PREDICTING FOOD PRICES IN KENYA

Table of Contents

Executive Summary

Food security is a cornerstone of Kenya's prosperity, relying on the agricultural sector contributing significantly to export earnings and supporting most of the population. This sector's stability is essential in ensuring a consistent food supply. Precise food price forecasting plays a crucial role in sustaining food security and aiding stakeholders within the sector.

Our research addresses Kenyan farmers and cash-crop retailers, providing predictive insights into future commodity prices. Key objectives include identifying price trends, segmenting regions, developing predictive models, and offering recommendations.

We adopted a comprehensive methodology, including data selection, cleaning, exploratory data analysis, crafting predictive models (SARIMA and LSTM), and drawing conclusions. Data was sourced from the World Food Programme (WFP) and integrated with external datasets to augment analysis and accuracy.

Our dataset spans various Kenyan regions, consisting of 14 columns. Notable findings include a wide price range, the impact of inflation rates, and regional preferences for food categories and commodities. Kenyans show a strong inclination toward consuming food items from categories such as cereals and tubers, pulses and nuts, and vegetables and fruits.

We applied data preprocessing, such as seasonal feature creation, cluster analysis, and feature selection before starting the modeling process. The LSTM model significantly outperformed the SARIMA model, becoming the preferred choice for predicting future commodity prices.

For maize, the LSTM model achieved an RMSE of 9.40 units, while for beans, it achieved an RMSE of 7.75 units, representing substantial improvements over SARIMA model which achieved 12.17 and 24.36units for maize and beans respectively.

The Baseline LSTM model was used to forecast maize and bean prices for the next 12 months as it also outperformed the hyperparameter tuned LSTM. This has offered valuable insights for stakeholders that can be used in decision-making in the agricultural sector.

In conclusion, our research contributes to the agricultural sector stability, enhances food security, and aids the prosperity of all stakeholders involved.

1.0 Introduction

Food security stands as the bedrock of a stable and prosperous society in Kenya. The nation's capacity to furnish its citizens with consistent access to safe and nourishing sustenance stands as a pivotal gauge of its overall well-being.

The agricultural sector takes center stage, contributing to 65% of the nation's export earnings and serving as the primary source of livelihood for over 80% of Kenya's population. Furthermore, this sector plays a pivotal role in enhancing nutrition by producing a diverse range of nutrient-rich foods, as underscored in a 2023 FAO article titled "Kenya at a Glance" [source: https://www.fao.org/kenya/fao-in-kenya/kenya-at-a-glance/en/].

Crop cultivation emerges as a driving force behind Kenya's economic advancement, offering sustenance, economic returns, employment opportunities, and a means to reduce foreign currency outlays through decreased reliance on imports.

According to the Kenya Food Directorate [source: https://food.agricultureauthority.go.ke/index.php/sectors/overview], as of September 2023, the food crops sub-sector contributes roughly 33% of the total agricultural GDP and plays a significant role in the nation's agricultural output.

Given the far-reaching impact of the agricultural sector on the economy, it becomes imperative to ensure the availability of precise food price forecasting and market analysis to serve the diverse stakeholders within this sector.

2.0 Research Problem and Objectives

The research targets Kenyan farmers and cash-crop retailers in the different Kenyan provinces. The research question for this study aims to address the following:

"How can the application of advanced data science methodologies contribute to meeting the critical needs of Kenyan farmers and retailers by providing predictive insights into future commodity prices, thus facilitating informed decision-making for planning, sales, and profit maximization?

In order to support this research study, below objectives were formulated and worked on by SokoSmart Analysts.

1. To identify key patterns and trends and relationships in the data:
    a. To identify how crop prices trends over time.
    b. To segment regions based on crops and prices i.e., coastal, inland, rural and urban.
    c. To analyze the different food categories and commodities.
    d. To determine Geospatial Analysis based of the latitude and longitude data. i.e., heatmaps
    e. To investigate potential correlations between food prices and other variables, such as weather conditions (rainfall) and other socio-economic factors.
2. To develop a predictive robust timeseries model that predicts the future prices of key agricultural commodities in Kenya.
3. To create a Market Basket Analysis for Retailers
4. To deploy a crop pricing model.
5. To provide recommendations on the outcomes of the project to our stakeholders

3.0 Methodology

SokoSmart employed the following approach to conduct the research:

1. Data Selection
2. Data Understanding and Cleaning
3. Conducting Exploratory Data Analysis to unearth patterns and revelations, including:
   - Univariate Analysis
   - Bivariate Analysis
   - Multivariate Analysis
4. Investigating and Crafting Models, involving:
   - Data Preprocessing and Feature Engineering
   - Model Development and Fine-tuning
   - Model Selection
5. Drawing Conclusions
6. Offering Recommendations

3.1 Data Selection

A dataset for the research was obtained from the World food programme (WFP) Price Database which can accessed on this link: https://data.humdata.org/dataset/wfp-food-prices-for-kenya.

The dataset covers a time period of 18 years from January 15, 2006 to February 15 2024.

Kenyan sources of the data for the WFP include the below: Arid Lands Resource Management Project (ALRMP), Energy Regulatory Commission, Energy Regulatory Commission (ERC), Energy and Petroleum Regulatory Authority (EPRA), FPMA, Kenya National Bureau of Statistics (KNBS), MOA, Ministry of Agriculture, Ministry of Agriculture (MOA), National Drought Management Authority (NDMA), Regional Agricultural Trade Intelligence Network via FAO: GIEWS, State Department of Agriculture, WFP, World Food Programme (WFP) Monitoring.

It is paramount to note that, external datasets covering inflation rates and weather patterns were included into the data world food programme dataset. The inflation rates was sourced from the Central Bank of Kenya https://www.centralbank.go.ke/inflation-rates/ while the weather patterns https://dataviz.vam.wfp.org/version2/climate-explorer

3.2 Data Understanding

Our primary dataset provides information on food commodity pricing across various regions in Kenya. It encompasses the following regions: *Coast, Eastern, Nairobi, North Eastern, Nyanza, Rift Valley, and Central* Kenya.

This dataset comprises 14 columns and 18,578 rows. The columns in the dataset are labeled as *date, admin1, admin2, market, latitude, longitude, category, commodity, unit, priceflag, pricetype, currency, price, and usdprice.*

There are a total of 62 distinct markets identified in the data, including names like *Mombasa, Kitui, Marsabit, Nairobi, Mandera, Kisumu, Lodwar (Turkana), Eldoret town (Uasin Gishu), Nakuru, Kilifi, Hola (Tana River), Garissa, Marigat (Baringo), Kajiado, Karatina (Nyeri), Vanga (Kwale), Kitui town (Kitui),* *Makueni, Wote town (Makueni), Kitengela (Kajiado), Garissa town (Garissa), Takaba (Mandera), Marigat town (Baringo), Illbissil Food Market (Kajiado), Wakulima (Nakuru), Lomut (West Pokot), Kongowea (Mombasa), Tala Centre Market (Machakos), Kangemi (Nairobi), Kibuye (Kisumu), Makutano (West Pokot), Kathonzweni (Makueni), Kaanwa (Tharaka Nithi), Wakulima (Nairobi), Dagahaley (Daadab), Bangladesh (Mombasa), Kalahari (Mombasa), Shonda (Mombasa), Kawangware (Nairobi), Kibra (Nairobi), Mathare (Nairobi), Mukuru (Nairobi), Hagadera (Daadab), Kalobeyei (Village 1), Kalobeyei (Village 2), Kalobeyei (Village 3), Isiolo town, Dandora (Nairobi), Wajir town, Ethiopia (Kakuma), HongKong (Kakuma), Kakuma 2, Kakuma 3, Kakuma 4, Lodwar town, Mogadishu (Kakuma), Junda (Mombasa), Moroto (Mombasa), Dadaab town, IFO (Daadab), Kisumu Ndogo (Mombasa), and Marsabit town.*

The dataset classifies food items into 9 general categories, which include *cereals and tubers, pulses and nuts, milk and dairy, oil and fats, non-food, meat, fish and eggs, miscellaneous food, and vegetables and fruits*. In total, there are 46 different commodities distributed among these categories.

Additionally, the dataset includes both Kenyan prices and USD prices within its columns, and there are other columns that provide further details regarding the type of price, such as whether it is wholesale or retail, actual or forecast.

3.3 Data Merging & Cleaning

To prepare for a more in-depth data analysis, the process of data cleaning and prepping involved the following steps:

- Integration: The inflation dataset from the Central Bank of Kenya and rainfall data was integrated with the primary WFP dataset. This merging was achieved by converting the "Date" column in both datasets into a datetime format and then setting them as the index columns for merging.
- Column Renaming: Columns were renamed to accurately represent the data. "Admin1" was relabeled as "Province," and "Admin2" was changed to "County."
- Elimination of Redundant Columns: Unnecessary columns were removed from the dataset to streamline and simplify the data structure.
- Removal of Zero-Value Entries: Entries with columns containing zero values were purged from the merged dataframe, thereby reducing the number of rows from 18,577 to 15,940 entries.

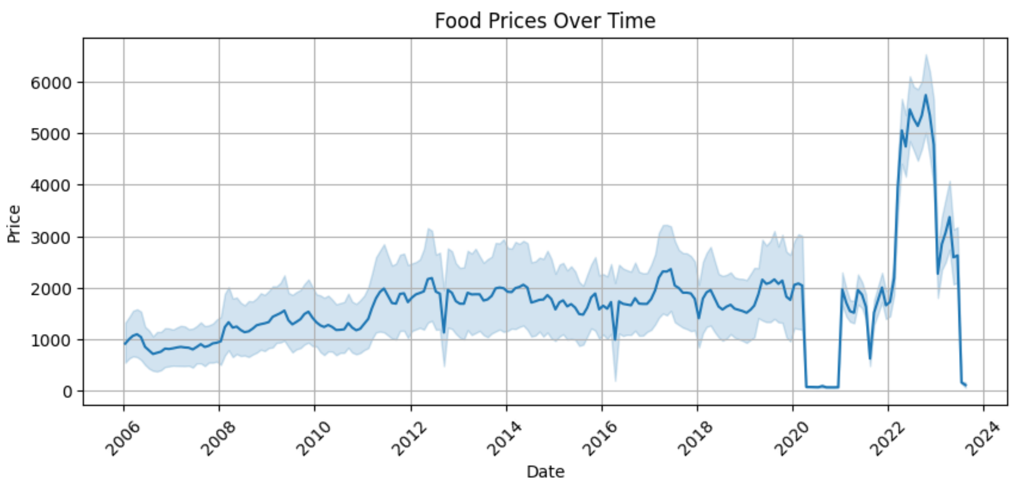
3.4 Exploratory Data Analysis

3.4.1 Key Findings: Univariate and Bivariate Analysis

1. **Commodity Pricing**:

In Kenya, the prices of most commodities span a wide range, but generally range between KES 57.20 and KES 3,000. On average, these commodities are priced at approximately KES 1,969.16. Notably, the standard deviation is approximately 2,996.66, indicating a substantial level of price variation and dispersion within the dataset. The lowest-priced commodity observed was valued at KES 5.00, while the highest reached KES 19,800.

Figure 1: Food prices over time 2006-2024

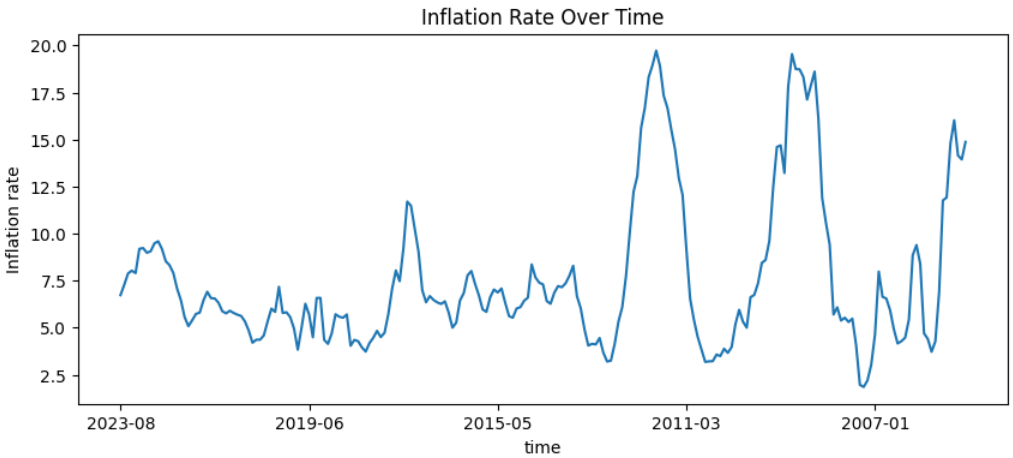


Food prices over time show there is a gradual upward trend in prices from 2007 to 2020, marked by occasional dips, followed by erratic price hikes in 2022 and 2023. Over the course of 14 years, starting from 2006, the prices have shown remarkable growth, transitioning from less than KES 1000 to well over KES 2000, reflecting a 100% increase.

2. **Impact of Inflation Rate:**

Over the years, Kenya's inflation rate has exhibited fluctuations, which, in turn, have had an impact on food prices. The inflation rate typically falls within the range of 5.7% to an upper quartile of 7.76%. Although the overall inflation rate has maintained a relatively stable average of 7.0%, food prices have shown some variability over time.

The inflation rate serves as a metric for gauging the pace at which the cost-of-living changes in Kenya and its repercussions on the purchasing power of the Kenyan currency. In the context of the current dataset, the inflation rate has oscillated between a minimum of 1.85% and a maximum of 19.72%.



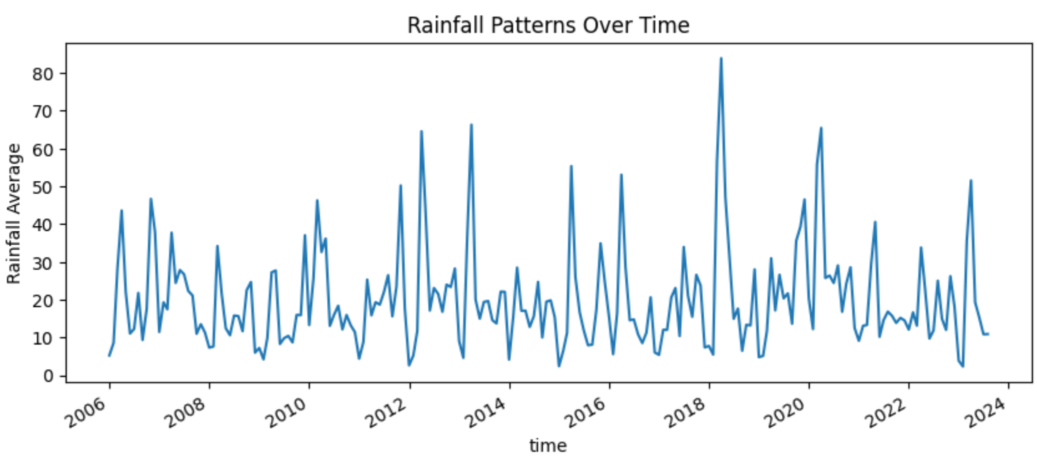
From 2013 to 2023, the inflation rate has shown a degree of stability, characterized by incremental shifts both upward and downward. Nevertheless, the years preceding 2013 are marked by heightened inflation volatility.

3. **Seasonal Influence:**

Seasonal pricing variations are influenced by different conditions and circumstances. In our case, we particularly focus on rainfall patterns which show the amount of precipitation in an area within a certain period of time.

The average annual rainfall has fluctuated within the range of 12.06mm to 22.49mm. The highest recorded rainfall level in the country was 83.90mm, while the lowest recorded rainfall was as low as 2.29mm.

The graph clearly illustrates the pronounced seasonality, showcasing the annual variability in rainfall. Notably, the period from 2018 to 2019 experienced a significant amount of rainfall, while in the years 2014 to 2015, the average rainfall levels were comparatively lower.



Maize prices typically exhibit a downward trend during the primary harvest season, which falls between September and December. However, they may experience an increase later in the year as available stocks diminish.

In the case of fruits and vegetables, their prices can exhibit substantial fluctuations in accordance with the seasons, often reaching their peak during off-season periods.

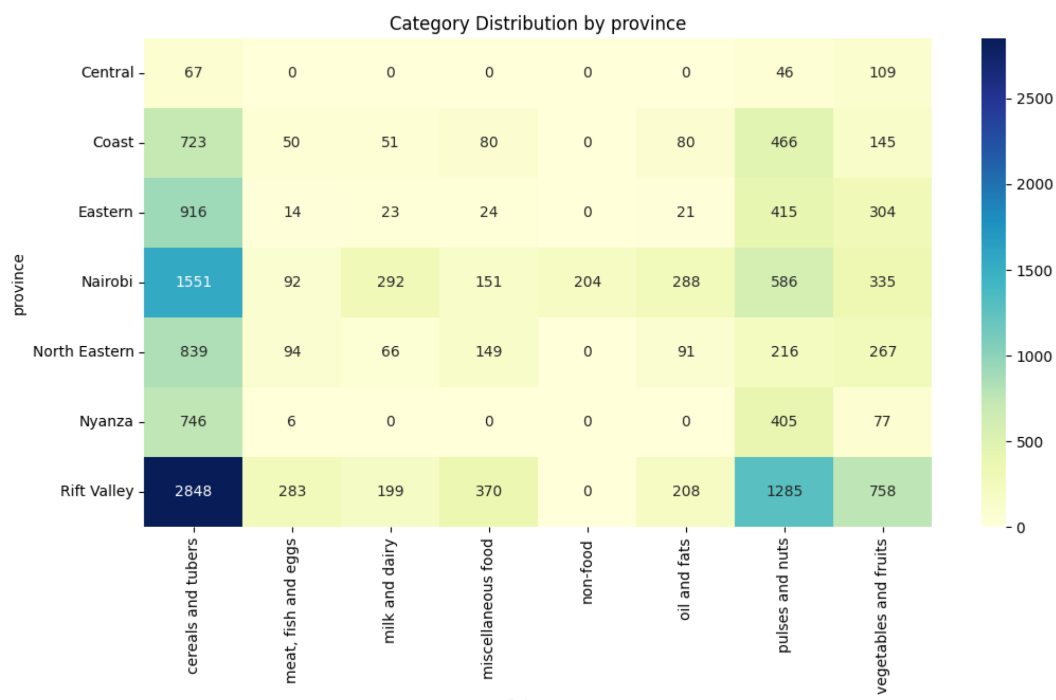
4. **Distribution by Provinces and Counties:**

The most frequently occurring provinces within the dataset are Rift Valley, Nairobi, and North Eastern. As for counties, the ones that appear most frequently are Nairobi, Turkana, Uasin Gishu, Garissa, and Mombasa.

Notably, the common markets in the data are Nairobi, Eldoret, Kisumu, Kitui, Mombasa, and Nakuru. This pattern can be attributed to the fact that these locations serve as major urban centers in Kenya and often function as de facto capital cities for their respective provinces.

Nairobi province stands out as having all commodities readily available, which can be attributed to its status as the capital city of Kenya. Other prominent provinces with a well-distributed presence of at least 7 out of 8 commodities are Coast, Eastern, North Eastern, and Rift Valley.

In contrast, provinces like Central and Nyanza primarily exhibit a high distribution in specific categories, focusing on cereals and tubers, pulses and nuts, and vegetables and fruits.

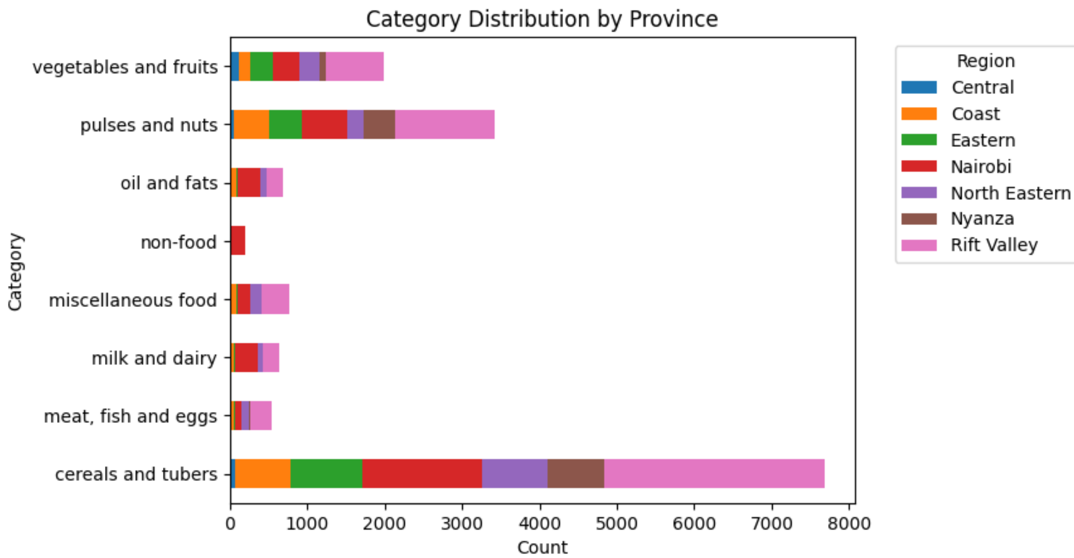


5. **Food Categories & commodities:**

The dataset highlights that Kenyans have a strong inclination toward consuming food items falling into categories such as cereals and tubers, pulses and nuts, as well as vegetables and fruits. This suggests that these particular food categories are readily available and affordable throughout the country.

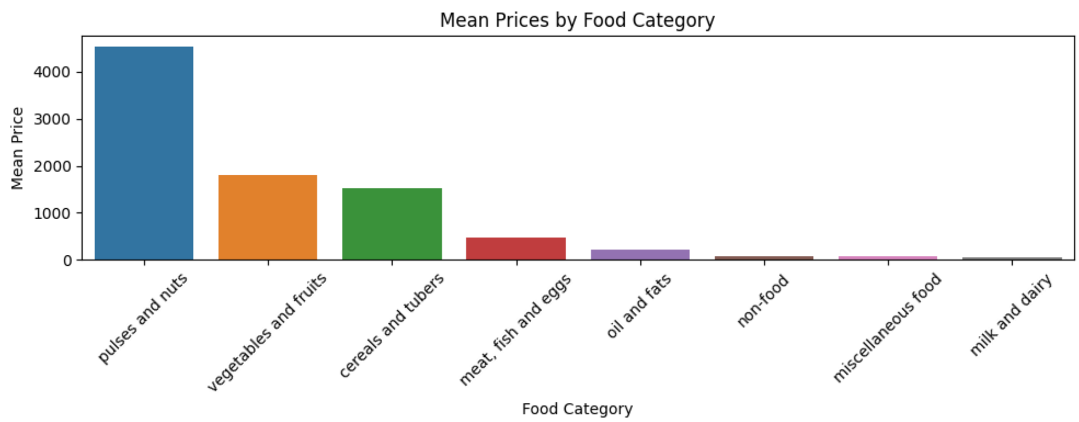
The dataset emphasizes the prominence of specific commodities, with maize, beans, potatoes, and sorghum being among the most prevalent. These commodities are widely consumed and commonly featured in the dataset.

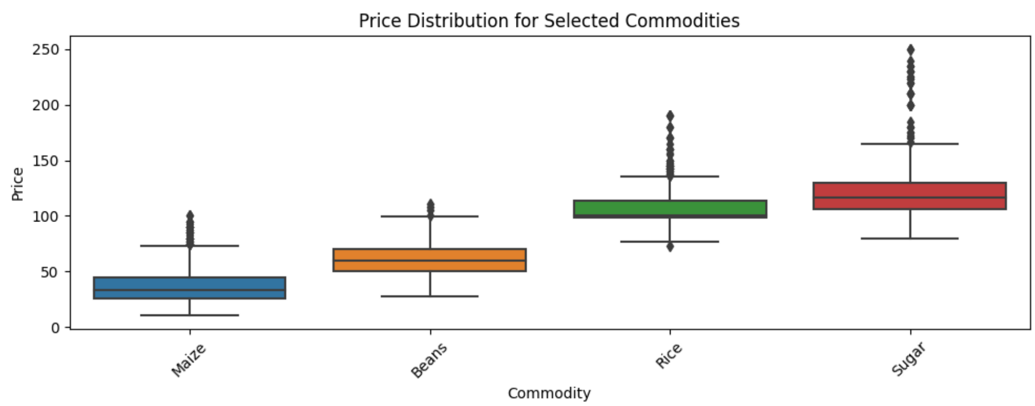
Cereals and tubers exhibit a strong presence across a diverse range of provinces, including Rift Valley, Nyanza, North Eastern, Nairobi, Eastern, Coast, and Central. Pulses and nuts closely follow in popularity, underlining the significance of these food categories in Kenyan diets.



The preliminary visualization unveils that among the food categories, "pulses and nuts" displayed the highest average prices, with "vegetables and fruits," "cereals and tubers," and "meat, fish, and eggs" following closely, among others.

Delving into individual food commodities, it became evident that sugar boasted a relatively high price, with rice trailing closely behind. It's worth mentioning that both sugar and rice exhibited a more significant number of outliers compared to other commodities. Conversely, maize registered the lowest price, with beans ranking closely behind.

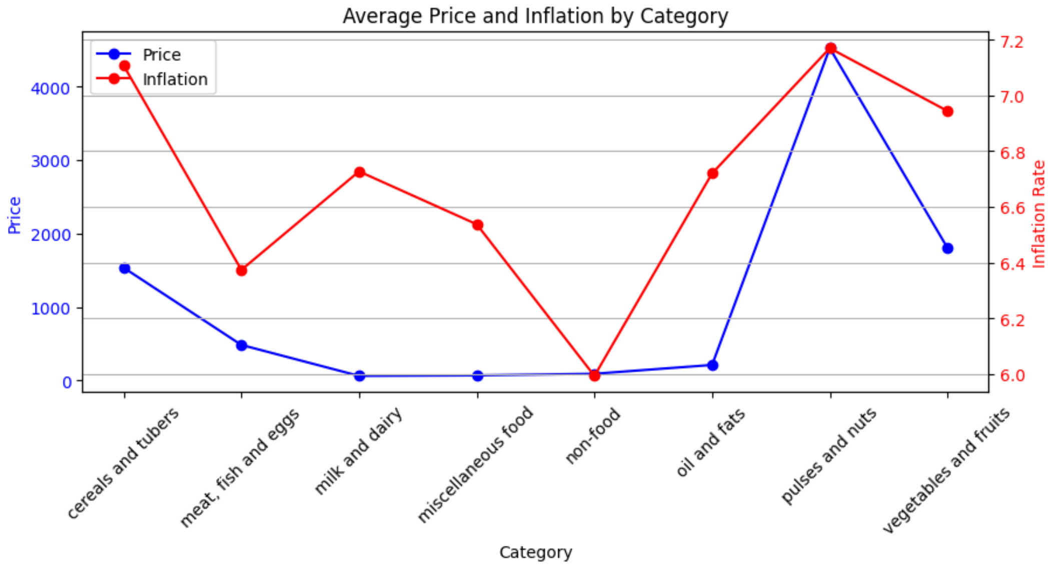




3.4.2 Key Findings: Multivariate Analysis

Upon scrutinizing the interrelationships between the various variables in the dataset simultaneously, several key observations emerged:

- The food category "pulses and nuts" consistently demonstrated higher prices across all provinces compared to other food categories.
- There has been a discernible upward trajectory in the prices of "pulses and nuts" and "cereals and tubers" over the years, while prices for categories like "milk and dairy" have generally remained relatively stable.
- The average prices across all food categories have exhibited an upward trend over time. However, the inflation rate varies across categories, with "pulses and nuts" and "cereals and tubers" experiencing the most pronounced inflationary pressures.


Average Price and Inflation by Category
Price
Inflation
4000
3000
2000
1000
0
Price
7.2
7.0
6.8
6.6
6.4
6.2
6.0
Inflation Rate
cereals and tubers
meat, fish and eggs
milk and dairy
miscellaneous food
non-food
oil and fats
pulses and nuts
vegetables and fruits
Category

3.4.3 Correlation Analysis

Moreover, SokoSmart employed the correlation coefficient to assess the magnitude and direction of the linear association between continuous variables. Notably, the variables exhibiting the strongest correlation, with a coefficient of 0.98, were the USD price and the price in KES.

SokoSmart made the decision to eliminate the USD price in the models and retain the KES price, primarily due to the fact that our primary audience comprises Kenyan farmers, retailers, and consumers who would find it more intuitive to comprehend the price in Kenyan shillings.

4.0 Crafting Predictive Models

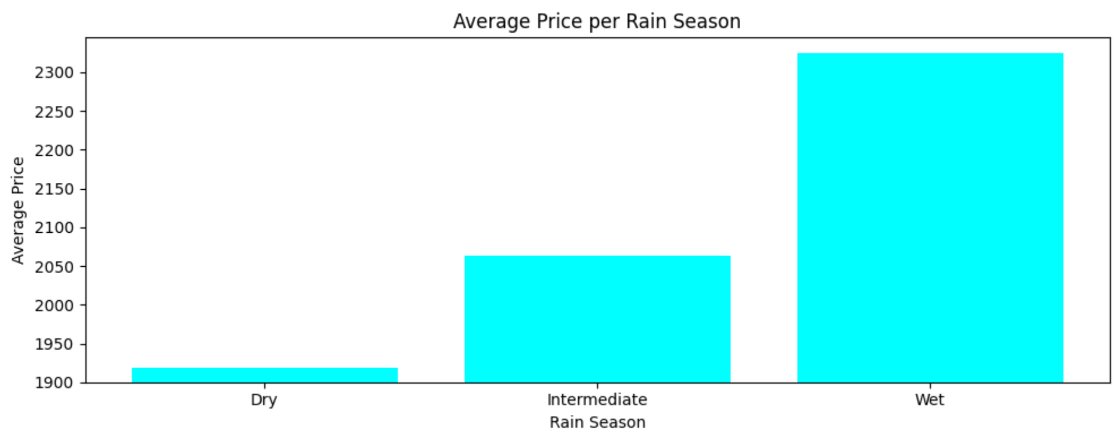
4.1 Data Preprocessing & Feature Engineering

In order to facilitate the creation of predictive models, several preprocessing and feature engineering steps were applied to the dataset:

4.1.1 Seasonal Feature Creation:

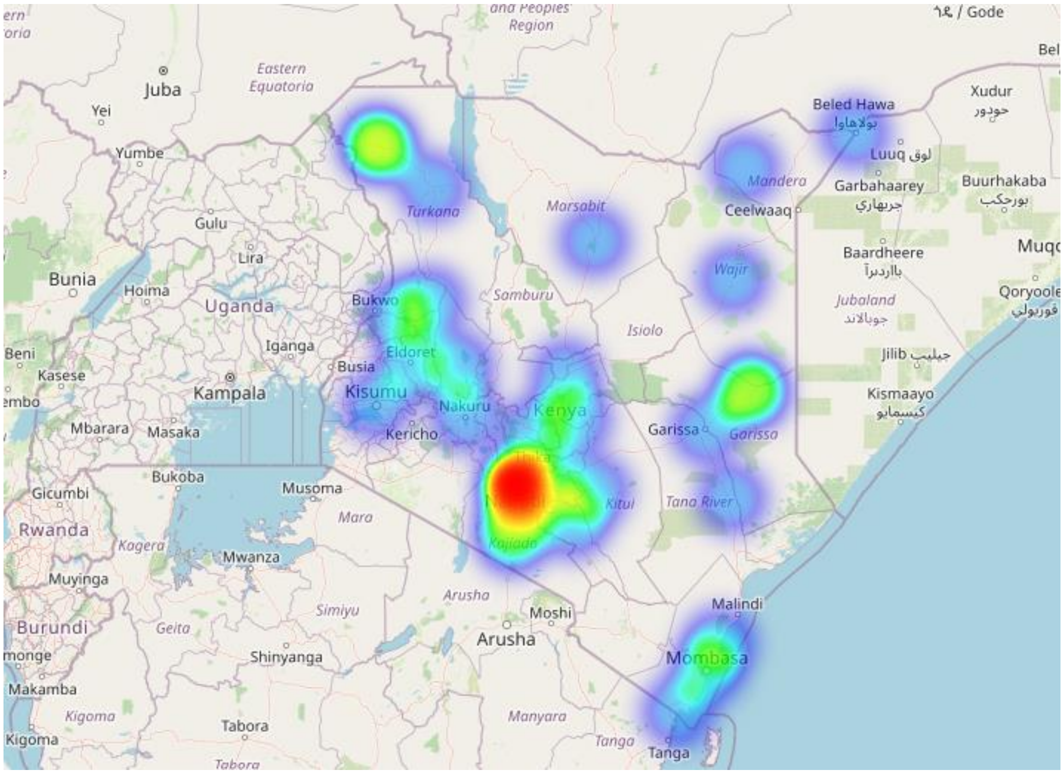
A new feature was introduced to represent the seasons, i.e., wet and dry seasons, based on average rainfall. This classification hinges on whether the average rainfall falls above a specified wet_threshold, below a designated dry_threshold, or within an intermediate range. The rationale behind this categorization is that most crops in Kenya require a minimum of 50 mm of monthly rainfall for proper growth. Months with less than 20 mm of rainfall are typically considered dry or arid, and thus, categorized as "dry."

An additional observation was made regarding commodity prices, showing that they tend to be higher during the wet seasons. This further underscores the significance of incorporating seasonality as a feature for predictive model.



4.1.2 Cluster Analysis:

Scatter plots were used to visualize the relationship between two continuous variables, namely "latitude" and "longitude." These plots revealed clusters of points with similar colors and sizes, indicating regions with comparable price levels. Additionally, regions were segmented based on the types of crops and their associated prices.



The above heatmap visually depicts the distribution and density of data points across the map, using latitude and longitude coordinates. Regions characterized by a greater concentration of data points will exhibit more vibrant and intense colors, appearing "hotter," whereas areas with fewer data points will display subdued and cooler colors.

4.1.3 Feature Selection:

Features were selected based on their importance for model building and predictive accuracy.

Factor analysis was employed to determine underlying patterns between latitude, longitude, price, usdprice and inflation rate. The Factor loading generated are (0, 1 and 2) which have been described below:

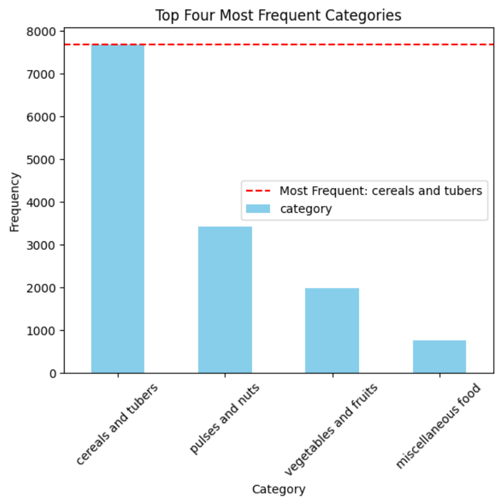
Factor 0: Focuses on price-related variables, with potential geographical influences. Price shows a strong positive relationship, suggesting higher prices in areas with greater latitude and longitude values.

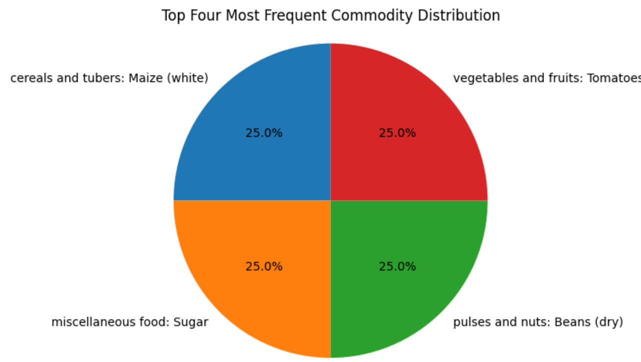
Factor 1: Represents a combination of latitude, longitude, and inflation rate, indicating that lower latitude and longitude regions tend to have lower inflation rates and slightly lower USD prices.

Factor 2: Primarily driven by the inflation rate, with a minor geographical component. Higher inflation areas are linked to lower USD prices.

- **Specific Commodity selection**

Out of the top four most frequent commodities, only maize and beans were suitable for modeling due to insufficient data. Unfortunately, tomatoes and sugar had to be excluded from the analysis.





4.2 Best Model Development and Fine-Tuning.

Two timeseries models were created for this project: a SARIMA model and LSTM model. Following the development and optimization of both models for both commodities, it was determined that the LSTM outperformed the SARIMA model in terms of future predictive accuracy.

This decision was also informed by how closely matched the model's future forecasts with the actual conditions observed in the Kenyan market. The decision was primarily driven by their practical alignment with real-world observations when comparing the current prices from the ministry of Agriculture and livestock development site https://amis.co.ke/site/market .

