WEATHER DATA ALFIDO TECH INTERNSHIP

In [1]:
```python
import pandas as pd
```

In [2]:
```python
dir(pd)
```

```
'read_clipboard',
'read_csv',
'read_excel',
'read_feather',
'read_fwf',
'read_gbq',
'read_hdf',
'read_html',
'read_json',
'read_orc',
'read_parquet',
'read_pickle',
'read_sas',
'read_spss',
'read_sql',
'read_sql_query',
'read_sql_table',
'read_stata',
'read_table',
'read_xml',
'reset_option',
```

In [3]:
```python
pd.read_csv('Weather Data.csv')
```

Out[3]:

	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	0.1	-2.7	81	30	9.7	100.13	Snow

In [7]:
```python
data = pd.read_csv('Weather Data.csv')
```

In [9]:
```python
import numpy as np
import matplotlib.pyplot as plt
```

In [13]:
```python
data.head(15)
```

Out[13]:

	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
5	1/1/2012 5:00	-1.4	-3.3	87	9	6.4	101.27	Fog
6	1/1/2012 6:00	-1.5	-3.1	89	7	6.4	101.29	Fog
7	1/1/2012 7:00	-1.4	-3.6	85	7	8.0	101.26	Fog
8	1/1/2012 8:00	-1.4	-3.6	85	9	8.0	101.23	Fog
9	1/1/2012 9:00	-1.3	-3.1	88	15	4.0	101.20	Fog
10	1/1/2012 10:00	-1.0	-2.3	91	9	1.2	101.15	Fog
11	1/1/2012 11:00	-0.5	-2.1	89	7	4.0	100.98	Fog
12	1/1/2012 12:00	-0.2	-2.0	88	9	4.8	100.79	Fog
13	1/1/2012 13:00	0.2	-1.7	87	13	4.8	100.58	Fog
14	1/1/2012 14:00	0.8	-1.1	87	20	4.8	100.31	Fog

In [14]:
```python
data.shape
```

Out[14]: (8784, 8)

In [15]:
```python
data.index
```

Out[15]: RangeIndex(start=0, stop=8784, step=1)

In [213]: `data.describe().transpose()`

Out[213]:

	count	mean	std	min	25%	50%	75%	max
Temp_C	8784.0	8.798144	11.687883	-23.30	0.10	9.30	18.80	33.00
Dew Point Temp_C	8784.0	2.555294	10.883072	-28.50	-5.90	3.30	11.80	24.40
Rel Hum_%	8784.0	67.431694	16.918881	18.00	56.00	68.00	81.00	100.00
Wind Speed_km/h	8784.0	14.945469	8.688696	0.00	9.00	13.00	20.00	83.00
Visibility_km	8784.0	27.664447	12.622688	0.20	24.10	25.00	25.00	48.30
Press_kPa	8784.0	101.051623	0.844005	97.52	100.56	101.07	101.59	103.65
year	8784.0	2012.000000	0.000000	2012.00	2012.00	2012.00	2012.00	2012.00
month	8784.0	6.513661	3.451430	1.00	4.00	7.00	10.00	12.00
day	8784.0	15.756831	8.812031	1.00	8.00	16.00	23.00	31.00

In [214]: `data.describe()`

Out[214]:

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	year	
count	8784.000000	8784.000000	8784.000000	8784.000000	8784.000000	8784.000000	8784.0	87
mean	8.798144	2.555294	67.431694	14.945469	27.664447	101.051623	2012.0	
std	11.687883	10.883072	16.918881	8.688696	12.622688	0.844005	0.0	
min	-23.300000	-28.500000	18.000000	0.000000	0.200000	97.520000	2012.0	
25%	0.100000	-5.900000	56.000000	9.000000	24.100000	100.560000	2012.0	
50%	9.300000	3.300000	68.000000	13.000000	25.000000	101.070000	2012.0	
75%	18.800000	11.800000	81.000000	20.000000	25.000000	101.590000	2012.0	
max	33.000000	24.400000	100.000000	83.000000	48.300000	103.650000	2012.0	

In [16]: `data.columns`

Out[16]:
```
Index(['Date/Time', 'Temp_C', 'Dew Point Temp_C', 'Rel Hum_%',
       'Wind Speed_km/h', 'Visibility_km', 'Press_kPa', 'Weather'],
      dtype='object')
```

```
In [18]: data.dtypes
```

```
Out[18]: 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
```

```
In [19]: data['Weather'].unique()
```

```
Out[19]: 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 [20]: dir(np)
```

```
         'sign',
         'signbit',
         'signedinteger',
         'sin',
         'sinc',
         'single',
         'singlecomplex',
         'sinh',
         'size',
         'sometrue',
         'sort',
         'sort_complex',
         'source',
         'spacing',
         'split',
         'sqrt',
         'square',
         'squeeze',
         'stack',
         'std',
```

In [21]:
```python
dir(plt)
```

```
'triplot',
'twinx',
'twiny',
'uninstall_repl_displayhook',
'violinplot',
'viridis',
'vlines',
'waitforbuttonpress',
'winter',
'xcorr',
'xkcd',
'xlabel',
'xlim',
'xscale',
'xticks',
'ylabel',
'ylim',
'yscale',
'yticks']
```

In [23]:
```python
data.nunique()
```

Out[23]:
```
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
```

In [25]:
```python
data['Weather'].nunique()
```

Out[25]: 50

In [26]:
```python
data['Dew Point Temp_C'].nunique()
```

Out[26]: 489

In [29]:
```python
data['Date/Time'].nunique()
```

Out[29]: 8784

In [45]: `data.count()`

Out[45]:
```
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
```

In [51]: `data['Weather'].value_counts()`

```
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 [57]: `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
```

In [59]:
```python
data['Wind Speed_km/h'].max()
```

Out[59]: 83

In [60]:
```python
data['Wind Speed_km/h'].min()
```

Out[60]: 0

In [61]:
```python
data['Temp_C'].max()
```

Out[61]: 33.0

In [62]:
```python
data['Temp_C'].min()
```

Out[62]: -23.3

In [149]:
```python
data.isna().sum()
```

Out[149]:
```
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
year                0
month               0
day                 0
dtype: int64
```

this above is to confirm that there are no missing or blanck values in the data set

In [88]:
```python
data['Date/Time'] = pd.to_datetime(data['Date/Time'])
```

In [76]:
```python
pd.to_datetime(data['Date/Time'])
```

Out[76]:
```
0       2012-01-01 00:00:00
1       2012-01-01 01:00:00
2       2012-01-01 02:00:00
3       2012-01-01 03:00:00
4       2012-01-01 04:00:00
                ...
8779    2012-12-31 19:00:00
8780    2012-12-31 20:00:00
8781    2012-12-31 21:00:00
8782    2012-12-31 22:00:00
8783    2012-12-31 23:00:00
Name: Date/Time, Length: 8784, dtype: datetime64[ns]
```

```
In [77]: data['Date/Time'].head(10)
```

```
Out[77]: 0    2012-01-01 00:00:00
         1    2012-01-01 01:00:00
         2    2012-01-01 02:00:00
         3    2012-01-01 03:00:00
         4    2012-01-01 04:00:00
         5    2012-01-01 05:00:00
         6    2012-01-01 06:00:00
         7    2012-01-01 07:00:00
         8    2012-01-01 08:00:00
         9    2012-01-01 09:00:00
         Name: Date/Time, dtype: datetime64[ns]
```

```
In [87]: data['Date/Time'].dt.to_period('t')
```

```
Out[87]: 0        2012-01-01 00:00
         1        2012-01-01 01:00
         2        2012-01-01 02:00
         3        2012-01-01 03:00
         4        2012-01-01 04:00
                       ...
         8779     2012-12-31 19:00
         8780     2012-12-31 20:00
         8781     2012-12-31 21:00
         8782     2012-12-31 22:00
         8783     2012-12-31 23:00
         Name: Date/Time, Length: 8784, dtype: period[T]
```

```
In [89]: data['Date/Time'].dt.to_period('M')
```

```
Out[89]: 0        2012-01
         1        2012-01
         2        2012-01
         3        2012-01
         4        2012-01
                     ...
         8779     2012-12
         8780     2012-12
         8781     2012-12
         8782     2012-12
         8783     2012-12
         Name: Date/Time, Length: 8784, dtype: period[M]
```

In [91]:
```python
data['Date/Time'].dt.to_period('y').head(10)
```

Out[91]:
```
0    2012
1    2012
2    2012
3    2012
4    2012
5    2012
6    2012
7    2012
8    2012
9    2012
Name: Date/Time, dtype: period[A-DEC]
```

In [110]:
```python
data['year'] = pd.to_datetime(data['Date/Time']).dt.year
data['month'] = pd.to_datetime(data['Date/Time']).dt.month
data['day'] = pd.to_datetime(data['Date/Time']).dt.day
```

In [105]:
```python
data['day'].head(5)
```

Out[105]:
```
0    1
1    1
2    1
3    1
4    1
Name: day, dtype: int64
```

In [106]:
```python
data['month'].head(5)
```

Out[106]:
```
0    1
1    1
2    1
3    1
4    1
Name: month, dtype: int64
```

In [107]:
```python
data['year'].head(5)
```

Out[107]:
```
0    2012
1    2012
2    2012
3    2012
4    2012
Name: year, dtype: int64
```

In [119]:
```python
data['Date/Time'].dt.year
```

Out[119]:
```
0        2012
1        2012
2        2012
3        2012
4        2012
         ...
8779     2012
8780     2012
8781     2012
8782     2012
8783     2012
Name: Date/Time, Length: 8784, dtype: int64
```

In [129]:
```python
data['Wind Speed_km/h'].unique()
```

Out[129]:
```
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],
      dtype=int64)
```

In [130]:
```python
data['Wind Speed_km/h'].nunique()
```

Out[130]: 34

In [133]:
```python
data['month'].unique()
```

Out[133]:
```
array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12], dtype=int64)
```

FILTRATION/ANALYSES

In [ ]:
```
ANALYSING RAINY WEATHER
```

In [158]:
```python
data.Weather=='Rain'
```

Out[158]:
```
0        False
1        False
2        False
3        False
4        False
         ...
8779     False
8780     False
8781     False
8782     False
8783     False
Name: Weather, Length: 8784, dtype: bool
```

In [139]: `data[data.Weather=='Rain'].head(10)`

Out[139]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	year
18	2012-01-01 18:00:00	3.8	1.0	82	15	12.9	99.74	Rain	2012
19	2012-01-01 19:00:00	3.1	1.3	88	15	12.9	99.68	Rain	2012
542	2012-01-23 14:00:00	2.7	-1.1	76	26	48.3	101.01	Rain	2012
543	2012-01-23 15:00:00	2.7	-0.9	77	30	24.1	100.86	Rain	2012
545	2012-01-23 17:00:00	3.1	-0.4	78	28	16.1	100.61	Rain	2012
552	2012-01-24 00:00:00	3.5	2.3	92	17	12.9	100.04	Rain	2012
554	2012-01-24 02:00:00	3.1	2.1	93	9	9.7	100.09	Rain	2012
555	2012-01-24 03:00:00	5.0	3.6	91	19	9.7	100.19	Rain	2012
566	2012-01-24 14:00:00	2.4	-2.4	71	35	24.1	100.75	Rain	2012
633	2012-01-27 09:00:00	0.7	-1.4	86	17	12.9	99.34	Rain	2012

In [157]: `data[data.Weather=='Rain'].value_counts().sum()`

Out[157]: 306

In [160]: `data.groupby('Weather').get_group('Rain')`

Out[160]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	
18	2012-01-01 18:00:00	3.8	1.0	82	15	12.9	99.74	Rain	2
19	2012-01-01 19:00:00	3.1	1.3	88	15	12.9	99.68	Rain	2
542	2012-01-23 14:00:00	2.7	-1.1	76	26	48.3	101.01	Rain	2
543	2012-01-23 15:00:00	2.7	-0.9	77	30	24.1	100.86	Rain	2
545	2012-01-23	3.1	-0.4	78	28	16.1	100.61	Rain	2

In [162]: `data.groupby('Weather').get_group('Rain').head(10)`

Out[162]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	year
18	2012-01-01 18:00:00	3.8	1.0	82	15	12.9	99.74	Rain	2012
19	2012-01-01 19:00:00	3.1	1.3	88	15	12.9	99.68	Rain	2012
542	2012-01-23 14:00:00	2.7	-1.1	76	26	48.3	101.01	Rain	2012
543	2012-01-23 15:00:00	2.7	-0.9	77	30	24.1	100.86	Rain	2012
545	2012-01-23 17:00:00	3.1	-0.4	78	28	16.1	100.61	Rain	2012
552	2012-01-24 00:00:00	3.5	2.3	92	17	12.9	100.04	Rain	2012
554	2012-01-24 02:00:00	3.1	2.1	93	9	9.7	100.09	Rain	2012
555	2012-01-24 03:00:00	5.0	3.6	91	19	9.7	100.19	Rain	2012
566	2012-01-24 14:00:00	2.4	-2.4	71	35	24.1	100.75	Rain	2012
633	2012-01-27 09:00:00	0.7	-1.4	86	17	12.9	99.34	Rain	2012

In [163]: `raingroup = data.groupby('Weather').get_group('Rain').head(10)`

In [164]: `raingroup`

Out[164]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	year
18	2012-01-01 18:00:00	3.8	1.0	82	15	12.9	99.74	Rain	2012
19	2012-01-01 19:00:00	3.1	1.3	88	15	12.9	99.68	Rain	2012
542	2012-01-23 14:00:00	2.7	-1.1	76	26	48.3	101.01	Rain	2012
543	2012-01-23 15:00:00	2.7	-0.9	77	30	24.1	100.86	Rain	2012
545	2012-01-23 17:00:00	3.1	-0.4	78	28	16.1	100.61	Rain	2012
552	2012-01-24 00:00:00	3.5	2.3	92	17	12.9	100.04	Rain	2012
554	2012-01-24 02:00:00	3.1	2.1	93	9	9.7	100.09	Rain	2012
555	2012-01-24 03:00:00	5.0	3.6	91	19	9.7	100.19	Rain	2012
566	2012-01-24 14:00:00	2.4	-2.4	71	35	24.1	100.75	Rain	2012
633	2012-01-27 09:00:00	0.7	-1.4	86	17	12.9	99.34	Rain	2012

FIND THE NUMBER OF TIMES RELATIVE HUMIDITY IS 80

In [167]: `data['Rel Hum_%']== 80`

Out[167]:
```
0       False
1       False
2       False
3       False
4       False
        ...
8779    False
8780    False
8781    False
8782    False
8783    False
Name: Rel Hum_%, Length: 8784, dtype: bool
```

In [169]:
```python
data[data['Rel Hum_%']==80]
```

Out[169]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	
98	2012-01-05 02:00:00	-5.4	-8.3	80	9	8.0	100.28	Snow	̇
140	2012-01-06 20:00:00	-10.2	-13.0	80	11	25.0	100.15	Cloudy	̇
145	2012-01-07 01:00:00	-8.5	-11.4	80	0	19.3	100.11	Cloudy	̇
149	2012-01-07 05:00:00	-6.9	-9.7	80	4	19.3	100.12	Cloudy	̇
150	2012-01-07	-6.7	-9.5	80	7	16.1	100.12	Cloudy	̇

In [177]:
```python
data.isnull().sum()
```

Out[177]:
```
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
year               0
month              0
day                0
dtype: int64
```

this above is to ascertain the number of blanck cells in each columns

In [178]:
```python
data.notnull().sum()
```

Out[178]:
```
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
year               8784
month              8784
day                8784
dtype: int64
```

In [ ]: TO RENAME THE WEATHER COLUMN TO WEATHER CONDITION

In [179]: `data.rename(columns={'Weather':'Weather Condition'}) #temporary`

Out[179]:

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

MEAN, SUM, STANDARD DEVIATION AND VARIANCE OF Press_kPa

In [186]: `data.Press_kPa.mean()`

Out[186]: 101.05162340619307

In [187]: `data.Press_kPa.sum()`

Out[187]: 887637.46

In [188]: `data.Press_kPa.std()`

Out[188]: 0.8440047459486474

In [190]: `data['Press_kPa'].var()`

Out[190]: 0.7123440111838408

Find all instances when fog was recorded

In [198]: `data[data.Weather=='Fog'].head(10)`

Out[198]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	year
0	2012-01-01 00:00:00	-1.8	-3.9	86	4	8.0	101.24	Fog	2012
1	2012-01-01 01:00:00	-1.8	-3.7	87	4	8.0	101.24	Fog	2012
4	2012-01-01 04:00:00	-1.5	-3.3	88	7	4.8	101.23	Fog	2012
5	2012-01-01 05:00:00	-1.4	-3.3	87	9	6.4	101.27	Fog	2012
6	2012-01-01 06:00:00	-1.5	-3.1	89	7	6.4	101.29	Fog	2012
7	2012-01-01 07:00:00	-1.4	-3.6	85	7	8.0	101.26	Fog	2012
8	2012-01-01 08:00:00	-1.4	-3.6	85	9	8.0	101.23	Fog	2012
9	2012-01-01 09:00:00	-1.3	-3.1	88	15	4.0	101.20	Fog	2012
10	2012-01-01 10:00:00	-1.0	-2.3	91	9	1.2	101.15	Fog	2012
11	2012-01-01 11:00:00	-0.5	-2.1	89	7	4.0	100.98	Fog	2012

In [200]: `data[data.Weather=='Fog'].tail(10)`

Out[200]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	yea
8336	2012-12-13 08:00:00	-6.0	-7.1	92	4	8.0	102.67	Fog	201
8337	2012-12-13 09:00:00	-4.7	-6.0	91	4	8.0	102.68	Fog	201
8338	2012-12-13 10:00:00	-2.3	-4.2	87	4	6.4	102.64	Fog	201
8483	2012-12-19 11:00:00	0.8	0.2	96	11	9.7	100.44	Fog	201
8715	2012-12-29 03:00:00	-13.4	-15.2	86	4	4.8	101.33	Fog	201
8716	2012-12-29 04:00:00	-16.0	-17.2	90	6	9.7	101.25	Fog	201
8717	2012-12-29 05:00:00	-14.8	-15.9	91	4	6.4	101.25	Fog	201
8718	2012-12-29 06:00:00	-13.8	-15.3	88	4	9.7	101.25	Fog	201
8719	2012-12-29 07:00:00	-14.8	-16.4	88	7	8.0	101.22	Fog	201
8722	2012-12-29 10:00:00	-12.0	-13.3	90	7	6.4	101.15	Fog	201

In [199]: `data['Weather'].str.contains('Fog')`

Out[199]:
```
0        True
1        True
2        True
3        True
4        True
         ...
8779    False
8780    False
8781    False
8782    False
8783    False
Name: Weather, Length: 8784, dtype: bool
```

In [201]: `data[data['Weather'].str.contains('Fog')].tail(10)`

Out[201]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	ye
8734	2012-12-29 22:00:00	-10.1	-11.6	89	15	2.4	100.20	Snow,Fog	20
8735	2012-12-29 23:00:00	-10.0	-12.0	85	20	6.4	100.19	Snow,Fog	20
8736	2012-12-30 00:00:00	-9.6	-11.3	87	13	3.2	100.23	Snow,Fog	20
8737	2012-12-30 01:00:00	-9.4	-10.5	92	9	2.4	100.22	Snow,Fog	20
8738	2012-12-30 02:00:00	-9.3	-10.4	92	9	4.0	100.28	Snow,Fog	20
8739	2012-12-30 03:00:00	-9.1	-10.4	90	11	3.6	100.30	Snow,Fog	20
8740	2012-12-30 04:00:00	-9.3	-10.6	90	13	9.7	100.28	Snow,Fog	20
8741	2012-12-30 05:00:00	-9.1	-10.4	90	11	4.0	100.32	Snow,Fog	20
8742	2012-12-30 06:00:00	-9.3	-10.8	89	17	8.0	100.39	Snow,Fog	20
8770	2012-12-31 10:00:00	-7.4	-8.9	89	4	6.4	101.05	Snow,Fog	20

In [204]: `data[data['Weather'].str.contains('Fog')].head(10)`

Out[204]:

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

to find all instances when wind speed is below 30 and relative humidity is 88

In [205]: `data[(data['Wind Speed_km/h']<30) & (data['Rel Hum_%']==88)]`

Out[205]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	
3	2012-01-01 03:00:00	-1.5	-3.2	88	6	4.0	101.27	Freezing Drizzle,Fog	2
4	2012-01-01 04:00:00	-1.5	-3.3	88	7	4.8	101.23	Fog	2
9	2012-01-01 09:00:00	-1.3	-3.1	88	15	4.0	101.20	Fog	2
12	2012-01-01 12:00:00	-0.2	-2.0	88	9	4.8	100.79	Fog	2
19	2012-01-01 19:00:00	3.1	1.3	88	15	12.9	99.68	Rain	2
...	...	...	...	...	...	...	...	...	
8720	2012-12-29 08:00:00	-13.6	-15.1	88	11	12.9	101.19	Cloudy	2
8721	2012-12-29 09:00:00	-13.2	-14.8	88	11	12.9	101.13	Cloudy	2
8730	2012-12-29 18:00:00	-9.3	-10.9	88	26	6.4	100.38	Snow,Fog	2
8772	2012-12-31 12:00:00	-5.8	-7.5	88	4	12.9	100.78	Snow	2
8778	2012-12-31 18:00:00	-1.3	-3.1	88	17	9.7	100.19	Snow	2

152 rows × 11 columns

In [ ]: `MAXIMUM AND MINIMUM VALUE OF EACH COLUMN IN RELATION TO WEATHER CONDITION`

In [207]:
```python
data.groupby('Weather').max()
```

Out[207]:

Weather	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Pre
Clear	2012-12-30 20:00:00	32.8	20.4	99	33	48.3	
Cloudy	2012-12-31 06:00:00	30.5	22.6	99	54	48.3	
Drizzle	2012-12-22 01:00:00	18.8	17.7	96	30	25.0	
Drizzle,Fog	2012-12-19 10:00:00	19.9	19.1	100	28	9.7	

In [211]:
```python
data.groupby('Weather').min()
```

Out[211]:

Weather	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Pre
Clear	2012-01-03 19:00:00	-23.3	-28.5	20	0	11.3	
Cloudy	2012-01-01 17:00:00	-21.4	-26.8	18	0	11.3	
Drizzle	2012-01-23 21:00:00	1.1	-0.2	74	0	6.4	
Drizzle,Fog	2012-01-23 20:00:00	0.0	-1.6	85	0	1.0	

