

**Step 1: Define the Business Problem**

**Business Problem**:

Agricultural producers, wholesalers, and retailers are facing challenges with an inefficient pricing structure and lack of transparency in the agricultural supply chain. As a result, producers are often unable to access real-time market data to set competitive and fair prices for their goods, while wholesalers and retailers are grappling with inconsistent price points and margins. This creates a situation where pricing is often disconnected from market realities, leading to inefficiencies, price volatility, and unsustainable practices across the supply chain. Additionally, policy decisions related to agricultural subsidies and consumer protection are hindered by the absence of clear pricing data and insights, making it difficult to effectively regulate the industry or support farmers and consumers.

To address these challenges, it is crucial to implement systems and strategies that improve price transparency, enable real-time data sharing, and foster data-driven pricing strategies to support more efficient and fair pricing across the agricultural supply chain. Additionally, these improvements will assist policymakers in making better-informed decisions to support the agricultural industry, ensure fair consumer pricing, and reduce inefficiencies.

The objectives are focused on improving the agricultural supply chain's efficiency and transparency. Key goals include enhancing price transparency to benefit all stakeholders and enabling data-driven pricing strategies for producers, wholesalers, and retailers. Additionally, insights from the data can support policy decisions related to agricultural subsidies, price regulations, and consumer protection. A specific point of interest is the notable events in 2015, particularly for beef production, which saw significant changes. Understanding what happened during this year could provide critical context for broader trends in the industry.

**Potential Research Questions:**

How do farm-level prices influence retail prices, what is the time lag between these changes, and how do price spreads across the supply chain vary for different byproducts?

The relationship between prices at the farm level and retail prices can be examined through statistical analysis, such as correlation and regression, to determine the extent of their influence. While farm-level prices are expected to have a direct impact on retail prices, this influence may be mediated by factors like wholesale costs, processing, transportation, and retailer markups. Additionally, external variables such as policy changes or shifts in consumer demand may weaken this relationship. To understand the time lag between changes in farm prices and their reflection at the retail level, time-series techniques, such as cross-correlation analysis, can be used. This analysis helps identify delayed responses in retail prices, which typically lag due to the time required for processing, distribution, and market adjustments. The lag period could vary depending on the supply chain's complexity and other influencing factors. Price spreads between farm and wholesale levels, as well as wholesale and retail levels, vary across different byproducts. By computing these spreads, it becomes possible to observe trends and assess variability for specific items like beef, pork, and poultry. Larger spreads may occur in products with higher processing or transportation costs, while variations in spreads over time could indicate market inefficiencies, seasonal demand shifts, or the effects of policies such as subsidies.

What seasonal trends and long-term changes exist in farm, wholesale, and retail prices, and are there consistent patterns in price spreads during specific months or seasons?

By examining seasonal fluctuations, such as higher turkey prices around Thanksgiving or increased beef prices during summer grilling season, we can determine whether these trends are consistent across years. This requires analyzing month-over-month price changes and visualizing seasonal peaks or troughs. Long-term trends can be assessed by observing year-over-year changes in prices for farm, wholesale, and retail levels to identify growth, decline, or stability over time. Additionally, price spreads—differences between farm-wholesale and wholesale-retail prices—can be evaluated for consistent seasonal patterns. These might reveal whether certain months or seasons are associated with narrower or wider spreads due to factors like demand spikes, supply shortages, or increased transportation costs during specific times of the year.

Which stage of the supply chain contributes most to price spreads, are there byproducts with disproportionate spreads, and do wholesale prices consistently pass through to retail prices without delays or inefficiencies?

The analysis focuses on identifying supply chain inefficiencies by examining which stage—Farm-Wholesale or Wholesale-Retail—contributes the most to price spreads for each byproduct. It also evaluates whether specific byproducts have disproportionately large or small spreads, indicating potential inefficiencies or unique characteristics in their supply chains. Additionally, the study assesses whether wholesale prices consistently influence retail prices or if delays and inefficiencies in price transmission exist, potentially due to factors like market structure or operational constraints.

Can retail prices be predicted based on farm and wholesale prices, and what factors are the strongest predictors of retail prices?

The analysis investigates whether retail prices can be accurately predicted using farm and wholesale prices. It also aims to identify the most influential factors, such as gross farm value, price spreads, and seasonality, in predicting retail prices. By leveraging machine learning or statistical models, the study seeks to uncover the key drivers of retail pricing and assess their predictive power.

**Datasets Where are you getting your data? Describe the data that you will use to solve the problem**

This data set provides monthly average price values, and the differences among those values, at the farm, wholesale, and retail stages of the production and marketing chain for all the meat produced by Choice-grade steers and hogs. This data set allows one to compare the values of animals at the farm, at the packing plant, and in the grocery store. The data set also includes wholesale and retail values for broilers.

<https://www.ers.usda.gov/data-products/meat-price-spreads.aspx>

These data sets are released monthly. Price spreads for Choice beef and pork are reported for last 6 years, 12 quarters, and 24 months. The retail price file provides monthly estimates for the last 24 months. The historical file provides monthly data since 1970.

Retail prices and values are based on data provided by the U.S. Department of Labor, Bureau of Labor Statistics. Livestock and wholesale meat and poultry prices are derived from U.S. Department of Agriculture, Agricultural Marketing Service reports.

People have used the retail values in this data set as a measure of the costs of beef and pork to consumers. The data also show the farmer’s share of the retail costs of Choice beef and pork.

Farm-to-consumer price spreads for other foods can be found in the [Price Spreads from Farm to Consumer](https://www.ers.usda.gov/data-products/price-spreads-from-farm-to-consumer/price-spreads-from-farm-to-consumer/) dataset.

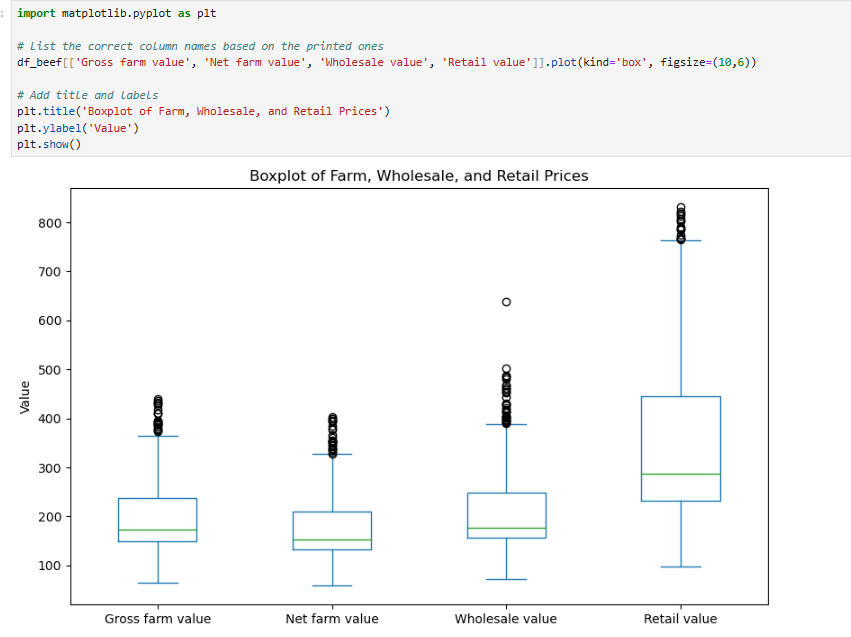
**Key Features**:

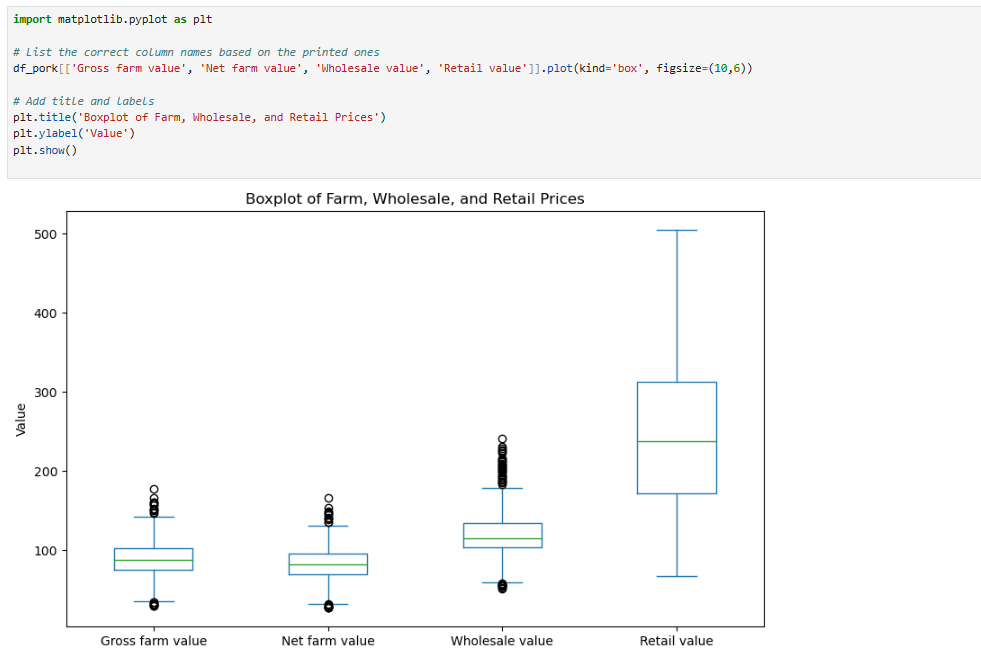
* **Date:** month/Year
* **ByProduct**
* **Gross farm value:** Initial value at the farm level.
* **Net farm value:** Farm-level value after deducting certain costs (e.g., transportation).
* **Wholesale value:** Value at the packing or processing stage.
* **Retail value:** Consumer-facing price in stores.
* **Total:** Likely a summation (you can confirm this) of all values or related costs.
* **Farm-wholesale:** Difference between wholesale and net farm values.
* **Wholesale-retail:** Difference between retail and wholesale values.

**Methods for analyzing the agricultural data set**

Analyzing the agricultural dataset involves several methods that provide valuable insights. Descriptive analysis is used to summarize and understand the data, identifying patterns and trends across variables. Time-series analysis takes advantage of the date column to explore temporal trends, seasonality, and long-term changes in prices and other metrics. Additionally, predictive modeling, through machine learning or statistical techniques, enables forecasting future values and uncovering relationships between variables. These methods collectively contribute to better decision-making and strategic planning based on the dataset's insights.

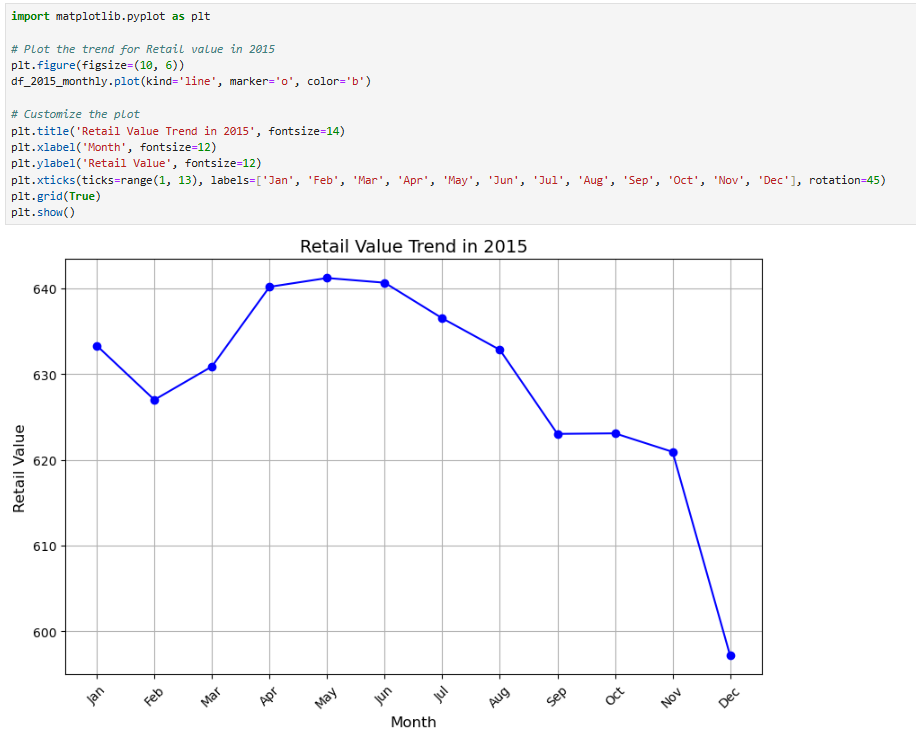
The comparison between beef and pork production reveals several key differences in terms of market dynamics, production processes, pricing trends, and consumer demand.





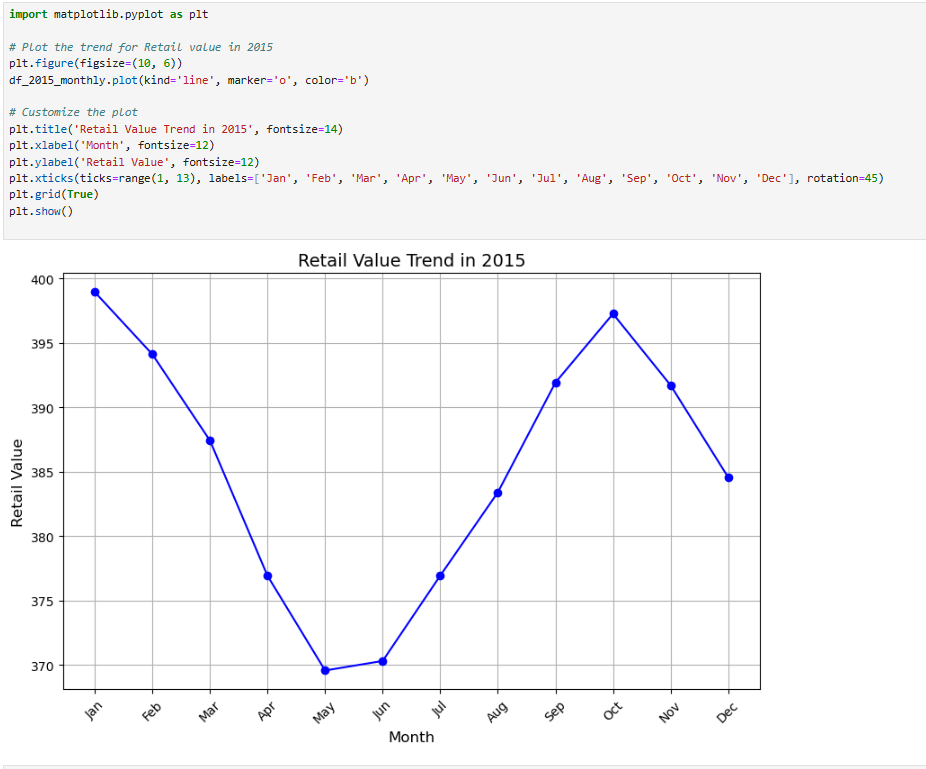
**Retail Trend in 2015 Beef**

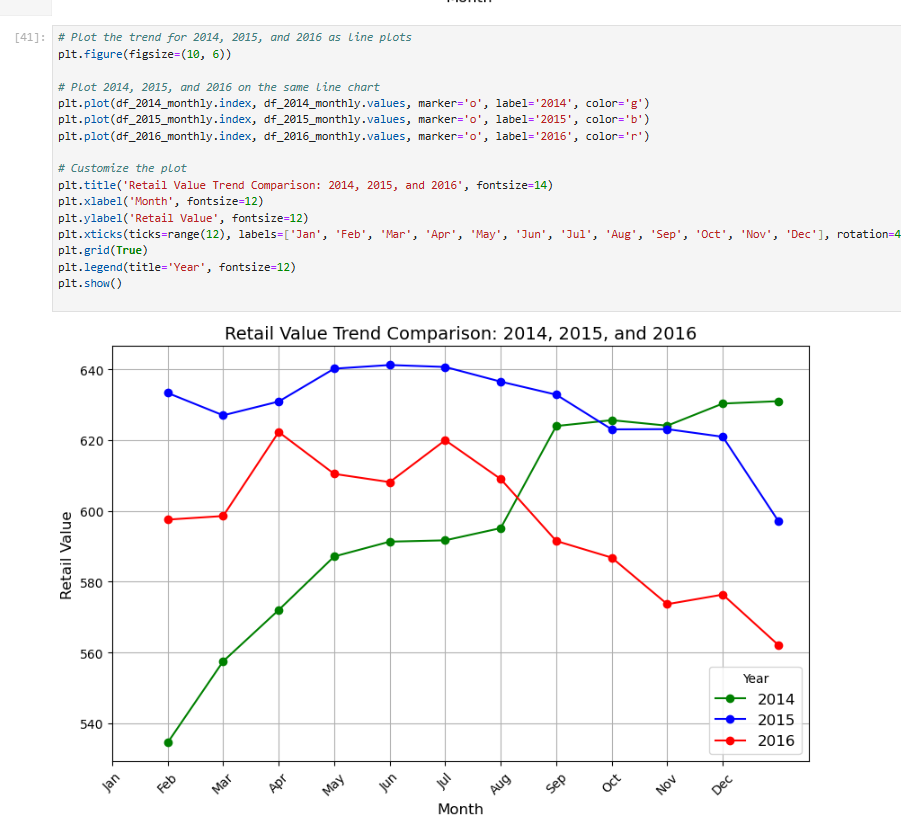
**Beef**: Beef prices tend to be more volatile due to supply constraints, higher production costs, and weather-related issues that affect feed availability. In years of strong demand or limited supply, beef prices can surge, as seen in 2015, which was a banner year for beef production.



**Retail Trend in 2015 Pork**

**Beef**: Beef prices tend to be more volatile due to supply constraints, higher production costs, and weather-related issues that affect feed availability. In years of strong demand or limited supply, beef prices can surge, as seen in 2015, which was a banner year for beef production.





**What happened in 2015 beef:**

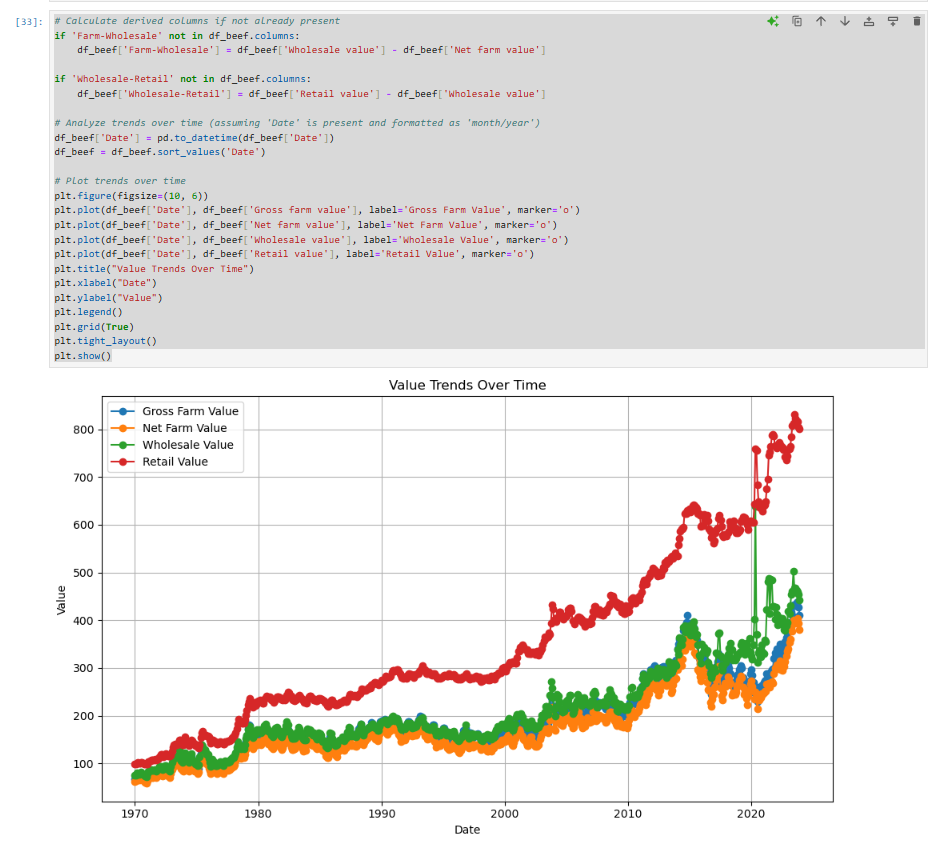
In 2015, the retail value of beef experienced a notable decline compared to 2014 and 2016. Starting with relatively high values in January, there was a gradual decrease throughout the year, particularly from July to December. This trend suggests challenges such as reduced demand, oversupply, or rising production costs. Compared to 2014, which showed steady increases, and 2016, which maintained consistency, the sharp decline in 2015 indicates a significant market disruption. Additionally, unlike the other years, where retail values showed recovery or stability towards the end of the year, 2015 continued its downward trend, highlighting sustained negative factors affecting the beef market.

In 2015, the U.S. beef industry experienced a notable increase in production driven by multiple factors. Following years of drought and reduced herd numbers, cattle inventories began recovering as ranchers rebuilt their herds, leading to higher production levels. Elevated beef prices, resulting from previous supply shortages, incentivized producers to increase output to capitalize on favorable market conditions. Domestic and international demand for beef, particularly in key export markets like Japan and South Korea, further supported the rise in production. Additionally, lower feed costs, such as corn and grains, made production more cost-effective for ranchers. This increase was also part of the natural cattle production cycle, as the industry moved into a phase of expansion after prior contraction. These combined factors contributed to significant growth in beef production during 2015.

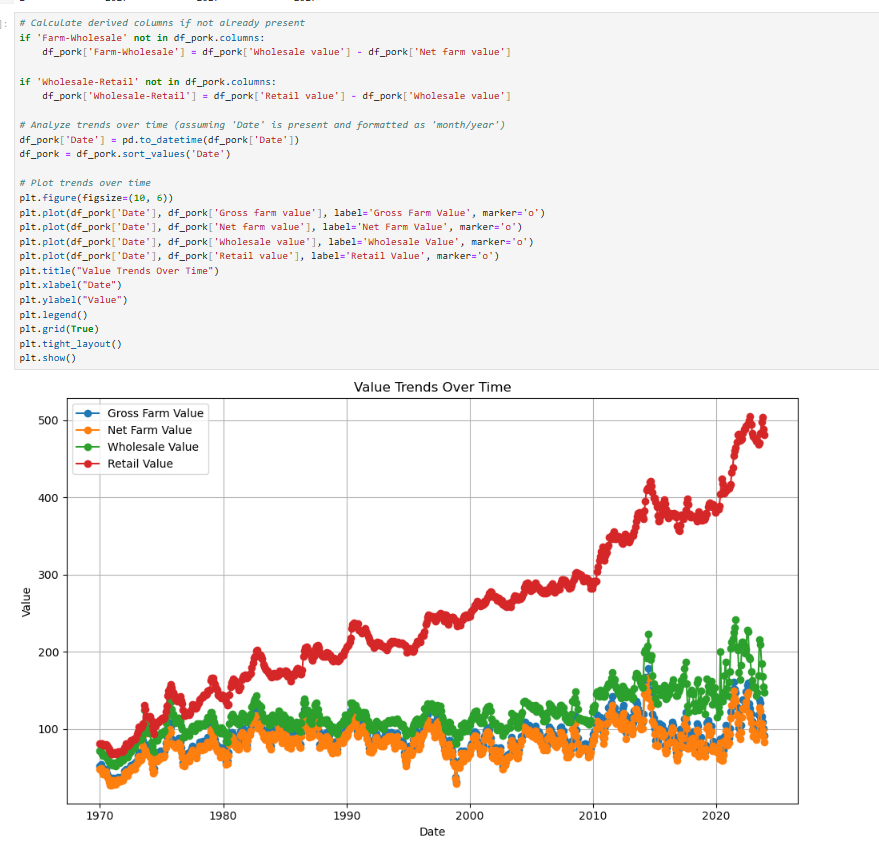
**Evaluating trends of time**

This graph illustrates the trends of various agricultural value metrics over time, including Gross Farm Value, Net Farm Value, Wholesale Value, and Retail Value. Notably, the Retail Value consistently outpaces the other metrics, showing a significant markup as products move through the supply chain. While the Gross Farm and Net Farm values show steady growth, they lag considerably behind the Wholesale and Retail values, highlighting disparities in pricing across the supply chain. Recent years also exhibit increased price volatility, particularly in the Retail and Wholesale values, which may indicate market instability or evolving pricing structures.

**Beef:**



**Pork:**



The comparison between pork and beef value trends reveals distinct patterns. Retail values for both products consistently exceed wholesale and farm-level values, reflecting significant markups as they move through the supply chain. Beef generally commands higher prices across all levels compared to pork, potentially due to differences in production costs, consumer demand, or market conditions. Both pork and beef exhibit steady growth in values over time, but beef shows higher volatility, particularly in retail pricing. This disparity suggests differing market dynamics, pricing structures, or supply chain challenges for the two commodities, underscoring the need for targeted strategies to enhance transparency and efficiency.

**Predictive Modelling:** After analyzing the historical trends of beef and pork values across the supply chain, it became evident that both commodities exhibit unique growth patterns and price volatilities. Beef consistently shows higher values compared to pork, with more pronounced fluctuations, particularly in retail prices, likely driven by higher production costs, consumer demand, and market dynamics. In contrast, pork demonstrates steadier growth and relatively smaller margins across the supply chain.

To address the need for future state predictions, predictive models were employed. Using Random Forest Regression and ARIMA time-series analysis, forecasts for future values were generated based on historical data and supply chain metrics such as Gross Farm Value, Net Farm Value, Wholesale Value, and Retail Value. These models revealed potential price trajectories and identified key factors influencing future pricing.

The predictions provide valuable insights for stakeholders to anticipate market changes, optimize pricing strategies, and make data-driven decisions. For policymakers, these forecasts can help shape targeted interventions to stabilize markets and support producers, wholesalers, and retailers in the agricultural supply chain.

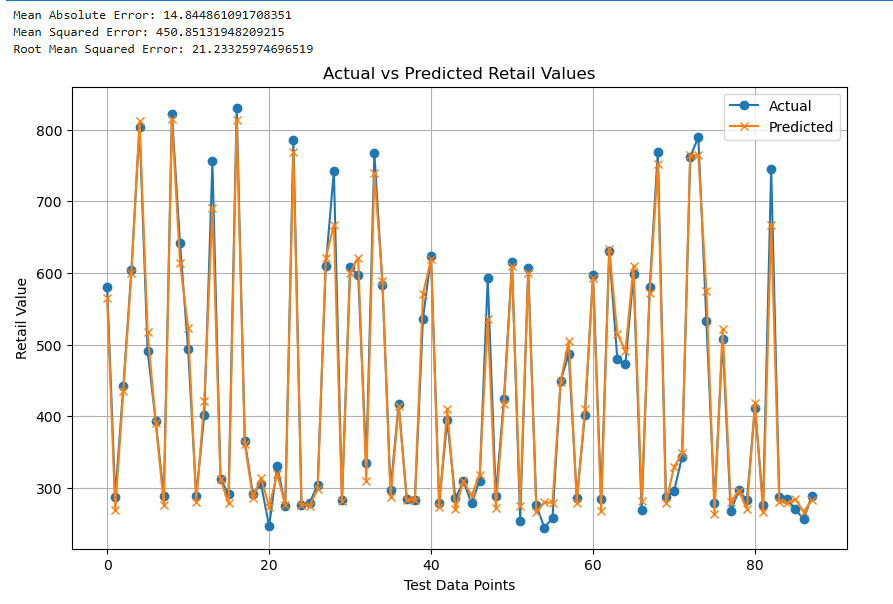
**Beef:**

The updated evaluation metrics indicate a significant improvement in the model’s performance:

1. **Mean Absolute Error (MAE):** 14.84
2. **Mean Squared Error (MSE):** 450.85
3. **Root Mean Squared Error (RMSE):** 21.23

**Summary:**

The model demonstrates solid predictive performance with minimal errors, indicating its effectiveness in capturing trends and relationships within the dataset. The low MAE and RMSE suggest that the model is accurate and suitable for forecasting future values or analyzing key trends in agricultural pricing.



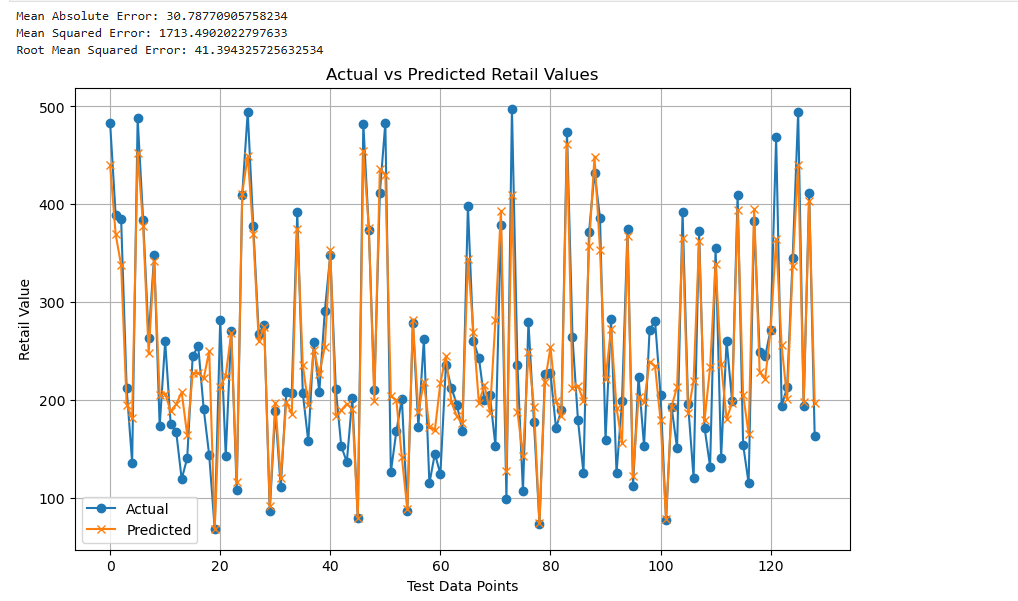
Pork:

The predictive model evaluation yielded the following error metrics:

1. **Mean Absolute Error (MAE):** 30.79
2. **Mean Squared Error (MSE):** 1713.49
3. **Root Mean Squared Error (RMSE):** 41.39

**Interpretation:**

The MAE and RMSE metrics indicate that the model delivers reasonably accurate predictions, though some deviations from the actual values are present. The higher RMSE compared to the MAE highlights the influence of a few significant deviations, suggesting that the model's performance is disproportionately affected by outliers or larger errors in predictions.



**Ethical Considerations**

When working with agricultural datasets, several ethical considerations must be addressed. Data privacy and confidentiality are paramount, requiring the anonymity of sensitive information, ensuring access permissions are in place, and securing data storage to prevent unauthorized breaches. Accuracy and bias must also be managed to avoid misrepresentation, ensuring findings are truthful and acknowledging limitations such as missing data or regional differences.

The impact on stakeholders, including producers, consumers, and the environment, must be carefully considered. For producers, findings should not support policies or actions that disadvantage farmers or unfairly consolidate market power. For consumers, the results should avoid contributing to price manipulation or reduced access to affordable food. Finally, the environmental implications of agricultural practices should be considered to ensure that findings do not unintentionally encourage unsustainable practices.

**Challenges**

Analyzing the agricultural dataset comes with several challenges. Data quality issues, such as missing or incomplete entries, can hinder accurate analysis. Ambiguities in the "Byproduct" column create difficulties in interpreting and categorizing the data effectively. Temporal factors add complexity, as agricultural processes are often influenced by seasonality, requiring advanced techniques to account for these patterns. Scalability poses another challenge; as the dataset grows, computational demands increase, potentially impacting performance. Lastly, establishing causation remains difficult, as strong correlations in the data do not necessarily imply direct causal relationships, making it essential to approach conclusions cautiously.

**References:**

<https://www.ers.usda.gov/data-products/meat-price-spreads>

**About the Dataset**

This data set provides monthly average price values, and the differences among those values, at the farm, wholesale, and retail stages of the production and marketing chain for all the meat produced by Choice-grade steers and hogs. This data set allows one to compare the values of animals at the farm, at the packing plant, and in the grocery store. The data set also includes wholesale and retail values for broilers. In addition, retail prices are provided for beef and pork cuts, turkey, whole chickens, eggs, and dairy products. These data are summary statistics meant to measure national average prices.

These data are released monthly. Price spreads for Choice beef and pork are reported for last 6 years, 12 quarters, and 24 months. The retail price file provides monthly estimates for the last 24 months. The historical file provides monthly data since 1970.

Retail prices and values are based on data provided by the U.S. Department of Labor, Bureau of Labor Statistics. Livestock and wholesale meat and poultry prices are derived from U.S. Department of Agriculture, Agricultural Marketing Service reports.

People have used the retail values in this data set as a measure of the costs of beef and pork to consumers. The data also show the farmer’s share of the retail costs of Choice beef and pork.

Answers to the Questions:

**Which stage of the supply chain contributes most to price spreads, are there byproducts with disproportionate spreads, and do wholesale prices consistently pass through to retail prices without delays or inefficiencies?**

**Stage of the Supply Chain Contributing Most to Price Spreads**

The most significant contribution to price spreads often comes from the **processing and distribution** stages. Here's why:

**Processing:** This stage involves slaughtering, butchering, and packaging the meat. The cost of processing (labor, equipment, regulatory compliance) tends to be high, contributing to a significant portion of the final price.

**Distribution:** The movement of the meat from processing plants to wholesalers, retailers, or foodservice can involve significant transportation costs, cold storage, and handling fees.

Price spreads are especially noticeable between **wholesale and retail** prices, where the wholesale price is typically lower due to bulk purchasing, and the retail price incorporates additional costs from the retailer (including overhead and profit margin).

**How do farm-level prices influence retail prices, what is the time lag between these changes, and how do price spreads across the supply chain vary for different byproducts?**

Farm-level prices serve as the foundation for meat pricing, but they don't always directly dictate retail prices. Here's how they influence the retail price:

**Initial Impact:** The farm-level price represents the cost paid to farmers for live animals. It is influenced by factors like supply and demand, feed costs, labor, and regulations. A rise or fall in farm-level prices can trigger a chain reaction in the supply chain that eventually affects retail prices.

**Cost Structure Influence:** Retail prices are influenced not only by farm-level prices but also by the additional costs added at each stage of the supply chain, including slaughter, processing, transportation, storage, and retail markups. As farm-level prices rise, processors may raise their prices, and retailers may adjust their prices to maintain margins.

**Retail Pricing Strategy:** Retailers often use complex pricing strategies (e.g., price signaling, dynamic pricing) that don’t immediately reflect farm-level price changes. They may absorb some cost increases to protect sales volume or wait until the cost changes become more consistent across the supply chain before adjusting.

**What seasonal trends and long-term changes exist in farm, wholesale, and retail prices, and are there consistent patterns in price spreads during specific months or seasons?**

Seasonal trends in meat prices are influenced by several factors, including consumer demand, weather conditions, feed costs, and market dynamics. These trends often result in consistent patterns of price changes throughout the year at each stage of the supply chain:

**Farm Prices:**

Spring and Summer: Generally, farm-level prices for meat (especially beef and pork) tend to be lower in the spring and early summer. This is often because animals are being raised and prepared for slaughter, and there’s a larger supply available after the winter months. For example, when cattle are finished and ready for market, the supply increases, leading to lower prices at the farm level.

Fall and Winter: Farm prices often increase in the fall and winter, particularly in anticipation of holidays (like Thanksgiving and Christmas), when consumer demand is higher for meats like turkey, beef, and pork. Additionally, extreme weather conditions such as droughts or cold temperatures can drive up costs for feed and transportation, increasing prices at the farm level.

Livestock Cycles: The cyclical nature of livestock production also affects farm prices. For example, pork and beef production cycles span several months, which means supply and prices can fluctuate based on production rates in previous years.

**Long-Term Changes in Prices and Price Spreads**

**Input Costs and Feed Prices:** Over the long term, changes in the cost of feed, fuel, and labor can drive up the cost of production at the farm level, which typically passes through the supply chain, affecting wholesale and retail prices. This is a key driver of price inflation over several years.

**Consumer Preferences:** Shifts in consumer demand for certain types of meat (e.g., plant-based alternatives, or organic and grass-fed beef) can lead to long-term changes in pricing patterns. As demand for premium or niche meats rises, those prices may increase at the farm, wholesale, and retail levels, affecting price spreads across the supply chain.

**Global Market Trends:** International trade policies, tariffs, and global demand for meat (e.g., from export markets like China) can influence price trends in the U.S. and other countries. For instance, when global demand for U.S. beef increases, farm-level prices may rise, impacting prices at all stages of the supply chain.

**Technological Advancements:** Over the longer term, technological changes in meat production, processing, and distribution can lead to cost reductions or efficiency gains. This can help stabilize or even reduce prices over time, though changes in consumer expectations (e.g., for more humane or sustainable practices) might offset some of these savings.