WEATHER PREDICTION

INTRODUCTION:

The `weather_prediction_dataset` comprises a comprehensive collection of weather-related data points essential for weather prediction tasks, featuring diverse atmospheric parameters such as temperature, humidity, wind speed, and cloud cover. This dataset serves as a valuable resource for both short-term forecasting and long-term climate analysis, providing insights into various weather patterns. Complementing this dataset, the `weather_prediction_bbq_labels` dataset offers binary labels indicating the suitability of weather conditions for outdoor barbecues. By leveraging the weather features from the `weather_prediction_dataset`, this dataset categorizes each data point as either conducive or non-conducive for barbecuing, empowering individuals to make informed decisions when planning outdoor activities. Together, these datasets facilitate a deeper understanding of weather dynamics and enable users to optimize their experiences based on weather forecasts.

PROBLEM STATEMENT:

The `weather_prediction_dataset` and `weather_prediction_bbq_labels` datasets address the need for accurate weather prediction and its impact on outdoor activities, particularly barbecuing. Weather conditions play a significant role in daily planning, outdoor events, and various industries such as agriculture, tourism, and transportation. However, predicting weather accurately remains a challenging task due to the complexity and variability of atmospheric phenomena. Additionally, individuals often face uncertainties when planning outdoor activities like barbecues, as weather conditions greatly influence the feasibility and enjoyment of such events. Therefore, the primary objective of this project is to develop predictive models using machine learning techniques to forecast weather conditions and determine the suitability of weather for barbecuing. By analyzing historical weather data and corresponding barbecue labels, we aim to build robust models capable of accurately predicting weather patterns and identifying optimal conditions for outdoor barbecues. This project aims to provide valuable insights for weather forecasting applications and empower individuals to make informed decisions when planning outdoor activities, ultimately enhancing their overall experience and well-being.

DATASET DESCRIPTION:

1.weather_prediction_dataset:

Based on the provided columns, the features for the `weather_prediction_dataset` include:

1. `DATE`: Date of the weather observation.
2. `MONTH`: Month of the weather observation.
3. `BASEL_cloud_cover`: Cloud cover in Basel.
4. `BASEL_humidity`: Humidity in Basel.
5. `BASEL_pressure`: Atmospheric pressure in Basel.
6. `BASEL_global_radiation`: Global radiation in Basel.
7. `BASEL_precipitation`: Precipitation in Basel.
8. `BASEL_sunshine`: Sunshine duration in Basel.
9. `BASEL_temp_mean`: Mean temperature in Basel.
10. `BASEL_temp_min`: Minimum temperature in Basel.
11. `BASEL_temp_max`: Maximum temperature in Basel.
12. `STOCKHOLM_wind_speed`: Wind speed in Stockholm.
13. `STOCKHOLM_humidity`: Humidity in Stockholm.
14. `STOCKHOLM_pressure`: Atmospheric pressure in Stockholm.
15. `STOCKHOLM_global_radiation`: Global radiation in Stockholm.
16. `STOCKHOLM_precipitation`: Precipitation in Stockholm.
17. `STOCKHOLM_temp_mean`: Mean temperature in Stockholm.
18. `STOCKHOLM_temp_min`: Minimum temperature in Stockholm.
19. `STOCKHOLM_temp_max`: Maximum temperature in Stockholm.
20. `TOURS_wind_speed`: Wind speed in Tours.
21. `TOURS_humidity`: Humidity in Tours.
22. `TOURS_pressure`: Atmospheric pressure in Tours.
23. `TOURS_global_radiation`: Global radiation in Tours.
24. `TOURS_precipitation`: Precipitation in Tours.
25. `TOURS_temp_mean`: Mean temperature in Tours.

26. `TOURS_temp_min`: Minimum temperature in Tours.

27. `TOURS_temp_max`: Maximum temperature in Tours.

2.weather_prediction_bbq_labels:

Based on the provided columns, the features for the `weather_prediction_bbq_labels` dataset include:

1. `DATE`: Date of the weather observation.
2. `BASEL_BBQ_weather`: BBQ suitability label for Basel.
3. `BUDAPEST_BBQ_weather`: BBQ suitability label for Budapest.
4. `DE_BBQ_weather`: BBQ suitability label for DE (assuming it represents Germany).
5. `DRESDEN_BBQ_weather`: BBQ suitability label for Dresden.
6. `DUSSELDORF_BBQ_weather`: BBQ suitability label for Dusseldorf.
7. `HEATHROW_BBQ_weather`: BBQ suitability label for Heathrow.
8. `KASSEL_BBQ_weather`: BBQ suitability label for Kassel.
9. `LJUBLJANA_BBQ_weather`: BBQ suitability label for Ljubljana.
10. `MAASTRICHT_BBQ_weather`: BBQ suitability label for Maastricht.
11. `MALMO_BBQ_weather`: BBQ suitability label for Malmo.
12. `MONTELIMAR_BBQ_weather`: BBQ suitability label for Montelimar.
13. `MUENCHEN_BBQ_weather`: BBQ suitability label for Munich.
14. `OSLO_BBQ_weather`: BBQ suitability label for Oslo.
15. `PERPIGNAN_BBQ_weather`: BBQ suitability label for Perpignan.
16. `SONNBLICK_BBQ_weather`: BBQ suitability label for Sonnblick.
17. `STOCKHOLM_BBQ_weather`: BBQ suitability label for Stockholm.
18. `TOURS_BBQ_weather`: BBQ suitability label for Tours.

DATA EXPLORATION:

Data exploration is an essential step in understanding the characteristics of your datasets and gaining insights into the underlying patterns and relationships. Here's a guide on how you can perform data exploration for the `weather_prediction_dataset` and `weather_prediction_bbq_labels` datasets:

Data Exploration for weather_prediction_dataset:

1. Summary Statistics:

   - Compute summary statistics such as mean, median, standard deviation, minimum, and maximum values for numerical features like temperature, humidity, wind speed, etc.

   - Identify any outliers or anomalies in the data.

2. Distribution Analysis:

   - Visualize the distributions of numerical features using histograms or kernel density plots to understand their spread and skewness.

   - For categorical features like month, create bar plots to visualize the frequency distribution of each category.

3. Correlation Analysis:

   - Compute the correlation matrix between numerical features to identify pairwise relationships.

   - Visualize the correlation matrix using a heatmap to identify strong correlations (positive or negative) between features.

4.Time Series Analysis:

   - If applicable (e.g., temperature over time), plot time series graphs to observe temporal trends and seasonality.

Data Exploration for weather_prediction_bbq_labels:

1. Class Distribution:

   - Determine the distribution of BBQ suitability labels (e.g., number of suitable vs. unsuitable days for barbecuing).

   - Visualize the class distribution using bar plots or pie charts.

2. Relationship with Weather Features:

- Explore how BBQ suitability labels vary with different weather features (e.g., temperature, precipitation, sunshine duration).

- Plot scatter plots or box plots to visualize the relationship between BBQ suitability and weather features.

3. Temporal Analysis:

- Analyze the temporal distribution of BBQ suitability labels over time (e.g., monthly or seasonal variations).

- Plot time series graphs or line plots to observe trends in BBQ suitability.

4. Comparison Across Locations:

- If applicable, compare BBQ suitability labels across different locations to identify regional variations.

- Plot side-by-side bar plots or box plots to compare BBQ suitability across locations.

Tools:

- Python libraries such as Pandas, Matplotlib, Seaborn, and Plotly can be used for data manipulation and visualization.

- Jupyter Notebooks and Google colab provide an interactive environment for conducting data exploration and documenting your analysis.

CODE AND IMPLEMENTATION:

Preprocessing:

1.Import the dataset(weather_prediction_dataset):

```
import pandas as pd

weather_data = pd.read_csv('weather_prediction_dataset.csv')

print(weather_data)
```

2. Import the dataset (weather_prediction_bbq_labels):

```
import pandas as pd

bbq_labels = pd.read_csv('weather_prediction_bbq_labels.csv')

df = bbq_labels.head()
print(df)
```

OUTPUT:

(weather_prediction_dataset)

```
          DATE  MONTH  BASEL_cloud_cover  BASEL_humidity  BASEL_pressure  \
0     20000101      1                  8            0.89          1.0286
1     20000102      1                  8            0.87          1.0318
2     20000103      1                  5            0.81          1.0314
3     20000104      1                  7            0.79          1.0262
4     20000105      1                  5            0.90          1.0246
...        ...    ...                ...             ...             ...
3649  20091228     12                  7            0.82          1.0084
3650  20091229     12                  7            0.92          1.0028
3651  20091230     12                  8            0.92          0.9979
3652  20091231     12                  7            0.93          0.9958
3653  20100101      1                  8            0.93          0.9965

      BASEL_global_radiation  BASEL_precipitation  BASEL_sunshine  \
0                       0.20                 0.03             0.0
1                       0.25                 0.00             0.0
2                       0.50                 0.00             3.7
3                       0.63                 0.35             6.9
4                       0.51                 0.07             3.7
...                      ...                  ...             ...
3649                    0.28                 0.42             0.3
3650                    0.22                 1.68             0.2
3651                    0.07                 1.54             0.0
3652                    0.17                 0.57             0.1
3653                    0.08                 0.56             0.0

      BASEL_temp_mean  BASEL_temp_min  ...  STOCKHOLM_temp_min  \
0                 2.9             1.6  ...                -9.3
1                 3.6             2.7  ...                 0.5
2                 2.2             0.1  ...                -1.0
3                 3.9             0.5  ...                 2.5
4                 6.0             3.8  ...                -1.8
...               ...             ...  ...                 ...
3649              3.2             1.0  ...                -2.7
3650              4.5             2.4  ...                -9.5
3651              8.5             7.5  ...               -12.5
3652              6.6             4.3  ...                -9.3
3653              2.9            -0.2  ...                -8.8
```

Weather_prediction_bbq_labels

```
       DATE  BASEL_BBQ_weather  BUDAPEST_BBQ_weather  DE_BBQ_weather  \
0  20000101              False                 False           False
1  20000102              False                 False           False
2  20000103              False                 False           False
3  20000104              False                 False           False
4  20000105              False                 False           False

   DRESDEN_BBQ_weather  DUSSELDORF_BBQ_weather  HEATHROW_BBQ_weather  \
0                False                   False                 False
1                False                   False                 False
2                False                   False                 False
3                False                   False                 False
4                False                   False                 False

   KASSEL_BBQ_weather  LJUBLJANA_BBQ_weather  MAASTRICHT_BBQ_weather  \
0               False                  False                   False
1               False                  False                   False
2               False                  False                   False
3               False                  False                   False
4               False                  False                   False

   MALMO_BBQ_weather  MONTELIMAR_BBQ_weather  MUENCHEN_BBQ_weather  \
0              False                   False                 False
1              False                   False                 False
2              False                   False                 False
3              False                   False                 False
4              False                   False                 False
```

3.Synthetic weather data

```
import numpy as np
import pandas as pd
np.random.seed(0)
num_samples = 1000
temperature = np.random.uniform(low=0, high=40, size=num_samples)
humidity = np.random.uniform(low=0, high=100, size=num_samples)
wind_speed = np.random.uniform(low=0, high=30, size=num_samples)
rainfall = temperature * 0.5 + humidity * 0.3 + np.random.normal(loc=0,
scale=5, size=num_samples)
weather_data = pd.DataFrame({
    'Temperature': temperature,
    'Humidity': humidity,
    'Wind_Speed': wind_speed,
    'Rainfall': rainfall
})
weather_data.to_csv('weather_prediction_dataset.csv', index=False)
weather_data.head()
```

OUTPUT:

	Temperature	Humidity	Wind_Speed	Rainfall
0	21.952540	59.288027	24.345554	37.764992
1	28.607575	1.006370	14.282520	8.630179
2	24.110535	47.582620	15.694680	28.912591
3	21.795327	70.877039	7.515618	34.207250
4	16.946192	4.397543	18.151291	7.696949

4.Merge dataset on “DATE”

```
merged_data = pd.merge(weather_data, bbq_labels, on='DATE')

merged_data .head()
```

OUTPUT:

	DATE	BASEL_BBQ_weather_x	BUDAPEST_BBQ_weather_x	DE_BBQ_weather_x	DRESDEN_BBQ_weather_x	DUSSELDORF_BBQ_weather_x
0	20000101	False	False	False	False	False
1	20000102	False	False	False	False	False
2	20000103	False	False	False	False	False
3	20000104	False	False	False	False	False
4	20000105	False	False	False	False	False

rows × 35 columns

5.Check for missing values

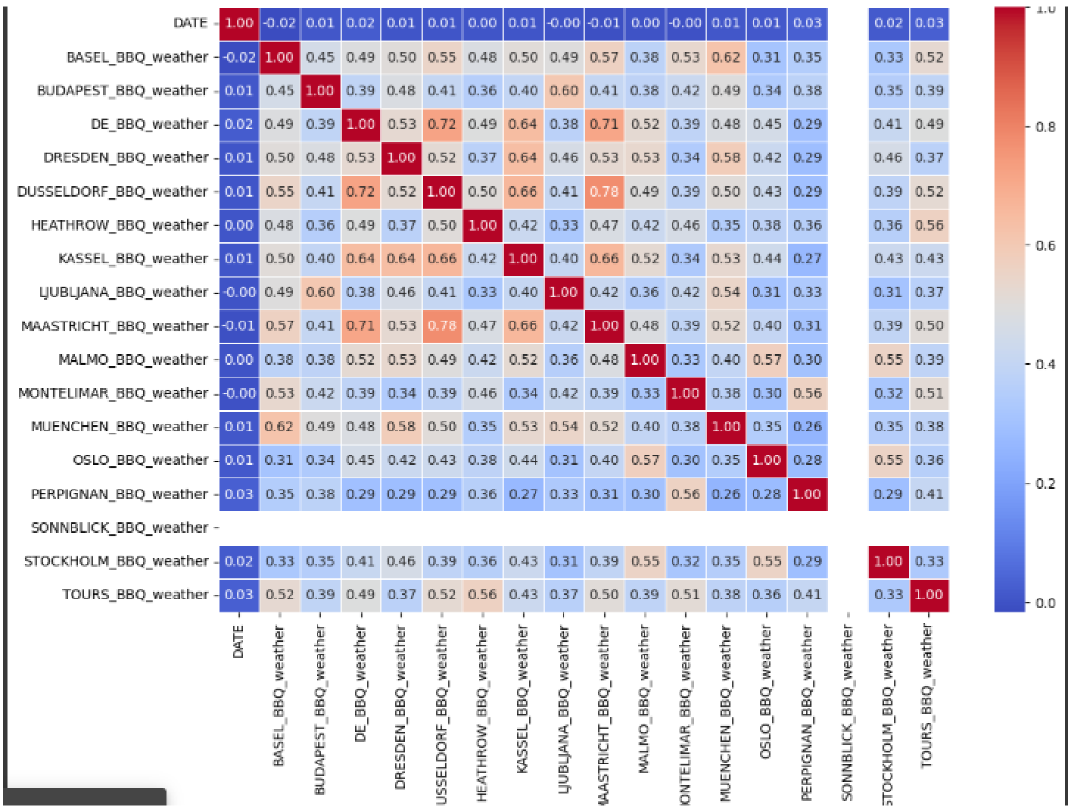
```
missing_counts = merged_data.isna().sum().sum()

missing_counts
```

OUTPUT:

6. Correlation heap map

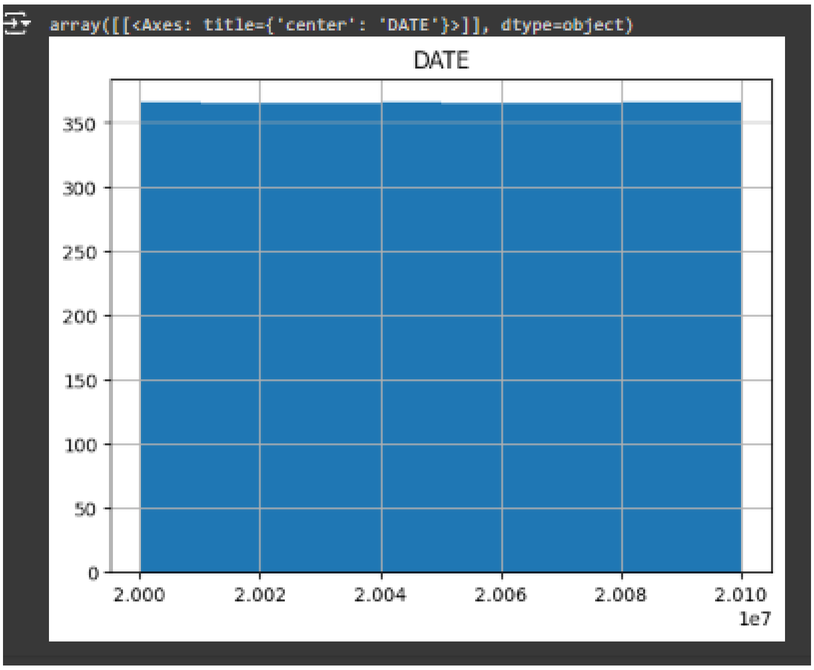
```
import seaborn as sns

import matplotlib.pyplot as plt


bbq_labels = pd.read_csv('weather_prediction_bbq_labels.csv')
correlation_matrix = bbq_labels.corr()
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f",
linewidths=.5)
plt.title('Correlation Heatmap')
plt.show()
```

OUTPUT:



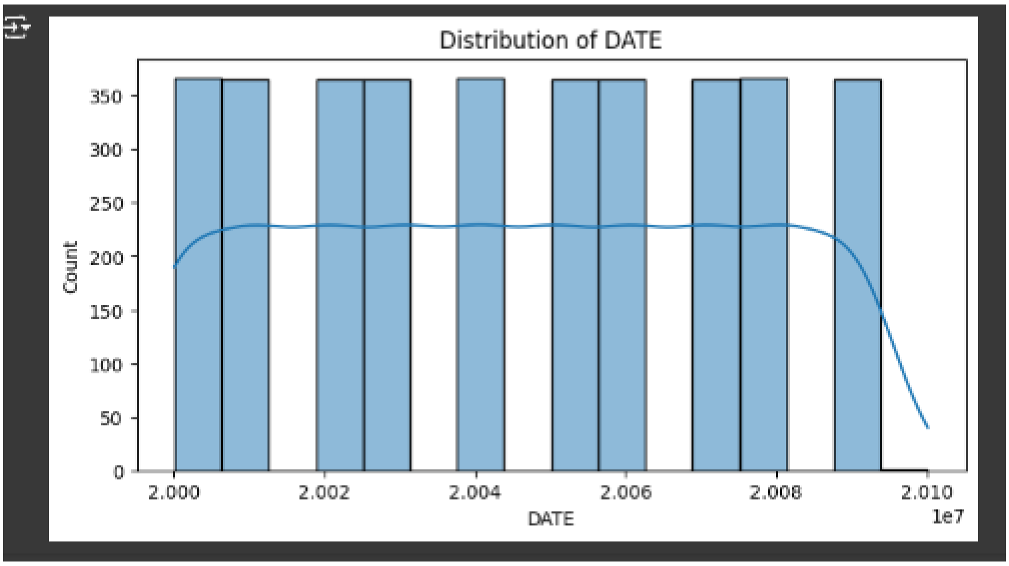
```
bbq_labels.hist()
```

OUTPUT:



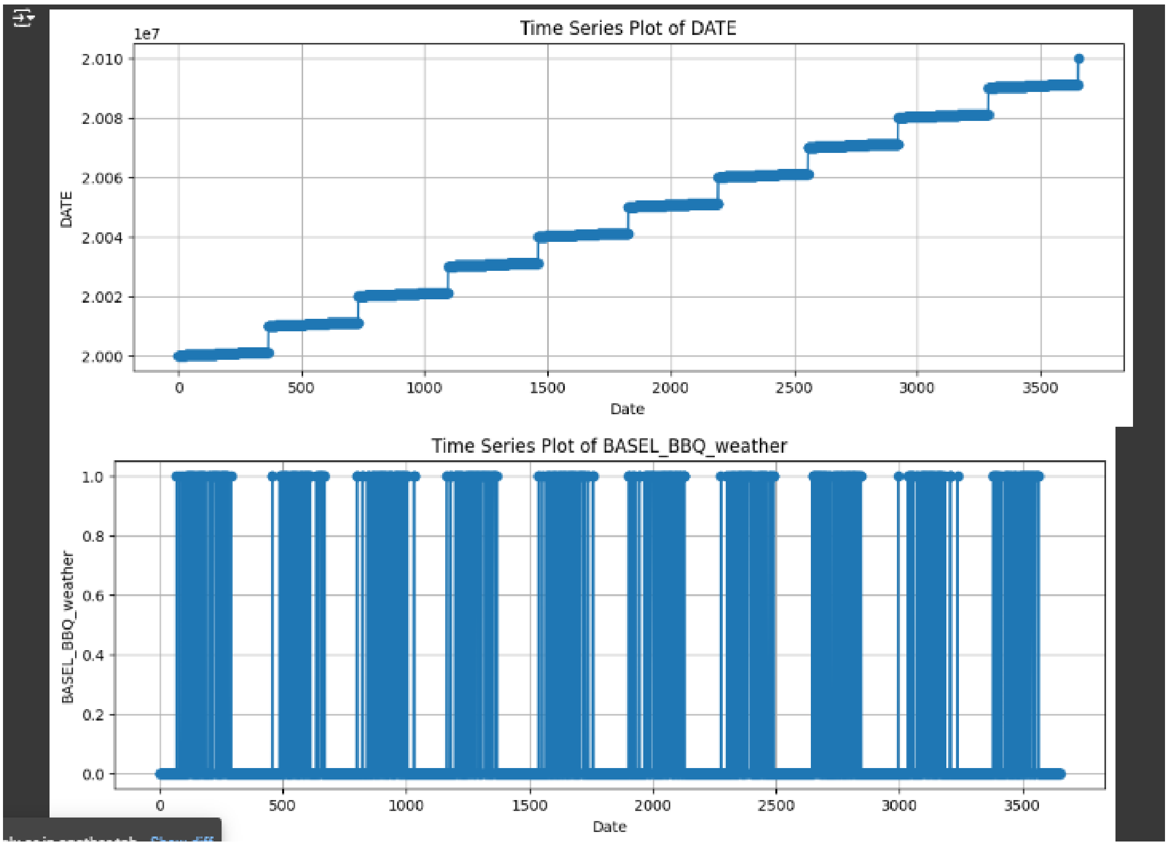

```
import matplotlib.pyplot as plt
for column in bbq_labels.select_dtypes(include=['float64',
'int64']).columns:
    plt.figure(figsize=(8, 4))
    sns.histplot(bbq_labels[column].dropna(), kde=True)
    plt.title(f'Distribution of {column}')
    plt.show()
```

OUTPUT:

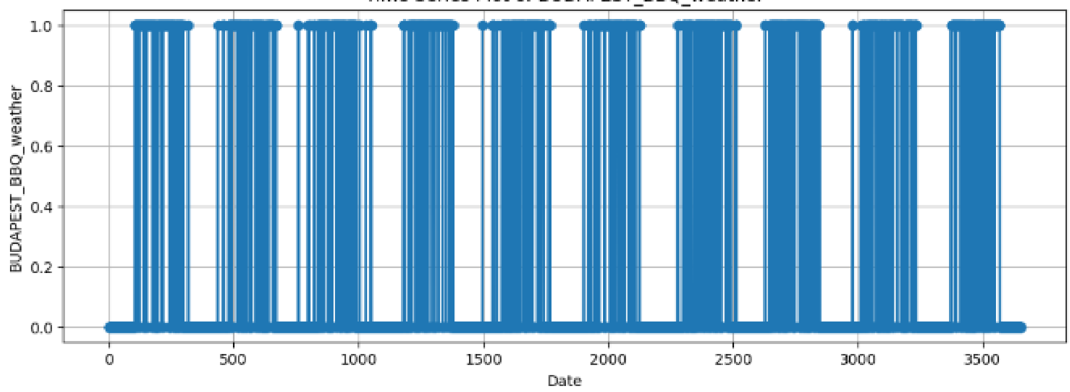


```
for column in bbq_labels.columns:
    plt.figure(figsize=(12, 4))
    plt.plot(bbq_labels.index, bbq_labels[column], marker='o',
linestyle='-')
    plt.title(f'Time Series Plot of {column}')
    plt.xlabel('Date')
    plt.ylabel(column)
    plt.grid(True)
    plt.show()
```

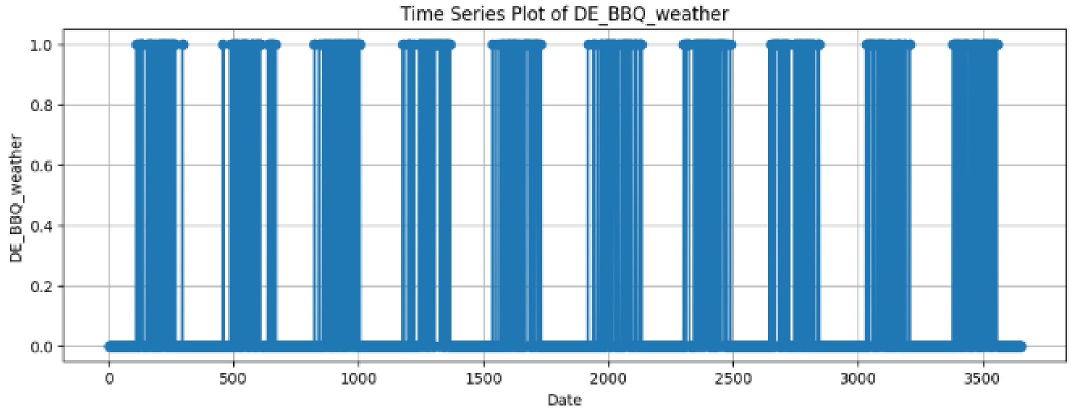
OUTPUT:



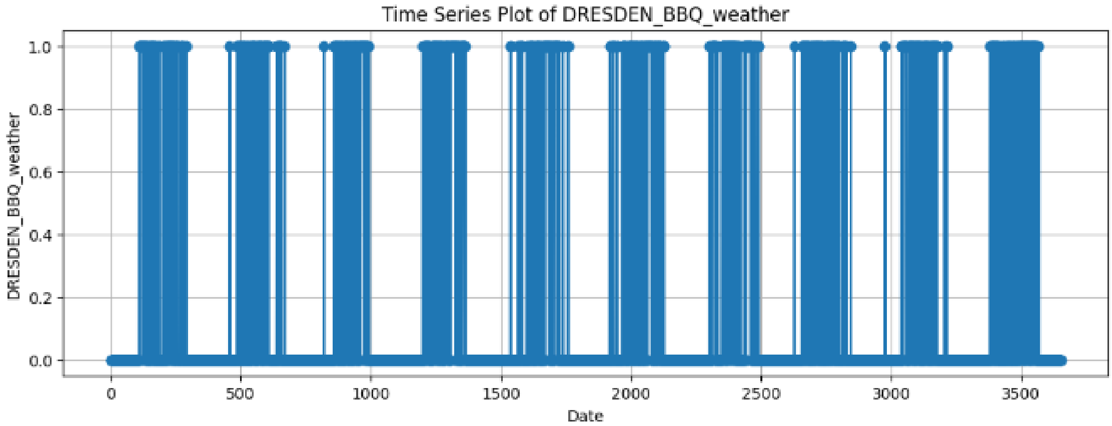
Time Series Plot of BUDAPEST_BBQ_weather



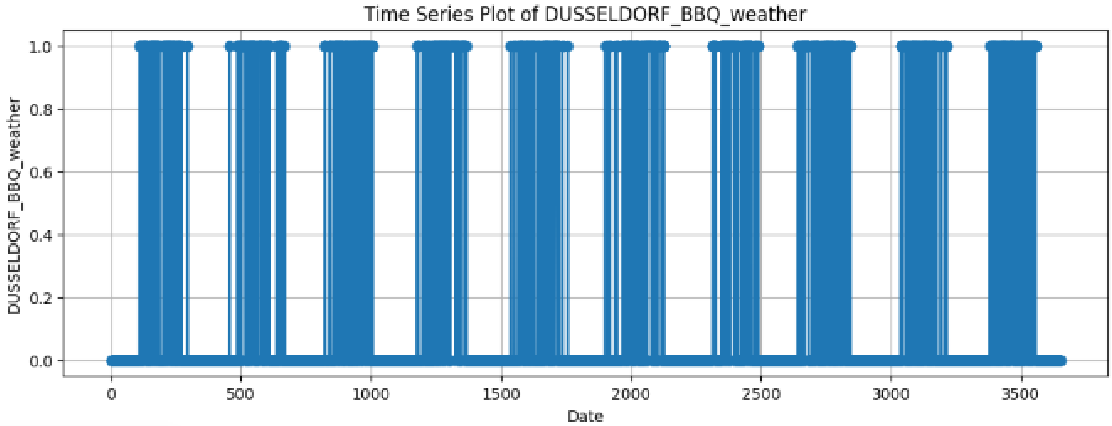
Time Series Plot of DE_BBQ_weather



Time Series Plot of DRESDEN_BBQ_weather



Time Series Plot of DUSSELDORF_BBQ_weather

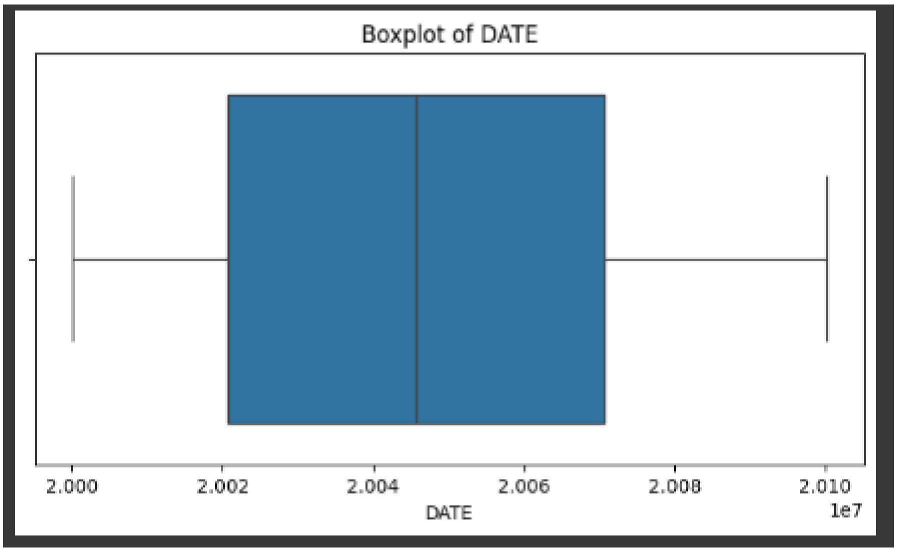


in another tab. Show diff

Time Series Plot of HEATHROW_BBQ_weather

```
for column in bbq_labels.select_dtypes(include=['float64',
'int64']).columns:
    plt.figure(figsize=(8, 4))
    sns.boxplot(x=bbq_labels[column])
    plt.title(f'Boxplot of {column}')
    plt.show()
```

OUTPUT:



CONCLUSION:

In conclusion, the weather_prediction_dataset offers a comprehensive array of weather-related data, encompassing essential atmospheric parameters crucial for weather prediction tasks. Through thorough data exploration, we gained valuable insights into the distribution, trends, and correlations within the dataset, enhancing our understanding of weather dynamics. Similarly, the weather_prediction_bbq_labels dataset, with its binary labels indicating barbecue suitability, complements the weather_prediction_dataset by providing actionable insights for outdoor activity planning. By analyzing the relationship between BBQ suitability and weather features, we can make informed decisions regarding outdoor events, considering weather factors for optimal planning. Moving forward, leveraging predictive modeling techniques and collaborating with stakeholders can further refine our understanding of weather patterns and improve decision-making processes for outdoor activities based on weather forecasts.

FUTURE WORK:

Future work for the weather_prediction_dataset involves leveraging advanced machine learning techniques to develop more accurate weather prediction models, incorporating additional weather features and exploring the integration of external datasets to enhance predictive performance. Furthermore, longitudinal studies could be conducted to analyze long-term weather trends and the impacts of climate change on weather patterns, providing valuable insights for climate research and adaptation strategies. Similarly, for the weather_prediction_bbq_labels dataset, future work could focus on refining predictive models to better forecast BBQ suitability based on weather conditions, considering factors such as temperature, precipitation, and sunshine duration. Collaborations with stakeholders and policymakers could lead to the development of decision-support tools for optimizing outdoor event planning, ensuring resilience to changing weather conditions and enhancing the overall outdoor experience.