Implement Regression With AWS

In [70]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [52]:
```
## Read the Reat estate data

df = pd.read_csv("https://nehagoudbaddam.s3.amazonaws.com/Data/Real+estate.csv")
print('Size of Reat Estate data frame is :',df.shape)
df.head()
```

```
Size of Reat Estate data frame is : (414, 8)
```

Out[52]:

	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 [53]:
```
## Preprocess the data

df.drop(['No'], axis = 1, inplace = True)
print("new shape:" ,df.shape)
```

```
new shape: (414, 7)
```

In [ ]:
```

```

In [71]:
```
## Implement Linear regression Model


X = df[["X1 transaction date","X2 house age","X3 distance to the nearest MRT stat
        "X4 number of convenience stores","X5 latitude","X6 longitude"]]

y = df["Y house price of unit area"]

X_train, X_test, y_train, y_test= train_test_split(X, y, test_size=0.2, random_st


linreg= LinearRegression()
linreg.fit(X_train,y_train)

y_pred = linreg.predict(X_test)

print("done")
```

```
done
```

In [72]:
```
## Evaluate the model
from sklearn.metrics import r2_score

score = r2_score(y_test,y_pred)
print('Accuracy :',score)
```

```
Accuracy : 0.5941247122614676
```

In [74]:
```
## Making predictions

Obs = pd.read_csv("https://nehagoudbaddam.s3.amazonaws.com/Data/RealestateObs.csv

newData = Obs.values

y_pred = logreg.predict(newData)
y_pred
```

Out[74]:
```
array([46.30788255, 47.02644092, 48.84775341])
```

Task 1: Execute the code properly with given sample data

Task 2: Explain what you analyzed in the code. Make a detailed report.

Task 3: Use any other dataset to run the tasks above again.

Task 4: Perform any other machine learning code with putting data in AWS S3.

In [ ]: