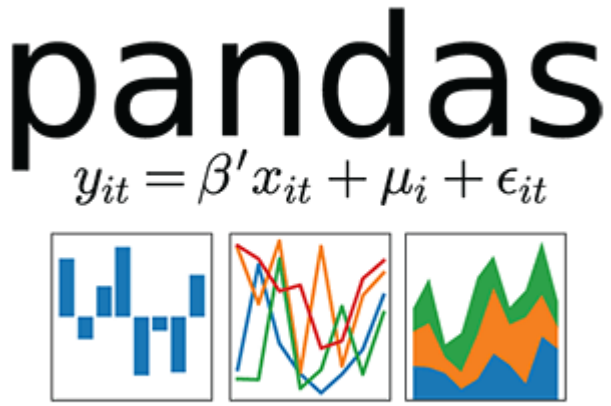

Outlines

- What is Pandas?
- What is Data Science?
- Pandas installation
- What is DataFrame?
- DataFrame Basics

What is Pandas

pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. Pandas is a python module that makes data science easy and effective. Click here for more information (https://pandas.pydata.org/)

What is Data Science?

Data science or data analytics is a process of analyzing large set of data points to get answers or questions related to that data set.

Pandas installation

Go to cmd and type pip install pandas

What is DataFrame?

Dataframe is a main object in Pandas. It is used to represent data with rows and columns (tabular or excel spreadsheet like data).

DataFrame Basics

In [2]:
```
# import pandas library
import pandas as pd
```

In [4]:
```
# Creating DataFrame
df = pd.read_csv('D:/Data_Science/My Github/Pandas-tutorial/Document/weather_data.
df
```

Out[4]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

In [5]:
```
# number of rows and columns
df.shape
```

Out[5]:
```
(6, 4)
```

In [6]:
```
# printing only few rows
df.head()
```

Out[6]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain

In [7]:
```
df.head(2)
```

Out[7]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny

In [8]:
```
# print the last 5 rows
df.tail()
```

Out[8]:

	day	temperature	windspeed	event
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

In [9]:
```
df.tail(2)
```

Out[9]:

	day	temperature	windspeed	event
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

In [10]:
```
# print row number 2 to 4
df[2:5]
```

Out[10]:

	day	temperature	windspeed	event
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain

In [11]:
```
# printing columns
df.columns
```

Out[11]:
```
Index(['day', 'temperature', 'windspeed', 'event'], dtype='object')
```

In [12]:
```
# print the individual column
df.day
```

Out[12]:
```
0    1/1/2017
1    1/2/2017
2    1/3/2017
3    1/4/2017
4    1/5/2017
5    1/6/2017
Name: day, dtype: object
```

In [13]:
```
df['event']
```

Out[13]:
```
0     Rain
1    Sunny
2     Snow
3     Snow
4     Rain
5    Sunny
Name: event, dtype: object
```

In [14]:
```
type(df['event'])
```

Out[14]:
```
pandas.core.series.Series
```

In [15]:
```
# Sometimes your DataFrame will have too many columns and you want to print only c
df[['event','day']]
```

Out[15]:

	event	day
0	Rain	1/1/2017
1	Sunny	1/2/2017
2	Snow	1/3/2017
3	Snow	1/4/2017
4	Rain	1/5/2017
5	Sunny	1/6/2017

In [16]:
```
# What was the maximum temperature?
df['temperature'].max()
```

Out[16]:
```
35
```

In [17]:
```
df['temperature'].mean()
```

Out[17]:
```
30.333333333333332
```

In [18]:
```
df['temperature'].std()
```

Out[18]:
```
3.8297084310253524
```

In [19]:
```
# print the statistics of Data set
df.describe()
```

Out[19]:

	temperature	windspeed
count	6.000000	6.000000
mean	30.333333	4.666667
std	3.829708	2.338090
min	24.000000	2.000000
25%	28.750000	2.500000
50%	31.500000	5.000000
75%	32.000000	6.750000
max	35.000000	7.000000

In [20]:
```
# How to conditionally select the data in DataFrame
df[df.temperature>=32]
```

Out[20]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
4	1/5/2017	32	4	Rain

In [21]:
```
df[df.temperature==df.temperature.max()]
```

Out[21]:

	day	temperature	windspeed	event
1	1/2/2017	35	7	Sunny

In [22]:
```
# What is the day when your temperature was maximum?
df['day'][df.temperature==df.temperature.max()]
```

Out[22]:
```
1    1/2/2017
Name: day, dtype: object
```

In [25]:
```
# What is the day and temperatre when your temperature was maximum?
df[['day','temperature']][df.temperature==df.temperature.max()]
```

Out[25]:

	day	temperature
1	1/2/2017	35

[Click here for more information about pandas series operations (https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.html)](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.html)

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

Out[26]:
```
RangeIndex(start=0, stop=6, step=1)
```

In [27]:
```
# Change the index from number to day
df.set_index('day')
```

Out[27]:

day	temperature	windspeed	event
1/1/2017	32	6	Rain
1/2/2017	35	7	Sunny
1/3/2017	28	2	Snow
1/4/2017	24	7	Snow
1/5/2017	32	4	Rain
1/6/2017	31	2	Sunny

In [28]:
```
df
```

Out[28]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

In [29]:
```
# with df.set_index('day') my original DataFrame is not modified with the syntax b
df.set_index('day',inplace=True)
```

In [31]:
```
# reset index
df.reset_index(inplace=True)
df
```

Out[31]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

Date	Author
2021-07-28	**Ehsan Zia**