Exploring Factors Affecting Student Placements: A Comprehensive Analysis of Demographics, Academic Performance, and Preparation Activities

[1]Javid Ahmad Khan ,[2]Pawan Kumar, [3]Sanjay Sood, [4]Raghav Gupta, [5]Sunny Gupta

[1]MSc (Information Technology) pursuing, Lovely Professional University, Phagwara, India

[2]School of Computer Applications, Lovely Professional University, Phagwara, India

[3,4]Division of Student Career Services, Lovely Professional University, Phagwara, India

[5]Senior Software Engineer, Mainbrainer Solutions Private Ltd, Mohali, India

Abstract.

When it comes to educational institutes, the performance of their students is largely represented in terms of their placements. Placements are very important for an educational institute for its brand building and for students it is quite necessary to launch their careers. These reasons have been a major reason to take up this analysis. This work is an attempt to identify the various factors that affect the placement status of students. The factors explored included demographic details, the prior academic performance, participation in placement preparation activities and performance in benchmarking tests. The methodology used was to frame research questions corresponding to factors of interest, collect data on students on these factors and analyze the relative performance of students in terms of placement percentage to derive insights. The outcomes of this analysis indicated that prior academic performance and participation in placement preparation classes have a significant impact on the final placement status of a student. Also, it was found that the gender or demography of students does not impact the placement status. The performance in benchmarking test was also found to have non-significant association with placement. This research is helpful for all the stakeholders in this domain including students, faculty, placement wing, management of the institute and recruiters as they get some key points out of this analysis.

Key Words: Student placement, Machine Learning, Data Analytics, Chi-Square test, Logistic Regression, Academic performance

1. Introduction

Quality placements hold great importance for students as well as educational institutions. It helps a student build a strong foundation for the professional career ahead. A good placement record gives a competitive edge to a college or university in the education market. The placement record is anticipated to be affected by several parameters like the academic performance of students, demography of students, dynamic changes in the job market and effectiveness of placement preparation strategies adopted by the concerned educational institute. The aim of this research is to analyze the different factors that affect the placement. The outcome of this research can help to anticipate how many and who are the students likely to get placed at the end of the placement season. By doing so, the higher authorities of an institution can identify the necessary steps to improve in the areas where students are lagging. Also, by analyzing different parameters, it can be figured out where is need to work on an individual student so that he or she also makes it into the cut-off. This kind of analysis has the potential to help the placement department or management to better plan the money to be spent on placement grooming.

This work discusses how to analyze the different factors that are anticipated to be affect the placement of students. Based on feedback from faculty, students, industry and placement team, the following research questions were formulated:

RQ1. Does the previous academic performance of a student affect the placement status?

RQ2. Does the gender of a student affect the placement status?

RQ3. Does the demography of a student affect the placement status?

RQ4. Does the percentage of attendance in placement preparatory (PEP) classes affect the placement status?

RQ5. Does the performance of a student in benchmarking tests affect the placement status?

The rest of the paper is organized as follows: Section 2 discusses the related work done in student placement analysis. Section 3 describes the dataset used in this study. Section 4 discusses outcomes or observations from the analysis in the form of answers to the research questions. Section 5 summarizes the conclusion and identifies possible lines for future work.

2. Related Work

The field of data science and machine learning has been buzzing around in the field of computing. Some of the recent applications include COVID-19 prediction systems taking clinical symptoms as input [1] and risk prediction in the stock market [2]. Applications specifically in the field of education data mining include predicting the academic performance of students [3] and predicting the joining behaviour of freshmen students enrolled at a university [4]. There are several works specifically focused on anticipating the likelihood of a student getting placed. The authors in [5] used machine learning algorithms to predict the placement status of MBA students studying at a University. The features used included academic performance, specialization and work experience. The authors in [6] used academic history, current academics, and socioeconomic background details of students to derive models for predicting the placement status of future batches. Explainable machine learning techniques have been used to understand which factors affect the placement status of engineering students and to analyze their relevant impact towards the placement status of students [7]. Attributes like the count of backlogs, the percentage in the qualifying examination and the current percentage have been used to develop model from data from past batches and to be used for predicting the placement status of future batches [8]. The authors in [9] used knowledge discovery and data mining to develop a model that can help teachers predict the placement class of a student, a step to be passed by students in Indonesia. The authors in [10] address the challenges associated with predicting employability. The challenges are associated due to ever-evolving labour market, the need for developing new competencies and variations in the data related to subjects. The authors in [11] evaluate the ICT-based school selection and placement systems adopted by Ghana Education Service in 2005 to improve the earlier manual system. An attempt to comprehend how graduates are hired has been made to develop a model for predicting students' employability index. The authors achieved an accuracy of more than 80% using the XGBoost algorithm [12].

3. Material (Data set) and Methods

Dataset: The subjects in this study consisted of students of MCA (Masters in Computer Applications), a postgraduate programme at a reputed university in North India. The features considered were demographic details, previous academic performance, attendance in placement preparatory classes and performance in different sections of the benchmarking test

(CoCubes). Table 1 compiles the different features along with their data type and a brief description.

Table 1. Dataset description

S.no	Attribute	Data type	Description
1	StudentID	Alphanumeric	A unique identifier for each student
2	State	Character	State which a student hail from
3	Gender	Character	Gender of a student (male or female)
4	10th Marks	Numeric	Percentage marks of a student in 10th Class
5	12th Marks	Numeric	Percentage marks of a student in 12th Class
6	Graduation Marks	Numeric	Percentage marks of a student in graduation
7	PEP Attendance	Numeric	Attendance % of a student in PEP classes
8	WET	Numeric	A section in the COCUBES test
9	ART	Numeric	Benchmarking test in COCUBES
10	EUT	Numeric	Benchmarking test in COCUBES
11	QUT	Numeric	Benchmarking test in COCUBES
12	WET	Numeric	Benchmarking test in COCUBES
13	Coding	Numeric	Benchmarking test in COCUBES
14	EAT	Numeric	Benchmarking test in COCUBES

Methods: For pre-processing and statistical significance, the following methods were used:

a) **Pre-processing**: A new attribute 'Throughout' was derived from given dataset. The value of this attribute was set to True if a student is throughout 1st class (at least 60%) else False. Moreover, missing values for 12th and Graduation marks were replaced with their mean values respectively.

b) **Statistical Significance:** For testing statistical significance of association, the Chi-Square test was used [13]. The Chi Square test for independence is used to determine if there is a significant association between two categorical variables. Performing this test involve defining the null hypothesis (H_0) and alternative hypothesis (H_1). The null hypothesis states that there is no association between the two categorical variables, while the alternative hypothesis states that there is a significant association. For testing statistical association between a continuous (independent) variable and categorical (dependent) variable, logistic regression was used [14]. The null hypothesis states that coefficient (beta) of the predictor variable in the logistic regression model is zero indicating no association with the categorical dependent variable.

4. Results and discussion

This section summarizes the outcome of the analysis in the form of answers to the research questions as framed in section 1:

RQ1. Does the previous academic performance of a student affect the placement status?

To answer this question, the idea was to analyze the impact of performance in 10th, 12th and Graduation separately. Four intervals, 90-100, 80-90, 70-80 and 60-70, were created for the performance of marks.

a) **Marks in 10th Vs Placement Status:** Table 2 compiles the placement percentage for each interval of marks in 10th class. As shown in figure 1, it can be observed that students who have got percentage in the range of 90-100 are having highest placement percentage.

Table 2. Marks in 10th Vs Placement %

Marks Slab in 10th	Count	Placed	Placement (%)
90-100	14	12	85.7
80-90	51	42	82.4
70-80	61	46	75.4
60-70	57	45	78.9

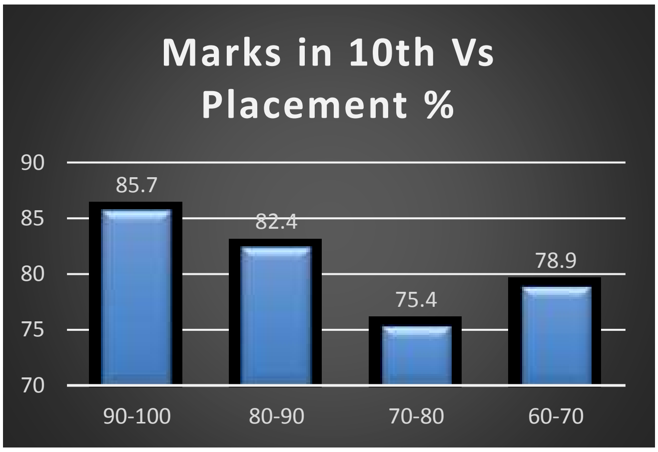

Figure 1. Marks in 10th Vs placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between 10th marks and placement status

H_1: There is a significant association between 10th marks and placement status

A test of independence using logistic regression was performed to examine the association between 10th marks percentage and the placement status of a student. The relation between these variables was found ***significant***, X^2 (1, N = 213) = 12.0155, p = .0005.

b) **Marks in 12th Vs Placement Status:** Table 3 compiles the placement percentage for each interval of marks in 12th class. It can be observed from figure 2 that students above

90 per cent are all getting placed. For the rest of the slabs for performance in 12th, the placement percentage is almost around 70%.

Table 3. Marks in 12th Vs Placement %			
Marks Slab in 12th	Count	Placed	Placement (%)
90-100	3	3	100.0
80-90	31	22	71.0
70-80	40	30	75.0
60-70	76	56	73.7

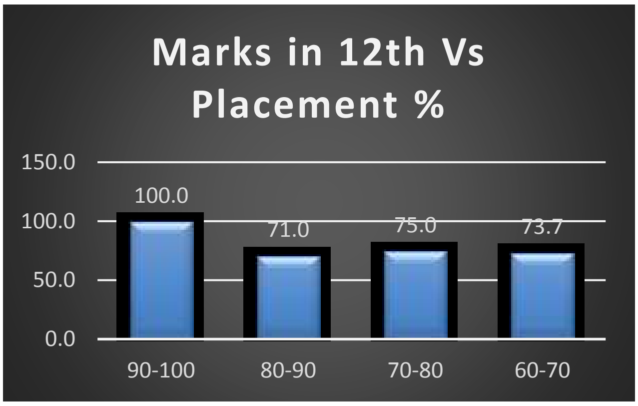

Figure 2. Marks in 12th Vs placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between 12th marks and placement status

H_1: There is a significant association between 12th marks and placement status

A test of independence using logistic regression was performed to examine the association between 12th marks percentage and the placement status of a student. The relation between these variables was found ***significant***, X^2 (1, N = 211) = 5.3642, p = .0038.

c) **Marks in Graduation Vs Placement Status:** Table 4 compiles the placement percentage for each interval of marks in Graduation. From figure 3, it is reflecting that those students who score above 90 have the highest chances of getting placed followed by those who score above 80.

Table 4. Graduation Marks Vs Placement %			
Marks Slab in 12th	Count	Placed	Placement (%)
90-100	3	3	100
80-90	22	18	81.8
70-80	101	67	66.3
60-70	72	51	70.8

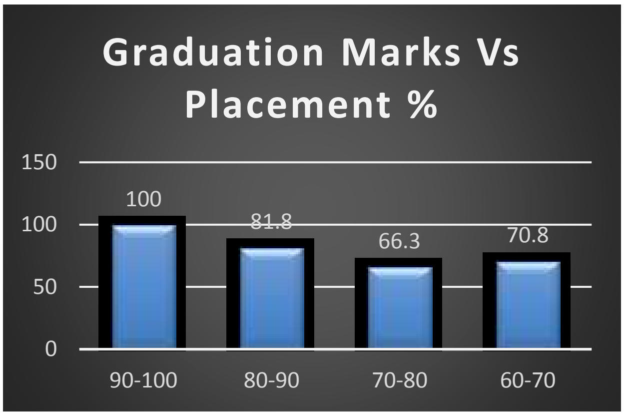

Figure 3. Graduation Marks Vs placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between Graduation marks and placement status

H_1: There is a significant association between Graduation marks and placement status

A test of independence using logistic regression was performed to examine the association between graduation marks percentage and the placement status of a student. The relation between these variables was found ***not significant***, X^2 (1, N = 187) = 2.0184, p = .1554.

d) **Throughout 1st class Vs Placement Status:** Table 5 analyzes the placement percentage of students who scored 60% marks throughout (10th, 12th and Graduation). As clear from figure 4, students who scored 60% throughout and students who scored first division (> 60%) in the 10th class have a higher placement percentage.

Table 5. Throughout 60% Vs Placement %			
Class	Count	Placed	Placement (%)
10th	183	145	79%
12th	150	111	74%
graduation	198	139	70%
Throughout 60%	129	101	78%

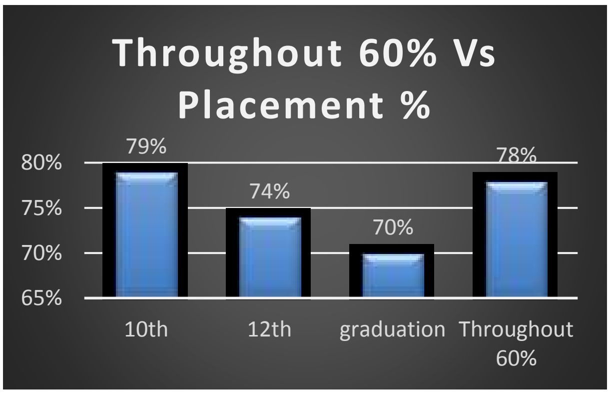

Figure 4. Throughout 60% Vs placement %

Statistical Significance: A chi-square test of independence was performed to examine the relation between student being consistent first class and the student placement status. The relation between these variables was found ***significant***, X^2 (1, N = 213) = 9.7318, p = .0018. So, students who are doing consistently good are likely to perform better in placement drives.

