STUDENT PLACEMENT ANALYSIS:

VISUALIZING SUCCESS FACTORS IN CAMPUS HIRING

Insights from Academic Performance to Placement Outcomes

PROJECT OVERVIEW

- Campus placements are a critical milestone for students and educational institutions, reflecting both academic achievement and employability. Understanding the factors that influence placement outcomes can help colleges enhance their training programs and guide students toward successful careers.
- Dataset:
- Real-world data of students' academic, demographic, and placement details.
- Size: 2,966 student records (after cleaning: 1,137 unique entries)

GOAL OF THE PROJECT

- Primary Goal:
- To identify and visualize the most significant factors that impact student placement success.
- Objectives:
- Understand the data structure and quality
- Clean and preprocess the data
- Perform exploratory data analysis (EDA)
- Draw actionable insights for students and institutions

TOOLS & TECHNOLOGIES USED

Python

- The main programming language for data analysis and visualization due to its flexibility and rich ecosystem.

Pandas

- Used for data loading, cleaning, manipulation, and exploration.
- Enabled efficient handling of tabular data and preprocessing tasks.

Matplotlib & Seaborn

- Powerful libraries for creating a wide range of visualizations.
- Used to generate bar charts, histograms, boxplots, heatmaps, and more for clear, insightful EDA.

Jupyter Notebook

- Provided an interactive environment for combining code, visualizations, and documentation in a single, shareable format.
- Facilitated step-by-step analysis and easy presentation of results.

DATA CLEANING & PREPROCESSING

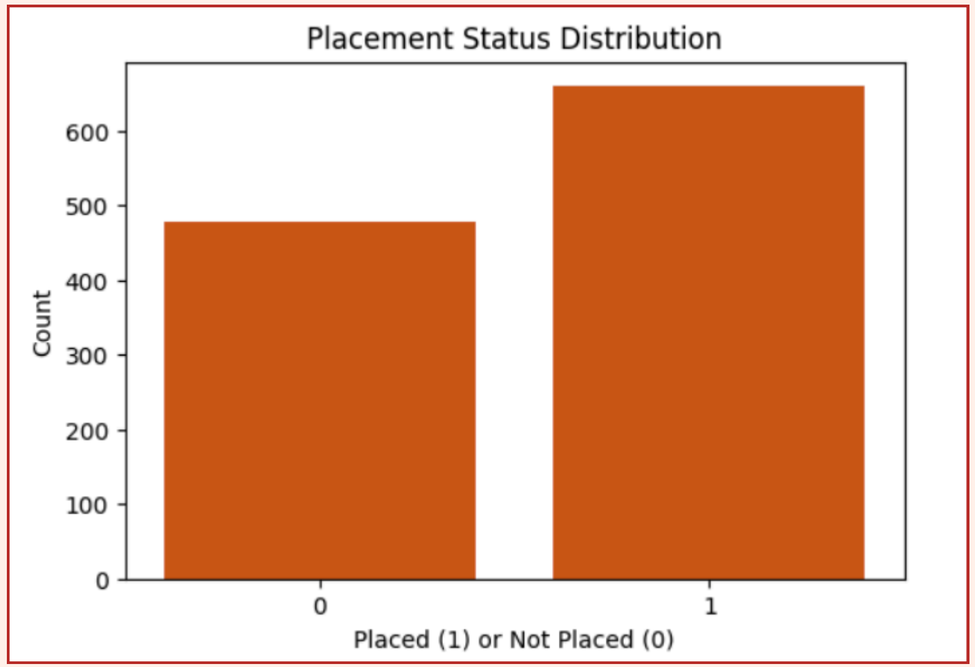
- Missing Values:
  - Checked and confirmed no missing values in any column.
- Duplicates:
  - Identified and removed 1,829 duplicate rows for data integrity.
- Data Types:
  - Converted categorical columns to category dtype for efficient analysis.
- Consistency Checks:
  - Verified unique values for all categorical columns (Gender, Hostel, Backlogs, Stream).

EDA HIGHLIGHTS

Target Variable Distribution:

- ~55% of students were placed; ~45% were not.

```
plt.figure(figsize=(6,4))
sns.countplot(x='PlacedOrNot', data=df)
plt.title('Placement Status Distribution')
plt.xlabel('Placed (1) or Not Placed (0)')
plt.ylabel('Count')
plt.show()
```

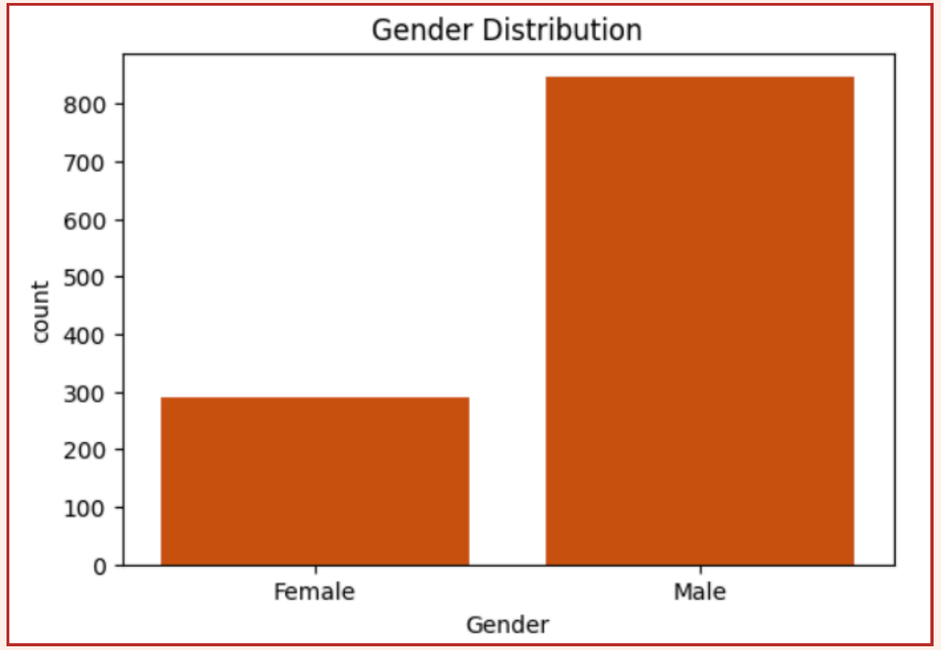


EDA HIGHLIGHTS

Gender Distribution:

- Both male and female students are represented.

```
plt.figure(figsize=(6,4))
sns.countplot(x='Gender', data=df)
plt.title('Gender Distribution')
plt.show()
```

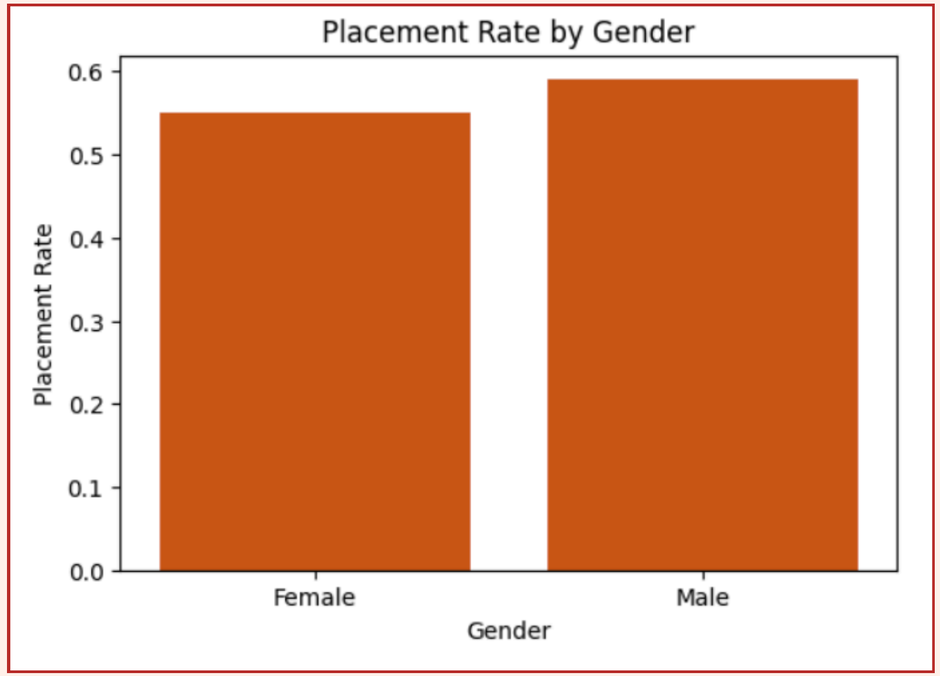


EDA HIGHLIGHTS

Placement Rate by Gender:

- Placement rates are slightly higher for males than females.

```
plt.figure(figsize=(6,4))
placement_gender = df.groupby('Gender')['PlacedOrNot'].mean().reset_index()
sns.barplot(x='Gender', y='PlacedOrNot', data=placement_gender)
plt.title('Placement Rate by Gender')
plt.ylabel('Placement Rate')
plt.show()
```

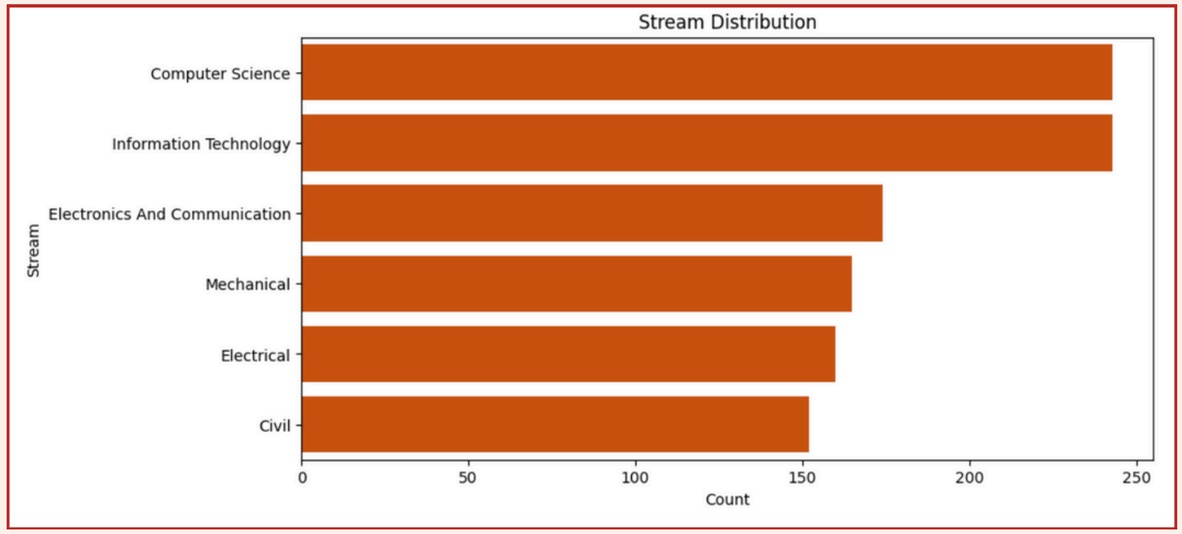


EDA HIGHLIGHTS

Stream Distribution:

- Streams include ECE, CSE, IT, Mechanical, Electrical, Civil.

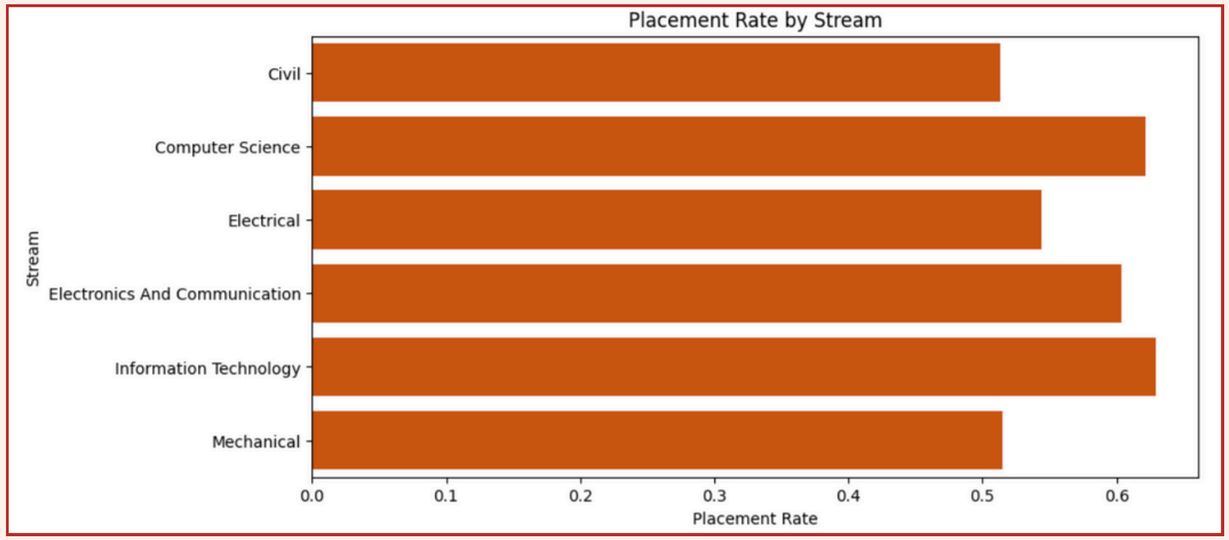
```
plt.figure(figsize=(10,5))
sns.countplot(y='Stream', data=df, order=df['Stream'].value_counts().index)
plt.title('Stream Distribution')
plt.xlabel('Count')
plt.ylabel('Stream')
plt.show()
```



EDA HIGHLIGHTS

Placement Rate by Stream:

- Some streams (e.g., CSE, ECE) have higher placement rates than others.

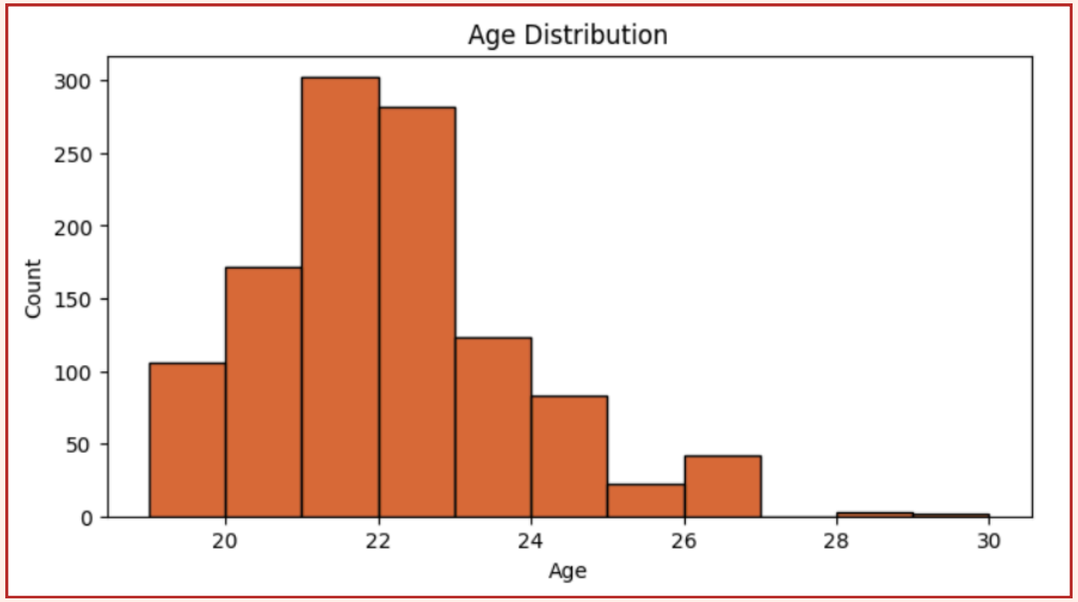



```
plt.figure(figsize=(10,5))
placement_stream = df.groupby('Stream')['PlacedOrNot'].mean().sort_values(ascending=False).reset_index()
sns.barplot(x='PlacedOrNot', y='Stream', data=placement_stream)
plt.title('Placement Rate by Stream')
plt.xlabel('Placement Rate')
plt.ylabel('Stream')
plt.show()
```

EDA HIGHLIGHTS

Age Distribution:

- Most students are aged 21–22.



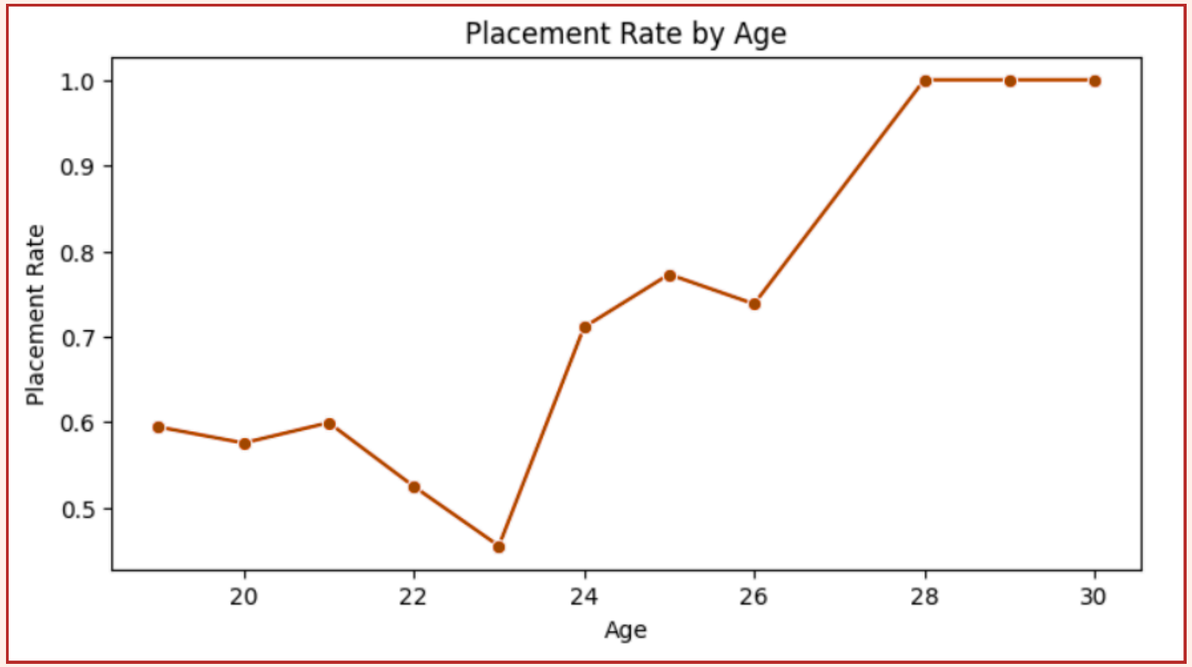

```
plt.figure(figsize=(8,4))
sns.histplot(df['Age'], bins=range(df['Age'].min(), df['Age'].max()+1), kde=False)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

EDA HIGHLIGHTS

Placement Rate by Age:

- Placement rates are consistent across age groups.

