

Assignment 7

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
df=pd.read_csv('/content/drive/MyDrive/DataSet/Real estate.csv')
df.head()
```

	No	X1 transaction date	X2 house age	X3 distance to the nearest MRT station	X4 number of convenience stores	X5 latitude	X6 longitude	Y house price of unit area
0	1	2012.917	32.0	84.87882	10	24.98298	121.54024	37.9
1	2	2012.917	19.5	306.59470	9	24.98034	121.53951	42.2
2	3	2013.583	13.3	561.98450	5	24.98746	121.54391	47.3
3	4	2013.500	13.3	561.98450	5	24.98746	121.54391	54.8

```
df.shape
```

```
(414, 8)
```

```
X=df.iloc[:,1:-1]
```

```
Y=df['Y house price of unit area']
```

```
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,random_state=3)
```

```
X_train
```



	X1 transaction date	X2 house age	X3 distance to the nearest MRT station	X4 number of convenience stores	X5 latitude	X6 longitude
75	2013.500	12.3	1360.1390	1	24.95204	121.54842
14	2013.500	13.2	1164.8380	4	24.99156	121.53406
397	2013.417	13.1	1164.8380	4	24.99156	121.53406
47	2013.583	35.9	640.7391	3	24.97563	121.53715
46	2013.417	21.7	463.9623	9	24.97030	121.54458
...
256	2012.667	14.6	339.2289	1	24.97519	121.53151
131	2013.500	4.0	2147.3760	3	24.96299	121.51284
249	2012.833	18.0	6306.1530	1	24.95743	121.47516
152	2013.333	12.0	1360.1390	1	24.95204	121.54842
362	2013.417	17.1	967.4000	4	24.98872	121.53408

```
from sklearn.linear_model import LinearRegression
model=LinearRegression()
```

```
model.fit(X_train,Y_train)
```

```
LinearRegression()
```

```
y_pred=model.predict(X_test)
```


```
from sklearn.metrics import r2_score
```

```
score=r2_score(Y_test,y_pred)
score
```

```
0.61013019865237
```

```
z=pd.DataFrame({'actual':Y_test,'predicted':y_pred})
```

```
Pz.head()
```

	actual	predicted	
161	39.6	41.400219	
130	37.5	44.632017	
227	40.2	43.212812	
103	45.7	44.031988	
240	28.8	33.343182	