

```
In [1]: # First we have to import the libraries which we are in need of
import pandas as pd

In [5]: # Here df_Data is the name of our data
df_Data = pd.read_csv("D:\Data.csv.zip")

In [ ]:

In [9]: # Here it will show all the data present
df_Data

Out[9]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
0  01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1  01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2  01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3  01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4  01-01-2012 04:00    -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 [10]: #Lets try to Analyze the Dataframes by using exploration commands
df_Data.head(11)

Out[10]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
0  01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1  01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2  01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3  01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4  01-01-2012 04:00    -1.5        -3.3      88         7         4.8    101.23      Fog
5  01-01-2012 05:00    -1.4        -3.3      87         9         6.4    101.27      Fog
6  01-01-2012 06:00    -1.5        -3.1      89         7         6.4    101.29      Fog
7  01-01-2012 07:00    -1.4        -3.6      85         7         8.0    101.26      Fog
8  01-01-2012 08:00    -1.4        -3.6      85         9         8.0    101.23      Fog
9  01-01-2012 09:00    -1.3        -3.1      88        15         4.0    101.20      Fog
10 01-01-2012 10:00    -1.0        -2.3      91         9         1.2    101.15      Fog

In [12]: # 2. .shape---- it shows the total number of rows and columns in the data set
df_Data.shape # here 8784 are number of rows and 8 are the no of columns

Out[12]:
(8784, 8)

In [13]: # 3. .index---- it provides the index of the data frame
df_Data.index

Out[13]:
RangeIndex(start=0, stop=8784, step=1)

In [15]: # 4. .columns --- shows the all columns name present in the data frame
df_Data.columns

Out[15]:
Index(['Date/Time', 'Temp_C', 'Dew Point Temp_C', 'Rel Hum_%', 'Wind Speed_kmh', 'Visibility_km', 'Press_kPa', 'Weather'],
      dtype='object')

In [18]: # 5. .dtypes ---- shows the datatype of each columns
df_Data.dtypes

Out[18]:
Date/Time      object
Temp_C          float64
Dew Point Temp_C  float64
Rel Hum_%        int64
Wind Speed_kmh   int64
Visibility_km     float64
Press_kPa        float64
weather          object
dtype: object

In [21]: # 6. .unique() ---it shows all the unique values present in the coloum, it is applied on the single column only, not on whole data frams column
# Syntax - df_nameofthefile['nameofthecolumn'].unique
df_Data['Weather'].unique()

Out[21]:
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 [22]: df_Data['Weather'].unique

Out[22]:
<bound method Series.unique of 0              Fog
1              Fog
2  Freezing Drizzle,Fog
3  Freezing Drizzle,Fog
4              Fog
...
8779          Snow
8780          Snow
8781          Snow
8782          Snow
8783          Snow
Name: Weather, Length: 8784, dtype: object>

In [25]: # .nunique() ---it shows total number unique values present in each coloum, it is applied on the single column only, aslo can be apply on the whole data frams column
df_Data.nunique()

Out[25]:
Date/Time      8784
Temp_C          533
Dew Point Temp_C  489
Rel Hum_%        63
Wind Speed_kmh   34
Visibility_km     24
Press_kPa        53
Weather         50
dtype: int64

In [27]: # here we use nunique command on Temp column
df_Data['Temp_C'].nunique()

Out[27]:
533

In [28]: # 7. .count --- tells about the non null values in each column , can be applied on each or whole data set columns
df_Data.count

Out[28]:
<bound method DataFrame.count of
0 01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1 01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2 01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3 01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4 01-01-2012 04:00    -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 x 8 columns>
Date/Time      8784
Temp_C          533
Dew Point Temp_C  489
Rel Hum_%        63
Wind Speed_kmh   34
Visibility_km     24
Press_kPa        53
Weather         50
dtype: int64

In [31]: # To find null values in Data frame
pd.isnull(df_Data.sum)

Out[31]:
False

In [36]: # .value_counts() --- it shows all the unique value with their count --- can be applied to each or whole data set coloms, means how much time that unique value is present in
df_Data['Weather'].value_counts() where mainly clear is present 2160 time in weather column,rain is presnt 306 time in weather column. Means the count of each is given

Out[36]:
Mainly Clear      2160
Mostly Cloudy     2069
Cloudy            1728
Clear             1326
Snow              390
Rain              386
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 [37]: df_Data.value_counts()

Out[37]:
Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog      1
3/29/2012 2:00     -2.3        -12.6    54        15        24.1    101.94      Clear    1
3/20/2012 7:00     5.5        -0.1      67         4         24.1    102.00      Clear    1
3/20/2012 8:00     8.0         0.6      60         6         24.1    102.08      Clear    1
3/20/2012 9:00    10.6         1.3      53         0         19.3    102.09      Mainly Clear 1
1/26/2012 3:00    -0.1        -12.5      71         0         25.0    101.97      Clear    1
1/26/2012 2:00    -0.2        -12.6      71         6         25.0    101.99      Clear    1
1/26/2012 23:00   -0.8        -9.1      77         2         6.4     100.81      Snow     1
1/26/2012 22:00   -5.3        -10.2     68        20        16.1    100.92      Snow     1
9/30/2012 9:00    11.7         9.9      89        22        19.3    100.84      Cloudy   1
Length: 8784, dtype: int64

In [40]: # .info() --tells us about the number of rows and columns present , how much memory is used , datatypes present in the data frames
# means it provides basic information about our data
df_Data.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_kmh/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 [41]: # Find all the unique "Wind speed" values in the data set

In [42]: df_Data

Out[42]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
0  01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1  01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2  01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3  01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4  01-01-2012 04:00    -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 [48]: df_Data.head(10)

Out[48]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
0  01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1  01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2  01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3  01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4  01-01-2012 04:00    -1.5        -3.3      88         7         4.8    101.23      Fog
5  01-01-2012 05:00    -1.4        -3.3      87         9         6.4    101.27      Fog
6  01-01-2012 06:00    -1.5        -3.1      89         7         6.4    101.29      Fog
7  01-01-2012 07:00    -1.4        -3.6      85         7         8.0    101.26      Fog
8  01-01-2012 08:00    -1.4        -3.6      85         9         8.0    101.23      Fog
9  01-01-2012 09:00    -1.3        -3.1      88        15         4.0    101.20      Fog

In [54]: df_Data['Press_kPa'].nunique() #Example

Out[54]:
518

In [55]: df_Data['Wind Speed_kmh/h'].nunique() #Example

Out[55]:
34

In [56]: df_Data['Wind Speed_kmh/h'].unique()

Out[56]:
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, 21],
      dtype=int64)

In [57]: # Find the number of time when the weather is exactly clear

In [61]: df_Data['Weather'].value_counts()

Out[61]:
Mainly Clear      2160
Mostly Cloudy     2069
Cloudy            1728
Clear             1326
Snow              390
Rain              386
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 [64]: # By using filtering command we will solve the above question
# Syntax is nameOfTheData[nameOfTheData.nameOfTheColumn == 'nameOfThekeyword']

In [68]: df_Data[df_Data.Weather == 'Clear']

Out[68]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
67  01-03-2012 19:00    -16.9        -24.8      50         24         25.0    101.74      Clear
114 01-05-2012 18:00    -7.1        -14.4     56         11         25.0    100.71      Clear
115 01-05-2012 19:00    -9.2        -15.4     61         7         25.0    100.80      Clear
116 01-05-2012 20:00    -9.8        -15.7     62         9         25.0    100.83      Clear
117 01-05-2012 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 [69]: df_Data[df_Data.Weather == 'Snow']

Out[69]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
55  01-03-2012 07:00    -14.0        -19.5     51         19         25.0    100.95      Snow
84  01-04-2012 12:00    -13.7        -19.7     63         11         24.1    101.25      Snow
86  01-04-2012 14:00    -11.3        -19.0     53         7         19.3    100.97      Snow
87  01-04-2012 15:00    -10.2        -16.3     61         11         9.7    100.89      Snow
88  01-04-2012 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

In [70]: # Now we will solve the question with the help of the group by command

In [72]: df_Data.groupby('Weather').get_group('Clear')

Out[72]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
67  01-03-2012 19:00    -16.9        -24.8      50         24         25.0    101.74      Clear
114 01-05-2012 18:00    -7.1        -14.4     56         11         25.0    100.71      Clear
115 01-05-2012 19:00    -9.2        -15.4     61         7         25.0    100.80      Clear
116 01-05-2012 20:00    -9.8        -15.7     62         9         25.0    100.83      Clear
117 01-05-2012 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 [73]: #Find the number of time when the wind speed was exactly 4km/h

In [92]: df_Data.head(10)

Out[92]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
0  01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1  01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2  01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3  01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4  01-01-2012 04:00    -1.5        -3.3      88         7         4.8    101.23      Fog
5  01-01-2012 05:00    -1.4        -3.3      87         9         6.4    101.27      Fog
6  01-01-2012 06:00    -1.5        -3.1      89         7         6.4    101.29      Fog
7  01-01-2012 07:00    -1.4        -3.6      85         7         8.0    101.26      Fog
8  01-01-2012 08:00    -1.4        -3.6      85         9         8.0    101.23      Fog
9  01-01-2012 09:00    -1.3        -3.1      88        15         4.0    101.20      Fog

In [93]: df_Data[df_Data['Wind Speed_kmh/h'] == 4]

Out[93]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
0  01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1  01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2  01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3  01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4  01-01-2012 04:00    -1.5        -3.3      88         7         4.8    101.23      Fog
...
...
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
8772 12/31/2012 12:00    -5.8         -7.5     88         4         12.9    100.78      Snow
8773 12/31/2012 13:00    -4.8         -6.6     86         4         12.9    100.63      Snow

474 rows × 8 columns

In [94]: df_Data.head()

Out[94]:
      Date/Time  Temp_C  Dew Point Temp_C  Rel Hum_%  Wind Speed_kmh  Visibility_km  Press_kPa  Weather
0  01-01-2012 00:00    -1.8        -3.9      86         4         8.0    101.24      Fog
1  01-01-2012 01:00    -1.8        -3.7      87         4         8.0    101.24      Fog
2  01-01-2012 02:00    -1.8        -3.4      89         7         4.0    101.26  Freezing Drizzle,Fog
3  01-01-2012 03:00    -1.5        -3.2      88         6         4.0    101.27  Freezing Drizzle,Fog
4  01-01-2012 04:00    -1.5        -3.3      88         7         4.8    101.23      Fog

In [ ]:
```