22MT0214 Multivariate Linear Regression

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from google.colab import files
al=files.upload()
```

Choose files house.csv

- **house.csv**(text/csv) - 283 bytes, last modified: 31/08/2022 - 100% done

```
Saving house.csv to house (1).csv
```

```
import io
df=pd.read_csv(io.BytesIO(al['house.csv']))
```

```
df.head(3)
```

	area	bedroom	age	price
0	2600	3	20	550000
1	3000	4	15	565000
2	3200	3	18	610000

```
df.tail(3)
```

	area	bedroom	age	price
11	5500	8	30	9500000
12	5700	9	31	9700000
13	6000	10	32	10000000

```
ds=df.sample(frac=1) #shuffling of data
ds.head()
```

	area	bedroom	age	price
4	4000	5	18	760000
7	4700	6	24	870000
0	2600	3	20	550000
13	6000	10	32	10000000
1	3000	4	15	565000

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import warnings
warnings.filterwarnings("ignore")
```

```
shape=ds.shape
print("Dataset contains {} rows and {} columns".format(shape[0],shape[1]))
```

```
Dataset contains 14 rows and 4 columns
```

```
x=ds.iloc[:,:3]
y=ds.iloc[:,3]
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.4,random_state=2)
```

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

```
LinearRegression()
```

```
y_pred=linreg.predict(x_test)
y_pred
```

```
array([  208455.67193672,  4157439.56102754,  1862864.79714682,
       12168998.94463582,  1951662.75671457,  6140196.05419258])
```

```
Accuracy=r2_score(y_test,y_pred)
Accuracy
```

```
0.7423088092102976
```

```
percent=Accuracy*100
print('accuracy:',percent,'%')
```

```
accuracy: 74.23088092102977 %
```

```
from sklearn.metrics import mean_squared_error,mean_absolute_error
mean_squared_error(y_test,y_pred)
```

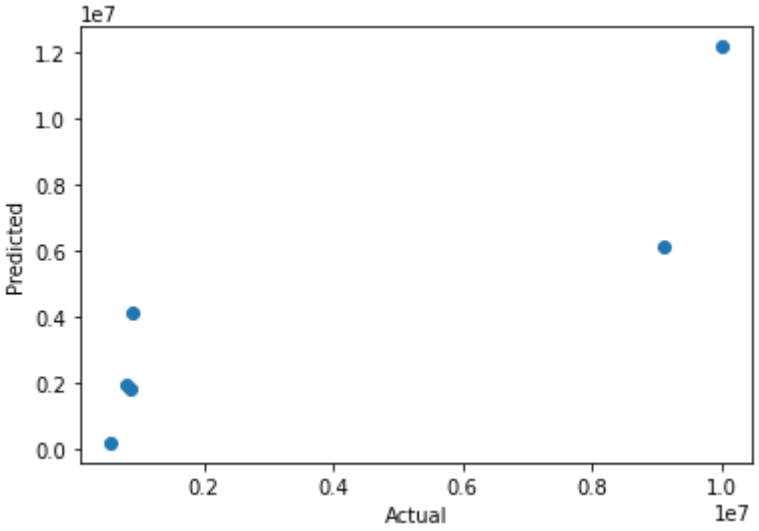
```
4425875728928.751
```

```
mean_absolute_error(y_test,y_pred)
```

```
1817885.7222325744
```

```
plt.scatter(y_test,y_pred);
```

```
plt.xlabel('Actual');
plt.ylabel('Predicted');
```




```
sns.regplot(x=y_test,y=y_pred,ci=None,color ='blue');
```