

My title*

My subtitle if needed

First author

Another author

November 14, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

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

2 Data

2.1 Overview

We use the statistical programming language SQL (R Core Team 2023)

.... Our data (**shelter?**).... Following (**tellingstories?**), we consider...

Overview text

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

2.3 Outcome variables

2.3.1 Price Levels

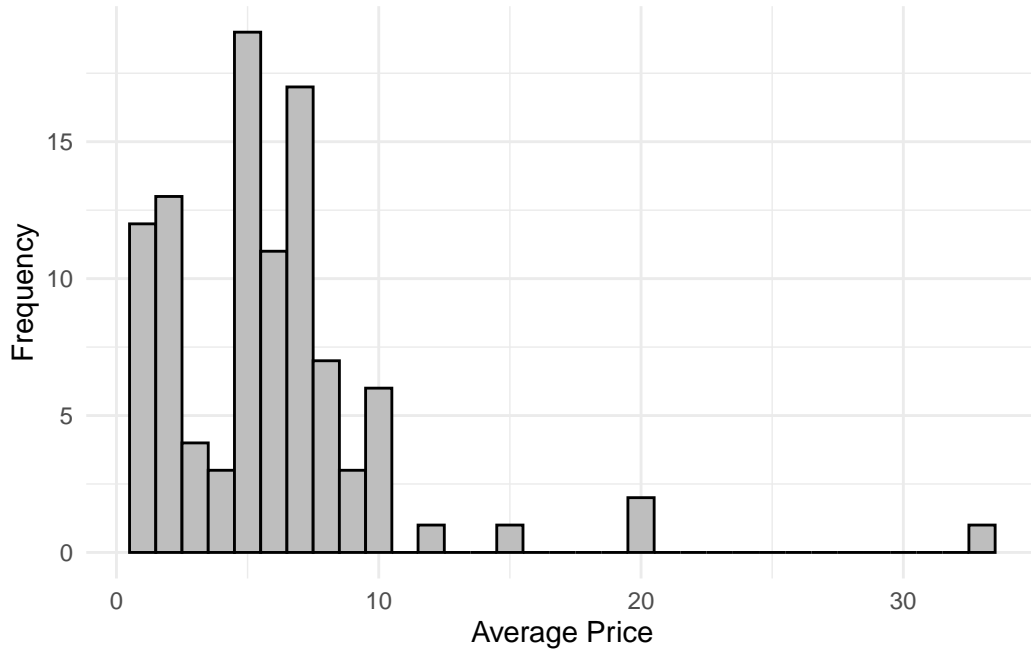


Figure 1: Price level distributions across all observations used in the analysis

2.4 Predictor variables

2.4.1 Type of Fruits: The Product that We Specified to Compare the Price Distribution

Figure 2 shows the frequency of observations for four fruit types: apples, mangoes, oranges, and pears. Apples have the highest count, with around 50 instances, possibly due to popularity or market demand. Pears follow with a moderate count of about 20. Mangoes and oranges have the lowest counts, around 10 each, possibly due to factors like seasonality or regional availability.

2.4.2 Type of Vendors

Figure 3 provides a visualization of the number of observations for each vendor in the dataset, specifically Loblaw's, NoFrills, and Voila. The x-axis represents these three vendors, while the y-axis shows the count of observations, likely indicating the number of unique products or

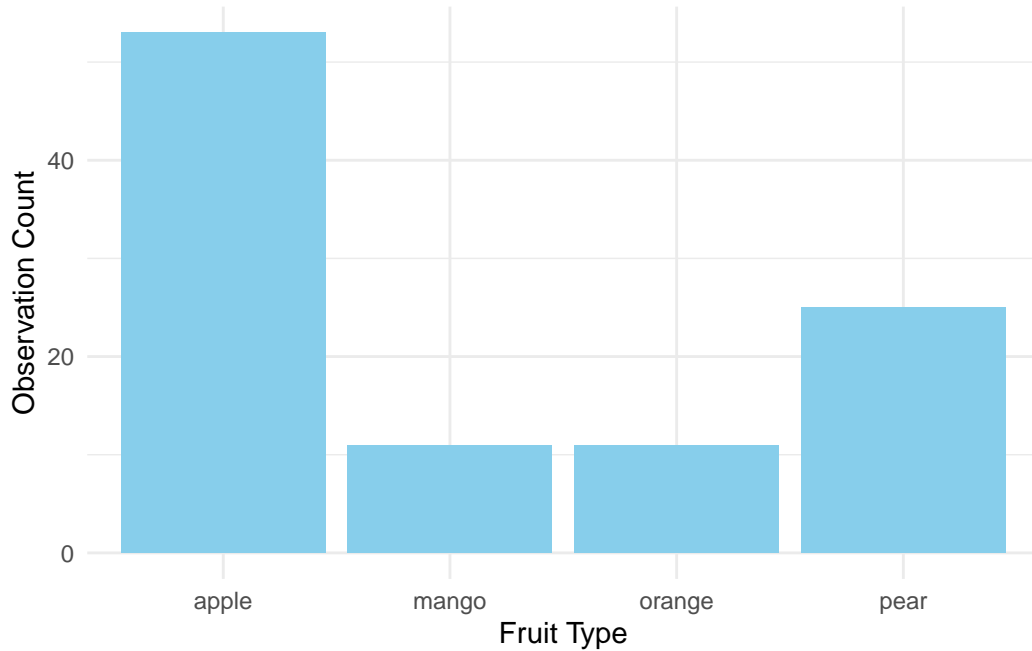


Figure 2: The distrubution of different categories of fruits used in the analysis

price points recorded for each vendor. Each bar is labeled with the exact count: Loblaws has 33 observations, NoFrills has 31, and Voila has the highest at 36. These counts are relatively similar across vendors, suggesting a fairly balanced representation in the data, although Voila has a slight edge in the number of entries. This balanced distribution is beneficial for comparative analysis, as it reduces the risk of one vendor disproportionately influencing results. The slight difference in counts could introduce a minor weight toward Voila in any statistical analysis, but overall, the dataset’s even distribution supports reliable comparisons. Thus, any averages or other metrics calculated by vendor are likely to be comparable across Loblaws, NoFrills, and Voila, minimizing selection bias due to unequal observation counts.

3 Results

Figure 4 illustrates the average prices of four fruits—apples, mangoes, oranges, and pears—across three different vendors: Loblaws, NoFrills, and Voila. Each fruit is represented by a distinct color, as indicated in the legend: red for apples, green for mangoes, blue for oranges, and purple for pears. The bar plot component of the chart displays the average price of each fruit at the respective vendors. Notably, apples are priced highest at Loblaws and lowest at Voila. Mangoes exhibit minimal price variation across the vendors. Oranges show a significant price difference, being most expensive at Loblaws and least expensive at Voila. Similarly,

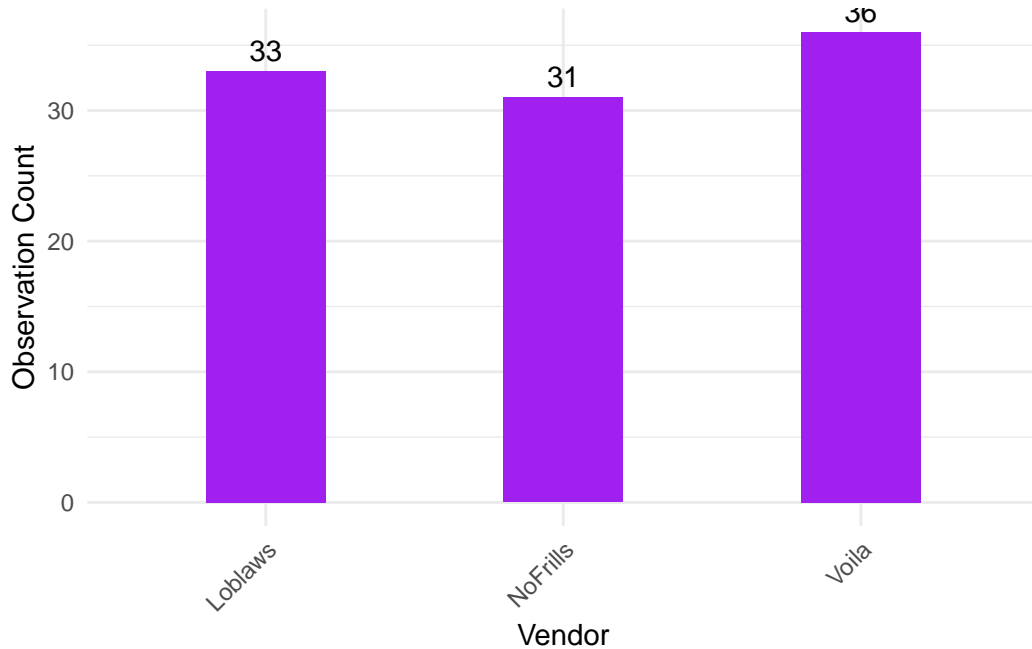


Figure 3: The distribution of vendors used in the analysis

pears are priced lowest at Voila. Superimposed on the bar plot is a line plot that connects the average prices of each fruit across the vendors, highlighting the price trends. The lines for apples, oranges, and pears demonstrate a downward trend from Loblaws to Voila, indicating decreasing prices. In contrast, the line for mangoes remains relatively flat, suggesting consistent pricing among the vendors. In summary, the chart effectively compares the pricing strategies of different vendors for various fruits, revealing that Voila generally offers lower prices for most fruits, while Loblaws tends to have higher prices. These visual trends provide insights into the pricing correlations among the vendors and set the stage for further analysis of potential causal factors influencing these pricing differences.

4 Discussion

4.1 Correlation vs. Causation

When analyzing the pricing of the same fruit in different supermarkets, it is very important to distinguish Correlation and Causation for understanding the root cause of price difference. We have observed that the prices of Loblaws apples, pears, orange and mango are always higher than those of other supermarkets (such as NoFrills and Voila), and the price of each fruit in Voila is always lower than that in other stores, which shows that there is a correlation between

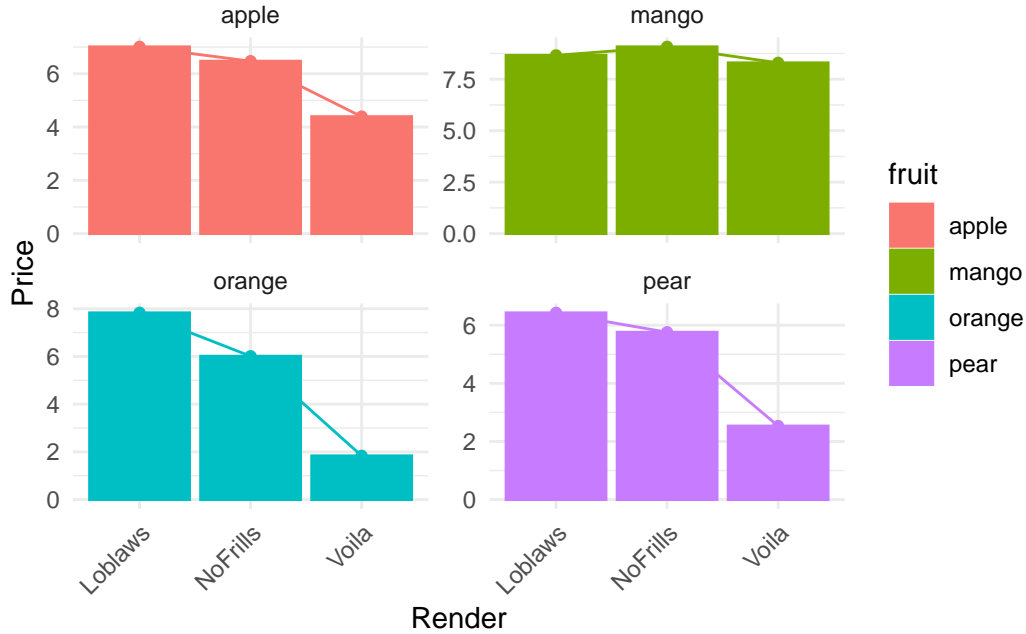


Figure 4: Price distribution of different kinds of fruits(apple, mango, orange, pear) by vendors

supermarket and fruit pricing. This correlation may be related to the market positioning, operating costs and supply chain management of supermarkets. For example, Loblaws is positioned in the high-end market and can provide better products and services to middle-class customers, so its pricing is higher. Voila mainly serves groups with limited budgets, so the price is lower. In addition, factors such as geographical location, rent and employee salary may lead to differences in operating costs, which in turn may affect pricing strategies. For example, Loblaws is located in the city center, the rent is high, and the salary of employees is relatively high because of the high service level. In supply chain management, for example, the procurement channels of different supermarkets and the efficiency of supply chain may lead to cost differences and affect the final selling price. In addition, to determine whether these factors really lead to price differences, further data and analysis are needed, such as market positioning research, cost structure analysis and supply chain evaluation. Through these in-depth analysis, we can understand the causal relationship of price difference more accurately.

4.2 Missing Data

In this dataset, missing data occurs because different supermarkets offer varying selections of fruit varieties. For example, some supermarkets may carry a wide range of apple varieties, while others only stock a few. As a result, when analyzing data across all supermarkets,

information on certain apple varieties may be incomplete, with specific types available only from certain vendors. Similar patterns may apply to other types of fruit, leading to gaps in the dataset. This missing data reflects differences in product offerings across vendors, and it's essential to be aware of and consider this incompleteness when conducting further analysis.

4.3 Sources of Bias

In this dataset, we study the price differences and potential changes of 4 fruits between different supermarkets. The price differences are due to the different positioning of each supermarket, Loblaw, NoFrills, and Voila.

4.3.1 Selection Bias

As mentioned before, selection bias in this dataset mainly comes from the differences in customer groups of different supermarkets. Each supermarket (such as Loblaws, NoFrills, and Voila) has a different market positioning and customer group. Customers who choose to shop at Loblaws usually prefer high-end goods or organic foods, while NoFrills customers may pay more attention to low prices and discounts. This customer preference leads to bias in the price data, that is, the price differences between different supermarkets not only reflect the market price of the product itself, but also are affected by customer selection behavior. Therefore, Loblaws' higher prices do not necessarily mean that all its fruits are priced higher than other supermarkets, but rather reflect that its customers are willing to pay a premium for higher quality or brand.

4.3.2 Measurement Bias

In this dataset, measurement bias may come from the different ways in which supermarkets record or define the average price (`avg_price`). For example, Loblaws may be promoting a specific brand or organic fruit, making the recorded average price lower than the actual price, while NoFrills may use a specific price for a single batch, ignoring other higher price ranges. These differences in measurement may cause distortions in price data, making our price comparisons between different supermarkets biased.

4.3.3 Confounding Bias

In this dataset, confounding bias may come from factors such as fruit type, quality, packaging type, seasonal changes, etc. For example, Loblaws may be more inclined to sell high-quality or organic apples, which are usually priced higher, and this tendency is not shown separately in the data. Therefore, the price difference between supermarkets may not be due to the supermarket itself, but to uncontrolled fruit quality or seasonal factors.

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/>.