OBJECTIVE

Here initially I have to do the data cleaning and then do the visualizations on the given data. By doing visualization itself we can find some properties in the data. Then I will try building some machine learning models on the given data. Then I will compare the accuracy scores I obtained with different data. From them, I will choose the model that gives the best accuracy score.

INTRODUCTION

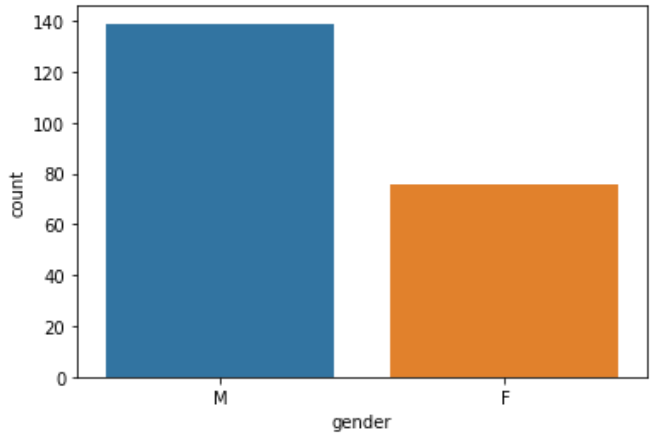
Here we have a dataset that contains placement data of students in a XYZ campus. It includes the educational information of each student, information about whether the student got placed or not, and the salary offered to the student.

The columns given in the dataset represents the following:

- sl_no :- serial number
- Gender:- gender identification M or F (Male or Female)
- ssc_p:- SSC (Senior Secondary School) percentage
- ssc_b:- information about SSC(Senior Secondary School) board
- hsc_p:- HSC(Higher Secondary School) percentage
- hsc_b:- information about HSC(Higher Secondary School) board
- hsc_s:- At which Branch the student completed the HSC?. Commerce, science, and art
- degree_p:- Under Graduation Degree percentage
- degree_t:- At which branch completed the Under Graduation degree? .science &technology and commerce &management.
- workex:- Whether the student have any work experience?. 'YES' or 'NO'
- etest_p:- test percentage
- specialisation:- branch specialization of the student. i.e. Marketing & Human Resource, Marketing &Finance
- mba_p:- MBA percentage obtained by the student
- Status:- About placement, he/she placed or not placed.

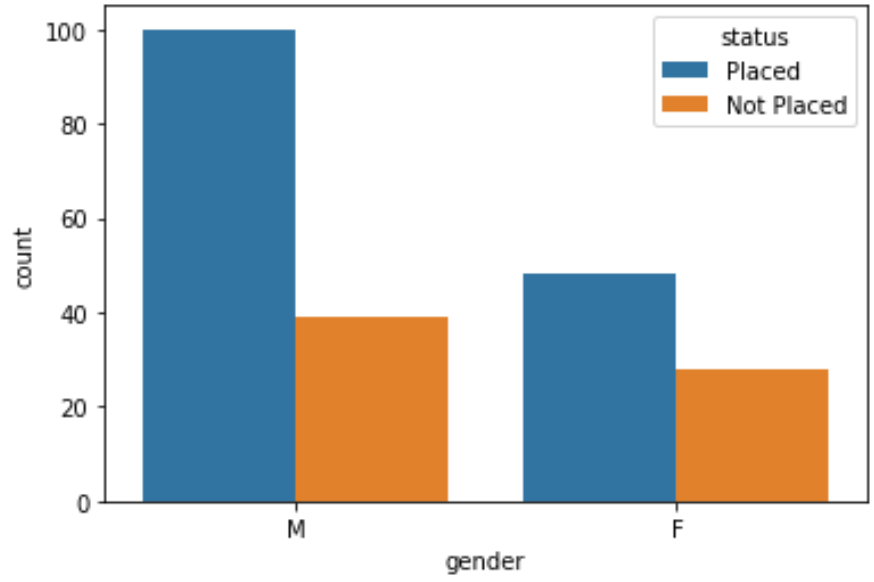
CHARTS, TABLES AND DIAGRAMS

PLOTTING THE DATA ON NUMBER OF MALES AND FEMALES IN OUR DATASET.



