Fire Safety Regulation Violations in Toronto: Analysis by Property Type and Inspection Duration for Inspections Closed in 2024*

My subtitle if needed

Yunkai Gu

December 2, 2024

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

Table of contents

---

*Code and data are available at: https://github.com/Kylie309/Toronto-Fire-Inspection.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

This paper uses Highrise Residential Fire Inspection Results dataset obtained from Open Data Toronto as the main source of data (City of Toronto 2024b). The portal is launched by City of Toronto to meet growing demand for open data. The raw dataset presents information of properties where Inspectors from Toronto Fire Services (TFS) have examined on, as well as whether the properties violated fire safety regulations and the details of violations if so.

Note that though the dataset named "Highrise" Residential Fire Inspection Results, expansion work completed in 2020 has allowed for inspection results for more prperty types to be presented in the dataset. The portal now not only shows residential high-rise buildings, but also shows "inspection matters for multi-unit residential occupancy types including high-rise, low-rise, small multi-unit buildings, rooming houses/group homes, hotels, motels, detention centres, and all applicable vulnerable occupancy buildings (hospitals, nursing homes, residential cares and group homes designated a VO)" (City of Toronto 2024a).

To ensure the analysis presented in this paper remains relevant and current, only inspection cases that have concluded their inspection files or enforcement proceedings within the year 2024 are selected. This is achieved by filtering the dataset to keep only rows where the 'INSPECTIONS_CLOSEDDATE' column displays a date that falls within the year 2024. This approach ensures that the data reflects the most recent developments, providing an up-to-date basis for analysis and maintaining a clear and accurate representation of the latest trends and outcomes in fire safety checks.

More detailed data cleaning process is included in Appendix B.

The analyses presented in this paper were conducted using R programming language (R Core Team 2023). The `tidyverse` packages (Wickham et al. 2019) were used in the process of data simulation, testing beforehand. After the original raw data was downloaded by using `tidyverse` package (Wickham et al. 2019), data cleaning process was done by using `tidyverse` package (Wickham et al. 2019), `lubridate` package (Grolemund and Wickham 2011), and `arrow` package (Richardson et al. 2024). We use `testthat` package (**citetest?**) to develop the test for structure and format of simulation and analysis data.

Then, models were constructed using `tidyverse` package (Wickham et al. 2019), `lubridate` package (Grolemund and Wickham 2011) and `rstanarm` (Goodrich et al. 2022) package. The model results were then presented by both the tables and graphs. Tables were made with `modelsummary` (Arel-Bundock 2022) package, `broom.mixed` (Bolker and Robinson 2024) package, `dplyr` (Wickham et al. 2023) package and `knitr` package. Graphs were made with `ggplot2` package (Wickham 2016) and `RColorBrewer` package (Neuwirth 2022).

Following Alexander (2023), this paper uses Bayesian modeling to predict the pass or fail status for safety checks of the properties in Toronto. The factors of region, property type and inspection time intervals from the dataset are mainly chosen and considered to examine their effects on the likelihood of violations for the properties.

2.2 Measurement

The dataset captures data related to the results of fire safety inspections conducted by Toronto Fire Services (TFS) within highrise residential buildings, which is guided by the Fire Protection and Prevention Act, 1997 (FPPA) and regulations under the FPPA, like the Ontario Fire Code (City of Toronto 2024a). It indicates properties where violations have been found by an Inspector which are required to be fixed for compliance at the time with the Ontario Fire Code as well as properties with no observable violations (City of Toronto 2024b).

The data collection process was conducted by TFS Fire Inspectors. Inspectors assess tangible aspects of fire safety based on observable and measurable attributes of the physical environment within the properties, such as proper storage of combustible materials, functionality of fire alarms and fire pumps, maintenance records for fire safety equipment and others in reference to the regulations under FPPA.

After the inspections were completed, the observations have been translated into records and written into collected data. The raw data entry presented by Open Data Toronto includes three main aspects: property identifier (such as address, property type), inspection information (such as inspection start and end dates), and inspection outcomes (pass or fail status for safety checks, and details of violations).

Note that this dataset only includes 'closed' cases, which are cases whose inspection processes have ended. The ongoing inspections were not available in the dataset, and therefore would lead to potential loss of information. This would be further evaluated in detail in the discussion section.

2.3 Outcome Variable

2.3.1 Violation

The outcome variable for the model is the violation status of the properties. In other words, it is the variable that represents whether the properties inspected have been found violating fire safety regulations by an Inspector or not.

In the raw dataset, columns 'VIOLATION_FIRE_CODE', 'VIOLATION_DESCRIPTION' and 'VIOLATIONS_ITEM_NUMBER' indicates the outcome of the inspections. The previous two provides detailed information on fire code under which violation was noted. The latter one presents the order number of violations by code, and shows 0 if no violations observed.

Therefore, a new column 'violation' is mutated to make the analysis procedure easier. It contains numeric values of 1 and 0, which indicates violation observed and no violation observed respectively. The outcome variable is thus binary.

Figure 1 presents the visualized counts of properties that were found violations versus those that were not. The height of each bar shows the total number of cases corresponding to each category. We could see that the orange bar is significantly taller, indicating that among all cases selected for analysis, violations were reported more frequently in the dataset (9829 cases out of 13193). On the other hand, the blue bar is shorter, indicating there were fewer cases without violations (3364 cases out of 13193).

In conclusion, the data contains a higher number of cases with violations (1) compared to cases where no violations were observed (0), suggesting that violations are common.
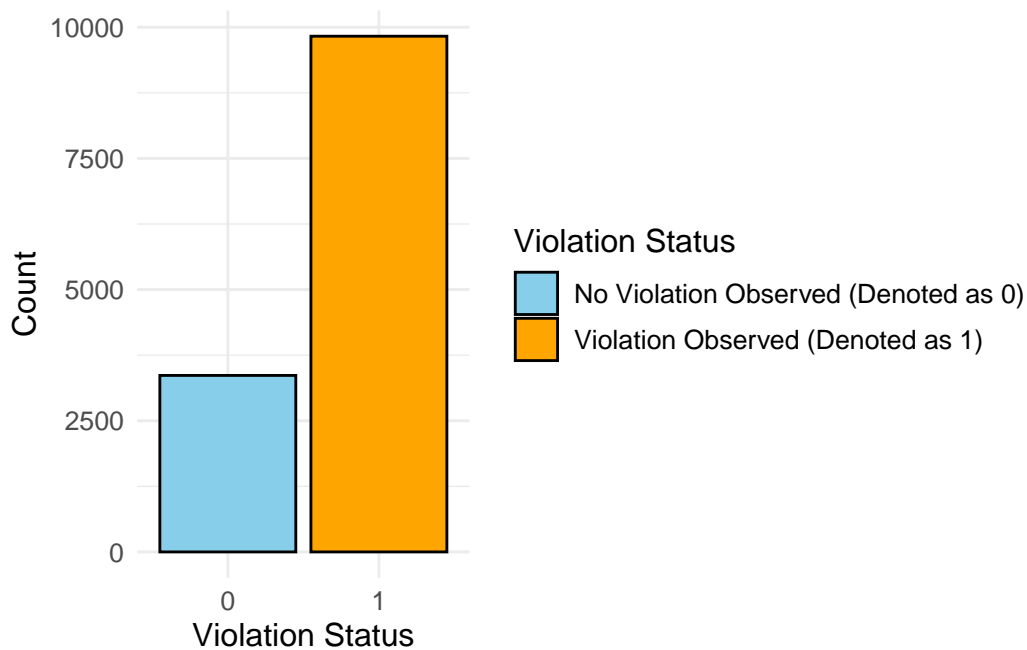

Figure 1: Counts of Properties that Were Found Violations versus Those that Were Not

2.4 Predictor Variables

2.4.1 Inspection Time Interval

One predictor included in the model is the inspection time interval. It refers to the duration of the fire inspections conducted by the TFS Inspector for each case.

In the raw dataset, there are two columns recording the date: 'INSPECTIONS_OPENDATE' and 'INSPECTIONS_CLOSEDDATE'. Previous one is the date TFS opened the inspection file, whereas the latter one is the date TFS closed inspection file or date or date enforcement proceedings ended (whichever is later).

During the data cleaning process, the date is converted into time interval to enable easier analysis. A new column named 'date_num' is mutated by calculating the time difference between the end time and the start time of the inspection, in the unit of days.

Figure 2 shows the distribution of the inspection time interval. X-axis represents the number of days of each inspection case, and Y-axis represents the counts. The distribution has peak at low intervals, and the highest bar is near 0 days, indicating that a large number of properties have very short inspection time intervals. The right-skewed inspection suggests that as the inspection interval increases, the corresponding counts decrease stably. The long tail of the right-hand side of the distribution means that there are properties with very long intervals (up to 1500 days or more).
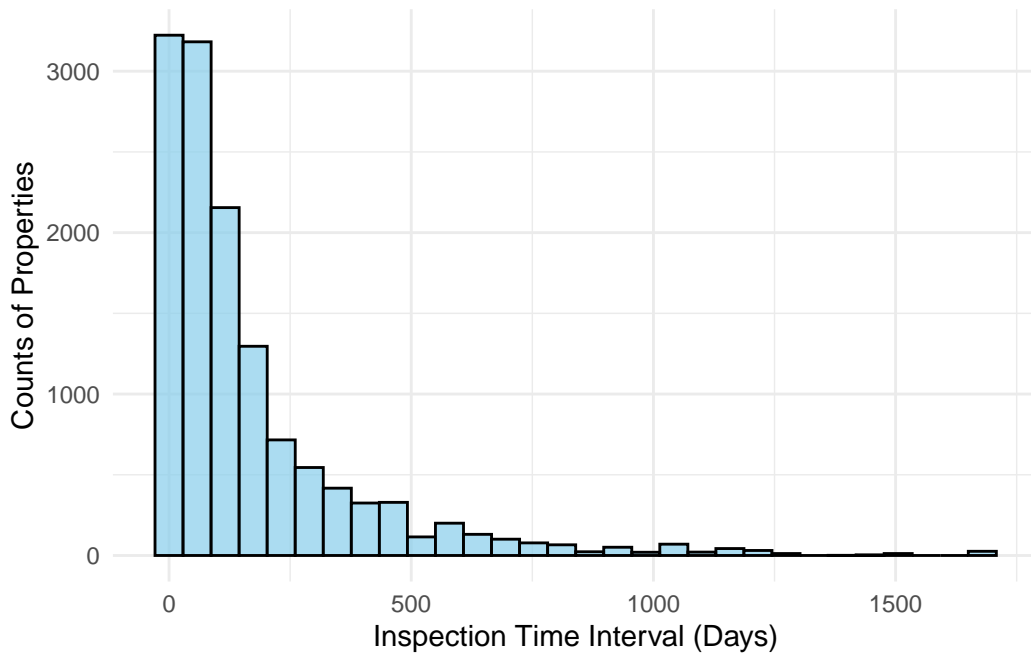
Figure 2: Distribution of Inspection Time Interval

2.4.2 Property Type

The second predictor variable included in the model is the property type. It refers to the occupancy type of the property where the inspection took place, and is directly contains in the raw dataset.

To visualize the variable, two bar plots are created separately in Figure 3. The left panel presents the number of occurrences of "High Rise" and "Low Rise", with the bar representing "High Rise" relatively taller, indicating a higher count compared to low rise. This suggests that the inspection results and data collection focus more heavily on high-rise buildings.

The right panel includes the other categories of properties. Specifically, the types include: "Detention," "Group Home," "Group Home (VO)," "Hospital," "Hotel & Motel," "Nursing Home," "Residential Care," and "Rooming House." We could see that "Rooming House" has the highest count among these property types, whereas categories of "Detention" and "Group Home" have relatively lower counts, indicating that they are less represented in the dataset.

In general, there is a diverse distribution across the property types, with some being far less frequent than others. High-rise properties dominate overall, while rooming houses stand out within other categories. The dataset might reflect an emphasis on inspecting high-rise and rooming houses more than other property types.

(a) High Rise versus Low Rise   (b) Other Types
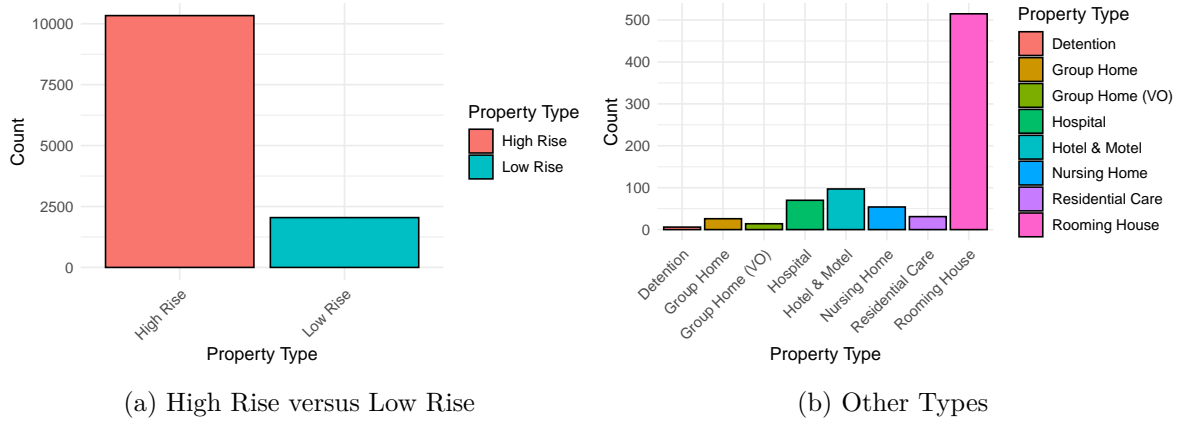
Figure 3: Counts of Property Types

2.5 Violation by Property Type

Figure 4 depicts faceted bar plots showing the distribution of violation statuses (0 for no violation, 1 for violation) across different property types. Each subplot corresponds to a different property type.

In general, property types of "Group Home," "High Rise," "Low Rise," and "Rooming House," have a higher proportion of violations (1). The type "Hotel & Motel" has balanced counts between 0 and 1, but violations (1) are slightly more frequent. These property types may require more attention or stricter inspections.

The other property types, especially "Detention" and "Nursing Home", show a much lower proportion of violations, suggesting fewer relative violation ratio.

Figure 4: Violation Status by Property Type

3 Model

The goal of the modelling strategy is twofold.

Firstly and most importantly, from the statistical standpoint, the goal is to estimate the the detection of violation from an inspection based on its time interval and the property type.
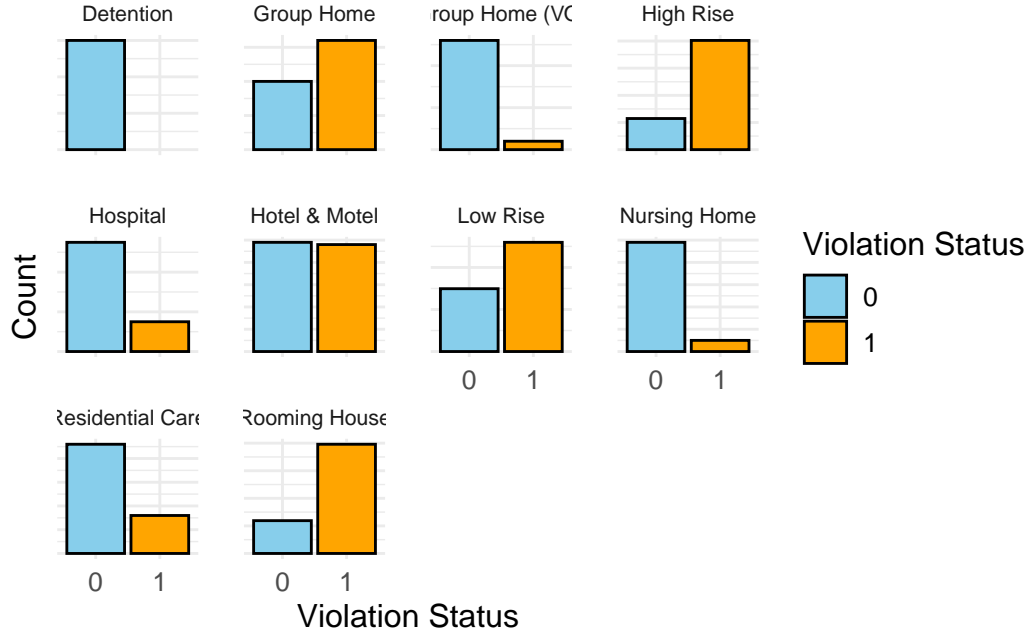
Secondly, from a more practical standpoint, by predicting which type of property with how long duration is more likely to be observed violations, more efficient inspection process could be constructed. Identifying certain property types that would more frequently violate the fire safety regulations could provide guidelines for interventions and help to establish targeted and focused actions towards them.

The model constructed in the paper is Bayesian logistics regression model designed to predict the violation status (denoted as violation) based on the fixed effect of inspection time interval (denoted as date_num) and property type (denoted as property_type). It is designed to predict the inspection results based on the two predictor variables.

Following sections define, explain and justify each model and the variables, as well as discuss underlying assumptions, potential limitations, software used to implement the model, and evidence of model validation and checking.

Background details and diagnostics of the model are included in Appendix C.

3.1 Model set-up

The Bayesian model was implemented using the R programming language (R Core Team 2023), specifically utilizing the **rstanarm** package of Goodrich et al. (2022). This package provides an interface for fitting Bayesian regression models using Stan, and the models are fit using the package.

Define y_i as the binary response variable for observation i. $y_i = 1$ indicates that the case has been observed violations, while $y_i = 0$ indicates that no violation occurs. It follows a Bernoulli distribution:

$$y_i = \text{Bernoulli}(p_i)$$

where p_i is the probability of a property that is observed to have violations of fire safety regulations.

Then define the inspection time interval for observation i as date_num and the property type as property_type.

The Bayesian model could be defined by the following mathematical expressions:

$$\text{logit}(p_i) = \beta_0 + \beta_1 \cdot \text{date}_i + \beta_2 \cdot \text{type}_i \tag{1}$$

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) \tag{2}$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$

where:

- $\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$ is the log-odds of p_i.
- β_0 is the intercept term of the Bayesian logistic regression, representing the baseline log-odds of a violation when 'date_num = 0' and 'property_type' is the reference category. Formula (2) specifies that β_0 follows normal distribution with mean equals 0 and variance equals 2.5^2.
- β_1 is the coefficient for predictor variable 'date_num', representing the change in log-odds of a violation per unit increase in 'date_num', the inspection time interval. Formula (3) specifies that β_1 follows normal distribution with mean equals 0 and variance equals 2.5^2.
- β_2 is the coefficient for predictor variable 'property_type', representing the change in log-odds of a violation per unit increase in 'property_type'. Formula (4) specifies that β_2 follows normal distribution with mean equals 0 and variance equals 2.5^2.

Formula (2), (3) and (4) specifies priors for the model. This prior reflects the assumption that coefficients are centered around 0, with a standard deviation of 2.5, allowing moderate variability while discouraging extreme values.

3.2 Model Assumptions and Limitations

The main, underlying assumptions and their limitations for the Bayesian model is discussed below:

Firstly, the model assumes that each observation is independent of others. This assumption suggests that violations at one property should not influence violations at another property unless explicitly modeled (e.g., through random effects).

If observations are clustered and the assumption is violated (for instance, multiple observations were conducted in the same region or by the same inspector), overestimation and underestimation of standard errors and outcomes may occur.

Secondly, the model assumes linearity in the logit. This assumption stated that the log-odds of the response variable ($\text{logit}(p_i)$) are a linear function of the predictors.

If violations change non-linearly over time, the model assumption would be violated, potentially resulting in misfit of the model.

Thirdly, the model assumes priors of $\text{Norma}(0, 2.5)$ for the coefficients. They were constructed on domain knowledge, and poorly chosen priors could make the estimation outcome biased.

Also, if new property_type levels are introduced, the model may not be no longer appropriate for the new situation, as fixed effects could not be feasible to compute estimates. Random effects, then, may be more appropriate.

3.3 Model Justification

Since the response variable is binary for each case and follows a Bernoulli distribution, it would be appropriate to use Bayesian logistic regression model to estimate it.

The predictor variable as 'date_num' is modeled as fixed effect in the model. There are various reasons for this. Firstly, As 'date_num' represents the number of days for the inspection takes place, it is a numeric variable. Modeling it as a fixed effect allows the model to estimate its direct, continuous influence on the response variable. Secondly, since it applies across all observations, regardless of property type, modeling it as a fixed effect could consider its impact consistently across the entire dataset. It also makes the coefficient of 'date_num' ($beta_1$) be easier to interpret, avoids adding complexity to the model.

The other predictor variable 'property_type' is also modeled as fixed effect. Since it is categorical variable, modeling it as fixed effect allow the model to estimate a separate coefficient

for each category, relative to a reference level. This would make it easier to compare between property types, in order to understand which types of properties would be observed to have violations. Moreover, since 'property_type' has only 10 levels, it is feasible to compute it as fixed effect and conduct estimation without overloading the model.

The prior of Normal$(0, 2.5)$ for every coefficients is also reasonable because it reflects prior knowledge that most predictors have small-to-moderate effects on the log-odds scale. This assumption would avoid overfitting to extreme values.

The above final model decision is made after considering several alternative models and variants. One alternative is a simple linear regression model (SLR), which assumes a constant linear relationship between the predictor (violation) and the response (date_num). Although the model is extremely straightforward and easy to interpret, it is far too simple for estimating and neglect the key variable of 'property_type'. More importantly, the response variable is binary, whereas SLR assumes a continuous response variable. MLR also assumes so, and therefore is considered and rejected because of the same reason.

Another alternative considered is the Bayesian model with natural splines (e.g. $f(\text{date}_i)$). While it could model complex, non-linear time trends and is appropriate for 'date_num', the expected model for this study does not primarily focus on exploratory analysis on non-linear relationships between date_num and violations. Non-linear functions would also overfit the data if not regularized properly.

While considering alternative Bayesian logistics model, the model with 'property_type' as fixed effect was also considered. This would capture variability across property types without explicitly estimating coefficients for each type, and does not provide explicit estimates for each property type. The model would be suitable if there are many property types, but the number of types is manageable to include it in the model as fixed effect.

Therefore, after considering many alternatives, the final model of Bayesian logistics model with 'date_num' and 'property_type' as fixed effect is chosen. It could directly provides explicit comparisons between property types without overfitting the model, and meet the focus of the research question.

4 Results

4.1 Model Summary

Our results of the Bayesian logistic regression model are summarized in Table 1. The table provides information of coefficients and main summary statistics.

The table firstly lists estimates of coefficients for all parameters. The first line suggests an intercept of -6.72 for the model, indicating very low baseline odds of a violation for the reference property type when date_num = 0. The second line presents that the estimated coefficient

for date_num is 0.02, which is very small and suggests that the inspection time interval has negligible effect on the violation outcome of observations.

Then the table shows the coefficients for every property type respectively. All property types show large positive coefficients, meaning that they increase the log-odds of violations and have higher risk of violation compared to the reference property type. For instance, for property type of "Group Home", it increases log-odds of violation by 5.88; for property type of "High Rise", it increases log-odds of violation by 6.21; for property type of "Hospital", it increases log-odds of violation by 4.67; etc. It could be observed that "Nursing Home" has the smallest coefficient, and thus has the smallest effect on violation. However, it still contribute positively to the log-odds.

The table also presents summary statistics of the model. One of the important results include $R^2 = 0.342$. R^2 is the proportion of the variation in the dependent variable that is predictable from the independent variables (Wikipedia contributors 2024). R-squared of 0.342 suggests that 34.2% of the variability in the response variable (violation) is explained by the model, which is relatively good for logistic regression. Moreover, the RMSE (Root Mean Squared Error) value of 0.32 shows the average deviation between predicted probabilities and observed outcomes. This value is low, indicates a satisfactory fit for the model.

4.2 Posterior Parameter Estimates

Table 2 shows more detailed information about the coefficients of each parameter. Column "Estimate" presents the mean posterior estimate for the effect of each parameter on the log-odds of the outcome (violation); "Std.Error" presents the standard deviationor the uncertainty in the estimate; "Conf.Low" and "Conf.High" are the lower and upper bounds of the 95% credible interval for the parameter, showing the range of plausible values.

Table 2: Posterior Parameter Estimates

Parameter	Estimate	Std.Error	Conf.Low	Conf.High
(Intercept)	-6.723	1.942	-10.962	-3.738
date_num	0.024	0.001	0.023	0.025
property_typeGroup Home	5.878	2.018	2.784	10.213
property_typeGroup Home (VO)	2.986	2.377	-0.880	7.429
property_typeHigh Rise	6.213	1.944	3.252	10.450
property_typeHospital	4.670	1.998	1.659	8.946
property_typeHotel & Motel	5.569	1.959	2.575	9.838
property_typeLow Rise	5.610	1.944	2.636	9.838
property_typeNursing Home	1.221	2.116	-2.036	5.562
property_typeResidential Care	4.450	2.018	1.379	8.752
property_typeRooming House	6.147	1.942	3.162	10.374

Table 1: Explanatory Models of Probability of Violation based on Inspection Time Interval and Property Type

	Model
(Intercept)	−6.72
	(1.94)
date_num	0.02
	(0.00)
property_typeGroup Home	5.88
	(2.02)
property_typeGroup Home (VO)	2.99
	(2.38)
property_typeHigh Rise	6.21
	(1.94)
property_typeHospital	4.67
	(2.00)
property_typeHotel & Motel	5.57
	(1.96)
property_typeLow Rise	5.61
	(1.94)
property_typeNursing Home	1.22
	(2.12)
property_typeResidential Care	4.45
	(2.02)
property_typeRooming House	6.15
	(1.94)
Num.Obs.	13 193
R2	0.342
Log.Lik.	−4948.270
ELPD	−4963.0
ELPD s.e.	105.2
LOOIC	9926.0
LOOIC s.e.	210.5
WAIC	9924.7
RMSE	0.32

Main interpretation for this code focuses on 95% confidence interval. In general, if the credible interval does not include 0, the effect of the variable is statistically significant, indicating that the model is confident about the relationship. Most of the parameters have credible intervals excluding value 0, and some intervals are far from 0, suggesting that there's strong evidence supporting their effect on the violation outcome. For instance, the property type of "High Rise" has 95% CI of [3.25, 10.45], showing that the model is confident about the significant effect of the increased contribution by high rise buildings to log-odds of violations.

On the other hand, a few parameters have 95% confidence interval including 0, such as property types of "Group Home", "Group Home (VO)", "Nursing Home". This represents that the effect of the variables on violations is not statistically significant, and the relationship is uncertain. Further data collection may be needed to better understand their relationship with violations.

Figure 5 visualizes the parameter estimate values and their credible intervals for easier presentation.
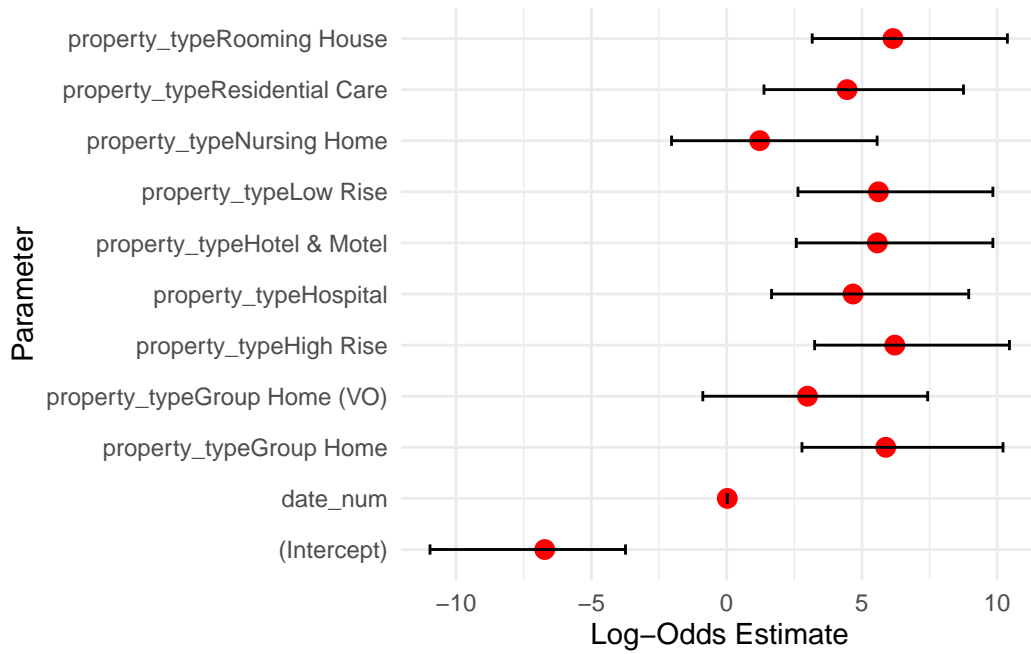

Figure 5: Parameter Estimates with 95% Credible Intervals

4.3 Distribution of Predicted Probabilities

Figure 6 plots the distribution of predicted probabilities for the model. X-axis represents the predicted probabilities (p_i) of the outcome for each observation in the dataset, and the values

of probabilities are bounded from 0 to 1. Y-axis represents the relative density of observations iwth particular predicted probabilities.The density curve on the graph represents the estimated density of predicted probabilities, and the curve's height reflects how concentrated the predicted probabilities are around specific values.

In general, the distribution shown by the plot has two peaks near 0 and 1, with very few probabilities in the intermediate range (0.25–0.75). This is a bimodal distribution, meaning that most observations are confidently predicted to either not have a violation ($p \approx 0$) or have a violation ($p \approx 1$). The sparse region in the intermediate range suggests that the model is making confident predictions, with minimal uncertainty for most observations.

The plot shows that the distribution has higher peak, which also means higher density, around $p = 1$. On the other hand, the peak around $p = 0$ is slightly lower. This indicates more confident predictions for observations with a high probability of a violation, meaning that the model predicts a larger proportion of observations with high confidence that a violation is likely to occur.

Therefore, the conclusion drawn from the graph is that: more observations are predicted to have a high probability (mostly $p > 0.9$) of being observed to have a violation.
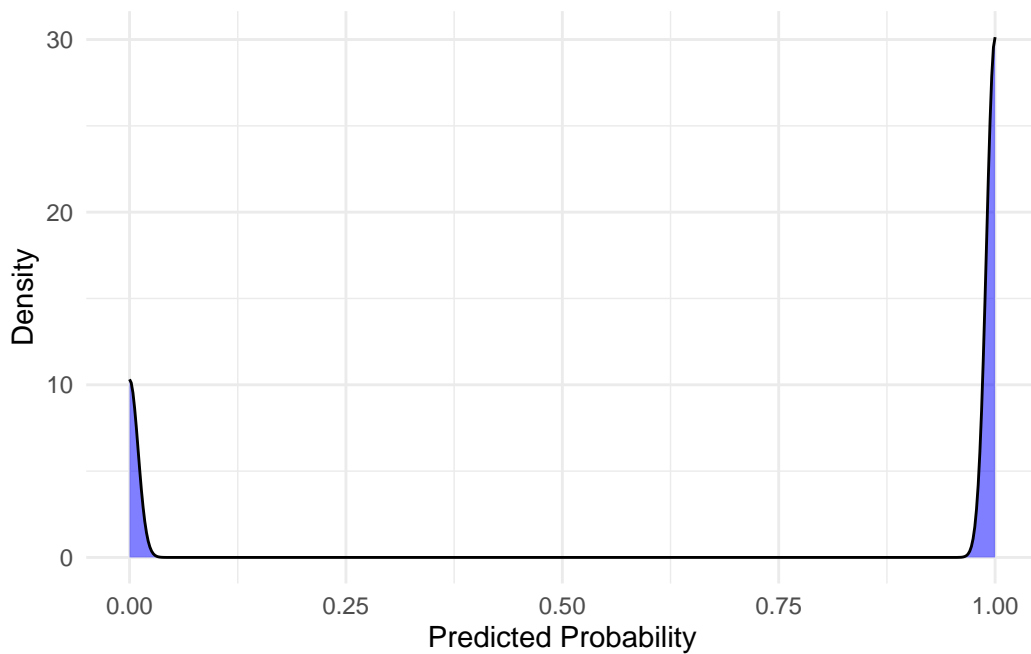

Figure 6: Distribution of Predicted Probabilities

5 Discussion

5.1 What is done in this paper?

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 What is something that we learn about the world?

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 What is another thing that we learn about the world?

5.4 Weaknesses and next steps

Firstly, the limitations in the process of measurement of the original dataset should be considered. Specifically, the limitations include:

- Observer Bias

Although inspectors conduct the violation detection procedure in reference to the same, standardized fire safety regulations (Ontario Fire Code, FPPA and Municipal Code, according to City of Toronto) to ensure consistency, their subjective judgment could still play a significant role in the examination process.

Various interpretations of the guidelines and different decision-making procedures made by the inspectors based on specific situations could impact the outcome of the inspection. Since the examinations have been thoroughly completed by human, the assessments could differ depending on the inspectors' focus and perspectives on certain regulations, influencing the observations recorded, even with clear rules.

- Data Completeness

The dataset does not capture all fire safety-related phenomena in Toronto, as it only contains the properties that have been inspected by the Inspectors from TFS. The properties that are not included in the list fail to provide any information, which would make the dataset not representative enough to conclude the situations of fire safety checks for the whole city.

Moreover, even for the properties that have been inspected by TFS, not all cases are recorded in the dataset. The raw dataset only includes 'closed' cases, which are specific cases where the inspection process has ended, and does not account for cases currently under review or in progress.

In other words, the properties whose initial observations have been made but further investigation or corrective actions are required, are excluded from the dataset. Therefore, the dataset does not represent the full scope of fire safety checks at any given time. The active cases would be neglected in analysis procedure, providing limited insights to identify fire safety concerns in real-time.

Appendix

A Surveys, Sampling, and Observational Data

A.1 Overview of the Idealized Fire Safety Resident Survey

Fire safety inspections in residential buildings provide a critical mechanism for ensuring compliance with fire safety regulations. However, the dataset presenting inspection outcomes and analysis based on it only reflects the violation status before the enforcement procedure ends. It fails to provide information about whether buildings will maintain compliance or revert to unsafe practices after the inspection process ends. Since frequent routine inspections are often expensive and therefore impossible, it is necessary to explore alternative methods for monitoring the fire safety status of the properties and evaluating the long-term effectiveness of previous inspections.

Introducing a resident survey fills these gaps by collecting the experiences and insights of those who are most directly affected by fire safety measures. Residents live in these buildings are familiar with the day-to-day reality of fire safety status, and could provide honest feedback and valuable perspective on whether inspections have resulted in improvements, or whether the properties conduct actions that violate regulations again.

This approach not only offers real-time updates of the violation status of the properties, but is also cost-effective. By incorporating resident surveys, the monitoring process becomes more sustainable, enhancing the overall understanding of fire safety compliance beyond the inspection period.

The fire resident survey is designed to achieve the following objectives:

- Assess Perceived Changes After Inspections: Determine if residents have noticed any improvements or changes in fire safety measures after the inspection and enforcement procedure ends, offering valuable feedback on the effectiveness of the interventions made.

- Identify Recurring Issues in Residential Buildings: Collect residents' observations and feedback on recurring violations of fire safety regulations in the properties, helping to highlight areas that may need further attention or action.

- Improve Policies Related to Fire Safety: Provide down-to-earth perspective from residents to improve regulatory practices related to fire safety, ensuring that policies are better aligned with the needs and experiences of those directly affected. This can also guide more effective resource allocation to address ongoing safety concerns.

- Arise Residents' Awareness of Fire Safety: Determine whether residents are aware that a fire safety inspection has taken place in their buildings, and to what extent residents recognizes the importance of fire safety. The survey, therefore, aimed to emphasize to residents that fire safety inspections are not just a property-level concern, but an issue

that affects everyone living in the building. By highlighting the collective responsibility for fire safety, the survey sought to encourage residents to take an active role in the ongoing maintenance and improvement of fire prevention measures, and to promote voluntary compliance with fire safety regulations, ensuring that all residents contribute to maintaining a safe living environment for everyone in the building.

A.2 Sampling Approach

A.2.1 Target Population

Target population refers to the population of the survey's interest. The target population for this fire safety feedback survey includes all residents of residential buildings in Toronto that have undergone fire safety inspections within the last 12 months, varying in property type and geographic location within the city.

This diverse target population ensures a broad range of perspectives on fire safety measures, reflecting various living conditions across different areas of the city.

A.2.2 Sampling Frame

The sampling frame or pool of potential respondents represents all the people in the population potentially to be sampled and invited to the survey (Stantcheva 2023). In this research, the sampling frame will be based on a list of properties from inspection records. The survey will be distributed by contacting the building management or tenant associations of those properties, who will then send the surveys to the residents for completion.

This process will be carried out in full compliance with privacy laws to ensure that residents' personal information is protected and that their participation remains confidential.

A.2.3 Stratified Random Sampling

To ensure the survey on fire safety compliance gathers reliable and representative data, a stratified random sampling method will be employed. The sampling approach follows the guidelines in "Survey Methodology" (Groves et al. 2004).

This approach divides the frame of population elements into meaningful subgroups, or "strata", which are "mutually exclusive groups of elements on a sampling frame" (Groves et al. 2004). Then, separate samples are drawn from each stratum, using randomization techniques (e.g. random number generators) to select participants within each stratum.

The strata for this study will include the following categories:

- Region: The geographic address of the property, ensuring representation from different areas within Toronto.

- Property Type: The occupancy type of the property, which includes various categories such as high-rise, low-rise, detention, group homes, group homes for vulnerable individuals (VO), hospitals, hotels & motels, nursing homes, residential care, and rooming houses.

This study plans to mount a survey collecting feedback of fire safety inspection of properties and expect that the results vary across different property types and regions in Toronto. Therefore, the strata will be defined based on key characteristics relevant to building types and location information.

A.2.4 Planned Sample and Sample Size

The planned or target sample is all the people that ideally to be included to complete the survey (Stantcheva 2023).

The sample size should be determined by both considering to achieve statistical significance and to meet resource constraints. After obtaining information about sample frame size, or the total number of residents across all strata, the proportion of the planned sample size allocated to each stratum should match the proportion of that stratum in the overall target population.

This approach ensures proportional representation from each stratum, reflecting the diversity of the population and improving the generalizability of the survey results. By maintaining this balance, the sample could provide a complete overview while meeting the constraints of available resources.

A.2.5 Potential Errors

- Coverage Error

Coverage error is the difference between the potential pool of respondents and the target population (Stantcheva 2023). In other words, the sampling frame does not fully represent the target population.

In this research, the surveys will be distributed to residents using online Google Forms. However, since the recruitment relies on digital methods, there is a potential risk of underrepresentation for properties in regions or with property types where a majority of residents lack internet access or digital literacy. This could result in skewed data, with certain populations being less likely to participate.

Additionally, since the sampling frame is based solely on official building management contact information, residents in informal or unregistered housing arrangements may be excluded from the survey. This limitation could further affect the representativeness of the sample, as it excludes individuals living in non-official housing setups who may also be affected by fire safety measures.

- Sampling Error

Sampling error refers to the difference between the planned sample and the sampling frame, that is, "the fact that you are drawing only a sample from the full sampling frame" (Stantcheva 2023).

In this research, despite randomization, it may still be possible that the selected individuals within a stratum do not represent the overall stratum accurately. This could occur due to certain individuals being overrepresented or underrepresented by chance. As a result, while randomization aims to minimize bias, sampling error could still influence the accuracy and precision of the findings.

- Non-response Error

Non-response error refers to the differences between the target sample and the actual sample (Stantcheva 2023).

In this research, the error could arise due to the residents not receiving or seeing the invitation of the survey. Since the surveys are distributed by building management or tenant associations through online platforms such as emails, there is a risk that the survey invitations may be overlooked or filtered into spam folders. Email-based distribution depends on residents regularly checking their inboxes, and not every resident has such habit. If the email is sent from an unfamiliar address or lacks clear identification, residents might dismiss it as junk or advertising mail, and ignore it entirely. Furthermore, some email systems automatically filter certain messages into spam folders, reducing the visibility of the survey and limiting participation.

Additionally, residents may ignore the survey invitation, potentially due to the lack of immediate rewards or incentives for participation. Without clear incentives, such as gift cards or other tangible rewards, residents may not feel motivated to take the time to complete the survey. This lack of incentive could result in lower response rates, particularly among residents who do not understand the survey's importance. As a result, the data collected might not fully capture the perspectives of all residents, leading to potential bias in the findings.

More detailed discussion and potential solutions would be discussed in Section A.3.2.

A.3 Survey

This survey will be conducted using Google Forms, which is an effective platform for data collection. It can be accessed by this link: Google Form Survey.

A.3.1 Survey Structure

Title:

Fire Safety Feedback Survey for Residential Properties in Toronto

Introduction:

We are conducting a survey to evaluate fire safety measures in your building following recent fire safety inspections conducted by Toronto Fire Services.

Your feedback will help us assess the effectiveness of inspections, identify any ongoing safety concerns, and improve fire safety standards.

Please note:

- Your answers will all be treated with complete confidentiality and anonymity.

- Participation in this survey is voluntary.

- We encourage you to provide honest and thoughtful responses.

- The survey is estimated to take about 10 minutes to complete.

- If you have any questions or concerns, feel free to contact kylie.gu@mail.utoronto.ca (Yunkai Gu).

Thank you for your valuable input. Your responses are essential for evaluating the outcomes and results of fire safety inspections by TFS.

Section 1: Residential Building Information

What is your city of residence?

- Toronto

- Other (End Survey)

What is your address? (Please specify to street number)

[Answer Box]

What is the occupancy type of your residential property?

- High Rise

- Low Rise

- Rooming House

- Group Home

- Group Home (VO)

- Hotels and Motels

- Detension

- Hospital

- Nursing Home

- Residential Cares

- Other [Answer Box]

- Unsure

What is your duration of residence? (Optional)

- Less than 1 year

- 1 to 3 years

- 3 to 5 years

- More than 5 years

Section 2: Awareness of Fire Safety Inspection

Were you aware that a fire safety inspection was recently conducted in your building?

- Yes

- No (Skip Next 2 Questions and Section 3)

If so, please indicate the source of awareness if available: (Optional; Select All That Apply)

- Notification from building management

- Signs posted in the building

- From neighbors/friends

- Other [Answer Box]

If so, please indicate the date of the inspection if available: (Optional; If you don't know, please leave the box blank.)

[Answer Box]

Section 3: Improvements and Changes After Inspection

Have you observed any improvements in fire safety facilities? (Select All That Apply)

- Fire alarms were tested or replaced.

- Emergency exits were cleared or repaired.

- Fire extinguishers were checked or replaced.

- Sprinkler systems were serviced.

- Other [answer box]

- No noticeable changes.

On a scale of 1 to 5, how effective do you think the changes have been in improving fire safety in your building?

[Scale 1-5]

Do you have any thoughts or suggestions regarding the fire safety inspection conducted by TFS? (Optional)

[Answer Box]

Section 4: Current Fire Safety Conditions

On a scale of 1 to 5, how confident are you in your building's current compliance with fire safety standards?

[Scale 1-5]

How is the fire alarm functionality in your building?

- Fire alarms are always functional.

- Fire alarms are sometimes not functional.

- Fire alarms are often not functional.

- Fire alarms are always not functional.

- Unsure.

- Other [Answer Box]

Are the fire alarms in your building regularly tested?

- Yes

- No

- Unsure

How is the emergency exit accessibility in your building?

- Emergency exits are always accessible.

- Emergency exits are sometimes blocked or locked.

- Emergency exits are often blocked or locked.

- Emergency exits are always blocked or locked.

- Unsure

- Other [Answer Box]

How is the fire extinguisher functionality in your building?

- Fire extinguishers are always functional.

- Fire extinguishers are sometimes not functional.

- Fire extinguishers are often not functional.

- Fire extinguishers are always not functional.

- Unsure.

- Other [Answer Box]

Are fire extinguishers easily accessible and appear to be in working condition in your building?

- Yes

- No

- Unsure

How is the sprinkler system functionality in your building?

- Sprinkler systems are always functional.

- Sprinkler systems are sometimes not functional.

- Sprinkler systems are often not functional.

- Sprinkler systems are always not functional.

- Unsure.

- Other [Answer Box]

Are the sprinkler systems in your building regularly maintained?

- Yes

- No

- Unsure

Have you noticed any fire safety concerns in your building?

- Yes

- No

If so, please describe the fire safety concerns you noticed. (Optional)

[Answer Box]

Section 5: Resident Experiences and Suggestions

Have you participated in any activities related to fire safety organized by the building management (e.g. a fire emergency preparedness training in your building)?

- Yes

- No

Have you received any notifications from the building management regarding fire safety inspection and/or awareness?

- Yes

- No

Do you have any suggestions for enhancing fire safety in your building? (Optional)

[Answer Box]

End Message:

Thank you for completing the Fire Safety Feedback Survey. Your responses will help us better evaluate the effectiveness of fire safety inspections by TFS, understand fire safety conditions in your residential building, and improve measures to protect residents like you.

If you have any further questions or concerns, please feel free to contact kylie.gu@mail.utoronto.ca.

A.3.2 Survey Design Considerations: Biases and Solutions

The designing process of the survey has drawn from the guidance in the paper "How to Run Surveys" (Stantcheva 2023). The following considerations address the potential biases in the proposed resident survey on fire safety compliance and suggest practical solutions:

- Response Bias

This type of bias occurs as respondents "try to take short cuts to minimize the cognitive load" (Stantcheva 2023). They may also be natural consequences of how people process information based on the serial position of alternatives and their visual presentation. Typical bias includes: moderacy biases, extreme response biases, and response Order biases. In other words, respondents may consistently select moderate or extreme options, or may be influenced by the order in which options are presented.

Stefanie Stantcheva suggests that randomizing the order of response options for questions with unordered (nominal) response options would be one of the solutions to the response biases. According to this recommendation, the fire safety feedback survey has been designed to randomly arrange the order of options for questions related to functionality and accessibility of fire safety facilities in the buildings in section 4. These questions followed a consistent and same structure and logic, making them particularly vulnerable to response biases due to the potential influence of the order in which options are presented. Randomizing the response options mitigates effects for such biases, enhancing the reliability of the data collected.

Stefanie Stantcheva also recommends avoiding long response lists, as they can overwhelm respondents and introduce bias. Therefore, most of the survey questions do not require respondents to read through extensive lists of alternatives and choices, unless necessary. This approach aims to enhance the clarity and make respondents feel less stressful in completing the survey.

Balanced scales, such as a 5-point Likert scale with clear and evenly distributed options, could also help reduce the negative effects of the response bias by providing a range of numeric, direct choices to the respondents. In accordance with this principle, the survey includes several scaled questions, such as "On a scale of 1 to 5, how effective do you think the changes have been in improving fire safety in your building?"

- Social Desirability Bias

Social Desirability Bias (SDB) typically stems from the desire of respondents to avoid embarrassment and project a favorable image to others, resulting in respondents not revealing their actual attitudes (Stantcheva 2023). Respondents may provide answers they believe are socially acceptable or align with what they perceive the surveyor expects.

Stefanie Stantcheva suggests that by assuring respondents of complete anonymity in the survey, the social desirability biases could be mitigated, since emphasizing the confidentiality of responses would reduce pressure for the respondents to conform.

According to this advice, the fire safety feedback survey has included special emphasis on anonymity at the beginning in the introduction section, clearly informing respondents that their answers "will all be treated with complete confidentiality and anonymity." This assurance is intended to encourage honest and open responses, helping to aviod situations where respondents provide socially desirable answers rather than their true opinions.

This fire safety feedback survey also uses neutral wording and avoids leading questions. For instance, instead of asking, "Do you think the building is safer now?", the survey asks, "Have you noticed any changes in safety measures?" This approach ensures that respondents are not influenced by biased or suggestive language, allowing them to provide more objective and accurate feedback based on their own observations and thoughts.

- Attrition or Non-response Bias

Not all residents will respond to the survey, and some may drop out midway, leading to missing data or biased outcomes. Attrition is a special case of non-response bias, which refers to the phenomenon of respondents dropping out of the study before completing it (Stantcheva 2023).

As recommended by Stefanie Stantcheva, the fire safety feedback survey is designed to provide a smooth respondent experience, ensuring that the survey is visually clear and user-friendly to reduce dropout rates. The survey only includes only necessary, non-complex questions to keep it concise and short, ensuring that respondents do not feel overwhelmed or impatient while completing it. By focusing on essential topics, the survey aims to encourage accurate responses without causing respondents to drop out.

Stefanie Stantcheva also suggests the survey provide strong incentives for the respondents, in order to encourage respondents to complete the survey thoroughly and maximize participation. This could include offering a variety of possible rewards that appeal to a broad range of people, preventing not only attrition but also selection biases.

If budget allows, offering cash or gifts as rewards could motivate the audience to participate, ensuring that the response rate remains high throughout the survey process.

B Additional data details

C Model details

C.1 Posterior Predictive Check

During the modeling process, posterior predictive checks (PPC) is used.

PPC is the comparison between what the fitted model predicts and the actual observed data, which validates whether the fitted model is compatible with the observed data. The aim is to detect if the model is inadequate to describe the data (Andrés López-Sepulcre ORCID iD 2024). This type of diagnostic visualizes the fit of the model by comparing the observed data (y) to the posterior predictive distributions (y_{rep}) generated by the model.

In this PPC plot, x-axis represents predicted probability of the binary outcome p_i, which is the probability of observing violation for a property during inspection; y-axis represents density or frequency of the predicted probabilities or aggregated counts.

The dark black line presents the distribution of the observed outcome (y), which is a smoothed estimate of the distribution of binary response 'violation'. On the other hand, the light blue lines represent the posterior predictive simulations, which are the density of the replicated data (y_{rep}) generated from the posterior predictive distribution of the fitted model.

Whether the black line matches well with the blue lines indicates how well the model fits the observed data. If the black line aligns well with the blue lines, it suggests that the model is performing well in capturing the actual data distribution, and vice versa.

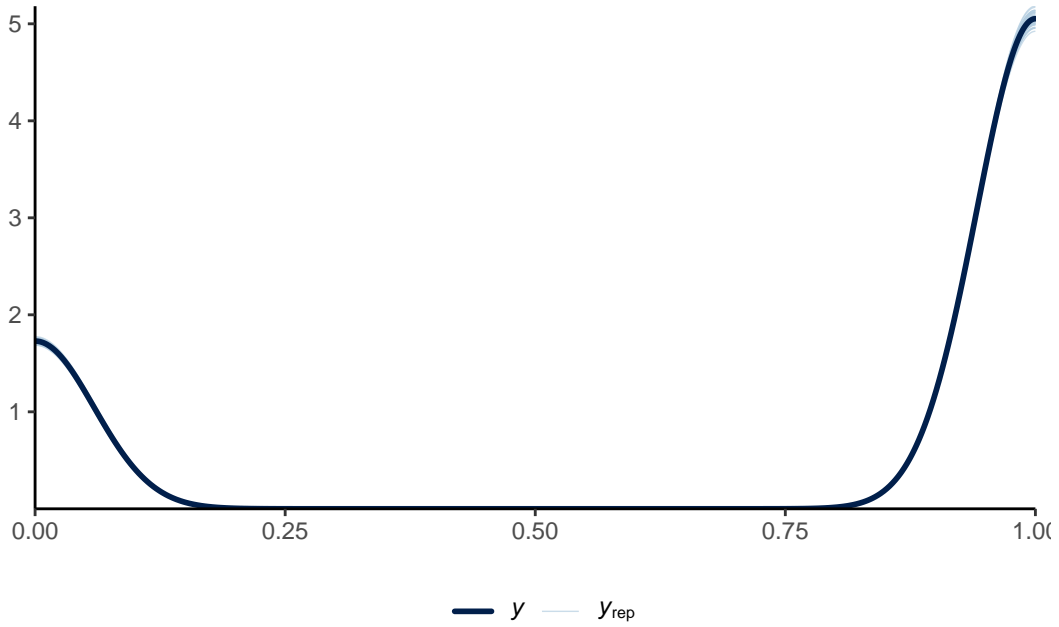Figure 7 shows the posterior predictive checks for the Bayesian model.

Figure 7: Posterior Prediction Check

The plot shows that the curves for y (observed data) and $y_r ep$ (posterior predictive distribution) align closely with each other, suggesting that the fix effects of two predictor variables capture the overall pattern of the observed data well, and the Bayesian model generates predictions consistent with the observed data. Lack of mismatches between curves mean that the model provides good fit of data. Near extreme probabilities of 0 and 1, two curves tightly follow each other, indicating that the model correctly predicts extreme outcomes.

C.2 Posterior vs Prior

Comparing the posterior distribution with the prior one is also necessary for model validation. It examines how the model fits and is affected by the data.

Prior distribution represents the beliefs about the parameter values before observing any data, which is constructed by our initial assumptions. In the plot, it is represented by horizontal lines with points centered at the prior mean (in this case, 0) in the right panel.

Posterior distribution reflects updated beliefs about the parameters after observing the data. It is represented by points (means or medians), which have colors representing the parameters respectively, and horizontal lines, which are credible intervals, derived from the data. The shorter horizontal lines, the narrower credible intervals, the more the reduced uncertainty for the parameter estimation after incorporating the data.
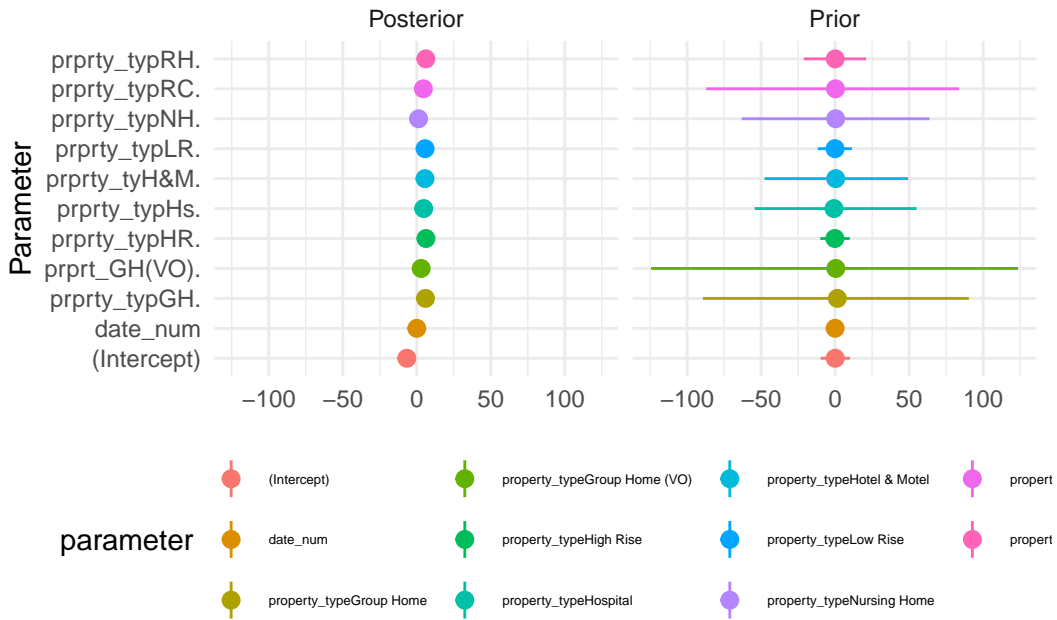

Figure 8: Comparing the Posterior with the Prior

Figure 8 compares the posterior with the prior distribution of the Bayesian model included in this paper, validating whether the data has a significant impact on the parameter estimates and whether the model is well-fit.

In general, each level shows a narrower posterior distribution than the prior, reflecting that the uncertainty decreases and the data provides substantial information to update the prior beliefs. Parameters with credible intervals far from zero in the posterior are strongly influenced by the

data and have meaningful effects, such as property type of 'Rooming House' (prprty_typRH) and 'Group Home' (prprty_typeGH).

Shown by the point representing 'Intercept', the posterior credible interval is narrower than the prior, suggesting strong data support for it. The negative posterior mean indicates that the baseline log-odds of a violation for the reference property type (when date_num = 0) is below 0 (i.e., a low probability of violation).

The posterior point for date_num is very narrow and centered near 0, suggesting that the data indicates little or no effect of inspection time interval on the log-odds of a violation. The credible interval includes 0, reinforcing the lack of a strong relationship between date_num and violations.

Each property type coefficient represents the deviation in log-odds from the reference category.

For instance, The points representing 'Rooming House' (prprty_RH), 'High Rise' (prprty_HR) and 'Group Home' (prprty_GH), have posterior distribution above 0, suggesting that rooming houses, high-rise buildings and group homes are more likely to have violations than the reference category. Narrow credible intervals indicate strong data support.

On the other hand, the point representing 'Nursing Home' (prprty_NH) have posteriors close to 0, suggesting their effects are small or negligible compared to the reference category. Narrow credible intervals still indicate strong data support.

The rest of dots representing different levels of 'property type' show various posterior means and credible intervals. Most of the points are close to 0 or above 0, providing information about effects of different property types respectively.

C.3 Diagnostics

C.3.1 Trace Plot

Trace plot shows samples from all the chains.

In the trance plot, x-axis represents the iteration numbers of the MCMC progress (in this case, up to 4000 iterations per chain), and y-axis displays the sampled values of the posterior distribution for each parameter. Each MCMC chain used during sampling is represented by a different color or line style..

If all chains appear to oscillate around the same mean, and overlap significantly, it suggests that the parameter estimates are stable around a consistent mean. It is also a good sign if the chains have overlap in most regions and move freely across the posterior distribution. There should also be no systematic patterns, trends or large jumps.
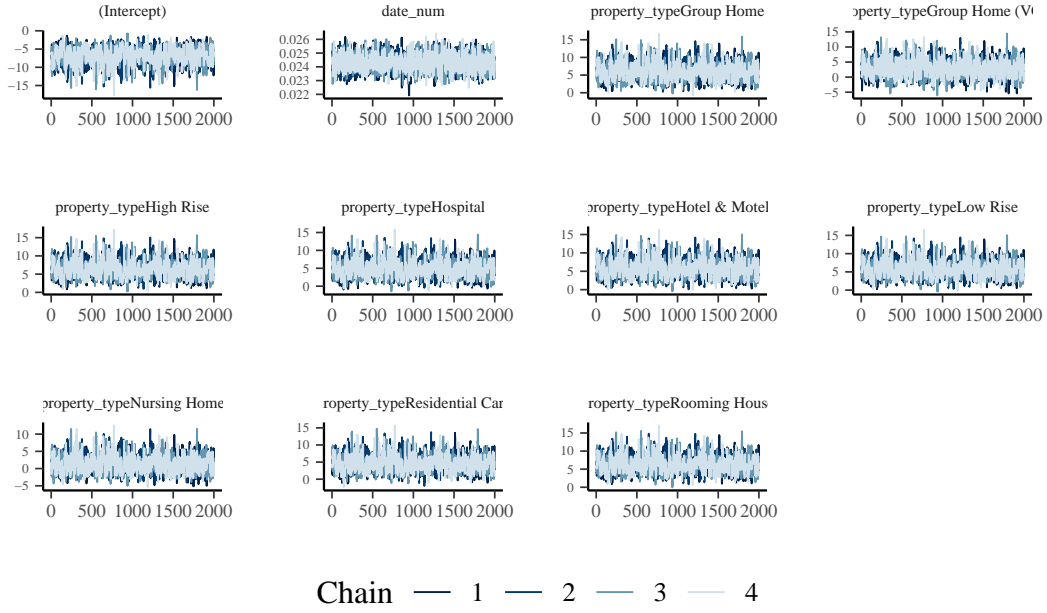
Figure 9: Checking the Convergence of the MCMC Algorithm by Trace Plot

Figure 9 shows the trace plot for the Bayesian model used in this paper. The chains for all parameters appear to mix well, with no signs of one chain dominating or exploring a different region of the parameter space. The parameters, especially 'Intercept' and 'date_num', show overlapping chains, which is a positive sign of convergence.

Moreover, most chains is stabilized and oscillates around a consistent mean. The consistent behavior across iterations indciates that the chains have reached stationarity. Also, there are no visible divergences or chain-specific trends, suggesting the sampling process is functioning correctly.

In general, the trace plots suggest that MCMC chains are converged, supported by good mixing, stationarity, and consistent sampling across chains, indicating that the posterior estimates for all parameters are likely reliable. Slight variability may be adjusted by more iterations, which aimed to gather additional independent samples.

C.3.2 R-hat Plot

$\hat{R}$ is a diagnostic measure that compares the variance within each chain to the variance between chains. This diagnostic plot is used to asses the convergence of MCMC chains.Ideally, the $\hat{R}$ should be close to 1, so that chains have mixed well and are sampling from the same posterior distribution. For $\hat{R} \leq 1.5$, we could say that convergence is acceptable.

Figure 10 is the R-hat plot for the Bayesian model used in this paper. The plot shows that the R-hat values for all parameters are below 1.05, suggesting that the chains have converged well and are sampling effectively from the posterior distribution. The posterior samples are likely reliable for inference, indicating that the results are robust and can be used for interpreting parameter estimates.
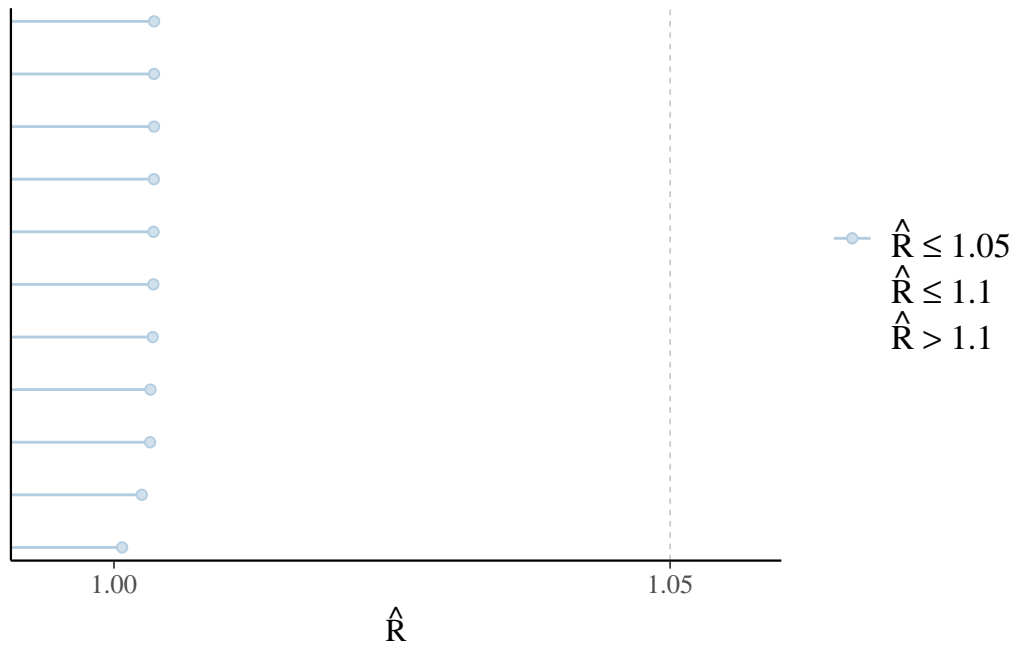

Figure 10: Checking the Convergence of the MCMC Algorithm by Rhat Plot

References

Alexander, Rohan. 2023. *Telling Stories with Data.* Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Andrés López-Sepulcre ORCID iD, Matthieu Bruneaux. 2024. "Posterior Predictive Checks." https://cran.r-project.org/web/packages/isotracer/vignettes/tutorial-100-posterior-predictive-checks.html.

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.

Bolker, Ben, and David Robinson. 2024. *Broom.mixed: Tidying Methods for Mixed Models.* https://CRAN.R-project.org/package=broom.mixed.

City of Toronto. 2024a. "Fire Inspection Results." https://www.toronto.ca/city-government/accountability-operations-customer-service/access-city-information-or-records/fire-inspection-results/#listing/eyJxdWVyeVN0cmluZyI6IiIsInRvcCI6MTB9/1.

———. 2024b. "Highrise Residential Fire Inspection Results." https://open.toronto.ca/dataset/highrise-residential-fire-inspection-results/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology.* Wiley Series in Survey Methodology. Wiley-Interscience.

Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes.* https://CRAN.R-project.org/package=RColorBrewer.

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.

Stantcheva, Stefanie. 2023. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." *Annual Review of Economics* 15: 205–34. https://doi.org/10.1146/annurev-economics-091622-010157.

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://CRAN.R-project.org/package=dplyr.

Wikipedia contributors. 2024. "Coefficient of Determination." https://en.wikipedia.org/wiki/

Coefficient_of_determination.