In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
hi=pd.read_excel("G:/house_price.xlsx")
```

In [3]:
```
hi
```

Out[3]:

	area	bedrooms	age	price
0	2600	3.0	20	550000
1	3000	4.0	15	565000
2	3200	NaN	18	610000
3	3600	3.0	30	595000
4	4000	5.0	8	760000

In [4]:
```
hi.bedrooms.median()
```

Out[4]:
```
3.5
```

In [5]:
```
import math
median_bedrooms = math.floor(hi.bedrooms.median())
median_bedrooms
```

Out[5]:
```
3
```

In [33]:
```
hi.bedrooms = hi.bedrooms.bfill()
hi
```

Out[33]:

	area	bedrooms	age	price
0	2600	3.0	20	550000
1	3000	4.0	15	565000
2	3200	0.0	18	610000
3	3600	3.0	30	595000
4	4000	5.0	8	760000

In [ ]:
```

```

In [7]:
```
H = hi.iloc[:,:-1].values
I = hi.iloc[:,-1].values
```

In [8]:
```
H
```

Out[8]:
```
array([[2.6e+03, 3.0e+00, 2.0e+01],
       [3.0e+03, 4.0e+00, 1.5e+01],
       [3.2e+03, 0.0e+00, 1.8e+01],
       [3.6e+03, 3.0e+00, 3.0e+01],
       [4.0e+03, 5.0e+00, 8.0e+00]])
```

In [9]:
```
I
```

Out[9]:
```
array([550000, 565000, 610000, 595000, 760000], dtype=int64)
```

In [10]:
```
from sklearn.model_selection import train_test_split
H_train,H_test, I_train, I_test  = train_test_split(H, I, test_size=0.2, random_s
```

In [11]:
```
H_train
```

Out[11]:
```
array([[2.6e+03, 3.0e+00, 2.0e+01],
       [3.0e+03, 4.0e+00, 1.5e+01],
       [3.6e+03, 3.0e+00, 3.0e+01],
       [4.0e+03, 5.0e+00, 8.0e+00]])
```

In [12]:
```
H_test
```

Out[12]:
```
array([[3200.,    0.,   18.]])
```

In [13]:
```
I_train
```

Out[13]:
```
array([550000, 565000, 595000, 760000], dtype=int64)
```

In [14]:
```
I_test
```

Out[14]:
```
array([610000], dtype=int64)
```

In [15]:
```
from sklearn.linear_model import LinearRegression
```

In [16]:
```
ref = LinearRegression()
```

In [17]:
```
ref.fit(H_train, I_train)
```

Out[17]:
```
LinearRegression()
```

In [18]:
```
ref.coef_
```

Out[18]:
```
array([    236.25, -175125.  ,  -19125.  ])
```

In [19]:
```
ref.intercept_
```

Out[19]:
```
843624.9999999993
```

In [20]:
```
I_pred = ref.predict(H_test)
```

In [21]:
```
I_pred
```

Out[21]:
```
array([1255375.])
```

In [22]:
```
H_test
```

Out[22]:
```
array([[3200.,    0.,   18.]])
```

In [23]:
```
ref.predict([[3100,4,16]])
```

Out[23]:
```
array([569500.])
```

In [24]:
```
ref.predict([[2500,3,10]])
```

Out[24]:
```
array([717625.])
```

In [25]:
```
ref.predict([[3300,4,13]])
```

Out[25]:
```
array([674125.])
```

In [26]:
```
ref.predict([[3800,3,20]])
```

Out[26]:
```
array([833500.])
```

In [27]:
```
import matplotlib.pyplot as plt
```

In [28]:
```
from sklearn import metrics
```

In [29]:
```
print("Mean absolute error is:",metrics.mean_absolute_error(I_test, I_pred))
print("Mean sqared error:",metrics.mean_squared_error(I_test, I_pred))
print("Root mean squared Error:", np.sqrt(metrics.mean_squared_error(I_test, I_pr
```

```
Mean absolute error is: 645375.0000000009
Mean sqared error: 416508890625.0012
Root mean squared Error: 645375.0000000009
```

In [30]:
```
hi.describe()
```

Out[30]:

	area	bedrooms	age	price
count	5.000000	5.000000	5.00000	5.000000
mean	3280.000000	3.000000	18.20000	616000.000000
std	540.370243	1.870829	8.01249	83919.604384
min	2600.000000	0.000000	8.00000	550000.000000
25%	3000.000000	3.000000	15.00000	565000.000000
50%	3200.000000	3.000000	18.00000	595000.000000
75%	3600.000000	4.000000	20.00000	610000.000000
max	4000.000000	5.000000	30.00000	760000.000000

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: