Factor Hair

Muthu Pandian G

August 26, 2019

Mini Project on PCA and Regression using Hair Dataset

OBJECTIVE OF THE PROJECT:

The objective of the project is to use the dataset 'Factor-Hair-Revised.csv to **build an optimum regression model to predict satisfaction**.

1. You are expected to perform exploratory data analysis on the dataset. Showcase some charts and graphs.
2. Check for outliers and missing values.
3. Is there evidence of multicollinearity?
4. Showcase your analysis
5. Perform simple linear regression for the dependent variable with every independent variable
6. Perform PCA/Factor analysis by extracting 4 factors.
7. Interpret the output and name the factors
8. Perform Multiple linear regression with customer satisfaction as dependent variables and the four factors as independent variables.
9. Comment on the Model output and validity. Your remarks should make it meaningful for everybody

Importing the Dataset:

```
setwd("D:/Great Lakes/Projects/REGRESSION AND PCA")
Hair <- read.csv("Factor-Hair-Revised.csv",header = TRUE)
attach(Hair)
```

Understanding the data

Structure of Data

```
str(Hair)

## 'data.frame':    100 obs. of  13 variables:
##  $ ID          : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
##  $ 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 ...
```

Column Names

```
names(Hair)
```

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

1. We have 13 variables and 100 observations.
2. We have 12 independent Variables 1 Dependent Variables - Customer Satisfaction
3. All the variables are of numeric data type except ID Variable which is Integer data type

Summary:

```
summary(Hair)
```

```
##        ID            ProdQual           Ecom          TechSup
##  Min.   :  1.00   Min.   : 5.000   Min.   :2.200   Min.   :1.300
##  1st Qu.: 25.75   1st Qu.: 6.575   1st Qu.:3.275   1st Qu.:4.250
##  Median : 50.50   Median : 8.000   Median :3.600   Median :5.400
##  Mean   : 50.50   Mean   : 7.810   Mean   :3.672   Mean   :5.365
##  3rd Qu.: 75.25   3rd Qu.: 9.100   3rd Qu.:3.925   3rd Qu.:6.625
##  Max.   :100.00   Max.   :10.000   Max.   :5.700   Max.   :8.500
##     CompRes       Advertising       ProdLine      SalesFImage
##  Min.   :2.600   Min.   :1.900   Min.   :2.300   Min.   :2.900
##  1st Qu.:4.600   1st Qu.:3.175   1st Qu.:4.700   1st Qu.:4.500
##  Median :5.450   Median :4.000   Median :5.750   Median :4.900
##  Mean   :5.442   Mean   :4.010   Mean   :5.805   Mean   :5.123
##  3rd Qu.:6.325   3rd Qu.:4.800   3rd Qu.:6.800   3rd Qu.:5.800
##  Max.   :7.800   Max.   :6.500   Max.   :8.400   Max.   :8.200
##    ComPricing      WartyClaim      OrdBilling       DelSpeed
##  Min.   :3.700   Min.   :4.100   Min.   :2.000   Min.   :1.600
##  1st Qu.:5.875   1st Qu.:5.400   1st Qu.:3.700   1st Qu.:3.400
##  Median :7.100   Median :6.100   Median :4.400   Median :3.900
##  Mean   :6.974   Mean   :6.043   Mean   :4.278   Mean   :3.886
##  3rd Qu.:8.400   3rd Qu.:6.600   3rd Qu.:4.800   3rd Qu.:4.425
```

```
## Max.   :9.900   Max.   :8.100   Max.   :6.700   Max.   :5.500  
##  Satisfaction  
## Min.   :4.700  
## 1st Qu.:6.000  
## Median :7.050  
## Mean   :6.918  
## 3rd Qu.:7.625  
## Max.   :9.900
```

Description of Variables

The data file Factor-Hair.csv contains 12 variables used for Market Segmentation in the context of Product Service Management.

Variable	Expansion
ProdQual	Product Quality
Ecom	E-Commerce
TechSup	Technical Support
CompRes	Complaint Resolution
Advertising	Advertising
ProdLine	Product Line
SalesFImage	Salesforce Image
ComPricing	Competitive Pricing
WartyClaim	Warranty & Claims
OrdBilling	Order & Billing
DelSpeed	Delivery Speed
Satisfaction	Customer Satisfaction

Data Preparation

Checking NA Values/ Missing Values

```
sapply(Hair,function(x) sum(is.na(x)))
```

```
##           ID     ProdQual         Ecom      TechSup      CompRes 
##            0            0            0            0            0 
##  Advertising     ProdLine  SalesFImage   ComPricing   WartyClaim 
##            0            0            0            0            0 
##   OrdBilling     DelSpeed Satisfaction 
##            0            0            0
```

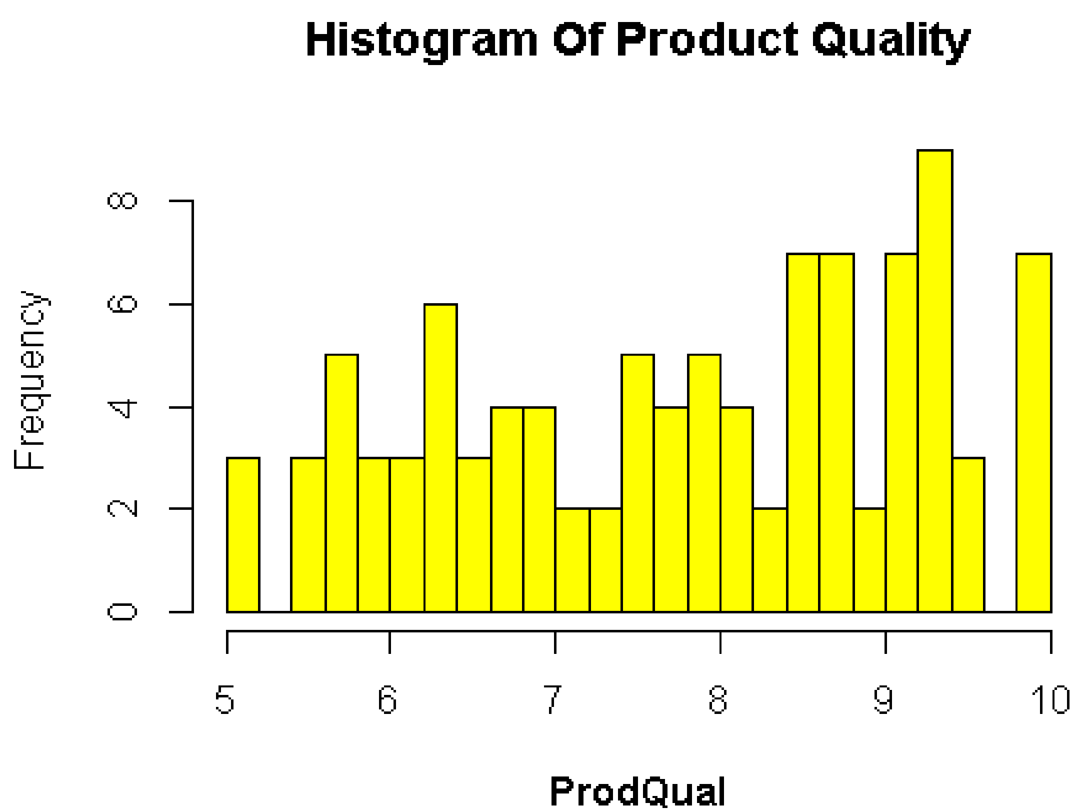
Exploratory Data Analysis

Univariate Analysis

Frequency Distribution of each Independent numerical Variable

Frequency Distribution of Product Quality

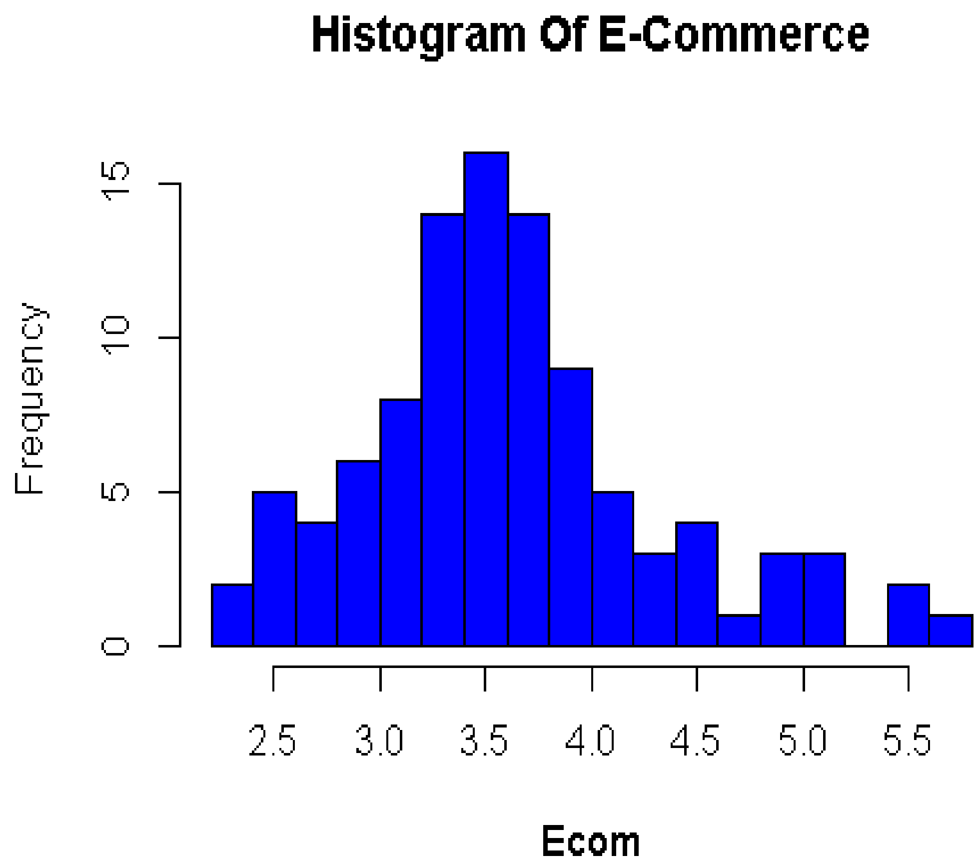
```
hist(ProdQual,col = "yellow",breaks = 20,main = "Histogram Of Product Quality
",xlab = expression(bold(ProdQual)))
```



- From the Product Quality plot, we can see that the the Product Quality ratings distribution lies in the range 5 - 10.
- Most of the Product Quality got ratings around 9-9.5. It means most of the Product Quality is very good.
- Another important thing to note is it's not normally distributed.

Frequency Distribution of E-Commerce

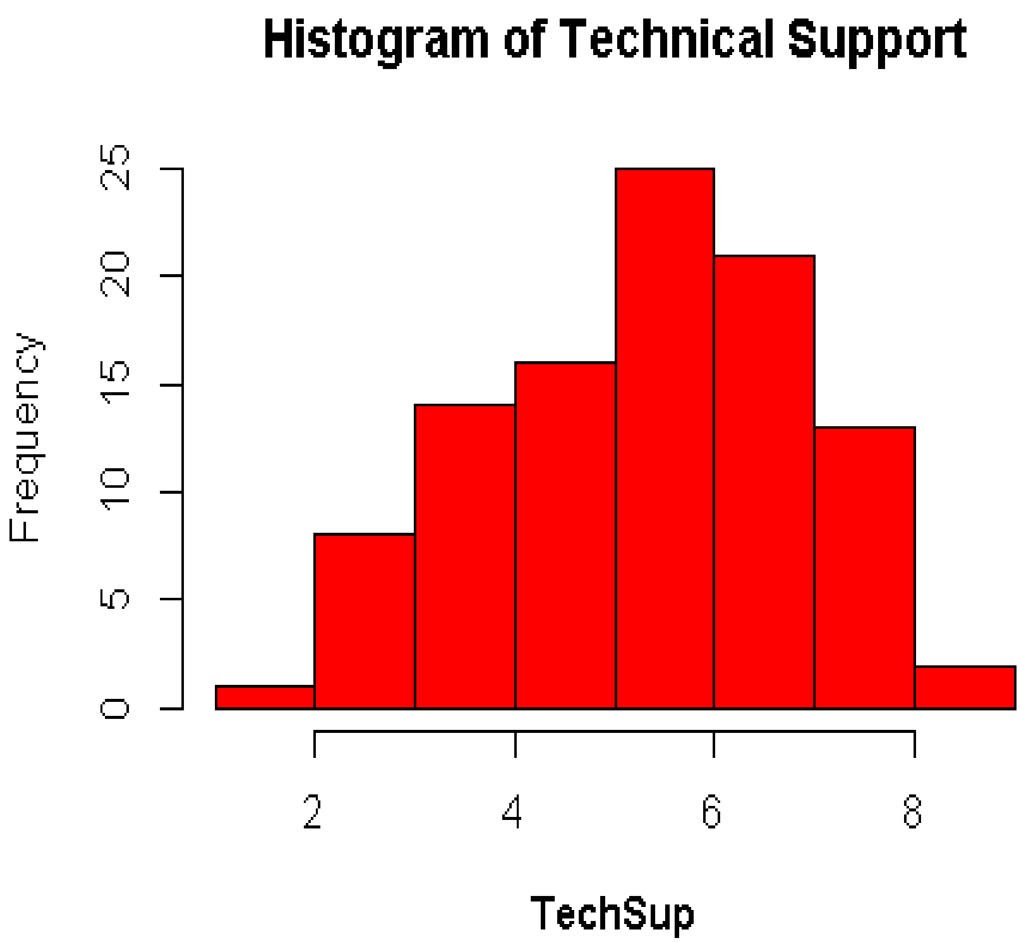
```
hist(Hair$Ecom,col = "Blue",breaks = 20,main = "Histogram Of E-Commerce",xlab
= expression(bold(Ecom)))
```



- Here, in the Ecommerce plot the Ecom rating distribution lies in the range of 2- 6.
- It's clear that the Ecom services provided is not good as we don't have a single rating above 6.
- It is almost normally distributed.

Frequency Distribution of Technical Support

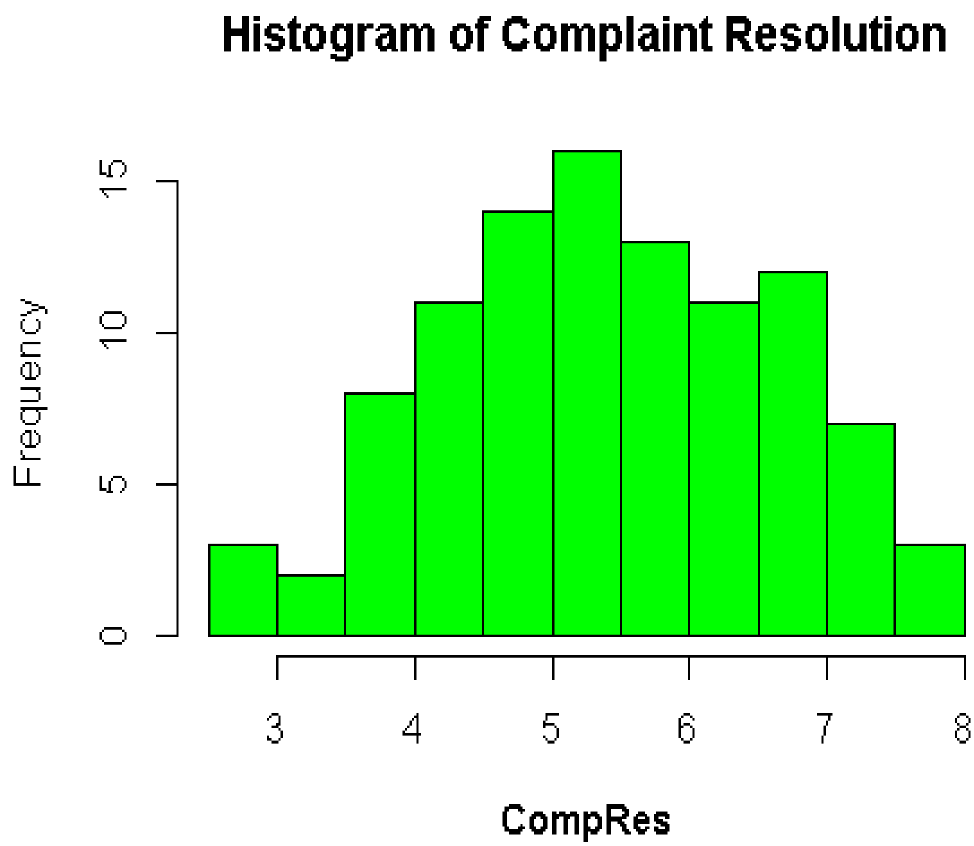
```
hist(TechSup,col = "Red",main = "Histogram of Technical Support",xlab = expre
ssion(bod(TechSup))
```



- Here, in the Technical Support plot the Tech Support ratings distribution range lies from 1 to 8.5 as most values tend to be around 5-7
- From this we can infer that the rating for Technical Support is pretty average and have an even spread.

Frequency Distribution of Complaint Resolution

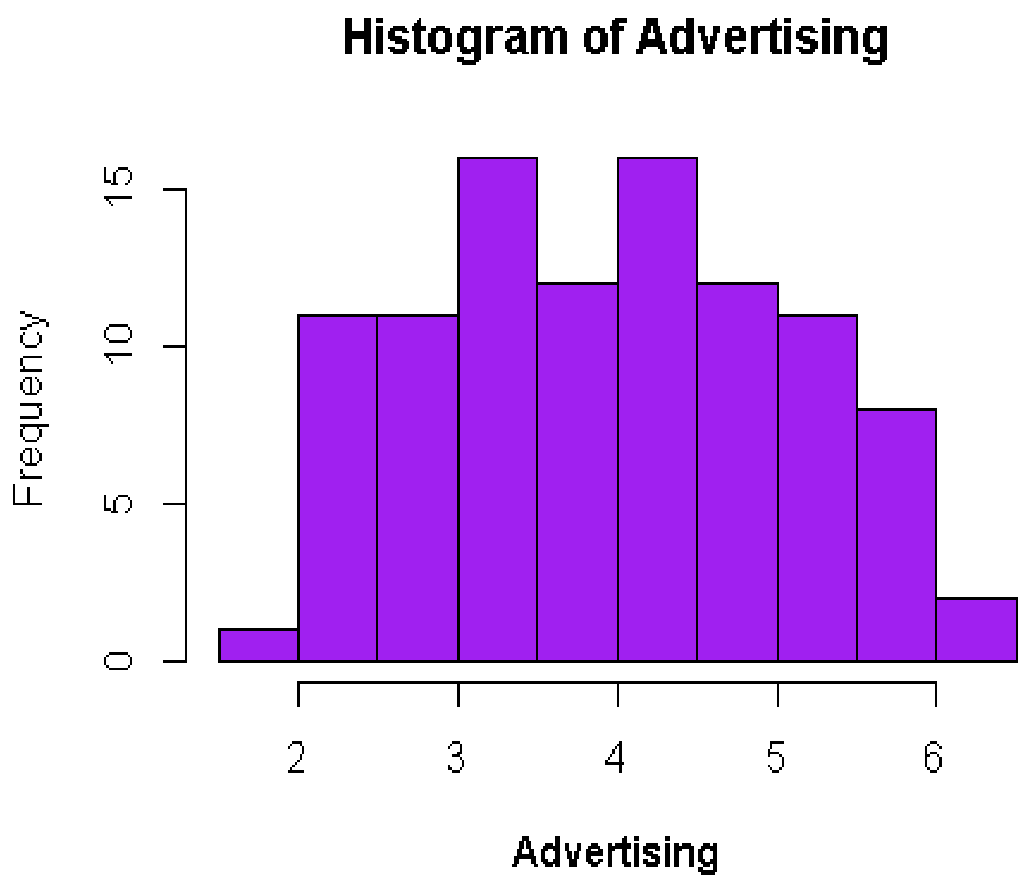
```
hist(CompRes,col = "Green",main = "Histogram of Complaint Resolution",xlab = 
expression(bold(CompRes)))
```



- From this plot, the Complaint Resolution ratings distribution range lies from 2 to 8 as most values tend to be around 5-6
- From this we can infer that the rating for Complaint Resolution is also pretty average.
- It is close to follow normally distribution.

Frequency Distribution of Advertising

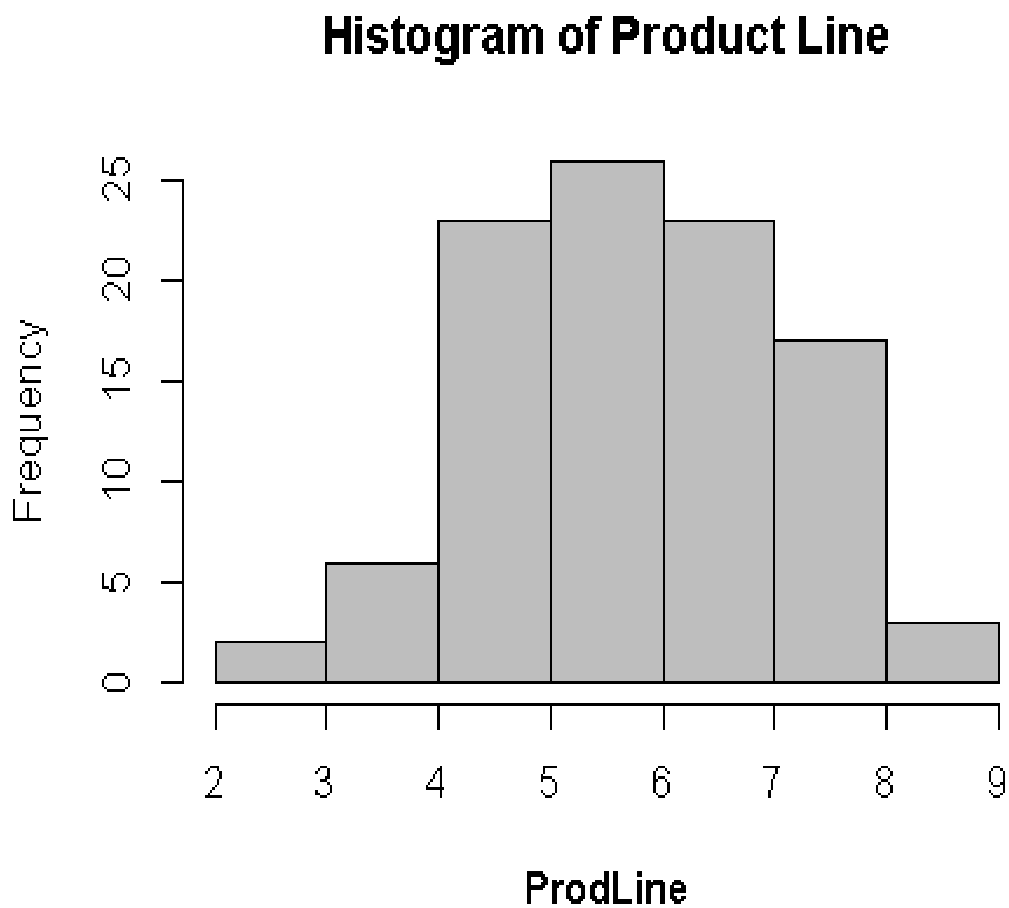
```
hist(Advertising,col = "Purple",main = "Histogram of Advertising",xlab = expr
ession(bold(Advertising)))
```



- Here, the Advertising ratings range lies from 1.9 to 6.5 as most values tend to be around 3.2 and 4.2
- From this we can infer that the rating for Advertising is also pretty average
- It has an even spread.

Frequency Distribution of Product Line

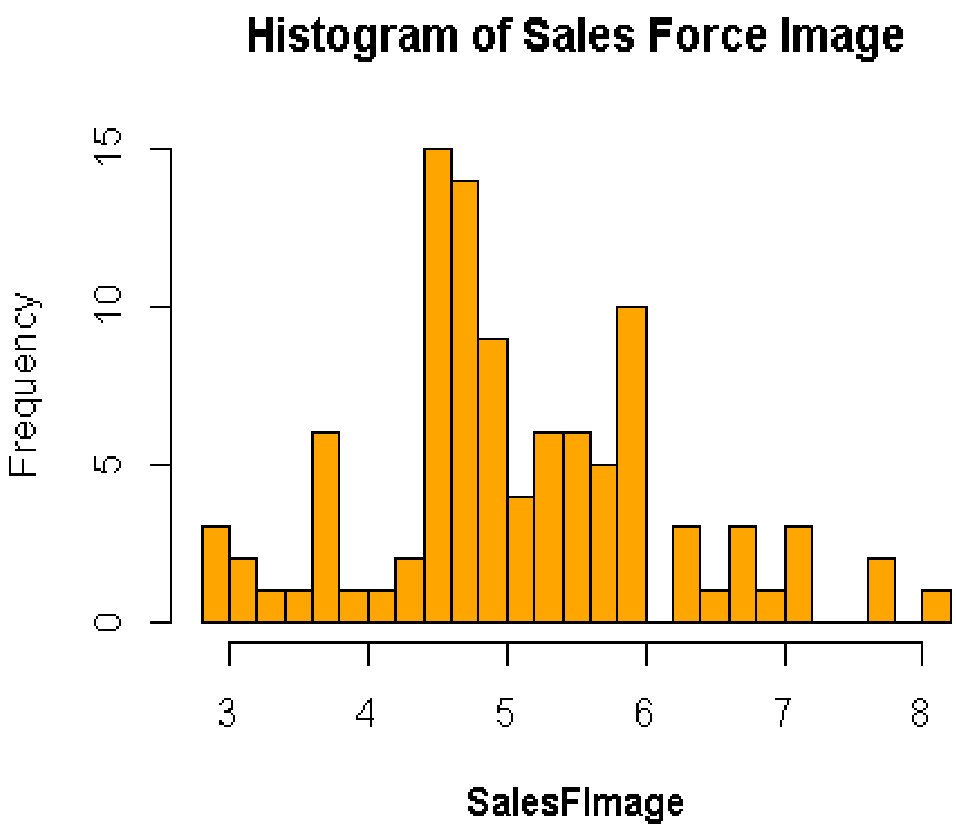
```
hist(ProdLine,col = "Grey",main = "Histogram of Product Line",xlab = expressi
on(bold(ProdLine)))
```



- From this plot, the Product Line ratings distribution range lies from 2 to 8.5 as most values tend to be around 5-6
- From this we can infer that the average rating for Product Line is around 5.5 which is average
- It is almost normally distributed

Frequency Distribution of Sales Force Image

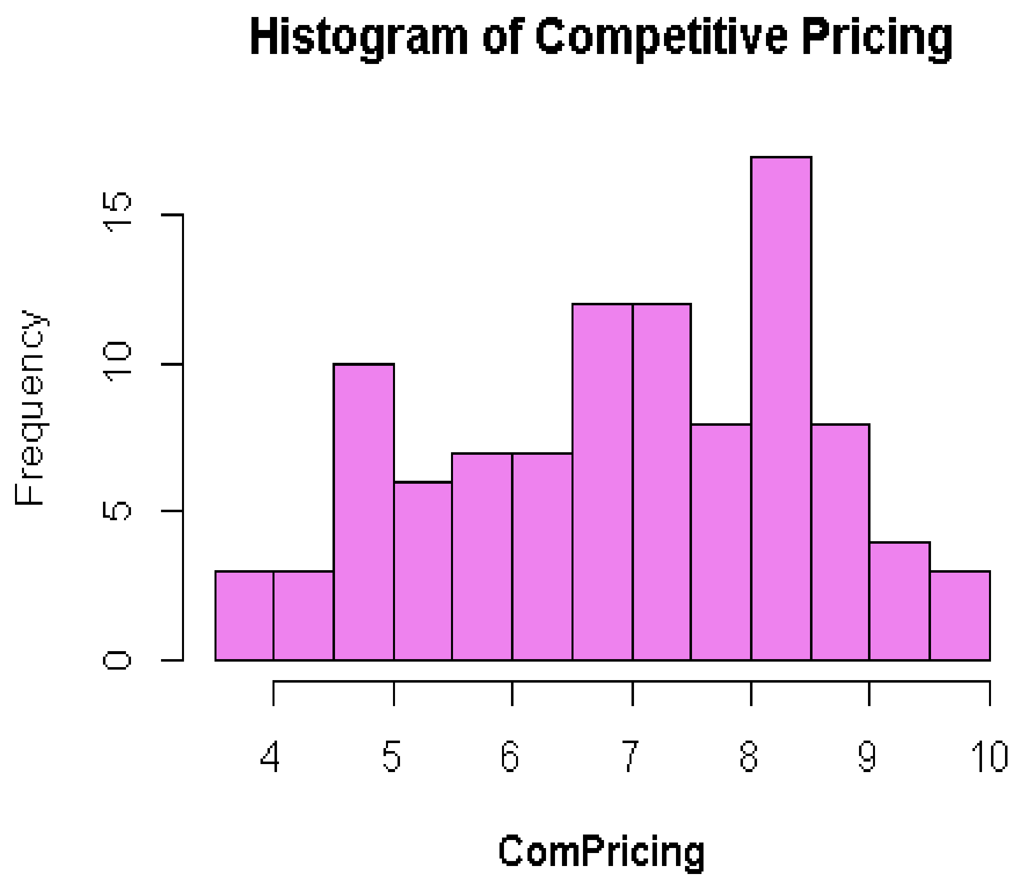
```
hist(SalesFImage,col = "Orange",breaks = 20,main = "Histogram of Sales Force
Image",xlab = expression(bold(SalesFImage)))
```



- Here, the Sales Force Image range lies from 2.9 to 8.2 as most values tend to be around 4.6.
- From this we can infer that the rating for Sales Force Image is also pretty average and it has an even spread.
- There is a possibility of Outlier also which we will check using boxplot later on.

