

Grocery stores in Canada*

What vendors hold the most unique items?

Hyunje Park, Charlie Zhang

November 21, 2024

This paper analyzes which Canadian grocery vendors offer the widest range of unique products using the Canadian grocery data from Project Hammer. It was discovered that Loblaws, Metro, and Walmart carry around 20,000 items each, while region-specialized vendors like Galleria and T&T offer fewer than 10,000.

Table of contents

1	Introduction	1
2	Data	2
3	Results	2
4	Discussion	3
4.1	Correlation vs Causation	3
4.2	Missing Data	3
4.3	Source of bias	4
5	Conclusion	4
	References	5

1 Introduction

Canada's diverse population supports a market rich in both Canadian and international goods. This paper analyzes which grocery vendor offers the most unique products, focusing on product

*Code and data are available at: <https://github.com/davidpxrk/most-common-vendors>

IDs (which define items) and vendor names. Summing unique product IDs reveals the number of items each vendor carries, providing insights into market competitiveness.

The paper is structured as follows: Data (Section 2) introduces the dataset; Results (Section 3) presents key findings; Discussion (Section 4) explores observations and limitations; and Conclusion (Section 5) summarizes the analysis.

2 Data

This report analyzes the Canadian Grocery Price dataset, collected by Jacob Filipp and hosted by Project Hammer (Filipp 2024). Of the three tables in the dataset, only the product table was used, focusing on the `id` (product identifier) and `vendor` (name of the vendor) variables.

First, the original dataset was downloaded from Project Hammer (Filipp 2024). Using SQL (ISO/IEC Joint Technical Committee 1987), the `id` and `vendor` variables were selected, and a table was created that containing the number of unique product IDs (`id`) for each vendor. This table was then exported as a `.csv` file. The exported `.csv` file was then analyzed in R Programming Language (R Core Team 2023) to visualize the relationship between vendors and their unique product counts, using the `ggplot2` package (Wickham 2016).

3 Results

Figure 1 shows the unique product counts for each Canadian grocery vendors. Overall, Loblaws, Metro and Walmart reigned at the top with over 20,000+ products, while international-focused chains such as Galleria or T&T had the least amount of unique products.

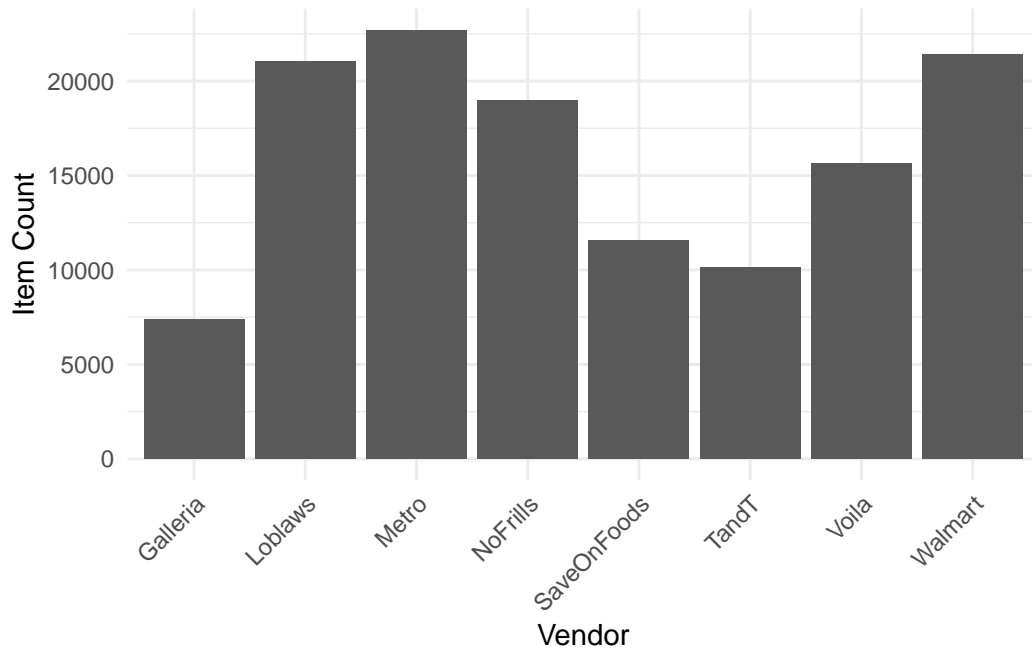


Figure 1: Number of Unique Products for each Grocery Vendor

4 Discussion

Figure 1 showed that big companies such as Metro and Walmart far passed Galleria and T&T in product variety, showing a competitive challenge for smaller vendors.

4.1 Correlation vs Causation

Larger vendors, like Walmart (the world's largest company by revenue) offer more unique items, while smaller vendors such as Galleria and T&T have a smaller revenue, which could explain the disparity in product variety.

4.2 Missing Data

The `upc` (universal product code) and `sku` (inventory identifier) variables were excluded from the analysis as they were missing for all products.

4.3 Source of bias

This raises the question of why vendors like Galleria and T&T lag in product variety. As Asian-fusion grocery chains, they rely on costly imports, which could limit their inventory. However, without data on product types, commenting on the efficiency and competition on these vendors is impossible.

5 Conclusion

In summary, this paper analyzed Canadian grocery vendors' product variety, finding Metro and Walmart lead with over 20,000 unique items, while chains like Galleria and T&T, focused on Asian products, have fewer than 10,000.

References

- Filipp, Jacob. 2024. *Project Hammer*. <https://jacobfilipp.com/hammer/>.
- ISO/IEC Joint Technical Committee. 1987. *SQL: Structured Query Language*. <https://www.iso.org/standard/63555.html>.
- 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>.