

Understanding Factors Contributing to DineSafe Violation Severity*

Utilizing multinomial logistic regression to visualize and understand components from June 2022 to June 2024 DineSafe data

Michelle Ji

November 19, 2024

This study aims to explore the relationship between various factors, such as minimum inspections per year and type of food establishment, and the severity of DineSafe violations in Toronto. Inspection data from June 2022 to June 2024 were compiled and analyzed through a multinomial logistic regression model. Key findings include specific types of food establishments, such as formal dining and fast food, and increasing the number of minimum inspections had higher likelihoods of a more severe infraction when compared to a baseline “minor” violation. Implications from this study can assist to raise awareness among both consumers and business owners, potentially leading to stricter adherence to health guidelines or policy changes aimed at improving and maintaining public health.

1 Introduction

Various public health measures have recently been put into place as a means to ensure public safety and sanitation. Specific to Toronto, DineSafe is Toronto Public Health’s food inspection program and is based on provincial and municipal regulations. Each establishment that processes, manufactures, prepares, stores, or serves food must be registered for inspection. As such, the inspection process and frequency will likely depend on the type of food establishment and the level of risk associated with the food handled. Each inspection follows a set of guidelines, often assessing the food temperature control, contamination, pest control, employee hygiene, maintenance of sanitation in public areas, and waste disposal (Food Safety Market 2024). Infraction levels are categorized by “minor”, “crucial”, and “significant”, in increasing severity order. Depending on the infraction level, results can vary from a verbal notice to closure of the establishment (Eastern Ontario Health Unit 2024). This paper examines the

*Code and data are available at: <https://github.com/michj3/dinesafe-toronto>

factors influencing the severity of DineSafe violations, addressing a gap in public knowledge to provide valuable insights for food establishments and policymakers. By understanding these factors, targeted actions can be taken to enhance public health outcomes for the citizens of Toronto.

The estimand in the paper is the relationship between factors such as the minimum number of inspections per year and the type of food establishment on the severity of a DineSafe violation. By analyzing the dataset and developing a multinomial logistic regression model, the primary objective is to identify the factors that influence the severity levels of DineSafe violations. This knowledge can raise awareness and guide employers, employees and policymakers to adopt improved habits or implement changes in policies. One main finding is that specific food establishments, such as formal dining, fast food, and industrial food preparation places, had a higher likelihood of having a “crucial” violation over a “minor” violation. This can lead to serious consequences, such as an increased likelihood of establishment closure. Therefore, it is crucial for business owners to be aware of these factors so they can take proactive measures to address potential issues and violations.

This paper is organized further into five main sections. Section 2 discusses how the dataset was measured and important variables of interest. Section 3 describes the model used for analysis and potential weaknesses. Section 4 explores the main findings of the model and Section 5 highlights real world implications and possible future courses of action. Section A discusses more in-depth how the data was cleaned for modeling and a discussion on observational data in this dataset, emphasizing its significance and the implications of the findings. Section A also includes a datasheet for the dataset.

2 Data

2.1 Measurement

The dataset was obtained through the City of Toronto’s OpenData Library (Gelfand 2022) and is titled “DineSafe” (OpenDataToronto 2024). As part of Toronto Public Health’s food safety program, DineSafe is a program that inspects all establishments preparing and serving food; these types of establishments range from food trucks to sit-down restaurants. Municipal health departments conduct inspections that are either routine, complaint-driven, or follow-up checks. Establishments will be evaluated on predefined criteria, such as cleanliness, food handling, and storage. Each inspection will result in a pass, conditional pass, or closed notice depending on the severity level of any violations and courses of action. Results, including the number and type of violations, are logged into Toronto’s municipal database, reflected in the DineSafe dataset through OpenDataToronto.

The dataset includes data containing inspection details of food establishments from June 2022 to June 2024. There were no other similar datasets. Table 1 details the first few observations of the cleaned dataset for analysis.

Table 1: First Few Observations of the DineSafe Dataset

Establishment Type	Minimum Inspections	Severity	Action
Industrial	2	M - Minor	Notice to Comply
Industrial	2	M - Minor	Notice to Comply
Fast Food	3	M - Minor	Notice to Comply
Fast Food	3	C - Crucial	Notice to Comply
Fast Food	3	M - Minor	Notice to Comply
Fast Food	3	S - Significant	Notice to Comply

The dataset was cleaned analyzed using the statistical programming software R and packages such as tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016), kable (Zhu 2021), opendata-toronto (Gelfand 2022), dplyr (Wickham et al. 2023), arrow (Richardson et al. 2024), vgam (Yee 2024), testthat (Wickham 2011), car (Fox and Weisberg 2019), and broom (Robinson, Hayes, and Couch 2024).

2.2 Variables

The analysis dataset is focused on 1 response variable, severity level, and 2 predictor variables, the minimum number of inspections per year and the food establishment type.

The severity level variable is a categorical variable with three levels: “C” for crucial, “S” for significant, and “M” for minor. The most severe infraction is “crucial” and the least severe is “minor”. Each level corresponds to the level of sanitation violation that the food establishment made and will correspond with different courses of action and outcome. Figure 1 details the distribution of the different infraction levels. From Figure 1, the most frequent type of infraction level is minor and crucial infractions of regulations are the least frequent. From a public health perspective, this is a positive sign, as it suggests that most food establishments maintain relatively high standards of sanitation. This is reassuring for consumers, as it indicates that food establishments are generally safe and comply with health regulations.

The variable focusing on the minimum number of inspections per year is a quantitative variable that ranges from 1 to 3 visits per year. Typically, the number of inspections will vary depending on the specific type of establishment, the food preparation processes, volume and type of food served and other related criteria (Toronto Public Health 2024). Figure 2 details the distribution of the number of inspections per year the food establishments in the dataset receive. Figure 2 highlights that most commonly, food establishments in Toronto receive around 3 inspections per year and rarely will receive only 1. This may relate to the complexity of certain restaurants or food establishments, which often require more thorough inspections due to multiple components, such as kitchen operations, food preparation, and serving processes. Generally, and rarely, only very basic establishments receive a single inspection per year.

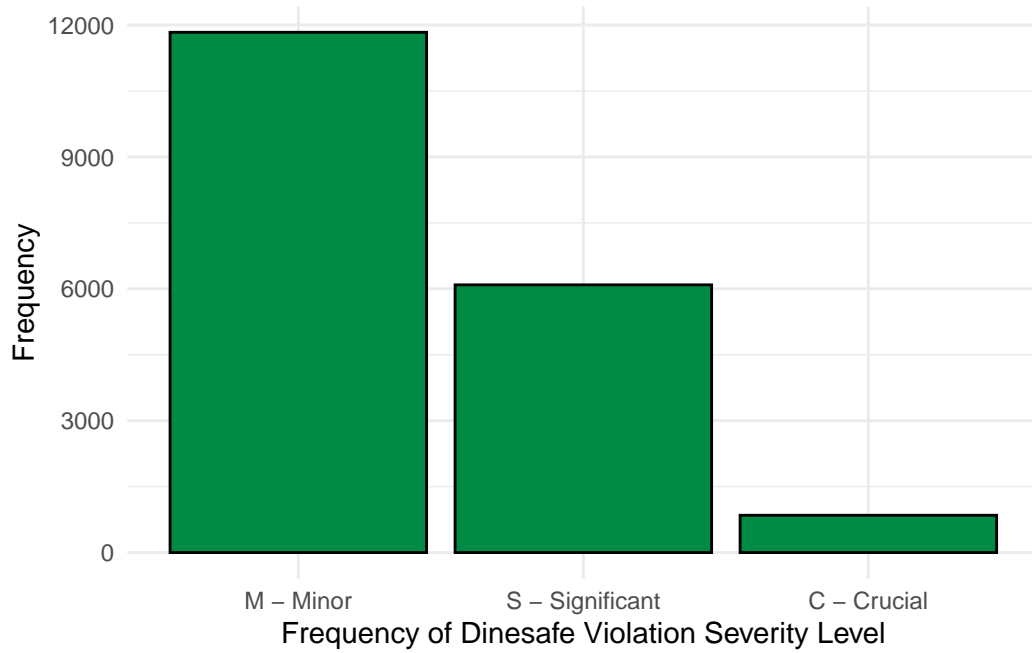


Figure 1: Distribution of severity infraction level

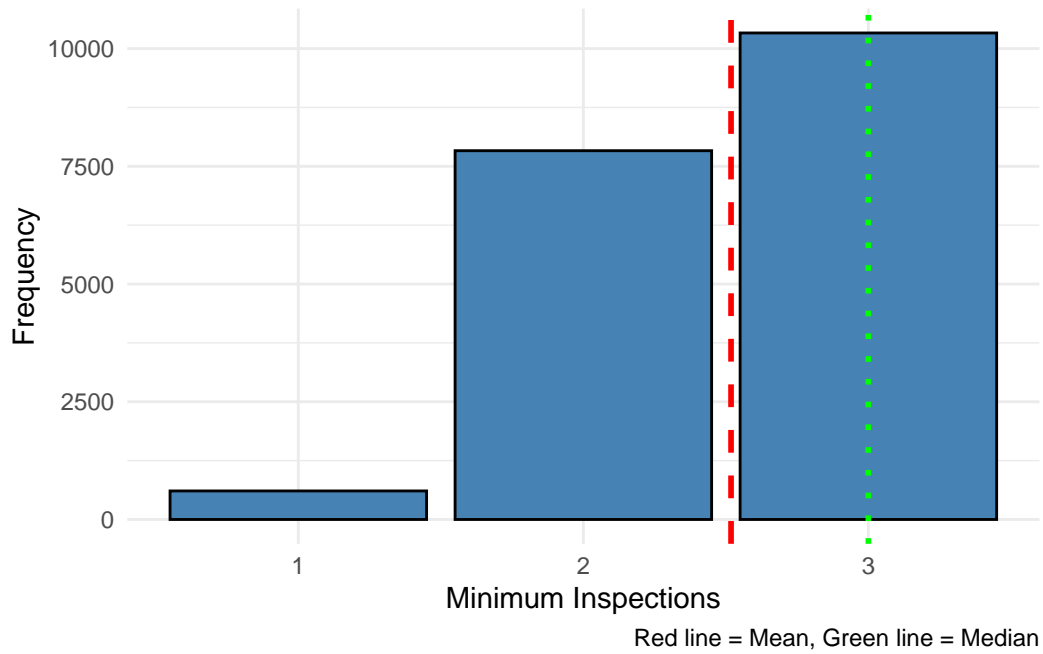


Figure 2: Distribution of minimum inspections per year, with mean and median.

Table 2 details the mean and median number of inspections of Toronto food establishment inspections. The variable focusing on food establishment type is a qualitative variable that contains the types of food establishments inspected, which include food trucks, serving kitchens, bar and cocktail lounges, and restaurants, to name a few.

Table 2: Mean and Median Number of Inspections by Establishment Type

Establishment Type	Mean Inspections	Median Inspections
All Types	2.518193	3

The variable focusing on food establishment type is a qualitative variable that contains the types of food establishments inspected, which include food trucks, serving kitchens, bar and cocktail lounges, and restaurants, to name a few. Figure 3 highlights the top ten most common food establishments in the dataset (OpenDataToronto 2024) that have been inspected. The most common types are traditional restaurants, food take out services, and supermarkets that serve and prepare food. It is important to note this, as we will later examine how different types of food establishments may correlate with the severity level of their violations.

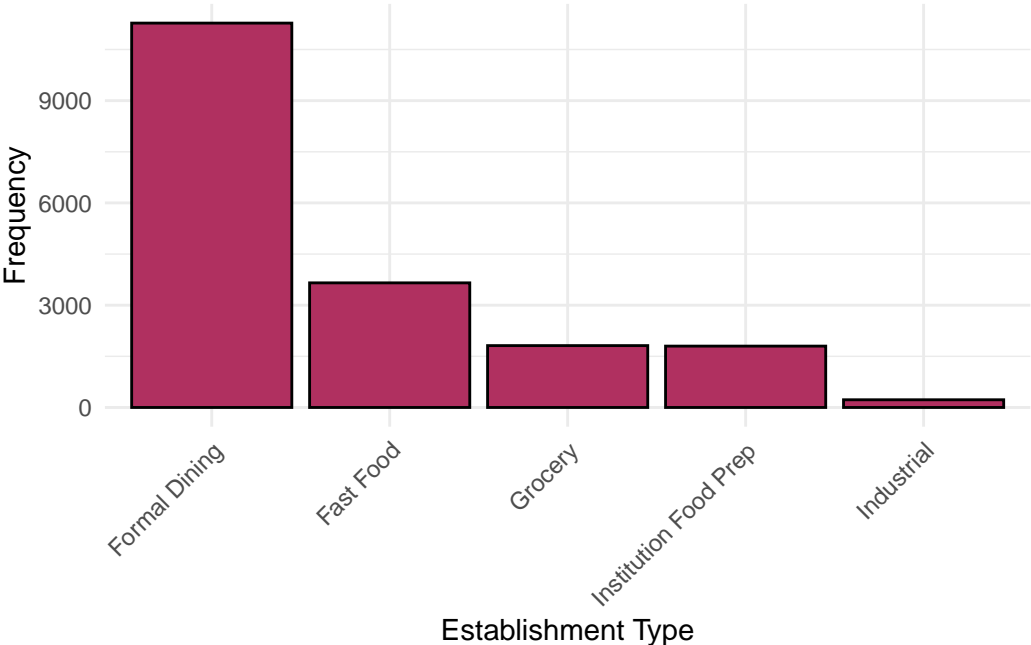


Figure 3: Top 10 Most Frequent Food Establishment Types

3 Model

3.1 Modeling

A multinomial logistic regression model was used to determine if the severity of a DineSafe violation had a relationship with the type of food establishment and minimum number of inspections per year an establishment receives. A multinomial logistic regression model is used on nominal outcome variables, in which the log odds of the outcomes are modeled as a linear combination of the predictor variables (UCLA Statistics 2024). Since the severity variable has multiple levels and is not dichotomous, a multinomial logistic regression model is used instead of a binary logistic model. Coefficients correspond to the log-odds of being in a specific category (either “C - Crucial” or “S - Significant”) relative to the reference category (“M - Minor”). Log odds is a way to describe the likelihood of an event happening; in this case, the event would be a different severity level relative to the “minor” severity level that uses a logarithmic scale.

The labels of “1” and “2” next to each of the predictor variables indicates what level severity we are evaluating the model at compared to the reference level, “minor”. “1” indicates comparing “crucial” to “minor”, whereas “2” indicates comparing “significant” to “minor”.

The multinomial logistic regression model used has two different levels and is as follows:

When minor is compared to crucial:

$$\log \left(\frac{P(\text{Severity} = \text{Significant})}{P(\text{Severity} = \text{Minor})} \right) = \beta_{0,1} + \beta_{1,1} \times \text{min inspections} + \quad (1)$$

$$\beta_{2,1} \times \text{establishmenttype:Fast Food} + \quad (2)$$

$$\beta_{3,1} \times \text{establishmenttype:Formal Dining} + \quad (3)$$

$$\beta_{4,1} \times \text{establishmenttype:Grocery} + \quad (4)$$

$$\beta_{5,1} \times \text{establishmenttype:Industrial} + \quad (5)$$

$$\beta_{6,1} \times \text{establishmenttype:Institution Food Prep} \quad (6)$$

- β_0 : The intercept of the baseline log-odds of being in the “Crucial” category relative to the “Minor” category of violation when all predictors are at their baseline or zero.
- β_1 : The coefficient corresponding to the change in the log-odds of an establishment being in the “Crucial” category, compared to the “Minor” category, for a one-unit increase in `min_inspections`.
- β_2 : The coefficient representing the change in the log-odds of an establishment being classified as “Crucial” (versus “Minor”) when the establishment is of type “Fast Food”.

- β_3 : The coefficient representing the change in the log-odds of an establishment being classified as “Crucial” (versus “Minor”) when the establishment is of type “Formal Dining”.
- β_4 : The coefficient representing the change in the log-odds of an establishment being classified as “Crucial” (versus “Minor”) when the establishment is of type “Grocery”.
- β_5 : The coefficient representing the change in the log-odds of an establishment being classified as “Crucial” (versus “Minor”) when the establishment is of type “Industrial”.
- β_6 : The coefficient representing the change in the log-odds of an establishment being classified as “Crucial” (versus “Minor”) when the establishment is of type “Institution”.

And when “significant” severity is compared to “minor”:

$$\log \left(\frac{P(\text{Severity} = \text{Significant})}{P(\text{Severity} = \text{Minor})} \right) = \beta_{0,2} + \beta_{1,2} \times \text{min inspections} + \quad (7)$$

$$\beta_{2,2} \times \text{establishmenttype:Fast Food} + \quad (8)$$

$$\beta_{3,2} \times \text{establishmenttype:Formal Dining} + \quad (9)$$

$$\beta_{4,2} \times \text{establishmenttype:Grocery} + \quad (10)$$

$$\beta_{5,2} \times \text{establishmenttype:Industrial} + \quad (11)$$

$$\beta_{6,2} \times \text{establishmenttype:Institution Food Prep} \quad (12)$$

- β_0 : The intercept of the baseline log-odds of being in the “Significant” category relative to the “Minor” violation category when all predictors are at their baseline or zero.
- β_1 : The coefficient corresponding to the change in the log-odds of an establishment being in the “Significant” category, compared to the “Minor” category, for a one-unit increase in `min_inspections`.
- β_2 : The coefficient that represents the change in the log-odds of an establishment being classified as “Significant” (versus “Minor”) when the establishment is of type “Fast Food”.
- β_3 : The coefficient that represents the change in the log-odds of an establishment being classified as “Significant” (versus “Minor”) when the establishment is of type “Formal Dining”.
- β_4 : The coefficient that represents the change in the log-odds of an establishment being classified as “Significant” (versus “Minor”) when the establishment is of type “Grocery”.
- β_5 : The coefficient that represents the change in the log-odds of an establishment being classified as “Significant” (versus “Minor”) when the establishment is of type “Industrial”.

- β_6 : The coefficient that represents the change in the log-odds of an establishment being classified as “Significant” (versus “Minor”) when the establishment is of type “Institution”.

For the purposes of the model’s simplicity, food establishments were categorized into 5 levels: fast food, formal dining, grocery, industrial, and institution food establishments. Without categorization, there would be too many levels of this categorical variable, making it challenging to draw meaningful conclusions within the context of the DineSafe program. Specifics are in Section A.1.

3.2 Model Validation

To check the validity of the model used, it’s important to check for independence of observations, linearity in log-odds, multicollinearity, and sufficient sample size. For independence of observations, independent observations are assumed since each observation in the dataset is a separate restaurant or individual violation of DineSafe. They are treated as separate violations or inspections and are unaffected by other inspections. One inspection’s outcome does not influence another establishment’s outcome and each inspection for the same establishment is treated as a stand-alone event or observation. To look at linearity in log-odds, we can use the numerical predictor variable, the minimum number of inspections per year for an establishment, and check the linearity of the relationship between that and the log-odds of the predicted probabilities from the multinomial model. This is shown in Figure 4, where there is a general linear relationship between the two variables.

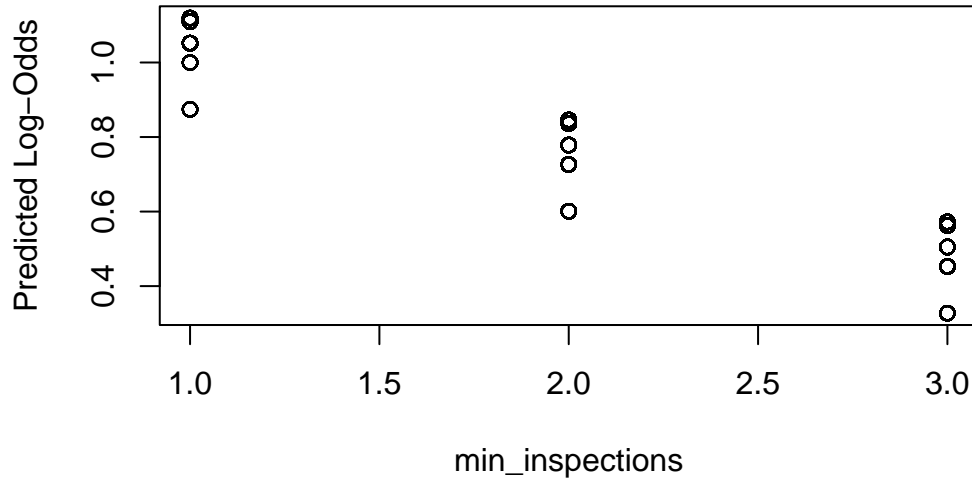


Figure 4: Linearity Check for Log-Odds

One thing to note is the distinct observations that form almost vertical lines. This should not be of concern, as there are only three possible values of minimum inspections and are integers: 1, 2, and 3. The overall trend of this plot is linear, which we can conclude this assumption

holds. To check for multicollinearity, we need to first convert the food establishment type variable into a dummy variable, holding it as a “numerical” value instead of categorical. Then, we can use the `cor()` function to look at correlation and thus, multicollinearity. The values of correlation between the variables are in Figure 8 in Section A.5. The variables do not seem to have high correlation with each other, indicating that our model likely does not have severe multicollinearity and its conclusions can be treated as valid. Finally, it’s important to note that for a model to be valid, there needs to be sufficient sample size. Our model uses a dataset (OpenDataToronto 2024) that has over 18,000 observations, so sample size is not a concern for model validity.

3.3 Model Limitations

Some limitations exist with the model used. First, the DineSafe program does not reveal how they make their inspections and if it is geographically coordinated. If certain types of establishments are clustered geographically and DineSafe inspectors inspect these clusters at the same time, the independence assumption might be violated, potentially biasing results. Additionally, since there are 46 different food establishment types in the dataset, categorizing them into 5 main categories made the modeling simpler and easier to draw conclusions. While categorizing establishment type helps manage the complexity, there may still be too many levels relative to the sample size for each category. This can lead to overfitting, especially if some categories have very few observations, reducing the model’s generalizability to new data. Finally, the model is built on data from the DineSafe program, which is specific to Toronto. Its findings may not generalize well to other contexts or geographic areas with different inspection practices or severity grading systems.

Other potential models include ordinal logistic regression or multilevel logistic regression. Ordinal logistic regression could be useful, as it accounts for the ordering of the response variable, but since there are multiple levels in the complex predictors, it may limit the model’s flexibility and interpretability. As for multilevel logistic regression, this model is useful when there is nested data, such as regional inspections that are clustered together. However, in the dataset, there is no geographical location for each observation. Additionally, multilevel models increase complexity and interpretive difficulty. For the purpose of drawing important public health conclusions for the DineSafe program, the multinomial model may draw clearer conclusions without the unnecessary complexity. The multinomial logistic regression model is a balanced choice, providing interpretable results and reasonable predictive power while accommodating categorical data, aligning with the needs of the DineSafe program.

4 Results

Model results are summarized in Table 3. The labels of “x 1” and “x 2” next to each of the predictor variables indicates what level severity we are evaluating the model at compared to

the reference level, “minor”. “x 1” indicates comparing “crucial” to “minor”, whereas “x 2” indicates comparing “significant” to “minor”.

Figure 5 illustrates the coefficients and their corresponding 95% confidence intervals for the predictor variables in the multinomial logistic regression model as a forest plot. A 95% confidence interval (CI) is a statistical range that is likely to contain the true value of a population parameter, in this case the regression coefficient, with 95% certainty. The forest plot allows visualization of the odds ratios with the confidence intervals, making it easier to interpret which predictors have a significant impact across different outcome classes. If the confidence interval for a coefficient includes 0, or the p-value is greater than 0.05, it suggests that there is no statistically significant effect of that predictor (establishment type or minimum inspections) on the response variable (severity) at the chosen level of significance of 0.05.

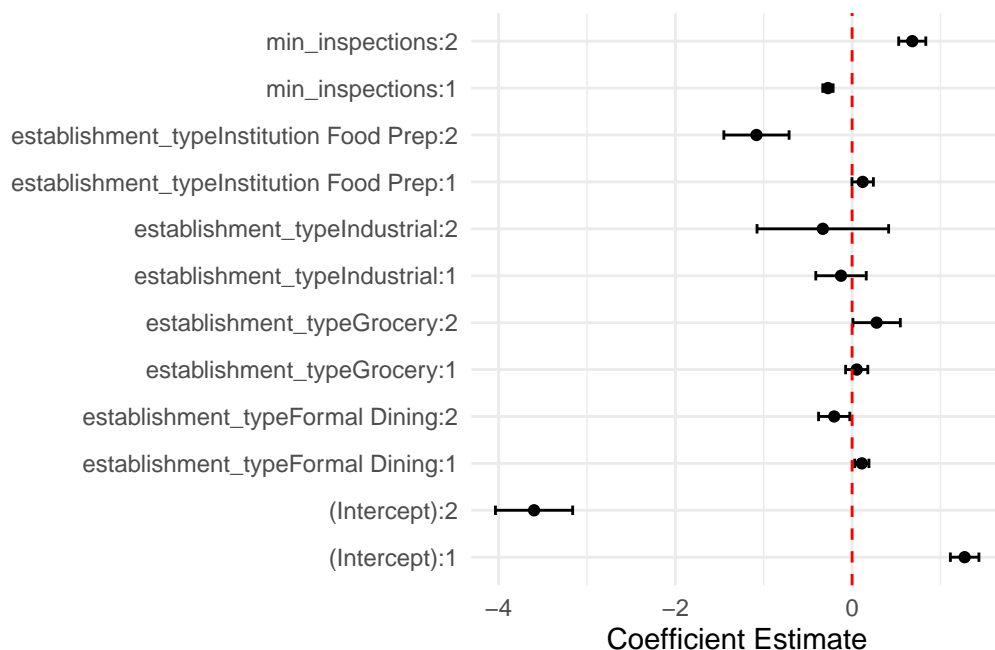


Figure 5: Forest Plot of Coefficient Estimates with 95% Confidence

From the Figure 5 and Table 3, the effect of minimum inspections per year and type of food establishment on severity level of a DineSafe violation can be analyzed. Analysis on the effect that minimum inspections per year has on severity has two levels. The coefficient for at the “crucial” level is -0.27228 with a very low p-value ($< 2e-16$), suggesting that as the number of minimum inspections increases, the log-odds of an establishment being classified as “crucial” compared to “minor” decrease significantly. This implies that higher minimum inspections are less likely to result in a “crucial” rating. When “significant” level is compared to the “minor” level, the coefficient 0.6808 with a very low p-value ($< 2e-16$), indicating that higher minimum inspections increase the log-odds of an establishment being classified as “significant” compared to “minor”.

Table 3: Multinomial Logistic Regression Model Summary

	(1)
(Intercept) \times 1	1.273 (0.082)
(Intercept) \times 2	-3.599 (0.223)
min_inspections \times 1	-0.273 (0.030)
min_inspections \times 2	0.681 (0.078)
establishment_typeFormal Dining \times 1	0.110 (0.041)
establishment_typeFormal Dining \times 2	-0.203 (0.090)
establishment_typeGrocery \times 1	0.052 (0.064)
establishment_typeGrocery \times 2	0.277 (0.137)
establishment_typeIndustrial \times 1	-0.126 (0.146)
establishment_typeIndustrial \times 2	-0.331 (0.380)
establishment_typeInstitution Food Prep \times 1	0.119 (0.062)
establishment_typeInstitution Food Prep \times 2	-1.082 (0.188)

Looking at formal dining establishments, the coefficient (2.1726) is statistically significant for “crucial”, suggesting that being classified as “formal dining” increases the log-odds of an establishment being rated crucial compared to minor. Institutional food preparation establishments have higher odds of being rated “crucial” relative to “minor” and is statistically significant, with a p value of 0.0466. Additionally, institutional food preparation establishments have statistically significant lower log-odds of being rated “significant” relative to “minor” (p value = $9e-09$). There are also a few marginally statistically significant differences. For example, grocery establishments have higher log-odds of being rated “crucial” relative to “minor” (p=0.0538). Fast food establishments also have marginally higher log-odds of being rated “crucial” compared to “minor” (p=0.0594).

Ultimately, model analysis shows two major results. First, the number of minimum inspections has a significant effect on the severity of a DineSafe violation, as the more inspections per year, the more likely a “significant” violation will occur, which is depicted by Figure 6.

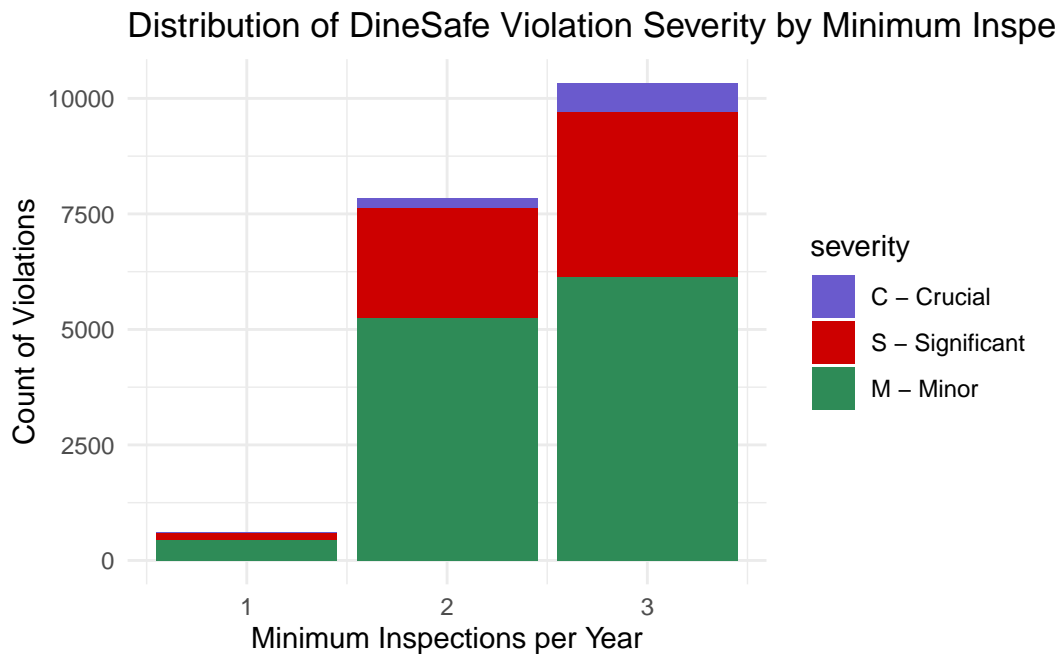


Figure 6: Forest Plot of Coefficient Estimates with 95% Confidence

Second, the type of food establishment also influences the severity level. Specifically, formal dining, fast food, institutional food preparation and grocery establishments have a higher likelihood of being rated as “crucial” than “minor”. This can be seen in Figure 7, where formal dining and institutional food preparation establishments have higher probabilities of having a “crucial” violation.

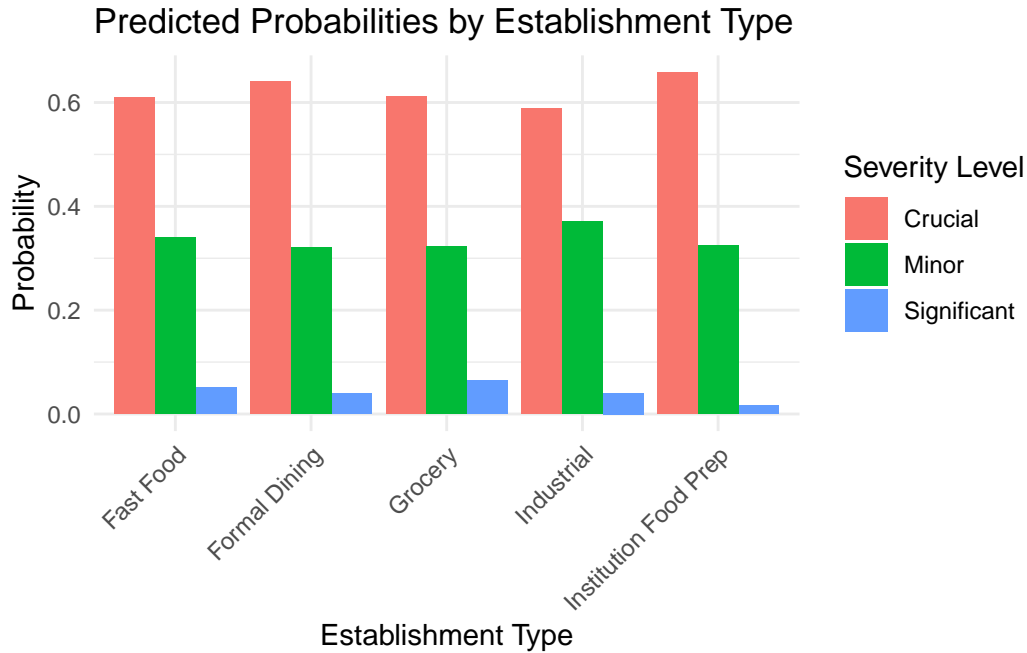


Figure 7: Predicted Probabilities by Establishment Type

5 Discussion

5.1 Minimum inspections per year and its association with severity of violation

From the model analysis, it is evident that a greater number of minimum inspections per year is associated with a higher likelihood of a significant violation. This finding carries several important implications for both public health and business operations. First, it can encourage business owners to minimize violations, as each follow-up inspection counts towards the total number of inspections. An increase in inspections or violations may result in the closure of an establishment, which not only negatively affects the business but also represents a loss of investment for the stakeholders involved. Furthermore, since businesses are required to display their DineSafe inspection results publicly, frequent negative changes in the rating can deter customers, leading to reduced patronage and further financial loss. Additionally, this insight may prompt businesses to implement more stringent sanitation practices and provide enhanced public health training for employees. Although these measures may be short-term costs, they can prove to be beneficial in the long run if they help prevent violations that can result in higher costs, thereby maintaining a good standing in public health inspections. From a public health perspective, the findings highlight the importance of motivating employees to adhere to health guidelines and sanitation protocols, as this not only helps reduce the likelihood of severe violations but also fosters a healthier work environment. This approach can contribute to more consistent compliance with health regulations, benefiting both employees and the business as

a whole.

5.2 Type of Establishment vs. Severity of Violation

The model suggests that formal dining, fast food, and institutional food preparation establishments are more likely to receive “crucial” ratings when compared to “minor” violations. For establishment owners and employees, this finding is particularly significant, as it highlights the importance of maintaining high standards of sanitation in these types of establishments. These establishments typically handle more raw food, such as meats and fresh produce, which increases the potential for foodborne contamination. Therefore, strict adherence to public health guidelines and proper food preparation practices is crucial in ensuring the safety of both employees and customers.

Improper handling of raw foods can lead to the spread of foodborne illnesses, such as salmonella and E. coli, which pose significant public health risks. A disease outbreak traced back to an establishment can severely damage its reputation, leading to public distrust and a decline in customers. The public health consequences are profound, but so are the business repercussions, potentially resulting in a loss of customers and revenue. Given these risks, employees in these establishments should be especially strict about sanitation practices and adhere to established health protocols. Awareness of the potential severity of violations may encourage staff to prioritize hygiene and food safety. Additionally, this information can drive the development of more vigorous employee training programs focused on sanitation practices and food safety. Not only will these programs help prevent severe violations and improve public health outcomes, but they will also positively impact business operations. Maintaining a strong DineSafe rating will be visible to the public, and establishments with higher ratings are likely to attract more customers, improving their overall business success. Therefore, investing in better sanitation practices is beneficial for both public health and long-term business sustainability.

The findings not only have implications for business owners but also for the inspectors responsible for monitoring food safety. Given the identified association between certain types of establishments—such as formal dining, fast food, and institutional food preparation—and a higher likelihood of receiving severe DineSafe violations, inspectors can use this information to adopt a more strict approach when inspecting these establishments. By recognizing that these types of businesses are more prone to severe violations, inspectors can focus their attention more closely on areas of higher risk, such as the handling of raw meats and produce, food storage, and sanitation practices. More rigorous inspection process not only minimizes the chances of a serious health violation occurring but also strengthens the overall public health framework by ensuring that establishments are held accountable for maintaining high standards of cleanliness and food safety. This can reinforce trust in the regulatory system and ensure that the DineSafe program remains a reliable indicator of a restaurant’s commitment to health and safety.

5.3 Weaknesses and Future Implications

While the results provide valuable insights for addressing public health practices in food establishments, the DineSafe model and its implications are specific to Toronto, as the data used in the analysis was collected from this city. Caution should be exercised when attempting to generalize these findings to other cities, as different regions may yield varying outcomes. Additionally, it is important to note that the model identifies associations between variables but does not establish causation, and should not be used as cause-and-effect ties. Moreover, potential biases in the data collection process cannot be ruled out. For example, individual inspector tendencies—such as one inspector being consistently stricter than others—could influence the severity of violations assigned, potentially skewing the results.

While all cities have similar programs to inspect public health and sanitation measures, one potential future step is to create more universal policies across a given province or country. More in-depth analysis could be done when practices and guidelines are standardized, providing more causation links that can be used to improve a region's public health. Also, public health initiatives and educational campaigns could be developed specifically for establishments that are prone to severe violations, educating both owners and employees on best practices for food safety. This could include targeted training on handling raw foods, sanitation, and compliance with food safety regulations. On the consumer side, making DineSafe data more accessible to the public, such as through a website or app, may also be a future step worth considering as a way to encourage consumers to make more informed decisions and incentivize businesses to strive for better performance.

A Appendix

A.1 Data cleaning

For the purpose of more simplistic modeling and easier interpretation, additional data cleaning steps were taken. After selecting for the variables of interest, food establishments were categorized into five categories. Before categorization, there were 46 types of food establishments. When modeling, this would translate to 46 categorical levels of the food establishment type variable, which would be too complex and redundant to interpret. As a result, the 46 types were organized into five types and shown below.

- Formal dining: restaurant, cocktail bar/beverage room, bakery, bake shop
- Fast food: food take out, food court, ice cream / yogurt, food cart, hot dog cart, catering vehicle, refreshment stand, mobile food prep
- Grocery: supermarket, fish shop, butcher shop, food store, food vending, flea market
- Institutional food preparation: college/uni food services, child care, institutional food service, cafeteria priv, cafeteria public, student nutrition site, secondary school food services, child care food prep, hospitals and health facilities, elementary school food services, hospital and health facilities, church banquet facility, Other Educational Facility Food Services, rest home, nursing home, serving kitchen, community kitchen, food caterer, boarding/lodging, food bank, serving kitchen, private club, centralized kitchen, retirement homes, commissary
- Industrial: food processing plant, food depot

A.2 Observational Data in the Context of DineSafe

Observational data, while important for understanding real-world phenomena, such as public health and food, comes with several limitations that can impact the robustness of analysis and the conclusions drawn. For example, one significant limitation is the potential for bias in data collection. In the context of DineSafe inspections, bias may arise through individual inspectors that differ in their rigor or interpretation of guidelines set out by Toronto Public Health. There may also be unconscious biases, which can result in inconsistent severity ratings. A study conducted on Dutch food safety services examined the sociological perspective of inspectors, highlighting how individual interpretations of their identity and professional responsibilities can lead to varying levels of rigor and standards in their work (Kleef 2014), ultimately reflecting potential biases in the rigor of inspections, which can affect how DineSafe data is examined and collected. Additionally, observational data often lacks randomization, meaning it may not adequately account for external factors influencing the results, such as differences in inspection timing or the socioeconomic context of a given establishment.

Observational datasets generally capture only correlations rather than causal relationships, making it difficult to definitively state that one factor directly influences the severity of violations. In the context of DineSafe, it is difficult to state that minimum inspections per year or type of establishment has a true casual relationship with the severity of a violation, although there is a statistically significant correlation. These limitations require careful interpretation of results and underscore the importance of supplementary methods, such as experimental designs or in-depth qualitative research, to validate findings and address gaps in understanding. However, that also poses an issue for DineSafe. Ethically, it would not be feasible to conduct experimental tests, such as A/B testing. the potential manipulation of inspection rigor between groups could lead to inconsistencies in public health protection. For example, if one group of establishments receives stricter inspections than another, consumers dining at less rigorously inspected establishments may face greater health risks that cannot be taken accountability for. Often, A/B testing requires double-blind experiments to account for any biases, raising an ethical concern for public food safety; businesses and consumers might unknowingly participate in an experimental setup, violating principles of informed consent, damaging trust in public health authorities and the integrity of the inspection system.

A.3 Idealized Survey

Another way to assess DineSafe as a public health program is through a consumer’s perspective. In 2020, a study was conducted to evaluate the effects of DineSafe’s disclosure system to customers (Barysheva 2020). In this study, a survey was conducted and advertised via social media platforms. The study consisted of 25 respondents of different age classes, all above the age of 18. While this survey gathered important data to analyze consumer’s knowledge on DineSafe, the sample size is fairly small to draw any definitive conclusions. Thus, in an idealized world with a high budget, a similar survey could be conducted at a larger and more representative scale. The idealized methodology for this survey to analyze consumer knowledge about DineSafe could involve quota sampling, with probability proportionate to the geographical location according to population data. For example, we could sample the Greater Toronto Area, however, this method ensures that the number of people polled is proportional to amount of people living in that region who would eat at those food establishments. Those in Downtown Toronto would have a higher probability of accessing and answering the survey compared to those in a less populated, uptown area.

After the sample has been selected with the above methods, the poll itself will be conducted through a survey sent to the chosen participants to reduce interviewer bias, or the effect of the interviewer’s views on the measured responses, which can affect results when face-to-face or phone interviews are utilized (Alexander 2023). To also reduce the social desirability effect, the survey must emphasize in the beginning that the respondents’ identity will be kept anonymous to the researchers and the final reported data (Stantcheva 2023). To encourage participation, a financial incentive should be offered for full completion of the poll. Displaying the progress bar and indicating the survey length at the very beginning also helps encourage participants

to complete the full questionnaire. Our poll would also be submitted to the appropriate ethics review boards, likely to those involved with Toronto Public Health, to increase confidence in our methodology.

A sample implementation of such a poll, created using Google form, can be found [here](#). Due to the limitations of Google form, the order of questions is consistent for all respondents. In an idealized methodology with a more professional surveying platform, the question order would be randomized when being presented to different respondents, with consideration for related questions must be kept together. This will help reduce response order bias, which is when the answer to a question is affected by the order in which it appeared in the survey (Stantcheva 2023). In this idealized survey, it would address similar topics of that found in the previous study, but reach a larger audience with more in-depth questions. As a result, public health officials can use both statistical modelling, like done in this paper, as well as these types of surveys to address knowledge gaps, both at the employer and consumer level, promoting improved public health in Toronto.

A.4 Datasheet for Dataset

The following datasheet is provided that provides essential, structured information about the DineSafe dataset (OpenDataToronto 2024). It draws information from Gebru et al. (2018) as a resource on the type of information included.

Motivation for dataset creation:

- Why was the dataset created?

The DineSafe dataset was created as a means to collect and store information on Toronto's food establishments and their overall public safety standards. The data is part of the Toronto Public Health's initiative to monitor and enforce sanitation and food safety standards across the city's food service sector.

- What tasks could the dataset be used for?

The DineSafe dataset can be used for public knowledge and for research purposes. The dataset should not be used for exploiting certain businesses or for personal financial gain.

- Who funded the creation of the dataset?

The dataset is part of Toronto Public Health's sanitation program and is funded by the City of Toronto. For more information, contact their email at: PublicHealth@toronto.ca

Dataset composition

- What are the instances?

Each observation or instance is based on a single food establishment in Toronto and its inspection record for a specific inspection date. There are no interactions between instances.

- What does each instance consist of?

In the raw data, there are 17 columns that describe each observation. This includes observation ID (or number), establishment ID, inspection ID, establishment name, type and address, the current status (pass, fail) of that establishment, minimum inspections per year for that specific establishment, any notes from the inspector, inspection date, outcome, amount fined, latitude and longitude of establishment, and a unique ID for that inspection. Everything is included in the raw dataset and does not rely on external resources for additional information.

- Are there recommended data splits or evaluation measures? What experiments were initially run on this dataset?

No recommendations for evaluating the data is mentioned by Toronto Public Health. No experiments are documented. This dataset likely serves as observational data for public knowledge and interest and does not serve as experimental data.

Data collection process

- How was the data collected?

Each observation in the data is a unique inspection result or violation of an establishment. An inspector will go to the establishment and note any violations based on DineSafe and city regulations. Each violation will be entered as a separate observation.

- Who was involved in the data collection process?

Individual inspectors were involved in the data collection process. These inspectors are employed by Toronto Public Health. Each inspector must be qualified based on Toronto's regulations and are compensated roughly \$43.58 - \$47.75 per hour (Jobs Toronto 2024). Inspectors are responsible for various inspections depending on their area of expertise, ranging from food sanitation to infectious diseases.

- How was the data associated with each instance acquired?

The data is directly observable by the inspector on site at the food establishment. The inspector will go to the physical food establishment and evaluate it based on DineSafe regulations, noting any violations they deem to be worthy of noting.

- Does the dataset contain all possible instances?

The dataset does not contain all possible instances. The dataset contains inspection information that was reported in the city of Toronto from June 2022 to June 2024. Therefore, it may not contain information on all food establishments in the city, only the ones that required inspection during this time period. The dataset also does not establish where the geographical boundary of “Toronto” is and may or may not include establishments in all of downtown and the Greater Toronto Area.

Data Processing

- What processing/cleaning was done?

For the analysis dataset in this paper, initial cleaning was done to select for the variables of interest. Additional cleaning was done to the variable, type of establishment, for the purpose of model simplicity. It is detailed in Section [A.1](#).

- Was the raw data saved in addition to the cleaned data?

Yes. The raw data, in csv form, was saved. Analysis data is saved in both .csv and .parquet forms. These saved datasets can be found in the repository [here](#) or linked on the first page of the paper.

- Notes about the processing software

Data was processed, cleaned, and analyzed using the statistical programming software, R and associated packages. More information is found in Section [2.1](#).

Dataset Distribution and Maintenance

- How is the dataset distributed?

The raw dataset is distributed on OpenDataToronto’s website, which can be found [here](#).

- Dataset license

The dataset is licensed under Open Government License, Toronto, which was developed through public consultation and a collaborative effort by the provincial and federal government. You are allowed to copy, modify, publish, translate, adapt, distribute or otherwise use the Information in any medium, mode or format for any lawful purpose. There are no fees associated with access or export. More information can be found [here](#).

- Who is supporting the dataset?

Toronto Public Health is supporting and maintaining the dataset. They can be contacted via email at PublicHealth@toronto.ca. The dataset is updated almost daily, or when there is new inspection data from an establishment.

Legal and Ethical Considerations

- Was the dataset informed about data collection?

Yes. The dataset includes information about food establishments and associated violations, which is public information. This dataset will not expose people or warrant legal action, as the food establishments are required to post their DineSafe rating and potential violations in a public space. This dataset is a reflection of that.

- Does the dataset unfairly advantage or disadvantage a particular social group?

No, it does not. The dataset is relaying information that has been given to the food establishments and serves as public information. The dataset does not contain sensitive or inappropriate information.

A.5 Figures

pdf

2

	min_inspections	establishment_typeFast Food	establishment_typeFormal Dining	establishment_typeGrocery	establishment_typeIndustrial	establishment_typeInstitution Food Prep
min_inspections	1	-0.03	0.15	-0.26	-0.06	0.07
establishment_typeFast Food	-0.03	1	-0.6	-0.16	-0.05	-0.16
establishment_typeFormal Dining	0.15	-0.6	1	-0.4	-0.14	-0.4
establishment_typeGrocery	-0.26	-0.16	-0.4	1	-0.04	-0.11
establishment_typeIndustrial	-0.06	-0.05	-0.14	-0.04	1	-0.04
establishment_typeInstitution Food Prep	0.07	-0.16	-0.4	-0.11	-0.04	1

Figure 8: Correlation Matrix Table

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Barysheva, Svitlana. 2020. “The Effect of DineSafe Food Safety Disclosure System on the Patronage of Customers.” <https://www.torontomu.ca/content/dam/occupational-public-health/PosterPDFs/2020/2020-svitlana-barysheva-food-dine-patronage.pdf>.
- Eastern Ontario Health Unit. 2024. “Food Premises Inspection Reports - FAQ.” <https://eohu.ca/en/advisories-reports/food-premises-inspection-reports-faq>.
- Food Safety Market. 2024. “Toronto Food Safety Regulations.” <https://www.foodsafetymarket.com/toronto-food-safety-regulations>.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. <https://www.john-fox.ca/Companion/>.
- Geburu, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2018. “Datasheets for Datasets.” <https://www.microsoft.com/en-us/research/uploads/prod/2019/01/1803.09010.pdf>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Jobs Toronto. 2024. “PUBLIC HEALTH INSPECTOR.” <https://jobs.toronto.ca/jobsatcity/job/Toronto-PUBLIC-HEALTH-INSPECTOR-ON-M5B-1W2/586461817/>.
- Kleef, Daphne van. 2014. *Inspections Services and Inter-Rater Reliability: Differentiating Professional Role Identities of Dutch Veterinary Inspectors*. *International Journal of Public Administration*. Vol. 38. <https://doi.org/doi.org/10.1080/01900692.2014.929589>.
- OpenDataToronto. 2024. *DineSafe Data*. <https://ckan0.cf.opendata.inter.prod-toronto.ca/dataset/b6b4f3fb-2e2c-47e7-931d-b87d22806948/resource/eda39233-4791-464e-98e6-094f51a01916/download/Dinesafe.csv>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Robinson, David, Alex Hayes, and Simon Couch. 2024. *Broom: Convert Statistical Objects into Tidy Tibbles*. <https://CRAN.R-project.org/package=broom>.
- Stantcheva, Stefanie. 2023. “How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible.” *Journal Article*. *Annual Review of Economics* 15 (Volume 15, 2023): 205–34. <https://doi.org/https://doi.org/10.1146/annurev-economics-091622-010157>.
- Toronto Public Health. 2024. “About DineSafe.” <https://www.toronto.ca/community-people/health-wellness-care/health-programs-advice/food-safety/dinesafe/about-dinesafe/>.
- UCLA Statistics. 2024. “Multinomial Logistic Regression.” <https://stats.oarc.ucla.edu/stata/dae/multinomiallogistic-regression/>.
- Wickham, Hadley. 2011. “Testthat: Get Started with Testing.” *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

- . 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 Golemund, et al. 2019. *Welcome to the tidyverse*. *Journal of Open Source Software*. Vol. 4. <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://CRAN.R-project.org/package=dplyr>.
- Yee, T. W. 2024. *VGAM: Vector Generalized Linear and Additive Models*. <https://CRAN.R-project.org/package=VGAM>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.