PROJECT REPORT

Project Title:
Placement Analysis using ML

INTRODUCTION:

Placement analysis is an important task for educational institutions and employers to understand the quality of students who are being placed. Machine learning can be used to analyze various factors such as grades, skills, and achievements to predict the probability of a student being placed. In this report, we will discuss the project of placement analysis using ML.

PROJECT OBJECTIVES:

The objective of this project is to predict the placement status of a student based on various factors such as academic performance, skills, and other achievements. The goal is to provide insights to educational institutions and employers to improve the placement rate and quality of students

DATASET:

The dataset used for this project contains information about students' academic performance, skills, and other achievements. The dataset includes details of 215 students having 15 features that are:

1.sl_no
2.gender
3.ssc_p
4.ssc_b
5.hsc_p
6.hsc_b
7.hsc_s

8.degree_p

9.degree_t

10.workex

11.etest_p

12.specialisation

13.mba_p

14.status

15.salary

EDA:

Before feeding the data to the ML model, we performed exploratory data analysis(EDA), which includes data cleaning, handling missing values, and encoding categorical variables and visualization. We also performed feature scaling to normalize the data.

The **Categorical cols** are: 'gender', 'ssc_b', 'hsc_b', 'hsc_s', 'degree_t', 'workex', 'specialisation', 'status
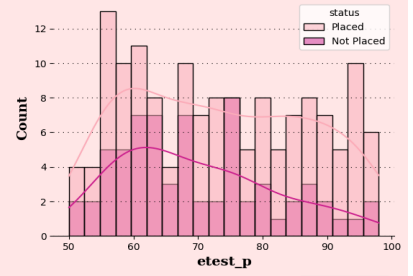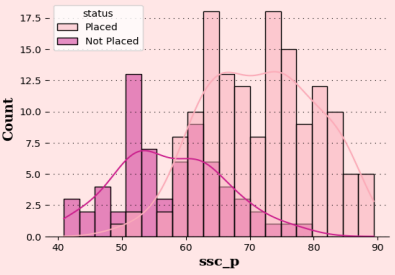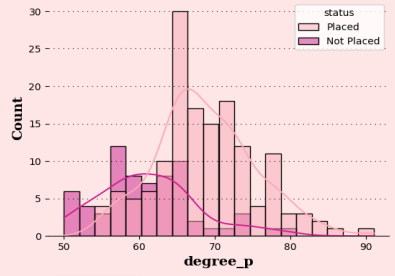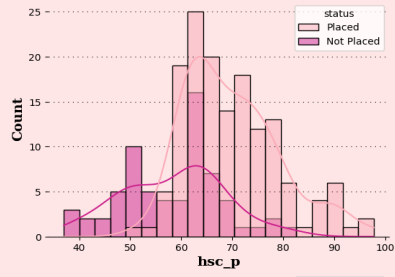
The **numerical cols** are : 'ssc_p', 'hsc_p', 'degree_p', 'etest_p', 'mba_p', 'salary'

The **target variable** is : 'status'

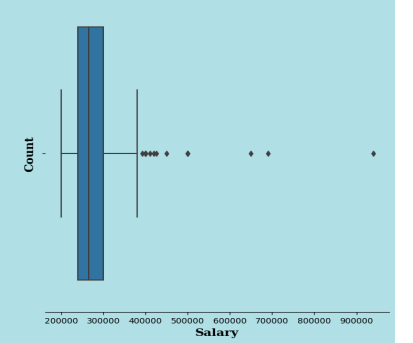Here are all the plots obtained during visualization:

Counting Categorical features and checking status among various categories

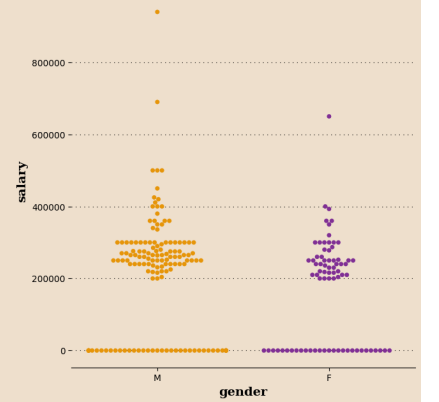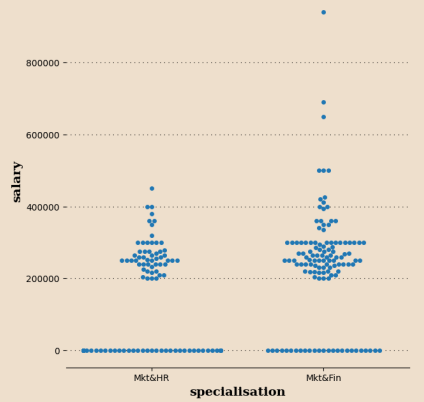**Understanding the target variable**
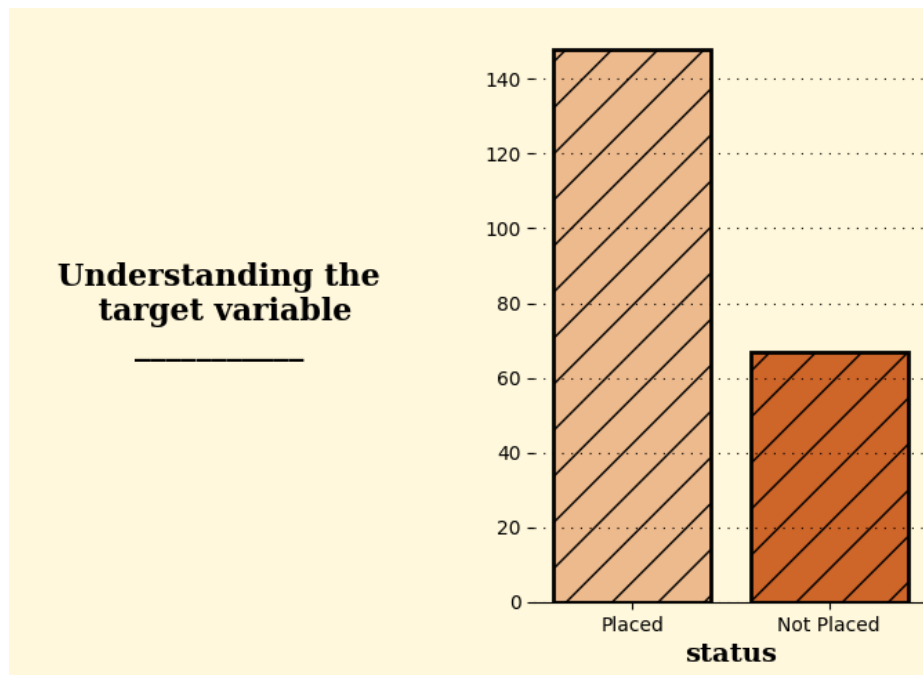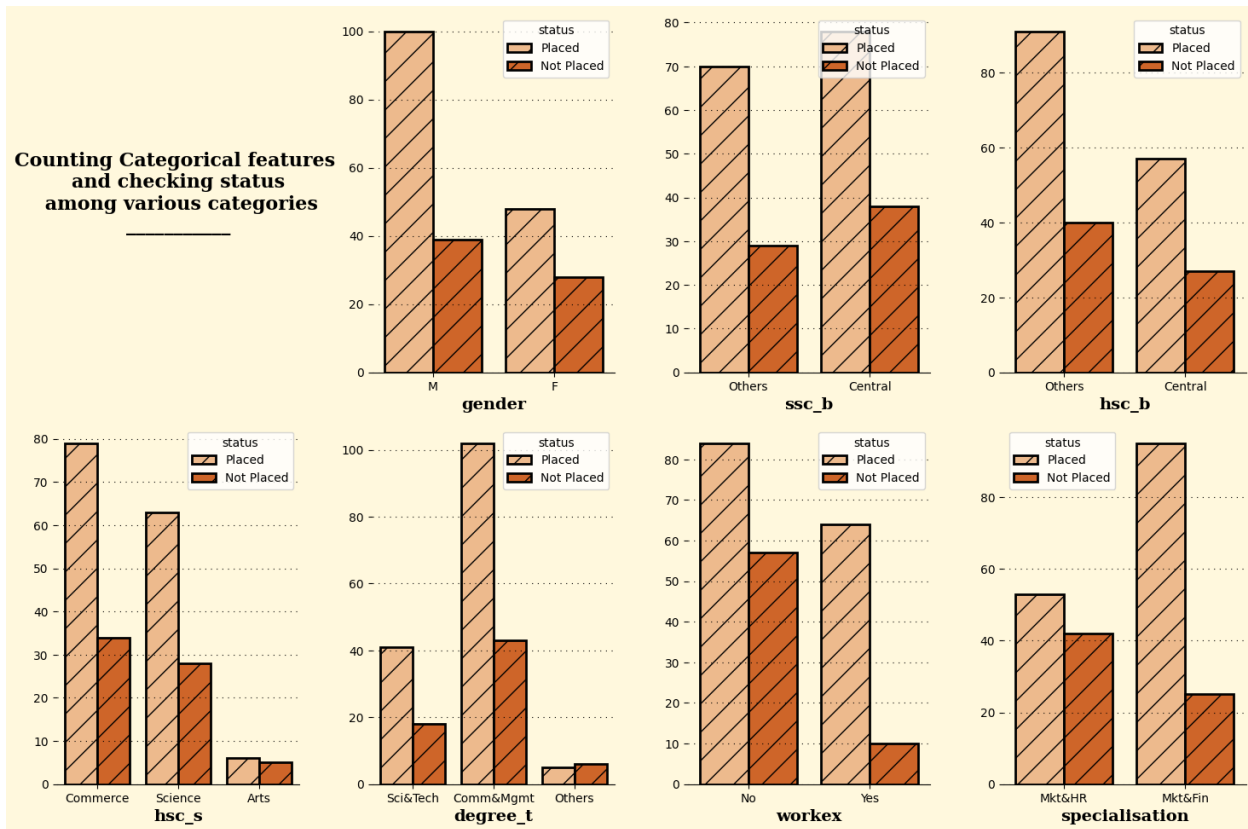
Understanding Continuous features
_____
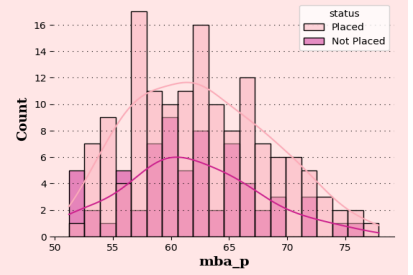
Salary analysis of Placed Students
_____
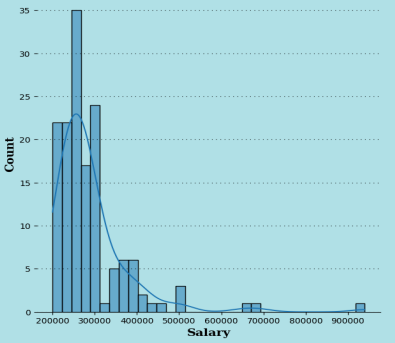
Salary analysis of Placed Students
_____

ML MODEL SELECTION:

We used various ML algorithms such as Logistic Regression, Random Forest, and Bernoulli Naive Bayes to predict the placement status of a student. We used the accuracy score as a metric to evaluate the performance of each model. After comparing the performance of each model, we selected the Random Forest model as it provided the highest accuracy score.

MODEL TRAINING AND TESTING:

We split the dataset into training and testing sets with a 80:20 ratio. We trained the Logistic Regression, Random Forest, and Bernoulli Naive Bayes model on the training set and evaluated its performance on the testing set. We also performed cross-validation to ensure that the model is not overfitting the training data.

RESULTS:

The Logistic Regression, Random Forest achieved an accuracy score of 84%, 77% respectively and Bernoulli Naive Bayes model achieved an accuracy score 86% on the testing set by cross-validation. This indicates that the model performed well to predict the placement status of a student based on various factors such as academic performance, skills, and other achievements.

CONCLUSION:

Placement analysis can help educational institutions and employers to improve the quality of students who are being placed. This project can be extended by incorporating more features such as personality traits and extracurricular activities to improve the accuracy of the model.