Mapping Grocery Pricing Dynamics: Vendor Strategies and Consumer Impact in Canada*

Price Variations: Walmart & NoFrills Lead in Affordability, Metro & Galleria Stay Premium

Jing Liang QISHENG YU Yuchen Yuan

November 21, 2024

This paper analyzes grocery pricing data from multiple Canadian vendors to explore differences in current and old prices. We find significant variation in pricing trends, with some vendors maintaining higher price levels while others exhibit more aggressive reductions over time. These findings highlight how vendor strategies impact consumer costs, offering insights into market dynamics and economic pressures. By uncovering these patterns, this study provides valuable information for consumers, policymakers, and regulators to promote affordability and transparency in grocery markets.

1 Introduction

Grocery pricing is central to discussions on inflation, economic inequality, and consumer behavior. Analyzing price variations across vendors, regions, and time offers key insights into market dynamics. This study examines Canadian grocery pricing data from Project Hammer (Filipp 2024), comparing old and current prices to identify vendor-specific trends. By addressing the lack of vendor-level price comparisons, this analysis informs both consumer choices and policy decisions. Using statistical and visualization techniques in R (R Core Team 2023), we uncover pricing patterns with implications for shoppers and regulators.

The study focuses on the average difference in pricing trends across vendors, measured as the ratio of old to current prices. This metric reflects vendors' pricing strategies and responses to economic pressures. We estimate average current and old prices for each vendor, accounting

---

for factors such as product category, location, and sampling frequency. This approach quantifies both absolute price levels and relative changes, offering a detailed view of grocery price trends.

Our findings reveal notable differences in vendor pricing. Metro and Galleria have the highest average prices, while Walmart and NoFrills rank the lowest. The old-to-current price ratios vary significantly, indicating diverse pricing strategies—some emphasizing affordability, others premium offerings. These trends show that consumers at higher-priced vendors may face significant cost differences for similar products, highlighting the need for transparent pricing data.

This analysis benefits consumers and policymakers alike. Consumers gain insights to make informed choices and save money amid inflation. Policymakers can use these findings to evaluate market competition, identify monopolistic tendencies, and advocate for fair pricing policies. As food security and economic inequality remain urgent issues, this study enhances transparency in grocery pricing, contributing to equitable access to essential goods.

The remainder of this paper is structured as follows. In Section 2, we describe the dataset used in this analysis and the measurement. In Section 3, we present the findings, focusing on inter-vendor comparisons and broader pricing trends. Section 4 contextualizes these results, addressing limitations and implications for stakeholders. By following this structure, we aim to provide a clear, systematic, and actionable analysis of grocery pricing in Canada.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) to perform all analyses and visualizations, leveraging its extensive libraries and tools for data manipulation, file path management, unit testing, and graphical representation, including `tidyverse`, `here`, and `testthat` (Wickham et al. 2019; Müller 2020; Wickham 2011). Our data (Filipp 2024) originates from a comprehensive dataset documenting grocery pricing trends across multiple vendors, providing a rich foundation for examining pricing dynamics and consumer behavior. Following the framework outlined by Alexander (2023), we consider both the narrative and analytical aspects of the data, ensuring that our interpretations are grounded in statistical rigor while remaining accessible to a broader audience. This approach allows us to highlight key insights and trends while addressing potential biases and limitations in the dataset.

2.2 Measurement

Measurement is the bridge that connects real-world phenomena to the structured entries in a dataset, encapsulating complex interactions in simplified, analyzable forms. In the context of

grocery pricing, the measurement process involves multiple steps, each of which influences the accuracy and reliability of the data. Prices, for example, are not static—they fluctuate based on promotions, regional differences, and supply chain dynamics. Capturing these variations in a single "average price" entry requires careful aggregation and standardization methods. Without a clear measurement protocol, such entries risk being incomplete or misleading representations of the actual phenomenon.

To collect pricing data, one might rely on automated web-scraping tools or manual entry systems. Each method introduces its own challenges. Automated systems are efficient but susceptible to errors from website formatting changes or missing data fields, while manual systems can introduce human errors or inconsistencies. Furthermore, the decision on what prices to include—such as whether to account for discounts, coupons, or bulk pricing—adds another layer of complexity. Each choice reflects an implicit judgment about what aspects of the pricing phenomenon are most relevant, potentially skewing the final dataset.

Beyond the mechanics of data collection, the temporal dimension of measurement is crucial. Prices captured at a specific point in time may not reflect longer-term trends, such as seasonal changes or economic shifts. For instance, measuring average prices during a promotional event might artificially lower the dataset's representation of a vendor's regular pricing. Incorporating mechanisms like time-weighted averages or periodic sampling ensures that the dataset remains representative of broader trends rather than momentary snapshots.

Finally, measurement depends on defining clear operational criteria for inclusion in the dataset. For example, what constitutes a vendor in this analysis? Should only national chains be included, or should smaller, independent stores also be represented? Similarly, decisions about which product categories to analyze—essentials like bread and milk versus luxury items— impact the generalizability of the dataset to different consumer demographics. These choices highlight the subjective nature of measurement and the importance of aligning it with the research question. By critically assessing how real-world phenomena are translated into dataset entries, we ensure that the data remains both valid and meaningful for analysis.

3 Results

Figure 1 visualizes the comparison between the average current and old prices for various grocery vendors in Canada, as analyzed using data inspired by Project Hammer. Vendors are ordered in descending order of their average current prices. The chart clearly demonstrates pricing trends, highlighting significant price differences among vendors, with "Metro" and "Galleria" showing the highest average current prices. Interestingly, Walmart and NoFrills, known for their low-cost branding, exhibit relatively lower prices compared to others. This visualization aligns with Project Hammer's goal of providing transparency in Canadian grocery pricing, capturing how price dynamics vary across vendors over time and offering valuable insights for consumers seeking cost-effective shopping options.
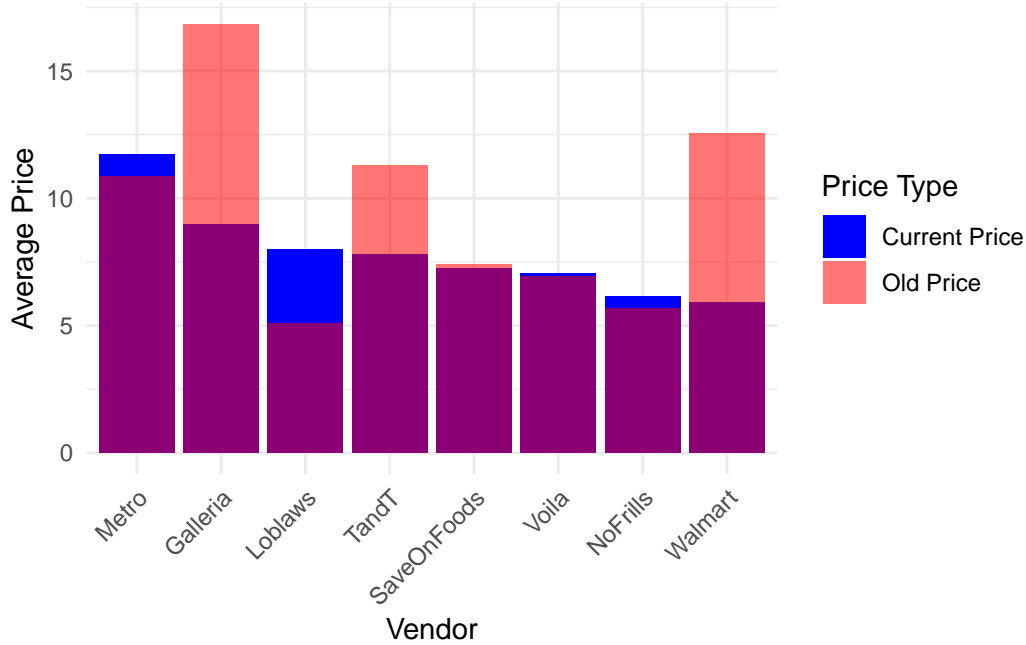
Figure 1: Comparison of Current and Old Prices by Vendor

4 Discussion

4.1 correlation vs. causation

One of the most critical aspects of interpreting relationships in data is distinguishing between correlation and causation. A correlation indicates that two variables move together in a certain pattern, but it does not inherently imply that one causes the other. This distinction is especially significant in observational studies, where underlying confounding factors often influence both variables. For example, in analyzing grocery prices across vendors, a correlation between a vendor's geographical location and average pricing might exist, but causation would require more robust evidence, such as controlled experiments or instrumental variables. Without careful consideration of confounders, one could erroneously conclude that location directly drives pricing, ignoring factors like vendor-specific strategies or local economic conditions.

From a broader perspective, understanding correlation versus causation involves grappling with concepts like temporal precedence and the directionality of relationships. For instance, the data may show that vendors with higher prices attract fewer customers, but the reverse could also be true: vendors with fewer customers might need to charge higher prices to cover operational costs. Establishing causation requires tools like mediation analyses, difference-in-differences, or propensity score matching, all of which account for potential confounders and ensure that observed relationships reflect true causative effects. This distinction becomes even

4

more nuanced when working with datasets where variables exhibit multicollinearity, which can obscure the independent effect of a single variable. Mastery over such methodologies allows researchers to move beyond superficial patterns and uncover the mechanisms driving observed outcomes, enhancing the interpretative rigor of their findings.

4.2 missing data

Missing data is an almost inevitable challenge in real-world datasets, yet its implications for analysis and conclusions can be profound. When exploring grocery price data, for instance, missing records for certain vendors or time periods could lead to biased averages, potentially distorting findings on pricing trends or inter-vendor comparisons. To mitigate this, it's essential to first identify the nature of the missingness: is the data missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR)? Each of these scenarios requires different approaches. For example, data that is MCAR might simply be removed without introducing bias, while MAR or MNAR often necessitates imputation techniques like multiple imputation or predictive modeling.

The choice of imputation method can also significantly influence the robustness of the results. Simple techniques like mean or median imputation may fail to capture the variability in the data, leading to underestimated standard errors and overly optimistic p-values. On the other hand, advanced methods like k-nearest neighbors or machine learning-based imputations preserve the underlying data structure but require more computational resources. Additionally, missing data can exacerbate biases if its presence correlates with other variables of interest. For instance, if prices from discount vendors are systematically missing due to limited reporting, the analysis might overestimate the overall average price, skewing consumer behavior insights. Effective handling of missing data is thus integral to ensuring that analyses remain unbiased, credible, and reflective of the true underlying patterns.

4.3 sources of bias

Bias is one of the most insidious threats to the validity of any analysis, often creeping in at various stages of data collection, processing, or interpretation. In the context of grocery price analysis, biases can arise from several sources, such as non-representative sampling, self-reporting errors, or systemic exclusions. For instance, if smaller or regional vendors are underrepresented in the dataset, the analysis might favor larger national chains, leading to conclusions that are not generalizable across all consumer demographics. This type of selection bias can significantly affect policy recommendations or consumer decisions derived from the study.

Another common source of bias is measurement error. If average prices are calculated from weekly or monthly snapshots rather than continuous data, they may not reflect real fluctuations, particularly for perishable goods or items with frequent promotions. Such inaccuracies

can lead to biased estimates of inter-vendor price differences. Furthermore, cognitive biases in interpretation, such as confirmation bias or anchoring, might lead analysts to overemphasize findings that align with preconceived notions about "premium" or "budget" vendors. Addressing these biases requires proactive strategies like ensuring diverse and representative sampling, cross-validating measurements, and rigorously testing assumptions through sensitivity analyses.

Lastly, publication bias poses a risk when disseminating findings. Studies highlighting stark price disparities or dramatic trends are more likely to gain attention than those reporting nuanced or incremental differences. This bias can mislead stakeholders by painting an incomplete or exaggerated picture of the market dynamics. By adopting transparent methodologies, preregistering analytical plans, and critically evaluating results, researchers can mitigate these biases, ensuring that their conclusions are both accurate and actionable. Bias, while inevitable to some extent, can be minimized through vigilance, methodological rigor, and a commitment to integrity in research practices.

References

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

Filipp, Jacob. 2024. "Project Hammer: Canadian Grocery Price Data." https://jacobfilipp.com/hammer/.

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/.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

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.