Project

Karthiheswar

Table of Contents

1 Project Objective

The objective of the report is to explore all the projects data set in Python and generate insights about the data set. This exploration report will consist of the following:

- Importing the dataset in Python
- Understanding the structure of dataset
- Graphical exploration
- Performing ANOVA analysis
- Hypothesis analysis
- Insights from the dataset
- Exploring the outliers in dataset
- Scaling the variables
- Principle Component Analysis

2 Fever

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

For A,
Null hypothesis Ho=u1=u2=2
Alternate hypothesis H1=u1!=u2

For B,
Null hypothesis Ho=u1=u2=2
Alternate hypothesis H1=u1!=u2

In pandas it is considered that A and B are continues variables as they are numerical. How ever they are categorical variable as per our dataset, so its difficult to find one way ANOVA

2.2 One-way ANOVA for variable 'A' with respect to the variable 'Relief'

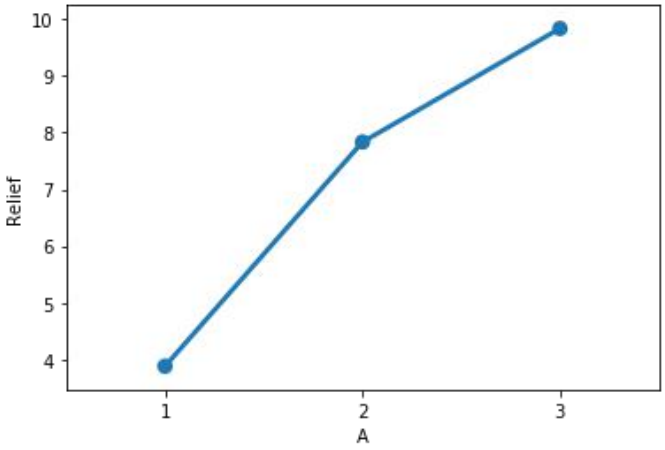
pvalue=4.578242e-07 is smaller than the level of significance α 0.05
So the null hypothesis is rejected based on the above observation for treatment 'A' with respect to 'Relief'

2.3 One-way ANOVA for variable 'B' with respect to the variable 'Relief'

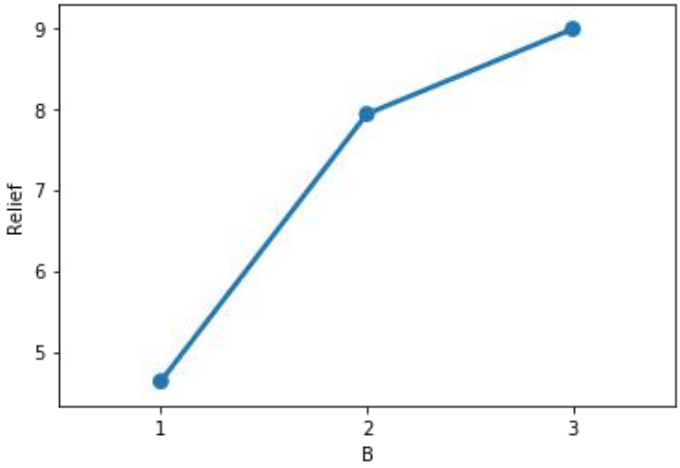
pvalue=0.00135 is smaller than the level of significance α 0.05
So the null hypothesis is rejected based on the above observation for treatment 'B' with respect to 'Relief'

2.4 Effects of one variable on another

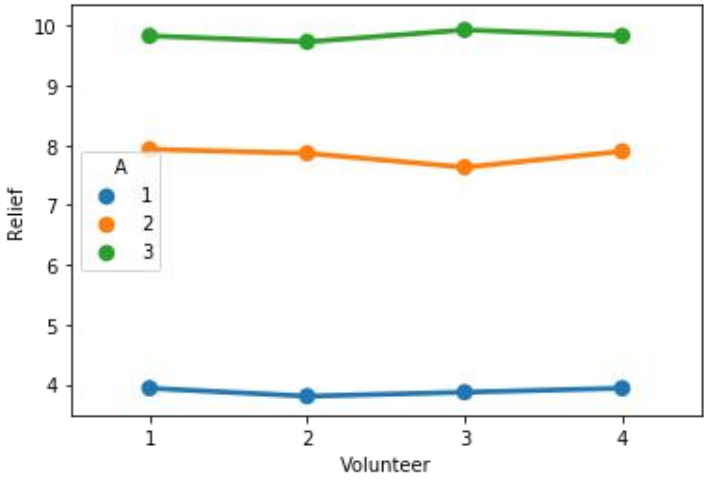
Interaction between treatment 'A' and 'Relief'



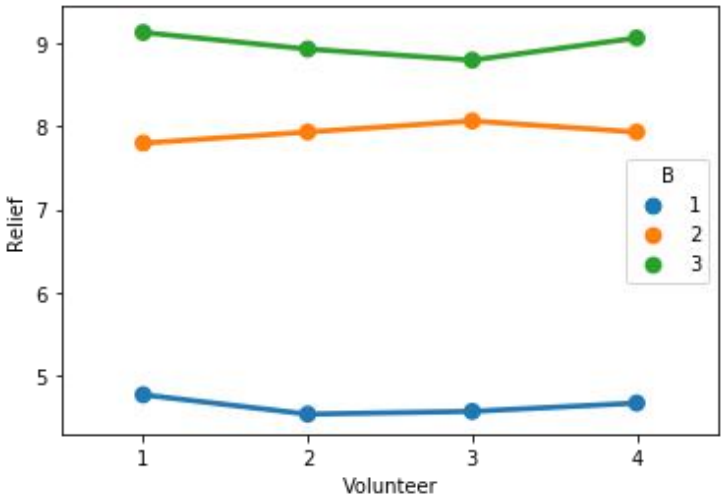
Interaction between treatment ‘B’ and ‘Relief’



Interaction between treatment ‘Volunteer’ and ‘Relief’ with respect A



Interaction between treatment ‘Volunteer’ and ‘Relief’ with respect B



2.5 Two-way ANOVA based on the different ingredients (variable 'A' & 'B') with the variable 'Relief'

pvalue=8.514029e-15(A) and 1.546749e-11(B) is smaller than the level of significance α 0.05
So the null hypothesis is rejected based on the above observation for treatment 'A' and 'B' with respect to 'Relief'

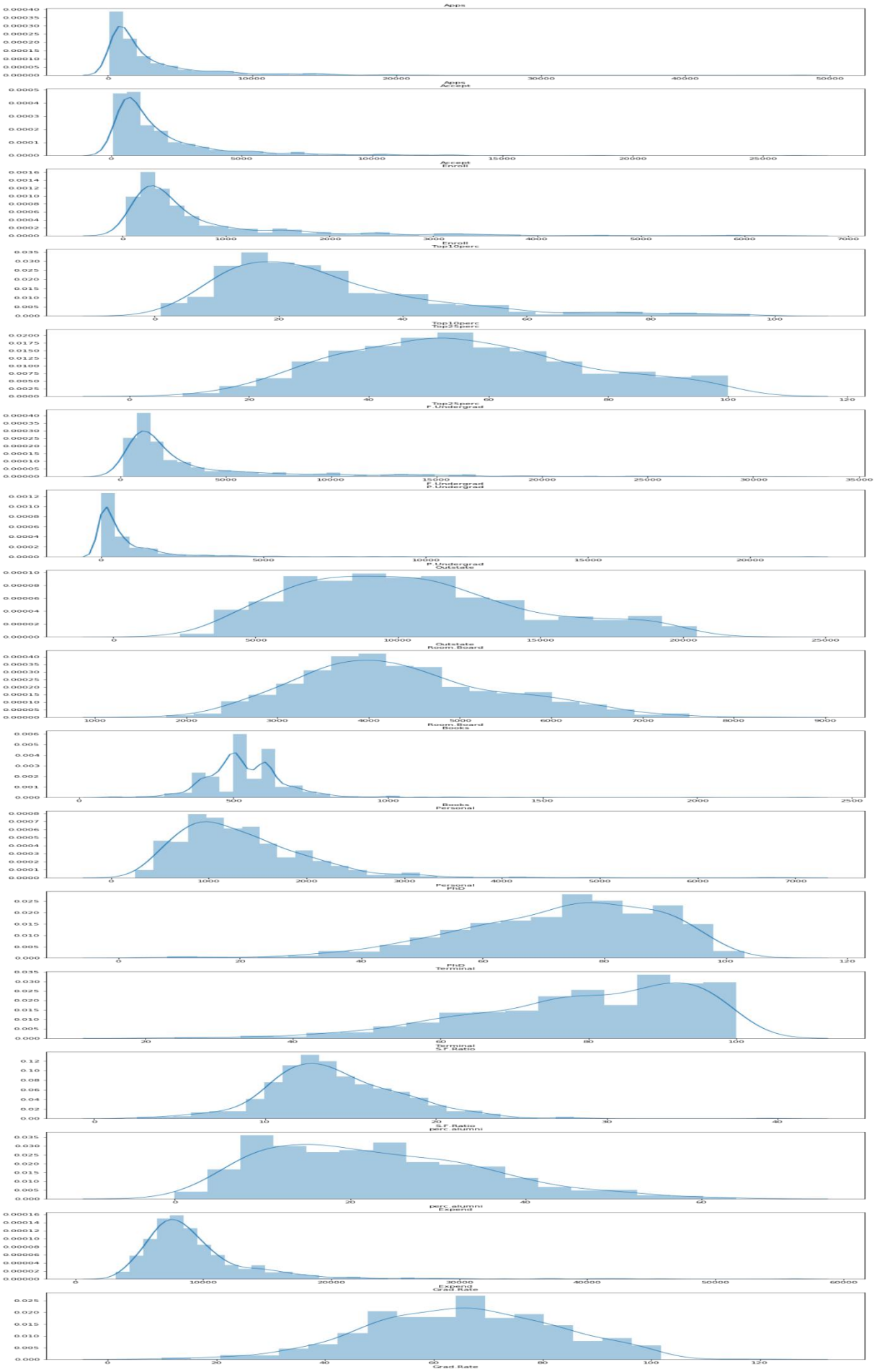
2.6 Business implications of performing ANOVA for Fever

The treatment of A and B doesn't make any significance difference individually and both combined on relief for those volunteers with this ANOVA analysis.

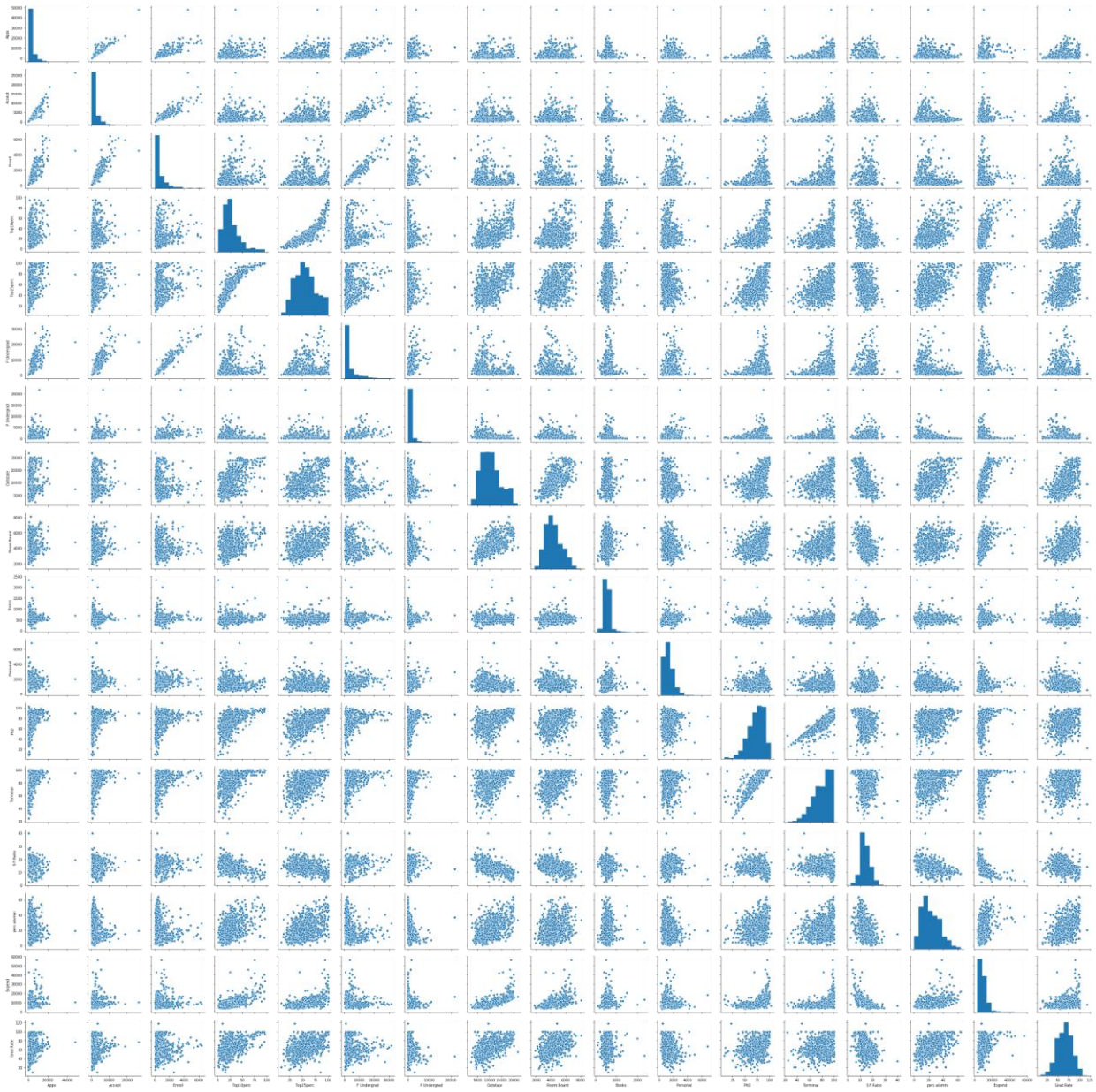
3 Education - Post 12th Standard

3.1 Perform Exploratory Data Analysis

Univariate analysis



Multivariate analysis



3.2 Scaling of variables

One type of scaling (Standard scaler)

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Termi
0	Abilene Christian University	-0.346882	-0.321205	-0.063509	-0.258583	-0.191827	-0.168116	-0.209207	-0.746356	-0.964905	-0.602312	1.270045	-0.163028	-0.115
1	Adelphi University	-0.210884	-0.038703	-0.288584	-0.655656	-1.353911	-0.209788	0.244307	0.457496	1.909208	1.215880	0.235515	-2.675646	-3.378
2	Adrian College	-0.406866	-0.376318	-0.478121	-0.315307	-0.292878	-0.549565	-0.497090	0.201305	-0.554317	-0.905344	-0.259582	-1.204845	-0.931
3	Agnes Scott College	-0.668261	-0.681682	-0.692427	1.840231	1.677612	-0.658079	-0.520752	0.626633	0.996791	-0.602312	-0.688173	1.185206	1.175
4	Alaska Pacific University	-0.726176	-0.764555	-0.780735	-0.655656	-0.596031	-0.711924	0.009005	-0.716508	-0.216723	1.518912	0.235515	0.204672	-0.523

Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
-0.115729	1.013776	-0.867574	-0.501910	-0.318252
-3.378176	-0.477704	-0.544572	0.166110	-0.551262
-0.931341	-0.300749	0.585935	-0.177290	-0.667767
1.175657	-1.615274	1.151188	1.792851	-0.376504
-0.523535	-0.553542	-1.675079	0.241803	-2.939613

Second type of scaling (Minmax scaler)

	Names	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal
0	Abilene Christian University	0.032887	0.044177	0.107913	0.231579	0.472527	0.087164	0.024548	0.263430	0.239596	0.157754	0.297710	0.652632	0.710526
1	Adelphi University	0.043842	0.070531	0.075035	0.157895	0.219780	0.080752	0.056148	0.513430	0.736129	0.291444	0.190840	0.221053	0.078947
2	Adrian College	0.028055	0.039036	0.047349	0.221053	0.450549	0.028473	0.004488	0.460227	0.310530	0.135472	0.139695	0.473684	0.552632
3	Agnes Scott College	0.006998	0.010549	0.016045	0.621053	0.879121	0.011776	0.002839	0.548554	0.578499	0.157754	0.095420	0.884211	0.960526
4	Alaska Pacific University	0.002333	0.002818	0.003146	0.157895	0.384615	0.003492	0.039753	0.269628	0.368852	0.313725	0.190840	0.715789	0.631579

S.F.Ratio	perc.alumni	Expend	Grad.Rate
0.418231	0.187500	0.072671	0.462963
0.260054	0.250000	0.138387	0.425926
0.278820	0.468750	0.104605	0.407407
0.139410	0.578125	0.298415	0.453704
0.252011	0.031250	0.145833	0.046296

3.3 Covariance and the correlation matrix

Covariance matrix

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD
Apps	0.006498	0.007099	0.009977	0.005072	0.006169	0.010108	0.002238	0.000840	0.002298	0.000786	0.001489	0.005413
Accept	0.007099	0.008714	0.012439	0.003336	0.005028	0.012564	0.002872	-0.000500	0.001467	0.000780	0.001939	0.005708
Enroll	0.009977	0.012439	0.021364	0.004921	0.007213	0.021708	0.005229	-0.004722	-0.001017	0.001212	0.004245	0.008327
Top10perc	0.005072	0.003336	0.004921	0.034480	0.036047	0.004039	-0.001364	0.021698	0.011925	0.001624	-0.001791	0.016973
Top25perc	0.006169	0.005028	0.007213	0.036047	0.047365	0.006683	-0.000813	0.022133	0.012472	0.001850	-0.001818	0.020419
F.Undergrad	0.010108	0.012564	0.021708	0.004039	0.006683	0.023704	0.006124	-0.006902	-0.001834	0.001309	0.005048	0.008424
P.Undergrad	0.002238	0.002872	0.005229	-0.001364	-0.000813	0.006124	0.004861	-0.003673	-0.000739	0.000417	0.002306	0.001787
Outstate	0.000840	-0.000500	-0.004722	0.021698	0.022133	-0.006902	-0.003673	0.043181	0.023503	0.000594	-0.006424	0.013679
Room.Board	0.002298	0.001467	-0.001017	0.011925	0.012472	-0.001834	-0.000739	0.023503	0.029885	0.001628	-0.003564	0.009781
Books	0.000786	0.000780	0.001212	0.001624	0.001850	0.001309	0.000417	0.000594	0.001628	0.005413	0.001364	0.000340
Personal	0.001489	0.001939	0.004245	-0.001791	-0.001818	0.005048	0.002306	-0.006424	-0.003564	0.001364	0.010685	-0.000194
PhD	0.005413	0.005708	0.008327	0.016973	0.020419	0.008424	0.001787	0.013679	0.009781	0.000340	-0.000194	0.029541
Terminal	0.005770	0.006104	0.008729	0.017666	0.022123	0.008948	0.001917	0.016423	0.012543	0.001425	-0.000613	0.028287
S.F.Ratio	0.000818	0.001746	0.003680	-0.007584	-0.006805	0.004570	0.001721	-0.012235	-0.006653	-0.000249	0.001496	-0.002381
perc.alumni	-0.001408	-0.002892	-0.005117	0.016376	0.017608	-0.006840	-0.003791	0.022783	0.009116	-0.000573	-0.005724	0.008287
Expend	0.002060	0.001146	0.000923	0.012081	0.011300	0.000283	-0.000574	0.013762	0.008538	0.000814	-0.000996	0.007322
Grad.Rate	0.001882	0.000999	-0.000519	0.014619	0.016521	-0.001929	-0.002850	0.018882	0.011684	0.000012	-0.004428	0.008339

Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
0.005770	0.000818	-0.001408	0.002060	0.001882
0.006104	0.001746	-0.002892	0.001146	0.000999
0.008729	0.003680	-0.005117	0.000923	-0.000519
0.017666	-0.007584	0.016376	0.012081	0.014619
0.022123	-0.006805	0.017608	0.011300	0.016521
0.008948	0.004570	-0.006840	0.000283	-0.001929
0.001917	0.001721	-0.003791	-0.000574	-0.002850
0.016423	-0.012235	0.022783	0.013762	0.018882
0.012543	-0.006653	0.009116	0.008538	0.011684
0.001425	-0.000249	-0.000573	0.000814	0.000012
-0.000613	0.001496	-0.005724	-0.000996	-0.004428
0.028287	-0.002381	0.008287	0.007322	0.008339
0.037526	-0.003291	0.010019	0.008367	0.008921
-0.003291	0.011262	-0.008279	-0.006099	-0.005177
0.010019	-0.008279	0.037489	0.007961	0.015118
0.008367	-0.006099	0.007961	0.009690	0.006111
0.008921	-0.005177	0.015118	0.006111	0.025298

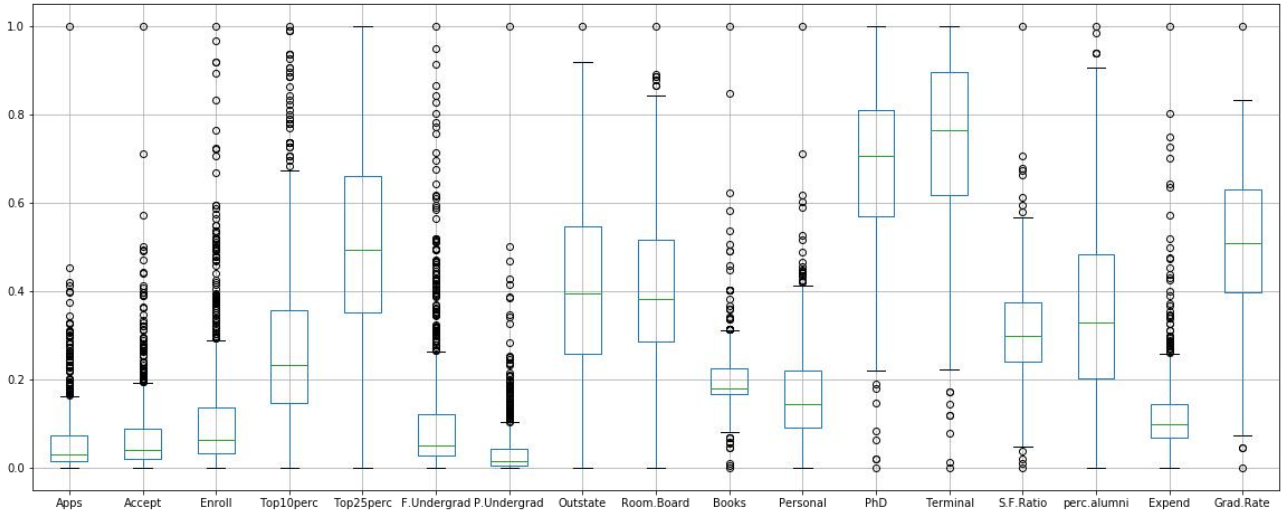
Correlation matrix

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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
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

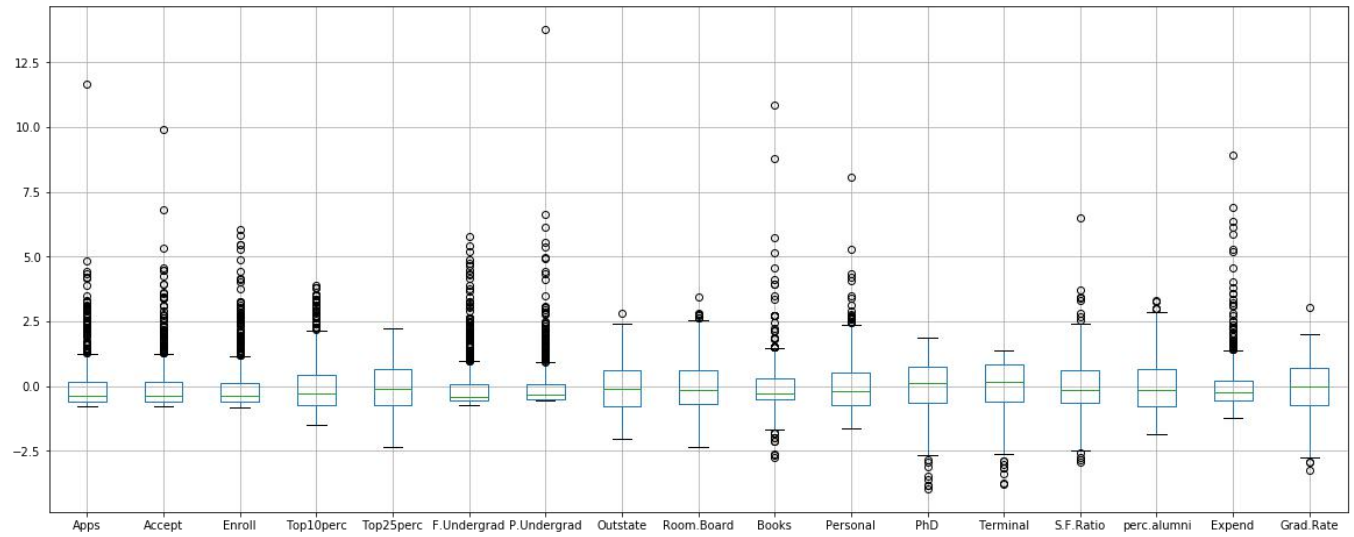
Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
0.369491	0.095633	-0.090226	0.259592	0.146755
0.337583	0.176229	-0.159990	0.124717	0.067313
0.308274	0.237271	-0.180794	0.064169	-0.022341
0.491135	-0.384875	0.455485	0.660913	0.494989
0.524749	-0.294629	0.417864	0.527447	0.477281
0.300019	0.279703	-0.229462	0.018652	-0.078773
0.141904	0.232531	-0.280792	-0.083568	-0.257001
0.407983	-0.554821	0.566262	0.672779	0.571290
0.374540	-0.362628	0.272363	0.501739	0.424942
0.099955	-0.031929	-0.040208	0.112409	0.001061
-0.030613	0.136345	-0.285968	-0.097892	-0.269344
0.849587	-0.130530	0.249009	0.432762	0.305038
1.000000	-0.160104	0.267130	0.438799	0.289527
-0.160104	1.000000	-0.402929	-0.583832	-0.306710
0.267130	-0.402929	1.000000	0.417712	0.490898
0.438799	-0.583832	0.417712	1.000000	0.390343
0.289527	-0.306710	0.490898	0.390343	1.000000

3.4 Dataset for outliers

Before scaling



After scaling



After scaling the outliers have also occurred before first inter quartile

3.5 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 Vectors

```
%s [[-2.48765602e-01  3.31598227e-01  6.30921033e-02 -2.81310530e-01
   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]]

Eigen Values

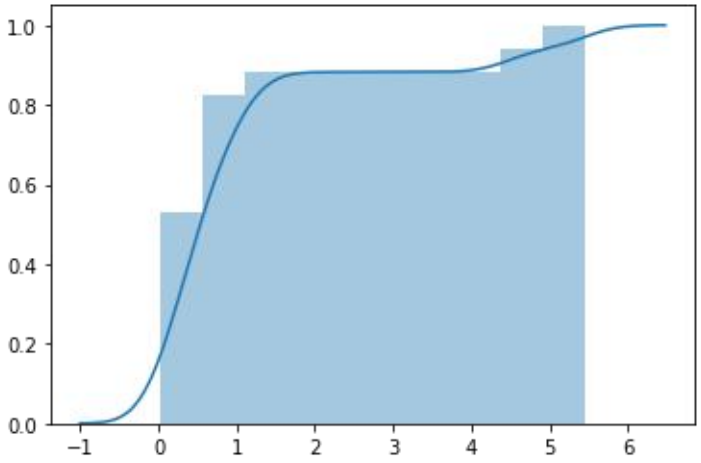
%s [5.45052162 4.48360686 1.17466761 1.00820573 0.93423123 0.84849117
0.6057878 0.58787222 0.53061262 0.4043029 0.02302787 0.03672545
0.31344588 0.08802464 0.1439785 0.16779415 0.22061096]

3.6 Explicit form of the first PC (in terms of Eigen Vectors)

The first PC in terms of eigen vector is 5.4505%

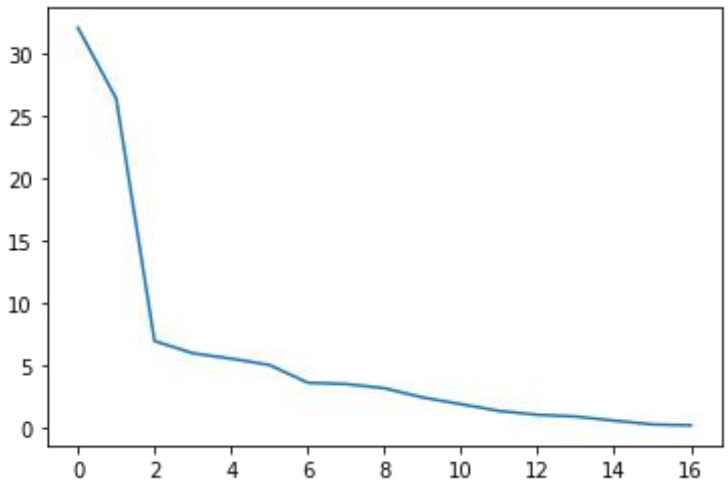
3.7 Principal Component scores into a data frame

Cumulative eigen values

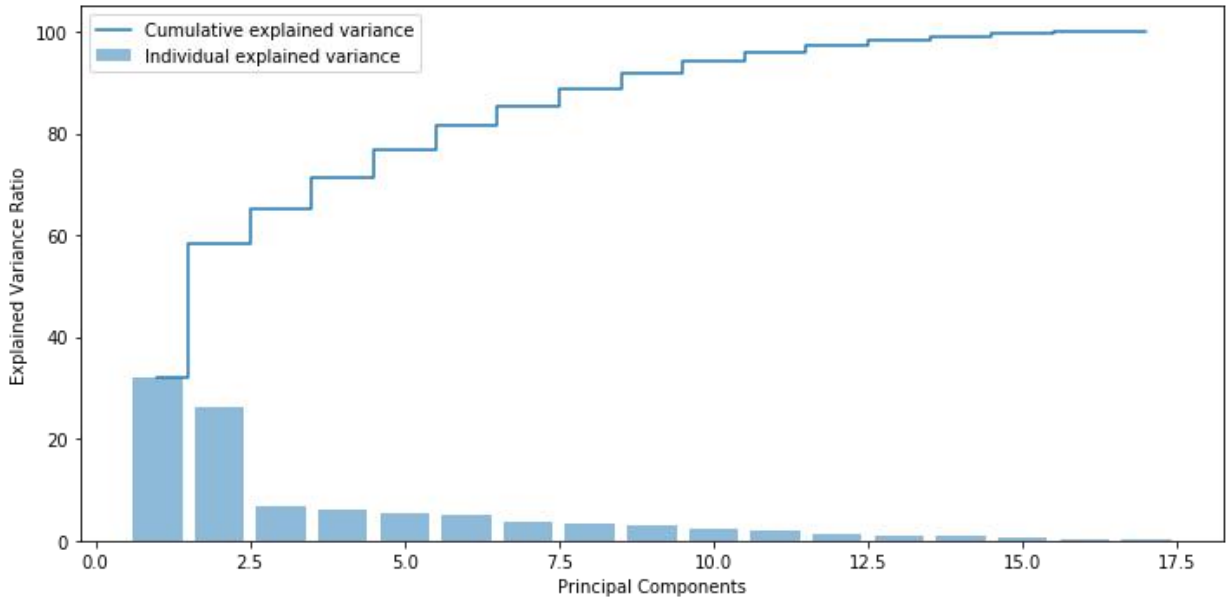


Cumulative Variance Explained [32.0206282 58.36084263 65.26175919 71.18474841 76.67315352
81.65785448 85.21672597 88.67034731 91.78758099 94.16277251
96.00419883 97.30024023 98.28599436 99.13183669 99.64896227
99.86471628 100.   ]

Plotting the variance



Visually we can observe that their is steep drop in variance explained with increase in number of PC's.



We are generating only 14 PCA dimensions (dimensionality reduction from 17 to 14), where eigen vector indicates the direction of the components

PCA Components

```
array([[ 0.2487656 ,  0.2076015 ,  0.17630359,  0.35427395,  0.34400128,
         0.15464096,  0.0264425 ,  0.29473642,  0.24903045,  0.06475752,
        -0.04252854,  0.31831287,  0.31705602, -0.17695789,  0.20508237,
         0.31890875,  0.25231565],
       [ 0.33159823,  0.37211675,  0.40372425, -0.08241182, -0.04477866,
         0.41767377,  0.31508783, -0.24964352, -0.13780888,  0.05634184,
         0.21992922,  0.05831132,  0.04642945,  0.24666528, -0.24659527,
        -0.13168986, -0.16924053],
       [-0.0630921 , -0.10124906, -0.08298557,  0.03505553, -0.02414794,
        -0.06139298,  0.13968172,  0.04659887,  0.14896739,  0.67741165,
         0.49972112, -0.12702837, -0.06603755, -0.2898484 , -0.14698927,
         0.22674398, -0.20806465],
       [ 0.28131053,  0.26781735,  0.16182677, -0.05154725, -0.10976654,
         0.10041234, -0.15855849,  0.13129136,  0.18499599,  0.08708922,
        -0.23071057, -0.53472483, -0.51944302, -0.16118949,  0.01731422,
         0.07927349,  0.26912907],
       [ 0.00574141,  0.05578609, -0.05569364, -0.39543434, -0.42653359,
        -0.04345437,  0.30238541,  0.222532  ,  0.56091947, -0.12728883,
        -0.22231102,  0.14016633,  0.20471973, -0.07938825, -0.21629741,
         0.07595812, -0.10926791],
       [-0.01623744,  0.00753468, -0.04255798, -0.0526928 ,  0.03309159,
        -0.04345423, -0.19119858, -0.03000039,  0.16275545,  0.64105495,
        -0.331398  ,  0.09125552,  0.15492765,  0.48704587, -0.04734001,
        -0.29811862,  0.21616331],
       [-0.04248635, -0.01294972, -0.02769289, -0.16133207, -0.11848556,
        -0.02507636,  0.06104235,  0.10852897,  0.20974423, -0.14969203,
         0.63379006, -0.00109641, -0.02847701,  0.21925936,  0.24332116,
```

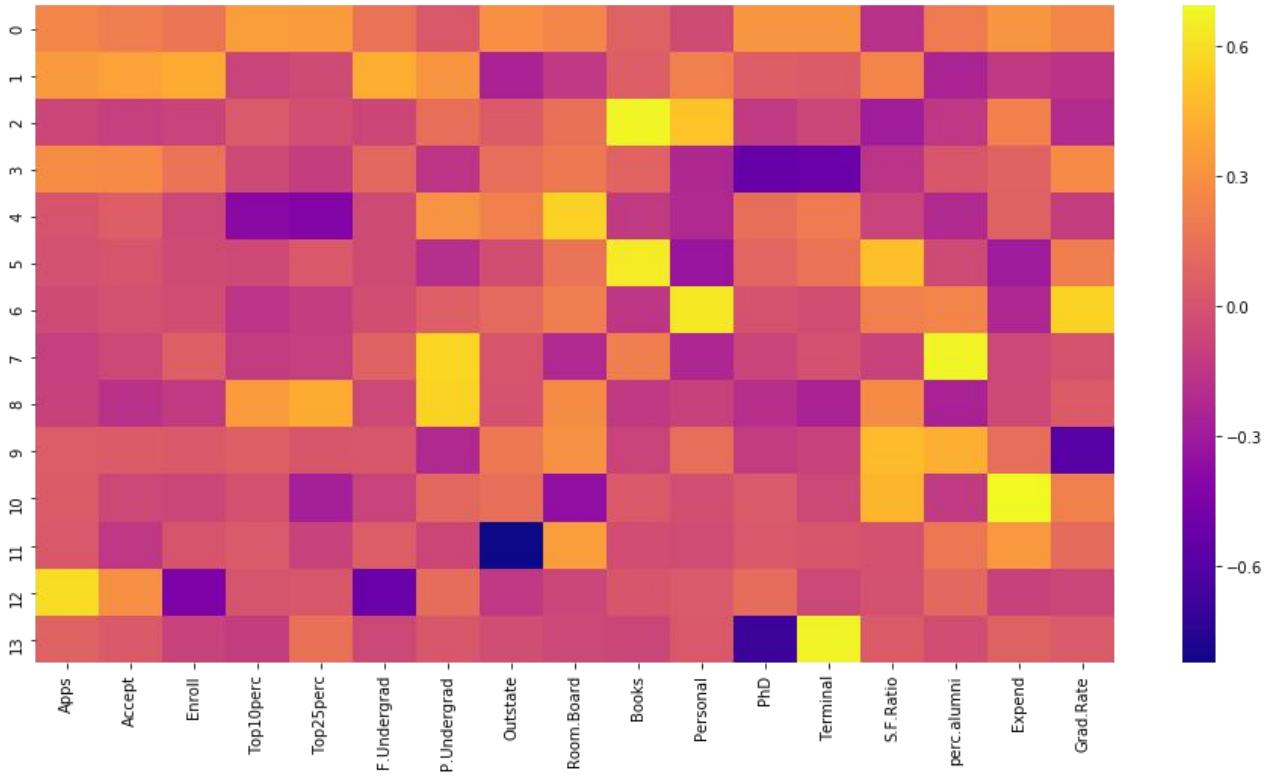
-0.22658448, 0.55994394],
[-0.1030904 , -0.05627096, 0.05866236, -0.12267803, -0.10249197,
 0.07888964, 0.57078382, 0.009846 , -0.22145344, 0.21329301,
 -0.23266084, -0.07704   , -0.01216133, -0.08360487, 0.67852365,
 -0.05415938, -0.00533554],
[-0.09022708, -0.17786481, -0.12856071, 0.34109986, 0.40371199,
 -0.05944192, 0.5606729 , -0.00457333, 0.27502255, -0.13366335,
 -0.09446889, -0.18518152, -0.2549382 , 0.27454438, -0.25533491,
 -0.04913888, 0.04190431],
[0.0525098 , 0.04114008, 0.03448791, 0.06402578, 0.01454923,
  0.02084718, -0.22310581, 0.18667536, 0.29832424, -0.08202922,
  0.13602762, -0.1234522 , -0.08857846, 0.47204525, 0.42299971,
  0.13228633, -0.59027107],
[0.04304621, -0.05840559, -0.06939888, -0.00810481, -0.27312847,
 -0.08115782, 0.10069332, 0.14322067, -0.35932173, 0.03194004,
 -0.01857847, 0.04037233, -0.0589734 , 0.44500073, -0.13072798,
  0.69208887, 0.219839],
[0.02407091, -0.14510245, 0.01114315, 0.0385543 , -0.08935156,
  0.05617677, -0.06353607, -0.82344378, 0.35455973, -0.02815937,
 -0.03926403, 0.02322243, 0.01648504, -0.01102621, 0.18266065,
  0.3259823 , 0.1221067],
[0.59583097, 0.2926424 , -0.44463821, 0.00102304, 0.02188388,
 -0.52362227, 0.12599765, -0.14185601, -0.06974859, 0.011438 ,
  0.03945474, 0.12769638, -0.05831347, -0.01771527, 0.10408809,
 -0.09374645, -0.06919698],
[0.0806328 , 0.03346743, -0.08569672, -0.10782819, 0.15174211,
 -0.05637288, 0.01928575, -0.03401154, -0.05842898, -0.06684946,
  0.02752862, -0.69112615, 0.67100861, 0.0413741 , -0.02715421,
  0.07312252, 0.03647674]])

The following table shows the dataframe of PCA components

	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.R
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	-0.176
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	0.246
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	-0.289
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	-0.161
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	-0.079

S.F.Ratio	perc.alumni	Expend	Grad.Rate
-0.176958	0.205082	0.318909	0.252316
0.246665	-0.246595	-0.131690	-0.169241
-0.289848	-0.146989	0.226744	-0.208065
-0.161189	0.017314	0.079273	0.269129
-0.079388	-0.216297	0.075958	-0.109268

Heatmap of PCA Component



3.8 Business implication of using the Principal Component Analysis for Education - Post 12th Standard

The data has been reduced its dimensions from 17 to 14 with respect to their cumulative variance.

4 Appendix A – Source Code

Karthiheswar_Adv Stat.ipynb