



TEXAS McCombs

MIS S381N Data Science Programming

Will You Find a Job?



Group 9

Abhilash Bagde, Rohan Chimne,

Michael Ovassapian, Isaac Tuan, Elmer Wang



Data Introduction

College Student Placement Factors Dataset

Source : kaggle

Author : Sahil Islam007

Description : Includes features like IQ, CGPA, internships, communication skills, and more

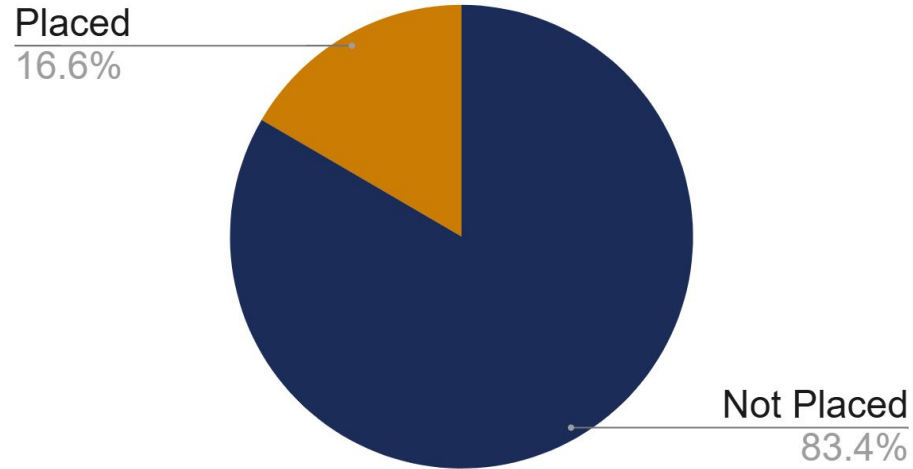
Data Structure : 10,000 rows, 9 columns

	College_ID	IQ	Prev_Sem_Result	CGPA	Internship_Experience	Extra_Curricular_Score	Communication_Skills	Projects_Completed	Placement
0	CLG0030	107	6.61	6.28	No	8	8	4	No
1	CLG0061	97	5.52	5.37	No	7	8	0	No
2	CLG0036	109	5.36	5.83	No	3	1	1	No
3	CLG0055	122	5.47	5.75	Yes	1	6	1	No
4	CLG0004	96	7.91	7.69	No	8	10	2	No
5	CLG0015	96	5.26	5.32	No	5	8	0	No

Features	Description	dtype
College_ID	Unique Student ID of the college	Object
IQ	Student's IQ score	int64
Prev_Sem_Result	GPA of previous semester (range: 5.0 to 10.0)	float64
CGPA	Cumulative GPA (range: 5.0 to 10.0)	float64
Internship_Experience	Whether the student had internship (Yes/No)	Object → Bool
Extra_Curricular_Score	Involvement in extracurriculars (from 0 to 10)	int64
Communication_Skills	Soft skill rating (from 1 to 10)	int64
Projects_Completed	Number of projects completed (0 to 5)	int64
Placement	Final placement result (Yes/No)	Object → Bool

Exploratory Data Analysis

Across the 10,000 students



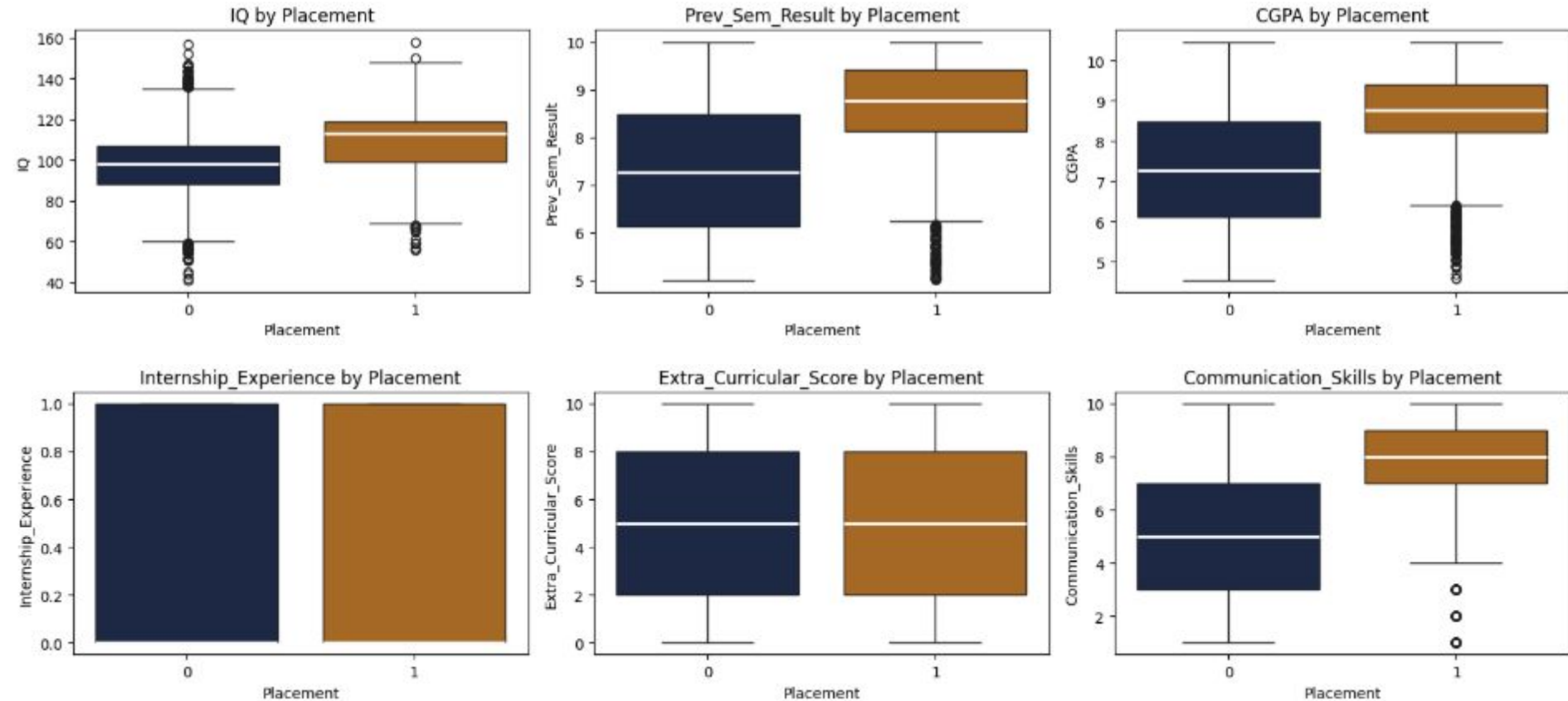
We calculated the baseline, i.e. the benchmark placement rate.

Out of all students:

- **16.6% were placed**
- **83.4% were not placed**

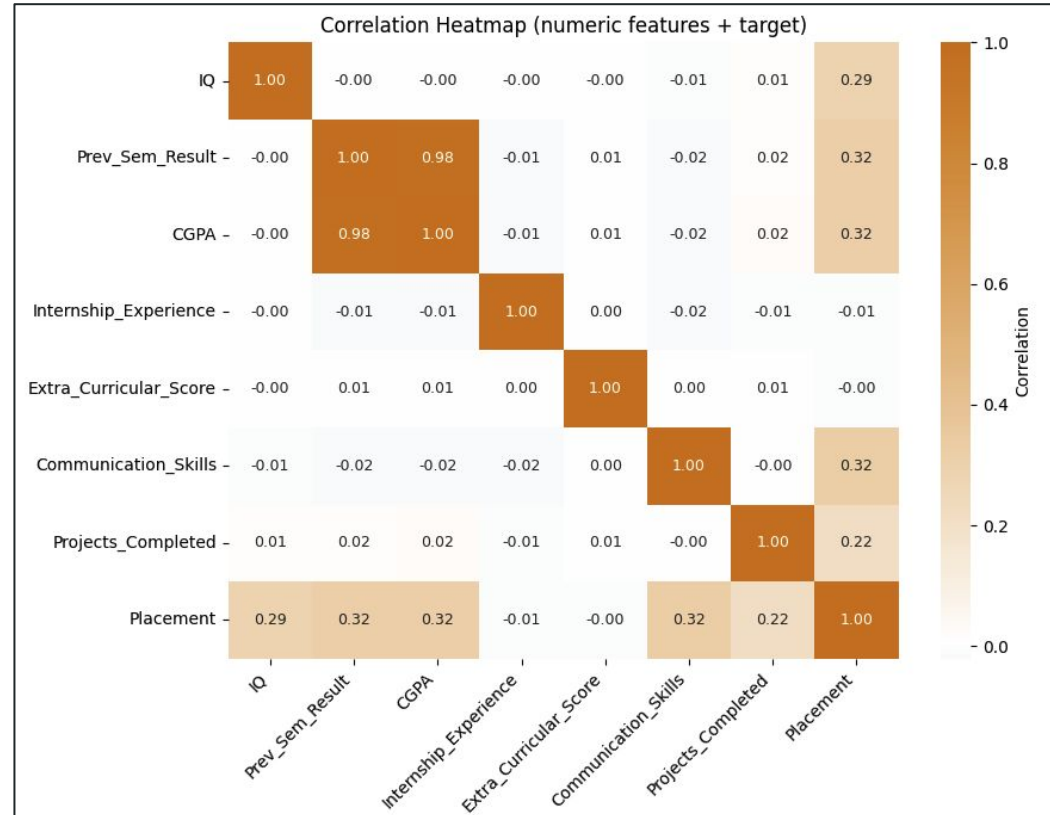


Feature Distributions: Placed vs. Not Placed Students

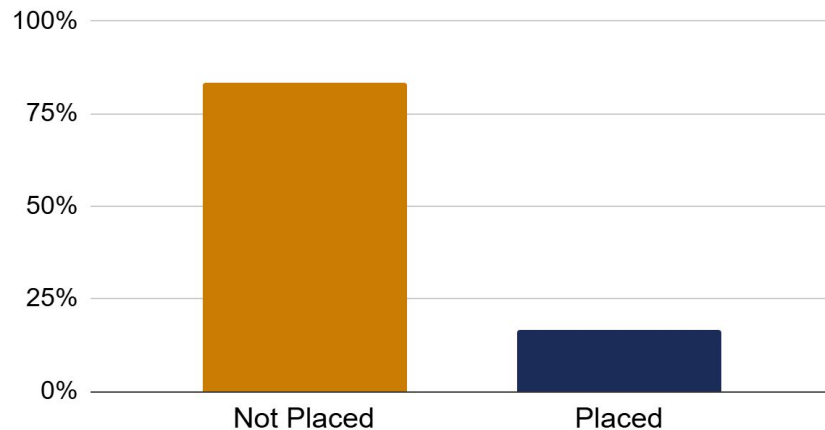


Correlation Analysis - Key Features Linked to Placement

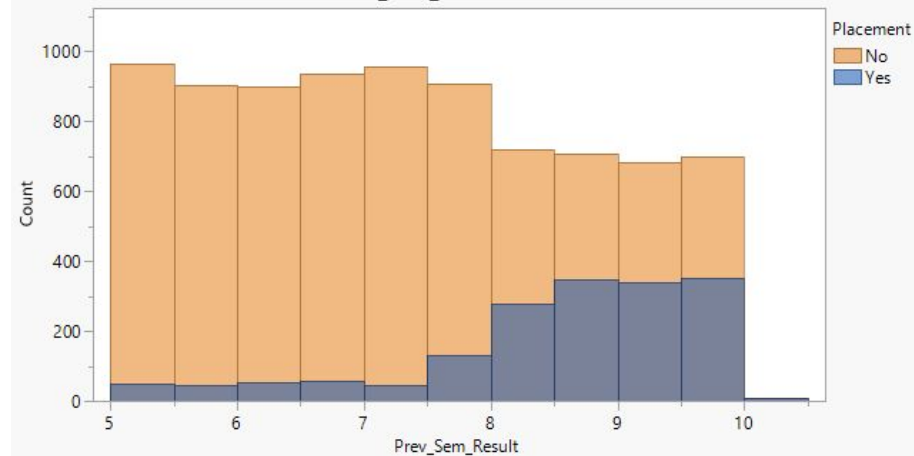
- Top factors influencing placement** – Communication Skills, CGPA, Previous Semester Result, and IQ show the strongest positive relationships.
- Minimal impact** – Internship Experience and Extracurricular Score show near-zero correlation with placement.



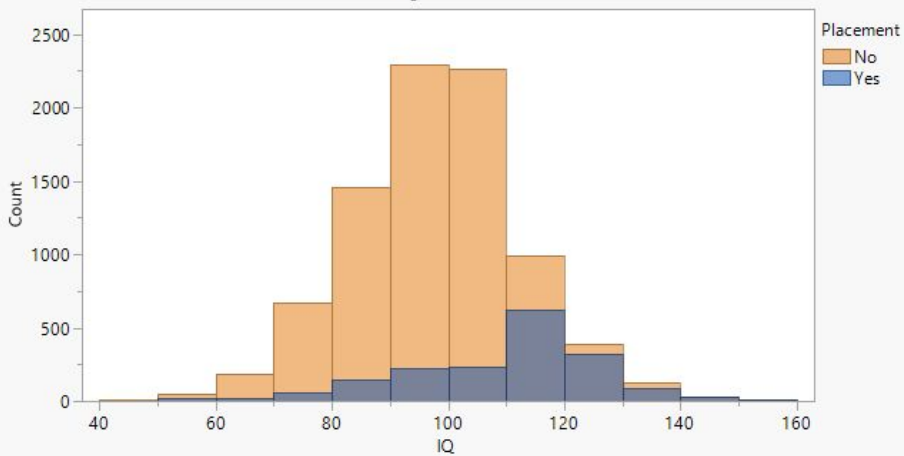
Placement



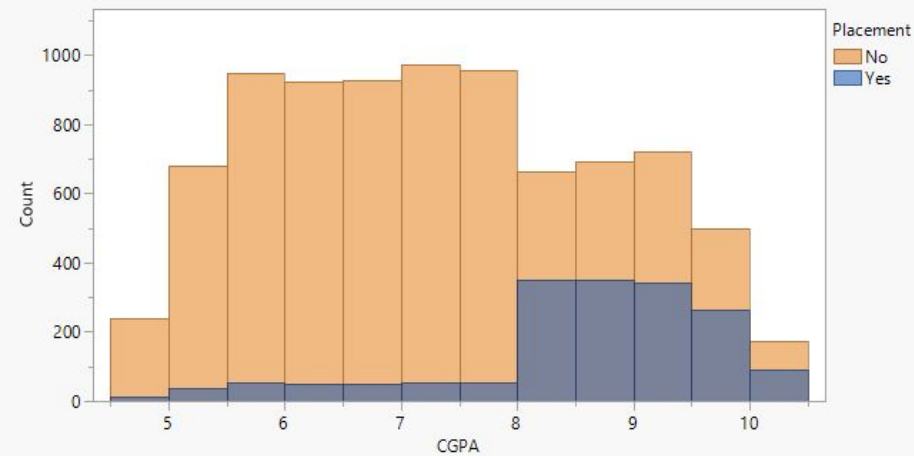
Prev_Sem_Result

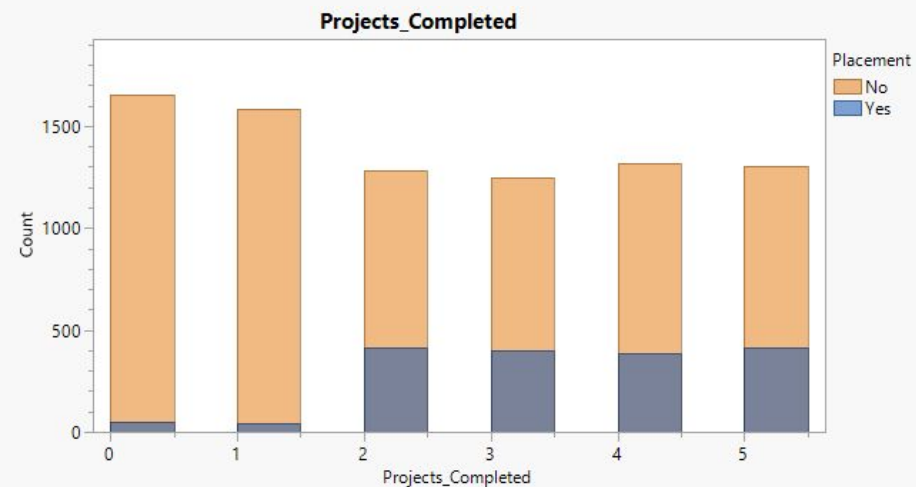
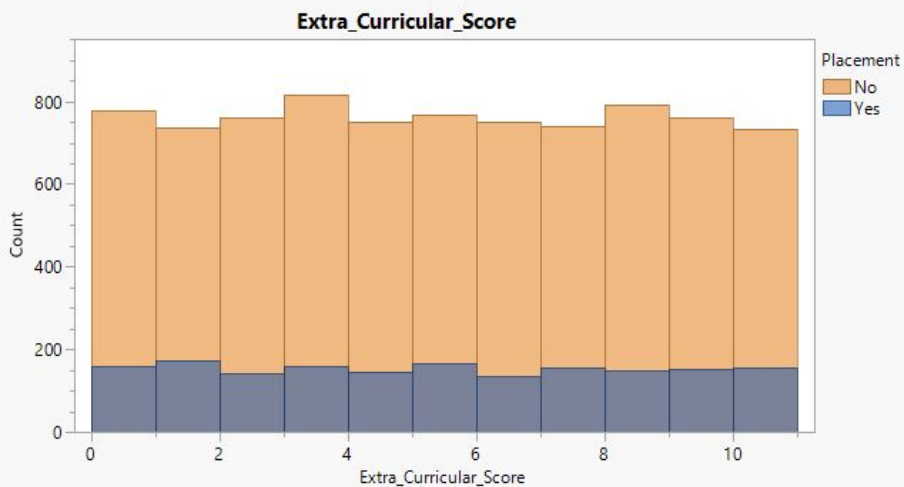
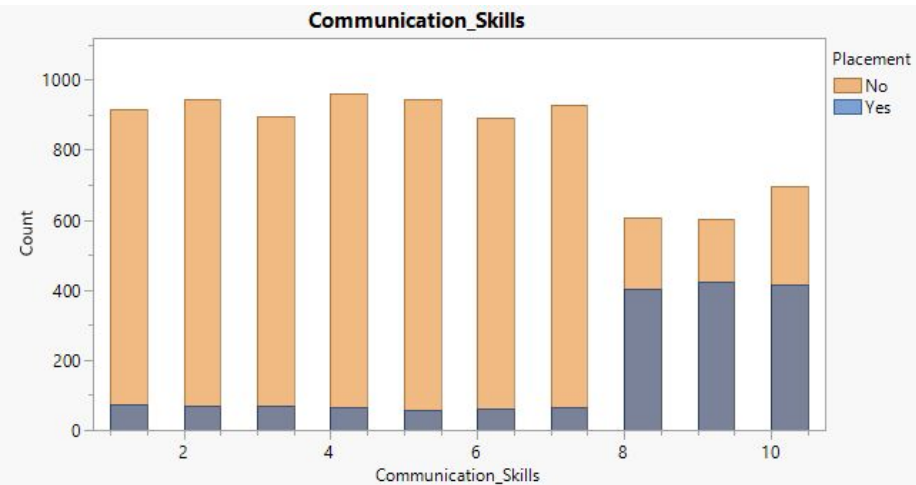
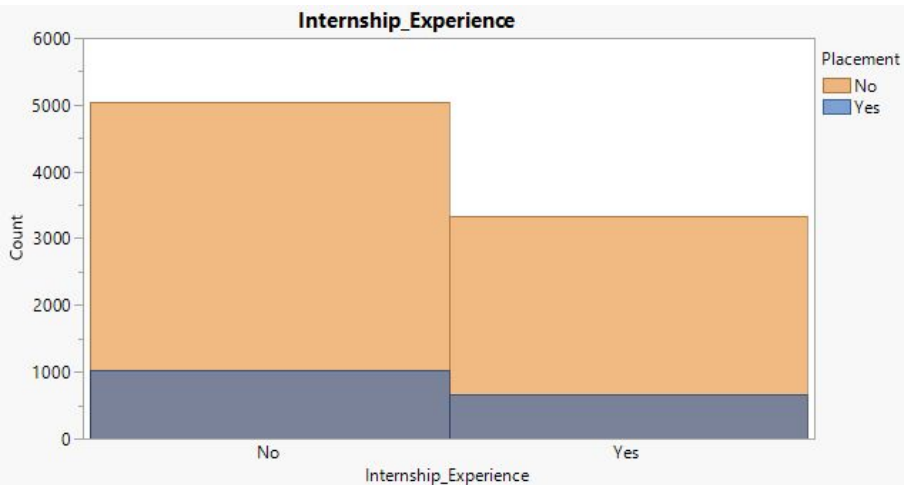


IQ



CGPA





Modeling Process

Logistic Regression

Split test size : 30%

Baseline accuracy : 83.2%

Training accuracy : 90.0%

Testing accuracy : 90.7%

Parameters	Coefficient	P > z
Internship_Experience	0.0371	0.639
IQ	1.6368	0.000
Prev_Sem_Result	0.3021	0.129
CGPA	1.5426	0.000
Extra_Curricular_Score	-0.0509	0.188
Communication_Skills	1.8801	0.000
Projects_Completed	1.1662	0.000

Nearest Neighbors

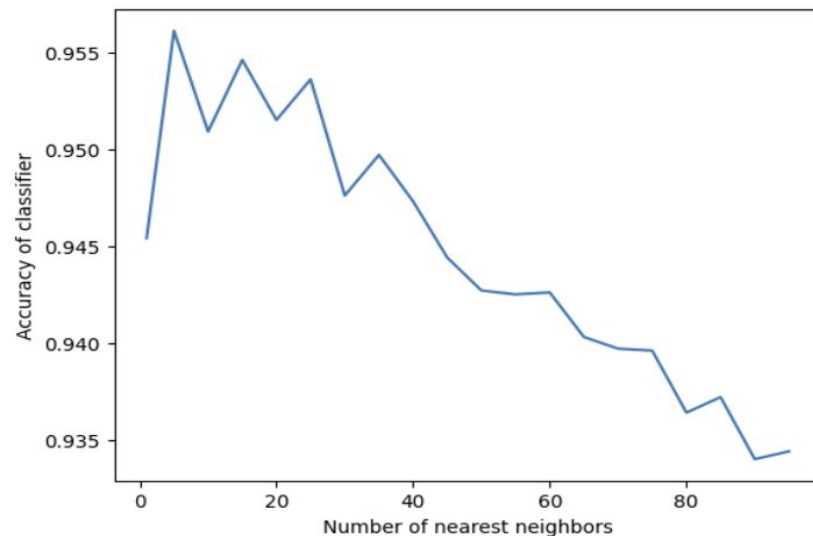
Split test size : 30%

Number of neighbors : 5

Weight : uniform, usual distance

Training accuracy : 96.5%

Testing accuracy : 95.7%



Classification Trees

Model	Settings	Training Accuracy	Test Accuracy
Decision Trees	criterion = 'entropy'	100%	100%
Bagging	criterion = 'entropy'	100%	100%
Random Forest	n_estimators = 10	100%	99.9%
Gradient Boosting	n_estimators = 200, max_depth = 4	100%	100%

Feature	IQ	Projects Completed	CGPA	Communication Skills	Others
Importance	0.3187	0.2535	0.2440	0.1838	0

Conclusion

01

Selecting Predictors

Plotted placed vs not placed for each variable

Looked for thresholds in placement rates

Checked model importance (Logistic Regression & Random Forest)

Kept only variables strong in both patterns + model

02

Top Predictors

Communication $\geq 8/10 \rightarrow$ biggest lift

≥ 2 projects \rightarrow clear jump in odds

1Q $\geq \sim 115 \rightarrow$ smaller lift beyond this point

GPA helps, but less once the top two are strong

03

Boosting Placement

Communication bootcamps : move 6-7 \rightarrow 8 +

Require ≥ 2 portfolio-grade projects (real clients / open-source)