Comparing Garlic and Leek prices in Major Superstores of Canada*

Shreya Sakura Noskor Khushaal Nandwani Prankit Bhardwaj
Veyasan Ragulan

November 14, 2024

We look at the average price of leek and garlic, from June 2024 to November 2024. We find that

1 Introduction

Canadian grocery prices have skyrocketed in the past few years, often surpassing inflation (Tahirali (2024)). This paper looks at key products provided by the big 8 vendors, and finds the most to least expensive deals. Current prices are also compared to previous records to see trends across vendors in these products.

2 Data

R by R Core Team (2023) was used to analyze the data along with its library Tidyverse by Wickham et al. (2019). We also used Python Software Foundation (2023) to convert the data from sqlite database to csv. Finally, SQLite by SQLite Development Team (2023) was used to clean and manupilate the data.

We got the data from https://jacobfilipp.com/hammer/ in the form of sqlite database. We selected to analyze Garlic and Leeks because it was available in different vendors with the same name. Average price was taken for the current price per month for every vendor and a table was created for the same. This table was then converted into a .csv file to analyze in R using Python, where the script for it can be found in /scripts/export.py.

SQL was used for data manupilation because it can handle larger datasets better than CSVs.

*Code and data are available at: https://github.com/NotSakura/ProjectHammerExcer.git.

```r
library(readr)
leek <- read_csv("../data/02-analysis_data/leeks.csv")
```

```
Rows: 20 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (2): vendor, month
dbl (1): avg_price

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
garlic <- read_csv("../data/02-analysis_data/garlic.csv")
```

```
Rows: 20 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (2): vendor, month
dbl (1): avg_price

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

3 Results

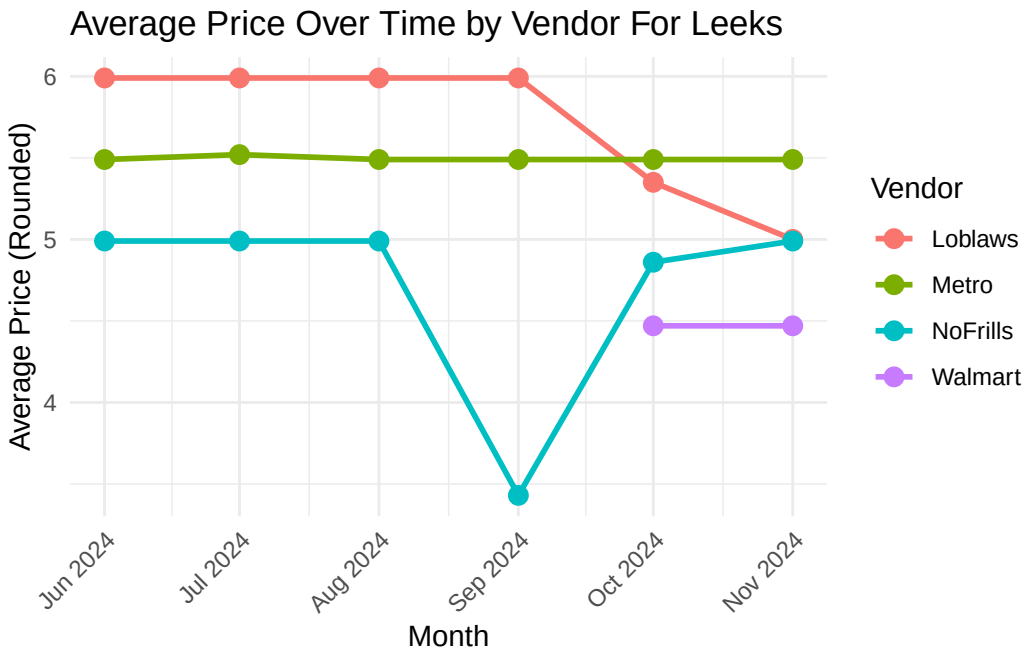
```r
library(ggplot2)
library(dplyr)
leek$Month <- as.Date(paste0(leek$month, "-01"))

# Round the Avg_Price to 2 decimal places
leek$Avg_Price <- round(leek$avg_price, 2)

# Create the plot
ggplot(leek, aes(x = Month, y = Avg_Price, color = vendor, group = vendor)) +
  geom_line(size = 1) +                  # Line graph
  geom_point(size = 3) +                 # Points on the line for visibility
  scale_x_date(date_labels = "%b %Y",    # Format the x-axis as month-year (e.g., Jan 2023)
               date_breaks = "1 month") +  # Set breaks to be 1 month
```

```
  labs(title = "Average Price Over Time by Vendor For Leeks",
       x = "Month",
       y = "Average Price (Rounded)",
       color = "Vendor") +                 # Label for the legend
  theme_minimal() +                       # Clean theme
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

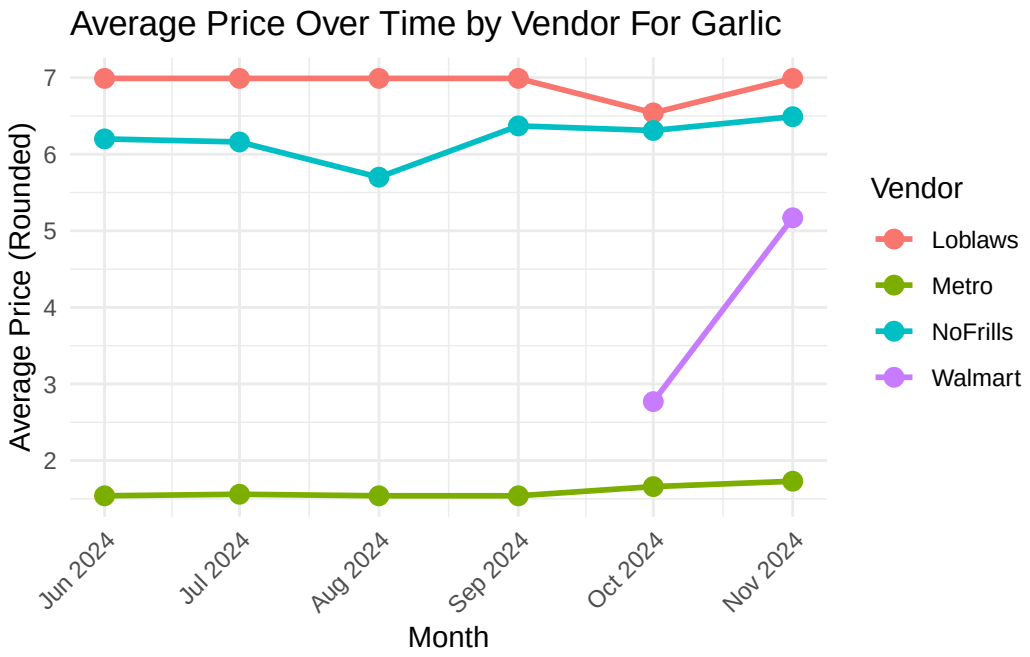



```
garlic$Month <- as.Date(paste0(leek$month, "-01"))

# Round the Avg_Price to 2 decimal places
garlic$Avg_Price <- round(garlic$avg_price, 2)

# Create the plot
ggplot(garlic, aes(x = Month, y = Avg_Price, color = vendor, group = vendor)) +
  geom_line(size = 1) +                   # Line graph
  geom_point(size = 3) +                  # Points on the line for visibility
  scale_x_date(date_labels = "%b %Y",     # Format the x-axis as month-year (e.g., Jan 2023)
               date_breaks = "1 month") +  # Set breaks to be 1 month
  labs(title = "Average Price Over Time by Vendor For Garlic",
```

```
      x = "Month",
      y = "Average Price (Rounded)",
      color = "Vendor") +                  # Label for the legend
theme_minimal() +                          # Clean theme
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



4 Discussion

5 Correlation vs. Causation

Correlation measures the relationship between two variables—how they move or change together. For example, if two variables (like hours studied and exam score) increase together, they have a positive correlation. If one variable increases while the other decreases (like hours of TV watched and exam score), they have a negative correlation. Importantly, correlation doesn't imply that one variable causes the change in the other; it only indicates that they tend to move in a related way.

In our graphs, we did not find any correlations among the variables. Each of the vendors changed their prices individually.

Causation implies that one event (the cause) directly affects another event (the effect). In other words, a change in one variable is responsible for the change in another. Causation is

a stronger claim than correlation and typically requires experimental or controlled studies to confirm because observational data alone often cannot definitively show causation due to other confounding factors.

Again there is no causation found in our graphs, and is not relevant.

6 Missing data

Project Hammer gets its information by scraping vendor website's for pricing. Vendors have their own APIs to track inventory and pricing across their stores, but these are not made available to the public. This is why on certain days and with certain vendors, the database lacks any information of product pricing, the attempt to scrape for prices on those days failed in some way. Filippa started this database using pricing on a small selection of products. This remained so until sometime in July, where he increased the amount of products he tracked. This means that only a select few products have approximately a year's worth of data available to them.

7 Source of Bias

The prices of garlic and leek in Walmart, Loblaws, and No Frills are used in this analysis. Of the two most significant sources of bias, possibly affecting this analysis, regional pricing and different product varieties stand out. Regional pricing may result in price variation due to a wide range of factors that include transportation cost, local supply and demand, and regional supplier agreements. For instance, prices of garlic and leek for the same vendor may vary between an urban and a rural setting and might make a vendor look more expensive simply because of the regions it operates within in the dataset. Product variety further complicates the comparison, as different vendors can have a variety of garlic and leeks they sell, such as organic, imported, or in bulk. These types often also carry different prices, and if one vendor mainly sells high premium or bulk produce, while another mainly sells regular, single-item offerings, it could skew average price comparisons. Trying to work out these biases by matching the prices within comparable regions and types of garlic and leeks, where possible, would go a long way toward comparison of pricing for each vendor in a far more fair manner.

References

Python Software Foundation. 2023. *Python: A Programming Language for General-Purpose Computing*. Wilmington, DE, USA: Python Software Foundation. https://www.python.org/.

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: A C Library that Implements a Self-Contained, Serverless, Zero-Configuration, Transactional SQL Database Engine*. SQLite Consortium. https://www.sqlite.org/.

Tahirali, Jesse. 2024. *Food Prices Continue to Outpace Inflation in Canada*. CTVNews. https://www.ctvnews.ca/business/food-prices-continue-to-outpace-inflation-in-canada-1.7074295.

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.