In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]:
```
df = pd.read_csv("cleaned_weather.csv")
```

In [5]:
```
df.isna().sum()
```

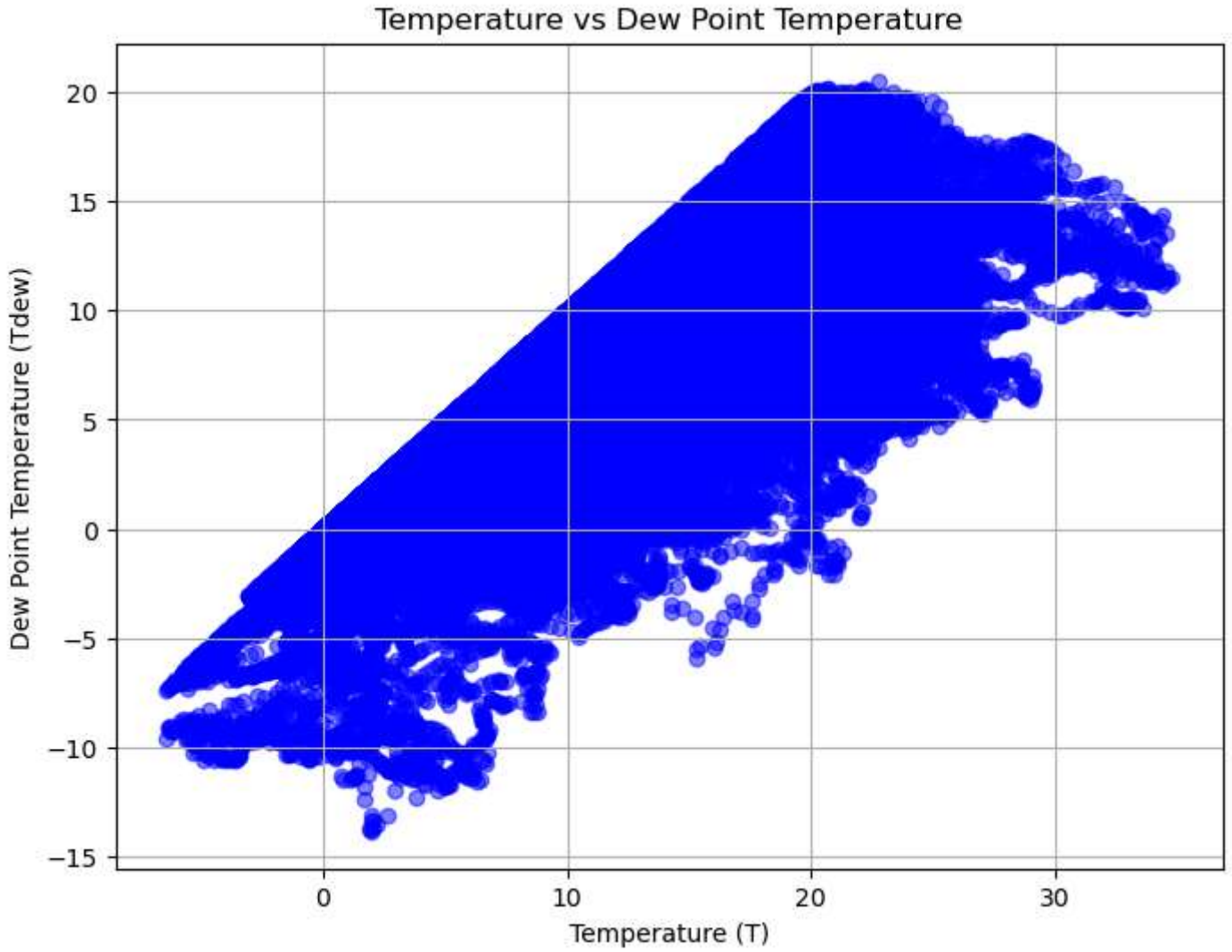
Out[5]:
```
date        0
p           0
T           0
Tpot        0
Tdew        0
rh          0
VPmax       0
VPact       0
VPdef       0
sh          0
H2OC        0
rho         0
wv          0
max. wv     0
wd          0
rain        0
raining     0
SWDR        0
PAR         0
max. PAR    0
Tlog        0
dtype: int64
```

In [7]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52696 entries, 0 to 52695
Data columns (total 21 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   date      52696 non-null  object 
 1   p         52696 non-null  float64
 2   T         52696 non-null  float64
 3   Tpot      52696 non-null  float64
 4   Tdew      52696 non-null  float64
 5   rh        52696 non-null  float64
 6   VPmax     52696 non-null  float64
 7   VPact     52696 non-null  float64
 8   VPdef     52696 non-null  float64
 9   sh        52696 non-null  float64
 10  H2OC      52696 non-null  float64
 11  rho       52696 non-null  float64
 12  wv        52696 non-null  float64
 13  max. wv   52696 non-null  float64
 14  wd        52696 non-null  float64
 15  rain      52696 non-null  float64
 16  raining   52696 non-null  float64
 17  SWDR      52696 non-null  float64
 18  PAR       52696 non-null  float64
 19  max. PAR  52696 non-null  float64
 20  Tlog      52696 non-null  float64
dtypes: float64(20), object(1)
memory usage: 8.4+ MB
```

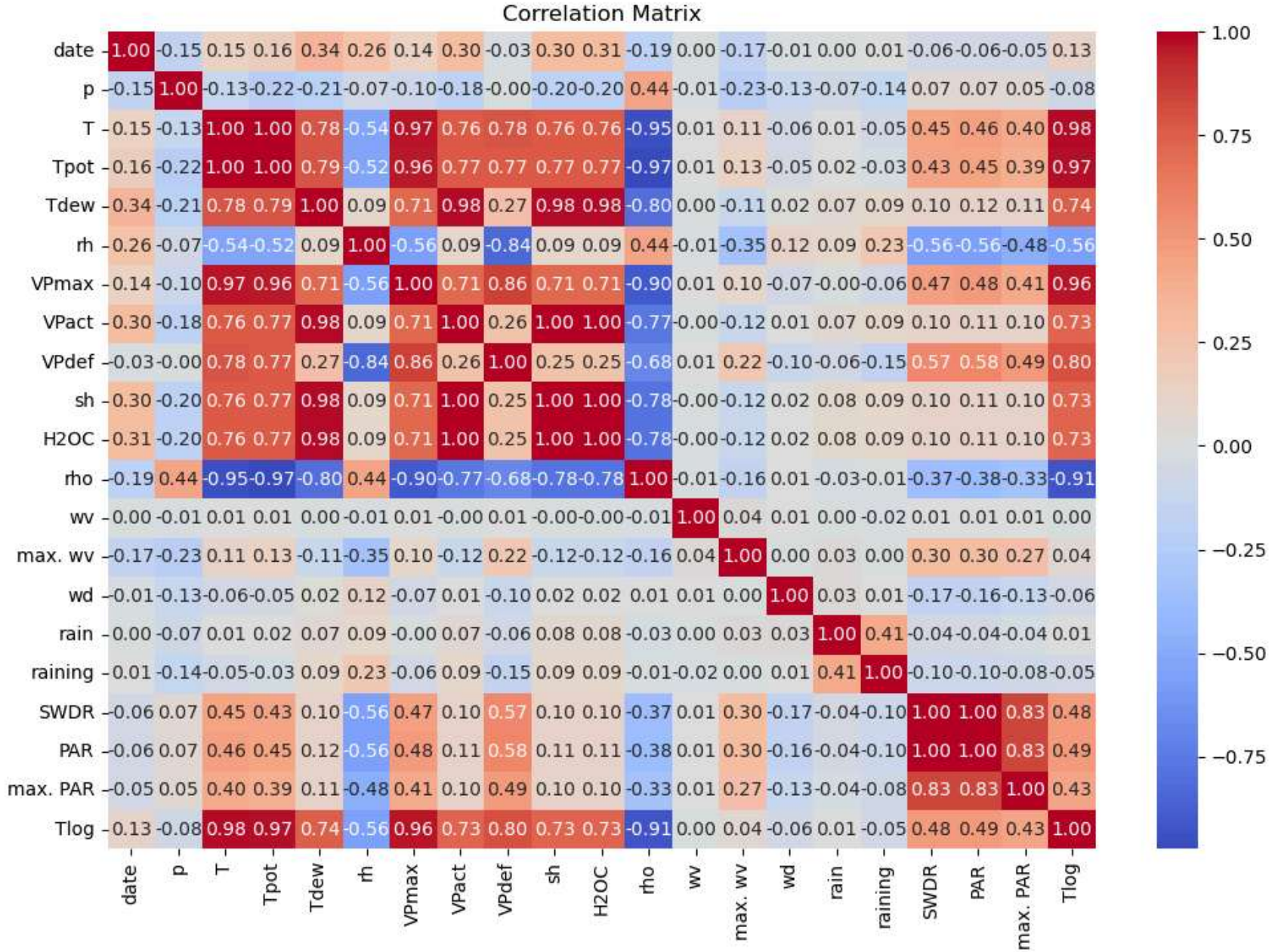
In [15]:
```
df['date'] = pd.to_datetime(df['date'])
```

In [9]:
```
plt.figure(figsize=(8, 6))
plt.scatter(df['T'], df['Tdew'], alpha=0.5, color='blue')
plt.title('Temperature vs Dew Point Temperature')
plt.xlabel('Temperature (T)')
plt.ylabel('Dew Point Temperature (Tdew)')
plt.grid(True)
plt.show()
```



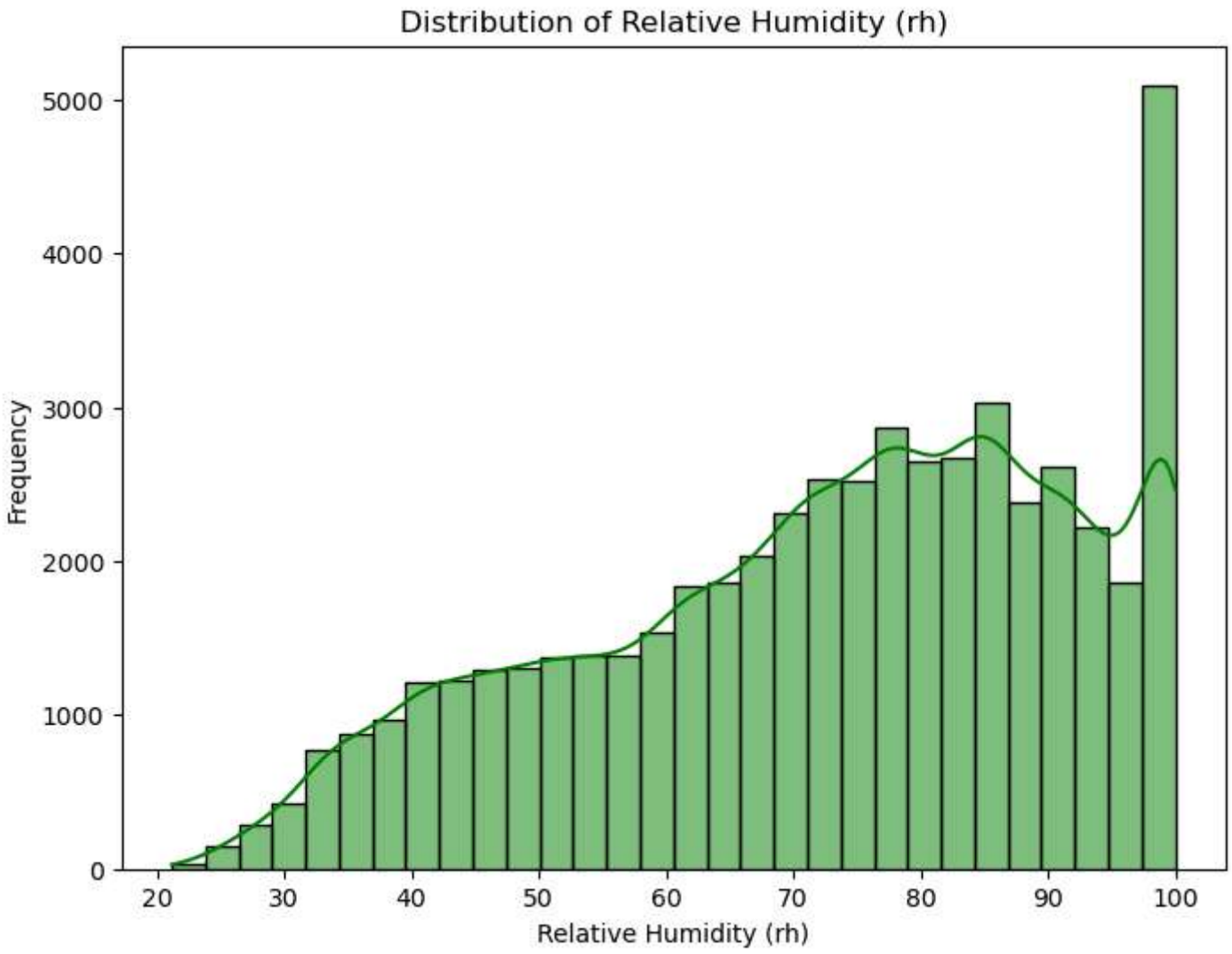
In [ ]:

In [17]:
```
plt.figure(figsize=(12, 8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```

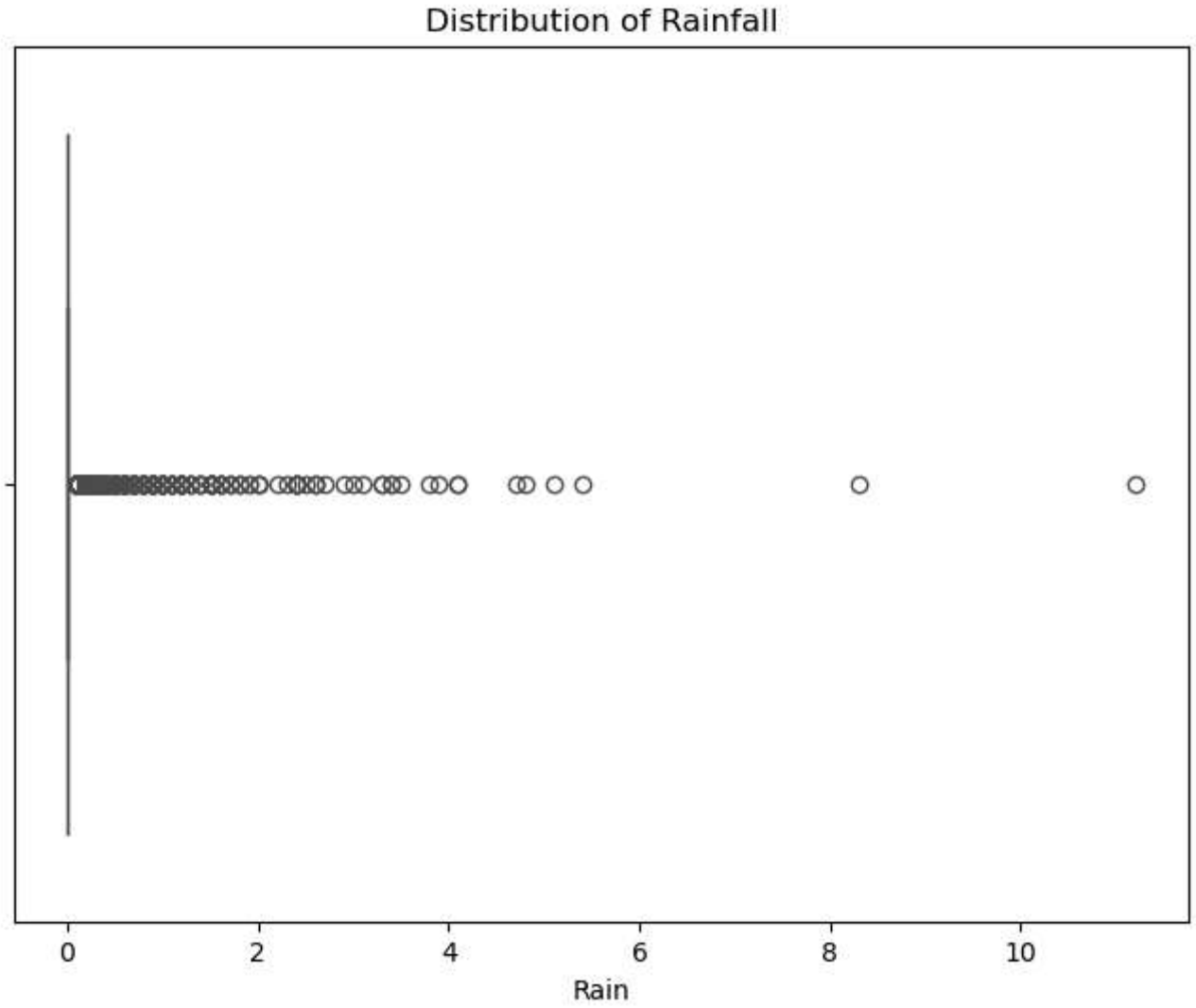


In [19]:
```
plt.figure(figsize=(8, 6))
sns.histplot(df['rh'], bins=30, kde=True, color='green')
plt.title('Distribution of Relative Humidity (rh)')
```

```
plt.xlabel('Relative Humidity (rh)')
plt.ylabel('Frequency')
plt.show()
```



In [21]:
```
plt.figure(figsize=(8, 6))
sns.boxplot(x=df['rain'], color='orange')
plt.title('Distribution of Rainfall')
plt.xlabel('Rain')
plt.show()
```

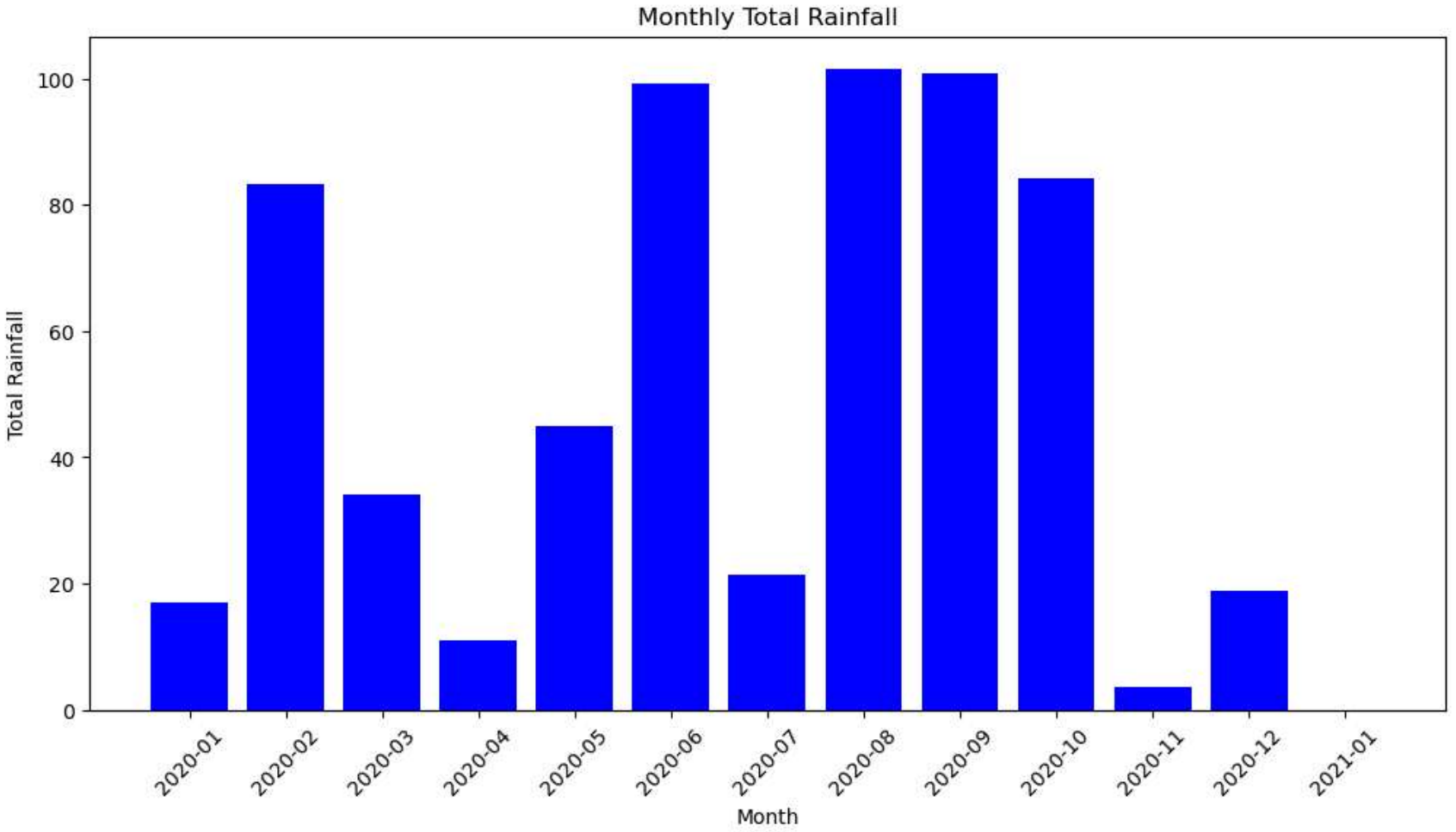


In [23]:
```
# Extract month from 'date'
df['month'] = df['date'].dt.to_period('M')

# Group by month and sum the 'rain'
monthly_rain = df.groupby('month')['rain'].sum().reset_index()

plt.figure(figsize=(12, 6))
plt.bar(monthly_rain['month'].astype(str), monthly_rain['rain'], color='blue')
plt.title('Monthly Total Rainfall')
```

```
plt.xlabel('Month')
plt.ylabel('Total Rainfall')
plt.xticks(rotation=45)
plt.show()
```



In [ ]: