In [ ]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
```

In [ ]:
```
# loading data set
data = pd.read_csv('C:\\Users\\huzai\\OneDrive\\Desktop\\data.csv')
```

In [ ]:
```
data.head()
```

Out[ ]:

	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 [ ]:
```
#Dependant and independant Variables
X = data[['X1 transaction date', 'X2 house age', 'X3 distance to the nearest MRT station',
          'X4 number of convenience stores', 'X5 latitude', 'X6 longitude']]
y = data['Y house price of unit area']
```

In [ ]:
```
#splitting for training and testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=0)
```

In [ ]:
```
regressor = LinearRegression()
```

In [ ]:
```
regressor.fit(X_train, y_train)
```

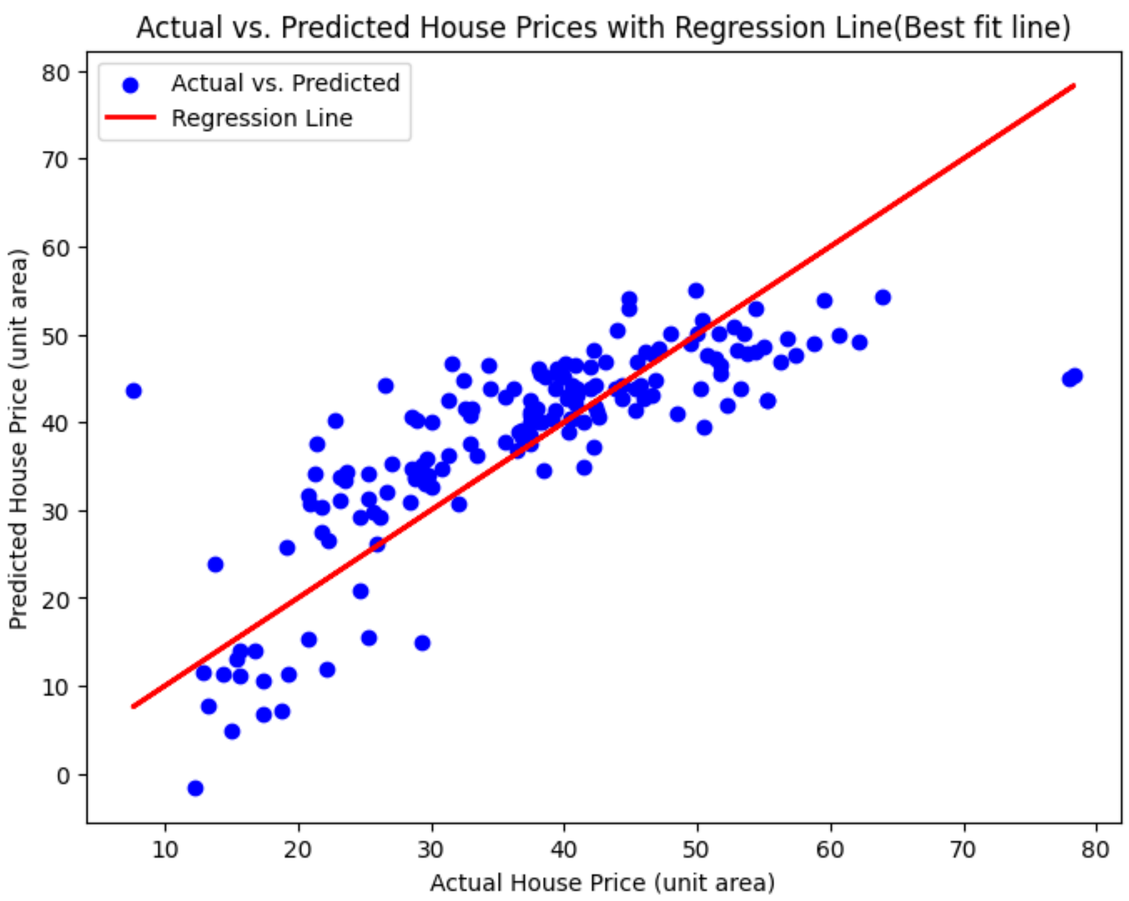
Out[ ]:
```
LinearRegression()
```

In [ ]:
```
y_pred = regressor.predict(X_test)
```

In [ ]:
```
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print("Mean Squared Error:", mse)
print("R-squared:", r2)
```

```
Mean Squared Error: 67.74242002373302
R-squared: 0.5963781925001514
```

In [ ]:
```
plt.figure(figsize=(8, 6))
plt.scatter(y_test, y_pred, color='blue', label='Actual vs. Predicted')
plt.plot(y_test, y_test, color='red', linewidth=2, label='Regression Line')
plt.title('Actual vs. Predicted House Prices with Regression Line(Best fit line)')
plt.xlabel('Actual House Price (unit area)')
plt.ylabel('Predicted House Price (unit area)')
plt.legend()
plt.show()
```



In [ ]:
```
result_df = pd.DataFrame({'Actual Price': y_test, 'Predicted Price': y_pred})
print(result_df.head(20))
```

```
     Actual Price  Predicted Price
356          45.3        41.294561
170          14.4        11.291442
224          46.0        42.690247
331          15.6        11.109610
306          50.2        43.852863
325          38.1        39.979873
150          48.5        41.006787
10           41.4        34.851242
21           51.6        50.110129
268          40.1        46.773232
316          42.0        46.394396
59           42.4        41.149964
402          28.5        40.699763
198          39.1        40.450744
348          53.7        47.823525
76           36.8        39.018088
264          40.6        40.376024
164          55.2        42.454432
12           39.3        41.317540
188          44.3        42.776466
```