**Campus Recruitment Project**

**The following are the features Present in dataset and I worked on this project to bring out good conclusions statistically with low bias and low variance :**

I Worked on Predicting the Placement Status and built good understanding over all the variables with **HIGHCHART visualization** tool. I worked completely on Classifying the Placement Status in the Data.

* 'ssc\_p' : 10th Grade percentage
* 'ssc\_b' : 10th Grade board
* 'hsc\_p' : 12th Grade percentage
* 'hsc\_b' : 12th Grade board
* 'hsc\_s' : Higher secondary stream
* 'degree\_p' : Undergraduate percentage
* 'degree\_t' : Undergraduate degree type
* 'workex' : Work experience
* 'etest\_p' : Placement test percentage
* 'specialisation' : MBA specialisation
* 'mba\_p' : MBA percentage
* 'status' : Hiring status
* ‘Salary’ : salary Offered

I have learnt how to visualize the variables with **HighChart method using Highchart packages**.And I have done Good EDA Part and learnt the techniques to visualize every category column with **HighChart with type BarPlot and Pie Chart.** And Drawn Good Conclusions with all categorical variables.

There is only one column with missing values i.e., Salary Column where distribution is not following Gaussian and the data is Right Skewed . Even after replacing with median the data is still right skewed, maybe there will be change in distribution if do the outliers treatment. But doing Outlier treatment in Salary column would may affect the change in bias and variance that might affect in the model Prediction

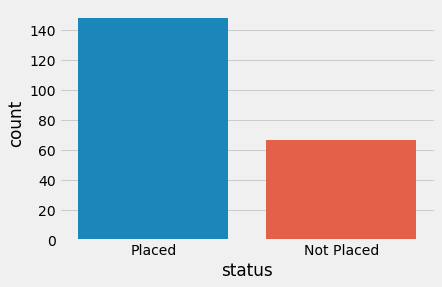
## The Gender column with Male:Female ratio is approximately 2 (1.86 nearly) . So for every 1 female candiate, there are 2 male candiates sitting for placements

## 

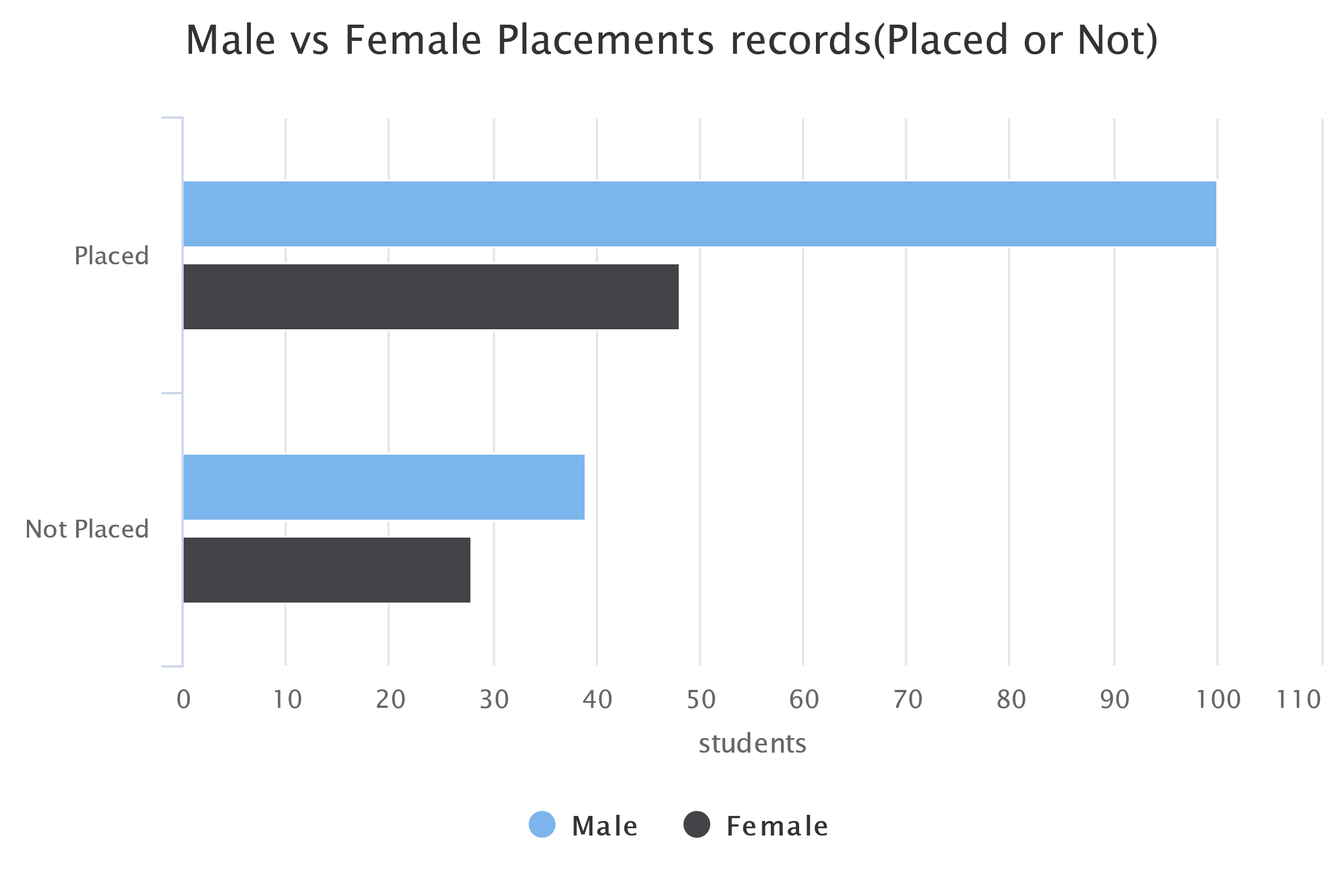
**Conclusion:**

**Conclusion:**

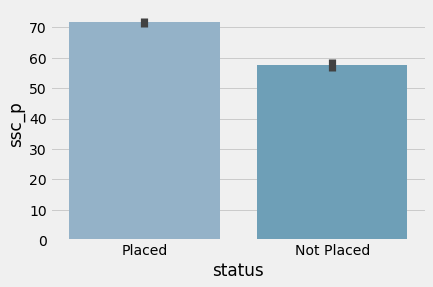
1. Target Variable Status is UnBalanced (Smote use)as it is the data collected from survey conducted. I did not used to use smote to balance the data

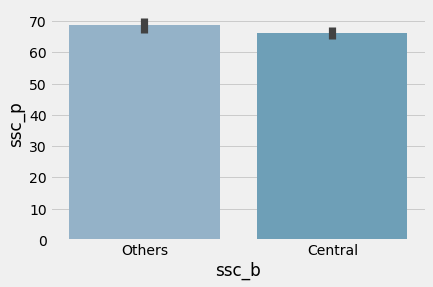


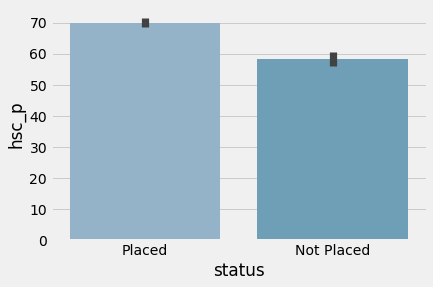
1. More male candidates got placed as compared to female candidates.



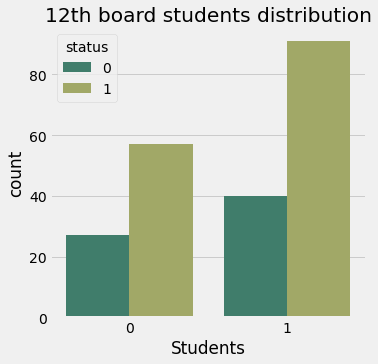
1. Male Candidates got higher CTCs as compared to female candidates.
2. Type of Board choosen does not have any effect on placements thus we can drop in preprocessing steps.

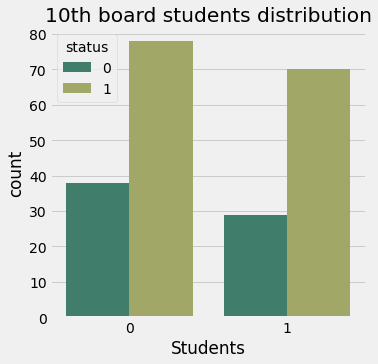




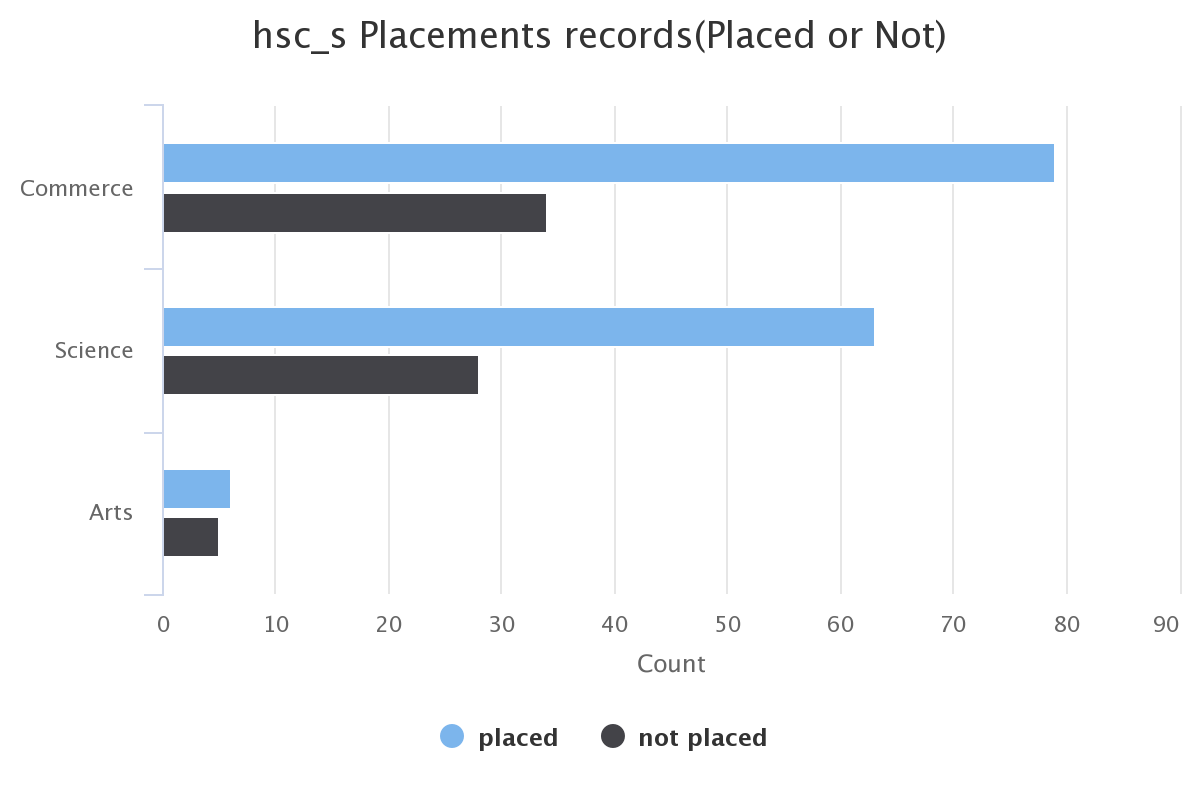


1. Most of the students preferred Central board in 10th grade whereas other boards in 12th grade.

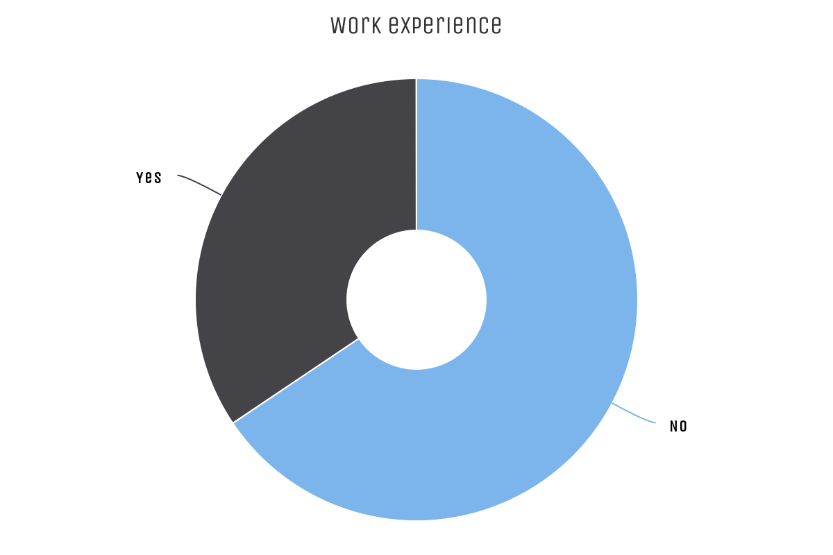


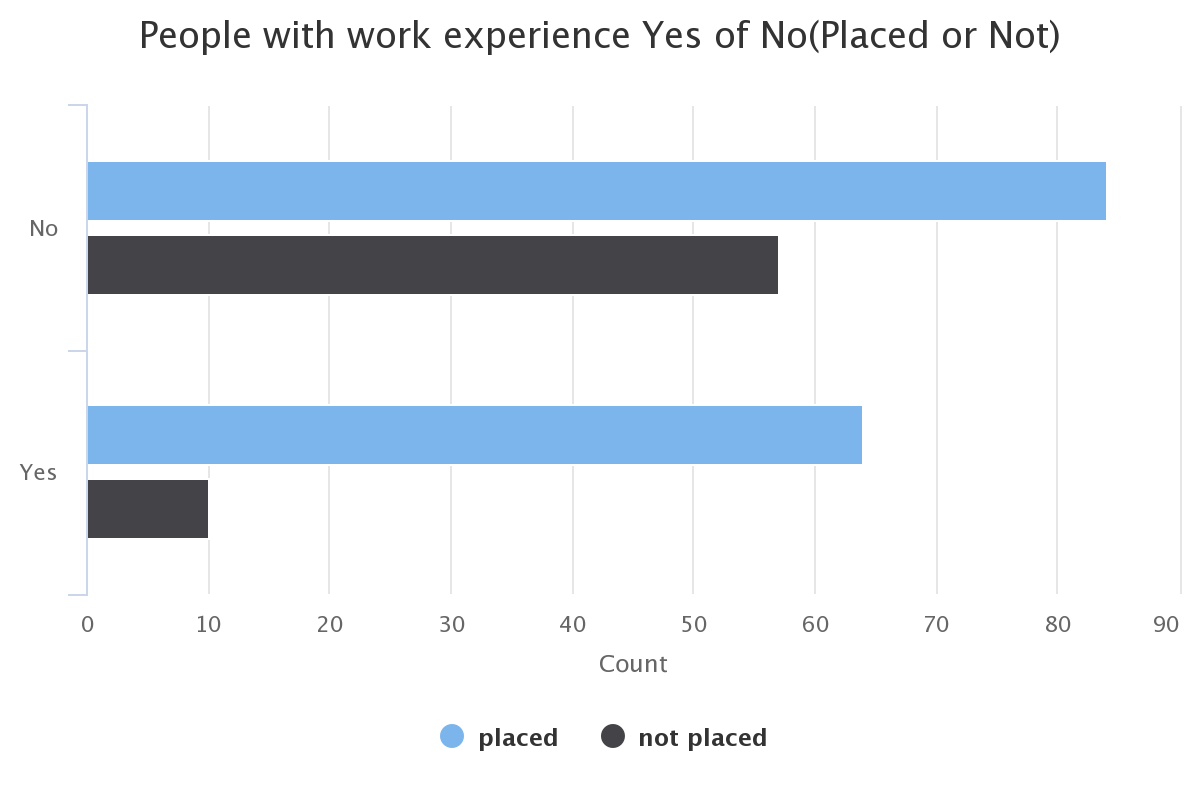


1. The most popular branch turns out to be commerce or maybe as most of students get average marks so they were admitted to got commerce on based of their marks. Science is the