Comparative Analysis of Discount Rates Across Grocery Vendors*

Ariel Xing Xinxiang Gao John Zhang Haoan Wang

November 21, 2024

This study examines discount rate patterns among major grocery vendors, analyzing how discount levels vary by vendor and over time. By comparing average discount rates and observing monthly trends, the research highlights notable differences in both the overall discount levels and monthly discounting patterns among vendors. Additionally, a deeper examination reveals evidence of Simpson's Paradox, where aggregated discount trends obscure unique, vendor-specific monthly patterns. These findings provide a clearer understanding of retailer discounting practices, illustrating how vendors balance pricing strategies to attract consumers. This research is important as it informs consumers and businesses on discount dynamics, supporting more informed decision-making in the retail market.

1 Introduction

In today's competitive retail market, discounting has become a prevalent strategy for attracting and retaining customers. Major grocery vendors utilize discounting not only to drive immediate sales but also to build customer loyalty over time. Understanding the patterns and trends in discount rates across various retailers can provide valuable information about broader pricing strategies and the economic factors influencing these strategies. However, while discount rates offer useful data for consumers, they also present complexities, especially when comparing multiple vendors with distinct business models and market approaches. This paper seeks to address these complexities by exploring discount trends among several major grocery vendors.

Prior research has examined the role of discounting in retail, with findings suggesting that discount rates impact consumer behavior, brand loyalty, and purchasing decisions. However,

*Code and data are available at: https://github.com/Jiaqi-Xing/Grocery-Prices-analysis few studies have systematically compared discounting practices across different grocery vendors over time. Additionally, aggregated discount data often obscures vendor-specific trends, leading to potentially misleading conclusions—a phenomenon known as Simpson's Paradox. This paradox occurs when trends observed in aggregated data contradict the patterns within individual groups, which is particularly relevant when analyzing discount rates across vendors with diverse pricing strategies. To address this gap, this paper investigates both aggregated and vendor-specific discount trends, revealing underlying patterns in grocery discounting.

This study utilizes a dataset containing monthly average discount rates from eight major grocery vendors. Through graphical analysis, we examine each vendor's average discount rate, track monthly changes, and identify instances of Simpson's Paradox within the data. Our key finding is that vendors exhibit distinct average discount levels and unique month-to-month patterns in discounting. This variability, obscured by aggregated trends, underscores the importance of detailed analyses when evaluating discount practices across retailers.

Understanding these trends is important for retailers, consumers, and policymakers as it sheds light on the dynamics of discount strategies. For consumers, knowing which vendors offer higher or more stable discounts can guide purchasing decisions. For retailers, recognizing competitive discounting patterns can inform pricing strategies to attract customers. For policymakers, identifying discount trends can reveal aspects of market competitiveness and economic trends within the retail sector. The remainder of this paper is structured as follows: in Section 2, we describe the dataset in detail, including the context, variables, and measurement approach used to capture real-world phenomena; in Section 3, we examine the results, presenting key findings through graphs, tables, and statistical analysis; in Section 4, we discuss the implications of our findings, addressing issues such as correlation versus causation, missing data, and potential sources of bias.

2 data

2.1 Overview

Our dataset contains detailed observations of product discount rates across eight grocery vendors: TandT, Walmart, SaveOnFoods, Metro, Galleria, Voila, Loblaws, NoFrills, over time. The primary goal of this dataset is to analyze vendor discount behavior and identify trends across various products and time periods. The data has been specifically organized to facilitate comparisons in average discount rates, with a focus on tracking fluctuations over daily and monthly timeframes for each vendor.

Our data source (Filipp 2024) was prepared using the SQLite programming language (SQLite Development Team 2023) to facilitate efficient handling and querying. Data simulation, testing, and visualization were performed using R (R Core Team 2023), with support from the packages readr (Wickham, Hester, and Bryan 2024), ggplot2 (Wickham 2016), tidyverse (Wickham et al. 2019), dplyr(Wickham et al. 2023), validate(van der Loo and de Jonge 2021). Inspired by the methods discussed in "Telling Stories with Data" (Alexander 2023), we applied Simpson's Paradox to explore complex trends and relationships in our analysis.

2.2 Measurement

The dataset comprises historical grocery prices compiled from the websites of top grocery retailers. Each price entry reflects the cost of individual products as displayed on each grocer's online platform, captured on specific dates to track price changes over time. Measurements include product attributes such as brand, category, size, and unit price, allowing for standardized comparisons across various retailers. The dataset spans multiple grocery categories, including fresh produce, dairy, meats, pantry staples, and household items, ensuring a broad representation of consumer goods. To measure seasonal or periodic price trends, data was recorded at regular intervals, such as weekly or monthly, across several years, providing insight into both short-term promotions and long-term pricing trends. Where applicable, discounts and promotions were documented alongside regular prices to capture fluctuations and offer a comprehensive view of each retailer's pricing strategy over time. This dataset serves as a valuable resource for analyzing price variability, inflation effects, and competitive pricing dynamics in the grocery market.

2.3 Variables

-Vendor: Identifies the grocery store or vendor offering the product. This variable allows for comparisons between different stores to determine which vendors generally offer higher or lower discounts.

-Product_ID: A unique identifier for each product. Although individual product names are not part of the analysis, the product ID ensures that each item is consistently tracked for discount changes over time.

-Current_Price: The present price of a product at a given point in time. This serves as a basis for calculating the discount rate.

-Old_Price: The original or previous price of the product, which is used to calculate the discount percentage.

-Nowatime: The specific date for each observation, enabling trend analysis over time.

Discount Rate: A constructed variable derived by calculating the percentage change between the old price and the current price. This variable provides insight into the magnitude of discounts and is essential for identifying which vendors offer the most significant price reductions.

Date: The specific date for each observation, enabling trend analysis over time.

2.4 Constructed Variables

- Average Discount Rate: This variable aggregates the average discount rate by vendor on a monthly, daily or total basis. It was constructed to help observe trends at a higher level and highlight any seasonal patterns.

3 Results

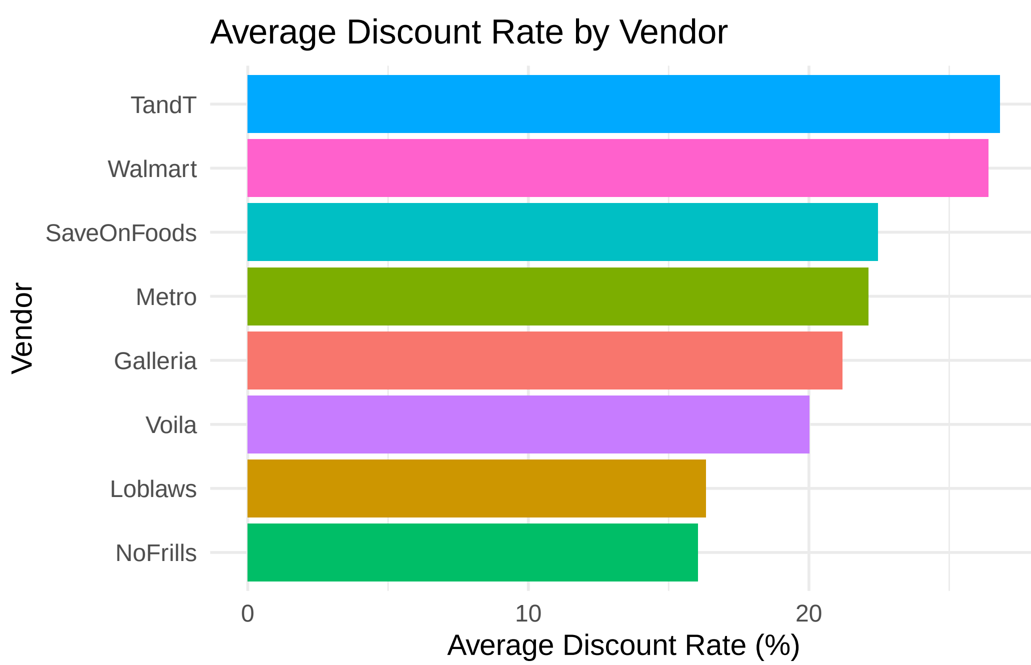

Figure 1: Average Discount Rate by Vendor: shows the average discount rate (%) in 2024 across various grocery vendors: TandT, Walmart, SaveOnFoods, Metro, Galleria, Voila, Loblaws and NoFrills

Figure 1 illustrates the average discount rate offered by each vendor. TandT and Walmart have the highest average discount rates, suggesting they tend to offer more substantial discounts than other vendors. On the other hand, NoFrills and Loblaws exhibit lower average discount rates, indicating a more conservative approach to discounting. The differences in average discount rates across vendors highlight varying discounting strategies that may be influenced by each vendor's pricing policies, target market, or competitive positioning.

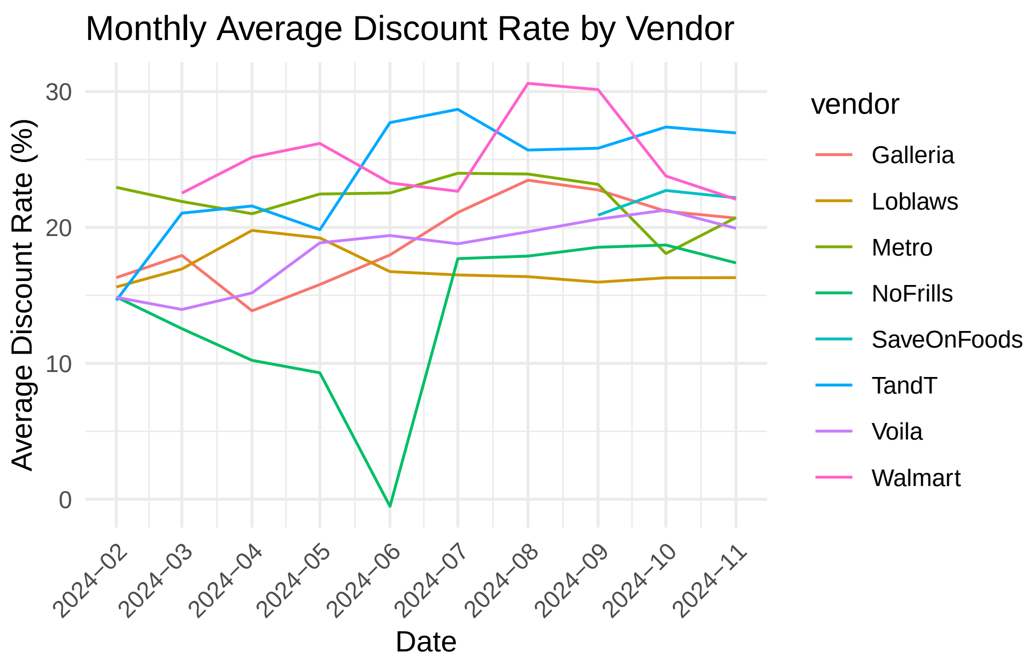

Figure 2: Monthly Average Discount Rate by Vendor: depicts the monthly average discount rate (%) for each vendor from February 2024 to November 2024.

Figure 2 tracks the monthly average discount rate for each vendor over the observed time period. We see considerable month-to-month variation in discount rates for certain vendors, particularly for SaveOnFoods, which shows a significant dip around July 2024, and Walmart, which reaches a peak around August 2024. The fluctuating patterns suggest that some vendors adjust their discount rates dynamically in response to factors such as seasonal demand, competitive pricing, or inventory levels. The chart provides a detailed view of how each vendor's discount strategy evolves over time, with some vendors maintaining more stable rates and others showing notable shifts across the months.

```
[1] "C"
```

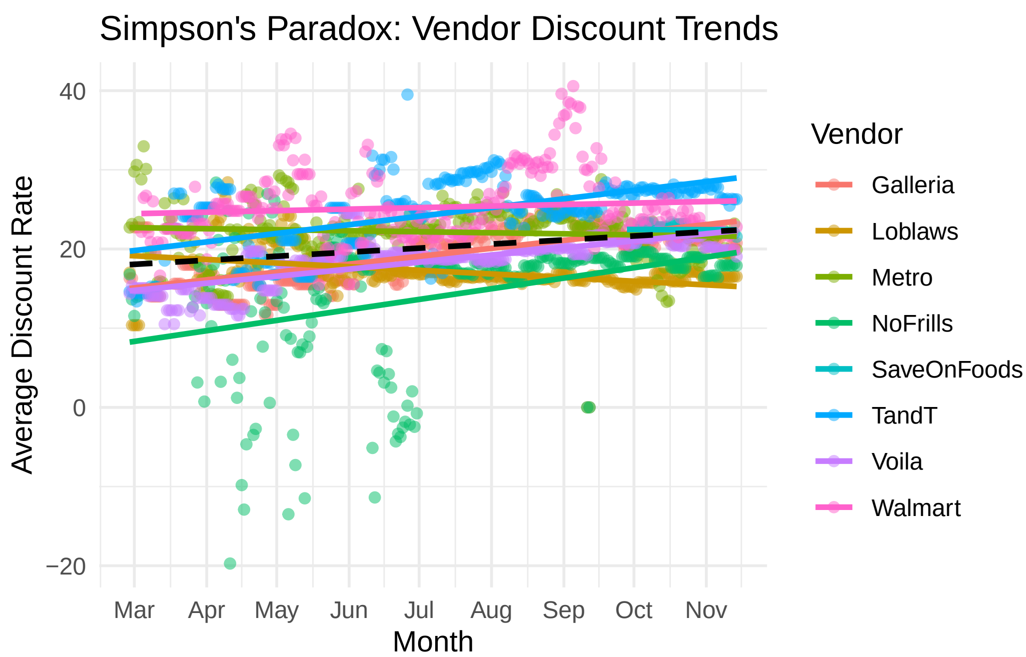

Figure 3: Trends in Vendor Discount Rates Over Time: Solid lines represent within-vendor trends, while the dashed black line indicates the overall trend across all vendors.

Figure 3 with trend lines illustrates the discount rates of various vendors over time, revealing an instance of Simpson's Paradox. While the overall trend may suggest a gradual increase or stable average discount rate, the individual vendor trends differ significantly. For example, some vendors exhibit a more consistent discount rate (such as Loblaws and Metro), while others, like TandT and Walmart, show more variability. This discrepancy between aggregate and individual trends highlights the importance of examining vendor-specific data to avoid drawing misleading conclusions from the aggregate trend alone.

4 Discussion

4.1 Correlation vs. Causation

The graphs reveal clear associations between vendors and their average discount rates, as well as month-to-month variations in discount trends. The bar chart in Figure 1 shows different average discount rates across vendors, with TandT and Walmart offering higher discounts on average than other vendors. Similarly, the line chart in Figure 2 shows monthly fluctuations for each vendor, suggesting that vendors adjust their discounts dynamically, potentially in response to market factors, seasonality, or promotional strategies. However, while these patterns suggest correlations between vendor identity and discount rate behaviors, they do not imply causation.

It is important to recognize that factors not captured in the dataset could influence discount trends. For instance, external factors like competitive pricing, supply chain costs, and consumer demand may impact a vendor's discounting strategy over time. Without data on these underlying factors, it is impossible to determine if the vendors' identities directly cause the observed discount patterns. As such, while the analysis highlights associations, we cannot conclude that being a specific vendor causes particular discounting behaviors. Further research with additional data would be required to explore causal mechanisms behind these trends.

4.2 Missing Data

The accuracy and representativeness of our analysis depend on the completeness of the data for each vendor over the observed period. If any data points were missing for certain months or vendors, this could impact the trends seen in Figure 2 and the average discount rates shown in Figure 1. For example, if a vendor's data were missing for a period when discounts were typically high or low, it could skew the average discount rate, leading to misrepresentations in the comparative analysis.

In Figure 3, which depicts Simpson's Paradox, missing data points could further exacerbate discrepancies between aggregate and vendor-specific trends. If discount records for certain vendors were inconsistent, this could make it difficult to discern actual trends. Handling missing data carefully is essential to avoid biases in our interpretation. In this analysis, any such gaps would affect our ability to capture accurate month-to-month fluctuations or correctly interpret vendor-specific trends. Future studies should strive to include complete datasets or apply robust imputation techniques to fill any missing values, ensuring that the analysis reflects a fuller picture of discount patterns.

4.3 Sources of Bias

Several potential sources of bias may have influenced the findings in these graphs. First, selection bias may arise if the dataset disproportionately represents vendors known for higher or lower discounting behaviors. For instance, if vendors with frequent promotions or seasonal discounts are overrepresented, this could artificially elevate the observed average discount rates. In this case, the trends might not reflect the discounting practices of grocery vendors with more stable or minimal discount policies.

Another potential bias is measurement bias. The consistency and methodology used to calculate discount rates could vary across vendors. For example, some vendors may record only specific types of discounts, such as seasonal promotions or store-wide discounts, while excluding others. This inconsistency could distort the average discount rates, especially if certain vendors include a broader range of discounts in their reporting.

Finally, Simpson's Paradox, as illustrated in Figure 3, indicates an aggregation bias in our analysis. While the aggregate trend might suggest an overall stability or gradual increase in discount rates, individual vendors display unique trends that differ from the aggregate. This phenomenon emphasizes the need to disaggregate data to understand vendor-specific behaviors accurately, as aggregation can obscure meaningful individual patterns.

Recognizing these biases is crucial to interpret the data cautiously. Selection, measurement, and aggregation biases can each impact the observed trends and conclusions. Future research should attempt to mitigate these biases by diversifying the sample of vendors, standardizing discount measurement methods, and disaggregating data where appropriate to provide a more accurate and representative analysis of discount trends in the grocery sector.

References

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

Filipp, Jacob. 2024. "Project Hammer." 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/.

SQLite Development Team. 2023. *SQLite Database Engine*. https://sqlite.org.

van der Loo, Mark P. J., and Edwin de Jonge. 2021. "Data Validation Infrastructure for R." *Journal of Statistical Software* 97 (10): 1–31. https://doi.org/10.18637/jss.v097.i10.

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

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