

Comparative Analysis of Milk and Bread Prices Across Canadian Retailers*

Ziqi Zhu

Yuanchen Miao

November 14, 2024

We analyze grocery pricing data from an SQL database, focusing on the prices of bread and milk across several major Canadian grocery stores: Save-On-Foods, Loblaws, Voila, Metro, No Frills, and Walmart. By comparing the mean prices and generating density plots for each product, the study examines pricing trends across retailers. Findings indicate that Walmart, Metro, and No Frills consistently offer lower-priced bread products, while Save-On-Foods, Voila, and Loblaws have higher price distributions, likely due to the inclusion of premium items. For milk, Walmart and Voila maintain a low-price range, reflected in a high density of low-cost options and a lower average price overall. These findings are helping public to decide which grocery store they should go for groceries by comparing price and shows the price competition of grocery stores

1 Introduction

This study provides a comparative analysis of bread and milk prices across major Canadian retailers, including Save-On-Foods, Loblaws, Voila, Metro, No Frills, and Walmart. By examining the average prices for these essential products, we aim to identify price variations and trends that may indicate different pricing strategies employed by each vendor. Bread and milk were selected for this analysis due to their role as staple goods, and can reflect the general affordability of grocery items in the market. The findings highlight notable price disparities, with Loblaws tending to have higher prices and Walmart offering relatively lower prices across retailers, and Walmart have a relatively lower across retailers which could influence consumer preferences and purchasing decisions across these retailers.

*Code and data are available at: https://github.com/zzq20010617/canadian_grocery_price_analyse.

2 Data

2.1 Overview

The data is come from SQL database from Jacob Filipp (Filipp 2024) and cleaned by SQL query. We use the statistical programming programming **language R** (R Core Team 2023), and packages **dplyr** (Wickham et al. 2023), **knitr** (Xie 2014), **tidyverse** (Wickham et al. 2019) and **ggplot2** (Wickham 2016) in this paper,

2.2 Measurement

The SQL dataset we use captures several key attributes: the vendor name, product name, unit size, brand, current price, and price per unit (standardized, such as per 100g or per liter). By standardizing the units and focusing on consistent product types (milk and bread in this paper), we aimed to minimize variability and allow for fair comparisons across retailers. The data is thus a structured representation of observed price points, capturing real-world pricing dynamics as discrete, comparable entries that facilitate analysis of price trends and disparities among vendors.

3 Results

Table 1: Average Price of Bread by Vendor in Dollars per 100g

Vendor	Average Price
SaveOnFoods	1.856258
Loblaws	1.817191
Voila	1.750362
Metro	1.511408
NoFrills	1.392128
Walmart	1.208197

As table Table 1 shows, Save-On-Foods has the highest average price for bread at approximately \$1.86 per unit, followed closely by Loblaws at \$1.82. Walmart has the lowest average price at around \$1.21, suggesting a competitive pricing strategy for bread among major retailers.

Table 2: Average Price of Milk by Vendor in Dollars per 100ml

Vendor	Average Price
Loblaws	2.425530
Metro	2.139882
NoFrills	2.049000
SaveOnFoods	1.996047
Walmart	1.781600
Voila	1.729788

The table Table 2 shows the average milk prices, with Loblaws once again having the highest average price at \$2.43 per unit, while Voila offers the lowest average milk price at \$1.73. Metro and No Frills are also on the higher end for milk prices, with averages above \$2.00.

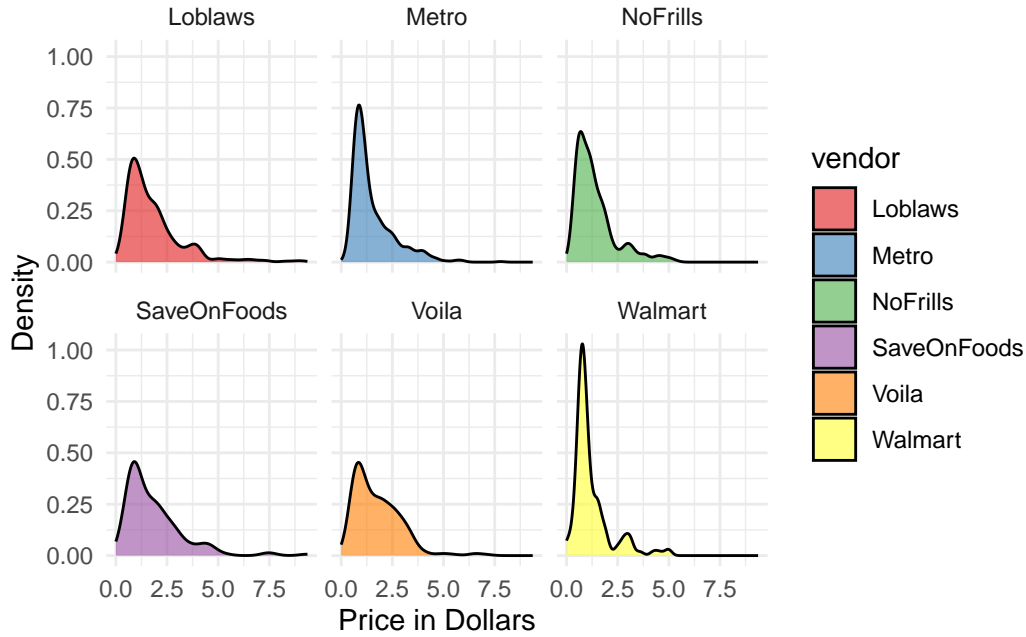


Figure 1: Density Plot of Bread Prices Across Vendors

The density plot of bread price Figure 1 shows Walmart, Metro, and No Frills appear to maintain the lowest price range for bread products, while Save On Foods, Voila, and Loblaws display slightly higher price distributions, likely due to a mix of standard and premium product offerings.

The density plot of milk prices Figure 2 indicates that Walmart and Voila appear to target lower-priced products, showing a higher density in the low-price range compared to other

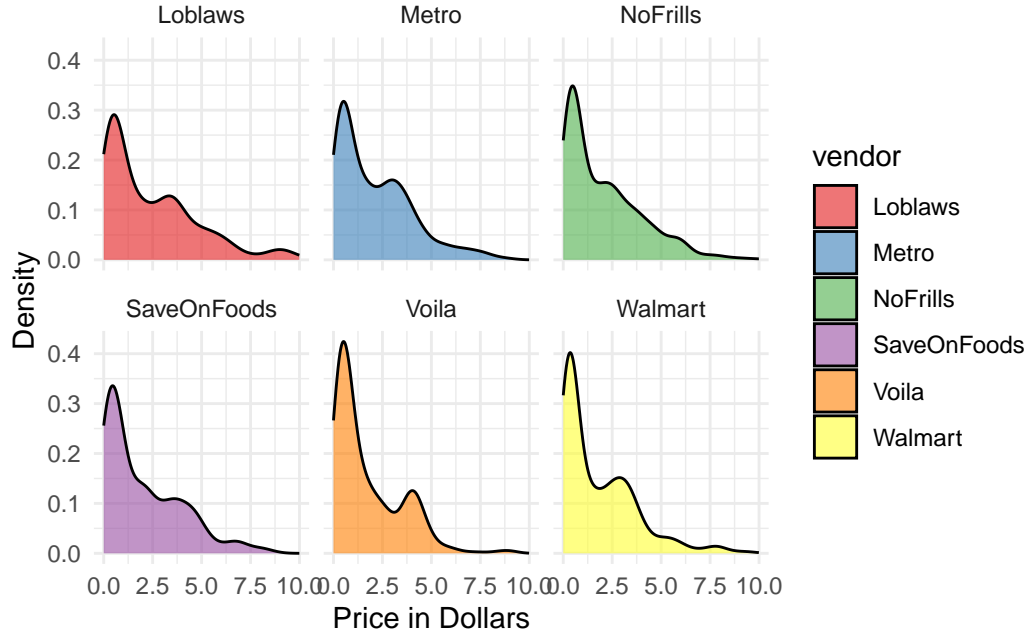


Figure 2: Density Plot of Milk Prices Across Vendors

retailers. This significant difference in density between low- and mid-priced products for these vendors contributes to a lower average price, as reflected in the summary table.

4 Discussion

4.1 Correlation and causation

In analyzing the pricing data of milk and bread across different grocery stores, our data show that Walmart and Voila generally offer lower prices for milk, which correlates with their positioning as discount retailers. The brand and quality of bread and milk that Walmart and Voila sells may be worse than the brand of milk and bread sells in Save-On-Foods, Loblaw's and Metro who show slightly higher price. Grocery retailers like Loblaw's, Save-On-Foods and Metro are including premium products. With different selling strategies of grocery stores they might have different cost of these products. We are not comparing the quality and brand of these retailers and this will be a factor that influences the price of milk and bread they sell.

4.2 Missing data

In our dataset, data on brands of milk and bread is missing, making it challenging to compare the quality of goods across brands. This lack of data introduces a limitation in our price analysis, as we cannot account for brand-related costs or directly assess price differences relative to product quality. Consequently, our price comparisons may be less accurate, potentially disadvantaging brands that offer higher-quality, premium goods, which are typically sold at higher prices. Without accounting for these brand variations in quality and associated costs, our analysis may inadvertently favor retailers with lower prices, even when those prices reflect differences in quality standards rather than simply cost efficiency. This limitation highlights the importance of complete data on brand quality when making fair comparisons across products and retailers.

4.3 Sources of bias

The data used in this analysis may contain several sources of bias. First, regional bias can affect the results, as retail prices vary by region due to transportation costs, local demand, and market competition. According to the data source, the prices are “online prices with a pickup location in North Toronto,” which may not reflect national pricing trends across all locations. Additionally, temporal bias is a concern, as prices fluctuate over time due to promotions, seasonal pricing, and supply chain factors. The unit price data we are using reflects the latest available prices in the dataset. If these prices were collected at a particular time or over a short period, such as during a promotional event, they may not represent typical prices across a broader period of time.

References

- Filipp, Jacob. 2024. “Hammer Database.” <https://jacobfilipp.com/hammer/>.
- 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 Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- 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>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.