

Building Evaluation Scores in Downtown Toronto*

Age, Size, Property Type, and Location Affect Score with Property Standards

Yiyi Yao

December 14, 2024

This paper analyzes building evaluation scores from Toronto’s RentSafeTO program to understand the factors influencing compliance with property standards. Using a Bayesian regression model, we examined the relationships between building characteristics such as year built, number of units, property type, and geographic location. We found that newer buildings, larger properties, certain property type, and those in specific regions are more likely to achieve higher evaluation scores. These findings highlight the importance of modern construction, economies of scale in property management, and localized enforcement practices in ensuring housing quality. This report can guide policymakers and property managers in improving urban housing conditions.

Table of contents

1	Introduction	2
1.1	Estimand	3
2	Data	3
2.1	Data source	4
2.2	Data Features	4
2.3	Data Measurement	4
2.4	Data limitation	5
2.5	Data Methodology	5
2.6	Data Visualization	6

*Code and data are available at: [https://github.com/Yaoee111/condo_evaluation].

3	Model	10
3.1	Model set-up	10
3.2	Model justification	11
3.3	Model limitation	12
3.4	Model Validation	12
4	Results	13
5	Discussion	14
5.1	Achievements	14
5.2	About Building Maintenance and Standards	15
5.3	Limitation	15
5.4	Next steps	16
	Appendix	17
A	Data details	17
A.1	Raw data	17
A.2	Selected variables	17
B	Model details	18
B.1	Posterior predictive check	18
B.2	Markov chain Monte Carlo Convergence Check	19
C	Survey	19
C.1	Overview	19
C.2	Sampling Approach	21
C.3	Recruitment of Respondents	22
C.4	Survey Design	22
C.4.1	Link	23
	References	23

1 Introduction

Cities like Toronto face significant challenges in maintaining safe and high-quality housing, especially as populations grow and infrastructure ages. In 2017, the city launched RentSafeTO, a bylaw enforcement program designed to ensure multi-unit residential buildings meet essential property standards. This program carefully assesses buildings' exteriors, mechanical systems, and common areas. While RentSafeTO provides a systematic approach to evaluating building conditions, questions remain about what factors most impact assessment scores. Our research examines the primary elements that affect building evaluations across downtown Toronto,

studying how a building’s age, size, type, and location influence compliance with property standards.

Using a Bayesian linear regression model, we analyzed the relationships between building characteristics and evaluation scores. Our findings highlight several important observations: Newer buildings and larger properties tend to receive higher evaluation scores. This is likely because of modern construction techniques and better management resources. Geographic location also plays a significant role, with some neighborhoods consistently performing better or worse than others.

These results provide meaningful guidance for policymakers and property managers. By understanding the factors that influence building compliance, we can target maintenance efforts more strategically, refine policy interventions, and improve the consistency of property evaluations.

Ultimately, this research aims to support better housing conditions and more effective enforcement of property standards across Toronto. By identifying the main drivers of building performance, we can work towards creating safer, more reliable rental housing for the city’s residents.

The remainder of this paper is organized into distinct sections. Section 2 provides an overview of the dataset used in our analysis, including tables and graphs to illustrate the characteristics of buildings and evaluation scores. Section 3 outlines the Bayesian model employed, explaining its formulation and justification. Section 4 presents the findings, using tables and visualizations to highlight the relationships between building characteristics and evaluation scores. Finally, Section 5 offers a detailed discussion of the results, focusing on how factors such as construction year, building size, and geographic location influence compliance with property standards.

1.1 Estimand

The estimand in this paper is the evaluation score of apartment buildings as a measure of their compliance with property standards under the RentSafeTO program. However, it is challenging to assess the exact evaluation scores for all apartment buildings in Toronto, as not all buildings are included in the RentSafeTO dataset. For instance, smaller buildings or those outside the program’s jurisdiction are excluded, limiting the scope of evaluation. Additionally, variability in inspections and data gaps may impact the accuracy of the scores. Therefore, in this paper, we estimate the estimand using a Bayesian linear regression model fitted with a subset of the RentSafeTO dataset.

2 Data

We use the statistical programming language R (R Core Team 2023) with packages `tidyverse` (Wickham et al. 2019), `rstanarm` (Brilleman et al. 2018), `knitr` (Xie 2014), `ggplot2` (Wick-

ham 2016), `sjPlot` (Lüdecke 2023), `patchwork` (Pedersen 2020), `ggdist` (Kay 2023), and follow Alexander (2023).

2.1 Data source

Data for this research comes from RentSafeTO program provided by Open Data Toronto (Gelfand (2020)). This dataset contains building evaluation scores for apartment buildings registered with the RentSafeTO program in Toronto until 2023. RentSafeTO (City of Toronto (n.d.)), established in 2017, is a bylaw enforcement initiative designed to ensure that apartment buildings with three or more storeys or at least 10 units meet required maintenance standards. Alongside evaluation details, the dataset provides information about the buildings, including their address, year built, and geographical coordinates. The RentSafeTO dataset was selected because it provides a comprehensive overview of apartment building maintenance standards in Toronto.

2.2 Data Features

The original RentSafeTo dataset, which shows in Appendix Section A, contains 3573 data entries and many variables. Since we only interested in data in downtown Toronto, this report will only explore and analyze through several data features. We chose these 8 variables: RSN, WARD, WARDNAME, YEAR.BUILT, PROPERTY.TYPE, CONFIRMED.STOREYS, CONFIRMED.UNITS, CURRENT.BUILDING.EVAL.SCORE. After data cleaning, there are 979 rows of data remained. Details of these selected variables are also in Section A. We could use such features to predict building condition in downtown Toronto.

2.3 Data Measurement

The evaluation scores in the dataset represent the extent to which buildings meet maintenance standards based on inspections. These inspections cover aspects such as building exteriors, mechanical systems, and common areas, and the results are recorded as numerical scores ranging from 0 to 100. The variable CURRENT.BUILDING.EVAL.SCORE is derived by combining reactive inspections (initiated by tenant complaints) and proactive inspections (routine checks conducted by the city). While the program’s structured evaluation process aims to ensure consistency, scores may still be influenced by inspectors’ subjective judgments or variations in enforcement across wards.

The variable YEAR.BUILT indicates the year each building was constructed, reflecting the age of the property. This information is sourced from municipal property records. However, older buildings may have incomplete or inaccurate records, which could affect the precision of this variable. The assumption underlying this variable is that newer buildings are likely

to comply better with modern construction standards, which could positively influence their evaluation scores.

CONFIRMED.UNITS represents the number of units within each building, as recorded in municipal databases. This variable reflects the size of the property and its capacity to house tenants. Larger buildings with more units might have better resources for property management, but they could also face challenges in maintaining compliance due to their scale. Similarly, the variable CONFIRMED.STOREYS captures the number of storeys in a building, which can indicate its vertical size and complexity. Both variables rely on official records, which are generally reliable but may occasionally contain errors or outdated information due to changes in building usage or renovations.