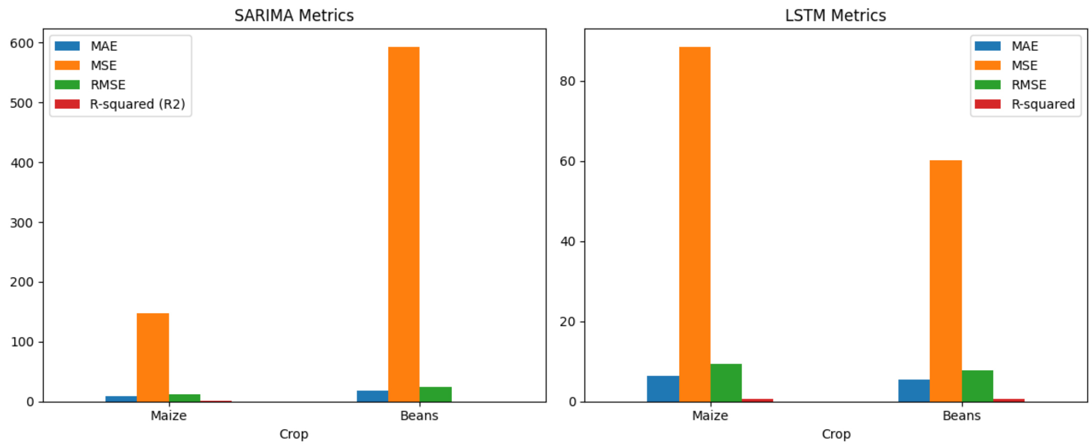
Additionally, the LSTM model has lower values for Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), as well as higher R-squared values, for both Maize and Beans. This suggests that the LSTM model's predictions are closer to the actual prices and it better explains the variance in the data.

SARIMA Model Metrics:				
Crop	**MAE**	**MSE**	**RMSE**	**R-squared**
Maize	9.301546	148.1971	12.17362	0.50903
Beans	17.74993	593.4815	24.36148	0.215182

LSTM Model Metrics:				
Crop	**MAE**	**MSE**	**RMSE**	**R-squared**
Maize	6.420924	88.4857	9.4066	0.57648
Beans	5.486654	60.066	7.75	0.712496

- *MAE (Mean Absolute Error): It represents the average magnitude of the errors between predicted and actual values. A lower MAE indicates better model accuracy.*
- *MSE (Mean Squared Error): This metric calculates the average of the squared differences between predicted and actual values. Smaller MSE values indicate improved prediction precision.*
- *RMSE is the square root of MSE and provides a measure of the average error in the same units as the original data.*

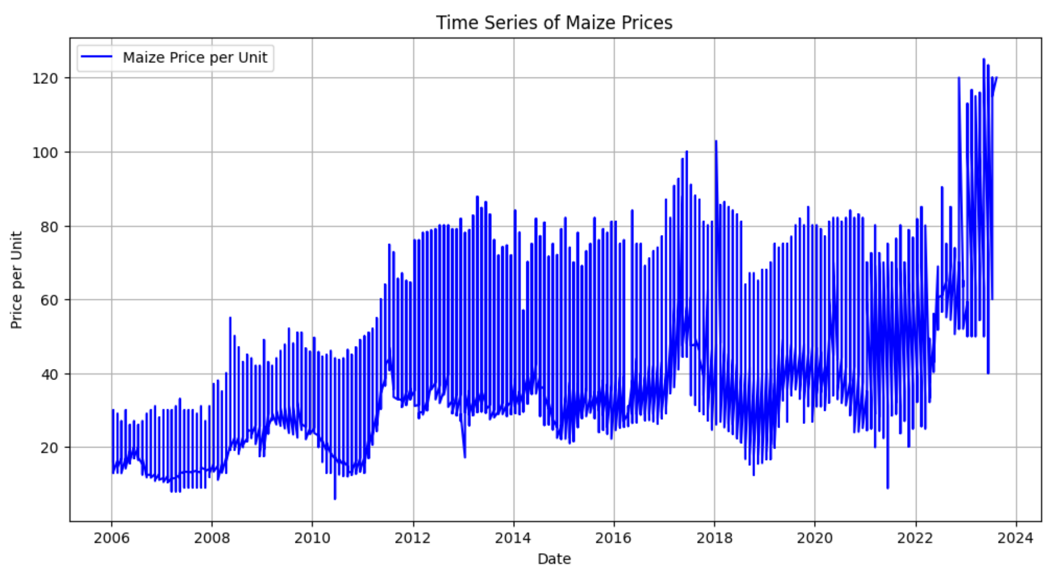
- *R-squared (R2): R-squared quantifies how well the model explains the variance in the data. A higher R2 indicates a better fit, with 1 being a perfect fit.*

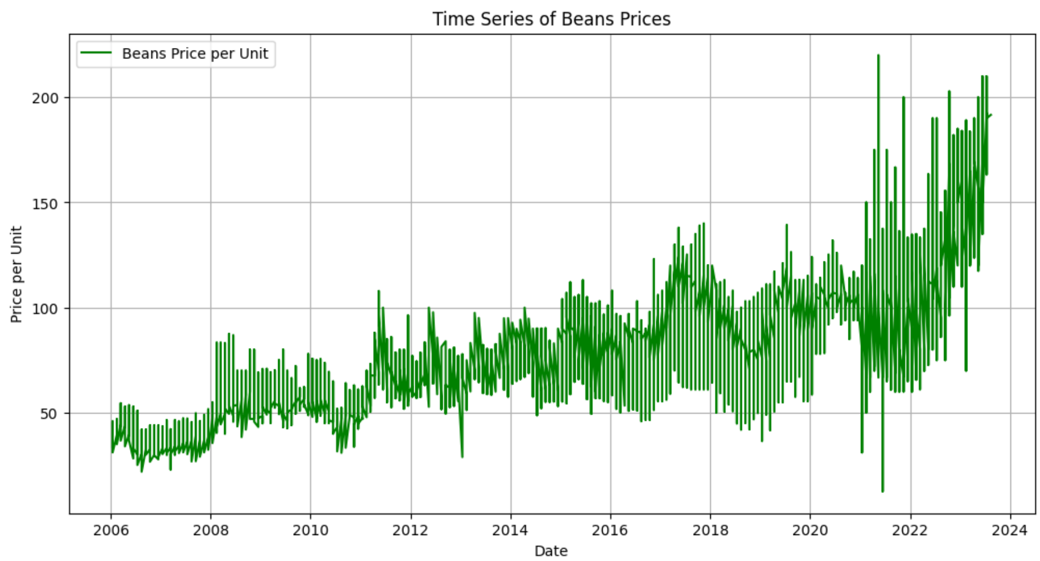


4.2.1 Timeseries Modeling

A time series was constructed for both maize and beans. It became apparent that maize prices have displayed a steady upward trend over time, with occasional short-term fluctuations. The trendline illustrates that the average wholesale price of maize has risen from approximately 20 KES per kilogram in 2006 to about 120 KES per kilogram in 2022.

Similarly, bean prices exhibited an upward trajectory over time, punctuated by periodic short-term variations. The trendline indicates that the average wholesale price of beans has increased from roughly 50 KES per kilogram in 2006 to approximately 200 KES per kilogram in 2022.



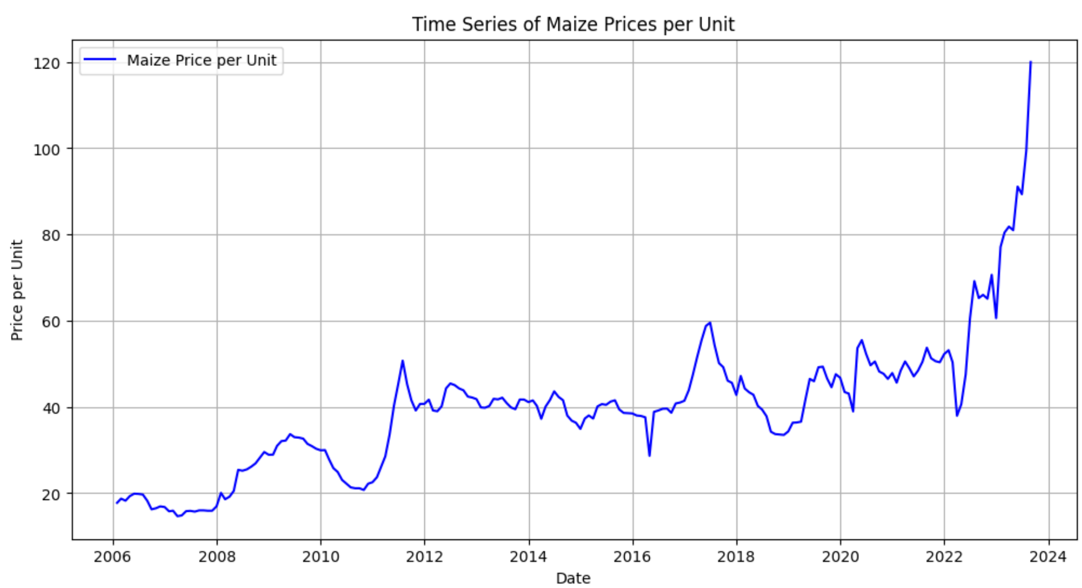


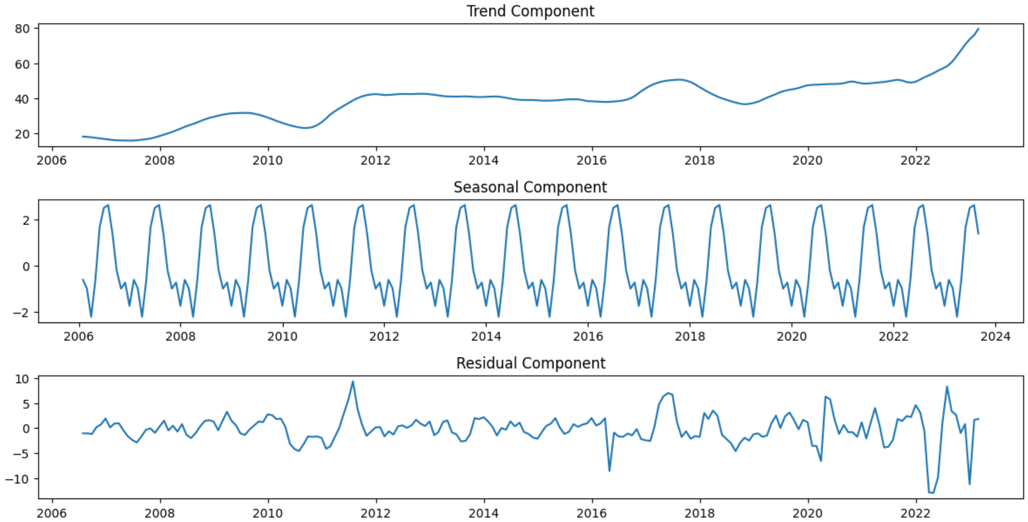
4.2.2 Best Model : LSTM

For both commodities, SokoSmart Analysts conducted a stationarity check by employing the Augmented Dickey-Fuller (ADF) test and implemented differencing to render the data stationary.

Next, SokoSmart decomposed the time series into its constituent components, including trend, seasonality, and residual elements.

Below is an example of Timeseries for maize prices per unit and its decomposition.





Additionally, the data was converted into numerical representation and missing values were imputed using linear interpolation.

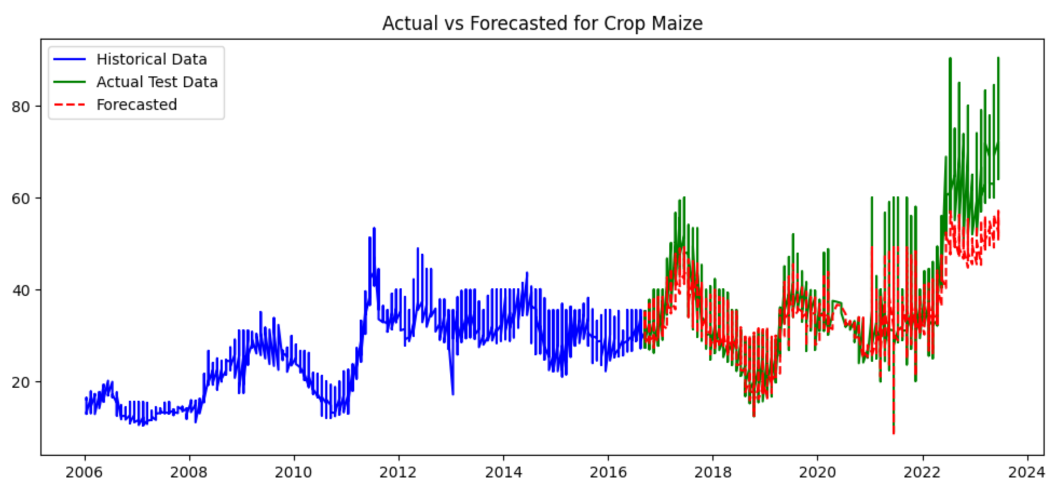
It is important to state that a training and test data set were created before the modelling.

The prices used for modeling were **wholesale unit prices** per each commodity to allow a true picture to be visualized and compared accurately regardless on the units provided.

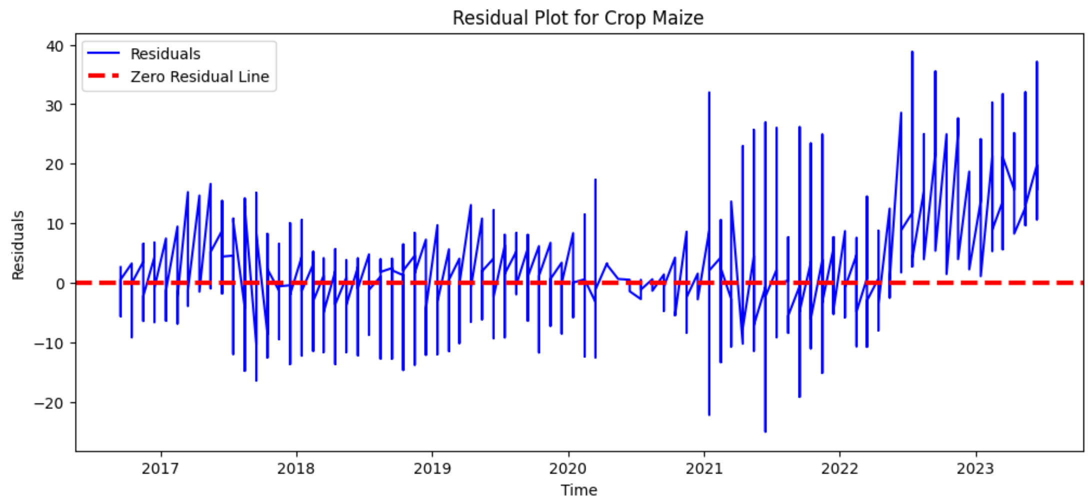
4.3 Baseline Models

4.3.1 Commodity 1: Maize – Baseline LSTM Model Results

The maize model has a Mean Squared Error (MSE) of 81.96, a Mean Absolute Error (MAE) of 6.18 and an R-squared (R2) value of 0.61. With an R2 of 0.61 for maize, this signifies that the model accounts for 61% of the price variation. In essence, the model effectively elucidates a substantial portion of the price fluctuation in maize crops.



In the case of maize, the below residual plot exhibits a number of outliers and discernible patterns within the residual values. Notably, there is a subtle upward trend in the residual values as time progresses. This implies a potential scenario where the model might be consistently underestimating maize crop prices in the later years.

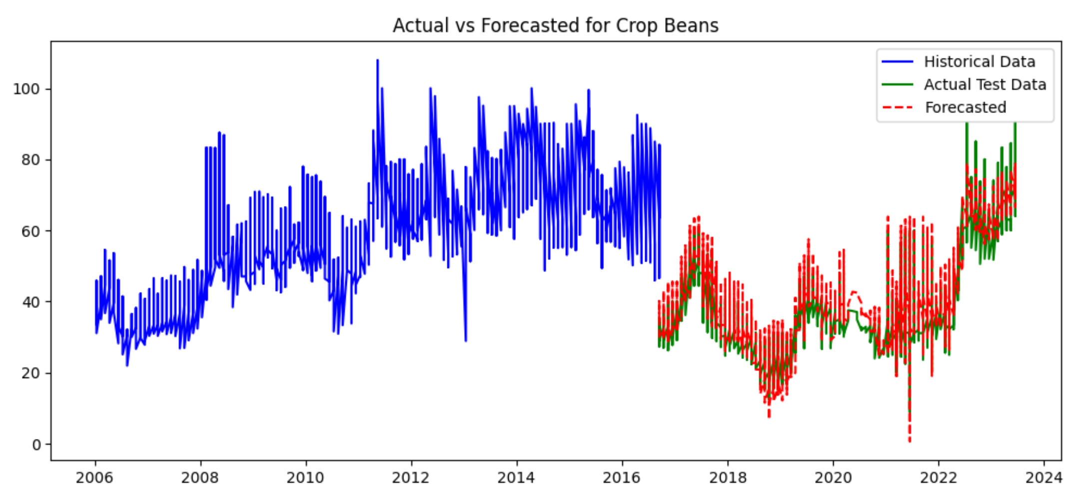


Mean Squared Error (MSE) for Crop Maize: 81.96
Mean Absolute Error for Crop Maize: 6.18
R-squared (R2) for Crop Maize: 0.61

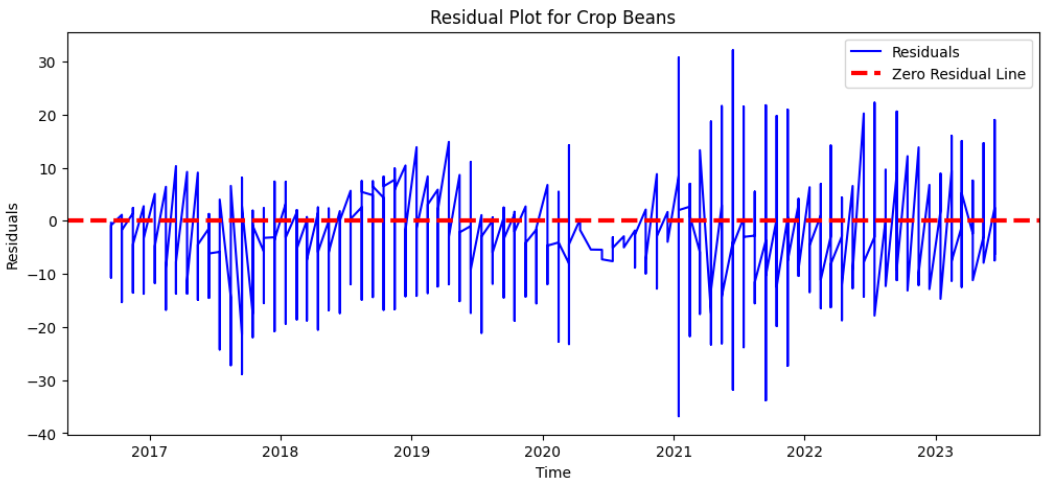
NB: A lower MSE indicates better performance while a higher R2 suggests that the model explains more of the variation in the data.

4.3.2 Commodity 2: Beans – Baseline LSTM Model Results

The beans baseline LSTM model exhibits superior performance in predicting bean prices compared to maize prices. The bean model boasts a Mean Squared Error (MSE) of 63.97, a Mean Absolute Error (MAE) of 5.97, and an R-squared (R2) value of 0.69.



The residual plot for beans below indicates that the model is performing well in price prediction. It reveals that the residual values are mostly centered around zero, with only a few outliers.



Mean Squared Error for Crop Beans: 63.97
Mean Absolute Error for Crop Beans: 5.97
R-squared (R2) for Crop Beans: 0.69

NB: A lower MSE indicates better performance while a higher R2 suggests that the model explains more of the variation in the data.

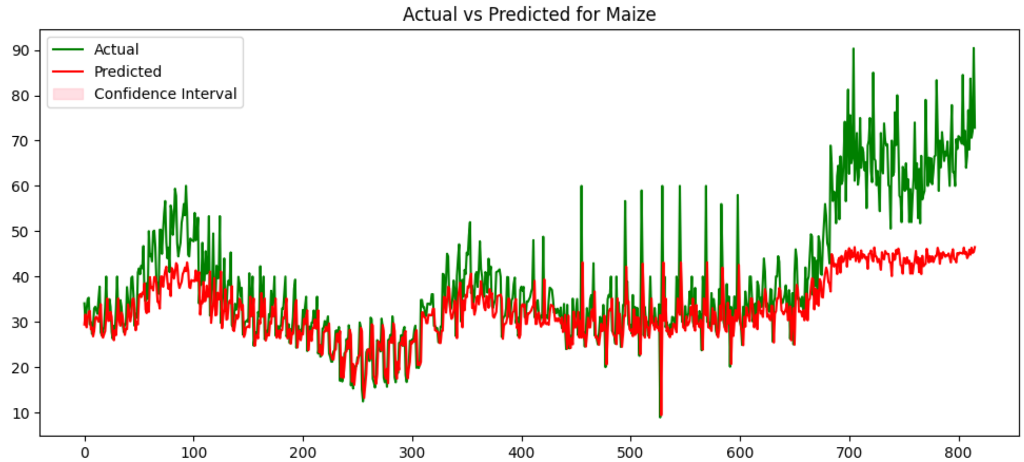
It was noted for both commodity LSTM baselines that the training loss decreases as it is trained more on more data. Theres a drastic decrease up to 2 epochs then the loss starts decreasing slowly.

4.4 Hyperparameter Tuning

In an effort to enhance the initial LSTM Model, SokoSmart analysts introduced the Monte Carlo Dropout and Early Stopping techniques. Interestingly, it was observed that fine-tuning the original models failed to yield better outcomes, and, quite unexpectedly, the **baseline models outperformed** their tuned counterparts.

4.4.1 Commodity 1: Maize – Hyperparameter tuning LSTM Results

The Mean Squared Error (MSE) of 105.13 indicates that, on average, the squared disparities between predicted and actual prices amount to 105.13. The Mean Absolute Error (MAE) of 6.89 signifies that, on average, the model's forecasts differ from actual prices by approximately 6.89 units. The R-squared (R2) value of 0.50 suggests that the model has the capability to account for 50% of the variability in maize prices, denoting a moderate alignment with the data.

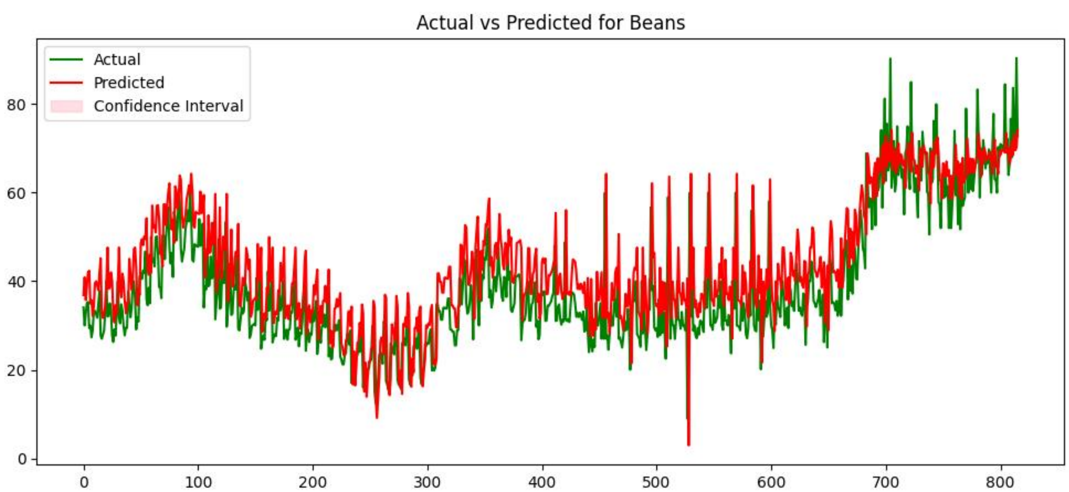


Mean Squared Error for Crop Maize: 105.13	
Mean Absolute Error for Crop Maize: 6.89	
R-squared (R2) for Crop Maize: 0.50	

NB: A lower MSE indicates better performance while a higher R2 suggests that the model explains more of the variation in the data.

4.4.2 Commodity 2: Beans – Hyperparameter tuning LSTM Results

The Mean Squared Error (MSE) of 76.49 reveals that, on average, the squared disparities between predicted and actual bean prices amount to 76.49 signifying a lower level of prediction error in comparison to maize. The Mean Absolute Error (MAE) of 6.61 implies that, on average, the model's forecasts deviate from actual bean prices by approximately 6.61 units. The R-squared (R2) value of 0.63 suggests that the model has the capability to account for 63% of the variability in bean prices, indicating a relatively robust fit to the data



The model developed using the finely-tuned parameters demonstrates superior predictive capabilities compared to the initial Beans - baseline model, as evidenced by the following reduced error metrics.

Mean Squared Error for Crop Beans: 76.49
Mean Absolute Error for Crop Beans: 6.61
R-squared (R2) for Crop Beans: 0.63

NB: A lower MSE indicates better performance while a higher R2 suggests that the model explains more of the variation in the data.

4.5 Forecasting

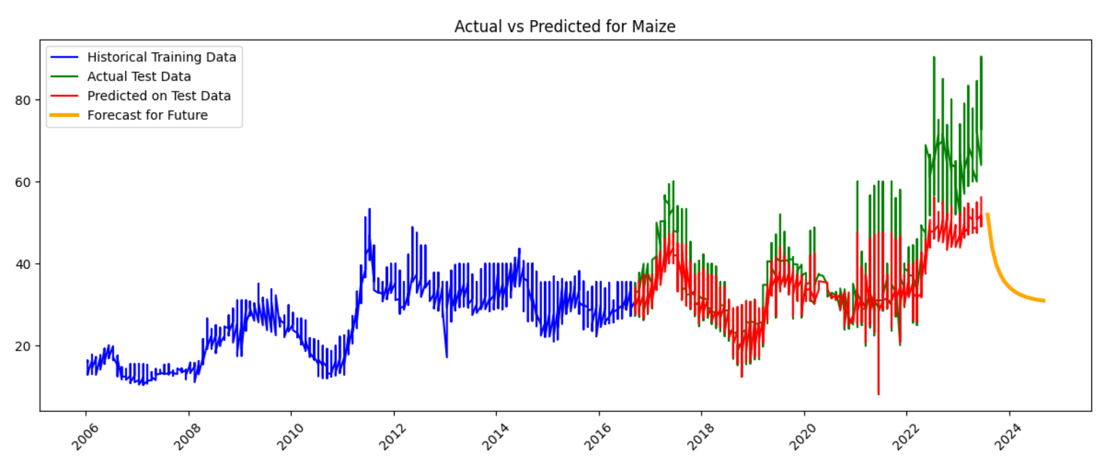
The basic LSTM model demonstrated superior performance compared to the optimized model, prompting the SokoSmart Analysts to employ it for forecasting the upcoming 12 months and contrasting these predictions with the current market values provided by the Ministry of Agriculture.

4.5.1 Commodity 1: Maize – Forecasting 12-month prices

The below output represents a 12-month forecast of wholesale Maize prices per unit, starting from July 31, 2023, to August 31, 2024. The forecasted prices gradually decrease over this period, suggesting an expected downward trend in Maize prices.

- *Overall trend:* Decreasing
- *Highest forecasted price* 51.92 (July 2023)
- *Lowest forecasted price* 31.00 (August 2024)
- *Average forecasted price* 35.59

In the context of maize forecasting, the model can explain 58% of the variability with an R-squared value of 0.58, indicating that it performs reasonably well in prediction, as evidenced by its low Mean Absolute Error (MAE) of 6.42 and Mean Squared Error (MSE) of 88.49.



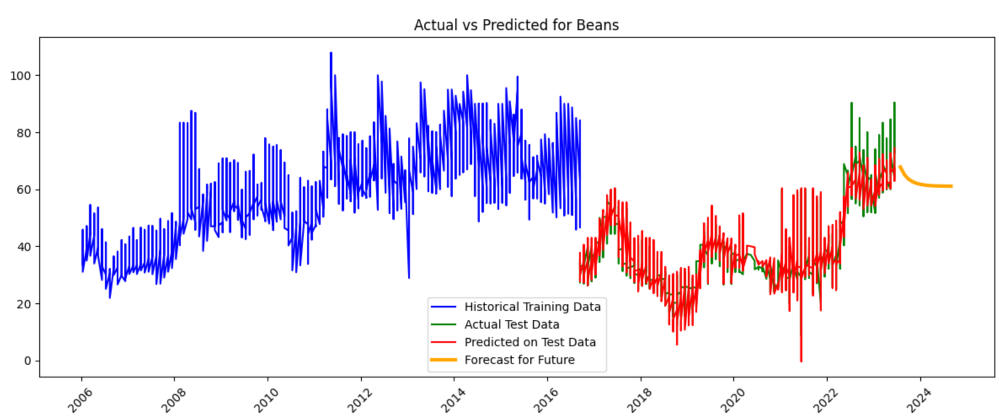
	Date	Forecasted_Price
0	31-07-23	51.922245
1	31-08-23	44.003323
2	30-09-23	39.813644
3	31-10-23	37.242111
4	30-11-23	35.525875
5	31-12-23	34.318111
6	31-01-24	33.437054
7	29-02-24	32.777687
8	31-03-24	32.274883
9	30-04-24	31.886019
10	31-05-24	31.582024
11	30-06-24	31.342379
12	31-07-24	31.152214
13	31-08-24	31.000542

4.5.2 Commodity 2: Beans – Forecasting 12-month Prices

The below output presents a 12-month forecast of Crop Maize prices per unit, starting from July 31, 2023, to August 31, 2024. Interestingly, the forecasted prices remain relatively stable and shows very little variation over the entire forecasting period. This suggests that the model expects maize prices to stay consistent with minor fluctuations, at least within the scope of this forecast.

- Overall trend: Decreasing
- Highest forecasted price: 67.83 (July 2023)
- Lowest forecasted price: 61.09 (June 2024)
- Average forecasted price: 62.38

In the case of beans forecasting, the model can explain 71% of the variability with an R-squared value of 0.71, signifying that it performs well in prediction, as indicated by its low Mean Absolute Error (MAE) of 5.49 and Mean Squared Error (MSE) of 60.07.

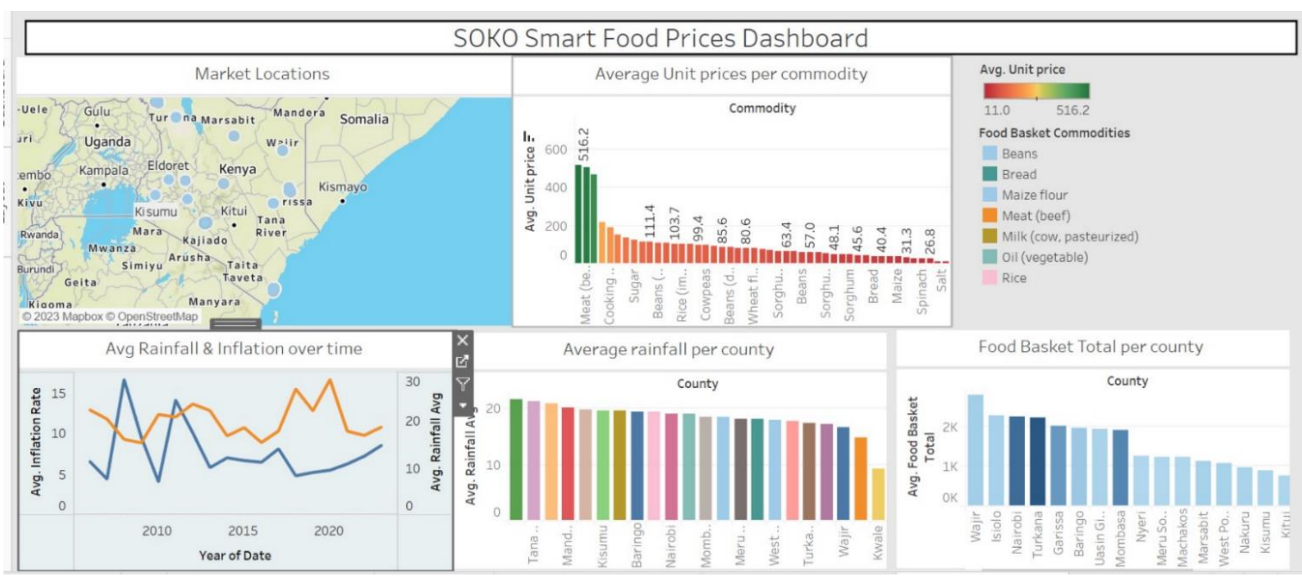


	Date	Forecasted_Price
0	31-07-23	67.835953
1	31-08-23	65.221024
2	30-09-23	63.701977
3	31-10-23	62.770588
4	30-11-23	62.181259
5	31-12-23	61.801086
6	31-01-24	61.552826
7	29-02-24	61.389408
8	31-03-24	61.281303
9	30-04-24	61.20953
10	31-05-24	61.16177
11	30-06-24	61.12994
12	31-07-24	61.108715
13	31-08-24	61.094551

4.6 Food Basket

According to the World Food program, a food basket consists of food commodities critical to maintaining the nutritional status people tailored to local preferences, demographic profile, activity levels, climatic conditions & local coping capacity.

Drawing upon this, we decided to choose the following items beans, bread, maize flour, beef, milk, oil, and rice for the food basket.



Maize Flour: Maize is a staple crop in Kenya, forming the basis of many traditional dishes. It's a primary source of carbohydrates, providing energy and satisfying a key dietary preference.

Beans: Beans are a rich source of plant-based protein, critical for balanced nutrition. In Kenya, they are widely consumed, and their inclusion helps meet the protein requirements of the population.

Beef: Beef is a valuable source of animal protein. In a country with varying dietary preferences and cultural traditions, it offers a protein option that's widely accepted.

Milk (Cow): Milk, particularly from cows, is an essential source of calcium and other nutrients. In many Kenyan communities, dairy is a fundamental part of the diet, and it helps in meeting the nutritional needs of children and adults.

Oil: Vegetable oil is important for providing essential fats and calories. Its inclusion ensures a balanced intake of nutrients and also aligns with cooking practices in Kenya.

Rice: While rice is not a Kenyan staple like maize, it offers dietary diversity and is well-accepted in urban areas. It complements the food basket, catering to different preferences and providing an alternative source of carbohydrates.

Analyzing the cost of this food basket in Kenya's various regions and counties, several insights emerge.

- Firstly, prices can significantly vary based on local market conditions and transportation costs.
- In urban areas, where demand is higher, prices may be slightly elevated.
- In contrast, more rural regions may have lower prices but may face challenges in consistent availability. The accessibility of these food items, especially in remote counties, can lead to fluctuations in prices.
- Additionally, seasonality plays a role, as agricultural harvests can impact the cost of staples like maize flour.

5.0 Drawing Conclusions

- In the realm of Kenyan food security, the ability to accurately predict food prices is of paramount importance. Our research, focusing on predicting food prices in Kenya, reveals valuable insights that can significantly impact various stakeholders within the agricultural sector.
- Key findings from the study encompass a comprehensive understanding of the dataset, covering a wide range of food commodities and regions. We discovered substantial price variability, a connection between inflation rate and food prices, and distinctive food category preferences among Kenyan consumers. These insights can guide market decisions and government policies to promote food security.
- Our research methodology, which included data preprocessing, feature engineering, and cluster analysis, allowed us to create predictive models. After careful evaluation, we selected the LSTM model as the preferred choice for forecasting food prices.
- Forecasting the future prices of maize and beans for the next 12 months, we provide valuable tools for decision-making within the agricultural sector. These predictions offer insights that can

aid Kenyan farmers, retailers, and consumers in planning and maximizing profits, ultimately contributing to food security in the nation.

6.0 Recommendations

Based on our research and predictive models, we offer the following recommendations to various stakeholders within the agricultural sector:

- **Farmers and Retailers**: Use the forecasted prices to plan your planting, harvesting, and stocking strategies. Being aware of future price trends can help optimize production and sales, ensuring better financial outcomes.

  Farmers may consider diversifying their crop portfolio to spread risk. Relying solely on one crop can render them vulnerable to market fluctuations.

  Retailers can explore hedging strategies to mitigate potential price risks. This may involve the use of forward contracts or other financial instruments.

- **Government and Policymakers**: Consider incorporating price forecasts into your food security and agricultural policies. These forecasts can help in managing food imports and exports, stabilizing prices, and ensuring a consistent food supply for the population. Given the anticipated increase in the price and beans in the coming months, the government should secure more food commodities (maize and beans) now before the price increases.

- **Consumers:** Be mindful of potential price fluctuations for maize and beans in the coming months. This awareness can assist you in making informed purchasing decisions and managing your food budget more effectively

By implementing these recommendations and continuously refining predictive models, Kenya can strengthen its food security, support its agricultural sector, and ensure a consistent and affordable food supply for its citizens.

7.0 References

https://www.fao.org/kenya/fao-in-kenya/kenya-at-a-glance/en/ : FAO in Kenya

https://food.agricultureauthority.go.ke/index.php/sectors/overview:

https://data.humdata.org/dataset/wfp-food-prices-for-kenya : World Food Programme

https://dataviz.vam.wfp.org/version2/climate-explorer : World Food Programme

https://www.centralbank.go.ke/inflation-rates/ : Central Bank of Kenya

https://amis.co.ke/site/market : Ministry of Agriculture and Livestock development

8.1 Acknowledgements

We extend our heartfelt gratitude to the dedicated team of SokoSmart Analysts whose unwavering efforts made this project a reality.

Their commitment to creating and deploying a powerful time series model for forecasting future prices has been instrumental in achieving our research objectives. Thank you for your hard work and dedication.