# Restaurants in Eastern Toronto More Likely to Cause Severe DineSafe Infractions*

Moohaeng Sohn

January 25, 2024

DineSafe is a program ran by Toronto Public Health which inspects establishments which serve food. We found that there are more severe infractions in the eastern part of Toronto, but ward's income had near-zero correlation to infraction rates. We used `opendatatoronto`'s datasets on DineSafe inspections and ward related data, along with various `R` packages to organize and visualize data. This insight could be useful for consumers who are looking for a safe dining experience.

## Table of contents

---

*Code and data used in this analysis is available at: https://github.com/alexsohn1126/fire_and_income_toronto

# 1  Introduction

Toronto is a city bustling with many different cultures mixed together. With many cultures, there are many restaurants in Toronto. In Greater Toronto Area, there are around 17000 restaurants that are operating. Restaurants give the owners ability to express their cultural and culinary creativity to almost 3 million people in the city of Toronto, and 6 million people in Greater Toronto Area (Toronto 2023c). Being a cultural hub and home to a variety of restaurants also means it is a challenge to keep a consistent health standard for every restaurant.

DineSafe program is ran by Toronto Public Health, and it runs randomized inspections for restaurants or any establishments that serve food to the general public. DineSafe inspections occur anywhere from once per year to 3 times per year, depending on whether the establishment is serving food that can possibly carry pathogenic organisms, or is serving food to a population that is high risk, such as hospital patients (Toronto Public Health 2019). There are 3 possible results to a DineSafe inspection. One can get a PASS, which indicate no or only minor infractions were found, or a CONDITIONAL PASS, meaning one or more significant infraction was observed, or a CLOSED, meaning one or more crucial infractions were found. For example, if an establishment were found to have 2 minor infractions and 1 crucial infraction, then they will be notified to close the establishment until the issue is fixed. Until a PASS re-inspection, they cannot reopen the establishment. Conditional pass means the establishment will be revisited in the near future, and all of the significant infractions must be fixed by then. These notices must be posted near the entrance of the restaurant, visible for anyone who enters the premise (Toronto Public Health 2019).

We have observed no significant correlation between ward income and DineSafe infraction rates, but we did observe eastern wards of Toronto being responsible for more severe infractions than minor infractions, while central wards are more likely to cause any infraction compared to other wards. Western wards seems to have the low DineSafe infraction rates regardless of infraction types compared to other wards.

In this paper, we will analyze DineSafe inspection data to see whether there are correlating factors which play into the inspection result of these establishments. This paper will be organized into these following sections: Data, Results, Discussion, and Conclusion. Data section will focus on how the data was obtained and what steps we took to clean the data. Results section will use the data and use various graphs and tables to showcase our data.

Discussion section will point out our findings and how we performed the analysis, and the possible shortcomings for the analysis we have done.

## 2 Data

All data was obtained by using Toronto's `opendatatoronto` R library (Gelfand 2022). Toronto hosts its open data portal website, where everyone can search and explore hundreds of datasets related to Toronto. We can use R statistical programming language (R Core Team 2022) with `opendatatoronto` library to download the datasets into R directly. We used R statistical programming language (R Core Team 2022) along with `openxlsx` (Schauberger and Walker 2023), `tidyverse` (Wickham et al. 2019), `sf` (Pebesma 2018) , `here` (Müller 2020), `knitr` (Xie 2023), `patchwork` (Pedersen 2024), `janitor` (Firke 2023) libraries to help save, process, then visualize the raw data. All the code was written within RStudio (Posit team 2023).

### 2.1 DineSafe Data

DineSafe dataset is published by Toronto Public Health (or TPH) (Toronto Public Health 2019). This dataset contains many variables for every inspection results, such as the location of the establishment in longitude and latitude, infraction severity (if any), and the date which the inspection was performed. There are 76827 inspections which we can analyze over as of January 21st, 2024. This dataset is updated daily. The date when the raw data was obtained for the project was on January 21st, 2024, and open data portal states that the dataset was last refreshed on Janurary 21st, 2024. DineSafe dataset contains inspections from January 20th, 2022 to January 19th, 2024, data from last 2 years.

### 2.2 Ward Bounds Data

City of Toronto is divided into 25 wards as of January 21st, 2024. This dataset is a map of all the wards in the City of Toronto, and is published by the City Clerk's Office. It is updated semi-annually, and was last updated on Jan 1st, 2024. This map data was given in a geojson format. The dataset allows us to query location with a longitude and a latitude and let us see which ward that location is in.

It is important that we use 25 ward map, not a 44 ward map that is also available on `opendatatoronto` because 25 ward map is the most commonly used ever since it came into use in 2018 (Toronto 2023b). There is also a 47 ward model, but it is very rarely used as it was not implemented, but it is still worth mentioning because it has been approved by the city council and there are 47 wards bounds data and profile data (which is the next section) (Toronto 2023a).

## 2.3 Ward Profile Data

This dataset contains 2011, 2016, and 2021 Census data and summarizes them into Toronto's 25 wards. This dataset was last refreshed on Jan 3rd, 2024, and was accessed for use on Jan 21st, 2024. We will only be using the Median household income of the wards from 2021 Census from this dataset. We decided to use 2021 Census data as it is the latest data available to us. Using older census data means we are using very recent DineSafe infraction data with misrepresented economic status of those wards. Again, we are using 25 ward version as it is the most up to date, and most commonly used.

## 2.4 Combined Data

We combined DineSafe dataset and ward profile data to produce a combined dataset. The combined dataset contains these columns: establishment's name, DineSafe infraction severity, the ward which the establishment is located in, and median household income of that ward. Table 3 shows the first few rows of the cleaned dataset. We carry over the 76,827 inspections that we had from DineSafe dataset, except now we have more information about which ward the establishment is located, and the median income of that ward. The process that was used during the combining process is documented in Section A.

First, we will group DineSafe inspection data by their restaurants. Grouping data by restaurant means an outlier restaurant which gets a lot of infractions will not have that big of an effect. Table 1 shows the number of restaurants in total within the ward, and number of restaurants which have received at least one minor, significant, or crucial infractions respectively. We can observe that there are most restaurants in ward 10, having 1999 restaurants, meanwhile ward 2 has the lowest number of restaurants. Starting with minor infractions, we can see that ward 25 has the lowest % of restaurants with a minor infraction (or more) at 23.1%. On the other hand, 77.8% of ward 8's restaurants had a DineSafe infraction, which is consistent with what we have observed in Figure 1. Ward 8 also has the highest significant and crucial infraction percentage. Ward 1 has the lowest % of significant infractions at 17.7%, and ward 2 has the lowest % of crucial infractions at 1.96%. It has to be noted that ward 2 also had the lowest number of restaurants, meaning this may not be reliable data.

Table 1: Total Number of Restaurants within that ward, Number of Restaurants with At Least One Minor, Significant, Crucial Infractions (Inf.) Received with their relative percentage to the total number of restaurant within their ward, from Jan 20, 2022 to Jan 19, 2024

| Ward | All Restaurants | Minor Inf. | (%) | Significant Inf. | (%) | Crucial Inf. | (%) |
|------|-----------------|------------|------|------------------|------|--------------|------|
| 1 | 623 | 211 | 33.9 | 110 | 17.7 | 25 | 4.01 |
| 2 | 255 | 135 | 52.9 | 65 | 25.5 | 5 | 1.96 |
| 3 | 849 | 328 | 38.6 | 153 | 18 | 23 | 2.71 |

| Ward | All Restaurants | Minor Inf. | (%) | Significant Inf. | (%) | Crucial Inf. | (%) |
|---|---|---|---|---|---|---|---|
| 4 | 682 | 353 | 51.8 | 237 | 34.8 | 43 | 6.3 |
| 5 | 557 | 252 | 45.2 | 158 | 28.4 | 37 | 6.64 |
| 6 | 581 | 339 | 58.3 | 226 | 38.9 | 56 | 9.64 |
| 7 | 513 | 241 | 47 | 99 | 19.3 | 24 | 4.68 |
| 8 | 567 | 441 | 77.8 | 348 | 61.4 | 106 | 18.7 |
| 9 | 788 | 325 | 41.2 | 200 | 25.4 | 27 | 3.43 |
| 10 | 1999 | 784 | 39.2 | 554 | 27.7 | 101 | 5.05 |
| 11 | 1376 | 877 | 63.7 | 710 | 51.6 | 97 | 7.05 |
| 12 | 499 | 305 | 61.1 | 246 | 49.3 | 71 | 14.2 |
| 13 | 1073 | 627 | 58.4 | 417 | 38.9 | 68 | 6.34 |
| 14 | 827 | 356 | 43 | 266 | 32.2 | 78 | 9.43 |
| 15 | 400 | 272 | 68 | 221 | 55.2 | 57 | 14.2 |
| 16 | 365 | 211 | 57.8 | 172 | 47.1 | 39 | 10.7 |
| 17 | 330 | 155 | 47 | 122 | 37 | 26 | 7.88 |
| 18 | 527 | 334 | 63.4 | 246 | 46.7 | 60 | 11.4 |
| 19 | 528 | 221 | 41.9 | 142 | 26.9 | 39 | 7.39 |
| 20 | 475 | 221 | 46.5 | 146 | 30.7 | 39 | 8.21 |
| 21 | 618 | 295 | 47.7 | 226 | 36.6 | 63 | 10.2 |
| 22 | 512 | 300 | 58.6 | 233 | 45.5 | 56 | 10.9 |
| 23 | 564 | 264 | 46.8 | 193 | 34.2 | 72 | 12.8 |
| 24 | 396 | 145 | 36.6 | 138 | 34.8 | 28 | 7.07 |
| 25 | 264 | 61 | 23.1 | 51 | 19.3 | 19 | 7.2 |
| Total | 16168 | 8053 | - | 5679 | - | 1259 | - |

Figure 1 gives an overview of what proportion of inspections led to an infraction being found for each ward. Here we can see that more severe infractions are usually rarer than less severe infractions. Ward Eglinton-Lawrence (Ward 8) had the highest proportion of all infractions, nearly 81% of inspection leading into some kind of infraction. Etobicoke-Lakeshore (Ward 3) had the lowest proportion of all infractions, only 34% of all inspections having an infraction. As for Minor infractions, Eglinton-Lawrence had the highest proportion, 49% of inspections leading to a minor infraction, and Scarborough-Rouge Park (Ward 25) had the lowest proportion with 21% of inspections observing a minor infraction. For Significant infractions, Eglinton-Lawrence had the highest proportion with 27%, and Etobicoke-Lakeshore had the lowest with 7%. Lastly, Scarborough North (Ward 23) had the highest proportion of crucial infractions with 5.8%, and Etobicoke Centre (Ward 2) had the lowest proportion of crucial infraction with 0.8%.

We can combine the ward map data alongside with the cleaned data to see whether we can spot a pattern in ward location and DineSafe infractions. Figure 2 shows what proportion of inspections led to an infraction (all, minor, significant, or crucial) for each ward. For all infractions, we can see that the wards near the center of Toronto has higher infraction
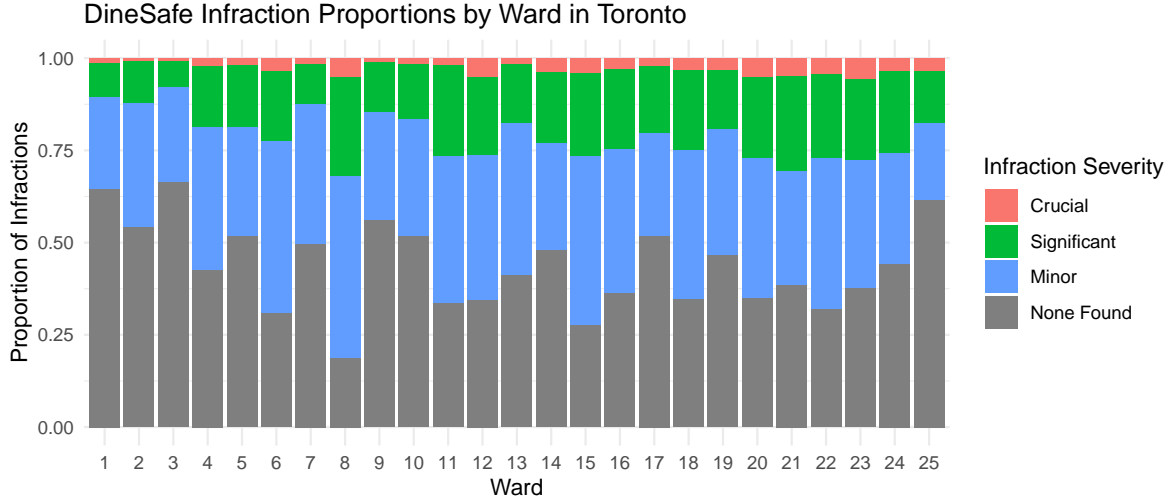
Figure 1: DineSafe Infraction Proportions by Ward in Toronto, from Jan 20, 2022 to Jan 19, 2024

proportions. But as the infraction severity increases from minor to crucial, we can observe that eastern wards is responsible for higher infraction proportions compared to other wards. This suggest there is something about eastern part of Toronto which causes these establishments to be caught with an infraction more than their western counterpart.

Figure 3 shows the infraction proportion in relation to the income of the wards. As shown in the figure, there does not seem to be a clear correlation between these two variables. The mean of all infraction proportions is 0.38, median is 0.33, standard deviation, how far the infraction proportions are spread, is 0.38. And the correlation coefficient, which shows to how strong the connection is between two variables, ranging from -1 which means it is very clear that as one variable increases, another decreases, and to 1 which means if one variable increases, another increases. between the income and the overall infraction proportion is -0.023 which is close to 0. This means there is almost no reason to believe that these two variables affect each other in any way.

## 3 Results

When we obtained the DineSafe dataset (Jan 21st, 2024), there are 76,827 DineSafe inspections across 13,936 different restaurants. Of those inspections, 32,227 inspections did not have any infractions, meaning around 42% of inspections passed without minor, significant, or crucial infractions. Out of 13,936 restaurants, about 6,057 restaurants did not have any infractions, 2,816 had Minor infractions at worst, 3,872 had Significant infractions at worst, and 1,191 restaurants had crucial infractions. Table 2 shows this. On average, there were 3.2 total

All Infractions

Minor Infractions

Significant Infractions

Crucial Infractions

Figure 2: Heat Map of Infraction Proportion for Each Ward using Toronto's 25-ward Model
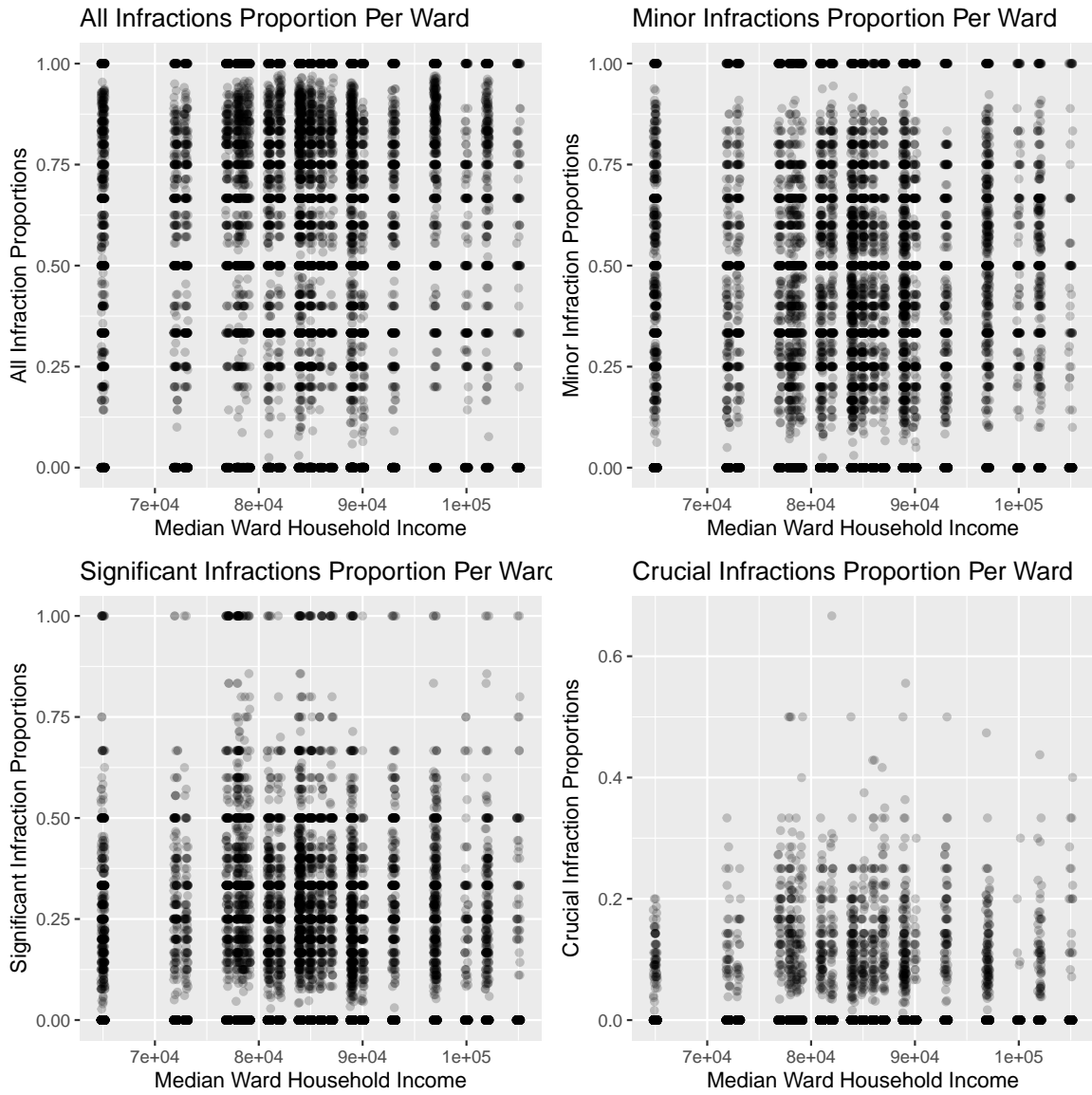
Figure 3: Median Ward Household Income Vs. Infraction Proportions Varying In Severity

infractions, 2.01 minor infractions, 1.03 significant infractions, 0.157 crucial infractions per restaurant. The standard deviation for average number of all infractions is 2.57.

Table 2: Count of Restaurants' Worst Infraction In Toronto

|  | All Restaurants | No Infractions | Minor Infraction at worst | Significant Infraction at worst | Cruicial Infraction at worst |
|---|---|---|---|---|---|
| Count | 13936 | 6057 | 2816 | 3872 | 1191 |

We also observed little to no correlation between ward median houshold income and DineSafe infraction rate. This shows that perhaps DineSafe is doing a good job keeping the standard fairly even across the city. But we did observe that while wards in the central parts of the city commit the most infractions, the eastern wards seem to have the high crucial infraction rates, which are more severe, even getting close to being the highest crucial infraction rate ward. This could raise some questions as to whether this is caused by the actual physical location's properties such as climate, or outside factors such as how thorough the inspections are, and the types of establishments that are located in those areas.

## 4 Discussion

### 4.1 Number of Restaurants with DineSafe Infractions

From Table 2, we can observe that nearly 10% of all establishments in Toronto has caused at least one crucial infraction, meaning establishments would have to close until the problem was fixed. This is somewhat worrying as that would mean one in 10 restaurants have been found of a crucial infraction which could cause illnesses.

There were surprisingly more establishments which have been found to have significant infractions at worst than minor infractions. This could be because those restaurants who have an infraction at all would have more than one infraction, one of them being significant, rather than having a sigular minor infraction. Or, it could be that some inspectors focused more on significant infractions rather than minor ones, as minor infractions do not affect what DineSafe result the restaurant gets. It is only when an establishment gets a significant or worse infraction which they get a conditional pass or a closed notice. Thus these inspectors are more focused on finding significant or crucial infractions.

## 4.2 DineSafe Infraction and Wards

Looking at maps from Figure 2, there seems to be a some sort of trend of where there were the highest infractions rates. For all infraction, it seems like wards that are close to the center of Toronto has higher general infraction rate than the outer wards. This could possibly be that these establishments are actually causing more infractions, or it could be that inspectors in that area could be stricter than those inspectors who inspect establishments in the outer wards. Another possibility could be that there are not that many establishments which serve hazardous food, and those establishment which don't serve possibly hazardous food are not inspected to the same standard as those who do.

As we look at different infraction severities however, it seems like the east side of the city causes more severe infraction compared to the west side. One could interpret this as the wards in the east side causing less infractions in numbers, but each infractions are more severe in the east. Again, this could be linked to any of the reasons stated previously, or it could be that mice/insect infestations occur more frequently in the eastern region of the city. Or, it could be that establishments in the eastern region have older buildings which are vulnerable to insects and rodent infestation.

## 4.3 DineSafe Infraction and Ward Income

There doesn't seem to be a clear correlation between DineSafe Infractions and a ward's median income. This is evident looking at Figure 3, the data points are very spread out, with no recognizable pattern. Therefore the ward's median income is not a good indicator of DineSafe infraction rates.

We observed the relationship between the ward income and DineSafe infraction because we thought there would be higher end establishments which can afford more experienced staff and equipment to clean their establishments better, possibly passing more DineSafe inspections without infractions. Clearly this does not seem to be the case. It could be that wards are too big, and it includes too many neighbourhoods thus containing restaurants from many variety of quality.

## 4.4 Weaknesses and next steps

DineSafe inspections are done by numerous inspectors, and different inspectors will be better at pointing out infraction than others. This means using DineSafe infractions to decide which restaurants are the cleanest, and which wards have the cleanest restaurants may not be the most appropriate. Not only that, there can be a seasonal difference too. For example, mice infestations are common during fall and winter seasons because they want to find a shelter for cold climate. Meaning if an inspection was done in fall or winter rather than spring or summer, an establishment might have a higher chance of getting caught with mice infestation. Having

any kind of mice infestation leads to a closed notice, as it is a crucial infraction (Toronto Public Health 2019). Also, ward income data is a snapshot of 2021 census data, meanwhile DineSafe dataset only contains inspection from January 20th, 2022 to January 19th, 2022. Therefore there is a discrepancy between those datasets for when the data was collected.

For future researches, sub-dividing wards into smaller neighbourhoods can allows us to pinpoint a clearer trend between DineSafe infractions and that neighbourhood's income. This could mitigate the problem with wards being too big and including too many variety of restaurants. We can also divide the restaurants into 3 categories: high-end, middle-end, low-end depending on the average price of the menu items in the restaurant, and see whether there is a clear trend between the quality of the establishments and DineSafe infraction rates.

# Appendix

## A  Combining and Cleaning Datasets

There are a lot of detailed information which we do not need for our purposes. We only want restaurant's name, severity of infractions, the ward where the restaurant is located in, and that ward's median income. To do this, we first use `filter` function from `tidyverse` (Wickham et al. 2019) on the ward census dataset to filter out the row which contained median income for each ward, and save it as a vector so the first element is the median income for ward 1, and second element for ward 2 and so on. This allows us to simply query what the median income is for a certain ward. We also make a custom function which will take in coordinate and spits out which ward the coordinate is located in.

Using these two things, we can start assembling our final dataset. First, we observed that some inspection's severity was marked as "`NA - Not Applicable`", and some were values `NA` which means that cell is completely empty. We keep the consistency by using the `mutate` function and changing every occurrence of "`NA - Not Applicable`" to `NA`. We use `mutate` again to add a `ward` column to the table, using `Map` function to apply the ward-finding function for each row's longitude-latitude combo. Then, using the median income info, we add another column using `mutate` to add in that ward's median income for every row. Then, we use the `select` function to select only the columns we want and save the cleaned data into a csv. Table 3 is a sample of the cleaned dataset.

## B  Cleaned Data Sample

Table 3: First Few Rows of the Cleaned Dataset

| Restaurant | DineSafe Infraction | Ward | Ward Median Income |
|---|---|---|---|
| NEW KANTAMANTO MARKET | NA | 7 | 73000 |
| NEW KANTAMANTO MARKET | M - Minor | 7 | 73000 |
| NEW KANTAMANTO MARKET | M - Minor | 7 | 73000 |
| NEW KANTAMANTO MARKET | NA | 7 | 73000 |
| # HASHTAG INDIA RESTAURANT | M - Minor | 16 | 78500 |
| # HASHTAG INDIA RESTAURANT | M - Minor | 16 | 78500 |

# References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. https://doi.org/10.32614/RJ-2018-009.

Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots.* https://CRAN.R-project.org/package=patchwork.

Posit team. 2023. *RStudio: Integrated Development Environment for r.* Boston, MA: Posit Software, PBC. http://www.posit.co/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schauberger, Philipp, and Alexander Walker. 2023. *Openxlsx: Read, Write and Edit Xlsx Files.* https://CRAN.R-project.org/package=openxlsx.

Toronto, City of. 2023a. "47-Ward Model (Ward Boundary Review, 2016)." *City of Toronto.* https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/ward-profiles/47-ward-model/.

———. 2023b. "Ward Profiles." https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/ward-profiles/.

———. 2023c. "Toronto at a Glance." *City of Toronto.* https://www.toronto.ca/city-government/data-research-maps/toronto-at-a-glance/.

Toronto Public Health. 2019. "DineSafe Inspection and Disclosure System." https://web.archive.org/web/20190624172604/https://www.toronto.ca/health/dinesafe/system.htm#crucial.

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.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.