Unpacking Breakfast: A Vendor-Based Analysis of Bacon, Eggs, and Bagels*

Jason Yang

December 3, 2024

This paper examines Canadian grocery prices for breakfast items (bacon, brown eggs, and bagels) across eight major vendors to identify which vendor offers the lowest prices and most frequent discounts. Using a dataset collected via web scraping, the study employs regression modeling to analyze vendor-specific prices, discount patterns, and seasonal trends. The findings reveal that TandT and Walmart consistently offer lower prices, while Metro provides the most frequent discounts, making it a strategic choice for cost-conscious consumers. These insights are valuable for helping families minimize grocery costs and informing policymakers aiming to foster competition and fairness in the Canadian grocery market.

Table of contents

*Code and data are available at: (https://github.com/jasonbot123/Canadian-Grocery-Price-Analyze).

1 Introduction

Grocery prices are a significant concern for consumers, particularly in light of recent economic challenges such as inflation and supply chain disruptions. In Canada, the rising cost of food has sparked public debate and policy discussions about the competitiveness of grocery vendors and their pricing strategies. For everyday breakfast staples like bacon, brown eggs, and bagels, understanding how vendors price their products and offer discounts is crucial for empowering consumers to make cost-effective choices. While data on grocery pricing exists, it often lacks the granularity needed to provide actionable insights. This study aims to address these challenges by leveraging data from Project Hammer, a comprehensive dataset of breakfast item prices collected across eight major Canadian grocery vendors.

This paper focuses on identifying which grocery vendor offers the lowest prices for breakfast items, analyzing pricing trends, discount patterns, and vendor-specific pricing strategies. Using a combination of observational data collected through web scraping and predictive modeling, this research fills a critical gap in the literature by providing a detailed comparison of grocery vendors' pricing behaviors. While existing studies often focus on general grocery pricing, few examine specific product categories or evaluate the role of discounts and seasonal trends in vendor choice. By narrowing the scope to breakfast items, this paper provides a focused and practical analysis relevant to both consumers and policymakers.

To achieve this, the study employs regression modeling to predict prices based on factors such as vendor, old price, discount status, and temporal trends. The analysis reveals key insights, including which vendors consistently offer lower prices, how discounts impact consumer costs, and whether seasonal variations affect pricing. The findings indicate that vendors like TandT, Walmart, and Metro generally provide lower prices for breakfast staples, while Save-On-Foods tends to charge a premium. Additionally, Metro emerges as the vendor with the highest discount frequency, making it a strategic choice for budget-conscious consumers.

The results of this study are important for several reasons. First, they offer practical recommendations for families looking to minimize grocery costs. Second, the analysis provides evidence to support policy interventions aimed at increasing competition and reducing potential collusion in the Canadian grocery sector. Lastly, the methodological framework, which integrates observational data with predictive modeling, can be applied to other product categories or markets.

1.1 Estimand paragraph

The estimand in this paper is the prices of typical breakfast items. However, it is challenging to determine exactly what people regularly consume for breakfast, given the diversity of dietary preferences and cultural practices. To narrow the focus, this paper analyzes three widely recognized breakfast staples: brown eggs, bacon, and bagels. These items were selected because they represent a balance of commonality, affordability, and versatility in breakfast meals. By focusing on these products, the paper aims to provide insights into pricing patterns, vendor competition, and promotional strategies for breakfast essentials across the Canadian grocery market.

2 Data

2.1 Data Source

This report leverages the dataset provided by Jacob Filipp for Project Hammer, focusing on analyzing grocery pricing trends„ vendor competition, and product availability. The dataset contains detailed information on products sold by Viola, T&T, Loblaws, No Frills, Metro, Galleria„ Walmart and Save-On-Foods. The dataset spans over a time period from February 28, 2024, to the latest load on November 26, 2024. The constant updated dataset allows for a robust time-based analysis of trends in grocery pricing, promotional activities, and stock availability.

2.1.1 Features

The original Canadian Grocery Price Data is provided in two CSV files: "hammer-4-product" and "hammer-4-raw". The "hammer-4-product" file contains metadata and product details, including variables such as id, vendor, concatted, units, product_name, and brand. The "hammer-4-raw" file contains time-series price data, with variables including nowtime, current_price, old_price, price_per_unit, other, and product_id. The two datasets were combined using the product_id variable, resulting in a merged dataset with 12,842,742 observations. The final dataset includes the following variables: nowtime, vendor, product_id, product_name, brand, units, current_price, old_price, price_per_unit, and other. The definition for these variables are shown in the appendix table (Table 1).

2.2 Data Measurement

This information was obtained through a screen-scrape of website user interfaces, which limits the available data to what is displayed publicly. As a result, some information that might be accessible through internal APIs, such as detailed inventory data or more granular time stamps, is missing.

The unit of measurement for numerical data, such as current_price and old_price, is in Canadian Dollars (CAD). For price_per_unit, the units vary depending on the product type (e.g., per kilogram, gram, pound, or per item). Other data, such as vendor and product_name, are categorical features with no numerical units.

2.3 Methodology

The original dataset was provided in two CSV files, which were merged based on the columns product_id and id, as these columns uniquely identify individual products. The merged dataset that contains 12842742 observations is then filtered into three separate datasets for bacon, brown eggs, and everything bagel using the dplyr package (Wickham et al. (2023)).

To address the issue of filtering product_name to include only bacon-related products (as this approach might inadvertently include items that contain bacon but are not the meat itself), we refined our method by excluding specific keywords that could misrepresent the data. These keywords include "dressing," "seasoning," "flavour," "snack," "Caesar," and similar terms. This approach was similarly applied to the brown egg dataset and the everything bagel dataset to ensure consistency and accuracy.

The cleaned data was written and read using the write_parquet and read_parquet functions from the arrow library (Richardson et al. (2024)). These functions enable efficient storage and retrieval of large datasets in the Parquet format, which is optimized for performance and storage. Other libraries like tidyverse (Wickham et al. (2019)), ggplot2 (Wickham (2016)), rstanarm (Goodrich et al. (2022)), here (Müller (2020)), modelsummary (Arel-Bundock (2022)), broom (Robinson, Hayes, and Couch (2024)), kableExtra (Zhu (2024)), and Metrics (Hamner and Frasco (2018)) was used for statistical analyzing.

2.4 Data Visualization

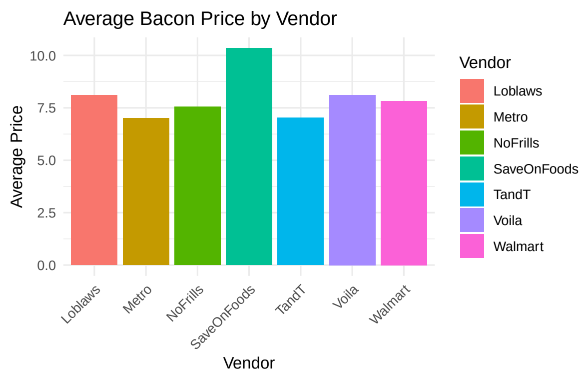


Figure 1: Average Bacon Price by Vendors.

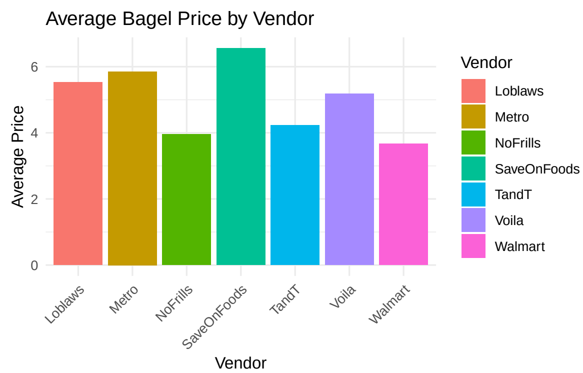


Figure 2: Average Bagels Price by Vendors.

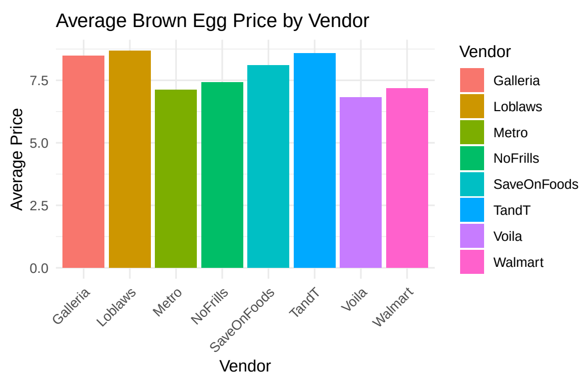


Figure 3: Average Brown Eggs Price by Vendors.

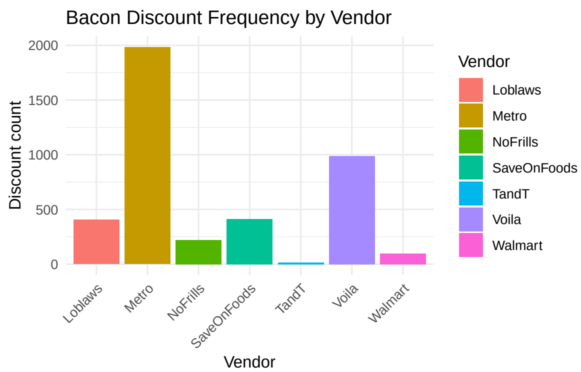


Figure 4: Discounts Freq for Bacon by vendors.

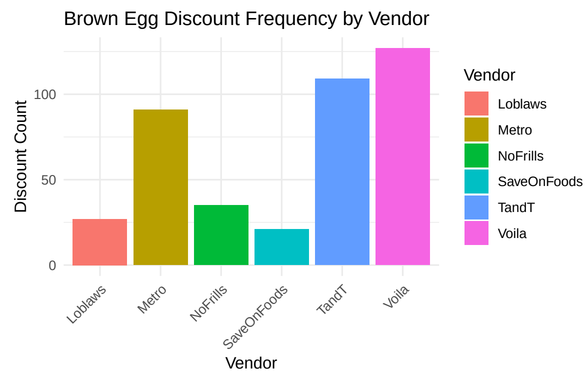


Figure 5: Discounts Freq for Brown Eggs by vendors.

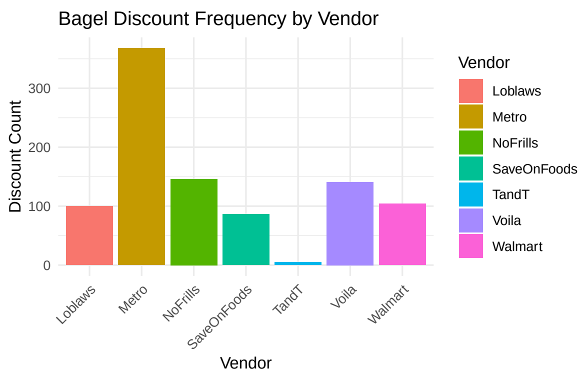


Figure 6: Discounts Freq for Bagels by vendors.

Figure 7: Bagel Trend for Current and Old Prices.

Figure 8: Bacon Trend for Current and Old Prices

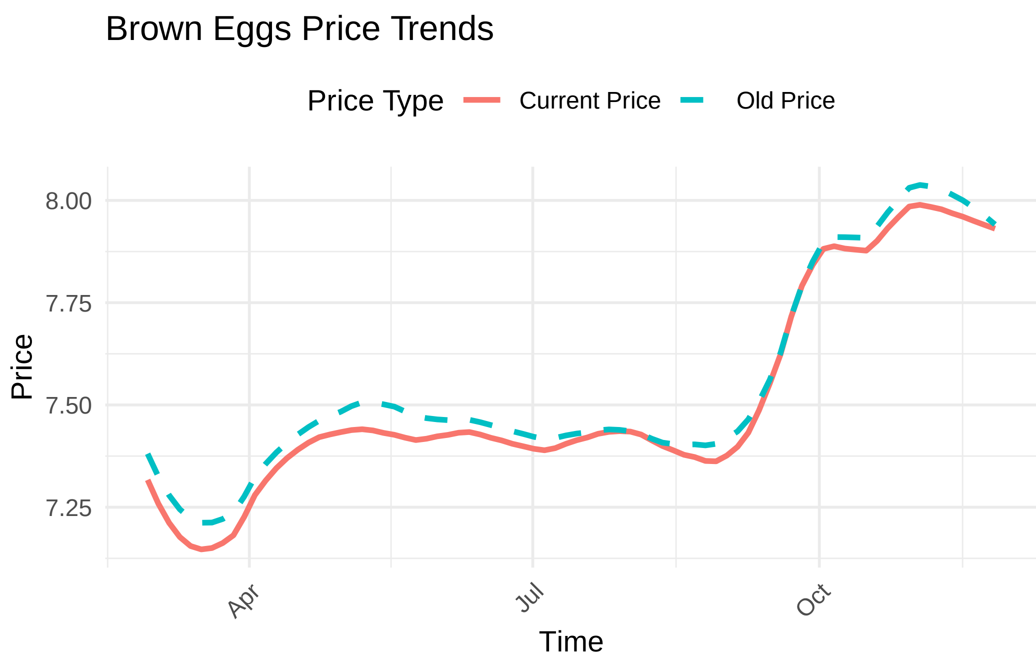


Figure 9: Brown Eggs Trend for Current and Old Prices.

3 Model

In our analysis, we utilized a linear regression model to predict the price of breakfast items based on predictor variables such as vendor, old_price, discount status, and month.

3.1 Model set-up

The model is formulated as follows:

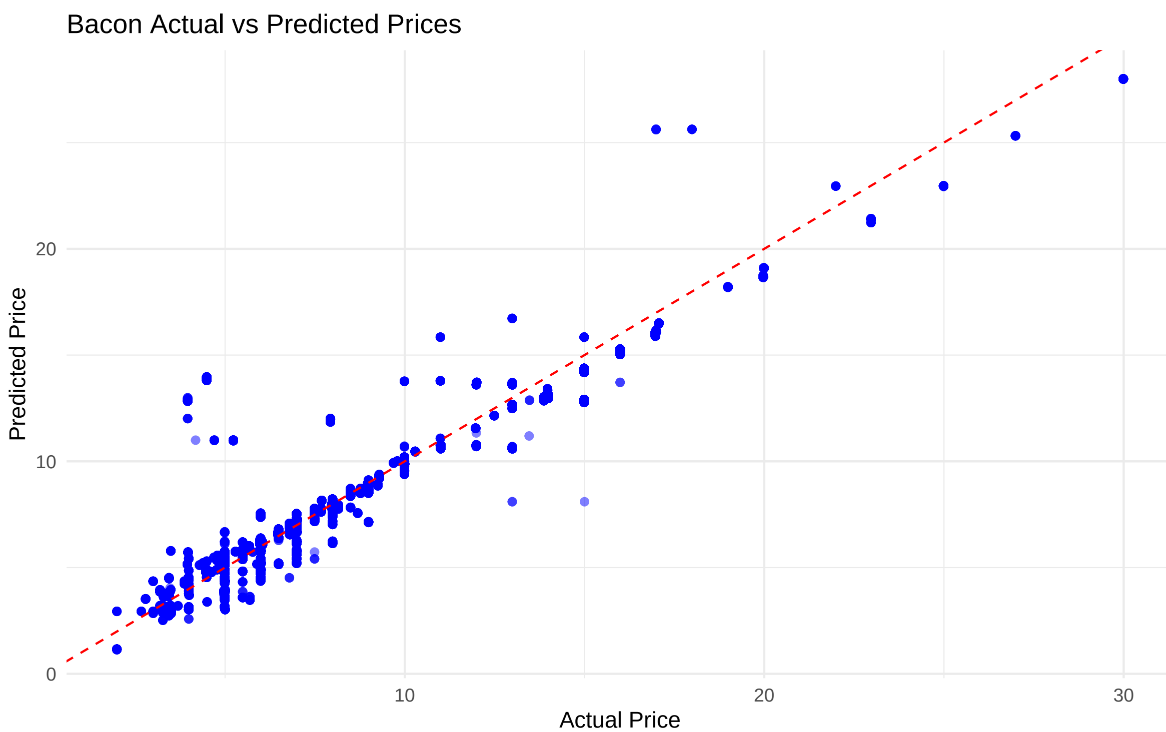
$$
\begin{aligned}
y_i = \beta_0 &+ \beta_1(\text{vendor}) + \beta_2(\text{old\_price}) && (1)\\
&+ \beta_3(\text{discounted}) + \beta_4(\text{month}) + \epsilon && (2)
\end{aligned}
$$

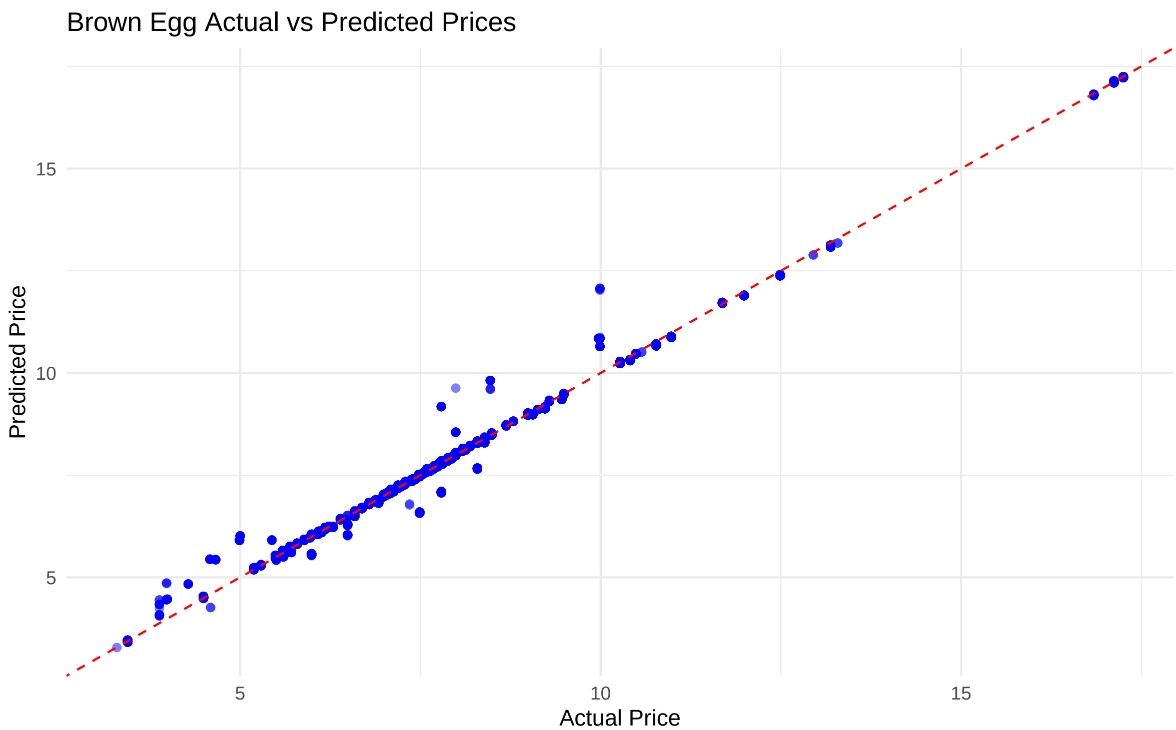
In this model, y_i represents the current_price of breakfast items, which is the dependent variable we aim to predict and analyze. This regression model estimates the current_price of breakfast items using key product and vendor characteristics as predictors. The model includes four predictor variables: vendor, old_price, month, and discounted.

The variable vendor captures pricing variations across different grocery chains, allowing for an evaluation of vendor-specific pricing. old_price, which represents the previous price of the product, is a crucial predictor for understanding the relationship between historical and current pricing trends. The variable month, a time-based factor, accounts for potential monthly pricing patterns, such as increased prices during holiday seasons or seasonal promotions. Lastly, the binary variable discounted indicates whether an item is currently on sale, capturing the average price reduction associated with discounts.

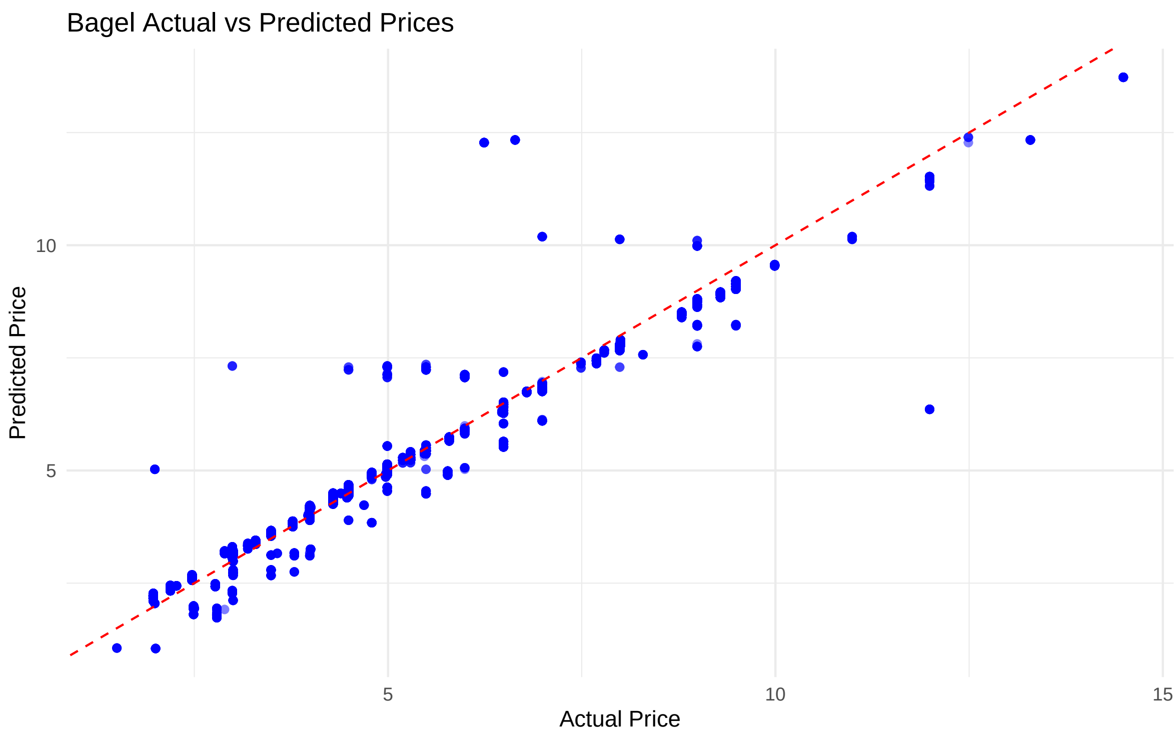
This regression model not only helps to predict prices but also serves as a tool for understanding pricing dynamics in the grocery market. It offers actionable insights into how vendors, historical prices, time factors, and promotional discounts influence the cost of breakfast essentials.

The model is implemented using R (R Core Team (2023)), utilizing the lm() function for linear regression. Preprocessing of the data is performed using the dplyr package. Visualizations and diagnostics are created using ggplot2.





Brown Egg Actual vs Predicted Prices
Predicted Price
15
10
5
5
10
15
Actual Price



Bagel Actual vs Predicted Prices
Predicted Price
10
5
5
10
15
Actual Price

3.2 Model justification

A regression model is appropriate for this analysis because it allows us to quantify the relationships between the response variable (current_price) and multiple predictor variables (vendor, old_price, month, and discounted). Regression analysis is particularly well-suited for predicting continuous outcomes, such as prices, while simultaneously evaluating the impact of multiple predictors. By using this approach, we can not only predict the current_price of breakfast items but also understand the relative importance of the factors driving price variation.

By using a linear regression model, we are assuming there is a linear relationship between current_price and predictors. While this may oversimplify certain relationships, it provides the general baseline for this analysis. Additionally, the observation are assumed to be independent, meaning the collection of one observation does not affect the chances of another observation being chosen. Lastly, the residuals are assumed to be normally distributed with constant variance.

In the raw data, many old_price values were empty or NA due to the absence of sales. During the data cleaning process, these missing old_price values were replaced with the corresponding current_price values. This decision ensures consistency and preserves the observations for analysis. For our assumptions, we treat these replaced values as representative for predictive analysis, assumptions stands. This approach aligns with the primary goal of the study: to identify the months and vendors offering the lowest prices for breakfast items. By retaining these observations, the model can provide a comprehensive understanding of pricing trends across different vendors and time periods while maintaining a sufficient sample size for analysis.

4 Results

The results for the regression models analyzing the prices of bacon, brown eggs, and bagels are summarized in Tables Table 2, Table 3, and Table 4, respectively.

The regression model for bacon indicates that old_price, vendor, and discounted significantly influence the current_price. For the regression model, the coefficient produced for each variable shows that old_price indicates a strong positive relationship between old_price and current_price.

The impact of discounts (discounted) was also significant in all models, with discounts consistently associated with lower prices. In the bacon model, discounted products were priced \$2.367 lower on average, while discounts reduced brown egg prices by \$1.056 and bagel prices by \$1.331. These findings emphasize the substantial role of discounts in determining grocery pricing.

Vendor effects varied across products. For bacon, Metro, NoFrills, TandT, Voila, and Walmart had significantly lower prices compared to the baseline vendor, with price reductions ranging from \$0.420 for Metro to \$0.135 for Walmart. Save-On-Foods, however, had higher prices, with a coefficient of 0.264. Similar trends were observed in the brown egg and bagel models, where TandT and Walmart consistently exhibited lower prices relative to the baseline vendor. Save-On-Foods continued to charge slightly higher prices for brown eggs, while its effect for bagels was not significant. This is reflected by the Figure 1, 2, and 3 as it shows Save-On-Foods has an higher average price across the three breakfast items compared to other vendors.

The variable month was included in the models to capture potential seasonal trends. For bacon and bagels, none of the monthly indicators were statistically significant, suggesting no meaningful seasonal variations in prices. However, in the brown egg model, certain months, such as May and October, showed small but significant price reductions, indicating minor seasonal effects on egg pricing.

In summary, the regression models demonstrate that old prices and discounts are the strongest and most consistent predictors of current prices across all three product categories. Vendor effects reveal significant price variations, with TandT and Walmart generally offering lower prices and Save-On-Foods often charging a premium. While seasonal effects were minimal for bacon and bagels, slight trends were observed in egg pricing, particularly in certain months.

5 Discussion

5.1 Which Grocer to Choose?

Which grocer is generally the cheapest for breakfast items like Bacon, Brown Eggs, and Bagels? Where should families buy breakfast items for the lowest price? Based on simple data visualization on these items, we can see that they are on average around the same price across the 8 vendors, with slight variation. However, for Bacon and Bagels, Save-On-Foods consistently exhibits higher average prices for bacon and bagels compared to other grocers. In contrast, TandT, WalMart, and NoFrills tend to offer lower prices, making them more budget-friendly options for families purchasing breakfast staples.

Additionally, Metro tends to offer the most deals for breakfast items compared to other grocers, as reflected in Figures 4, 5, and 6. Metro has a significantly higher discount frequency for bacon and bagels than other grocers, making it an excellent choice for budget-conscious families looking for deals on these items. For brown eggs, however, Voila and TandT show more frequent discounts than Metro.

That said, if families prefer to avoid the hassle of shopping at multiple grocers for different items, Metro emerges as the best all-around option. It combines frequent discounts with competitive average prices across bacon, brown eggs, and bagels, making it a convenient and cost-effective choice for families seeking to save on breakfast staples.

These claims are further supported by the regression models, which indicate that Metro and other vendors such as TandT and Walmart consistently had significantly lower prices compared to grocers like Loblaws and Save-On-Foods. For instance, in the regression results for bacon and bagels, Metro demonstrated significantly reduced prices relative to the baseline vendor. Meanwhile, Save-On-Foods exhibited higher average prices, particularly for bacon and bagels, reaffirming its role as a more premium-priced option.

Moreover, the model highlights the influence of discounts on pricing, with discounted items showing substantial price reductions across all products. Metro's higher discount frequency for bacon and bagels, along with its relatively low average prices, makes it a standout option for budget-conscious families.

When considering the combined effect of discounts, vendor pricing strategies, and convenience, Metro remains the optimal choice for families aiming to purchase all breakfast items—bacon, brown eggs, and bagels—at a single store while benefiting from competitive pricing and frequent deals. However, families with specific preferences for brown eggs or bagels may still find TandT or Walmart as appealing alternatives for targeted savings.

5.2 Limitations and Weaknesses

A significant portion of the old_price column was missing or left blank due to the absence of sales. These missing values were replaced with current_price, under the assumption that items without an old_price were sold at their regular price. While this approach ensures completeness of the dataset, it may oversimplify real-world pricing dynamics, as the absence of a sale does not necessarily mean that old_price and current_price are identical.

The regression models focused on a select few predictors—old_price, vendor, discounted, and month. While these variables capture key aspects of pricing behavior, other potentially relevant factors, such as product brand, regional differences, or marketing promotions, were not included. Their exclusion may limit the model's ability to fully explain price variability.

The dataset represents a snapshot of grocery prices in a specific geographic and temporal context, spanning February 2024 to November 2024. Pricing dynamics in other regions, time periods, or under different economic conditions (e.g., inflation, supply chain disruptions) may differ significantly, limiting the applicability of these findings to broader contexts.

5.3 Future Steps

Future research should consider incorporating additional predictors, such as brand or geographic location, to enhance the models' accuracy. However, analyzing all brands for bacon may over complicate the modeling process. Instead, future studies could focus on specific brands that are particularly popular among consumers. Data on preferred brands could be gathered through surveys, which would provide valuable insights into consumer preferences.

This approach is similar to the methodology applied to the bagel data in this study. The selection of bagels for analysis was informed by examining product names and identifying trends in popularity, such as the prevalence of “everything bagel” products across vendors.

In addition, future research could broaden the scope of analysis by including more granular data, such as daily price changes and more specific time frames. This would enable a deeper understanding of temporal pricing trends and discount patterns, capturing dynamic fluctuations that occur over shorter periods. Such advancements would provide a more nuanced perspective on grocery pricing behavior and offer greater value for both researchers and consumers.

Table 1: Selected feature of the Grocery dataset

Data_feature	Data_definition
nowtime	Timestamp indicating when the data was gathered.
vendor	One of the 7 grocery vendors (Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart Canada, Save-On-Foods).
product_id	An internal ID for a product - it represents 1 unique product at 1 vendor. It will allow you to track prices on the same product over time.
product_name	Name of the Product.
brand	Name of the Product Brand.
units	Units (grams, kg, number of items in package).
current_price	Price at time of extract.
old_price	An old struck-out price.
other	Other details that appeared on the listing. (Out of Stock, SALE, Best Sellet, Low on Stock, $5.00 MIN 2).

Appendix

A Data

A.1 Data Feature

Table 1 shows the variable features of the Canadian Grocery Price Data

B Model details

B.1 Model Summaries

Table 2: Summary of the Bacon Regression Model

Term	Estimate	Std. Error	t Value	p Value
(Intercept)	1.079	1.022	1.056	0.291
old_price	0.890	0.002	437.490	< 0.001
vendorMetro	-0.420	0.024	-17.311	< 0.001
vendorNoFrills	-0.121	0.027	-4.440	< 0.001
vendorSaveOnFoods	0.264	0.037	7.072	< 0.001
vendorTandT	-0.233	0.049	-4.768	< 0.001
vendorVoila	-0.074	0.023	-3.203	0.001
vendorWalmart	-0.135	0.027	-5.063	< 0.001
discounted	-2.367	0.019	-127.181	< 0.001
month06	0.029	1.022	0.029	0.977
month07	0.006	1.022	0.005	0.996
month08	-0.152	1.022	-0.149	0.882
month09	-0.058	1.022	-0.056	0.955
month10	-0.032	1.022	-0.032	0.975
month11	-0.044	1.022	-0.043	0.966

Table 3: Summary of the Brown Eggs Regression Model

Term	Estimate	Std. Error	t Value	p Value
(Intercept)	0.041	0.016	2.524	0.012
old_price	0.999	0.001	1686.093	< 0.001
vendorLoblaws	0.009	0.007	1.211	0.226
vendorMetro	0.032	0.007	4.341	< 0.001
vendorNoFrills	-0.014	0.007	-1.966	0.049
vendorSaveOnFoods	0.037	0.009	3.991	< 0.001
vendorTandT	-0.085	0.008	-10.899	< 0.001
vendorVoila	0.011	0.007	1.484	0.138
vendorWalmart	0.001	0.008	0.146	0.884
discounted	-1.056	0.007	-162.059	< 0.001
month03	-0.024	0.015	-1.629	0.103
month04	-0.001	0.015	-0.045	0.964
month05	-0.045	0.015	-3.073	0.002
month06	-0.037	0.015	-2.539	0.011
month07	-0.038	0.015	-2.581	0.01
month08	-0.027	0.015	-1.877	0.061
month09	-0.029	0.015	-1.969	0.049
month10	-0.048	0.015	-3.276	0.001
month11	-0.026	0.015	-1.794	0.073

Table 4: Summary of the Bagel Regression Model

Term	Estimate	Std. Error	t Value	p Value
(Intercept)	0.621	0.571	1.088	0.277
old_price	0.917	0.003	284.476	< 0.001
vendorMetro	-0.123	0.019	-6.303	< 0.001
vendorNoFrills	-0.098	0.021	-4.577	< 0.001
vendorSaveOnFoods	-0.036	0.035	-1.032	0.302
vendorTandT	-0.037	0.116	-0.317	0.751
vendorVoila	-0.108	0.021	-5.138	< 0.001
vendorWalmart	-0.146	0.022	-6.578	< 0.001
discounted	-1.331	0.021	-64.103	< 0.001
month05	0.331	0.601	0.551	0.582
month06	-0.116	0.570	-0.203	0.839
month07	-0.090	0.570	-0.158	0.874
month08	-0.115	0.570	-0.202	0.84
month09	-0.176	0.570	-0.309	0.758
month10	-0.057	0.570	-0.101	0.92
month11	-0.117	0.570	-0.205	0.838

B.2 Surveys, Sampling, and Observational Data

The research question focuses on identifying which vendor offers the lowest prices for breakfast items—bacon, brown eggs, and bagels—across eight Canadian grocery vendors. While the observational data collected through screen-scraping provides a foundation understanding of pricing trends, incorporating survey data could address some gaps and enhance the analysis. Surveys could capture consumer preferences, purchasing behaviors, and perceptions that are not directly observable in the scraped data, offering complementary insights into vendor choice and its relationship to pricing strategies.

B.2.1 Surveys as an Additional Data Source

If provided with a $100,000 budget to incorporate surveys as a complementary method alongside observational data collected through screen-scraping, the funds could be used to gain valuable insights into consumer preferences, behaviors, and perceptions that are not directly observable from grocery websites. Surveys could be designed to ask respondents to rank their preferred brands of bacon, bagels, and eggs, providing a clearer picture of demand-side dynamics. Furthermore, surveys could investigate the factors driving vendor choice, such as proximity, price, or product quality, offering a deeper understanding of pricing patterns observed in the dataset. This integration of consumer insights would enhance the analysis by bridging the gap between observed prices and the underlying consumer behaviors shaping them. Lastly, the survey can collect information on whether consumers prefer to purchase all breakfast items from a single vendor or shop at multiple stores to secure better deals. This data would help explore consumer behavior related to convenience versus cost savings, including their willingness to travel farther for lower prices. Additionally, this investigation would provide insights into the influence of vendor location on consumer choices, highlighting how proximity impacts shopping decisions alongside pricing strategies. The budget could also be utilized to purchase discount coupons from all eight vendors and offer them to respondents, allowing for an experimental approach to validate their stated preferences. By observing which vendor respondents are most likely to choose when presented with equivalent discounts, the study can confirm whether their answers about preferred vendors are consistent with actual behavior. This method would help mitigate potential biases introduced by self-reported preferences, ensuring that the survey findings are more reliable and reflective of true consumer decision-making.

B.2.2 Proposed Survey Methodology

Sample Questions:

1. *Which vendor do you typically shop at for breakfast items like brown eggs, bagels, and bacon.*
2. *Which brands do you prefer for these items?*
3. *What makes you want to shop at these vendors? Is it because of convenience?*
4. *Would you travel to a different store if it offered a 10% discount on breakfast items?*

Our target population are Canadian consumers who shop for groceries, ideally spanning diverse demographics (e.g.„ age, income levels, geographic locations). For sampling, we are going to use a stratified random sampling approach to ensure representation from various regions and consumer groups. A sample size of at least 1,000 respondents for the best representation.

We will employ a combination of online survey panels, social media outreach, and targeted telephone recruitment to ensure broad participation and accessibility, particularly for harder-to-reach demographics such as older adults and rural residents. These methods are tailored to include individuals who typically drive to grocery stores for their shopping. For example, the survey will explore whether consumers are willing to drive an additional 5 kilometers (e.g., from a store 20 km away to one 25 km away) in exchange for a 10% discount. This approach will help assess the trade-offs consumers make between convenience and cost savings, offering valuable insights into the role of location in grocery vendor choice.

B.2.3 Challenges and Considerations

Responses may still be subject to social desirability bias, where participants report what they believe to be the "correct" answer rather than their actual behavior. Additionally, since the dataset only includes eight vendors, there is the possibility that respondents may shop at other grocery stores not included in the dataset, such as Costco or smaller, independent retailers. This limitation highlights the potential for unrepresented vendor preferences and behaviors that could influence the overall findings. To address these gaps, the survey could include open-ended questions ask participants to name additional vendors they visit for breakfast items.

As a result, if respondents' preferred vendor is not among the eight vendors listed in the dataset, they will still be eligible to receive gift cards as prizes for completing the survey. This approach ensures that respondents are not incentivized to provide biased answers by selecting a vendor from the list solely to receive a discount coupon. By decoupling the reward from the vendor selection, this strategy minimizes response bias and encourages participants to answer questions honestly, even if their preferred vendor is not represented in the dataset. This not only enhances the reliability of the survey data but also acknowledges the diversity of grocery shopping preferences.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Hamner, Ben, and Michael Frasco. 2018. *Metrics: Evaluation Metrics for Machine Learning*. https://CRAN.R-project.org/package=Metrics.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Robinson, David, Alex Hayes, and Simon Couch. 2024. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.