PROJECT :- 3
Weather Data Analysis

DECLARATION:-

I declare that this Weather Data Analysis is my original work and has been completed to the best of my abilities. All data used in this analysis has been sourced from reliable and authorized channels, and any external references or sources have been duly cited. The findings, visualizations, and recommendations presented in this report are based on the data analysis conducted and are aimed at providing actionable insights into weather patterns, trends, and extreme events.

I affirm that the techniques and methodologies applied in this analysis adhere to standard data analysis practices and ethical guidelines. This analysis has been conducted with integrity and objectivity, ensuring that the conclusions drawn are unbiased and supported by the data.

Any resemblance to existing works is purely coincidental, and any errors or inaccuracies in the analysis are unintentional. I take full responsibility for the content and outcomes of this analysis and am committed to making any necessary corrections if discrepancies are identified.

By submitting this analysis, I agree to abide by the principles of academic and professional integrity, and I understand the importance of maintaining high standards of honesty and transparency in all analytical endeavors.

PROBLEM STATEMENT:-

Weather Data Analysis

Use a dataset containing historical weather data, including temperature, precipitation, and wind speed.

Analyze seasonal weather patterns and trends over time.

Identify correlations between weather variables (e.g., temperature and precipitation). Visualize weather data using line graphs, heatmaps, or box plots.

Extract insights about climate trends or extreme weather events from the analysis.

Introduction to Weather Data Analysis

Overview

Weather data analysis is a crucial aspect of meteorology, environmental science, and various other fields that rely on accurate weather predictions and climate trend analysis. By studying historical weather data, which typically includes variables such as temperature, precipitation, wind speed, humidity, and pressure, we can gain valuable insights into seasonal patterns, long-term climate trends, and extreme weather events. These insights are essential for preparing for future weather conditions, understanding the impacts of climate change, and making informed decisions in agriculture, urban planning, disaster management, and other sectors.

Importance of Weather Data Analysis

The analysis of weather data serves several critical functions:

1. Predicting Future Weather: By analyzing historical data, meteorologists can develop models to predict future weather conditions. This predictive capability is vital for various sectors, including agriculture, transportation, and emergency services.

2. Understanding Climate Trends: Long-term weather data allows scientists to identify and analyze climate trends. This understanding helps in addressing global challenges such as climate change, global warming, and the increasing frequency of extreme weather events.

3. Disaster Preparedness and Management: Accurate weather data analysis can predict natural disasters such as hurricanes, floods, and droughts. Early warnings enable authorities to prepare and respond more effectively, potentially saving lives and reducing economic losses.

4. Optimizing Agriculture: Farmers rely on weather forecasts to make crucial decisions about planting, irrigation, and harvesting. Analyzing weather data helps in developing more resilient agricultural practices and mitigating the impacts of adverse weather conditions on crop yields.

5. Urban Planning and Infrastructure Development: Understanding weather patterns is essential for designing resilient infrastructure and urban landscapes. Weather data analysis informs the construction of buildings, roads, and drainage systems that can withstand extreme weather events.

Key Components of Weather Data Analysis

1. Data Collection and Preparation: The first step involves gathering historical weather data from reliable sources such as meteorological stations, satellites, and weather databases. Data preparation includes cleaning, formatting, and handling missing values to ensure the dataset is ready for analysis.

2. Descriptive Statistics: Calculating basic statistics (mean, median, standard deviation) provides a summary of the weather data, helping to understand its overall distribution and variability.

3. Data Visualization: Visual tools like line graphs, heatmaps, and box plots are used to represent weather data graphically. These visualizations make it easier to identify patterns, trends, and anomalies.

4. Trend Analysis: This involves examining weather data over time to identify seasonal patterns, long-term trends, and changes in weather variables. Trend analysis is crucial for understanding the effects of climate change and predicting future conditions.

5. Correlation Analysis: Investigating the relationships between different weather variables (e.g., temperature and precipitation) can reveal how these factors interact and influence each other. Correlation analysis helps in identifying key drivers of weather patterns.

6. Insights and Conclusions: Based on the analysis, we can extract valuable insights about climate trends, seasonal variations, and extreme weather events. These insights are critical for developing strategies to mitigate the impacts of adverse weather conditions.

Applications of Weather Data Analysis

1. Agriculture: Farmers can optimize planting schedules, irrigation plans, and pest control measures based on weather forecasts and trends. This optimization helps in maximizing crop yields and minimizing losses due to adverse weather conditions.

2. Energy Sector: Weather data analysis is used to predict energy demand, particularly in heating and cooling systems. It also helps in optimizing the operation of renewable energy sources like solar and wind power.

3. Transportation: Weather forecasts are essential for ensuring the safety and efficiency of transportation systems. Airlines, shipping companies, and road transport services use weather data to plan routes and schedules.

4. Public Health: Understanding weather patterns can help predict and manage health risks related to

extreme temperatures, air quality, and the spread of vector-borne diseases.

5. Insurance: Insurance companies use weather data analysis to assess risk and set premiums for policies related to natural disasters, crop insurance, and other weather-related claims.

Conclusion

Weather data analysis is a powerful tool that provides critical insights into our environment and helps us prepare for and adapt to changing weather conditions. By leveraging historical weather data, we can improve our understanding of climate trends, enhance predictive capabilities, and develop strategies to mitigate the impacts of extreme weather events. Whether for agriculture, urban planning, disaster management, or everyday decision-making, weather data analysis plays a vital role in shaping a resilient and informed society.

In the following sections, we will delve into the specifics of weather data analysis, including data cleaning, visualization, trend analysis, and the extraction of actionable insights. Through careful examination of historical weather data, we aim to uncover patterns and trends that can inform better decision-making and contribute to our understanding of climate dynamics.

DATASET DESCRIPTION

The weather dataset is a time-series data set with per hour information about the weather conditions at a particular locaton. It records the following information;

- **Temperature:** The measure of how hot or cold the atmosphere is at a specific time and location, typically recorded in degrees Celsius (°C).
- **Dew Point Temperature:** The temperature at which air becomes saturated with moisture and dew forms, indicating the amount of moisture in the air.
- **Relative Humidity:** The percentage of moisture in the air relative to the maximum amount of moisture the air can hold at that temperature, crucial for understanding precipitation likelihood.
- **Wind Speed:** The speed at which air is moving horizontally through the atmosphere, measured in kilometers per hour (km/h)
- **Visibility:** The distance one can clearly see, measured in kilometers (km) important for transportation safety.
- **Pressure:** The force exerted by the weight of the air above a specific point, measured in kPa, indicating weather patterns.
- **Conditions:** A qualitative description of the overall weather at a specific time, such as sunny, cloudy, or rainy, providing a quick summary of the weather.

IMPORTING RELEVANT MODULES

[1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/weather-data/Weather Data.csv
```

[2]:

```
warnings.filterwarnings('ignore')
```

LOADING THE DATASETS

[3]:

```
data = pd.read_csv('/kaggle/input/weather-data/Weather Data.csv')
```

[4]:

```
data.head()
```

[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
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

GETTING INFORMATION ON THE DATAFRAME

[5]:

```
data.shape
```

[5]: (8784, 8)

[6]:

```python
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
```

[7]:

```python
data.describe()
```

[7]:

	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

DATA CLEANING

[8]:

```
# Checking for null values in the dataset
data.isnull().sum()
```

[8]:

```
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
```

[9]:

```
# Spliting the date and time into seperate columns
data[['Date', 'Time']] = data['Date/Time'].str.split(' ', expand = True)
```

[10]:

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

[10]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	Date	Time
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog	1/1/2012	0:00
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog	1/1/2012	1:00

[11]:

```python
#Convert date column to date
data['Date'] = pd.to_datetime(data['Date'])
```

[12]:

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

[12]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	Date	Time
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog	2012-01-01	0:00
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog	2012-01-01	1:00

[13]:

```python
# Getting seasons and seperating the month into a new column
def get_season(month):
    if month >=3 and month <= 5:
        return 'Spring'
    elif month>= 6 and month <= 8:
        return 'Summer'
    elif month >= 9 and month <= 11:
        return 'Autumn'
    else:
        return 'Winter'

data['Month'] = data['Date'].dt.month
data['Season'] = data['Month'].apply(get_season)
```

14]:

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

14]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	Date	Time	Month	Season
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog	2012-01-01	0:00	1	Winter
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog	2012-01-01	1:00	1	Winter

16]:

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

16]:

	Date/Time	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa	Weather	Date	Time	Month	Season
0	1/1/2012 0:00	-1.8	-3.9	86	4	8.0	101.24	Fog	2012-01-01	0:00	January	Winter
1	1/1/2012 1:00	-1.8	-3.7	87	4	8.0	101.24	Fog	2012-01-01	1:00	January	Winter

EXPLORATORY DATA ANALYSIS

[21]:
```python
# Checking for duplicate values
data.duplicated().sum()
```

[21]: 0

[22]:
```python
#Checking all the unique values in my dataset
data.nunique()
```

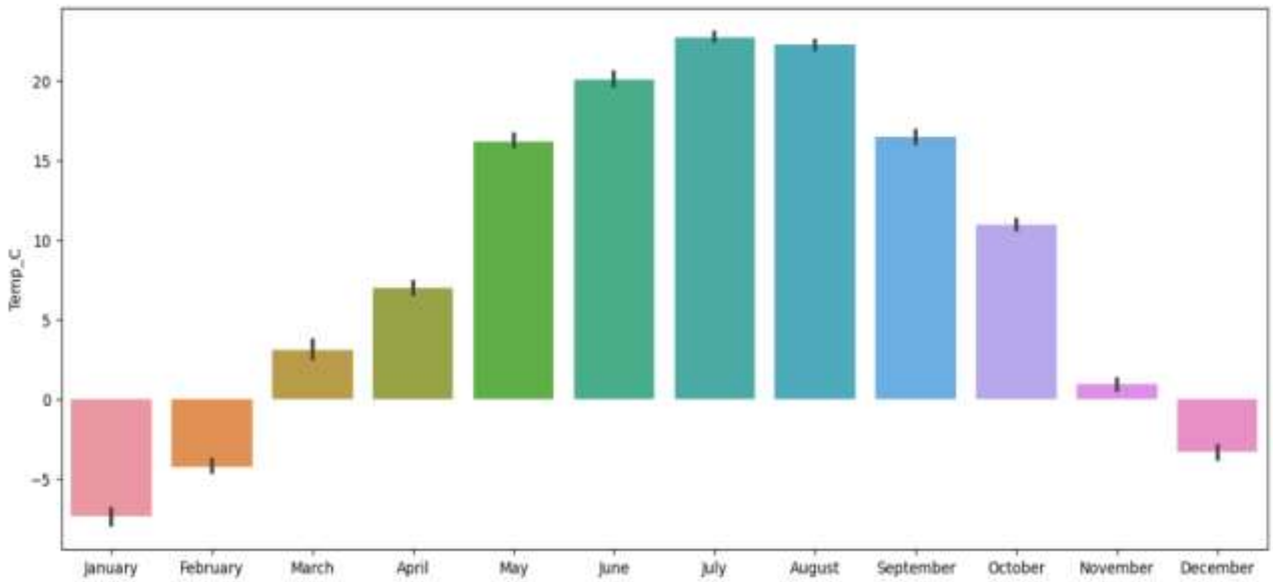
[22]:
```
Date                   366
Month                   12
Time                    24
Temp_C                 533
Dew Point Temp_C       489
Rel Hum_%               83
Wind Speed_km/h         34
Visibility_km           24
Press_kPa              518
Weather Condition       50
Season                   4
dtype: int64
```

Temperature Distribution by Month

[23]:

```
plt.figure(figsize=(15,6))
sns.barplot(x= 'Month', y= 'Temp_C', data= data)
```

[23]: <Axes: xlabel='Month', ylabel='Temp_C'>



[24]:

```
# Checking the frequency of each weather condition
data['Weather Condition'].value_counts()
```

```
[24]: Weather Condition
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
Name: count, dtype: int64
```

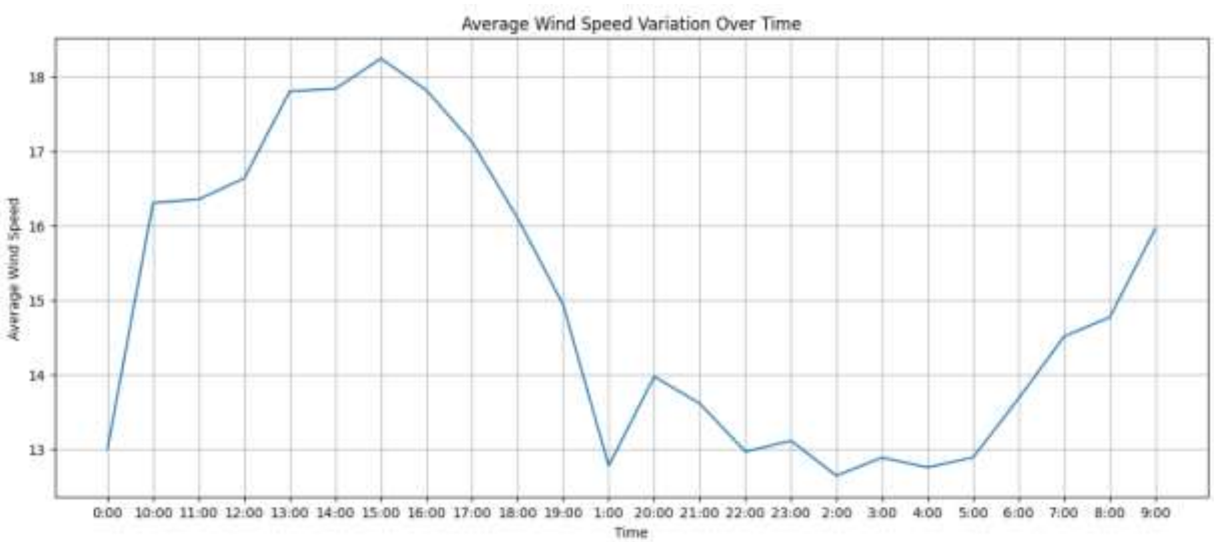
Relationship between Windspeed and Time

[27]:

```
# Checking if there's any relationship between windspeed and time

avg_windspeed = data.groupby('Time')['Wind Speed_km/h'].mean()
plt.figure(figsize=(15, 6))

plt.plot(avg_windspeed.index, avg_windspeed.values)
plt.xlabel('Time')
plt.ylabel('Average Wind Speed')
plt.title('Average Wind Speed Variation Over Time')
plt.grid()
```



[28]:

```
# What is the mean 'Visibility' in the dataset?

mean_visibility = data['Visibility_km'].mean().round(2)
print(f'The mean visibility of the data is {mean_visibility}km')
```

```
The mean visibility of the data is 27.66km
```

Correlation between Weather Variables

[37]:

```
correlation_matrix = data.corr(numeric_only= True)
```

[38]:

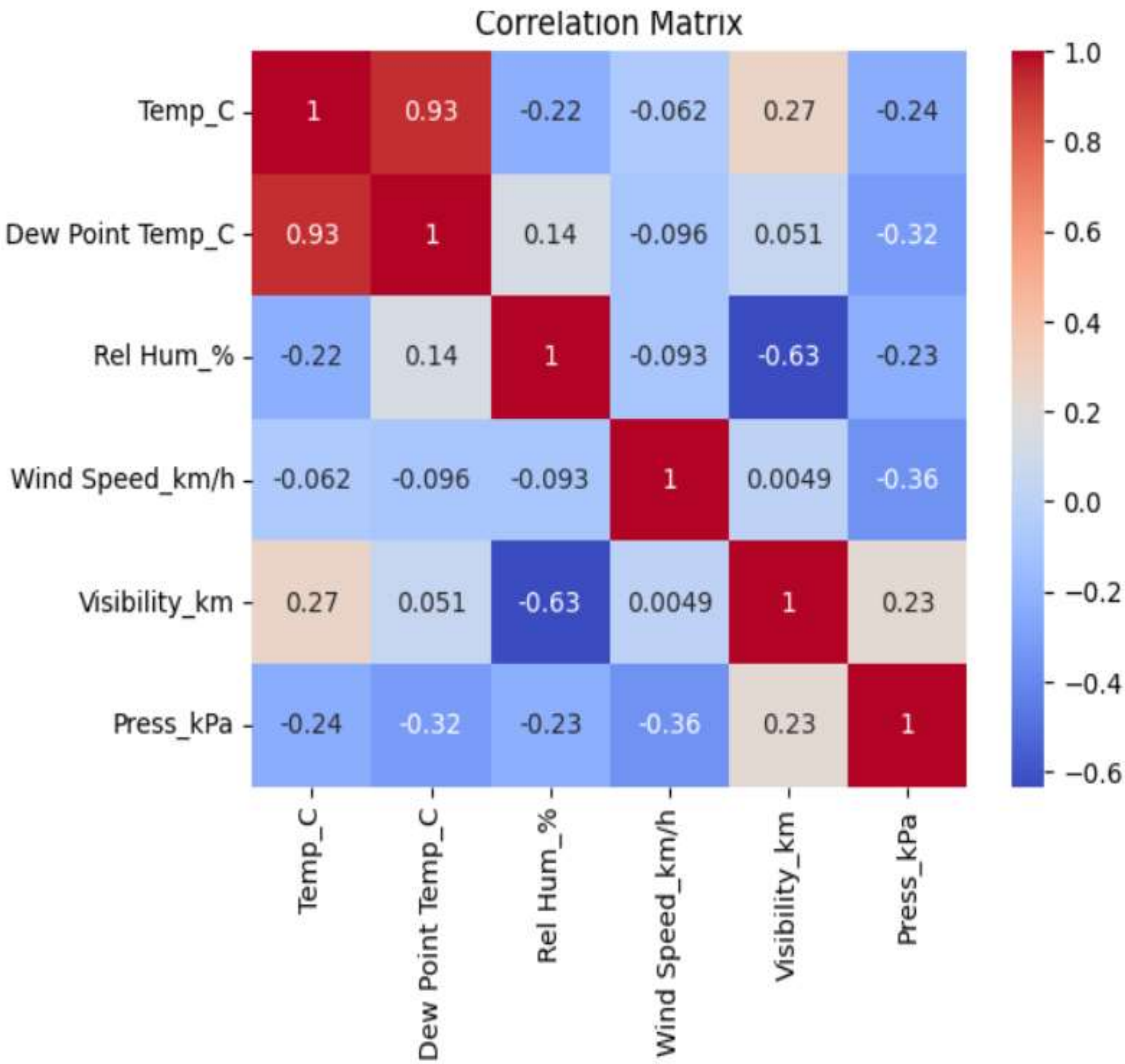
```
correlation_matrix.head()
```

[38]:

	Temp_C	Dew Point Temp_C	Rel Hum_%	Wind Speed_km/h	Visibility_km	Press_kPa
Temp_C	1.000000	0.932714	-0.220182	-0.061876	0.273455	-0.236389
Dew Point Temp_C	0.932714	1.000000	0.139494	-0.095685	0.050813	-0.320616
Rel Hum_%	-0.220182	0.139494	1.000000	-0.092743	-0.633683	-0.231424
Wind Speed_km/h	-0.061876	-0.095685	-0.092743	1.000000	0.004883	-0.356613
Visibility_km	0.273455	0.050813	-0.633683	0.004883	1.000000	0.231847

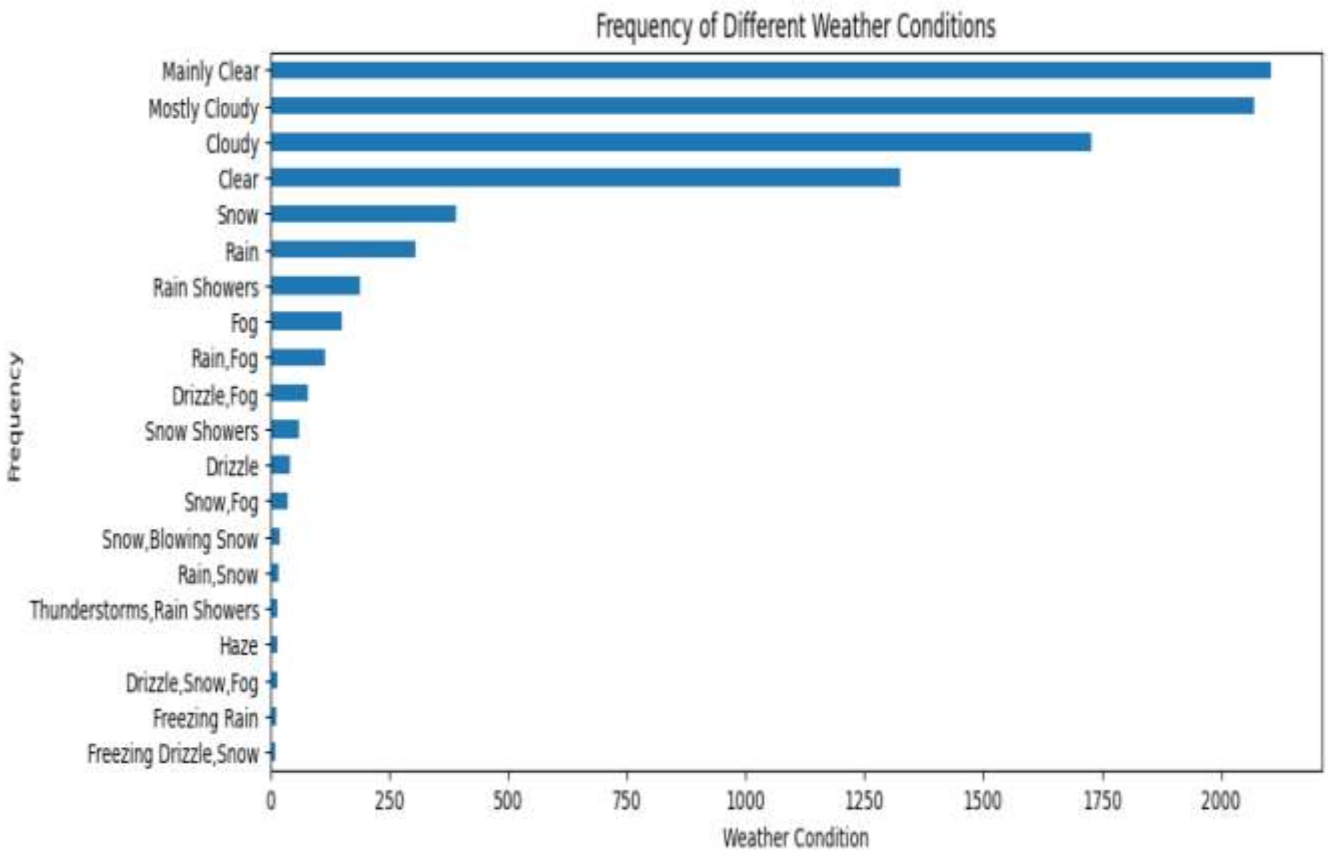
[39]:

```
sns.heatmap(correlation_matrix, annot = True, cmap = 'coolwarm')
plt.title('Correlation Matrix')
```



Most Frequent Weather Conditions

```
weather_counts = data['Weather Condition'].value_counts(ascending=True)[-20:]
weather_counts.plot(kind='barh', figsize=(12, 5))
plt.title('Frequency of Different Weather Conditions')
plt.xlabel('Weather Condition')
plt.ylabel('Frequency')
plt.show()
```

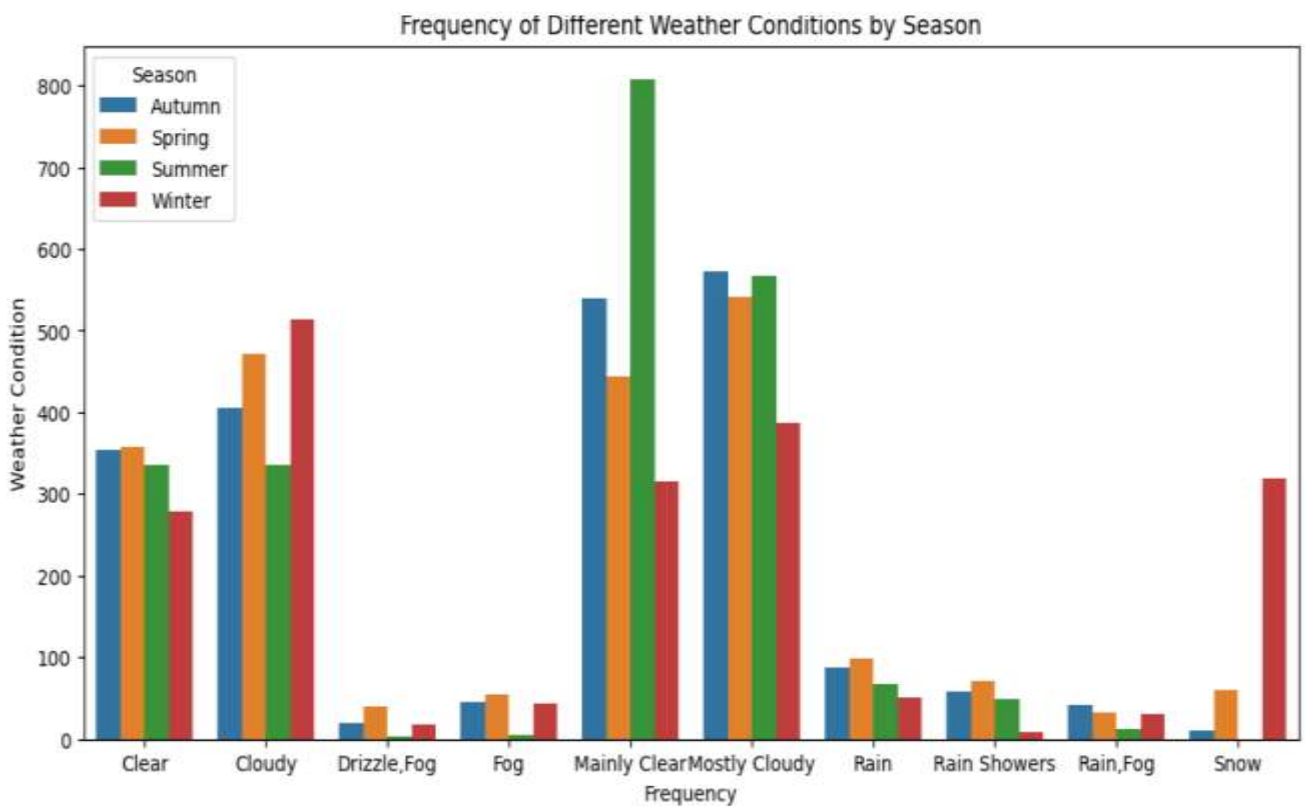


```python
# Grouping the data by 'Season' and 'Weather Condition' and count the occurrences
grouped_data = data.groupby(['Season', 'Weather Condition']).size().reset_index(name='Count')

# Getting the top 20 weather conditions across all seasons
top_conditions = grouped_data.groupby('Weather Condition')['Count'].sum().nlargest(10).index

# Filter the original grouped data to include only the top conditions
filtered_data = grouped_data[grouped_data['Weather Condition'].isin(top_conditions)]

# Plotting
plt.figure(figsize=(12, 6))
sns.barplot(data=filtered_data, x='Weather Condition', y='Count', hue='Season', ci=None)
plt.title('Frequency of Different Weather Conditions by Season')
plt.xlabel('Frequency')
plt.ylabel('Weather Condition')
plt.legend(title='Season')
plt.show()
```



[42]:

```python
data.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	64.497738	10.557315	30.153243	101.587443
Cloudy	7.970544	2.375810	69.592593	16.127315	26.625752	100.911441
Drizzle	7.353659	5.504878	88.243902	16.097561	17.931707	100.435366
Drizzle,Fog	8.067500	7.033750	93.275000	11.862500	5.257500	100.786625
Drizzle,Ice Pellets,Fog	0.400000	-0.700000	92.000000	20.000000	4.000000	100.790000
Drizzle,Snow	1.050000	0.150000	93.500000	14.000000	10.500000	100.890000
Drizzle,Snow,Fog	0.693333	0.120000	95.866667	15.533333	5.513333	99.281333
Fog	4.303333	3.159333	92.286667	7.946667	6.248000	101.184067
Freezing Drizzle	-5.657143	-8.000000	83.571429	16.571429	9.200000	100.202857
Freezing Drizzle,Fog	-2.533333	-4.183333	88.500000	17.000000	5.266667	100.441667
Freezing Drizzle,Haze	-5.433333	-8.000000	82.000000	10.333333	2.666667	100.316667
Freezing Drizzle,Snow	-5.109091	-7.072727	86.090909	16.272727	5.872727	100.520909
Freezing Fog	-7.575000	-9.250000	87.750000	4.750000	0.650000	102.320000
Freezing Rain	-3.885714	-6.078571	84.642857	19.214286	8.242857	99.647143
Freezing Rain,Fog	-2.225000	-3.750000	89.500000	15.500000	7.550000	99.945000
Freezing Rain,Haze	-4.900000	-7.450000	82.500000	7.500000	2.400000	100.375000
Freezing Rain,Ice Pellets,Fog	-2.600000	-3.700000	92.000000	28.000000	8.000000	100.950000
Freezing Rain,Snow Grains	-5.000000	-7.300000	84.000000	32.000000	4.800000	98.560000
Haze	-0.200000	-2.975000	81.625000	10.437500	7.831250	101.482500
Mainly Clear	12.558927	4.581671	60.667142	14.144824	34.264862	101.248832
Moderate Rain,Fog	1.700000	0.800000	94.000000	17.000000	6.400000	99.980000
Moderate Snow	-5.525000	-7.250000	87.750000	33.750000	0.750000	100.275000
Moderate Snow,Blowing Snow	-5.450000	-6.500000	92.500000	40.000000	0.600000	100.570000
Mostly Cloudy	10.574287	3.131174	62.102465	15.813920	31.253842	101.025288
Rain	9.786275	7.042810	83.624183	19.254902	18.856536	100.233333
Rain Showers	13.722340	9.187766	75.159574	17.132979	22.816489	100.404043
Rain Showers,Fog	12.800000	12.100000	96.000000	13.000000	6.400000	99.830000
Rain Showers,Snow Showers	2.150000	-1.500000	76.500000	22.500000	21.700000	101.100000
Rain,Fog	8.273276	7.219828	93.189655	14.793103	6.873276	100.500862
Rain,Haze	4.633333	2.066667	83.333333	11.666667	6.700000	100.540000