In [1]:
```
import pandas as pd
from sklearn.linear_model import LinearRegression
import numpy as np
```

In [3]:
```
data=pd.read_csv("C:\\Users\\mohit\\OneDrive\\Data Science\\Placement.csv")
df=pd.DataFrame(data)
df.head()
```

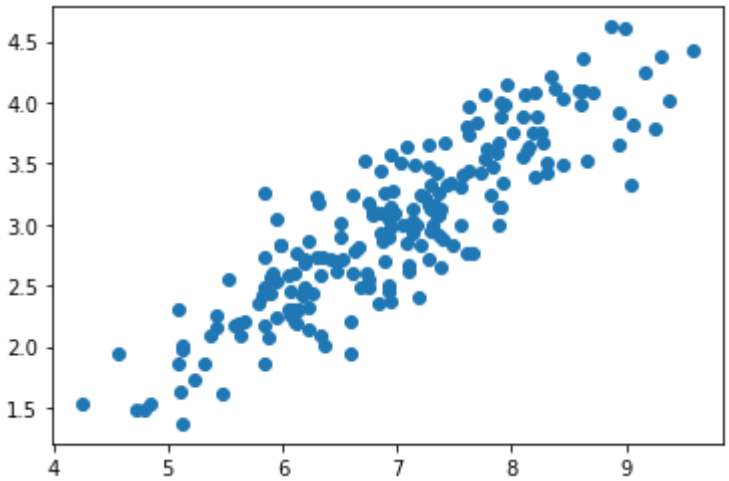
Out[3]:

	cgpa	package
0	6.89	3.26
1	5.12	1.98
2	7.82	3.25
3	7.42	3.67
4	6.94	3.57

In [4]:
```
cgpa=np.array(df["cgpa"]).reshape(-1,1)
package=np.array(df["package"]).reshape(-1,1)
```

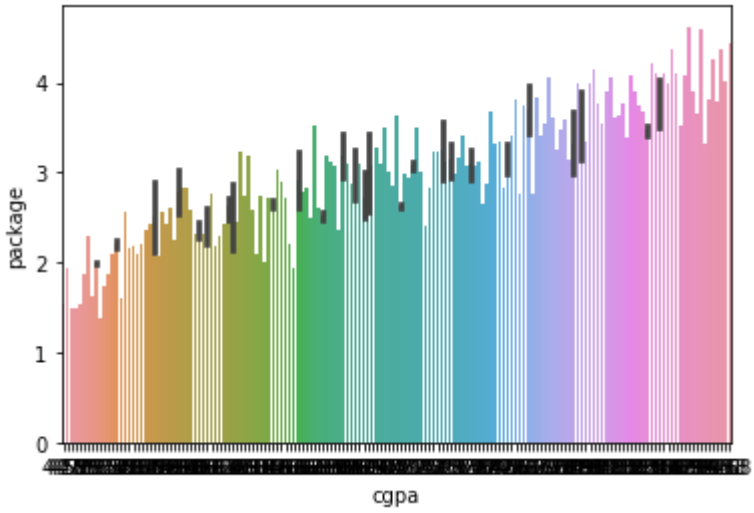
In [5]:
```
import matplotlib.pyplot as plt
import seaborn as sns

plt.scatter(df["cgpa"],df["package"])
```

Out[5]:
```
<matplotlib.collections.PathCollection at 0x1e0a8fa4a90>
```



In [6]:
```
sns.barplot(x=df["cgpa"],y=df["package"])
```

Out[6]:
```
<AxesSubplot:xlabel='cgpa', ylabel='package'>
```



In [7]:

```
q1,q3=df["cgpa"].quantile([0.25,0.75])
iqr=q3-q1

lower_bound=q1-1.5*(iqr)
upper_bound=q3+1.5*(iqr)

for i in range(len(df["cgpa"])):
    if df["cgpa"][i]>upper_bound or df["cgpa"][i]<lower_bound:
        print(df["cgpa"],df["package"])
        df["cgpa"].pop(i)
        df["package"].pop(i)
```

Observation : No outlier found.

In [10]:

```
regressor=LinearRegression()
regressor.fit(cgpa,package)

x=[]
new=float(input("Enter Cgpa to predict package : "))
x.append(new)

cgpa_new=np.array(x).reshape(-1,1)

package_predict=regressor.predict(cgpa_new)
print("Your predicted package for cgpa {} : {}".format(new,package_predict[0][0]))
```

```
Enter Cgpa to predict package : 7.8
Your predicted package for cgpa 7.8 : 3.4571341531355433
```

In [ ]:

/