RQ2. Does the gender of a student affect the placement status?

To answer this question, gender-wise placement percentage was computed. Table 6 compiles gender-wise placement percentage. As clear from figure 5, the placement status of students is not affected by his/her gender.

Table 6. Gender vs percentage

Gender	count	Placed	Placed (%)
Male	161	113	70.2
Female	52	37	71.2

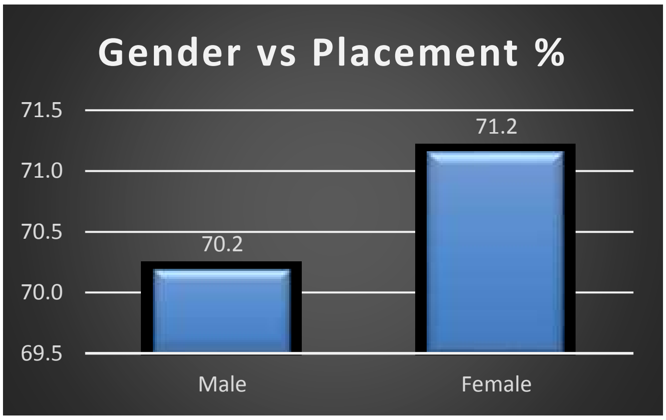

Figure 5. Gender-wise placement %

Statistical Significance: The following hypothesis was framed to test the significance between gender and placement status:

H_0: There is no association between gender and placement status

H_1: There is a significant association between gender and placement status

A chi-square test of independence was performed to examine the relation between student gender and the student placement status. The relation between these variables was ***not significant***, X^2 (1, N = 213) = 0.0177, p = .894. So, male and female students are likely to perform equally well. It is an indication that the university is providing equal opportunity to both genders in terms of placement grooming as well as placement opportunities.

RQ3. Does the demography of a student affect the placement status?

To answer this question, a state-wise count of students and placement percentage was computed. Only those states were picked where at least five students were there. Table 7 compiles the state-wise count of students and placement percentage. As clear from figure 6, Uttar Pradesh, Bihar and Odisha are the leading states in terms of student count. However, the top three states in terms of placement percentage are Delhi, Haryana and Andhra Pradesh. Also, students from Jharkhand performed miserably as no student could get placed.

Table 7. State-wise Placement %

State	Count	Placed	Placement %
Uttar Pradesh	49	38	77.6
Bihar	45	29	64.4
Odisha	19	11	57.9
Punjab	16	11	68.8
Rajasthan	15	8	53.3
Uttarakhand	11	7	63.3
Jammu & Kashmir	11	9	81.3
Haryana	9	8	88.9
Delhi	6	6	100.0
Jharkhand	6	6	0.0
Andhra Pradesh	6	5	83.0
Madhya Pradesh	5	2	40.0

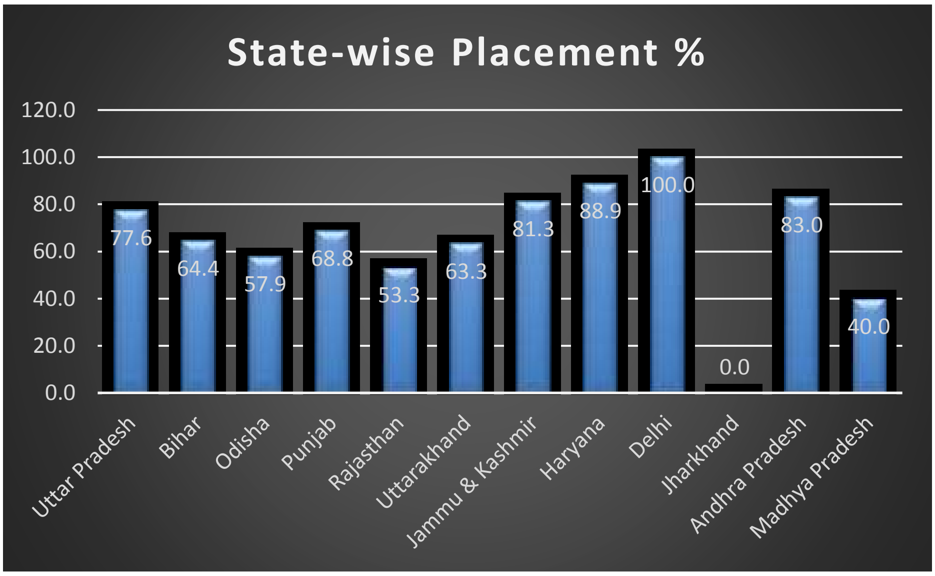

Figure 6. State-wise placement %

Test of Statistical Significance: The following hypothesis was framed:

H_0: There is no association between student demography and placement status

H_1: There is a significant association between student demography and placement status

A chi-square test of independence was performed to examine the relation between student demography and the student placement status. The relation between these variables was ***not significant***, X^2 (1, N = 144) = 4.619, p = .3287 at 5% level of significance. It is an indication that there is no demography-based biasedness in placement grooming and opportunities.

RQ4. Does the percentage of attendance in placement preparatory (PEP) classes affect the placement status?

To answer this question, students were classified into three slabs (<50%, >=50% and <70%, >=70%) based on their attendance percentage. Table 8 compiles slab-wise count and placement percentage. From figure 7, it is clear that placement percentage increases with the increase in PEP attendance percentage.

Table 8. PEP Attendance Vs Placement %

PEP Attendance % Slab	Count	Placed	Placed %
<50%	27	12	44.4
>=50% & <75%	19	14	73.7
>=75%	142	108	76.1

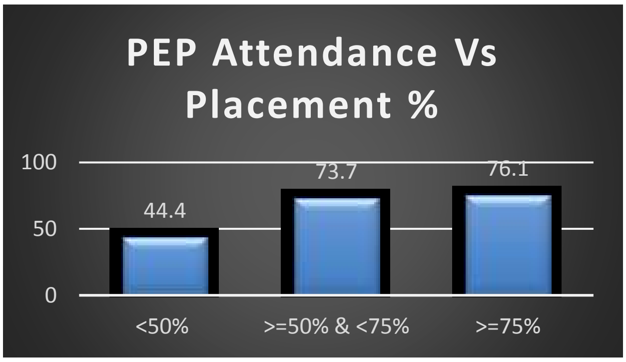

Figure 7. PEP Attendance % Vs Placement %

Test of Statistical Significance: The following hypothesis was framed:

H_0: There is no association between PEP attendance and placement status

H_1: There is a significant association between PEP attendance and placement status

A chi-square test of independence was performed to examine the relation between PEP attendance and the student placement status. The chi-square statistic, X^2 (1, N = 188) = 11.1333. The p-value is .003823. The result is **significant** at $p < .05$. So, PEP classes are helpful for students to increase their chances of getting placed.

RQ5. Does the performance of a student in benchmarking tests affect the placement status?

The benchmarking test was CoCubes in our case. This benchmarking test has different sections corresponding to the selection process of different recruit companies. The sections of Cocubes that were analyzed include ART, EUT, QUT, WET, Coding and EAT. To answer this question, we analyzed the placement % vs performance in every section of this test. Students were classified into four performance slabs (<40%, 40-60%, 60-80% and >80%).

a) **ART Vs Placement Status:** Table 9 compiles placement percentages against different performance slabs for ART section. When we look at figure 8, what we find is that those students who have scored greater than 80% in the ART slab have got all been placed.

Table 9. ART Slab Vs Placement %

ART Slab	Count	placed	Placement %
<=40%	58	36	62.1
>40% & <=60%	94	70	74.5
>60% & <=80%	29	21	72.4
>80%	3	3	100.0

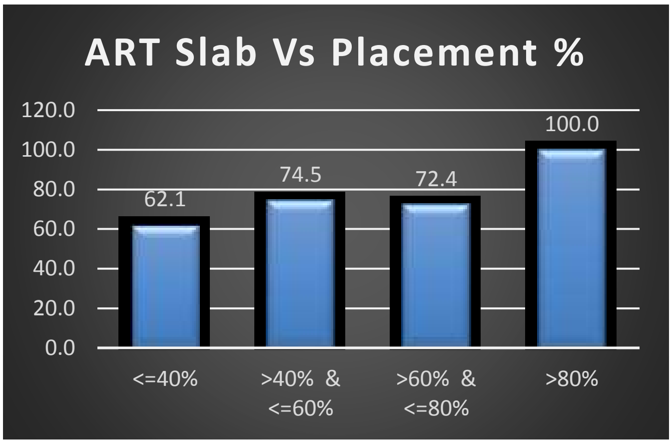

Figure 8. ART Slab Vs Placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between ART marks and placement status

H_1: There is a significant association between ART marks and placement status

A test of independence using logistic regression was performed to examine the association between ART marks and the placement status of a student. The relation between these variables was found ***not significant***, X^2 (1, N = 185) = .3641, p = .5462.

b) **EUT Vs Placement Status:** Table 10 compiles placement percentages against different performance slabs for EUT section. As shown in figure 9, the placement percentage is not showing a regular increasing trend. However, the placement % for students scoring above 80% is 100%.

Table 10. EUT Slab Vs Placement %

EUT Slab	Count	placed	Placement %
<=40%	31	21	67.7
>40% - <=60%	71	44	62.0
>60% - <=80%	79	61	77.2
>80%	4	4	100.0

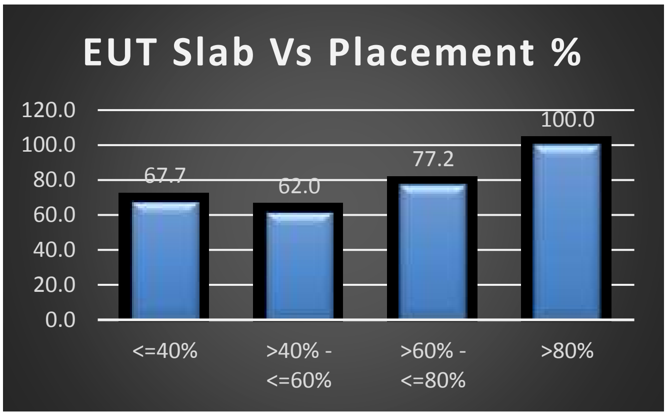

Figure 9. EUT Slab Vs Placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between EUT marks and placement status

H_1: There is a significant association between EUT marks and placement status

A test of independence using logistic regression was performed to examine the association between EUT marks and the placement status of a student. The relation between these variables was found ***not significant***, X^2 (1, N = 185) = .0252, p = .8740.

c) **QUT Vs Placement Status:** Table 11 compiles placement percentages against different performance slabs for QUT section. As shown in figure 10, as the QUT Slab percentage increases the placement percentage also increases.

Table 11. QUT Slab Vs Placement %

QUT SLAB	Count	placed	Placement %
<=40%	106	75	70.8
>40% - <=60%	56	40	71.7
>60% - <=80%	16	12	75.0
>80%	4	3	75.0

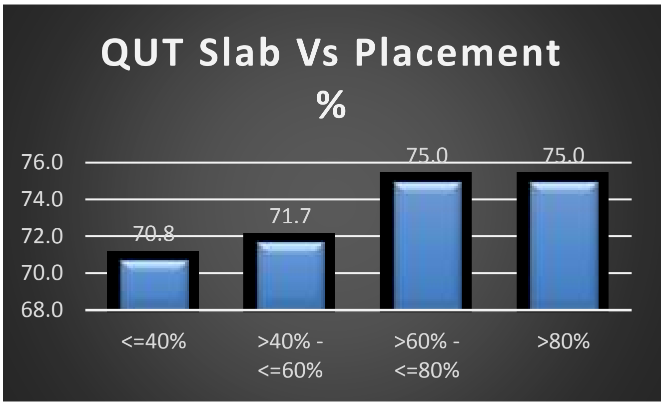

Figure 10. QUT Slab Vs Placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between QUT marks and placement status

H_1: There is a significant association between QUT marks and placement status

A test of independence using logistic regression was performed to examine the association between QUT marks and the placement status of a student. The relation between these variables was found ***not significant***, X^2 (1, N = 185) = .0493, p = .8244.

d) **WET Vs Placement Status:** Table 12 compiles placement percentages against different performance slabs for WET section. As shown in figure 11, the WET slab is showing some irregular trends. Those who scored between (>40% - <=60%) have got highest placement percentage.

Table 12. WET Slab Vs Placement %

WET slab	Count	placed	Placement %
<=40%	45	27	60.0
>40% - <=60%	63	51	81.0
>60% - <=80%	60	44	73.3
>80%	13	8	61.5

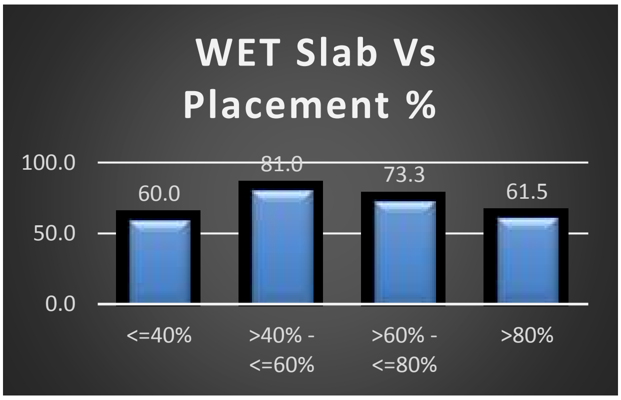

Figure 11**.** WET Slab Vs Placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between WET marks and placement status

H_1: There is a significant association between WET marks and placement status

A test of independence using logistic regression was performed to examine the association between WET marks and the placement status of a student. The relation between these variables was found ***not significant***, X^2 (1, N = 185) = .0674, p = .7952.

e) **Coding Vs Placement Status:** Table 13 compiles placement percentages against different performance slabs for Coding section. As shown in figure 12, we are finding out that overall the graph is showing an upward trend. Though at the end, there is a slight drop.

Table 13. Coding Slab Vs Placement %

Coding Slab	Count	placed	Placement %
<=40%	132	90	68.2
>40% - <= 60%	21	17	81.0
>60% - <= 80%	1	1	100.0
>80%	26	21	80.8

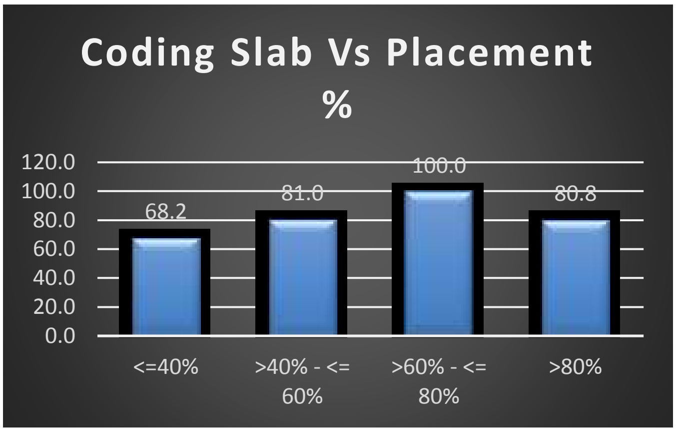

Figure 12. Coding Slab Vs Placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between Coding marks and placement status

H_1: There is a significant association between Coding marks and placement status

A test of independence using logistic regression was performed to examine the association between Coding marks and the placement status of a student. The relation between these variables was found ***not significant***, X^2 (1, N = 185) = .0561, p = .8128.

f) **EAT Vs Placement Status:** Table 14 compiles placement percentages against different performance slabs for EAT section. For EAT, as shown in figure 13, we are finding a regular trend. As the EAT slab percentage increases the placement percentage also increases.

Table 14. EAT Slab Vs Placement %

EAT SLAB	Count	placed	Placement %
<=40%	81	56	69.1
>40% - <=60%	63	45	71.4
>60% - <= 80%	33	25	75.8
>80%	4	4	100.0

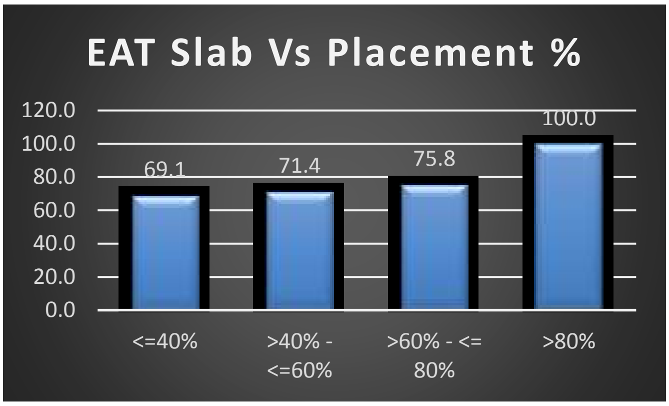

Figure 13. EAT Slab Vs Placement %

Statistical Significance: The following hypothesis was framed:

H_0: There is no significant association between EAT marks and placement status

H_1: There is a significant association between EAT marks and placement status

A test of independence using logistic regression was performed to examine the association between EAT marks and the placement status of a student. The relation between these variables was found ***not significant***, X^2 (1, N = 185) = 1.003, p = .3172.

5. Conclusion and future work

First, the prior academic performance affects the placement status of a student. A statistically significant association was observed between academic performance (in 10th and 12th) and placement. Marks in graduation was not having a significant association with placement status. Also, the students who scored throughout 1st class were likely to have better placement percentage. Second, the gender of a student does not affect the placement as both genders are performing almost equally well (70% for males and 71% for females). This indicates that the training activities and placement opportunities are at par for both genders. Third, it was found that although students coming from Delhi, Haryana and Andhra Pradesh are performing well, the overall association between demography and placement was not significant. Fourth, it was found the participation in placement preparation classes is very important as a significant association was observed between PEP attendance and placement. Lastly, a thorough analysis of performance in ART, QUT, EUT, WET, CODING and EAT sections of the COCUBES benchmarking test showed that the performance in these sections was not having a significant association with placement status. This gives a direction for further investigation that the benchmarking test performance may not be true indicator of chances of a student getting placed.

There are several interesting directions to extend this research. Does the academic background during graduation (computer science or non-computer science) affect the placement status? What is the placement status of those students who avail of scholarships from the college at the time of admission? Are they performing relatively better in terms of placement percentage? Another question we will like to explore in the future is the impact of attendance in regular courses (apart from PEP attendance) on performance in placement drives. We also plan to use machine learning techniques to develop predictive models that can be used to predict the placement status of a student well in advance.

7. References

[1] Rai, K. (2022, October). Predicting Covid-19 based on symptoms using machine learning techniques implemented in Python. In *AIP Conference Proceedings* (Vol. 2555, No. 1, p. 050002). AIP Publishing LLC.

[2] Singh, B., Henge, S. K., Sharma, A., Menaka, C., Kumar, P., Mandal, S. K., & Debtera, B. (2022). ML-Based Interconnected Affecting Factors with Supporting Matrices for Assessment of Risk in Stock Market. *Wireless Communications and Mobile Computing*, *2022*.

[3] Kumar, P., & Sharma, M. (2020). Predicting Academic performance of international students using machine learning techniques and human interpretable explanations using LIME—A case study of an Indian University. In *International Conference on Innovative Computing and Communications* (pp. 289-303). Springer, Singapore.

[4] Kumar, P., Kumar, V., & Sobti, R. (2020, July). Predicting joining behaviour of freshmen students using machine learning–A case study. In *2020 International Conference on Computational Performance Evaluation (ComPE)* (pp. 141-145). IEEE.

[5] Rai, K. (2022). STUDENTS' PLACEMENT PREDICTION USING MACHINE LEARNING ALGORITHMS. *South Asia*, *8*(5).

[6] Kumar, P., Sharma, M., & Sood, S. (2019). Anticipating placement status of students using machine learning. *J Gujarat Res Soc*, *21*(6), 374-8588.

[7] Pawan Kumar, Manmohan Sharma, Sandeep Kaur, Sanjay Sood, "Identifying factors affecting placement status of engineering students using explainable machine learning", Journal of Emerging Technologies and Innovative Research (JETIR), Volume 5, Issue 12, December 2018.

[8] Manvitha, P., & Swaroopa, N. (2019). Campus placement prediction using supervised machine learning techniques. *Int. J. Appl. Eng. Res*, *14*(9), 2188-2191.

[9] Pratiwi, O. N. (2013, August). Predicting student placement class using data mining. In *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (pp. 618-621). IEEE.

[10] Mezhoudi, N., Alghamdi, R., Aljunaid, R., Krichna, G., & Düştegör, D. (2021). Employability prediction: a survey of current approaches, research challenges and applications. *Journal of Ambient Intelligence and Humanized Computing*, 1-17.

[11] Owusu, A., & Nettey, J. N. A. (2022). Evaluation of computerized school selection and placement system in Ghana using fit and viability theory. *Education and Information Technologies*, 1-28.

[12] Sharma, P., Anand, D., & Kapoor, N. (2022, July). Student's Employability Indexing Using Machine Learning Approach. In *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)* (pp. 242-249). IEEE.

[13] McHugh M. L. (2013). The chi-square test of independence. Biochemia medica, 23(2), 143–149. https://doi.org/10.11613/bm.2013.018

[14] Bonney, G. E. (1987). Logistic Regression for Dependent Binary Observations. Biometrics, 43(4), 951–973. https://doi.org/10.2307/2531548