Advanced Statistics

Table of Contents

Problem Statement: ***1***

Analysis: **1**

Preparing the Data: **2**

Hypothesis Testing: 3

Let us perform one-way ANOVA for variable 'A' 3

Let us perform one-way ANOVA for variable 'B' 3

Conclusion: 4

Problem Statement: ***4***

Analysis: **4**

Bivariate Analysis: 10

Scale the variables: 10

Comparison between covariance and the correlation matrix: 11

Check the dataset for outliers before and after scaling 12

Build the covariance matrix, eigenvalues and eigenvector 17

Explicit form of the first PC: 19

Cumulative values of the eigenvalues: 19

spree plot. 19

Perform PCA and export the data of the Principal Component scores into a data frame 19

Business implication of using the Principal Component Analysis for this case study: 20

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

Analysis:

First let's look at the dataset, there are four columns (excluding the index column), With the active ingredients specified as A & B and volunteers randomly put in and the relief time in hours.

Information Classification: General

	A	B	Volunteer	Relief
0	1	1	1	2.4
1	1	1	2	2.7
2	1	1	3	2.3
3	1	1	4	2.5
4	1	2	1	4.6

Preparing the Data:

Let's analyze the data and look for any issues with data. First let's check how data is provided. There are 36 entries in the data and 4 columns, and the type of the data are

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 4 columns):
A            36 non-null category
B            36 non-null category
Volunteer    36 non-null category
Relief       36 non-null float64
dtypes: category(3), float64(1)
memory usage: 924.0 bytes
```

As we can clearly see except Relief column rest of the columns are categorical. Also, we can see there are no null values in the Data, let's describe the data, as Relief is the only variable we can able to describe it alone rest all are categorical.

	Relief
count	36.000000
mean	7.183333
std	3.272090
min	2.300000
25%	4.675000
50%	6.000000
75%	9.325000
max	13.500000

Information Classification: General

Hypothesis Testing:

Now let us state the Null and Alternate Hypothesis for conducting one-way ANOVA for both the variables 'A'.

Null Hypothesis $H0$: The mean Relief hours for the Hay fever using active ingredient 'A' are same between all the volunteers

Alternate Hypothesis HA : The mean Relief hours for the hay fever using active ingredient 'A' is different for at least one of the Volunteer

Now let us state the Null and Alternate Hypothesis for conducting one-way ANOVA for both the variables 'B'.

Null Hypothesis $H0$: The mean Relief hours for the Hay fever using active ingredient 'B' are same between all the volunteers

Alternate Hypothesis HA : The mean Relief hours for the hay fever using active ingredient 'B' is different for at least one of the Volunteer

Let us perform one-way ANOVA for variable 'A' with respect to the variable 'Relief' and state whether the Null Hypothesis is accepted or rejected based on the ANOVA results

	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

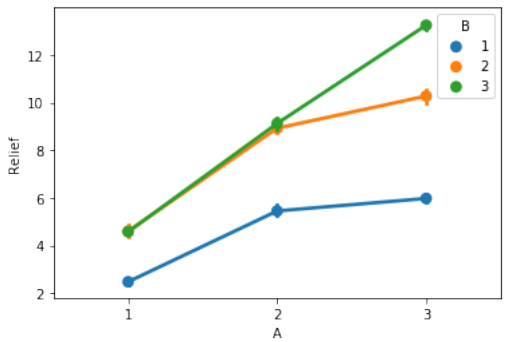
With the test result we came to a conclusion the **the p value is less than the 5% significant value so rejecting the null hypothesis for Active ingredient A**

Let us perform one-way ANOVA for variable 'B' with respect to the variable 'Relief' and state whether the Null Hypothesis is accepted or rejected based on the ANOVA results

	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

With the test result we came to a conclusion the **the p value is less than the 5% significant value so rejecting the null hypothesis for Active ingredient B**

Now let us analyse the effects of one variable on another with the help of an interaction plot.



We can see there is interaction between Ingredient A & B at 2 places

Now Let's perform two way ANOVA on the dataset with both the active ingredients and Relief hours. Below is the result of the analysis

	df	sum_sq	mean_sq	F	PR(>F)
C(A)	2.0	220.02	110.010000	109.832850	8.514029e-15
C(B)	2.0	123.66	61.830000	61.730435	1.546749e-11
Residual	31.0	31.05	1.001613	NaN	NaN

Since the p value in both of the above scenarios are less than α (0.05), we can say that we reject the null hypothesis ($H0$).

Conclusion:

By doing the hypothesis testing on both the active ingredient and the Relief time and also on each volunteers we can clearly say that the active ingredients work different way on different volunteers and also each volunteer is reacting in different ways for different active ingredients. We have come to the conclusion based on the p value we derived using both one way and two way ANOVA.

Problem Statement:

The dataset Education - Post 12th Standard.csv is a dataset which contains the names of various colleges. This particular case study is based on various parameters of various institutions. You are expected to do Principal Component Analysis for this case study according to the instructions given in the following rubric.

Analysis:

First let's look at the dataset to understand the data provided to us.

```
RangeIndex: 777 entries, 0 to 776
Data columns (total 18 columns):
Names           777 non-null object
Apps            777 non-null int64
Accept          777 non-null int64
Enroll          777 non-null int64
Top10perc       777 non-null int64
Top25perc       777 non-null int64
F.Undergrad     777 non-null int64
P.Undergrad     777 non-null int64
Outstate        777 non-null int64
Room.Board      777 non-null int64
Books           777 non-null int64
Personal        777 non-null int64
PhD             777 non-null int64
Terminal        777 non-null int64
S.F.Ratio       777 non-null float64
perc.alumni     777 non-null int64
Expend          777 non-null int64
Grad.Rate       777 non-null int64
dtypes: float64(1), int64(16), object(1)
memory usage: 109.4+ KB
```

There are totally 18 columns available everything seems like continuous variable except *Names* there are 777 records available in each column, and there are no null values in the dataset.

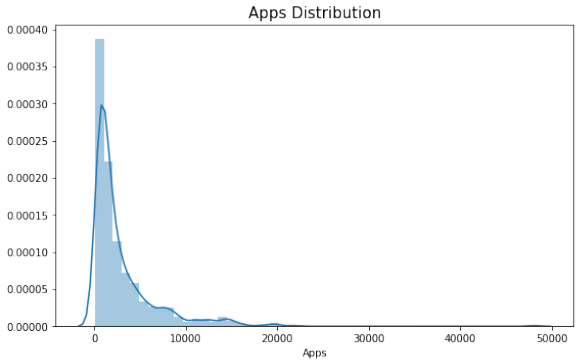
Let's describe the data

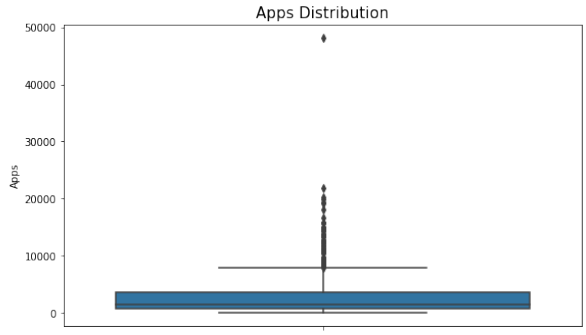
	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal
count	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000
mean	3001.638353	2018.804376	779.972973	27.558559	55.796654	3699.907336	855.298584	10440.669241	4357.526384	549.380952	1340.642214
std	3870.201484	2451.113971	929.176190	17.640364	19.804778	4850.420531	1522.431887	4023.016484	1096.696416	165.105360	677.071454
min	81.000000	72.000000	35.000000	1.000000	9.000000	139.000000	1.000000	2340.000000	1780.000000	96.000000	250.000000
25%	776.000000	604.000000	242.000000	15.000000	41.000000	992.000000	95.000000	7320.000000	3597.000000	470.000000	850.000000
50%	1558.000000	1110.000000	434.000000	23.000000	54.000000	1707.000000	353.000000	9990.000000	4200.000000	500.000000	1200.000000
75%	3624.000000	2424.000000	902.000000	35.000000	69.000000	4005.000000	967.000000	12925.000000	5050.000000	600.000000	1700.000000
max	48094.000000	26330.000000	6392.000000	96.000000	100.000000	31643.000000	21836.000000	21700.000000	8124.000000	2340.000000	6800.000000

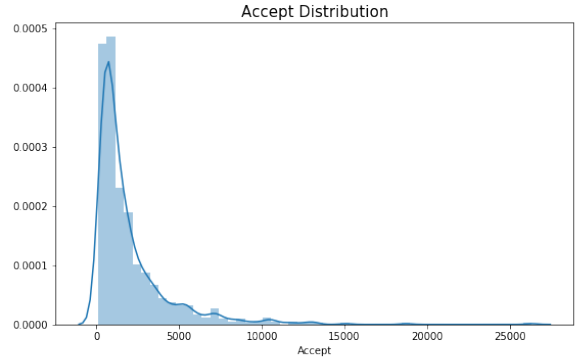
PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
777.000000	777.000000	777.000000	777.000000	777.000000	777.00000
72.660232	79.702703	14.089704	22.743887	9660.171171	65.46332
16.328155	14.722359	3.958349	12.391801	5221.768440	17.17771
8.000000	24.000000	2.500000	0.000000	3186.000000	10.00000
62.000000	71.000000	11.500000	13.000000	6751.000000	53.00000
75.000000	82.000000	13.600000	21.000000	8377.000000	65.00000
85.000000	92.000000	16.500000	31.000000	10830.000000	78.00000
103.000000	100.000000	39.800000	64.000000	56233.000000	118.00000

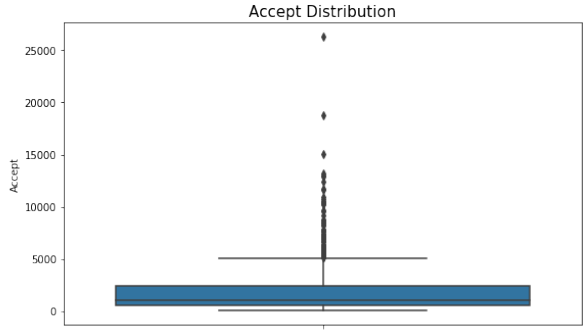
Information Classification: General

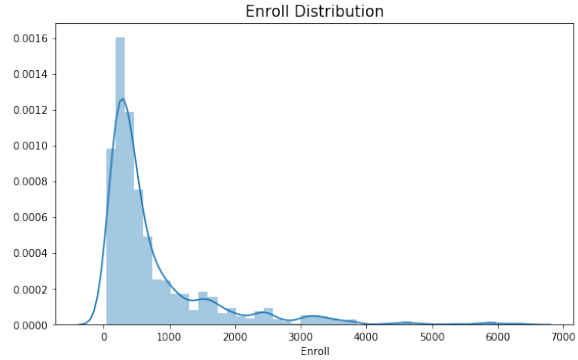
There are no duplicates in the data and no null values. Now let's start with univariate analysis for each values separately.

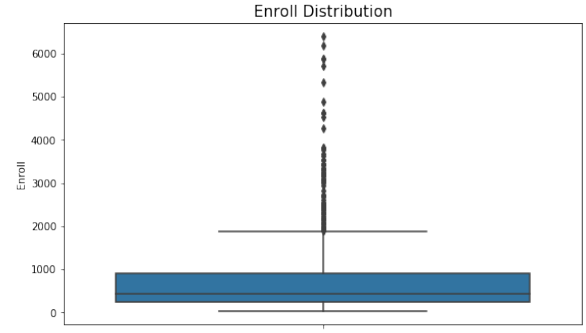


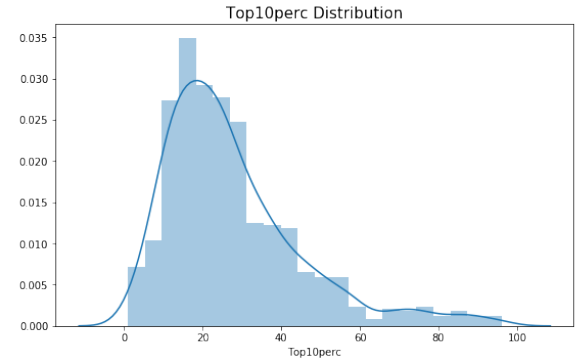


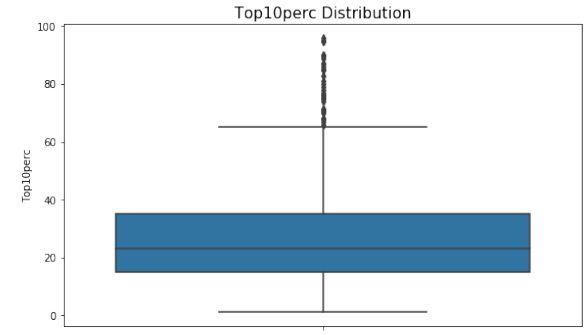


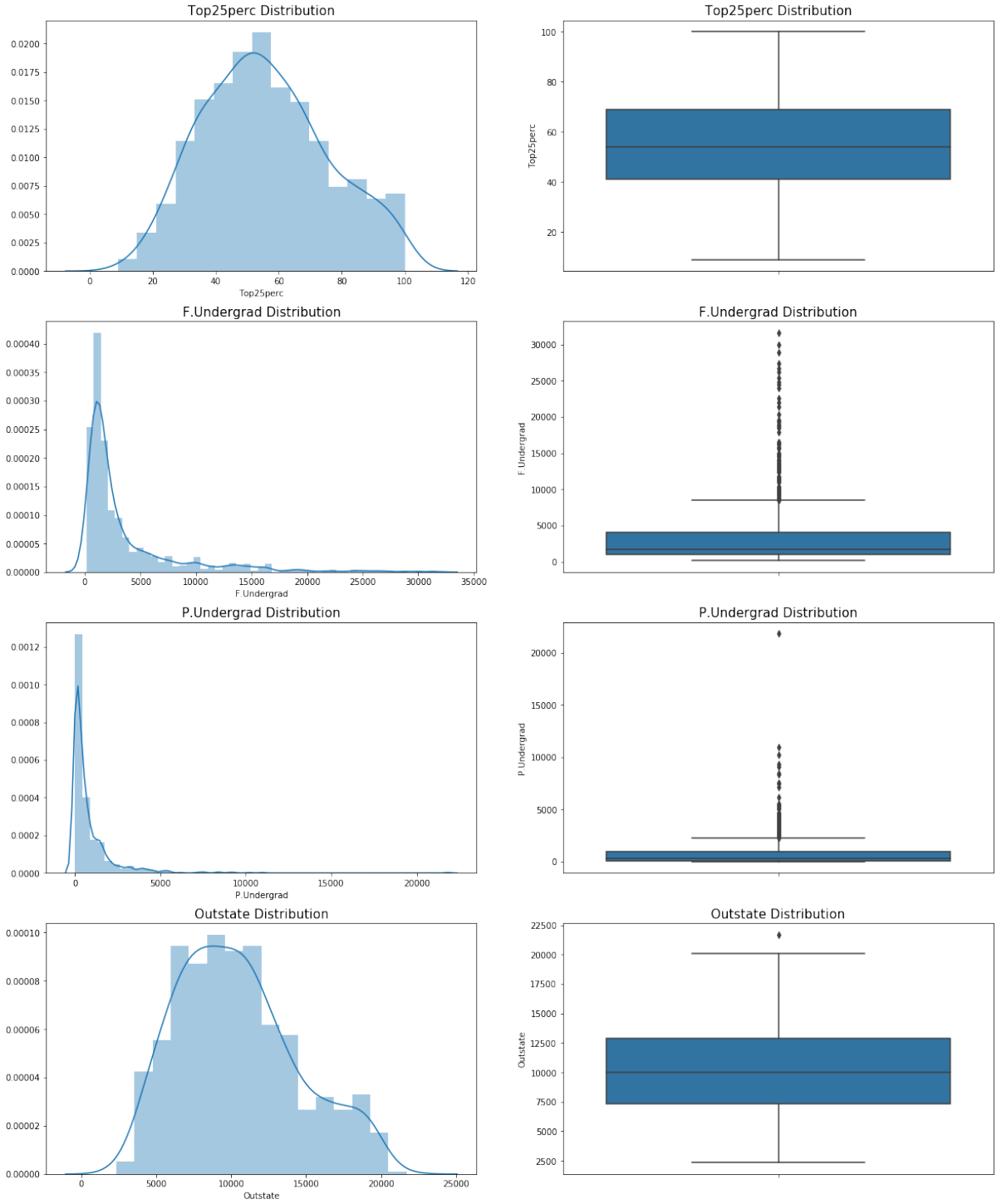


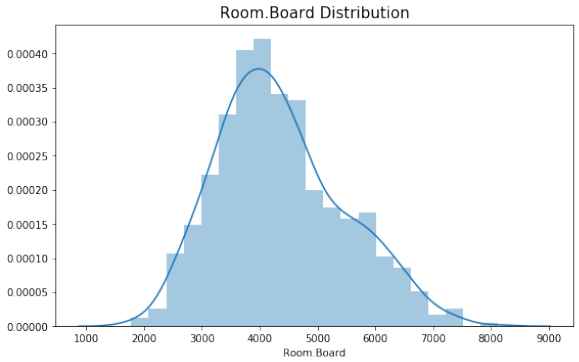


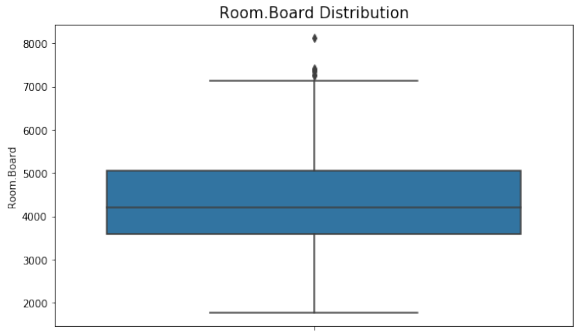


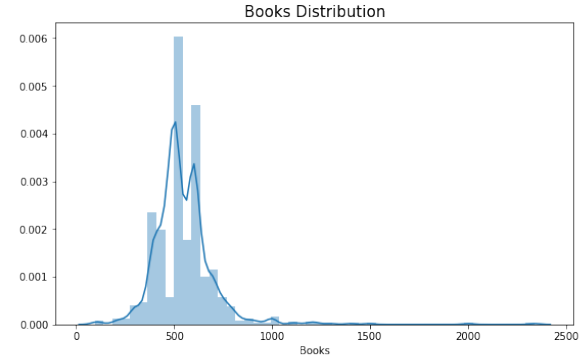


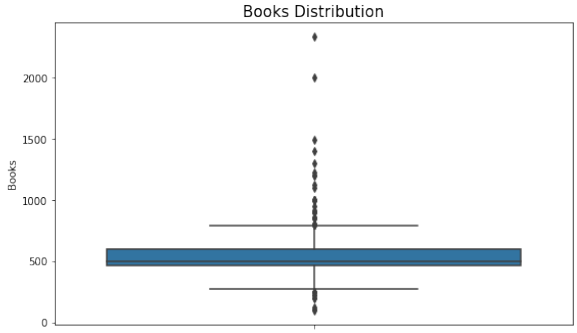


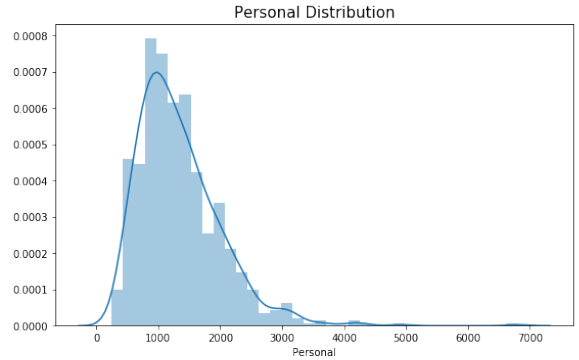

Top25perc Distribution
Top25perc
Top25perc Distribution
Top25perc
F.Undergrad Distribution
F.Undergrad
F.Undergrad Distribution
F.Undergrad
P.Undergrad Distribution
P.Undergrad
P.Undergrad Distribution
P.Undergrad
Outstate Distribution
Outstate
Outstate Distribution
Outstate

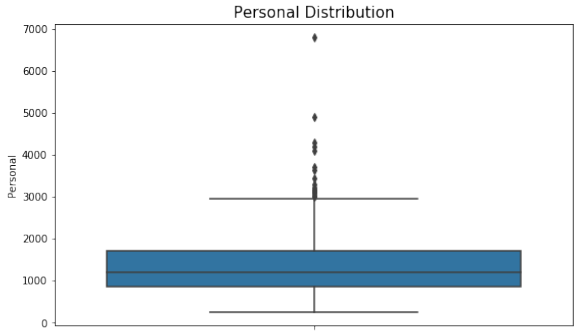


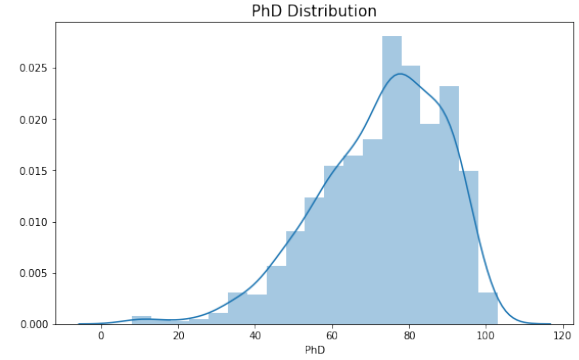


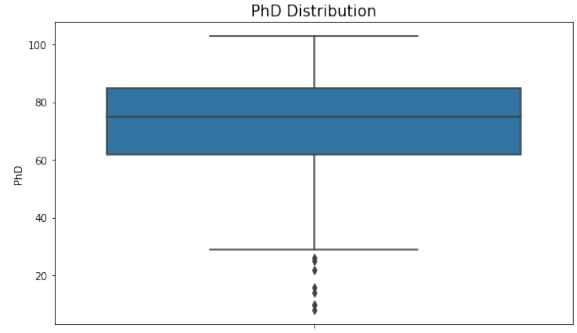




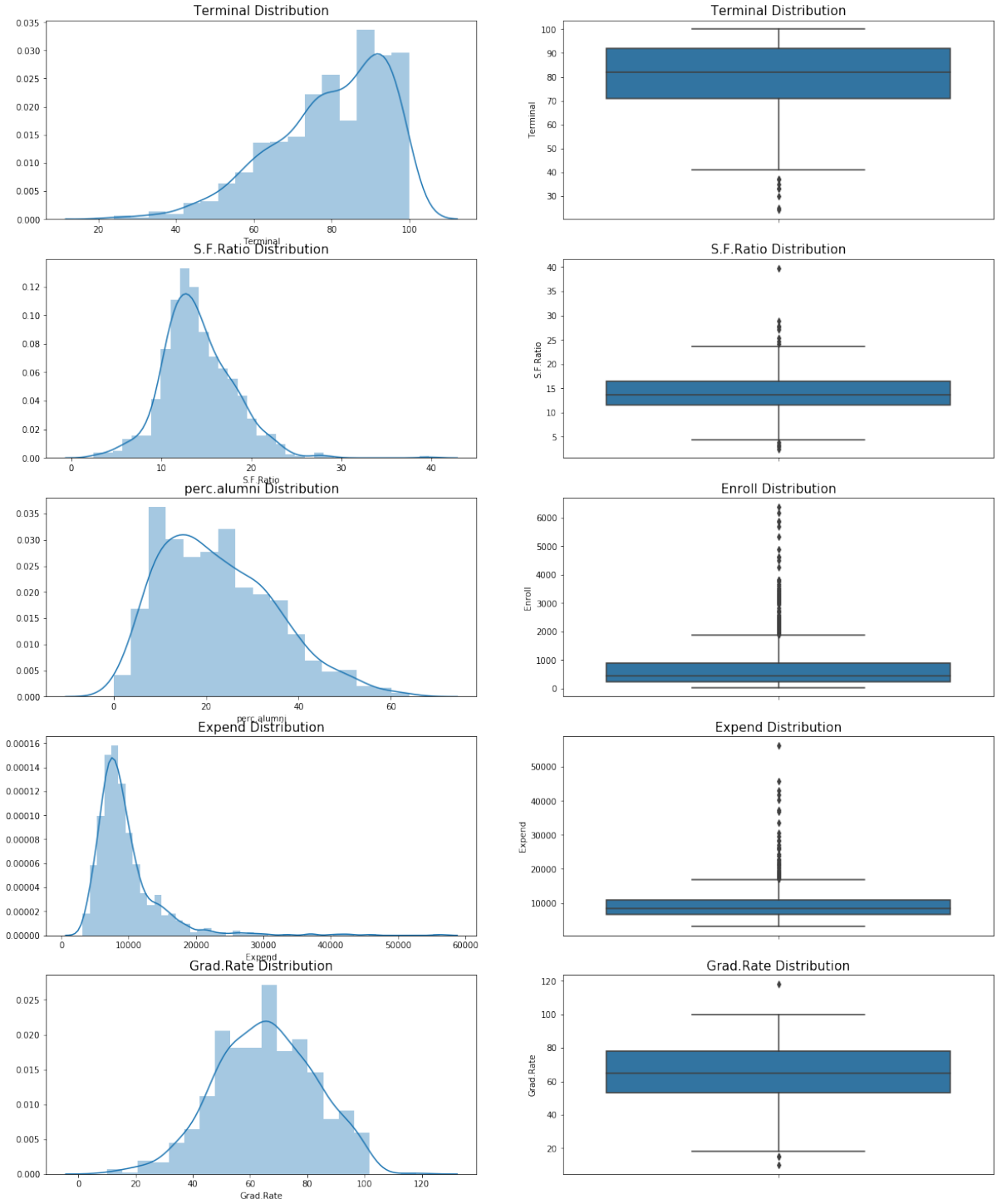






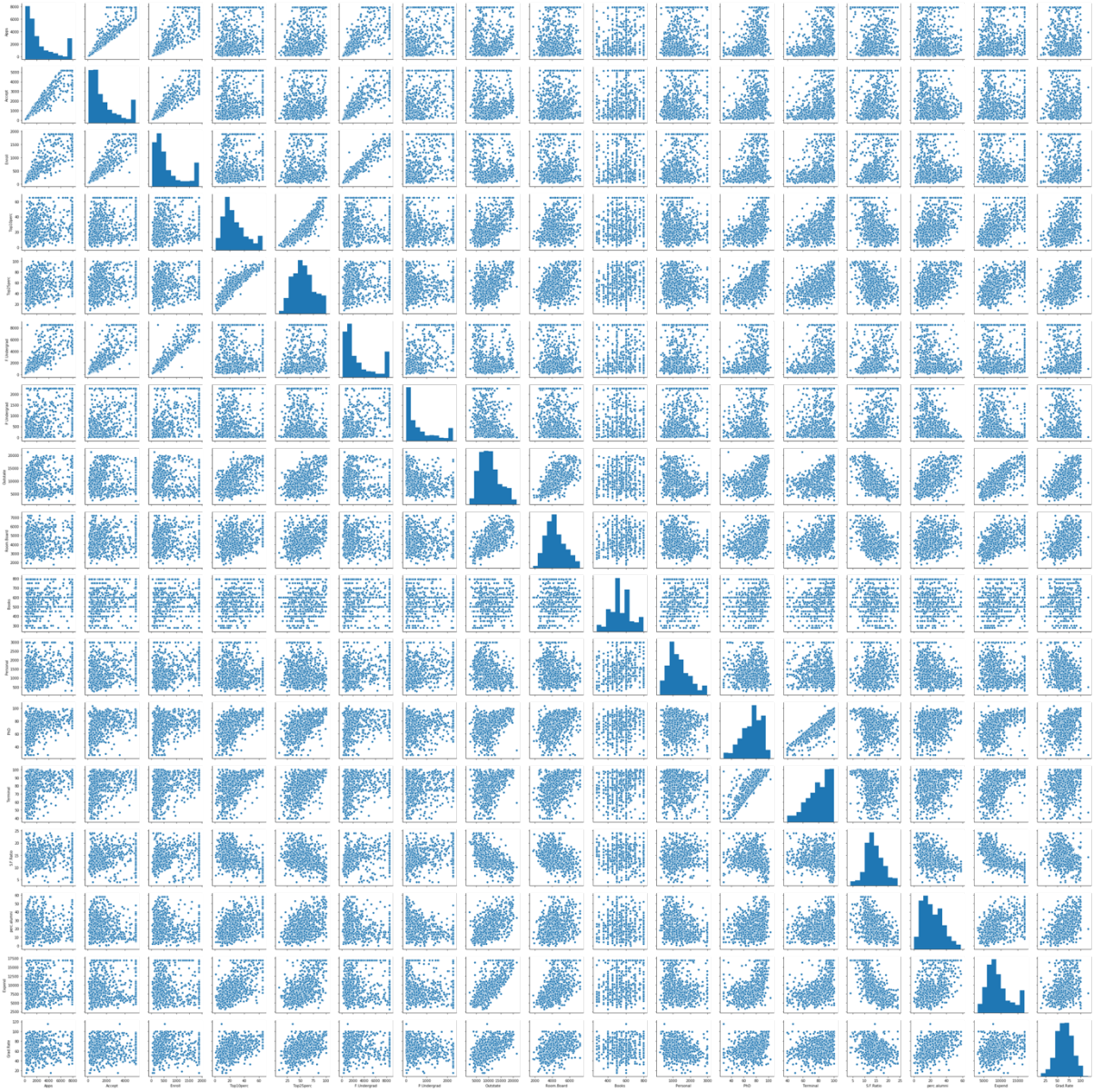


Information Classification: General



From above plots we can easily identify that the outliers are there in almost all the fields.

Bivariate Analysis:



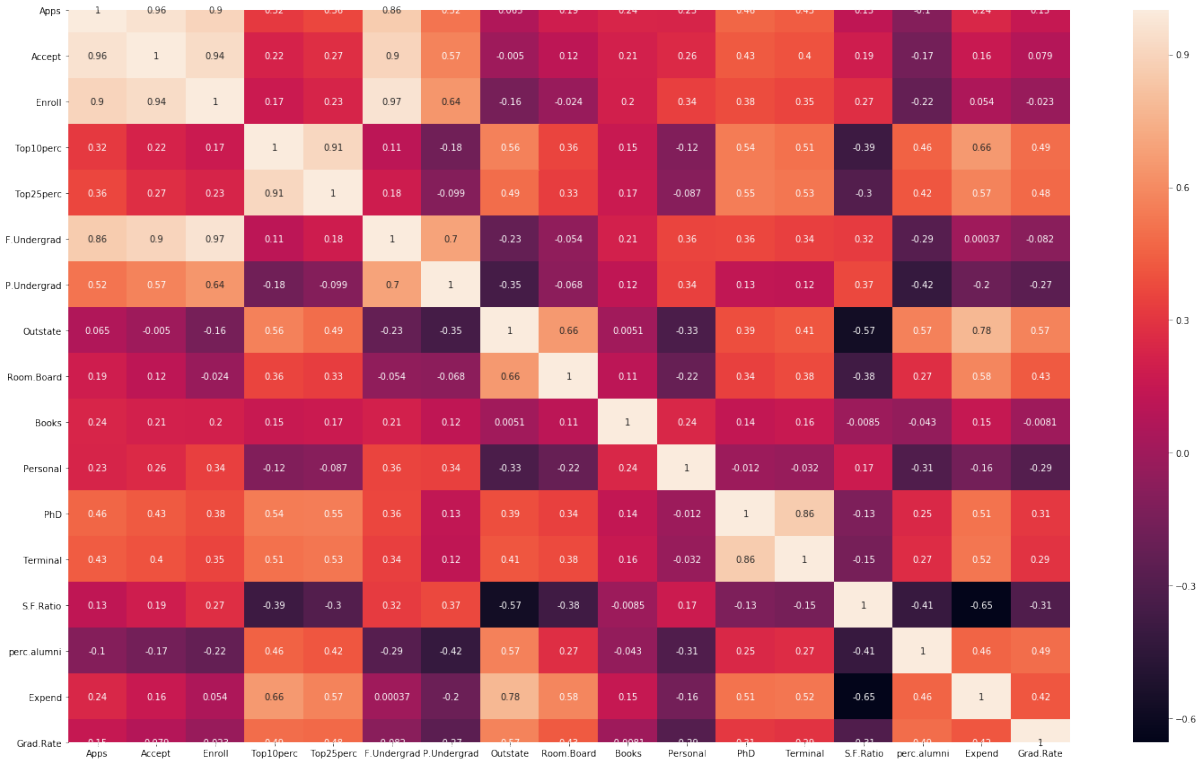
Scale the variables:

Let us the scale the value to bring all the fields into common scale we are going to use STANDARDSCALER from sklearn package.
The standardscaler standardize features by removing the mean and scaling to unit variance.

The standard score of a sample x is calculated as:
z = (x - u) / s
