Ontario Renting: Localtion Difference & Pets are Effecting Rent

Yan Wang

21/12/2020

Abstract

This project focuses on how location difference (mainly about Downtown Toronto, Scarborough, and Mississauga) in Ontario and wether the tenant have a pet could affect the rent.Statistical models such as linear regressions are used in the analyses. The results lead to the conclusion that both location difference and pets have influences on rent in Ontario.

Keywords

Keywords: rent, Toronto, Mississauga, Scarborough, pet, apartment, condo, size, bedroom, bathroom, den.

Itroduction

With the special time of COVID-19 period, major students are staying at home for quarantine and having online courses in rental apartments. Most of them live in Downtown Toronto with the purpose to be more convenience to get to university. However, is living in Downtown become necessary in this period? Obviously not. There are many other options for students to choose since there is no need to attend the class in person. Both Scarborough(along 401 highway) area and downtown Mississauga are popular for people to live with larger supermarkets, convenience transportation, and diverse restaurants. In addition, you can find a larger room size in these two places since apartments in Downtown Toronto are all small-scale for the reasons of the limit land for residential buildings. More and more people are consider to settle down in these two locations rather than in Downtown.

One another factor that effecting rent is whether the tenant carring a pet. Some of rental apartments have strict regulation for the pets not only from land owner, also from the appartment management.

We would like to find the relationship between location difference and pets on rent in Ontario. Given that room size, room amount may be related, and pet species may be related. In the following part of the report I will introduce our the data, the statistical models that help to visualize the relationship. Eventually we will give a conclusion based on our analysis result, explaining how location difference and pets have influences on rent in Ontario. More detailed information about the variables and models will be further explained in the sections after.

Methology

Source

We chose to analyze the data from Toronto Rentals, the rent information provided by apartment/condo management department and the house owner.

Target population

The target population in the data includes all available listings in three different areas.

Smapling methodology and Sample

By observing data, I remove the variable "parking" since it is missing lots of values. The data has simple random sample without replacement of records was performed in each stratum. Respondents were then randomly selected from the apartments/condos in different locations that registered in the system.

Selected variables

In order to analyze the affection on rent we select the variables below:
bedroom: Number of bedroom in the suite.
bathroom: Number of bedroom in the suite.
size: Suite size in square feet.
den: Number of bedroom in the suite.
pet: Whether the pet is allowed in suite.
location: Location of the suite. "T" represents Toronto, "M" represents Mississauga and "S" represents Scarborough.

Data

```
## # A tibble: 6 x 8
##      id  rent bedroom bathroom `size(sqf)`   den pet   location
##   <dbl> <dbl>   <dbl>    <dbl>       <dbl> <dbl> <chr> <chr>
## 1    71  1895       1      1           605     0 NO    T
## 2   313  1567       1      1           685     0 NO    S
## 3   223  2505       3      1.5        1215     0 YES   S
## 4    54  1750       1      1           679     0 NO    T
## 5   298  1525       2      1           840     0 NO    S
## 6   177  1989       2      1           805     0 NO    M
```

The rental information are randomly selected around or along
(1) University of Toronto in variable location T,
(2) Square one zone in variable location M,
(3) 401 highway in variable location S.
The variable "pet" are separate as YES and NO provide by owners and apartments.
The variable "rent","bedroom","bathroom","size","den" are numerical variable. Since the data are including both categorical and numerical variable, we use 0 to represent do not allow to have a pet, and 1 represents allow to have a pet.

Model

```
## # A tibble: 8 x 7
##   term         estimate std_error statistic p_value lower_ci upper_ci
##   <chr>           <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept     897.       74.2       12.1    0       751.    1043.
## 2 locationS     -68.3      50.1       -1.36   0.174 -167.       30.3
## 3 locationT     299.       50.6        5.9    0       199.     398.
## 4 bedroom       325.       35.9        9.07   0       255.     396.
## 5 bathroom      317.       50.8        6.24   0       217.     417.
## 6 den           137.       62.3        2.21   0.028   14.9    260.
## 7 petYES         82.3      49.6        1.66   0.098  -15.3    180.
## 8 `size(sqf)`     0.188     0.061      3.09   0.002    0.068    0.307
```

In order to determine the affections to rent, we the variable "location", "bedroom", "bathroom", "den", "pet", and "size". Since the "location" and "pet" are categorical variables, we have the reference group location in Mississauga as intercept from the linear model, and petYES count as 1 and petNO count as 0. The reason to choose a linear regression model to test the relationship of location and rent is to estimate the final rent after considering all the factors that may effect the rent. Our goal is find out how each variable effects the rent in Great Toronto area.

The model is:

$$\hat{rent} = \beta_0 + \beta_1 X_S \ + \beta_2 X_T + \beta_3 X_1 + \beta_4 X_2 + \beta_5 X_3 + \beta_6 X_P + \beta_7 X_4 + \epsilon_i$$

$\hat{rent}$:represents the rent of the suite in the model.

X_S:represent the suite located in Scarborough.

X_T: represent the suite located in Toronto.

X_1:represents number of bedroom in the suite.

X_2:represents number of bathroom in the suite.

X_3:represents number of den in the suite.

X_P:represents the suite allows pet.

X_4:represents the size of the suite.

$$\beta_0 = 896.69774, \beta_1 = -68.31753, \beta_2 = 298.50183, \beta_3 = 325.11566, \beta_4 = 317.30697$$

$$\beta_5 = 137.43215, \beta_6 = 82.28467, \beta_7 = 0.18763$$

$$\hat{rent} = 896.69774 - 68.31753X_S \ + 298.50183X_T + 325.11566X_1 + 317.30697X_2$$

$$+137.43215X_3 + 82.28467X_P + 0.18763X_4 + \epsilon_i$$

When X_S the location is at Scarborough the average rent decrease 68.31753, given other predictors hold constant.
When X_T the location is at Toronto, the average rent increase 298.50183, given other predictors hold constant.
When X_1 increase by one unit, the average rent increase 325.11566, given other predictors hold constant.
When X_2 increase by one unit, the average rent decrease 317.30697, given other predictors hold constant.
When X_3 increase by one unit, the average rent increase 137.43215, given other predictors hold constant.
When X_P allows to have a pet in suite, the average rent increase 82.28467, given other predictors hold constant.
When X_4 increase by one unit, the average rent increase 0.18763, given other predictors hold constant.

Based on the linear regression model and t-test result, given 5% significant level, the significant variables are intercept, locationT, bedroom, bathroom, den and size since their p-values are less than 0.05 and reject the null hypothesis H_0=0.

Next, we use AIC and BIC to test the model.
AIC:

```
## Start:  AIC=3516.28
## rent ~ location + bedroom + bathroom + den + pet + `size(sqf)`
##
##               Df Sum of Sq      RSS    AIC
## <none>                     35018067 3516.3
## - pet          1    330296 35348364 3517.1
## - den          1    584205 35602273 3519.2
## - `size(sqf)`  1   1148175 36166242 3524.0
## - bathroom     1   4671981 39690048 3551.8
## - location     2   6952901 41970968 3566.6
## - bedroom      1   9858444 44876511 3588.7

##
## Call:
## lm(formula = rent ~ location + bedroom + bathroom + den + pet +
##     `size(sqf)`, data = df)
##
## Coefficients:
## (Intercept)    locationS    locationT      bedroom     bathroom          den
##    896.6977     -68.3175     298.5018     325.1157     317.3070     137.4322
##      petYES  `size(sqf)`
##     82.2847       0.1876
```

BIC:

```
## Start:  AIC=3517.86
## rent ~ location + bedroom + bathroom + den + pet + `size(sqf)`
##
##               Df Sum of Sq      RSS    AIC
## <none>                     35018067 3517.9
## - pet          1    330296 35348364 3518.5
## - den          1    584205 35602273 3520.6
## - `size(sqf)`  1   1148175 36166242 3525.3
## - bathroom     1   4671981 39690048 3553.2
## - location     2   6952901 41970968 3567.8
## - bedroom      1   9858444 44876511 3590.1

##
## Call:
## lm(formula = rent ~ location + bedroom + bathroom + den + pet +
##     `size(sqf)`, data = df)
##
## Coefficients:
## (Intercept)    locationS    locationT      bedroom     bathroom          den
##    896.6977     -68.3175     298.5018     325.1157     317.3070     137.4322
##      petYES  `size(sqf)`
##     82.2847       0.1876
```

Both AIC and BIC shows the full model, which is

$$\hat{rent} = 896.69774 - 68.31753X_S \ + 298.50183X_T + 325.11566X_1 + 317.30697X_2$$

$$+137.43215X_3 + 82.28467X_P + 0.18763X_4 + \epsilon_i$$

By observing the raw data, I found that the room size also has a huge differences in different locations

```
## # A tibble: 2 x 7
##   term          estimate std_error statistic p_value lower_ci upper_ci
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept     547.     101.          5.42        0   347.     748.
## 2 `size(sqf)`     2.31     0.134     17.3          0     2.05     2.58
```

```
## # A tibble: 2 x 7
##   term          estimate std_error statistic p_value lower_ci upper_ci
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept   1331.          92.5      14.4        0 1147.     1515.
## 2 `size(sqf)`    0.852        0.1       8.56       0    0.655     1.05
```

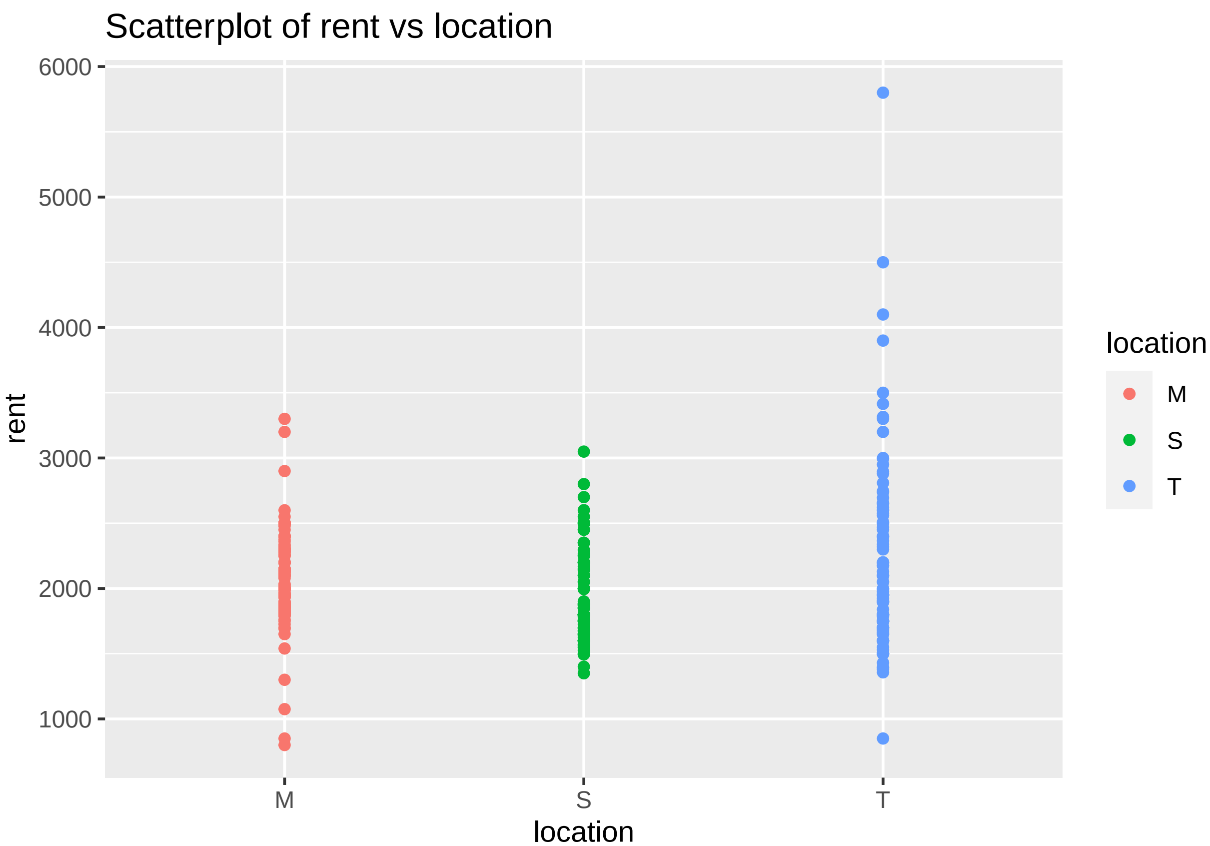
```
## # A tibble: 2 x 7
##   term          estimate std_error statistic p_value lower_ci upper_ci
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept    1864.      67.3       27.7    0     1730.     1998.
## 2 `size(sqf)`     0.08     0.066      1.21   0.228   -0.051     0.21
```

The first graph is location in Toronto, the second graph is location in Mississauga, and the third one is in Scarborough. By comparing the coefficient of the size, location in Toronto increases the most by 2.311 dollar in rent by increasing 1 unit square feet of the room, and Scarborough increases the least by 0.080 dollar in rent by increasing 1 unit square feet of the room. Mississauga increases 0.852 dollar in rent by increasing 1 unit square feet.

Result

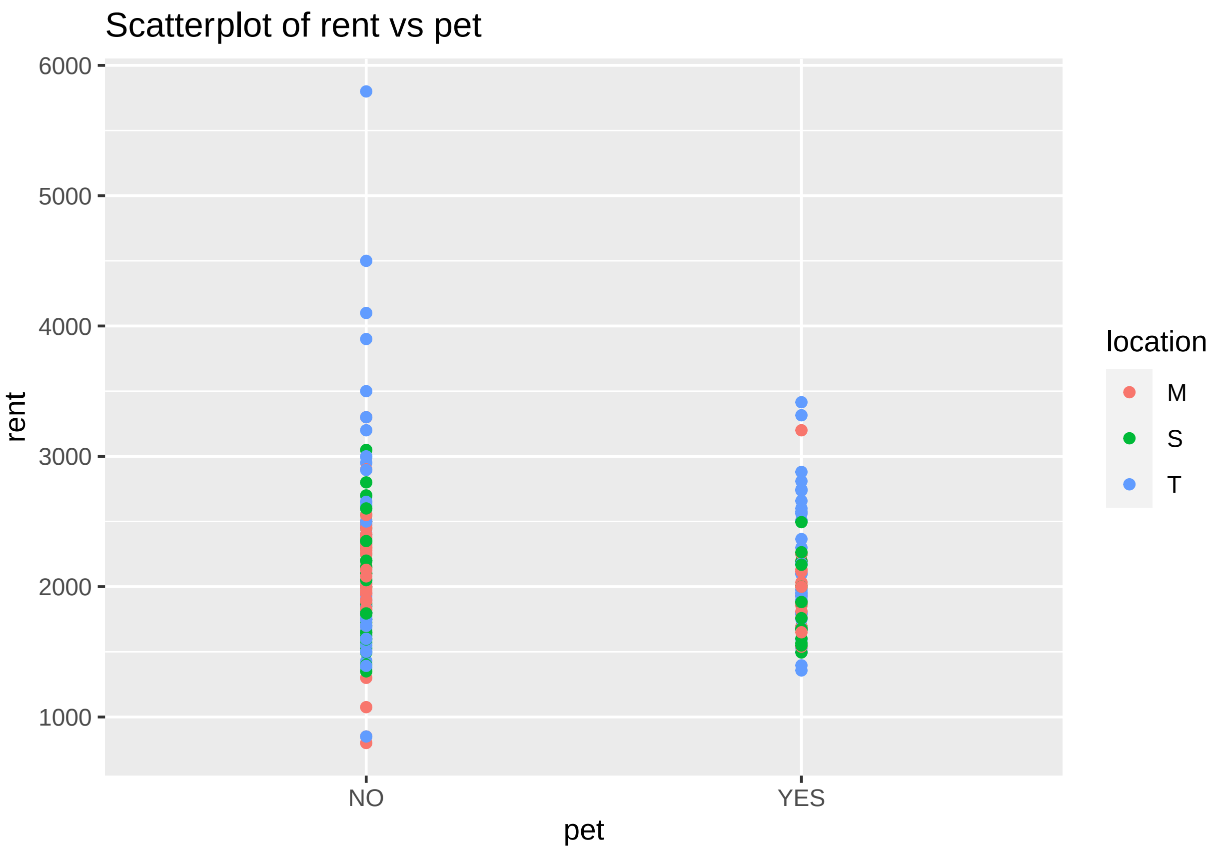
Firstly, we create a multiple linear model for rent with predictors.

Figure1



By figure1 we observe that rent in Toronto has a wider range, and rent in Scarborough has narrower range. We get the mean rent 2182.708 in Toronto, 2083.58 in Mississauga and 1933.362 in Scarborough. Toronto has the highest mean rent, and Scarborough has the lowest mean rent.

Figure2



By figure2 we observe that rent with no pet has a wider range, and rent with pet has narrower range in all Toronto, Scarborough and Mississauga. The result shows that with pet restriction has more choices than without pet restriction.

By the linear model regression, we clearly get that location and pet regulation do effect the rent in Great Toronto area. Scarborough has the lowest rent with decreasing 68.3175 in average compare to other places. As mentioned before, we count without pet as 0, and with pet as 1. The model shows the rent increases 82.2847 on average, in all locations.

By other three models of different location and how the size will effect the rent, we can get the result that the large the house in Toronto, the more expensive rent, however the same size suite in Scarborough will be much cheapper.

Discussion

Summary

The idea in the beginning is to find the relationship between different locations and rent, and how pet regulation will effect the rent. Then I clean the data and make sure the data is randomly select and the response are valid. Next we use linear regression model to test the accurate number of the rent and confirm the positive regression. Then we separate the data by location and test the linear regression separately to investigate the relationship between house size and rent.

Conlusion

The model represents that living in downtown Toronto has a higher living cost from the basic housing. In addition, living in downtown Toronto is better to rent a small size apartment since the 1 unit of square feet of size is 2 times the amount compare to the same 1 unit square in Mississauga and almost 30 times in Scarborough. By observing the raw data in the original website, the major residence for rent are the houses, or basement with a large room size in Scarborough, however in downtown Toronto, it is hard to rent a house or basement around the university of Toronto and the only choice is to rent a small condo. Living in downtown Mississauga is also very expensive, but there still have house to rent for another option. Beside, the cost for 1 unit square of feet in Mississauga is still cheaper than in Toronto. Using model to calculate the rent, we can get the result is if you are going to live in Toronto, than it is cheaper to rent a small condo, and if you are going to live in Scarborough, it is better to rent a large condo.
Pet regulation is always a hard question. By collecting the data,if we carrying a pet, the choices are less and the buildings are old. However, if we selecting apartments without a pet, the option is a lot more. The model also mentioned that if we carrying a pet, the rent will increase 82.285 in total every month, that is also a huge a mount if you are going to rent for 2 or more years.

Key features, strenths and weakness

The dataset collected a large amount of data for each selected respondent as well as some information about household, including rent, bedroom, bathroom,size,den,pet,location.

In order to avoid invalid answers, this data has eliminated all missing value. This methodology the accuracy of the data set. The large amount of data also enables statistician to explore various topic to rent in Canada.

The model still have the weakness part like the number of bedroom and bathroom. In downtown Toronto, most condo have 2 bedrooms and 2 bathrooms at most, but in Mississauga and Scarborough, there exist house layout with 3 or more rooms.

The future step I would like to take is to have a survey through those who already settle down in Downtown Toronto, Downtown Mississauga, and Scarborough(alone 401 highway) and investigate by three major parts, first is the rent, second is the house size and third is the score of satisfaction of the place you currently living.

Appendix

https://github.com/yyanwang1/304-Final

Reference

[1]H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. [2]RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/ [3]Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686 [4]Yihui Xie andJ.J. Allaire and Garrett Grolemund (2018). R Markdown:The Definitive Guide. Chapman and Hall/CRC. URL https://bookdown.org/yihui/rmarkdown. [5]Yihui Xie (2020). tinytex: Helper Functions to Install and Maintain 'TeX Live', and Compile 'LaTeX' Documents. R package version 0.21. [6]"Toronto, ON Apartments, Condos & Houses For Rent." TorontoRentals.com, www.torontorentals.com/toronto.