Learning Pandas

> link for the course : https://bit.ly/3vmaoHe

```
In [ ]:    #installing pandas and Importing it
           # pip install pandas

           import pandas as pd
```

```
In [ ]:    #importing the data fram

           df = pd.read_csv('D:\\Study\\DataScience\\py\\pandas\\1_intro\\nyc_weather.csv')
           df
```

Out[ ]:

	EST	Temperature	DewPoint	Humidity	Sea Level PressureIn	VisibilityMiles	WindSpeedMPH	Precipitat
0	1/1/2016	38	23	52	30.03	10	8.0	
1	1/2/2016	36	18	46	30.02	10	7.0	
2	1/3/2016	40	21	47	29.86	10	8.0	
3	1/4/2016	25	9	44	30.05	10	9.0	
4	1/5/2016	20	-3	41	30.57	10	5.0	
5	1/6/2016	33	4	35	30.50	10	4.0	
6	1/7/2016	39	11	33	30.28	10	2.0	
7	1/8/2016	39	29	64	30.20	10	4.0	
8	1/9/2016	44	38	77	30.16	9	8.0	
9	1/10/2016	50	46	71	29.59	4	NaN	
10	1/11/2016	33	8	37	29.92	10	NaN	
11	1/12/2016	35	15	53	29.85	10	6.0	
12	1/13/2016	26	4	42	29.94	10	10.0	
13	1/14/2016	30	12	47	29.95	10	5.0	
14	1/15/2016	43	31	62	29.82	9	5.0	
15	1/16/2016	47	37	70	29.52	8	7.0	
16	1/17/2016	36	23	66	29.78	8	6.0	
17	1/18/2016	25	6	53	29.83	9	12.0	
18	1/19/2016	22	3	42	30.03	10	11.0	
19	1/20/2016	32	15	49	30.13	10	6.0	
20	1/21/2016	31	11	45	30.15	10	6.0	

	EST	Temperature	DewPoint	Humidity	Sea Level PressureIn	VisibilityMiles	WindSpeedMPH	Precipitat
21	1/22/2016	26	6	41	30.21	9	NaN	
22	1/23/2016	26	21	78	29.77	1	16.0	
23	1/24/2016	28	11	53	29.92	8	6.0	
24	1/25/2016	34	18	54	30.25	10	3.0	
25	1/26/2016	43	29	56	30.03	10	7.0	
26	1/27/2016	41	22	45	30.03	10	7.0	
27	1/28/2016	37	20	51	29.90	10	5.0	
28	1/29/2016	36	21	50	29.58	10	8.0	
29	1/30/2016	34	16	46	30.01	10	7.0	
30	1/31/2016	46	28	52	29.90	10	5.0	

In [ ]:
```python
#To find the maximinm temprature
df['Temperature'].max()
```

Out[ ]:
```
50
```

In [ ]:
```python
#On which dates there was raining?

df['EST'][df['Events'] == 'Rain']
```

Out[ ]:
```
8      1/9/2016
9      1/10/2016
15     1/16/2016
26     1/27/2016
Name: EST, dtype: object
```

In [ ]:
```python
#what was the average wind speed

print(df['WindSpeedMPH'].mean())

#this is because there was NAN values in dataset, Now we have to do Data munging or Dat

df['WindSpeedMPH'].fillna(0, inplace=True)
df
```

```
6.225806451612903
```

Out[ ]:
	EST	Temperature	DewPoint	Humidity	Sea Level PressureIn	VisibilityMiles	WindSpeedMPH	Precipitat
0	1/1/2016	38	23	52	30.03	10	8.0	
1	1/2/2016	36	18	46	30.02	10	7.0	
2	1/3/2016	40	21	47	29.86	10	8.0	

