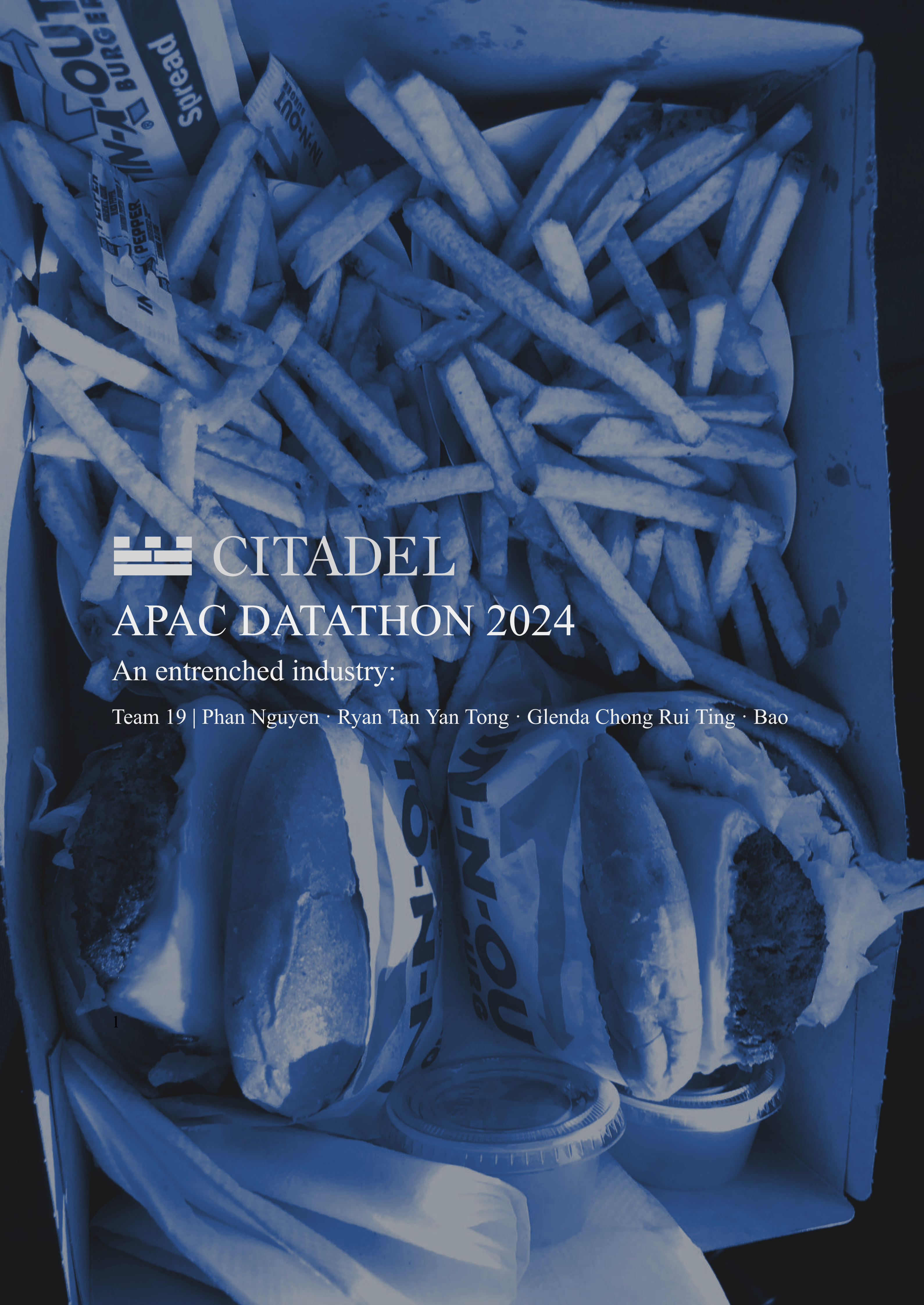


CITADEL

APAC DATATHON 2024

An entrenched industry:

Team 19 | Phan Nguyen · Ryan Tan Yan Tong · Glenda Chong Rui Ting · Bao

1 Executive Summary

1.1 Research Motivation

The American relationship with processed food is multifaceted and complex, intertwining cultural, economic and health considerations. Despite the growing awareness of the negative health implications associated with consumption of processed foods, their prevalence continues to rise, contributing to alarming obesity rates and health-related issues, with approximately 678,000 Americans succumbing to such conditions annually [5].

Yet, the United States (US) still remains largely reluctant and wary to enact substantive measures to reduce public consumption of processed foods, as evidenced by the scarcity of policies targeting ultra-processed food, only 25 policies implemented within the span of 29 years [8]. The processed food industry plays a significant role in the American economy, employing millions across various sectors and driving substantial financial investments [7].

So how exactly does the processed food industry influence our health and the economy? This study seeks to demystify the entrenchment of the industry of processed foods in the United States economy and shed light on the true impacts of the excessive consumption of processed foods in the United States. More specifically, we will investigate the obesity trends across different demographic populations and locations, as well as the relationship between the processed food industry and stock prices.

1.2 Key Findings

In this report, we present an analysis on the obesity rates among the adult and student populations in the United States, with the application of SARIMAX models, feature engineering and financial modelling to forecast meat production levels and market capitalisations of key processed food companies.

The upward trend in the obesity rate has been clearly demonstrated by the growth in the meat industry, including both production and cold storage quantities, highlighting the current and forecasted state of health in the United States should the trend continue.

The food industry thrives during tough economic periods. Market capitalization of pre processed food companies saw huge spikes during 2009, 2011, and most significantly 2020. This corresponds with the Great Recession, the European debt crisis and the Covid-19 pandemic, respectively.

Finally, meat supply and demand is highly seasonal. In particular, demand for beef, chicken and pork is at peak during summer and drops during winter. However, rather paradoxically, the average slaughtered weight for these animals follow an opposite trend.

Through this analysis, we identified key breakout levels that policymakers could target to limit the continued excessive growth of such processed food companies.

2 Technical Exposition

2.1 General Approach

In order to identify important insights, we first conducted exploratory data analysis on each dataset. The important insights from various datasets that correspond to various facets of the food sector will then be combined to create findings that result in challenges, solutions, and conclusions.

In the initial stage, we made an effort to gather data while organising and purifying it. Plotting serves as a tool for us to observe patterns and anomalies as well as to pinpoint important characteristics that could be crucial.

The first strategy was to identify the obesity situations in the United States, broken down by states and regions, and then use historical data to estimate predictions.

Next, we shift our attention to cold storage, meat production, and stock prices. We attempt to analyse meat production using relevant stock tickers in order to determine which had the greatest influence on the sector and output. The relationship between the two most significant meat datasets—Cold Storage and Meat Production—is then determined. This discovery will yield some accurate details regarding the steps in the production process that occur between actual production and reserve—cold storage for later use.

To help with our subsequent prediction and discoveries, we also add new features to both meat production and stock prices. Furthermore, we employed ARIMA and SARIMAX models to predict our own features in order to obtain insights.

Finally, we use a number of key criteria, with XGBOOST, to predict our own financial aspects in order to capture the future image of the scenarios — that is, the terrible reality of obesity and the state of the economy.

2.2 Data Collection and Cleaning

Our exploration of the datasets revealed significant missing data, particularly in meat production, cold storage, and slaughter counts for the pre-1960s period. Since recent data holds greater predictive power and relevance for our analysis, we made the decision to remove these missing data points. This approach focuses on the most impactful information for the Exploratory Data Analysis (EDA) stage.

Furthermore, we encountered attributes with limited value for our analysis. In the "Nutrition Physical Activity and Obesity Data.csv" file, There are columns like "Data_Value_Unit" (data in this column are consistently empty), "Data_Value_Footnote_Symbol" and "Data_Value_Footnote" (primarily indicators). Our strategy involves removing these redundant attributes and prioritising those with stronger predictive weight. This ensures a cleaner dataset focused on the most relevant information for our analysis.

Details on specific data usage methods will be provided in each subsequent section.

2.2.1 Description of Dataset

We attempt this by utilising those provided dataset, including:

- **all_stock_and_etfs.csv (1):** Prices for stocks associated with processed foods and index-tracking ETFs on the stock market from initial listing through current day.
- **Meat_Stats_Meat_Production.csv (2):** Red meat and poultry production measures the total quantity of red meat and poultry produced within a specified time frame
- **Meat_Stats_Cold_Storage.csv (3):** Quantity of meat products (both red meat and poultry) that is held in frozen storage at a specific point in time
- **Meat_Stats_Slaughter_Counts.csv (4):** Livestock and poultry slaughter represents the number of animals that are processed and slaughtered for meat production
- **Meat_Stats_Slaughter_Weights.csv (5):** Live Weights: The weight of livestock or poultry at the time of slaughter before any processing. Dressed Weights: The weight of the carcass after slaughter and removal of inedible parts
- **all_commodities.csv (6):** Prices for relevant commodities traded on the stock market from initial listing through current day.
- **Nutrition_Physical_Activity_and_Obesity_Data.csv (7):** Information on obesity and reported physical activity in the US at a state level from 2001-2022.
- **acs_5yr_est_selected_economic_characteristics_2010-2022.csv (8):** The American Community Survey (ACS) is an ongoing survey conducted by the US Census Bureau that provides vital information on a yearly basis about various social and economic characteristics across demographics.

2.3 Obesity Trends Analysis

In this section, we will delve deeper into the health landscape of the United States, in particular, the obesity trends, seeking to confirm established trends while uncovering new insights.

2.3.1 Obesity Trends Across Years

Firstly, we analysed the obesity rates across the years for adults aged 18 years and above and students in Grades 9-12. We only considered the data related to obesity, specifically data containing the Question as 'Percent of adults aged 18 years and older who have obesity' and 'Percent of students in grades 9-12 who have obesity'. We did not include overweight classification as part of the obesity trends because overweight is classified as pre-obesity with a BMI of 25-29.9 kg/m^2, while obesity is defined as a BMI ≥ 30kg/m^2 according to the World Health Organisation [9]. As seen from the results in Figure 1, the obesity rates for both the adult and student populations have been steadily increasing from 2001 to 2022 (where data is available). This coincides with the trend from multiple literature reviews discussed earlier. Moreover, the obesity rates among the adults always outweigh that among the students between 2011 and 2019, with the obesity rates of the adults being approximately double that of students. In fact, the obesity rates among the adults and students surpassed 30% after 2016 and 15% after 2017.

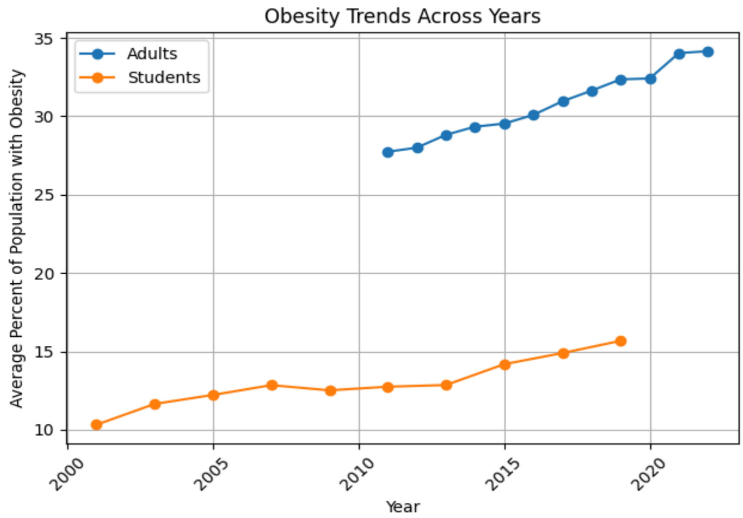


Figure 1: Obesity Rates of Adults Aged 18 and above and Students In Grades 9-12 Across Years 2000 and 2019

2.3.2 Obesity Trends Across All Locations in the United States

Next, we wanted to investigate if the obesity trends differ across the various locations in the United States, considering that it is a huge country. According to Figure 2, it is evident that obesity trends persist consistently across all the states and territories in the dataset, with the average percentage of population with obesity being above 25%. Notably, in Minnesota, Oregon, and Washington, where no data of students were available, the average percent of the adult population with obesity is one of the highest above 25%, compared to the other locations.

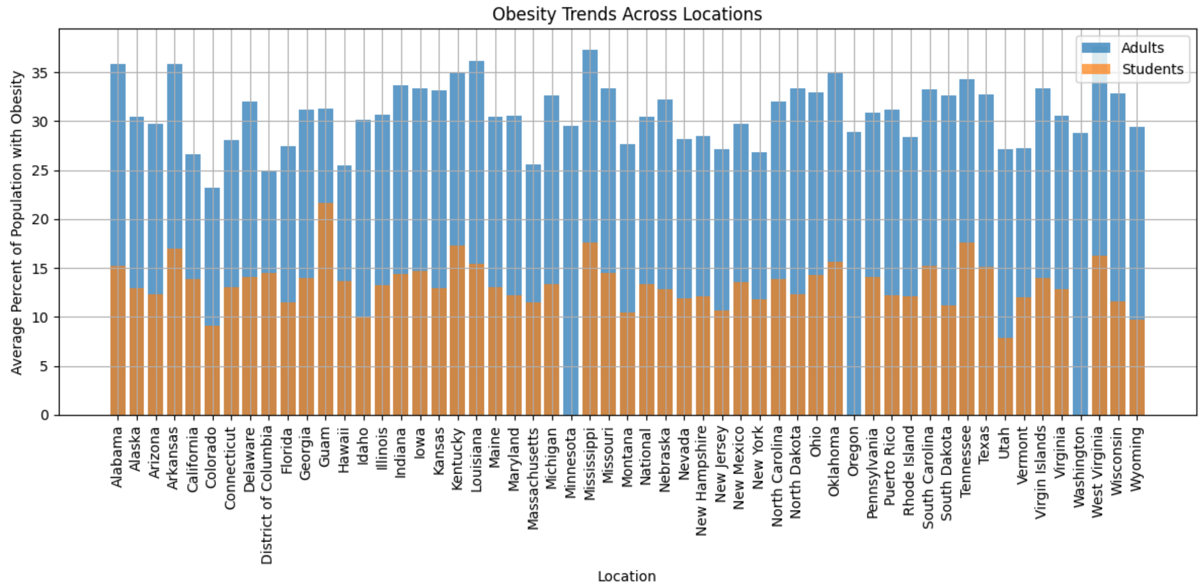


Figure 2: Average Percent of Population with Obesity For Each Location in the Dataset

2.3.3 Obesity Trends Across Regions in the United States

Since each location has very similar results from Section 2.3.2, we went one step further to categorise the locations into regions, like South, West, Northeast and Midwest, in order to investigate if there were any regions with a significant difference in obesity trends among the others.

To do so, we created a new column called 'Region', which contains the mapping of each location to a region, based on information obtained from Google Search. Territories and non-state locations such as 'District of Columbia', 'Puerto Rico' and 'Virgin Islands' were grouped under the 'Other' region category. From Figure 3, we can conclude that each region has a similar average percentage of population with obesity, with all regions surpassing the 25% threshold.. The South region has the highest average percentage, with the Midwest region closely following as the 2nd highest. Both regions exhibit an average population exceeding 30%, emphasising the notable prevalence of obesity in these areas.

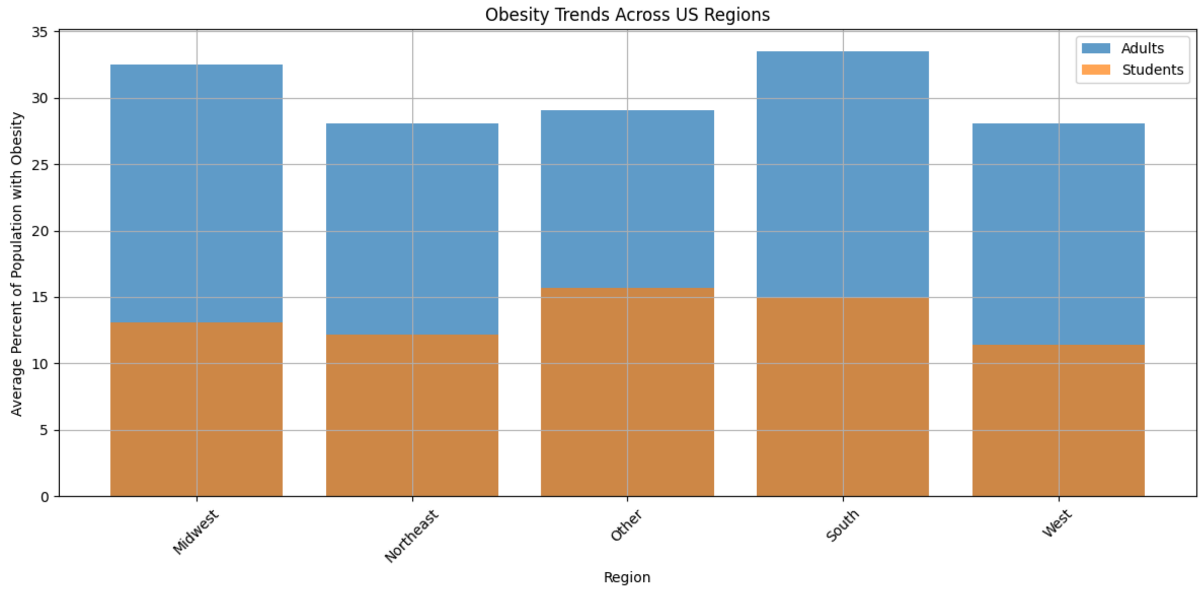


Figure 3: Average Percent of Population with Obesity For Each Region in the United States

2.3.4 Forecasting Obesity Trends

Given the obesity trend across years, we observed a linear relationship between obesity rate and time. As such, we proceeded to perform linear regression to quantify and model the trend to forecast how the average percent of population with obesity will be in the future. According to Figures 4 and 5, we can infer that there is a strong positive correlation between the average percent of population with obesity for both adults and students and time, given by the high R-value of 0.99 and 0.95 respectively.

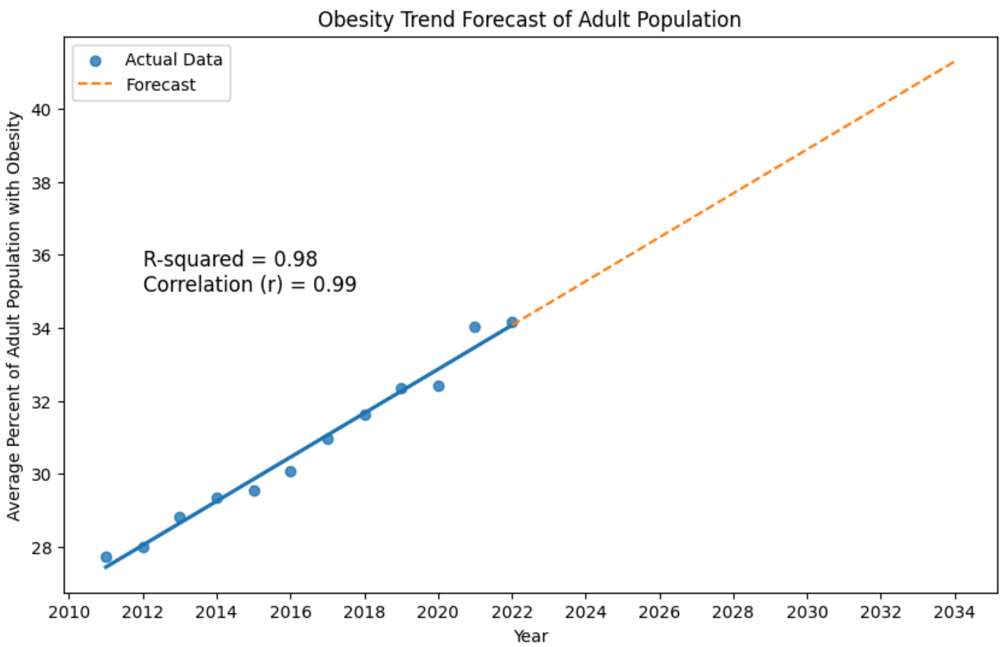


Figure 4: Obesity Trend Forecast of Adult Population

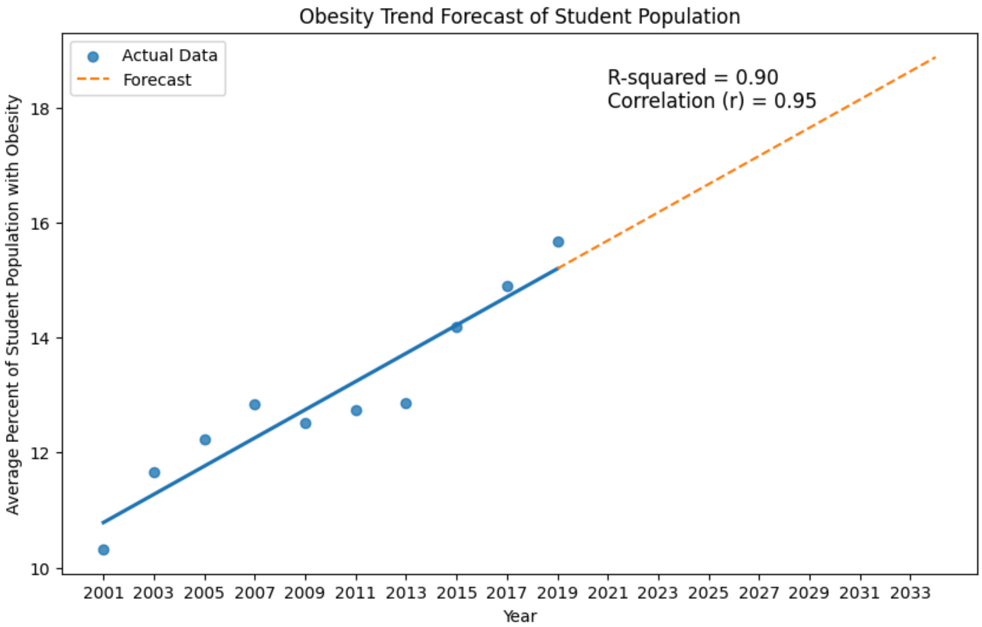


Figure 5: Obesity Trend Forecast of Student Population

2.4 Stock Price & Production

Obesity is one of the most obvious and commonly associated impacts of processed food consumption. However, stock prices of various companies are also being affected by the processed food industry, possibly through meat production levels. As such, we decided to embark on an investigation to analyse how stock prices are truly impacted by the processed food industry in the United States.

2.4.1 Choosing the right meats

To quantify the impact of the culling of meat production on the wider economy, we chose to focus on the stock prices of certain stocks. Analysing the various companies and what their main businesses are, we decided to focus our efforts on …. Drawing from a smaller sample size of stocks, we examined the correlation of the stock prices with the various meat production levels. Unsurprisingly, from Figure 6, we can see that the top three meats associated with these fast food companies across the board were Broilers, Pork and Beef, barring those chains that rely less on beef. As such, we narrowed the discussion down to these three meats.

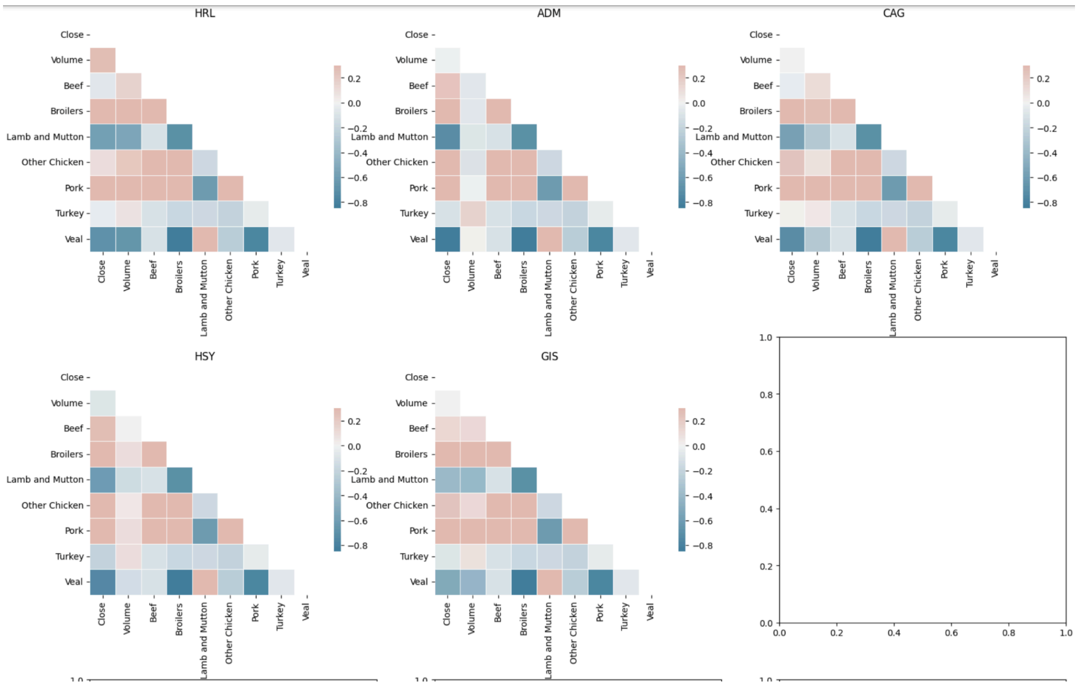


Figure 6: Correlation Between Selected Stocks and Production Levels of Various Meat

Additionally, plotting the time series of cold storage levels, as shown in Figure 7, further supports the argument – the meat cold storage levels of Broilers, Beef and Pork have consistently been the top three, with Turkey production levels leading close behind, albeit following a trend.

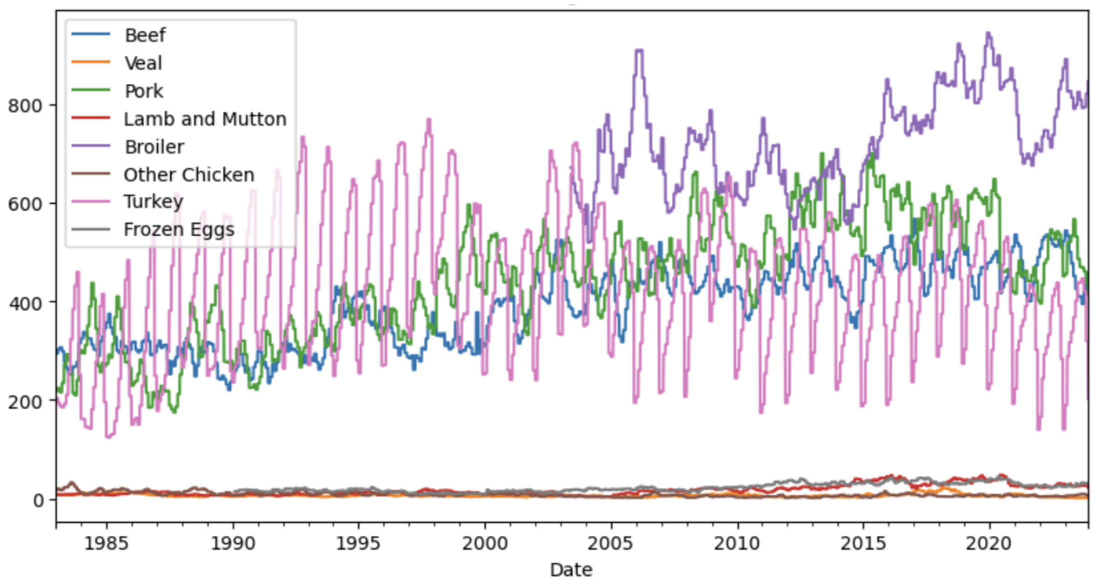


Figure 7: Time Series of Weight for All Animals

Finally, a notable outlier in the dataset that we decided to use is 'Frozen Eggs'. Driven by prior knowledge of the exorbitant use of frozen eggs by fast food companies, we further delved into this 'Animal' type and also found a strong rising trend over the years (Figure 8).

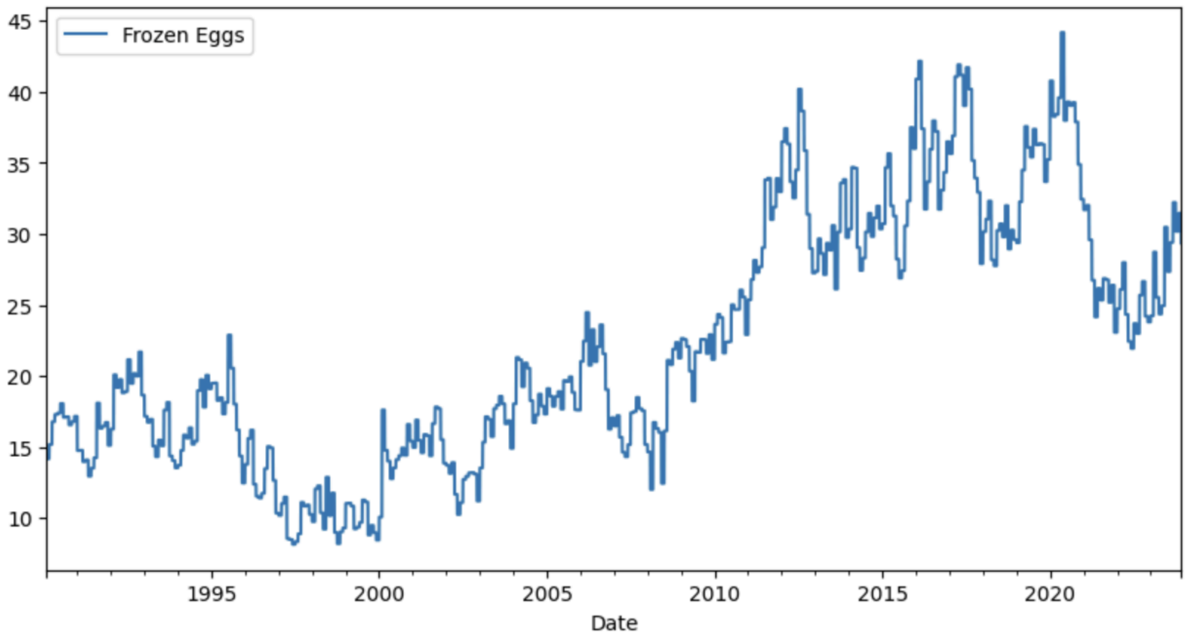


Figure 8: Time Series of Weight for Animal Type 'Frozen Eggs'

2.4.2 Understanding the relation between Cold Storage and Meat Production

Given this 'Cold Storage' dataset, it was outlined that the values could reveal insights into the supply and demand dynamics in the market. Our first instinct was to make these datasets more understandable; if cold storage levels of the same meat fell in any given month, the meat must have been sold; if cold storage levels of the same meat rose in any given month, there must have been overproduction. Therefore, we sought to uncover whether there were correlations between the meat production and cold storage datasets.

As seen from Figure 9, for **beef**, there was a moderate correlation of 0.458 between Production and Cold Storage Weight values. The correlation for **broilers** was slightly stronger at 0.553. Finally, for pork, this correlation was the strongest at 0.727.

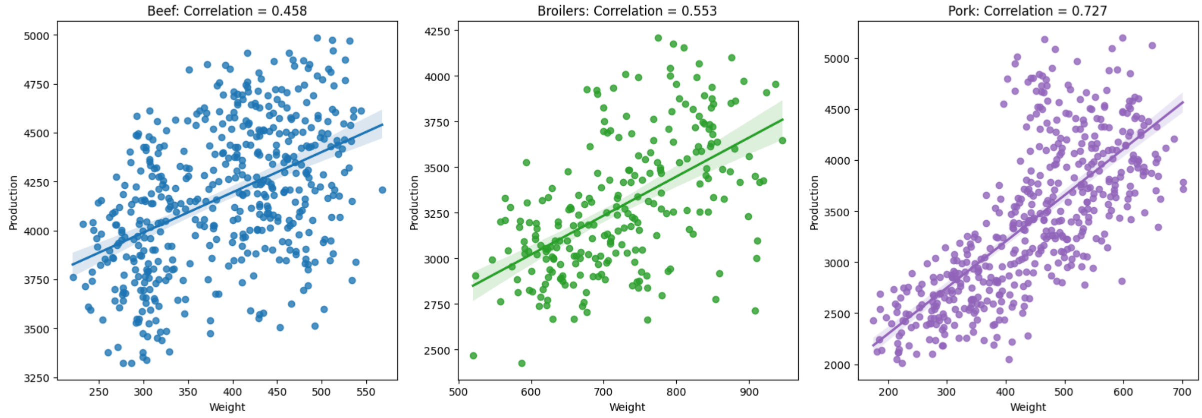


Figure 9: Correlation of Beef, Broilers and Pork Cold Storage Weight Versus Production

However, these relations are unsurprising – with higher production levels, there should be more meat in cold storage. Rather, given that the shelf life of these meats are roughly around 8 to 9 months, we computed a new column representing the increase in cold storage, which is likely closer to the true value of new meat added to the cold storage for the month.

This was calculated by the following formula:

$$Increase\ in\ Cold\ Storage = Cold\ Storage\ Level_{Latest\ Month} - (Cold\ Storage\ Level_{Previous\ Month} - Cold\ Storage\ Level_{Average\ of\ past\ 8\ months})$$

Then we plot the relationship between Meat Cold Storage Weight Increased and Meat Production

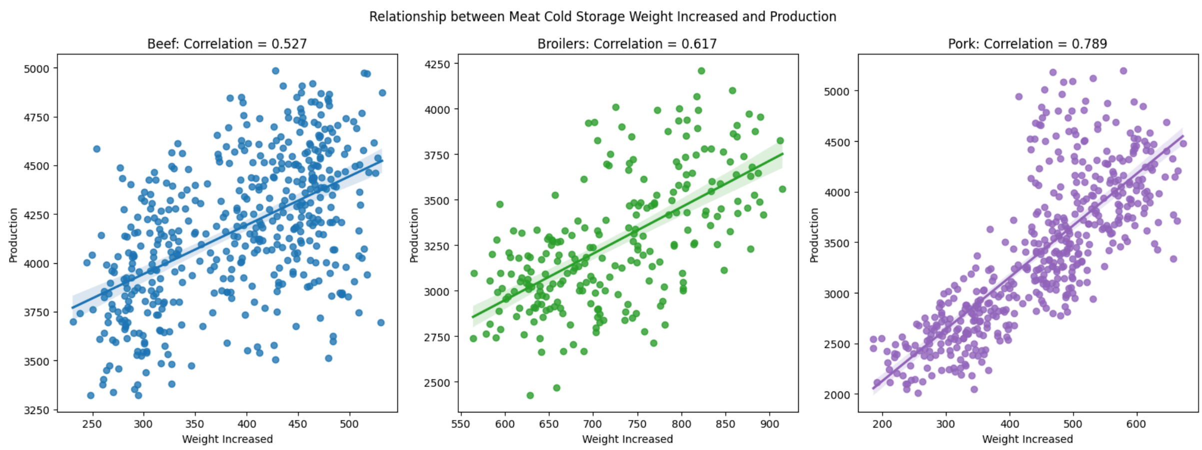


Figure 10: Correlation of beef, broilers and pork cold storage increased weight vs production

This improved the correlation measure significantly to 0.527, 0.617 and 0.789 for Beef, Broilers and Pork respectively.

Those results suggest the need for a new feature that represents the two factors, meat production and cold storage, to reduce dimensionality for the models. Therefore, we introduce a new feature called **pseudo demand.** Pseudo demand is calculated as follow:

$$Pseudo\ Demand = Meat\ Production - Increase\ in\ Cold\ Storage$$

2.4.3 Meat Production Levels as an extension of Slaughter Weights and Counts

As a sanity check, we computed a new level by Slaughter Weight * Slaughter Count to understand if Meat Production Levels = Slaughter Weight * Slaughter Count.

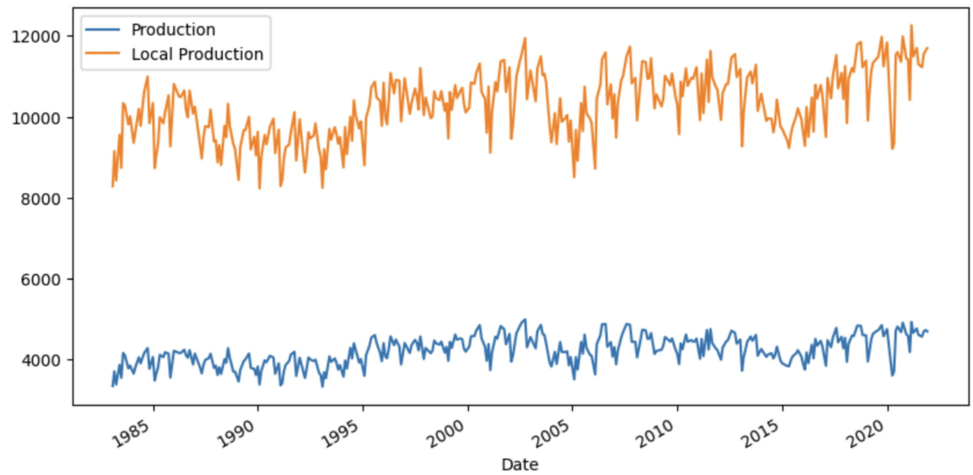


Figure 11: Time Series of Production and Local Production for Beef

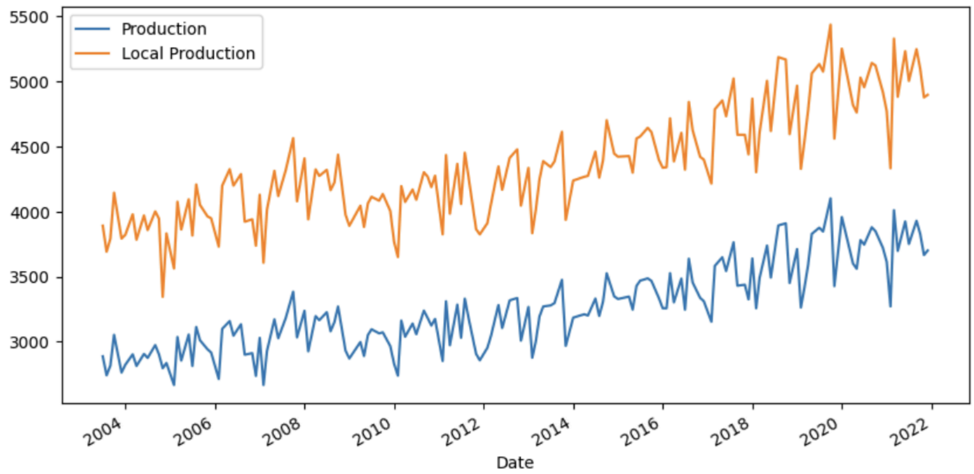


Figure 12: Time Series of Production and Local Production for Broilers

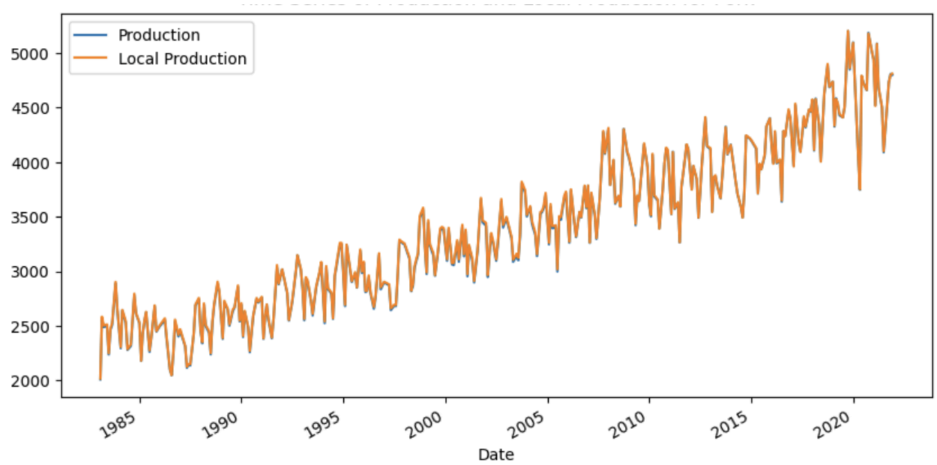


Figure 13: Time Series of Production and Local Production for Pork

Plotting the time series graphs (Figures 11, 12 and 13), the relationship becomes clearer. Even barring the difference in magnitudes between the computed 'Local Production' levels and 'Production' levels given, the ratio between the two is roughly constant, even 1 to 1 for pork. Hence, although 'slaughter weights' is likely to bring in new information, the further inclusion of slaughter counts may introduce multicollinearity in the model instead.

2.4.4 Understanding the relationship between meat production levels and stock prices

2.4.4.1 Calculate Market Capitalisation Estimator

Since the stock prices include various tickers, with a long time period (from 1999 to 2024), and different volumes & prices, we decided to calculate the market capitalization for each ticker base on a monthly basis, and use feature engineering to create an extensive indicator base on those capitalizations.

With the traditional method, to calculating the market capitalization of a stock (on a monthly basis), we calculated using the formula:

$$Market\ Cap\ =\ Share\ Price\ \times\ Outstanding\ Shares$$

However, we only have access to the Volume from the original dataset, which is the number of shares traded for that stock on that day, we switched our approach to the method below:

(i) Calculate Average Daily Volume for each month: The time frame we want to derive is on a monthly basis, hence we sum up the daily trading volumes for each trading day in the month and divide that by the number of trading days in that month. This can be calculated as:

$$Average\ Daily\ Volume = \frac{\sum_{i=1}^{number\ of\ trading\ days\ in\ month} trading\ volumes\ in\ day\ i}{number\ of\ trading\ days}$$

(ii) Estimate Turnover Ratio: Estimate the turnover ratio, which represents how many times the total shares outstanding have been traded during the month. This can be calculated as:

$$Turnover\ Ratio = \frac{Average\ Daily\ Volume}{Total\ Shares\ Outstanding\ in\ that\ month}$$

(iii) Estimate Total Shares Traded: Using the turnover ratio, estimate the total shares traded during the month. This can be calculated as:

$$Total\ Daily\ Shares\ Traded\ =\ Turnover\ Ratio\ \times\ Total\ Outstanding\ Shares$$

(iv) Calculate Market Capitalization Estimation Factor: Finally, we calculate the market cap estimation factor using the closing share price for the month. This can be calculated as:

$$Monthly\ Market\ Cap\ estimator\ =\ Closing\ Price\ of\ Month\ \times\ Total\ Shares\ Traded$$

2.4.4.2 Calculate Customised Index base on derived Market Capitalisation Estimator

(i) First strategy - Equal Weight:

In this first strategy, we build an "equal weight" portfolio as an estimator. This is constructed by taking every Ticker that is available at a certain timestamp (monthly) and compute the average market capitalisation of those stocks:

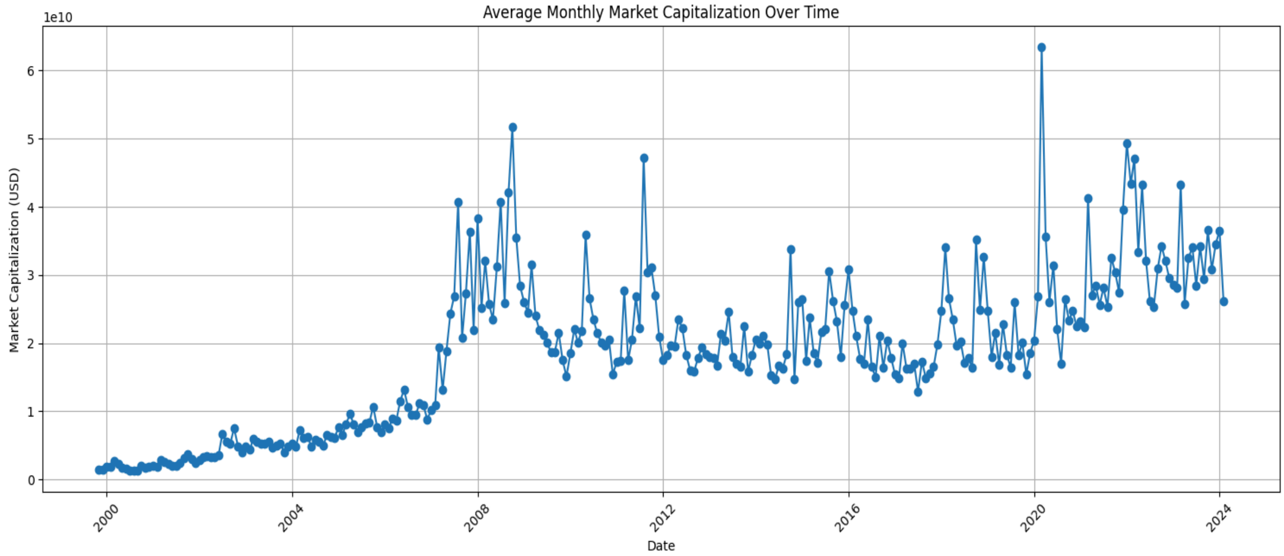


Figure 14: Graph of Average Monthly Market Capitalisation Over Time

We smoothen the plot in Figure 14 to Figure 15 for clearer demonstration.

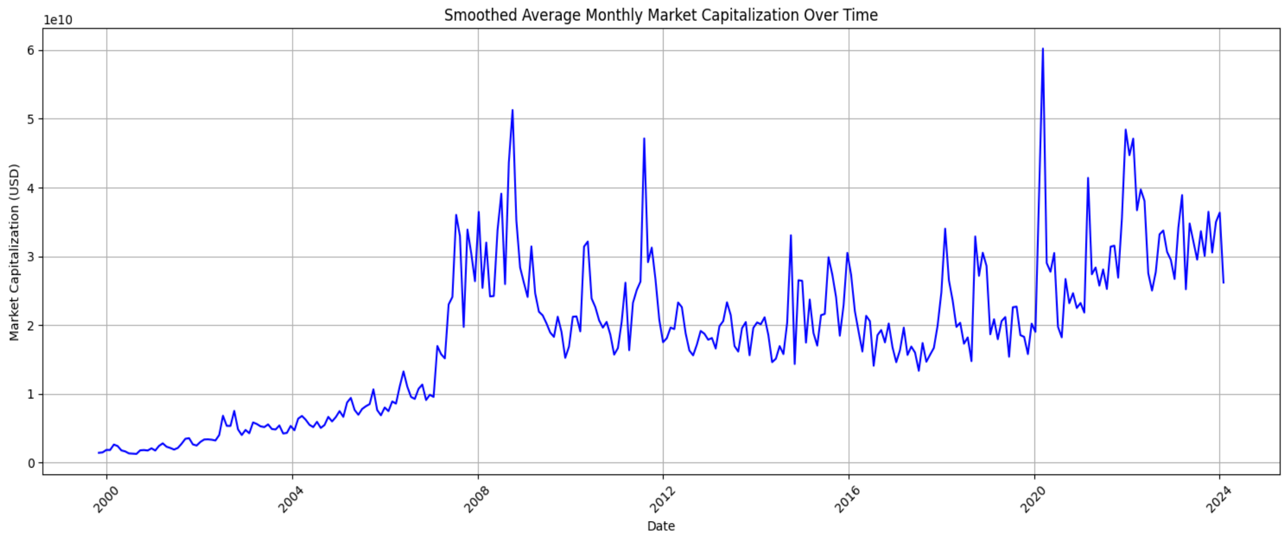


Figure 15: Smoothed Graph of Average Monthly Market Capitalisation Over Time

Our analysis of the constructed portfolio reveals several significant surges throughout the period. These include 2009, late 2011, and most notably, 2020. The emergence of new stocks, particularly in the meat production and cold storage sectors, could potentially explain these increases. We will explore this correlation in more detail later in this section.

(ii) Second Strategy - Rank by Market Capitalisation Estimator:

We decided to construct a portfolio that resembles the mechanism of the S&P 500 index: The Stock that has higher market capitalization is weighted more than those with lower market capitalization, which presents their dominance in market performance.

First, we normalise the stocks by calculate the weight assign to Stock i:

$$Weight_{i} = \frac{Market\ Capitalisation\ of\ Stock\ i}{\sum_{j=1}^{Number\ of\ Stock} Market\ Capitalisation\ of\ Stock\ j}$$

Next, we calculate the Weighted Market Capitalisation:

$$Weighted\ Market\ Capitalisation = \sum_{i=1}^{Number\ of\ Stocks} (Market\ Capitalisation\ of\ Stock\ i\ *\ Weight_{i})$$

This is demonstrated in the following plot:

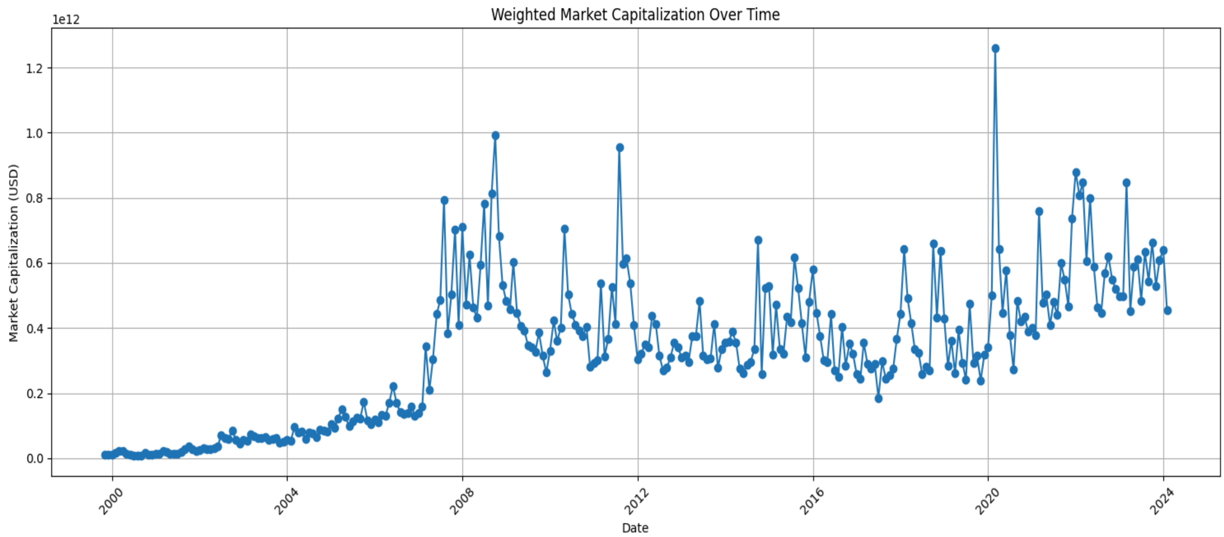


Figure 16: Graph of Weighted Market Capitalisation Over Time

Similarly,, we smooth Figure 16 out for clearer illustration, into Figure 17 as shown below.

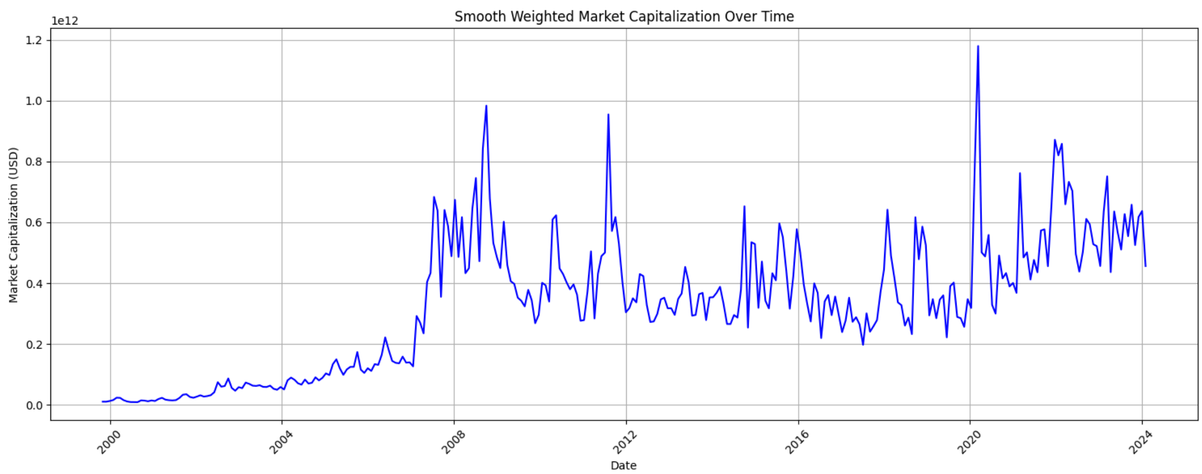


Figure 17: Smoothed Graph of Weighted Market Capitalisation Over Time

While the trend remains similar to the equally weighted strategy, this approach results in significantly higher and more volatile market capitalizations.

2.4.4.3 Factual derivation from processed food stocks:

Our analysis of the market capitalization for this portfolio reveals several significant surges throughout the period. The years 2009, late 2011, and most notably, 2020, stand out as periods of exceptional spark. By delving deeper into these surges and examining them in the context of major world events, we can gain valuable insights into the factors influencing the processed food industry.

The initial surge in 2009 coincides with the aftermath of a major global economic crisis [3]. This recession had a significant impact on the food industry, with disruptions to supply chains and rising costs of basic necessities. As consumers tightened their belts and prioritised essential spending, the processed food industry experienced a relative advantage. These readily available, often shelf-stable food options became more attractive due to their affordability and convenience. This explains the observed jump in the market capitalization of processed food companies during this challenging economic period.

Fast forward to late 2011, and we encounter another surge in market capitalization. While the specific triggers for this increase require further investigation, it's possible that other global events or industry-specific developments influenced consumer behaviour and investment decisions.

However, the most dramatic surge in market capitalization occurred in 2020, coinciding with the emergence of the COVID-19 pandemic. This global health crisis caused widespread economic damage, yet the processed food industry experienced an unprecedented boom. Several factors likely contributed to this phenomenon. Concerns about potential food shortages due to supply chain disruptions may have led to increased stockpiling of processed foods. Lockdowns and social distancing measures limited people's ability or willingness to visit grocery stores frequently. This shift in consumer behaviour led to a reliance on convenient, pre-packaged options offered by the processed food industry. Furthermore, with restaurants closed and access to fresh ingredients limited, many individuals resorted to readily available processed foods for ease and convenience during a time of social isolation and disruption. Hence, the pandemic triggered a food crisis marked by scarcity [4]. This, coupled with the aforementioned psychological factors, likely fueled an unprecedented demand for readily available processed food.

The observed correlation between these major world events and market capitalization surges in the processed food industry warrants further investigation. A deeper analysis can help solidify these connections, uncover any additional contributing factors, and provide valuable insights into the complex relationship between global events, consumer behaviour, and market forces within the food industry.

2.4.5 Forecasting Pseudo Demand for the various meats

Analysing the graphs for beef, pork and broilers separately, we discovered seasonality in the production schedules. The ‘Demand Per Pax’ seems to increase steadily and sharply during the summer before sharply dropping once winter comes around, as seen in Figures 18, 19 and 20. We postulate this could be associated with harsher conditions for animals and thus, overall lower yields of meat.

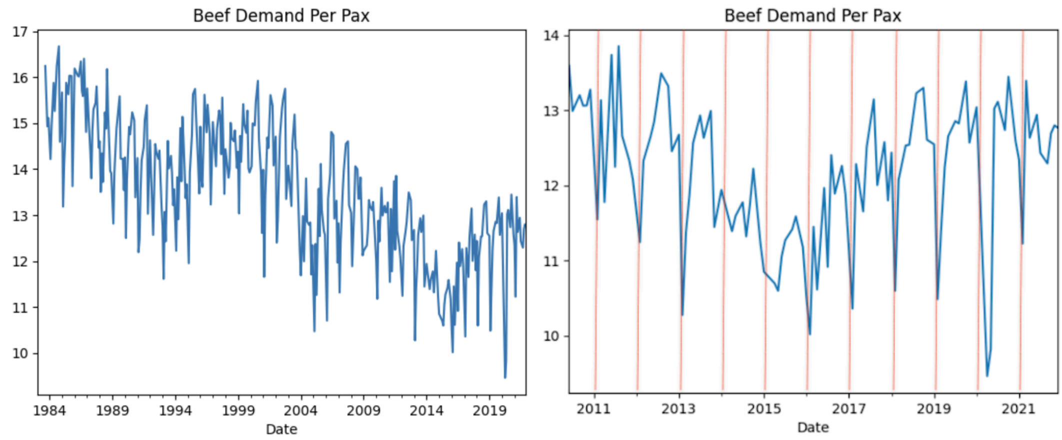


Figure 18: Current (Left) and Forecasted (Right) Pseudo Demand For Beef

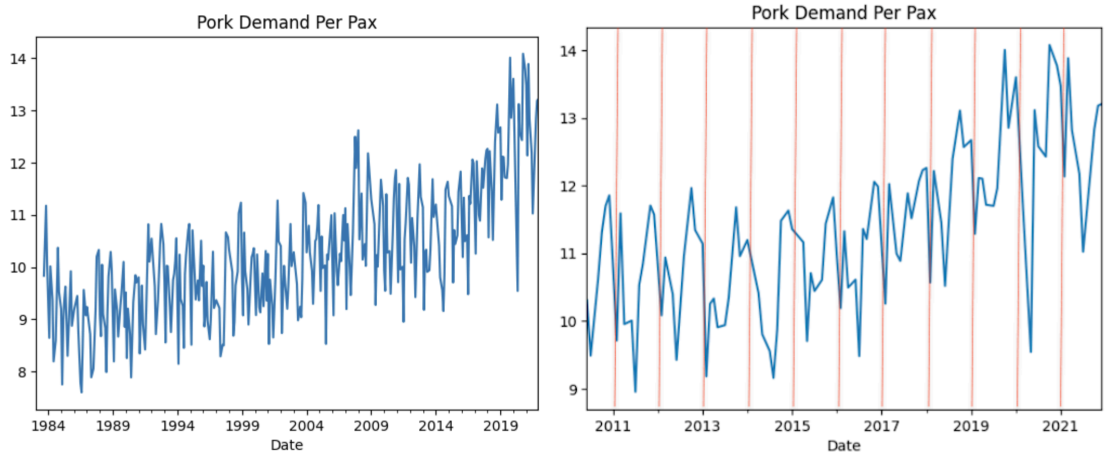


Figure 19: Current (Left) and Forecasted (Right) Pseudo Demand For Pork

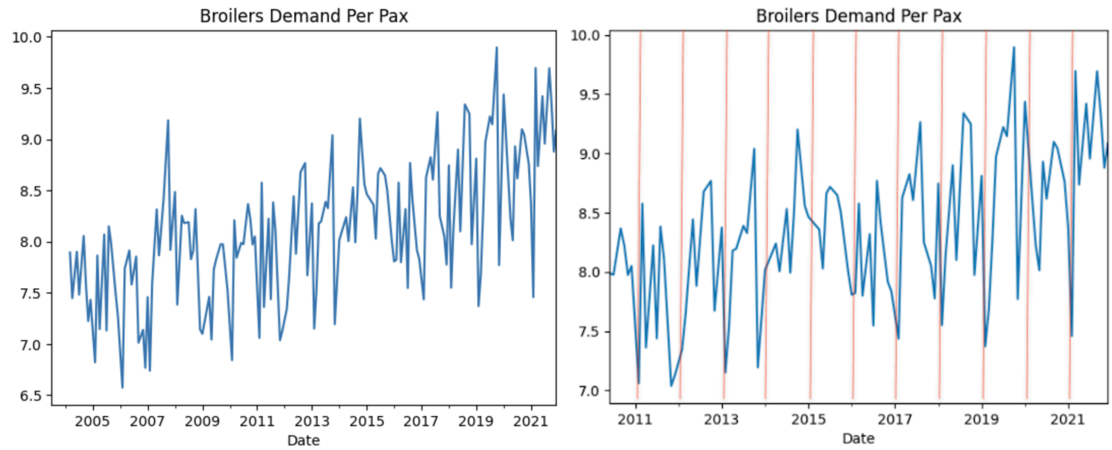


Figure 20: Current (Left) and Forecasted (Right) Pseudo Demand For Broilers

As such, we chose to model the demand per pax using SARIMAX that takes into account seasonal trends. Using the optimal weights for p, d, q, P, D, Q and m determined by pmd.auto_arima(), we trained the SARIMAX models for the individual meat types, achieving moderate success. The residuals of the models for all the meats seem to largely be normally distributed with predictions that align with the seasonality of the production levels as well.

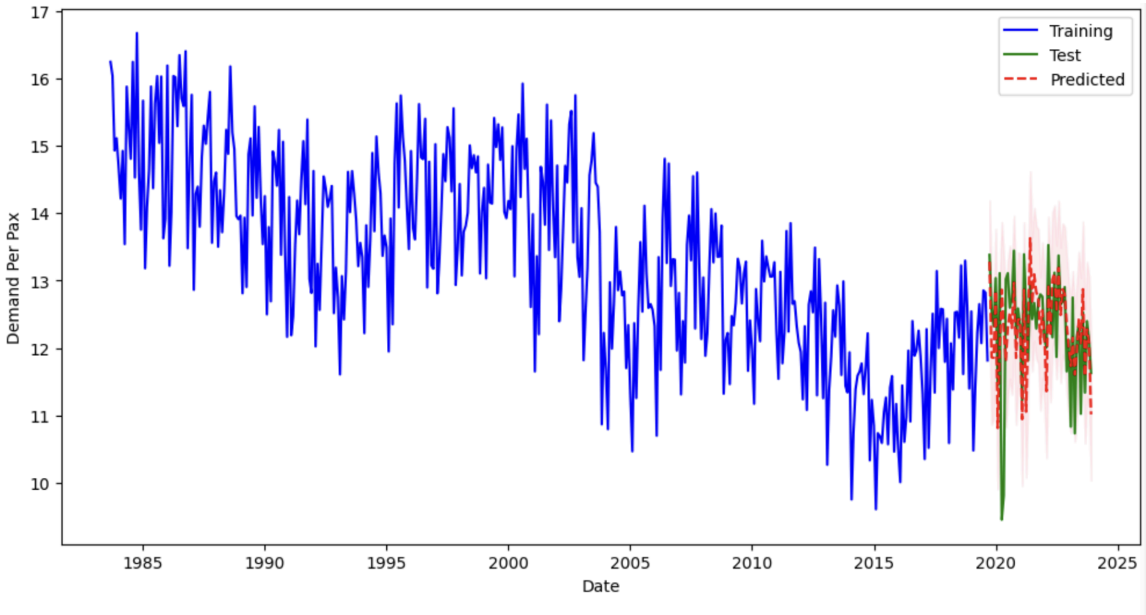


Figure 21: ARIMA Predictions of Demand Per Pax for Beef with Confidence Intervals

To observe the predictions of the test data clearer, Figure 22 shows the zoomed-in ARIMA Predictions of the period between 2020 and 2024.

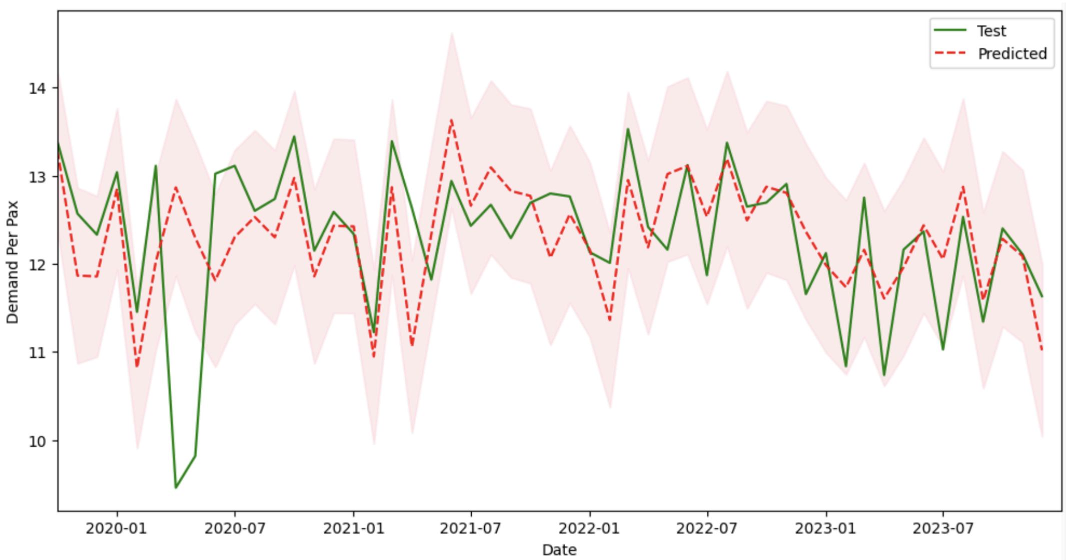


Figure 22: Zoomed-in ARIMA Predictions of Demand Per Pax for Beef between 2020 and 2024

SARIMAX Results

Dep. Variable:	demand_per_pax	**No. Observations:**	483
Model:	SARIMAX(2, 1, 4)x(2, 0, [1, 2], 12)	**Log Likelihood**	-576.562
Date:	Sun, 24 Mar 2024	**AIC**	1175.123
Time:	20:50:52	**BIC**	1221.103
Sample:	09-30-1983	**HQIC**	1193.192
	- 11-30-2023		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-1.1568	0.002	-765.670	0.000	-1.160	-1.154
ar.L2	-0.9976	0.002	-581.000	0.000	-1.001	-0.994
ma.L1	0.4544	0.039	11.698	0.000	0.378	0.531
ma.L2	0.1544	0.036	4.289	0.000	0.084	0.225
ma.L3	-0.7307	0.034	-21.274	0.000	-0.798	-0.663
ma.L4	-0.0368	0.036	-1.020	0.308	-0.108	0.034
ar.S.L12	0.1414	0.048	2.968	0.003	0.048	0.235
ar.S.L24	0.8581	0.047	18.267	0.000	0.766	0.950
ma.S.L12	0.1048	0.078	1.347	0.178	-0.048	0.257
ma.S.L24	-0.8175	0.065	-12.542	0.000	-0.945	-0.690
sigma2	0.2482	0.016	15.187	0.000	0.216	0.280

Ljung-Box (L1) (Q):	0.70	**Jarque-Bera (JB):**	411.84
Prob(Q):	0.40	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.59	**Skew:**	-0.88
Prob(H) (two-sided):	0.00	**Kurtosis:**	7.16

Figure 23: Statistical SARIMAX Results for Beef

Figures 24 and 25 are the graphs showing the ARIMA Predictions of Demand Per Pax for Pork.

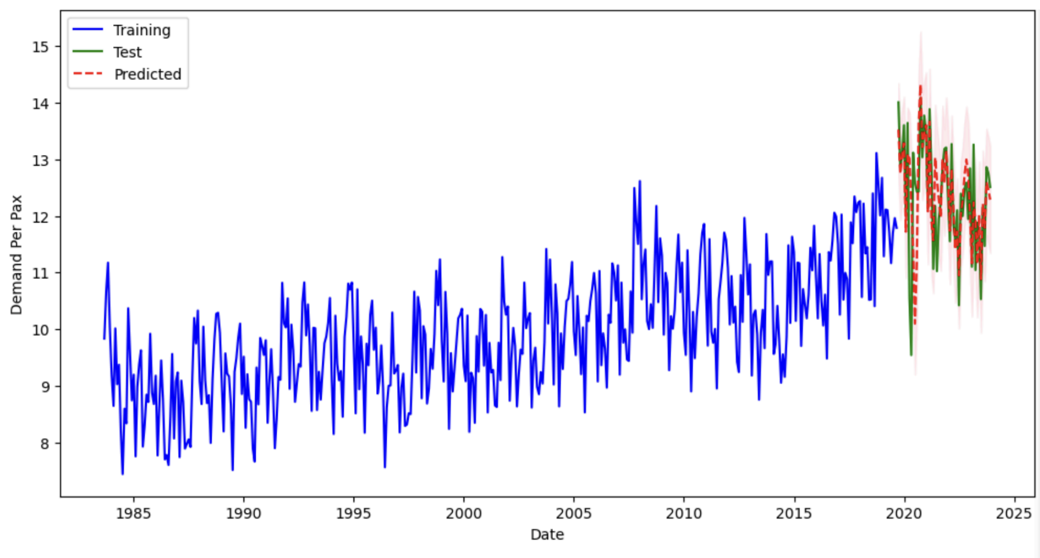


Figure 24: ARIMA Predictions of Demand Per Pax for Pork with Confidence Intervals

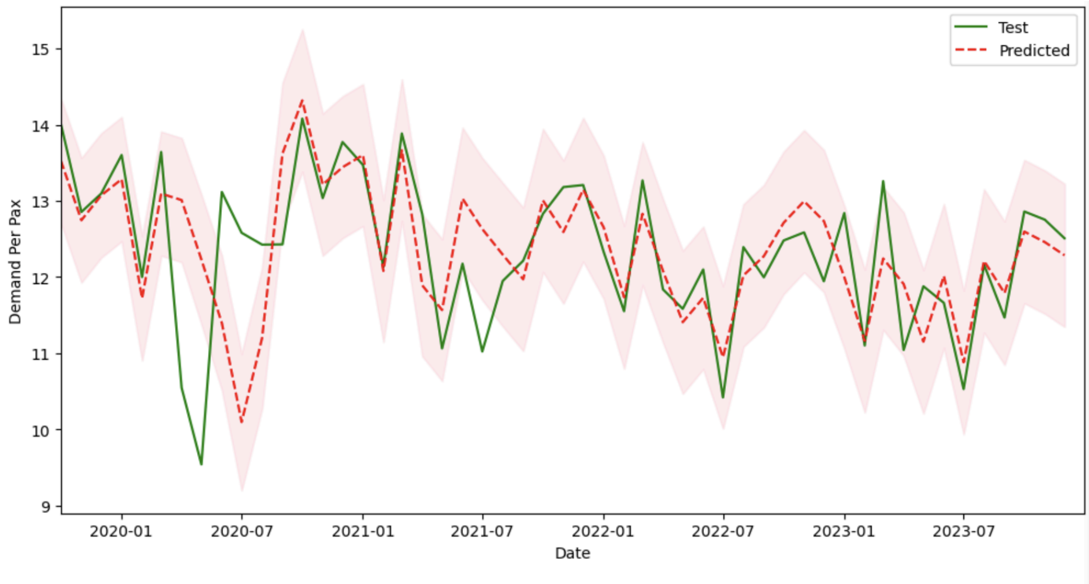


Figure 25: Zoomed-in ARIMA Predictions of Demand Per Pax for Pork between 2020 and 2024

SARIMAX Results

Dep. Variable:	demand_per_pax	**No. Observations:**	483
Model:	SARIMAX(2, 1, 1)x(2, 0, [1, 2], 12)	**Log Likelihood**	-527.486
Date:	Sun, 24 Mar 2024	**AIC**	1070.973
Time:	20:57:59	**BIC**	1104.413
Sample:	09-30-1983	**HQIC**	1084.114
	- 11-30-2023		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
ar.L1	-0.8031	0.049	-16.361	0.000	-0.899	-0.707
ar.L2	-0.5649	0.044	-12.790	0.000	-0.651	-0.478
ma.L1	-0.0407	0.057	-0.714	0.475	-0.152	0.071
ar.S.L12	0.7852	0.125	6.259	0.000	0.539	1.031
ar.S.L24	0.2094	0.124	1.683	0.092	-0.034	0.453
ma.S.L12	-0.4468	0.126	-3.559	0.000	-0.693	-0.201
ma.S.L24	-0.4200	0.110	-3.820	0.000	-0.635	-0.204
sigma2	0.2286	0.011	21.029	0.000	0.207	0.250

Ljung-Box (L1) (Q):	0.01	**Jarque-Bera (JB):**	346.64
Prob(Q):	0.90	**Prob(JB):**	0.00
Heteroskedasticity (H):	2.09	**Skew:**	-0.58
Prob(H) (two-sided):	0.00	**Kurtosis:**	6.99

Figure 26: Statistical SARIMAX Results for Pork

Figures 27 and 28 are the graphs showing the ARIMA Predictions of Demand Per Pax for Broilers.

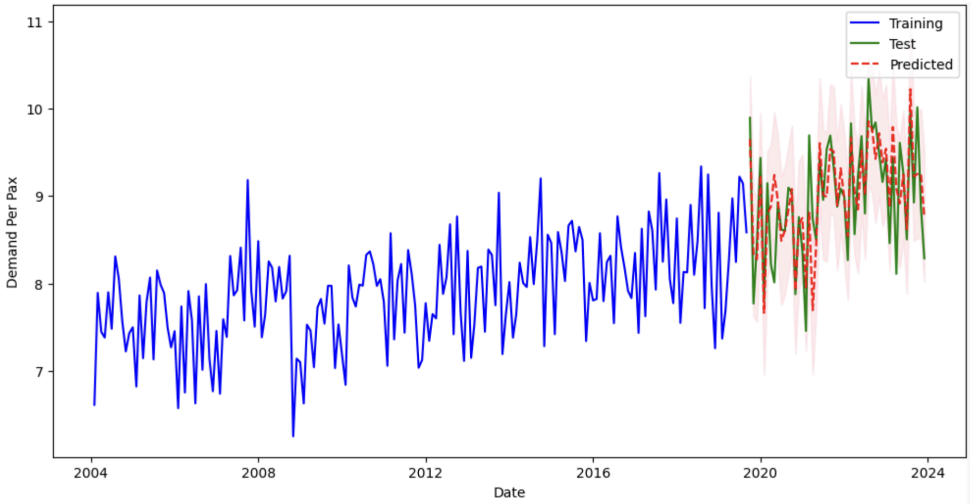


Figure 27: ARIMA Predictions of Demand Per Pax for Broilers with Confidence Intervals

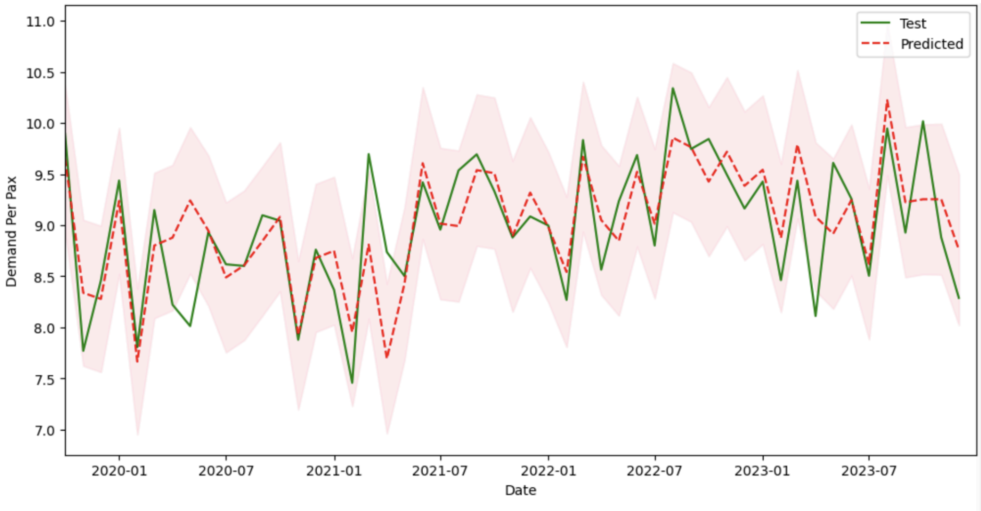


Figure 28: Zoomed-in ARIMA Predictions of Demand Per Pax for Broilers between 2020 and2024

SARIMAX Results

Dep. Variable:	demand_per_pax	**No. Observations:**	238
Model:	SARIMAX(3, 1, 0)x(2, 0, 0, 12)	**Log Likelihood**	-109.818
Date:	Sun, 24 Mar 2024	**AIC**	231.635
Time:	21:02:04	**BIC**	252.444
Sample:	02-29-2004	**HQIC**	240.023
	- 11-30-2023		
Covariance Type:	opg		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
ar.L1	-0.8608	0.063	-13.710	0.000	-0.984	-0.738
ar.L2	-0.5523	0.086	-6.423	0.000	-0.721	-0.384
ar.L3	0.1348	0.069	1.960	0.050	3.42e-05	0.270
ar.S.L12	0.6063	0.066	9.245	0.000	0.478	0.735
ar.S.L24	0.1415	0.074	1.914	0.056	-0.003	0.286
sigma2	0.1415	0.012	11.328	0.000	0.117	0.166

Ljung-Box (L1) (Q):	0.04	**Jarque-Bera (JB):**	2.52
Prob(Q):	0.85	**Prob(JB):**	0.28
Heteroskedasticity (H):	1.19	**Skew:**	-0.15
Prob(H) (two-sided):	0.45	**Kurtosis:**	3.40

Figure 29: Statistical SARIMAX Results for Broilers

In particular, the MAPE (Mean Absolute Percentage Errors) of the models were 4.71%, 4.91% and 3.61% for Beef, Pork and Broilers respectively. Although there were notable periods whereby the meat production levels exceeded the confidence intervals for the three meats, this period represents a common anomaly – COVID-19, which caused meat production levels to fall dramatically in a short amount of time, regardless of the type of meat.

Forecasting demand per pax until 2030, we find a rising trend for all three types of meat, with wider confidence intervals for Beef and Broilers (Figures 30, 31 and 32).

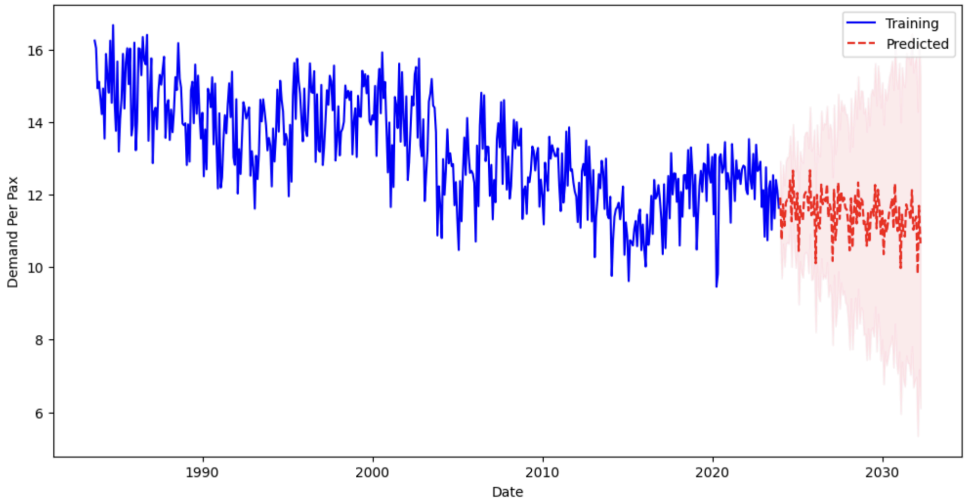


Figure 30: ARIMA Predictions of Demand Per Pax for Beef Until 2030

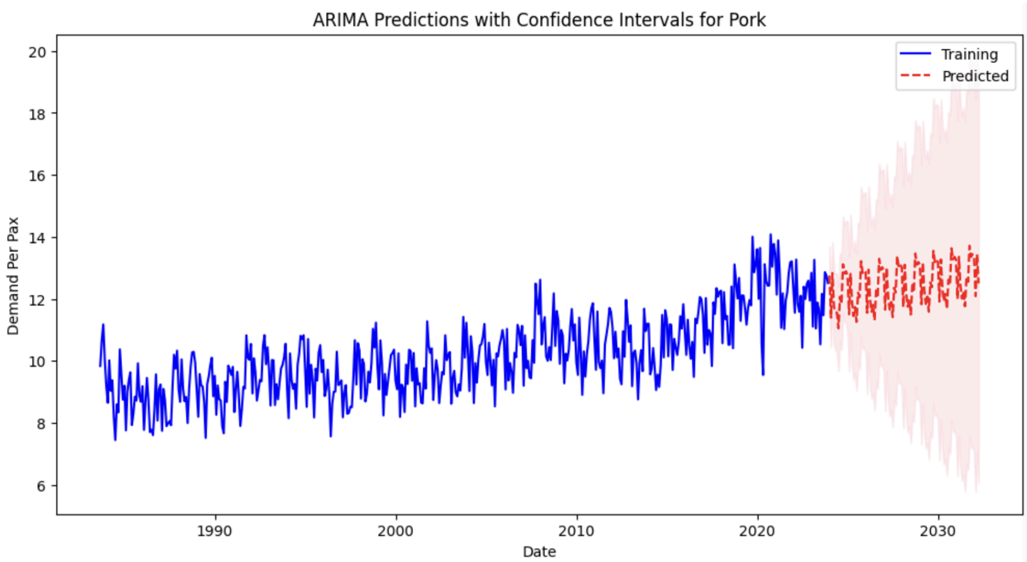


Figure 31: ARIMA Predictions of Demand Per Pax for Pork Until 2030

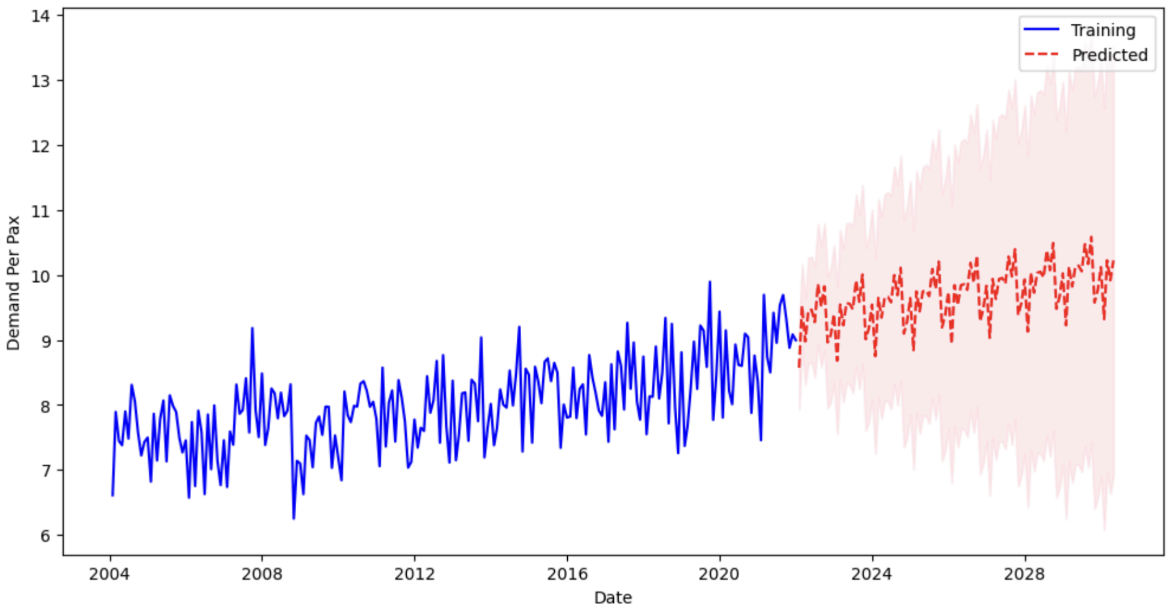
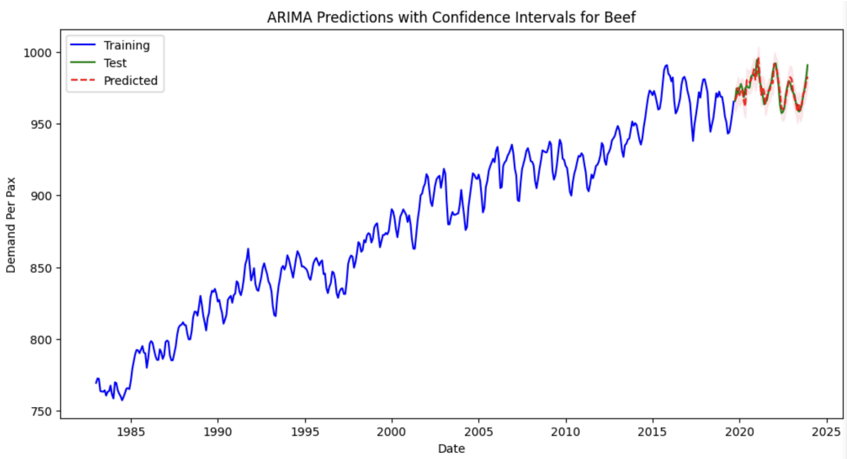


Figure 32: ARIMA Predictions of Demand Per Pax for Broilers Until 2030

2.4.6 Forecasting Slaughter Weights for the Various Meats

Once again, while analysing the graphs for beef, pork and broilers separately, seasonality in the slaughter weights is observed, even more prominent than the seasonality in the previous section. However, the seasonality observed among these graphs of slaughter weights is exactly inverse that of the previous graphs. This is a curious phenomenon - for which deeper research offers a variety of reasons for. Firstly, farmers usually feed animals more during the winter to stave off sickness [9]. Furthermore, higher daily maximum and minimum temperatures (during the summer) was also shown to lead to lower carcass weights and poorer growth [1]. Finally, evaporation of water from the carcass during the summer further reduces the carcass weights during summer . Overall, this results in a relatively predictable cycle of peaks and troughs that are very nicely outlined by the SARIMAX model. The following graphs represent the relatively successful results from our modelling.



SARIMAX Results

Dep. Variable:	Weight	**No. Observations:**	491
Model:	SARIMAX(1, 1, 2)x(2, 0, [1], 12)	**Log Likelihood**	-1384.912
Date:	Sun, 24 Mar 2024	**AIC**	2783.824
Time:	21:16:05	**BIC**	2813.185
Sample:	01-31-1983	**HQIC**	2795.355
	- 11-30-2023		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.8995	0.034	26.395	0.000	0.833	0.966
ma.L1	-0.6928	0.051	-13.466	0.000	-0.794	-0.592
ma.L2	-0.2782	0.043	-6.398	0.000	-0.363	-0.193
ar.S.L12	1.1399	0.078	14.635	0.000	0.987	1.293
ar.S.L24	-0.1667	0.069	-2.402	0.016	-0.303	-0.031
ma.S.L12	-0.8015	0.055	-14.494	0.000	-0.910	-0.693
sigma2	16.3389	1.031	15.848	0.000	14.318	18.360

Ljung-Box (L1) (Q):	0.01	**Jarque-Bera (JB):**	3.27
Prob(Q):	0.91	**Prob(JB):**	0.20
Heteroskedasticity (H):	0.88	**Skew:**	-0.19
Prob(H) (two-sided):	0.41	**Kurtosis:**	3.14

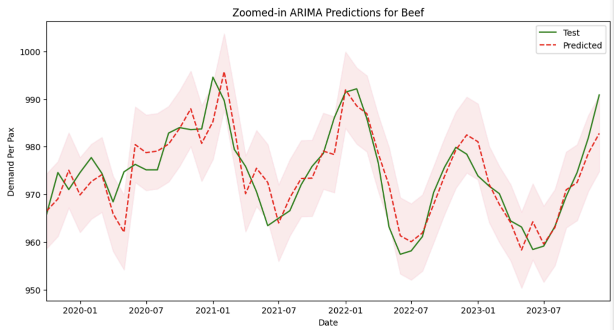
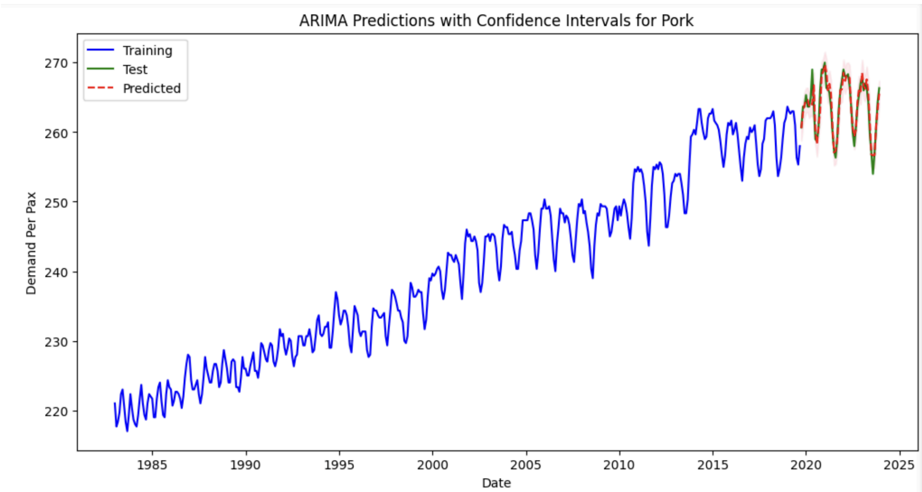


Figure 33: ARIMA Predictions of Slaughter Weights for Beef Until 2025



SARIMAX Results

Dep. Variable:	Weight	**No. Observations:**	491
Model:	SARIMAX(1, 1, 2)x(1, 0, [1], 12)	**Log Likelihood**	-678.509
Date:	Sun, 24 Mar 2024	**AIC**	1369.018
Time:	21:26:54	**BIC**	1394.184
Sample:	01-31-1983	**HQIC**	1378.901
	- 11-30-2023		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.8662	0.029	30.005	0.000	0.810	0.923
ma.L1	-0.8317	0.037	-22.414	0.000	-0.904	-0.759
ma.L2	-0.1500	0.038	-3.932	0.000	-0.225	-0.075
ar.S.L12	0.9914	0.005	194.325	0.000	0.981	1.001
ma.S.L12	-0.7554	0.037	-20.522	0.000	-0.828	-0.683
sigma2	0.8866	0.042	21.250	0.000	0.805	0.968

Ljung-Box (L1) (Q):	0.03	**Jarque-Bera (JB):**	67.56
Prob(Q):	0.86	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.50	**Skew:**	0.24
Prob(H) (two-sided):	0.01	**Kurtosis:**	4.75

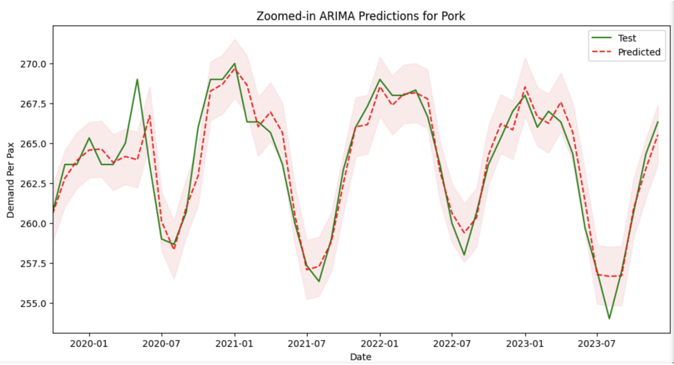
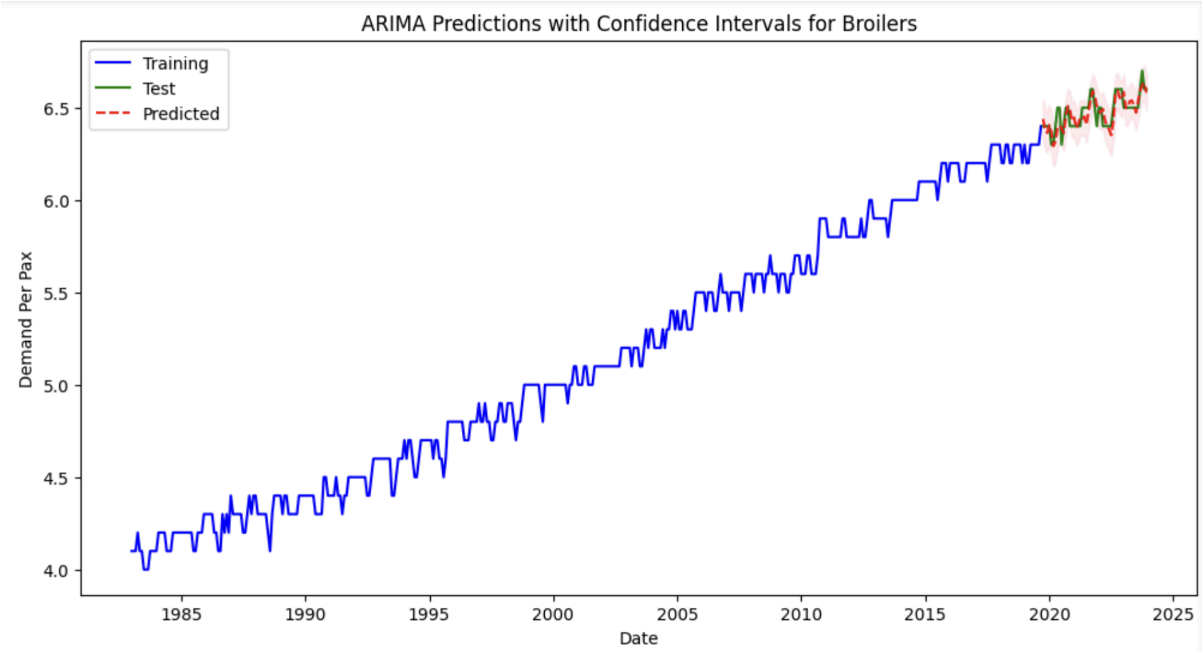
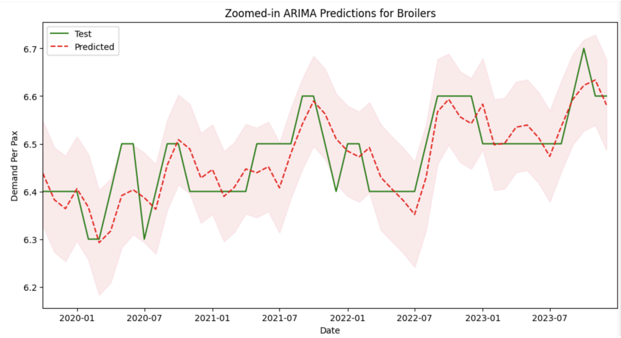


Figure 34: ARIMA Predictions of Slaughter Weights for Pork Until 2025





SARIMAX Results

Dep. Variable:	Weight	**No. Observations:**	491
Model:	SARIMAX(1, 1, 1)x(1, 0, 1, 12)	**Log Likelihood**	780.376
Date:	Sun, 24 Mar 2024	**AIC**	-1550.752
Time:	21:34:21	**BIC**	-1529.780
Sample:	01-31-1983	**HQIC**	-1542.516
	- 11-30-2023		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.3384	0.054	6.230	0.000	0.232	0.445
ma.L1	-0.8769	0.031	-28.616	0.000	-0.937	-0.817
ar.S.L12	0.9917	0.005	181.811	0.000	0.981	1.002
ma.S.L12	-0.8796	0.033	-26.791	0.000	-0.944	-0.815
sigma2	0.0023	0.000	14.302	0.000	0.002	0.003

Ljung-Box (L1) (Q):	0.48	**Jarque-Bera (JB):**	0.94
Prob(Q):	0.49	**Prob(JB):**	0.63
Heteroskedasticity (H):	0.92	**Skew:**	-0.08
Prob(H) (two-sided):	0.61	**Kurtosis:**	2.85

Figure 33: ARIMA Predictions of Slaughter Weights for Broilers Until 2025

Given these slaughter weights follow a more predictable pattern, we observe the SARIMAX models here excelling, with MAPE of 0.36%, 0.35% and 0.65% for Beef, Pork and Broilers respectively. This gives us greater confidence in using the model in predicting future slaughter weights, as we do in the following part from Figures 34, 35 and 36 to predict slaughter weights until roughly 2032.

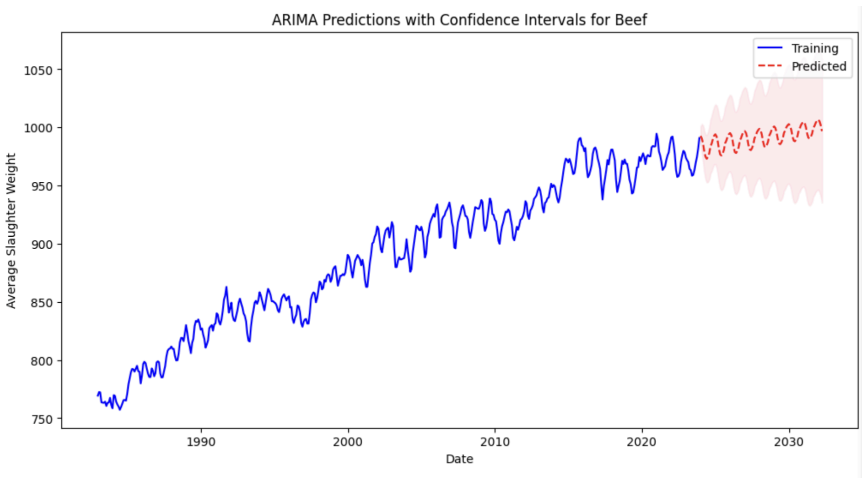


Figure 34: ARIMA Predictions of Slaughter Weights for Beef Until 2032

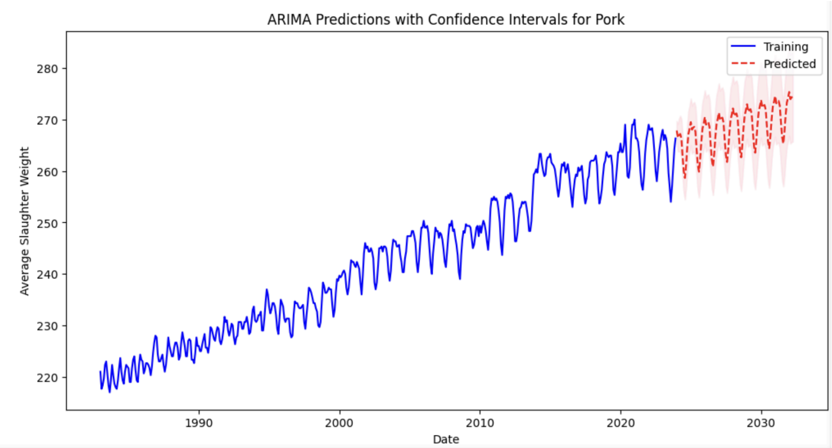


Figure 35: ARIMA Predictions of Slaughter Weights for Pork Until 2032

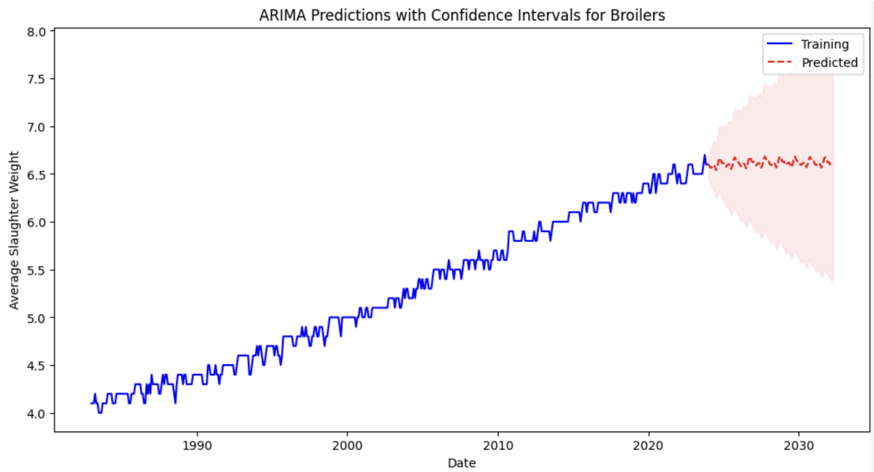


Figure 36: ARIMA Predictions of Slaughter Weights for Broilers Until 2032

2.4.7 Forecasting Market Capitalization

Given the aforementioned discovered relations between the market capitalisations and meat production levels, forecasting the market capitalisation levels of the various companies represents an extension of the forecast of the meat production levels and slaughter weights. To fully understand the strength of predictions of the various factors, we employed a gradient boosting model. As we are dealing with time series data, we trained and validated the model using time series split technique to ensure no data leakage. With a model achieving validated Mean Squared Log Error of 0.16, we forecasted the market capitalization of preprocessed food companies as shown in Figure 37.

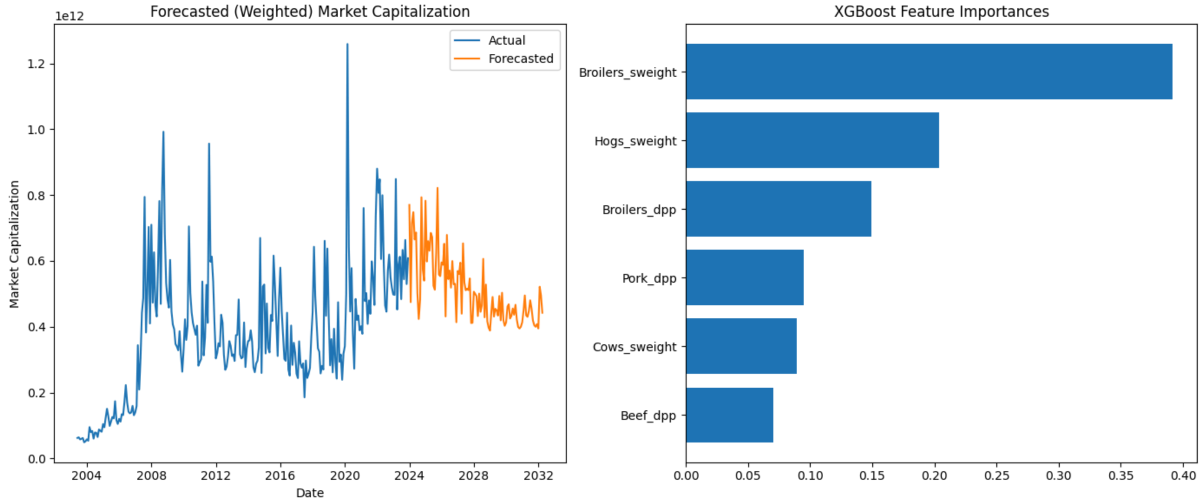


Figure 37: Graph showing Forecasted Market Capitalisation and Relative Feature Importance of XGBoost Model

Rather unsurprisingly, the slaughter weights of broilers play the biggest role in determining prices. Following that, the forecasted market capitalisation of these fast food companies does not look too hopeful.

2.6 The Effects of Altering Meat Production

We can derive two key insights from this research - which meat to focus our efforts on and the given intervention methods.

Firstly, given these trends into the future, a reduction in meat production levels for pork and beef could prove to be less detrimental to the strength of these fast food companies than meat production levels for broilers since the meat production level of broilers are forecasted to stagnate for the near future. This would be important for ensuring that these fast food companies, which represent a lifeline for many Americans in terms of their livelihoods, are not excessively affected.

Secondly, a potential policy that the data points towards would be limiting the production levels of certain meats in the future without affecting current production levels. For meat production levels, these could be limiting levels for Broilers to 10 pounds per resident of the United States to avoid overproduction in the future. For slaughter weights, these could be limiting the weights of carcasses of beef to under 1000 pounds and pork to under around 270 pounds. In particular, Beef and Pork production levels are not forecasted to grow significantly while broiler weights face a similar predicament, limiting potential successes of such a policy.

3 Conclusion

In conclusion, our comprehensive analysis delves into the intricate relationship between the processed food industry, obesity rates, and the economy in the United States. Through various methodologies including SARIMAX models, feature engineering, and financial modelling, we have derived significant findings that underscore the urgent need for attention and action.

Firstly, our investigation into obesity trends reveals a troubling upward trajectory, closely correlated with the growth in the meat industry. This observation highlights the critical importance of addressing the excessive consumption of processed foods for the overall health and well-being of the population.

Moreover, our examination of market capitalizations of key processed food companies unveils a concerning pattern of spikes during economic downturns, indicating the industry's resilience and profitability even during challenging times. This underscores the economic influence and significance of the processed food sector, further complicating efforts to implement substantive measures to curb consumption.

Additionally, our analysis of meat production and stock prices reveals seasonal fluctuations and complex interdependencies within the industry, shedding light on the intricate dynamics at play.

It is imperative for policymakers to take decisive action to limit the continued excessive growth of the processed food industry. Targeted interventions aimed at promoting healthier dietary habits and mitigating the economic incentives driving the industry's expansion are urgently needed to address the intertwined challenges of obesity and economic dependency.

Overall, our study provides valuable insights into the multifaceted nature of the processed food industry and its impacts on health and the economy, paving the way for informed decision-making and policy formulation aimed at fostering a healthier and more sustainable future for all Americans.

References

[1] Bunning H., & Wall. E. (2022). The effect of weather on beef carcass and growth traits. Retrieved from https://www.sciencedirect.com/science/article/pii/S1751731122002142

[2] Jeannine P. Schweihofer, M. S. U. E. (2022). Carcass dressing percentage and cooler shrink. Retrieved from https://www.canr.msu.edu/news/carcass_dressing_percentage_and_cooler_shrink

[3] Jenkins, R. H., Vamos, E. P., Taylor-Robinson, D., Millett, C., & Laverty, A. A. (2021). Impacts of the 2008 Great Recession on dietary intake: A systematic review and meta-analysis. International Journal of Behavioral Nutrition and Physical Activity, 18(57). https://doi.org/10.1186/s12966-021-01125-8

[4] Kakaei, H., Nourmoradi, H., Bakhtiyari, S., Jalilian, M., & Mirzaei, A. (2022). Effect of COVID-19 on food security, hunger, and food crisis. COVID-19 and the Sustainable Development Goals, 3–29. https://doi.org/10.1016/B978-0-323-91307-2.00005-5

[5] Mande, J. (2023, May 23). Processed foods are making us sick. it's time for the FDA and USDA to step in. Harvard Public Health Magazine. https://harvardpublichealth.org/policy-practice/processed-foods-make-us-sick-its-time-for-government-action/

[6] New Media Retailer. (2024). Winter feeding and weight management. Retrieved from https://familyfarmandgarden.com/blog/75199/winter-feeding-and-weight-management

[7] OECD. (2021, January 11). Making Better Policies for Food Systems. 6. The contribution of the processed food sector to the triple challenge | Making Better Policies for Food Systems | OECD iLibrary. https://www.oecd-ilibrary.org/sites/15ae7a3c-en/index.html?itemId=%2Fcontent%2Fcomponent%2F15ae7a3c-en

[8] Pomeranz, J. L., Mande, J. R., & Mozaffarian, D. (2023, July 13). U.S. policies addressing ultraprocessed foods, 1980-2022. American journal of preventive medicine. https://pubmed.ncbi.nlm.nih.gov/37451324/

[9] World Obesity. (n.d.). Obesity classification. Obesity Classification. https://www.worldobesity.org/about/about-obesity/obesity-classification