The variable PROPERTY.TYPE.CODE categorizes buildings into private properties, social housing, or Toronto Community Housing Corporation (TCHC) properties. This classification is based on ownership and management type, which can significantly influence the availability of resources for maintenance and compliance. Private properties typically have more financial resources for upkeep, while social housing and TCHC properties may face funding constraints, which could affect their evaluation scores.

WARD indicates the geographic location of the building, based on Toronto’s ward boundaries. This variable allows us to capture local factors such as neighborhood policies, socioeconomic conditions, or differences in enforcement practices. Geographic location is an important consideration because it reflects the broader context within which a building operates, including variations in infrastructure quality and local government support. The dataset associates each building with its respective ward at the time of inspection, ensuring geographic consistency.

2.4 Data limitation

The data represents evaluations of buildings registered under RentSafeTO, so it does not cover all apartment buildings in Toronto. This limits its ability to represent the full state of rental housing in the city. Additionally, the evaluation process can be affected by variability in inspector judgments, leading to potential inconsistencies in scores. Gaps in the data, such as 0 scores or blanks, further complicate analysis and reduce the dataset’s completeness. Moreover, the dataset primarily focuses on physical and structural conditions, neglecting tenant perspectives or broader social factors that may influence overall building quality. These limitations highlight the need to interpret the data with caution when making broader inferences.

2.5 Data Methodology

The original dataset contains information on all registered apartment buildings in the Toronto area, along with the data needed for their evaluation and scoring. This comprehensive data includes various attributes related to building conditions and maintenance. However, since our primary research objective is to predict the conditions of apartment buildings specifically

in the downtown area, we narrowed our focus to the variables most relevant to this goal. To achieve this, we carefully selected key features, like YEAR.BUILT, PROPERTY.TYPE, CONFIRMED.UNITS, CURRENT.BUILDING.EVAL.SCORE, and clean the original dataset, ensuring it aligns with the scope and requirements of our analysis.

2.6 Data Visualization

Figure 1 illustrates the relationship between CURRENT.BUILDING.EVAL.SCORE and YEAR.BUILT. Each point represents a building, with its year of construction on the x-axis and its evaluation score on the y-axis. The plot includes a trend line generated through linear regression to highlight any general trends. The positive trend line suggests that buildings constructed more recently tend to have slightly higher evaluation scores, although the relationship is not very strong. This could indicate that newer buildings are generally better maintained or constructed to meet higher standards. However, there is considerable variability in the scores for buildings of all ages, as shown by the spread of points around the trend line.

Outliers in the older buildings with high scores could reflect cases where older properties have been well-maintained or renovated. Conversely, some newer buildings with lower scores might indicate specific issues despite their recent construction.

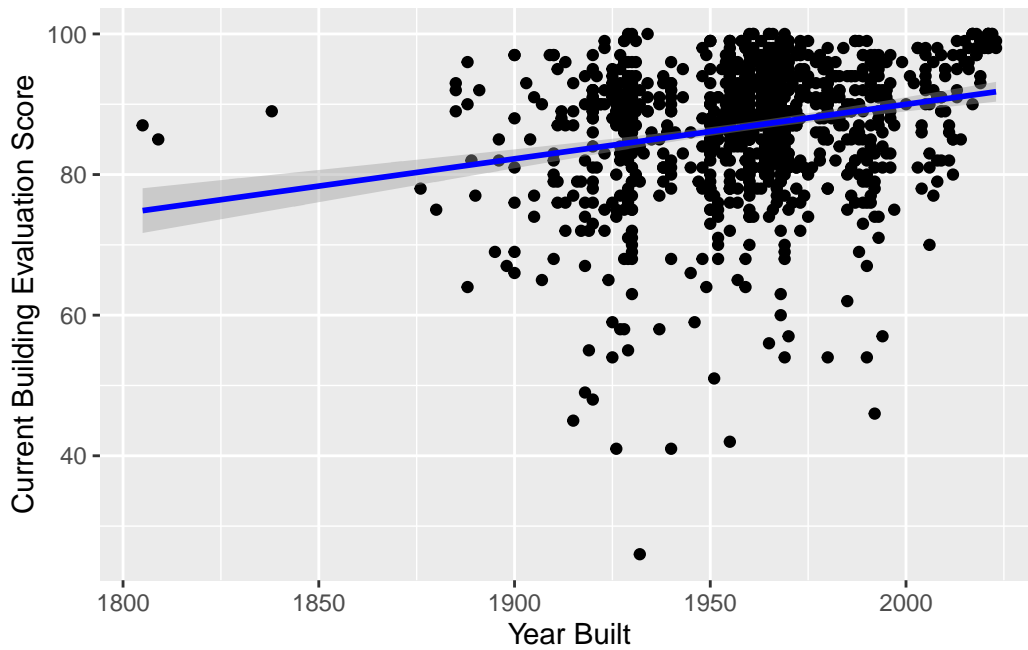


Figure 1: Relationship between Evaluation Score and Year Built

Figure 2 compares CURRENT.BUILDING.EVAL.SCORE across different PROPERTY.TYPE categories. Each box represents the distribution of scores for a specific property type (e.g., Private, Social Housing, TCHC). The graph shows the median score as well as the range of scores within each category. All property types have relatively high median scores, but slight variations can be observed. TCHC and Social Housing buildings have marginally higher medians than Private properties, suggesting potentially better maintenance or stricter compliance standards. Private properties exhibit the largest range of scores, with several outliers at the lower end. This indicates that some private buildings may significantly underperform in terms of maintenance and compliance. Social Housing and TCHC properties have narrower score distributions, implying more consistent maintenance or management practices. The lower outliers in Private properties highlight a subset of buildings that fall well below the typical evaluation scores, potentially requiring closer attention or intervention.

The graph suggests that Social Housing and TCHC properties are generally more consistent in maintaining standards, while Private properties show greater variability and occasional poor performance.

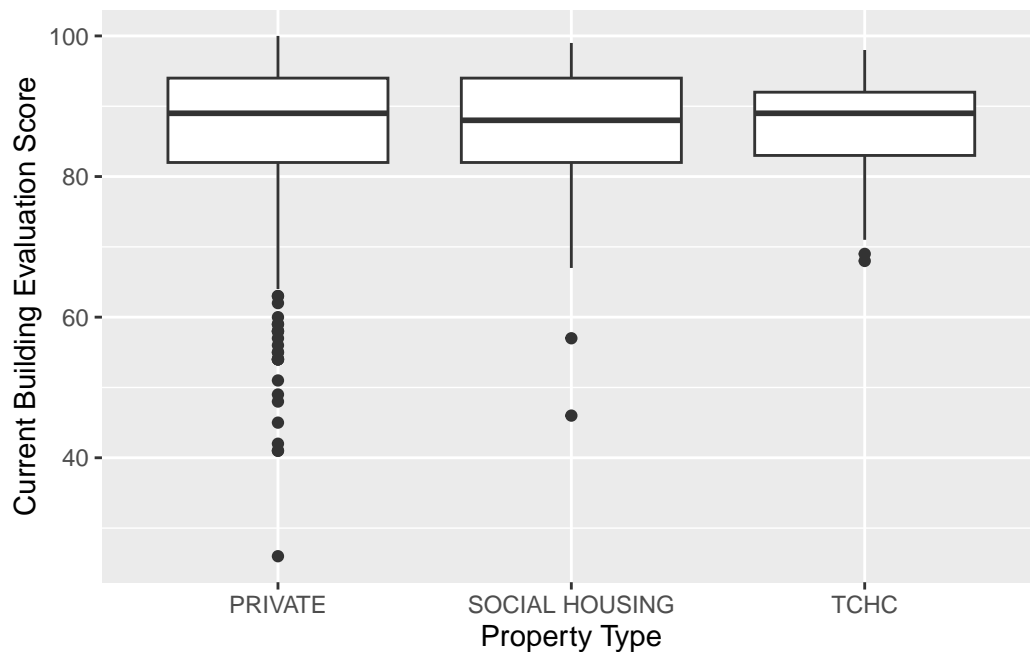


Figure 2: Evaluation Score by Property Type

Figure 3 explores the relationships between building evaluation scores, the number of units, and the number of storeys in Toronto apartment buildings. Graph A illustrates the relationship between the evaluation score and the number of units in a building, while Graph B highlights the relationship between the evaluation score and the number of storeys.

Graph A shows a general trend in the relationship between the number of units and evaluation

scores. Buildings with more units tend to have a wide range of scores, suggesting that the number of units alone does not directly predict a building’s quality or condition. However, there are clusters of higher scores associated with moderate unit counts, which might indicate a balance between resource allocation and maintenance efforts in mid-sized buildings. The linear regression line provides an overview of the trend, but the scatter suggests considerable variability, indicating other factors may also play significant roles in determining building scores.

Graph B examines the relationship between the number of storeys and building evaluation scores. Buildings with fewer storeys generally display consistent scores, often in the higher range, likely due to simpler maintenance and construction dynamics. Conversely, taller buildings show more variability in scores, with some achieving high evaluations while others fall significantly lower. The regression line suggests a modest trend, but, similar to the first graph, the data points indicate substantial variability within different building heights.

Together, Figure 3 highlight key patterns in how building size, as measured by units and storeys, relates to evaluation scores. While larger and taller buildings often exhibit more variability in their scores, mid-sized and low-rise buildings show more consistent evaluations. These findings suggest that building design and scale may influence maintenance practices and quality assessments.

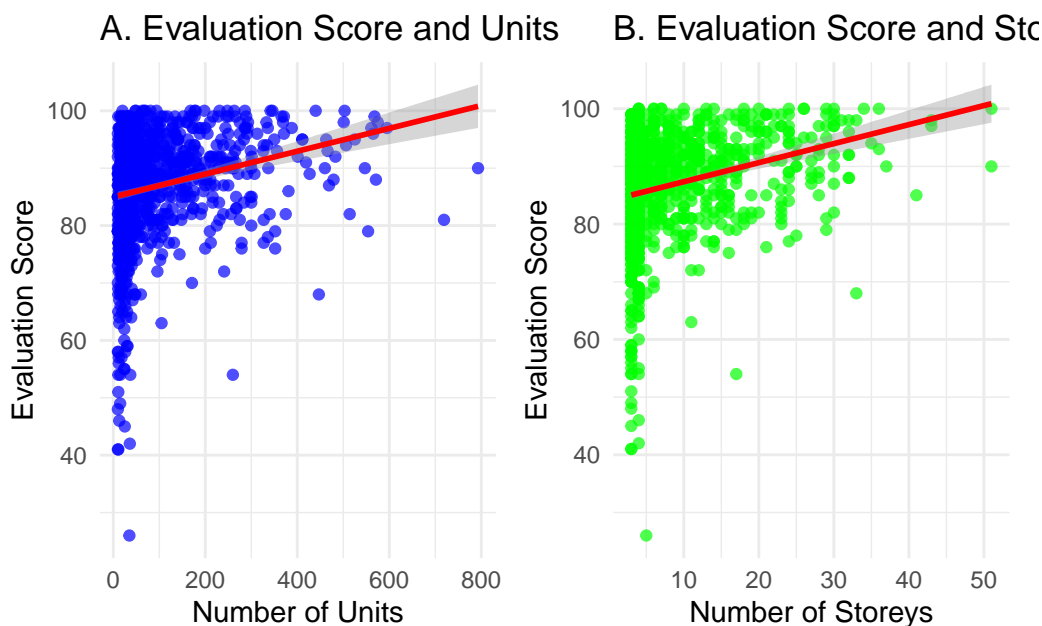


Figure 3: Relationships Between Evaluation Score, Units, and Storeys

Figure 4 shows the distribution of current building evaluation scores across six wards in Toronto, represented by their respective numbers on the x-axis. Each ward corresponds to a specific area: Ward 9 (Davenport), Ward 10 (Spadina-Fort York), Ward 11 (University-

Rosedale), Ward 12 (Toronto-St. Paul's), Ward 13 (Toronto Centre), and Ward 14 (Toronto-Danforth). The y-axis represents the evaluation scores, providing insights into the quality or condition of buildings in each ward.

Wards such as Spadina-Fort York (10) and Toronto Centre (13) show relatively higher median scores, suggesting overall better building conditions compared to wards like Davenport (9), which has a slightly lower median score.

Red points below the whiskers indicate buildings with evaluation scores significantly below the rest. For instance, Davenport (9), University-Rosedale (11), and Toronto-St. Paul's (12) exhibit several outliers with very low scores, potentially indicating localized issues with specific buildings in these areas.

Wards like Toronto Centre (13) and Spadina-Fort York (10) exhibit tighter score distributions, suggesting more uniform building quality across the ward. In contrast, wards like Toronto-St. Paul's (12) and University-Rosedale (11) display a wider spread, reflecting greater variability in building conditions.

The majority of the scores for all wards fall between 70 and 90, with exceptions in the form of outliers. This indicates a relatively consistent baseline of building conditions across these areas, though outliers and variability suggest that some buildings deviate significantly from the norm.

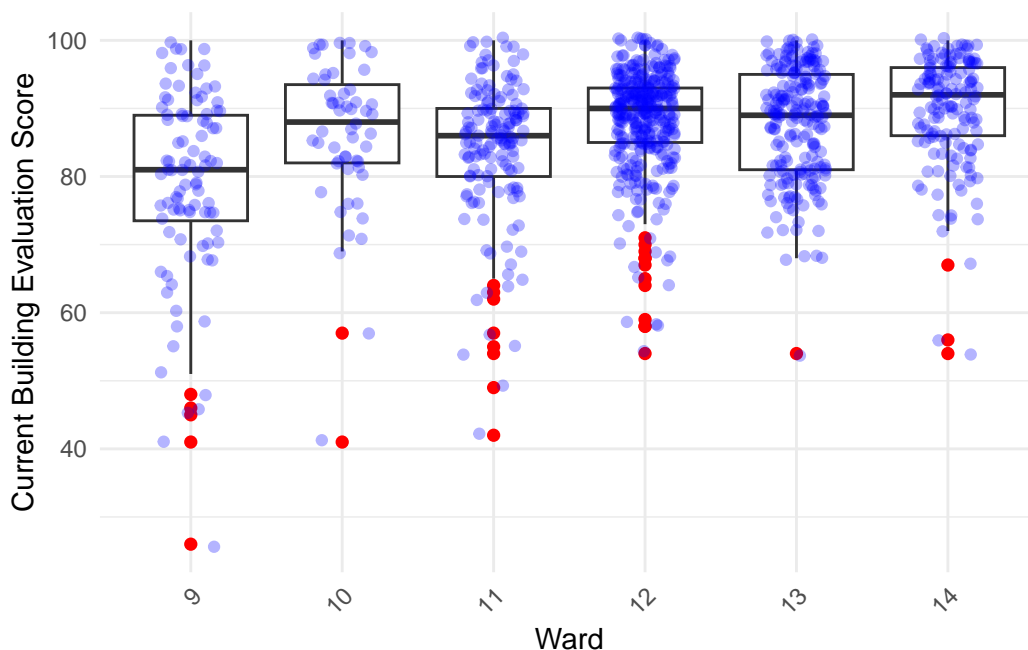


Figure 4: Distribution of Building Evaluation Scores Across Wards

3 Model

In our analysis, we utilized a Bayesian logistic regression model to examine the relationship between apartment evaluation and other building factors — built year, number of units, location (ward), and property type. Background details and diagnostics are included in Appendix Section B.

3.1 Model set-up

The model is formulated as follows:

$$\begin{aligned} \text{CURRENT.BUILDING.EVAL.SCORE}_i | \mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) \\ \mu_i &= \beta_0 + \beta_1 \text{YEAR.BUILT}_i + \beta_2 \text{PROPERTY.TYPE.CODE}_i + \beta_3 \text{CONFIRMED.STOREYS}_i \\ &\quad + \beta_4 \text{CONFIRMED.UNITS}_i + \beta_5 \text{WARD}_i \\ \beta_{0,1,\dots,5} &\sim \text{Normal}(0, 2.5) \\ \sigma &\sim \text{Exponential}(1) \end{aligned}$$

In this model, $\text{CURRENT.BUILDING.EVAL.SCORE}_i$ represents the evaluation score of building i . The mean evaluation score, μ_i , is modeled as a linear combination of the intercept β_0 , the effect of the year the building was built $\beta_1 \times \text{YEAR.BUILT}_i$, the effect of the building's property type $\beta_2 \times \text{PROPERTY.TYPE.CODE}_i$, the effect of the number of storeys $\beta_3 \times \text{CONFIRMED.STOREYS}_i$, the effect of the number of units $\beta_4 \times \text{CONFIRMED.UNITS}_i$, and the effect of the ward $\beta_5 \times \text{WARD}_i$. Here, YEAR.BUILT_i is a continuous predictor, while $\text{PROPERTY.TYPE.CODE}_i$ and WARD_i are categorical predictors that capture variability due to building type and location, respectively.

The intercept β_0 represents the baseline evaluation score when all predictors are at their reference or zero levels. The coefficients $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ represent the effects of each predictor on the evaluation score. For example, β_1 captures how the evaluation score changes with each additional year since the building was constructed, while β_3 quantifies the effect of an additional storey on the score. The categorical variables, $\text{PROPERTY.TYPE.CODE}_i$ and WARD_i , allow the model to account for systematic differences in scores due to property type and location.

Each coefficient $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ is assigned a weakly informative prior, $\text{Normal}(0, 2.5)$, which reflects a prior belief that most plausible values for the effects are moderate, but the prior allows for larger effects if supported by the data. This regularization helps prevent overfitting, especially in cases where predictors may be correlated or sample sizes are limited. The residual standard deviation σ , representing the variability in evaluation scores not explained by the predictors, is modeled using an $\text{Exponential}(1)$ prior, favoring smaller values of variability while still accommodating the possibility of larger residual noise.

This Bayesian framework provides several advantages. It captures the uncertainty in parameter estimates by modeling them probabilistically and allows for regularization through the use of weakly informative priors. By including predictors like $\text{PROPERTY.TYPE.CODE}_i$ and WARD_i , the model accounts for potential heterogeneity in evaluation scores arising from building type and location-specific factors, such as local policies, maintenance practices, or regional infrastructure quality. Compared to traditional frequentist approaches, this Bayesian model ensures robust inference by explicitly modeling uncertainty and leveraging prior information, resulting in more reliable and interpretable results for analyzing building evaluation scores in Toronto.

We run the model in R (R Core Team 2023) using the `rstanarm` (Brilleman et al. 2018) package.

3.2 Model justification

Regarding the relationship between building characteristics and evaluation scores, we hypothesize that newer buildings are more likely to receive higher evaluation scores. This assumption is based on the premise that newer buildings typically adhere to updated construction codes, utilize modern materials, and are less affected by wear and tear compared to older buildings. In contrast, older buildings may experience structural or aesthetic degradation over time, leading to lower scores unless significant maintenance or renovation efforts are undertaken.

The number of storeys is expected to positively correlate with evaluation scores. Taller buildings often cater to higher population densities and are generally built to more stringent standards due to the complexities associated with high-rise construction. Furthermore, these buildings may attract greater investment in terms of maintenance and infrastructure to ensure compliance with regulations and to retain their functionality over time.

Similarly, we anticipate that the number of units in a building will positively impact evaluation scores. Larger buildings with more units often benefit from economies of scale, allowing for structured property management systems, dedicated maintenance teams, and better allocation of resources. These factors can contribute to consistent upkeep and adherence to property standards. Conversely, smaller buildings may face resource constraints or lack formal management systems, which could result in lower scores.

The inclusion of property type as a predictor captures differences between building categories, such as private properties, social housing, and TCHC buildings. Each category may face unique challenges and standards based on ownership structure, funding, and operational priorities, which can influence evaluation scores. For example, private buildings may have more financial resources for maintenance, while social housing may face budgetary constraints that impact building conditions.

Finally, the inclusion of ward as a predictor accounts for geographic variability in evaluation scores. Wards represent different neighborhoods in Toronto, each potentially influenced by

unique local factors such as socioeconomic conditions, zoning policies, infrastructure quality, and inspection practices. By modeling the effect of wards, the model ensures that unobserved geographic and policy-related factors are captured, improving the robustness of the predictions.

This Bayesian framework was chosen because it allows for the integration of prior knowledge while providing robust uncertainty estimates for each parameter. The weakly informative priors help prevent overfitting, especially when predictors like `PROPERTY.TYPE.CODEi` and `WARDi` introduce categorical variability. By including these key building characteristics and accounting for geographic variability, the model offers a nuanced understanding of how structural and contextual factors influence building evaluation scores.

3.3 Model limitation

While the model provides valuable insights, it has several limitations and relies on assumptions that may not hold in all contexts. First, the assumption of a linear relationship between predictors (e.g., ‘`YEAR.BUILT`’ and ‘`CONFIRMED.UNITS`’) and the evaluation score may oversimplify complex interactions. For instance, the effect of a building’s age on evaluation scores may vary depending on factors like renovation history or maintenance practices, which are not captured in the dataset. Second, the model assumes that the random effect for ‘`WARD`’ fully accounts for geographic variability, but unobserved confounders such as neighborhood socioeconomic status or differences in enforcement rigor may influence evaluation outcomes. Additionally, the use of standardized continuous predictors assumes that their scales are comparable, which may obscure meaningful differences in their effects. Finally, the model is context-specific and tailored to Toronto’s RentSafeTO program, limiting its generalizability to other regions with different building standards, inspection processes, or housing markets. These limitations suggest the need for cautious interpretation and highlight areas for potential refinement in future analyses.

3.4 Model Validation

For posterior predictive checks, in Figure 5, the alignment of the posterior predictive distribution with the observed building evaluation scores suggests that the model accurately captures the patterns in the data. This indicates the reliability of the model in representing the relationship between building characteristics and evaluation scores. Additionally, Figure 6 compares the posterior to the prior distributions, revealing significant parameter shifts, particularly for ‘`YEAR.BUILT`’ and ‘`CONFIRMED.UNITS`’. These shifts demonstrate that the data provided strong evidence to update the priors, though minimal changes for some parameters may reflect either limited data informativeness or initial alignment with the priors.

For Markov chain Monte Carlo (MCMC) convergence checks, Figure 7 shows trace plots for the intercept and key predictors. These plots indicate good mixing and consistent exploration of

the posterior distribution, suggesting reliable convergence for these parameters. Furthermore, Figure 8 provides the Gelman-Rubin diagnostic (\hat{R}), confirming that all \hat{R} values are below 1.05. This ensures that the chains have converged and the samples are representative of the posterior distribution.

Additional details about the model and validation processes are provided in Appendix Section B.

4 Results

Our results are summarized in Table 1. The analysis found the relationships between building characteristics and evaluation scores under Toronto’s RentSafeTO program. To establish a baseline, the model uses reference groups for categorical variables. The intercept (-27.459) represents the estimated evaluation score for buildings in the reference ward (likely Ward 9) with fewer than 10 units and without other distinguishing features.

The positive coefficient for `YEAR.BUILT` (0.055) indicates that newer buildings generally achieve higher evaluation scores. This suggests that each additional year of construction is associated with a small increase in the score.

The coefficient for `CONFIRMED.STOREYS` (0.234) shows that taller buildings tend to have slightly higher scores, possibly because such properties are better maintained or built to stricter design requirements. Similarly, the number of units (`CONFIRMED.UNITS`) has a small positive coefficient (0.002), suggesting that larger buildings may benefit from more structured management practices and resources, contributing to better compliance.

For `PROPERTY.TYPE.CODE`, the negative coefficients (-0.547 for Code 2 and -0.301 for Code 3) indicate that certain property types, social housing and TCHC, tend to have lower scores compared to private properties. This difference may stem from funding or maintenance constraints associated with non-private ownership types.

Geographic location, `WARD`, shows significant variability. For example, buildings in Ward 12 (6.189) and Ward 14 (6.968) have higher scores compared to the reference ward. This suggests localized factors, such as better enforcement of property standards or more resources allocated for building maintenance. In contrast, some wards show less pronounced increases, which might reflect differences in policies, socioeconomic conditions, or inspection practices across neighborhoods.

In summary, newer buildings generally perform better. Larger buildings, with more units and storeys, tend to achieve higher scores. Private properties have higher scores compared to social housing and TCHC. Geographic location also matters, as buildings in certain wards consistently score higher. These factors together provide a clear picture of what influences compliance with property standards in Toronto.

Table 1: Explanatory model Building Evaluation

Predictors	Estimates	Evaluation Score	
		std. Error	CI (100%)
(Intercept)	-27.459	23.896	-131.556 – 55.213
YEAR.BUILT	0.055	0.012	0.012 – 0.108
PROPERTY.TYPE.CODE:	-0.547	0.991	-3.706 – 2.969
PROPERTY.TYPE.CODE2			
PROPERTY.TYPE.CODE:	-0.301	0.935	-3.555 – 3.308
PROPERTY.TYPE.CODE3			
CONFIRMED.STOREYS	0.234	0.076	-0.055 – 0.497
CONFIRMED.UNITS	0.002	0.005	-0.015 – 0.020
WARD: WARD10	1.278	1.270	-2.609 – 5.165
WARD: WARD11	1.730	0.985	-1.586 – 5.180
WARD: WARD12	6.189	0.849	2.969 – 9.238
WARD: WARD13	4.527	0.910	1.305 – 7.892
WARD: WARD14	6.968	0.990	3.620 – 10.521
Observations	979		
R ² Bayes	0.154		

5 Discussion

5.1 Achievements

Our research investigated the factors influencing property standard compliance in Toronto using the RentSafeTO dataset. We employed a Bayesian linear regression model to examine how building characteristics—including construction year, number of units, and geographic location—impact evaluation scores.

The study focused on downtown Toronto, carefully analyzing a subset of the dataset to identify the most significant predictors of building performance. Our approach incorporated statistical methods that account for neighborhood-specific variations, allowing for a nuanced understanding of property standards.

Key findings emerged from our analysis:

- **Building Age and Size Matter:** Newer buildings and larger properties consistently received higher evaluation scores. This suggests that construction quality and management resources play a substantial role in maintaining property standards.
- **Geographic Variations:** Evaluation scores showed meaningful differences across different wards, highlighting the importance of local context in property maintenance.

- **Modeling Approach:** By using Bayesian linear regression with random effects and weakly informative priors, we developed a robust statistical framework for understanding complex housing data.

Visualization techniques, including regression plots, boxplots, and credibility interval representations, provided clear insights into the relationships between building characteristics and evaluation scores.

Our work demonstrates the power of sophisticated statistical modeling in understanding urban housing challenges. By breaking down the complex factors that influence building evaluations, we provide actionable information for property managers, policymakers, and urban planners working to improve housing standards in Toronto.

5.2 About Building Maintenance and Standards

From this analysis, we gain valuable insights into the factors that influence building maintenance and compliance with property standards. Specifically, we learn that newer buildings tend to perform better in evaluations, suggesting that modern construction practices and adherence to updated regulations play a significant role in meeting maintenance standards. This highlights the importance of updating building codes and encouraging renovations to improve compliance in older properties.

Additionally, the positive relationship between the number of units in a building and its evaluation score underscores the role of scale in property management. Larger buildings, often equipped with more structured management systems and dedicated resources, appear better positioned to meet maintenance standards compared to smaller properties. This finding suggests that economies of scale in management can lead to better overall building conditions.

Lastly, the variability in scores across different wards reveals how geographic and neighborhood factors influence compliance. Local policies, enforcement rigor, and neighborhood socioeconomic conditions likely contribute to these differences, underscoring the need for context-sensitive strategies to improve housing standards across the city.

5.3 Limitation

There are several limitations to consider. First, the RentSafeTO dataset only includes buildings registered under the program, excluding smaller buildings and those not covered by the bylaw. As a result, the findings may not generalize to all rental housing in Toronto, particularly for properties outside the program's scope.

Additionally, the dataset primarily focuses on physical and structural conditions, leaving out important contextual factors such as tenant satisfaction, management practices, or broader social influences on building quality. These unobserved variables may introduce bias or limit

the comprehensiveness of the analysis. Moreover, evaluation scores rely on inspections conducted by different bylaw enforcement officers, which may lead to variability due to subjective judgment or inconsistencies in enforcement practices.

The Bayesian model employed in this analysis assumes linear relationships between predictors, such as YEAR.BUILT and CONFIRMED.UNITS, and evaluation scores. While this simplification is useful for interpretability, it may overlook more complex interactions, such as the combined effects of building age and maintenance practices. Furthermore, missing values in key variables, such as evaluation scores and building characteristics, pose challenges for analysis. For instance, blank or zero scores may reflect non-evaluated cases rather than true building conditions, potentially skewing the results.

Finally, the results are context-specific and tailored to the RentSafeTO program in Toronto. They may not apply to other regions with different building standards, enforcement practices, or housing markets. These limitations emphasize the need for cautious interpretation of the findings and highlight areas for future research, such as integrating qualitative data on tenant experiences or examining the impact of inspector variability on evaluation outcomes.

5.4 Next steps

First, expanding the dataset to include a broader range of buildings, including smaller rental properties and those not registered with RentSafeTO, would provide a more comprehensive understanding of building conditions across Toronto. Incorporating data from other regions could also help generalize the findings and explore how different enforcement practices or housing markets affect evaluation outcomes.

Second, integrating additional variables, such as tenant satisfaction surveys, management practices, and neighborhood socioeconomic factors, would provide a more holistic view of what influences building quality. This could help identify factors beyond physical structure that contribute to compliance with property standards. Collecting longitudinal data on building conditions over time could also help explore the effects of renovations, policy changes, or aging on evaluation scores.

Third, standardizing inspection practices and incorporating measures to assess variability between inspectors could improve the reliability of evaluation scores. Training programs and inter-rater reliability checks could be implemented to ensure consistency in scoring across different inspectors and regions.

Finally, applying more flexible modeling approaches could capture complex interactions between predictors. For example, hierarchical or non-linear models might better reflect how building age, size, and geographic factors interact to influence evaluation scores. Exploring machine learning approaches could also help uncover hidden patterns in the data while providing robust predictions.

Appendix

A Data details

A.1 Raw data

Table 2 shows the preview of the raw RentSafeTo dataset.

Table 2: Preview of the raw RentSafeTo dataset

ID	RSN	Year Registered	Year Built	Year Evaluated	Property Type	Ward	Ward Name
1	4153014	2017	1930	2023	PRIVATE	04	Parkdale-High Park
2	4153013	2017	1973	2024	PRIVATE	04	Parkdale-High Park
3	4153012	2017	1958	2024	PRIVATE	04	Parkdale-High Park
4	4153026	2017	1917	2023	PRIVATE	04	Parkdale-High Park
5	4287030	2018	1917	2024	PRIVATE	04	Parkdale-High Park

A.2 Selected variables

We choose several variables for our report. The “RSN” (Record Serial Number) is the ID number for a building in RentSafeTO Open Data Sets. The “WARD” variable indicates the location where the building is located, represented numerically. Wards are geographical divisions within the city, and their names are specified in the “WARDNAME” variable. The “YEAR.BUILT” variable records the year the building was constructed. The “PROPERTY.TYPE” variable describes the type of property ownership (private, social housing, and TCHC). “CONFIRMED.STOREYS” denotes the number of floors in the building, while “CONFIRMED.UNITS” represents the total number of housing units in the building.

Finally, the “CURRENT.BUILDING.EVAL.SCORE” is an assessment of the building’s compliance with property standards. This score is derived by combining reactive scores (responses to complaints or issues) and proactive scores (regular inspections or maintenance reports). It serves as a comprehensive measure of the building’s condition and adherence to local property regulations.

Particularly, ‘WARD’ is the number from the 25-ward system (City of Toronto (n.d.)) which is used to mark different regions in Toronto. In this report, we choose 9, 10, 11, 12, 13, and 14, which are ‘Davenport’, ‘Spadina-Fort York’, ‘University-Rosedale’, ‘Toronto-St. Paul’s’, ‘Toronto Centre’, and ‘Toronto-Danforth’.

B Model details

B.1 Posterior predictive check

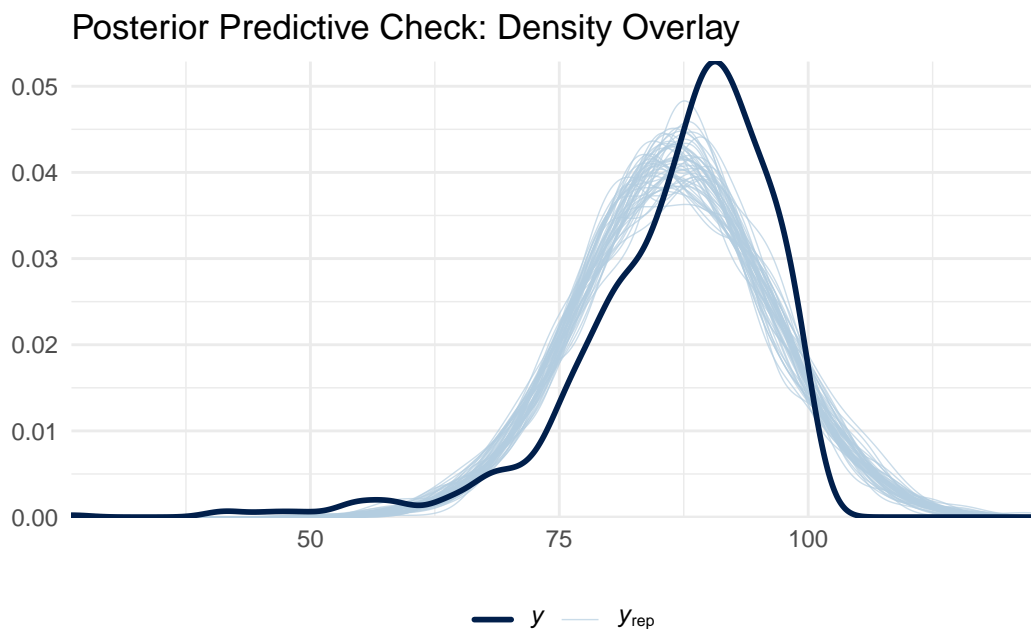


Figure 5: Posterior predictive distribution for Bayesian linear model

In Figure 5, we compare the model's predicted values to the observed data. This involves overlaying the posterior predictive distribution on the observed data distribution. The purpose of this check is to evaluate whether the model can reproduce patterns and variability seen in the actual evaluation scores. The results indicate a good fit, as the predicted and observed distributions align closely, suggesting the model captures the underlying structure of the data effectively. This validation supports the model's reliability in representing the relationship between building characteristics and evaluation scores.

Figure 6 compares the posterior with the prior distributions for key parameters in the Bayesian model, including the intercept, 'YEAR.BUILT', and 'CONFIRMED.UNITS'. The plot uses overlapping density curves to illustrate how the data influenced the parameter estimates. The posterior distributions shift away from the priors, indicating that the data provided substantial evidence to update the parameter values. For instance, the posterior for 'YEAR.BUILT' narrows and centers around values informed by the observed data, reflecting its significant role in predicting evaluation scores.

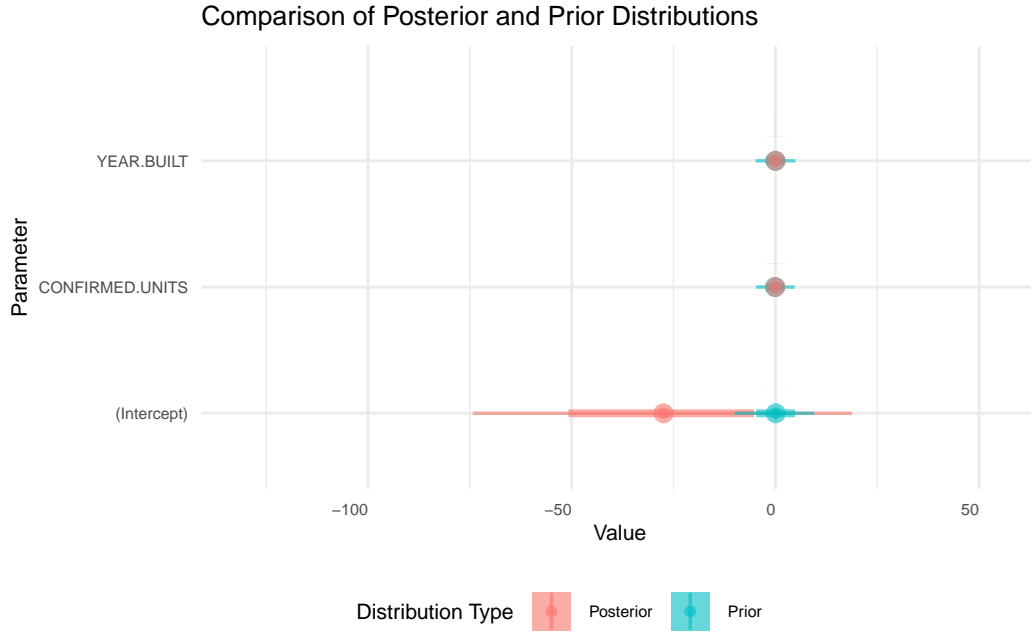


Figure 6: Comparison of posterior and prior distributions for Bayesian linear model

B.2 Markov chain Monte Carlo Convergence Check

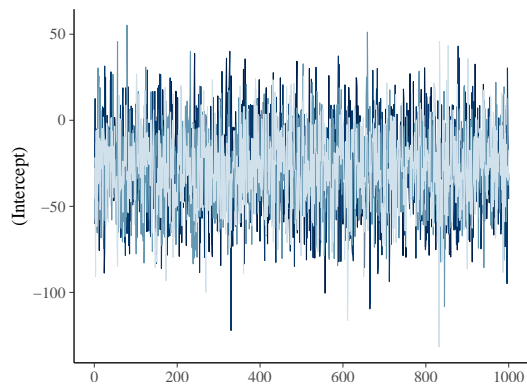
Figure 7 is a trace plot. It shows the sampling paths for key fixed effects, including the intercept, ‘YEAR.BUILT’, and ‘CONFIRMED.UNITS’, across all Markov chains. The plots demonstrate good mixing, with chains overlapping and exploring the posterior distribution consistently without any visible drifts or irregular patterns. This suggests that the sampling process converged for these parameters, providing reliable estimates for the posterior distributions.

Figure 8 is a Rhat plot. It shows the Gelman-Rubin diagnostic \hat{R} for all model parameters, which assesses convergence across Markov chains. All \hat{R} values are close to 1 and below the commonly accepted threshold of 1.05, indicating that the chains have converged, and the samples are representative of the posterior distribution. This suggests that the model has achieved satisfactory convergence, and the parameter estimates can be trusted for inference.

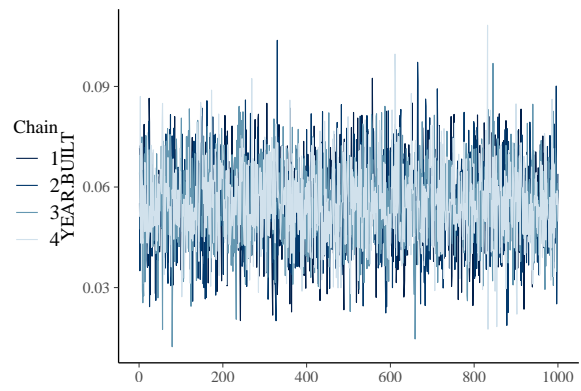
C Survey

C.1 Overview

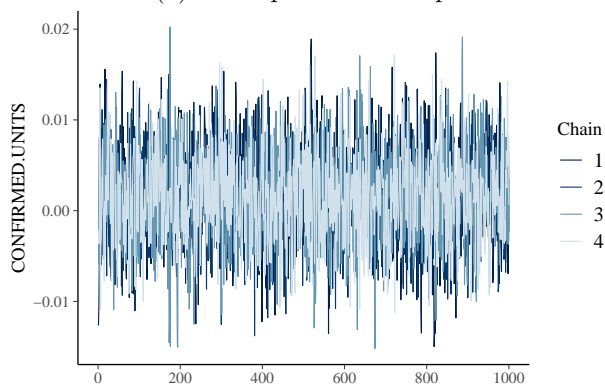
The goal of this survey is to gather residents’ opinions about the condition and maintenance of their apartment buildings, complementing the RentSafeTO evaluation system. By incorporat-



(a) Trace plot of Intercept



(b) Trace plot of YEAR.BUILT



(c) Trace plot of CONFIRMED.UNITS

Figure 7: Trace plot of intercept and predictors

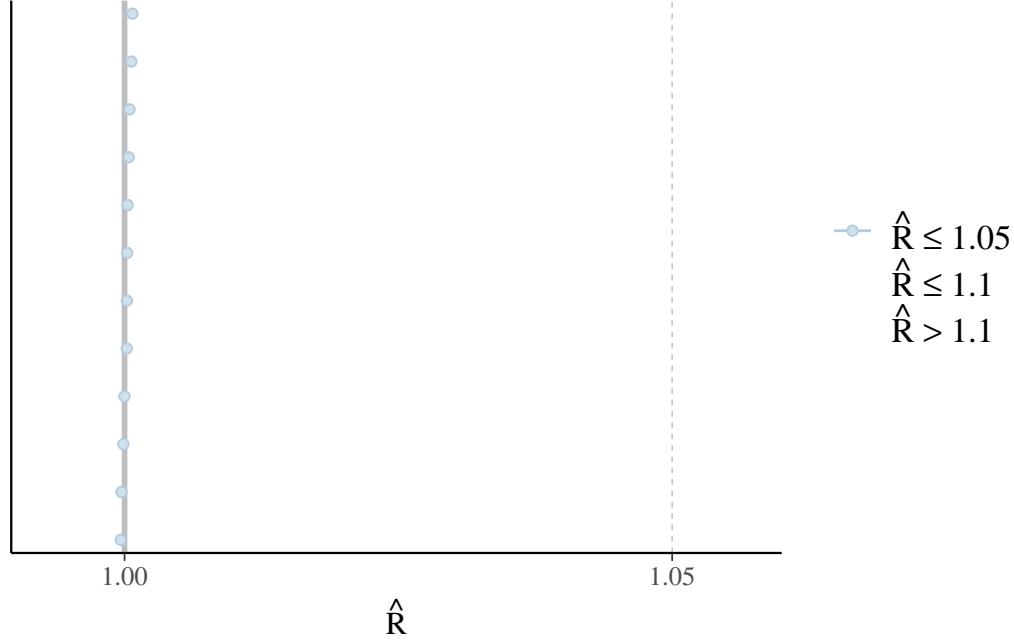


Figure 8: Rhat plot

ing tenant perspectives, the survey aims to provide a holistic understanding of building quality and identify areas for improvement. This feedback will help property managers, policymakers, and enforcement agencies address tenant concerns, prioritize maintenance efforts, and enhance housing standards across Toronto.

The survey focuses on key aspects of apartment living, including cleanliness, safety, maintenance responsiveness, and overall satisfaction. The collected data will be analyzed alongside the RentSafeTO evaluation scores to explore correlations between tenant perceptions and objective building assessments.

C.2 Sampling Approach

Target Population: The target population includes residents of multi-unit residential buildings registered under the RentSafeTO program in Toronto. This ensures alignment with the buildings already evaluated by the program.

Sample Size: To ensure representativeness, the survey will aim to collect responses from approximately 5,000 residents, distributed proportionally across all wards in Toronto. The sample size is calculated to provide reliable insights with a margin of error of $\pm 3\%$ at a 95% confidence level.

Sampling Technique: A stratified random sampling approach will be used: **Strata:** The sampling will be stratified by ward to ensure geographic diversity. **Random Sampling:** Within

each ward, a random sample of buildings will be selected, and a set number of residents per building will be invited to participate.

Geographic Coverage: The survey will cover all 25 wards in Toronto, ensuring equitable representation of diverse neighborhoods and building types. Special attention will be given to include a mix of private, social housing, and Toronto Community Housing Corporation (TCHC) properties.

C.3 Recruitment of Respondents

Outreach Channels: Email Invitations: Contact information will be obtained from property managers, with consent, to send email invitations to residents. Posters and Flyers: Informational posters and flyers will be placed in common areas of selected buildings. Community Partnerships: Collaboration with local tenant associations and housing organizations to encourage participation.

Incentives: To boost response rates, participants will be offered a chance to win one of 50 \$25 gift cards through a random draw. The incentive will be clearly communicated during recruitment.

Ensuring Anonymity: To encourage honest responses, the survey will be anonymous, with no identifying information collected. This will be emphasized in all recruitment materials.

C.4 Survey Design

Survey Format The survey will be conducted online using Google Forms, with an option for paper forms for residents without internet access. It will take approximately 10 minutes to complete.

Sections and Question Types Demographics:

Age group (checkboxes) Tenure in the building (dropdown menu: less than 1 year, 1–5 years, more than 5 years) Housing type (radio buttons: private, social housing, TCHC) Building Condition:

“How would you rate the cleanliness of your building?” (Likert scale: 1–5) “Are common areas, such as hallways and lobbies, well-maintained?” (Yes/No) Safety and Security:

“Do you feel safe in your building?” (Likert scale: 1–5) “Are security measures, such as cameras or locks, adequate?” (Yes/No) Maintenance and Responsiveness:

“How satisfied are you with the responsiveness of building management to maintenance requests?” (Likert scale: 1–5) “How long does it typically take for maintenance requests to be resolved?” (multiple choice: less than a week, 1–2 weeks, more than 2 weeks) Overall Satisfaction:

“How satisfied are you overall with living in your building?” (Likert scale: 1–5) “What is the most pressing issue you face as a tenant?” (open-ended) Suggestions for Improvement:

“What changes would improve your living experience?” (open-ended) Pilot Testing The survey will be pilot-tested with 50 respondents to ensure clarity and effectiveness. Feedback will be incorporated to refine the questions and structure.

C.4.1 Link

https://docs.google.com/forms/d/e/1FAIpQLSdN_4gdZUWW0m9YNXgFPfU_DjmRexdUMZNHmX1INj0s2

References

- Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in R*. Chapman; Hall/CRC.
- Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Bueros Novik, and R Wolfe. 2018. “Joint Longitudinal and Time-to-Event Models via Stan.” https://github.com/stan-dev/stanco_n_talks/.
- City of Toronto. n.d. “RentSafeTO Evaluation Tool.” <https://www.toronto.ca/community-people/housing-shelter/rental-housing-tenant-information/rental-housing-standards/apartment-building-standards/rentsafeto-evaluation-tool/>.
- . n.d. “Ward Profiles.” <https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/ward-profiles/>.
- Gelfand, Sharla. 2020. *opendatatoronto: Access the City of Toronto Open Data Portal*. <https://cran.r-project.org/package=opendatatoronto>.
- Kay, Matthew. 2023. *Ggdist: Visualizations of Distributions and Uncertainty*. <https://mjskay.github.io/ggdist/>.
- Lüdecke, Daniel. 2023. *sjPlot: Data Visualization for Statistics in Social Science*. <https://CRAN.R-project.org/package=sjPlot>.
- Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- 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 Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Roger D. Peng Victoria Stodden Friedrich Leisch. Chapman; Hall/CRC.