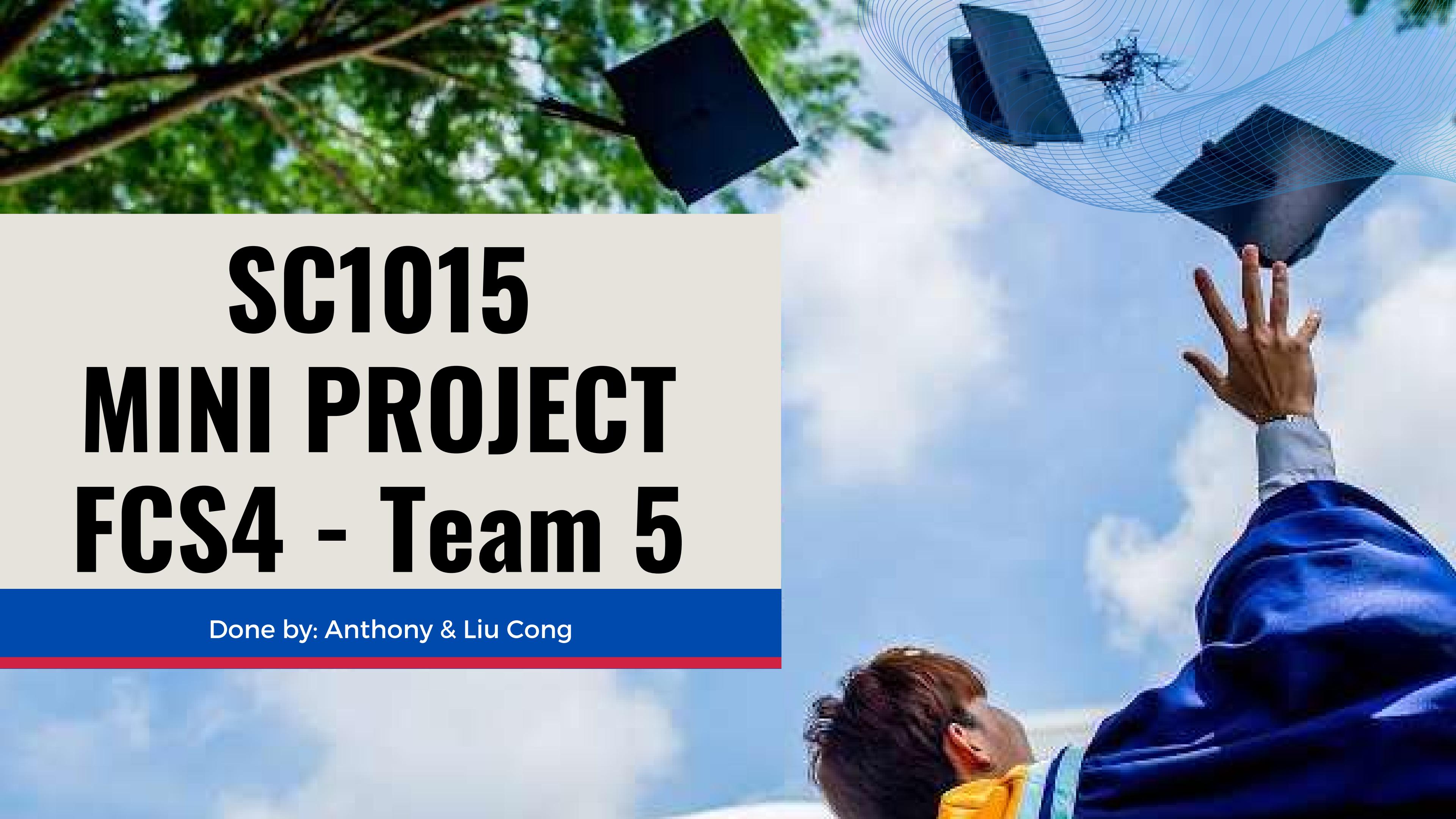


# **SC1015**

# **MINI PROJECT**

# **FCS4 - Team 5**

Done by: Anthony & Liu Cong



# EMERGENCE OF AI





WHAT  
DO YOU  
THINK?

# ARE AI REPLACING US?



# OUR PROJECT

## THEREFORE...

We aim to predict the employability of fresh graduates with Bachelor's, from Computer Science and IT related background, in the US

# 01 FINDING DATASETS

# STEP 1: Finding Datasets



- The 1st dataset: 700 fresh graduates with their placement status
- The 2nd dataset: The ranking of universities from across 114 locations

# STEP 1: Finding Datasets

jobData												
	<b>id</b>	<b>name</b>	<b>gender</b>	<b>age</b>	<b>degree</b>	<b>stream</b>	<b>college_name</b>	<b>placement_status</b>	<b>salary</b>	<b>gpa</b>	<b>years_of_experience</b>	
0	1	John Doe	Male	25	Bachelor's	Computer Science	Harvard University	Placed	60000	3.7	2.0	
1	2	Jane Smith	Female	24	Bachelor's	Electrical Engineering	Massachusetts Institute of Technology	Placed	65000	3.6	1.0	
2	3	Michael Johnson	Male	26	Bachelor's	Mechanical Engineering	Stanford University	Placed	58000	3.8	3.0	
3	4	Emily Davis	Female	23	Bachelor's	Information Technology	Yale University	Not Placed	0	3.5	2.0	
4	5	David Brown	Male	24	Bachelor's	Computer Science	Princeton University	Placed	62000	3.9	2.0	
...	...	...	...	...	...	...	...	...	...	...	...	
695	696	Lucas Taylor	Male	23	Bachelor's	Computer Science	University of Washington	Placed	67000	3.8	3.0	
696	697	Emma Martinez	Female	26	Bachelor's	Electronics and Communication	University of California--Berkeley	Placed	66000	3.9	3.0	
697	698	Aiden Davis	Male	24	Bachelor's	Computer Science	University of Illinois--Urbana-Champaign	Placed	65000	3.8	3.0	
698	699	Mia Wilson	Female	23	Bachelor's	Electrical Engineering	University of Colorado--Boulder	Placed	66000	3.7	2.0	
699	700	Jack Garcia	Male	26	Bachelor's	Information Technology	University of North Carolina--Chapel Hill	Not Placed	0	3.6	1.0	
700 rows × 11 columns												

The 1st dataset

# STEP 1: Finding Datasets

jobData.describe()					
	<b>id</b>	<b>age</b>	<b>salary</b>	<b>gpa</b>	<b>years_of_experience</b>
<b>count</b>	700.000000	700.000000	700.000000	700.000000	699.000000
<b>mean</b>	350.500000	24.411429	52474.285714	3.750429	2.177396
<b>std</b>	202.21688	1.164268	25160.331005	0.121212	0.779393
<b>min</b>	1.00000	23.000000	0.000000	3.400000	1.000000
<b>25%</b>	175.75000	23.000000	61000.000000	3.700000	2.000000
<b>50%</b>	350.50000	24.000000	64000.000000	3.800000	2.000000
<b>75%</b>	525.25000	26.000000	66000.000000	3.900000	3.000000
<b>max</b>	700.000000	26.000000	68000.000000	3.900000	3.000000

The 1st dataset

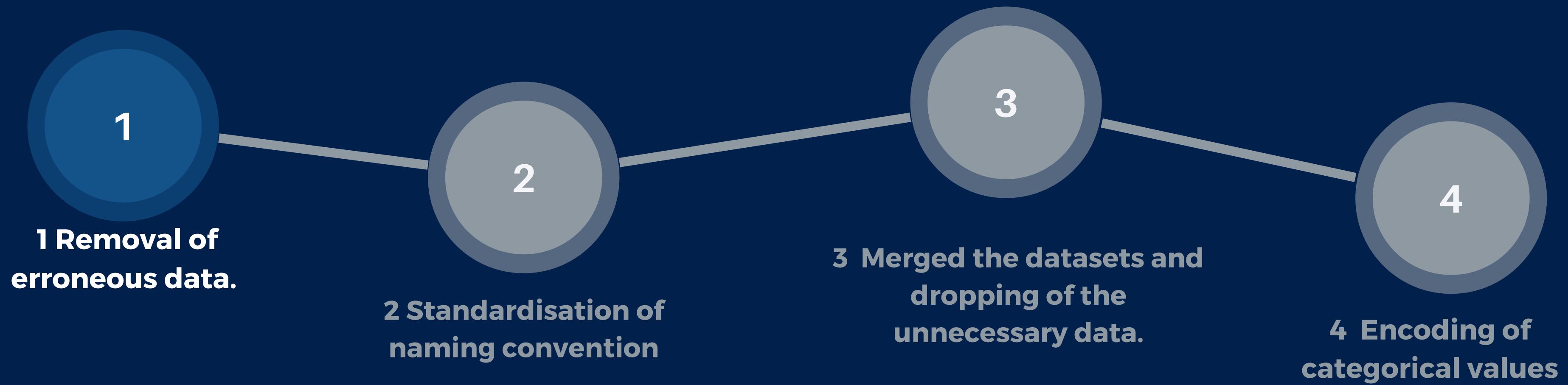
# STEP 1: Finding Datasets

uniRankings.head()																
	2024 RANK	2023 RANK	Institution Name	Country Code	Country	SIZE	FOCUS	RES.	AGE	STATUS	...	International Faculty Rank	International Students Score	International Students Rank	International Research Network Score	Inte...
0	rank display	rank display2	institution	location code	location	size	focus	research	age band	status	...	ifr rank	isr score	isr rank	irn score	...
1	1	1	Massachusetts Institute of Technology (MIT)	US	United States	M	CO	VH	5	B	...	56	88.2	128	94.3	...
2	2	2	University of Cambridge	UK	United Kingdom	L	FC	VH	5	A	...	64	95.8	85	99.9	...
3	3	4	University of Oxford	UK	United Kingdom	L	FC	VH	5	A	...	110	98.2	60	100.0	...
4	4	5	Harvard University	US	United States	L	FC	VH	5	B	...	210	66.8	223	100.0	...

The 2nd dataset

# 02 PREPARATION AND CLEANING OF DATA

# STEP 2: Preparation



# STEP 2: Preparation

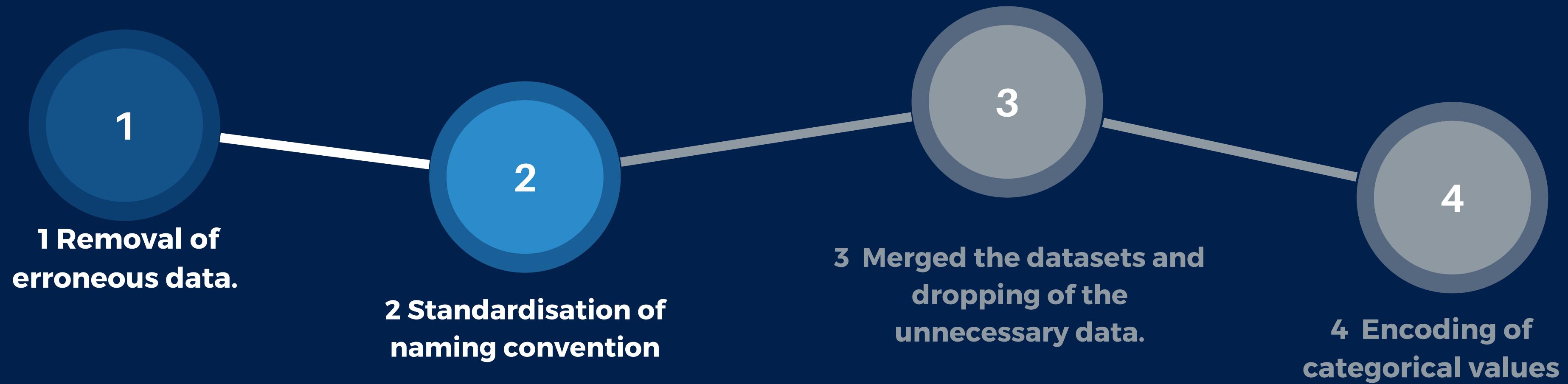
```
jobData.isnull().sum()  
  
id          0  
name        0  
gender      0  
age          0  
degree       0  
stream       0  
college_name 0  
placement_status 0  
salary        0  
gpa           0  
years_of_experience 1  
dtype: int64
```

```
jobData[jobData['years_of_experience'].isnull()] # seeing the null row  
jobData=jobData.dropna() # dropping the null row  
jobData
```

0	1	John Doe	Male	25	Bachelor's	Computer Science	Harvard University	Placed	60000	3.7	2.0
1	2	Jane Smith	Female	24	Bachelor's	Electrical Engineering	Massachusetts Institute of Technology	Placed	65000	3.6	1.0
2	3	Michael Johnson	Male	26	Bachelor's	Mechanical Engineering	Stanford University	Placed	58000	3.8	3.0
3	4	Emily Davis	Female	23	Bachelor's	Information Technology	Yale University	Not Placed	0	3.5	2.0
4	5	David Brown	Male	24	Bachelor's	Computer Science	Princeton University	Placed	62000	3.9	2.0
...	...	...	...	...	...	...	...	...	...	...	...
695	696	Lucas Taylor	Male	23	Bachelor's	Computer Science	University of Washington	Placed	67000	3.8	3.0
696	697	Emma Martinez	Female	26	Bachelor's	Electronics and Communication	University of California--Berkeley	Placed	66000	3.9	3.0
697	698	Aiden Davis	Male	24	Bachelor's	Computer Science	University of Illinois--Urbana-Champaign	Placed	65000	3.8	3.0
698	699	Mia Wilson	Female	23	Bachelor's	Electrical Engineering	University of Colorado--Boulder	Placed	66000	3.7	2.0
699	700	Jack Garcia	Male	26	Bachelor's	Information Technology	University of North Carolina--Chapel Hill	Not Placed	0	3.6	1.0

699 rows x 11 columns

# STEP 2: Preparation & Cleaning



# STEP 2: Preparation & Cleaning

**Dataset 1:**

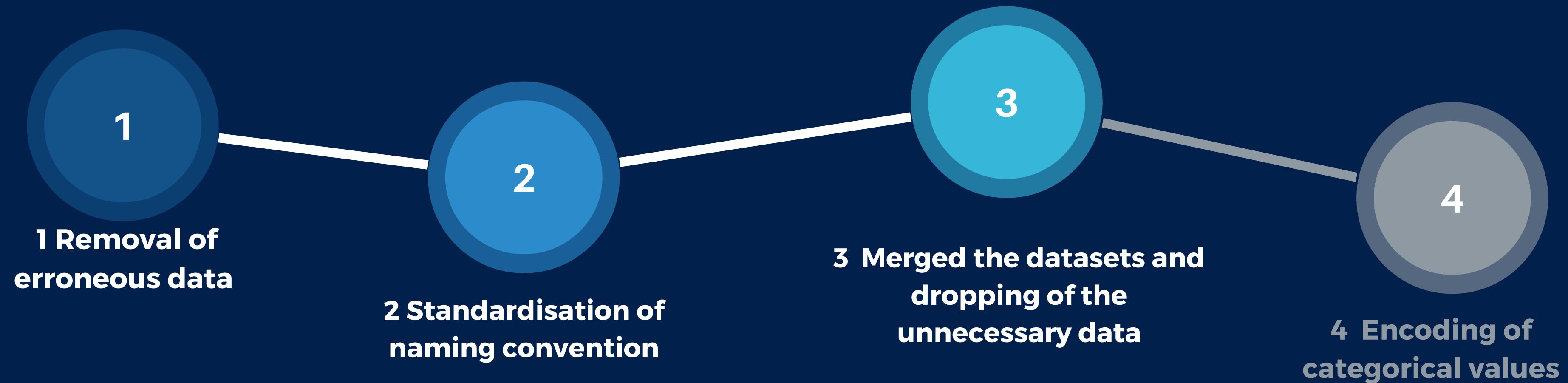
- Massachusetts Institute of Technology
- University of California--Los Angeles
- University of Illinois--Urbana-Champaign



**Dataset 2:**

- Massachusetts Institute of Technology (MIT)
- University of California, Los Angeles (UCLA)
- University of Illinois at Urbana-Champaign

# STEP 2: Preparation & Cleaning



# STEP 2: Preparation & Cleaning

	<b>id</b>	<b>name</b>	<b>gender</b>	<b>age</b>	<b>degree</b>	<b>stream</b>	<b>college_name</b>	<b>placement_status</b>	<b>salary</b>	<b>gpa</b>	<b>years_of_experience</b>
0	1	John Doe	Male	25	Bachelor's	Computer Science	Harvard University	Placed	60000	3.7	2.0
1	2	Jane Smith	Female	24	Bachelor's	Electrical Engineering	Massachusetts Institute of Technology	Placed	65000	3.6	1.0
2	3	Michael Johnson	Male	26	Bachelor's	Mechanical Engineering	Stanford University	Placed	58000	3.8	3.0
3	4	Emily Davis	Female	23	Bachelor's	Information Technology	Yale University	Not Placed	0	3.5	2.0
4	5	David Brown	Male	24	Bachelor's	Computer Science	Princeton University	Placed	62000	3.9	2.0

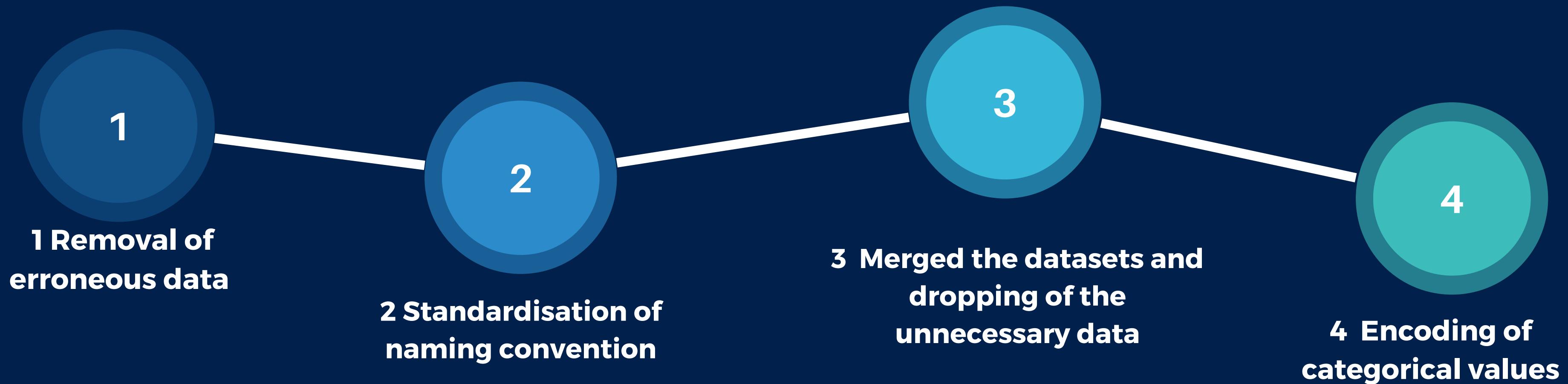
	<b>id</b>	<b>name</b>	<b>gender</b>	<b>age</b>	<b>degree</b>	<b>stream</b>	<b>college_name</b>	<b>placement_status</b>	<b>salary</b>	<b>gpa</b>	<b>years_of_experience</b>	<b>ranking</b>
	1	John Doe	Male	25	Bachelor's	Computer Science	Harvard University	Placed	60000	3.7	2.0	4
	2	Jane Smith	Female	24	Bachelor's	Electrical Engineering	Massachusetts Institute of Technology	Placed	65000	3.6	1.0	1
	3	Michael Johnson	Male	26	Bachelor's	Mechanical Engineering	Stanford University	Placed	58000	3.8	3.0	5
	4	Emily Davis	Female	23	Bachelor's	Information Technology	Yale University	Not Placed	0	3.5	2.0	16
	5	David Brown	Male	24	Bachelor's	Computer Science	Princeton University	Placed	62000	3.9	2.0	17

# STEP 2: Preparation & Cleaning

id	name	gender	age	degree	stream	college_name	placement_status	salary	gpa	years_of_experience	ranking
1	John Doe	Male	25	Bachelor's	Computer Science	Harvard University	Placed	60000	3.7	2.0	4
2	Jane Smith	Female	24	Bachelor's	Electrical Engineering	Massachusetts Institute of Technology	Placed	65000	3.6	1.0	1
3	Michael Johnson	Male	26	Bachelor's	Mechanical Engineering	Stanford University	Placed	58000	3.8	3.0	5
4	Emily Davis	Female	23	Bachelor's	Information Technology	Yale University	Not Placed	0	3.5	2.0	16
5	David Brown	Male	24	Bachelor's	Computer Science	Princeton University	Placed	62000	3.9	2.0	17

	id	gender	age	stream	college_name	placement_status	salary	gpa	years_of_experience	ranking
0	1	Male	25	Computer Science	Harvard University	Placed	60000	3.7	2.0	4
3	4	Female	23	Information Technology	Yale University	Not Placed	0	3.5	2.0	16
4	5	Male	24	Computer Science	Princeton University	Placed	62000	3.9	2.0	17
6	7	Male	26	Information Technology	California Institute of Technology	Placed	59000	3.8	3.0	15
7	8	Female	24	Computer Science	University of Chicago	Not Placed	0	3.6	2.0	11

# STEP 2: Preparation & Cleaning



# STEP 2: Preparation & Cleaning

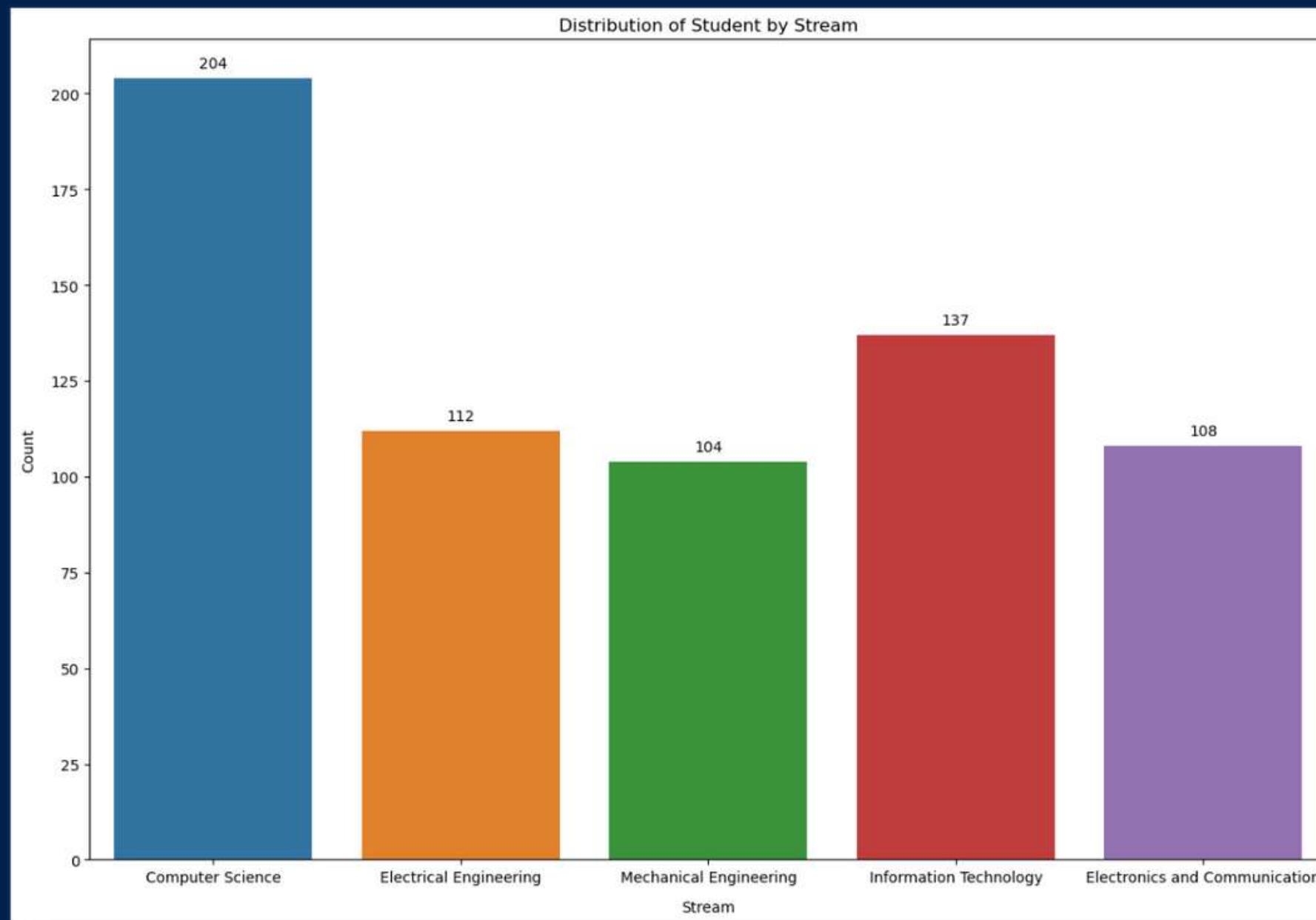
	<b>id</b>	<b>gender</b>	<b>age</b>	<b>stream</b>	<b>college_name</b>	<b>placement_status</b>	<b>salary</b>	<b>gpa</b>	<b>years_of_experience</b>	<b>ranking</b>
0	1	1	25	0	4	1	60000	3.7	2.0	4
3	4	0	23	1	33	0	0	3.5	2.0	16
4	5	1	24	0	5	1	62000	3.9	2.0	17
6	7	1	26	1	1	1	59000	3.8	3.0	15
7	8	0	24	0	12	0	0	3.6	2.0	11
...	...	...	...	...	...	...	...	...	...	...
691	692	1	25	0	29	1	63000	3.6	1.0	520
693	694	1	26	1	21	1	67000	3.9	3.0	33
695	696	1	23	0	31	1	67000	3.8	3.0	63
697	698	1	24	0	18	1	65000	3.8	3.0	64
699	700	1	26	1	23	0	0	3.6	1.0	132

341 rows × 10 columns

# 03 INSIGHTS GENERATION

# Basic Insights

## Distribution of Student by Stream

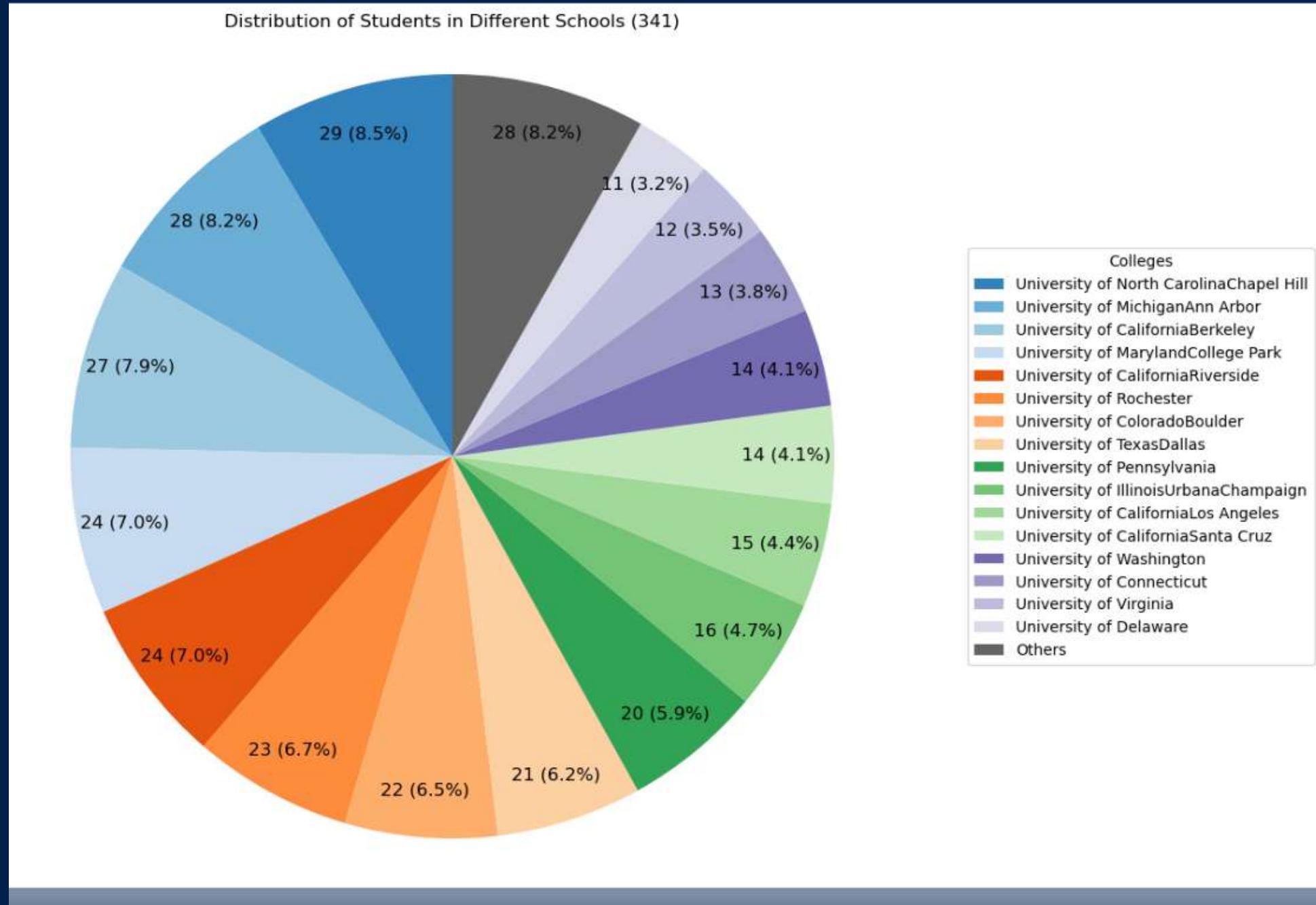


Primarily focusing on the data from these 2 streams:

- Computer Science: 204 Students
- Information Technology: 137 Students

# Basic Insights

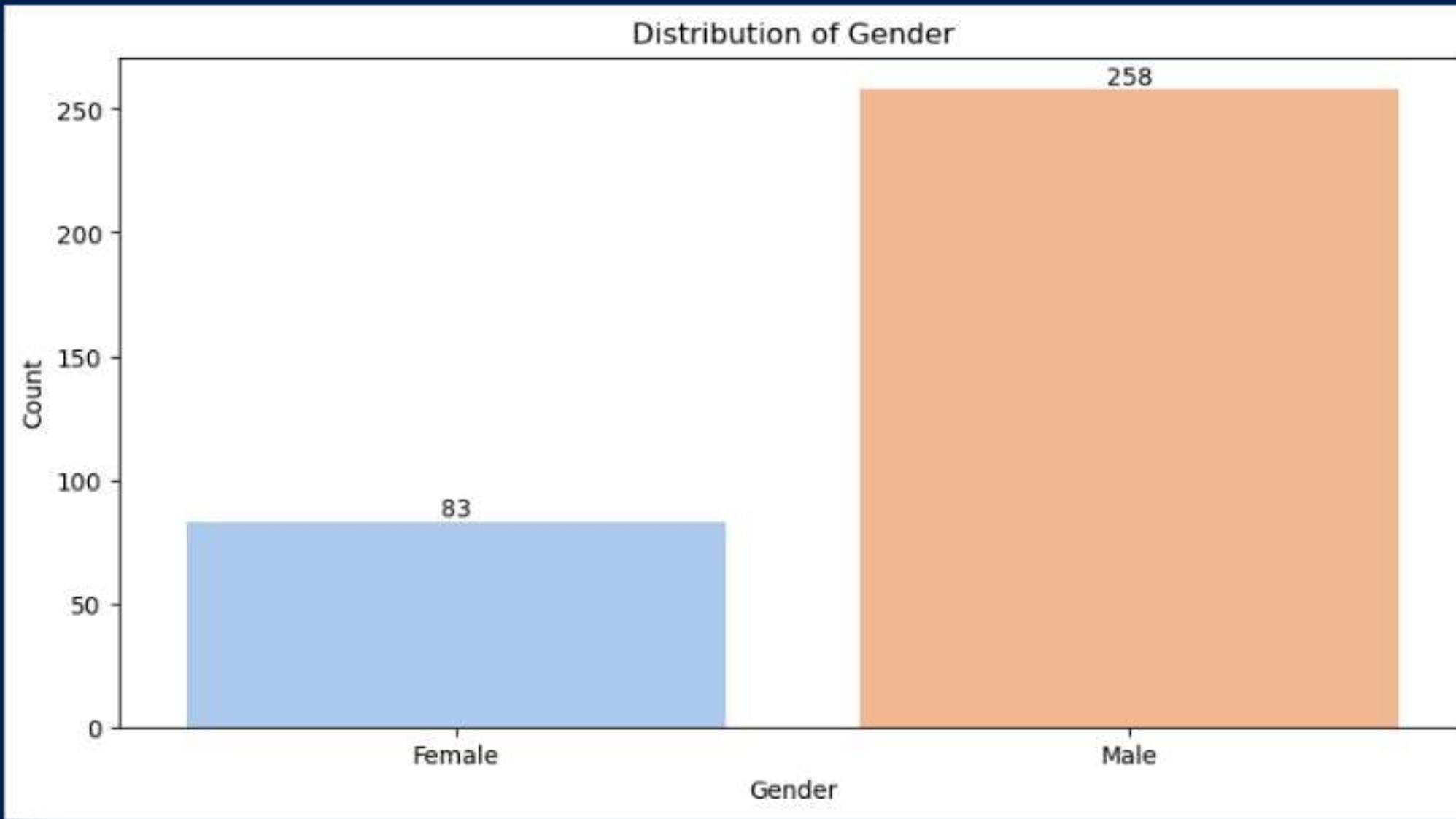
## Distribution by Colleges



For simplicity, colleges that fell below the threshold of 3% are grouped together for easy visualisation

# Basic Insights

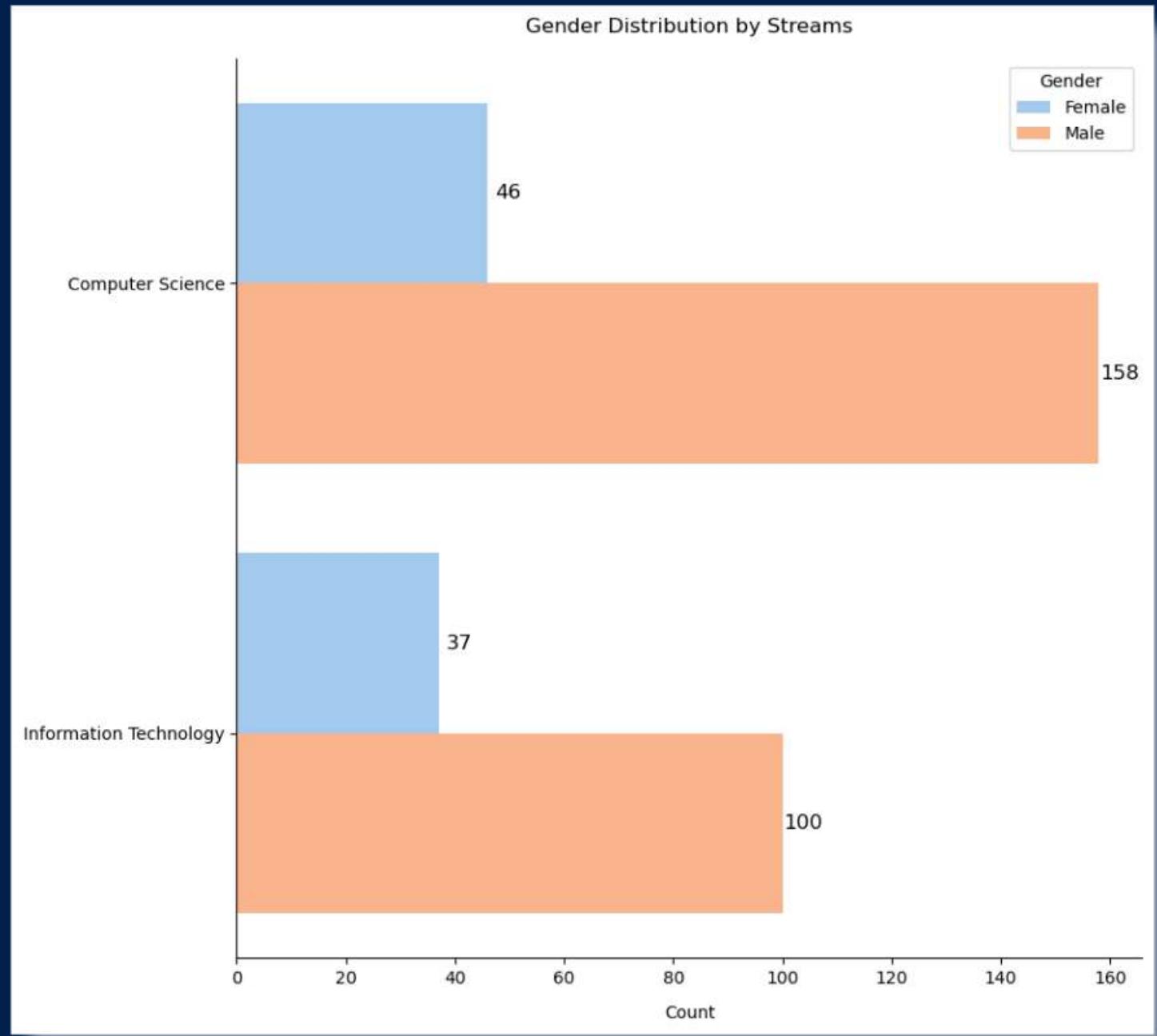
## Distribution by Gender (Dataset)



- Out of 341 graduates:
  - 83 identified as female
  - 258 identified as male
- Discovered a gender imbalance in the dataset

# Basic Insights

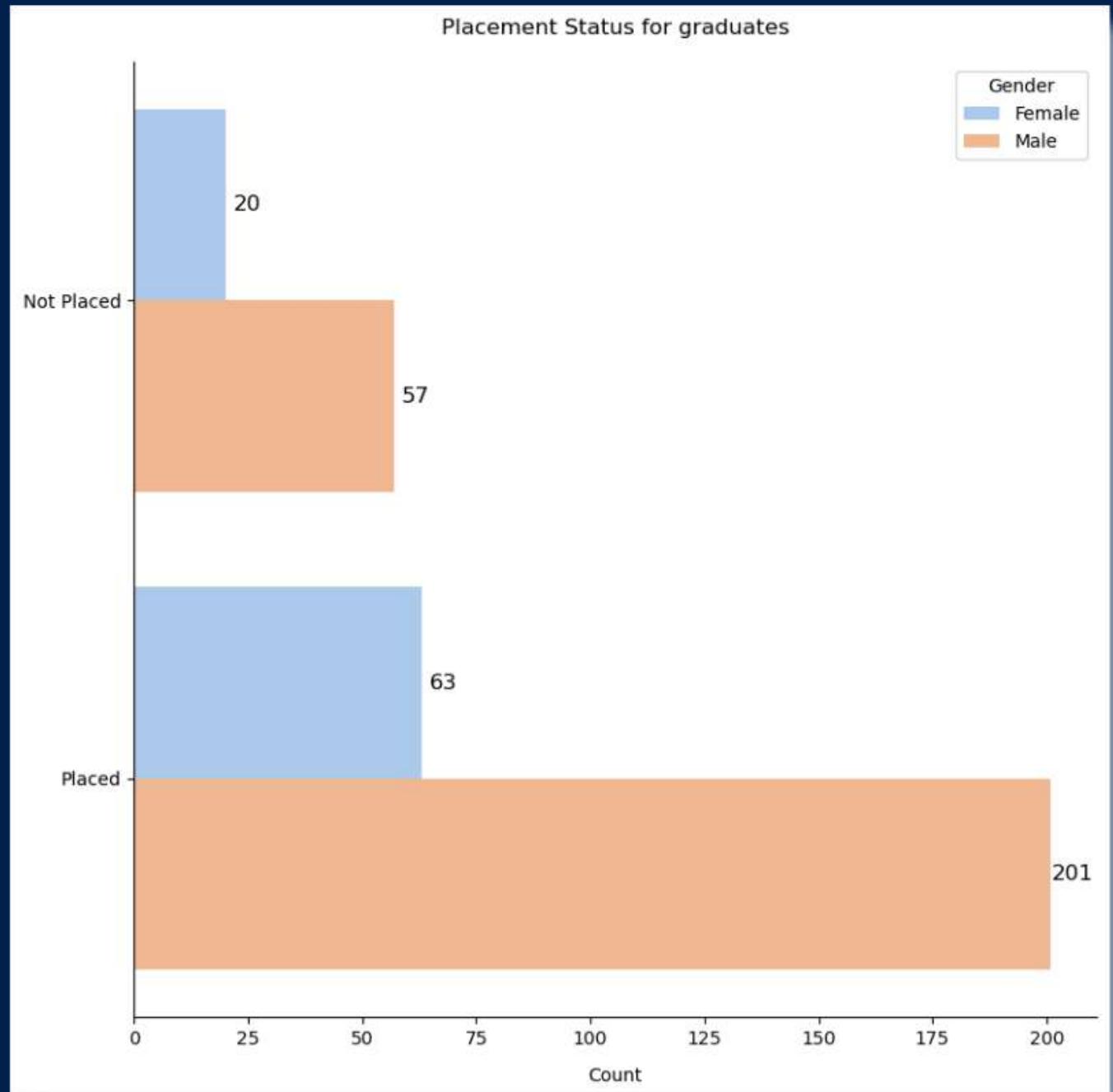
## Distribution by Gender (Streams)



- **46 Female in Computer Science, 37 Female in Information Technology**
- **158 Male in Computer Science, 100 Male in Information Technology**

# Basic Insights

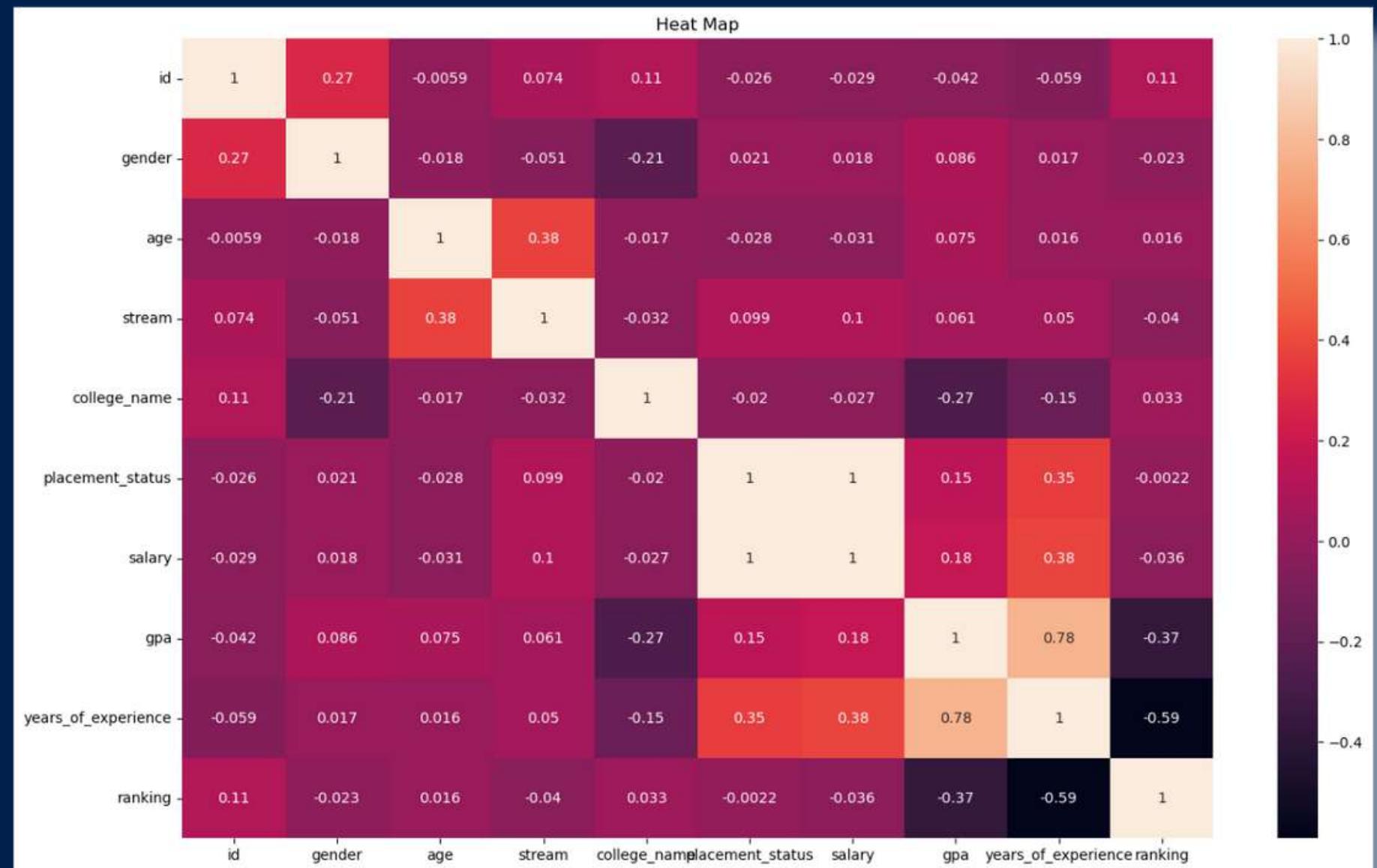
## Distribution by Gender (Placement Status)



- 24% of Female graduates do not have a placement
- 22% of Male graduates do not have a placement

# Exploratory Analysis

## Heat Map



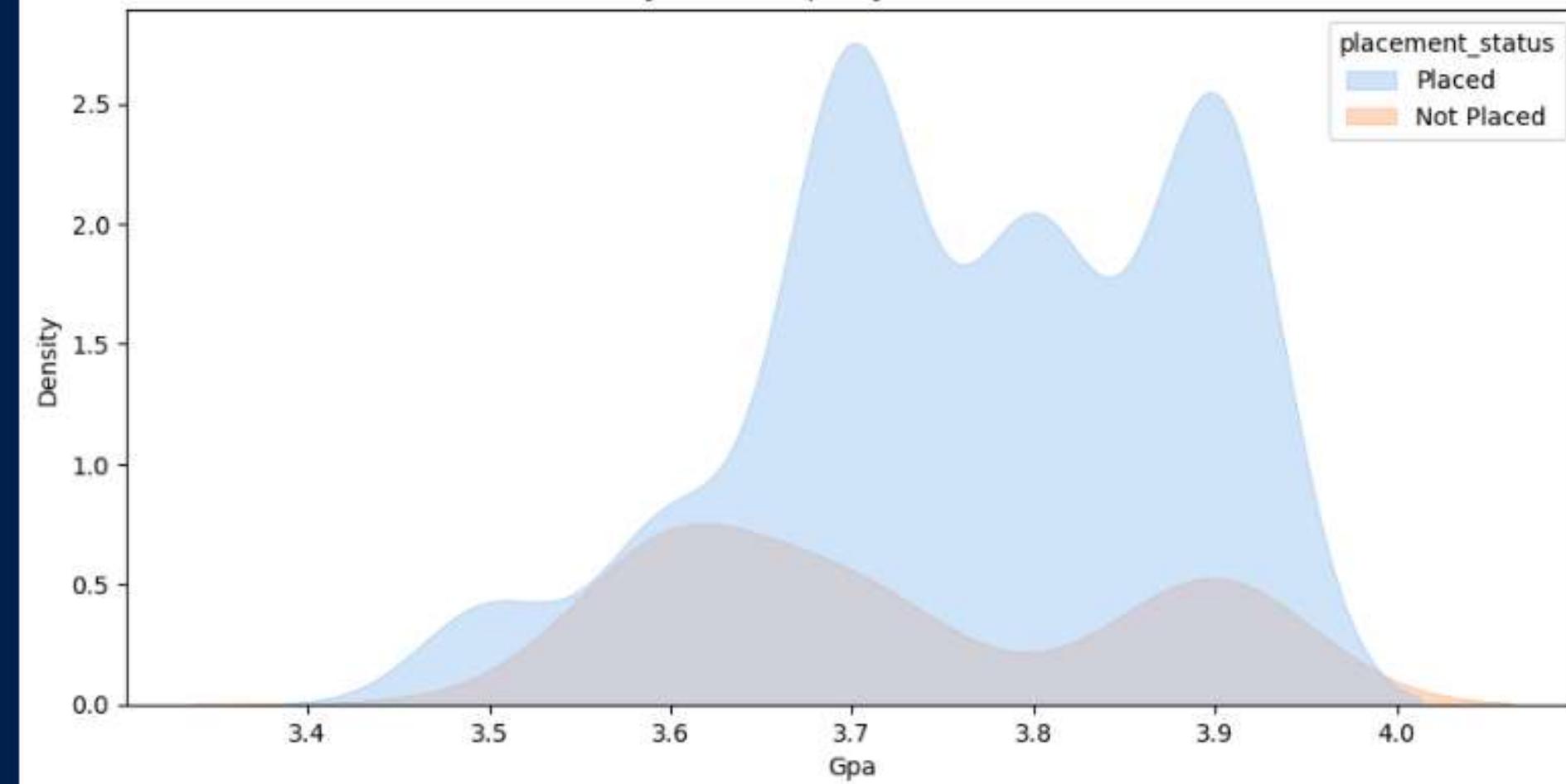
## Correlation to Placement Status

- Gender, Stream, GPA (0.15) and Years of Experience (0.38) are positively correlated
- College Name and Age are negatively correlated

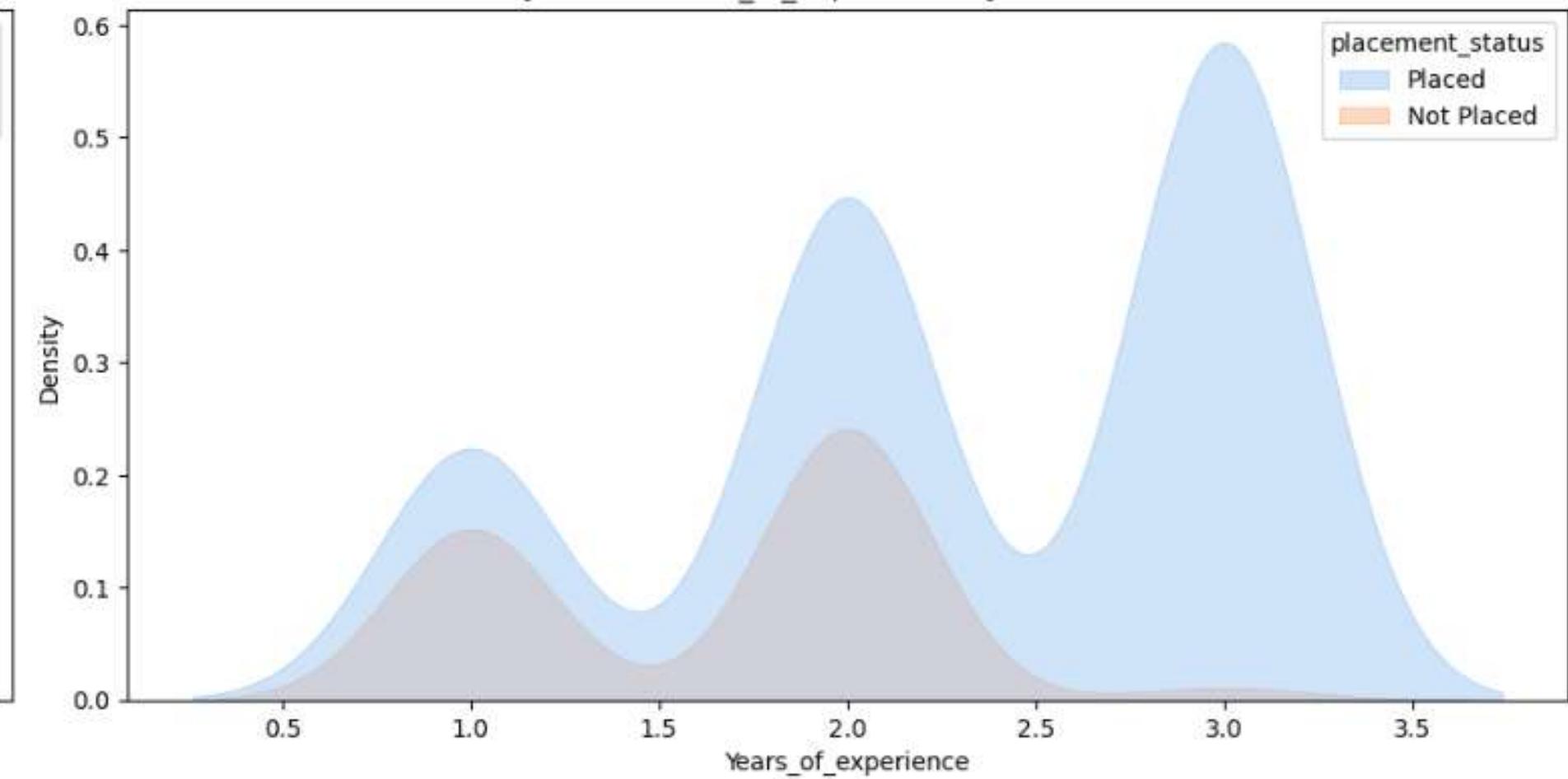
# Exploratory Analysis

## KDE

Density Plot for Gpa by Placement Status



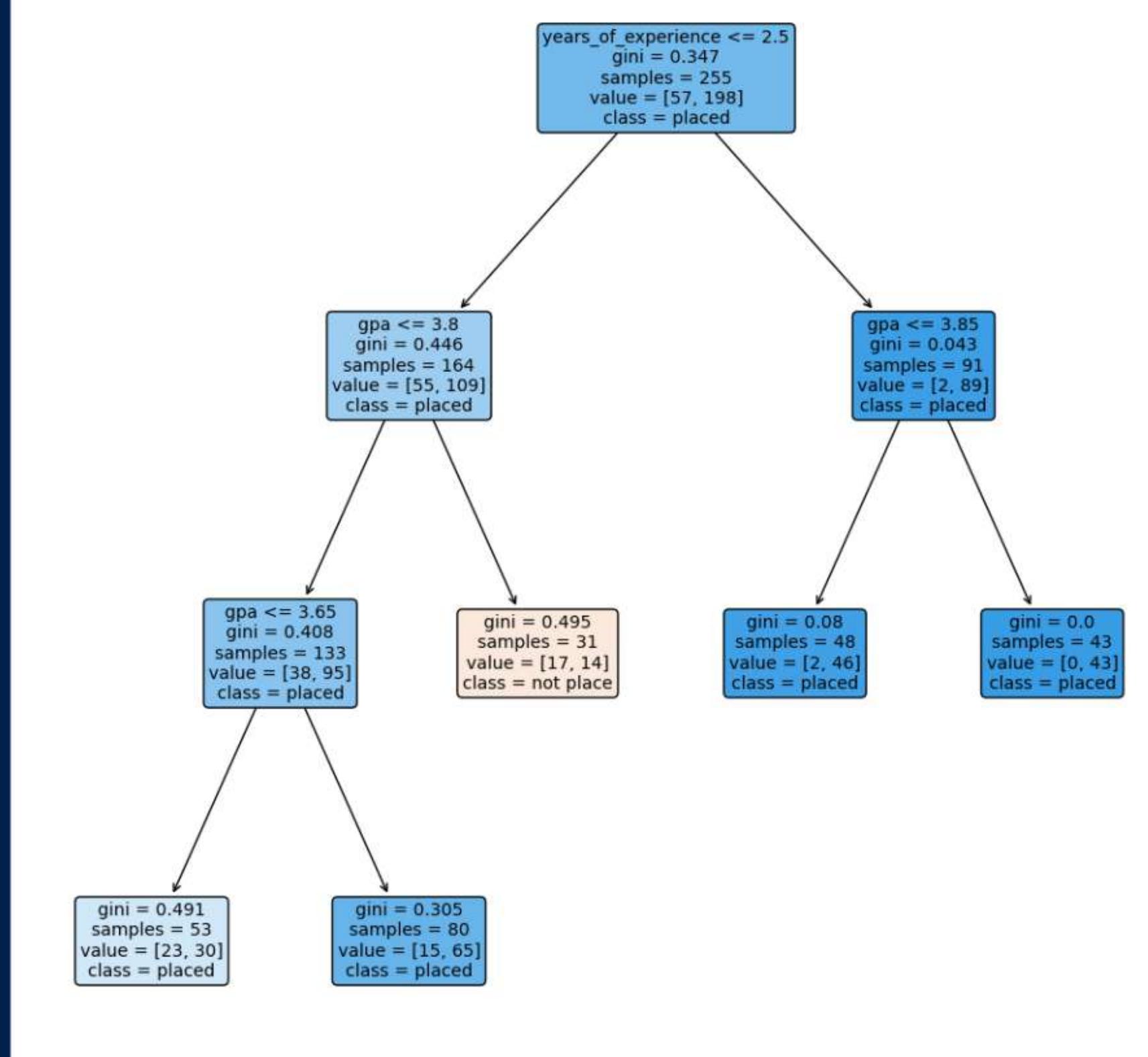
Density Plot for Years\_of\_experience by Placement Status



**Majority with Placement Status has a GPA  $\geq 3.65$  and/or Year of Experience  $\geq 1.5$**

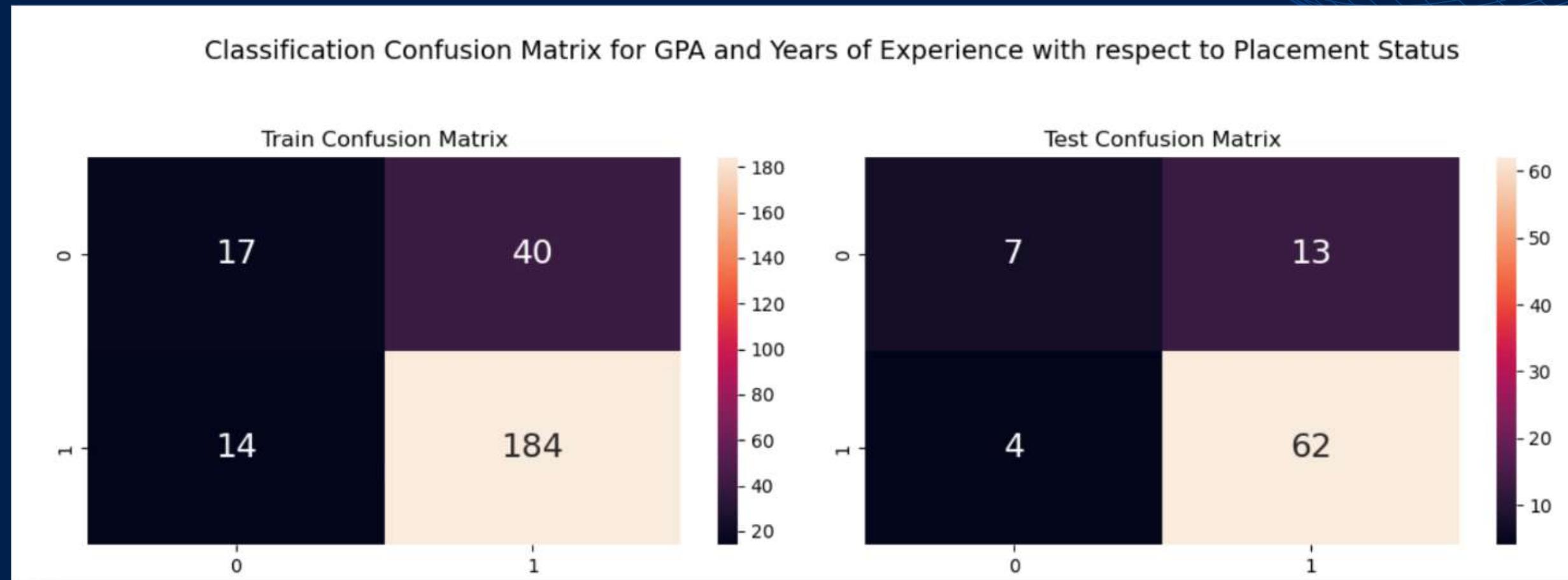
# Exploratory Analysis

## Decision Tree Classifier model



- **Years of Experience  $\geq 2.5$  and GPA  $> 3.8$  will have be placed**

# Exploratory Analysis



Goodness of Fit of Model  
Classification Accuracy

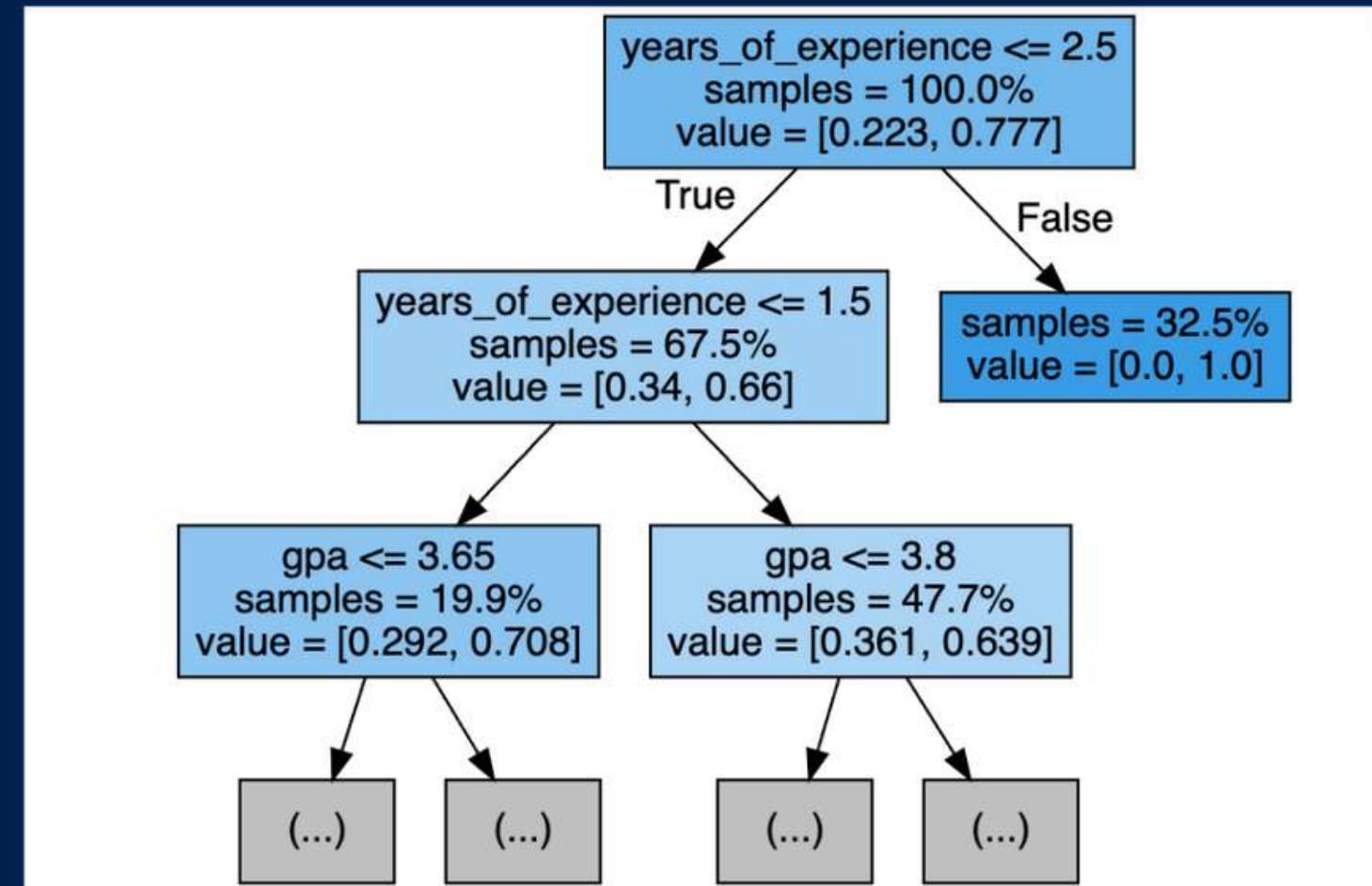
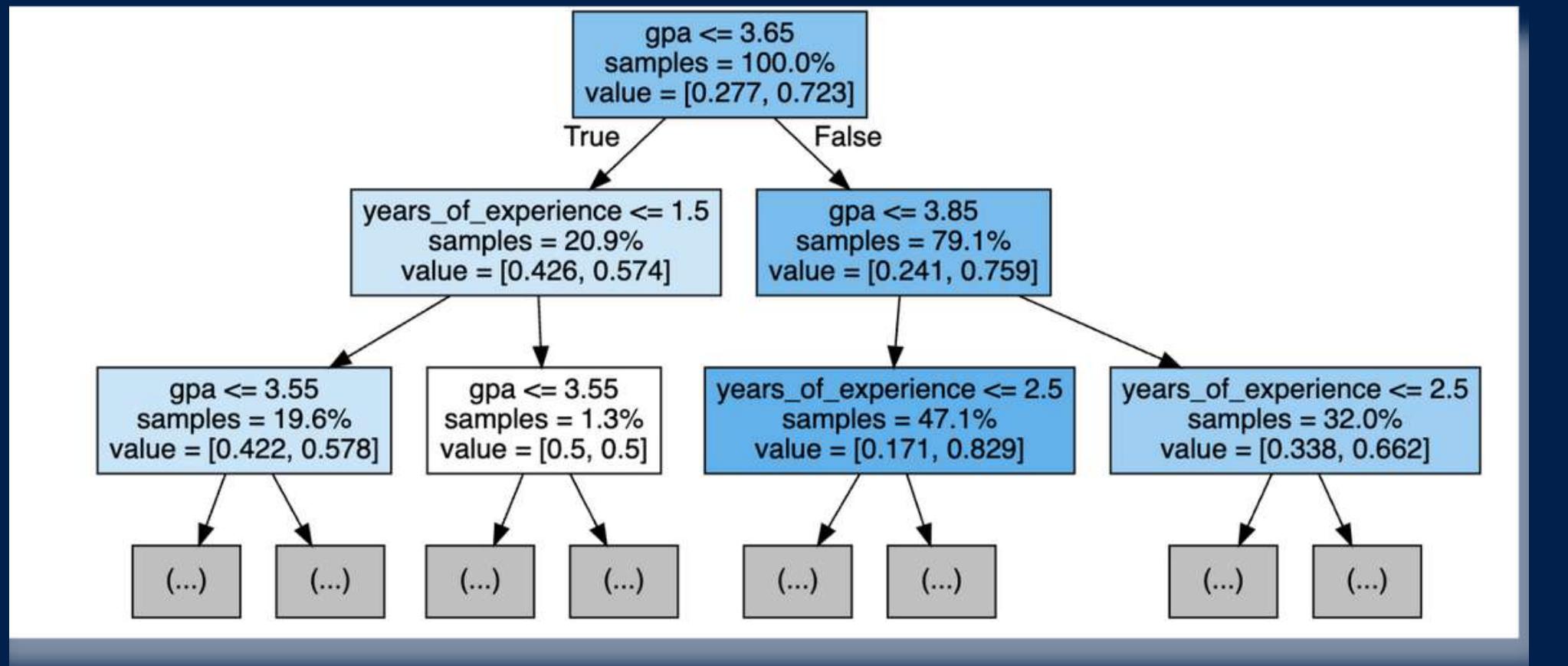
Goodness of Fit of Model  
Classification Accuracy

Train Dataset  
: 0.788235294117647

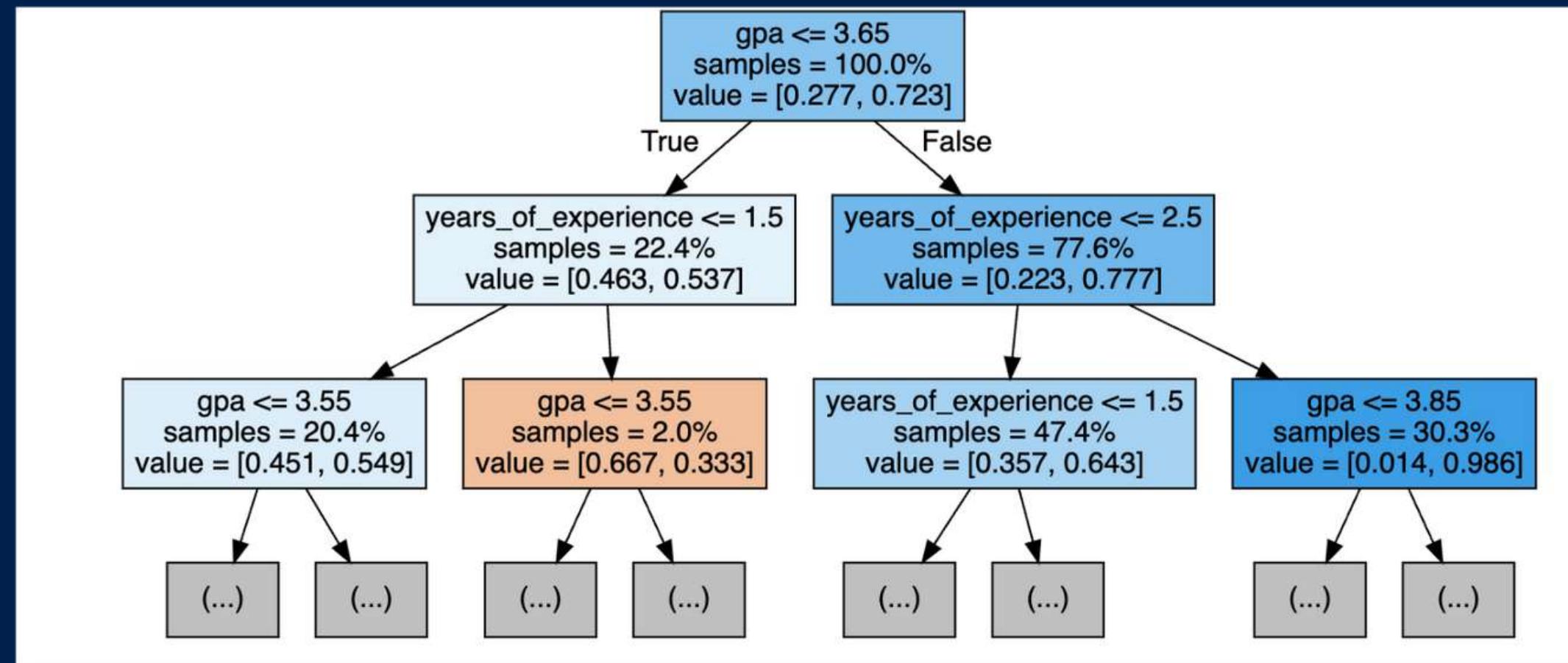
Test Dataset  
: 0.8023255813953488

Overall the Decision Tree Classifier attained a fairly good classification result

# Exploratory Analysis Random Forest Classification Model

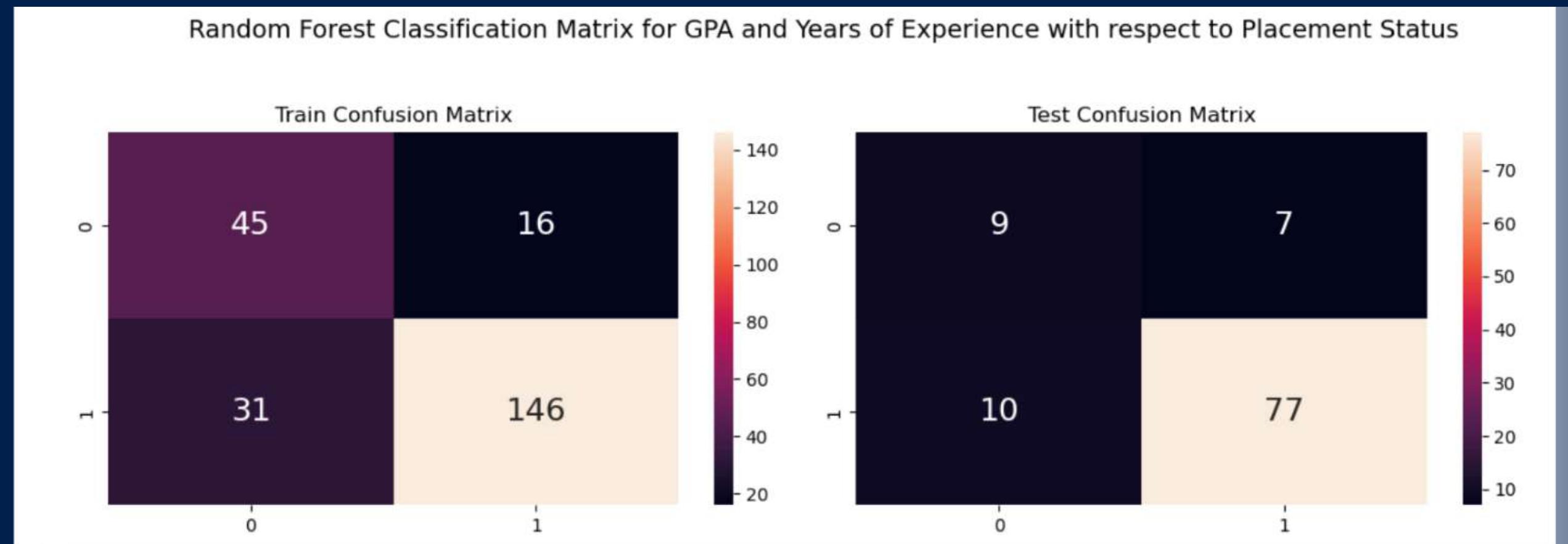


# Exploratory Analysis



Clear that those with higher GPA and work experience will likely have a placement

# Exploratory Analysis

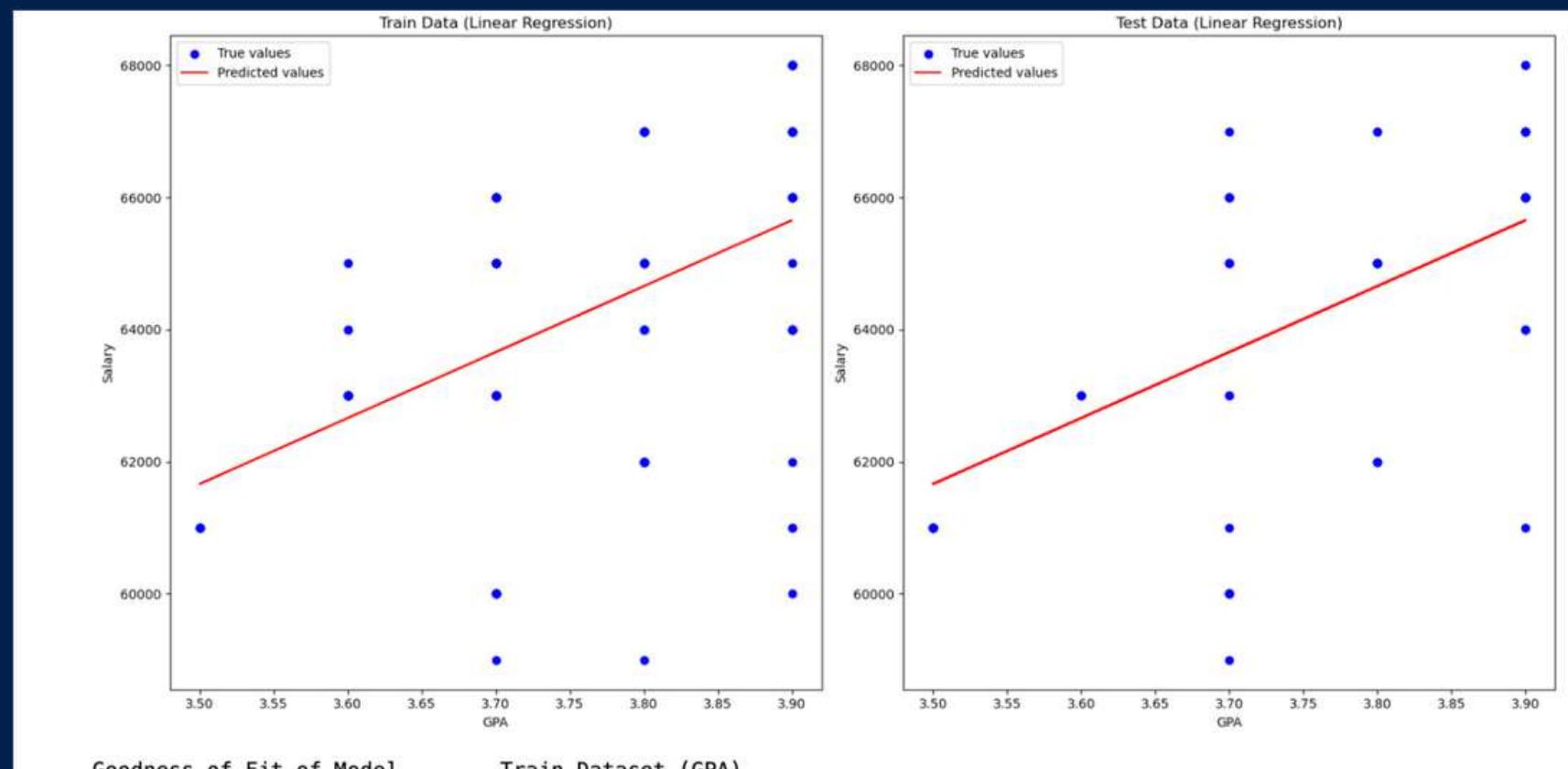


```
Random Forest Classifier Accuracy for train data: 0.8025210084033614  
Random Forest Classifier Accuracy for test data: 0.8349514563106796
```

Overall the Random Forest Classifier attained a better classification result as compared to Decision Tree Classifier due to its benefits

# Exploratory Analysis

## Linear Regression

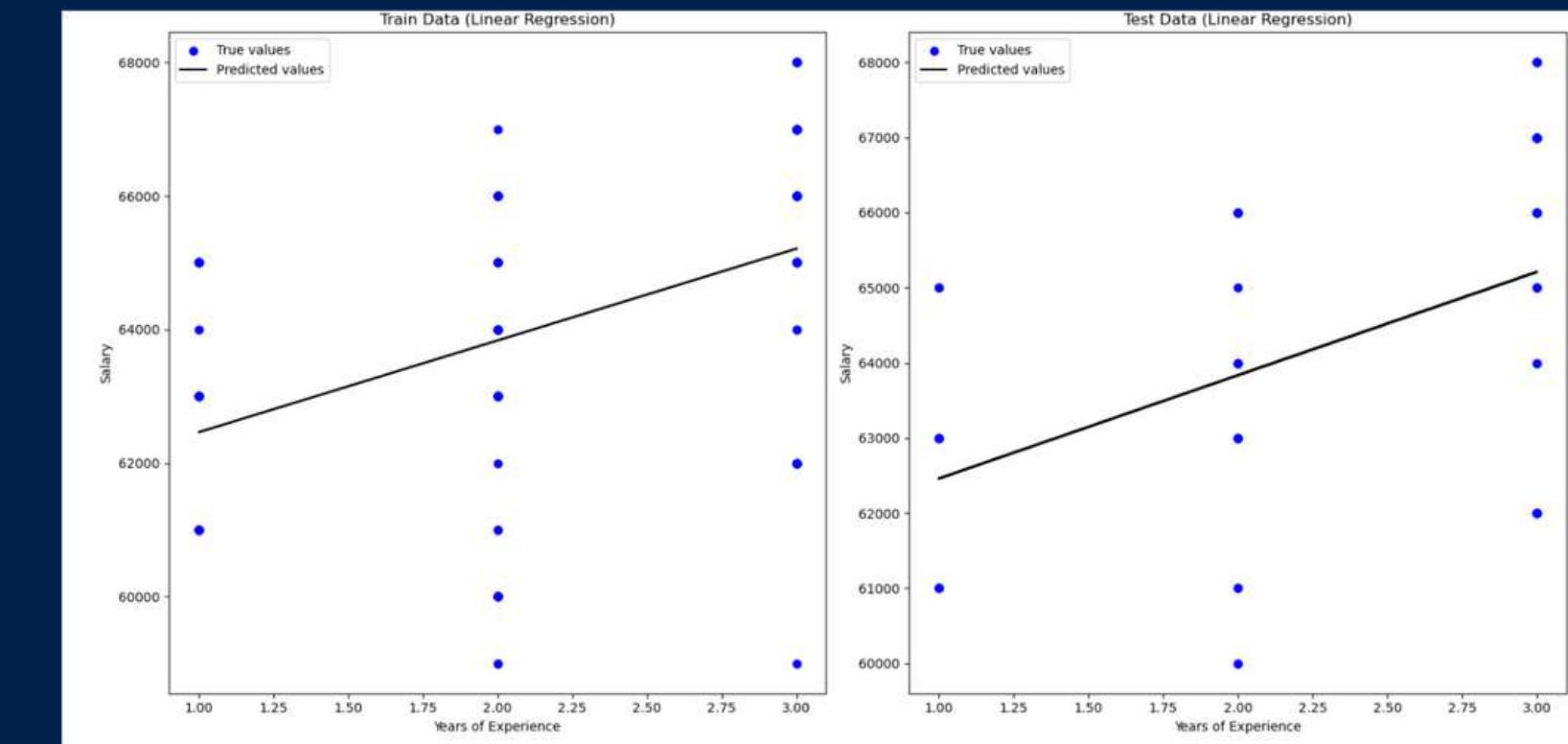


Goodness of Fit of Model  
Explained Variance ( $R^2$ )  
Mean Squared Error (MSE)

Goodness of Fit of Model  
Explained Variance ( $R^2$ )  
Mean Squared Error (MSE)

Train Dataset (GPA)  
: 0.23103913217027627  
: 3831545.040437954

Test Dataset (GPA)  
: 0.40948814463837147  
: 3625086.667636664



Goodness of Fit of Model  
Explained Variance ( $R^2$ )  
Mean Squared Error (MSE)

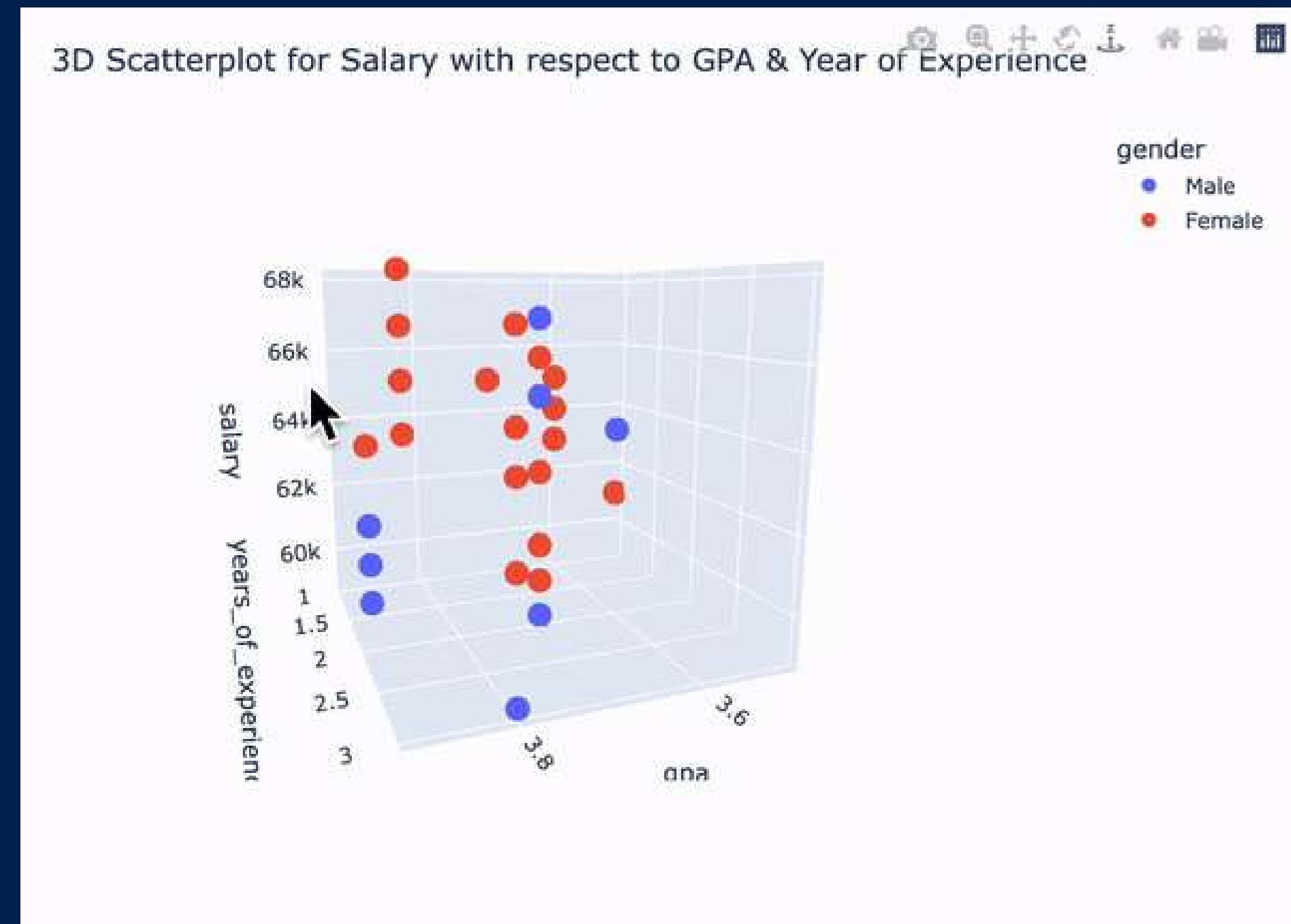
Goodness of Fit of Model  
Explained Variance ( $R^2$ )  
Mean Squared Error (MSE)

Train Dataset (Years of Experience)  
: 0.20198960218852746  
: 4278039.212509988

Test Dataset (Years of Experience)  
: 0.2323871735373102  
: 3771276.8822607957

# Exploratory Analysis

## 3D Scatterplot



**Higher GPA and/or Years of Experience = Higher Salary**

# SUMMARY

- 1) College ranking, gender, and age have minimal impact on placement**
- 2) GPA and experience influence placement and salary**



# RECOMMENDATION

Go Beyond GPA:

## Factors Influencing Placement

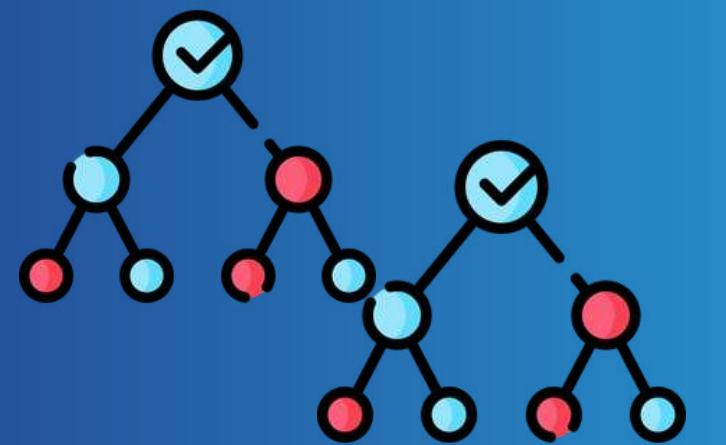
- Side projects
- Internships
- Hackathons
- Leetcode practice



# Key takeaway



1. Utilizing Github for CI/CD



2. Application of Random  
Forest Classification Model



3. Significance of Data  
Preparation

# THANK YOU!



# RESOURCES

**Arvidsson, J. (2023b, July 31). QS World University Rankings 2024. Kaggle.**  
<https://www.kaggle.com/datasets/joebeachcapital/qs-world-university-rankings-2024>

**Shafi, A. (2023, February 24). Random Forest classification with Scikit-Learn.**  
DataCamp. <https://www.datacamp.com/tutorial/random-forests-classifier-python>

**Mahad, M. (2024, March 3). Job placement dataset. Kaggle.**  
<https://www.kaggle.com/datasets/mahad049/job-placement-dataset/data>