

Exploring the Price of Egg Dozens*

Liam Wall

November 21, 2024

Understanding the pricing behavior of large and small food vendors is essential, particularly for staple goods like eggs. By using SQL and R to clean, manipulate, and visualize data on price changes for a dozen eggs over a six-month period, we identified vendor-specific trends. Our analysis reveals that larger vendors tend to maintain stable prices, while smaller vendors experience more frequent price changes to remain competitive. Additionally, smaller discount vendors are often priced higher than their competitors. These findings shed light on the different strategies vendors employ based on their market position and challenge common perceptions about discount vendors.

1 Introduction

The grocery retail industry in Canada is dominated by a few key players, each employing distinct pricing strategies to maintain market share in a competitive environment. Among these, Loblaw's has consistently held the top spot, commanding an estimated 29% share of the grocery retail market in 2022. Walmart, a multinational retail giant, followed with roughly 8% of the market. These two vendors, alongside competitors like Metro Inc. and Sobeys (owner of the online grocery platform Voila), shape the pricing landscape for essential goods such as eggs (Ozbun 2022).

Metro Inc., one of Loblaw's primary competitors, has made significant financial strides in recent years. In 2022, Metro achieved a record net income of 922 million Canadian dollars, with the majority of its stores located in Québec. Meanwhile, Walmart and Loblaw's remain the most popular grocery retailers in Canada, with Walmart being a regular destination for nearly half of Canadian shoppers in 2019 (Ozbun 2022). These retailers' pricing strategies are critical in influencing consumer behavior and defining market trends.

This paper examines the pricing behavior of these key vendors with a specific focus on the price of a dozen eggs. Using a dataset covering price changes over a six-month period, this study

*Code and data are available at: <https://github.com/Lwall02/Project-Hammer>

explores vendor-specific trends and pricing strategies. The findings reveal how large vendors, such as Loblaws and Walmart, exhibit price stability, while smaller or more specialized vendors, including discount retailers like No Frills, experience greater variability in pricing. These insights provide a deeper understanding of the competitive dynamics within the Canadian grocery industry and challenge common assumptions about discount pricing.

This paper was made using the open source programming language R Core Team (2023). The dataset used in this study was sourced from (Filipp 2024) and cleaned using SQL for consistency and analysis. SQL’s query capabilities allowed for efficient manipulation of paired price data, which was then visualized in R using the tidyverse package (Wickham et al. 2019) and ggplot2 (Wickham 2016).

2 Data

The dataset used in this paper hails from Project Hammer. Project Hammer is a personal project by Jacob Filipp in an effort to drive competition while reducing collusion in the Canadian grocery market. The basis for this project is compiling a historical database of the prices of any and all Canadian goods that are made available by Canada’s largest food vendors. Filipp makes this database available to the public in CSV, SQL, and Excel in easy to use formats. Furthermore, Filipp encourages the use of this database for academic analysis and even legal action. (Filipp 2024)

The information in Project Hammer’s historical grocery price is from a screen scrape of a website UI. Although the actual website is never referenced, any entry in the dataset is from said website with the price scraped from the pick up in store option for a specific neighborhood in Toronto (also not mentioned). Being that is not using an API or a more reliable method, this comes with many instances of missing information. There are a total of 129,420 different products from eight vendors (Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart Canada, and Save-On-Foods). For these products, the webscraping has collected 12,027,666 instances of prices for these goods. Each individual instance is the price of one product at a time recorded down to the nearest minute.

For each entry in the dataset there is the following information: - the date and time the information was collected for this product - the vendor supplying the product - a unique product identification number for the purposes of this dataset - the product name as scraped from the website - the units of the product (items per package, weight, volume, etc.) - current price at the time of extracting the information - an “old price” or the original price if the product is on sale at the time of extracting - price per unit if the unit is sufficient for this calculation - a note which may or may not provide additional information like “Out of Stock”, minimum purchase specifications, etc.

2.1 Data Specific to Egg Dozens

For the purposes of this paper, we look only at the price of a dozen eggs. In order to do this, all data manipulation was completed using SQL. We found a total of 438 entries relating to egg dozens that contained sufficient information in order to examine a change in price over time. To do this we filtered the data down to entries where “eggs” is mentioned in the product name and “12” is mentioned in the units. Although one may assume this cuts out a lot of information relevant to this paper, it does not. There were zero entries where the word “dozen” appeared in the units column. Furthermore we used “eggs” instead of “egg” because “egg” allows for a multitude of other products that include phrases like “veggies.” We needed to specify both “eggs” and “12” because there are many instances of the price of eggs in quantities other than a dozen. For example there are instances of “eggs” appearing in the product name while “12” does not appear in the units; “duck eggs” and “veggies” both contain “egg” but neither will contain 12. Lastly, we only looked at the oldest and most recent instance for each individual product based on the date listed.

Using the above criteria we found 452 entries. By manually searching through the data we found instances where some entries only have price information for a singular date. We also found instances of products that were not egg dozens. For example “duck eggs” with units “312g” which was included based on the search criteria. Also a pack of 30 eggs which had the units “\$12.49” which also was included. After this manual double checking of the dataset we found exactly 438 entries of egg dozens where there is two listed prices at dates that are more than a day apart. That is exactly two entries each for 219 instances of egg dozens.

With these 438 entries, still using SQL, we then combined the data such that both the “old” price and “new” price are in their own column as well as adding a change in price column for each product. Bringing the total number of rows to 219. That is 219 different egg dozen products at a certain vendor with an old and recent price along with their accompanying dates and its change in price.

3 Measurement

Although previously mentioned, the data is collected by web scraping a website that is not mentioned. It is however mentioned that all prices and products are listed for the nearby stores in a specific Toronto neighborhood and using the pick-up-in-store option. In this way we have consistency of the extracted data since there are no location differences and the store’s extra fees are not included in the pick-up-in-store option. That is the prices for each vendor are specific to that location and should reflect the pricing behaviors due to competition between these vendors. As opposed to vendors far apart that do not compete with each other.

Another aspect of the measurement in this paper is the price. The web scraping tool takes the current price and units at the time of scraping. It is important to consider products with the same units and their corresponding price because there can be many differences in how

the product is listed to the public. For this reason we made sure only to consider dozens of eggs where the price pertains to the whole dozen and not an individual egg. The price is always the exact price you would see if you were physically shopping in the store due to the pick-up-in-store option. Added online fees are not apart of this dataset.

4 Results

Now to look at some of the findings from this data. First we look at a collection of all the data faceted by each vendor Figure 1. This graph shows a faceted bar graph showing the frequency of different changes in price for a dozen eggs.

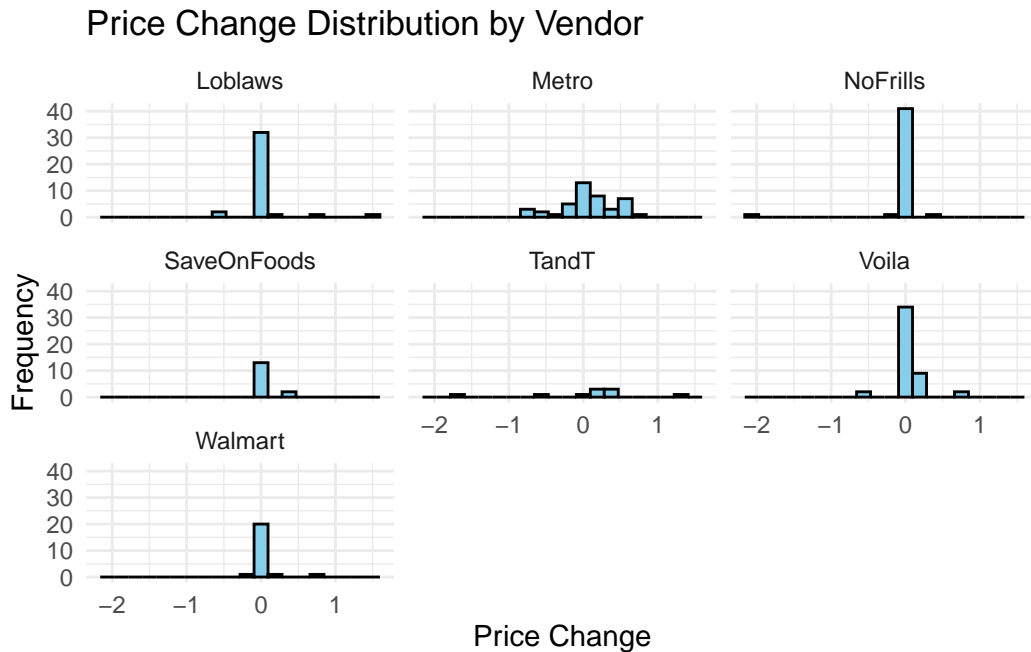


Figure 1: This graph shows the change in price for a dozen eggs for each vendor.

In this next graph we see the aggregated change in prices over all vendors regardless of vendor Figure 2. The change in prices are grouped into bars where each bar accounts for plus or minus 10 cents around the value which it lies above. For example the bar above a \$0.00 price change accounts for those prices changes contained in the interval -\$0.10 to \$0.10.

This next figure shows a more detailed look into the severity of price changes each vendor employs and their frequency Figure 3. We see a box plot showing the interquartile range of on the price changes and the vendor's outliers. We can look at Figure 4 to get an idea of the time span of the data collected for each of the vendors.



Figure 2: This graph shows the frequency of all changes in price for the 219 products (dozen eggs).

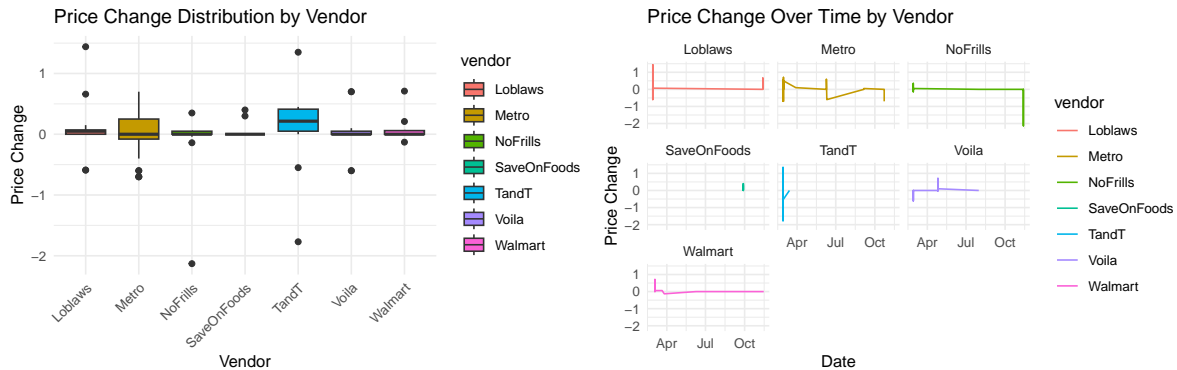


Figure 3: This graph shows the distribution of the vendor's change in price for a dozen eggs using a box plot. The box represents the interquartile range with the lines as its extremes. And the dots are the outliers.

Figure 4: This line graph shows the price changes in a dozen eggs for each vendor. The line tracks all prices of the vendor's products which is why it is not a continuous line but rather the aggregated look at all the prices of a dozen eggs and how they change.

Lastly, let us look at a table, Figure 5, showing the typical statistics we are curious in when examining data. This table shows the average change in price per vendor, along with it's minimum and maximum change in price as well as the number of products we have information for per vendor.

vendor	Mean Change	Median	Minimum	Maximum	Occurences	Average Price
TandT	0.092	0.215	-1.77	1.35	10	7.46
SaveOnFoods	0.047	0.000	0.00	0.40	15	6.88
Walmart	0.047	0.000	-0.13	0.71	23	6.34
Loblaws	0.056	0.050	-0.59	1.44	37	6.75
Metro	0.054	0.000	-0.70	0.70	43	6.61
NoFrills	-0.026	0.000	-2.13	0.35	44	6.49
Voila	0.026	0.000	-0.60	0.70	47	6.21

Figure 5: This table shows the mean, median, minimum, and maximum in changes in price for a dozen eggs for each vendor. It also shows the number of products we are tracking for each vendor and the average price of a dozen eggs at each vendor.

5 Discussion

What is most obvious from these results is that each vendor is somewhat unique and there is no exact matches despite the fact some vendors are selling the exact same products. It is difficult to make assumptions from such a small time period and in the case where some vendors have sparse data, but nonetheless there are still takeaways.

What is clear is that Loblaws, No Frills, Walmart, and Voila maintain more stable trends over time, indicating consistency in pricing strategies. Whereas Metro and T&T display more dynamic fluctuations in price change. SaveOnFoods has very sparse data. What we can takeaway from this is the notable difference in size of company between those with stable and unstable fluctuations. The vendors with stable trends over this time period are generally the larger corporations like Walmart, Sobeys (owns Voila) and Loblaws. Metro is a direct competitor with Loblaws and is clearly trying a different pricing strategy to maintain such competition. Large corporations are more immune to smaller issues, like supply chain issues, in short periods of time which shows in their lack of fluctuation in egg dozen prices. However smaller vendors may feel the effects much quicker and have to respond quicker in order to bolster sales. Hence we see more fluctuation in smaller period of time.

Another takeaway from the results above is that T&T and SaveOnFoods, two vendors generally considered to be more of discount establishment, have the highest average price of a dozen of

eggs. The lowest average price of a dozen eggs is from Voila (an online grocery vendor) and Walmart. This is not a surprising find that these large corporations are able to provide the cheapest option but it does show that discount vendors are not necessarily cheaper when you expect them to be for staple goods like eggs.

Lastly, if we look at the box plot we can get a nice overview of which vendors experience less price change. If we also take into account the number of products we are considering for each vendor it is clear that NoFrills, Voila, and Walmart have the most consistent prices. Although Loblaw's shows little change in price over time, any change that does happen is larger compared to the previously mentioned vendors. Hence why on the table Loblaw's has the second highest average change in price but in the line graph it seems completely stable. We can also see that Metro, a direct competitor to Loblaw's, actually has a smaller average change in price than Loblaw's. It is that there are many more smaller changes in price that makes Metro seem like it has a lot of fluctuation in price. The less common but large changes in price have a lot more effect on the average than the more often but smaller changes.

In conclusion, I would shop at Walmart or the online vendor Voila if I wanted to be able to confidently predict my grocery bill cost. They not only have small average changes in price over time, but also offer the cheapest product on average.

5.1 Correlation Versus Causation

While this paper explores trends in the price changes of a dozen eggs across various vendors, it is essential to recognize the limitations of inferring causation from these observed correlations. For example, vendors with smaller market shares, such as T&T or No Frills, exhibit more frequent price changes compared to larger retailers like Loblaw's or Walmart. However, this observation does not necessarily mean that a smaller market share causes more volatile pricing.

5.2 Missing Data

Missing data plays an important role in these findings because we cannot be certain all occurrences for a dozen eggs were considered in the data set I used. There could have been more occurrences with different values of prices that could have changed the outcome. Especially since only 15 occurrences were found for SaveOnFoods and 47 for Voila, any missing occurrences for SaveOnFoods could have had a large effect on its results.

We also had to discard information that did not end up in our dataset based off of our search conditions. Some data with NULL values were not useful for finding average prices or change in price and therefore had to be discarded. In the case that these data points had a numeric value for the old or current price, that is still valuable information to this paper but unusable.

5.3 Sources of Bias

One of the most obvious sources of bias is that not all vendors are represented equally. Some vendors have many more data point for egg dozens than others so we can't get a full picture of what their pricing behavior is really like. Another source of bias is the location of these vendors. Based on their location they may undergo more or less hardships in areas like supply chain. They could also be closer or further from populated areas which would effect their prices. This source of biased is minimized since we are taking vendors that are close together so ideally they should all be experiencing these issues to a similar degree.

There is also the idea that some vendors would sell a different type of food. For example, Loblaw's may be considered a little higher classed than Walmart, so therefore maybe the base price of the eggs they choose to sell is higher in the first place. With that comes many other factors that could effect the price of eggs that is out of their control.

Lastly, there is a large bias on the results in this paper in that this only spans 6 months. It is hard to get a full picture of the pricing behavior of a company only looking at one product over 6 months.

Ultimately, this paper lacks real ability to be able to determine any information about Canadian groceries with great confidence. Simply looking at one product over 6 months may not be sufficient to make actual conclusions. In future research, more staple goods should be considered with a priority and being able to track them over a longer period of time.

References

- Filipp, Jacob. 2024. *SQLITE Canadian Grocery Price Data*. Project Hammer. <https://jacobfilipp.com/hammer/>.
- Ozbun, T. 2022. *Leading Grocery Retailers in Canada in 2022, by Market Share*. Statista. <https://www.statista.com/statistics/481019/leading-grocery-retailers-by-market-share-canada/>.
- 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 Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.