Evaluating Average Grocery Discounts by Store and Month in Canada in 2024*

Fangning Zhang Cher Ning-Li Kevin Roe

November 14, 2024

In a period of high grocery prices, grocery price discounts have helped lower costs of living for Canadians. The paper evaluates average grocery discounts by store and month and found that the highest average discounts occured in September and at Walmart at $15.37. The results of the paper informs shoppers' spending patterns to find the best deal across different grocers.

1 Introduction

Since the end of the pandemic with rising inflation, grocery prices have been a contentious topic of concern for Canadians. While grocery chains have raised prices, they also have introduced various discounts to encourage shopping.

The database explores grocery price discounts from February, 28th, 2024 to November 14th, 2024, across various vendors such as NoFrills, Walmart, and SaveOnFoods. Based on our results the month of September had the highest average discounts at $4.18. In addition, Walmart had the highest average discounts in 2024 at $6.26. Finally, Walmart in September had the highest average discounts at $15.37. These results signal that Walmart is in-line with their corporate strategy and September is the most popular month for higher discounts potentially because of back-to-school sales.

Understanding the results of the paper suggests when what month and grocery chains has the highest average discount for each month, allowing us to make better shopping decisions.

The paper is structured as follows: Section 2 focuses on the data and measurement methods; Section 3 focuses on the results; Section 4 evaluates correlation vs causation, missing data and different sources of bias.

*Code and data are available at: https://github.com/Kanghyunroe/grocery_discount.

2 Data

The dataset was obtained from Filipp (2024), and contains pricing information about products offered across 8 major grocers across Canada — Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart and Save-On-Foods. The pricing data was compiled from the vendor's websites directly. One table, named Product, contained information about the product name and vendor. Another table, named Raw, tracked all the price changes of these products, including the original price, new price, and time of price change each time it occurred. The time frame that the data covers is from February 28, 2024 until November 14, 2024 around noon.

Using SQL, the data was cleaned and organized to locate patterns. This dataset contained many entries where data was incorrectly entered, often with the word "was" under the variable tracking original price. There are also events that are not price changes, such as out of stock notices, that are recorded in Raw. These events are distinguished by a null value in the old_price column. Both of these occurences were filtered out using SQL. To find patterns, the average discount amount for each month and each vendor was also calculated using SQL.

We used the statistical programming language `R` (R Core Team 2023) and packages `tidyverse` (Wickham et al. 2019), `here` (Müller 2020), and `ggplot` (Wickham 2016) to produce graphs to better visualize the data cleaned by SQL.

3 Results

Figure 1 shows how average nominal discounts varies for each month, highlighting that September had the highest discounts in 2024.

Figure 2 shows the average discount amounts for each vendor, showing that Walmart had the highest discounts in 2024, on average.

Figure 3 breaks down the averaged discount amounts further into each month and vendor, showing the increase in discount amount observed in September is mainly from Walmart.

4 Discussion

4.1 Correlation vs Causation

It is important to distinguish correlation and causation for our results. Even though September saw the highest discounts, it is not the fact that it being September caused lower prices. Moreover, Figure 3 shows that Walmart significantly pulls up the average discounts showed in Figure 1, indicating that Walmart might have more aggressive discounts in September than other companies to capture more customers during back-to-school sales. While there are multiple relationships to explore, we cannot claim that the time of the year causes higher

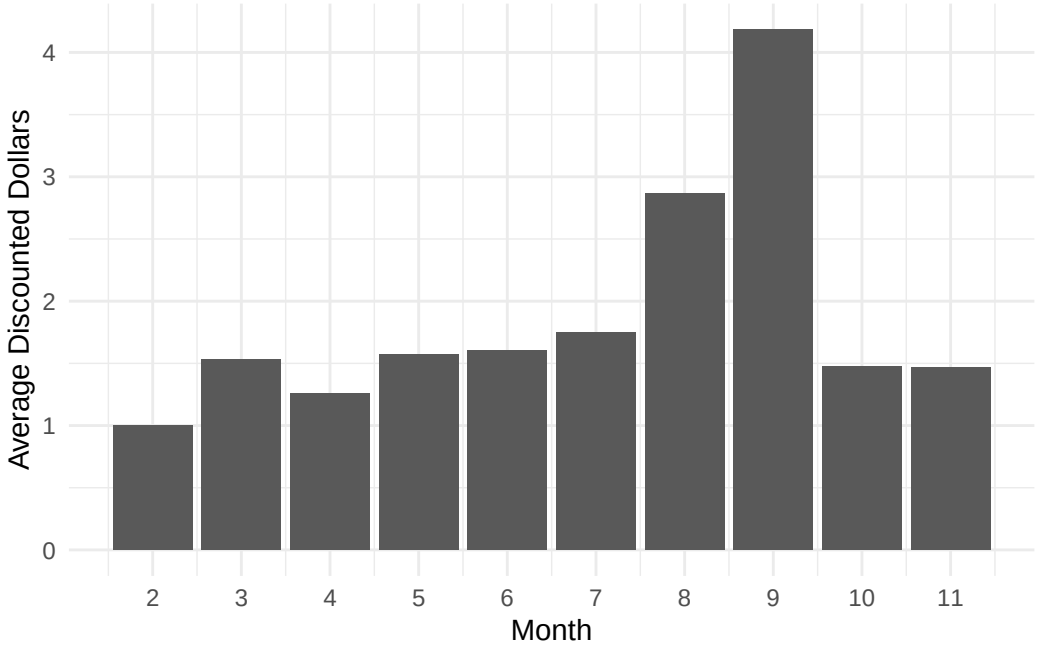

Figure 1: Average Discounted Dollar Amount Each Month

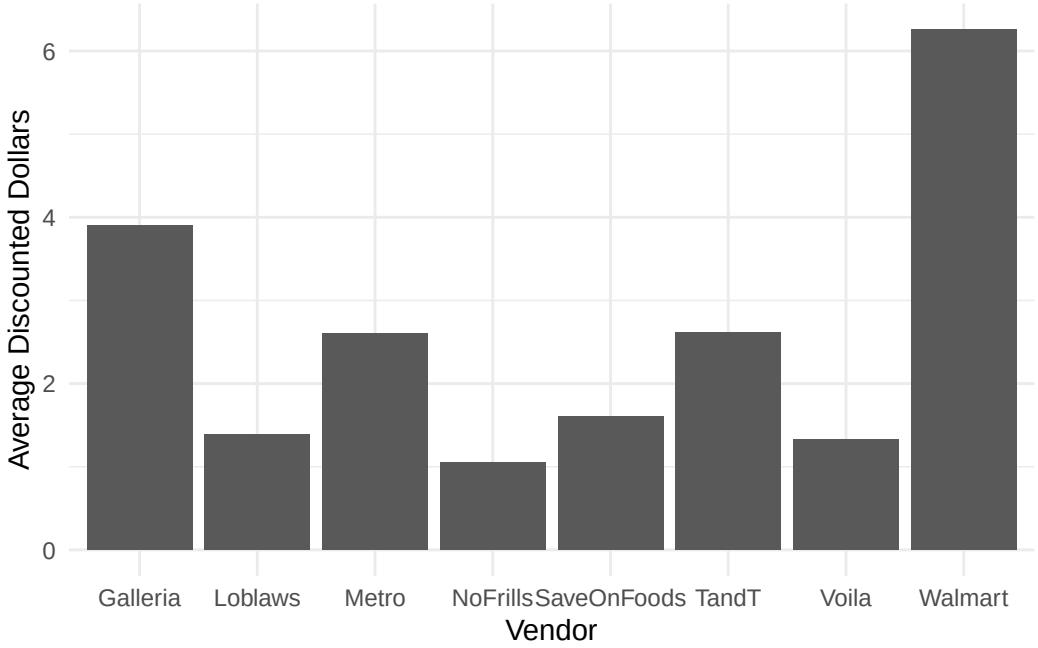

Figure 2: Average Discounted Dollar Amount Provided by Each Vendor

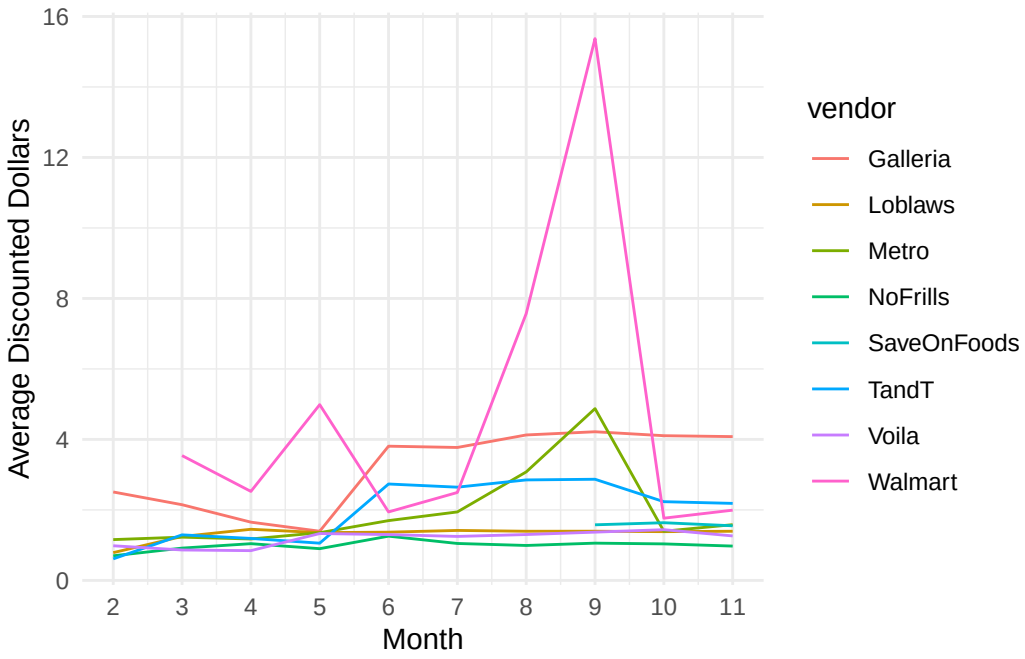

Figure 3: Average Discounted Dollar Amount Provided by Each Vendor Thoughout the Months

grocery discounts. Further analysis is needed to explore what predictor variables influence grocery discounts, which would provide unique insights.

4.2 Missing Data

The dataset only records data from February 28th, 2024 to November 11th, 2024. The problem is then that the February entry in Figure 1 does not accurately reflect the average grocery discount. For future study, I'd want to have a greater variation of our data across multiple years so we can find a more accurate the average grocery discount for each month.

Moreover, the database had a lot of errors where either discounts were not listed, not all the entries were discounts, and the discount values were listed in text. All three errors made certain values unusable, leading us to filter them out of the dataset.

4.3 Sources of Bias

There is significant measurement bias in the dataset because the method of recording discounts either via text or numerically introduces bias in the dataset. This hurts the validity of our results because multiple data points had to be removed to allow for interpretation. In

addition, without further research, one must be careful in establishing that there is a relationship between September and higher discounts. While we observed that 2024 September saw higher discounts, our initial hypothesis on back-to-school sales driving the discount might be due to our confirmation bias. However, further research is needed to validate the conclusion. Therefore, it is important to check our preconceived biases and not let it taint our data.

References

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

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.