Mini Project 2: Principal Component Analysis and Factor Analysis

Author- Amit KULKARNI

Table of Contents

1. Project Objective…….……..3.

2. Assumptions……...……3.

3. Exploratory Data Analysis…………………………………………………………………………………………………….....3.

3.1 Environmental set up………………………………………………………………………………………………………..3.

3.2 Variable Identification ………………………………………………………………………………………………………..3.

3.3 Univariate Analysis …………………………………………………………………………………………………….…3.

4. Provided Information of the Case………………………………………………………………………...…………………4.

5. Solution…….5.

5.1 Is there an evidence of multicollinearity? ……………………………………………………………………………….…5.

5.2 Performing factor analysis for up to 4 factors…………………………………………………………………………….…6.

5.3 Names of the factor ………………………………………………………………………………………………………..…9.

5.4 Regression using the 4 factors and Model Testing ………………………………………………………………………..…9.

6. Conclusion ……….…..10.

1. *Project Objective*:

The objective is to explore the Hair dataset in R and get insight about the dataset. The insight will include

i. To perform the regression analysis amongst the dependent and independent variables
ii. To check if there exists any multicollinearity
iii. To reduce the dimensions using the Principal Component Analysis and Factor Analysis
iv. To overall validate the model if it is Good Fit

2. *Assumptions: No specific assumptions done*

3. *Exploratory Data Analysis*: Various tools, techniques and data visualizations are used like

a. Linear Modelling
b. Identify multicollinearity via the corrplot diagram'
c. Identify the Eigenvalues via the Scree Plot
d. Identify the PCA up to 4 variables
e. Identify the factors using the FA diagram
f. Calculate the correlation using the factors identified
g. Test the model on independent test data and arrive at R-Squared values

3.1: Environmental Set Up and Data Imports:

a) Installation of the required packages and calling them via the library functions

i. Nfactors
ii. Car
iii. caTools
iv. psych
v. corrplot

b) R installed to calculate and evaluate the PCA and FA model

c) **Set Up of Working Directory**: All the codes, dataset files are stored on the below path which is being set up as the Working Directory

C:\Users\Amit Kulkarni\Documents\R Programming\Advance Statistic\project 2

d) **Import and Read the Dataset**: The given dataset is in the .csv format. Hence the command read.csv is used to import the file.

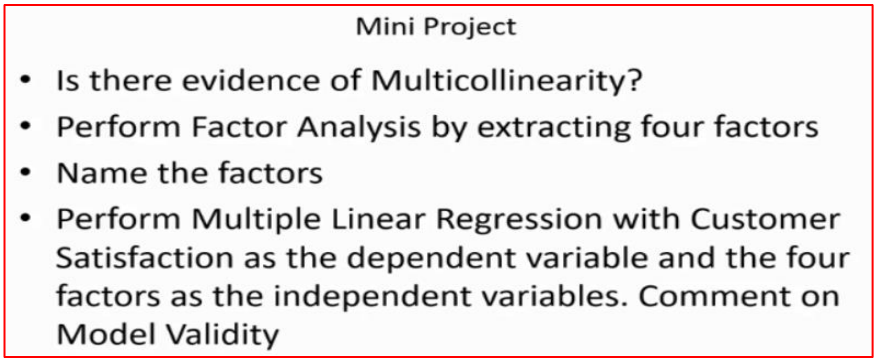
3.2: **Variable Identification – Inferences:** There are a total of 11 variables assumed to be independent of each other and 1 dependent variable. Test will be conducted to identify if there are 11 independent variables or there exists any multicollinearity between the independent variables.

3.3. **Univariate Analysis:** Univariate analysis of all the variables is given in Section 4 below.

4. *Provided information of the case:*

ID	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed	Satisfaction
1	8.5	3.9	2.5	5.9	4.8	4.9	6	6.8	4.7	5	3.7	8.2
2	8.2	2.7	5.1	7.2	3.4	7.9	3.1	5.3	5.5	3.9	4.9	5.7
3	9.2	3.4	5.6	5.6	5.4	7.4	5.8	4.5	6.2	5.4	4.5	8.9
4	6.4	3.3	7	3.7	4.7	4.7	4.5	8.8	7	4.3	3	4.8
5	9	3.4	5.2	4.6	2.2	6	4.5	6.8	6.1	4.5	3.5	7.1
6	6.5	2.8	3.1	4.1	4	4.3	3.7	8.5	5.1	3.6	3.3	4.7
7	6.9	3.7	5	2.6	2.1	2.3	5.4	8.9	4.8	2.1	2	5.7
8	6.2	3.3	3.9	4.8	4.6	3.6	5.1	6.9	5.4	4.3	3.7	6.3
9	5.8	3.6	5.1	6.7	3.7	5.9	5.8	9.3	5.9	4.4	4.6	7
10	6.4	4.5	5.1	6.1	4.7	5.7	5.7	8.4	5.4	4.1	4.4	5.5
11	8.7	3.2	4.6	4.8	2.7	6.8	4.6	6.8	5.8	3.8	4	7.4
12	6.1	4.9	6.3	3.9	4.4	3.9	6.4	8.2	5.8	3	3.2	6
13	9.5	5.6	4.6	6.9	5	6.9	6.6	7.6	6.5	5.1	4.4	8.4
14	9.2	3.9	5.7	5.5	2.4	8.4	4.8	7.1	6.7	4.5	4.2	7.6
15	6.3	4.5	4.7	6.9	4.5	6.8	5.9	8.8	6	4.8	5.2	8
16	8.7	3.2	4	6.8	3.2	7.8	3.8	4.9	6.1	4.3	4.5	6.6
17	5.7	4	6.7	6	3.3	5.5	5.1	6.2	6.7	4.2	4.5	6.4
18	5.9	4.1	5.5	7.2	3.5	6.4	5.5	8.4	6.2	5.7	4.8	7.4
19	5.6	3.4	5.1	6.4	3.7	5.7	5.6	9.1	5.4	5	4.5	6.8
20	9.1	4.5	3.6	6.4	5.3	5.3	7.1	8.4	5.8	4.5	4.4	7.6
21	5.2	3.8	7.1	5.2	3.9	4.3	5	8.4	7.1	3.3	3.3	5.4
22	9.6	5.7	6.8	5.9	5.4	8.3	7.8	4.5	6.4	4.3	4.3	9.9
23	8.6	3.6	7.4	5.1	3.5	7.3	4.7	3.7	6.7	4.8	4	7
24	9.3	2.4	2.6	7.2	2.2	7.2	4.5	6.2	6.4	6.7	4.5	8.6
25	6	4.1	5.3	4.7	3.5	5.3	5.3	8	6.5	4.7	4	4.8
26	6.4	3.6	6.6	6.1	4	3.9	5.3	7.1	6.1	5.6	3.9	6.6
27	8.5	3	7.2	5.8	4.1	7.6	3.7	4.8	6.9	5.3	4.4	6.3
28	7	3.3	5.4	5.5	2.6	4.8	4.2	9	6.5	4.3	3.7	5.4
29	8.5	3	5.7	6	2.3	7.6	3.7	4.8	5.8	5.7	4.4	6.3
30	7.6	3.6	3	4	5.1	4.2	4.6	7.7	4.9	4.7	3.5	5.4
31	6.9	3.4	8.5	4.3	4.5	6.4	4.7	5.2	7.7	3.7	3.3	6.1
32	8.1	2.5	7.2	4.5	2.3	5.1	3.8	6.6	6.8	3	3	6.4
33	6.7	3.7	6.5	5.3	5.3	5.1	4.9	9.2	5.7	3.5	3.4	5.4
34	8	3.3	6.1	5.7	5.5	4.6	4.7	8.7	5.9	4.7	4.2	7.3
35	6.7	4	5.2	3.9	3	5.4	6.8	8.4	6.2	2.5	3.5	6.3
36	8.7	3.2	6.1	4.3	3.5	6.1	2.9	5.6	6.1	3.1	2.5	5.4
37	9	3.4	5.9	4.6	3.9	6	4.5	6.8	6.4	3.9	3.5	7.1
38	9.6	4.1	6.2	7.3	2.9	7.7	5.5	7.7	6.1	5.2	4.9	8.7
39	8.2	3.6	3.9	6.2	5.8	4.9	5	9	5.2	4.7	4.5	7.6
40	6.1	4.9	3	4.8	5.1	3.9	6.4	8.2	5.1	4.5	3.2	6
41	8.3	3.4	3.3	5.5	3.1	4.6	5.2	9.1	4.1	4.6	3.9	7
42	9.4	3.8	4.7	5.4	3.8	6.5	4.9	8.5	4.9	4.1	4.1	7.6
43	9.3	5.1	4.6	6.8	5.8	6.6	6.3	7.4	5.1	4.6	4.3	8.9
44	5.1	5.1	6.6	6.9	4.4	5.4	7.8	5.9	7.2	4.9	4.5	7.6
45	8	2.5	4.7	7.1	3.6	7.7	3	5.2	5.1	4.3	4.7	5.5
46	5.9	4.1	5.7	5.9	5.8	6.4	5.5	8.4	6.4	5.2	4.8	7.4
47	10	4.3	7.1	6.3	2.9	5.4	4.5	3.8	6.7	5	3.5	7.1
48	5.7	3.8	6.8	7.5	5.7	5.7	6	8.2	6.6	6.5	5.2	7.6
49	9.9	3.7	3.7	6.1	4.2	7	6.7	6.8	5.9	4.5	3.9	8.7
