isua

campus placement data analysis

Source: Kaggle

Sia Bhowmick

Part of Data Analysis Portfolio

Contents

[1 Introduction 2](#_Toc44373900)

[1.1 Overview 2](#_Toc44373901)

[1.2 Preliminary Inspection of Variables 2](#_Toc44373902)

[2 Analysis 3](#_Toc44373903)

[2.1 Exploratory Data Analysis 3](#_Toc44373904)

[3 Conclusion 20](#_Toc44373905)

# 1 Introduction

## 1.1 Overview

The dataset studied here consists of campus placement data of students studying MBA in Business Analytics at Jain University, Bangalore, India. It contains details on the placement status of the students, specializations, and percentages for secondary and higher secondary, undergraduate and MBA degrees, work experience, and salary offered to the students who were placed.

The goal of this analysis is to understand the placement process in the field of Business Analytics in Bangalore, India by answering the following questions.

1. For the candidates who were placed, does the salary depend on the different exam percentages / test scores?
2. Does the salary depend on the degree and specialization?
3. Which factor / factors influenced a candidate getting placed?
4. Which degree / specialization is in demand?
5. Play with data conducting all statistical tests.

## 1.2 Preliminary Inspection of Variables

The dataset used for this analysis is the [placement data from Kaggle](https://www.kaggle.com/benroshan/factors-affecting-campus-placement) with 215 observations and the following 15 variables.

* sl\_no: Serial Number of the observations for each student
* gender: Male = ‘M’; Female = ‘F’
* ssc\_p: Secondary Education percentage (10th grade)
* ssc\_b: Board of Education for Secondary (Central / Others)
* hsc\_p: Higher Secondary Education percentage (12th grade)
* hsc\_b: Board of Education for Higher Secondary (Central / Others)
* hsc\_s: Specialization in Higher Secondary Education (Commerce / Science / Other)
* degree\_p: Degree Percentage
* degree\_t: Under Graduation degree type (Field of education)
* workex: Record of previous work experience (Yes / No)
* etest\_p: Employability test percentage (conducted by college)
* specialisation: Post Graduation (MBA) specialization (Mkt&Fn / Mkt &HR)
* mba\_p: MBA percentage
* status: Status of placement (Placed / Not Placed)
* salary: Annual salary offered by corporate to candidates who were placed

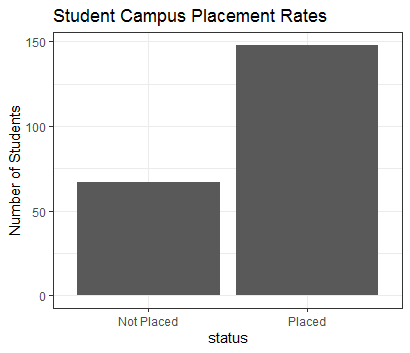
Based on the information above, the variable sl\_no is not going to be to useful in the analysis, and therefore, can be eliminated during model fit. Out of the remaining variables, six are numeric and eight are categorical. The only missing data in this dataset belong to the salary column for those students who were not placed.

# 2 Analysis

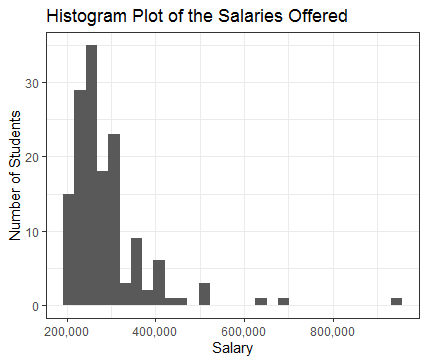
The dataset will be first inspected visually and then statistically using in R in order to answer the questions asked in section 1.1.

## 2.1 Exploratory Data Analysis

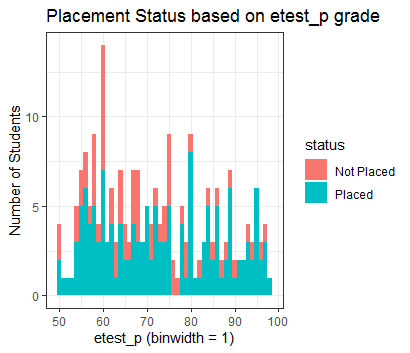
1. 1st question – **What is the placement rate?**
2. Placement is recorded by the categorical variable status.
3. Use ggplot() and geom\_bar() with x = status.



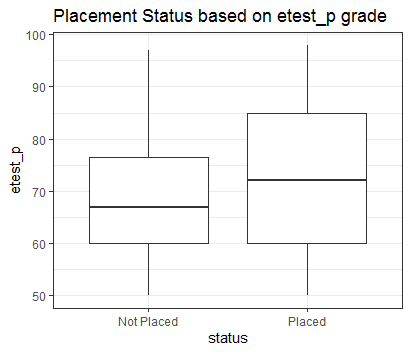
1. **Remark 1:** It is noticed that more students were placed than not placed
2. Obtain the numeric percentages using the table() function.
3. Not Placed: ~31%, Placed: ~69%.
4. 2nd question – **What is the numeric range of the salaries offered?**
5. Use the histogram option with ggplot() and scale the x-axis so that the grid marks do not display in scientific notation.



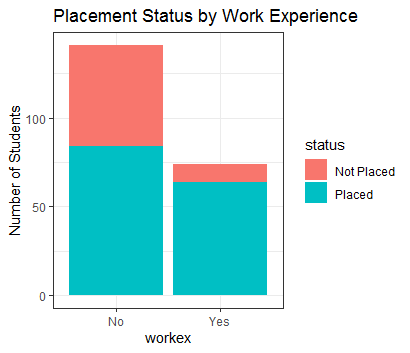
1. **Remark 2:** In the report we had mentioned the presence of salaries that were not available for students who were not placed. The histogram plot of salaries generated a warning sign for this reason, pointing out 67 observations without the salary information. As a verification, if we divide 67 by 215, we get ~31%, same as generated from earlier results. This is a way to confirm that we are not missing any salary information on students who were placed. The salary offered ranges from roughly Rs.200,000 to Rs.950,000. And the median salary is around Rs.265,000.
2. 3rd question – **What is the status by etest\_p grade?**
3. Hypothesis is that the higher the etest\_p grade the higher the chance of being placed.



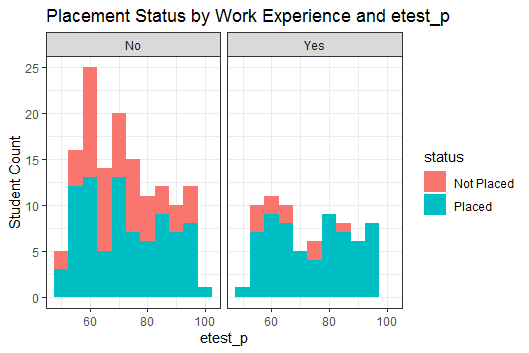
1. **Remark 3a:** Histogram plot for etest\_p with fill=status shows that in general for higher etest\_p scores, a greater percentage of students were placed than not placed. However, there are anomalies where the hypothesis does not hold. For example, students with etest\_p from 75-77 were not placed, and those with score between 53-57 has a higher placement percentage.



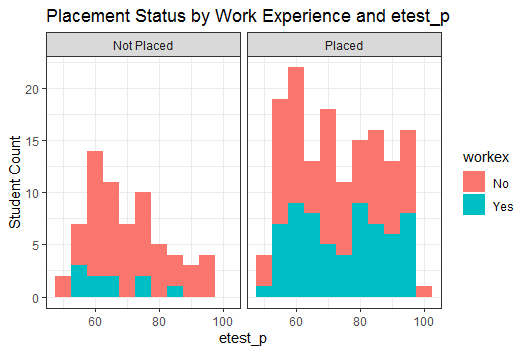
1. **Remark 3b:** Boxplot shows the etest\_p score for students who were not placed in the range 60-76, and for those placed in the range 60-85. Therefore, it is certain that students with high etest\_p scores were placed, but those with scores 60-76 who were also placed, there must be other factors that influenced it.
2. 4th question - **What is the status by work experience?**
3. Hypothesis is that candidates with previous work experiences are likely to be placed.



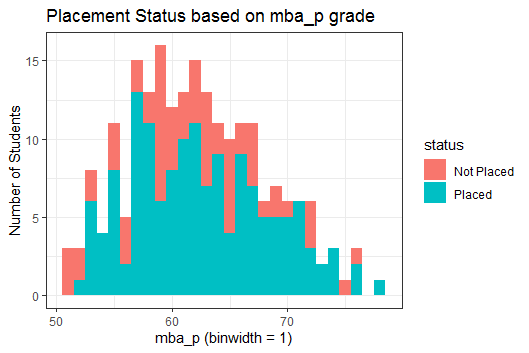
1. **Remark 4:** The bar plot shows that almost 60% of students without work experience were placed, whereas almost 90% with work experience were. This result matches with our hypothesis. We also note that there are more students without work experience than with.
2. Now, perhaps we can test the effect of work experience and etest\_p on placement status.
3. 5th question - **How does the status depend on both etest\_p and work experience?**



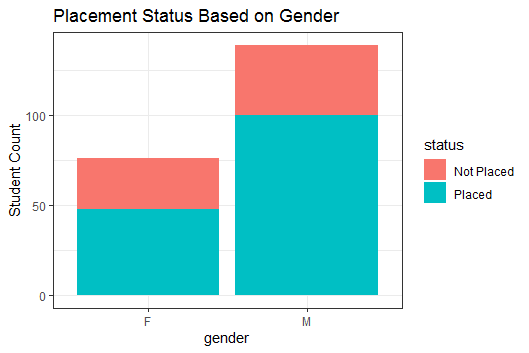
1. **Remark 5a:** This first plot's layout is confusing and does not help us to visually determine the relationship asked in the question.



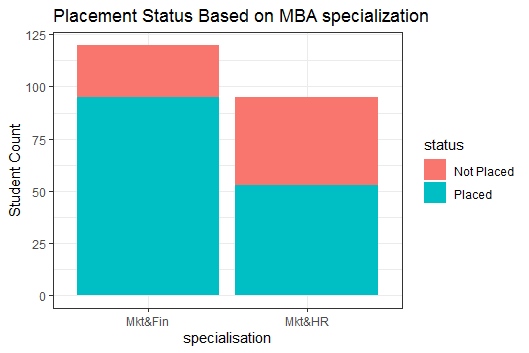
1. **Remark 5b:** From the second plot it is much clear that a very small percentage of students with previous work experience were not placed, and more among these students had low etest\_p score than high. However, for the students who were placed, etest\_p scores do not seem to have any effect on the placement decision.
2. 6th question – **What is the effect of MBA percentage on placement status?**
3. Hypothesis is that mba\_p is an influence in some way.



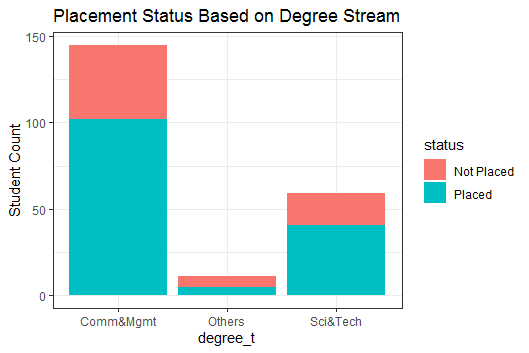
1. **Remark 6:** The plot shows the hypothesis to hold for the lowest and the highest mba\_p scores, but for the scores in between the visual evidence is not strong. Therefore, we cannot correctly conclude that the hypothesis holds.
2. 7th question – **Did gender play a role?**
3. Hypothesis is no.



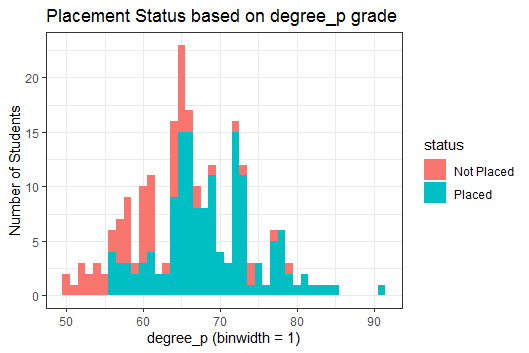
1. **Remark 7:** The bar plot shows that there are less female than male candidates. And we also see that similar percentages of male and female candidates were placed and not placed. Hypothesis holds true.
2. 8th question – **Does MBA specialization have any effect on status?**



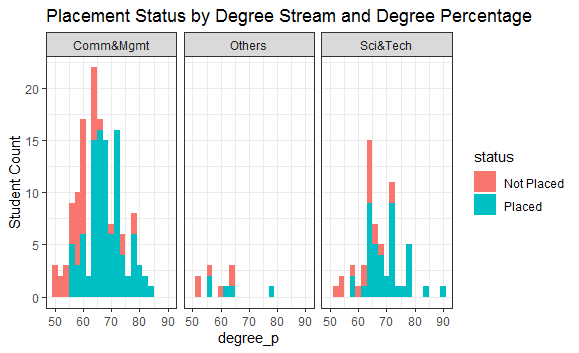
1. **Remark 8:** We see that more students specialized in Mkt&Fin than Mkt&HR. In addition, we see that a more candidates with Mkt&Fin were hired.
2. 9th question – **Does the degree stream have any effect on status?**



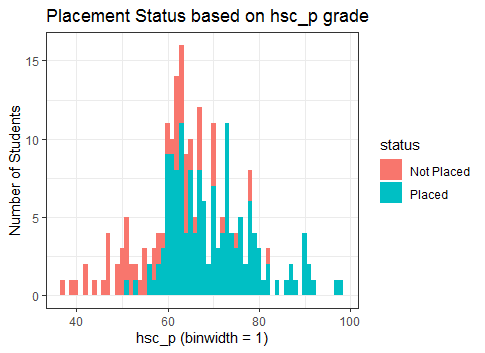
1. **Remark 9:** We see more students from Comm&Mgmt being placed, than Sci&Tech, and very few from the Others degree stream.
2. 10th question – **Does the degree percentage have any effect on status?**



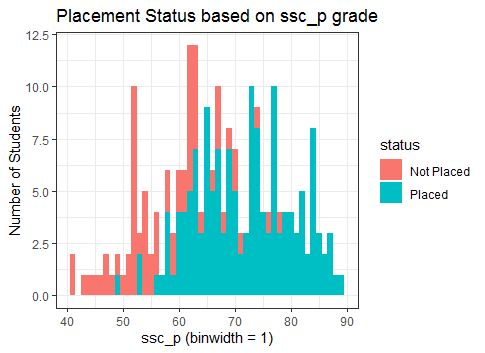
1. **Remark 10:** The histogram shows that overall candidates with higher degree percentage received placement.
2. 11th question – **How does both degree stream and degree percentage reflect on status?**



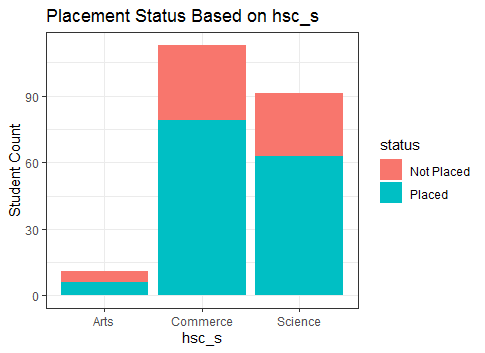
1. **Remark 11:** We see the overall trends observed for questions 9 and 10. The plot shows that candidates with higher degree percentages were preferred, regardless of their degree streams. Degree in Comm&Mgmt has the highest number of students among all degree streams.
2. Note that the preference in Comm&Mgmt stream and MBA specialization in Mkt&Fin could be due to the type of companies that took part in the placement process than the usefulness of the degrees themselves, and that larger number of MBA students preferred this degree and specialization.
3. 12th question – **How does the hsc\_p affect status?**



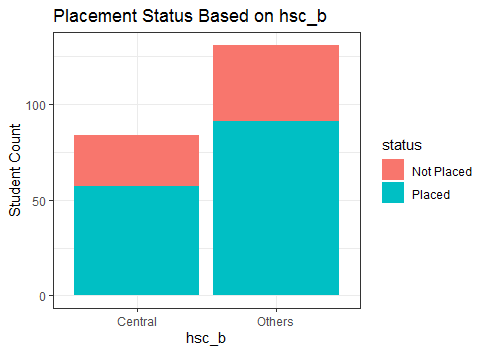
1. **Remark 12:** Candidates with higher hsc percentages were preferred for placement.
2. 13th question – **How does the ssc\_p affect status?**

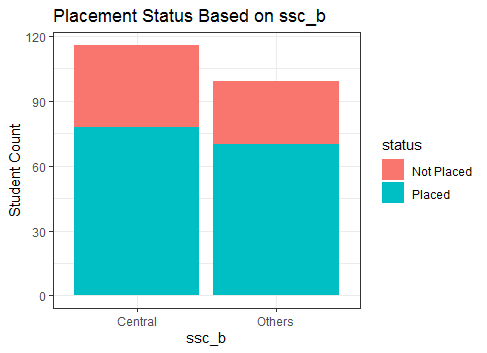


1. **Remark 13:** Candidates with higher ssc percentages were preferred for placement.
2. 14th question – We saw that hsc\_p has a positive correlation with status. **Now let us explore how hsc\_s might affect the status.**

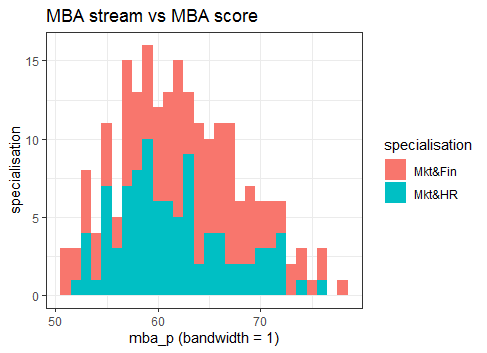


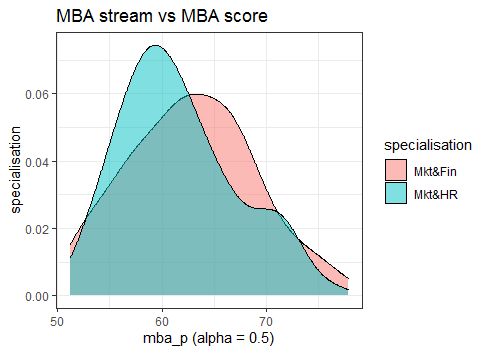
1. **Remark 14:** We see that that most candidates are in the Commerce stream followed by the Science stream, and very few students are in the Arts stream. But we also see that a higher percentage of Commerce and Science students were placed than Arts, for which stream the percentage placed is approximately 55%.
2. 15th question – **Do the hsc and ssc boards have any effect?**



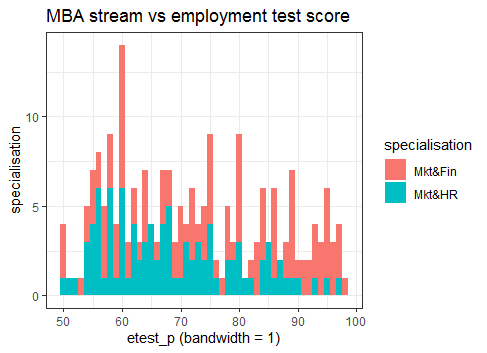


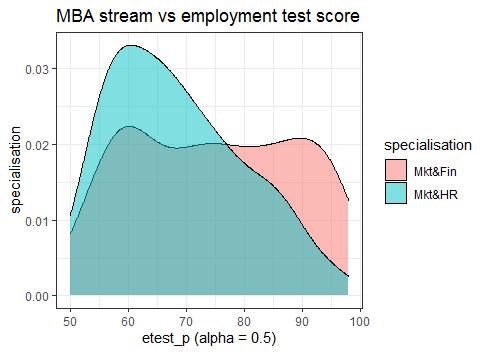
1. **Remark 15:** There does not seem to be any evidence to suggest one board is preferred than the other.
2. We noticed a higher preference for candidates who specialized in Mkt&Fin, and we also noticed that students with higher test scores in degree\_p, hsc\_p, and ssc\_p were preferred.
3. 16th question – **How does the MBA score vary over different specializations?**



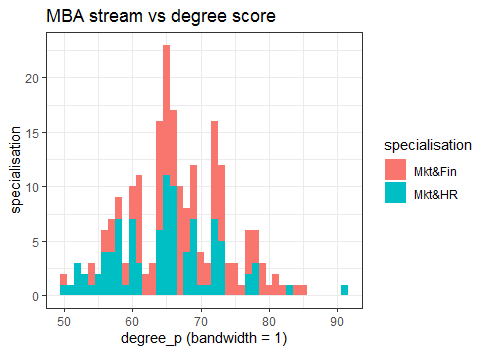


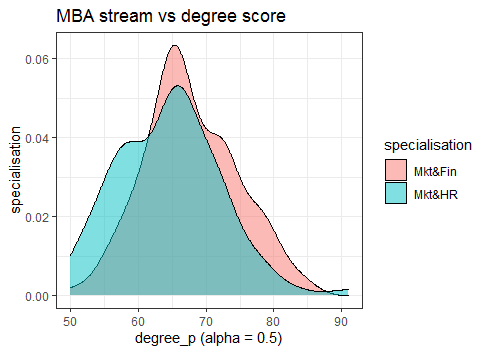
1. **Remark 16:** The MBA score trend for both specializations seem to be following the same trend. But the density plot shows more candidates in Mkt&HR had lower scores.
2. 17th question – **How does the etest score vary for MBA students based on their specialization?**



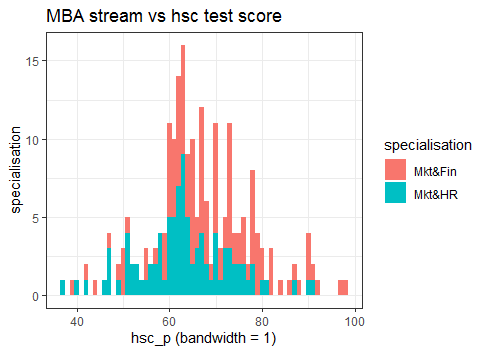


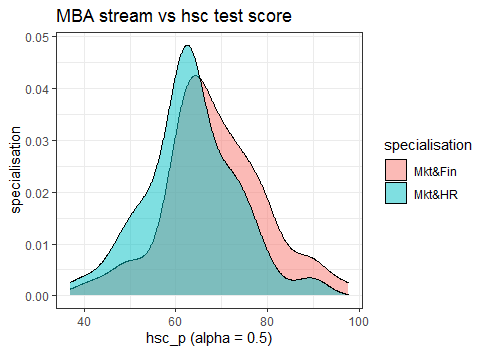
1. **Remark 17:** We see that more students in Mkt&HR had lower employment test scores.
2. 18th question - **Did students who specialized in Mkt&HR score poorly in these tests?**



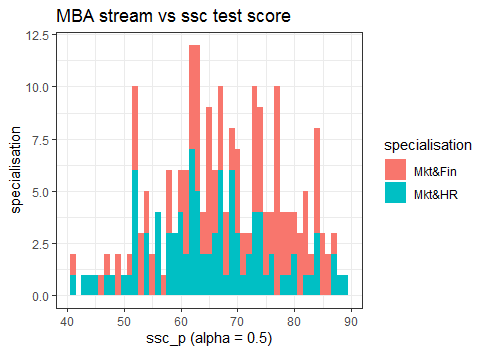


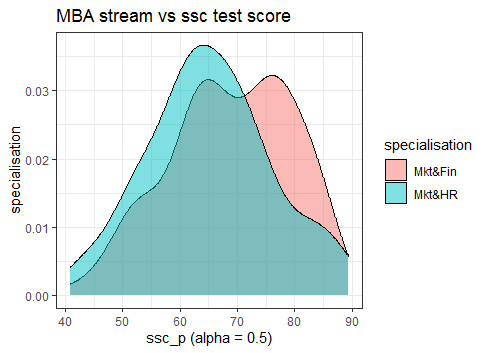
1. **Remark 18:** The histogram plot shows that students with Mkt&HR had lower degree scores, and more students specializing in Mkt&Fin scored higher.
2. 19th question – **How are the hsc\_p test scores for candidates’ MBA specialization?**



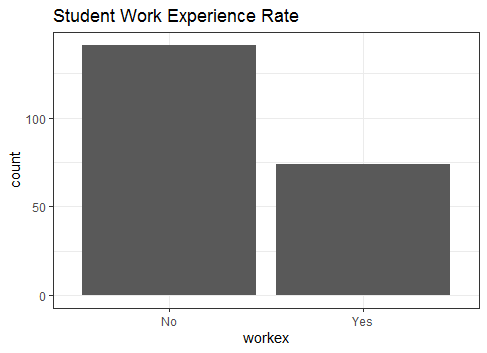


1. **Remark 19:** The difference in test scores for hsc\_p was not noticeable for students in different MBA specialization.
2. 20th question – **How are the ssc\_p test scores for candidates’ MBA specialization?**

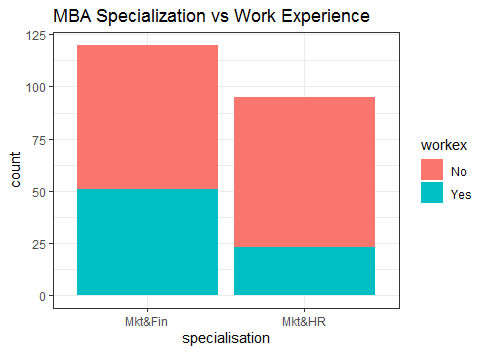




1. **Remark 20:** Although a higher number of Mkt&Fn students scored between 75-90, we can conclude no noticeable difference based on students’ specializations as more students are enrolled in Mkt&Fin than Mkt&HR.
2. 21st question – **Do more students have work experience?**



1. **Remark 21:** We see that more students with work experience than not at the time of the campus placement event.
2. 22nd question – **Is there a relationship between MBA specialization and work experience?**



1. Remark 22: We see that more students in Mkt&Fin with work experience than those in the Mkt&HR specialization.

# 3 Conclusion

1. More students were placed than not placed (69% vs. 31%)
2. The salary offered ranges from approximately Rs. 200,000 to Rs. 950,000, and the median salary is around Rs. 265,000.
3. Inspection of the numeric variables
   1. The effect of all test scores on the placement status were individually inspected.
   2. The employment test scores did not seem to have any definite effect on the placement decision. The box-and-whiskers plot showed that students with really high test scores were definitely placed, but there were no distinction for those whose scores were in the low to mid range.
   3. The MBA test score did not seem to be a strong factor in placement decision.
   4. Candidates with higher percentages in degree exam, higher secondary exam, and secondary exam were preferred for placement.
4. Inspection of the categorical variables
   1. Inspecting the effect of previous work experience on placement decision showed a positive correlation. Almost 90% of students with work experience were placed, whereas for those without work experience the placement rate is around 60%.
   2. Student’s gender did not play a role in the selection process as the bar plot showed that around 70% of students from each gender were placed.
   3. It was noted that students who specialized in Marketing and Finance were preferred to those who specialized in Marketing and HR.
   4. There were no observable preferences towards a stream or school board.
5. Overlap of categorical and numeric variables
   1. Inspecting the combined effect of employment test score and previous work experience on the placement decision showed that a very small percentage of students who were not placed had previous work experience, and the employment test scores of students not placed were skewed left.
   2. Inspecting the combined effect of degree stream and degree percentage showed candidates with higher degree percentages were preferred, regardless of their degree streams. But it was noted that most students were in Commerce and Management, followed by Science and Technology, and very few in other streams.

To summarize, in general, students with previous work experiences and higher test scores in secondary, higher secondary, and degree final exams were preferred. It was noted that students who specialized in Marketing and Finance were preferred to those who specialized in Marketing and HR. It turned out that the latter group of students had consistently lower scores for MBA test score, employment test score, and degree test score, and a lesser number of these students had work experience. This shows that test scores and previous work experiences were two important deciding factors for the placement decision.