where u is the mean of the training samples or zero if with_mean=False, and s is the standard deviation of the training samples or one if with_std=False.

Centring and scaling happen independently on each feature by computing the relevant statistics on the samples in the training set. Mean and standard deviation are then stored to be used on later data using transform.
Standardization of a dataset is a common requirement for many machine learning estimators: they might behave badly if the individual features do not more or less look like standard normally distributed data. *In our dataset most of the data's are not normally distributed we need apply Scaling*.

We have generated the heatmap with correlation between values. And we can able to see there are few values corelated like Application and acceptance and lot of candidates are enrolling for Full time Under Graduation



And We can also see Expenditure is least corelated with Student/Faculty Ratio.

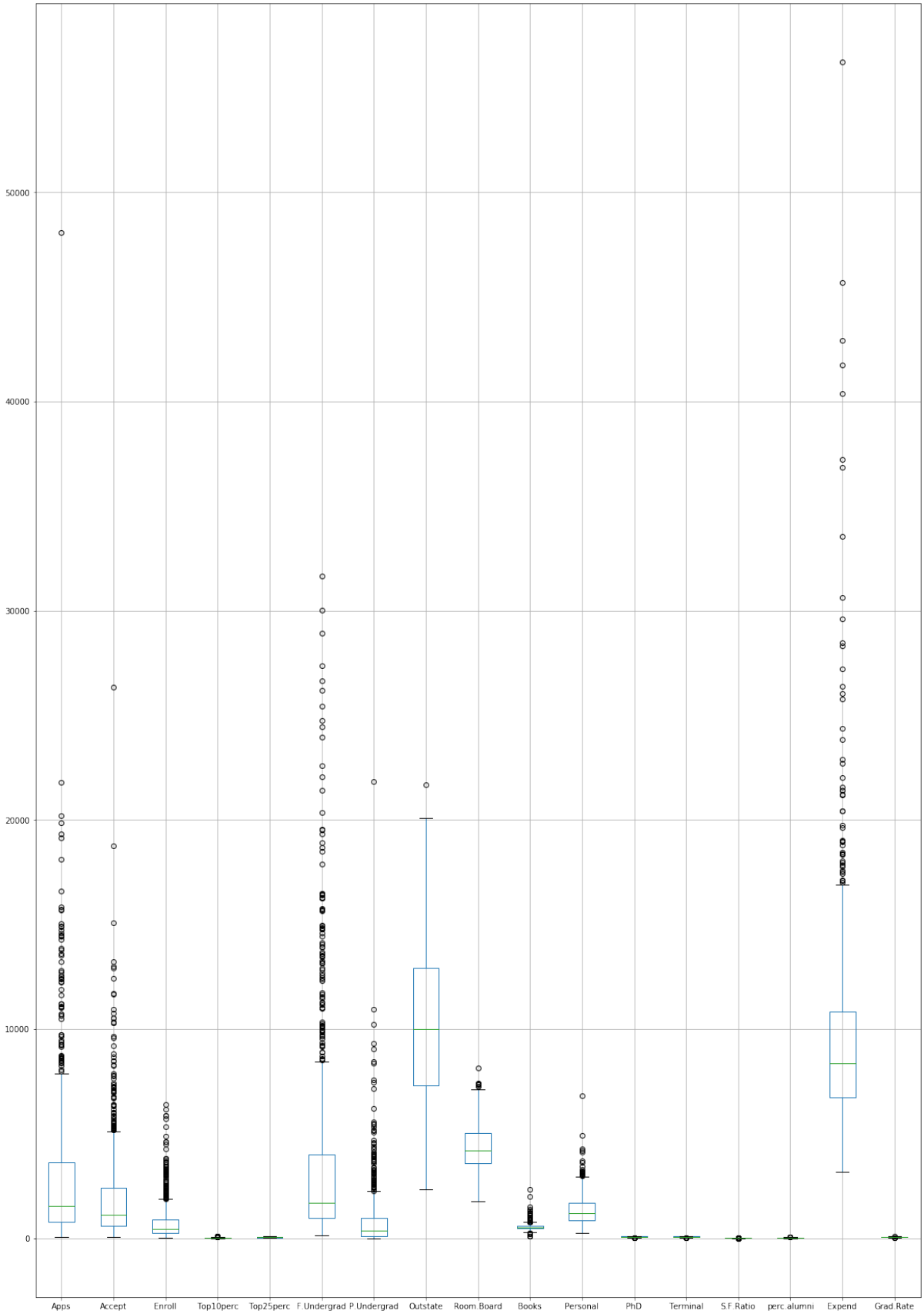
Comparison between covariance and the correlation matrix:

The correlation matrix values will not change even when we scale the values on the other hand the value of covariance is affected by the change in scale of the variables. This is because we divide the value of covariance by the product of standard deviations which have the same units. Another difference between covariance and correlation is the range of values that they can assume. While correlation coefficients lie between -1 and +1, covariance can take any value between -∞ and +∞.
"Covariance" indicates the direction of the linear relationship between variables. "Correlation" on the other hand measures both the strength and direction of the linear relationship between two variables. Correlation is a function of the covariance.

You can obtain the correlation coefficient of two variables by dividing the covariance of these variables by the product of the standard deviations of the same values.

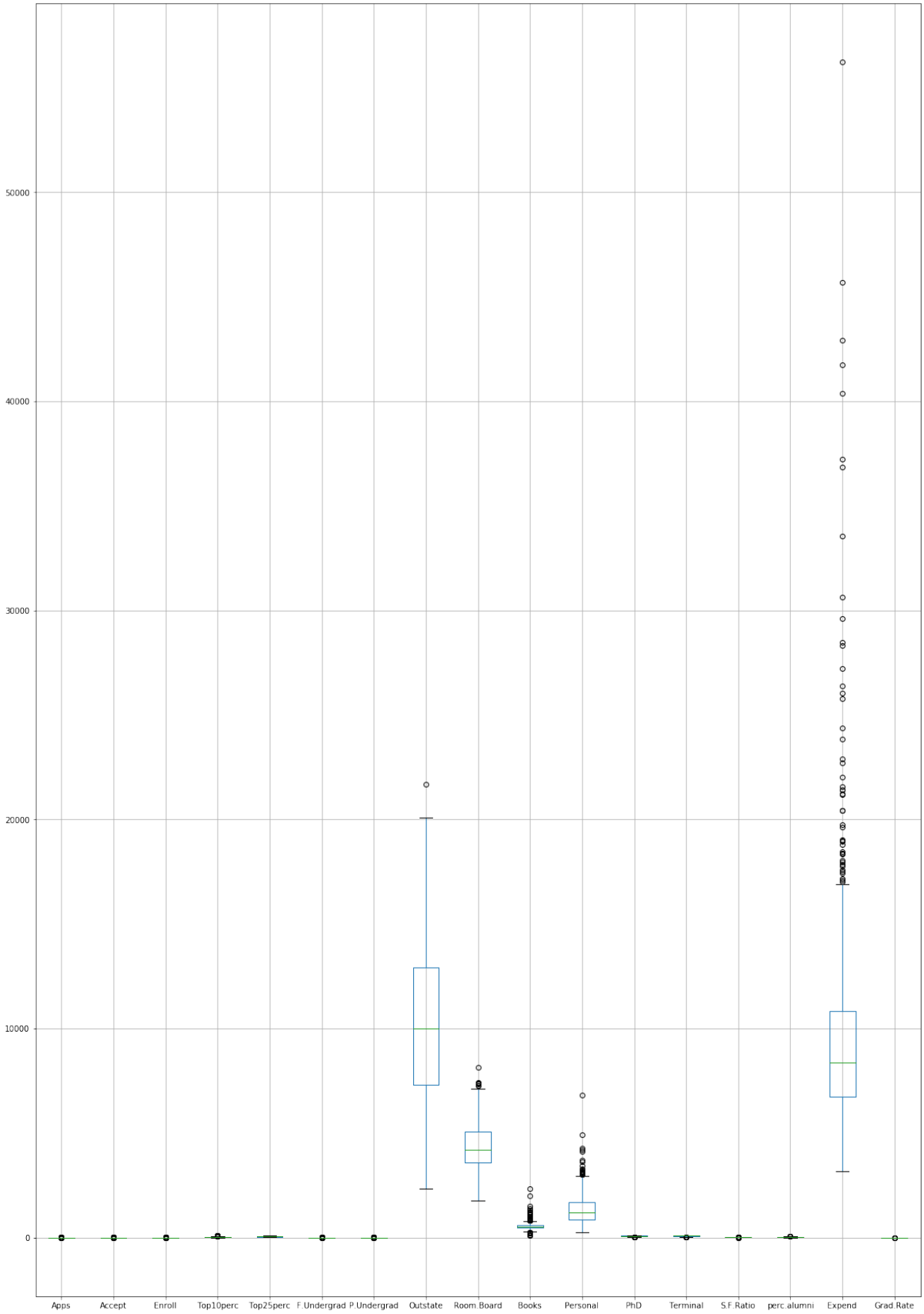
Check the dataset for outliers before and after scaling

Before Scaling



After Scaling

Information Classification: General



Information Classification: General

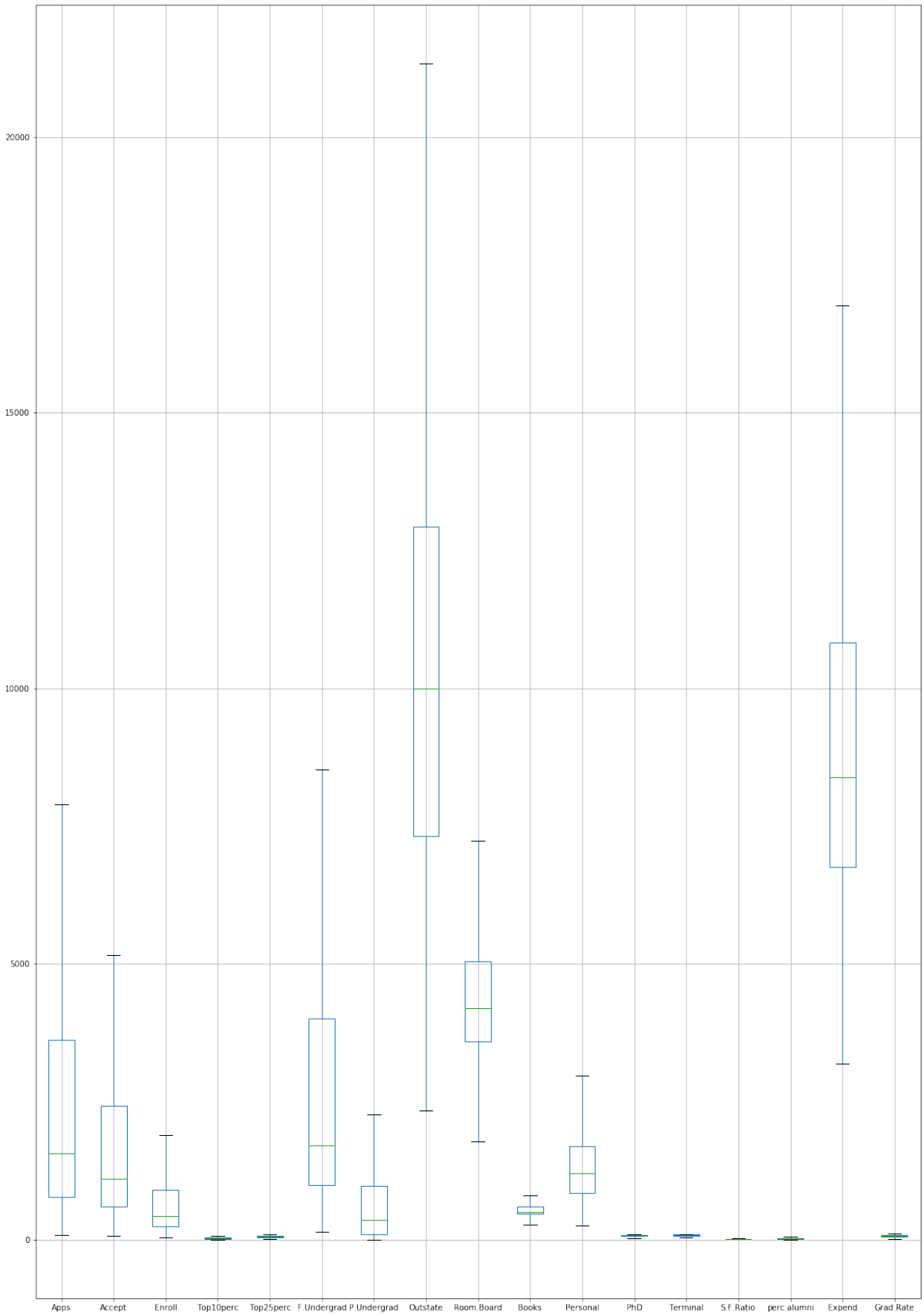
When we see the data we clearly see some outliers are still there in Expenditure, Books, Personal, Room Board and Out state columns lets go ahead and remove them. The approach we going to use is find the Inter Quartile Range (IQR) then with the help of 25th and 75th percentile we can use the use these formulas to find the upper and lower range

lower_range= Q1-(1.5 * IQR)

upper_range= Q3+(1.5 * IQR)

Where Q1 is 25th percentile and Q3 is 75th percentile. One important thing to note is we need to sort the columns before doing this. After removing the outliers using the above approach let's look at the visual representation of it.

Information Classification: General



Information Classification: General

Build the covariance matrix, eigenvalues and eigenvector

```
Covariance Matrix
%s [[ 1.00128866  0.94466636  0.84791332  0.33927032  0.35209304  0.81554018
   0.3987775   0.05022367  0.16515151  0.13272942  0.17896117  0.39120081
   0.36996762  0.09575627 -0.09034216  0.2599265   0.14694372]
 [ 0.94466636  1.00128866  0.91281145  0.19269493  0.24779465  0.87534985
   0.44183938 -0.02578774  0.09101577  0.11367165  0.20124767  0.35621633
   0.3380184   0.17645611 -0.16019604  0.12487773  0.06739929]
 [ 0.84791332  0.91281145  1.00128866  0.18152715  0.2270373   0.96588274
   0.51372977 -0.1556777  -0.04028353  0.11285614  0.28129148  0.33189629
   0.30867133  0.23757707 -0.18102711  0.06425192 -0.02236983]
 [ 0.33927032  0.19269493  0.18152715  1.00128866  0.89314445  0.1414708
  -0.10549205  0.5630552   0.37195909  0.1190116  -0.09343665  0.53251337
   0.49176793 -0.38537048  0.45607223  0.6617651   0.49562711]
 [ 0.35209304  0.24779465  0.2270373   0.89314445  1.00128866  0.19970167
  -0.05364569  0.49002449  0.33191707  0.115676   -0.08091441  0.54656564
   0.52542506 -0.29500852  0.41840277  0.52812713  0.47789622]
 [ 0.81554018  0.87534985  0.96588274  0.1414708   0.19970167  1.00128866
   0.57124738 -0.21602002 -0.06897917  0.11569867  0.31760831  0.3187472
   0.30040557  0.28006379 -0.22975792  0.01867565 -0.07887464]
 [ 0.3987775   0.44183938  0.51372977 -0.10549205 -0.05364569  0.57124738
   1.00128866 -0.25383901 -0.06140453  0.08130416  0.32029384  0.14930637
   0.14208644  0.23283016 -0.28115421 -0.08367612 -0.25733218]
 [ 0.05022367 -0.02578774 -0.1556777   0.5630552   0.49002449 -0.21602002
  -0.25383901  1.00128866  0.65509951  0.03890494 -0.29947232  0.38347594
   0.40850895 -0.55553625  0.56699214  0.6736456   0.57202613]
 [ 0.16515151  0.09101577 -0.04028353  0.37195909  0.33191707 -0.06897917
  -0.06140453  0.65509951  1.00128866  0.12812787 -0.19968518  0.32962651
   0.3750222  -0.36309504  0.27271444  0.50238599  0.42548915]
 [ 0.13272942  0.11367165  0.11285614  0.1190116   0.115676    0.11569867
   0.08130416  0.03890494  0.12812787  1.00128866  0.17952581  0.0269404
   0.10008351 -0.03197042 -0.04025955  0.11255393  0.00106226]
 [ 0.17896117  0.20124767  0.28129148 -0.09343665 -0.08091441  0.31760831
   0.32029384 -0.29947232 -0.19968518  0.17952581  1.00128866 -0.01094989
  -0.03065256  0.13652054 -0.2863366  -0.09801804 -0.26969106]
 [ 0.39120081  0.35621633  0.33189629  0.53251337  0.54656564  0.3187472
   0.14930637  0.38347594  0.32962651  0.0269404  -0.01094989  1.00128866
   0.85068186 -0.13069832  0.24932955  0.43331936  0.30543094]
 [ 0.36996762  0.3380184   0.30867133  0.49176793  0.52542506  0.30040557
   0.14208644  0.40850895  0.3750222   0.10008351 -0.03065256  0.85068186
   1.00128866 -0.16031027  0.26747453  0.43936469  0.28990033]
 [ 0.09575627  0.17645611  0.23757707 -0.38537048 -0.29500852  0.28006379
   0.23283016 -0.55553625 -0.36309504 -0.03197042  0.13652054 -0.13069832
  -0.16031027  1.00128866 -0.4034484  -0.5845844  -0.30710565]
 [-0.09034216 -0.16019604 -0.18102711  0.45607223  0.41840277 -0.22975792
  -0.28115421  0.56699214  0.27271444 -0.04025955 -0.2863366   0.24932955
   0.26747453 -0.4034484   1.00128866  0.41825001  0.49153016]
 [ 0.2599265   0.12487773  0.06425192  0.6617651   0.52812713  0.01867565
  -0.08367612  0.6736456   0.50238599  0.11255393 -0.09801804  0.43331936
   0.43936469 -0.5845844   0.41825001  1.00128866  0.39084571]
 [ 0.14694372  0.06739929 -0.02236983  0.49562711  0.47789622 -0.07887464
  -0.25733218  0.57202613  0.42548915  0.00106226 -0.26969106  0.30543094
   0.28990033 -0.30710565  0.49153016  0.39084571  1.00128866]]
```

```
 Eigen Values
 %s [5.45052162 4.48360686 1.17466761 1.00820573 0.93423123 0.8484
 0.6057878  0.58787222 0.53061262 0.4043029  0.02302787 0.03672545
 0.31344588 0.08802464 0.1439785  0.16779415 0.22061096]
Eigen Vectors
 %s [[-2.48765602e-01  3.31598227e-01  6.30921033e-02 -2.81310530e
   5.74140964e-03  1.62374420e-02  4.24863486e-02  1.03090398e-01
   9.02270802e-02 -5.25098025e-02  3.58970400e-01 -4.59139498e-01
   4.30462074e-02 -1.33405806e-01  8.06328039e-02 -5.95830975e-01
   2.40709086e-02]
 [-2.07601502e-01  3.72116750e-01  1.01249056e-01 -2.67817346e-01
   5.57860920e-02 -7.53468452e-03  1.29497196e-02  5.62709623e-02
   1.77864814e-01 -4.11400844e-02 -5.43427250e-01  5.18568789e-01
  -5.84055850e-02  1.45497511e-01  3.34674281e-02 -2.92642398e-01
  -1.45102446e-01]
 [-1.76303592e-01  4.03724252e-01  8.29855709e-02 -1.61826771e-01
  -5.56936353e-02  4.25579803e-02  2.76928937e-02 -5.86623552e-02
   1.28560713e-01 -3.44879147e-02  6.09651110e-01  4.04318439e-01
  -6.93988831e-02 -2.95896092e-02 -8.56967180e-02  4.44638207e-01
   1.11431545e-02]
 [-3.54273947e-01 -8.24118211e-02 -3.50555339e-02  5.15472524e-02
  -3.95434345e-01  5.26927980e-02  1.61332069e-01  1.22678028e-01
  -3.41099863e-01 -6.40257785e-02 -1.44986329e-01  1.48738723e-01
  -8.10481404e-03 -6.97722522e-01 -1.07828189e-01 -1.02303616e-03
   3.85543001e-02]
 [-3.44001279e-01 -4.47786551e-02  2.41479376e-02  1.09766541e-01
  -4.26533594e-01 -3.30915896e-02  1.18485556e-01  1.02491967e-01
  -4.03711989e-01 -1.45492289e-02  8.03478445e-02 -5.18683400e-02
  -2.73128469e-01  6.17274818e-01  1.51742110e-01 -2.18838802e-02
  -8.93515563e-02]
 [-1.54640962e-01  4.17673774e-01  6.13929764e-02 -1.00412335e-01
  -4.34543659e-02  4.34542349e-02  2.50763629e-02 -7.88896442e-02
   5.94419181e-02 -2.08471834e-02 -4.14705279e-01 -5.60363054e-01
  -8.11578181e-02 -9.91640992e-03 -5.63728817e-02  5.23622267e-01
   5.61767721e-02]
 [-2.64425045e-02  3.15087830e-01 -1.39681716e-01  1.58558487e-01
   3.02385408e-01  1.91198583e-01 -6.10423460e-02 -5.70783816e-01
  -5.60672902e-01  2.23105808e-01  9.01788964e-03  5.27313042e-02
   1.00693324e-01 -2.09515982e-02  1.92857500e-02 -1.25997650e-01
  -6.35360730e-02]
 [-2.94736419e-01 -2.49643522e-01 -4.65988731e-02 -1.31291364e-01
   2.22532003e-01  3.00003910e-02 -1.08528966e-01 -9.84599754e-03
   4.57332880e-03 -1.86675363e-01  5.08995918e-02 -1.01594830e-01
   1.43220673e-01 -3.83544794e-02 -3.40115407e-02  1.41856014e-01
  -8.23443779e-01]
 [-2.49030449e-01 -1.37808883e-01 -1.48967389e-01 -1.84995991e-01
   5.60919470e-01 -1.62755446e-01 -2.09744235e-01  2.21453442e-01
  -2.75022548e-01 -2.98324237e-01  1.14639620e-03  2.59293381e-02
  -3.59321731e-01 -3.40197083e-03 -5.84289756e-02  6.97485854e-02
   3.54559731e-01]
 [-6.47575181e-02  5.63418434e-02 -6.77411649e-01 -8.70892205e-02
  -1.27288825e-01 -6.41054950e-01  1.49692034e-01 -2.13293009e-01
   1.33663353e-01  8.20292186e-02  7.72631963e-04 -2.88282896e-03
   3.19400370e-02  9.43887925e-03 -6.68494643e-02 -1.14379958e-02
  -2.81593679e-02]
 [ 4.25285386e-02  2.19929218e-01 -4.99721120e-01  2.30710568e-01
  -2.22311021e-01  3.31398003e-01 -6.33790064e-01  2.32660840e-01
   9.44688900e-02 -1.36027616e-01 -1.11433396e-03  1.28904022e-02
  -1.85784733e-02  3.09001353e-03  2.75286207e-02 -3.94547417e-02
  -3.92640266e-02]
 [-3.18312875e-01  5.83113174e-02  1.27028371e-01  5.34724832e-01
   1.40166326e-01 -9.12555212e-02  1.09641298e-03  7.70400002e-02
   1.85181525e-01  1.23452200e-01  1.38133366e-02 -2.98075465e-02
   4.03723253e-02  1.12055599e-01 -6.91126145e-01 -1.27696382e-01
   2.32224316e-02]
 [-3.17056016e-01  4.64294477e-02  6.60375454e-02  5.19443019e-01
   2.04719730e-01 -1.54927646e-01  2.84770105e-02  1.21613297e-02
   2.54938198e-01  8.85784627e-02  6.20932749e-03  2.70759809e-02
  -5.89734026e-02 -1.58909651e-01  6.71008607e-01  5.83134662e-02
   1.64850420e-02]
 [ 1.76957895e-01  2.46665277e-01  2.89848401e-01  1.61189487e-01
  -7.93882496e-02 -4.87045875e-01 -2.19259358e-01  8.36048735e-02
  -2.74544380e-01 -4.72045249e-01 -2.22215182e-03  2.12476294e-02
   4.45000727e-01  2.08991284e-02  4.13740967e-02  1.77152700e-02
  -1.10262122e-02]
 [-2.05082369e-01 -2.46595274e-01  1.46989274e-01 -1.73142230e-02
  -2.16297411e-01  4.73400144e-02 -2.43321156e-01 -6.78523654e-01
   2.55334907e-01 -4.22999706e-01 -1.91869743e-02 -3.33406243e-03
  -1.30727978e-01  8.41789410e-03 -2.71542091e-02 -1.04088088e-01
   1.82660654e-01]
 [-3.18908750e-01 -1.31689865e-01 -2.26743985e-01 -7.92734946e-02
   7.59581203e-02  2.98118619e-01  2.26584481e-01  5.41593771e-02
   4.91388809e-02 -1.32286331e-01 -3.53098218e-02  4.38803230e-02
   6.92088870e-01  2.27742017e-01  7.31225166e-02  9.37464497e-02
   3.25982295e-01]
 [-2.52315654e-01 -1.69240532e-01  2.08064649e-01 -2.69129066e-01
  -1.09267913e-01 -2.16163313e-01 -5.59943937e-01  5.33553891e-03
  -4.19043052e-02  5.90271067e-01 -1.30710024e-02  5.00844705e-03
   2.19839000e-01  3.39433604e-03  3.64767385e-02  6.91969778e-02
   1.22106697e-01]]
```

Information Classification: General

Explicit form of the first PC:

The first principal component, i.e. the eigenvector which goes the largest value of λ, is the direction along which the data have the most variance

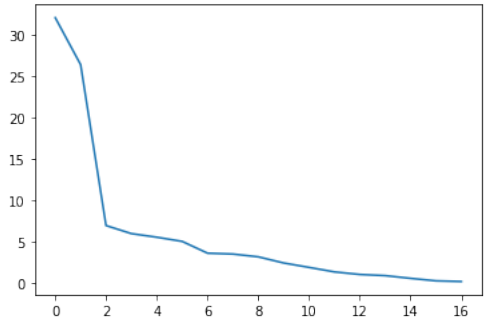
Cumulative values of the eigenvalues:

There are 4 Eigen Values whose values are greater than 1. Now we will derive cumulative distribution the result is

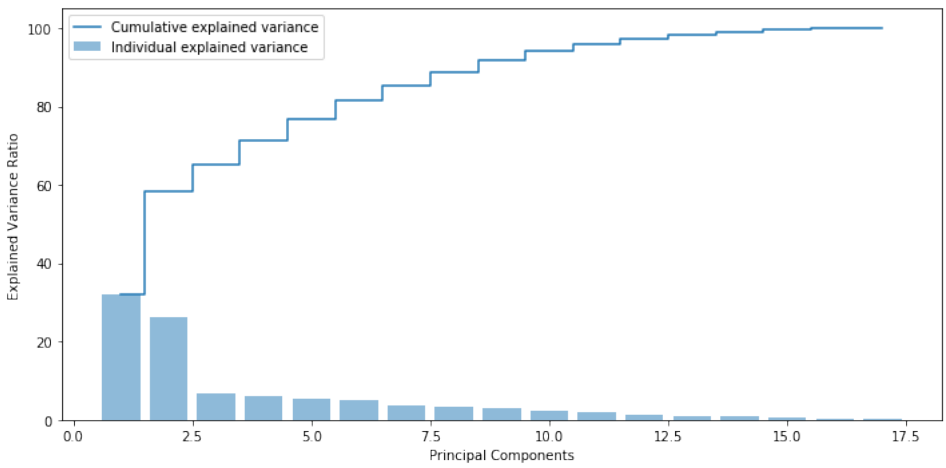
```
Cumulative Variance Explained [ 32.0206282   58.36084263  65.26175919  71.18474841  76.67315352
  81.65785448  85.21672597  88.67034731  91.78758099  94.16277251
  96.00419883  97.30024023  98.28599436  99.13183669  99.64896227
  99.86471628 100.        ]
```

With the result we can say that with 76components we can build more than 80% of the data. Let's visualise this

spree plot.



Visually we can observe that there is steep drop in variance explained with increase in number of PC's. Now we will proceed with 6 components here we can increase or decrease based on the requirements

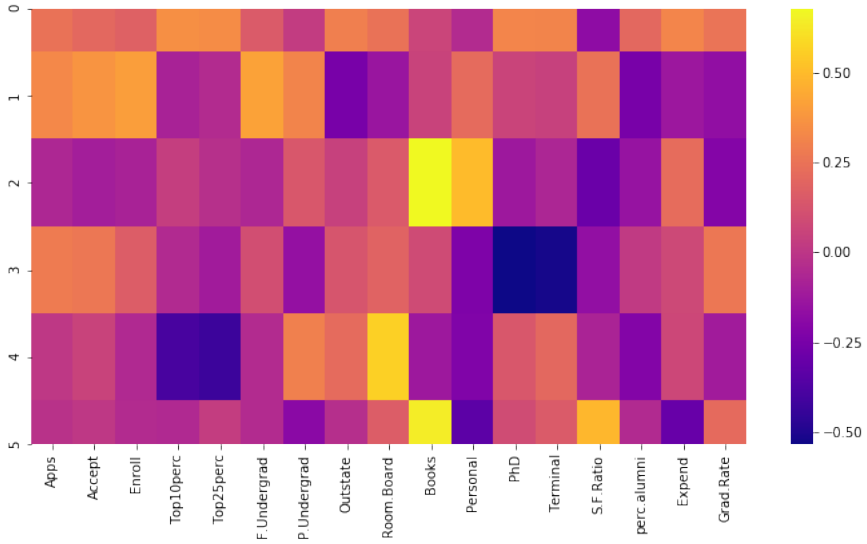


Perform PCA and export the data of the Principal Component scores into a data frame

After performing PCA and exporting the data in to a Data Frame the snapshot of the data is available below

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal
0	0.248766	0.207602	0.176304	0.354274	0.344001	0.154641	0.026443	0.294736	0.249030	0.064758	-0.042529	0.318313	0.317056
1	0.331598	0.372117	0.403724	-0.082412	-0.044779	0.417674	0.315088	-0.249644	-0.137809	0.056342	0.219929	0.058311	0.046429
2	-0.063092	-0.101249	-0.082986	0.035056	-0.024148	-0.061393	0.139682	0.046599	0.148967	0.677412	0.499721	-0.127028	-0.066038
3	0.281311	0.267817	0.161827	-0.051547	-0.109767	0.100412	-0.158558	0.131291	0.184996	0.087089	-0.230711	-0.534725	-0.519443
4	0.005741	0.055786	-0.055694	-0.395434	-0.426534	-0.043454	0.302385	0.222532	0.560919	-0.127289	-0.222311	0.140166	0.204720

With above data we have derived a heatmap to view the correlation



Business implication of using the Principal Component Analysis for this case study:

PCA is a statistical technique that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables.

The business implication for using PCA is we can able to reduce the dimension to analyse the data. In our case we have 18 columns to analyse the data that is 18 dimensions which will be complicated. To reduce complexity we have applied PCA and reduced the dimension to cover the 80% of the data.