

Analysis of Grocery Price Dynamics*

A Study of Pricing Strategies and Vendor Comparisons

Chenika Bukes

November 21, 2024

This paper analyzes pricing data from various grocery vendors to uncover insights into pricing strategies, including average prices across vendors and changes in sale prices over time. We use SQL for data preparation and cleaning, with visualization and analysis conducted in R.

1 Introduction

This study investigates grocery prices from various vendors, focusing on the frequency of price changes, comparisons across brands, and temporal trends. We analyze price dynamics using time-series data to answer questions about vendor pricing strategies. The study uses R (R Core Team (2023)) for visualization with package with the aid of the following packages: tidyverse (Wickham et al. (2019)), here (Müller (2020)), lubridate (Grolemund and Wickham (2011)) and RSQLite (Müller et al. (2024)) for initial data preparation.

The main goals are to identify patterns in average prices, determine if vendors show coordinated price trends, and uncover how sale prices evolve. The remainder of this paper is structured as follows. Section 2 provides an overview of data sources and preparation; Section 2.2 outlines measurement processes; and Section 3 presents findings. The Discussion will cover insights, limitations, and next steps.

2 Data

2.1 Overview

The raw-data was downloaded from Project Hammer (Filipp (2024)) and downloaded as SQLite. The dataset includes time-series price data, cleaned through SQL to focus on non-null

*Code and data are available at: https://github.com/chenikabukes/project_hammer.

prices, vendors, and product identifiers.

2.2 Measurement

Price data were collected from different vendors through web scraping. SQL scripts ensured data consistency by standardizing price fields and transforming missing or erroneous entries.

2.3 Outcome variables

The primary outcome variable is the price of products over time, enabling comparisons across vendors.

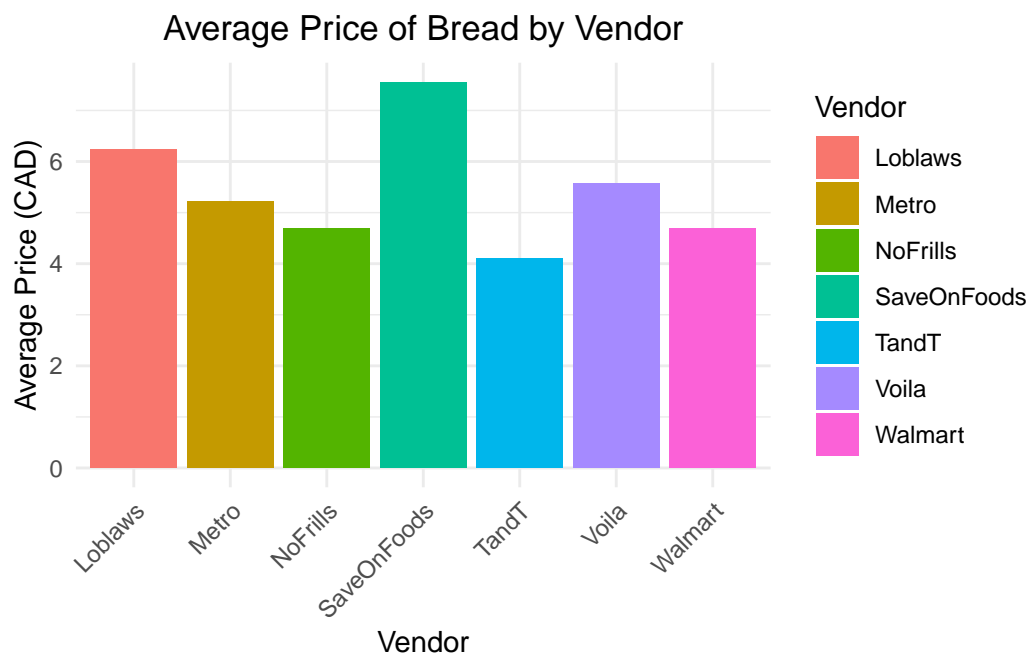


Figure 1: Average Price of Bread by Vendor

The bar chart in Figure 1 shows average bread prices per Vendor. This preliminary analysis indicates significant price variation, suggesting differences in vendor pricing strategies.

2.4 Temporal Analysis of Bread Prices

To gain insights into price trends over time, we examined the monthly average prices of bread products across various vendors. This analysis helps reveal any seasonal trends or price adjustments by specific vendors over the months.

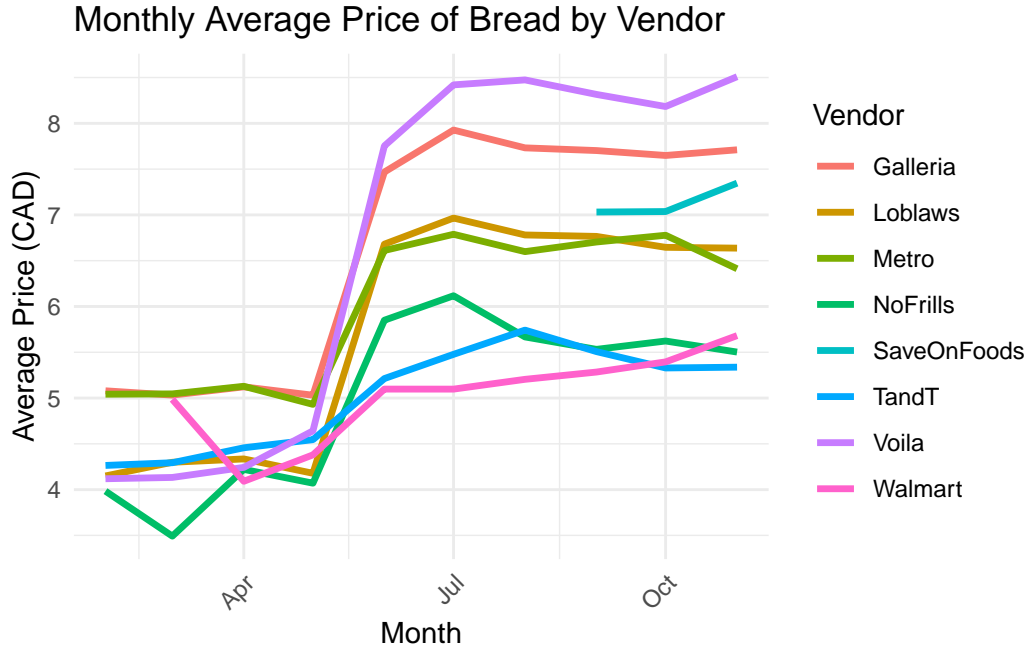


Figure 2: Monthly Average Price of Bread by Vendor

From February to November, we observe in Figure 2 that the average price of bread experiences a noticeable increase across most vendors, with a steep rise occurring between March and May. This rise may indicate a seasonal adjustment or market-driven price increase that affected all vendors simultaneously. Notably, Voila and Galleria exhibit the highest average prices throughout the timeframe, peaking at over 8 CAD per unit by November. This positions these vendors as the premium-priced outlets for bread, which may correlate with specific branding or target consumer demographics.

On the other hand, No Frills and T&T maintain relatively lower prices in comparison to other vendors, with No Frills being one of the most consistently affordable options. The trendline for No Frills remains more stable, indicating a strategy of price consistency, likely aimed at budget-conscious consumers. Metro and Walmart follow similar pricing trends, though their fluctuations are more moderate, with slight dips observed in late summer months, suggesting possible price adjustments during low-demand periods.

2.5 Predictor variables

The predictor variables include product brand, vendor, and price dynamics across time points. These help us identify trends in vendor-specific pricing and sale events.

3 Results

Our model reveals significant differences in pricing strategies across vendors. Results in Figure 1 suggest that vendor and brand effects are both statistically significant predictors of price. Additionally, analysis of monthly average bread prices, as shown in Figure 2, illustrates distinct pricing strategies across vendors, with notable variations in price stability, seasonal adjustments, and overall price levels.

4 Discussion

4.1 Correlation vs. Causation

While our analysis highlights correlations in price patterns across vendors, these do not necessarily imply causation. The observed price increases from March to May, for instance, are consistent across most vendors, suggesting potential market-driven factors such as increased production costs or seasonal demand. However, it remains unclear if these changes were directly influenced by shared supply chain costs, vendor-specific pricing strategies, or other external economic pressures. Furthermore, the higher price levels observed at vendors like Voila and Galleria might correlate with a premium pricing strategy, but this does not imply that their prices caused other vendors, such as No Frills and T&T, to adopt lower prices. These patterns may simply reflect each vendor's strategic positioning within the market.

4.2 Missing Data

The dataset contains some missing data due to web scraping limitations, particularly around sale prices and certain timestamps. This may result in incomplete price trends for specific vendors or products, potentially impacting the accuracy of our monthly average price calculations. For instance, if data were missing for sale prices during high-demand periods, the average prices could be artificially inflated, obscuring the true pricing strategy of a vendor. Ensuring more consistent data collection across all vendors and time periods would improve the reliability of our findings and reduce any skew caused by missing values.

4.3 Sources of Bias

Bias could arise from differences in vendor reporting or inconsistencies in data scraping. For example, vendors may display prices differently across regions, which could influence the scraped data if not properly accounted for. Additionally, vendors may implement promotional pricing at different times, which could lead to apparent discrepancies in the data that do not accurately reflect real differences in price trends. For example, Walmart and Save-On-Foods show price fluctuations likely due to promotional cycles, while No Frills maintains a more stable

price trend. These differences in pricing practices may introduce bias when comparing vendors. Furthermore, brand-specific promotions may not be fully captured, which could impact the average prices displayed in our results.

4.4 Weaknesses and next steps

Future research should expand to more regions to capture geographical price variation. Additionally, increasing the frequency of data collection would improve accuracy in sale duration analysis.

References

- Filipp, Jacob. 2024. *Project Hammer*. Toronto, Canada. <https://jacobfilipp.com/hammer/>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- 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/>.
- 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>.