Climate Resilience in Agricultural Production: A Time Series Analysis

Introduction:

Agricultural production is susceptible to the impacts of climate variability, posing significant challenges to food security and livelihoods worldwide. This project aims to analyze a real-world dataset spanning over a decade to understand the complex interactions between weather patterns, agricultural practices, and crop yields. By identifying key drivers of crop productivity and assessing climate resilience strategies, we aim to provide actionable insights for stakeholders to enhance sustainability and adaptability in agriculture.

Objectives:

1. **Identify Long-term Trends**: Analyze historical data to identify long-term trends and patterns in temperature, precipitation, and crop yields across different regions.

2. **Assess Seasonal Variability**: Explore seasonal variations in weather patterns and crop yields, considering factors such as planting and harvesting seasons.

3. **Evaluate Climate Impact:** Quantify the impact of weather variability on crop productivity, including the effects of extreme weather events such as droughts, floods, and heatwaves.

4. **Investigate Adaptation Strategies:** Assess the effectiveness of various adaptation strategies, including irrigation techniques, crop rotation, and pest management, in mitigating the adverse effects of climate change.

5. **Predict Future Scenarios**: Develop time series forecasting models to predict future crop yields under different climate scenarios, enabling stakeholders to anticipate and plan for future challenges.

6. **Recommend Policy Interventions:** Provide evidence-based recommendations to policymakers, agricultural organizations, and farmers for implementing sustainable practices and policies to enhance climate resilience in agriculture.

**Methodology**

1**. Data Collection** Gather a comprehensive dataset containing monthly observations of weather variables (temperature, precipitation), crop yields, soil moisture, pest infestation, agricultural practices, and market prices across multiple regions.

2**. Data Preprocessing:** First read the data

**Missing Value :** There are 8 missing values in the Temperature (°C) column

**Handle Missing value**  :Fill the missing value with mean.

**Outliers detection:**

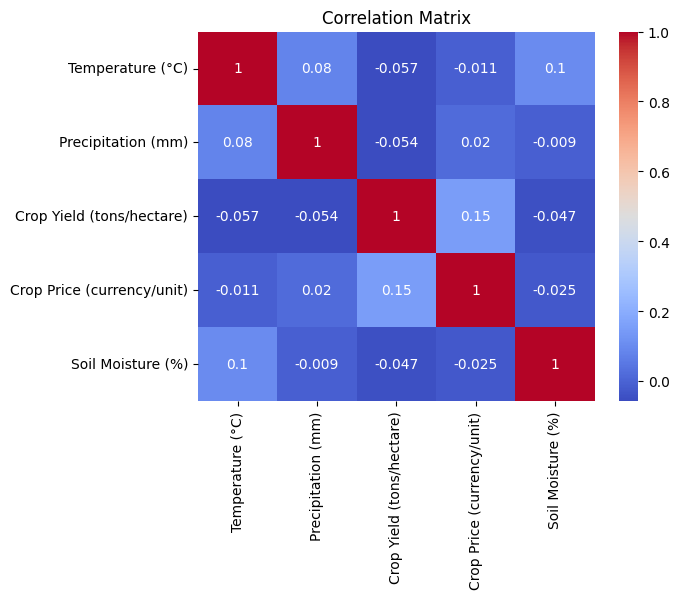
i)**using Z score**

ii) **using Box plot**: In Temperature (°C) there are some outliers

**Remove outliers:**

Remove the outliers In Temperature (°C) columns

Plot the correaltion matrix



**mean Crop Price (currency/unit) for each group (A, B, C)**

Region A 10.220707

Region B 10.797556

Region C 9.694191

So ,Region B has the highest crop Price for each group

**mean Crop Yield (tons/hectare) for each group (A, B, C)**

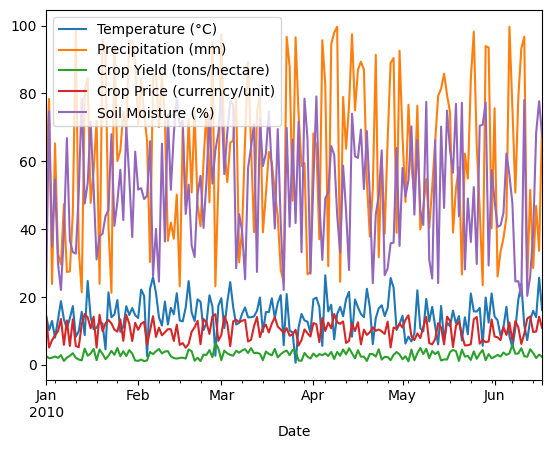
Region A 3.025097

Region B 2.834710

Region C 3.035678

Region C has highest Crop yield

**Plot the variables respect to date**



**Mean Agricultural Practice Crop Price (currency/unit) for each group Conventional ,Organic**

Conventional 9.792619

Organic 10.585264

Organic Agricultural Practice which has highest value.

**Mean Agricultural Practice crop Yield (tons/hectare) for each group Conventional ,Organic**

Conventional 2.972108

Organic 2.970469

**Conventional** Agricultural Practicewhich has Conventional crop Yield (tons/hectare)

**Where temerature is low and soil moisture is low there pest infaction cahmnce is high**

High 44.630506

Low 52.963807

Medium 50.055177

**Is there a significant difference Crop Price (currency/unit)in between the different Region ?**

ANOVA F-statistic: 2.1281291070746535

ANOVA p-value: 0.12231640567205836

The F-statistic indicates that there is some difference in crop prices among the regions, but the magnitude of this difference may not be very large relative to the variability within the regions.

The p-value (0.122) is greater than the typical significance level of 0.05. This suggests that there is not enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is not sufficient evidence to conclude that there are significant differences in crop prices among regions A, B, and C.

**Is there a significant difference Crop Yield (tons/hectare) in between the different Region**

**ANOVA F-statistic: 0.5186389795379699**

**ANOVA p-value: 0.5962975096353611**

The F-statistic is relatively low, indicating that the difference in crop yield between the regions is small compared to the variability within the regions.

The p-value (0.596) is greater than the typical significance level of 0.05. This suggests that there is not enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is not sufficient evidence to conclude that there are significant differences in crop yield between the regions.

Based on these results, we fail to reject the null hypothesis, indicating that there is no significant difference in crop yield between regions A, B, and C according to the ANOVA test. However, it's important to consider other factors and the context of the analysis when interpreting these results. Additionally, verifying the assumptions of ANOVA is crucial to ensure the validity of the analysis.

**Is there a significant difference Temperature (°C) in between the different Pest Infestation**

**ANOVA F-statistic: 4.310155442701172**

**ANOVA p-value: 0.014975513085669093**

The F-statistic is relatively high, suggesting that there may be a notable difference in temperature between the different levels of pest infestation.

The p-value (0.015) is less than the typical significance level of 0.05. This indicates that there is enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is sufficient evidence to conclude that there are significant differences in temperature between the different levels of pest infestation.

**Is there a significant difference Soil Moisture (%) in between the different Pest Infestation**

**ANOVA F-statistic: 2.264609115748912**

**ANOVA p-value: 0.10709008279554443**

The F-statistic is moderate, suggesting that there might be some difference in soil moisture between the different levels of pest infestation, but it's not very substantial compared to the variation within the groups.

The p-value (0.107) is greater than the typical significance level of 0.05. This indicates that there is not enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is not sufficient evidence to conclude that there are significant differences in soil moisture between the different levels of pest infestation.

**Ttest\_indResult(statistic=0.009249539429829291, pvalue=0.9926311495240384) in between Agricultural Practice andCrop Yield (tons/hectare)**

The p-value (0.9926) is much greater than the typical significance level of 0.05. Therefore, there is not enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is not sufficient evidence to conclude that there is a significant difference in crop yield between different agricultural practices.

**Ttest\_indResult(statistic=-1.8077414893568624, pvalue=0.07245708745668966) Agricultural Practice andCrop Price (currency/unit)**

The p-value (0.0725) is greater than the typical significance level of 0.05. Therefore, there is not enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is not sufficient evidence to conclude that there is a significant difference in crop prices between different agricultural practices.

**Correlation between Crop Yield (tons/hectare) and Crop Price (currency/unit): 0.1529448773813911**

The correlation coefficient of 0.1529 suggests a relatively weak positive correlation between Crop Yield and Crop Price. This means that there is a slight tendency for higher crop yields to be associated with higher crop prices, but the relationship is not very strong.

**Chi-square statistic: 336.0**

**Chi-square p-value: 0.45897617528775 between crop price and region**

The chi-square statistic is 336.0, indicating a substantial discrepancy between the observed frequencies of Crop Price and Region compared to what would be expected if there were no association between these variables.

However, the p-value associated with the chi-square statistic is 0.4589. This value is greater than the typical significance level of 0.05, suggesting that there is not enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is not sufficient evidence to conclude that there is a significant association between Crop Price and Region.

**Kruskal-Wallis H-statistic: 4.006306095892724**

**Kruskal-Wallis p-value: 0.1349092366253987 between region A,b,c and agricultural practice**

The Kruskal-Wallis H-statistic is 4.0063, suggesting some difference in ranks among the groups (Region A, B, C and Agricultural Practice).

However, the p-value associated with the Kruskal-Wallis test is 0.1349. This value is greater than the typical significance level of 0.05, suggesting that there is not enough evidence to reject the null hypothesis at the 0.05 significance level. In other words, there is not sufficient evidence to conclude that there is a significant difference in distributions between Region and Agricultural Practice.

**95% Confidence Interval for the mean Crop Price (currency/unit): (9.762740497967854, 10.63401564306768)**

With 95% confidence, we estimate that the true mean Crop Price (currency/unit) lies within the interval (9.762740497967854, 10.63401564306768). In other words, if we were to take multiple samples and calculate 95% confidence intervals for the mean Crop Price from each sample, we would expect the true mean to fall within this interval for approximately 95% of those samples.

This confidence interval provides a range estimate for the true mean Crop Price, giving us an idea of the precision of our sample estimate. It's a useful measure in statistical inference for understanding the uncertainty associated with estimating population parameters from sample

**Linear Regression Slope: -0.6926505829277767**

**Linear Regression Intercept: 53.44114768119642**

**Linear Regression R-squared value: 0.0022347847747832203**

**Linear Regression p-value: 0.5428556768396888**

The negative slope (-0.6927) suggests a decreasing relationship between the independent and dependent variables.

The intercept (53.4411) represents the value of the dependent variable when the independent variable is zero, but in many cases, this value may not be meaningful depending on the context.

The low R-squared value (0.0022) indicates that the independent variable explains only a very small proportion of the variance in the dependent variable.

The high p-value (0.5429) suggests that the relationship between the independent and dependent variables is not statistically significant at the typical significance level of 0.05.

**Correlation between Temperature (°C) and crop yield: -0.06, p-value: 0.4613**

**Correlation between Precipitation (mm) and crop yield: -0.05, p-value: 0.4849**

**Correlation between Soil Moisture (%) and crop yield: -0.05, p-value: 0.5429**

**T-test between Conventional and other practices: t-statistic: 0.01, p-value: 0.9926**

**T-test between Organic and other practices: t-statistic: -0.01, p-value: 0.9926**

Temperature (°C) and Crop Yield: There is a weak negative correlation between temperature and crop yield, but it is not statistically significant (p-value > 0.05). This suggests that there may be a slight tendency for crop yield to decrease with increasing temperature, but this relationship is not strong enough to be considered significant.

Precipitation (mm) and Crop Yield: Similarly, there is a weak negative correlation between precipitation and crop yield, but it is also not statistically significant (p-value > 0.05). This indicates that there may be a slight tendency for crop yield to decrease with increasing precipitation, but again, this relationship is not strong enough to be considered significant.

Soil Moisture (%) and Crop Yield: The correlation between soil moisture and crop yield is also weak and not statistically significant (p-value > 0.05). This suggests that there is no significant relationship between soil moisture levels and crop yield in this analysis.

T-tests between Different Practices: Both t-tests comparing conventional and organic practices with other practices show very small t-statistics close to zero, indicating very little difference between these practices in terms of crop yield. Furthermore, the p-values for both tests are very high (close to 1), indicating that there is no statistically significant difference in crop yield between conventional, organic, and other practices.

**Training Random Forest...**

**Random Forest - MAE: 0.96, MSE: 1.45**

**Training Gradient Boosting...**

**Gradient Boosting - MAE: 1.04, MSE: 1.83**

**Training Neural Network...**

**Neural Network - MAE: 1.81, MSE: 4.21**

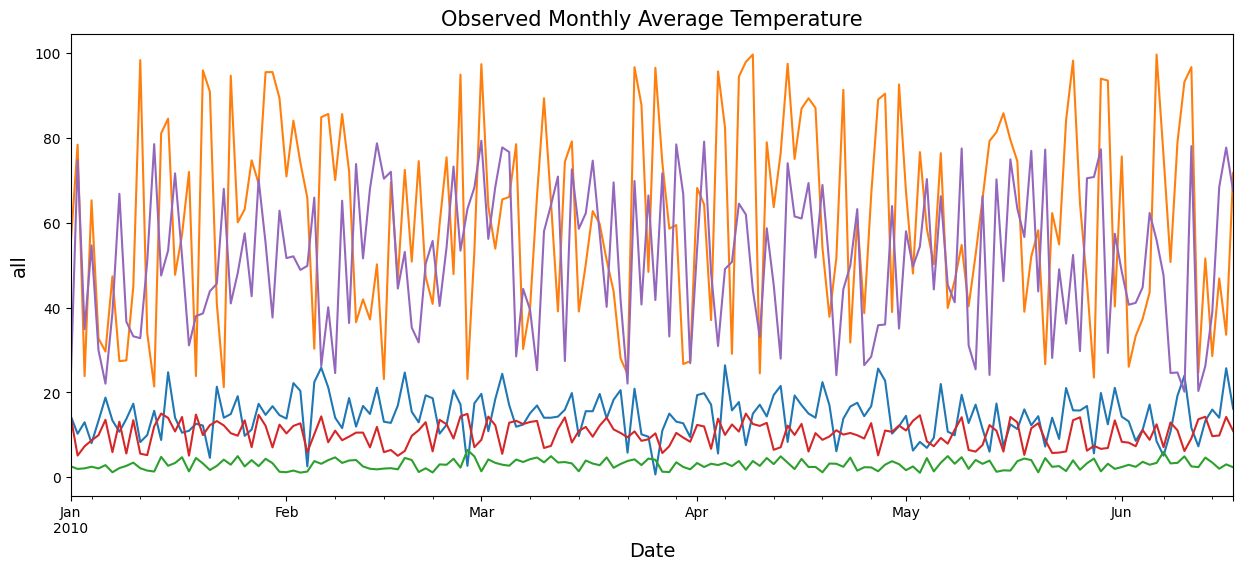
The Gradient Boosting model has slightly higher MAE and MSE compared to Random Forest but lower than the Neural Network.

The Neural Network model has the highest MAE and MSE among the three models, indicating it may have more prediction errors compared to the other two models.

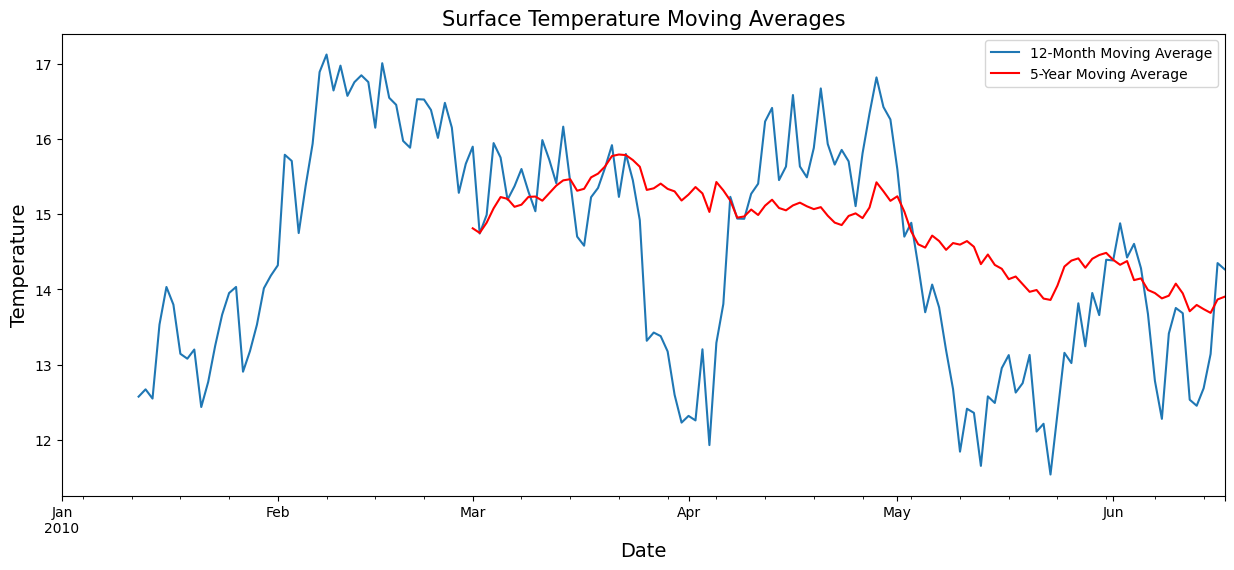
The **Random Forest** model seems to be the most favorable choice.

**The value of 0.148 indicates a very weak positive correlation. This means that as one variable (crop price) tends to increase, the other variable (crop yield) also tends to increase, but the relationship is very weak.**

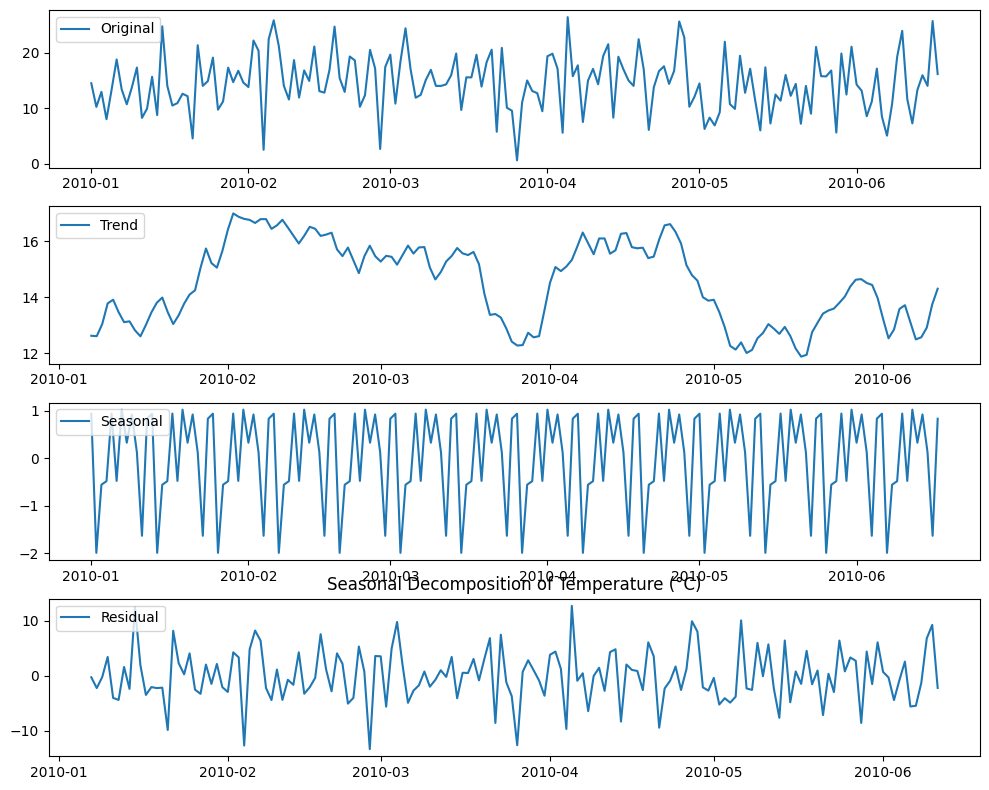
**Moving Average**

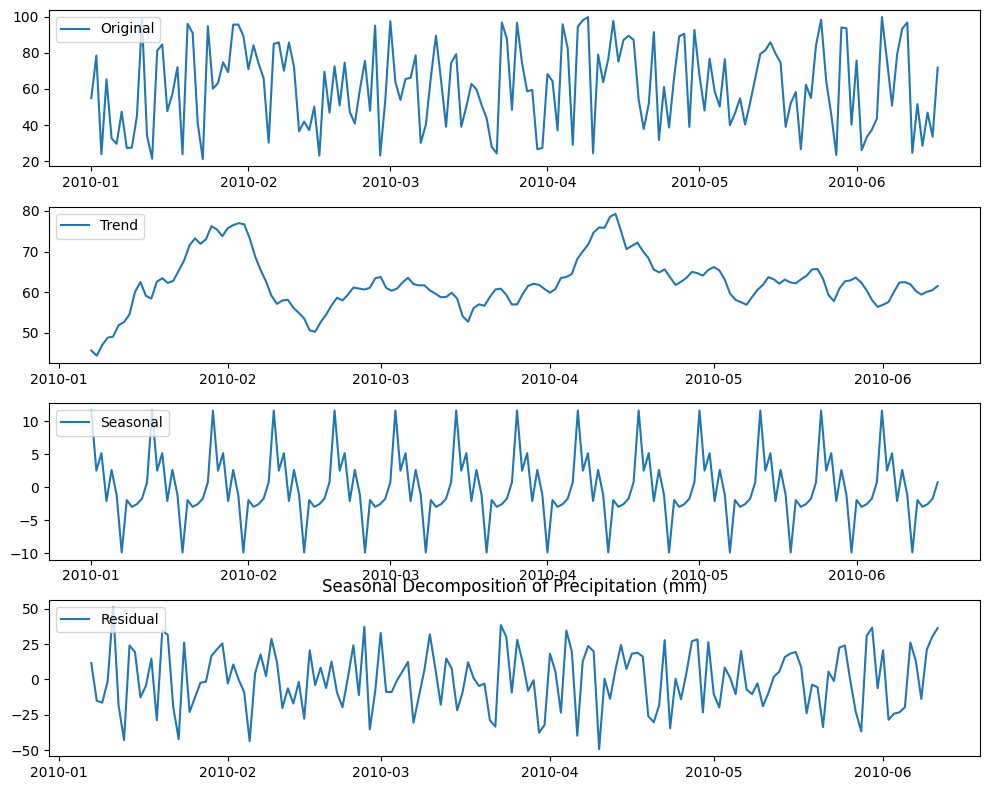
****

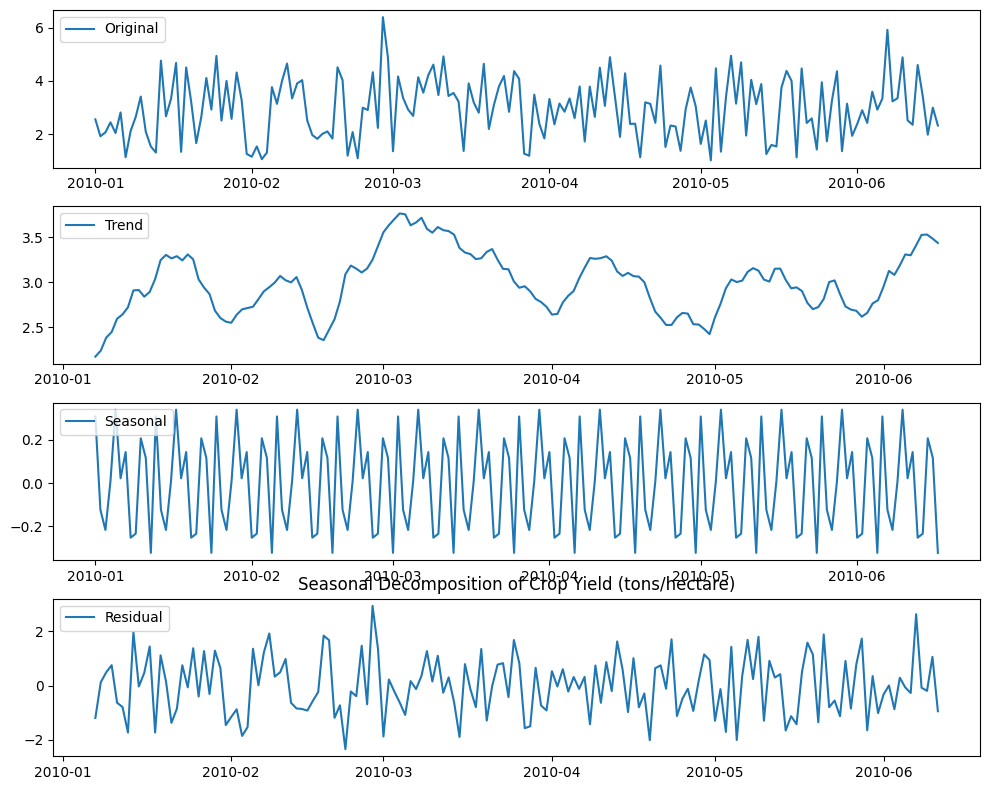
**Temperature (°C) moving average**

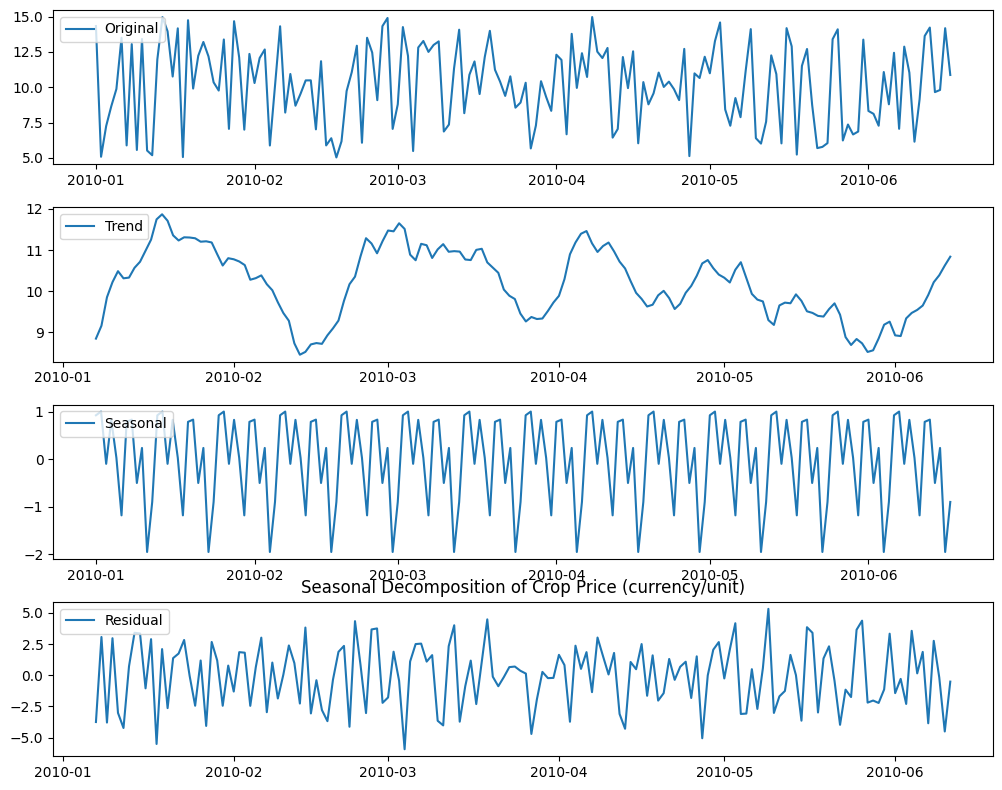
****

**find the trend,seasonal,residual on the Crop Yield (tons/hectare)**

****

****

****

****

**Checking for Stationarity of Data:**

A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. We need to ensure that the time series is stationary before using it to train a model.

**Augmented Dickey-Fuller (ADF) Test:**

Augmented Dickey-Fuller (ADF) statistics is one of the more widely used statistical test to check whether the time series is stationary or non-stationary.

It uses an autoregressive model and optimizes an information criterion across multiple different lag values.

Null Hypothesis: Series is not stationary.

Alternate Hypothesis: Series is stationary.

The p-value is less than the threshold of 0.05, and the ADF statistic is close to the critical values. Therefore, the time series is indeed stationary.