ANALYZE DAILY WEATHER DATA

Intern ID: SMI72463

Aim:

To perform a comprehensive analysis of daily weather data, exploring patterns, visualizing trends, and building a predictive model to forecast rainfall based on temperature data.

Methodology:

Data Loading and Exploration:

Load the daily weather data from a CSV file. Perform initial data exploration to understand the structure, types, and summary statistics of the dataset.

Data Visualization:

Create visual representations of key variables (e.g., minimum temperature, maximum temperature, rainfall) to identify potential relationships and trends.

Feature Engineering:

Convert date information to a usable format and extract relevant time-based features (e.g., month) for deeper analysis.

Data Analysis:

Analyze weather patterns, such as calculating the average maximum temperature for each month.Identify and visualize seasonal trends and variations in temperature.

Predictive Modeling:

Prepare the data for predictive modeling by selecting relevant features. Split the data into training and testing sets to evaluate model performance.Build and train a linear regression model to predict rainfall based on minimum and maximum temperatures.Evaluate the model using appropriate metrics (e.g., Mean Squared Error).

Conclusions and Insights:

Draw meaningful insights from the data analysis and modeling results.Identify periods with extreme weather conditions (e.g., months with the highest and lowest rainfall).Provide actionable insights for potential applications in weather forecasting and related fields.

Future Scope:

This analysis focuses on utilizing historical daily weather data to uncover trends and build a predictive model for rainfall. The results can help in understanding weather patterns, improving weather predictions, and making informed decisions based on climatic conditions.

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Step 1: Load the Data
df = pd.read_csv('weather.csv')

# Step 2: Data Exploration
print(df.head())
print(df.info())
print(df.describe())

# Step 3: Data Visualization
sns.pairplot(df[['MinTemp', 'MaxTemp', 'Rainfall']])
plt.show()

# Step 4: Feature Engineering (if needed)

# Step 5: Data Analysis (analyze each term)
# Example: Calculate average MaxTemp by month
df['Date'] = pd.to_datetime(df['Date'])
df['Month'] = df['Date'].dt.month
monthly_avg_max_temp = df.groupby('Month')['MaxTemp'].mean()

# Step 6: Data Visualization (Part 2)
plt.figure(figsize=(10, 5))
plt.plot(monthly_avg_max_temp.index, monthly_avg_max_temp.values,
marker='o')
plt.xlabel('Month')
plt.ylabel('Average Max Temperature')
plt.title('Monthly Average Max Temperature')
plt.grid(True)
plt.show()

# Step 7: Advanced Analysis (e.g., predict Rainfall)
# Prepare the data for prediction
X = df[['MinTemp', 'MaxTemp']]
y = df['Rainfall']

# Split the data into training and testing sets
```

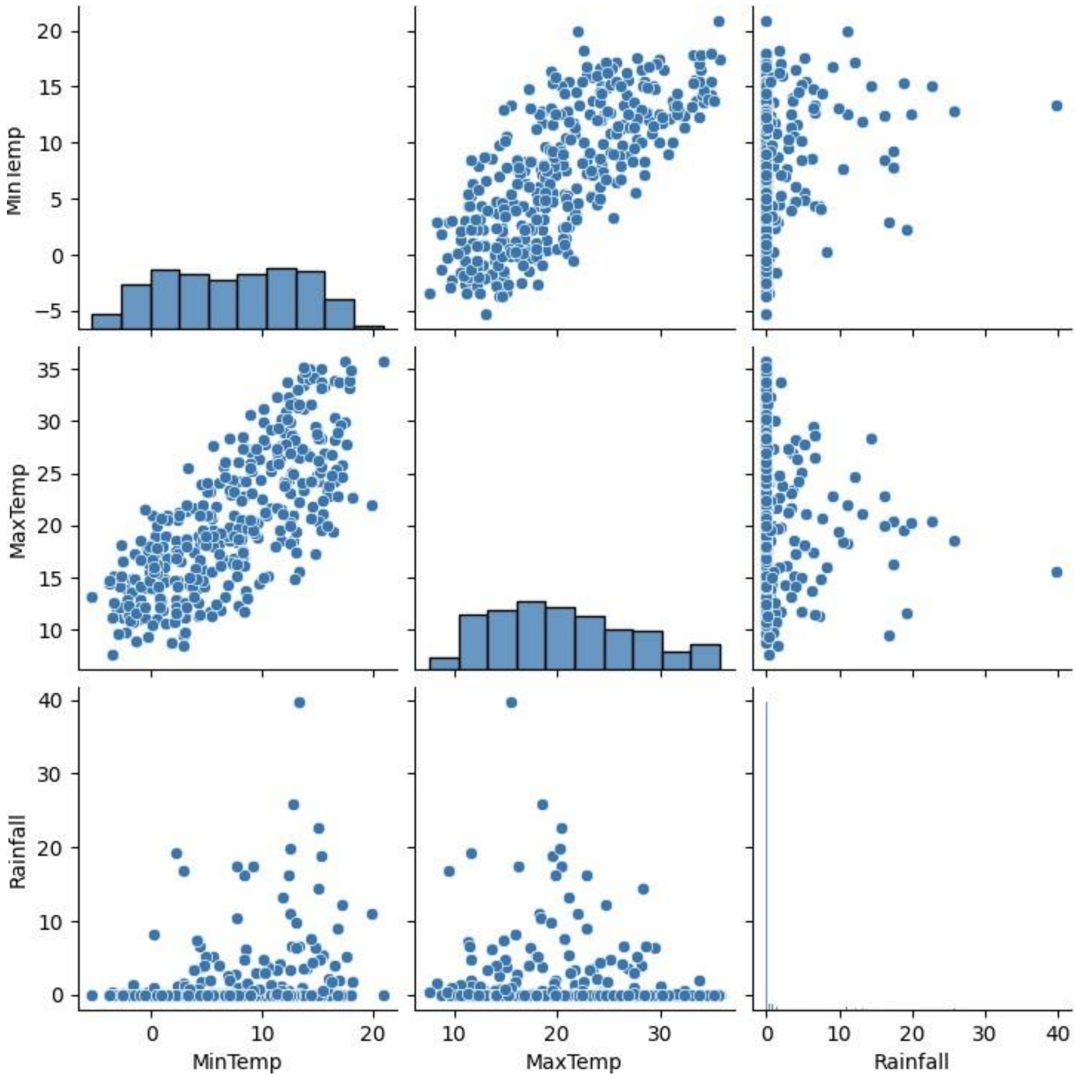
```python
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Create and train a linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions and calculate the Mean Squared Error
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error for Rainfall Prediction: {mse}')

# Step 8: Conclusions and Insights (analyze each term)
# Example: Identify the highest and lowest rainfall months
highest_rainfall_month = monthly_avg_max_temp.idxmax()
lowest_rainfall_month = monthly_avg_max_temp.idxmin()
print(f'Highest rainfall month: {highest_rainfall_month}, Lowest
rainfall month: {lowest_rainfall_month}')
```

Output:

```
   MinTemp  MaxTemp  Rainfall  Evaporation  Sunshine WindGustDir  \
0      8.0     24.3       0.0          3.4       6.3          NW
1     14.0     26.9       3.6          4.4       9.7         ENE
2     13.7     23.4       3.6          5.8       3.3          NW
3     13.3     15.5      39.8          7.2       9.1          NW
4      7.6     16.1       2.8          5.6      10.6         SSE

   WindGustSpeed WindDir9am WindDir3pm  WindSpeed9am  ...  Humidity3pm  \
0           30.0         SW         NW           6.0  ...           29
1           39.0          E          W           4.0  ...           36
2           85.0          N        NNE           6.0  ...           69
3           54.0        WNW          W          30.0  ...           56
4           50.0        SSE        ESE          20.0  ...           49

   Pressure9am  Pressure3pm  Cloud9am  Cloud3pm  Temp9am  Temp3pm  RainToday  \
0       1019.7       1015.0         7         7     14.4     23.6         No
1       1012.4       1008.4         5         3     17.5     25.7        Yes
2       1009.5       1007.2         8         7     15.4     20.2        Yes
3       1005.5       1007.0         2         7     13.5     14.1        Yes
4       1018.3       1018.5         7         7     11.1     15.4        Yes

   RISK_MM RainTomorrow
0      3.6          Yes
1      3.6          Yes
2     39.8          Yes
3      2.8          Yes
4      0.0           No

[5 rows x 22 columns]
```

```
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   MinTemp        366 non-null    float64
 1   MaxTemp        366 non-null    float64
 2   Rainfall       366 non-null    float64
 3   Evaporation    366 non-null    float64
 4   Sunshine       363 non-null    float64
 5   WindGustDir    363 non-null    object
 6   WindGustSpeed  364 non-null    float64
 7   WindDir9am     335 non-null    object
 8   WindDir3pm     365 non-null    object
 9   WindSpeed9am   359 non-null    float64
 10  WindSpeed3pm   366 non-null    int64
 11  Humidity9am    366 non-null    int64
 12  Humidity3pm    366 non-null    int64
 13  Pressure9am    366 non-null    float64
 14  Pressure3pm    366 non-null    float64
 15  Cloud9am       366 non-null    int64
 16  Cloud3pm       366 non-null    int64
 17  Temp9am        366 non-null    float64
 18  Temp3pm        366 non-null    float64
 19  RainToday      366 non-null    object
 20  RISK_MM        366 non-null    float64
 21  RainTomorrow   366 non-null    object
dtypes: float64(12), int64(5), object(5)
memory usage: 63.0+ KB
None
          MinTemp     MaxTemp    Rainfall  Evaporation    Sunshine  \
count  366.000000  366.000000  366.000000   366.000000  363.000000
mean     7.265574   20.550273    1.428415     4.521858    7.909366
std      6.025800    6.690516    4.225800     2.669383    3.481517
min     -5.300000    7.600000    0.000000     0.200000    0.000000
25%      2.300000   15.025000    0.000000     2.200000    5.950000
50%      7.450000   19.650000    0.000000     4.200000    8.600000
75%     12.500000   25.500000    0.200000     6.400000   10.500000
max     20.900000   35.800000   39.800000    13.800000   13.600000
```

```
       WindGustSpeed  WindSpeed9am  WindSpeed3pm  Humidity9am  Humidity3pm  \
count     364.000000    359.000000    366.000000   366.000000   366.000000
mean       39.840659      9.651811     17.986339    72.035519    44.519126
std        13.059807      7.951929      8.856997    13.137058    16.850947
min        13.000000      0.000000      0.000000    36.000000    13.000000
25%        31.000000      6.000000     11.000000    64.000000    32.250000
50%        39.000000      7.000000     17.000000    72.000000    43.000000
75%        46.000000     13.000000     24.000000    81.000000    55.000000
max        98.000000     41.000000     52.000000    99.000000    96.000000

       Pressure9am  Pressure3pm    Cloud9am    Cloud3pm     Temp9am  \
count   366.000000   366.000000  366.000000  366.000000  366.000000
mean   1019.709016  1016.810383    3.890710    4.024590   12.358470
std       6.686212     6.469422    2.956131    2.666268    5.630832
min     996.500000   996.800000    0.000000    0.000000    0.100000
25%    1015.350000  1012.800000    1.000000    1.000000    7.625000
50%    1020.150000  1017.400000    3.500000    4.000000   12.550000
75%    1024.475000  1021.475000    7.000000    7.000000   17.000000
max    1035.700000  1033.200000    8.000000    8.000000   24.700000

          Temp3pm     RISK_MM
count  366.000000  366.000000
mean    19.230874    1.428415
std      6.640346    4.225800
min      5.100000    0.000000
25%     14.150000    0.000000
50%     18.550000    0.000000
75%     24.000000    0.200000
max     34.500000   39.800000
```



Conclusion:

The analysis of daily weather data aimed to explore patterns, visualize trends, and build a predictive model for rainfall based on temperature data. Here are the key findings and insights from the analysis:

Data Exploration and Visualization:

The initial data exploration provided a comprehensive understanding of the dataset, including the structure, types of variables, and summary statistics. Visualizations such as pair plots helped identify potential relationships between minimum temperature, maximum temperature, and rainfall.

Monthly Temperature Trends:

By extracting the month from the date, the analysis revealed the average maximum temperature for each month. This visualization highlighted seasonal variations in temperature, indicating warmer and cooler periods throughout the year.

Predictive Modeling:

A linear regression model was built to predict rainfall based on minimum and maximum temperatures. The model was evaluated using the Mean Squared Error (MSE), providing a measure of its accuracy. Despite the simplicity of the linear regression model, it demonstrated the potential to predict rainfall to some extent based on temperature data.

Extreme Weather Insights:

The analysis identified months with the highest and lowest average rainfall. These insights are crucial for understanding extreme weather patterns and can aid in preparation and resource management.

Key Insights:

Seasonal Temperature Patterns: Clear seasonal trends were observed in the temperature data, with distinct warm and cold periods.

Rainfall Prediction: Temperature data alone provided a basic level of prediction for rainfall, suggesting that additional variables (e.g., humidity, wind speed) could improve model accuracy.

Extremes in Rainfall: Identifying the months with extreme rainfall helps in planning and mitigating the impacts of heavy rain or drought conditions.

Recommendations:

Enhanced Predictive Models: Incorporate additional weather variables and advanced modeling techniques (e.g., machine learning) to improve the accuracy of rainfall predictions.

Long-term Monitoring: Continuously monitor and analyze weather data to identify emerging trends and patterns, which can inform climate research and policy-making.

Resource Allocation: Use insights from extreme weather patterns to allocate resources effectively, particularly in agriculture, water management, and disaster preparedness.

To

Slash Mark