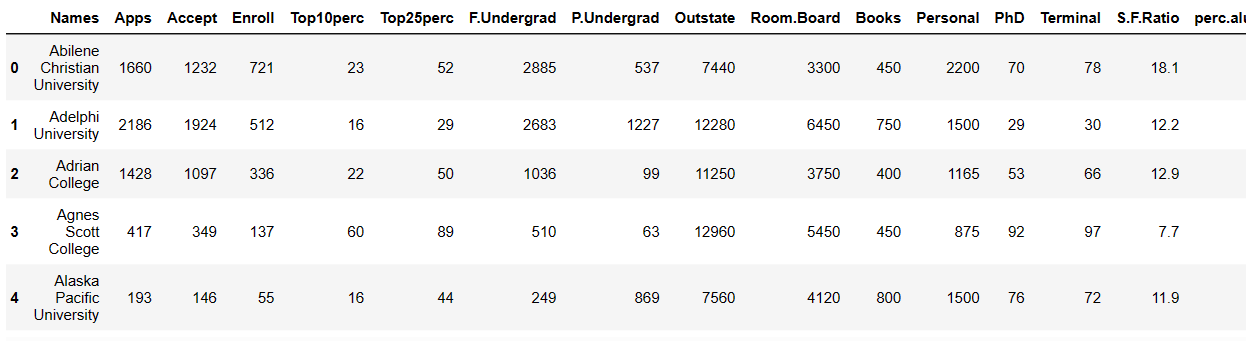
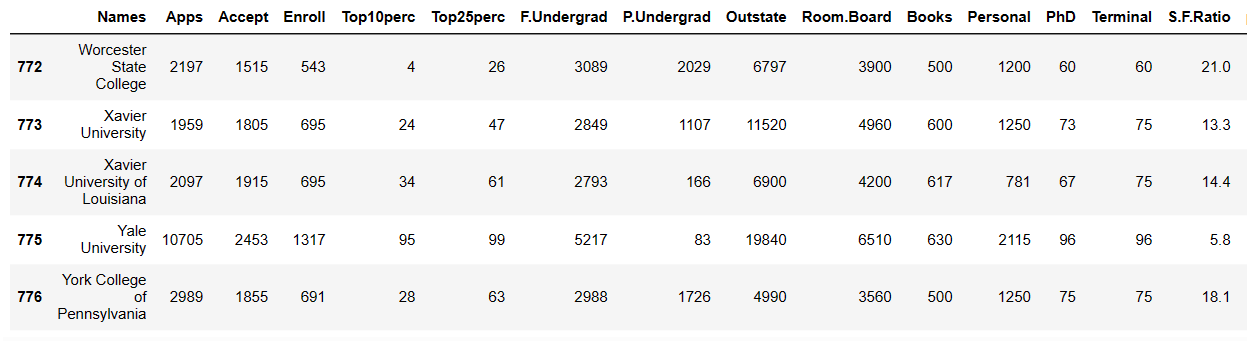
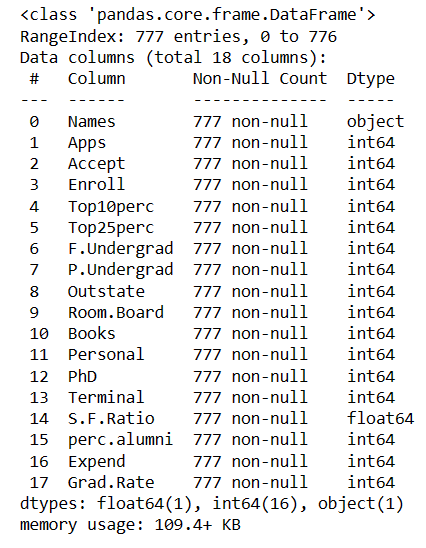
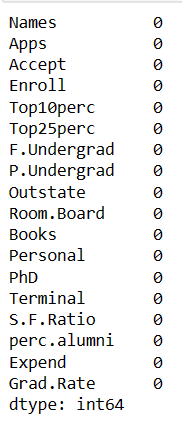
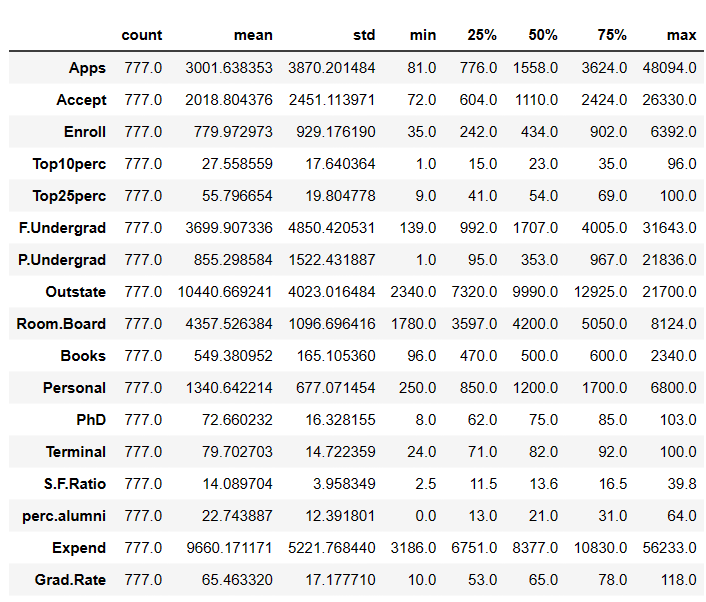
**SMDM Extended Project**

**Part 2**

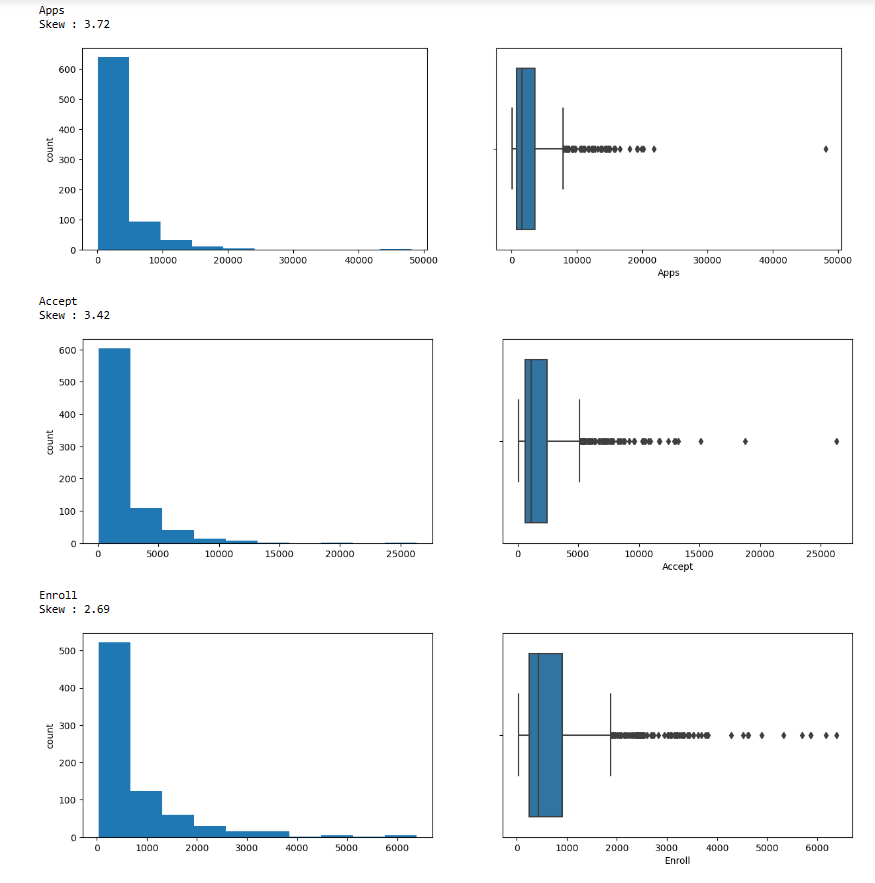
**The dataset Education - Post 12th Standard.csv (attached) contains information on various colleges. The data dictionary of the 'Education - Post 12th Standard.csv' can be found in the following file: Data Dictionary.xlsx (attached)Perform Exploratory Data Analysis [Univariate, Bivariate, and Multivariate analysis to be performed]. What insight do you draw from the EDA.**

**Ans:-**

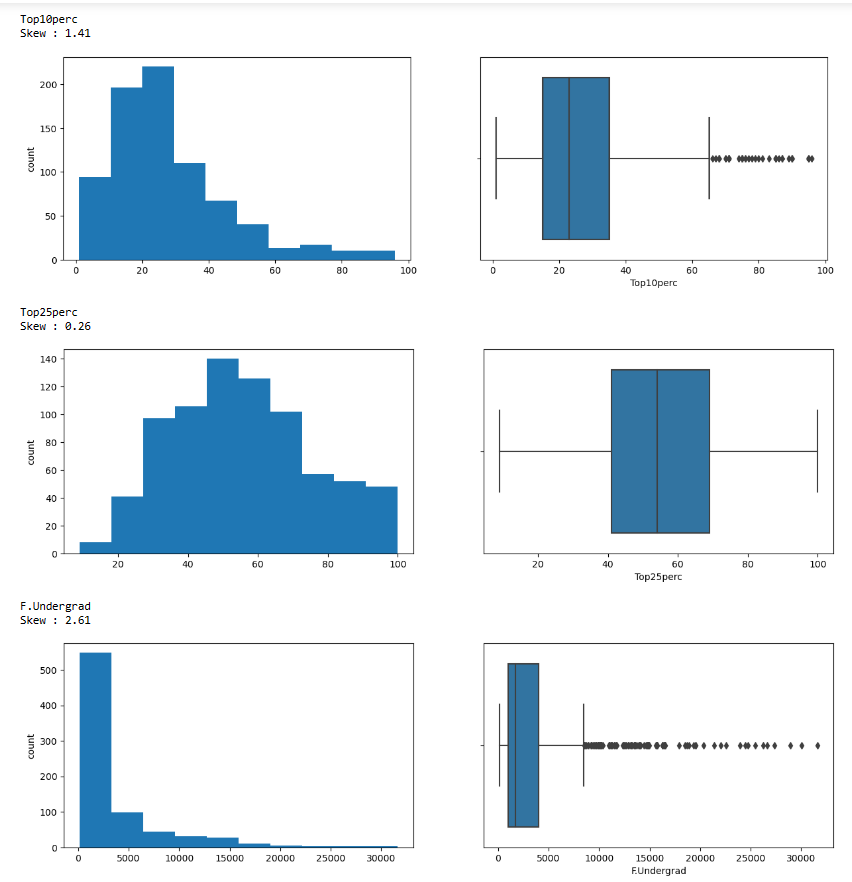
1. We have loaded required library and loaded the dataset Education - Post 12th Standard.
2. We have 777 rows and 18 columns.
3. We have We have viewed first and last few rows using head() and tail() functions respectively.
4. 
5. 
6. We can view dataset information using info() function
7. 
8. We have total 16 variables with integer data type, 1 variable with float datatype and 1 variable with object datatype.
9. We have no null value and no duplicate records.
10. 
11. We can describe data using Describe().
12. 
13. Observation:
14. We can see there total 777 rows in each fields with no null value.
15. We can find average data,minimum,maximum , total count of each field from above graph.

Univariate Ananlysis

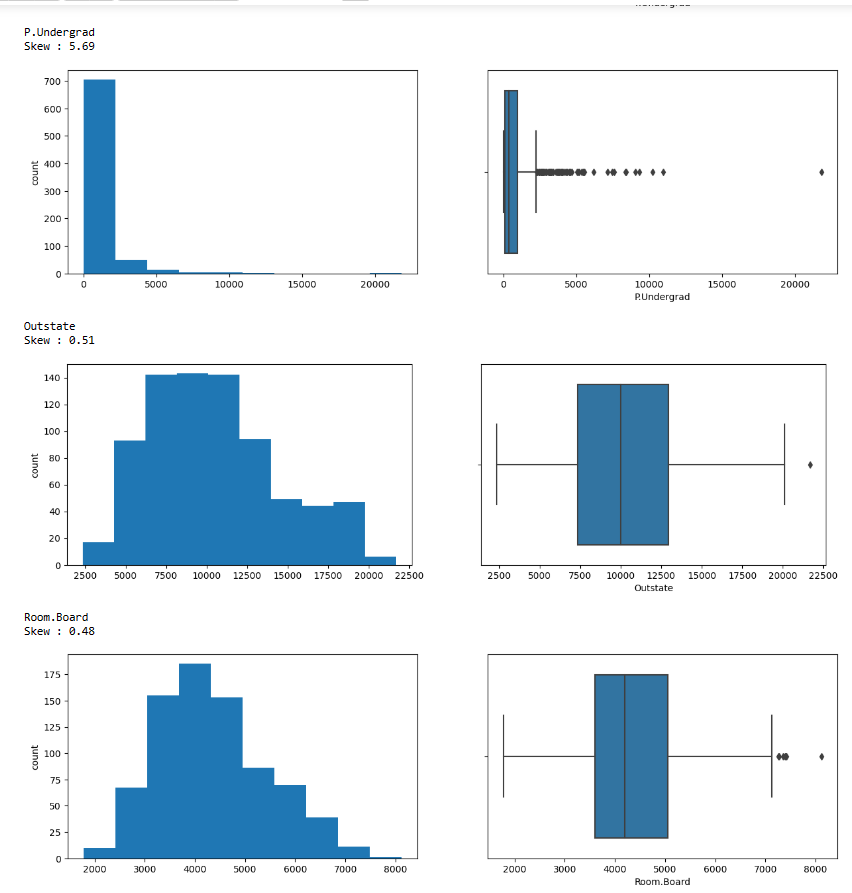
1. For Numeric values



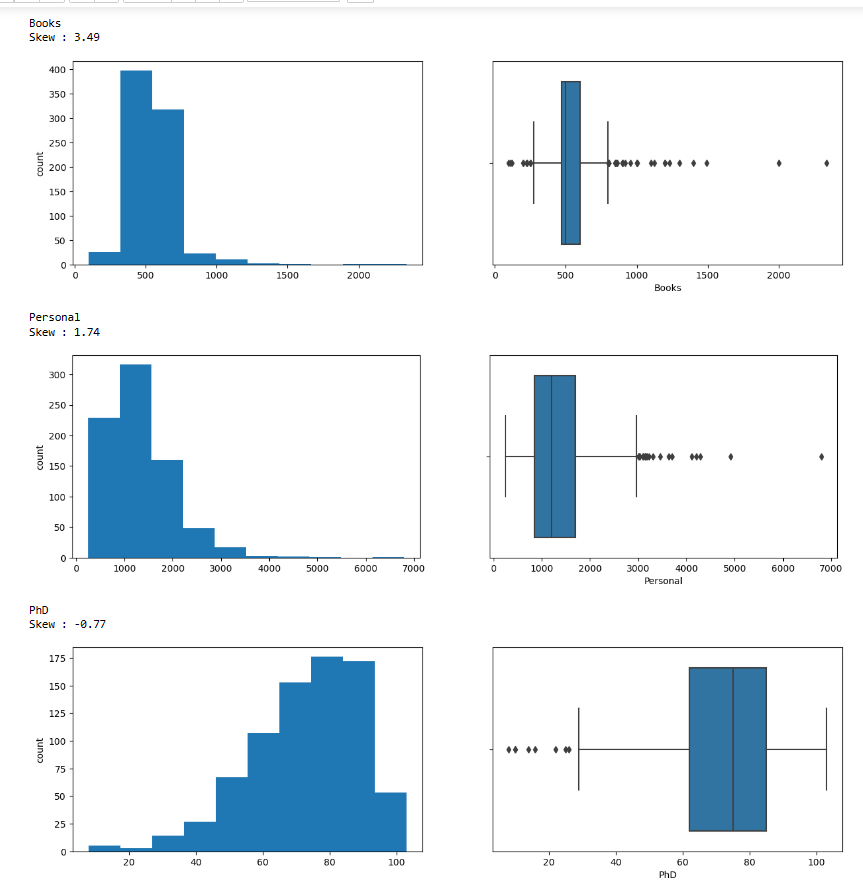
1. Boxplot of Apps,Accept and Enroll variable shows that they have Outliers.
2. All the three variables are right skewed.
3. The highest number of application is 50000,Highest number of Accepted Applications is around 28k and highest number of enrolled students is around 64k.



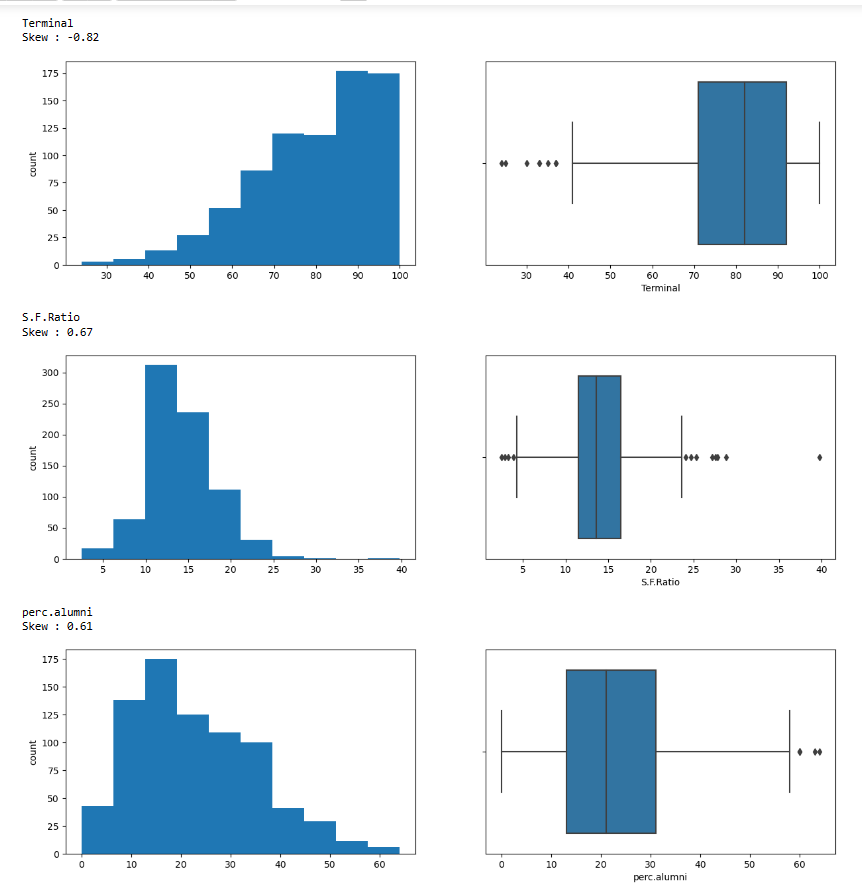
1. Top10perc and F.Undergrad have outliers in them. They are both right skewed.
2. Top25perc has no outlier. It has skewness of 0.26 . The distribution is not perfectly normal.
3. Highest number of full time graduates is around 33k.



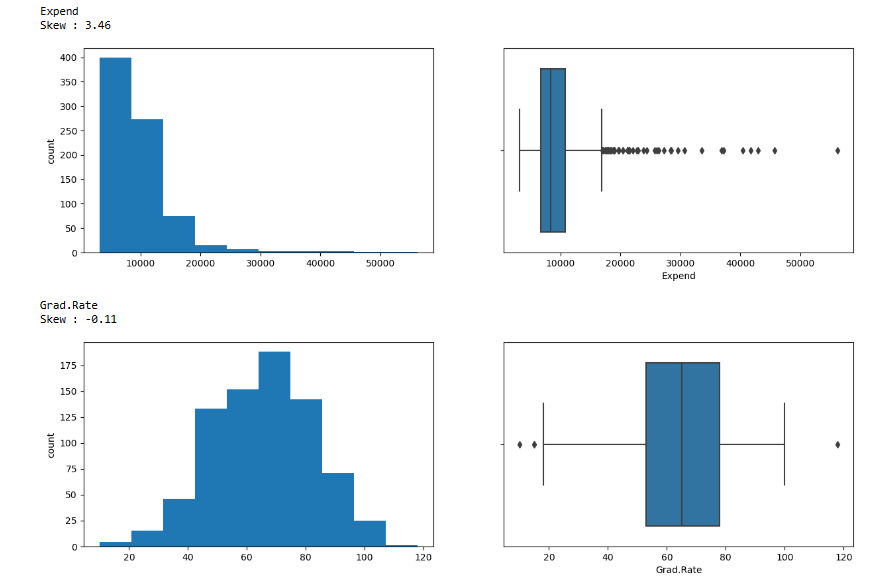
1. Number of part time undergraduate students,Outstate and Room.Board have outliers.
2. Outstate variable and Room.Board have very few outliers is not very skewed with skewness of +0.51 and +0.48 respectively.
3. Outstate variable and Room.Board does not have perfect normal distribution.
4. P.Undergrad variable is highly right skewed with skewness of +5.69.



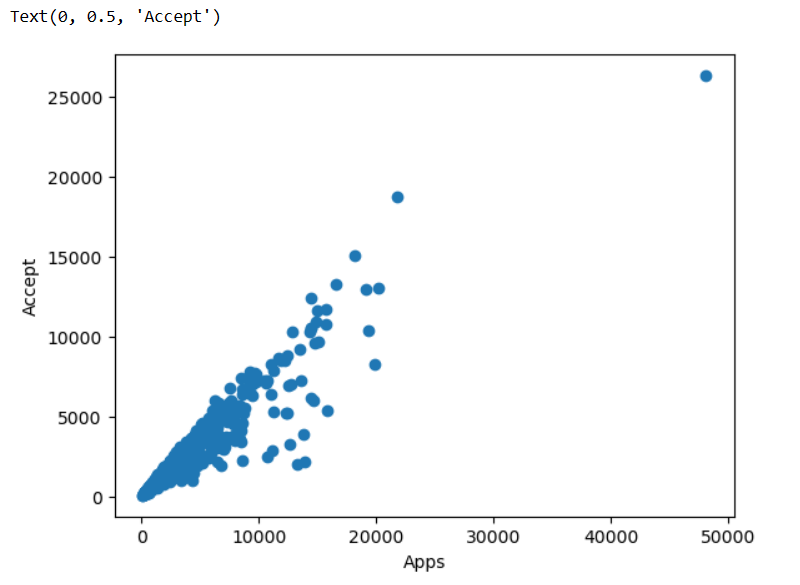
1. Books,Personal and PhD variables have outliers.
2. Books and Personal are Right skewed. PhD is left skewed with skewness of -0.77 .
3. Highest number of books are around 2500 and highest Personal is around 7k.



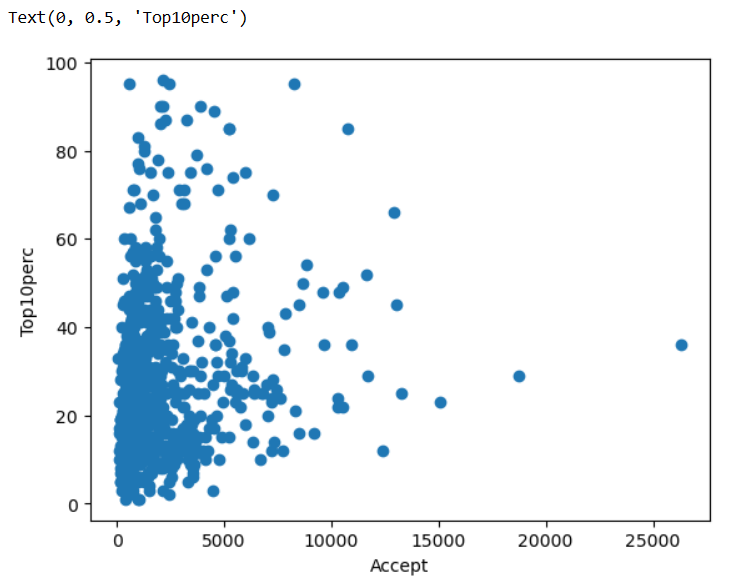
1. Terminal,S.F.Ratio and perc.alumni have outliers.
2. Terminal variable is negatively skewed.
3. Remaining two are right skewed.
4. Highest S.F.Ratio is around 40 and highest perc.alumini is around 63.



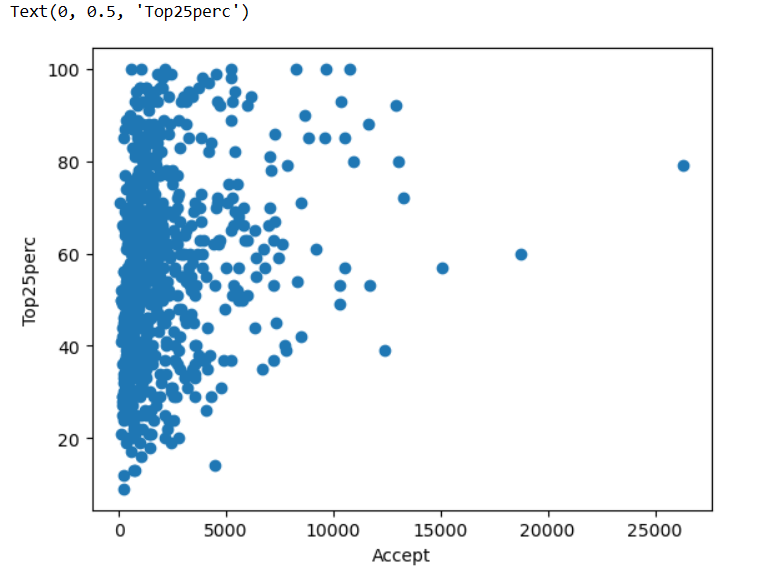
1. Expend and Grad.Rate have outliers.
2. Expend is right skewed with +3.46 and Grad.Rate is Left skewed with -0.11
3. Highest expenditure per student is around 57k.
4. Grad.Rate lies between 15 to 120.
5. For Bivariate Analysis



We can say that number of Accepted Application is positively correlated with number of Applications.

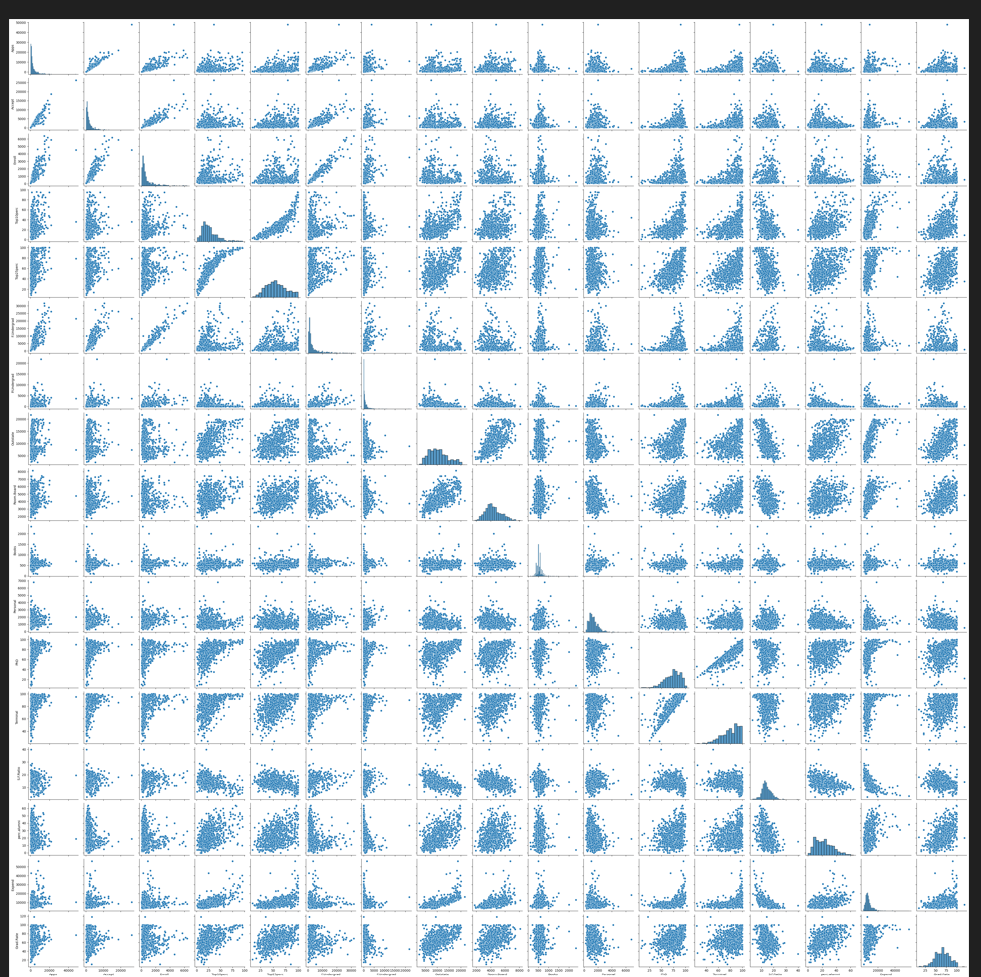


There is no relation between Top10perc and Accepted applications.

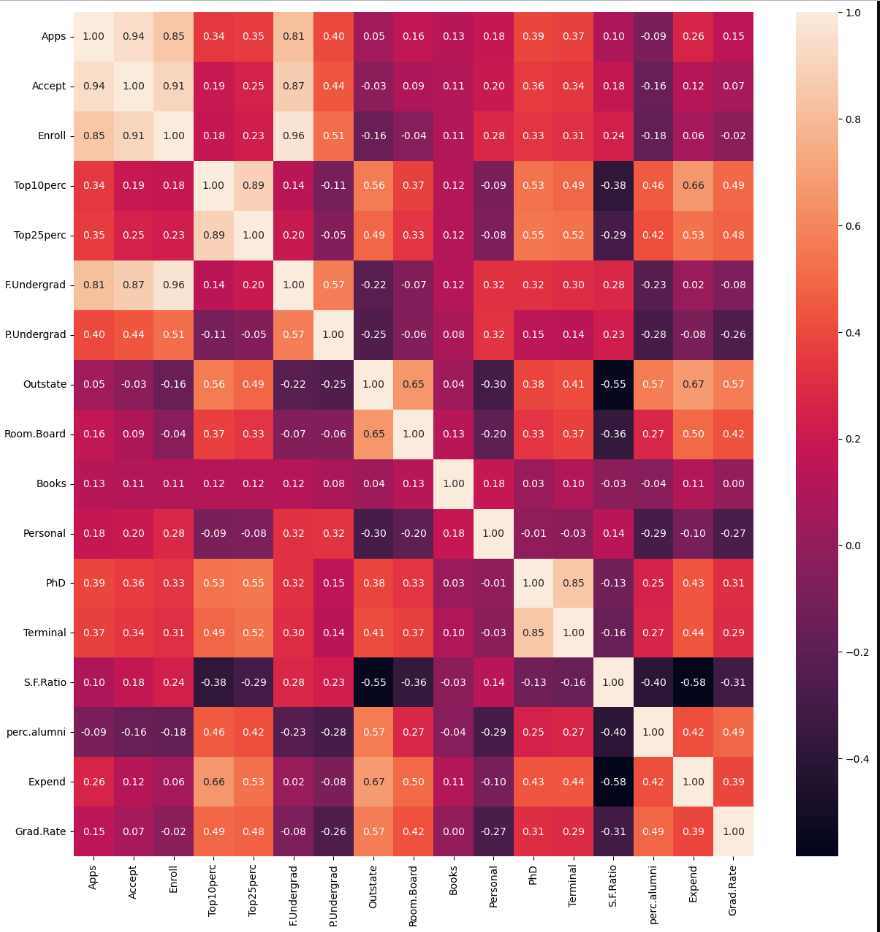


There is no relation between Top25perc and Accepted applications.

1. Multivariate Analysis.



With the help of pairplot we can understand relationship between different variables.Most of the variables are not related to each other. Very few variables are positively related to each other.



1. Apps has high positive correlation with Accept,Enroll and F.Undergrad.
2. Top10perc is highely correlated with Top25perc. It means that students who have topped in 10th are highly likely to top in 12th.
3. Outstate is mildly correlated with Room.Board.Outstate students are likely to take Room for Boarding.
4. phD and terminal are correrated positively.
5. Expend and S.F.Ratio is mildly negatively correlated to each other.S.F.Ratio is negatively related to OutState.

There are outliers in some of the variables.

