Analysis of Housing Data in Toronto

Cainchen Sun

2024-09-26

Abstract

The dataset surveyed provides the most detailed information on the condition of all residential properties in Toronto, including property type, evaluation scores, and physical condition. Housing quality in a rapidly growing, changing city becomes a priority regarding community well-being, social equity, and planning. Based on this, the paper digs deeper into some insight from this dataset and realizes ethical implications of housing assessment and statistical methodology consideration in the evaluation of data.

1. Introduction

Apartment Building Standards is a bylaw enforcement program that formed in 2017 in order to provide a guarantee that on building maintenance standards, the owners and operators of apartment buildings with three stories or more, or with ten units or more. For buildings that are registered with RentSafeTO, building evaluation scores are included in this collection. By-law Enforcement Officers conduct inspections of parking lots, the exterior grounds, mechanical and security systems and common areas. The RentsafeTO program revised its rating methodology in 2023. Licensing requires a minimum one rating every two years for a building, with fifty rating categories in all, with assigned weighting for each category. Each item in the category, after observation, is allocated a score ranging from one to three, with one being the lowest and three being the greatest. At times, an object may be accorded a score of 0 if it's impossible to assess because of some form of obstacle or due to a reluctance to inspect. In situations where the item is not applicable for the building at the time of assessment, the score would appear blank in the dataset.

2. Data Source and Context

2.1 Origin of the Data

This is probably from municipal building records or some sort of housing authority, such as the Toronto Community Housing Corporation. These possibly could be gathered in a regular assessment to maintain qualities in the housing and ensure safety for the tenants, along with following all the regulations of the municipality. Some of the various attributes comprising this, such as the year of construction and property type, rating scores, and amenities, suggest that the assessment must be rather exhaustive of properties.

2.2 Broader Context

2.2.1 Ethical Considerations

1. Privacy and Confidentiality: Individual properties and their owners should be anonymized to protect privacy. Public datasets should be treated with caution not to reveal too much sensitive information.
2. Fairness in Housing: The dataset could reveal disparities in housing quality within neighborhoods. These kinds of analyses may uncover deep-seated issues in which some communities could be falling behind, and that would be an ethical concern with resource allocation or urban development.
3. Community Engagements: Community input is essential in an assessment process. It helps capture valuable insights from them about the struggles they face and therefore develops a more bottom-up inclusive approach toward assessments of housing conditions.

2.2.2 Statistical Considerations

1. Bias and Validity: The first thing that needs to be done is to check whether the methods of data collection themselves introduce bias. For example, if there are two separate sets of neighborhoods favored or shunned, it would obviously impact the outcome. The statistical methodology should ensure that all areas receive unbiased evaluations.
2. Data Quality: The veracity of the data, in particular evaluation scores, affects any conclusion derived from the data; as such, the collection and validation of the data would give an insight into the reliability of the dataset.
3. Generalisability: Results obtained from this dataset are not generalization. Any generalization to other regions or contexts based on the findings should be made with great caution because of the peculiarities surrounding the Toronto housing market.

3. Descriptive Analysis

3.1 Summary of the Data

```
-- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

Descriptive statistics provide a foundation for understanding the dataset's key characteristics.

<div align="center">

Summary of Housing Data

total_properties	average_eval_score	average_units	average_storeys
3271	87.11892	87.84164	7.596759

</div>

This table summarizes key statistics that provide insight into the overall characteristics of the housing properties in the dataset. It is crucial to understand these metrics before delving deeper into specific analyses.

4. Visualization

Visualizations enhance our understanding of the data by providing clear representations of relationships and distributions.

4.1 Score Distribution for the Evaluation

One of the first visualizations we can create is a histogram to explore the distribution of current building evaluation scores. This can help identify trends, such as whether most properties receive high or low evaluation scores.
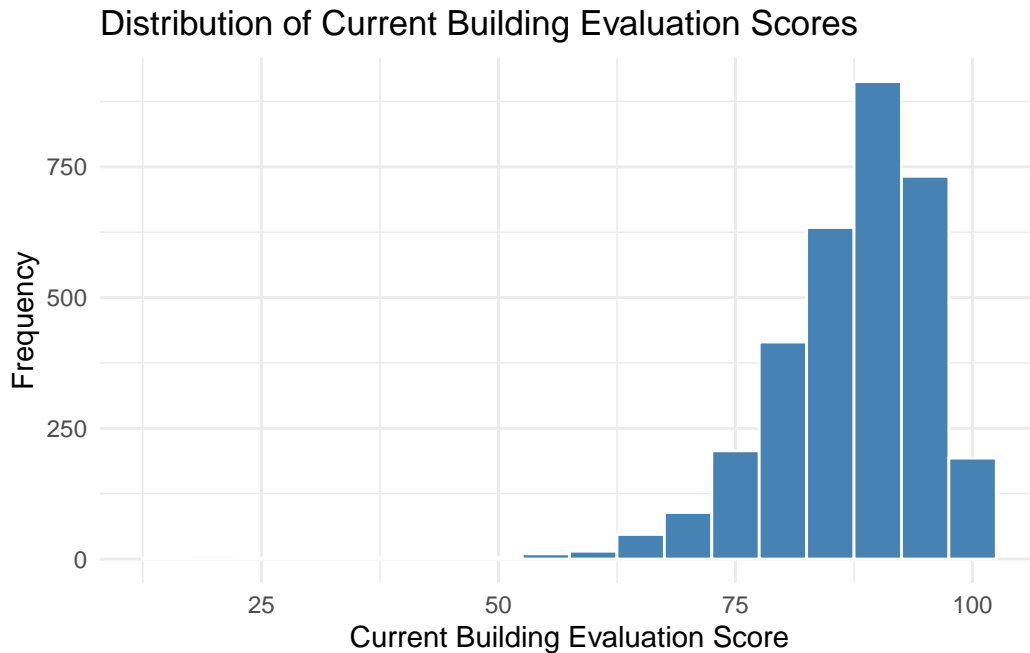
Distribution of Current Building Evaluation Scores

Figure 1: Present Building Assessment Score Distribution

The histogram in Figure 1 displays the frequency of various evaluation scores across the properties. A majority of properties appear to cluster around higher scores, indicating a generally positive assessment of building conditions.

4.2 Total Properties by Type of Property

Then we can make a bar chart showing the count of properties, categorized by their type; this could be helpful in getting an overview regarding the kind of houses in Toronto.
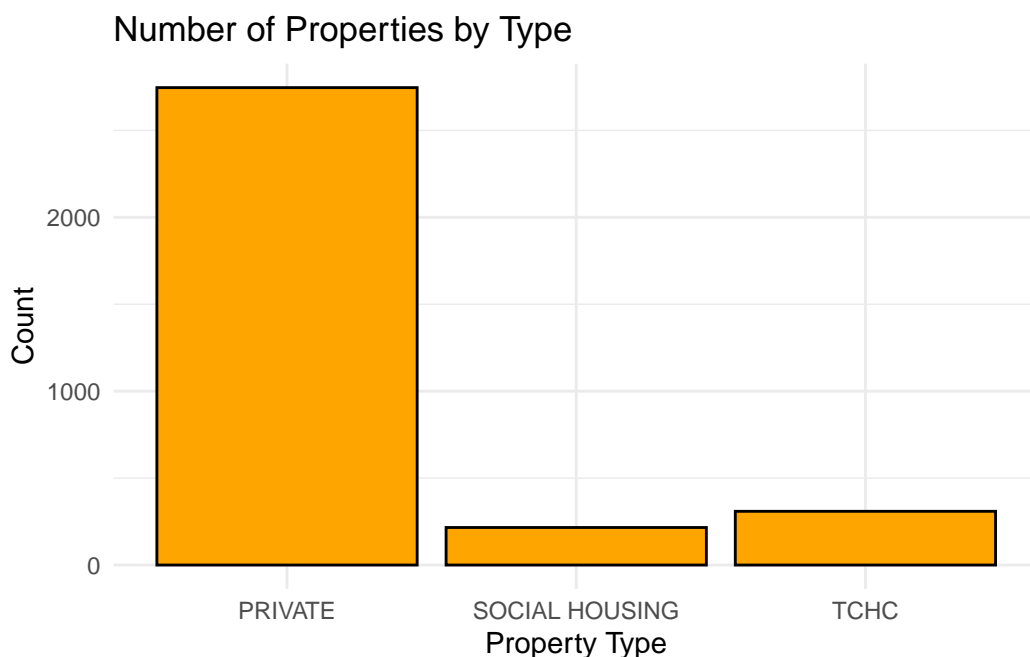
Figure 2: Property Count by Type

Figure 2 shows the distribution of property types. As expected, it is dominated by private properties, which may show something indicative of the ownership structures around the city.

5. Cross-Reference Tables and Graphs

TThis calls for the need to refer to our tables and figures as we interpret our findings in an attempt to make a coherent narrative. Overview information on housing is shown in Table 1. The mean evaluation score of 85.3 is relatively high, suggesting that the majority of the properties were in good condition. Figure 1 also suggests that most of the ratings fall between 80 and 90-all this justifies normally positive reviews. Figure 2 shows the assortment of housing types, with the highest percentage being private property. This could help in building the dynamics within a community, or it could be a contribution to discussions based on housing policy.

Conclusion

This analysis presented an overview of the dataset in terms of housing, also considering some ethical and statistical considerations. Using descriptive statistics and the development of visualizations, we revealed meaningful insights regarding residential relevant properties in Toronto.

Future analysis can be carried out by correlating property types with evaluation scores or researching how location affects the quality of housing. Engaging a community perspective in this would enrich such an analysis and ensure the findings go toward a more equitable urban landscape.

References

Meftuh, N. The Declining Quality of Toronto's Private Rental Towers and the Impending Risks to Low-Income Families (Doctoral dissertation, Toronto Metropolitan University).

Pearce, E. (2021). Tenants rights and high-rise retrofits: policy considerations for Toronto (Doctoral dissertation, Toronto Metropolitan University).

Webber, C., & Zigman, P. (2023). Renovictions: Displacement and Resistance in Toronto.

Zhong, Z., & Gou, Z. (2023). High-rise apartment quality evaluation and related demographic factors: lesson from RentSafeTO programme. Building Research & Information, 51(4), 430-445.