Grocery stores in Canada*

What vendors hold the most items?

Hyunje Park, Charlie Zhang

November 14, 2024

This paper analyzes the Canadian Grocery Price data from Canadian grocery vendors, which was gathered by Jacob Filipp and hosted publicly by Project Hammer. It uses information on items, and the name of vendors to determine how many vendors hold a unique grocery item. It was discovered that vendors such as Loblaws, Metro and Walmart held the most unique items at around 20,000 items, while international-focused vendors such as Galleria and T&T held the least amount of items below 10,000.

Table of contents

*Code and data are available at: https://github.com/davidpxrk/cost-of-living-toronto

1 Introduction

Canada is home to many international people around the world, where many people of different cultures, backgrounds, ethnicity come together to create a unity. With this in mind, it can be said that a lot different products are sold in Canada; both Canadian and international goods.

This paper aims to analyze what grocery vendor holds the most unique product. This analysis focuses on two key metrics; the id of the product (the id that defines the product), and the name of the grocery vendor. The summation of the unique ID of the products can be used to determine how many items each vendor holds, and can help vendors compete in the grocery markets.

The paper is organized as follows: Section 2 (Data) introduces the grocery dataset used in this paper. Section 3 (Results) highlights important insights that were derived from the data analysis. Section 4 (Discussion) discusses the observations from Section 3 and some of the limitations that come with this analysis. Finally, Section 5 (Conclusion) draws a summary of the key findings.

2 Data

This report uses the Canadian Grocery Price dataset, gathered by Jacob Filipp and hosted by Project Hammer. The dataset, contained 3 different tables, but the `product` table was specificially used for this analysis, as the other 2 tables were of no relevance.

The `product` table contained 9 variables; most notably `id` and `vendor`. These two variables were important in this analysis, which captured the information on the product id, and the vendor that item was sold at. Data manipulation was conducted through `SQL` (Technical Committee 1)/ SC 32 (Subcommittee 32) 2023), to select the variables `id`, `vendor`, and find the number of unique items each vendor sold. This information was exported as a `.csv` file, and the data analysis was conducted through `R Programming Language` (R Core Team 2023), `ggplot2` (Wickham 2016),.

3 Results

Figure 1 shows the number of unique items for each Canadian grocery vendor. 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.

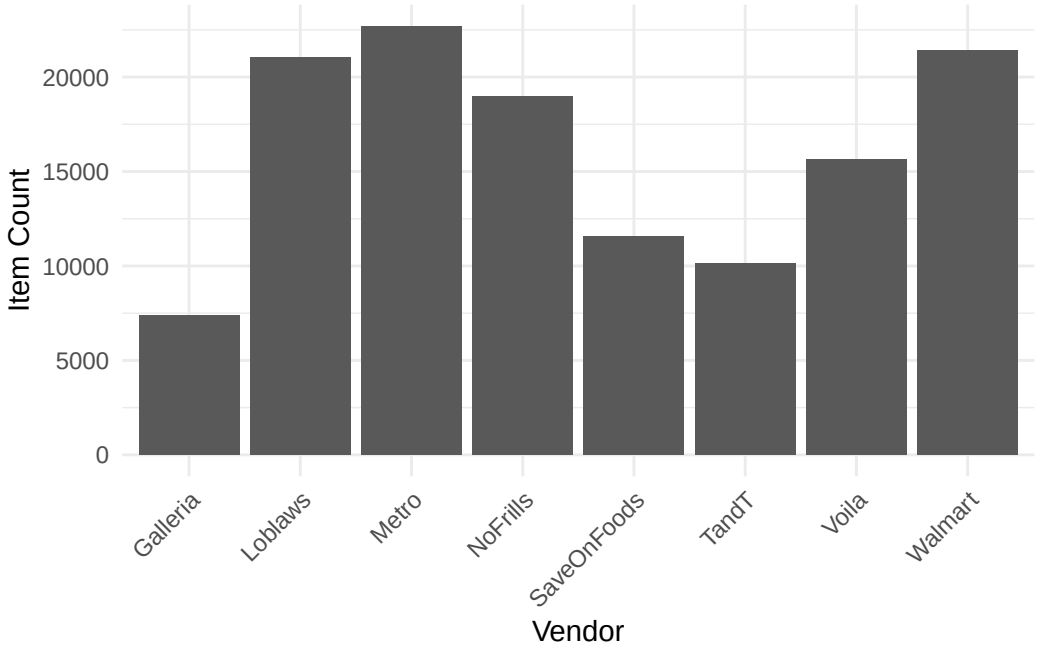

Figure 1: Number of Unique Products for each Grocery Vendor

4 Discussion

The graph above show that vendors such as Metro and Walmart far exceeded the number of products compared to vendors such as Galleria, or T&T, highlighting a problem of competition for these vendors.

4.1 Correlation vs Causation

Larger vendors, tend to have more unique items. Walmart, for example is the biggest chain in the United States, and is also the largest company in the world by revenue, whereas vendors such as Galleria or T&T are much larger in scale. Hence, this could explain the huge difference in item catalogue.

4.2 Missing Data

There were some information on products. For example, there was a variable for `upc`, a universally global code of a product, which was missing for every product, hence this attribute was left out of the analysis. Similarly, `sku`, which is an identifier used by retailers for inventory management was left out due to the same reason.

4.3 Source of bias

However, this raises the question; what disadvantages do these vendors have that cause them to lag behind on products? For example, Galleria and T&T are both Asian-fusion grocery chains that rely heavily on foreign products, which could make imports very expensive, potentially explaining the lack of items that have. Without information on the types of products, commenting on the efficiency and competition on these vendors is impossible.

5 Conclusion

In summary, this paper investigated what grocery vendors in Canada held the most unique products. It showed that Metro and Walmart reigned on top, with over 20,000+ unique products, and Asian-product focused chains such as Galleria or T&T had the least amount of unique products, below 10,000.

6 Statement on LLMs

LLMs and other generative AI tools were not used in the making of this paper.

References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Technical Committee 1)/ SC 32 (Subcommittee 32), ISO/IEC JTC 1 (Joint. 2023. *Information Technology - Database Languages SQL*. https://www.iso.org/standard/76583.html.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.