Problem 1: Hay Fever Case Problem

24.01.2021

Peeyush Vardhan
Ph: +91-9515456727
Email: peeyushvrdhn@gmail.com

Problem Statement

A research laboratory was developing a new compound for the relief of severe cases of hay fever. In an experiment with 36 volunteers, the amounts of the two active ingredients (A & B) in the compound were varied at three levels each. Randomization was used in assigning four volunteers to each of the nine treatments. The data on hours of relief can be found in the following .csv file: Fever.csv

Problem 1.1

- State the Null and Alternate Hypothesis for conducting one-way ANOVA for both the variables 'A' and 'B' individually.

Solution:

The hypothesis of the One-way ANOVA of Relief' variable with the ingredient 'A' variable

- H0: The means of 'Relief' variable with respect to each amount of ingredient A is equal
- H1: At least one of the means of the 'Relief' variable with respect to each amount of ingredient A is unequal.

The hypothesis of the One-way ANOVA of Relief' variable with the ingredient 'B' variable

- H0: The means of 'Relief' variable with respect to each amount of ingredient B is equal
- H1: At least one of the means of the 'Relief' variable with respect to each amount of ingredient B is unequal.

Problem 1.2

- Perform one-way ANOVA for variable 'A' with respect to the variable 'Relief'. State whether the Null Hypothesis is accepted or rejected based on the ANOVA results.

Solution:

- **Formula = 'Relief ~ C(A)'**

	df	sum_sq	mean_sq	F	PR(>F)
C(A)	2.0	220.02	110.010000	23.465387	4.578242e-07
Residual	33.0	154.71	4.688182	NaN	NaN

Insight: p-value is smaller than choosen alpha level α = 0.05, so null hypothesis can be rejected.

Problem 1.3

- Perform one-way ANOVA for variable 'B' with respect to the variable 'Relief'. State whether the Null Hypothesis is accepted or rejected based on the ANOVA results.

Solution:

- **Formula = 'Relief ~ C(B)'**

	df	sum_sq	mean_sq	F	PR(>F)
C(B)	2.0	123.66	61.830000	8.126777	0.00135
Residual	33.0	251.07	7.608182	NaN	NaN

Insight: p-value is greater than chosen alpha level α = 0.05, so the null hypothesis can not be rejected. Hence, we accept the Null Hypothesis.

Problem 1.4

- Analyze the effects of one variable on another with the help of an interaction plot. What is an interaction between two treatments?

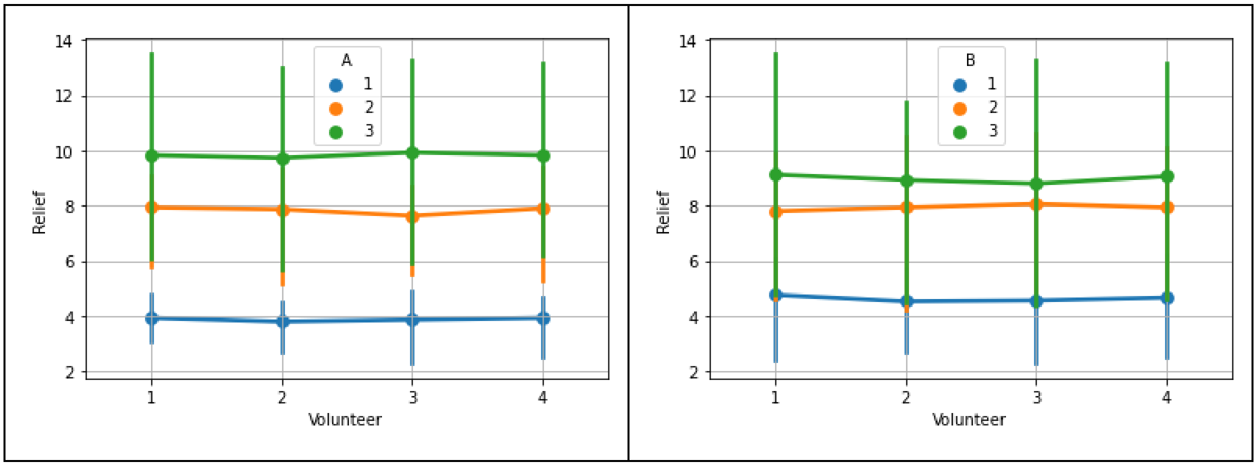
Solution:

Analysis of the effects of one variable on another with the help of an interaction plot.

Formula:

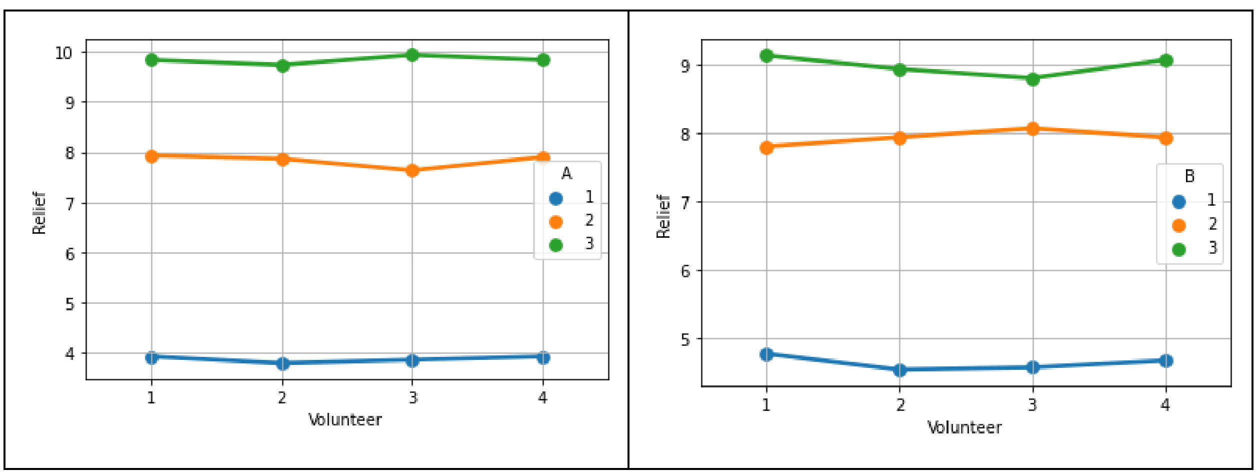
sns.pointplot(x = 'Volunteer', y = 'Relief', hue='A', data=df)

sns.pointplot(x = 'Volunteer', y = 'Relief', hue='B', data=df)



sns.pointplot(x= 'Volunteer', y= 'Relief', hue='A', data=df,ci= None)

sns.pointplot(x= 'Volunteer', y= 'Relief', hue='B', data=df,ci= None)



Insights from the graphs (Pairplots):

- The second amount variation for both ingredients A and B have very strong interaction with the third variation.

- In the case of Ingredient A and Volunteer 2, the highest interaction for 2nd and 3rd amount variation can be observed while amount 1 has no interaction from other amount variation. The relief is also the lowest from amount 1.
- We can observe the strong evidence of interaction between the variation 2 and 3. For volunteer 3 the interaction is strongest and the difference in terms of relief is less than one. While looking into amount 1 of ingredient B the interaction is weakest.

Problem 1.5

- Perform a two-way ANOVA based on the different ingredients (variable 'A' & 'B') with the variable 'Relief' and state your results.

Solution:

- H0: The means of the 'Relief' variable with respect to each ingredient category and their interaction is equal.
- H1: At least one of the means of the 'Relief' variable with respect to each ingredient category and their interaction is unequal.

Formula:

model=ols('Relief ~ C(A)+C(B)+C(A):C(B)',data=df).fit()

aov_table=anova_lm(model,type=2)

	df	sum_sq	mean_sq	F	PR(>F)
C(A)	2.0	220.020	110.010000	1827.858462	1.514043e-29
C(B)	2.0	123.660	61.830000	1027.329231	3.348751e-26
C(A):C(B)	4.0	29.425	7.356250	122.226923	6.972083e-17
Residual	27.0	1.625	0.060185	NaN	NaN

Insights:

- After performing two-way Anova on the different variable (A and B) along with the interaction of variables with Relief variable, it is found that there is no significant value as all values of P in each case is greater than the value of alpha α=0.05

Problem 1.6

- Mention the business implications of performing ANOVA for this particular case study.

Solution:

The analysis of the variables of the given data and the respective results leads to two significant conclusions on a business level. These conclusions are as follows:

Quality and Cost Comparison:

- In the graph, it can be seen that the amount three of both ingredients A and B is most effective for relief in fever. The best quality of a medicine can be made while using the third amount and it can be sold at a premium cost. In case of a severe case of fever, this dose can be implemented.
- It can also be seen that both ingredients A and B for amount 1 (marked in blue in the graph) is not effective in terms of relief from fever.
- The different combination can be compared for the medicine for the ingredients A and B

Production Optimization:

- The different combination of Ingredients can be used as per the requirement of treatment, effectivity, and also market requirements. The A/B test can be performed for better outcomes from the medicine.
- For different types of patients, the different or specific doses could be produced on the basis of predictive analysis.

Problem 2: Education - Post 12th Standard

24.01.2021

Problem Statement

The dataset Education - Post 12th Standard.csv is a dataset that contains the names of various colleges. This particular case study is based on various parameters of various institutions. You are expected to do a Principal Component Analysis for this case study according to the instructions given in the following rubric. The data dictionary of the 'Education - Post 12th Standard.csv' can be found in the following file: Data Dictionary.xlsx.

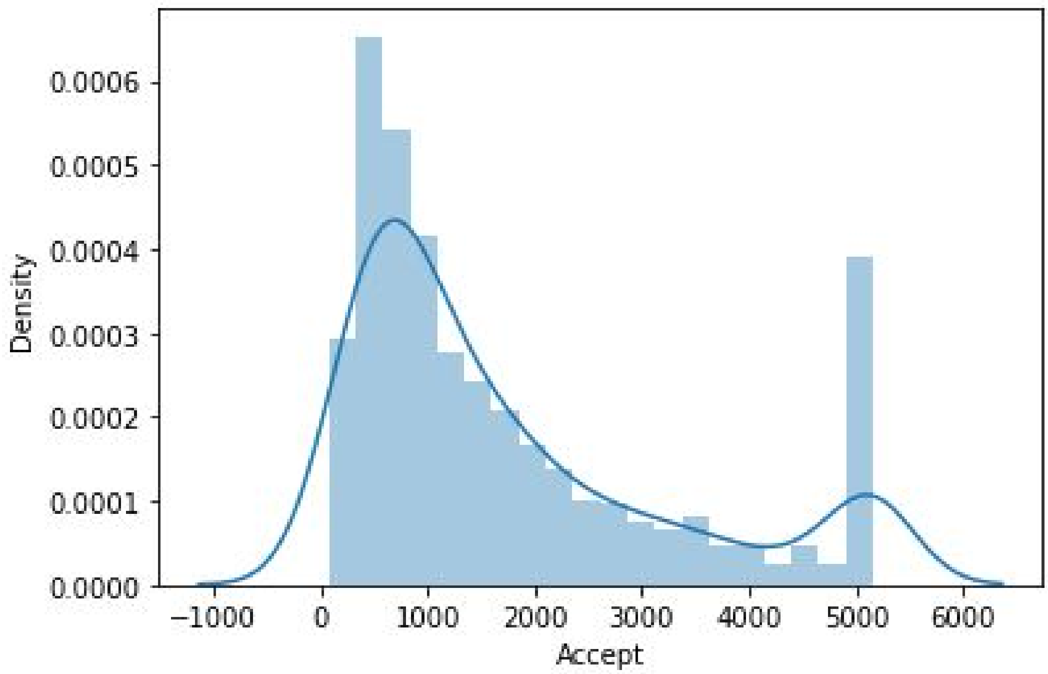
Problem 2.1

- Perform Exploratory Data Analysis [both univariate and multivariate analysis to be performed]. The inferences drawn from this should be properly documented.

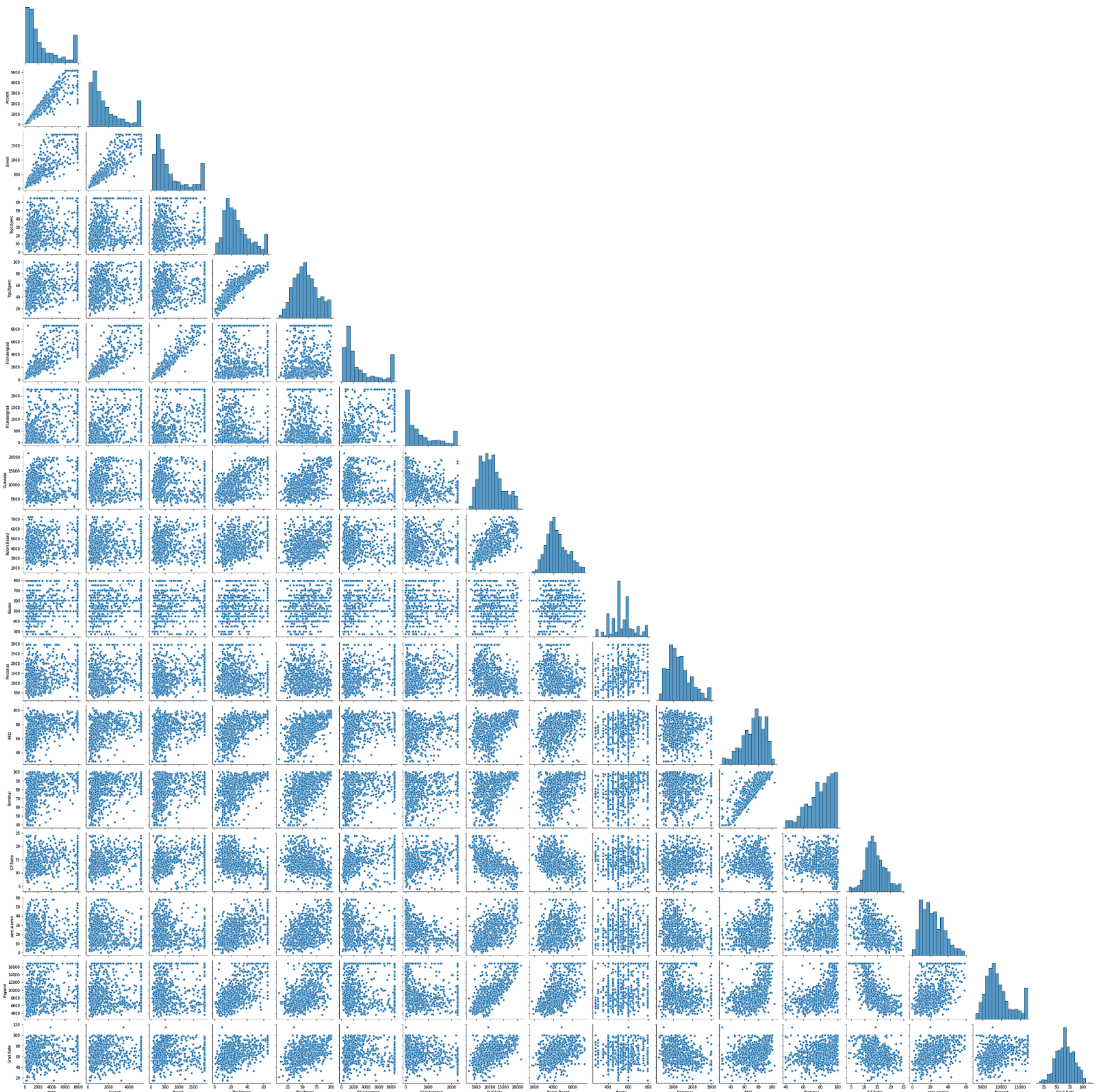
Solution:

Univariate Analysis:

<AxesSubplot:xlabel='Accept', ylabel='Density'>



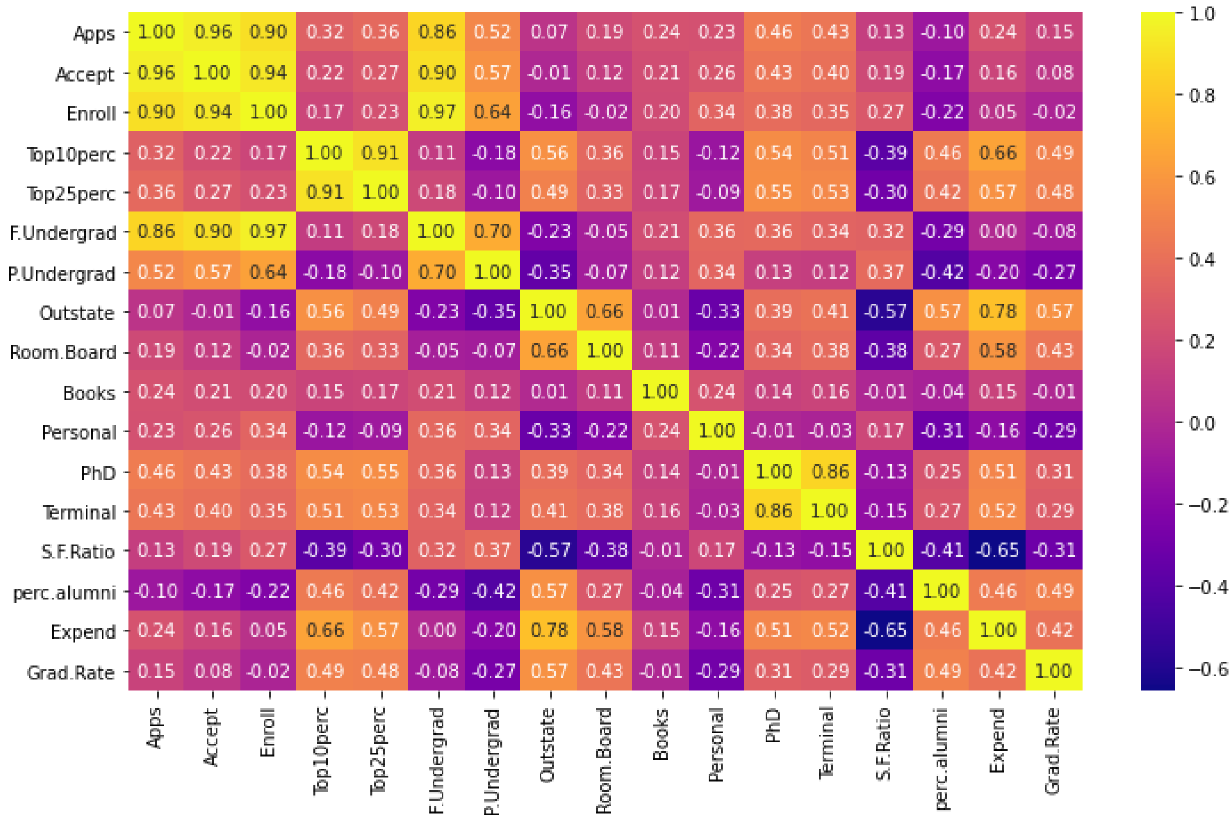
<u>Multivariate Analysis:</u>



Insights:

- The univariate analysis shows that the acceptance of the applications are increasing over the time and this would lead to a better key performance in terms of producing competitive top ranked students.
- The multivariate analysis shows the outliers and the growth in each column. It displays the rate of increment and decrement in each dataset.

Heatmap:



Problem 2.2

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

Solution:

Please refer to the attached Python notebook to look into the codes, analysis and detailed results

Standard Scaled Data:

The standard scaled data can be found in the below-given chart:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
0	Abilene Christian University	-0.37649	-0.33783	0.10638	-0.24678	-0.191827	-0.018769	-0.166083	-0.74648	-0.96832	450	1.4385	-0.17405	-0.12324	1.070602	-0.87047	-0.63092	-0.31921
1	Adelphi University	-0.1592	0.116744	-0.26044	-0.69629	-1.353911	-0.093626	0.797856	0.457762	1.92168	750	0.289289	-2.74573	-2.78507	-0.48951	-0.54573	0.396097	-0.55269
2	Adrian College	-0.47234	-0.42651	-0.56934	-0.310996	-0.292878	-0.703966	-0.777974	0.201488	-0.55547	400	-0.26069	-1.24035	-0.9529	-0.30441	0.590864	-0.13185	-0.66944
3	Agnes Scott College	-0.88999	-0.91787	-0.91861	2.129202	1.677612	-0.898889	-0.828267	0.626954	1.004218	450	-0.73679	1.205884	1.190391	-1.67943	1.159159	2.28794	-0.37758
4	Alaska Pacific University	-0.98253	-1.05122	-1.06253	-0.69629	-0.596031	-0.99561	0.297726	-0.71662	-0.21601	795	0.289289	0.202299	-0.53807	-0.56884	-1.68232	0.512468	-2.91676

Encoding:
Encoded data are retrieved post identification of dummies. The chart is given below:

dummies = pd.get_dummies(df[["Apps", "Accept","Enroll","Top10perc","Top25perc","Outstate","Personal","Expend","Grad.Rate"]],columns =["Apps", "Accept","Enroll","Top10perc","Top25perc","Outstate","Personal","Expend","Grad.Rate"], prefix = ["Apps", "Accept","Enroll","Top10perc","Top25perc","Outstate","Personal","Expend","Grad.Rate"],drop_first=True).head()

Apps_-1.0209517429285273	Apps_-1.0040140462152114	Apps_-1.0002960152293616	Apps_-0.999469786121395	Apps_-0.9932730678116454	Apps_-0.9903812659337623	Apps_-0.9833583185160459	Apps_-0.9825320894080793	Apps_-0.9788140584222296	Apps_-0.9746829128823965	...	Grad.Rate_1.5486984329168878	Grad.Rate_1.6070704281813273	Grad.Rate_1.6654424234457668	Grad.Rate_1.7238144187102065	Grad.Rate_1.782186413974646	Grad.Rate_1.8405584092390856	Grad.Rate_1.8989304045035251	Grad.Rate_1.9573023997679646	Grad.Rate_2.0156743950324043	Grad.Rate_2.920440321631218	
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0	0	0	0

Problem 2.3

- Comment on the comparison between covariance and the correlation matrix.

Solution:

Please refer to the attached Python notebook to look into the codes, analysis, and detailed results

<u>Comparing Correlation and Covariance Matrix:</u>

Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate	
Apps	1	0.955307	0.896883	0.321342	0.364491	0.861002	0.519823	0.065337	0.187475	0.236138	0.229948	0.463924	0.434478	0.126411	-0.10116	0.242935	0.150803
Accept	0.955307	1	0.935277	0.223298	0.273681	0.897034	0.572691	-0.005002	0.119586	0.208705	0.256346	0.427341	0.403409	0.188506	-0.16552	0.161808	0.078982
Enroll	0.896883	0.935277	1	0.171756	0.230434	0.967302	0.641595	-0.155655	-0.02385	0.202057	0.339348	0.38154	0.354379	0.274269	-0.22272	0.054221	-0.02325
Top10perc	0.321342	0.223298	0.171756	1	0.913875	0.111215	-0.18001	0.56216	0.357366	0.153452	-0.11673	0.544048	0.506748	-0.387926	0.455797	0.657039	0.49367
Top25perc	0.364491	0.273681	0.230434	0.913875	1	0.181196	-0.0993	0.489569	0.330987	0.169761	-0.08681	0.551461	0.527654	-0.297233	0.416832	0.572905	0.478985
F.Undergrad	0.861002	0.897034	0.967302	0.111215	0.181196	1	0.69613	-0.226166	-0.05448	0.207879	0.359783	0.361564	0.335054	0.324504	-0.28546	0.000371	-0.08224
P.Undergrad	0.519823	0.572691	0.641595	-0.180009	-0.099295	0.69613	1	-0.354216	-0.06764	0.122529	0.344053	0.127663	0.122152	0.370607	-0.41933	-0.20193	-0.26516
Outstate	0.065337	-0.005	-0.155655	0.56216	0.489569	-0.226166	-0.35422	1	0.655489	0.00511	-0.32561	0.391321	0.412579	-0.573683	0.565736	0.775328	0.572458
Room.Board	0.187475	0.119586	-0.023846	0.357366	0.330987	-0.054476	-0.06764	0.655489	1	0.108924	-0.21955	0.341469	0.37927	-0.37643	0.272393	0.580622	0.42579
Books	0.236138	0.208705	0.202057	0.153452	0.169761	0.207879	0.122529	0.00511	0.108924	1	0.239863	0.13639	0.159318	-0.008536	-0.04283	0.149983	-0.00805
Personal	0.229948	0.256346	0.339348	-0.11673	-0.08681	0.359783	0.344053	-0.325609	-0.21955	0.239863	1	-0.01168	-0.03197	0.173913	-0.30575	-0.16327	-0.29089
PhD	0.463924	0.427341	0.38154	0.544048	0.551461	0.361564	0.127663	0.391321	0.341469	0.13639	-0.01168	1	0.862928	-0.12939	0.248877	0.510529	0.310019
Terminal	0.434478	0.403409	0.354379	0.506748	0.527654	0.335054	0.122152	0.412579	0.37927	0.159318	-0.03197	0.862928	1	-0.150993	0.266033	0.524068	0.292803
S.F.Ratio	0.126411	0.188506	0.274269	-0.387926	-0.297233	0.324504	0.370607	-0.573683	-0.37643	-0.00854	0.173913	-0.12939	-0.15099	1	-0.4121	-0.65438	-0.30853
perc.alumni	-0.10116	-0.16552	-0.222723	0.455797	0.416832	-0.285457	-0.41933	0.565736	0.272393	-0.04283	-0.30575	0.248877	0.266033	-0.412101	1	0.462922	0.491408
Expend	0.242935	0.161808	0.054221	0.657039	0.572905	0.000371	-0.20193	0.775328	0.580622	0.149983	-0.16327	0.510529	0.524068	-0.654376	0.462922	1	0.415291
Grad.Rate	0.150803	0.078982	-0.023251	0.49367	0.478985	-0.082239	-0.26516	0.572458	0.42579	-0.00805	-0.29089	0.310019	0.292803	-0.308525	0.491408	0.415291	1

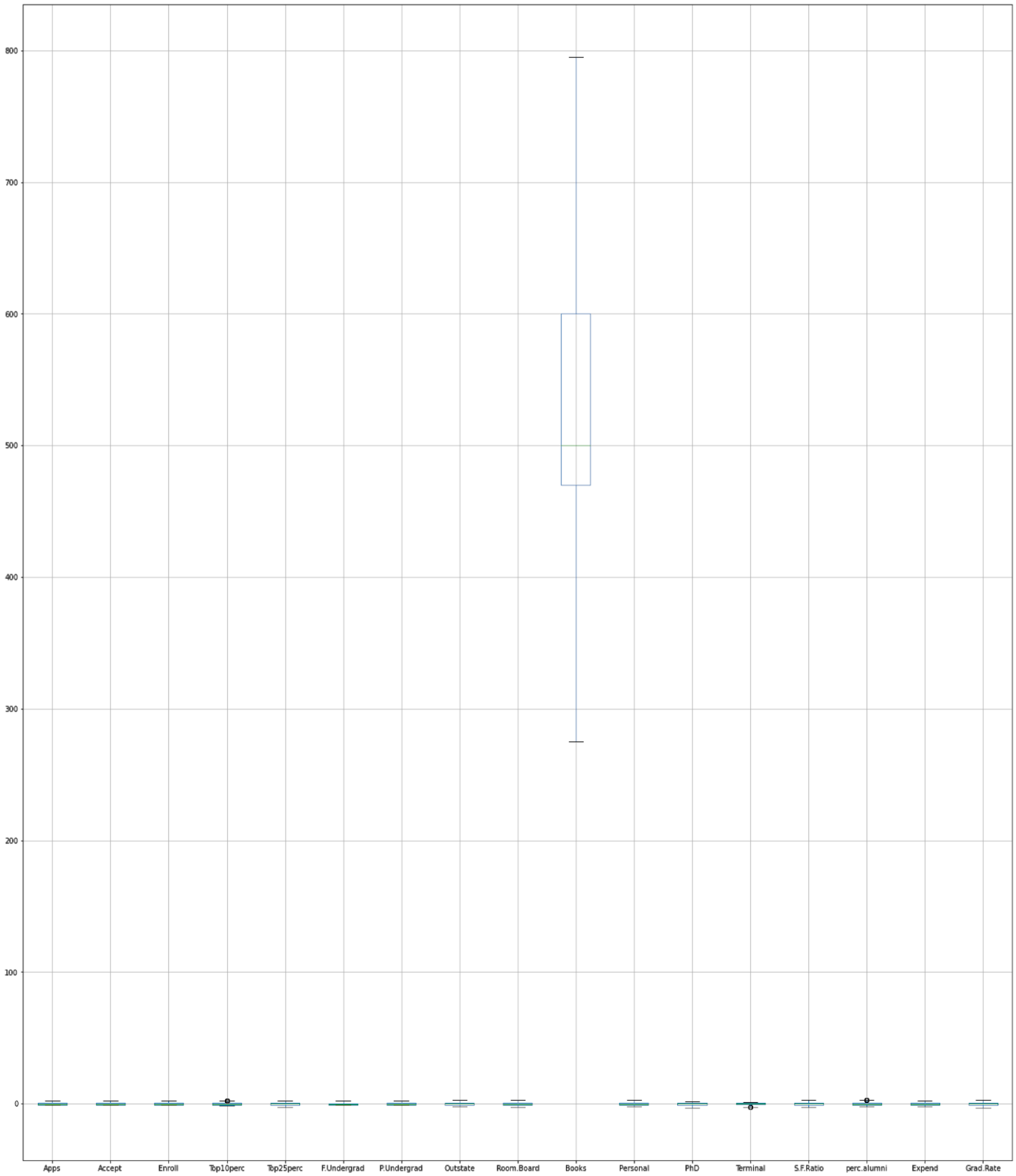
Problem 2.4

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

Solution:

Please refer to the attached Python notebook to look into the codes, analysis and detailed results

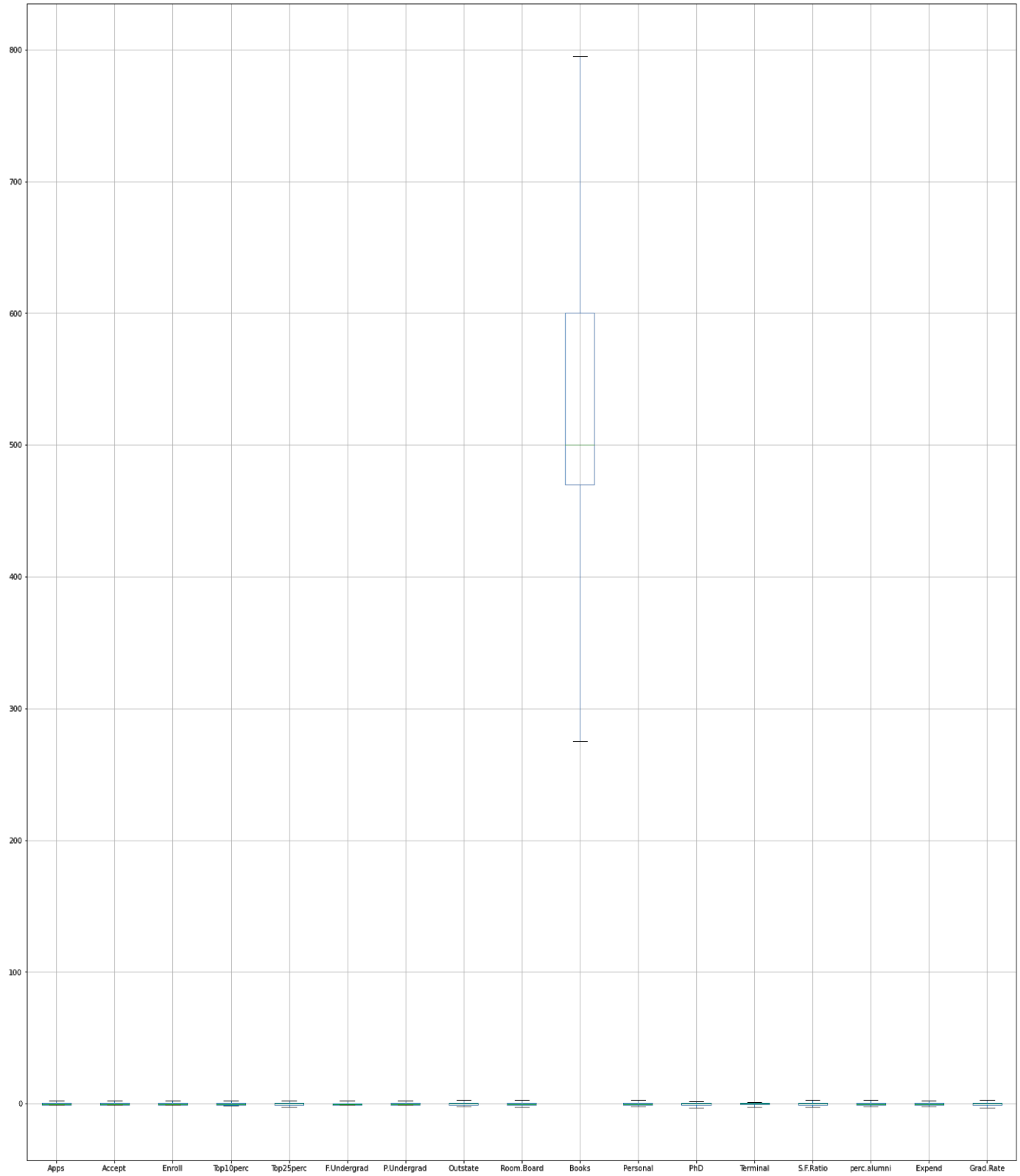
<u>Checking outliers of the current dataset:</u>


800
700
600
500
400
300
200
100
0
Apps
Accept
Enroll
Top10perc
Top25perc
F.Undergrad
P.Undergrad
Outstate
Room.Board
Books
Personal
PhD
Terminal
S.F.Ratio
perc.alumni
Expend
Grad.Rate

Handling Outliers or Outlier Treatment

Codes are given in the Jupiter Notebook

Checking outliers after treatment:



Problem 2.5

- Build the covariance matrix, eigenvalues, and eigenvector.

Solution:

Please refer to the attached Python notebook to look into the codes, analysis and detailed results

Covariance matrix:

```
%s [[ 1.00128866e+00  9.56537704e-01  8.98039052e-01  3.21756324e-01
   3.64960691e-01  8.62111140e-01  5.20492952e-01  6.54209711e-02
   1.87717056e-01  2.36441941e-01  2.30243993e-01  4.64521757e-01
   4.35037784e-01  1.26573895e-01 -1.01288006e-01  2.43248206e-01
   1.50997775e-01]
 [ 9.56537704e-01  1.00128866e+00  9.36482483e-01  2.23586208e-01
   2.74033187e-01  8.98189799e-01  5.73428908e-01 -5.00874847e-03
   1.19740419e-01  2.08974091e-01  2.56676290e-01  4.27891234e-01
   4.03929238e-01  1.88748711e-01 -1.65728801e-01  1.62016688e-01
   7.90839722e-02]
 [ 8.98039052e-01  9.36482483e-01  1.00128866e+00  1.71977357e-01
   2.30730728e-01  9.68548601e-01  6.42421828e-01 -1.55856056e-01
  -2.38762560e-02  2.02317274e-01  3.39785395e-01  3.82031198e-01
   3.54835877e-01  2.74622251e-01 -2.23009677e-01  5.42906862e-02
  -2.32810071e-02]
 [ 3.21756324e-01  2.23586208e-01  1.71977357e-01  1.00128866e+00
   9.15052977e-01  1.11358019e-01 -1.80240778e-01  5.62884044e-01
   3.57826139e-01  1.53650150e-01 -1.16880152e-01  5.44748764e-01
   5.07401238e-01 -3.88425719e-01  4.56384036e-01  6.57885921e-01
   4.94306540e-01]
 [ 3.64960691e-01  2.74033187e-01  2.30730728e-01  9.15052977e-01
   1.00128866e+00  1.81429267e-01 -9.94231153e-02  4.90200034e-01
```

```
  3.31413314e-01  1.69979808e-01 -8.69219644e-02  5.52172085e-01
  5.28333659e-01 -2.97616423e-01  4.17369123e-01  5.73643193e-01
  4.79601950e-01]
 [ 8.62111140e-01  8.98189799e-01  9.68548601e-01  1.11358019e-01
  1.81429267e-01  1.00128866e+00  6.97027420e-01 -2.26457040e-01
 -5.45459528e-02  2.08147257e-01  3.60246460e-01  3.62030390e-01
  3.35485771e-01  3.24921933e-01 -2.85825062e-01  3.71119607e-04
 -8.23447851e-02]
 [ 5.20492952e-01  5.73428908e-01  6.42421828e-01 -1.80240778e-01
 -9.94231153e-02  6.97027420e-01  1.00128866e+00 -3.54672874e-01
 -6.77252009e-02  1.22686416e-01  3.44495974e-01  1.27827147e-01
  1.22309141e-01  3.71084841e-01 -4.19874031e-01 -2.02189396e-01
 -2.65499420e-01]
 [ 6.54209711e-02 -5.00874847e-03 -1.55856056e-01  5.62884044e-01
  4.90200034e-01 -2.26457040e-01 -3.54672874e-01  1.00128866e+00
  6.56333564e-01  5.11656377e-03 -3.26028927e-01  3.91824814e-01
  4.13110264e-01 -5.74421963e-01  5.66465309e-01  7.76326650e-01
  5.73195743e-01]
 [ 1.87717056e-01  1.19740419e-01 -2.38762560e-02  3.57826139e-01
  3.31413314e-01 -5.45459528e-02 -6.77252009e-02  6.56333564e-01
  1.00128866e+00  1.09064551e-01 -2.19837042e-01  3.41908577e-01
  3.79759015e-01 -3.76915472e-01  2.72743761e-01  5.81370284e-01
  4.26338910e-01]
 [ 2.36441941e-01  2.08974091e-01  2.02317274e-01  1.53650150e-01
  1.69979808e-01  2.08147257e-01  1.22686416e-01  5.11656377e-03
  1.09064551e-01  1.00128866e+00  2.40172145e-01  1.36566243e-01
  1.59523091e-01 -8.54689129e-03 -4.28870629e-02  1.50176551e-01
 -8.06107505e-03]
 [ 2.30243993e-01  2.56676290e-01  3.39785395e-01 -1.16880152e-01
 -8.69219644e-02  3.60246460e-01  3.44495974e-01 -3.26028927e-01
```

-2.19837042e-01 2.40172145e-01 1.00128866e+00 -1.16986124e-02
-3.20117803e-02 1.74136664e-01 -3.06146886e-01 -1.63481407e-01
-2.91268705e-01]
[4.64521757e-01 4.27891234e-01 3.82031198e-01 5.44748764e-01
5.52172085e-01 3.62030390e-01 1.27827147e-01 3.91824814e-01
3.41908577e-01 1.36566243e-01 -1.16986124e-02 1.00128866e+00
8.64040263e-01 -1.29556494e-01 2.49197779e-01 5.11186852e-01
3.10418895e-01]
[4.35037784e-01 4.03929238e-01 3.54835877e-01 5.07401238e-01
5.28333659e-01 3.35485771e-01 1.22309141e-01 4.13110264e-01
3.79759015e-01 1.59523091e-01 -3.20117803e-02 8.64040263e-01
1.00128866e+00 -1.51187934e-01 2.66375402e-01 5.24743500e-01
2.93180212e-01]
[1.26573895e-01 1.88748711e-01 2.74622251e-01 -3.88425719e-01
-2.97616423e-01 3.24921933e-01 3.71084841e-01 -5.74421963e-01
-3.76915472e-01 -8.54689129e-03 1.74136664e-01 -1.29556494e-01
-1.51187934e-01 1.00128866e+00 -4.12632056e-01 -6.55219504e-01
-3.08922187e-01]
[-1.01288006e-01 -1.65728801e-01 -2.23009677e-01 4.56384036e-01
4.17369123e-01 -2.85825062e-01 -4.19874031e-01 5.66465309e-01
2.72743761e-01 -4.28870629e-02 -3.06146886e-01 2.49197779e-01
2.66375402e-01 -4.12632056e-01 1.00128866e+00 4.63518674e-01
4.92040760e-01]
[2.43248206e-01 1.62016688e-01 5.42906862e-02 6.57885921e-01
5.73643193e-01 3.71119607e-04 -2.02189396e-01 7.76326650e-01
5.81370284e-01 1.50176551e-01 -1.63481407e-01 5.11186852e-01
5.24743500e-01 -6.55219504e-01 4.63518674e-01 1.00128866e+00
4.15826026e-01]
[1.50997775e-01 7.90839722e-02 -2.32810071e-02 4.94306540e-01
4.79601950e-01 -8.23447851e-02 -2.65499420e-01 5.73195743e-01

4.26338910e-01 -8.06107505e-03 -2.91268705e-01 3.10418895e-01
2.93180212e-01 -3.08922187e-01 4.92040760e-01 4.15826026e-01
1.00128866e+00]]

Eigen Values

%s [5.6625219 4.89470815 1.12636744 1.00397659 0.87218426 0.7657541
0.58491404 0.5445048 0.42352336 0.38101777 0.24701456 0.02239369
0.03789395 0.14726392 0.13434483 0.09883384 0.07469003]

Eigen Vectors

%s [[-2.62171542e-01 3.14136258e-01 -8.10177245e-02 9.87761685e-02
2.19898081e-01 -2.18800617e-03 2.83715076e-02 8.99498102e-02
-1.30566998e-01 1.56464458e-01 8.62132843e-02 -1.82169814e-01
5.99137640e-01 -8.99775288e-02 -8.88697944e-02 5.49428396e-01
5.41453698e-03]
[-2.30562461e-01 3.44623583e-01 -1.07658626e-01 1.18140437e-01
1.89634940e-01 1.65212882e-02 1.29584896e-02 1.37606312e-01
-1.42275847e-01 1.49209799e-01 4.25899061e-02 3.91041719e-01
-6.61496927e-01 -1.58861886e-01 -4.37945938e-02 2.91572312e-01
1.44582845e-02]
[-1.89276397e-01 3.82813322e-01 -8.55296892e-02 9.30717094e-03
1.62314818e-01 6.80794143e-02 1.52403625e-02 1.44216938e-01
-5.08712481e-02 6.48997860e-02 4.38408622e-02 -7.16684935e-01
-2.33235272e-01 3.53988202e-02 6.19241658e-02 -4.17001280e-01
-4.97908902e-02]
[-3.38874521e-01 -9.93191661e-02 7.88293849e-02 -3.69115031e-01
1.57211016e-01 8.88656824e-02 2.57455284e-01 -2.89538833e-01
1.22467790e-01 3.58776186e-02 -1.77837341e-03 5.62053913e-02
-2.21448729e-02 3.92277722e-02 -6.99599977e-02 8.79767299e-03
-7.23645373e-01]
[-3.34690532e-01 -5.95055011e-02 5.07938247e-02 -4.16824361e-01
1.44449474e-01 2.76268979e-02 2.39038849e-01 -3.45643551e-01
1.93936316e-01 -6.41786425e-03 1.02127328e-01 -1.96735274e-02
-3.22646978e-02 -1.45621999e-01 9.70282598e-02 -1.07779150e-02
6.55464648e-01]
[-1.63293010e-01 3.98636372e-01 -7.37077827e-02 1.39504424e-02
1.02728468e-01 5.16468727e-02 3.11751439e-02 1.08748900e-01
-1.45452749e-03 1.63981359e-04 3.49993487e-02 5.42774834e-01
3.67681187e-01 1.33555923e-01 8.71753137e-02 -5.70683843e-01
2.53059904e-02]

```
[-2.24797091e-02  3.57550046e-01 -4.03568700e-02  2.25351078e-01
 -9.56790178e-02  2.45375721e-02  1.00138971e-02 -1.23841696e-01
  6.34774326e-01 -5.46346279e-01 -2.52107094e-01 -2.95029745e-02
 -2.62494456e-02 -5.02487566e-02 -4.45537493e-02  1.46321060e-01
 -3.97146972e-02]
[-2.83547285e-01 -2.51863617e-01 -1.49394795e-02  2.62975384e-01
  3.72750885e-02  2.03860462e-02 -9.45370782e-02 -1.12721477e-02
  8.36648339e-03  2.31799759e-01 -5.93433149e-01 -1.03393587e-03
  8.14247697e-02 -5.60392799e-01 -6.72405494e-02 -2.11561014e-01
 -1.59275617e-03]
[-2.44186588e-01 -1.31909124e-01  2.11379165e-02  5.80894132e-01
 -6.91080879e-02 -2.37267409e-01 -9.45210745e-02 -3.89639465e-01
  2.20526518e-01  2.55107620e-01  4.75297296e-01 -9.85725168e-03
 -2.67779296e-02  1.07365653e-01 -1.77715010e-02 -1.00935084e-01
 -2.82578388e-02]
[-9.67082754e-02  9.39739472e-02  6.97121128e-01 -3.61562884e-02
  3.54056654e-02 -6.38604997e-01  1.11193334e-01  2.39817267e-01
 -2.10246624e-02 -9.11624912e-02 -4.35697999e-02 -4.36086500e-03
 -1.04624246e-02 -5.16224550e-02 -3.54343707e-02 -2.86384228e-02
 -8.06259380e-03]
[ 3.52299594e-02  2.32439594e-01  5.30972806e-01 -1.14982973e-01
 -4.75358244e-04  3.81495854e-01 -6.39418106e-01 -2.77206569e-01
 -1.73715184e-02  1.27647512e-01 -1.51627393e-02  1.08725257e-02
 -4.54572099e-03 -9.39409228e-03  1.18604404e-02  3.38197909e-02
  1.42590097e-03]
[-3.26410696e-01  5.51390195e-02 -8.11134044e-02 -1.47260891e-01
 -5.50786546e-01 -3.34444832e-03 -8.92320786e-02  3.42628480e-02
 -1.66510079e-01 -1.00975002e-01  3.91865961e-02 -1.33146759e-02
 -1.25137966e-02  7.16590441e-02 -7.02656469e-01 -6.38096394e-02
  8.31471932e-02]
[-3.23115980e-01  4.30332048e-02 -5.89785929e-02 -8.90079921e-02
 -5.90407136e-01 -3.54121294e-02 -9.16985445e-02  9.03076644e-02
 -1.12609034e-01 -8.60363025e-02  8.48575651e-02 -7.38135022e-03
  1.79275275e-02 -1.63820871e-01  6.62488717e-01  9.85019644e-02
 -1.13374007e-01]
[ 1.63151642e-01  2.59804556e-01 -2.74150657e-01 -2.59486122e-01
 -1.42842546e-01 -4.68752604e-01 -1.52864837e-01 -2.42807562e-01
  1.53685343e-01  4.70527925e-01 -3.63042716e-01 -8.85797314e-03
 -1.83059753e-02  2.39902591e-01  4.79006197e-02  6.19970446e-02
  3.83160891e-03]
[-1.86610828e-01 -2.57092552e-01 -1.03715887e-01 -2.23982467e-01
  1.28215768e-01 -1.25669415e-02 -3.91400512e-01  5.66073056e-01
  5.39235753e-01  1.47628917e-01  1.73918533e-01  2.40534190e-02
```

```
 8.03169296e-05  4.89753356e-02 -3.58875507e-02  2.80805469e-02
 -7.32598621e-03]
[-3.28955847e-01 -1.60008951e-01  1.84205687e-01  2.13756140e-01
 -2.24240837e-02  2.31562325e-01  1.50501305e-01  1.18823549e-01
 -2.42371616e-02  8.04154875e-02 -3.93722676e-01 -1.05658769e-02
 -5.60069250e-02  6.90417042e-01  1.26667522e-01  1.28739213e-01
  1.45099786e-01]
[-2.38822447e-01 -1.67523664e-01 -2.45335837e-01 -3.61915064e-02
  3.56843227e-01 -3.13556243e-01 -4.68641965e-01 -1.80458508e-01
 -3.15812873e-01 -4.88415259e-01 -8.72638706e-02  2.51028410e-03
 -1.48410810e-02  1.59332164e-01  6.30737002e-02 -7.09643331e-03
 -3.29024228e-03]]
```

Problem 2.6

- Write the explicit form of the first PC (in terms of Eigen Vectors).

Solution:

Explicit form of the first PC in terms of Eigen Vectors

```
%s [[-2.62171542e-01  3.14136258e-01 -8.10177245e-02  9.87761685e-02
  2.19898081e-01 -2.18800617e-03  2.83715076e-02  8.99498102e-02
 -1.30566998e-01  1.56464458e-01  8.62132843e-02 -1.82169814e-01
  5.99137640e-01 -8.99775288e-02 -8.88697944e-02  5.49428396e-01
  5.41453698e-03]
[-2.30562461e-01  3.44623583e-01 -1.07658626e-01  1.18140437e-01
  1.89634940e-01  1.65212882e-02  1.29584896e-02  1.37606312e-01
 -1.42275847e-01  1.49209799e-01  4.25899061e-02  3.91041719e-01
 -6.61496927e-01 -1.58861886e-01 -4.37945938e-02  2.91572312e-01
  1.44582845e-02]
[-1.89276397e-01  3.82813322e-01 -8.55296892e-02  9.30717094e-03
  1.62314818e-01  6.80794143e-02  1.52403625e-02  1.44216938e-01
 -5.08712481e-02  6.48997860e-02  4.38408622e-02 -7.16684935e-01
 -2.33235272e-01  3.53988202e-02  6.19241658e-02 -4.17001280e-01
 -4.97908902e-02]
[-3.38874521e-01 -9.93191661e-02  7.88293849e-02 -3.69115031e-01
  1.57211016e-01  8.88656824e-02  2.57455284e-01 -2.89538833e-01
  1.22467790e-01  3.58776186e-02 -1.77837341e-03  5.62053913e-02
 -2.21448729e-02  3.92277722e-02 -6.99599977e-02  8.79767299e-03
 -7.23645373e-01]
[-3.34690532e-01 -5.95055011e-02  5.07938247e-02 -4.16824361e-01
```

```
 1.44449474e-01  2.76268979e-02  2.39038849e-01 -3.45643551e-01
 1.93936316e-01 -6.41786425e-03  1.02127328e-01 -1.96735274e-02
-3.22646978e-02 -1.45621999e-01  9.70282598e-02 -1.07779150e-02
 6.55464648e-01]
[-1.63293010e-01  3.98636372e-01 -7.37077827e-02  1.39504424e-02
 1.02728468e-01  5.16468727e-02  3.11751439e-02  1.08748900e-01
-1.45452749e-03  1.63981359e-04  3.49993487e-02  5.42774834e-01
 3.67681187e-01  1.33555923e-01  8.71753137e-02 -5.70683843e-01
 2.53059904e-02]
[-2.24797091e-02  3.57550046e-01 -4.03568700e-02  2.25351078e-01
-9.56790178e-02  2.45375721e-02  1.00138971e-02 -1.23841696e-01
 6.34774326e-01 -5.46346279e-01 -2.52107094e-01 -2.95029745e-02
-2.62494456e-02 -5.02487566e-02 -4.45537493e-02  1.46321060e-01
-3.97146972e-02]
[-2.83547285e-01 -2.51863617e-01 -1.49394795e-02  2.62975384e-01
 3.72750885e-02  2.03860462e-02 -9.45370782e-02 -1.12721477e-02
 8.36648339e-03  2.31799759e-01 -5.93433149e-01 -1.03393587e-03
 8.14247697e-02 -5.60392799e-01 -6.72405494e-02 -2.11561014e-01
-1.59275617e-03]
[-2.44186588e-01 -1.31909124e-01  2.11379165e-02  5.80894132e-01
-6.91080879e-02 -2.37267409e-01 -9.45210745e-02 -3.89639465e-01
 2.20526518e-01  2.55107620e-01  4.75297296e-01 -9.85725168e-03
-2.67779296e-02  1.07365653e-01 -1.77715010e-02 -1.00935084e-01
-2.82578388e-02]
[-9.67082754e-02  9.39739472e-02  6.97121128e-01 -3.61562884e-02
 3.54056654e-02 -6.38604997e-01  1.11193334e-01  2.39817267e-01
-2.10246624e-02 -9.11624912e-02 -4.35697999e-02 -4.36086500e-03
-1.04624246e-02 -5.16224550e-02 -3.54343707e-02 -2.86384228e-02
-8.06259380e-03]
[ 3.52299594e-02  2.32439594e-01  5.30972806e-01 -1.14982973e-01
-4.75358244e-04  3.81495854e-01 -6.39418106e-01 -2.77206569e-01
-1.73715184e-02  1.27647512e-01 -1.51627393e-02  1.08725257e-02
-4.54572099e-03 -9.39409228e-03  1.18604404e-02  3.38197909e-02
 1.42590097e-03]
[-3.26410696e-01  5.51390195e-02 -8.11134044e-02 -1.47260891e-01
-5.50786546e-01 -3.34444832e-03 -8.92320786e-02  3.42628480e-02
-1.66510079e-01 -1.00975002e-01  3.91865961e-02 -1.33146759e-02
-1.25137966e-02  7.16590441e-02 -7.02656469e-01 -6.38096394e-02
 8.31471932e-02]
[-3.23115980e-01  4.30332048e-02 -5.89785929e-02 -8.90079921e-02
-5.90407136e-01 -3.54121294e-02 -9.16985445e-02  9.03076644e-02
```

```
 -1.12609034e-01 -8.60363025e-02  8.48575651e-02 -7.38135022e-03
  1.79275275e-02 -1.63820871e-01  6.62488717e-01  9.85019644e-02
 -1.13374007e-01]
[ 1.63151642e-01  2.59804556e-01 -2.74150657e-01 -2.59486122e-01
 -1.42842546e-01 -4.68752604e-01 -1.52864837e-01 -2.42807562e-01
  1.53685343e-01  4.70527925e-01 -3.63042716e-01 -8.85797314e-03
 -1.83059753e-02  2.39902591e-01  4.79006197e-02  6.19970446e-02
  3.83160891e-03]
[-1.86610828e-01 -2.57092552e-01 -1.03715887e-01 -2.23982467e-01
  1.28215768e-01 -1.25669415e-02 -3.91400512e-01  5.66073056e-01
  5.39235753e-01  1.47628917e-01  1.73918533e-01  2.40534190e-02
  8.03169296e-05  4.89753356e-02 -3.58875507e-02  2.80805469e-02
 -7.32598621e-03]
[-3.28955847e-01 -1.60008951e-01  1.84205687e-01  2.13756140e-01
 -2.24240837e-02  2.31562325e-01  1.50501305e-01  1.18823549e-01
 -2.42371616e-02  8.04154875e-02 -3.93722676e-01 -1.05658769e-02
 -5.60069250e-02  6.90417042e-01  1.26667522e-01  1.28739213e-01
  1.45099786e-01]
[-2.38822447e-01 -1.67523664e-01 -2.45335837e-01 -3.61915064e-02
  3.56843227e-01 -3.13556243e-01 -4.68641965e-01 -1.80458508e-01
 -3.15812873e-01 -4.88415259e-01 -8.72638706e-02  2.51028410e-03
 -1.48410810e-02  1.59332164e-01  6.30737002e-02 -7.09643331e-03
 -3.29024228e-03]]
```

Problem 2.7

- 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.

Solution:

Please refer to the attached Python notebook to look into the codes, analysis and detailed results

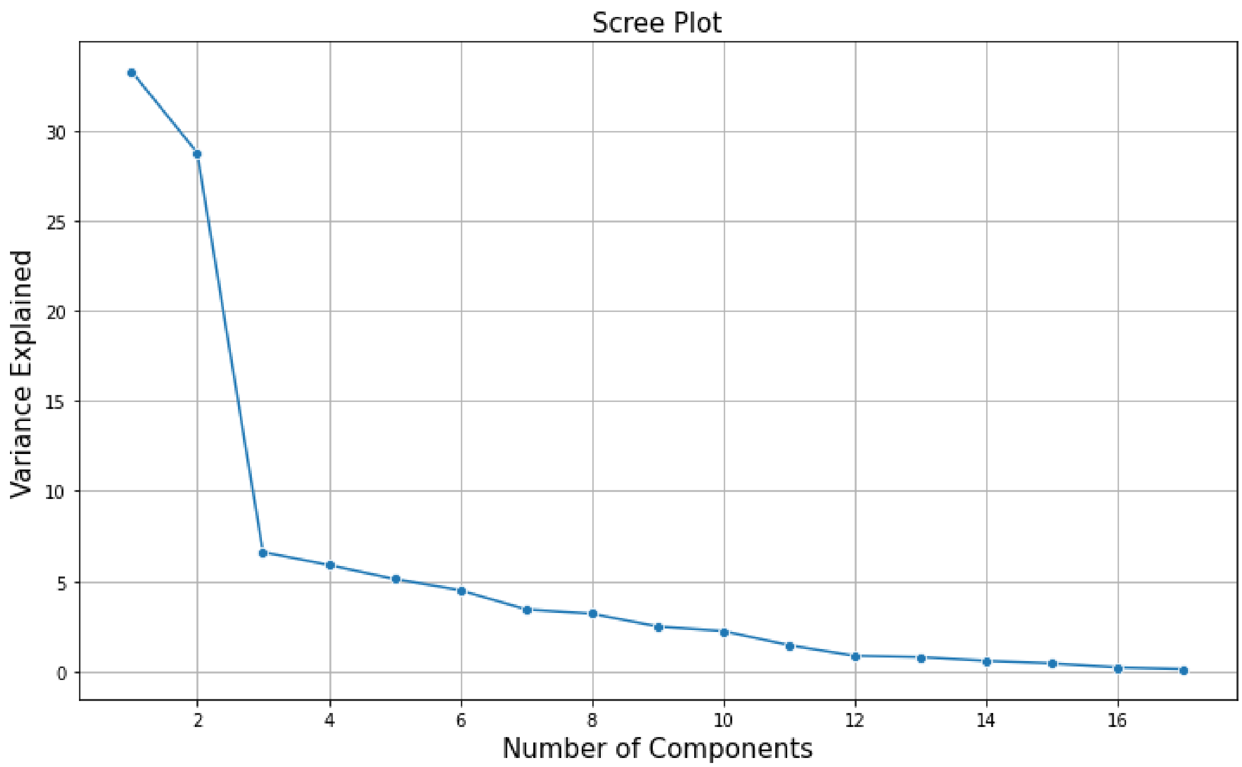
Cumulative Variance Explained

[33.26608367 62.02142867 68.63859223 74.53673619 79.66062886

 84.15926753 87.59551019 90.79435736 93.28246491 95.52086136

96.97201814 97.83716159 98.62640821 99.20703552 99.64582321
99.86844192 100.]

Scree Plot



Principle Component Analysis:

Array ([[-1.60249937, -1.80467542, -1.60828258, ..., -0.57688266,
6.57095201, -0.47739306],
[0.99368301, -0.07041506, -1.3827921 , ..., 0.01779842,
-1.18493016, 1.04394671],
[0.03004474, 2.12212631, -0.50151217, ..., 0.32215983,
1.32596477, -1.42543857],
[-1.00842101, 3.13898354, -0.03638515, ..., -0.58723704,
0.07771533, -1.30026532]])

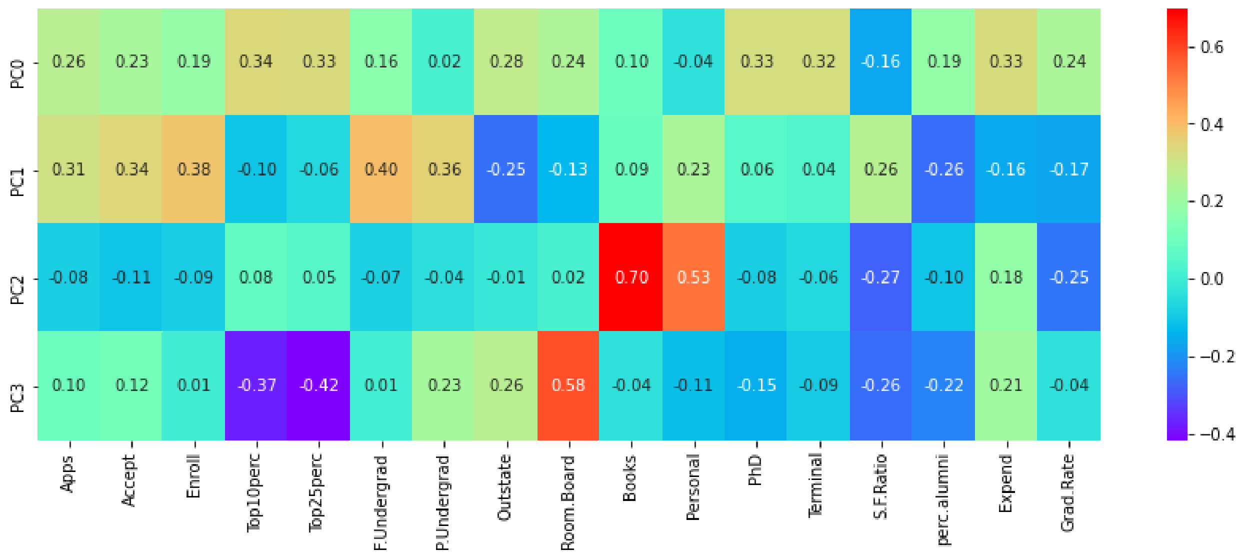
PCA Components:

Array ([[0.26217154, 0.23056246, 0.1892764 , 0.33887452, 0.33469053,
0.16329301, 0.02247971, 0.28354729, 0.24418659, 0.09670828,
-0.03522996, 0.3264107 , 0.32311598, -0.16315164, 0.18661083,
0.32895585, 0.23882245],
[0.31413626, 0.34462358, 0.38281332, -0.09931917, -0.0595055 ,
0.39863637, 0.35755005, -0.25186362, -0.13190912, 0.09397395,
0.23243959, 0.05513902, 0.0430332 , 0.25980456, -0.25709255,
-0.16000895, -0.16752366],
[-0.08101774, -0.10765861, -0.08552969, 0.07882921, 0.050794 ,
-0.07370778, -0.04035688, -0.01493945, 0.02113791, 0.69712113,
0.53097281, -0.0811134 , -0.0589786 , -0.27415067, -0.10371589,
0.18420568, -0.24533585],
[0.09877621, 0.11814021, 0.00930744, -0.36910855, -0.41683051,
0.01395012, 0.22535143, 0.26297418, 0.58089453, -0.03615632,
-0.11498298, -0.14726105, -0.08900775, -0.25948557, -0.22398226,
0.21375641, -0.03619118]])

Variance Ratio:
Array ([0.33266084, 0.28755345, 0.06617164, 0.05898144])
Or: Cumulative percentage
Array ([33.3, 62.1, 68.7, 74.6])

Note: *The Cumulative percentage describes the percentage of variance accounted for by the n components.*
In the above array we see that the first feature explains 33.3% of the variance within our data set while the first two explain 62.1 and so on. If we employ 4 features we capture ~ 75% (round figure) of the variance within the dataset, thus we gain very little by implementing an additional feature.

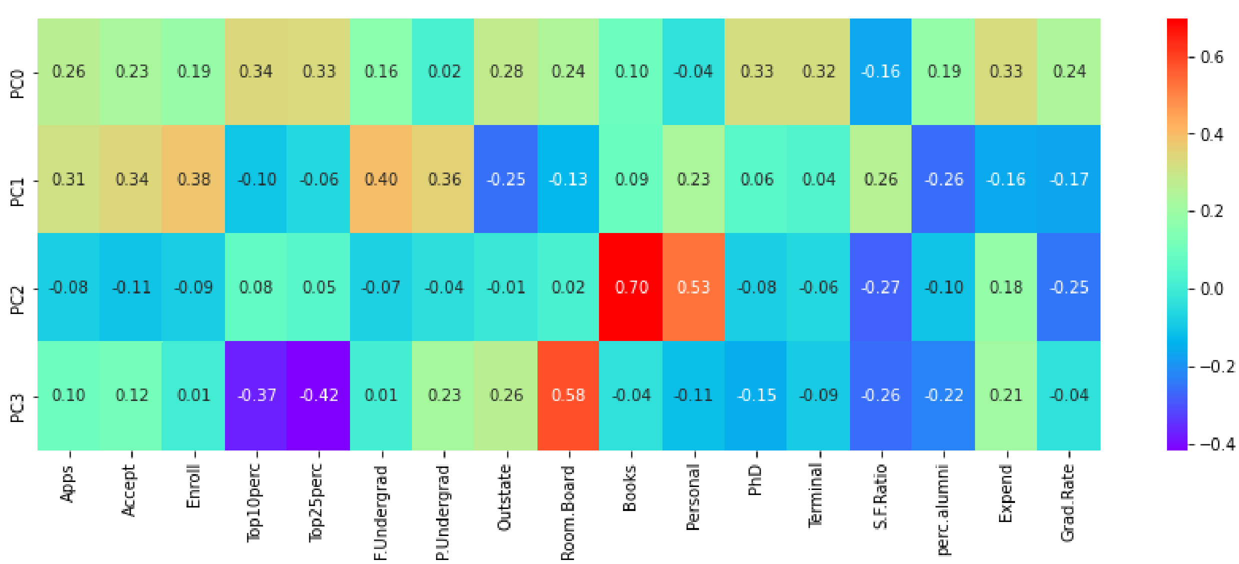
Correlation between components and features



Problem 2.8

- Mention the business implication of using the Principal Component Analysis for this case study.

Looking at the PCA done for the dataset given and we have drawn the below given heatmap.



Looking at the heatmap, the color bar basically represents the correlation between the various features and the principal component itself.

In the current status quo PC2 shows the highest correlation with the Number of applications, number of accepted applications and number of new students enrolled but the negative correlation with the estimated book cost of students and personal expenses of the students.

Post clear observation of the report, we can conclude that there'd be 3 major action items which would decide a growth of the business:

1. Strength and focused area
2. Highly improvement required area
3. Focused result driven area

Let's look at the each action items:

1. **Strength and focused area:**
   The ratio of the applications and enrollment is quite impressive and those need to be maintained. This would require strengthening the lacking areas where the major improvements are needed
2. **Highly improvement required area**
   Post enrollment the facility assurance of the students is lacking in PC2 and PC3, this would display a direct impact on the enrollment. However, if these areas are strengthened on a priority then the student enrollment would be constant.
3. **Focused result driven area**
   The student-faculty ratio, percentage of faculty with PhD and alumni donation are some of the key areas where all the colleges are performing well. These are the driving factors of the high enrollment and it would keep going well if the above two mentioned areas are taken care properly.