

Using linear regression to prove the relationship between apartment age and the evaluation score of the characteristics of the apartment*

Yuqiao Lin

4/27/2022

Abstract

Several previous research indicated that the aging plumbing systems, floor structure, and heating systems lead to indoor noise decreasing the insulation performance of walls respectively. The data set is pulled from open data Toronto and analyzed using multiple linear regression with the corrected assumptions and low variance inflation factor. Unlike the previous research, apartments with higher age did not actually affect the inspection scores of these mentioned aspects of the apartments in Toronto. In this paper, I only apply power transformation using the Box-Cox method and most of the responses failed to be explained using the reduced model. Using a different stratification method or alternative transformation may improve the model

Introduction

When people prepare to rent an apartment, there are many factors to consider. The interior amenities of the apartment and the surrounding environment are especially important factors. The surrounding environment includes but is not limited to supermarkets, transportation, hospitals, etc., while the interior amenities of the apartment include decoration, ceilings, heating and ventilation systems, etc. However, the age at which the apartment was built is also worth considering. First, residents living in old apartments often complain about indoor noise. Oh (2014) investigated these noise contributing factors: floor impact noise, plumbing noise, sound insulation between units and traffic noise levels. The results show that floor structure and plumbing systems influence noise in apartments. Oh (2014) also claims the acoustic performance of the new apartments is much better than that of the first-generation apartments and the gap is widening. Oh (2014) believes that although it is relatively simple to remove and rebuild flooring structures and plumbing to reduce indoor noise level, there are still but many old apartments that need to be renovated to meet modern living standards. Secondly, Yoon et al. (2020) note that heating system in old apartment buildings has poor performance and high energy consumption. In addition, Choi and Lee (2015) also made a comparison between newly built apartments and old apartments over 40 years old and pointed out that the insulation performance of walls would decline with aging, resulting in an increased risk of condensation.

Although the problem can be solved by improving the exterior walls and windows or replacing them with more efficient equipment, apartment residents may not like the replacement of more efficient equipment or the renovation of the exterior walls. This is because these measures take a long time to implement. Meanwhile, apartment residents may feel uncomfortable or inconvenient during the replacement period (Song and Choi, 2012). However, this relationship is previously shown only in Korea. In this paper, I attempt to reveal the similar relationship between apartment building age and inspection of the apartment building in Toronto through the data provided by Open Data Toronto.

*Code and data are available at: <https://github.com/Yukiu-Lam/-Appartment-further-research>.

Methods and Results

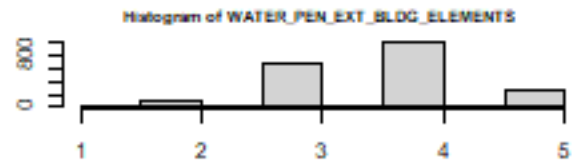
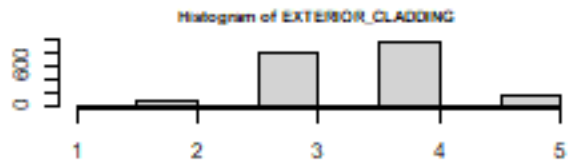
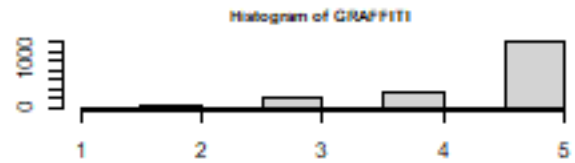
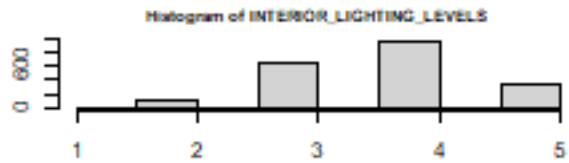
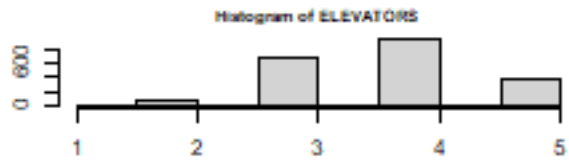
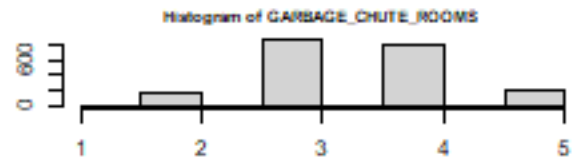
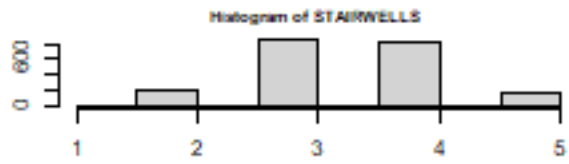
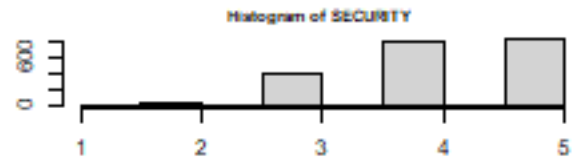
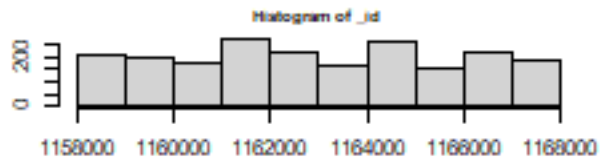
Apartment Building Evaluation from Open Data Toronto is used in this paper (opendatatoronto, 2022). The Apartment Building Evaluation Program named RentSafeTO, works to ensure that apartment owners and operators meet building maintenance standards through measures such as assessments (City of Toronto, 2022). A Bylaw Enforcement Officer will complete inspections of the apartment building, including but not limited to facilities, common areas, elevators, garbage and recycling management, lighting, mechanical systems, security systems, etc.(City of Toronto, 2022). The program applies to apartment buildings with three or more storeys and 10 or more units, hence I assume the apartment discussed in this paper will be defined with more than 3 storeys and 10 units.

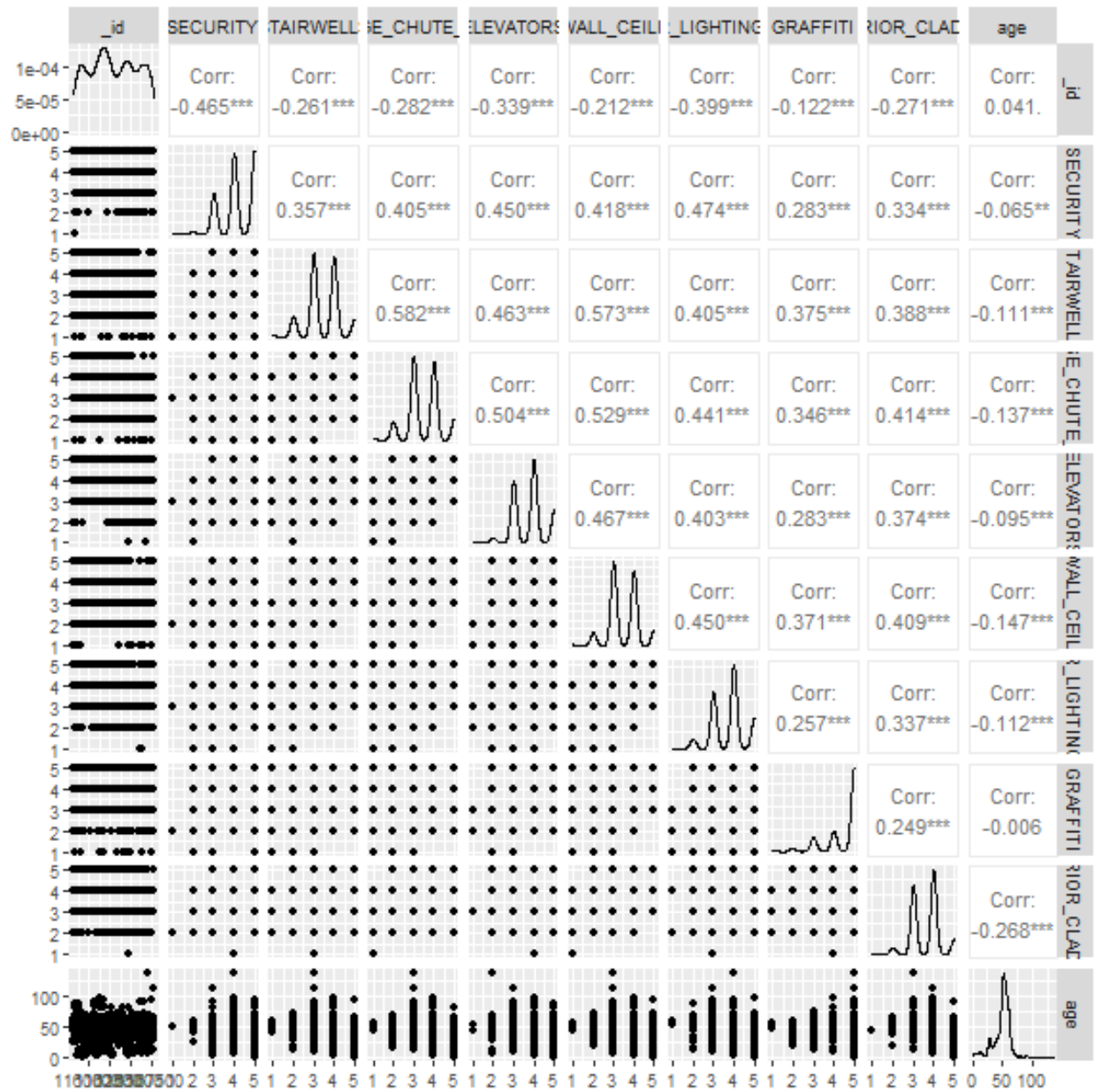
The RentSafeTO team collected the data and Municipal Licensing & Standards published the data, which was updated on February 19, 2022. The raw data includes 25 wards in the Greater Toronto Area, evaluation score, built years, confirmed units and storeys and others (opendatatoronto, 2022). R, tidyverse, knitr and dplyr are utilized to clean and extract the necessary data. Package GGally (Barret Schloerke et al., 2021) is used for making pair plots among the predictors to check both conditions and four assumption of linear regression. Package car (John Fox et al., 2021) is used for power transformation of response and predictors if needed using Box-Cox method. Last, package leaps (Thomas Lumley, 2020) is used for model selection by exhaustive search.

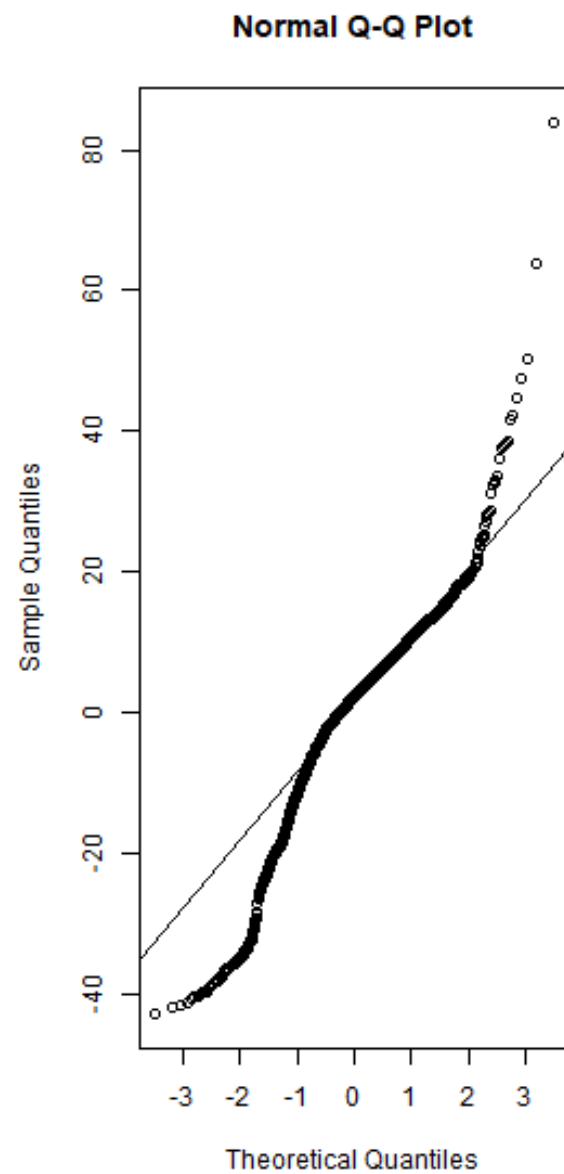
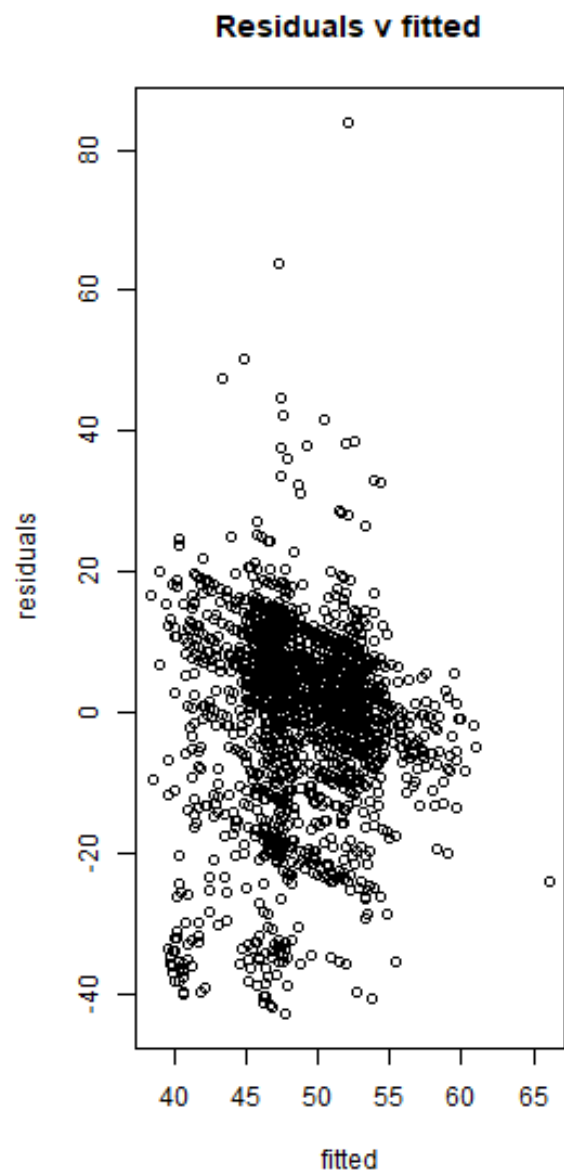
First, the dataset is separate randomly into in 50:50. The one part of the dataset is used for training and the another part is used for testing. The exploratory data analysis of predictors' mean and standard deviation are similar in the training dataset and testing dataset (Table 1). In the first full model, the exploratory data analysis show the number of higher score of security and graffiti are much more than the number of the score below 3. The other predictors are distributed normally (Figure 1). The full model's conditional mean response is a single function of a linear combination of the predictors. Each predictors' condition mean is a linear function with another predictor by checking the scatter plot of all the predictors and a plot of the response against the fitted values made by function in package GGally (Barret Schloerke et al., 2021). Although the assumption of linearity of the relation ship is not violated, this model is still needed to be corrected using the Box-cox method (Figure 2 and 3). After I fixed the full model and create model 1, I check all the conditions and assumptions. There is a slight violation of normality by viewing the Q-Q plot. The patterns of residual plots illustrate no discernible pattern to provide evidence to show the violation of linearity, uncorrelated errors and constant variance (Figure 4 and 5). Both multiple R-squared and adjusted R-squared, however, are both low, 0.07079 and 0.06672 respectively. Then, using the Variance inflation factor measure the amount of multicollinearity of the fixed model. None of the predictors have a larger value than the cut-off value. Multicollinearity is satisfied (Table 2).

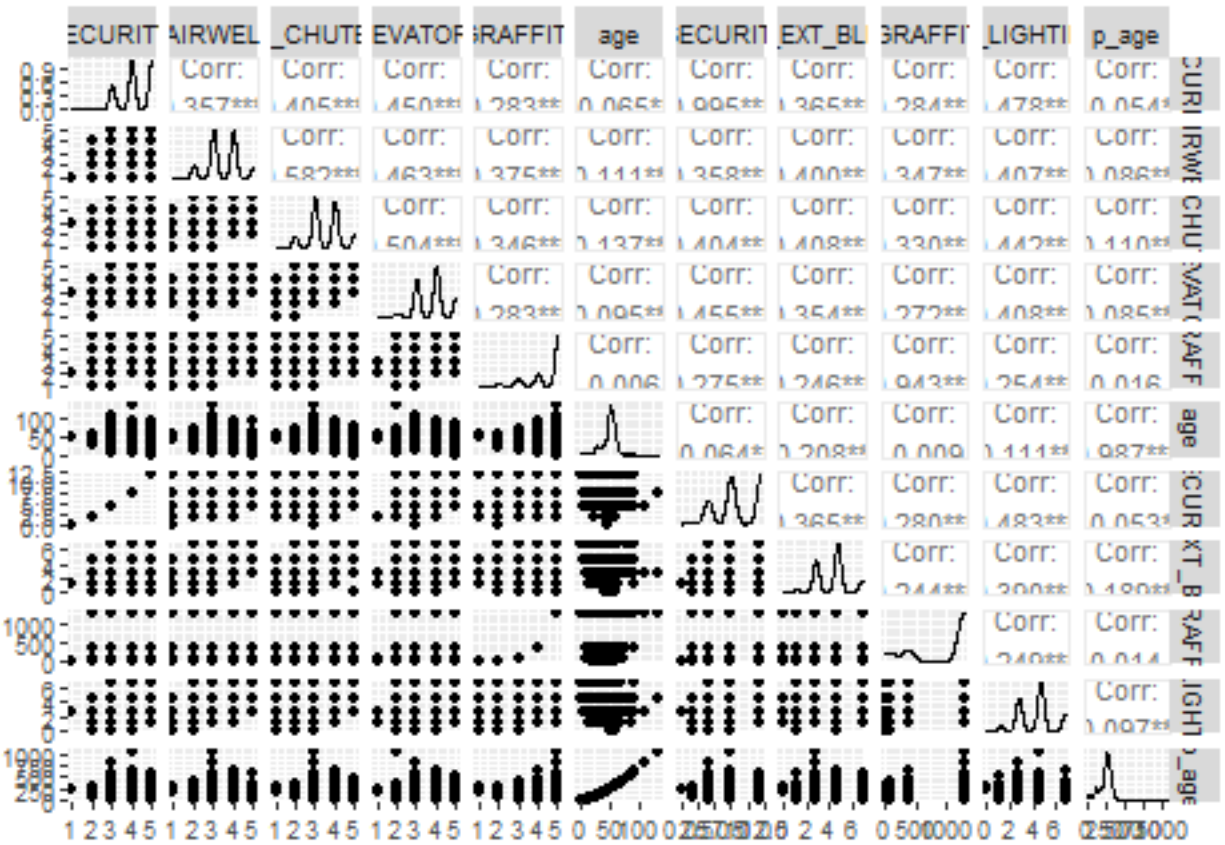
Variable	mean (s.d.) in training	mean (s.d.) in test
_id	1.1629569×10^6 (2827.128)	1.1631358×10^6 (2830.197)
SECURITY	4.174 (0.793)	4.142 (0.799)
STAIRWELLS	3.448 (0.825)	3.469 (0.817)
GARBAGE_CHUTE_ROOMS	3.493 (0.826)	3.491 (0.796)
ELEVATORS	3.787 (0.767)	3.791 (0.784)
INTERIOR_WALL_CEILING_FLOOR	3.479 (0.771)	3.488 (0.763)
INTERIOR_LIGHTING_LEVELS	3.747 (0.797)	3.736 (0.809)
GRAFFITI	4.502 (0.836)	4.505 (0.844)
EXTERIOR_CLADDING	3.608 (0.701)	3.607 (0.709)
WATER_PEN_EXT_BLDG_ELEMENTS	3.701 (0.759)	3.704 (0.737)

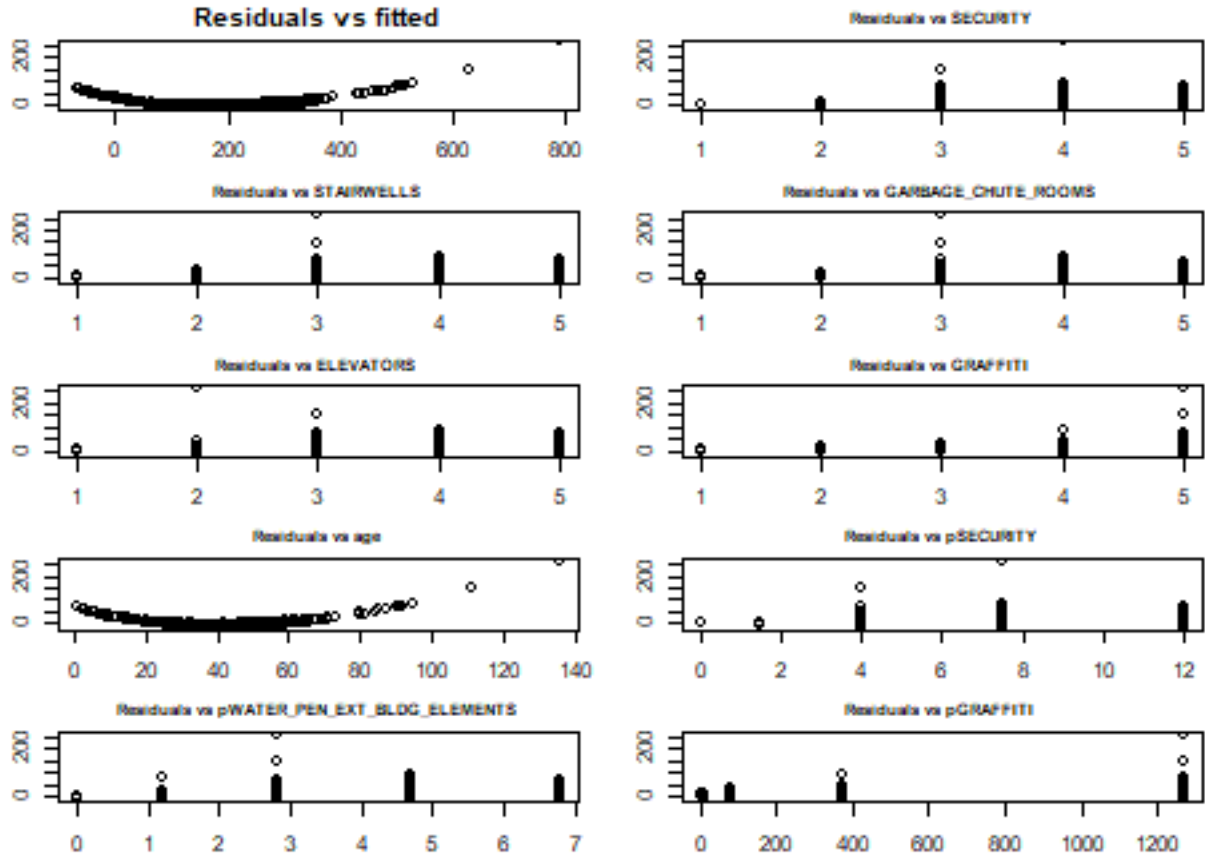
Table1: Comparison of both mean and standard deviation of predictors between training dataset and testing dataset.



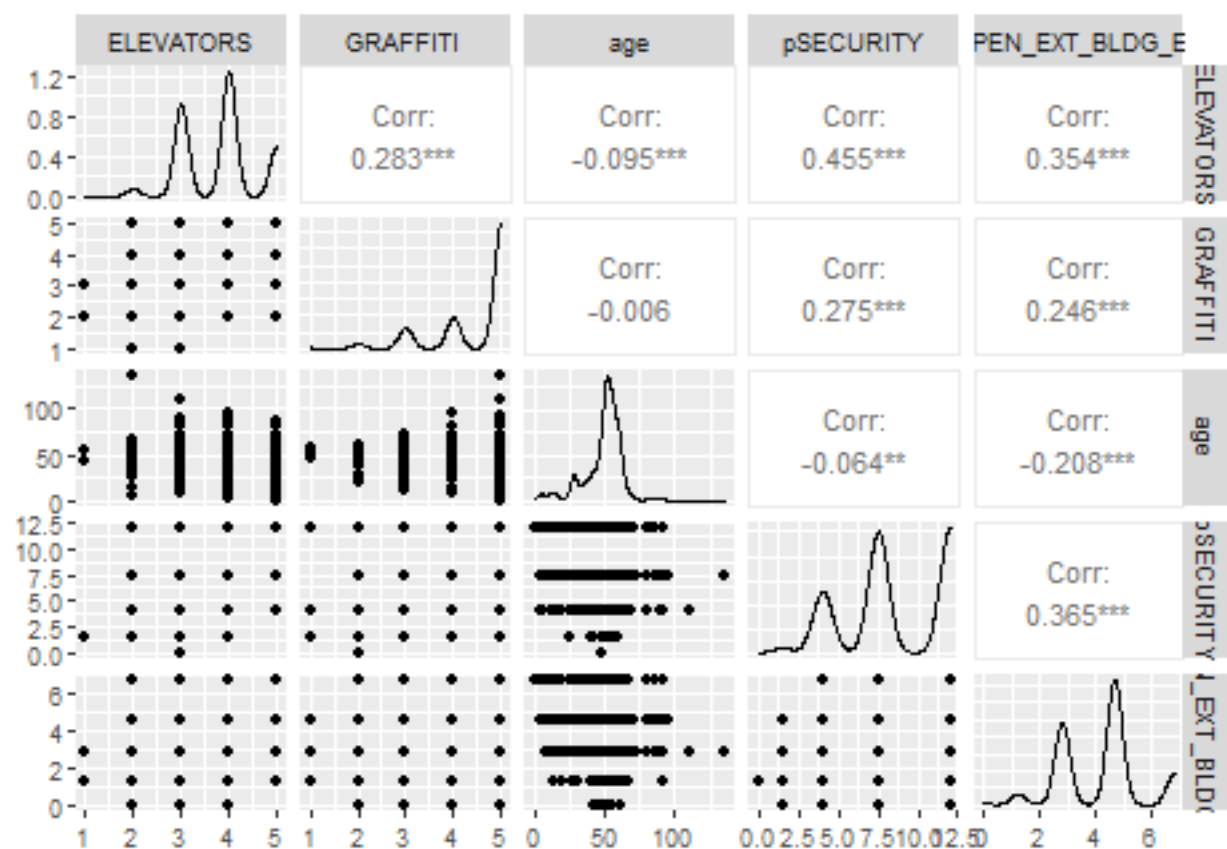


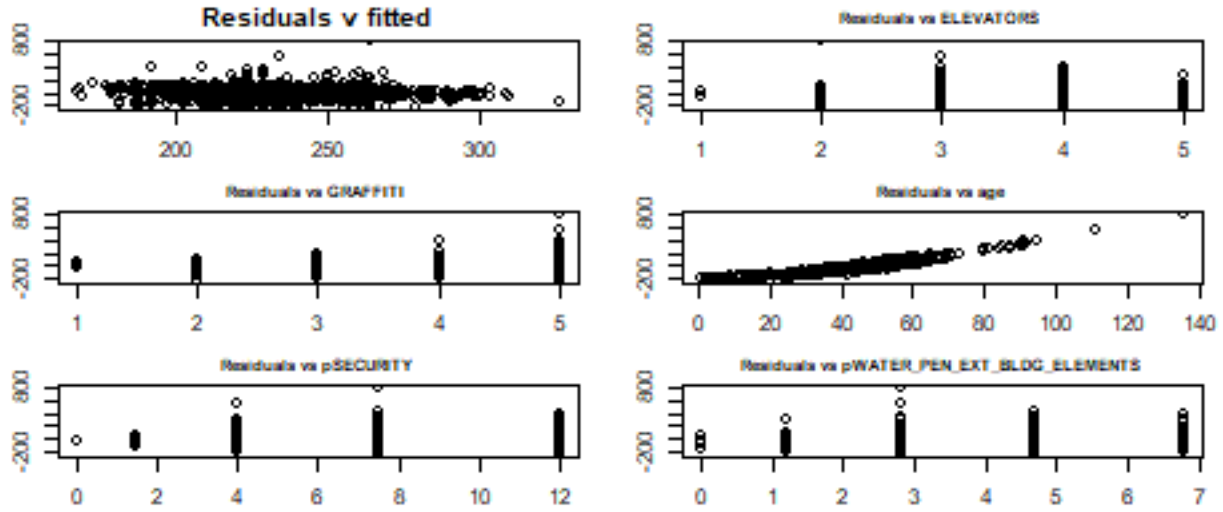




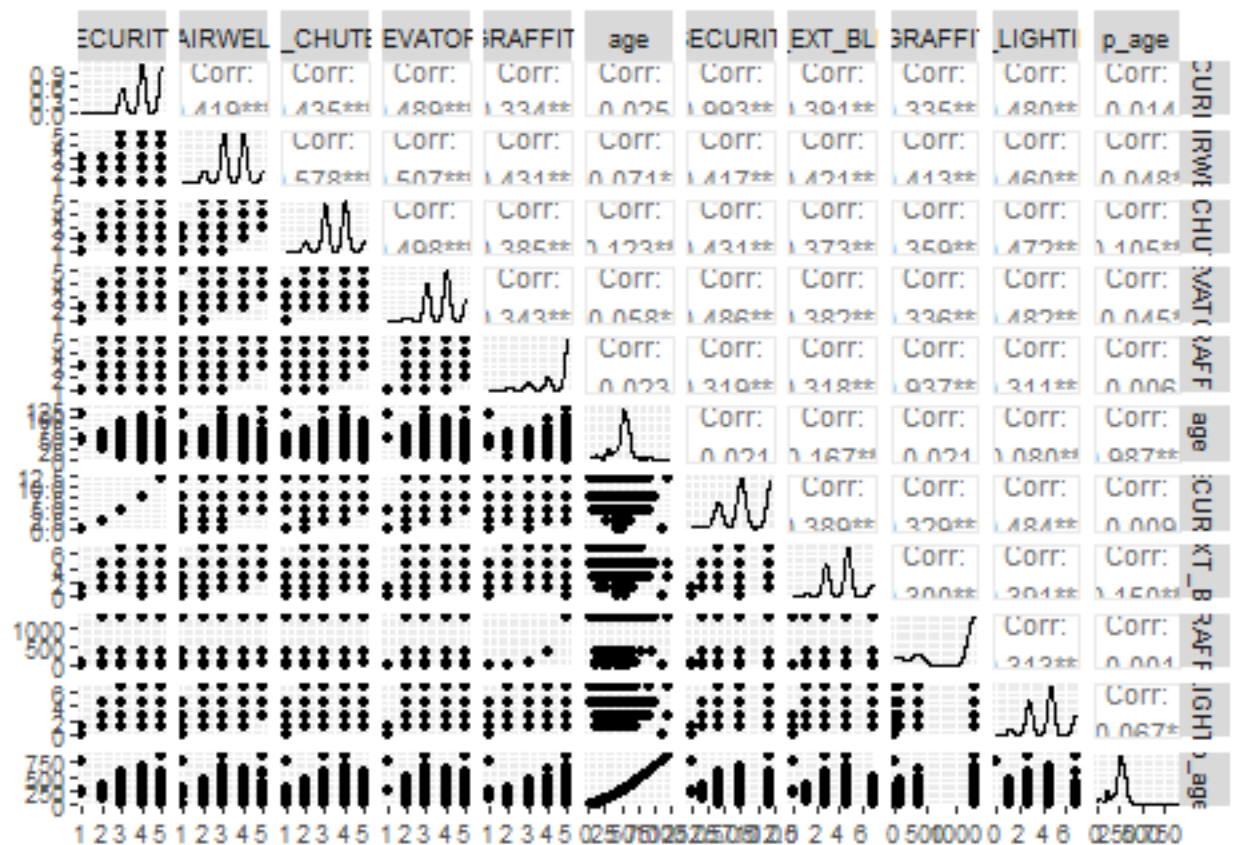


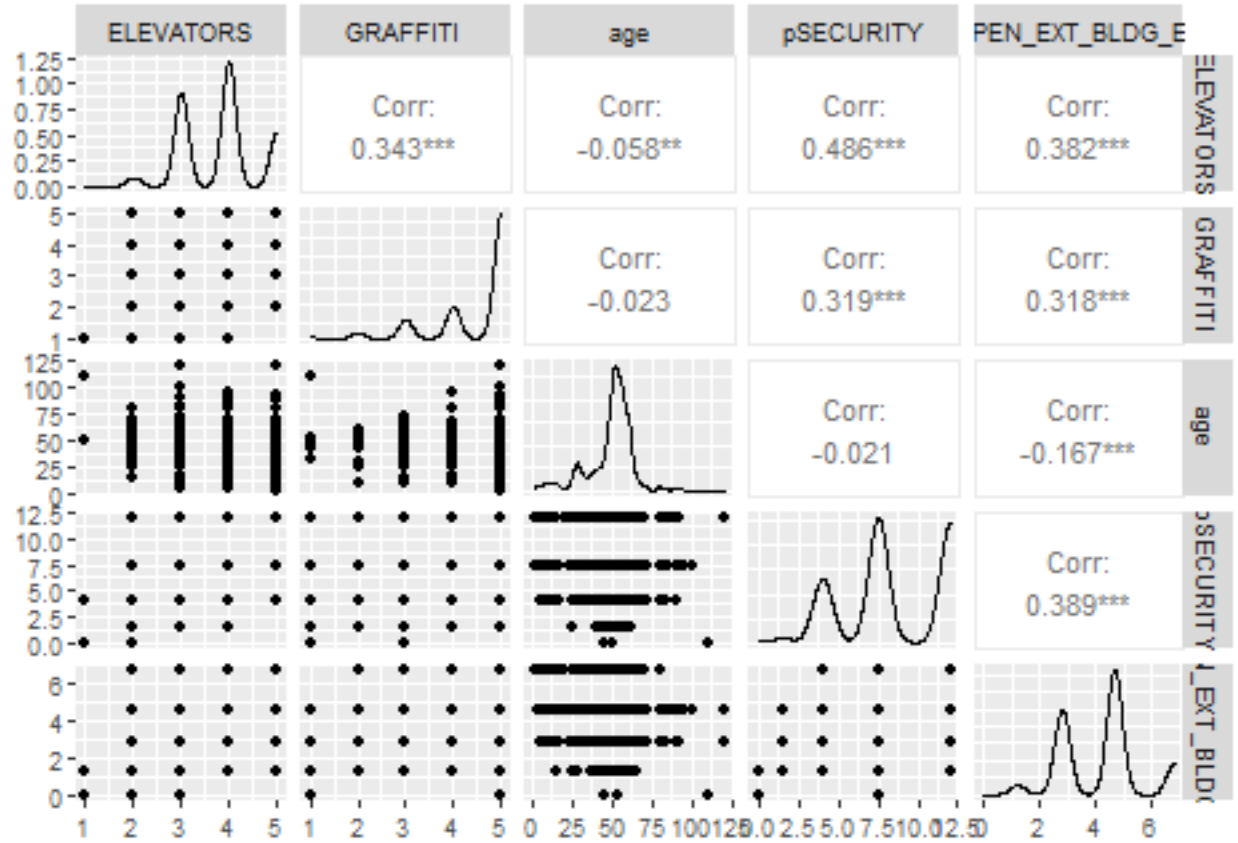
Additionally, the function in package leaps is used for checking all possible subsets of predictors and picking the one have the score of interior wall ceiling floor, the score of exterior cladding, the score of power transformed security, the score of power transformed water penetration of external elements of a building and the score of power transformed graffiti as predictors. This is because the model includes all significant predictors reported by previous studies with higher adjust r square and fewer predictors compared to the other possible subset with relatively higher adjust r square (using function in package leaps). The final model satisfies most of the conditions and assumptions except for a slight violation of normality (Figures 6 and 7). The maximum Variance inflation factor is lower than the cut-off value proving the absence of multicollinearity (Table 2).





Before fitting the training testing data, both the full test model and the reduced test model are checked for the conditions and assumptions. Both models satisfy most of the conditions and assumptions except for a slight violation of normality according to Figures 8 and 9. These test models have a lower maximum Variance inflation factor than the cut-off value, which indicates that predictors do not have multicollinearity (Table 2).





The Hhat matrix element is used to detect leverage points whereas the standardized residuals are used to detect outliers. In addition, using cook's distance, DFFITS (difference in fitted values) and DFBETAS (difference in betas) to detect influential Observations. For all 4 models, the numbers of outliers are smaller than 10. Only 7 observations are leverage point in the reduced test model. But the others have more than 100 leverage points. According to Table 2, there are 0 influential observations detected by Cook's distance while nearly 80-100 observations are detected by DFFITS in both full models (full Model (train) and full Model(test 1)) and approximately 30 observations in both reduced models Table 2.

Discussions

Because both the full and reduced model do not violate the conditions and assumptions, and the partial F test indicates no significant difference with a p-value > 0.05 (0.7588), the reduced model is better than the full model. However, regarding to the Table 3, the adjusted R squares are all low in 4 models. The reduced model have different the significant predictors compared to the test model. Although the coefficient of the predictor of the score of interior wall ceiling and power transformed score of security have different coefficient and standard error, the change are acceptable due to the range of 2 times of standard error. The number of influential observations is acceptable based on the fact that there are over 2000 observations in the training dataset. Hence, the training reduced model is validated.

However, this reduced model is failed to explain the relationship between apartment age and the evaluation score of the characteristics of the apartment because only 0.07 % of observation can be predicted using the reduced model (Table 3). The relationship cannot be proved using these dataset provided by open data Toronto. The dataset is limited to the Great Toronto area. Further study and other dataset is needed for checking the relationship of the other location. In this reduced model, there is still a slight violation of normality using power transformation. There may be another transformation used to fix this problem and fit the dataset better with a higher adjusted R square.

Characteristic	full Model (Train)	full Model (Test)	reduced Model (Train)	reduced Model (Test)
Largest VIF value	96.344011	81.0894969	1.7285254	1.7212321
# Cook's D	0	0	0	0
# DFFITS	79	103	26	37
Violations	slight normality	slight normality	slight normality	slight normality
Intercept	342.661364 ± 12.164075 (*)	329.539920 ± 12.753134 (*)	343.340525 ± 11.008862 (*)	330.093775 ± 11.477377 (*)
STAIRWELLS	2.513354 ± 3.084711	5.709266 ± 3.354806	-	-
GARBAGE	-2.442031 ± 3.086117	-10.199523 ± 3.323726 (*)	-	-
CHUTE ROOMS	0.703932 ± 3.077719	3.910775 ± 3.272446	-	-
ELEVATORS	-5.433211 ± 3.258344	-1.856081 ± 3.631531	-5.643014 ± 2.864353 (*)	-2.655935 ± 3.160849
INTERIOR WALL CEILING FLOOR	-25.723613 ± 3.517570 (*)	-25.380281 ± 3.639546 (*)	-25.848498 ± 3.476023 (*)	-25.826534 ± 3.595249 (*)
CLADDING	1.283108 ± 0.726700	2.337170 ± 0.772265 (*)	1.083463 ± 0.671663	2.244433 ± 0.718270 (*)
p SECURITY	-4.592518 ± 1.730997 (*)	-4.536230 ± 1.839124 (*)	-4.773412 ± 1.697428 (*)	-4.360110 ± 1.823755 (*)
p WATER PENETRATION	0.016081 ± 0.004147 (*)	0.010985 ± 0.004573 (*)	0.016090 ± 0.004075 (*)	0.011180 ± 0.004499 (*)
p GRAFFITI	-1.351833 ± 1.523867	-0.647220 ± 1.637279	-	-
p INTERIOR LIGHTING LEVELS				

Table 2: Summary of characteristics of two candidate models in the training and test datasets. Model 1 uses the score of stairwells, the score of garbage chute rooms, the score of elevators, the score of interior wall ceiling floor, the score of exterior cladding, the power transformed score of security, the power transformed score of water penetration of external elements of a building, the power transformed score of graffiti and the power transformed score of interior lighting levels as predictors, while Model 2 uses the score of interior wall ceiling floor, the score of exterior cladding, the power transformed score of security, the power transformed score of water penetration of external elements of a building and the power transformed score of graffiti as predictors. Response is $\frac{age^{1.5}-1}{1.5}$ in both models. Coefficients are presented as estimate \pm 2SE (* = significant t-test at $\alpha = 0.05$).

full model (train)	full model (test)	reduced model (train)	reduced model (train)	
adjusted R square	0.9754192	0.9753865	0.0676864	0.0502791

Table 3: the adjusted R square of each model

References

- City of Toronto. (2021, October 6). *Rentsafeto Building Evaluations and Audits*. City of Toronto. Retrieved February 20, 2022, from <https://www.toronto.ca/community-people/housing-shelter/rental-housing-tenant-information/rental-housing-standards/apartment-building-standards/rentsafeto-for-building-owners/rentsafeto-building-evaluations-and-audits/>
- Choi, D. S., & Lee, M. E. (2015). Condensation prevention performance assessment taking into account thermal insulation performance degradation due to aging for apartment housing. *KIEAE Journal*, 15(6), 11–18. <https://doi.org/10.12813/kieae.2015.15.6.011>
- Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., et al. (2021). car: Companion to Applied Regression. <https://cran.r-project.org/web/packages/car/index.html>
- Lumley, T. (2020). leaps: Regression Subset Selection. <https://cran.r-project.org/web/packages/leaps/index.html>
- Oh, Y. K. (2014). An assessment model for the indoor noise environment of aged apartment houses. *Journal of Asian Architecture and Building Engineering*, 13(2), 445–451. <https://doi.org/10.3130/jaabe.13.445>
- Opendatatoronto. (2022, February 19). *Apartment Building Evaluation*. City of Toronto Open Data Portal. Retrieved February 20, 2022, from <https://open.toronto.ca/dataset/apartment-building-evaluation/>
- Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., et al. (2021). GGally: Extension to ‘ggplot2’. <https://cran.r-project.org/web/packages/GGally/index.html>
- Song, D., & Choi, Y. J. (2012). Effect of building regulation on energy consumption in residential buildings in Korea. *Renewable and Sustainable Energy Reviews*, 16(1), 1074–1081. <https://doi.org/10.1016/j.rser.2011.10.008>
- Yoon, Y. B., Seo, B., Koh, B. B., & Cho, S. (2020). Heating Energy savings potential from retrofitting old apartments with an advanced double-skin façade system in cold climate. *Frontiers in Energy*, 14(2), 224–240. <https://doi.org/10.1007/s11708-020-0801-1>

Appendix

Extract of the questions from @gebru2021datasheets

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
 - The dataset was created to enable evaluate the building maintenance standards. All prospective tenants have right to know the details about the most recent building evaluation.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
 - RentSafeTO
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
 - Published by: Municipal Licensing & Standards

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
 - Unique row identifier, ID number for a building, year of the building evaluation scores, year that the building was built i, PROPERTY TYPE, the ward that the building is located in, name of the ward, building address, number of storeys in a building, number of units in a building, date the evaluation was conducted on, overall score of the building, score is used to determine whether an audit, number of items that were evaluated, condition of the entrance and/or lobby, condition of the entrance doors and windows in a building, condition of the security system(s), condition of the stairwells, the laundry room(s), the condition of the internal guards and handrails, condition of the garbage/chute rooms, the condition of the garbage bin storage room or outdoor enclosure area, the condition of the elevator(s), the condition of the storage areas/lockers, the condition of internal walls, ceilings and floors, the condition of internal lighting levels, the severity of graffiti, the condition of the exterior grounds, the condition of the exterior walkways, the condition of the balcony guards, the condition of water penetration of external elements, the condition of the parking areas, the condition of other facilities in a building, the grid that the building is located in, the latitude, the longitude, projected X coordinate, projected Y coordinate
2. *How many instances are there in total (of each type, if appropriate)?*
 - 40
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
 - The dataset contain all possible instances. The dataset is not a sample.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
 - The evaluation score of each item is range from 1-5. The total score is out of 100. Ward name is a string. The others are number.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
 - The target is each apartment being evaluated.
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
 - The missing information is randomly distributed. The reason is unknown.
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*

- The overall 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)
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
 - I splits the dataset into two part with identical number of apartments randomly. One is the training data. The another on is testing data.
 9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
 - No
 10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*
 - No guarantees that they will exist, and remain constant, over time. No official archival versions of the complete dataset. No restrictions (for example, licenses, fees) associated with any of the external resources.
 11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
 - Not contain data that might be considered confidential
 12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
 - No
 13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
 - No
 14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
 - No
 15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
 - No

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*
 - A Bylaw Enforcement Officer is assigned for each building evaluation and building owner/operators are notified when their building requires an evaluation. The evaluation is scheduled so that the property owner, or a designate can be present to provide the bylaw enforcement officer access to locked common areas and/or facility amenities.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
 - unknown
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic,*

probabilistic with specific sampling probabilities)?

- No
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
 - A Bylaw Enforcement Officer
 5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
 - The timeframe is contained in the dataset and vary.
 6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
 - No
 7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
 - City of Toronto open data portal
 8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*
 - Yes, Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented
 9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
 - Yes, everyone can use their data.
 10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
 - Unknown
 11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*
 - Ensures that tenants live in safe, well-maintained buildings through RentSafeTO

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*
 - The interest variable is select out. The variable of apartment age is created.
2. *Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data.*
 - Yes, <https://github.com/Yukiu-Lam/-Apartment-further-research>
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
 - R studio

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
 - Yes, I used it to answer another similar research question before.
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
 - <https://github.com/Yukiu-Lam/Apartment>

3. *What (other) tasks could the dataset be used for?*
 - No
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
 - No
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
 - No

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
 - Open data
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
 - City of Toronto open data portal website
3. *When will the dataset be distributed?*
 - anytime
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
 - No
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
 - No
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
 - No

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
 - RentSafeTO
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
 - RentSafeTO@toronto.ca
3. *Is there an erratum? If so, please provide a link or other access point.*
 - No
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
 - Refreshed Daily
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
 - No
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
 - Yes, the old version can be download at the same website.

7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*

- No