Introduction to Pandas dataframe

Data frame is a main object in pandas. It is used to represent data with rows and columns

Data frame is a datastructure represent the data in tabular or excel spread sheet like data)

creating dataframe:

In [0]:
```
import pandas as pd
df = pd.read_csv("weather_data.csv")   #read weather.csv data
df
```

Out[0]:

	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 [0]:
```
#list of tuples

weather_data = [('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')
               ]
df = pd.DataFrame(weather_data, columns=['day', 'temperature', 'windspeed', 'e
vent'])
df
```

Out[0]:

	day	temp	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 [0]:
```
#get dimentions of the table

df.shape   #total number of rows and columns
```

Out[0]: (6, 4)

In [0]:
```
#if you want to see initial some rows then use head command (default 5 rows)
df.head()
```

Out[0]:

	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 [0]:
```
#if you want to see last few rows then use tail command (default last 5 rows w
ill print)
df.tail()
```

Out[0]:

	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 [0]:
```
#slicing
df[2:5]
```

Out[0]:

	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 [0]:
```
df.columns   #print columns in a table
```

Out[0]:
```
Index(['day', 'temperature', 'windspeed', 'event'], dtype='object')
```

In [0]:
```
df.day      #print particular column data
```

Out[0]:
```
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 [0]:
```
#another way of accessing column
df['day'] #df.day (both are same)
```

Out[0]:
```
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 [0]:
```
#get 2 or more columns
df[['day', 'event']]
```

Out[0]:

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

In [0]:
```
#get all temperatures
df['temperature']
```

Out[0]:
```
0    32
1    35
2    28
3    24
4    32
5    31
Name: temperature, dtype: int64
```

In [0]:
```
#print max temperature
df['temperature'].max()
```

Out[0]:
```
35
```

In [0]:
```
#print max temperature
df['temperature'].min()
```

Out[0]:
```
24
```

In [0]:
```
#print max temperature
df['temperature'].describe()
```

Out[0]:
```
count     6.000000
mean     30.333333
std       3.829708
min      24.000000
25%      28.750000
50%      31.500000
75%      32.000000
max      35.000000
Name: temperature, dtype: float64
```

In [0]:
```
# select rows which has maximum temperature
df[df.temperature == df.temperature.max()]
```

Out[0]:

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

In [0]:
```
#select only day column which has maximum temperature
df.day[df.temperature == df.temperature.max()]
```

Out[0]:
```
1    1/2/2017
Name: day, dtype: object
```