Frequency Distribution of Competitive Pricing

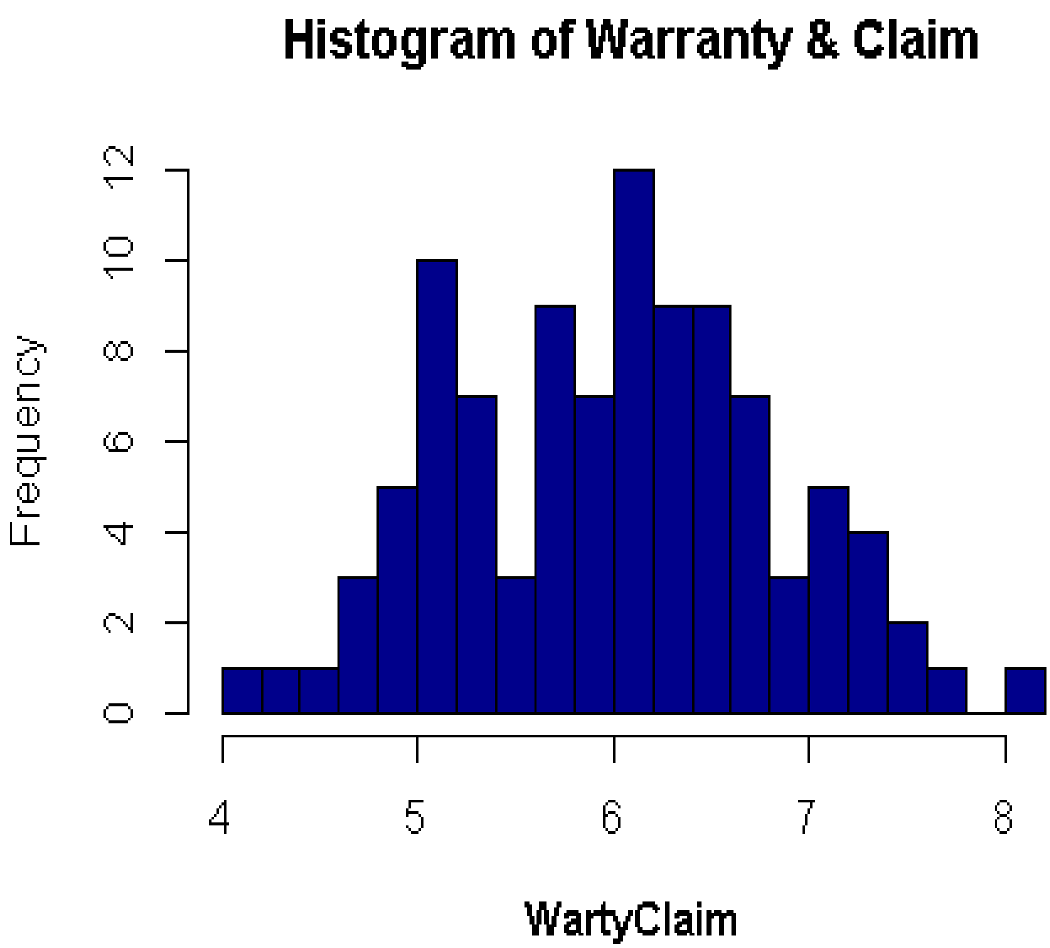
```
hist(ComPricing,col = "Violet",breaks = 20,main = "Histogram of Competitive P
ricing",xlab = expression(bold(ComPricing)))
```



- From the Plot, the Competitive Pricing range lies from 3.7 to 9.9 as most values tend to be around 8.2.
- We can Infer that the pricing range is good as most value lies above 6 rating.
- It’s an even Spread

Frequency Distribution of Warranty & Claim

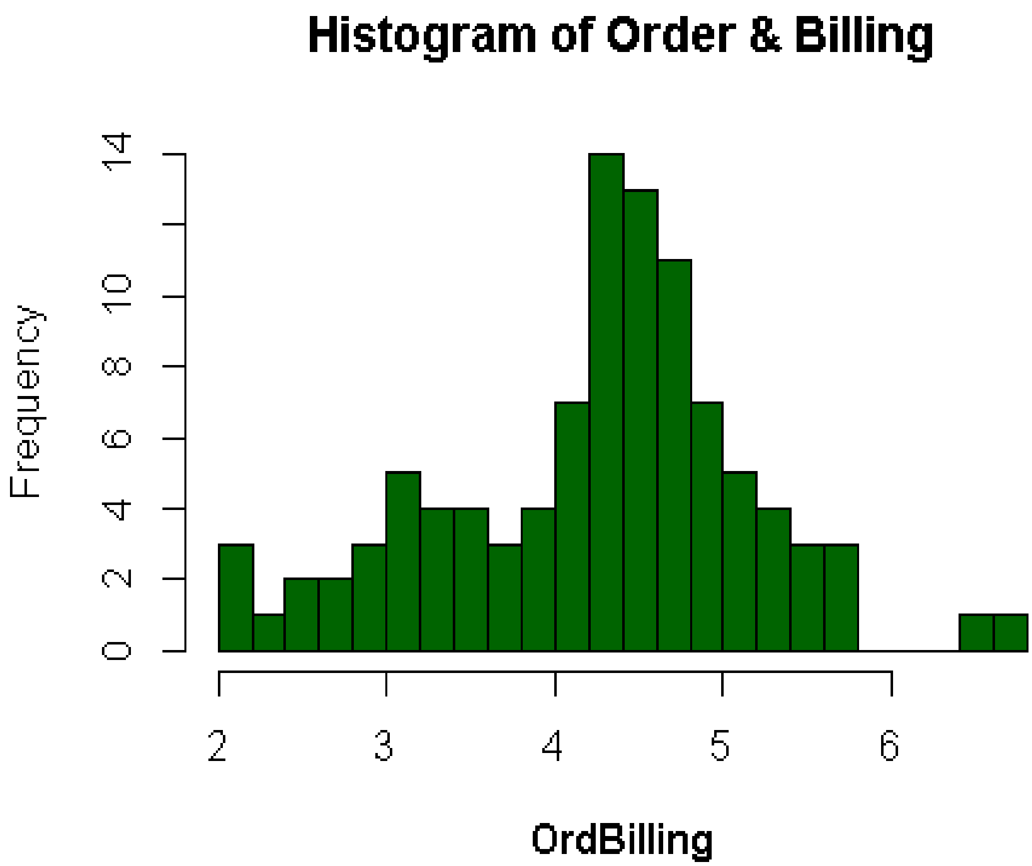
```
hist(WartyClaim,col = "Dark Blue",breaks = 20,main = "Histogram of Warranty &
Claim",xlab = expression(bold(WartyClaim)))
```



- From the Plot, the Warranty & Claim range lies from 4.1 to 8.1 as most values tend to be around 6.
- It's an evenly Spread allover

Frequency Distribution of Order & Billing

```r
hist(OrdBilling,col = "Dark Green",breaks = 20,main = "Histogram of Order & B
illing",xlab = expression(bold(OrdBilling)))
```



- Here, the Order and Billing range lies from 2 to 6.7 as most values tend to be around 4.2.
- Clearly the Order and Billing range is below average as most values lies under 5.
- There is a possibility of Outlier also which we will check using boxplot later on.

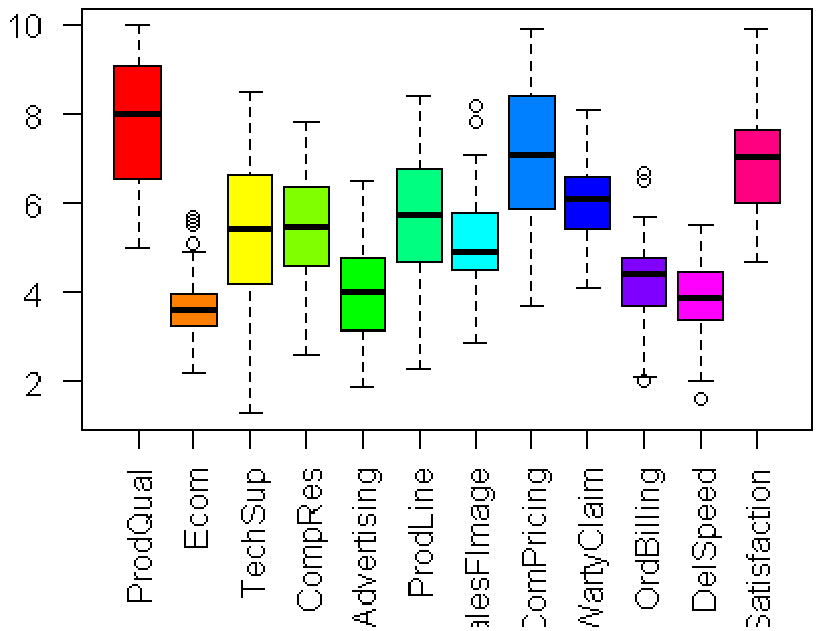
Frequency Distribution of Delivery Speed

```
hist(DelSpeed,col = "Dark Grey",breaks = 20,main = "Histogram of Delivery Spe
ed",xlab = expression(bold(DelSpeed)))
```

- Here, the Delivery Speed range lies from 1.6 to 5.5 as most values tend to be around 3.9.
- Clearly the Delivery Speed range is below average as most values lies under 5.
- There is a possibility of Outlier also which we will check using boxplot later on.

Outlier Detection

```r
boxplot(Hair[,2:13],col=rainbow(length(Hair[,2:13])),las=2)
```



- From the Plot it's Clear that the Variables E-Commerce, Sales Force Image, Order and Billing, Delivery Speed has Outliers, which we also seen in Histogram Plot.
- We need to deal with the outlier as it will affect our Model interpretability/Validity.
- We need to find the Outlier values using the formula (Q1-1.5IQR, Q1+1.5IQR)

Outlier Treatment

There are different methods in dealing with Outliers, one of the methods is **Winsorizing technique** which replaces the outliers with the allowable or Benchmark Range.

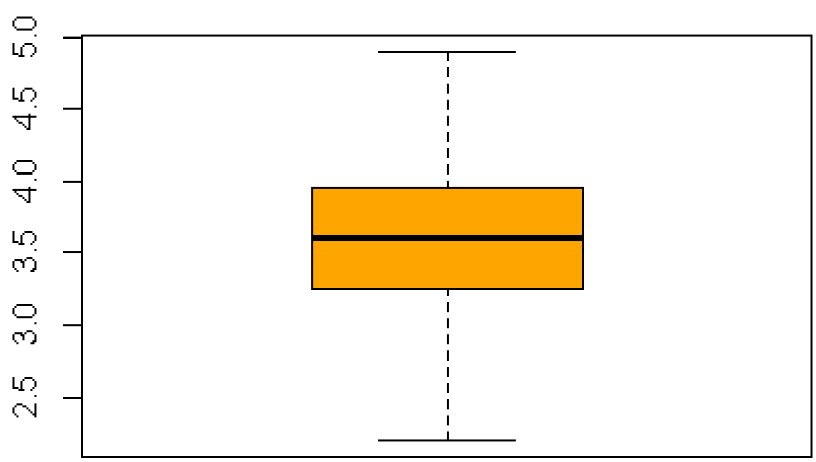
Outlier Formula = (Q1-1.5IQR, Q1+1.5IQR)

```r
summary(Hair$Ecom)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.200   3.275   3.600   3.672   3.925   5.700
```

```
benchmark <- 3.925+1.5*IQR(Hair$Ecom)
print(benchmark)

## [1] 4.9
```

Note: `Benchmark is the upper allowable Limit`

```
Hair$Ecom[Hair$Ecom>4.9] <- benchmark
summary(Hair$Ecom)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.200   3.275   3.600   3.645   3.925   4.900

boxplot(Hair$Ecom,col = "orange")
```

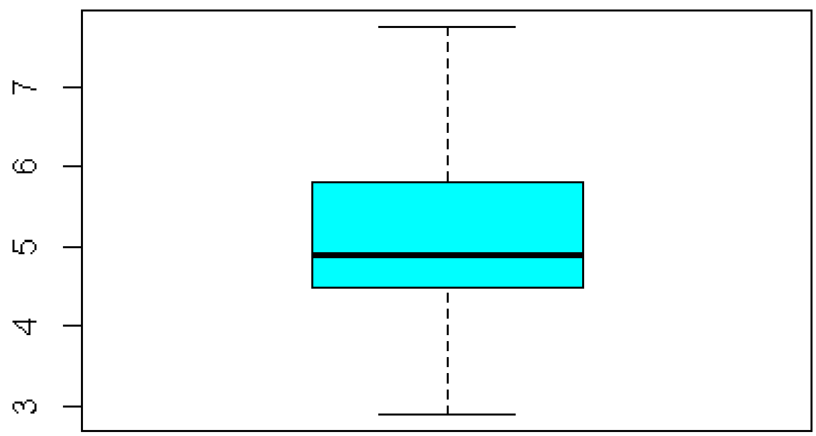


- We have replaced the Outliers with acceptable or allowable range which is evident in the boxplot

Now Let's do the same process for all other variables which has outliers,

```
summary(Hair$SalesFImage)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.900   4.500   4.900   5.123   5.800   8.200

benchmark1 <- 5.800+1.5*IQR(Hair$SalesFImage)
print(benchmark1)

## [1] 7.75

Hair$SalesFImage[Hair$SalesFImage>7.75] <- benchmark1
summary(Hair$SalesFImage)
```

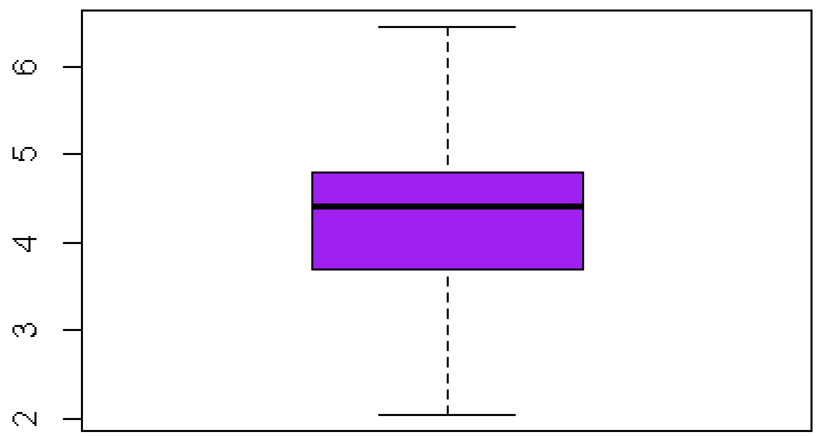
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   2.900   4.500   4.900   5.117   5.800   7.750

boxplot(Hair$SalesFImage,col = "cyan")
```



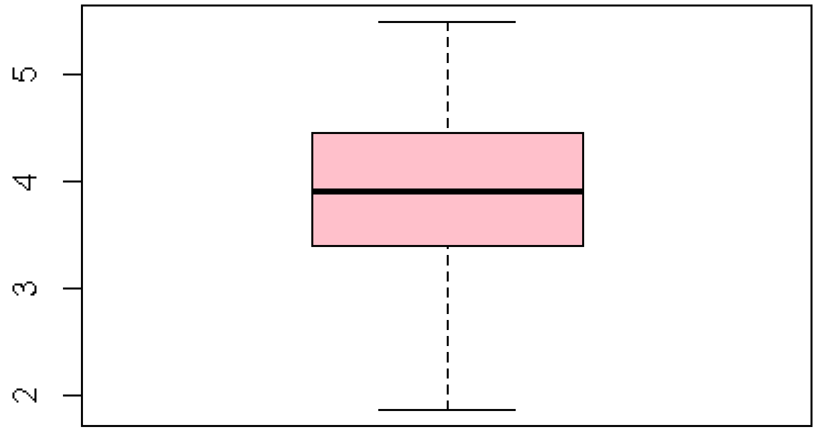

```
summary(Hair$OrdBilling)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   2.000   3.700   4.400   4.278   4.800   6.700

benchmark2 <- 4.800+1.5*IQR(Hair$OrdBilling)
print(benchmark2)

## [1] 6.45

Hair$OrdBilling[Hair$OrdBilling>6.45] <- benchmark2
l.benchmark2 <- 3.700-1.5*IQR(Hair$OrdBilling)
print(l.benchmark2)

## [1] 2.05
```

Note: l. benchmark2 means lower benchmark/ acceptable range as Order & Billing has outliers in both sides

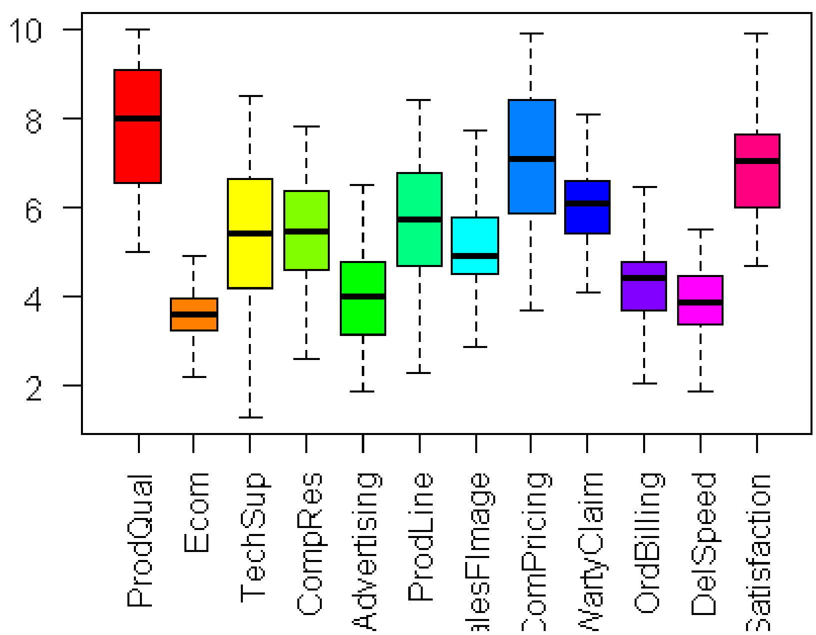
```
Hair$OrdBilling[Hair$OrdBilling<2.05] <- l.benchmark2
summary(Hair$OrdBilling)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   2.050   3.700   4.400   4.276   4.800   6.450

boxplot(Hair$OrdBilling,col = "Purple")
```




```
summary(Hair$DelSpeed)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.600   3.400   3.900   3.886   4.425   5.500

l.benchmark3 <- 3.400-1.5*IQR(Hair$DelSpeed)
print(l.benchmark3)

## [1] 1.8625

Hair$DelSpeed[Hair$DelSpeed<1.8625] <- l.benchmark3
summary(Hair$DelSpeed)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.863   3.400   3.900   3.889   4.425   5.500

boxplot(Hair$DelSpeed,col = "Pink")
```




```
boxplot(Hair[,2:13],col=rainbow(length(Hair[,2:13])),las = 2)
```



Multicollinearity

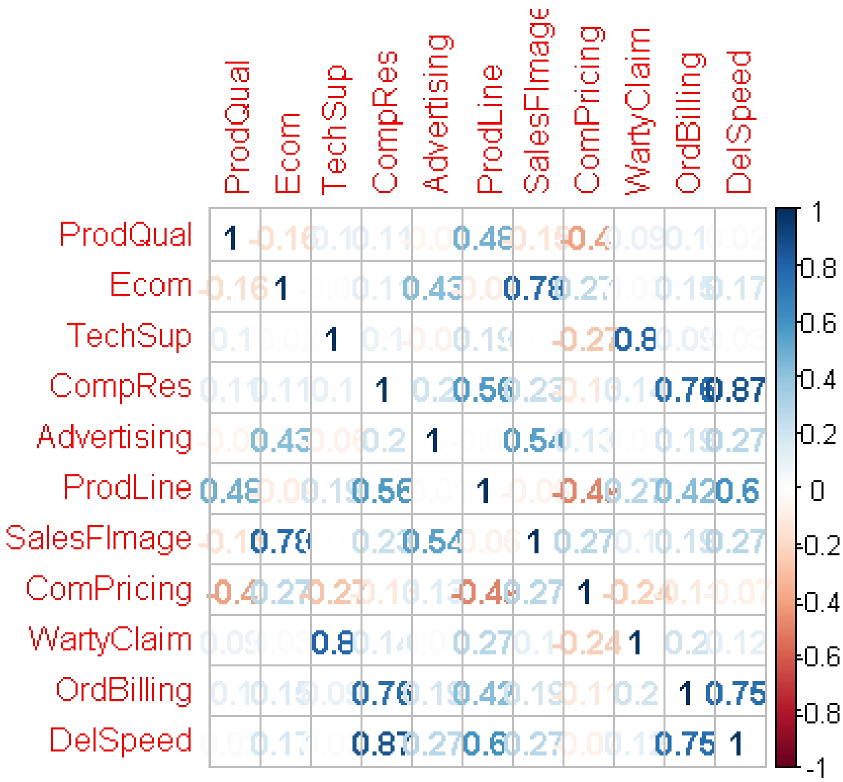
Now Let's Check Whether there is evidence of Multicollinearity

```
library(corrplot)

## corrplot 0.84 loaded

Cor <- cor(Hair[,c(-1,-13)])
print(Cor)

##                 ProdQual        Ecom      TechSup    CompRes Advertising
## ProdQual      1.00000000 -0.16090224  0.095600454  0.1063700 -0.05347313
## Ecom         -0.16090224  1.00000000 -0.019688320  0.1102404  0.42590967
## TechSup       0.09560045 -0.01968832  1.000000000  0.0966566 -0.06287007
## CompRes       0.10637000  0.11024041  0.096656598  1.0000000  0.19691685
## Advertising -0.05347313  0.42590967 -0.062870067  0.1969168  1.00000000
## ProdLine      0.47749341 -0.09467730  0.192625457  0.5614170 -0.01155082
## SalesFImage -0.14649835  0.77911050  0.009836493  0.2266473  0.54292307
## ComPricing  -0.40128188  0.26792718 -0.270786682 -0.1279543  0.13421689
## WartyClaim    0.08831231  0.02736822  0.797167926  0.1404083  0.01079207
## OrdBilling    0.10249539  0.14767899  0.085443472  0.7579955  0.18800483
## DelSpeed      0.02433152  0.17037182  0.028897804  0.8688463  0.27297319


##                 ProdLine  SalesFImage ComPricing  WartyClaim  OrdBilling
## ProdQual      0.47749341 -0.146498352 -0.4012819  0.08831231  0.10249539
## Ecom         -0.09467730  0.779110501  0.2679272  0.02736822  0.14767899
## TechSup       0.19262546  0.009836493 -0.2707867  0.79716793  0.08544347
## CompRes       0.56141695  0.226647292 -0.1279543  0.14040830  0.75799548
## Advertising -0.01155082  0.542923066  0.1342169  0.01079207  0.18800483
## ProdLine      1.00000000 -0.062583619 -0.4949484  0.27307753  0.42387000
## SalesFImage -0.06258362  1.000000000  0.2712463  0.10095251  0.19469492
## ComPricing  -0.49494840  0.271246288  1.0000000 -0.24498605 -0.11331825
## WartyClaim    0.27307753  0.100952514 -0.2449861  1.00000000  0.19810636
## OrdBilling    0.42387000  0.194694920 -0.1133182  0.19810636  1.00000000
## DelSpeed      0.60027249  0.271212797 -0.0702886  0.11616846  0.75229753


##                 DelSpeed
## ProdQual      0.02433152
## Ecom          0.17037182
## TechSup       0.02889780
## CompRes       0.86884630
## Advertising   0.27297319
## ProdLine      0.60027249
## SalesFImage   0.27121280
## ComPricing   -0.07028860
## WartyClaim    0.11616846
## OrdBilling    0.75229753
## DelSpeed      1.00000000
```

From the output, we can clearly see there is presence of **High Multi-Collinearity** between all the Independent Variables with each other

Let's check this with Corrplot

```
corrplot(Cor,method = "number")
```



- ✓ From the Plot, it's evident that there is High Multi-Collinearity between all the Independent Variables with each other
- ✓ Let's build a simple Linear Regression model for the dependent Variable with every independent variable and interpret the significance

Initial Simple Linear Regression analysis

Satisfaction Vs Product Quality model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 3.67593 + 0.41512 (Product Quality)
- ✓ Variance Explained by the model is 23% (0.2365)
- ✓ Product Quality is highly Significant @ 100% confidence Interval
- ✓ Product Quality has high impact on dependent variable Satisfaction

Satisfaction Vs E-Commerce model

```
model2 <- lm(Satisfaction~Hair$Ecom)
summary(model2)
```

```
## 
## Call:
## lm(formula = Satisfaction ~ Hair$Ecom)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.36925 -0.80338  0.06082  0.68785  2.41386 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   5.2679     0.6776   7.774 7.67e-12 ***
## Hair$Ecom     0.4527     0.1832   2.472   0.0152 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.162 on 98 degrees of freedom
## Multiple R-squared:  0.05868,    Adjusted R-squared:  0.04908 
## F-statistic: 6.109 on 1 and 98 DF,  p-value: 0.01517
```

Inference:

- ✓ Regression Equation Yhat = 5.2679 + 0.4427 (E-Commerce)
- ✓ Variance Explained by the model is 5% (0.05868) E-Commerce is not highly Significant
- ✓ E-commerce doesn't have high impact on dependent variable Satisfaction

Satisfaction Vs Technical Support model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 6.44757 + 0.08768 (TechSup)
- ✓ Variance Explained by the model is 1% (0.01268)
- ✓ Technical Support is **not Significant** Technical Support doesn't have any major impact on dependent variable Satisfaction

Satisfaction Vs Complaint Resolution model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 3.68005 + 0.59499 (CompRes)
- ✓ Variance Explained by the model is 36% (0.3639)
- ✓ Complaint Resolution is highly Significant @ 100% confidence Interval
- ✓ Complaint Resolution also has high impact on dependent variable Satisfaction

Satisfaction Vs Advertising model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 5.6259 + 0.3222 (Advertising)
- ✓ Variance Explained by the model is 9% (0.09282)
- ✓ Advertising is highly Significant @ 99% confidence Interval
- ✓ Advertising also has impact on dependent variable Satisfaction

Satisfaction Vs Product Line model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 4.02203 + 0.49887 (ProdLine)
- ✓ Variance Explained by the model is 30% (0.3031)
- ✓ Product Line is highly Significant @ 100% confidence Interval
- ✓ Product Line also has high impact on dependent variable Satisfaction

Satisfaction Vs Sales Figure Image model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 4.02836 + 0.56466 (SalesFImage)
- ✓ Variance Explained by the model is 25% (0.2511)
- ✓ Salesforce Image is highly Significant @ 100% confidence Interval
- ✓ Salesforce Image also has high impact on dependent variable Satisfaction

Satisfaction Vs Competitive Price model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 8.03856 - 0.16068 (ComPricing)
- ✓ Variance Explained by the model is 4% (0.04339)
- ✓ Competitive Pricing is highly Significant @ 95% confidence Interval
- ✓ Competitive Pricing also has some significant impact on dependent variable Satisfaction

Satisfaction Vs Warranty and Claim model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 5.3581 + 0.2581 (WartyClaim)
- ✓ Variance Explained by the model is 3% (0.03152)
- ✓ Warranty Claim is highly Significant @ 90% confidence Interval
- ✓ Warranty Claim also has less impact on dependent variable Satisfaction as it is less than the confidence level of 95%

Satisfaction Vs Order & Billing model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 4.0267 + 0.6762 (OrdBilling)
- ✓ Variance Explained by the model is 27% (0.2718)
- ✓ Order and Billing is highly Significant @ 100% confidence Interval
- ✓ Order and Billing also has high impact on dependent variable Satisfaction

Satisfaction Vs Delivery Speed model

```
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
```

Inference:

- ✓ Regression Equation Yhat = 3.2352 + 0.9471 (Delivery Speed)
- ✓ Variance Explained by the model is 33% (0.333)
- ✓ Delivery Speed is highly Significant @ 100% confidence Interval
- ✓ Delivery Speed also has high impact on dependent variable Satisfaction

Inference Out of all the Models

- ✓ Technical Support doesn't have much impact on the Satisfaction level of the customer
- ✓ All other independent Variables have significant impact on the dependent variable
- ✓ Notably, DelSpeed, SalesFimage, ProdQual, CompRes, ProdLine are Highly Significant and Impactful Variables

Since the pressence of multi-collinearity is evident, it's time to act on it by using one of the methods namely **Principal Component and Factor Analysis**

PCA/ Factor Analysis

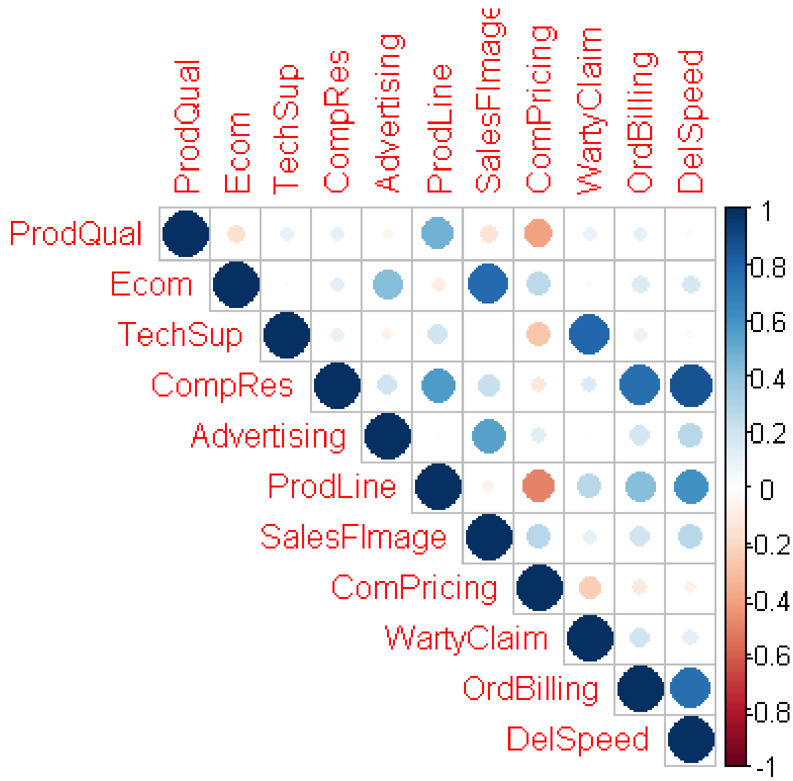
1st Step: The correlation matrix to check

```
Cor <- cor(Hair[,c(-1,-13)])
print(Cor)

##                ProdQual        Ecom      TechSup    CompRes Advertising
## ProdQual     1.00000000 -0.16090224  0.095600454  0.1063700 -0.05347313
## Ecom        -0.16090224  1.00000000 -0.019688320  0.1102404  0.42590967
## TechSup      0.09560045 -0.01968832  1.000000000  0.0966566 -0.06287007
## CompRes      0.10637000  0.11024041  0.096656598  1.0000000  0.19691685
## Advertising -0.05347313  0.42590967 -0.062870067  0.1969168  1.00000000
## ProdLine     0.47749341 -0.09467730  0.192625457  0.5614170 -0.01155082
## SalesFImage -0.14649835  0.77911050  0.009836493  0.2266473  0.54292307
## ComPricing  -0.40128188  0.26792718 -0.270786682 -0.1279543  0.13421689
## WartyClaim   0.08831231  0.02736822  0.797167926  0.1404083  0.01079207
## OrdBilling   0.10249539  0.14767899  0.085443472  0.7579955  0.18800483
## DelSpeed     0.02433152  0.17037182  0.028897804  0.8688463  0.27297319


##                ProdLine  SalesFImage ComPricing  WartyClaim  OrdBilling
## ProdQual     0.47749341 -0.146498352 -0.4012819  0.08831231  0.10249539
## Ecom        -0.09467730  0.779110501  0.2679272  0.02736822  0.14767899
## TechSup      0.19262546  0.009836493 -0.2707867  0.79716793  0.08544347
## CompRes      0.56141695  0.226647292 -0.1279543  0.14040830  0.75799548
## Advertising -0.01155082  0.542923066  0.1342169  0.01079207  0.18800483
## ProdLine     1.00000000 -0.062583619 -0.4949484  0.27307753  0.42387000
## SalesFImage -0.06258362  1.000000000  0.2712463  0.10095251  0.19469492
## ComPricing  -0.49494840  0.271246288  1.0000000 -0.24498605 -0.11331825
## WartyClaim   0.27307753  0.100952514 -0.2449861  1.00000000  0.19810636
## OrdBilling   0.42387000  0.194694920 -0.1133182  0.19810636  1.00000000
## DelSpeed     0.60027249  0.271212797 -0.0702886  0.11616846  0.75229753


##                DelSpeed
## ProdQual     0.02433152
## Ecom         0.17037182
## TechSup      0.02889780
## CompRes      0.86884630
```

```
## Advertising  0.27297319
## ProdLine     0.60027249
## SalesFImage  0.27121280
## ComPricing  -0.07028860
## WartyClaim   0.11616846
## OrdBilling   0.75229753
## DelSpeed     1.00000000
```

Its clear that all the variables are highly corelated with each other as we already saw it while checking multi-collinearity

```
library(corrplot)
corrplot(Cor,method = "circle",type = "upper")
```



Inference from Correlation Plot

Correlation coefficients greater than 0.3 in absolute value are indicative of acceptable correlations.

- Product Quality and Product Line are positively Correlated (0.477)
- Product Quality and Competitive Pricing are negatively Correlated (-0.401)
- E-commerce and Sales Force Image are very highly positively correlated (0.792)
- E-commerce and Advertising are positively correlated (0.429)
- Technical Support and Warranty Claim are Highly Positively Correlated (0.797)

- Complaint Resolution and Delivery Speed are highly positively correlated (0.865)
- Complaint Resolution and Order Billing are Highly positively Correlated (0.757)
- Complaint Resolution and Product Line are positively Correlated (0.561)
- Advertising and Sales Force Image are positively Correlated (0.542)
- Product Line and Competitive Pricing are negatively Correlated (-0.495)
- Product Line and Delivery Speed are positively Correlated (0.601)
- Product Line and Order & Billing are positively Correlated (0.424)
- Order Billing and Delivery Speed are Highly positively Correlated (0.751)

Step 2 - Factor Extraction using Kaizer Rule

Based on Correlation Matrix we confirmed the presence of Correlation and there is need for Factor Analysis

Now we need to figure out how many Factors are needed using scree plot

To decide on how many factors, we need to represent the data, we use 2 statistical criteria:

1. Eigen Values
2. The Scree Plot.

Eigen Value Computation

```
library(nFactors)

ev <- eigen(cor(Cor))
print(ev)
```

```
## eigen() decomposition
## $values
##  [1]  5.171396e+00  3.167489e+00  1.312628e+00  7.979254e-01  3.226377e-01
##  [6]  1.209992e-01  4.693274e-02  3.911689e-02  1.355644e-02  7.318543e-03
## [11] -3.347761e-16
##
## $vectors
##             [,1]        [,2]        [,3]        [,4]        [,5]
##  [1,]  0.3170823  0.09759696  0.39037492  0.52786804 -0.14969915
##  [2,] -0.3637262 -0.14126284 -0.25926863  0.27327668 -0.53153525
##  [3,]  0.2328364  0.33339566 -0.51688296 -0.03542737  0.04857354
##  [4,]  0.2349776 -0.45880184 -0.11050789 -0.13480920 -0.01723775
##  [5,] -0.2856075 -0.23990040 -0.16039359  0.47099886  0.75652337
##  [6,]  0.3954224 -0.17689530  0.04008828  0.22174943 -0.12400733
##  [7,] -0.3535673 -0.18594465 -0.31046063  0.28673590 -0.31538408
##  [8,] -0.3910260 -0.01735731  0.13340918 -0.47741347  0.03128249
##  [9,]  0.2387844  0.28826740 -0.56869880 -0.01813723  0.05808985
```

```
## [10,]  0.2215037 -0.44556926 -0.15144445 -0.17146797 -0.02212617
## [11,]  0.1892386 -0.49347436 -0.11461445 -0.12791607  0.01971759
##                [,6]        [,7]         [,8]        [,9]       [,10]
##  [1,]  0.352164830 -0.28575525 -0.091886090 -0.20301432  0.27675761
##  [2,]  0.027620334  0.39486310  0.342144535 -0.34626089  0.07375836
##  [3,]  0.027310857 -0.25823124  0.518003366  0.30562806  0.13437988
##  [4,] -0.048211448 -0.47374162  0.209385082 -0.47321721 -0.45551794
##  [5,]  0.009821877  0.07568799  0.049459993 -0.06214891 -0.01250684
##  [6,] -0.568555678  0.30434626 -0.172956358  0.25505142 -0.25327211
##  [7,] -0.019233413 -0.49583380 -0.396156906  0.38786619 -0.06582369
##  [8,] -0.027704544 -0.11689986 -0.214139051 -0.10573562  0.12188344
##  [9,]  0.044479994  0.12488780 -0.563022247 -0.44311443  0.04176640
## [10,]  0.689721018  0.30652804 -0.100790914  0.30586889 -0.09296676
## [11,] -0.264506668 -0.06232270 -0.003947722 -0.03324835  0.77364825
##               [,11]
##  [1,]  0.323353999
##  [2,]  0.155687058
##  [3,]  0.341509678
##  [4,]  0.039686907
##  [5,]  0.169387660
##  [6,]  0.413760074
##  [7,] -0.066408111
##  [8,]  0.716779523
##  [9,]  0.009892816
## [10,]  0.132461249
## [11,] -0.131466615
```

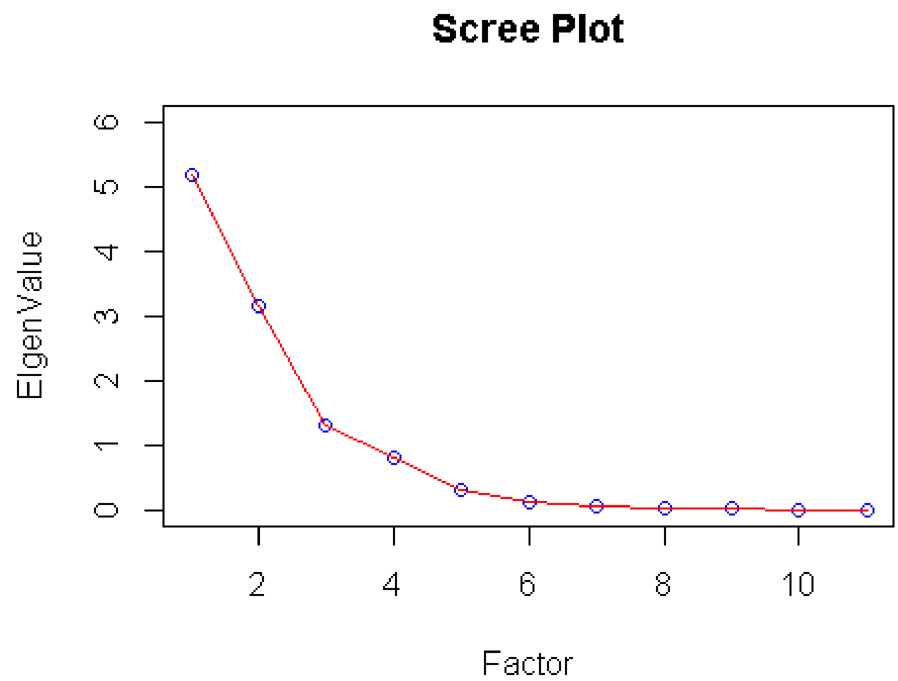
```
EIgenValue <- ev$values
print(EIgenValue)
```

```
##  [1]  5.171396e+00  3.167489e+00  1.312628e+00  7.979254e-01  3.226377e-01
##  [6]  1.209992e-01  4.693274e-02  3.911689e-02  1.355644e-02  7.318543e-03
## [11] -3.347761e-16
```

- ✓ The determination of the number of factors is usually done by considering only factors with Eigen values greater than 1.
- ✓ Factors with a variance less than 1 are no better than a single variable, since each variable is expected to have a variance of 1.
- ✓ Here we have 4 eigen Values Greater than 1 so we can have 4 Factors.
- ✓ Let's check it with scree plot also.

Scree Plot

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



In Scree Plot we can clearly see the decreasing trend in the Eigen Value as the Factor Increases.

As per **Kaizer Normalization Rule**,

"Any Eigen Value less than 1 is not a much of a Value to be considered"

- So as per this rule we can omit Eigen Values which are Less than 1.
- i.e., we have 4 Factors to be extracted (as it is Given in Problem Statement)

Now let's find out the Component matrix using Principle Component Analysis

```
library(psych)
Unrotate <- principal(Hair[,c(-1,-13)],nfactors = 4,rotate = "none")
print(Unrotate,digits = 3)

## Principal Components Analysis
## Call: principal(r = Hair[, c(-1, -13)], nfactors = 4, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                 PC1    PC2    PC3    PC4    h2     u2   com
## ProdQual      0.255 -0.493 -0.088  0.676 0.772 0.2280 2.19
## Ecom          0.247  0.741  0.296  0.267 0.770 0.2305 1.86
## TechSup       0.298 -0.364  0.798 -0.188 0.894 0.1062 1.85
## CompRes       0.874  0.051 -0.267 -0.213 0.884 0.1164 1.32
## Advertising   0.326  0.585  0.118  0.346 0.582 0.4178 2.37
## ProdLine      0.724 -0.439 -0.153  0.211 0.786 0.2143 1.96
## SalesFImage   0.351  0.758  0.308  0.244 0.853 0.1471 2.03
## ComPricing   -0.291  0.659 -0.060 -0.342 0.640 0.3602 1.95
## WartyClaim    0.399 -0.300  0.783 -0.178 0.893 0.1070 1.95
```

