Analyzing Toronto's Dining Scene: Public Food Safety and Economic Indicators in Focus*

Tianen (Evan) Hao

January 21, 2024

Nowadays, as the market economy continues to thrive, people have more diverse choices of edible products and are also more aware of their potential health risks. This report uses the open-source statistical programming language R to analyze a data set of food safety inspections at Toronto dining establishments and further uses additional data sets to correlate food safety trends with an economic indicator such as the Consumer Price Index (CPI). The analysis shows a limited reduction in sanitation infractions; this potentially means food safety compliance is improving in the City of Toronto. Additionally, the analysis suggests some potential links between economic downturns and increases in health infractions.

Table of contents

1 **Introduction** **2**

2 **Dataset Explanation** **2**
  2.1 Dinesafe 3
  2.2 Toronto's Dashboard Key Indicators 3

3 **Data Analysis Result** **4**
  3.1 Toronto Dine Safe Pass Status 4
  3.2 Infraction Data 5
  3.3 Infraction Type Assessment 5
  3.4 Actions from Infraction Companies 7
  3.5 Potential Relationship between CPI and Number of Infractions 8

4 **Discussion and Conclusion** **9**

---

*Code and data are available at: https://github.com/ShadyEvan4830/Toronto_Dinesafe_DataAnalysis.git

1 Introduction

Toronto's vibrant dining scene is a cornerstone of the city's culture and economy; therefore, residents and visitors need strict health screenings to ensure food safety standards. Nevertheless, the public often overlooks that dining establishments must adhere to strict health standards.

Over the years, public health and safety in the food service industry have received increasing attention since foodborne illness can have severe and sometimes fatal consequences. Each year, Toronto Public Health monitors, inspects, and reports on health and safety standards at food establishments across the city; this data is fully publicly available to allow the public to understand the current situation. Nonetheless, among the existing studies, there are few professional analysis of Toronto restaurant food safety data, or a lack of research that directly correlates economic factors with food safety standards.

This report conducts an in-depth analysis of Dinesafe data provided by the City of Toronto's open data website OpenDataToronto (Gelfand 2022). The author uses the open-source statistical programming language R (R Core Team 2024) to display trends in the data, focusing on key indicators such as inspection pass rate and changes in infraction details. In addition, this report further examines how economic changes indicated by external factors such as the Consumer Price Index (CPI) are related to changes in public health standards within the food industry. Through analysis, results found that restaurants are more likely to violate food safety regulations, and there is a correlation between CPI fluctuations and infraction behaviours. The trend shows that the number of infraction behaviours will increase when CPI decreases.

Based on the above research, the report aim to provide the public with further nuanced insights, inform policy decisions, and ultimately contribute to the ongoing conversation about public health and safety in urban food landscapes.

2 Dataset Explanation

In this report, the author used the open-source statistical programming language R (R Core Team 2024) and library packages include `AER`(Kleiber and Zeileis 2008), `future`(Bengtsson

Table 1: Sample of DineSafe Data

Establishment Status	Minimum Inspections Per Year	Severity
Pass	2	NA
Pass	2	M - Minor
Pass	2	M - Minor
Pass	2	NA
Pass	3	M - Minor
Pass	3	M - Minor

2021), `gircreds`(Csárdi 2022), `knitr`(Xie 2014), `lintr`(Hester Year of publication), `renv`(Ushey and Wickham 2023), `reprex`(Bryan et al. 2024), `styler`(Müller and Walthert 2023), `tictoc`(Izrailev 2023), `tidyverse`(Wickham et al. 2019), `tinytex`(Xie 2020), and `usethis`(Wickham et al. 2023) to conduct in-depth analysis from multiple aspects.

2.1 Dinesafe

The Dinesafe dataset is publicly available from the City of Toronto (Gelfand 2022) and has many related variables. In this report, some key variables are mainly selected, including inspection dates, establishment type, establishment status, severity, action, and minimum inspections per year, as Table 1 shown. Inspection dates represent the time range for data set inspection; this report mainly selects the range from January 2022 to January 2024 for spot inspection analysis. Establishment type represents the relevant business type when Toronto Public Health inspects, such as a restaurant, food store or food takeout. Establishment status mainly reflects whether an enterprise has passed the food safety inspection. In this data set, only relevant enterprises that successfully passed or passed conditionally are displayed. Although it does not directly reflect the situation of failed companies, the severity variable in the data set shows that these companies that passed also have infraction problems in different situations, and the action variable records the different actions taken by these companies after infractions. Finally, minimum inspections per year serve as a reference for analyzing food safety status and corporate behaviour. Considering the hugeness of the dataset data, please refer to A.

2.2 Toronto's Dashboard Key Indicators

In addition, this report hopes to find the connection between public health trends and society from a broader perspective. To this end, the report further searched for Toronto's Dashboard Key Indicators dataset through Gelfand (2022). This dataset contains many economic-related indicators that the Toronto government is open to the public; the author selected the relevant period between January 2022 and January 2023 of the Consumer Price Index (CPI), including

numerical values and years that are consistent with Dinesafe for analysis. Considering the hugeness of the dataset data, please refer to B.

3 Data Analysis Result

3.1 Toronto Dine Safe Pass Status

The report extracted the complete data through R and classified the number of institutions by Establishment Status in the Dinesafe dataset (Toronto 2021), specifically showing two categories: "Conditional Pass" and "Pass" as Figure 1 shown. Taken together, the count for the immediate "Pass" category is significantly higher with number of 76436, indicated by the higher bar on the right, indicating that more businesses are fully compliant with food safety inspection standards than those 369 businesses receiving a "Conditional Pass." "Conditional pass" indicates that certain businesses must address specific issues to fully comply with health standards. This visual demonstrates that most businesses inspected adhere to the required health and safety guidelines.

Figure 1: Number of Establishments by Status

Table 2: Total Count of different Severity levels

Severity	Total Count
Crucial	2182
Minor	28030
Not Applicable	3338
Significant	14379

3.2 Infraction Data

In one of the companies that passed above, the DineSafe data set shows many companies whose food safety inspections were judged to be infractions. The report analyzes the extent of these total infractions companies Table 2. The results show that most of the businesses that have the amount of 28030 with violations are at the "Minor" level; this means that their violations of the hygiene assessment are minor and can be improved and reach the passing standard in a relatively faster time, which is a positive representation of Infractions The overall trend can improve faster. Despite this, data shows that 14379 companies still face a "Significant" level of sanitation infractions; this means they may need to spend more time and money to improve sanitation conditions to provide better protection to the public.

3.3 Infraction Type Assessment

Since there are too many types of companies, the report selected the three most numerous categories for bar chart comparison Figure 2. They are Supermarket, Restaurant, and Food Take Out. Each category is divided into four violation severity levels: Minor, Significance, Crucial and Not Applicable (NA), which are further divided. This report referred to the relevant codes of the "Top 50 Ggplot2 Visualization" (Kassambara 2024) website when writing the code for this bar chart.

The analysis found that the category level of Minor has the highest number of infractions among all types of places, and the frequency of infractions in restaurants is exceptionally high. This suggests that while minor infractional behaviours are common, they are usually not serious and can be corrected with minimal intervention.

Compared to minor infractions, the number of infractions belonging to NA is relatively tiny. This category represents situations where standard infraction categories do not apply, perhaps due to specific exemptions or non-standard operations within the enterprise.

Significantly fewer infractions were labelled major than minor ones, which may indicate that more serious health risks are less likely to occur. Nonetheless, the presence of these infractions indicates the need for improvements to ensure public health.

The last infraction, Crucial, is the least common, indicating relatively few cases that pose an immediate and serious health risk. However, their presence, especially in restaurants and supermarkets, highlights the importance of strict and regular inspections to prevent such significant risks to public health.
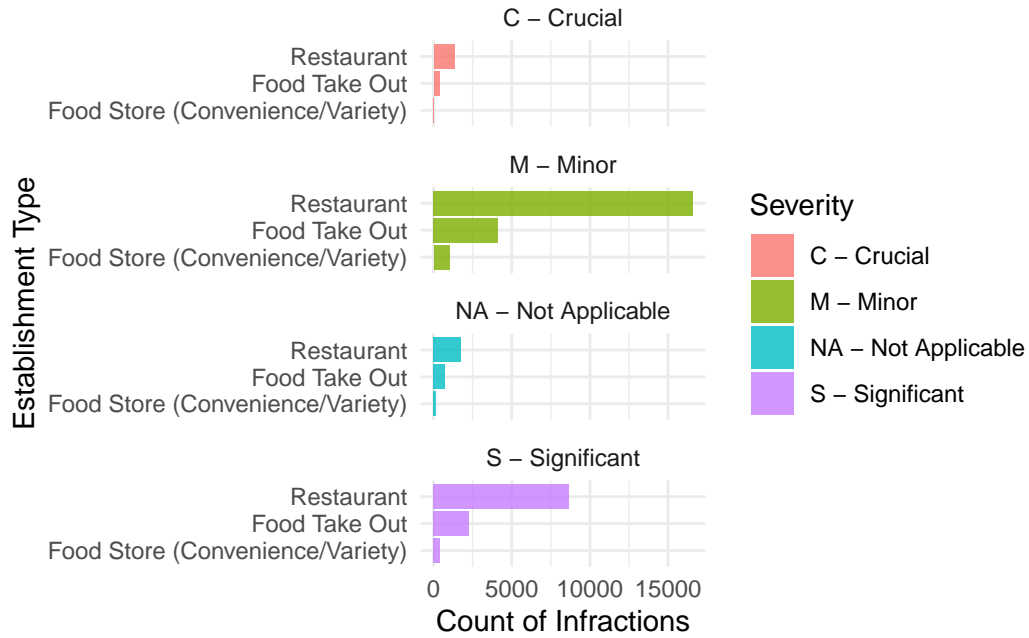

Figure 2: Distribution of Food Inspection Infractions by Severity

3.4 Actions from Infraction Companies

To further evaluate the actions these violating companies took after being flagged, the report used the "Minor" level with the most violations for analysis and created a bar chart to more clearly visualize the data Figure 3. Please note that the coding of this chart refers to the "Top 50 Ggplot2 Visualization" (Kassambara 2024) website. The chart is dominated by the green proportion, representing the behaviour of "notice to comply." This suggests that in most cases of minor level, immediate closure or fines are not the first action Toronto Public Health considers; instead, businesses are allowed to make amends within a certain period.
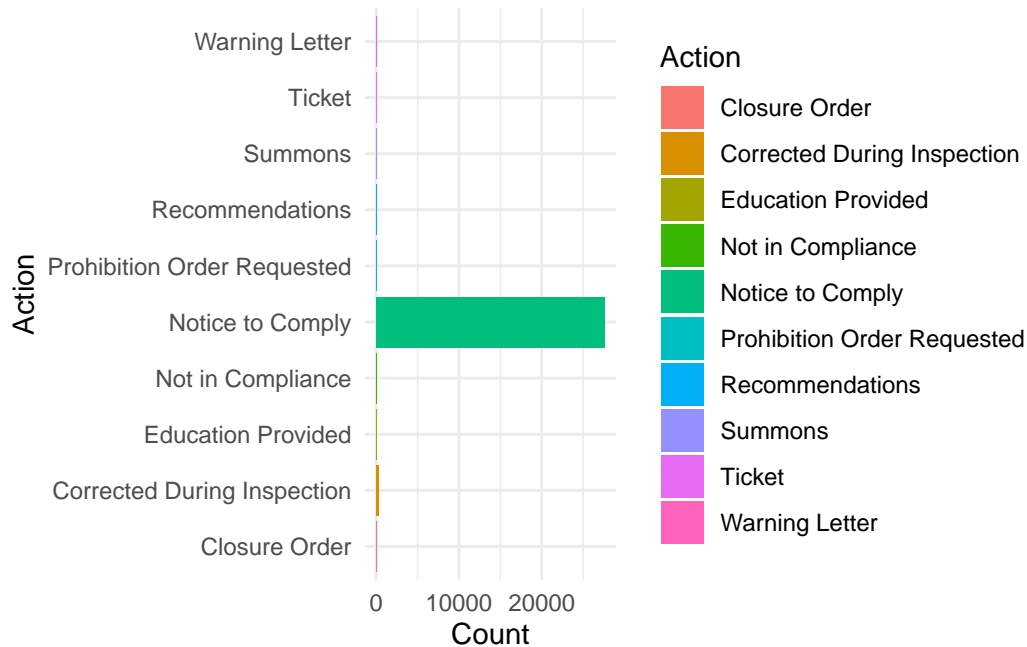

Figure 3: Distribution of Food Inspection Infractions by Severity

3.5 Potential Relationship between CPI and Number of Infractions

Finally, in order to observe food safety issues from a broader social perspective, the report chose to use the Consumer Price Index (CPI) for further comparison. CPI is an economic indicator that measures changes in the inflation rate or cost of living in a country or region (Goldberg and Campa 2010). In the early stages of writing this report, the AI language model ChatGPT 4 (OpenAI 2023) provided the author with inspiration for relevant economic indicators; the author finally selected CPI as a suitable indicator. The relevant data for this CPI comes from Toronto's Dashboard Key Indicators data set published by the City of Toronto Gelfand (2022). The report generated two line charts Figure 4, showing the CPI trend and the infraction number trend, respectively. Observation shows that when CPI decreases, corporate hygiene violations will increase. Although there may not be a direct causal relationship between the two, their trends may be affected by similar external factors. For example, a decline in CPI is usually a sign of a recession (Goldberg and Campa 2010), indicating that Toronto residents will potentially have more financial difficulties; therefore, businesses will spend less money on maintaining food safety conditions to stabilize the economic situation.
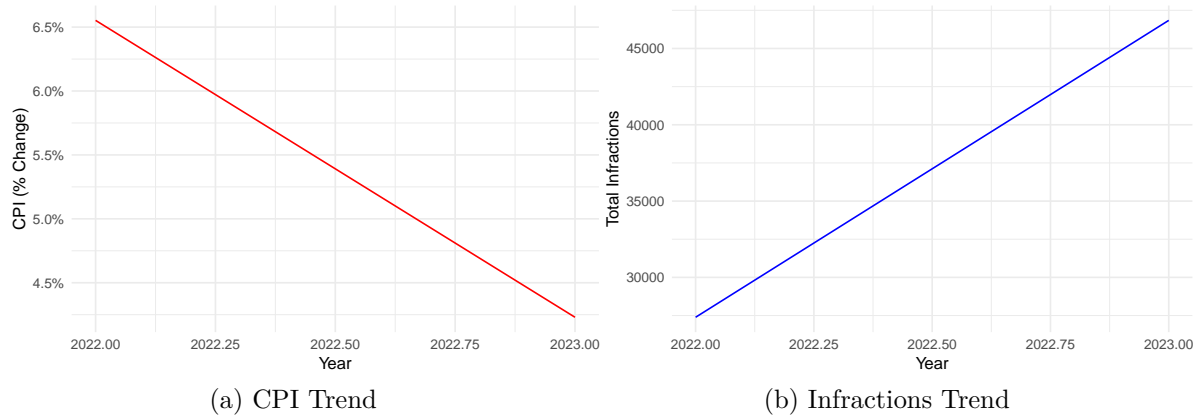

(a) CPI Trend   (b) Infractions Trend

Figure 4: Trend comparison of CPI and Infractions

4 Discussion and Conclusion

This report examines the state of public food safety in Toronto using two different data sets provided by Toronto City Hall: Dinesafe and Toronto's Dashboard Key Indicators through the statistical programming language R (R Core Team 2024) to conduct a comprehensive analysis of Toronto restaurant hygiene pass rates and Infractions. The analysis focused on an in-depth study of the DineSafe data and concluded that the number of infractions showed a clear downward trend within the constraints of the dataset. It is worth noting that the data shows that most enterprises have passed or conditionally passed the assessment of Toronto Public Health. The vast majority of enterprises marked as infractions are at the "Minor" level; they are usually notified of rectification without further punishment.

The high compliance rates above may reflect the effectiveness of Toronto Public Health's strategy; however, there are still some companies labelled as "Significance" and "Crucial" in the data. Although the number of severe levels of these food safety conditions is small, relevant administrative departments and the public must continue to be vigilant about the need and take appropriate measures to contain it.

Beyond this, our comparative analysis of DineSafe and Toronto's Dashboard Key Indicators containing the CPI data set further illuminates the interplay between economic indicators and public health outcomes. A potential economic downturn reflected in the fall in the Consumer Price Index could lead to tighter budgets within food businesses, potentially compromising hygiene standards, according to analysis from a data visualization made in R.

Despite some valid findings, the author acknowledges the limitations inherent in this report and the two datasets it used. First, the report could not examine these data sets provided by the City of Toronto for potential biases and incomplete content that could distort interpretations of actual trends. Secondly, CPI is a complex indicator that may be affected by various economic activities and policy changes. These factors have not been fully considered in this report. Therefore, future research efforts should use broader data sets and more precise analyses to provide a more convincing picture of public food safety in Toronto.

Appendix

A Clean Dinesafe Dataset

Considering the huge content of the data set, this appendix contains a preview of the cleaned data set that has been used for the research. Original data is avaliable at https://github.com/ShadyEvan4830/Tonronto_Dinesafe_DataAnalysis.git

Establishment_Type	Establishment_Status	Min_Inspections_Per_Year	Severity
Food Depot	Pass	2	NA
Food Depot	Pass	2	M - Minor
Food Depot	Pass	2	M - Minor
Food Depot	Pass	2	NA
Food Take Out	Pass	3	M - Minor
Food Take Out	Pass	3	M - Minor
Food Take Out	Pass	3	M - Minor
Food Take Out	Pass	3	M - Minor
Food Take Out	Pass	3	M - Minor
Food Take Out	Pass	3	C - Crucial
Food Take Out	Pass	3	M - Minor
Food Take Out	Pass	3	S - Significant
Food Take Out	Pass	3	S - Significant
Food Take Out	Pass	3	NA
Food Take Out	Pass	3	S - Significant
Food Take Out	Pass	3	M - Minor
Food Take Out	Pass	3	NA - Not Applicable
Food Take Out	Pass	3	C - Crucial
Food Take Out	Pass	3	S - Significant
Food Take Out	Pass	3	S - Significant

Year	CPI_Value
2008	0.0231
2008	0.0183
2008	0.0109
2008	0.0126
2008	0.0189
2008	0.0280
2008	0.0379
2008	0.0371
2008	0.0370
2008	0.0271
2008	0.0225
2008	0.0171
2009	0.0163
2009	0.0171
2009	0.0206
2009	0.0080
2009	0.0053
2009	0.0018
2009	-0.0113
2009	-0.0096

B Clean CPI Dataset

Considering the huge content of the data set, this appendix contains a preview of the cleaned data set that has been used for the research. Original data is avaliable at https://github.com/ShadyEvan4830/Tonronto_Dinesafe_DataAnalysis.git

References

Bengtsson, Henrik. 2021. "A Unifying Framework for Parallel and Distributed Processing in r Using Futures." *The R Journal* 13 (2): 208–27. https://doi.org/10.32614/RJ-2021-048.

Bryan, Jennifer, Jim Hester, David Robinson, Hadley Wickham, and Christophe Dervieux. 2024. *Reprex: Prepare Reproducible Example Code via the Clipboard.* https://reprex.tidyverse.org.

Csárdi, Gábor. 2022. *Gitcreds: Query 'Git' Credentials from 'r'.* https://gitcreds.r-lib.org/.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://sharlagelfand.github.io/opendatatoronto/.

Goldberg, Linda S, and José Manuel Campa. 2010. "The Sensitivity of the CPI to Exchange Rates: Distribution Margins, Imported Inputs, and Trade Exposure." *The Review of Economics and Statistics* 92 (2): 392–407.

Hester, Jim. Year of publication. *Lintr: A 'Linter' for r Code.* https://CRAN.R-project.org/package=lintr.

Izrailev, Sergei. 2023. *Tictoc: Functions for Timing r Scripts, as Well as Implementations of "Stack" and "StackList" Structures.* https://github.com/jabiru/tictoc.

Kassambara, Alboukadel. 2024. "Top 50 Ggplot2 Visualizations - Master List (with Full r Code)." https://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html.

Kleiber, Christian, and Achim Zeileis. 2008. *Applied Econometrics with R.* New York: Springer-Verlag. https://CRAN.R-project.org/package=AER.

Müller, Kirill, and Lorenz Walthert. 2023. *Styler: Non-Invasive Pretty Printing of r Code.* https://github.com/r-lib/styler.

OpenAI. 2023. "ChatGPT-4." https://openai.com/.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Toronto, Canada: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto, City of. 2021. "DineSafe." https://open.toronto.ca/dataset/dinesafe/.

Ushey, Kevin, and Hadley Wickham. 2023. *Renv: Project Environments.* https://rstudio.github.io/renv/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jennifer Bryan, Malcolm Barrett, and Andy Teucher. 2023. *Usethis: Automate Package and Project Setup.* https://usethis.r-lib.org.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

———. 2020. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents.* https://CRAN.R-project.org/package=tinytex.