Exploring venues in Toronto, Canada and the correlation with Rental Price

Iris Yeung

15 January 2020

1.   Problem

This project purposed to find out what factors are affecting the rental price of housing in Toronto and to analyze the rental price in a statistical approach. A quantitative model could be built by the finding to provide insight and forecast future pricing movement.

2.   Introduction

2.1 Background

Toronto is the most populous city in Canada, with an urban agglomeration of over 9,000,000 people, it is home to the Toronto Stock Exchange, meanwhile economy is highly diversified in technology, design and financial services. With these highly developed natures and the huge size population, the demand for housing has never decreased. Analysis of the rental price in a statistical approach would benefit people who looking for renting a house in Toronto with more completed information and helping them make a better decision.

2.2   Interested audience

The target audience for such a project is a person who would like to rent a house in Toronto, can use the research from this project to quickly select places that suit their budget and requirement. As well as the house owner who would like to lease their house would use this research to set their rental price for the house.

3.   Data

3.1   Data Source

This report mainly used a Dataset 'Toronto Apartment Rental Price' by Raja CSP in Kaggle on 2018 for information of rental price and apartments, as well as a Python API 'Foursquare' for geography information, and data collected and cleaned by the researcher. In this project, factors including room numbers, location and nearly facilities would be analyzed, compared and segmented. Correlation between rental prices and factors would be calculated. At the end of the project, a quantitative model would be built to provide further insight and forecasted the rental price.

From Foursquare API (http://developer.foursquare.com/api), researcher retrieved the following for each venue:

- Name: The name of the venue

- Location: The latitude and longitude of the venue

- Category: The category type of the venue

From 'Toronto Apartment Rental Price' data set (https://www.kaggle.com/rajacsp/toronto-apartment-price), research retrieved the following for each house:

- Address: Full address of rental apartment

- Bedroom number: The number bedroom/bedrooms the apartment has

- Bathroom number: The number bathroom/bathrooms the apartment has

- Location: The latitude and longitude of the apartment

- Rental Price: The rental price for the apartment per month

4.   Methodology and Analysis

In the beginning, data will be clean and prepare, the apartment data set will be loaded and narrow down to the downtown area of Toronto, also, outliner which having extreme higher and lowers rental price will be removed. Second, the location data from the Foursquare API for all venue up to a distance of 4 kilometers from downtown Toronto will be extracted, using this, the venue information including location, name and category will be fetched.

After cleaning and preparing the data, the following steps will be performed. Firstly, the researcher will apply some basic exploratory analysis to both data set. For the location of each apartment and borough on the map. And visually inspect some values in the data with charts. Thereafter, reduce the number of the data set and transform into data frame format. Finally, build the model by cluster analysis to find the desired result.