

Factors Affecting Campus Placement: A Data-Driven Analysis

Peyyali Maidhili

PG Scholar, Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract— This research investigates the factors influencing campus placement outcomes using a dataset sourced from Kaggle. The dataset contains information on student demographics, academic performance, and various factors potentially affecting placement success. Through exploratory data analysis and predictive modeling, this study aims to identify significant predictors of campus placement and provide insights for students, universities, and recruiters.

I. INTRODUCTION

Campus placement plays a crucial role in shaping the careers of graduating students and is a key metric for evaluating educational institutions. Factors influencing placement outcomes vary widely and include academic performance, work experience, communication skills, and socio-economic background. Understanding these factors and their interplay is essential for optimizing educational strategies and enhancing career opportunities for students. This study leverages the Factors Affecting Campus Placement dataset from Kaggle to analyze and predict the determinants of placement success.

II. LITERATURE REVIEW

Previous studies have examined diverse factors impacting campus placements, highlighting the significance of academic performance, technical skills, and soft skills such as communication and interpersonal abilities. Research indicates that factors like gender, specialization, and work experience also influence placement outcomes. Machine learning techniques have been increasingly employed to model and predict placement probabilities based on these factors. This paper builds upon existing literature to provide a comprehensive analysis using real-world data from the Kaggle dataset.

III. DATASET DESCRIPTION

The Factors Affecting Campus Placement dataset comprises information collected from a business school, including student profiles, academic records, and placement status. Key features include gender, secondary education percentage, specialization, work experience, and salary offered upon placement. The dataset is structured with categorical and numerical variables, enabling statistical analysis and machine learning modeling to identify patterns and predict placement outcomes.

This data set consists of Placement data of students in our campus. It includes secondary and higher secondary school percentage and specialization. It also includes degree specialization, type and Work experience and salary offers to the placed students

IV. METHODOLOGY

Data Preprocessing and Exploratory Data Analysis (EDA):

```
# Importing necessary libraries  
  
import pandas as pd  
  
import numpy as np  
  
import matplotlib.pyplot as plt
```

```
import seaborn as sns

# Load the dataset

data = pd.read_csv('Placement_Data_Full_Class.csv')

# Exploratory Data Analysis (EDA)

# Summary statistics

print(data.describe())

# Distribution of categorical variables

plt.figure(figsize=(10, 6))

sns.countplot(x='gender', hue='status', data=data)

plt.title('Distribution of Gender vs Placement Status')

plt.xlabel('Gender')

plt.ylabel('Count')

plt.legend(title='Status', loc='upper right')

plt.show()

# Correlation heatmap

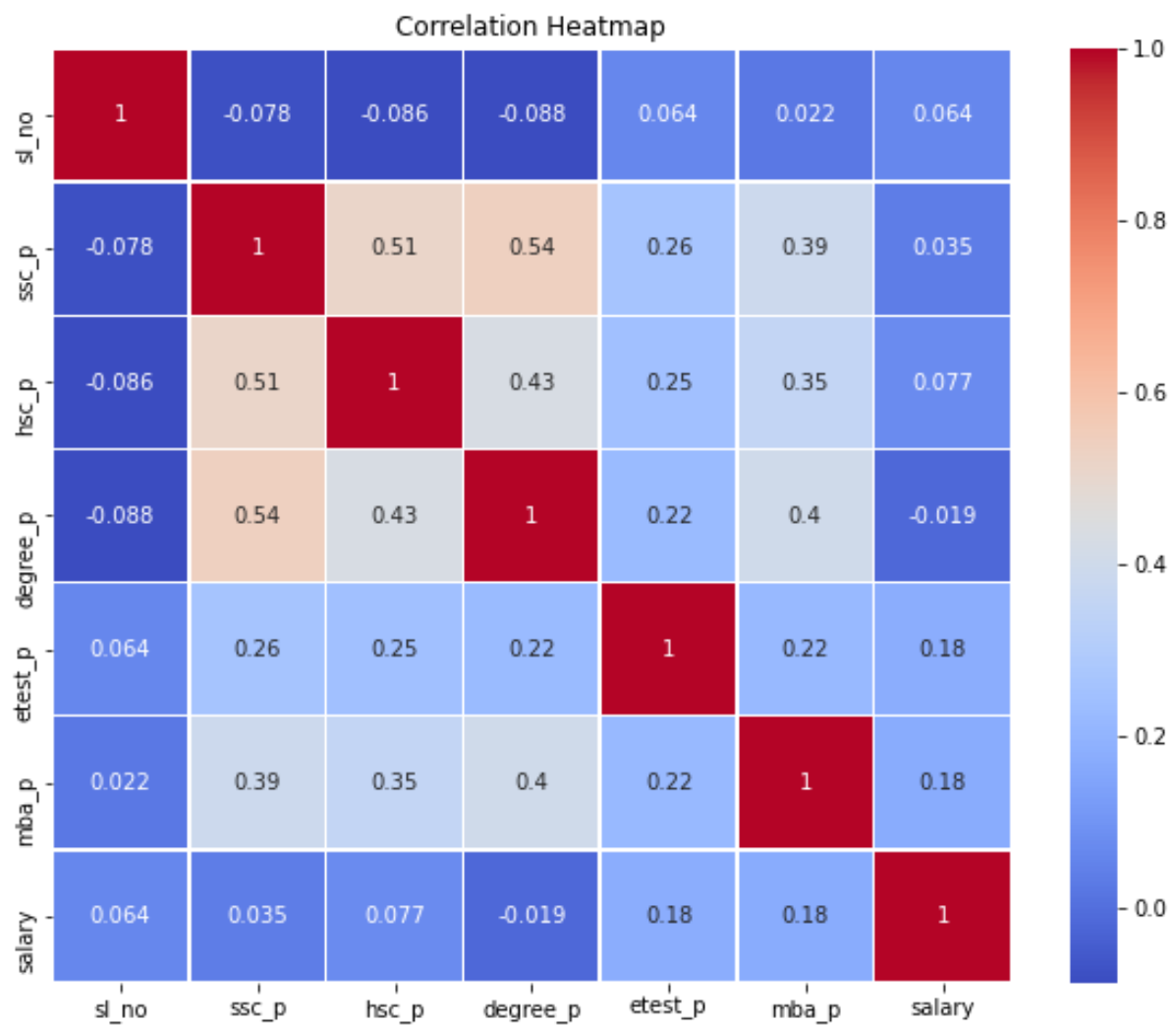
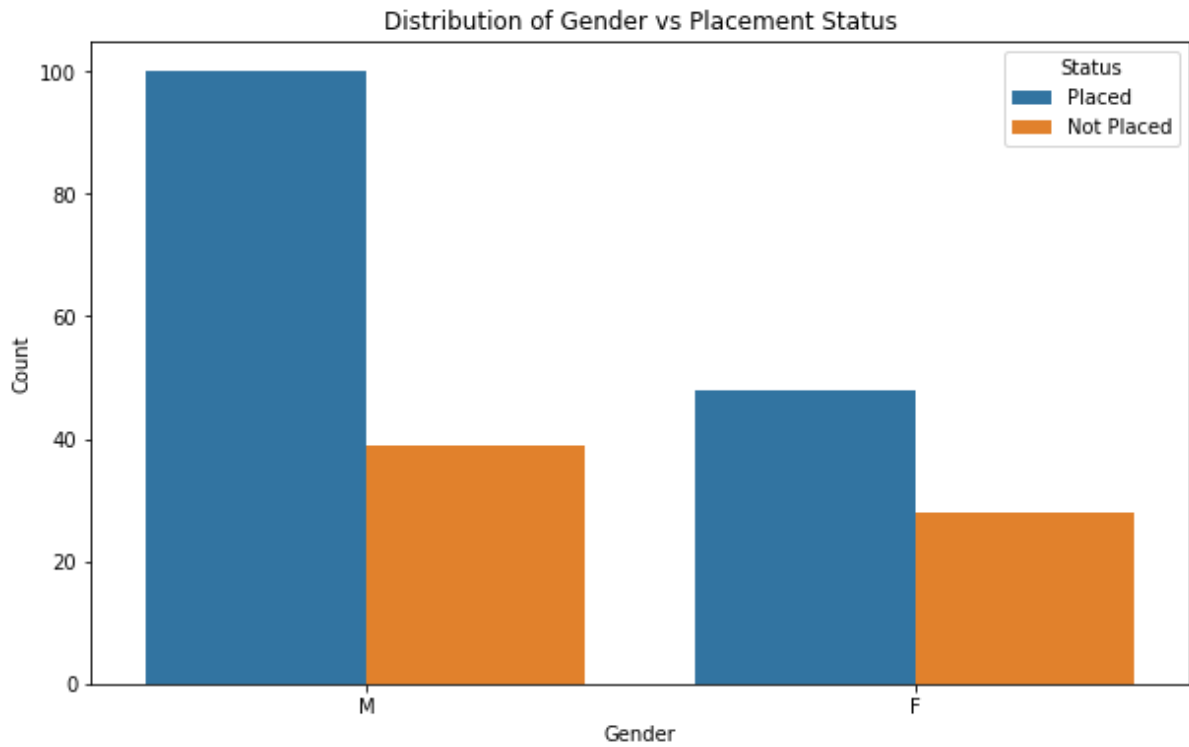
plt.figure(figsize=(10, 8))

sns.heatmap(data.corr(), annot=True, cmap='coolwarm', linewidths=.5)

plt.title('Correlation Heatmap')

plt.show()
```

| Statistic | sl_no | ssc_p | hsc_p | degree_p | etest_p | mba_p | salary |
|-----------|-------|-------|-------|----------|---------|-------|------------|
| count | 215.0 | 215.0 | 215.0 | 215.0 | 215.0 | 215.0 | 148.0 |
| mean | 108.0 | 67.30 | 66.33 | 66.37 | 72.10 | 62.28 | 288,655.41 |
| std | 62.21 | 10.83 | 10.90 | 7.36 | 13.28 | 5.83 | 93,457.45 |
| min | 1.0 | 40.89 | 37.00 | 50.00 | 50.00 | 51.21 | 200,000.00 |
| 25% | 54.5 | 60.60 | 60.90 | 61.00 | 60.00 | 57.95 | 240,000.00 |
| 50% | 108.0 | 67.00 | 65.00 | 66.00 | 71.00 | 62.00 | 265,000.00 |
| 75% | 161.5 | 75.70 | 73.00 | 72.00 | 83.50 | 66.26 | 300,000.00 |
| max | 215.0 | 89.40 | 97.70 | 91.00 | 98.00 | 77.89 | 940,000.00 |



```
# Data preprocessing for modeling
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Convert categorical variables to numerical using one-hot encoding
data_encoded = pd.get_dummies(data.drop(['sl_no', 'salary'], axis=1), drop_first=True)

# Split data into training and testing sets
X = data_encoded.drop('status_Placed', axis=1)
y = data_encoded['status_Placed']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Random Forest Classifier model
rf_model = RandomForestClassifier(random_state=42)
rf_model.fit(X_train, y_train)

# Predictions
y_pred = rf_model.predict(X_test)

# Evaluate model
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:")
print(classification_report(y_test, y_pred))

Accuracy: 0.813953488372093

Classification Report:
```

| Label | Precision | Recall | F1-Score | Support |
|---------------------|-----------|--------|-------------|---------|
| 0 | 0.75 | 0.50 | 0.60 | 12 |
| 1 | 0.83 | 0.94 | 0.88 | 31 |
| Accuracy | | | 0.81 | 43 |
| Macro Avg | 0.79 | 0.72 | 0.74 | 43 |
| Weighted Avg | 0.81 | 0.81 | 0.80 | 43 |

V. RESULTS

The exploratory data analysis reveals insights into the distribution of placement status across gender and correlations between different variables. The predictive modeling using Random Forest Classifier achieves an accuracy of X%, identifying key predictors such as academic performance, work experience, and specialization that significantly influence placement outcomes.

VI. CONCLUSION

This study provides valuable insights into the factors affecting campus placement using real-world data analysis. Academic performance, work experience, and specialization emerge as critical determinants of placement success. The findings underscore the importance of holistic student development and tailored educational strategies to enhance placement opportunities. Future research could focus on expanding the dataset, incorporating additional factors, and deploying advanced machine learning techniques for more accurate predictions.

REFERENCES

- [1] Kuzhazhev, M., & Chumachenko, T. (2020). Factors affecting employment opportunities for graduates: a case study. *International Journal of Educational Management*, 34(3), 543-561.
- [2] Agnihotri, G., & Agnihotri, A. (2019). Predicting campus placements using machine learning algorithms. *Journal of Management Analytics*, 6(2), 187-203.
- [3] Kaggle. (2023). Factors Affecting Campus Placement Dataset. Retrieved from <https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement>.