second most popular and the least popular is arts

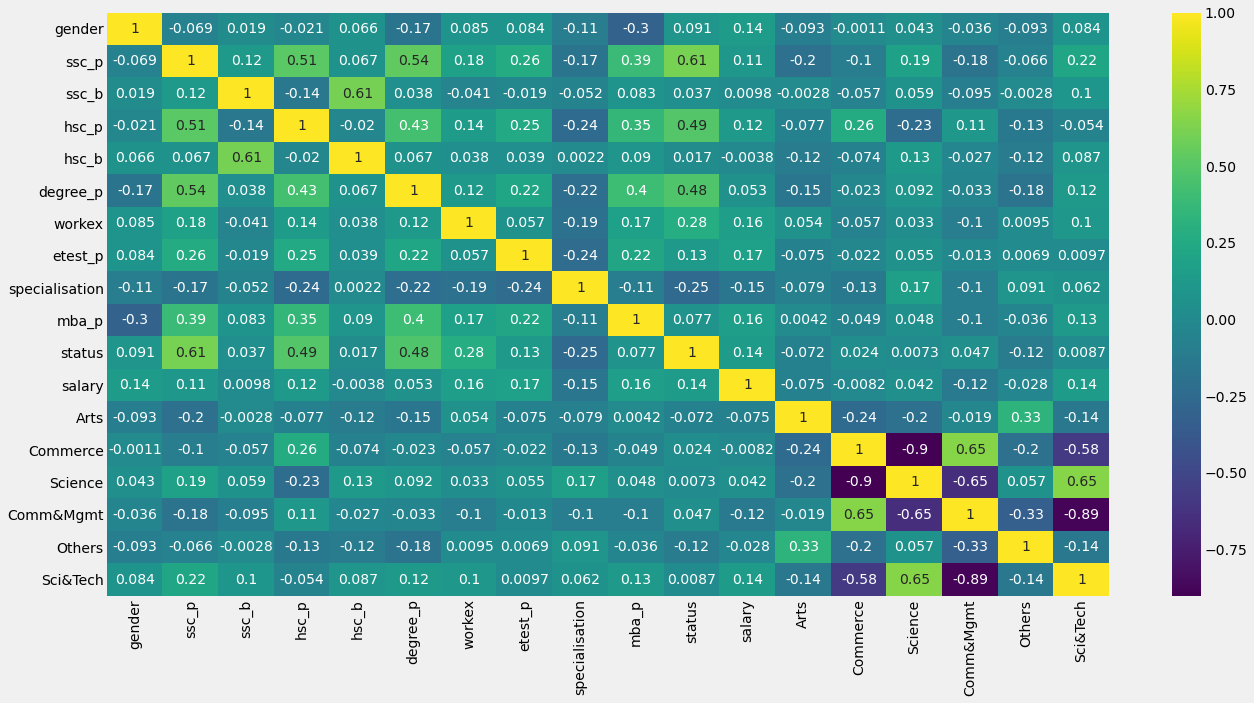


1. Candidates with higher percentages have better chance of placements.
2. Work experience with placement status Yes: 74, No : 141

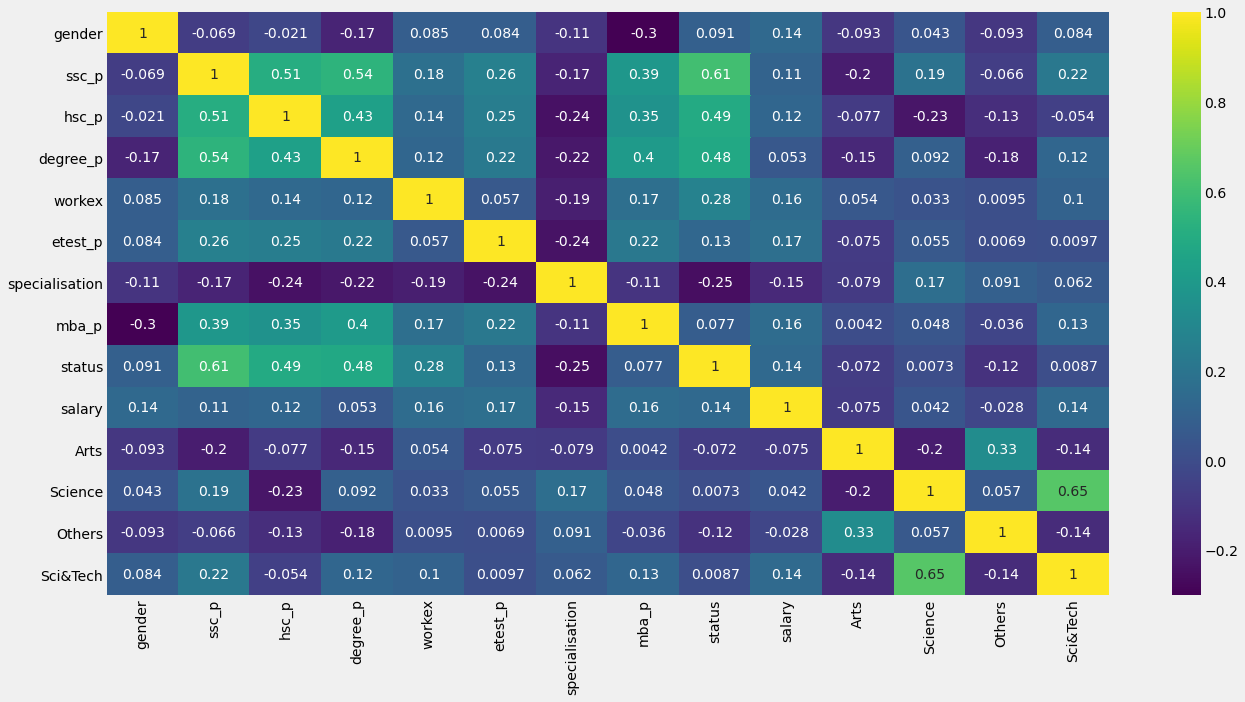




1. Choosing Science and Commerce as Specialisation seems to have perk when it comes to placments.
2. 86% of proportion who had experience has got placed, where as there is only 59.57% of proportion got placed with no experience.
3. Commerce is the most popular branch among other Specialisations. And candidates with Commerce and Science backdrop got placed
4. Choosing Sci&Tech and Comm&Mngmt as degree will fetch you higher CTCs.I took the threshold value for correlation as **"0.65"** As there is some bad correlations that need to be omit out. "Comm&Mgmt" vs "Sci&Tech", One need to be removed out of these two as there is bad correlation. Similarily Between Commerce and science one need to be removed as there is high correlation. Examination boards can be dropped as well, since they don't seem to give us any extra information.
5. Constructed a correlation matrix after scaling the data



Dropped few features having correlation more than 0.65(threshold)



# **Logistic Regression :**

# I got a test\_score of **89.6000001%** and an **0.952 AUC score** where it is proved that my model is **95.2%** correctly predicted which is great.

# **KNN Model:**

# Here I was able train data with some more performance out of the model by tuning to a **better K value at K = 8** where I achieved **95.8 Percent accuracy**

**Submitted by :**

**Sai Nikhil K**

**Guidance By:**

**Prof . Dr. Vinod Kumar Murthi**