ML (CSE(AIML) - 5th Semester)

SUPRATIM NAG/AIML/22/057

Linear Regression --- In Personal Dataset

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
```

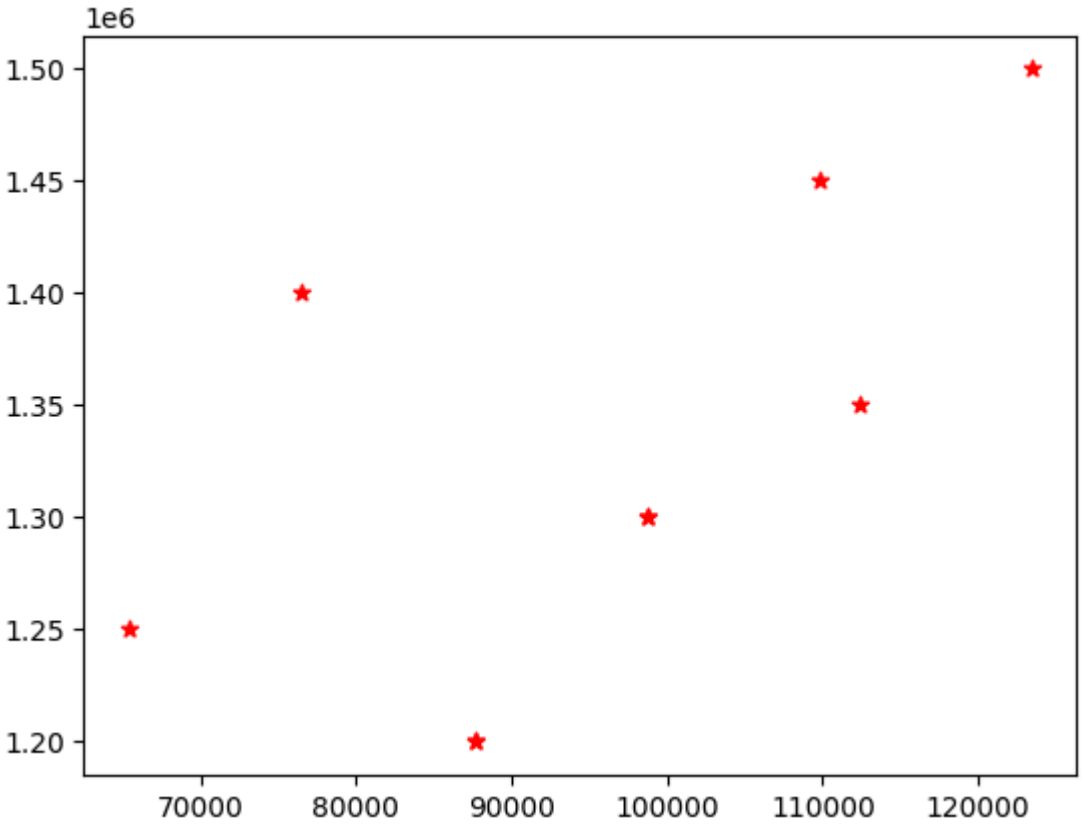
In [ ]:
```
file_path='../Personal_Datasets/housing.csv'
df=pd.read_csv(file_path)
df
```

Out[ ]:

	avg_area_income	Price
0	123456.78900	1500000.988
1	87654.32110	1200000.099
2	98765.43211	1300000.568
3	76543.21099	1400000.123
4	65432.10988	1250000.235
5	112345.67890	1350000.457
6	109876.54320	1450000.321
7	98765.43211	1300000.568
8	87654.32110	1200000.099

In [ ]:
```
plt.scatter(df.avg_area_income,df.Price,color='red',marker='*')
```

Out[ ]:
```
<matplotlib.collections.PathCollection at 0x22052838210>
```



In [ ]:
```
reg=linear_model.LinearRegression()
reg.fit(df[['avg_area_income']],df.Price)
```

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

In [ ]:
```
reg.coef_
```

Out[ ]:
```
array([3.63436562])
```

In [ ]:
```
reg.intercept_
```

Out[ ]:
```
980294.9146517538
```

In [ ]:
```
file_paths='../Personal_Datasets/housing1.csv'
d=pd.read_csv(file_paths)
d.head()
```

Out[ ]:

	avg_area_income
0	76543.21099
1	65432.10988
2	76543.21099
3	65432.10988
4	109876.54320

In [ ]:
```
p=reg.predict(d)
p
d['price']=p
plt.xlabel('avg_area_income',fontsize=20)
plt.ylabel('Price',fontsize=20)
plt.scatter(df.avg_area_income,df.Price,color='red',marker='*')
plt.plot(df.avg_area_income,reg.predict(df[['avg_area_income']]),color='blue')
```

Out[ ]:
```
[<matplotlib.lines.Line2D at 0x22052b0e6d0>]
```