Mini Project – Factor-Hair

Table of Content

1. Project Objective

The objective of this report work is to analyze the customer satisfaction on the basis of various variables or parameters viz. Product Quality, E-commerce, Technical Support, Complaint Resolution, Advertising, Product Line, Sales force Image, Competitive Pricing, Warranty and Claims, Order and Billing and Delivery Speed and to generate the insights from the data. The report will consist of the following:

- Importing the data set in R.
- Analyzing data set.
- Graphical representation.
- Building Simple Linear Regression Model.
- Identification of multi-colinearity in the sample data.
- PCA/FA analysis, factoring the sample data into 4 factors.
- Building Multiple Linear Regression Model.
- Predicting the validity of the model.

2. Exploratory Data Analysis – Step by step approach

- Importing the data set in R.
- Checking the evidence of multi-colinearity.
- Building regression model.
- PCA/FA method to identify the factors with high colinearity.

2.1Environment Setup and Data Import

2.1.1 Install necessary packages and Invoke Libraries

This section install the necessary packages and libraries required to read the data file.

2.1.2 Set up working directory

Setting up of a working directory in the initializing stage of the R session helps the user to import and export data files as well as code files easily.

2.1.3 Import and Read the data set

In this section the data file was read by the R. If it is in .csv format, read.csv command is used likewise if data file is in excel format read_excel command is used to read the data file.

2.2 Variable Identification

str() – to observe the structure of sample data.

corrplot() – to plot the correlation value.

EigenValue() – to calculate the eigen value.

read.csv() – to read the data file.

boxplot() – to plot the boxplot.

Scree() – to plot the eigen values.

principal() – to find the principal components.

lm() – to build the linear regression model.

vif() – to calculate the variation inflation factor.

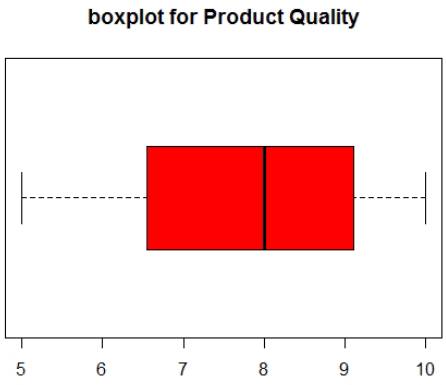
2.2.1 Variable Identification – Inferences

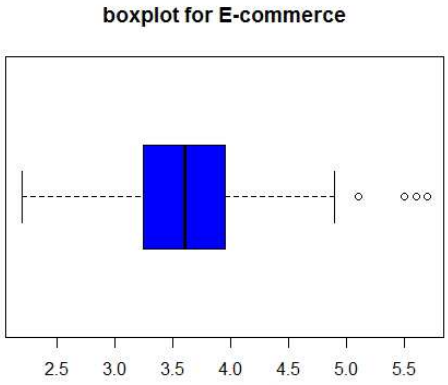
Problem-1

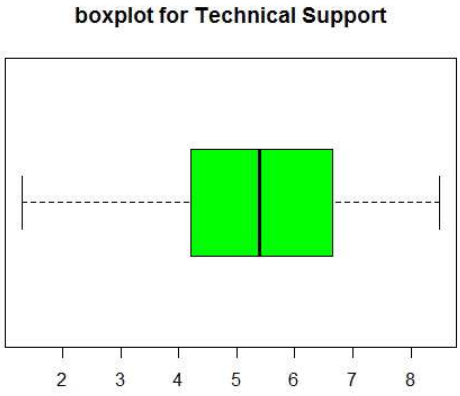
```
       ID            ProdQual          Ecom          TechSup         CompRes        Advertising   
 Min.   :  1.00   Min.   :5.000   Min.   :2.200   Min.   :1.300   Min.   :2.600   Min.   :1.900  
 1st Qu.: 25.75   1st Qu.: 6.575   1st Qu.:3.275   1st Qu.:4.250   1st Qu.:4.600   1st Qu.:3.175  
 Median : 50.50   Median : 8.000   Median :3.600   Median :5.400   Median :5.450   Median :4.000  
 Mean   : 50.50   Mean   : 7.810   Mean   :3.672   Mean   :5.365   Mean   :5.442   Mean   :4.010  
 3rd Qu.: 75.25   3rd Qu.: 9.100   3rd Qu.:3.925   3rd Qu.:6.625   3rd Qu.:6.325   3rd Qu.:4.800  
 Max.   :100.00   Max.   :10.000   Max.   :5.700   Max.   :8.500   Max.   :7.800   Max.   :6.500  
     ProdLine        SalesFImage      ComPricing      WartyClaim      OrdBilling       DelSpeed    
 Min.   :2.300   Min.   :2.900   Min.   :3.700   Min.   :4.100   Min.   :2.000   Min. :1.600  
 1st Qu.:4.700   1st Qu.:4.500   1st Qu.:5.875   1st Qu.:5.400   1st Qu.:3.700   1st Qu.:3.400  
 Median :5.750   Median :4.900   Median :7.100   Median :6.100   Median :4.400   Median :3.900  
 Mean   :5.805   Mean   :5.123   Mean   :6.974   Mean   :6.043   Mean   :4.278   Mean   :3.886  
 3rd Qu.:6.800   3rd Qu.:5.800   3rd Qu.:8.400   3rd Qu.:6.600   3rd Qu.:4.800   3rd Qu.:4.425  
 Max.   :8.400   Max.   :8.200   Max.   :9.900   Max.   :8.100   Max.   :6.700   Max.   :5.500  
```

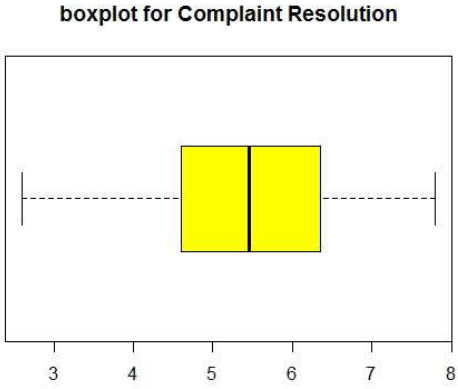
Satisfaction

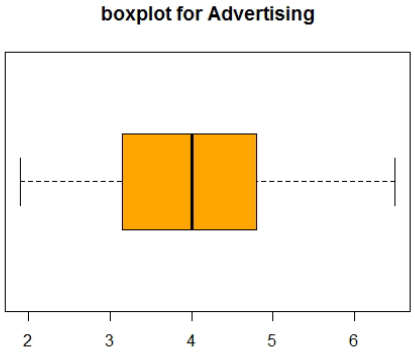
```
Min.   :4.700
1st Qu.:6.000
Median :7.050
Mean   :6.918
3rd Qu.:7.625
Max.   :9.900
```

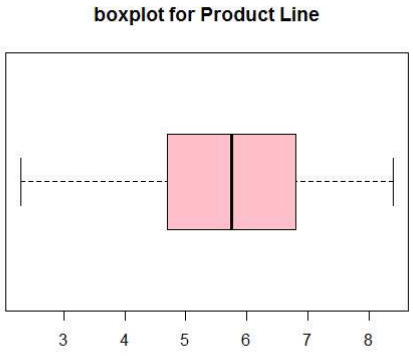


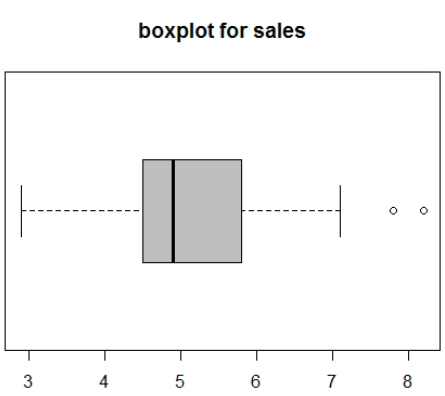


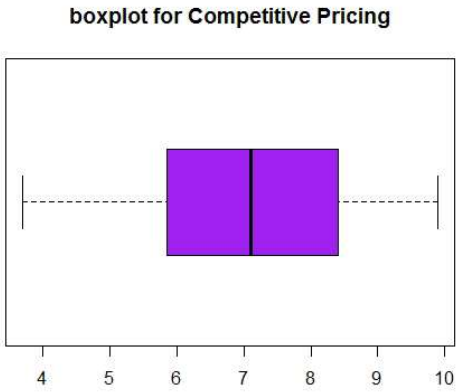


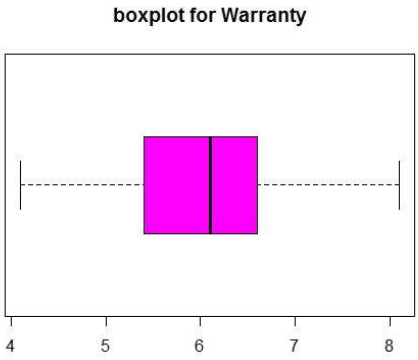


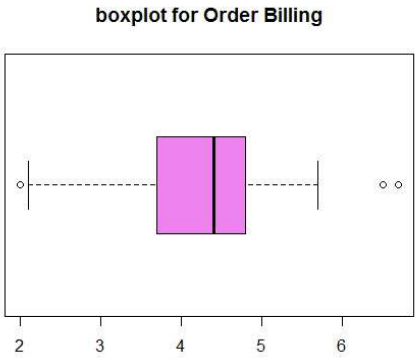


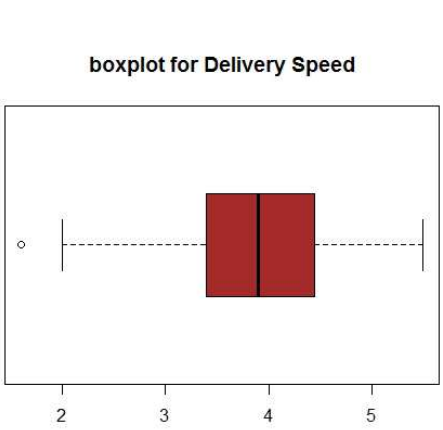


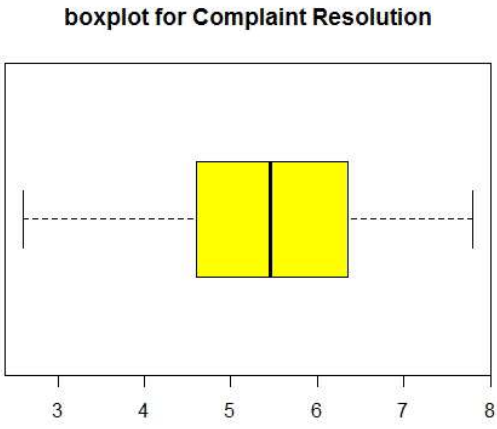

boxplot for sales
3
4
5
6
7
8

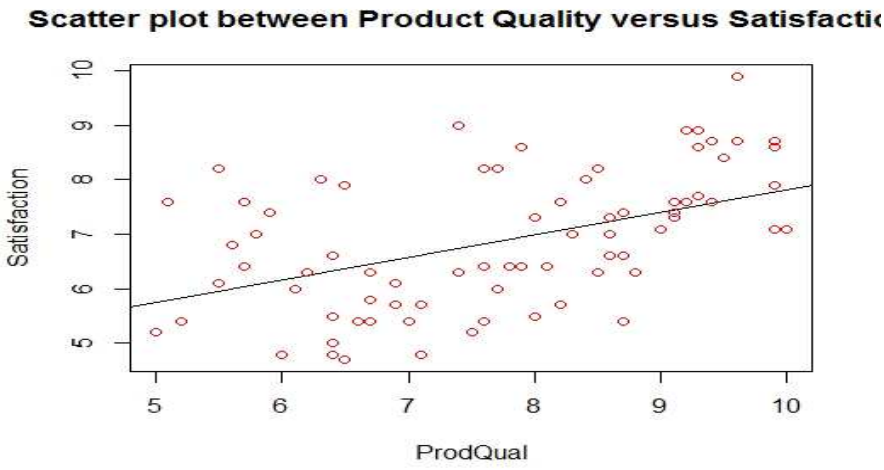

boxplot for Competitive Pricing
4
5
6
7
8
9
10

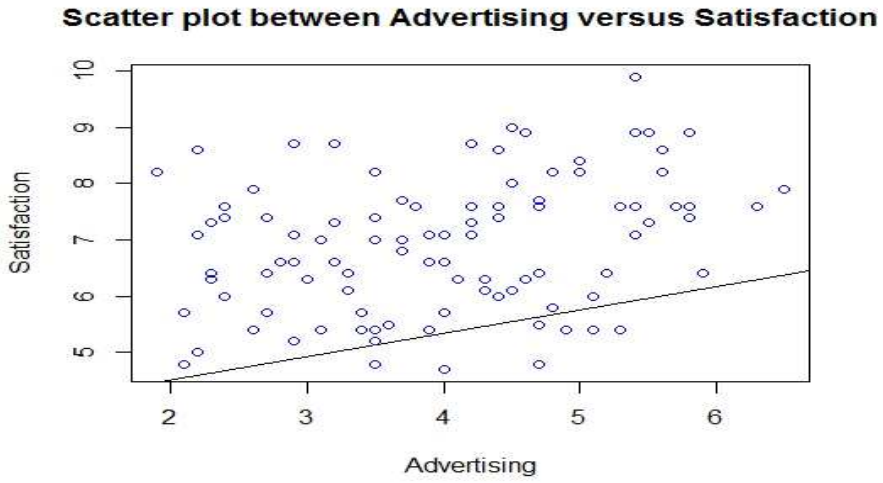

boxplot for Warranty
4
5
6
7
8


boxplot for Order Billing
2
3
4
5
6


boxplot for Delivery Speed
2
3
4
5


boxplot for Complaint Resolution
3
4
5
6
7
8





In the given sample data, boxplot of all the variables are plotted which shows the evidence of outlier in the data E-commerce, Sales Image Factor and Order billing.

Problem – 2

The simple linear regression model was build with each independent variable, results obtained shows multiple R squared value as:

ProdQual = 23.65%

Ecom = 7.994%

CompRes = 36.39%

TechSup = 1.268%

ProdLine = 30.31%

Advertising = 9.282%

SalesFImage = 25.025%

ComPricing = 4.339%

WartyClaim = 3.152%

OrdBilling = 27.22%

DelSpeed = 33.3%

Multiple R squared value shows that there is evidence of multi-colinearity in the data. However the percentage variance among the variables is not much higher, variables such as ProdQual, CompRes, ProdLine, SalesFImage, OrdBilling and DelSpeed shows certain amount of multi-colinearity.

Problem-3

Simple Linear regression model was build with each independent variable:

```
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



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



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



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



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



Call:
```

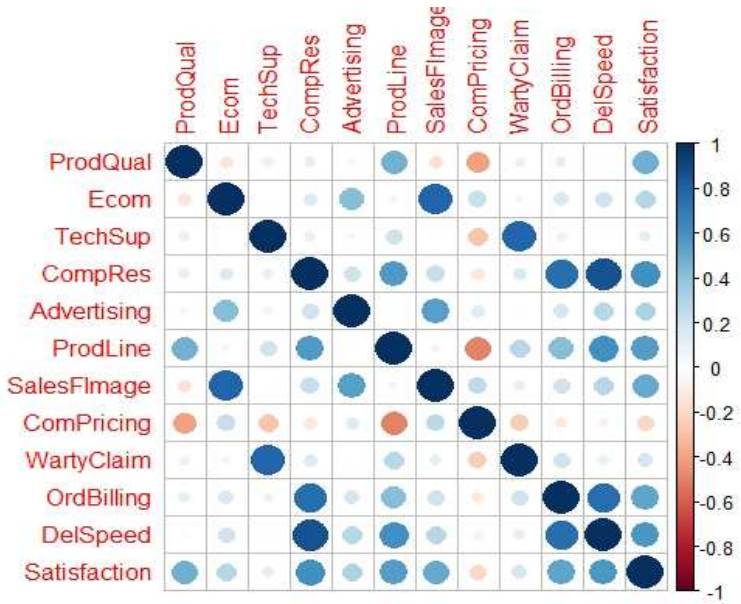
```
lm(formula = Satisfaction ~ ProdLine)

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



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



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



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



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



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
```

Problem-4

PCA or Factor Analysis without considering target variable

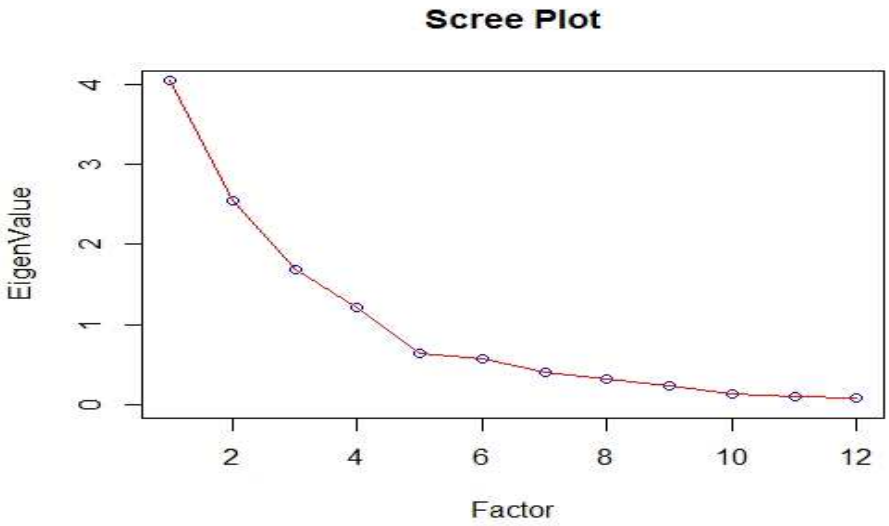
Correlation Plot



Eigen Values

4.04285997 2.55292440 1.69222417 1.21754639 0.63596293 0.56853132 0.40282774 0.32448016 0.23613948 0.14422355 0.09913845 0.08314143

Scree Plot



Unrotated Principal Component Factors

```
Call: principal(r = mydata, nfactors = 4, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix
                 PC1     PC2     PC3     PC4     h2      u2   com
ProdQual      0.3188 -0.5003 -0.0957  0.6776 0.8202 0.17978 2.370
Ecom          0.3341  0.7040  0.3077  0.2166 0.7488 0.25119 2.083
TechSup       0.2516 -0.3807  0.8017 -0.1980 0.8902 0.10976 1.802
CompRes       0.8498  0.0021 -0.2558 -0.3086 0.8828 0.11717 1.457
Advertising   0.3635  0.5708  0.1169  0.2273 0.5232 0.47675 2.165
ProdLine      0.7094 -0.4758 -0.1447  0.1104 0.7628 0.23717 1.907
SalesFImage   0.4382  0.7428  0.3134  0.2201 0.8905 0.10952 2.244
ComPricing   -0.2711  0.6675 -0.0672 -0.2657 0.5942 0.40583 1.690
WartyClaim    0.3519 -0.3214  0.7876 -0.2092 0.8912 0.10876 1.924
OrdBilling    0.7801  0.0145 -0.2017 -0.3384 0.7640 0.23603 1.515
DelSpeed      0.8492  0.0870 -0.2836 -0.3199 0.9114 0.08856 1.547
Satisfaction  0.8305  0.0382 -0.0374  0.3654 0.8261 0.17393 1.383

                       PC1    PC2    PC3    PC4
SS loadings         4.0429 2.5529 1.6922 1.2175
Proportion Var      0.3369 0.2127 0.1410 0.1015
Cumulative Var      0.3369 0.5496 0.6907 0.7921
Proportion Explained  0.4253 0.2686 0.1780 0.1281
Cumulative Proportion 0.4253 0.6939 0.8719 1.0000

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

The root mean square of the residuals (RMSR) is  0.0551
 with the empirical chi square  40.1477  with prob <  0.02062

Fit based upon off diagonal values = 0.9754
```

Rotated Principal Component Factors

```
Call: principal(r = mydata, nfactors = 4, rotate = "varimax")
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
```

```
Mean item complexity =  1.4
Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is  0.0551
 with the empirical chi square  40.1477  with prob <  0.02062

Fit based upon off diagonal values = 0.9754
```

Problem-5

Muliple Linear Regression

Initial regression model

Summary

```
lm(formula = Satisfaction ~ ., data = mydata)

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

Residual standard error: 0.5623 on 88 degrees of freedom
Multiple R-squared:  0.8021,   Adjusted R-squared:  0.7774
F-statistic: 32.43 on 11 and 88 DF,  p-value: < 2.2e-16
```

Variation Inflation Factor of initial regression model

```
ProdQual  Ecom     TechSup  CompRes  Advertising ProdLine SalesFImage  ComPricing
1.635797  2.756694 2.976796 4.730448 1.508933    3.488185  3.439420    1.635000

WartyClaim  OrdBilling  DelSpeed
3.198337    2.902999    6.516014
```

Factor analysis using principal axis factoring 4 factors

Correlation value after removing target variable (Satisfaction) from the sample data

	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage
ProdQual	1.00000000	-0.1371632174	0.0956004542	0.1063700	-0.05347313	0.47749341	-0.15181287
Ecom	-0.13716322	1.0000000000	0.0008667887	0.1401793	0.42989071	-0.05268784	0.79154371
TechSup	0.09560045	0.0008667887	1.0000000000	0.0966566	-0.06287007	0.19262546	0.01699054
CompRes	0.10637000	0.1401792611	0.0966565978	1.0000000	0.19691685	0.56141695	0.22975176
Advertising	-0.05347313	0.4298907110	-0.0628700668	0.1969168	1.00000000	-0.01155082	0.54220366
ProdLine	0.47749341	-0.0526878383	0.1926254565	0.5614170	-0.01155082	1.00000000	-0.06131553
SalesFImage	-0.15181287	0.7915437115	0.0169905395	0.2297518	0.54220366	-0.06131553	1.00000000
ComPricing	-0.40128188	0.2294624014	-0.2707866821	-0.1279543	0.13421689	-0.49494840	0.26459655
WartyClaim	0.08831231	0.0518981915	0.7971679258	0.1404083	0.01079207	0.27307753	0.10745534
OrdBilling	0.10430307	0.1561473316	0.0801018246	0.7568686	0.18423559	0.42440825	0.19512741
DelSpeed	0.02771800	0.1916360683	0.0254406935	0.8650917	0.27586308	0.60185021	0.27155126

	ComPricing	WartyClaim	OrdBilling	DelSpeed
ProdQual	-0.40128188	0.08831231	0.10430307	0.02771800
Ecom	0.22946240	0.05189819	0.15614733	0.19163607
TechSup	-0.27078668	0.79716793	0.08010182	0.02544069
CompRes	-0.12795425	0.14040830	0.75686859	0.86509170
Advertising	0.13421689	0.01079207	0.18423559	0.27586308
ProdLine	-0.49494840	0.27307753	0.42440825	0.60185021
SalesFImage	0.26459655	0.10745534	0.19512741	0.27155126
ComPricing	1.00000000	-0.24498605	-0.11456703	-0.07287173
WartyClaim	-0.24498605	1.00000000	0.19706512	0.10939460
OrdBilling	-0.11456703	0.19706512	1.00000000	0.75100307
DelSpeed	-0.07287173	0.10939460	0.75100307	1.00000000

Eigen values of sample data without considering target variable

4.04285997 2.55292440 1.69222417 1.21754639 0.63596293 0.56853132 0.40282774 0.32448016
0.23613948 0.14422355 0.09913845 0.08314143

Factor analysis using principal axis factoring 4 factors

```
Factor Analysis using method =  pa
Call: fa(r = mydata1, nfactors = 4, rotate = "none", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
              PA1   PA2   PA3   PA4   h2    u2 com
ProdQual     0.20 -0.41 -0.06  0.46 0.42 0.576 2.4
Ecom         0.29  0.66  0.27  0.22 0.64 0.362 2.0
TechSup      0.28 -0.38  0.74 -0.17 0.79 0.205 1.9
CompRes      0.86  0.01 -0.26 -0.18 0.84 0.157 1.3
Advertising  0.29  0.46  0.08  0.13 0.31 0.686 1.9
ProdLine     0.69 -0.45 -0.14  0.31 0.80 0.200 2.3
SalesFImage  0.39  0.80  0.35  0.25 0.98 0.021 2.1
ComPricing  -0.23  0.55 -0.04 -0.29 0.44 0.557 1.9
WartyClaim   0.38 -0.32  0.74 -0.15 0.81 0.186 2.0
```

```
OrdBilling   0.75  0.02 -0.18 -0.18 0.62 0.378 1.2
DelSpeed     0.90  0.10 -0.30 -0.20 0.94 0.058 1.4

                       PA1  PA2  PA3  PA4
SS loadings           3.21 2.22 1.50 0.68
Proportion Var        0.29 0.20 0.14 0.06
Cumulative Var        0.29 0.49 0.63 0.69
Proportion Explained  0.42 0.29 0.20 0.09
Cumulative Proportion 0.42 0.71 0.91 1.00

Mean item complexity =  1.9
Test of the hypothesis that 4 factors are sufficient.

The degrees of freedom for the null model are  55  and the objective function was  6.55
with Chi Square of  619.27
The degrees of freedom for the model are 17  and the objective function was  0.33

The root mean square of the residuals (RMSR) is  0.02
The df corrected root mean square of the residuals is  0.03

The harmonic number of observations is  100 with the empirical chi square  3.19  with
prob <  1
The total number of observations was  100  with Likelihood Chi Square =  30.27  with prob
<  0.024

Tucker Lewis Index of factoring reliability =  0.921
RMSEA index =  0.096  and the 90 % confidence intervals are  0.032 0.139
BIC =  -48.01
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                PA1  PA2  PA3  PA4
Correlation of (regression) scores with factors   0.98 0.97 0.95 0.88
Multiple R square of scores with factors          0.96 0.95 0.91 0.78
Minimum correlation of possible factor scores     0.92 0.90 0.82 0.56


Call: fa(r = mydata1, nfactors = 4, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
              PA1   PA2   PA3   PA4   h2    u2 com
ProdQual     0.02 -0.07  0.02  0.65 0.42 0.576 1.0
Ecom         0.07  0.79  0.03 -0.11 0.64 0.362 1.1
TechSup      0.02 -0.03  0.88  0.12 0.79 0.205 1.0
CompRes      0.90  0.13  0.05  0.13 0.84 0.157 1.1
Advertising  0.17  0.53 -0.04 -0.06 0.31 0.686 1.2
ProdLine     0.53 -0.04  0.13  0.71 0.80 0.200 1.9
SalesFImage  0.12  0.97  0.06 -0.13 0.98 0.021 1.1
ComPricing  -0.08  0.21 -0.21 -0.59 0.44 0.557 1.6
WartyClaim   0.10  0.06  0.89  0.13 0.81 0.186 1.1
OrdBilling   0.77  0.13  0.09  0.09 0.62 0.378 1.1
DelSpeed     0.95  0.19  0.00  0.09 0.94 0.058 1.1

                       PA1  PA2  PA3  PA4
SS loadings           2.63 1.97 1.64 1.37
Proportion Var        0.24 0.18 0.15 0.12
Cumulative Var        0.24 0.42 0.57 0.69
Proportion Explained  0.35 0.26 0.22 0.18
Cumulative Proportion 0.35 0.60 0.82 1.00
```

```
Mean item complexity =  1.2
Test of the hypothesis that 4 factors are sufficient.

The degrees of freedom for the null model are  55  and the objective function was  6.55
with Chi Square of  619.27
The degrees of freedom for the model are 17  and the objective function was  0.33

The root mean square of the residuals (RMSR) is  0.02
The df corrected root mean square of the residuals is  0.03

The harmonic number of observations is  100 with the empirical chi square  3.19  with
prob <  1
The total number of observations was  100  with Likelihood Chi Square =  30.27  with prob
<  0.024

Tucker Lewis Index of factoring reliability =  0.921
RMSEA index =  0.096  and the 90 % confidence intervals are  0.032 0.139
BIC =  -48.01
Fit based upon off diagonal values = 1
Measures of factor score adequacy
                                                 PA1  PA2  PA3  PA4
Correlation of (regression) scores with factors   0.98 0.99 0.94 0.88
Multiple R square of scores with factors          0.96 0.97 0.88 0.78
Minimum correlation of possible factor scores     0.93 0.94 0.77 0.55
```

Binding target variable (Satisfaction) and factors

Head Values

```
   Satisfaction  Rating      Price      OtherCompanyPrice  CustomerSupport
1           8.2 -0.1338871  0.9175166      -1.719604873      0.09135411
2           5.7  1.6297604 -2.0090053      -0.596361722      0.65808192
3           8.9  0.3637658  0.8361736       0.002979966      1.37548765
4           4.8 -1.2225230 -0.5491336       1.245473305     -0.64421384
5           7.1 -0.4854209 -0.4276223      -0.026980304      0.47360747
6           4.7 -0.5950924 -1.3035333      -1.183019401     -0.95913571
```

Building regression model on train data (70%)

```
lm(formula = Satisfaction ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6857 -0.4018  0.1051  0.4027  1.2036

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        6.92625    0.08263  83.827  < 2e-16 ***
Rating             0.62022    0.08408   7.377 3.73e-10 ***
Price              0.57735    0.08047   7.175 8.50e-10 ***
OtherCompanyPrice  0.09567    0.08667   1.104    0.274
CustomerSupport    0.66562    0.09374   7.101 1.15e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6814 on 65 degrees of freedom
Multiple R-squared:  0.7079,   Adjusted R-squared:   0.69
F-statistic: 39.39 on 4 and 65 DF,  p-value: < 2.2e-16
```

Testing the trained regression model

	Satisfaction	Rating	Price	OtherCompanyPrice	CustomerSupport	Satisfaction.predict
1	8.2	-0.13388710	0.9175166	-1.719604873	0.09135411	7.269232
2	5.7	1.62976040	-2.0090053	-0.596361722	0.65808192	7.158146
3	8.9	0.36376581	0.8361736	0.002979966	1.37548765	8.550469
4	4.8	-1.22252302	-0.5491336	1.245473305	-0.64421384	5.541333
5	7.1	-0.48542093	-0.4276223	-0.026980304	0.47360747	6.690958
7	5.7	-2.52885363	0.3883688	-0.603275803	-1.29659025	4.661277
14	7.6	0.18944710	-0.1200159	0.341391428	1.43748733	7.963941
21	5.4	-0.49662748	-0.4051355	1.413398115	-1.42620085	5.570249
23	7.0	-0.08593028	-0.2064799	0.954813784	1.29099542	7.704405
27	6.3	0.69040218	-1.2767621	0.980470048	0.97700863	7.361437

4. Conclusion

- This project report was build to analyze the impact of various factors such as ProdQual, ECom, TechSup, CompRes, Advertising, ProdLine, SalesFImage, ComPricing, WartyClaim, OrdBilling and DelSpeed on Customer Satisfaction.

- Exploratory data analysis shows the presence of outlier in the given sample data.

- The results showed the existence multi-colinearity among variables which is found to be not very significant as multiple R squared value is very low.

- Simple linear regression model was build which shows that p-value is low, which means our null hypothesis i.e. all betas (β) are zero is overwhelmingly rejected. The multiple R squared value shows the evidence of multi-colinearity.

- Principal component analysis or factor analysis method is applied to obtained the factors of the variable that are correlated. As per the Keiser rule all the variables which have eigen value greater than one are selected, from the data available and shown above four factors were obtained. **The four factors are named as Rating, Price, OtherCompanyPrice and CustomerSupport.**

- Multiple Linear regression Model was build considering Satisfaction as the dependent variable using the trained data (70% of the sample data).

- Having build the multiple linear regression model, it was tested on test data (30% of the sample data). Results obtained showed that the build model produces output which is approximately equal to the trained data output.

5. Appendix-A: Source Code

```
1.  ### Install Packages and Loading Library
2.  library(readr)
3.  ### Setting working directory
4.  setwd("D:/great learning/advanced statistics/Project-2")
5.  mydata_Hair = read.csv("Factor-Hair-Revised.csv", header=TRUE)
6.  attach(mydata_Hair)
7.  mydata_Hair
8.  ### Exploratory Data Analysis
9.  summary(mydata_Hair)
10. str(mydata_Hair)
11. boxplot(ProdQual, horizontal = TRUE, col = "Red", main="boxplot for Product Quality")
12. boxplot(Ecom, horizontal = TRUE, col = "Blue", main="boxplot for E-commerce")
13. boxplot(TechSup, horizontal = TRUE, col = "Green", main="boxplot for Technical Support")
14. boxplot(CompRes, horizontal = TRUE, col = "Yellow", main="boxplot for Complaint Resolution")
15. boxplot(Advertising, horizontal = TRUE, col = "Orange", main="boxplot for Advertising")
16. boxplot(ProdLine, horizontal = TRUE, col = "Pink", main="boxplot for Product Line")
17. boxplot(SalesFImage, horizontal = TRUE, col = "Grey", main="boxplot for sales")
```

```
boxplot(ComPricing, horizontal = TRUE, col = "Purple", main="boxplot for Competitive Pricing")
boxplot(WartyClaim, horizontal = TRUE, col = "Magenta", main="boxplot for Warranty")
boxplot(OrdBilling, horizontal = TRUE, col = "violet", main="boxplot for Order Billing")
boxplot(DelSpeed, horizontal = TRUE, col = "Brown", main="boxplot for Delivery Speed")
boxplot(Satisfaction, horizontal = TRUE, col = "Ivory", main="boxplot for Customer satisfaction")
plot(ProdQual,Satisfaction, col= "Red", abline(lm(Satisfaction~ProdQual)), main = "Scatter plot between Product Quality versus Satisfaction")
plot(Advertising,Satisfaction, col= "Blue", abline(lm(Satisfaction~ProdQual)), main = "Scatter plot between Advertising versus Satisfaction")
interaction.plot(ProdQual, Advertising, Satisfaction, col = c("Red","Blue"), main = "Interaction between Product Quality and Advertising")
### Simple Linear Regression Model and multicollinearity
Model1 = lm(Satisfaction~ProdQual)
summary(Model1)
Model2 = lm(Satisfaction~Ecom)
summary(Model2)
Model3 = lm(Satisfaction~TechSup)
summary(Model3)
Model4 = lm(Satisfaction~CompRes)
summary(Model4)
Model5 = lm(Satisfaction~Advertising)
summary(Model5)
Model6 = lm(Satisfaction~ProdLine)
```

```
summary(Model6)
Model7 = lm(Satisfaction~SalesFImage)
summary(Model7)
Model8 = lm(Satisfaction~ComPricing)
summary(Model8)
Model9 = lm(Satisfaction~WartyClaim)
summary(Model9)
Model10 = lm(Satisfaction~OrdBilling)
summary(Model10)
Model11 = lm(Satisfaction~DelSpeed)
summary(Model11)
### PCA or Factor Analysis without considering target variable
install.packages("nFactors")
install.packages("corrplot")
library(corrplot)
library(nFactors)
mydata = mydata_Hair[, 2:13]   ### Removing ID varaible
mydata
mydata_corr = cor(mydata)
mydata_corr
corrplot(mydata_corr)
ev = eigen(cor(mydata))
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")
library(psych)
unrotate = principal(mydata, nfactors = 4, rotate = "none")
print(unrotate,digits = 4)
unrotateProfile = plot(unrotate, row.names(unrotate$loadings))
Rotate = principal(mydata, nfactors = 4, rotate = "varimax")
print(Rotate, digits = 4)
RotateProfile = plot(Rotate, row.names(Rotate$loadings),cex = 1.0)
Rotate$scores
factor.scope(mydata, f = Rotate$loadings, method = "Harman")
### Multiple Linear Regression
install.packages("car")
library(car)
### Building initial regression model
MLM0 = lm(Satisfaction~.,mydata)
summary(MLM0)
vif(MLM0)
### Factor analysis using principal axis factoring 4 factors
install.packages("nFactors")
library(nFactors)
mydata1 = mydata[-c(12)]    ### Removing Satisfaction from mydata
mydata1
names(mydata1)
mydata1corr = cor(mydata1)
mydata1corr
ev1 = eigen(cor(mydata1))
ev1
EigenValue1 = ev$values
```

```
EigenValue1
solution = fa(r = mydata1, nfactors = 4, rotate = "none", fm = "pa")
solution
solution1 = fa(r = mydata1, nfactors = 4, rotate = "varimax", fm = "pa")
solution1
solution1$scores
### Binding target variable satisfaction and factors
mydata_bind = cbind(mydata[12], solution1$scores)
mydata_bind
### Labelling data
names(mydata_bind) = c("Satisfaction", "Rating", "Price", "OtherCompanyPrice", "CustomerSupport")
head(mydata_bind)
### Splitting data 70:30
set.seed(100)
indices = sample(1:nrow(mydata_bind), 0.70*nrow(mydata_bind))
train = mydata_bind[indices,]
test = mydata_bind[-indices,]
### Building regression on train data
MLM1 = lm(Satisfaction~.,data = train)
summary(MLM1)
vif(MLM1)
### Mean square error and RMSE
mse1 = mean(MLM1$residuals^2)
mse1
rmse1 = sqrt(mse1)
rmse1
### Testing the trained regression model
```

```
MLM2 = predict(MLM1, newdata = test, type = "response")
MLM2
test$Satisfaction.predict = MLM2
head(test,10)
### Mean square error and RMSE
mse2 = mean((test$Satisfaction-test$Satisfaction.predict)^2)
mse2
rmse2 = sqrt(mse2)
rmse2
```