WEATHER DATASET BIG DATA ANALYSIS #

This dataset is in CSV format and we will solve this dataset using pandas dataFrame.

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

In [4]:
```python
#reading the CSV file through the pandas read command.
df = pd.read_csv(r"E:\MY  Data analytics project\Weather Data.csv")
```

In [5]:
```python
#printing the  imported dataset.
df
```

Out[5]:

	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
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 23:00	0.0	-2.1	86	30	11.3	99.89	Snow

8784 rows × 8 columns

In [6]:
```python
#show the first 10 rows in the data frame.
df.head(10)
```

Out[6]:

	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
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
5	1/1/2012 5:00	-1.4	-3.3	87	9	6.4	101.27	Fog
6	1/1/2012 6:00	-1.5	-3.1	89	7	6.4	101.29	Fog
7	1/1/2012 7:00	-1.4	-3.6	85	7	8.0	101.26	Fog
8	1/1/2012 8:00	-1.4	-3.6	85	9	8.0	101.23	Fog
9	1/1/2012 9:00	-1.3	-3.1	88	15	4.0	101.20	Fog

In [7]:
```python
#printing the shape of the dataframe.
df.shape
```

Out[7]:
```
(8784, 8)
```

In [8]:
```python
#Printing the index of the dataframe.
df.index
```

Out[8]:
```
RangeIndex(start=0, stop=8784, step=1)
```

In [12]:
```python
#showing all the columns name which is present in the dataframe.
df.columns
```

Out[12]:
```
Index(['Date/Time', 'Temp_C', 'Dew Point Temp_C', 'Rel Hum_%',
       'Wind Speed_km/h', 'Visibility_km', 'Press_kPa', 'Weather'],
      dtype='object')
```

In [13]:
```python
#Printing the data type of the dataframe.
df.dtypes
```

Out[13]:
```
Date/Time            object
Temp_C              float64
Dew Point Temp_C    float64
Rel Hum_%             int64
Wind Speed_km/h       int64
Visibility_km       float64
Press_kPa           float64
Weather              object
dtype: object
```

In [14]:
```python
#printing Unique value in the data frame
```

```python
df["Weather"].unique()
```

Out[14]: array(['Fog', 'Freezing Drizzle,Fog', 'Mostly Cloudy', 'Cloudy', 'Rain',
       'Rain Showers', 'Mainly Clear', 'Snow Showers', 'Snow', 'Clear',
       'Freezing Rain,Fog', 'Freezing Rain', 'Freezing Drizzle',
       'Rain,Snow', 'Moderate Snow', 'Freezing Drizzle,Snow',
       'Freezing Rain,Snow Grains', 'Snow,Blowing Snow', 'Freezing Fog',
       'Haze', 'Rain,Fog', 'Drizzle,Fog', 'Drizzle',
       'Freezing Drizzle,Haze', 'Freezing Rain,Haze', 'Snow,Haze',
       'Snow,Fog', 'Snow,Ice Pellets', 'Rain,Haze', 'Thunderstorms,Rain',
       'Thunderstorms,Rain Showers', 'Thunderstorms,Heavy Rain Showers',
       'Thunderstorms,Rain Showers,Fog', 'Thunderstorms',
       'Thunderstorms,Rain,Fog',
       'Thunderstorms,Moderate Rain Showers,Fog', 'Rain Showers,Fog',
       'Rain Showers,Snow Showers', 'Snow Pellets', 'Rain,Snow,Fog',
       'Moderate Rain,Fog', 'Freezing Rain,Ice Pellets,Fog',
       'Drizzle,Ice Pellets,Fog', 'Drizzle,Snow', 'Rain,Ice Pellets',
       'Drizzle,Snow,Fog', 'Rain,Snow Grains', 'Rain,Snow,Ice Pellets',
       'Snow Showers,Fog', 'Moderate Snow,Blowing Snow'], dtype=object)

In [15]:
```python
df.nunique()
```

Out[15]:
```
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             50
dtype: int64
```

In [16]:
```python
df.count()
```

Out[16]:
```
Date/Time          8784
Temp_C             8784
Dew Point Temp_C   8784
Rel Hum_%          8784
Wind Speed_km/h    8784
Visibility_km      8784
Press_kPa          8784
Weather            8784
dtype: int64
```

In [17]:
```python
df.count
```

```
Out[17]:  <bound method DataFrame.count of              Date/Time  Temp_C  Dew Point Temp_C
          Rel Hum_%  Wind Speed_km/h  \
          0           1/1/2012 0:00    -1.8            -3.9       86              4
          1           1/1/2012 1:00    -1.8            -3.7       87              4
          2           1/1/2012 2:00    -1.8            -3.4       89              7
          3           1/1/2012 3:00    -1.5            -3.2       88              6
          4           1/1/2012 4:00    -1.5            -3.3       88              7
          ...                 ...       ...             ...      ...            ...
          8779  12/31/2012 19:00      0.1            -2.7       81             30
          8780  12/31/2012 20:00      0.2            -2.4       83             24
          8781  12/31/2012 21:00     -0.5            -1.5       93             28
          8782  12/31/2012 22:00     -0.2            -1.8       89             28
          8783  12/31/2012 23:00      0.0            -2.1       86             30

                Visibility_km  Press_kPa                Weather
          0               8.0     101.24                    Fog
          1               8.0     101.24                    Fog
          2               4.0     101.26  Freezing Drizzle,Fog
          3               4.0     101.27  Freezing Drizzle,Fog
          4               4.8     101.23                    Fog
          ...             ...        ...                    ...
          8779            9.7     100.13                   Snow
          8780            9.7     100.03                   Snow
          8781            4.8      99.95                   Snow
          8782            9.7      99.91                   Snow
          8783           11.3      99.89                   Snow

          [8784 rows x 8 columns]>
```

In [18]: `df["Weather"].value_counts()`

```
Out[18]:  Mainly Clear                                       2106
          Mostly Cloudy                                      2069
          Cloudy                                             1728
          Clear                                              1326
          Snow                                                390
          Rain                                                306
          Rain Showers                                        188
          Fog                                                 150
          Rain,Fog                                            116
          Drizzle,Fog                                          80
          Snow Showers                                         60
          Drizzle                                              41
          Snow,Fog                                             37
          Snow,Blowing Snow                                    19
          Rain,Snow                                            18
          Thunderstorms,Rain Showers                           16
          Haze                                                 16
          Drizzle,Snow,Fog                                     15
          Freezing Rain                                        14
          Freezing Drizzle,Snow                                11
          Freezing Drizzle                                      7
          Snow,Ice Pellets                                      6
          Freezing Drizzle,Fog                                  6
          Snow,Haze                                             5
          Freezing Fog                                          4
          Snow Showers,Fog                                      4
          Moderate Snow                                         4
          Rain,Snow,Ice Pellets                                 4
          Freezing Rain,Fog                                     4
          Freezing Drizzle,Haze                                 3
          Rain,Haze                                             3
          Thunderstorms,Rain                                    3
          Thunderstorms,Rain Showers,Fog                        3
          Freezing Rain,Haze                                    2
          Drizzle,Snow                                          2
          Rain Showers,Snow Showers                             2
          Thunderstorms                                         2
          Moderate Snow,Blowing Snow                            2
          Rain Showers,Fog                                      1
          Thunderstorms,Moderate Rain Showers,Fog              1
          Snow Pellets                                          1
          Rain,Snow,Fog                                         1
          Moderate Rain,Fog                                     1
          Freezing Rain,Ice Pellets,Fog                         1
          Drizzle,Ice Pellets,Fog                               1
          Thunderstorms,Rain,Fog                                1
          Rain,Ice Pellets                                      1
          Rain,Snow Grains                                      1
          Thunderstorms,Heavy Rain Showers                      1
          Freezing Rain,Snow Grains                             1
          Name: Weather, dtype: int64

In [19]:  df.info()
```

```
<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          8784 non-null   object
dtypes: float64(4), int64(2), object(2)
memory usage: 549.1+ KB
```

In [20]: `df["Wind Speed_km/h"].nunique()`

Out[20]: 34

In [21]: `df["Wind Speed_km/h"].unique()`

Out[21]:
```
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)
```

In [22]: `df.head(2)`

Out[22]:

	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

In [23]: `df.Weather.value_counts()`

```
Out[23]:   Mainly Clear                                    2106
           Mostly Cloudy                                   2069
           Cloudy                                          1728
           Clear                                           1326
           Snow                                             390
           Rain                                             306
           Rain Showers                                     188
           Fog                                              150
           Rain,Fog                                         116
           Drizzle,Fog                                       80
           Snow Showers                                      60
           Drizzle                                           41
           Snow,Fog                                          37
           Snow,Blowing Snow                                 19
           Rain,Snow                                         18
           Thunderstorms,Rain Showers                        16
           Haze                                              16
           Drizzle,Snow,Fog                                  15
           Freezing Rain                                     14
           Freezing Drizzle,Snow                             11
           Freezing Drizzle                                   7
           Snow,Ice Pellets                                   6
           Freezing Drizzle,Fog                               6
           Snow,Haze                                          5
           Freezing Fog                                       4
           Snow Showers,Fog                                   4
           Moderate Snow                                      4
           Rain,Snow,Ice Pellets                              4
           Freezing Rain,Fog                                  4
           Freezing Drizzle,Haze                              3
           Rain,Haze                                          3
           Thunderstorms,Rain                                 3
           Thunderstorms,Rain Showers,Fog                     3
           Freezing Rain,Haze                                 2
           Drizzle,Snow                                       2
           Rain Showers,Snow Showers                          2
           Thunderstorms                                      2
           Moderate Snow,Blowing Snow                         2
           Rain Showers,Fog                                   1
           Thunderstorms,Moderate Rain Showers,Fog            1
           Snow Pellets                                       1
           Rain,Snow,Fog                                      1
           Moderate Rain,Fog                                  1
           Freezing Rain,Ice Pellets,Fog                      1
           Drizzle,Ice Pellets,Fog                            1
           Thunderstorms,Rain,Fog                             1
           Rain,Ice Pellets                                   1
           Rain,Snow Grains                                   1
           Thunderstorms,Heavy Rain Showers                   1
           Freezing Rain,Snow Grains                          1
           Name: Weather, dtype: int64
```

In [24]:
```python
df[df.Weather=="Clear"]
```

	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

In [25]:
```python
df.groupby("Weather").get_group("Clear").head(10)
```

	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
241	1/11/2012 1:00	-10.7	-17.8	56	17	25.0	101.49	Clear
242	1/11/2012 2:00	-12.0	-18.9	56	19	25.0	101.57	Clear
243	1/11/2012 3:00	-12.7	-19.4	57	19	25.0	101.64	Clear
244	1/11/2012 4:00	-13.4	-20.1	57	17	25.0	101.66	Clear
344	1/15/2012 8:00	-23.3	-28.5	62	7	24.1	102.45	Clear

```python
df[df["Wind Speed_km/h"]==4].head(10)
```

Out[26]:

	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
147	1/7/2012 3:00	-7.8	-10.8	79	4	19.3	100.12	Cloudy
149	1/7/2012 5:00	-6.9	-9.7	80	4	19.3	100.12	Cloudy
164	1/7/2012 20:00	-1.8	-3.7	87	4	9.7	100.48	Snow
166	1/7/2012 22:00	-1.5	-3.0	89	4	4.0	100.54	Fog
194	1/9/2012 2:00	-9.0	-14.1	66	4	25.0	102.15	Mostly Cloudy

In [29]:
```python
df[(df["Wind Speed_km/h"]==4)& (df["Weather"]=="Snow")]
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
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
164	1/7/2012 20:00	-1.8	-3.7	87	4	9.7	100.48	Snow
789	2/2/2012 21:00	-7.2	-11.0	74	4	25.0	102.26	Snow
1087	2/15/2012 7:00	-0.3	-1.6	91	4	6.4	101.51	Snow
1088	2/15/2012 8:00	0.1	-1.5	89	4	4.8	101.57	Snow
1089	2/15/2012 9:00	0.6	-1.2	88	4	3.2	101.61	Snow
1389	2/27/2012 21:00	-7.3	-9.7	83	4	25.0	101.29	Snow
1391	2/27/2012 23:00	-6.8	-9.1	84	4	6.4	101.38	Snow
8207	12/7/2012 23:00	1.5	-2.2	76	4	19.3	101.22	Snow
8208	12/8/2012 0:00	1.8	-1.4	79	4	25.0	101.20	Snow
8507	12/20/2012 11:00	-1.1	-4.0	81	4	24.1	101.92	Snow
8769	12/31/2012 9:00	-8.1	-9.6	89	4	2.4	101.09	Snow
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

In [30]:
```python
df[(df["Visibility_km"]>24) |(df["Weather"]=="Snow")]
```

Out[30]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
20	1/1/2012 20:00	3.2	1.3	87	19	25.0	99.50	Cloudy
21	1/1/2012 21:00	4.0	1.7	85	20	25.0	99.39	Cloudy
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
...	...	...	...	...	...	...	...	...
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

7573 rows × 8 columns

In [ ]: