Vendor-Based Pricing Dynamics of Eggs in Canada*

A Comparative Analysis of White and Brown Egg Market Prices

Colin Sihan Yang Lexun Yu Siddharth Gowda

November 14, 2024

This study provides an empirical analysis of the price dynamics of white and brown eggs across various vendors in Canada, derived from a robust dataset obtained through Project Hammer. By meticulously cleaning and processing the data with advanced statistical techniques in R, we uncover significant insights into the pricing variations between different egg types and vendors. The research identifies and compares the current and historical prices, facilitating an understanding of market trends at the vendor level without external economic or regulatory influence. This approach allows for a granular analysis of factors directly influencing egg prices at the retail level, providing a valuable resource for industry stakeholders aiming to strategize based on competitive pricing and consumer preferences. Our findings serve as a preliminary step towards a comprehensive exploration of micro-level market behaviors within the Canadian egg industry.

1 Introduction

Egg prices have fluctuated in recent years due to various economic and environmental factors (Bundale 2023). In Canada, egg pricing and marketing are regulated by the government. According to Ministry of Agriculture, Food and Agribusiness and Ministry of Rural Affairs (2023), the Canadian Hatching Egg Producers (CHEP) determines annual national production levels for hatching eggs with input from an advisory committee composed of industry and government representatives. These numbers are then allocated to provinces and adjusted throughout the year to meet demand. However, as Filipp (2024) noted, research on Canadian grocery food prices remains limited.

*Code and data are available at: https://github.com/yulexun/projecthammer.

This paper examines the factors influencing egg prices in Canada. This approach aims to identify areas that could explain recent price trends, providing a foundation for future studies on the Canadian egg market.

Our analysis shows that while the relationship between egg product price and per-unit cost is weak, it strengthens significantly when egg quantity is factored in. Some brands, like Rabbit River Farms, are consistently more expensive than others, such as President's Choice. Wide price ranges for most brands suggest they offer diverse product lines, addressing various consumer needs.

The remainder of this paper is structured as follows: Section 2 provides an overview of the data. We then present our results in Section 3 and discuss the implications, limitations, and future research directions in Section 4.

The data gathering and analysis is done in R (R Core Team 2023) with the following packages: knitr (Xie 2014), tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016), dplyr (Wickham et al. 2023), arrow (Richardson et al. 2024), here (Müller 2020), modelsummary (Arel-Bundock 2022) and lubridate (Grolemund and Wickham 2011).

2 Data

2.1 Measurement

The raw data is obtained in SQLite form from Filipp (2024)'s website named Project Hammer. The raw dataset contains two tables. The first table contains various product, product names, and the vendor of the products. The second table contains old and current price, as well as price per unit.

2.2 Cleaning

We clean the data with the clean_data.sql script and the exploratory_data_analysis R script. We retrieve data related to "White Eggs" and "Brown Eggs" from two tables, `raw` and `product`, with the following steps:

- Selects various columns including `nowtime`, `vendor`, `product_id`, `product_name`, `brand`, `current_price`, `old_price`, `units`, and `other`.
- Converts `price_per_unit` to a numeric format, stripping out symbols (like `$` and `/item`) and renaming it as `price_per_unit_numeric`.
- Joins `raw` and `product` tables on `product_id`.
- Filters for products containing "White Eggs" or "Brown Eggs" in their names and ensures `price_per_unit` is not `NULL`.

2.3 Summary

Figure 1a displays the distribution of vendor counts clearly, revealing a significant variation in the frequency of different vendors. From the visualization, it's evident that Voila dominates with the highest count, significantly outpacing other vendors like SaveOnFoods, NoFrills, Metro, and Loblaws, which have a more moderate presence. This dominance of Voila suggests a strong market presence or a higher transaction volume compared to the others. The much lower count for SaveOnFoods indicates it has a smaller market share or fewer transactions. Such trends highlight the market dynamics and competitive landscape among these vendors, which could be critical for strategic business analyses and decision-making in related sectors.

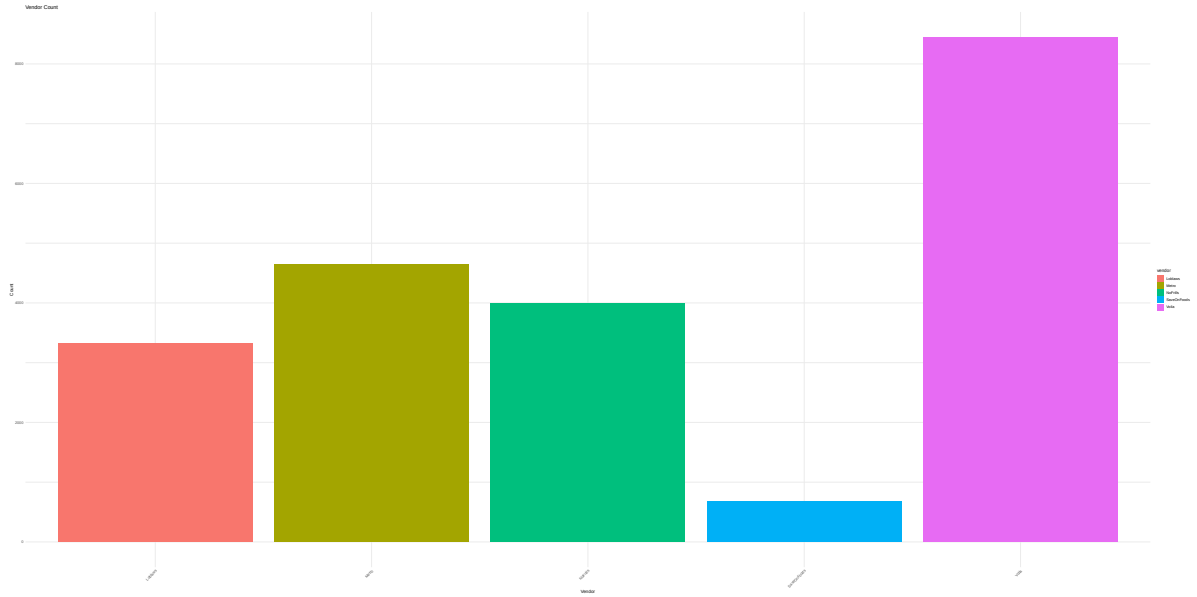
Figure 1b The brands "Darvonda" and "Excelsior Farms" have a far greater count compared to the rest signifying them as the dominant brands in this dataset. The rest of the brands exhibit frequencies spread out over a wide spectrum but are on the whole very sparse, with a number of brands being barely in existence. This gives the impression that a few brands lead with a larger market share, perhaps because of consumer preference or successful distribution strategies, while several others have only niche market segments. The visual dispersion from left to right also indicates that the market is not only concentrated around the major players, but several small players have also a very little share in the market.

3 Results

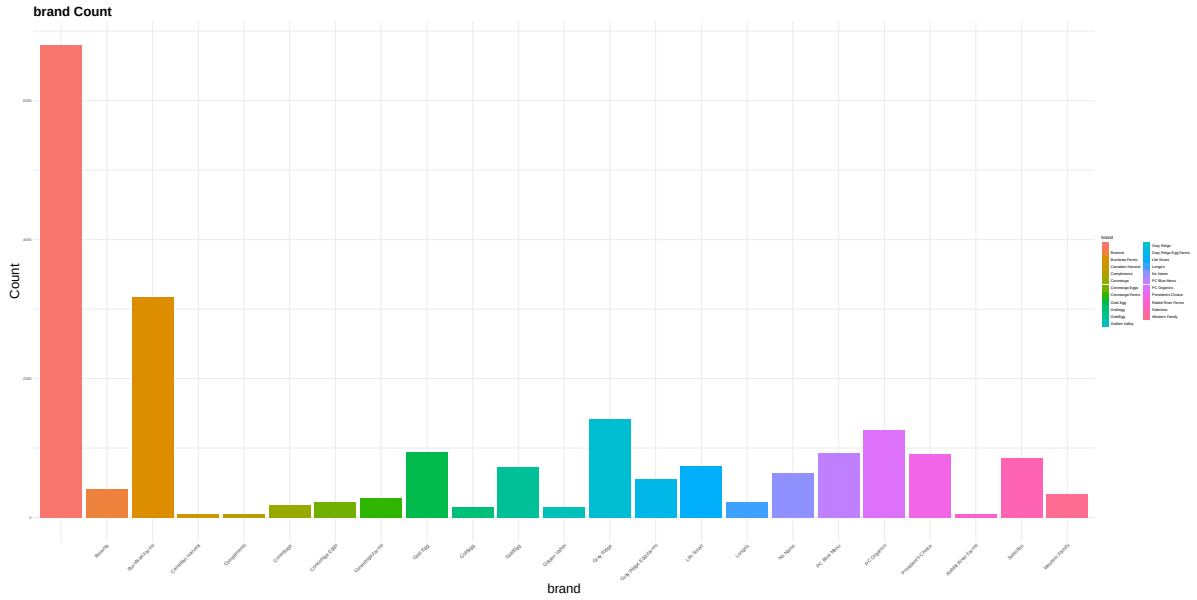
```
# A tibble: 6 x 9
  nowtime             vendor product_id product_name   brand current_price units
  <chr>               <chr>       <int> <chr>          <chr>         <dbl> <chr>
1 2024-06-22 10:55:00 Voila        2518 Gold Egg Omeg~ ""             5.69 12 p~
2 2024-06-10 23:58:00 Voila        2518 Gold Egg Omeg~ ""             5.69 12 p~
3 2024-06-11 18:55:00 Voila        2518 Gold Egg Omeg~ ""             5.69 12 p~
4 2024-06-12 11:53:00 Voila        2518 Gold Egg Omeg~ ""             5.69 12 p~
5 2024-06-13 10:46:00 Voila        2518 Gold Egg Omeg~ ""             5.69 12 p~
6 2024-06-14 11:02:00 Voila        2518 Gold Egg Omeg~ ""             5.69 12 p~
# i 2 more variables: price_per_unit_numeric <dbl>, numerical_units <int>
```

From Figure 2, there appears to be a weak linear relationship between current egg product price and the price per a single egg unit. However, after including egg quanitity, the relationship between current egg product price and egg quantity as a predictor of cost of an egg per a single unit is quite strong.

Figure 3 shows that some brands have more expensive egg quantities than others. For example, Rabbit River Farms is more expensive than President's choice. Also, it is obvious that most of the brands offer a variety of products since the IQR of their prices are quite large.



(a) bar chart for vendor



(b) bar chart for brand

Figure 1: categorical variable bar chart

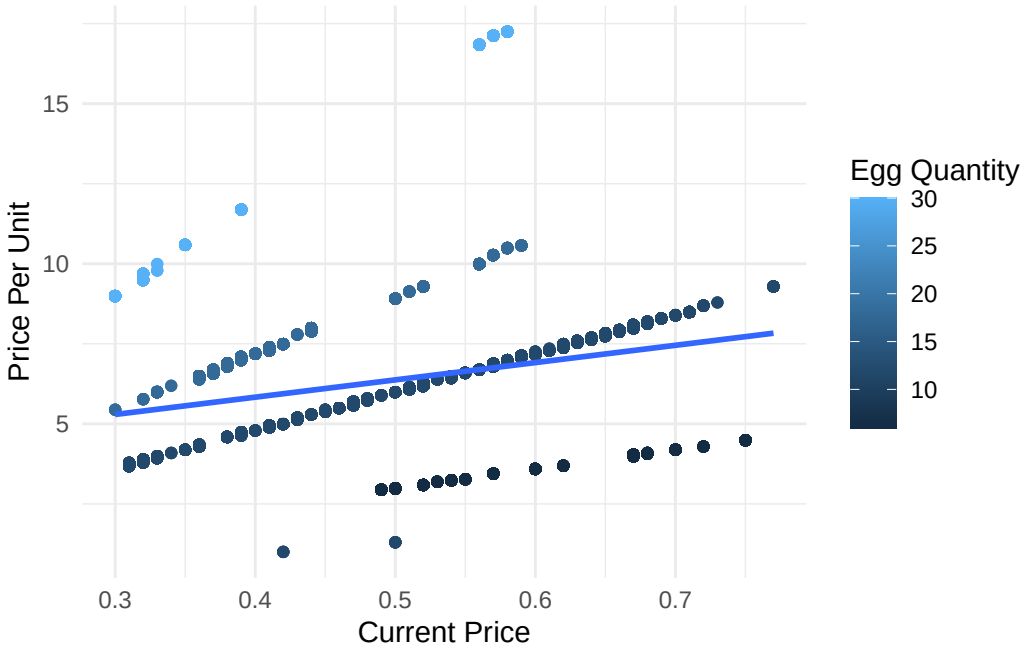


Figure 2: Higher Quantity Egg Products Are More Expensive

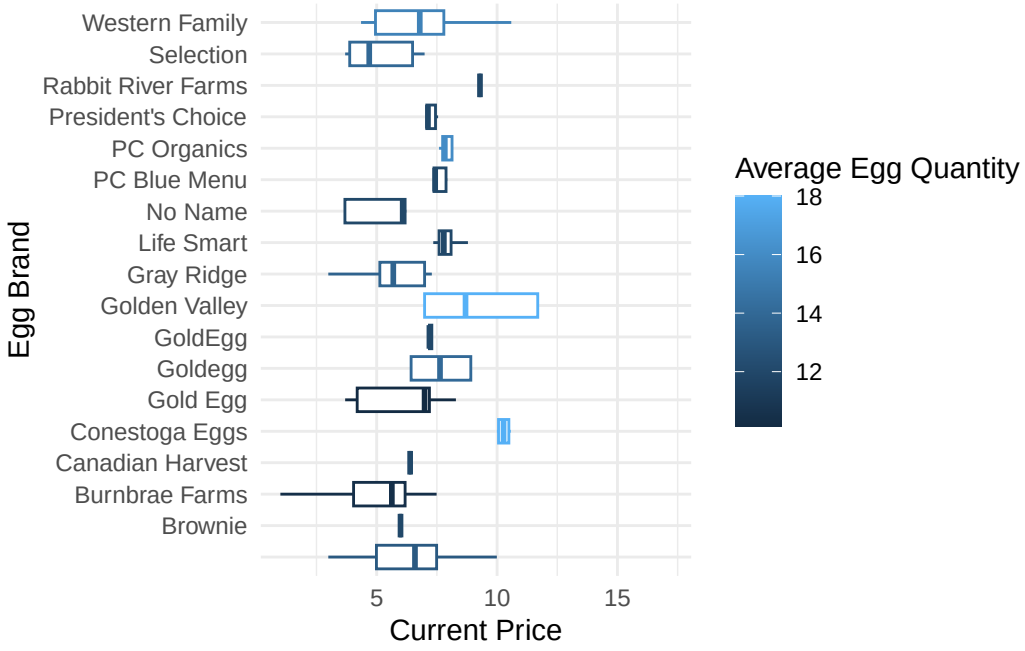


Figure 3: Some Egg Brands are More Expensive than Others

4 Discussion

4.1 Correlation vs. Causation

From Figure 3, some companies eggs are more important than other, for instance Gray Ridge is more expensive than Burnbrae Farms. However, this is not true since Golden Valley sells more eggs per one batch than Burnbrae Farms. While this is not a traditional cause of correlation vs. causation this is a similar siutation. Since only considering the relationship between the egg brand and the price per batch, we might think some brands are more expensive than others, but when considering the egg qunatity and price per single egg (Figure 2), we realize a different relationship olds true.

4.2 Missing Data

In this analysis, the old egg price variable was removed. This is because almost all of the data entries did not include this column. This limited our analysis since it made it quite difficult to analyze how specfic egg product prices changed over time.

4.3 Sources of Bias

In Figure 2, by only looking at the linear model lines, there does not appear to be a strong relationship between current egg price and the price per unit. However, by looking at the graph in general, there appear to be four different positive, strong, linear relationships between current egg price and price per egg. This suggests that there is another variable affecting the reltionship, which based on the graph is the quantity of eggs in that batch. Thus, most conclusions regarding current egg price must be interpreted in a manner that accounts for the number of eggs in that batch.

4.4 Weaknesses And Next Steps

The biggest issue in this paper is the limitation of the data. The team had limited numerical data and it was only regarding egg price and egg quantity. In the future, we would gather purchase data to see how changes in egg prices could change consumer spending habits. We would also like to incorporate other historical data like interest changes, tax changes, and how those affect consumer egg spending habits and egg prices.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bundale, Brett. 2023. "Why Canada Has Avoided Egg Shortages, Major Price Spikes Seen in U.S. - National | Globalnews.ca." *Global News*. https://globalnews.ca/news/9457616/egg-prices-shortage-canada-us/.

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

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Ministry of Agriculture, Food and Agribusiness and Ministry of Rural Affairs. 2023. "Supply Management Systems for Eggs, Poultry and Dairy." *Ontario.ca*. http://www.ontario.ca/page/supply-management-systems-eggs-poultry-and-dairy.

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/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

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.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.