

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 [CITATION].

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 [CITATION]. 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.

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. [CITATION]

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 [CITATION FOR 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 contains “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 there 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 scrapping 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 scrapping. 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

5 Discussion

6 References