Train And Test Data

Date	19 September 2022
Team ID	PNT2022TMID25121
Project Name	Project - Crude Oil Price Prediction
Maximum Marks	4 Marks

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
ds=pd.read_excel(r"C:\Users\Dhyalan\Desktop\Crude Oil Prices Daily.xlsx")
ds.head()
```

Out[1]:

	Date	Closing Value
0	1986-01-02	25.56
1	1986-01-03	26.00
2	1986-01-06	26.53
3	1986-01-07	25.85
4	1986-01-08	25.87

In [2]:

```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

# get the locations
X = ds.iloc[:, :-1]
y = ds.iloc[:, -1]

# split the dataset
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.05, random_state=0)
```

In [5]:

```
print(X_train, X_test, y_train, y_test)
```

```
           Date
1940 1993-08-11
2270 1994-12-01
2500 1995-10-30
572  1988-04-07
7144 2014-04-29
...         ...
4373 2003-04-17
7891 2017-03-30
4859 2005-03-31
3264 1998-11-10
2732 1996-10-01

[7811 rows x 1 columns]            Date
5993 2009-10-02
7764 2016-09-30
7937 2017-06-05
7986 2017-08-11
2402 1995-06-12
...         ...
6706 2012-08-01
5489 2007-10-03
7663 2016-05-15
396  1987-07-30
8206 2018-06-15

[412 rows x 1 columns] 1940     17.87
2270     17.77
2500     17.67
572      17.05
7144    101.56
         ...
4373     30.10
7891     50.35
4859     55.31
3264     13.54
2732     24.35
Name: Closing Value, Length: 7811, dtype: float64 5993    69.80
7764    48.24
7937    47.40
7986    48.82
2402    18.87
        ...
6706    88.99
5489    79.97
7663    46.80
396     21.47
8206    65.01
Name: Closing Value, Length: 412, dtype: float64
```

In [ ]:

```
```