L2

Pip is a tool that installs Python packages. We can install pandas like below.

In [ ]:
```
#!pip install pandas
```

Now, we can import pandas. It's a widespread convention to import pandas with alias as pd.

In [13]:
```
import pandas as pd
import numpy as np
```

`read_csv()` method from pandas reads a tabular data into a pandas DataFrame. DataFrame is a data structure that pandas uses that has rows and columns and their corresponding labels

In [5]:
```
DIS = pd.read_csv('../Data/DIS.csv')
```

Check the DataFrame if it was read correctly

In [6]:
```
DIS
```

Out[6]:

	Date	Open	High	Low	Close	Adj Close	Volume
0	1/2/1962	0.092908	0.096026	0.092908	0.092908	0.035517	817400
1	1/3/1962	0.092908	0.094467	0.092908	0.094155	0.035994	778500
2	1/4/1962	0.094155	0.094467	0.093532	0.094155	0.035994	934200
3	1/5/1962	0.094155	0.094779	0.093844	0.094467	0.036113	934200
4	1/8/1962	0.094467	0.095714	0.092285	0.094155	0.035994	1245600
...	...	...	...	...	...	...	...
14594	12/23/2019	145.910004	146.330002	144.330002	144.679993	144.679993	9314000
14595	12/24/2019	144.580002	145.429993	144.449997	145.289993	145.289993	3508500
14596	12/26/2019	145.399994	145.860001	145.169998	145.699997	145.699997	4422000
14597	12/27/2019	146.050003	146.509995	145.449997	145.750000	145.750000	5495300
14598	12/30/2019	145.750000	145.869995	143.399994	143.770004	143.770004	6602800

numpy.dtype shows data type. Data type 'o' is an object.

In [16]:
```
DIS.Date.dtype
```

Out[16]:
```
dtype('O')
```

Make a column with name 'date_format'. 'to_datetime' converts argument to datetime. Inside the function 'to_datetime', give 'format'. You can specify the format of the date you use. For example,

'%d/%m/%Y' or '%m/%d/%Y'.

In [9]:
```
1 DIS['date_format'] = pd.to_datetime(DIS['Date'], format="%m/%d/%Y")
```

Let's chect date_format column in DIS data.

In [10]:
```
1 DIS.date_format
```

Out[10]:
```
0       1962-01-02
1       1962-01-03
2       1962-01-04
3       1962-01-05
4       1962-01-08
           ...
14594   2019-12-23
14595   2019-12-24
14596   2019-12-26
14597   2019-12-27
14598   2019-12-30
Name: date_format, Length: 14599, dtype: datetime64[ns]
```

Let's check data type of date_format. '<M8[ns]' is a specific type of datetime.

In [11]:
```
1 DIS.date_format.dtype
```

Out[11]:
```
dtype('<M8[ns]')
```

Let's export csv data to file. 'index = False' will let you export the dataframe without index.

In [12]:
```
1 DIS.to_csv("DIS_data.csv", index = False)
```

In [ ]:
```
1
```