Data Spliting Into Train And Test

PNT2022TMID32512

```
import numpy as np import pandas as pd import seaborn
as sns import matplotlib.pyplot as plt
ds=pd.read_csv(r"/content/Crude-Oil-Prices-Daily.csv")
ds.head()
```

```
       Date  Closing Value
0  1/2/1986          25.56
1  1/3/1986          26.00
2  1/6/1986          26.53
3  1/7/1986          25.85 4  1/8/1986          25.87
```

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

X = ds.iloc[:, :-1] y
= ds.iloc[:, -1]

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.05, random_state=0) print(X_train)
```

```
            Date
1940   8/11/1993
2270   12/1/1994
2500   10/30/1995
572    4/7/1988 7144
4/29/2014 ...
... 4373
4/17/2003
7891   3/30/2017
4859   3/31/2005
3264   11/10/1998
2732   10/1/1996

[7811 rows x 1 columns] print(X_test)

           Date
5993   10/2/2009
7764   9/30/2016 7937
6/5/2017
7986   8/11/2017 2402
6/12/1995
...          ...
6706   8/1/2012
5489  10/3/2007
```

```
7663  5/15/2016
396   7/30/1987
8206  6/15/2018

[412 rows x 1 columns] print(y_train)

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

print(y_test)

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
```