

Background

To start up a business, conducting research on how local businesses operate is essential for making critical decisions for a business model. It not only prevents potential business risks but also allows business owners to adjust their plans based on the current market quickly. In this capstone project, two cities will be explored. The downtown Toronto and downtown New York (Manhattan).

Problems

What is the top 5 and the least favorite 5 of the business categories and their venues? How similar and different they are? What is the best K value when KMeans applied? Provide general recommendations for a start-up business owner after analysis.

Data

The data will be obtained from Foursquare API, extracting the data that has the radius of 2000 meters from the geographical coordination of downtown Toronto and downtown New York (Manhattan). Analyzing the obtained data as a free account user from Foursquare API based on questions from Background.

Methodology

Applying geocoder to get the geographical location of Toronto and New York first. Those two cities are the target that we are interested in.

Connecting the Foursquare API to get the value of venue, venue latitude, venue longitude and venue category within 2000 meters, from the geographical locations of two cities we obtained from the previous step. In reality, the area of downtown Toronto and New York is not regular shape. Therefore, we only consider portions of downtown areas in those cities.

Interestingly, the unique venue categories in downtown Toronto and New York are very close. Downtown Toronto has 61 unique venue categories whereas downtown New York is 62.

The top 5 venue categories in those two cities are shown below:

Downtown Toronto	
1	Coffee shop
2	Cafe
3	Park
4	Gym
5	Hotel

Table 1.0

Downtown New York	
1	Asian Restaurant
2	Hotel
3	Pizza Place
4	Bakery
5	Bookstore

Table 1.1

Toronto downtown top 5 venue categories distribution map

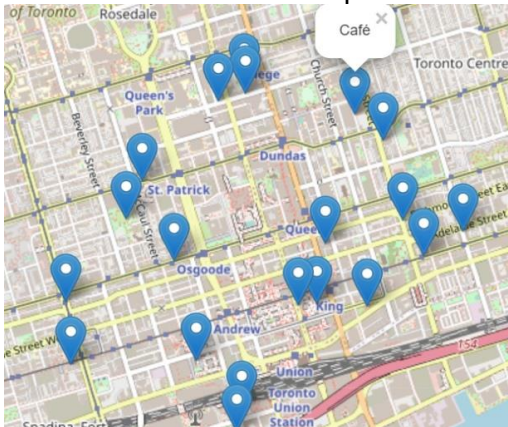


Fig1.0

New York downtown (Manhattan) top 5 venue categories distribution map

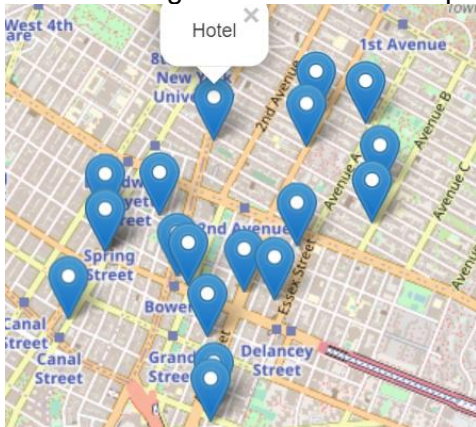


Fig1.1

The least favorite 5 venue categories in those two cities are shown below:

Downtown Toronto	
1	American Restaurant
2	Arts & Crafts Store
3	Baseball Stadium
4	Basketball Stadium
5	Beer Bar

Table 2.0

Downtown New York	
1	Argentinian Restaurant
2	Austrian Restaurant
3	Bar
4	Boxing Gym
5	Breakfast spot

Table 2.1

Toronto downtown the least 5 venue categories distribution map

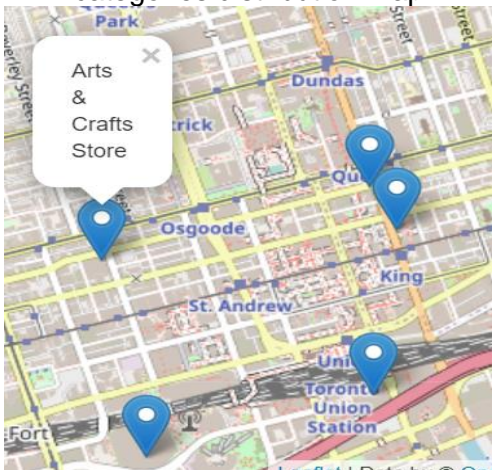


Fig 2.0

New York downtown (Manhattan) the least 5 venue categories distribution map

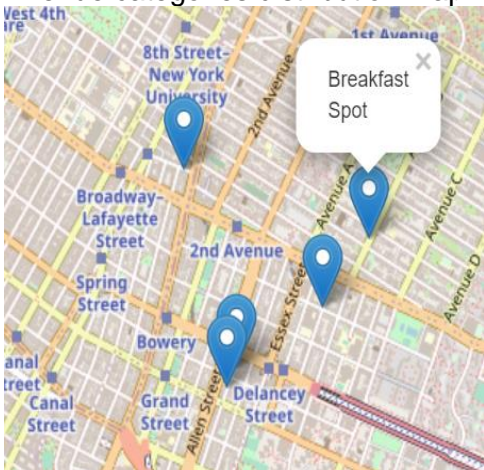


Fig2.1

Applying Silhouette Coefficient to find the best K value for each of two cities based on geographical coordination distribution. The optimal K value appears at the peak of the graph.

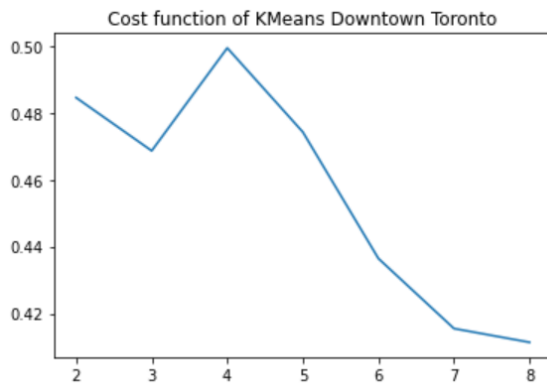


Fig 3.0

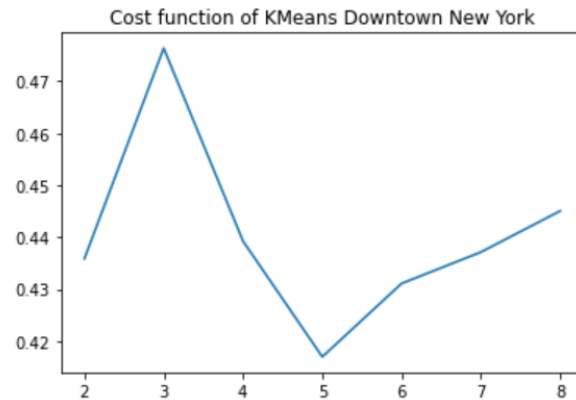


Fig 3.1

Applying KMeans with the same random state and its best K value for each city in the map.

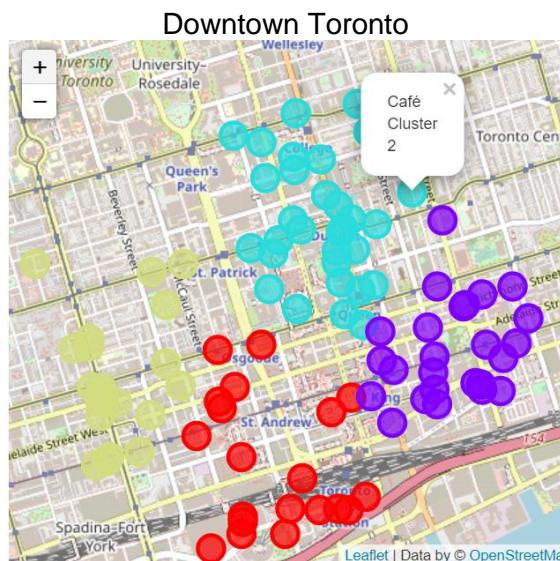


Fig 4.0

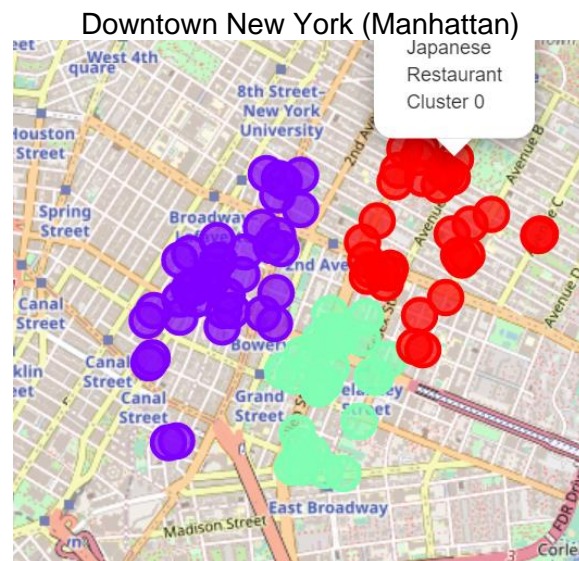


Fig 4.1

Results

From Table 1.0 and Table 1.1, Manhattan has more different varieties of business than that in Toronto.

Most of the popular business categories in Toronto in the table are food services (Coffee shop, Cafe, hotel) and retailer (Gym). Surprisingly, the park is the top 3 that local people would like to visit. On the other hand, Manhattan contains food services (Asian restaurant, hotel, Pizza Place, Bakery) and retailer (Bookstore).

Two cities have preference in food services. This implies that local people like various types of food services. As a newcomer to join this business category, it would be hard to compete with other local business owners. However, it may cost less to start a business in this category since there are so many competitors here, the business model is fully developed.

To have a better position for a newcomer, it is better to choose less common business types in food services, such as coffee shops, pizza shops, restaurants etc. Otherwise, the newcomer has exclusive ingredients that local food service does not offer, or the special strategies of business operation to exceed local food business services to earn a seat.

From table 2.0 and table 2.1, it shows the differences of five least favorite business types in Toronto and Manhattan.

In Toronto, it contains three business categories. They are food services (American restaurant, Beer Bar), retailers (Arts & Crafts) and investors (Basketball Stadium, Baseball Stadium). However, it only has two business categories in Manhattan. Food services (Argentinian Restaurant, Austrian Restaurant, Bar, Breakfast spot) and investor (Boxing gym).

From the distribution types of business categories, both cities have less favor in certain food services types and investors. Food services, there are foods that are not so popular for local people in both cities. However, it could be an opportunity for start-up business owners to improve their recipe but retain the uniqueness of food culture itself. Therefore, they can differentiate themselves with unique selling points to compete with local business. For investors, it mainly depends on the budget and risk assessment of the business that is evaluated by the business owner. It is not recommended for new start-up business owners to launch a big project even if budget is not an issue. The reason for that is the investment contains many uncertainties if the business owner is not familiar with relevant industry.

Interestingly, the clustering made by KMeans in both cities' venue points. They look similar as the possible number of clustering shown on Fig 2.0 and Fig 2.1. This may be due to coincidence. The number of the best or the worst cases are interesting.

Discussion

If we look closer to the Fig 4.0 and Figure 4.1, two cities downtown KMean distribution map, each cluster contains food services, drinks, coffee shops, or restaurants. For a high density of clusters, it is easy to see different food services types by crossing a few blocks in the street. This reflects those area has developed and became a specific commercial street.

For newcomers to start their business, the cost will be extremely expensive to operate business since those streets are the major stream of people. It is recommended to provide fast-moving consumer goods if start-up owners decide to do their business in those areas. This allows the business owners to have a better chance to get more customers. Alternatively, newcomers can provide services on healthy industry, such as stress release counselling, yoga, or spa. The reason for this recommendation is that people who work in downtown areas of those two cities have a fast-paced working environment everyday. As a human being, we not only need the nutrition from the food for our body needs, but also we need to alleviate the tension of our minds. Partnership with local businesses would be another choice for newcomers. Developing business relationships and creating an industry chain for the needs of local people.

Conclusion

Those suggestions just provide brief ideas about possible ways to launch businesses in the downtown area of Toronto and New York. Bias may apply since the data sample is not big enough.

Additionally, there are some limitations to access specific data types. For an example, the operation of a restaurant (tips, operating hours, etc.) to have further investigation. However, for learning purposes on this project, this project achieved the application of KMeans and data visualization on a map.

Reference

1. <https://www.expowest.com/en/register/business-categories.html>
2. <https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c>
3. <https://developer.foursquare.com/>
4. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>