Canadian Bread Price Variations Across Vendors*

An Observational Study on Pricing Trends and Data Limitations

Amy Jin   Diana Shen   Maggie Zhang   Claire Chang

November 14, 2024

This study analyzes the average price of bread across various vendors to understand pricing trends and highlight differences in consumer costs. By examining observational data, we found notable variations in bread prices between vendors, with some consistently offering higher or lower prices. These findings shed light on how vendor practices, product offerings, and possible external factors can impact bread affordability. Understanding these price variations is important for consumers seeking cost-effective options and for stakeholders aiming to address pricing disparities in everyday essentials.

1 Introduction

Bread is a staple food item across households, and understanding price variations among different vendors can provide insights into the accessibility and affordability of basic food products. This study aims to examine the average price of bread across various vendors in Canada to identify patterns and potential disparities in pricing. Using data on both overall average prices and per-unit prices (where available), we assess how bread pricing differs between vendors and highlight the factors that may contribute to these differences.

The primary estimand in this analysis is the average price of bread offered by each vendor. Additionally, for vendors that provide per-unit pricing, the analysis includes the average unit price, which helps standardize cost comparisons across varying product sizes and quantities. The focus on vendor-based pricing provides a granular view of how pricing might vary by store brand and supply chain factors.

Our results indicate a noticeable range in average bread prices across vendors. For instance, Walmart exhibits the highest average bread price among the surveyed vendors, while Voila

---

*Code and data are available at: https://github.com/aj3616/Canadian_Groceries_SQL/tree/main

offers the lowest average price. However, the availability of unit prices is limited, with only Walmart providing sufficient data to calculate an average per-unit price. This lack of unit price data highlights a gap in the dataset, suggesting inconsistencies in reporting or data availability across different vendors.

Understanding these variations in bread prices matters for consumers and policymakers alike. For consumers, awareness of price disparities between vendors can guide more cost-effective shopping decisions. For policymakers, identifying patterns in food pricing could support initiatives aimed at improving food affordability and accessibility. By shedding light on vendor-based pricing differences, this study contributes to broader discussions on food cost dynamics and potential areas for consumer advocacy.

The remainder of this paper is structured as follows. Section 2 discusses the data used for this analysis, including key variables and sources, with particular attention to the quality metrics that affect polling accuracy. Section 3 contains the graphs and tables of our findings. Finally, our Section 4 discusses the implications of our findings and suggests directions for future research by addressing three critical aspects: correlation vs. causation, missing data, and sources of bias.

2 Data

2.1 Overview

We use the programming language R (R Core Team 2023) to analyze and visualize our data, which allows us to make clear graphs and summaries. Our data comes from a recent study on grocery prices in Canada (Filipp 2024), providing detailed information about bread prices at different stores. Following the approach in Telling Stories with Data (Alexander 2023), we look at price patterns by store and consider what they mean for shoppers and decision-makers. We also use R packages to support our analysis and visualization tasks. The tidyverse package (Wickham et al. 2019) provided a suite of tools for data manipulation, and ggplot2 (Wickham 2016) enabled us to create visualizations. For generating tables, we used gt (Iannone et al. 2024). The readr package (Wickham, Hester, and Bryan 2024) facilitated efficient data import, while here (Müller 2020) helped manage file paths for a reproducible workflow.

2.2 Measurement

This study translates real-world bread prices from various Canadian vendors into a structured dataset for analysis. Each entry represents the average price of bread at a vendor, aggregated from observed prices to facilitate comparison across stores. Where available, unit prices (e.g., per kilogram or loaf) were also recorded to allow standardized comparisons despite differences in packaging sizes.

The data captures typical prices without including promotional discounts, providing a snapshot of regular bread costs across vendors. However, limitations arise due to missing unit prices from some vendors and the dataset's single-time snapshot, which may not account for seasonal fluctuations. Despite these limitations, this dataset offers valuable insights into bread pricing patterns across vendors.

3 Results

The chart Figure 1 demonstrates the average price for bread among different vendors This shows that Walmart is the vendor of the highest price of bread.
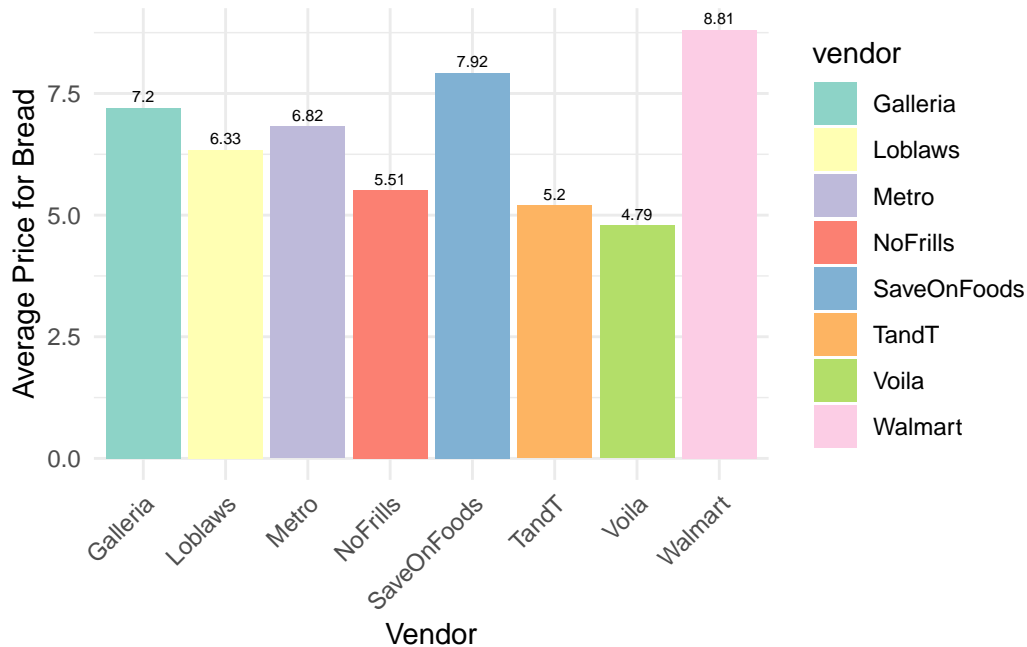

Figure 1: Average Price for Bread amoung different vender. This show that Walmart is the vendor of the highest price of bread.

4 Discussion

4.1 Correlation vs. Causation

In analyzing the pricing data of bread across different vendors, we are constantly differentiating correlation and causation to draw non-misleading conclusions. In this case, we are primarily

examining correlations between variables, but without enough evidence and experimental design, we are conscious not to overstate causal relationships. For example, our analysis shows that bread prices are higher at certain vendors like Walmart compared to Voila. However, we are aware that this is only a correlation; the price differences could arise from a variety of reasons, including business models, and promotional offerings, rather than a direct cause-and-effect relationship. If we were to argue that price increases in bread are caused by rising transportation costs from the vendor to the selling company, we would need to include more information in our data set and introduce transportation costs as control variables to observe changes in the dependent variable, which in this case is price.

4.2 Missing data

Missing data is an issue in our current data set and introduces challenges in conducting a comprehensive analysis. Specifically, the data set contains missing values in key variables, such as: Brand Information for Bread: Some records lack information on the bread brand, which limits the ability to analyze price differences between brands. Complete brand data could reveal patterns specific to certain brands and help us understand whether certain brands tend to be priced higher or lower than others. Without this information, the analysis is restricted to vendor-based comparisons, rather than offering insights into brand-specific pricing strategies. This is why we chose to rely on average pricing for our final conclusion. Old Price Data for Bread: Another key piece of missing data is the previous price for certain bread items. This prevents us from assessing how vendors adjust bread prices over time, which could be important in understanding pricing trends. The lack of this historical price data limits our ability to observe seasonal changes, inflation effects, or shifts in supply chain costs that might influence pricing behaviour. But despite these data being missing, the current analysis remains useful for identifying broad price trends across vendors. We assume that the missing data is missing at random, which means we assume every vendor has an almost equal chance of having missing data. Therefore, as a result, the analysis should still offer meaningful insights. However, it is important to acknowledge that the missingness has an impact.

4.3 Sources of bias

The data set spans from February 28, 2024, to the latest available load, which may not represent pricing trends over a long enough period to make definitive conclusions. For example, if a vendor undergoes a significant financial restructuring this year, this might influence how they adjust their pricing strategy. Therefore, the data set may only reflect prices at a specific point in time, limiting the ability to identify broader trends or seasonal price fluctuations. Vendor: The dataset consists of prices from only eight specific vendors: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods. In a competitive market, these vendors may have different pricing strategies, which could directly impact our findings. For example, vendors like Walmart may offer lower prices due to their larger customer base, while stores like

Voila may price their products higher due to quality perceptions and product selection. These inherent differences between vendors may introduce a bias, meaning that the observed price discrepancies could be driven more by each vendor's market position than by actual price differences for bread. Missing Data: As discussed in the Missing Data subsection, missing data for certain variables could also introduce bias. We are assuming that the missing data is randomly distributed, but if that is not the case, the missing data could skew the results. For example, if missing data disproportionately affects certain bread types at specific vendors (such as a store with a location where consumers tend to prefer one specific type of bread), this could lead to misleading conclusions. Missing data could also affect our ability to make comparisons between vendors, especially if missing values are not randomly distributed. Time Period: The dataset spans from February 28, 2024, to the latest available load, which may not represent pricing trends over a long enough period to make definitive conclusions. For example, if a vendor undergoes a significant financial restructuring thi

References

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

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

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy. 2024. *Gt: Easily Create Presentation-Ready Display Tables.* https://CRAN.R-project.org/package=gt.

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. 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. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data.* https://CRAN.R-project.org/package=readr.