College Student Placement Analysis

Name: Devanshu Sawarkar

PRN: 22070521090

Semester: 7 Section: C

Github: <https://github.com/DevanshuSawarkar/College_Student_Placement_Analysis>

1. **Introduction**

This project explores a dataset containing academic, personal, and placement details of students from a business school. The objective is to understand what factors influence a student’s likelihood of being placed and to prepare the dataset for predictive modeling using machine learning techniques.

1. **Dataset Overview**

* Dataset Name: college\_student\_placement\_dataset.csv
* Source: <https://www.kaggle.com/datasets/sahilislam007/college-student-placement-factors-dataset>
* Rows: 10000
* Columns: 10
  1. Dataset Description

This dataset simulates the academic and professional profiles of 10,000 college students, focusing on factors that influence placement outcomes. It includes features like IQ, academic performance, CGPA, internships, communication skills, and more.

The dataset is ideal for:

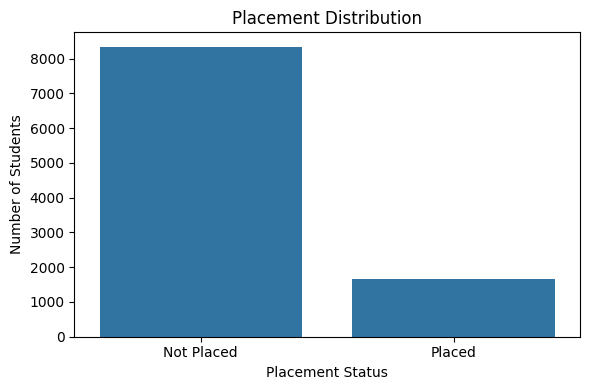
* Predictive modeling of placement outcomes
* Educational exercises in classification
* Feature importance analysis
* End-to-end machine learning projects
  1. Columns Description

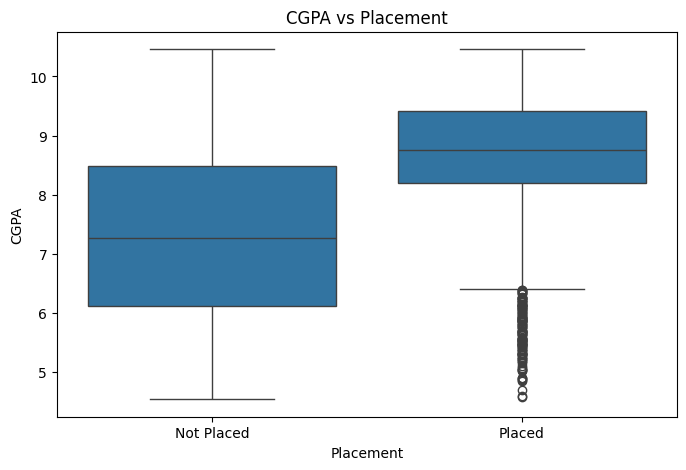
|  |  |
| --- | --- |
| **Column Name** | **Description** |
| College\_ID | Unique ID of the college (e.g., CLG0001 to CLG0100) |
| IQ Student’s | IQ score (normally distributed around 100) |
| Prev\_Sem\_Result | GPA from the previous semester (range: 5.0 to 10.0) |
| CGPA | Cumulative Grade Point Average (range: ~5.0 to 10.0) |
| Academic\_Performance | Annual academic rating (scale: 1 to 10) |
| Internship\_Experience | Whether the student has completed any internship (Yes/No) |
| Extra\_Curricular\_Score | Involvement in extracurriculars (score from 0 to 10) |
| Communication\_Skills | Soft skill rating (scale: 1 to 10) |
| Projects\_Completed | Number of academic/technical projects completed (0 to 5) |
| Placement | Final placement result (Yes = Placed, No = Not Placed) |

1. **Data Cleaning**

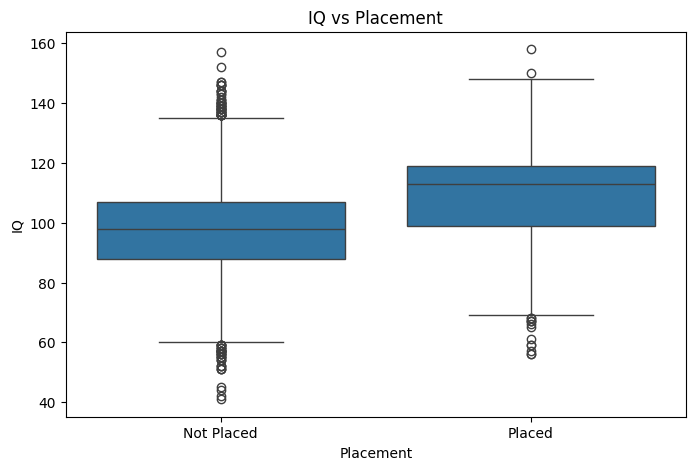
* Null Values: No null values found.
* Duplicate Values: No duplicate values found.
* Dropped ‘College\_ID’: As it has no predictive power.
* Encoded ‘InternshiP\_Experience’ and ‘Placement’ ‘Yes/No’ to binary (‘1/0’).

1. **Exploratory Data Analysis (EDA)**

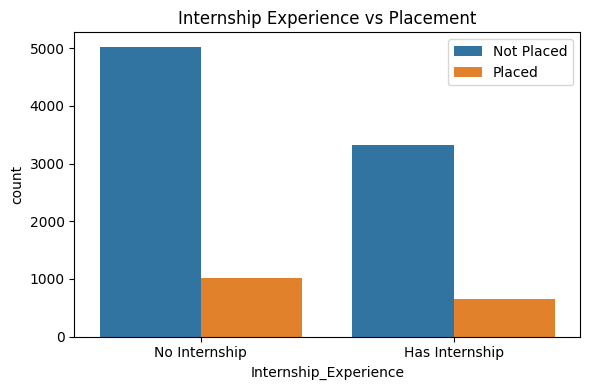
* Placement Status:
* A majority of students are not placed (about 83%).
* CGPA’s impact on placement:



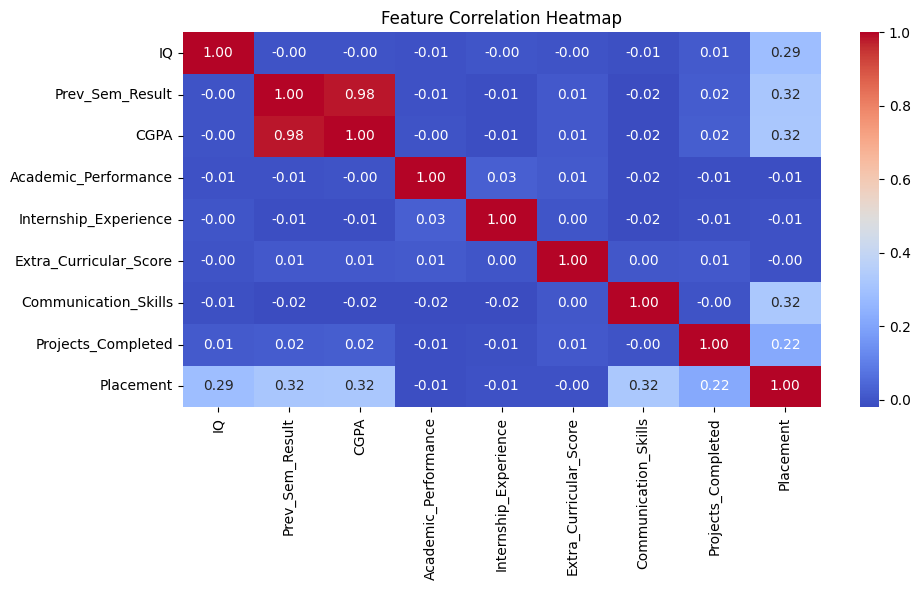
* Students with higher CGPA are more likely to be placed.
* Student’s IQ impact on placement:



* Similar trend: higher IQ may correspond with higher placement chances.
* Internship experience impact on placement:



* Students with internship experience (Yes) are more frequently placed.
* Heatmap of Feature Correlation:



* 1. Univariant Analysis Summary

Univariate analysis focused on each variable independently:

* + IQ and CGPA are normally distributed with mild right skew.
  + Academic Performance, Communication Skills, and Extra-Curricular Score show balanced spread.
  + Placement Outcome (target) is imbalanced:
    - Around 83% students are not placed, and only 17% are placed.
  + No missing values or duplicates were found.
  + Internship Experience is binary (Yes/No) and was encoded for modeling.
  1. Multivariant Analysis Summary

Multivariate analysis explored interactions between variables and their effect on placement:

* + CGPA vs Placement:
    - Students with higher CGPA were more likely to be placed.
    - Clear distinction in median CGPA between placed vs not placed students.
  + IQ vs Placement:
    - Students with higher IQ tended to have higher placement probability.
  + Internship Experience:
    - Strongly correlated with placement.
    - Students with internships had significantly higher placement rates.
  + Feature Correlation Heatmap:
    - Positive correlation among IQ, CGPA, Academic Performance, and Placement.
    - All features show varying but meaningful contributions to placement.

1. **Key Insights**

* Higher CGPA and IQ scores positively correlate with placement.
* Internship Experience greatly improves placement chances.
* Academic Performance, Communication Skills, and Projects also play a role.
* The target variable 'Placement' is imbalanced (majority not placed).

1. **Machine Learning Model Suggestion**

Based on EDA, Random Forest Classifier on 'Placement' will be the best suited. Random Forest Classifier is suitable due to:

* Ability to handle both numeric and categorical features.
* Handles imbalanced classes well.
* Provides feature importance.

1. **Conclusion**

This dataset analysis reveals that academic performance, work experience, and specialization significantly impact placement outcomes. Using EDA, we’ve identified meaningful patterns and prepared the dataset for ML modeling.

1. **References**

* Dataset: <https://www.kaggle.com/datasets/sahilislam007/college-student-placement-factors-dataset>
* Tools: Python (Pandas, Matplotlib, Seaborn)
* Github: <https://github.com/DevanshuSawarkar/College_Student_Placement_Analysis>