

Weather Data Analysis

```
In [1]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [2]: df = pd.read_csv("C:\\\\Users\\\\Dell\\\\OneDrive\\\\Desktop\\\\STATISTICS\\\\Data analysis\\\\pr
```

Exploratory Data Analysis

```
In [3]: df.head()
```

```
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

```
◀ ▶
```

```
In [4]: df.isna().sum() # no null values exist
```

```
Out[4]: 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
```

```
In [5]: df.describe()
```

```
Out[5]:
```

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

```
In [6]: # Simple and colorful box plots  
# Select only numeric columns  
numeric_cols = df.select_dtypes(include=['float64', 'int64']).columns.tolist()
```

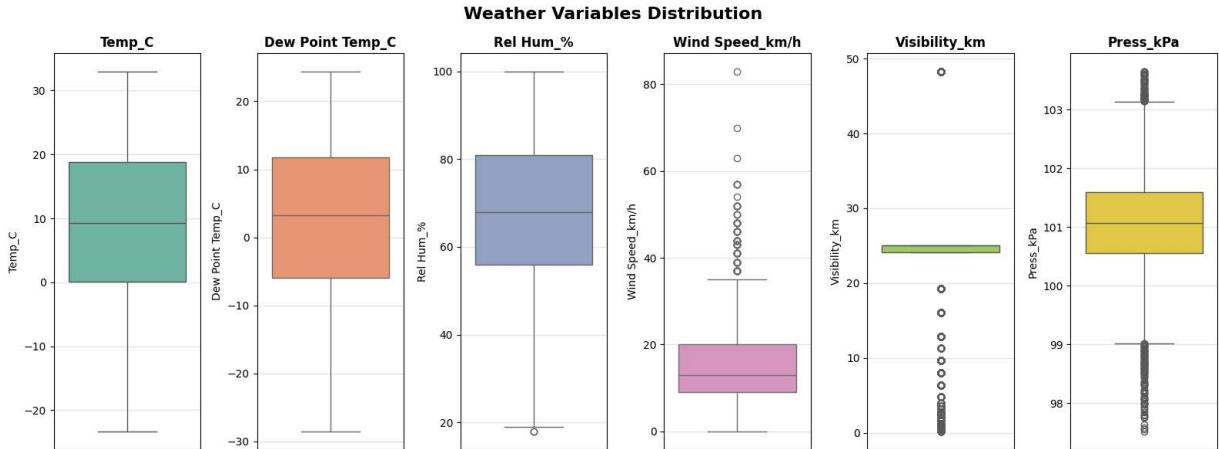
```

fig, axes = plt.subplots(1, len(numeric_cols), figsize=(16, 6))

for i, col in enumerate(numeric_cols):
    sns.boxplot(y=df[col], ax=axes[i], color=sns.color_palette("Set2")[i])
    axes[i].set_title(col, fontsize=12, fontweight='bold')
    axes[i].set_xlabel('')
    axes[i].tick_params(axis='x', bottom=False, labelbottom=False)
    axes[i].grid(True, alpha=0.3, axis='y')

plt.suptitle('Weather Variables Distribution', fontsize=16, fontweight='bold')
plt.tight_layout()
plt.show()

```



In [7]: `df.shape`

Out[7]: (8784, 8)

In [8]: `df.dtypes`

```

Out[8]: Date/Time      str
Temp_C          float64
Dew Point Temp_C  float64
Rel Hum_%       int64
Wind Speed_km/h  int64
Visibility_km    float64
Press_kPa        float64
Weather          str
dtype: object

```

In [9]: `df.nunique()`

```

Out[9]: 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

```

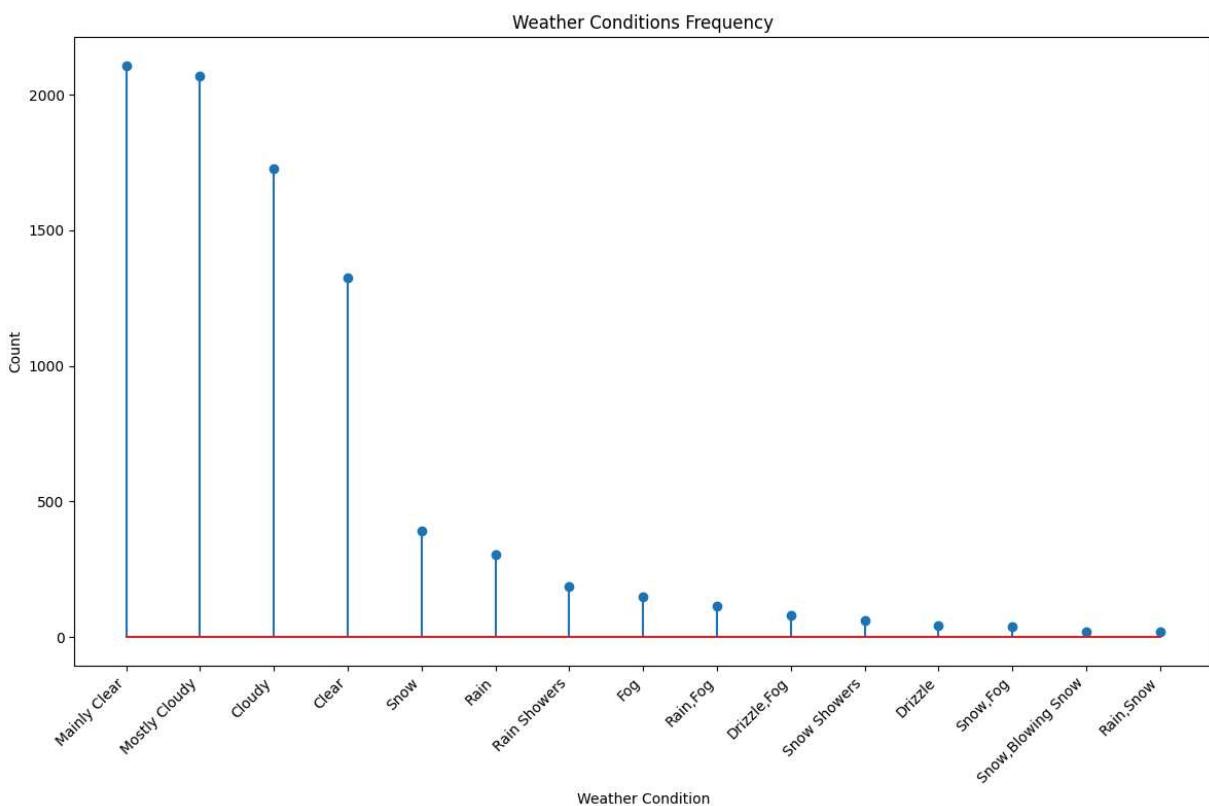
In [10]: `df['Weather'].unique()`

```
Out[10]: <ArrowStringArray>
[               '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']
Length: 50, dtype: str
```

Weather Conditions

```
In [11]: weather_counts=df['Weather'].value_counts()
```

```
In [12]: plt.figure(figsize=(12, 8))
plt.stem(weather_counts.head(15).index, weather_counts.head(15).values)
plt.title('Weather Conditions Frequency')
plt.xlabel('Weather Condition')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



```
In [13]: # How many times Weather is Exactly Clear
print(sum(df['Weather']=='Clear'))
df.groupby("Weather").get_group('Clear')
```

1326

Out[13]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weath
67	1/3/2012 19:00	-16.9	-24.8	50	24	25.0	101.74	Cle
114	1/5/2012 18:00	-7.1	-14.4	56	11	25.0	100.71	Cle
115	1/5/2012 19:00	-9.2	-15.4	61	7	25.0	100.80	Cle
116	1/5/2012 20:00	-9.8	-15.7	62	9	25.0	100.83	Cle
117	1/5/2012 21:00	-9.0	-14.8	63	13	25.0	100.83	Cle
...
8646	12/26/2012 6:00	-13.4	-14.8	89	4	25.0	102.47	Cle
8698	12/28/2012 10:00	-6.1	-8.6	82	19	24.1	101.27	Cle
8713	12/29/2012 1:00	-11.9	-13.6	87	11	25.0	101.31	Cle
8714	12/29/2012 2:00	-11.8	-13.1	90	13	25.0	101.33	Cle
8756	12/30/2012 20:00	-13.8	-16.5	80	24	25.0	101.52	Cle

1326 rows × 8 columns

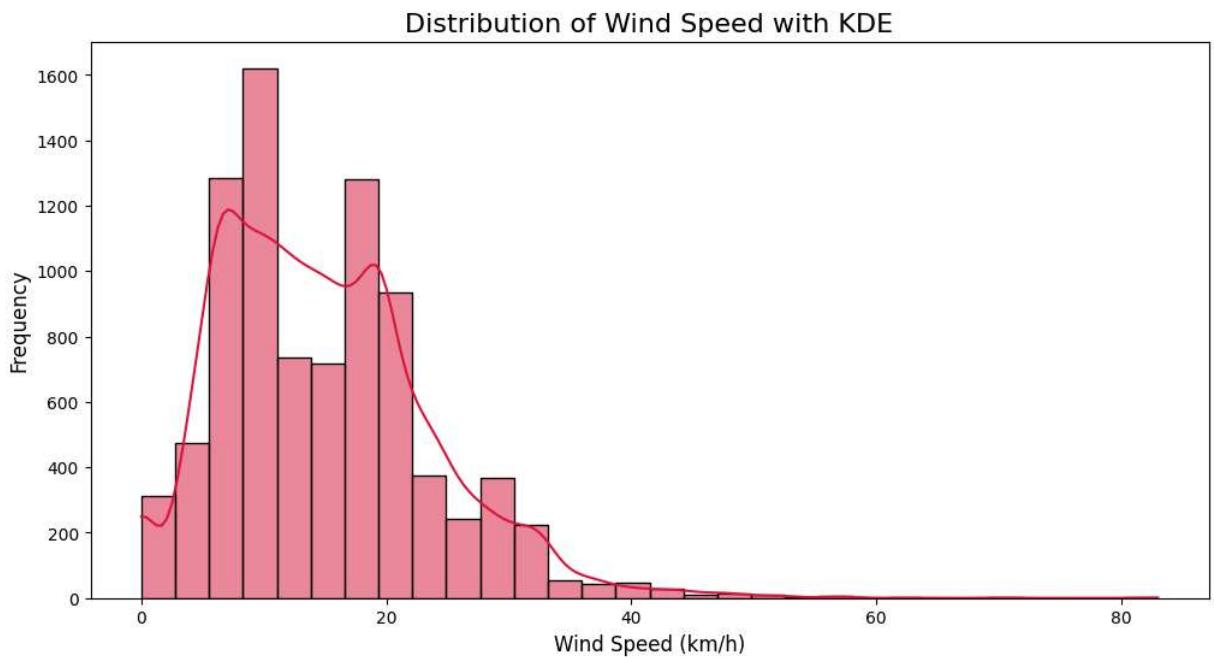


```
In [14]: # How many times Snow was recorded
sum(df['Weather'].str.contains('Snow'))
```

Out[14]: 583

Wind Speed Distribution

```
In [15]: plt.figure(figsize=(12, 6))
sns.histplot(data=df, x='Wind Speed_km/h', bins=30, kde=True, color='crimson')
plt.title('Distribution of Wind Speed with KDE', fontsize=16)
plt.xlabel('Wind Speed (km/h)', fontsize=12)
plt.ylabel('Frequency', fontsize=12)
plt.show()
```



```
In [16]: # How many times Wind Speed is exactly 43km/h
print(sum(df['Wind Speed_km/h']==43))
df[df['Wind Speed_km/h']==43]
```

13

Out[16]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weath
34	1/2/2012 10:00	1.5	-4.1	66	43	24.1	99.18	Mos Cloud
417	1/18/2012 9:00	-11.2	-17.2	61	43	24.1	100.65	Mair Cle
675	1/29/2012 3:00	-0.2	-6.0	65	43	25.0	100.34	Mair Cle
676	1/29/2012 4:00	-0.9	-7.0	63	43	25.0	100.47	Mair Cle
677	1/29/2012 5:00	-1.5	-7.8	62	43	25.0	100.59	Mair Cle
1494	3/3/2012 6:00	1.5	-1.6	80	43	25.0	99.02	Mair Cle
2576	4/17/2012 8:00	10.2	0.6	51	43	48.3	101.42	Mair Cle
2727	4/23/2012 15:00	5.7	2.4	79	43	24.1	98.90	Cloud
2730	4/23/2012 18:00	5.9	2.8	80	43	24.1	98.66	Cloud
5827	8/30/2012 19:00	25.5	16.3	57	43	25.0	100.49	Mos Cloud
6044	9/8/2012 20:00	19.2	13.2	68	43	25.0	99.51	Mair Cle
8536	12/21/2012 16:00	1.4	0.2	92	43	9.7	97.93	Rain
8669	12/27/2012 5:00	-7.5	-9.1	88	43	1.2	100.95	Snow



In [17]:

```
# Wind Speed > 50 and Visibility=25
df[(df['Wind Speed_km/h']>50) & (df['Visibility_km']==25)]
```

Out[17]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weat
409	1/18/2012 1:00	3.7	-2.1	66	83	25.0	98.36	Mo Clo
410	1/18/2012 2:00	0.5	-4.0	72	70	25.0	98.62	Mo Clo
411	1/18/2012 3:00	-3.5	-8.2	70	57	25.0	99.00	Mo Clo
672	1/29/2012 0:00	1.5	-5.0	62	57	25.0	99.96	Ma C
673	1/29/2012 1:00	0.9	-5.3	63	52	25.0	100.09	Ma C
674	1/29/2012 2:00	0.3	-5.8	64	52	25.0	100.21	Ma C
1510	3/3/2012 22:00	2.8	-3.4	64	57	25.0	99.07	Sr
2715	4/23/2012 3:00	1.7	-0.7	84	52	25.0	100.11	Rain,Sr
7268	10/29/2012 20:00	14.3	9.7	74	52	25.0	99.55	Clo
7271	10/29/2012 23:00	14.1	10.0	76	54	25.0	99.33	Clo
7273	10/30/2012 1:00	14.6	10.4	76	52	25.0	99.23	Mo Clo

Mean of each column against each Weather

In [18]:

```
numeric_cols = df.select_dtypes(include=['float64', 'int64']).columns
weather_group = df.groupby('Weather')[numeric_cols]
weather_means=weather_group.mean()
weather_means
```

Out[18]:

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Precipitation_mm
Weather						
Clear	6.825716	0.089367	64.497738	10.557315	30.153243	100
Cloudy	7.970544	2.375810	69.592593	16.127315	26.625752	100
Drizzle	7.353659	5.504878	88.243902	16.097561	17.931707	100
Drizzle,Fog	8.067500	7.033750	93.275000	11.862500	5.257500	100
Drizzle,Ice Pellets,Fog	0.400000	-0.700000	92.000000	20.000000	4.000000	100
Drizzle,Snow	1.050000	0.150000	93.500000	14.000000	10.500000	100
Drizzle,Snow,Fog	0.693333	0.120000	95.866667	15.533333	5.513333	95
Fog	4.303333	3.159333	92.286667	7.946667	6.248000	100
Freezing Drizzle	-5.657143	-8.000000	83.571429	16.571429	9.200000	100
Freezing Drizzle,Fog	-2.533333	-4.183333	88.500000	17.000000	5.266667	100
Freezing Drizzle,Haze	-5.433333	-8.000000	82.000000	10.333333	2.666667	100
Freezing Drizzle,Snow	-5.109091	-7.072727	86.090909	16.272727	5.872727	100
Freezing Fog	-7.575000	-9.250000	87.750000	4.750000	0.650000	100
Freezing Rain	-3.885714	-6.078571	84.642857	19.214286	8.242857	95
Freezing Rain,Fog	-2.225000	-3.750000	89.500000	15.500000	7.550000	95
Freezing Rain,Haze	-4.900000	-7.450000	82.500000	7.500000	2.400000	100
Freezing Rain,Ice Pellets,Fog	-2.600000	-3.700000	92.000000	28.000000	8.000000	100
Freezing Rain,Snow Grains	-5.000000	-7.300000	84.000000	32.000000	4.800000	95
Haze	-0.200000	-2.975000	81.625000	10.437500	7.831250	100
Mainly Clear	12.558927	4.581671	60.667142	14.144824	34.264862	100
Moderate Rain,Fog	1.700000	0.800000	94.000000	17.000000	6.400000	95
Moderate Snow	-5.525000	-7.250000	87.750000	33.750000	0.750000	100
Moderate Snow,Blowing Snow	-5.450000	-6.500000	92.500000	40.000000	0.600000	100
Mostly Cloudy	10.574287	3.131174	62.102465	15.813920	31.253842	100
Rain	9.786275	7.042810	83.624183	19.254902	18.856536	100
Rain Showers	13.722340	9.187766	75.159574	17.132979	22.816489	100
Rain Showers,Fog	12.800000	12.100000	96.000000	13.000000	6.400000	95
Rain Showers,Snow Showers	2.150000	-1.500000	76.500000	22.500000	21.700000	100
Rain,Fog	8.273276	7.219828	93.189655	14.793103	6.873276	100
Rain,Haze	4.633333	2.066667	83.333333	11.666667	6.700000	100
Rain,Ice Pellets	0.600000	-0.600000	92.000000	24.000000	9.700000	100
Rain,Snow	1.055556	-0.566667	89.000000	28.388889	11.672222	95
Rain,Snow Grains	1.900000	-2.100000	75.000000	26.000000	25.000000	100
Rain,Snow,Fog	0.800000	0.300000	96.000000	9.000000	6.400000	100
Rain,Snow,Ice Pellets	1.100000	-0.175000	91.500000	23.250000	6.000000	100
Snow	-4.524103	-7.623333	79.307692	20.038462	11.171795	100

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Precipitation_mm
Weather						
Snow Pellets	0.700000	-6.400000	59.000000	35.000000	2.400000	95.0
Snow Showers	-3.506667	-7.866667	72.350000	19.233333	20.158333	100.0
Snow Showers,Fog	-10.675000	-11.900000	90.750000	13.750000	7.025000	100.0
Snow,Blowing Snow	-5.410526	-7.621053	84.473684	34.842105	4.105263	95.0
Snow,Fog	-5.075676	-6.364865	90.675676	17.324324	4.537838	100.0
Snow,Haze	-4.020000	-6.860000	80.600000	5.000000	4.640000	100.0
Snow,Ice Pellets	-1.883333	-3.666667	87.666667	23.833333	7.416667	100.0
Thunderstorms	24.150000	19.750000	77.000000	7.500000	24.550000	100.0
Thunderstorms,Heavy Rain Showers	10.900000	9.000000	88.000000	9.000000	2.400000	100.0
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	100.0
Thunderstorms,Rain Showers	20.037500	17.618750	86.375000	18.312500	15.893750	100.0
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

Temperature Distribution by Weather Condition

```
In [19]: plt.figure(figsize=(10, 20))
sns.boxplot(data=df, x='Temp_C', y='Weather')
plt.title('Temperature Distribution by Weather Condition', fontsize=16, fontweight='bold')
plt.xlabel('Temperature (°C)')
plt.ylabel('Weather Condition')
plt.grid(True, axis='x', alpha=1, linestyle='--', linewidth=0.5, color='teal')
plt.tight_layout()
plt.show()
```

