```
from google.colab import drive
drive.mount("/content/drive")
```

```
Mounted at /content/drive
```

```
%cd /content/drive/MyDrive/ML LAB/Week 2
```

```
/content/drive/MyDrive/ML LAB/Week 2
```

```
import numpy as np
import pandas as pd
import sklearn as sk
import sklearn.linear_model as slm
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
data = pd.read_csv('/content/drive/MyDrive/ML LAB/Week 2/housing.csv')
data.head()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	

```
data_refine = data.drop ('ocean_proximity',axis = 1)
data_refine.info ()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20640 entries, 0 to 20639
Data columns (total 9 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   longitude           20640 non-null  float64
 1   latitude            20640 non-null  float64
 2   housing_median_age  20640 non-null  float64
 3   total_rooms         20640 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20640 non-null  float64
 6   households          20640 non-null  float64
 7   median_income       20640 non-null  float64
 8   median_house_value  20640 non-null  float64
dtypes: float64(9)
memory usage: 1.4 MB
```

```
data_refine = data_refine.dropna(axis = 0)
data_refine.info ()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 20433 entries, 0 to 20639
Data columns (total 9 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   longitude           20433 non-null  float64
 1   latitude            20433 non-null  float64
 2   housing_median_age  20433 non-null  float64
 3   total_rooms         20433 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20433 non-null  float64
 6   households          20433 non-null  float64
 7   median_income       20433 non-null  float64
 8   median_house_value  20433 non-null  float64
dtypes: float64(9)
memory usage: 1.6 MB
```

```
X = data_refine.drop('median_house_value', axis = 1)
y = data_refine['median_house_value']
X.info()
print(y.shape)
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 20433 entries, 0 to 20639
Data columns (total 8 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   longitude           20433 non-null  float64
 1   latitude            20433 non-null  float64
 2   housing_median_age  20433 non-null  float64
 3   total_rooms         20433 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20433 non-null  float64
 6   households          20433 non-null  float64
 7   median_income       20433 non-null  float64
dtypes: float64(8)
memory usage: 1.4 MB
(20433,)
```

```
LR = slm.LinearRegression()
X_train, X_test, Y_train, Y_test = train_test_split(X,y,test_size = 0.25)
print(X_train.shape)
print(Y_train.shape)
print(X_test.shape)
print(Y_test.shape)
LR.fit(X_train,Y_train)
predict = LR.predict(X_test)
print("Predicted Value : ",predict[0])
print("Actual Value : ",Y_test.values[0])
LR.score(X_test,Y_test)
```

```
(15324, 8)
(15324,)
(5109, 8)
(5109,)
Predicted Value :  260754.34949170426
Accepted Value :  242700.0
0.6465463623994534
```

```
gr = pd.DataFrame({'Predicted':predict,'Actual':Y_test})
gr = gr.reset_index()
gr = gr.drop(['index'],axis=1)
plt.plot(gr[:1000])
plt.legend(['Actual','Predicted'])
```

```
<matplotlib.legend.Legend at 0x78d3dede6980>
```