Exploring Apartment Building Maintenance and Management Trends*

An Analysis of 2019-2023 RentSafeTO Apartment Building Evaluation Data

itf author

March 27, 2024

This report explores trends in apartment building maintenance and management using data from RentSafeTO, a program aimed at ensuring compliance with building maintenance standards in Toronto. Through exploratory data analysis (EDA), it was found that privately owned and social housing buildings generally exhibit higher building scores compared to Toronto Community Housing Corporation (TCHC) buildings. Furthermore, a positive relationship between graffiti scores and overall building scores was observed for privately owned and TCHC buildings. Additionally, older buildings tend to receive lower scores, indicating a negative correlation between building age and scores. Subsequent linear regression analysis quantified the impact of building age on building scores, revealing a decrease of 0.11 units in score for each additional year of building age. These findings underscore the importance of proactive maintenance efforts in ensuring the quality of apartment buildings and provide valuable insights for policymakers and building managers in enhancing building maintenance standards.

Table of contents

*Code and data are available at: https://github.com/cthierst/marital_status_politics_gss_analysis.git

1 Introduction

Access to adequate housing is a fundamental aspect of human well-being, alongside essentials like food, clothing, and transportation. The quality and comfort of housing significantly impact individuals' overall quality of life. In this report,a public data set on OpenToronto about *Apartment Building Evaluation* will be evaluated to quantify the problem.

The Apartment Building Standards program, known as RentSafeTO, plays a crucial role in ensuring that apartment building owners and operators adhere to maintenance standards. This program applies to buildings with three or more storeys and 10 or more units, aiming to safeguard tenants' living conditions. More detailed information about this program named *RentSafeTO* can be learned by visiting its website.

The primary objective of this paper is to investigate the factors that affect the overall score of apartment buildings, as assessed by RentSafeTO. We examine various variables, including property type, graffiti severity, number of confirmed units, and building age. By analyzing the relationships between these factors and the building scores, we aim to uncover insights into the maintenance and management trends prevalent in Toronto's apartment buildings. Through data visualization techniques and linear regression modeling, we aim to provide a comprehensive understanding of the drivers behind building scores.

My research seeks to answer the question: what factors influence the evaluation of apartment buildings, and do they have a positive or negative impact? We examine various factors such as property type and building age, among others. In urban environments, the maintenance and management of apartment buildings play a crucial role in ensuring the well-being and safety of residents. The assessment of these buildings not only reflects their physical conditions but also sheds light on broader trends in housing management practices. In this report, we delve into exploring the trends in apartment building maintenance and management using data sourced from RentSafeTO, a program aimed at enforcing building maintenance standards in Toronto. Our analysis focuses on identifying factors that influence building scores, thereby contributing to a better understanding of the determinants of building quality and maintenance practices. By leveraging open questionnaire data, our analysis focuses on understanding residents' concerns regarding living comfort. These insights can inform governmental planning efforts, facilitating the provision of convenient services and facilities to communities. Ultimately, our research aims to contribute to enhancing residents' overall happiness and well-being.

The remainder of this paper is structured as follows. Section 2 will primarily introduce the data used in the research, including its sources, data cleaning procedures, basic statistics, and explanations of key variables. Section 3 will delve into exploratory data analysis using the cleaned dataset, aiming to derive valuable yet intuitive insights. Section 4 will employ linear models to investigate the quantitative relationships affecting the evaluation score. Finally, a discussion will be provided on the strengths and limitations of the report.

In summary, this paper contributes to the existing literature by providing insights into the factors influencing apartment building scores and their implications for housing management. Through our analysis, we aim to facilitate informed decision-making and enhance the quality of apartment living in urban environments.

2 Data

2.1 Data Management

This paper uses the R statistical programming language (R Core Team 2023), along with several packages, tidyverse (Wickham et al. 2019), janitor (Firke 2021), here (Müller 2020), and (Wickham et al. 2022). Figures in this paper were created using the packages ggplot2 (Wickham 2016) and the tables were created using knitr (Xie 2023) and kableExtra(Zhu 2021). Combinations of figures were created using the patchwork package (Pedersen 2022).

2.2 Data Source

This dataset contains building evaluation scores for buildings registered with RentSafeTO. Before 2023, buildings must undergo evaluation at least once every three years. During evaluations, Bylaw Enforcement Officers inspect common areas, mechanical and security systems,

Table 1: Distribution of Records in Each Year Evaluated

Year	Total # of Records	Proprotion in Analysis
2019	1561	25.02%
2020	1469	23.55%
2021	1461	23.42%
2023	1746	27.99%

parking and exterior grounds. Each item is inspected and assigned a score from one to five, with one being the lowest and five being the highest. In 2023, RentsafeTO updated its evaluation process. Buildings must undergo evaluation at least once every two years, increasing the number of evaluation categories to 50 and allocating a weight to the category. The dataset before 2023 and after 2023 are downloaded as two csv files from the website, and the file before 2023 contains 11760 records and 40 cloumns, as for the file after 2023, if contains 1758 records with 71 columns.

The dataset is downloaded from the Open Toronto website.

2.3 Data Cleaning

Since the two datasets contains different variable sets and may contain errors, missing values, or inconsistent information, these issues can affect the accuracy and reliability of the analysis. Before formal data analysis and exploration, we need to perform data cleaning, filter out suitable variables from the original data set and merge the two data sets into a single data set for analysis. At the same time, I hope to process a new variable house age from the original data set. This variable needs to subtract the year the house was built from the year the data was collected. This step must also be completed in the data preprocessing step.

After performing necessary cleaning and missing value processing on the data, we obtained the final sample data set. The data set contains a variety of variables for different years (representing when the sector was assessed), and the distribution is shown in Table 1.

$$AGE = YEAR\_EVALUATED - YEAR\_BUILT \tag{1}$$

2.4 Key Features

The evaluation score is the estimand of interest, which is the overall score of the building. The score is the sum total of each item that was evaluated. The formula to calculate scores is as follows: sum of all assigned scores during the evaluation / (number of unique items reviewed *5). The other independent variables are house age, property type, confirmed units and graffiti.

Table 2: Variable Descriptions

Variable	Variable Description	Measurement
SCORE	This is the overall score of the building.	Numeric
PROPERTY_TYPE	This field informs the users of a building.	Categorical
GRAFFITI	This score represents the severity of graffiti in a building.	Numeric
YEAR_BUILT	This is the year that the building was built in.	Year
YEAR_EVALUATED	This represents the year of the building evaluation scores.	Year
AGE	YEAR_EVALUATED minus YEAR_BUILT.	Numeric

And description of these variables are in Table 2, which is gained from the official website of this data source.

The variable *Property_Type* refers to the ownership type of the building. In RentSafeTO's dataset, this variable reflects whether the building is privately owned, owned by the Toronto Community Housing Corporation (TCHC), or owned by another assisted, social or supportive housing provider. This variable is crucial to our understanding of how different types of buildings are maintained and managed.TCHC is the housing agency of the City of Toronto and is responsible for managing most of the city's public housing. TCHCs provide low-rent housing for low-income families and individuals, so their maintenance and management may be affected by government regulations and funding constraints. Other Assisted, Social, or Supportive Housing Provider: This category includes assisted, social, or supportive housing provided by other organizations or agencies, such as non-profit organizations, social service agencies or supportive housing agency. These buildings may have special management and service models to meet the needs of specific groups.

We can find in Figure 1 that the distribution of various property types in each evaluation year from the figure below. We can find that Private accounts for the largest proportion, accounting for more than 78% of the samples in each year from Table 3.

Table 3: Percentage of Different Property Types in Each Year Evaluated

YEAR_EVALUATED	PRIVATE	SOCIAL HOUSING	TCHC
2019	83.92 %	7.5 %	8.58 %
2020	86.86 %	4.83 %	8.3 %
2021	78.78 %	7.8 %	13.42 %
2023	83.62 %	7.22 %	9.16 %

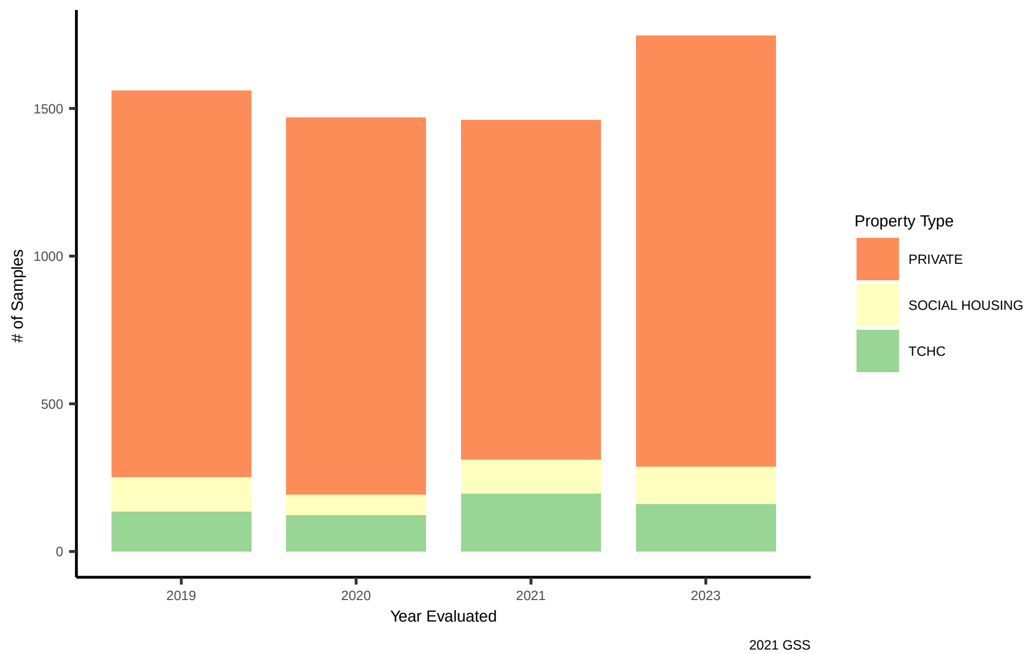


Figure 1: 2019-2023 Comparison of Sample's Property Type Against Year Evaluated

3 Exploratory Data Analysis and Results

Exploratory Data Analysis (EDA) plays a pivotal role in unraveling the underlying patterns, relationships, and insights hidden within a dataset. In this section, we embark on a journey of discovery, utilizing a plethora of statistical graphics and visualizations to delve into the intricacies of our data. Through a series of subsections, we aim to unravel the relationships between variables, uncover trends, and identify potential patterns that may inform subsequent analyses.

3.1 Privately owned and social housing buildings generally exhibit higher building scores compared to TCHC.

In this section of analysis, we utilized boxplots to explore the relationship between building scores and property types across four evaluation years (2019, 2020, 2021, and 2023). A boxplot is a commonly used statistical graph that displays the distribution of data, showing the median, quartiles, and potential outliers. Through the boxplots, we visually compared the distribution of scores across different property types, allowing us to discern any notable differences.

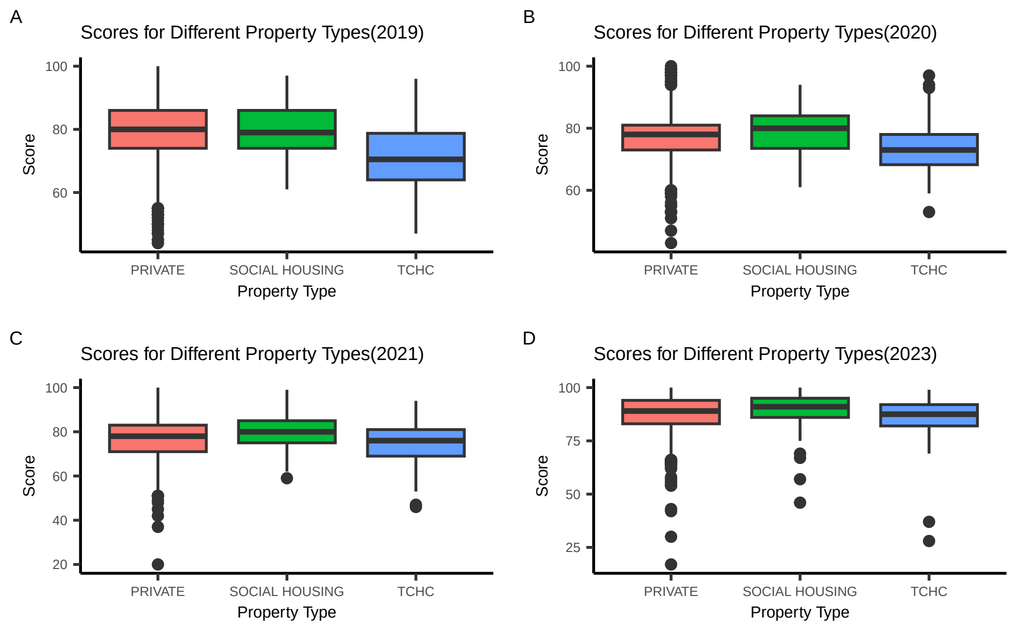


Figure 2: 2019-2022 Comparison of Building's Score Against Property Type

The boxplot Figure 2 analysis yielded a straightforward conclusion: buildings categorized as privately owned and social housing tended to have higher overall scores compared to those owned by the Toronto Community Housing Corporation (TCHC). This observation aligns with our initial expectations but also prompts further investigation and consideration.

The analysis of average building scores across different property types over the four-year period in Table 4 reveals a notable increase in scores in 2023 compared to previous years (2019-2021). This discrepancy can be attributed to the modification of evaluation rules by RentSafeTO in 2023, which incorporated additional dimensions and attributes for building assessment and introduced a new scoring methodology. As a result, the overall distribution of scores in 2023 differs significantly from previous years. Therefore, by controlling for the year evaluated in our analysis, we mitigate the potential measurement error arising from variations in scoring rules over time.

The underlying reasons for this discrepancy could be multifaceted. Firstly, TCHC provides low-

Table 4: Mean Score for Different Property Types in each Year Evaluated

YEAR_EVALUATED	PRIVATE	SOCIAL HOUSING	TCHC
2019	79.35	79.44	70.60
2020	77.34	78.70	73.83
2021	76.68	79.44	75.27
2023	87.51	89.44	86.31

rent housing for low-income families and individuals, resulting in relatively lower rental costs. This may lead to limited investments in maintenance and management for TCHC buildings, consequently impacting their overall scores. Additionally, privately owned and social housing buildings may benefit from more proactive maintenance and management efforts by private owners or social organizations to maintain property values and living environments(Ferrell 2018).

However, it's important to note that this conclusion is based on observed data patterns and does not imply that privately owned and social housing buildings are inherently of higher quality than those owned by TCHC. Further analysis is required to explore other potential influencing factors such as building age, geographical location, etc., and to provide a more nuanced interpretation of the data. Such analysis will contribute to a more comprehensive understanding of the causal relationships underlying building scores, informing future decision-making and policy formulation.

3.2 For privately owned and TCHC buildings, there is a positive relationship between graffiti scores and overall building scores.

In the subsequent analysis using pivot tables, I examined the relationship between building scores and graffiti scores across different property types and evaluation years. From Table 5, it is evident that for privately owned houses and those owned by the Toronto Community Housing Corporation (TCHC), higher graffiti scores corresponded to higher overall scores during evaluations in 2019-2021. Conversely, for social housing, this relationship was only prominent in the evaluation of 2020, while being less apparent in other years.

These findings suggest that the impact of graffiti scores on building scores varies across different property types and evaluation years. For privately owned and TCHC buildings, the positive correlation between graffiti scores and overall scores could be indicative of proactive maintenance efforts to address graffiti and maintain overall building quality. However, the less consistent relationship observed for social housing may reflect different management approaches or varying levels of emphasis on graffiti removal and building maintenance across different evaluation periods.

Table 5: Mean Score for Different Property Types against Graffiti Score in each Year Evaluated

PROPERTY_TYPE	YEAR_EVALUATED	2	3	4	5	1	0
PRIVATE	2019	60.00	66.85	73.66	80.76	NA	NA
PRIVATE	2020	67.90	70.67	74.18	78.38	NA	NA
PRIVATE	2021	66.33	65.54	70.80	78.17	35.50	NA
PRIVATE	2023	83.50	88.32	NA	NA	78.21	43.25
SOCIAL HOUSING	2019	82.00	72.57	69.08	81.32	NA	NA
SOCIAL HOUSING	2020	67.80	77.50	75.08	80.90	NA	NA
SOCIAL HOUSING	2021	65.00	79.22	69.56	80.55	NA	NA
SOCIAL HOUSING	2023	85.88	90.84	NA	NA	83.29	NA
TCHC	2019	61.67	62.41	70.22	76.80	56.57	NA
TCHC	2020	64.50	70.26	72.77	75.81	NA	NA
TCHC	2021	64.00	68.43	72.61	78.98	57.00	NA
TCHC	2023	86.92	87.35	NA	NA	84.95	32.50

Possible reasons behind these variations could include differences in property management strategies, resource allocation for maintenance and cleaning, or community engagement initiatives aimed at addressing graffiti issues. Additionally, external factors such as changes in neighborhood dynamics or vandalism trends may also influence the observed relationships between graffiti scores and building scores. Overall, these findings underscore the complex interplay between building maintenance practices, property management approaches, and external factors in shaping building evaluation outcomes across different property types and evaluation years.

3.3 Older buildings tend to receive lower scores.

Through scatter plot analysis, it is evident from Figure Figure 3 that for privately owned houses, there exists a significant negative correlation between building scores and age, indicating that older buildings tend to receive lower scores. This observation aligns with our intuition that newer properties are often preferred for residence. However, anomalies are observed for buildings with ages exceeding 150 years, as their scores deviate from this trend and appear higher. This anomaly may be attributed to the historical significance of these older buildings, coupled with diligent maintenance efforts by owners, resulting in elevated scores.

Conversely, for social housing, no clear trend between building scores and age is discernible. However, for buildings owned by the Toronto Community Housing Corporation (TCHC), a noticeable negative correlation exists, and it is apparent from Figure 3 that TCHC buildings generally have newer ages.

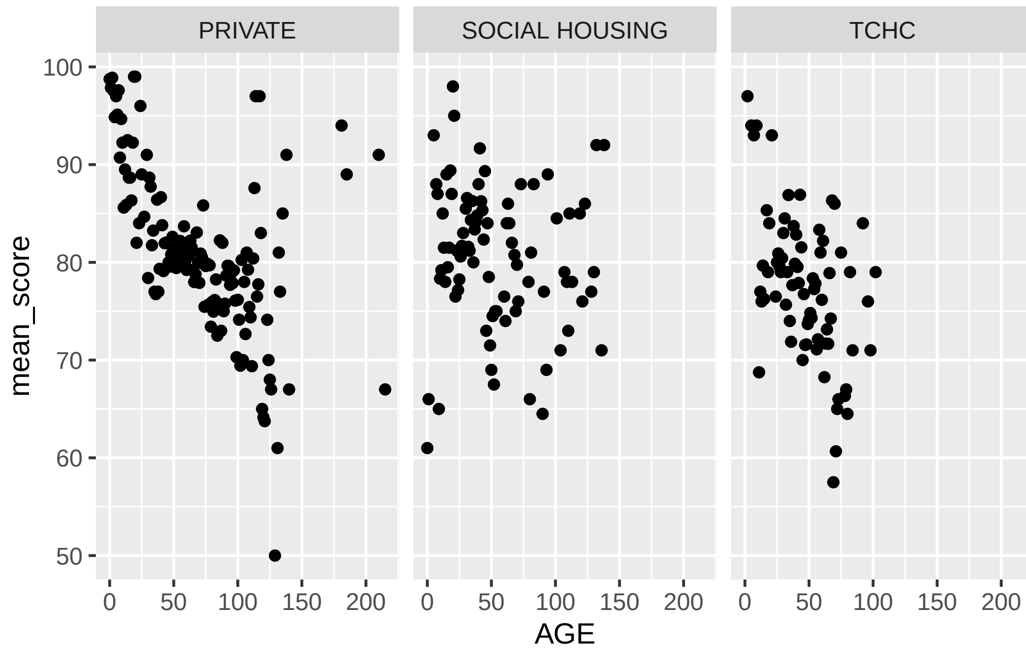


Figure 3: Mean Score Grouped by Age for Different Property Types

These findings underscore the complex relationship between building age and scores, with various factors potentially influencing the observed patterns. For privately owned houses, the negative correlation may be attributed to preferences for newer properties and the impact of diligent maintenance practices. However, anomalies observed for exceptionally old buildings suggest the influence of historical significance and exceptional maintenance efforts. Conversely, the absence of a clear trend for social housing buildings may be indicative of varied management approaches and maintenance standards. Overall, these insights highlight the nuanced interplay between building age, maintenance practices, and historical significance in shaping building evaluation outcomes.

4 Model

In light of the negative correlation observed between building age and scores during the Exploratory Data Analysis (EDA) phase, we further delve into the relationship through linear regression analysis. In this Model section, our aim is to construct a linear regression model to quantify the impact of building age on building scores. By delineating the model's formulation, justifications, and rationale, we seek to provide a comprehensive understanding of the underlying relationship between these variables.

Table 6: Mean Score for Different Property Types against Graffiti Score in each Year Evaluated

Table 7: Model Summary

term	estimate	std.error	statistic	p.value
(Intercept)	87.68	1.01	87.02	0
AGE	-0.11	0.01	-8.52	0

4.1 Model set-up

Define the model formula as:

$$SCORE = \beta_0 + \beta_1 AGE + \epsilon. \tag{2}$$

4.2 Model Result

We run the model in R (R Core Team 2023) using the `lm()` command. And the model result is show in Figure 4 and Table 6.

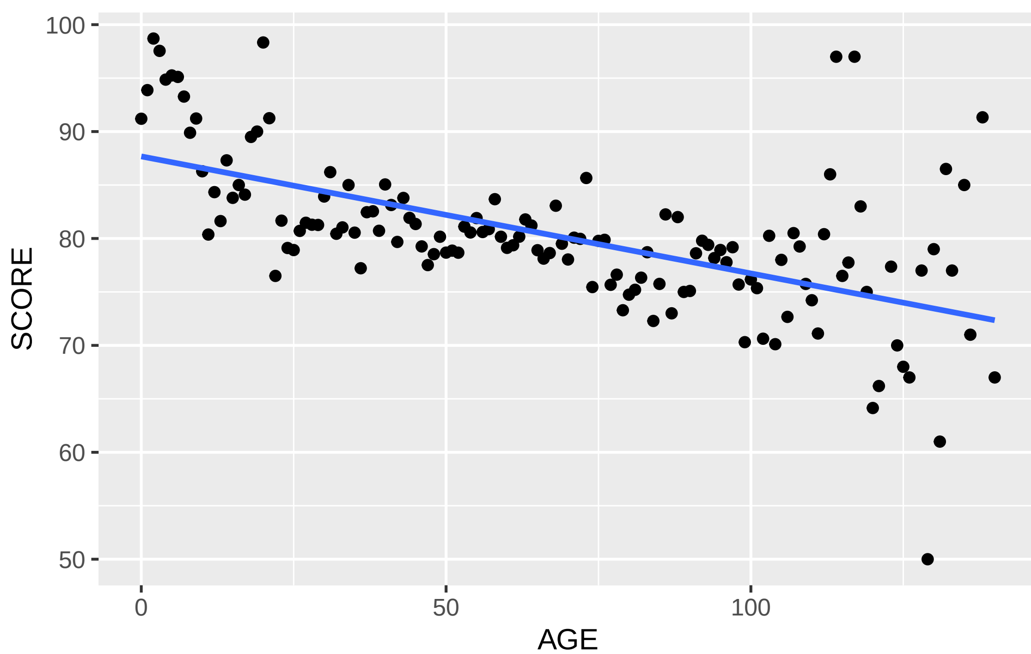


Figure 4: Mean Score Grouped by Age for Different Property Types

The conclusion drawn from the linear regression model is that there exists a statistically significant linear relationship between the independent variable (age) and the dependent variable (score). The fitted regression line accurately captures the trend observed in the scatter plot, suggesting that age is a significant predictor of building scores.

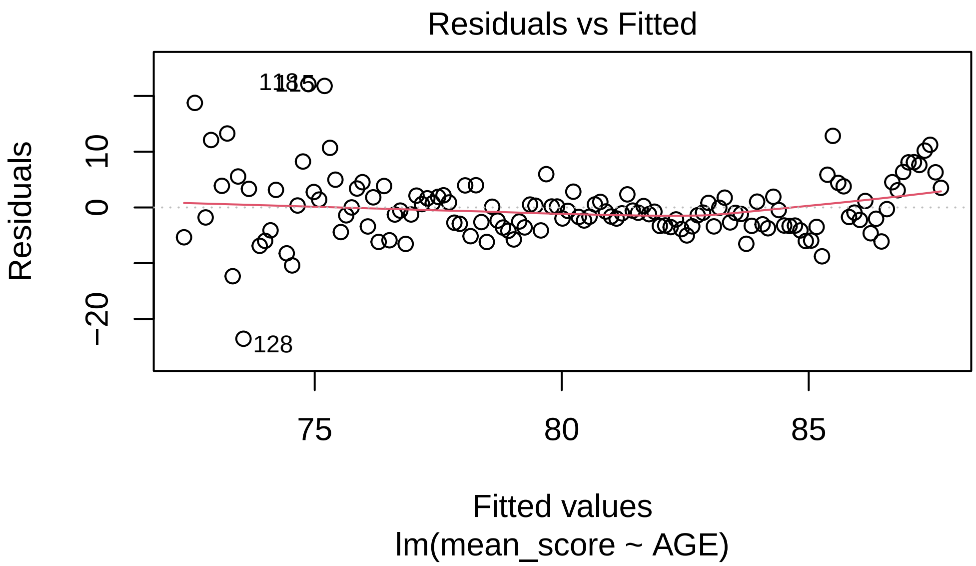
The formula

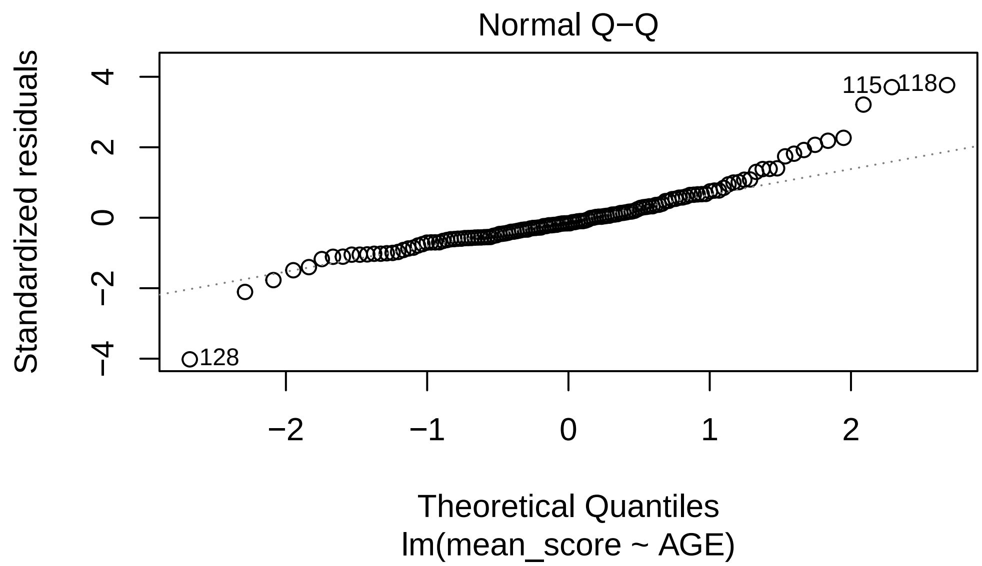
$$SCORE = 87.68 - 0.11 \times AGE$$

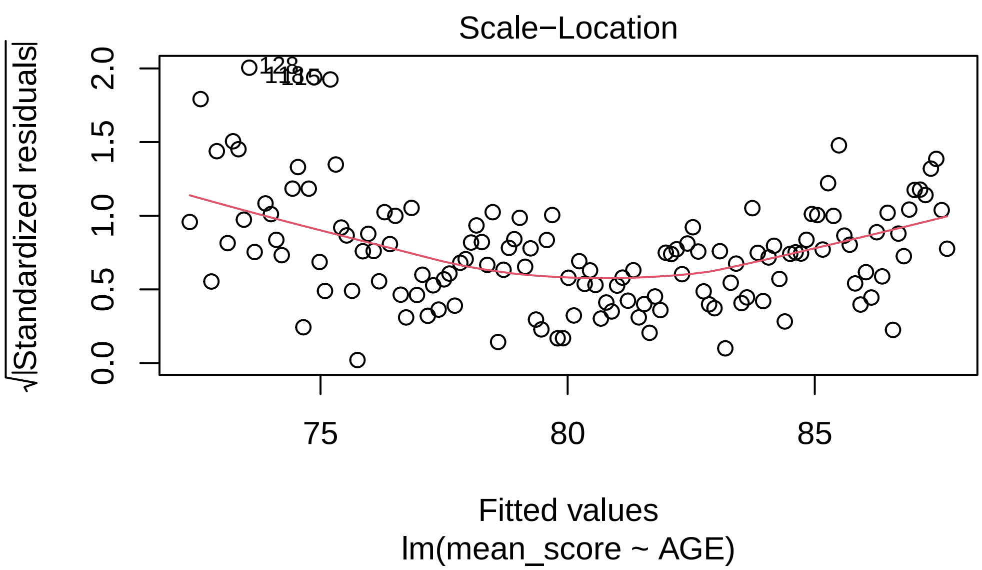
epresents the linear regression equation obtained from the model. 87.68 represents the intercept of the regression line, indicating the estimated score when the age of the building is zero. -0.11 is the coefficient of the age variable, representing the rate of change in the predicted score for each unit increase in the age of the building.

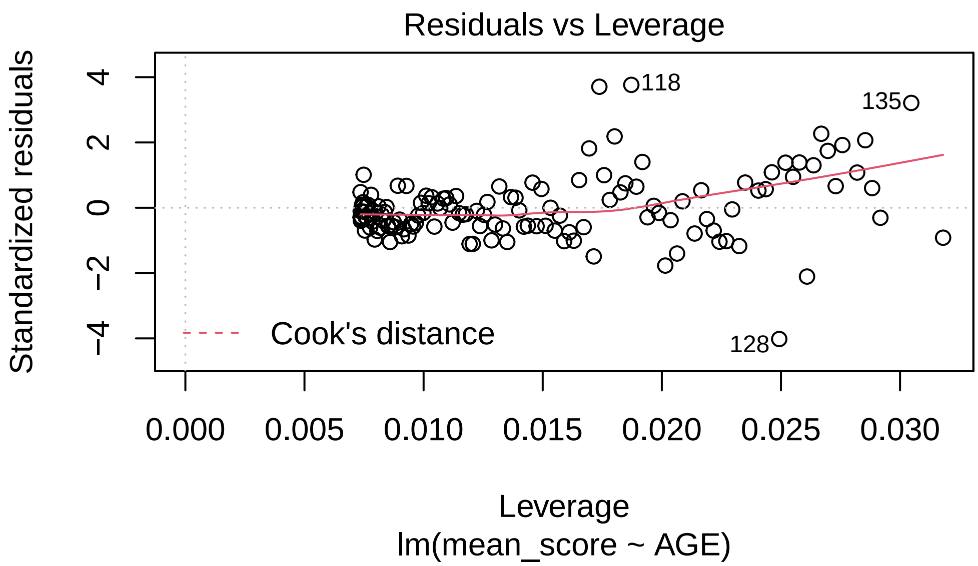
4.3 Model justification

During the validation and justification process of the linear regression model, we used the `plot()` function to examine the model's fit. By observing the residual plots, we found that the residuals exhibit a random distribution pattern as the predicted values increase, without any apparent systematic patterns. This suggests that the model may adhere to the assumptions of linear regression. Additionally, no clear heteroscedasticity was observed in the scatter plot of residuals against predicted values. Taken together, these results indicate that our linear regression model has passed validation and can reasonably account for the variability in the target variable.




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mean_score ~ AGE)
115
118
128


Scale−Location
√|Standardized residuals|
Fitted values
lm(mean_score ~ AGE)
128



5 Discussion

5.1 Main Findings

We summarize the key findings obtained through exploratory data analysis (EDA) and linear regression analysis:

1. Exploratory Data Analysis (EDA) Findings:

- Privately owned and social housing buildings generally exhibit higher building scores compared to TCHC.
- For privately owned and TCHC buildings, there is a positive relationship between graffiti scores and overall building scores.
- Older buildings tend to receive lower scores.

2. Linear Regression Analysis:

- Through linear regression analysis, we quantified the impact of building age (Age) on building scores. The regression equation obtained is:

$$SCORE = 87.68 - 0.11 \times AGE.$$

- This equation indicates that for each additional year of building age, the building score is expected to decrease by 0.11 units.

5.2 Limitations and Weaknesses

1. Regional Limitation: The analysis is confined to apartment buildings in Toronto due to the dataset's source being RentSafeTO. Consequently, the findings may not be generalizable to apartment buildings in other cities or regions.
2. Limited Variables: The dataset may lack certain variables that could provide deeper insights into building maintenance and management. For instance, factors such as tenant demographics, building materials, or maintenance budgets could influence building scores but are not included in the analysis.
3. Model Assumptions: While linear regression was employed to quantify the relationship between building age and scores, it assumes a linear relationship between the variables and may not capture potential non-linear associations. Alternative modeling approaches, such as polynomial regression or non-linear models, could be explored to better capture the complexity of the relationship.
4. External Factors: The analysis may not account for external factors that could influence building scores, such as economic conditions, neighborhood characteristics, or building management practices. Failure to consider these factors could limit the comprehensiveness of the analysis and the interpretation of results.

Addressing these limitations and weaknesses could enhance the robustness and applicability of the findings, providing a more comprehensive understanding of apartment building maintenance and management trends.

5.3 Future Research

1. Multi-City Comparison: Future research could extend the analysis beyond Toronto to include apartment buildings in other cities or regions. Comparing building maintenance and management practices across different geographical areas could provide valuable insights into regional variations and best practices.
2. Inclusion of Additional Variables: Expanding the dataset to include a wider range of variables, such as tenant demographics, building materials, maintenance budgets, and neighborhood characteristics, could enrich the analysis. Exploring the impact of these factors on building scores could offer a more comprehensive understanding of building maintenance dynamics.
3. Exploration of Non-linear Relationships: Investigating non-linear relationships between building age, graffiti scores, and building scores could provide a more nuanced understanding of their interplay. Future research could explore the use of advanced modeling techniques, such as nonlinear regression or machine learning algorithms, to capture complex relationships more accurately.

4. Policy Evaluation: Assessing the impact of RentSafeTO and other building maintenance standards programs on building scores and overall building quality could be an important area for future research. Evaluating the effectiveness of policy interventions and identifying areas for improvement could inform policy development and implementation strategies(Foster et al. 2022).

Appendix

A Linear Regression Model

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The relationship is assumed to be linear, meaning that changes in the independent variables are associated with linear changes in the dependent variable. The general form of a linear regression model with one independent variable can be expressed as(Su, Yan, and Tsai 2012):

$$y = \beta_0 + \beta_1 x + \epsilon.$$

Where - y is the dependent variable (the variable we want to predict). - x is the independent variable (the variable that influences y). - β_0 is the intercept, which represents the value of y when $x = 0$. - β_1 is the slope, which represents the change in y fro one-unit change in x. - ϵ is the error term.

The goal of linear regression is to estimate the coefficients that best fit the data, typically by minimizing the sum of the squared differences between the observed and predicted values of y. Once the coefficients are estimated, the model can be used to make predictions about the dependent variable based on new values of the independent variables.

References

Ferrell, David L. 2018. "Livable for All Ages: Evaluating Perceptions of Community in an Intergenerational Context."

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Foster, Sarah, Paula Hooper, Anthony Duckworth, and Julian Bolleter. 2022. "An Evaluation of the Policy and Practice of Designing and Implementing Healthy Apartment Design Standards in Three Australian Cities." *Building and Environment* 207: 108493.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Pedersen, Thomas Lin. 2022. *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/.

Su, Xiaogang, Xin Yan, and Chih-Ling Tsai. 2012. "Linear Regression." *Wiley Interdisciplinary Reviews: Computational Statistics* 4 (3): 275–94.

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, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

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

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.