Project Summary:

As in our model the adjusted R-squared is 0.811, meaning that independent variables explain 81.1% of the variance of the dependent variable, only 4 variables are significant out of 11 independent variables.

The p-value of the F-statistic is less than 0.05(level of Significance), which means our model is significant. This means that, at least, one of the predictor variables is significantly related to the outcome variable.

Our model equation can be written as:

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          6.89123    0.08065  85.441  < 2e-16 ***
Order.fulfillment.KPI  0.62966    0.07759   8.115 1.82e-11 ***
Brand.Image          0.48568    0.08798   5.521 6.34e-07 ***
After.Sales.Support -0.01397    0.07812  -0.179    0.859
Product.Differentiator  0.46061    0.08192   5.623 4.27e-07 ***
---
```

i.e

Customer Satisfaction = 6.89123+0.62966*(Order Fulfillment KPI)+0.48568*(Brand Image)+(-0.01397)*(After Sales Support)+0.46061*(Product differentiator)

In this Problem, we applied Factor Analysis to reduce the dimensionality of a dataset and then we used multiple linear regression on the dimensionally reduced columns for further analysis/predictions. Topics Covered in this Calculation Sheet and attachments are:

1. Checked for Multicollinearity
2. Factor Analysis
3. Clustering and naming the Factors
4. Perform Multiple Linear Regression with new variables.
5. Testing the model after training the data.

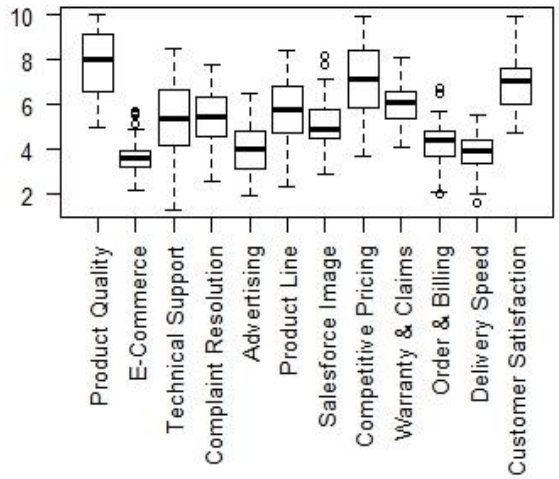
The Detailed Analysis step by step description is attached herewith in PDF and HTML format

Besides R notebook is also attached

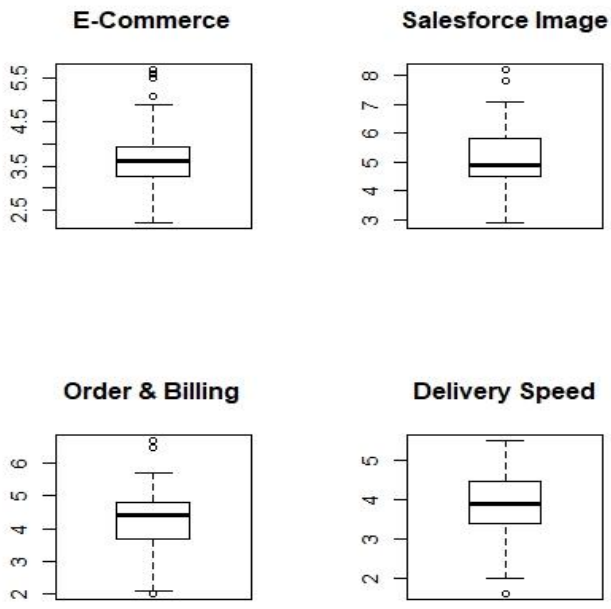
This File contains the Plots and graphs used in the statistical estimations

Graph 1:

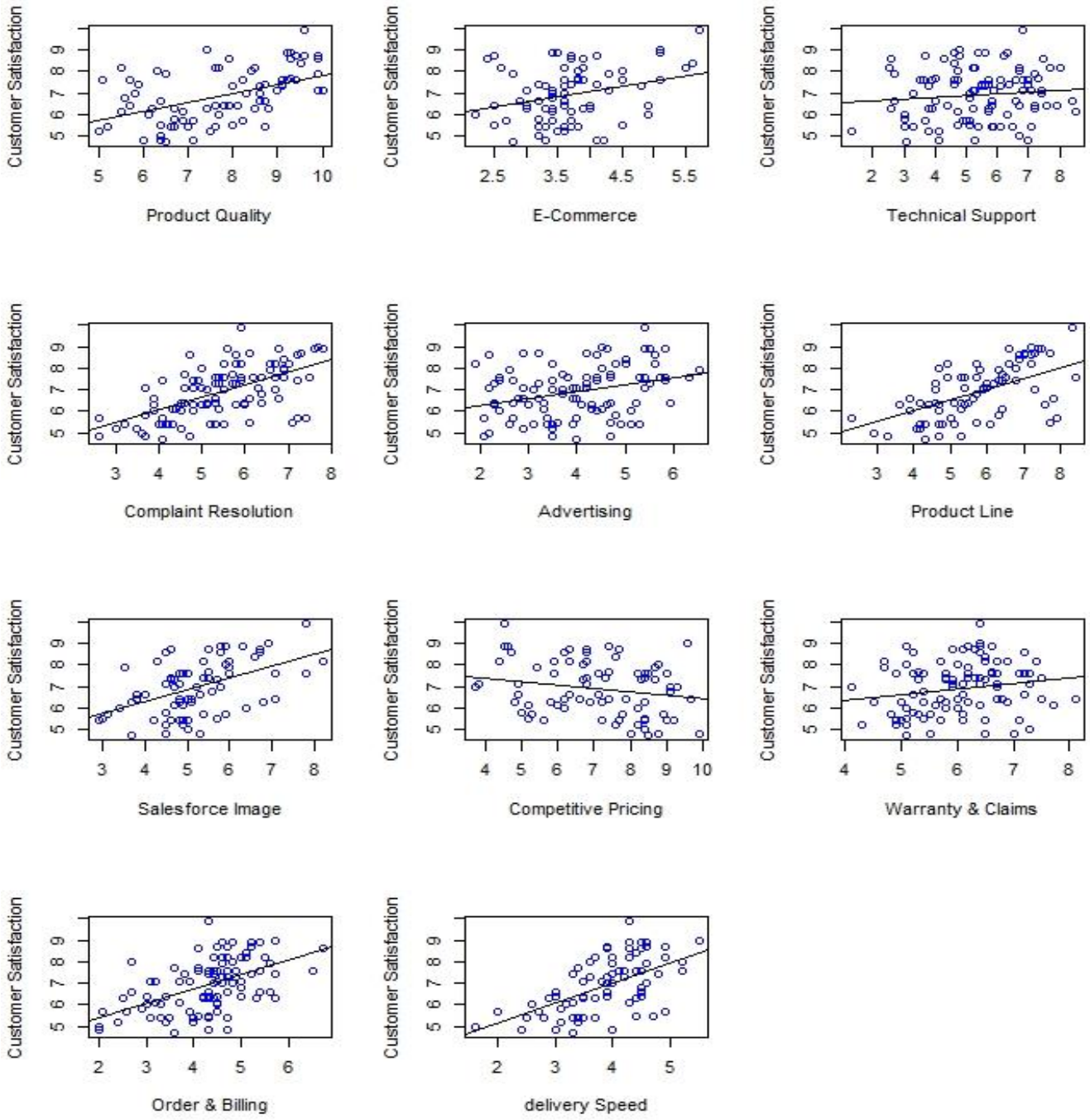
Box plot of all the independent and the dependent variable to identify the outliers



Graph2: Box plot of the Variables having some incidence of outliers

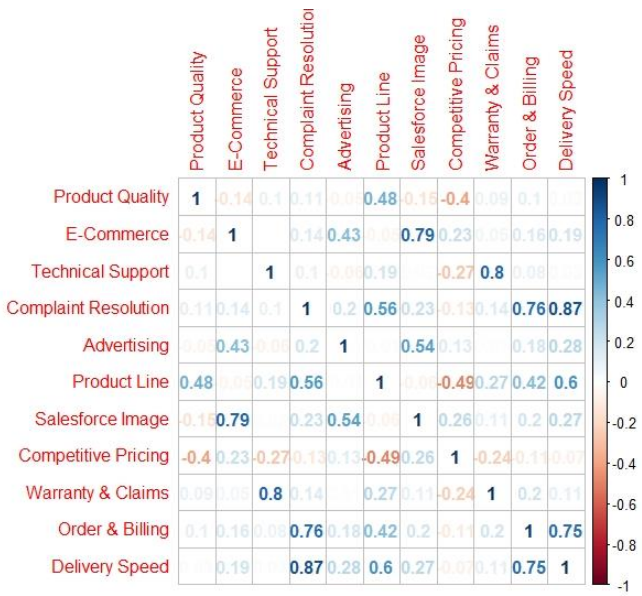


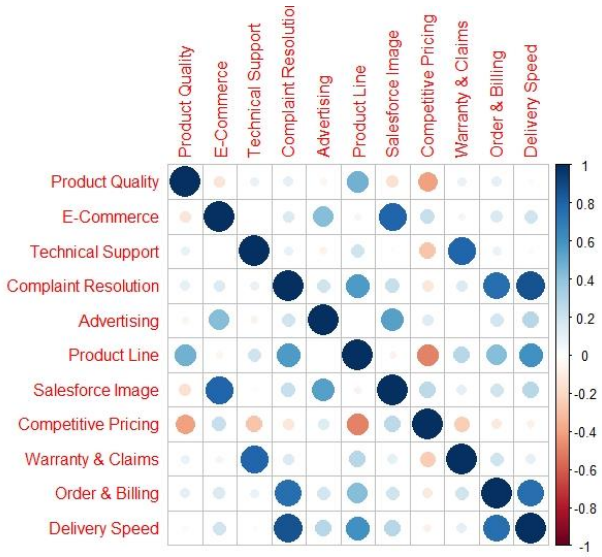
Graph 3 : Scatter Plot of all the independent Variables against the dependent / target variable i.e. “Customer Satisfaction and the best fit line.



Graph 4: Correlation plot of all the variables except the dependent variable

To identify the multicollinearity



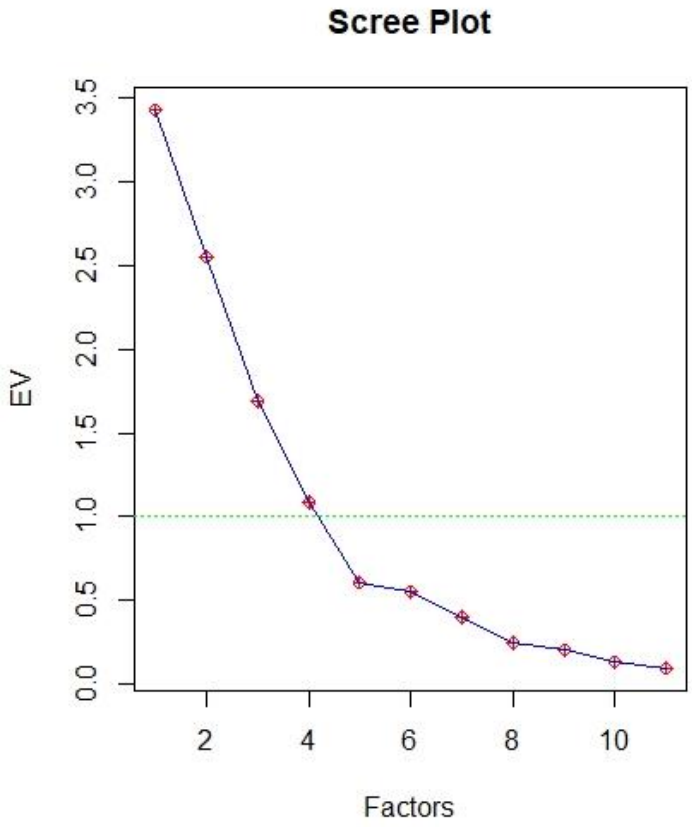


Observation:

Significant amount of multicollinearity is observed

Graph 5:

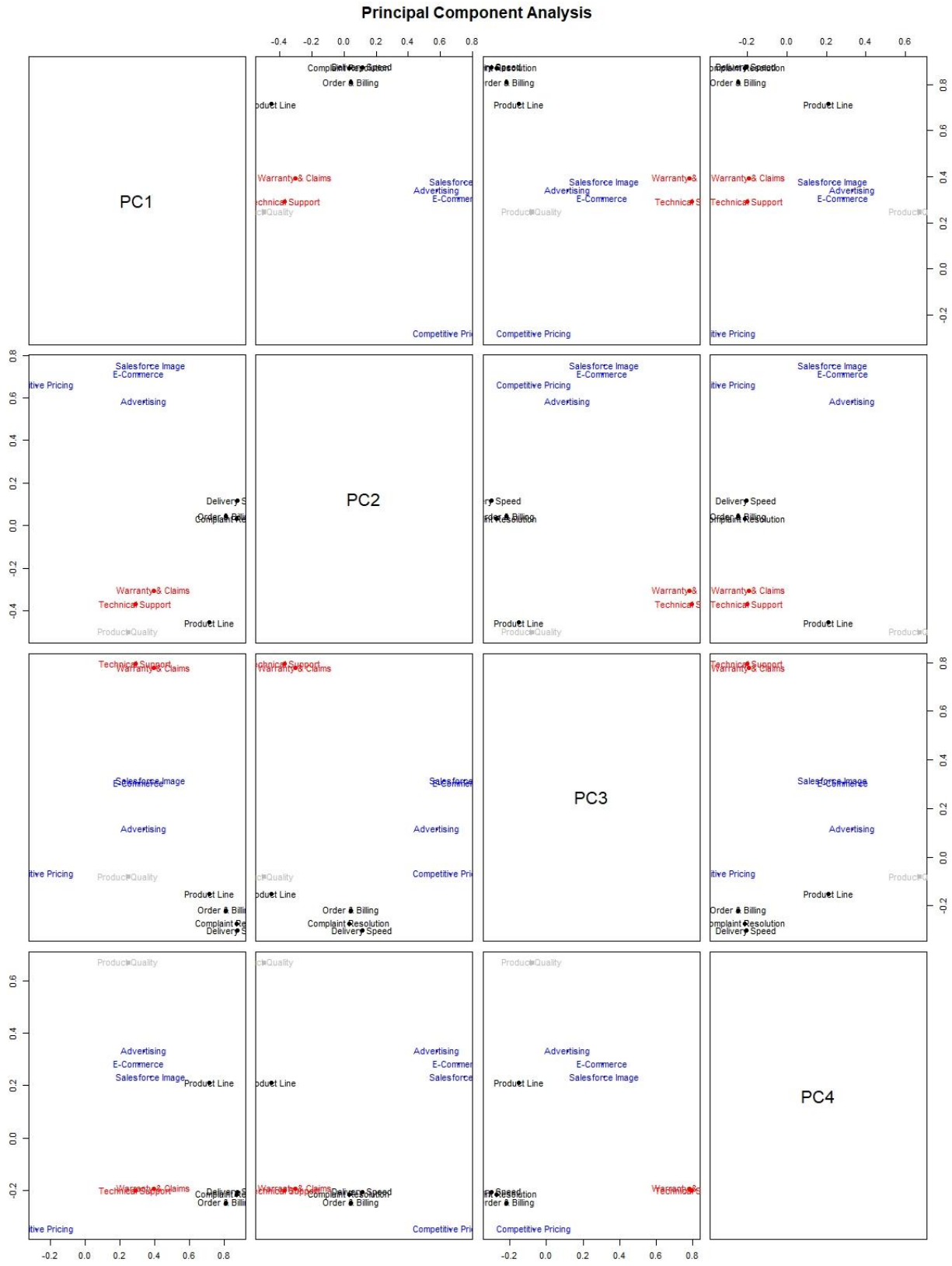
The next step in graphical representation is the Scree plot to identify and extract the factors that affect the target variable the most



Observation: As visible from the Graph four factors lie above the eigen value of 1, so four variables are enough to explain most of the variation.

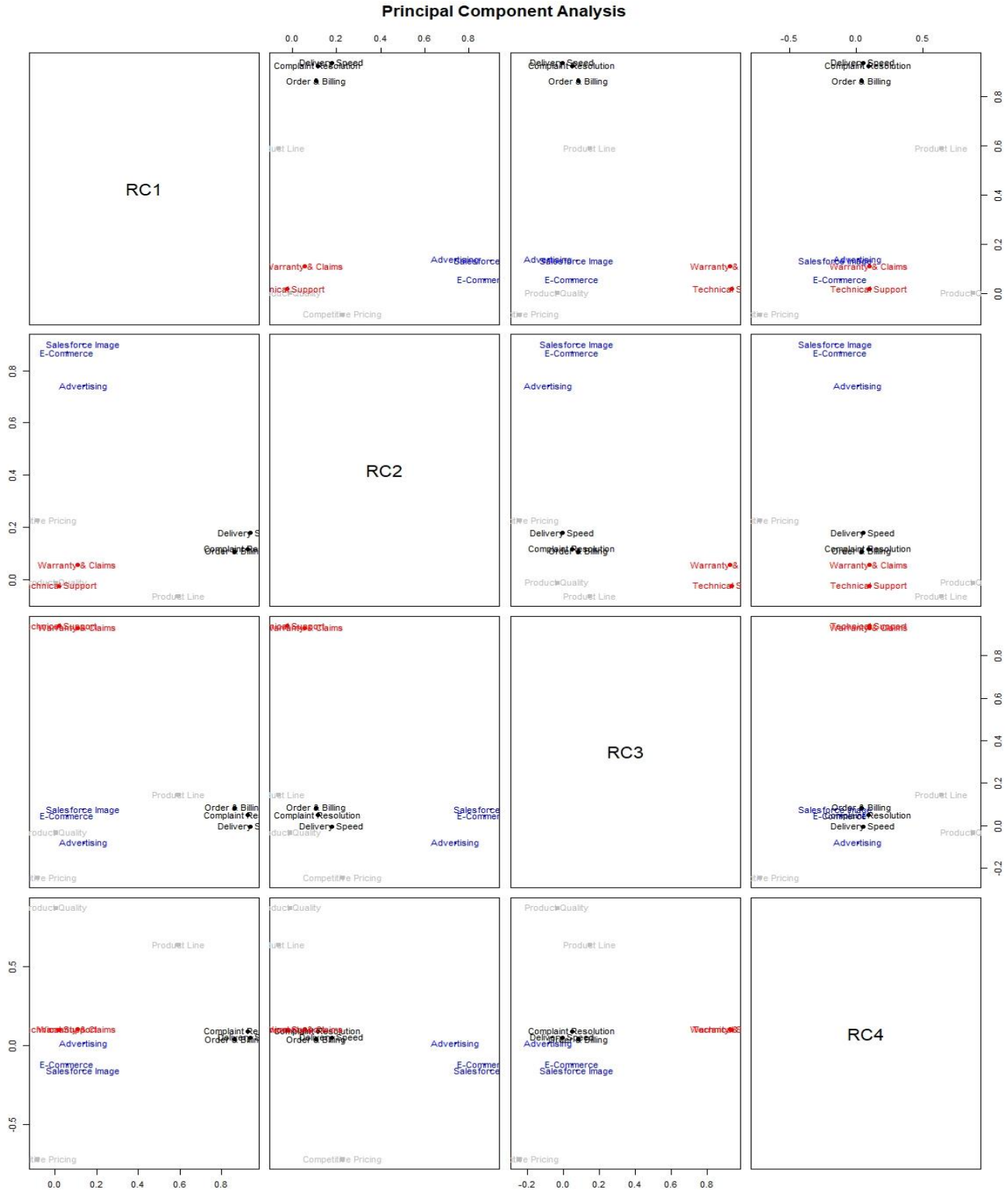
The scree plot helps us to determine the optimal number of components. As we see that first four variables have eigen values higher than one, so we choose these four variables hencefourth.

Graph6: Next step is PCA / FA analysis



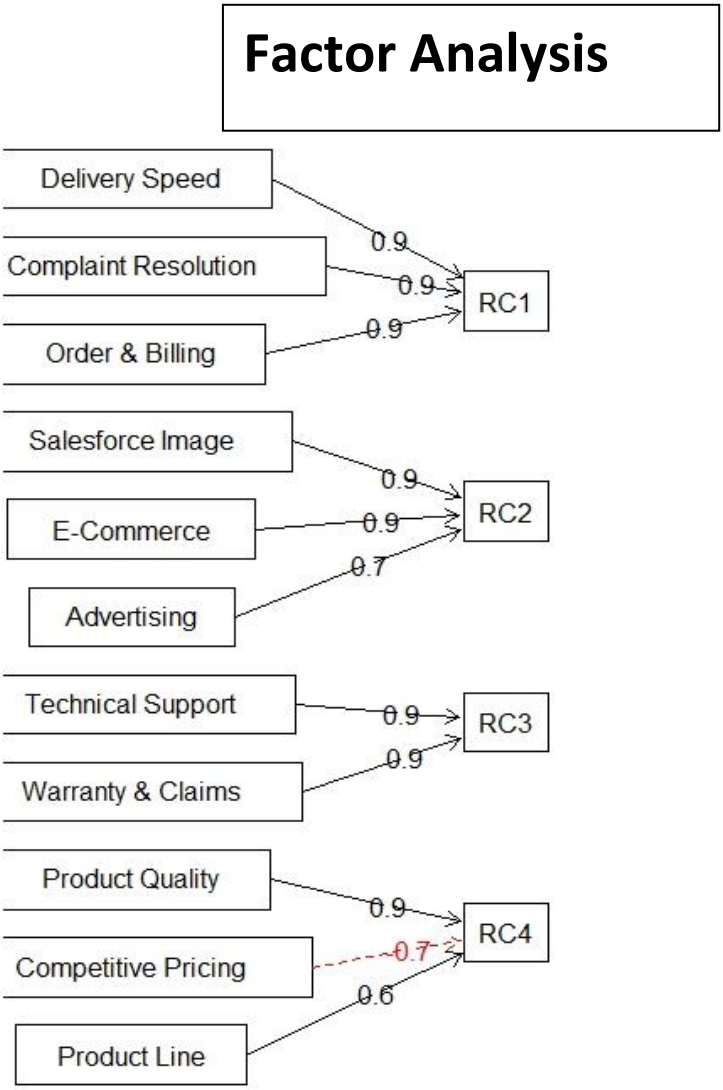
Observation: The plot of loadings of the four factor or cluster variables before rotation. The variables in red show the negative correlation and in blue the positive correlation. Still some ambiguity is their, so need to conduct an orthogonal rotation on the variables. So that the values closer to zero are pushed to zero and we have a clear picture.

Graph 7: Plot after rotation and applying Varimax rotation i.e. orthogonal rotation



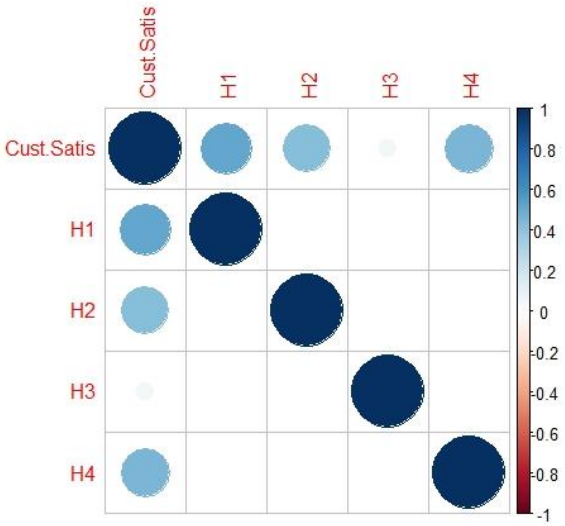
Observation: As observed the values / loadings has been pushed to the boundaries and now we have a clear picture of the cluster variables attributed to each of the four components or factors.

Graph 8: FA diagram is helpful in interpreting maximum loading on each of the factors



Observation: This rotated FA diagram shows good amount of loadings on each of the factors which is satisfactory to proceed forward.

Graph 8:



This correlation plot of the Factors / Components vs. the Dependent variable i.e. “Customer Satisfaction” , shows that no multicollinearity is existing in the present dimensionally reduced dataset. So all set to device the correlation and train and test the data.

The variables mentioned above have been clustered as follows:

H1, “Order Fulfillment KPI” is cluster of "Delivery Speed", "Complaint Resolution" and "Order $ Billing”

H2, “Brand Image” , cluster of variables is "Salesforce Image", "E-Commerce" & "Advertising" ;

H3, ”After Sales Support”, cluster of variables is "Technical Support" & "Warranty & Claims;

H4, “Product Differentiator” cluster of variables is "Product Quality", "Competitive Pricing" & "Product Line”

Next step is Calculating VIF

VIF Values:

H1	Order Fulfillment KPI	1.007
H2	Brand Image	1.034
H3	After Sales Support	1.009
H4	Product Differentiator	1.042

The Variance Inflation Factor (VIF) measures the impact of collinearity among the independent variables in a regression model

VIF is 1/Tolerance, it is always greater than or equal to 1. Small tolerance value indicates that the variables under consideration are almost a perfect linear combination of the independent variables.