DATA ANALYTICS PROJECT

The Weather Dataset

Here, The weather dataset is a time-series data set with per-hour information about the weather condiitons at a particular location. It records temperature, Dew point temperature, relative humidity, wind speed, visibility, pressure and conditions.

```
import pandas as pd

data = pd.read_csv("/file.csv")

data
```

	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

How to analyze Data Frame?

.head()

```
data.head()
```

	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

.shape()

```
data.shape
```

```
(8784, 8)
```

.index

```
data.index
```

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

.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')
```

.dtypes

```
data.dtypes
```

	0
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

.Uniques

```
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

```
data.nunique()
```

	0
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

```
data.count()
```

	0
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

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

Weather	count
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

dtype: int64

.info

```
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
```

Q1. Find all the unique 'Wind speed' values in the data?

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

```
34
```

Q2. Find the number of times when the "Weather is exactly Clear"?

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

Q3. Find the number 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_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

Q4. Find out all the Null values in the data?

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

	0
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

Q5. Rename the column name 'Weather' of the dataframe to 'Weather COndition'.

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

Q6. What is the mean 'Visibility'?

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

```
27.664446721311478
```

Q7. What is the standard deviation of 'Pressure' in this data?

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

```
0.8440047459486474
```

Q8. What is the Variance of Relative Humidity in this data?

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

```
286.2485501984998
```

Q9. Find all instances when snow was recorded?

```
data.groupby('Weather Condition').get_group('Snow')
```

	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

```
data[data['Weather Condition'].str.contains('Snow')]
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
41	1/2/2012 17:00	-2.1	-9.5	57	22	25.0	99.66	Snow Showers
44	1/2/2012 20:00	-5.6	-13.4	54	24	25.0	100.07	Snow Showers
45	1/2/2012 21:00	-5.8	-12.8	58	26	25.0	100.15	Snow Showers
47	1/2/2012 23:00	-7.4	-14.1	59	17	19.3	100.27	Snow Showers
48	1/3/2012 0:00	-9.0	-16.0	57	28	25.0	100.35	Snow Showers
...	...	...	...	...	...	...	...	...
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

583 rows × 8 columns

Q10. Find all instances when 'Wind Speed is baove 24' and 'Visibilty 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_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

Q11. What is the Mean Value of each column aganist eash 'Weather Condition'?

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

	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

```
new_var = data.groupby('Weather Condition').mean(numeric_only=True)
new_var
```

Weather Condition	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Pres
Clear	6.825716	0.089367	64.497738	10.557315	30.153243	101.5
Cloudy	7.970544	2.375810	69.592593	16.127315	26.625752	100.9
Drizzle	7.353659	5.504878	88.243902	16.097561	17.931707	100.4
Drizzle,Fog	8.067500	7.033750	93.275000	11.862500	5.257500	100.7
Drizzle,Ice Pellets,Fog	0.400000	-0.700000	92.000000	20.000000	4.000000	100.7
Drizzle,Snow	1.050000	0.150000	93.500000	14.000000	10.500000	100.8
Drizzle,Snow,Fog	0.693333	0.120000	95.866667	15.533333	5.513333	99.2
Fog	4.303333	3.159333	92.286667	7.946667	6.248000	101.1
Freezing Drizzle	-5.657143	-8.000000	83.571429	16.571429	9.200000	100.2
Freezing Drizzle,Fog	-2.533333	-4.183333	88.500000	17.000000	5.266667	100.4
Freezing Drizzle,Haze	-5.433333	-8.000000	82.000000	10.333333	2.666667	100.3
Freezing Drizzle,Snow	-5.109091	-7.072727	86.090909	16.272727	5.872727	100.5
Freezing Fog	-7.575000	-9.250000	87.750000	4.750000	0.650000	102.3
Freezing Rain	-3.885714	-6.078571	84.642857	19.214286	8.242857	99.6
Freezing Rain,Fog	-2.225000	-3.750000	89.500000	15.500000	7.550000	99.9
Freezing Rain,Haze	-4.900000	-7.450000	82.500000	7.500000	2.400000	100.3
Freezing Rain,Ice Pellets,Fog	-2.600000	-3.700000	92.000000	28.000000	8.000000	100.9
Freezing Rain,Snow Grains	-5.000000	-7.300000	84.000000	32.000000	4.800000	98.5
Haze	-0.200000	-2.975000	81.625000	10.437500	7.831250	101.4
Mainly Clear	12.558927	4.581671	60.667142	14.144824	34.264862	101.2
Moderate Rain,Fog	1.700000	0.800000	94.000000	17.000000	6.400000	99.9
Moderate Snow	-5.525000	-7.250000	87.750000	33.750000	0.750000	100.2
Moderate Snow,Blowing Snow	-5.450000	-6.500000	92.500000	40.000000	0.600000	100.5
Mostly Cloudy	10.574287	3.131174	62.102465	15.813920	31.253842	101.0
Rain	9.786275	7.042810	83.624183	19.254902	18.856536	100.2
Rain Showers	13.722340	9.187766	75.159574	17.132979	22.816489	100.4
Rain Showers,Fog	12.800000	12.100000	96.000000	13.000000	6.400000	99.8
Rain Showers,Snow	2.150000	1.500000	76.500000	22.500000	21.700000	101.1

Showers	2.150000	-1.500000	76.500000	22.500000	21.700000	101.1
Rain,Fog	8.273276	7.219828	93.189655	14.793103	6.873276	100.5
Rain,Haze	4.633333	2.066667	83.333333	11.666667	6.700000	100.5
Rain,Ice Pellets	0.600000	-0.600000	92.000000	24.000000	9.700000	100.1
Rain,Snow	1.055556	-0.566667	89.000000	28.388889	11.672222	99.9
Rain,Snow Grains	1.900000	-2.100000	75.000000	26.000000	25.000000	100.6
Rain,Snow,Fog	0.800000	0.300000	96.000000	9.000000	6.400000	100.7
Rain,Snow,Ice Pellets	1.100000	-0.175000	91.500000	23.250000	6.000000	100.1
Snow	-4.524103	-7.623333	79.307692	20.038462	11.171795	100.5
Snow Pellets	0.700000	-6.400000	59.000000	35.000000	2.400000	99.7
Snow Showers	-3.506667	-7.866667	72.350000	19.233333	20.158333	100.9
Snow Showers,Fog	-10.675000	-11.900000	90.750000	13.750000	7.025000	101.2
Snow,Blowing Snow	-5.410526	-7.621053	84.473684	34.842105	4.105263	99.7
Snow,Fog	-5.075676	-6.364865	90.675676	17.324324	4.537838	100.6
Snow,Haze	-4.020000	-6.860000	80.600000	5.000000	4.640000	100.7
Snow,Ice Pellets	-1.883333	-3.666667	87.666667	23.833333	7.416667	100.5
Thunderstorms	24.150000	19.750000	77.000000	7.500000	24.550000	100.2
Thunderstorms,Heavy Rain Showers	10.900000	9.000000	88.000000	9.000000	2.400000	100.2
Thunderstorms,Moderate Rain Showers,Fog	19.600000	18.500000	93.000000	15.000000	3.200000	100.0
Thunderstorms,Rain	20.433333	18.533333	89.000000	15.666667	19.833333	100.4
Thunderstorms,Rain Showers	20.037500	17.618750	86.375000	18.312500	15.893750	100.2
Thunderstorms,Rain Showers,Fog	21.600000	18.700000	84.000000	19.666667	9.700000	100.0
Thunderstorms,Rain,Fog	20.600000	18.600000	88.000000	19.000000	4.800000	100.0