50	7.9	3.9	4.3	5.8	4.4	6.9	5.8	4.7	5.2	4.1	4.3	8.6
51	6.7	3.6	5.9	4.2	3.4	4.7	4.8	7.2	5.7	4	2.8	5.4
52	8.2	2.7	3.7	7.4	2.7	7.9	3.1	5.3	5.3	4.5	4.9	5.7
53	9.4	2.5	4.8	6.1	3.2	7.3	4.6	6.3	6.3	4.7	4.6	8.7
54	6.9	3.4	5.7	4.4	3.3	6.4	4.7	5.2	6.4	3.2	3.3	6.1
55	8	3.3	3.8	5.8	3.2	4.6	4.7	8.7	5.3	4.9	4.2	7.3
56	9.3	3.8	7.3	5.7	3.7	6.4	5.5	7.4	6.6	4.1	3.4	7.7
57	7.4	5.1	4.8	7.7	4.5	7.2	6.9	9.6	6.4	5.7	5.5	9
58	7.6	3.6	5.2	5.8	5.6	6.6	5.4	4.4	6.7	4.6	4	8.2
59	10	4.3	5.3	3.7	4.2	5.4	4.5	3.8	6.7	3.7	3.5	7.1
60	9.9	2.8	7.2	6.9	2.6	5.8	3.5	5.4	6.2	5.6	4	7.9
61	8.7	3.2	8.4	6.1	2.8	7.8	3.8	4.9	7.2	5.4	4.5	6.6
62	8.4	3.8	6.7	5	4.5	4.7	5.9	6.7	5.1	2.7	3.6	8
63	8.8	3.9	3.8	5.1	4.3	4.7	4.8	5.8	5	4.4	2.9	6.3
64	7.7	2.2	6.3	4.5	2.4	4.7	3.4	6.2	6	3.3	2.6	6
65	6.6	3.6	5.8	4.1	4.9	4.7	4.8	7.2	6.5	3.5	2.8	5.4
66	5.7	3.8	3.5	6.7	5.4	5.7	6	8.2	5.4	4.7	5.2	7.6
67	5.7	4	7.9	6.4	2.7	5.5	5.1	6.2	7.5	5	4.5	6.4
68	5.5	3.7	4.7	5.4	4.3	5.3	4.9	6	5.6	4.5	4.3	6.1
69	7.5	3.5	3.8	3.5	2.9	4.1	4.5	7.6	5.1	4	3.4	5.2
70	6.4	3.6	2.7	5.3	3.9	3.9	5.3	7.1	5.2	4.7	3.9	6.6
71	9.1	4.5	6.1	5.9	6.3	5.3	7.1	8.4	7.1	5.4	4.4	7.6
72	6.7	3.2	3	3.7	4.8	6.3	4.5	5	5.2	2.9	3.1	5.8
73	6.5	4.3	2.7	6.6	6.5	6.3	6	8.7	4.7	4.6	4.6	7.9
74	9.9	3.7	7.5	4.7	5.6	7	6.7	6.8	7.2	4.1	3.9	8.6
75	8.5	3.9	5.3	5.5	5	4.9	6	6.8	5.7	4.4	3.7	8.2
76	9.9	3	6.8	5	5.4	5.9	4.8	4.9	7.3	3.1	3.8	7.1
77	7.6	3.6	7.6	4.6	4.7	4.6	5	7.4	8.1	4.5	3.9	6.4
78	9.4	3.8	7	6.2	4.7	6.5	4.9	8.5	7.3	4.3	4.1	7.6
79	9.3	3.5	6.3	7.6	5.5	7.5	5.9	4.6	6.6	5.2	4.6	8.9
80	7.1	3.4	4.9	4.1	4	5	5.9	7.8	6.1	2.6	2.7	5.7
81	9.9	3	7.4	4.8	4	5.9	4.8	4.9	5.9	3.2	3.8	7.1
82	8.7	3.2	6.4	4.9	2.4	6.8	4.6	6.8	6.3	4.3	4	7.4
83	8.6	2.9	5.8	3.9	2.9	5.6	4	6.3	6.1	2.7	3	6.6
84	6.4	3.2	6.7	3.6	2.2	2.9	5	8.4	7.3	2	1.6	5
85	7.7	2.6	6.7	6.6	1.9	7.2	4.3	5.9	6.5	4.7	4.3	8.2
86	7.5	3.5	4.1	4.5	3.5	4.1	4.5	7.6	4.9	3.4	3.4	5.2
87	5	3.6	1.3	3	3.5	4.2	4.9	8.2	4.3	2.4	3.1	5.2
88	7.7	2.6	8	6.7	3.5	7.2	4.3	5.9	6.9	5.1	4.3	8.2
89	9.1	3.6	5.5	5.4	4.2	6.2	4.6	8.3	6.5	4.6	3.9	7.3
90	5.5	5.5	7.7	7	5.6	5.7	8.2	6.3	7.4	5.5	4.9	8.2
91	9.1	3.7	7	4.1	4.4	6.3	5.4	7.3	7.5	4.4	3.3	7.4
92	7.1	4.2	4.1	2.6	2.1	3.3	4.5	9.9	5.5	2	2.4	4.8
93	9.2	3.9	4.6	5.3	4.2	8.4	4.8	7.1	6.2	4.4	4.2	7.6
94	9.3	3.5	5.4	7.8	4.6	7.5	5.9	4.6	6.4	4.8	4.6	8.9
95	9.3	3.8	4	4.6	4.7	6.4	5.5	7.4	5.3	3.6	3.4	7.7
96	8.6	4.8	5.6	5.3	2.3	6	5.7	6.7	5.8	4.9	3.6	7.3
97	7.4	3.4	2.6	5	4.1	4.4	4.8	7.2	4.5	4.2	3.7	6.3
98	8.7	3.2	3.3	3.2	3.1	6.1	2.9	5.6	5	3.1	2.5	5.4
99	7.8	4.9	5.8	5.3	5.2	5.3	7.1	7.9	6	4.3	3.9	6.4
100	7.9	3	4.4	5.1	5.9	4.2	4.8	9.7	5.7	3.4	3.5	6.4

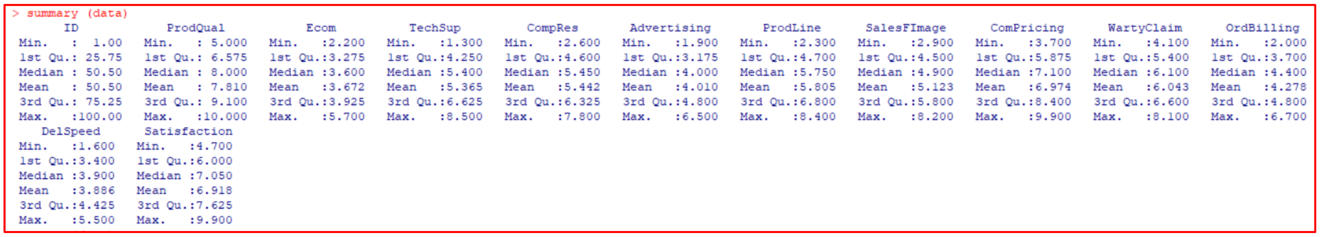
It is expected to calculate the following



Mini Project

- Is there evidence of Multicollinearity?
- Perform Factor Analysis by extracting four factors
- Name the factors
- Perform Multiple Linear Regression with Customer Satisfaction as the dependent variable and the four factors as the independent variables. Comment on Model Validity

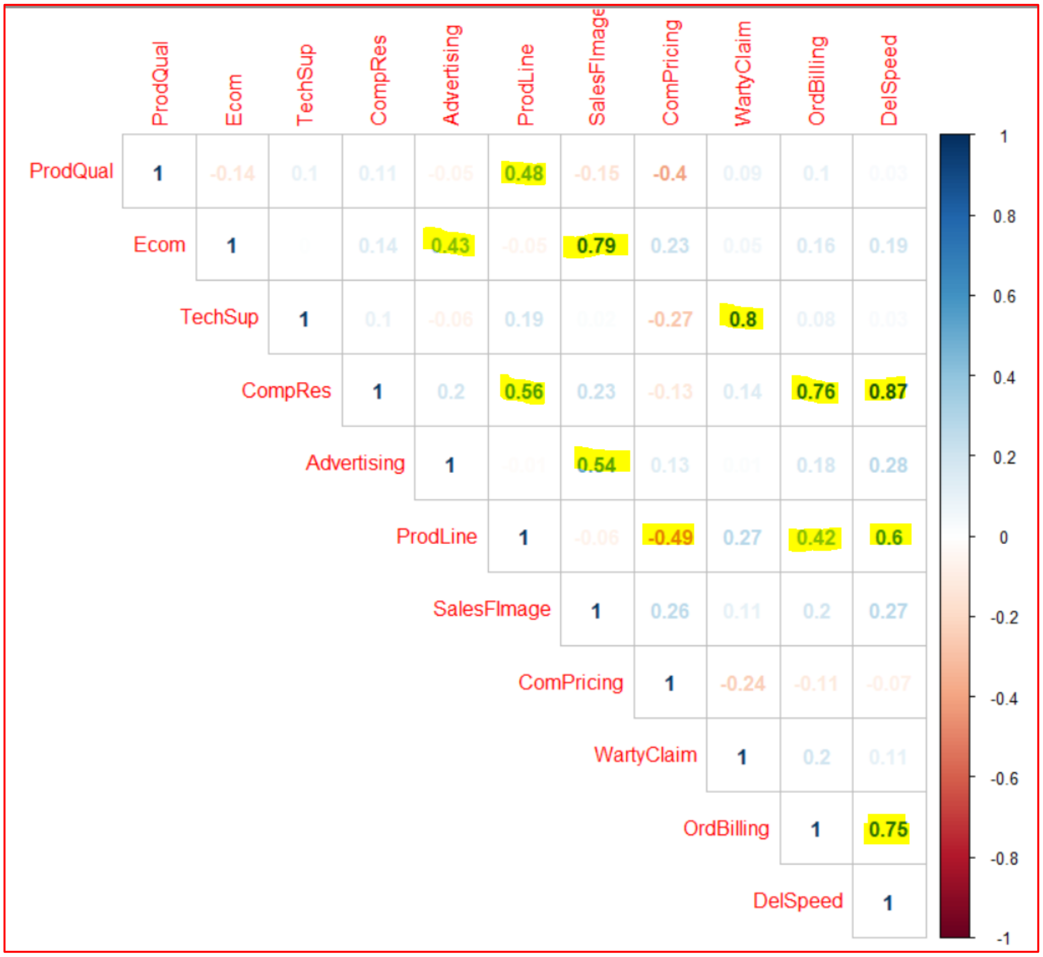
Univariate analysis:



```
> summary (data)
       ID            ProdQual            Ecom            TechSup          CompRes        Advertising       ProdLine       SalesFImage      ComPricing      WartyClaim      OrdBilling
 Min.   :  1.00   Min.   : 5.000   Min.   :2.200   Min.   :1.300   Min.   :2.600   Min.   :1.900   Min.   :2.300   Min.   :2.900   Min.   :3.700   Min.   :4.100   Min.   :2.000
 1st Qu.: 25.75   1st Qu.: 6.575   1st Qu.:3.275   1st Qu.:4.250   1st Qu.:4.600   1st Qu.:3.175   1st Qu.:4.700   1st Qu.:4.500   1st Qu.:5.875   1st Qu.:5.400   1st Qu.:3.700
 Median : 50.50   Median : 8.000   Median :3.600   Median :5.400   Median :5.450   Median :4.000   Median :5.750   Median :4.900   Median :7.100   Median :6.100   Median :4.400
 Mean   : 50.50   Mean   : 7.810   Mean   :3.672   Mean   :5.365   Mean   :5.442   Mean   :4.010   Mean   :5.805   Mean   :5.123   Mean   :6.974   Mean   :6.043   Mean   :4.278
 3rd Qu.: 75.25   3rd Qu.: 9.100   3rd Qu.:3.925   3rd Qu.:6.625   3rd Qu.:6.325   3rd Qu.:4.800   3rd Qu.:6.800   3rd Qu.:5.800   3rd Qu.:8.400   3rd Qu.:6.600   3rd Qu.:4.800
 Max.   :100.00   Max.   :10.000   Max.   :5.700   Max.   :8.500   Max.   :7.800   Max.   :6.500   Max.   :8.400   Max.   :8.200   Max.   :9.900   Max.   :8.100   Max.   :6.700
    DelSpeed       Satisfaction
 Min.   :1.600   Min.   :4.700
 1st Qu.:3.400   1st Qu.:6.000
 Median :3.900   Median :7.050
 Mean   :3.886   Mean   :6.918
 3rd Qu.:4.425   3rd Qu.:7.625
 Max.   :5.500   Max.   :9.900
```

5. *Solution:*

5.1: Is there evidence of Multicollinearity?



	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed
ProdQual	1	-0.14	0.1	0.11	-0.05	0.48	-0.15	-0.4	0.09	0.1	0.03
Ecom		1	0	0.14	0.43	-0.05	0.79	0.23	0.05	0.16	0.19
TechSup			1	0.1	-0.06	0.19	0.02	-0.27	0.8	0.08	0.03
CompRes				1	0.2	0.56	0.23	-0.13	0.14	0.76	0.87
Advertising					1	-0.01	0.54	0.13	0.01	0.18	0.28
ProdLine						1	-0.06	-0.49	0.27	0.42	0.6
SalesFImage							1	0.26	0.11	0.2	0.27
ComPricing								1	-0.24	-0.11	-0.07
WartyClaim									1	0.2	0.11
OrdBilling										1	0.75
DelSpeed											1

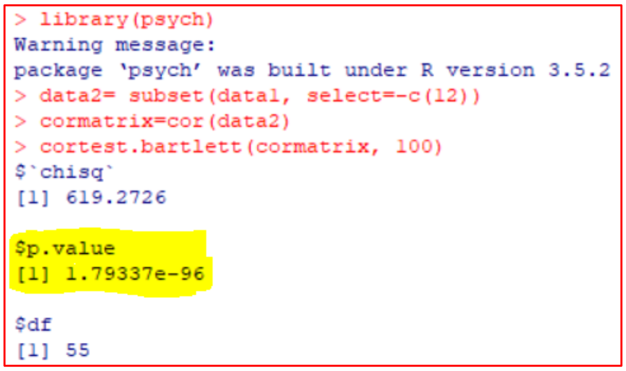
The highlighted values in yellow above shows that there exists multicollinearity amongst the variables like

i. Ecom to Sales Force Image
ii. Complaints Raised to warranty claims
iii. Complaints raised to Order and Billing
iv. Complaints Raised to Delivery Speed
v. Product Line to Delivery Speed
vi. Product Quality to Product Line

5.2: Perform Factor Analysis by extracting four factors

Before the Factor Analysis is done below tests done to be sure that the dataset qualifies for a PCA and a Factor Analysis.

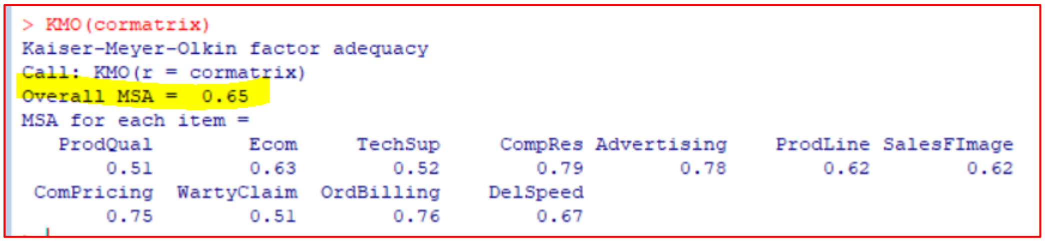
Test 1: Bartlett test will be done to check if the data qualifies for a Principal Component Analysis



```
> library(psych)
Warning message:
package 'psych' was built under R version 3.5.2
> data2= subset(data1, select=-c(12))
> cormatrix=cor(data2)
> cortest.bartlett(cormatrix, 100)
$`chisq`
[1] 619.2726

$p.value
[1] 1.79337e-96

$df
[1] 55
```

Since the p value is less than 0.05, it can be concluded that the dataset qualifies for a PCA analysis

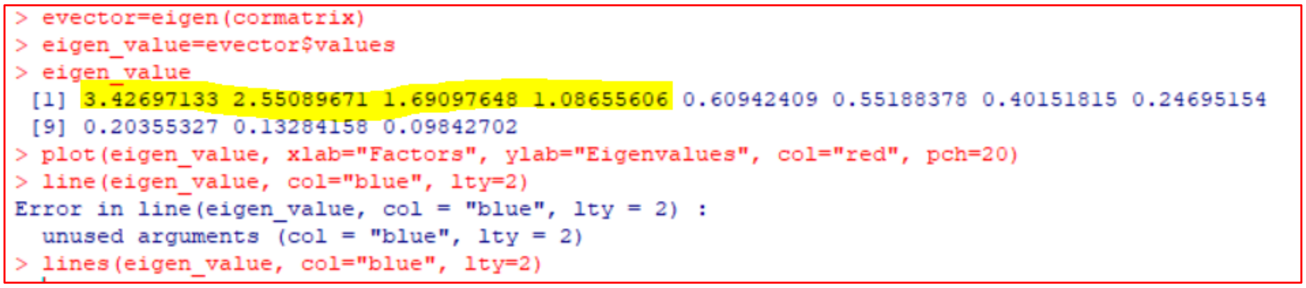
Test2: KMO test will be done to check if the data qualifies for a Factor Analysis



```
> KMO(cormatrix)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = cormatrix)
Overall MSA =  0.65
MSA for each item = 
   ProdQual        Ecom     TechSup     CompRes Advertising     ProdLine SalesFImage 
       0.51        0.63        0.52        0.79        0.78         0.62        0.62 
 ComPricing  WartyClaim  OrdBilling    DelSpeed 
       0.75        0.51        0.76        0.67 
```

Since the overall MSA >0.5 the dataset qualifies for Factor Analysis

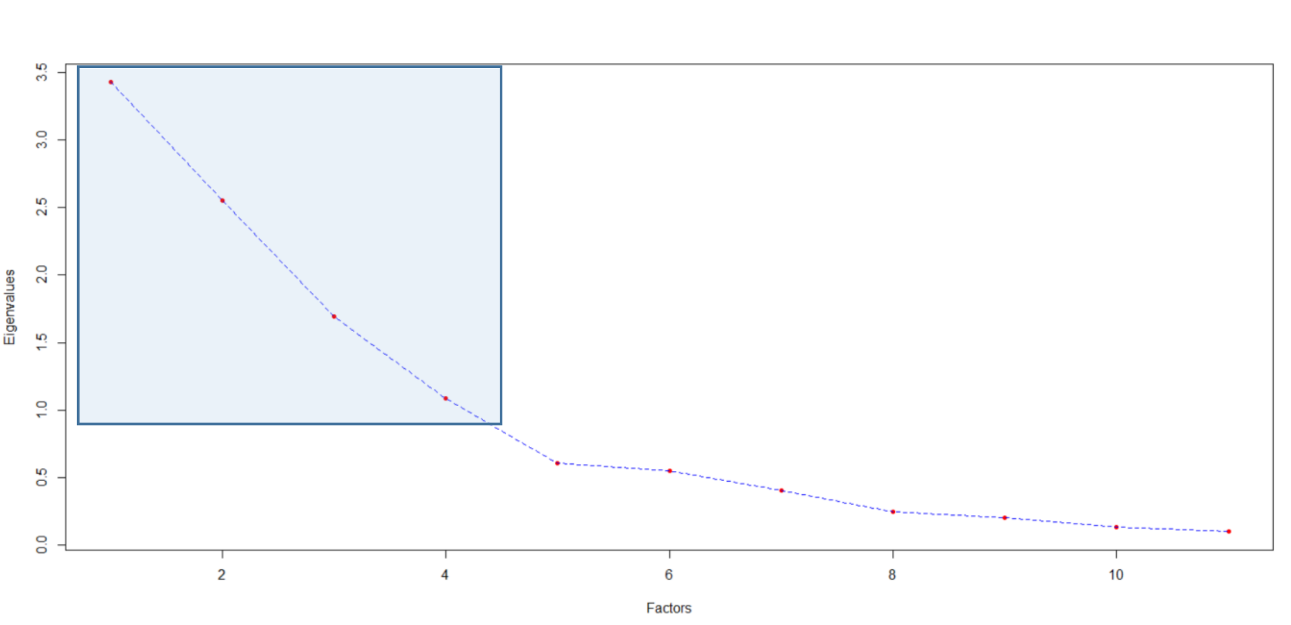
Calculation of the Eigenvalue



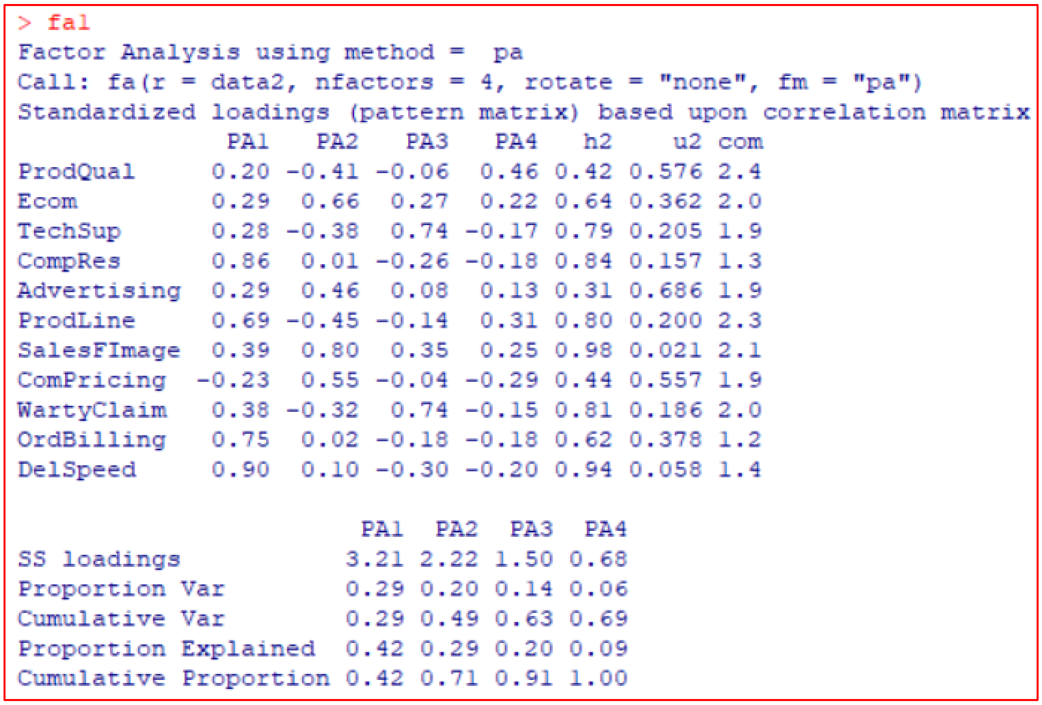
```
> evector=eigen(cormatrix)
> eigen_value=evector$values
> eigen_value
 [1] 3.42697133 2.55089671 1.69097648 1.08655606 0.60942409 0.55188378 0.40151815 0.24695154
 [9] 0.20355327 0.13284158 0.09842702
> plot(eigen_value, xlab="Factors", ylab="Eigenvalues", col="red", pch=20)
> line(eigen_value, col="blue", lty=2)
Error in line(eigen_value, col = "blue", lty = 2) : 
  unused arguments (col = "blue", lty = 2)
> lines(eigen_value, col="blue", lty=2)
```

All the Eigenvalues greater than 1 will be considered as the Principal components

Scree Plot depicting the Eigenvalues in the blue box below. This shows that there are 4 factors to be considered for PCA and FA analysis since all the these 4 values lie above 1

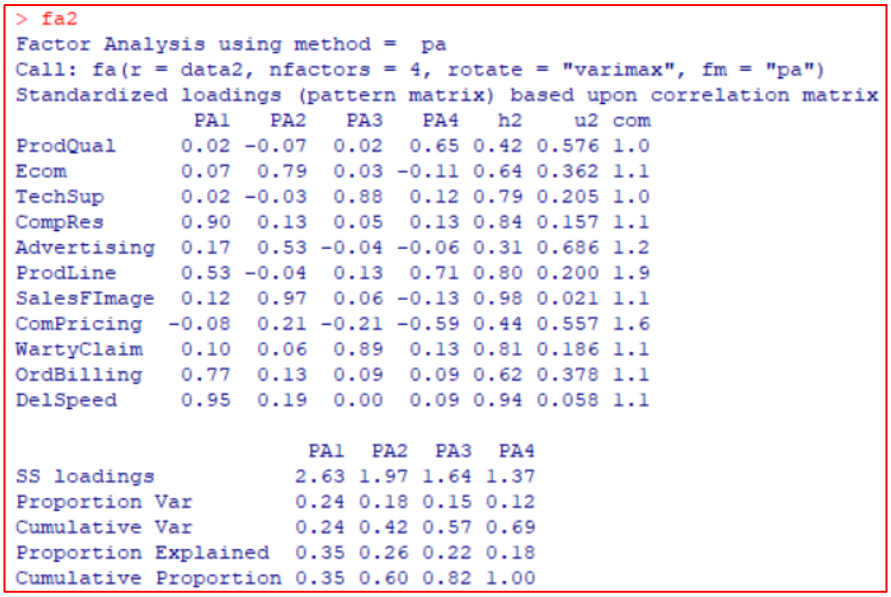


Unrotated Values:

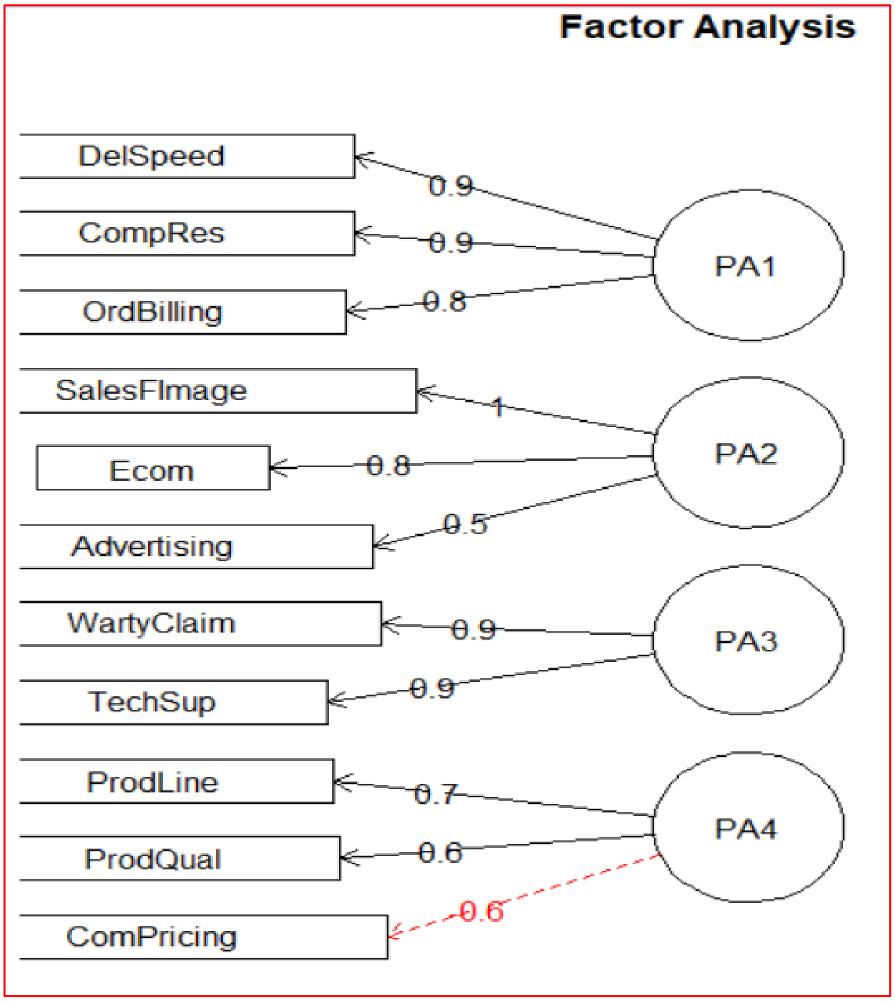


```
> fa1
Factor Analysis using method =  pa
Call: fa(r = data2, nfactors = 4, rotate = "none", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
              PA1   PA2   PA3   PA4   h2    u2 com
ProdQual     0.20 -0.41 -0.06  0.46 0.42 0.576 2.4
Ecom         0.29  0.66  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
```

Rotated values



```
> fa2
Factor Analysis using method =  pa
Call: fa(r = data2, 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
```



FA Diagram

The FA diagram shows the 4 major factors

5.3: Name the Factors

As can be seen

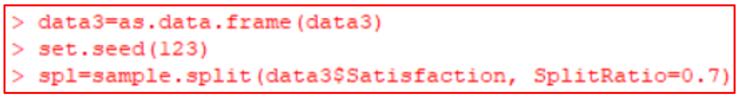
1. The first factor is pertaining to pre and during the sales of the product like delivery time, correct and accurate billing and any complaint resolution
2. Secondly activities and facilities pertaining to advertising impact since it can be observed variables like sales force images, E-commerce, advertising are grouped
3. Thirdly post sales services like warranty claims, technical support matters
4. Lastly the Product experience like the quality, variety of products impact and competitive pricing

To summarize, the factors impacting in chronology are

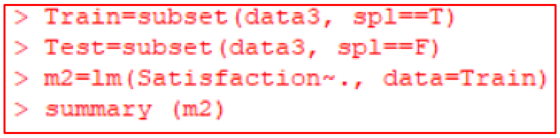
a. Pre-sales
b. Advertising and marketing
c. Post sale services
d. Product Experience

5.4: Perform multiple Liner Regression with Customer Satisfaction as the dependent variable and the 4 factors as the independent variables. Comment on Model validity

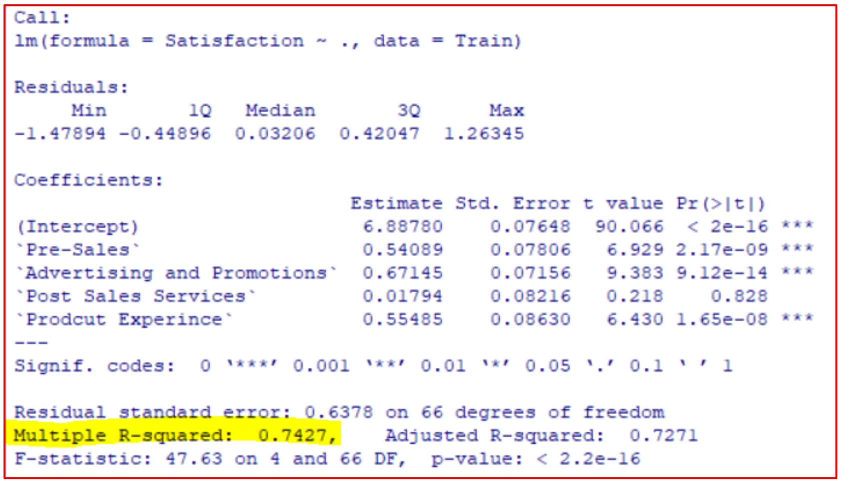
A multiple Linear Regression for Satisfaction and 4 factors is shown below with the coding done in R



```
> data3=as.data.frame(data3)
> set.seed(123)
> spl=sample.split(data3$Satisfaction, SplitRatio=0.7)
```



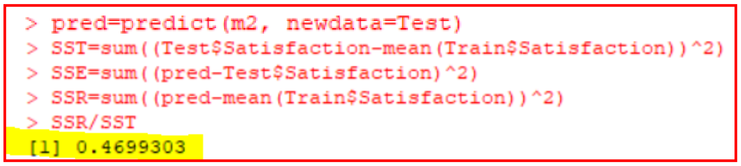
```
> Train=subset(data3, spl==T)
> Test=subset(data3, spl==F)
> m2=lm(Satisfaction~., data=Train)
> summary (m2)
```



```
Call:
lm(formula = Satisfaction ~ ., data = Train)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.47894 -0.44896  0.03206  0.42047  1.26345 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   6.88780    0.07648  90.066  < 2e-16 ***
`Pre-Sales`                   0.54089    0.07806   6.929 2.17e-09 ***
`Advertising and Promotions`  0.67145    0.07156   9.383 9.12e-14 ***
`Post Sales Services`         0.01794    0.08216   0.218    0.828    
`Prodcut Experince`           0.55485    0.08630   6.430 1.65e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6378 on 66 degrees of freedom
Multiple R-squared:  0.7427,	Adjusted R-squared:  0.7271 
F-statistic: 47.63 on 4 and 66 DF,  p-value: < 2.2e-16
```

An R-squared value of 74.27% depicts strong correlation between Satisfaction and the 4 Factors that are pre-sales, advertising, post sales and product experience.

To test the validity of the Model below is performed



```
> pred=predict(m2, newdata=Test)
> SST=sum((Test$Satisfaction-mean(Train$Satisfaction))^2)
> SSE=sum((pred-Test$Satisfaction)^2)
> SSR=sum((pred-mean(Train$Satisfaction))^2)
> SSR/SST
[1] 0.4699303
```

6. *Conclusion:*

To test the validity of the model, data was split into 2 parts.

In the initial part where the data is train and a correlation was identified to be at ~74%

Later this model was tested on an independent data and this gave an R-squared value of ~47% which is way less than 74% thus depicting that the model is not a Good Fit i.e. it is depicting a weak model.

Since there is a steep decrease in the R-squared value in the test data as compared to R-squared of the train data a further judgmental decision to consider the model can be taken by the business or new model testing with new variables and more data collected will need to be done

--------------------End of the report----------------