

Report Toronto Airbnb and Nearby Venues

1. Introduction

Airbnb is an online marketplace where millions of hosts and travellers can create a free account so they can list their space or book accommodations anywhere in the world. It covers more than 100,000 cities and 220 countries worldwide.

Airbnb has been successful because it has managed to replace hotels as a traditional method of accommodation for travellers worldwide.

Being one of the most popular tourist destination in Canada, Toronto's AirBnb market has been growing with an increased number of hosts joining and listing their homes for rent.

For this Capstone Project the business need is to provide the travellers with relevant information about Airbnb offers and venues in Toronto.

The listing price and amenities are an important factor in choosing a rental, but location is also important. Using FourSquare API venues data to compare different neighbourhoods could be a good opportunity to choose, for example, an area with good restaurants or concert halls or shopping centres.

This project aims to cluster places to stay in Toronto top venues of interest nearby.

2. Data acquisition

2.a Data sources

There are three data sets used in this project.

The first one is the data source for Toronto's Airbnb is provided by website: [Inside Airbnb](#).

The data for Toronto's neighbourhoods is extracted from [Wikipedia](#) and it was used in the week3 assignment. This data set is going to be used in combination with the Geo-spatial data from [Week3](#) to find the postal code for each neighbourhood.

2.2 Data cleaning and selection

The Toronto's airbnb data set has **74 columns** and **18265 entries**. For the purpose of this project I used only the following columns in order to get insight about the price listings, which neighbourhood has the most amount of listings, neighbourhoods, etc.:

'neighbourhood_cleansed', 'latitude', 'longitude', 'room_type', 'price'

I selected only top three neighbourhoods with the most listings and used them to find venues of interest in the area, then I found the postal codes for each listing in order to find what borough corresponds to each listing.

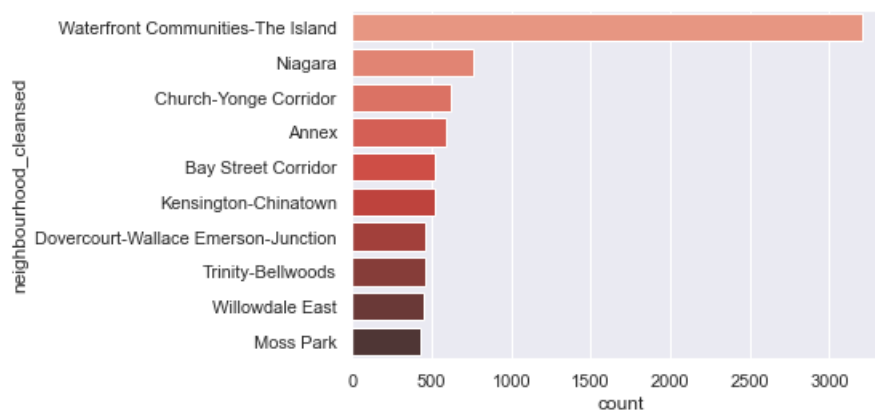
Dataset of Toronto is used to filter neighbourhoods and their locations in the selected borough.

The final dataset of listings includes neighborhoods, location of each property, room type and price.

Using the **FourSquare API** I retrieved information about each neighbourhood venue categories, such as: restaurants, shopping centres, museums, parks, recreation places, etc.

3. Methodology

I visualized the Toronto's airbnb price distribution for the top 10 neighbourhoods with the most listings



and calculate the average price:

Waterfront Communities-The Island	195.546729
Niagara	187.755263
Annex	177.899317
Moss Park	149.810624
Bay Street Corridor	143.007707
Trinity-Bellwoods	128.799564
Kensington-Chinatown	128.550388

Church-Yonge Corridor	128.360709
Willowdale East	113.933036
Dovercourt-Wallace Emerson-Junction	100.848485

Because of the big number of the listings, I decided to use only the first three neighbourhoods that have over 500 listings each.

I used Google geocoder API to obtain the postal code for each listing, which were more than 4500 and the process was very slow.

I saved the results into a local file to use it for the further developing of the Jupiter notebook.

I kept only the first three characters of the postal codes and I appended them to the listings dataset.

Then, I merged this dataset with the Toronto's neighbourhoods (used in week 3 's assignment) dataset, on the postal code column, to get the borough for each listing.

All the airbnb's listings from the top three neighbourhoods were in two boroughs:

Downtown Toronto and West Toronto.

I applied Foursquare API to find latitude and longitude of **Downtown Toronto** borough. Latitude and longitude of Downtown Toronto is a key factor of this project. Obtaining the location of Downtown Toronto helps carry out map of the borough along with its neighbourhoods. Interactive maps in the project are rendered using Folium library.

Next, calling requests obtain venues data within the neighbourhoods.

Merging the neighbourhoods data, listings data and venues data results in a complete data including neighbourhoods and their locations; listings and their locations; venues, venue categories and their locations.

The next step is to apply one-hot encoding to get dummies values associated with each venue category. Counting the occurrence of each venue category in each neighbourhood, then applies mean() method to get the frequency of occurrence of each venue category. Sorting the values of frequencies obtained in ascending order to get the most common venue category of each neighbourhood.

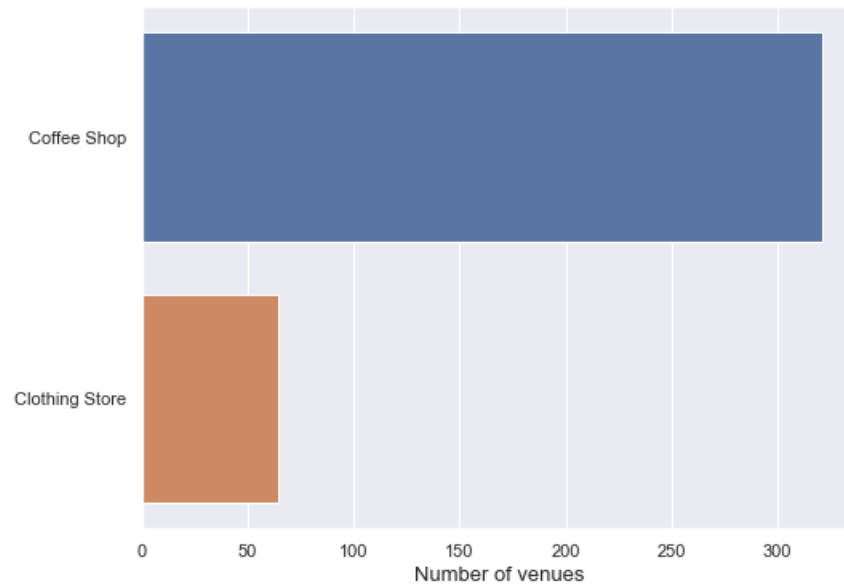
The final step is to cluster the venue categories by location of listings. In this step, k-means clustering method is applied as the number of clusters set to 5. The result obtained 5 clusters which labelled from 0 to 4.

Cluster 1 is picked since it meets the project's interest. Create an interactive map of all listings that are in cluster 1.

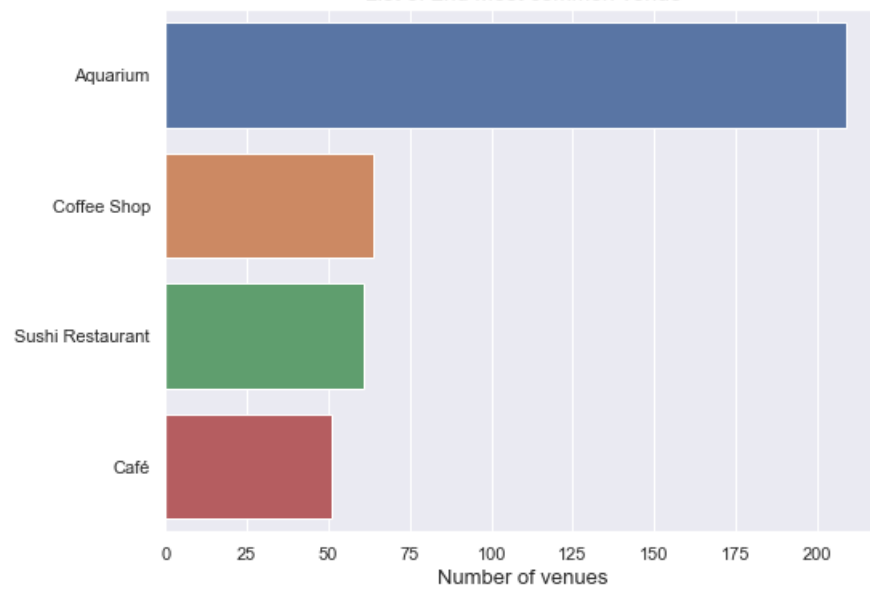
4. Results

There are five clusters of venue categories. Our goal is to find venue places of interest, so the cluster labelled 1 is picked. Cluster 1 results following number of venues categories in Downtown Toronto:

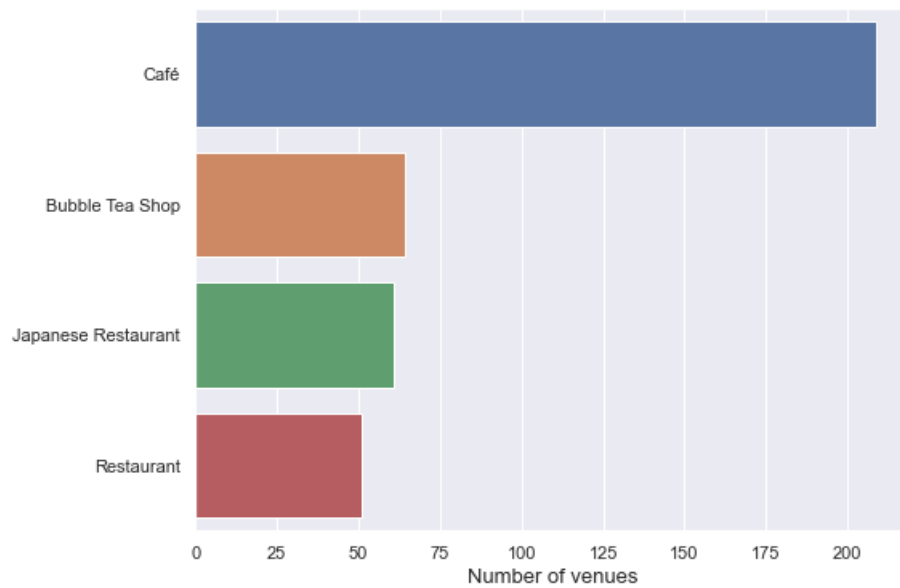
List of 1st Most common venue



List of 2nd Most common venue



List of 3rd Most common venue



The project reaches its goal that is to find places to stay in Downtown Toronto which is also near venue places of interest by clustering data.

5. Discussion

This project was a good opportunity to work with Foursquare API, Folium library, geocoders, and clustering method in machine learning.

The K-Means algorithm is an unsupervised learning technique and simple algorithm capable of clustering this kind of dataset very quickly and efficiently, often in just a few iterations.

Choosing different k-clusters effectively affects the result. The project approaches the k-means method to clustering as the end goal, but does not intentionally find the best k in terms of k-means.

6. Conclusion

In this project, I was able to find locations of places to rent in Downtown Toronto airbnb where are also near venue places of interest using Foursquare API, Folium, and machine learning approaches. I carried out interactive map for places to rent. I obtained lists of the most common food places in Toronto downtown.