```
import pandas as pd
import numpy as np

df=pd.read_csv('https://raw.githubusercontent.com/sithukhant1/train-valid-test/main/data/netflix-stock-price-prediction.csv')
```

+ Code + Text

```
df
```

| | Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|---|
| **0** | 2018-02-05 | 262.000000 | 267.899994 | 250.029999 | 254.259995 | 254.259995 | 11896100 |
| **1** | 2018-02-06 | 247.699997 | 266.700012 | 245.000000 | 265.720001 | 265.720001 | 12595800 |
| **2** | 2018-02-07 | 266.579987 | 272.450012 | 264.329987 | 264.559998 | 264.559998 | 8981500 |
| **3** | 2018-02-08 | 267.079987 | 267.619995 | 250.000000 | 250.100006 | 250.100006 | 9306700 |
| **4** | 2018-02-09 | 253.850006 | 255.800003 | 236.110001 | 249.470001 | 249.470001 | 16906900 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **1004** | 2022-01-31 | 401.970001 | 427.700012 | 398.200012 | 427.140015 | 427.140015 | 20047500 |
| **1005** | 2022-02-01 | 432.959991 | 458.480011 | 425.540009 | 457.130005 | 457.130005 | 22542300 |
| **1006** | 2022-02-02 | 448.250000 | 451.980011 | 426.480011 | 429.480011 | 429.480011 | 14346000 |
| **1007** | 2022-02-03 | 421.440002 | 429.260010 | 404.279999 | 405.600006 | 405.600006 | 9905200 |
| **1008** | 2022-02-04 | 407.309998 | 412.769989 | 396.640015 | 410.170013 | 410.170013 | 7782400 |

1009 rows × 7 columns

```
df=df.sort_values(by='Date',ascending=True)
```

```
df
```

| | Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|---|
| **0** | 2018-02-05 | 262.000000 | 267.899994 | 250.029999 | 254.259995 | 254.259995 | 11896100 |
| **1** | 2018-02-06 | 247.699997 | 266.700012 | 245.000000 | 265.720001 | 265.720001 | 12595800 |
| **2** | 2018-02-07 | 266.579987 | 272.450012 | 264.329987 | 264.559998 | 264.559998 | 8981500 |
| **3** | 2018-02-08 | 267.079987 | 267.619995 | 250.000000 | 250.100006 | 250.100006 | 9306700 |
| **4** | 2018-02-09 | 253.850006 | 255.800003 | 236.110001 | 249.470001 | 249.470001 | 16906900 |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **1004** | 2022-01-31 | 401.970001 | 427.700012 | 398.200012 | 427.140015 | 427.140015 | 20047500 |
| **1005** | 2022-02-01 | 432.959991 | 458.480011 | 425.540009 | 457.130005 | 457.130005 | 22542300 |
| **1006** | 2022-02-02 | 448.250000 | 451.980011 | 426.480011 | 429.480011 | 429.480011 | 14346000 |
| **1007** | 2022-02-03 | 421.440002 | 429.260010 | 404.279999 | 405.600006 | 405.600006 | 9905200 |
| **1008** | 2022-02-04 | 407.309998 | 412.769989 | 396.640015 | 410.170013 | 410.170013 | 7782400 |

1009 rows × 7 columns

```
x=df.drop('Volume',axis=1)
y=df['Volume']
```

```
x
```

| | Date | Open | High | Low | Close | Adj Close |
|---|---|---|---|---|---|---|
| **0** | 2018-02-05 | 262.000000 | 267.899994 | 250.029999 | 254.259995 | 254.259995 |
| **1** | 2018-02-06 | 247.699997 | 266.700012 | 245.000000 | 265.720001 | 265.720001 |
| **2** | 2018-02-07 | 266.579987 | 272.450012 | 264.329987 | 264.559998 | 264.559998 |
| **3** | 2018-02-08 | 267.079987 | 267.619995 | 250.000000 | 250.100006 | 250.100006 |
| **4** | 2018-02-09 | 253.850006 | 255.800003 | 236.110001 | 249.470001 | 249.470001 |
| **...** | ... | ... | ... | ... | ... | ... |
| **1004** | 2022-01-31 | 401.970001 | 427.700012 | 398.200012 | 427.140015 | 427.140015 |
| **1005** | 2022-02-01 | 432.959991 | 458.480011 | 425.540009 | 457.130005 | 457.130005 |
| **1006** | 2022-02-02 | 448.250000 | 451.980011 | 426.480011 | 429.480011 | 429.480011 |
| **1007** | 2022-02-03 | 421.440002 | 429.260010 | 404.279999 | 405.600006 | 405.600006 |
| **1008** | 2022-02-04 | 407.309998 | 412.769989 | 396.640015 | 410.170013 | 410.170013 |

1009 rows × 6 columns

```
..
```

```
y
```

```
0       11896100
1       12595800
2        8981500
3        9306700
4       16906900
          ...
1004    20047500
1005    22542300
1006    14346000
1007     9905200
1008     7782400
Name: Volume, Length: 1009, dtype: int64
```

```
from sklearn.model_selection import train_test_split

x_train,x_valid_test,y_train,y_valid_test=train_test_split(x,y,test_size=0.4)


x_valid,x_test,y_valid,y_test=train_test_split(x_valid_test,y_valid_test,test_size=0.5)

x_train,y_train=x[:800],y[:800]

x_valid,y_valid=x[801:901],y[801:901]


x_test,t_valid=x[902:],y[902:]

print(len(x_train),len(x_valid),len(x_test))
```

```
800 100 107
```