

Experiment 2

Aim: Data Visualization and Exploratory Data Analysis using Matplotlib and Seaborn.

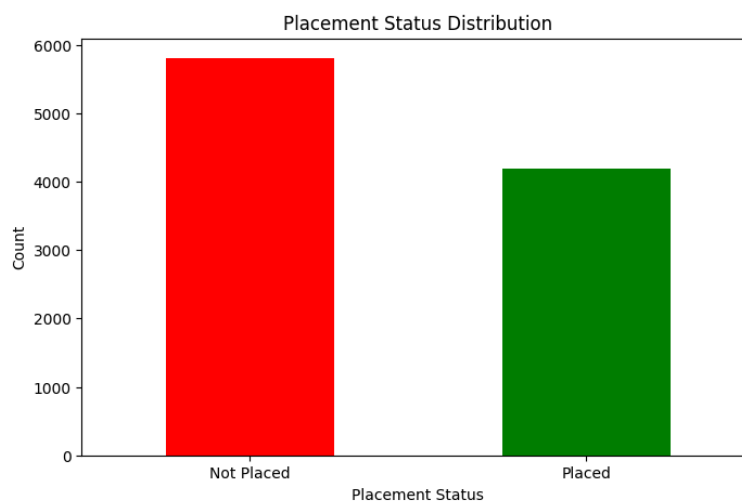
Theory: Exploratory Data Analysis (EDA) and visualization techniques help us understand trends, patterns, and correlations in data. Matplotlib is a widely used library for creating various types of graphs, while Seaborn provides a high-level interface for statistical visualizations. Through bar graphs, scatter plots, histograms, and heatmaps, we can gain insights into student academic performance and its impact on placement outcomes.

1. Bar Graph and Contingency Table

a. Bar Graph for Placement Status Distribution

A bar graph is useful for showing placement trends among students.

```
plt.figure(figsize=(8, 5))
data['PlacementStatus_Placed'].value_counts().plot(kind='bar', color=['red', 'green'])
plt.title("Placement Status Distribution")
plt.xlabel("Placement Status")
plt.ylabel("Count")
plt.xticks([0, 1], ["Not Placed", "Placed"], rotation=0)
plt.show()
```



Observation: The graph shows that around 4000 students have successfully secured placements. However, a significant proportion remains unplaced, indicating potential gaps in skill development or industry expectations.

b. Contingency Table for Placement and Training Participation

A contingency table helps analyze the relationship between placement training and placement status.

```
contingency_table = pd.crosstab(data['PlacementTraining_Yes'], data['PlacementStatus_Placed'])  
print(contingency_table)
```

PlacementStatus_Placed	False	True
PlacementTraining_Yes		
False	2264	418
True	3539	3779

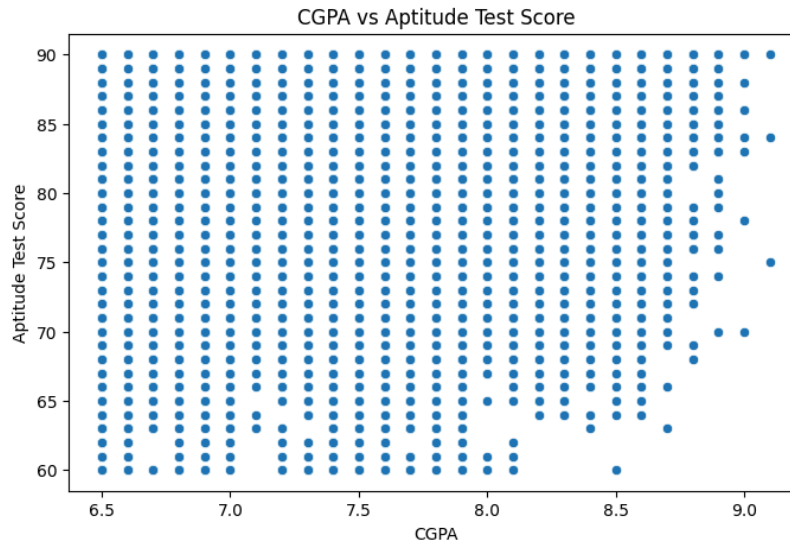
Observation: The table reveals that students who participated in placement training had a significantly higher placement rate. This confirms that additional coaching and guidance play a crucial role in job readiness. Students who opted out of training faced difficulties in securing jobs, highlighting the importance of structured preparation.

2. Scatter Plot, Box Plot, and Heatmap

a. Scatter Plot for CGPA vs Aptitude Test Score

A scatter plot helps in visualizing the relationship between academic CGPA and aptitude test scores.

```
data = pd.read_csv("new_data.csv")  
plt.figure(figsize=(8, 5))  
sns.scatterplot(x=data['CGPA'], y=data['AptitudeTestScore'])  
plt.title("CGPA vs Aptitude Test Score")  
plt.xlabel("CGPA")  
plt.ylabel("Aptitude Test Score")  
plt.show()
```

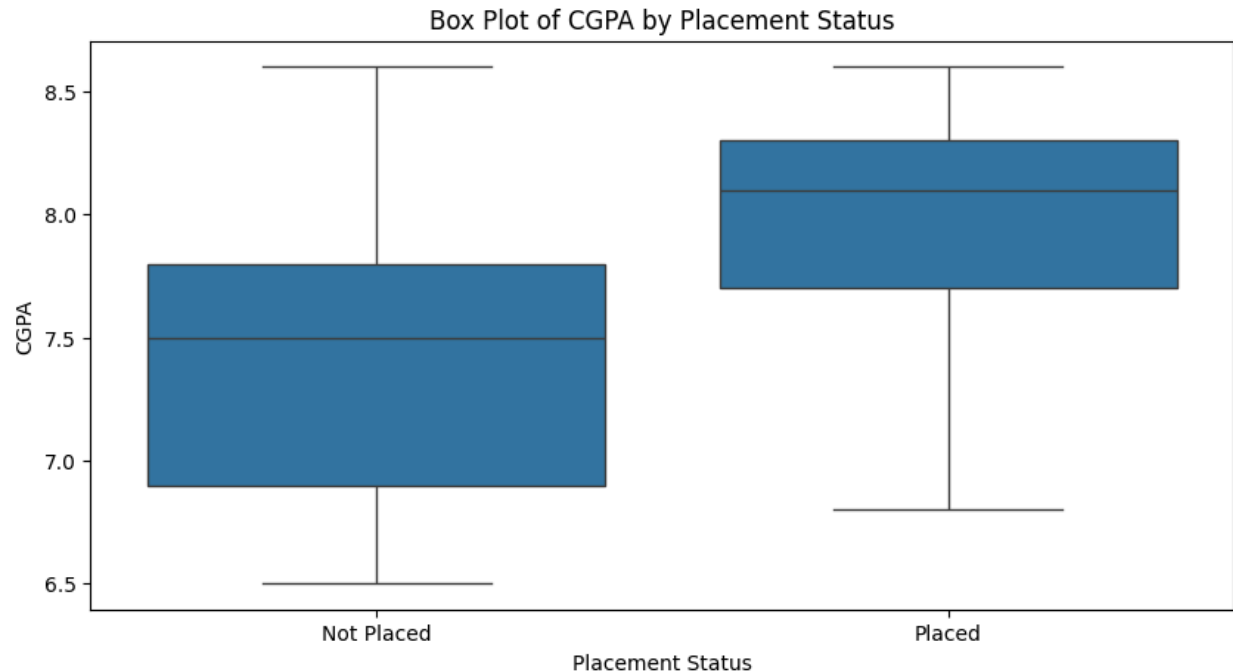


Observation: The scatter plot shows a moderate positive correlation between CGPA and aptitude test scores. Students with higher CGPA tend to perform better in aptitude tests. However, there are some students with high CGPA but lower aptitude scores, indicating that academic excellence does not always translate into better test performance. Similarly, some students with lower CGPA have performed well in aptitude tests, suggesting strong problem-solving abilities despite lower grades.

b. Box Plot for CGPA and PLACEMENT

A box plot helps visualize the distribution of internships and projects.

```
plt.figure(figsize=(10, 5))
sns.boxplot(x=df_filtered["PlacementStatus_Placed"], y=df_filtered["CGPA"])
plt.title("Box Plot of CGPA by Placement Status")
plt.xticks(ticks=[0, 1], labels=["Not Placed", "Placed"])
plt.xlabel("Placement Status")
plt.ylabel("CGPA")
plt.show()
```

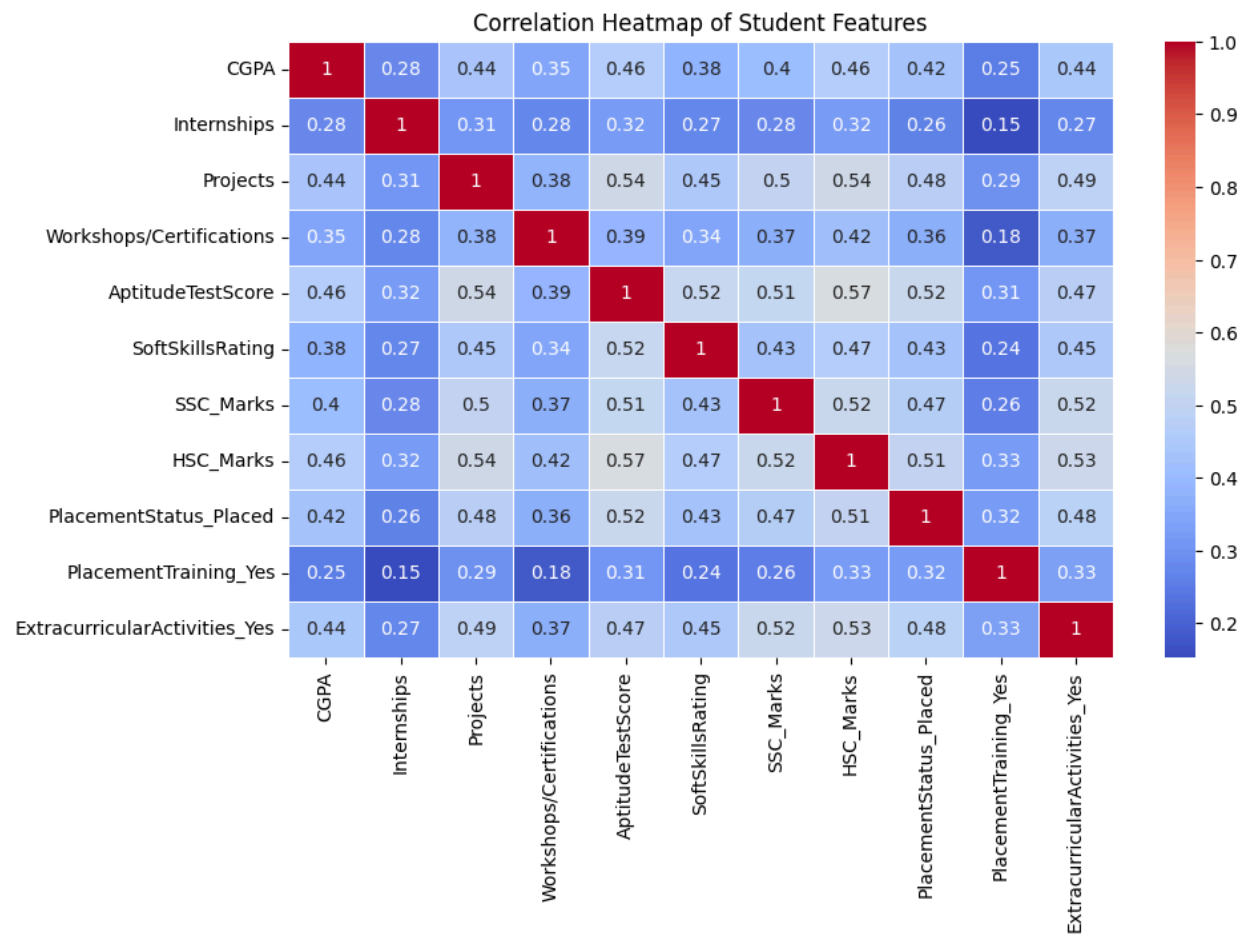


Observation: The box plot shows that placed students generally have a higher and more consistent CGPA, while not placed students have lower and more varied CGPA. Some with lower CGPA still got placed, suggesting other factors like projects and internships influence placements.

c. Heatmap for Correlation Analysis

A heatmap helps visualize correlations between various academic and placement-related factors.

```
plt.figure(figsize=(10, 6))
sns.heatmap(data.corr(), annot=True, cmap='coolwarm', linewidths=0.5)
plt.title("Correlation Heatmap of Student Features")
plt.show()
```



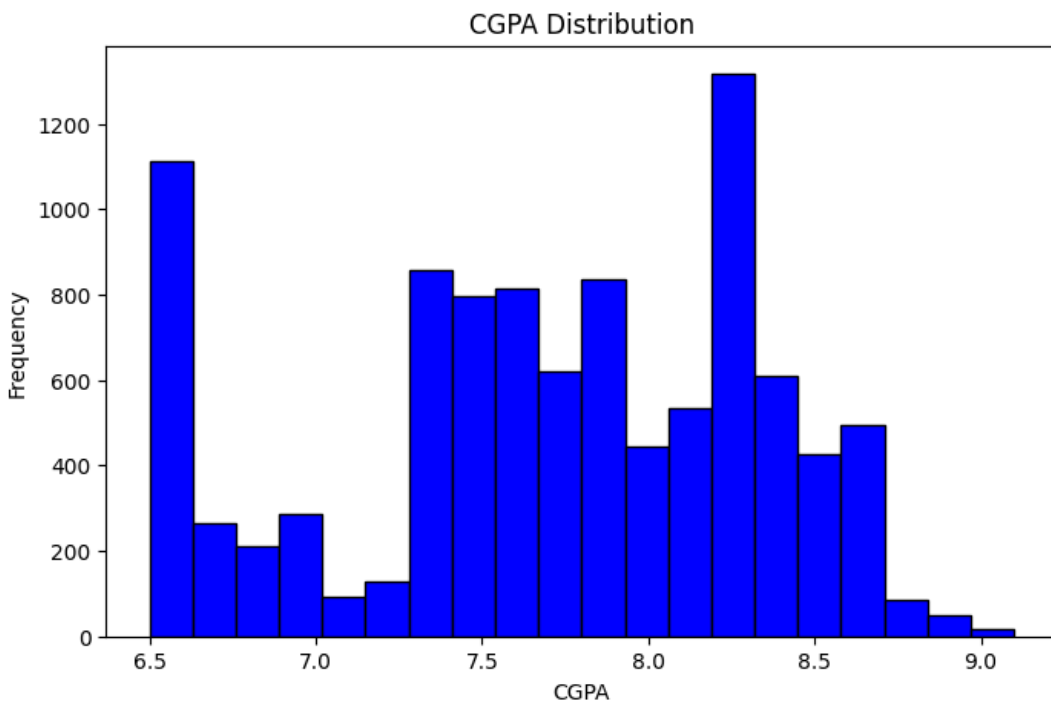
Observation: Strong positive correlations are observed between CGPA, SSC, and HSC marks, indicating that students who perform well in early academics tend to maintain good grades throughout. Placement status has a noticeable dependence on CGPA and training participation, confirming that higher academic performance and additional placement training significantly boost placement chances. Surprisingly, soft skills rating does not show a strong correlation with placement, suggesting that employers may prioritize technical skills over soft skills.

3. Histogram and Normalized Histogram

a. Histogram for CGPA Distribution

A histogram represents the frequency distribution of CGPA values among students.

```
plt.figure(figsize=(8, 5))
plt.hist(data['CGPA'], bins=20, color='blue', edgecolor='black')
plt.title("CGPA Distribution")
plt.xlabel("CGPA")
plt.ylabel("Frequency")
plt.show()
```

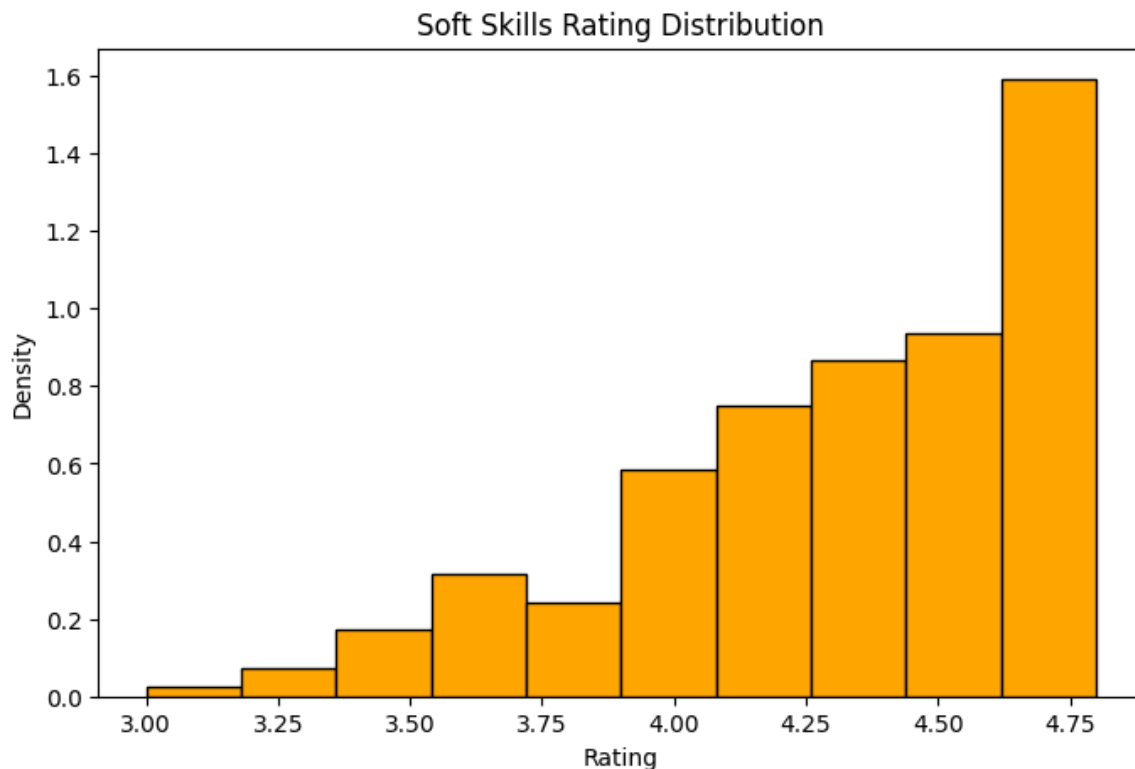


Observation: The histogram indicates that most students have a CGPA between 7.0 and 8.5, with a few having extreme high or low values. A small group of students with CGPA below 6.0 may face difficulties in placement, as many companies set a minimum CGPA criterion for eligibility.

b. Normalized Histogram for Soft Skills Rating

A normalized histogram represents probability densities instead of raw counts.

```
plt.figure(figsize=(8, 5))
plt.hist(data['SoftSkillsRating'], bins=10, color='orange', edgecolor='black', density=True)
plt.title("Soft Skills Rating Distribution")
plt.xlabel("Rating")
plt.ylabel("Density")
plt.show()
```



Observation: The majority of students have a soft skills rating between 4.0 and 5.0, indicating a high level of competency in communication and teamwork skills. However, a small fraction of students with low soft skills ratings may struggle in interviews and teamwork-based roles, emphasizing the need for soft skills development alongside technical proficiency.

Conclusion

This experiment successfully demonstrated the application of Matplotlib and Seaborn for data visualization and exploratory analysis in student placement data.

By looking at different student-related features, we gained insights into factors influencing placements, such as CGPA, aptitude scores, placement training, and hands-on experience through internships and projects. We also found that students with lower soft skills ratings struggled despite good academic performance, which highlights the importance of communication and teamwork.

Using scatter plots, heatmaps, and histograms helped visualize the relationships and distributions in the data.