Project- 3
Data analysis from a weather dataset by Pandas

Nur Hossain Shawon

Data Analyst

Data Preprocessing

```
data.info()
```

✓ 0.1s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8784 entries, 0 to 8783
Data columns (total 8 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Date/Time           8784 non-null   object
 1   Temp_C              8784 non-null   float64
 2   Dew Point Temp_C    8784 non-null   float64
 3   Rel Hum_%           8784 non-null   int64
 4   Wind Speed_km/h     8784 non-null   int64
 5   Visibility_km       8784 non-null   float64
 6   Press_kPa           8784 non-null   float64
 7   Weather condition   8784 non-null   object
dtypes: float64(4), int64(2), object(2)
memory usage: 549.1+ KB
```

Find all the unique 'Wind speed' values in the data.

```
data.nunique()
```

[80] ✓ 0.0s

```
Date/Time                  8784
Temp_C                      533
Dew Point Temp_C            489
Rel Hum_%                    83
Wind Speed_km/h              34
Visibility_km                24
Press_kPa                   518
Weather condition            50
dtype: int64
```

Find the number of times when the 'Weather is exactly clear'

```
#Groupby
data.groupby('Weather condition').get_group('Clear')
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Clear
114	1/5/2012 18:00	-7.1	-14.4	56	11	25.0	100.71	Clear
115	1/5/2012 19:00	-9.2	-15.4	61	7	25.0	100.80	Clear
116	1/5/2012 20:00	-9.8	-15.7	62	9	25.0	100.83	Clear
117	1/5/2012 21:00	-9.0	-14.8	63	13	25.0	100.83	Clear
...	...	...	...	...	...	...	...	...
8646	12/26/2012 6:00	-13.4	-14.8	89	4	25.0	102.47	Clear
8698	12/28/2012 10:00	-6.1	-8.6	82	19	24.1	101.27	Clear
8713	12/29/2012 1:00	-11.9	-13.6	87	11	25.0	101.31	Clear
8714	12/29/2012 2:00	-11.8	-13.1	90	13	25.0	101.33	Clear
8756	12/30/2012 20:00	-13.8	-16.5	80	24	25.0	101.52	Clear

Find the number of time when the wind speed was exactly 4 KM/H

```
data.groupby('Wind Speed_km/h').get_group(4)
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog
96	1/5/2012 0:00	-8.8	-11.7	79	4	9.7	100.32	Snow
101	1/5/2012 5:00	-7.0	-9.5	82	4	4.0	100.19	Snow
146	1/7/2012 2:00	-8.1	-11.1	79	4	19.3	100.15	Cloudy
...	...	...	...	...	...	...	...	...
8768	12/31/2012 8:00	-8.6	-10.3	87	4	3.2	101.14	Snow Showers
8769	12/31/2012 9:00	-8.1	-9.6	89	4	2.4	101.09	Snow
8770	12/31/2012 10:00	-7.4	-8.9	89	4	6.4	101.05	Snow,Fog
8772	12/31/2012 12:00	-5.8	-7.5	88	4	12.9	100.78	Snow
8773	12/31/2012 13:00	-4.6	-6.6	86	4	12.9	100.63	Snow

Find the null values in the data

```
data.isnull().sum()
```

✓ 0.0s

```
Date/Time                0
Temp_C                   0
Dew Point Temp_C         0
Rel Hum_%                0
Wind Speed_km/h          0
Visibility_km            0
Press_kPa                0
Weather condition        0
dtype: int64
```

Rename the column name Weather of the data frame to weather condition

```
data.rename(columns= {'Weather' : 'Weather condition'}, inplace=True)
data
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog
2	1/1/2012 2:00	-1.8	-3.4	89	7	4.0	101.26	Freezing Drizzle,Fog
3	1/1/2012 3:00	-1.5	-3.2	88	6	4.0	101.27	Freezing Drizzle,Fog
4	1/1/2012 4:00	-1.5	-3.3	88	7	4.8	101.23	Fog
...	...	...	...	...	...	...	...	...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/31/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	12/31/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/31/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow
8783	12/31/2012 [illegible]	0.0	-2.1	86	30	11.3	99.89	Snow

What is the called visibility column?

```
data.Visibility_km.mean()
#Finding mean in a column
```

✓ 0.0s

```
27.664446721311478
```

What is the standard deviation of 'Pressure' in the data?

```
data.Press_kPa.std()
```

✓ 0.0s

```
0.8440047459486483
```

What is the variance of relative humidity in the data?

```
data['Rel Hum_%'].var()
#If there is no space in column name we may use dataframe.column name.
#BUt if there is any space we should use dataframe['column name]
```

✓ 0.0s

```
286.24855019850196
```

Q9) Find all instance when 'Snow' was recorded.**

```
data[data['Weather condition'].str.contains('Snow')]
#There is lots of complex data like snow ice pellet, snow bowling snow.
#str.contain can collect all snow data from there
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
41	1/2/2012 17:00	-2.1	-9.5	57	22	25.0	99.66	Snow Showers
44	1/2/2012 20:00	-5.6	-13.4	54	24	25.0	100.07	Snow Showers
45	1/2/2012 21:00	-5.8	-12.8	58	26	25.0	100.15	Snow Showers
47	1/2/2012 23:00	-7.4	-14.1	59	17	19.3	100.27	Snow Showers
48	1/3/2012 0:00	-9.0	-16.0	57	28	25.0	100.35	Snow Showers
...	...	...	...	...	...	...	...	...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.13	Snow
8780	12/31/2012 20:00	0.2	-2.4	83	24	9.7	100.03	Snow
8781	12/31/2012 21:00	-0.5	-1.5	93	28	4.8	99.95	Snow
8782	12/31/2012 22:00	-0.2	-1.8	89	28	9.7	99.91	Snow

Find all instance when 'Wind Speed is above 24' and 'Visibility is 25.

```python
data[(data['Wind Speed_km/h'] >24) & (data['Visibility_km']==25)]
#At first two condition was specified and then filtered the data
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
23	1/1/2012 23:00	5.3	2.0	79	30	25.0	99.31	Cloudy
24	1/2/2012 0:00	5.2	1.5	77	35	25.0	99.26	Rain Showers
25	1/2/2012 1:00	4.6	0.0	72	39	25.0	99.26	Cloudy
26	1/2/2012 2:00	3.9	-0.9	71	32	25.0	99.26	Mostly Cloudy
27	1/2/2012 3:00	3.7	-1.5	69	33	25.0	99.30	Mostly Cloudy
...	...	...	...	...	...	...	...	...
8705	12/28/2012 17:00	-8.6	-12.0	76	26	25.0	101.34	Mainly Clear
8753	12/30/2012 17:00	-12.1	-15.8	74	28	25.0	101.26	Mainly Clear
8755	12/30/2012 19:00	-13.4	-16.5	77	26	25.0	101.47	Mainly Clear
8759	12/30/2012 23:00	-12.1	-15.1	78	28	25.0	101.52	Mostly Cloudy
8760	12/31/2012 0:00	-11.1	-14.4	77	26	25.0	101.51	Cloudy

What is the min and max value pf each column against each 'Weather Condition'?

```
data.groupby('Weather condition').min()
```

✓ 0.1s Python

Weather condition	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Clear	1/11/2012 1:00	-23.3	-28.5	20	0	11.3	99.52
Cloudy	1/1/2012 17:00	-21.4	-26.8	18	0	11.3	98.39
Drizzle	1/23/2012 21:00	1.1	-0.2	74	0	6.4	97.84
Drizzle,Fog	1/23/2012 20:00	0.0	-1.6	85	0	1.0	98.65
Drizzle,Ice Pellets,Fog	12/17/2012 9:00	0.4	-0.7	92	20	4.0	100.79
Drizzle,Snow	12/17/2012 15:00	0.9	0.1	92	9	9.7	100.63
Drizzle,Snow,Fog	12/18/2012 21:00	0.3	-0.1	92	7	2.4	97.79
Fog	1/1/2012 0:00	-16.0	-17.2	80	0	0.2	98.31
Freezing Drizzle	1/13/2012 10:00	-9.0	-12.2	78	6	4.8	98.44
Freezing Drizzle,Fog	1/1/2012 2:00	-6.4	-9.0	82	6	3.6	98.74

