Business Report
Advance Statistics

Arnab Ghosal

30 April 2021

Problem - ANOVA

The staff of a service centre for electrical appliances include three technicians who specialise in repairing three widely used electrical appliances by three different manufacturers. It was desired to study the effects of Technician and Manufacturer on the service time. Each technician was randomly assigned five repair jobs on each manufacturer's appliance and the time to complete each job (in minutes) was recorded. The data for this particular experiment is thus attached.

Questions:

1) State the Null and Alternate Hypothesis for conducting one-way ANOVA for both the variables 'Manufacturer' and 'Technician' individually. – 3 points

Hypothesis for conducting **one-way ANOVA** for **Manufacturer** with respect to **Service time**

- HO: There is no significant effect in terms of Service time based on **Manufacturer**. (**HO : µ1=µ2**) (**Null Hypothesis**)
- Ha: There is a significant effect in terms of Service time based on **Manufacturer**. (**HO : µ1!=µ2**) (**Alternative Hypothesis**)

Hypothesis for conducting **one-way ANOVA** for **Technician** with respect to **Service time**

- HO: There is no significant effect in terms of Service time based on **Technician**. (**HO : µ1=µ2**) (**Null Hypothesis**)
- Ha: There is a significant effect in terms of Service time based on **Technician**. (**HO : µ1!=µ2**) (**Alternative Hypothesis**)

2) Perform one-way ANOVA for variable 'Manufacturer' with respect to the variable 'Service Time'. State whether the Null Hypothesis is accepted or rejected based on the ANOVA results. - 3 points

	df	sum_sq	mean_sq	F	PR(>F)
C(Manufacturer)	2.0	28.311111	14.155556	0.191029	0.826822
Residual	42.0	3112.266667	74.101587	NaN	NaN

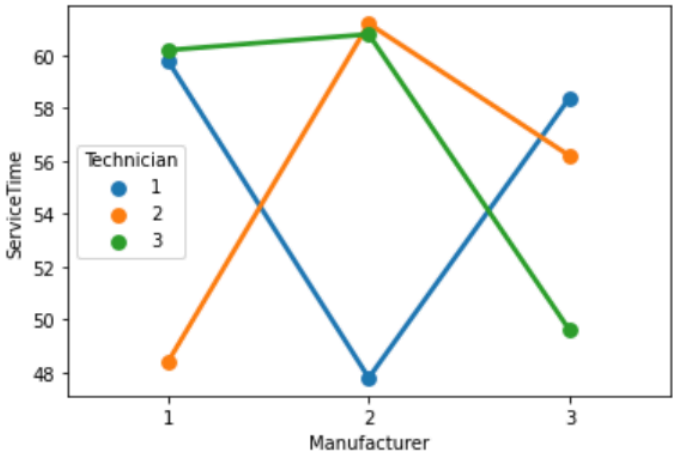
From this above **ANOVA table**, we can see that here **P value** is **greater than (P > 0.05) level of significance** , hence **we fail to reject the null hypothesis**.

3) Perform one-way ANOVA for variable 'Technician' with respect to the variable 'Service Time'. State whether the Null Hypothesis is accepted or rejected based on the ANOVA results. - 3 points

	df	sum_sq	mean_sq	F	PR(>F)
C(Technician)	2.0	24.577778	12.288889	0.16564	0.847902
Residual	42.0	3116.000000	74.190476	NaN	NaN

From this above **ANOVA table**, we can see that here **P value** is **greater than (P > 0.05) level of significance** , hence **we fail to reject the null hypothesis**.

4) Analyse the effects of one variable on another with the help of an interaction plot. What is an interaction between two treatments?



- **Technician 2** with **Manufacturer 1** & **Technician 1** with **Manufacturer 2** takes almost same service time (which is close to **48 mins**).
- **Technician 1** & **Technician 3** with **Manufacturer 1** takes almost similar service time (which is ~**60 mins**).
- **Technician 2** & **Technician 3** with **Manufacturer 2** takes almost similar service time (which is little over **60 mins**).

- **Technician 1** takes maximum service time with **Manufacturer 1** (which is ~**60 mins**) and minimum service time with **Manufacturer 2** (which is ~**48 mins**).

- **Technician 2** takes maximum service time with **Manufacturer 2** (which is little over **60 mins**) & minimum service time with **Manufacturer 1** (which is ~**48 mins**).

- **Technician 3** takes maximum service time with **Manufacturer 2** (which is ~**60 mins**) & minimum service time with **Manufacturer 3** (which is little lesser than **50 mins**).

5) Perform a two-way ANOVA based on the variables 'Manufacturer' & 'Technician' with respect to the variable 'Service Time' and state your results. - 5 points

*Performing **two-way ANOVA without interaction** of two variables (**Manufacturer & Technician**)*

	df	sum_sq	mean_sq	F	PR(>F)
C(Manufacturer)	2.0	28.311111	14.155556	0.183381	0.833147
C(Technician)	2.0	24.577778	12.288889	0.159199	0.853365
Residual	40.0	3087.688889	77.192222	NaN	NaN

*Performing **two-way ANOVA with interaction** of two variables (**Manufacturer & Technician**)*

	df	sum_sq	mean_sq	F	\
C(Manufacturer)	2.0	28.311111	14.155556	0.272164	
C(Technician)	2.0	24.577778	12.288889	0.236274	
C(Manufacturer):C(Technician)	4.0	1215.288889	303.822222	5.841487	
Residual	36.0	1872.400000	52.011111	NaN	

	PR(>F)
C(Manufacturer)	0.763283
C(Technician)	0.790779
C(Manufacturer):C(Technician)	0.000994
Residual	NaN

From the above **two-way ANOVA table** we can see for both variable "**Manufacturer**" & "**Technician**" **P value is greater than (P > 0.05) level of significance** , hence **we fail to reject the null hypothesis here** , so we can conclude that "**Manufacturer**" & "**Technician**" are not really a significant factor in terms of deciding the hypothesis.

But if we look at the **interaction** between "**Manufacturer**" *&* "**Technician**" , in this case **P value < 0.05** (**level of significance**), Hence **we reject null hypothesis** . We can understand the **interaction** b/w "**Manufacturer**" *&* "**Technician**" playing a major role in terms of deciding the hypothesis .

6) Mention the business implications of performing ANOVA for this particular case study. - 5 points

- **Technician 1** takes maximum service time with **Manufacturer 1** (which is ~**60 mins**) and minimum service time with **Manufacturer 2** (which is ~**48 mins**).

- **Technician 2** takes maximum service time with **Manufacturer 2** (which is little over **60 mins**) *&* minimum service time with **Manufacturer 1** (which is ~**48 mins**).

- **Technician 3** takes maximum service time with **Manufacturer 2** (which is ~**60 mins**) *&* minimum service time with **Manufacturer 3** (which is little lesser than **50 mins**).

- **Manufacturer 3** looks like **most efficient** among all 3 manufacturer as on an avg. all 3 technician takes comparatively **lesser service time** .

- **Manufacturer 2** looks like the **least efficient** as out of 3 there are 2 technician takes **maximum service time** with **Manufacturer 2** equipments.

- **Technician 2** looks **most efficient** as he/she takes comparatively **lesser service time** with all 3 manufacturers .

- **Technician 3** looks least efficient as he/she takes **comparatively more service time** with all 3 manufacturers .

- **Interaction** b/w "**Manufacturer**" *&* "**Technician**" playing a major role in

Problem - PCA

The 'Hair Salon.csv' dataset contains various variables used for the context of Market Segmentation. This particular case study is based on various parameters of a salon chain of hair products. You are expected to do Principal Component Analysis for this case study according to the instructions given in the following rubric.

