

Canadian grocery analysis*

Jin Zhang Yitang Siyuan Lu

November 14, 2024

In this paper, we collect historical grocery data from top Canadian grocery sector, clean it, and analyze the pricing strategies. Many items have remained the same price or become cheaper, showing that the grocery vendors tend to use competitive pricing strategies. The research can help drive more competition and reduce collusion in the Canadian grocery sector.

Table of contents

| | | |
|----------|-----------------------|----------|
| 1 | Introduction | 2 |
| 2 | Data | 2 |
| 2.1 | Overview | 2 |
| 2.2 | Measurement | 3 |
| 2.3 | Graphs | 4 |
| 3 | Results | 6 |
| 4 | Discussion | 7 |
| | References | 8 |

*Code and data are available at: <https://github.com/YiTang2/Canadian-grocery-price.git>

1 Introduction

The Canadian grocery sector is a dynamic marketplace characterized by a variety of pricing strategies that reflect the intense competition and efforts to avoid collusion among vendors. This paper aims to illuminate these pricing strategies by analyzing historical data from top Canadian grocery chains, focusing on how prices have changed over time and how different vendors position themselves in the market. Through this research, we seek to contribute to the broader discourse on market dynamics in retail, providing insights that could help drive more competition and reduce collusion across the sector.

This paper explores the relationship between current and old prices across various vendors, revealing that many products have either maintained their price point or become cheaper. Such findings suggest that Canadian grocery vendors frequently employ competitive pricing tactics, possibly as a response to market demand dynamics or as a strategy to capture greater market share. These pricing behaviors are crucial for understanding the competitive landscape and for policymakers aiming to foster fair competition.

The significance of this study lies in its potential to impact both economic research and practical policy-making. By providing a detailed analysis of pricing strategies and their effectiveness, this study offers valuable insights that could help shape strategies to enhance competitive practices and reduce anti-competitive behavior in the grocery sector.

The remainder of this paper is structured as follows. Section 2 discusses the overview and variables of the data, including graphs and analysis. Section 3 explains the findings of the paper. Section 4 consists references of all data, software, and material. Finally, the Discussion section explores the implications of our findings, the study's limitations, and potential avenues for future research.

2 Data

2.1 Overview

We got the Canadian Grocery Price Data from Jacob Filipp(Filipp 2024). We used R programming language (R Core Team 2023) to analyze the data and plot the graphs for this paper. We also used SQL (9075:2016(E) 2016) to clean the raw data and generated the analysis dataset. The folder structure for this paper follows the startup folder created by (Wickham et al. 2019a). It is very helpful in keeping everything organized. I also referenced many of the examples and techniques provided by Telling Stories with Data (Alexander 2023), which helped me understand how to visualize the data and communicate the findings effectively. In addition, the `tidyverse` (Wickham et al. 2019b) packages is essential for simplifying data cleaning and analysis.`styler` (Müller and Walthert 2024) used in arranging code style. These tools help to organize and present data efficiently. I also plotted the graphs with `ggplot2`

(Wickham 2016) and read the csv file with `here` (Müller 2020). We also use the `comma` function from the `scales`(Wickham, Pedersen, and Seidel 2023) library to format the data, thereby avoiding scientific notation in our charts. We used the `knitr`(Xie 2014) to show the table.

All datas aim to foster competition and reduce market collusion in the Canadian grocery sector by assembling a database of historical grocery prices from prominent vendors. This dataset serves academic and legal purposes, enhancing transparency in grocery pricing strategies. The data encompasses prices from eight major grocery vendors in Canada, including Voila, TandT, Loblaws, and others. Below is the description for the variable we are taking to research:

- vendor: One of the 7 grocery vendors.
- current_price: Price at time of extract.
- old_price: An “old” struck-out price. This indicates that there was a Sale on. This is how you can differentiate a price-drop advertised as a Sale vs. a “quiet” decrease without a Sale.
- units: Units (grams, kg, number of items in package). May be blank for some vendors/products.

The dataset facilitates analysis through visualizations like price trends over time for products such as Becel Margarine and Sourdough Bread. Each visualization aids in comparing pricing strategies across different vendors and times. Exploratory data analysis provides insights into pricing patterns, with summary statistics highlighting average prices, standard deviations, and price ranges. Comparative analysis across vendors or periods can spotlight trends or anomalies in pricing behaviors.

2.2 Measurement

The measurement process involves converting observed grocery pricing phenomena into a structured dataset. This transformation starts with the collection of price data from various Canadian grocery vendors’ websites through screen-scraping techniques. Each price observation captures the ‘current price’ at the time of data collection and the ‘old price’, which indicates the previous price point before any discounts or changes. Prices are specifically for “in-store pickup” in Toronto, collected via screen-scraping of vendor websites. This method captures customer-facing prices but may miss details available in internal APIs. This approach allows us to trace price fluctuations over time and analyze market behaviors.

To ensure accuracy and relevance, the data is collected for items available for in-store pickup in a specific Toronto neighborhood, providing a consistent geographical context. The ‘units’ associated with each product, such as grams or kilograms, are noted to standardize price comparisons. This meticulous approach to data measurement facilitates a comprehensive analysis of pricing strategies across different vendors, revealing trends and potential collusive behaviors within the market.

2.3 Graphs

Some of our data is of current price (Figure 1), from JacobFilipp Filipp (2024).

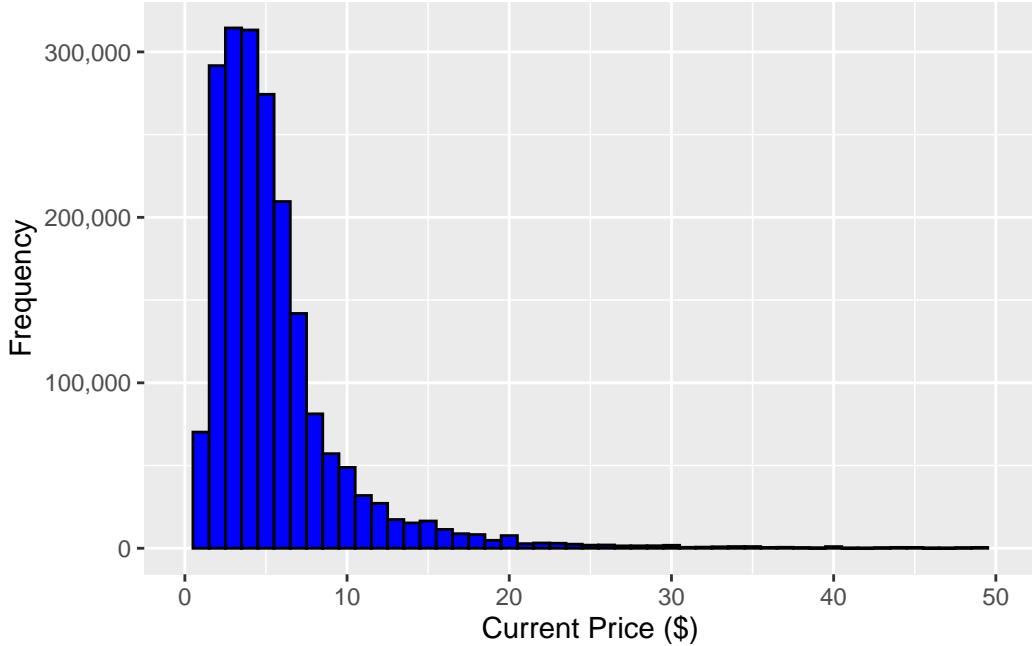


Figure 1: Distribution of Current Prices in Canadian Grocery Sector

Figure 1 shows the distribution of current prices in the Canadian grocery sector. The x-axis of this histogram represents the range of current prices in Canadian dollars, while the y-axis indicates the frequency of each price point. This visualization allows us to see that most grocery items are priced under \$10, showcasing a significant decline in frequency as prices increase. This distribution shows a concentration of more affordable items, with fewer high-priced items, indicating a price-sensitive market where cheaper products dominate the sales volume.

Some of our data is of price (Figure 1), from JacobFilipp Filipp (2024).

Figure 2 shows the logarithmic distribution of prices by different vendors in the Canadian grocery sector, using a box plot for each vendor. The x-axis categorizes the vendors, and the y-axis shows the logarithm of current prices, which shows the wide range of prices and highlighting the differences in price distributions across vendors. Each box plot provides the median, interquartile range (IQR), and potential outliers, revealing variability in pricing strategies among vendors. For instance, some vendors show a wider range of prices, indicating a diverse product mix, while others have more compact distributions, suggesting more consistent pricing across their products.

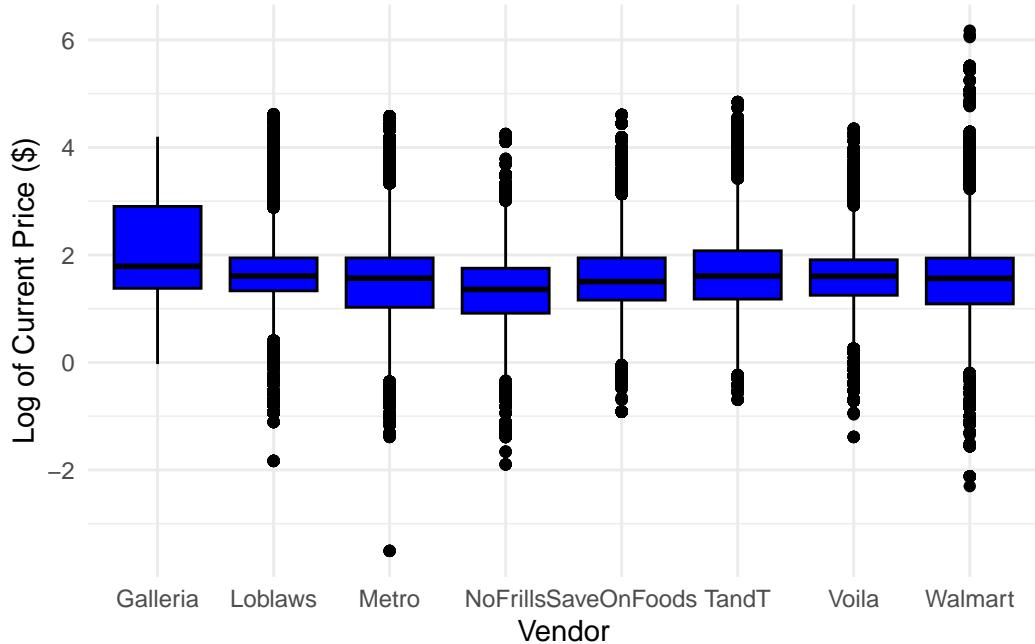


Figure 2: Distribution of Prices by Vendor in Canadian Grocery Sector

Some of our data is of comparison of price (Figure 3), from JacobFilipp Filipp (2024).

Figure 3 visualizes the relationship between old and current prices of grocery items in a scatter plot. The x-axis represents the old prices, and the y-axis the current prices, both in Canadian dollars. The diagonal red dashed line indicates where the old and current prices are equal. Points below this line represent price drops, while points above indicate price increases. This graph reveals a significant cluster of points along and below the line, suggesting that many items have remained the same price or become cheaper. This could indicate competitive pricing tactics or responses to market demand dynamics.



Figure 3: Comparison of Old and Current Prices in Canadian Grocery Sector

3 Results

| Mean | Median | SD | Min | Max |
|----------|--------|----------|-----|--------|
| 6.755484 | 4.79 | 9.957137 | 0 | 479.99 |

Summary Statistics for Current Prices in the Canadian Grocery Sector

(table-summary-statistics-current-price?) for current prices in the Canadian grocery sector reveal a diverse and competitive pricing landscape. The average price, represented by the mean, alongside the median, provides insight into typical consumer costs, suggesting a market geared towards affordability. The standard deviation highlights significant variability in prices, indicating a range of low to high-cost options available to consumers. The minimum and maximum values delineate the full spectrum of pricing, from the most economical to premium offerings, underscoring the market's breadth in catering to varied consumer needs and preferences. These statistics collectively demonstrate a dynamic market where vendors employ different pricing strategies to attract a broad consumer base.

| vendor | Mean | Median | SD | Min | Max |
|-------------|----------|----------|-----------|-----------|----------|
| Galleria | 2.238724 | 1.943049 | 0.8883970 | 0.6780335 | 4.214495 |
| Loblaws | 1.842668 | 1.790091 | 0.5191117 | 0.1484200 | 4.626442 |
| Metro | 1.820652 | 1.756132 | 0.7173577 | 0.0295588 | 4.589955 |
| NoFrills | 1.608753 | 1.587192 | 0.4952125 | 0.1397619 | 4.263102 |
| SaveOnFoods | 1.767844 | 1.704748 | 0.5158906 | 0.3364722 | 4.615022 |
| TandT | 1.928242 | 1.790091 | 0.7188706 | 0.4054651 | 4.851952 |
| Voila | 1.770481 | 1.786747 | 0.4636231 | 0.0000000 | 4.360675 |
| Walmart | 1.779771 | 1.752672 | 0.5741976 | 0.0953102 | 6.175846 |

Summary Statistics for vendor Prices in the Canadian Grocery Sector

The summary statistics for log-transformed current prices by vendor illustrate the varied pricing strategies across different grocery vendors in the Canadian market. The mean and median provide insights into the central tendency of prices on a logarithmic scale, indicating how each vendor positions itself price-wise within the market. The standard deviation reflects the diversity in each vendor's product pricing, revealing the range from basic to premium products they offer. The minimum and maximum values highlight the extremities of pricing for each vendor, capturing the cheapest and most expensive options available. These statistics underscore the competitive dynamics within the grocery sector, showing how vendors differentiate themselves through pricing to cater to different segments of the consumer market.

4 Discussion

This paper systematically analyzes the Canadian grocery market's pricing behaviors by comparing current and historical prices across various vendors. Through visual and statistical analysis, it identifies patterns and strategies in the pricing mechanisms employed by these retailers. The findings illuminate a highly competitive market environment where grocery vendors aggressively price their products to capture consumer interest. The prevalence of lower-priced goods underscores a strategic emphasis on affordability, potentially a response to consumer demand for value in everyday purchases. Differential pricing strategies among vendors reveal diverse approaches to market segmentation. Some vendors opt for a broad range of product prices, possibly catering to varied consumer segments, while others maintain tighter price clusters, indicating a focus on specific market niches. The primary limitation of this study arises from its data collection method—screen scraping—which might not capture all relevant pricing details or temporary promotional offers. This could lead to incomplete or skewed insights into the actual pricing dynamics. Future research could enhance the robustness of findings by integrating direct API access for real-time pricing data and expanding the geographical scope of data collection. Further investigation could also explore the impact of external economic factors on pricing strategies, providing a more comprehensive understanding of market forces and consumer behavior.

References

- 9075:2016(E), ISO/IEC. 2016. “Structured Query Language.” International Standard.
- Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r*. Chapman; Hall/CRC.
- Filipp, Jacob. 2024. “Hammer Dataset.” <https://jacobfilipp.com/hammer/>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Müller, Kirill, and Lorenz Walthert. 2024. *Styler: Non-Invasive Pretty Printing of r Code*. <https://CRAN.R-project.org/package=styler>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- 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. 2019a. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- , et al. 2019b. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Thomas Lin Pedersen, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.