Problem 2 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 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.

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

Data set :

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
Abilene Christian University	1660	1232	721	23	52	2885	537	7440	3300	450	2200	70	78	18.1	12	7041	60
Adelphi University	2186	1924	512	16	29	2683	1227	12280	6450	750	1500	29	30	12.2	16	10527	56
Adrian College	1428	1097	336	22	50	1036	99	11250	3750	400	1165	53	66	12.9	30	8735	54
Agnes Scott College	417	349	137	60	89	510	63	12960	5450	450	875	92	97	7.7	37	19016	59
Alaska Pacific University	193	146	55	16	44	249	869	7560	4120	800	1500	76	72	11.9	2	10922	15
Albertson College	587	479	158	38	62	678	41	13500	3335	500	675	67	73	9.4	11	9727	55
Albertus Magnus College	353	340	103	17	45	416	230	13290	5720	500	1500	90	93	11.5	26	8861	63
Albion College	1899	1720	489	37	68	1594	32	13868	4826	450	850	89	100	13.7	37	11487	73
Albright College	1038	839	227	30	63	973	306	15595	4400	300	500	79	84	11.3	23	11644	80
Alderson-Broaddus College	582	498	172	21	44	799	78	10468	3380	660	1800	40	41	11.5	15	8991	52
Alfred University	1732	1425	472	37	75	1830	110	16548	5406	500	600	82	88	11.3	31	10932	73
Allegheny College	2652	1900	484	44	77	1707	44	17080	4440	400	600	73	91	9.9	41	11711	76
Allentown Coll. of St. Francis de Sales	1179	780	290	38	64	1130	638	9690	4785	600	1000	60	84	13.3	21	7940	74
Alma College	1267	1080	385	44	73	1306	28	12572	4552	400	400	79	87	15.3	32	9305	68
Alverno College	494	313	157	23	46	1317	1235	8352	3640	650	2449	36	69	11.1	26	8127	55
American International College	1420	1093	220	9	22	1018	287	8700	4780	450	1400	78	84	14.7	19	7355	69
Amherst College	4302	992	418	83	96	1593	5	19760	5300	660	1598	93	98	8.4	63	21424	100
Anderson University	1216	908	423	19	40	1819	281	10100	3520	550	1100	48	61	12.1	14	7994	59
Andrews University	1130	704	322	14	23	1586	326	9996	3090	900	1320	62	66	11.5	18	10908	46
Angelo State University	3540	2001	1016	24	54	4190	1512	5130	3592	500	2000	60	62	23.1	5	4010	34
Antioch University	713	661	252	25	44	712	23	15476	3336	400	1100	69	82	11.3	35	42926	48
Appalachian State University	7313	4664	1910	20	63	9940	1035	6806	2540	96	2000	83	96	18.3	14	5854	70
Aquinas College	619	516	219	20	51	1251	767	11208	4124	350	1615	55	65	12.7	25	6584	65
Arizona State University Main campus	12809	10308	3761	24	49	22593	7585	7434	4850	700	2100	88	93	18.9	5	4602	48
Arkansas College (Lyon College)	708	334	166	46	74	530	182	8644	3922	500	800	79	88	12.6	24	14579	54

We are provided with the above data set of 777 rows and 18 columns. Of the above columns, one column is float data type , one column is object data type and remaining columns are of integer data type.

There are no Null values in all the columns.

```
 0   Names          777 non-null    object
 1   Apps           777 non-null    int64
 2   Accept         777 non-null    int64
 3   Enroll         777 non-null    int64
 4   Top10perc      777 non-null    int64
 5   Top25perc      777 non-null    int64
 6   F.Undergrad    777 non-null    int64
 7   P.Undergrad    777 non-null    int64
 8   Outstate       777 non-null    int64
 9   Room.Board     777 non-null    int64
 10  Books          777 non-null    int64
 11  Personal       777 non-null    int64
 12  PhD            777 non-null    int64
 13  Terminal       777 non-null    int64
 14  S.F.Ratio      777 non-null    float64
 15  perc.alumni    777 non-null    int64
 16  Expend         777 non-null    int64
 17  Grad.Rate      777 non-null    int64
dtypes: float64(1), int64(16), object(1)
```

Descriptive statistics for the dataset:

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
count	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.000000	777.00000
mean	3001.638353	2018.804376	779.972973	27.558559	55.796654	3699.907336	855.298584	10440.669241	4357.526384	549.380952	1340.642214	72.660232	79.702703	14.089704	22.743887	9660.171171	65.46332
std	3870.201484	2451.113971	929.176190	17.640364	19.804778	4850.420531	1522.431887	4023.016484	1096.696416	165.105360	677.071454	16.328155	14.722359	3.958349	12.391801	5221.768440	17.17771
min	81.000000	72.000000	35.000000	1.000000	9.000000	139.000000	1.000000	2340.000000	1780.000000	96.000000	250.000000	8.000000	24.000000	2.500000	0.000000	3186.000000	10.00000
25%	776.000000	604.000000	242.000000	15.000000	41.000000	992.000000	95.000000	7320.000000	3597.000000	470.000000	850.000000	62.000000	71.000000	11.500000	13.000000	6751.000000	53.00000
50%	1558.000000	1110.000000	434.000000	23.000000	54.000000	1707.000000	353.000000	9990.000000	4200.000000	500.000000	1200.000000	75.000000	82.000000	13.600000	21.000000	8377.000000	65.00000
75%	3624.000000	2424.000000	902.000000	35.000000	69.000000	4005.000000	967.000000	12925.000000	5050.000000	600.000000	1700.000000	85.000000	92.000000	16.500000	31.000000	10830.000000	78.00000
max	48094.000000	26330.000000	6392.000000	96.000000	100.000000	31643.000000	21836.000000	21700.000000	8124.000000	2340.000000	6800.000000	103.000000	100.000000	39.800000	64.000000	56233.000000	118.00000

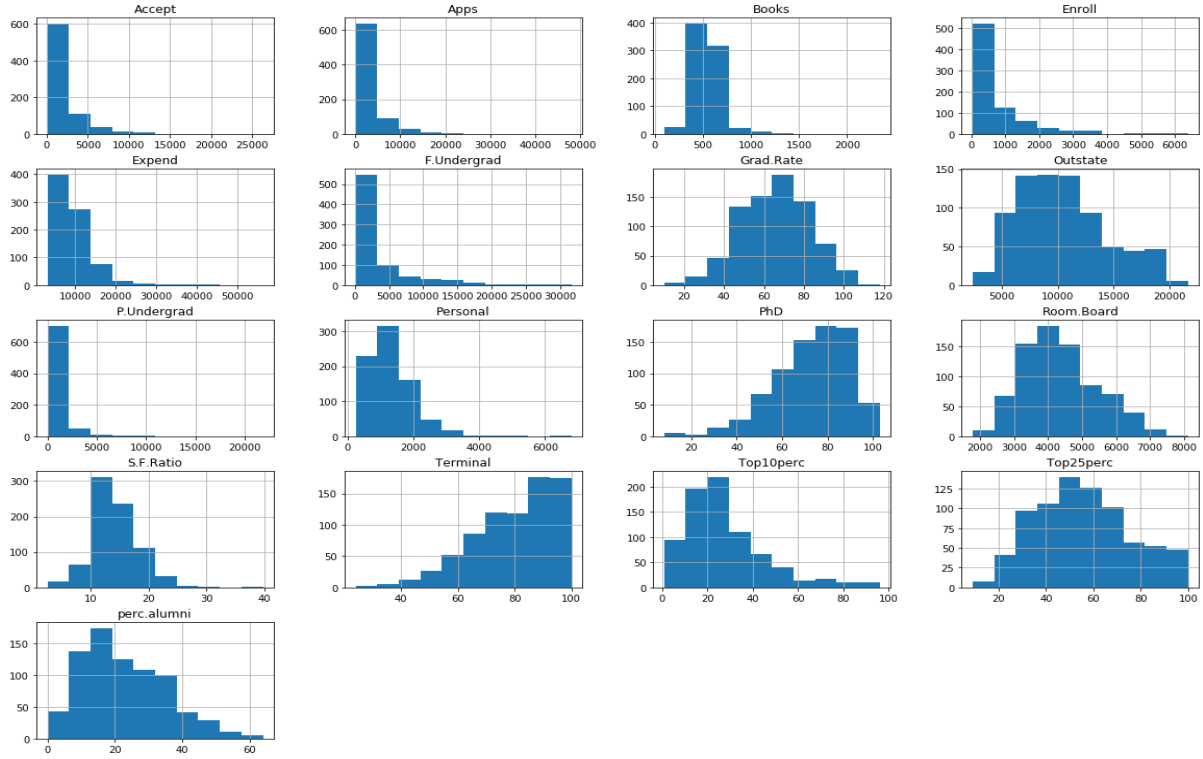
We have Columns **'Names'** as **categorical type** data and column **'S.F.Ratio'** as **Float type** and remaining all columns are **Integer type** data.

As per the details resulted from the descriptive statistics of the dataset, we can find that:

All the variables have a **value count** of **777** with **no null values**.

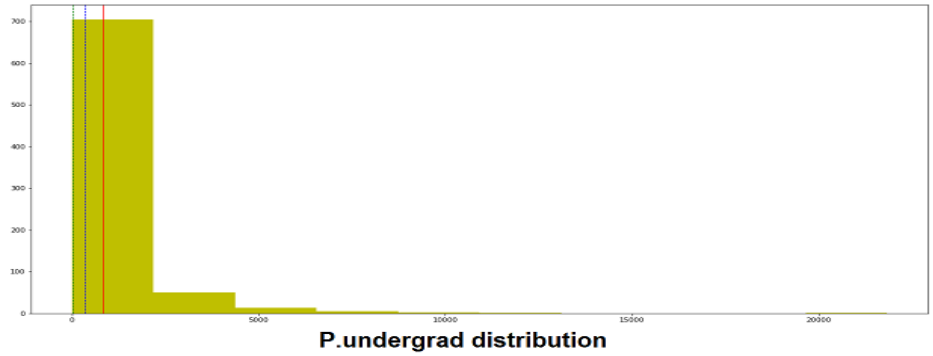
Of the entire dataset, column **'Expand'** has **highest max** value of 56233 and column **'per.alumni'** has **least min** value of 0.00

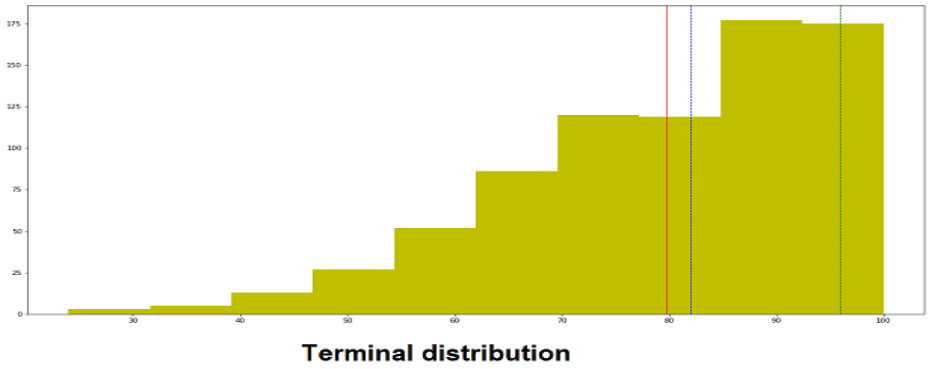
Columns **'Outstate'** and **'S.F.Ratio'** have highest mean value - 10440.669241 and least mean value - 14.089704 respectively.



From the above histograms of the variables , we can see that majority of the variables are not symmetrical.

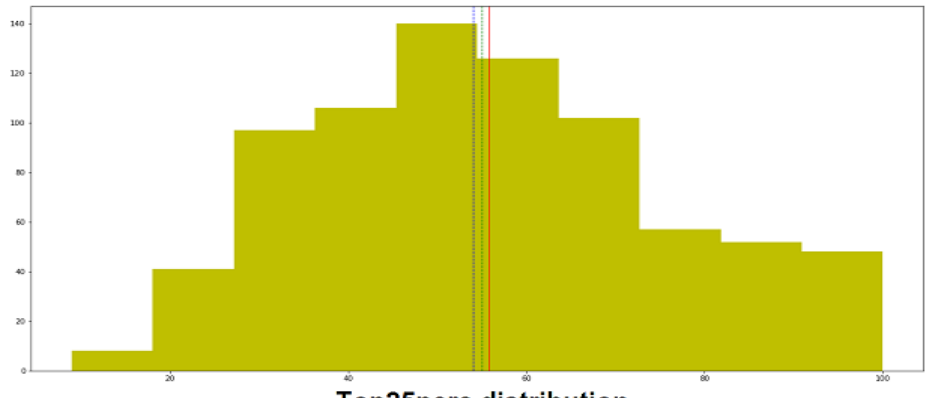
Among all the variables , Variable **'P.Undergrad'** is highly **right skewed** (skew = 5.692353) and Varible **'Terminal'** is highly **left skewed** (skew = -0.816542).

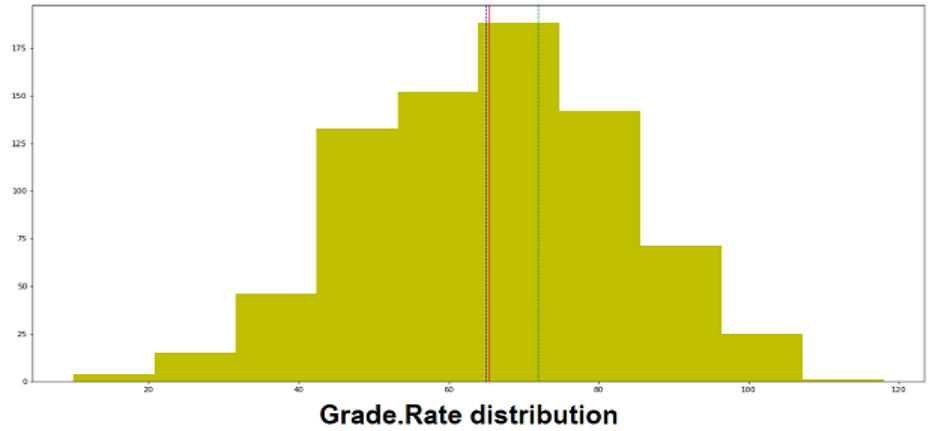




Red = mean
Blue = median
Green = mode

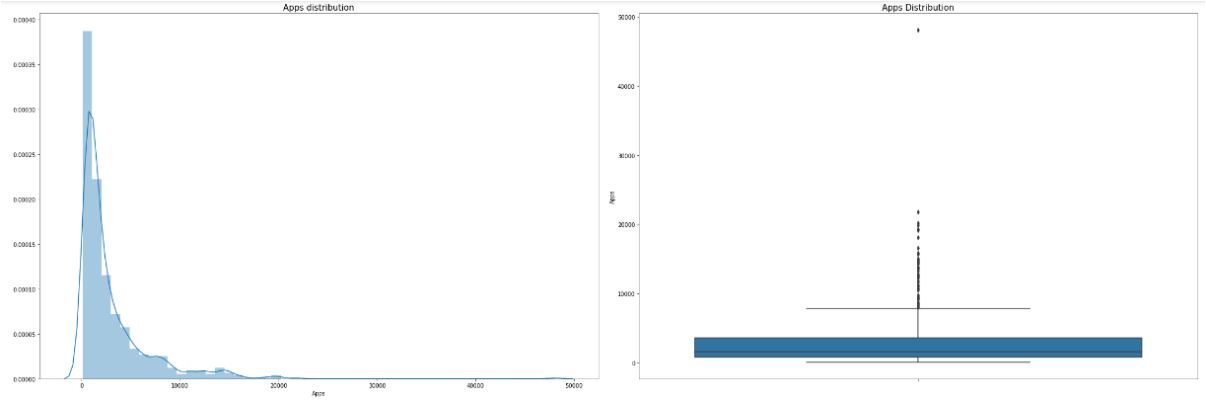
Variables **'Top25perc'** and **'Grad.Rate'** are having almost symmetrical distributions.



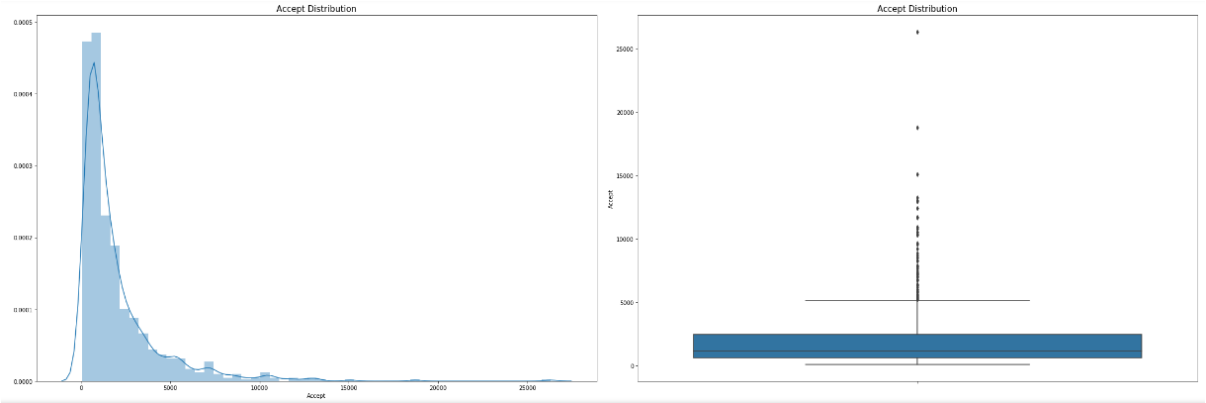


Red = mean
Blue = median
Green = mode

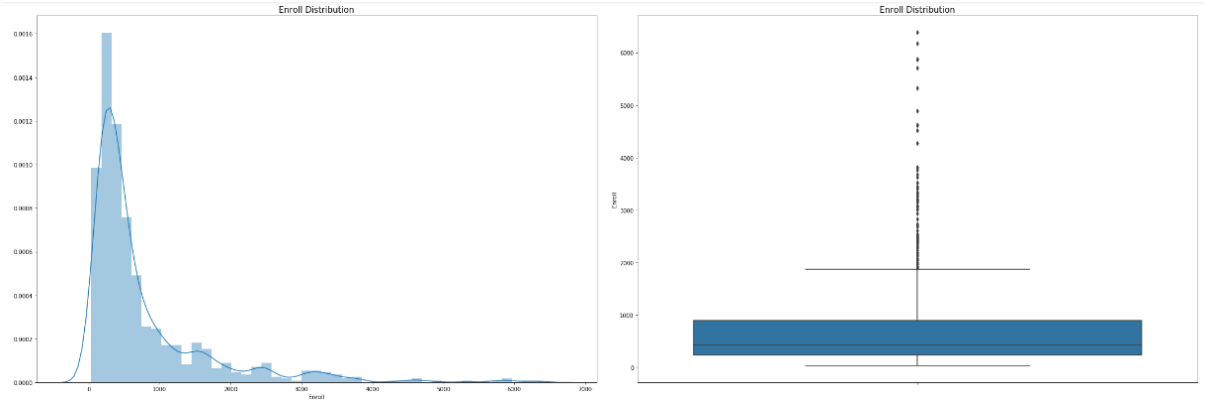
Variable – **'Apps'**
Skew – Positive
Outliers presence – Yes



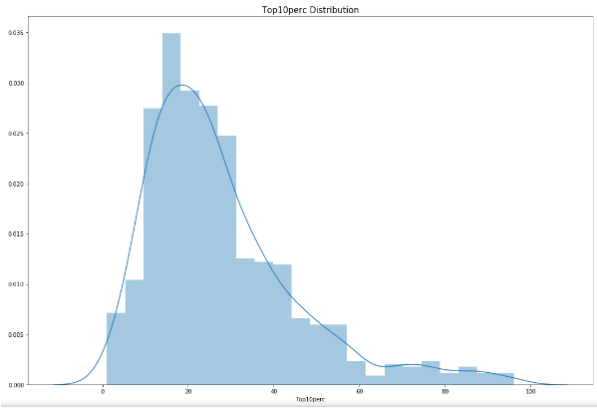
Variable – **'Accept'**
Skew – Positive
Outliers presence – Yes

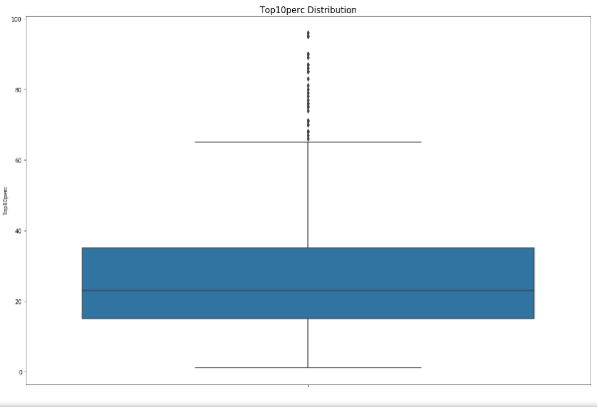


Variable – **'Enrol'**
Skew – Positive
Outliers presence – Yes

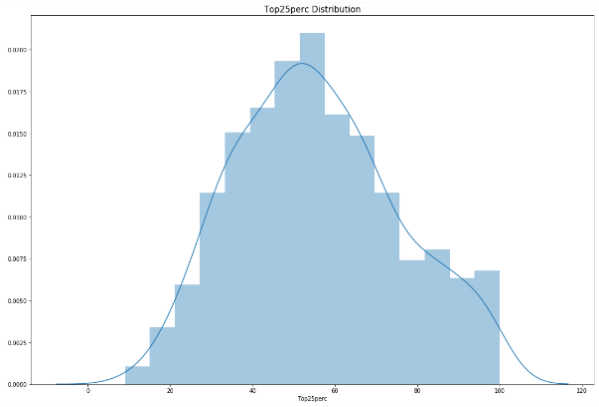


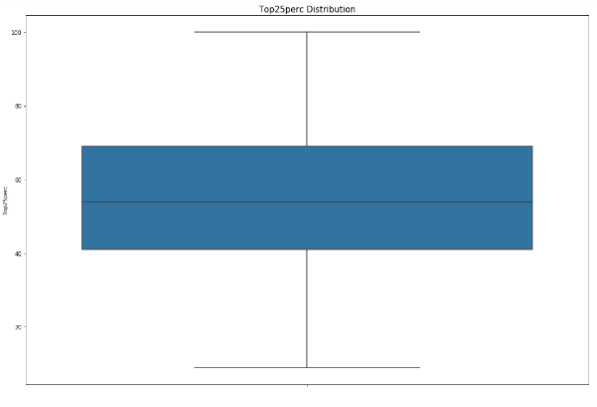
Variable – **'Top10perc'**
Skew – Positive
Outliers presence – Yes



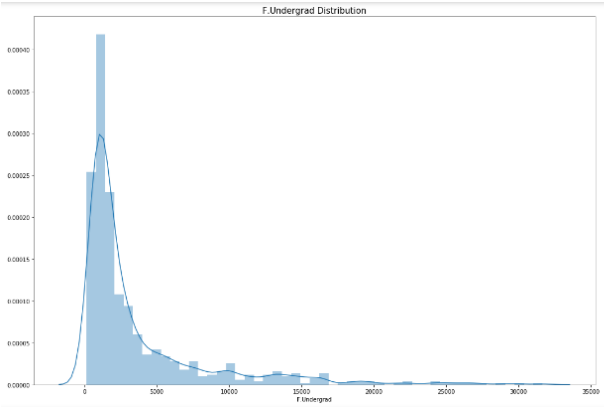


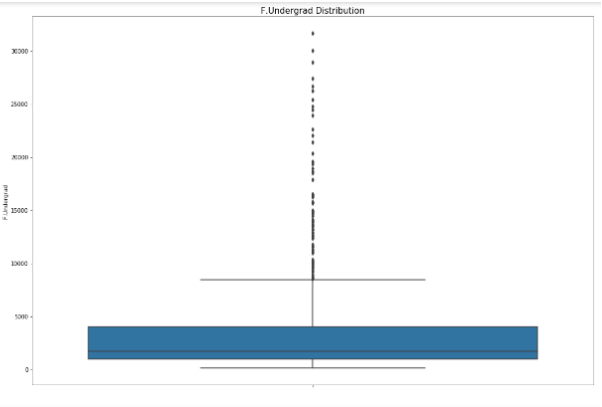
Variable – **'Top10perc'**
Skew – Positive
Outliers presence – No



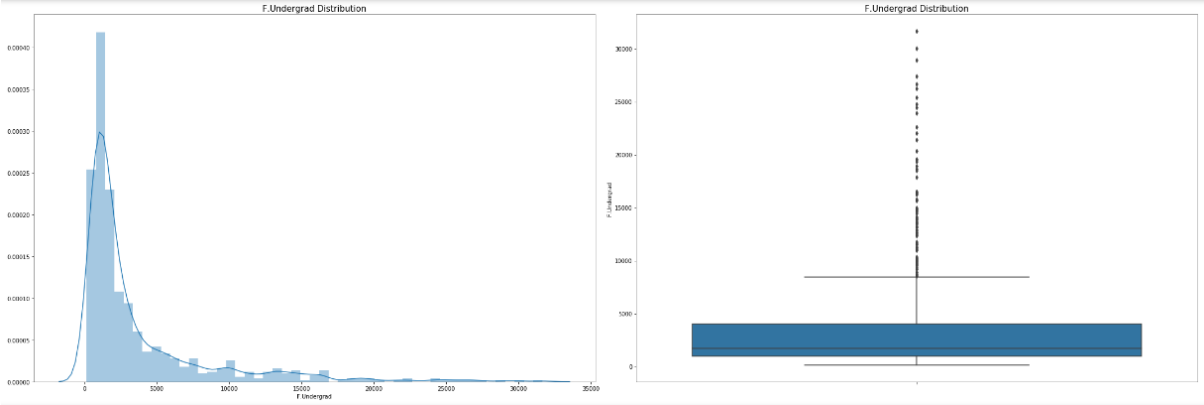


Variable – **'Top10perc'**
Skew – Positive
Outliers presence – Yes

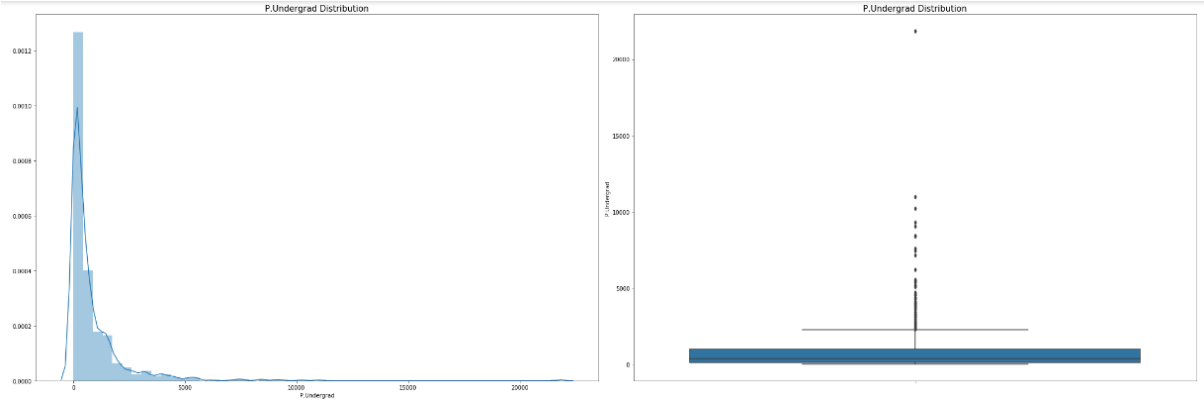




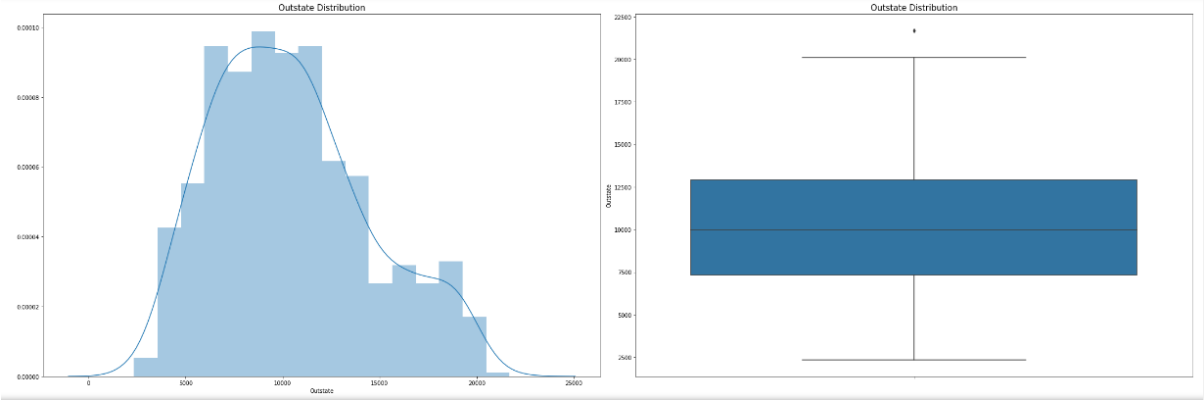
Variable – **'F.UnderGrad'**
Skew – Positive
Outliers presence – Yes



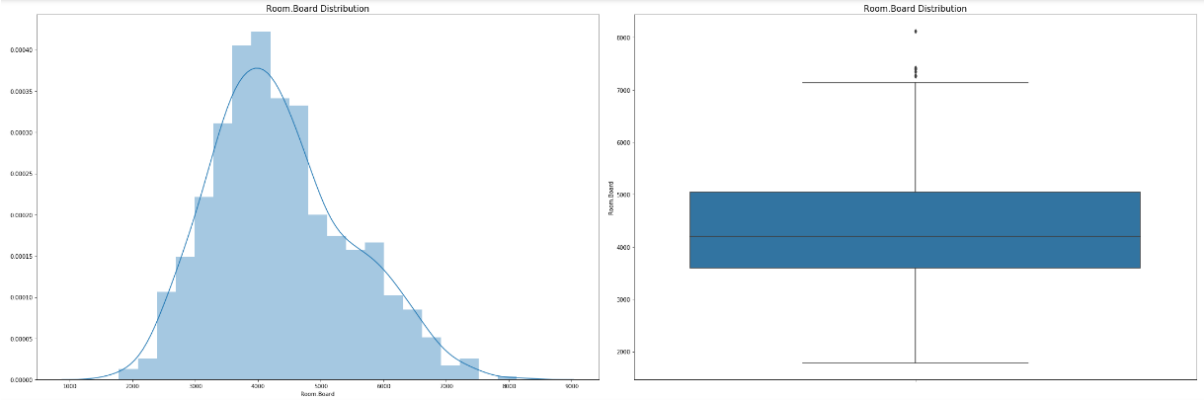
Variable – **'P.Undergrad'**
Skew – Positive
Outliers presence – Yes



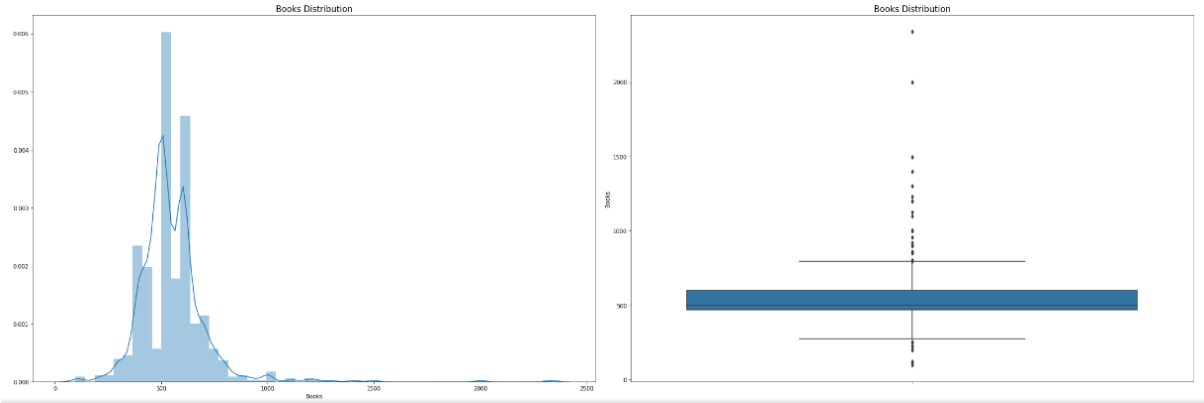
Variable – **'Outstate'**
Skew – Positive
Outliers presence – Yes



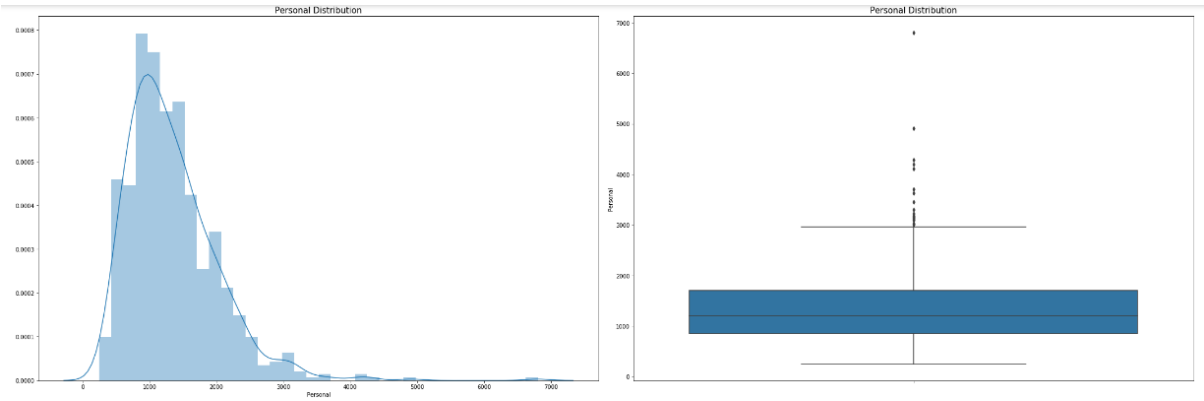
Variable – **'Room.Board'**
Skew – Positive
Outliers presence – Yes



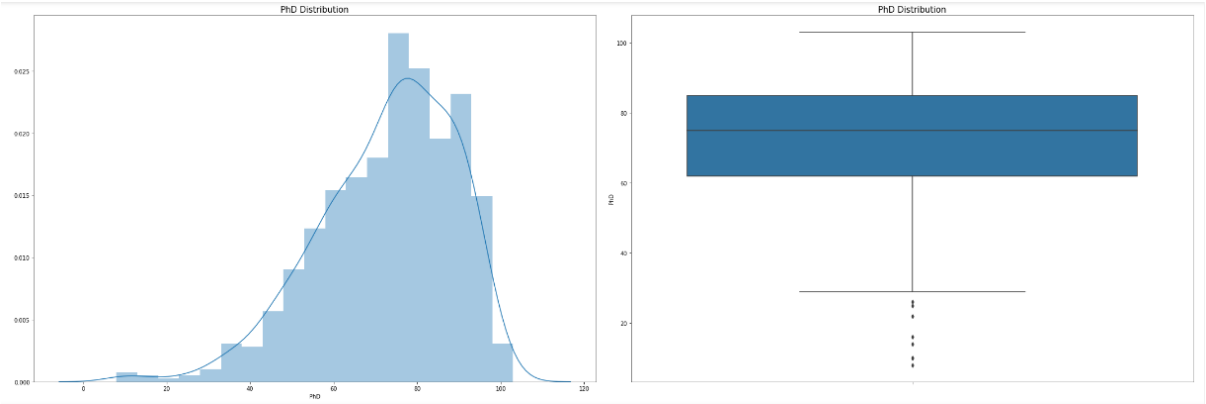
Variable – **'Books'**
Skew – Positive
Outliers presence – Yes



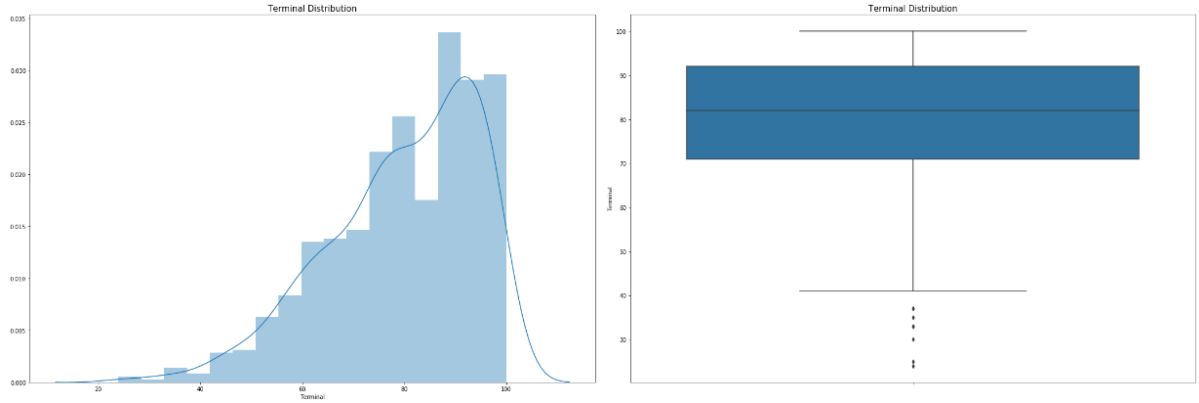
Variable – **'Personal Distribution'**
Skew – Positive
Outliers presence – Yes



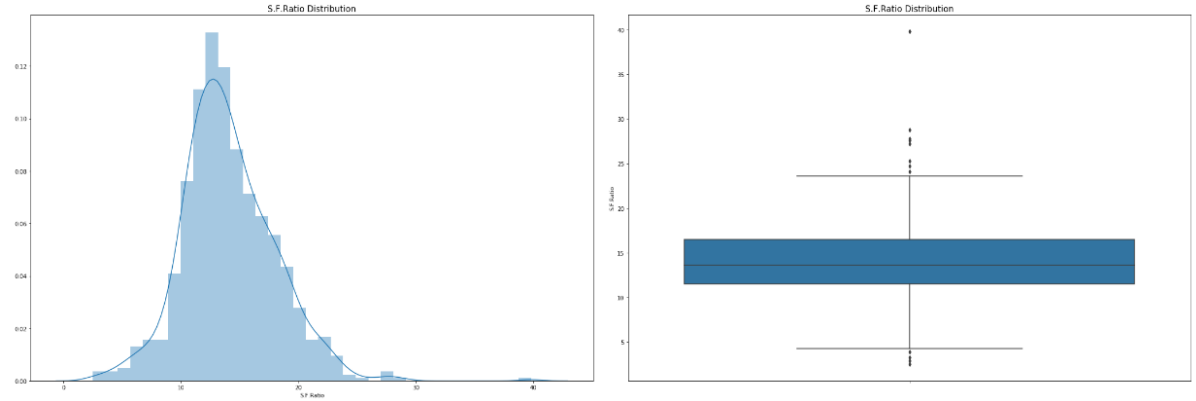
Variable – **'PhD Distribution'**
Skew – Negative
Outliers presence – Yes



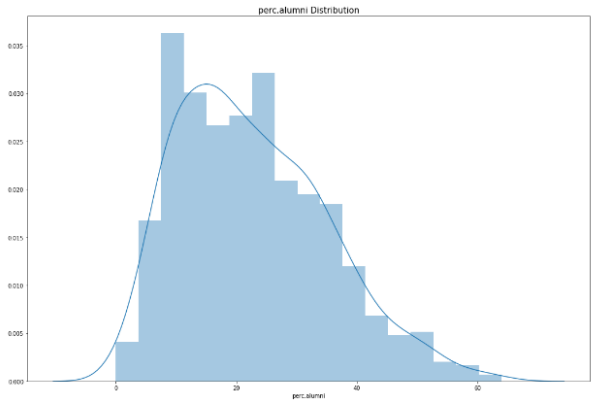
Variable – **'Terminal Distribution'**
Skew – Negative
Outliers presence – Yes

