Grocery Pricing Strategies: An Analysis of Vendor Positioning and Market Dynamics*

Examining Sales Frequency, Price Distribution, and Average Price Levels Across Major Grocery Chains

Andy Jiang Rui Hu Onon Burentuvshin

November 14, 2024

This paper examines grocery pricing strategies across major vendors to understand how they position themselves in a competitive market. By analyzing sales frequency, price distribution, and average price levels, we identify distinct approaches used by each vendor to attract customers. Our findings reveal that some vendors, like Metro, rely on frequent promotions to appeal to a broad customer base, while others, such as No Frills and Walmart, prioritize everyday low pricing to cater to budget-conscious shoppers. This analysis highlights the varied tactics grocery vendors use to target different consumer segments, providing insights into the competitive dynamics that shape grocery pricing and customer choices.

Table of contents

*Code and data are available at: [https://github.com/AndyYanxunJiang/grocery-price-analysis.git).

1 Introduction

In recent years, competitive pricing and promotional strategies among grocery vendors have created a rich landscape for data analysis. By examining grocery price data collected across multiple vendors, we can gain insights into various aspects of vendor pricing, such as frequency of sales, price distribution, and average price levels. This analysis focuses on several major grocery vendors, comparing their sales frequency, price ranges, and average prices to better understand their market positioning and strategies.

This study investigates the frequency of sales events, the distribution of product prices, and the average price by vendor, highlighting the distinct approaches taken by each vendor to attract and retain customers. Through visualizations such as bar plots and box plots, the analysis reveals trends in vendor pricing strategies, showcasing how some vendors prioritize frequent discounts, while others maintain a steady low-price approach. The findings suggest that vendors like Metro employ high-frequency sales with a broad price range, while No Frills and Walmart focus on consistently low prices. These patterns reflect the varied strategies used by grocery vendors to appeal to different consumer segments.

This paper is organized as follows: Section 2 outlines the dataset and its measurement of real-world pricing strategies. Section 3 presents key findings on sales frequency, price distribution, and average price, revealing distinct vendor pricing strategies. Section 4 addresses considerations such as correlation versus causation, missing data, and sources of bias, which are essential for interpreting the analysis accurately.

2 Data

We use the statistical programming language R (R Core Team 2023) and several libraries, including `tidyverse` (Wickham et al. 2019), `DBI` (R Special Interest Group on Databases (R-SIG-DB), Wickham, and Müller 2024), `RSQLite` (Müller et al. 2024), and `ggplot2` (Wickham 2016). The data was sourced from Project Hammer (Filipp 2024), a project database created by Jacob Filipp that compiles grocery price data from various vendors.

2.1 Measurement

The dataset used in this analysis captures real-world grocery pricing and promotional strategies from multiple vendors. Each entry, sourced from Project Hammer (Filipp 2024), represents a product listing, with variables such as `current_price`, `old_price`, `vendor`, and `nowtime`. These fields translate vendor pricing strategies into data points: `current_price` reflects the active selling price, while `old_price` denotes any discount or promotional pricing. The vendor field identifies the grocery chain, and `nowtime` timestamps each entry, allowing us to analyze trends over time. This structure enables an organized examination of pricing tactics across vendors by converting complex market behaviors into measurable data.

2.2 Overview

The key attributes of the data include:

Current Price: The price of each product at the time of data collection, which helps in understanding price distribution across vendors. Sales Frequency: The frequency of discounted prices, indicating how often each vendor holds sales or promotional events. Vendor: The specific grocery store or vendor from which the pricing data was collected, allowing for comparisons between different grocery chains. Using this dataset, we analyzed the pricing and promotional strategies of each vendor, focusing on three main dimensions: frequency of sales, price distribution, and average price by vendor.

Section 3 presents our findings on these aspects and highlights the distinct strategies employed by different vendors.

3 Results

Our results on the frequency of sales by vendor is illustrated in Figure 1, while Figure 2 shows the distribution of prices across vendors. Additionally, Figure 3 highlights the average price by vendor.

3.1 Frequency of Sales by Vendor

In Figure 1, Metro stands out with a noticeably higher frequency of sales compared to all other vendors. This suggests that Metro might be employing an aggressive pricing strategy, using frequent sales promotions to attract a large customer base. In contrast, Loblaws and Voila follow at a second tier, with sales frequencies approximately half that of Metro. This moderate level of sales activity indicates a competitive, but less aggressive, approach. No Frills, Save-On-Foods, and Walmart have similar and relatively low frequencies, at around half the

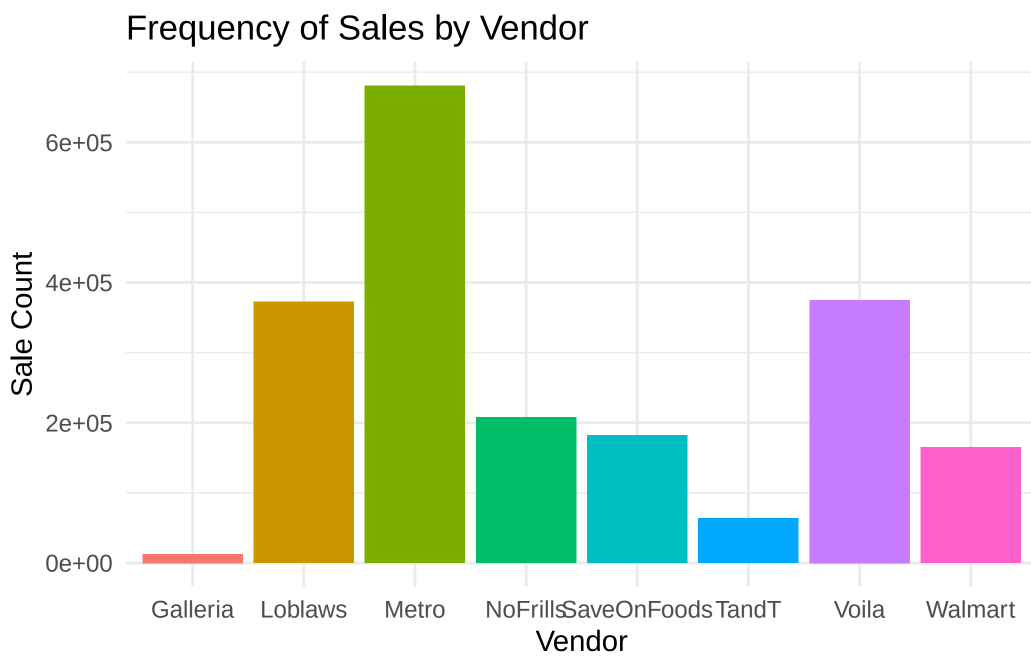

Figure 1: The Frequency of Sales by Vendor

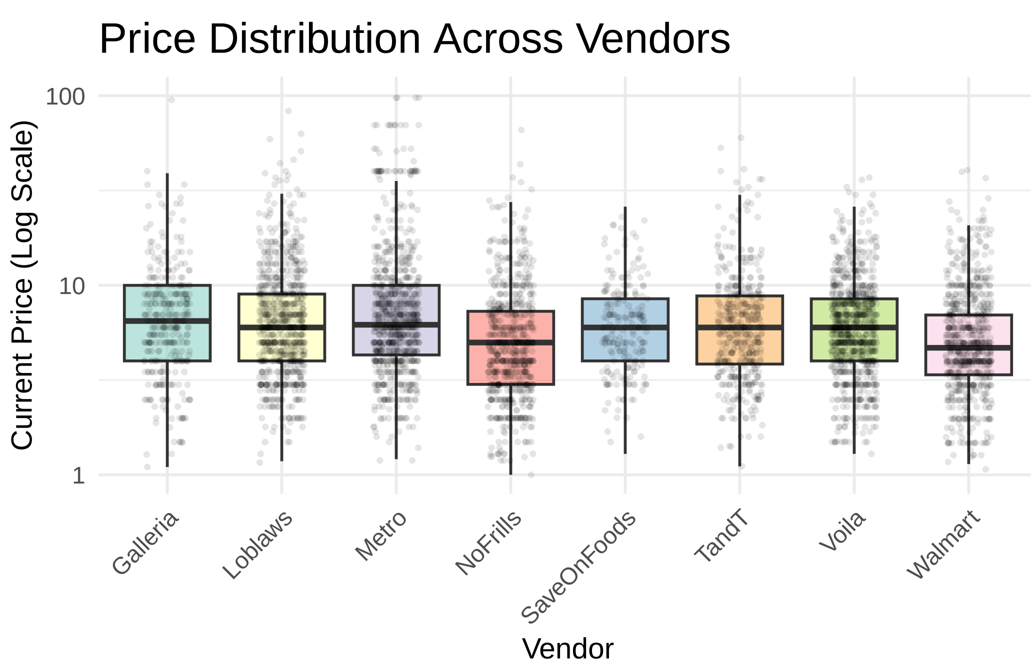

Figure 2: The Price Distribution Across Vendors

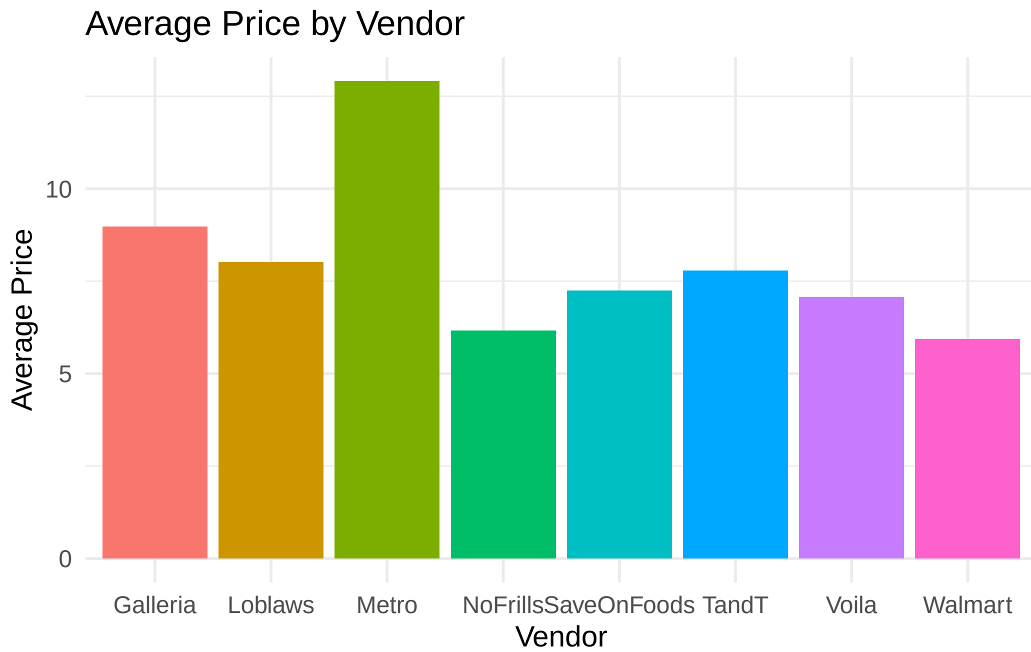

Figure 3: The Average Price by Vendor

level of Loblaws and Voila. This low frequency suggests that these vendors may focus less on frequent promotional events and more on maintaining stable pricing or lower base prices.

3.2 Price Distribution Across Vendors

Figure 2 reveals further distinctions in pricing strategy and positioning among the vendors. No Frills and Walmart show consistently lower price distributions, aligning with a budget-friendly image. This pattern suggests that these vendors aim to attract price-sensitive customers by offering lower-cost items, with minimal overlap into higher price ranges. Metro, on the other hand, displays a much wider price range, with a significant number of higher price points that are not seen in other vendors. This broad range suggests that Metro caters to a wide spectrum of customers, from budget-conscious shoppers to those willing to pay for premium products. Other vendors, such as Loblaws and Voila, show moderate price distributions, which are less spread out than Metro's but still cover a range that suggests a balanced approach rather than strictly budget or premium positioning.

3.3 Average Price by Vendor

Figure 3 aligns with the previous observations, highlighting Metro's position as the highest-priced vendor on average. This high average price, combined with Metro's wide distribution range, reinforces the notion that Metro offers a variety of products, including higher-priced or premium items. Galleria follows Metro with a relatively high average price, indicating a potential focus on specialty or premium goods. No Frills and Walmart emerge as the lowest-priced vendors on average, solidifying their positioning as budget-friendly options. The other vendors, such as Loblaws, Save-On-Foods, T&T, and Voila, occupy a middle ground, with average prices slightly above the lowest-priced vendors and below Galleria, suggesting a balanced pricing approach that caters to both cost-conscious and mid-range consumers.

4 Discussion

The results of this analysis provide valuable insights into the pricing and promotional strategies of various grocery vendors, revealing distinct approaches across vendors like Metro, No Frills, Walmart, and others. However, to interpret these findings accurately, it is essential to address certain methodological considerations. These include distinguishing between correlation and causation, handling missing data, and recognizing potential sources of bias. These factors are critical to refining the conclusions drawn from the study and are discussed in detail below.

4.1 Correlation vs. Causation

Our analysis revealed interesting correlations, such as Metro's high frequency of sales being associated with a broad price range, suggesting a dual approach that includes both budget and premium offerings. Additionally, No Frills and Walmart showed consistently low prices and lower sales frequency, which aligns with a strategy that focuses on everyday low pricing. However, while these correlations are clear, they do not imply causation. For instance, Metro's high sales frequency and diverse price range do not necessarily mean that frequent sales lead to a broader price distribution or that a wider range causes more frequent promotions. Both metrics could be influenced by other factors, such as Metro's target market, inventory diversity, or management's pricing strategy, which may independently affect both the frequency of sales and price diversity.

Similarly, vendors like No Frills and Walmart, with lower prices and fewer promotional events, might reflect a commitment to stable, low-cost offerings rather than suggesting that low prices cause a lower sales frequency. These observed patterns highlight potential correlations in vendor strategies but do not establish a causal link. Further research with controlled variables would be required to determine whether these strategies actively drive changes in other metrics or if they are simply co-occurring features of each vendor's broader approach.

4.2 Missing Data

The dataset contained missing values in key fields, such as `current_price` and `old_price`, which affected parts of the analysis, particularly in calculating accurate price averages and sales frequency. For instance, missing price data during high promotional periods for certain vendors could lead to an underestimation of their true sales frequency. This is particularly relevant for vendors like Metro, where sales frequency was a significant observation. If promotional pricing data was incomplete, Metro's high frequency of sales might appear inflated or underestimated, depending on when the data was missing. Similarly, missing data in the `current_price` field for premium items could cause an underrepresentation of higher prices in the dataset, potentially leading to a lower calculated average price than is accurate. This would especially impact vendors with premium offerings, like Galleria, where higher average prices were noted.

To mitigate the effects of missing data, we filtered out incomplete rows in critical analyses, ensuring that results were based only on fully recorded entries. However, this approach cannot account for the representativeness of missing data over time and across vendors, and therefore, the conclusions drawn may be limited by these gaps. Addressing missing data remains essential for improving the accuracy of this type of analysis in future research.

4.3 Sources of Bias

The dataset is subject to several potential biases that may affect the analysis and interpretation of grocery pricing data. First, the data includes only a select number of vendors, possibly excluding smaller, independent stores that may have different pricing and promotional strategies. This limitation could skew our understanding of average pricing, as larger vendors with more data points (e.g., Metro, Walmart) might dominate the analysis, potentially overshadowing strategies that smaller vendors might use. Additionally, the data is region-specific and does not account for regional variations in grocery pricing, meaning that the findings may not fully represent pricing trends across different areas outside of our target population of vendors within Toronto.

Another important source of bias stems from the method of data collection. The data was gathered from online listings, which may differ from in-store pricing and promotions. For instance, in-store promotions or discounts exclusive to loyalty program members would not be captured, potentially underestimating the frequency and diversity of promotions offered by each vendor. This could particularly affect vendors like No Frills and Walmart, which tend to have stable, low prices and may offer in-store promotions not reflected online. Furthermore, this method of data collection raises the problem of measurement bias, pertaining to in-store pick-up prices. These pick-up prices may differ from those that are available at in-store location or online delivery. The non-random exclusion of such prices bias our estimates for average prices and trends thereof. For instance, consumers choosing in-store pickup often differ in preferences, price sensitivity, and shopping habits from those opting for delivery. Thus, this method may only partially capture vendor strategies, especially regarding promotions and discounts, leading to an incomplete view of pricing dynamics.

The dataset has evolved to capture a broader spectrum of consumer habits by progressively including a wider variety of goods in the basket (Filipp 2024). Initially focused on a smaller set of products, the basket later expanded, aiming to better represent the diversity of grocery consumption. However, this inconsistency in product range over time introduces a potential source of bias, particularly in the analysis of price distribution and trends. Including a larger variety of goods can shift the average prices and potentially alter the price distribution across vendors. For instance, if the initial basket emphasized staple goods with relatively stable prices, while later data incorporated specialty items with higher price variability, observed price trends might reflect changes in product range rather than actual market shifts. This inconsistency potentially introduces a bias that renders any perceived trends could stem from the evolution of the dataset itself rather than from underlying price changes. Adjusting for this by creating comparable baskets over time or normalizing based on product categories could help mitigate this bias and provide a clearer understanding of true price patterns.

These biases should be carefully considered when interpreting the results, as they may influence conclusions about vendor positioning and market strategy. While this analysis highlights general trends across major vendors, understanding the full scope of grocery pricing dynamics would require addressing these biases in future research, potentially by expanding the dataset to include more vendors, regions, and in-store data.

References

Filipp, Jacob. 2024. "Project Hammer." *Jacobfilipp*. https://jacobfilipp.com/hammer/.

Müller, Kirill, Hadley Wickham, David A. James, and Seth Falcon. 2024. *RSQLite: SQLite Interface for r*. https://CRAN.R-project.org/package=RSQLite.

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

R Special Interest Group on Databases (R-SIG-DB), Hadley Wickham, and Kirill Müller. 2024. *DBI: R Database Interface*. https://CRAN.R-project.org/package=DBI.

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. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.