STUDENT PLACEMENT PREDICTION SYSTEM

A DATA SCIENCE AND ANALYTICS PROJECT REPORT

---

1. Introduction

Campus placements play a vital role in evaluating the academic and professional readiness of students. Educational institutions aim to understand the factors that influence student placement outcomes in order to improve training, curriculum design, and career guidance. With the availability of historical placement data, data science and machine learning techniques can be used to analyze patterns and predict placement outcomes effectively.

This project focuses on building a Student Placement Prediction System using data analysis, machine learning models, and business intelligence dashboards. The system helps identify key factors such as academic performance, work experience, gender, and specialization that influence placement decisions.

---

2. Objectives of the Project

The main objectives of this project are:

- To analyze historical student placement data
- To identify key factors influencing placement outcomes
- To build machine learning models to predict placement status
- To compare model performance using evaluation metrics
- To visualize insights using Python and Power BI dashboards
- To assist institutions in early identification of students requiring placement support

---

3. Dataset Description

The dataset used in this project is a Campus Placement dataset consisting of 215 student records. It contains academic, demographic, and placement-related attributes.

Key attributes include:

- Gender
- Secondary School Percentage (SSC)
- Higher Secondary Percentage (HSC)
- Degree Percentage
- MBA Percentage
- Specialization
- Work Experience
- Placement Status (Placed / Not Placed)

The dataset was cleaned and preprocessed before analysis.

---

4. Data Preprocessing

Data preprocessing was performed using Python and involved the following steps:

- Handling missing values
- Dropping irrelevant or redundant columns
- Encoding categorical variables such as gender, specialization, and work experience
- Converting placement status into binary form for modeling
- Exporting the cleaned dataset for Power BI visualization

The final cleaned dataset was saved as placement_cleaned.csv.

---

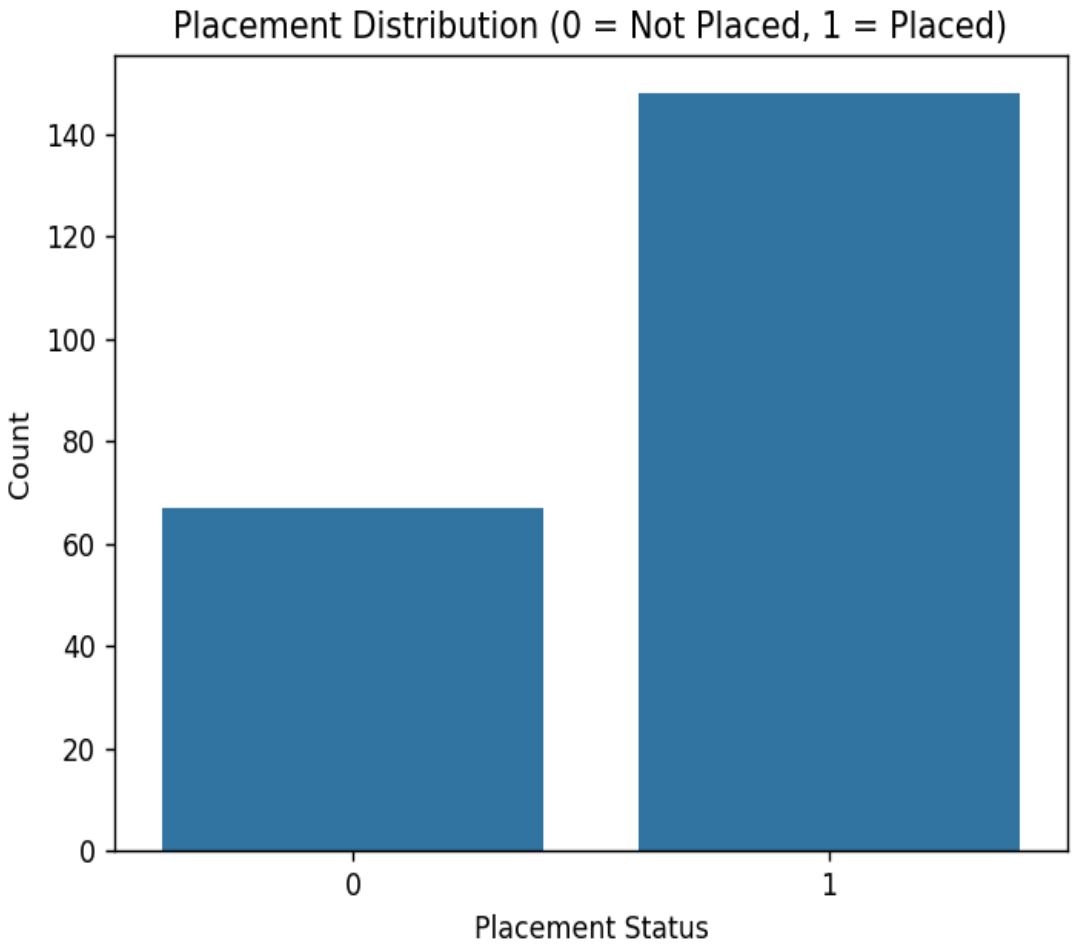
5. Exploratory Data Analysis (EDA)

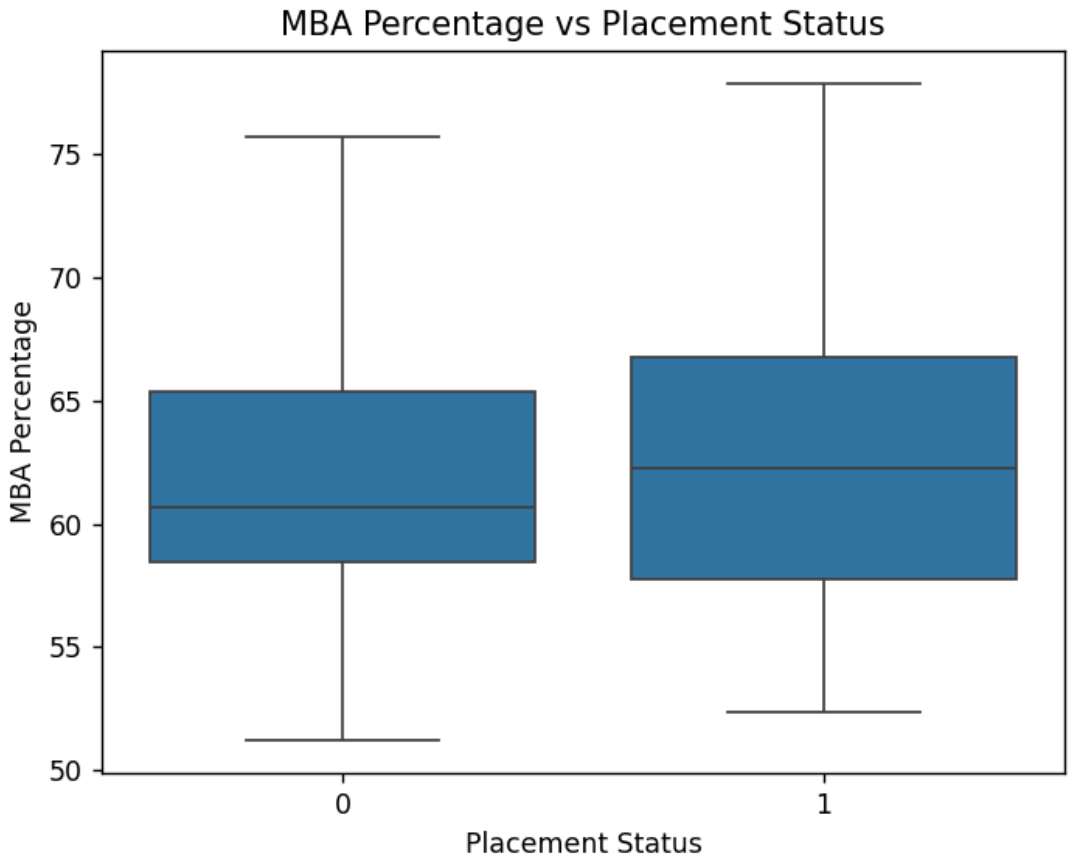
Exploratory Data Analysis was conducted to understand trends and relationships within the data.

The following Python-based visualizations were created:

- Placement distribution
- Gender-wise placement analysis

- Average academic scores by placement status
- Degree and MBA percentage comparison
- Specialization-wise placement trends





These visualizations helped identify early patterns such as higher academic scores and work experience being linked to placement success.

---

6. Machine Learning Models

Two machine learning models were developed and evaluated:

- Logistic Regression
- Random Forest Classifier

The dataset was split into training and testing sets. Models were trained using scikit-learn and evaluated using standard classification metrics.

---

7. Model Evaluation

The models were evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1-score

Results summary:

- Logistic Regression achieved approximately 86 percent accuracy
- Random Forest showed slightly lower performance

Logistic Regression performed better due to the linear and structured nature of the dataset.

---

8. Feature Importance Analysis

Feature importance analysis revealed the most influential factors affecting placement:

Top contributing features:

- Work Experience
- Gender
- SSC Percentage
- Degree Percentage
- Specialization

This confirms that industry exposure and academic performance are the strongest predictors of placement.

---

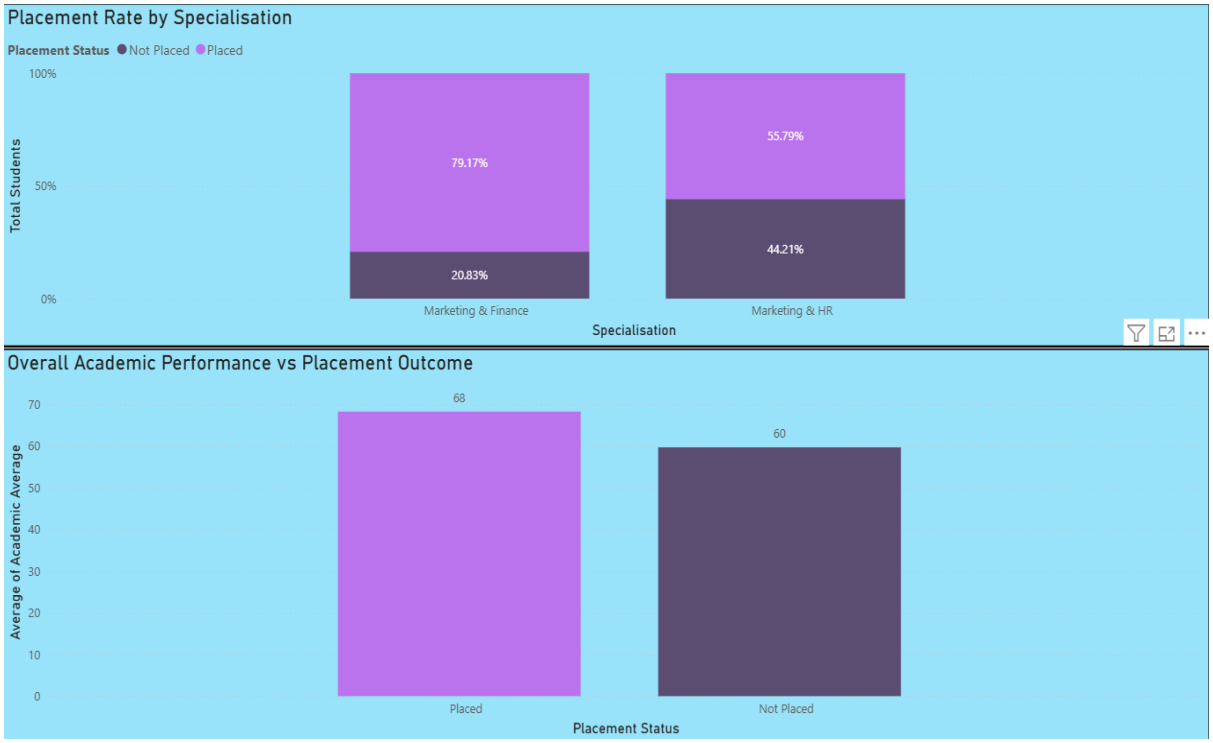
9. Power BI Dashboard

An interactive Power BI dashboard was created using the cleaned dataset to provide business-level insights.

Dashboard features include:

- Total students, placed students, and placement percentage KPIs

- Placement distribution donut chart
- Gender-wise placement analysis using 100 percent stacked bar chart
- Average SSC, Degree, and MBA percentage by placement status
- Placement rate by specialization
- Work experience impact on placement



The dashboard allows users to filter by gender, specialization, and work experience for dynamic analysis.

10.Key Insights and Findings

- Approximately 69 percent of students were placed
- Students with higher academic scores had better placement outcomes
- Work experience significantly increased placement probability
- Marketing and Finance specialization showed higher placement rates
- Gender had comparatively lower impact than academic and experiential factors
- Logistic Regression proved to be the most effective prediction model

---

11.Applications of the System

- Early identification of students at risk of not getting placed
- Better placement training and mentoring programs
- Data-driven decision-making for academic institutions

- Career guidance based on student profile analysis

12.Conclusion

The Student Placement Prediction System successfully demonstrates how data science, machine learning, and business intelligence tools can be combined to analyze and predict student placement outcomes. The integration of Python-based analysis with Power BI dashboards provides both technical accuracy and business interpretability. The project can be further extended by incorporating additional datasets, real-time data, or advanced models.

13.Future Scope

- Integration with real-time student academic systems
- Use of advanced ensemble or deep learning models
- Inclusion of soft skills and internship duration data
- Deployment as a web-based placement analytics system

14.Tools and Technologies Used

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- Scikit-learn
- Power BI
- Jupyter Notebook