AgriSenseAi

**Market Recommender and Price Predicting System for Maize, Tomatoes, and Avocados in Kenya**



## **1. Project Team Members**

This project is a **group capstone project** developed by:

* **Knight Mbithe**
* **Enid Kibet**
* **Joy Gitau**
* **John Louis Kirigo**
* **Caleb Kipkech**

## 

## **2. Business Understanding**

### Overview

Agriculture is the backbone of Kenya’s economy, contributing over **20% of GDP** and employing more than **40% of the workforce**. However, **smallholder farmers and traders face significant challenges**, including **market access limitations, price volatility, and unpredictable climatic conditions**. These factors lead to **income instability, post-harvest losses, and inefficient supply chains**.

Traditional market decision-making in Kenya **relies on manual surveys and informal networks**, which are often **inaccurate, outdated, and lack predictive power**. As a result, farmers and traders struggle to determine the **best markets to sell their produce** and to anticipate **fluctuations in crop prices**.

### Problem Statement

Kenyan farmers and traders **lack reliable market intelligence** to guide pricing and market selection decisions. Without accurate **price forecasts and demand analytics**, they suffer from **unstable incomes, post-harvest losses, and inefficient supply chains**.

To address this, AgriSense proposes a **deep learning-powered market recommender and price forecasting system** that leverages:

* **Historical price data**
* **Weather conditions**
* **NDVI (vegetation health index)**
* **Supply-demand factors**
* **Transport costs and infrastructure quality**
* **Macroeconomic indicators (inflation, exchange rates, fuel costs)**

### Objectives

#### **Main Objective:**

To develop a **deep learning-powered market recommender and price prediction system** integrating **historical price data, weather conditions, NDVI index, supply-demand factors, transport costs, and macroeconomic indicators** to optimize decision-making for farmers and traders.

#### **Specific Objectives:**

* **Market Price Prediction** – Build a deep learning model to predict future prices of **maize, tomatoes, and avocados** by utilizing **historical pricing data, climate variables, and demand-supply trends**.
* **Market Recommender System** – Develop a system to suggest the best markets based on **demand-supply dynamics, price trends, transport costs, and road infrastructure quality**.
* **Real-Time Data Integration** – Integrate real-time data sources on market price feeds to enhance prediction accuracy.
* **User-Friendly Dashboard** – Create an accessible dashboard to provide **real-time price forecasts, market recommendations, and insights based on macroeconomic conditions**.

### Success Criteria

* **Prediction Accuracy:** Models should achieve at least **85% accuracy** (evaluated using **MAPE and RMSE**).
* **Market Recommendation Efficiency:** The recommender should correctly identify **high-profit markets at least 85% of the time** (evaluated using **precision, recall, and F1-score**).
* **Pilot Testing:** Conduct a small-scale test with selected farmers or traders to measure usability and impact.
* **Deployment Feasibility:** Ensure trained models are ready for integration into a user-friendly platform.

## **3. Data Understanding**

### 3.1 Data Sources

* **FAOSTAT**:<https://www.fao.org/faostat/en/#data/>
* **KAMIS**:<https://amis.co.ke/site/market>
* **CBK Monthly Economic Indicators Reports**:<https://www.centralbank.go.ke/monthly-economic-indicators/>
* **Ministry of Roads and Transport Annual Report**: [https://www.transport.go.ke](https://www.transport.go.ke/)
* **Kenya Urban Roads Authority (KURA) Annual Reports**: [https://kura.go.ke](https://kura.go.ke/)
* **Kenya Rural Roads Authority (KeRRA) Project Reports**:<https://kerra.go.ke>

The data was sourced and compiled by the project team. The data was retrieved from multiple sources, cleaned, and merged into a unified dataset. The process included:

* **Data Extraction**: Collecting data from FAOSTAT, KAMIS, CBK Monthly Reports, and other government sources.
* **Data Cleaning**: Handling missing values, standardizing formats, and removing inconsistencies.
* **Data Integration**: Merging datasets on common attributes such as date, location, and market indicators.
* **Feature Engineering**: Creating relevant columns that enhance model performance.

### 3.2 Dataset Structure

| **Column** | **Description** |
| --- | --- |
| **Date** | Market transaction date |
| **Price (KES/kg)** | Crop price (Maize, Tomato, Avocado) |
| **Market Demand Factor** | Ratio of demand to supply for each market |
| **NDVI Index** | Vegetation health index (0-1, higher = healthier crops) |
| **Market** | The specific market where the transaction took place |
| **County** | Geographical location of the market |
| **Retail Price** | Retail price for consumers |
| **Temperature (°C)** | The average temperature on the transaction date |
| **Humidity (%)** | Humidity level (%) on the transaction date |
| **Rainfall (mm)** | Rainfall recorded in the region on that day |
| **Crop** | Type of crop (Maize, Tomato, Avocado) |
| **Month** | Months of the transaction |
| **Year** | Year of the transaction |
| **Season** | Agricultural season (e.g., Long Rains, Short Rains) |
| **Average Transport Cost (KES/km)** | Estimated cost of transportation per kilometer |
| **Road Infrastructure Quality Index** | Quality of roads affecting transport efficiency |
| **Inflation Rate (%)** | Overall inflation in the economy |
| **Food Inflation (%)** | Inflation specific to food prices |
| **Fuel Inflation (%)** | Inflation in fuel prices affecting transport costs |
| **Exchange Rate (KES/USD)** | Exchange rate influencing market stability |

### 4. Challenges Experienced

Throughout the project, we encountered several challenges:

* **Real-Time Price Forecasting:** We were unable to implement real-time price forecasting as we could not successfully link to the KAMIS website for live price updates.
* **Data Collection Difficulties:** Sourcing relevant data was challenging as it was not readily available from centralized sources.
* **Meteorological Data Access:** We were unable to acquire climate data from the meteorological department, which limited the accuracy of our climate-based predictions.
* **Limited Time Constraints:** Given the constraints of our capstone project timeline, we focused on only three crops—maize, tomatoes, and avocados—rather than a broader range of agricultural products.

### 5. Notebook 1 – Data Preparation & Exploration

### 5.1 Data Preparation

1. **Loading the Data**The process begins by reading the raw data from a CSV file into a Pandas DataFrame. This dataset contains essential market information such as dates, product types, prices, volumes, and market locations. Loading the data ensures that all the necessary information is available for subsequent analysis.
2. **Initial Data Inspection**Once the data is loaded, the next step is to inspect its structure and contents:
   * **Previewing the Data:** The first few rows are examined to understand the overall structure.
   * **Checking Dimensions:** The number of rows and columns is verified to understand the dataset's scale.
   * **Assessing Data Types and Completeness:** Detailed information on the data types and any missing values is reviewed.
   * **Generating Summary Statistics:** Key numerical statistics (such as mean, median, and standard deviation) are calculated to identify any anomalies or irregular patterns early in the process.
3. **Handling Missing Values**Missing values are identified and addressed to prevent them from skewing the analysis:
   * **Identification:** The dataset is checked for missing values in each column.
   * **Imputation:** For numeric columns (such as price), missing values are replaced with the median value. For categorical columns (such as market location), missing values are filled in using the mode, ensuring that the dataset remains complete without discarding valuable information.
4. **Checking for Duplicates**Although this step typically involves checking for and removing duplicate entries, it was found that the dataset contained no duplicates. This means that every record in the dataset is unique, which helps maintain the integrity of the analysis.
5. **Data Type Conversion**To ensure that each column is in the proper format for analysis:
   * **Date Conversion:** The date column is converted from a string format to a datetime object, which allows for time-based analysis.
   * **Numeric Conversion:** Columns that represent numbers but may have been read as strings are converted to numeric types.
   * **Categorical Conversion:** Textual data such as product names is converted to a categorical data type, which helps with both efficiency and clarity in further analysis.
6. **Outlier Detection and Treatment**Outliers can distort analysis, so it’s important to identify and manage them:
   * **Visual Inspection:** A box plot is used to visualize the spread of the price data, helping to identify any extreme values.
   * **IQR Method:** The interquartile range (IQR) method is applied to filter out any values that fall outside a defined acceptable range, thus minimizing the impact of potential anomalies on the analysis.
7. **Standardizing Text Data**Finally, textual data is standardized to ensure consistency:
   * **Uniformity:** All text entries (for example, product names) are converted to lowercase and stripped of extra spaces. This step prevents discrepancies due to variations in capitalization or formatting.
   * **Column Renaming:** Columns may be renamed to more intuitive names, ensuring clarity and consistency in later stages of analysis.

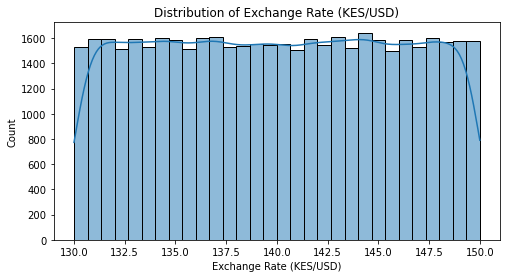
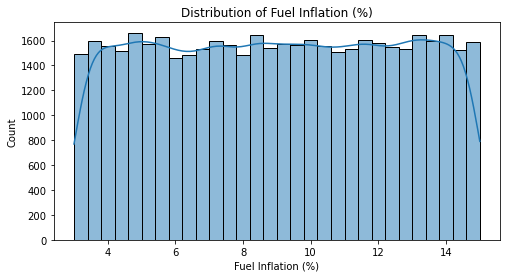
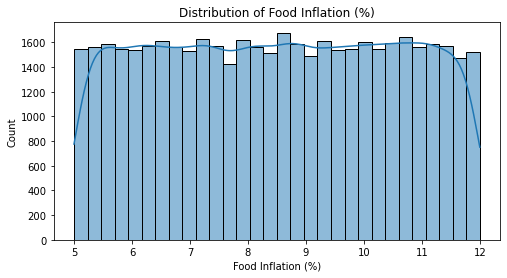
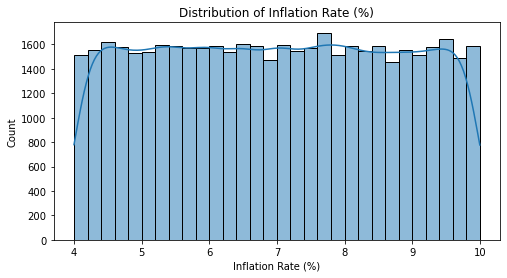
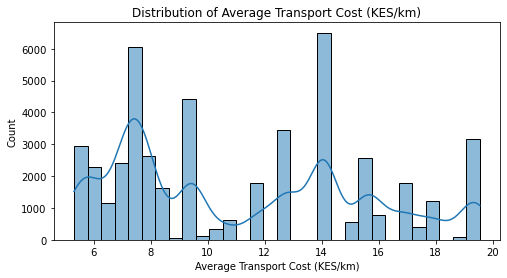
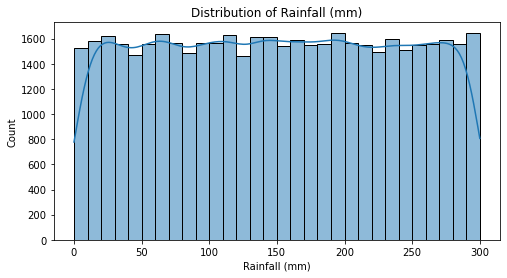
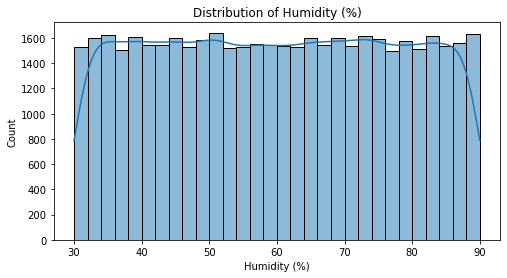
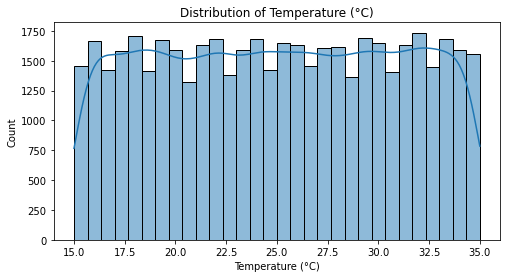
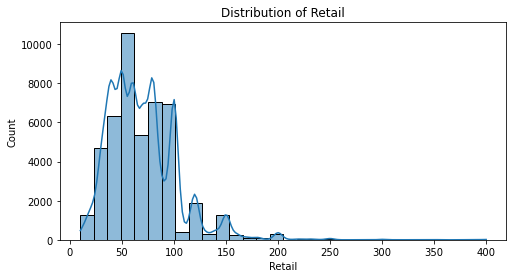
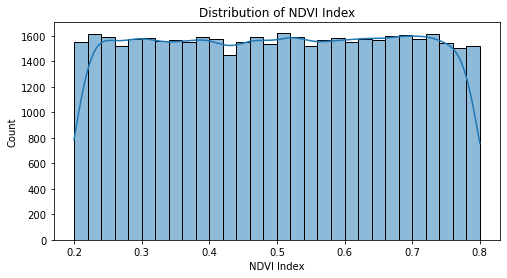
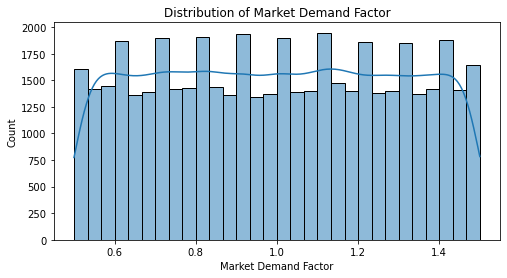
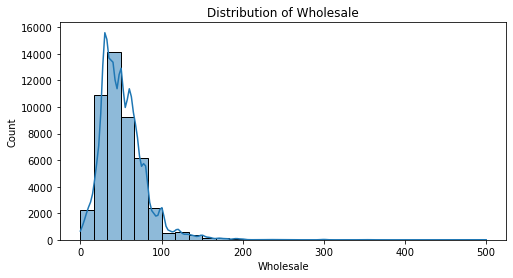
Each of these steps is designed to systematically transform raw market data into a clean, consistent, and analysis-ready dataset, forming a solid foundation for further exploratory data analysis, model building, and ultimately, the development of a robust market recommender and price forecasting system.

### 5.2 Data Analysis

#### **5.2.1 Univariate Analysis** The objective of univariate analysis is to understand the behavior of each variable in the dataset individually.

* **Histograms:**

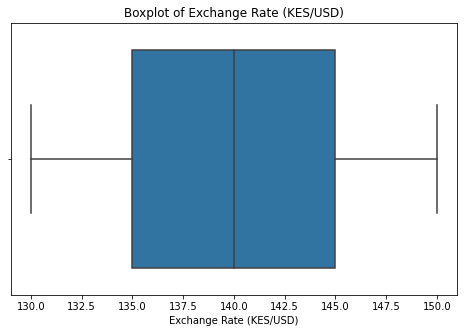
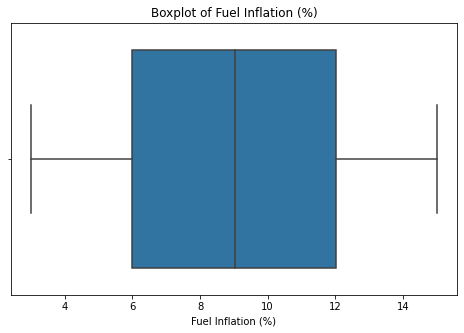
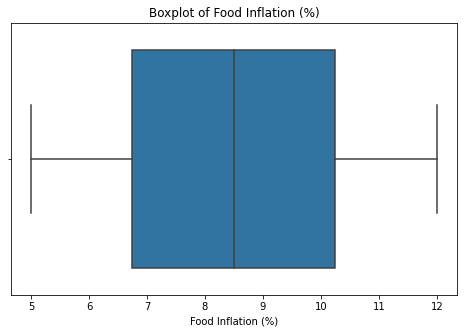
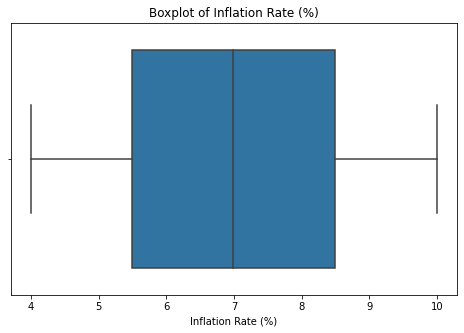
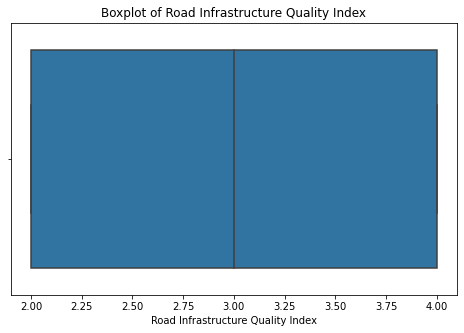
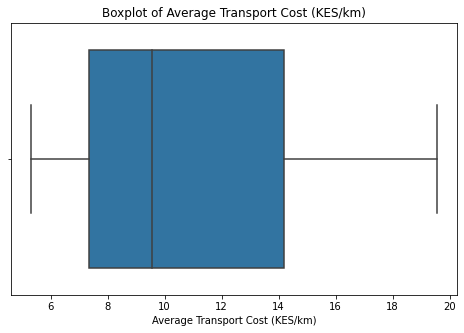
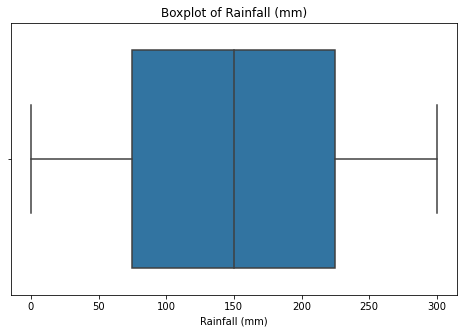
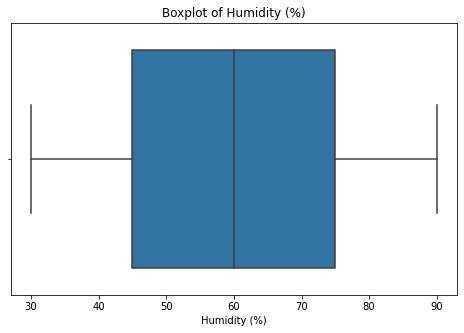
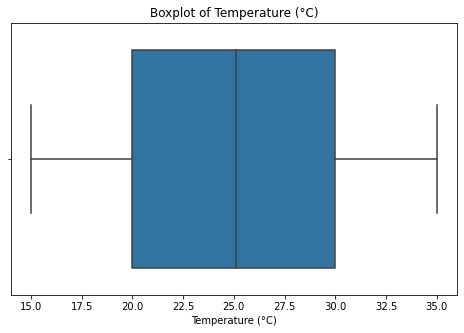
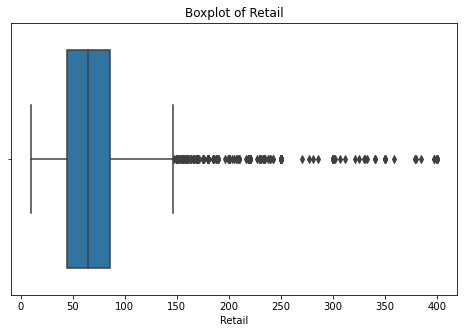
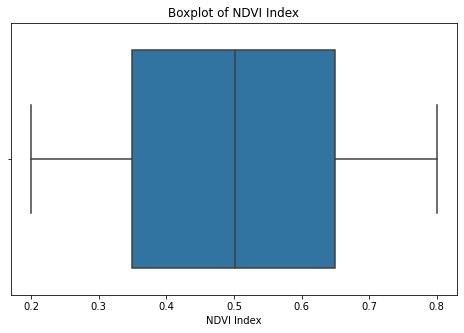
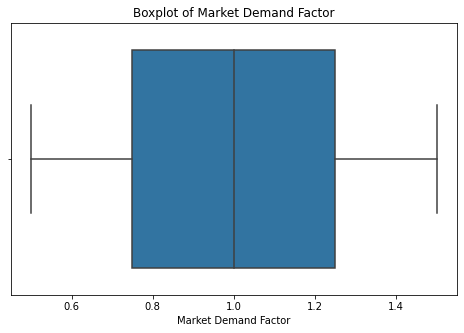
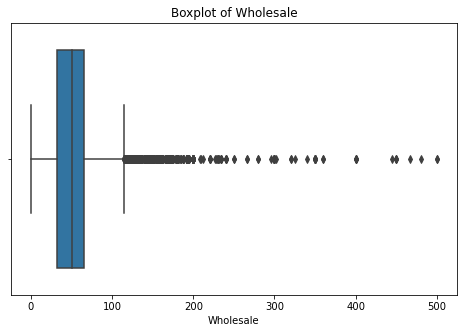
Histograms were created for key numerical variables. These plots display the frequency distribution of each variable and include a kernel density estimate (KDE) overlay to highlight the underlying probability density.



The observation is that wholesale and retail price distributions are right-skewed, indicating most transactions occur at lower price levels, with occasional high-priced outliers. Market demand is relatively uniform with some outliers suggesting instances of excess supply or acute shortages, while the NDVI index points to generally consistent crop health across regions—with deviations possibly indicating localized climate stress. Weather conditions are mostly stable, although sporadic extreme events may impact crop yields. Furthermore, transport costs show multiple peaks due to regional differences, and road infrastructure quality exhibits a bimodal distribution, highlighting a divide between well-maintained and poorly maintained roads that affects transport efficiency.

On the economic side, inflation rates—including those for general, food, and fuel inflation—display varying patterns, and the bimodal exchange rate distribution underscores periods of stability followed by depreciation. These economic fluctuations, combined with factors such as poor infrastructure and weather variability, contribute significantly to market price volatility. Overall, these insights emphasize the importance of integrating multiple data sources into a robust market recommender and price forecasting system to aid farmers and traders in making informed decisions in a dynamic agricultural market.

* **Boxplots:**

****

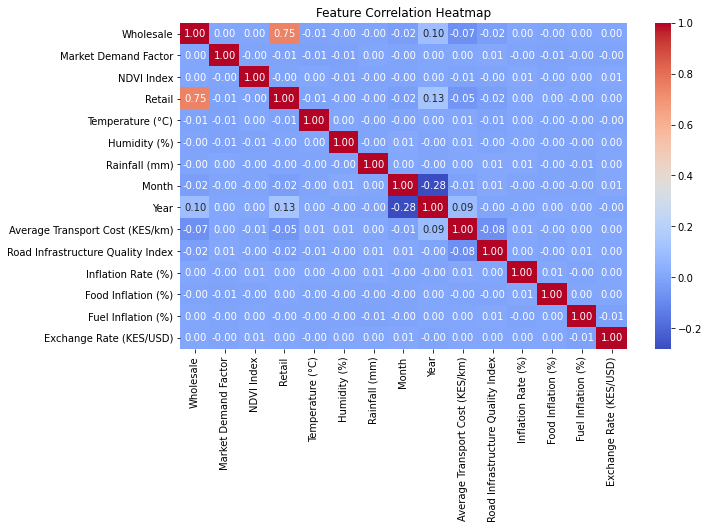
The boxplot analysis reveals distinct patterns across various economic, environmental and infrastructure metrics. For variables such as wholesale and retail prices, inflation rates, food & fuel inflation and the exchange rate, the data exhibits a right-skewed distribution with moderate to wide interquartile ranges and notable high-end outliers. This suggests that while the majority of values are relatively low, occasional extreme spikes possibly due to supply chain disruptions, policy shifts or global market pressures can have a significant impact. In contrast, indicators such as the market demand factor, NDVI index and temperature show more symmetrical distributions with limited outliers, implying stable conditions and less volatility in these areas.

On the other hand, metrics such as humidity, rainfall and average transport cost display greater variability with numerous outliers and highly skewed distributions. This indicates that these factors can fluctuate dramatically depending on regional or temporal conditions, potentially affecting crop yields and overall market accessibility. The road infrastructure quality index, while moderately variable, also points to disparities that could hinder market connectivity in areas with poorer conditions. Overall, the analysis underscores the need for targeted interventions ranging from economic stabilization measures to infrastructure improvements and adaptive agricultural practices to address the identified extremes and ensure a more balanced, predictable operating environment.

#### **5.2.1 Bivariate Analysis**

The bivariate analysis aims to explore and quantify the relationships between pairs of variables. Its primary objectives are to identify associations, understand correlations, and detect patterns that could inform further analysis or predictive modeling.

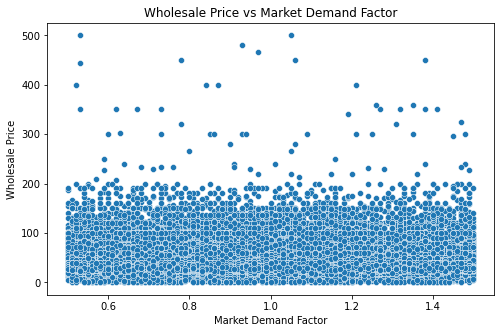
* **Correlation Patterns Among Variables**

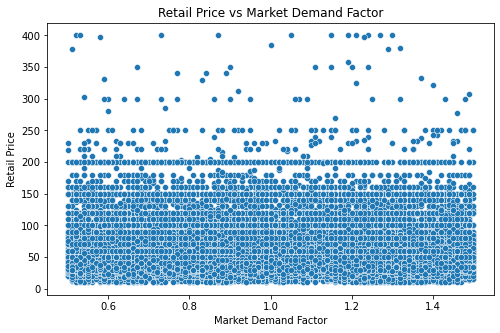
****

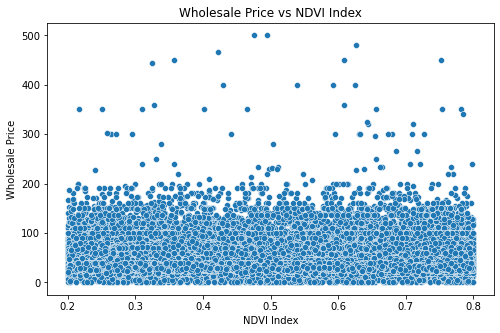
The correlation heatmap reveals strong positive relationships between wholesale and retail prices, indicating that fluctuations at the wholesale level directly impact consumer pricing. Additionally, food inflation exhibits a near-perfect correlation with general inflation and fuel prices, emphasizing the role of rising transport and production costs in driving food prices higher. The relationship between fuel inflation and food inflation highlights the sensitivity of agricultural markets to changes in fuel costs, as transportation is a critical factor in price determination.

