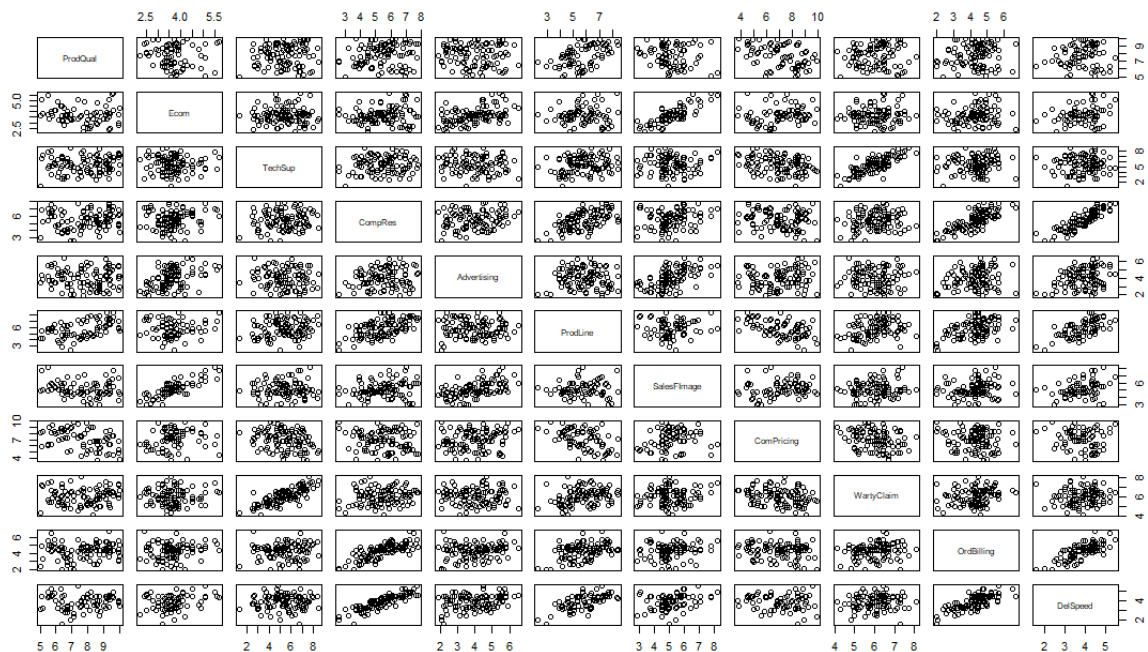


Customer Satisfaction Survey

Q. Is there evidence of multicollinearity?

The given dataset consists of 11 input variables and 1 dependent variable. To check for multicollinearity of the model we have to see the covariance matrix for the given 11 input variables.

ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFmImage	ComPricing	WartyClaim	OrdBilling	DelSpeed
1	-0.137163217	0.095600454	0.10637	-0.05347313	0.47749341	-0.15181287	-0.40128188	0.08831231	0.10430307	0.027718
-0.13716322	1	0.000866789	0.1401793	0.42989071	-0.05268784	0.79154371	0.2294624	0.05189819	0.15614733	0.19163607
0.09560045	0.000866789	1	0.0966566	-0.06287007	0.19262546	0.01699054	-0.27078668	0.79716793	0.08010182	0.02544069
0.10637	0.140179261	0.096656598	1	0.19691685	0.56141695	0.22975176	-0.12795425	0.1404083	0.75686859	0.8650917
-0.05347313	0.429890711	-0.062870067	0.1969168	1	-0.01155082	0.54220366	0.13421689	0.01079207	0.18423559	0.27586308
0.47749341	-0.052687838	0.192625457	0.561417	-0.01155082	1	-0.06131553	-0.4949484	0.27307753	0.42440825	0.60185021
-0.15181287	0.791543712	0.01699054	0.2297518	0.54220366	-0.06131553	1	0.26459655	0.10745534	0.19512741	0.27155126
-0.40128188	0.229462401	-0.270786682	-0.1279543	0.13421689	-0.4949484	0.26459655	1	-0.24498605	-0.11456703	-0.07287173
0.08831231	0.051898192	0.797167926	0.1404083	0.01079207	0.27307753	0.10745534	-0.24498605	1	0.19706512	0.1093946
0.10430307	0.156147332	0.080101825	0.7568686	0.18423559	0.42440825	0.19512741	-0.11456703	0.19706512	1	0.75100307
0.027718	0.191636068	0.025440694	0.8650917	0.27586308	0.60185021	0.27155126	-0.07287173	0.1093946	0.75100307	1



As we can see from the covariance matrix there is a strong evidence of multicollinearity among variables. The graphs also show a strong linear relationship among the variables.

Q. Perform factor analysis by extracting 4 factors

To perform the factor analysis, we will have to extract the eigenvectors and eigenvalues for the said covariation matrix. The eigenvectors and eigenvalues are as follows.

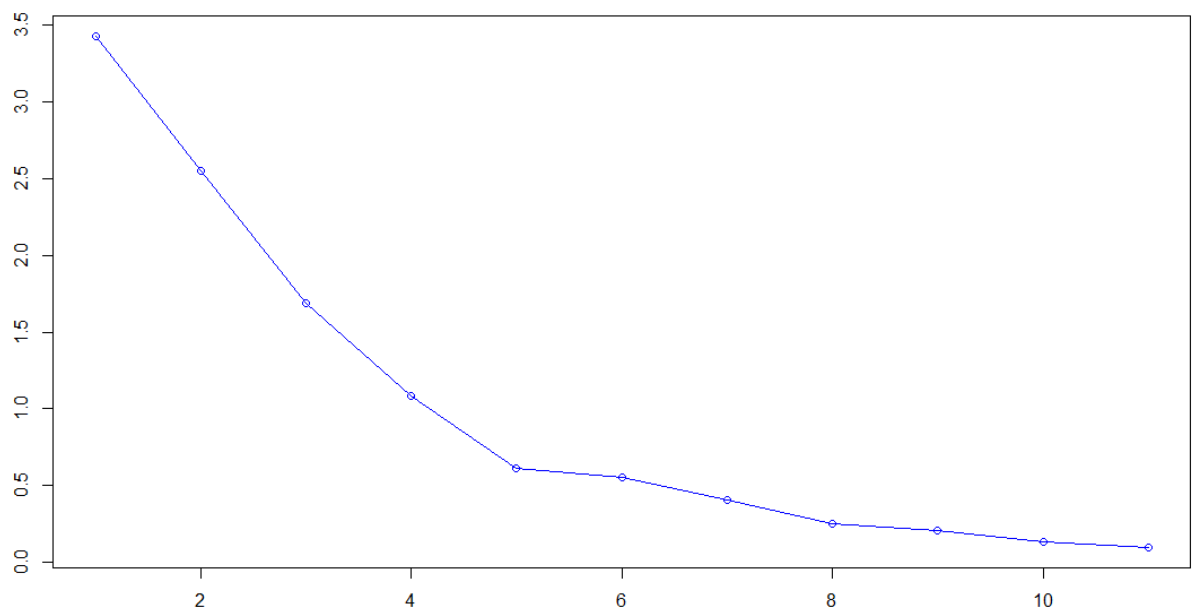
Eigenvalues

3.426971	2.550897	1.690976	1.086556	0.609424	0.551884	0.401518	0.246952	0.203553	0.132842	0.098427
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

Eigenvectors

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]
[1,]	-0.13379	0.313498	0.062272	0.643136	0.231666	0.56457	-0.19164	0.135473	0.031328	-0.0666	-0.18279
[2,]	-0.16595	-0.44651	-0.23525	0.27238	0.422288	-0.26326	-0.05963	-0.12203	-0.54251	-0.28156	-0.06234
[3,]	-0.15769	0.230967	-0.61095	-0.19339	-0.02396	0.108769	0.0172	0.46471	-0.3593	0.388171	0.05193
[4,]	-0.47068	-0.01944	0.210351	-0.20632	0.028657	0.028152	0.0085	0.513398	0.093248	-0.53467	0.362534
[5,]	-0.18374	-0.36366	-0.0881	0.317895	-0.80387	0.200569	0.06307	-0.05348	-0.15468	-0.03716	0.081187
[6,]	-0.38677	0.284781	0.116279	0.202902	0.116674	-0.0982	0.608148	-0.33321	-0.08416	0.234798	0.385078
[7,]	-0.20367	-0.4707	-0.24134	0.222177	0.204373	-0.10497	-0.00144	0.169107	0.644899	0.353412	0.084699
[8,]	0.151689	-0.41346	0.053045	-0.33354	0.248926	0.709736	0.308249	-0.09883	-0.09414	0.045182	0.102958
[9,]	-0.21293	0.191672	-0.59856	-0.1853	-0.03293	0.13984	0.03064	-0.44354	0.317566	-0.43535	-0.12893
[10,]	-0.43722	-0.0264	0.16893	-0.23685	0.026754	0.11948	-0.65932	-0.36602	-0.09907	0.303865	0.194151
[11,]	-0.47309	-0.07305	0.232625	-0.19733	-0.03543	-0.0298	0.234239	0.065391	-0.02189	0.120104	-0.77563

The scree plot for the eigenvalues is as follows.



Beyond the first 4 values the eigenvalues are less than 1 and hence they can be ignored.

To find the principal components we shall find the percentage variance explained by the components.

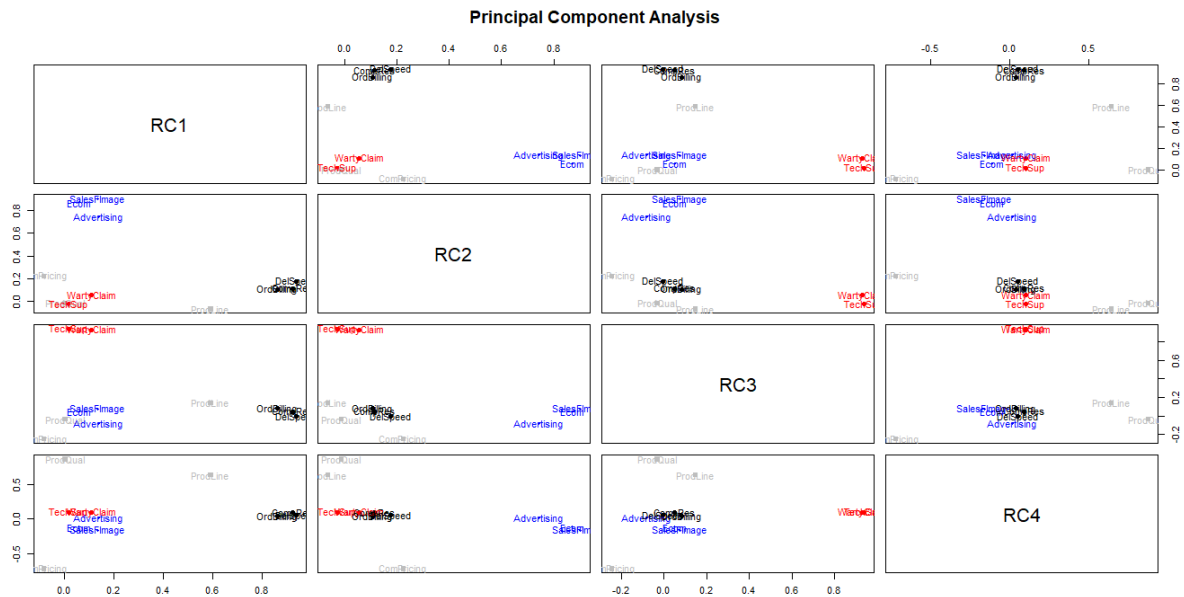
	PC1	PC2	PC3	PC4	h2	u2	com
ProdQual	0.25	-0.5	-0.08	0.67	0.77	0.232	2.2
Ecom	0.31	0.71	0.31	0.28	0.78	0.223	2.1
TechSup	0.29	-0.37	0.79	-0.2	0.89	0.107	1.9
CompRes	0.87	0.03	-0.27	-0.22	0.88	0.119	1.3
Advertising	0.34	0.58	0.11	0.33	0.58	0.424	2.4
ProdLine	0.72	-0.45	-0.15	0.21	0.79	0.213	2
SalesFImage	0.38	0.75	0.31	0.23	0.86	0.141	2.1
ComPricing	-0.28	0.66	-0.07	-0.35	0.64	0.359	1.9
WartyClaim	0.39	-0.31	0.78	-0.19	0.89	0.108	2
OrdBilling	0.81	0.04	-0.22	-0.25	0.77	0.234	1.3
DelSpeed	0.88	0.12	-0.3	-0.21	0.91	0.086	1.4

Since the variance proportion is not very well understood in the normal principal components, we shall opt for the varimax rotation.

	RC1	RC2	RC3	RC4	h2	u2	com
ProdQual	0	-0.01	-0.03	0.88	0.77	0.232	1
Ecom	0.06	0.87	0.05	-0.12	0.78	0.223	1.1
TechSup	0.02	-0.02	0.94	0.1	0.89	0.107	1
CompRes	0.93	0.12	0.05	0.09	0.88	0.119	1.1
Advertising	0.14	0.74	-0.08	0.01	0.58	0.424	1.1
ProdLine	0.59	-0.06	0.15	0.64	0.79	0.213	2.1
SalesFImage	0.13	0.9	0.08	-0.16	0.86	0.141	1.1
ComPricing	-0.09	0.23	-0.25	-0.72	0.64	0.359	1.5
WartyClaim	0.11	0.05	0.93	0.1	0.89	0.108	1.1
OrdBilling	0.86	0.11	0.08	0.04	0.77	0.234	1.1
DelSpeed	0.94	0.18	0	0.05	0.91	0.086	1.1

As we can see that the demarcations among the variables are clearly defined.

	RC1	RC2	RC3	RC4
SS loadings	2.89	2.23	1.86	1.77
Proportion Var	0.26	0.2	0.17	0.16
Cumulative Var	0.26	0.47	0.63	0.8
Proportion Explained	0.33	0.26	0.21	0.2
Cumulative Proportion	0.33	0.59	0.8	1



Together these 4 factors account for 80% variation in the data.

Q. Name the factors

As shown above the independent variables can be collapsed into 4 factors, namely:

1. Customer Convenience: Complaint Resolution, Order and Billing and Delivery Speed
2. Customer Reach: E-commerce, Advertising and Sales Image
3. Customer Support: Technical Support and Warranty and Claims
4. Product Saleability: Product Quality, Product Line and Competitive Pricing

Q. Perform Multiple Linear Regression with Customer Satisfaction as the dependent variable and the four factors as the independent variables. Comment on the model validity.

After collapsing the variables into the main factors, the data scores for the principal components are as follows:

	Cust. Convenience	Cust. Reach	Cust. Support	Prod. Saleability	Satisfaction
1	0.12749104	0.7698686	-1.878446273	0.36648477	8.2
2	1.22166663	-1.64586166	-0.61403001	0.81306481	5.7
3	0.6158214	0.58000368	0.003689252	1.56997685	8.9
4	-0.84462665	-0.27192183	1.267493254	-1.25416452	4.8
5	-0.3197943	-0.83406501	-0.008096627	0.44753766	7.1
6	-0.64702925	-1.06726829	-1.303198892	-1.05277921	4.7
7	-2.62679851	-0.24588272	-0.555423494	-1.2260147	5.7
8	-0.27936394	-0.15732039	-0.749311481	-1.01464175	6.3
9	1.05151341	-0.17228834	-0.092252815	-1.65809634	7
10	0.42875382	0.76353272	-0.450377116	-0.89116595	5.5
11	-0.13580761	-0.76759698	-0.463706767	0.6063414	7.4
12	-1.45030579	1.35959912	0.437785016	-1.06981053	6
13	0.62461823	2.11311565	-0.168284409	0.87466736	8.4
14	0.42724294	-0.40405102	0.432245882	0.90236591	7.6
15	1.43869881	0.66394839	-0.268050576	-1.04431806	8

16	0.91969055	-1.05791159	-0.556847385	1.16667179	6.6
17	0.52182175	-0.31959634	1.106009732	-1.03228845	6.4
18	1.71349224	-0.16356534	0.254874808	-1.47834954	7.4
19	1.16101062	-0.41943765	-0.375574495	-1.76167798	6.8
20	0.29327394	1.77627892	-0.950139113	0.24112808	7.6
21	-0.61501848	-0.17897273	1.52594354	-1.83178487	5.4
22	-0.11282553	2.83382456	0.634265462	2.24434088	9.9
23	0.08062	-0.35141218	1.141318858	1.33498913	7
24	1.94944755	-1.67141336	-0.859208476	0.50283683	8.6
25	0.11534004	-0.01629685	0.47184192	-1.25041487	4.8
26	0.57499258	-0.24490397	0.62429286	-1.3543536	6.6
27	0.82896381	-0.98564797	1.042612499	0.921637	6.3
28	0.11695051	-1.10728007	0.379702318	-1.35959873	5.4
29	1.15812632	-1.60628019	-0.055788125	0.79531052	6.3
30	-0.50739097	0.16192496	-1.551322987	-0.30617006	5.4
31	-0.81074131	-0.17909238	2.256638942	0.21624964	6.1
32	-1.07438259	-1.60132074	1.186706049	-0.07026025	6.4
33	-0.49992323	0.30576561	0.157100923	-0.9702076	5.4
34	0.27885747	0.07142401	-0.032941868	-0.65628441	7.3
35	-1.21092268	0.61247373	0.27577366	-0.68907425	6.3
36	-1.37569442	-1.0590106	0.277541003	1.02901615	5.4
37	-0.62476762	-0.24359504	0.310901127	0.66051905	7.1
38	1.36407521	0.03533514	0.111220579	0.58229289	8.7
39	0.60127495	0.47053204	-1.291508459	-0.44567425	7.6
40	-0.58595295	1.48246242	-1.184474889	-1.03900017	6
41	0.19167763	-0.38987441	-1.981705114	-0.59621998	7
42	0.04337736	0.09038218	-1.165712378	0.53711635	7.6
43	0.40978439	1.9582198	-1.094672035	0.98888677	8.9
44	0.77547735	1.61343935	1.512055016	-1.1492399	7.6
45	1.26977129	-1.77421869	-0.982794252	0.73741113	5.5
46	1.06006213	0.67869812	0.324241314	-1.10289754	7.4
47	-0.12283972	-0.09120895	0.996132311	1.41658476	7.1
48	2.09832312	0.46224836	0.840138645	-1.68134357	7.6
49	0.1560411	0.8820225	-0.8352767	1.29848126	8.7
50	0.22982346	0.50302016	-0.877037378	1.03687279	8.6
51	-0.9418317	-0.37565064	0.19417445	-0.65267018	5.4
52	1.56112818	-1.90837771	-1.17649658	0.72135781	5.7
53	0.86011758	-1.08934973	-0.24143124	0.87182584	8.7
54	-0.81818435	-0.52905894	0.539901007	0.33090833	6.1
55	0.54057306	-0.67964718	-1.060702696	-0.81493134	7.3
56	-0.36862437	0.28299033	0.917529711	0.60437604	7.7
57	1.97865621	1.43218345	-0.085319811	-0.83928511	9
58	0.20552648	0.51721871	0.347543516	0.85780222	8.2
59	-1.34118399	0.55279292	0.326579529	1.94033636	7.1

60	0.85269365	-1.57772836	0.565957142	0.74035745	7.9
61	0.9933519	-1.26473291	1.700148685	0.79107349	6.6
62	-1.10480994	0.70911509	-0.152796271	0.39572776	8
63	-0.75921278	0.26001089	-1.188441475	0.78014681	6.3
64	-1.09474826	-1.95079477	0.427161087	-0.14850194	6
65	-1.20922892	0.15287985	0.577570622	-0.51556079	5.4
66	1.34313803	0.53659415	-1.039141561	-1.24941075	7.6
67	0.90215965	-0.58791187	2.06239035	-1.31875384	6.4
68	0.42318247	-0.24798003	-0.301264201	-0.84662237	6.1
69	-0.87487795	-0.60376193	-0.997620068	-0.52944051	5.2
70	0.14372369	-0.15149397	-1.275988102	-1.00015303	6.6
71	0.34387385	2.05641521	0.68634614	0.09426189	7.6
72	-1.16028876	-0.18463387	-1.205197353	0.71392258	5.8
73	0.9262035	1.31556747	-1.869872622	-0.55887325	7.9
74	-0.56659595	1.40049678	1.226627789	1.34965616	8.6
75	-0.29927186	0.87194345	-0.29462564	0.30300903	8.2
76	-0.89076271	0.23334622	1.037887857	1.61337977	7.1
77	-0.35535699	0.14354788	2.057316893	-0.63270298	6.4
78	0.21054781	0.3421826	1.073262401	0.30917078	7.6
79	1.12960563	0.64023318	0.441396478	1.46536309	8.9
80	-1.53178615	0.28775431	0.032504303	-0.31110748	5.7
81	-0.84995072	-0.24812793	0.45256285	1.53107516	7.1
82	0.02821132	-0.91638751	0.493585747	0.40440014	7.4
83	-1.39215814	-0.98489128	0.20760994	0.62550901	6.6
84	-2.48589153	-0.73564594	1.633547463	-1.4448807	5
85	1.0034756	-1.78211709	0.797684019	-0.01141758	8.2
86	-0.82905678	-0.41939997	-1.080457442	-0.45156381	5.2
87	-1.42542804	-0.29820535	-2.155317026	-1.27019948	5.2
88	1.0707665	-1.29822928	1.400760179	0.04006707	8.2
89	0.08823132	-0.05909838	0.1342287	0.2351372	7.3
90	1.07621515	2.37671168	1.892951438	-1.0134198	8.2
91	-0.78483349	0.46274897	1.391773475	0.61318828	7.4
92	-2.3479307	-0.26426141	-0.534487111	-1.18940207	4.8
93	0.29898878	0.20636519	-0.37141607	1.20810631	7.6
94	1.10722906	0.37021414	0.053771549	1.44542651	8.9
95	-0.79676401	0.71175008	-1.087719898	1.06131961	7.7
96	-0.11270919	0.39627233	0.048312077	0.3476712	7.3
97	-0.20833274	-0.2526409	-1.880921516	-0.3208168	6.3
98	-1.58596201	-1.12347151	-1.337515839	1.23670188	5.4
99	-0.32827278	1.90243479	0.140227444	-0.12061112	6.4
100	-0.6274407	0.21100398	-0.748923176	-0.69590553	6.4

On performing the multiple linear regression, we observe

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.918	0.07089	97.589	< 2E-16	***
Cust. Conv.	0.61805	0.07125	8.675	1.12E-13	***
Cust. Reach	0.50973	0.07125	7.155	1.74E-10	***
Cust. Supp.	0.06714	0.07125	0.942	0.348	
Prod. Sale.	0.54032	0.07125	7.584	2.24E-11	***

Residual standard error: 0.7089 on 95 degrees of freedom

Multiple R-squared: 0.6605, Adjusted R-squared: 0.6462

F-statistic: 46.21 on 4 and 95 DF, p-value: < 2.2e-16

With the given p-value there is overwhelming evidence that a regression exists in the population and it can be predicted based on the said variables.

Analysis of Variance Table					
Response:	Customer Satisfaction				
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Cust. Conv.	1	37.816	37.816	75.253	1.12E-13 ***
Cust. Reach	1	25.723	25.723	51.188	1.74E-10 ***
Cust. Supp.	1	0.446	0.446	0.888	0.3484
Prod. Sale.	1	28.903	28.903	57.515	2.25E-11 ***
Residuals	95	47.74	0.503		

If we estimate the linear regression to be given by $\hat{Y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + c$.

$\hat{Y}(\text{est.}) = 6.918 + 0.61805 * (\text{Customer Convenience}) + 0.50973 * (\text{Customer Reach}) + 0.54032 * (\text{Product Saleability})$

Our hypothesis for the validity of the model is as follows:

H_0 : all betas=0, H_a : at least one beta \neq 0. Since the p-value=2.2e-16, we can say with confidence that the alternative hypothesis is true and that at least one beta \neq 0.

The given model is able to explain 64.62% of the variations in customer satisfaction. The regression has 4 degrees of freedom.

If we consider the individual slope coefficients for the given variables and make the following hypothesis

For Customer Convenience

H_0 :Beta1=0

H_a :Beta1 \neq 0

Since p-value=1.12E-13 the null hypothesis can be rejected.

For Customer Reach

H_0 :Beta1=0

H_a :Beta1 \neq 0

Since $p\text{-value}=1.74\text{E-}10$ the null hypothesis can be rejected.

For Customer Support

$H_0:\text{Beta1}=0$

$H_a:\text{Beta1}\neq 0$

Since $p\text{-value}=0.3484$ the null hypothesis cannot be rejected.

For Product Saleability

$H_0:\text{Beta1}=0$

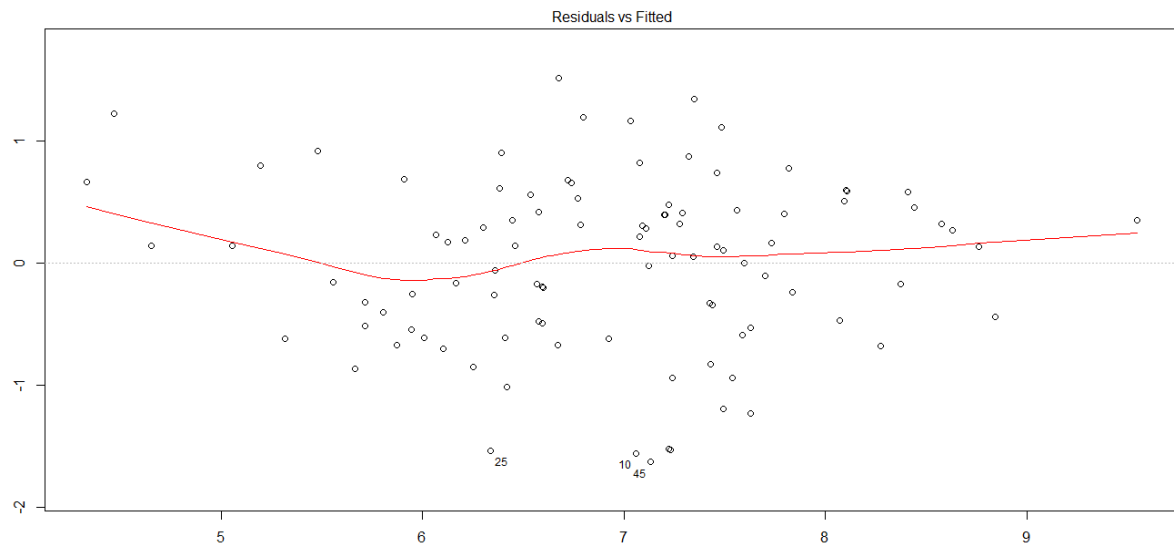
$H_a:\text{Beta1}\neq 0$

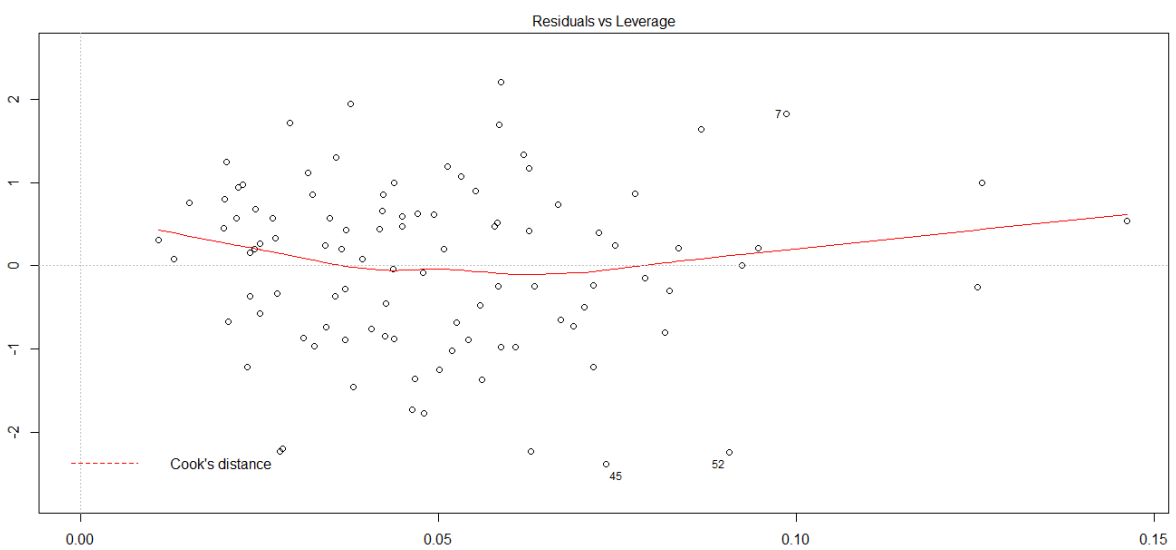
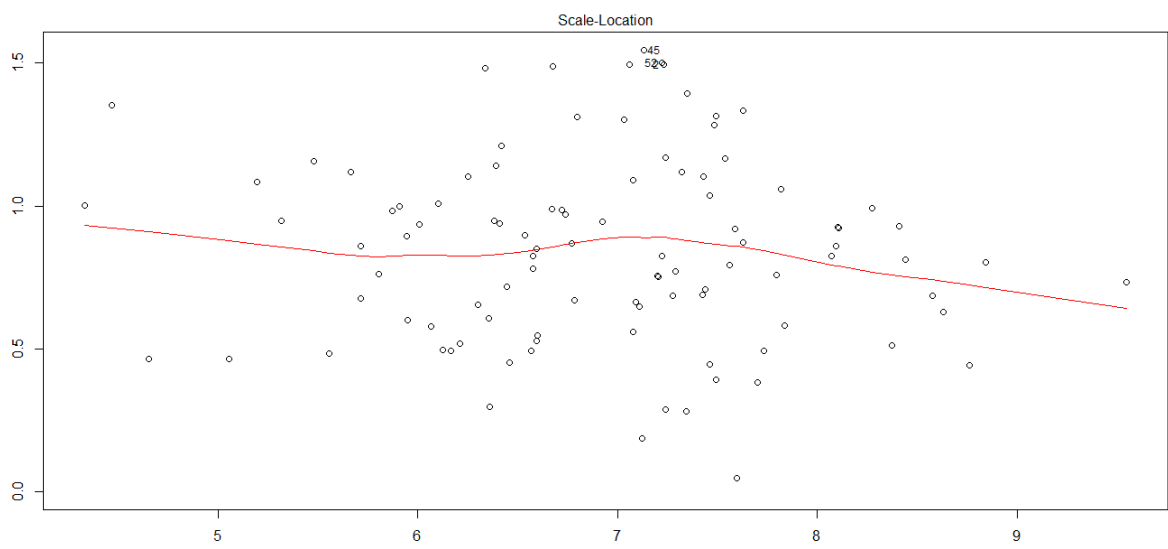
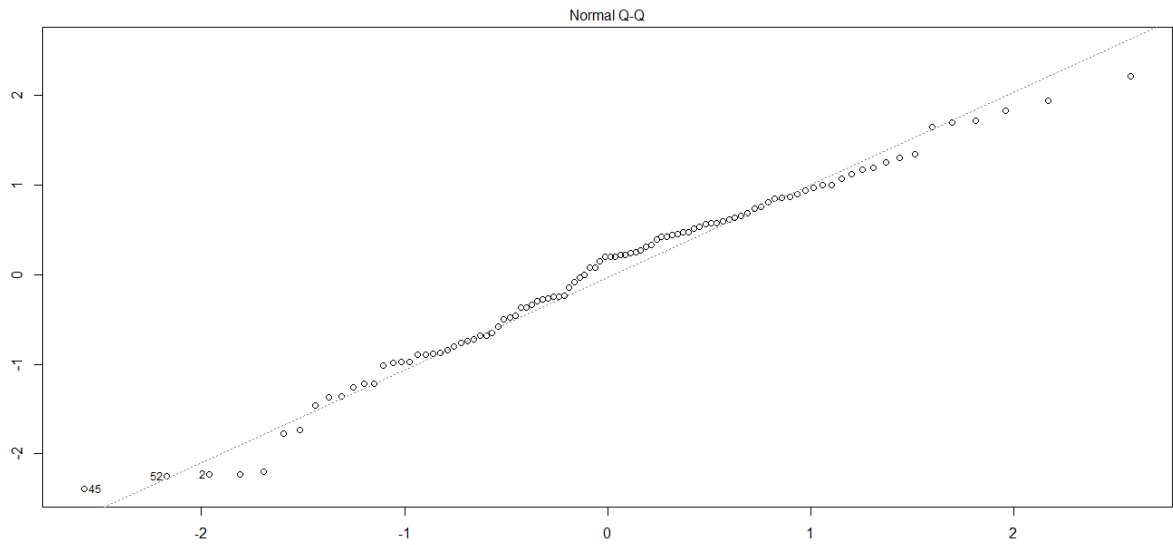
Since $p\text{-value}=2.25\text{E-}11$ the null hypothesis can be rejected.

We can conclude that a regression model exists in the population. The linear model of customer satisfaction that is linearly dependent on Customer Convenience, Customer Reach and Product Saleability is robust and statistically valid.

The individual coefficients of intercept, Customer Convenience, Customer Reach and Product Saleability are highly significant given their very low p-values for the t-statistics, which are much less than the significance level of 5%. Hence slope factors exist for Customer Convenience, Customer Reach, Product Saleability and intercept. Customer support is not significant in customer satisfaction.

The following graphs show the fit of the linear regression model with the actual values.

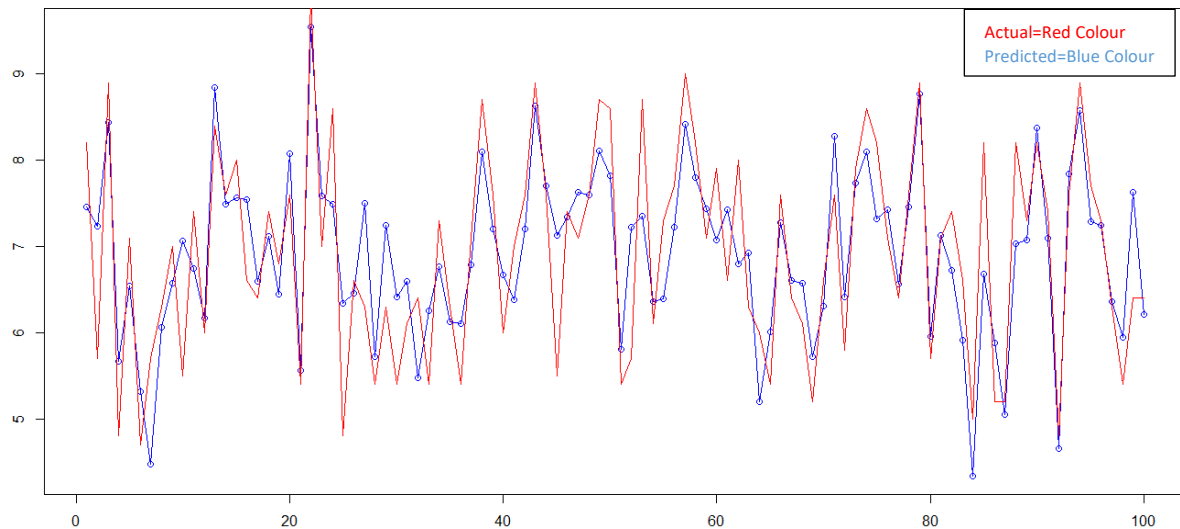




The estimates for the confidence interval for the coefficients are as follows.

Conf. Int.	2.50%	97.50%
(Intercept)	6.777268	7.058732
Cust. Conv.	0.476606	0.759488
Cust. Reach	0.368293	0.651175
Cust. Supp.	-0.07431	0.208577
Prod. Sale.	0.398878	0.68176

To observe the overall backtracking ability of the model we shall superimpose the actual and the predicted values for customer satisfaction.



As we can see that the predicted values are matching with the actual values at most of the points. We can conclude that model is able to predict the actual behaviour quite accurately.

Conclusion: We have observed that the given dataset contains information on 12 variables. We found an overwhelming evidence of correlation among the given variables. Thus, on performing the Principal Component Analysis we were able to reduce the independent variables to 4 variables. These variables were able to account for 80% variation in the data. We also observed that the principal components were able to account for the individual variables. Later by performing the multiple linear regression we were able to see a strong evidence of linear regression existing in the population. On performing the linear regression, we were able to see that the model successfully explained 64.62% variations in the data. The linear regression showed the relationship between the fitted and the residual values. The model also explained the conformance of the predicted values with the actual values.