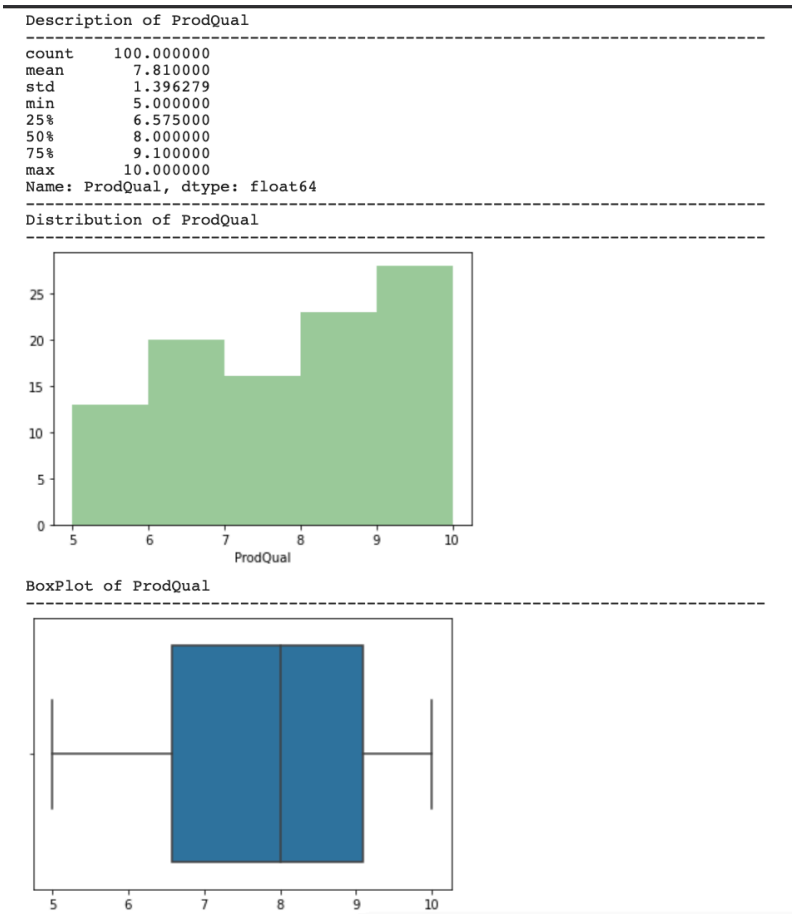
Note: *This particular dataset contains the target variable satisfaction as well. Please do drop this variable before doing Principal Component Analysis.*

Questions:

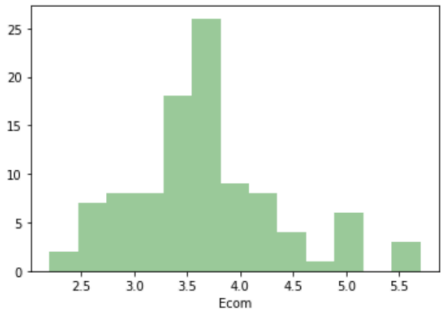
1) Perform Exploratory Data Analysis [both univariate and multivariate analysis to be performed]. The inferences drawn from this should be properly documented. – 5 points

Univariate Analysis :

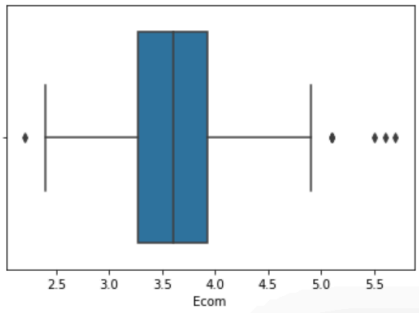
Univariate analysis is the simplest form of analysing data. "Uni" means "one", so in other words our data takes only one variable at a time. It doesn't deal with causes or relationships (unlike regression) and it's major purpose is to describe; It takes data (individual column) separately , summarises that data and finds patterns in the data.



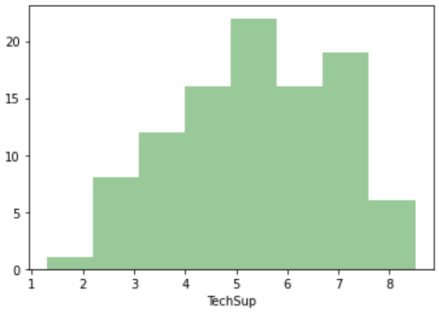
```
Description of Ecom
--------------------------------------------------------------------------------
count    100.000000
mean       3.672000
std        0.700516
min        2.200000
25%        3.275000
50%        3.600000
75%        3.925000
max        5.700000
Name: Ecom, dtype: float64
--------------------------------------------------------------------------------
Distribution of Ecom
--------------------------------------------------------------------------------
```



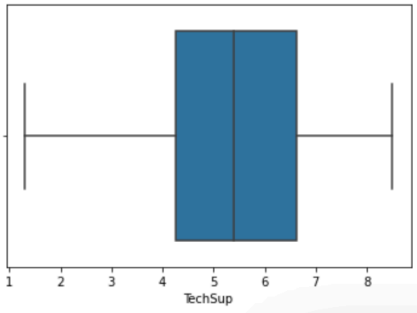

```
BoxPlot of Ecom
--------------------------------------------------------------------------------
```



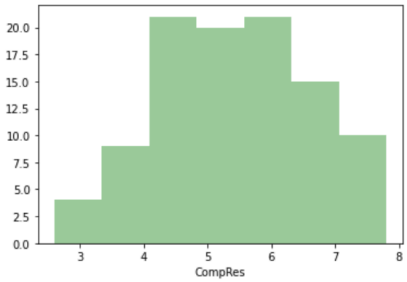

```
Description of TechSup
--------------------------------------------------------------------------------
count    100.000000
mean       5.365000
std        1.530457
min        1.300000
25%        4.250000
50%        5.400000
75%        6.625000
max        8.500000
Name: TechSup, dtype: float64
--------------------------------------------------------------------------------
Distribution of TechSup
--------------------------------------------------------------------------------
```



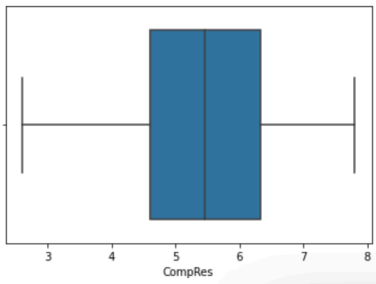

```
BoxPlot of TechSup
--------------------------------------------------------------------------------
```



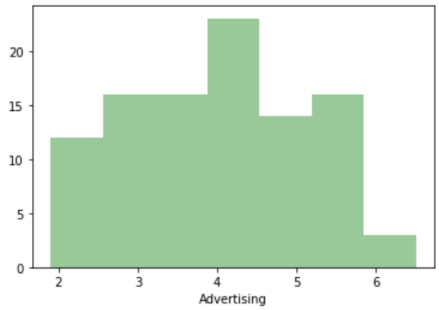
```
Description of CompRes
--------------------------------------------------------------------------------
count    100.000000
mean       5.442000
std        1.208403
min        2.600000
25%        4.600000
50%        5.450000
75%        6.325000
max        7.800000
Name: CompRes, dtype: float64
--------------------------------------------------------------------------------
Distribution of CompRes
--------------------------------------------------------------------------------
```



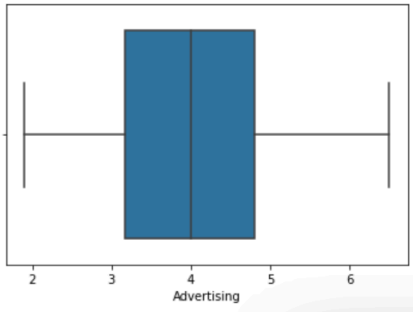

```
BoxPlot of CompRes
--------------------------------------------------------------------------------
```



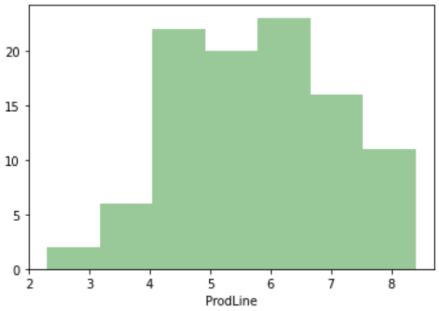

```
Description of Advertising
--------------------------------------------------------------------------------
count    100.000000
mean       4.010000
std        1.126943
min        1.900000
25%        3.175000
50%        4.000000
75%        4.800000
max        6.500000
Name: Advertising, dtype: float64
--------------------------------------------------------------------------------
Distribution of Advertising
--------------------------------------------------------------------------------
```



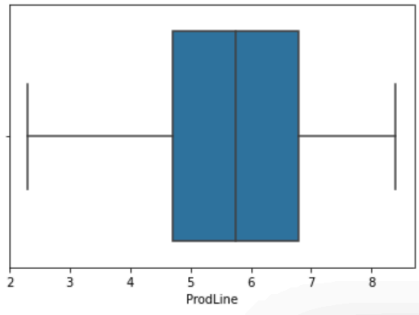

```
BoxPlot of Advertising
--------------------------------------------------------------------------------
```



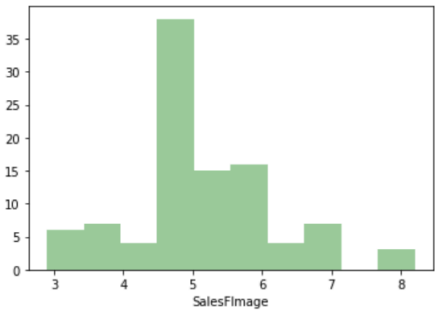
```
Description of ProdLine
--------------------------------------------------------------------------------
count    100.000000
mean       5.805000
std        1.315285
min        2.300000
25%        4.700000
50%        5.750000
75%        6.800000
max        8.400000
Name: ProdLine, dtype: float64
--------------------------------------------------------------------------------
Distribution of ProdLine
--------------------------------------------------------------------------------
```



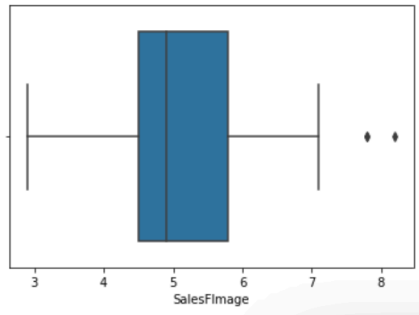

```
BoxPlot of ProdLine
--------------------------------------------------------------------------------
```



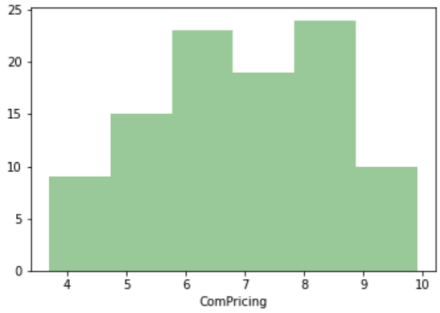

```
Description of SalesFImage
--------------------------------------------------------------------------------
count    100.00000
mean       5.12300
std        1.07232
min        2.90000
25%        4.50000
50%        4.90000
75%        5.80000
max        8.20000
Name: SalesFImage, dtype: float64
--------------------------------------------------------------------------------
Distribution of SalesFImage
--------------------------------------------------------------------------------
```



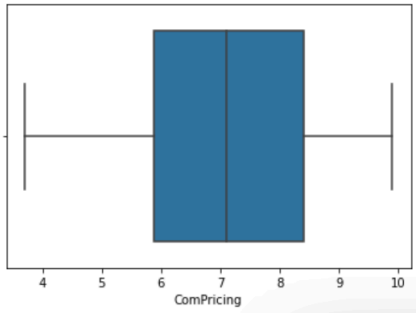

```
BoxPlot of SalesFImage
--------------------------------------------------------------------------------
```



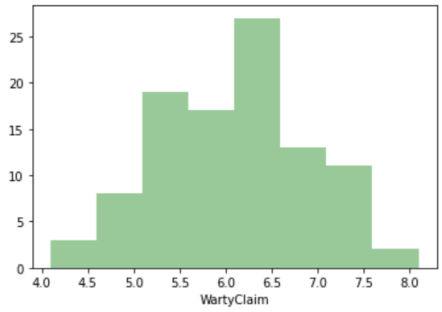
```
Description of ComPricing
--------------------------------------------------------------------------------
count    100.000000
mean       6.974000
std        1.545055
min        3.700000
25%        5.875000
50%        7.100000
75%        8.400000
max        9.900000
Name: ComPricing, dtype: float64
--------------------------------------------------------------------------------
Distribution of ComPricing
--------------------------------------------------------------------------------
```



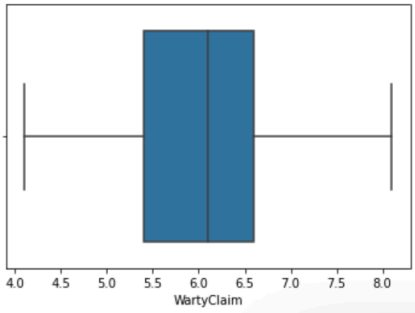

```
BoxPlot of ComPricing
--------------------------------------------------------------------------------
```



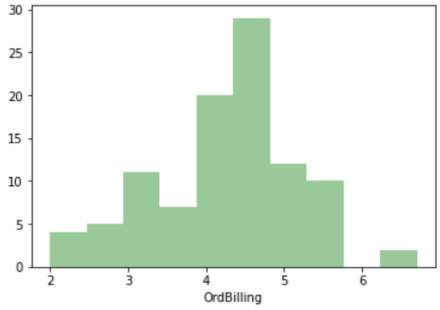

```
Description of WartyClaim
--------------------------------------------------------------------------------
count    100.000000
mean       6.043000
std        0.819738
min        4.100000
25%        5.400000
50%        6.100000
75%        6.600000
max        8.100000
Name: WartyClaim, dtype: float64
--------------------------------------------------------------------------------
Distribution of WartyClaim
--------------------------------------------------------------------------------
```



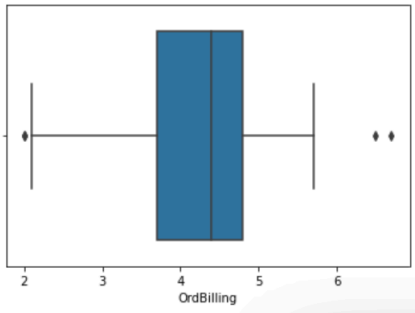

```
BoxPlot of WartyClaim
--------------------------------------------------------------------------------
```



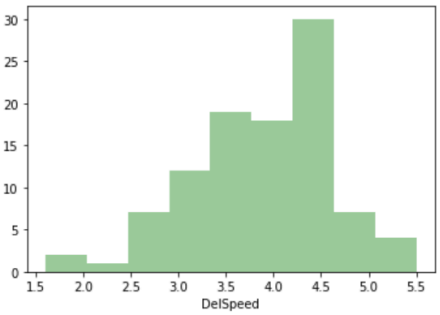
```
Description of OrdBilling
--------------------------------------------------------------------------------
count    100.00000
mean       4.27800
std        0.92884
min        2.00000
25%        3.70000
50%        4.40000
75%        4.80000
max        6.70000
Name: OrdBilling, dtype: float64
--------------------------------------------------------------------------------
Distribution of OrdBilling
--------------------------------------------------------------------------------
```



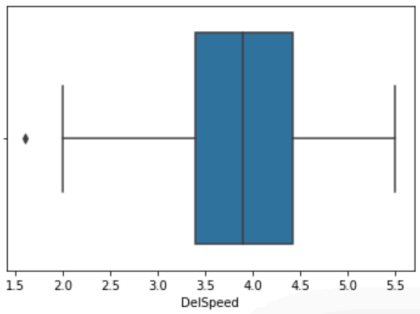

```
BoxPlot of OrdBilling
--------------------------------------------------------------------------------
```




```
Description of DelSpeed
--------------------------------------------------------------------------------
count    100.000000
mean       3.886000
std        0.734437
min        1.600000
25%        3.400000
50%        3.900000
75%        4.425000
max        5.500000
Name: DelSpeed, dtype: float64
--------------------------------------------------------------------------------
Distribution of DelSpeed
--------------------------------------------------------------------------------
```




```
BoxPlot of DelSpeed
--------------------------------------------------------------------------------
```



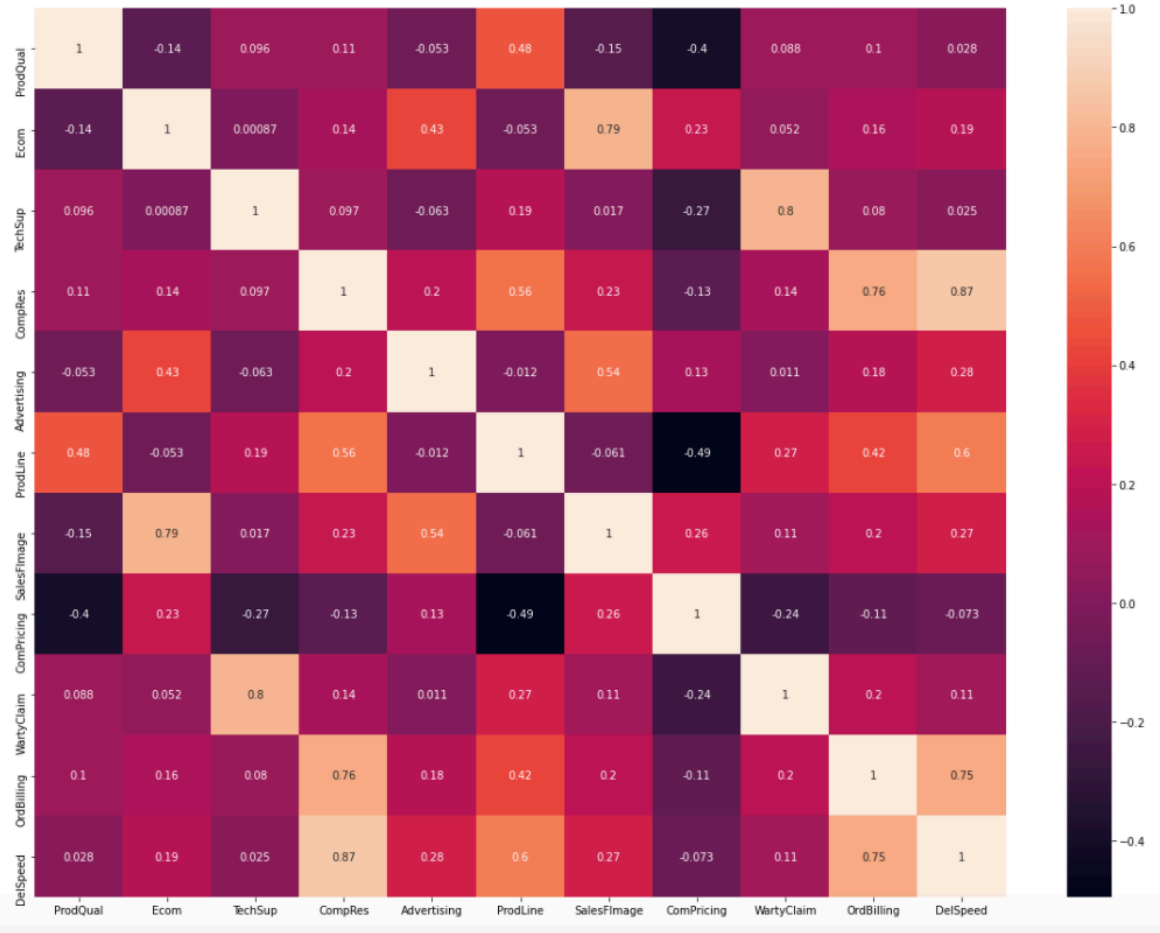
Multivariate Analysis :

Multivariate means involving multiple dependent variables resulting in one outcome. For example, we cannot predict the weather of any year based on the season. There are multiple factors like pollution, humidity, precipitation, etc.

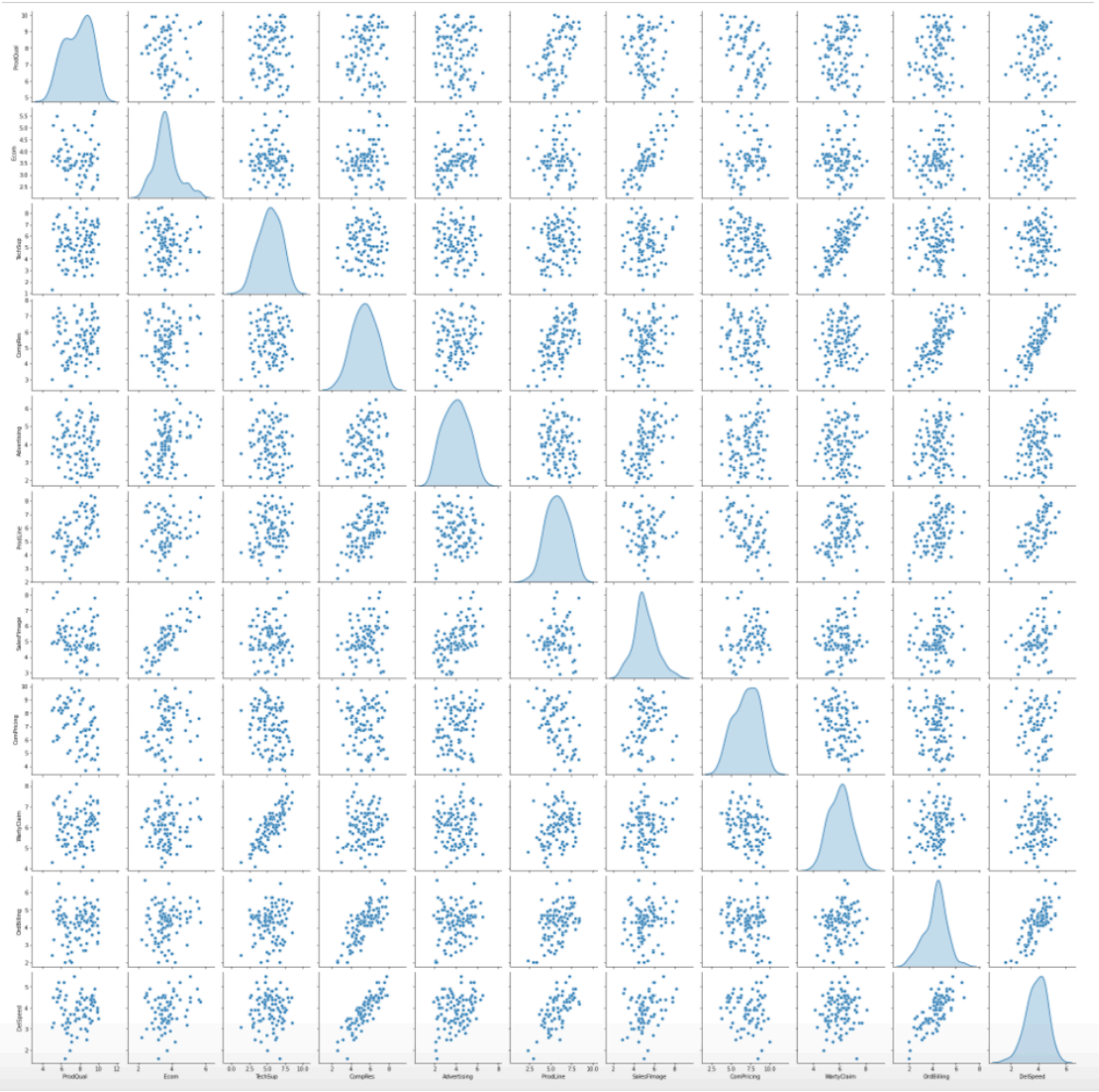
To derive **Multivariate Analysis** we have implemented **correlation heat map** and **pair plot**.

Below are the images for the same .

Correlation -Heat-map



Pair-plot



Inferences drawn from the above analysis:

1. Data consists of 100 different Hair salons information with 11 different market segmentation. (**Excluding "ID" & "Satisfaction"**)

2. There are no missing & duplicate values in the data set.

3. All Salons are identified by the unique ID which is mentioned by "**ID**" column.

4. All columns are **numerical** fields.

5. Almost all columns have some amount difference b/w **mean** and **median** values, which shows **data set has some amount of skewness**.

6. Data set **has few outliers**.

7. **High correlation** observed b/w **warranty claim & Technical Support** , as those comes under backend support , hence both are inter related .

8. **High correlation** observed b/w **Delivery Speed & Complaint Resolution**

9. **High correlation** observed b/w **Salesforce Image & E-commerce** , as those comes under marketing and foreground support of the business , hence both are inter related .

10. Potentially **better correlation** observed b/w **OrderBilling & Complaint Resolution** .

2) Scale the variables and write the inference for using the type of scaling function for this case study. - 3 points

Post scaling the dataset is as below :

	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed
0	0.496660	0.327114	-1.881421	0.380922	0.704543	-0.691530	0.821973	-0.113185	-1.646582	0.781230	-0.254531
1	0.280721	-1.394538	-0.174023	1.462141	-0.544014	1.600835	-1.896068	-1.088915	-0.665744	-0.409009	1.387605
2	1.000518	-0.390241	0.154322	0.131410	1.239639	1.218774	0.634522	-1.609304	0.192489	1.214044	0.840226
3	-1.014914	-0.533712	1.073690	-1.448834	0.615361	-0.844354	-0.583910	1.187789	1.173327	0.023805	-1.212443
4	0.856559	-0.390241	-0.108354	-0.700298	-1.614207	0.149004	-0.583910	-0.113185	0.069885	0.240212	-0.528220

Basically if all the columns in a table or dataset are not in a same magnitude then its really difficult to compare them and come to a conclusion , so **by scaling them we bring their magnitude into a certain range where we can compare them and come to conclusion , derive business implications.**

Inferences for using this scaling mechanism :

1. Z-scores normalisation are a way to compare results from a test to a "**normal**" population and bring them to a **same comparable scale or magnitude.**

2. **Z Score** can be used to compare raw scores that are taken from different tests .

3. Z score takes into account both the mean value and the variability in a set of raw scores.

4. Z Score always assume a normal distribution. Here in this dataset / case study we can see minimal skewness in terms of data distribution in few columns ,hence we used this.

5. Z score mechanism is simple and **easy to implement** .

6. Z score provides **better computation**. Computationally, it can **speed up the calculation** .

7. Z score provide **better interpretation** of the data.¶

3) Comment on the comparison between covariance and the correlation matrix after scaling. - 2 points

Without Scaling the correlation matrix :

	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed
ProdQual	1.000000	-0.137163	0.095600	0.106370	-0.053473	0.477493	-0.151813	-0.401282	0.088312	0.104303	0.027718
Ecom	-0.137163	1.000000	0.000867	0.140179	0.429891	-0.052688	0.791544	0.229462	0.051898	0.156147	0.191636
TechSup	0.095600	0.000867	1.000000	0.096657	-0.062870	0.192625	0.016991	-0.270787	0.797168	0.080102	0.025441
CompRes	0.106370	0.140179	0.096657	1.000000	0.196917	0.561417	0.229752	-0.127954	0.140408	0.756869	0.865092
Advertising	-0.053473	0.429891	-0.062870	0.196917	1.000000	-0.011551	0.542204	0.134217	0.010792	0.184236	0.275863
ProdLine	0.477493	-0.052688	0.192625	0.561417	-0.011551	1.000000	-0.061316	-0.494948	0.273078	0.424408	0.601850
SalesFImage	-0.151813	0.791544	0.016991	0.229752	0.542204	-0.061316	1.000000	0.264597	0.107455	0.195127	0.271551
ComPricing	-0.401282	0.229462	-0.270787	-0.127954	0.134217	-0.494948	0.264597	1.000000	-0.244986	-0.114567	-0.072872
WartyClaim	0.088312	0.051898	0.797168	0.140408	0.010792	0.273078	0.107455	-0.244986	1.000000	0.197065	0.109395
OrdBilling	0.104303	0.156147	0.080102	0.756869	0.184236	0.424408	0.195127	-0.114567	0.197065	1.000000	0.751003
DelSpeed	0.027718	0.191636	0.025441	0.865092	0.275863	0.601850	0.271551	-0.072872	0.109395	0.751003	1.000000

With Scaling the correlation matrix :

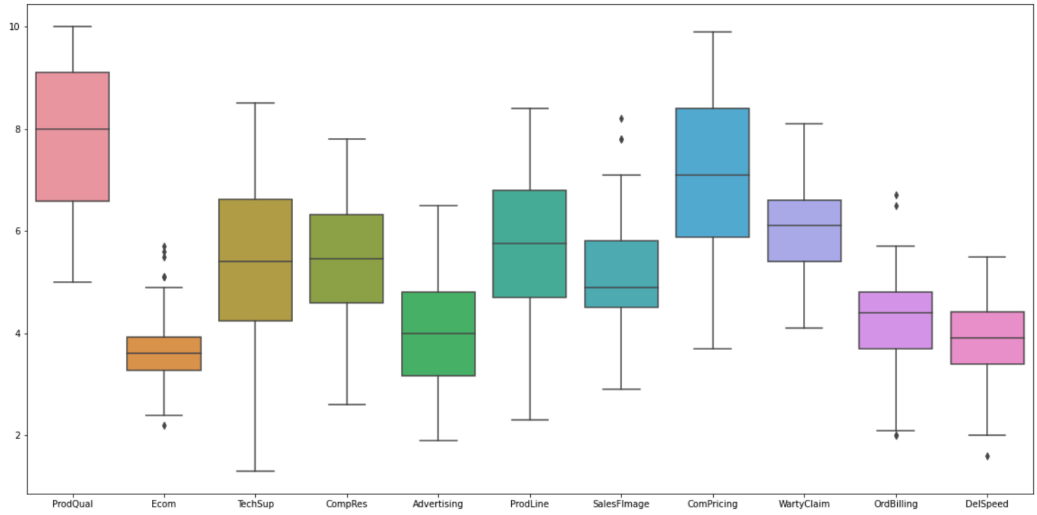
	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed
ProdQual	1.000000	-0.137163	0.095600	0.106370	-0.053473	0.477493	-0.151813	-0.401282	0.088312	0.104303	0.027718
Ecom	-0.137163	1.000000	0.000867	0.140179	0.429891	-0.052688	0.791544	0.229462	0.051898	0.156147	0.191636
TechSup	0.095600	0.000867	1.000000	0.096657	-0.062870	0.192625	0.016991	-0.270787	0.797168	0.080102	0.025441
CompRes	0.106370	0.140179	0.096657	1.000000	0.196917	0.561417	0.229752	-0.127954	0.140408	0.756869	0.865092
Advertising	-0.053473	0.429891	-0.062870	0.196917	1.000000	-0.011551	0.542204	0.134217	0.010792	0.184236	0.275863
ProdLine	0.477493	-0.052688	0.192625	0.561417	-0.011551	1.000000	-0.061316	-0.494948	0.273078	0.424408	0.601850
SalesFImage	-0.151813	0.791544	0.016991	0.229752	0.542204	-0.061316	1.000000	0.264597	0.107455	0.195127	0.271551
ComPricing	-0.401282	0.229462	-0.270787	-0.127954	0.134217	-0.494948	0.264597	1.000000	-0.244986	-0.114567	-0.072872
WartyClaim	0.088312	0.051898	0.797168	0.140408	0.010792	0.273078	0.107455	-0.244986	1.000000	0.197065	0.109395
OrdBilling	0.104303	0.156147	0.080102	0.756869	0.184236	0.424408	0.195127	-0.114567	0.197065	1.000000	0.751003
DelSpeed	0.027718	0.191636	0.025441	0.865092	0.275863	0.601850	0.271551	-0.072872	0.109395	0.751003	1.000000

Covariance tells us the direction of the linear relationship between two variables. **Correlation** indicates the measures of both the strength and direction of the linear relationship between two variables. Correlation is function of the covariance. We can find the correlation coefficient of two variables by dividing the covariance of these variables by the product of the standard deviations of the same values. Here before & after Scaling correlation matrix gives same result.

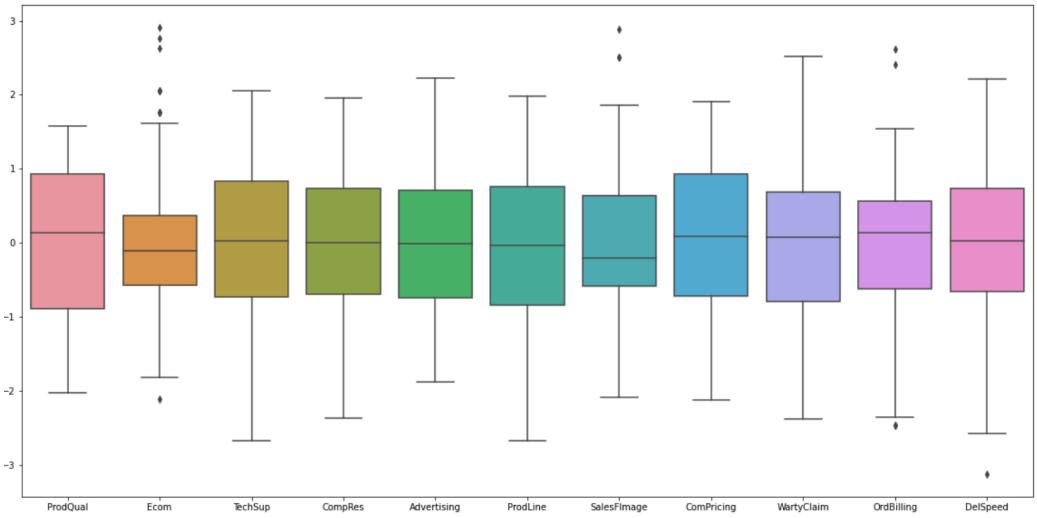
So in one word we can conclude saying **correlation matrix is just a scaled derivative of the covariance matrix**.¶

4) Check the dataset for outliers before and after scaling. Draw your inferences from this exercise. - 3 points

Checking outliers **before scaling** :



Checking outliers **after scaling** :



Inferences drawn from above analysis:

1. By scaling, all variables are having same standard deviation, So all the variables have the same weight/magnitude and the same resulting in PCA ,by which the **dataset has a better interpretation and easy to compare and derive business implications and conclusions** .

2. Post scaling the data is **more normally distributed** .

5) Build the covariance matrix, eigenvalues and eigenvector. - 4 points

Covariance Matrix :

```
Covariance Matrix
%s [[ 1.01010101e+00 -1.38548704e-01  9.65661154e-02  1.07444445e-01
  -5.40132667e-02  4.82316579e-01 -1.53346338e-01 -4.05335236e-01
   8.92043497e-02  1.05356640e-01  2.79979825e-02]
 [-1.38548704e-01  1.01010101e+00  8.75544162e-04  1.41595213e-01
   4.34233041e-01 -5.32200387e-02  7.99539102e-01  2.31780203e-01
   5.24224157e-02  1.57724577e-01  1.93571786e-01]
 [ 9.65661154e-02  8.75544162e-04  1.01010101e+00  9.76329270e-02
  -6.35051180e-02  1.94571168e-01  1.71621612e-02 -2.73521901e-01
   8.05220127e-01  8.09109340e-02  2.56976702e-02]
 [ 1.07444445e-01  1.41595213e-01  9.76329270e-02  1.01010101e+00
   1.98905906e-01  5.67087831e-01  2.32072486e-01 -1.29246720e-01
   1.41826562e-01  7.64513729e-01  8.73829997e-01]
 [-5.40132667e-02  4.34233041e-01 -6.35051180e-02  1.98905906e-01
   1.01010101e+00 -1.16674936e-02  5.47680463e-01  1.35572620e-01
   1.09010852e-02  1.86096560e-01  2.78649579e-01]
 [ 4.82316579e-01 -5.32200387e-02  1.94571168e-01  5.67087831e-01
  -1.16674936e-02  1.01010101e+00 -6.19348764e-02 -4.99947880e-01
   2.75835887e-01  4.28695202e-01  6.07929503e-01]
 [-1.53346338e-01  7.99539102e-01  1.71621612e-02  2.32072486e-01
   5.47680463e-01 -6.19348764e-02  1.01010101e+00  2.67269246e-01
   1.08540752e-01  1.97098390e-01  2.74294201e-01]
 [-4.05335236e-01  2.31780203e-01 -2.73521901e-01 -1.29246720e-01
   1.35572620e-01 -4.99947880e-01  2.67269246e-01  1.01010101e+00
  -2.47460661e-01 -1.15724268e-01 -7.36078070e-02]
 [ 8.92043497e-02  5.24224157e-02  8.05220127e-01  1.41826562e-01
   1.09010852e-02  2.75835887e-01  1.08540752e-01 -2.47460661e-01
   1.01010101e+00  1.99055678e-01  1.10499598e-01]
 [ 1.05356640e-01  1.57724577e-01  8.09109340e-02  7.64513729e-01
   1.86096560e-01  4.28695202e-01  1.97098390e-01 -1.15724268e-01
   1.99055678e-01  1.01010101e+00  7.58588957e-01]
 [ 2.79979825e-02  1.93571786e-01  2.56976702e-02  8.73829997e-01
   2.78649579e-01  6.07929503e-01  2.74294201e-01 -7.36078070e-02
   1.10499598e-01  7.58588957e-01  1.01010101e+00]]
```

Eigen Vectors :

```
Eigen Vectors :

 [[ 0.13378962 -0.31349802  0.06227164  0.6431362  -0.2316662  -0.56456996
   0.19164132  0.18279209  0.06659717 -0.13547311  0.0313281 ]
 [ 0.16595278  0.44650918 -0.23524791  0.27238033 -0.42228844  0.26325703
   0.05962621  0.06233863  0.28155772  0.12202642 -0.54251104]
 [ 0.15769263 -0.23096734 -0.61095105 -0.19339314  0.02395667 -0.10876896
  -0.01719992 -0.05192956 -0.3881709  -0.46470964 -0.35929961]
 [ 0.47068359  0.01944394  0.21035078 -0.20632037 -0.02865743 -0.02815231
  -0.0084996  -0.36253352  0.53467243 -0.51339754  0.09324751]
 [ 0.18373495  0.36366471 -0.08809705  0.31789448  0.80387024 -0.20056937
  -0.06306962 -0.08118684  0.03715799  0.05347713 -0.15468169]
 [ 0.38676517 -0.28478056  0.11627864  0.20290226 -0.11667416  0.09819533
  -0.60814755 -0.38507778 -0.23479794  0.3332071  -0.08415534]
 [ 0.2036696   0.47069599 -0.2413421   0.22217722 -0.20437283  0.10497225
   0.00143735 -0.08469869 -0.35341191 -0.16910665  0.64489911]
 [-0.15168864  0.4134565   0.05304529 -0.33354348 -0.24892601 -0.70973595
  -0.30824887 -0.10295751 -0.04518224  0.09883227 -0.09414389]
 [ 0.21293363 -0.19167191 -0.59856398 -0.18530205  0.03292706 -0.13983966
  -0.03064024  0.12893245  0.43534752  0.4435404   0.31756604]
 [ 0.43721774  0.02639905  0.16892981 -0.23685365 -0.02675377 -0.11947974
   0.65931989 -0.19415064 -0.30386545  0.36601754 -0.09907265]
 [ 0.47308914  0.07305172  0.23262477 -0.1973299   0.03543294  0.02979992
  -0.23423927  0.77563222 -0.12010386 -0.06539059 -0.02188514]]
```

Eigen Values :

```
Eigen Values :

[3.4615872  2.57666335 1.70805705 1.09753137 0.61557989 0.55745836
0.40557389 0.09942123 0.13418341 0.249446   0.20560936]
```

6) Write the explicit form of the first PC (in terms of Eigen Vectors) – 5 points

In terms of Eigen Vectors derived First PC is : **eig_vecs[0]**

Derived value of **first PC** as below :

```
array([ 0.13378962, -0.31349802,  0.06227164,  0.6431362 , -0.2316662 ,
       -0.56456996,  0.19164132,  0.18279209,  0.06659717, -0.13547311,
        0.0313281 ])
```

7) Discuss the cumulative values of the eigenvalues. How does it help you to decide on the optimum number of principal components? What do the eigenvectors indicate? Perform PCA and export the data of the Principal Component scores into a data frame. – 10 points

For **cumulative eigenvalues**, we calculated the **cumulative sum** of **eigenvalues** such that the total **sum** is 100% .

cumulative eigenvalues :

```
Cumulative Variance : [ 31.1542848   54.34425491  69.71676832  79.59455066  85.1347697
  90.15189496  93.80205993  96.04707397  97.89755822  99.10520892
 100.        ]
```

From the above cumulative values of the eigenvalues, we can see that around **6 principal components** explained over **90% (90.15%)** of the variance. Thus, the **optimum number of principal components can be 6**

Eigen Vectors indicate the direction of the principal components, we can multiply the original data by the eigenvectors to re-organise/re-orient the data onto the new axis/dimension.

Performing **PCA** :

```
# NOTE - we are generating only 6 PCA dimensions as its
covering over ~90% data
```

```
array([[ 7.95507962e-02, -1.10096634e+00, -2.19706653e+00,
         1.56293289e+00,  7.67570385e-01,  2.90862177e+00,
         5.29319132e+00,  1.47659077e+00, -6.13947615e-01,
        -4.23660076e-01,  5.76252311e-01,  1.86757037e+00,
        -2.66029481e+00, -1.15437973e+00, -1.98252867e+00,
        -1.19534642e+00, -6.29210603e-01, -1.94912563e+00,
        -4.40654330e-01, -1.18679105e+00,  1.32903312e+00,
        -3.07501457e+00, -1.22862294e+00, -1.95215660e+00,
         2.77114349e-01, -3.38635080e-01, -1.76471720e+00,
         1.03510170e+00, -1.29479142e+00,  1.66262708e+00,
         1.44608491e-01,  2.11462740e+00,  1.06970726e+00,
        -1.09408090e-01,  1.80536022e+00,  2.12248520e+00,
         6.22051111e-01, -2.57411754e+00, -3.44966983e-01,
         1.21539637e+00,  1.25487586e+00,  1.67930781e-01,
        -1.82095895e+00, -2.33269255e+00, -8.73288881e-01,
        -1.65347263e+00, -1.03184895e+00, -3.12115015e+00,
        -1.07236222e+00, -7.97675229e-01,  1.98945764e+00,
```

```
 -1.15121892e+00, -1.07789784e+00,  1.16797753e+00,
  5.28522663e-01, -3.76893195e-01, -3.50943905e+00,
 -1.28569623e+00,  5.74368371e-01, -1.10762365e+00,
 -2.11898365e+00,  1.19627748e+00,  1.22409809e+00,
  2.78237859e+00,  1.83048999e+00, -1.25224924e+00,
 -1.39907113e+00,  9.18897030e-02,  2.55191074e+00,
  1.05553043e+00, -2.17896719e+00,  2.17537864e+00,
 -1.01608393e+00, -1.29736966e+00, -6.38034682e-02,
 -1.30860701e-01, -2.06993579e-01, -1.25207434e+00,
 -3.21564763e+00,  2.42613744e+00,  4.32888801e-01,
  3.27874823e-02,  2.36190826e+00,  4.37978758e+00,
 -9.28110037e-01,  2.36712173e+00,  4.23846478e+00,
 -1.65771553e+00, -3.02524564e-01, -3.53671209e+00,
 -6.44755632e-02,  4.82862408e+00, -1.07928030e+00,
 -2.81182077e+00,  8.07527247e-01, -2.73668121e-01,
  1.60818602e+00,  3.19577568e+00, -6.20888188e-01,
  1.63523181e+00],
[ 1.54319843e+00, -2.42029823e+00, -7.27440443e-01,
  1.71366473e-01, -1.42811141e+00,  3.09386635e-01,
  1.05748117e+00,  1.11108334e+00,  1.37947271e+00,
  1.98154125e+00, -1.26161660e+00,  2.39193283e+00,
  1.86358088e+00, -1.52960009e+00,  1.92826246e+00,
 -2.06387739e+00,  1.53350405e-02,  1.06318630e+00,
  1.31777080e+00,  2.42132796e+00,  6.90057800e-01,
  1.06660984e+00, -2.22454236e+00, -2.03020668e+00,
  8.95044261e-01,  6.44411179e-01, -2.55342530e+00,
 -2.86166486e-01, -2.63654169e+00,  1.24602693e+00,
 -1.56163320e+00, -2.49819204e+00,  1.17585682e+00,
  7.11613293e-01,  1.21832759e+00, -2.39858791e+00,
 -1.07056775e+00, -5.22825851e-01,  1.63997306e+00,
  3.34138683e+00,  1.06936086e+00,  2.00993138e-01,
  2.03011293e+00,  2.27659170e+00, -2.32142943e+00,
  1.69649585e+00, -1.91387396e+00,  1.72528176e+00,
  3.01340467e-01,  1.02765812e-01,  2.81247685e-02,
 -2.36760950e+00, -1.99023980e+00, -1.23330313e+00,
  4.62843812e-01, -6.77235220e-01,  2.59315555e+00,
 -3.29110097e-01, -1.29841702e+00, -2.86825408e+00,
 -3.09937222e+00,  5.55633856e-01,  1.80791202e-01,
 -2.47159379e+00,  3.48448035e-01,  2.34615380e+00,
 -5.19948268e-01,  6.34828142e-01,  2.35346818e-01,
  1.37636158e+00,  2.07859219e+00, -2.98430763e-01,
  3.06798825e+00, -1.60530098e-01,  9.24769356e-01,
 -1.73470304e+00, -2.80686494e-01, -4.01311216e-01,
 -7.67806275e-01,  5.91771751e-01, -1.95114344e+00,
 -1.73352795e+00, -1.90286414e+00, -4.31082542e-01,
 -2.53944144e+00,  4.30472714e-01,  1.85761911e+00,
 -2.29865293e+00, -3.53607099e-01,  2.89542887e+00,
 -7.11686170e-01,  9.96088555e-01, -6.67563071e-01,
 -8.84384584e-01,  4.21043085e-01,  1.36511701e-01,
```

```
  9.25498892e-01, -1.86292048e+00,  2.34984871e+00,
  1.25986453e+00],
[ 1.89504643e+00,  2.04552113e+00,  1.66800258e-01,
 -1.82717898e+00,  2.34356001e-01,  1.53270624e+00,
 -6.44860661e-01,  7.05904921e-01,  5.75067006e-01,
  3.36887578e-01,  8.24678612e-01, -1.85612007e+00,
 -3.27460628e-01, -4.60492743e-02,  6.57158467e-01,
  1.61067072e+00, -9.37041129e-01,  5.15860773e-01,
  1.04646171e+00,  4.92412433e-01, -2.08768076e+00,
 -1.80445175e+00, -1.01140869e+00,  2.66954855e+00,
 -5.60705768e-01, -4.19165019e-01, -2.95052784e-01,
 -5.03929353e-03,  1.36353767e+00,  1.41964173e+00,
 -2.86593652e+00, -1.20557205e+00, -6.21359653e-01,
  9.74658028e-02, -1.21388678e+00, -4.72136843e-01,
 -5.09886027e-01,  5.70675516e-01,  1.53009034e+00,
  3.58208421e-01,  2.46199587e+00,  1.34326736e+00,
  6.90472951e-01, -2.09300371e+00,  2.53572770e+00,
 -2.10688283e-01, -1.05020890e+00, -2.35960507e-01,
  7.45863156e-01,  9.69951636e-01, -5.71589670e-01,
  2.95364353e+00,  1.21597286e+00, -7.68149012e-01,
  1.69531843e+00, -1.29697927e+00,  4.07203668e-01,
 -4.49043076e-01, -1.12267032e+00,  4.92719345e-01,
 -8.51862928e-01, -6.39030093e-01,  9.20911431e-01,
 -2.13284504e-01, -1.34978559e+00,  1.52325302e+00,
 -1.74223193e+00,  5.93115811e-01,  9.17044242e-01,
  1.50940777e+00, -1.46616460e+00,  9.24921430e-01,
  1.98239226e+00, -2.16078876e+00, -1.57565406e-01,
 -1.59701287e+00, -2.61240168e+00, -1.23498251e+00,
 -1.09824841e-01, -9.32331700e-01, -7.17038785e-01,
 -1.31967348e-01, -4.60882225e-01, -2.86721735e+00,
  3.36241959e-01,  9.61592950e-01,  1.77787733e+00,
 -5.13970785e-01, -6.72268487e-02, -2.68798142e+00,
 -2.11394201e+00, -5.21283494e-01,  5.67521502e-01,
  4.30619794e-01,  6.18622940e-01, -2.51745365e-01,
  2.11327746e+00,  1.29747695e+00, -1.12785588e+00,
  4.02550692e-01],
[ 1.16811885e+00, -4.27083409e-01,  1.31031245e+00,
 -1.19224019e+00,  6.95254671e-02, -7.46605491e-01,
  2.84698493e-02, -5.67127307e-01, -1.76926421e+00,
 -3.65410091e-01,  2.93518604e-01,  2.17253308e-01,
  1.55164811e+00,  2.40411597e-01, -9.61216138e-01,
  2.37646120e-01, -1.49440592e+00, -1.96589252e+00,
 -1.93096267e+00,  1.22974396e+00, -1.76782014e+00,
  3.03690733e+00,  5.33451351e-01, -8.84126573e-01,
 -1.18645614e+00, -1.59482832e+00, -3.53908543e-01,
 -1.78131550e+00, -5.60646711e-01,  4.71961213e-01,
 -2.71835072e-01, -7.88805195e-01, -4.87116117e-01,
 -5.82455639e-01,  1.14812235e-01,  7.29745518e-01,
  5.48201090e-01, -5.19362238e-02,  2.79337598e-02,
```

```
 4.54974686e-01, -1.62084987e-01,  7.95937008e-01,
 1.91525424e+00, -8.53346292e-01, -4.60931712e-01,
-1.03344488e+00,  8.43015769e-01, -2.13351288e+00,
 1.65409894e+00,  1.24462436e+00, -4.14930214e-01,
-5.90391007e-01, -8.29745885e-02,  1.49883673e-01,
-8.74525437e-01,  4.90518487e-01, -6.73362957e-01,
 7.61785884e-01,  2.21949850e+00, -6.58726600e-01,
-8.46256685e-01,  1.11418112e+00,  1.37513768e+00,
-7.94207060e-01, -5.97121560e-02, -9.28249420e-01,
-2.27232735e+00, -8.65995328e-01, -1.06592613e-01,
-5.55476126e-01,  7.55549358e-01,  1.25717946e+00,
 3.99327442e-01,  1.61581524e+00,  8.64441940e-01,
 1.42959047e+00, -9.01959791e-01,  2.52049136e-02,
 9.39947493e-01,  4.46666440e-01,  1.28323707e+00,
-2.79029152e-01,  4.70011384e-01, -1.07365103e+00,
-1.48214626e+00,  5.30217110e-02, -9.99462453e-03,
-1.40418390e+00,  8.74763815e-02, -5.92713043e-01,
 6.00982836e-01, -6.00064284e-02,  1.06396606e+00,
 9.12722813e-01,  1.80536987e+00,  4.99108366e-01,
 2.43062360e-01,  1.41060767e+00,  9.14580433e-01,
-4.25025076e-03],
[-1.13909362e-01, -5.50452952e-01, -1.06179705e+00,
 -9.39628800e-01,  1.20649782e+00, -8.48740962e-01,
  1.30354697e+00, -1.08644238e+00,  3.67066922e-01,
  1.19661729e-01,  7.49130007e-01,  4.07158227e-01,
  1.24928384e+00,  1.67482738e+00,  4.30128765e-01,
  5.41609431e-02,  1.32701348e-01,  6.92885395e-01,
  1.59271545e-01,  3.98434161e-01, -2.81279852e-01,
  8.19774364e-01, -8.12451759e-02,  7.66453478e-01,
  4.49703762e-01, -3.55560498e-01, -8.71437733e-01,
  7.04092382e-01,  5.03357452e-01, -9.05877847e-01,
 -1.11480830e+00,  1.11461013e-01, -8.44427386e-01,
 -1.17352371e+00,  1.18048379e+00, -4.01908101e-01,
 -5.27470360e-02,  1.71865607e+00, -9.80460239e-01,
  5.03681702e-02,  9.41824447e-01,  7.99653831e-01,
  2.67385962e-01,  3.93287868e-01, -8.60201815e-01,
 -1.01269434e+00,  8.78305462e-01, -1.12413522e+00,
  6.63098788e-01, -2.33270429e-01,  1.05247276e-01,
  3.62124736e-03,  5.15267914e-02, -1.70072284e-01,
  5.43955483e-01,  7.14128211e-01,  1.34667654e+00,
 -1.52298647e+00, -1.24966714e-01,  2.97138281e-01,
  2.42456790e-01, -1.94681755e-01, -1.45249408e-01,
 -3.11205319e-01, -1.03437819e+00, -8.79876136e-01,
  5.44367182e-01, -8.07893008e-01,  5.01814684e-01,
 -2.31302099e-01, -3.96154025e-01, -1.40432478e+00,
 -1.03303487e+00, -4.97790633e-01, -3.68615047e-01,
 -1.55815146e+00, -7.67844815e-01,  4.62172385e-02,
 -1.00209423e+00, -8.32459298e-02, -5.09266350e-01,
  9.32558739e-01,  6.86424644e-02,  6.54124241e-01,
```

```
 6.17969596e-01,  8.14867091e-02, -7.40017572e-02,
-5.51728254e-01,  1.72089032e-01, -1.48491704e-02,
 1.78743472e-02,  1.68211961e+00,  4.14212520e-01,
-3.41452026e-01,  6.09200494e-02,  2.15921163e+00,
-3.46204929e-01, -5.29038082e-02,  3.64281590e-01,
-1.49647947e+00],
[ 8.63641435e-02,  4.75587801e-01,  2.44818551e-01,
-9.57207069e-01, -2.51605701e-01, -3.34014092e-01,
 1.07495110e-01,  4.94166065e-01, -1.27711327e-01,
 3.10581091e-01, -2.43084577e-02,  6.53872522e-01,
-3.40346787e-01, -2.39824677e-01,  1.53633645e-01,
 6.60101313e-01,  1.25458391e+00,  2.28158515e-01,
-5.32221077e-02, -8.06406045e-01,  1.48346036e-01,
 1.22229995e+00,  1.01224524e+00, -5.54412011e-01,
 3.88021152e-01,  8.01580850e-02,  5.05000025e-02,
-7.53337245e-01,  6.11812740e-01, -3.15526396e-01,
 5.71005926e-01, -3.74744335e-01, -8.09793988e-01,
-1.48889174e+00,  4.66976006e-01,  4.92806850e-02,
-5.79555603e-01, -7.13870551e-01, -1.31931355e+00,
 6.69392521e-01, -7.65231141e-01, -9.58345929e-01,
-2.47775603e-01,  1.92360449e+00,  5.06495067e-01,
-1.36745760e-01,  5.53708470e-01, -3.63965567e-01,
-4.47677381e-01,  1.43051056e+00,  3.55294953e-01,
 6.52836007e-01, -6.16577698e-01,  1.27083537e+00,
-8.38527623e-01, -8.53198731e-01, -5.06274860e-01,
 9.02769642e-01,  6.78686529e-01, -7.18234187e-01,
 1.02954562e-01,  1.09982128e-01,  3.01877838e-01,
-8.17399953e-02,  6.45876657e-02,  3.82585645e-01,
 1.02624373e+00,  1.39988634e+00,  1.16381930e-01,
 6.66003031e-01, -1.49137932e+00,  1.51902890e+00,
 3.17152689e-02, -1.10791707e+00, -2.33923327e-01,
-5.77786763e-01, -8.85195284e-01, -1.73968329e+00,
 5.97638285e-02,  4.57163707e-02, -1.38436121e-01,
-2.51923409e-01, -6.56344509e-02, -2.36400865e-01,
 2.98272814e-01,  7.60740852e-02,  1.35467340e+00,
-2.03414153e-01, -1.34747166e+00,  1.39822134e+00,
-1.04435498e+00, -2.86609327e-01, -3.79889880e-01,
 3.66354185e-01, -4.83079004e-01,  5.44903381e-01,
 2.81136277e-01,  5.35178314e-01, -3.85227261e-02,
-1.80567536e+00]])
```

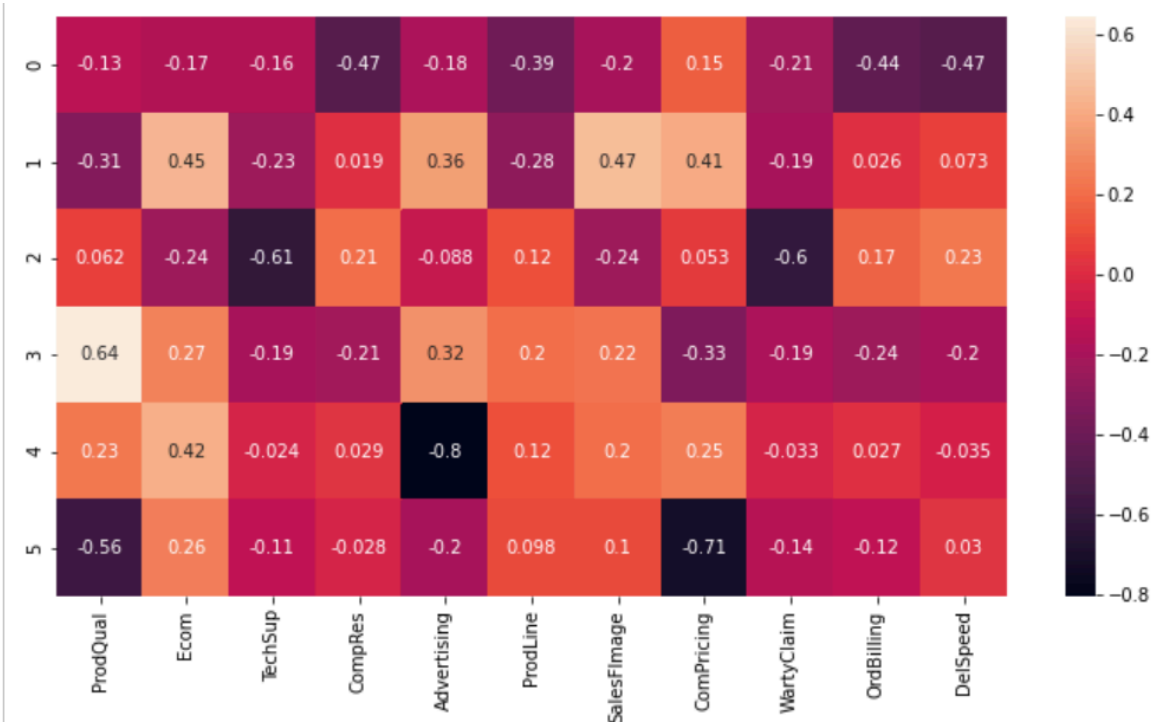
Generating **PCA components** :

```
array([[-0.13378962, -0.16595278, -0.15769263, -0.47068359, -0.18373495,
        -0.38676517, -0.2036696 ,  0.15168864, -0.21293363, -0.43721774,
        -0.47308914],
       [-0.31349802,  0.44650918, -0.23096734,  0.01944394,  0.36366471,
        -0.28478056,  0.47069599,  0.4134565 , -0.19167191,  0.02639905,
         0.07305172],
       [ 0.06227164, -0.23524791, -0.61095105,  0.21035078, -0.08809705,
         0.11627864, -0.2413421 ,  0.05304529, -0.59856398,  0.16892981,
         0.23262477],
       [ 0.6431362 ,  0.27238033, -0.19339314, -0.20632037,  0.31789448,
         0.20290226,  0.22217722, -0.33354348, -0.18530205, -0.23685365,
        -0.1973299 ],
       [ 0.2316662 ,  0.42228844, -0.02395667,  0.02865743, -0.80387024,
         0.11667416,  0.20437283,  0.24892601, -0.03292706,  0.02675377,
        -0.03543294],
       [-0.56456996,  0.26325703, -0.10876896, -0.02815231, -0.20056937,
         0.09819533,  0.10497225, -0.70973595, -0.13983966, -0.11947974,
         0.02979992]])
```

Loading **PCA components** into a **Data-frame :**

	ProdQual	Ecom	TechSup	CompRes	Advertising	ProdLine	SalesFImage	ComPricing	WartyClaim	OrdBilling	DelSpeed
0	-0.133790	-0.165953	-0.157693	-0.470684	-0.183735	-0.386765	-0.203670	0.151689	-0.212934	-0.437218	-0.473089
1	-0.313498	0.446509	-0.230967	0.019444	0.363665	-0.284781	0.470696	0.413457	-0.191672	0.026399	0.073052
2	0.062272	-0.235248	-0.610951	0.210351	-0.088097	0.116279	-0.241342	0.053045	-0.598564	0.168930	0.232625
3	0.643136	0.272380	-0.193393	-0.206320	0.317894	0.202902	0.222177	-0.333543	-0.185302	-0.236854	-0.197330
4	0.231666	0.422288	-0.023957	0.028657	-0.803870	0.116674	0.204373	0.248926	-0.032927	0.026754	-0.035433

Putting the **PCA data** into a **heat-map** for better analysis :



8) Mention the business implication of using the Principal Component Analysis for this case study. – 5 points

PC1- Explains highlights about ComplaintsResolution ,OrderBilling , Delivery Speed & Satisfaction are big concern here , need to focus .

PC2- Explains great growth over E-commerce ,SalesforceImage & competitive Pricing .

PC3- Explains that Technical Support & WarrantyClaim have very big concern , need to improve the same .

PC4- Explains that Product quality has a huge improvement. Though in this PC backend support system need to be improved especially OrderBilling , Complaint resolution & Delivery Speed.

PC5- We can see great growth over the marketing strategies like Advertisement but need to focus on the competitive pricing to make the business more impactful.

PC6- Highlights severe drop in E-commerce , need to focus on the same , on the other side we can see great growth over competitive pricing . Advertising also has a fair amount of growth but it has dropped a bit from previous one .

PC7- Explains about severe concern on Product line , need to focus on the same to make the business more successful , on the contrary there is a huge growth in the backend support in terms of OrderBilling which is been a concern for a long time , that is been addressed here .

THE END