```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.mlab as mlab
import seaborn as sns


df=pd.read_csv("/content/DengueAndClimateBangladesh.csv")


print(f"First 5 rows: {df.head()}")
print("<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>")
print(f"Info of Dataset: {df.info()}")
print("<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>")
print(f"Data Describe: {df.describe()}")
print("<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>")
print(f"Corelation of Dataset: {df.corr()}")
```

```
    First 5 rows:    YEAR  MONTH       MIN        MAX   HUMIDITY  RAINFALL  DENGUE
0   2008      1  12.993929  25.059539  78.824885  1.286636       0
1   2008      2  13.658286  26.383448  72.602956  0.687685       0
2   2008      3  20.411250  31.367558  76.903226  0.974194       0
3   2008      4  22.771889  34.031619  73.943810  0.980952       0
4   2008      5  23.945625  34.171060  77.360369  7.021198       0
    <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 134 entries, 0 to 133
    Data columns (total 7 columns):
     #   Column    Non-Null Count  Dtype
    ---  ------    --------------  -----
     0   YEAR      134 non-null    int64
     1   MONTH     134 non-null    int64
     2   MIN       134 non-null    float64
     3   MAX       134 non-null    float64
     4   HUMIDITY  134 non-null    float64
     5   RAINFALL  134 non-null    float64
     6   DENGUE    134 non-null    int64
    dtypes: float64(4), int64(3)
    memory usage: 7.5 KB
    Info of Dataset: None
    <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>
    Data Describe:              YEAR       MONTH         MIN         MAX    HUMIDITY  \
    count  134.000000  134.000000  134.000000  134.000000  134.000000
    mean   2013.089552    6.425373   20.943415   30.848480   80.115577
    std       3.233896    3.493015    4.830249    2.851090    5.253859
    min    2008.000000    1.000000   10.597946   23.522676   67.548387
    25%    2010.000000    3.000000   16.404490   29.277935   77.188214
    50%    2013.000000    6.000000   22.935495   31.987028   80.093810
    75%    2016.000000    9.000000   25.278970   32.676912   84.778111
    max    2019.000000   12.000000   26.485821   35.769333   88.384071

             RAINFALL       DENGUE
    count  134.000000   134.000000
    mean    27.840582   211.582090
    std     98.393795   451.214606
    min      0.000000     0.000000
    25%      0.547765     0.000000
    50%      6.008095    36.000000
    75%     12.774048   187.000000
    max    689.161290  3087.000000
    <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>
    Corelation of Dataset:              YEAR     MONTH       MIN       MAX  HUMIDITY  RAINFALL    DENGUE
    YEAR      1.000000 -0.039341 -0.026401 -0.042889  0.008876  0.335181  0.372338
    MONTH    -0.039341  1.000000  0.273338  0.084662  0.530021  0.021895  0.316265
    MIN      -0.026401  0.273338  1.000000  0.879903  0.502352  0.231499  0.301316
    MAX      -0.042889  0.084662  0.879903  1.000000  0.087856  0.064593  0.177811
    HUMIDITY  0.008876  0.530021  0.502352  0.087856  1.000000  0.094968  0.275363
    RAINFALL  0.335181  0.021895  0.231499  0.064593  0.094968  1.000000  0.368675
    DENGUE    0.372338  0.316265  0.301316  0.177811  0.275363  0.368675  1.000000
```

```python
df.isnull().sum()
```

```
    YEAR        0
    MONTH       0
    MIN         0
    MAX         0
    HUMIDITY    0
    RAINFALL    0
    DENGUE      0
    dtype: int64
```

```python
# Histogram of the DENGUE variable
plt.hist(df['DENGUE'], bins=10)
plt.xlabel('Number of Dengue Cases')
plt.ylabel('Frequency')
```

```
plt.title('Distribution of Dengue Cases')
plt.show()
```
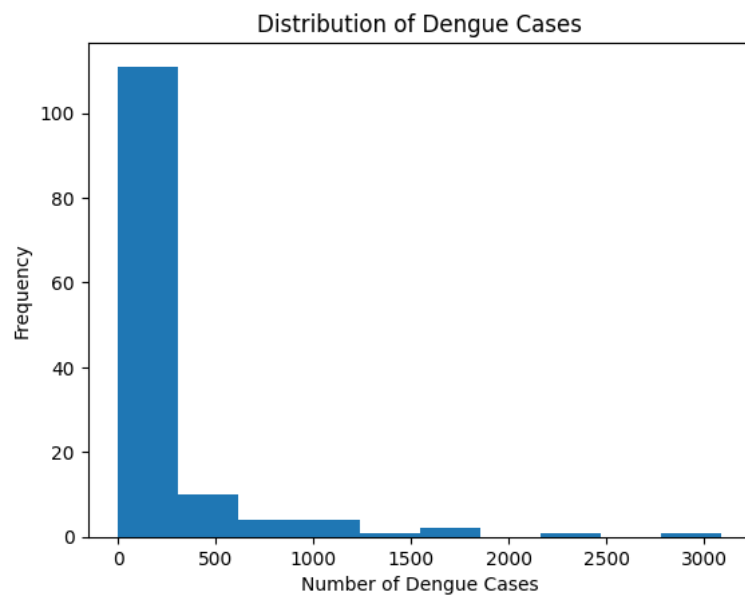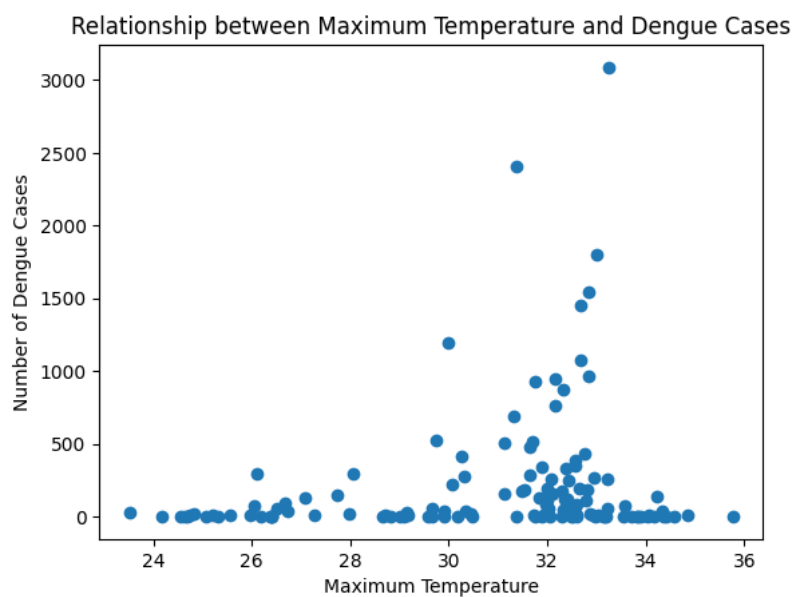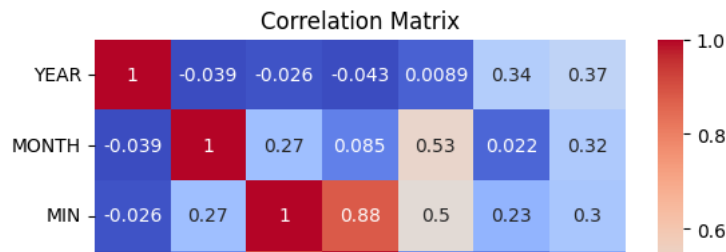

Distribution of Dengue Cases

```
plt.scatter(df['MAX'], df['DENGUE'])
plt.xlabel('Maximum Temperature')
plt.ylabel('Number of Dengue Cases')
plt.title('Relationship between Maximum Temperature and Dengue Cases')
plt.show()
```


Relationship between Maximum Temperature and Dengue Cases

```
# Correlation matrix heatmap
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

```
sns.boxplot(x='MONTH', y='DENGUE', data=df)
plt.xlabel('Month')
plt.ylabel('Number of Dengue Cases')
plt.title('Distribution of Dengue Cases Across Months')
plt.show()
```
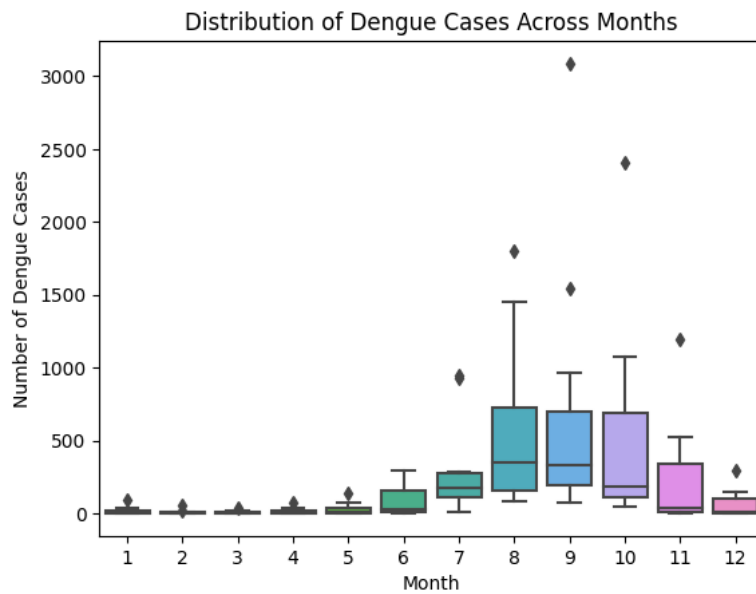


```
df_yearly = df.groupby('YEAR')['DENGUE'].mean().reset_index()
sns.barplot(x='YEAR', y='DENGUE', data=df_yearly)
plt.xlabel('Year')
plt.ylabel('Average Number of Dengue Cases')
plt.title('Average Dengue Cases by Year')
plt.show()
```
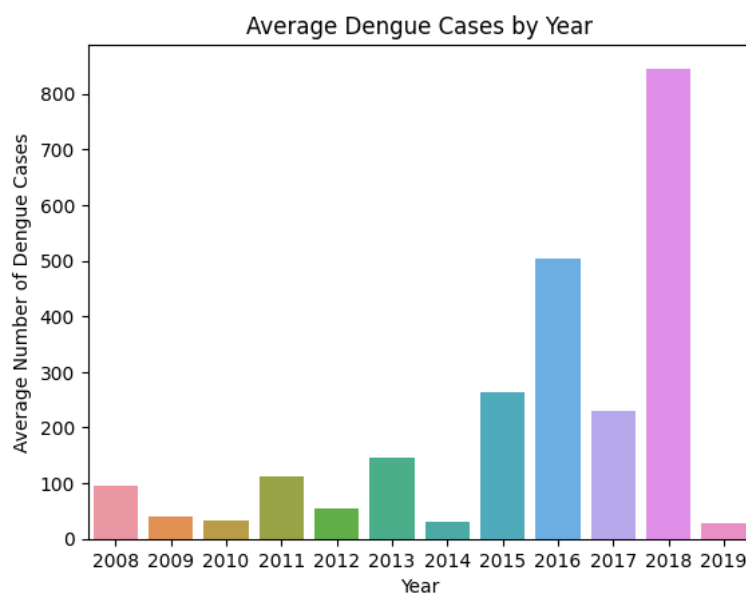


```
df['DATE'] = pd.to_datetime(df['YEAR'].astype(str) + '-' + df['MONTH'].astype(str), format='%Y-%m')
df_monthly = df.groupby('DATE')['DENGUE'].sum().reset_index()
plt.plot(df_monthly['DATE'], df_monthly['DENGUE'])
plt.xlabel('Date')
plt.ylabel('Number of Dengue Cases')
```

```
plt.title('Trend of Dengue Cases Over Time')
plt.xticks(rotation=45)
plt.show()
```
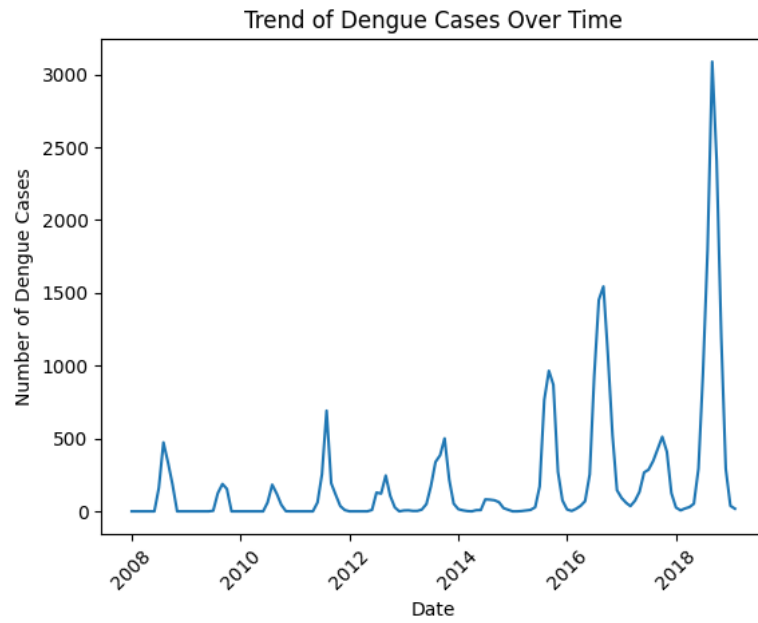


```
sns.pairplot(df[['MIN', 'MAX', 'HUMIDITY', 'RAINFALL', 'DENGUE']])
plt.show()
```
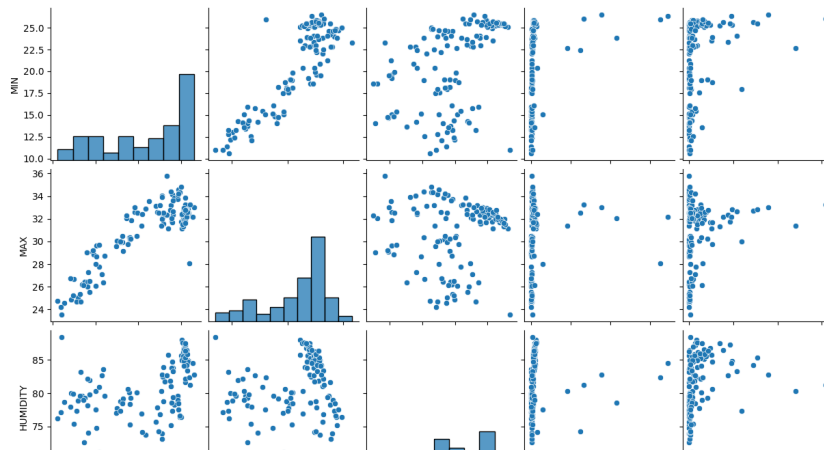
```
df_monthly_avg_rainfall = df.groupby('MONTH')['RAINFALL'].mean().reset_index()
sns.barplot(x='MONTH', y='RAINFALL', data=df_monthly_avg_rainfall)
plt.xlabel('Month')
plt.ylabel('Average Rainfall')
plt.title('Average Rainfall by Month')
plt.show()
```
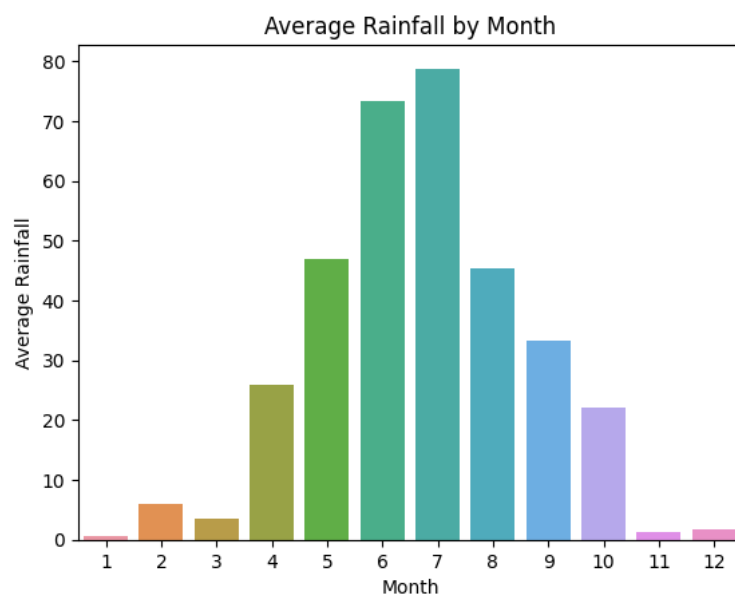


```
df_monthly_yearly = df.groupby(['YEAR', 'MONTH'])['DENGUE'].sum().reset_index()
sns.barplot(x='MONTH', y='DENGUE', hue='YEAR', data=df_monthly_yearly)
plt.xlabel('Month')
plt.ylabel('Number of Dengue Cases')
plt.title('Distribution of Dengue Cases by Month and Year')
plt.legend(title='Year')
plt.show()
```

Distribution of Dengue Cases by Month and Year

```
plt.figure(figsize=(16, 6))
plt.title('Time series plot of DENGUE')
ax = sns.lineplot(x="YEAR", y="DENGUE",ci=10,color="blue", label="DENGUE",data=df)
```

```
<ipython-input-14-983fbe6538c9>:3: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=('ci', 10)` for the same effect.

  ax = sns.lineplot(x="YEAR", y="DENGUE",ci=10,color="blue", label="DENGUE",data=d
```
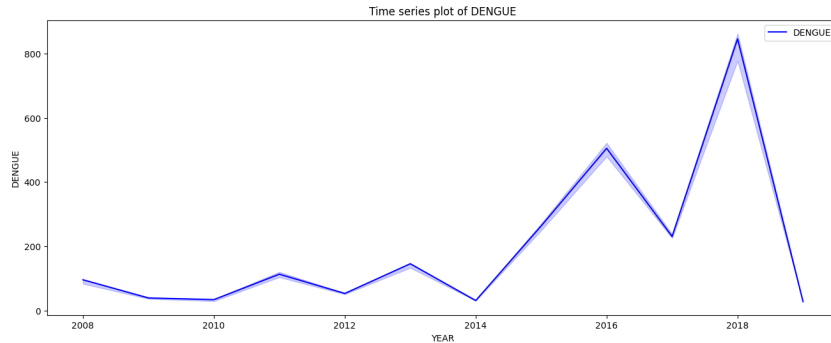


```
df['TEMP_RANGE'] = pd.cut(df['MAX'], bins=[0, 20, 25, 30, 35], labels=['<20', '20-25', '25-30', '30-35'])
sns.violinplot(x='TEMP_RANGE', y='DENGUE', data=df)
plt.xlabel('Temperature Range (MAX)')
plt.ylabel('Number of Dengue Cases')
plt.title('Distribution of Dengue Cases by Temperature Range')
plt.show()
```


Distribution of Dengue Cases by Temperature Range
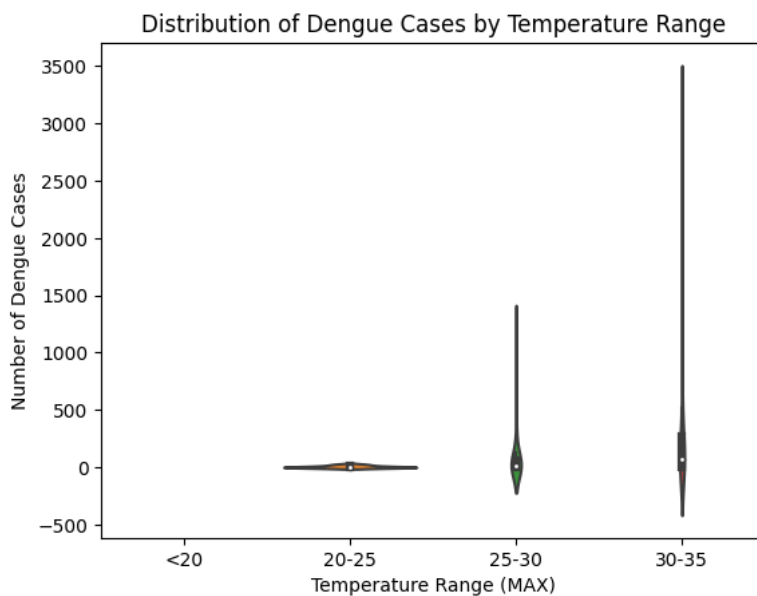
```
df_humidity = df.groupby('HUMIDITY')['DENGUE'].sum().reset_index()
sns.barplot(x='HUMIDITY', y='DENGUE', data=df_humidity)
plt.xlabel('Humidity')
plt.ylabel('Number of Dengue Cases')
plt.title('Distribution of Dengue Cases by Humidity')
plt.xticks(rotation=45)
```

```
plt.show()
```


Distribution of Dengue Cases by Humidity
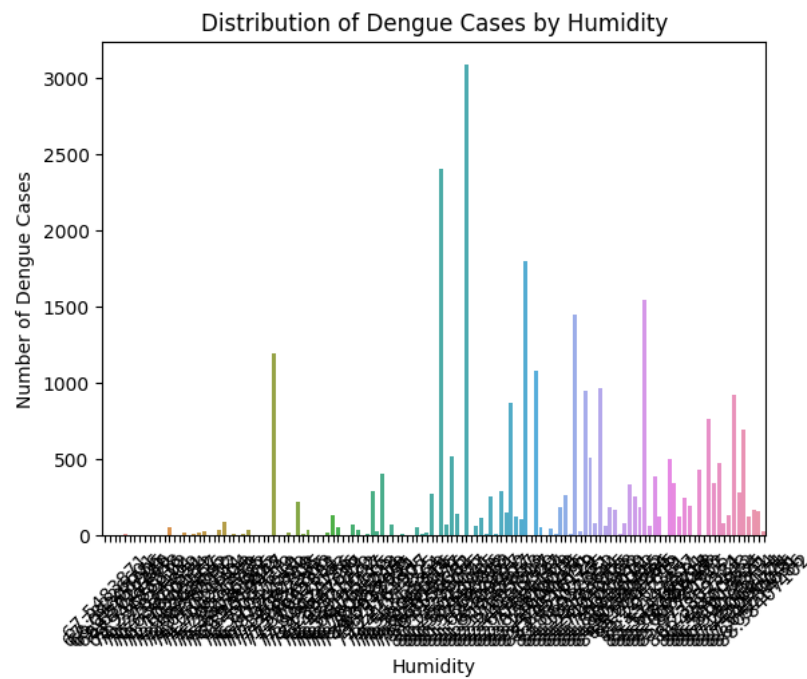
```
plt.figure(figsize=(20,12),facecolor='red',edgecolor='blue')
df.hist()
plt.show()
```

```
<Figure size 2000x1200 with 0 Axes>
```