Analyzing Food Prices in Canada: Trends, Regional Disparities, and Key Insights

Daksh Patel, Michaela Foo
CMPT 353 Fall 2024

1 Introduction

This study explores the **factors driving the rapid increase in food prices** across Canada between the years 2017 and 2024. It also aims to identify regions and food products that have experienced significant price changes compared to national trends, providing a deeper understanding of the underlying dynamics that shape these variations.

1.1 Inital Concept and Refinement

The original plan was to study how the cost of making a sandwich has changed over the past 10 years. However, limited data availability and the complexity of the topic led us to expand our focus. Instead, we analyzed the average price changes of staple food items across Canada and its provinces. This broader approach allowed us to explore bigger questions, like identifying times and places with unusual price increases and examining factors that might have influenced these changes, such as COVID-19 lockdowns, shifts in U.S. policies, and demographic changes like population growth and income levels.

1.2 Stakeholders and Relevance

Considering the perspectives of various stakeholder groups—**government, non-profit, and industry**—has been essential in shaping the questions driving this analysis. *Government agencies* can use insights into regional price increases and their impact on households to develop targeted policies and broader economic strategies. *Non-profit organizations*, such as food banks and social aid groups, can leverage the findings to advocate for improved food accessibility and better allocate resources by focusing on the causes of price spikes and the affordability of healthy foods across regions. *Industry stakeholders*, including retailers and grocers, can use the trends to refine pricing strategies, manage inventory, and adapt to seasonal fluctuations and shifting market demands. By aligning our analysis with the priorities of these groups, we ensured that the findings are both meaningful and actionable for a diverse audience.

2 Results and Findings

2.1 Data Source

The data for this study was sourced from Statistics Canada, a trusted and comprehensive provider of national and provincial statistics. For our analysis, we used a single CSV file containing monthly average prices for staple food products spanning January 2017 to August 2024. This data set covers all Canadian provinces (excluding territories) and includes the national averages calculated between provinces.

2.2 Data Structure and Attributes

The raw dataset consisted of 111,229 rows, providing a detailed breakdown of the monthly price data for each food item observed in the provinces. The key columns included the date (`REF_DATE`), the province (`GEO`), the product name (`Products`), and the corresponding price values (`VALUE`). Other columns in the raw dataset (e.g., `Unit of Measurement`) were redundant and removed during the cleaning process.

3 Data Preparation

3.1 Cleaning Processes

The data preparation process was streamlined due to the reliability and quality data of Statistics Canada. Initial cleaning focused on standardizing and filtering the data to enhance its usability for analysis. Dates were reformatted into a consistent and usable structure, and the dataset was filtered to focus on staple food items relevant to the study. To facilitate temporal analysis, aggregations were performed to calculate yearly and quarterly averages for prices and incomes. This cleaning process ensured the data was organized and prepared for detailed analysis while maintaining its integrity for accurate results.

3.2 Categorization and Enhancements

To add to the analytical value of the dataset, additional columns were added to segment the timeline into specific periods, such as Pre-COVID, During-COVID, and Post-COVID. Provinces were also grouped into broader regional categories —West Coast, East Coast, or Interior, and Urban or Rural Provinces—to provide a structured framework for geographic analysis. These classifications helped create a clearer context for examining both temporal and regional trends. Additionally, custom product groupings (meats, dairy, fruits, etc.) were for category-specific insights into food price variations.

To broaden the scope of the analysis, we merged supplementary datasets from Statistics Canada. These additional datasets linked food prices with population and income trends, providing a deeper understanding of how demographic and economic factors influenced price variations. Connections to global events, such as U.S. presidential terms (Pre-Trump, During-Trump, Post-Trump), were also examined to assess potential external influences. The cleaned and organized datasets were stored in structured directories to ensure accessibility and reproducibility while preserving the original raw data for reference. However, due to file size limitations, some larger datasets were excluded from the repository, though their details were documented for transparency.

4 Analysis Techniques

We structured our analysis around two primary categories—temporal factors and geographic/demographic factors—to provide distinct perspectives on the data and enable a comprehensive examination of trends and patterns.

4.1 Statistical Methods

Trend analysis was employed to identify long-term patterns and anomalies in food prices over the study period. This technique allowed us to highlight periods of irregular growth or decline, particularly during significant events like the COVID-19 pandemic and different U.S. presidential terms. Examining these trends provided valuable insights into the temporal dynamics of food price changes.

Analysis of Variance (ANOVA) was utilized to test for statistically significant differences in price changes across various time intervals and regions. It allowed us to evaluate whether price disparities between provinces or grouped geographic regions (e.g., East Coast, West Coast, and Interior) were due to real differences or random variation. **Tukey posthoc tests** followed ANOVA to pinpoint specific differences between groups, offering a more detailed understanding of how and where prices varied.

Correlation analysis was conducted to explore the relationships between food prices and demographic factors, such as income levels and population density. This method provided insights into how economic and social variables influenced price trends across Canada.

Non-parametric tests, including the Mann-Whitney U and Kruskal-Wallis tests, were applied to examine price differences across regions and urban versus rural areas. These methods can handle non-normal data distributions efficiently, ensuring accurate results even when traditional parametric assumptions are not met.

4.2 Visualization Approaches

Using *time-series* graphs we were able to track trends in food prices over time, identifying patterns, spikes, and anomalies across categories and regions. *Heatmaps* were used to highlight regional price differences, providing clear comparisons across provinces. Additional visualizations, such as line graphs and scatter plots, were generated to further explore relationships and trends in the data. All generated plots are available in our GitHub repository git@github.sfu.ca:dpa45/CMPT353-Project.git under the folder `output/geo_png`.

5 Results and Findings

5.1 Temporal

The temporal analysis focused on three key questions: whether food prices increased faster during the COVID-19 pandemic, during Donald Trump's first presidential term, and whether certain food items exhibited significant price increases during specific years or months. Given the global economic disruptions caused by COVID-19, the data was divided into three periods—before, during, and after the pandemic—using the CIHI Pandemic Timeline to define the "during COVID" period as January 2020 to May 2022. Similarly, for Trump's presidency, the data was divided into pre-, during-, and post-presidency periods spanning January 2017 to January 2021.

A three-way ANOVA test was applied to examine whether the mean rates of price changes differed significantly across these time periods. This method helped identify overarching patterns, while Tukey post-hoc tests provided granular insights into variations in month-over-month (MoM) and year-over-year (YoY) price changes for individual food items. For both COVID-19 and Trump's presidency, the results revealed no statistically significant differences in price changes across the defined periods, suggesting that external factors, rather than specific events or policies, influenced long-term price trends.

A broader trend analysis showed that food prices rose steadily by approximately 3–5% annually between 2018 and 2021, with notable exceptions. Meat and vegetables experienced a sharp ˜13% increase between 2018 and 2019, as illustrated in Figure 1. Additionally, all food categories saw a significant spike from 2021 to 2022, followed by a gradual decline over the next two years. By 2023–2024, many categories experienced an average price drop of ˜5%. Interestingly, while visual trends suggested more pronounced price increases for meat and vegetables, statistical tests, including t-tests and Tukey comparisons, consistently returned p-values above 0.05. This discrepancy likely reflects variability in category sizes, where extreme price changes in a few items may distort overall trends that statistical tests adjust for.

In conclusion,these findings demonstrated that while COVID-19 and Trump's presidency are often cited as causes for rising food prices in Canada, the data suggests that the rate prices were increasing both before, after, and during these periods were not significantly different. The visual trends in specific categories like meat and vegetables point to potential short-term anomalies rather than sustained, statistically significant changes, highlighting the importance of sufficient dataset sizes and statistical testing alongside visual interpretation.

5.2 Geographic and Demographic

This analysis examines how population density, income levels, and geographic factors influence food pricing across Canada. By identifying patterns through correlations and statistical tests, it provides insights into how pricing strategies and accessibility are shaped.

5.2.1 Population and Pricing

Population-related factors revealed significant correlations with product pricing, reflecting the interplay of demand and supply chain efficiency. For instance, *Seafood* (0.52, $p < 0.001$) and *Meats* (0.35, $p = 0.0016$) showed strong positive correlations, indicating that densely populated areas tend to have higher prices for pre-

mium goods. *Nuts* (-0.38, $p < 0.001$), *Fruits* (-0.40, $p < 0.001$), and *Veggies* (-0.39, $p < 0.001$) have shown negative correlations, suggesting that economies of scale and more efficient supply chains drive lower prices in these regions. Meanwhile, categories such as *Canned* (-0.02, $p = 0.88$) and *Frozen* (0.06, $p = 0.58$) showed stable pricing, unaffected by population density.

5.2.2 Income and Pricing

Income levels were shown to strongly influenced the pricing of premium goods. For example, *Meats* (0.78) and *Seafood* (0.67) demonstrated strong positive correlations, reflecting higher demand in wealthier regions. Moderate correlations for *Drinks* (0.47) and *Veggies* (0.43) suggest income impacts these categories to a lesser extent.

5.2.3 Urban vs. Rural Pricing

The Mann-Whitney U test revealed no statistically significant differences between urban and rural prices, with p-values consistently at 1.0. However, observable trends showed that urban areas had higher prices for items like meat, dairy, and fish—*Chicken thigh, per kilogram*, for example, cost 11.15% more in urban regions. Rural areas benefit from lower prices on bulk goods such as *Potatoes, 4.54 kilograms*, which were 12.8% cheaper.

5.2.4 Regional Differences

The Kruskal-Wallis test identified significant regional price disparities for categories like *Nuts* ($H = 12.545$, $p = 0.00188$) and *Seafood* ($H = 7.335$, $p = 0.0255$), reflecting the influence of local dietary habits, production capacities, and distribution networks. However, other categories, such as *Baby Food* and *Dairy*, exhibited no significant regional differences, with p-values exceeding 0.05, indicating uniform pricing across regions.

While urban areas and wealthier regions experience higher prices for premium goods, economies of scale in densely populated areas commonly reduce prices for basic staple items like fruits and vegetables. These insights offer opportunities to optimize supply chains, address regional disparities, and develop targeted pricing strategies to improve food accessibility across Canada.

6 Limitations and Insights

Our analysis faced several limitations that impacted the scope and depth of our findings. One significant issue was the exclusion of data from Canadian territories, limiting the exclusivity of the study. This prevented us from examining food pricing trends in these regions, potentially overlooking unique patterns or disparities. Additionally, the dataset covered a restricted time range from January 2017 to August 2024, which did not allow for a full comparison with the most recent price trends or historical data beyond this period. The dataset also contained redundant items, reducing the data on unique food categories. Some categories, such as specialty products, were underrepresented, making it difficult to draw accurate conclusions about these groups. Furthermore, the pre-averaged nature of the data from Statistics Canada limited our ability to analyze raw data or explore micro-level trends.

6.1 Challenges Encountered

During the analysis, the biggest challenge we had was working with the aggregated data. This made it difficult to assess individual item trends or variations within broader categories. Additionally, the lack of information about supply chain factors, such as transportation costs or distribution networks, limited our ability to connect price trends to external economic variables. Finally, the file sizes of supplementary datasets made it challenging to incorporate all potential variables without compromising performance or accessibility.

6.2 Improvements and Retrospective Actions

If given more time, we would focus on obtaining raw, unaggregated data to gain more detailed insights into price variations at a product-specific level. Access to raw data would enable

more accurate statistical testing, leading to a clearer picture of the results. We had intended to use machine learning models to predict future trends based on historical data; however, due to time constraints, we were unable to complete that section of our research. Incorporating such models could have significantly enhanced the practical relevance of our findings. In retrospect, exploring additional external datasets (supply chain costs, consumer behaviour, e.g.) could have been added. For instance, integrating datasets on weather patterns might have revealed factors that influence regional price differences. .

7 Conclusion

This project provided valuable insights into food price trends across Canada, emphasizing the temporal, geographic, and demographic factors that shape pricing. The findings reveal a complex interplay of economic and regional dynamics, highlighting how population size, income levels, and geographic location influence food affordability. These results have significant practical applications across multiple sectors. Policymakers can address regional disparities by implementing targeted subsidies to enhance affordability in urban areas and strengthen rural supply chains. Retailers can refine pricing and inventory strategies to better cater to the demands of urban and wealthy consumers while maintaining competitive pricing for staples in rural markets. Non-profit organizations, such as food banks, can prioritize resources in high-cost regions and focus on vulnerable populations. Moreover, this study highlights the importance of taking action to address food pricing disparities as well as improving supply chains in Canada.