	EST	Temperature	DewPoint	Humidity	Sea Level PressureIn	VisibilityMiles	WindSpeedMPH	Precipitat
3	1/4/2016	25	9	44	30.05	10	9.0	
4	1/5/2016	20	-3	41	30.57	10	5.0	
5	1/6/2016	33	4	35	30.50	10	4.0	
6	1/7/2016	39	11	33	30.28	10	2.0	
7	1/8/2016	39	29	64	30.20	10	4.0	
8	1/9/2016	44	38	77	30.16	9	8.0	
9	1/10/2016	50	46	71	29.59	4	0.0	
10	1/11/2016	33	8	37	29.92	10	0.0	
11	1/12/2016	35	15	53	29.85	10	6.0	
12	1/13/2016	26	4	42	29.94	10	10.0	
13	1/14/2016	30	12	47	29.95	10	5.0	
14	1/15/2016	43	31	62	29.82	9	5.0	
15	1/16/2016	47	37	70	29.52	8	7.0	
16	1/17/2016	36	23	66	29.78	8	6.0	
17	1/18/2016	25	6	53	29.83	9	12.0	
18	1/19/2016	22	3	42	30.03	10	11.0	
19	1/20/2016	32	15	49	30.13	10	6.0	
20	1/21/2016	31	11	45	30.15	10	6.0	
21	1/22/2016	26	6	41	30.21	9	0.0	
22	1/23/2016	26	21	78	29.77	1	16.0	
23	1/24/2016	28	11	53	29.92	8	6.0	
24	1/25/2016	34	18	54	30.25	10	3.0	
25	1/26/2016	43	29	56	30.03	10	7.0	
26	1/27/2016	41	22	45	30.03	10	7.0	
27	1/28/2016	37	20	51	29.90	10	5.0	
28	1/29/2016	36	21	50	29.58	10	8.0	
29	1/30/2016	34	16	46	30.01	10	7.0	
30	1/31/2016	46	28	52	29.90	10	5.0	

In [ ]:
```python
#now the data in WindSpeedMPH Column is cleaned now we can find the average speed

df['WindSpeedMPH'].mean()
```

Out[ ]: 6.225806451612903

In [ ]:

Out[ ]: 6.225806451612903

In [ ]:

Learning the DataFrame in Pandas

In [ ]:
```python
# importing the pandas

import pandas as pd
```

In [ ]:
```python
# Importing the dataset

df = pd.read_csv("D:\\Study\DataScience\\Learning_Datascience\\Learning_Pandas\\Codebas
df
```

Out[ ]:

	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 [ ]:
```python
#Shape of the dataframe

df.shape

rows, cols = df.shape
```

In [ ]:
```python
print('Rows : ', rows, ' Cols: ', cols)
```

Rows : 6 Cols: 4

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

Out[ ]:

	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 [ ]:
```python
df.tail()
```

```
Out[ ]:
```

	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 [ ]:
df[2:5]
```

```
Out[ ]:
```

	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 [ ]:
#headings of the cols

df.columns
```

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

```
In [ ]:
df.day
```

```
Out[ ]:
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 [ ]:
df['day']
```

```
Out[ ]:
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 [ ]:
type(df['event'])
```

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

```
In [ ]:
type(df)
```

Out[ ]: `pandas.core.frame.DataFrame`

In [ ]:
```python
#for printing custome columns

df[['event', 'day']]
```

Out[ ]:
	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 [ ]:
```python
# now we are doing the analysis

# Now we can find the Max/Min temperatures and also the average temperature

maximumTemperature = df['temperature'].max()
minimumTemperature = df['temperature'].min()
averageTemperature = df['temperature'].mean()

print('Max Temp: ', maximumTemperature,'\nMin Temp: ', minimumTemperature, '\nAvg Temp:
```

```
Max Temp:  35
Min Temp:  24
Avg Temp:  30.333333333333332
```

In [ ]:
```python
df
```

Out[ ]:
	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 [ ]:
```python
df[['day', 'temperature']][df['event'] == 'Sunny']
```

Out[ ]:
	day	temperature
1	1/2/2017	35
5	1/6/2017	31

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

Out[ ]:

	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 [ ]:   df
```

Out[ ]:

	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 [ ]:   # as we can see the we have index staring form 0 to 5 but if we want that our dataframe
          # to the day column we can do it so that we can give a date as an index and recieve the

          # we can see the index by

          df.index
```

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

```
In [ ]:   #to set the index of the df to day

          df.set_index('day', inplace=True)
          df
```

Out[ ]:

	temperature	windspeed	event
day			
1/1/2017	32	6	Rain

	temperature	windspeed	event
day			
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 [ ]:
```python
df.loc['1/1/2017']
```

Out[ ]:
```
temperature       32
windspeed          6
event           Rain
Name: 1/1/2017, dtype: object
```

In [ ]:
```python
# Now to reset the indexes to 0 to 5

df.reset_index(inplace = True)
df
```

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

In [ ]: