

1. Introduction

The objective of this assignment was to create visualizations about university placement data of students in a particular college in India. Job hunting is an extremely stressful prospect, especially for fresh graduates in India. Given adverse macroeconomic conditions, coupled with stiff competition from their peers and people already employed in the workforce, being able to visualize specific queries to understand and improve an individual's employability could prove helpful.

2. Datasets

2.1 Main Dataset

The dataset sourced for this project was the Campus Recruitment Dataset, containing student data from 216 students that obtained job offers prior to graduation.

The dataset contains 15 columns, as explained in the section below.

ID	Column data type/range/	Data Type	Data Properties	Description
1	sl_no	Int	Categorical	Serial Number of candidate
2	Gender	String	Categorical	Gender of candidate (Male='M', Female='F')
3	Ssc_b	String	Categorical	Board of education cluster that candidate was part of for sec. school education
4	hsc_b	Float	Quantitative	Secondary education percentage score obtained by candidate
5	hsc_p	Float	Quantitative	Higher secondary education percentage score obtained by candidate
6	hsc_b	String	Categorical	Board of education cluster that candidate was part of for high school education
7	hsc_s	String	Categorical	Specialization in Higher Secondary Education
8	degree_p	Float	Quantitative	Undergrad degree Percentage score obtained by candidate
9	degree_t	String	Categorical	Undergraduate degree type chosen by candidate
10	workex	String	Categorical	Presence of prior work experience for candidate
11	etest_p	Float	Quantitative	Employability test percentage score obtained by candidate
12	specialisation	String	Categorical	Post-Graduation (MBA) specialization selected by candidate
13	mba_p	Float	Quantitative	MBA percentage score obtained by candidate
14	status	Boolean	Categorical	Status of placement- Placed/Not placed
15	salary	Int	Quantitative	Salary offered by corporate to candidates

2.2 Pre-processing

There were no NA values within the dataset and all entries were appropriately populated.

3. Matplotlib & Seaborn

This task was completed using Python 3.9 with the following important libraries.

- Seaborn 0.11.2
- Matplotlib 3.5.1
- Pandas 1.5.0

Plots were developed with a combination of Seaborn and Matplotlib whilst Pandas was used for reading of csv files containing the data. A jupyter notebook was created containing the 4 queries arranged sequentially. All the associated code with each of the plots was documented along with the notebook cells.

3. Queries

3.1 Query 1: How many candidates have work prior work experience, and does prior work experience improve the changes of obtaining a work placement?

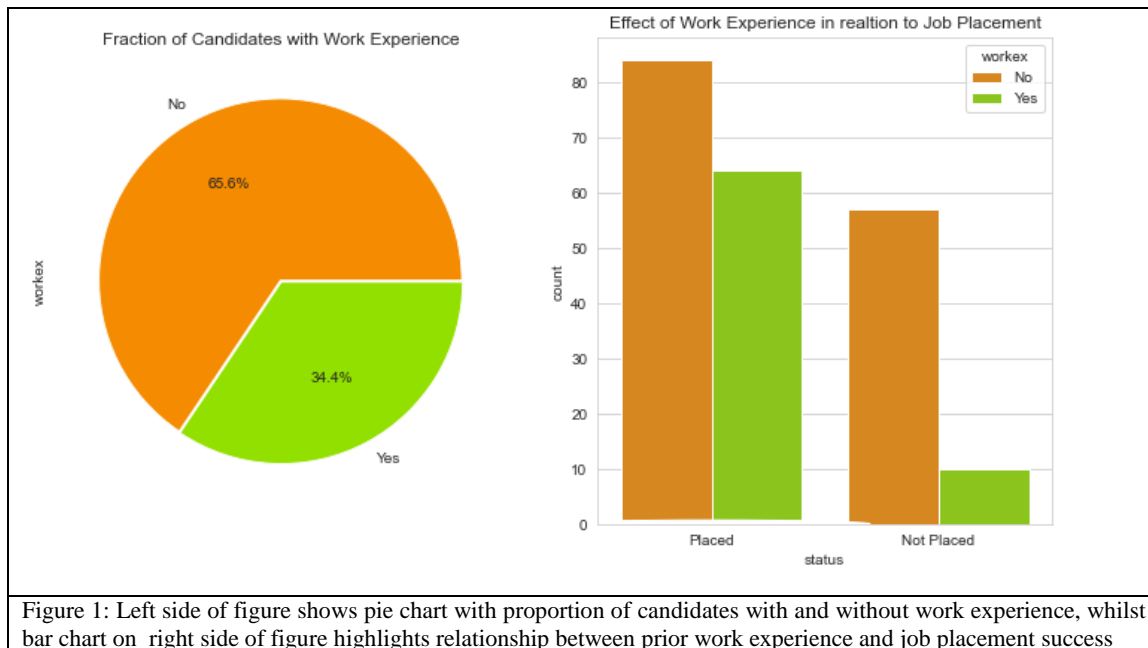


Figure 1: Left side of figure shows pie chart with proportion of candidates with and without work experience, whilst bar chart on right side of figure highlights relationship between prior work experience and job placement success

Table 1: Data used in the pie chart and bar chart plots in Figure 1

Data	Data Type	Encoding	Note
Prior Work Experience	Categorical	Colour	Candidates who had prior work experience were coloured green while candidates with no work experience were coloured orange
Fraction of candidates (left)	Quantitative	Size of slice	For the left chart, the size of the respective pie slices indicates the percentage of candidates with and without prior work experience
Candidate Count (right)	Quantitative	Height of bars	For the right chart, the height of each bar within the bar chart indicates the number of candidates within a specific combination of successful job placement and presence of prior work experience.

3.2 Query 2: Does the test scores of candidates affect their job placement opportunities?

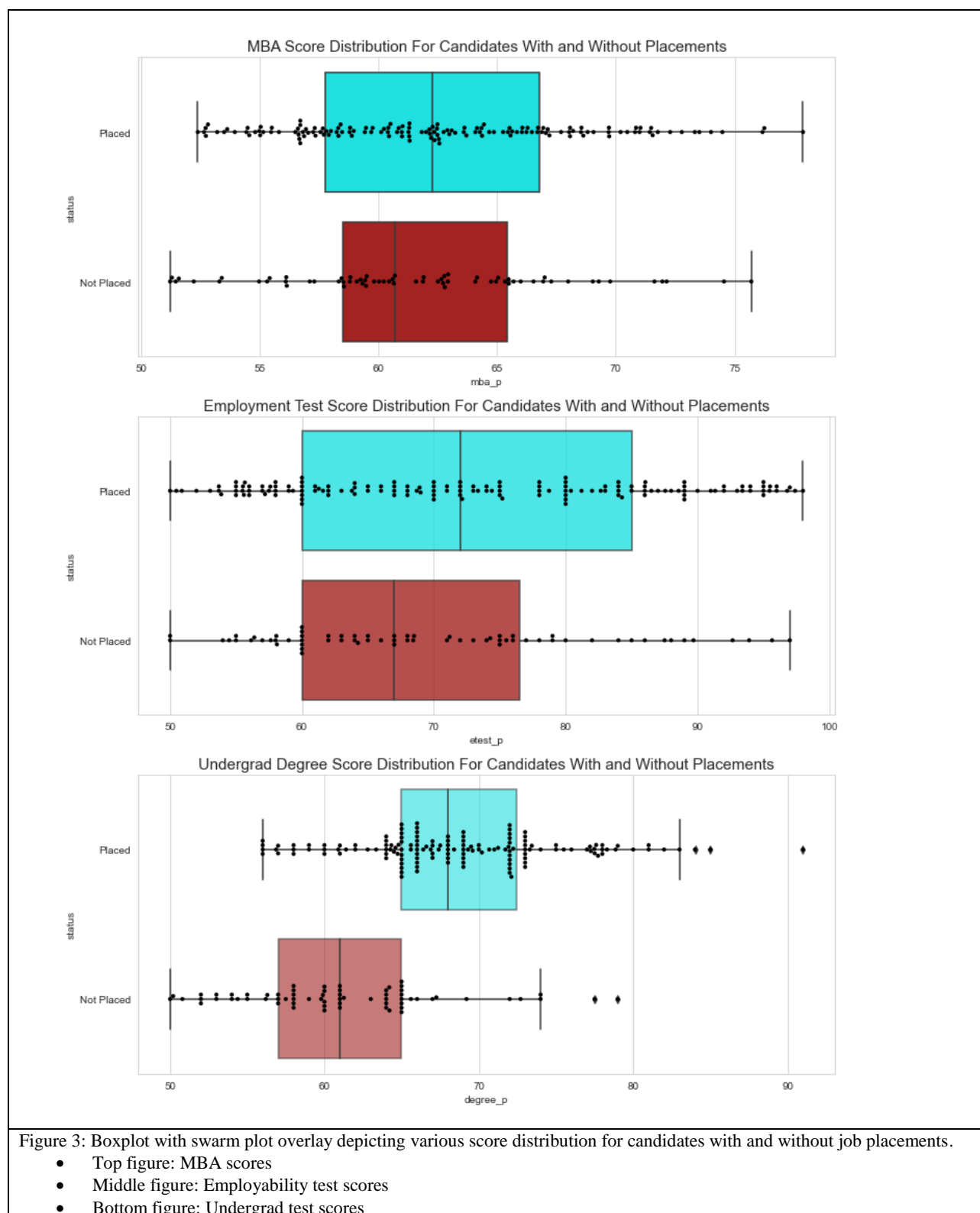


Table 2: Data used in the boxplot with swarmplot overlays in Figure 2

Data	Data Type	Encoding	Note
Successful Placement	Categorical	Colour	Candidates that were offered placements were highlighted in blue, whilst candidates that were not offered job placements were highlighted in orange
Number of Candidates	Quantitative	Position (y-axis)	The wider the number spread of candidates on the y-axis, the greater the number of candidates with the same score for a given test/grade and placement status
MBA Score (Top) Employability Score (Mid) Undergrad Score (Bottom)	Quantitative	Position (x-axis)	The higher the test/grade scores, the further towards the right the individual would be located on the x-axis.

