In [1]:

```
# import all the required libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [2]:

```
# get data
df=pd.read_csv("job_placement.csv")
df
```

Out[2]:

	id	name	gender	age	degree	stream	college_name	placement_status	salar
0	1	John Doe	Male	25	Bachelor's	Computer Science	Harvard University	Placed	6000
1	2	Jane Smith	Female	24	Bachelor's	Electrical Engineering	Massachusetts Institute of Technology	Placed	6500
2	3	Michael Johnson	Male	26	Bachelor's	Mechanical Engineering	Stanford University	Placed	5800
3	4	Emily Davis	Female	23	Bachelor's	Information Technology	Yale University	Not Placed	
4	5	David Brown	Male	24	Bachelor's	Computer Science	Princeton University	Placed	6200
...	...	...	...	...	...	...	...	...	
695	696	Lucas Taylor	Male	23	Bachelor's	Computer Science	University of Washington	Placed	6700
696	697	Emma Martinez	Female	26	Bachelor's	Electronics and Communication	University of California--Berkeley	Placed	6600
697	698	Aiden Davis	Male	24	Bachelor's	Computer Science	University of Illinois--Urbana-Champaign	Placed	6500
698	699	Mia Wilson	Female	23	Bachelor's	Electrical Engineering	University of Colorado--Boulder	Placed	6600
699	700	Jack Garcia	Male	26	Bachelor's	Information Technology	University of North Carolina--Chapel Hill	Not Placed	

700 rows × 11 columns

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 699
Data columns (total 11 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   id                   700 non-null    int64  
 1   name                 700 non-null    object 
 2   gender               700 non-null    object 
 3   age                  700 non-null    int64  
 4   degree               700 non-null    object 
 5   stream               700 non-null    object 
 6   college_name         700 non-null    object 
 7   placement_status     700 non-null    object 
 8   salary               700 non-null    int64  
 9   gpa                  700 non-null    float64
 10  years_of_experience  699 non-null    float64
dtypes: float64(2), int64(3), object(6)
memory usage: 60.3+ KB
```

In [4]:

```
# to check null
df.isnull().sum()
```

Out[4]:

```
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
```

In [5]:

```
# drop the row having null value
df.dropna(inplace=True)
```

In [6]:

```
df.info()
```

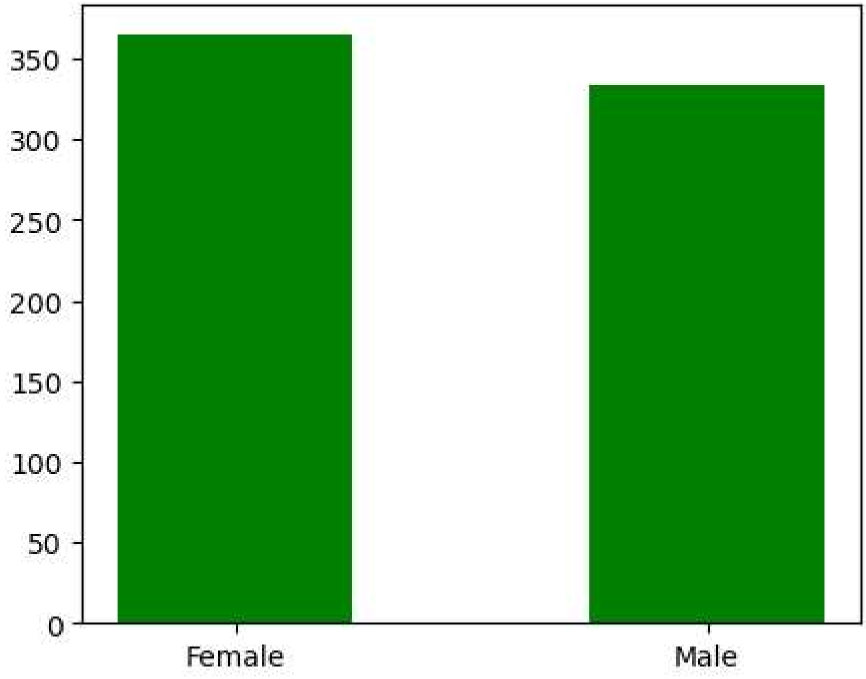
```
<class 'pandas.core.frame.DataFrame'>
Index: 699 entries, 0 to 699
Data columns (total 11 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   id                   699 non-null    int64  
 1   name                 699 non-null    object 
 2   gender               699 non-null    object 
 3   age                  699 non-null    int64  
 4   degree               699 non-null    object 
 5   stream               699 non-null    object 
 6   college_name         699 non-null    object 
 7   placement_status     699 non-null    object 
 8   salary               699 non-null    int64  
 9   gpa                  699 non-null    float64
 10  years_of_experience  699 non-null    float64
dtypes: float64(2), int64(3), object(6)
memory usage: 65.5+ KB
```

In [7]:

```
gender=df["gender"].value_counts()
gender
```

Out[7]:
```
gender
Female    365
Male      334
Name: count, dtype: int64
```

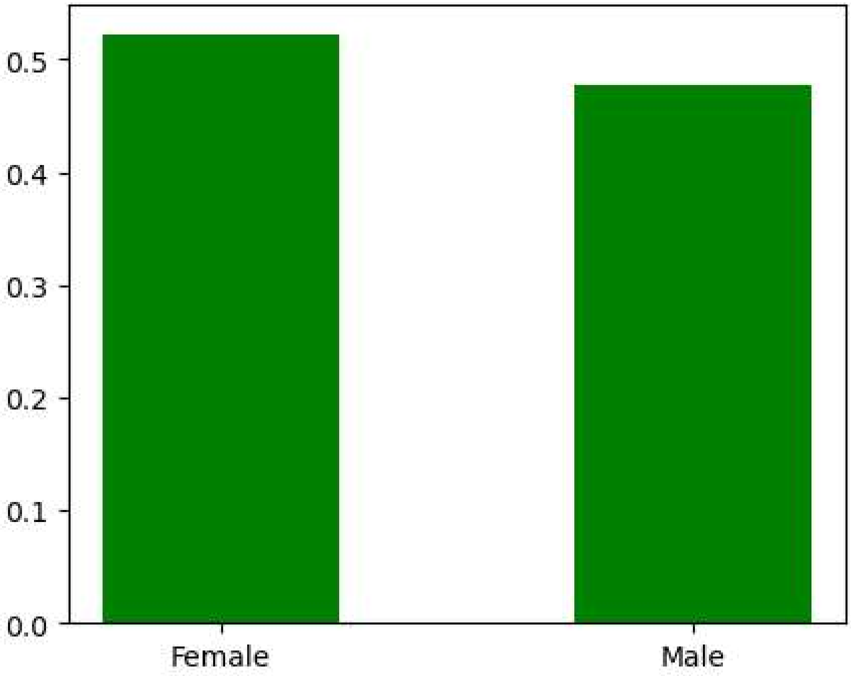
In [8]:
```
plt.figure(figsize=(5,4))
plt.bar(gender.keys(),gender,color="g",width=0.5)
plt.show()
```



In [9]:
```
gender=df["gender"].value_counts(normalize=True)
gp=gender*100
gp
```

Out[9]:
```
gender
Female    52.217454
Male      47.782546
Name: proportion, dtype: float64
```

In [10]:
```
plt.figure(figsize=(5,4))
plt.bar(gp.keys(),gender,color="g",width=0.5)
plt.show()
```



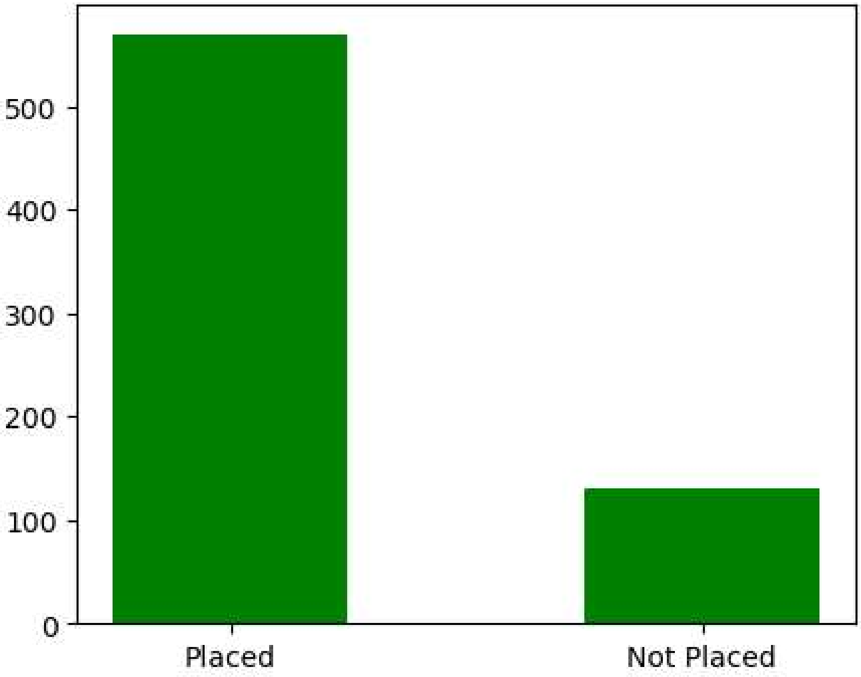
```
In [11]: df.columns
```

```
Out[11]: Index(['id', 'name', 'gender', 'age', 'degree', 'stream', 'college_name',
                'placement_status', 'salary', 'gpa', 'years_of_experience'],
               dtype='object')
```

```
In [12]: placement=df["placement_status"].value_counts()
         placement
```

```
Out[12]: placement_status
         Placed        569
         Not Placed    130
         Name: count, dtype: int64
```

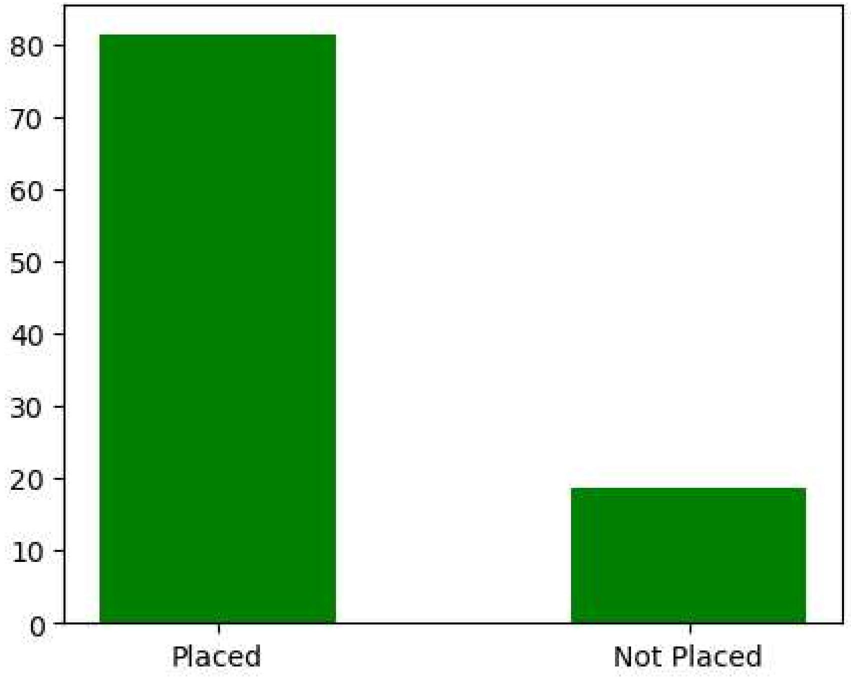
```
In [13]: plt.figure(figsize=(5,4))
         plt.bar(placement.keys(),placement,color="g",width=0.5)
         plt.show()
```



```
In [14]: placement=df["placement_status"].value_counts(normalize=True)
         pp=placement*100
         pp
```

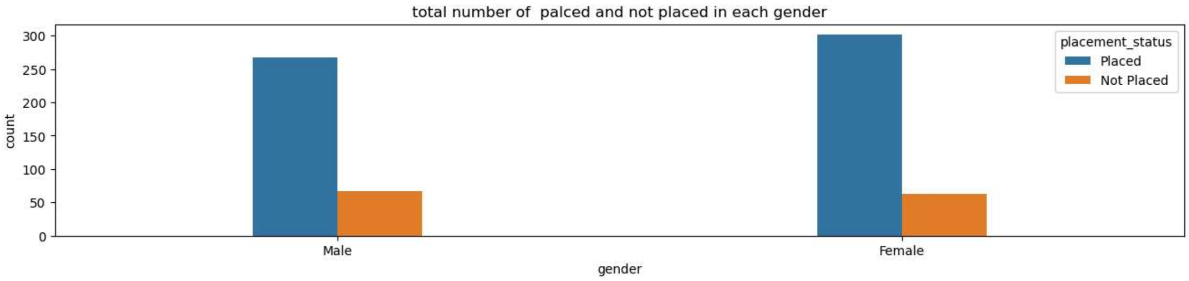
```
Out[14]: placement_status
         Placed        81.402003
         Not Placed    18.597997
         Name: proportion, dtype: float64
```

```
In [15]: plt.figure(figsize=(5,4))
         plt.bar(pp.keys(),pp,color="g",width=0.5)
         plt.show()
```



```
In [16]: plt.figure(figsize=(16,3))
         sns.countplot(x="gender",data=df,width=0.3,hue="placement_status")
         plt.title("total number of  palced and not placed in each gender")
```

```
Out[16]: Text(0.5, 1.0, 'total number of  palced and not placed in each gender')
```



```
In [17]: df["stream"].value_counts()
```

```
Out[17]: stream
         Computer Science                  214
         Information Technology            152
         Electrical Engineering            112
         Electronics and Communication     111
         Mechanical Engineering            110
         Name: count, dtype: int64
```

```
In [18]: df1=df.groupby("stream")["placement_status"].value_counts()
         df1
```

Out[18]:

```
stream                          placement_status
Computer Science                Placed              161
                                Not Placed           53
Electrical Engineering          Placed               90
                                Not Placed           22
Electronics and Communication   Placed              101
                                Not Placed           10
Information Technology          Placed              128
                                Not Placed           24
Mechanical Engineering          Placed               89
                                Not Placed           21
Name: count, dtype: int64
```

In [19]:

```
df1=df.groupby("stream")["placement_status"].value_counts(normalize=True)
df1
```

Out[19]:

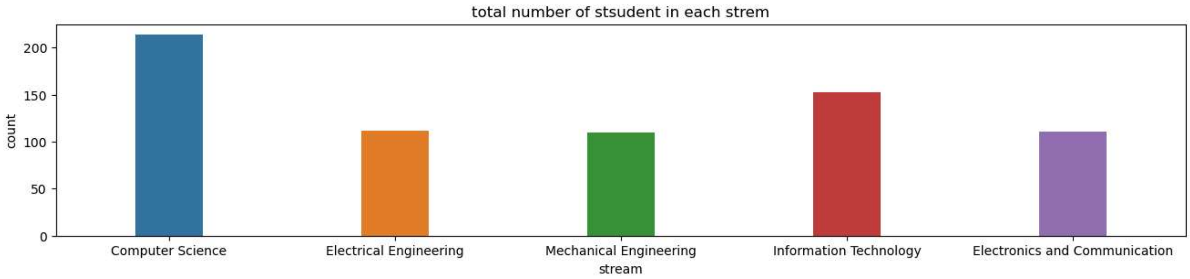
```
stream                          placement_status
Computer Science                Placed              0.752336
                                Not Placed          0.247664
Electrical Engineering          Placed              0.803571
                                Not Placed          0.196429
Electronics and Communication   Placed              0.909910
                                Not Placed          0.090090
Information Technology          Placed              0.842105
                                Not Placed          0.157895
Mechanical Engineering          Placed              0.809091
                                Not Placed          0.190909
Name: proportion, dtype: float64
```

In [20]:

```
plt.figure(figsize=(16,3))
sns.countplot(x="stream",data=df,width=0.3)
plt.title("total number of stsudent in each strem")
```

Out[20]:

```
Text(0.5, 1.0, 'total number of stsudent in each strem')
```

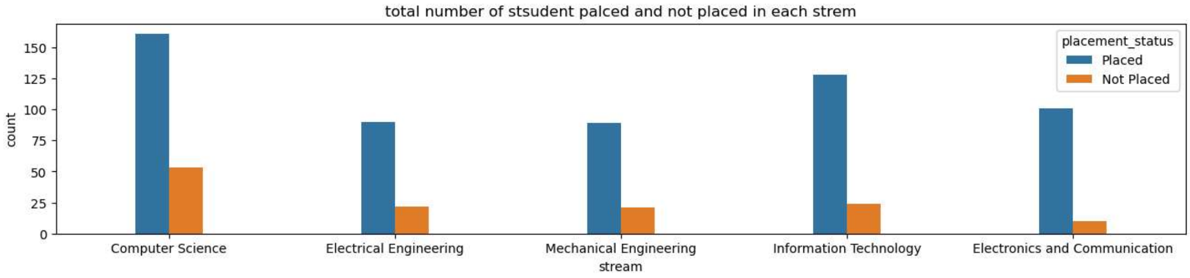


In [21]:

```
plt.figure(figsize=(16,3))
sns.countplot(x="stream",data=df,width=0.3,hue="placement_status")
plt.title("total number of stsudent palced and not placed in each strem")
```

Out[21]:

```
Text(0.5, 1.0, 'total number of stsudent palced and not placed in each strem')
```



In [22]:

```
df[["stream"]].value_counts(normalize=True)
```

Out[22]:
```
stream
Computer Science                    0.306152
Information Technology              0.217454
Electrical Engineering              0.160229
Electronics and Communication       0.158798
Mechanical Engineering              0.157368
Name: proportion, dtype: float64
```

In [23]:
```
df["years_of_experience"]=df.years_of_experience.apply(lambda i:int(i))
df
```

Out[23]:

	id	name	gender	age	degree	stream	college_name	placement_status	salar
0	1	John Doe	Male	25	Bachelor's	Computer Science	Harvard University	Placed	6000
1	2	Jane Smith	Female	24	Bachelor's	Electrical Engineering	Massachusetts Institute of Technology	Placed	6500
2	3	Michael Johnson	Male	26	Bachelor's	Mechanical Engineering	Stanford University	Placed	5800
3	4	Emily Davis	Female	23	Bachelor's	Information Technology	Yale University	Not Placed	
4	5	David Brown	Male	24	Bachelor's	Computer Science	Princeton University	Placed	6200
...	...	...	...	...	...	...	...	...	
695	696	Lucas Taylor	Male	23	Bachelor's	Computer Science	University of Washington	Placed	6700
696	697	Emma Martinez	Female	26	Bachelor's	Electronics and Communication	University of California--Berkeley	Placed	6600
697	698	Aiden Davis	Male	24	Bachelor's	Computer Science	University of Illinois--Urbana-Champaign	Placed	6500
698	699	Mia Wilson	Female	23	Bachelor's	Electrical Engineering	University of Colorado--Boulder	Placed	6600
699	700	Jack Garcia	Male	26	Bachelor's	Information Technology	University of North Carolina--Chapel Hill	Not Placed	

699 rows × 11 columns

In [24]:
```
df1=df["college_name"].value_counts().sort_values(ascending=False)
df1.head()
```

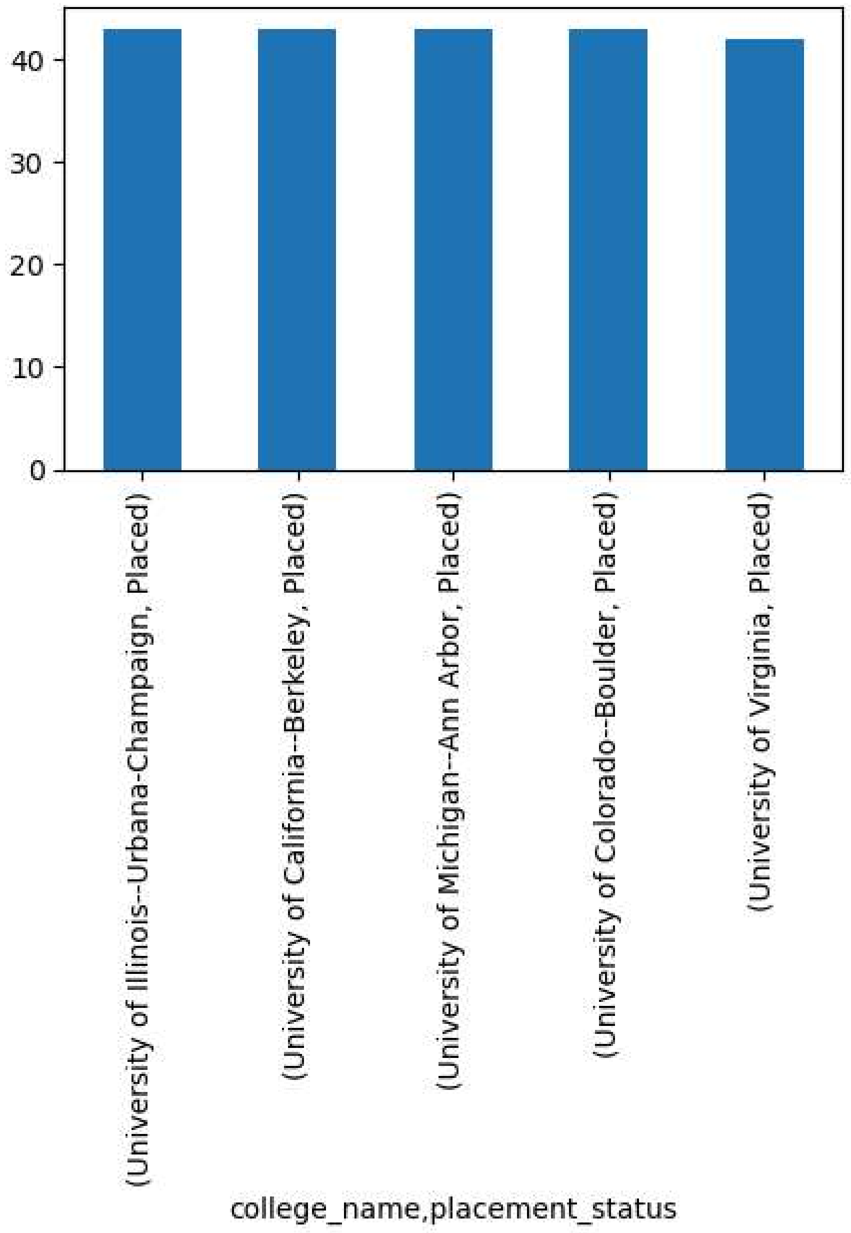
Out[24]:
```
college_name
University of Michigan--Ann Arbor              43
University of California--Berkeley             43
University of Colorado--Boulder                43
University of Illinois--Urbana-Champaign       43
University of Virginia                         43
Name: count, dtype: int64
```

In [25]:
```
df1=df.groupby("college_name")["placement_status"].value_counts().sort_values(ascer
df1=df1[:5]
df1
```

Out[25]:
```
college_name                              placement_status
University of Illinois--Urbana-Champaign  Placed              43
University of California--Berkeley        Placed              43
University of Michigan--Ann Arbor         Placed              43
University of Colorado--Boulder           Placed              43
University of Virginia                    Placed              42
Name: count, dtype: int64
```

In [26]:
```
plt.figure(figsize=(5,3))
df1.plot(kind="bar")
```

Out[26]:
```
<Axes: xlabel='college_name,placement_status'>
```



In [27]:
```
# average age of male and female
df.groupby("gender")["age"].mean()
```

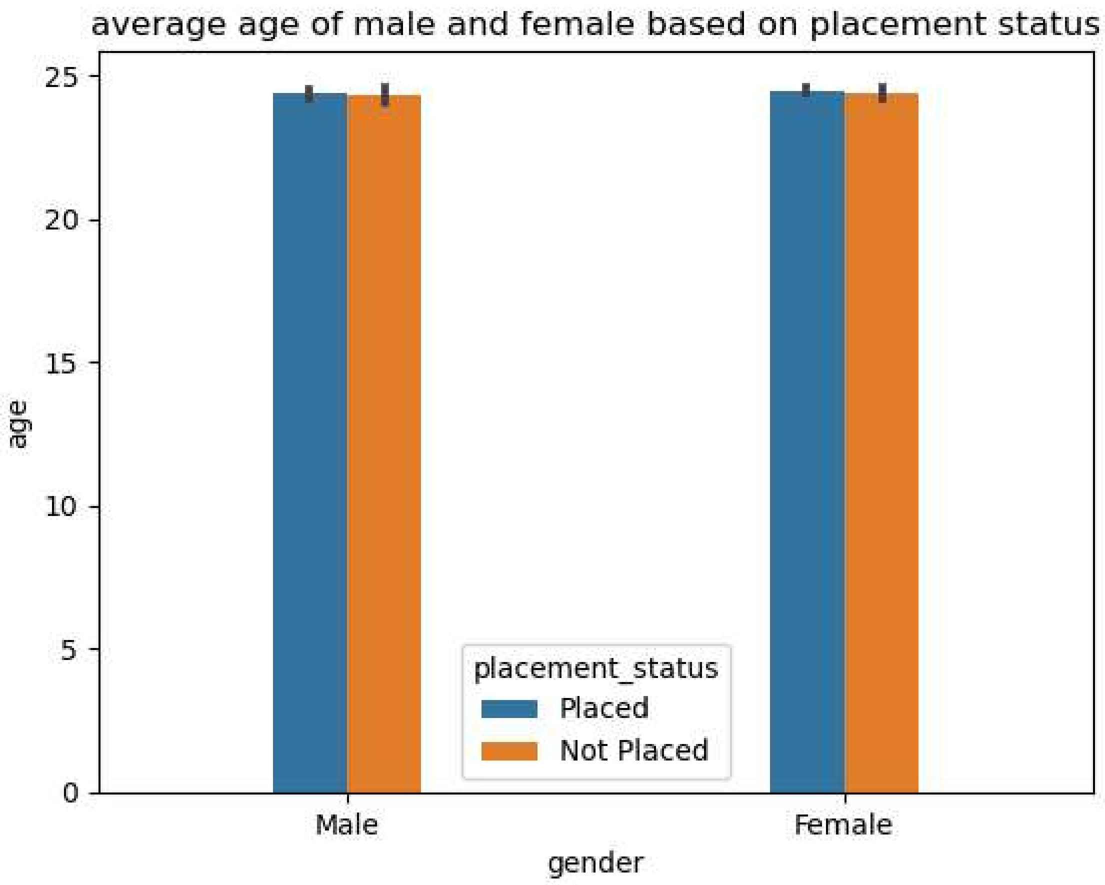
Out[27]:
```
gender
Female    24.457534
Male      24.362275
Name: age, dtype: float64
```

In [28]:
```
df.groupby("placement_status")["age"].mean()
```

Out[28]:
```
placement_status
Not Placed    24.353846
Placed        24.425308
Name: age, dtype: float64
```

In [29]:
```
sns.barplot(x="gender",y="age",data=df,hue="placement_status",width=0.3)
plt.title("average age of male and female based on placement status")
```

Out[29]:
```
Text(0.5, 1.0, 'average age of male and female based on placement status')
```

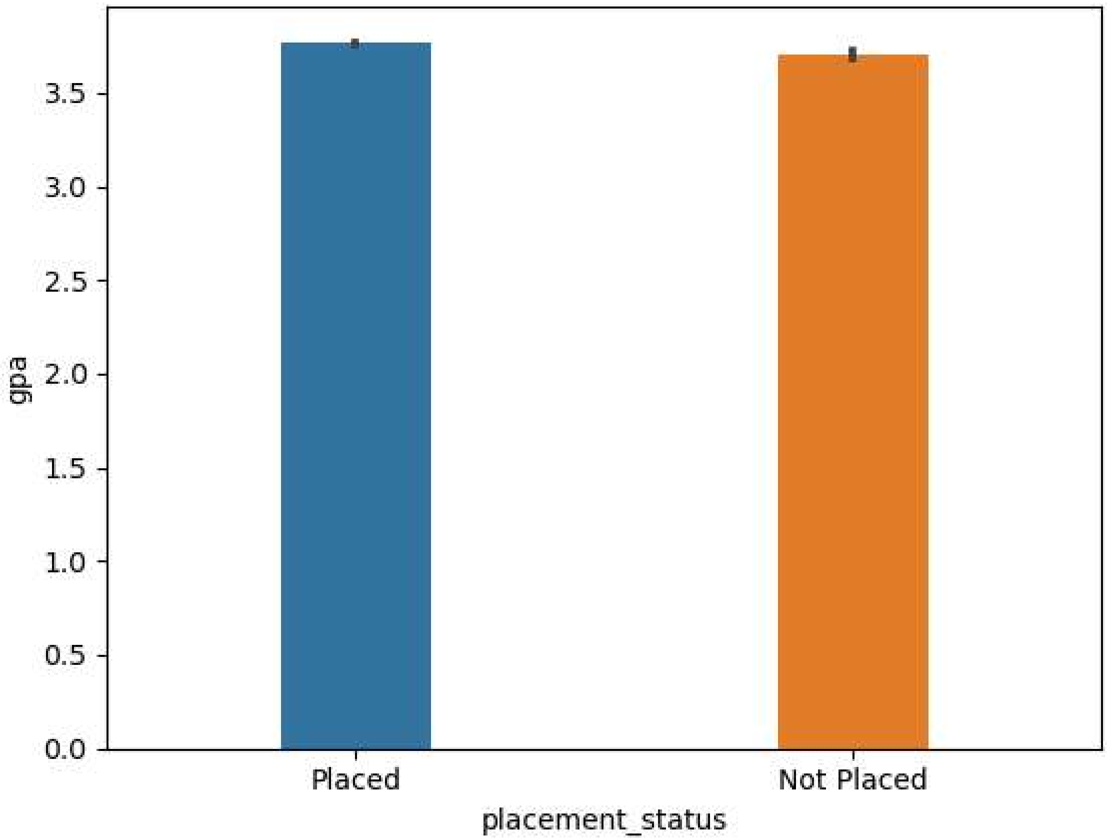


In [30]:
```
# average gpa based on their placement
df.groupby("placement_status")["gpa"].mean()
```

Out[30]:
```
placement_status
Not Placed    3.702308
Placed        3.761511
Name: gpa, dtype: float64
```

In [31]:
```
sns.barplot(x="placement_status",y="gpa",data=df,width=0.3)
```

Out[31]:
```
<Axes: xlabel='placement_status', ylabel='gpa'>
```



In [32]:

```
# average yoe based on their placement
df.groupby("placement_status")["years_of_experience"].mean()
```

Out[32]:

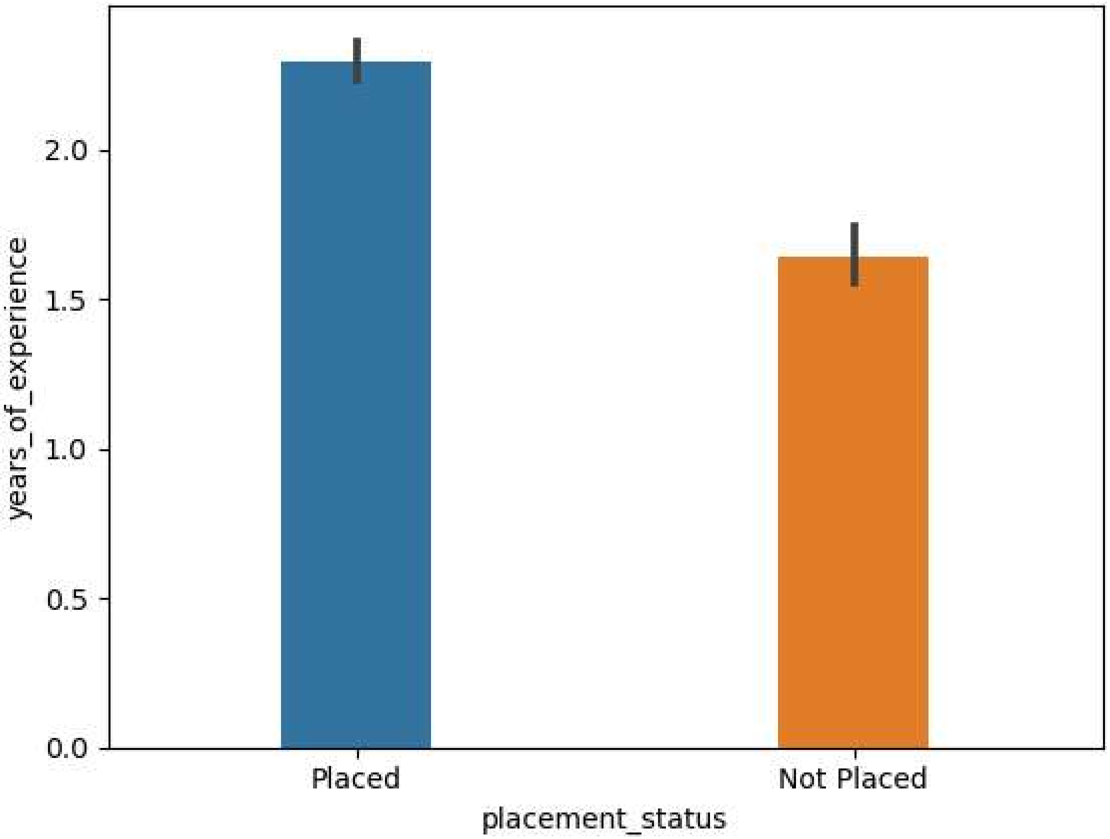
```
placement_status
Not Placed    1.646154
Placed        2.298770
Name: years_of_experience, dtype: float64
```

In [33]:

```
sns.barplot(x="placement_status",y="years_of_experience",data=df,width=0.3)
```

Out[33]:

```
<Axes: xlabel='placement_status', ylabel='years_of_experience'>
```

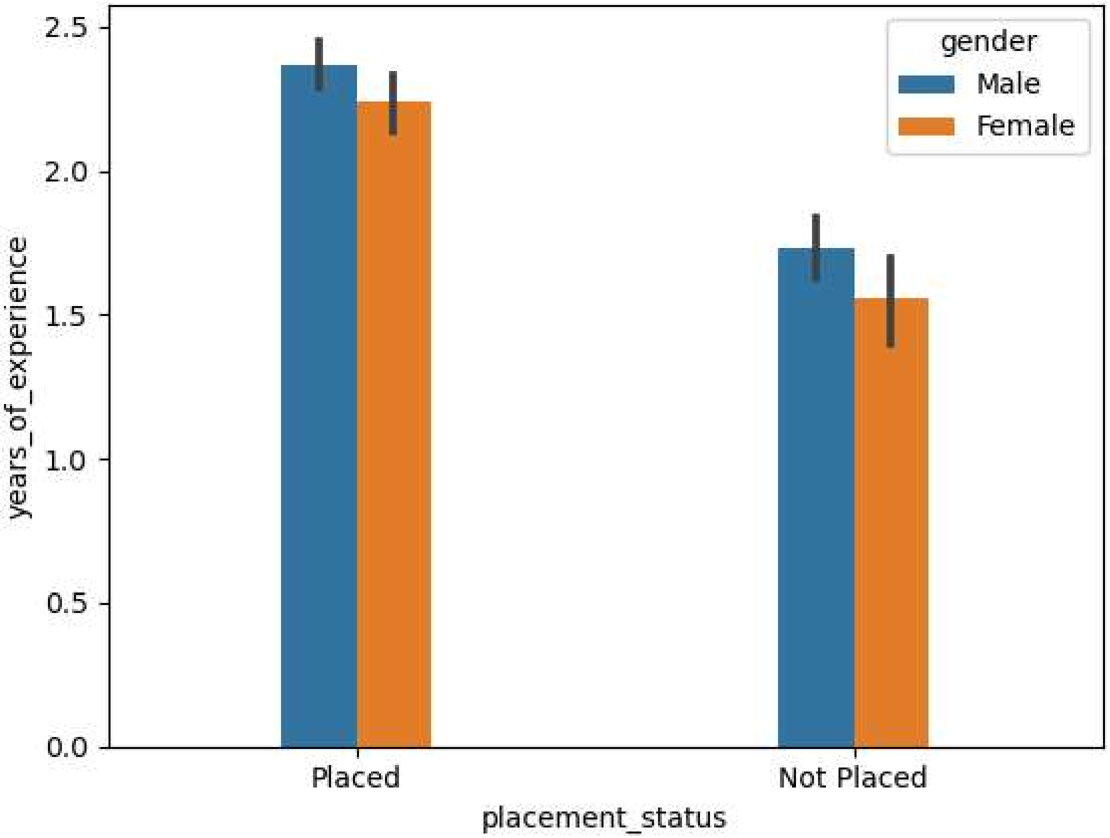


In [34]:

```
sns.barplot(x="placement_status",y="years_of_experience",data=df,width=0.3,hue="ger
```

Out[34]:

```
<Axes: xlabel='placement_status', ylabel='years_of_experience'>
```



In [ ]: