Comparative Analysis of Sandwich Ingredient Prices Across Grocery Stores*

Ruiying Li Lorina Yang

November 14, 2024

This paper examines the cost variability of making a simple sandwich—composed of white bread and turkey ham across 3 grocery stores: Metro, Voila, Walmart, using data from the Hammer dataset. We find significant differences in ingredient prices, revealing which stores offer the most cost-effective options for budget-conscious shoppers. Our analysis highlights the practical impact of price disparities on consumer spending. Understanding these differences empowers consumers to make more informed decisions, ultimately promoting financial awareness and smarter shopping habits.

1 Introduction

Grocery prices have a direct impact on household budgets, making it crucial for consumers to know where they can find the best deals. This study examines the cost of making a standard sandwich—composed of white bread and turkey ham—across three major grocery stores: Metro, Voila, and Walmart. Despite the growing availability of price data, there remains a need for direct, transparent comparisons to inform consumer decisions and highlight pricing trends.

Using the Hammer dataset, we cleaned and standardized price information with SQL and analyzed it using R/Python(R Core Team 2023) with packages tidyverse(Wickham et al. 2016) and dplyr(Wickham 2023a) . Our findings reveal significant price variations between stores, offering practical insights for budget-conscious shoppers. By understanding these differences, consumers can make more informed choices, and our analysis underscores the importance of monitoring grocery price dynamics.

*Code and data are available at: https://github.com/Lorina-Y/Hammer.git

The paper is organized as follows: we describe the data and measurement, present our findings in result section, and discuss 1) correlation vs. causation; 2) missing data; 3) sources of bias in discussion section and conclude with the practical implications of our results.

2 Data

This study uses data from the Hammer dataset (Filipp Year), which provides detailed pricing information for grocery products across several major stores, including Metro, Voila, and Walmart. The product we analyzed are two main ingredients for making sandwich: Ham and White bread, and analyzing data through SQL and graphing data with R code(R Core Team 2023) and ggplot(Wickham 2023b) The two origin datasets has 15 variables in total, and this paper only interested in the following 4 variables: id, product name, price per unit and 3 vendors. Through this data, this study aim to analyzed the comparison of the price of two sandwich ingredients in specific 3 grocery stores to find the most cost-effective way of making a sandwich. # Measurement The Hammer dataset represents a snapshot of grocery prices at specific points in time. It translates real-world grocery pricing dynamics into structured data for analysis. The dataset use in this paper simplifies this complexity by providing records for product prices of white bread and turkey ham at Metro, Voila, and Walmart. To ensure comparability across Metro, Voila, and Walmart, the dataset we use contain standardized price per unit for white bread and turkey ham to dollars per 100g, and it is draw by record the price of sale in each grocery store.

3 Result

The bar chart(Figure 1) shows the price per unit of turkey ham across Metro, Voila, and Walmart, revealing significant differences: Metro and Voila have higher average prices at around $2.7 and $2.5 per 100g, respectively, while Walmart offers a much cheaper option at $1.83 per 100g. These results highlight substantial price variability among vendors, suggesting that Walmart may be the more cost-effective choice for consumers seeking to save money on turkey ham.

The bar chart(Figure 2) illustrates the price per unit of white bread at Metro, Voila, and Walmart. Metro and Voila have similar, relatively low unit prices, both just above $1 per unit, while Walmart's price is significantly higher, exceeding $3 per unit. This striking difference highlights Walmart's considerably higher pricing for white bread compared to the other two vendors, which could have important implications for consumer purchasing decisions.

By add the price for bread and turkey ham for each store, we found price of sandwich is around $3.65 in Metro, $3.52 in Voila and $5.33 in Walmart. So Voila offere cheapest sandwich.

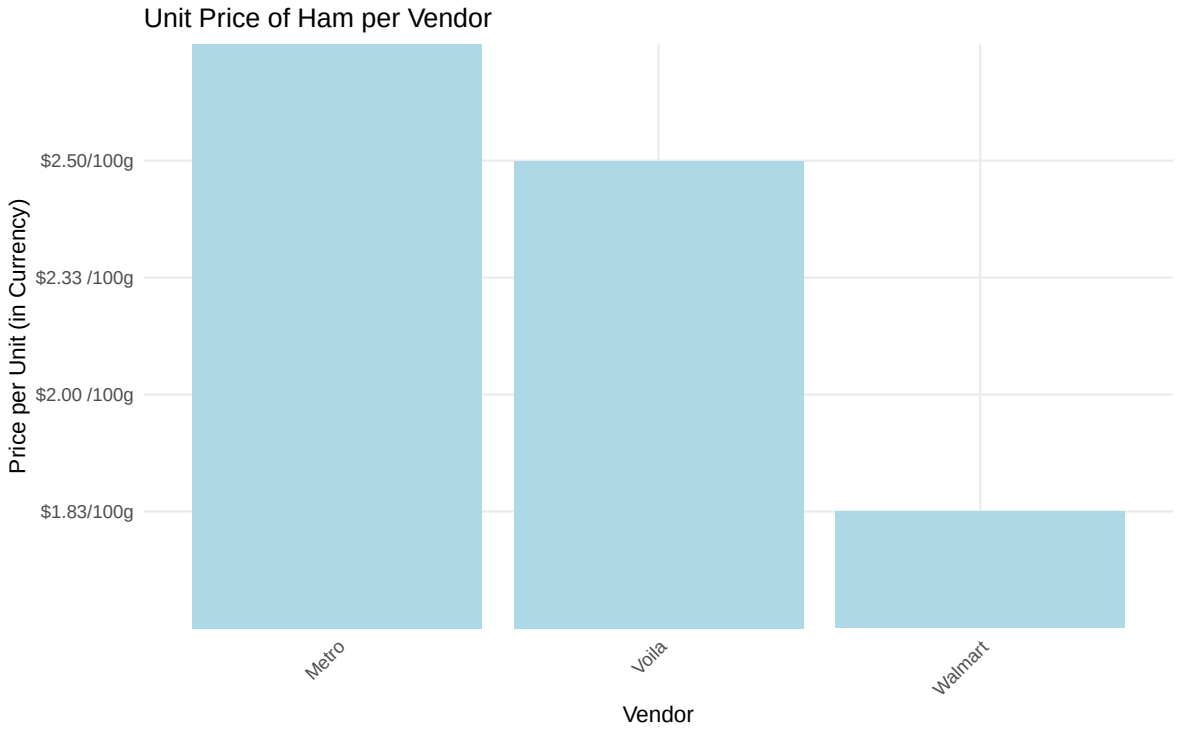

Figure 1: A Bar chart compares the price per unit of turkey ham across Metro, Voila, and Walmart

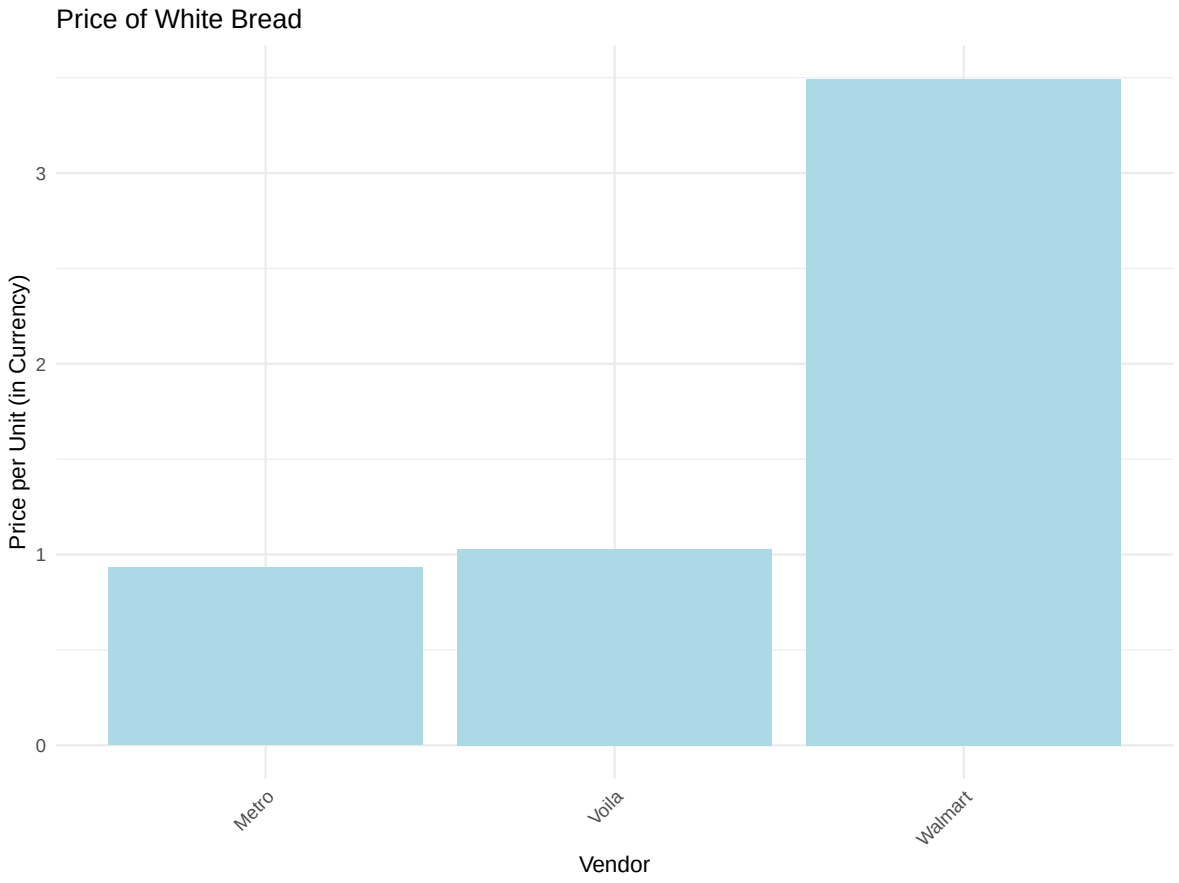

Figure 2: A Bar chart illustrating the price per unit of white bread across three grocery vendors: Metro, Voila, and Walmart

4 Discussion

4.1 correlation vs. causation

Our results show significant differences in the average prices of white bread and turkey ham across Metro, Voila, and Walmart (see Results section). However, it is crucial to emphasize that these findings only reflect correlations and not causal relationships. The observed price variations may be influenced by factors such as vendor-specific supply chain agreements, store locations, or target demographics, rather than deliberate pricing strategies. Additionally, other variables like seasonal promotions or regional economic conditions could play a role in these price disparities but were not included in our analysis. Therefore, while we can compare and describe price trends, we cannot draw causal inferences from these data.

4.2 missing data

As discussed in the Data and Measurement sections, the Hammer dataset provided comprehensive pricing information; however, potential gaps remain. Missing entries for certain product variations or incomplete price data could have affected the accuracy of our computed averages. Although our SQL-based cleaning process aimed to standardize and organize the dataset, some missing values might still introduce bias into our findings. Addressing these data gaps through imputation methods or additional data collection would strengthen the reliability of future analyses.

4.3 sources of bias

Several biases may have impacted our analysis. Selection bias is a concern, as the dataset may overrepresent certain product types or price ranges, especially if premium or discounted items were disproportionately included. This could affect the generalizability of our conclusions. Additionally, measurement bias might arise from differences in packaging sizes and unit formats, despite our efforts to standardize prices to dollars per 100g (see Measurement section). Finally, temporal bias should be considered, as our data represents a single point in time. Grocery prices fluctuate, and our analysis does not account for these temporal variations, limiting the scope of our findings. These biases highlight the importance of cautious interpretation and suggest directions for future research to improve the robustness of grocery price comparisons.

5 Conclusion

This study examined the cost variability of making a standard sandwich with white bread and turkey ham across Metro, Voila, and Walmart, using the Hammer dataset to compare prices after standardization. Our results show that Walmart offers lower prices for turkey ham but

significantly higher prices for white bread compared to Metro and Voila, and Voila offer the cheapest sandwich—composed of white bread and turkey ham across the 3 store.

Reference

Filipp, Jacob. Year. "Hammer Grocery Dataset." 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 et al. 2016. "The Tidyverse." https://www.tidyverse.org.

Wickham, Hadley. 2023a. "Dplyr: A Grammar of Data Manipulation." https://cran.r-project.org/web/packages/dplyr/index.html.

———. 2023b. "Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics." https://cran.r-project.org/web/packages/ggplot2/index.html.