Exploring the Relationship Between Product Variety and Package Size Across Vendors: A Correlational Analysis Revealing a Positive Association*

Claire Ma

November 14, 2024

This study investigates the relationship between unique product count and average package size across various retail vendors. Using a correlational analysis, we aimed to determine whether vendors with a broader product selection tend to offer larger average package sizes. Our findings reveal a positive correlation between product count and package size, indicating that vendors with higher product diversity, such as Walmart, generally provide larger packages, potentially reflecting a focus on bulk items or a diverse inventory strategy. However, this correlation does not imply causation, as additional factors, including product types and target demographics, may influence this relationship. The analysis also discusses limitations such as missing data and potential sources of bias. This preliminary investigation provides insights into retail packaging and inventory strategies, highlighting the need for further research to better understand the drivers behind these trends.

1 Introduction

Understanding the relationship between a vendor's product diversity and packaging practices can offer valuable insights into consumer preferences, inventory strategies, and market positioning. In retail, the number of unique products a vendor offers may reflect its target demographic, customer demands, and its approach to product variety. Additionally, average package size can influence purchasing decisions, shipping costs, and environmental impact. Examining how

*A GitHub Repository containing all data, R code, and other files used in this investigation is located here: https://github.com/ClaireMa0311/Canadian-Grocery.git

these factors correlate provides an opportunity to explore potential patterns across different vendors.

This analysis aims to investigate the relationship between the number of unique products offered by various vendors and their average package size. By comparing vendors based on these two metrics, we can observe whether vendors with a higher product count tend to offer larger or smaller average package sizes. However, while correlation may highlight patterns, establishing causation requires caution. Multiple factors, such as product type, vendor size, and consumer needs, could confound this relationship.

This study also addresses potential challenges in the analysis, including missing data, sources of bias, and limitations in interpreting correlation versus causation. These considerations are critical for ensuring that any observed trends are not only accurate but also meaningful in the context of the retail landscape. Through this exploration, we aim to uncover preliminary insights into the interplay between product variety and packaging size, providing a foundation for further, more detailed investigations.

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2, Section 2.2, Section 3.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023). Some of the R code used are from Alexander (2023). Our raw data (Admin 2024) a 2024 Canadian Grocery Price Data, including 8 vendors in the dataset: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart and Save-On-Foods.

```
Call:
lm(formula = avg_package_size_kg ~ unique_product_count, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.17898 -0.03206  0.02924  0.06459  0.07993

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.081e-02  5.262e-02   0.396   0.7061
unique_product_count 2.296e-05  9.376e-06   2.449   0.0499 *
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09478 on 6 degrees of freedom
Multiple R-squared:  0.4998,    Adjusted R-squared:  0.4165
F-statistic: 5.996 on 1 and 6 DF,  p-value: 0.04988
```

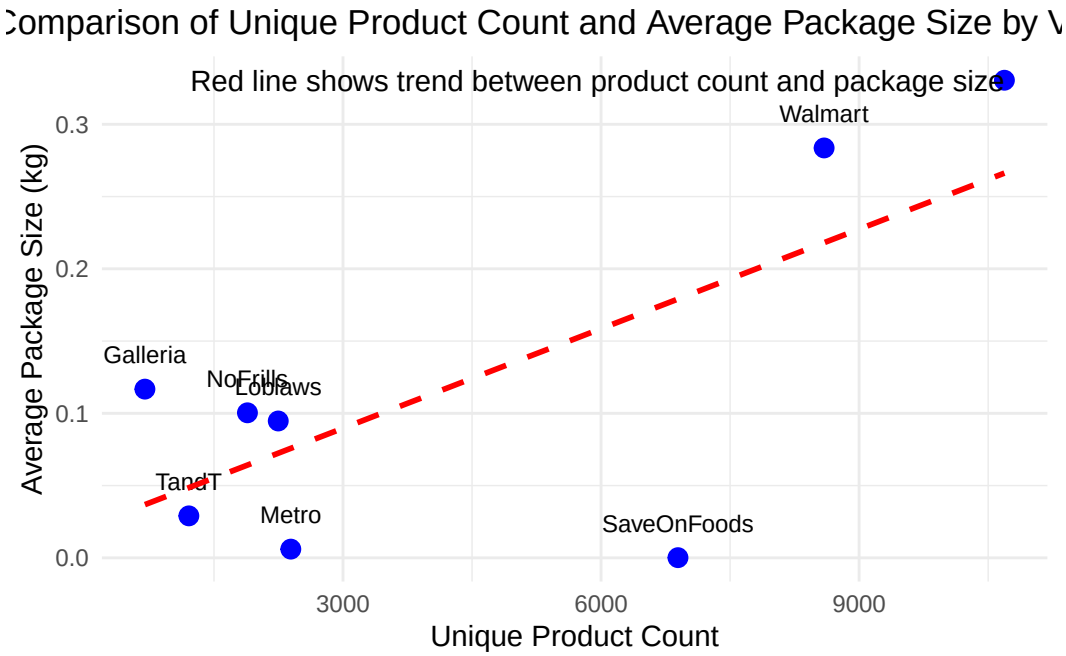

Figure 1: Correlation Between Product Count and Average Package Size

2.2 Results

The analysis reveals a positive correlation between the number of unique products offered by vendors and their average package size. As shown in Figure 1, vendors with a higher product count, such as Walmart, tend to have larger average package sizes compared to those with fewer unique products, like Galleria and NoFrills. The trendline in Figure 1 highlights this relationship, suggesting that as the variety of products increases, vendors may lean toward larger average package sizes. This pattern could indicate that vendors with a wide range of products are more likely to offer bulk items or larger packaging options, possibly to meet the needs of a diverse customer base or to support a more varied inventory strategy.

However, while the correlation is evident, it does not imply causation. Other factors not represented in this dataset, such as the types of products offered (e.g., bulk versus single-serve items), vendor size, or customer demographics, may also influence the observed relationship. Additionally, there is variation among vendors, with some vendors showing lower average

package sizes despite a relatively high product count, as seen with SaveOnFoods. This variation suggests that while there is a general trend, individual vendor strategies and product mixes can impact packaging decisions.

These findings offer a preliminary understanding of how product variety and package size relate, providing insights into inventory strategies across different vendors. Further research with additional variables could help clarify the underlying factors contributing to this positive correlation.

3 Discussion

1) Correlation vs. Causation

Figure 1 shows a positive correlation between the number of unique products offered by each vendor and the average package size in kilograms. The red dashed trendline suggests that vendors with higher product counts tend to offer larger average package sizes. However, this correlation does not imply causation. We cannot conclude from this data alone that increasing the number of products directly causes an increase in average package size. This trend may instead be influenced by other underlying factors, such as the type of products each vendor prioritizes (e.g., bulk items versus smaller packaged goods) or differences in customer demographics and preferences. Without experimental or longitudinal data, it is difficult to establish a causal relationship between product count and package size.

2) Missing Data

One limitation in interpreting this analysis is the potential impact of missing data. Figure 1 only includes a subset of vendors, and we do not have detailed information on the product categories or the actual distribution of package sizes within each vendor's inventory. If some vendors or product categories are underrepresented, this could skew the observed trend. For example, vendors with a focus on bulk products may naturally have larger average package sizes, which could distort the overall correlation when compared to vendors with a focus on smaller, individual items. Additional data, such as the breakdown of product categories and more vendors, would provide a more comprehensive view.

3) Sources of Bias

Several sources of bias may affect the analysis. First, selection bias may be present if the vendors included in the dataset are not representative of the overall market. If only larger vendors or specific types of stores (e.g., discount stores or high-end grocers) are represented, the results may not generalize to all types of vendors. Additionally, measurement bias could occur if package size is calculated differently across vendors, potentially leading to inconsistencies. Finally, confounding variables—such as regional factors, vendor target demographics, or marketing strategies—may influence both the product count and the average package size

independently. Without accounting for these confounding factors, the observed correlation may not accurately reflect a true relationship between product count and package size.

References

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

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

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