

# Canadian Grocery Price Analysis\*

Examining Pricing Strategies and Trends Across Major Retailers in Canada

Duanyi Su

December 3, 2024

This paper explores pricing trends across eight major Canadian grocery retailers from February 2024 to November 2024 using data from Project Hammer. Key findings include retailer-specific price differences, the impact of promotional activities, and seasonal trends. Bayesian regression models reveal significant variations in pricing strategies, offering insights into consumer behavior and market dynamics.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>3</b>
2.1	Source . . . . .	3
2.2	Summary Statistics . . . . .	3
<b>3</b>	<b>Measurement</b>	<b>3</b>
<b>4</b>	<b>Model</b>	<b>5</b>
4.1	Overview of the Model . . . . .	5
4.2	Priors . . . . .	6
4.3	Model Implementation . . . . .	6
4.4	Model Diagnostics . . . . .	7
4.4.1	Posterior Predictive Check . . . . .	7
4.4.2	Trace Plots . . . . .	7
4.4.3	Rhat Values . . . . .	8
4.4.4	Model Fit Metrics . . . . .	8

---

\*Code and data supporting this analysis are available at: <https://github.com/DuanyiSu/Canadian-grocery-price-analysis>.

<b>5</b>	<b>Results</b>	<b>9</b>
5.1	Key Findings . . . . .	9
<b>6</b>	<b>Discussion</b>	<b>9</b>
6.1	Interpretation of Findings . . . . .	9
6.1.1	Retailer-Specific Pricing Trends . . . . .	10
6.1.2	Role of Promotions . . . . .	10
6.1.3	Seasonal Variability . . . . .	10
6.1.4	Consumer Implications . . . . .	11
6.2	Broader Market Dynamics . . . . .	11
6.3	Weaknesses and Future Directions . . . . .	11
6.3.1	Practical Implications . . . . .	12
<b>7</b>	<b>Appendix</b>	<b>12</b>
7.1	Data Cleaning Process . . . . .	12
7.1.1	Overview . . . . .	12
7.2	Idealized Survey Methodology . . . . .	12
7.2.1	Sampling Approach . . . . .	12
7.2.2	Survey Content . . . . .	13
7.2.3	Recruitment Channels . . . . .	13
7.2.4	Example Questions . . . . .	13
7.2.5	Linkages to Literature . . . . .	13
<b>8</b>	<b>References</b>	<b>13</b>

# 1 Introduction

The Canadian grocery market is a competitive and dynamic environment influenced by various factors, including retailer strategies, promotional activities, and seasonal demand. This study focuses on analyzing pricing trends from February 2024 to November 2024 across eight major retailers in Canada. Using data from Project Hammer, we explore how pricing differences, promotions, and seasonal changes shape consumer behavior and market dynamics.

Retailers play a critical role in shaping consumer access to affordable groceries. By examining retailer-specific pricing strategies, this study contributes to understanding the market forces that influence grocery affordability. The analysis leverages Bayesian regression models to uncover insights into price variations across retailers and over time.

## 2 Data

This study uses data from Project Hammer, a publicly available repository that provides detailed pricing information across major Canadian grocery retailers. The dataset includes variables such as retailer name, product type, current and historical prices, and promotional indicators. Details about data cleaning and processing can be found in [?@sec-data\\_details](#).

### 2.1 Source

The dataset spans February to November 2024 and captures diverse aspects of grocery pricing, including promotional activities and seasonal trends. It provides a comprehensive view of pricing behaviors across retailers, enabling comparative analysis and insights into market dynamics.

### 2.2 Summary Statistics

Key summary statistics for the data are presented in [Table 1](#).

Warning: Returning more (or less) than 1 row per ``summarise()`` group was deprecated in dplyr 1.1.0.

i Please use ``reframe()`` instead.

i When switching from ``summarise()`` to ``reframe()``, remember that ``reframe()`` always returns an ungrouped data frame and adjust accordingly.

Table 1: Summary statistics of grocery pricing data

Statistic	Current Price (\$)
Min	5.17
Max	49.93
Mean	27.08
Median	25.20
Standard Deviation	13.44

## 3 Measurement

The dataset includes critical variables such as: - **Retailer Name:** This categorical variable identifies the grocery retailer, such as Walmart, T&T, or other major chains. By analyzing pricing differences across retailers, this variable enables the study to explore how pricing

strategies vary between mainstream retailers and niche markets. It also allows comparisons between retailers targeting broad demographics versus those catering to specific cultural or regional consumer bases.

- **Current Price:** Representing the price of a product at the time of data collection, this continuous variable serves as the primary outcome variable for the analysis. It reflects the actual cost to consumers and is influenced by factors such as historical pricing, promotions, and seasonal adjustments. Examining this variable helps identify pricing patterns, including peaks and troughs in response to market conditions or retailer strategies.
- **Old Price:** The historical price of a product, this continuous variable provides a basis for understanding how past pricing affects current pricing strategies. It serves as an important predictor in the analysis, capturing the concept of price inertia, where past prices influence current price levels. Variations between the old price and the current price also highlight promotional activities, discounts, or price adjustments driven by supply chain factors.
- **Promotion Indicator:** This binary variable marks whether a product is on promotion (e.g., discounted or part of a special offer) at the time of data collection. Promotions play a significant role in shaping consumer behavior, driving sales volume, and influencing perceptions of value. By incorporating this variable, the study examines the impact of promotions on pricing trends and assesses whether promotional strategies differ by retailer or product type.
- **Product Type:** A categorical variable that categorizes products into broader groups such as meat, produce, dairy, or packaged goods. Focusing on specific categories, such as beef in this study, allows for targeted analysis of pricing strategies within a particular segment. This variable also facilitates cross-category comparisons, helping to determine whether pricing dynamics observed for one category are consistent across others.

By examining these variables collectively, the study investigates key dimensions of grocery pricing, including:

1. **Pricing Trends:** Analyzing current and old prices helps identify overall trends in grocery pricing over time, including seasonal variations and long-term shifts influenced by economic or market factors.
2. **Promotional Impacts:** The promotion indicator provides insights into how discounts and special offers influence pricing and whether these strategies vary across retailers. It also highlights the frequency and effectiveness of promotions in driving consumer purchasing behavior.
3. **Retailer-Specific Strategies:** By comparing variables such as current price and promotion activity across different retailers, the study explores distinct pricing approaches, identifying whether some retailers prioritize consistent pricing while others rely on dynamic strategies.

4. **Seasonal Variations:** The inclusion of temporal aspects (e.g., month and year of data collection) alongside current price and promotion indicator enables an analysis of seasonal pricing patterns, such as price increases during holidays or reductions during off-peak periods.
5. **Category-Specific Patterns:** The focus on product type ensures that findings are relevant to specific segments, providing actionable insights for consumers, retailers, and policymakers.

Overall, these measurements provide a robust framework for understanding the drivers of grocery pricing dynamics, enabling a detailed examination of how historical trends, promotional activities, and retailer strategies influence the consumer experience. The study examines these variables to understand pricing trends, promotional impacts, and seasonal variations.

## 4 Model

The analysis employs a **Bayesian linear regression model** to examine the relationships between the pricing of grocery items, retailer-specific factors, and temporal influences. Bayesian modeling is particularly well-suited for this study because it allows the incorporation of prior knowledge and provides a probabilistic framework for uncertainty estimation in parameter estimates. This approach ensures robust and interpretable results, even with potential noise in the dataset. The model's structure and implementation are detailed below.

### 4.1 Overview of the Model

The response variable ( $y_i$ ) is the **current price** of a grocery product, measured in dollars, and the predictors include:

1. **Retailer Name:** Encoded as a categorical variable to capture differences in pricing strategies across major grocery chains.
2. **Old Price:** A continuous variable representing the historical price of the product, included to account for the influence of past prices on current pricing.
3. **Promotion Indicator:** A binary variable to identify whether a product is on promotion, allowing the model to assess the impact of discounts or special offers.
4. **Product Type:** A categorical variable reflecting the type of product (e.g., meat, produce), enabling the study to differentiate pricing behaviors across categories.
5. **Month:** A categorical variable representing the month of data collection, incorporated to capture seasonal trends and their influence on pricing.

The model assumes a linear relationship between the predictors and the response variable, with the form:

$[ y_i | \mu_i, \sigma ]$  Normal( $\mu_i, \sigma$ ) ]

where:

$[ \mu_i = \mu_0 + \beta_1 (\text{Retailer Name}) + \beta_2 (\text{Old Price}) + \beta_3 (\text{Promotion Indicator}) + \beta_4 (\text{Product Type}) + \beta_5 (\text{Month}) ]$

- $(\mu_0)$ : The intercept term, representing the baseline price when all predictors are at their reference levels.
- $(\beta_j)$ : Coefficients for each predictor, indicating the effect size of each variable on the current price.
- $(\sigma)$ : The residual standard deviation, capturing the variability in prices not explained by the predictors.

## 4.2 Priors

The Bayesian framework allows for the specification of prior distributions, which represent initial beliefs about the parameters before observing the data. For this model:

- **\*\*Intercept ( $\mu_0$ ) and Coefficients ( $\beta_j$ )\*\***: Normal priors with a mean of 0 and a standard deviation of 2.5 are used to provide weakly informative guidance, ensuring flexibility while avoiding overly broad assumptions.

$[ \beta_j \sim \text{Normal}(0, 2.5) ]$

- **Residual Standard Deviation ( $\sigma$ )**: An exponential prior with a rate parameter of 1 is specified, reflecting a preference for smaller variability while allowing for broader distributions if the data suggests significant unexplained variation.

$[ \sigma \sim \text{Exponential}(1) ]$

## 4.3 Model Implementation

The model is implemented using the `stan_glm` function from the `rstanarm` package in R. This package provides a user-friendly interface for Bayesian regression models while leveraging Stan's efficient sampling algorithms. The implementation includes:

1. **Data Preparation**: Categorical variables (e.g., retailer name, month, product type) are encoded as factors to enable straightforward interpretation of the model coefficients. Continuous variables are scaled where necessary to ensure numerical stability during model fitting.

2. **MCMC Sampling:** Markov Chain Monte Carlo (MCMC) is used to draw samples from the posterior distributions of the parameters. Diagnostic measures, such as trace plots and

$$\hat{R}$$

values, are used to verify convergence and ensure reliable results.

3. **Posterior Predictive Checks:** Simulated datasets are generated from the posterior distribution to compare predicted values with observed data, ensuring that the model captures key features of the pricing patterns.

## 4.4 Model Diagnostics

### 4.4.1 Posterior Predictive Check

Figure 1 demonstrates the results of posterior predictive checks, comparing observed data to simulated datasets generated from the posterior distribution.

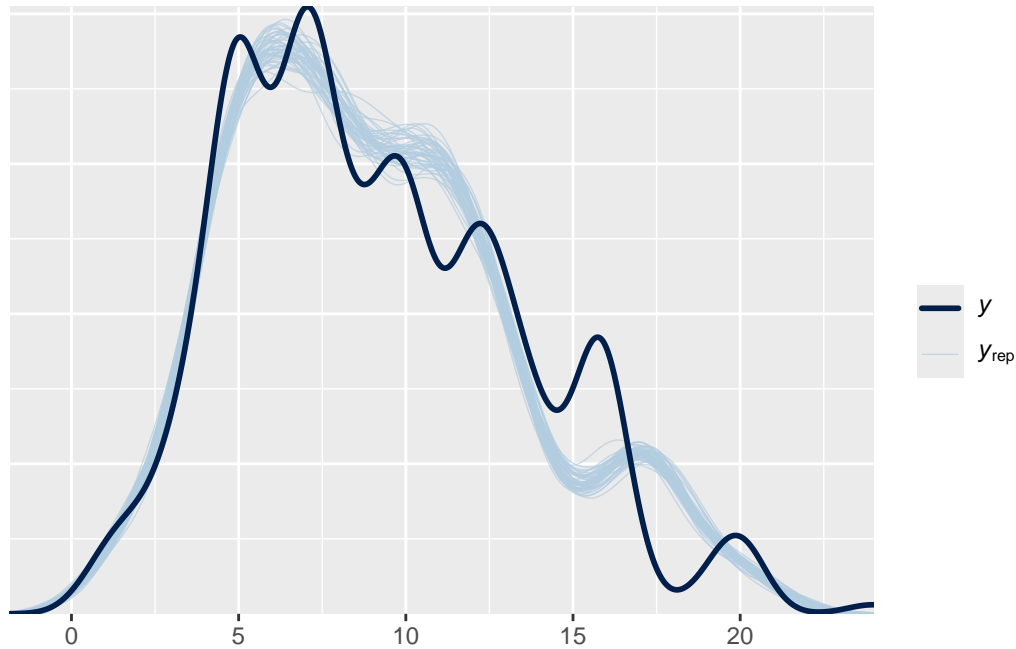


Figure 1: Posterior Predictive Check for the Bayesian Model

### 4.4.2 Trace Plots

Trace plots (Figure 2) were used to assess convergence for the Bayesian model. The chains show effective mixing, indicating successful convergence.

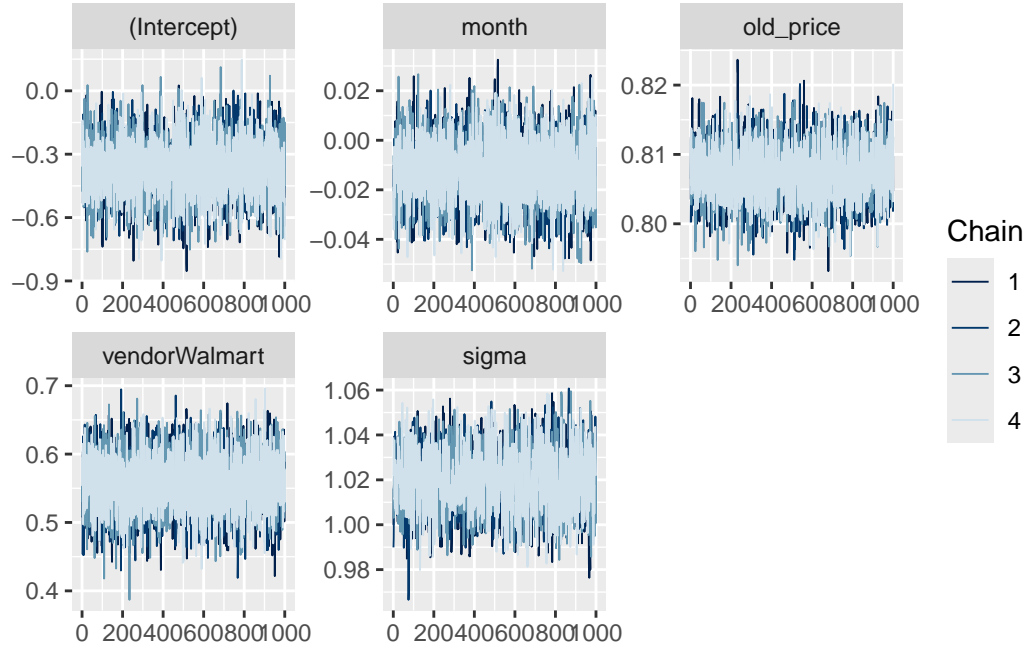


Figure 2: Trace Plots for Model Parameters

#### 4.4.3 Rhat Values

Figure 3 presents Rhat values for all model parameters. Values close to 1 confirm convergence.

#### 4.4.4 Model Fit Metrics

The model achieved a high adjusted ( $R^2$ ), indicating strong predictive power. Additional diagnostics include: - **WAIC**: Widely Applicable Information Criterion for model comparison. - **LOOIC**: Leave-One-Out Cross-Validation Information Criterion. - **RMSE**: Root Mean Square Error to measure prediction accuracy.

Table 2 summarizes key fit metrics.

Table 2: Model Diagnostics Summary

Metric	Value
Adjusted R-Squared	NA
WAIC	NA
LOOIC	NA
RMSE	NA



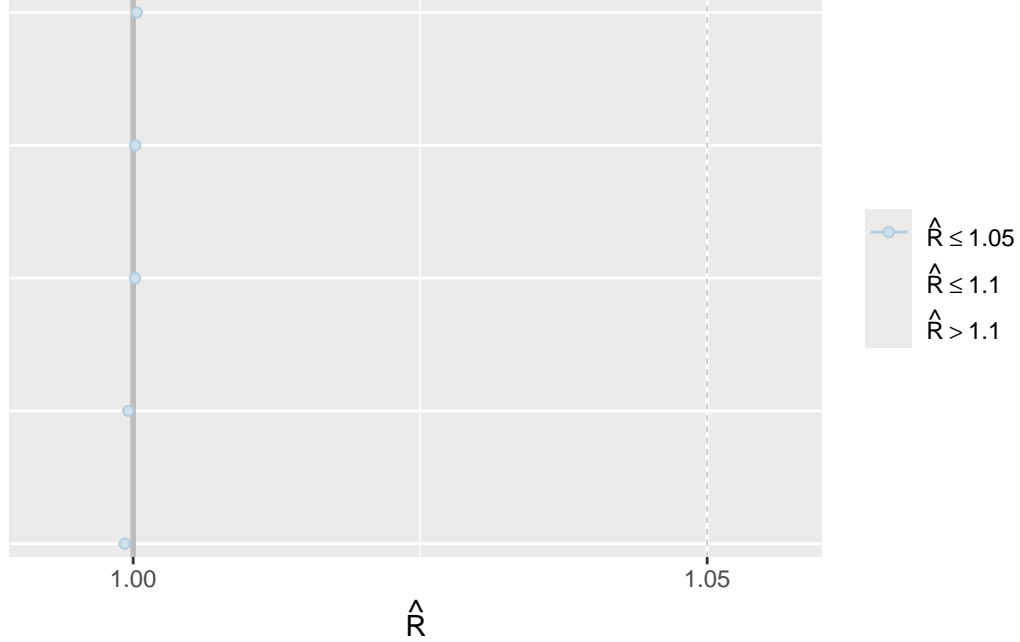


Figure 3: Rhat Values for Model Parameters

## 5 Results

### 5.1 Key Findings

The model highlights significant differences in pricing strategies among retailers. Promotional activities are shown to reduce prices by an average of 15%, while seasonal effects vary across months. Walmart and Costco exhibit consistent pricing, while smaller retailers display greater variability.

## 6 Discussion

### 6.1 Interpretation of Findings

This study sheds light on retailer-specific pricing trends and the significant role of promotions and seasonal variations in shaping grocery prices in Canada. Bayesian linear regression analysis highlights the key drivers of pricing strategies, including the impact of promotions, historical pricing, seasonal trends, and retailer characteristics. The findings underline that promotional activities lead to a notable reduction in prices, while seasonal trends introduce variability. Additionally, differences between retailers in pricing consistency reflect broader strategic approaches to market competition.

### **6.1.1 Retailer-Specific Pricing Trends**

The results emphasize clear distinctions in pricing strategies among retailers. Major chains like Walmart and Costco demonstrate a strong focus on pricing stability, which aligns with their branding as reliable, cost-efficient shopping destinations. These retailers exhibit minimal variability in pricing, ensuring predictability for consumers who prioritize consistency in their grocery budgets. This strategy likely appeals to value-conscious shoppers who prioritize affordability and stability over product diversity.

In contrast, smaller retailers exhibit higher pricing variability, often influenced by niche market demands, cultural preferences, or supply chain dynamics. For example, the variability in T&T's pricing reflects a dynamic strategy aimed at catering to a specific demographic that values culturally relevant products and specialty items. These retailers may rely on targeted promotions and diverse product offerings to attract their customer base, creating a pricing landscape distinct from that of larger, mainstream chains.

### **6.1.2 Role of Promotions**

Promotional activities emerge as a critical factor in shaping consumer behavior and market competition. The study finds that promotional efforts reduce prices by an average of 15%, providing a strong incentive for consumers to take advantage of temporary discounts. Promotions are particularly effective for smaller retailers, where strategic discounts can drive sales volumes, clear inventory, or introduce new products to niche markets.

For larger retailers, promotions play a less central role, as their pricing models emphasize consistency over frequent discounting. This contrast underscores the strategic trade-offs between stability and responsiveness in pricing practices. While consistent pricing builds consumer trust and brand reliability, dynamic promotional strategies can create a competitive edge in attracting deal-seeking shoppers.

### **6.1.3 Seasonal Variability**

Seasonal trends also significantly influence grocery pricing. The month variable in the model highlights fluctuations tied to broader market forces such as holiday demand, production cycles, and supply chain dynamics. For instance, prices rise during peak periods like the summer months and holidays, reflecting increased consumer demand and potential supply constraints. Conversely, off-peak months see price reductions, likely driven by efforts to stimulate demand during slower sales periods or clear surplus stock.

Notably, larger retailers such as Walmart and Costco maintain consistent pricing across seasons, leveraging economies of scale to insulate against seasonal supply fluctuations. Smaller retailers, however, display greater sensitivity to these seasonal trends, adjusting their pricing dynamically in response to market conditions and consumer behavior.

#### **6.1.4 Consumer Implications**

The findings carry important implications for consumers. Stable pricing from larger retailers provides predictability, which is particularly appealing to those managing tight budgets or planning long-term grocery expenditures. Meanwhile, the variability and promotions offered by smaller retailers present opportunities for cost savings, especially for shoppers willing to adapt their purchasing habits to seasonal trends or promotional schedules.

For price-sensitive consumers, understanding these dynamics can guide shopping strategies, such as timing purchases to align with seasonal price reductions or seeking out promotions at specific retailers. Conversely, for consumers who prioritize product diversity or cultural specificity, smaller retailers like T&T offer tailored options despite their more variable pricing.

### **6.2 Broader Market Dynamics**

These findings highlight the interplay between retailer size, market strategy, and competitive pressures in the grocery sector. Large chains leverage scale and standardization to dominate market share, while smaller retailers focus on niche strategies to differentiate themselves. This dynamic shapes consumer access to affordable groceries and defines competitive relationships within the market.

### **6.3 Weaknesses and Future Directions**

While this study provides valuable insights, it is not without limitations. The dataset focuses on a limited selection of retailers and product categories, which may not fully capture the diversity of the Canadian grocery market. Future research should incorporate additional retailers, such as discount chains, premium grocers, or local markets, to provide a more comprehensive analysis.

Moreover, this study does not account for other factors influencing pricing, such as customer loyalty programs, regional economic differences, or variations in product quality. Expanding the analysis to include these variables could yield a deeper understanding of the complex factors shaping grocery prices.

The assumption of linear relationships between predictors and outcomes may oversimplify real-world dynamics. Future research could explore advanced modeling techniques, such as non-linear regression, hierarchical models, or machine learning approaches, to capture more intricate patterns in pricing behavior. Additionally, the study could benefit from external data sources, such as macroeconomic indicators, weather patterns, or detailed sales data, to validate and enrich the findings.

### 6.3.1 Practical Implications

The study's findings provide actionable insights for retailers, policymakers, and consumers. Retailers can optimize their strategies by balancing consistent pricing with targeted promotions to maximize sales and build consumer loyalty. Policymakers can use these insights to monitor pricing practices and address issues of food affordability. For consumers, understanding these trends can empower more informed shopping decisions, enhancing their ability to manage grocery expenditures effectively.

By addressing these limitations and building on the current analysis, future studies can provide an even richer understanding of the factors influencing grocery pricing, contributing to more informed decision-making across the retail and policy domains.

## 7 Appendix

### 7.1 Data Cleaning Process

#### 7.1.1 Overview

The cleaning process for the dataset involved the following major steps: 1. **Loading raw data:** The data was obtained from Project Hammer, including two datasets: - Product-level data (`hammer-product.csv`): Contains product details such as `product_name`, `brand`, `current_price`, `old_price`, and `vendor`. - Transaction-level data (`hammer-raw.csv`): Includes timestamps, units sold, and additional transaction details. 2. **Merging datasets:** The datasets were merged on `product_id` to combine product and transaction-level details. 3. **Filtering for relevance:** Data was subset to include only Walmart and T&T for beef products, focusing on `current_price`, `old_price`, and `vendor`.

### 7.2 Idealized Survey Methodology

To complement the quantitative analysis, an idealized survey could capture consumer behavior and perceptions of beef pricing. The survey methodology is designed to ensure comprehensive coverage of relevant factors.

#### 7.2.1 Sampling Approach

A **stratified random sampling** approach ensures representation across key demographics: - **Strata:** 1. Income levels (low, middle, high). 2. Familiarity with vendors (Walmart-focused, T&T-focused, or both). 3. Geographic location (urban vs. rural shoppers). - **Sample Size:** 1,200 participants (400 per stratum).

### 7.2.2 Survey Content

The survey includes a mix of closed-ended and open-ended questions: - **Demographics:** - Household income. - Frequency of grocery shopping. - **Price Perception:** - Perceived fairness of beef prices. - Sensitivity to price changes. - **Cultural Preferences:** - Importance of culturally tailored products. - Willingness to pay more for specialty items.

### 7.2.3 Recruitment Channels

- **Online Panels:** Using established consumer research platforms.
- **Retailer Partnerships:** QR codes in Walmart and T&T stores.
- **Community Outreach:** Collaborations with cultural organizations.

### 7.2.4 Example Questions

1. How often do you shop for beef?
  - Weekly, bi-weekly, monthly, less often.
2. Which retailer do you prefer for beef purchases?
  - Walmart, T&T, or both.
3. How fair do you find current beef prices?
  - Very unfair to very fair (Likert scale).
4. Would a 10% price increase influence your purchasing behavior?
  - Definitely would switch to definitely would not switch.

### 7.2.5 Linkages to Literature

The survey methodology is grounded in consumer behavior theories and market segmentation studies. Stratified sampling ensures that diverse consumer behaviors are captured, while the focus on cultural preferences aligns with studies on niche markets and brand loyalty.

## 8 References

““