

# Toronto Dinesafe Open data\*

A public health environment that is constantly improving

Tianen (Evan) Hao

January 14, 2024

This report will analyze publicly available data made available to the public by Toronto City Hall. Hygiene indicators in public catering have been under serious scrutiny from the public for many years. Excellent catering hygiene conditions will drive the common development of society and economy.

## 1 Introduction

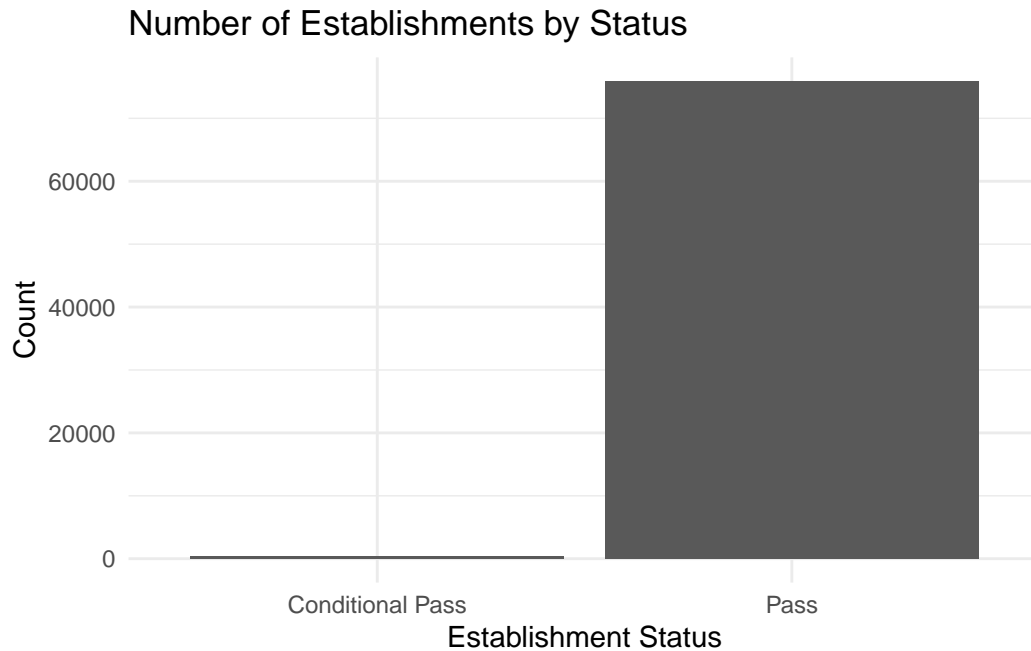
Toronto's vibrant dining scene is a cornerstone of the city's culture and economy, which is why residents and visitors require strict health screenings to ensure public safety. This report examines the Dinesafe data set provided by Toronto Public Data through the data analysis software R to conduct a comprehensive analysis of Toronto restaurant hygiene pass rates and Infractions. We further focus on the lowest inspection counts that are most frequent each year, revealing patterns that shed light on the city's public health standards.

## 2 Data Analysis

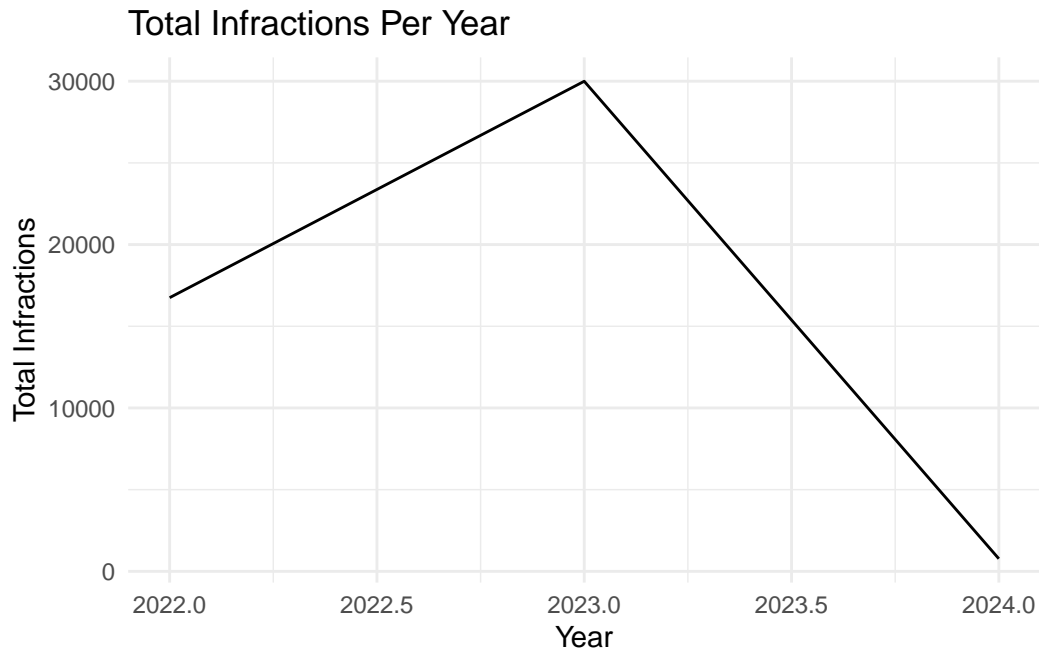
The Dinesafe dataset is publicly available from the City of Toronto (Toronto 2021) and covers several years, with key variables such as establishment type, inspection dates, and infraction details. Prior to analysis, we resolved missing data and extracted relevant temporal components for annual trend analysis.

---

\*Code and data are available at: [https://github.com/ShadyEvan4830/DataAnalysis\\_TorontoOpenData.git](https://github.com/ShadyEvan4830/DataAnalysis_TorontoOpenData.git).



We extracted the complete data through R and classified the number of institutions by Establishment Status in Dinesafe dataset (Toronto 2021), specifically showing two categories: “Conditional Pass” and “Pass”. Taken together the count for the immediate “Pass” category is significantly higher, indicated by the higher bar on the right, indicating that more businesses are fully compliant with health inspection standards than those receiving a “Conditional Pass”. “Conditional pass” indicates that a certain number of businesses must address certain issues to fully comply with health standards. This visual demonstrates that the majority of businesses inspected are adhering to the required health and safety guidelines, which is very good news for the Toronto public.



Furthermore, we created a line chart by analyzing the number of Total Infractions in the data set over the years. It can be clearly seen from the chart that the total number of violations judged by catering companies each year has begun to show a downward trend after the early cycle of 2023; this shows that Toronto’s public health safety is becoming excellent and more secure for people’s health.

```
# A tibble: 3 x 2
  Severity    Total_Count
  <chr>         <int>
1 Crucial         2155
2 Minor        27792
3 Significant    14260
```

Finally, we conduct statistical analysis on the extent of these Total Infractions companies. The results show that most of the businesses with violations are at the “Minor” level (27792); 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 14,260 companies still face “Significant” level of sanitation infractions; this means that they may need to spend more time and cost to improve sanitation conditions to provide better protection to the public.

### 3 Discussion and Conclusion

In summary, the purpose of this report is to examine the Dinesafe data set provided by Toronto Public Data (Toronto 2021) through the statistical programming language R (R Core Team 2024) to conduct a comprehensive analysis of Toronto restaurant hygiene pass rates and Infractions. Include packages Kleiber and Zeileis (2008), Bengtsson (2021), Csárdi (2022), (**knitr?**), Hester (Year of publication), Ushey and Wickham (2023), Bryan et al. (2024), Müller and Walthert (2023), Izrailev (2023), Wickham et al. (2019), Xie (2020), and Wickham et al. (2023).

Analysis of data sets through R and analysis of trends over time provide a valuable perspective for assessing the effectiveness of public health policies. We found that the number of infractions in Toronto restaurants is declining year over year, which means more restaurants are meeting health standards and open to the public. The highest proportion of restaurants that failed were rated “minor,” indicating that more restaurants have the opportunity to improve their hygiene in a short period of time to comply with Toronto Health standards. Taken together, the above information is a very positive feedback for Toronto residents and tourists because Toronto’s public restaurants are becoming healthy and standardized and increasing in number to promote economic development.

Through the analysis of the Torotno DineSafe data set, cleaning and categorizing the “severity” data allows us to conduct more granular analysis, such as in this report, making it more clear to highlight the extent of specific areas of food safety that require improvement. Distributed so that people can better analyze the hygiene status of the restaurant industry.

While this report provides a handful of solid insights, it also acknowledges the limitations of the dataset, such as potential reporting bias and missing values that may exist in the dataset. Future research could incorporate longitudinal data analysis to capture the evolution of compliance and noncompliance. In addition, this report is only a simple analysis produced for INF312 TUT2, missing many more aspects and details of description and judgment; therefore, if more accurate results are needed for the current hygiene and health of restaurants in Toronto, a more detailed and professional analysis is required.

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