University of Essex

School of Mathematics, Statistics and Actuarial Science

MA981 DISSERTATION

Forecasting Wholesale Meat Prices: A Time Series Analysis On USDA Sub Meat-Type Price Trends

BHARATH YUVARAJ
2316529

Supervisor: **Dr Tao Gao**

September 18, 2024
Colchester

Abstract

Agricultural Commodity price forecasting is critical as people like stakeholders in food and agriculture industries, are informed about the decisions in production, trading, and policy making. This study is more on time series analysis of wholesale meat prices in the United States and the data is from the U.S. Department of Agriculture (USDA) ranging from 2000 to 2024. The research employs different kinds of forecasting models, such as Seasonal Autoregressive Integrated Moving Average (SARIMA), Random Forest, Gradient Boosting, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) neural network, to determine the most effective methods for accurate predictions.

In this study external factors that might be affecting the meat price were implemented, such as weather data and economic indicators. By comparing the predictive R2 score and RMSE error on different models, we aim to identify the most reliable methods for forecasting meat prices in the U.S. market

Our findings reveal that the LSTM model with external features generally outperformed other models, achieving the highest R2 score (0.8962) and the lowest RMSE (0.01571) for the 'Boxed Beef Cutout' subtype. Similarly, the SVR model without external features showed strong performance for the 'Imported Boneless Beef, 90 Percent, Frozen' subtype, while the Gradient Boosting model demonstrated robustness across multiple subtypes, particularly for Loins, 1/4" Trimmed Vacuum-Packed and Drumsticks. This research provides more insights into effective forecasting methods and understanding the practical implications for market participants and policymakers in the agriculture sector.

Keywords: *Time series, USDA, Forecasting, Wholesale Meat Price, Meat Sub types, CPI, LSTM, SVR, SARIMA, Random Forest, Gradient Boosting, R2 Score, RMSE.*

Acknowledgement

I would like to express my sincere gratitude and warm thanks to Dr. Tao Gao, my supervisor, for his invaluable guidance, support, and encouragement throughout this dissertation. His expertise and insights along with constructive feedback have been instrumental in shaping the direction and outcomes of this research. I would like to thank the University of Essex and the Department of Mathematics, Statistics, and Actuarial Science for their invaluable support throughout my dissertation.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

Agriculture is the backbone of human civilization and has played a very important role in the development of different societies worldwide, as it is one of the oldest human activities. The term agriculture oscillates from the cultivation of crops and rearing of livestock, providing food production, fiber, medical plants, and other products that are essential to human health. Agriculture farming in small sectors is now leading to multi-industrial-scale production systems feeding billions of people around the world. Technological advancements, scientific research, and socio-economic factors make agriculture a vital sector for food security and contribute to socio-economic development and environmental awareness.

Meat Type	**Meat Sub-Type**
Beef, Central U.S.	Boxed beef cutout / select 1-3 (600-900) lbs
Beef, Central U.S.	Imported boneless beef, 90 percent, frozen
Pork, Central U.S.	Pork cutout composite
Pork, Central U.S.	Loins, 1/4", trimmed vacuum-packed
Turkeys, Central U.S.	Drumsticks
Turkeys, Central U.S.	Wings, full cut

Table 1.1: Different Meat Type and Sub-Type

The wholesale meat market plays an important role in global food sectors, impacting consumer costs and a wide range of the agricultural economy. In past years the wholesale meat prices of various meat cuts have been completely fluctuating due to major shifts in supply and demand, economic policies global market trends, and unexpected events like pandemics, which introduce sudden and unpredictable price shocks. For example, the COVID-19 pandemic has profoundly impacted global commodity markets, including the meat industry, leading to unprecedented price volatility. To forecast this price accurately, the United States Department of Agriculture (USDA) always monitors these prices, providing essential data that help stakeholders make informed decisions on schedules and pricing strategies, stabilizing the supply chains and reducing waste from meat. Thus, the nature of wholesale meat price is very important, associated with price volatility, and enhances market stability. To enhance the stability of meat prices few authors [1] used dynamic ensemble learning and spectral clustering to forecast beef and lamb prices, the authors used analytical methods in handling price volatility [1]. Similarly, [4] analyzed meat price volatility and spillovers in Finland, the importance of understanding regional factors for accurate forecasting [4]. [6] examined global meat consumption trends and their economic relationships, providing insights into the broader economic forces affecting meat prices.

This research was done on the dataset USDA Livestock and Meat Domestic Data which gives more detailed information including historical monthly prices from 2000 January to 2024 April with three meat types along with six sub-meat cuts as shown in table 1.1, allowing for a comprehensive analysis across different meat types providing a valuable resource for developing the predictive models. The analysis was done on advanced forecasting predictive models like the seasonal Auto Regressive Integrated Moving Average (SARIMA) model captures both non-seasonal to predict future price trends based on historical data. Machine learning predictive models such as Random Forest, Gradient Boosting, and Support Vector Regression (SVR) are very powerful and can handle very large datasets and capture complex, non-linear relationships between different kinds of variables. Whereas Long Short-Term Memory (LSTM) is a type of deep learning model, that works well in sequential data analysis and long-term dependencies, which are important for accurate forecasting. This report includes a different pricing pattern, forecasting models, and an evaluation of potential future trends. This approach is informed by recent works such as [2] forecasted pork prices, using an LSTM-based forecasting model that incorporates price fluctuation forecasting and attention mechanism to improve model accuracy [2]. [9] reviewed methodologies for food price forecasting, offering insights into effective approaches for modeling meat prices [9].

While doing more research on this meat price forecasting identified the complex interplay between internal market dynamics and external factors. By integrating multiple external features such as feed grain prices, climate data, and macroeconomic indicators like inflation and unemployment rates into these predictive models, researchers have developed more accurate forecasts even in significant market disruptions. The study by [15] utilizes multiple regression and machine learning methods to predict meat cuts and carcass traits, reflecting the complexity of the factors affecting meat prices [15]. Similarly, external economic conditions like unemployment rates and disaster occurrences contribute to price volatility, as discussed by previous research [14]. The integration of these external features not only enhances the predictive power of these models but also provides a better understanding of the factors driving meat prices in a globalized economy. [8] applied deep learning techniques to analyze the impact of COVID-19 on commodity markets, illustrating how external shocks can influence forecasting outcomes [8].

The structure of this research paper follows as: The introduction provides a brief analysis of wholesale meat price background, objectives, significance, and overview of this dataset. The Literature Review of existing research on different meat price forecasting and the time series. The methodology section outlines the research design, data sources (USDA sub-meat prices), and various forecasting models used for this analysis. The Result section tells the best-performing time series analysis models and insights on price trends. The Conclusions talk about key findings contributing to the field and talk about future works. References give additional documentation and links to the research paper.

This analysis reveals significant patterns in wholesale meat prices for selected cuts from 2000 January to 2024 April. This analysis will give insights into future price movements and their impact on the industry. By understanding these trends, the stakeholders can participate in market price analysis, along with strategies and manage risk related to price volatility and supply chain disruptions

CHAPTER 2

Literature Review

2.1 Advanced Forecasting Techniques

Yuan *et al.* [1] found challenges in predicting beef and lamb prices due to the high volatility. The authors applied spectral gathering on prices of beef and lamb data for grouping similar patterns which allowed the model to treat each cluster differently for improving the accuracy. They followed to combine several machine learning models (SVR, LSTM, Xgboost, and LightGBM) into the dynamic ensemble. This ensemble was adjusted based on the performance of the individual models on each cluster. later this ensemble learning model was trained on the final clustered data to adapt different price behaviours within each cluster to make up for the drawbacks of the lagging market of Qinghai's livestock industry. To conclude this paper spectral clustering with ensemble learning gives very good forecasting values with unstable agriculture prices.

Zhao *et al.* [2] developed an LSTM-based pork price forecasting method that uses price fluctuation forecasting and an attention mechanism to improve accuracy. To forecast pork prices, the authors experimented with various combinations of machine learning and deep learning algorithms, such as Xgboost, SVM, Random Forest, LSTM, and BERT. The LSTM model is optimized and accuracy is increased through the use of the attention mechanism. The suggested LSTM-based method outperforms other models in forecasting fluctuations in pork prices, with RMSE = 1.57, MAE = 1.28, and MAPE = 2.83, according to the results. The extension of fluctuation forecasting and an attention mechanism enhances the forecasting effect. Understanding how to apply complex techniques to the forecasting process is made easier with the help of these insights.

Amalia *et al.* [7] discussed which of the two RNN variants LSTM and GRU is a better forecasting model for predicting the daily prices of three important Indonesian commodities: rice, chicken eggs, and broiler meat. Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were used in the author's evaluation along with techniques such as LSTM and GRU models with the same hyperparameters. Based on two evaluation standards, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), the outcomes showed that the GRU model achieves higher accuracy in predicting the daily prices of rice, broiler meat, and chicken eggs. Besides that, the GRU model completed the computation 20 seconds faster than the LSTM.

Fahrudin *et al.* [11] inform the planning strategies for DKUKMPP, which stands for Dinas Koperasi Usaha Kecil Menengah Perdagangan dan Perindustrian Kota Cirebon, or the people of Cirebon City, Fahrudin provided accurate predictions of the prices of basic food commodities. The five important interrelated stages of the waterfall software development process observation, literature study, design, implementation, verification, and maintenance as well as the Long Short Term Memory (LSTM) method, were employed. The study leads to the fact that the DKUKMPPs decision-making can be assisted by using the LSTM algorithm to forecast the prices of staple food commodities in Cirebon City.

Anderson *et al.* [15] Using non-invasive in vivo measurements, Anderson studied the effectiveness of multiple linear regression and machine learning techniques to predict commercial meat cuts and carcass traits in lambs. To predict carcass attributes and commercial meat cuts, the author employed support vector machines, Bayesian networks, multiple linear regression, and artificial neural networks. Overall, the feature selection process is required to obtain more accurate results, regardless of the model applied, and the support vector machine algorithm is a valuable technique for carcass traits and commercial meat cuts prediction. Moreover, the size of the data in the study indicates that when used as a pre-selection tool for input variables in support vector machine models, the Bayesian network model is a helpful tool for offering better prediction accuracy.

2.2 Price Volatility and Market Dynamics

Abdallah *et al.* [4] monitored the volatility overflow along the Finland meat chain and looked into the price volatility of producer and consumer meat prices. The volatility of the first price level to the second price level is analyzed by the authors using the Generalized Autoregressive Conditional Heteroskedasticity and Kroner (GARCH-BEKK) model. The author also attempted to check for stationary behaviour using the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests, as well as autocorrelation and heteroskedasticity using the ARCH test. The study concludes that the BEKK-GARCH model captures the volatility interconnections between producer and consumer prices and is appropriate for explaining the conditional variance of the data. The results have implications for our information on the volatility of meat prices and the necessity of government intervention to safeguard the retail and farm sectors.

Xu *et al.* [5] In Yunnan Province, China, the author looked into the price fluctuations of the hog industry chain and suggestions for the industry's sustainable growth. The authors analyzed monthly data on piglet, live pig, pork, and pig feed prices in the hog industry chain in Yunnan Province from 2001 to 2017 by using the Census-X12 seasonal adjustment method and HP filter. The author also used the HP filter decomposition method to analyze price fluctuations in the hog industry chain. The study's results indicate that the costs of piglets, live pigs, pork, and pig feed were subject to rapid fluctuations, randomness, and external shocks. Price cycles and correlations between piglets and live pigs were strong, and there was a very quick downward transmission of price movements from piglets and live pigs to pork.

Sadefo *et al.* [8] aims were to: build deep learning models for intraday commodity price prediction, conduct an empirical investigation of deep learning tools for computational finance, and examine the influence of COVID-19 on intraday commodity prices. The correlation and causal relationship between commodity prices and COVID-19 data are examined by the authors using statistical models, such as the Granger causality test, Pearson correlation coefficient, ARIMA-WBF model, and LSTM model. According to the study's findings, the LSTM model has been accurate in predicting commodity prices, and the hybrid ARIMA-WBF model is accurate in COVID-19 cases. The report advises investors and countries participating in these markets to focus their financial resources on durable commodities like silver.

Calvia *et al.* [14] to better understand the historical behavior, coordination, and co-movements of meat commodities, such as beef, lamb, pork, and poultry, the author analyzed the short-term fluctuations of global meat commodity prices from January 1980 to October 2023. The authors estimated the cyclical behavior of poultry prices using a harmonic model and the MBBQ algorithm to identify peaks and troughs in cycles of meat prices. The study concludes that there is a common cycle in meat prices, which has been observed historically and during recent global crises. The study also emphasizes how vital it is to recognize changes in the price of meat for planning purposes, including food security and the production and consumption of meat.

2.3 Market Trends and Consumption Patterns

Wickramarachchi *et al.* [3] analyzed Srilanka poultry product prices. The Colombo Consumer Price Index (CCPI) was used to convert nominal prices into real prices. This study used complex models, such as SARIMA and ARIMA, to examine how different poultry product prices changed over time. Maximum Likelihood Estimates (MLEs) of the parameters are used to fit the models, and the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are used to examine the goodness-of-fit. Model prediction success is assessed with the Mean Absolute Percentage Error (MAPE). For broiler chicken, the results showed ARIMA (3,1,1), for curry chicken, ARIMA (0,1,2), and for brown and white eggs, SARIMA (2,1,2) (0,0,2). where price forecasts by models are accurate. The importance of understanding the variables, like inflation and exchange rates, influencing price trends is important to make well-informed decisions regarding industry.

Whitton *et al.* [6] examined the changes in meat consumption between 2000 and 2019 in 35 different countries. The authors aimed to group countries based on their patterns of meat consumption and investigated the relationships between changes in GDP and changes in meat consumption using hierarchical cluster analysis in addition to using linear regression models to investigate the relationship between GDP per capita and meat consumption. The findings indicated that consumption of chicken has increased while consumption of beef and sheep has decreased in certain countries, and changes in GDP are correlated with changes in meat consumption in some countries. In conclusion, there are differences in meat consumption between nations, and the trends in meat consumption are influenced by consumer appetite, historical technological industrialization, and population growth.

Akbar, M *et al.*[10] tells that the growth trend of local chicken meat production in Pakpak Bharat Regency is highlighted by authors who also identify the factors influencing the supply and demand for local chicken meat. To determine the variables influencing the supply and demand of local chicken meat, the study applies multiple linear regression analysis to time series data ranging from 2010 to 2019. The study concludes that some variables, such as pricing, population, income, and production costs, have an impact on the supply and demand for local chicken meat.

2.4 Practical Applications and Risk Management

Mattos *et al.* [12] mentioned the possibility of a new beef futures contract based on the (BBCO) Boxed Beef Cut Out index, which could be useful in minimizing price risks for several different cuts of beef. Using the BBCO index and live cattle futures contracts, the study estimates hedge ratios and hedging effectiveness measures for a selection of meat cuts and an equally weighted portfolio at 1-, 6-, and 12-week horizons using unconditional and conditional hedging procedures. This indicates that the BBCO index may offer more opportunities for hedging wholesale meat cut prices and that a futures contract based on the index may be a more useful instrument for managing price risk in the beef industry than the live cattle futures contract.

Luke *et al.* [13] discussed the inverse almost ideal demand system (IAIDS) method of analyzing US demand for various pork primal cuts and other meats while taking seasonal supply variation into account. The analysis utilized monthly estimates of the disappearance of primal cuts of pork, beef, and chicken, as well as ham, loin, and other pork, per capita in the United States. The study's results suggest that the belly's quantity flexibility is less than one in absolute value, proving that the belly is an essential meat cut. The remaining primal cuts and beef display scale flexibility that suggests luxury goods, while the other pork and chicken scale flexibility that indicates necessity goods.

2.5 Methodologies for Forecasting

Peshevski *et al.* [9] used cutting-edge machine-learning techniques to create a comprehensive framework for modeling and analyzing trends in food prices in twelve selected European countries. The training data was from April 2013 to December 2019 and the lead time was from January 2020 to January 2022. The authors used models such as Xgboost. When predicting food prices from January 2020 to January 2022 and from January 2021 to January 2023, the Xgboost model achieves an average R2 of 0.85 and 0.64, respectively. The Xgboost Regressor model's potential to capture complex price dynamics is demonstrated by this approach's conclusion, but it also demonstrates its limitations when it comes to handling outlier years.

CHAPTER 3

Methodologies

Figure 3.1 presents the Methodology Research Design. Each step is to be discussed in the remaining of the dissertation.

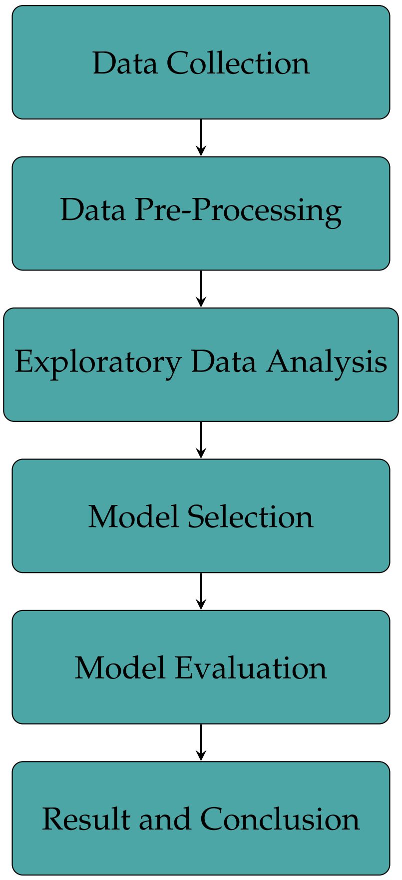

Figure 3.1: Steps Involved in the Methodology Section.

3.1 Source and Nature of the Data

Period	Meat Type	Meat Sub-Type	Price Dollars/cwt/doz	Year	Month	YearMonth
2000-01-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	114.74	2000	1	2000-01
2000-02-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	112.59	2000	2	2000-02
2000-03-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	118.42	2000	3	2000-03
2000-04-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	123.45	2000	4	2000-04
2000-05-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	124.88	2000	5	2000-05
2000-06-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	123.30	2000	6	2000-06
2000-07-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	115.85	2000	7	2000-07
2000-08-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	111.20	2000	8	2000-08
2000-09-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	108.68	2000	9	2000-09
2000-10-01	Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	112.58	2000	10	2000-10

Table 3.1: The Wholesale meat dataset from the year 2000 to 2024

This dataset is taken from USDA Livestock and Meat Domestic Data component of the USDA Livestock, Dairy, and Poultry Outlook. In this Dataset, each column represents a monthly time point of historical records from 2000 January to 2024 April with wholesale prices for a wide range of meat products, where each time point represents a specific month and year when the price data was recorded as shown in the table 3.1. This analysis gives monthly price trends in identifying both short-term fluctuations and long-term patterns. In this dataset, each row encompasses several meat products, categorized broadly into types and sub-types. Some of the key categories of types and sub-types along with the descriptions are shown in table 1.1. The prices of each meat at the wholesale level, are expressed in dollars per hundredweight (cwt). This unit of measure, commonly used in the meat industry, allows for standardized comparisons across different meat types and markets. This dataset is publicly available ensuring that the information is very accurate, reliable, and consistent across all periods. Each entry in the dataset represents the real-time wholesale market price that is related to economic conditions, supply chain dynamics, and consumer demand for the specific price.

This data set comes with some important points that need to be highlighted while doing the analysis:

1. (1, 2, 3) = [1]12-city composite wholesale price discontinued in January 2013, [2]The Georgia Dock Poultry Report was discontinued by the Georgia Department of Agriculture in December 2016, [3]Northeast parts prices were discontinued in August 2022.

Meat Type	Meat Sub Type	Description
Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	Beef cuts, 600-900 lbs, Choice 1-3 grade, central US
Beef, Central U.S.	Boxed beef cutout / select 1-3 (600-900) lbs	Beef cuts, 600-900 lbs, Select 1-3 grade, central US
Beef, Central U.S.	Boneless beef, 90% fresh	90% lean beef, fresh, central US
Beef, Central U.S.	Imported boneless beef, 90% frozen	90% lean beef, imported, frozen, central US
Pork, Central U.S.	Pork cutout composite	Composite price, various pork cuts, central US
Pork, Central U.S.	Loins, 1/4", trimmed vacuum-packed	Pork loins, 1/4" trim, vacuum-packed, central US
Pork, Central U.S.	Bellies, 12-14 lb, skin on	Pork bellies, 12-14 lbs, skin on, central US
Pork, Central U.S.	Hams, 23-27 lb, trimmed selected	Hams, 23-27 lbs, trimmed, selected, central US
Pork, Central U.S.	Trimmings, 72% combo	72% lean pork trimmings mix, central US
Lamb, East Coast	55-65 lb, Choice	Lamb cuts, 55-65 lbs, Choice grade, East Coast
Turkeys	Hens, 8-16 lb	Female turkeys, 8-16 lbs, avg. 17 lbs live
Turkeys	Toms, 16-24 lb	Male turkeys, 16-24 lbs, avg. 41 lbs live
Turkeys	Breast, 4-8 lb	Turkey breast cuts, 4-8 lbs, boneless
Turkeys	Drumsticks	Lower turkey leg, rich flavor
Turkeys	Wings, full cut	Entire wing section, for appetizers
Eggs, Grade A, large	Combined regional	Composite price, large Grade A eggs, various regions
Eggs, Grade A, large	New York	Price for large Grade A eggs, New York market
Broilers	National composite	Composite price, broiler cuts, nationwide
Broilers	12-city composite[1]	Price average, 12 major cities
Broilers	Georgia dock[2]	Price index, Georgia
Broilers	NE parts[3] / Breast, boneless	Boneless breast, Northeast
Broilers	NE parts[3] / Breast, ribs on	Breast with ribs, Northeast
Broilers	NE parts[3] / Legs, whole	Whole legs, Northeast
Broilers	NE parts[3] / Leg quarters	Thigh and drumstick, Northeast
Broilers	National parts / Breast, boneless	Boneless breast, nationwide
Broilers	National parts / Breast, ribs on	Breast with ribs, nationwide
Broilers	National parts / Legs, whole	Whole legs, nationwide
Broilers	National parts / Leg quarters	Thigh and drumstick, nationwide

Table 3.2: Descriptions for Various Wholesale Meat Products

2. 0 = missing data or no quotes

3. cwt = hundredweight, lb = pound. Bl = boneless

4. It is important to note that several key prices have been discontinued over the years, which may impact long-term trend analysis. The discontinuation of the 12-city composite wholesale price in 2013, the Georgia Dock Poultry Report in 2016, and Northeast parts prices in 2022 represent significant changes in how poultry prices have been reported. These changes may create challenges in comparing current prices with historical data and should be considered when interpreting long-term trends.

3.2 Data Pre-Processing

Data Pre-Processing is the task of cleaning and transforming raw data to make it suitable for analysis and modeling. Data Pre-Processing is very important for the time series dataset as it holds important information, especially when dealing with economic data like USDA wholesale meat price. The common problems associated with time series are unstructured timestamps, missing values, and outliers. Effective Pre-Processing ensures data is clean and ready for analysis which leads to accurate results. This section gives a detailed analysis of the steps involved in time series data for the USDA dataset:

1. Data Cleaning
2. Feature Engineering
3. Data Integration
4. Data Scaling/Normalization

3.2.1 Data Cleaning

Data cleaning is a very important step in time series analysis, ensuring the quality of the dataset before further analysis or modeling. This section outlines a systematic approach to data cleaning, starting from loading the dataset to preparing it for analysis.

Loading the Wholesale Meat Price dataset

After loading the Wholesale meat price dataset, it contains (8176 Rows, and 6 Columns). So, each column is named Period, Meat Type, Meat Sub-Type, Year, Month, and combined year month which gives a brief analysis of how the dataset is structured. The period column is in (YYYY-MM-DD) format and the meat type and sub-meat type consist of unique meat data from the year 2000 to 2024 and the price column is measured in dollars per hundredweight (cwt). The data appears to be consistent in its structure and complete for the period year shown.

Loading All the external Features datasets

1. The weather data is loaded, and it contains (295 Rows, 9 Columns). The dataset includes daily weather observations, with variables such as date, maximum temperature (TMAX), minimum temperature (TMIN), precipitation (PRCP), and snowfall (SNOW).

2. The Disaster data is loaded and contains (291Rows, 6 Columns). The dataset includes declaration Date, incident Type, incidentBeginDate, and incidentEndDate. These columns capture the declaration date of the disaster, the type of incident, and the start and end dates of each incident.

3. The inflation rate data is loaded and contains (293 Rows, 7 Columns). The columns contain information about the inflation rate, including the period (month or quarter) and year.

4. The unemployment rate data is loaded and contains (294 Rows, 7 Columns). This dataset includes columns for the year, period (representing the month or other time segments), and the unemployment rate value.

5. The GDP data is loaded and contains (293 Rows, and 5 Columns). This data contains information on the monthly real GDP index, which serves as a key economic indicator for analyzing the performance of the U.S. economy.

6. The corn price data is loaded and contains (292 Rows, and 4 Columns). This data contains monthly corn price information along with corresponding year and month details.

7. The sorghum price data is loaded and contains (292 Rows, and 4 Columns). This dataset contains monthly price information for sorghum, along with corresponding year and month data.

Checking basic information about the dataset

1. `df.head()` function in pandas, which gives the first few rows of the dataset a look if the data is printing properly or not.

2. `df.tail()` function in pandas, which gives the last few rows of the dataset have look if the data is printing properly or not.

3. `df.info()` function in pandas, which provides a summary of the dataset's structure. The meat type columns (object) contain a categorical column, similarly, the meat sub-type columns (object) also contain a categorical column, price contains (float64) indicates decimal numbers, appropriate for prices, year column (int32) indicates integer values, suitable for years. Similarly follow the same structure for month and year and month columns. The merged dataset contains data on both wholesale meat prices and all the external feature combinations in the same dataset named merged data. The merged data structure is given as meat sub-type columns (object) which also contain categorical columns, price contains (float64)

indicates decimal numbers, year column (int32) indicates integer values, suitable for years, YearMonth (int32), PRCP (float64), SNOW (float64), TAVG (float64), Disaster Count (float64), CPI (float64), Unemployment Rate (float64), GDP Value (float64), Corn Price (float64), Sorghum Price (float64) as these values are decimal.

4. `df.unique()` function in pandas, which provides a comprehensive overview of the unique values in each column. So, the unique values in meat type are (Beef, Central US, Pork, Central US, Lamb, East Coast, Broilers, Turkeys, Eggs, Grade A, large)

Checking Duplicates

`df.duplicated().sum()` checks for duplicate rows in the dataset and returns the total number of duplicates. There are 0 duplicate values in the dataset. This means each row represents a unique combination of Meat Type, Meat Sub-Type, Price, Year, Month, and Year Month. The dataset suggests there is consistent monthly data without any repeated entries, which is ideal for time series analysis. In merged data, there are 0 duplicate values in the dataset. This means each row represents a unique combination of Meat Type, Meat Sub-Type, Price, Year, Month, Year Month, PRCP, SNOW, TAVG, Disaster Count, CPI, Unemployment Rate, GDP Value, Corn Price, Sorghum Price.

Checking Null Values

1. `df.is null().sum()` checks for missing values (null or NaN) in each column of the dataset and returns the count of missing values for each column. There are 0 (null or NaN) values in the dataset as shown in table 3.3. The dataset is complete with no gaps.

2. On doing more analysis on (null or NaN) values found out that there are many 0 price values for the meat subtype in the dataset as shown in table 3.4. So the number of null values is checked by writing a code. The main reason that there are many zero values in the dataset is that several key prices have been discontinued over the years like the 12-city composite wholesale price in 2013, the Georgia Dock Poultry Report in 2016, and Northeast parts prices in 2022 represent significant changes in poultry prices. One more reason is that data integrity like missing to

Attribute	Null Values (Count)
Meat Type	0
Meat Sub-Type	0
Price Dollars/cwt/doz	0
Year	0
Month	0
YearMonth	0

Table 3.3: Null Values Check in Dataset

add some value in the price column may impact the prediction of future prices from all these categories. The reason for not taking zero values in this time series dataset if we are interpolating or taking mean or median values that will not give accurate price predictions for each meat subtype.

3. So, by taking into consideration all these points another code is written as shown in the table 3.5 by taking only the meat subtypes with non-zero values in the "Price Dollars/cwt/doz" column, which enhances data reliability and ensures continuity in price trends this approach not only provides a strong foundation for forecasting models, seasonal adjustments, and price index calculations but also guides our strategy for handling missing data in other subtypes, and enables more accurate cross-sectional comparisons across different periods and product categories. By focusing on these non-zero value subtypes, the risk of artificial price fluctuations skews our analysis, which is particularly good for trend analysis, seasonality detection, and the development of reliable economic indicators in the meat industry. This approach strengthens our time series models but also provides valuable insights into market dynamics and product-specific pricing patterns, ultimately contributing to more informed decision-making in both academic research and practical market analysis contexts.

4. So finally, after taking all these points selected for each meat type, two sub-meat types were included in the analysis. The chosen sub Meat types are:

   - Boxed Beef Cutout / Select 1-3 (600-900) LBS
   - Imported Boneless Beef, 90 Percent, Frozen

Meat Sub-Type	**Zero Price Counts**
12-city composite	136
55-65 lb, Choice	90
Bellies, 12-14 lb, skin on	47
Boneless beef, 90 percent, fresh	2
Breast, 4-8 lb	2
Georgia dock[a]	96
Hams, 23-27 lb, trimmed selected	1
National composite	156
National parts / Breast, boneless	272
National parts / Breast, ribs on	272
National parts / Leg quarters	272
National parts / Legs, whole	272
Northeast parts[3] / Breast, boneless	20
Northeast parts[3] / Breast, ribs on	20
Northeast parts[3] / Leg quarters	20
Northeast parts[3] / Legs, whole	20

Table 3.4: Zero values count in the wholesale meat dataset

- Pork Cutout Composite
- Loins, 1/4", Trimmed Vacuum-Packed
- Drumsticks
- Wings, Full Cut

This approach allows for a better examination of price dynamics within a specific meat category, potentially revealing intra-category variations and providing deeper insights into market segmentation. These selected sub-meat types were constantly employed across all models and analytical techniques used in this research.

Meat Sub-Type with no zero values
Boxed beef cutout/choice 1-3 (600-900) lbs
Boxed beef cutout/select 1-3 (600-900) lbs
Drumsticks
Hens, 8-16 lb
Imported boneless beef, 90 percent frozen
Loins, 1/4, trimmed vacuum-packed
New York
Pork cutout composite
Toms, 16-24 lb
Trimmings, 72 percent combo
Wings, full cut

Table 3.5: Meat Sub-Types with no zero values

Checking Null Values for external factors

`merged data.is null().sum()` checks for missing values (null or NaN) in each column of the dataset and returns the count of missing values for each column. There are many (null or NaN) values in the dataset as shown in table 3.6. There are null values in PRCP (476), SNOW (3668), TAVG (476), and Disaster Count (112). To fill null values in columns like PRCP, SNOW, and TAVG the forward fill method (ffill). This technique propagates the last valid observation forward to fill gaps, ensuring continuity in the time series data. The assumption here is that precipitation, snowfall, and average temperature patterns do not change drastically between months, so using the last known value provides a reasonable estimate for the missing data. While for disaster Count are filled with zeros. The reason behind this decision is that the absence of data likely indicates no disasters were declared during those months. Hence, filling with zero accurately reflects the situation rather than leaving the data missing.

3.2.2 Feature Engineering

Feature engineering is the process of selecting, transforming, and creating new features from raw data to improve the performance of machine learning models. Features are

Column	Null Values
Meat Type	0
Meat Sub-Type	0
Price Dollars/cwt/doz	0
Year	0
Month	0
YearMonth	0
PRCP	476
SNOW	3668
TAVG	476
Disaster Count	112
CPI	0
Unemployment Rate	0
GDP Value	0
Corn Price	0
Sorghum Price	0

Table 3.6: Null Values in External Feature Dataset

the input variables that models use to make predictions. In time series analysis, feature engineering involves creating features that capture the temporal structure, trends, seasonality, and other relevant patterns in the data.

In the wholesale dataset, there are four columns which are Period, Meat Type, Meat Sub Type, and Price columns. From the Period column, two other columns called year and month are created, the year and month columns then are combined to create another new column YearMonth as shown in table 3.1. Having unique columns in both datasets like YearMonth will make it easier to merge the wholesale meat price data into all the external features data and capture them in one single dataset that is the meat_merged dataset. While analyzing all the external feature datasets there were many issues because the datasets were not in the correct format. Adding or removing columns and also changing the string to date time format whereas aggregating two columns and forming new columns out of it were some things that needed to be changed before forming one complete dataset. So here are the examples that have been worked on each external feature dataset.

In weather data, the `DATE` column is converted from a string format to a Date Time object. This conversion allows for easy extraction of the year and month components. A new column `TAVG` is created to represent the daily average temperature, calculated as the mean of the maximum and minimum temperatures. The `DATE` column is converted to a monthly period using the `to_period('M')` method and then back to a timestamp. This resampling allows for aggregation of the daily data into monthly summaries. After which, all the variables like maximum temperature (`TMAX`), minimum temperature (`TMIN`), precipitation (`PRCP`), and snowfall (`SNOW`) are converted to monthly format. Finally, additional columns were created for Year, Month, and a combined Year-Month, as these would help in merging the wholesale meat data into weather data, using common columns in both datasets.

In the disaster data, the `declaration Date`, `incidentBeginDate`, and `incidentEndDate` columns are converted from string format to date-time objects. The dataset is filtered to include only disaster declarations made between January 1, 2000, and April 31, 2024, to match the date range of the wholesale meat price dataset. The `declaration Date` column is converted to a monthly period using the `to_period('M')` method, facilitating the accumulation of incidents by month. The dataset is then grouped by month, and several key metrics are calculated: `Disaster Count` is calculated based on the number of disaster incidents declared in each month; `DisasterStartDate` represents the earliest incident start date within each month, and `DisasterEndDate` represents the latest incident end date within each month. Finally, the `incident Type` column is renamed as `Disaster Count`, `incidentBeginDate` is renamed as `DisasterBeginDate`, and `incidentEndDate` is renamed as `DisasterEndDate`, with all these columns now reflecting monthly analysis. Further, columns for Year, Month, and a combined Year-Month were created to facilitate the merging of the wholesale meat data with the weather data, using these common columns across both datasets.

The Inflation rate data contains some unwanted periods labeled as 'S01' and 'S02', which represent semi-annual or other non-monthly periods. These periods are removed to focus on the analysis exclusively of monthly data. A new Date column is created by combining the Year and Period columns. The Period column contains a string with the format "MXX" where "XX" is the month number. The leading "M" is stripped off, and the remaining part is zero-padded to ensure it forms a valid two-digit month number. The concatenated string is then converted into a date-time format. The new columns for Year and Month are derived from the Date column. The Value column, which represents the Consumer Price Index (CPI), is renamed to CPI. A Year Month column is created by concatenating the Year and Month columns. The final columns are the Year, Month, CPI, Date, and Year Month columns.

The Unemployment Rate dataset includes some periods labeled as 'S01' and 'S02', which likely represent semi-annual or non-monthly data points. A new Date column is generated by combining the Year and Period columns. The Period column is processed to extract the month number by stripping the leading character and padding it to ensure it forms a valid two-digit month number. The concatenated string is then converted into a date-time object, allowing for chronological analysis and manipulation of the data. From the Date column, new columns for Year and Month are derived. The Value column, which contains the unemployment rate, is renamed to Unemployment Rate. A new Year Month column is created by concatenating the Year and Month columns, formatted as YYYY-MM. This identifier is essential for monthly trend analysis and for merging this dataset with others on a time basis.

In GDP data the Date column, which represents the period for each GDP measurement, is converted to a date-time format. New columns for Year and Month are derived from the Date column. The Monthly Real GDP Index column, which holds the GDP values, is renamed to GDP Value. A Year Month column is created by concatenating the Year and Month columns, formatted as YYYY-MM. This identifier is essential for monthly trend analysis and for merging this dataset with others on a time basis.

In the Corn price dataset, the `Month` column contains abbreviated month names (e.g., 'Jan', 'Feb'). To standardize these entries, a mapping dictionary `month_map` is created to convert each month's abbreviation to its corresponding two-digit numerical value. The `Month` column in the `Corn_Price` data frame is updated by replacing the month abbreviations with their corresponding numerical values using the `month_map` dictionary. A new column `Year Month` is created by concatenating the `Year` and `Month` columns. The `Year` is converted to a string, and the `Month` is ensured to be a two-digit string using zero-padding where necessary. This `Year Month` identifier serves as a unique reference for each month-year combination, facilitating time series analysis and data merging with other datasets. The `Price` column, which contains the corn price values, is renamed to `Corn Price` to provide a more descriptive and clearer label. A new data frame `final_df` is created by selecting only the necessary columns `Year`, `Month`, `Year Month`, and `Corn Price`.

In the `Sorghum_Price` dataset, the `Month` column contains abbreviated month names (e.g., 'Jan', 'Feb'). To standardize these entries, a mapping dictionary `month_map` is defined to convert each month's abbreviation to its corresponding two-digit numerical value. The `Month` column in the `Sorghum_Price` data frame is updated by replacing the month abbreviations with their corresponding numerical values using the `month_map` dictionary. A new column `Year Month` is created by concatenating the `Year` and `Month` columns. The `Year` is converted to a string and the `Month` is ensured to be a two-digit string using zero-padding where necessary. This `Year Month` identifier serves as a unique reference for each month-year combination, facilitating time series analysis and data merging with other datasets. The `Price` column, which contains the sorghum price values, is renamed to `Sorghum Price` for better clarity. A new data frame `Sorghum_Price_Final` is created by selecting only the essential columns: `Year`, `Month`, `Year Month`, and `Sorghum Price`.

3.2.3 Data Integration

Data Integration is a technique of combining data from different sources and arranging all the datasets in a proper format. Before merging all the datasets and doing more analysis on the wholesale meat data identified as certain meat types like pork, beef, and poultry which were influenced by external factors like weather data, disaster data, inflation rates, unemployment rates, GDP values, and commodity prices (corn and sorghum) as referred in research papers [3, 6]. By integrating these datasets based on a common temporal dimension ('YearMonth'), we aim to capture the influence of these external factors on meat prices, enabling a more comprehensive time series analysis.

The merging process was performed using the Pd.merge function with a how='left' join, ensuring that all rows from the meat price data are retained, even if there are missing values in the external datasets. Firstly, the meat price data is merged with the weather data, like 'PRCP', 'SNOW', and 'TAVG' variables, and indexed by 'Year-

Month'. This states that weather conditions might affect meat production and pricing. Secondly, disaster data is merged, adding the 'Disaster Count' variable along with indexed by 'YearMonth' which reflects the number of disasters declared in each month. Thirdly inflation rate data is merged, adding the 'CPI' variable along with indexed by 'YearMonth'. Next unemployment data is merged, adding in the 'Unemployment Rate' variable along with indexed by 'YearMonth'. The GDP data is merged, adding the 'GDP Value' variable along with indexed 'YearMonth'. GDP is a key economic indicator that reflects the overall economic activity, which can impact meat prices through changes in demand and supply. Next corn price data is merged, adding the 'Corn Price' variable along with indexed 'YearMonth'. Seventh sorghum price data is merged, adding the 'Sorghum Price' variable along with indexed by 'YearMonth'. Corn and Sorghum are very important animal feed, and their price can similarly affect meat production costs.

Finally, after merging the complete dataset as shown in table 3.7 the total number (8176 Rows, 15 Columns) of the dataset is named merged_data.

Meat Type	Meat Sub-Type	Price Dollars/cwt/doz	Year	Month	YearMonth	PRCP	SNOW	TAVG	Disaster Count	CPI	Unemployment Rate	GDP Value	Corn Price	Sorghum Price
Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	114.74	2000	1	2000-01	0.143600	0.133333	29.566667	137.0	168.800	4.0	13812.379328	1.91	1.60
Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	112.59	2000	2	2000-02	0.278621	0.000000	41.327586	183.0	169.800	4.1	13871.521334	1.98	1.71
Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	118.42	2000	3	2000-03	0.074516	0.000000	46.839286	11.0	171.200	4.0	13951.023825	2.03	1.80
Beef, Central U.S.	Boxed beef cutout/choice 1-3 (600-900) lbs	123.45	2000	4	2000-04	0.074286	0.000000	52.326923	28.0	171.300	3.8	14133.222640	2.03	1.81

Table 3.7: Merged Meat Dataset with External Influencing Factors

3.2.4 Data Scaling/Normalisation

Min-max scaling, also known as normalization, is a technique commonly used in data prepossessing. To scale the meat_data_price, use the Min-max Scalar from the sklearn, prepossessing module. This technique is very useful and re-scales the data within a specific range, between 0 and 1. Then fit _transform first computes the minimum and maximum values for each feature in the dataset by (fit step) and then scales the data accordingly by (transform step) as shown in formula (3.1). The result is stored in the meat_data_price dataset, where all the feature values have been re-scaled to fall within the range of [0, 1].

This min-max scaling works well on categorical variables, Ordinal variables, and Binary variables. This technique is very useful in machine learning models like gradient descent and neural networks which perform faster when the input feature is scaled to [0,1] and when visualizing data, features that are on the same scale make it easier to compare and understand the visualizations better.

By [17], the formula for Min-Max Scaling can be written by

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}\,. \tag{3.1}$$

In this formula:

- x is the original value of a feature.
- $\min(x)$ is the minimum value of that feature in the dataset.
- $\max(x)$ is the maximum value of that feature in the dataset.
- x' is the scaled value of the feature between 0 and 1.

3.3 Exploratory Data Analysis

The Exploratory Data Analysis section provides a comprehensive overview of the data related to various meat sub-types. It includes statistical summaries and different visualizations, to understand the characteristics, trends, and market dynamics of each meat product.

3.3.1 Descriptive statistics is summary table

Meat Sub-Type	**Count**	**Mean**	**Std**	**Min**	**Max**	**25%**	**50%**	**75%**
Boxed beef cutout / select 1-3 (600-900) lbs	292	178.55	52.68	101.46	386.24	134.34	176.18	211.83
Drumsticks	292	61.16	24.62	17.99	118.33	42.90	59.99	74.10
Imported boneless beef, 90 percent, frozen	292	183.18	58.11	89.43	307.63	131.83	196.60	225.14
Loins, 1/4", trimmed vacuum-packed	292	108.39	17.40	72.22	192.07	96.10	108.12	117.80
Pork cutout composite	292	78.52	16.57	45.88	133.58	67.10	76.45	87.51
Wings, full cut	292	55.07	25.55	12.74	142.00	34.76	50.34	70.54

Table 3.8: Descriptive Statistics Summary Table

Descriptive statistics is a summary table that explains each feature from the collection of structured data. Each metric describes the dataset as a measure of central tendency (mean, median), a measure of dispersion (range, standard deviation), and a measure of position (percentiles, quartile) all of them are summarized adequately, and the so-called five number summaries. By referring to descriptive summary table 3.8 six meat subtypes are Boxed beef cutout select, imported boneless beef frozen, pork cutout composite, lions vacuum packed, Drumsticks and wings, full cut, and all meat sub-types have 292 observations, indicating sizes across all categories are same. The mean represents the average price of each meat sub-type over the observed period. A higher mean suggests a higher average price, indicating either a higher quality or increased demand. The standard deviation tells how much variation exists from the average mean. A high standard deviation indicates a wide range of prices, suggesting volatility. A low standard deviation suggests stable pricing. The min and max values show the range within which price values for each meat sub-type can fluctuate. Percentiles give a detailed picture of the data distribution and provide insights into price concentration.

Starting from Boxed Beef Cutout / Select 1-3 (600-900) LBS, the Mean price is 178.55 this suggests a moderately high price point, which indicates the demand level and cost in the market. The standard Deviation is 52.68 this implies that there is more variability in price fluctuations and seasonal factors. The range between the min value of 101.46 and Max of 386.24 suggests that prices can vary over time. Percentiles are 134.34, 176.18, and 211.83, the spread of values across the quartiles suggests that 50% of the data is concentrated between approximately 134.34 and 211.83. This indicates the typical price range, which is useful for setting pricing strategies or forecasting future prices.

In Imported Boneless Beef, 90 Percent, Frozen, the Mean price is 183.18 which highest among all meat sub-types, which tells that the cost of this product is very high. The Standard Deviation is 58.11 suggesting prices are highly volatile. The Min value is 89.43 and the Max value is 307.63 suggesting high volatility in price. Percentiles (131.83, 196.6, 225.14): These indicate that 50% of the values are between 131.83 and 225.14, suggesting that prices are frequently on the higher end.

In Pork Cutout Composite, the mean price is 78.52. This price suggests that pork cutout is moderately priced, less than beef but higher than drumsticks and wings. The standard deviation is 16.57, which suggests low, stable pricing and consistent demand. The range between the minimum value of 45.88 and the maximum value of 133.58 indicates that few price fluctuations are seen. Percentile values 67.10, 76.45, and 87.51 show that prices tend to be clustered, making it useful for predicting this product's price.

In Loins, the Mean price is 108.39 these are positioned as a mid-range meat product.

The Standard Deviation of 17.40 suggests stable pricing with less volatility compared to more expensive cuts like imported beef. The range between the Min value is 72.22 and the Max value is 192.0 suggesting that mid-range products with moderate price stability. Percentiles are 96.09, 108.12, and 117.8 these values indicate a tight clustering of prices, most sales occur within this range.

In Drumsticks the Mean Price is 61.16 which suggests that it is one of the more affordable meat options within this price range. The Standard Deviation is 24.62 there is variability in prices that tend to be more stable among other meat types. The range between Min values of 17.99 and Max values of 118.33 is very low compared to other meat types, low-cost options in higher demand situations. The percentiles are 42.90,59.99,74.10, in this 50% of prices are between 42.90 and 74.10 suggesting a narrow range of prices.

In Wings, Full Cut, the Mean price is 55.07 lowest average price among all meat types. The Standard Deviation is 25.54 not as high as another more expensive meat type. The range between the minimum value of 12.74 and the maximum value of 142.00 which is a very low minimum and high maximum suggests higher price demand. The percentiles 34.76,50.34,70.54 suggest maintaining a low to moderate price range.

3.3.2 Histogram Plot for different price distributions

The histogram plots 3.2 tell how the price distributions are spread for each meat subtype in the dataset. The plots give a better understanding of frequency distributions of observed prices over time and give insights into different price ranges, central tendency, and variability of prices for each meat subtype. Let's analyze each plot individually to understand the distribution characteristics.

Boxed Beef Cutout / Select 1-3 (600-900) lbs

The Plot of the Boxed Beef Cutout shows a right-skewed distribution with a long tail extending towards the higher prices end. The prices are mostly concentrated in the range between 100 and 250 cwt and very few prices reach up to 400 cwt. This range of prices indicates considerable price variability in the market for this meat sub-type. The long tail on the right indicates that most prices are moderate, there have been occasional spikes in price, but due to market conditions such as supply shortages or increased demand.

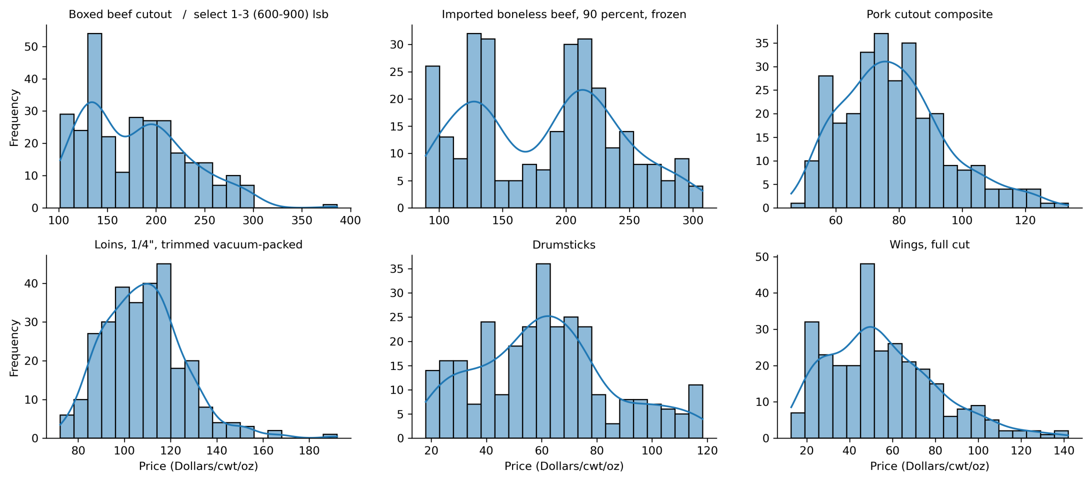

Figure 3.2: Histogram Plot for Price Distribution

Imported Boneless Beef, 90 Percent, Frozen

The Plot Imported Boneless Beef Shows bimodal distribution with two peaks, one around 125 cwt and another around 200 cwt. The prices range from 100 to 300 cwt, with many fluctuations. There is no single price range that dominates over another price, the market sees it as two different clusters of pricing regimes. The spread between the peaks could reflect varying demand or differences in product specifications.

Pork Cutout Composite

The plot of pork cutout composite shows a slight right-skewed distribution where most of the prices range between 60 and 100 cwt. Most prices have hover around 80cwt. Compared with other sub-types of the pork composite, it has moderate variability and very few extreme values. This distribution suggests that more stable market prices for pork cutouts are due to supply and demand pricing mechanisms.

Loins, 1/4", Trimmed Vacuum-Packed

The plot of Loins shows a normal distribution, with most prices falling within a narrower range. The peak price range is around 110 and 120 cwt. Compared to another subtype there is very little variability in loins which is indicated by a bell-shaped curve. This distribution suggests a very stable pricing movement, due to stable demand, and a controlled supply chain.

Drumsticks

The plot of Drumsticks shows multi-modal distributions with a peak around 50 cwt and another peak increasing around 75 cwt. Most prices are concentrated between 30 and 100 cwt. The variability is very high indicating multiple peaks with a wider range of prices. This distribution suggests that customer segments might prefer different pricing points and the supply chain could affect prices.

Wings, Full Cut

The plot of Wings shows a right-skewed distribution with prices between 20 and 75 cwt. The peak price is around 50 cwt. There is moderate variability with some higher prices up to 120 cwt. This distribution indicates occasional price surges, due to supply shortages and demand increases, and market disruption.

3.3.3 Historical wholesale price trend for all meat sub type

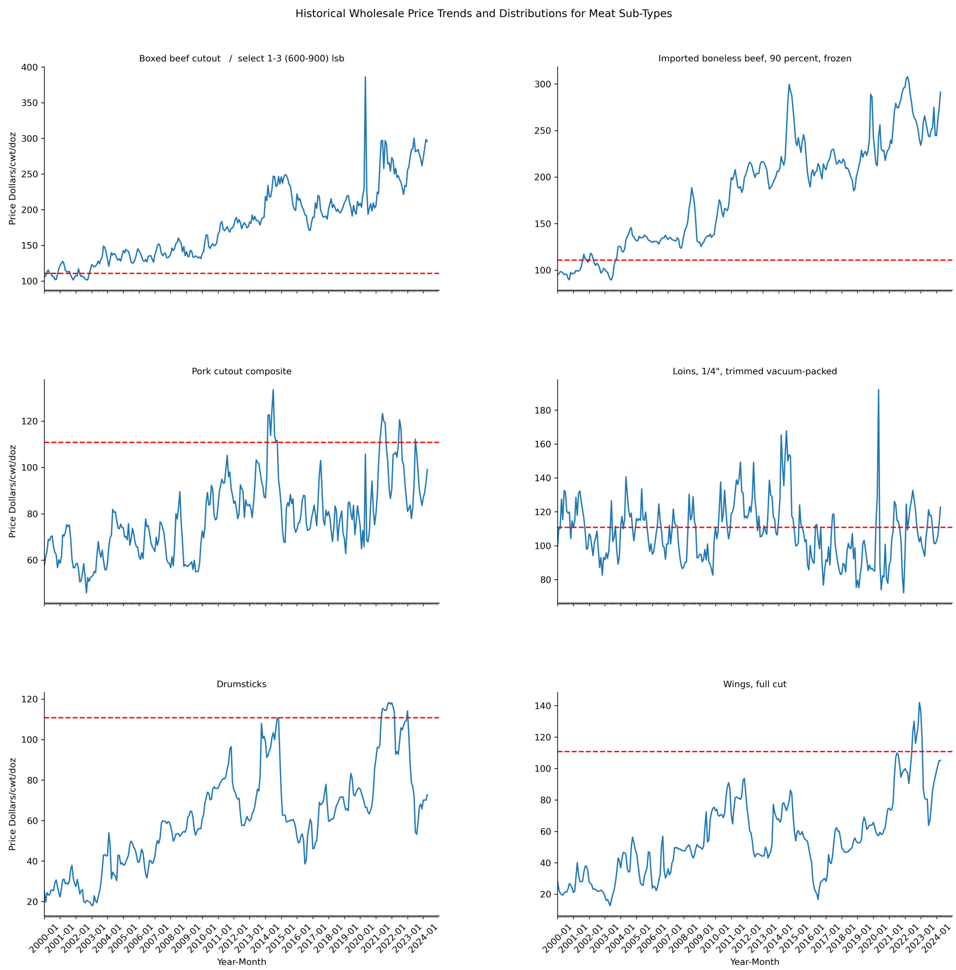

Figure 3.3: Historical Price Trend For All Meat Type

Boxed Beef Cutout / Select 1-3 (600-900) lbs

The price of boxed beef cutouts shows an upward trend over time as shown in the figure 3.3. There are many significant spikes visible in the plot, it all started in the year 2013 when the price rose sharply to around 200 cwt, and another significant spike occurred from 2019 to 2020 when the price reached over 300 cwt in mid-2020. This price rise is due to the COVID-19 pandemic causing a sharp in meat prices. After mid-2020 prices remain 250cwt.There are low price ranges from the year 2000 to 2012 fluctuating between 100 and 150 cwt. This is the current trend of this Boxed Beef Cutout suggesting continuous demand and ongoing supply challenges.

Imported Boneless Beef, 90 Percent, Frozen

The price movement of Imported Boneless Beef shows an upward trend over time as shown in figure 3.3. The spikes visible in the plot, all started in the year 2014 when the price rose sharply to around 220 cwt this could be due to global economic factors, the price remained volatile but generally upward until another spike in the year 2020 when the price reached over 300 cwt in mid-2020. This price rise is due to the COVID-19 pandemic causing a sharp in meat prices. After mid-2020 prices remain above 250cwt to 300cwt.The price remained low 150cwt from the year 2000 to 2008 after this there was a slight upwards that began in the year 2009. This is the current trend of this Imported Boneless Beef suggesting prices have consistently moved higher, reflecting increased demand for imported beef and global supply chain issues.

Pork Cutout Composite

The price of Pork Cutout Composite fluctuates within a relatively narrow range. There are many significant spikes visible in the plot as shown in the 3.3, it all started in the year 2014 when the price rise sharply to around 120 cwt, prices fluctuated significantly but generally maintained levels between 80 and 110 cwt until a sharp peak in 2020, again likely due to COVID-19-related disruptions. After mid-2020 prices remain above 80 and 110 cwt. The price remained low from 60 cwt to 80 cwt from the year 2000 to 2012, which is the current trend of this Pork Cutout Composite where the overall trend appears relatively stable with periodic fluctuations.

Loins, 1/4", Trimmed Vacuum-Packed

The price of Loins highly fluctuated without a clear upward or downward trend. There are a few short-lived spikes in the plot as shown in the 3.3, particularly in the year 2013 with the price of 160 cwt and again in 2020 at 180 cwt. These spikes are likely due to short-term supply disruptions or sudden changes in demand. Prices were relatively low, fluctuating around 90 to 110 cwt from the year 2000 to 2012, which is the current trend of the loin where the overall trend is a stable price point over time despite occasional shocks.

Drumsticks

The price of Drumsticks shows an upward trend. The significant spikes are visible in the plot as shown in the 3.3, with a sharp spike around the year 2015, reaching 80 cwt. This is due to increased demand for production. Prices continue to rise, peaking around 2020 at 120 due to pandemic-induced supply chain disruptions. After peaking in 2020, prices fell sharply but stabilized above 60 cwt, still higher than the levels seen in the early 2010s. The price remained low, between 20 cwt to 40 cwt from the year 2000 to 2008, this is the current trend of this drumstick where the overall trend tells a sharp rise and subsequent fall indicating a product that is price sensitive.

Wings, Full Cut

The price of Wings shows a gradual increase trend over time. There are many significant spikes visible in the plot as shown in the 3.3, a noticeable spike begins around 2011, when the price rises sharply to around 100 cwt. Another significant spike occurs from 2019 to 2020, with the price reaching over 140 cwt at its peak in mid-2020. This surge is likely related to the COVID-19 pandemic, which disrupted supply chains, causing sharp increases in meat prices. The price was low, at 20 cwt from the years 2003 and 2016, this is the current trend of these Wings where the overall trend reflects increasing demand, cost pressures, or supply shortages.

3.3.4 Handling Seasonality and Trend

In time series forecasting models the decomposition method is used to find patterns in a series of historical data. The three main components are trend, seasonality, and residuals. These methods help in finding the patterns of the dataset that are random or unexplained. For this analysis, the additive decomposition method assumes the fluctuation around the trends is constant over time as seen in the figure 3.4 each meat subtype.

Beef Cutout / Select 1-3 (600-900) LBS analysis on the seasonality trend

Observed Data: This is the actual price trend of Beef Cutout / Select 1-3 (600-900) over the years. An upward trend is seen from year 2000 to 2024. The price was around

Figure 3.4: Beef Cutout seasonality trend

100-150 in 2000 and later at the end, it came up to 250-300 in the year 2024. The sudden spike is visible around the year 2020 prices reaching up to 400 due to supply chain disruption during the COVID-19 pandemic.

Trend: An upward trend is visible, indicating a steady gradual increase in prices over time. As can be seen in Figure 3.4, there is a spike in the year 2022. This is due to the COVID-19 impact that affected meat prices. The trend is relatively constant from the year 2000 to 2010, with a price range around 125-150 and slowly a stepper increase is seen from 2010 to 2015 with a price of 225 and a slight dip is seen from 2016 to 2018. After 2020 there is a sharp increase due to pandemic issues.

Seasonality: There is a Clear seasonal present, showing recurring patterns within each year, and cyclic patterns are seen with peaks occurring at regular intervals. The seasonal fluctuations remain constant, ranging between 0.96 and 1.06. This suggests that seasonal factors influenced beef prices by about 5% all the years.

Residual: The residual plot 3.4 shows the irregular fluctuations of trend and seasonality. Most of the residual falls between 0.9 and 1.1 tell that the model captures the majority of price movements. In this plot, the spike is visible around 2020, with observed price movements and smaller spikes throughout telling about the price volatility.

Imported Boneless Beef, 90 Percent, Frozen analysis on the seasonality trend

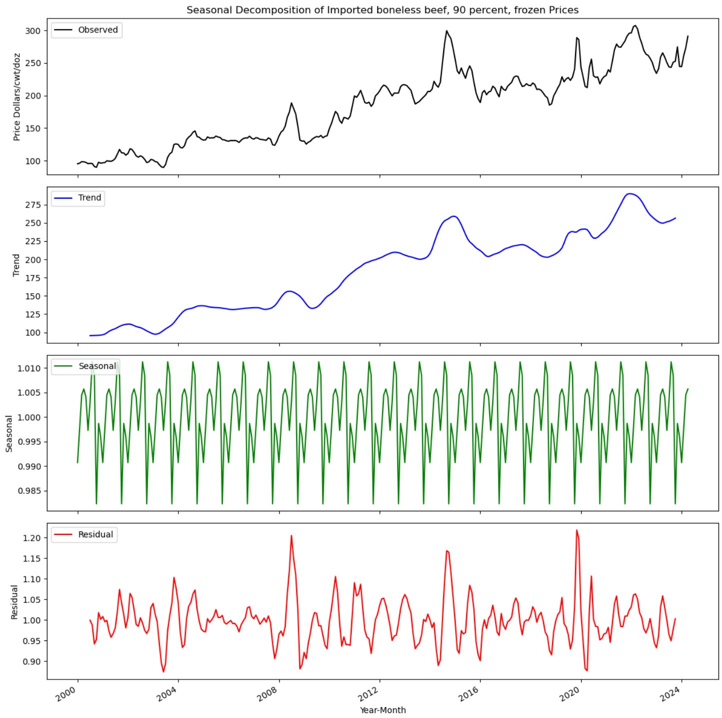

Figure 3.5: Imported Boneless Beef seasonality trend

Observed Data: This is the actual price trend of Imported Boneless Beef, 90 Percent, Frozen over years. An upward trend is seen in plot 3.5 from the year 2000 to 2024. The price fluctuations over the year are around 100-300. Multiple price spikes throughout the series, notably around 2013 and 2020.

Trend: There is an upward trend visible, with many fluctuations seen around the years. As you see in the plot 3.5 that is a gradual increase from 2000 to 2010 and again a sharp rise in the year 2013 followed by a decline period. Again, a significant rise in the year 2019 peaking around 2020-2021 with a price range of 275 to 300.

Seasonality: There is a Clear seasonal present, with Slight changes in the amplitude of seasonal effects over time, suggesting some evolution in seasonal patterns. The seasonal fluctuations remain constant, ranging between 0.985 and 1.010.

Residual: Residuals mostly fall between 0.8 and 1.2, with some extreme values reaching beyond these limits. Significant spikes were seen in plot 3.5 residuals, particularly around 2011 and 2020-2021, corresponding to observed price surges.

Pork Cutout Composite analysis on the seasonality trend

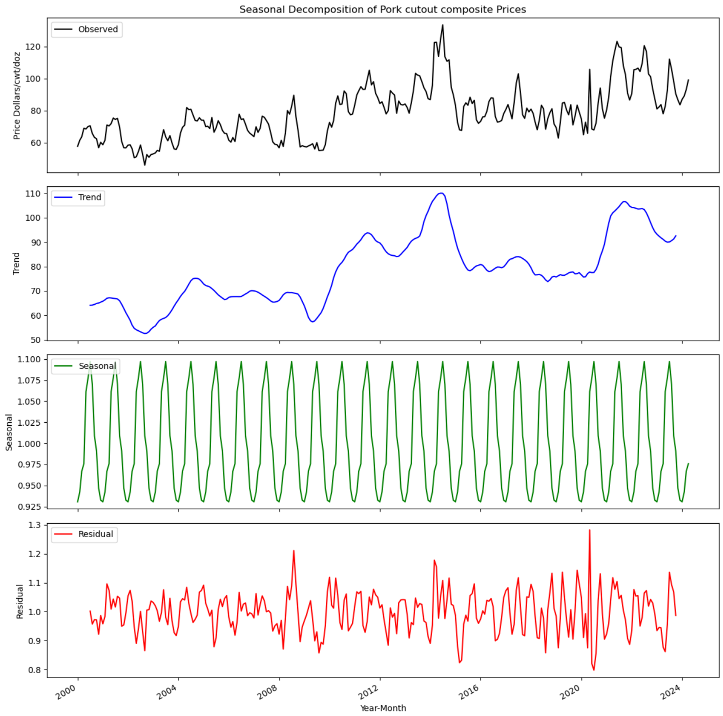

Figure 3.6: Pork Cutout Composite seasonality trend

Observed Data: This is the actual price trend of Pork Cutout Composite over the years. An upward trend is seen from year 2000 to 2024. The price fluctuations between about 50 and 150, with a notable spike reaching nearly 200. A significant price spike is visible around 2014-2015.

Trend: There is an upward trend visible, with many fluctuations seen around the years. As you see in the plot 3.6 in the years 2000-2010 gradual increase with minor fluctuations and in the year 2010-2015 steeper upward trend, with a sharp peak around 2014-2015, in the year 2015-2024 continued upward trend but with increased variability. The trend suggests a persistent long-term increase in pork cutout composite prices.

Seasonality: There is a Clear seasonal present, and the seasonal pattern remains stable throughout the observed period. The seasonal fluctuations remain constant, ranging between 0.925 and 1.100. Regular peaks and troughs suggest a predictable annual cycle in pork cutout prices.

Residual: Residuals mostly fall between 0.8 and 1.3, with some larger deviations. This plot 3.5 shows periods of increased volatility, particularly around the 2014-2015 price spike.

Loins, 1/4", Trimmed Vacuum-Packed analysis on the seasonality trend

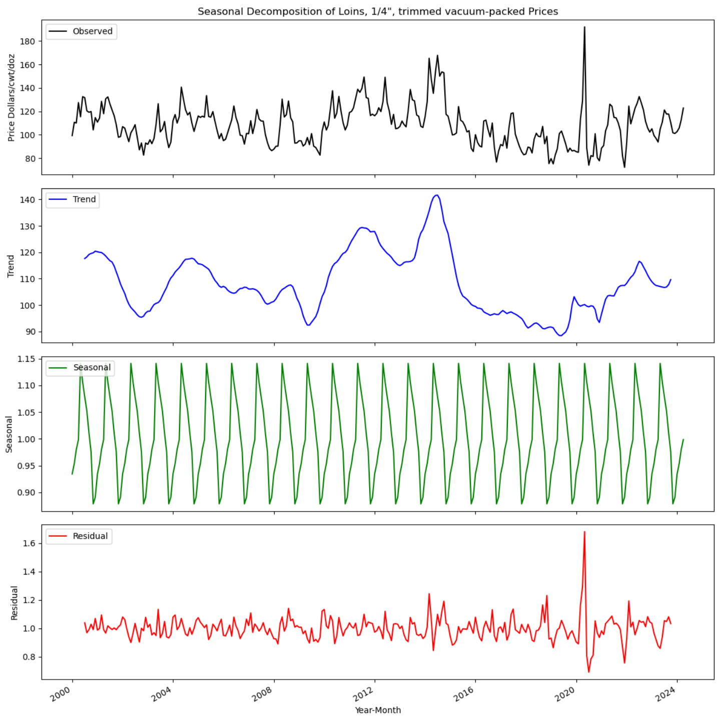

Figure 3.7: Pork Loins seasonality trend

Observed Data: This is the actual price trend Loins, 1/4", Trimmed Vacuum-Packed over years. An upward trend is seen from year 2000 to 2024. The price fluctuations between about 50 and 150, with a notable spike reaching nearly 200. A significant price spike is visible around 2014-2015 and 2020. There are short-term fluctuations, indicating a volatile market.

Trend: There is an upward trend visible, with many fluctuations seen around the years. As you see in the plot 3.7 in the years 2000-2010 gradual increase with minor fluctuations and in the year 2010-2015 steeper upward trend, with a sharp peak around 2014-2015, in the year 2015-2024 variable trend with periods of decline and recovery Despite short-term variability, the overall trend suggests a long-term increase in prices.

Seasonality: There is a Clear seasonal present, and the seasonal pattern remains stable throughout the observed period. The seasonal fluctuations remain constant, ranging between 0.90 and 1.15. Peaks occur at regular intervals, likely corresponding to annual cycles in supply and demand.

Residual: Residuals mostly fall between 0.8 and 1.6, with some larger deviations. This plot 3.5 shows periods of increased volatility, particularly around the 2014-2015 price spike.

Drumsticks analysis on the seasonality trend

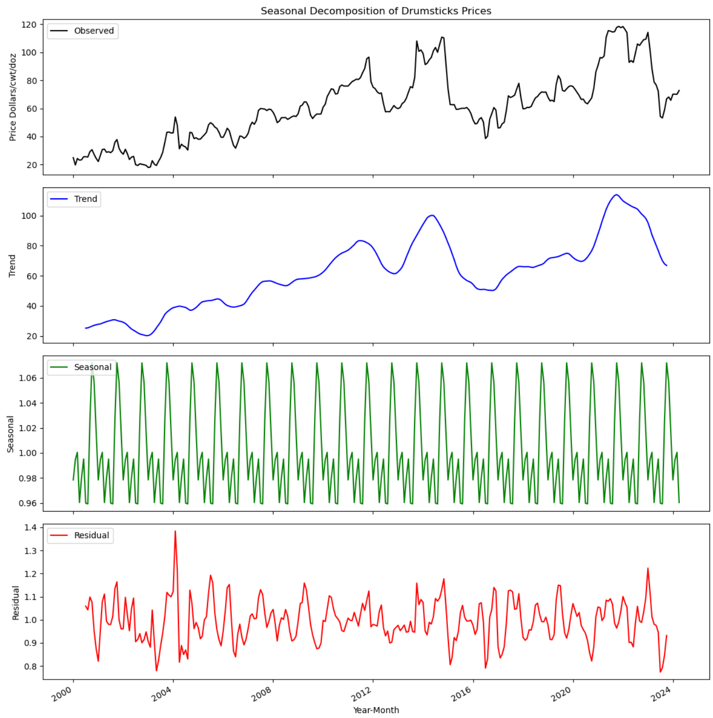

Figure 3.8: Drumsticks seasonality trend

Observed Data: This is the actual price trend of Drumsticks over the years. An upward trend is seen from year 2000 to 2024. The price fluctuations between about 40 and 140, with a notable spike reaching nearly 140. A significant price spike is visible around the years 2015-2016. Several smaller spikes occur throughout the series, indicating periodic market shocks.

Trend: There is an upward trend visible, with many fluctuations seen around the years. As you see in the plot 3.8 in the years 2000-2010 stable prices with a slight upward trend and in the years 2010-2015 an increase in price growth, in the years 2015-2024 continued upward trend with increased volatility.

Seasonality: There is a Clear seasonal present, and the seasonal pattern remains stable throughout the observed period. The seasonal fluctuations remain constant, ranging between 0.96 and 1.06. Peaks occur at regular intervals, likely corresponding to annual cycles in supply and demand.

Residual: Residuals mostly fall between 0.8 and 1.4, with some larger deviations.

This plot 3.8 large positive spike coincides with the years 2003 and 2004 and 2015-2016 price surge, indicating an extraordinary market event not fully captured by the trend and seasonal components. Several smaller spikes throughout the series suggest periodic market shocks or short-term supply-demand imbalances.

Wings, Full Cut analysis on the seasonality trend

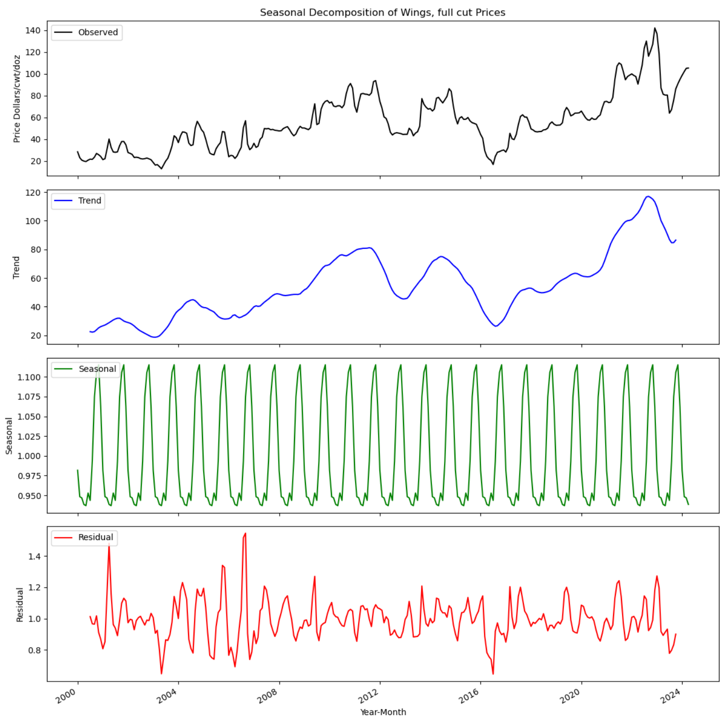

Figure 3.9: Wings seasonality trend

Observed Data: This is the actual price trend Wings, Full Cut over years. An upward trend is seen from year 2000 to 2024. The price fluctuations were about 20 and 140, with a notable spike reaching nearly 140. A significant price spike is visible around the years 2022-2023. Several smaller spikes occur throughout the series, indicating periodic market shocks.

Trend: There is an upward trend visible, with many fluctuations seen around the years. As you see in the plot 3.9 in the years 2000-2010 gradual upward trend with minor fluctuations. and in the years 2010-2015 steeper price growth, in the year 2015-2024 continued the upward trend with increased volatility and a slight leveling off towards the end.

Seasonality: There is a Clear seasonal present, and the seasonal pattern remains stable throughout the observed period. The seasonal fluctuations remain constant, ranging between 0.950 and 1.100. Peaks occur at regular intervals, likely corresponding to annual cycles in supply and demand.

Residual: Residuals mostly fall between 0.5 and 1.6. This plot 3.9 has a large positive spike, in the years 2004 to 2007 indicating an extraordinary market event like external factors.

3.3.5 Correlation plot on the external feature

The correlation matrix gives a brief understanding of how different economic and environmental variables affect the prices of different meat products. The correlations are shown in the form of heat maps, where the color represents the strength and direction of correlation between two variables, and the values are in the range -1 to 1. where -1 indicates negative correlation (if one variable increases the other decreases proportionally), 0 indicates no correlation between the variables and 1 indicates positive correlation (one variable increase, the other increases proportionally) as shown in figure 3.10.

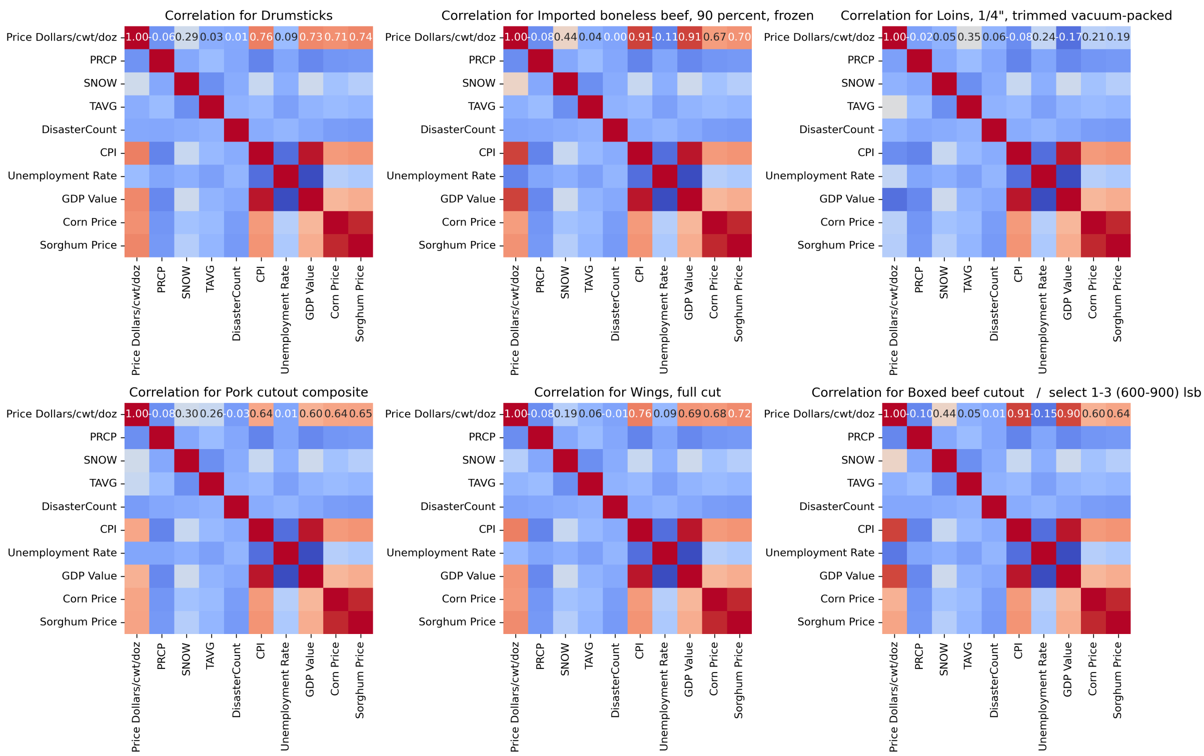

Figure 3.10: Correlation plot Heat Map For all Meat sub Types

Meat Sub Types	Top Correlated External Factor	Correlation Coefficient
Drumsticks	CPI	0.7566
Imported boneless beef, 90 percent, frozen	CPI	0.9071
Loins, 1/4", trimmed vacuum-packed	TAVG	0.3504
Pork cutout composite	Sorghum Price	0.6485
Wings, full cut	CPI	0.7594
Boxed beef cutout / select 1-3 (600-900) lbs	CPI	0.9098

Table 3.9: Summary of Top Correlated External Factors for Meat Sub Types

Boxed Beef Cutout / Select 1-3 (600-900) lbs

In boxed beef cutout price there is a strong positive correlation with CPI (0.91) as shown in table 3.9 which indicates inflation plays a major part in the boxed beef cutout. If the price level increases, the cost of production and distribution also needs to have a higher price. There are moderate correlations with GDP (0.60) which tells economic growth is also influenced by beef pricing strategies. Another strong correlation with feed grains like Corn and Sorghum (0.60 and 0.64) tells feed cost is a very crucial part of cattle raising. There is one more moderated correlation which is the unemployment rate (0.50), economic conditions affecting employment also influence beef prices, possibly through changes in demand or supply chain disruptions. Weather and disaster counts have weak correlations, suggesting they do not directly impact the price of boxed beef cutouts.

Imported Boneless Beef, 90 Percent, Frozen

By analyzing the correlations plot of Imported Boneless Beef a strong positive correlation with CPI (0.91) as shown in table 3.9 tells inflation impacts the price of beef products. As levels increase the cost of packaging, transporting, and feed cost also rises. The equally strong correlation with the unemployment rate (0.91) tells economic conditions affecting employment also influence beef prices, possibly through changes in demand or supply chain disruptions. The moderate correlation with feed grains (0.67 and 0.70) shows the cost of corn and sorghum rises along with the cost of beef production with higher prices. Weather and GDP have weak correlations where factors do not directly influence the price of beef products.

Pork Cutout Composite

The correlations plot of the Pork Cutout Composite has strong positive correlations with CPI (0.64) and GDP (0.60) tells pork prices are influenced by inflation and economic growth. As the cost of pork production increases, which leads to higher prices and GDP indicates as the economy grows the consumer demand for pork price moves upwards. There is another correlation between feed prices (0.64 and 0.65) as shown in table 3.9 important factors to determine the pork prices. The positive correlation with the unemployment rate (0.64) is due to complex market dynamics where higher unemployment affects the supply chain or demand leading to price increases. Weather and disaster count have a weak correlation suggesting they do not impact pork prices.

Loins, 1/4", Trimmed Vacuum-Packed

By analysing the correlations plot of Loins there is a weak correlation between CPI (0.06) and GDP (0.24) which shows that loin prices are lower from broader economic conditions and feed costs. This indicates the very stable demand for loins very sensitive to economic fluctuations and fewer variables compared to other meat products. The minimal impact from weather where TAVG (0.35) as shown in table 3.9 and disaster count tells loins prices are driven by factors that are not captured in these variables such as long-term contracts or market dynamics.

Drumsticks

By analyzing the correlations plot of Drumsticks a strong positive correlation with CPI (0.76) as shown in table 3.9 and (0.74) suggests that as inflation and economic growth increase, so does the price of drumsticks. This is because higher inflation raises the cost of production, and economic growth typically boosts consumer demand, driving prices up. The strong correlation with feed grains (0.73 and 0.74) indicates that the cost of poultry is more dependent on feed cost which directly impacts drumsticks prices. The positive correlation with the unemployment rate (0.73) could reflect supply chain disruptions or changes in consumer behavior during economic downturns, leading to higher prices. Weather and disaster count, however, show weak correlations, suggesting minimal direct impact from these environmental factors on drumstick prices.

Wings, Full Cut

By analyzing the correlations plot of Wings there is a strong positive correlation with CPI (0.76) as shown in table 3.9 indicates that inflation has a significant impact on wing prices, as higher inflation drives up the cost of production. The moderate correlation with GDP (0.69) economic growth that demands in increasing the wings prices. The strong correlation with feed grains (0.68 and 0.72) is the importance of feed cost in pricing wings as poultry production is heavily dependent on feed. Importance of feed costs in the pricing of wings, as poultry production is heavily dependent on feed. The weak correlation with the unemployment rate (0.09) suggests that wing prices are less sensitive to changes in employment levels. Weather and disaster count also show weak correlations, indicating minimal direct impact on wing prices from these factors.

3.4 Models Selection

In this Research, we aim to forecast the future prices of various meat sub-types using time series data. To achieve accurate price predictions, we employ five different models: Seasonal Autoregressive Integrated Moving Average (SARIMA), Random Forest, Gradient Boosting, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks. Each of these models was selected based on its unique capabilities and advantages in handling different aspects of time series data, like seasonality, non-linearity, complex patterns, and long-term dependencies. The combination of these models allows us to explore different approaches and choose the most effective method for forecasting meat prices.

3.4.1 Framework of SARIMA model

The SARIMA model follows as, implementation, and evaluation to forecast the prices of various meat subtypes. The analysis started by assessing the stationary of the time series data for each meat subtype and applied necessary transformations like optimal parameters using ACF and PACF plots utilized SARIMA models for forecasting the price.

Stationary Tests and Data Preparation

The time series data can be classified into stationary and non-stationary. Stationary is an important property, as some models work well when the data is stationary. However, time series data often possesses the non-stationary property. By working on the SARIMA model, the first step is to check the number of differences required to make the time series stationary for each meat type. A time series is considered stationary when its mean, variance, and autocorrelation structure remain constant all the time, based on this it is a very critical assumption for the SARIMA model as it mostly relies on the patterns to make accurate predictions. The two important tests were performed to check the stationary of time series on different meat subtypes follows as:

1. Augmented Dickey-Fuller (ADF) Test:

This statistical test checks the null hypothesis (H0) which is the unit root, indicating whether the time series data for each meat subtype is non-stationary. The alternative hypothesis (H1) is where the time series data for each meat subtype is stationary or trend stationary. If p-values are below 0.05 suggests that time series data for each meat sub-type is stationary.

2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test:

This is also a statistical test that checks the stationary of the time series. Like the (ADF) test for unit root (non-stationary), the KPSS test got the reverse null hypothesis (H0) that tests whether the time series is stationary around deterministic trends. So, in this test Null hypothesis (H0) is stationary and the alternative hypothesis (H1) is non-stationary. If p-values are below 0.05 suggests that time series data for each meat subtype is non-stationary.

Applied Differencing to Achieve Stationary

To make each meat sub-type time series data stationary, the differencing method was applied in both test cases. The role of differencing is to remove trend and seasonality from the meat sub-type time series data and stabilize the mean and variance from the dataset. By doing this process it helps in removing the time-dependent structure and making the series stationary and ready for predictive modeling. For example, in this Meat price dataset, the first-order differencing ($d = 1$) was very effective in making stationary for most meat subtypes, as it removes trends and stabilizes the mean from the dataset. After differencing was done, the data was reevaluated using the ADF And KPSS tests which confirmed stationary in the transformed series.

ACF and PACF Plots Analysis for Parameter Selection

After Data becomes stationary plotting of ACF and PACF is a must as they will be giving exact parameters that will help the model to get good accuracy. Below is how each plot defines its parameters:

ACF Plot: This plot shows the correlations between the time series value and their lagged values. This plot also helps to identify the MA (q) showing the lags at which autocorrelation significantly differs from zero.

PACF plot: This plot shows the correlations between the time series value and their lagged values after removing the effects of the previous lags. This plot helps identify the AR (p) component by showing the lags at which partial autocorrelation cuts off.

Introduction to the SARIMA Model

The Seasonal ARIMA (SARIMA) model is an extension of the ARIMA model by adding seasonal components and incorporating exogenous variables. It can capture both seasonal and non-seasonal patterns in the time series data while adjusting for external influences such as weather conditions, economic indicators, and commodity prices. To break down each component:

$$SARIMA\,\underbrace{(p,d,q)}_{non-seasonal}\,\underbrace{(P,D,Q)_m}_{seasonal}$$

Figure 3.11: SARIMA Parameters [16]

Working of the SARIMA Model

ARIMA Component:

Auto regressive (AR): The Auto regressive model, denoted as AR(p), makes predictions based on previously observed values. It can be expressed as:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t \tag{3.2}$$

where:

- X_t represents the observation at time t,
- ϕ_i (for $i = 1, 2, \ldots, p$) are the autoregressive coefficients,
- p specifies the number of previous data points to consider,

- e_t denotes the error term at time t.

Integrated (I): The d-component (I) represents the number of times the series needs to be differenced to make it stationary. In this case, the dataset is set to first-order differencing (d=1) which was very effective in making stationary for most meat subtypes.

Moving Average (MA): The Moving Average model, denoted as MA(q), adjusts the model based on the average prediction errors from previous q observations. It can be expressed as:

$$X_t = \mu + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \tag{3.3}$$

where:

- X_t represents the observation at time t,
- μ is the mean of the series,
- θ_i (for $i = 1, 2, \ldots, q$) are the moving average coefficients,
- q specifies the number of previous error terms to include in the model,
- e_t denotes the error term at time t.

Non-Seasonal Component:

Additional parameters (p, d, q) where p: Trend autoregression order, d: Trend difference order., q: Trend moving average order.

Seasonal Component:

Additional parameters (P, D, Q) where P: Seasonal autoregressive order, D: Seasonal difference order, Q: Seasonal moving average order, m: The number of time steps for a single seasonal period.

Creating Input Lag values for all the machine learning models

In traditional machine learning models, lag values are arranged in a 2D feature matrix format, where each row is a sample, and each column represents a lagged value in deep neural networks like the LSTM model, the input data must be formatted into 3D arrays (samples, time steps, features) to take advantage of their sequential learning capabilities.

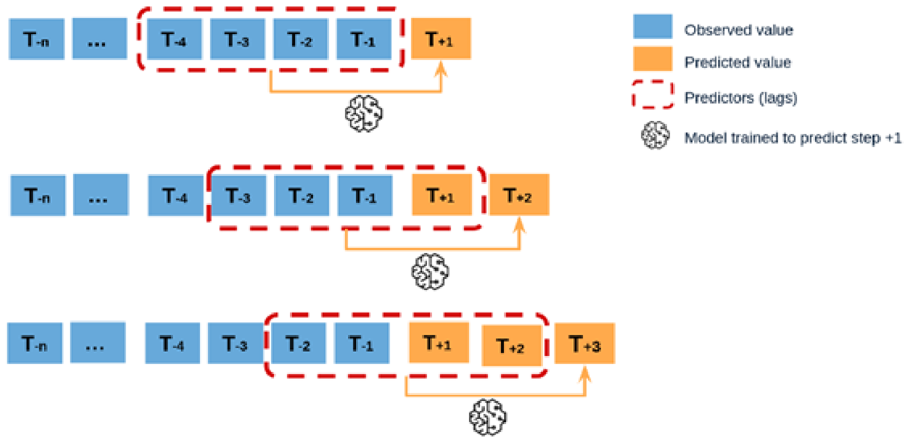

Figure 3.12: Lags Creations [19]

Lag values in time series analysis represent past observations of the target variable used as features to predict future values. If we are forecasting the price of meat for the next month, lag values might include the prices from the previous month, two months ago, and three months ago. These lagged values help the model understand patterns in the data, such as trends, seasonality, and auto-regressive relationships, which are crucial for making accurate predictions as seen in figure 3.12.

In time series forecasting, a key step in preparing the dataset was creating the input features. This process involves rearranging the historical price data to create a supervised learning problem where past price values (lags) predict future values, for example in all my machine learning models the array arrangement functions were used to create a sequence of past prices (lags) for each meat subtype. This method takes an integer i (lag size) and creates an input-output pair (x, y) where x contains past, i price values and y contains the price value to be predicted. This process was repeated for varying lag sizes (ranging from 1 to 50) to identify the optimal lag that maximized model performance. The other approach enhances these sequences by integrating additional external features such as weather, consumer price index, and commodity prices, which are normalized and combined with the historical data to provide a more contextual understanding and improve the model's predictive accuracy by incorporating influential factors beyond past trends.

3.4.2 Random Forest Regression Models

Random Forest is a supervised machine-learning model which is used for regression and classification problems. This model uses the strategies of combining the ensemble learning method by creating multiple decision trees on many subsets of the data during training and merges them to get stable mean predictions. For regression tasks predicting the prices of various meat sub-types like Boxed beef cutouts, Loins, and Drumsticks where the model handles many input features, such as historical price data, seasonality factors, and external economic features. It helps in capturing the relationship between these inputs and the target variable (meat price) by training multiple decision trees and averaging (Voting) their predictions. Each tree in the forest learns to make predictions based on a unique subset, which makes the overall model less likely to be biased toward a specific pattern in the data.

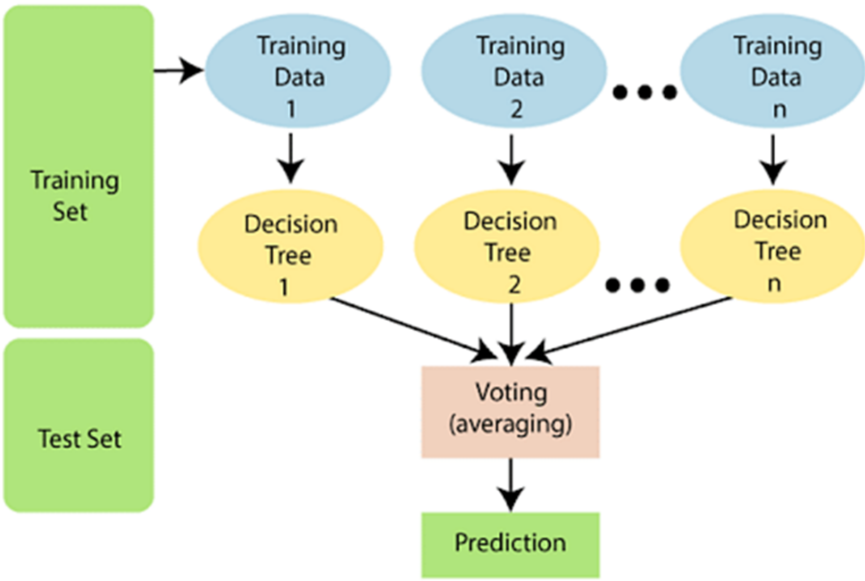

Figure 3.13: General Working Of Random Forest [20]

General working Random Forest Algorithm:

Step 1: Bootstrap Sampling or Bagging

From Figure 3.13 this algorithm starts creating multiple subsets from the sample training data. Each subset is created by random sampling with a replacement called bagging, where a few data points may be included multiple times where others might not be included at all.

Step 2: Boosting

This process works by combining weak learners into strong learners by creating sequential models, where the final model that has the highest accuracy is called boosting.

Step 3: Decision Tree Building

For each bootstrap sample, a decision tree is built. In traditional decision trees, the Random Forest will only consider a random subset of features at each split, which introduces more diversity to the trees.

Step 4: Prediction Aggregation

For regression tasks like predicting meat prices, all the trees are averaged to produce the result.

Step 5: Feature Importance

Random Forest can also provide a measure of feature importance, where each node of the tree is randomly selected as a feature which indicates how much each feature contributes to the model's predictions. The tree chooses out of all particularly useful for understanding which factors (e.g., previous price, economic indicators) mostly influence the price of each meat sub-type.

3.4.3 Gradient Boosting Regression Model

Gradient Boosting is well known for its powerful machine-learning technique used in classification and regression problems. It also uses an ensemble learning method where it sequentially uses many decision trees and each tree in the model learns to improve upon the mistakes of the previous trees, capturing intricate patterns and relationships in the data. It does this by minimizing a specific loss function using gradient descent. This model is very effective for regression tasks with complex nonlinear relationships like predicting the prices of different meat sub-types influenced by various market factors.

General working Gradient Boosting Regression Algorithm:

Step 1: Initial Model Prediction

Gradient Boosting starts with a simple model as the initial prediction, that is Weak classifier as shown in the figure 3.14. It is called a decision tree on its own but can be combined to form a strong model.

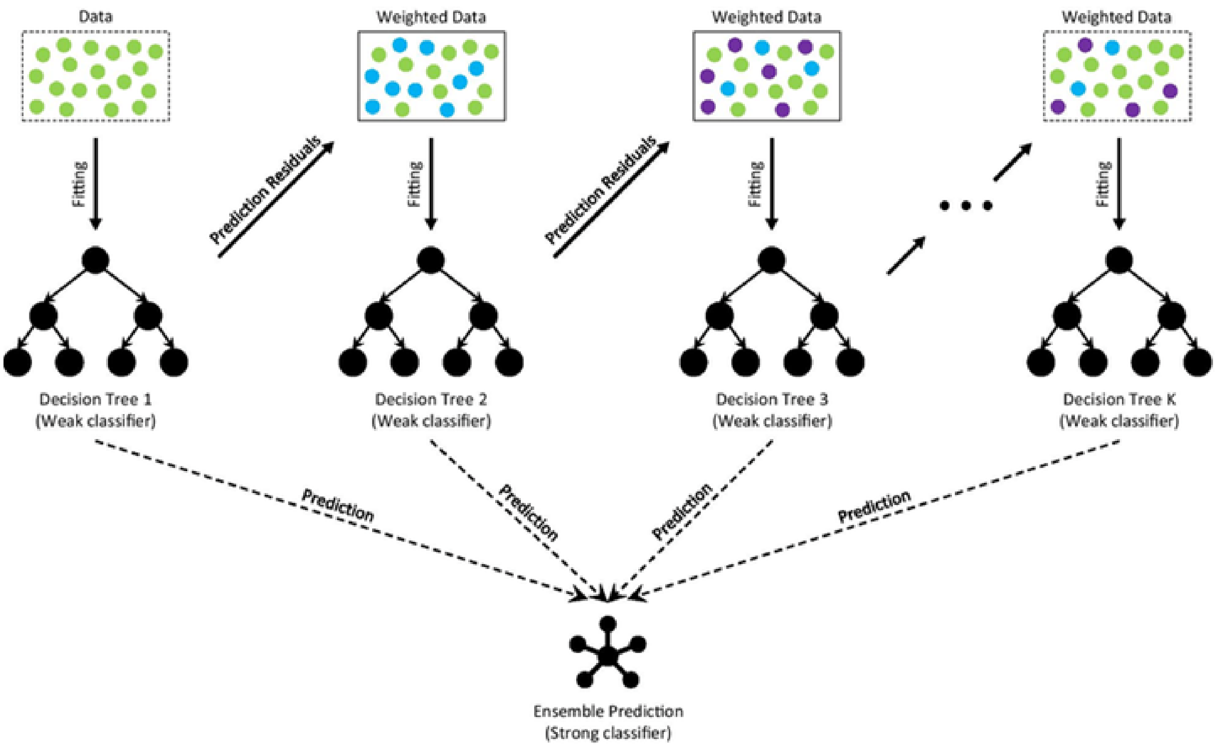

Figure 3.14: Gradient Boosting Regression Working [24]

Step 2: Sequential Tree Building

For regression tasks, each tree is built to predict the residuals (errors) of the previous trees, thereby correcting the mistakes sequentially.

Step 3: Learning Rate

A parameter that controls the contribution of each tree just by adding them, and is scaled by a small number called the learning rate (typically between 0.01 and 0.3) before being added. This scaling ensures that each tree contributes only a fraction of its predictions to the final model.

Step 3: Regularization Techniques

Regularization methods such as limiting tree depth improve the model's generalization capabilities.

Step 3: Final prediction

The final prediction is obtained by summing up the predictions from all the weak learners and considering the learning rate. For regression tasks, this sum represents the predicted target value.

3.4.4 Support Vector Regression (SVR)

Support Vector Regression (SVR) is also a powerful supervised machine learning model that works on the principle of support vector machines (SVM) in regression problems. This model is useful for handling nonlinear relationships with input continuous target variables while minimizing the prediction error in time series data, like those observed in meat prices influenced by a variety of economic, market, and seasonal factors. SVR model uses kernel functions to transform the input features (historical prices) into a higher-dimensional space by adding a linear hyperplane to the data and the model tries to find the best-fitting hyperplane that has the maximum margin of tolerance epsilon around it, minimizing the errors while ignoring data points within this margin.

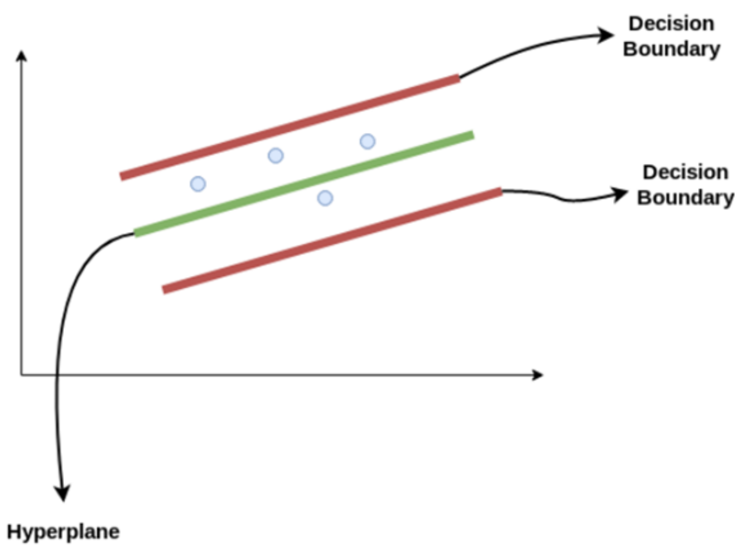

Figure 3.15: Working Of SVR Model [23]

General working Support Vector Regression Algorithm:

Step 1

From the figure 3.15, the hyperplane with the green line in the graph is a best-fit line that can predict the prices of different meat subtypes accurately and minimize the error. A hyperplane is a linear regression that only fits a straight line where most of the data points (prices) lie close to [23].

By [23] hyperplane (our best-fit line) can be written as:

$$Y = wx + b \tag{3.4}$$

- Y represents the predicted price,
- w represents the weights or coefficients that determine the slope of the line,
- b is the bias (intercept).

Step 2

The two red lines in the graph are decision boundaries where these are drawn from a certain distance from the hyperplane, creating a tube around it. The distance is given as a, and the lines are drawn at distances $+a$ and $-a$ from the hyperplane. This a in the text is essentially referred to as epsilon, which allows flexibility for the prices to deviate slightly from predicted values [23]. For example, if the price of drumstick is around \$1.50 per cwt, epsilon defines how much higher or lower than \$1.50 per cwt you are allowed to be before considering it an error.

By [23] decision boundaries can be written as:

$$wx + b = +a \tag{3.5}$$

$$wx + b = -a \tag{3.6}$$

Hence, we will take only those points within the decision boundary that have the least error rate or are within the Margin of Tolerance. This will give us a better-fitting model.

3.4.5 LSTM (Long Short Term Memory)

The LSTM (Long Short-Term Memory) model has become a powerful tool in artificial intelligence and deep learning. It is a type of Recurrent Neural Network (RNN) that handles sequential data with long-term dependencies making it ideal for sequence prediction tasks, whereas the traditional neural LSTM network has a memory component that stores important information for long periods to help in predicting future prices based on historical price trends. For example, it can learn seasonality effects on drumstick meat prices and capture long-term market trends affecting loan prices.

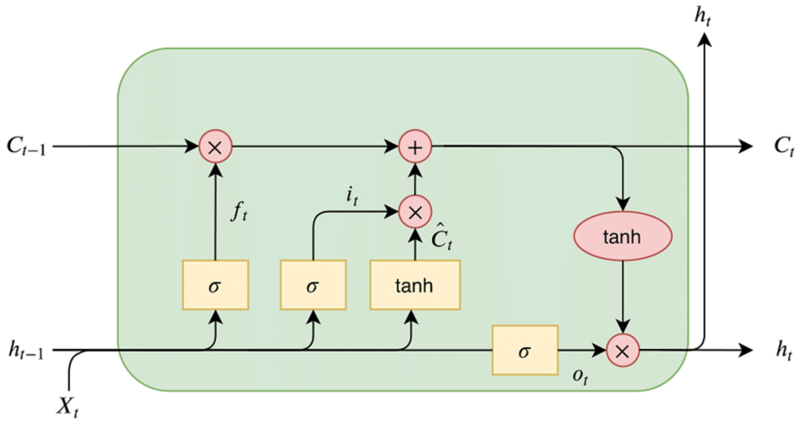

Figure 3.16: LSTM Model Working [21]

General working LSTM Algorithm:

Memory cells: So, while implementing the meat sub-type data into the LSTM model the memory cell keeps all the important information about the past meat prices and how they are affected by all external factors over time.

Hidden State: This is an internal state with short-term memory as it passes from one time step to the next carrying information that is immediately very relevant in making predictions.

Three Gates: The three parts of LSTM units are known as gates which control the flow of information in and out of the memory cell. Well, the first gate is known as the forget gate whereas the second gate is known as the input gate and the last one is the output gate. In the LSTM model units of these three gates and memory cells are known as layers of neurons in the feed-forward neural network, with each neuron having a hidden layer and current state.

Forget Gate (ft): In this gate it decides which part of the information from the previous cell state should be kept or forgotten. For example, if the meat price 12 months earlier is not important for predicting the next month's price the forget gate might give an output value close to 0 saying that information should be forgotten [22].

By [22] LSTM forget gate equation can be written as :

$$f_t = \sigma(x_t \cdot U_f + H_{t-1} \cdot W_f) \tag{3.7}$$

- Here, x_t is the input at time t (e.g., the current month's price and external features),

- H_{t-1} is the hidden state from the previous month.
- U_f is the weight associated with the input,
- W_f is the weight associated with the hidden state,
- The sigmoid function σ produces a value between 0 and 1 that decides how much of the past information to retain.

Input Gate: Basically, this gate works like any new information of the current time step that will be added to the cell state. For example, if the current month's external factor (like CPI suddenly increases in price) will be important for predicting future prices the input gate will allow this information to be added to the memory cell.

By [22] LSTM Input gate equation can be written as:

$$i_t = \sigma(x_t \cdot U_i + H_{t-1} \cdot W_i) \tag{3.8}$$

- Here, x_t is the input at time t (e.g., the current month's price and external features),
- H_{t-1} is the hidden state from the previous month.
- U_i is the weight associated with the input,
- W_i is the weight associated with the hidden state,
- The sigmoid function σ produces a value between 0 and 1 that decides how much of the past information to retain.

By [22] LSTM New Information equation can be written as :

$$N_t = \tanh(x_t \cdot U_c + H_{t-1} \cdot W_c) \quad \text{(New Information)} \tag{3.9}$$

Update Cell State: In this process the cell state will be updated by combining forget and input gates. For Example, if the forget gate decides to forget 80% of the previous cell state and the input gate decides to add 50% of the new cell state this will be updated in the cell state by combining these two gates.

By [22] LSTM Update cell state equation can be written as :

$$C_t = f_t \cdot C_{t-1} + i_t \cdot N_t \tag{3.10}$$

- C_{t-1} is the cell state at the current timestamp, and the others are the values we have calculated previously.

Output Gate: This gate follows which part of the cell state should be output as a new hidden state and will be used for the next time step. For example, if the model predicts that current external factors like a sharp drop in TAVG will heavily influence future prices, the output gate will reflect this price [22].

By [22] LSTM Output gate equation can be written as :

$$O_t = \sigma(x_t \cdot U_O + H_{t-1} \cdot W_O) \tag{3.11}$$

- H_{t-1} The hidden state is then updated and passed to the next LSTM cell to continue learning from the sequence.

3.5 Model Evaluation

3.5.1 R2 Score (Coefficient of Determination):

The R2 score is known for its proportion of variance in the dependent variable that is predictable from the independent variables. It tells how well the model's predictions match the actual data points. The R2 score ranges from 0 to 1 and follows a few points:

1. If the R2 score is close to 1 that indicates that the model captures most of the variance in the data, which tells it a good fit which is predicting the data points very accurately.

2. If the R2 score is 0 or close to 0 this tells that the model does not capture any variance in the data effectively and further requires optimization.

By [25] Formula of R2 Score can written as:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} \tag{3.12}$$

where:

- RSS represents the sum of squares of residuals,

- TSS represents the total sum of squares.

3.5.2 Root Mean Squared Error (RMSE):

RMSE measures the average magnitude of the prediction errors (i.e. the differences between predicted and actual values). It is the square root of the mean of the squared errors.

RMSE (Root Mean Squared Error) measures the average magnitude of the prediction error (i.e. difference between predicted and actual values). It is the square root of mean squared errors. The lower RMSE value is a better model fit for the dataset. The RMSE follows a few points:

1. A lower RMSE value tells more accurately the model works with fewer errors in the predictions dataset.

2. RMSE is very sensitive to outliers meaning larger errors will have a disproportionate impact on this metric.

By [25] Formula of RMSE can written as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{3.13}$$

where:

- $\sum$ is a symbol that means "sum",
- P_i is the predicted value for the i-th observation,
- O_i is the observed value for the i-th observation,
- n is the sample size.

CHAPTER 4

Results

In this chapter, we try to provide the result of the method discussed in the previous chapter based on metrics such as R2-Score and RMSE values. To enhance the accuracy of our predictions, hyperparameter tuning was performed for each model, involving a systematic search for optimal parameter combinations using techniques like grid search, while also incorporating external factors, such as seasonal trends and economic indicators, thereby allowing us to compare the models' predictive capabilities and identify the most reliable model for forecasting future price movements.

4.1 SARIMA Model Results

The SARIMA model helps to forecast the prices of various meat sub-types over a 5-month horizon. It includes the outcomes of stationary tests, differencing, initial model building, hyperparameter tuning, and the impact of incorporating external features.

Stationary Tests Output

To determine if the time series data for each meat sub-type was stationary, we performed two tests: the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. Test output on stationary or non-stationary:

4.1.1 Stationary Test Output:

- **Boxed Beef Cutout:** ADF Test: p-value = 0.95, KPSS Test: p-value = 0.01. The series is likely non-stationary.
- **Imported Boneless Beef:** ADF Test: p-value = 0.80, KPSS Test: p-value = 0.01. The series is likely non-stationary.
- **Pork Cutout Composite:** ADF Test: p-value = 0.36, KPSS Test: p-value = 0.01. The series is likely non-stationary.
- **Loins, 1/4", Trimmed Vacuum-Packed:** ADF Test: p-value = 0.09, KPSS Test: p-value = 0.01. The series is likely non-stationary.
- **Drumsticks:** ADF Test: p-value = 0.18, KPSS Test: p-value = 0.01. The series is likely non-stationary.
- **Wings, Full Cut:** ADF Test: p-value = 0.28, KPSS Test: p-value = 0.01. The series is likely non-stationary.

4.1.2 Evaluation of Stationary Tests Results

By analyzing output for all meat sub-types, the time series was likely to be non-stationary, as indicated by the high p-values in the ADF test and low p-values in the KPSS test. Therefore, differencing was applied to make the data stationary

4.1.3 Differencing Output

To convert the non-stationary series into stationary ones, we applied first-order differencing and tested again using ADF and KPSS tests.

- **Boxed Beef Cutout / Select 1-3 (600-900 lbs):** First-order Differencing: ADF Test: p-value = 5.68×10^{-7}, KPSS Test: p-value = 0.1. The series is likely stationary.
- **Imported Boneless Beef, 90 Percent, Frozen:** First-order Differencing: ADF Test: p-value = 1.82×10^{-8}, KPSS Test: p-value = 0.1. The series is likely stationary.
- **Pork Cutout Composite:** First-order Differencing: ADF Test: p-value = 6.61×10^{-15}, KPSS Test: p-value = 0.1. The series is likely stationary.

- **Loins, 1/4", Trimmed Vacuum-Packed:** First-order Differencing: ADF Test: p-value = 6.53×10^{-10}, KPSS Test: p-value = 0.1. The series is likely stationary.

- **Drumsticks:** First-order Differencing: ADF Test: p-value = 1.30×10^{-21}, KPSS Test: p-value = 0.1. The series is likely stationary.

- **Wings, Full Cut:** First-order Differencing: ADF Test: p-value = 5.73×10^{-24}, KPSS Test: p-value = 0.1. The series is likely stationary.

4.1.4 Evaluation of Differencing Results

First-order differencing was sufficient to make all meat sub-types stationary, as indicated by the low p-values in the ADF test and higher p-values in the KPSS test. Further differencing was not necessary.

4.1.5 ACF and PACF Output for Each Meat Sub type:

After achieving stationary, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were used to identify appropriate parameters for the SARIMA model (p, d, q, P, D, Q, m).

Boxed Beef Cutout:

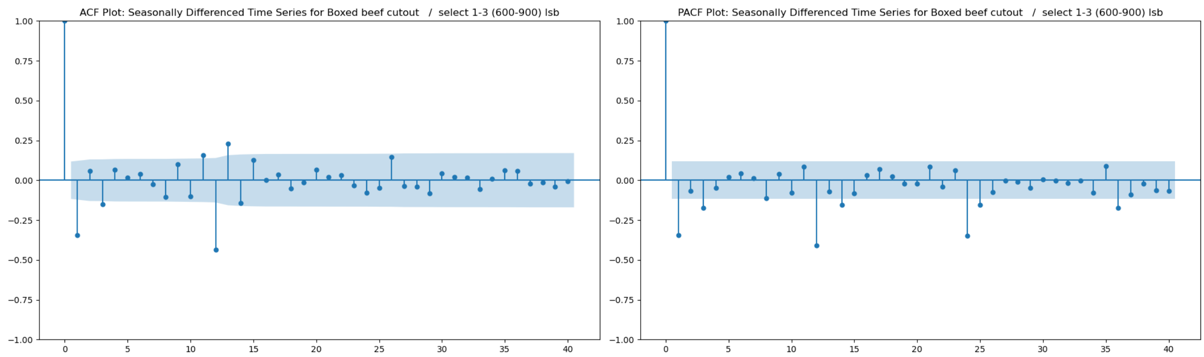

Figure 4.1: ACF AND PACF plot for Boxed Beef Cutout

- **ACF:** Significant spike seen in the ACF figure 4.1 at lag 1 and 2, indicating a possible Moving Average (MA) order of 2. The spike at lag 1 is strong and negative, suggesting the first lag could drive the MA component, but the spike at lag 2 of MA can be considered.

- **PACF:** Sharp cutoff after lag 1 seen in the PACF figure 4.1 suggesting an Auto-Regressive (AR) order of 1.
- **Selected Parameters:** $p = 1$, $d = 1$, $q = 2$.

Imported Boneless Beef:

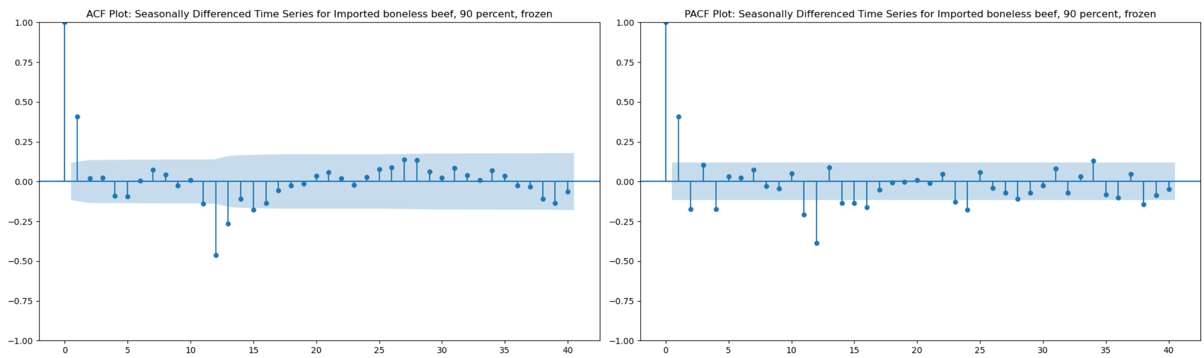

Figure 4.2: ACF AND PACF plot for Imported Boneless Beef

- **ACF:** There is a spike at lag 1 seen in the ACF figure 4.2 indicating strong negative autocorrelation. This suggests an MA order of 1.
- **PACF:** There is a strong negative partial autocorrelation at lag 1 sharp cutoff seen in the PACF figure 4.2 . This indicates an AR order of 1.
- **Selected Parameters:** $p = 1$, $d = 1$, $q = 1$.

Pork composite cutout:

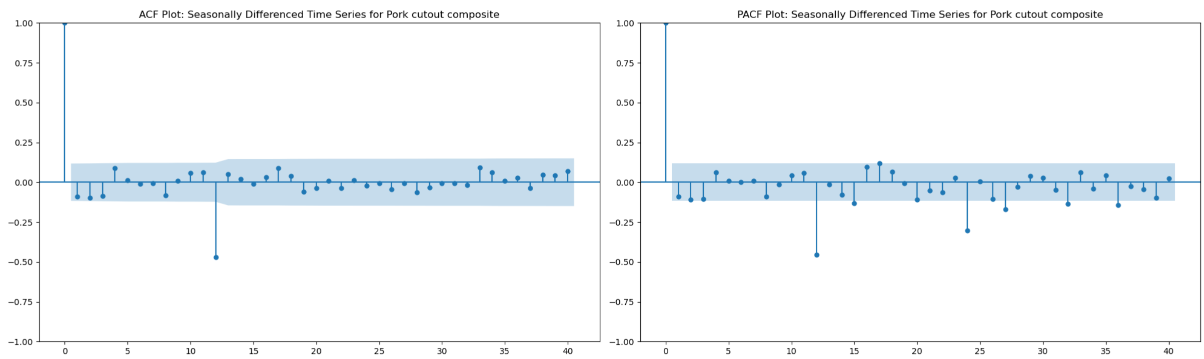

Figure 4.3: ACF AND PACF plot for Pork composite cutout

- **ACF:** There are spikes at early lags, particularly around lag 12 as seen in the ACF figure 4.3 indicating some remaining seasonality or autocorrelation that may require a seasonal MA term.

- **PACF:** The PACF figure 4.3 shows a significant spike at lag 12 with a sharp cutoff afterward, suggesting the presence of a seasonal AR component.

- **Selected Parameters:**$p = 1$, $d = 1$, $q = 0$, $P = 1$, $D = 1$, $Q = 1$ because there is seasonal order around lag 12

loins:

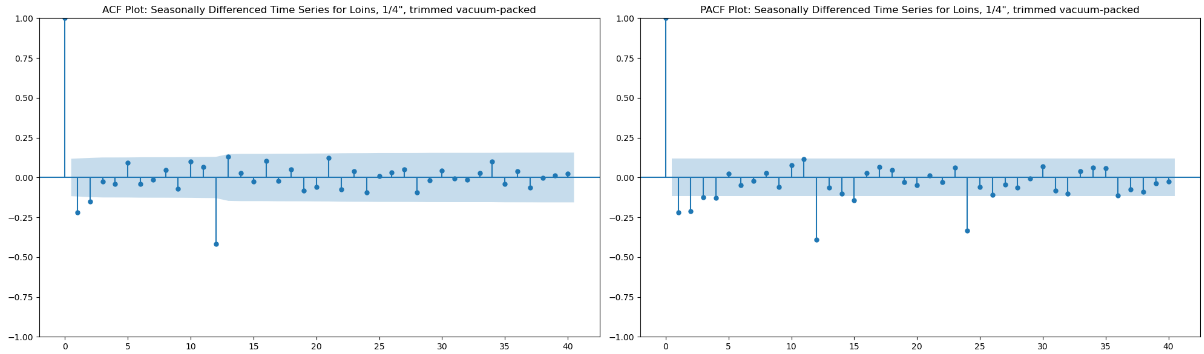

Figure 4.4: ACF AND PACF plot for loins

- **ACF:** There is a significant spike at lag 1 as seen in the ACF figure 4.4 followed by a sharp decline with most of the subsequent lags within the confidence bounds.

- **PACF:** There is a significant spike at lag 1 with a sharp cutoff afterward as seen in the PACF figure 4.4 and all other lags remaining within the confidence bounds.

- **Selected Parameters:**$p = 1$, $d = 1$, $q = 1$.

Drumstick:

- **ACF:** There are noticeable spikes around lag 12 as seen in the ACF figure 4.5 indicating possible seasonality. However, the spikes are within the confidence bounds for most other lags, suggesting limited autocorrelation at those lags. This may suggest a seasonal moving average (MA) component.

- **PACF:**: There is a significant spike at lag 12 with a sharp cutoff afterward as seen in the PACF figure 4.5 indicating a potential seasonal autoregressive (AR) order.

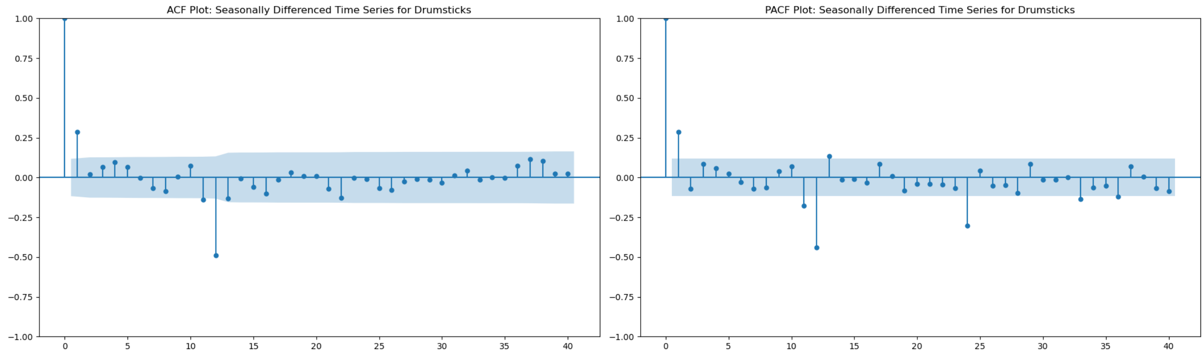

Figure 4.5: ACF AND PACF plot for Drumstick

- **Selected Parameters:** $p = 1$, $d = 1$, $q = 0$, $P = 1$, $D = 1$, $Q = 1$ because there is seasonal order around lag 12

Wings:

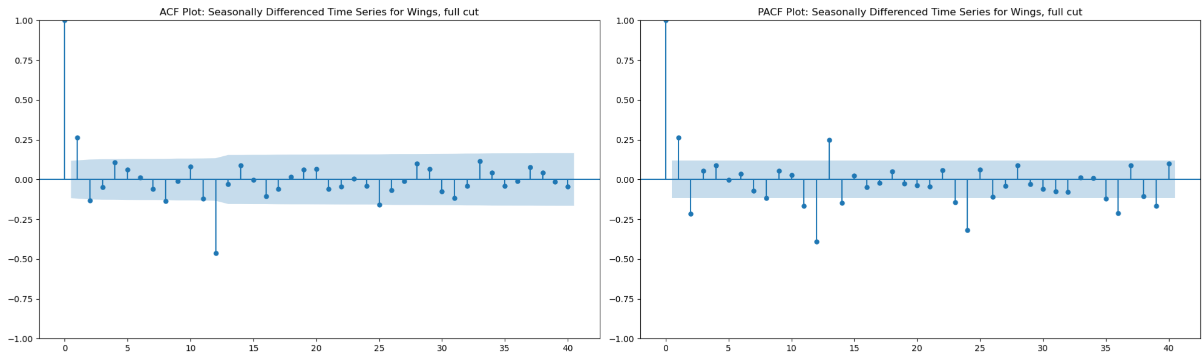

Figure 4.6: ACF AND PACF plot for Wings

- **ACF:** Spikes are observed in the ACF figure 4.6 around lags 12 and 24 suggesting strong seasonal autocorrelation. This indicates a seasonal MA order.
- **PACF:** The PACF figure 4.6 indicates significant spikes at lag 12, with a sharp decline, thereafter, indicating a possible seasonal AR component.
- **Selected Parameters:** $p = 1$, $d = 1$, $q = 0$, $P = 1$, $D = 1$, $Q = 1$ because there is seasonal order around lag 12

4.1.6 ACF and PACF plots Result

Based on the ACF and PACF plots, initial model parameters were selected for each meat sub-type. However, initial performance was low for some types, indicating the need for further hyperparameter tuning optimization.

4.1.7 Hyperparameter Tuning Process

Hyperparameter tuning was conducted using a grid search to identify the optimal combination of parameters like (p, d, q) and (P, D, Q, m) that maximized the R2 score and minimized the Root Mean Squared Error (RMSE). The results for each meat sub-type are summarised below:

Parameter Grid:

AR order (p): [0, 1, 2]

Differencing order (d): [1]

MA order (q): [0, 1, 2]

Seasonal AR (P), MA (Q), Differencing (D): [1]

Seasonal period (m): [12, 24, 36, 48]

4.1.8 Model Forecasting with Hyperparameter Tuning Results:

The optimal parameters were selected for each meat subtype based on the highest R2 score and RMSE value.

Meat Sub-Type	Best P	Best Q	Seasonal m	Best R^2 Score	Best RMSE
Boxed beef cutout / select 1-3 (600-900 lbs)	1	0	36	0.8370	0.0197
Imported boneless beef, 90 percent, frozen	0	0	46	0.6317	0.0495
Pork cutout composite	1	0	10	0.8062	0.0268
Loins, 1/4", trimmed vacuum-packed	1	1	48	0.9771	0.0100
Drumsticks	1	1	35	0.7889	0.0104
Wings, full cut	0	0	38	0.8084	0.0134

Table 4.1: Final Output of Hyperparameter Tuning Using SARIMA Model for All Meat Sub-Types

Boxed Beef Cutout / Select 1-3 (600-900 lbs):

The model's best parameters were identified as $P = 1$, $Q = 0$, and Seasonal $m = 36$. This strong R2 score of(0.8370) and an RMSE of (0.0197), indicate a good fit of the model to the data and a relatively low forecast error.

Imported Boneless Beef, 90 Percent, Frozen:

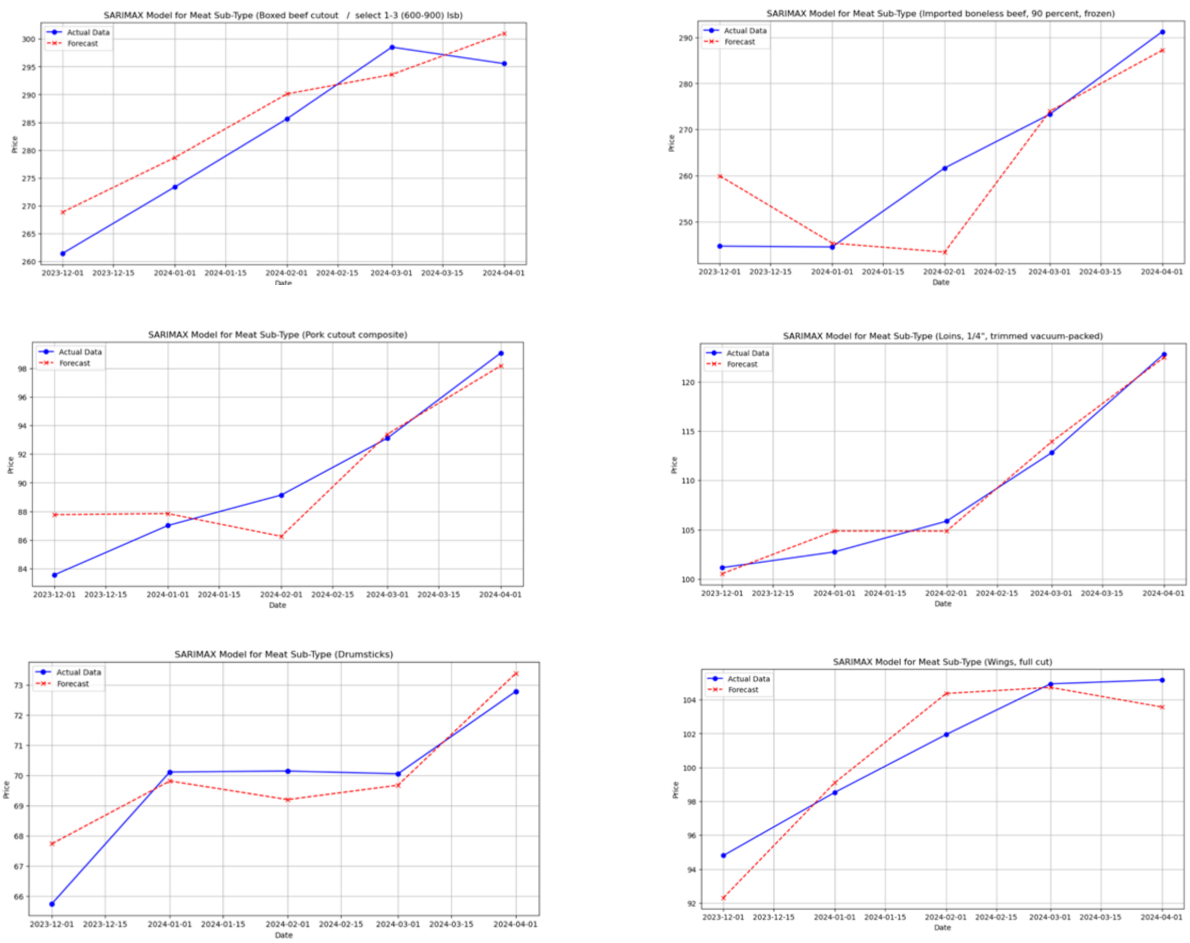

Figure 4.7: Comparison of actual versus predicted prices for different meat subtypes over the last 5 months using SARIMA model.

For this meat type, the optimal parameters were $P = 0$, $Q = 0$, and Seasonal $m = 46$, resulting in an R2 score of (0.6317) and an RMSE of (0.0495). While the R2 score is moderate, it suggests that the model captures some of the variance in the data.

Pork Cutout Composite:

The model parameters were $P = 1$, $Q = 0$, and Seasonal $m = 10$. This combination provided a robust R2 score of (0.8062) and an RMSE of (0.0268), indicating good predictive accuracy.

Loins, 1/4", Trimmed Vacuum-Packed:

The model achieved an R2 score of (0.9771) with an RMSE of 0.0100 using parameters $P = 1$, $Q = 1$, and Seasonal $m = 48$. This suggests that the model very accurately captured the patterns in the data for this meat type.

Drumsticks:

The optimal parameters were $P = 1$, $Q = 1$, and Seasonal $m = 35$. The resulting R2 score was (0.7889) with an RMSE of (0.0104), indicating a relatively good fit.

Wings, Full Cut:

The best parameters were $P = 0$, $Q = 0$, and Seasonal $m = 38$, yielding an R2 score of (0.8084) and an RMSE of (0.0134), which indicates a strong predictive performance.

4.1.9 Incorporating External Features Results

To make the SARIMA model performance more powerful, top external variables were incorporated as shown in the table 3.9 and they are:

Weather Conditions: Average Temperature

Economic Indicators: Consumer Price Index (CPI),

Commodity Prices: sorghum prices.

Meat Sub-Type	Best P	Best Q	Seasonal m	Best R^2 Score	Best RMSE
Boxed beef cutout / select 1-3 (600-900 lbs)	1	0	34	0.8942	0.0159
Imported boneless beef, 90 percent, frozen	0	0	46	0.6872	0.0456
Pork cutout composite	1	0	9	0.9136	0.0179
Loins, 1/4", trimmed vacuum-packed	0	1	19	0.9264	0.0180
Drumsticks	0	1	35	0.7664	0.0109
Wings, full cut	1	1	38	0.8920	0.0101

Table 4.2: Final Output of Hyperparameter Tuning Using SARIMA Model with External Features for All Meat Sub-Types

Boxed Beef Cutout / Select 1-3 (600-900 lbs):

By adding external features, the best parameters were $P = 1$, $Q = 0$, and Seasonal $m = 34$. This configuration improved the R2 score to 0.8942 and reduced the RMSE to 0.0159, indicating a significant enhancement in predictive accuracy.

Imported Boneless Beef, 90 Percent, Frozen:

The model's performance improved with an R2 score of 0.6872 and an RMSE of 0.0456 using the same parameters as before ($P = 0$, $Q = 0$, Seasonal $m = 46$). This indicates a moderate improvement in accuracy due to external features.

Pork Cutout Composite:

The external features led to a substantial improvement, with the best parameters ($P = 1$, $Q = 0$, Seasonal $m = 9$) yielding an R2 score of 0.9136 and an RMSE of

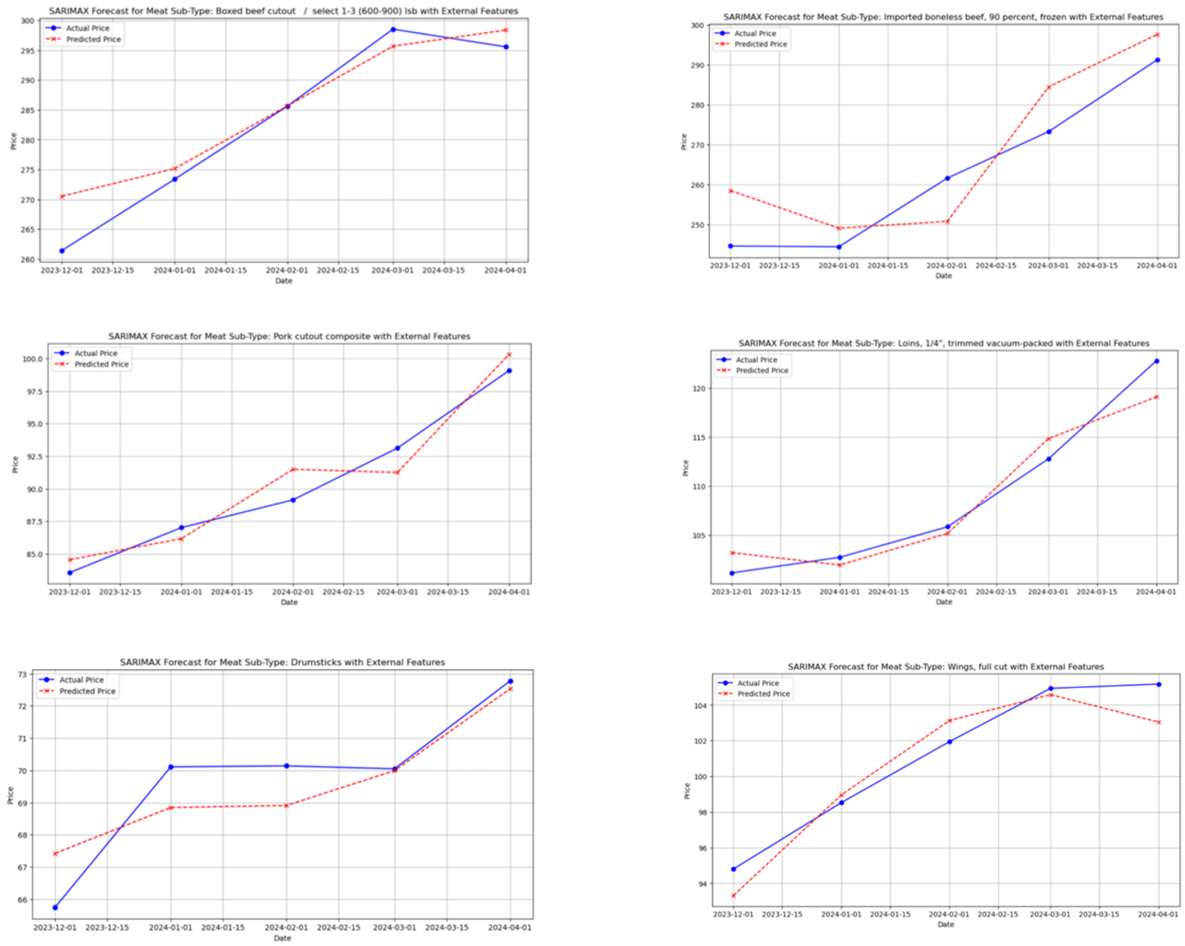

Figure 4.8: Comparison of actual versus predicted prices for different meat subtypes over the last 5 months using SARIMA model with External Features

0.0179. This significant improvement suggests that external factors play a crucial role in forecasting this meat type.

Loins, 1/4", Trimmed Vacuum-Packed:

The best model configuration changed to $P = 0$, $Q = 1$, and Seasonal $m = 19$, with a slightly reduced R2 score of 0.9264 and an RMSE of 0.0180. While the R2 score decreased marginally, the model remained highly accurate.

Drumsticks:

The best parameters with external features were $P = 0$, $Q = 1$, and Seasonal $m = 35$. However, the R2 score decreased slightly to 0.7664, and the RMSE increased marginally to 0.0109, suggesting potential overfitting or noise introduction.

Wings, Full Cut: With external features, the optimal parameters were $P = 1$, $Q = 1$, and Seasonal $m = 38$, resulting in a notable improvement in the R2 score to 0.8920 and a reduced RMSE of 0.0101.

4.1.10 Comparison of Models with and without External Features

The comparison of the models with and without external features revealed several insights:

Improvement in Performance with External Features:

For some meat types, such as Boxed Beef Cutout and Pork Cutout Composite, adding external features significantly improved the R2 score and reduced RMSE. For instance, the R2 score for Boxed Beef Cutout increased from 0.8370 to 0.8942, and the RMSE decreased from 0.0197 to 0.0159. Similarly, Pork Cutout Composite saw its R2 score improve from 0.8062 to 0.9136, with RMSE reducing from 0.0268 to 0.0179. This suggests that external factors like economic indicators and weather conditions play a crucial role in predicting price movements for these meat types.

The R2 Score for Boxed Beef Cutout increased from 0.8370 to 0.8942, and the RMSE decreased from 0.0197 to 0.0159.

Similarly, Pork Cutout Composite saw its R2 score improve from 0.8062 to 0.9136, with RMSE reducing from 0.0268 to 0.0179.

Mixed Results for Some Meat Types:

For other meat types, such as Imported Boneless Beef and Drumsticks, the improvement seen was moderate. Although there was a slight increase in R2 scores, the more significant result was the reduction in RMSE, indicating improved forecast accuracy. For example, the R2 score for Imported Boneless Beef improved from 0.6317 to 0.6872, with RMSE decreasing from 0.0495 to 0.0456.

Performance Degradation for Some Meat Types:

A few meat types, such as Drumsticks, slightly decreased the R2 score from 0.7889 to 0.7664, coupled with a minimal change in RMSE. This suggests that the inclusion of external features might have introduced some noise or led to overfitting in these cases, potentially complicating the model without adding substantial predictive power.

4.2 Random Forest Results

The Random Forest model forecasts the prices of various meat sub-types over a 5-month horizon. It includes the outcomes of hyperparameter tuning and the impact of incorporating external features.

4.2.1 Hyperparameter Tuning Process

The following hyperparameters were tuned to optimize the Random Forest model:

Number of Estimators (n_estimators): A Grid Search was tried on several trees in the forest on multiple combinations of randomly selected parameter values like [2, 4, 5, 8, 10, 15, 20, 30] to determine the optimal number.

Lag Size (i): This represents the number of past time steps used as input features. The range tested was from 1 to 50 lags to identify which provided the best forecast accuracy.

4.2.2 Model Forecasting with Hyperparameter Tuning Results:

The optimal parameters were selected for each meat subtype based on the highest R2 score and RMSE value.

Meat Type	Best Lags	Best n_estimators	Best R2 Score (Without External)	Best RMSE (Without External)
Boxed beef cutout (select 1-3, 600-900 lbs)	49	20	0.5266	0.0335
Imported boneless beef, 90 percent, frozen	38	2	0.5359	0.0556
Pork cutout composite	48	10	0.9136	0.0179
Loins (1/4", trimmed, vacuum-packed)	48	4	0.9751	0.0105
Drumsticks	49	8	0.7438	0.0114
Wings (full cut)	14	2	0.0010	0.0306

Table 4.3: Final Output of Hyperparameter Tuning Using Random Forest Model for All Meat Sub-Types

Boxed Beef Cutout (Select 1-3, 600-900 lbs):

The model uses 49 lags and 20 n_estimators, suggesting that it relies heavily on long historical records. The moderate R2 score (0.5266) and RMSE (0.0335)suggest that the model explains half of the variance there is still a need for improvement in getting predictive accuracy.

Imported Boneless Beef, 90 Percent, Frozen:

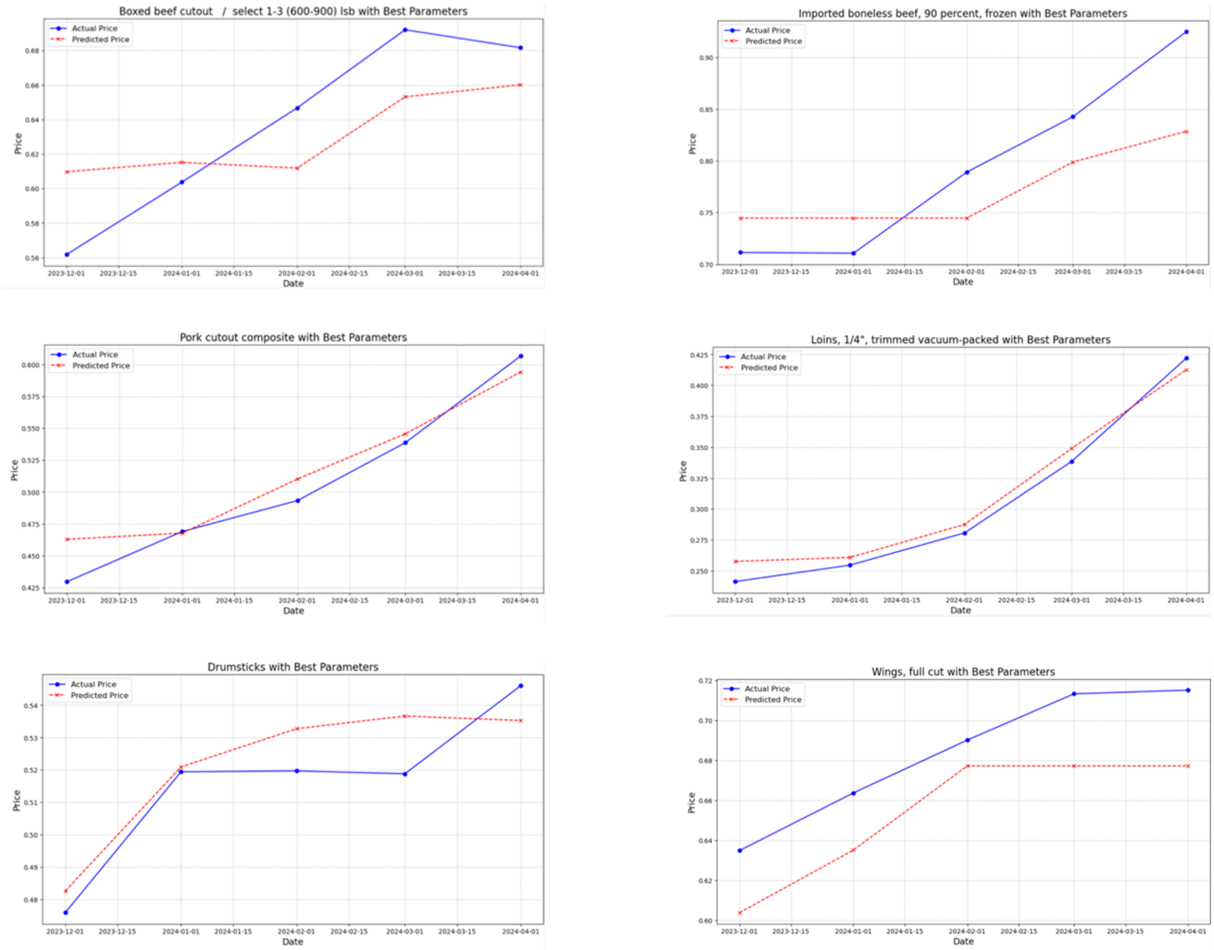

Figure 4.9: Comparison of actual versus predicted prices for different meat subtypes over the last 5 months Using Random Forest Model.

The model uses 38 lags and only 2 n_estimators, indicating that past prices impact this meat subtype more, but it uses a very low number of trees. The R2 score (0.5359) reflects moderate explanatory power, with the relatively high RMSE (0.0556) pointing to prediction errors.

Pork Cutout Composite:

With 48 lags and 10 n_estimators, the model captures a long-term price trend effectively, explaining the variance R2 score (0.9136). The low RMSE (0.0179) indicates that the model achieves high prediction accuracy.

Loins (1/4", Trimmed, Vacuum-Packed):

This model utilizes 48 lags and 4 n_estimators, capturing the price pattern effectively over a long history with a small number of trees. The high R2 score (0.9751) and very low RMSE (0.0105) show excellent model performance, suggesting the chosen parameters are well-suited to this meat type.

Drumsticks:

The model uses 49 lags and 8 n_estimators, indicating a reliance on long price history and moderate model complexity to achieve a good R2 score (0.7438). The low RMSE (0.0114) suggests reasonable predictive accuracy.

Wings (Full Cut):

The model uses only 14 lags and 2 n_estimators, suggesting a limited dependency on historical data. The extremely low R2 score (0.0010) indicates that the model fails to capture the underlying price dynamics, resulting in poor performance.

4.2.3 Incorporating External Features Results

To make the Random Forest model performance more powerful, top external variables were incorporated as shown in the table 3.9 and they are:

Weather Conditions: Average Temperature

Economic Indicators: Consumer Price Index (CPI),

Commodity Prices: sorghum prices.

Meat Type	Best Lags	Best n_estimators	Best R2 Score (With External)	Best RMSE (With External)
Boxed beef cutout (select 1-3, 600-900 lbs)	17	4	0.5223	0.0337
Imported boneless beef, 90 percent, frozen	30	4	0.5761	0.0531
Pork cutout composite	6	8	0.8869	0.0205
Loins (1/4", trimmed, vacuum-packed)	6	2	0.8244	0.0278
Drumsticks	7	4	0.7264	0.0118
Wings (full cut)	3	5	0.6797	0.0173

Table 4.4: Final Output of Hyperparameter Tuning Using Random Forest Model with External Feature for All Meat Sub-Types.

Boxed Beef Cutout (Select 1-3, 600-900 lbs):

With external features, the model reduces historical data to 17 lags and 4 n_estimators. suggesting that external factors provide valuable predictive information. However, the R2 score (0.5223) is slightly lower, indicating that external features do not significantly enhance prediction accuracy.

Imported Boneless Beef, 90 Percent, Frozen:

The model uses 30 lags and 4 n_estimators, showing moderate dependence on both past data and external factors. The R2 score improves to 0.5761, and the RMSE decreases to 0.0531, suggesting that external variables help explain more variance and reduce errors.

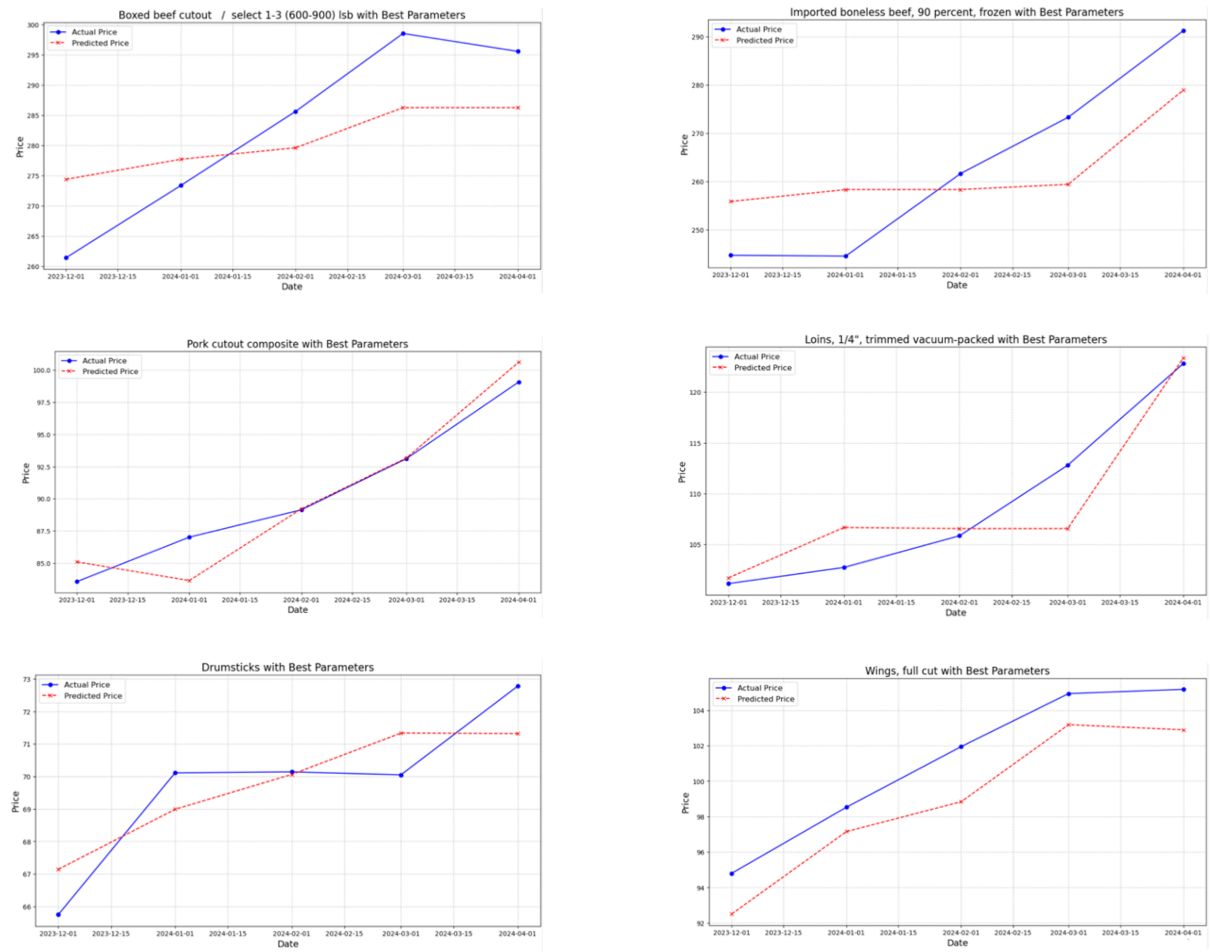

Figure 4.10: Comparison of actual versus predicted prices for different meat Sub types over the last 5 months Using Random Forest Model With External Feature.

Pork Cutout Composite:

When external features are added, the model reduces the number of lags significantly to 6 but increases the number of n_estimators to 8. The R2 score decreases slightly to 0.8869, and RMSE increases to 0.0205, indicating a minor decline in performance, potentially due to less relevant external variables.

Loins (1/4", Trimmed, Vacuum-Packed):

The model reduces both lags to 6 and n_estimators to 2, reflecting a simpler model with external features. The R2 score drops substantially to 0.8244, and the RMSE increases to 0.0278, suggesting that the external variables might not be as relevant for this meat type.

Drumsticks:

The model reduces the number of lags to 7 and n_estimators to 4, indicating that external features are moderately helpful. The R2 score decreases slightly to 0.7264, and

RMSE increases to 0.0118, reflecting minimal change in performance.

Wings (Full Cut):

With external features, the model reduces lags further to 3 and increases n_estimators to 5. The R2 score improves dramatically to 0.6797, and RMSE decreases to 0.0173, indicating that external features significantly enhance predictive power for this meat type.

4.2.4 Comparison Analysis of Models With and Without External Features

With External Features:

For Imported boneless beef, 90 percent, frozen and Wings, full cut the model saw a noticeable increase in R2 Score and a reduction in RMSE, demonstrating that external factors like weather, economic conditions, and feed prices have a significant influence on these subtypes.

Without External Features:

The highest R2 Score (0.9751) was observed for Loins, 1/4", trimmed, vacuum-packed without external features, suggesting that the historical price data itself was sufficient to capture the trend for this meat type.

4.3 Gradient Boosting Regression Results

The Gradient Boosting model forecasts the prices of various meat sub-types over a 5-month horizon. It includes the outcomes of hyperparameter tuning and the impact of incorporating external features

4.3.1 Hyperparameter Tuning Process

The following hyperparameters were tuned to optimize the Gradient Boosting model:

Lag Size (i): This represents the number of past time steps used as input features. The range tested was from 1 to 50 lags to identify which provided the best forecast accuracy.

Number of Trees (n_estimators): The Grid search was tried on several trees that need to be added to the model. More trees lead to better performance and also lead to overfitting. So randomly selected parameter values like [2, 4, 5, 8, 10, 15, 20] to get optimal values.

Max Depth (max_depth): The Grid search was tried on maximum depth to control the complexity of the base learners. A deeper tree can capture more complex patterns but is more likely to overfit the training data. So randomly selected parameter values [5, 10, 15, 20, 25, 30, 35, 40] were run on multiple values to get an optimal number that gives the best parameters.

Learning Rate (learning rate): The Grid search was tried on each new tree. It is a small number of positive numbers. A lower learning rate requires more iteration but can lead to better generalization.

4.3.2 Model Forecasting with Hyperparameter Tuning Results:

The optimal parameters were selected for each meat subtype based on the highest R2 score and RMSE value.

Meat Type	Best Lags	Best n_estimators	Max Depth	Learning Rate	Best R2 Score (With External)	Best RMSE (With External)
Boxed beef cutout (select 1-3, 600-900 lbs)	32	10	7	0.5	0.6066	0.0306
Imported boneless beef, 90 percent, frozen	42	10	2	0.6	0.1576	0.0749
Pork cutout composite	13	20	10	0.1	0.9476	0.0139
Loins (1/4", trimmed, vacuum-packed)	19	10	10	0.7	0.9881	0.0072
Drumsticks	13	2	10	0.6	0.8530	0.0086
Wings (full cut)	-	-	-	-	0	0

Table 4.5: Final Output of Hyperparameter Tuning Using Gradient Boosting Model for All Meat Sub-Types

Boxed Beef Cutout / Select 1-3 (600-900) lbs

The model used hyperparameters for this meat subtype like 32 lags, 10 estimators, a maximum depth of 5, and a learning rate of 0.7, achieving an R2 score of (0.6067) and an RMSE score of (0.0306). This suggests a moderate fit, as the model predicted the target variable reasonably well without considering any external parameters.

Imported Boneless Beef, 90% Frozen

This meat subtype used parameters like 42 lags, 2 estimators, a maximum depth of 10, and a learning rate of 0.6, producing a low R2 score of (0.1576)and a high RMSE score of (0.0749). This indicates very poor predictive power and the model could not

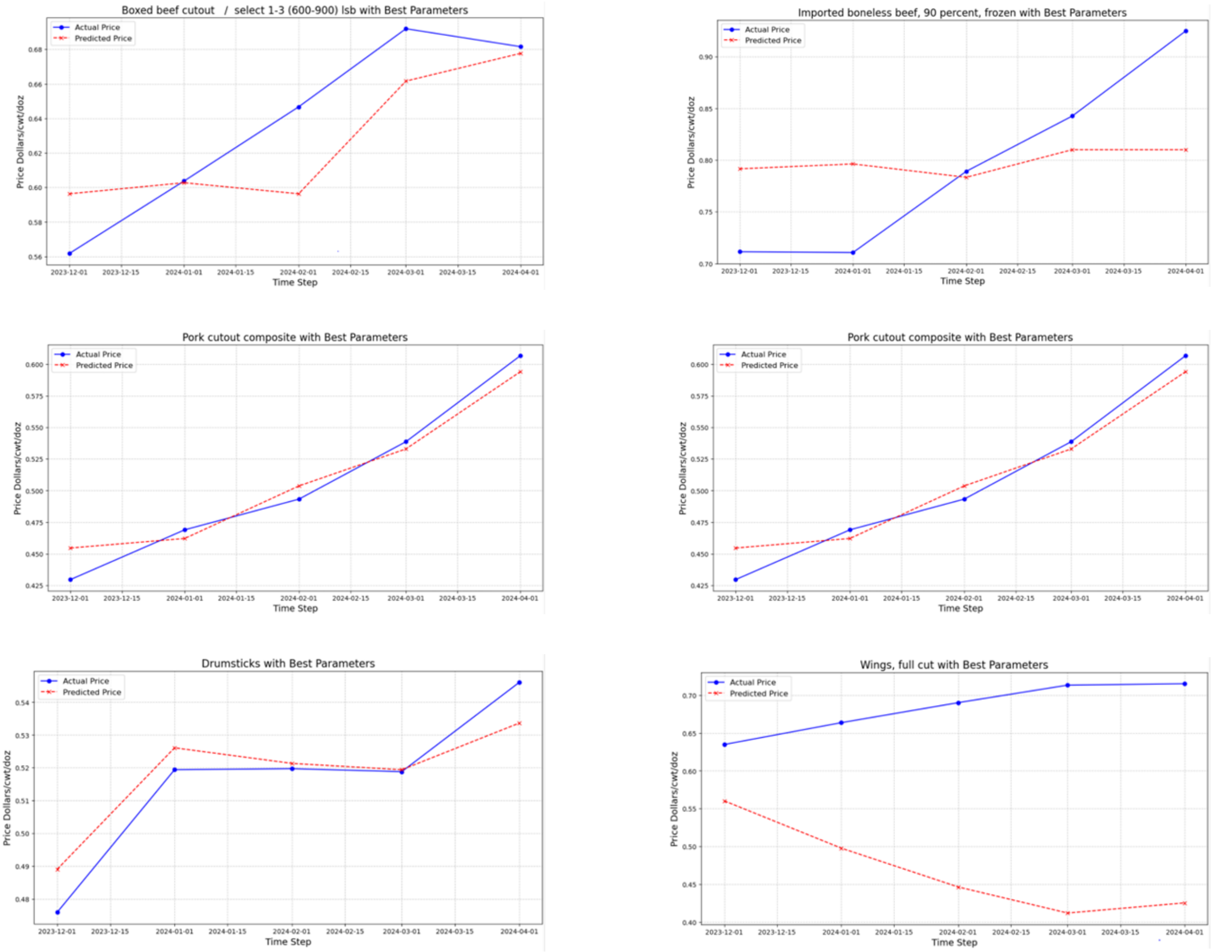

Figure 4.11: Comparison of actual versus predicted prices for different meat types over the last 5 months using Gradient Boosting Model

capture any patterns effectively.

Pork Cutout Composite

Initially, the model used 13 lags, 20 estimators, a maximum depth of 10, and a learning rate of 0.1, resulting in a high R2 score of 0.9477 and a low RMSE of 0.0139. This indicates a strong predictive model, suggesting that the internal data (such as lagged values and past prices) were sufficient for accurate prediction without the need for external factors.

Loins, 1/4", Trimmed Vacuum-Packed

The model, which employed 19 lags, 10 estimators, a maximum depth of 10, and a learning rate of 0.7, performed excellently with an R2 score of (0.9881) and an RMSE score of (0.0072). This suggests that the model was highly accurate and stable predictable of internal patterns.

Drumsticks

The model with 13 lags, 2 estimators, a maximum depth of 10, and a learning rate of 0.6, achieved a high R2 score of 0.9436, indicating good predictive accuracy. However historical data make its predictions accurately.

Wings, Cut Out / Full Cut

By analyzing for 5 months, the model failed to predict accurately, resulting in an R2 score of 0 and an RMSE of 0. This suggests that the parameters did not contribute effectively and may have introduced significant noise or conflicting signals, leading to model failure.

4.3.3 Incorporating External Features Results:

To make the Gradient Boosting model performance more powerful, top external variables were incorporated as shown in the table 3.9 and they are:

Weather Conditions: Average Temperature

Economic Indicators: Consumer Price Index (CPI),

Commodity Prices: sorghum prices.

Meat Type	Best Lags	Best n_estimators	Max Depth	Learning Rate	Best R2 Score (With External)	Best RMSE (With External)
Boxed beef cutout (select 1-3, 600-900 lbs)	5	20	5	0.9	0.6226	0.0299
Imported boneless beef, 90 percent, frozen	1	4	20	0.7	0.7223	0.0429
Pork cutout composite	14	10	5	0.9	0.9556	0.0128
Loins (1/4", trimmed, vacuum-packed)	2	8	10	0.7	0.9293	0.0017
Drumsticks	4	2	10	0.6	0.7981	0.0101
Wings (full cut)	-	-	-	-	0	0

Table 4.6: Final Output of Hyperparameter Tuning Using Gradient Boosting with External Feature for All Meat Sub-Types

Boxed Beef Cutout / Select 1-3 (600-900) lbs

The model performance with parameters, with 5 lags, 20 estimators, a maximum depth of 5, and a higher learning rate of 0.9, the performance slightly improved, resulting in an R2 score of (0.6226) and a reduced RMSE of (0.0299). The addition of external parameters, and seasonal trend considerations, may have helped refine the model's accuracy over this specific period.

Imported Boneless Beef, 90% Frozen

In this meat subtype with parameters 1 lag, 4 estimators, a maximum depth of 20, and a learning rate of 0.7, the R2 score improved significantly to (0.7223), and the RMSE was reduced to (0.0430). There are a lot of changes seen after adding external features

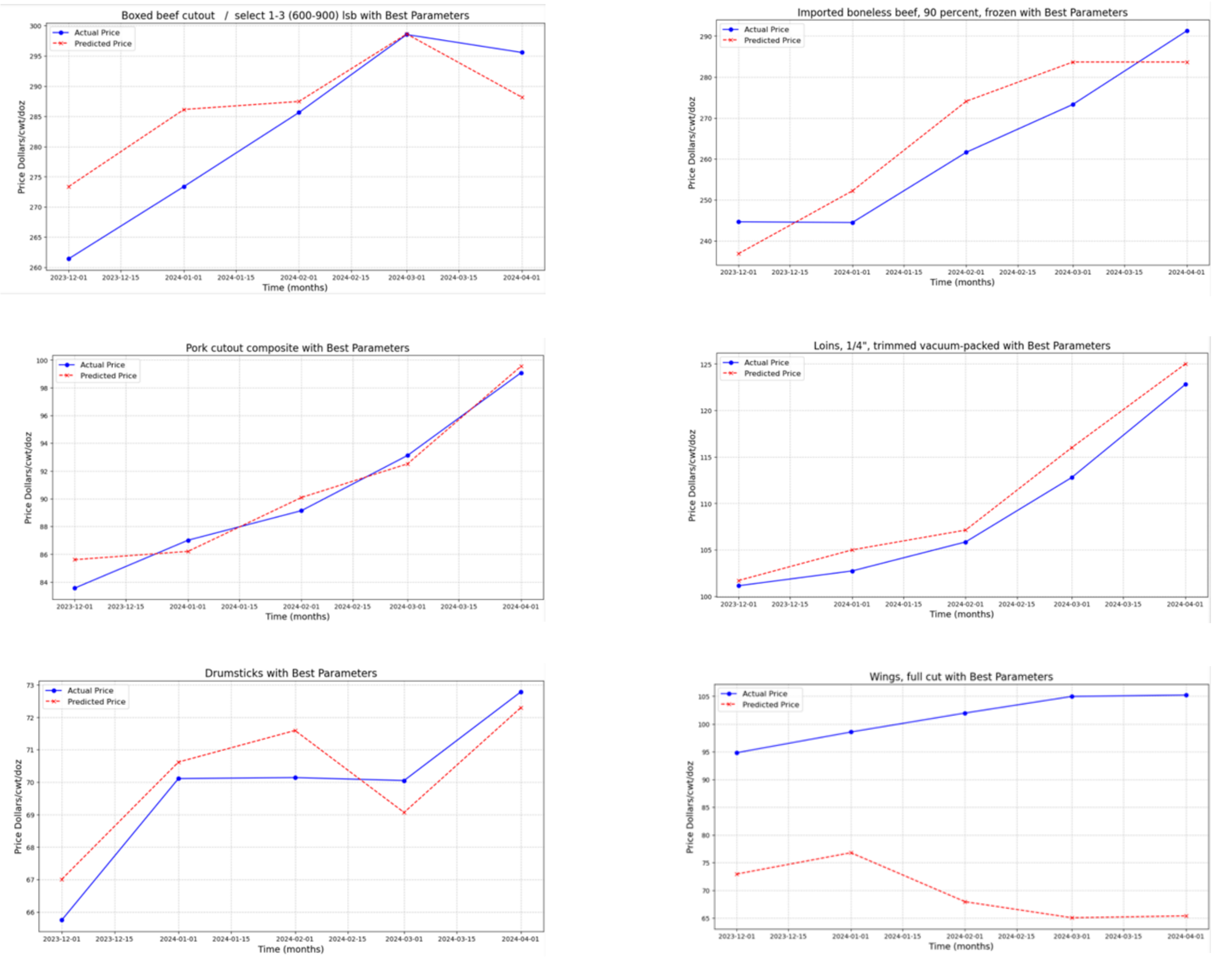

Figure 4.12: Comparison of actual versus predicted prices for different meat types over the last 5 months Using Gradient Boosting Model with External Features

like seasonal trends, economic indicators, and market conditions that improved the model's accuracy.

Pork Cutout Composite

For this meat type using parameters like 14 lags, 10 estimators, a maximum depth of 5, and a learning rate of 0.9, further improved the R2 score to (0.9556) and decreased the RMSE to ()0.0128. This shows that with an already strong model, the addition of external parameters like demand forecasts or economic indicators provided marginal gains in prediction accuracy.

Loins, 1/4", Trimmed Vacuum-Packed

When optimized for the 5-month prediction, using 2 lags, 8 estimators, a maximum depth of 10, and a learning rate of 0.7, the model maintained a high R2 score of (0.9294) and an RMSE of (0.0176). The slight decrease in performance could imply that external parameters, such as packaging or processing changes, had minimal impact or possibly added noise to the predictions.

Drumsticks

The hyperparameter used for this meat subtype is 4 lags, 2 estimators, a maximum depth of 10, and a learning rate of 0.6, the R2 score dropped to (0.7982), though the RMSE remained low at (0.0101). By introducing external features the accuracy of the model is reduced with very less variability.

Wings, Cut Out / Full Cut

This meat subtype failed to predict accurately resulting in R2 0 and RMSE 0. This suggests that after including external features also the model failed to incorporate more noise data.

4.3.4 Comparison Analysis of Models with and Without External Features

With External Features:

For Imported Boneless Beef, 90 Percent, Frozen and Boxed Beef Cutout / Select 1-3 (600-900) lbs, there was a marked improvement in model performance. The R2 Score for Imported Boneless Beef, 90 Percent, Frozen increased from 0.1576 to 0.7223, and the RMSE dropped from 0.0749 to 0.0429, highlighting the significant impact of external variables such as economic conditions, supply chain factors, and market demand. Similarly, the slight improvement for Boxed Beef Cutout indicates that external features, while not critical, add value in certain contexts.

Without External Features:

The highest R2 Score of 0.9881 was observed for Loins, suggesting that the historical price data alone effectively captures the trend for this meat type. Additionally, models for Drumstick and Wings, Full Cut performed better without external features, with Drumstick achieving an R2 Score of 0.9436 and Wings, Full Cut maintaining a reasonable R2 Score of 0.7433. This indicates that, for these types, adding external data does not necessarily enhance performance and might introduce irrelevant complexity.

4.4 Support Vector Regression (SVR) Results

The Support Vector Regression model forecasts the prices of various meat sub-types over a 5-month horizon. It includes the outcomes of hyperparameter tuning and the impact of incorporating external features.

4.4.1 Hyperparameter Tuning Process

The following hyperparameters were tuned to optimize the Support Vector Regression model:

Lag Size (i): This represents the number of past time steps used as input features. The range tested was from 1 to 50 lags to identify which provided the best forecast accuracy.

C (Regularization Parameter): This Parameter C takes controls between achieving low training error and low testing error which helps to avoid the overfitting to the model.

Epsilon : The parameter defines the margin of tolerance where no penalty is given for errors. This is the width of the "tube" around the regression line within which errors are ignored.

4.4.2 Model Forecasting with Hyperparameter Tuning Results:

The optimal parameters were selected for each meat subtype based on the highest R2 score and RMSE value.

Meat Type	Best Lags	Best C	Best Epsilon	Best R2 Score (With External)	Best RMSE (With External)
Boxed beef cutout (select 1-3, 600-900 lbs)	24	200	0.0001	0.8896	0.0162
Imported boneless beef, 90 percent, frozen	6	1000	0.005	0.8416	0.0325
Pork cutout composite	5	200	0.005	0.9328	0.0158
Loins (1/4", trimmed, vacuum-packed)	10	100	0.0001	0.9760	0.0103
Drumsticks	13	0,5	0.1	0.8208	0.0095
Wings (full cut)	-	-	-	0	0

Table 4.7: Final Output of Hyperparameter Tuning Using SVR Model for All Meat Sub-Types

Boxed Beef Cutout / Select 1-3 (600-900) lbs

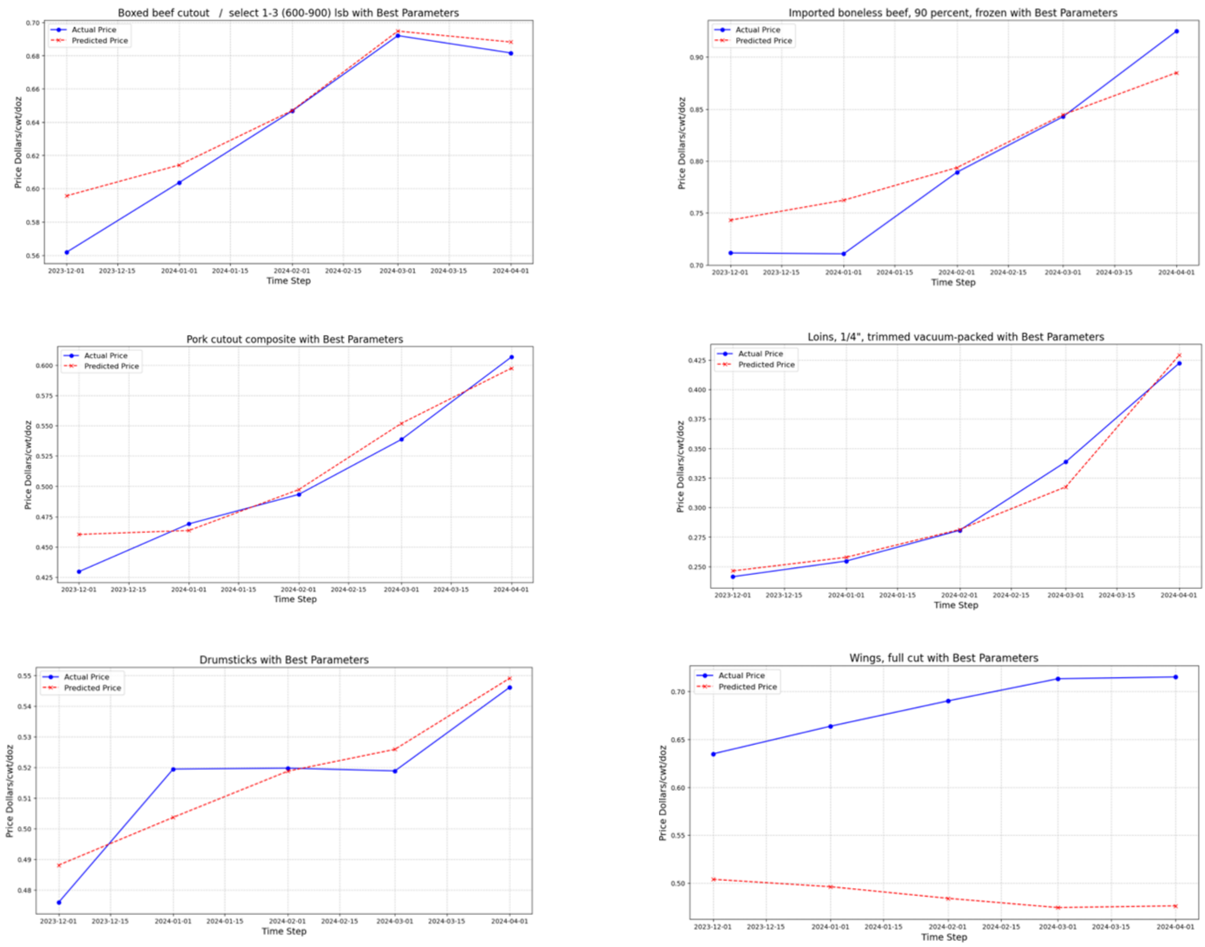

Figure 4.13: Comparison of actual versus predicted prices for different meat types over the last 5 months Using SVR Model.

The parameters used for this meat subtype are lags 24, C 200, and Epsilon 0.0001 with an R2 score of(0.8896) and low RMSE (0.0162), indicating strong predictive performance using only historical data.

Imported Boneless Beef, 90 Percent, Frozen

The meat sub-type uses parameters of lags 6, C 1000, and Epsilon 0.005 with an R2 score of (0.8416) and the RMSE (0.0325). This indicates that the model was able to capture price trends with historical data alone.

Pork Cutout Composite

The model performed well with an R2 score of (0.9328) and low RMSE(0.0158), this suggests that with a very low lags value(5) and Epsilon value (0.005), the model was able to capture the historical price trend easily.

Loins, 1/4", Trimmed Vacuum-Packed

The model achieved the highest R2 score of (0.9760) and the lowest RMSE of(0.0103).

The parameters like lags 13, C 100, and epsilon of 0.0001 have suggested that the price trend for this subtype was highly predictable using historical data.

Drumsticks

The model performance was reasonably good with an R2 Score of 0.8208 and RMSE of 0.0095.

The model performance for this meat subtype was reasonably good with an R2 score of (0.8208) and RMSE of (0.0095).

Wings, Full Cut

The model was unable to provide meaningful predictions (R2 Score = 0, RMSE = 0).

4.4.3 Incorporating External Features Results

To make the SVR model performance more powerful, top external variables were incorporated as shown in table 3.9 and they are:

Weather Conditions: Average Temperature

Economic Indicators: Consumer Price Index (CPI),

Commodity Prices: sorghum prices.

Meat Type	Best Lags	Best C	Best Epsilon	Best R2 Score (With External)	Best RMSE (With External)
Boxed beef cutout (select 1-3, 600-900 lbs)	11	500	0.0001	0.6042	0.0307
Imported boneless beef, 90 percent, frozen	1	10	0.05	0.7059	0.0442
Pork cutout composite	4	1	0.5	0.8918	0.0200
Loins (1/4", trimmed, vacuum-packed)	8	50	0.0001	0.8060	0.0292
Drumsticks	14	0.5	0.1	0.7971	0.0101
Wings (full cut)	2	5	0.0001	0.9187	0.0087

Table 4.8: Final Output of Hyperparameter Tuning Using SVR Model with External Feature for all Meat Sub-Types

Boxed Beef Cutout / Select 1-3 (600-900) lbs

The performance metrics slightly decreased the R2 Score (0.6042) and RMSE (0.0307). The addition of external factors did not enhance the model's ability to predict price changes, suggesting that the internal patterns in historical data were already sufficient.

Imported Boneless Beef, 90 Percent, Frozen

A slight decrease in R2 Score (0.7059) and an increase in RMSE (0.0442) were observed. This indicates that the external features did not significantly contribute to improving the model's performance for this subtype.

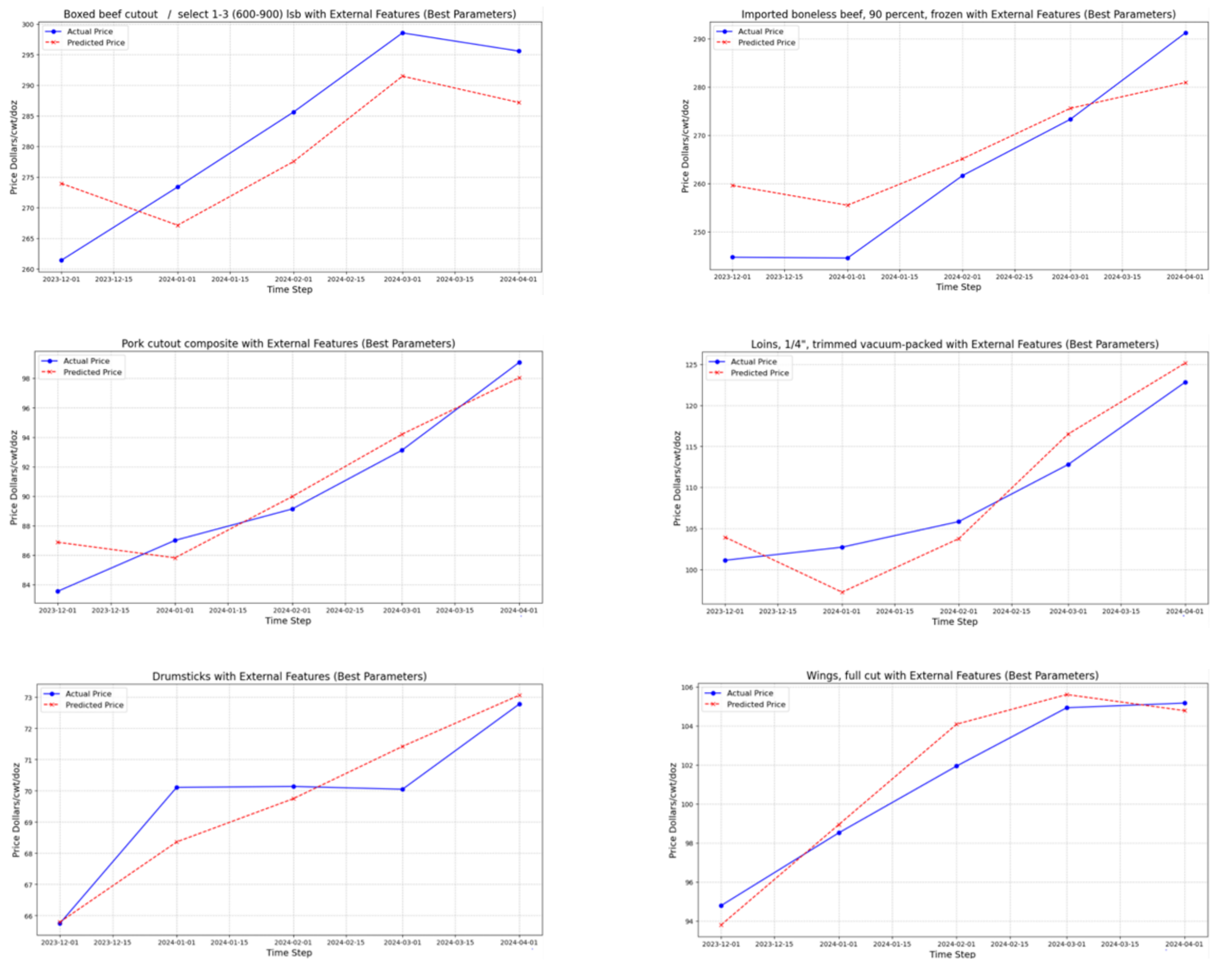

Figure 4.14: Comparison of actual versus predicted prices for different meat types over the last 5 Months Using SVR with External Features.

Pork Cutout Composite The model performance for this meat subtype with parameters like lags 4, C 1, and Epsilon 0.5 with R2 score (0.8918) and RMSE (0.0200) shows that external feature did not give good predictive accuracy to this meat subtype.

Loins, 1/4", Trimmed Vacuum-Packed

The hyperparameters used for this meat subtypes are lags 8, C 50, and Epsilon 0.0001 with R2 score (0.8060) and RMSE (0.0292), adding external features reduced the model accuracy, and without external features, the model performed well.

5. Drumsticks

A slight decrease in the R2 Score to (0.7971) was observed, but the RMSE remained nearly the same (0.0101). This suggests that the external features neither significantly helped nor hindered the prediction accuracy.

6. Wings, Full Cut

The model achieved a significant performance improvement with an R2 score of

0.9187 and the lowest RMSE (0.0087). The external factors like temperature and economic indicators greatly enhanced the model's predictive capability for this subtype.

4.4.4 Comparison Analysis of Models with and Without External Features

With External Features: Wings, full cut saw significant improvement when external features were included. This indicates that external factors like weather, economic conditions, and feed prices are crucial predictors for this subtype.

Without External Features: Boxed beef cutout / select 1-3 (600-900) lbs, Imported boneless beef, 90 percent, frozen, Pork cutout composite, and Loins, 1/4, trimmed vacuum-packed showed better performance without external features. The historical price data alone was sufficient to capture their price trends accurately.

Mixed Results: For Drumsticks the inclusion of external features did not significantly impact model performance, suggesting that while external features may add some value, their contribution is not substantial enough to affect the overall prediction accuracy.

4.5 Long Short Term Memory (LSTM) Results

The LSTM model forecasts the prices of various meat sub-types over a 5-month horizon. It includes the outcomes of hyperparameter tuning and the impact of incorporating external features

4.5.1 Hyperparameter Tuning Process

The following hyperparameters were tuned to optimize the LSTM model:

Lag Size (i): This represents the number of past time steps used as input features. The range tested was from 21 to 50 lags as the LSTM model has long-term dependencies that identify and provide the best forecast accuracy.

Units: These parameters tell more about the number of units (neurons) in each LSTM layer. The values are tested from the range (50,100,150) units of memory cells that can maintain information in memory for an extended period.

Learning Rate: This parameter helps to control the step size after each iteration and move towards the minimum loss function. The values are taken as (0.001,0.005,0.01) randomly and also tell model weights concerning to loss function

Batch Size: Batch size is the number of training used in one iteration before model weights get updated. The values used for testing are (16,32,64) are randomly selected and were able to learn from performing well using this batch size.

4.5.2 Model Forecasting with Hyperparameter Tuning Results:

The optimal parameters were selected for each meat subtype based on the highest R2 score and RMSE value.

Meat Type	Best Lags	Best Units	Best Learning Rate	Best Batch Size	Best R2 Score (With External)	Best RMSE (With External)
Boxed Beef Cutout	43	150	0.001	32	0.6990	0.04248
Loins, 1/4" Trimmed Vacuum-Packed	38	100	0.01	16	0.8376	0.01117
Drumsticks	21	150	0.001	64	0.8017	0.01571
Imported Boneless Beef, 90 Percent, Frozen	37	50	0.001	16	0.6892	0.04548
Pork Cutout Composite	41	50	0.005	16	0.8549	0.02320
Wing, Cutout	32	50	0.001	16	0.9467	0.0070

Table 4.9: Final Output of Hyperparameter Tuning Using LSTM Model for All Meat Sub-Types

Boxed Beef Cutout / Select 1-3 (600-900) lbs

The Boxed Beef Cutout model performs with a moderate R2 score (0.6990) and RMSE (0.4248). The choice of 43 lags and 150 units suggests the model uses a substantial amount of historical data and a moderate batch size to balance training efficiency and prediction accuracy.

The Boxed beef Cutout with the best hyperparameter values like 43 lags,150 units, learning rate of 0.001, and batch size of 32 has a moderate R2 score of (0.6990) and RMSE (0.4248). The use of 43 lags and 150 units shows that the model uses more historical data and moderate training efficiency for predicting the values accurately.

Imported Boneless Beef, 90 Percent, Frozen

This model shows moderate performance with a low R2 score of (0.6892) and RMSE (0.04548). The use of 37 lags and a very small batch size indicated that the model has limited complexity and works well for a very small dataset.

Pork Cutout Composite

The pork composite with parameters like 41 lags,50 units, a learning rate of 0.001, and batch size 16 has a high R2 score (0.8549) and RMSE (0.02320). This indicates that

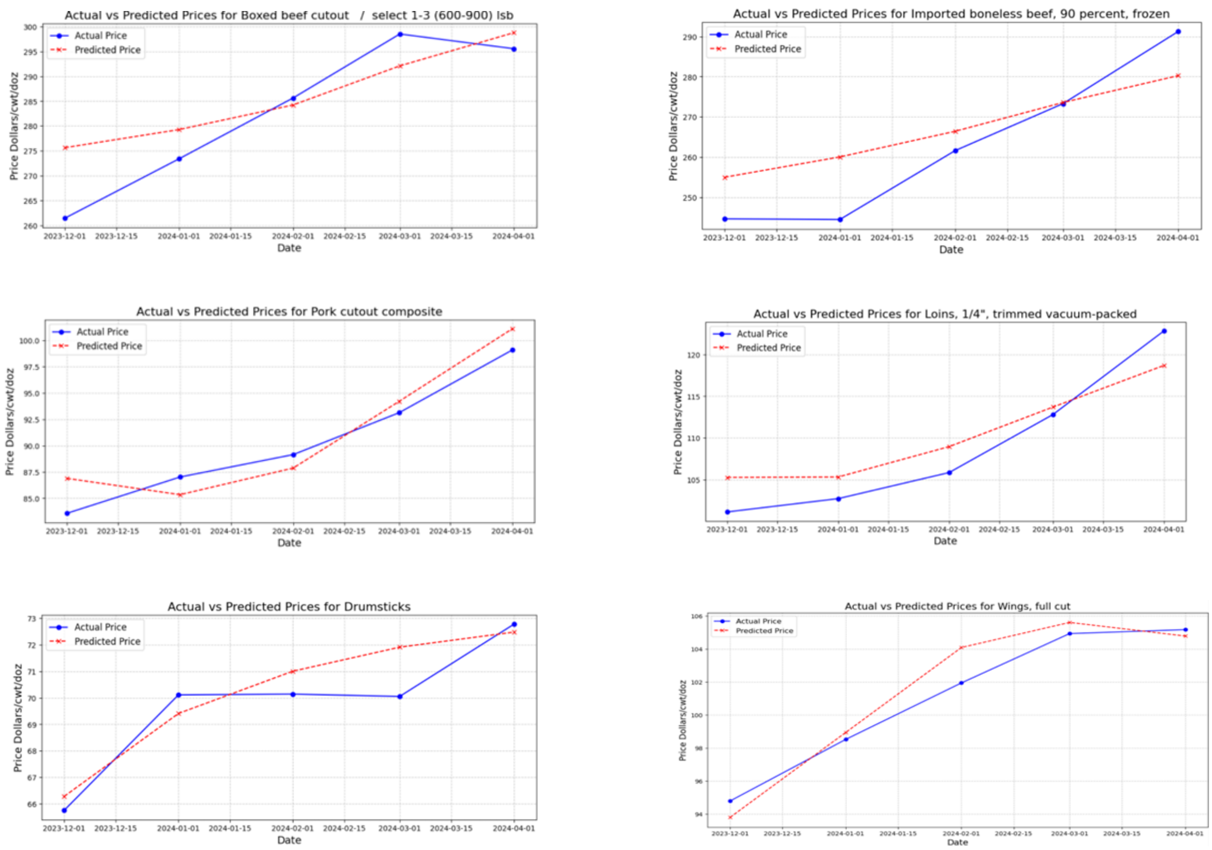

Figure 4.15: Comparison of actual versus predicted prices for different meat subtypes over the last 5 months Using LSTM Model.

this model has an effective balance of historical data and model training.

. Loins, 1/4", Trimmed Vacuum-Packed

The best parameters for this meat type are 38 lags,100 units,0.01 learning rate, and 16 batch sizes with a high R2 score of (0.8376) and low RMSE(0.01117) This suggests that it works efficiently with a moderate number of lags and smaller batch sizes for good predictive performance.

Drumsticks

The Drumsticks subtype has achieved a strong R2 score (0.8017) and low RMSE (0.01571). The smaller lags (21) and larger batch size (64) tell only fewer historical data points were captured but more training on larger batches

Wings, Full Cut

This model performs the best with an excellent R2 score of (0.9467) and the lowest RMSE value (0.0070). The model uses 32 lags and a small batch size (16) by effectively capturing the trends and minimal error in the model for this meat type.

4.5.3 Incorporating External Features Results:

To make the LSTM model performance more powerful, top external variables were incorporated as shown in table 3.9 and they are:

Weather Conditions: Average Temperature

Economic Indicators: Consumer Price Index (CPI),

Commodity Prices: sorghum prices.

Meat Type	Best Lags	Best Units	Best Learning Rate	Best Batch Size	Best R2 Score (With External)	Best RMSE (With External)
Boxed Beef Cutout	30	150	0.001	32	0.8962	0.01571
Loins, 1/4" Trimmed Vacuum-Packed	22	50	0.01	64	0.9440	0.01569
Drumsticks	44	150	0.005	16	0.5808	0.01459
Imported Boneless Beef, 90 Percent, Frozen	49	150	0.001	16	0.6833	0.04591
Pork Cutout Composite	43	50	0.01	16	0.9615	0.01459
Wing, Cutout	44	100	0.01	32	0.6652	0.04234

Table 4.10: Final Output of Hyperparameter Tuning Using LSTM model with External Feature for All Meat Sub-Types

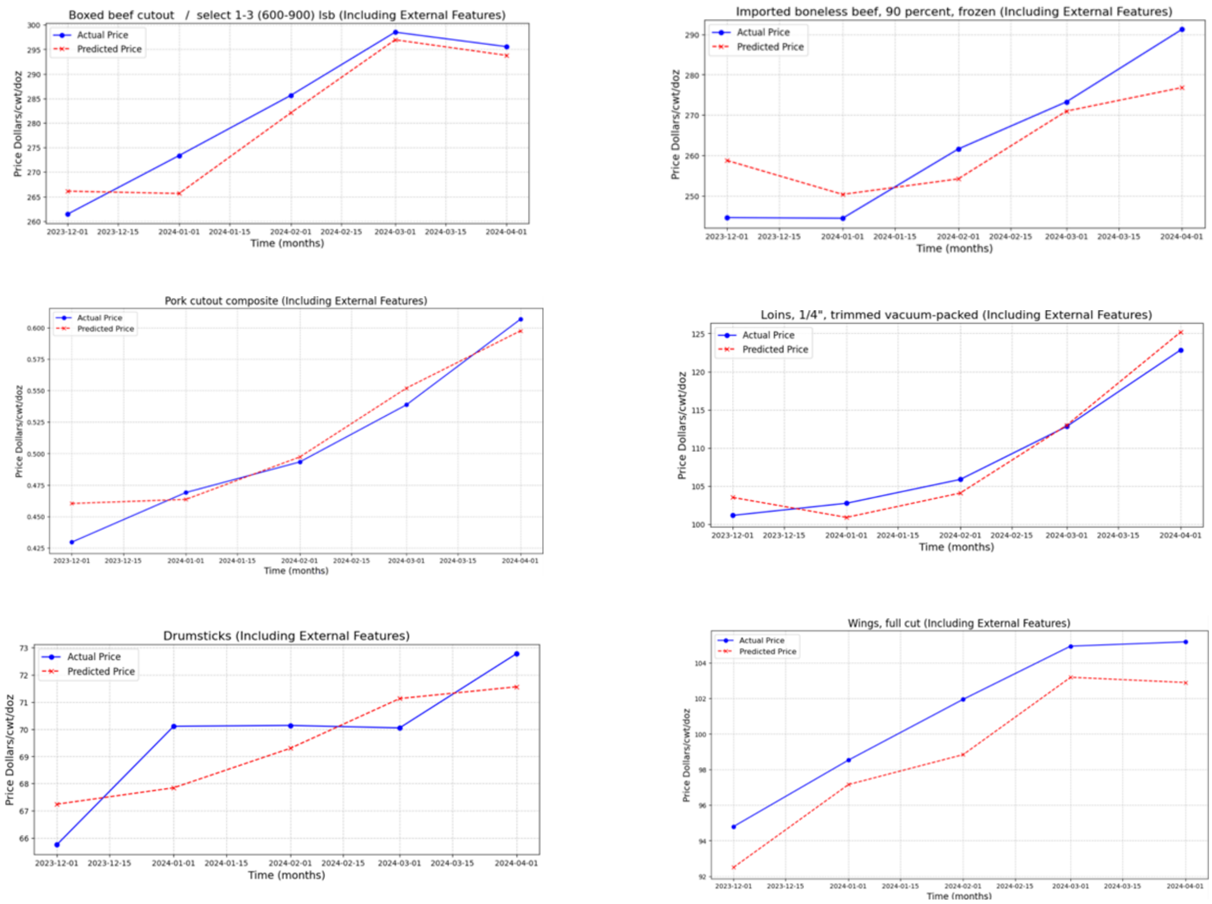

Figure 4.16: Comparison of actual versus predicted prices for different meat subtypes over the last 5 months Using LSTM Model With External Features.

Boxed Beef Cutout / Select 1-3 (600-900) lbs

Adding external features has improved the R2 score (0.8962) and reduced RMSE (0.01571). The decrease in I lags to 30 helps external features capture only relevant information more effectively.

Imported Boneless Beef, 90 Percent, Frozen

The R2 score has been reduced slightly (0.6833) and increased in RMSE (0.04591). This tells higher lags (49) suggesting external features may introduce complexity without improving model accuracy,

Pork Cutout Composite

The pork cutout with the best parameters of 43 lags,50 units,0.01 learning rate, and 32 batch size has more benefits from external features showing an R2 score (0.9615) and low RMSE (0.01459). This tells the batch size and lag values have enhanced the model's performance effectively.

Loins, 1/4", Trimmed Vacuum-Packed

The loins meat subtype with the best parameters are 22 lags,50 units, learning rate 0.01, and batch size 64 with a good R2 score of (0.9440) and low RMSE (0.01569). This model suggests that reduced lags and decreased units tell external features and also help in predicting good predictive values.

Drumsticks

The model shows a decreased R2 score (0.5808) with external features, despite having a higher number of lags. The higher learning rate unchanged and batch size suggest the model might be overfitting.

Wings, Full Cut

The wings cutout performed well without the external feature of capturing the highest R2 score (0.9467) and lowest RMSE value (0.0070). After adding the external feature there is a complete decrease in the R2 score of (0.6652) and an increase in RMSE (0.04234). This suggests that external features are not performing well for this meat dub type.

4.5.4 Comparison Analysis with and Without External Features

With External Features:

The adding external features improved the performance of the Wing, Cutout model. This indicates that for some sub-types, external factors provide critical insights that enhance predictive accuracy.

Without External Features:

Models such as Boxed Beef Cutout, Imported Boneless Beef, 90 Percent, Frozen, Pork Cutout Composite, and Loins, 1/4" Trimmed Vacuum-Packed performed better without external features. This suggests that for these sub-types, historical price data alone is sufficient for accurate predictions.

Mixed Results:

For Drumsticks, external features are very limited and also add some value, but they are not crucial for improving the model's performance significantly.

CHAPTER 5

Conclusions

In this research, the evaluation and performance of many forecasting models like SARIMA, Random Forest, Gradient Boosting, SVR, and LSTM both with and without external features to predict the prices of various meat sub-types are seen. The main objective was to identify which model and features provide the most accurate forecast price for each of the meat sub-types.

5.1 Key Findings

Our results reveal several important patterns regarding model performance across different meat sub-types and are summarised in Table 5.1.

Meat Sub type (Without External)	**SARIMA R^2 (Without External)**	**SARIMA RMSE (Without External)**	**RF R^2 (Without External)**	**RF RMSE (Without External)**	**GB R^2 (Without External)**	**GB RMSE (Without External)**	**SVR R^2 (Without External)**	**SVR RMSE (Without External)**	**LSTM R^2 (Without External)**	**LSTM RMSE (Without External)**
Boxed Beef Cutout	0.837	0.0197	0.5266	0.0335	0.6066	0.0306	0.8896	0.0162	0.699	0.0425
Imported Boneless Beef, 90% Frozen	0.6317	0.0495	0.5359	0.0556	0.1576	0.0749	0.8416	0.0325	0.6892	0.0455
Pork Cutout Composite	0.8062	0.0268	0.9136	0.0179	0.9476	0.0139	0.9328	0.0158	0.8549	0.0232
Loins, 1/4" Trimmed Vacuum-Packed	0.9771	0.01	0.9751	0.0105	0.9881	0.0072	0.976	0.0103	0.8376	0.0112
Drumsticks	0.7889	0.0104	0.7438	0.0114	0.853	0.0086	0.8208	0.0095	0.8017	0.0157
Wings, Full Cut	0.8084	0.0134	0.001	0.0306	0	0	0	0	0.9467	0.007

Table 5.1: Model Performance Comparison for Different Meat Sub types (Without External Features).

Boxed Beef Cutout: The LSTM model with external features with R2 score (0.8962) and RMSE (0.0157), making it the best performer for this meat subtype. From the LSTM model, we can see the complex dependencies in time series data, which benefits from adding external features such as economic indicators and seasonal trends. Other models like **SARIMA and SVR with external features** also performed well but LSTM

Meat Sub type (With External)	**SARIMA R^2 (With External)**	**SARIMA RMSE (With External)**	**RF R^2 (With External)**	**RF RMSE (With External)**	**GB R^2 (With External)**	**GB RMSE (With External)**	**SVR R^2 (With External)**	**SVR RMSE (With External)**	**LSTM R^2 (With External)**	**LSTM RMSE (With External)**
Boxed Beef Cutout	0.8942	0.0159	0.5223	0.0337	0.6226	0.0299	0.6042	0.0307	0.8962	0.0157
Imported Boneless Beef, 90% Frozen	0.6872	0.0456	0.5761	0.0531	0.7223	0.0429	0.7059	0.0442	0.6833	0.0459
Pork Cutout Composite	0.9136	0.0179	0.8869	0.0205	0.9556	0.0128	0.8918	0.0200	0.9615	0.0146
Loins, 1/4" Trimmed Vacuum-Packed	0.9264	0.0180	0.8244	0.0278	0.9293	0.0017	0.8060	0.0292	0.9440	0.0157
Drumsticks	0.7664	0.0109	0.7264	0.0118	0.7981	0.0101	0.7971	0.0101	0.5808	0.0146
Wings, Full Cut	0.8920	0.0101	0.6797	0.0173	0.9187	0.0087	0.0000	0.0000	0.6652	0.0423

Table 5.2: Model Performance Comparison for Different Meat Sub types (With External Features)

performed better due to its ability to handle sequential data more effectively.

Imported Boneless Beef, 90 Percent, Frozen: The **SVR model without external features** with R2 score of (0.8416) and RMSE (0.0325). The SVR model focuses on predicting continuous outputs rather than classifying data points and it captures non-linear relationships that benefit from the external data and provides additional power for the meat sub-types. Likewise, other models like LSTM and Gradient Boosting performed well, but SVR had advanced flexibility and adaptability in handling data patterns which is more advantageous to this model.

Pork Cutout Composite: LSTM Model with external features achieved the highest R2 score (0.9615) and lowest RMSE (0.0146). This model captured the long-term dependencies and non-linear relationships which gave additional power to external features to capture these price movements accurately. Moreover, other models like gradient boosting performed well not as compared to the LSTM model.

Loins, 1/4" Trimmed Vacuum-Packed: The Gradient Boosting model without external features performed achieving the highest R2 score (0.9881) and very low RMSE (0.0072) values among all other meat subtypes and this model tells that it recognized the patterns of prices very accurately without any external features data. There is a slight reduction in performance when external features were added which tells that these features added complexity which did not contribute significantly to the predictive power.

Drumsticks: The **Gradient Boosting model without external feature** performed best for Drumstick with an R2 score of (0.8530) and RMSE (0.0086). This indicates that the gradient boosting model captured the important factors with the price dynamic of this subtype even without external data. **The LSTM model without external features** also showed reasonable performance and capability to learn from sequential data, although it did not perform more than the gradient boosting model

Wings, Full Cut: The **LSTM without external feature** has shown the highest R2 score (0.9467) and the lowest RMSE (0.007), making it the most effective model for this subtype. Without external features, LSTM models perform well for wings, which tells that this feature is very critical content in the historical price data alone. Like other **LSTM model without external features** models like SARIMA and Gradient Boosting With External Features performed well but the LSTM model without external features was able to capture more accurate values.

5.1.1 Interpretation of Results

Overall, this research tells that **LSTM Model** with external features performed well apart from all other models across most of the sub-types. This tells models are particularly affected by external information to improve the forecasting accuracy. The LSTM Model without external information also performed well only using the Historical information to capture the forecasting accuracy. The **LSTM model** strength is to handle complex dependencies and patterns, which work well for time series data. So ideally LSTM works best with a great R2 score and RMSE values for this Meat Price Predictions. The **SVR model** without external information performed well for imported boneless beef meat subtype and is known for handling non-linear relationships that benefit from a broader range of data variation.

Other meat sub-types like Loins and Drumstick **Gradient Boosting without external feature** have shown superior performance, highlighting their capacity for non-linear relationships within the data itself. This finding indicates that the choice of the model selections and adding external features should be carefully analyzed based on data patterns of each meat sub types.

5.1.2 Future Research

To enhance this model more future research analysis should be integrated by adding more border limits of external features, such as macroeconomic indicators, consumer behavior metrics, and International trade statistics to give model accuracy more precise. Additionally developing hybrid models that combine the strength of different approaches (e.g. LSTM with Gradient Boosting) could provide even more better forecasting capabilities. Fine-tuning hyperparameters by adding more parameters for each model would help to understand the model's performance with a good R2 score and low RMSE value. Further adding more agriculture products or commodities would provide generalized insights into price dynamics across different markets.

Bibliography

[1] Yuan, J., Hao, J., Liu, M., Wu, D., & Li, J. (2022). A dynamic ensemble learning approach with spectral clustering for beef and lamb prices prediction. *Procedia Computer Science*, 214, 1190-1197. https://doi.org/10.1016/j.procs.2022.11.295.

[2] Zhao, S., Lin, X., & Weng, X. (2022). A Method for Forecasting the Pork Price Based on Fluctuation Forecasting and Attention Mechanism. *International Conference on Machine Learning and Cybernetics (ICMLC)*, Japan, 2022, pp. 18-24. https://doi.org/10.1109/ICMLC56445.2022.9941318.

[3] Wickramarachchi, A., Herath, K., Jayasinghe-Mudalige, U., Edirisinghe, J., Udugama, M., Lokuge, N., & Wijesuriya, W. (2017). An Analysis of Price Behavior of Major Poultry Products in Sri Lanka. *Journal of Agricultural Sciences*, 12, 138. https://doi.org/10.4038/jas.v12i2.8231.

[4] Ben Abdallah, M., Fekete Farkas, M., & Lakner, Z. (2020). Analysis of meat price volatility and volatility spillovers in Finland.*Agricultural Economics (Zemedelska ekonomika)*, 66(2), 84-91. https://doi.org/10.17221/158/2019-agricecon.

[5] Xu, K., Dou, Y., & Chen, J. (2020). Analysis of price fluctuation characteristics of hog industry chain in Yunnan Province: Based on livestock data 2001-2017. *International Conference on Modern Education and Information Management (ICMEIM)*, Dalian, China, 2020, pp. 579-582. https://doi.org/10.1109/ICMEIM51375.2020.00133.

[6] Whitton, C., Bogueva, D., Marinova, D., & Phillips, C.J.C. (2021). Are We Approaching Peak Meat Consumption? Analysis of Meat Consumption from 2000 to 2019 in 35 Countries and Its Relationship to Gross Domestic Product. *Animals*, 11(12), 3466. https://doi.org/10.3390/ani11123466.

[7] Amalia, S., Dhini, A., & Zulkarnain. (2022). Indonesia's Food Commodity Price Forecasting using Recurrent Neural Networks. *International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*, Kochi, India, pp. 1-6. https://doi.org/10.1109/IC3SIS54991.2022.9885249.

[8] Sadefo Kamdem, J., Bandolo Essomba, R., & Njong Berinyuy, J. (2020). Deep learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities. *Chaos, Solitons and Fractals*, 140, 110215. https://doi.org/10.1016/j.chaos.2020.110215.

[9] Peshevski, D., Todorovska, A., Trajkovikj, F., Hristov, N., Trajanoska, M., Dobreva, J., Stojanov, R., & Trajanov, D. (2023). Methodology for food prices forecasting. *IEEE Big Data Conference*, 4539-4547. https://doi.org/10.1109/BigData59044.2023.10386082.

[10] Akbar, M., Hasnudi., Supriana, T. (2021). Analysis Factors Affecting the Demand and Supply of Free-range Chicken Meat in Pakpak Bharat Regency. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 4(1), 986-998. https://doi.org/10.33258/birci.v4i1.1718.

[11] Fahrudin, R., Kusnadi, K., & Lukita. (2024). Price Prediction System of Basic Commodities Using Long Short-Term Memory Method: Analysis and Implementation for Future Projections. *Journal Research of Social Science, Economics, and Management*, 3(7), 1617. https://doi.org/10.59141/jrssem.v3i7.629.

[12] Mattos, F., Garcia, P., Leuthold, R., & Hahn, T. (2003). The Feasibility of a Boxed Beef Futures Contract: Hedging Wholesale Beef Cuts. *Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, St. Louis, MO. http://www.farmdoc.uiuc.edu/nccc134.

[13] Luke, J. R., Tonsor, G. T., & Brown, D. S. (2024). Wholesale pork demand: Understanding primal-level heterogeneity. *Journal of Commodity Markets*, 34, 100402. https://doi.org/10.1016/j.jcomm.2024.100402.

[14] Calvia, M. (2024). Beef, lamb, pork and poultry meat commodity prices: Historical fluctuations and synchronisation with a focus on recent global crises.

Agricultural Economics - Czech, 70(1), 24–33. https://doi.org/10.17221/361/2023-AGRICECON.

[15] Anderson Antonio Carvalho Alves, A., Chaparro Pinzon, A., Magalhães da Costa, R., Santos da Silva, M., Mendes Vieira, E. H., Barbosa de Mendonça, I., de Sena Sales Viana, V., & Braga Lõbo, R. N. (2019). Multiple regression and machine learning based methods for carcass traits and saleable meat cuts prediction using non-invasive in vivo measurements in commercial lambs. *Small Ruminant Research*, 171, 49–56. https://doi.org/10.1016/j.smallrumres.2018.12.008.

[16] https://medium.com/@jihyun.kim423/exploring-the-seasonal-arima-sarima-model-for-forecasting-differences-from-arima-e30c3488e5f6.

[17] https://medium.com/@poojaviveksingh/all-about-min-max-scaling-c7da4e0044c5.

[18] https://www.visual-design.net/post/time-series-analysis-arma-arima-sarima.

[19] https://cienciadedatos.net/documentos/py27-time-series-forecasting-python-scikitlearn.html.

[20] https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm.

[21] https://www.projectpro.io/article/lstm-model/832.

[22] https://www.analyticsvidhya.com/blog/2022/01/the-complete-lstm-tutorial-with-implementation/

[23] https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning/

[24] https://www.researchgate.net/figure/The-architecture-of-Gradient-Boosting-Decision-Tree_fig2_356698772

[25] https://www.statology.org/rmse-vs-r-squared/