Analyzing Food Safety Compliance in Toronto: Identifying Current Hazards and Challenges*

Several Factors Affecting Local Food Safety in Toronto

Pengyu Sui

April 2, 2024

This research conducts an in-depth investigation into the current state of food safety compliance in Toronto, pinpointing the present condition of food safety in the city. By utilizing a dataset that includes various dining establishments and their inspection results, the study employs statistical methods for analysis and visualization, including the development of logistic regression models. The goal is to highlight areas requiring immediate attention or policy adjustments through a detailed examination of inspection reports, types of violations, frequencies, dates, and the nature of the violations. The findings suggest that the government needs to increase inspection frequency based on factors such as different seasons, types of food businesses, and the nature of violations to enhance local food safety, which is crucial for the health of every resident in Toronto.

1 Introduction

Food safety is a critical issue that affects the health of every resident in Toronto. It involves a variety of factors, including the freshness and hygiene of ingredients, as well as compliance with the expiration dates of products sold in supermarkets. Unfortunately, some suppliers engage in practices such as improper food storage and maintaining substandard sanitary conditions, which compromise food safety standards. Conducting a comprehensive study of the current state of food safety in Toronto is crucial. Our research will provide the public and regulatory authorities with a clearer understanding of the challenges and risks associated with food safety in the city, and facilitate policy adjustments and improvements to further enhance Toronto's food safety.

*Code and data are available at: https://github.com/simon0202sui/Analyzing-Food-Safety-Compliance-in-Toronto.git.

The focus of this study is on the Toronto Open Dataset, which documents the results of food safety inspections at a variety of establishments. The main objective is to explore the factors that influence the level of compliance in these establishments. The estimated objective of this study is the impact of various operational and environmental factors on the likelihood of an establishment's compliance with food safety regulations. This includes analyzing aspects such as the type of food establishment, frequency of inspections, and record of previous violations. we leverage the official dataset provided by Open Data Toronto (Gelfand 2022) to conduct an in-depth analysis of food safety compliance across the city's dining establishments. approach involves utilizing R (R Core Team 2023), to meticulously clean and preprocess the dataset, ensuring the reliability and validity of subsequent analyses.By employing a combination of descriptive statistics and predictive modeling techniques, including logistic regression, this paper seeks to estimate the influence of these factors on compliance rates. The insights gained aim to pinpoint specific risk factors and recommend targeted interventions for improving compliance, thereby enhancing public health safety in Toronto.

This paper is divided into four sections following the introduction, excluding the introduction itself. In Section 2 , we examine the dataset used for the report, including an introduction to the column names in the dataset, tables for cleaning the raw data, and summary statistics. In Section 3, we will design and explain the models used for analyzing the data, detailing the significance of each variable. In Section 4, we will visualize the outcomes of the statistical and logistic regression models through charts to aid our analysis. Finally, in Section 5, based on the analysis, we will offer three recommendations regarding the current state of food safety in Toronto. We will also highlight the limitations of our report and suggest possible improvements for future research.

2 Data

2.1 Data Source

The dataset used in this study comes from the DineSafe program managed by Toronto Public Health, available through Open Data Toronto (Gelfand 2022). DineSafe is an inspection program that targets all establishments providing and preparing food in Toronto. The significance of this dataset lies in its comprehensive coverage of food safety compliance, offering insights into the operational standards of local food enterprises. The main entity of the dataset is a CSV file named DineSafe, which contains operational information about various dining establishments, including the results of each inspection (categorized as pass, conditional pass, or closure), compliance status, and the regulatory efforts of related government agencies, involving inspection frequency and penalties. This dataset is crucial for assessing the current state of food safety in Toronto and identifying areas needing improvement. Although similar datasets exist from other jurisdictions or private entities, the DineSafe dataset was chosen for its direct relevance to Toronto, its authoritative nature, assured accuracy, and the detailed granularity of the inspection data it provides. Local focus is essential for formulating policy recommendations and interventions tailored to Toronto's unique food safety challenges. The data collection process is conducted by Toronto Public Health through regular administrative inspections of dining establishments, ensuring the data's reliability and high reference value. The report relies on this dataset and uses the statistical software R (R Core Team 2023). For data cleaning and preparation, the tidyverse package (Wickham et al. 2019) was used, and for modeling, the rstanarm package (Goodrich et al. 2020) was employed. Additionally, rstanarm was used for citation (Goodrich et al. 2020), ggplot2 for visualizations (Wickham 2016), and modelsummary for providing concise model overviews (Arel-Bundock 2022).

2.2 Dataset characteristics

The dataset is a complete CSV file encapsulating various indicators pertinent to food safety analyses. The primary variables include:

- Establishment Type:Categorizes each food service entity into various types, such as restaurants, cafes, fast food outlets, and food trucks. This classification assists in assessing the specific risks associated with different service models.
- Inspection Date: Records the dates on which food safety inspections were conducted, allowing for temporal analysis of compliance and identification of any seasonal trends in food safety issues.
- Establishment Status: Indicates the outcome of each inspection (e.g., pass, conditional pass). This variable is crucial for evaluating the compliance levels of food establishments.
- Severity of Violations: Details the severity of any violations found during inspections, categorized into minor, significant, or critical. This helps in understanding the extent of non-compliance and potential health risks.
- Min. Inspections Per Year: This column typically refers to the minimum number of inspections that an establishment, such as a restaurant or food service provider, is required to undergo annually.

2.3 Data Clean

Data cleaning is the first step in my data analysis process. It ensures the quality and accuracy of the data and provides a reliable foundation for subsequent analyses. During this process, I loaded the data using the read_csv function from the readr package (Wickham, Hester, and RStudio 2024) in R (R Core Team 2023) and renamed variables for clarity and interpretability. Then, I performed comprehensive cleaning and modification steps using the mutate and select functions from the dplyr package (Wickham et al. 2023). I filtered, processed, and corrected the raw data to eliminate potential errors and inconsistencies, thus ensuring the integrity and usability of the data. Using kable (Xie 2023) to generate a summary chart (Table 1) to

Table 1: summarize Table

Establishment Status	Establishment Type	Inspection Date	Severity	Min. Inspections
Pass	Restaurant	2022-06-27	C - Crucial	
Pass	Supermarket	2022-09-14	C - Crucial	
Pass	Restaurant	2022-09-08	C - Crucial	
Pass	Food Store (Convenience/Variety)	2022-07-15	C - Crucial	
Pass	Food Store (Convenience/Variety)	2022-07-15	M - Minor	
Pass	Butcher Shop	2023-08-18	M - Minor	

visualize the first few columns of the cleaned dataset, I observed that the column names of the variables for subsequent study are fully presented in the chart, there are no missing entries, and the data can be used for subsequent analyses.

3 Model

3.1 Model Setup

Through the data analysis presented in this paper, we discovered that there is a correlation between Establishment Type, Severity, the minimum number of inspections per year and the result (pass, conditional pass).To further analyze the relationships, we constructed a logistic regression model to model the relationship.

We define the model as follows:

$$Y_i \mid \pi_i \sim \text{Bernoulli}(\pi_i) \tag{1}$$

$$\text{logit}(\pi_i) = \alpha + \beta_1 \times \text{Establishment\_Type}_i + \beta_2 \times \text{Severity}_i + \beta_3 \times \text{Min\_Inspections\_Per\_Year}_i \tag{2}$$

$$\alpha \sim \text{Normal}(0, 10) \tag{3}$$

$$\beta_j \sim \text{Normal}(0, 10) \quad \text{for } j = 1, 2, 3 \tag{4}$$

$$\tag{5}$$

Here's a breakdown of the components of the model:

- Y_i: The outcome variable represents whether a food establishment passes (1) or Condition passes (0) an inspection.
- π_i: The probability of the establishment i passing the inspection.
- $\text{Bernoulli}(\pi_i)$: Indicates that the outcome variable Y_i follows a Bernoulli distribution with probability π_i.
- $\text{logit}(\pi_i)$: The logit function transforms the probability π_i to the log-odds scale.
- α: The intercept term, representing the log-odds of passing the inspection when all predictor variables are zero.
- $\beta_1, \beta_2, \beta_3$: The coefficients corresponding to the predictors Establishment Type, Severity of infractions, and Minimum Inspections Per Year, respectively.
- $\text{Normal}(0, 10)$: The prior distribution for the intercept and coefficients, assumed to be normally distributed with a mean of 0 and a standard deviation of 10.

The logistic regression model allows us to estimate the impact of each predictor variable on the likelihood of a food establishment passing an inspection, while accounting for the potential correlation between predictors. By analyzing the estimated coefficients, we can identify the relative importance of each predictor and make informed decisions regarding food safety measures and regulatory interventions.

3.1.1 Model justification

In our analysis, we assume that the outcome variable `Y_i`, conditional on the probability `pi_i`, follows a Bernoulli distribution, which is appropriate for binary response data. The binary response in our context could represent a food establishment passing or failing an inspection.

The probability `pi_i` of the establishment `i` passing the inspection is modeled using the logistic function, which is the inverse of the logit function. The logit function, given by the equation `log( pi_i / (1 - pi_i) )`, linearly relates the log-odds of the probability `pi_i` to the predictors. The model includes an intercept `alpha` and three predictors: Establishment Type (`Establishment_Type_i`), Severity of infractions (`Severity_i`), and the minimum number of inspections per year (`Min_Inspections_Per_Year_i`).

The intercept `alpha` represents the log-odds of a food establishment passing the inspection when all other variables are zero. The coefficients `beta_1`, `beta_2`, and `beta_3` are the effects of each predictor variable on the log-odds scale. Specifically, `beta_1` represents the effect of the Establishment Type on the log-odds of passing, `beta_2` for the Severity of violations, and `beta_3` for the Minimum Inspections Per Year.

These coefficients are estimated from the data and are assumed to come from a Normal distribution with a mean of 0 and a standard deviation of 10, as expressed by `beta_j ~ Normal(0, 10)`. This expresses a prior belief that, before observing the data, we expect the effect sizes to be small, with a wide range of plausible values given the standard deviation.

The notation `beta_j` for `j=1,2,3` indicates that this model includes three separate `beta` coefficients, each corresponding to one of the three predictor variables. These coefficients determine the direction and strength of the relationship between each predictor and the outcome, with positive values indicating higher odds of passing the inspection and negative values indicating lower odds.

By analyzing the sign and magnitude of these coefficients, we can interpret the importance and influence of each predictor on the likelihood that a food establishment will pass the inspection, allowing for targeted interventions based on the identified risk factors.

4 Results

4.1 Variables of Interest

We organized the timeline within the dataset to track the monthly variation in the number of restaurant violations from 2022 to 2024, in order to validate my hypothesis about the seasonal impact on food safety in Toronto. Figure 1, a line graph (Figure 1), shows the trend in the number of restaurant violations over time and provides a general overview. There are two noteworthy observations: first, there is a significant increase in violations in 2024 compared to 2022 and 2023. This trend indicates that food safety violations are worsening, which may

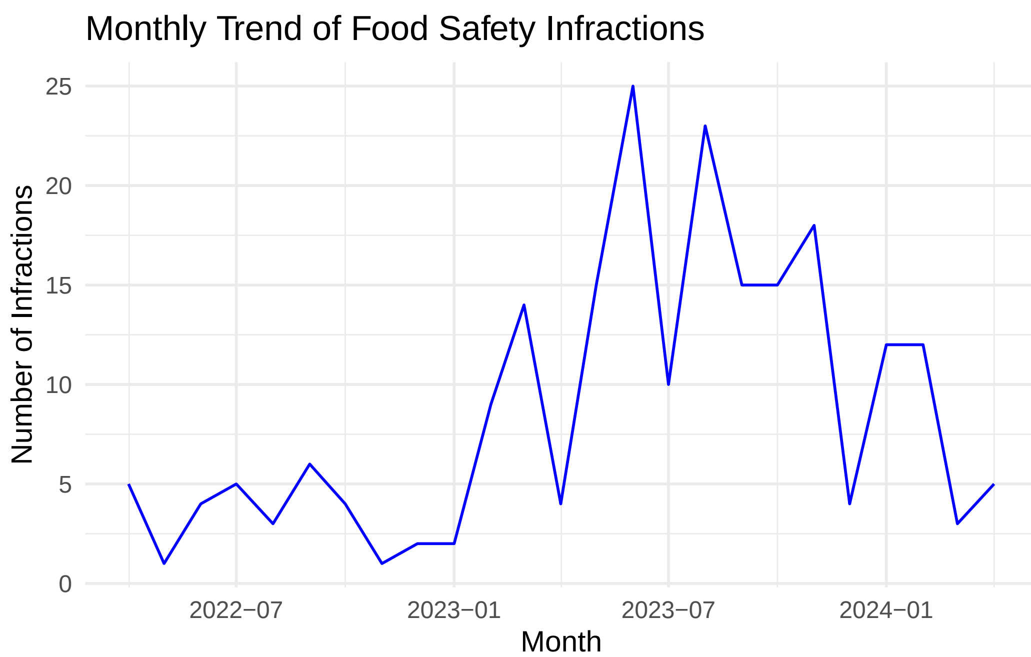

Figure 1: Summarize data by month

suggest a need for stricter safety measures or policy intervention. Second, both in 2023 and 2024, there is a pattern of seasonal fluctuations—peaks and troughs—that repeats annually, which could imply that certain times of the year are associated with a higher rate of violations. This may be related to periods of increased food service activity, such as holidays or festivals. Additionally, the most significant peaks occur in the middle of the year and at the beginning of the next year. The mid-year peak could be associated with summer activities when food-related events and outdoor dining are likely to increase. The early-year peak could be related to the winter holiday season, which traditionally sees an increase in dining out.

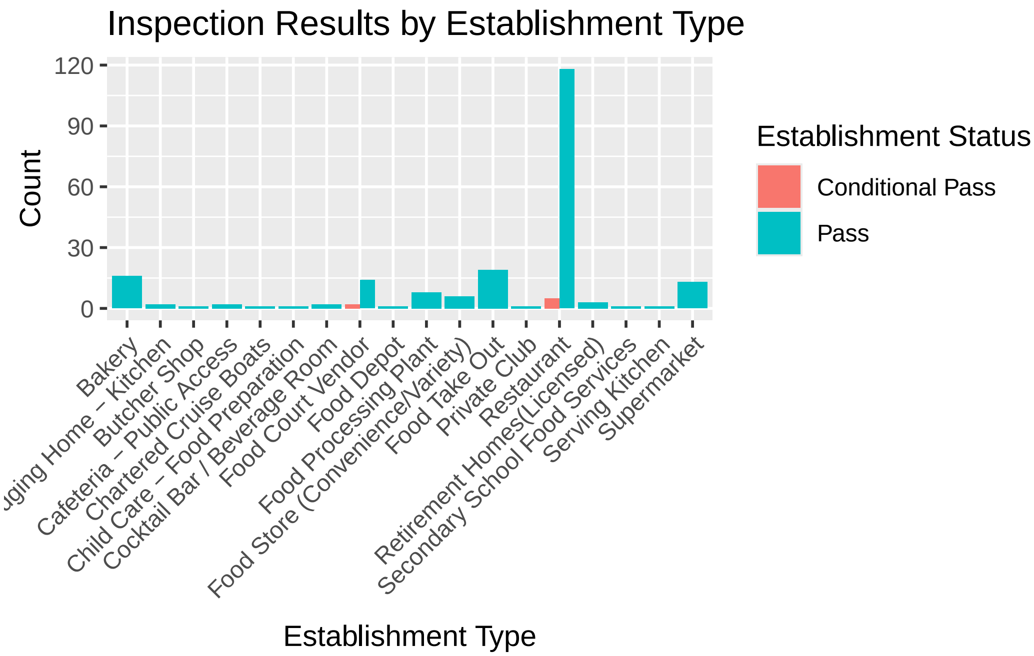

Figure 2: Relationship between Establishment Type and Establishment Status

This graph (Figure 2) focuses on the outcomes of food safety inspections across different types of dining establishments, revealing that the "Restaurant" category has a noticeably higher rate of both passes and conditional passes. This indicates that restaurants are either inspected more frequently or are simply more numerous compared to other types of establishments. Concurrently, there is a considerable number of conditional passes in the categories of "Restaurants," "Food Take Out," and "Private Clubs." A conditional pass typically signifies minor infractions that need to be addressed but are not severe enough to necessitate immediate closure. Other establishment types, such as "Bakeries," "Food Stores," and "Cocktail Bars," have relatively fewer inspections, which might reflect their smaller scale of operation or different inspection frequency policies. It is evident that Toronto's regulatory authorities are concentrating their inspection efforts on restaurants, and there is insufficient scrutiny of other food establishments like mobile food carts and small eateries, which may require additional focus or assistance to meet food safety standards.

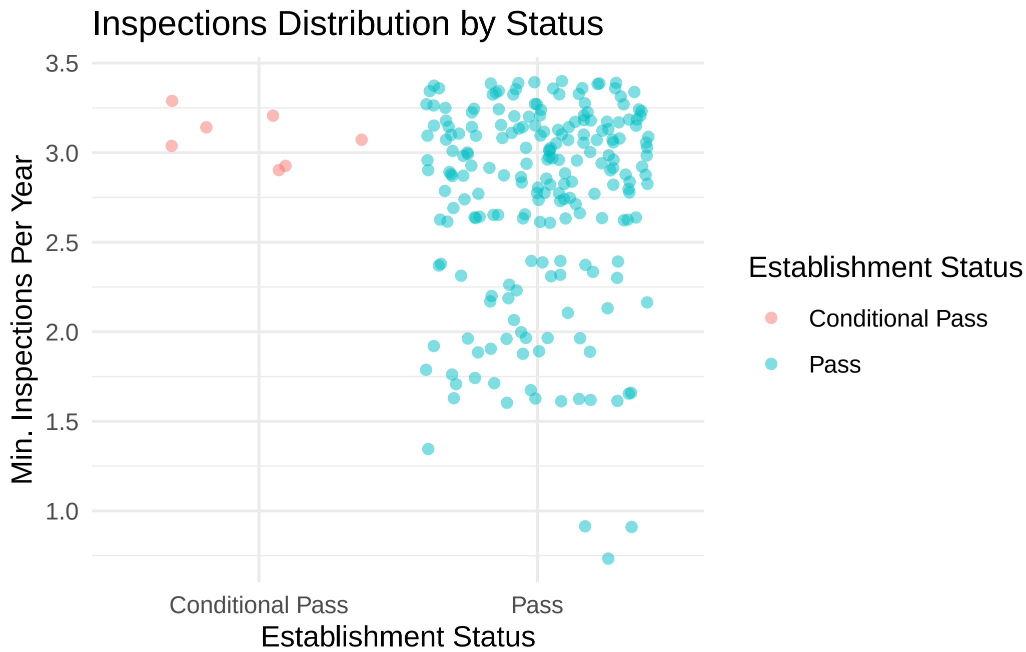

Figure 3: Relationship between wing Min. Inspections Per Year and Establishment Status

Scatter plot (Figure 3) titled "Inspections Distribution by Status" illustrates the relationship between the status of establishments following inspections—either 'Pass' or 'Conditional Pass'—and the minimum number of inspections per year. A 'Conditional Pass' suggests that violations were identified during the inspection, which were rectified before final approval was granted. It is observable that there are more points for the 'Pass' status, indicating that the majority of establishments comply with food safety standards during inspections. There are outliers with a higher frequency of inspections, possibly reflecting a targeted approach towards establishments that had prior violations or are considered high-risk. Notably, the 'Conditional Pass' status, marked in red, is less common and appears after a higher threshold of inspections, specifically after 2.5 inspections per year. Establishments with fewer than 2.5 inspections per year show no 'Conditional Pass' outcomes, suggesting that an increase in the minimum number of inspections could lead to the identification of more violations.

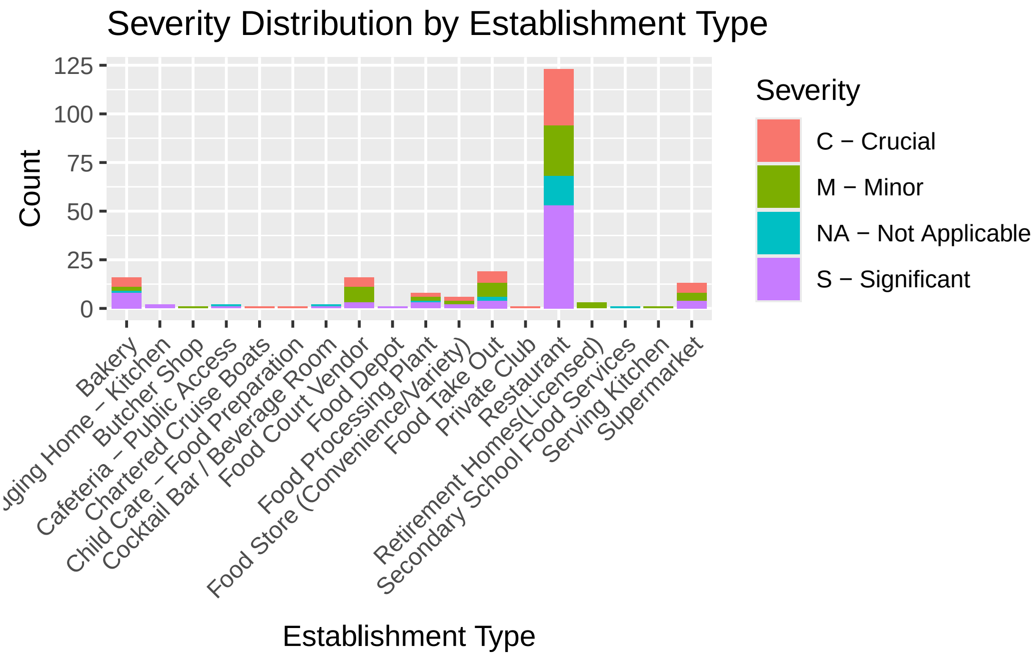

Figure 4: Relationship between Establishment Type and Severity

This figure (Figure 4) "Severity Distribution by Establishment Type" depicts the frequency of food safety violations across different types of establishments, organized by the severity of the infractions, which are color-coded: 'Crucial' (red), 'Minor' (green), 'Not Applicable' (blue), and 'Significant' (purple). The chart demonstrates that most violations are minor, as indicated by the predominance of green. However, attention is drawn to the categories featuring red and purple bars, signifying more severe infractions. Notably, establishments such as supermarkets, take-outs, restaurants, mobile food carts, and bakeries are prominent, suggesting that these categories have more complex issues that require additional investigative support.

4.2 Model result

ACCording to table (Table 2),the analysis of our Bayesian logistic regression model reveals several key insights into the factors affecting compliance rates across various food service establishments. Notably, Food Court Vendors and Restaurants exhibit significant negative coefficients (-3.087 and -2.070, respectively), suggesting that these types of establishments generally have lower compliance rates, likely due to challenges in maintaining food safety standards. Conversely, Food Take Out services show a positive coefficient (0.771), indicating higher compliance, possibly due to simpler operational processes. Severity levels also play a crucial role, with Minor and Significant severities showing negative impacts on compliance rates, as coefficients for SeverityMMMinor and SeveritySMSignificant are -1.264 and -1.781, respectively. This indicates that more severe violations are associated with lower compliance rates. Additionally, interaction terms like b_Establishment_TypeFoodCourtVendor × SeverityMMMinor and b_Establishment_TypeRestaurant × SeverityMMMinor further underscore the challenges faced by these establishments under strict inspection conditions, with coefficients of -1.532 and -1.595. These findings highlight the importance of targeted regulatory strategies to improve food safety practices across differing service types and severity levels.

Table 2: Model summary

	Result
b_Intercept	16.344
b_Establishment_TypeBoardingDLodgingHomeMKitchen	0.417
b_Establishment_TypeButcherShop	0.239
b_Establishment_TypeCafeteriaMPublicAccess	0.366
b_Establishment_TypeCharteredCruiseBoats	0.012
b_Establishment_TypeChildCareMFoodPreparation	0.037
b_Establishment_TypeCocktailBarDBeverageRoom	0.148
b_Establishment_TypeFoodCourtVendor	−3.087
b_Establishment_TypeFoodDepot	0.074
b_Establishment_TypeFoodProcessingPlant	0.737
b_Establishment_TypeFoodStoreConvenienceDVariety	0.185
b_Establishment_TypeFoodTakeOut	0.771
b_Establishment_TypePrivateClub	0.011
b_Establishment_TypeRestaurant	−2.070
b_Establishment_TypeRetirementHomesLicensed	0.382
b_Establishment_TypeSecondarySchoolFoodServices	0.035
b_Establishment_TypeServingKitchen	0.243
b_Establishment_TypeSupermarket	0.512
b_SeverityMMMinor	−1.264
b_SeverityNAMNotApplicable	0.941
b_SeveritySMSignificant	−1.781
b_Min_Inspections_Per_Year	−2.558
b_Establishment_TypeBoardingDLodgingHomeMKitchen × SeverityMMMinor	0.007
b_Establishment_TypeButcherShop × SeverityMMMinor	0.233
b_Establishment_TypeCafeteriaMPublicAccess × SeverityMMMinor	0.023
b_Establishment_TypeCharteredCruiseBoats × SeverityMMMinor	0.024
b_Establishment_TypeChildCareMFoodPreparation × SeverityMMMinor	−0.004
b_Establishment_TypeCocktailBarDBeverageRoom × SeverityMMMinor	0.003
b_Establishment_TypeFoodCourtVendor × SeverityMMMinor	−1.532
b_Establishment_TypeFoodDepot × SeverityMMMinor	0.041
b_Establishment_TypeFoodProcessingPlant × SeverityMMMinor	0.263
b_Establishment_TypeFoodStoreConvenienceDVariety × SeverityMMMinor	0.072
b_Establishment_TypeFoodTakeOut × SeverityMMMinor	0.256
b_Establishment_TypePrivateClub × SeverityMMMinor	−0.051
b_Establishment_TypeRestaurant × SeverityMMMinor	−1.595
b_Establishment_TypeRetirementHomesLicensed × SeverityMMMinor	0.379
b_Establishment_TypeSecondarySchoolFoodServices × SeverityMMMinor	0.022
b_Establishment_TypeServingKitchen × SeverityMMMinor	0.222
b_Establishment_TypeSupermarket × SeverityMMMinor	0.085
b_Establishment_TypeBoardingDLodgingHomeMKitchen × SeverityNAMNotApplicable	−0.006
b_Establishment_TypeButcherShop × SeverityNAMNotApplicable	−0.001
b_Establishment_TypeCafeteriaMPublicAccess × SeverityNAMNotApplicable	0.103
b_Establishment_TypeCharteredCruiseBoats × SeverityNAMNotApplicable	0.005
b_Establishment_TypeChildCareMFoodPreparation × SeverityNAMNotApplicable	0.004
b_Establishment_TypeCocktailBarDBeverageRoom × SeverityNAMNotApplicable	0.023
b_Establishment_TypeFoodCourtVendor × SeverityNAMNotApplicable	−0.003
b_Establishment_TypeFoodDepot × SeverityNAMNotApplicable	−0.003
b_Establishment_TypeFoodProcessingPlant × SeverityNAMNotApplicable	0.101
b_Establishment_TypeFoodStoreConvenienceDVariety × SeverityNAMNotApplicable	0.019
b_Establishment_TypeFoodTakeOut × SeverityNAMNotApplicable	0.142

The visualization of the top 7 effect sizes and their 95% confidence intervals from our logistic regression model reveals several impactful predictors. According to figure (Figure 5), the intercept stands out with a substantial positive value, suggesting a high baseline likelihood of passing food safety inspections when other variables are at zero. Notably, food court vendors are significantly less likely to pass an inspection, as indicated by a large negative coefficient. Moreover, an increased frequency of inspections per year correlates with a lower probability of passing, possibly reflecting stricter scrutiny or insufficient time for establishments to address issues. Moreover, an increased frequency of inspections per year correlates with a lower probability of passing, possibly reflecting stricter scrutiny or insufficient time for establishments to address issues. The interaction terms, especially for restaurants and food court vendors when coupled with significant severity violations, suggest a compounded negative impact on passing inspections. Restaurants also display a negative coefficient independently, further emphasizing the challenges they face. Lastly, the presence of significant severity violations () markedly decreases the likelihood of passing, underscoring the stringent penalties for severe infractions.

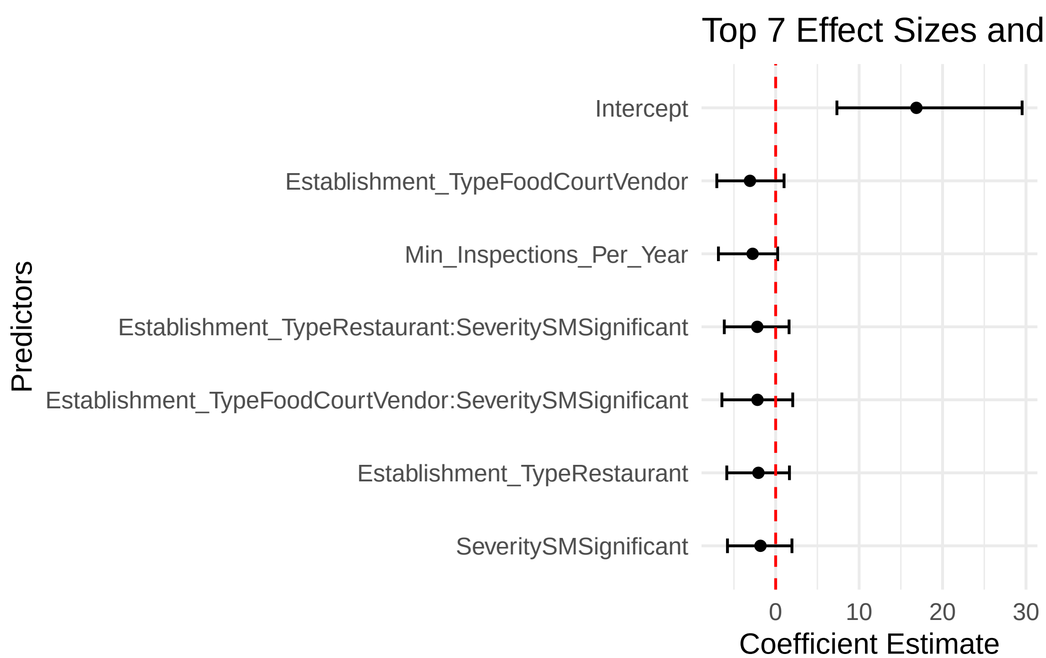

Figure 5: Top 7 effect

5 Discussion

5.1 Toronto Food Service Food Safety Survey and Regulatory Policies Need to be Adapted to the Type of Food Establishment

The correlation between food safety and the operational modes of providers in Toronto is significant. For instance, restaurants, being fixed establishments with permanent structures, face challenges in ensuring food safety due to the large number of establishments and limited inspection personnel. On the other hand, mobile food vendors, often found on the streets, predominantly serve fast food like hot dogs and hamburgers. The affordability and simplicity of their production environments often lead to hygiene issues.

It's noteworthy that supermarkets also pose significant food safety risks. In some Toronto supermarkets, it's not uncommon to find products like milk that have expired by a day, possibly due to negligence on the part of store managers. Conversely, suppliers such as food warehouses demonstrate commendable pass rates and favorable inspection results, indicating that Toronto's food storage and transportation processes are reliable.

It is recommended that relevant authorities increase regulatory oversight on enterprises with frequent food safety issues, including restaurants, mobile food vendors, and supermarkets. This proactive approach can help mitigate risks and enhance overall food safety standards in Toronto.

5.2 The number of overall Toronto food safety inspections per year needs to increase

Based on our research analysis, a significant number of food-related enterprises reveal issues only after more than 2.5 inspections, while the probability of passing on the first inspection is very high, close to one hundred percent. This underscores the importance of inspection frequency, as more frequent inspections of food establishments are conducive to improving the overall safety standards of Toronto's restaurant industry and food supply. According to the results depicted in the charts, given the considerable number of enterprises found to engage in non-compliant behavior, we recommend that relevant authorities increase the minimum number of inspections to at least three times. This measure will effectively address food safety concerns in Toronto.

5.3 Toronto government should adjust food safety regulation for different seasons

According to the research findings, the Toronto government should adjust its inspection system based on seasons. For example, during periods of high temperatures when food safety issues are more common, or during holidays when there is increased customer traffic, temporarily increasing the frequency of random inspections could effectively alert some businesses to hygiene issues. As the number of violations detected increases over time, indicating worsening food safety problems in Toronto, it is necessary to increase the severity of penalties and inspection frequency to warn food companies. Food safety concerns affect the health of every Toronto citizen.

5.4 Weaknesses and next Steps

5.4.1 Weakness

According to the research findings, the Toronto government should adjust its inspection system based on seasons. For example, during periods of high temperatures when food safety issues are more common, or during holidays when there is increased customer traffic, temporarily increasing the frequency of random inspections could effectively alert some businesses to hygiene issues. As the number of violations detected increases over time, indicating worsening food safety problems in Toronto, it is necessary to increase the severity of penalties and inspection frequency to warn food companies. Food safety concerns affect the health of every Toronto citizen.

5.4.2 Next Steps

The next step could involve exploring how different categories of restaurants are penalized for violations to assess the efficiency of the penalty mechanism in reducing violations, along with analyzing the final review results. Additionally, refining the logistic regression model and incorporating information on the horizontal and vertical coordinates of each restaurant could facilitate the creation of a violation map. This map would allow for the analysis of areas with a high incidence of food safety issues.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

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. 2020. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm.

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

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 RStudio. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

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