The Cost of Convenience: High Prices Across Grocery Chains*

Loblaws and Metro Lead as the Priciest Options in Grocery Shopping

Sophia Brothers Bruce Zhang Ruizi Liu Deyi Kong
Yuechen Zhang Yingke He

November 14, 2024

This paper explores price trends across different grocery vendors using observational data, focusing on identifying chains with the highest prices across various products. SQL was utilized for data preparation, and R was employed to create visualizations highlighting price distributions by vendor. The analysis reveals that Loblaws and Metro consistently emerge as the priciest options in grocery shopping. Discussions address data limitations, such as missing values and potential biases, offering insights into price variability within the grocery market.

Table of contents

*Code and data are available at: https://github.com/sophiabrothers1/groceryprices.

1 Introduction

The grocery market is marked by dynamic pricing across various vendors and product types, making it essential to understand price patterns for consumer insight and economic analysis. This paper investigates grocery prices across major retail chains, focusing on the most expensive vendors to highlight disparities in pricing strategies. Using SQL (PISO/IEC 2016) for data extraction and cleaning, combined with R (R Core Team 2023) for visualization, this analysis explores average price trends and identifies specific products and chains that command the highest prices.

By examining data from Canadian grocery sectors (Filipp 2024), including Loblaws and Metro, this study reveals significant insights into the grocery market's pricing structure. The results show that Loblaws and Metro consistently rank as the priciest options for a wide array of products, revealing potential factors behind these higher prices, including brand reputation and regional market influence. Through an analysis of price distributions across chains, this paper provides a valuable understanding of cost variability in the grocery industry.

The structure of this paper is organized as follows: Section 2 discusses data sources, SQL data preparation, and visualization methodologies. The main findings are presented in Section 3, where Loblaws and Metro are shown to lead as the most expensive chains. Section 4 addresses data limitations, such as missing values and potential observational biases.

2 Data

This data was obtained from an online source of grocery prices related data (Filipp 2024). The analysis made use of R Programming language (R Core Team 2023), including packages dplyr (Wickham, François, et al. 2023), tibble (Wickham, Vaughan, et al. 2023), here (RStudio 2024), RSQLite (Lang 2024), testthat (Wickham 2024), DBBI (Team 2024), and ggplot2 (Wickham et al. 2023). The dataset displays the price of different products for different vendors. The vendors include a series of grocery chains including Metro, Loblaws, Walmart, T&T, and a few more.

The dataset includes:

- **Vendor**: One of the seven major grocery vendors included in the analysis, each representing a significant presence in the grocery market.

- **Product Name**: The specific name or label of the grocery product, allowing for accurate identification and comparison across vendors.

- **Current Price**: The price of the product at the time the data was collected, reflecting real-time pricing and enabling analysis of cost variations across vendors.

The query cleaned the data by filtering for products available from at least three vendors to ensure relevance and by ranking product prices by vendor to identify the most expensive listings accurately. The query also joins the product table and raw table on the product_id/id columns.

The data visualized in Figure 1 highlights the overall frequency of vendors appearing in the dataset, with Metro and Loblaws showing a dominant presence. This suggests a trend where these two vendors not only stock a wide range of products but also maintain higher prices across multiple categories. Furthermore, Figure 2 illustrates the distribution of product prices by vendor, focusing on items priced under $250. The distribution charts reveal that Metro and Loblaws consistently have a higher density of products in the upper price ranges compared to competitors like Walmart, NoFrills, and T&T.

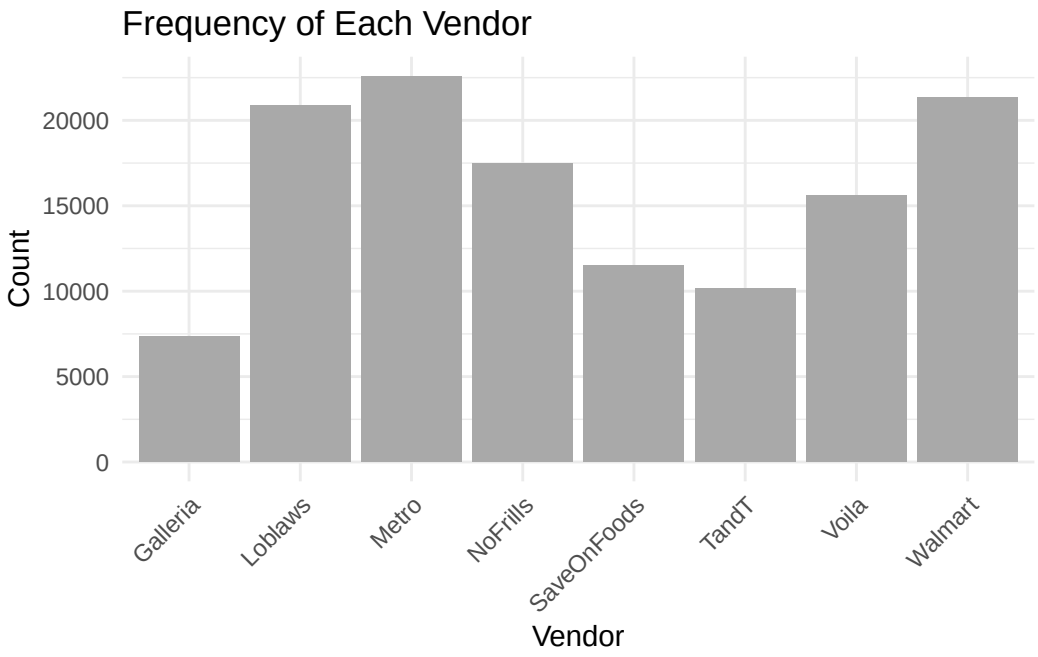

Figure 1: Shows the frequency in which different vendors appear in the dataset

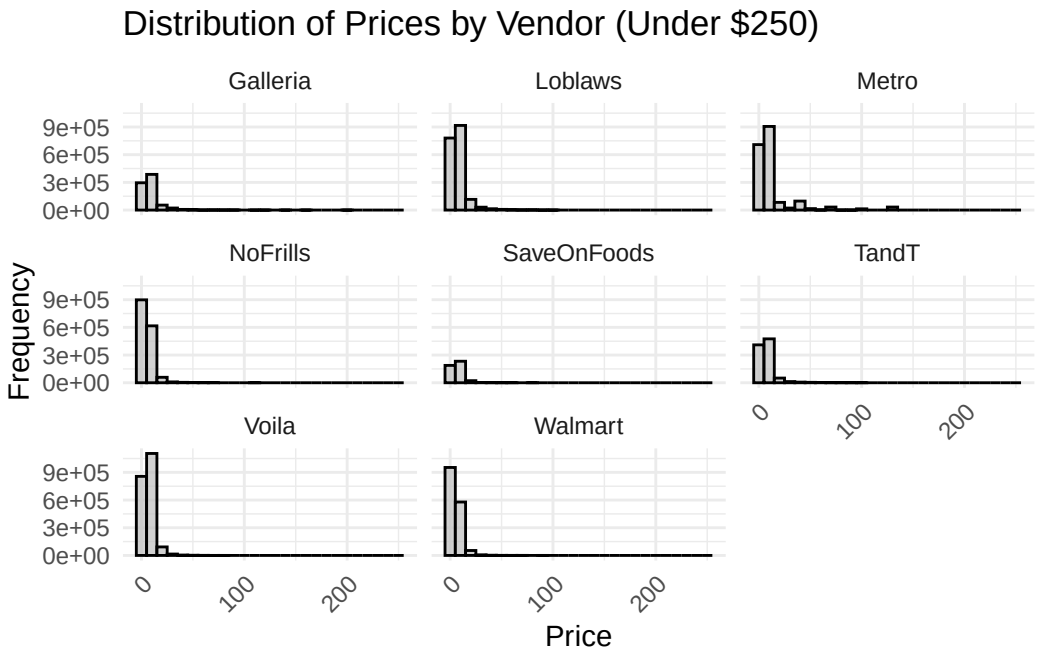

Figure 2: Visualizes the distribution of price by vendor

3 Results

Data analysis conducted using SQL and R revealed notable trends in grocery pricing across various vendors. As shown in Figure 3, Metro and Loblaws consistently emerged as the vendors with the highest counts of most expensive products. Specifically, Metro was identified as the most expensive vendor for products carried by three or more vendors a total of 28,609 times, while Loblaws followed closely behind with 27,561 occurrences. This pattern reinforces the observation that Metro and Loblaws implement premium pricing strategies, potentially targeting consumers who prioritize brand reputation and convenience over cost. These findings suggest that Metro and Loblaws may aim to position themselves as higher-quality or luxury grocery options, appealing to a more brand-loyal customer base. In contrast, other vendors with fewer high-priced counts may focus on affordability to attract cost-conscious shoppers. Overall, the visualizations highlight a distinct variation in pricing strategies among grocery vendors, with Metro and Loblaws positioned as the premium-priced options in the market.

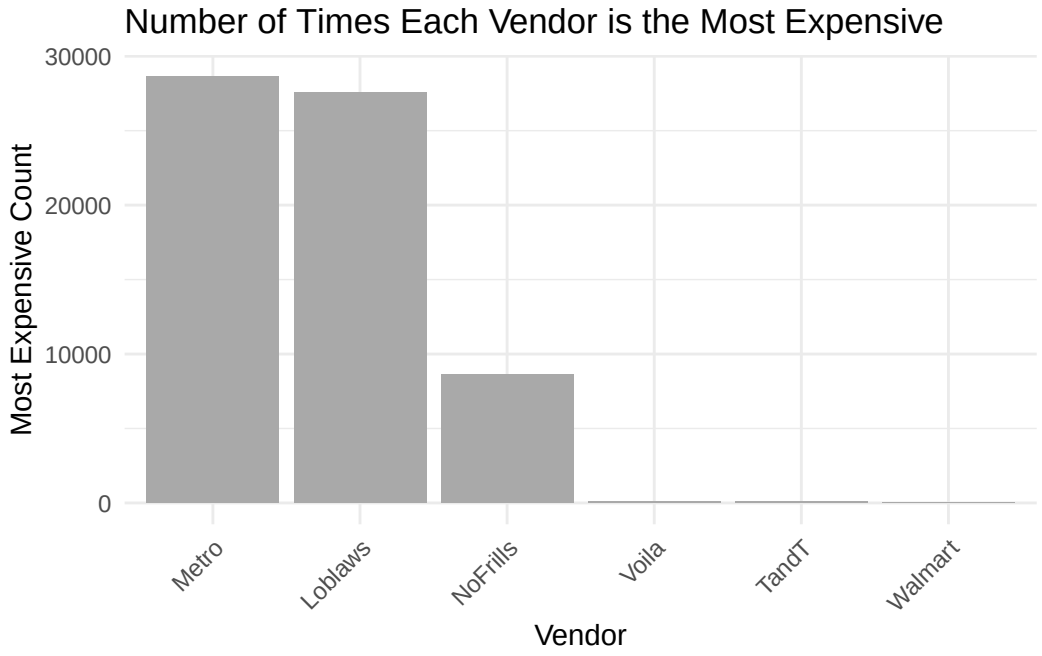

Figure 3: Illustrates the number of times a product was found to be the most expensive at a specific vendor. Products must be cross-carried by three vendors to count.

4 Discussion

The analysis of grocery prices across vendors reveals clear trends, with Loblaws and Metro consistently ranking as the most expensive chains for a variety of products. This finding suggests that these vendors may follow premium pricing strategies, potentially due to factors like brand reputation and market positioning. The price distributions show that, while there is variation among vendors, Loblaws and Metro maintain higher average prices across a broad range of items, setting them apart from competitors. These trends highlights grocery market pricing strategies, revealing how certain vendors target specific consumer demographics and shopping habits.

4.1 Loblaws and Metro are Pricier than other Grocery Chains

Loblaws and Metro are more expensive than other grocery chains based on the number of products that are most expensive at these chains. This can demonstrate real-world trends of pricing, as these chains are large and often located in urbanized regions. However, we cannot conclude a causation between location, size, and price. Many limitations of this paper highlighted below show that further research is needed to make meaningful, causal conclusions.

4.2 Correlation vs. Causation

While we observe differences in pricing across vendors, these differences do not imply causative factors without further experimental controls. In order to conclude causation, it is important to have an experimental design that involves the manipulation of variables. This dataset is obtained based on observed price trends and does not invovle variable manipulation. Similarly, vendors may set prices based on external market trends or supply chain factors not captured in this dataset. Further data collection and analysis including variable manipulation must be done in order to conclude causation.

4.3 Missing Data

Missing data is a common challenge in grocery datasets. It is unlikely that every aspect of each data point is collected perfectly, especially in a grocery store setting where the main purpose is to sell products instead of data collection. For example, some vendors may not report product prices consistently, leading to potential biases in the average prices. Handling missing data effectively, such as by imputation or exclusion, is essential for future analyses.

4.4 Sources of Bias

Several potential biases may influence these results. Vendor pricing strategies vary based on store policies, regional preferences, or consumer behavior. The pricing may be influenced by the abundance of the grocery chain in specific locations. For example, more urbanized regions may have more larger grocery chain stores and also have higher pricing compared to rural areas. Furthermore, online-only data may exclude regional in-store promotions and other timnely changes in pricing, limiting the generalizability of our findings.

5 Conclusion

The data in this analysis present a general idea of the pricing in grocery chains, summarizing the information on which chains have the most expensive pricing. The analysis breaks down the number of most expensive products by grocery chain, offering a broad overview of which chains have the highest occurence of being the most expensive. Future research could expand on this analysis by incorporating additional data on consumer preferences and regional factors.

References

Filipp, Jacob. 2024. "Hammer: A Toolkit for Quantitative Data Processing." https://jacobfilipp.com/hammer/.

Lang, Duncan Temple. 2024. "RSQLite: SQLite Interface for r." https://CRAN.R-project.org/package=RSQLite.

PISO/IEC. 2016. *Information Technology — Database Languages — SQL — Part 1: Framework (SQL/Framework)*. https://www.iso.org/standard/663555.html.

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

RStudio. 2024. "Here: A Simple Approach to File Paths." https://CRAN.R-project.org/package=here.

Team, R Core. 2024. "DBI: Database Interface." https://CRAN.R-project.org/package=DBI.

Wickham, Hadley et al. 2023. *ggplot2: Elegant Graphics for Data Analysis*. https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley. 2024. "Testthat: Testing r Code." https://CRAN.R-project.org/package=testthat.

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.

Wickham, Hadley, Davis Vaughan, et al. 2023. *tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.