Patterns for DineSafe Infractions in Various Toronto Wards: A Visual Analysis*

Moohaeng Sohn

January 23, 2024

Dinesafe is a program ran by Toronto Public Health which inspects establishments which serve and provide food. We found that there are more severe infractions in the eastern part of Toronto, but ward's income did not have any 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 shows that some wards have higher infraction rate than other wards.

Table of contents

---

*Code and data are 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 GTA, there are around 17000 restaurants that are operating. Restaurants give the owners ability to express their cultural and culinary creativity to almost 30 million people of Toronto, and more to those who comes from outside of the city. Being a cultural hub and home to a variety of restaurants also mean it is a challenge to keep a consistent health standard for every restaurant. And the number of restaurants growing [TODO:ADD CITATION], Toronto needed a system to keep all the restaurants safe for the customers to eat in.

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 [TODO:ADD CITATION]. 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 [TODO: ADD CITATION].

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) 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) [TODO: CITE TPH]. This dataset contains 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 around 77000 rows, meaning there are 77000 inspection 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. The data was given in a csv file, and we used `write_csv` function to save the raw data. Dinesafe dataset contains inspections from January 20th, 2022 to January 19th, 2024.

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. We use this data to query which ward this establishment was located in. Details on how this was done will be in Section 2.4. A sample of what Ward Bounds data looks like is shown in Figure 3.

3

2.3 Ward Profile Data

This dataset contains 2021 Census data and summarizes them into 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 income of the wards from this dataset.

2.4 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` [TODO: CITE TIDYVERSE] 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.

3 Results

3.1 Worst Dinesafe Infraction Per Restaurant

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 1 shows this with the percentages compared to all restaurants. On average, there were 3.2 total 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 1: Count and Percentage of Different Worst Infraction by Restaurant

|	All Restaurants	No Infractions	Minor Infraction at worst	Significant Infraction at worst	Cruicial Infraction at worst
Count	13936	6057	2816	3872	1191
Percent	100%	43.46 %	20.21 %	27.78 %	8.55 %

3.2 Dinesafe Infractions Per Ward

We can use `sf` package to visualize what percentage of inspections led to an infraction for each ward. Ward Eglinton-Lawrence (Ward 8) had the highest rate of all infractions, nearly 81% of inspection leading into some kind of infraction. Etobicoke-Lakeshore (Ward 3) had the lowest rate of all infractions, only 34% of all inspections having an infraction. As for Minor infractions, Eglinton-Lawrence had the highest rate, 49% of inspections leading to a minor infraction, and Scarborough-Rouge Park (Ward 25) had the lowest rate with 21% of inspections observing a minor infraction. For Significant infractions, Eglinton-Lawrence had the highest rate with 27%, and Etobicoke-Lakeshore had the lowest with 7%. Lastly, Scarborough North (Ward 23) had the highest rate of crucial infractions with 5.8%, and Etobicoke Centre (Ward 2) had the lowest rate of crucial infraction with 0.8%. Refer to Figure 1 for a map visualizing this data.

3.3 Dinesafe Infractions Versus Ward Income

We created Figure 2 to see whether there are any relationship between infraction occurrence rate and ward income, but there doesn't seem to be a clear trend for all 4 scatter plots which highlight all, minor, significant, and crucial infraction rates versus ward income for which those infractions were found at. Table 2 contains the correlation between ward income and various infraction types, and it shows very little correlation, strongest correlation coefficient being 0.15.
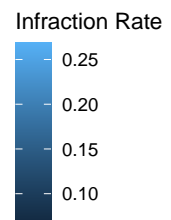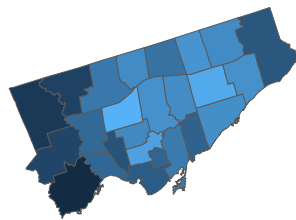
All Infractions
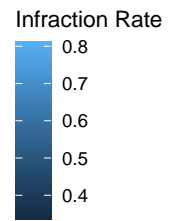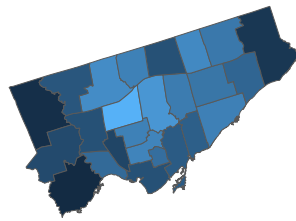
Minor Infractions
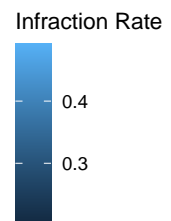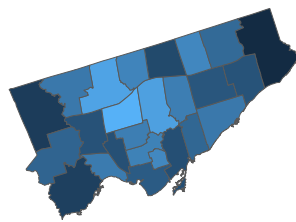
Significant Infractions
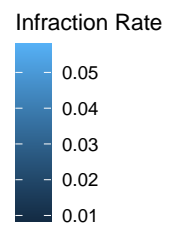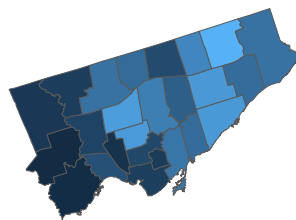
Crucial Infractions

Figure 1: Infraction Percentage for Each Ward from 2022 to 2024

Table 2: Correlation of Income and Various Infractions

Infraction type	Correlation to ward Income
All Infraction	-0.0772076
Minor Infraction	-0.1120716
Significant Infraction	-0.0647607
Crucial Infraction	0.1484907

4 Discussion

4.1 Number of Restaurants with Dinesafe Infractions

From Table 1, 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 1, 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
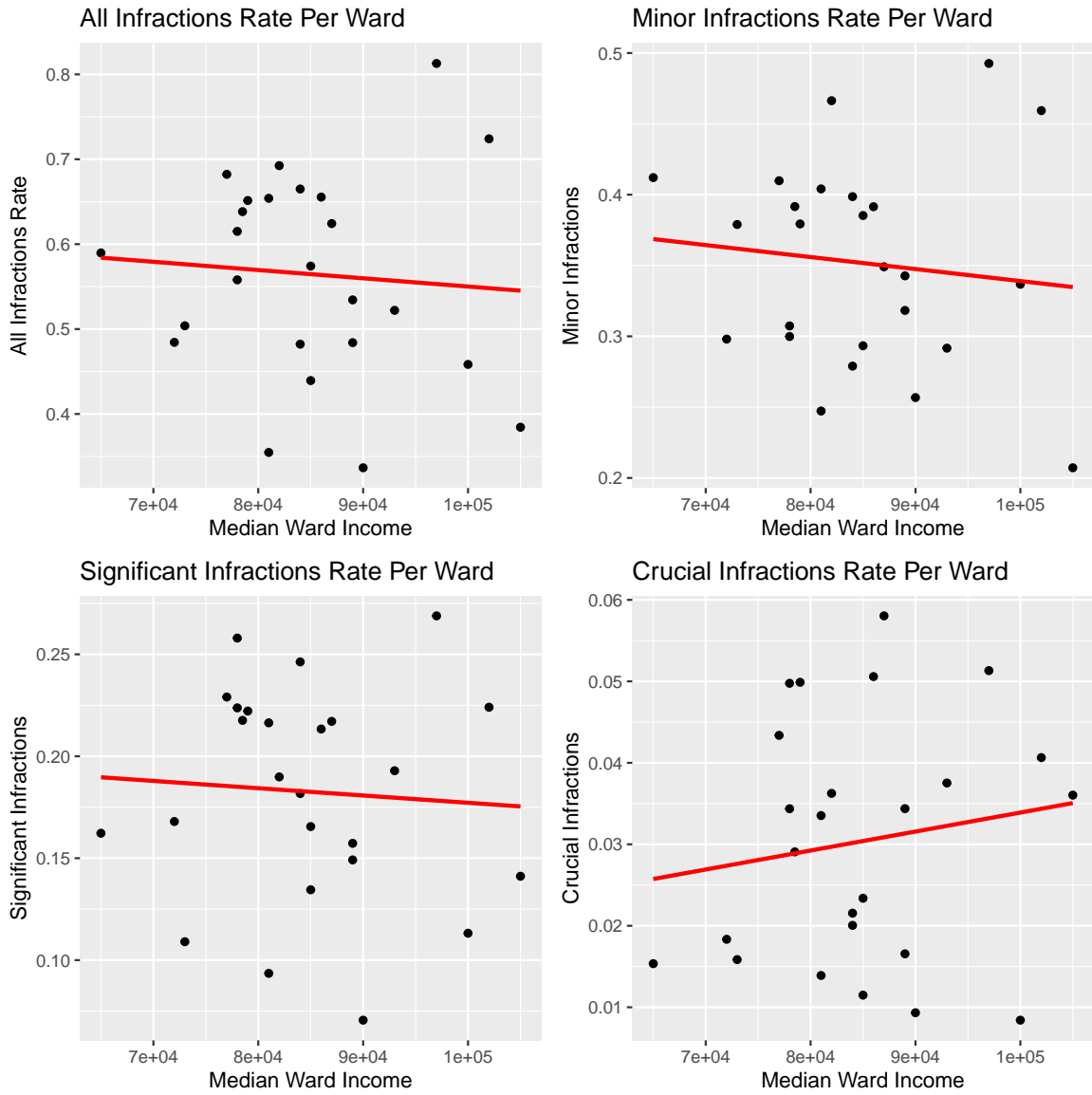
Figure 2: Median Ward Household Income Vs. Infractions Ratio Varying In Severity

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 Table 2 and Figure 2, the correlation coefficient is positive or negative depending on which severity of infraction we look at, and the data points are very spread out. 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 the most common during fall and winter season [TODO: ADD SOURCE HERE]. 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 [TODO: Add archive.org page about this]. 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.
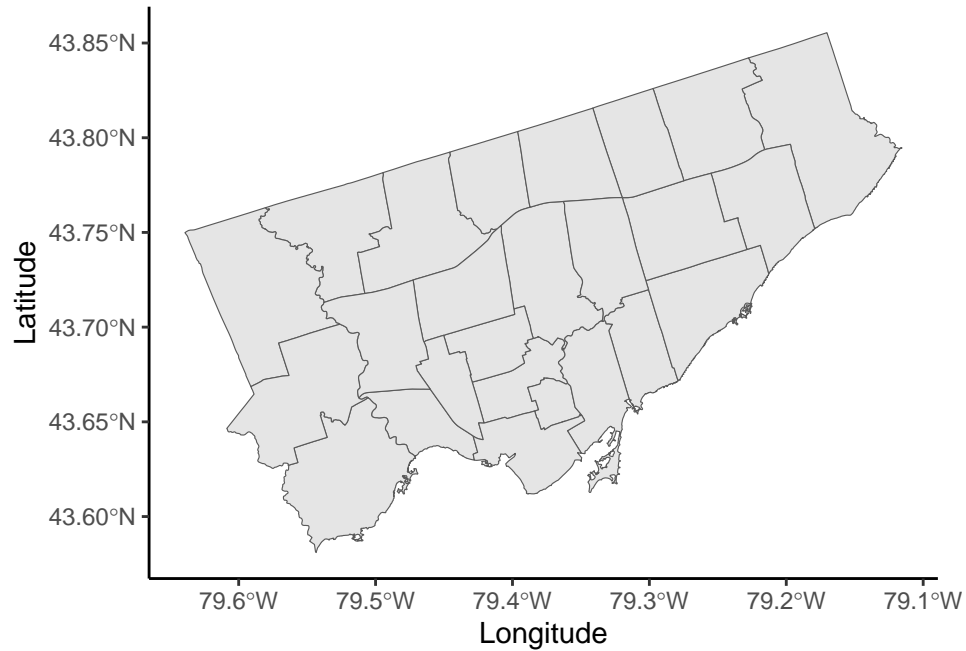
Appendix

A Ward Map Visualization

Figure 3: Visualization of Ward Map Data

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

Restaurant	Dinesafe Infraction	Ward	Ward Median Income
# HASHTAG INDIA RESTAURANT	M - Minor	16	78500
# HASHTAG INDIA RESTAURANT	M - Minor	16	78500

References

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.

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/.