Vendor Price Correlation Analysis

Jacob Gilbert Aman Rana

2024-11-14

Abstract

This study analyzes the correlations in product pricing between eight major vendors using data from Project Hammer, with a focus on correlation versus causation, missing data, and potential biases. The findings aim to better understand competitive behaviors and pricing strategies.

Table of contents

Data Preparation 1
Results 2
Correlation v Causation 2
Missing data 2
Sources of Bias 2

The aim of this study is to analyze the correlations in product pricing between eight major vendors: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods, using data from Project Hammer ("https://jacobfilipp.com/hammer/"). By exploring the relationship btwn the pricing of, we hope to uncover patterns that might help explain competitive behaviors, pricing strategies, and the impact of broader economic factors on retail prices.

This paper specifically focuses on assessing the correlations between the prices of dairy products at each of the listed vendors, attempting to discern whether a relationship exists and how strong that relationship is. While correlation can often hint at similar trends or suggest price competition, it is important to exercise caution when drawing conclusions about the underlying causes. For this reason, we dedicate a sub-section to the distinction between correlation and causation, highlighting the dangers of misinterpretation.

Ultimately, this paper aims to provide a comprehensive analysis of price relationships across these eight vendors. "result"

This analysis uses SQlite (Team (2024)) for data manipulation and Python (Zelle (2024)) to display the regression.

Data Preparation

We begin by creating a csv for our downstream parsing. Observing the database we find 1L of Heinz ketchup has ids 716683 and 1704661. We filter the database of products to just ketchup, drop rows which have a SALE price, and then save the table to a csv file.

An example of the data is shown below:

Table 1

vendor date	Loblaws	Voila
2024-09-23	5.99	6.49

vendor date	Loblaws	Voila
2024-09-24	5.99	6.49
2024-09-25	5.99	6.49
2024-09-27	5.99	6.49
2024-09-28	5.99	6.49

Results

`np.float64(0.09167290674014908)`

The correlation coefficent when looking at the price of 1L of Heinz Ketchup between Loblaws and Voila is 0.09, this is a weak positive correlation. We ccan interpret this as the prices of goods move together, but weakly.

Correlation v Causation

Correlation is a measure of the relationship between two variables, such as milk prices across different vendors. While high correlation can indicate that prices move in similar ways, it does not imply that one vendor's pricing directly causes changes in another's. Factors like regional supply chain disruptions, seasonal demand, or broader market trends could all lead to similar pricing patterns without any direct causal link between vendors.

To understand causation, one must look beyond mere correlation and consider controlled experiments or advanced econometric models that account for external influences. In this analysis, while we identify price similarities, attributing them to causation would be misleading without deeper scrutiny. One method of looking at causation would be diffrence in difference analysis to indentify what effect a treatment (such as competitors pricing increasing) may have. Thus, it is essential to recognize that correlation highlights relationships but does not inherently explain the reasons behind them.

Missing data

In the context of this study, one significant challenge is the absence of data from other potential sources of the same products, such as smaller local vendors or specialized online retailers. These sources can have a substantial impact on overall market dynamics, yet they are not represented in the current dataset, potentially leading to incomplete or biased insights. Another important consideration is the absence of competitor pricing data. Competitors not included in the dataset could influence the pricing strategies of the vendors analyzed. The exclusion of these other sources means that our analysis may overemphasize the influence of the major vendors included in the dataset. Prices at smaller vendors may differ significantly due to factors such as lower overhead costs, niche market positioning, or different supply chains. The absence of this data could lead to a skewed understanding of price correlations, making it seem as though the major vendors are more interconnected than they might be in a broader context. To address these gaps, future research should consider incorporating a wider variety of data sources to provide a more comprehensive view of the market and improve the robustness of the conclusions drawn.

Sources of Bias

Main baises are likely to arise from the missing data mentioned previously. By focusing solely on major vendors like Voila, T&T, Loblaws, and others, we may inadvertently ignore market forces that influence smaller competitors, leading to an incomplete picture of the overall market dynamics. This selection bias can create a distorted view of price competition and consumer behavior. The unawarness of relationships with competitors allows more selection bias also. One source of bias could be from unobserved influences on price and sales. for example advrtising campaigns of discounts that influence prices in partcular stores. his

could lead to observing a downturn n prices in one stor that i not fitting with the other stores. This could giv our regression unreliable results.

Team, SQLite Development. 2024. "SQLite Documentation." https://www.sqlite.org/docs.html.

Zelle, John M. 2024. *Python Programming: An Introduction to Computer Science*. Franklin, Beedle & Associates Inc.