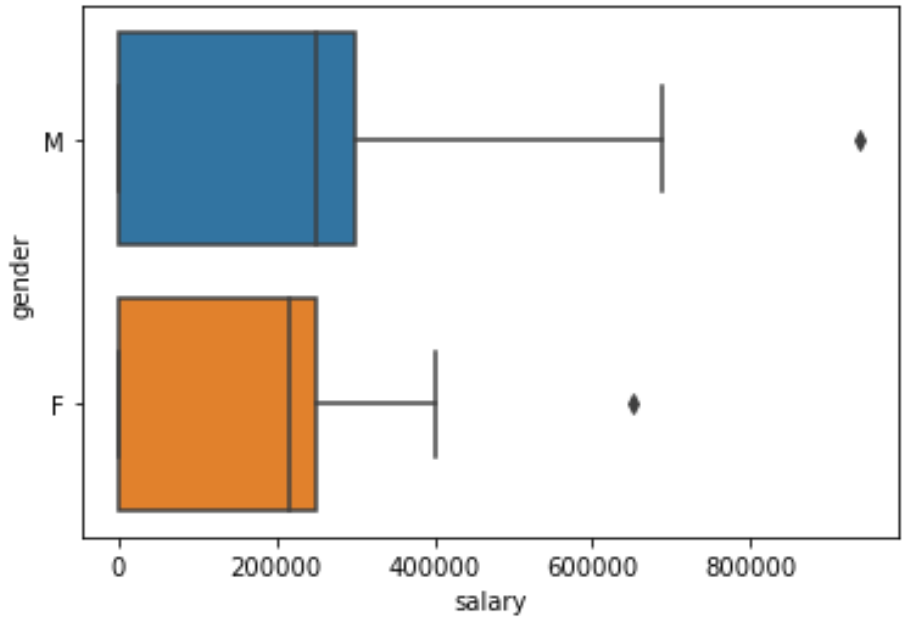
- From this plot, we can say that the number of males in the dataset is almost double the total number of females.

PLOT CREATED BASED ON WHETHER EACH STUDENT FROM DIFFERENT GENDERS GOT PLACED OR NOT



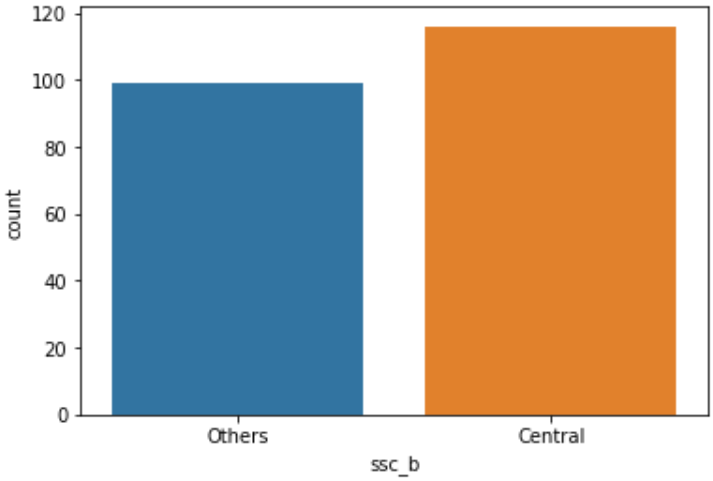
- From this plot, we can say that the number of female students that are not placed are almost half as the number of female students that are placed.
- Among the Male students the number of students placed is significantly higher. The number of students not placed is not reaching the halfway mark of the number of students that are placed.

BOXPLOT CREATED BASED ON SALARY OFFERED TO STUDENTS IN DIFFERENT GENDER.



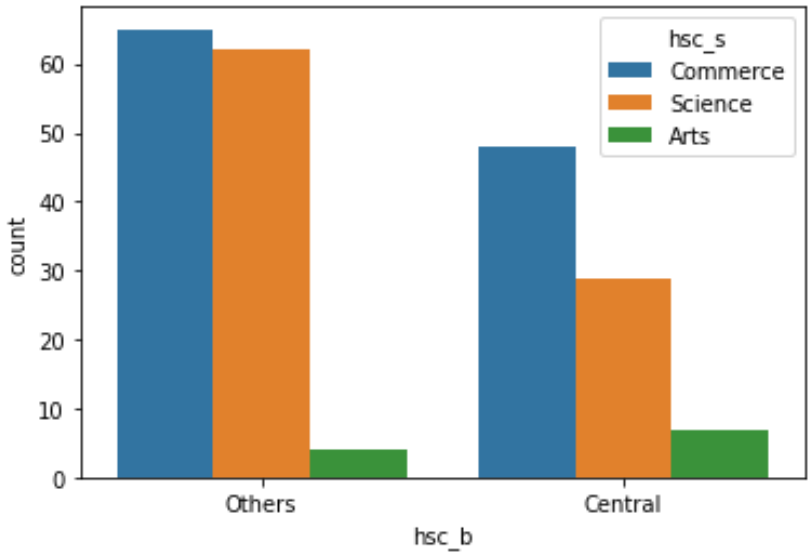
- Here we can see the highest salary is offered to a male candidate.
- Among the salaries offered to male students there is only one outlier.
- Among the salaries offered to female students there is also only one outlier.

PLOT CREATED BASED ON DATA OF STUDENTS FROM DIFFERENT SSC BOARDS.



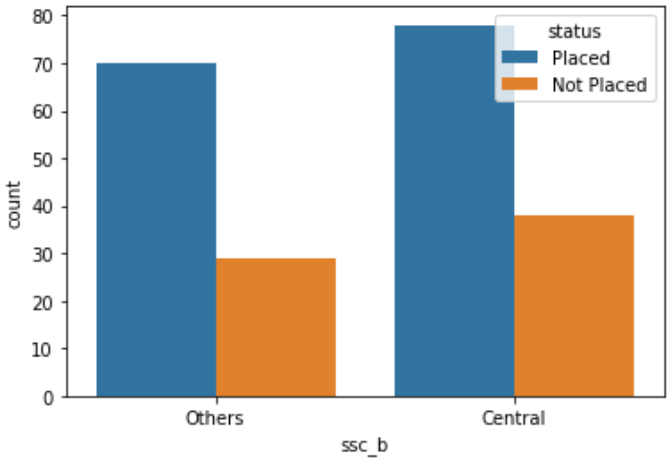
- From this, we can say that most of the students have completed their senior secondary schools from the SSC board.

PLOT CREATED BASED ON DATA OF STUDENTS FROM DIFFERENT HSC BOARDS.



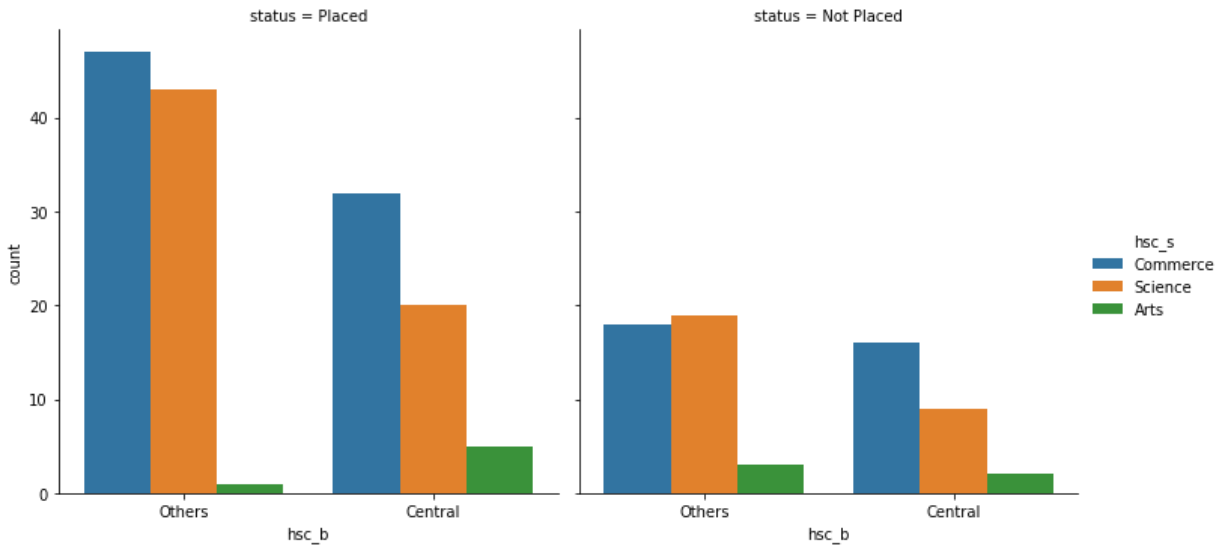
- From this, we can say that most of the students have completed their HSC from other boards than Central boards and most of the students are from the commerce branch and science branch.

PLOT CREATED BASED ON DATA OF STUDENTS FROM WHICH SSC BOARDS GOT MORE PLACEMENTS.



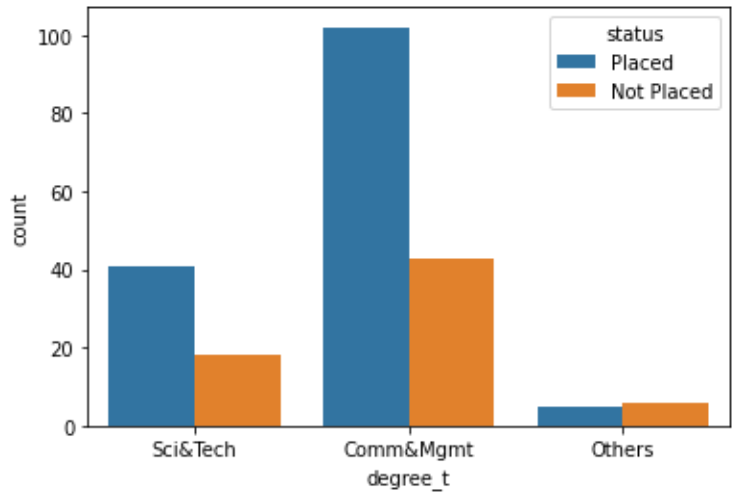
- We can see that more students are from central boards and the number of students who are placed and from central boards are higher than the number of students who are placed and not from central boards.

PLOT CREATED BASED ON DATA OF STUDENTS FROM DIFFERENT BOARDS, BRANCH AND THEIR PLACEMENT STATUS.



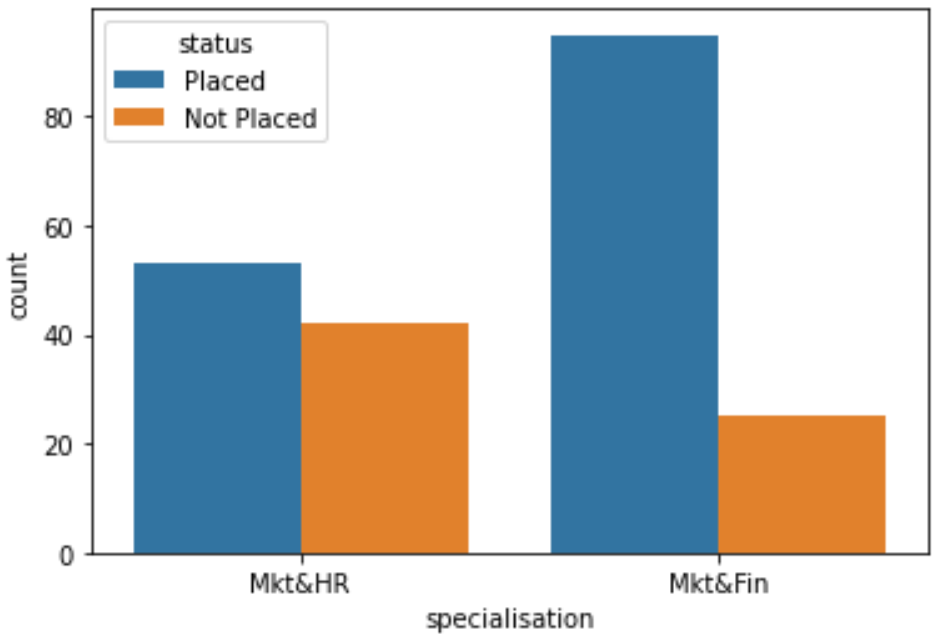
- From this, we can say that among the students who have got placed, more students have completed their HSC from other boards than central board and most of them are from Commerce and Science backgrounds.

PLOT CREATED BASED ON DATA OF STUDENT'S UG DEGREE AND PLACEMENT STATUS



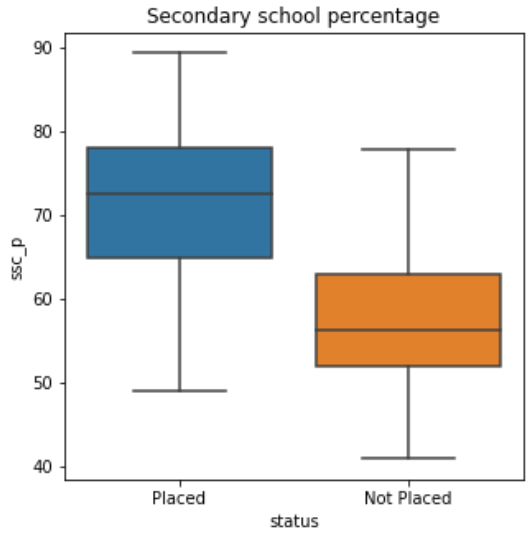
- From this, we can say that the number of students who have completed their UG degree in commerce and management have got placed are more than the number of students who have completed their UG degree in science and technology.

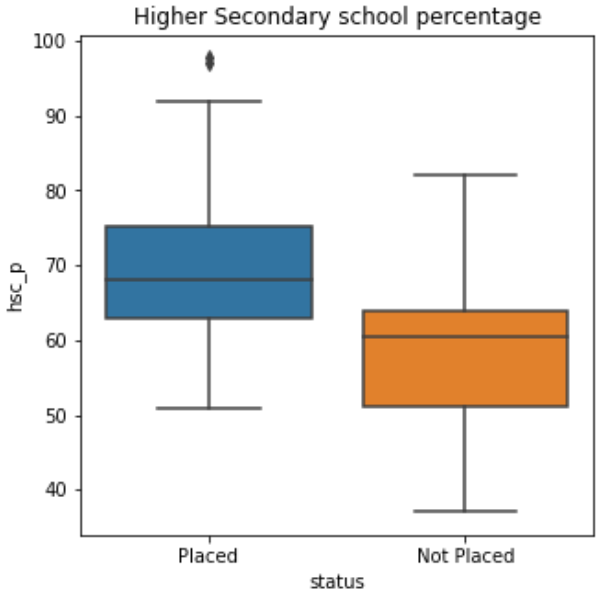
PLOT CREATED BASED ON THE DATA OF PLACEMENT STATUS OF STUDENTS FROM DIFFERENT SPECIALIZATIONS

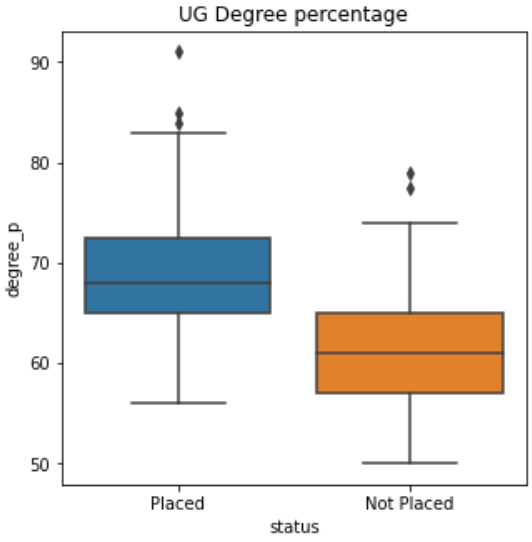


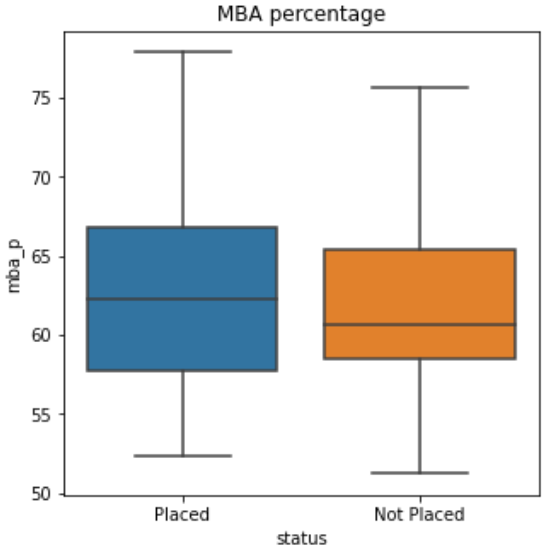
- From this, we can say that students in marketing and finance specializations were more in number and also they were higher in number of getting placed.

PLOTS COMPARING PERCENTAGES OBTAINED DURING DIFFERENT GRADUATIONS AND THE PLACEMENT STATUS

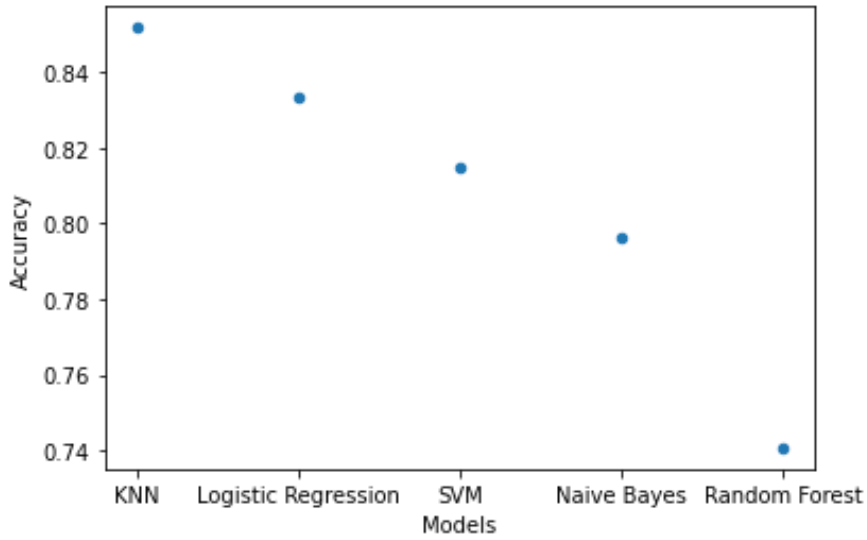



Higher Secondary school percentage
hsc_p
100
90
80
70
60
50
40
Placed
Not Placed
status


UG Degree percentage
degree_p
90
80
70
60
50
Placed
Not Placed
status


MBA percentage
mba_p
75
70
65
60
55
50
Placed
Not Placed
status

PLOT CREATED AFTER EVALUATING THE PERFORMANCE OF DIFFERENT MACHINE LEARNING MODELS TO PREDICT THE PLACEMENT STATUS BASED ON OTHER VARIABLES.



- From this we can say that KNN is the best model we can use for this dataset.

CONCLUSION

We were able to create a KNN (k-nearest neighbors) model for this dataset and we were able to find that our model have an accuracy of 0.8518518518518519.