Entertainment Venues & House Sales Prices
Data Analysis of Toronto Neighborhood

Applied Data Science Capstone Project

Peiwen Zhong

1. Introduction

1.1 Background

With the improvement in the economy and our life quality, when choosing to buy a property, people are paying more and more attention to the entertainment amenities in the community, especially young adult buyers. They would like to have easy access to different kinds of entertainment and cultural activities. And also, the house price in that neighborhood is affordable, too. For example, they may want to be close to places like museums, art galleries, and movie theaters, where they can enjoy the nightlife or have fun during the weekend. They believe a community near these amenities is more convenient and people living there usually younger and more energetic.

1.2 Problem

The problem I am going to solve is, to help young property buyers find out which communities in Toronto have more kinds of entertainment venues and its average house price is reasonable. I believe this could be a community reference for them and help them narrow down the targets when they are doing property search.

1.3 Interest

Young adult buyers would be interested to know where they can purchase a property that they can afford, and also have different kinds of entertainment amenities in the community. People who are helping these young buyers for property search, like property agents, will be interested to this, too.

2. Data

2.1 Data sources

In order to solve this problem, the data we need is:

- List of Toronto Neighborhoods and its postal code, latitude and the longitude coordinates.
- Entertainment Venues of each neighborhood within 500 meters
- Average house sales price of each neighborhood

Based on the data requirement, we can obtain the data by:

- obtain postal codes of Toronto Neighborhoods on the wiki page “List of postal code of Canada”.
- use the Geocoder Python package to get the latitude and the longitude coordinates of each neighborhood from the postal code.
- use Foursquare location data API to search for specific categories related to entertainment amenities for each neighborhood, within 1000 meters.
- obtain sales price from Toronto Real Estate and calculate the average house sales price of each Toronto neighborhood. As there are not average sale price data available for each Toronto neighborhood, we will have to calculate the average price by ourselves and store it in a csv file.

2.2 Data cleaning

The postal codes of Toronto Neighborhoods on the wiki page have some columns with “Not assigned”, we have to clean these data. We only process the cells that have an assigned borough and ignore cells with a borough that is “Not assigned”. If a cell has a borough but a “Not assigned” neighborhood, then we assign the neighborhood value as the same as the borough.

As more than one neighborhood can exist in one postal code area, there are multiple neighborhoods with same postal code value in the wiki page data. We combine them into one row with the neighborhoods separated with a comma.
We obtain current sale house of each neighborhood from Toronto Real Estate and calculate their average price based on their postal code. We only process records with correct format of price and postal code.

2.3 Entertainment selection

We use Foursquare location data API to search for specific categories related to entertainment amenities for each neighborhood, within 1000 meters. According to the category list of Foursquare API, we choose the following categories: Amphitheater, Aquarium, Arcade, Art Gallery, Concert Hall, Movie Theater, Museum, Public Art, Stadium, Bar, Lounge, Nightclub.