# 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.

# 3. Methodology

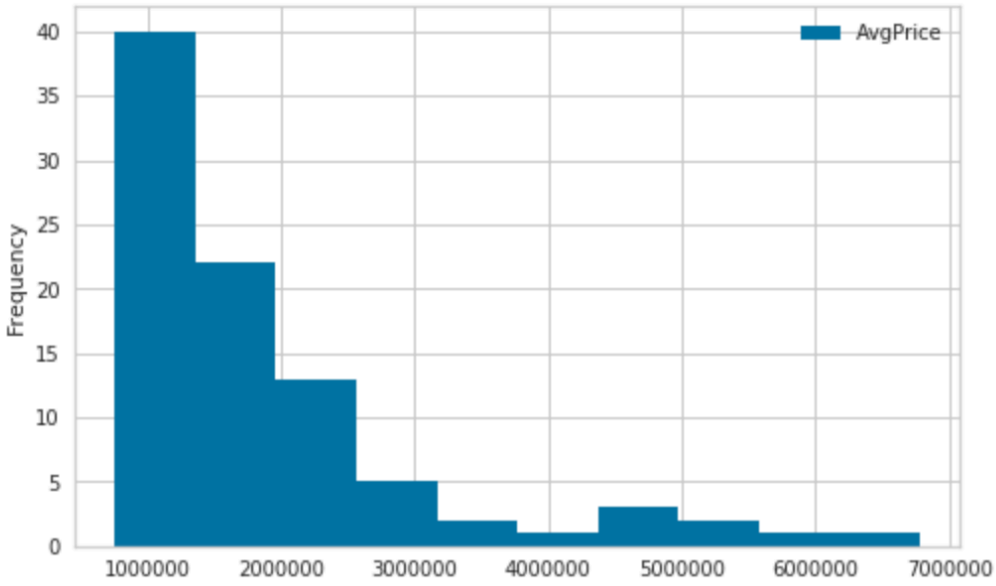
3.1 Main data

After data collection and cleaning, we get 103 neighborhood data of Toronto. Then we use Geocoder Python package to get the latitude and the longitude coordinates of each neighborhood from the postal code. By merging coordinates and average house price with the data we get from the wiki page, the main data now includes postal code, borough name, neighborhood name, latitude, longitude, average price and current available house number.

| | Postcode | Borough | Neighborhood | Latitude | Longitude | AvgPrice | count |
|---|---|---|---|---|---|---|---|
| 0 | M3A | North York | Parkwoods | 43.753259 | -79.329656 | 1.635100e+06 | 30 |
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 | 1.238611e+06 | 9 |
| 2 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 1.897599e+06 | 13 |
| 3 | M6A | North York | Lawrence Heights,Lawrence Manor | 43.718518 | -79.464763 | 1.514723e+06 | 22 |
| 4 | M9A | Etobicoke | Islington Avenue | 43.667856 | -79.532242 | 2.972532e+06 | 56 |

## 3.2 Average house price label

As part of the problem we want to solve is which neighborhood house price is more affordable, we can define the range by its histogram:




As it shows in above histogram, we can define the ranges by average house price as below:

- $1,000,000-$1,500,000: “Low Level HSP”
- $1,500,000–$2,500,000: “Mid-1 Level HSP”
- $2,500,000–$4,000,000: “Mid-2 Level HSP”
- $4,000,000–$5,500,000: “High-1 Level HSP”
- $5,500,000 and up: “High-2 Level HSP”

We then add this house price label to the main data:

| | Postcode | Borough | Neighborhood | Latitude | Longitude | AvgPrice | count | Price Labels |
|---|---|---|---|---|---|---|---|---|
| 0 | M3A | North York | Parkwoods | 43.753259 | -79.329656 | 1.635100e+06 | 30 | Mid-1 Level HSP |
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 | 1.238611e+06 | 9 | Low Level HSP |
| 2 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 1.897599e+06 | 13 | Mid-1 Level HSP |
| 3 | M6A | North York | Lawrence Heights,Lawrence Manor | 43.718518 | -79.464763 | 1.514723e+06 | 22 | Mid-1 Level HSP |
| 4 | M9A | Etobicoke | Islington Avenue | 43.667856 | -79.532242 | 2.972532e+06 | 56 | Mid-2 Level HSP |

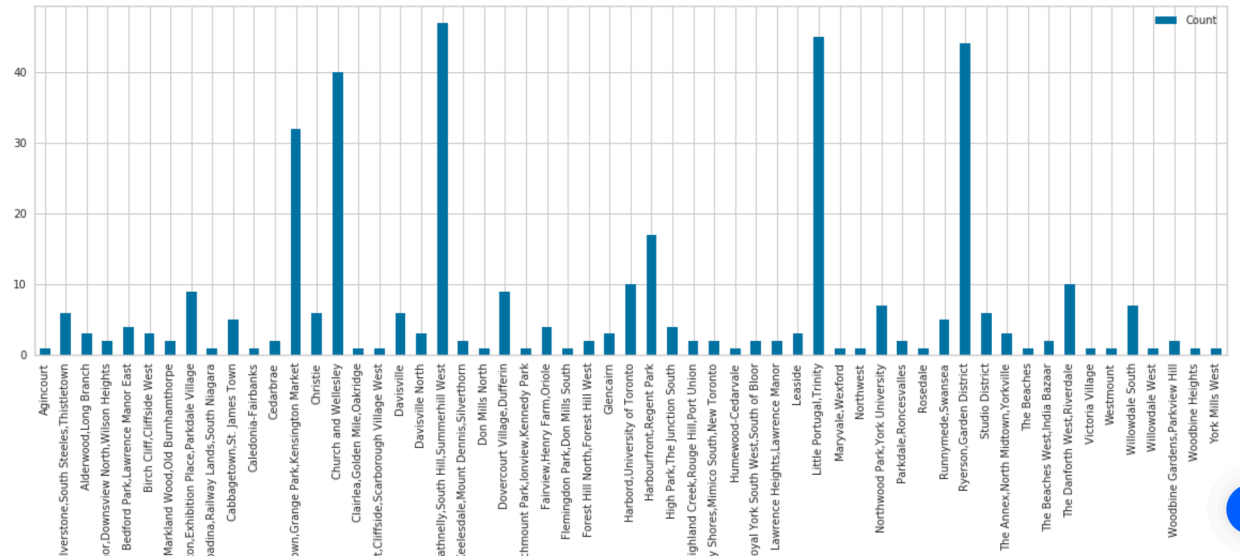
## 3.3 Entertainment Venues

I used the Foursquare API to explore the neighborhood and get its entertainment venues with 500-meter radius and 100 limits from their given latitude and longitude. 383 venues in total and 61 unique venues categories are returned from Foursquare API. Here is a head of the list adding venue name, venue category, venue latitude and longitude from Foursquare API.

| | Postcode | Borough | Neighborhood | Latitude | Longitude | AvgPrice | Venue Id | Venue Name | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M4A | North York | Victoria Village | 43.725882 | -79.315572 | 0 1.635100e+06 1 1.238611e+06 2 1.... | 4c633acb86b6be9a61268e34 | Victoria Village Arena | 43.723481 | -79.315635 | Hockey Arena |
| 1 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 0 1.635100e+06 1 1.238611e+06 2 1.... | 4ad8d551f964a5201f1521e3 | Vistek | 43.657046 | -79.359667 | Electronics Store |
| 2 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 0 1.635100e+06 1 1.238611e+06 2 1.... | 4ade8ea8f964a5205a7621e3 | Berkeley Church | 43.655123 | -79.365873 | Event Space |
| 3 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 0 1.635100e+06 1 1.238611e+06 2 1.... | 4af21e78f964a520fae521e3 | Arta Gallery | 43.650022 | -79.361222 | Art Gallery |
| 4 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 0 1.635100e+06 1 1.238611e+06 2 1.... | 4beb1543415e20a10cdde5bb | Enoch Turner Schoolhouse | 43.652873 | -79.361672 | History Museum |

We need to note that there are some neighborhood has no entertainment venues returned from Foursquare API. These neighborhoods are not included in later clustering.

We calculate total number of entertainment venues in each neighborhood. As shown below, we can see Downtown Toronto M5T, Downton Toronto M4Y, Central Toronto M4V, West Toronto M6J and Downtown Toronto M5B has over 30 entertainment venues. These neighborhoods have more choices of entertainment. West Toronto M6K, West Toronto M6H, Downtown Toronto M5S, Downtown Toronto M5A, East Tonton M4K has close to 10 entertainments, they can be alternate choices.




## 3.4 Clustering

From the venues returned from Foursquare API, I calculated 5 top entertainment venue categories in each neighborhood.

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Agincourt | Lounge | Social Club | Theater | Concert Hall | Speakeasy |
| 1 | Albion Gardens,Beaumond Heights,Humbergate,Jam... | Movie Theater | Indie Movie Theater | Nightclub | Concert Hall | Speakeasy |
| 2 | Alderwood,Long Branch | Pub | Bar | Social Club | Theater | Concert Hall |
| 3 | Bathurst Manor,Downsview North,Wilson Heights | Art Gallery | Bar | Social Club | Theater | Concert Hall |
| 4 | Bedford Park,Lawrence Manor East | Bar | Art Gallery | Pub | Social Club | Gastropub |

I used K-means algorithm to cluster the neighborhoods. I also use elbow method to experiment K-means cluster number and the optimal k is 5. So, I cluster neighborhoods to 5 cluster. Here is the data with cluster label.

| | Postcode | Borough | Neighborhood | Latitude | Longitude | AvgPrice | count | Price Labels | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 | 1.238611e+06 | 9 | Low Level HSP | 3 | Hockey Arena | Nightclub | Concert Hall | Speakeasy | Wine Bar |
| 2 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 1.897599e+06 | 13 | Mid-1 Level HSP | 0 | Art Gallery | Pub | Event Space | Café | Electronics Store |
| 3 | M6A | North York | Lawrence Heights,Lawrence Manor | 43.718518 | -79.464763 | 1.514723e+06 | 22 | Mid-1 Level HSP | 0 | Event Space | Sports Bar | Social Club | Theater | Concert Hall |
| 6 | M3B | North York | Don Mills North | 43.745906 | -79.352188 | 4.572644e+06 | 54 | High-1 Level HSP | 1 | Hookah Bar | Social Club | Theater | Concert Hall | Speakeasy |
| 7 | M4B | East York | Woodbine Gardens,Parkview Hill | 43.706397 | -79.309937 | 1.291514e+06 | 20 | Low Level HSP | 0 | Bar | Winery | Social Club | Theater | Concert Hall |

Based on the first most common venues in each cluster, we can define each cluster as:

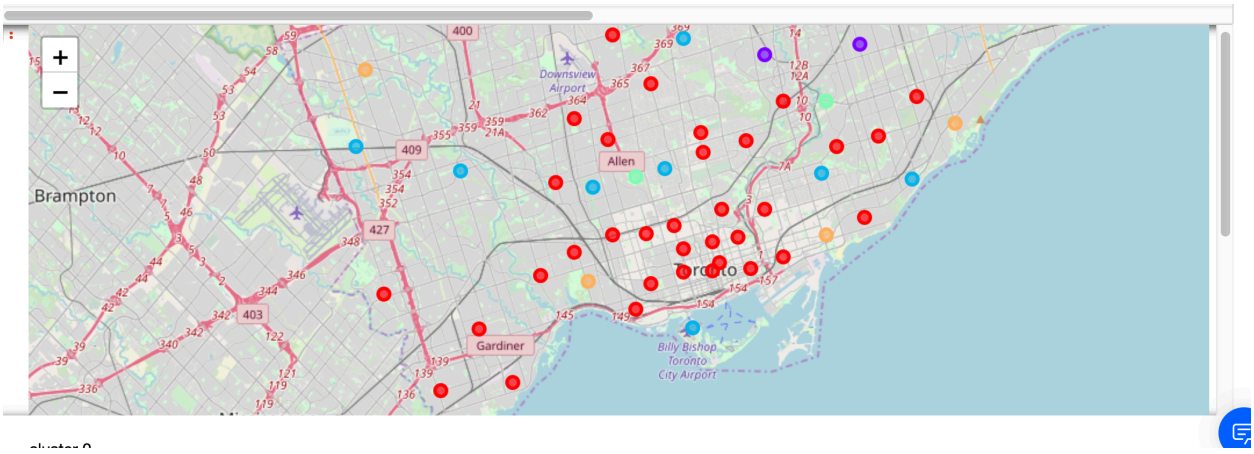
- Cluster 0: Multiple Entertainment Venues
- Cluster 1: Hookah Bar Venues
- Cluster 2: Bar Venues
- Cluster 3: Hockey Arena Venues
- Cluster 4: Movie Theater Venues

## 4. Results

The final data merged with Cluster Label and Price Label:

| | Postcode | Borough | Neighborhood | Latitude | Longitude | AvgPrice | count | Price Labels | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 | 1.238611e+06 | 9 | Low Level HSP | 3 | Hockey Arena | Nightclub | Concert Hall | Speakeasy | Wine Bar |
| 2 | M5A | Downtown Toronto | Harbourfront,Regent Park | 43.654260 | -79.360636 | 1.897599e+06 | 13 | Mid-1 Level HSP | 0 | Art Gallery | Pub | Event Space | Café | Electronics Store |
| 3 | M6A | North York | Lawrence Heights,Lawrence Manor | 43.718518 | -79.464763 | 1.514723e+06 | 22 | Mid-1 Level HSP | 0 | Event Space | Sports Bar | Social Club | Theater | Concert Hall |
| 6 | M3B | North York | Don Mills North | 43.745906 | -79.352188 | 4.572644e+06 | 54 | High-1 Level HSP | 1 | Hookah Bar | Social Club | Theater | Concert Hall | Speakeasy |
| 7 | M4B | East York | Woodbine Gardens,Parkview Hill | 43.706397 | -79.309937 | 1.291514e+06 | 20 | Low Level HSP | 0 | Bar | Winery | Social Club | Theater | Concert Hall |

The clustered map of Toronto Neighborhood:




As we want to know which neighborhoods have more entertainment venues, we will focus on Cluster 0: “Multiple Entertainment Venues” (red markers). We could

see most “Multiple Entertainment Venues” neighborhoods are near downtown of Toronto. The central of the city is usually crowded so it is not surprised that there are different kinds of entertainments there. There are other neighborhoods located near highway or railway. When the traffic is convenient, it will bring more people so there are more entertainments.

Also, as of the house price, from the price label, there are 21 neighborhoods in Mid-1 Level HSP, 15 in Low Level HSP, 2 in Mid-2 Level HSP, 2 in High-2 Level HSP and 1 in High-1 Level HSP.

| | Postcode | Borough | Neighborhood | Latitude | Longitude | AvgPrice | count | Price Labels | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **76** | M4V | Central Toronto | Deer Park,Forest Hill SE,Rathnelly,South Hill,... | 43.653226 | -79.383184 | 5.307793e+06 | 31 | High-1 Level HSP | 0 | Bar | Gastropub | Pub | Art Gallery |
| **12** | M3C | North York | Flemingdon Park,Don Mills South | 43.725900 | -79.340923 | 5.667300e+06 | 10 | High-2 Level HSP | 0 | Concert Hall | Social Club | Nightclub | Speakeasy |
| **81** | M4W | Downtown Toronto | Rosedale | 43.679563 | -79.377529 | 6.782176e+06 | 17 | High-2 Level HSP | 0 | Art Gallery | Social Club | Theater | Concert Hall |
| **36** | M1L | Scarborough | Clairlea,Golden Mile,Oakridge | 43.711112 | -79.284577 | 9.137000e+05 | 41 | Low Level HSP | 0 | Soccer Field | Social Club | Theater | Concert Hall |
| **68** | M1S | Scarborough | Agincourt | 43.794200 | -79.262029 | 1.205497e+06 | 36 | Low Level HSP | 0 | Lounge | Social Club | Theater | Concert Hall |

The most expensive neighborhoods (High-1, High-2 Level) in cluster 0 are in central Toronto M4V, North York M3C and Downtown Toronto M4W. Mid-2 Level has only one neighborhood: Central Toronto M5R, Downtown Toronto M4Y. Especially, Central Toronto M4V and Downtown Toronto M4Y have over 40 entertainment venues in total. These neighborhoods are not likely affordable for young buyers.

We could see most neighborhoods are in mid-1 Level. They are most in Downtown Toronto, Central Toronto, North York. A few are also in West Toronto, Etobicoke. With enough budget, young buyers can consider those in mid-1 Level and closer to downtown Toronto.

There are neighborhoods in Low Level, and they are most in Scarborough and Etobicoke. North York, West York, East Toronto and York all have one. Young

buyers can consider these with limit budget. Especially those in Scarborough and Etobicoke (Humber Bay Shores, Mimico South, New Toronto), as they are close to highway or railway.

## 5. Discussion

There are some neighborhood has no entertainment venues returned from Foursquare API. These neighborhoods are not included in the clustering. It is because there are no entertainment venues within 500 meters radius. Young Buyers who cares about entertainment venues should avoid these. Or we can increase the radius to include more neighborhoods.

There is no average house price for each neighborhood available, so we have to calculate it using current sale houses on the market. It may not be very accurate. Using sold data is a better choice but it is a bit difficult to get directly from Toronto Real Estate. We can use it when it is available to public.

## 6. Conclusion

Purpose of this project is to help young property buyers find out which communities in Toronto are close to entertainment venues and its average house price is reasonable. We collect neighborhood data, average house price data and entertainment venues data from different data sources and using these data to cluster neighborhoods to five clusters. Based on their 1$^{st}$ common entertainment venue category, we define the five clusters. “Multiple Entertainment Venues” neighborhood with low level or mid-1 level price is a good choice for the young buyers.

With above analysis, young buyers can purchase a property within their budget while considering entertainment venues in the community. As house data and

venues will change each year, similar methods can be used with new data for future analysis.