Alyakbar Ahmed Sheikh RegNo. BCS-03-0003/2022

Assignment 1 Question

1. USING THE HOUSING PRICING DATA ATTACHED, CREATE A LINEAR REGRESSION TO PREDICT THE HOUSE PRICES ANALYZE THE PERFORMANCE OF THE MODEL USING MSE, MAE MAE AND RMSE Real estate.csv importing the libraries needded

loading the data

In [ ]:

```
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
import pandas as pd
df = pd.read_csv('Real estate.csv')
print(df)
```

```
      No  X1 transaction date  X2 house age  \
0      1             2012.917          32.0  
1      2             2012.917          19.5  
2      3             2013.583          13.3  
3      4             2013.500          13.3  
4      5             2012.833           5.0  
..   ...                  ...           ...  
409  410             2013.000          13.7  
410  411             2012.667           5.6  
411  412             2013.250          18.8  
412  413             2013.000           8.1  
413  414             2013.500           6.5  

     X3 distance to the nearest MRT station  X4 number of convenience stores  \
0                                  84.87882                               10   
1                                 306.59470                                9   
2                                 561.98450                                5   
3                                 561.98450                                5   
4                                 390.56840                                5   
..                                      ...                              ...   
409                              4082.01500                                0   
410                                90.45606                                9   
411                               390.96960                                7   
412                               104.81010                                5   
413                                90.45606                                9   

     X5 latitude  X6 longitude  Y house price of unit area  
0       24.98298     121.54024                        37.9  
1       24.98034     121.53951                        42.2  
2       24.98746     121.54391                        47.3  
3       24.98746     121.54391                        54.8  
4       24.97937     121.54245                        43.1  
..           ...           ...                         ...  
409     24.94155     121.50381                        15.4  
410     24.97433     121.54310                        50.0  
411     24.97923     121.53986                        40.6  
412     24.96674     121.54067                        52.5  
413     24.97433     121.54310                        63.9  

[414 rows x 8 columns]
```

Exploring the data or look inside it

```
In [ ]: print(df.head())
        print(df.info())
```

```
   No  X1 transaction date  X2 house age  \
0   1             2012.917          32.0   
1   2             2012.917          19.5   
2   3             2013.583          13.3   
3   4             2013.500          13.3   
4   5             2012.833           5.0   

   X3 distance to the nearest MRT station  X4 number of convenience stores  \
0                                84.87882                               10   
1                               306.59470                                9   
2                               561.98450                                5   
3                               561.98450                                5   
4                               390.56840                                5   

   X5 latitude  X6 longitude  Y house price of unit area  
0     24.98298     121.54024                        37.9  
1     24.98034     121.53951                        42.2  
2     24.98746     121.54391                        47.3  
3     24.98746     121.54391                        54.8  
4     24.97937     121.54245                        43.1  
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 414 entries, 0 to 413
Data columns (total 8 columns):
 #   Column                                  Non-Null Count  Dtype  
---  ------                                  --------------  -----  
 0   No                                      414 non-null    int64  
 1   X1 transaction date                     414 non-null    float64
 2   X2 house age                            414 non-null    float64
 3   X3 distance to the nearest MRT station  414 non-null    float64
 4   X4 number of convenience stores         414 non-null    int64  
 5   X5 latitude                             414 non-null    float64
 6   X6 longitude                            414 non-null    float64
 7   Y house price of unit area              414 non-null    float64
dtypes: float64(6), int64(2)
memory usage: 26.0 KB
None
```

seperating the data into variables

```
In [ ]: X = df.drop(['No', 'Y house price of unit area'], axis=1)
        y = df['Y house price of unit area']
```

Training and testing the sets

```
In [ ]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_sta
```

creating linear regression

```
In [ ]: model = LinearRegression()
        model.fit(X_train, y_train)
```

Out[ ]:

```
▾ LinearRegression ⓘ ?
LinearRegression()
```

Making predictions

In [ ]:

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

Calculating the Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)

In [ ]:

```
mse = mean_squared_error(y_test, predictions)
mae = mean_absolute_error(y_test, predictions)
rmse = np.sqrt(mse)

print("Mean Squared Error (MSE):", mse)
print("Mean Absolute Error (MAE):", mae)
print("Root Mean Squared Error (RMSE):", rmse)
```

```
Mean Squared Error (MSE): 53.50561912450505
Mean Absolute Error (MAE): 5.305355690074272
Root Mean Squared Error (RMSE): 7.314753524521866
```