PREDICTING CUSTOMER SATISFACTION USING MULTIPLE REGRESSION MODEL

Objective

To use the given information to build an optimum regression model to predict satisfaction.

Assumptions

1. Customer Satisfaction is the dependent variable and all the other factors are independent of each other
2. There is no multicollinearity between variables
3. Linear relationship exist between Customer Satisfaction and other variables

DataSet

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

Here Customer Satisfaction is the dependent variable. The dataset has 100 records with 11 independent variables

Exploratory data analysis on the dataset

Structure of data:

'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 ...

Data Summary :

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

Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim
Min. :1.900	Min. :2.300	Min. :2.900	Min. :3.700	Min. :4.100
1st Qu.:3.175	1st Qu.:4.700	1st Qu.:4.500	1st Qu.:5.875	1st Qu.:5.400
Median :4.000	Median :5.750	Median :4.900	Median :7.100	Median :6.100
Mean :4.010	Mean :5.805	Mean :5.123	Mean :6.974	Mean :6.043
3rd Qu.:4.800	3rd Qu.:6.800	3rd Qu.:5.800	3rd Qu.:8.400	3rd Qu.:6.600
Max. :6.500	Max. :8.400	Max. :8.200	Max. :9.900	Max. :8.100

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

- Variable ID is a unique number given to each customer and has no relevance while doing multiple regression modelling. Hence we can ignore Variable ID during our analysis.
- We have checked for missing values (NAs) in the data-set, and could not find any missing values.

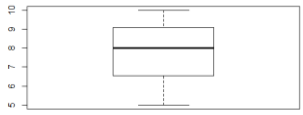
Variable Analysis :

ProdQual :

summary(mydata$ProdQual)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5.000   6.575   8.000   7.810   9.100  10.000
```

There is no outliers present in the data

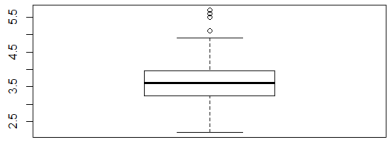


Ecom :

summary(mydata$Ecom)

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

There is possibility of presence of outliers



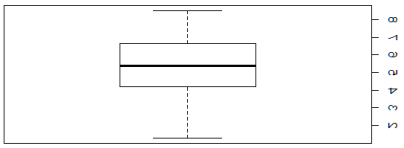
ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing
12	6.1	4.9	6.3	3.9	4.4	3.9	6.4	8.2
13	9.5	5.6	4.6	6.9	5.0	6.9	6.6	7.6
22	9.6	5.7	6.8	5.9	5.4	8.3	7.8	4.5
24	9.3	2.4	2.6	7.2	2.2	7.2	4.5	6.2
32	8.1	2.5	7.2	4.5	2.3	5.1	3.8	6.6
40	6.1	4.9	3.0	4.8	5.1	3.9	6.4	8.2
43	9.3	5.1	4.6	6.8	5.8	6.6	6.3	7.4
44	5.1	5.1	6.6	6.9	4.4	5.4	7.8	5.9
45	8.0	2.5	4.7	7.1	3.6	7.7	3.0	5.2
53	9.4	2.5	4.8	6.1	3.2	7.3	4.6	6.3
57	7.4	5.1	4.8	7.7	4.5	7.2	6.9	9.6
64	7.7	2.2	6.3	4.5	2.4	4.7	3.4	6.2
85	7.7	2.6	6.7	6.6	1.9	7.2	4.3	5.9
88	7.7	2.6	8.0	6.7	3.5	7.2	4.3	5.9
90	5.5	5.5	7.7	7.0	5.6	5.7	8.2	6.3
96	8.6	4.8	5.6	5.3	2.3	6.0	5.7	6.7
99	7.8	4.9	5.8	5.3	5.2	5.3	7.1	7.9

TechSup :

summary(mydata$TechSup)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.300   4.250   5.400   5.365   6.625   8.500 
```

There is no outliers present in the data

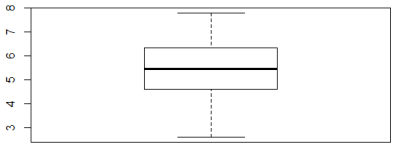


CompRes :

summary(mydata$CompRes)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  2.600   4.600   5.450   5.442   6.325   7.800 
```

There is no outliers present in the data

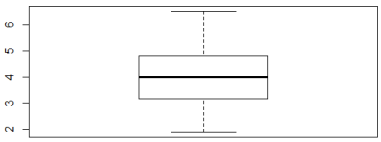


Advertising :

summary(mydata$Advertising)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.900   3.175   4.000   4.010   4.800   6.500 
```

There is no outliers present in the data

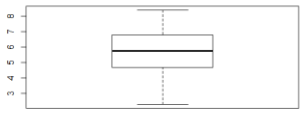


ProdLine :

summary(mydata$ProdLine)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.300   4.700   5.750   5.805   6.800   8.400
```

There is no outliers present in the data

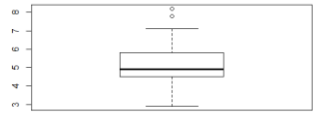


SalesFImage :

summary(mydata$SalesFImage)

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

There is possibility of presence of outliers



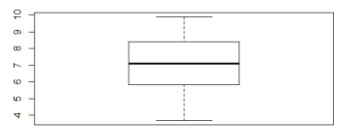
	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing
2	8.2	2.7	5.1	7.2	3.4	7.9	3.1	5.3
22	9.6	5.7	6.8	5.9	5.4	8.3	7.8	4.5
36	8.7	3.2	6.1	4.3	3.5	6.1	2.9	5.6
44	5.1	5.1	6.6	6.9	4.4	5.4	7.8	5.9
45	8.0	2.5	4.7	7.1	3.6	7.7	3.0	5.2
52	8.2	2.7	3.7	7.4	2.7	7.9	3.1	5.3
90	5.5	5.5	7.7	7.0	5.6	5.7	8.2	6.3
98	8.7	3.2	3.3	3.2	3.1	6.1	2.9	5.6

ComPricing :

summary(mydata$ComPricing)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.700   5.875   7.100   6.974   8.400   9.900
```

There is no outliers present in the data

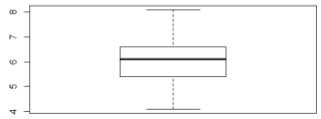


WartyClaim :

summary(mydata$WartyClaim)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  4.100   5.400   6.100   6.043   6.600   8.100
```

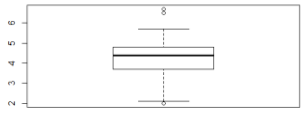
There is no outliers present in the data



OrdBilling :

summary(mydata$OrdBilling)

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



There is possibility of presence of outliers

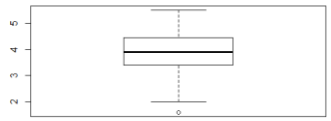
ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing
7	6.9	3.7	5.0	2.6	2.1	2.3	5.4	8.9
24	9.3	2.4	2.6	7.2	2.2	7.2	4.5	6.2
35	6.7	4.0	5.2	3.9	3.0	5.4	6.8	8.4
48	5.7	3.8	6.8	7.5	5.7	5.7	6.0	8.2
80	7.1	3.4	4.9	4.1	4.0	5.0	5.9	7.8
84	6.4	3.2	6.7	3.6	2.2	2.9	5.0	8.4
87	5.0	3.6	1.3	3.0	3.5	4.2	4.9	8.2
92	7.1	4.2	4.1	2.6	2.1	3.3	4.5	9.9

DelSpeed :

summary(mydata$DelSpeed)

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

There is possibility of presence of outliers



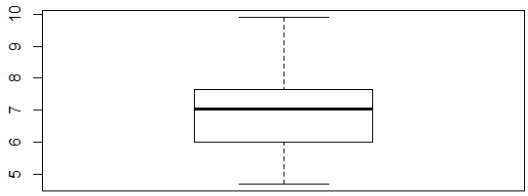

```
   ProdQual Ecom TechSup CompRes Advertising ProdLine SalesFImage ComPricing
7       6.9  3.7    5.0     2.6         2.1      2.3         5.4       8.9
57      7.4  5.1    4.8     7.7         4.5      7.2         6.9       9.6
84      6.4  3.2    6.7     3.6         2.2      2.9         5.0       8.4
```

Satisfaction :

summary(mydata$Satisfaction)

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

There is no outliers present in the data



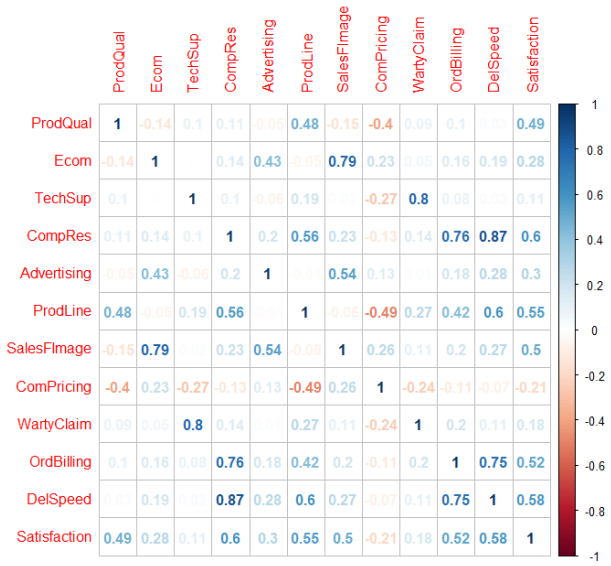
We shall check and plot the Correlation among the variables.

	ProdQual	Ecom	TechSup	CompRes	Advertising
ProdQual	1.00000000	-0.1371632174	0.0956004542	0.1063700	-0.05347313
Ecom	-0.13716322	1.0000000000	0.0008667887	0.1401793	0.42989071
TechSup	0.09560045	0.0008667887	1.0000000000	0.0966566	-0.06287007
CompRes	0.10637000	0.1401792611	0.0966565978	1.0000000	0.19691685
Advertising	-0.05347313	0.4298907110	-0.0628700668	0.1969168	1.00000000
ProdLine	0.47749341	-0.0526878383	0.1926254565	0.5614170	-0.01155082
SalesFImage	-0.15181287	0.7915437115	0.0169905395	0.2297518	0.54220366
ComPricing	-0.40128188	0.2294624014	-0.2707866821	-0.1279543	0.13421689
WartyClaim	0.08831231	0.0518981915	0.7971679258	0.1404083	0.01079207
OrdBilling	0.10430307	0.1561473316	0.0801018246	0.7568686	0.18423559
DelSpeed	0.02771800	0.1916360683	0.0254406935	0.8650917	0.27586308
Satisfaction	0.48632500	0.2827450147	0.1125971788	0.6032626	0.30466947

	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling
ProdQual	0.47749341	-0.15181287	-0.40128188	0.08831231	0.10430307
Ecom	-0.05268784	0.79154371	0.22946240	0.05189819	0.15614733
TechSup	0.19262546	0.01699054	-0.27078668	0.79716793	0.08010182
CompRes	0.56141695	0.22975176	-0.12795425	0.14040830	0.75686859
Advertising	-0.01155082	0.54220366	0.13421689	0.01079207	0.18423559
ProdLine	1.00000000	-0.06131553	-0.49494840	0.27307753	0.42440825
SalesFImage	-0.06131553	1.00000000	0.26459655	0.10745534	0.19512741
ComPricing	-0.49494840	0.26459655	1.00000000	-0.24498605	-0.11456703
WartyClaim	0.27307753	0.10745534	-0.24498605	1.00000000	0.19706512
OrdBilling	0.42440825	0.19512741	-0.11456703	0.19706512	1.00000000
DelSpeed	0.60185021	0.27155126	-0.07287173	0.10939460	0.75100307
Satisfaction	0.55054594	0.50020531	-0.20829569	0.17754482	0.52173191

	DelSpeed	Satisfaction
ProdQual	0.02771800	0.4863250
Ecom	0.19163607	0.2827450
TechSup	0.02544069	0.1125972
CompRes	0.86509170	0.6032626
Advertising	0.27586308	0.3046695
ProdLine	0.60185021	0.5505459
SalesFImage	0.27155126	0.5002053
ComPricing	-0.07287173	-0.2082957
WartyClaim	0.10939460	0.1775448
OrdBilling	0.75100307	0.5217319
DelSpeed	1.00000000	0.5770423
Satisfaction	0.57704227	1.0000000

The correlation matrix is given below:



As we can see from the above correlation matrix:

1. CompRes and DelSpeed are highly correlated
2. OrdBilling and CompRes are highly correlated
3. WartyClaim and TechSup are highly correlated
4. OrdBilling and DelSpeed are highly correlated
5. Ecom and SalesFImage are highly correlated

Thus, we can assume and tell that there exists a high degree of multicollinearity between the independent variables.

Simple Linear Regression with Satisfaction as the dependent variable and all the other variables in the data frame as the independent variables:

model1=lm (Satisfaction~ProdQual)

> summary (model1)

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

model2=lm(Satisfaction~Ecom)

> summary (model2)

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

model3=lm (Satisfaction~TechSup)

> summary (model3)

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

model4=lm(Satisfaction~CompRes)

```
> summary(model4)

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
```

model5=lm(Satisfaction~Advertising)

```
> summary(model5)

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
```

model6=lm(Satisfaction~ProdLine)

```
> summary(model6)

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
```

model7=lm (Satisfaction~SalesFImage)

```
> summary (model7)

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
```

model8=lm(Satisfaction~ComPricing)

```
> summary(model8)

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
```

model9=lm (Satisfaction~WartyClaim)

```
> summary (model9)

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
```

model10=lm(Satisfaction~OrdBilling)

```
> summary(model10)

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
```

model11=lm(Satisfaction~DelSpeed)

```
> summary(model11)

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
```

Regression Model using all Independent Variables

model12 =lm(Satisfaction~., mydata1)

> summary (model12)

```
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
```

Adjusted R-squared: 0.7774, means independent variables explain 78% of the variance of the dependent variable.
p-value of the F-statistic is less than 0.05 (alpha), which means this model is significant. This means, at least, one of the independent variables is related to the response variable.
This model equation is written as:
Satisfaction = -0.66 + 0.37*ProdQual -0.44*Ecom + 0.034*TechSup + 0.16*CompRes -0.02*Advertising + 0.14ProdLine + 0.80*SalesFImage-0.038*CompPricing -0.10*WartyClaim + 0.14*OrdBilling + 0.16*DelSpeed

Principal Component Analysis/ Factor Analysis

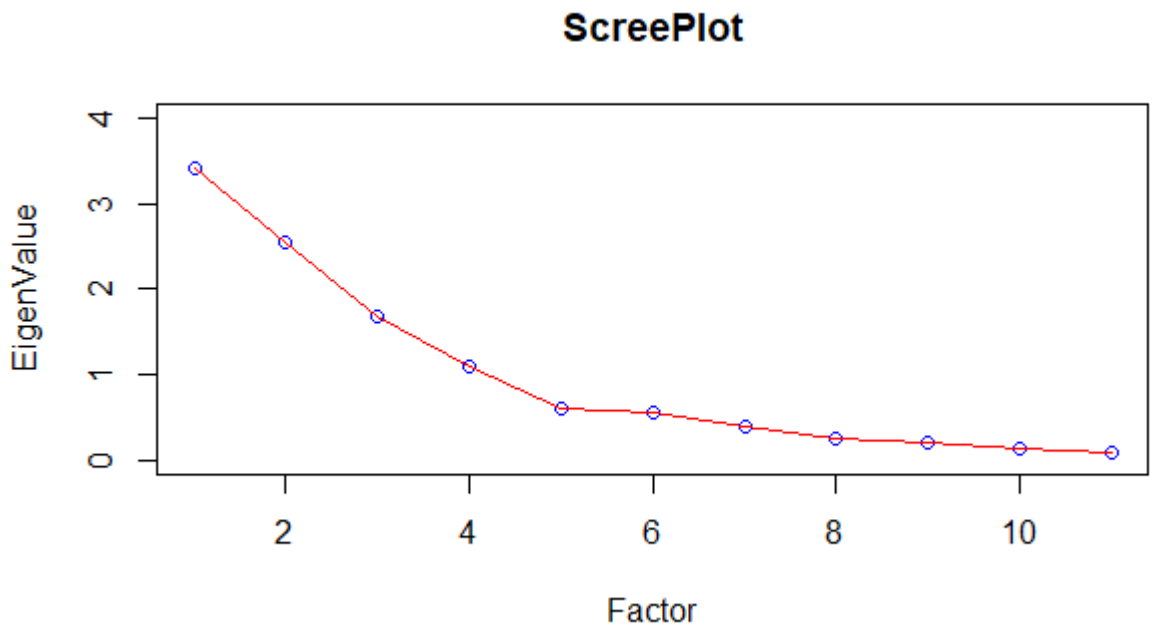
Performed Kaiser Meyer Olkin factor analysis

Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = datamatrix)
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

As MSA > 0.5, we can run Factor Analysis on this data.
Scree plot using Plot function : To find the number of factors in a correlation matrix is to plot the "scree" plot of eigenvalues. Sharp bend in the plot suggest the appropriate number of components.



In the graph, after factor 4 there is a sharp change in the slope in the scree plot. This shows that after factor 4 the total variance is significantly low. So, we can go ahead with 4 factors.

Let's now use 4 factors to perform the factor analysis.

```
Factor Analysis using method =  pa
Call: fa(r = mydata2, nfactors = 4, rotate = "none", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
              PA1    PA2    PA3    PA4    h2     u2  com
ProdQual     0.201 -0.408 -0.058  0.463 0.424 0.5757 2.40
Ecom         0.290  0.659  0.270  0.216 0.638 0.3618 2.00
TechSup      0.278 -0.381  0.738 -0.166 0.795 0.2054 1.95
CompRes      0.862  0.012 -0.255 -0.184 0.843 0.1572 1.27
Advertising  0.286  0.457  0.082  0.129 0.314 0.6858 1.95
ProdLine     0.689 -0.453 -0.142  0.315 0.800 0.1997 2.30
SalesFImage  0.395  0.801  0.346  0.251 0.979 0.0208 2.11
ComPricing  -0.232  0.553 -0.044 -0.286 0.443 0.5567 1.91
WartyClaim   0.379 -0.324  0.735 -0.153 0.814 0.1865 2.04
OrdBilling   0.747  0.021 -0.175 -0.181 0.622 0.3782 1.23
DelSpeed     0.895  0.098 -0.303 -0.198 0.942 0.0580 1.36

                       PA1   PA2   PA3   PA4
SS loadings          3.215 2.223 1.499 0.678
Proportion Var       0.292 0.202 0.136 0.062
Cumulative Var       0.292 0.494 0.631 0.692
Proportion Explained 0.422 0.292 0.197 0.089
Cumulative Proportion 0.422 0.714 0.911 1.000

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  6.553 with C
hi Square of  619.273
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  0.017
The df corrected root mean square of the residuals is  0.031

The harmonic number of observations is  100 with the empirical chi square  3.189  with prob <
1
The total number of observations was  100  with Likelihood Chi Square =  30.273  with prob <  0.
0244

Tucker Lewis Index of factoring reliability =  0.9215
RMSEA index =  0.0878  and the 90 % confidence intervals are  0.0317 0.1393
BIC =  -48.015
Fit based upon off diagonal values = 0.997
```

Measures of factor score adequacy

	PA1	PA2	PA3	PA4
Correlation of (regression) scores with factors	0.981	0.974	0.953	0.883
Multiple R square of scores with factors	0.962	0.948	0.908	0.779
Minimum correlation of possible factor scores	0.923	0.897	0.816	0.558

Let's use orthogonal rotation (varimax) because in orthogonal rotation the rotated factors will remain uncorrelated:

Factor analysis using fa method:
Factor Analysis using method = pa
Call: fa(r = mydata2, nfactors = 4, rotate = "varimax", fm = "pa")

Standardized loadings (pattern matrix) based upon correlation matrix

	PA1	PA2	PA3	PA4	h2	u2	com
ProdQual	0.024	-0.070	0.016	0.647	0.424	0.5757	1.03
Ecom	0.068	0.787	0.028	-0.113	0.638	0.3618	1.06
TechSup	0.020	-0.025	0.883	0.116	0.795	0.2054	1.04
CompRes	0.898	0.130	0.054	0.132	0.843	0.1572	1.09
Advertising	0.166	0.530	-0.043	-0.062	0.314	0.6858	1.24
ProdLine	0.525	-0.035	0.127	0.712	0.800	0.1997	1.92
SalesFImage	0.115	0.971	0.063	-0.135	0.979	0.0208	1.08
ComPricing	-0.076	0.213	-0.209	-0.590	0.443	0.5567	1.57
WartyClaim	0.103	0.057	0.885	0.128	0.814	0.1865	1.08
OrdBilling	0.768	0.127	0.088	0.089	0.622	0.3782	1.11
DelSpeed	0.949	0.185	-0.005	0.087	0.942	0.0580	1.09

	PA1	PA2	PA3	PA4
SS loadings	2.635	1.967	1.641	1.371
Proportion Var	0.240	0.179	0.149	0.125
Cumulative Var	0.240	0.418	0.568	0.692
Proportion Explained	0.346	0.258	0.215	0.180
Cumulative Proportion	0.346	0.604	0.820	1.000

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 6.553 with Chi Square of 619.273.
The degrees of freedom for the model are 17 and the objective function was 0.33

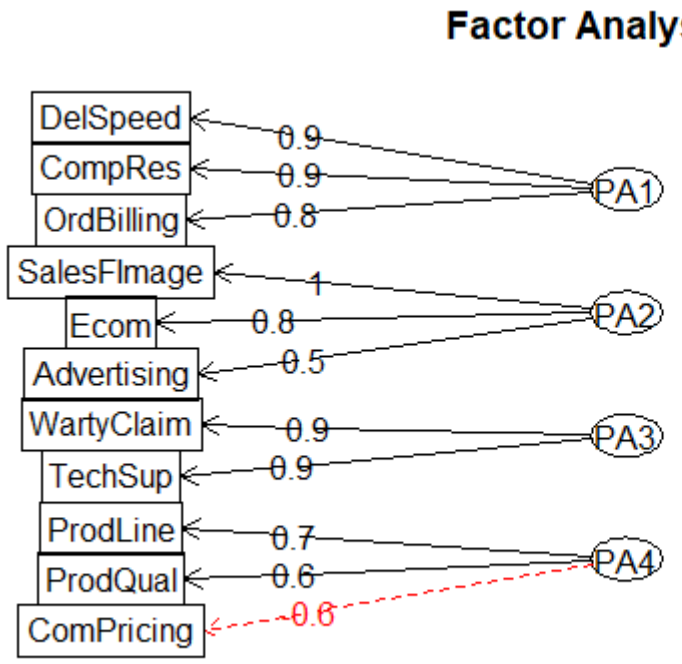
The root mean square of the residuals (RMSR) is 0.017

The df corrected root mean square of the residuals is 0.031

The harmonic number of observations is 100 with the empirical chi square 3.189 with prob < 1
The total number of observations was 100 with Likelihood Chi Square = 30.273 with prob < 0.02
44

Tucker Lewis Index of factoring reliability = 0.9215
RMSEA index = 0.0878 and the 90 % confidence intervals are 0.0317 0.1393
BIC = -48.015
Fit based upon off diagonal values = 0.997
Measures of factor score adequacy

	PA1	PA2	PA3	PA4
Correlation of (regression) scores with factors	0.982	0.986	0.940	0.882
Multiple R square of scores with factors	0.964	0.972	0.883	0.777
Minimum correlation of possible factor scores	0.928	0.945	0.766	0.554



We will combine the dependent variables and the factor scores into a dataset and label them.

	Customer_Satisfaction	Product_Purchase	Marketing	Post_deal	Positioning
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

Now, we run the Multiple Linear Regression model with Satisfaction as the independent variable and the 4 factors (from Factor Analysis) as the dependent variables.

```
lm(formula = Customer_Satisfaction ~ ., data = regdata)

Residuals:
   Min     1Q  Median     3Q    Max
-1.7125 -0.4708  0.1024  0.4158  1.3483

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.91800    0.06696 103.317  < 2e-16 ***
Product_Purchase  0.57963    0.06857   8.453 3.32e-13 ***
Marketing        0.61978    0.06834   9.070 1.61e-14 ***
Post_deal        0.05692    0.07173   0.794    0.429
Positioning      0.61168    0.07656   7.990 3.16e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6696 on 95 degrees of freedom
Multiple R-squared:  0.6971,             Adjusted R-squared:  0.6844
F-statistic: 54.66 on 4 and 95 DF,  p-value: < 2.2e-16
```

Here Adjusted R-squared is 0.6844. So, we can conclude that the multiple regression model built on the dataset with factor analysis explains almost 68% of variation in the dependent variable. The model is fit enough to be used and is valid.
Based on the model, variables like Complaint Resolution, Order & Billing and Delivery Speed explains the purchase aspects, E-commerce, Sales Force Image and Advertising explains marketing aspects. Product Quality, Product Line and Competitive Pricing explains the positioning of the product in the market. Warranty & claims and Technical Support explains the post deal aspects.
Most of variation in the dependent variable ‘Customer Satisfaction’ can be explained by the above-mentioned factors. Hence insight for the service provider is to focus on these aspects for bringing high levels of ‘Customer Satisfaction’.

Reference

Great learning video lectures and mentoring sessions