

Toronto a NY venues

Introduction

The problem:

Determine which city is the best to open new stores and which one is the best to invest in already existed ones, which city is the best to invest on a coffee shop, bakery, etc. What are the most common venues and therefore the ones with the biggest amount of competition.

Target audience

The target audience are owners of small business and investors that want to support small new business such as coffee shops, bakeries, etc, in Toronto and New York. The location of a venue is important for its success, this project will provide understanding of the different locations of New York and Toronto. By understanding the locations of the cities, investors will be able to determine which locations are the best for the different venues and therefore investors can decide on support venues or where to open a second venue of an already supported store. By understanding the venues surrounding their business, owners of small businesses can determine which new products and services they can offer to bring more clients. A cafe surrounded by fitness studios may add low-carb options. A bakery in a business office area may add a variety of coffee to go.

Data

The data we will use in this project is:

- Name of the neighborhoods;
- geographical coordinates (latitude and longitude) of the neighborhood;
- 100 venues around the geographical coordinates of neighborhoods,

The data of the first item can be obtained by Wikipedia or other sources. The data of the second item can be obtained by the use of geocoder data or other sources. The data of the third item can be obtained by the use of Foursquare location data.

Methodology

We process and prepare data of the financial districts of the cities and their surroundings. Due to the amount of people visiting the financial district, we can study all these locations at the same time to understand the preferences and opportunities for each city. For the other locations, the study requires a different approach.

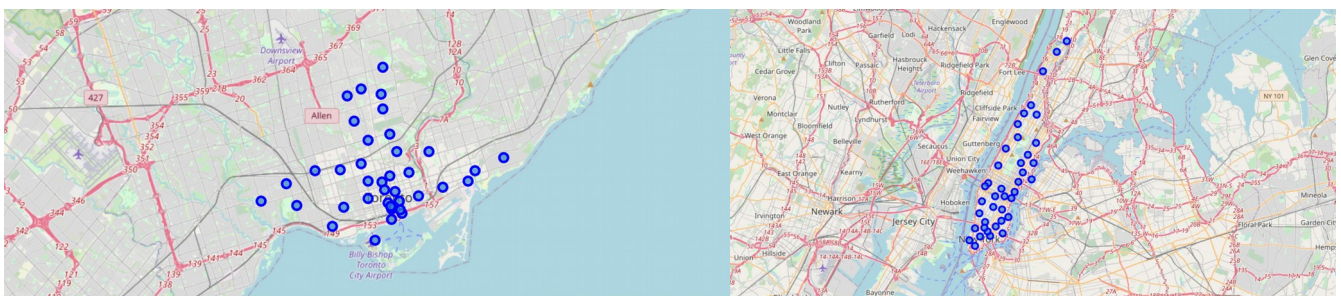
We picture our data on a map, to understand the locations and the mobility inside them. This will be fundamental at the moment we conclude opportunities for new venues. The distribution of the neighborhoods tells us which neighborhoods are closer to the financial district, therefore we can understand which neighborhoods have more volume of people during the day.

We get the top 100 venues per neighborhood in a 500 meters radius, a bounded distance for people looking for a coffee or bakery. From this information we can get the most common venues in the district, and the different kind of venues.

We get the top 5 most common venues per neighborhood, this will be useful to study those neighborhoods that are isolated points, i.e. has too few in common with the other neighborhoods. Those neighborhoods required their own analysis.

We cluster the neighborhoods based on the similarities of their venues. From these clusters we can infer how homogeneous are the cities. Due to the similarity of the neighborhoods in each cluster, we can study the neighborhoods by studying the clusters. This provides a better understanding of the neighborhoods and their interactions as elements of the same cluster.

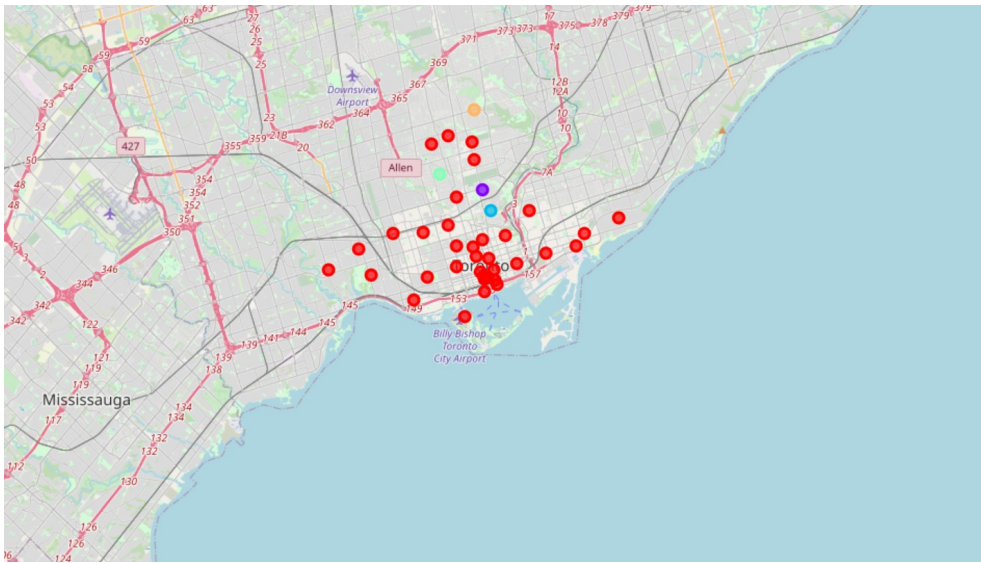
Results



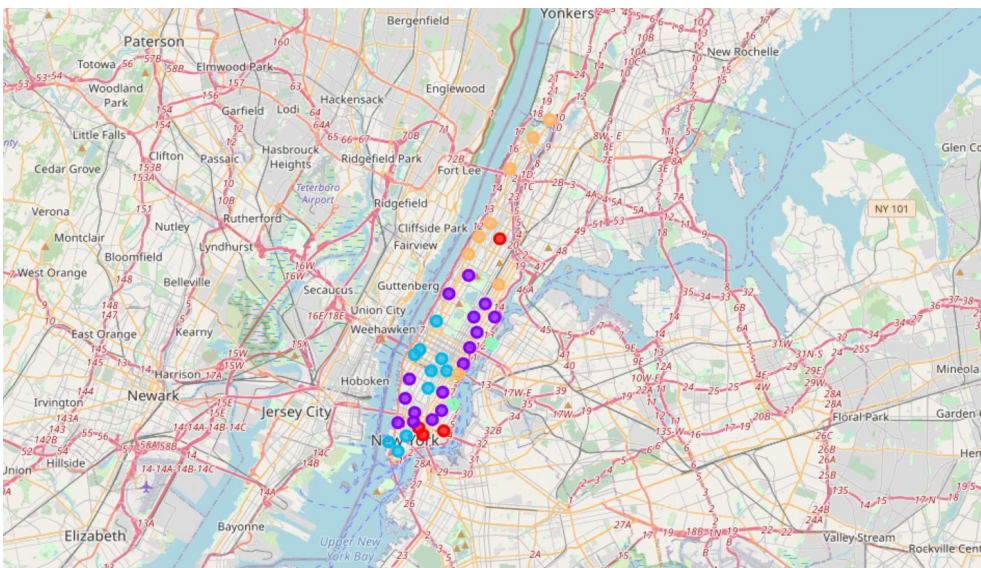
In the picture on the left, we can see the distribution of the neighborhoods around the financial district in Toronto. We only studied the neighborhoods with the word "Toronto" in the borough. On the picture on the right we can see the distribution of the neighborhoods around the financial district in Toronto. We only studied the borough "Manhattan".

As we can see in the images that the distribution of the neighborhoods in Toronto has not a well defined shape. Meanwhile the distribution of the neighborhoods in New York has a well defined shape limited by water.

There are 232 different kind of venues in the neighborhoods of Toronto and a total of 321 in New York.



Toronto clusters: In the picture above, we can see the distribution of the five clusters in Toronto, each color represents a cluster. As a result of clustering the neighborhoods, we can study all the neighborhoods in red as a big uniform neighborhood. At the same time, the other clusters are very small containing at most one neighborhood. Therefore the neighborhoods that are not in the red cluster requires a study as an isolation neighborhood. On the other hand, due to the size of the red cluster, we can obtain a big picture of the venues in Toronto by study only the red cluster.



Manhattan (New York) clusters: In the picture above, we can see the distribution of the five clusters in Manhattan, each color represents a cluster. As a result we see that there are no isolated neighborhoods in the sense of similarities, but there are neighborhoods which are isolated in the sense of distance to the other elements of the cluster. Due to this particularities, compare to Toronto, Manhattan is not homogeneous and therefore we can not study Manhattan by only study one of the clusters. Even more, decision on the red cluster need a more deep study depending on the neighborhood (there is a red neighborhood very far from the other red neighborhoods).

Discussion

The first observation is that Toronto is a very homogeneous city with one cluster covering almost the complete city. In this cluster the most common venues are related to gastronomy, and a more detail analysis is required to determine what are the best investment option in this cluster. The second observation is that Manhattan is very heterogeneous, this makes more difficult to start a new venue, but it makes easier to determine new products or services to offer (base on the fact that different kind of venues implies people commuting between stores). After determined what is the best investment strategy for each city, we can finish this section by giving an analysis of the small clusters of Toronto.

Cluster 1

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
Moore Park, Summerhill East	1	Playground	Tennis Court	Women's Store	Dance Studio	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run	----Moore Park, Summerhill East---- venue freq 0 Playground 0.5 1 Tennis Court 0.5 2 Yoga Studio 0.0 3 Movie Theater 0.0 4 Market 0.0

As we can see in this cluster we only have playgrounds and Tennis Courts. Therefore, this neighborhood is about physical activities and open a store about sport equipment is a good idea.

Cluster 2

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
Rosedale	2	Park	Playground	Trail	Dance Studio	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run	----Rosedale---- venue freq 0 Park 0.50 1 Playground 0.25 2 Trail 0.25 3 Yoga Studio 0.00 4 Movie Theater 0.00

As we can see in this cluster we only have playgrounds and parks. Therefore it is a family neighborhood with a kids friendly point of view, open a toy store would be a good idea.

Cluster 3

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
Forest Hill North & West	3	Jewelry Store	Trail	Mexican Restaurant	Sushi Restaurant	Women's Store	Department Store	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Donut Shop	----Forest Hill North & West---- venue freq 0 Jewelry Store 0.25 1 Trail 0.25 2 Mexican Restaurant 0.25 3 Sushi Restaurant 0.25 4 Yoga Studio 0.00

This cluster jewelry plus restaurants cluster, therefore a wine store fits the needs of the cluster.

Cluster 4

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	
Lawrence Park	4	Park	Swim School	Bus Line	Falafel Restaurant	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Donut Shop	Doner Restaurant	Dog Run	----Lawrence Park---- venue freq 0 Park 0.33 1 Bus Line 0.33 2 Swim School 0.33 3 Yoga Studio 0.00 4 Moroccan Restaurant 0.00

As we can see in this cluster we only have parks and swimming pools. Therefore, this neighborhood is about physical activities and open a store about sport equipment is a good idea.

Conclusion

Toronto is a more homogeneous city in terms of venues, and with 232 different kind of venues it is a more attractive city to investors in new kind of venues compare to New York, which is more heterogeneous and with 321 different kind of venues the opportunities for new venues are smaller. Generally speaking new kind of venues have more possibilities of success in Toronto than in Manhattan. On the other hand due to the homogeneity of the city, the location of a venue is not an important factor at the moment of deciding which new service a bakery or cafe could offer in Toronto. Clearly Mahattan is on the other side of the spectrum, it would be harder to start a new venue in Manhattan, but the data tells us that people commute more in Manhattan due to the heterogeneity of the area and therefore it is more convenient to invest on small stores to offer new services than to invest in new venues. To decide which new service to offer and where in Manhattan or what kind of new venue to open in Toronto, a more detail analysis is required, depending on the ambitions of the investment. Nevertheless as it was discussed above, the analysis is simpler for some clusters in Toronto.