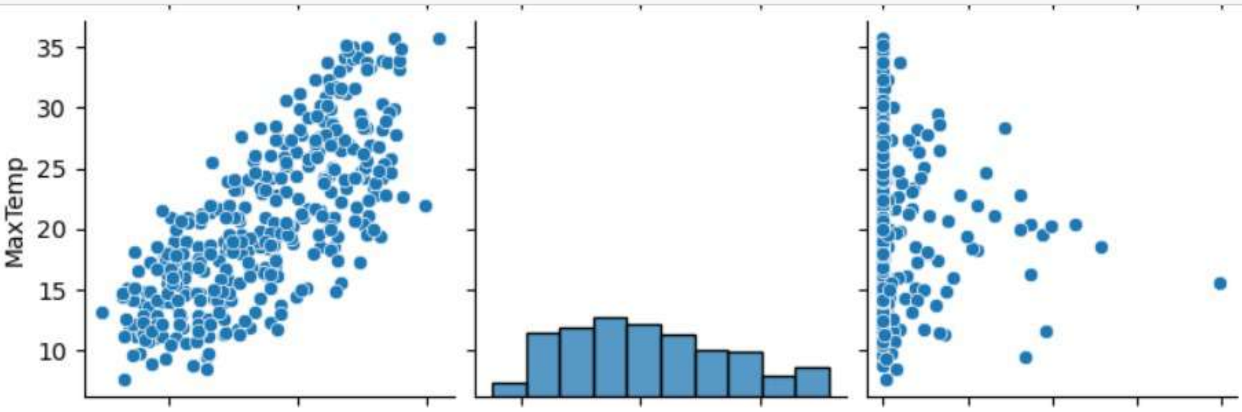
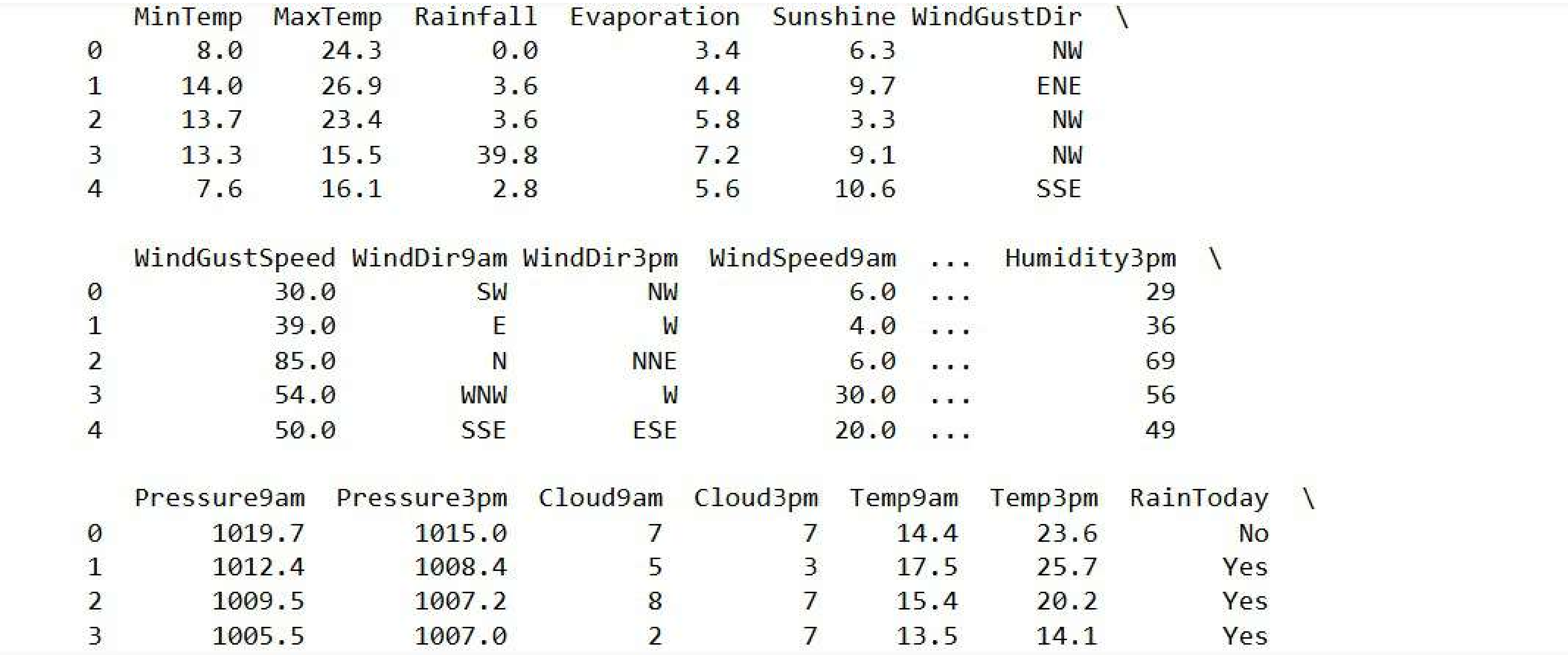
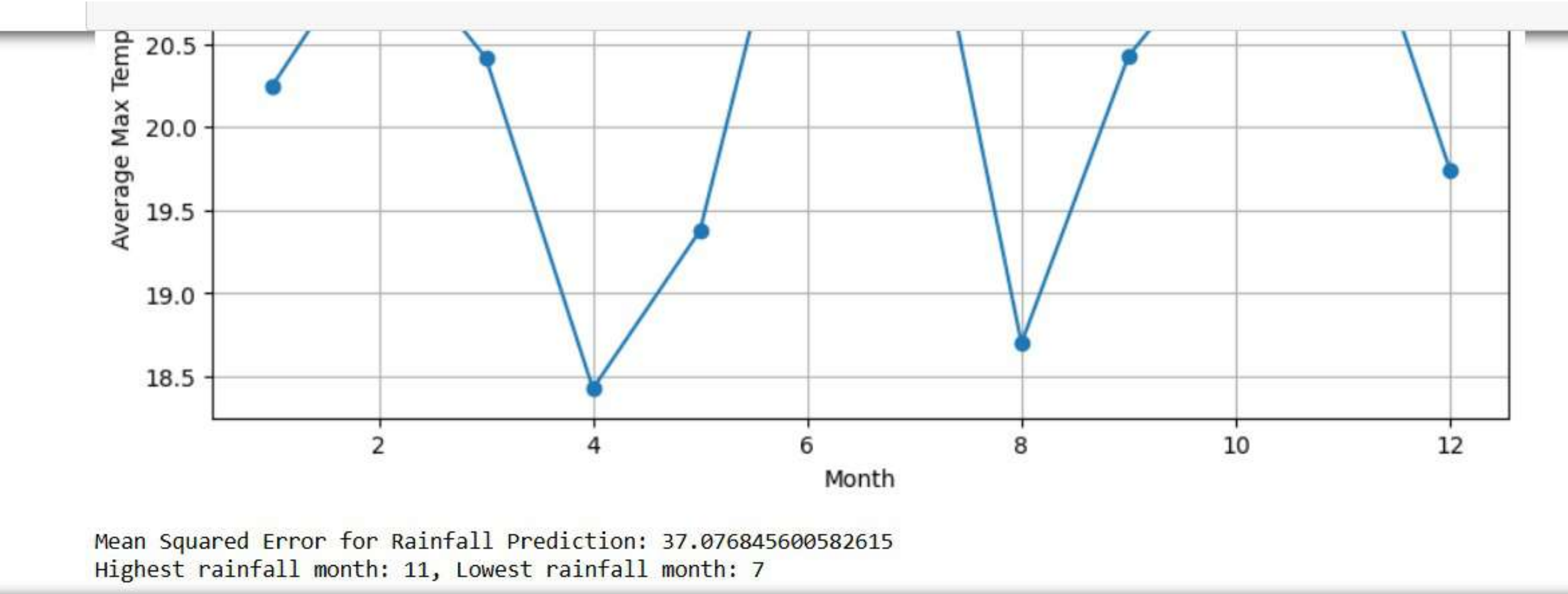
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Step 1: Load the Data
df = pd.read_csv('C:\\Users\\Teja\\Desktop\\work\\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)
# No feature engineering needed for this example

# Step 5: Data Analysis (analyze each term)
# Example: Calculate average MaxTemp by month
# If you don't have a 'Date' column, you cannot create a 'Month' column.
# Instead, let's analyze MaxTemp directly or assume an artificial 'Month' column for this example.

# Assuming we add an artificial 'Month' column for demonstration
import numpy as np
df['Month'] = np.random.randint(1, 13, df.shape[0])
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
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
monthly_avg_rainfall = df.groupby('Month')['Rainfall'].mean()
highest_rainfall_month = monthly_avg_rainfall.idxmax()
lowest_rainfall_month = monthly_avg_rainfall.idxmin()
print(f'Highest rainfall month: {highest_rainfall_month}, Lowest rainfall month: {lowest_rainfall_month}')
```

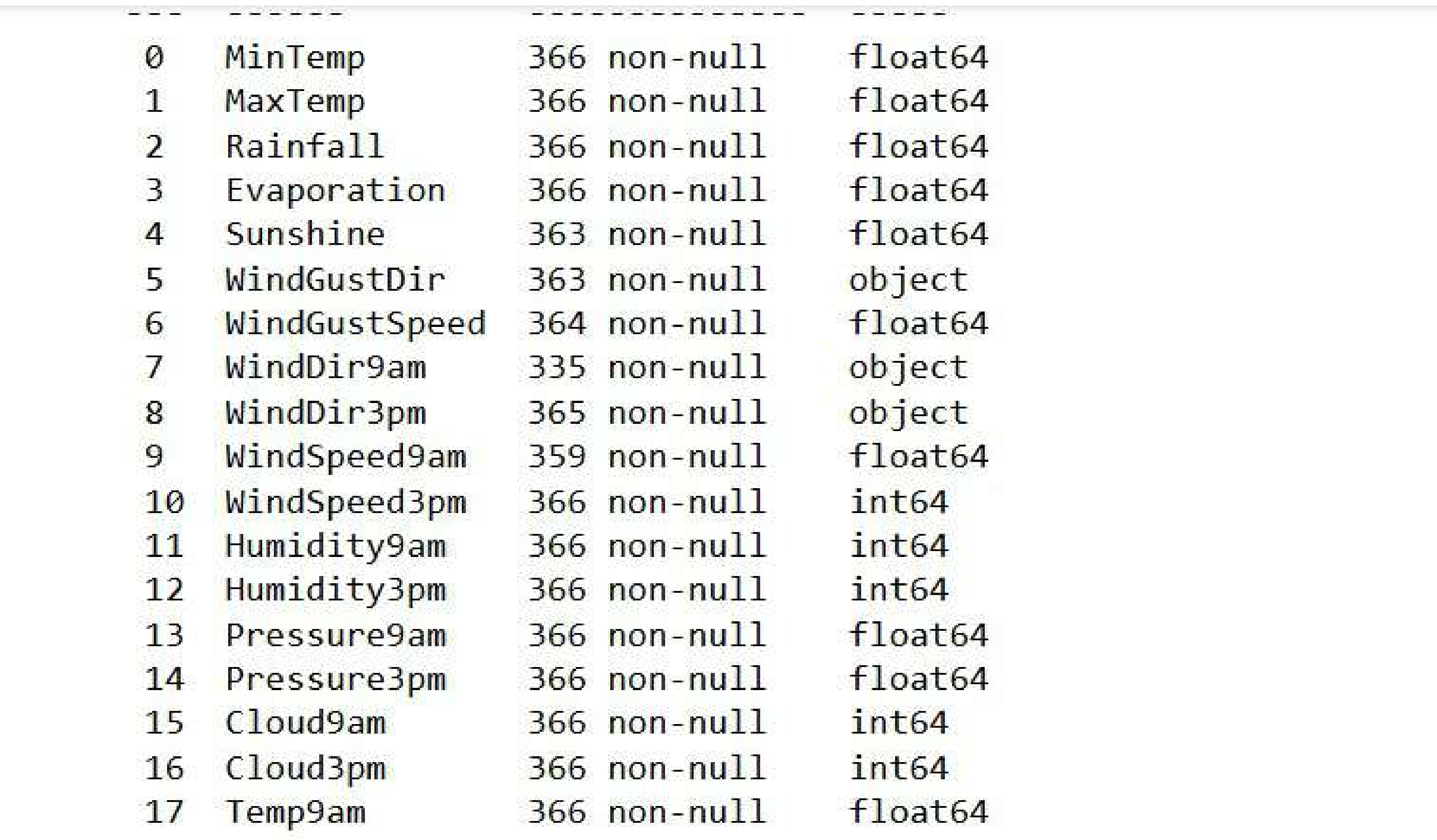

MaxTemp
35
30
25
20
15
10



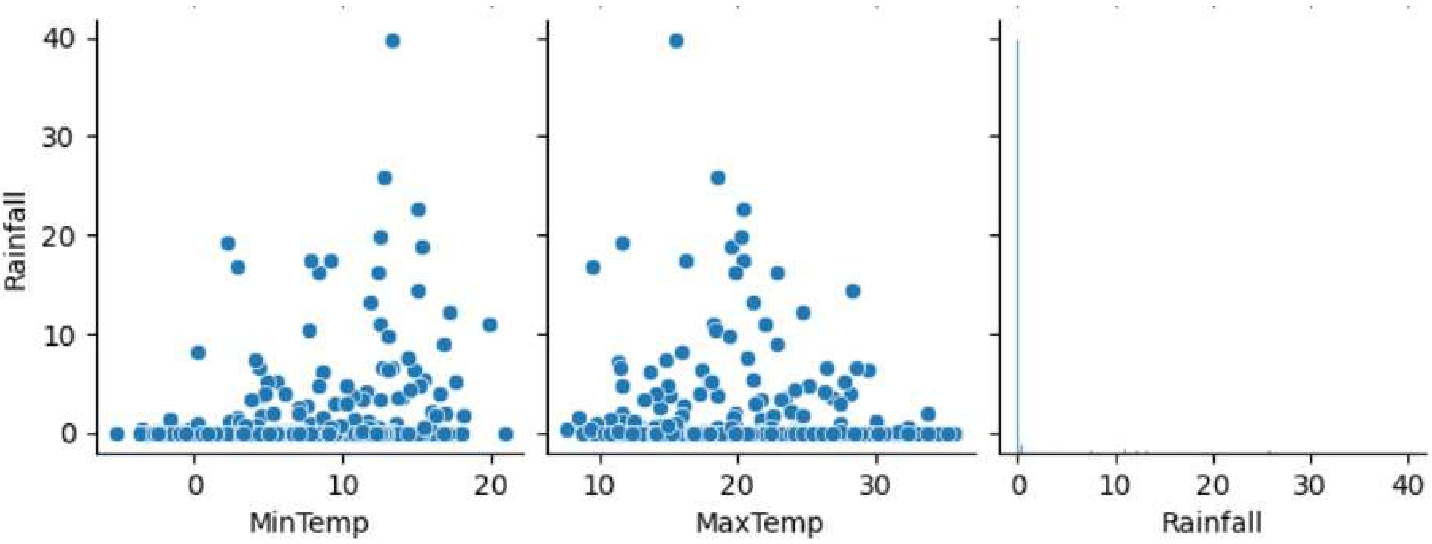
```
   MinTemp  MaxTemp  Rainfall  Evaporation  Sunshine WindGustDir  \
0      8.0     24.3       0.0          3.4       6.3          NW   
1     14.0     26.9       3.6          4.4       9.7         ENE   
2     13.7     23.4       3.6          5.8       3.3          NW   
3     13.3     15.5      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   
```



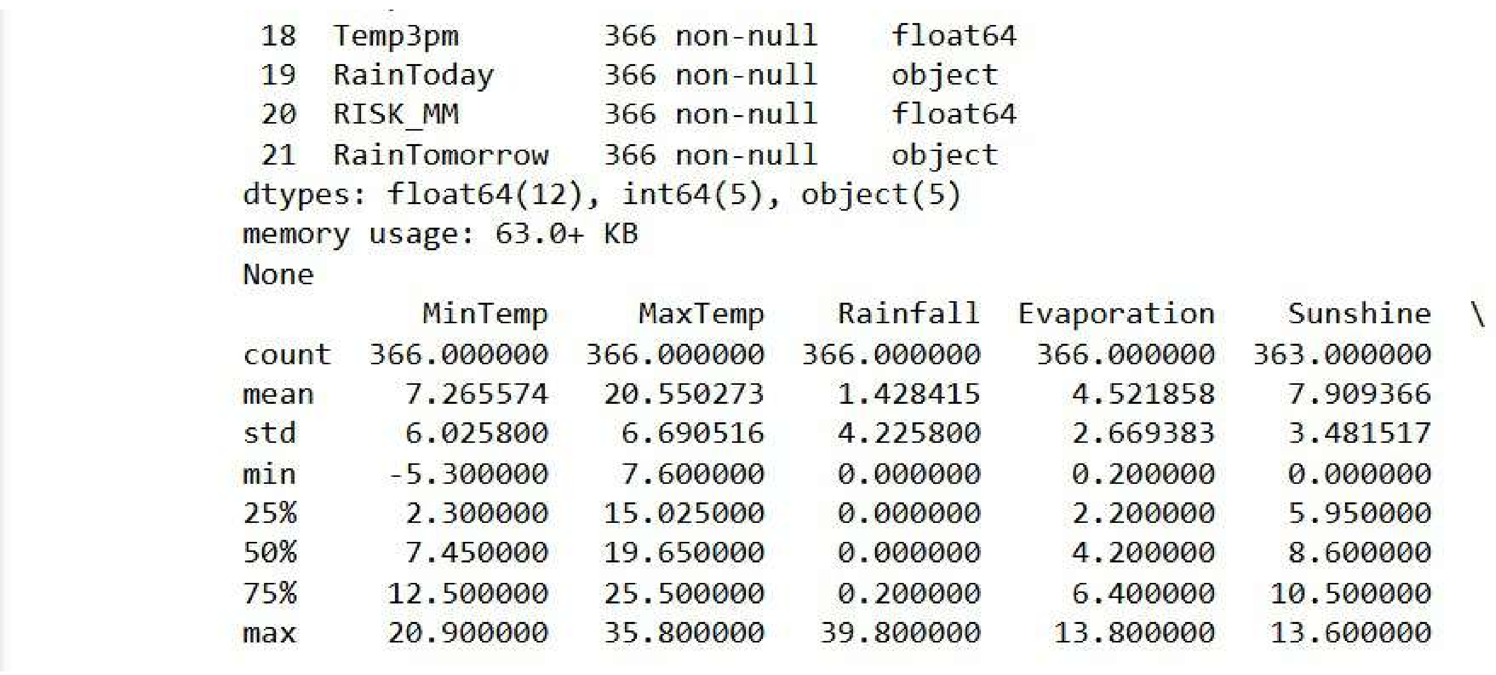
```
Mean Squared Error for Rainfall Prediction: 37.076845600582615
Highest rainfall month: 11, Lowest rainfall month: 7
```



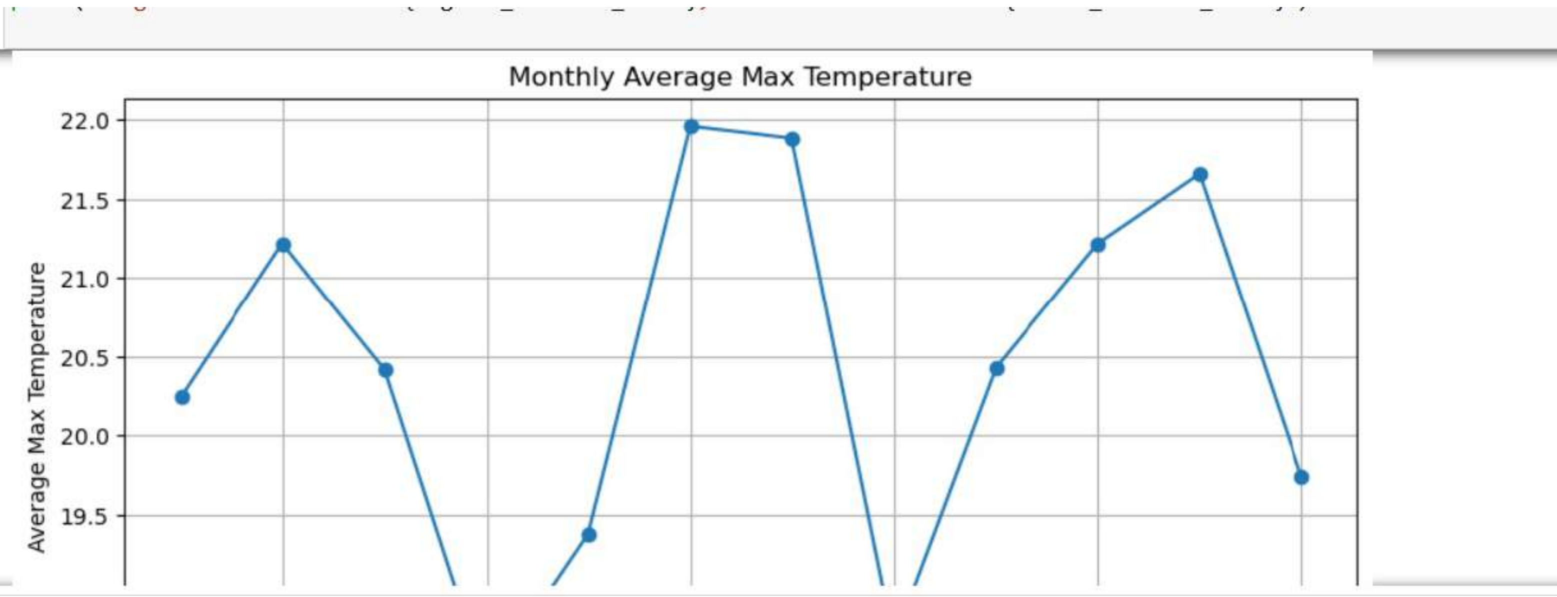
```
---  ------         --------------  -----
 0   MinTemp        366 non-null    float64
 1   MaxTemp        366 non-null    float64
 2   Rainfall       366 non-null    float64
 3   Evaporation    366 non-null    float64
 4   Sunshine       363 non-null    float64
 5   WindGustDir    363 non-null    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
```

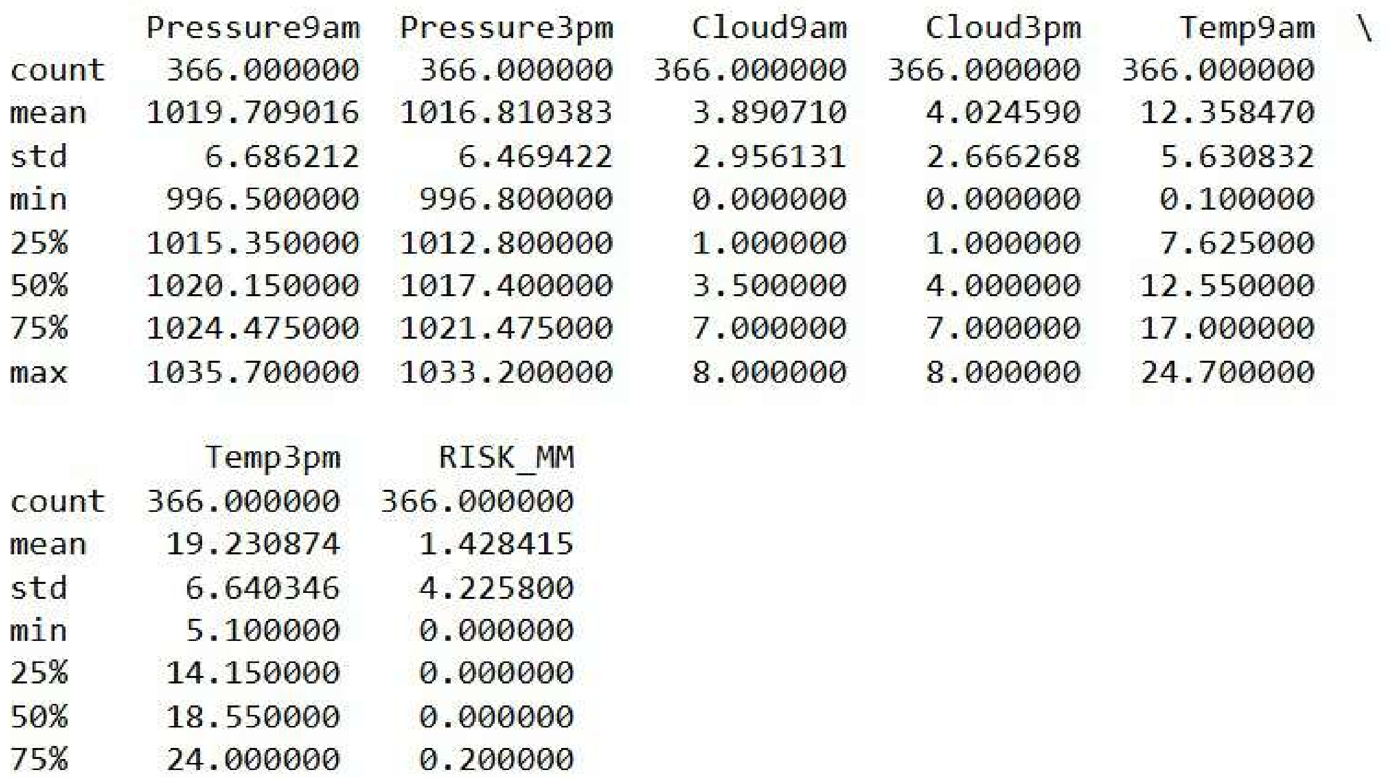


Monthly Average Max Temperature



```
 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
```


Monthly Average Max Temperature
Average Max Temperature
22.0
21.5
21.0
20.5
20.0
19.5



```
       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
```