

# Analyzing Canadian Grocery Price Trends to Identify Market Patterns and Consumer Insights

Mingrui Li

2024-11-14

## Abstract

The Canadian grocery market has dealt with increasing concerns regarding price practices and market concentration, highlighting the importance of transparent and accessible pricing data. In light of this report, we will be making use of Project Hammer, a publicly available historic grocery price data set to examine grocery pricing practises from eight major Canadian grocery vendors: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods. The paper examines product varieties, price change over time, and price trend changes by vendors, extracting data through SQL and analyzing it using Python. Some key results that the analysis brings out relate to striking changes in prices and a detailed breakdown of products by vendors-a proxy for changes in market dynamics underlying the variations in consumer prices. It emphasizes that price transparency is vital for good competition in the grocery market and for informed consumer choice, while it also tends to make use of missing values and the biases inherent in observational data.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data</b>	<b>3</b>
<b>3</b>	<b>Measurement</b>	<b>4</b>
<b>4</b>	<b>Results</b>	<b>4</b>
4.1	Distribution of Price per Item . . . . .	5
4.2	Price Changes Over Time . . . . .	8
4.3	Vendor Product Variety . . . . .	10
<b>5</b>	<b>Discussion</b>	<b>11</b>
5.1	Correlation v.s. Causation . . . . .	11
5.2	Missing Data . . . . .	11
5.3	Sources of Bias . . . . .	11
	<b>References</b>	<b>12</b>

# 1 Introduction

Over the past years, the Canadian grocery market has been involved in controversies due to potential inflation in prices and concentration of a few strong retailers. These dynamics have thus raised debates on the need for better price transparency and full availability of grocery price data. It is based on this realization that Project Hammer was started to bridge this gap by compiling a publicly available database of historical grocery prices across major Canadian vendors. This dataset includes everything from grocery items to a wide range of vendors, including, but not limited to, Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods. It covers data starting February 28, 2024, onward.

This project utilizes data from Project Hammer to analyze grocery pricing trends, find out the fluctuation in product prices over time, and analyze product variety across vendors.

Data extraction was done in SQL for better manipulation, and the extracted data were analyzed in-depth by using Python. Key metrics considered in this regard are average prices per item, how their prices change over time, and how many different products each vendor offers. This analysis aims to reveal certain price behaviors and patterns of product availability that may be indicative of general market trends or strategies by the particular vendors themselves. This analysis is a good insight into grocery price behaviors, but it also navigates challenges on how to interpret correlation versus causation, handle the data when missing values exist, and possible biases inherent in the dataset. These will have implications for raising awareness of grocery price patterns across Canada, for informing consumer behaviors, and for contributing to policy discussions about creating a competitive and transparent grocery marketplace.

(Filipp 2024)

## 2 Data

The dataset used for this analysis originates from Project Hammer, an initiative aimed at increasing competition and reducing collusion within the Canadian grocery sector. The project's goal is to compile a comprehensive database of historical grocery prices from various top grocers' websites, making the data accessible for academic and legal purposes.

The dataset includes grocery price information from eight major Canadian vendors: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods. The data is collected from February 28, 2024, to the latest available update, providing time-series information on a range of grocery products.

The full data is available in two primary formats:

-CSV Files: The dataset is divided into two CSV files. The “product” file contains metadata about each product, while the “raw” file includes the time-series price data. These files can be joined using the id and product\_id columns to analyze prices over time.

-SQLite Database: For more extensive analysis, the entire dataset is also available as a single SQLite file. This database format facilitates efficient querying and filtering of data, which is particularly useful for large-scale analyses.

For this project, the data was accessed through the CSV format, with supplementary queries conducted in the SQLite database to examine pricing trends and vendor product offerings.

(Filipp 2024)

### 3 Measurement

The following metrics shall form the basis for analysis in order to achieve the objective of assessing grocery pricing trends and product availability on a vendor-specific basis:

**Average Price per Product:** The data was analyzed to calculate the average price for every product across vendors. This gives an idea of the range of prices that usually exist but also informs of both high-end and budget-friendly items.

**The trend of prices over time:** This is determined by comparing the initial and final recorded prices of a given product. This actually highlights items with a big increase or drop in price, pointing to possible inflationary pressures, seasonality, or promotional activities.

**Product Variety by Vendor:** Variety in the count of unique products offered from each vendor was measured in this analysis. This metric will describe the breadth of inventory across vendors and how this may influence customer shopping preferences, influencing vendor competitiveness.

The extractions and aggregation have been done in SQL, while visualizations were performed using Python to understand the pattern observed in the price set better.

(*Information Technology — Database Languages — SQL* 2023)

### 4 Results

The data shows a total of *102039* unique products across vendors. The following shows the head of all or extracted data from database.

Average Price per Item DataFrame:

	product_name	avg_price_per_item
0	Acorn Squash 1 Count	4.722026
1	Anise (Fennel) 1 Count	4.411569
2	Anjou Pears Large 1 Count	1.264194
3	Apple Fuji 1 Count	1.818177
4	Apple Granny Smith Large 1 Count	1.408956

Price Flow DataFrame:

	product_name	start_price	end_price
0	1% Yogurt Tubes, Bubble Gum Special Edition, K...	5.49	3.69
1	100% Butter Blueberry and Cheese Pastry	1.99	1.99
2	AROY-D QUAIL EGGS IN WATER 425G	4.49	4.49
3	ASSI COOKED SWEET BUN 520G	5.99	5.99
4	Ap Porto SQ. Egg Roll Pan	20.99	20.99

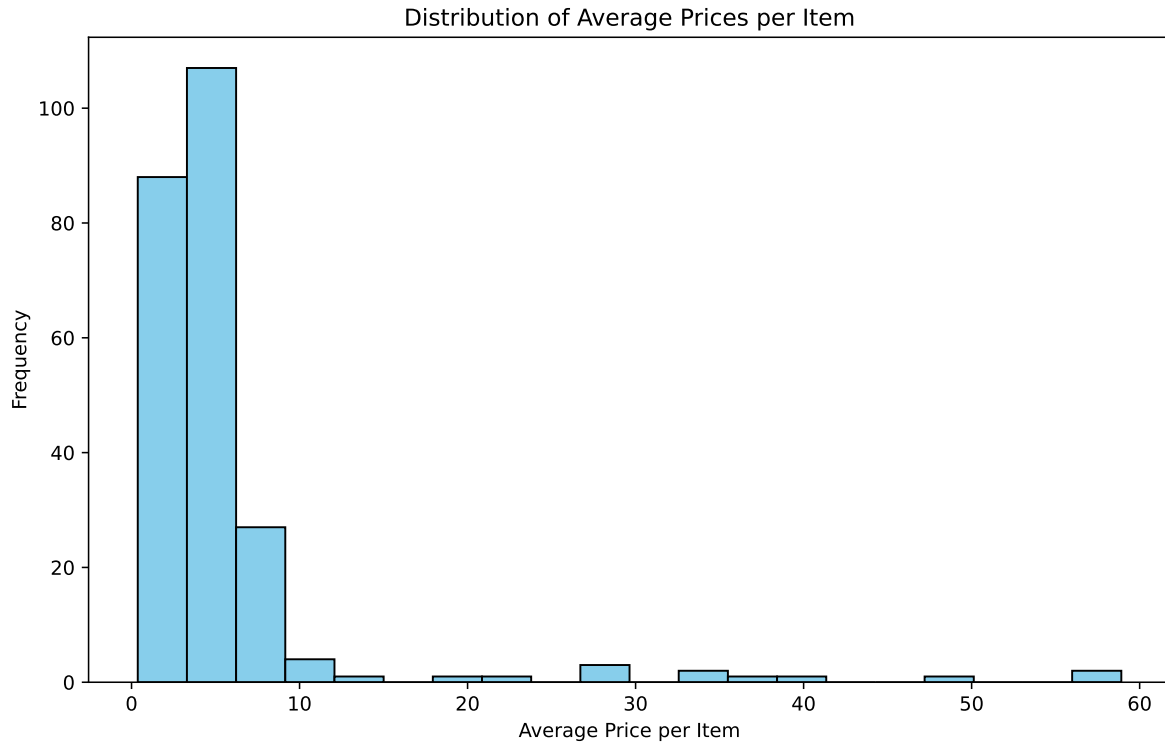
Vendor Count DataFrame:

	Vendor	Count
	Galleria	6104
0	Loblaws	18361
1	Metro	18895
2	NoFrills	15404
3	SaveOnFoods	10772
4	TandT	10148

(Rossum and Foundation 2024)

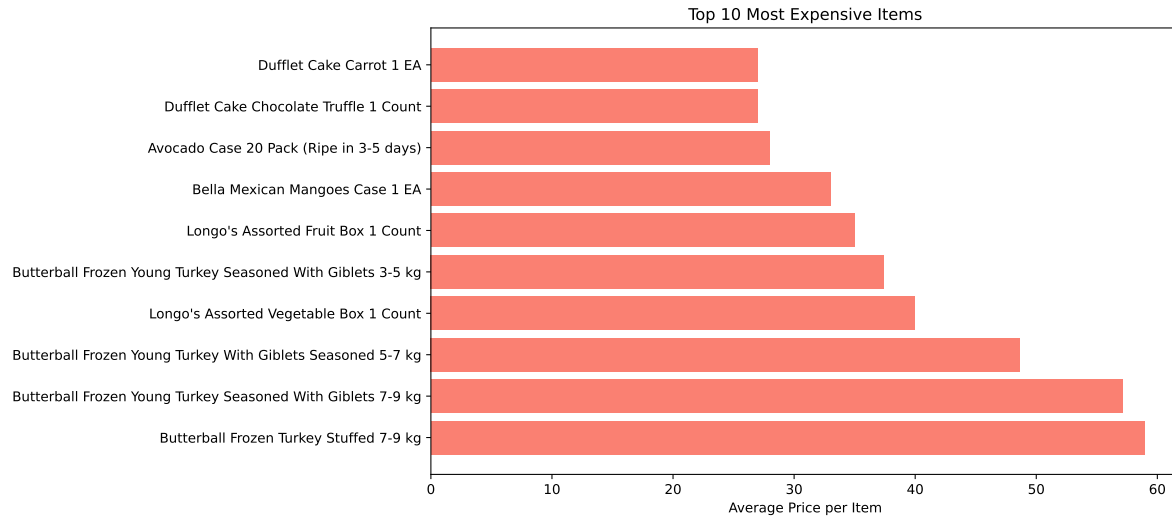
## 4.1 Distribution of Price per Item

This subsection we would analyze prices based on unit types provides insight into pricing strategies.



(Rossum and Foundation 2024)

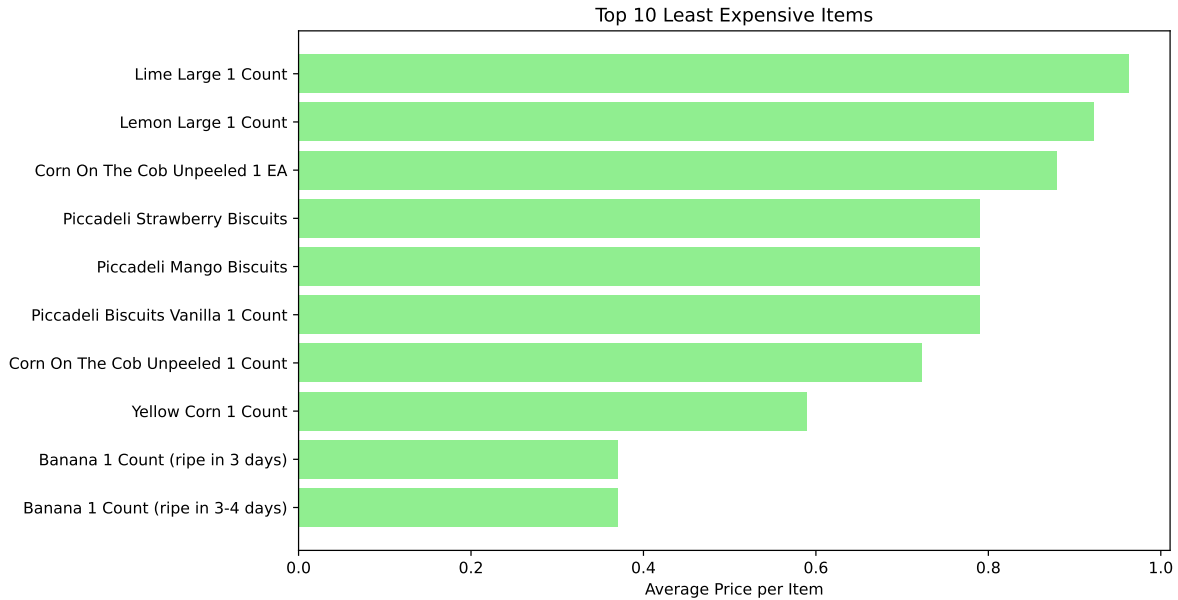
The histogram above shows how the average price per item of all products can be modeled. It follows that most of the items are below the 10 dollars price, especially in a range from 0 to 5 dollars, which underlines that the majority of products are quite cheap. Only a few products are categorized in prices beyond 20\$ and reach up to 60\$. This is a skewed distribution, with high prices representing the outliers and a general pricing strategy based on low, accessible prices for most of their products. A histogram of this kind puts in focus the affordability of the products in this dataset, while just a few premium items drive the top of the price spectrum.



(Rossum and Foundation 2024)

**Top 10 Things That Are The Most Expensive** The following bar chart represents the top 10 items according to their average price per item. This indicates that high-end items like Butterball Frozen Turkeys, Longo's Assorted Vegetable and Fruit Boxes, and Avocado cases are expensive items in the range of close to \$20 to over \$50.

It would appear from the chart that specialty or seasonal items—larger fruit ensembles or quality turkeys—are much more highly priced. These products satisfy particular markets that have specific needs or large quantities and thus aim at buyers ready to pay more for convenience or high quality.



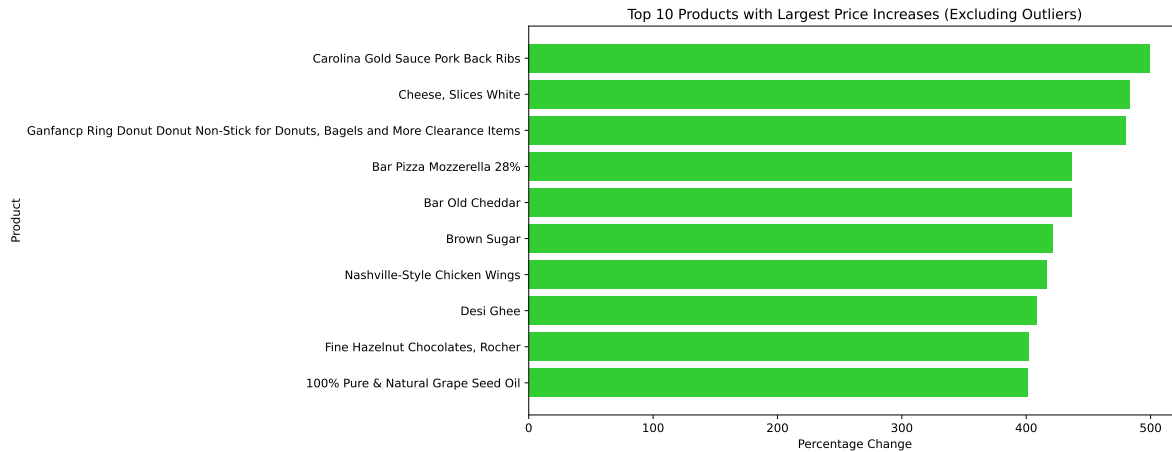
(Rossum and Foundation 2024)

**Top 10 Least Expensive Items** This is further emphasized by the following bar chart on the top 10 least expensive items, including products with a low average price per item: for example, meat and most fresh produce items, bananas at \$0.28, lemons at \$0.25, limes at \$0.24, and corn at \$0.25. Most of these items were less than \$1.00 per unit, which is normally a typical price for everyday items or staple products. Furthermore, there was a lot of Piccadeli biscuits that may give one an inclination toward a snack product that is low-priced within the product line. This chart shows that basic fruits and vegetables are kept at a low price to meet the consumer's expectation of affordability for such staples. Emphasizing more affordable basics makes grocery staples more accessible, especially for items that have more frequent purchases.

## 4.2 Price Changes Over Time

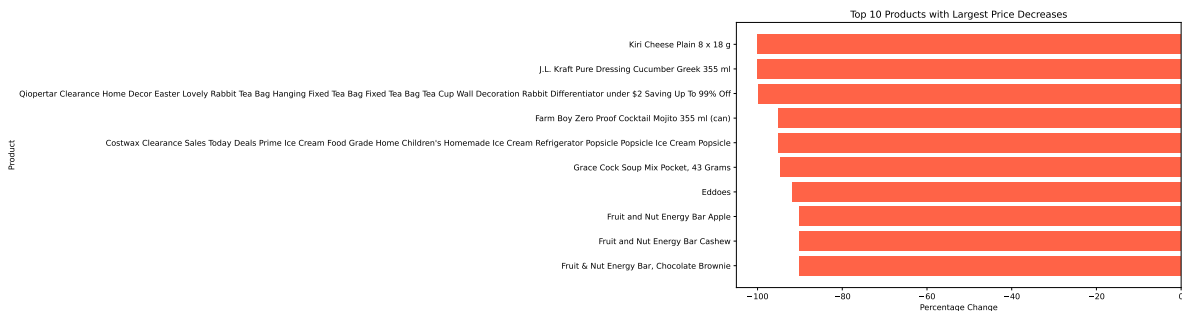
The following bar charts show the top 10 products with the highest percentage increases and decreases in price. These findings highlight significant price volatility among certain products, which could impact consumer spending.





(Rossum and Foundation 2024)

Above is a bar chart of the top 10 products with the largest price increases, after the cleaning process to remove extreme outliers. Products such as “Carolina Gold Sauce Pork Back Ribs” and “Cheese, Slices White” reveal very high percentage increases, with values close to the upper 500% range. It would thus seem that part of the increases in grocery staple items have been relatively large and may be seasonal in nature, tied to supply chain disruption, or inflationary pressures in manufacturing. This chart reflects a relatively high level of price volatility in an array of items, the majority of which pertain to specialty or processed foods that could affect what the consumer purchases.



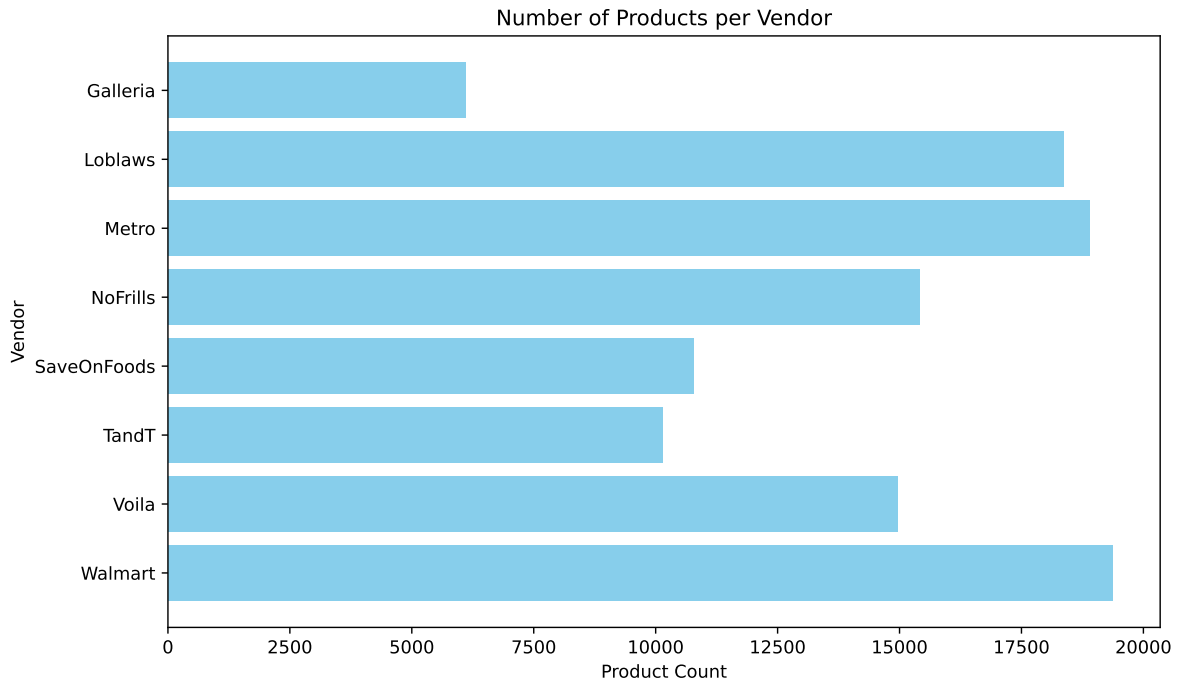
(Rossum and Foundation 2024)

This bar chart shows the top 10 products of the biggest price reductions. Similarly, a almost than 100% drop in items like “Kiri Cheese Plain” and “J.L. Kraft Pure Dressing Cucumber Greek” indicates that these items are being significantly marked down or cleared out. The other listed items, which include energy bars and soup mixes, reveal quite a significant plunge in their values. These huge markdowns can come from having too much stock in store, upcoming

expiration dates, or promotions to clear the items from their inventories. The chart indicates aggressive price cuts to trim inventories on select items, adding to shopper bargains.

### 4.3 Vendor Product Variety

The bar chart below compares the number of products available from each vendor, showcasing their relative variety in product offerings.



(Rossum and Foundation 2024)

The bar chart provides an overview of the product variety by each vendor, thus showing the difference in the size of inventories of major Canadian grocery retailers. Walmart, Metro, and Loblaws have the top product count, offering upwards of a vast 18,000 item selection each, thus underlying their leading positions as complete, one-stop shopping destinations. This wide range attracts a broad clientele that can be served with different needs in a number of product lines. In comparison, other vendors like Galleria and SaveOnFoods maintain a relatively small product portfolio, presumably focusing their attentions on narrower markets or more focused presentations that might reflect specific regional or cultural interests. These differences in product variety reflect diverse business strategies, where larger vendors compete on the basis of broad assortments and convenience, while smaller vendors may attract customers through specialization or unique product offerings. Therefore, both variety and focus will be potent positions in the competitive game of the grocery marketplace.

## 5 Discussion

### 5.1 Correlation v.s. Causation

When comparing changes in prices among all grocery products, one must be very careful to separate mere correlations from actual causations. While data shows a trend-like scenario, for instance, increased prices in certain categories of products, these findings are no more than observations of correlations, not confirmed causes. A highly increased price in certain items may correlate to seasonal demand shifts, inflation pressures, or supply chain disruptions; however, that would not confirm these as direct causes. This would require controlled experiments or more contextual data that regularly allows one to eliminate the influence of everything but the influence of interest on pricing. We observe that observational data, like ours, is by its nature limited to capturing associations and does not provide the experimental controls needed to establish causal relations. For this reason, though very important insights are gained from the analysis into the patterns of pricing, the conclusions should not be regarded as evidence of mechanisms of causation. Additional context could enhance understanding, possibly involving controlled factors or other variables, which may explain drivers behind the line in observed pricing trends.

### 5.2 Missing Data

The problem of missing values is inherent in any observation dataset; this problem is further aggravated when the source of such data is dynamic, as will be expected with online grocery prices. Gaps in this dataset can take several forms, including intermittent collections, unavailable products, and technical problems in data scraping. These missing values, while removed through SQL, may affect analytical accuracy, particularly in those calculations that are sensitive to time, such as the computation of trends of prices over time. These excluded non-numeric and incomplete values in critical fields to make sure data integrity is maintained in each record. However, such analysis may contain some sort of bias since, during the analysis of complete cases, products or vendors may have systematic missing values taken out. This may underrepresent items for which availability is temporary or those that are regionally exclusive. Future research may explore data imputation or sampling adjustments for more comprehensive insights into the same. Approaches such as data imputation or alternative sampling might give a full and accurate representation of the pricing behavior across products.

### 5.3 Sources of Bias

Several sources of bias may affect the interpretations and conclusions that could be drawn with this dataset. First, selection bias is an issue because this data set could include only products available online for certain vendors and may not reflect the entirety of product sets offered in physical stores. This could well limit the data so that budget or region-specific items which are

sold primarily in-store are underrepresented. Second, temporal bias is likely to occur: grocery prices are subject to fluctuation in response to sales and promotion activities and seasonal demand; prices analyzed at a point, or over a short period, may miss broader pricing trends or cyclical patterns. The final assemblage of risks is associated with confirmation bias, whereby an exponential feeling of volatility might be seen when the analysis of products with the most volatile prices is centered on, such that any stable or products showing a minor increase in price might go unnoticed. These biases can be prevented in further research by expanding and balancing the dataset to include online and in-store prices over longer time spans. The result would be a far more detailed grocery pricing pattern, which would reduce the influence of biases inherent in selective data samplings and short-term analyses.

## References

- Filipp, Jacob. 2024. “Project Hammer: Canadian Grocery Price Data.” Project Hammer. <https://jacobfilipp.com/hammer/>.
- Information Technology — Database Languages — SQL*. 2023. International Organization for Standardization. <https://www.iso.org/standard/76583.html>.
- Rossum, Guido van, and The Python Software Foundation. 2024. *Python Documentation*. <https://docs.python.org/3/>.