



STUDENT PLACEMENT PREDICTOR



Group Members

1. Sumit Kumar (58)
 2. Ritik Kumar (46)
- 

Detailed Overview



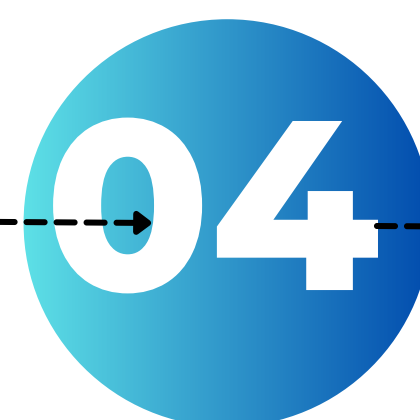
**PROBLEM
STATEMENT**



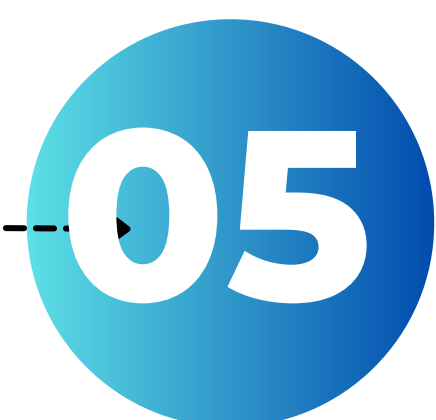
**Data
Preprocessing**



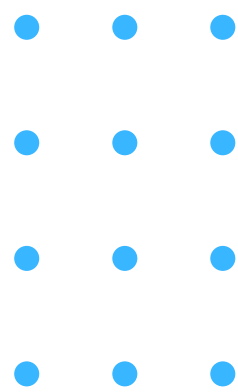
**DATA
VISUALIZATION**



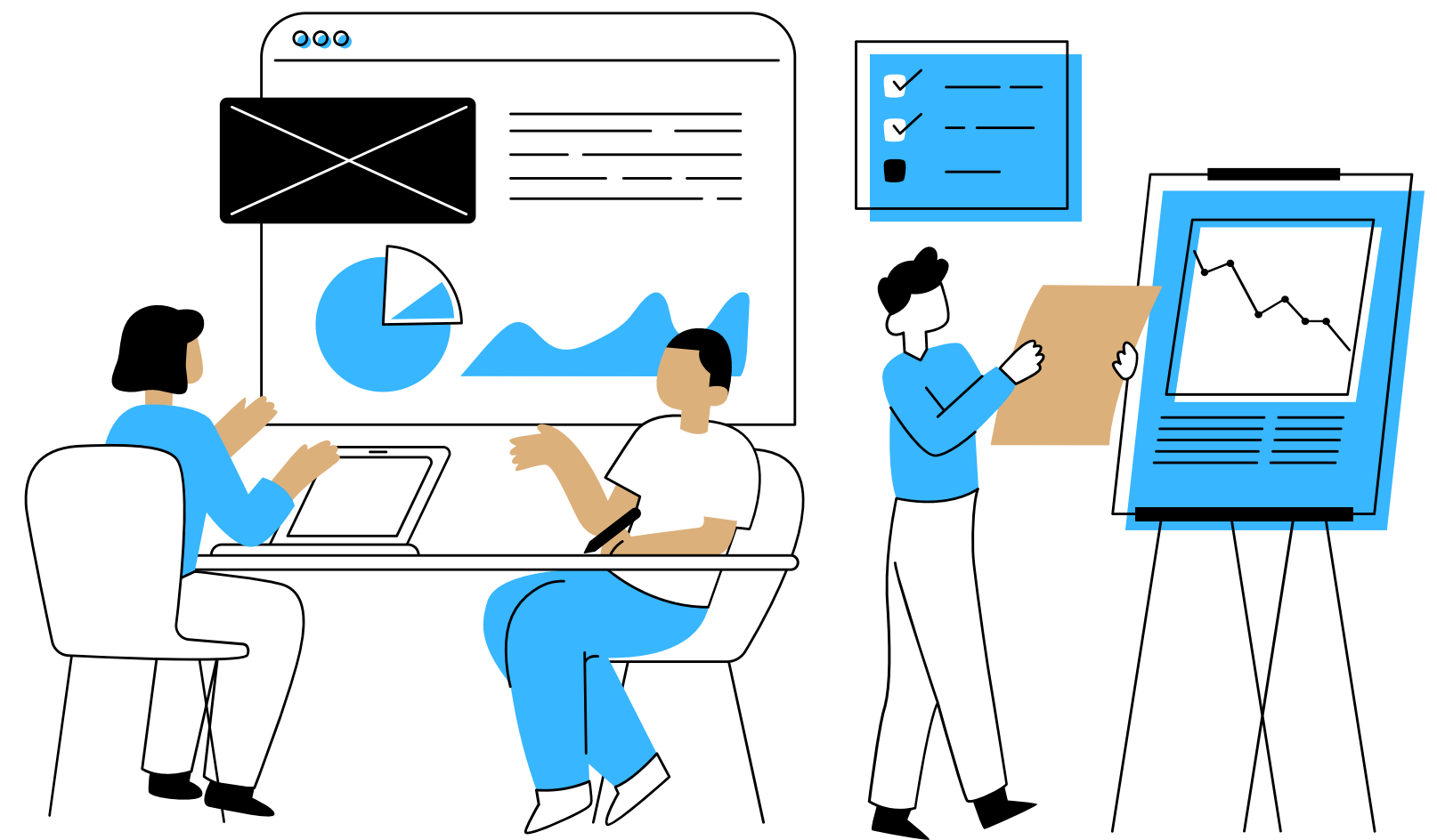
MODELS



**CONCLUSION &
FUTURE WORK**



PROBLEM STATEMENT



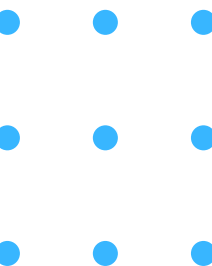
The primary challenge for learning institutions today is enhancing student placement performance. To address this, educational institutes seek more efficient technology, with a focus on improving the quality of knowledge related to educational processes and managerial systems. Machine learning techniques are proposed to extract insights from operational and historical data within educational databases, offering a promising approach to tackling this challenge.

DATASET INFORMATION

The dataset used for training as well as testing was obtained from the Kaggle. It contains 10000 rows and 11 features.

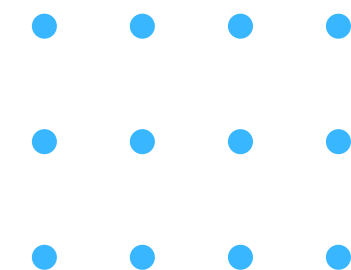
Features

- Internship
- Projects
- Workshops/Certificates
- Soft Skills assessments
- Academic performance in Secondary School Certificate (SSC)
- ExtracurricularActivities
- PlacementTraining
- Aptitude Scores
- Higher Secondary Certificate Marks.
- CGPA



DATA PRE- PROCESSING

Data preprocessing is a crucial step, encompassing the cleaning and transformation of raw data to render it suitable for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task.



**Dropped “StudentId” column because it not relevant for our data analysis.
StudentId does affect our model or is not significant for analysis**

All the data in the dataset is in int or float except for 3 columns/variables that are of object data type:

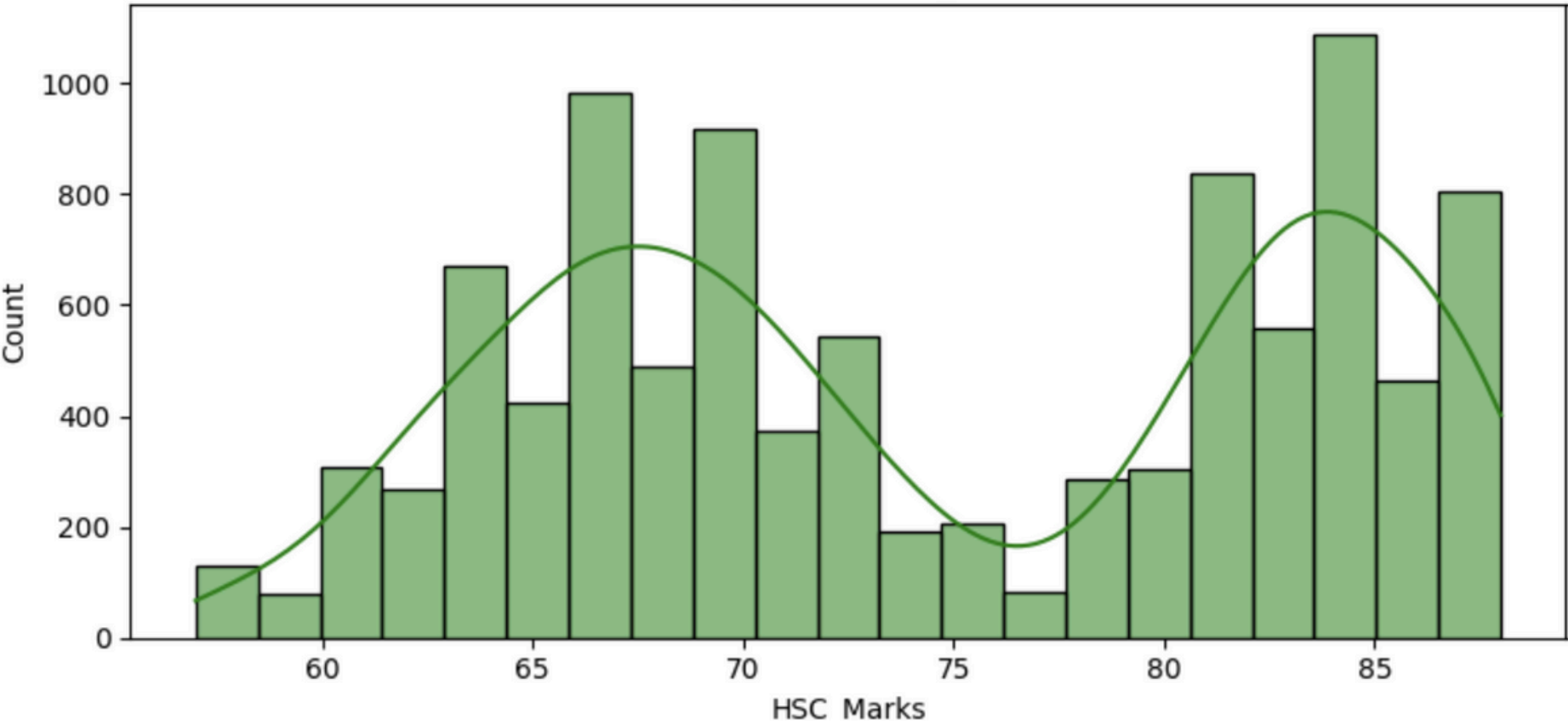
- **ExtracurricularActivities**
- **PlacementTraining**
- **Placement Status**

In sci-kit-learn, the LabelEncoder is a utility class used to encode categorical labels into numerical labels. It essentially converts categorical data (text labels) into numerical labels so that machine learning algorithms can handle them more effectively.

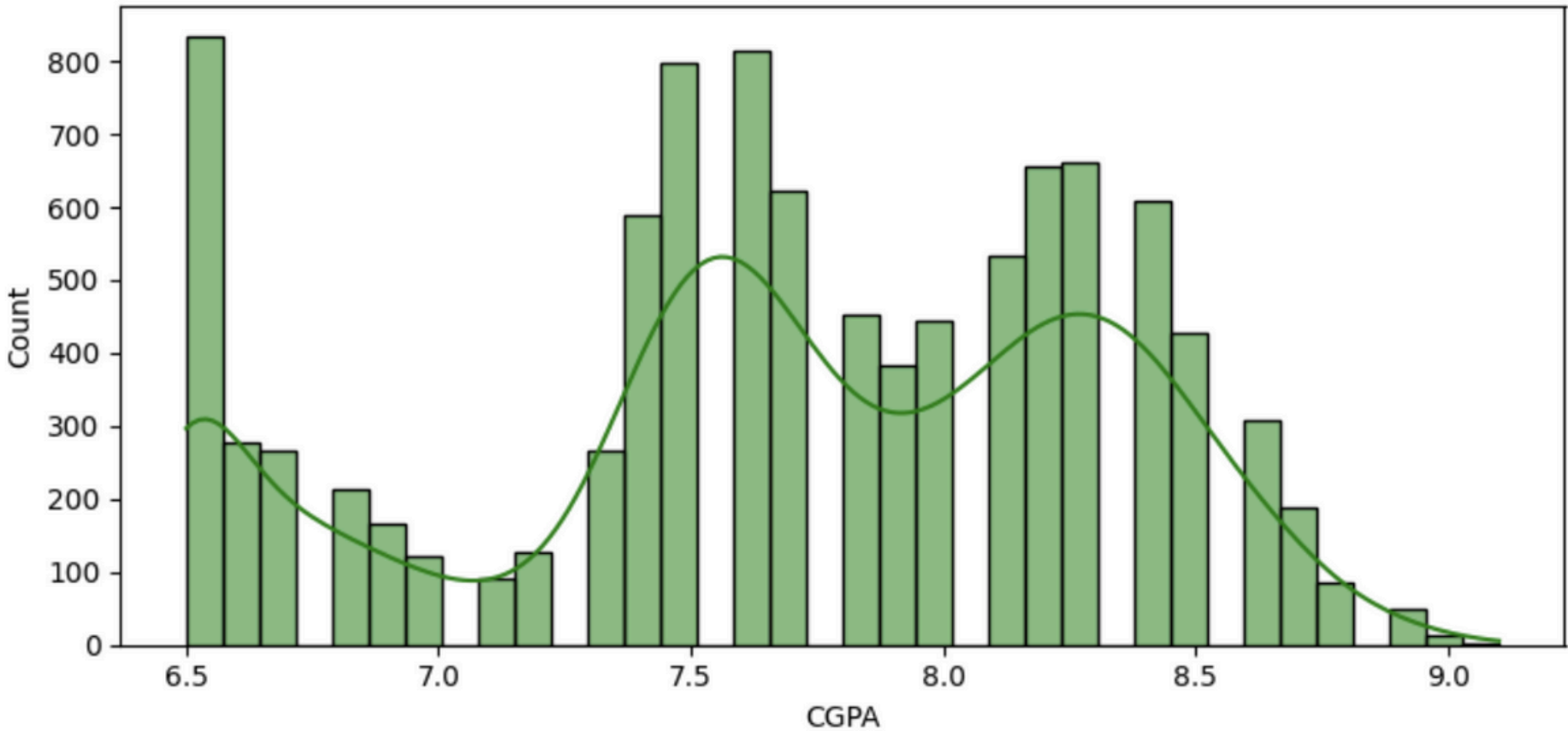
For instance, if you have a categorical feature like "Yes" and "No". The LabelEncoder would assign them numerical labels, such as 0 and 1, respectively.

DATA VISUALIZATION

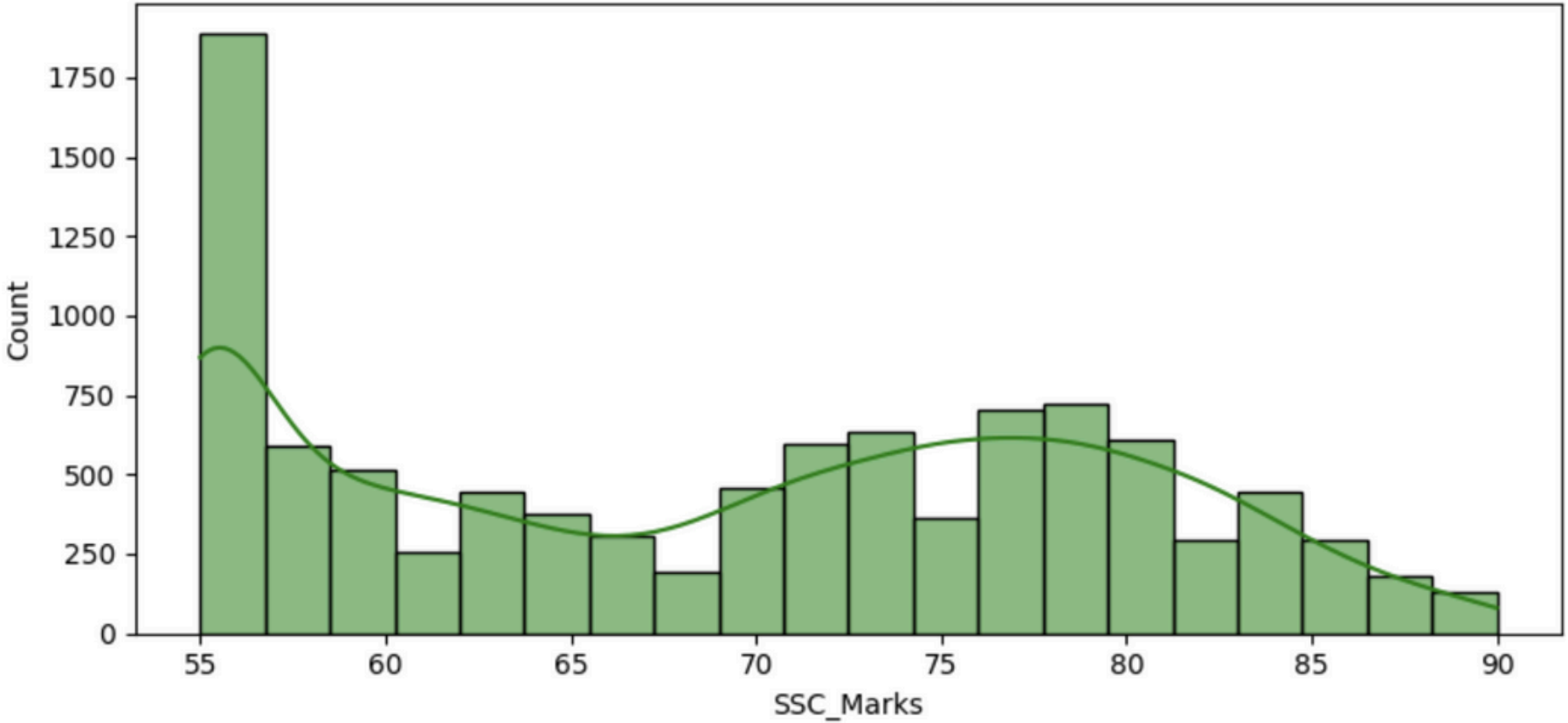
Histogram of HSC_Marks



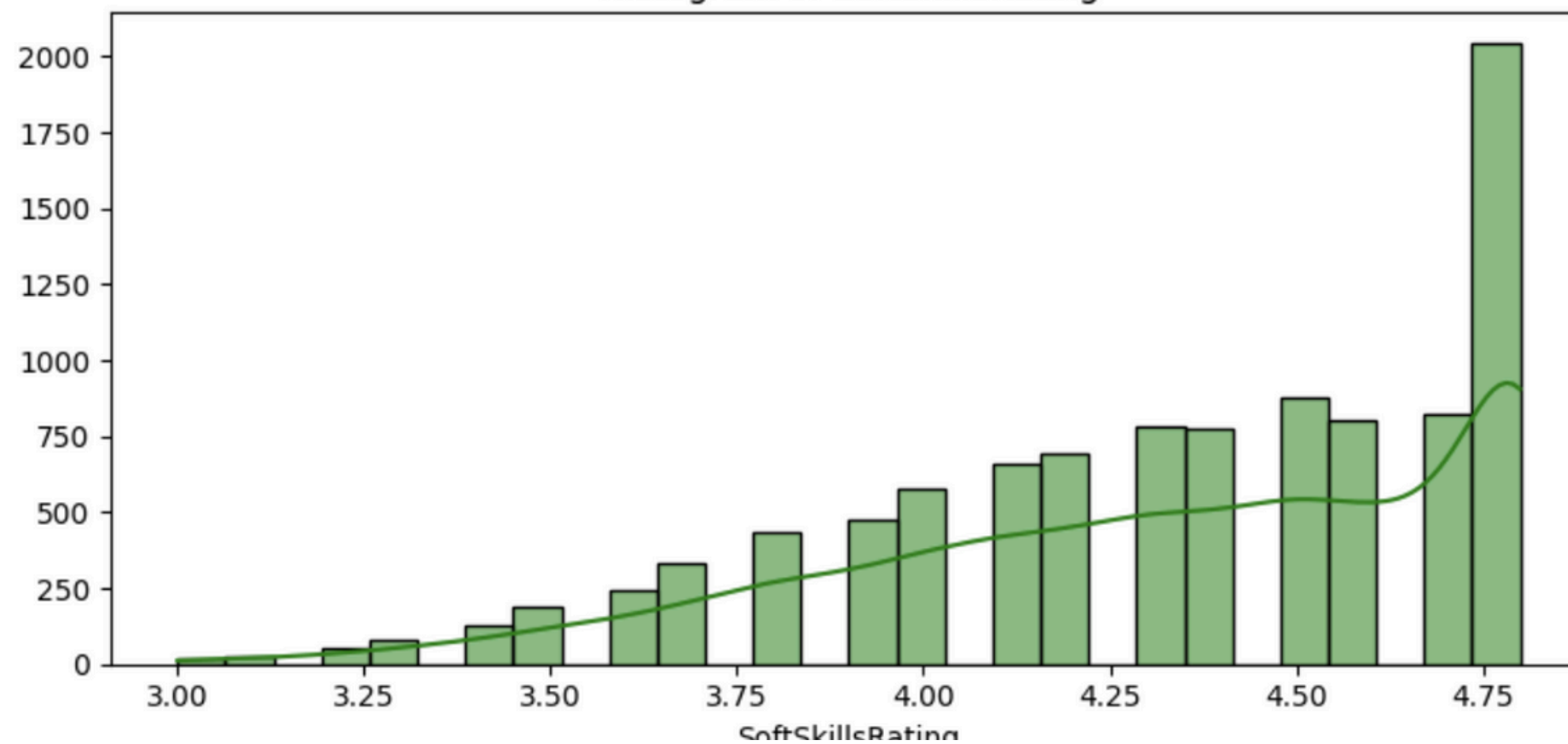
Histogram of CGPA



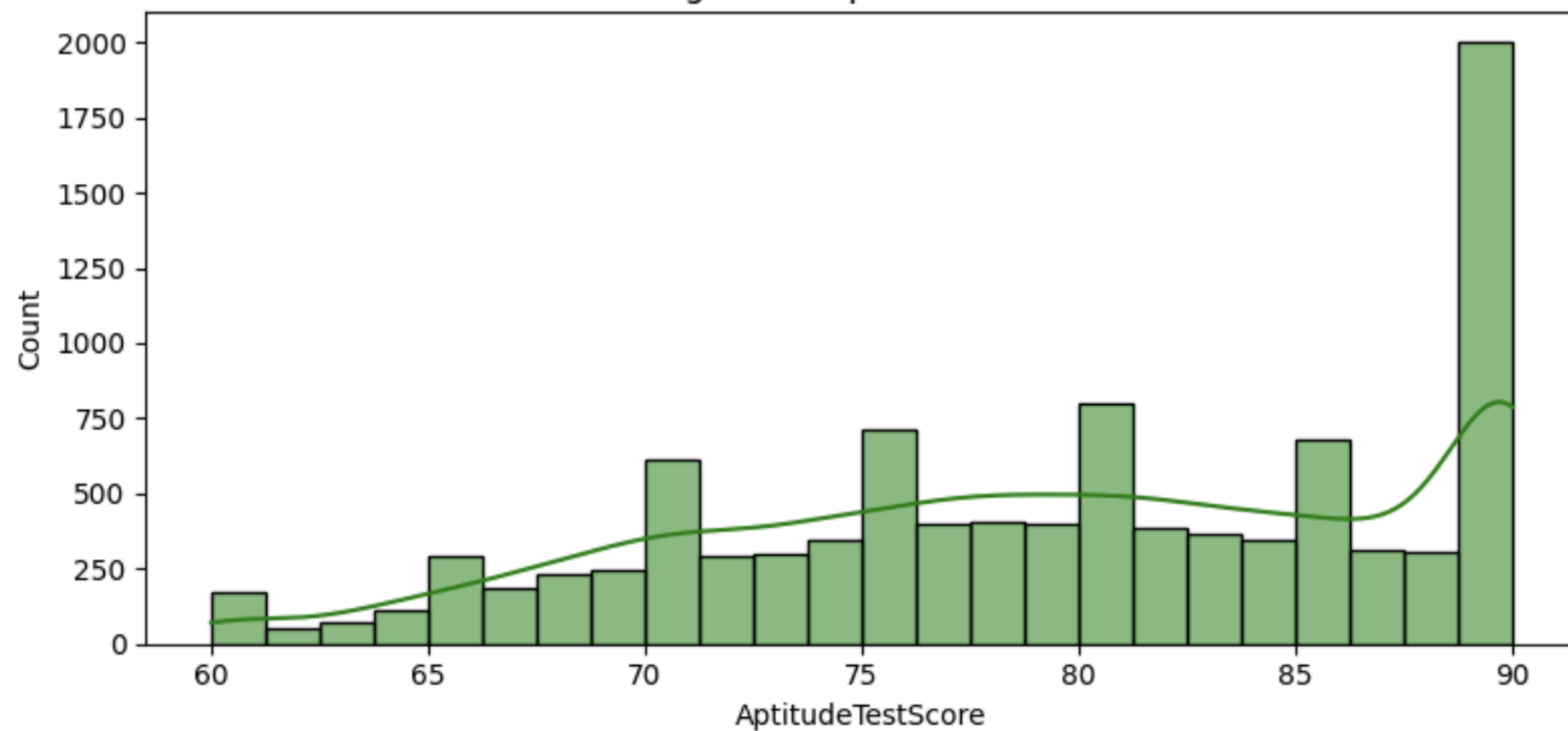
Histogram of SSC_Marks



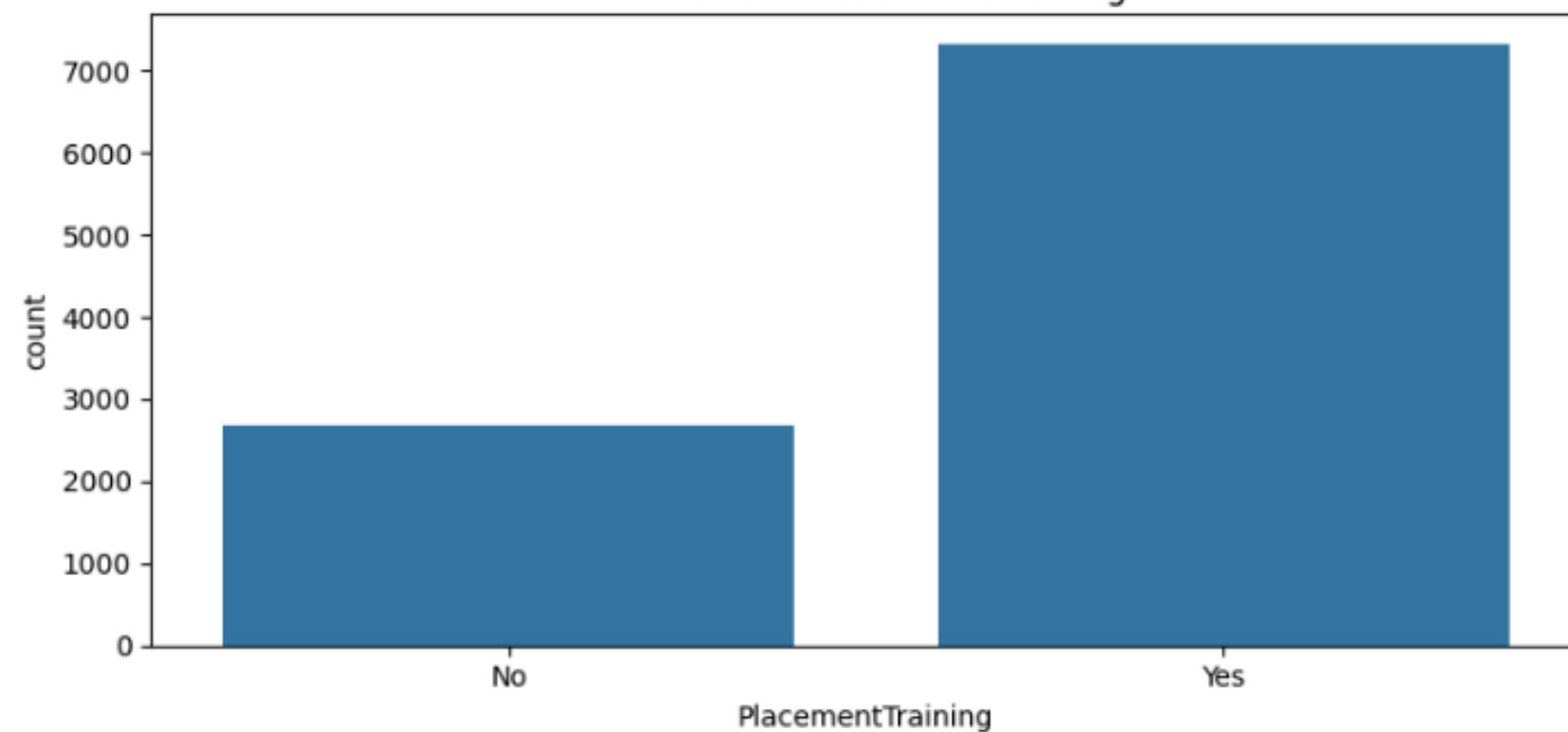
Histogram of SoftSkillsRating



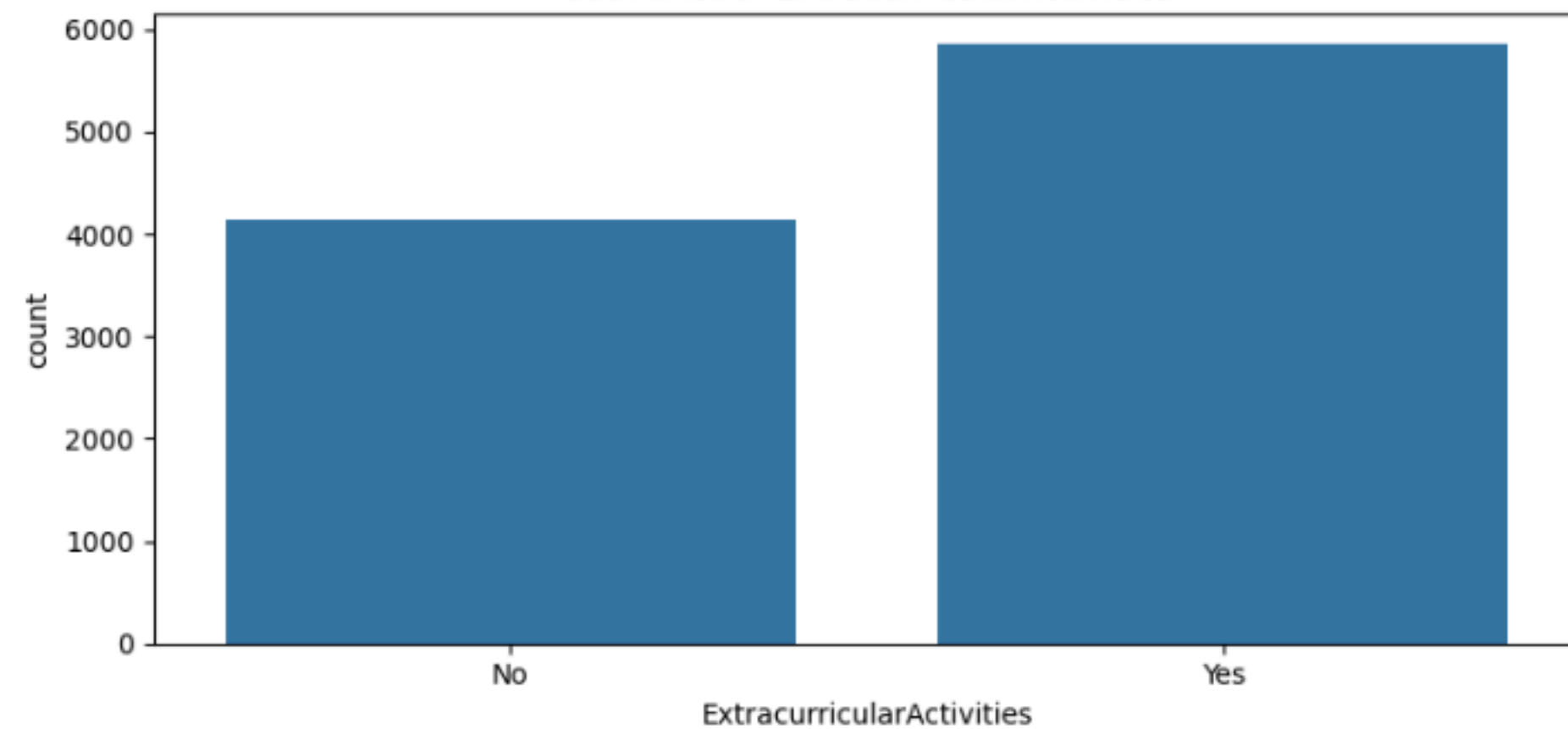
Histogram of AptitudeTestScore



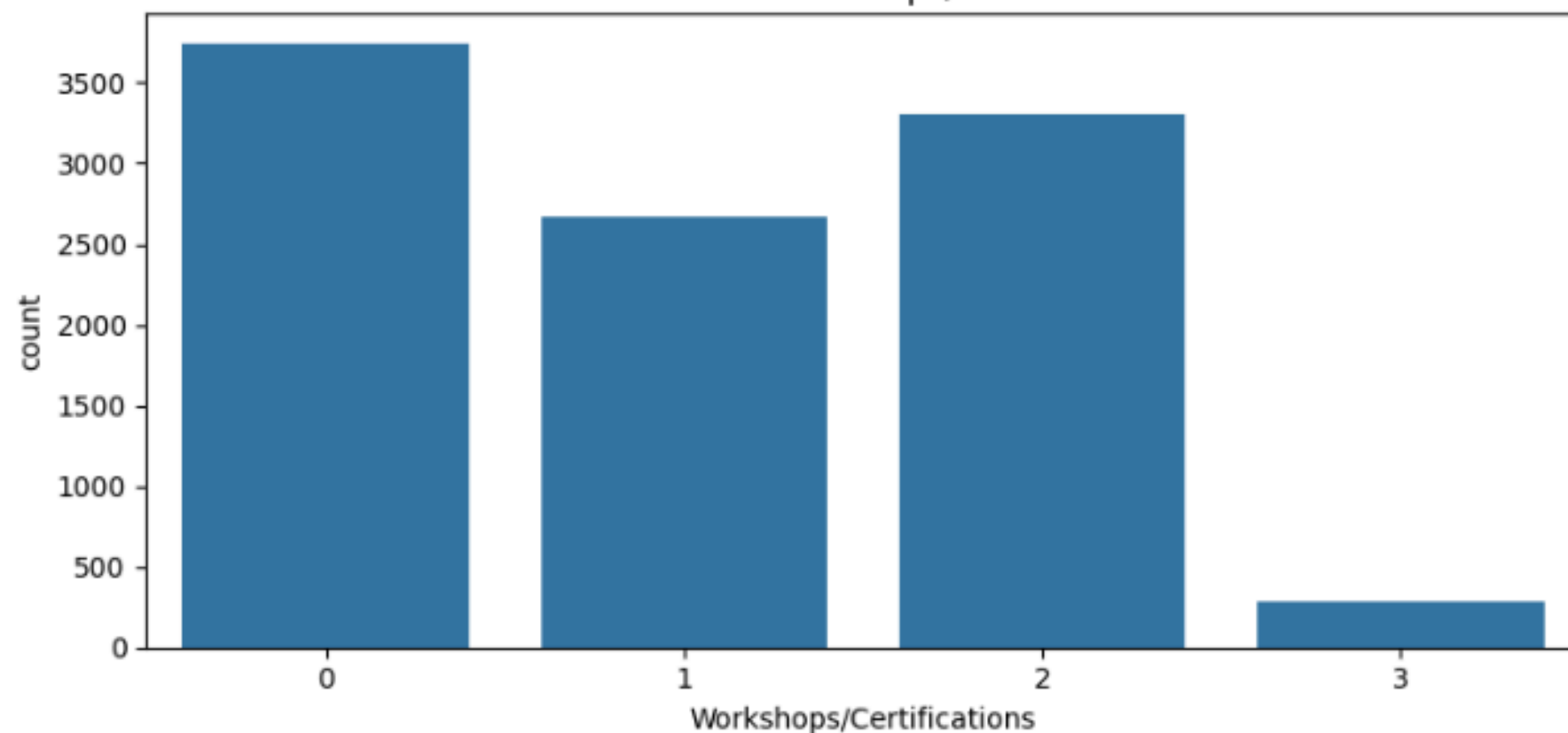
CountPlot of PlacementTraining



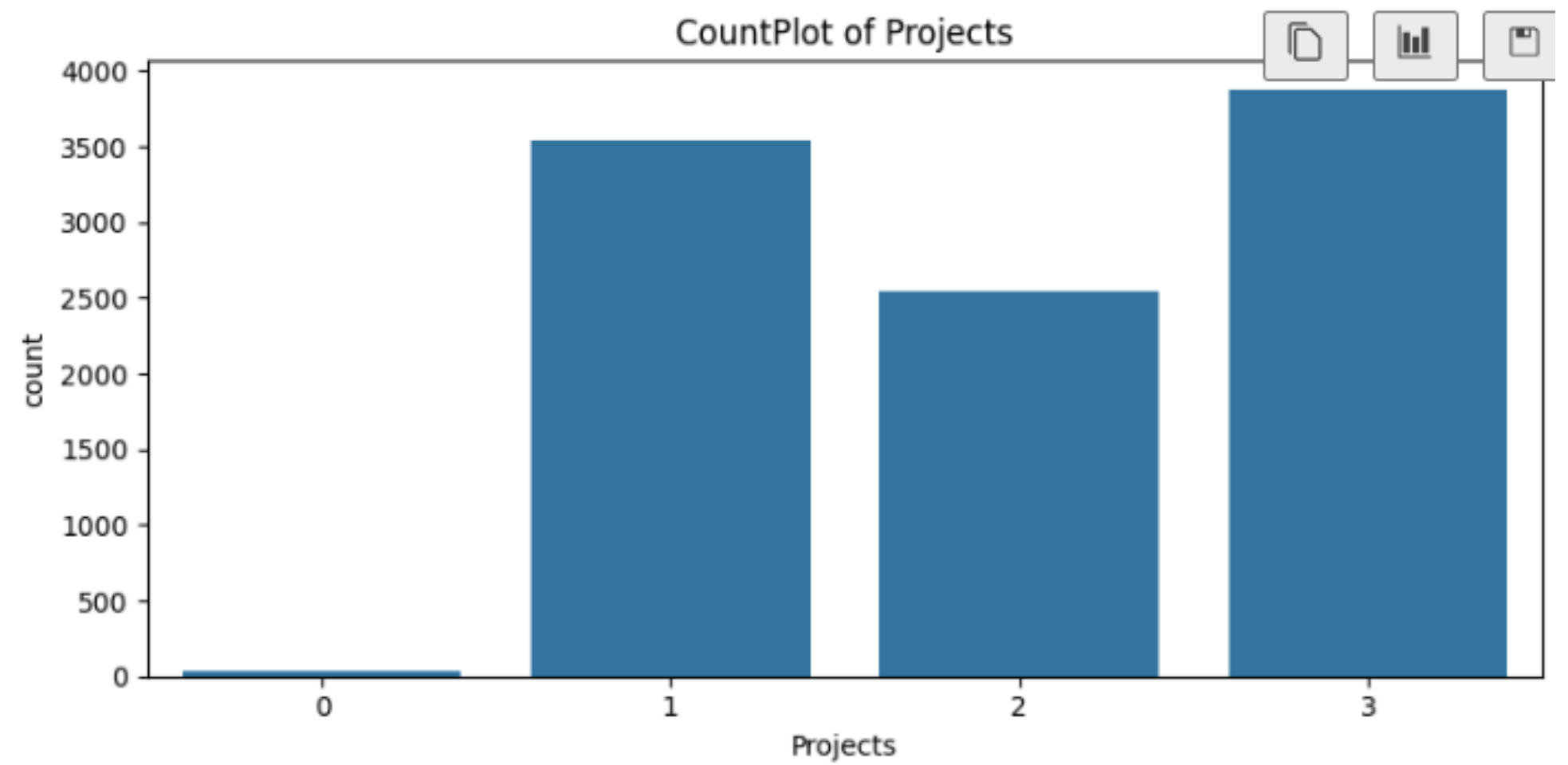
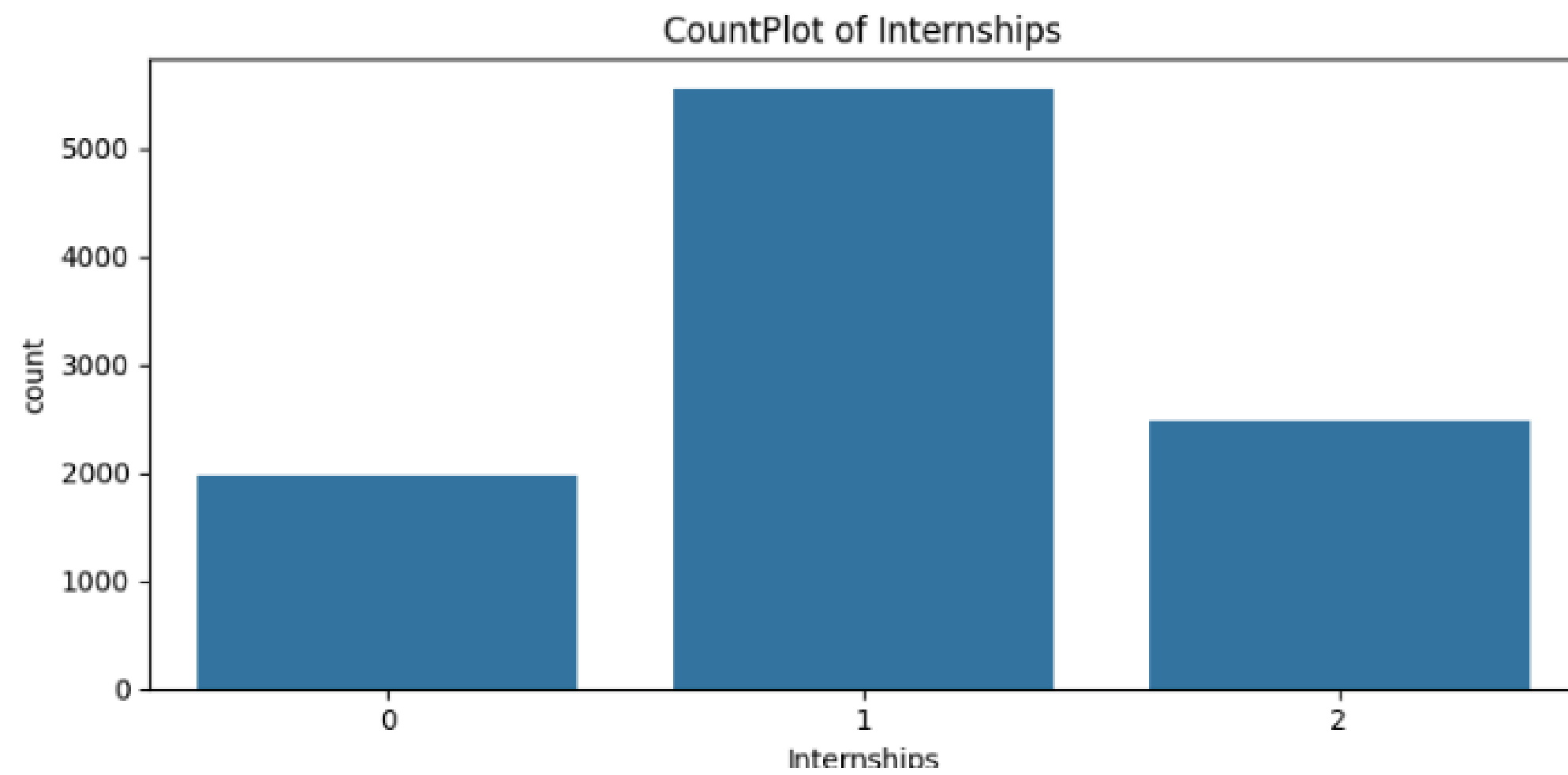
CountPlot of ExtracurricularActivities



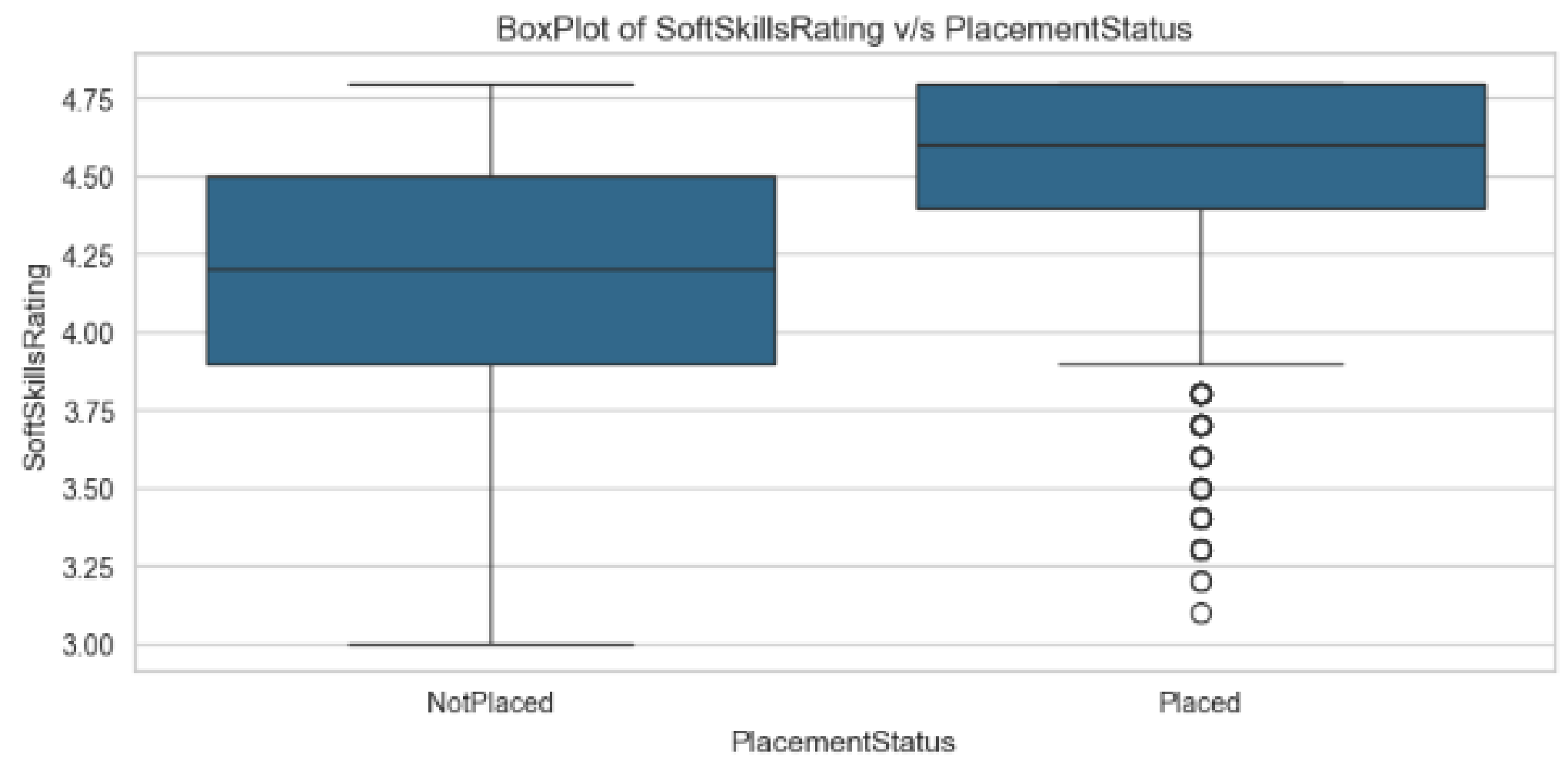
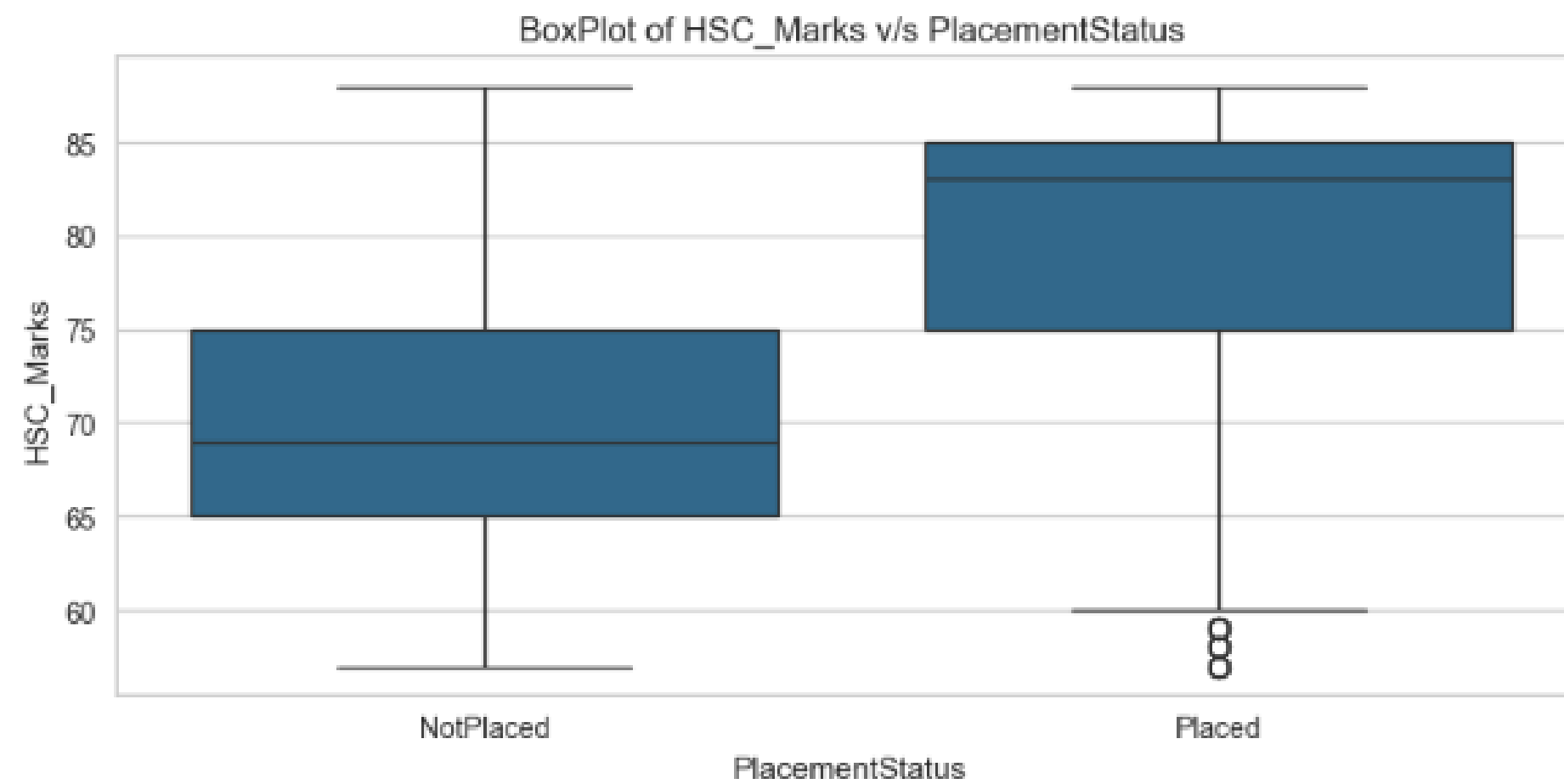
CountPlot of Workshops/Certifications

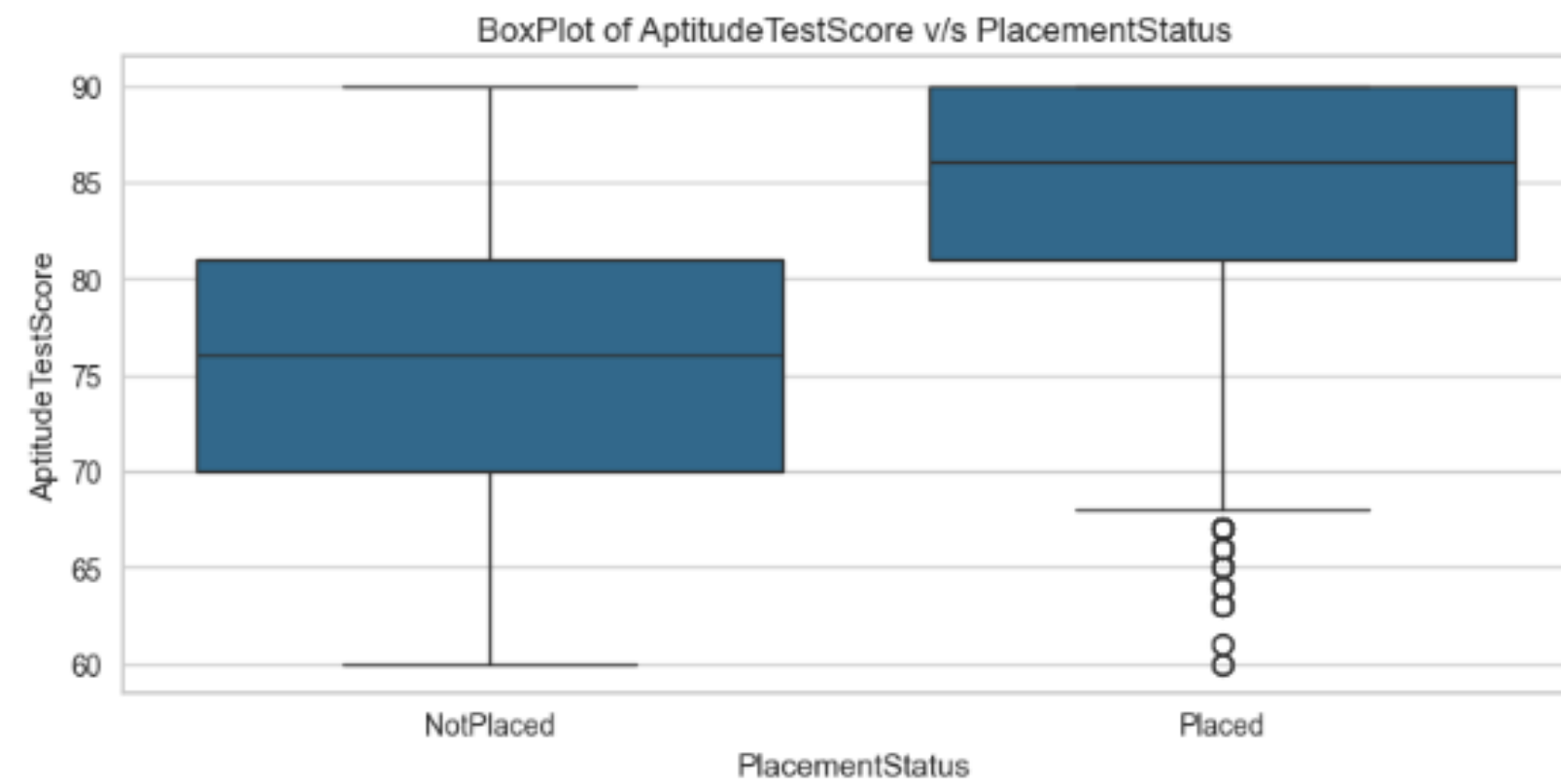
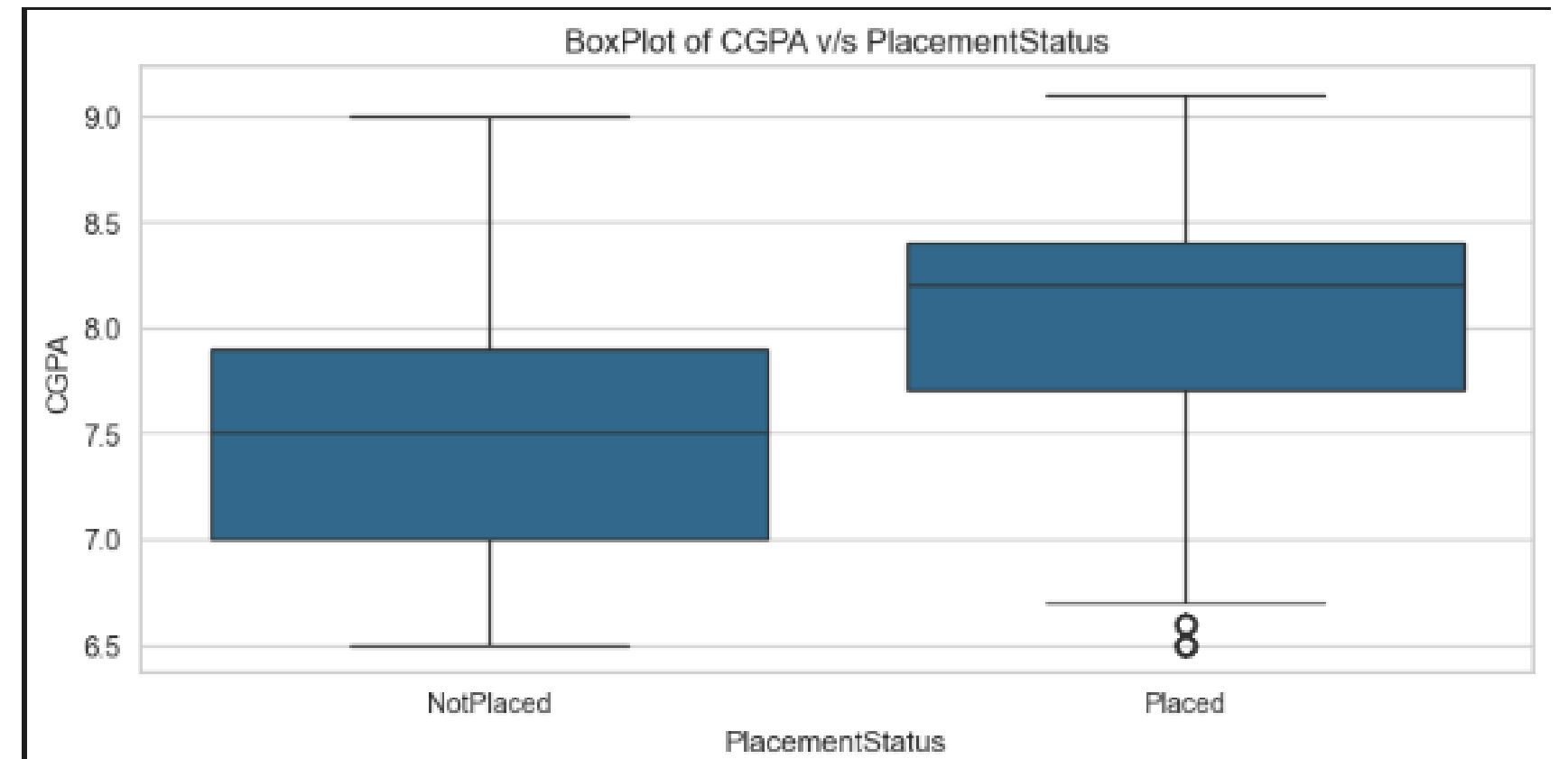
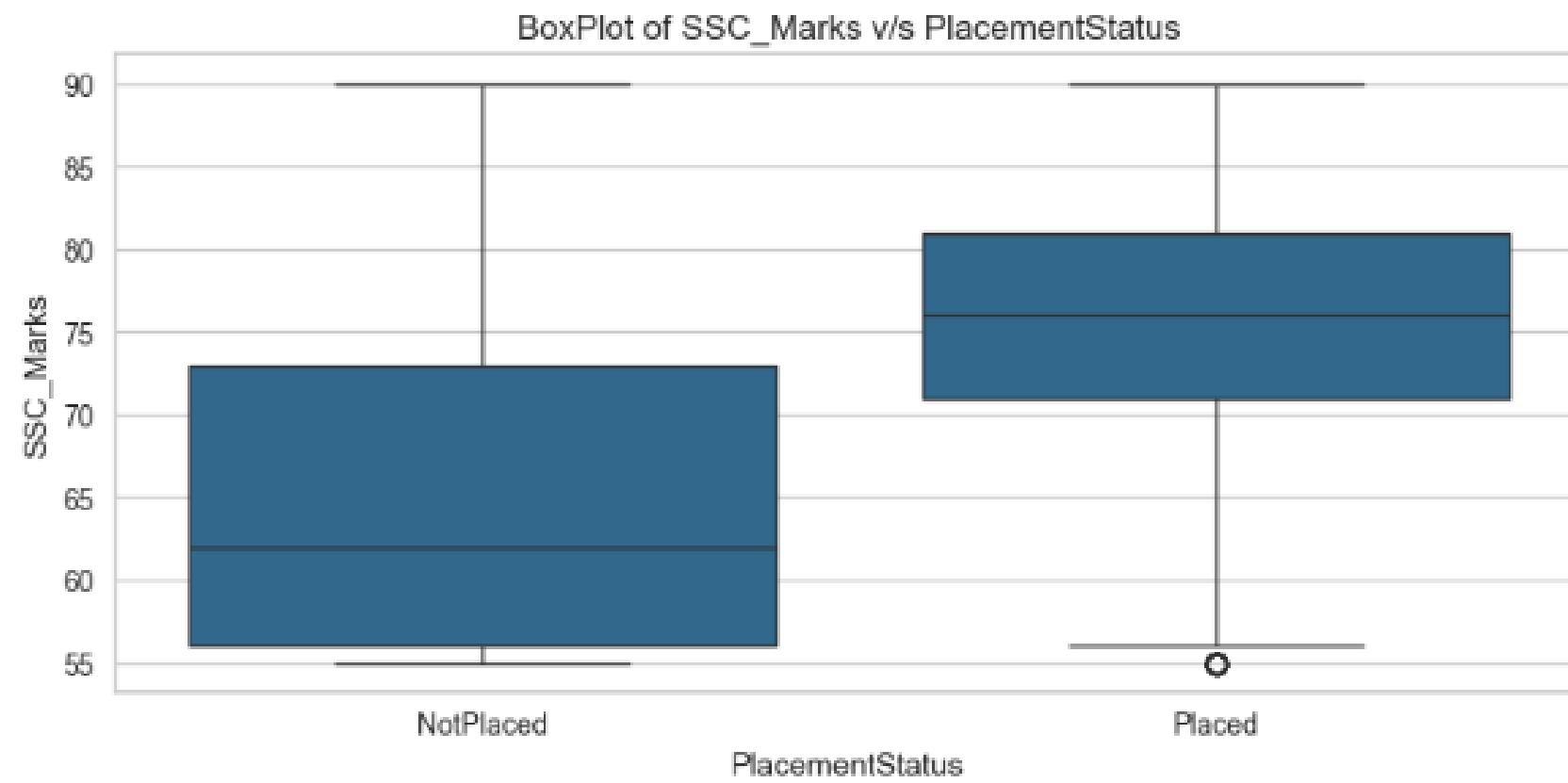


CountPlot of PlacementTraining



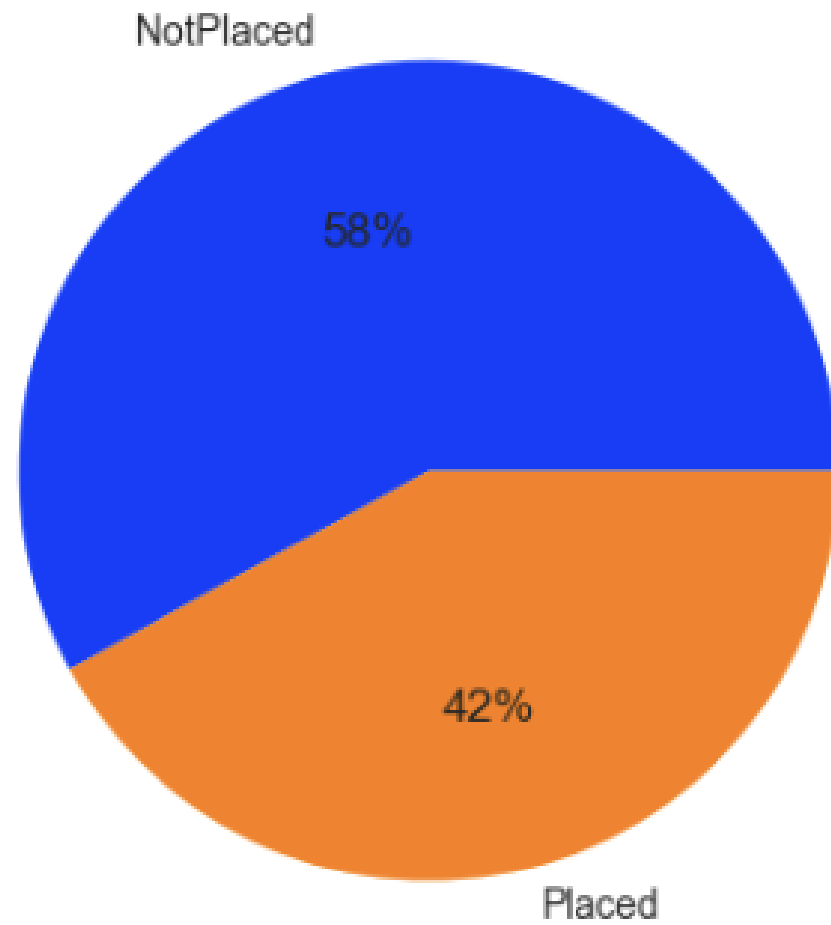
CountPlot of ExtracurricularActivities



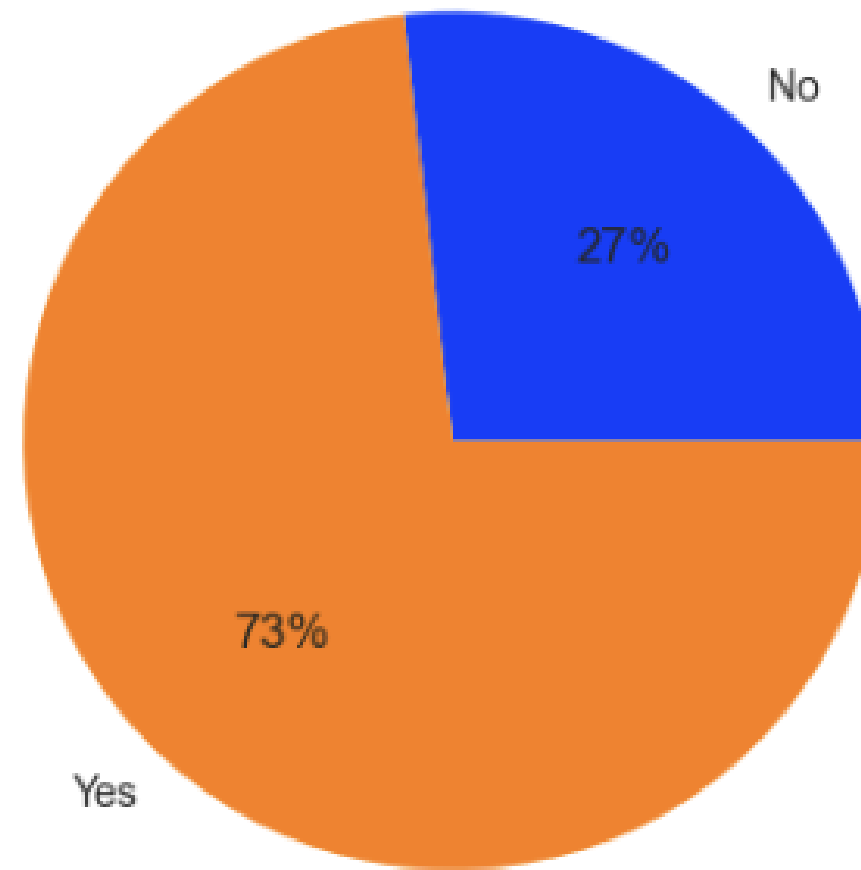


Percentage Distribution

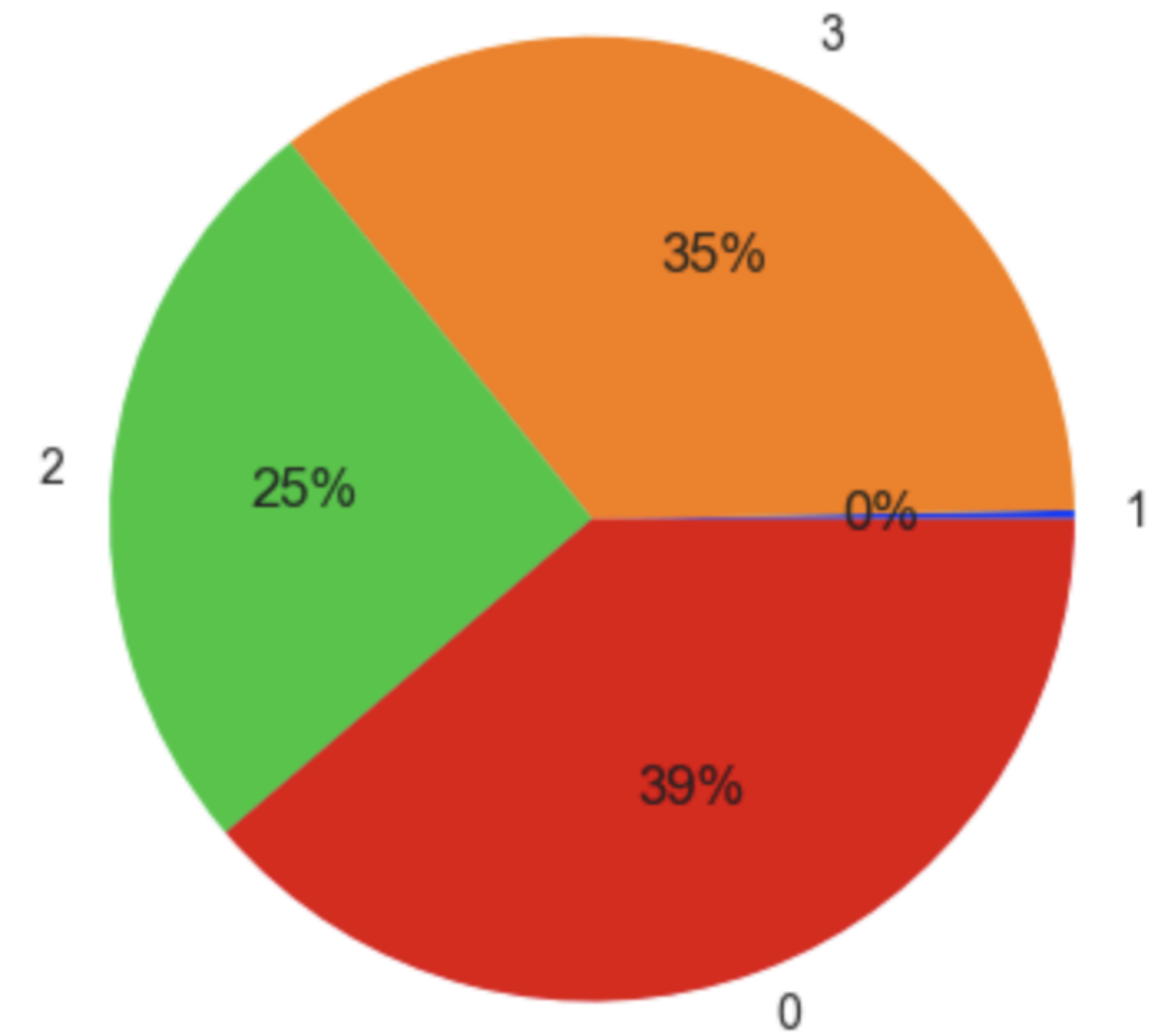
Pie Chart For PlacementStatus



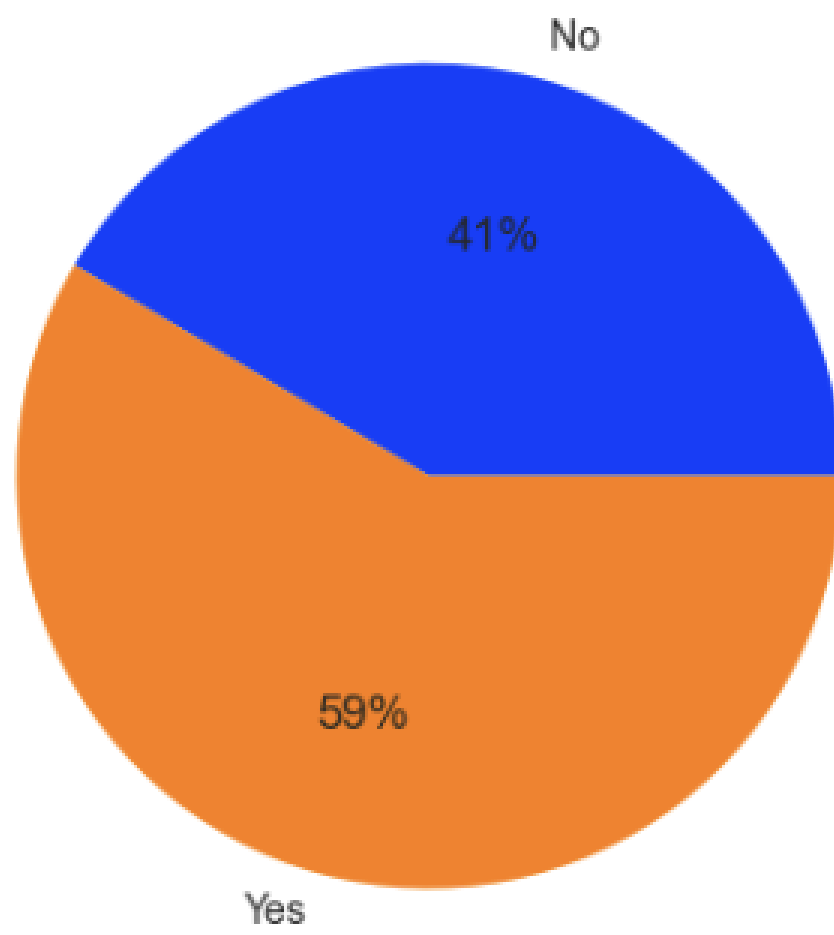
Pie Chart For PlacementTraining



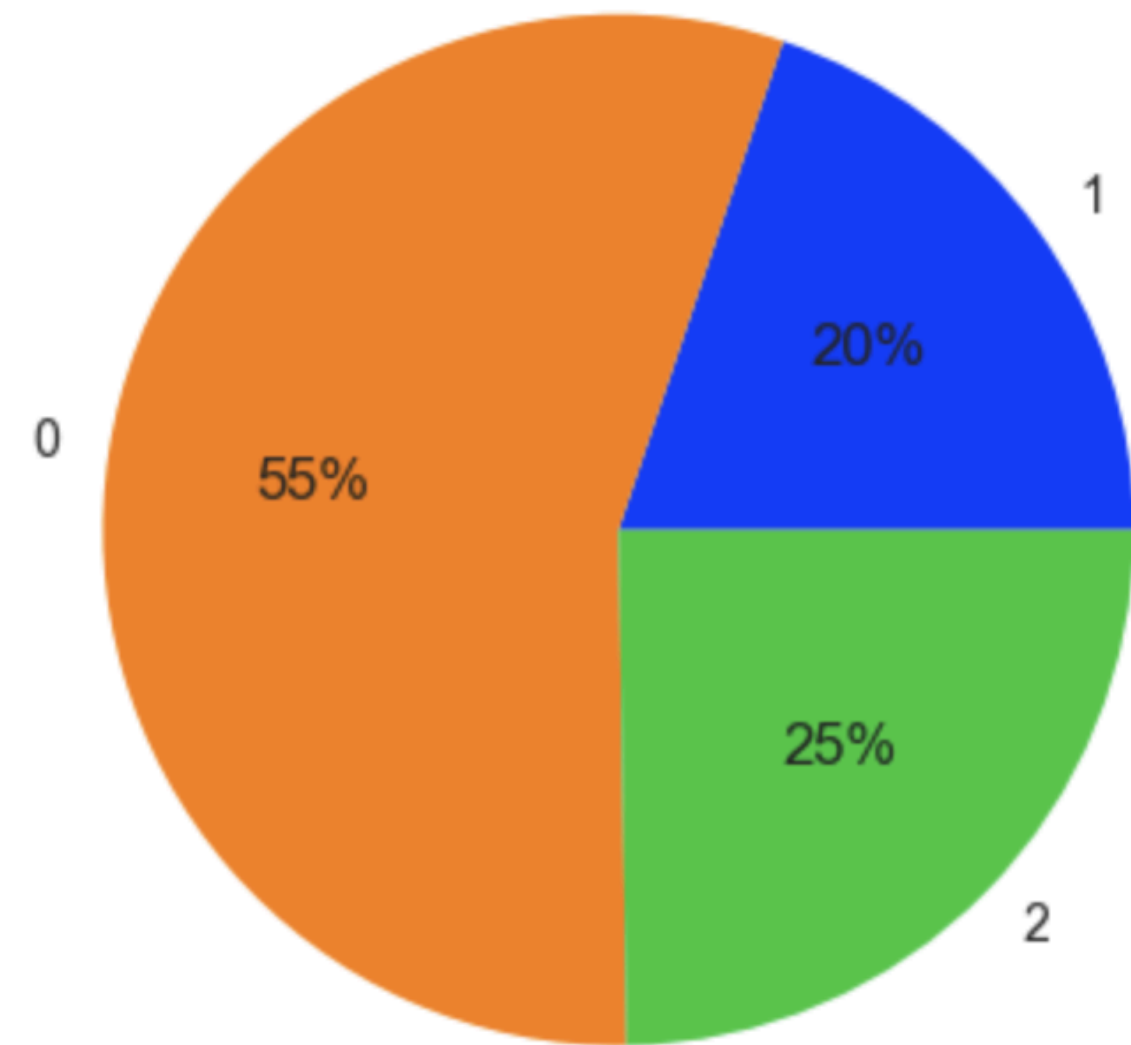
Pie Chart For Projects

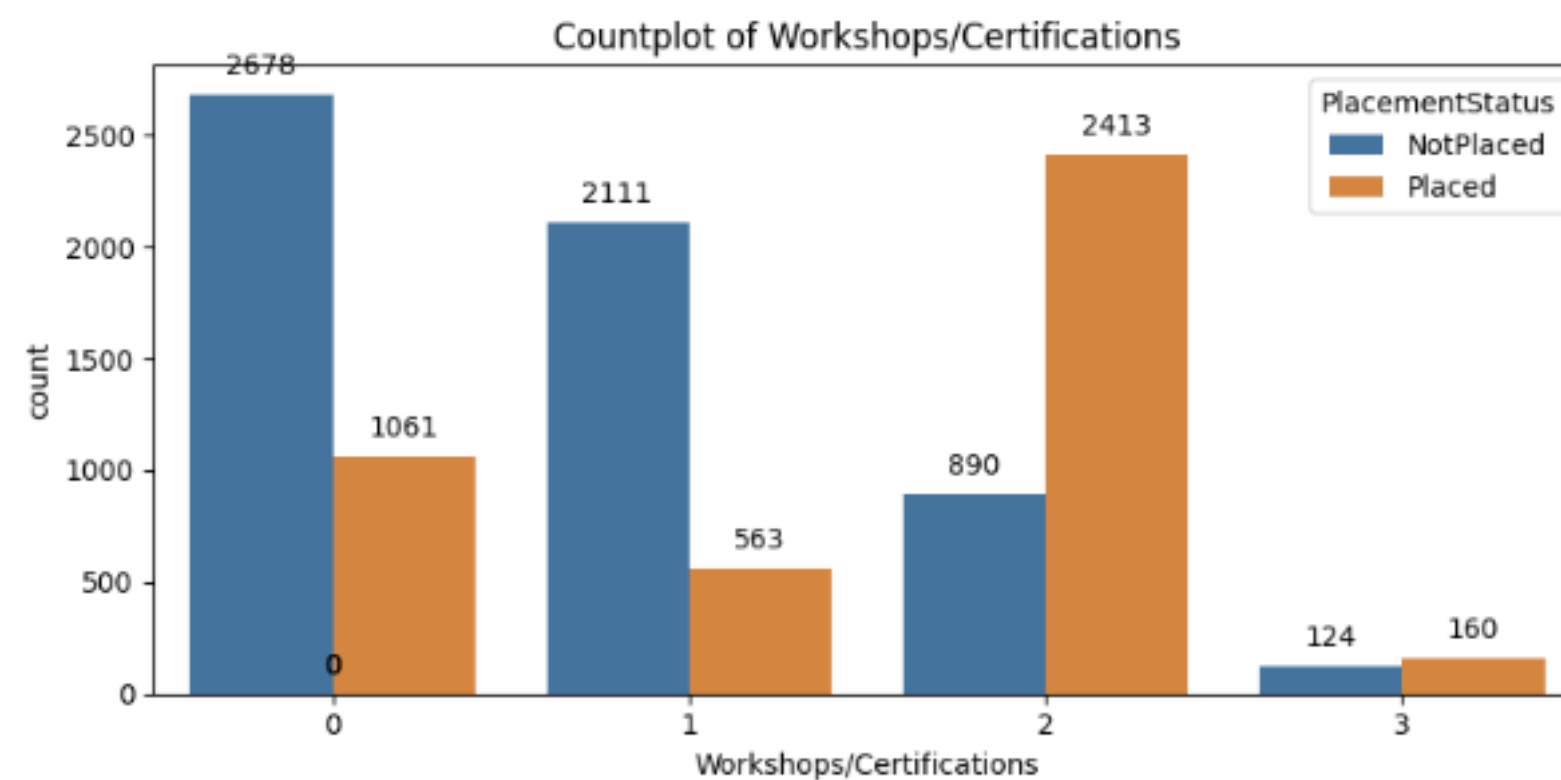
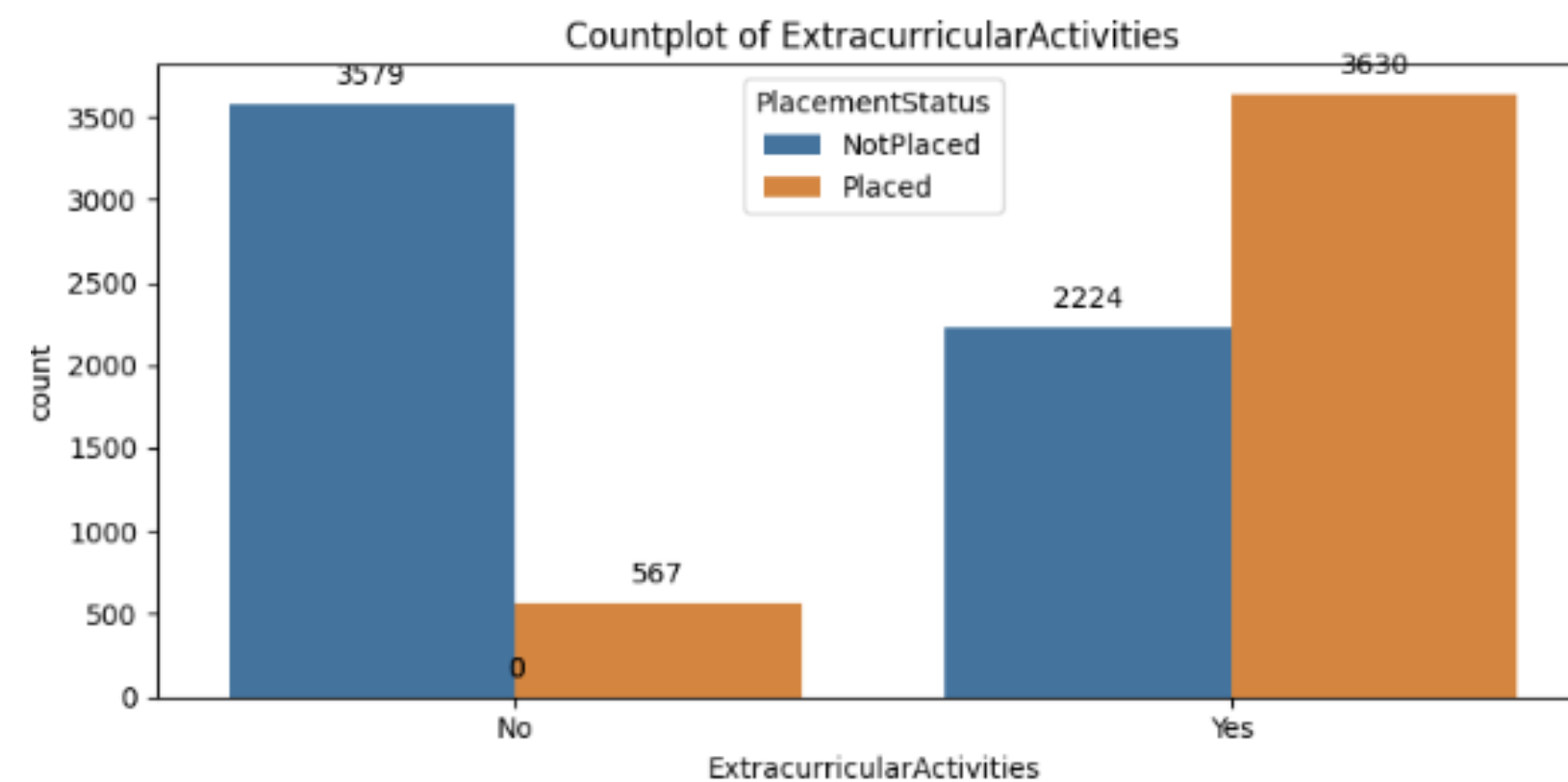
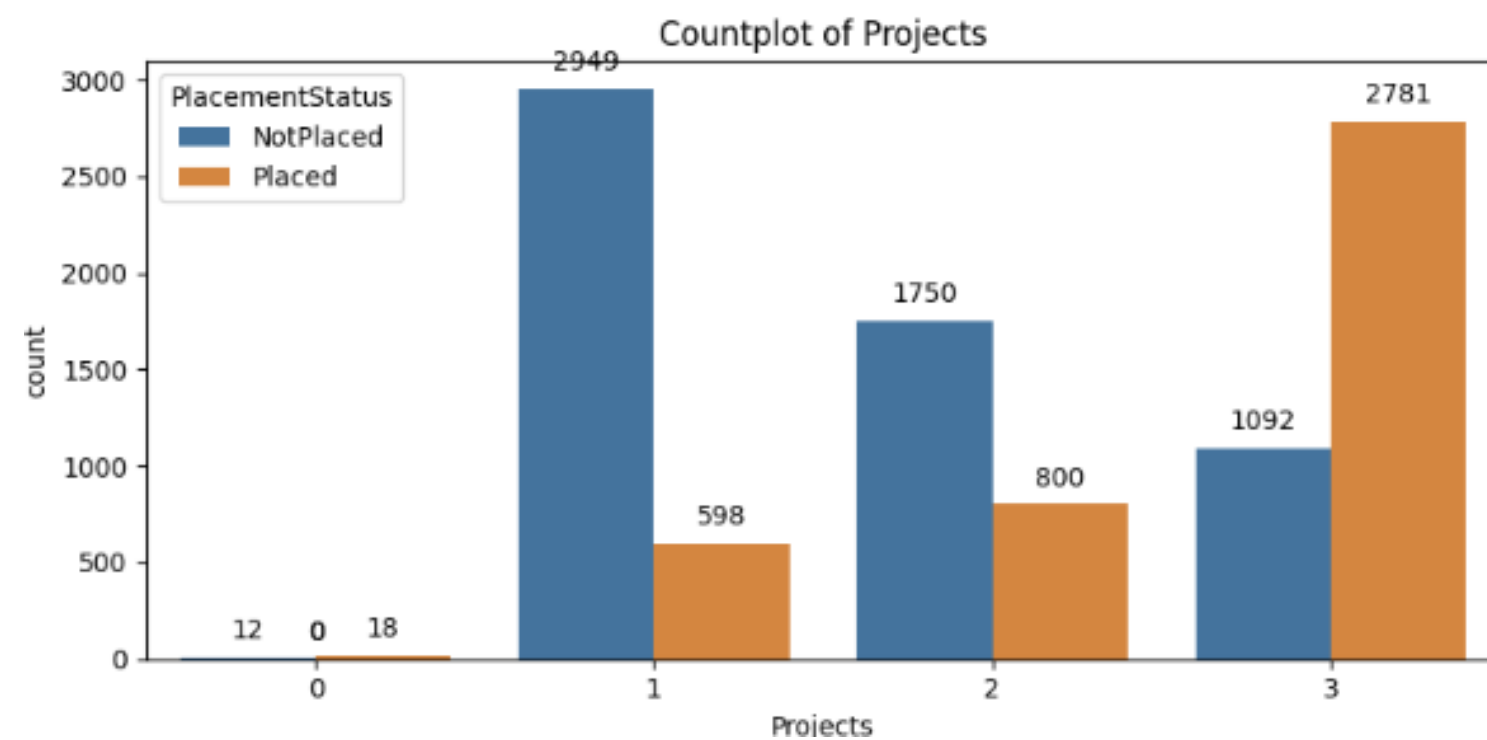


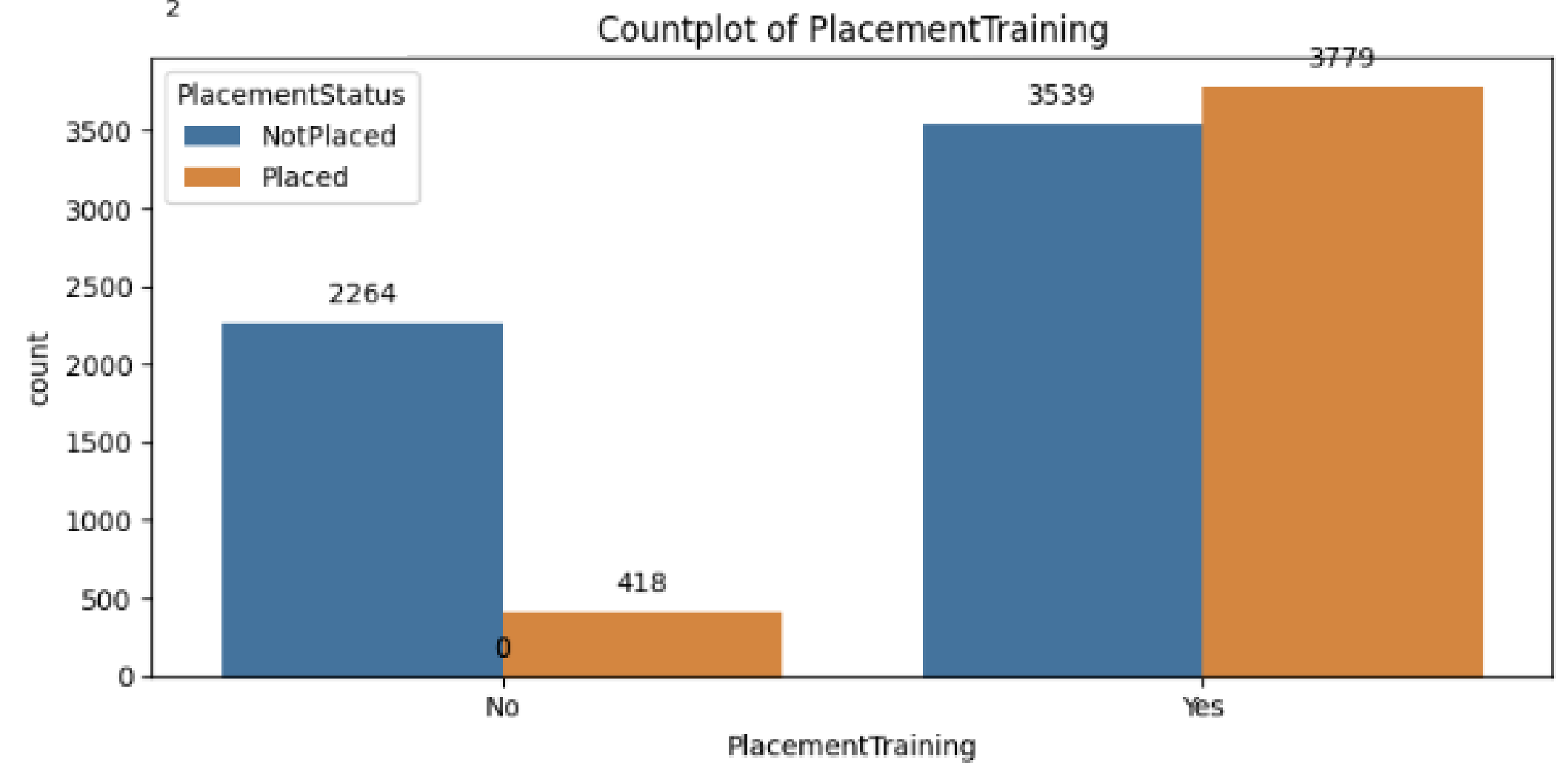
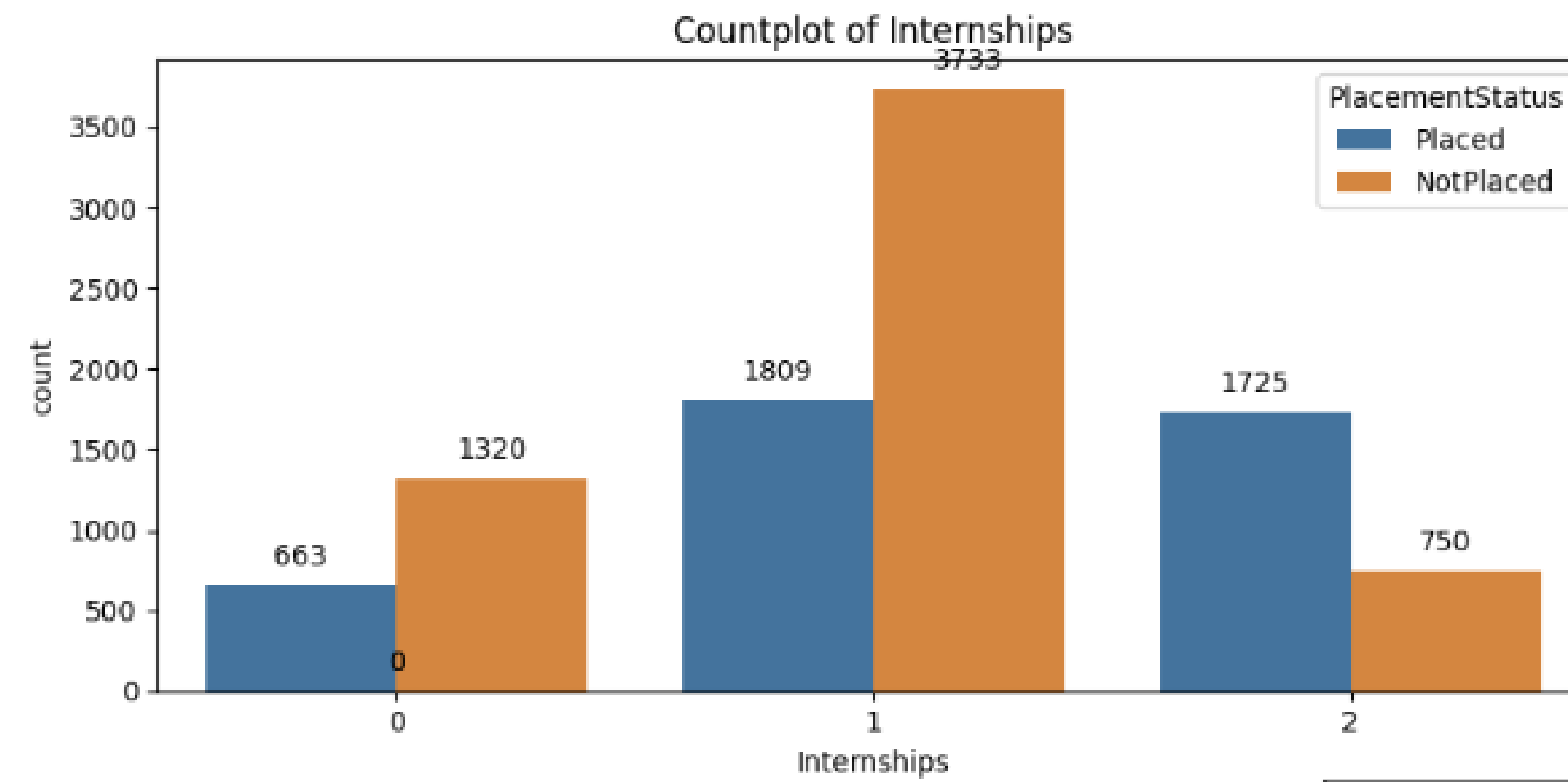
Pie Chart For ExtracurricularActivities



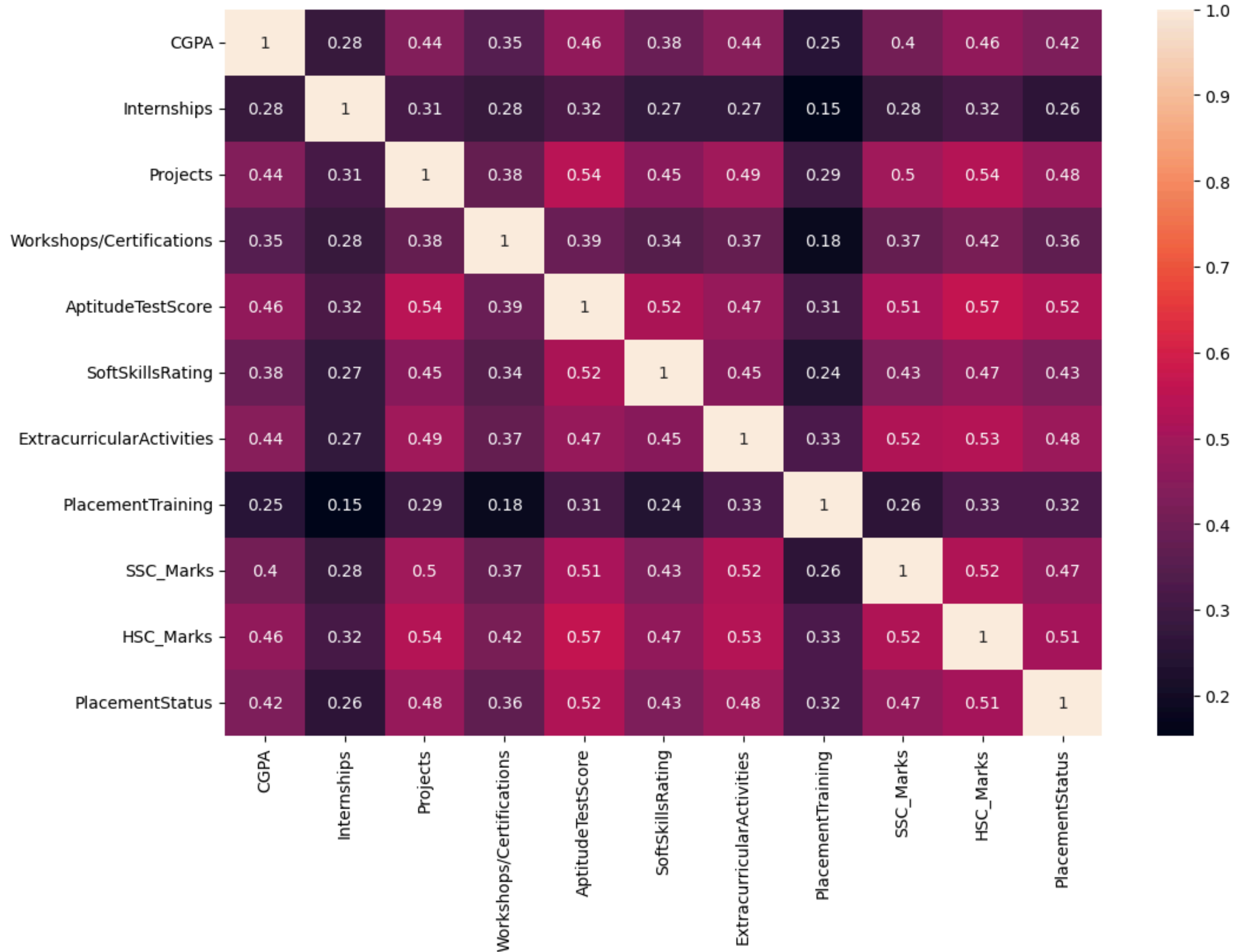
Pie Chart For Internships







HeatMap



Models used for making predictions

In our project, we have used 3 models to conduct a comparative study, aiming to analyze differences in accuracy. Models that we have used are:

- Logistic Regression
- Decision Tree
- K- Nearest Neighbor

Logistic Regression

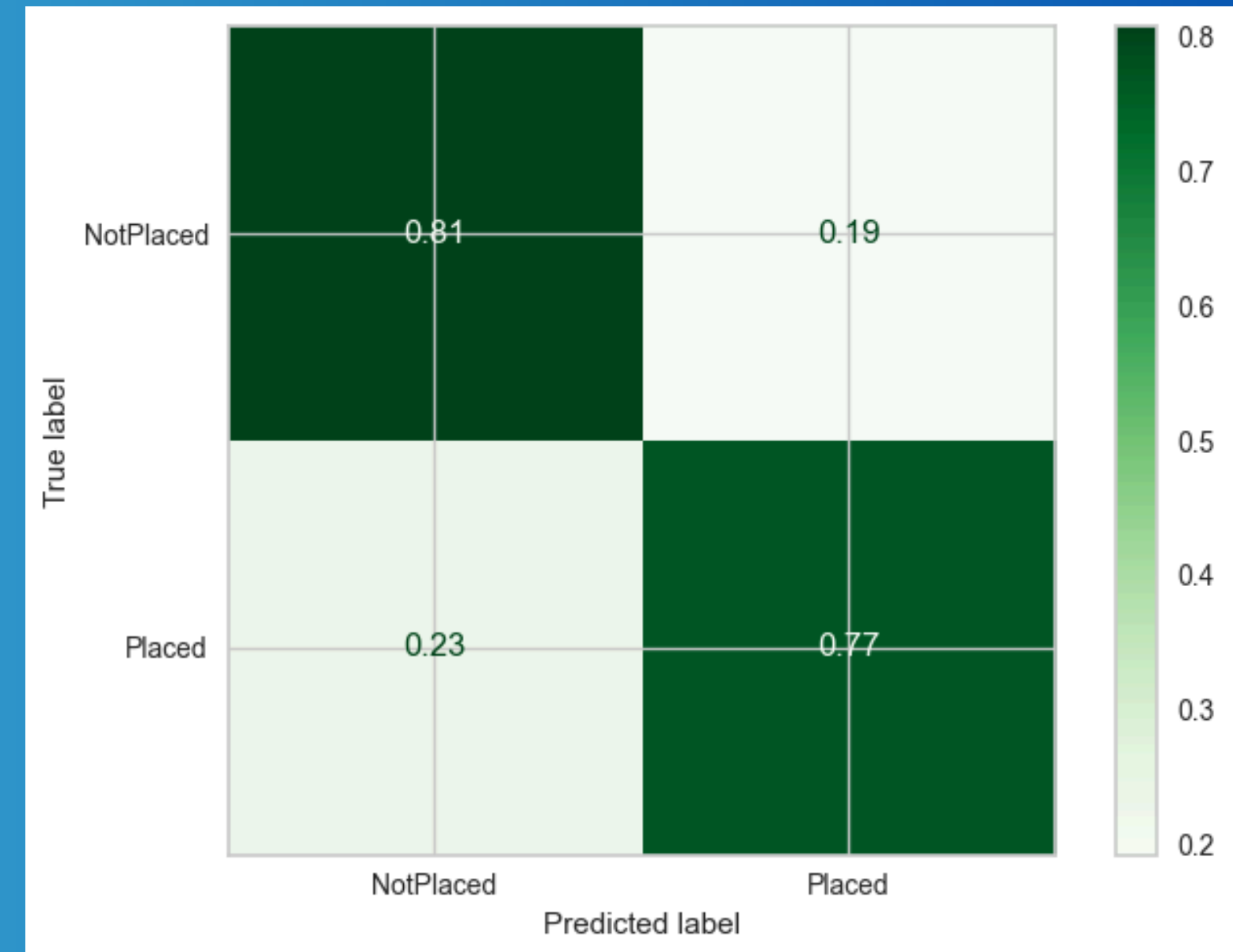
Logistic regression is a statistical method used for binary classification tasks. It predicts the probability of a categorical outcome by fitting data to a logistic function, transforming values into probabilities between 0 and 1.

-

This model scored an accuracy of approximately 79.28%.

Confusion Matrix

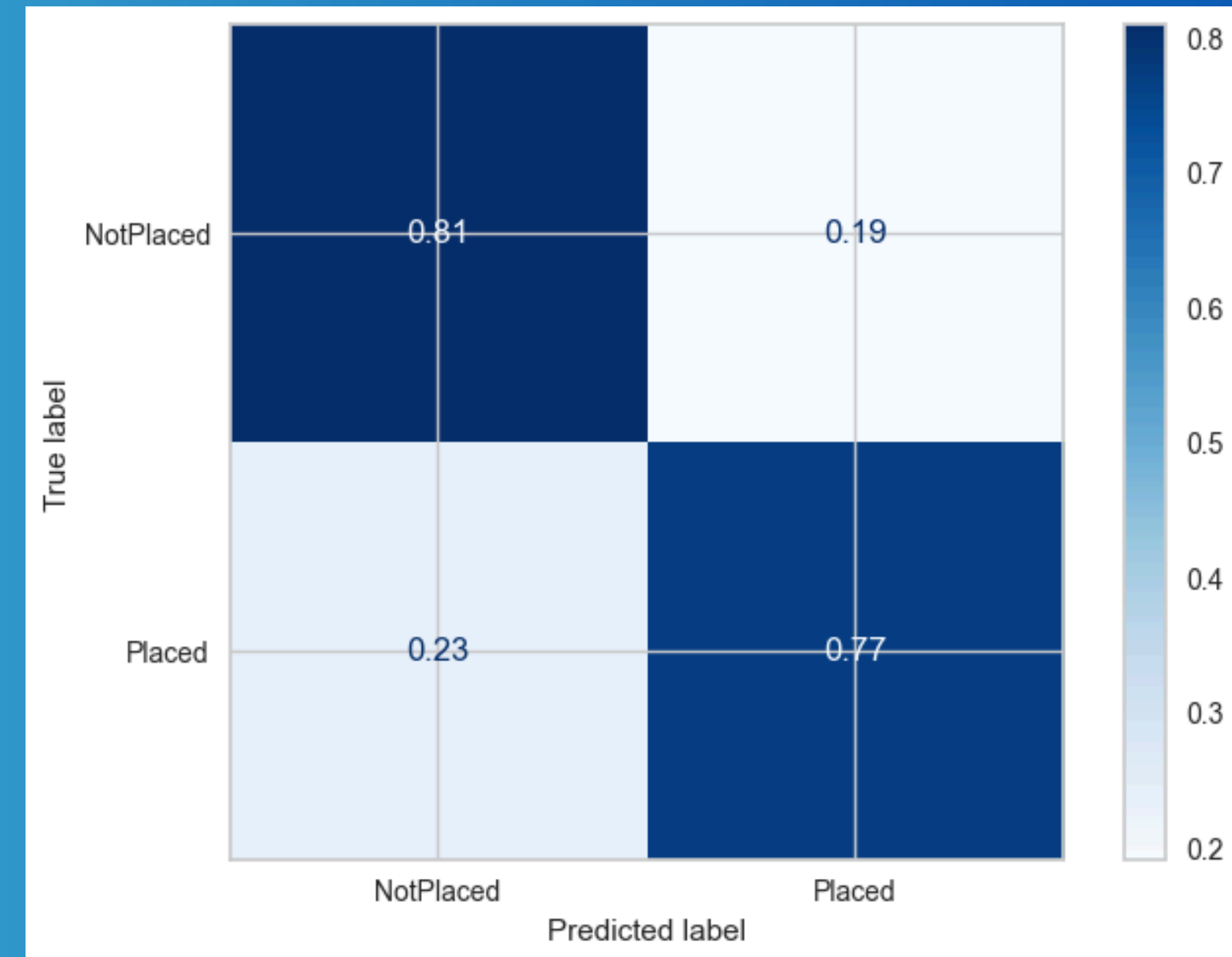
	Precision	Recall	F1-score	Support
Not Placed	0.83	0.81	0.82	1471
Placed	0.74	0.77	0.75	1029
Accuracy			0.7928	2500
Macro Avg	0.79	0.79	0.79	2500
Weighted Avg	0.80	0.79	0.79	2500



With Normalization

Confusion Matrix

	Precision	Recall	F1-score	Support
Not Placed	0.83	0.81	0.82	1471
Placed	0.74	0.77	0.75	1029
Accuracy			0.7924	2500
Macro Avg	0.79	0.79	0.79	2500
Weighted Avg	0.79	0.79	0.79	2500



Accuracy of Model 79.24%



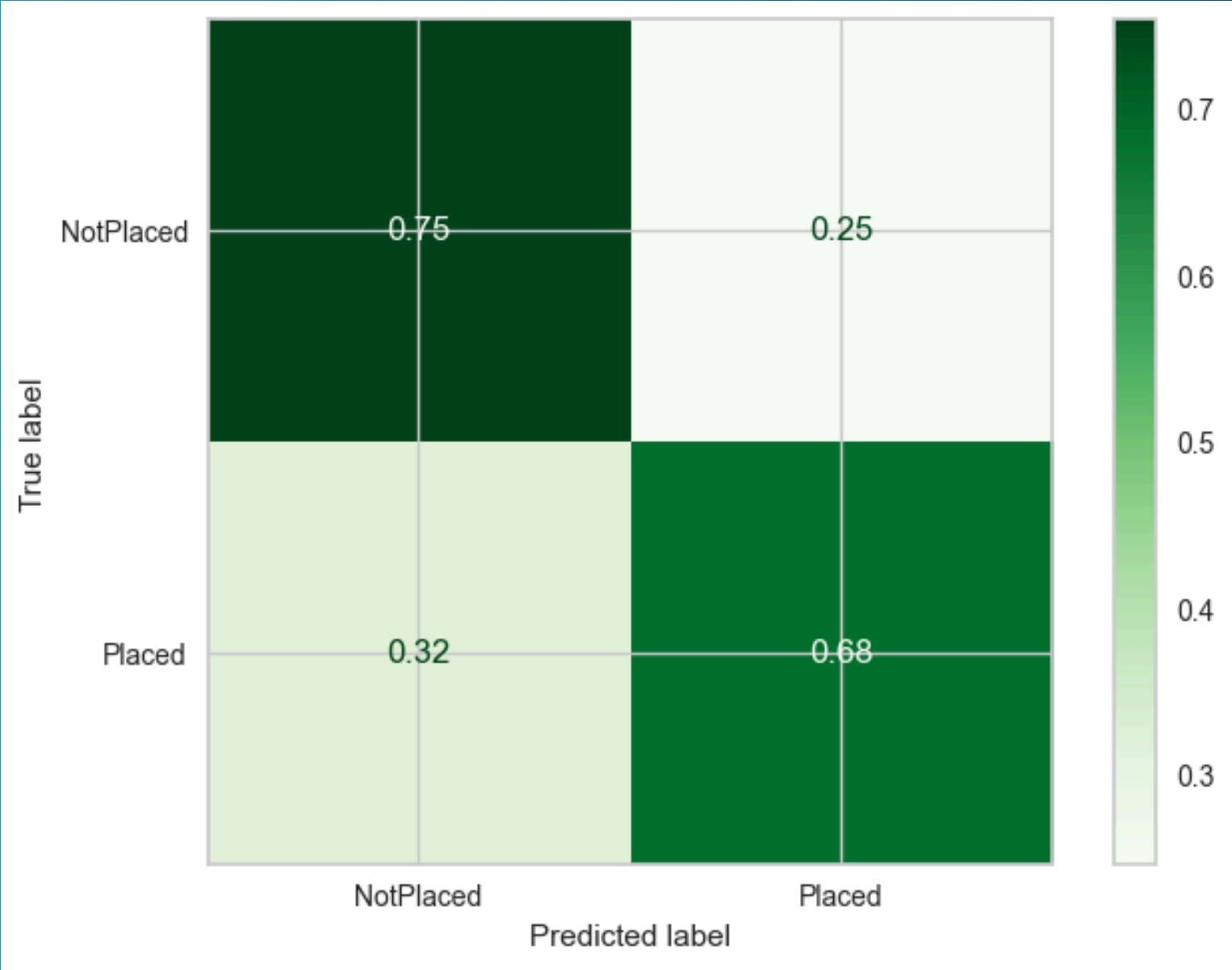
Decision Tree

In machine learning, a decision tree is a model that resembles a tree-like structure used for classification and regression tasks. It starts with a root node that represents the entire dataset and splits the data into branches based on specific features and their values. This process continues down the tree until reaching leaf nodes, which provide final decisions or predictions.

- **This model scored an accuracy of approximately 72.52%.**

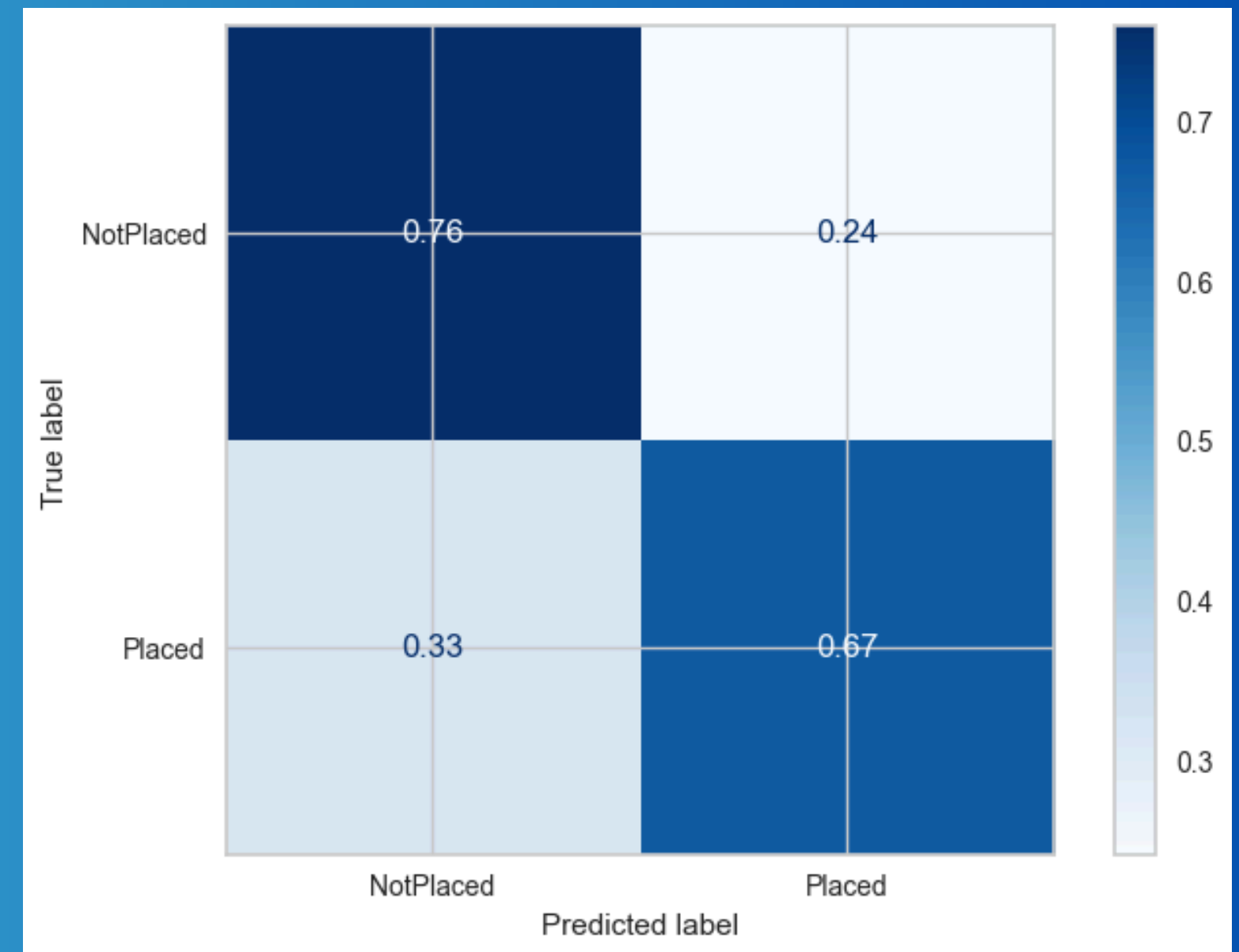
Confusion Matrix

	Precision	Recall	F1-score	Support
Not Placed	0.77	0.75	0.76	1471
Placed	0.66	0.68	0.67	1029
Accuracy			0.7252	2500
Macro Avg	0.72	0.72	0.72	2500
Weighted Avg	0.73	0.72	0.73	2500



With Normalization Confusion Matrix

	Precision	Recall	F1-score	Support
Not Placed	0.77	0.76	0.76	1471
Placed	0.66	0.67	0.67	1029
Accuracy			0.7248	2500
Macro Avg	0.72	0.72	0.72	2500
Weighted Avg	0.73	0.72	0.72	2500



Accuracy Of Model Is 72.48%



K Neighbour Classifier

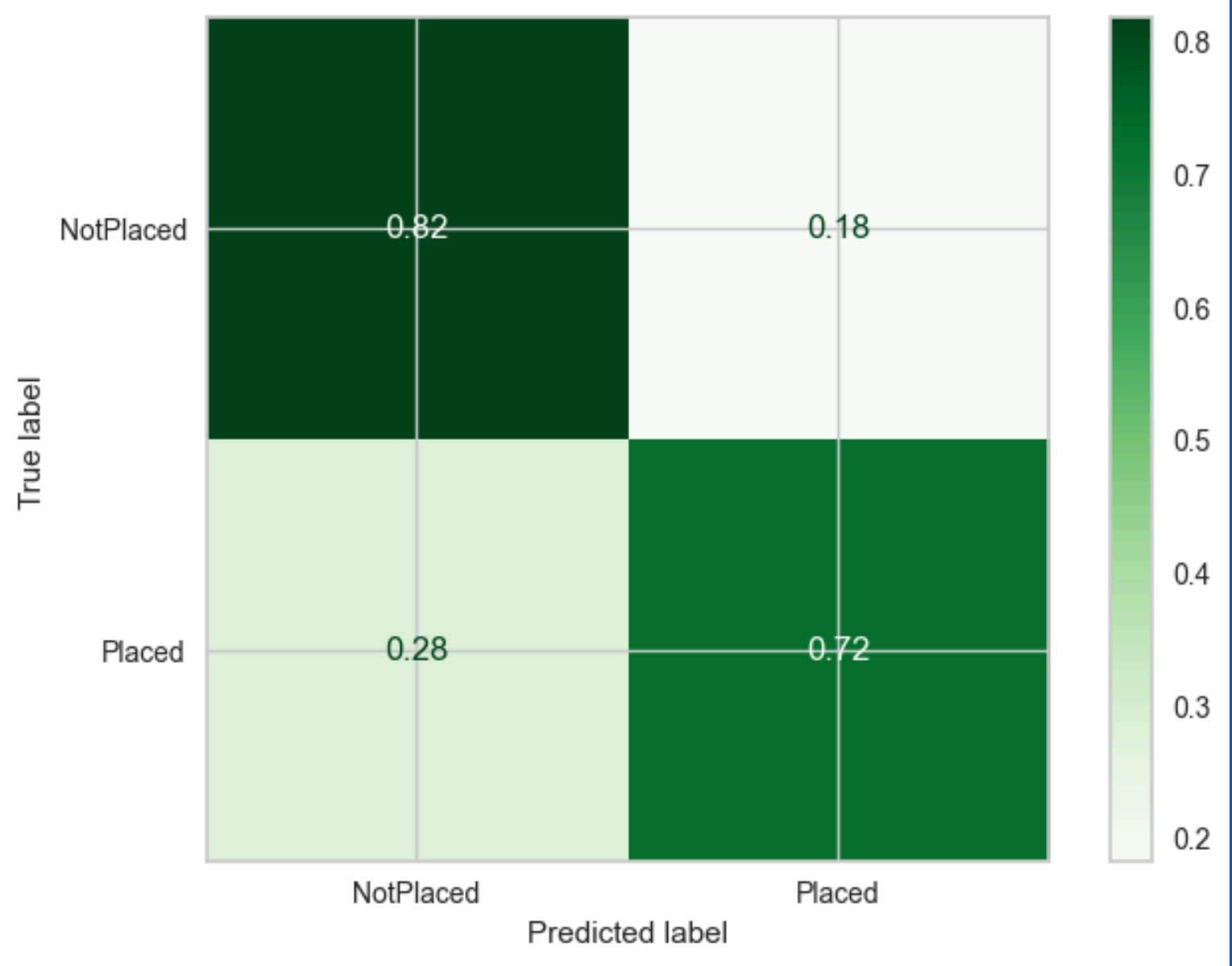
The K-Nearest Neighbors (KNN) Classifier is a straightforward yet effective supervised machine learning algorithm for classification tasks. It operates on the principle of similarity, classifying new data points based on their proximity to labeled data points in the training set. The algorithm determines the class of a new data point by identifying its K nearest neighbors in the feature space using a distance metric (like Euclidean distance).

This model scored an accuracy of approximately 77.80%.



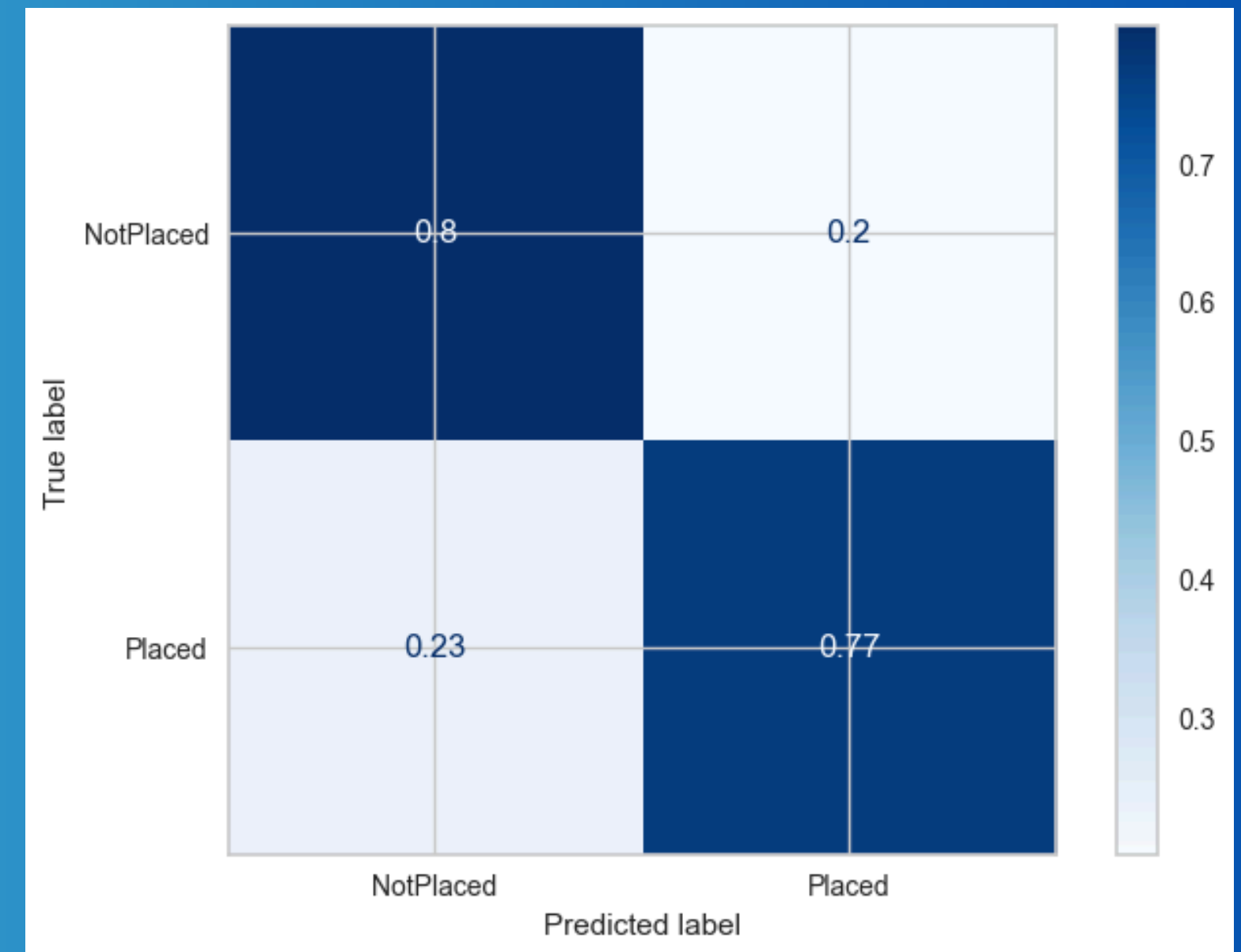
Confusion Matrix

	Precision	Recall	F1-score	Support
Not Placed	0.81	0.82	0.81	1471
Placed	0.73	0.72	0.73	1029
Accuracy			0.7780	2500
Macro Avg	0.77	0.77	0.77	2500
Weighted Avg	0.78	0.78	0.78	2500



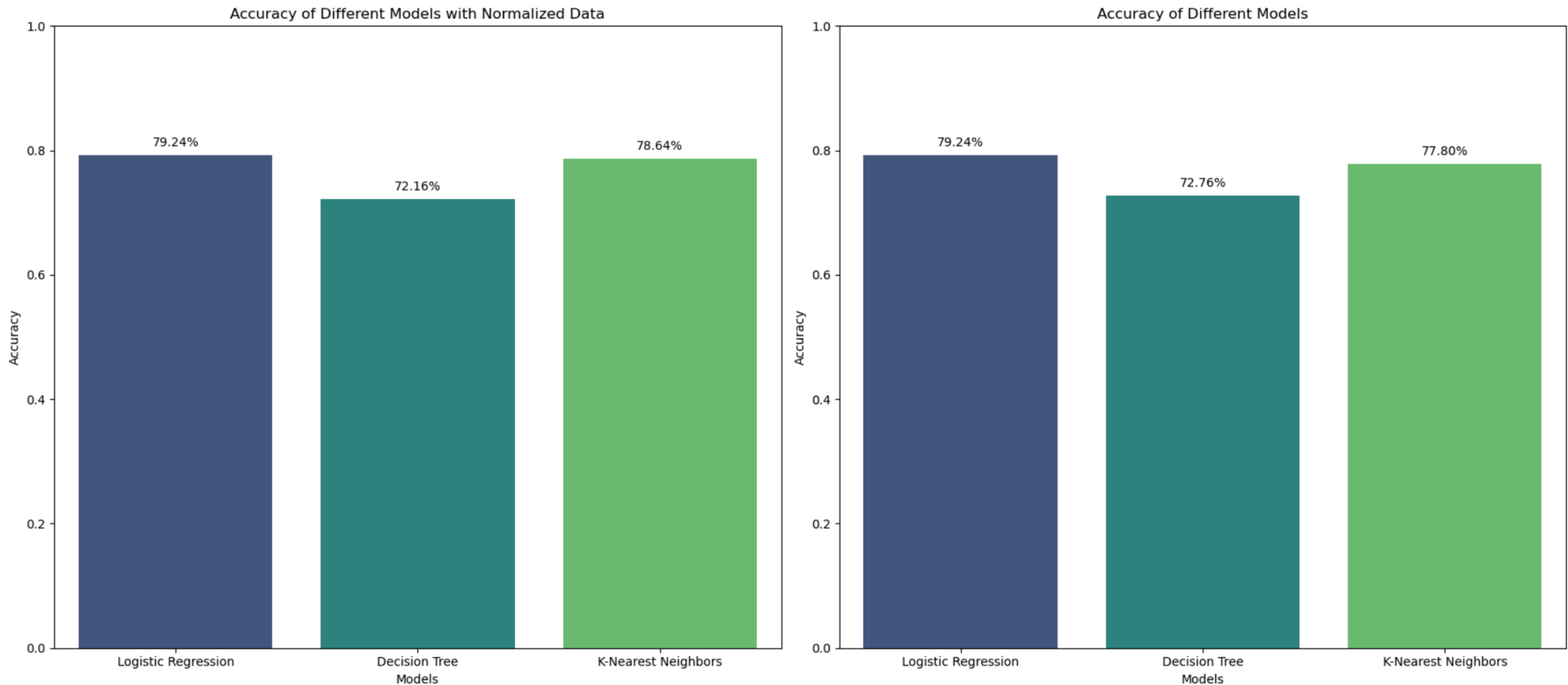
With Normalization Confusion Matrix

	Precision	Recall	F1-score	Support
Not Placed	0.83	0.80	0.81	1471
Placed	0.73	0.77	0.75	1029
Accuracy			0.7864	2500
Macro Avg	0.78	0.78	0.78	2500
Weighted Avg	0.79	0.79	0.79	2500



Accuracy Of Model Is 78.64%

Comparision Between Models



Conclusion & Future Work

- Primary objective achieved: Successfully predicted students' placement status post-final year using four classification algorithms (LR, DTC, KNN).
- In Future work we can apply this predictor Model on real time application instead of synthetic data.
- System's efficacy: Elevates institution's placement rates. Enhances institution's reputation.
- Signifies a substantial advancement in classification techniques for placement prediction.
- Stands as a pivotal tool to improve placement prediction methodologies significantly.

References and bibliography

- Senthil Kumar Thangavel, Divya Bharathi P and Abijith Sankar: Student Placement Analyzer: A Recommendation System Using Machine Learning 2017 International Conference on Advanced Computing and Communication Systems.
- Student Placement Prediction Model: A Data Mining Perspective for Outcome-Based Education System Article in International Journal of Recent Technology And Engineering(IRJTE) September 2019:-By Abhishek Rao, NMAM Institute of Technology
- 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) A Review on Student Placement Chance Prediction
- 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) Campus Placement Predictive Analysis using Machine Learning
- <https://www.kaggle.com/datasets/chandhurubaskar/campus-placement-data-for-engineering-colleges/code?datasetId=3678158&sortBy=voteCount>



THANK YOU

