In [60]:
```
import pandas as a
import numpy as np

import matplotlib.pyplot as plt

from  sklearn.model_selection import train_test_split
from  sklearn.linear_model import LinearRegression
from  sklearn.metrics import r2_score
```

In [61]:
```
dataset=a.read_csv("C://Users//Exam//Downloads//archive//Realestate.csv")
dataset.head()
```

Out[61]:

	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
4	5	2012.833	5.0	390.56840	5	24.97937	121.54245	43.1

In [62]:
```
X=dataset['X3 distance to the nearest MRT station']
y=dataset['Y house price of unit area']
```

In [63]:
```
X
```

Out[63]:
```
0        84.87882
1       306.59470
2       561.98450
3       561.98450
4       390.56840
          ...
409    4082.01500
410      90.45606
411     390.96960
412     104.81010
413      90.45606
Name: X3 distance to the nearest MRT station, Length: 414, dtype: float64
```

In [64]:
```
y
```

Out[64]:
```
0      37.9
1      42.2
2      47.3
3      54.8
4      43.1
       ...
409    15.4
410    50.0
411    40.6
412    52.5
413    63.9
Name: Y house price of unit area, Length: 414, dtype: float64
```

In [65]:
```
x_train,x_test,y_train,y_test=train_test_split(X,y,test_size=0.3,shuffle=True)
```

In [66]:
```
x_train,x_test,y_train,y_test
```

Out[66]:
```
(266    1783.1800
 196     707.9067
 106     189.5181
 60     1931.2070
 362     967.4000
          ...
 245     639.6198
 206     379.5575
 209     175.6294
 92     2469.6450
 367    1828.3190
 Name: X3 distance to the nearest MRT station, Length: 289, dtype: float64,
 291      56.47425
 296    1144.43600
 412     104.81010
 359    2408.99300
 197     126.72860
           ...
 275      23.38284
 46      463.96230
 342      90.45606
 33      323.65500
 165    1236.56400
 Name: X3 distance to the nearest MRT station, Length: 125, dtype: float64,
 266    23.7
 196    36.6
 106    47.1
 60     21.3
 362    40.0
        ...
 245    40.8
 206    44.0
 209    40.9
 92     21.8
 367    20.9
 Name: Y house price of unit area, Length: 289, dtype: float64,
 291    54.4
 296    34.1
 412    52.5
 359    24.7
 197    48.2
        ...
 275    49.7
 46     42.0
 342    53.5
 33     49.3
 165    30.6
 Name: Y house price of unit area, Length: 125, dtype: float64)
```

In [67]:
```
X_train = x_train.values.reshape(-1, 1)
X_test = x_test.values.reshape(-1, 1)
```

In [68]:
```
y_train = y_train.values.reshape(-1, 1)
y_test = y_test.values.reshape(-1, 1)
```

In [69]:
```
reg=LinearRegression()
reg.fit(X_train,y_train)
print(reg.score(X_test,y_test))
```

```
0.39185565729535077
```

In [70]:
```
y_predict=reg.predict(X_test)
y_predict
```

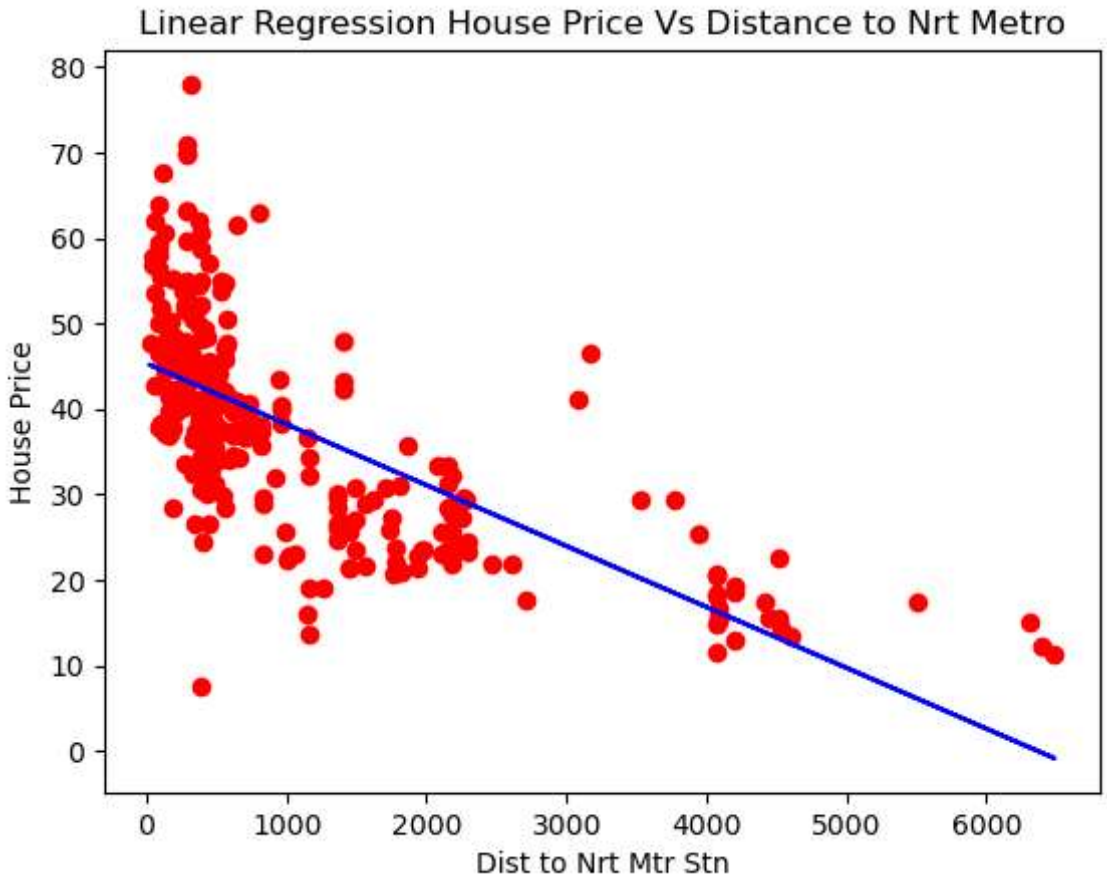
Out[70]:
```
array([[44.92496593],
       [37.17323212],
       [44.58057271],
       [28.16325613],
       [44.42440327],
       [16.24296464],
       [41.83693247],
       [42.97627363],
       [42.59646949],
       [43.54159795],
       [44.68284533],
       [28.16325613],
       [43.2659044 ],
       [41.82019441],
       [33.09233516],
       [33.12483231],
       [41.52632425],
       [35.33797312],
       [43.97331488],
       [43.39725602],
       [43.2659044 ],
       [40.06968878],
       [39.98308932],
       [44.68284533],
       [43.54159795],
       [41.82019441],
       [43.99153281],
       [43.14285626],
       [23.39532336],
       [16.24296464],
       [26.26823258],
       [43.2659044 ],
       [44.30249656],
       [43.84256176],
       [43.23973426],
       [43.48776853],
       [37.06622878],
       [44.11517588],
       [33.12483231],
       [43.38451081],
       [43.86410422],
       [27.73111028],
       [35.07746153],
       [38.45349877],
       [43.2659044 ],
       [39.93130976],
       [23.3831111 ],
       [40.93425551],
       [12.73914327],
       [16.26146828],
       [41.55268036],
       [44.54398296],
       [42.44019745],
       [40.95159849],
       [38.23258226],
       [42.01975371],
       [16.35288923],
       [39.84534158],
       [43.33823889],
       [38.13654003],
```

```
       [42.93671132],
       [29.75829556],
       [41.04621144],
       [42.68219044],
       [34.25330498],
       [43.88885007],
       [ 6.05404243],
       [44.01399154],
       [41.81003629],
       [30.02727878],
       [35.63634699],
       [40.74811473],
       [40.20949849],
       [36.31613667],
       [44.11517588],
       [13.19101114],
       [42.59232345],
       [41.62617765],
       [42.54454316],
       [32.49133379],
       [33.09233516],
       [43.99845974],
       [41.1169713 ],
       [43.2659044 ],
       [35.63634699],
       [43.92276581],
       [43.47726129],
       [44.68284533],
       [42.97548489],
       [42.30246539],
       [34.95477606],
       [42.54454316],
       [13.12452771],
       [37.08530242],
       [31.50662236],
       [18.39061952],
       [44.68284533],
       [43.84256176],
       [43.9480546 ],
       [37.07667404],
       [43.9480546 ],
       [41.31393353],
       [43.0210436 ],
       [16.24296464],
       [42.82752631],
       [42.54454316],
       [34.04331681],
       [44.00615831],
       [42.70437201],
       [42.54168461],
       [30.02727878],
       [42.93296428],
       [43.54159795],
       [41.82019441],
       [43.9480546 ],
       [33.61740534],
       [43.52769564],
       [43.37499109],
       [15.85639032],
       [44.58057271],
```

```
       [45.16074242],
       [42.0216112 ],
       [44.68284533],
       [43.02130153],
       [36.51681958]])
```

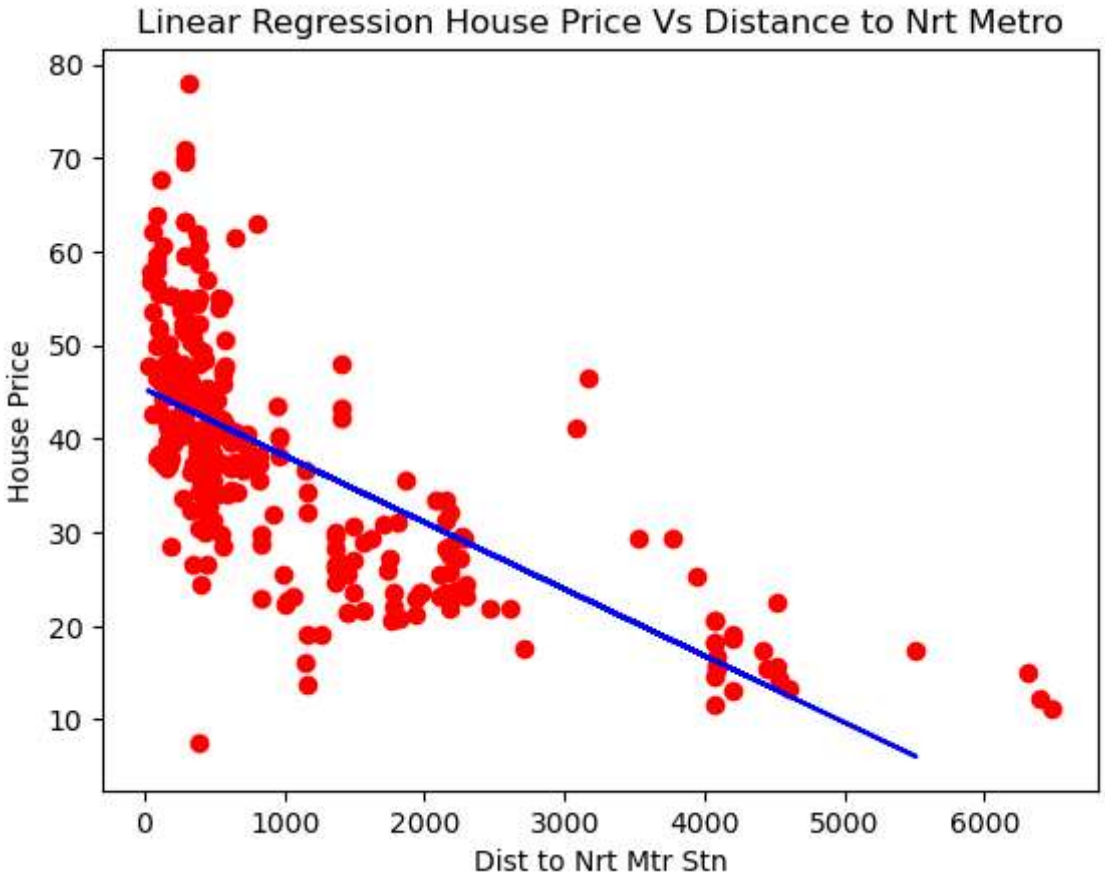
In [71]:
```
plt.scatter(X_train,y_train, color='red')
plt.plot(X_train, reg.predict(X_train), color='blue')
plt.title("Linear Regression House Price Vs Distance to Nrt Metro")
plt.xlabel("Dist to Nrt Mtr Stn")
plt.ylabel("House Price")
plt.show
```

Out[71]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [72]:
```
plt.scatter(X_train,y_train, color='red')
plt.plot(X_test, reg.predict(X_test), color='blue')
plt.title("Linear Regression House Price Vs Distance to Nrt Metro")
plt.xlabel("Dist to Nrt Mtr Stn")
plt.ylabel("House Price")
plt.show
```

Out[72]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```


Linear Regression House Price Vs Distance to Nrt Metro
House Price
Dist to Nrt Mtr Stn
80
70
60
50
40
30
20
10
0
1000
2000
3000
4000
5000
6000

In [ ]:

In [ ]: