

8_Feb_EDA

February 9, 2025

1 Name: Tufan Kundu

2 Reg No. : 24MDT0184

2.1 February EDA Lab

2.2 Time Series analysis on Weather Data

2.3 Importing the necessary libraries

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

2.4 Importing the dataset

```
[4]: df = pd.read_csv("Weather Data.csv")
df
```

```
[4]:
```

	Date/Time	Temp_C	Dew Point	Temp_C	Rel Hum_%	Wind Speed_km/h	\
0	1/1/2012 0:00	-1.8		-3.9	86		4
1	1/1/2012 1:00	-1.8		-3.7	87		4
2	1/1/2012 2:00	-1.8		-3.4	89		7
3	1/1/2012 3:00	-1.5		-3.2	88		6
4	1/1/2012 4:00	-1.5		-3.3	88		7
...	
8779	12/31/2012 19:00	0.1		-2.7	81		30
8780	12/31/2012 20:00	0.2		-2.4	83		24
8781	12/31/2012 21:00	-0.5		-1.5	93		28
8782	12/31/2012 22:00	-0.2		-1.8	89		28
8783	12/31/2012 23:00	0.0		-2.1	86		30

	Visibility_km	Press_kPa	Weather
0	8.0	101.24	Fog
1	8.0	101.24	Fog
2	4.0	101.26	Freezing Drizzle,Fog
3	4.0	101.27	Freezing Drizzle,Fog

4	4.8	101.23	Fog
...
8779	9.7	100.13	Snow
8780	9.7	100.03	Snow
8781	4.8	99.95	Snow
8782	9.7	99.91	Snow
8783	11.3	99.89	Snow

[8784 rows x 8 columns]

```
[5]: 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
```

2.5 Checking for the missing values

```
[6]: df.isna().sum()
```

```
[6]: 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
```

Inference: - No missing values present in the dataset

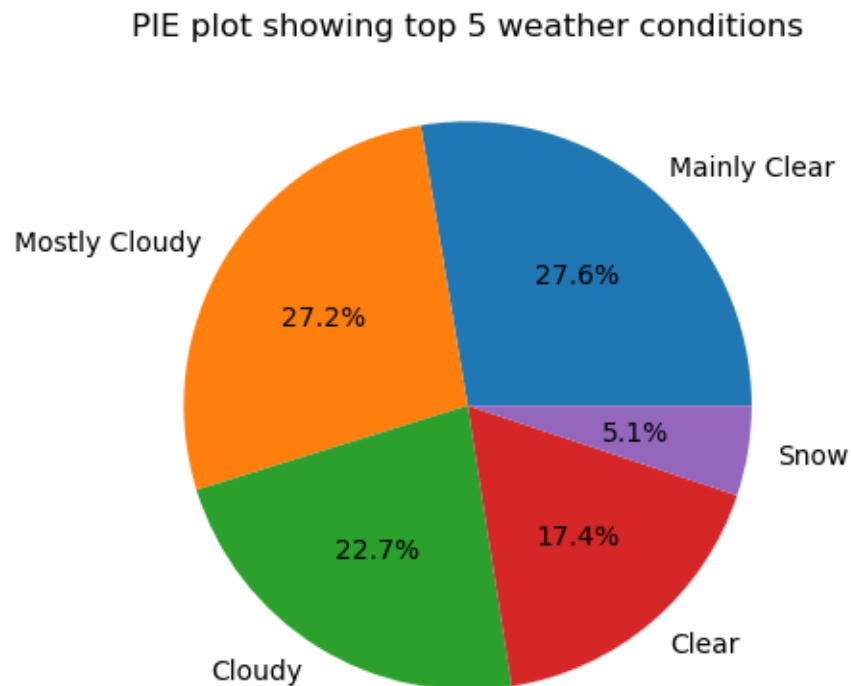
2.6 Checking for duplicates

```
[7]: df.duplicated().sum()
```

[7]: 0

Inference: - No duplicate values present in the dataset

```
[8]: df['Weather'].value_counts().nlargest(5).plot(kind = 'pie', autopct='%1.1f%%')
plt.title("PIE plot showing top 5 weather conditions")
plt.ylabel("")
plt.show()
```



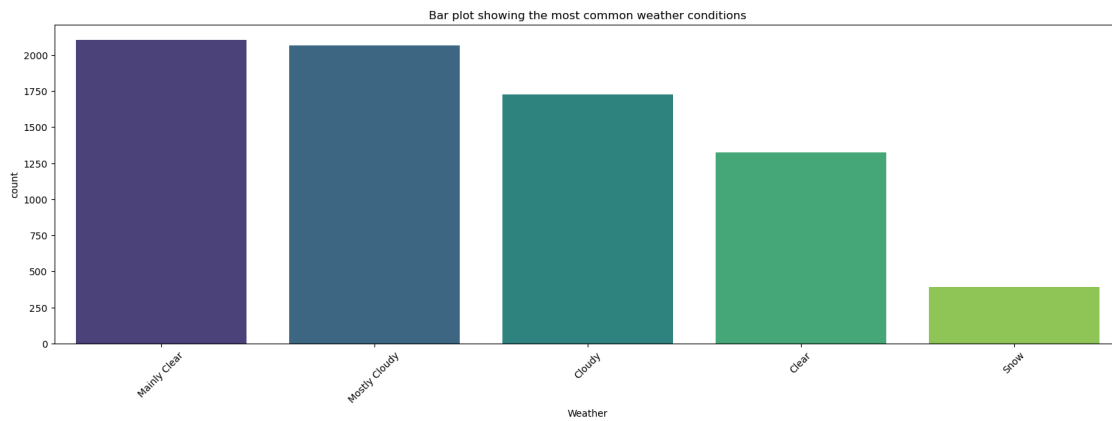
Inference: - The weather is mostly either clear or cloudy followed by 5% snowy weather

```
[9]: plt.figure(figsize=(20,6))
k = df['Weather'].value_counts().nlargest(5)
sns.barplot(data = k,palette='viridis')
plt.xticks(rotation = 45)
plt.title("Bar plot showing the most common weather conditions")
plt.show()
```

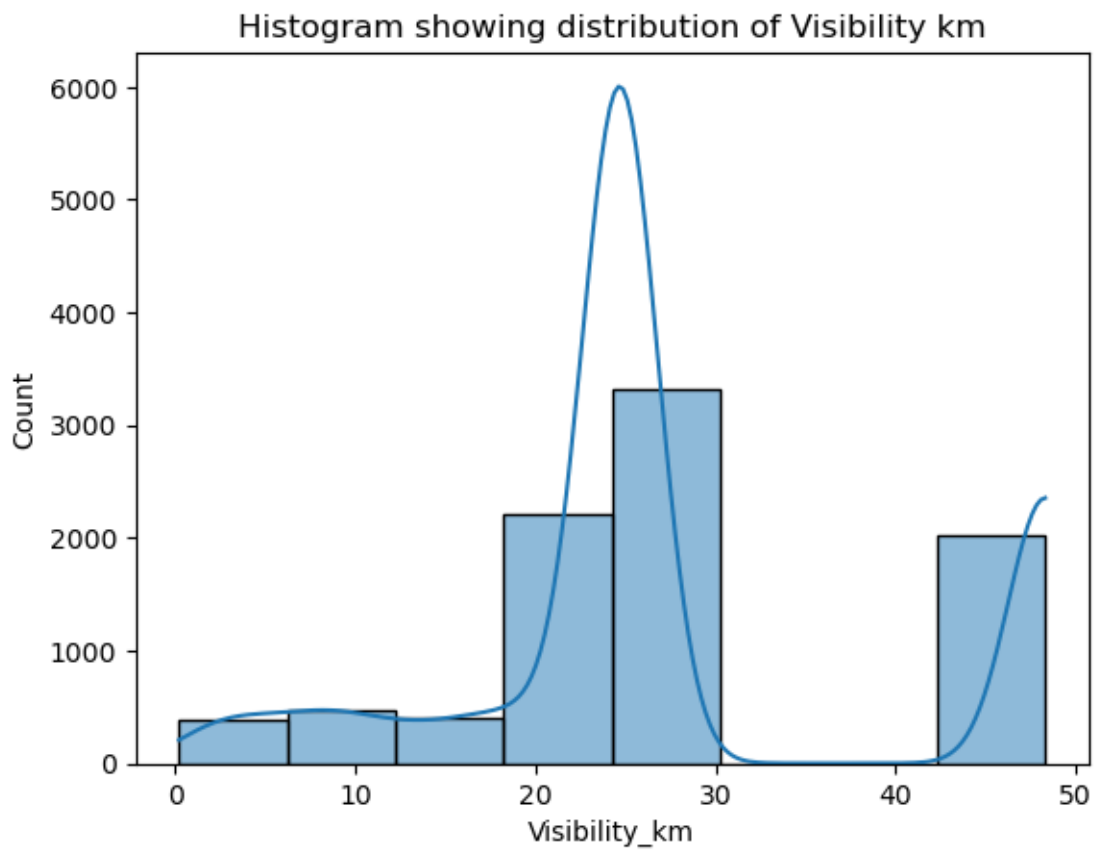
C:\Users\TUFAN\AppData\Local\Temp\ipykernel_6572\2648648754.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data = k,palette='viridis')
```

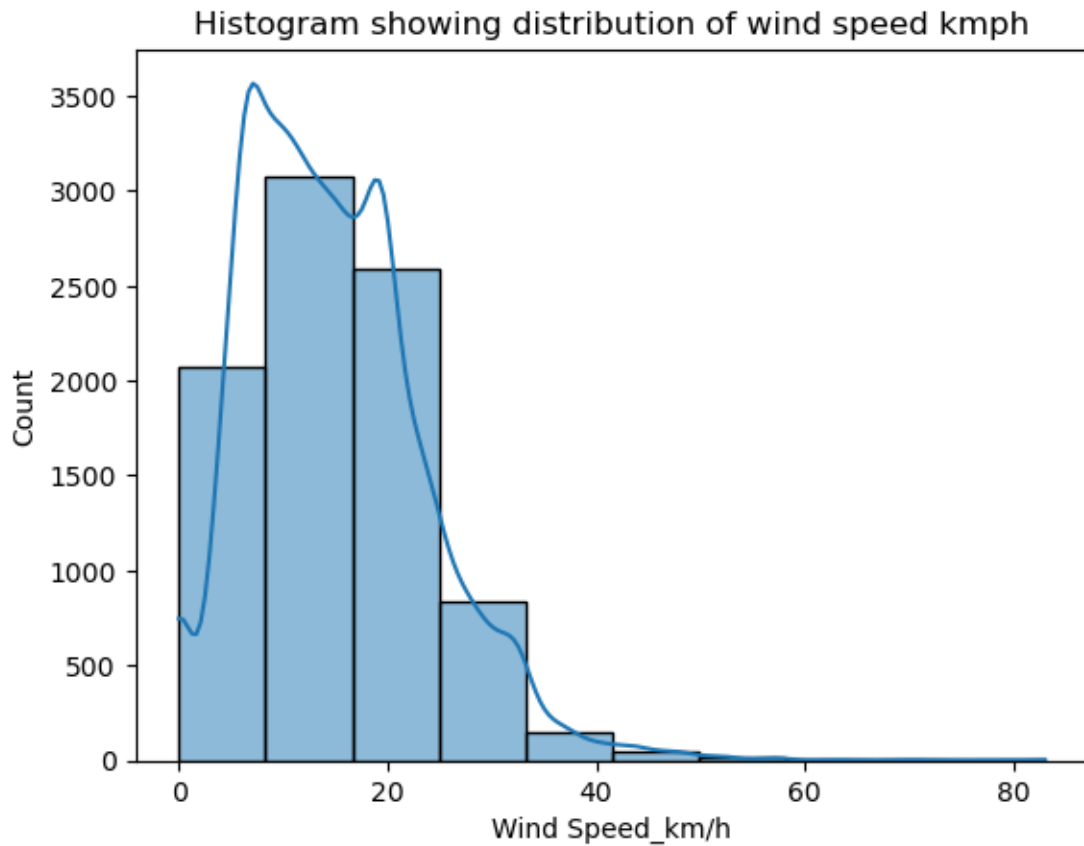


```
[10]: sns.histplot(df['Visibility_km'],kde=True,bins=8)  
plt.title("Histogram showing distribution of Visibility km")  
plt.show()
```



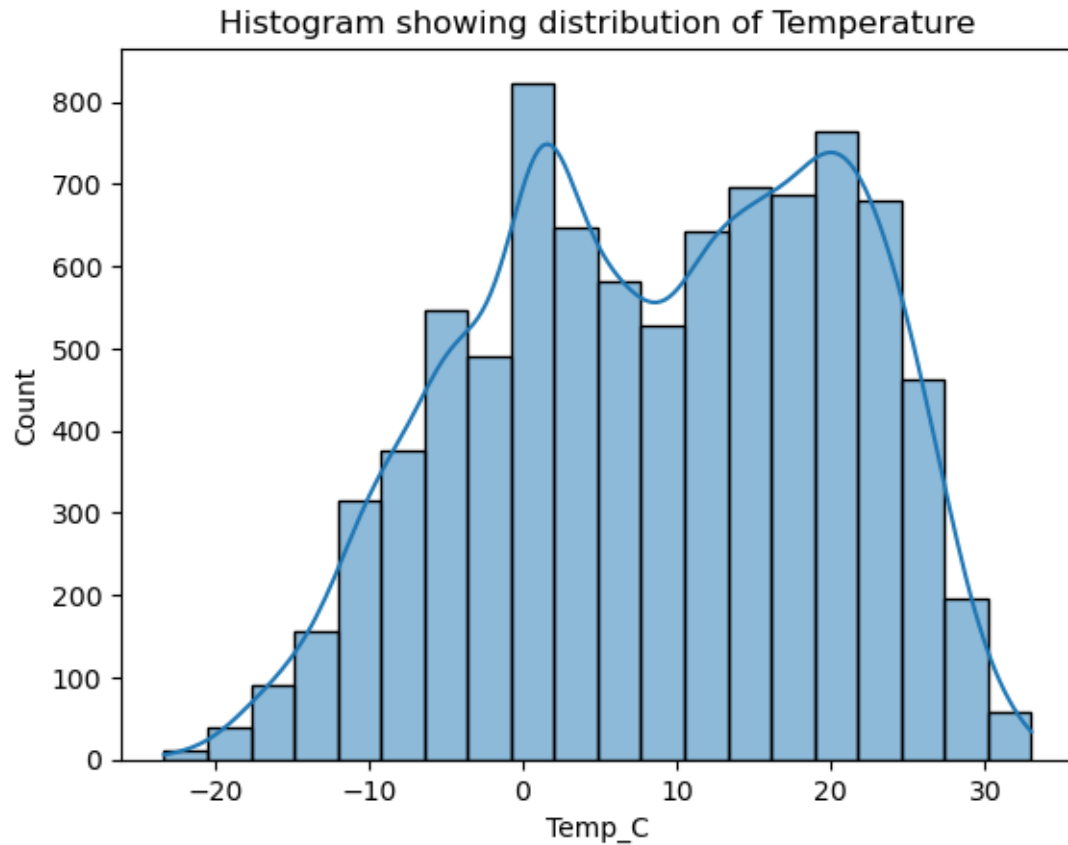
Inference: - Most visible range is between 20-30 km

```
[11]: sns.histplot(df['Wind Speed_km/h'],kde=True,bins=10)
plt.title("Histogram showing distribution of wind speed kmph")
plt.show()
```



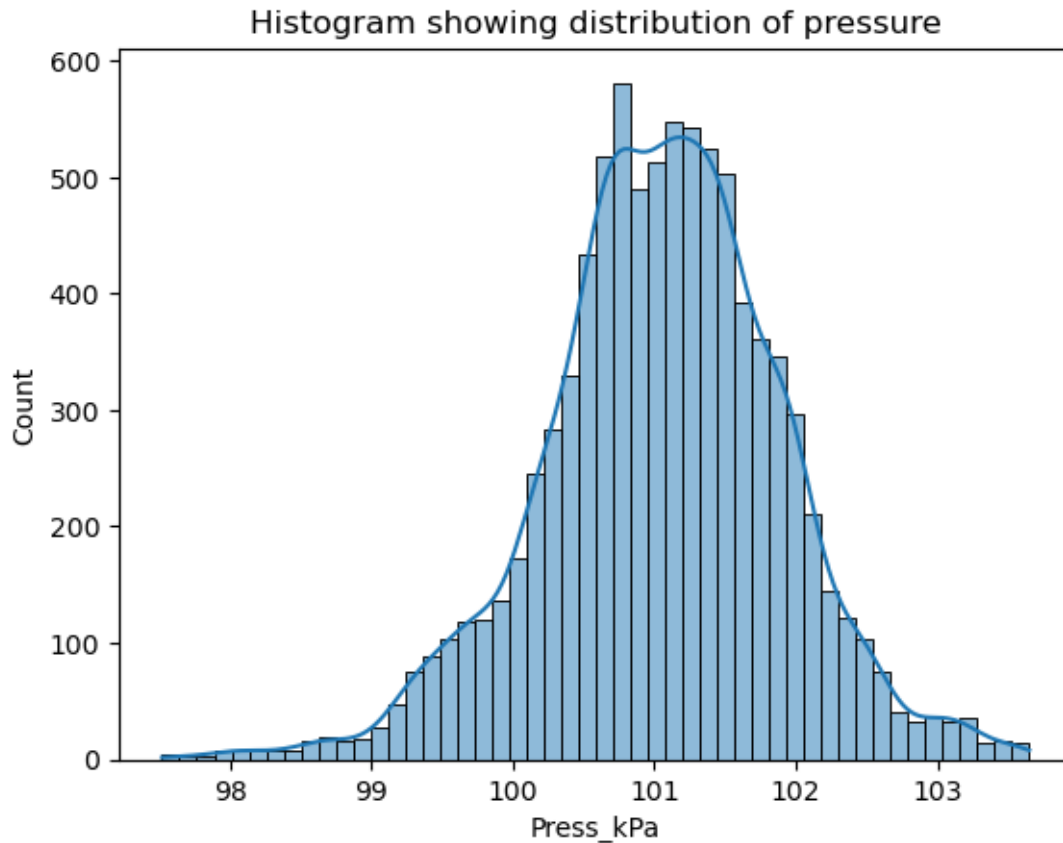
Inference: - Mostly windspeed is around 8-20 kmph

```
[12]: sns.histplot(df['Temp_C'],kde=True,bins=20)
plt.title("Histogram showing distribution of Temperature")
plt.show()
```



Inference: - the temperature is mostly in the cooler side, -1 to 2 degree Celcius

```
[13]: sns.histplot(df['Press_kPa'],kde=True,bins=50)
plt.title("Histogram showing distribution of pressure")
plt.show()
```



Inference: - Average pressure ranges between 100-102 K pascal

```
[14]: df.corr(numeric_only='True')
```

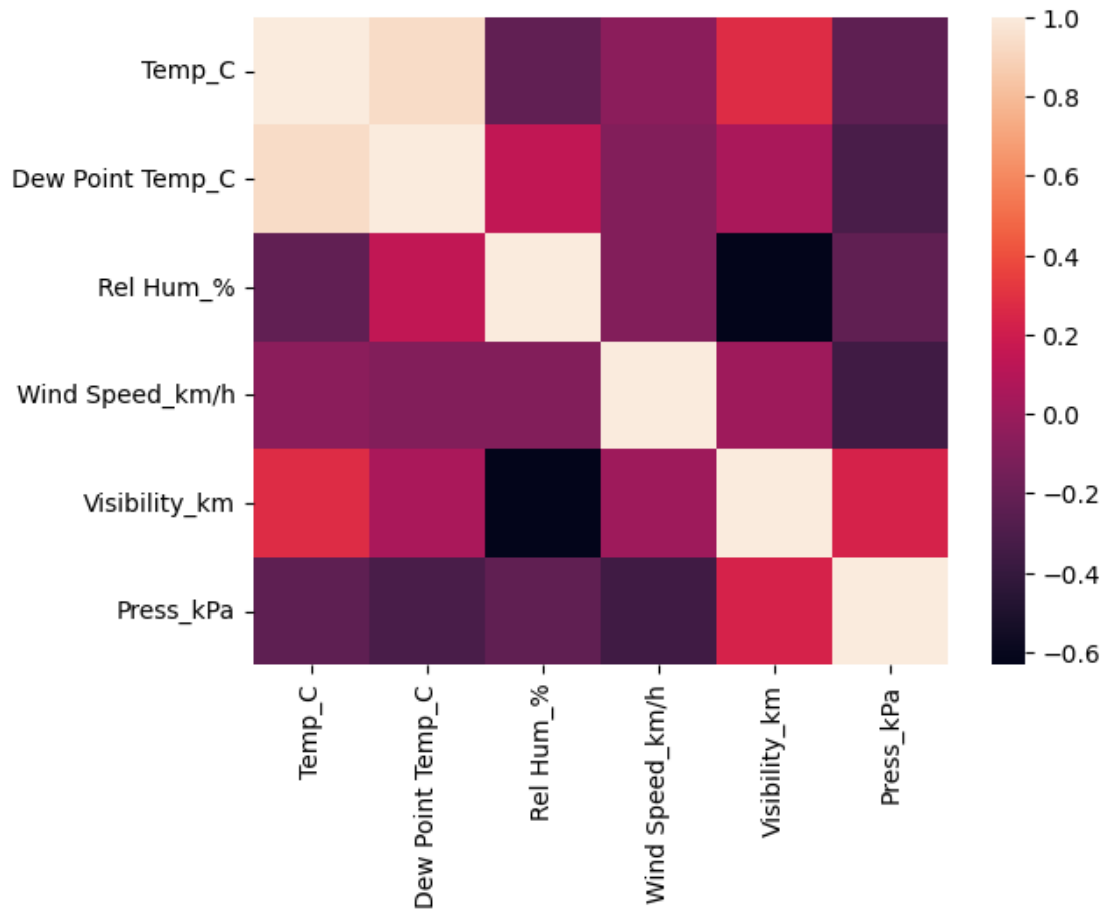
```
[14]:
```

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	\
Temp_C	1.000000	0.932714	-0.220182	-0.061876	
Dew Point Temp_C	0.932714	1.000000	0.139494	-0.095685	
Rel Hum_%	-0.220182	0.139494	1.000000	-0.092743	
Wind Speed_km/h	-0.061876	-0.095685	-0.092743	1.000000	
Visibility_km	0.273455	0.050813	-0.633683	0.004883	
Press_kPa	-0.236389	-0.320616	-0.231424	-0.356613	

	Visibility_km	Press_kPa
Temp_C	0.273455	-0.236389
Dew Point Temp_C	0.050813	-0.320616
Rel Hum_%	-0.633683	-0.231424
Wind Speed_km/h	0.004883	-0.356613
Visibility_km	1.000000	0.231847
Press_kPa	0.231847	1.000000

```
[69]: sns.heatmap(df.corr(numeric_only='True'))
```

```
[69]: <Axes: >
```



Inference: - Dew Point and temperature has a very high positive correlation - relative humidity and visibility has very high negative correlation

```
[15]: top_2_weather = df['Weather'].value_counts().nlargest(2).index
df_filtered = df[df['Weather'].isin(top_2_weather)]
df_filtered
```

```
[15]:
```

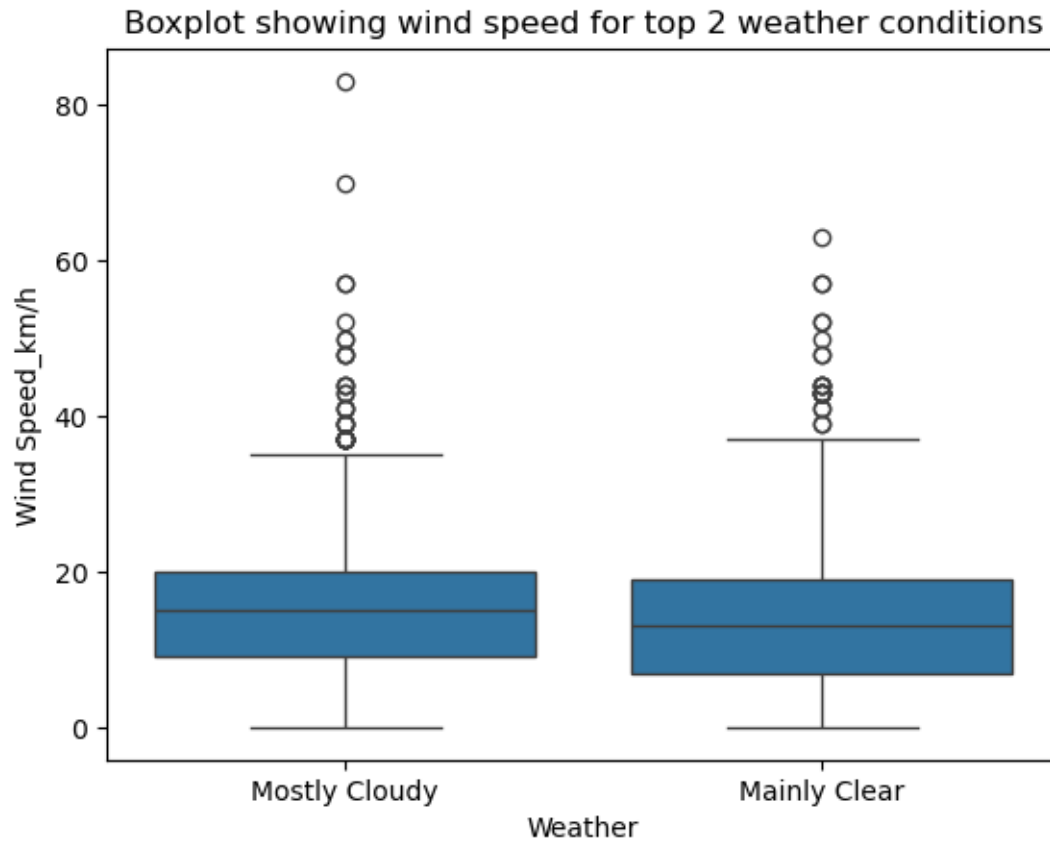
	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	\
16	1/1/2012 16:00	2.6	-0.2	82	13	
26	1/2/2012 2:00	3.9	-0.9	71	32	
27	1/2/2012 3:00	3.7	-1.5	69	33	
28	1/2/2012 4:00	2.9	-2.3	69	32	
29	1/2/2012 5:00	2.6	-2.3	70	32	
...	
8755	12/30/2012 19:00	-13.4	-16.5	77	26	

8757	12/30/2012 21:00	-13.8	-16.5	80	20
8758	12/30/2012 22:00	-13.7	-16.3	81	19
8759	12/30/2012 23:00	-12.1	-15.1	78	28
8763	12/31/2012 3:00	-11.8	-14.4	81	6

	Visibility_km	Press_kPa	Weather
16	12.9	99.93	Mostly Cloudy
26	25.0	99.26	Mostly Cloudy
27	25.0	99.30	Mostly Cloudy
28	25.0	99.26	Mostly Cloudy
29	25.0	99.21	Mostly Cloudy
...
8755	25.0	101.47	Mainly Clear
8757	25.0	101.50	Mainly Clear
8758	25.0	101.54	Mainly Clear
8759	25.0	101.52	Mostly Cloudy
8763	25.0	101.42	Mostly Cloudy

[4175 rows x 8 columns]

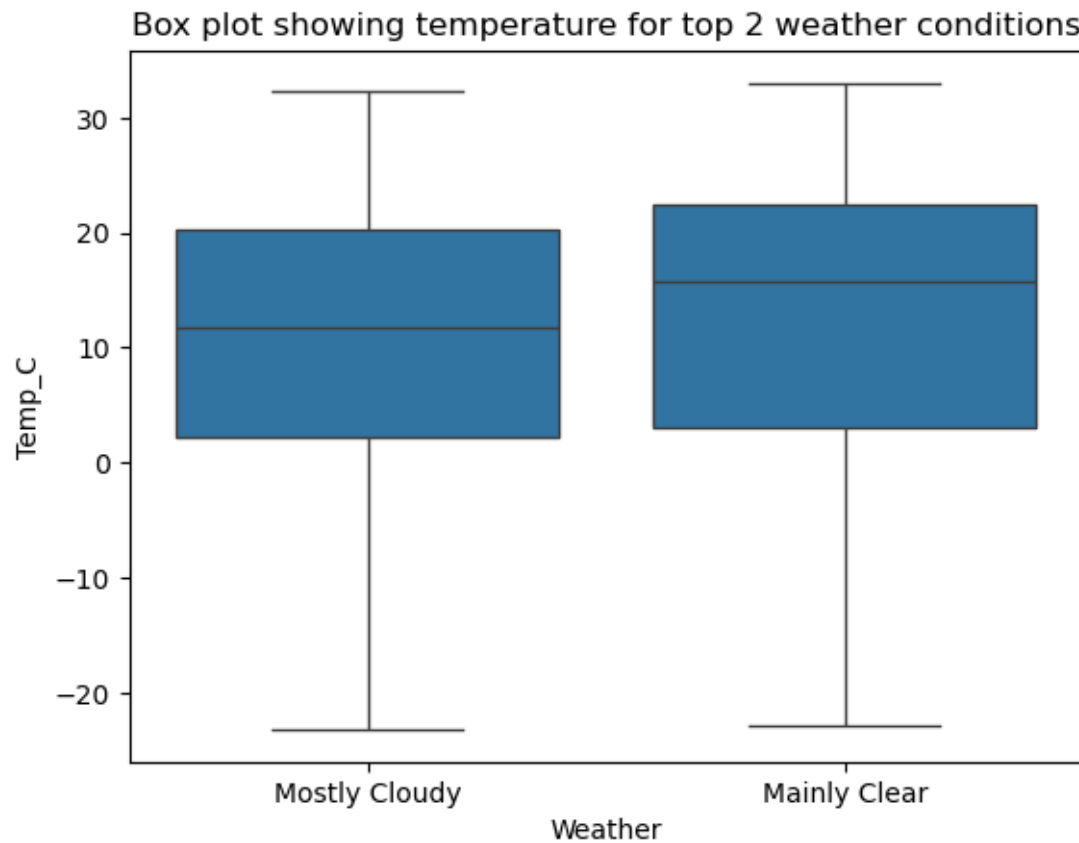
```
[17]: sns.boxplot(x='Weather', y = 'Wind Speed_km/h', data = df_filtered)
plt.title("Boxplot showing wind speed for top 2 weather conditions")
plt.show()
```



2.7 Inference:

- average windspeed ins around 10-20 kmph with some outliers in both the weather

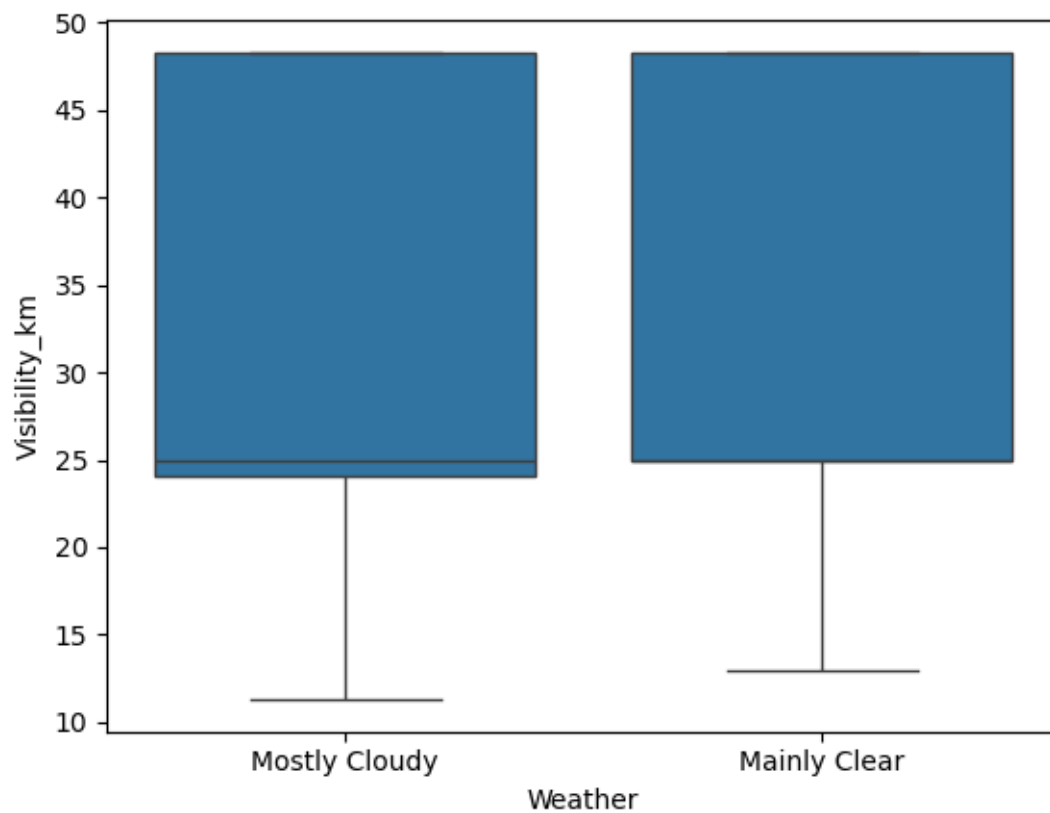
```
[18]: sns.boxplot(x='Weather', y = 'Temp_C', data = df_filtered)
plt.title("Box plot showing temperature for top 2 weather conditions")
plt.show()
```



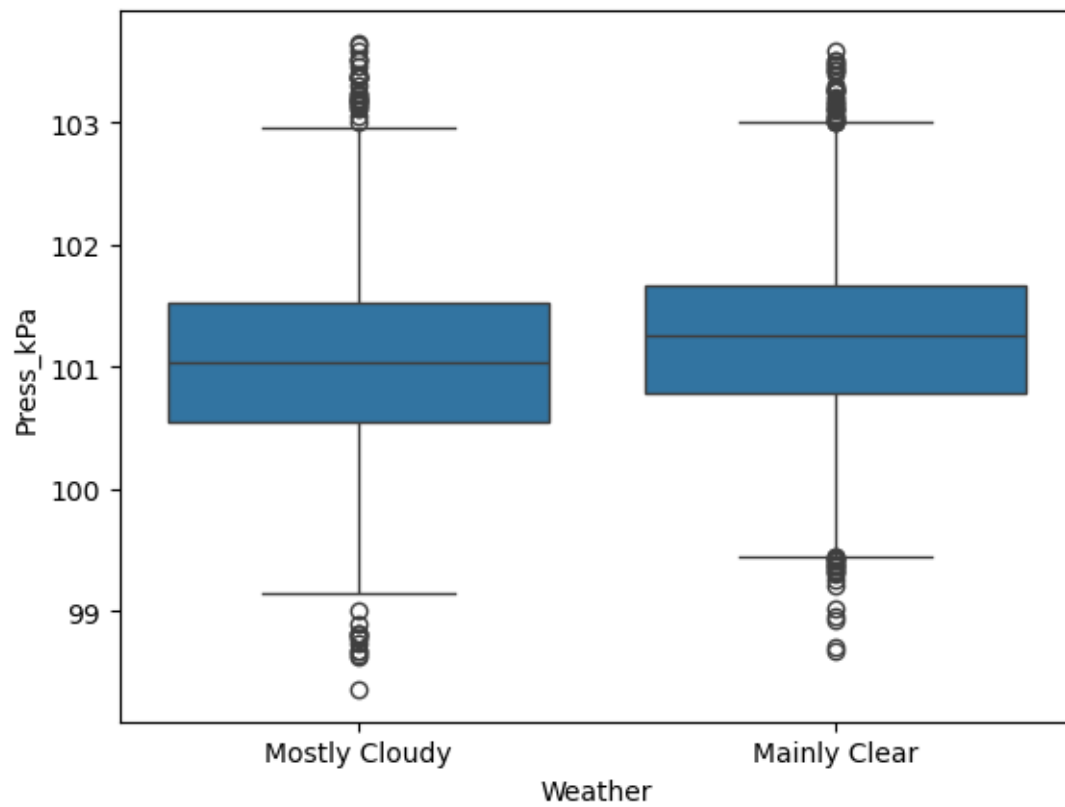
2.8 Inference:

This box plot shows average weather in both the weather condition ranges between 1 to 20 degrees celsius

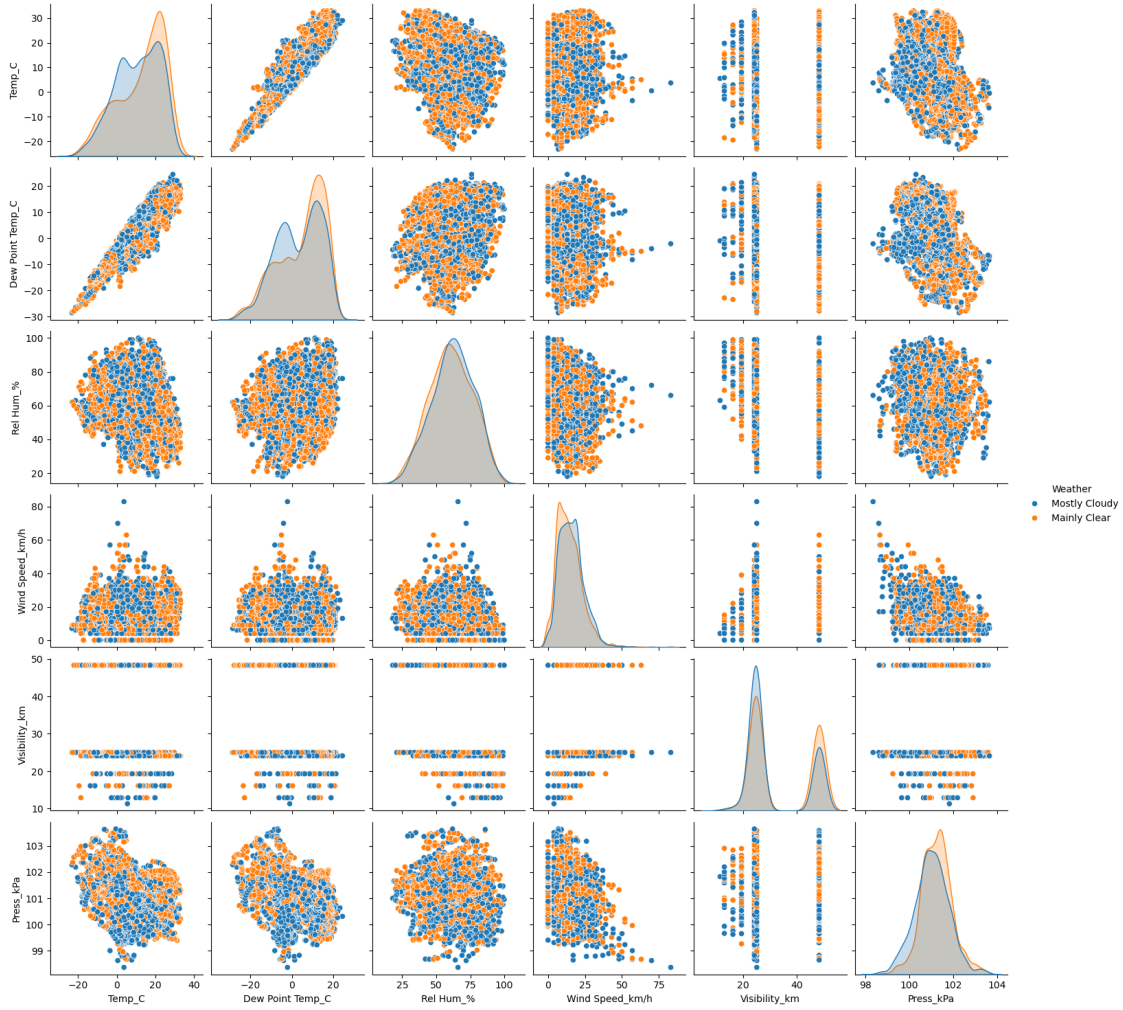
```
[88]: sns.boxplot(x='Weather', y = 'Visibility_km', data = df_filtered)
plt.show()
```



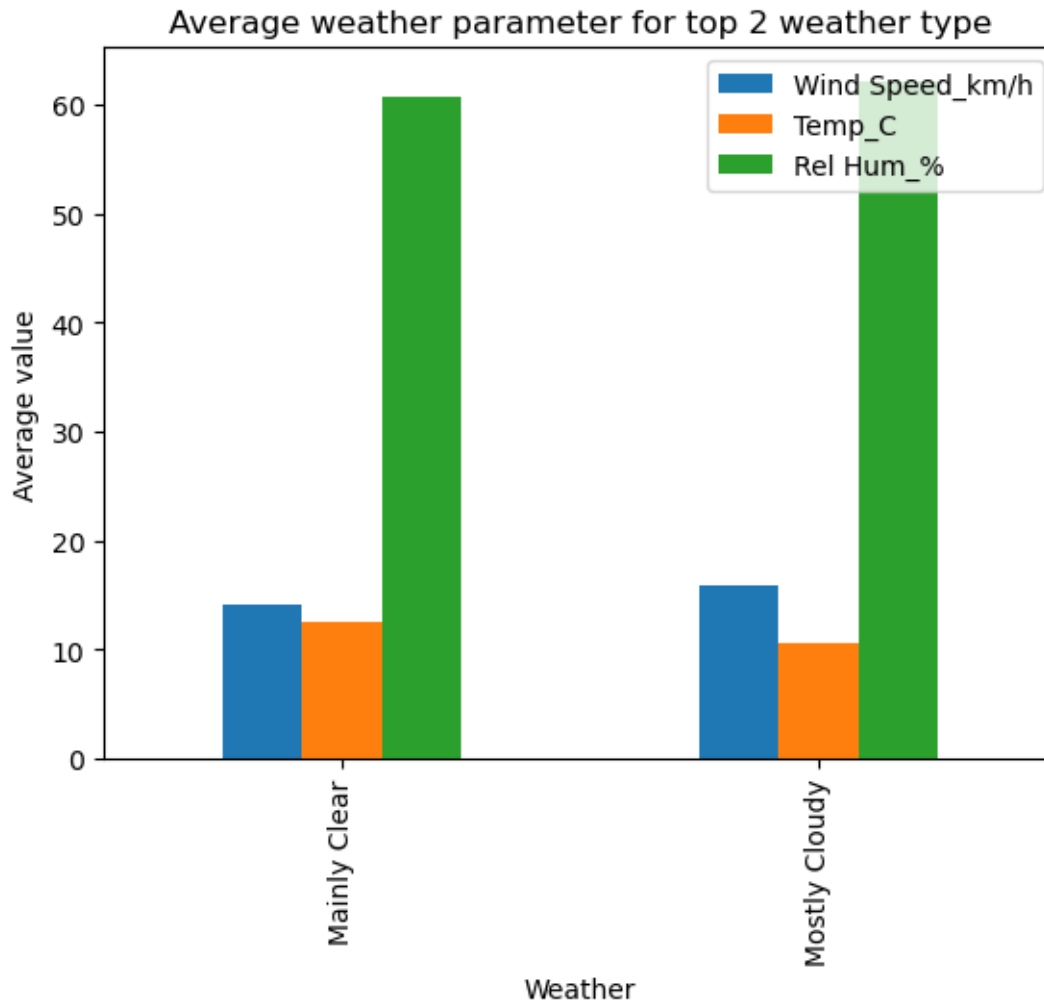
```
[90]: sns.boxplot(x='Weather', y = 'Press_kPa', data = df_filtered)
plt.show()
```



```
[93]: sns.pairplot(df_filtered,hue = 'Weather')  
plt.show()
```



```
[94]: df_grouped = df_filtered.groupby('Weather')[['Wind Speed_kmh', 'Temp_C', 'Rel_
Hum_%']].mean()
df_grouped.plot(kind = 'bar')
plt.title("Average weather parameter for top 2 weather type")
plt.ylabel("Average value")
plt.show()
```



2.9 Inference:

- on a cloudy day the temperature is lesser and the wind speed is more

2.9.1 Grouping by Weather Condition & Computing Summary Statistics

```
[19]: df.groupby('Weather')[['Temp_C', 'Wind Speed_km/h', 'Rel Hum_%', 'Press_kPa']].
      ↪agg(['mean', 'median', 'std'])
```

```
[19]:
```

	Temp_C \		
	mean	median	std
Weather			
Clear	6.825716	7.50	12.132015
Cloudy	7.970544	7.35	11.000283
Drizzle	7.353659	5.40	5.297126
Drizzle,Fog	8.067500	9.15	4.970286

Drizzle,Ice Pellets,Fog	0.400000	0.40	NaN
Drizzle,Snow	1.050000	1.05	0.212132
Drizzle,Snow,Fog	0.693333	0.60	0.317280
Fog	4.303333	2.65	8.038802
Freezing Drizzle	-5.657143	-5.50	2.024728
Freezing Drizzle,Fog	-2.533333	-1.85	2.122891
Freezing Drizzle,Haze	-5.433333	-5.50	0.404145
Freezing Drizzle,Snow	-5.109091	-4.40	1.705552
Freezing Fog	-7.575000	-5.60	9.214255
Freezing Rain	-3.885714	-4.80	2.112183
Freezing Rain,Fog	-2.225000	-1.45	2.772935
Freezing Rain,Haze	-4.900000	-4.90	0.000000
Freezing Rain,Ice Pellets,Fog	-2.600000	-2.60	NaN
Freezing Rain,Snow Grains	-5.000000	-5.00	NaN
Haze	-0.200000	-0.25	6.292535
Mainly Clear	12.558927	15.80	12.348807
Moderate Rain,Fog	1.700000	1.70	NaN
Moderate Snow	-5.525000	-5.45	0.613052
Moderate Snow,Blowing Snow	-5.450000	-5.45	0.070711
Mostly Cloudy	10.574287	11.80	11.312839
Rain	9.786275	8.90	6.404164
Rain Showers	13.722340	14.95	6.977575
Rain Showers,Fog	12.800000	12.80	NaN
Rain Showers,Snow Showers	2.150000	2.15	0.070711
Rain,Fog	8.273276	7.90	6.037209
Rain,Haze	4.633333	4.40	0.776745
Rain,Ice Pellets	0.600000	0.60	NaN
Rain,Snow	1.055556	1.05	0.279121
Rain,Snow Grains	1.900000	1.90	NaN
Rain,Snow,Fog	0.800000	0.80	NaN
Rain,Snow,Ice Pellets	1.100000	1.10	0.163299
Snow	-4.524103	-4.80	4.155435
Snow Pellets	0.700000	0.70	NaN
Snow Showers	-3.506667	-3.70	4.322776
Snow Showers,Fog	-10.675000	-10.70	0.556028
Snow,Blowing Snow	-5.410526	-5.00	2.854995
Snow,Fog	-5.075676	-7.40	4.668714
Snow,Haze	-4.020000	-4.00	0.294958
Snow,Ice Pellets	-1.883333	-2.55	2.017341
Thunderstorms	24.150000	24.15	3.606245
Thunderstorms,Heavy Rain Showers	10.900000	10.90	NaN
Thunderstorms,Moderate Rain Showers,Fog	19.600000	19.60	NaN
Thunderstorms,Rain	20.433333	20.60	0.960902
Thunderstorms,Rain Showers	20.037500	20.85	4.203947
Thunderstorms,Rain Showers,Fog	21.600000	22.40	1.835756
Thunderstorms,Rain,Fog	20.600000	20.60	NaN

	Wind Speed_km/h			\
	mean	median	std	
Weather				
Clear	10.557315	9.0	6.725291	
Cloudy	16.127315	15.0	8.416282	
Drizzle	16.097561	17.0	6.456798	
Drizzle,Fog	11.862500	11.0	6.289923	
Drizzle,Ice Pellets,Fog	20.000000	20.0	NaN	
Drizzle,Snow	14.000000	14.0	7.071068	
Drizzle,Snow,Fog	15.533333	11.0	9.287985	
Fog	7.946667	7.0	4.572733	
Freezing Drizzle	16.571429	19.0	7.502381	
Freezing Drizzle,Fog	17.000000	14.0	11.541230	
Freezing Drizzle,Haze	10.333333	11.0	1.154701	
Freezing Drizzle,Snow	16.272727	19.0	6.574054	
Freezing Fog	4.750000	5.0	3.774917	
Freezing Rain	19.214286	18.5	6.040613	
Freezing Rain,Fog	15.500000	14.5	9.949874	
Freezing Rain,Haze	7.500000	7.5	2.121320	
Freezing Rain,Ice Pellets,Fog	28.000000	28.0	NaN	
Freezing Rain,Snow Grains	32.000000	32.0	NaN	
Haze	10.437500	10.0	4.486554	
Mainly Clear	14.144824	13.0	8.359296	
Moderate Rain,Fog	17.000000	17.0	NaN	
Moderate Snow	33.750000	35.0	5.737305	
Moderate Snow,Blowing Snow	40.000000	40.0	1.414214	
Mostly Cloudy	15.813920	15.0	8.403605	
Rain	19.254902	19.0	9.667949	
Rain Showers	17.132979	17.0	8.376717	
Rain Showers,Fog	13.000000	13.0	NaN	
Rain Showers,Snow Showers	22.500000	22.5	7.778175	
Rain,Fog	14.793103	15.0	6.794521	
Rain,Haze	11.666667	11.0	5.033223	
Rain,Ice Pellets	24.000000	24.0	NaN	
Rain,Snow	28.388889	24.0	12.476514	
Rain,Snow Grains	26.000000	26.0	NaN	
Rain,Snow,Fog	9.000000	9.0	NaN	
Rain,Snow,Ice Pellets	23.250000	24.0	4.856267	
Snow	20.038462	20.0	10.282441	
Snow Pellets	35.000000	35.0	NaN	
Snow Showers	19.233333	20.0	9.388718	
Snow Showers,Fog	13.750000	13.0	6.396614	
Snow,Blowing Snow	34.842105	35.0	6.256225	
Snow,Fog	17.324324	19.0	8.107110	
Snow,Haze	5.000000	4.0	6.164414	
Snow,Ice Pellets	23.833333	24.0	5.115336	
Thunderstorms	7.500000	7.5	10.606602	

Thunderstorms,Heavy Rain Showers	9.000000	9.0	NaN
Thunderstorms,Moderate Rain Showers,Fog	15.000000	15.0	NaN
Thunderstorms,Rain	15.666667	13.0	13.203535
Thunderstorms,Rain Showers	18.312500	17.0	7.002083
Thunderstorms,Rain Showers,Fog	19.666667	17.0	14.189198
Thunderstorms,Rain,Fog	19.000000	19.0	NaN

	Rel Hum_%		\
	mean	median	std
Weather			
Clear	64.497738	65.0	16.171895
Cloudy	69.592593	71.0	14.915147
Drizzle	88.243902	90.0	4.978858
Drizzle,Fog	93.275000	94.0	4.047112
Drizzle,Ice Pellets,Fog	92.000000	92.0	NaN
Drizzle,Snow	93.500000	93.5	2.121320
Drizzle,Snow,Fog	95.866667	97.0	2.166850
Fog	92.286667	92.0	4.501061
Freezing Drizzle	83.571429	83.0	4.649629
Freezing Drizzle,Fog	88.500000	88.5	4.277850
Freezing Drizzle,Haze	82.000000	82.0	1.000000
Freezing Drizzle,Snow	86.090909	85.0	4.206056
Freezing Fog	87.750000	90.5	13.351030
Freezing Rain	84.642857	83.5	3.002746
Freezing Rain,Fog	89.500000	91.5	5.196152
Freezing Rain,Haze	82.500000	82.5	0.707107
Freezing Rain,Ice Pellets,Fog	92.000000	92.0	NaN
Freezing Rain,Snow Grains	84.000000	84.0	NaN
Haze	81.625000	83.0	4.716991
Mainly Clear	60.667142	60.0	15.773872
Moderate Rain,Fog	94.000000	94.0	NaN
Moderate Snow	87.750000	87.5	5.500000
Moderate Snow,Blowing Snow	92.500000	92.5	0.707107
Mostly Cloudy	62.102465	62.0	15.424275
Rain	83.624183	86.0	9.784528
Rain Showers	75.159574	77.0	12.310426
Rain Showers,Fog	96.000000	96.0	NaN
Rain Showers,Snow Showers	76.500000	76.5	2.121320
Rain,Fog	93.189655	94.0	4.434869
Rain,Haze	83.333333	83.0	2.516611
Rain,Ice Pellets	92.000000	92.0	NaN
Rain,Snow	89.000000	89.5	4.325030
Rain,Snow Grains	75.000000	75.0	NaN
Rain,Snow,Fog	96.000000	96.0	NaN
Rain,Snow,Ice Pellets	91.500000	92.0	2.516611
Snow	79.307692	81.0	9.311391
Snow Pellets	59.000000	59.0	NaN

Snow Showers	72.350000	72.5	10.616305
Snow Showers,Fog	90.750000	91.0	1.500000
Snow,Blowing Snow	84.473684	88.0	7.290545
Snow,Fog	90.675676	90.0	4.307965
Snow,Haze	80.600000	81.0	0.547723
Snow,Ice Pellets	87.666667	90.0	6.153590
Thunderstorms	77.000000	77.0	14.142136
Thunderstorms,Heavy Rain Showers	88.000000	88.0	NaN
Thunderstorms,Moderate Rain Showers,Fog	93.000000	93.0	NaN
Thunderstorms,Rain	89.000000	91.0	5.291503
Thunderstorms,Rain Showers	86.375000	87.5	8.562515
Thunderstorms,Rain Showers,Fog	84.000000	81.0	6.082763
Thunderstorms,Rain,Fog	88.000000	88.0	NaN

	Press_kPa		
	mean	median	std
Weather			
Clear	101.587443	101.560	0.668658
Cloudy	100.911441	100.900	0.813979
Drizzle	100.435366	100.610	0.710785
Drizzle,Fog	100.786625	100.765	0.803076
Drizzle,Ice Pellets,Fog	100.790000	100.790	NaN
Drizzle,Snow	100.890000	100.890	0.367696
Drizzle,Snow,Fog	99.281333	99.650	0.905056
Fog	101.184067	101.245	0.877428
Freezing Drizzle	100.202857	100.420	0.854726
Freezing Drizzle,Fog	100.441667	100.640	0.944784
Freezing Drizzle,Haze	100.316667	100.310	0.040415
Freezing Drizzle,Snow	100.520909	100.750	0.667959
Freezing Fog	102.320000	102.230	0.374789
Freezing Rain	99.647143	99.840	0.894861
Freezing Rain,Fog	99.945000	100.225	1.146429
Freezing Rain,Haze	100.375000	100.375	0.049497
Freezing Rain,Ice Pellets,Fog	100.950000	100.950	NaN
Freezing Rain,Snow Grains	98.560000	98.560	NaN
Haze	101.482500	101.120	0.907160
Mainly Clear	101.248832	101.250	0.701775
Moderate Rain,Fog	99.980000	99.980	NaN
Moderate Snow	100.275000	100.275	0.444710
Moderate Snow,Blowing Snow	100.570000	100.570	0.098995
Mostly Cloudy	101.025288	101.030	0.786758
Rain	100.233333	100.330	0.788261
Rain Showers	100.404043	100.495	0.766391
Rain Showers,Fog	99.830000	99.830	NaN
Rain Showers,Snow Showers	101.100000	101.100	0.014142
Rain,Fog	100.500862	100.555	0.618777
Rain,Haze	100.540000	100.510	0.060828

Rain,Ice Pellets	100.120000	100.120	NaN
Rain,Snow	99.951111	99.820	0.724292
Rain,Snow Grains	100.600000	100.600	NaN
Rain,Snow,Fog	100.730000	100.730	NaN
Rain,Snow,Ice Pellets	100.105000	100.050	0.267145
Snow	100.536103	100.625	1.002490
Snow Pellets	99.700000	99.700	NaN
Snow Showers	100.963500	100.980	0.825917
Snow Showers,Fog	101.292500	101.010	0.840253
Snow,Blowing Snow	99.704737	100.010	0.838871
Snow,Fog	100.688649	100.600	0.584476
Snow,Haze	100.782000	100.780	0.148223
Snow,Ice Pellets	100.548333	100.885	0.622042
Thunderstorms	100.230000	100.230	0.551543
Thunderstorms,Heavy Rain Showers	100.260000	100.260	NaN
Thunderstorms,Moderate Rain Showers,Fog	100.010000	100.010	NaN
Thunderstorms,Rain	100.420000	100.240	0.355949
Thunderstorms,Rain Showers	100.233750	100.205	0.424576
Thunderstorms,Rain Showers,Fog	100.063333	99.840	0.503620
Thunderstorms,Rain,Fog	100.080000	100.080	NaN

2.10 Converting Date/Time column to DateTime format

```
[21]: df['Date/Time'] = pd.to_datetime(df['Date/Time'])
```

```
[22]: df
```

```
[22]:
```

	Date/Time	Temp_C	Dew Point	Temp_C	Rel Hum_%	\
0	2012-01-01 00:00:00	-1.8		-3.9	86	
1	2012-01-01 01:00:00	-1.8		-3.7	87	
2	2012-01-01 02:00:00	-1.8		-3.4	89	
3	2012-01-01 03:00:00	-1.5		-3.2	88	
4	2012-01-01 04:00:00	-1.5		-3.3	88	
...	
8779	2012-12-31 19:00:00	0.1		-2.7	81	
8780	2012-12-31 20:00:00	0.2		-2.4	83	
8781	2012-12-31 21:00:00	-0.5		-1.5	93	
8782	2012-12-31 22:00:00	-0.2		-1.8	89	
8783	2012-12-31 23:00:00	0.0		-2.1	86	

	Wind Speed_km/h	Visibility_km	Press_kPa	Weather
0	4	8.0	101.24	Fog
1	4	8.0	101.24	Fog
2	7	4.0	101.26	Freezing Drizzle,Fog
3	6	4.0	101.27	Freezing Drizzle,Fog
4	7	4.8	101.23	Fog
...

8779	30	9.7	100.13	Snow
8780	24	9.7	100.03	Snow
8781	28	4.8	99.95	Snow
8782	28	9.7	99.91	Snow
8783	30	11.3	99.89	Snow

[8784 rows x 8 columns]

```
[35]: df['Year'] = df['Date/Time'].dt.year
df['Month'] = df['Date/Time'].dt.month
df['Day'] = df['Date/Time'].dt.day
df['Weekday'] = df['Date/Time'].dt.day_name()
df['Hour'] = df['Date/Time'].dt.hour
df['Minute'] = df['Date/Time'].dt.minute
df['Second'] = df['Date/Time'].dt.second

df
```

```
[35]:
```

	Date/Time	Temp_C	Dew Point	Temp_C	Rel Hum_%	\
0	2012-01-01 00:00:00	-1.8		-3.9	86	
1	2012-01-01 01:00:00	-1.8		-3.7	87	
2	2012-01-01 02:00:00	-1.8		-3.4	89	
3	2012-01-01 03:00:00	-1.5		-3.2	88	
4	2012-01-01 04:00:00	-1.5		-3.3	88	
...
8779	2012-12-31 19:00:00	0.1		-2.7	81	
8780	2012-12-31 20:00:00	0.2		-2.4	83	
8781	2012-12-31 21:00:00	-0.5		-1.5	93	
8782	2012-12-31 22:00:00	-0.2		-1.8	89	
8783	2012-12-31 23:00:00	0.0		-2.1	86	

	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	Year	\
0	4	8.0	101.24	Fog	2012	
1	4	8.0	101.24	Fog	2012	
2	7	4.0	101.26	Freezing Drizzle,Fog	2012	
3	6	4.0	101.27	Freezing Drizzle,Fog	2012	
4	7	4.8	101.23	Fog	2012	
...
8779	30	9.7	100.13	Snow	2012	
8780	24	9.7	100.03	Snow	2012	
8781	28	4.8	99.95	Snow	2012	
8782	28	9.7	99.91	Snow	2012	
8783	30	11.3	99.89	Snow	2012	

	Month	Day	Hour	Minute	Second	Temp_MA7	Humidity_MA7	Weekday
0	1	1	0	0	0	NaN	NaN	Sunday
1	1	1	1	0	0	NaN	NaN	Sunday

2	1	1	2	0	0	NaN	NaN	Sunday
3	1	1	3	0	0	NaN	NaN	Sunday
4	1	1	4	0	0	NaN	NaN	Sunday
...
8779	12	31	19	0	0	-2.000000	84.285714	Monday
8780	12	31	20	0	0	-1.314286	83.857143	Monday
8781	12	31	21	0	0	-0.900000	85.142857	Monday
8782	12	31	22	0	0	-0.600000	85.857143	Monday
8783	12	31	23	0	0	-0.400000	86.428571	Monday

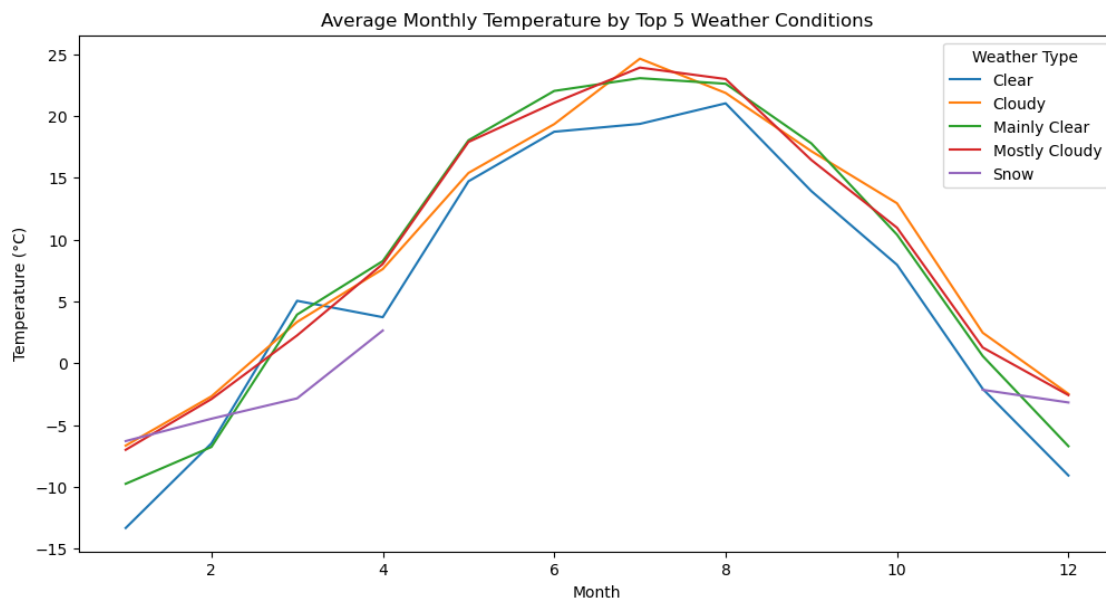
[8784 rows x 17 columns]

```
[ ]: # Selecting the top 5 most common weather conditions
top_5_weather = df['Weather'].value_counts().nlargest(5).index

# Filtering dataset for only these top 5 weather types
df_top5 = df[df['Weather'].isin(top_5_weather)]

# Group by Month and Weather for temperature analysis
df_top5.groupby(['Month', 'Weather'])['Temp_C'].mean().unstack().
    plot(figsize=(12,6))

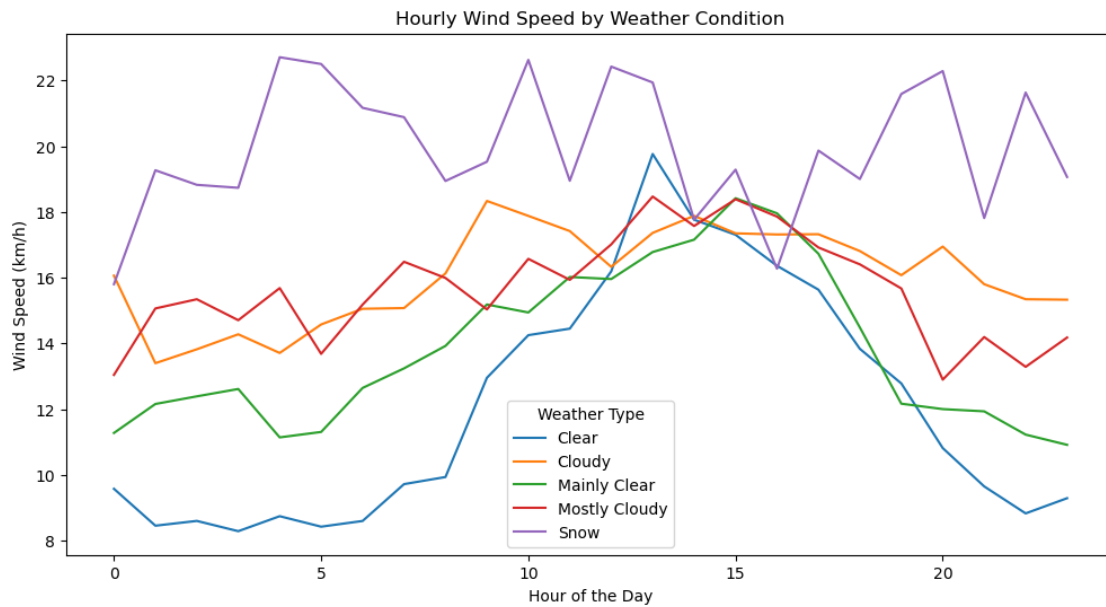
plt.title("Average Monthly Temperature by Top 5 Weather Conditions")
plt.xlabel("Month")
plt.ylabel("Temperature (°C)")
plt.legend(title="Weather Type")
plt.show()
```



2.10.1 Interpretation:

- Helps understand how weather change month to month

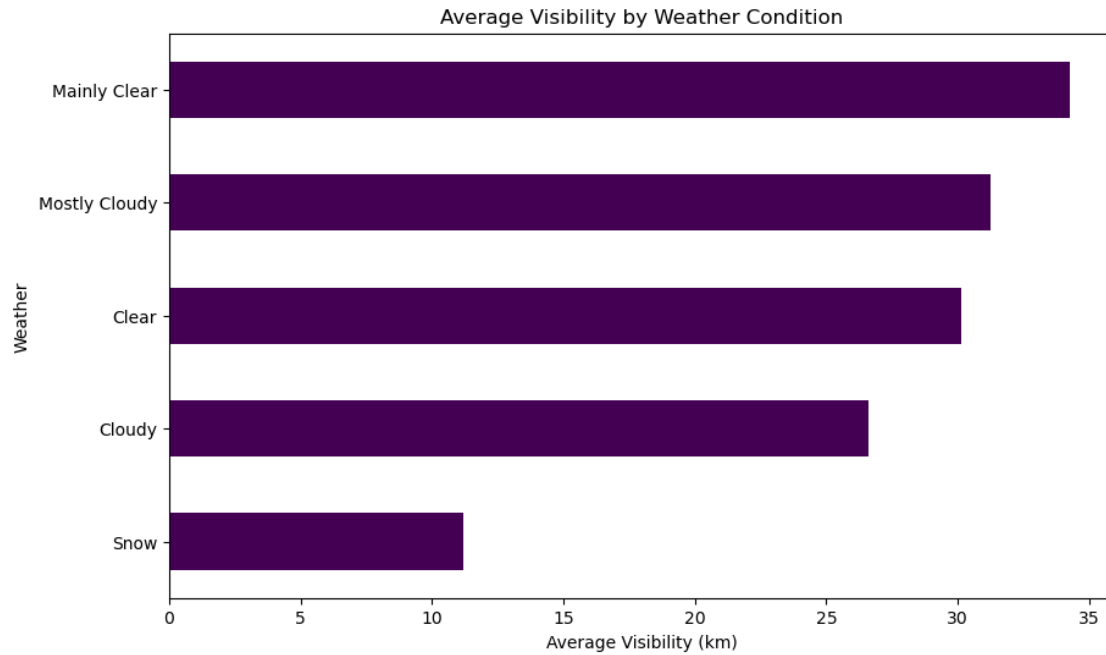
```
[28]: df_top5.groupby(['Hour', 'Weather'])['Wind Speed_kmh'].mean().unstack().  
      ↪ plot(figsize=(12,6))  
plt.title("Hourly Wind Speed by Weather Condition")  
plt.xlabel("Hour of the Day")  
plt.ylabel("Wind Speed (km/h)")  
plt.legend(title="Weather Type")  
plt.show()
```



2.10.2 Inference:

- snowy weather has the highest windspeed

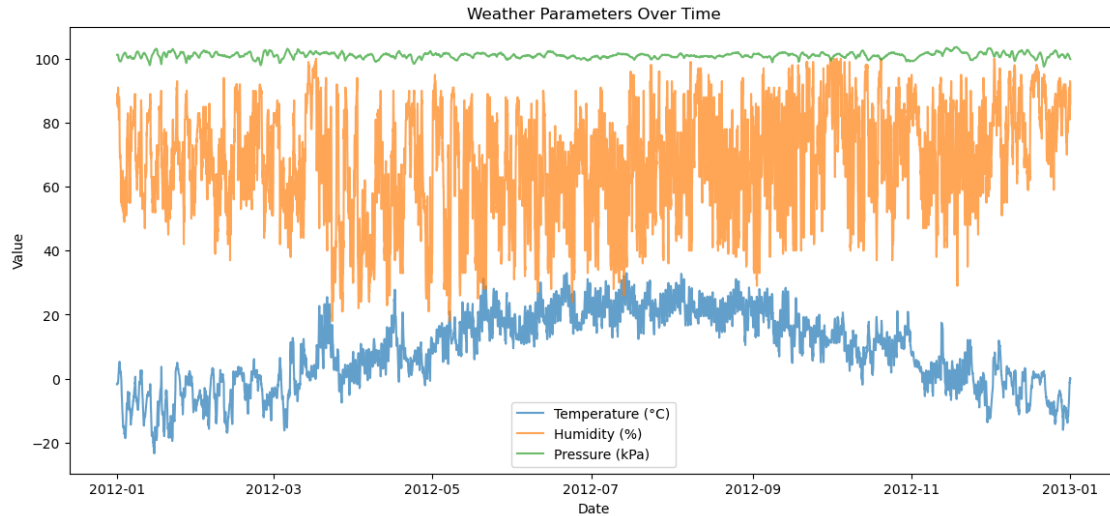
```
[42]: df_top5.groupby('Weather')['Visibility_km'].mean().sort_values().  
      ↪ plot(kind='barh', figsize=(10,6), colormap='viridis')  
plt.title("Average Visibility by Weather Condition")  
plt.xlabel("Average Visibility (km)")  
plt.show()
```



2.10.3 Inference:

- snowy weather has the lowest visibility

```
[32]: plt.figure(figsize=(14,6))
plt.plot(df['Date/Time'], df['Temp_C'], label='Temperature (°C)', alpha=0.7)
plt.plot(df['Date/Time'], df['Rel Hum_%'], label='Humidity (%)', alpha=0.7)
plt.plot(df['Date/Time'], df['Press_kPa'], label='Pressure (kPa)', alpha=0.7)
plt.legend()
plt.xlabel("Date")
plt.ylabel("Value")
plt.title("Weather Parameters Over Time")
plt.show()
```

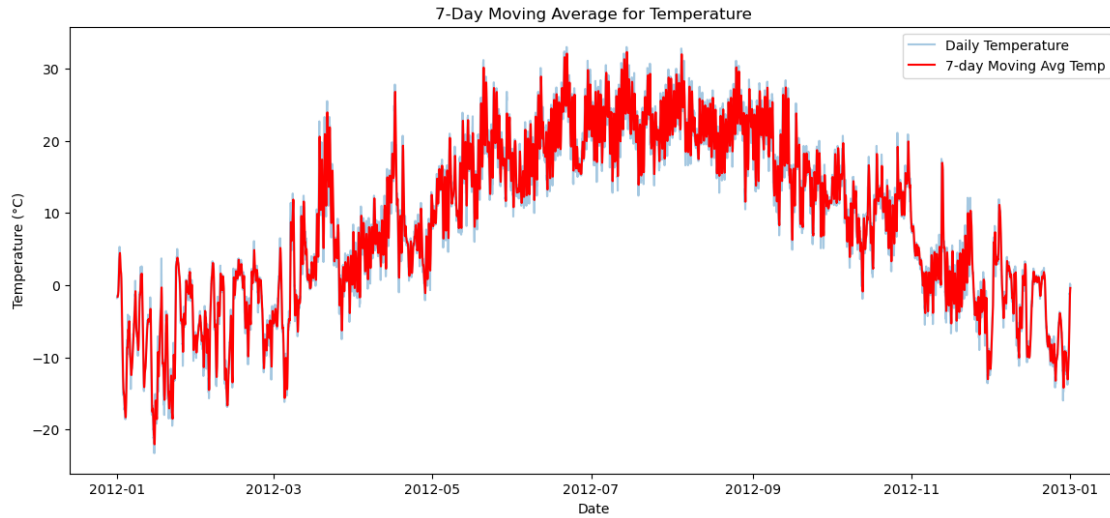



2.10.4 Inference:

- shows how weather parameters have changed over time
- helps identifying seasonal patterns

```
[34]: df['Temp_MA7'] = df['Temp_C'].rolling(window=7).mean()

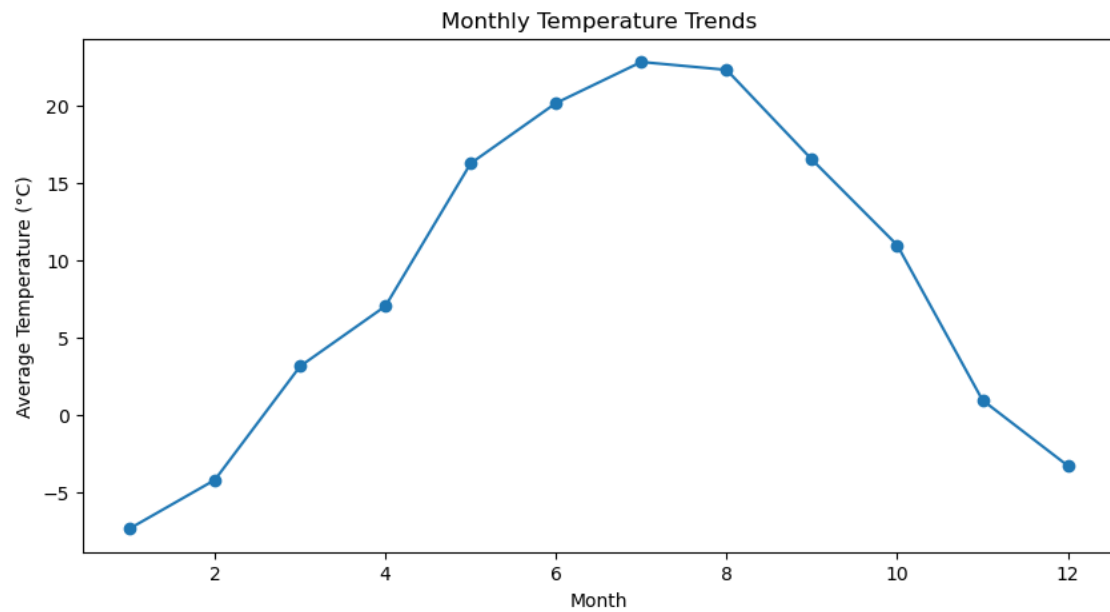
plt.figure(figsize=(14,6))
plt.plot(df['Date/Time'], df['Temp_C'], alpha=0.4, label="Daily Temperature")
plt.plot(df['Date/Time'], df['Temp_MA7'], label="7-day Moving Avg Temp",
         color='red')
plt.legend()
plt.xlabel("Date")
plt.ylabel("Temperature (°C)")
plt.title("7-Day Moving Average for Temperature")
plt.show()
```



2.10.5 Interpretation:

- Helps smooth out short-term fluctuations and highlight long-term trends.

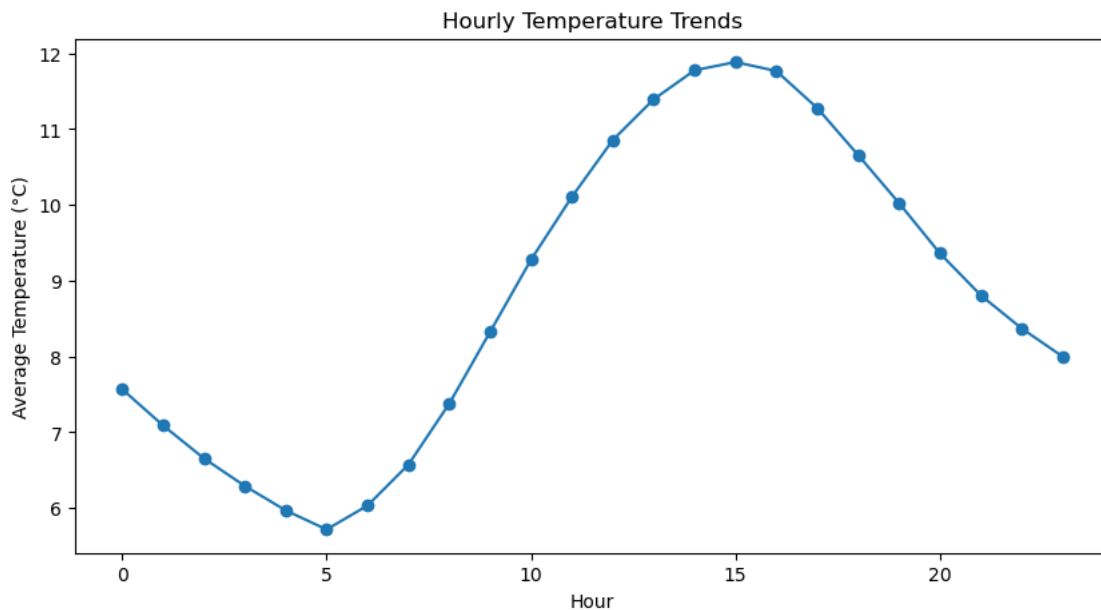
```
[36]: df.groupby('Month')['Temp_C'].mean().plot(kind='line', marker='o',
        figsize=(10,5))
plt.xlabel("Month")
plt.ylabel("Average Temperature (°C)")
plt.title("Monthly Temperature Trends")
plt.show()
```



2.10.6 Inference:

- In the month of June July August the temperature reaches maximum

```
[37]: df.groupby('Hour')['Temp_C'].mean().plot(kind='line', marker='o',  
      ↪figsize=(10,5))  
plt.xlabel("Hour")  
plt.ylabel("Average Temperature (°C)")  
plt.title("Hourly Temperature Trends")  
plt.show()
```



2.10.7 Inference:

- Shows how the average temperature changes in a day

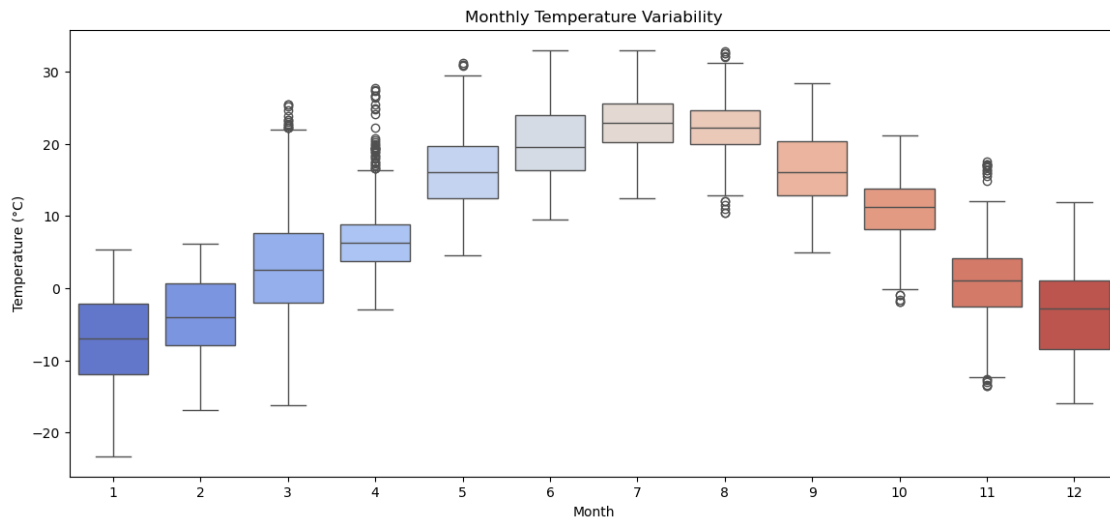
```
[39]: plt.figure(figsize=(14,6))  
sns.boxplot(x='Month', y='Temp_C', data=df, palette="coolwarm")  
plt.xlabel("Month")  
plt.ylabel("Temperature (°C)")  
plt.title("Monthly Temperature Variability")  
plt.show()
```

C:\Users\TUFAN\AppData\Local\Temp\ipykernel_6572\165339793.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same

effect.

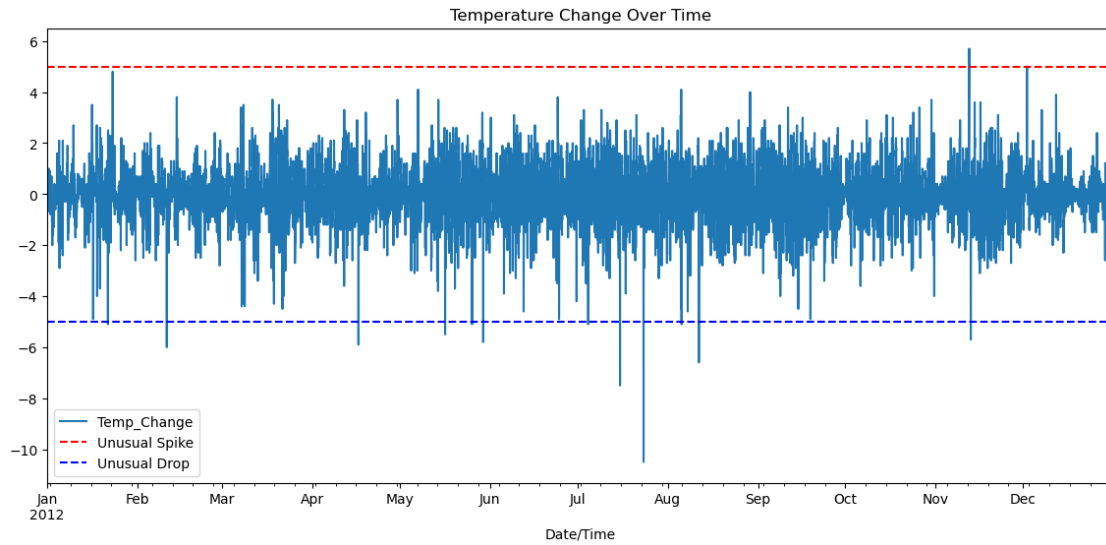
```
sns.boxplot(x='Month', y='Temp_C', data=df, palette="coolwarm")
```



2.10.8 Inference:

- Shows temperature variations within each month

```
[ ]: df['Temp_Change'] = df['Temp_C'].diff() # Calculate the change in temperature_
      ↪ from the previous row
df['Temp_Change'].plot(figsize=(14,6), title="Temperature Change Over Time")
plt.axhline(y=5, color='r', linestyle='--', label='Unusual Spike')
plt.axhline(y=-5, color='b', linestyle='--', label='Unusual Drop')
plt.legend()
plt.show()
```



2.10.9 Inference:

- In the month of July and August there is huge temperature fluctuations