Project Weather Data

In [3]:
```
import pandas as pd
```

In [4]:
```
data =  pd.read_csv(r"C:\Users\946851\Downloads\1. Weather Data.csv")
```

1)Find All the wind speed values in the data

In [5]:
```
data['Wind Speed_km/h'].unique()
```

Out[5]:
```
array([ 4,  7,  6,  9, 15, 13, 20, 22, 19, 24, 30, 35, 39, 32, 33, 26, 44,
       43, 48, 37, 28, 17, 11,  0, 83, 70, 57, 46, 41, 52, 50, 63, 54,  2],
      dtype=int64)
```

2)Find Number of times Weather is clear

In [6]:
```
#data['Weather'].value_counts()
#data.groupby('Weather').get_group('Clear')
data[data['Weather'] == 'Clear']
```

Out[6]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
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

1326 rows × 8 columns

3) Find the number of times when the 'Wind Speed was exactly 4 km/h'.

In [7]:
```
data[data['Wind Speed_km/h'] == 4]
```

Out[7]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
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

474 rows × 8 columns

4) Find out all the Null Values in the data.

In [8]:
```
#sum or use count
#data.notnull().sum()
data.isnull().sum()
```

Out[8]:
```
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             0
dtype: int64
```

5)Rename the column name 'Weather' of the dataframe to 'Weather Condition'.

In [9]:
```
data.rename(columns = {'Weather': 'Weather Condition'},inplace=True)
```

6) What is the mean 'Visibility' ?

In [10]:
```
data.Visibility_km.mean()
```

Out[10]:
```
27.664446721311478
```

7) What is the Standard Deviation of 'Pressure' in this data?

In [11]:
```
data.Press_kPa.std()
```

Out[11]:
```
0.8440047459486483
```

8) What is the Variance of 'Relative Humidity' in this data ?

In [12]:
```
data['Rel Hum_%'].var()
```

Out[12]:
```
286.24855019850196
```

9) Find all instances when 'Snow' was recorded.

In [14]:
```
data[data['Weather Condition'] == "Snow"]
```

Out[14]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
55	1/3/2012 7:00	-14.0	-19.5	63	19	25.0	100.95	Snow
84	1/4/2012 12:00	-13.7	-21.7	51	11	24.1	101.25	Snow
86	1/4/2012 14:00	-11.3	-19.0	53	7	19.3	100.97	Snow
87	1/4/2012 15:00	-10.2	-16.3	61	11	9.7	100.89	Snow
88	1/4/2012 16:00	-9.4	-15.5	61	13	19.3	100.79	Snow
...	...	...	...	...	...	...	...	...
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 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

390 rows × 8 columns

10) Find all instances when 'Wind Speed is above 24' and 'Visibility is 25'.

In [15]:
```
data[(data['Wind Speed_km/h'] > 25) & (data['Visibility_km'] == 25)]
```

Out[15]:

	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

308 rows × 8 columns

11) What is the Mean value of each column against each 'Weather Condition ?

In [ ]:
```
data.groupby('Weather Condition').mean()
```

12) What is the Minimum & Maximum value of each column against each 'Weather Condition ?

In [ ]:
```
#data.groupby('Weather Condition').describe()
#data.groupby('Weather Condition').min()
data.groupby('Weather Condition').max()
```

Show all the Records where Weather Condition is Fog.

In [ ]:
```
#data[data['Weather Condition'] == 'Fog']
data['Weather Condition'].value_counts()
```

14) Find all instances when 'Weather is Clear' or 'Visibility is above 40'.

In [28]:
```
data[(data['Weather Condition'] == "Clear") & (data['Visibility_km'] > 40)]
```

Out[28]:

	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

313 rows × 8 columns

In [16]:
```
data.head()
```

Out[16]:

	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

15) Find all instances when :

A. 'Weather is Clear' and 'Relative Humidity is greater than 50' or B. 'Visibility is above 40'

In [29]:
```
data[((data['Weather Condition'] == "Clear") & (data['Rel Hum_%'] > 50))  |  (data['Visibility_km'] > 40)]
```

Out[29]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
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
110	1/5/2012 14:00	-5.1	-10.7	65	22	48.3	100.46	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

2921 rows × 8 columns

In [ ]:
```

```