

# Comparing the similarities of different cities across the globe

Priya Dwivedi

May 2020

## 1 Introduction

An ill planned city can be a pain for people living in that city. For example, it is important to have enough public space, parks so that families can spend their quality time or to do any exercise. Similarly, for young people, it becomes important to have enough bar, pub, cafe where they can spend their quality night life. Until now, people just move to the city where they find good jobs. However, living in a city without enough amenities can be mentally challenging.

At the same time, municipalities are trying to attract young and skilled labours to their city in order to improve the city's economic activity. Therefore, the department of marketing works together with the urban planner to improve the city and utilise the space efficiently.

To help the urban planner or the person who is looking for a new city to move, we will examine the city's neighborhood and assess its similarities with other cities across the globe. In this project, we will answer the following question:

- Which of the cities falls in the same cluster?
- Which of the venue categories are dominated in a particular city or cluster?
- What actions need to be taken in order to improve the resident lifestyle of that particular city?

We will be investigating multiple cities from Europe, Asia and Australia. This analysis will be performed using Foursquare API.

The targeted audience for this project will be the urban planner as well as the people looking for a new city to move.

## 2 Data

After defining the problem, the next step for data science project is to find the relevant data in order to solve the problem. This step is detailed in the sub-section Data acquisition. Furthermore, this will be followed by data pre-processing.

### 2.1 Data acquisition

First, we chose the 4 cities from Australia: Sydney, Melbourne, Brisbane, and Perth, 4 cities from India: Delhi, Mumbai, Ahmedabad, and Bangalore, 5 cities from Europe: Amsterdam, Rotterdam, Cologne, Munich and Berlin. Then we used their names and countries as an input to geopy python library. This gave us the latitude and longitude for each city.

Moreover, we used foursquare service to extract the information about the nearby venues or venue categories in a city. Foursquare API is location based service with diverse information about venues. The latitudes and longitudes were used as an input to Foursquare API in order to explore the nearby venues. We chose to explore the area around the city center by defining the radius of 8 km and limited the number of venues to 100.

For example, if  $x$ ,  $y$  and  $z$  are the cities to be investigated, the following steps were performed to extract the relevant data.

1. Found the latitude and longitude of cities using geopy python library.
2. These latitudes and longitudes were the input for foursquare API. Using foursquare API, explored the nearby venues of the city.
3. Clean and pre-process the data.

In the following sub-section, we will detail about the pre-processing steps.

## 2.2 Data pre-processing

The collected data via Foursquare API had venue name, latitude, longitude and venue category. After dropping the non-significant columns e.g., venue name, latitude and longitude and adding the significant column, e.g., city name, we created a dataframe containing the cities and venue categories. These venue categories were then converted to dummy variables corresponding to each city. We grouped the data by cities and took average of dummy variables corresponding to the same cities. The resulted dataframe was used for further analysis.

## 3 Methodology

Before we begin the analysis of the data, let us visualise the selected cities on the world map as shown in Fig. 1. For each city, we extracted the nearby venues and venue categories. The higher the frequency of a venue category, the dominant the presence of the category is. In order to know which categories are dominantly present in a city, we found the top 15th most common venue categories.

We checked the most common venue categories for Amsterdam, Delhi and Sydney separately. Fig. 2 shows the most common venues for Amsterdam. Netherlands is a country known for much more tolerant policy of drugs than rest of the countries. Coffee shops in the Netherlands should not be confused with the American version where you will be served with hot coffee and baked goods. Coffee shops act as a legal dispensaries for marijuana. The strong presence of coffee shops in the bar plot indicate its popularity and demand. Amsterdam is a very popular city for tourists, therefore seeing hotels as 2nd most common venue is not surprising for us. Note that parks and canals are at the bottom of the list. It suggests the city center is hugely crowded with either tourists or shops. Fig. 3 shows the most common venues for Delhi. Indians being famous for their food, Indian restaurants are on top of the list, is completely logical. However, parks/gardens are something that this city lacks. Fig. 4 shows the most common venues for Sydney. The interesting thing to note here is that park are the most common venues. Even being the most populous city in Australia, they have maintained a balanced combination of economical activity and public spaces.

Until now, we have analysed 3 cities from different continents. Each city

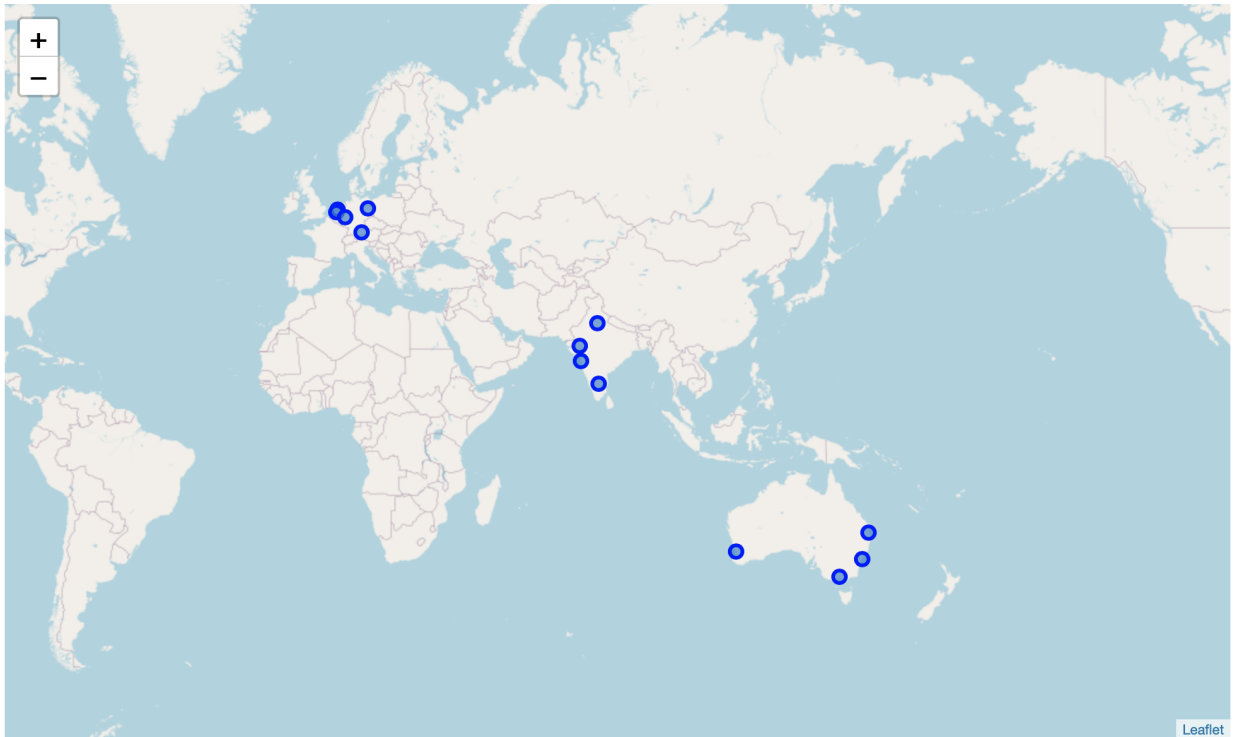


Figure 1: Selected cities on the world map. We have selected 4 cities from Australia, 4 cities from India and 5 cities from Europe.

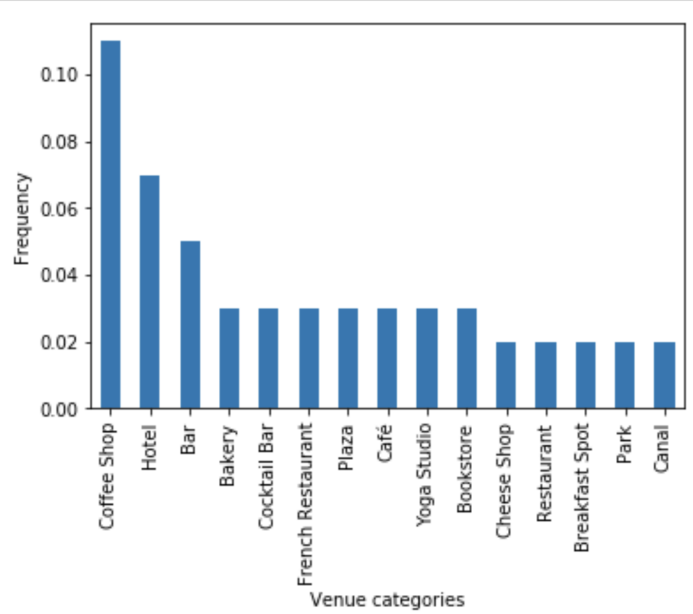


Figure 2: Bar plot for top 15th most common venue categories in Amsterdam

has its own type. We would like to see the same analysis for each city. This will give us the insights about the city as well as the opportunity to improve the city infrastructure. For example, we can clearly see that there no parks in the top 15th most common venue in Delhi. This indicate an opportunity to create some space for parks for healthy and happy lifestyle of the resident. Therefore, we performed the similar analysis for all the selected cities, the results as shown in Fig. 5. However, by closely observing these cities and most common venues, we could not figure out the closely resembling cities. Then comes the importance of machine learning algorithms.

The most obvious choice of machine learning algorithm is k-mean clustering. This algorithm calculates the distance of two data points and creates few clusters for each data point based on the calculated distance. After several iterations of this method, it divides the data points in few clusters and data points in each cluster share large amount of similarities.

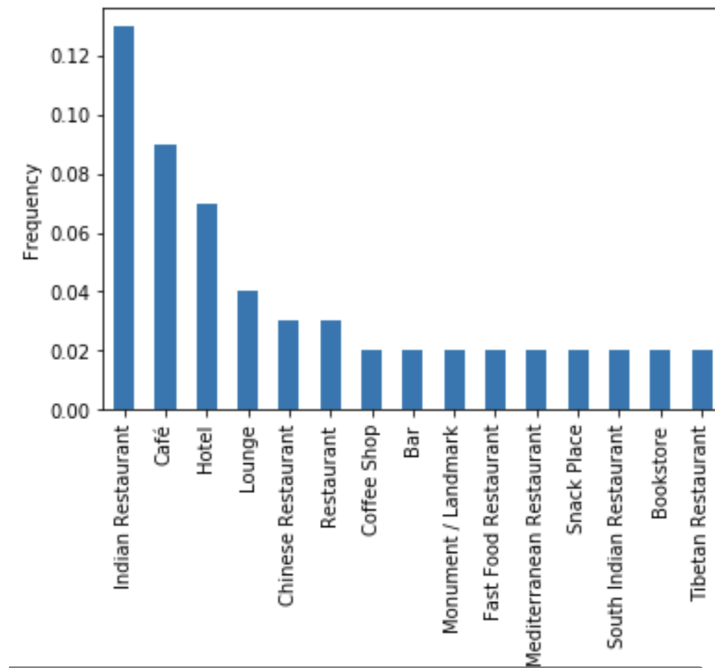


Figure 3: Bar plot for top 15th most common venue categories in Delhi

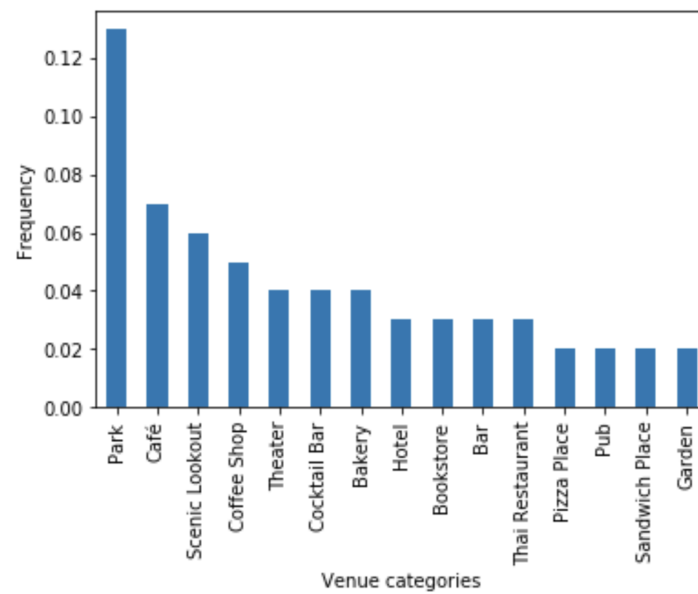


Figure 4: Bar plot for top 15th most common venue categories in Sydney

city	latitude	longitude	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
Perth	-31.952712	115.860480	2	Café	Coffee Shop	Park	Bakery	Italian Restaurant	Hotel	Vietnamese Restaurant	Asian Restaurant	Ice Cream Shop	Gym
Sydney	-33.854816	151.216454	2	Park	Café	Scenic Lookout	Coffee Shop	Theater	Cocktail Bar	Bakery	Hotel	Bookstore	Bar
Melbourne	-37.814218	144.963161	2	Café	Coffee Shop	Park	Bar	Cocktail Bar	Ice Cream Shop	Plaza	Hotel	Bakery	Japanese Restaurant
Brisbane	-27.468968	153.023499	0	Coffee Shop	Park	Beer Bar	Brewery	Hotel	Café	Italian Restaurant	Bar	Farmers Market	Scenic Lookout
Amsterdam	52.372760	4.893604	0	Coffee Shop	Hotel	Bar	Bakery	Cocktail Bar	French Restaurant	Plaza	Café	Yoga Studio	Bookstore
Rotterdam	51.922896	4.463173	0	Coffee Shop	Café	Bakery	Hotel	Restaurant	Bar	Park	Movie Theater	Sandwich Place	Deli, Bodega
Berlin	52.517037	13.388860	0	Coffee Shop	Park	Bookstore	Ice Cream Shop	Bakery	Wine Bar	Monument / Landmark	Café	Gourmet Shop	Art Gallery
Munich	48.137108	11.575382	2	Café	Plaza	Ice Cream Shop	Cocktail Bar	Gourmet Shop	Hotel	Coffee Shop	Department Store	Beer Garden	Italian Restaurant
Cologne	50.938361	6.959974	2	Park	Café	Coffee Shop	Burger Joint	Italian Restaurant	Plaza	Bar	Hotel	Ice Cream Shop	Restaurant
Delhi	28.651718	77.221939	1	Indian Restaurant	Café	Hotel	Lounge	Chinese Restaurant	Restaurant	Coffee Shop	Bar	Monument / Landmark	Fast Food Restaurant
Mumbai	18.938771	72.835335	1	Indian Restaurant	Hotel	Dessert Shop	Café	Fast Food Restaurant	Cricket Ground	Scenic Lookout	Pizza Place	Ice Cream Shop	Seafood Restaurant
Bangalore	12.979120	77.591300	1	Hotel	Ice Cream Shop	Indian Restaurant	Bakery	Brewery	Italian Restaurant	Department Store	Burger Joint	Lounge	Pub
Ahmedabad	23.021624	72.579707	1	Café	Indian Restaurant	Restaurant	Hotel	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Dessert Shop	Pizza Place	Coffee Shop	Multiple

Figure 5: This table shows the most common venues for all the selected cities.

## 4 Results

In Fig. 9, we have shown the results after running k-mean clustering algorithm. The pre-defined number of clusters were 3. The clusters after running the k-mean clustering algorithms are shown in the Table 1. In cluster 0 as

Cluster 0	Cluster 1	Cluster 2
Brisbane	Delhi	Perth
Amsterdam	Mumbai	Sydney
Rotterdam	Ahmedabad	Melbourne
Berlin	Bangalore	Cologne
		Munich

Table 1: Clusters formed after running k-mean cluster algorithm.

shown in Fig. 6, the most dominant venues are coffee shop, Park, hotel, brewery and bar. Apart from these venues, they also have yoga studio, monument / landmark and art gallery. In cluster 1 as shown in Fig. 7, the most dominant venues are Indian restaurant, hotel and cafe. However, there no park or scenic lookout in the 8 km range. It is possible that park and scenic lookout can be at the outskirts of city or there are none. This can be an opportunity for the stakeholders to look into details and improve the city lifestyle by taking actions. In cluster 2 as shown in Fig. 8, The most dominant venues in this clusters are cafe, park/ scenic lookout.

city	latitude	longitude	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
Brisbane	-27.468968	153.023499	0	Coffee Shop	Park	Beer Bar	Brewery	Hotel	Café	Italian Restaurant	Bar	Farmers Market	Scenic Lookout
Amsterdam	52.372760	4.893604	0	Coffee Shop	Hotel	Bar	Bakery	Cocktail Bar	French Restaurant	Plaza	Café	Yoga Studio	Bookstore
Rotterdam	51.922896	4.463173	0	Coffee Shop	Café	Bakery	Hotel	Restaurant	Bar	Park	Movie Theater	Sandwich Place	Deli / Bodega
Berlin	52.517037	13.388860	0	Coffee Shop	Park	Bookstore	Ice Cream Shop	Bakery	Wine Bar	Monument / Landmark	Café	Gourmet Shop	Art Gallery

Figure 6: Most common venues for cluster 0.



city	latitude	longitude	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	1
Delhi	28.651718	77.221939	1	Indian Restaurant	Café	Hotel	Lounge	Chinese Restaurant	Restaurant	Fast Food Restaurant	Snack Place	Monument / Landmark	Bookstore	F
Mumbai	18.938771	72.835335	1	Indian Restaurant	Hotel	Dessert Shop	Café	Scenic Lookout	Fast Food Restaurant	Pizza Place	Ice Cream Shop	Cricket Ground	Theater	
Bangalore	12.979120	77.591300	1	Hotel	Ice Cream Shop	Indian Restaurant	Bakery	Pub	Brewery	Department Store	Lounge	Burger Joint	Italian Restaurant	
Ahmedabad	23.021624	72.579707	1	Café	Indian Restaurant	Restaurant	Hotel	Vegetarian / Vegan Restaurant	Dessert Shop	Fast Food Restaurant	Pizza Place	Coffee Shop	Multiplex	

Figure 7: Most common venues for cluster 1.

city	latitude	longitude	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
Perth	-31.952712	115.860480	2	Café	Coffee Shop	Park	Bakery	Asian Restaurant	Italian Restaurant	Hotel	Ice Cream Shop	Vietnamese Restaurant	Gym
Sydney	-33.854816	151.216454	2	Park	Café	Scenic Lookout	Coffee Shop	Theater	Cocktail Bar	Bakery	Garden	Bar	Hotel
Melbourne	-37.814218	144.963161	2	Café	Coffee Shop	Park	Bar	Cocktail Bar	Plaza	Ice Cream Shop	Thai Restaurant	Bakery	Gym / Fitness Center
Munich	48.137108	11.575382	2	Café	Plaza	Ice Cream Shop	Cocktail Bar	Gourmet Shop	Coffee Shop	Department Store	Beach	Italian Restaurant	Beer Garden
Cologne	50.938361	6.959974	2	Park	Café	Coffee Shop	Italian Restaurant	Plaza	Burger Joint	Bar	Hotel	Ice Cream Shop	French Restaurant

Figure 8: Most common venues for cluster 2.

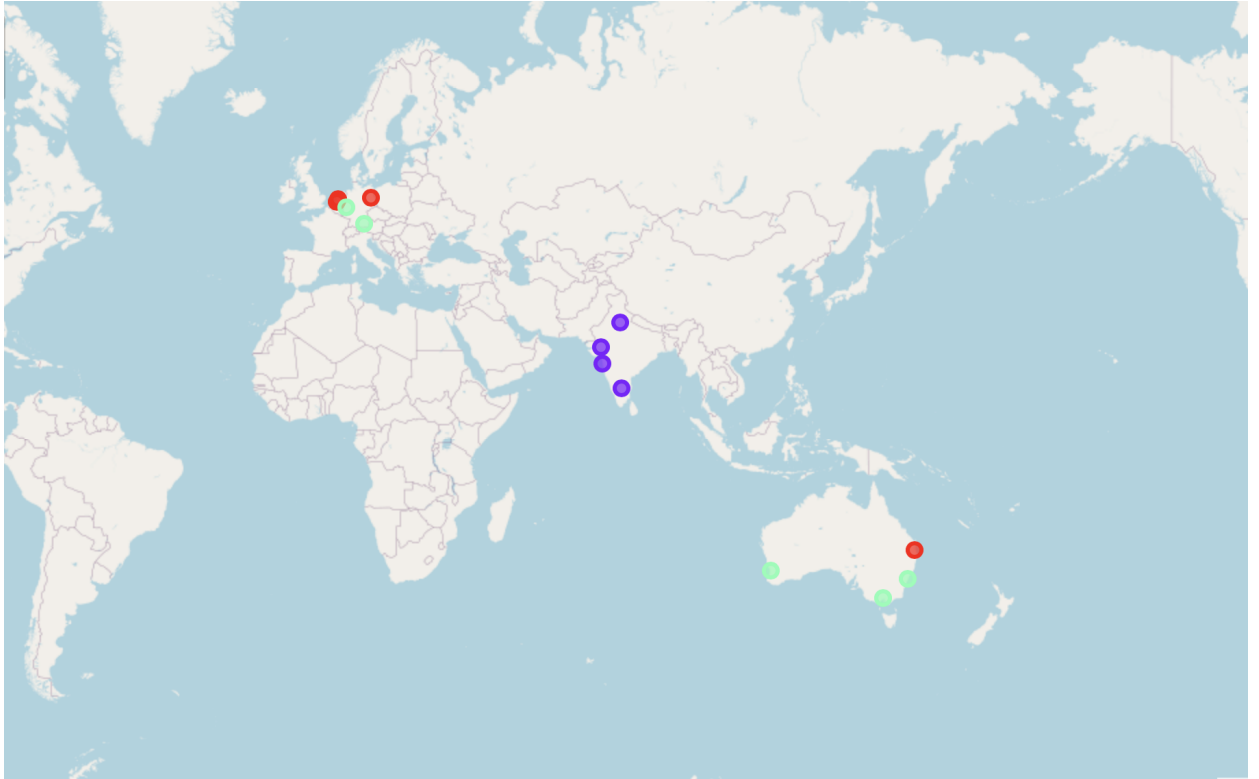


Figure 9: Clusters after running k-mean algorithm.

## 5 Discussions

In this work, we have not considered the influence of the population and population density of the cities. This can certainly influence our results as highly populated city might be bigger in size as well. Where as, we have consider 8 km of radius for our study. It is possible that a large fraction of population lives out side of this circle. Hence, the results of this study can be improved by considering the cities of similar population density.

## 6 Conclusions

In this project, we compared the similarities of different cities across the globe. This was done by using python library called geopy and foursquare API. We extracted the nearby venues of each city within a radius of 8 km.

Then, we applied k-mean clustering algorithms to the data set in order to find the clusters with similar cities. This gives an important insight about the cities. The results in this work can be further improved by considering the population density.