Data Science Capstone Project

Finding Similar Neighborhoods between New York and Toronto

Table of contents

1. Introduction : Business Problem

Living in a big city provides a number of options in terms of neighborhood choices. When selecting a neighborhood to live in a person might have a number of considerations, including rental costs, housing prices, transportation , walkability, restaurants in neighborhood, Gyms, running routes etc.

If someone is relocating from one city to another while they might have a good idea of the neighborhoods in one city they might be unfamiliar with the neighborhoods in another city. In this case study we are going to look at neighborhoods in New York and try to find similar neighborhoods in Toronto as an example.

While the intention for this Capstone project is only to do this for two sample cities, this can easily be extended to add additional cities for comparison

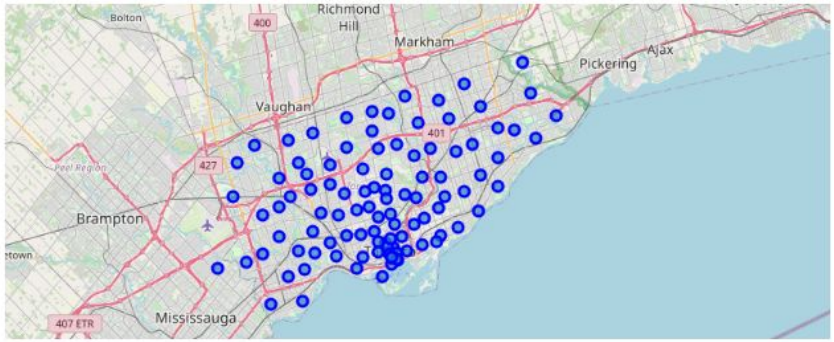
2. Data

To group neighborhoods into different categories and find comparable neighborhoods from the other city, we can use the data from foursquare API to find venues type and number of venues in each neighborhood.

We first gather neighbourhood data for both Toronto and and New York for along with zip code and latitude and longitude data. We then plot this on a map to get an idea of the spread. Below is the map of all the neighborhoods we'll be looking at for New york and Toronto.

New York Neighborhoods

Toronto Neighborhoods



3. Exploratory Data Analysis

Our analysis of Venue Types after combining neighborhoods from both cities provide the following distribution.

Venue Type	Number of Venues
Arts & Entertainment	446
College & University	16
Food	6389
Nightlife Spot	768
Outdoors & Recreation	1172
Professional & Other Places	100
Residence	10
Shop & Service	2717
Travel & Transport	471
Events	0

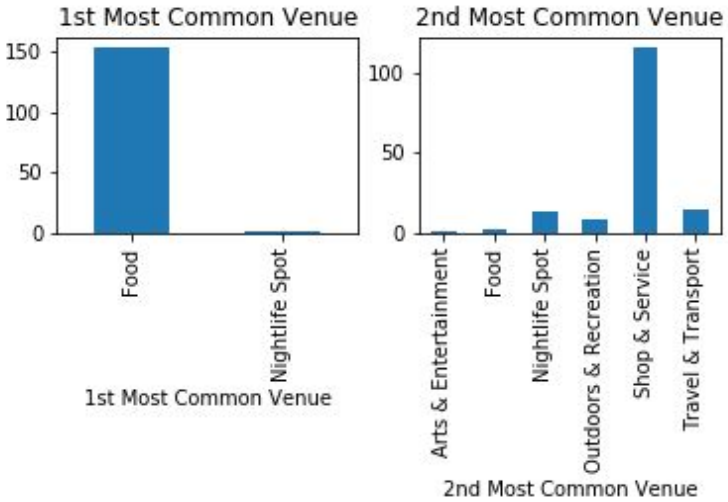
4. Methodology

We run K-means clustering on the neighbourhoods across the two cities and classify them into six different clusters. Providing us with the most similar neighbourhoods based on the make-up of the type of venues in each neighborhood. The six neighborhood clusters end up with the following distribution based on the venue type make up in these neighborhoods

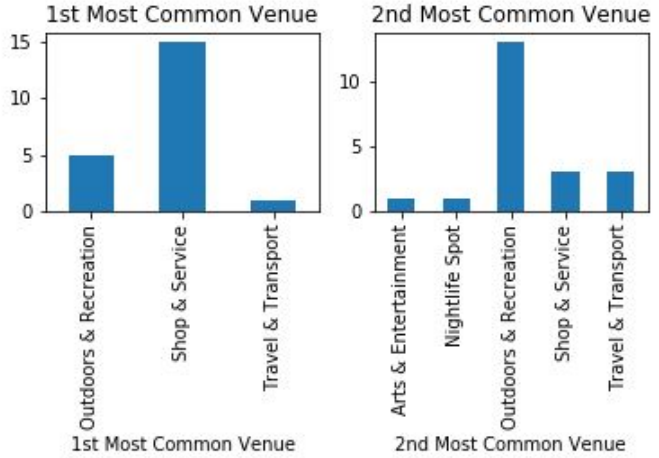
Cluster Label	Number of Neighborhoods
Cluster 1	156
Cluster 2	21
Cluster 3	14
Cluster 4	14
Cluster 5	51
Cluster 6	148

5. Analysis

Cluster1 :Total Number of Neighborhoods: 156. Based on the first and second most common venue types in these neighborhoods we can classify Cluster 1 as Shopping districts with good food options and some nightlife.



Cluster 2: Total Number of Neighborhoods: 21
We can classify Cluster 2 Neighborhoods ar primarily have lot of outdoor and recreation with some shopping & service.

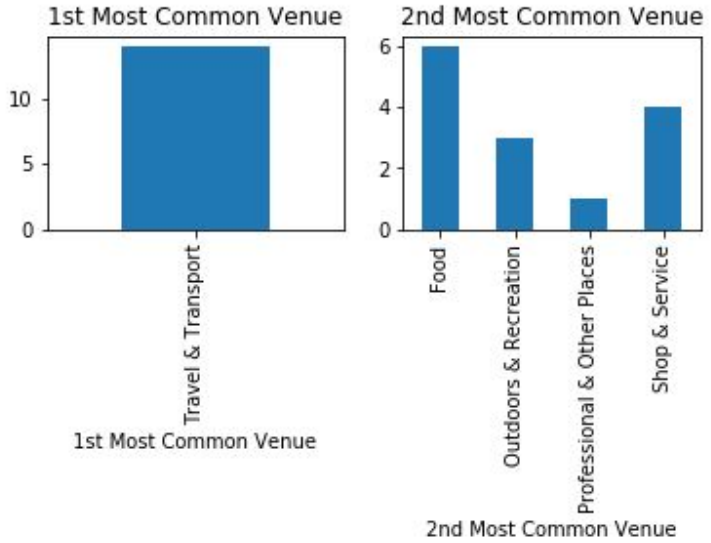


5. Analysis

Cluster 3

Total Number of Neighborhoods: 14

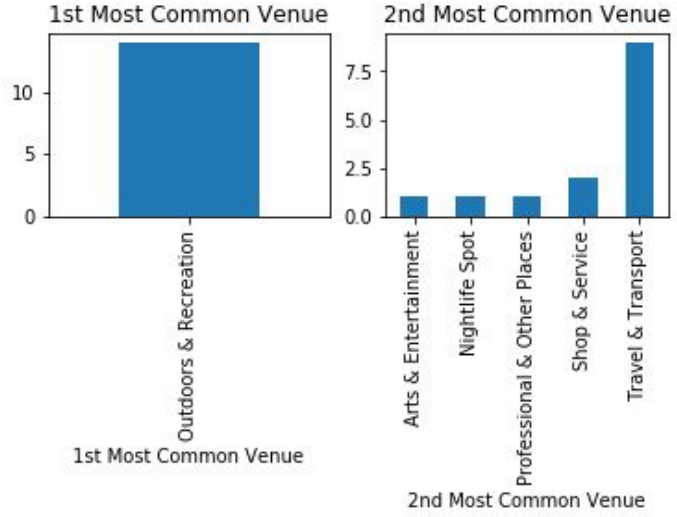
Cluster 3 Neighborhoods can be classified as Travel and Transportation Neighborhoods.



Cluster 4

Total Number of Neighborhoods: 14

Cluster 4 Neighborhoods can be classified as Outdoors & Recreation with travel and transport. This Mix would indicate more Isolated neighborhoods with parks.

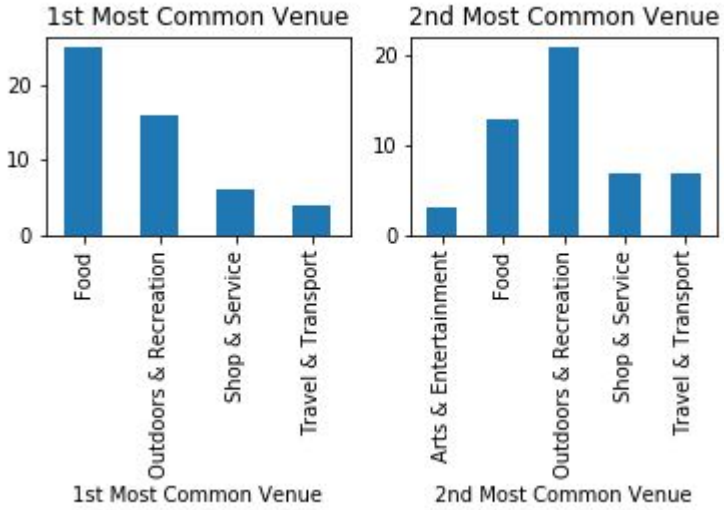


5. Analysis

Cluster 5

Total Number of Neighborhoods: 51

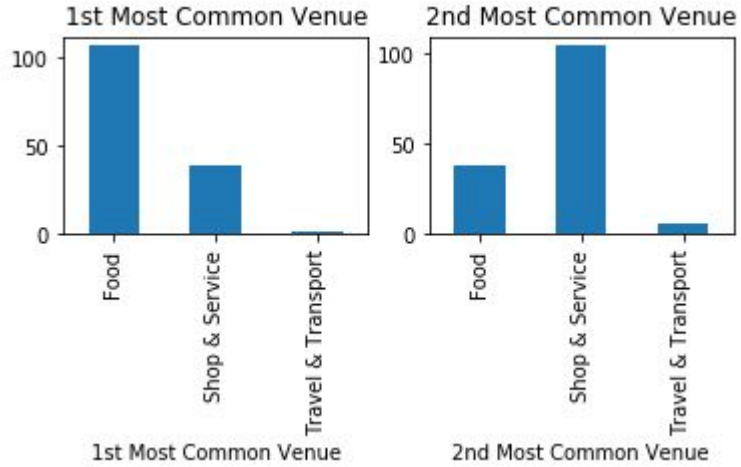
Cluster5 seem to be neighborhoods with a good mix of all venues with a number of outdoor and recreation areas.



Cluster 6

Total Number of Neighborhoods: 148

These Neighborhoods would seem to have great dining and shopping venues.



6. Results and Discussion

Based on our Clustering we have come up with this set of six neighborhoods.

Cluster Number	Neighborhood Types
Cluster 1	Primarily Shopping districts with good food options and some nightlife.
Cluster 2	Good Outdoor and recreation with some shopping & service.
Cluster 3	Travel and Transportation
Cluster 4	Outdoors & Recreation with travel and transport. This Mix would indicate more Isolated neighborhoods with parks.
Cluster 5	Good mix of all venues with a number of outdoor and recreation areas.
Cluster 6	Great for dining and shopping venues.

7.Conclusion

We used the number and type of venues to come up with similar neighborhoods between the cities of Toronto and New York. This allows us to get a good idea of which neighborhoods would have a similar feel in a new city. This can be further expanded in the future by adding rental rates, property prices walkscore etc.