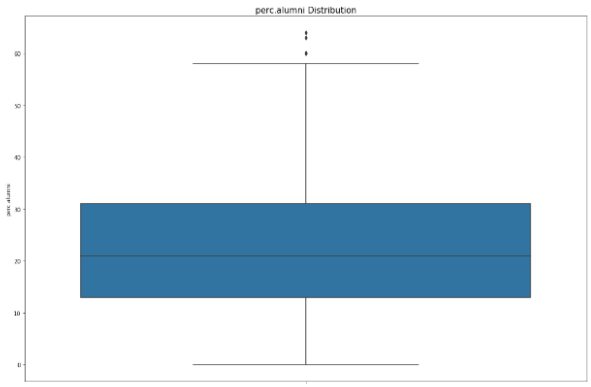


Variable – **'S.F.Ratio Distribution'**
Skew – Positive
Outliers presence – Yes

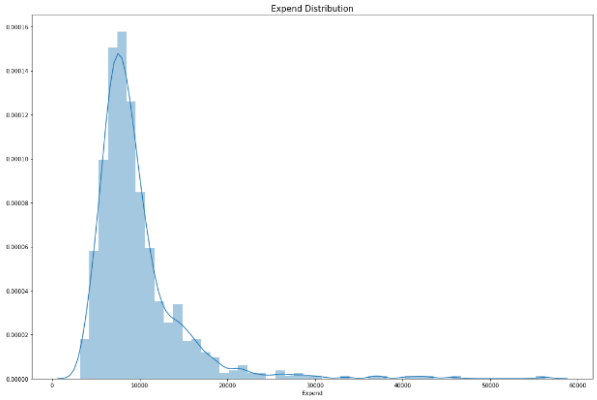


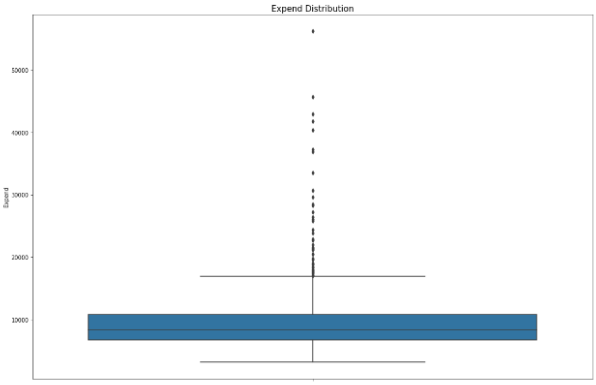
Variable – **'perci.alumni Distribution'**
Skew – Positive
Outliers presence – Yes



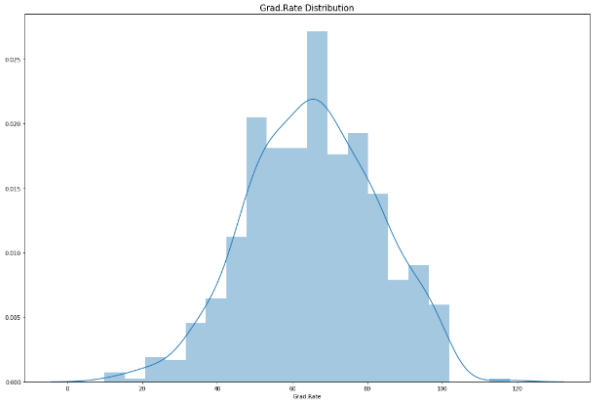


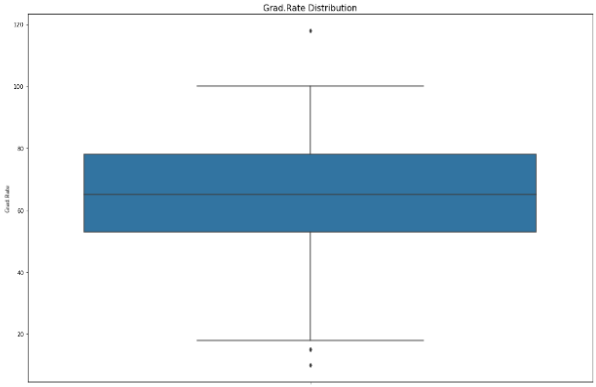
Variable – **'Expand Distribution'**
Skew – Positive
Outliers presence – Yes





Variable – **'Grade.Rate Distribution'**
Skew – Positive
Outliers presence – Yes





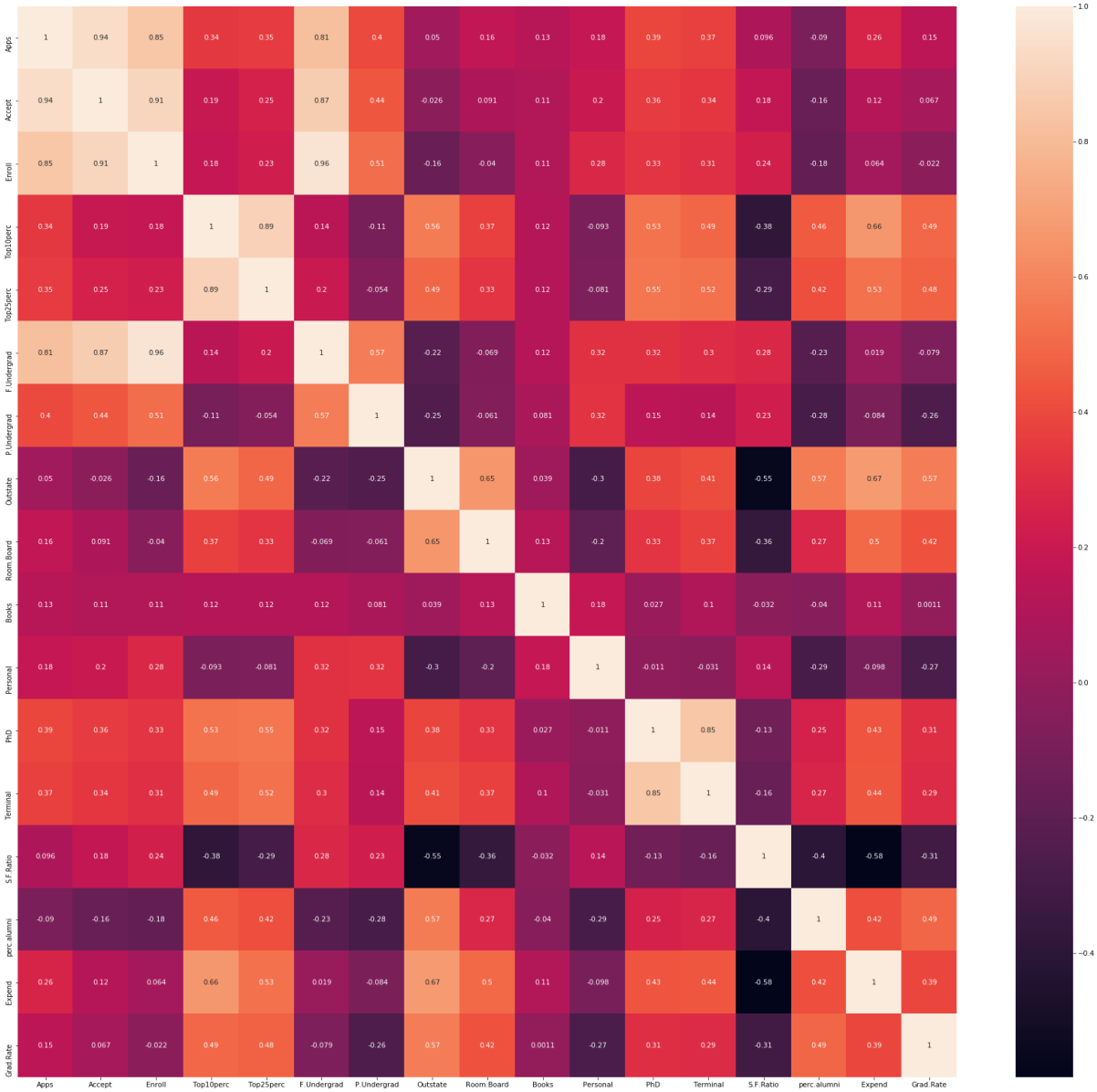
From above data, Expect Variable '**Top10perc'** , remaining all the variables have outliers in their data.

Multivariate Analysis:

We have the following correlation among the different variables given in the dataset.

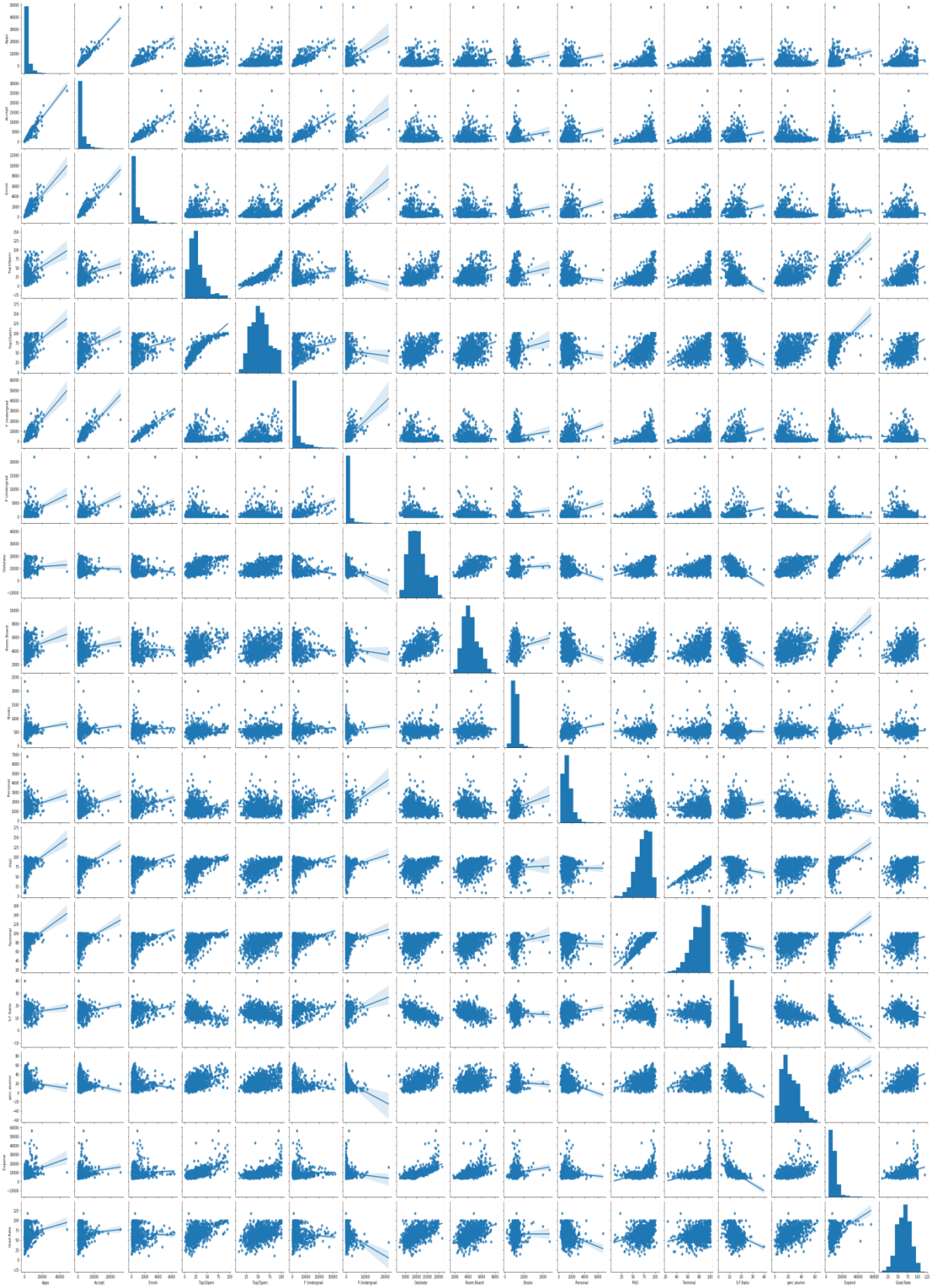
	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
Apps	1.000000	0.943451	0.846822	0.338834	0.351640	0.814491	0.398264	0.050159	0.164939	0.132559	0.178731	0.390697	0.369491	0.095633	-0.090226	0.259592	0.146755
Accept	0.943451	1.000000	0.911637	0.192447	0.247476	0.874223	0.441271	-0.025755	0.090899	0.113525	0.200989	0.355758	0.337583	0.176229	-0.159990	0.124717	0.067313
Enroll	0.846822	0.911637	1.000000	0.181294	0.226745	0.964640	0.513069	-0.155477	-0.040232	0.112711	0.280929	0.331469	0.308274	0.237271	-0.180794	0.064169	-0.022341
Top10perc	0.338834	0.192447	0.181294	1.000000	0.891995	0.141289	-0.105356	0.562331	0.371480	0.118858	-0.093316	0.531828	0.491135	-0.384875	0.455485	0.660913	0.494989
Top25perc	0.351640	0.247476	0.226745	0.891995	1.000000	0.199445	-0.053577	0.489394	0.331490	0.115527	-0.080810	0.545862	0.524749	-0.294629	0.417864	0.527447	0.477281
F.Undergrad	0.814491	0.874223	0.964640	0.141289	0.199445	1.000000	0.570512	-0.215742	-0.068890	0.115550	0.317200	0.318337	0.300019	0.279703	-0.229462	0.018652	-0.078773
P.Undergrad	0.398264	0.441271	0.513069	-0.105356	-0.053577	0.570512	1.000000	-0.253512	-0.061326	0.081200	0.319882	0.149114	0.141904	0.232531	-0.280792	-0.083568	-0.257001
Outstate	0.050159	-0.025755	-0.155477	0.562331	0.489394	-0.215742	-0.253512	1.000000	0.654256	0.038855	-0.299087	0.382982	0.407983	-0.554821	0.566262	0.672779	0.571290
Room.Board	0.164939	0.090899	-0.040232	0.371480	0.331490	-0.068890	-0.061326	0.654256	1.000000	0.127963	-0.199428	0.329202	0.374540	-0.362628	0.272363	0.501739	0.424942
Books	0.132559	0.113525	0.112711	0.118858	0.115527	0.115550	0.081200	0.038855	0.127963	1.000000	0.179295	0.026906	0.099955	-0.031929	-0.040208	0.112409	0.001061
Personal	0.178731	0.200989	0.280929	-0.093316	-0.080810	0.317200	0.319882	-0.299087	-0.199428	0.179295	1.000000	-0.010936	-0.030613	0.136345	-0.285968	-0.097892	-0.269344
PhD	0.390697	0.355758	0.331469	0.531828	0.545862	0.318337	0.149114	0.382982	0.329202	0.026906	-0.010936	1.000000	0.849587	-0.130530	0.249009	0.432762	0.305038
Terminal	0.369491	0.337583	0.308274	0.491135	0.524749	0.300019	0.141904	0.407983	0.374540	0.099955	-0.030613	0.849587	1.000000	-0.160104	0.267130	0.438799	0.289527
S.F.Ratio	0.095633	0.176229	0.237271	-0.384875	-0.294629	0.279703	0.232531	-0.554821	-0.362628	-0.031929	0.136345	-0.130530	-0.160104	1.000000	-0.402929	-0.583832	-0.306710
perc.alumni	-0.090226	-0.159990	-0.180794	0.455485	0.417864	-0.229462	-0.280792	0.566262	0.272363	-0.040208	-0.285968	0.249009	0.267130	-0.402929	1.000000	0.417712	0.490898
Expend	0.259592	0.124717	0.064169	0.660913	0.527447	0.018652	-0.083568	0.672779	0.501739	0.112409	-0.097892	0.432762	0.438799	-0.583832	0.417712	1.000000	0.390343
Grad.Rate	0.146755	0.067313	-0.022341	0.494989	0.477281	-0.078773	-0.257001	0.571290	0.424942	0.001061	-0.269344	0.305038	0.289527	-0.306710	0.490898	0.390343	1.000000

HeatMap:



From the above map , we can see that many columns are co-related to each other and there is **highest positive correlatio n** (0.964640) between variables **'F.undergrand'** and **'Enroll'** . Also there is **highest negative correlation**(- 0.010936) betw een variables **'PhD'** and **'Personal'**.

Pairplot:



In the above plot scatter diagrams are plotted for all the numerical columns in the dataset. From the visual representation , we can understand the degree of correlation between any two columns of the given dataset.

Variables **'Accept' , 'Enroll' , 'F.Undergrad'** show high positive linear correlation with Variable **'App'**.

Variables **'Enroll' , 'F.Undergrad'** show high positive linear correlation with Variable '**Accept'**.

Variable **'F.Undergrad'** shows high positive linear correlation with Variable **'Enroll'**.

Variable **'Top25perc'** shows high positive linear correlation with Variable **'Top10perc'**.

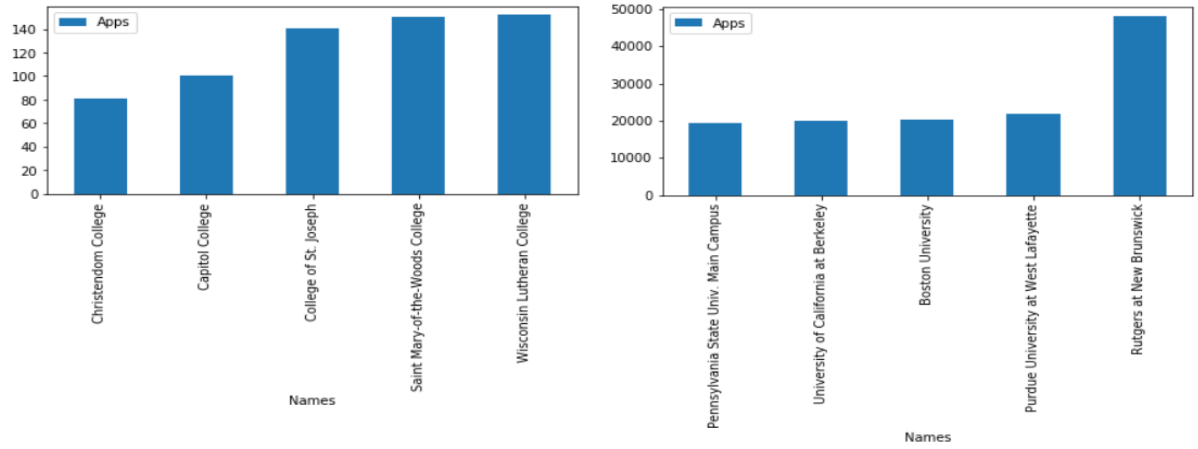
Variables**'Outstate','per.alumni','Grad.Rate'** show high negative linear correlation with Variable **'P.Undergrad'**.

Variable **'Terminal'** shows high positive linear correlation with Variable **'PhD'**.

Variable **'Expend'** shows high negative linear correlation with Variable **'S.F.Ratio'**.

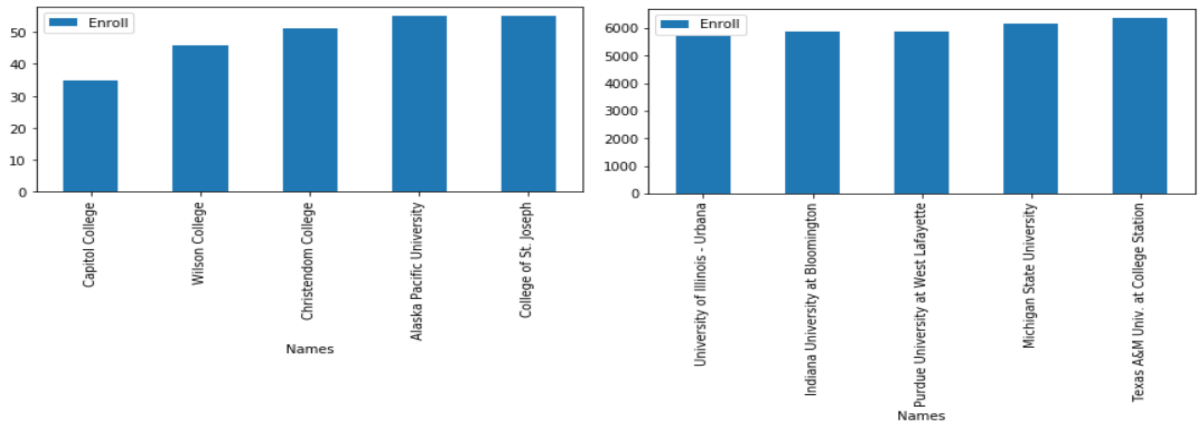
Number of applications received :

Among all the universities and colleges , **Rutgers at New Brunswick** received **highest** no of applications and **Christendom College** received **least** no of applications.



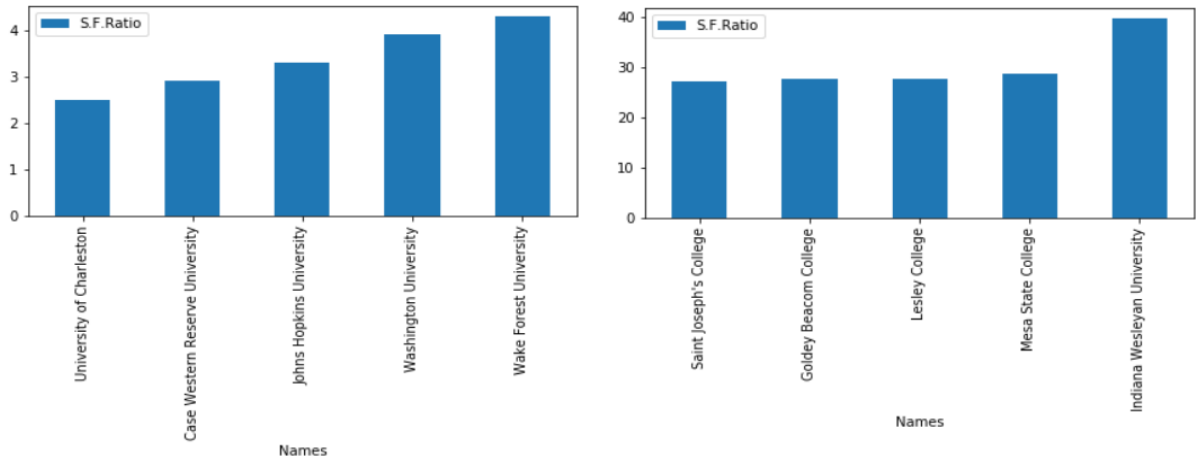
Number of new students enrolled:

Among all the universities and colleges , **Texas A&M Univ. at College Station** received **highest** no of enrolment of new students and **Capitol College** has **least** no of enrolment of new students.



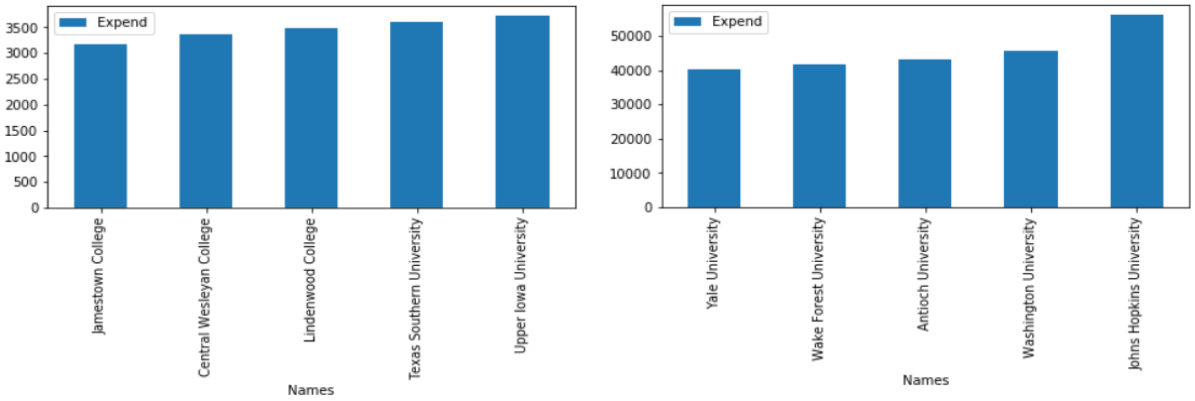
S.F.Ratio: Student/faculty ratio:

Among all the universities and colleges , **Indiana Wesleyan University** has **highest** Student/faculty ratio and **University of Charleston** has **least** Student/faculty ratio.



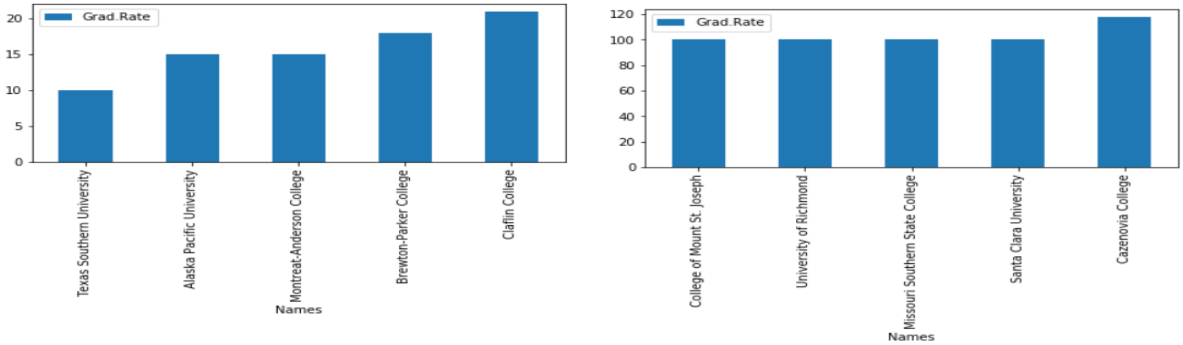
Expend: The Instructional expenditure per student:

Among all the universities and colleges , **Johns Hopkins University** has **highest** expenditure per student and **Jamestown College** has **least** expenditure per student.



Grad.Rate: Graduation rate:

Among all the universities and colleges , **Cazenovia College** has **highest** expenditure per student and **Texas Southern University** has **least** expenditure per student.



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

As most of the variables are highly skewed towards either right or left and also there is much difference in the magnitude of value in the dataset between values among different variables. We prefer to use z scale method to minimize the effect of both skewness and variability due to difference between high and low magnitude value of the data.

Scaled Variables :

	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	-0.346882	-0.321205	-0.063509	-0.258583	-0.191827	-0.168116	-0.209207	-0.746356	-0.964905	-0.602312	1.270045	-0.163028	-0.115729	1.013776	-0.867574	-0.501910	-0.318252
1	-0.210884	-0.038703	-0.288584	-0.656656	-1.353911	-0.209788	0.244307	0.457496	1.909208	1.215880	0.235515	-2.675646	-3.378176	-0.477704	-0.544572	0.166110	-0.551262
2	-0.406866	-0.376318	-0.478121	-0.315307	-0.292878	-0.549565	-0.497090	0.201305	-0.554317	-0.905344	-0.259582	-1.204845	-0.931341	-0.300749	0.585935	-0.177290	-0.667767
3	-0.668261	-0.681682	-0.692427	1.840231	1.677612	-0.658079	-0.520752	0.626633	0.996791	-0.602312	-0.688173	1.185206	1.175657	-1.615274	1.151188	1.792851	-0.376504
4	-0.726176	-0.764555	-0.780735	-0.656656	-0.596031	-0.711924	0.009005	-0.716508	-0.216723	1.518912	0.235515	0.204672	-0.523535	-0.553542	-1.675079	0.241803	-2.939613

2.3) Comment on the comparison between covariance and the correlation matrix.

Both covariance and the correlation matrix gives the same result. Covariance matrix can be used when the variable scales are similar and correlation matrix when variables are on different scales .

Covariance indicates the direction of the linear relationship between variables.

Correlation 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.

Correlation:

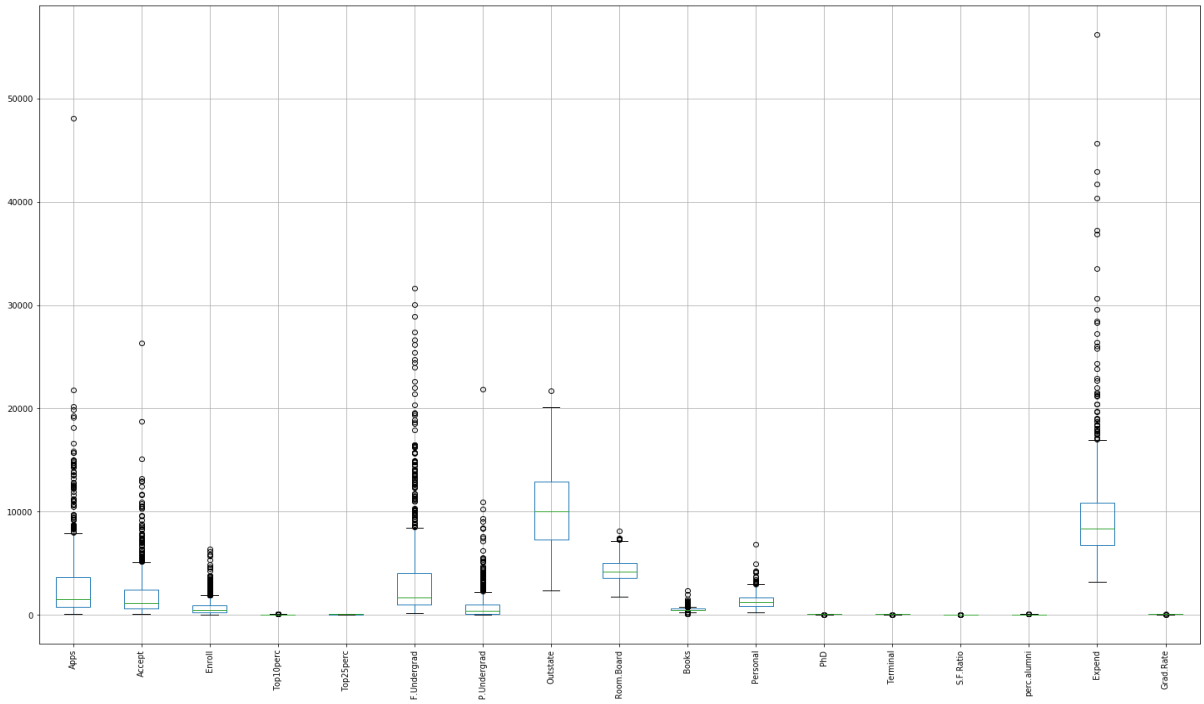
	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
Apps	1.000000	0.955307	0.896883	0.321342	0.364491	0.861002	0.519823	0.065337	0.187475	0.236138	0.229948	0.463924	0.434478	0.126411	-0.101158	0.242935	0.150803
Accept	0.955307	1.000000	0.935277	0.223298	0.273681	0.897034	0.572691	-0.005002	0.119586	0.208705	0.256346	0.427341	0.403409	0.188506	-0.165516	0.161808	0.078982
Enroll	0.896883	0.935277	1.000000	0.171756	0.230434	0.967302	0.641595	-0.155655	-0.023846	0.202057	0.339348	0.381540	0.354379	0.274269	-0.222723	0.054221	-0.023251
Top10perc	0.321342	0.223298	0.171756	1.000000	0.913875	0.111215	-0.180009	0.562160	0.357366	0.153452	-0.116730	0.544048	0.506748	-0.387926	0.455797	0.657039	0.493670
Top25perc	0.364491	0.273681	0.230434	0.913875	1.000000	0.181196	-0.099295	0.489569	0.330987	0.169761	-0.086810	0.551461	0.527654	-0.297233	0.416832	0.572905	0.478985
F.Undergrad	0.861002	0.897034	0.967302	0.111215	0.181196	1.000000	0.696130	-0.226166	-0.054476	0.207879	0.359783	0.361564	0.335054	0.324504	-0.285457	0.000371	-0.082239
P.Undergrad	0.519823	0.572691	0.641595	-0.180009	-0.099295	0.696130	1.000000	-0.354216	-0.067638	0.122529	0.344053	0.127663	0.122152	0.370607	-0.419334	-0.201929	-0.265158
Outstate	0.065337	-0.005002	-0.155655	0.562160	0.489569	-0.226166	-0.354216	1.000000	0.655489	0.005110	-0.325609	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	-0.067638	0.655489	1.000000	0.108924	-0.219554	0.341469	0.379270	-0.376430	0.272393	0.580622	0.425790
Books	0.236138	0.208705	0.202057	0.153452	0.169761	0.207879	0.122529	0.005110	0.108924	1.000000	0.239863	0.136390	0.159318	-0.008536	-0.042832	0.149983	-0.008051
Personal	0.229948	0.256346	0.339348	-0.116730	-0.086810	0.359783	0.344053	-0.325609	-0.219554	0.239863	1.000000	-0.011684	-0.031971	0.173913	-0.305753	-0.163271	-0.290894
PhD	0.463924	0.427341	0.381540	0.544048	0.551461	0.361564	0.127663	0.391321	0.341469	0.136390	-0.011684	1.000000	0.862928	-0.129390	0.248877	0.510529	0.310019
Terminal	0.434478	0.403409	0.354379	0.506748	0.527654	0.335054	0.122152	0.412579	0.379270	0.159318	-0.031971	0.862928	1.000000	-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	-0.376430	-0.008536	0.173913	-0.129390	-0.150993	1.000000	-0.412101	-0.654376	-0.308525
perc.alumni	-0.101158	-0.165516	-0.222723	0.455797	0.416832	-0.285457	-0.419334	0.565736	0.272393	-0.042832	-0.305753	0.248877	0.266033	-0.412101	1.000000	0.462922	0.491408
Expend	0.242935	0.161808	0.054221	0.657039	0.572905	0.000371	-0.201929	0.775328	0.580622	0.149983	-0.163271	0.510529	0.524068	-0.654376	0.462922	1.000000	0.415291
Grad.Rate	0.150803	0.078982	-0.023251	0.493670	0.478985	-0.082239	-0.265158	0.572458	0.425790	-0.008051	-0.290894	0.310019	0.292803	-0.308525	0.491408	0.415291	1.000000

Covariance:

	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	1.001289	0.956538	0.898039	0.321756	0.364961	0.862111	0.520493	0.065421	0.187717	0.236442	0.230244	0.464522	0.435038	0.126574	-0.101288	0.243248	0.150998
Accept	0.956538	1.001289	0.936482	0.223586	0.274033	0.898190	0.573429	-0.005009	0.119740	0.208974	0.256676	0.427891	0.403929	0.188749	-0.165729	0.162017	0.079084
Enroll	0.898039	0.936482	1.001289	0.171977	0.230731	0.968549	0.642422	-0.155856	-0.023876	0.202317	0.339785	0.382031	0.354836	0.274622	-0.223010	0.054291	-0.023281
Top10perc	0.321756	0.223586	0.171977	1.001289	0.915053	0.111358	-0.180241	0.562884	0.357826	0.153650	-0.116880	0.544749	0.507401	-0.388426	0.456384	0.657886	0.494307
Top25perc	0.364961	0.274033	0.230731	0.915053	1.001289	0.181429	-0.099423	0.490200	0.331413	0.169980	-0.086922	0.552172	0.528334	-0.297616	0.417369	0.573643	0.479602
F.Undergrad	0.862111	0.898190	0.968549	0.111358	0.181429	1.001289	0.697027	-0.226457	-0.054546	0.208147	0.360246	0.362030	0.335486	0.324922	-0.285825	0.000371	-0.082345
P.Undergrad	0.520493	0.573429	0.642422	-0.180241	-0.099423	0.697027	1.001289	-0.354673	-0.067725	0.122686	0.344496	0.127827	0.122309	0.371085	-0.419874	-0.202189	-0.265499
Outstate	0.065421	-0.005009	-0.155856	0.562884	0.490200	-0.226457	-0.354673	1.001289	0.656334	0.005117	-0.326029	0.391825	0.413110	-0.574422	0.566465	0.776327	0.573196
Room.Board	0.187717	0.119740	-0.023876	0.357826	0.331413	-0.054546	-0.067725	0.656334	1.001289	0.109065	-0.219837	0.341909	0.379759	-0.376915	0.272744	0.581370	0.426339
Books	0.236442	0.208974	0.202317	0.153650	0.169980	0.208147	0.122686	0.005117	0.109065	1.001289	0.240172	0.136566	0.159523	-0.008547	-0.042887	0.150177	-0.008061
Personal	0.230244	0.256676	0.339785	-0.116880	-0.086922	0.360246	0.344496	-0.326029	-0.219837	0.240172	1.001289	-0.011699	-0.032012	0.174137	-0.306147	-0.163481	-0.291269
PhD	0.464522	0.427891	0.382031	0.544749	0.552172	0.362030	0.127827	0.391825	0.341909	0.136566	-0.011699	1.001289	0.864040	-0.129556	0.249198	0.511187	0.310419
Terminal	0.435038	0.403929	0.354836	0.507401	0.528334	0.335486	0.122309	0.413110	0.379759	0.159523	-0.032012	0.864040	1.001289	-0.151188	0.266375	0.524744	0.293180
S.F.Ratio	0.126574	0.188749	0.274622	-0.388426	-0.297616	0.324922	0.371085	-0.574422	-0.376915	-0.008547	0.174137	-0.129556	-0.151188	1.001289	-0.412632	-0.655220	-0.308922
perc.alumni	-0.101288	-0.165729	-0.223010	0.456384	0.417369	-0.285825	-0.419874	0.566465	0.272744	-0.042887	-0.306147	0.249198	0.266375	-0.412632	1.001289	0.463519	0.492041
Expend	0.243248	0.162017	0.054291	0.657886	0.573643	0.000371	-0.202189	0.776327	0.581370	0.150177	-0.163481	0.511187	0.524744	-0.655220	0.463519	1.001289	0.415826
Grad.Rate	0.150998	0.079084	-0.023281	0.494307	0.479602	-0.082345	-0.265499	0.573196	0.426339	-0.008061	-0.291269	0.310419	0.293180	-0.308922	0.492041	0.415826	1.001289

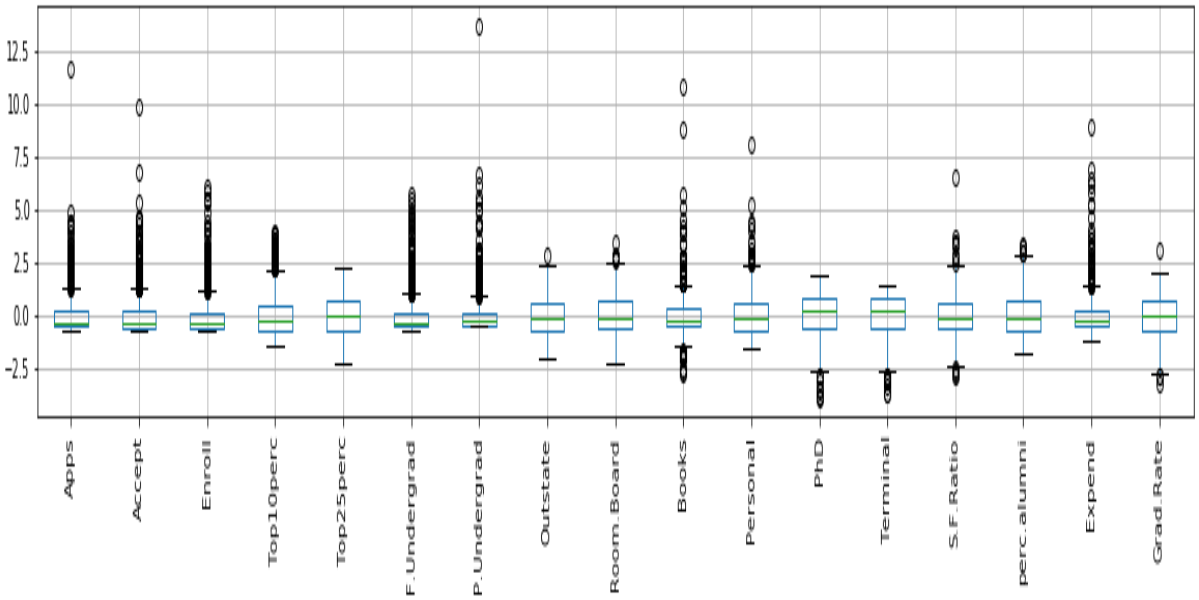
2.4) Check the dataset for outliers before and after scaling. Draw your inferences from this exercise.

Box Plot of original data set :



From the given data set , Except variable **'Top25perc'** , remaining all the variables are having outliers.

Box plot of data after scaling:



Even after scaling the data, we can see similar type of box plot for all the variables in data. Hence we can conclude that scaling has no affect on outliers.

2.5) Build the covariance matrix, eigenvalues, and eigenvector.