```
## OrdBilling   0.813  0.069 -0.204 -0.238 0.765 0.2350 1.32
## DelSpeed     0.878  0.139 -0.289 -0.208 0.916 0.0836 1.39
##
##                         PC1   PC2   PC3   PC4
## SS loadings           3.409 2.586 1.677 1.082
## Proportion Var        0.310 0.235 0.152 0.098
## Cumulative Var        0.310 0.545 0.697 0.796
## Proportion Explained  0.389 0.295 0.192 0.124
## Cumulative Proportion 0.389 0.685 0.876 1.000
##
## Mean item complexity =  1.8
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.059
##  with the empirical chi square  38.833  with prob <  0.00189
##
## Fit based upon off diagonal values = 0.968
```

Insights

- Eigen Value (SS Loadings) represents the largest Variance Summarized / reduction Proportion
- Variance explains how much variance is explained by each Component
- PC1 explains 31.2%, PC2 explains 23.2%, PC3 explains 15.4%, PC4 explains 9.9% of Variance
- **Cumulative Proportion** explains total Variance explained by all the components
- Totally we have explained 100 % of information with 4 Variables. That is the hallmark of Factor Analysis
- **Communality** is the common Variance captured by all the Factors Communality gives how much variance explained by all the 4 Factors of every Variable.
- eg.,91.4% of variance of Delivery Speed is explained by all the 4 Factors
- The Interpretability of Components in Unrotated Factor Loading is Difficult as many variables have High Loading (Corelated) in each Factors.
- It creates an amalgamation Effect of Variables in all the Factor

Step 3 - Factor Rotation

- Un-rotated factors are typically not very interpretable (most factors are correlated with many variables).
- Factors are rotated to make them more meaningful and easier to interpret (each variable is associated with a minimal number of factors).

- Different rotation methods may result in the identification of somewhat different factors.
- The most popular rotational method is **Varimax rotations.**
- Varimax use orthogonal rotations yielding uncorrelated factors/components. Varimax attempts to minimize the number of variables that have high loadings on a factor. This enhances the interpretability of the factors.

```
Rotate <- principal(Hair[,c(-1,-13)],nfactors = 4,rotate = "varimax")
print(Rotate,digits = 3)

## Principal Components Analysis
## Call: principal(r = Hair[, c(-1, -13)], nfactors = 4, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                RC1    RC2    RC3    RC4    h2     u2  com
## ProdQual     0.000 -0.012 -0.030  0.878 0.772 0.2280 1.00
## Ecom         0.042  0.862  0.031 -0.155 0.770 0.2305 1.07
## TechSup      0.020 -0.030  0.940  0.098 0.894 0.1062 1.02
## CompRes      0.928  0.107  0.047  0.091 0.884 0.1164 1.05
## Advertising  0.141  0.746 -0.075  0.029 0.582 0.4178 1.10
## ProdLine     0.593 -0.076  0.145  0.638 0.786 0.2143 2.13
## SalesFImage  0.135  0.898  0.072 -0.151 0.853 0.1471 1.12
## ComPricing  -0.087  0.241 -0.245 -0.717 0.640 0.3602 1.51
## WartyClaim   0.112  0.049  0.932  0.102 0.893 0.1070 1.06
## OrdBilling   0.862  0.111  0.086  0.039 0.765 0.2350 1.06
## DelSpeed     0.941  0.171  0.000  0.048 0.916 0.0836 1.07
##
##                         RC1   RC2   RC3   RC4
## SS loadings           2.902 2.226 1.854 1.772
## Proportion Var        0.264 0.202 0.169 0.161
## Cumulative Var        0.264 0.466 0.635 0.796
## Proportion Explained  0.331 0.254 0.212 0.202
## Cumulative Proportion 0.331 0.586 0.798 1.000
##
## Mean item complexity =  1.2
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.059
##  with the empirical chi square  38.833  with prob <  0.00189
##
## Fit based upon off diagonal values = 0.968
```

Interpretating Factors

After rotation,

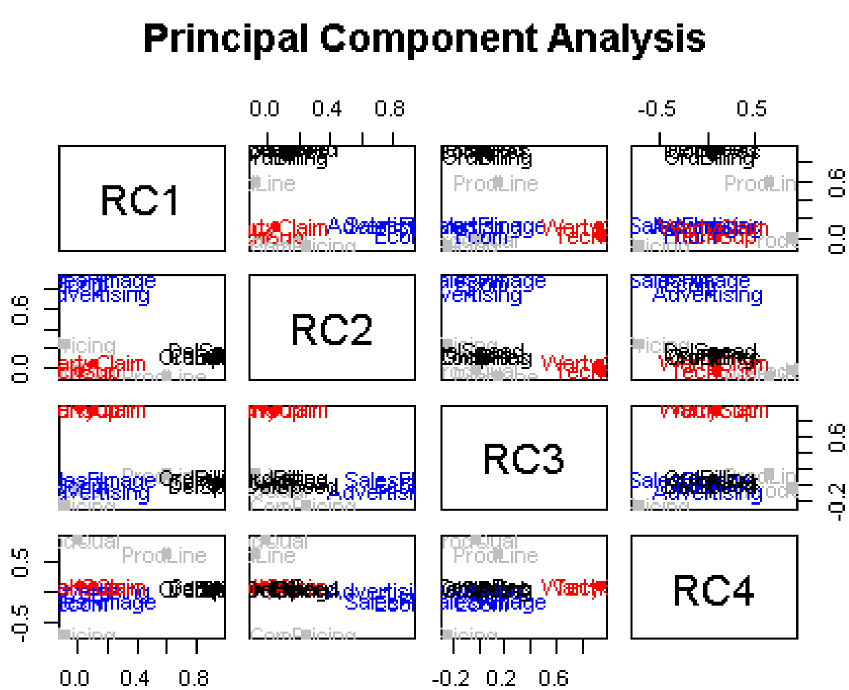
1. To identify factors, we need to group variables that have larger loadings for the same factor
2. Name the factors based on the meaning of Variables.
3. The Factor Names are Purely Intuitional Variables.

Labelling the Factors

- **Factor 1** is strongly correlated with Complaint Resolution (0.926), Order and Billing (0.864), Delivery Speed (0.938)
- This means this factor represents people who care about the "Customer Services". Therefore, we can Call it **Customer Service Factor**.
- In **Factor 2** is more correlated with E-Commerce (0.871), Salesforce Image (0.900), Advertising (0.742), competitive pricing (0.226)
- This means this factor represents people who care about the "Marketing Strategies". Therefore, we can Call it **Marketing Factor.**
- **Factor 3** is explaining a lot about Technical Support (0.939), warranty claim (0.931)
- This means this factor represents people who care about the "Customer Support". Therefore, we can Call it **Customer Support Factor.**
- **Factor 4** is explaining more about Product Quality (0.876), Product Line (0.642)
- This means this factor represents people who care about the "Product and its quality". Therefore, we can Call it **Product Factor.**

Let's Check the Clustering of our Factors Visually

```
plot(Rotate,row.names(Rotate$loadings),cex=1.0)
```



Multiple Linear Regression analysis using the factors as Independent Variables

In order to do Multiple Linear Regression using the Factors as Independent Variable first we need to Combine our Factor Scores and Dependent Variable and form a Data Frame.

```
MLR <- data.frame(Rotate$scores,Satisfaction)
MLR.model <- lm(MLR$Satisfaction~RC1+RC2+RC3+RC4,data = MLR)
summary(MLR.model)

##
## Call:
## lm(formula = MLR$Satisfaction ~ RC1 + RC2 + RC3 + RC4, data = MLR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6168 -0.5051  0.1029  0.4565  1.5340
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.91800    0.07086  97.632  < 2e-16 ***
## RC1          0.62185    0.07121   8.732 8.45e-14 ***
## RC2          0.49922    0.07121   7.010 3.44e-10 ***
## RC3          0.06961    0.07121   0.978    0.331
## RC4          0.54583    0.07121   7.665 1.52e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7086 on 95 degrees of freedom
## Multiple R-squared:  0.6608, Adjusted R-squared:  0.6465
## F-statistic: 46.27 on 4 and 95 DF,  p-value: < 2.2e-16
```

Model Interpretation

Regression Equation Yhat = 6.91800+0.61805(x1) +0.50973(x2) +0.06714(x3) +0.54032(x4) where,

X1 - Customer Service Factor,

X2 - Marketing Factor

X3 - Customer Support Factor

X4 - Product Factor

Understanding Slope

- If the Customer Service Factor score is increased by 1 unit, then the Satisfaction Rating is estimated to increase by 0.62 rating keeping Marketing Factor, Customer Support Factor and Product Factor as constant.

- ❖ If the Marketing Factor Score is increased by 1 unit, then the Satisfaction Rating is estimated to increase by 0.50 rating keeping Customer Service Factor score, Customer Support Factor and Product Factor as constant.
- ❖ If the Customer Support Factor Score is increased by 1 unit, then the Satisfaction Rating is estimated to increase by 0.067 rating keeping Customer Service Factor, Marketing Factor and Product Factor as constant.
- ❖ If the Product Factor Score is increased by 1 unit, then the Satisfaction Rating is estimated to increase by 0.54 rating keeping Customer Service Factor, Marketing Factor and Customer Support Factor as constant.

From the MLR model summary it's clear that the **Customer Support Factor** is **very less significant** as its is less impactfull in predicting Dependent Variable Satisfaction.

Model Performance Metrics

R-Square Interpretation

- ❖ Multiple R- Squared - 0.6605 implies 66.05% of the Variation in Satisfaction Rating is explained by the 4 Factor Scores.

Significance of the Model (Is there a Regression in the Population)

- ❖ Regression has 4 DF and Total has 99 DF.
- ❖ Hence error or residual has 99-4 = 95 DF Probability(F>46.21) = 2.2e-16(P-Value) is much smaller than alpha of 5% level
- ❖ Hence Reject the Null hypothesis of all betas are zero.
- ❖ Conclude at least one beta is non-zero and hence accept the alternative hypothesis.
- ❖ Therefore, overall **there is Overwhelming Evidence that regression model exists in the Population.**
- ❖ Individual Coefficients are **highly significant except Customer Support (RC3)** as evidence by the T-Stats thats have extremely low P Values each one of them less than alpha of 5% except Customer Support as its P Values are greater than alpha of 5%
- ❖ **All Factors Considered in the Regression Model is Robust** and can be used for Predictive Analytics.