Similarity of the 10 most multicultural cities

Bo Jiang

10 June 2020

1. Introduction

1.1 Background

With the rapid development of transportation technology, travelling around the world became an easier thing for most of the people. Successive waves of immigrants have created diverse cultural environments full of incredible restaurants, festivals and art scenes. ‘Some cities manage to fit the entire world within them’, said Cultural Trip.

1.2 Problem

Sometimes, we might be wondering, of all those large multicultural cities, which factor plays the most important role during the formation of the cities. And of all those cities, could one city be similar with another or all cities are different?

1.3 Interest

Tourists would be interested in this study and can refer to this article to decide their next destination. The municipal administration department of those cities might also be interested as they’ll learn from the successfully planning of other cities.

2. Data

There’re many large multicultural cities in the world, to make things easier, I got the 10 most multicultural cities from an article found in Cultural Trip website.

City Name	Population	Multicultural Symbols
Toronto	2.7 m	Half the population born outside Canada, with 200 ethnic groups, over 140 languages
Sydney	5.3 m	Approximately 250 languages
Singapore	5.7 m	Chinese, Malay, Indian and other ethnic groups, 4 official languages
Sao Paulo	12m	Immigrants from Africa, Italy, Lebanon
San Francisco	9.6m	Immigrants from China, Germany, Italy, Mexico and Indian.
New York City	8.2m	More languages spoken in the Queens than anywhere else in the world
Paris	12.5m	14% to 20% are immigrants
Log Angeles	18m	People from 140 countries and speaking 86 languages
London	8.9m	1/3 Londoner are foreign-born, over 200 languages are spoken
Amsterdam	8.1m	178 different cultural backgrounds

Data Source: Cultural Trip, Wikipedia

Following tools or libraries will be used to extracted/generate the required information:

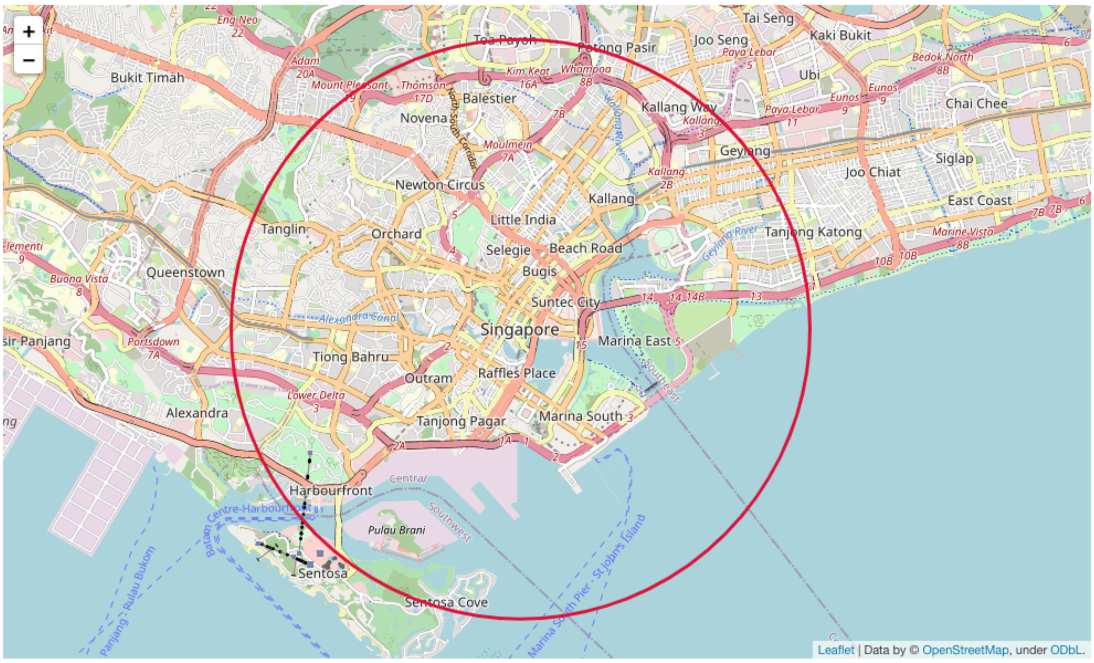
- **Python Geocoder Library** will be used to obtain the coordinates of the cities, and use as the Centre Point of the cities
- The nearby venues around the center area will be obtained using **Foursquare API**.
- **Folium** will be used to visualize the areas that we're going to study

3. Methodology

I'll compare the 10 most common venues of the 10 cities, and use K-means Clustering to group the similar cities.

3.1 Set the target areas

As all these cities are huge cities, I will only analyze the center areas, which is within 5km from the Center Point of the cities.



3.2 Get all the nearby venues

Using Foursqure API, I can get the nearby venues within 5km of the Centre Point.

	Name	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Toronto, Canada	43.64869	-79.38544	Byblos Toronto	43.647615	-79.388381	Mediterranean Restaurant
1	Toronto, Canada	43.64869	-79.38544	Downtown Toronto	43.653232	-79.385296	Neighborhood
2	Toronto, Canada	43.64869	-79.38544	Adelaide Club Toronto	43.649279	-79.381921	Gym / Fitness Center
3	Toronto, Canada	43.64869	-79.38544	Delta Hotels by Marriott Toronto	43.642882	-79.383949	Hotel
4	Toronto, Canada	43.64869	-79.38544	Nathan Phillips Square	43.652270	-79.383516	Plaza

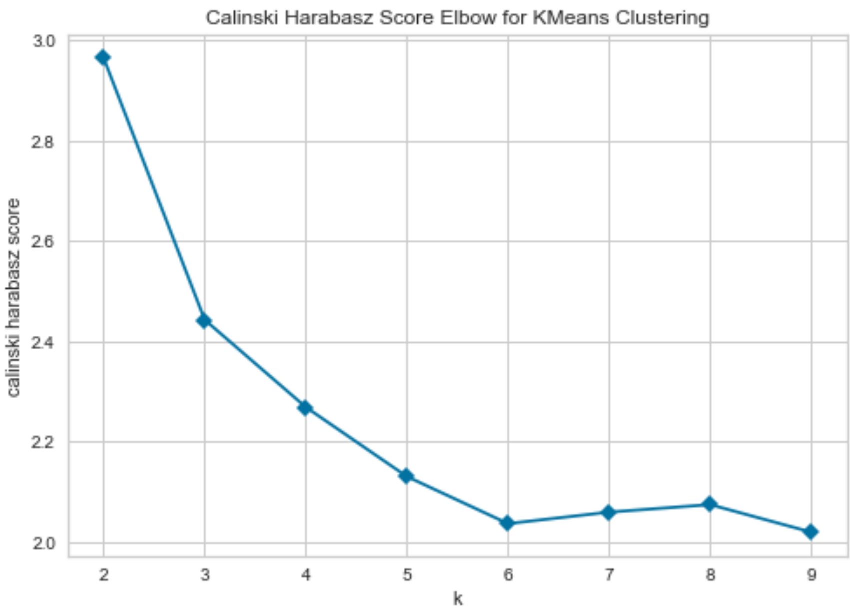
3.3 Select 10 most common categories as features

Once we have all the venues for all the 10 cities, I found that there're total **219** unique Venue Categories. Next I will group the data by City Name and by taking the mean of the frequency of each Venue Category, then select only the 10 most common categories.

	Name	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
0	Amsterdam, Netherlands	Hotel	Coffee Shop	Ice Cream Shop	Bookstore	Canal	Bar	Breakfast Spot	Cocktail Bar	Bakery	Plaza
1	London, UK	Hotel	Theater	Park	Plaza	Grocery Store	Hotel Bar	Wine Bar	Art Gallery	History Museum	Bookstore
2	Los Angeles, US	Coffee Shop	Brewery	American Restaurant	Taco Place	Plaza	Theater	Bakery	Italian Restaurant	Art Gallery	Concert Hall
3	New York City, US	Park	Memorial Site	Bakery	Italian Restaurant	Theater	Garden	Hotel	Ice Cream Shop	Yoga Studio	Dance Studio
4	Paris, France	Plaza	Hotel	Ice Cream Shop	Bookstore	Garden	Wine Bar	Art Museum	Sandwich Place	Fountain	Coffee Shop
5	San Francisco, US	Bakery	Coffee Shop	Grocery Store	Pizza Place	Yoga Studio	Park	Gym / Fitness Center	Dance Studio	Marijuana Dispensary	Restaurant
6	Singapore	Hotel	Waterfront	Coffee Shop	Wine Bar	Event Space	Park	Performing Arts Venue	Art Gallery	Garden	Bistro
7	Sydney, Australia	Café	Park	Cocktail Bar	Hotel	Coffee Shop	Thai Restaurant	Ice Cream Shop	Theater	Bakery	Bookstore
8	São Paulo, Brazil	Ice Cream Shop	Italian Restaurant	Art Museum	Theater	Pizza Place	Bookstore	Park	Restaurant	Hotel	Dance Studio
9	Toronto, Canada	Coffee Shop	Park	Bakery	Café	Dessert Shop	Sandwich Place	Restaurant	Dance Studio	Japanese Restaurant	Farmers Market

3.4 Use Elbow Method the determine the K value

Using Elbow Method, I decided to use 6 as the K value in the following K-Means Clustering.



3.5 Use K-means Clustering to group cities

Using 6 as K value, I run K-means Clustering and got following result.

	Name	Cluster Labels
0	Toronto, Canada	2
1	Sydney, Australia	3
2	Singapore	0
3	São Paulo, Brazil	1
4	San Francisco, US	2
5	New York City, US	2
6	Paris, France	4
7	Los Angeles, US	5
8	London, UK	0
9	Amsterdam, Netherlands	4

4. Result

From the analysis, I found that Northern American cities, Toronto, San Francisco, New York City are more similar, except Los Angeles. European cities, Paris and Armsterdam, are similar to each other. Singapore and London belong to the same cluster. Other cities, Sao Paulo and Sydney both belong to its own cluster.

5. Discussion

As we only have 10 cities for the clustering, from the result, we can roughly group the cities based on their locations. For example, Northern American, Latin American, Oceanian, European, etc. For Singapore and London cluster, it might because Singapore used to be part of UK and is still having close relationship with UK.

6. Conclusion

There might be many factors affected the formation of large multicultural cities, location is definitely one of most important one. Other than location, history also plays important part.

7. Future directions

7.1 Getting more cities

More cities need to be added for the further study, so that more factors can be found for the formation of large multicultural cities.

7.2 Dividing the area

Because there’s a limitation in Foursquare API for the number of nearby venues returned, the next step will be dividing the center area (5km from Centre Point) into smaller circular areas with radius of 1000 meters. For each circular area, use 1200 meters (to cover the gap among circulars) as radius and 100 as the max number of venues returned. Then with the combination of all the circular areas, and removing duplicated venues, the dataset will be completed.

7.3 Use more accurate coordinates

As I just use the coordinates obtained from Python Geocoder library as the center point of the city, this might be inaccurate. Further analysis will be using more accurate coordinates as the center point and also adjust the radius of each city based on the city area.

7.4 Grouping venue categories

The Venue Categories can also be grouped to reduce the features, for example, “Ice Cream Shop” can be grouped into “Desert Shop”,