kaviyadevi 20106064

In [66]:
```
#to import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [67]:
```
#to import dataset
data=pd.read_csv(r"C:\Users\user\Downloads\13_placement - 13_placement.csv")
data
```

Out[67]:

	cgpa	placement_exam_marks	placed
0	7.19	26	1
1	7.46	38	1
2	7.54	40	1
3	6.42	8	1
4	7.23	17	0
...	...	...	...
995	8.87	44	1
996	9.12	65	1
997	4.89	34	0
998	8.62	46	1
999	4.90	10	1

1000 rows × 3 columns

DATA PREPROCESSING AND CLEANING

In [68]:
```
data.head()
```

Out[68]:

	cgpa	placement_exam_marks	placed
0	7.19	26	1
1	7.46	38	1
2	7.54	40	1
3	6.42	8	1
4	7.23	17	0

In [69]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 3 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   cgpa                  1000 non-null   float64
 1   placement_exam_marks  1000 non-null   int64  
 2   placed                1000 non-null   int64  
dtypes: float64(1), int64(2)
memory usage: 23.6 KB
```

In [70]:

```
data.describe()
```

Out[70]:

	cgpa	placement_exam_marks	placed
count	1000.000000	1000.000000	1000.000000
mean	6.961240	32.225000	0.489000
std	0.615898	19.130822	0.500129
min	4.890000	0.000000	0.000000
25%	6.550000	17.000000	0.000000
50%	6.960000	28.000000	0.000000
75%	7.370000	44.000000	1.000000
max	9.120000	100.000000	1.000000

In [71]:

```
data.columns
```

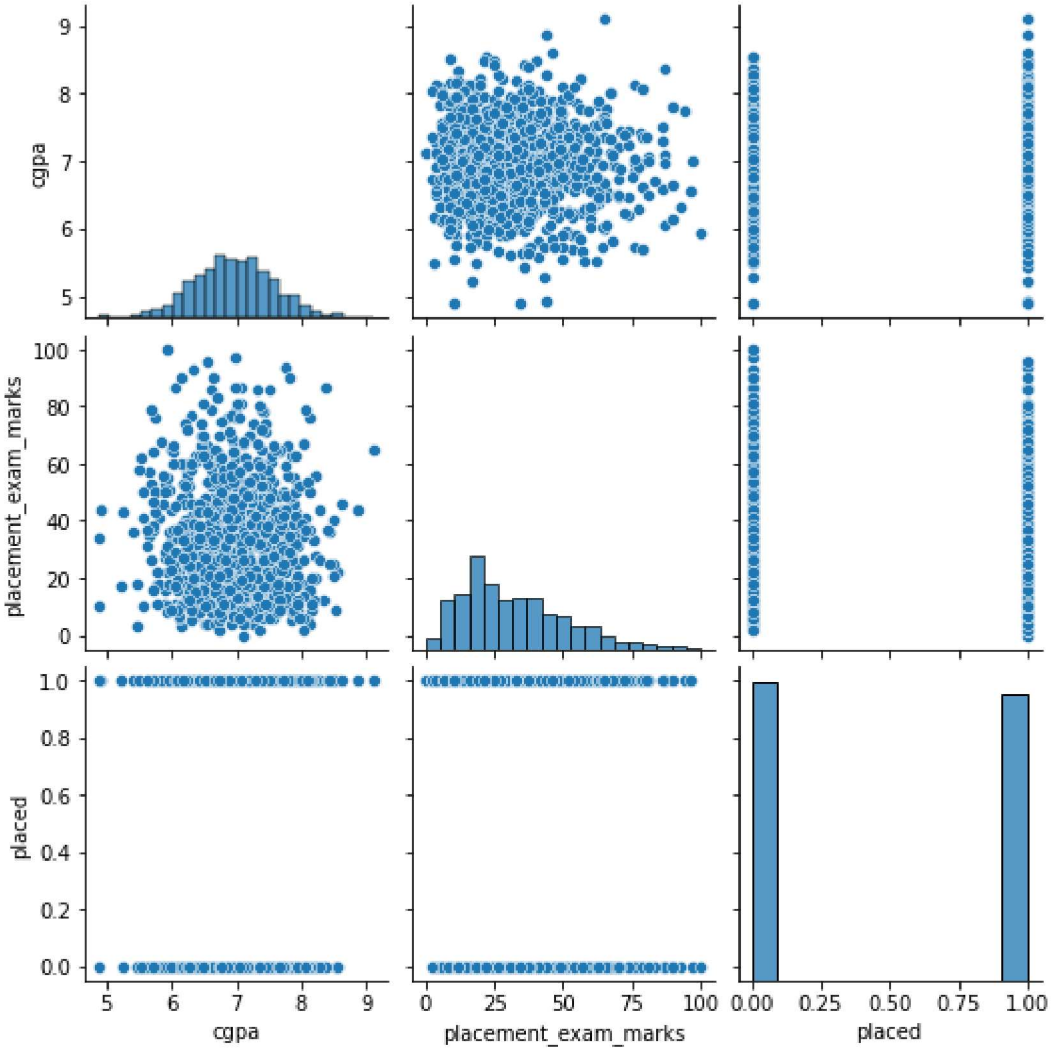
Out[71]:

```
Index(['cgpa', 'placement_exam_marks', 'placed'], dtype='object')
```

EDA and DATA VISUALIZATION

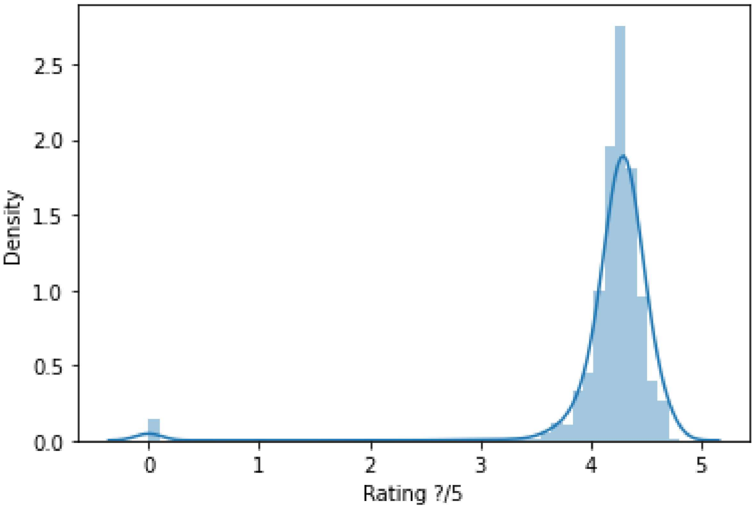
```
In [72]: sns.pairplot(data)
```

```
Out[72]: <seaborn.axisgrid.PairGrid at 0x139f28b4160>
```



```
In [45]: sns.distplot(data['Rating ?/5'])
```

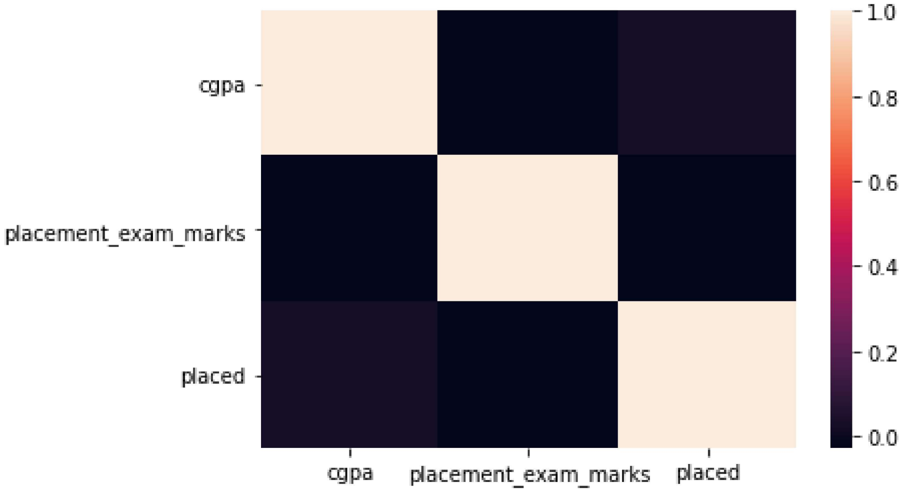
```
Out[45]: <AxesSubplot:xlabel='Rating ?/5', ylabel='Density'>
```



```
In [77]: df=data[['cgpa', 'placement_exam_marks', 'placed']]
```

```
In [78]: sns.heatmap(df.corr())
```

```
Out[78]: <AxesSubplot:>
```



MODEL TRAINING

```
In [110]: x=df[['placement_exam_marks','cgpa']]
          y=df[['placed']]
```

In [111]:
```
#to split my dataset into trainning and test

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

In [112]:
```
from sklearn.linear_model import LinearRegression

lr=LinearRegression()
lr.fit(x_train,y_train)
```

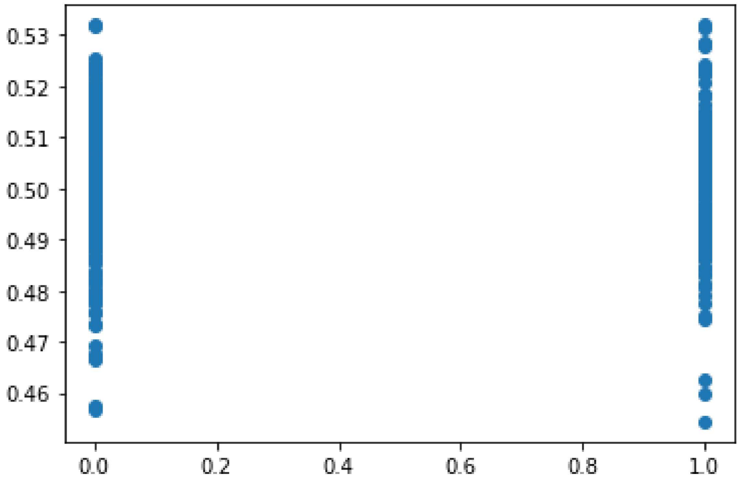
Out[112]:
```
LinearRegression()
```

In [113]:
```
#to find intercept
print(lr.intercept_)
```

```
[0.34338165]
```

In [107]:
```
prediction = lr.predict(x_test)
plt.scatter(y_test,prediction)
```

Out[107]:
```
<matplotlib.collections.PathCollection at 0x139f33480a0>
```



In [108]:
```
print(lr.score(x_test,y_test))
```

```
-0.003248468391156889
```