CSE3505 Foundations of Data Analytics J-Component

**Project Title: Food price forecasting**

Team Members

**Sanjeev N S 20MIA1058**

**Sriharish R 20MIA1027**

## S P Mithun 20MIA1038

# Abstract

The objective of this study is to develop the time series model so the model can give a forecast value of rice price using various models. Forecasting food price is an action to help government in monitor and control it. Price monitoring is conducted to achieve the good development of domestic trade. In this case, price monitoring is done to maintain the price stability so that it will not harm producers and consumers.

## Table of Contents

[Abstract 1](#_Toc3734)

[1. Background 3](#_Toc3735)

[2. Problem Statement 4](#_Toc3736)

[3. Aim and Objectives 4](#_Toc3737)

[4. Significance of the Study 4](#_Toc3738)

[5. Scope of the Study 4](#_Toc3739)

[6. Research Methodology 5](#_Toc3740)

[7. Experimental Analysis 5](#_Toc3741)

[References 8](#_Toc3742)

# Background

Predicting the prices of essential food commodities like chickpeas, oil, mustard, potatoes, rice, sugar, wheat, and onions is of significant importance due to its far-reaching impact on various aspects of society. These predictions have critical implications for global food security, economic stability, and individual livelihoods. For instance, predicting chickpeas and other staples is vital for farmers and agribusinesses to make informed planting and investment decisions. The prices of edible oils, like mustard, impact both the cost of food production and consumers' budgets. The cost of potatoes, rice, sugar, wheat, and onions is closely linked to the affordability of staple diets, affecting the nutrition and well-being of populations. Predictions enable governments and policymakers to respond proactively to potential food crises and implement measures to mitigate the effects of price fluctuations. Accurate forecasting is also crucial for international trade, aiding in the management of food imports and exports, and it influences investors' decisions in the agricultural and food industry.

Food price prediction is a crucial and multidimensional aspect of the global food supply chain, which plays a significant role in the daily lives of people around the world. It involves the use of data analysis, statistical modeling, and machine learning techniques to anticipate and forecast the future prices of various food products. This predictive endeavor is driven by a multitude of factors, including economic conditions, weather patterns, agricultural practices, transportation costs, consumer demand, and geopolitical events.

The use of advanced technologies, such as big data analytics and artificial intelligence, has greatly enhanced the precision and timeliness of food price predictions. These tools allow for the analysis of vast datasets, including historical price records, weather patterns, market trends, and geopolitical developments. With this information, machine learning algorithms can generate forecasts that help businesses optimize their operations, governments make informed policy decisions, and consumers plan their budgets.

In this era of global interconnectedness and climate change, accurate food price prediction is more critical than ever. It not only contributes to the sustainability of the food supply but also has far-reaching implications for economic stability, social well-being, and environmental conservation. As such, ongoing research and innovation in the field of food price prediction remain essential to addressing the challenges and opportunities associated with food security and market stability.

# Problem Statement

Develop a food price forecasting model for India that can predict the prices of various food commodities over a specific time horizon (e.g., monthly or quarterly)

# Aim and Objectives

The objective of this project is to build and find a optimal predictive model for forecasting food prices in India. This model will provide reliable price forecasts for various food commodities, allowing stakeholders to make informed decisions and plan accordingly.

# Significance of the Study

The study on food price prediction using various models, including ARIMA, SARIMA, exponential smoothing, and regression models, holds significant importance in the fields of economics and agriculture. These models collectively provide valuable tools for accurate food price forecasting, which is of utmost importance to various stakeholders.

Farmers can utilize these models to make informed decisions about planting and production, reducing the risks associated with overproduction or food shortages. For policymakers, having multiple models at their disposal offers the flexibility to proactively respond to potential food crises and to implement measures that stabilize prices and ensure food security. Additionally, consumers benefit from the ability to access more predictable and affordable food prices, contributing to their economic well-being and nutritional access.

The significance of employing multiple models, including ARIMA, SARIMA, exponential smoothing, and regression, lies in their capacity to analyze and model time series data effectively. By considering historical food price trends, these models enable more precise predictions, which can significantly impact the planning and decision-making processes. Moreover, their application can reduce market uncertainties and facilitate more efficient resource allocation, promoting economic stability, and enhancing food supply chain management.

In summary, the study on food price prediction using a range of models, including ARIMA, SARIMA, exponential smoothing, and regression, has far-reaching implications for food security, economic stability, and informed decision-making across the agricultural and food industries. The goal is to determine the best model among these options for accurate forecasting, ensuring that the results are reliable and beneficial to all stakeholders involved.

# Scope of the Study

The scope of this study, which incorporates multiple models, including ARIMA, SARIMA, exponential smoothing, and regression, is both extensive and adaptable. It encompasses a diverse array of food commodities and markets, allowing for the analysis of essential staples, oils, and various other food items. The scope extends to regional and global levels, enabling a comprehensive assessment of the distinct challenges and variables affecting different markets and regions. This research serves as a valuable resource for governments, businesses, and organizations involved in the food supply chain, offering insights into price dynamics, market trends, and strategies for risk management.

Furthermore, the scope of this study expands to encompass the integration of a wide range of data sources and factors that influence food prices. This includes economic indicators, weather patterns, consumer demand, and geopolitical events, among others. It allows for the development of adaptable and accurate models that can evolve with changing circumstances. Additionally, the study can serve as a foundation for further research into advanced predictive techniques, broadening the scope of knowledge in the field of food price forecasting. This research may lead to innovative solutions for addressing food security challenges on a global scale while identifying the best model among the options considered for the most reliable forecasting outcomes.

# Research Methodology

Methodology deployed involves key processes such as the selection of target data, preprocessing the chosen data, transforming the data into a structured and comprehensible format, balancing the dataset, implementing supervised learning techniques and evaluating the machine learning performance using evaluation measures. These steps comprise a comprehensive methodology for conducting the study on food price prediction using the ARIMA, SARIMA, exponential smoothing, and regression models.

The first crucial step involves the selection of target data, where researchers identify the specific food commodities, geographical regions, and time periods to be analyzed. This selection process ensures that the study is focused and relevant to the problem at hand.

Following this, the pre-processing of chosen data is undertaken to clean and prepare the data for analysis. This includes addressing missing values, outliers, and any inconsistencies in the dataset, ensuring that the data is of high quality and suitable for analysis.

The next phase involves transforming the data into a structured and comprehensible format. Data transformation might include feature engineering to create new variables, aggregating data to different time intervals, or scaling numerical values. This step is critical to facilitate accurate modeling.

Furthermore, balancing the dataset is a necessary step to address any class imbalances, ensuring that the model is not skewed toward predicting one outcome more than others. This helps maintain the model's predictive accuracy across different food commodities.

The implementation of supervised learning techniques is a key part of the methodology. Researchers use ARIMA, SARIMA, exponential smoothing, and regression models to make predictions based on historical data and various exogenous factors influencing food prices.

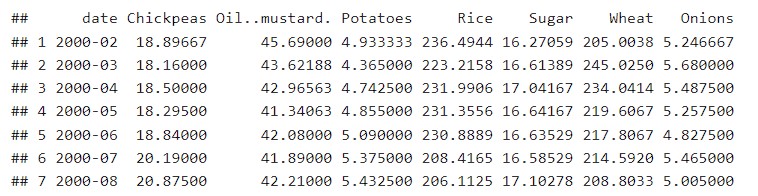
Finally, evaluating all the machine learning performance using appropriate evaluation measures is essential to assess the model's accuracy. This step enables researchers to determine the model's effectiveness and its ability to provide reliable food price predictions.

This part is done in 2 ways. First RMSE (Root Mean Squared Error) of all four models are found and select the best model for predicting food prices based on the RMSE values. Second error rate is found for the traditional forecasting models (ARIMA,SARIMA,exponential smoothing).error rate is found by taking difference of forecasted values of year 2019 with validation dataset which already has prices of the year 2019. Based on less error rate ,model is chosen

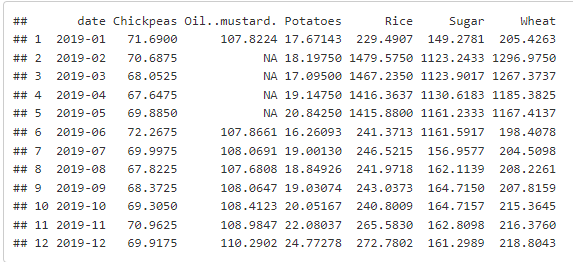
In conclusion, this comprehensive methodology covers the critical processes involved in conducting a study on food price prediction, ensuring that the research is based on sound data selection, preprocessing, modeling, and evaluation principles.

# Experimental Analysis

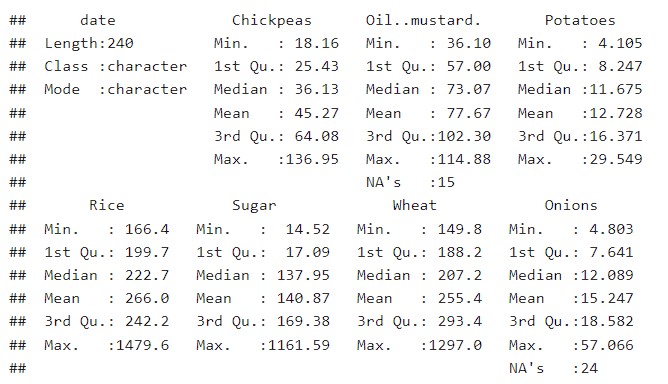
Here is the data that will be used in our prediction



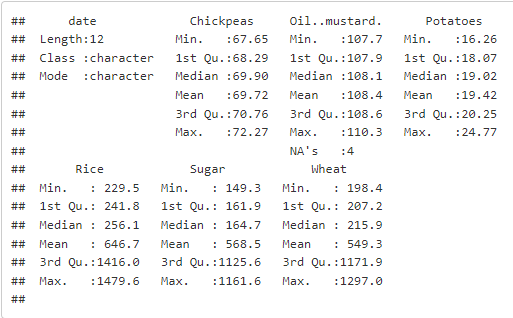
Validation dataset



Summary statistics of the data



Validation dataset



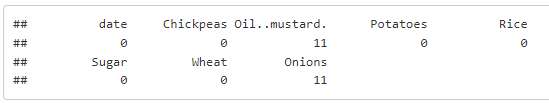
The summary statistics data presents a snapshot of food commodity prices over a specific time period. The data is structured in a tabular format, with columns representing various food items, such as Chickpeas, Mustard Oil, Potatoes, Rice, Sugar, Wheat, and Onions. It also includes the 'date' column, which presumably indicates the time or date of each recorded price.

The data provides valuable insights into the distribution of prices for each food commodity. Notably, it includes summary statistics, such as minimum, maximum, and quartiles, which offer a glimpse into the price ranges and variations of these commodities. For instance, Chickpeas have a minimum price of 18.16 and a maximum of 136.95, suggesting a significant price range. Similarly, Rice has a remarkably high maximum price of 1479.6, indicating potential outliers or extreme price fluctuations.

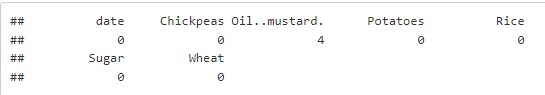
It's also evident from the data that some values are missing, denoted by "NA's," which might require further data cleaning and imputation to ensure the accuracy of any subsequent analyses or predictions.

In summary, this data provides a foundation for conducting a comprehensive analysis of food commodity prices. It reveals the range and distribution of prices for various commodities, enabling researchers or analysts to explore trends, correlations, and potentially construct predictive models for these essential food items. However, addressing missing data and conducting further exploratory analysis would be necessary before any in-depth insights or predictions can be drawn from this dataset.

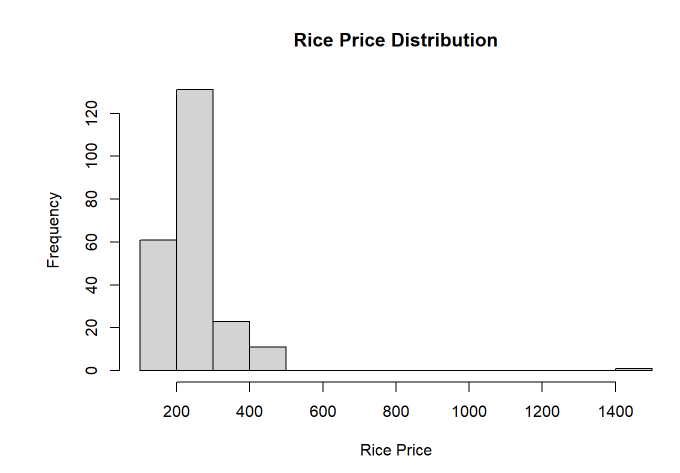
**Missing Value analysis**

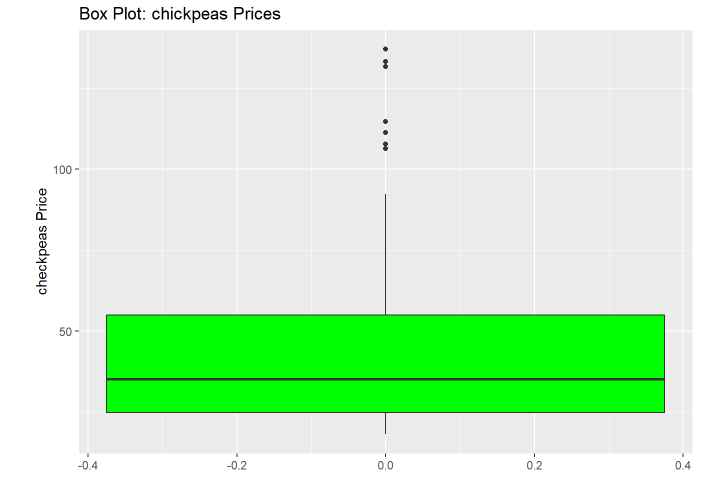
****

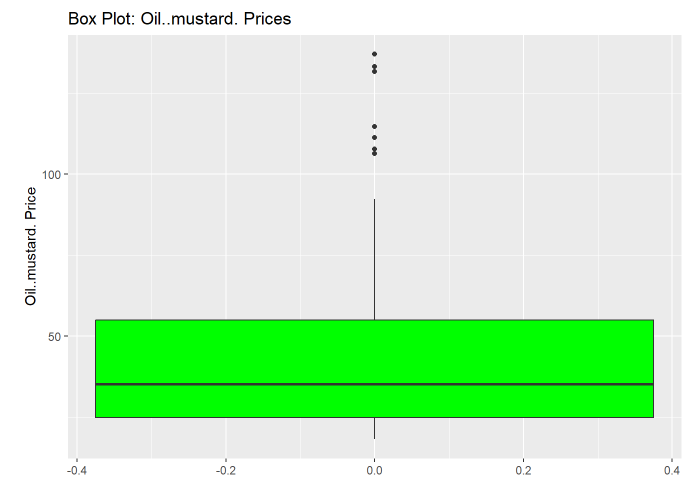
Validation dataset

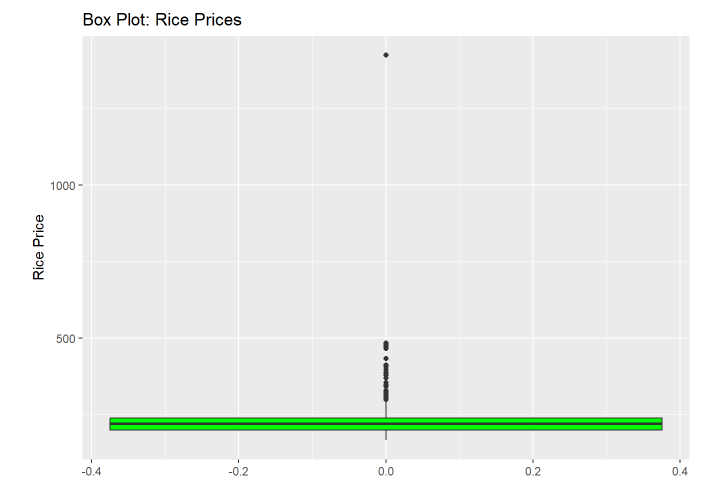
****

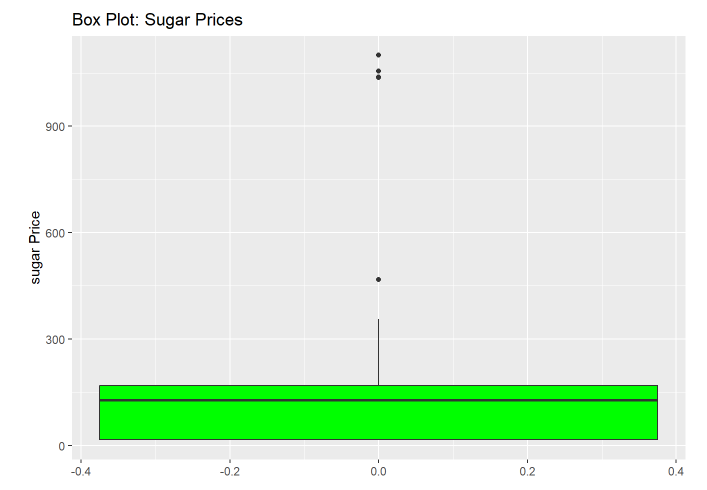
**Sample Visualization of onion price distribution and other prices**





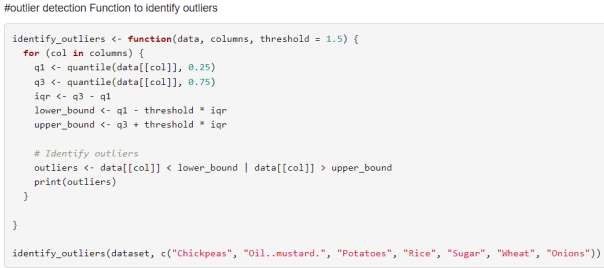




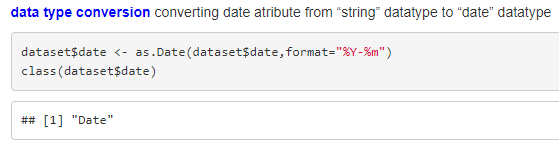


From these visualizations, we can infer that except Oil. Mustard price, all other price has outliers in it.next part in EDA is to identify outliers and impute it with mean

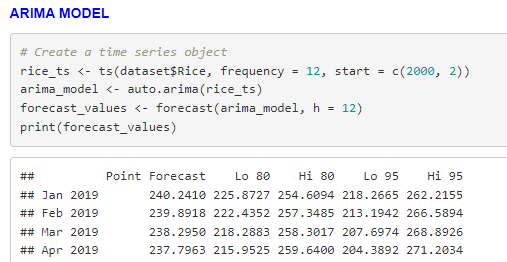
Detecting Outliers

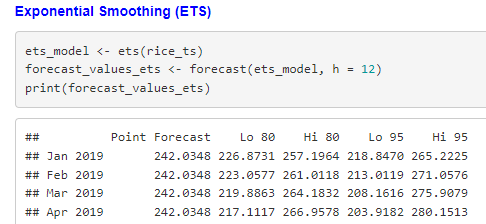


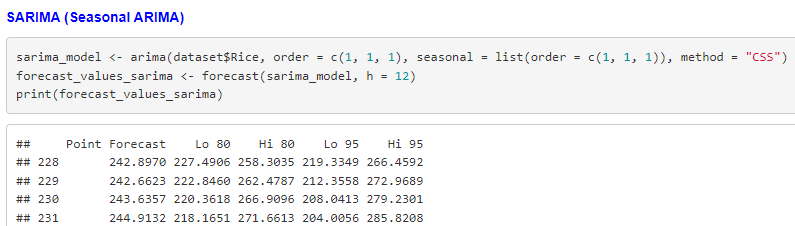


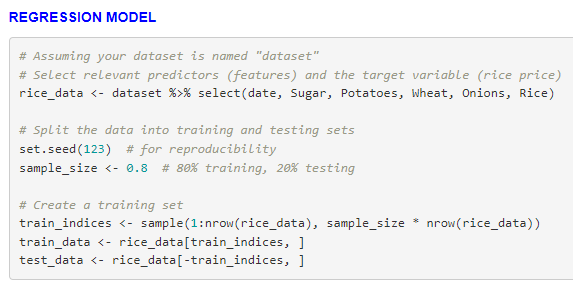


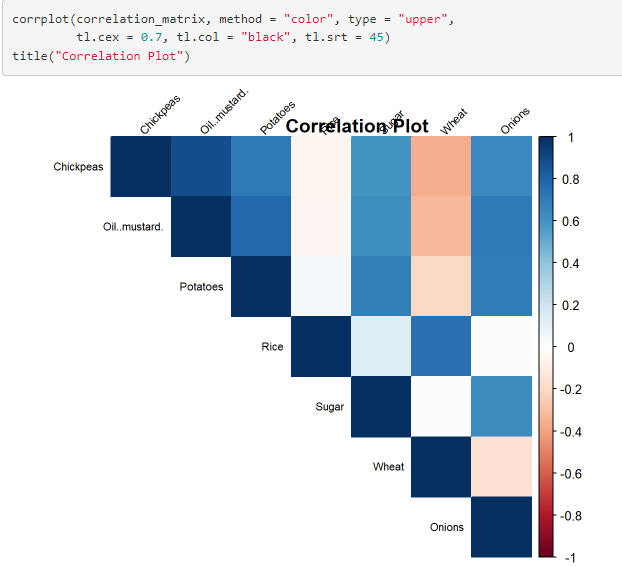
This convert the string data type to date data type.



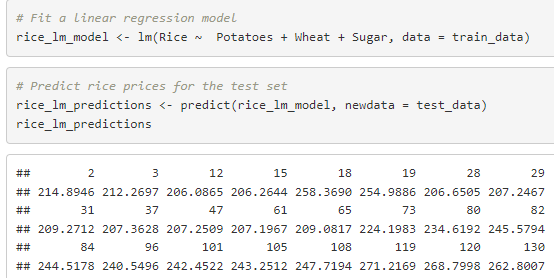


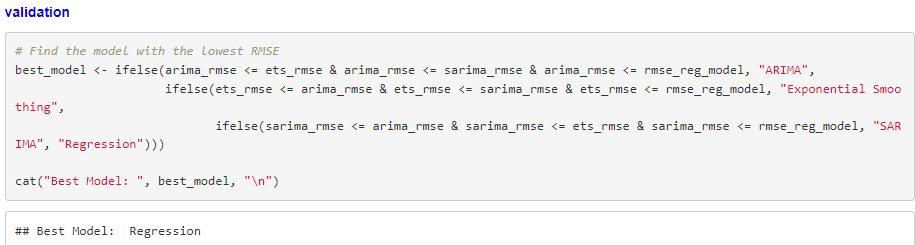


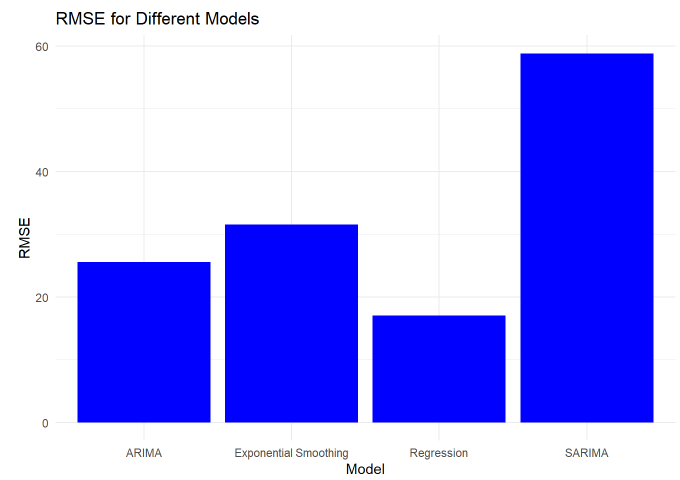




From the correlation matrix, for predicting rice , we can use potatoes, sugar, onion and Wheat as independent variable as they are positively correlated

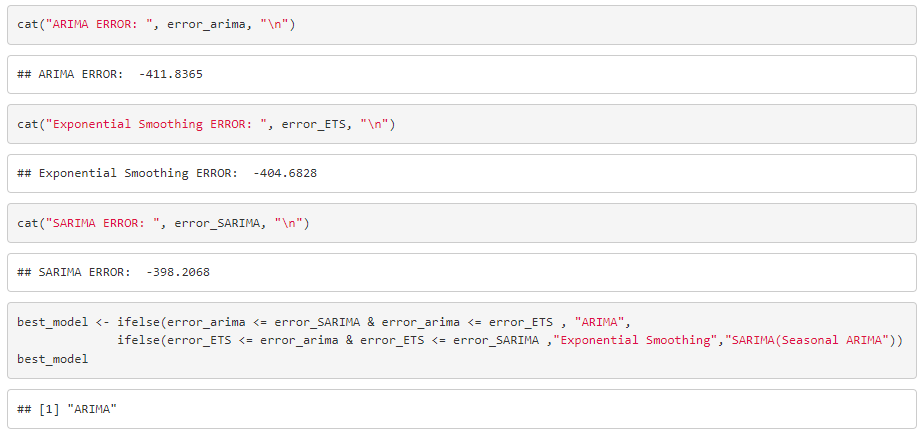


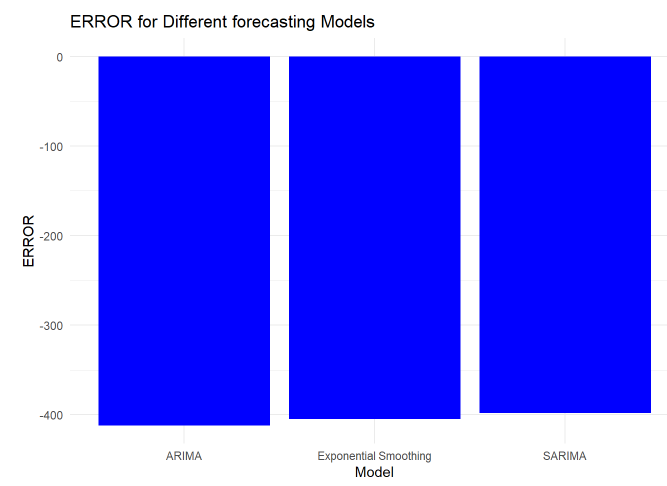




We can absovere visually that regression model has less rmse than other models,so regresson model is a best model to predict the prices

**Comparision of forecasting models**





We can observe visually that ARIMA model has less error than other models, so ARIMA model is a best model to forecast the prices

# Conclusion:-

Based on many experimental analysis on the dataset, regression model comes out as an best model for predicting food prices based on other prices. ARIMA comes out to be a best model in traditional forecasting models. Also if more model is needed, models like Prophet, Autoregressive (AR) Models, and Dynamic Regression Models are recommended

# References

1. Wang, Y., Ye, X. and Huo, Y., 2011, July. Prediction of household food retail prices based on ARIMA Model. In 2011 International Conference on Multimedia Technology (pp. 2301-2305). IEEE.
2. Ohyver, M. and Pudjihastuti, H., 2018. Arima model for forecasting the price of medium quality rice to anticipate price fluctuations. Procedia Computer Science, 135, pp.707-711.
3. Li, S. and Li, R., 2017. Comparison of forecasting energy consumption in Shandong, China Using the ARIMA model, GM model, and ARIMA-GM model. Sustainability, 9(7), p.1181.
4. Xia, Y., 2007. Construction and application of ARIMA models in seasonal agricultural product price prediction [J]. Science and Comprehensive Research of Agricultural System, 23(1), pp.8992.
5. Ndunagu, J.N., Aderemi, E.H., Jimoh, R.G. and Awotunde, J.B., 2022, November. Time Series: Predicting Nigerian Food Prices using ARIMA Model and R-Programming. In 2022 5th Information Technology for Education and Development (ITED) (pp. 1-6). IEEE.