3.4 Query 3: Given the high job placement of students with work experience and high undergrad scores, is there any observable relationship between the 3 variables?

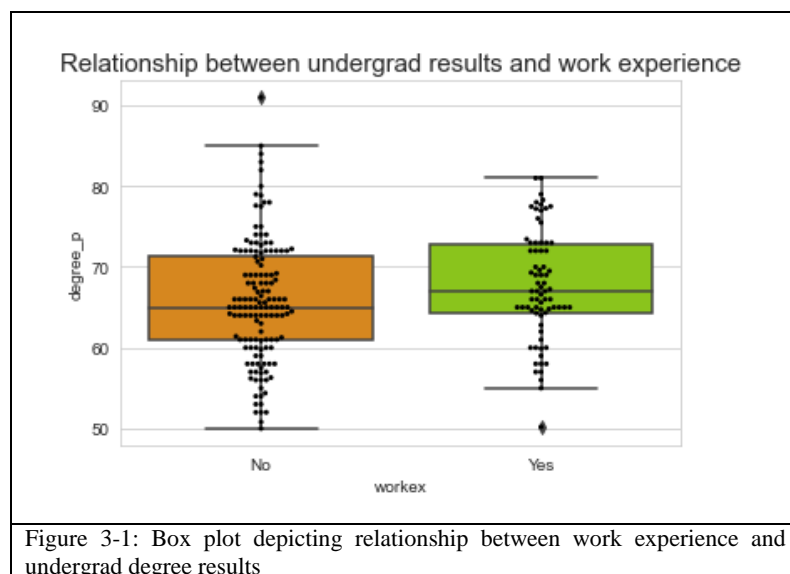


Table 3-1: Data used in the box plots in Figure 3-1

Data	Data Type	Encoding	Note
Prior Work Experience	Categorical	Colour	Candidates who had prior work experience were coloured green while candidates with no work experience were coloured orange
Candidate Undergrad Scores	Quantitative	Y-axis position	Candidates with higher undergrad scores were represented with points in the y-axis
Candidate Count	Quantitative	X-axis width	Candidates with the same score for each category were represented on a similar horizontal plane – so candidates



Table 3-2: Data used in the violin chart and bar chart plots in Figure 3-2

Data	Data Type	Encoding	Note
Prior Work Experience	Categorical	Colour	Candidates who had prior work experience were coloured green while candidates with no work experience were coloured orange
Quantity of candidates	Quantitative	Width	For each category of prior work and placement status, the higher the number of individuals with the undergrad degree score, the wider the plot in the x-axis
Undergrad Degree Score	Quantitative	Height	The higher the position on the y-axis, the greater the undergrad degree score obtained by a particular candidate.
Job Placement Obtained	Categorical	Position	Candidates who obtained job placements were represented in the left side of the chart whilst candidates who did not obtain job placements were represented in the right side of the chart.

3.4 Query 4: Of the candidates that were offered jobs, did their specialization or gender affect their salaries?

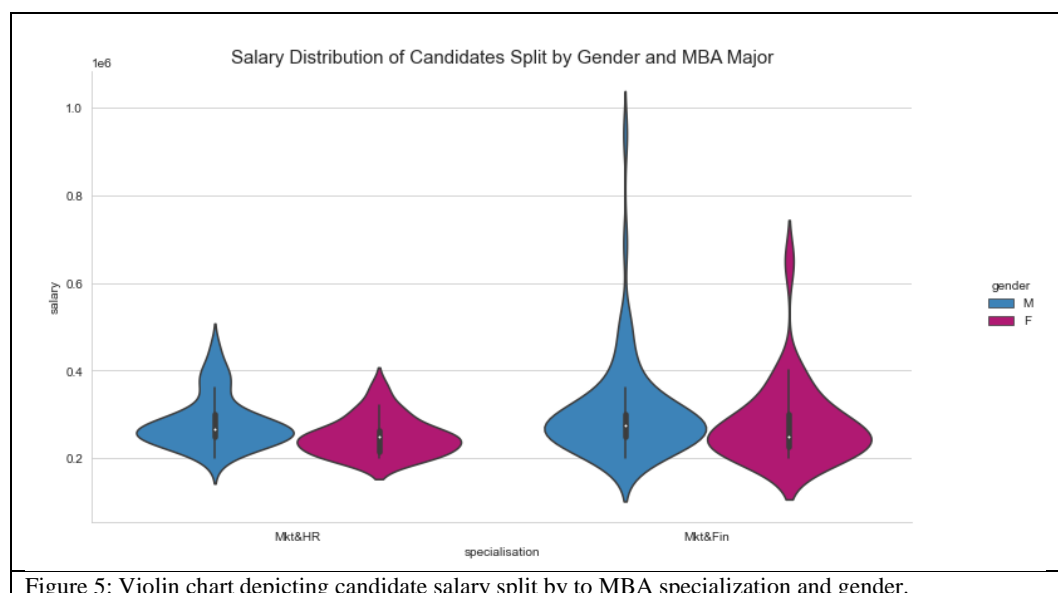


Table 5: Data used in the Violin chart in Figure 5

Data	Data Type	Encoding	Note
Gender	Categorical	Colour	Male candidates represented within the blue plot whilst female candidates represented within the burgundy plot
Quantity of candidates	Quantitative	Width	For each gender and specialization, the higher the number of individuals with the same salary, the wider the plot in the x-axis
Salary	Quantitative	Height	The higher the position on the y-axis, the greater the salary offered to a particular candidate.
Specialization	Categorical	Position	Candidates with MBA majors in Marketing & HR were represented in the left side of the chart whilst candidates with MBA majors in Marketing & Finance were represented in the right side of the chart.

4. Analysis

4.1 Query 1: How prior work experience affects obtaining work placements

The first pair of visualizations presented here is a pie chart (left) of the proportion of all MBA students with and without prior work experience and a bar chart (right) that highlights relationship between prior work experience and job placement success. From the plot on the left we note that approximately of students have prior work experience whilst the remaining students do not have any work experience. From the figure on the right, we note that a disproportionate number of students with prior job experience have placement offers as compared to students with no work experience. This indicates that prior job experience is important in securing job placements after completing an MBA.

4.2 Query 2: Relationship between test scores and placement opportunities

This query is answered by 3 visualizations that cover the relationship between MBA scores, Employability Test scores and Undergrad test scores versus whether candidates obtained job placements. From the boxplots in each of the 3 figures, we note that that the median score for candidates with job placements is consistently higher than for candidates without job placements. That being said, the median, 25th and 75th percentile score difference between candidates with and without job placements was most pronounced for the Undergrad degree scores. For the other 2 scenarios, there was significant overlap between both sets of complementary boxplots. This indicates that students who perform well in their undergrad have higher chances of getting work placements after their MBAs.

4.3 Query 3: Given the results of Queries 1 and 2, is there any relation between high job placements versus work experience and high undergrad scores

The first visualization in this section is a simple boxplot with a swarmplot overlay of the relationship between Undergrad results and prior work experience. This visualization shows that students with prior work experience have a slightly higher median score. But looking at the individual point distribution, we do not see any conclusive difference based on the points located within the box itself.

The second visualization is a violin plot that combines the undergrad score, work experience and job placement status of all candidates. Here we note that when comparing between students with placements and without placements, the main contributor to obtaining job placements is higher undergrad results and not prior work experience.

5. Citations

1. CMS Business School: DG160 (2019) *Campus Recruitment*, *Github*. Available at: https://github.com/DG1606/CMS-R-2020/blob/master/Placement_Data_Full_Class.csv (Accessed: March 11, 2023).
2. Seaborn Catplots (no date) *seaborn.catplot* - *seaborn 0.12.2 documentation*. Available at <https://seaborn.pydata.org/generated/seaborn.catplot.html> (Accessed: March 15, 2023).