

# Do Restaurants in A study of relationship between neighbourhood income and DineSafe infractions\*

Moohaeng Sohn

January 22, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>3</b>
2.1	Dinesafe Data . . . . .	3
2.2	Ward Bounds Data . . . . .	3
2.3	Ward Profile Data . . . . .	4
2.4	Combining and Cleaning Datasets . . . . .	4
<b>3</b>	<b>Results</b>	<b>4</b>
3.1	Worst Dinesafe Infraction Per Restaurant . . . . .	4
3.2	Dinesafe Infractions Per Ward . . . . .	5
3.3	Dinesafe Infractions Versus Ward Income . . . . .	5
<b>4</b>	<b>Discussion</b>	<b>8</b>
4.2	Second discussion point . . . . .	8
4.3	Third discussion point . . . . .	8
4.4	Weaknesses and next steps . . . . .	8
	<b>Appendix</b>	<b>9</b>
<b>A</b>	<b>Ward Map Visualization</b>	<b>9</b>

---

\*Code and data are available at: [https://github.com/alexsohn1126/fire\\_and\\_income\\_toronto](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 come 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 [TODO: CITE OPENDATATORONTO]. 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 [TODO: CITE R HERE] with `opendatatoronto` library to download the dataset into R directly. We used R statistical programming language [TODO: CITE R HERE] along with `openxlsx` [TODO: CITE openxlsx], `tidyverse` [TODO: CITE TIDYVERSE], `sf` [TODO: CITE sf], `here` [TODO: CITE here], `kableExtra` [TODO: CITE KABLEEXTRA], `patchwork` [TODO: CITE PATCHWORK] libraries to help save and process the raw data. All the code was written within RStudio [TODO: CITE RSTUDIO].

### 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 January 21st, 2024. The data was given in a csv file, and we used `write_csv` function to save the raw data.

### 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.

## 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. Figure 4 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	Crucial Infraction at worst
Count	13936	6057	2816	3872	1191
Percentage	100%	43.46 %	20.21 %	27.78 %	8.55 %

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

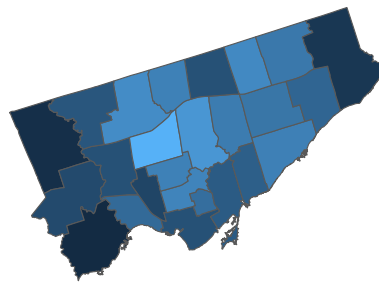
### 3.2 Dinesafe Infractions Per Ward

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.

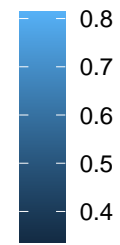
### 3.3 Dinesafe Infractions Versus Ward Income

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.

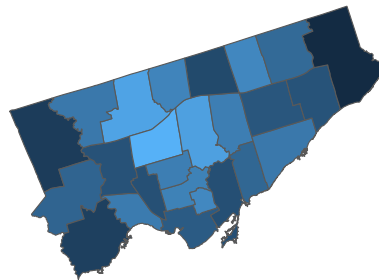
All Infractions



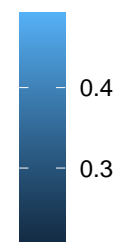
Infraction Ratio



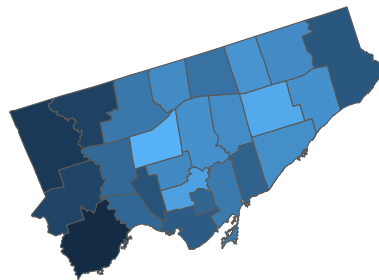
Minor Infractions



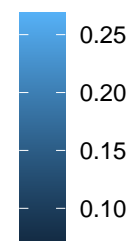
Infraction Ratio



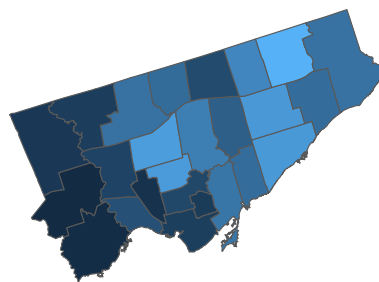
Significant Infractions



Infraction Ratio



Crucial Infractions



Infraction Ratio

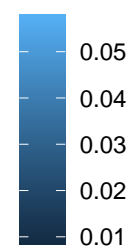


Figure 1: Infraction Percentage for Each Ward

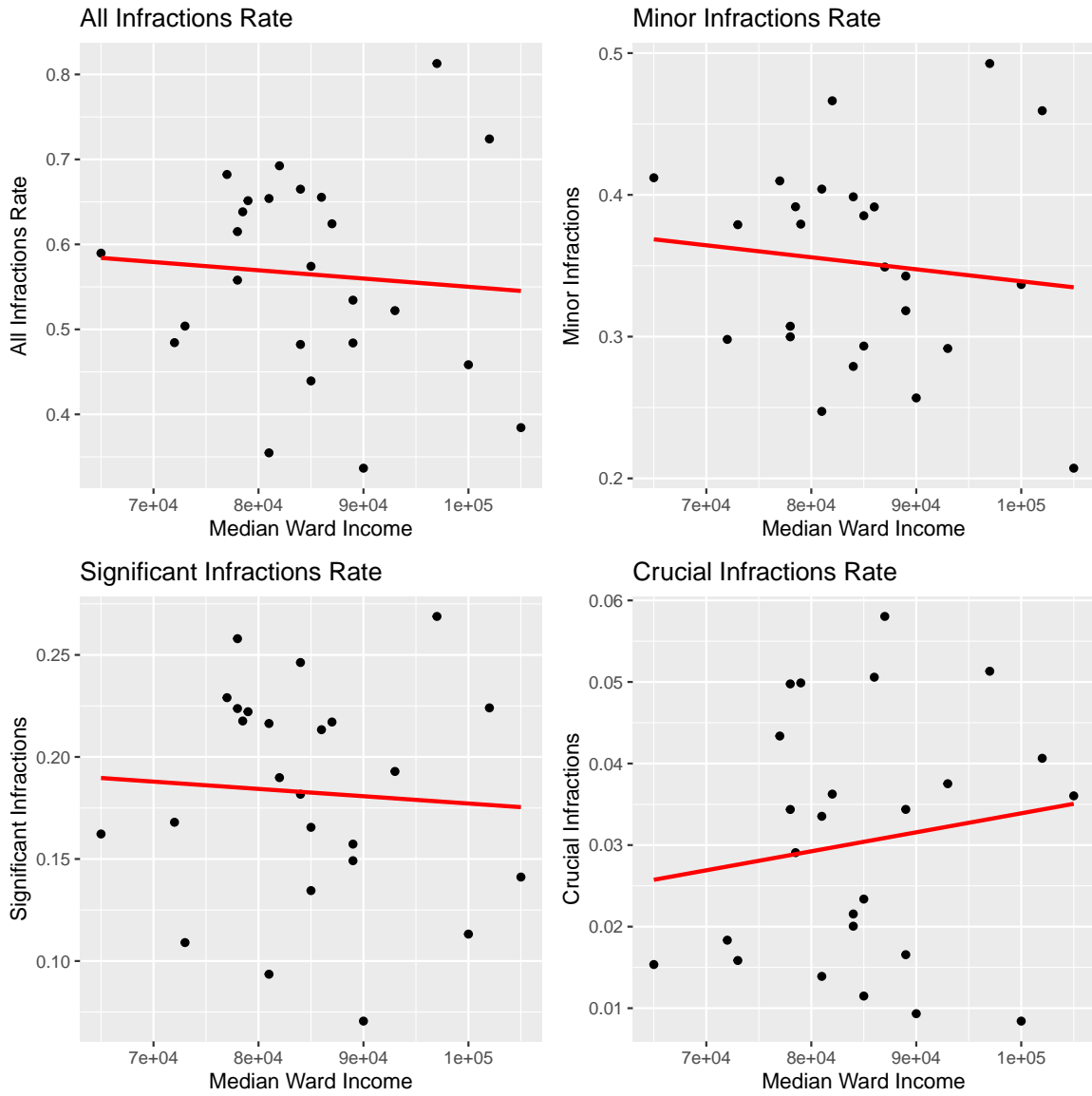


Figure 2: Median Ward Household Income Vs. Infractions Ratio Varying In Severity

## **4 Discussion**

### **4.1**

### **4.2 Second discussion point**

### **4.3 Third discussion point**

### **4.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.



## Appendix

### A Ward Map Visualization

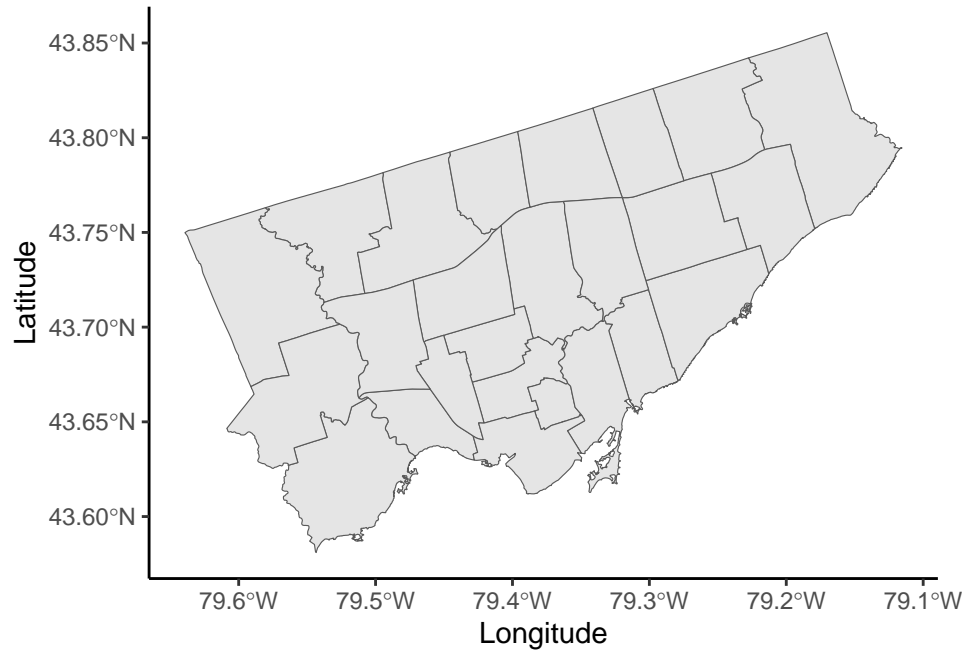


Figure 3: Visualization of Ward Map Data

### B Cleaned Data Sample

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

Figure 4: First Few Rows of the Cleaned Dataset

## C References