```
plt.figure(figsize=(8,4))
placement_age = df.groupby('Age')['PlacedOrNot'].mean().reset_index()
sns.lineplot(x='Age', y='PlacedOrNot', data=placement_age, marker='o')
plt.title('Placement Rate by Age')
plt.ylabel('Placement Rate')
plt.xlabel('Age')
plt.show()
```

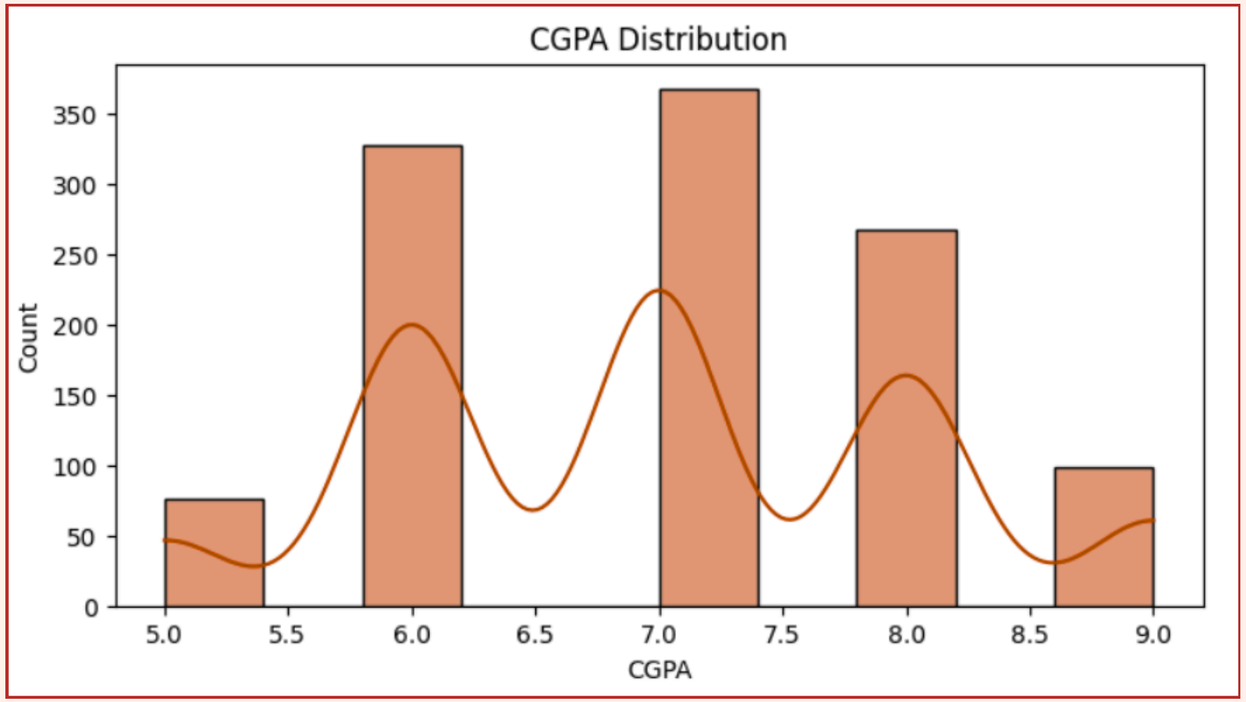


EDA HIGHLIGHTS

CGPA Distribution:

- CGPA ranges from 5 to 9, with most students scoring 6–8.

```
plt.figure(figsize=(8,4))
sns.histplot(df['CGPA'], bins=10, kde=True)
plt.title('CGPA Distribution')
plt.xlabel('CGPA')
plt.ylabel('Count')
plt.show()
```

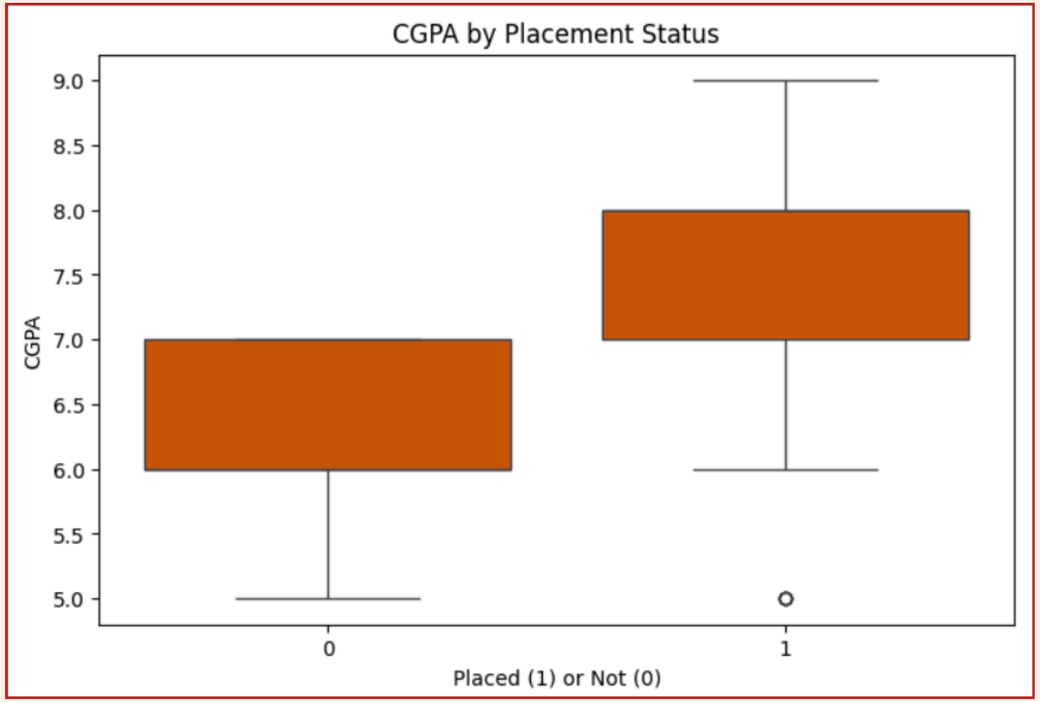


EDA HIGHLIGHTS

CGPA vs Placement:

- Higher CGPA generally correlates with higher placement rates.

```
plt.figure(figsize=(8,5))
sns.boxplot(x='PlacedOrNot', y='CGPA', data=df)
plt.title('CGPA by Placement Status')
plt.xlabel('Placed (1) or Not (0)')
plt.ylabel('CGPA')
plt.show()
```

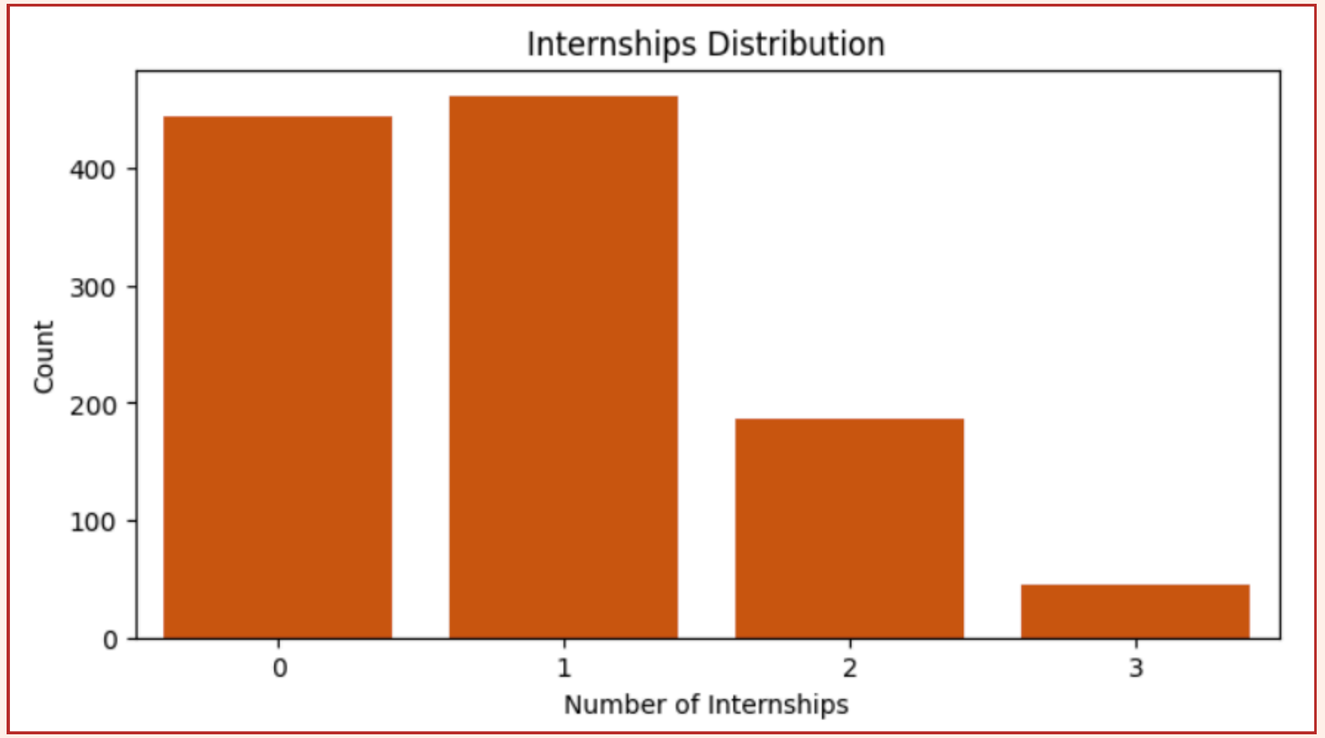


EDA HIGHLIGHTS

Internships Distribution:

- Most students have 0–1 internship; few have 2 or more.

```
plt.figure(figsize=(8,4))
sns.countplot(x='Internships', data=df)
plt.title('Internships Distribution')
plt.xlabel('Number of Internships')
plt.ylabel('Count')
plt.show()
```

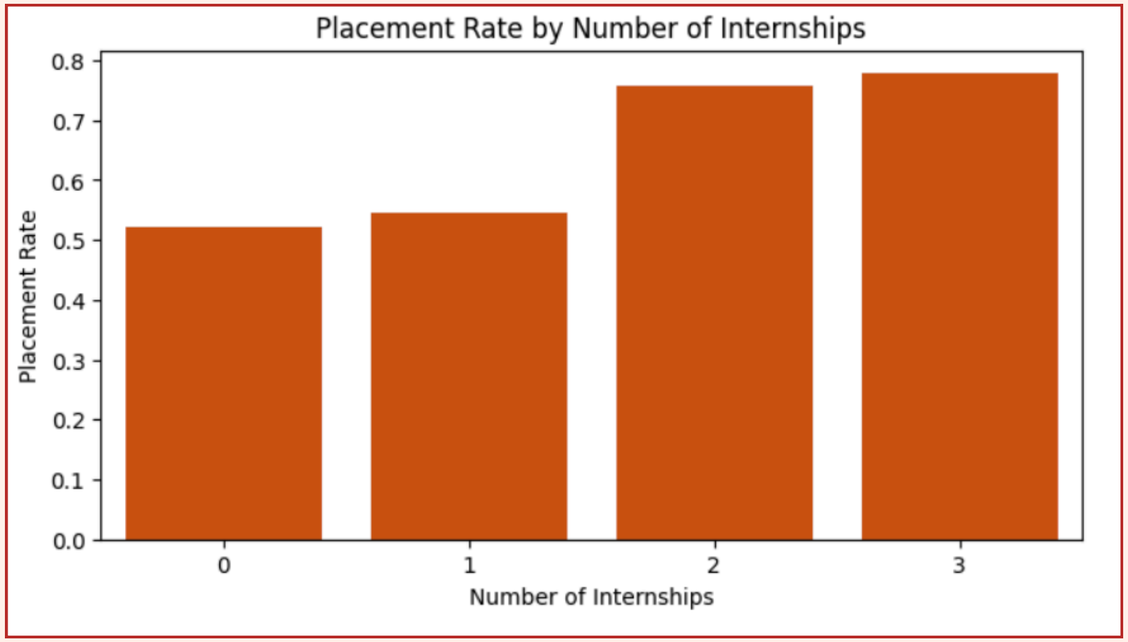


EDA HIGHLIGHTS

Placement Rate by Internships:

- Placement probability increases with the number of internships.

```
plt.figure(figsize=(8,4))
placement_internships = df.groupby('Internships')['PlacedOrNot'].mean().reset_index()
sns.barplot(x='Internships', y='PlacedOrNot', data=placement_internships)
plt.title('Placement Rate by Number of Internships')
plt.ylabel('Placement Rate')
plt.xlabel('Number of Internships')
plt.show()
```

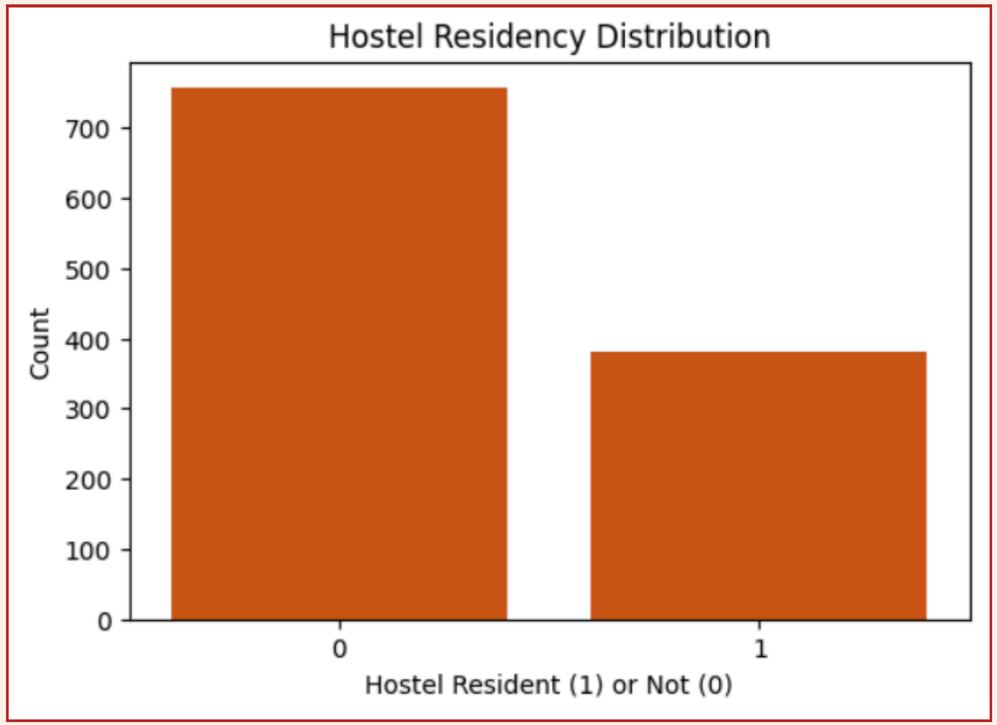


EDA HIGHLIGHTS

Hostel Residency:

- About 27% of students are hostel residents.

```
plt.figure(figsize=(6,4))
sns.countplot(x='Hostel', data=df)
plt.title('Hostel Residency Distribution')
plt.xlabel('Hostel Resident (1) or Not (0)')
plt.ylabel('Count')
plt.show()
```

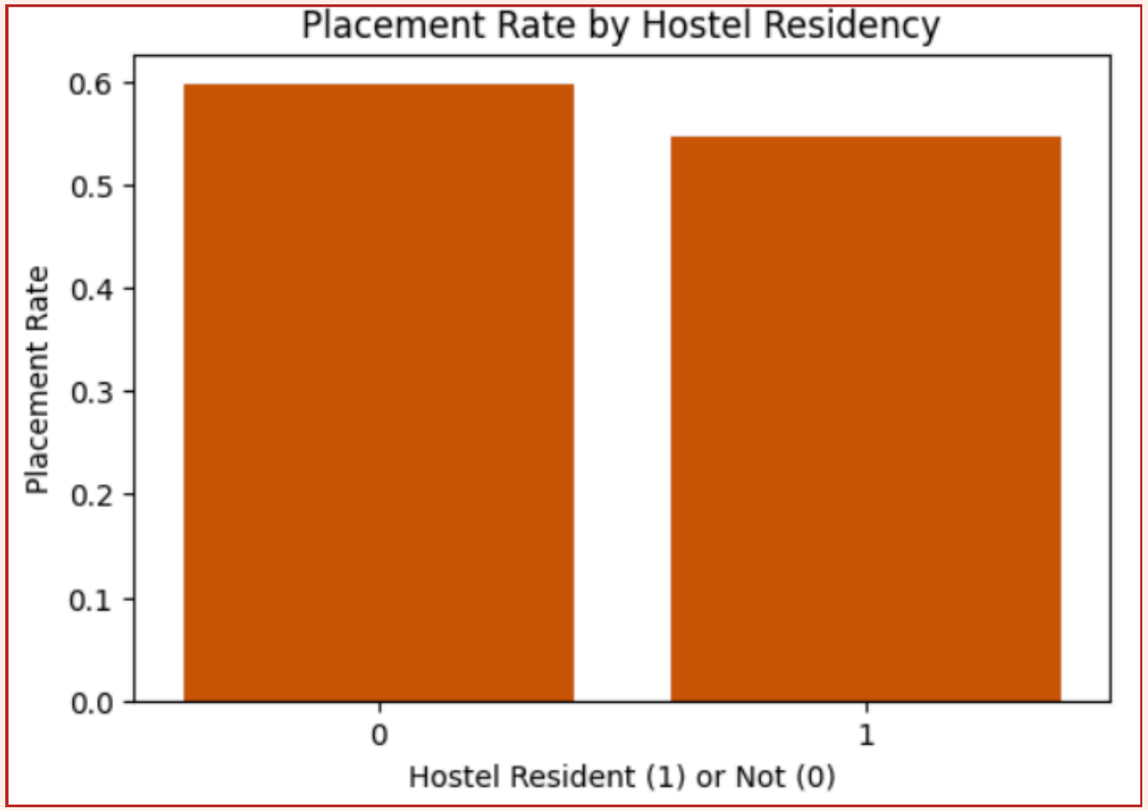


EDA HIGHLIGHTS

Placement Rate by Hostel:

- Hostel residency shows a slight positive correlation with placement.

```
plt.figure(figsize=(6,4))
placement_hostel = df.groupby('Hostel')['PlacedOrNot'].mean().reset_index()
sns.barplot(x='Hostel', y='PlacedOrNot', data=placement_hostel)
plt.title('Placement Rate by Hostel Residency')
plt.ylabel('Placement Rate')
plt.xlabel('Hostel Resident (1) or Not (0)')
plt.show()
```

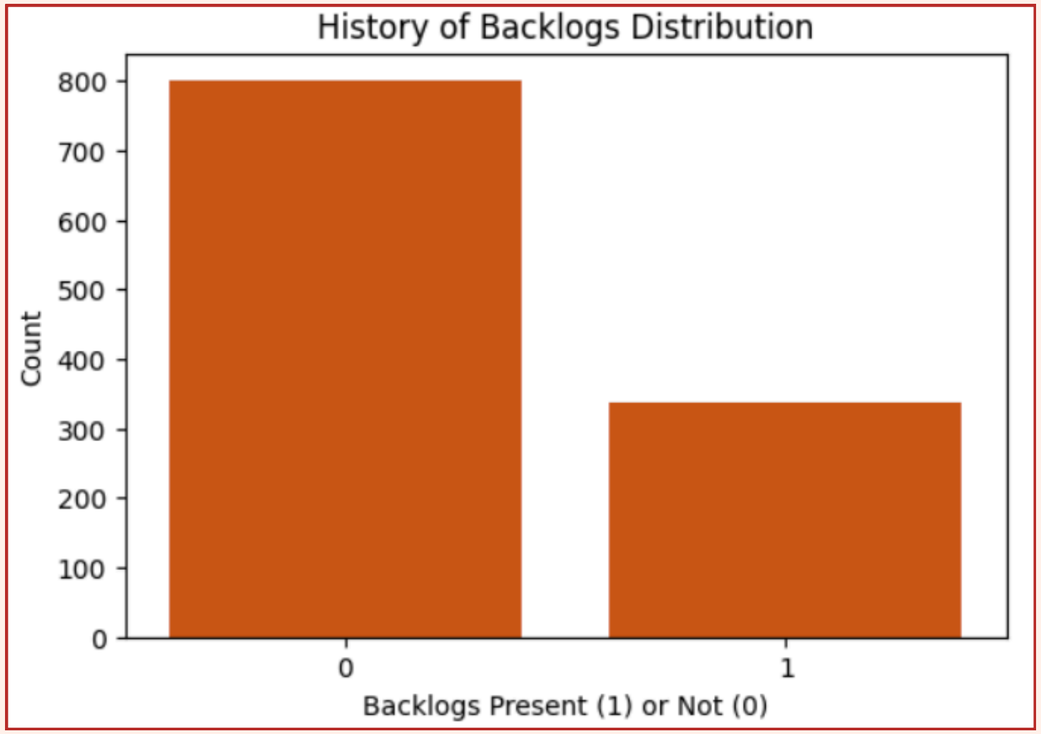


EDA HIGHLIGHTS

History of Backlogs:

- ~19% of students have a history of backlogs.

```
plt.figure(figsize=(6,4))
sns.countplot(x='HistoryOfBacklogs', data=df)
plt.title('History of Backlogs Distribution')
plt.xlabel('Backlogs Present (1) or Not (0)')
plt.ylabel('Count')
plt.show()
```

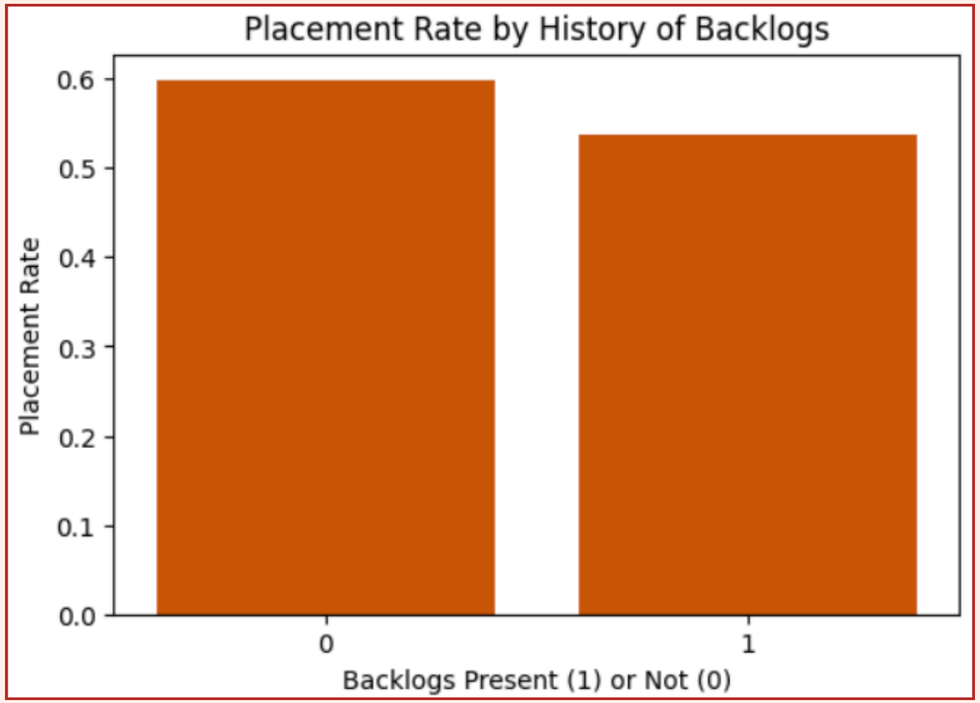


EDA HIGHLIGHTS

Placement Rate by Backlogs:

- Students with backlogs have significantly lower placement rates.

```
plt.figure(figsize=(6,4))
placement_backlogs = df.groupby('HistoryOfBacklogs')['PlacedOrNot'].mean().reset_index()
sns.barplot(x='HistoryOfBacklogs', y='PlacedOrNot', data=placement_backlogs)
plt.title('Placement Rate by History of Backlogs')
plt.ylabel('Placement Rate')
plt.xlabel('Backlogs Present (1) or Not (0)')
plt.show()
```



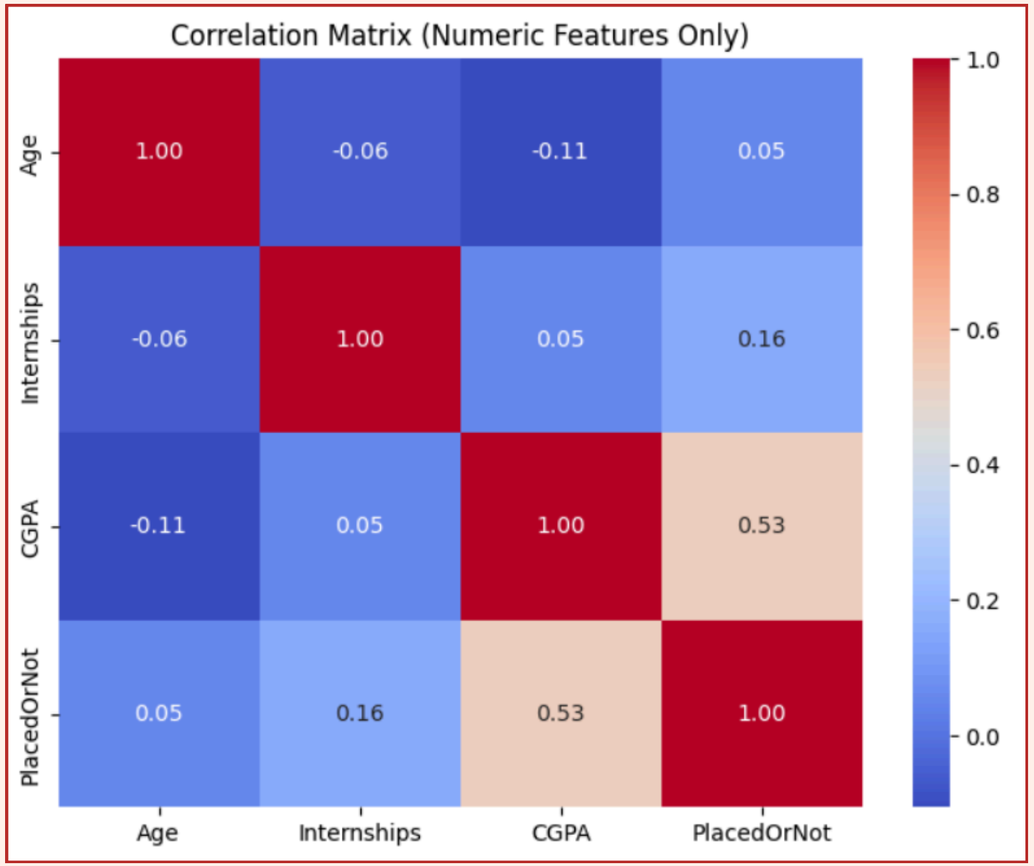
EDA HIGHLIGHTS

Correlation Analysis:

- Strongest predictors: CGPA, Internships, Backlogs.

```
numeric_df = df.select_dtypes(include=['number'])
corr = numeric_df.corr()

plt.figure(figsize=(8,6))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix (Numeric Features Only)')
plt.show()
```

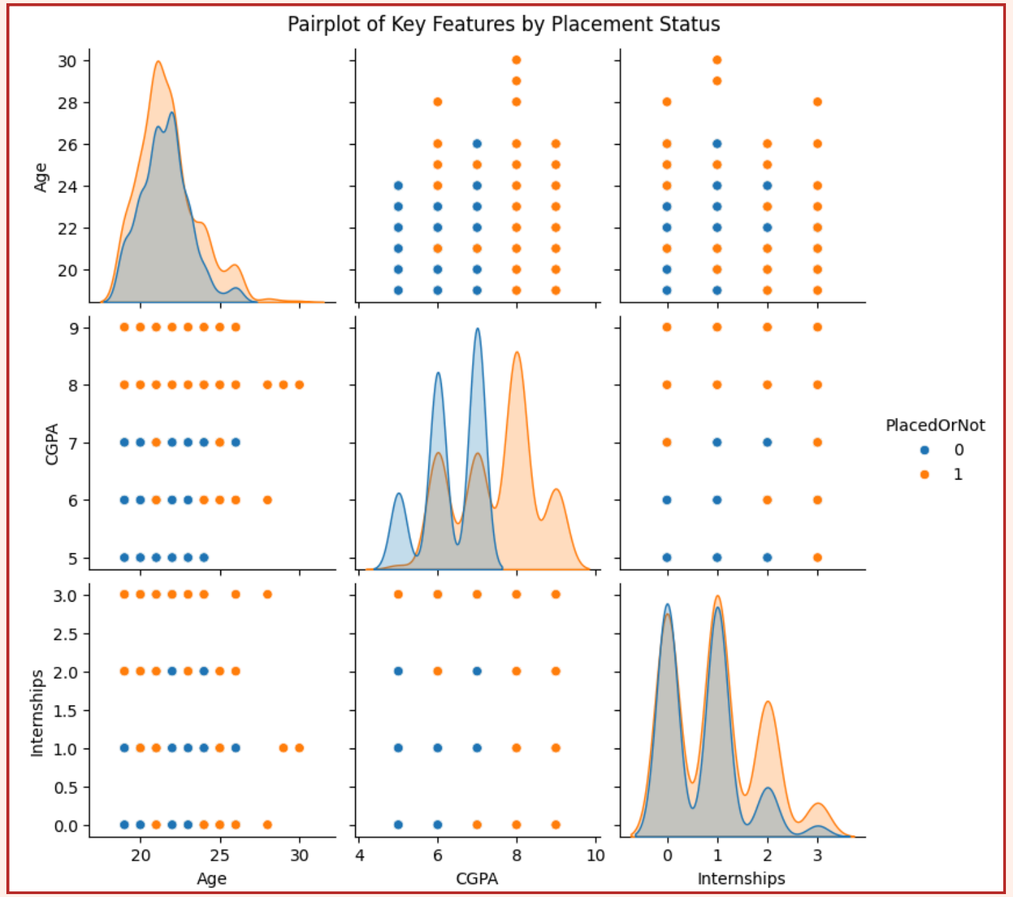


EDA HIGHLIGHTS

Key Features by Placement:

- Placed students have higher CGPA and more internships.
- Age has little impact on placement.
- CGPA and internships are strong placement indicators.

```
sns.pairplot(df, hue='PlacedOrNot', vars=['Age', 'CGPA', 'Internships'])
plt.suptitle('Pairplot of Key Features by Placement Status', y=1.02)
plt.show()
```



CONCLUSION

- Key Takeaways:
- Academic performance (CGPA) and practical experience (internships) are critical for placement.
- Backlogs negatively impact placement chances.
- Stream of study and hostel residency have moderate influence.

- Recommendations:
- Encourage internships and maintain good academic standing.
- Provide additional support for students with backlogs.

THANK YOU

bhaskarpal.official@gmail.com