DATA VISUALIZATION

In [216]:
```python
import seaborn as sns
```

```
In [237]:  plt.figure(figsize=(8, 6))
           sns.histplot(data['Rel Hum_%'], bins=20, color='red', kde=True)
           plt.xlabel('Rel Hum_%')
           plt.ylabel('Frequency')
           plt.title('Rel Hum_%')
           plt.tight_layout()
           plt.show()
```
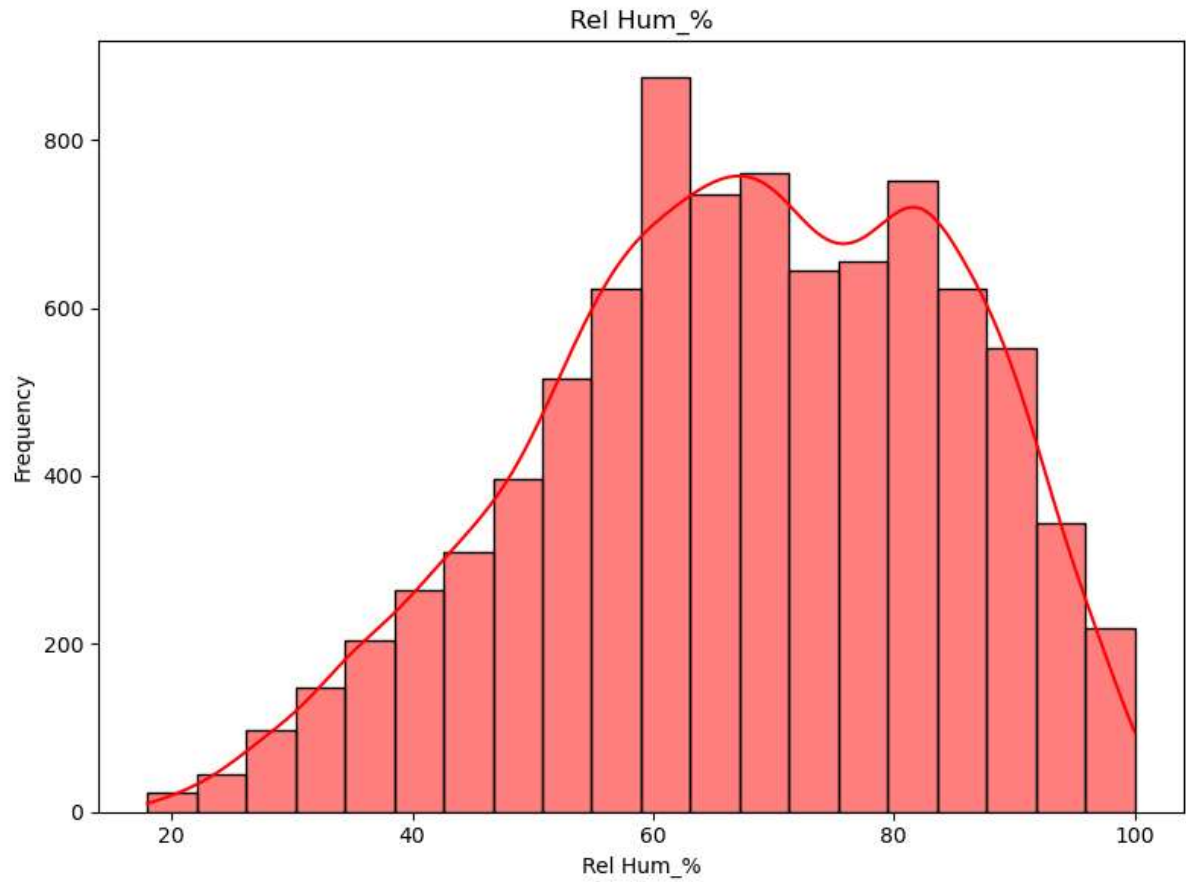
In [229]:
```python
plt.figure(figsize=(8, 6))
sns.boxplot(x=data['Wind Speed_km/h'], color='green')
plt.xlabel('Wind Speed_km/h)')
plt.title('Wind Speed Distribution')
plt.tight_layout()
plt.show()
```


Wind Speed Distribution
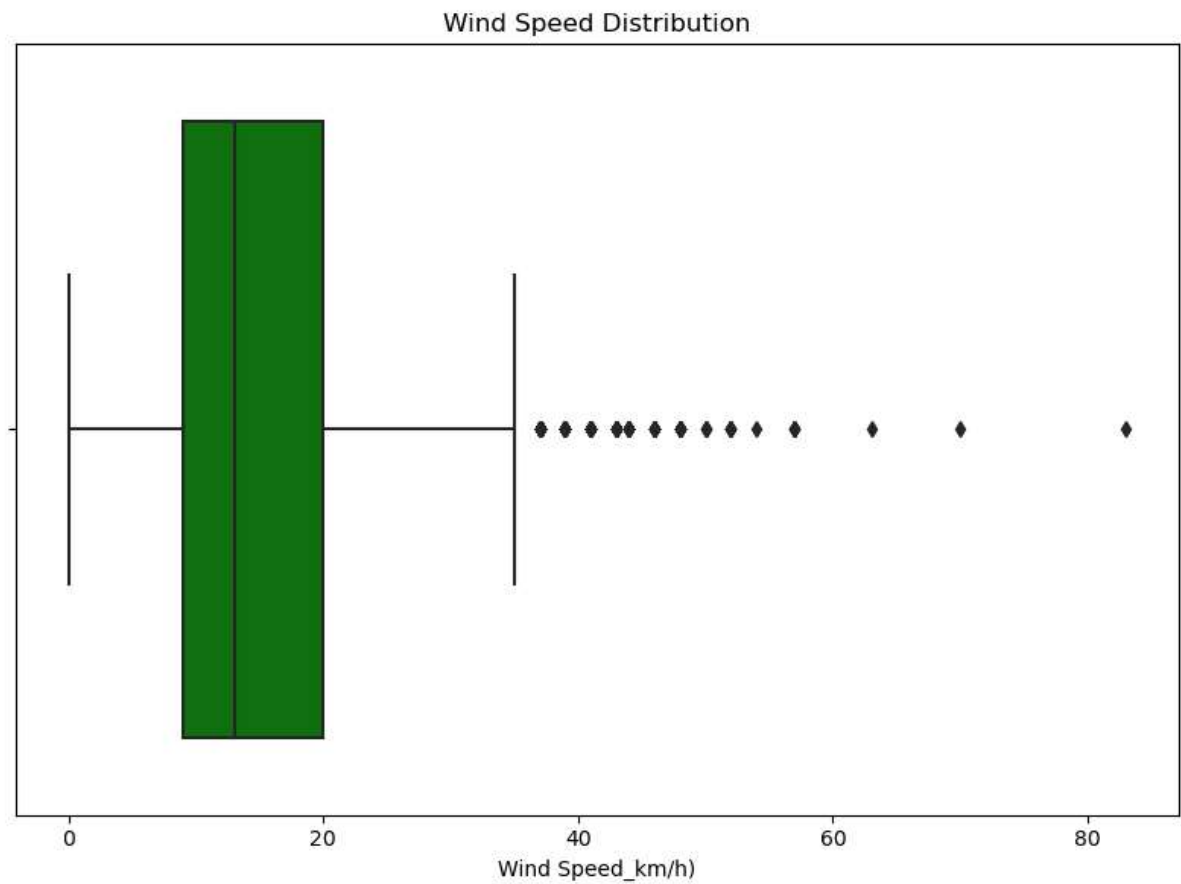
In [234]:
```python
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Temp_C', y='Dew Point Temp_C', data=data, color='orange')
plt.xlabel('Temperature (°C)')
plt.ylabel('Dew Point Temperature (°C)')
plt.title('Temperature vs. Dew Point Temperature')
plt.tight_layout()
plt.show()
```
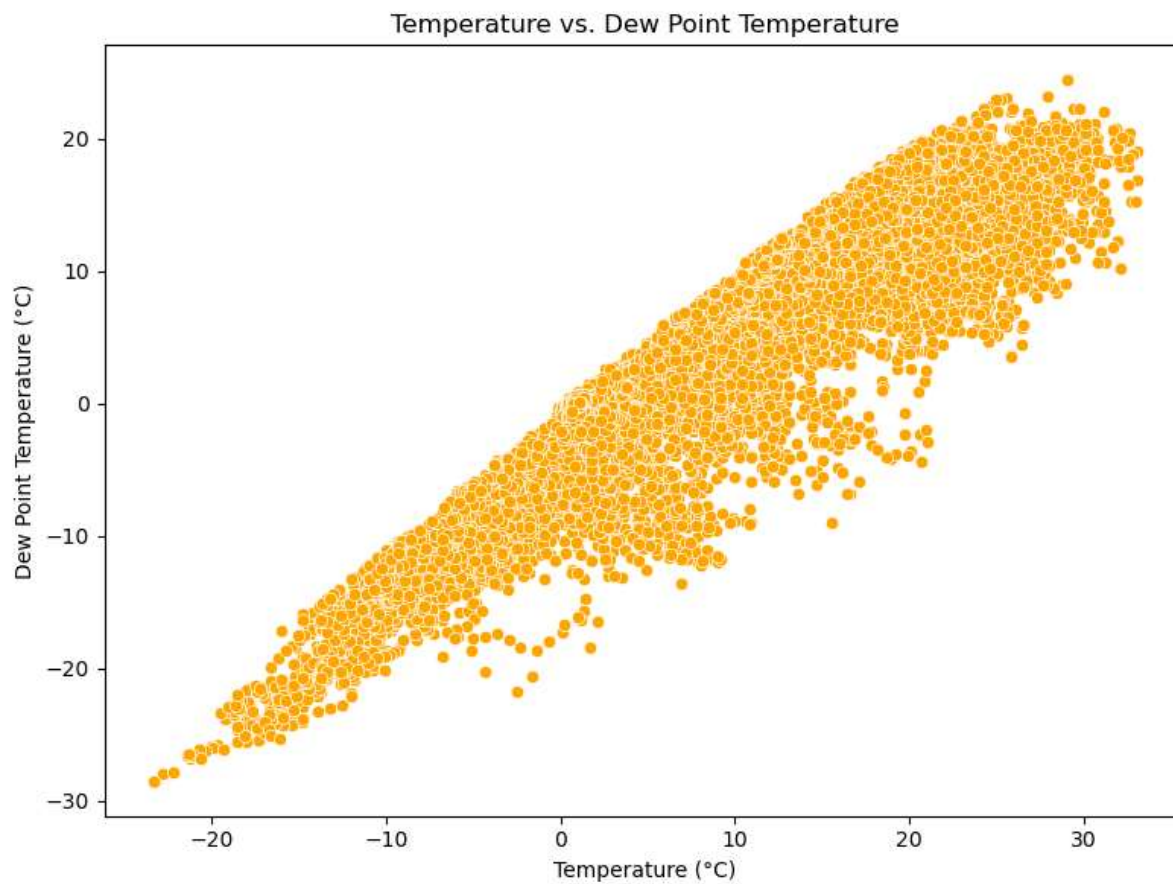
In [247]:
```python
plt.figure(figsize=(10, 6))
sns.countplot(x='Weather', data=data, palette='viridis')
plt.xlabel('Weather Condition')
plt.ylabel('Count')
plt.title('Weather Condition Counts')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```
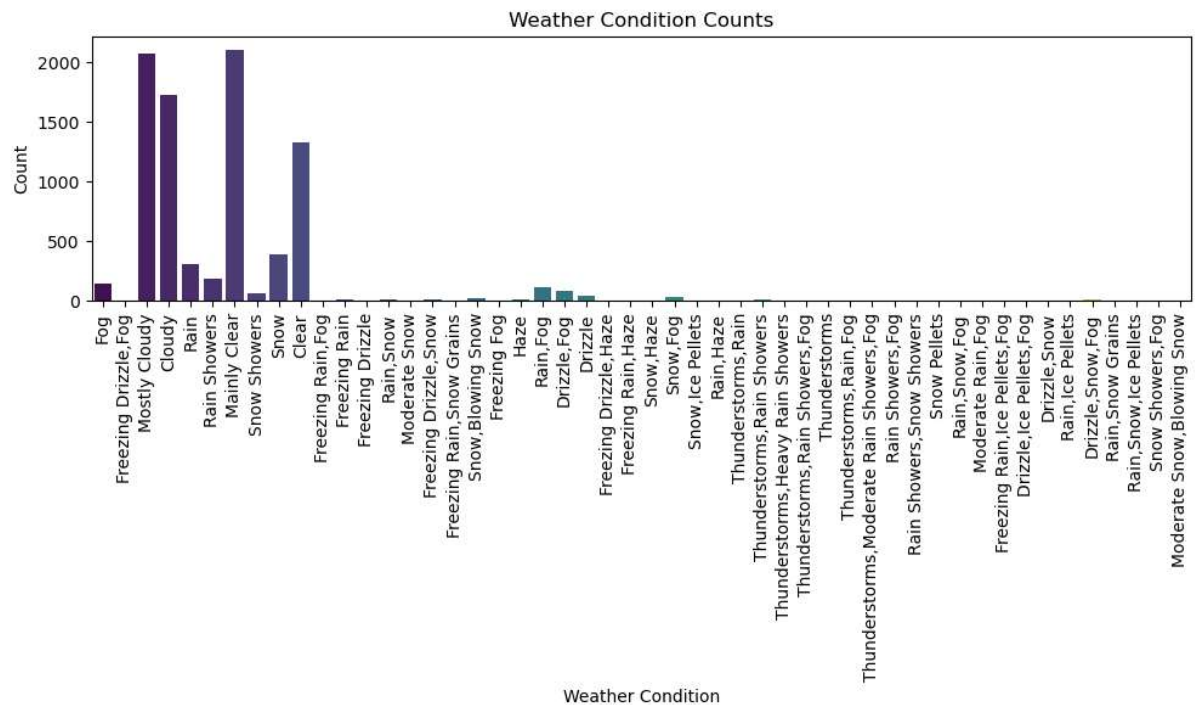

TIME BASED ANALYSIS

In [249]:
```python
data['Year'] = data['Date/Time'].dt.year
data['Month'] = data['Date/Time'].dt.month
data['Day'] = data['Date/Time'].dt.day
data['Hour'] = data['Date/Time'].dt.hour
```

In [256]:
```python
plt.figure(figsize=(12, 4))
sns.lineplot(x='Month', y='Temp_C', data=data.groupby('Month')['Temp_C'].mean()
plt.title('Average Temperature Over Months')
plt.xlabel('Month')
plt.ylabel('Average Temperature (°C)')
plt.show()
```

Average Temperature Over Months
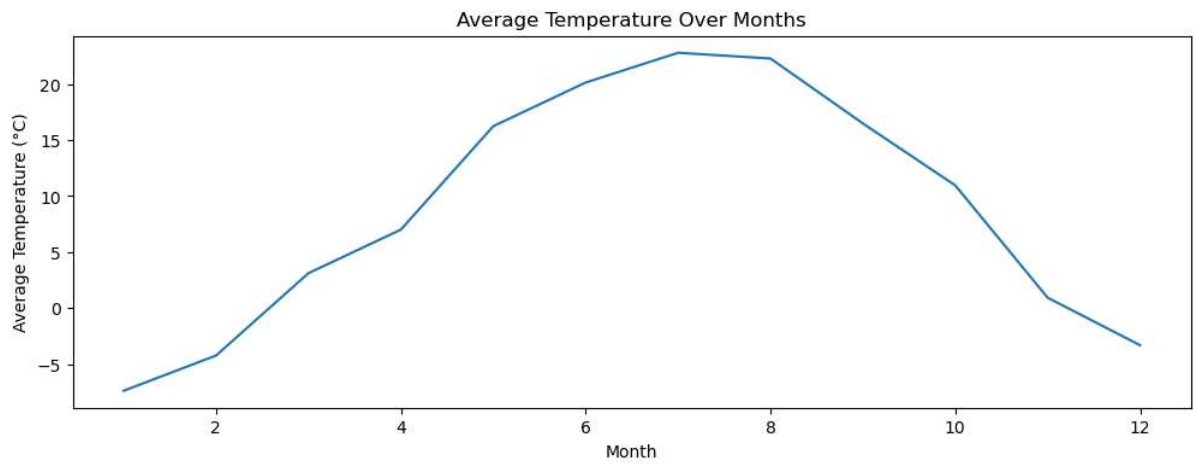
In [266]:
```python
plt.figure(figsize=(10, 6))
sns.histplot(data['Wind Speed_km/h'], bins=20, color='skyblue', kde=True)
plt.title('Distribution of Wind Speeds')
plt.xlabel('Wind Speed (km/h)')
plt.ylabel('Frequency')
plt.show()
```


Distribution of Wind Speeds
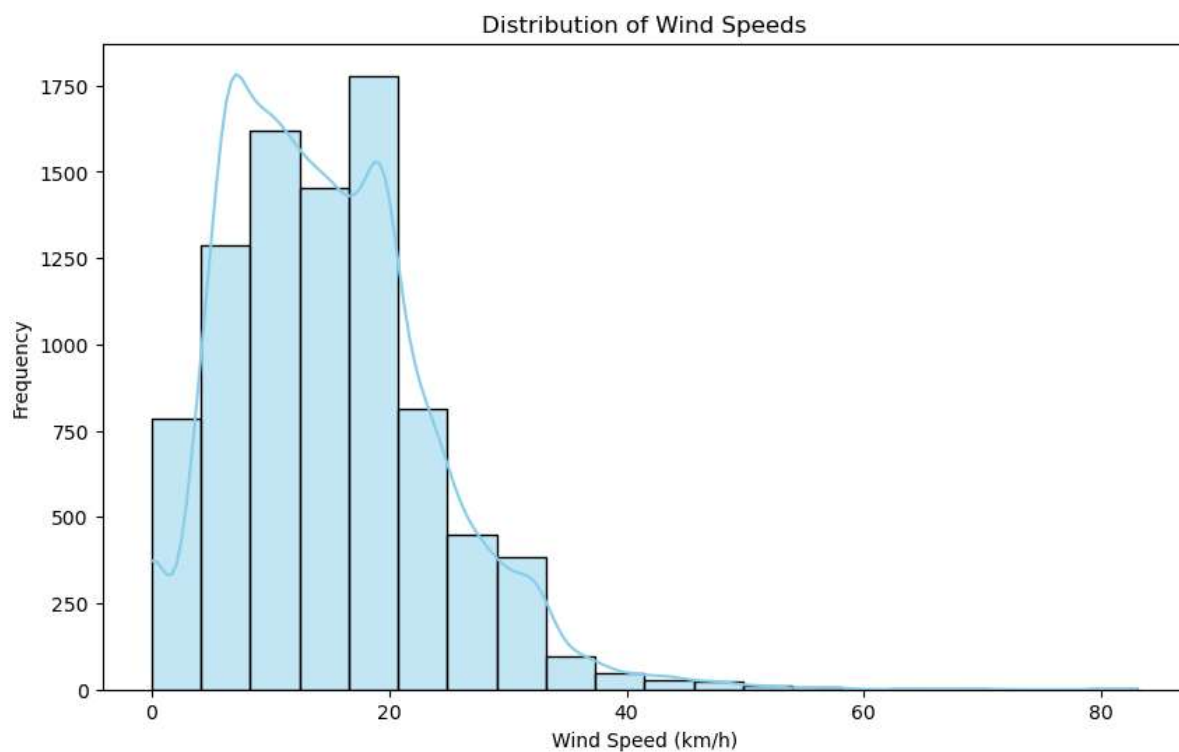
In [267]:
```python
plt.figure(figsize=(12, 6))
sns.lineplot(x='Date/Time', y='Wind Speed_km/h', data=data)
plt.title('Wind Speed Over Time')
plt.xlabel('Date/Time')
plt.ylabel('Wind Speed (km/h)')
plt.xticks(rotation=45)
plt.show()
```
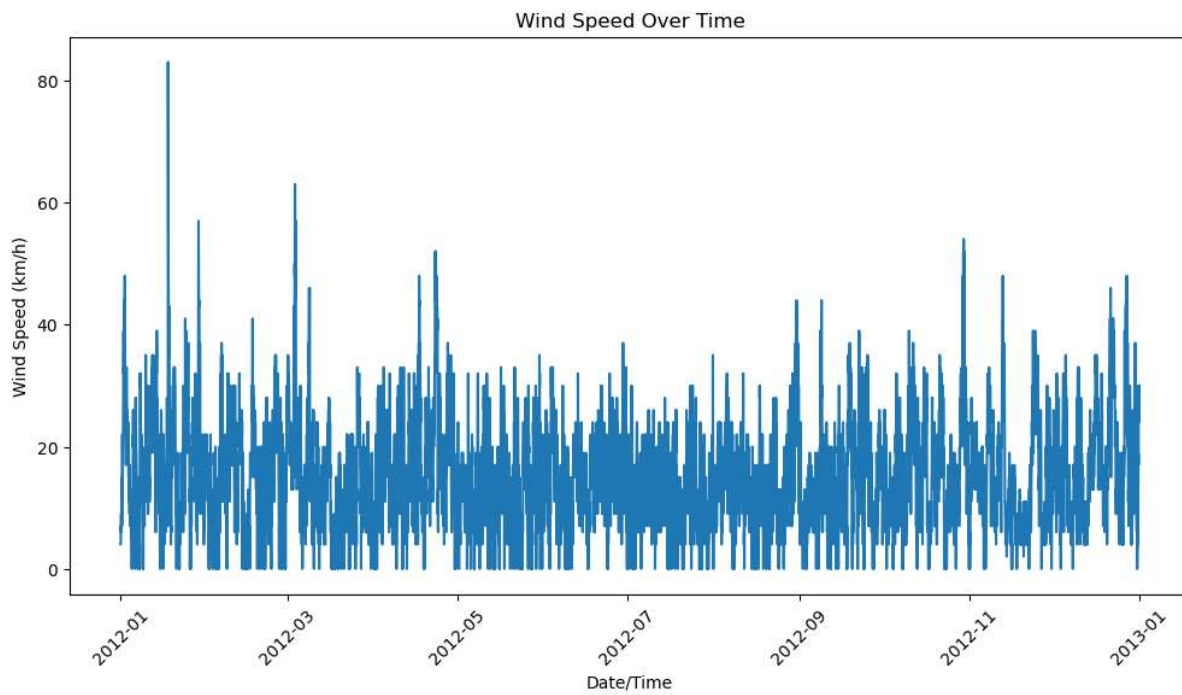

In [269]:
```python
data.to_csv('alfidotech internship project I')
```

In [ ]: