Lowest Product Prices across Canadian Groceries*

Analysis the Price per Units of Selected Product across 8 Canadian Groceries

Yun Chu   Yiyue Deng   Tiarui Fu   Yanzun Jiang   Felix Li

Wen Han Zhao

November 14, 2024

This study compares the pricing of various grocers to determine which consistently offers the lowest prices across comparable product families. By analyzing data collected from market surveys over several months, we assess price differences in categories such as fresh produce, packaged goods, and household items. The results show that discount grocers generally offer the lowest prices, with regional stores occasionally providing competitive pricing in select categories. These findings provide valuable insights for consumers looking to maximize savings and for businesses aiming to refine pricing strategies. The results shows that Loblaws have the lowest price per unit and No frills have the second lowest price per unit for selected product.

Table of contents

---

*Code and data are available at: https://github.com/younazhao/Canadian-Grocery-Analysis.

1 Introduction

Canada's grocery sector, with its high market concentration and rising consumer prices, faces increasing scrutiny over limited competition and potential collusion. The recent Competition Bureau study highlights an urgent need for pricing transparency, yet consumers lack clear, data-driven insights into which grocers consistently offer the lowest prices across core product categories. To address this, Project Hammer has compiled an unprecedented, time-stamped database of historical prices from eight major Canadian grocers—Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods—with the aim of promoting fairer competition and enhancing consumer choice. While the dataset is extensive, a structured analysis is essential to identify which retailer consistently provides the best value.

This study fills that gap by rigorously analyzing Project Hammer's data from February to November 2024 to answer the question: Which grocer is generally the cheapest across all comparable product families? Using standardized methodologies to calculate and compare average prices across thousands of items, we identify the retailer with the lowest overall pricing and examine category-specific competitive patterns, providing a comprehensive view of grocery pricing dynamics.

2 Data

2.1 Overview

The data is from Filipp (2024) which contains Canadian Grocery Price Data from February 28, 2024 to today November 14, 2024. The 8 vendors in the dataset contains: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart and Save-On-Foods. The variables key to this paper are price per units, vendors, product_id, and product name. We used Wickham et al. (2023) and R Core Team (2023) to clean the data and perform analysis.

Price per Unit: This represents the price for each product, normalized by its unit of measurement (e.g., per kilogram, per liter, or per item). This variable is pivotal for our analysis of price comparison across vendors and is used to identify the lowest price for specific products across all vendors.

Vendors: The vendor variable identifies the specific retailer selling the product. By analyzing prices across different vendors, we can examine competitive pricing strategies and pricing trends for the same products offered by different retailers.

Product ID: The product_id is an internal identifier that uniquely represents a product at a specific vendor. This allows us to track individual products over time. However, since product_id is not consistent across vendors, we focus on the product_name to ensure accurate cross-vendor comparisons.

Product Name: The product_name variable represents the name of the product, often including details like brand and packaging size. Since product names may vary slightly across vendors (e.g., "Banana Organic" at one vendor and "Organic Banana" at another), we will standardize product names where necessary for comparison.

2.2 Measurement

The measurement of data in this analysis will focus on the price per unit variable, which represents the cost of a product per its unit of measurement (e.g., per kilogram, per liter, or per item). This metric allows for a standardized comparison of prices across different Canadian grocery vendors. By analyzing the price per unit, we can identify the lowest price for each product offered by various vendors, enabling us to assess price variations and competitive pricing strategies within the market.
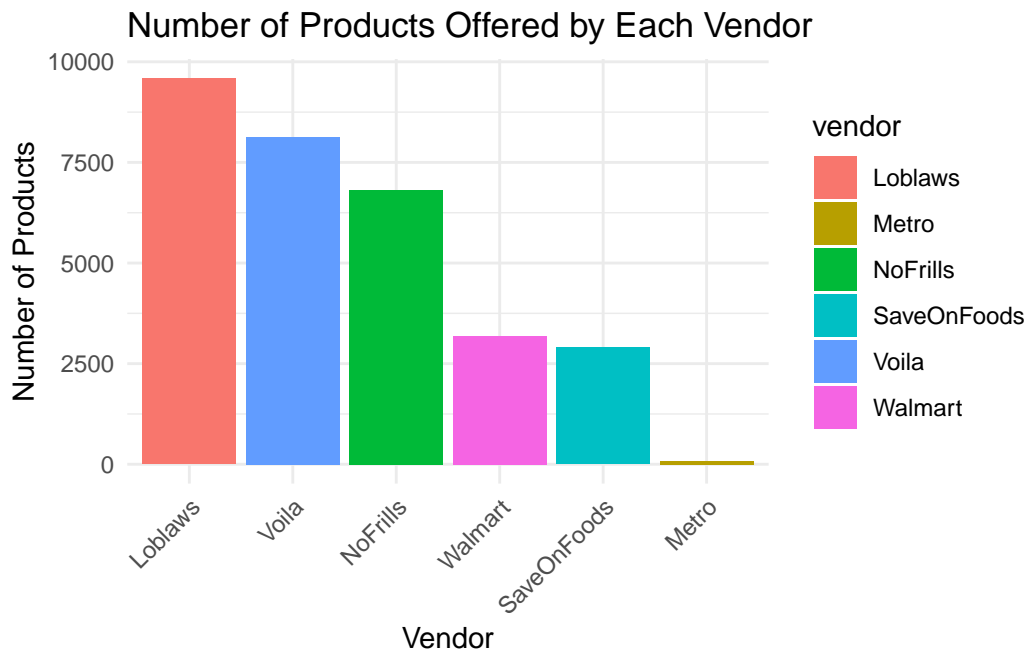
2.3 Overview of different vendors

Figure 1: Number of Products Offereed by Each Vendors

3 Results

The analysis of unit prices for milk and juice across five vendors — Loblaws, NoFrills, SaveOn-Foods, Voila, and Walmart, and eggplant across three of them — Loblaws, NoFrills, and Voila - reveals that Loblaws offers the lowest prices for milk and juice, which are both 10 cents per 100 ml. While eggplants using (per item) as the unit might not be comparable, NoFrills offers the lowest price, whichi is 1.29 dollars per lb (29 cents per 100 g). This indicates that Loblaws is the best choice for milk and juice, while NoFrills is the best choice for eggplants, offering the greatest savings on a wide range of groceries.
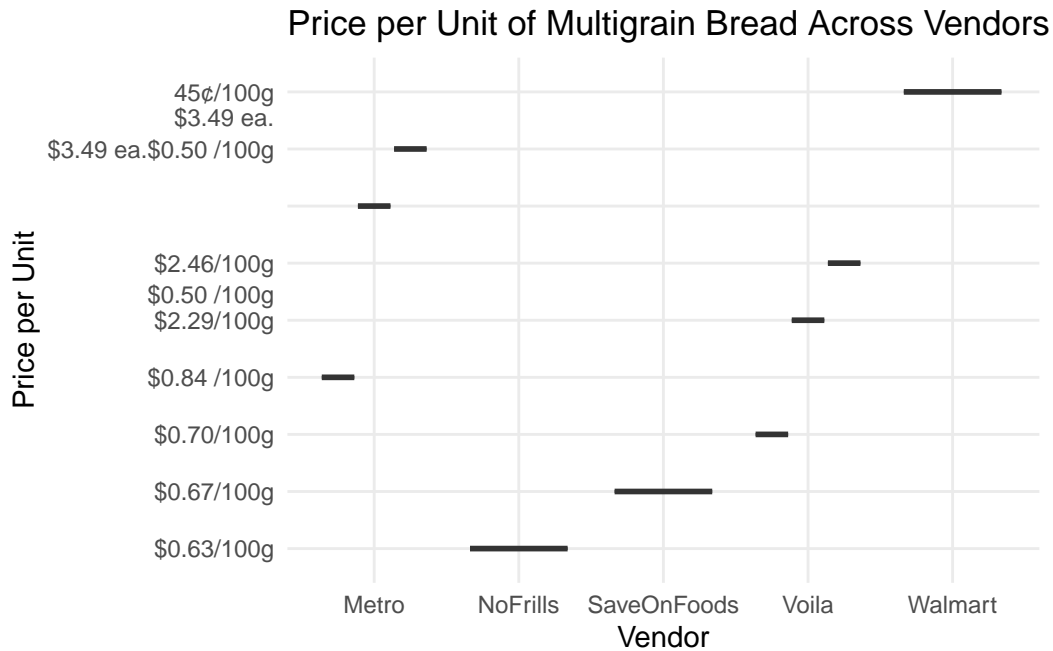
Price per Unit of Multigrain Bread Across Vendors

Figure 2: Price per Unit of Milk Across Vendors

4 Discussion

4.1 1. Correlation vs. Causation

The observed trend in prices over time may suggest a relationship between time and changes in the price of Nature Valley products. However, it is crucial to distinguish between **correlation** and **causation**. The price changes might coincide with certain periods (e.g., sales promotions or seasonal demand fluctuations), but this does not necessarily imply that time directly influences the price adjustments. Other underlying factors, such as marketing strategies, supply
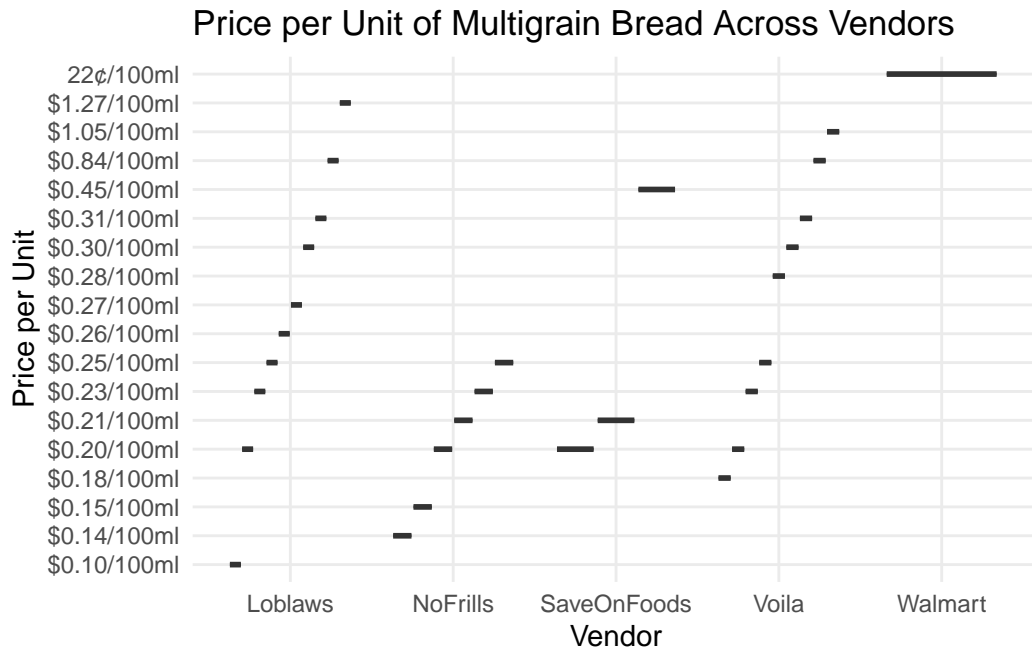
Price per Unit of Multigrain Bread Across Vendors

Figure 3: Price per Unit of 100% Juice Acrss Vendors

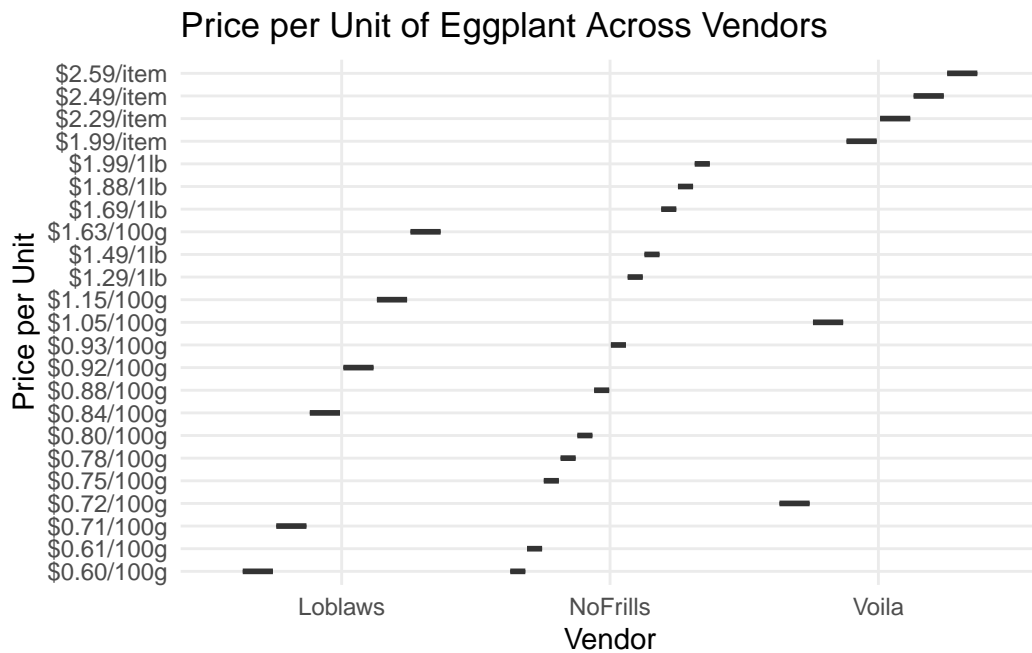## Price per Unit of Eggplant Across Vendors

Figure 4: Price per units of Eggplant across vendors

chain issues, or changes in consumer demand, could also be driving these variations. Therefore, while the graph shows a correlation between time and price, asserting a causal relationship would require deeper analysis, including controlling for other variables.

4.2 2. Missing Data

Missing data could affect the analysis and interpretation of the price trend. In this dataset, any gaps in the time series (e.g., missing records for certain dates) could lead to misleading conclusions about price stability or changes. For instance, a price drop or spike could be missed entirely if data is missing for a key period. Imputation techniques or sensitivity analyses may be necessary to address these gaps and provide a more robust understanding of the price trend.

4.3 3. Sources of Bias

There are several potential sources of bias in this analysis: - **Selection Bias**: The data only includes records from a specific vendor (e.g., Voila), which may not be representative of the overall market. Prices from other retailers or platforms could vary significantly. - **Temporal Bias**: The analysis spans a limited timeframe, which may not capture long-term trends or seasonal variations adequately. A longer dataset might reveal different patterns. - **Measurement Bias**: If the pricing data were collected manually or extracted from different sources, inconsistencies could introduce measurement errors. Variability in product packaging (e.g., different sizes or promotional bundles) might also distort the price comparisons.

These biases should be acknowledged when interpreting the results, and future studies could aim to include a broader range of vendors, a longer time series, and more rigorous data validation methods.

References

Filipp, Jacob. 2024. "Hammer | FAQ." https://jacobfilipp.com/hammer/#FAQ.

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