In [1]:
```python
import pandas as pd
```

In [21]:
```python
# Read the csv file

df = pd.read_csv("D:\\Data Analytics Current Batch\\4_April_2024_batch3\\pand
df
```

Out[21]:

	Date	Closing price	Return
0	1/1/2020	100	0.010000
1	2/1/2020	120	0.200000
2	3/1/2020	130	0.083333
3	4/1/2020	98	-0.246154
4	5/1/2020	50	-0.489796
5	6/1/2020	102	1.040000
6	7/1/2020	104	0.019608
7	8/1/2020	150	0.442308
8	9/1/2020	160	0.066667
9	10/1/2020	109	-0.318750
10	11/1/2020	95	-0.128440

In [5]:
```python
df.head(1)
```

Out[5]:

	Date	Closing price	Return
0	1/1/2020	100	0.01

In [6]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 3 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Date           11 non-null     object
 1   Closing price  11 non-null     int64
 2   Return         11 non-null     float64
dtypes: float64(1), int64(1), object(1)
memory usage: 392.0+ bytes
```

In [7]:
```python
df.describe()
```

Out[7]:

	Closing price	Return
count	11.000000	11.000000
mean	110.727273	0.061707
std	29.604361	0.413015
min	50.000000	-0.489796
25%	99.000000	-0.187297
50%	104.000000	0.019608
75%	125.000000	0.141667
max	160.000000	1.040000

In [9]:
```python
df['Closing price']
```

Out[9]:
```
0      100
1      120
2      130
3       98
4       50
5      102
6      104
7      150
8      160
9      109
10      95
Name: Closing price, dtype: int64
```

In [10]:
```python
df.get('Closing price')
```

Out[10]:
```
0      100
1      120
2      130
3       98
4       50
5      102
6      104
7      150
8      160
9      109
10      95
Name: Closing price, dtype: int64
```

In [12]: `df.Return`

Out[12]:
```
0      0.010000
1      0.200000
2      0.083333
3     -0.246154
4     -0.489796
5      1.040000
6      0.019608
7      0.442308
8      0.066667
9     -0.318750
10    -0.128440
Name: Return, dtype: float64
```

In [15]: `df.loc[[0,1]]`

Out[15]:

	Date	Closing price	Return
0	1/1/2020	100	0.01
1	2/1/2020	120	0.20

In [ ]:

In [16]: `df.iat[1,1]`

Out[16]: 120

In [19]: `df.at[1,'Closing price']`

Out[19]: 120

In [ ]: `# what is the difference between iat and at ?`

In [20]:
```
df.drop(columns = 'Closing price')
```

Out[20]:

	Date	Return
0	1/1/2020	0.010000
1	2/1/2020	0.200000
2	3/1/2020	0.083333
3	4/1/2020	-0.246154
4	5/1/2020	-0.489796
5	6/1/2020	1.040000
6	7/1/2020	0.019608
7	8/1/2020	0.442308
8	9/1/2020	0.066667
9	10/1/2020	-0.318750
10	11/1/2020	-0.128440

In [ ]:
```
### Set Index
```

In [22]:
```
df
```

Out[22]:

	Date	Closing price	Return
0	1/1/2020	100	0.010000
1	2/1/2020	120	0.200000
2	3/1/2020	130	0.083333
3	4/1/2020	98	-0.246154
4	5/1/2020	50	-0.489796
5	6/1/2020	102	1.040000
6	7/1/2020	104	0.019608
7	8/1/2020	150	0.442308
8	9/1/2020	160	0.066667
9	10/1/2020	109	-0.318750
10	11/1/2020	95	-0.128440

In [23]:
```
df.index
```

Out[23]: RangeIndex(start=0, stop=11, step=1)

```
In [24]: df.set_index('Date',inplace=True)
```

```
In [25]: df
```

Out[25]:

Date	Closing price	Return
1/1/2020	100	0.010000
2/1/2020	120	0.200000
3/1/2020	130	0.083333
4/1/2020	98	-0.246154
5/1/2020	50	-0.489796
6/1/2020	102	1.040000
7/1/2020	104	0.019608
8/1/2020	150	0.442308
9/1/2020	160	0.066667
10/1/2020	109	-0.318750
11/1/2020	95	-0.128440

```
In [26]: df.index
```

```
Out[26]: Index(['1/1/2020', '2/1/2020', '3/1/2020', '4/1/2020', '5/1/2020', '6/1/202
         0',
                '7/1/2020', '8/1/2020', '9/1/2020', '10/1/2020', '11/1/2020'],
               dtype='object', name='Date')
```

```
In [27]: df.loc['1/1/2020']
```

```
Out[27]: Closing price    100.00
         Return             0.01
         Name: 1/1/2020, dtype: float64
```

```
In [28]: df.reset_index(inplace=True)
```

In [29]: `df`

Out[29]:

	Date	Closing price	Return
0	1/1/2020	100	0.010000
1	2/1/2020	120	0.200000
2	3/1/2020	130	0.083333
3	4/1/2020	98	-0.246154
4	5/1/2020	50	-0.489796
5	6/1/2020	102	1.040000
6	7/1/2020	104	0.019608
7	8/1/2020	150	0.442308
8	9/1/2020	160	0.066667
9	10/1/2020	109	-0.318750
10	11/1/2020	95	-0.128440

In [30]: `df.index`

Out[30]: `RangeIndex(start=0, stop=11, step=1)`

In [ ]: `### Missing value`

In [33]: `df_ = pd.read_csv('dataset\\weather.csv')`

In [34]: `df_`

Out[34]:

	Day	Temperature	Windspeed	Event
0	01-09-2023	32.0	6.0	Rain
1	01-09-2023	32.0	6.0	Rain
2	02-09-2023	NaN	7.0	Snow
3	03-09-2023	28.0	8.0	NaN
4	04-09-2023	NaN	NaN	Sunny
5	05-09-2023	33.0	9.0	NaN
6	06-09-2023	NaN	NaN	Cloudy
7	07-09-2023	36.0	11.0	Rain
8	07-09-2023	36.0	11.0	Rain

In [ ]:
```
## Delete

    ## drop
    ## dropna
    ### drop_duplicates

## FillIn

    # FillIn
```

In [38]:
```
df_.drop_duplicates(inplace=True)
```

In [39]:
```
df_
```

Out[39]:

	Day	Temperature	Windspeed	Event
0	01-09-2023	32.0	6.0	Rain
2	02-09-2023	NaN	7.0	Snow
3	03-09-2023	28.0	8.0	NaN
4	04-09-2023	NaN	NaN	Sunny
5	05-09-2023	33.0	9.0	NaN
6	06-09-2023	NaN	NaN	Cloudy
7	07-09-2023	36.0	11.0	Rain

In [40]:
```
df_.dropna()
```

Out[40]:

	Day	Temperature	Windspeed	Event
0	01-09-2023	32.0	6.0	Rain
7	07-09-2023	36.0	11.0	Rain

In [41]:
```
df_.drop(columns= 'Day')
```

Out[41]:

	Temperature	Windspeed	Event
0	32.0	6.0	Rain
2	NaN	7.0	Snow
3	28.0	8.0	NaN
4	NaN	NaN	Sunny
5	33.0	9.0	NaN
6	NaN	NaN	Cloudy
7	36.0	11.0	Rain

In [42]: df

Out[42]:

	Day	Temperature	Windspeed	Event
0	01-09-2023	32.0	6.0	Rain
2	02-09-2023	NaN	7.0	Snow
3	03-09-2023	28.0	8.0	NaN
4	04-09-2023	NaN	NaN	Sunny
5	05-09-2023	33.0	9.0	NaN
6	06-09-2023	NaN	NaN	Cloudy
7	07-09-2023	36.0	11.0	Rain

In [ ]: