Food Safety Insights: A statistical Analysis behind the Safety Practices of Toronto Dining*

Aaron Xiaozhou Liu

April 17, 2024

This paper focuses on the DineSafe dataset from OpenDataToronto to explore the food safety inspections that took place in Toronto. The use of temporal and geospatial analysis helps us to find the reason behind the seasonal trends, spatial patterns, and hotspots of non-compliance. It also gives us an idea of the current status of food safety in the city. We also compare inspection results by facility types, isolate factors determining compliance, monitor rating variation trend, and analyze inconsistencies between official information and public perception data. The results of our research provide various beneficiaries, including policy makers, officials of public health, and foodservice industry stakeholders, with useful information that can be used to improve food safety regulations and guarantee secure dining experiences to residents of Toronto.

Table of contents

1 Introduction **3**

2 Data **4**
2.1 Data collection and cleaning 4
2.2 Interpreting the data 4
2.3 Variables of interest 4
2.3.1 Infraction Types 5
2.3.2 Infraction severity 5
2.3.3 Amount Fined 6

*Code and data are available at https://github.com/Pix3ls126/Toronto_food_safety

3 Model 6
3.1 Model set-up 8
3.1.1 Amount fined model 8
3.1.2 Inspection count model 8
3.2 Model Justification 8

4 Results 9

5 Discussion 10
5.1 Shortcomings and next steps 11

A Appendix 12
A.1 Datasheet 12

Bibliography 20

1 Introduction

Food safety is a crucial issue in the culinary scene of Toronto compared to other urban regions because it is a home to many restaurants that draw guests and local residents. While the diversity of restaurants in the city is indicative of it being a culturally luring place, concern with food quality and safety is also of equal significance. The food safety inspections, those are done by the officials of the Toronto Public Health department under DineSafe system (Market 2024), have great value as a tool of control over the compliance with health requirements and for protecting the public health. DineSafe represents one of the crucial strong points since the system works through OpenDataToronto which permits for students and researchers that want to learn the food environment of Toronto in more details to have access to such information.

This study will investigate the DineSafe data set, mainly focusing on the temporal and spatial trends associated with food safety inspections in different neighborhoods in Toronto. Through studying the trends of food safety compliance across years and locations, we intend to detect the efficiency of existing food safety systems and point out the areas for improvements. Furthermore, we will dig into the factors that may affect the outcome of an inspection, like establishment type, size, and frequency of inspection giving the in-depth picture of the determinants of compliance within the city's food service industry.

With the overarching goal of enhancing food safety standards and fostering transparency within Toronto's dining establishments, this research endeavors to provide evidence-based insights for policymakers, public health officials, and stakeholders. By analyzing the patterns emerging from DineSafe inspections, we aim to contribute to informed decision-making and the development of targeted interventions to address potential gaps in food safety regulation enforcement. Ultimately, our study seeks to promote a culture of accountability and consumer trust, ensuring that Toronto's vibrant culinary landscape remains a safe and enjoyable experience for all. In Section 2, we will cover the data and methods of analysis that will be used. We will visualize the results in Section 4, and finally interpret the results in Section 5.

2 Data

2.1 Data collection and cleaning

All data collection and analysis was done using statistical computing and data visualization program R (R Core Team 2023) and Rstudio IDE (RStudio Team 2020) to help streamline the workflow. The data used for this paper was found in the opendatatoronto library and downloaded using the R package opendatatoronto (Gelfand 2022). All the analysis was done using the R program and the following supporting packages tidyverse (Wickham et al. 2019), Janitor (Firke 2023), dplyer (Wickham et al. 2023), ggplot2 (Wickham 2016), knitr (Xie 2023), readr (Wickham, Hester, and Bryan 2024), arrow(Richardson et al. 2024), zoo(Zeileis and Grothendieck 2005), rstanarm(Goodrich et al. 2024), and here (Müller 2020). We will dive into more detail about the data collection, cleaning, and analysis in the following sections of the paper.

2.2 Interpreting the data

Once we had our data, obtained using the opendatatoronto (Gelfand 2022) R package. From this dataset, we see there are many columns of information. We will be ommiting some information, such as the longitude and latitude of different restaurants, as this information isn't relevant for most people, and using an address is much more practical when finding location. Other information such as establishment status and minimum inspection years are also omitted from this research, as the information they provide are relatively consistent among all establishments. The more important features of the dataset that can be potentially explored such as infraction details and infraction severity, which we will dive more in depth into exploring. Below is a table of all the attributes of the dataset and a brief explanation of each.

2.3 Variables of interest

Now that we have a better understanding of the information our dataset is providing us, we can better decide what variables we want to choose to further explore. For the purposes of this paper, we will be investigating infraction details, infraction severity, and establishment type. These variables will show what kind of infractions to food safety most establishments may more commonly commit, and among these different infractions, what kind of infractions are taken most seriously by DineSafe. Another variable that will be the inspection dates, where we will explore trends these inspections over time. Lastly, we will look at the area of establishments where the inspections take place and see if any conclusions can be drawn on hot spots of safety infractions.

2.3.1 Infraction Types

During an inspection by DineSafe, an establishment must satisfy the inspectors in the following categories ("Toronto Food Safety Regulations" 2024):

- Food temperature control
- Protecting food from contamination
- Maintenance and sanitation of food contact surfaces and equipment
- Maintenance and sanitation of non-food contact surfaces and equipment
- Maintenance and sanitation of washrooms
- Storage and removal of waste
- Pest control

Of these regulations, if the restaurant fails to meet any of the above requirements, them they are issued an infraction. Infractions have different levels of severity, which will be covered in the next subsection. Gathering data on this column would let us see what kinds of infractions happen more often than others, indicating what actions might need to be taken to help mitigate further occurrences of the same kind of infraction.

2.3.2 Infraction severity

Of the previously mentioned infractions, infractions are also labelled with a severity based on how bad the infraction is. This is split among three levels of severity, being minor, significant, and crucial infractions ("Toronto Food Safety Regulations" 2024).

Minor Infractions

- Pose minimal health risks, such as chefs not wearing hair ties.
- Can still receive an overall pass in the inspection
- Infraction must be addressed promptly.

Significant Infractions

- Violations that pose more significant health risks, such as dirty food surfaces
- Can receive a conditional pass, but must resolve the violation within 24-48 hours

Crucial Infractions

- Crucial infractions pose major health concerns such as food contamination and lack of safe water
- Must be dealt with immediately
- If left unchecked, the establishment will be issued a closure notice

Using this, we can get a general idea of what types of infractions are given what level of severity. We can use this information to draw on conclusions for what establishments should be aware of and make sure that they don't have any infractions that are considered severe enough to be issued a closure notice.

2.3.3 Amount Fined

The final variable of interest that will be explored is to take a look at the amount fined. We will investigate this with relation to the amount of infractions, as well as relate it to the severity of infractions. We expect to see some correlation between the severity of the infraction compared to amount fined, as more the more dangerous the infraction, the more an establishment would be fined due to negligence.

3 Model

Here, we will discuss the model used in this analysis. A simple Linear regression model is used where we will compare the the amount establishments are fined for infractions per month, and the total number of infractions per month.

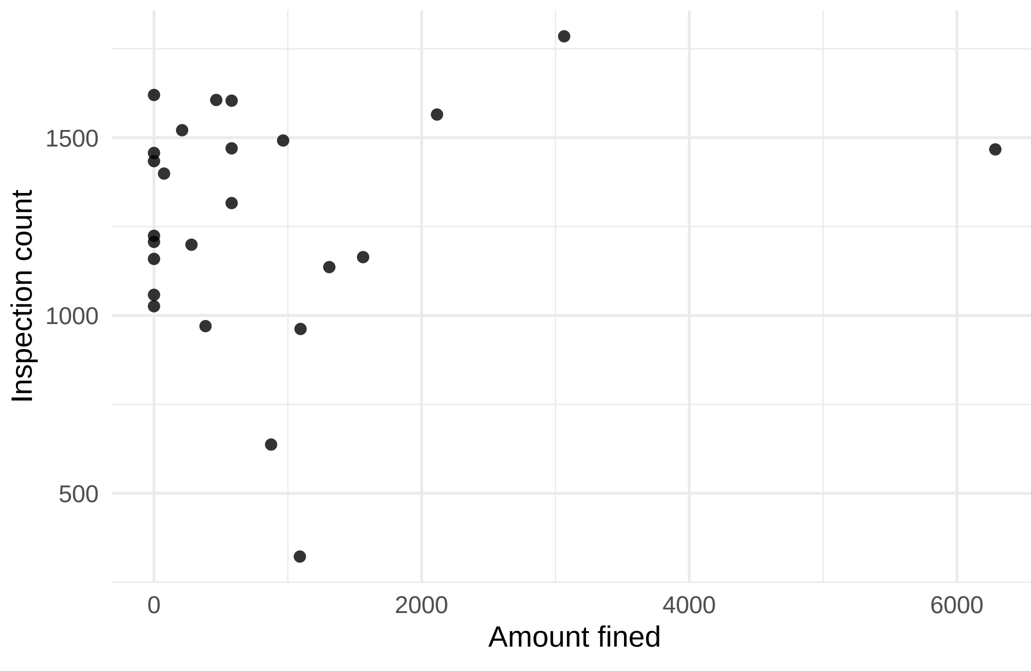

Figure 1: Amount fined relative to infraction count

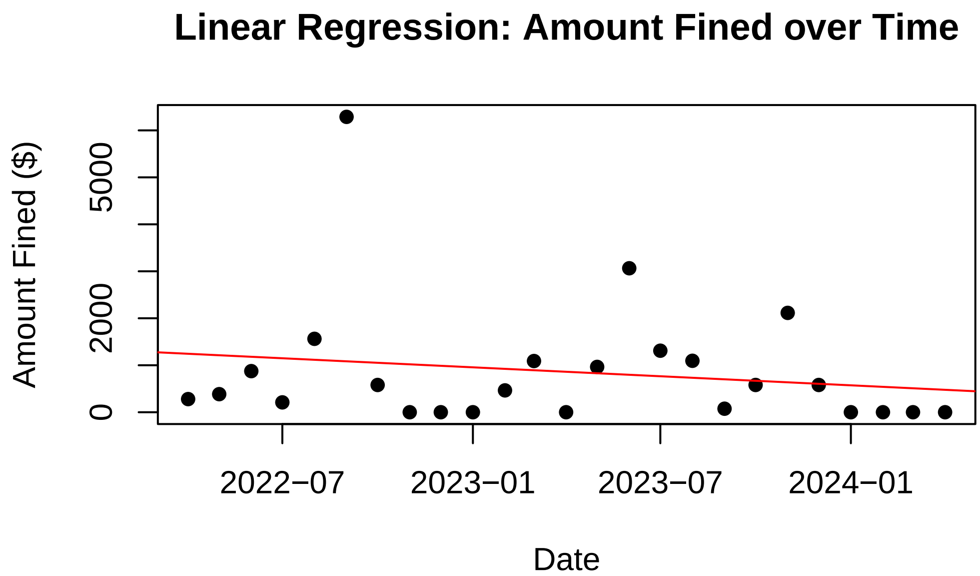

Figure 2: Amount fined over time

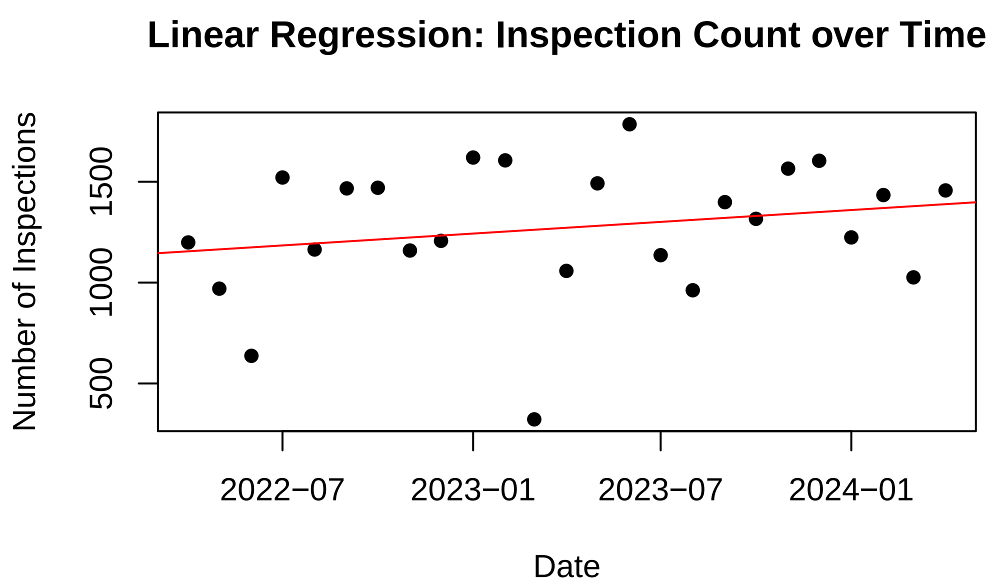

Figure 3: Infraction counts over time

3.1 Model set-up

For our model, we do two simple linear regressions. For the first one, we use date and the amount fined by DineSafe, and the second compared the total number of inspections conducted by DineSafe per month during the time period of our dataset.

3.1.1 Amount fined model

Using simple linear regression, we define the month y_i as the amount fined as the amount fined.

$$\begin{aligned} y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1) \\ \mu_i &= \beta_0 + \beta_1 \times \text{Date}_i + \epsilon && (2) \\ \beta_0 &\sim \text{Normal}(\mu_0, \sigma_0) && (3) \\ \beta_1 &\sim \text{Normal}(\mu_1, \sigma_1) && (4) \\ \sigma &\sim \text{Exponential}(\lambda) && (5) \end{aligned}$$

3.1.2 Inspection count model

Similar to the above, we can model the number of inspections by having y_i be the total number of inspections.

$$\begin{aligned} y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (6) \\ \mu_i &= \beta_0 + \beta_1 \times \text{Date}_i + \epsilon && (7) \\ \beta_0 &\sim \text{Normal}(\mu_0, \sigma_0) && (8) \\ \beta_1 &\sim \text{Normal}(\mu_1, \sigma_1) && (9) \\ \sigma &\sim \text{Exponential}(\lambda) && (10) \end{aligned}$$

3.2 Model Justification

Due to the nature of the dataset, as it is rather simple and there are limited amount of features to be graphed, a simple linear regression model would be the most appropriate model to use. We use this model to compare the dates with the features of amount fined and inspection counts, and find any relationship between these features. We expect to see that amount fined and inspections to both be relatively consistent over time, and the relation between amount fined is directly correlated with the number of inspections, as a higher number of inspections would mean a more infractions, which then lead to more fines.

4 Results

In Figure 1, we can see the plot of amount fined and inspection count, where we see that generally most cases dont result in fines, due there being a large number of points clustered at $0 in fines, even with over a thousand inspections being conducted. This suggests that there isn't any relation between the amount fined and the number of infractions per month.

Individually, we see that amount fined and inspection over time both remain relatively constant, with then only decreasing slightly or increasing slightly respectively. These can be observed in Figure 2, and Figure 3, where we see a downward trend in the total dollars fined per month from establishments, but we do see an increasing trend of total number of inspections per month conducted by DineSafe.

Some further analyses were conducted using other features of the dataset, which can be seen below in Figure 4

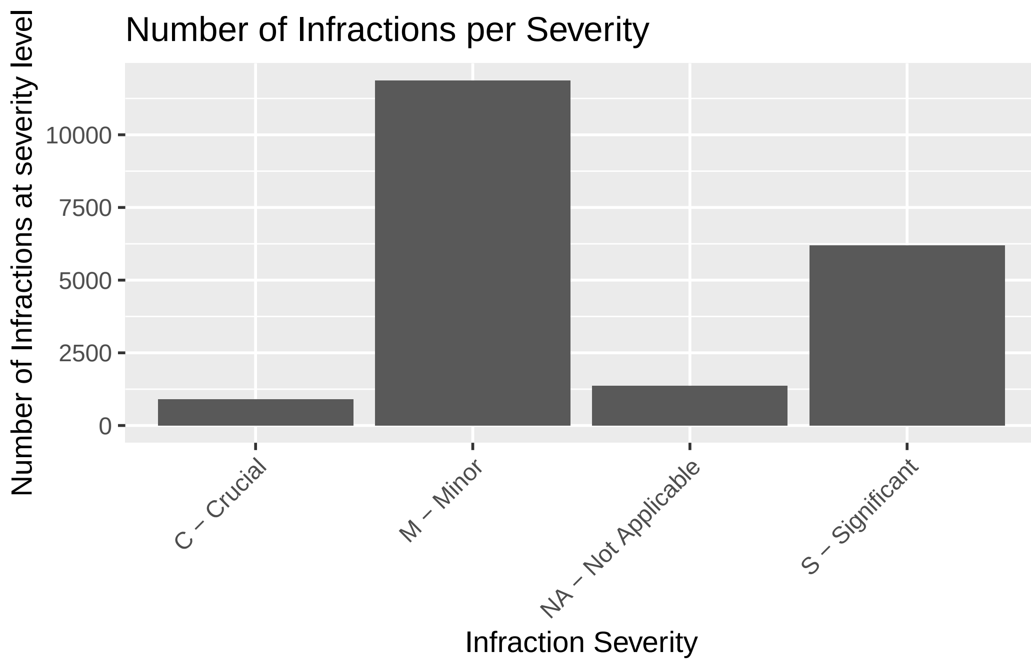

Figure 4: Inspections by establishment type

In the above graph, we see that the majority of all inspections result in infractions that are usually minor. However, we can also see that there is a large number of significant infractions as well. There are relatively few infractions that are given a crucial level of severity, and also a few that aren't classified under any severity level.

5 Discussion

From our discoveries in Figure 1, we can come to the conclusion that there isn't any relation between amount establishments are fined with how many inspections are found per month. We can see a cluster of points between a thousand to a thousand five hundred inspections, with many of them totaling zero dollars fined that month, while there are a few months with significantly less inspections conducted, while still totaling over a one thousand dollars in fines. This is likely because not all infractions will result in fines. Considering the our results in Figure 4, we see that the majority of infractions are usually just minor infractions. We know that minor infractions only usually result in a fine if the infraction is still ongoing in a re inspection, which usually takes place about twenty-four to forty-eight hours after the original ("Toronto Food Safety Regulations" 2024). This would mean that most establishments are quick to fix any broken regulations to ensure that any infractions that were found on the initial check are not present during the second round of inspection as to avoid having to pay any fines. Similarly with any infractions at a severity level of significant. Establishments are usually allowed to continue business with a conditional pass, and must pass a second inspection or face legal action ("Toronto Food Safety Regulations" 2024). As such, these locations would likely rectify these errors as quickly as possible to avoid and future troubles and pass the second round of checks.

In Figure 2, there are noticeable rises and falls in the graph, we usually see that there are months where there are no fines, but then total fines per month will slowly increase until it peaks and falls again. A reason for this is possibly due to these periods being around peak tourist travelling time. According to an article about travelling and visiting Toronto, the best times to visit Toronto are the periods of June to September and December to February, the former as it's the best time of year for travel, and the latter for lower prices. During these periods of time, restaurants would likely see a spike in customers. Due to the increase in business, it is possible for restaurant employees to unintentionally forget about the regulations, leading to them getting reported and receiving an infraction. Another issue would likely be that the hotter temperature in the summer leads to food spoiling faster. At temperatures between 90 degrees Fahrenheit, it is only safe to leave food out for one hour before throwing it away. These factors likely lead to more reports of poor food safety during the summer, and the spike in total fines we see in the plot diagram. Overall, we can see that the trend of total fines per month is on a downwards trend, meaning that food establishment safety is overall improving over time.

We can observe an overall increase in inspections conducted by DineSafe over the period of time in our plot, opposite of the trend we observe in the amount fined in Figure 2, where it is on a downward trend. It was expected for the trend to remain constant, as most restaurants have a minimum number of years before inspection, and all restaurants must be inspected to ensure food safety for customers. The period of time of this study, from April 2022, to 2024, is around the period of time when covid restrictions were relaxed in Ontario(Comartin and Ilunga 2022). Once the people were able to freely go out and eat at restaurants again, which were largely vacant during lock down with strict customer limits and people using delivery services instead of dining in, the resurgence of people dining in likely resulted in reports to DineSafe about food safety, as more and more people start dining in again and food safety must be taken more seriously.

5.1 Shortcomings and next steps

While the data set provided useful information for analysis, there was a lot of redundant information included that needed to be filtered out. Secondly, a lot of the information provided couldn't be utilized in the analysis. For example, while looking at some trends, such as location, it was hard to find a way to group establishments in certain areas, as the two features we had to identify a location were address, and longitude and latitude coordinates. Addresses were unique to each establishment, and since there are thousands of entries in the data set, it was impossible to graph this information. Lat/long on the other hand had a similar problem. While it's possible to strip each combination of coordinates to generalize the location, they do not describe where each establishment is, as the only way to find out what a long/lat coordinate is referring to is to go on a map and navigate to the location.

Another issue that arises with this data set is that there are not many features to use that can be graphed. Many of the features are string inputs, which means the only information we can graph about them is the number of occurrences within the data set. While this can still provide insights in the research, it is difficult to find relations between different features of the data set.

Some next steps to take would be to look for other data sets that could be related to the topic, for example a data set that goes into more details about restaurant situation, such as general location, such as if a restaurant is located within china town. Another variable of interest is the number of customers a restaurant sees per day, and seeing if there is any relation between the number of customers and infractions, or if the number of infractions potentially impacts a restaurants number of customers.

A Appendix

A.1 Datasheet

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
    - The dataset was created for the purposes of analyzing how restaurants fair with regards to food safety inspections. The dataset was acquired from OpenDataToronto, and the goal is to examine trends within the results of these safety inspections done by DineSafe.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
    - Aaron Liu, a student at the University of Toronto.
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
    - No funding was received
4. *Any other comments?*
    - No

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
    - The contents of the dataset are the restaurants that were evaluated by DineSafe and the results of that inspection, listing the date, location, and infraction types.
2. *How many instances are there in total (of each type, if appropriate)?*
    - About 20000 instances
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

- The dataset is a cumulation of all the inspections done by DineSafe in restaurants across the city of Toronto, and as such, is representative of the larger set.

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

   - The instances consist of the results of the inspection

5. *Is there a label or target associated with each instance? If so, please provide a description.*

   - Yes, there are unique inspection IDs

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

   - No

7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*

   - Yes, certain instances have same inspection IDs

8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*

   - No

9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

   - No

10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

    - The data is self contained

11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*

- No, the data which the dataset is based off of is publicly available on OpenData-Toronto

12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*

    - No

13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*

    - No

14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*

    - No

15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*

    - No

16. *Any other comments?*

    - No

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

    - The data was taken from OpenDataToronto using the OpenDataToronto R package

2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*

    - Downloaded using R and the OpenDataToronto package

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
   - The dataset is the whole set.
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
   - Employees of DineSafe
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
   - Over the period of 2024 to 2024
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
   - No
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
   - Obtained via third-party
8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*
   - No
9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
   - No
10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
    - Consent was not needed as the data was openly available

11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

    - No

12. *Any other comments?*

    - No

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

    - Data was cleaned, no missing values needed processing. Selected all rows that had infractions

2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*

    - Yes

3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

    - R was the only software used

4. *Any other comments?*

    - No

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*

    - Unsure

2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*

    - no

3. *What (other) tasks could the dataset be used for?*

    - Other types of analysis can be done on the dataset other than what was done in this paper

4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
   - No
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
   - No
6. *Any other comments?*
   - No

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
   - Available on github
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
   - Github
3. *When will the dataset be distributed?*
   - The data is already available for use on github
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
   - No
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
   - No

6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*

    - No

7. *Any other comments?*

    - No

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*

    - Aaron Liu

2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*

    - As the following email: aaronxiaozhou.liu@mail.utoronto.ca

3. *Is there an erratum? If so, please provide a link or other access point.*

    - No

4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*

    - Not likely

5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*

    - No

6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*

    - No

7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*

    - Fork the repository on github

8. *Any other comments?*

    - No

Bibliography

Comartin, Michael, and Gloria Ilunga. 2022. "Ontario Provides Timeline to Lift All COVID-19–Related Restrictions." https://ogletree.com/insights-resources/blog-posts/ontario-provides-timeline-to-lift-all-covid-19-related-restrictions/.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

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

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Market, Food Safety. 2024. "DineSafe." https://www.toronto.ca/community-people/health-wellness-care/health-programs-advice/food-safety/dinesafe/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. http://www.rstudio.com/.

"Toronto Food Safety Regulations." 2024. https://www.foodsafetymarket.com/toronto-food-safety-regulations.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

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, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

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

Zeileis, Achim, and Gabor Grothendieck. 2005. "Zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. https://doi.org/10.18637/jss.v014.i06.