# Analysis of data from every month’s average Temperature and average Rainfall from 1901 to 2019 in Bangladesh

**Snippet of First few Columns of Data**

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Year | Month | Temperatu | Rain | | | |
| 2 | 1901 | 1 | 16.98 | 18.54 | | | |
| 3 | 1901 | 2 | 19.9 | 16.25 | | | |
| 4 | 1901 | 3 | 24.32 | 70.8 | | | |
| 5 | 1901 | 4 | 28.18 | 66.16 | | | |
| 6 | 1901 | 5 | 27.89 | 267.22 | | | |
| 7 | 1901 | 6 | 28.89 | 341.04 | | | |
| 8 | 1901 | 7 | 28.33 | 540.91 | | | |
| 9 | 1901 | 8 | 27.92 | 493.21 | | | |
| 10 | 1901 | 9 | 27.61 | 291.55 | | | |
| 11 | 1901 | 10 | 27.09 | 199.17 | | | |
| 12 | 1901 | 11 | 22.17 | 126.29 | | | |
| 13 | 1901 | 12 | 18.56 | 1.69 | | | |
| 14 | 1902 | 1 | 18.55 | 1.29 | | | |
| 15 | 1902 | 2 | 20.13 | 0.15 | | | |
| 16 | 1902 | 3 | 25.55 | 62.77 | | | |
| 17 | 1902 | 4 | 26.56 | 229.59 | | | |
| 18 | 1902 | 5 | 27.32 | 302.2 | | | |

Download the dataset from here.

**Here are the summary statistics for the temperature and rainfall data from 1901 to 2019:**

**Temperature (°C)**

- **Count:** 1428 (number of observations)
- **Mean:** 25.08°C (average monthly temperature over the period)
- **Standard Deviation (std):** 3.68°C (variation in the monthly temperatures)
- **Minimum (min):** 16.80°C (lowest recorded monthly average temperature)
- **25th percentile (25%):** 22.12°C (25% of the monthly temperatures are below this value)
- **Median (50%):** 27.02°C (middle value of the monthly temperatures)
- **75th Percentile (75%):** 28.03°C (75% of the monthly temperatures are below this value)
- **Maximum (max):** 29.53°C (highest recorded monthly average temperature)

**Rainfall (mm)**

- **Count:** 1428 (number of observations)
- **Mean:** 202.80 mm (average monthly rainfall over the period)
- **Standard Deviation (std):** 201.97 mm (variation in the monthly rainfall amounts)
- **Minimum (min):** 0.00 mm (lowest recorded monthly rainfall)

- **25th percentile (25%):** 18.60 mm (25% of the monthly rainfall amounts are below this value) **Median (50%):** 146.46 mm (middle value of the monthly rainfall amounts)
- **75th Percentile (75%):** 345.54 mm (75% of the monthly rainfall amounts are below this value)
- **Maximum (max):** 1012.02 mm (highest recorded monthly rainfall)

These statistics provide a comprehensive summary of the central tendency, dispersion, and shape of the distribution of the temperature and rainfall datasets. If you have any other questions or need further analysis, feel free to ask.

**Descriptive Statistics and Correlation**

**Descriptive Statistics:**

- **Temperature:** The average monthly temperature is around 25.08˚C, with a standard deviation of 3.68˚C. The temperatures ranged from 16.8˚C to 29.53˚C.
- **Rain:** The average monthly rainfall is about 202.8202.8 mm, with a standard deviation of 201.97 mm. The rainfall ranged from 00 mm to 1012.02 mm.

**Correlation:**

The Pearson correlation coefficient between temperature and rain is approximately 0.7020.702, with a very low p-value ($3.29\times10-2123.29\times10-212$), indicating a statistically significant positive relationship.

# Regression Analysis:

**Regression Model Summary:**

- **Dependent Variable (Dep. Variable):** Rain
- **R-squared ($R^2$):** 0.492 (Approximately 49.2% of the variance in Rain is explained by the model.)
- **Model:** Ordinary Least Squares (OLS) regression
- **Adj. R-squared (Adjusted $R^2$):** 0.492
- **F-statistic:** 1383
- **Prob (F-statistic):** 3.29e-212 (Very low p-value, indicating statistical significance)
- **Method:** Least Squares
- **Date:** Tue, 12 Sep 2023
- **Time:** 20:10:38
- **Log-Likelihood:** -9121.6
- **No. Observations:** 1428
- **AIC (Akaike Information Criterion):** 1.825e+04 (Lower values are preferred)
- **Df Residuals:** 1426
- **BIC (Bayesian Information Criterion):** 1.826e+04 (Lower values are preferred)
- **Df Model:** 1
- **Covariance Type:** Nonrobust

**Coefficient Estimates:**

- **Intercept (const):** -762.4500
- **Temperature:** 38.4884

**Standard Errors:**

- **Intercept (std err):** 26.231
- **Temperature (std err):** 1.035

**t-statistics:**

- **Intercept (t):** -29.067
- **Temperature (t):** 37.192

**P-values:**

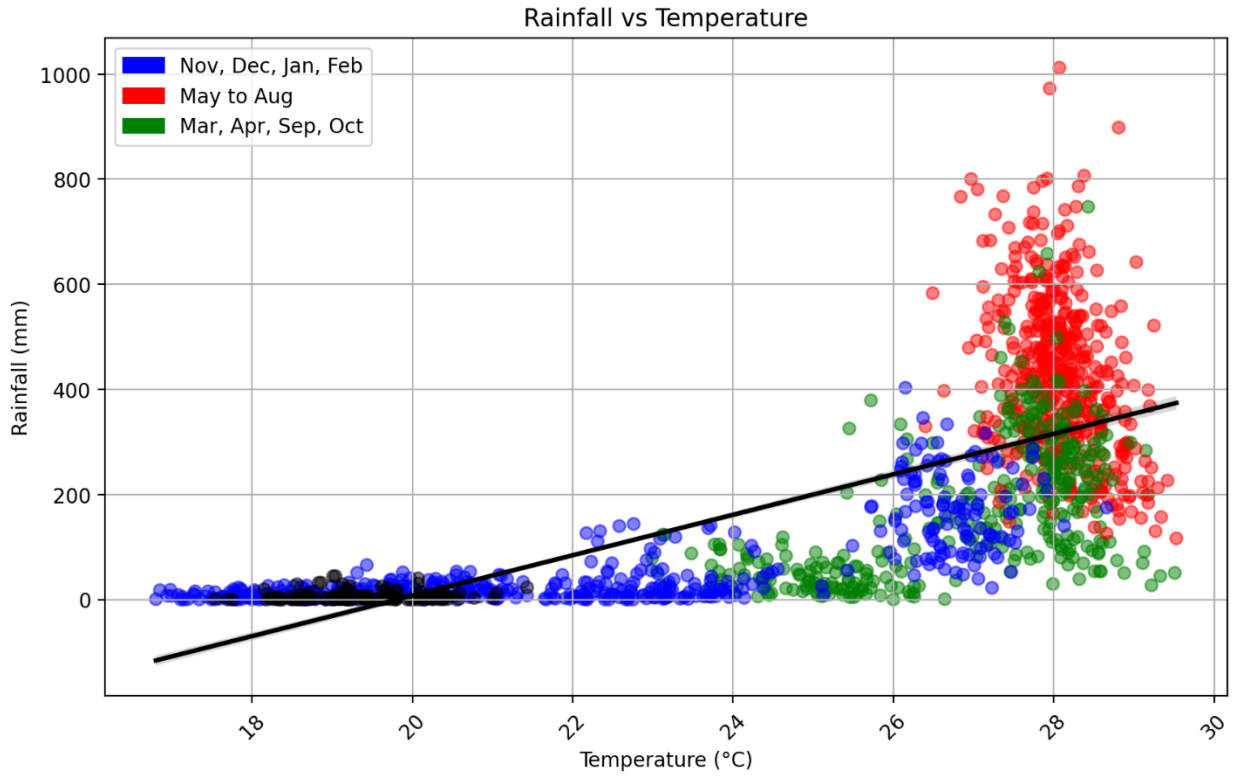
- **Intercept (P>|t|):** 0.000
- **Temperature (P>|t|):** 0.000

**Confidence Intervals (95%):**

- **Intercept:** [-813.905, -710.995]
- **Temperature:** [36.458, 40.518]

**Residual Diagnostics:**

- **Omnibus:** 162.157
- **Durbin-Watson:** 1.278
- **Prob(Omnibus):** 0.000 (Very low p-value, indicating non-normality of residuals)
- **Jarque-Bera (JB):** 247.191
- **Skew:** 0.809
- **Prob(JB):** 2.10e-54 (Very low p-value, indicating non-normality of residuals)
- **Kurtosis:** 4.239
- **Cond. No. (Condition Number):** 175




In this graph:

- Blue dots represent data from November, December, January, and February. The Colder Months.
- Red dots represent data from May to August. The Warmer months.
- Green dots represent data from March, April, September, and October. The months that are not too cold nor too hot.

**Explanations of Regression Model Summary:**

1. **Dependent Variable**: The variable we're trying to predict is "Rain."
2. **R-squared**: About 49.2% of the variability in Rain is explained by the model, which suggests a moderate explanatory power.
3. **Adjusted R-squared**: Same as R-squared, indicating that we have only one predictor.
4. **F-statistic**: A very high value (1383), and the p-value for the F-statistic is practically zero, suggesting the model is statistically significant.
5. **Observations**: We have 1428 observations in the dataset.
6. **AIC and BIC**: These criteria for model fit indicate lower values are better; however, without a comparison model, these are less informative.

**Coefficient Estimates**

1. **Intercept**: -762.45
2. **Temperature**: 38.4884

This means that the model predicts that for each 1-unit increase in temperature, rain increases by approximately 38.49 units.

**Hypothesis Testing**

1. **Intercept and Temperature p-values**: Both are practically zero, indicating they are statistically significant predictors.
2. **t-statistics**: Both the intercept and the temperature have high t-values, reinforcing their significance.

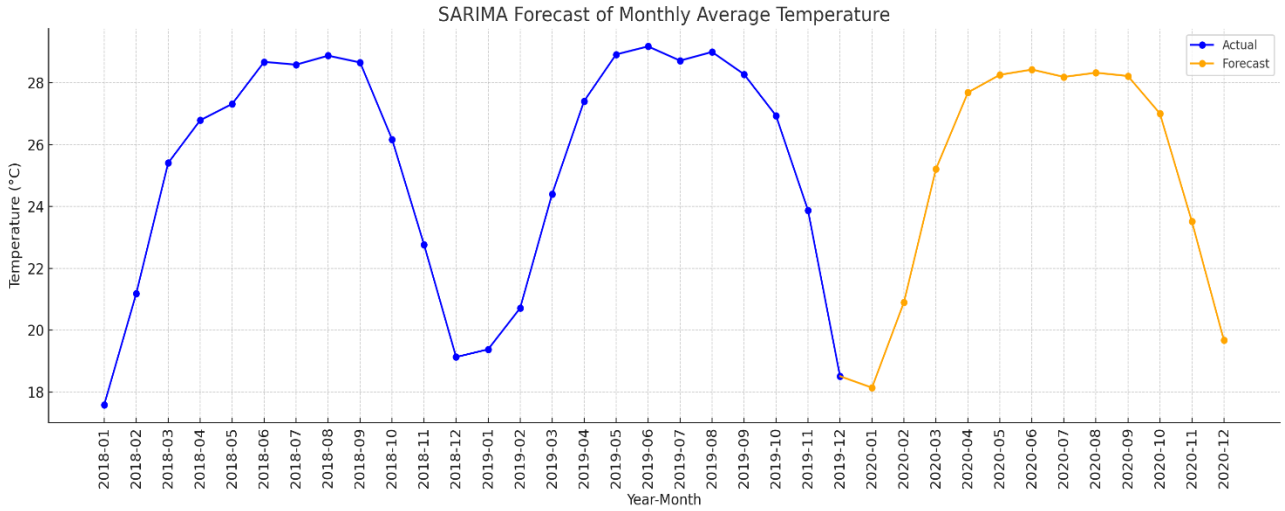
**Confidence Intervals**

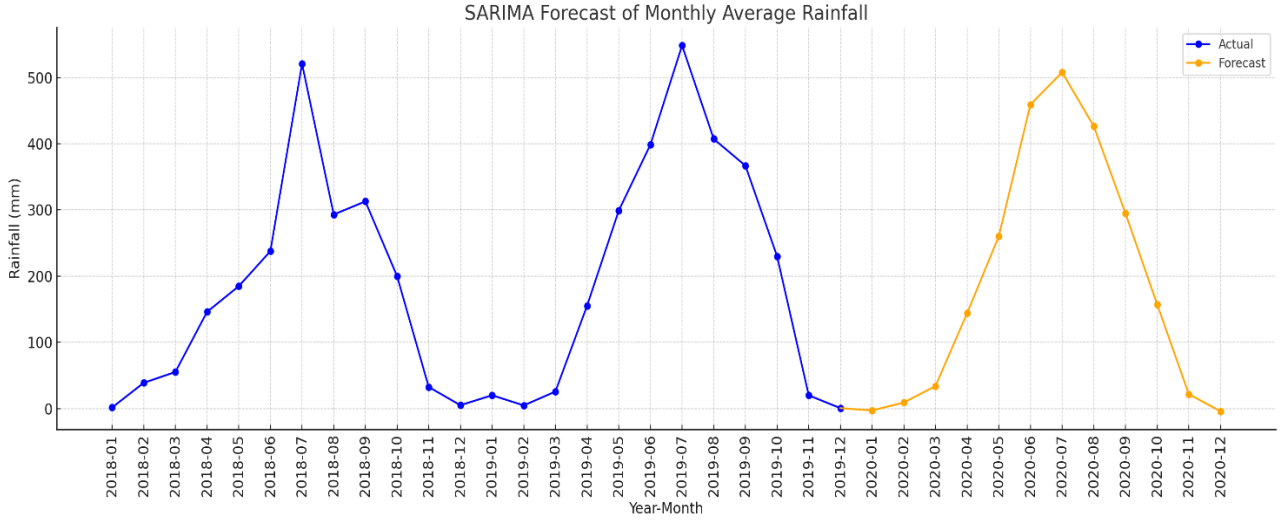
- The 95% confidence intervals for the intercept and temperature coefficients don't contain zero, which is another sign that they are statistically significant.

**Residual Diagnostics**

1. **Omnibus and Jarque-Bera Tests**: Indicate that the residuals are not normally distributed.
2. **Durbin-Watson**: Indicates that there is some evidence of positive autocorrelation, but it's not severe.
3. **Condition Number**: 175, which is not extremely high, suggesting that multicollinearity is likely not a concern.

**Forecasting 2020 every month Temperature and Rainfall using Seasonal ARIMA model (SARIMA):**

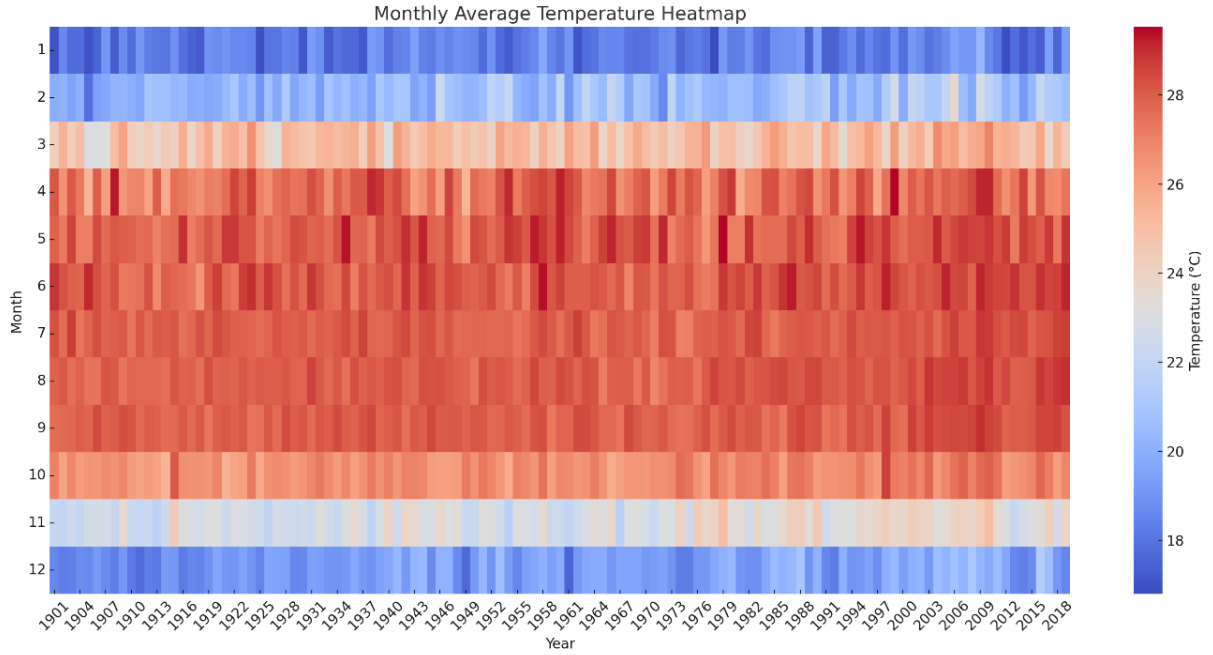


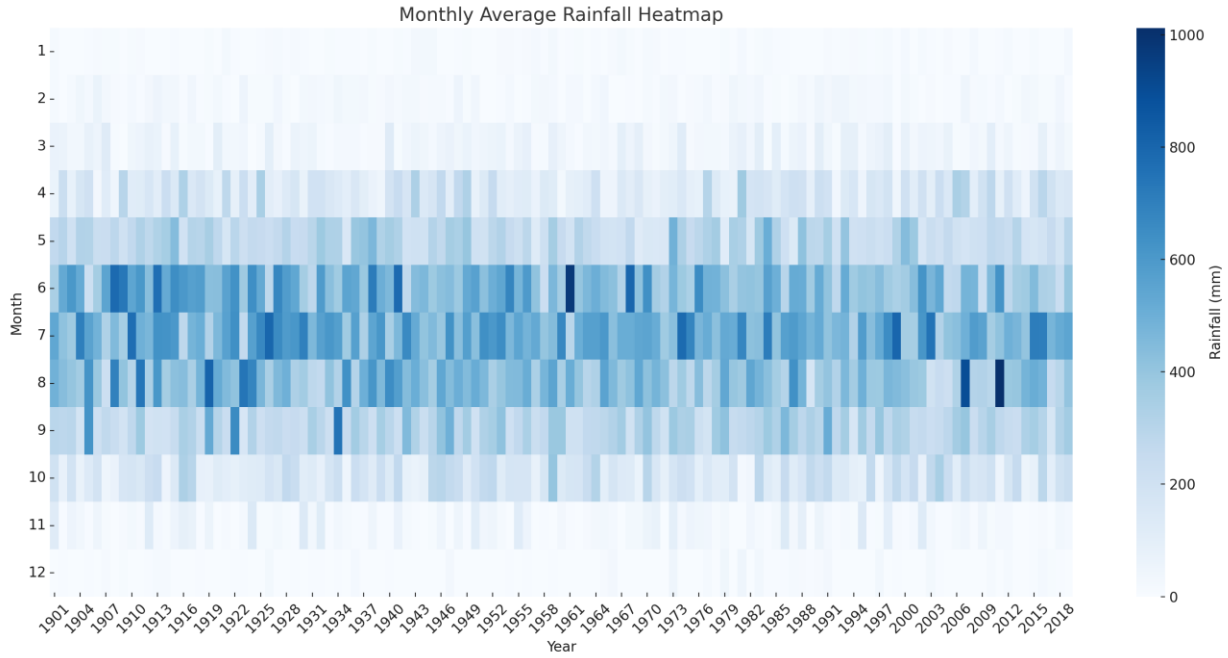

Using the data from 1901 to 2019 every month's average temperature and rainfall, we implemented SARIMA model to forecast 2019's every month's average temperature and rainfall.

## Heatmap of Temperature and Rainfall:

Y axis represents the months and X axis Represents the Years




Darker red means higher temperatures and darker blue means lower temperatures.



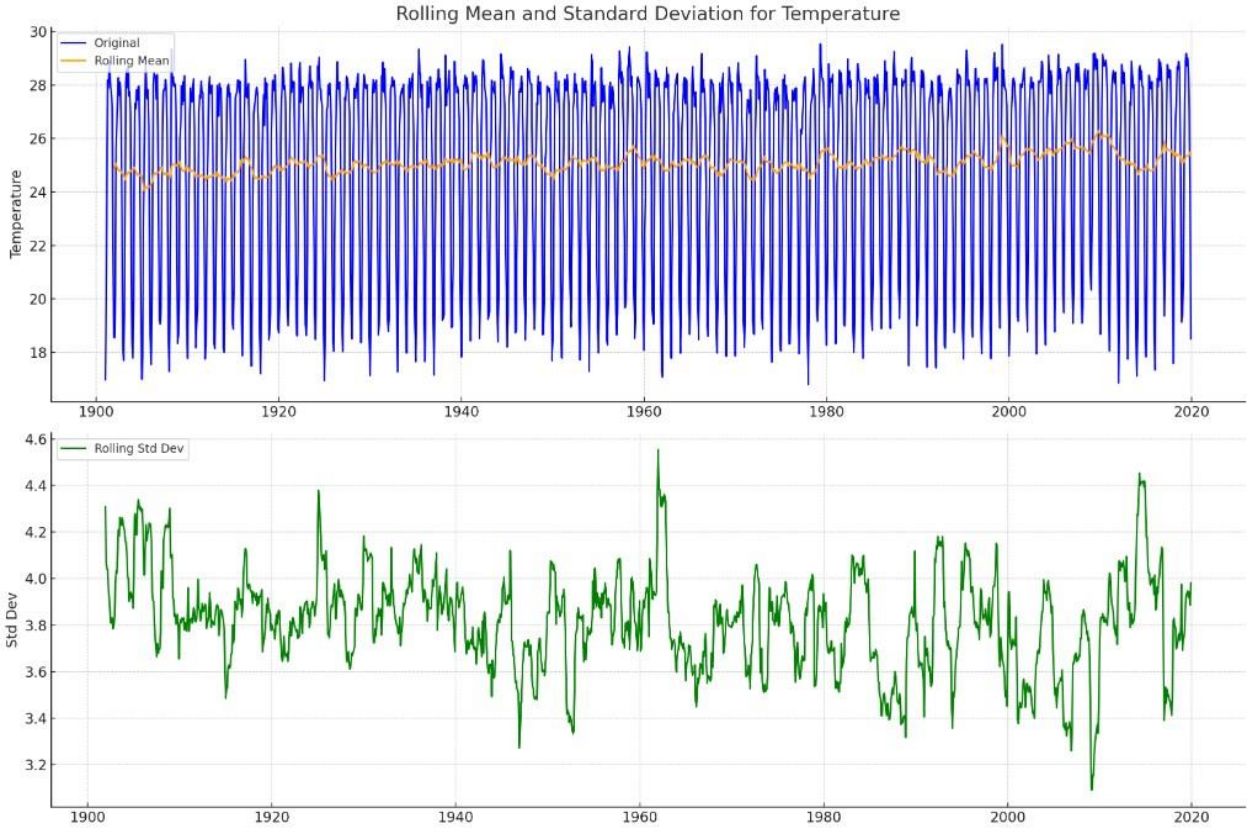

Darker blue means higher rainfall.

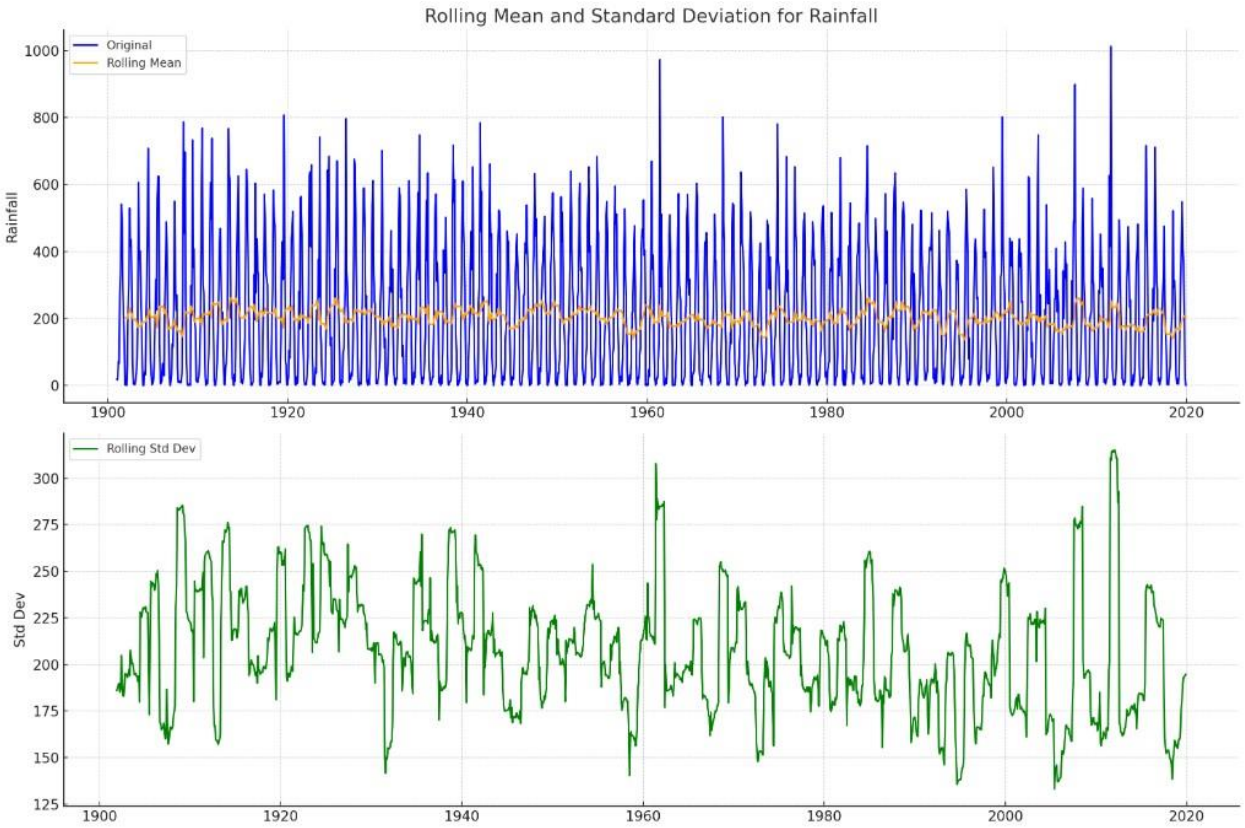
## Rolling Statistics

The rolling statistics are plotted above for both temperature and rainfall. In each case, the following are displayed:

1. **Original Data (Blue):** The original time series data.
2. **Rolling Mean (Orange):** The rolling mean calculated with a window of 12 months, which smooths out the data and helps identify the trend.
3. **Rolling Standard Deviation (Green):** The rolling standard deviation calculated with the same window, which provides insights into the variability of the data.

We can observe that the rolling mean and standard deviation are not constant over time for both temperature and rainfall, indicating non-stationarity in the time series data.



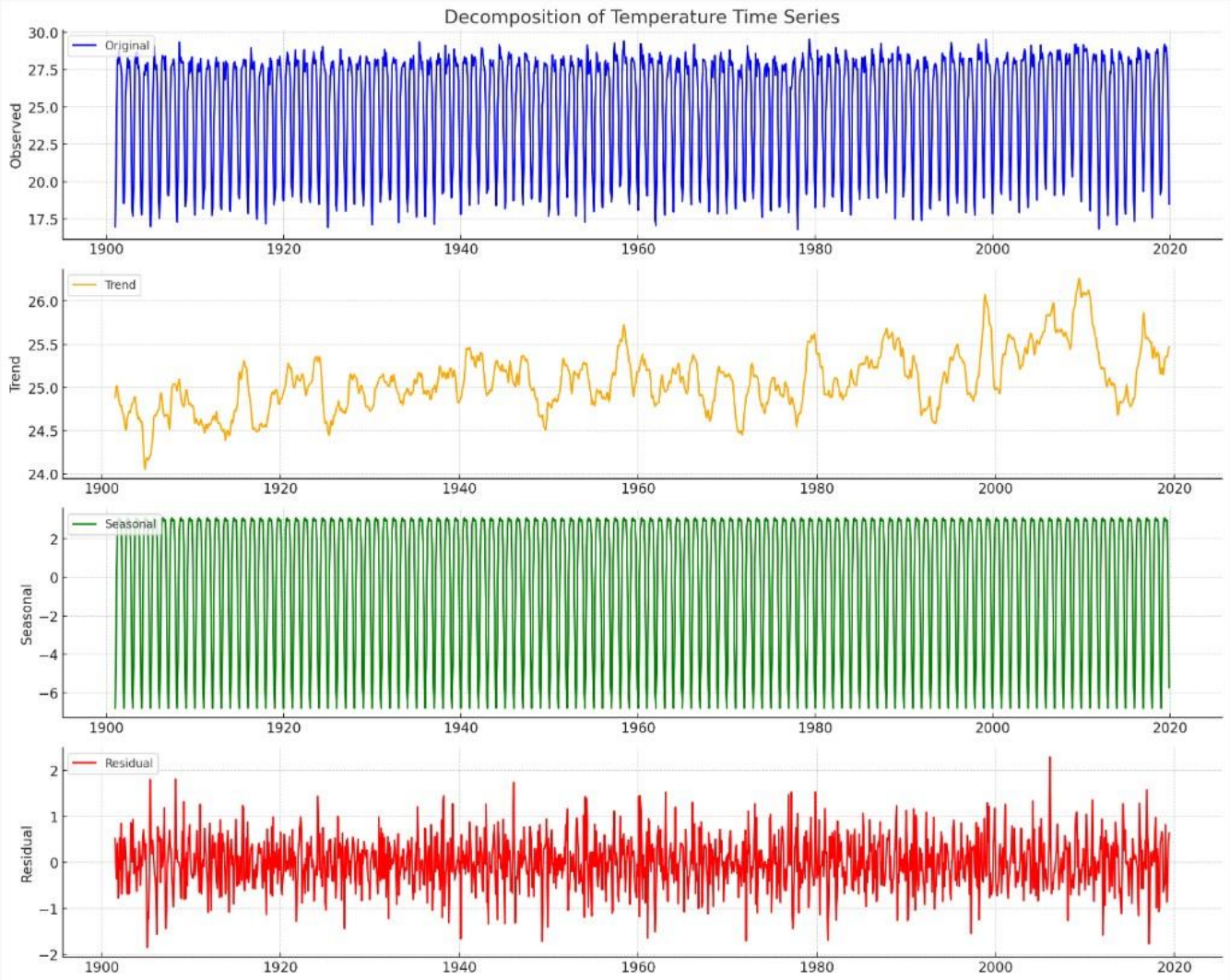

Rolling Mean and Standard Deviation for Rainfall
Original
Rolling Mean
Rainfall
1000
800
600
400
200
0
1900
1920
1940
1960
1980
2000
2020
Rolling Std Dev
Std Dev
300
275
250
225
200
175
150
125
1900
1920
1940
1960
1980
2000
2020

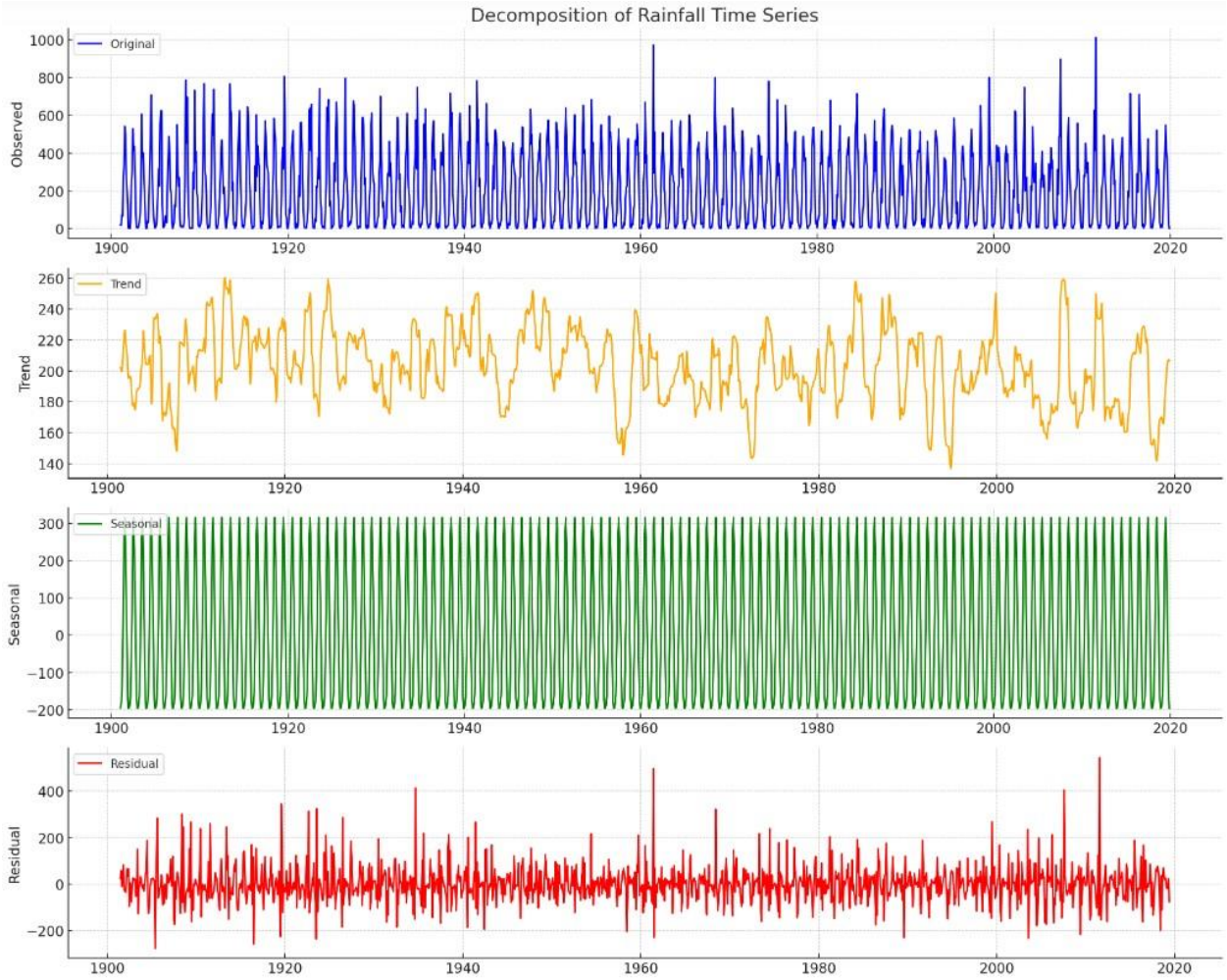
## Seasonal Decomposition

The seasonal decomposition of both the temperature and rainfall time series data is shown above. Each time series is decomposed into the following components:

1. **Observed (Blue)**: The original data.
2. **Trend (Orange)**: The underlying trend in the data.
3. **Seasonal (Green)**: The seasonal patterns.
4. **Residual (Red)**: The residual errors, representing what's left after removing the trend and seasonal components.

For both temperature and rainfall, the trend and seasonal patterns are evident.




Decomposition of Rainfall Time Series
Original
Observed
Trend
Trend
Seasonal
Seasonal
Residual
Residual
1900
1920
1940
1960
1980
2000
2020