Covariance matrix :

1.00E+00	9.57E-01	8.98E-01	3.22E-01	3.65E-01	8.62E-01	5.20E-01	6.54E-02	1.88E-01	2.36E-01	2.30E-01	4.65E-01	4.35E-01	1.27E-01	-1.01E-01	2.43E-01	1.51E-01
9.57E-01	1.00E+00	9.36E-01	2.24E-01	2.74E-01	8.98E-01	5.73E-01	-5.01E-03	1.20E-01	2.09E-01	2.57E-01	4.28E-01	4.04E-01	1.89E-01	-1.66E-01	1.62E-01	7.91E-02
8.98E-01	9.36E-01	1.00E+00	1.72E-01	2.31E-01	9.69E-01	6.42E-01	-1.56E-01	-2.39E-02	2.02E-01	3.40E-01	3.82E-01	3.55E-01	2.75E-01	-2.23E-01	5.43E-02	-2.33E-02
3.22E-01	2.24E-01	1.72E-01	1.00E+00	9.15E-01	1.11E-01	-1.80E-01	5.63E-01	3.58E-01	1.54E-01	-1.17E-01	5.45E-01	5.07E-01	-3.88E-01	4.56E-01	6.58E-01	4.94E-01
3.65E-01	2.74E-01	2.31E-01	9.15E-01	1.00E+00	1.81E-01	-9.94E-02	4.90E-01	3.31E-01	1.70E-01	-8.69E-02	5.52E-01	5.28E-01	-2.98E-01	4.17E-01	5.74E-01	4.80E-01
8.62E-01	8.98E-01	9.69E-01	1.11E-01	1.81E-01	1.00E+00	6.97E-01	-2.26E-01	-5.45E-02	2.08E-01	3.60E-01	3.62E-01	3.35E-01	3.25E-01	-2.86E-01	3.71E-04	-8.23E-02
5.20E-01	5.73E-01	6.42E-01	-1.80E-01	-9.94E-02	6.97E-01	1.00E+00	-3.55E-01	-6.77E-02	1.23E-01	3.44E-01	1.28E-01	1.22E-01	3.71E-01	-4.20E-01	-2.02E-01	-2.65E-01
6.54E-02	-5.01E-03	-1.56E-01	5.63E-01	4.90E-01	-2.26E-01	-3.55E-01	1.00E+00	6.56E-01	5.12E-03	-3.26E-01	3.92E-01	4.13E-01	-5.74E-01	5.66E-01	7.76E-01	5.73E-01
1.88E-01	1.20E-01	-2.39E-02	3.58E-01	3.31E-01	-5.45E-02	-6.77E-02	6.56E-01	1.00E+00	1.09E-01	-2.20E-01	3.42E-01	3.80E-01	-3.77E-01	2.73E-01	5.81E-01	4.26E-01
2.36E-01	2.09E-01	2.02E-01	1.54E-01	1.70E-01	2.08E-01	1.23E-01	5.12E-03	1.09E-01	1.00E+00	2.40E-01	1.37E-01	1.60E-01	-8.55E-03	-4.29E-02	1.50E-01	-8.06E-03
2.30E-01	2.57E-01	3.40E-01	-1.17E-01	-8.69E-02	3.60E-01	3.44E-01	-3.26E-01	-2.20E-01	2.40E-01	1.00E+00	-1.17E-02	-3.20E-02	1.74E-01	-3.06E-01	-1.63E-01	-2.91E-01
4.65E-01	4.28E-01	3.82E-01	5.45E-01	5.52E-01	3.62E-01	1.28E-01	3.92E-01	3.42E-01	1.37E-01	-1.17E-02	1.00E+00	8.64E-01	-1.30E-01	2.49E-01	5.11E-01	3.10E-01
4.35E-01	4.04E-01	3.55E-01	5.07E-01	5.28E-01	3.35E-01	1.22E-01	4.13E-01	3.80E-01	1.60E-01	-3.20E-02	8.64E-01	1.00E+00	-1.51E-01	2.66E-01	5.25E-01	2.93E-01
1.27E-01	1.89E-01	2.75E-01	-3.88E-01	-2.98E-01	3.25E-01	3.71E-01	-5.74E-01	-3.77E-01	-8.55E-03	1.74E-01	-1.30E-01	-1.51E-01	1.00E+00	-4.13E-01	-6.55E-01	-3.09E-01
-1.01E-01	-1.66E-01	-2.23E-01	4.56E-01	4.17E-01	-2.86E-01	-4.20E-01	5.66E-01	2.73E-01	-4.29E-02	-3.06E-01	2.49E-01	2.66E-01	-4.13E-01	1.00E+00	4.64E-01	4.92E-01
2.43E-01	1.62E-01	5.43E-02	6.58E-01	5.74E-01	3.71E-04	-2.02E-01	7.76E-01	5.81E-01	1.50E-01	-1.63E-01	5.11E-01	5.25E-01	-6.55E-01	4.64E-01	1.00E+00	4.16E-01
1.51E-01	7.91E-02	-2.33E-02	4.94E-01	4.80E-01	-8.23E-02	-2.65E-01	5.73E-01	4.26E-01	-8.06E-03	-2.91E-01	3.10E-01	2.93E-01	-3.09E-01	4.92E-01	4.16E-01	1.00E+00

Eigen values:

5.662522	4.894708	1.126367	1.003977	0.872184	0.765754	0.584914	0.544505	0.423523	0.381018	0.247015	0.022394	0.037894	0.147264	0.134345	0.098834	0.07469

Eigen vector:

-2.62E-01	3.14E-01	8.10E-02	-9.88E-02	-2.20E-01	2.19E-03	-2.84E-02	-8.99E-02	1.31E-01	-1.56E-01	-8.62E-02	1.82E-01	-5.99E-01	9.00E-02	8.89E-02	5.49E-01	5.41E-03
-2.31E-01	3.45E-01	1.08E-01	-1.18E-01	-1.90E-01	-1.65E-02	-1.30E-02	-1.38E-01	1.42E-01	-1.49E-01	-4.26E-02	-3.91E-01	6.61E-01	1.59E-01	4.38E-02	2.92E-01	1.45E-02
-1.89E-01	3.83E-01	8.55E-02	-9.31E-03	-1.62E-01	-6.81E-02	-1.52E-02	-1.44E-01	5.09E-02	-6.49E-02	-4.38E-02	7.17E-01	2.33E-01	-3.54E-02	-6.19E-02	-4.17E-01	-4.98E-02
-3.39E-01	-9.93E-02	-7.88E-02	3.69E-01	-1.57E-01	-8.89E-02	-2.57E-01	2.90E-01	-1.22E-01	-3.59E-02	1.78E-03	-5.62E-02	2.21E-02	-3.92E-02	7.00E-02	8.80E-03	-7.24E-01
-3.35E-01	-5.95E-02	-5.08E-02	4.17E-01	-1.44E-01	-2.76E-02	-2.39E-01	3.46E-01	-1.94E-01	6.42E-03	-1.02E-01	1.97E-02	3.23E-02	1.46E-01	-9.70E-02	-1.08E-02	6.55E-01
-1.63E-01	3.99E-01	7.37E-02	-1.40E-02	-1.03E-01	-5.16E-02	-3.12E-02	-1.09E-01	1.45E-03	-1.64E-04	-3.50E-02	-5.43E-01	-3.68E-01	-1.34E-01	-8.72E-02	-5.71E-01	2.53E-02
-2.25E-02	3.58E-01	4.04E-02	-2.25E-01	9.57E-02	-2.45E-02	-1.00E-02	1.24E-01	-6.35E-01	5.46E-01	2.52E-01	2.95E-02	2.62E-02	5.02E-02	4.46E-02	1.46E-01	-3.97E-02
-2.84E-01	-2.52E-01	1.49E-02	-2.63E-01	-3.73E-02	-2.04E-02	9.45E-02	1.13E-02	-8.37E-03	-2.32E-01	5.93E-01	1.03E-03	-8.14E-02	5.60E-01	6.72E-02	-2.12E-01	-1.59E-03
-2.44E-01	-1.32E-01	-2.11E-02	-5.81E-01	6.91E-02	2.37E-01	9.45E-02	3.90E-01	-2.21E-01	-2.55E-01	-4.75E-01	9.86E-03	2.68E-02	-1.07E-01	1.78E-02	-1.01E-01	-2.83E-02
-9.67E-02	9.40E-02	-6.97E-01	3.62E-02	-3.54E-02	6.39E-01	-1.11E-01	-2.40E-01	2.10E-02	9.12E-02	4.36E-02	4.36E-03	1.05E-02	5.16E-02	3.54E-02	-2.86E-02	-8.06E-03
3.52E-02	2.32E-01	-5.31E-01	1.15E-01	4.75E-04	-3.81E-01	6.39E-01	2.77E-01	1.74E-02	-1.28E-01	1.52E-02	-1.09E-02	4.55E-03	9.39E-03	-1.19E-02	3.38E-02	1.43E-03
-3.26E-01	5.51E-02	8.11E-02	1.47E-01	5.51E-01	3.34E-03	8.92E-02	-3.43E-02	1.67E-01	1.01E-01	-3.92E-02	1.33E-02	1.25E-02	-7.17E-02	7.03E-01	-6.38E-02	8.31E-02
-3.23E-01	4.30E-02	5.90E-02	8.90E-02	5.90E-01	3.54E-02	9.17E-02	-9.03E-02	1.13E-01	8.60E-02	-8.49E-02	7.38E-03	-1.79E-02	1.64E-01	-6.62E-01	9.85E-02	-1.13E-01
1.63E-01	2.60E-01	2.74E-01	2.59E-01	1.43E-01	4.69E-01	1.53E-01	2.43E-01	-1.54E-01	-4.71E-01	3.63E-01	8.86E-03	1.83E-02	-2.40E-01	-4.79E-02	6.20E-02	3.83E-03
-1.87E-01	-2.57E-01	1.04E-01	2.24E-01	-1.28E-01	1.26E-02	3.91E-01	-5.66E-01	-5.39E-01	-1.48E-01	-1.74E-01	-2.41E-02	-8.03E-05	-4.90E-02	3.59E-02	2.81E-02	-7.33E-03
-3.29E-01	-1.60E-01	-1.84E-01	-2.14E-01	2.24E-02	-2.32E-01	-1.51E-01	-1.19E-01	2.42E-02	-8.04E-02	3.94E-01	1.06E-02	5.60E-02	-6.90E-01	-1.27E-01	1.29E-01	1.45E-01
-2.39E-01	-1.68E-01	2.45E-01	3.62E-02	-3.57E-01	3.14E-01	4.69E-01	1.80E-01	3.16E-01	4.88E-01	8.73E-02	-2.51E-03	1.48E-02	-1.59E-01	-6.31E-02	-7.10E-03	-3.29E-03

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

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
Scaled variable Data	a	-3.76E-01	-3.38E-01	1.06E-01	-2.47E-01	-1.92E-01	-1.88E-02	-1.66E-01	-7.46E-01	-9.68E-01	-7.77E-01	1.44E+00	-1.74E-01	-1.23E-01	1.07E+00	-8.70E-01	-6.31E-01	-3.19E-01
	b	-1.59E-01	1.17E-01	-2.60E-01	-6.96E-01	-1.35E+00	-9.36E-02	7.98E-01	4.58E-01	1.92E+00	1.83E+00	2.89E-01	-2.75E+00	-2.79E+00	-4.90E-01	-5.46E-01	3.96E-01	-5.53E-01
	c	-4.72E-01	-4.27E-01	-5.69E-01	-3.11E-01	-2.93E-01	-7.04E-01	-7.78E-01	2.01E-01	-5.55E-01	-1.21E+00	-2.61E-01	-1.24E+00	-9.53E-01	-3.04E-01	5.91E-01	-1.32E-01	-6.69E-01
	d	-8.90E-01	-9.18E-01	-9.19E-01	2.13E+00	1.68E+00	-8.99E-01	-8.28E-01	6.27E-01	1.00E+00	-7.77E-01	-7.37E-01	1.21E+00	1.19E+00	-1.68E+00	1.16E+00	2.29E+00	-3.78E-01
	e	-9.83E-01	-1.05E+00	-1.06E+00	-6.96E-01	-5.96E-01	-9.96E-01	2.98E-01	-7.17E-01	-2.16E-01	2.22E+00	2.89E-01	2.02E-01	-5.38E-01	-5.69E-01	-1.68E+00	5.12E-01	-2.92E+00

Eigen vector	f	-2.62E-01	3.14E-01	8.10E-02	-9.88E-02	-2.20E-01	2.19E-03	-2.84E-02	-8.99E-02	1.31E-01	-1.56E-01	-8.62E-02	1.82E-01	-5.99E-01	9.00E-02	8.89E-02	5.49E-01	5.41E-03
	g	-2.31E-01	3.45E-01	1.08E-01	-1.18E-01	-1.90E-01	-1.65E-02	-1.30E-02	-1.38E-01	1.42E-01	-1.49E-01	-4.26E-02	-3.91E-01	6.61E-01	1.59E-01	4.38E-02	2.92E-01	1.45E-02
	h	-1.89E-01	3.83E-01	8.55E-02	-9.31E-03	-1.62E-01	-6.81E-02	-1.52E-02	-1.44E-01	5.09E-02	-6.49E-02	-4.38E-02	7.17E-01	2.33E-01	-3.54E-02	-6.19E-02	-4.17E-01	-4.98E-02
	i	-3.39E-01	-9.93E-02	-7.88E-02	3.69E-01	-1.57E-01	-8.89E-02	-2.57E-01	2.90E-01	-1.22E-01	-3.59E-02	1.78E-03	-5.62E-02	2.21E-02	-3.92E-02	7.00E-02	8.80E-03	-7.24E-01
	j	-3.35E-01	-5.95E-02	-5.08E-02	4.17E-01	-1.44E-01	-2.76E-02	-2.39E-01	3.46E-01	-1.94E-01	6.42E-03	-1.02E-01	1.97E-02	3.23E-02	1.46E-01	-9.70E-02	-1.08E-02	6.55E-01
	k	-1.63E-01	3.99E-01	7.37E-02	-1.40E-02	-1.03E-01	-5.16E-02	-3.12E-02	-1.09E-01	1.45E-03	-1.64E-04	-3.50E-02	-5.43E-01	-3.68E-01	-1.34E-01	-8.72E-02	-5.71E-01	2.53E-02
	l	-2.25E-02	3.58E-01	4.04E-02	-2.25E-01	9.57E-02	-2.45E-02	-1.00E-02	1.24E-01	-6.35E-01	5.46E-01	2.52E-01	2.95E-02	2.62E-02	5.02E-02	4.46E-02	1.46E-01	-3.97E-02
	m	-2.84E-01	-2.52E-01	1.49E-02	-2.63E-01	-3.73E-02	-2.04E-02	9.45E-02	1.13E-02	-8.37E-03	-2.32E-01	5.93E-01	1.03E-03	-8.14E-02	5.60E-01	6.72E-02	-2.12E-01	-1.59E-03
	n	-2.44E-01	-1.32E-01	-2.11E-02	-5.81E-01	6.91E-02	2.37E-01	9.45E-02	3.90E-01	-2.21E-01	-2.55E-01	-4.75E-01	9.86E-03	2.68E-02	-1.07E-01	1.78E-02	-1.01E-01	-2.83E-02
	o	-9.67E-02	9.40E-02	-6.97E-01	3.62E-02	-3.54E-02	6.39E-01	-1.11E-01	-2.40E-01	2.10E-02	9.12E-02	4.36E-02	4.36E-03	1.05E-02	5.16E-02	3.54E-02	-2.86E-02	-8.06E-03
	p	3.52E-02	2.32E-01	-5.31E-01	1.15E-01	4.75E-04	-3.81E-01	6.39E-01	2.77E-01	1.74E-02	-1.28E-01	1.52E-02	-1.09E-02	4.55E-03	9.39E-03	-1.19E-02	3.38E-02	1.43E-03
	q	-3.26E-01	5.51E-02	8.11E-02	1.47E-01	5.51E-01	3.34E-03	8.92E-02	-3.43E-02	1.67E-01	1.01E-01	-3.92E-02	1.33E-02	1.25E-02	-7.17E-02	7.03E-01	-6.38E-02	8.31E-02
	r	-3.23E-01	4.30E-02	5.90E-02	8.90E-02	5.90E-01	3.54E-02	9.17E-02	-9.03E-02	1.13E-01	8.60E-02	-8.49E-02	7.38E-03	-1.79E-02	1.64E-01	-6.62E-01	9.85E-02	-1.13E-01
	s	1.63E-01	2.60E-01	2.74E-01	2.59E-01	1.43E-01	4.69E-01	1.53E-01	2.43E-01	-1.54E-01	-4.71E-01	3.63E-01	8.86E-03	1.83E-02	-2.40E-01	-4.79E-02	6.20E-02	3.83E-03
	t	-1.87E-01	-2.57E-01	1.04E-01	2.24E-01	-1.28E-01	1.26E-02	3.91E-01	-5.66E-01	-5.39E-01	-1.48E-01	-1.74E-01	-2.41E-02	-8.03E-05	-4.90E-02	3.59E-02	2.81E-02	-7.33E-03
	u	-3.29E-01	-1.60E-01	-1.84E-01	-2.14E-01	2.24E-02	-2.32E-01	-1.51E-01	-1.19E-01	2.42E-02	-8.04E-02	3.94E-01	1.06E-02	5.60E-02	-6.90E-01	-1.27E-01	1.29E-01	1.45E-01
	v	-2.39E-01	-1.68E-01	2.45E-01	3.62E-02	-3.57E-01	3.14E-01	4.69E-01	1.80E-01	3.16E-01	4.88E-01	8.73E-02	-2.51E-03	1.48E-02	-1.59E-01	-6.31E-02	-7.10E-03	-3.29E-03

PC1 = a1*f1+a2*g1+a3*h1+a4*i1+a5*j1+a6*k1+a7*l1+a8*m1+a9*n1+a10*o1+a11*p1+a12*q1+a13*r1+a14*s1+a15*t1+a16*u1+a17*v1

Similarly remaining PC1 values can be derived by summing up the values obtained by multiplying values in rows b,c, d etc., with values in column1.

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.

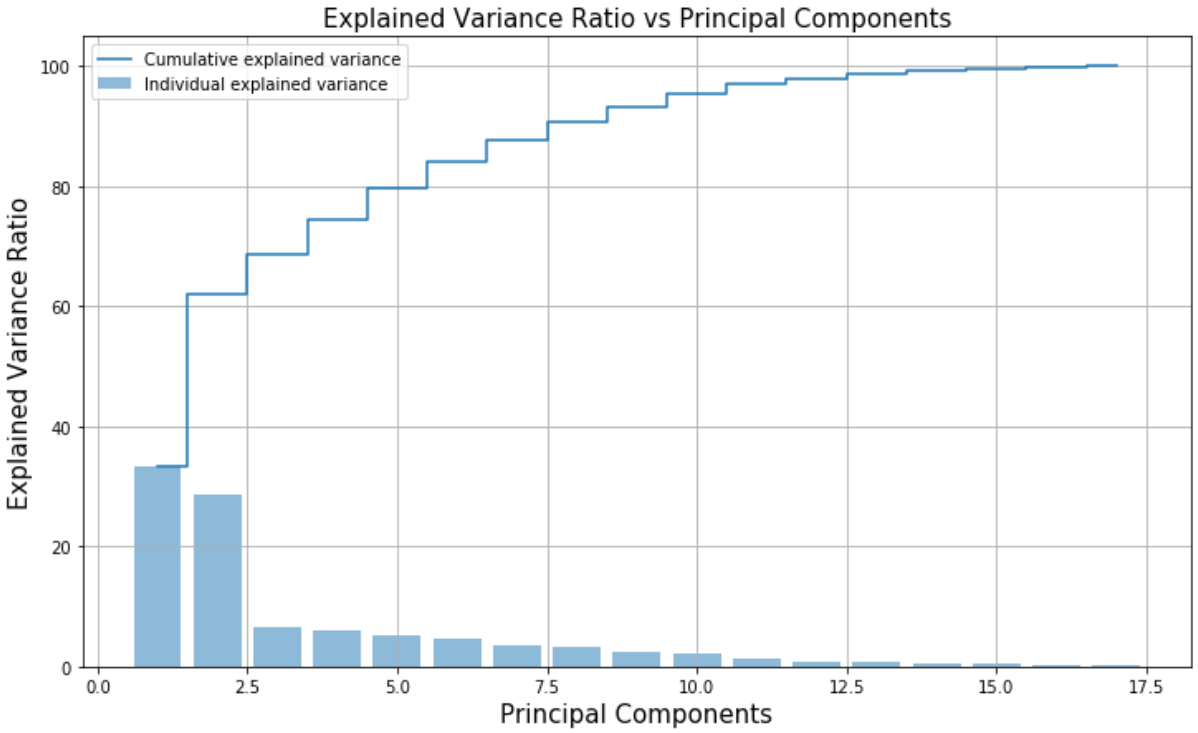
```
Eigen Values
%s [5.6625219  4.89470815 1.12636744 1.00397659 0.87218426 0.7657541
0.58491404 0.5445048  0.42352336 0.38101777 0.24701456 0.02239369
0.03789395 0.14726392 0.13434483 0.09883384 0.07469003]
```

```
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.        ]
```

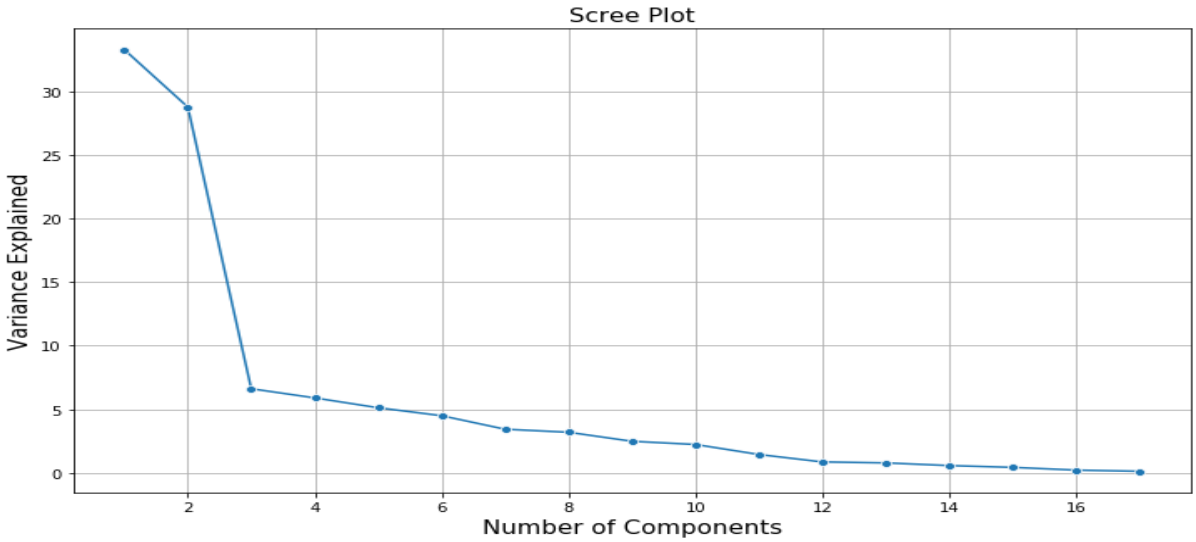
We can see from the above values that PC1 can explain upto 33.36% variability of the data.

Cumulative of PC1 & PC2 can explain upto 62.02 % variability of the data.

Cumulative of PC1 to PC8 can explain upto 90.79% variability of the data.



From above plot, we can understand the Cumulative variance and Individual varaiance explained by the Principal Components. By combining all the Principal Components, we are getting the 100 % cumulative variance.



We can see that there is steep drop in variance along with the increase in PC's.

We can see that there are totally 4 eigenvalues greater than 1.

So ,we will proceed with 4 components here. But depending on requirement we can take 8 components to look upto 90% variation.

Eigen vectors denote the weightage for the Principal Components where as Eigen values denote the Variance

PCA :

```
[[-1.60249937,  0.99368301,  0.03004472, -1.00842227],
 [-1.80467546, -0.07041499,  2.12213038,  3.13893348],
 [-1.60828257, -1.38279212, -0.50151303, -0.03637215],
 ...,
 [-0.57688268,  0.01779846,  0.32216192, -0.58726295],
 [ 6.570952  , -1.18493014,  1.32596593,  0.07770606],
 [-0.47739307,  1.04394672, -1.42543772, -1.3002754 ]])
```

PC0 **PC1** **PC2** **PC3**

We have the above list of values against individual Principal Components.

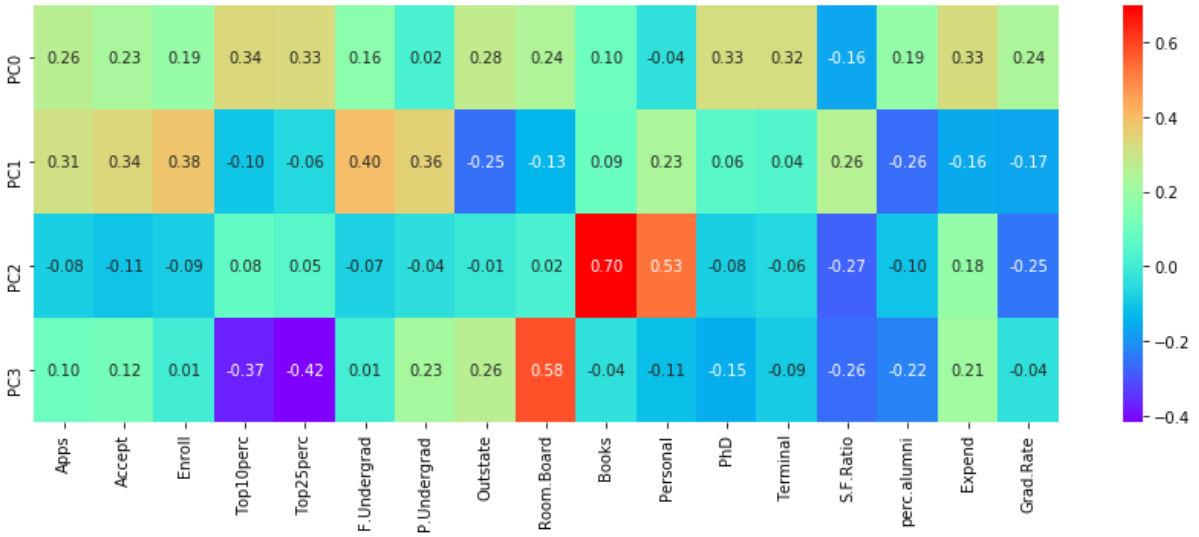
Principal Component scores data frame for four PC's

	0	1	2	3
0	-1.602499	0.993683	0.030045	-1.008422
1	-1.804675	-0.070415	2.122126	3.138943
2	-1.608283	-1.382792	-0.501512	-0.036374
3	2.803644	-3.367395	0.367768	-0.632913
4	-2.200868	-0.099348	3.122524	0.657705
...	...	...	...	...
772	-3.395392	1.995628	-0.744776	0.800067
773	0.319750	-0.314944	0.013597	0.653857
774	-0.576883	0.017798	0.322160	-0.587258
775	6.570952	-1.184930	1.325964	0.077707
776	-0.477393	1.043947	-1.425438	-1.300274

2.8) Mention the business implication of using the Principal Component Analysis for this case study.
[Hint: **Write Interpretations of the Principal Components Obtained]**

The data set given comprises 17 columns of continous variables and we have used PCA to create principal components to reduce the dimensionality of data, while keeping as much variation as possible.

This PCA helped us to reduce multidimensional data to lower dimensions while retaining most of the information.



By drawing heatmap between all the variables and the principal components , we can understand the relation between various variables and PC's.

		Variables	
		High Positive relation	**High negative relation**
PC's	**PC0**	Top10perc, Top25perc , PhD , Terminal, Expend	S.F.Ratio, Personal
	PC1	Enroll, F.Undergrad, P.Undergrad, Apps, Accept	perc.alumni, Outstate
	PC2	Books, Personal	S.F.Ratio, Grad.Rate
	PC3	Room.Board	Top10perc, Top25perc

So, we are able to understand the 75% variability of data associated the given dataset by just looking at the four Principal Components. This result will really help us in model building and analysis.