

```
import pandas as pd
```

```
data = pd.read_csv('/content/sample_data/1. Weather Data.csv')
```

data



	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.
2	1/1/2012 2:00	-1.8	-3.4	89	7	4.0	101.
3	1/1/2012 3:00	-1.5	-3.2	88	6	4.0	101.
4	1/1/2012 4:00	-1.5	-3.3	88	7	4.8	101.
...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.

How to Analyze DataFrame?

▼ .head()

it shows the first n rows in the data.(by default n=5)

```
data.head()
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24
1	1/1/2012 0:00	-1.8	-3.7	87	4	8.0	101.24

▼ .shape

it shows the total of rows and no of columns of the dataframe

```
data.shape
```

```
(8784, 8)
```

▼ .index

this attribute provides the index of the dataframe

```
data.index
```

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

▼ .columns

it shows the name of each columns

```
data.columns
```

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

▼ .dtype

it shows the data type of each columns

```
data.dtypes
```

```
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
```

▼ .unique()

in a column,it shows at the unique values .it can be applied on a single column only.not on the whole dataframe

```
data['Weather'].unique()
```

```
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)
```

▼ .nunique()

it shows the total no of unique values in each columns.it can be applied on a single columns as well as on whole dataframe.

```
data.nunique()
```

```

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

```

▼ .count

it shows the total no of non-null in each column.it can be applied on a single column as well as whole dataframe.

```
data.count()
```

```

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

```

▼ .value_counts

in a columns ,it shows all the unique values with their count .it can be applied on sigle column only.

```
data['Weather'].value_counts()
```

```

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	

▼ .info()

provides basic information about the dataframe.

```
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_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
```

Q).1 Find all the unique "wind speed " values in the data

```
data.head(2)
```

Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
1/1/2012						

```
data.nunique()
```

```
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
```

```
data['Wind Speed_km/h'].nunique()
```

```
34
```

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

```
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]`

Q) 2.find the number of times when the "Weather is exactly clear"

```
data.Weather.value_counts()
```

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

```

```
data[data.Weather=='Clear']
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.
114	1/5/2012 18:00	-7.1	-14.4	56	11	25.0	100.
115	1/5/2012 19:00	-9.2	-15.4	61	7	25.0	100.
116	1/5/2012 20:00	-9.8	-15.7	62	9	25.0	100.
117	1/5/2012 21:00	-9.0	-14.8	63	13	25.0	100.
...
8646	12/26/2012 6:00	-13.4	-14.8	89	4	25.0	102.



```
data.groupby('Weather').get_group('Clear')
```


	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.
114	1/5/2012 18:00	-7.1	-14.4	56	11	25.0	100.
115	1/5/2012 19:00	-9.2	-15.4	61	7	25.0	100.
116	1/5/2012	0.0	15.7	60	0	25.0	100.

Q) 3. find the numbers of times when the 'wind speed was exactly 4 km/h'

```
data[data['Wind Speed_km/h']==4]
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.
96	1/5/2012 0:00	-8.8	-11.7	79	4	9.7	100.
101	1/5/2012 5:00	-7.0	-9.5	82	4	4.0	100.
146	1/7/2012 2:00	-8.1	-11.1	79	4	19.3	100.
...
8768	12/31/2012 8:00	-8.6	-10.3	87	4	3.2	101.

Q) 4.find out the null values in the data

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

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

```
data.notnull().sum()
```

```
Date/Time      8784
Temp_C         8784
Dew Point Temp_C 8784
Rel Hum_%      8784
Wind Speed_kmh 8784
Visibility_km   8784
Press_kPa      8784
Weather        8784
dtype: int64
```

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

```
data.rename(columns={'Weather': 'Weather Condition'})
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.

Q) 6.What is the mean 'visibility' ?

2	2:00	-1.8	-3.4	89	7	4.0	101.
---	------	------	------	----	---	-----	------

```
data.Visibility_km.mean()
```

```
27.66444672131151
```

4	1/1/2012	-1.5	-3.3	88	7	4.8	101
---	----------	------	------	----	---	-----	-----

Q)7.what is the standard Deviation of 'pressure' in this data?

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

```
0.8440047459486474
```

Q)8.what is the variance of 'Relative Humidity' in this data?

```
data.head(2)
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
	1/1/2012						

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

```
286.2485501984998
```

Q)9.Find all instances when 'snow' was recorded

```
data.head(2)
data['Weather'].value_counts()
```

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

```

```
data[data['Weather']=='Snow']
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
55	1/3/2012 7:00	-14.0	-19.5	63	19	25.0	100.
84	1/4/2012 12:00	-13.7	-21.7	51	11	24.1	101.
86	1/4/2012 14:00	-11.3	-19.0	53	7	19.3	100.
87	1/4/2012 15:00	-10.2	-16.3	61	11	9.7	100.
88	1/4/2012 16:00	-9.4	-15.5	61	13	19.3	100.
...
8779	12/31/2012 19:00	0.1	-2.7	81	30	9.7	100.

Q)10.Find all Instances when 'Wind speed is above 24 ' and 'visibility is 25'

```
data[(data['Wind Speed_km/h']>24) & (data['Visibility_km']==25)]
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
23	1/1/2012 23:00	5.3	2.0	79	30	25.0	99.
24	1/2/2012 0:00	5.2	1.5	77	35	25.0	99.
25	1/2/2012 1:00	4.6	0.0	72	39	25.0	99.
26	1/2/2012 2:00	3.9	-0.9	71	32	25.0	99.

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

17.00

```
data.groupby('Weather').mean()
```

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visib:
Weather					
Clear	6.825716	0.089367	64.497738	10.557315	3
Cloudy	7.970544	2.375810	69.592593	16.127315	2
Drizzle	7.353659	5.504878	88.243902	16.097561	1
Drizzle,Fog	8.067500	7.033750	93.275000	11.862500	
Drizzle,Ice Pellets,Fog	0.400000	-0.700000	92.000000	20.000000	
Drizzle,Snow	1.050000	0.150000	93.500000	14.000000	1
Drizzle,Snow,Fog	0.693333	0.120000	95.866667	15.533333	
Fog	4.303333	3.159333	92.286667	7.946667	
Freezing Drizzle	-5.657143	-8.000000	83.571429	16.571429	
Freezing Drizzle,Fog	-2.533333	-4.183333	88.500000	17.000000	
Freezing Drizzle,Haze	-5.433333	-8.000000	82.000000	10.333333	
Freezing Drizzle,Snow	-5.109091	-7.072727	86.090909	16.272727	
Freezing Fog	-7.575000	-9.250000	87.750000	4.750000	
Freezing Rain	-3.885714	-6.078571	84.642857	19.214286	
Freezing Rain,Fog	-2.225000	-3.750000	89.500000	15.500000	
Freezing Rain,Haze	-4.900000	-7.450000	82.500000	7.500000	
Freezing Rain,Ice Pellets,Fog	-2.600000	-3.700000	92.000000	28.000000	
Freezing Rain,Snow Grains	-5.000000	-7.300000	84.000000	32.000000	
Haze	-0.200000	-2.975000	81.625000	10.437500	
Mainly Clear	12.558927	4.581671	60.667142	14.144824	3
Moderate Rain,Fog	1.700000	0.800000	94.000000	17.000000	
Moderate Snow	-5.525000	-7.250000	87.750000	33.750000	
Moderate Snow,Blowing Snow	-5.450000	-6.500000	92.500000	40.000000	
Mostly Cloudy	10.574287	3.131174	62.102465	15.813920	3
Rain	9.786275	7.042810	83.624183	19.254902	1
Rain Showers	13.722340	9.187766	75.150574	17.132070	2

Rain,Showers	10.722070	9.107700	70.100074	17.102070	2
Rain,Showers,Fog	12.800000	12.100000	96.000000	13.000000	
Rain,Showers,Snow Showers	2.150000	-1.500000	76.500000	22.500000	2
Rain,Fog	8.273276	7.219828	93.189655	14.793103	
Rain,Haze	4.633333	2.066667	83.333333	11.666667	
Rain,Ice Pellets	0.600000	-0.600000	92.000000	24.000000	
Rain,Snow	1.055556	-0.566667	89.000000	28.388889	1
Rain,Snow,Showers	1.000000	0.000000	75.000000	20.000000	0

Q)12.What is the minimum & Maximum value of each column against each 'Weather condition'?

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


	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visi
Weather						
Clear	1/11/2012 1:00	-23.3	-28.5	20	0	
Cloudy	1/1/2012 17:00	-21.4	-26.8	18	0	
Drizzle	1/23/2012 21:00	1.1	-0.2	74	0	
Drizzle,Fog	1/23/2012 20:00	0.0	-1.6	85	0	
Drizzle,Ice Pellets,Fog	12/17/2012 9:00	0.4	-0.7	92	20	
Drizzle,Snow	12/17/2012 15:00	0.9	0.1	92	9	
Drizzle,Snow,Fog	12/18/2012 21:00	0.3	-0.1	92	7	
Fog	1/1/2012 0:00	-16.0	-17.2	80	0	
Freezing Drizzle	1/13/2012 10:00	-9.0	-12.2	78	6	
Freezing Drizzle,Fog	1/1/2012 2:00	-6.4	-9.0	82	6	
Freezing Drizzle,Haze	2/1/2012 11:00	-5.8	-8.3	81	9	
Freezing Drizzle,Snow	1/13/2012 3:00	-8.3	-10.4	79	6	
Freezing Fog	1/22/2012 6:00	-19.0	-22.9	71	0	
Freezing Rain	1/13/2012 11:00	-6.5	-9.0	81	7	
Freezing Rain,Fog	1/17/2012 23:00	-6.1	-8.7	82	7	
Freezing Rain,Haze	2/1/2012 14:00	-4.9	-7.5	82	6	
Freezing Rain,Ice Pellets,Fog	12/17/2012 3:00	-2.6	-3.7	92	28	

Freezing Rain,Snow Grains	1/13/2012 9:00	-5.0	-7.3	84	32
Haze	1/22/2012 12:00	-11.5	-16.0	68	0
Mainly Clear	1/10/2012 11:00	-22.8	-28.0	20	0
Moderate Rain,Fog	12/10/2012 8:00	1.7	0.8	94	17

```
data.groupby('Weather').max()
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visi
Weather						
Clear	9/9/2012 5:00	32.8	20.4	99	33	
Cloudy	9/9/2012 23:00	30.5	22.6	99	54	
Drizzle	9/30/2012 3:00	18.8	17.7	96	30	
Drizzle,Fog	9/30/2012 2:00	19.9	19.1	100	28	
Drizzle,Ice Pellets,Fog	12/17/2012 9:00	0.4	-0.7	92	20	
Drizzle,Snow	12/19/2012 18:00	1.2	0.2	95	19	
Drizzle,Snow,Fog	12/22/2012 3:00	1.1	0.6	98	32	
Fog	9/22/2012 0:00	20.8	19.6	100	22	
Freezing Drizzle	2/1/2012 5:00	-2.3	-3.3	93	26	
Freezing Drizzle,Fog	12/10/2012 5:00	-0.3	-2.3	94	33	
Freezing Drizzle,Haze	2/1/2012 13:00	-5.0	-7.7	83	11	
Freezing Drizzle,Snow	3/2/2012 12:00	-3.3	-4.6	94	24	
Freezing Fog	3/17/2012 6:00	-0.1	-0.3	99	9	
Freezing Rain	2/1/2012 7:00	0.3	-1.7	92	28	
Freezing Rain,Fog	12/17/2012 1:00	0.1	-0.9	93	26	
Freezing Rain,Haze	2/1/2012 15:00	-4.9	-7.4	83	9	
Freezing Rain,Ice Pellets,Fog	12/17/2012 3:00	-2.6	-3.7	92	28	

☺ ☺

UU

```
data[data['Weather']=='Fog']
```

◀
3/13/2012
55
20
86
17
▶


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

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.
106	1/5/2012 10:00	-6.0	-10.0	73	17	48.3	100.
107	1/5/2012 11:00	-5.6	-10.2	70	22	48.3	100.
108	1/5/2012 12:00	-4.7	-9.6	69	20	48.3	100.
	1/5/2012						

Q)15.find 'Weather is clear' and 'Relative Humidity is greater than 50'or 'visibility is above 40'

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

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_k
106	1/5/2012 10:00	-6.0	-10.0	73	17	48.3	100.
107	1/5/2012 11:00	-5.6	-10.2	70	22	48.3	100.
108	1/5/2012 12:00	-4.7	-9.6	69	20	48.3	100.
109	1/5/2012 13:00	-4.4	-9.7	66	26	48.3	100.
110	1/5/2012 14:00	-5.1	-10.7	65	22	48.3	100.
...
8749	12/30/2012 13:00	-12.4	-16.2	73	37	48.3	100.

✓

0s

completed at 12:10 AM

●

✕