Market Segmentation Customer Satisfaction-Product Service Management

K. R. R. Gandhi

***Abstract**: It is very relevant for a better management to calculate some satisfied index of consumers of a particular market, and of a consumer service. To achieve this, we add study of PCA, factor and regression to our data points and then by normalization we create a best suited model. Eventually, the model explains the effectiveness of a given set of customers ' happy list. All these measurements are given in a systematic order to measure happy index, including the source code.*

Keywords: PCA, Linear Regression, Factor Analysis

I. INTRODUCTION

As we all know, it is more important to measure customer satisfaction in any company. This calculation can be named as a happy index, and the essence of future business will be informed. Indeed, if the happy index is quite big, the great sales and profits quality in the company will lead. Therefore, many companies are quite happy to know the positive index of their client.

The happy index addresses the following: 1) It is easier to keep satisfied customers than to gain new ones 2) Customer satisfaction is even more critical than cost 3) Customer satisfaction puts your product in front of your rivals 4) Customer satisfaction encourages the retention of customers 5) Customer satisfaction encourages loyalty to the customer 6) Customer satisfaction decreases word of mouth negative.

We may understand the importance of having a balanced index of customer satisfaction to answer these all questions. As we said in the introduction, the following analysis would calculate all of these.

To calculate any happy index, we need to know the target, assumptions (if any), factor analysis, data import, packages and libraries (may be), variable classification, inferences, uni-variate analysis to lead correlation and variables normalization, and then PCA & Factor analysis by regression. As mentioned above, we will explain the approach in detail to complete the goal.

As we mentioned in the abstract, there are many methods and techniques available for calculating happy index using ANOVA or PCA, regression, and factor analysis.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author

Dr. K. R. R. Gandhi*, Professor, GITAM Institute of Management, GITAM (Deemed to be University), Visakhapatnam, India. Email: themath8@gmail.com

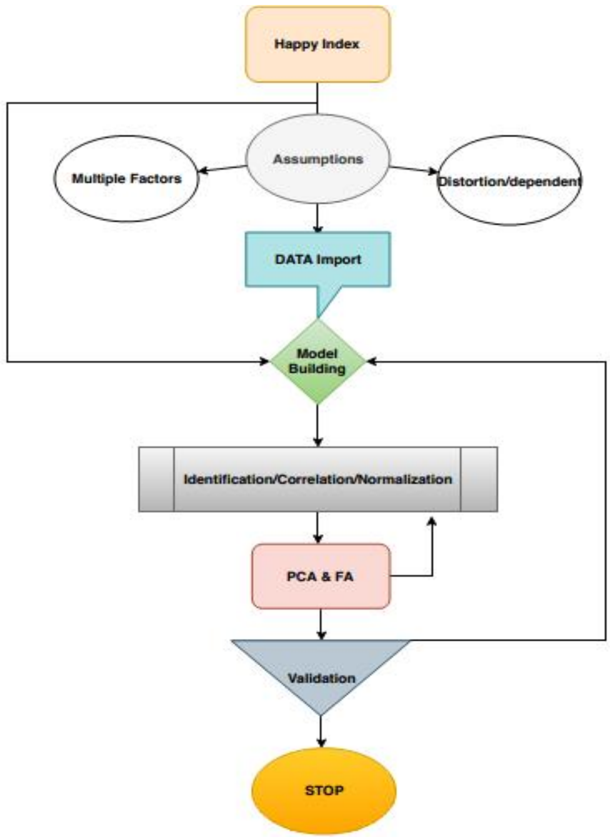
Nevertheless, to conclude section XII, we have given the systematic approach from sections II to XI.

Remember that at the end of the article, readers will understand the appendix of the source code type. We are not interested in definitions, as many of us found the definitions/meanings are understood by default.

II. PROPOSED METHODOLOGY

In this section, the proposed methodology for achieving the objective is discussed. As requested, we will add multiple factors and variables depending on the assumption. We will then work on the imported data set to build a model, keeping these assumptions in mind.

Under PCA & FA, we can think about correlation and standardization in order to know our goal. If not, we can change our setup from validation to reconstructing the model without losing the generality of our assumptions and the goal.



The above-mentioned flowchart gives the exact idea of finding a happy index or customer satisfaction for any particular segment of the market.

III. OBJECTIVE WITH ASSUMPTIONS

The following are some of the main objects that we assume as part of the methodology set out in Section II.

The report's main goal is to explore and build a linear data set model ("Factor-Hair-Revised") in R and to obtain insights from the data set. This report of exploration will consist of:

• Import in R • Multi-collinearity proof • Descriptive statistics • Factor analysis / PCA and factor recognition • Multi-linear model and validity development • Graphical exploration

Our assumptions are as follows:

a. Multiple factors influence the customer satisfaction score are the data provided.

b. All 12 variables influencing the dependent variable are presumed to be taken without distortion from a valid source.

IV. DATA IMPORT

We have imported data from a company to calculate through a certain origin. It's very hard to share or write all the data fields here. Nevertheless, the data in the CSV file can be found by reading our appendix at the end of the document. We are asking all our readers to follow a system or process for finding a happy index, but not the data submitted.

Required packages were installed in this section and related libraries were invoked. All packages at the same locations improve the readability of software Packages used: dplyr, psych, car, foreign lattice, MASS, ggplot2, caret

It makes it easier to import and export data files and code files by creating a working directory to begin the R session. The working directory is basically the location / folder on the PC where we have the project-related data, keys, etc. For source code, please refer to Appendix A (given at the end of this paper).

The data set is available in csv format. The command ' read.csv ' is therefore used to import the file. Please refer to the source code in Appendix A (given at the end of this paper).

Section V to XI will provide an idea of Identification, Correlation and Normalization by model. Often, we can consider the styles of testing under this model and re-construct the model when the validation fails. In this paper, we have made it clear that the model and analysis should be developed in accordance with our methodology described in Section-II.

V. MODEL VALIDATION: LINEAR REGRESSION AND FACTOR ANALYSIS

PCA does not make any such assumptions in order to identify the relevant factors performed Principle Component Analysis (PCA) or Factor Analysis from the data set containing large data values and FA assumes that there are few common factors driving the data variation. Using satisfaction as dependent variable and others as independent variable, the multi-linear regression model was created after the factors were identified.

1. Setting up system and importing DATA
2. Identification of variable
3. Analysis (Univariate)

VI. IDENTIFICATION OF VARIABLE

dim(mydata): Dim function is used to determine the total number of rows and columns. The data shown results have 100 rows and the function of 13 column names gives us information on the column names in the data.

attach(mydata): Attached to the search route is the file. This means that when evaluating a variable, the database is searched by R, so that objects can be accessed in the database simply by giving their names.

>str(mydata): Display the internal structure of a data.frame object R compactly: 100 obs. Of the thirteen 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 ...
```

Note that, all columns in the set of data are numeric except f or the int ID.

head(mydata):

```
ID ProdQualEcomTechSupCompRes Advertising ProdLineS
alesFImageComPricingWartyClaimOrdBillingDelSpeed
1 1    8.5  3.9    2.5    5.9      4.8    4.9     6.0     6.8
     4.7      5.0     3.7
2 2    8.2  2.7    5.1    7.2      3.4    7.9     3.1     5.3
     5.5      3.9     4.9
3 3    9.2  3.4    5.6    5.6      5.4    7.4     5.8     4.5
     6.2      5.4     4.5
4 4    6.4  3.3    7.0    3.7      4.7    4.7     4.5     8.8
     7.0      4.3     3.0
5 5    9.0  3.4    5.2    4.6      2.2    6.0     4.5     6.8
     6.1      4.5     3.5
6 6    6.5  2.8    3.1    4.1      4.0    4.3     3.7     8.5
     5.1      3.6     3.3
```

Retrieval Number: B6423129219/2019©BEIESP
DOI: 10.35940/ijitee.B6423.129219
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Satisfaction
1 8.2
2 5.7
3 8.9
4 4.8
5 7.1
6 4.7

>tail(mydata):

ID ProdQualEcomTechSupCompRes Advertising ProdLineSalesFImageComPricingWartyClaimOrdBillingDelSpeed
95 95 9.3 3.8 4.0 4.6 4.7 6.4 5.5
7.4 5.3 3.6 3.4
96 96 8.6 4.8 5.6 5.3 2.3 6.0 5.7
6.7 5.8 4.9 3.6
97 97 7.4 3.4 2.6 5.0 4.1 4.4 4.8
7.2 4.5 4.2 3.7
98 98 8.7 3.2 3.3 3.2 3.1 6.1 2.9
5.6 5.0 3.1 2.5
99 99 7.8 4.9 5.8 5.3 5.2 5.3 7.1
7.9 6.0 4.3 3.9
100 100 7.9 3.0 4.4 5.1 5.9 4.2 4.8
9.7 5.7 3.4 3.5
Satisfaction
95 7.7
96 7.3
97 6.3
98 5.4
99 6.4
100 6.4

>names(mydata):

[1] "ID" "ProdQual" "Ecom" "TechSup" "CompRes" "Advertising" "ProdLine" "SalesFImage"
[9] "ComPricing" "WartyClaim" "OrdBilling" "DelSpeed" "Satisfaction"

VII. CORRELATION MATRIX

>cor(mydata[,2:13]):

ProdQualEcomTechSupCompResAdvertising ProdLineSalesFImageComPricingWartyClaim

ProdQual 1.00000000 -0.1371632174 0.0956004542 0.1063700 -0.05347313 0.47749341 -0.15181287 -0.40128188 0.08831231
Ecom -0.13716322 1.0000000000 0.0008667887 0.1401793 0.42989071 -0.05268784 0.79154371 0.22946240 0.05189819
TechSup0.09560045 0.0008667887 1.0000000000 0.0966566 -0.06287007 0.19262546 0.01699054 -0.27078668 0.79716793
CompRes0.10637000 0.1401792611 0.0966565978 1.0000000 0.19691685 0.56141695 0.22975176 -0.12795425 0.14040830
Advertising -0.05347313 0.4298907110 -0.0628700668 0.1969168 1.00000000 -0.01155082 0.54220366 0.13421689 0.01079207
ProdLine 0.47749341 -0.0526878383 0.1926254565 0.5614170 -0.01155082 1.00000000 -0.06131553 -0.49494840 0.27307753
SalesFImage -0.15181287 0.7915437115 0.0169905395 0.2297518 0.54220366 -0.06131553 1.00000000 0.26459655 0.10745534
ComPricing -0.40128188 0.2294624014 -0.2707866821 -0.1279543 0.13421689 -0.49494840 0.26459655 1.00000000 -0.24498605
WartyClaim0.08831231 0.0518981915 0.7971679258 0.1404083 0.01079207 0.27307753 0.10745534 -0.24498605 1.00000000
OrdBilling0.10430307 0.1561473316 0.0801018246 0.7568686 0.18423559 0.42440825 0.19512741 -0.11456703 0.19706512
DelSpeed0.02771800 0.1916360683 0.0254406935 0.8650917 0.27586308 0.60185021 0.27155126 -0.07287173 0.10939460
Satisfaction 0.48632500 0.2827450147 0.1125971788 0.6032626 0.30466947 0.55054594 0.50020531 -0.20829569 0.17754482
OrdBillingDelSpeed Satisfaction
ProdQual0.10430307 0.02771800 0.4863250
Ecom0.15614733 0.19163607 0.2827450
TechSup0.08010182 0.02544069 0.1125972
CompRes0.75686859 0.86509170 0.6032626
Advertising 0.18423559 0.27586308 0.3046695
ProdLine0.42440825 0.60185021 0.5505459
SalesFImage0.19512741 0.27155126 0.5002053
ComPricing -0.11456703 -0.07287173 -0.2082957
WartyClaim0.19706512 0.10939460 0.1775448
OrdBilling1.00000000 0.75100307 0.5217319
DelSpeed0.75100307 1.00000000 0.5770423
Satisfaction 0.52173191 0.57704227 1.0000000

>names(mydata):

[1] "ID" "ProdQual" "Ecom" "TechSup" "CompRes" "Advertising" "ProdLine" "SalesFImage"
[9] "ComPricing" "WartyClaim" "OrdBilling" "DelSpeed" "Satisfaction"

cor(mydata[,2:13])

ProdQualEcomTechSupCompResAdvertising ProdLineSalesFImageComPricingWartyClaim

>mydatanum=mydata[-c(1,13)]

>names(mydatanum):
[1] "ProdQual" "Ecom" "TechSup" "CompRes" "Advertising" "ProdLine" "SalesFImage" "ComPricing" "WartyClaim"
[10] "OrdBilling" "DelSpeed"

VIII. ANALYSIS (UNIVARIATE)

Summary is a generic function used to summarize the results of the different fitting functions of the model. The function invokes specific methods that depend on the first Argument class.

> summary(mydatanum)

ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing
Min. : 5.000	Min. :2.200	Min. :1.300	Min. :2.600	Min. :1.900	Min. :2.300	Min. :2.900	Min. :3.700
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
Median : 8.000	Median :3.600	Median :5.400	Median :5.450	Median :4.000	Median :5.750	Median :4.900	Median :7.100
Mean : 7.810	Mean :3.672	Mean :5.365	Mean :5.442	Mean :4.010	Mean :5.805	Mean :5.123	Mean :6.974
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
Max. :10.000	Max. :5.700	Max. :8.500	Max. :7.800	Max. :6.500	Max. :8.400	Max. :8.200	Max. :9.900

WartyClaim	OrdBilling	DelSpeed
Min. :4.100	Min. :2.000	Min. :1.600
1st Qu.:5.400	1st Qu.:3.700	1st Qu.:3.400
Median :6.100	Median :4.400	Median :3.900
Mean :6.043	Mean :4.278	Mean :3.886
3rd Qu.:6.600	3rd Qu.:4.800	3rd Qu.:4.425
Max. :8.100	Max. :6.700	Max. :5.500

Description of the data shows median and average is very similar, usually the data are distributed.

>mydataCorr<- cor(mydatanum)

>mydataCorr

>corrdf<- data.frame(mydataCorr)

>write.csv(corrdf,"correlationmatrix.csv")

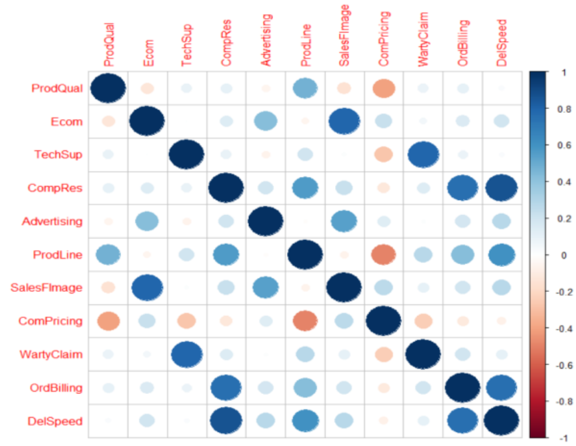
IX. CORRELATION AND NORMALIZATION

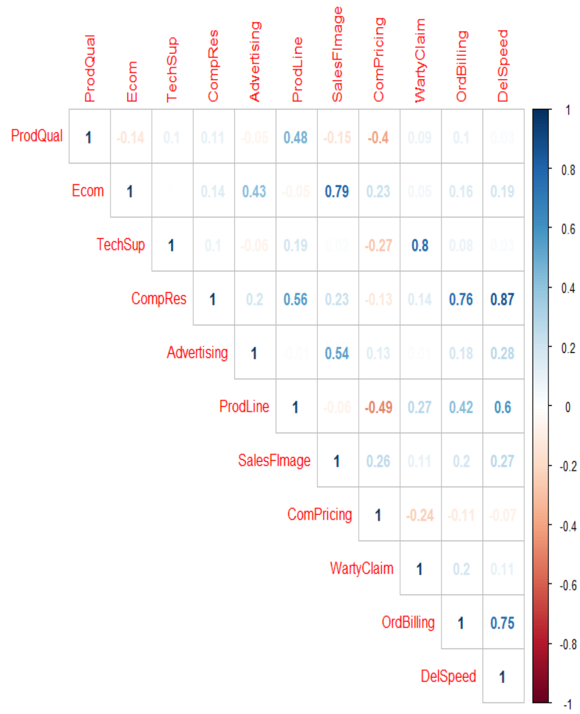
#Shapiro test
>shapiro.test(mydatanum$ProdQual) Shapiro-Wilk normality test data: mydatanum$ProdQual W = 0.94972, p-value = 0.0007953
>shapiro.test(mydatanum$Ecom) Shapiro-Wilk normality test data: mydatanum$Ecom W = 0.95852, p-value = 0.003157 In above 2 cases P- value is less than 0.05 so it is normal.
>shapiro.test(mydatanum$TechSup) Shapiro-Wilk normality test data: mydatanum$TechSup W = 0.98626, p-value = 0.39 P-value is greater than 0.05 so it is not normal.

	ProdQual	Ecom	TechSup	CompRes	Advertisin	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed
ProdQual	1	-0.13716	0.0956	0.10637	-0.05347	0.477493	-0.151812874	-0.401281884	0.088312306	0.104303	0.027718
Ecom	-0.13716	1	0.000867	0.140179	0.429891	-0.05269	0.791543711	0.229462401	0.051898192	0.156147	0.19163607
TechSup	0.0956	0.000867	1	0.096657	-0.06287	0.192625	0.01699054	-0.270786682	0.797167926	0.080102	0.02544069
CompRes	0.10637	0.140179	0.096657	1	0.196917	0.561417	0.229751761	-0.127954253	0.140408297	0.756869	0.8650917
Advertisin	-0.05347	0.429891	-0.06287	0.196917	1	-0.01155	0.542203658	0.134216894	0.010792074	0.184236	0.27586308
ProdLine	0.477493	-0.05269	0.192625	0.561417	-0.01155	1	-0.061315528	-0.494948402	0.273077528	0.424408	0.60185021
SalesFIma	-0.15181	0.791544	0.016991	0.229752	0.542204	-0.06132	1	0.264596554	0.107455345	0.195127	0.27155126
ComPricin	-0.40128	0.229462	-0.27079	-0.12795	0.134217	-0.49495	0.264596554	1	-0.244986054	-0.11457	-0.0728717
WartyClai	0.088312	0.051898	0.797168	0.140408	0.010792	0.273078	0.107455345	-0.244986054	1	0.197065	0.1093946
OrdBilling	0.104303	0.156147	0.080102	0.756869	0.184236	0.424408	0.195127406	-0.114567026	0.197065121	1	0.75100307
DelSpeed	0.027718	0.191636	0.025441	0.865092	0.275863	0.60185	0.271551259	-0.072871729	0.109394602	0.751003	1

We can see in the table above that diagonal elements are highly correlated with value as 1 and are distributed between -1 and + 1.

Another way to identify the connection is to use the corrplot below.





In the above graphs, the symptoms of multi-colinearity can be clearly seen as certain variables are highly correlated. If you can see the plot of the ring, the darker the color comparison will be, as salesFimage and Ecom are associated, like Compres vs delspeed and much more.

If they are negatively correlated, the hue will be orange and then blue if it is positively correlated. The two different types of plots show the number and circle (upper type). So enough evidence of causation.

X. PCA Vs FACTOR ANALYSIS

Before operating, certain packages and library must be installed:

```
install.packages("car")
install.packages("nortest")
install.packages("GPArotation")
library(car)
library(foreign)
library(lattice)
library(MASS)
library(GPArotation)
library(psych)
```

Bartlett sphercity test for checking the data dimen sion reduction possibility or PCA:

```
>cortest.bartlett(mydataCorr,nrow(mydata))

$chisq
[1] 619.2726

$p.value
[1] 1.79337e-96

$df
[1] 55
```

P-value is less than 0.05 then it is possible to redue the size o r PCA in our case

#finding out the Eigen values and vectors:

```
>A <-eigen(mydataCorr)
>eigenvalues<- A$values
>eigenvector<- A$vectors
>print(eigenvalues,digits = 3)

 [1] 3.4270 2.5509 1.6910 1.0866 0.6094 0.5519 0.4015 0.2470 0.2036 0.1328 0.0984
```

As per Kaiser rule we need to consider eigen values > 1 hence PC1-PC4 which contains maximum possible inf ormation. This will help us in understanding the variance.

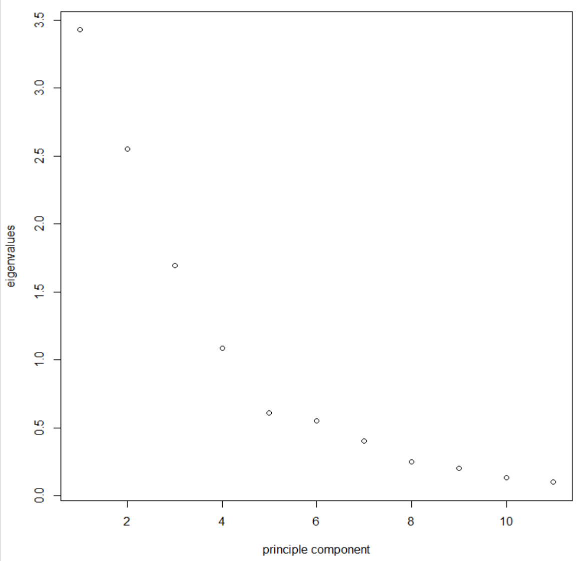
```
>part.pca<- eigenvalues/sum(eigenvalues)*100
>part.pca

 [1] 31.1542848 23.1899701 15.3725134  9.8777823  5.5402190  5.0171253  3.6501650  2.2450140  1.8504843  1.2076507  0.8947911
```

The first 4 factors are having the maximum infor mation and rest are ignored as they contain very less informat ion about variance.

Plotting scree graphs:

```
>plot(eigenvalues, types ="lines", xlab ="principle component", ylab ="eigenvalues")
```



#Component loading and PCA with rotation/unrotate:

```
>PC4_unrotate <- principal(mydatanum, nfactors = 4, rotate = "none")

>PC4_unrotate
```

Principal Components Analysis

```
Call: principal(r = mydatanum, nfactors = 4, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix

PC1   PC2   PC3   PC4   h2    u2 com
ProdQual     0.25 -0.50 -0.08  0.67 0.77 0.232 2.2
Ecom0.31  0.71  0.31  0.28 0.78 0.223 2.1
TechSup      0.29 -0.37  0.79 -0.20 0.89 0.107 1.9
CompRes0.87  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 2.0
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 2.0
OrdBilling0.81  0.04 -0.22 -0.25 0.77 0.234 1.3
DelSpeed0.88  0.12 -0.30 -0.21 0.91 0.086 1.4

PC1  PC2  PC3  PC4
SS loadings           3.43 2.55 1.69 1.09
Proportion Var         0.31 0.23 0.15 0.10
Cumulative Var          0.31 0.54 0.70 0.80
Proportion Explained  0.39 0.29 0.19 0.12
Cumulative Proportion 0.39 0.68 0.88 1.00

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

The root mean square of the residuals (RMSR) is  0.06
with the empirical chi square  39.02  with prob<  0.0018
Fit based upon off diagonal values = 0.97
```

```
>PC4_unrotate$loadings

Loadings:
          PC1    PC2    PC3    PC4
ProdQual    0.248 -0.501        0.670
Ecom0.307  0.713  0.306  0.284
TechSup     0.292 -0.369  0.794 -0.202
CompRes     0.871         -0.274 -0.215
Advertising  0.340  0.581  0.115  0.331
ProdLine    0.716 -0.455 -0.151  0.212
SalesFImage  0.377  0.752  0.314  0.232
ComPricing  -0.281  0.660         -0.348
WartyClaim   0.394 -0.306  0.778 -0.193
OrdBilling   0.809         -0.220 -0.247
DelSpeed0.876  0.117 -0.302 -0.206

               PC1   PC2   PC3   PC4
SS loadings    3.427 2.551 1.691 1.087
Proportion Var 0.312 0.232 0.154 0.099
Cumulative Var 0.312 0.543 0.697 0.796
```

If we add PC1 to PC3, the cumulative variable will be equiva lent to PC4, which accounts for 79.6% of the data variance. This means that all four factors are approx. 80% that will cor rectly tell about the rest of the data 20% can be taken from ot her factors that are discarded above.

```
>PC4_rotate <- principal(mydatanum, nfactors = 4, rotate = "varimax")

>PC4_rotate

Principal Components Analysis:

Call: principal(r = mydatanum, nfactors = 4, rotate = "varima x")
Standardized loadings (pattern matrix) based upon correlatio n matrix

RC1   RC2   RC3   RC4   h2    u2 com
ProdQual    0.00 -0.01 -0.03  0.88 0.77 0.232 1.0
Ecom0.06  0.87  0.05 -0.12 0.78 0.223 1.1
TechSup     0.02 -0.02  0.94  0.10 0.89 0.107 1.0
CompRes0.93  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  0.90  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
WartyClaim0.11  0.05  0.93  0.10 0.89 0.108 1.1
OrdBilling0.86  0.11  0.08  0.04 0.77 0.234 1.1
DelSpeed0.94  0.18  0.00  0.05 0.91 0.086 1.1

RC1  RC2  RC3  RC4
SS loadings           2.89 2.23 1.86 1.77
Proportion Var        0.26 0.20 0.17 0.16
Cumulative Var        0.26 0.47 0.63 0.80
Proportion Explained  0.33 0.26 0.21 0.20
Cumulative Proportion 0.33 0.59 0.80 1.00

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

The root mean square of the residuals (RMSR) = 0.06
with the empirical chi square  39.02  with prob<  0.0018

Fit based upon off diagonal values = 0.97

>PC4_rotate$loadings

Loadings:
          RC1    RC2    RC3    RC4
ProdQual                       0.876
Ecom               0.871        -0.117
TechSup0.939  0.101
CompRes0.926  0.116
Advertising  0.139  0.742
ProdLine     0.591         0.146  0.642
SalesFImage  0.133  0.900         -0.159
ComPricing          0.226 -0.246 -0.723
WartyClaim   0.110         0.931  0.102
OrdBilling0.864  0.107
DelSpeed0.938  0.177

               RC1   RC2   RC3   RC4
SS loadings    2.893 2.234 1.856 1.774
Proportion Var 0.263 0.203 0.169 0.161
Cumulative Var 0.263 0.466 0.635 0.796
```

When you equate the unrotate vs rotation aggregate function, the variance stays the same, i.e. 79.6 percent. And here RC4 and RC1 have the total parameter that influences the variabili ty.

```
>newdf= PC4_rotate$scores
>newdf= as.data.frame(newdf)
>reg_PCA = cbind(mydata$Satisfaction, newdf)
>names(reg_PCA)
[1] "mydata$Satisfaction"

"RC1"             "RC2"             "RC3"             "RC4"
```

It involves the ' satisfaction ' dependent variable and other 4 variables that are most likely to affect the data set variability.

```
>names(reg_PCA)= c("Satisfaction","1","2","3","4")
>head(reg_PCA)
  Satisfaction         1          2            3          4
1          8.2  0.1274910  0.7698686 -1.878446273  0.3664848
2          5.7  1.2216666 -1.6458617 -0.614030010  0.8130648
3          8.9  0.6158214  0.5800037  0.003689252  1.5699769
4          4.8 -0.8446267 -0.2719218  1.267493254 -1.2541645
5          7.1 -0.3197943 -0.8340650 -0.008096627  0.4475377
6          4.7 -0.6470292 -1.0672683 -1.303198892 -1.0527792
```

Factor Analysis:

In factor analysis, the same adequacy of KMO is tested along with own values and the calculation of vectors as present in PCA. So move directly to the role of FA:

```
>KMO(r=mydataCorr)
```

Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = mydataCorr)
Overall MSA = 0.65
MSA for each item =
ProdQualEcomTechSupCompRes Advertising ProdLineSalesFImageComPricingWartyClaimOrdBilling
0.51 0.63 0.52 0.79 0.78 0.62
0.62 0.75 0.51 0.76
DelSpeed
0.67

factor analysis without rotation:

	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

Factor analysis with rotation:

```
>solution1 <- fa(r=mydataCorr, nfactors = 4, rotate = "varimax", fm ="pa")
>solution1
```

Factor Analysis using method = pa
Call: fa(r = mydataCorr, 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

In this case, the average parameter for both cases is 0.69, which means that the variance of 69 percent is measured using the four variables in the data set. The mixture of other variables remains the variability. All four factors are highly correlated, as you can check the PA1 to PA4 matrix with scores of 0.98, 0.99, 0.94, and 0.88.

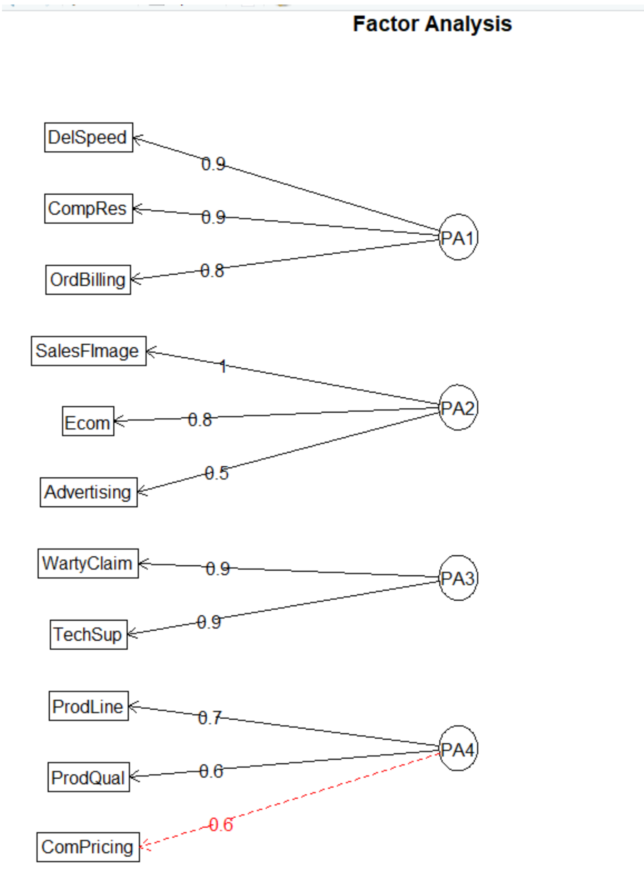
```
>names(reg_fa)

[1] "mydata$Satisfaction" "RC1"      "RC2"          "RC3"
 "RC4"
```

>names(reg_fa)= c("Satisfaction","5","6","7","8")
>head(reg_fa)

	Satisfaction	5	6	7	8
1	8.2	0.1274910	0.7698686	-1.878446273	0.3664848
2	5.7	1.2216666	-1.6458617	-0.614030010	0.8130648
3	8.9	0.6158214	0.5800037	0.003689252	1.5699769
4	4.8	-0.8446267	-0.2719218	1.267493254	-1.2541645
5	7.1	-0.3197943	-0.8340650	-0.008096627	0.4475377
6	4.7	-0.6470292	-1.0672683	-1.303198892	-1.0527792

It includes the matrix of all factors affecting the variability and the dependent variable ' satisfaction ' the diagram below shows the relationship between PA1 and PA4 with various factors present in the dataset, the red color 0.6 indicates the negative relationship between the detailed and PA4.



XI. MULTI LINEAR REGRESSION MODEL

>model1=lm(Satisfaction ~1,data=reg_fa)
>summary(model1)

Call:
lm(formula = Satisfaction ~ 1, data = reg_fa)

Residuals:
Min 1Q Median 3Q Max
-2.218 -0.918 0.132 0.707 2.982

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.9180	0.1192	58.05	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.192 on 99 degrees of freedom

Here we can see that the model is very stable and important with a confidence level of 97.5 percent. Even the p-value is less than 0.05 so the null hypothesis is rejected and the model accepted. In order to accurately calculate the dependent variable satisfaction rate in the data set, the variable frequency, compression and order billing independent variables are highly significant.

>model3=lm(Satisfaction ~ 3,data=reg_fa)

>summary(model3)

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

It is less significant and p is higher than 0.05 and therefore accepts the null hypothesis and rejects the linear equation. The level of significance is also lower. Consequently, variables ' wartyclaim ' and ' Techsup ' do not influence the dependent variable in the model and are therefore less likely to affect the variance.

XII. OBSERVATIONS

The following findings were noted in the PCA and FA versions of our model. As a result, the affected variables are calculated and can therefore be expected to have more score under the happy section scale.

1. All variables are taken into account for the identification of the four variables in factor analysis and PCA.

2. As shown in the factor analysis diagram, the four factors RC1 to RC4 directly affect the variability of the dataset.

3. Mainly PA1 consisting of "Del pace, compress, and order billing" is the most important compared to other factors that affect the design.

4. The linear model1 has a 97.5 percent confidence level and a high value.

5. The model is accurate and the variables of PA1 are most likely to affect the score of the relevant dependent variable.

XIII. CONCLUSIONS & FUTURE SCOPE

In this article, we discussed the calculation of PCA and PA within order to know the customer satisfaction of the segment by calling as happy index. The causal factors are evaluated and the RC1 to RC4 variability can be seen under this model.

If we have more RC's (may be up to RC10), we need to modify our flow chart as well as our PCA to measure the index. Readers can work with more RC's or identify the pitfalls of the model, or build a new model through validation.

We believe that some more factors should be added under our assumption (given in section-II) when we have more RCs. As a result, readers can work under these parameters to measure the effect.

APPENDIX (SOURCE CODE)

```
setwd("D:/RSMDM ")
mydata=read.csv("Factor-Hair-Revised.csv",header =
TRUE)
mydata
attach(mydata)
#Descriptive Statistics
dim(mydata)

names(mydata)
summary(mydata)
str(mydata)
head(mydata)
mydatanum=mydata[-c(1,13)]
names(mydatanum)
summary(mydatanum)

sd(ProdQual)
sd(Ecom)
shapiro.test(mydatanum$ProdQual)
shapiro.test(mydatanum$Ecom)
shapiro.test(mydatanum$TechSup)
library(corrplot)
mydataCorr<- cor(mydatanum)
mydataCorr
corrdf<- data.frame(mydataCorr)
write.csv(corrdf,"correlationmatrix.csv")
corrplot(mydataCorr)
corrplot(mydataCorr, method = 'number', type = "upper")
```

#-Principle component analysis#

```
install.packages("car")
install.packages("nortest")
install.packages("GPArotation")
library(car)
library(foreign)
library(lattice)
library(MASS)
library(GPArotation)
library(psych)

cortest.bartlett(mydataCorr,nrow(mydata))
A <-eigen(mydataCorr)
eigenvalues<- A$values
eigenvector<- A$vectors
print(eigenvalues,digits = 3)
eigenvector
part.pca<- eigenvalues/sum(eigenvalues)*100
part.pca
plot(eigenvalues, types ="lines", xlab ="principle
component", ylab ="eigenvalues")
PC4_unrotate <- principal(mydatanum, nfactors = 4, rotate =
"none")
PC4_unrotate
PC4_unrotate$loadings
PC4_rotate <- principal(mydatanum, nfactors = 4, rotate =
"varimax")
PC4_rotate
PC4_rotate$loadings
PC4_rotate$scores
dim(PC4_rotate$scores)
dim(mydata)
dim(mydatanum)
newdf= PC4_rotate$scores
newdf= as.data.frame(newdf)
reg_PCA = cbind(mydata$Satisfaction, newdf)
reg_PCA
names(reg_PCA)
names(reg_PCA)= c("Satisfaction","1","2","3","4")
head(reg_PCA)

mydataCorr<-(mydatanum)
mydataCorr
KMO(r=mydataCorr)

A <-eigen(mydataCorr)
eigenvalues<- A$values
eigenvector<- A$vectors
eigenvalues
plot( eigenvalues, types = "lines", xlab = "principal
component", ylab = "eigenvalues")
solution<- fa(r=mydataCorr, nfactors = 4, rotate = "none", fm
="pa")
solution
solution1 <- fa(r=mydataCorr, nfactors = 4, rotate =
"varimax", fm ="pa")
solution1
fa.diagram(solution1, sample=FALSE)
newdf1 = PC4_rotate$scores
newdf1 = as.data.frame(newdf1)
reg_fa = cbind(mydata$Satisfaction, newdf1)
names(reg_fa)
names(reg_fa)= c("Satisfaction","5","6","7","8")
head(reg_fa)

model1=lm(Satisfaction ~1,data=reg_fa)
summary(model1)
model2=lm(Satisfaction ~2,data=reg_fa)
summary(model2)
```

```
model3=lm(Satisfaction ~ 3,data=reg_fa)
summary(model3)
model4=lm(Satisfaction ~ 4,data=reg_fa)
summary(model4)
confint(model1,"1")
confint(model2,"2")
confint(model3,"3")
```


ACKNOWLEDGMENT

I am very grateful to Honorable Vice Chancellor **Prof. K. Siva Rama Krishna,** GITAM (Deemed to be University), and my Director, GITAM Institute of Management (Deemed to be University), **Prof. Y. Gouthama Rao** for their continuous encouragement to prepare this article.

REFERENCES

1. Alessandro Ruggieri, Cecilia Silvestri "Quality and Customer Satisfaction: Relationship and Customer Behavior. A Case Study," International Journal of Managerial Studies and Research (IJMSR), Volume 2, Issue 10, November 2014, PP 22-35
2. Pratheepkanth Puwanenthiren, Ymwgpk Udurawana "Market Segmentation and Its Impact on Customer Satisfaction with Especial Reference to Commercial Bank of Ceylon PLC," Marketing Information & Strategies: Core for Development, Scitech Publications (India) Pvt Ltd (January 1, 2014) 17 Pages
3. Wang Xia, Zhao Ping, Wang Gao, Liu Jia "Market segmentation based on customer satisfaction-loyalty links," Frontiers of Business Research in China May 2007, Volume 1, Issue 2, pp 211–221
4. Festus M Epetimehin "Market Segmentation: A Tool for Improving Customer Satisfaction and Retention in Insurance Service Delivery," Journal of Emerging Trends in Economics and Management Sciences (JETEMS) 2(1):62-67

AUTHORS PROFILE

Dr. K. R. R. Gandhi, Professor of Operations, GITAM Institute of Management, GITAM University, Visakhapatnam. Dr. Gandhi is a renowned Mathematician and is a Resource person in Mathematics for Oxford University Press. His passion for Pure Mathematics started at an early age and his love for Number Theory has grown exponentially since then. He did immense contributions in the field of Number theory and has published several papers and monographs in quite a few international journals, and few patents on his name. Dr. Gandhi is always seen doing research in applications of Mathematics, especially in Data Science. When asked what he does in his spare time, he smiles and says "More Research". He is a reviewer, referee and editorial member of several international journals. He has been conferred the best teacher award by JNTUK in 2012, and best researcher by RBI in 2013 and in 2014. Dr. Gandhi has vast administrative experience and held eminent positions at Govt. of Andhra Pradesh, such as Member of State Panning Board; and Vice President, Voice of Big Data. Dr. Gandhi has been the life member of Calcutta Mathematical Society (CMS), Indian Mathematical Society (IMS), Ramanujan Mathematical Society (RMS), and member in SOMASS.