Find all instance when weather is clear or visibility is above 40.

```
data[(data['Weather condition'] == 'Clear') | (data['Visibility_km']>40)]
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Clear
106	1/5/2012 10:00	-6.0	-10.0	73	17	48.3	100.45	Mainly Clear
107	1/5/2012 11:00	-5.6	-10.2	70	22	48.3	100.41	Mainly Clear
108	1/5/2012 12:00	-4.7	-9.6	69	20	48.3	100.38	Mainly Clear
109	1/5/2012 13:00	-4.4	-9.7	66	26	48.3	100.40	Mainly Clear
...	...	...	...	...	...	...	...	...
8749	12/30/2012 13:00	-12.4	-16.2	73	37	48.3	100.92	Mostly Cloudy
8750	12/30/2012 14:00	-11.8	-16.1	70	37	48.3	100.96	Mainly Clear
8751	12/30/2012 15:00	-11.3	-15.6	70	32	48.3	101.05	Mainly Clear
8752	12/30/2012 16:00	-11.4	-15.5	72	26	48.3	101.15	Mainly Clear
8756	12/30/2012 20:00	-13.8	-16.5	80	24	25.0	101.52	Clear

Find all instance when weather is clear and Relative humidity is grater than 50.

```
data[(data['Weather condition']=='Clear') & (data['Rel Hum_%']>50)]
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
114	1/5/2012 18:00	-7.1	-14.4	56	11	25.0	100.71	Clear
115	1/5/2012 19:00	-9.2	-15.4	61	7	25.0	100.80	Clear
116	1/5/2012 20:00	-9.8	-15.7	62	9	25.0	100.83	Clear
117	1/5/2012 21:00	-9.0	-14.8	63	13	25.0	100.83	Clear
241	1/11/2012 1:00	-10.7	-17.8	56	17	25.0	101.49	Clear
...	...	...	...	...	...	...	...	...
8646	12/26/2012 6:00	-13.4	-14.8	89	4	25.0	102.47	Clear
8698	12/28/2012 10:00	-6.1	-8.6	82	19	24.1	101.27	Clear
8713	12/29/2012 1:00	-11.9	-13.6	87	11	25.0	101.31	Clear
8714	12/29/2012 2:00	-11.8	-13.1	90	13	25.0	101.33	Clear
8756	12/30/2012 20:00	-13.8	-16.5	80	24	25.0	101.52	Clear

Q15) Find all instance when weather is clear and visibility is grater than 40.

```
data[(data['Weather condition']=='Clear') & (data['Visibility_km']>40)]
```

✓ 0.0s Python

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather condition
351	1/15/2012 15:00	-15.4	-22.8	53	24	48.3	102.71	Clear
352	1/15/2012 16:00	-15.1	-22.8	52	24	48.3	102.79	Clear
425	1/18/2012 17:00	-11.3	-18.8	54	26	48.3	101.54	Clear
440	1/19/2012 8:00	-13.7	-18.4	68	19	48.3	101.84	Clear
441	1/19/2012 9:00	-12.7	-17.2	69	17	48.3	101.73	Clear
...	...	...	...	...	...	...	...	...
8384	12/15/2012 8:00	-10.7	-15.6	67	13	48.3	102.69	Clear
8385	12/15/2012 9:00	-10.4	-15.9	64	19	48.3	102.74	Clear
8389	12/15/2012 13:00	-8.4	-14.7	60	19	48.3	102.64	Clear
8631	12/25/2012 15:00	-7.1	-13.7	59	17	48.3	101.98	Clear
8632	12/25/2012 16:00	-7.5	-13.9	60	11	48.3	102.03	Clear

Thank You

/nurhossain.shawon.79

/nurhossainshawon/

Nur Hossain Shawon

Data Analyst