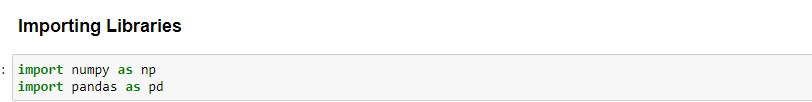
REPORT

****

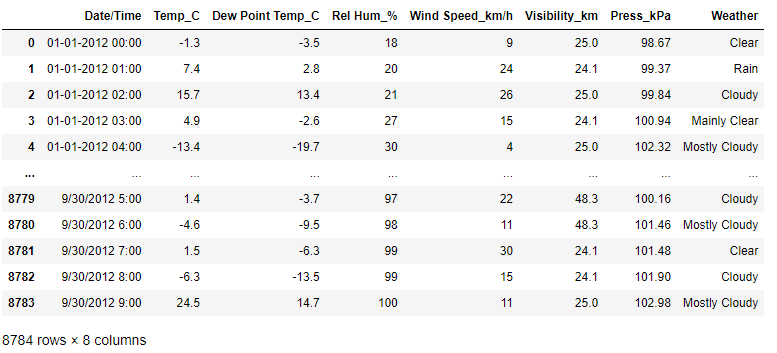
**NumPy (Numerical Python):-**

**NumPy is a fundamental library for numerical computations in Python. It provides support for larger,multi-dimensional arrays and matrices, as well as a wide range of mathematical functions to operate on these arrays efficiently.**

**Pandas:-**

**Pandas is a data manipulation and analysis library. It provides data structures and functions for efficiently working with structured data, such as tabular data in the form of Data Frames.**

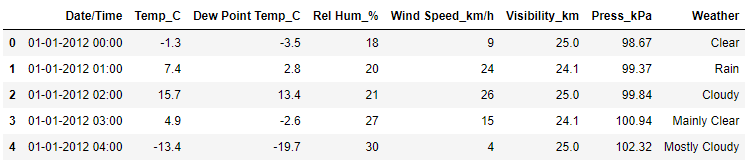
**Importing Dataset:-**

****

Importing data from Comma-Separated Values (CSV) files into a Pandas DataFrame. CSV files are commonly used to store structured data in a plain text format, where values are separated by commas

**Analyzing the Data:-**

****

****

It shows First 5 rows in the dataset (By default number of rows =5)

**data.shape**

(8784, 8)

It shows total number of rows and columns of the data Frame

**data.index**

RangeIndex(start=0, stop=8784, step=1)

This attribute Provide the index of the Data Frame

**data.dtypes**

Date/Time object

Temp\_C float64

Dew Point Temp\_C float64

Rel Hum\_% int64

Wind Speed\_km/h int64

Visibility\_km float64

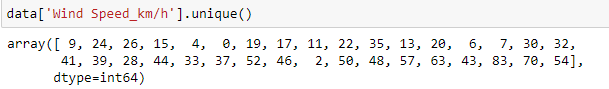
Press\_kPa float64

Weather object

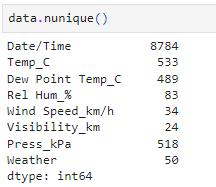
dtype: object

IT shows datatype of each column

**1.Find all the unique 'Wind Speed' values in the data.**



In a column, it shows all the unique values. It can be appliednon a single column only. Not on the whole data frame



Count the number of distinct values within a Pandas Series or DataFrame column. It helps in understanding the diversity of data and is commonly used in data exploration and data quality assessment tasks.

### 2. Find the number of times when the 'Weather is exactly Clear'.

### 

### value\_counts function in Pandas is a powerful tool for understanding the frequency distribution of categorical data within a Pandas Series or DataFrame column. It is commonly used in data exploration, data quality assessment, and data visualization tasks.

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

### 

### 

For finding the number of times when the ‘Wind Speed was exactly 4 km/h**'** I use ‘**value\_counts()**’ function on ‘Wind Speed’ column and value is ‘4’.

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

### 

For finding the all the null values in the data I use ‘**isnull()**’ function on df data and ‘**.sum()**’ function to get number of null values in each column.

### 5. Rename the column name 'Weather' of the data frame to 'Weather Condition'

### 

### This code tells renaming operation does not modify the original data frame but creates a new data frame with the modified column name. Here weather column is renamed as weather condition.

### 6. What is the mean 'Visibility' ?

### 

For calculating the mean ‘visibility’ I use ‘**mean()**’ function on ‘Visibility’ column.

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

### 

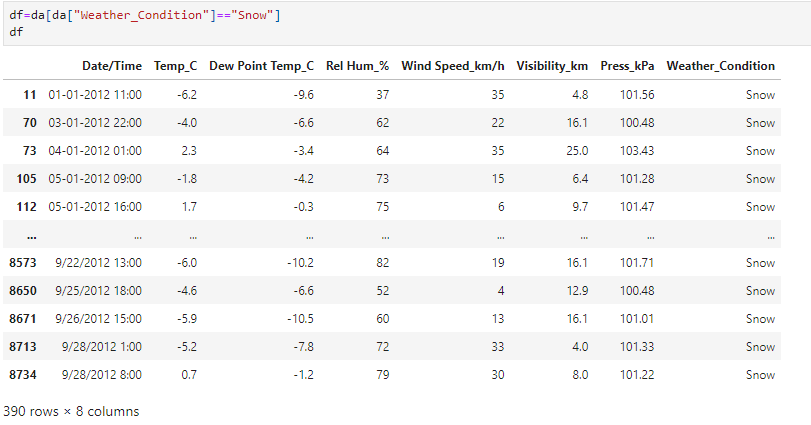
For calculating the Standard Deviation of ‘Pressure’ I use ‘**std()**’ function on ‘Pressure’ column.

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

### 

For calculating the Variance of ‘Relative Humidity’ I use ‘**var()**’ function on ‘Relative Humidity’ column.

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

****

For finding the all instances when ‘Snow’ was recorded I use ‘**==**’ comparison operator on ‘Weather Condition’ column and the value is Snow.

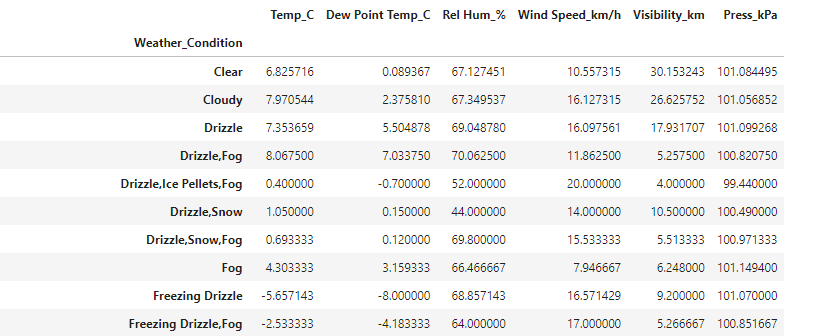
### 10. Find all instances when 'Wind Speed is above 24' and 'Visibility is 25'

### 

For finding the all instances when ‘Wind Speed is above 24 and visibility is 25’ I use ‘**>**’ comparison operator on ‘Wind speed’ column with the value is 24 and **‘==**’ comparison operator on ‘Visibility’ column with the value is 25 and also ‘**&**’ logical operator to satisfy the both conditions.

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

****

****

For finding the mean value of each column against each ‘Weather Condition’ I use ‘**groupby()**’ function on ‘Weather condition’ column and for mean I take aggregate function ‘**mean()**’.

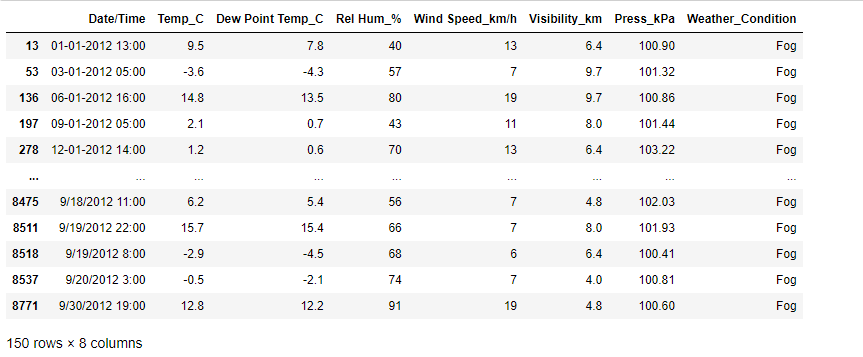
### 12. What is the Minimum & Maximum value of each column against each 'Weather Condition ?

### 

For finding the minimum & maximum value of each column against each ‘Weather Condition’ I use ‘**groupby()**’ function on ‘Weather condition’ column and for minimum and maximum I take aggregate function ‘**agg([’min’,’max’])’**.

### 13. Show all the Records where Weather Condition is Fog.





For showing the all records where ‘Weather Condition’ is Fog I use ‘**==**’ comparison operator on ‘Weather Condition’ column and the value is Fog.

### 14. Find all instances when 'Weather is Clear' or 'Visibility is above 40'.[¶](http://localhost:8888/notebooks/Desktop/innomatics/Data_Analysis%20In%20Weather%20Dataset.ipynb#14.Find-all-instances-when-'Weather-is-Clear'-or-'Visibility-is-above-40'.)



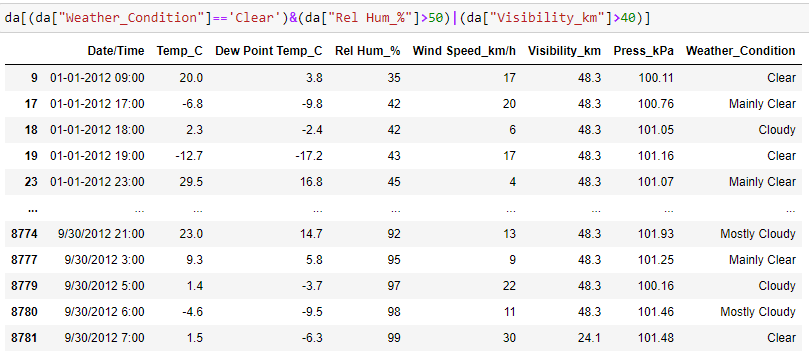
For finding the all instances when ‘Weather is Clear or visibility is above 40’ I use ‘**==**’ comparison operator on ‘Weather Condition’ column with the value is Clear and **‘>**’ comparison operator on ‘Visibility’ column with the value is 40 and also **‘|**’ logical operator to satisfy the either one condition.

### 15. Find all instances when:

A. 'Weather is clear' and 'Relative Humidity is greater than 50'

or

B. 'Visibility is above 40'



It computes data frame containing only the rows from the "Weather", " Rel Hum\_%", and "Visibility\_ km" columns where either the weather is 'Clear' and relative humidity is greater than 50% or visibility is greater than 40 km.

In [91]: