PROJECT REPORT

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

CODE

df['Wind Speed_km/h'].unique()

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

```
array([ 9, 24, 26, 15,  4,  0, 19, 17, 11, 22, 35, 13, 20,  6,  7, 30, 32,
       41, 39, 28, 44, 33, 37, 52, 46,  2, 50, 48, 57, 63, 43, 83, 70, 54],
      dtype=int64)
```

Explanation

The result shows the unique values in the column Wind Speed_km/h

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

CODE

df['Weather'].value_counts().Clear

```
df['Weather'].value_counts().Clear
```

```
1326
```

Explanation

The result shows the count of values in the Weather column which is clear

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

CODE

df['Wind Speed_km/h'].value_counts()[4]

```
df['Wind Speed_km/h'].value_counts()[4]
```

```
474
```

Explanation

The result shows the count of values in the Wind Speed_km/h column which is “4Km/h”

Q. 4) Find out all the Null Values in the data.

CODE

df.info()

df.isnull().sum()

```
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
```

```
df.isnull().sum()
```

```
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 Condition   0
dtype: int64
```

Explanation

The result shows that there are no null values in the given dataset.

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

CODE

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

df

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

	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	-1.3	-3.5	18	9	25.0	98.67	Clear
1	1/1/2012 1:00	7.4	2.8	20	24	24.1	99.37	Rain
2	1/1/2012 2:00	15.7	13.4	21	26	25.0	99.84	Cloudy
3	1/1/2012 3:00	4.9	-2.6	27	15	24.1	100.94	Mainly Clear
4	1/1/2012 4:00	-13.4	-19.7	30	4	25.0	102.32	Mostly Cloudy
...	...	...	...	...	...	...	...	...
8779	9/30/2012 5:00	1.4	-3.7	97	22	48.3	100.16	Cloudy
8780	9/30/2012 6:00	-4.6	-9.5	98	11	48.3	101.46	Mostly Cloudy
8781	9/30/2012 7:00	1.5	-6.3	99	30	24.1	101.48	Clear
8782	9/30/2012 8:00	-6.3	-13.5	99	15	24.1	101.90	Cloudy
8783	9/30/2012 9:00	24.5	14.7	100	11	25.0	102.98	Mostly Cloudy

8784 rows × 8 columns

Explanation

The result shows that the name of “Weather” column changed to “Weather Condition”

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

CODE

df['Visibility_km'].mean()

```
df['Visibility_km'].mean()
```

```
27.664446721311478
```

Explanation

The result shows that the mean of column “Visibility_km”

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

CODE

df['Press_kPa'].std()

```
df['Press_kPa'].std()
```

```
0.8440047459486459
```

Explanation

The result shows that the standard deviation of the column “Press_kPa”

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

CODE

df['Rel Hum_%'].var()

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

```
286.2485501985015
```

Explanation

The result shows that the variance of the column “Rel Hum_%”

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

CODE

df[df['Weather']=='Snow']

```
df[df['Weather Condition']=='Snow']
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
11	1/1/2012 11:00	-6.2	-9.6	37	35	4.8	101.56	Snow
70	3/1/2012 22:00	-4.0	-6.6	62	22	16.1	100.48	Snow
73	4/1/2012 1:00	2.3	-3.4	64	35	25.0	103.43	Snow
105	5/1/2012 9:00	-1.8	-4.2	73	15	6.4	101.28	Snow
112	5/1/2012 16:00	1.7	-0.3	75	6	9.7	101.47	Snow
...	...	...	...	...	...	...	...	...
8573	9/22/2012 13:00	-6.0	-10.2	82	19	16.1	101.71	Snow
8650	9/25/2012 18:00	-4.6	-6.6	52	4	12.9	100.48	Snow
8671	9/26/2012 15:00	-5.9	-10.5	60	13	16.1	101.01	Snow
8713	9/28/2012 1:00	-5.2	-7.8	72	33	4.0	101.33	Snow
8734	9/28/2012 8:00	0.7	-1.2	79	30	8.0	101.22	Snow

390 rows × 8 columns

Explanation

The result shows that the rows which has the Weather Condition is “Snow”

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

CODE

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

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

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
2	1/1/2012 2:00	15.7	13.4	21	26	25.0	99.84	Cloudy
73	4/1/2012 1:00	2.3	-3.4	64	35	25.0	103.43	Snow
126	6/1/2012 6:00	10.0	5.4	77	39	25.0	101.30	Cloudy
158	7/1/2012 14:00	1.9	-2.1	87	26	25.0	100.87	Rain,Snow Grains
184	8/1/2012 16:00	14.2	9.2	35	44	25.0	99.49	Mostly Cloudy
...	...	...	...	...	...	...	...	...
8707	9/27/2012 5:00	-1.0	-6.0	70	33	25.0	98.56	Mostly Cloudy
8714	9/28/2012 10:00	2.6	0.3	72	26	25.0	101.60	Rain
8738	9/29/2012 10:00	22.8	12.3	80	28	25.0	101.60	Mostly Cloudy
8745	9/29/2012 17:00	-10.3	-12.9	82	28	25.0	102.16	Cloudy
8776	9/30/2012 23:00	19.2	13.2	93	43	25.0	101.60	Mainly Clear

308 rows × 8 columns

Explanation

The result shows that the rows which has ‘Wind Speed is above 24' and 'Visibility is 25'

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

CODE

df.groupby('Weather Condition').mean()

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

Weather Condition	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Clear	6.825716	0.089367	67.127451	10.557315	30.153243	101.084495
Cloudy	7.970544	2.375810	67.349537	16.127315	26.625752	101.056852
Drizzle	7.353659	5.504878	69.048780	16.097561	17.931707	101.099268
Drizzle,Fog	8.067500	7.033750	70.062500	11.862500	5.257500	100.820750
Drizzle,Ice Pellets,Fog	0.400000	-0.700000	52.000000	20.000000	4.000000	99.440000
Drizzle,Snow	1.050000	0.150000	44.000000	14.000000	10.500000	100.490000
Drizzle,Snow,Fog	0.693333	0.120000	69.800000	15.533333	5.513333	100.971333
Fog	4.303333	3.159333	66.466667	7.946667	6.248000	101.149400
Freezing Drizzle	-5.657143	-8.000000	68.857143	16.571429	9.200000	101.070000
Freezing Drizzle,Fog	-2.533333	-4.183333	64.000000	17.000000	5.266667	100.851667
Freezing Drizzle,Haze	-5.433333	-8.000000	63.333333	10.333333	2.666667	101.136667
Freezing Drizzle,Snow	-5.109091	-7.072727	62.454545	16.272727	5.872727	100.380909
Freezing Fog	-7.575000	-9.250000	68.000000	4.750000	0.650000	101.222500
Freezing Rain	-3.885714	-6.078571	60.785714	19.214286	8.242857	101.500714
Freezing Rain,Fog	-2.225000	-3.750000	52.750000	15.500000	7.550000	100.267500
Freezing Rain,Haze	-4.900000	-7.450000	63.000000	7.500000	2.400000	100.265000
Freezing Rain,Ice Pellets,Fog	-2.600000	-3.700000	65.000000	28.000000	8.000000	98.330000
Freezing Rain,Snow Grains	-5.000000	-7.300000	92.000000	32.000000	4.800000	102.520000
Haze	-0.200000	-2.975000	69.625000	10.437500	7.831250	100.805625
Mainly Clear	12.558927	4.581671	68.020893	14.144824	34.264862	101.040940

Explanation

The result shows the mean of each column grouped by the Weather condition

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

CODE

df.groupby('Weather Condition').min()

df.groupby('Weather Condition').max()

```
df.groupby('Weather Condition').min().head()
```

Weather Condition	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Clear	1/1/2012 0:00	-23.3	-28.5	18	0	11.3	97.75
Cloudy	1/1/2012 10:00	-21.4	-26.8	20	0	11.3	97.52
Drizzle	1/18/2012 16:00	1.1	-0.2	37	0	6.4	98.29
Drizzle,Fog	1/28/2012 12:00	0.0	-1.6	38	0	1.0	98.32
Drizzle,Ice Pellets,Fog	7/24/2012 5:00	0.4	-0.7	52	20	4.0	99.44

```
df.groupby('Weather Condition').max().head()
```

Weather Condition	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Clear	9/9/2012 4:00	32.8	20.4	100	33	48.3	103.63
Cloudy	9/9/2012 6:00	30.5	22.6	100	54	48.3	103.52
Drizzle	9/15/2012 22:00	18.8	17.7	97	30	25.0	103.58
Drizzle,Fog	9/6/2012 10:00	19.9	19.1	98	28	9.7	103.56
Drizzle,Ice Pellets,Fog	7/24/2012 5:00	0.4	-0.7	52	20	4.0	99.44

Explanation

The result shows the minimum and maximum of each column grouped by the Weather condition

Q. 13) Show all the Records where Weather Condition is Fog.

CODE

df[df['Weather Condition']=='Fog']

```
df[df['Weather Condition']=='Fog']
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
13	1/1/2012 13:00	9.5	7.8	40	13	6.4	100.90	Fog
53	3/1/2012 5:00	-3.6	-4.3	57	7	9.7	101.32	Fog
136	6/1/2012 16:00	14.8	13.5	80	19	9.7	100.86	Fog
197	9/1/2012 5:00	2.1	0.7	43	11	8.0	101.44	Fog
278	12/1/2012 14:00	1.2	0.6	70	13	6.4	103.22	Fog
...	...	...	...	...	...	...	...	...
8475	9/18/2012 11:00	6.2	5.4	56	7	4.8	102.03	Fog
8511	9/19/2012 22:00	15.7	15.4	66	7	8.0	101.93	Fog
8518	9/19/2012 8:00	-2.9	-4.5	68	6	6.4	100.41	Fog
8537	9/20/2012 3:00	-0.5	-2.1	74	7	4.0	100.81	Fog
8771	9/30/2012 19:00	12.8	12.2	91	19	4.8	100.60	Fog

150 rows × 8 columns

Explanation

The result shows the rows which has the Weather Condition is “Fog”

Q. 14) Find all instances when 'Weather is Clear' or 'Visibility is above 40'.

CODE

df[(df['Weather Condition']=='Clear') | (df['Visibility_km']>40)]

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

	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	-1.3	-3.5	18	9	25.0	98.67	Clear
9	1/1/2012 9:00	20.0	3.8	35	17	48.3	100.11	Clear
16	1/1/2012 16:00	23.8	17.6	42	9	25.0	100.52	Clear
17	1/1/2012 17:00	-6.8	-9.8	42	20	48.3	100.76	Mainly Clear
18	1/1/2012 18:00	2.3	-2.4	42	6	48.3	101.05	Cloudy
...	...	...	...	...	...	...	...	...
8774	9/30/2012 21:00	23.0	14.7	92	13	48.3	101.93	Mostly Cloudy
8777	9/30/2012 3:00	9.3	5.8	95	9	48.3	101.25	Mainly Clear
8779	9/30/2012 5:00	1.4	-3.7	97	22	48.3	100.16	Cloudy
8780	9/30/2012 6:00	-4.6	-9.5	98	11	48.3	101.46	Mostly Cloudy
8781	9/30/2012 7:00	1.5	-6.3	99	30	24.1	101.48	Clear

3027 rows × 8 columns

Explanation

The result shows all instances when 'Weather is Clear' or 'Visibility is above 40'.

Q. 15) Find all instances when :

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

or

B. 'Visibility is above 40'

CODE

df[(df['Weather Condition']=='Clear') & (df['Rel Hum_%'] >50) | (df['Visibility_km']> 40)]

```
df[(df['Weather Condition']=='Clear' ) & (df['Rel Hum_%'] >50) | (df['Visibility_km']> 40)]
```

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather Condition
9	1/1/2012 9:00	20.0	3.8	35	17	48.3	100.11	Clear
17	1/1/2012 17:00	-6.8	-9.8	42	20	48.3	100.76	Mainly Clear
18	1/1/2012 18:00	2.3	-2.4	42	6	48.3	101.05	Cloudy
19	1/1/2012 19:00	-12.7	-17.2	43	17	48.3	101.16	Clear
23	1/1/2012 23:00	29.5	16.8	45	4	48.3	101.07	Mainly Clear
...	...	...	...	...	...	...	...	...
8774	9/30/2012 21:00	23.0	14.7	92	13	48.3	101.93	Mostly Cloudy
8777	9/30/2012 3:00	9.3	5.8	95	9	48.3	101.25	Mainly Clear
8779	9/30/2012 5:00	1.4	-3.7	97	22	48.3	100.16	Cloudy
8780	9/30/2012 6:00	-4.6	-9.5	98	11	48.3	101.46	Mostly Cloudy
8781	9/30/2012 7:00	1.5	-6.3	99	30	24.1	101.48	Clear

2864 rows × 8 columns

Explanation

The result shows the all instances when 'Weather is Clear' and 'Relative Humidity is greater than 50' or 'Visibility is above 40'.