  48.922 3.3e-10 ***
## Residuals 98 93.802   0.957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inferences: By performing simple linear regression we get the intercepts and slopess of each independant variable(X) with the dependant variable(Yhat), which we may substitute in the equation Yhat = Intercept + Slope(X). This gives us the amount of increase in Yhat for unit increase of X. e.g. In "model1", the equation is Yhat = 3.676 + 0.415(X). This means that Satisfaction increases by 0.415 times for unit increase in Product Quality when all other independent variables remain unchanged . We can also get to know the significance level of the model from the P-value, which is in this case 2.901e-07 making it much smaller than 0.001 and denoting the model as highly signnficant and a Robust Model, and also from the Significance codes given by the stars"***" beside the table. Hence, in this case we reject the null Hypothesis(Ho) and accept the Alternative Hypothesis(Ha) that atleast one alternative Beta exists and there is a regression model in the population. The Multiple R-squared value(0.2365) Implies how much of the variation in Customer Satisfaction(in this case 23.65% of the variation) is explained by the independant variable(In this case Product Quality).

4.2.Pincipal Component Analysis/ Factor Analysis

Barlett Sphericity Test for possible dimensional reduction.

```
print(cortest.bartlett(corr.matrix, nrow(data)))
```

```
## $chisq
## [1] 769.5441
##
## $p.value
## [1] 1.736073e-120
##
## $df
## [1] 66
```

Inferences: from the test we can see that the P-value(1.736073e-120) is very small and we can go ahead with dimensional reduction and principal component analysis.

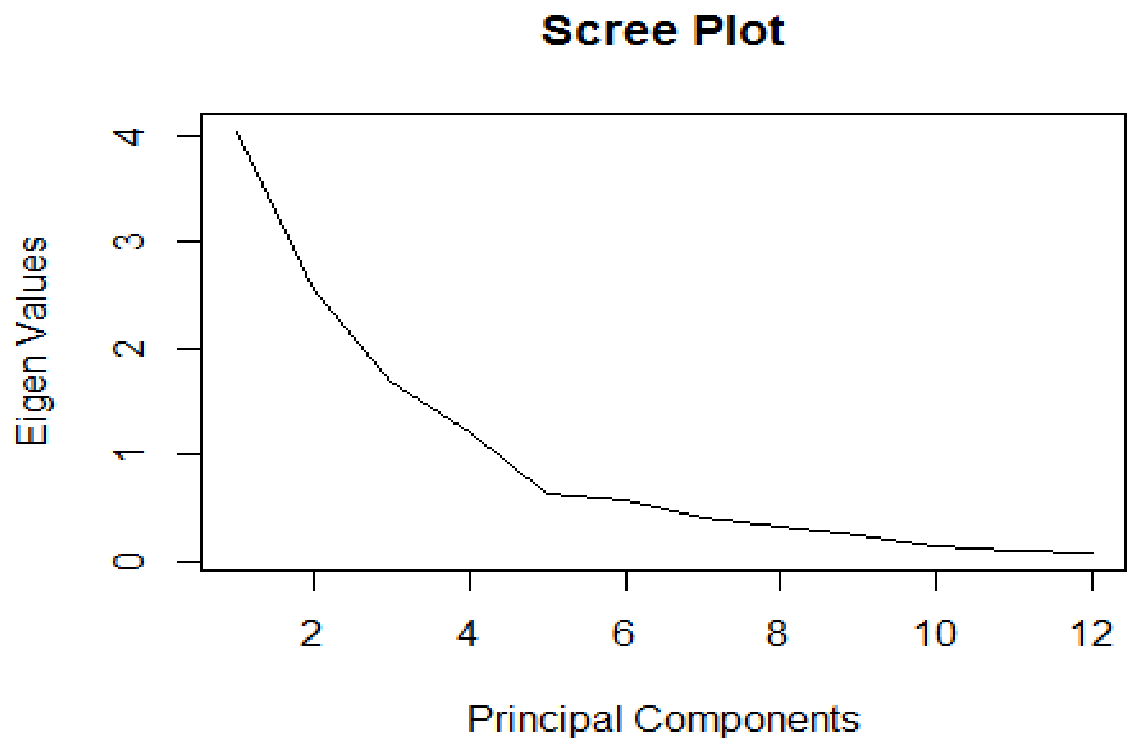
Eigen Value Calculation

```
EV = eigen(corr.matrix)
eigenvalues = EV$values
eigenvalues
```

```
##  [1] 4.04280509 2.55300833 1.69217838 1.21722584 0.63591869 0.56896713
##  [7] 0.40323059 0.32361712 0.23663377 0.14396982 0.09915782 0.08328741
```

```
plot(eigenvalues, type = "lines", xlab="Principal Components", ylab="Eigen
Values", main = "Scree Plot")
```

```
## Warning in plot.xy(xy, type, ...): plot type 'lines' will be truncated to
## first character
```



We Calculate the Eigen vales for the correlation matrix to determine the number of principal components. Following the Kaiser rule we select only components having value greater than “1” i.e. the first four components. We also draw a scree plot to determine the number of components using the Elbow rule. Here also we see the elbow bend at 4. Hence, we choose 4 factors.

```
PCA = principal(data, nfactors = length(data),rotate="none")
PCA

## Principal Components Analysis
## Call: principal(r = data, nfactors = length(data), rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10
## ProdQual      0.32 -0.50 -0.10  0.68  0.20  0.28 -0.08  0.19  0.09 -0.08
## Ecom          0.33  0.70  0.31  0.22  0.15 -0.35 -0.01  0.29  0.10 -0.01
## TechSup       0.25 -0.38  0.80 -0.20  0.03  0.09  0.01 -0.05  0.27  0.16
## CompRes       0.85  0.00 -0.26 -0.31  0.03  0.01  0.00 -0.09  0.21 -0.22
## Advertising   0.36  0.57  0.12  0.23 -0.61  0.32  0.05  0.07  0.02  0.01
## ProdLine      0.71 -0.48 -0.14  0.11 -0.02 -0.15  0.40  0.13 -0.11  0.10
## SalesFImage   0.44  0.74  0.31  0.22  0.11 -0.13 -0.01 -0.19 -0.08 -0.02
## ComPricing   -0.27  0.67 -0.07 -0.27  0.39  0.44  0.22  0.09 -0.01  0.03
## WartyClaim    0.35 -0.32  0.79 -0.21  0.02  0.11  0.03  0.06 -0.25 -0.17
## OrdBilling    0.78  0.01 -0.20 -0.34  0.04  0.07 -0.40  0.19 -0.12  0.12
## DelSpeed      0.85  0.09 -0.28 -0.32 -0.05 -0.02  0.15  0.00  0.04  0.02
## Satisfaction  0.83  0.04 -0.04  0.37  0.18  0.11 -0.05 -0.30 -0.07  0.08
```

```
##              PC11  PC12 h2       u2 com
## ProdQual     -0.07  0.06  1  7.8e-16 3.4
## Ecom          0.01 -0.10  1  5.6e-16 3.4
## TechSup       0.01  0.03  1  8.9e-16 2.2
## CompRes       0.12  0.02  1  1.9e-15 1.9
## Advertising   0.03 -0.01  1 -6.7e-16 3.7
## ProdLine      0.11  0.05  1  2.6e-15 3.0
## SalesFImage  -0.02  0.19  1  5.6e-16 2.8
## ComPricing    0.03  0.00  1  1.3e-15 3.7
## WartyClaim   -0.03 -0.05  1  7.8e-16 2.4
## OrdBilling    0.05  0.04  1  8.9e-16 2.4
## DelSpeed     -0.25 -0.02  1  1.9e-15 1.8
## Satisfaction  0.03 -0.15  1  8.9e-16 2.0
##
##                        PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9 PC10
## SS loadings           4.04 2.55 1.69 1.22 0.64 0.57 0.40 0.32 0.24 0.14
## Proportion Var        0.34 0.21 0.14 0.10 0.05 0.05 0.03 0.03 0.02 0.01
## Cumulative Var        0.34 0.55 0.69 0.79 0.85 0.89 0.93 0.95 0.97 0.98
## Proportion Explained  0.34 0.21 0.14 0.10 0.05 0.05 0.03 0.03 0.02 0.01
## Cumulative Proportion 0.34 0.55 0.69 0.79 0.85 0.89 0.93 0.95 0.97 0.98
##                       PC11 PC12
## SS loadings           0.10 0.08
## Proportion Var        0.01 0.01
## Cumulative Var        0.99 1.00
## Proportion Explained  0.01 0.01
## Cumulative Proportion 0.99 1.00
##
## Mean item complexity =  2.7
## Test of the hypothesis that 12 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0
##  with the empirical chi square  0  with prob <  NA
##
## Fit based upon off diagonal values = 1
```

We carry out the principal component analysis to check the loadings in each of the components, and the percentage of variation explained by the components. Here the first four components can explain 79% of the variation.

Kaiser-Meyer-Olkin (KMO) Test : For finding Measure of Sampling Adequacy

```
KMO(corr.matrix)
```

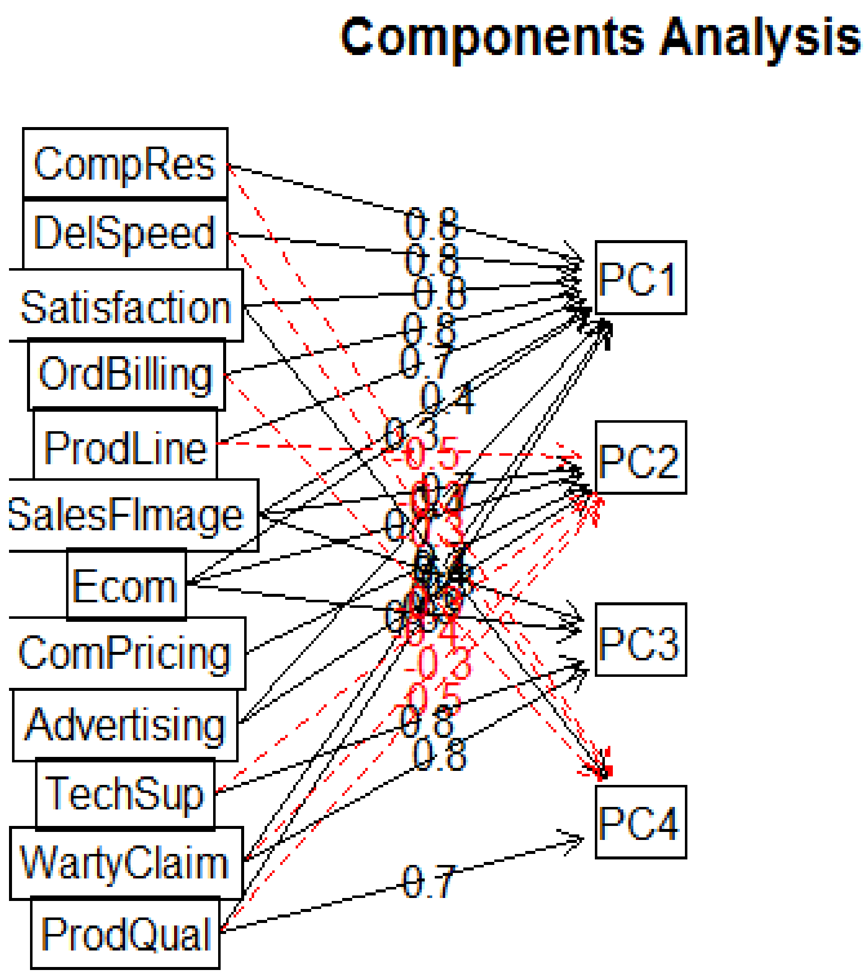
```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = corr.matrix)
## Overall MSA =  0.66
## MSA for each item =
##     ProdQual         Ecom      TechSup      CompRes  Advertising
##         0.49         0.59         0.52         0.83         0.83
##     ProdLine  SalesFImage   ComPricing   WartyClaim   OrdBilling
```

```
##         0.70         0.52         0.77         0.52         0.79
##     DelSpeed Satisfaction
##         0.72         0.66
```

Infernces: Overall Sample Adequacy is 66% hence

We do Principal Component Analysis with 4 components.

```
pca1 = principal(data, nfactors = 4, rotate = "none")
pca1

## Principal Components Analysis
## Call: principal(r = data, nfactors = 4, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                PC1   PC2   PC3   PC4   h2    u2 com
## ProdQual      0.32 -0.50 -0.10  0.68 0.82 0.180 2.4
## Ecom          0.33  0.70  0.31  0.22 0.75 0.251 2.1
## TechSup       0.25 -0.38  0.80 -0.20 0.89 0.110 1.8
## CompRes       0.85  0.00 -0.26 -0.31 0.88 0.117 1.5
## Advertising   0.36  0.57  0.12  0.23 0.52 0.477 2.2
## ProdLine      0.71 -0.48 -0.14  0.11 0.76 0.237 1.9
## SalesFImage   0.44  0.74  0.31  0.22 0.89 0.110 2.2
## ComPricing   -0.27  0.67 -0.07 -0.27 0.59 0.406 1.7
## WartyClaim    0.35 -0.32  0.79 -0.21 0.89 0.109 1.9
## OrdBilling    0.78  0.01 -0.20 -0.34 0.76 0.236 1.5
## DelSpeed      0.85  0.09 -0.28 -0.32 0.91 0.089 1.5
## Satisfaction  0.83  0.04 -0.04  0.37 0.83 0.174 1.4
##
##                        PC1  PC2  PC3  PC4
## SS loadings           4.04 2.55 1.69 1.22
## Proportion Var        0.34 0.21 0.14 0.10
## Cumulative Var        0.34 0.55 0.69 0.79
## Proportion Explained  0.43 0.27 0.18 0.13
## Cumulative Proportion 0.43 0.69 0.87 1.00
##
## Mean item complexity =  1.8
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  40.15  with prob <  0.021
##
## Fit based upon off diagonal values = 0.98

fa.diagram(pca1, simple = FALSE)
```

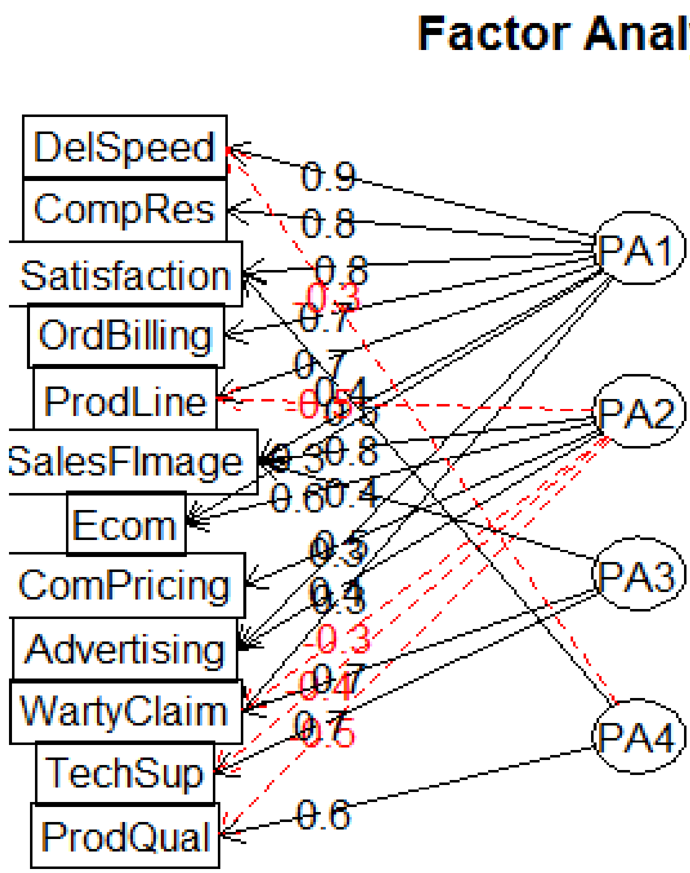


Infereneces: Here we see the factor loadings and that 79.2 % of the variation can be explained by the 4 components. We see in the diagram all the variables that go into the components.

We perform Factor Analysis with 4 factors.

```
fa1 = fa(data, nfactors = 4, rotate = "none", fm = "pa")

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs
## = np.obs, : The estimated weights for the factor scores are probably
## incorrect. Try a different factor extraction method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate =
## rotate, : An ultra-Heywood case was detected. Examine the results
carefully

fa1

## Factor Analysis using method =  pa
## Call: fa(r = data, nfactors = 4, rotate = "none", fm = "pa")
```

```
## Standardized loadings (pattern matrix) based upon correlation matrix
##                PA1   PA2   PA3   PA4   h2     u2 com
## ProdQual      0.29 -0.46 -0.07  0.61 0.67  0.332 2.4
## Ecom          0.32  0.63  0.25  0.13 0.57  0.426 1.9
## TechSup       0.24 -0.38  0.74 -0.17 0.78  0.223 1.9
## CompRes       0.84 -0.03 -0.25 -0.27 0.85  0.152 1.4
## Advertising   0.32  0.44  0.08  0.09 0.31  0.689 2.0
## ProdLine      0.67 -0.46 -0.12  0.13 0.68  0.315 1.9
## SalesFImage   0.47  0.82  0.35  0.23 1.06 -0.064 2.2
## ComPricing   -0.23  0.54 -0.05 -0.19 0.39  0.614 1.6
## WartyClaim    0.34 -0.33  0.75 -0.19 0.82  0.175 2.0
## OrdBilling    0.73 -0.01 -0.17 -0.24 0.62  0.381 1.3
## DelSpeed      0.86  0.06 -0.30 -0.33 0.95  0.055 1.5
## Satisfaction  0.81  0.03 -0.02  0.38 0.81  0.191 1.4
##
##                        PA1  PA2  PA3  PA4
## SS loadings           3.82 2.24 1.50 0.95
## Proportion Var        0.32 0.19 0.13 0.08
## Cumulative Var        0.32 0.51 0.63 0.71
## Proportion Explained  0.45 0.26 0.18 0.11
## Cumulative Proportion 0.45 0.71 0.89 1.00
##
## Mean item complexity =  1.8
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are  66  and the objective
function was  8.17 with Chi Square of  769.64
## The degrees of freedom for the model are 24  and the objective function
was  0.54
##
## The root mean square of the residuals (RMSR) is  0.02
## The df corrected root mean square of the residuals is  0.04
##
## The harmonic number of observations is  100 with the empirical chi square
6.19  with prob <  1
## The total number of observations was  100  with Likelihood Chi Square =
49.22  with prob <  0.0018
##
## Tucker Lewis Index of factoring reliability =  0.898
## RMSEA index =  0.111  and the 90 % confidence intervals are  0.061 0.144
## BIC =  -61.3
## Fit based upon off diagonal values = 1

fa.diagram(fa1, simple = FALSE)
```



Inferences: Here we see the factor loadings and that 71 % of the variation can be explained by the 4 components. We see the factors in the diagram. We may name them “Customer Service”, “Online Marketing and Sales”, “Grievence Redressal” and “Product Quality”.

4.3.Multiple Linear Regression

null Hypothesis(Ho) = all Betas are equal

alternative Hypothesis(Ha) = atleast one alternative Beta exists.

```
fadata= data.frame(fa1$scores)
attach(fadata)

MLM = lm(Satisfaction~PA1+PA2+PA3+PA4)
summary(MLM)

##
## Call:
## lm(formula = Satisfaction ~ PA1 + PA2 + PA3 + PA4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09015 -0.25036  0.04828  0.35577  0.75893
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.91800    0.04370 158.291   <2e-16 ***
## PA1          1.00957    0.04458  22.645   <2e-16 ***
## PA2         -0.04392    0.04342  -1.011    0.314
## PA3         -0.06618    0.04559  -1.452    0.150
## PA4          0.56700    0.04779  11.864   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.437 on 95 degrees of freedom
## Multiple R-squared:  0.871,  Adjusted R-squared:  0.8655
## F-statistic: 160.3 on 4 and 95 DF,  p-value: < 2.2e-16

anova(MLM)

## Analysis of Variance Table
##
## Response: Satisfaction
##           Df Sum Sq Mean Sq  F value Pr(>F)
## PA1        1 95.462  95.462 499.7848 <2e-16 ***
## PA2        1  0.009   0.009   0.0451 0.8324
## PA3        1  0.126   0.126   0.6618 0.4180
## PA4        1 26.885  26.885 140.7524 <2e-16 ***
## Residuals 95 18.146   0.191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inferences: By performing multiple linear regression we get the intercepts and slopess of the independant variables(PA1,PA2,PA3,PA4) with the dependant variable(Satisfaction), which we may substitute in the equation Yhat = Intercept + Slope1(X1) + Slope2(X2) + ….. In "MLM", the equation is Satisfaction = 6.918 + 1.00957(PA1) + (-0.04392)PA2 + (-0.06618)PA3 + 0.56700(PA4). This means that Satisfaction increases by 1.00957 times for unit increase in Customer Service(PA1). We can also get to know the significance level of the model from the P-value, which is in this case < 2.2e-16 making it much smaller than 0.001 and denoting the model as highly signnficant and a Robust Model. Hence, in this case we reject the null Hypothesis(Ho) and accept the Alternative Hypothesis(Ha) that atleast one alternative Beta exists. Also from the Significance codes given by the stars"***" beside the table we can see that only Customer Service and Product Quality are shown as highly significant. The co-efficient of Determination(R-squared) in this case the Adjusted R-squared value tells us that 86.55% of the variation in Satisfaction is explained by the model.

We test the confidence interval of the model.

```
confint(MLM)

##                   2.5 %     97.5 %
## (Intercept)  6.8312359 7.00476405
## PA1          0.9210651 1.09808294
## PA2         -0.1301130 0.04228232
## PA3         -0.1566858 0.02432750
## PA4          0.4721176 0.66187445
```

Hence we can say that the intercept will lie between 6.831 and 7.004 with 97.5 % confidence. Similarly for the independant variables.

We use the backtracking ability to determine the robustness of the model and analyze the goodness of fit.

```
testdata = data.frame(PA1 = 0.1, PA2 = 0.12, PA3 = 0.06, PA4 = 0.08)

Prediction = predict(MLM, testdata)

Prediction = predict(MLM, testdata, interval = "confidence")
Prediction

##        fit      lwr      upr
## 1 7.055076 6.966921 7.143232

Predicted = predict(MLM)
Actual = Satisfaction


Backtrack = data.frame(Actual, Predicted)
Backtrack

##     Actual Predicted
## 1      8.2  7.836021
```

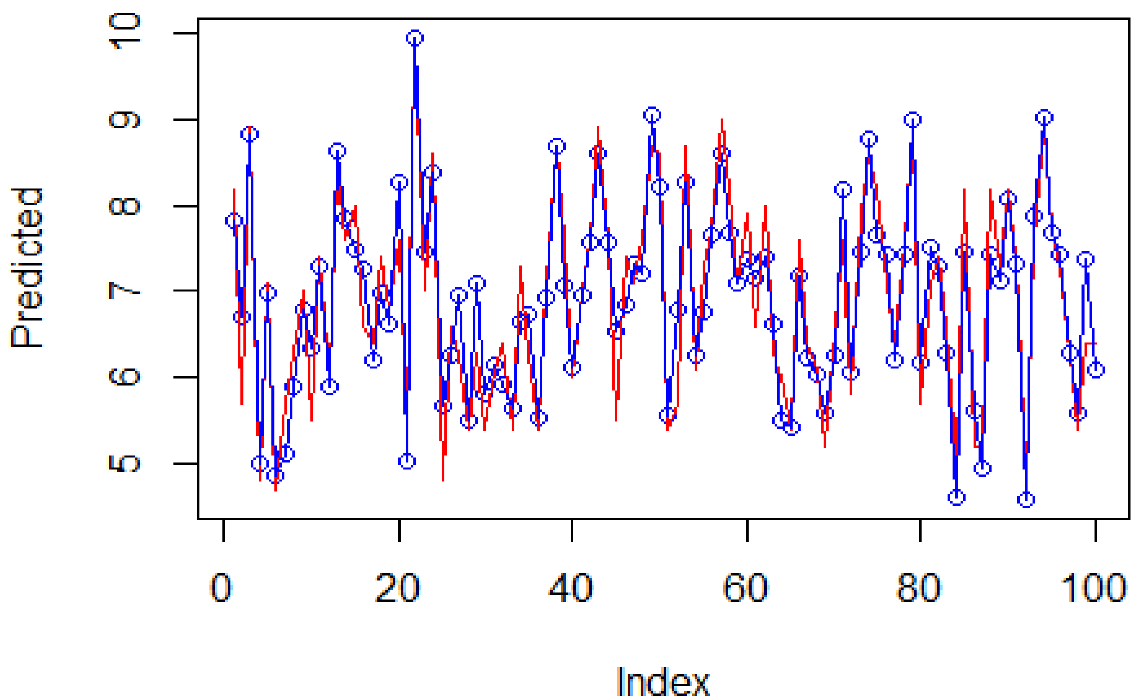
```
## 2       5.7  6.703456
## 3       8.9  8.843100
## 4       4.8  5.007135
## 5       7.1  6.999824
## 6       4.7  4.853575
## 7       5.7  5.103462
## 8       6.3  5.891856
## 9       7.0  6.783291
## 10      5.5  6.347160
## 11      7.4  7.307009
## 12      6.0  5.887717
## 13      8.4  8.620037
## 14      7.6  7.852248
## 15      8.0  7.474719
## 16      6.6  7.270262
## 17      6.4  6.210722
## 18      7.4  6.999540
## 19      6.8  6.612450
## 20      7.6  8.279319
## 21      5.4  5.038437
## 22      9.9  9.949232
## 23      7.0  7.462028
## 24      8.6  8.372474
## 25      4.8  5.672120
## 26      6.6  6.246158
## 27      6.3  6.956354
## 28      5.4  5.504210
## 29      6.3  7.101239
## 30      5.4  5.811862
## 31      6.1  6.158635
## 32      6.4  5.919328
## 33      5.4  5.649731
## 34      7.3  6.642323
## 35      6.3  6.739827
## 36      5.4  5.535765
## 37      7.1  6.919981
## 38      8.7  8.686578
## 39      7.6  7.077701
## 40      6.0  6.119404
## 41      7.0  6.947348
## 42      7.6  7.565538
## 43      8.9  8.607323
## 44      7.6  7.560469
## 45      5.5  6.552751
## 46      7.4  6.854253
## 47      7.1  7.324843
## 48      7.6  7.218262
## 49      8.7  9.057903
## 50      8.6  8.215701
## 51      5.4  5.560874
```

```
## 52      5.7  6.790153
## 53      8.7  8.271788
## 54      6.1  6.268753
## 55      7.3  6.750636
## 56      7.7  7.656089
## 57      9.0  8.598612
## 58      8.2  7.691361
## 59      7.1  7.099739
## 60      7.9  7.386871
## 61      6.6  7.164487
## 62      8.0  7.393884
## 63      6.3  6.616814
## 64      6.0  5.511481
## 65      5.4  5.432239
## 66      7.6  7.186907
## 67      6.4  6.219325
## 68      6.1  6.030118
## 69      5.2  5.586153
## 70      6.6  6.255860
## 71      7.6  8.174553
## 72      5.8  6.065144
## 73      7.9  7.446760
## 74      8.6  8.771071
## 75      8.2  7.669004
## 76      7.1  7.419365
## 77      6.4  6.192422
## 78      7.6  7.428675
## 79      8.9  9.009633
## 80      5.7  6.186735
## 81      7.1  7.509699
## 82      7.4  7.288334
## 83      6.6  6.283700
## 84      5.0  4.599362
## 85      8.2  7.468102
## 86      5.2  5.615580
## 87      5.2  4.944348
## 88      8.2  7.441075
## 89      7.3  7.114315
## 90      8.2  8.077354
## 91      7.4  7.312730
## 92      4.8  4.586756
## 93      7.6  7.867144
## 94      8.9  9.030612
## 95      7.7  7.680975
## 96      7.3  7.439503
## 97      6.3  6.298790
## 98      5.4  5.582461
## 99      6.4  7.370488
## 100     6.4  6.102484
```

We can see a table with the actual and predicted values of Satisfaction.

To visualise this properly we use a chart.

```
plot(Actual, col = "Red")

plot(Predicted, col = "Blue")
lines(Actual, col = "Red")
lines(Predicted, col = "Blue")
```



5.Conclusion: Therefore, Regression exists in the population and the model is robust. We can conclude that the model is a good fit and the predicted values are very close to the actual values hence fortifying the robustness.