Renting an Apartment at the Toronto Community Housing Corporation and Social Housing. Is it a sweet or sour deal?

Anusha Champaneria, Dina Abu Ghosh, Ahmed Lugya

09 February 2020

Abstract

We use Apartment Building Evaluation dataset from Toronto's Open Data Portal (Sharla, 2019) to analyze the security levels of various property types in Toronto from 2017 to 2019. Using R programming (R Core Team,2019) we find that buildings owned privately have better security conditions compared to social housing or Toronto Community Housing Corporation. Our findings identify opportunities for social housing providers and the Toronto Community Housing Corporation to improve their security conditions in the future.It additionally provides further clarity to renters in terms of which type of property they should look into.

Introduction

Toronto is ranked 6th among the 11 most expensive cities to live in 2020 including London UK, Auckland NZ, Victoria CA, San Jose US and San Francisco US (Ben, 2020). Despite the increase in rental fees, apartment security still remains an important factor that renters consider when looking for a new place to live. Additionally, apartment building owners are not always aware of their security measures which can be a barrier for renters. We will conduct an analysis that will indicate security measures for each apartment type which can allow renters to better identify which building type is best suitable for their safety, and alert building owners to know whether they need to upgrade their security system.

Research question: Which type of apartment building; private, social housing, or toronto community housing corporation, evaluated in the past two years, has the highest and lowest security measure in Toronto?

Data Analysis

The Apartment Building Evaluation dataset shows the evaluation of three types of apartment buildings in Toronto, providing 32 attributes. In this dataset the evaluation is done on a scale from 1 to 5 where 1 is the worst and 5 being the best. Building aspects that are evaluated include; balcony guards, condition of the elevator, entrance doors and windows, lobby, exterior cladding, exterior grounds, exterior walkways, garbage bin storage area, chute rooms, severity of graffiti, interior lighting levels, interior walls, ceilings, & floors, internal guards and handrails, laundry rooms, other facilities, parking area, security, stairwells, storage lockers, and water penetration.

Additionally, the dataset includes the confirmed number of storeys in the building, number of reviewed areas done in a single evaluation, the ward in which the building is located in (25 ward system), year building was built in, and overall score of the building based on review during evaluation. Furthermore it provides the date of evaluation, registration ID number, building address, and registration year.

To address our research question, we decided to graph attributes 'security' and 'property type'. Property type attribute includes 2883 private property, 240 social housing (governmental), and 327 Toronto community housing corporation (TCHC)(Table @ref(tab:Table3)). Figure 1@ref(fig:Figure1) shows that majority property has security ranked between 3 and 5. Private has 48% of its property ranked at security level 5 whereas social housing has 9% and TCHC has 7% of their property ranked at 5 (Table @ref(tab:Table1)). However, at security level 4, private property and social housing both have only 35% of their property at this level and TCHC has 55% of its property of this level (Table @ref(tab:Table1)). The dramatic difference between security level 4 and 5 in these property types(Table @ref(tab:Table2)) shows us that private apartment building owners have invested the best security in most of their buildings, whereas social housing and TCHC owners have invested more in moderate security in most of their buildings. On average, private property building is the most secure apartment building type in Toronto, following TCHC, and social housing being the least secure. This could be due to the fact that private apartment building owners accumulate higher revenue because their rent charges are more expensive than a public or government property. It would be advised for the government to increase security in social housing as their security is dramatically low compared to the other property types.

Table 1: Frequency of property type for each security

X	Security	Property_type	Frequency
1	1	PRIVATE	1
2	2	PRIVATE	43
3	2	SOCIAL HOUSING	1
4	2	TCHC	8
5	3	PRIVATE	368
6	3	SOCIAL HOUSING	29
7	3	TCHC	63
8	4	PRIVATE	1002
9	4	SOCIAL HOUSING	83
10	4	TCHC	166
11	5	PRIVATE	1326
12	5	SOCIAL HOUSING	106
13	5	TCHC	67
14	NA	PRIVATE	143
15	NA	SOCIAL HOUSING	21
16	NA	TCHC	23

Table 2: Summary security scores for property type

X	.Security	Number	percent	valid_percent
1	1	1	0%	0%
2	2	52	2%	2%
3	3	460	13%	14%
4	4	1251	36%	38%
5	5	1499	43%	46%
6	NA	187	5%	-
7	Total	3450	-	-

Table 3: Total number of each property type

Property_Type	Sum
PRIVATE	2883
SOCIAL HOUSING	240
TCHC	327

Graph

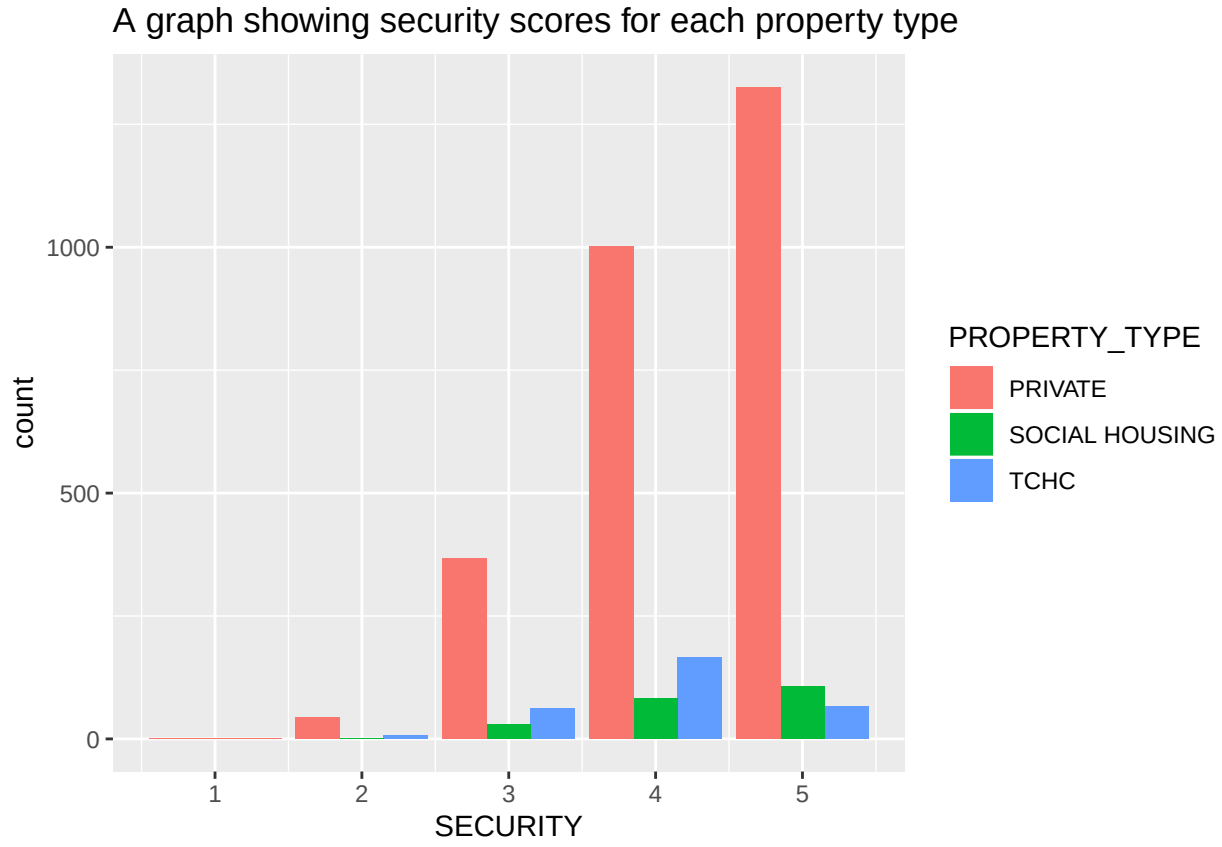

Figure 1:

Ethics

Informed consent is extremely crucial when collecting data (Thomas et al., 2017) , however, it is not mentioned anywhere whether informed consent was obtained. The dataset contains sensitive data such as address and locations. The fact that it is accessible to the public can pose a safety threat to the residents. A person with malicious intent who intends to commit theft can easily look at the properties with the lowest security measure and use the address that is readily provided to get to the property. This is just an example to depict how sensitive data can be missed with ill intent.

The intent of the study has to be listed as part of research ethics (Thomas et al., 2017), which in this case is to help owners know where their property stands in terms of security and accessibility in comparison to other properties. This helps aid with improving inadequate safety measures and implementing better ones.

When looking at the dataset from Toronto's Open Data Portal (Sharla Gelfand, 2019) the intent of the study is clearly stated which is complaint with research ethics.

Shortcomings

The dataset had a total of 12692 missing values. This consisted of 143 security missing values for private property type, 21 security missing values for social housing, and 23 security missing values for TCHC Table @ref(tab:Table1). These values could have made a significant difference in the percentage of property within each security level and possibly increasing the number of social housing property at the higher security rank.

One of the main limitations, is that the fiscal year that represents the year of the evaluation is not the same as program year that the evaluation is completed in (Sharla, 2019)

If the overall evaluation score is relatively low, then an audit has to be conducted. The timing in which the follow up audit is conducted is very important, as the scores of certain attributes can change over a long period of time.

Biases: The number and focus of attributes that the score is based on, could be not fully comprehensive. As some attributes such as "graffiti" which its ranking scale states whether there is graffiti or not. In certain cases, graffiti is used as a form of art and is not viewed as an act of vandalism. Which can be quite misleading as some buildings have intentional graffiti, which is supervised and monitored by the property owner so it would be informative to simply associate the presence of graffiti with a direct lower score.

APPENDIX

Source Code

Gather Apartments data

```
#### Contact details ####
# Title: Get data from Open data Toronto
# Purpose: This script gets data from Open Data Portal and saves it to inputs.
# Author: Dina, Anusha, Ahmed
# Last updated: 09 February 2020
# License: MIT License.

#### Set up workspace ####
library(rvest)
library(tidyverse)
library(opendatatoronto)
library(dplyr)

# get package
package <- show_package("4ef82789-e038-44ef-a478-a8f3590c3eb1")
package

# get all resources for this package
resources <- list_package_resources("4ef82789-e038-44ef-a478-a8f3590c3eb1")

# identify datastore resources; by default, Toronto Open Data sets datastore resource format to CSV for
datastore_resources <- filter(resources, tolower(format) %in% c('csv', 'geojson'))

# load the first datastore resource as a sample
data <- filter(datastore_resources, row_number()==1) %>% get_resource()
data

#Save the dataset to inputs
write.csv(data, "inputs/raw_apartment_building_data.csv")
```

Clean Apartments data

```
#### Contact details ####
# Title: Data cleaning
# Purpose: This script gets data from inputs and does some cleaning.
# Author: Dina, Anusha, Ahmed
# Last updated: 09 February 2020
# License: MIT License.

library(tidyverse)
library(rvest)

clean_apartment_data <- read.csv("inputs/raw_apartment_building_data.csv")
```

```
#### Clean data ####
summary(clean_apartment_data)
sum(is.na(clean_apartment_data))
#### Save the data ####
write.csv(clean_apartment_data, "outputs/data/clean_data.csv")
```

Exploratory Data Analysis

```
#### Contact details ####
# Title: Perform an analysis on the data
# Purpose: This script performs an analysis on the dataset.
# Author: Dina, Anusha, Ahmed
# Last updated: 09 February 2020
# License: MIT License.

#### Set up workspace ####
library(janitor)
library(tidyverse)

apartment_dataEDA<- read.csv("outputs/data/clean_data.csv")

#### Summary stats ####
skimr::skim(apartment_dataEDA)

table(apartment_dataEDA$SECURITY) %>%
  knitr::kable(col.names = c("Security", "Frequency"))

summary(apartment_dataEDA)
```

Make graphs and tables

```
#### Contact details ####
# Title: Graphs and Tables
# Purpose: This script draws graphs and tables out of the dataset.
# Author: Dina, Anusha, Ahmed
# Last updated: 09 February 2020
# License: MIT License.

#### Set up workspace ####
library(janitor)
library(tidyverse)

# Calculate the security score per property type
score_count <- read.csv("outputs/data/Clean_data.csv")
counts <- ddply(score_count, .(score_count$SECURITY, score_count$PROPERTY_TYPE), nrow)
names(counts) <- c("Security", "Property_type", "Frequency")
counts
```

```r
#### Make tables ####
# Make table1
count <- counts %>%
  pull() %>%
  tabyl() %>%
  adorn_pct_formatting(digits = 0) %>%
  adorn_totals()
# Make table2
counttable2 <- score_count %>%
  pull(SECURITY) %>%
  tabyl() %>%
  adorn_pct_formatting(digits = 0) %>%
  adorn_totals()

#rename tables
counttable2 <-
  counttable2 %>%
  rename(`.` = "Security",
         Number = Score,
         Percent = percent)
summary_property_type <- summary(score_count$PROPERTY_TYPE)
summary_property_type

# Make a graph
score_count %>%
ggplot( aes(x = SECURITY, fill = PROPERTY_TYPE)) +
geom_bar(position = "dodge") +
ggtitle("A graph showing security scores for each property type")

#write to a csv file named counts4
write.csv(counts, "outputs/tables/counts4.csv")
#write to a csv file named counts5
write.csv(counttable2, "outputs/tables/counts5.csv")
#write to a csv file named counts6
write.csv(summary_property_type, "outputs/tables/counts6.csv")
```

References

- Yihui Xie (2019). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.26.
- Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963
- Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595
- Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2019). skimr: Compact and Flexible Summaries of Data. R package version 2.0.2. https://CRAN.R-project.org/package=skimr
- Hadley Wickham and Lionel Henry (2019). tidyr: Tidy Messy Data. R package version 1.0.0. https://CRAN.R-project.org/package=tidyr
- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
- B. Hofner (2019). papeR: A Toolbox for Writing Pretty Papers and Reports, R package version 1.0-4, https://CRAN.R-project.org/package=papeR.
- Firke, Sam (2020). janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 1.2.1. https://CRAN.R-project.org/package=janitor.
- Müller, Kirill (2017). here: A Simpler Way to Find Your Files. R package version 0.1. https://CRAN.R-project.org/package=here.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
- Wickham, Hadley, et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686.
- Wickham, Hadley (2019). rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3.5. https://CRAN.R-project.org/package=rvest.
- Sharla Gelfand (2019). opendatatoronto: Access the City of Toronto Open Data Portal. R package version 0.1.1. https://CRAN.R-project.org/package=opendatatoronto
- Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr
- Jessica Myles. Average rent in Toronto since 2000, (November 11, 2017) Retrieved from https://www.torontorentals.com/blog/average-rent-in-toronto-since-2000
- LowestRates.ca (2020) Cost of living in Toronto in 2020. Retrieved from https://www.lowestrates.ca/cost-of-living/toronto
- Ben Mack(January 20, 2020) The 11 most expensive cities to live in around the world in 2020.Retrieved from https://www.insider.com/most-expensive-cities-to-live-in-2020-1#6-toronto-canada-86-6
- Jessica Myles (November 11, 2017). Average Rent in Toronto Since 2000. Retrieved from https://www.torontorentals.com/blog/average-rent-in-toronto-since-2000
- Thomas, D., Pastrana, S., Hutchings, A., Clayton, R. and Beresford, A. (2017). Ethical issues in research using datasets of illicit origin. Cambridge Cybercrime Centre, Computer Laboratory, p.2.