Beef Pricing Dynamics in the Canadian Grocery Market: A Comparative Analysis of Walmart and T&T*

Evidence of Cultural and Strategic Influences on Beef Pricing

Ziheng Zhong

November 28, 2024

This study analyzes beef pricing dynamics in the Canadian retail market, focusing on two vendors: Walmart and T&T. Using Bayesian linear regression, we found that historical pricing and vendor-specific strategies significantly influence current beef prices, with Walmart displaying more price stability and T&T showing greater variability. These findings reveal how both market-wide and vendor-specific factors shape beef pricing, offering insights that could help retailers refine their pricing strategies and support policymakers in addressing food affordability. This paper provides a clearer understanding of the interplay between cultural preferences and retail strategies in determining food prices.

Table of contents

*Code and data supporting this analysis is available at: Link to repository.

1 Introduction

The pricing of grocery items, particularly beef, plays an important role in understanding consumer behavior and market trends within the Canadian retail sector. As a staple in many households, beef pricing is shaped by various factors, including vendor strategies (McGoldrick and Marks 1986), historical trends, and seasonal variations (Soysal and Krishnamurthi 2012). This study focuses on examining how major retailers adjust beef prices over time and how these adjustments reflect broader economic and cultural patterns. With growing concerns around food affordability and access, analyzing beef pricing dynamics is essential for both consumers and stakeholders involved in food distribution and policy-making.

This paper centers on two significant vendors, Walmart and T&T, to investigate how beef pricing differs across retail environments. Walmart, a global retailer, exemplifies standardized pricing strategies (Collins, n.d.), while T&T, catering primarily to Asian communities, represents a more targeted market segment (Hii 2007). By analyzing data from these vendors, the study addresses an often-overlooked aspect of how cultural preferences and retail strategies influence pricing. Previous research has focused on general trends in food pricing but has not fully considered the distinctions between mainstream retail chains and culturally specific stores. This analysis provides a detailed comparison of beef pricing patterns in these two diverse settings.

The estimand of this study is the current price of beef, determined by factors such as vendor type, historical pricing, etc. By estimating how these variables impact the current pricing, the paper aims to provide a comprehensive understanding of the pricing mechanisms that affect retail beef prices in Canada.

Using a Bayesian linear regression model, the study examines how historical pricing, vendor-specific strategies, and seasonal factors impact current beef prices. The analysis indicates that past prices strongly predict current pricing, suggesting a high degree of consistency. Differences in pricing strategies between Walmart and T&T further highlight distinct market approaches, with Walmart focusing on stable pricing and T&T displaying greater variability, potentially to address specific consumer preferences. These results provide a clearer understanding of pricing strategies in a diverse retail landscape and offer insights into factors that shape affordability and market behavior.

The paper is structured as follows: Section 2 describes the data collection and cleaning methods, as well as the outcome and predictor variables used in the analysis. Section 3 introduces the forecasting models and explains their selection for predicting beef prices. Section 4 presents the key findings, including vendor-specific pricing effects and seasonal trends. Lastly, Section 5 interprets these results, comparing vendor strategies, identifying significant patterns, and discussing the study's limitations.

2 Data

This project is motivated and guided by Rohan Alexander and his book (Alexander 2023). Data used in this paper was cleaned, analyzed and modeled with the programming language R (R Core Team 2023). Also with support of additional packages in R: `readr` (Wickham, Hadley and others 2023b), `ggplot2` (Wickham, Hadley 2023a), `tidyverse` (Wickham, Hadley and Bryan, Jennifer and others 2023), `dplyr` (Wickham, Hadley and others 2023a), `here` (Müller, Kirill and Bryan, Jennifer 2023), `knitr` (Xie, Yihui 2023a), `kableExtra` (Zhu, Hao 2023), `rstanarm` (Goodrich, Ben and Gabry, Jonah and Ali, Iram and Brilleman, Sam 2023), `modelsummary` (Arel-Bundock, Vincent 2023), `lme4` (Bates, Douglas and Mächler, Martin and Bolker, Ben and Walker, Steve 2023), `tinytex` (Xie, Yihui 2023b), `reshape2` (Wickham, Hadley 2023b), `arrow` (Neal, Weston and Urbanek, Simon and others 2023).

Details about data cleaning and variable selection can be found in Appendix A.

2.1 Source

The dataset for this research was obtained from Hammer, a publicly accessible repository offering pricing information from a variety of retail chains. It includes details on product attributes such as name, vendor, current and old prices, available units, and the month associated with certain prices. The dataset focuses on consumer pricing trends across different vendors, enabling an analysis of how factors like vendor type, historical pricing, and seasonal changes affect current prices. This provides a basis for examining retail market behaviors and assessing vendor pricing strategies.

The dataset contains both categorical and numerical variables. Important variables include 'vendor,' which identifies the retailer (e.g., Walmart, Galleria), and 'product_name,' which specifies the item being analyzed. Variables like 'current_price' and 'old_price' help track pricing changes, while temporal variables such as 'month' allow for the identification of seasonal pricing patterns. Summary statistics and visualizations have been generated to illustrate the distributions and relationships of these variables. These graphs include frequency distributions for vendors, trends in price changes, and comparisons between old and current prices. Relationships such as those between 'vendor' and 'current_price' are highlighted to present a detailed view of the dataset.

Alternative datasets, such as proprietary retail databases or other consumer purchasing records, were considered for this analysis. However, these were not selected due to limitations in accessibility, licensing restrictions, or insufficient detail in product-level attributes. The Hammer dataset was chosen because it provides detailed pricing data necessary for analyzing product-level trends across various vendors. It also enabled the creation of derived variables, such as 'price_per_unit,' which facilitated more meaningful comparisons between products. Data cleaning efforts included addressing missing values in the 'old_price' variable by replacing them with the ‘current_price,’ ensuring consistency and completeness in the analysis.

2.2 Measurement

The data for Project Hammer, focused on historical grocery prices, does not explicitly describe its collection methods. However, automated web scraping is a feasible approach for gathering this information. This method extracts structured information from retailer websites, capturing details such as product names, prices, brands, and unit measurements. The process involves developing scripts or utilizing tools that navigate websites, identify relevant HTML elements, and store the data in structured formats like CSV files or databases. To maintain accuracy and consistency, scraping schedules can be automated at regular intervals, such as weekly or monthly, to monitor price changes over time. Moreover, including user-agent strings and introducing delays in the scraping process helps simulate human browsing behavior and reduce the risk of triggering website anti-scraping mechanisms.

This research translates real-world phenomena into structured data entries within the dataset to analyze pricing behavior effectively. Variables such as ‘vendor,’ ‘current_price,’ and ‘month’ capture consumer purchasing patterns and their relationships with retail environments. The ‘vendor’ variable identifies the retailer, allowing an analysis of how retail settings influence pricing strategies. The ‘current_price’ and ‘old_price’ variables document pricing trends, enabling the measurement of changes in consumer costs over time and their connections to economic activity.

The ‘month’ variable introduces a temporal aspect, capturing the influence of seasonal patterns on pricing. Derived variables, such as ‘price_per_unit,’ standardize product comparisons by accounting for differences in package sizes or quantities. This transformation of consumer behavior and retail dynamics into structured data ensures that the analysis remains tied to real-world market activities, enhancing the understanding of pricing trends and behaviors across various retail contexts.

Additional details about the dataset are provided in the datasheet, which can be accessed through the repository associated with this paper.

2.3 Outcome variables

The outcome variable for this study, current_price, is the central focus for analyzing product pricing patterns in the Canadian grocery market. It represents the price of beef at the time of data collection, providing a direct measure of how prices vary across different retail settings. By examining current_price, the study captures both short-term price changes and longer-term strategies used by vendors.

This section outlines the characteristics of the current_price variable, including its distribution and observed patterns, supported by summary statistics and visualizations. Key statistics, such as the mean, median, minimum, and maximum values, describe the central tendency and range of beef prices across vendors. Measures of variability, such as the standard deviation, illustrate the extent of price dispersion, showing how consistently prices are set across products and time. This descriptive analysis establishes a basis for identifying significant patterns in beef pricing.

2.3.1 Current Price

This variable represents the price of the product at the time of data collection and serves as the foundation for analyzing how various factors influence pricing. By modeling current_price, the study examines the impact of historical pricing, vendor type, and seasonal patterns on current market prices. Understanding this variable is essential for identifying pricing strategies used by vendors and their effects on consumer expenses.

To present an overview of the outcome variable, a table is provided (Table 1). This table includes key statistics such as the minimum, maximum, mean, median, and standard deviation of the current_price variable. These measures summarize the distribution and variability of product prices, helping to characterize overall pricing patterns in the dataset.

Table 1: Summary statistics for the outcome variable (current_price)

Statistic	Value
Min	0.770
Max	23.980
Mean	9.112
Median	8.580
Standard Deviation	4.327

Table 1 highlighting key measures of central tendency and variability. The data ranges from a minimum value of 0.77 to a maximum of 23.98, illustrating a broad spread. The mean value is 9.11, indicating that most beef prices are concentrated around this range. The median is 8.58, suggesting that half of the observations fall below this value. The proximity of the mean and median implies a relatively symmetric distribution. The standard deviation of 4.33 reflects moderate variability, indicating that most values are distributed within approximately 4.33 units of the mean, providing a clearer understanding of the data's spread.

In addition to the summary table, Figure 1 visualizes the distribution of current_price, showing common price ranges and the overall spread. This graph helps identify any skewness or clustering in the data, which could indicate specific pricing patterns. Combined, the table and histogram provide a detailed overview of the outcome variable, illustrating how product prices differ among vendors and change over time.

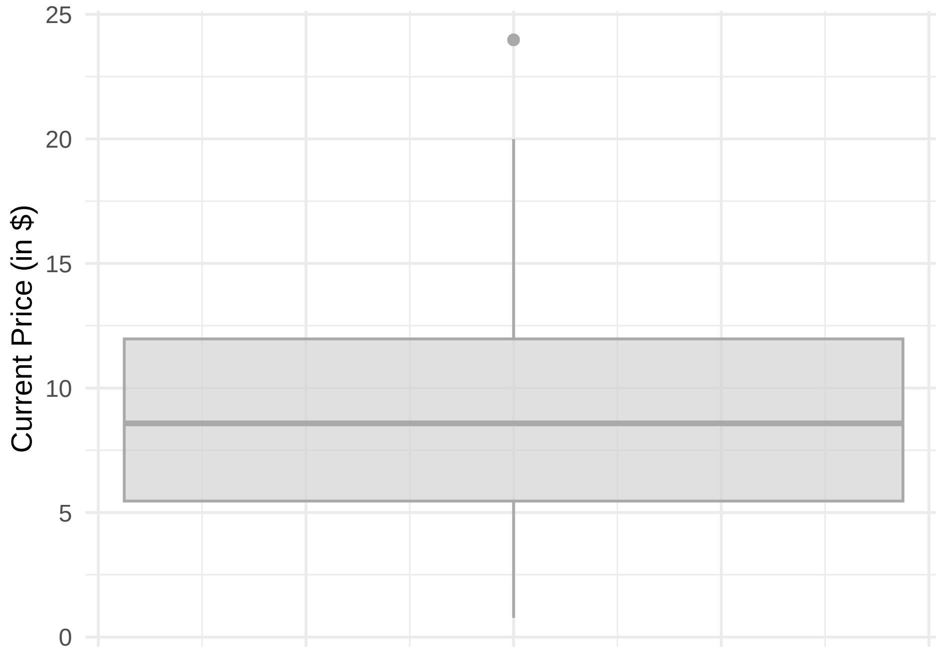

Figure 1: Boxplot of the distribution of current prices

The Figure 1 provides an overview of the distribution of current_price for the products in the dataset. This visualization highlights key features such as the median price, interquartile range (IQR), and any potential outliers in the pricing data.

The median, represented by the horizontal line within the box, indicates the central tendency of product prices, showing where most values are centered. The box itself represents the IQR, which captures the range within which the middle 50% of prices fall, offering a view of price concentration. The whiskers extending from the box depict the range of prices outside the IQR, excluding extreme values. In this case, no significant outliers were detected, suggesting that product prices are generally within a consistent range.

The relatively flat boxplot, combined with a narrow IQR, indicates limited variability in current_price. The small difference between the lower and upper bounds of the box suggests that product pricing is consistent, with most prices clustering around the median. This consistency could reflect standardized pricing strategies among vendors or a lack of variation in the types of products sold. Overall, the boxplot effectively summarizes the distribution of current_price and provides a clear view of pricing patterns across vendors.

2.4 Predictor variables

The predictor variables in this study—month, old_price, and vendor—each contribute to explaining variations in current_price and provide a structured view of the factors influencing beef pricing across retail settings. These variables reflect temporal trends, the influence of historical pricing, and differences in vendor strategies, collectively aiding in the understanding of product pricing in the Canadian grocery market.

The month variable represents the temporal aspect of pricing, enabling the analysis of seasonal patterns in beef prices. Treating month as a categorical variable allows for the identification of months with higher or lower prices, potentially tied to changes in demand due to holidays, promotions, or seasonal supply fluctuations. For instance, months associated with holidays like Thanksgiving or cultural festivals may show increased demand, which could lead to price adjustments. This temporal perspective helps to assess whether vendors apply seasonal pricing strategies that influence current_price over time.

2.4.1 Month

The month variable serves as a predictor to account for seasonal effects on pricing. Including month allows the analysis to identify whether certain times of the year are associated with price fluctuations. Factors such as holidays, promotional events, or changes in seasonal demand may contribute to these variations. Table 2 presents a breakdown of the observations for each month, showing the frequency of data collection across the year. This summary helps identify any months with disproportionately high or low representation, which could impact the interpretation of seasonal pricing patterns.

Table 2: Summary statistics for the predictor variable (month)

Month	Count
6	433
7	514
8	624
9	535
10	578
11	342

The Table 2 presents the number of observations recorded for each month, spanning June to November. The highest count is in August (624), while the lowest is in November (342). This variation may reflect increased market activity or data collection efforts during the summer, possibly due to higher consumer demand during barbecue season. Conversely, the decline in November could be linked to reduced data collection or changes in consumer purchasing patterns. Since data collection began in early 2024, the dataset does not represent the entire year, limiting the ability to evaluate pricing trends across all seasons. These monthly differences highlight the importance of accounting for seasonal factors when analyzing pricing patterns.

2.4.2 Old Price

old_price is a key predictor variable that represents a product's historical price. This variable is used to examine how past pricing influences current pricing strategies, indicating whether products have experienced discounts or price increases. Table 3 summarizes the key statistics for old_price, including its minimum, maximum, mean, median, and standard deviation. Analyzing the distribution of old_price allows us to evaluate whether historical prices significantly differ from current prices and whether adjustments, such as discounts, are applied consistently across products. The relationship between old_price and current_price is further illustrated in Chart 4, providing a visual representation of how historical pricing affects current price decisions.

Table 3: Summary statistics for the predictor variable (old_price)

Statistic	Value
Min	0.770
Max	23.980
Mean	9.112
Median	8.580
Standard Deviation	4.327

The Table 3 summarizes the statistics for the predictor variable old_price, representing the historical price of beef. The minimum value is $1.87, and the maximum value is $25.99, reflecting a wide range of historical prices for beef products. The mean value of $11.48 indicates that, on average, beef products were priced near this level in the past. The median of $10.97 shows that half of the products had historical prices below this value, suggesting a slightly skewed distribution. The standard deviation of $5.32 reflects moderate variability, indicating that while many products were priced close to the average, some were significantly higher or lower. These statistics help describe the variation and central tendency in historical beef pricing.

2.4.3 Vendor

The vendor variable is a categorical predictor that captures differences in pricing behavior across retail chains. It enables the analysis of how pricing strategies differ among vendors. Table 4 lists the number of observations for each vendor, helping to evaluate the representation of each retail chain in the dataset and determine whether certain vendors have a larger impact on the overall analysis.

Table 4: Count of observations for each vendor

Vendor	Count
TandT	1330
Walmart	1696

The Table 4 presents the number of observations for each vendor, with Walmart accounting for 1,696 and T&T for 1,330. While there is a difference in the number of observations, the counts are close enough to allow for a meaningful comparison between the two vendors. This indicates that while Walmart may offer a slightly wider range or maintain more consistent availability of beef products, T&T's dataset is still robust enough to analyze pricing patterns effectively. This balance ensures the analysis reflects variations between a mainstream retailer and a culturally focused store without being significantly affected by unequal sample sizes.

3 Model

Background details and diagnostics are included in Appendix B.

3.1 Overview

To model the current price of beef, y_i, at time i, we apply a Bayesian linear regression model using the `stan_glm` function from the R package `rstanarm`. The response variable, y_i, represents the current price of beef in dollars. The predictors include x_1, x_2, and x_3, corresponding to the month, old price, and vendor, respectively. Each component of the model is defined and explained in the following sections.

3.2 Set-up

Define y_i as the current price of beef at time i, expressed in dollars. The predictors are represented as follows: x_1 is the month (a categorical variable), x_2 is the old price, and x_3 is the vendor (a categorical variable).

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \tag{2}$$

In this model: - α denotes the intercept term, representing the baseline price of beef when all predictors are at their reference levels.

- β_1 represents the effect of the month on the current price, capturing seasonal patterns that may influence beef prices. The month is treated as a categorical variable, recognizing that demand and supply conditions can vary by month due to factors such as holidays, weather, or promotional periods.
- β_2 reflects the effect of the old price on the current price, accounting for the influence of historical pricing. This term helps identify whether past prices have a consistent impact on current pricing decisions.
- β_3 captures the effect of the vendor, also treated as a categorical variable. Different vendors may apply unique pricing strategies, and including this variable allows us to model variations specific to each vendor.

We assume normal priors for the model coefficients and intercept, with mean 0 and standard deviation 2.5:

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\beta_3 \sim \text{Normal}(0, 2.5) \tag{6}$$

These priors are designed to be weakly informative, offering guidance while allowing the data to determine the outcomes. A standard deviation of 2.5 reflects the expectation that most effects are likely to fall within a reasonable range, ensuring the priors are flexible without imposing overly restrictive assumptions.

For the residual standard deviation, σ, we assume an Exponential(1) prior:

$$\sigma \sim \text{Exponential}(1) \tag{7}$$

This prior reflects our belief that the standard deviation should be positive and allows flexibility, while preferring smaller values over larger ones, consistent with the expectation of modest variability around the mean prediction.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich, Ben and Gabry, Jonah and Ali, Iram and Brilleman, Sam (2023). We use the default priors from `rstanarm`.

3.3 Justification

The selected predictors strike a balance between relevance and simplicity for this scenario. Including `month` and `vendor` as categorical variables accounts for the influence of seasonal changes and vendor-specific pricing strategies on the current price of beef. The use of `old_price` reflects the concept of price inertia, where past prices are likely to affect current pricing patterns.

The model avoids being overly simplistic or unnecessarily complex by incorporating key predictors without introducing excessive variables that could risk overfitting. Treating `month` and `vendor` as categorical variables ensures the model captures relevant distinctions without arbitrary groupings, maintaining both clarity and parsimony.

The model is implemented using `stan_glm` from the `rstanarm` package, which provides an accessible interface to Stan for Bayesian modeling. Stan's efficient sampling algorithms ensure reliable convergence and accurate estimation of posterior distributions, making it a robust choice for this analysis.

3.3.1 Assumptions and Limitations

- **Linearity**: The model assumes a linear relationship between the predictors (month, old_price, vendor) and the response variable (current_price). This assumption simplifies the analysis and enhances interpretability, but it may limit the model's ability to capture more complex relationships. For example, interactions or threshold effects between predictors and the response variable would not be reflected in a purely linear framework. If these nonlinear relationships exist but are not modeled, the resulting estimates and predictions may be less accurate, potentially leading to incorrect conclusions about factors affecting beef pricing.
- **Normality of Residuals**: The residuals are assumed to follow a normal distribution, a requirement for making valid statistical inferences about model parameters. If this assumption is violated—such as when residuals are skewed or have heavy tails—it could result in biased estimates, unreliable confidence intervals, and less dependable predictions. Non-normal residuals might suggest missing variables in the model or that a different modeling approach, such as a generalized linear model, might better fit the data.
- **Priors**: The Bayesian framework employs weakly informative priors to allow the data to primarily determine parameter estimates. While this provides flexibility, it also means that stronger prior knowledge of pricing patterns could have been used to improve precision. Using more specific priors, informed by domain knowledge, might enhance the model's performance by producing more stable estimates. However, weakly informative priors can make the model more sensitive to data quality and outliers, as they do not strongly influence the parameter estimates. In cases where the data are limited or noisy, stronger priors could help stabilize the model's outputs.

3.3.2 Validation and Diagnostics

- **Posterior Predictive Checks**: Posterior predictive checks were conducted to evaluate how well the model represents the variation in the data. These checks compare simulated data generated from the posterior distribution to the observed data to identify any discrepancies. Assessing the alignment between these distributions helps determine whether the model captures the underlying patterns in the data. A close match indicates that the model is appropriately capturing the structure of the data, while notable differences may suggest issues such as omitted variables or an incorrect model specification. These checks are a key step in validating the model's ability to generalize beyond the observed dataset.
- **Convergence**: Model convergence was assessed using trace plots and the R-hat statistic, with all R-hat values near 1, confirming successful convergence. Trace plots illustrate how effectively the Markov Chain Monte Carlo (MCMC) sampling algorithm explores the parameter space. Well-mixed chains with stable trajectories suggest the model has reached a stationary distribution. The R-hat statistic, also known as the Gelman-Rubin diagnostic, compares between-chain and within-chain variance, with values close to 1 indicating that all chains have converged to the same target distribution. Together, these diagnostics ensure the reliability of the parameter estimates and confirm that sampling issues did not influence the results.
- **Alternative Models**: Simpler linear models excluding vendor or seasonal variables were also tested. These models showed notably lower predictive performance, highlighting the importance of including both vendor and seasonal factors for accurate modeling. Excluding these variables reduced the model's fit, as indicated by metrics like R-squared and predictive accuracy. This comparison reinforces the necessity of accounting for vendor-specific strategies and seasonal variations to effectively model beef pricing. The evaluation of alternative models confirmed that the selected predictors provide a better representation of the factors influencing price changes in the dataset.

In conclusion, the chosen model effectively captures how various factors influence the current price of beef. The Bayesian framework incorporates uncertainty and prior knowledge, while the selection of priors and predictors ensures the model aligns well with the available data and research objectives.

4 Results

4.1 Charting

Figure 2 illustrates the distribution of current beef prices in the dataset, depicting how prices are spread across different levels. The horizontal axis represents beef prices in dollars, ranging

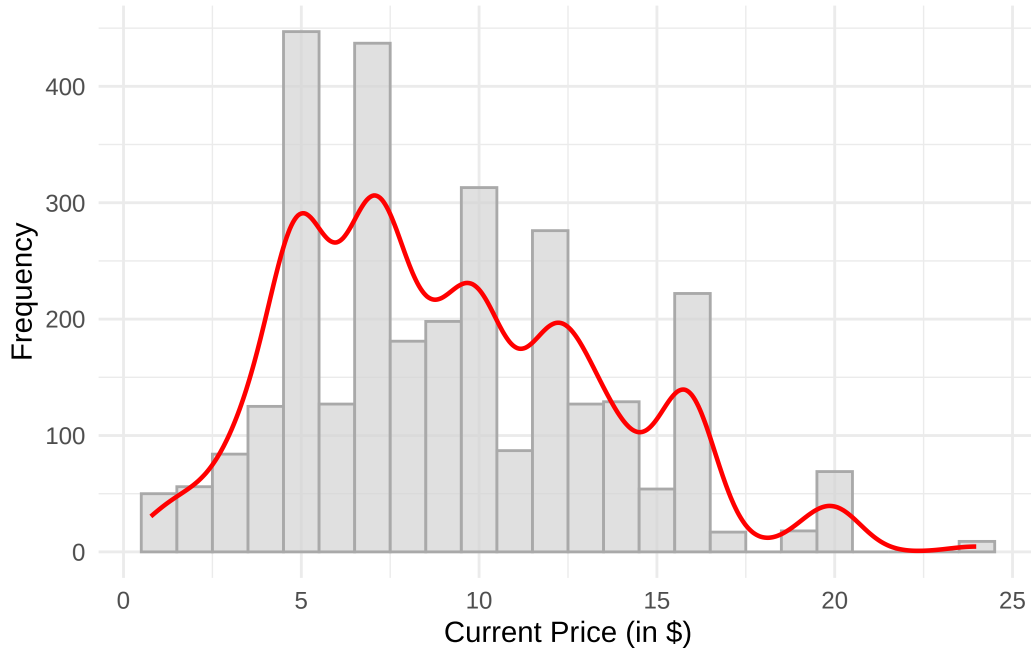

Figure 2: Distribution of Current Prices

from 0 to approximately 25, while the vertical axis shows the frequency, or the number of observations, within each price range. The grey bars represent the histogram, visualizing the frequency distribution of beef prices.

The histogram shows that the most common price range is between $4 and $8, with noticeable peaks around $5 and $8. This indicates that a significant portion of beef products falls within this range, suggesting a central pricing cluster in the market. Smaller peaks around $13 and $15 point to secondary groupings where beef prices are also relatively frequent. Beyond $15, the frequency declines sharply, indicating that higher-priced beef products are less common and that most items are concentrated within a lower price range.

The overlaid red line represents a smoothed density curve, highlighting the overall trend in the distribution. The curve closely aligns with the histogram, showing a skewed pattern with a higher concentration of prices in the lower range and a gradual decrease as prices rise. This distribution suggests that consumers may prefer lower-priced beef products or that pricing aligns with affordability constraints. The multiple peaks in the curve indicate distinct pricing tiers, potentially reflecting differences in product quality, cut types, or retailer-specific pricing strategies.

Figure 3 shows the distribution of price differences for beef products sold by T&T and Walmart, highlighting variations in pricing between these retailers. The y-axis represents the price

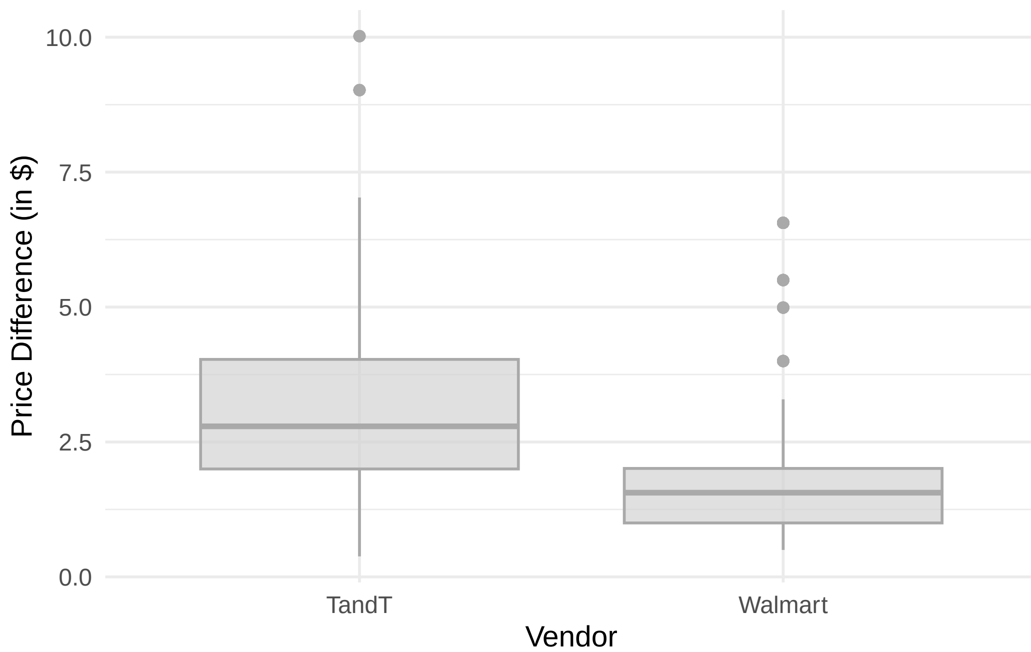

Figure 3: Price Difference by Vendor

difference in dollars, while the x-axis lists the two vendors. The boxplots display the range, median, and variability of price differences for each retailer.

For T&T, the boxplot indicates a broader range of price differences, with the interquartile range (IQR) spanning approximately $1.5 to $4. The median price difference is about $2.5, suggesting that many products cluster around this value. T&T also shows greater variability, with outliers extending beyond $7.5 and $10. These outliers suggest that certain products have significantly larger price differences, possibly reflecting diverse pricing strategies or variations in product offerings.

In comparison, Walmart's boxplot shows a narrower distribution of price differences. The IQR ranges from roughly $1 to $2.5, with a median of around $1.5. This tighter range indicates that Walmart's beef pricing differences are more consistent, reflecting a standardized pricing strategy. While there are a few outliers above $4, these deviations are less pronounced than those seen for T&T.

The boxplots emphasize a clear difference in pricing approaches: T&T exhibits more variability and larger outliers, potentially due to a broader product assortment or targeted market strategies. Walmart, on the other hand, maintains more uniform pricing, offering greater predictability for consumers. These patterns reflect how each retailer positions its beef products in the market.

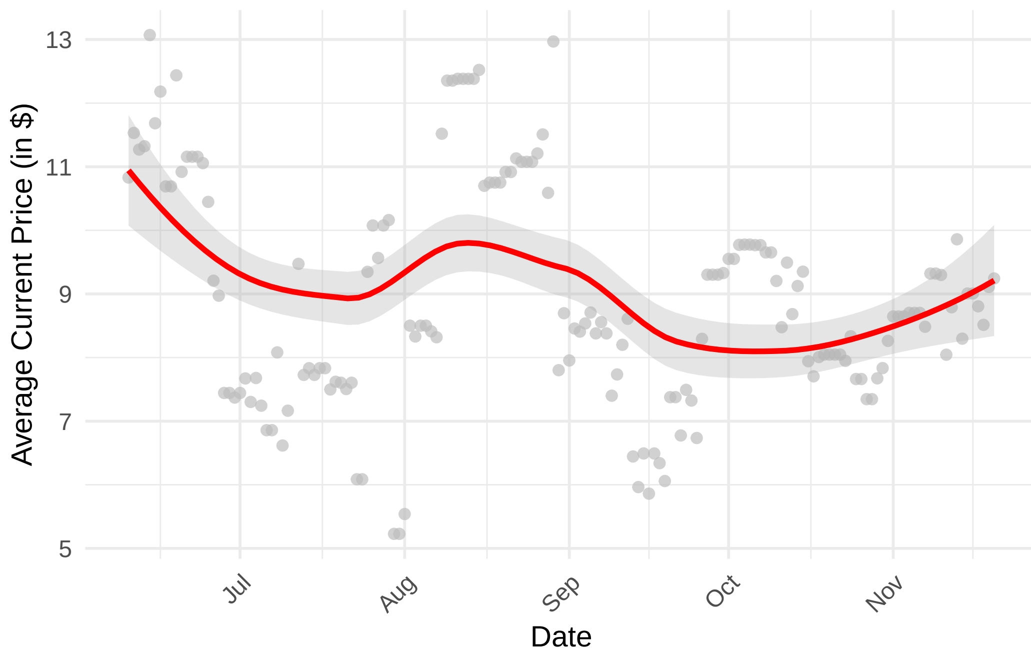

Figure 4: Average Daily Current Price Over Time

Figure 4 displays the trend in average beef prices from July to November, highlighting fluctuations in the market during this period. The y-axis represents the average current price in dollars, while the x-axis shows the progression of dates.

The chart shows a general downward trend in average prices starting in early July, when prices were around $11, and decreasing to approximately $7 by the end of the month. This decline could reflect changes in market dynamics, such as seasonal shifts in demand or adjustments in supply. In August, prices display moderate fluctuations, with occasional spikes above $9, possibly due to short-term disruptions or promotional pricing events.

September and October show a more consistent decline, with average prices dropping below $6 at times. This extended period of lower prices may point to factors like an oversupply or reduced consumer demand. Toward the end of October, prices begin to recover, showing a gradual upward trend into early November and stabilizing between $8 and $9. The shaded region around the line, representing confidence intervals, widens at the edges of the time range, indicating increased uncertainty in the estimates.

This pattern suggests that beef prices during this timeframe were shaped by seasonal trends, retailer pricing strategies, and possible supply chain fluctuations. The observed variability and eventual stabilization highlight the dynamic nature of the market and emphasize the importance of tracking temporal pricing trends to better understand market behavior and inform retailer or policy decisions on pricing strategies and affordability.

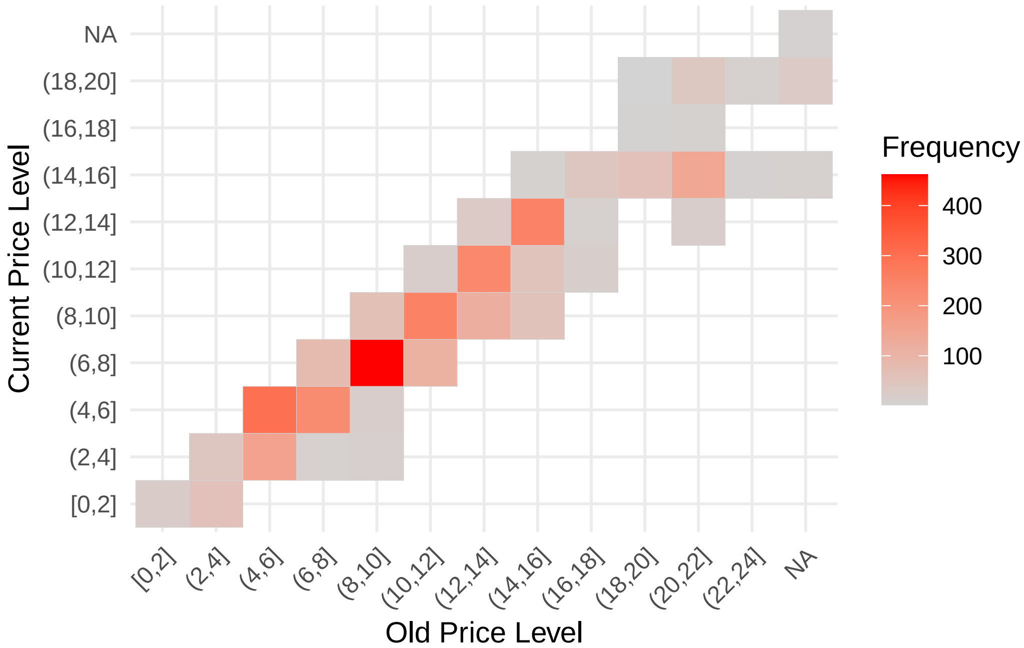

Figure 5: Relationship Between Old and Current Beef Price Levels

Figure 5 presents a heatmap showing the relationship between old and current beef prices, illustrating how pricing has changed over time across various products. The x-axis represents old price levels, and the y-axis represents current price levels, with each cell indicating the frequency of products at specific combinations of these levels. The intensity of the cell color, ranging from light red to dark red, reflects the density of observations, as described by the legend on the right.

The heatmap shows a clear diagonal pattern, indicating that higher old price levels correspond closely to higher current price levels. This alignment highlights a strong positive correlation between historical and current prices, suggesting that most products have undergone proportional pricing adjustments. The darker cells along the diagonal represent the most frequent combinations of old and current prices, pointing to consistent adjustments influenced by factors such as market stability, inflation, or standardized vendor practices.

In contrast, off-diagonal cells, which are lightly shaded or empty, represent rare occurrences of significant deviations from the proportional trend. These sparse areas indicate that sharp discounts or large price increases are infrequent, reinforcing the observation that pricing tends to remain stable and closely tied to historical values.

Overall, the heatmap provides an aggregated view of the consistency in beef pricing across different conditions. By grouping pricing data into distinct levels, it visually emphasizes the alignment between old and current prices, highlighting systematic adjustments and the rarity of irregular deviations, which may be influenced by vendor strategies or external market factors.

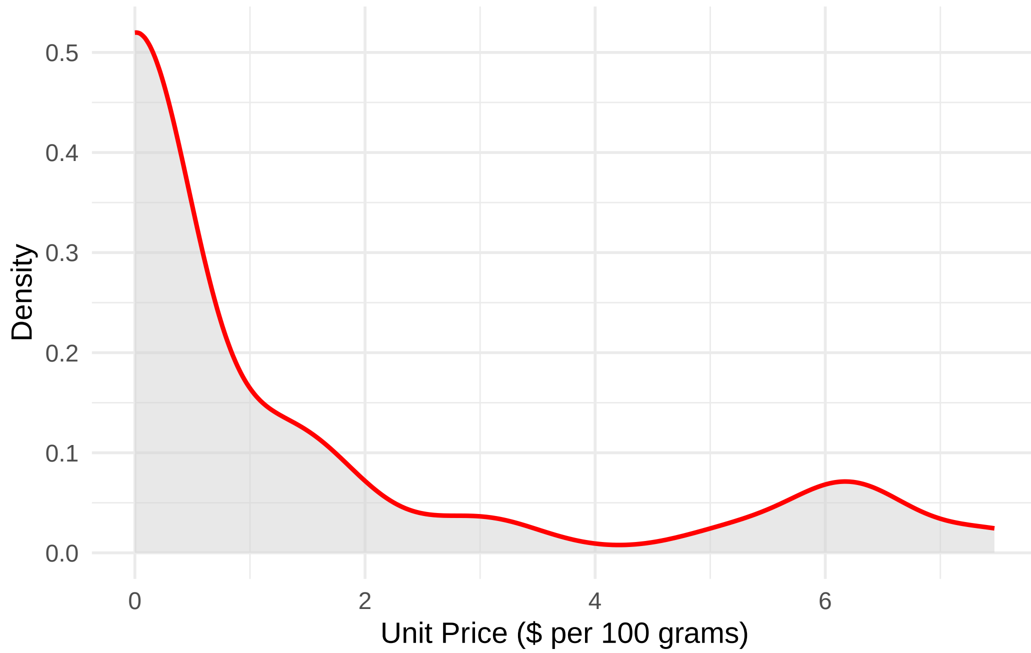

Figure 6: Distribution of Unit Prices

Figure 6 displays the distribution of unit prices for beef, measured in $ per 100 grams. The x-axis shows unit prices, while the y-axis represents density, indicating the proportion of observations across different price ranges. The grey area under the density curve highlights the relative concentration of unit prices, while the red curve provides a smoothed estimate of the overall distribution.

A notable feature of this figure is the high density near $0, indicating that a large number of observations in the dataset have very low unit prices. This concentration could reflect promotions, bulk discounts, or potential data issues that may require further review. Beyond this initial spike, the density decreases sharply, with smaller secondary peaks appearing around $2, $4, and $6 per 100 grams. These peaks may correspond to different pricing tiers influenced by factors such as product quality, cut type, or retailer-specific strategies.

The distribution is heavily right-skewed, with the majority of observations clustered at lower price points and fewer cases of higher-priced beef products. This pattern suggests that the market is dominated by affordable beef options catering to a broad consumer base, while premium products occupy a smaller share. The presence of multiple peaks indicates segmentation within the market, driven by variations in product attributes or retailer pricing approaches. This highlights the coexistence of budget-friendly and premium offerings, reflecting diverse consumer preferences and pricing strategies in the Canadian grocery market.

Table 5: Beef model results

	Beef model (Bayesian)
(Intercept)	-0.37
	(0.14)
month	-0.01
	(0.01)
old_price	0.81
	(0.00)
vendorWalmart	0.56
	(0.04)
Num.Obs.	3026
R2	0.945
R2 Adj.	0.944
Log.Lik.	-4349.672
ELPD	-4354.7
ELPD s.e.	58.7
LOOIC	8709.5
LOOIC s.e.	117.3
WAIC	8709.4
RMSE	1.02

4.2 Modeling

Table 5 presents the outcomes of the Bayesian model used to analyze beef prices, highlighting parameter estimates and diagnostic measures. The model evaluates the relationship between the current price of beef and predictors, including the old price, vendor, and month of observation.

The estimated intercept of -0.37, while seemingly unusual, serves as a reference point within the model and should be interpreted in the context of the predictors. The "month" coefficient of -0.01 indicates a small downward trend in current prices over time, possibly reflecting seasonal influences or market adjustments. The "old_price" coefficient, at 0.81, demonstrates a strong positive relationship, indicating that higher historical prices correspond to proportionally higher current prices, suggesting consistent pricing patterns over time. Additionally, the "vendorWalmart" coefficient of 0.56 shows that Walmart's beef prices are, on average, higher than those at T&T, consistent with observed vendor-specific pricing strategies.

In terms of model fit, the R-squared value of 0.945 and adjusted R-squared of 0.944 indicate that the model accounts for a significant portion of the variability in beef prices. Metrics such as the negative log-likelihood (-4349.672) and expected log pointwise predictive density (ELPD) of -4354.7 further support the model's ability to capture key patterns in the data. Predictive measures, including the leave-one-out information criterion (LOOIC) and Widely Applicable Information Criterion (WAIC), both approximately 8709, confirm the model's reliability in making predictions. The root mean square error (RMSE) of 1.02 reflects minimal average deviation between predicted and observed values, underscoring the model's accuracy.

The results demonstrate a strong, consistent relationship between historical and current pricing while highlighting differences in pricing strategies between vendors. Using a Bayesian framework allows for uncertainty to be incorporated into the analysis, providing robust parameter estimates. The high explained variance and reliable predictive performance affirm the model's effectiveness in analyzing and forecasting beef price trends, offering a clear understanding of market behaviors for both consumers and stakeholders.

5 Discussion

5.1 Interpretation of Findings

The findings of this study provide a detailed look into the factors influencing beef pricing in the Canadian grocery market. Using Bayesian linear regression, the results demonstrate that historical pricing (old_price) and vendor-specific strategies are significant predictors of current prices. The strong positive relationship between old_price and current_price suggests a high level of consistency in pricing practices, likely influenced by factors such as inflation or vendor strategies. Walmart and T&T exhibit distinct pricing approaches, with Walmart showing more consistent pricing while T&T displays greater variability, possibly due to its focus on specialty goods catering to specific cultural preferences. Seasonal trends, represented by the month variable, indicate small but noticeable temporal price variations, likely linked to supply and demand fluctuations throughout the year.

Visualizations in the previous section (Section 4), along with statistical summaries, reinforce these findings. The heatmap in Figure 5 demonstrates a strong alignment between old and current prices, highlighting the significant influence of historical costs on current pricing. Factors such as production expenses, inflation, and market demand appear to affect historical and current prices in similar ways. The high R-squared value of the Bayesian model (Table 5) further confirms that the selected predictors capture the majority of variation in beef prices, emphasizing their importance. Additional observations can be drawn from these results, as discussed below.

5.2 Vendor-Specific Pricing Strategies

A key observation is the clear difference in pricing strategies between Walmart and T&T. Walmart's pricing strategy, as reflected by its narrower boxplot range (Figure 3), emphasizes consistency and predictability, aligning with its branding as a retailer offering everyday low prices (Schepers 2021). This approach likely appeals to price-sensitive consumers seeking stability and affordability. In contrast, T&T shows a wider range of price differences, which may reflect a dynamic pricing strategy tailored to its specialty products, diverse supply chain, and specific customer demographic (Hii 2007). The variability in T&T's pricing could stem from promotional efforts, targeted price differentiation, or a wider range of product qualities and types.

Understanding these strategies is essential for interpreting pricing data. Walmart's standardized pricing is likely to attract a broad customer base, while T&T's variability might appeal to niche markets looking for specialty goods. Future research could expand on these findings by incorporating product-level details, customer demographics, or promotional data to better understand the drivers behind these pricing differences.

5.3 Price Stability and Market Forces

The study highlights a general stability in beef pricing, with adjustments largely proportional to historical prices. This stability may suggest a well-functioning market where historical prices act as an anchor for setting current prices, potentially influenced by supplier contracts or steady production costs. The modest decline in prices observed in certain months could reflect seasonal promotions or changes in demand, driven by factors like holiday sales (Elsbree 2022) or shifts in meat consumption habits.

The clustering of prices at the lower end of the range suggests that the majority of beef products cater to value-conscious consumers, which may reflect broader economic trends or retailer strategies to maintain affordability. Larger players like Walmart might use economies of scale to keep prices competitive, shaping market dynamics as a result. To build on these findings, future studies could investigate how external events, such as economic downturns or supply chain disruptions (Chen et al. 2019), affect this pricing stability and whether similar patterns are observed in other product categories.

5.4 Cultural Preferences and Product Differentiation

Cultural preferences play a significant role in beef pricing, as evidenced by the differences between T&T and Walmart. T&T's pricing patterns, which are more varied, may reflect its focus on serving an Asian customer base with specific culinary needs (Shannon and Mandhachitara 2005). The diversity in its product cuts and grades could contribute to these patterns, as sourcing requirements and preparation methods may differ based on cultural preferences. This may lead to higher variability in prices.

For policymakers and retailers addressing food affordability, these findings underscore the importance of considering cultural factors in consumption patterns. Retailers like T&T, which cater to specific demographic groups, might need tailored pricing strategies, while larger chains like Walmart could focus on maintaining uniform pricing to appeal to a broader market.

5.5 Promotional Strategies and Consumer Behavior

Promotional strategies also play a role in shaping beef prices. T&T's greater variability could reflect frequent promotional activities aimed at specific customer segments. Promotions may be used to move inventory, introduce new products, or compete with other retailers (Pilar Martínez Ruiz and Mollá Descals 2008). This contrasts with Walmart's approach, which prioritizes stable pricing over frequent discounts.

Consumer preferences likely influence these strategies. Price-sensitive shoppers may gravitate toward Walmart for its predictable pricing, while those seeking specialty cuts or products may prefer T&T, even with its price fluctuations. Future research could examine the impact of promotions by analyzing temporal pricing data alongside promotional calendars or sales events to better understand how short-term discounts affect pricing patterns and consumer loyalty.

5.6 Weaknesses and Future Directions

While this study provides a useful analysis of beef pricing dynamics, it has some limitations. First, the dataset includes only two vendors, Walmart and T&T, which may not fully represent the diversity of the Canadian grocery market. Expanding the study to include additional retailers, such as discount chains, premium grocers, or independent stores (Anselmsson, Johansson, and Persson 2007), could offer a broader perspective. Additionally, the analysis does not account for factors like promotions, product quality, or customer loyalty programs (Ho et al. 2009), which could significantly influence pricing. Future models incorporating these variables could provide a deeper understanding of pricing behavior.

The assumption of a linear relationship between predictors and the outcome may oversimplify the complexity of real-world pricing. Future studies could explore more flexible modeling techniques, such as non-linear models or machine learning approaches, to better capture these relationships. Incorporating external data sources, such as economic indicators, weather patterns, or detailed sales records, could further validate and enrich the findings.

Finally, the assumptions of constant variance and normality of residuals may not hold perfectly in pricing data, which often exhibit heteroscedasticity or other deviations. Employing more robust regression techniques or heteroscedasticity-consistent estimators could improve model accuracy and reliability. An interesting avenue for future research would be to explore the impact of shifts in consumer preferences, such as increased demand for plant-based alternatives, on beef pricing trends, providing insight into evolving market dynamics.

Appendix

A Data details

A.1 Cleaning methods

In this data cleaning process, the goal was to load raw product and transaction data, merge them, filter and transform the necessary columns, and finally save the cleaned data for further analysis. The process began with loading two CSV files: "hammer-4-product.csv" and "hammer-4-raw.csv". These files contained information about different products and transactions related to them. After loading the data, the next step was to merge the two datasets to create a combined dataset for a more complete product information, as Table 6 shown.

Table 6: Raw Merged Data

Time	Vendor	Product ID	Product Name	Brand	Current Price	Old Price	Units	Price per Unit
2024-06-22 10:35:00	Voila	3	Apples Ambrosia	NA	5.49	6.49	1.36kg	$0.40/100g
2024-06-10 23:15:00	Voila	3	Apples Ambrosia	NA	5.99	6.49	1.36kg	$0.44/100g
2024-06-11 18:36:00	Voila	3	Apples Ambrosia	NA	5.99	6.49	1.36kg	$0.44/100g
2024-06-12 11:39:00	Voila	3	Apples Ambrosia	NA	5.99	6.49	1.36kg	$0.44/100g
2024-06-13 10:23:00	Voila	3	Apples Ambrosia	NA	6.49	NA	1.36kg	$0.48/100g
2024-06-14 10:39:00	Voila	3	Apples Ambrosia	NA	6.49	NA	1.36kg	$0.48/100g

The merging of the two datasets was accomplished using an inner join on the `product_id` and `id` columns. This allowed for the selection of relevant columns, such as product details (`product_name`, `brand`), vendor information, and pricing (`current_price`, `old_price`, and `price_per_unit`). The goal was to retain only those data attributes that were essential for subsequent analyses.

Once the datasets were merged, the next stage was to focus specifically on price-per-unit data (`price_per_unit`) and create a subset, `ppu_data`, which only included vendors Walmart and TandT. Additionally, this subset included only those records where the product names contained "beef", helping to narrow down the dataset to beef-related items. Only the `vendor` and `price_per_unit` columns were selected to generate this subset, which would later be saved as a separate parquet file for analysis.

The main cleaning process involved creating a cleaned dataset (Table 7), `cleaned_data`, by filtering out only those records from the merged dataset that belonged to Walmart and TandT. The data was further refined by selecting important columns, such as transaction time (`nowtime`), vendor, product pricing, and product details. Several transformations were then applied to ensure the data was suitable for analysis. For example, the `nowtime` column was parsed to extract `year`, `month`, and `day` values, which provided a more granular breakdown of transaction time. The `current_price` and `old_price` columns were converted to numeric values using `parse_number()` to handle any non-numeric characters.

Table 7: Cleaned Data

vendor	current_price	old_price	product_name	price_per_unit	year	month	day
TandT	10.99	13.43	T&T Japanese Curry Beef (450g)	0	2024	6	11
TandT	10.99	13.43	T&T Japanese Curry Beef (450g)	0	2024	6	11
TandT	10.99	11.99	T&T Japanese Curry Beef (450g)	0	2024	6	12
TandT	10.99	11.99	T&T Japanese Curry Beef (450g)	0	2024	6	13
TandT	7.99	9.99	Shirakiku Cooked Sukiyaki Beef (250g)	0	2024	7	12
TandT	7.99	9.99	Shirakiku Cooked Sukiyaki Beef (250g)	0	2024	7	13

For the `price_per_unit` column, the values were cleaned to ensure consistency. The dollar sign was extracted using `str_extract()` and subsequently removed using `str_remove()`. The result was then converted to numeric values for accurate comparison and analysis. In cases where `price_per_unit` was missing, the `ifelse()` function was used to assign a default value of 0, ensuring no missing data points could interfere with subsequent operations.

To focus specifically on beef products, the `cleaned_data` dataset was filtered to include only those rows where `product_name` contained "beef" while excluding records with terms like "flavour", "vermicelli", "rice", and other similar terms that indicated non-beef products. This filtering step was crucial to ensure the final dataset contained only relevant beef-related products, reducing noise from irrelevant items. The `nowtime` column, which was no longer needed, was dropped, and rows containing any `NA` values were also removed using `drop_na()`.

Finally, the cleaned data was saved into parquet files: `ppu_data.parquet` for the price-per-unit data subset and `beef_data.parquet` for the full cleaned dataset.

A.2 Vendor Choice

The selection of Walmart and T&T as representative vendors for this analysis was driven by their distinctive characteristics and market coverage, which together offer a comprehensive view of beef pricing dynamics. Walmart is a prominent multinational retailer with a broad presence across North America, making it a well-suited representative of mainstream, large-scale grocery retailing. Its extensive reach and emphasis on standardized pricing provide valuable insights into the general market trends and pricing behaviors that are accessible to a wide segment of consumers. By including Walmart in the analysis, we can capture a perspective that is reflective of the typical shopping experience for many consumers, characterized by competitive pricing and a broad product selection.

On the other hand, T&T Supermarket represents a niche segment of the market, catering specifically to Asian communities and consumers seeking specialty goods that align with Asian culinary traditions. T&T's focus on products tailored to Asian tastes, along with its unique supply chains and vendor relationships, provides an important contrast to Walmart. The inclusion of T&T allows for an exploration of how cultural preferences and niche market positioning influence pricing. By analyzing T&T, we gain insights into pricing practices that reflect the demands of a distinct consumer group and how these differ from the broader market.

The combination of Walmart and T&T enables the analysis to explore both mainstream and culturally specific market behaviors. This choice of vendors thus enriches the study by accounting for variations in consumer preferences, pricing strategies, and market positioning. Walmart's emphasis on scale and cost-efficiency contrasts with T&T's specialization and cultural targeting, allowing us to draw meaningful comparisons in terms of beef pricing strategies between a general and a specialized market context. Such a nuanced understanding is essential for capturing the diversity within the beef retail market, providing a more holistic view of the factors driving beef prices in different retail settings.

Table 8: Price per Unit Data for TandT and Walmart

(a) Raw Data

Vendor	Price per Unit
TandT	NA
TandT	NA
TandT	NA
TandT	NA
TandT	NA
TandT	NA
Walmart	$1.31/100g
Walmart	$1.31/100g
Walmart	$1.31/100g
Walmart	$1.31/100g
Walmart	$1.31/100g
Walmart	$1.31/100g

(b) Cleaned Data

Vendor	Price per Unit
TandT	0.0
TandT	0.0
TandT	0.0
TandT	0.0
TandT	0.0
TandT	0.0
Walmart	1.5
Walmart	1.5
Walmart	1.5
Walmart	1.5
Walmart	1.5
Walmart	1.5

A.3 Price per unit

In cleaning the `price_per_unit` variable, I began by extracting the numerical value using `str_extract()`, which allowed me to identify the dollar amounts in the original string (`\\$[0-9\\.]+`), as shown in Table 8a. This step ensured that only valid price values, starting with a dollar sign, were captured, removing any extraneous text or symbols.

Next, I removed the dollar sign itself with `str_remove("\\$")` to convert the extracted value into a format suitable for numerical operations. After this, I used `as.numeric()` to convert the string into a numeric type, allowing for further quantitative analysis.

To handle missing values, I applied `ifelse(is.na(price_per_unit), 0, price_per_unit)`, which replaced any `NA` values in the `price_per_unit` column with `0`. This step ensured that all rows had a value for `price_per_unit`, preventing issues during subsequent analysis. Using zero as the replacement helped maintain continuity in the dataset, although I acknowledge that this choice could imply that missing values are negligible, which might require additional context or consideration during interpretation. The price_per_unit column in the dataset after cleaning looks like Table 8b.

B Model details

B.1 Posterior predictive check

In Figure 7, we implement a posterior predictive check. This plot shows the observed data distribution (in dark lines) compared with simulated datasets from the posterior (in light grey lines). The purpose of this check is to evaluate how well the model can reproduce the observed data. The close alignment between the simulated posterior predictive distributions and the observed data suggests that the model is accurately capturing the underlying patterns in the data. Specifically, we observe that the peaks and troughs in the observed data are well represented by the posterior samples, indicating that the model is capable of capturing key features such as the general shape and spread of the data distribution. The variability across the posterior predictive samples is consistent with the observed variability, suggesting no significant model misfit.

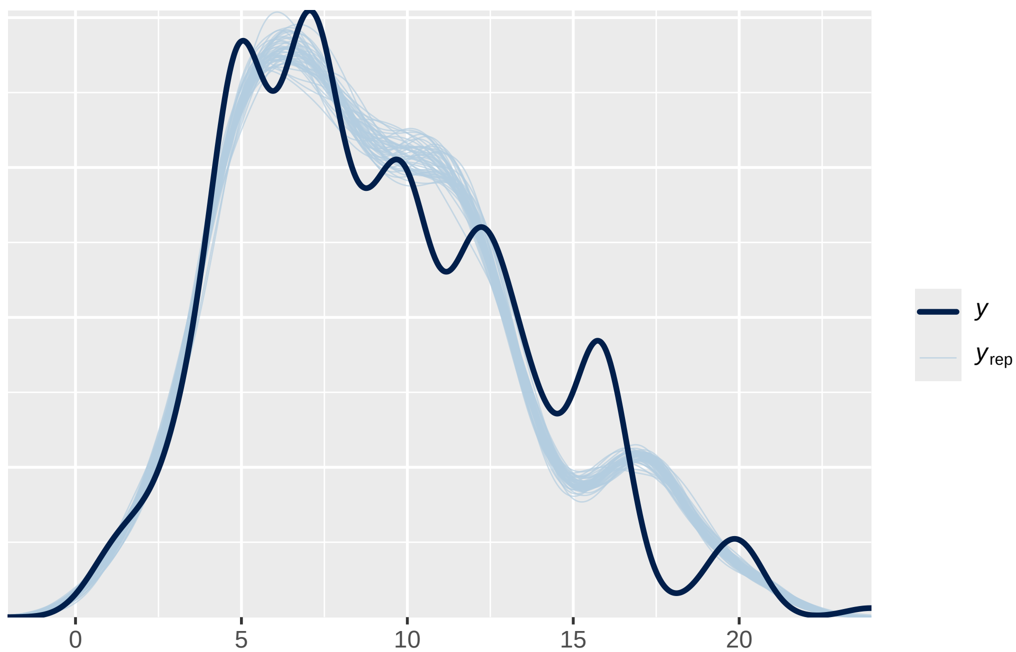

Figure 7: Posterior prediction check result

In Figure 8, we compare the posterior with the prior distributions for each parameter. The posterior distribution (shown on the left) is more concentrated compared to the prior (shown on the right), which demonstrates how the data has influenced and updated our beliefs about the model parameters. For instance, the posterior distribution for the intercept has shrunk considerably compared to its prior, indicating that the data has substantially informed our estimate of the baseline price level. Similarly, for the other parameters (month, old_price, sigma, and vendorWalmart), the posteriors are less dispersed than the priors, suggesting a meaningful reduction in uncertainty. Notably, the posterior intervals are much narrower, especially for the intercept and old price parameters, indicating that the data has provided strong information about these effects. The results demonstrate that the priors were sufficiently weak, allowing the data to dominate the inference and significantly refine the parameter estimates.

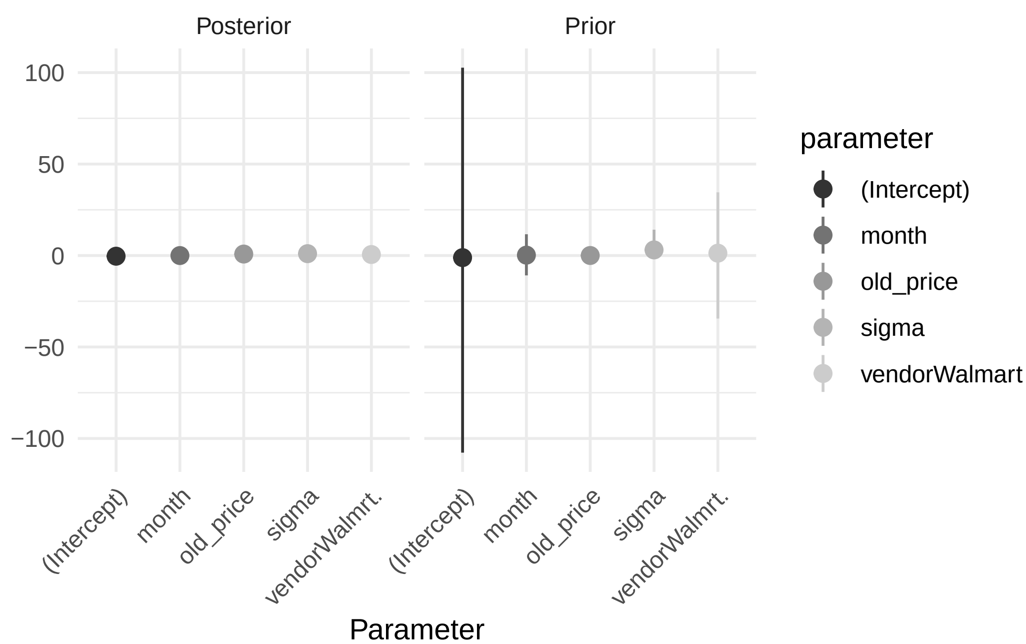

Figure 8: Comparing the posterior with the prior

B.2 Diagnostics

In Figure 9, we present a trace plot. This plot shows the sampled values for each parameter across the MCMC iterations. The trace plot helps assess whether the chains have converged to a stationary distribution. In these plots, we see that each parameter's chains are well mixed, with no discernible trends, indicating that the MCMC algorithm has adequately explored the parameter space. The lack of significant upward or downward drift suggests that all chains are consistently sampling from the target distribution, which is a strong indicator of convergence.

Specifically, the chains for each parameter appear to oscillate around a constant mean, and the overlap between the four chains is extensive. This well-mixed appearance across all parameters, including the intercept, month, old price, vendorWalmart, and sigma, gives us confidence that our model has reached convergence, with each chain sampling from similar regions of the posterior distribution.

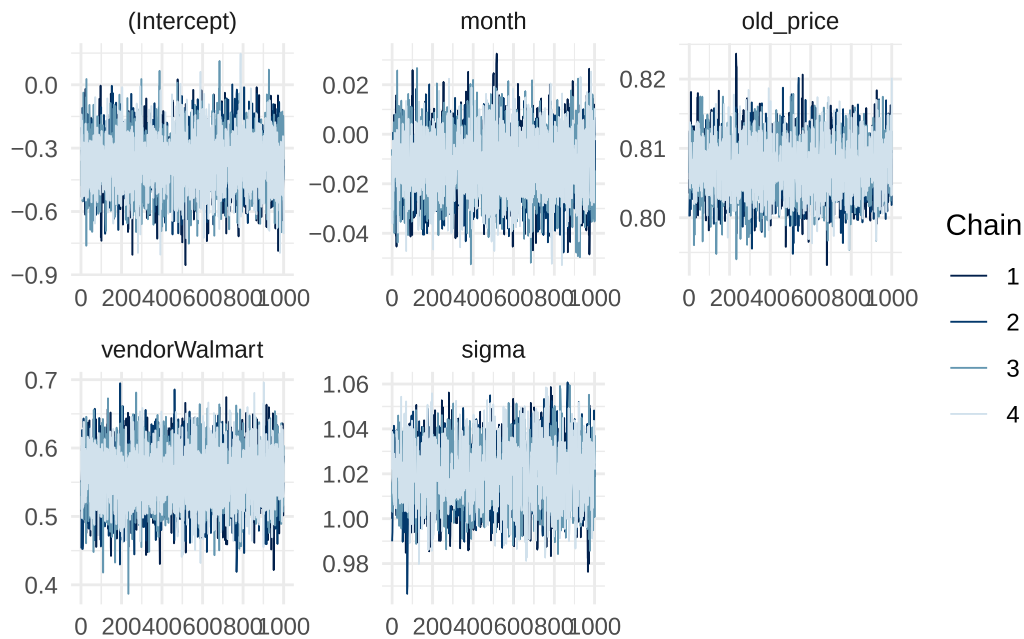

Figure 9: Trace plot

In Figure 10, we present an Rhat plot. The Rhat statistic (also known as the potential scale reduction factor) measures the ratio of the variance between chains to the variance within chains, and values close to 1 indicate that the chains have converged. In this plot, we see that all Rhat values for the parameters are very close to 1, specifically below the threshold of 1.05, suggesting that convergence has been achieved across all parameters.

The fact that all Rhat values are below 1.05 suggests that the different chains are in agreement about the underlying posterior distributions. This indicates that our model's parameter estimates are stable, and the MCMC algorithm has effectively converged. As such, we can be confident in the reliability of the estimates provided by our Bayesian model.

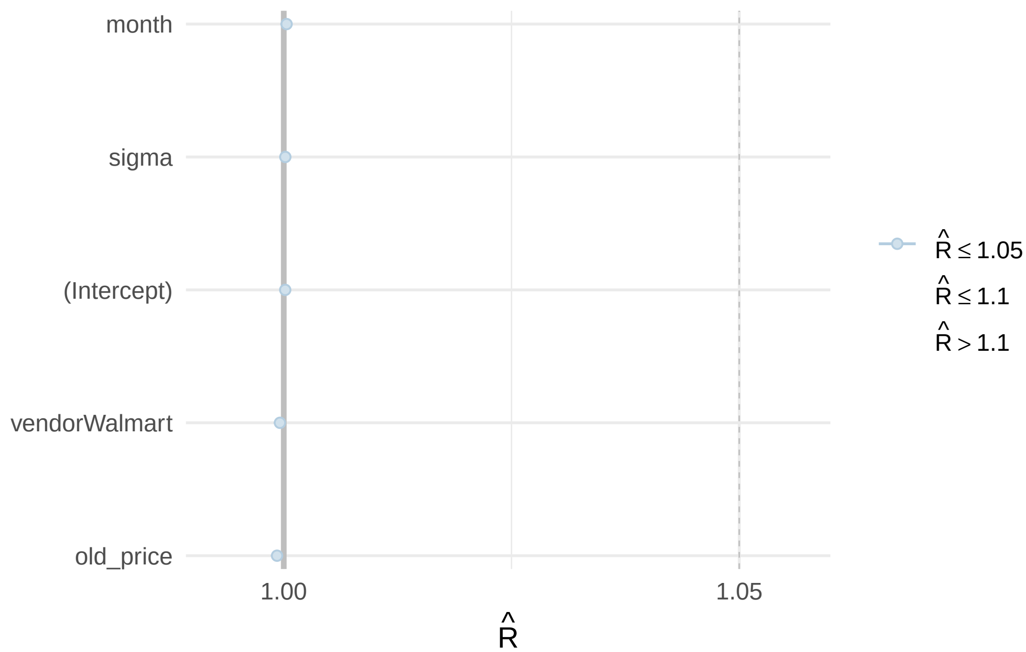

Figure 10: Rhat plot

C Idealized Methodology for a Survey on Beef Pricing

C.1 Overview

To complement the quantitative analysis presented in this paper, an idealized survey methodology could provide direct insights into consumer behavior and perceptions regarding beef pricing at Walmart and T&T. This survey would aim to identify factors influencing purchasing decisions, perceptions of price fairness, and preferences for vendor-specific attributes, offering a holistic understanding of the dynamics underpinning beef pricing in the Canadian grocery market.

C.2 Sampling Approach

A **stratified random sampling** method would be employed to ensure comprehensive representation across key demographic groups, including age, income levels, and cultural background. This approach aims to capture the diversity in shopping preferences and behaviors, particularly as they relate to beef pricing in the Canadian grocery market. By employing stratified random sampling, the survey would not only provide insights into how different demographic segments make purchasing decisions but also allow for more granular comparisons between consumer groups, which is critical for understanding pricing dynamics in diverse retail environments.

1. **Region**: The sample would be divided into urban and rural shoppers across Canada. This stratification would ensure representation of different population densities, as consumer preferences and shopping patterns can differ significantly between urban and rural areas. Urban shoppers, for example, might have more convenient access to both Walmart and T&T locations, whereas rural shoppers may be limited to fewer options or specific local vendors. This distinction is crucial for understanding geographical influences on grocery purchasing behavior, particularly in terms of access, convenience, and price sensitivity.

2. **Vendor familiarity**: The survey would categorize respondents into three groups: regular Walmart shoppers, regular T&T shoppers, and those who shop at both retailers. Regular Walmart shoppers may prioritize price consistency and availability of products, whereas regular T&T shoppers might be more focused on product variety and specialty items that cater to cultural preferences. Including those who shop at both retailers would provide insights into consumer behaviors that shift between general and specialty retail environments, offering a comparative view of how brand familiarity and loyalty shape beef purchasing decisions.

3. **Income brackets**: Respondents would be classified into low, middle, and high-income households. This stratification is essential for analyzing the impact of financial constraints on beef purchasing patterns. Lower-income households might be more sensitive

to price fluctuations and promotions, often choosing products based on affordability. Middle-income households may exhibit a mix of price sensitivity and preference for product quality, whereas high-income households could prioritize premium product quality over price. By segmenting the respondents into income brackets, the survey could identify how economic factors influence both the frequency and type of beef products purchased at Walmart and T&T, revealing the varying levels of price elasticity among different socioeconomic groups.

A sample size of approximately 1,200 respondents would be targeted, with 400 participants in each stratum (Walmart-focused, T&T-focused, and dual shoppers). This stratification would allow for meaningful comparisons between vendor-specific perceptions and strategies.

C.3 Survey Structure

The survey would include a mix of **closed-ended** and **open-ended** questions.

C.3.1 Question Types

1. **Demographics and Shopping Habits**
    - Frequency of grocery shopping (weekly, bi-weekly, etc.).
    - Preferred vendor for beef purchases and reasons for preference.
    - Average budget allocated for beef purchases.
2. **Price Perception**
    - Rating the perceived fairness of beef prices on a Likert scale.
    - Comparison of beef prices at Walmart and T&T (e.g., cheaper, equivalent, more expensive).
    - Sensitivity to price changes (e.g., a 10% increase in beef prices would lead to...).
3. **Cultural and Quality Preferences**
    - Importance of cultural alignment in beef products (specific cuts, preparation styles, etc.).
    - Perceived quality differences between Walmart and T&T beef products.
    - Willingness to pay a premium for culturally tailored products.
4. **External Factors**
    - Awareness of seasonal price trends.
    - Impact of promotions on purchasing decisions.
    - Responses to supply chain disruptions (e.g., alternative purchasing strategies).

C.3.2 Question List

1. How often do you shop for groceries?
   - Weekly
   - Bi-weekly
   - Monthly
   - Less often
2. Which retailer do you prefer for buying beef products?
   - Walmart
   - T&T
   - Both
3. What factors influence your choice of retailer for beef purchases? (Select all that apply)
   - Price
   - Quality
   - Availability of specific cuts
   - Cultural preferences
   - Convenience
   - Promotions
4. On average, how much do you spend on beef per shopping trip?
   - Less than $20
   - $20 - $50
   - $50 - $100
   - More than $100
5. How fair do you find the current beef prices at Walmart/T&T?
   - Very unfair
   - Somewhat unfair
   - Neutral
   - Somewhat fair
   - Very fair
6. How would you rate the quality of beef at Walmart compared to T&T?
   - Much lower quality
   - Slightly lower quality
   - About the same quality
   - Slightly higher quality
   - Much higher quality
7. If beef prices increased by 10%, would you consider switching to an alternative product or retailer?

- Definitely would switch
- Probably would switch
- Might switch
- Probably would not switch
- Definitely would not switch

8. How important is it for you that beef products align with your cultural preferences (e.g., specific cuts, preparation styles)?
    - Not important at all
    - Slightly important
    - Moderately important
    - Very important
    - Extremely important
9. Are you willing to pay more for beef products that are culturally tailored to your preferences?
    - Not willing at all
    - Slightly willing
    - Moderately willing
    - Very willing
    - Extremely willing
10. How aware are you of seasonal price changes in beef products?
    - Not aware at all
    - Slightly aware
    - Moderately aware
    - Very aware
    - Extremely aware
11. How do promotions impact your decision to purchase beef?
    - No impact
    - Slight impact
    - Moderate impact
    - Significant impact
    - Very significant impact
12. How do you typically respond to supply chain disruptions affecting beef availability?
    - Purchase alternative products
    - Purchase from a different retailer
    - Reduce beef consumption
    - No change in purchasing behavior

C.4 Recruitment Strategy

Participants would be recruited using **multichannel outreach**, including:

1. **Online Panels**: The recruitment process would involve partnering with established Canadian consumer research platforms to access existing online panels. These platforms typically have large databases of participants that are segmented based on various demographics such as age, income, and geographic region, allowing for targeted recruitment. By collaborating with these platforms, the survey can include participants from diverse regions—spanning urban centers like Toronto and Vancouver, as well as rural areas across provinces. Online panels are advantageous as they provide quick access to a variety of respondents, increasing the efficiency of data collection while allowing for quotas to be set to ensure representativeness across the stratification variables (e.g., income brackets, vendor familiarity).

2. **Retailer Partnerships**: Collaborating directly with Walmart and T&T stores would allow for in-store recruitment, making it possible to capture insights from shoppers at the point of purchase. A key aspect of this strategy would involve placing QR codes on receipts or at store exits, inviting shoppers to participate in the survey by scanning the code with their mobile devices. This in-store recruitment tactic not only encourages immediate participation from shoppers but also provides the opportunity to capture data from consumers while their purchasing experience is still fresh. Additionally, leveraging in-store displays, posters, or handouts could help draw more attention to the survey. Walmart and T&T staff could also provide gentle prompts to shoppers, further promoting survey participation. This approach would help ensure that insights are gathered directly from those actively engaged with these retailers, offering real-time perspectives on beef purchasing behavior.

3. **Community Outreach**: To reach underrepresented groups, particularly T&T shoppers from culturally diverse backgrounds, targeted community outreach would be conducted in collaboration with cultural organizations, community centers, and local events. Engaging with organizations that have strong connections to the Asian community—such as cultural associations, language schools, or ethnic community centers—would help ensure that the sample includes individuals who may not be as reachable through online panels or in-store promotions alone. Additionally, outreach could include attending cultural festivals, placing ads in community newsletters, or utilizing social media channels that serve specific cultural groups. This grassroots strategy would be particularly useful in accessing populations that may have different shopping patterns or dietary needs, thereby providing richer insights into how cultural preferences influence beef purchasing decisions. Engaging directly with cultural communities would help ensure a higher level of trust and response rate from these groups, who might otherwise be hesitant to participate in market research.

C.5 Linkage to Literature

The survey design for this study draws on established methodologies in consumer behavior and retail pricing research. By employing **stratified random sampling**, this study ensures that the key factors influencing purchasing decisions, such as income, cultural background, and shopping preferences, are effectively captured. This approach is informed by existing literature emphasizing the importance of **representative sampling** to understand market segmentation, particularly in heterogeneous consumer markets (Foedermayr and Diamantopoulos 2008). Stratified sampling is particularly beneficial in capturing the variations in consumer preferences that may be linked to demographic differences, as suggested by **demographic-based segmentation theories** in consumer research.

The use of **multichannel recruitment** has also been shown to be effective in increasing participation rates, especially when attempting to reach a diverse population base (Subramaniam et al. 2024). Studies have demonstrated that combining online recruitment with community engagement enhances the inclusiveness of the survey sample, thereby reducing selection bias and ensuring that underrepresented groups are adequately represented. This is particularly relevant for this study, which aims to understand pricing strategies across different consumer segments, including culturally specific shoppers who might not be well-represented through online-only recruitment.

The inclusion of both **vendor-specific** and **cultural factors** in the survey aligns with research that highlights the role of retail environments and cultural preferences in shaping consumer behavior. Literature on cultural consumption patterns suggests that different retail formats, such as mainstream versus culturally specialized stores, influence purchasing decisions based on familiarity, trust, and cultural alignment (Keller 1993). Therefore, the stratification variables and recruitment strategies are designed to capture these nuanced consumer preferences, providing a comprehensive understanding of how pricing dynamics differ across vendor types and cultural backgrounds.

Moreover, the focus on **vendor familiarity** builds on findings from **brand loyalty** literature, which suggests that consumers' familiarity with a brand or retailer can significantly impact their price sensitivity and shopping behavior (Ailawadi and Keller 2004). By including vendor familiarity as a key stratification criterion, the survey is structured to better understand how loyalty and brand perception contribute to the pricing dynamics of beef in different retail contexts.

In summary, the survey design leverages established theories and methodologies from consumer behavior, market segmentation, and cultural studies to ensure that the data collected is robust, representative, and capable of answering the core research questions about beef pricing dynamics in the Canadian retail market.

References

Ailawadi, Kusum L, and Kevin Lane Keller. 2004. "Understanding Retail Branding: Conceptual Insights and Research Priorities." *Journal of Retailing* 80 (4): 331–42.

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Anselmsson, Johan, Ulf Johansson, and Niklas Persson. 2007. "Understanding Price Premium for Grocery Products: A Conceptual Model of Customer-Based Brand Equity." *Journal of Product & Brand Management* 16 (6): 401–14.

Arel-Bundock, Vincent. 2023. *modelsummary: Data Tables and Model Summaries*. https://CRAN.R-project.org/package=modelsummary.

Bates, Douglas and Mächler, Martin and Bolker, Ben and Walker, Steve. 2023. *lme4: Linear Mixed-Effects Models using 'Eigen' and S4*. https://CRAN.R-project.org/package=lme4.

Chen, Xu, Shuyao Wu, Xiaojun Wang, and Dong Li. 2019. "Optimal Pricing Strategy for the Perishable Food Supply Chain." *International Journal of Production Research* 57 (9): 2755–68.

Collins, Jane. n.d. "WALMART, AMERICAN CONSUMER."

Elsbree, Claire. 2022. "Black Friday Pricing Behavior at Walmart."

Foedermayr, Eva K, and Adamantios Diamantopoulos. 2008. "Market Segmentation in Practice: Review of Empirical Studies, Methodological Assessment, and Agenda for Future Research." *Journal of Strategic Marketing* 16 (3): 223–65.

Goodrich, Ben and Gabry, Jonah and Ali, Iram and Brilleman, Sam. 2023. *rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm/.

Hii, Rebecca. 2007. "Changes in Chinese Supermarket Development in the Toronto CMA: A Case Study of t & t Supermarket." PhD thesis, Toronto Metropolitan University.

Ho, Richard, Leo Huang, Stanley Huang, Tina Lee, Alexander Rosten, and Christopher S Tang. 2009. "An Approach to Develop Effective Customer Loyalty Programs: The VIP Program at t&t Supermarkets Inc." *Managing Service Quality: An International Journal* 19 (6): 702–20.

Keller, KL. 1993. "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity." *The Journal of Marketing*.

McGoldrick, Peter J, and Helen J Marks. 1986. "How Many Grocery Prices Do Shoppers Really Know?" *Retail and Distribution Management* 14 (1): 24–27.

Müller, Kirill and Bryan, Jennifer. 2023. *here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Neal, Weston and Urbanek, Simon and others. 2023. *arrow: Integration to Apache Arrow*. https://CRAN.R-project.org/package=arrow.

Pilar Martínez Ruiz, María, and Alejandro Mollá Descals. 2008. "Retail Price Promotion Influences for Product Varieties in Grocery Stores: Evidence from Spain." *International Journal of Retail & Distribution Management* 36 (6): 494–517.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schepers, Bram. 2021. "Walmart, Always Low Prices?" PhD thesis.

Shannon, Randall, and Rujirutana Mandhachitara. 2005. "Private-Label Grocery Shopping Attitude and Behaviour: A Cross-Cultural Study." *Journal of Brand Management* 12 (6): 461–74.

Soysal, Gonca P, and Lakshman Krishnamurthi. 2012. "Demand Dynamics in the Seasonal Goods Industry: An Empirical Analysis." *Marketing Science* 31 (2): 293–316.

Subramaniam, Shasileka, Sagunthala Ragunathan, Mustaqim Bin Wali Abdul Khalid, B Tharumadurai Velayudhan, and Muhamad Huzaifi Faiq Bin Alias. 2024. "TO ENHANCING RECRUITMENT OUTREACH: STRATEGIC INTERVENTIONS FOR EXPANDED CANDIDATE ENGAGEMENT." In *INTERNATIONAL ACTION RESEARCH CONFERENCE (IARC) 2024 CONFERENCE PROCEEDINGS*, 2.

Wickham, Hadley. 2023a. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

———. 2023b. *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package*. https://CRAN.R-project.org/package=reshape2.

Wickham, Hadley and Bryan, Jennifer and others. 2023. *tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley and others. 2023a. *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

———. 2023b. *readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2023a. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. https://CRAN.R-project.org/package=knitr.

———. 2023b. *tinytex: Helper Functions to Install and Maintain 'TeX Live', and Compile 'LaTeX' Documents*. https://CRAN.R-project.org/package=tinytex.

Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'kable()'*. https://CRAN.R-project.org/package=kableExtra.