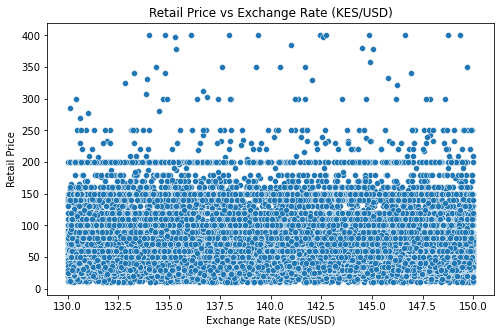
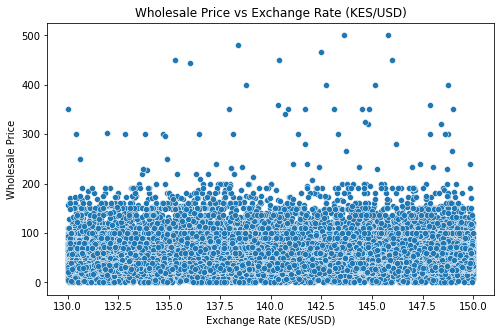
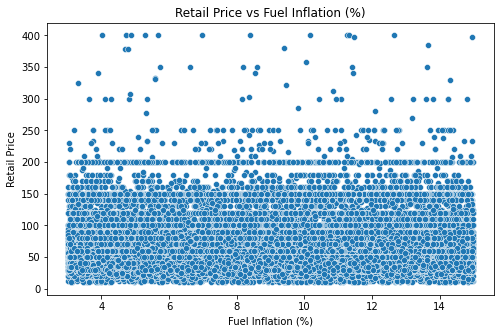
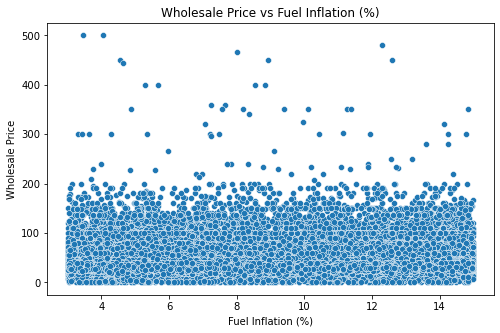
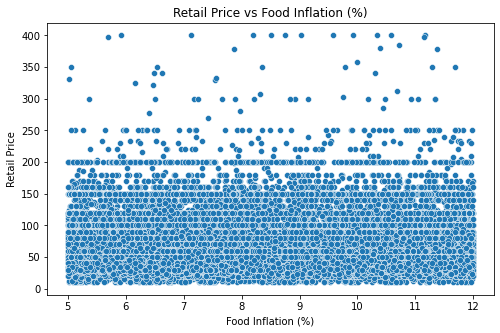
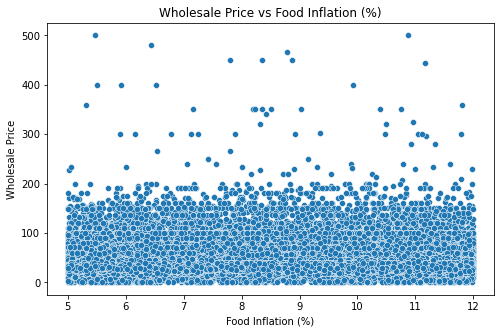
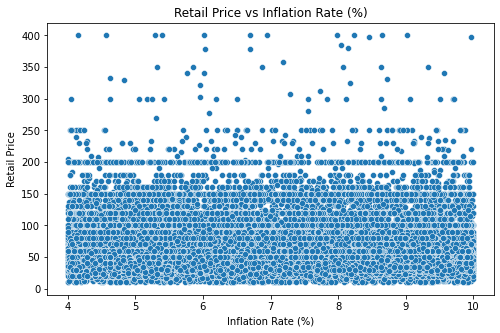
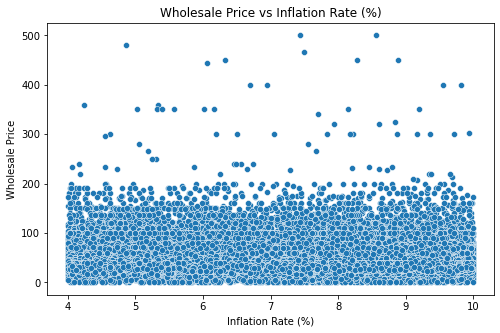
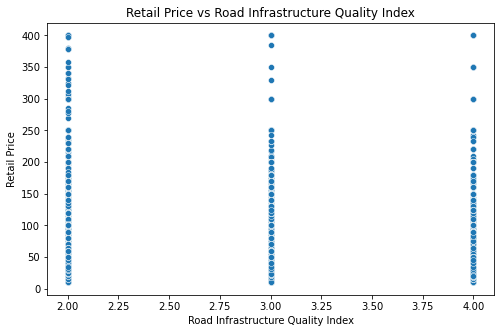
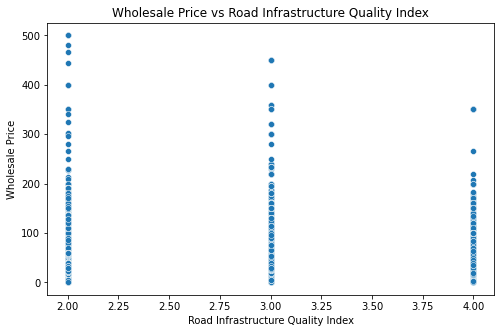
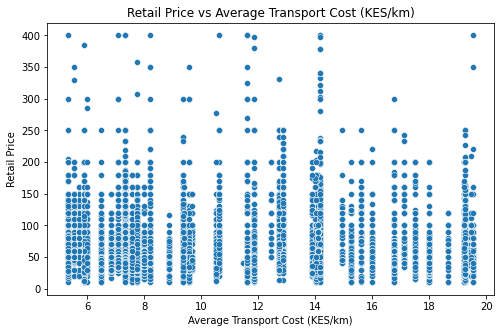
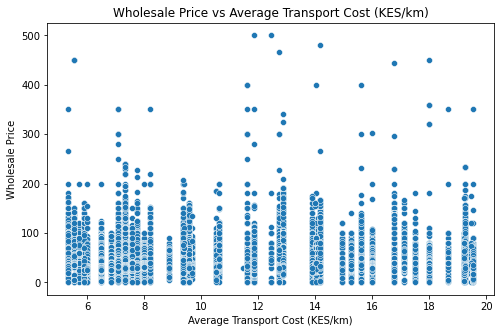
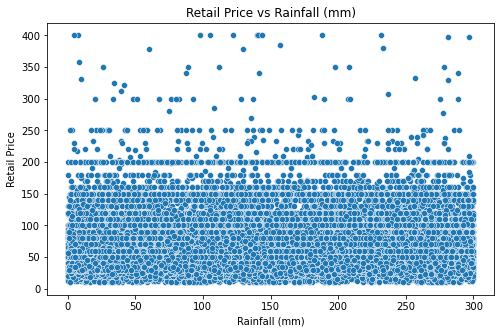
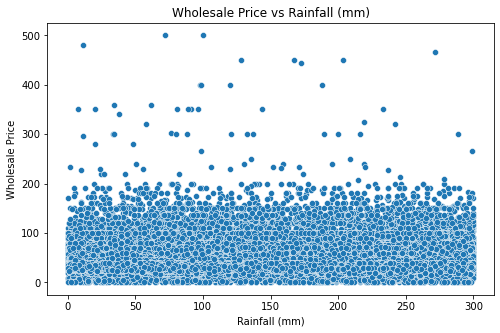
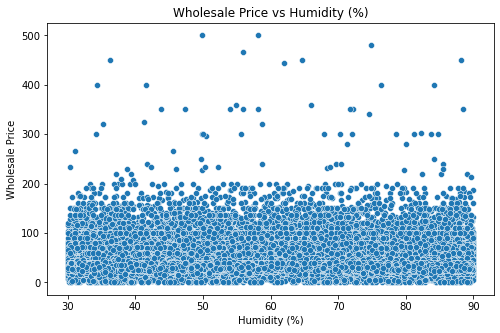
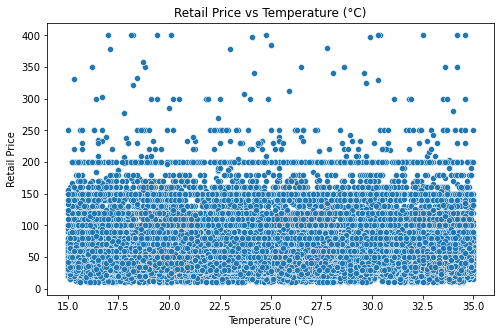
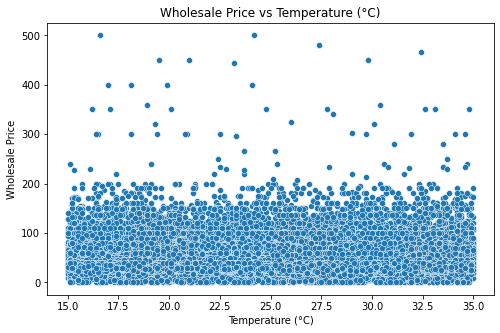
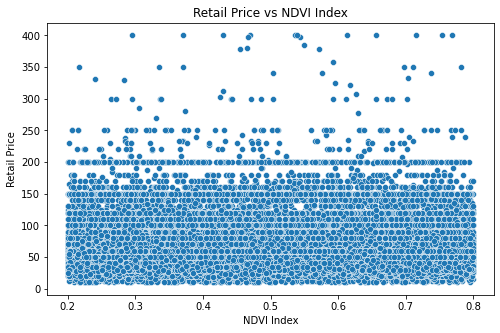
On the other hand, weak correlations between climatic factors (temperature, humidity, and rainfall) and food prices suggest that short-term weather variations have limited direct effects on pricing. However, long-term climate trends could still impact agricultural productivity. The negative correlation between inflation and NDVI implies that rising costs may limit investments in farm inputs, affecting vegetation health. Lastly, the exchange rate shows weak associations with most agricultural variables, suggesting that currency fluctuations primarily influence imported farm inputs rather than direct pricing within local markets.

* **Price vs. Other Variables**

****

****



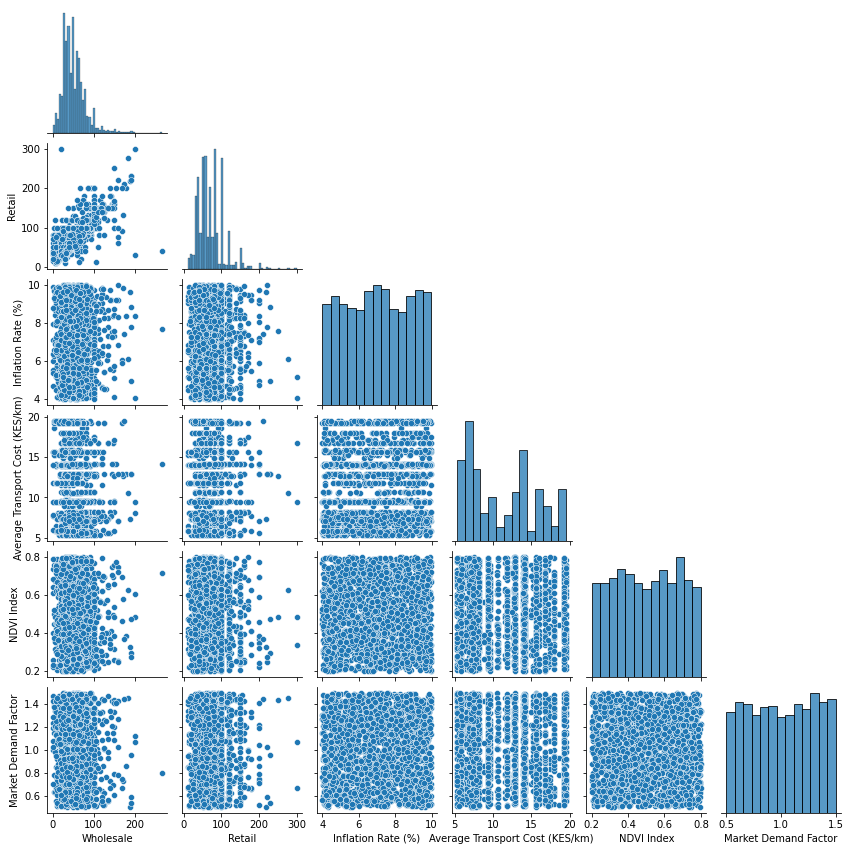


No single factor consistently explains price variations in the agricultural market. While food inflation is strongly linked to food prices, variables such as market demand, NDVI, temperature, humidity, and rainfall show little to no direct correlation with pricing, implying that their effects might be more indirect or long-term. Similarly, transport cost and road infrastructure quality do not exhibit clear, predictable trends with prices, suggesting that additional economic and logistical factors are at play. Inflation rate and fuel inflation, marked by some outliers, indicate that broader economic conditions can influence prices, whereas the exchange rate shows only a weak indirect impact through imported inputs. Overall, these findings highlight the complex interplay of multiple factors, where outliers may signal external shocks, supply chain inefficiencies, or regional disparities.

#### **5.2.2 Multivariate Analysis**

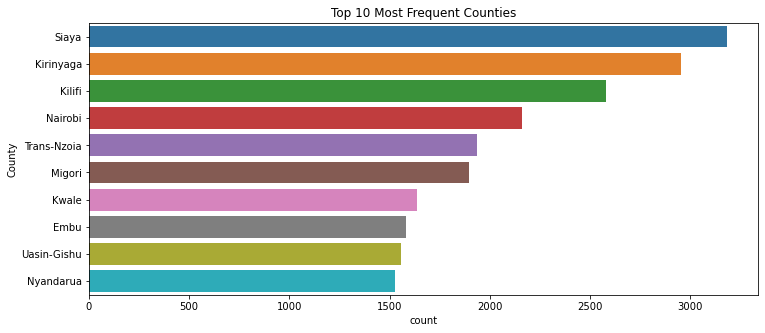
The objective of multivariate analysis is to explore and understand the complex interrelationships among several variables simultaneously. It helps in determining how different factors interplay to influence outcomes, guiding feature selection for deep learning models and refining predictive insights.

* **Pair Plots**

****

The pairplot offers valuable insights into how numerical variables interrelate. It shows that wholesale and retail prices are strongly correlated, meaning that when wholesale prices rise, retail prices tend to follow suit. In contrast, neither the Market Demand Factor nor the NDVI Index appears to have a strong direct link to price variations, implying that other elements might be influencing market prices. The relationships between transport costs and inflation rates with prices are more dispersed, indicating indirect effects. Additionally, the presence of outliers in price-related data points to occasional spikes driven by external economic or supply chain factors. Note that for clarity, only a subset of numerical columns was selected for this analysis.

* **Categorical Analysis: Market & County**

****

The bar charts display the top ten markets and counties based on frequency. In the market ranking, the most common ones are Aram, Kitale Municipality Market, Eldoret Main, Diani Market, Nakuru Wakulima, Embu Town, Daraja Mbili, Gongoni, Ngarubani Market, and Nyamakima, with Aram standing out as the clear leader. The other markets show relatively similar frequencies. Similarly, the county chart highlights Siaya, Kirinyaga, Kilifi, Nairobi, Trans-Nzoia, Migori, Kwale, Embu, Uasin-Gishu, and Nyandarua, with Siaya emerging as the most dominant, and the remaining counties exhibiting a gradual decline in frequency. These insights provide a clear picture of regional market activity and economic trends across different locations.

For further interactive exploration of our data analysis, please visit our Tableau dashboard available at [Tableau Dashboard Link].

### 5.3 Feature Engineering & Selection

**5.3.1 Feature Engineering**

This section outlines the steps taken to prepare and enhance the dataset for price prediction. The initial phase focused on feature engineering. First, temporal features were extracted from the date information. The date was broken down into components such as year, month, week number, day, and the day of the week. A binary indicator was then created to distinguish weekends from weekdays. Additionally, a custom seasonal classification was applied based on the month. For example, specific months were grouped under categories like "Long Rains," "Cold Season," "Short Rains," and "Dry Season." These seasonal labels were subsequently transformed into separate binary variables, allowing the model to capture seasonal patterns in agricultural prices.

Next, historical trends were incorporated by generating lag features. This involved calculating previous values for wholesale and retail prices over various periods (such as 1 day, 7 days, and 30 days) and computing rolling statistics (mean and standard deviation) over 7-day and 30-day intervals. These features are designed to capture both short-term and medium-term trends that may influence future prices. To ensure continuity, any missing entries resulting from these operations were filled using methods that propagate previous or subsequent values.

Further enhancements included the creation of market-based features. These were developed to capture the dynamics of individual markets by computing the percentage change in market demand, monthly average prices, and measures of price volatility. Direct comparisons between wholesale and retail prices were also derived by calculating both the difference and the ratio of these values. In parallel, crop-specific features were generated for different crops (such as avocado, tomatoes, and maize) by establishing lag features over one-month and three-month intervals and calculating seasonal average prices. Any gaps in these crop-related trends were addressed by carrying forward the last observed value or substituting with the crop-specific average.

Once the feature engineering was complete, categorical attributes like Crop, Season, Market, and County were transformed into numerical representations through encoding processes. The original textual columns were then removed, and the mappings from the transformation were preserved for future use.

The final phase involved examining the relationships among all features to identify and reduce multicollinearity. A statistical analysis was performed to determine the variance inflation factor (VIF) for each variable. Variables with excessively high VIF values, which indicate strong correlations among themselves, were removed from the dataset. This step ensured that the remaining features were more independent and less redundant, thereby enhancing the stability of subsequent predictive models.

To assess the relevance of the engineered features for predicting prices, ensemble regression techniques were applied separately for wholesale and retail price predictions. The importance of each feature was evaluated, revealing that factors such as the difference and ratio between wholesale and retail prices were particularly influential. Based on these insights, features with negligible impact—like the indicator for weekends—were dropped from the final dataset. The refined dataset, which now contains both the selected predictors and target variables for wholesale and retail prices, was then saved for further modeling and prediction tasks.

**5.3.1 Feature Selection**

The objective in this section is to refine the set of predictors that will be used for a market recommender system. The approach begins by considering all the available features from the cleaned dataset as inputs and using the market identifier as the target outcome. A Random Forest classifier, known for its ability to handle complex interactions and rank feature importance, is employed to assess how strongly each predictor contributes to correctly identifying a market. The reasoning behind this is that by understanding which variables most influence the market categorization, we can better focus on the factors that truly drive market behavior.

After the classifier is trained, the features are ranked according to their importance scores. This ranking process reveals that certain factors—such as retail and wholesale volatility, average prices, and key identifiers like Market\_ID and County\_ID—play a crucial role in the prediction process. On the other hand, variables such as the weekend indicator, along with some weather-related factors (rainfall, temperature, humidity, and NDVI Index), exhibit lower importance. The reasoning for removing these less relevant features is to reduce noise and avoid unnecessary complexity in the model, ultimately leading to a more efficient and interpretable system. The final, streamlined dataset containing the most pertinent features is then saved, ensuring that future predictive tasks are built on a robust and focused set of inputs for market recommendation.

### 6. Notebook 2 – Modeling & Evaluation

### 6.1 Price Prediction Model

For price prediction, the process began with a carefully prepared dataset enriched with temporal features, lag variables, rolling statistics, and market-specific indicators, which formed the basis for developing two deep learning models—one for wholesale prices and one for retail prices. The model architectures were optimized using an automated hyperparameter tuning process, specifically the Hyperband algorithm from the Keras Tuner library. This approach defined a search space that included the number of units in dense layers, dropout usage and rates (ranging from 0.1 to 0.5), the option to add extra layers, and learning rates sampled logarithmically between 1e-4 and 1e-2. Hyperband dynamically allocated training resources by evaluating models based on the validation mean absolute error (val\_mae), allowing promising configurations to train for more epochs while quickly terminating those with poor performance. The data was split into training and test sets to ensure robust evaluation.

For the final models, the Wholesale Price Model achieved a Mean Absolute Error (MAE) of 4.56, a Root Mean Squared Error (RMSE) of 7.89, and an R² score of 0.92, while the Retail Price Model achieved an MAE of 5.66, an RMSE of 9.47, and an R² score of 0.92. These metrics indicate that both models are highly effective in capturing the underlying trends in the data. Additionally, to better understand the impact of individual features on the predictions, a sample of 100 observations was analyzed using SHAP-based explanations, which generated summary plots that highlight each predictor's contribution.

### 6.2 Market Recommender Model

For market recommendation, the focus shifts to predicting the correct market based on a distinct set of features. The process begins with a separate, cleaned, and encoded dataset that is balanced using synthetic oversampling to ensure fair representation across different market classes, followed by dimensionality reduction with Principal Component Analysis (PCA) after standard scaling. A deep learning classifier is then constructed with multiple dense layers, enhanced by techniques such as batch normalization and dropout, and outputs probabilities across market classes via a softmax layer. The model is trained with early stopping and evaluated using classification accuracy as well as detailed precision, recall, and f1-scores for each market class. The final model achieved an impressive overall accuracy of approximately 97.22% when mapping predicted market identifiers back to market names, demonstrating that it effectively captures the complex relationships in the data that determine market dynamics. Finally, predictions from the market recommendation model, along with wholesale and retail price forecasts generated by their respective models, are saved as CSV files, and all trained models and their preprocessing pipelines are stored for deployment in a real-time decision support system.

### 7. Deployment We developed a user-friendly interface by combining FastAPI and Streamlit. FastAPI handled the backend logic, providing APIs for price predictions and market recommendations by processing user inputs and fetching data from the saved prediction csv files. Streamlit was used to create an intuitive frontend where users could interact with the application, selecting crops, counties, and price types to receive predictions and market suggestions. We deployed the app on Streamlit Community Cloud by linking the project from GitHub, enabling easy access for farmers through a public URL. This approach allowed us to seamlessly integrate backend calculations with a simple, interactive UI for farmers.

Here is the link to the user-friendly interface at:<https://agrisenseai-project-4wckgd8cno4p4vjuc2rhme.streamlit.app/>.

**8. Conclusions**Based on our extensive development and rigorous evaluation, the AgriSenseAI system has achieved significant success in addressing the challenges faced by Kenyan farmers and traders. Below are the key outcomes:

* Successfully developed an integrated deep learning system that leverages historical price data, weather information, vegetation indices, supply-demand dynamics, transport costs, and macroeconomic indicators to enhance market decision-making.
* Achieved high predictive accuracy in price forecasting, with the wholesale model recording an MAE of 4.56, RMSE of 7.89, and R² of 0.92, and the retail model obtaining an MAE of 5.66, RMSE of 9.47, and R² of 0.92.
* Constructed a robust market recommender that effectively predicts optimal market locations, achieving an overall classification accuracy of approximately 97.22%.
* Delivered a user-friendly interactive dashboard using Streamlit, enabling farmers and traders to easily access real-time price forecasts, market recommendations, and visual insights.
* Deployed the system on Streamlit Community Cloud, ensuring a scalable, reliable production environment that supports seamless real-time decision support for the agricultural sector in Kenya.

**9. Recommendations**

* **Enhance Data Collection:** Collaborate with government agencies and data providers to secure real-time, granular meteorological data.
* **Expand Crop Coverage:** Extend the system to include a broader range of crops to improve the robustness of market recommendations and price predictions.
* **Improve Model Interpretability:** Integrate advanced explainable AI techniques (e.g., SHAP analyses) to continuously validate model insights and build user trust.
* **Strengthen Infrastructure:** Refine the user interface and deployment strategy to ensure the system remains scalable and responsive, especially as additional data streams are integrated.

**10. Challenges**

* **Real-Time Price Forecasting:** Encountered difficulties in establishing a reliable connection with the KAMIS website for live price updates.
* **Data Collection Challenges:** Faced issues with sourcing relevant data due to limited centralized data availability, necessitating the integration of multiple data sources.
* **Meteorological Data Access:** Inability to acquire comprehensive, high-resolution climate data from the meteorological department limited the accuracy of climate-based predictions.
* **Data Quality and Consistency:** Addressing inconsistencies, missing values, and varying data formats required extensive cleaning and preprocessing efforts.
* **Integration of Heterogeneous Data Sources:** Merging and harmonizing diverse datasets from various governmental and institutional sources proved complex and resource-intensive.

**11. Next Steps**

* **Model Refinement:** Update and retrain the deep learning models by incorporating additional features and more diverse data sources to further improve prediction accuracy.
* **Real-Time Data Integration:** Establish a connection with the KAMIS website to integrate live price data into our system, enabling dynamic and up-to-date price forecasting.
* **User Interface Enhancement:** Continue refining the Streamlit dashboard with advanced visualization and filtering options to enhance user experience and accessibility.
* **Pilot Deployment and Feedback:** Initiate a pilot deployment with selected farmers and traders to gather real-world feedback and validate system performance.
* **Monitoring and Maintenance:** Implement a robust monitoring framework to continuously track model performance, data quality, and system uptime, with regular updates and retraining based on new data trends.
* **Process Review:** Conduct periodic reviews of the entire pipeline—from data collection through deployment—to identify and implement improvements that adapt to evolving market dynamics.