

Influence of Culture in City Venues

Rafael Ferreira

May 2nd, 2020

1. Introduction

Living in a multicultural world, the way cities offer entertainment to citizens and tourists is getting more globalized. Nevertheless, cities around the world keep their distinct way of presenting themselves and sharing their particular culture.

A group of social scientists, the stakeholders of this project, want to answer the following question:

Does culture influence the kind of venues in a City?

To address this issue, the locations data of three major cities will be used, all three represent different continents and cultures:

- **Tokyo** - Representing Asia, one of the biggest cities in the continent.
- **Barcelona** - Representing Europe, a large city in the south of Spain with a differentiated culture.
- **New York** - More specifically, Queens' area, representing America.

2. Data Processing

For Tokyo, we will get information about the neighborhoods [1] and append the location data via Nominatim module in Geopy. For Barcelona we will use the same method [2]. Lastly, for New York we will use the json file [3] with New York spatial data, getting Queens' neighborhoods data and respective locations. The data processing and wrangling details can be consulted in my GitHub repository for this study [4].

The datasets with location and data have the following structure (Table 1).

	District	Neighborhood	Latitude	Longitude
0	Chiyoda	Nagatacho	35.675618	139.743469
1	Chiyoda	Kasumigaseki	35.674054	139.750972
2	Chiyoda	Otemachi	35.686757	139.763616
3	Chiyoda	Marunouchi	35.680656	139.765222
4	Chiyoda	Akihabara	35.699736	139.771250

Table 1 - Location dataset example for Tokyo

Once the location data is computed, we use the python Folium library to visualize it in a map with geographic details (Figure 1).

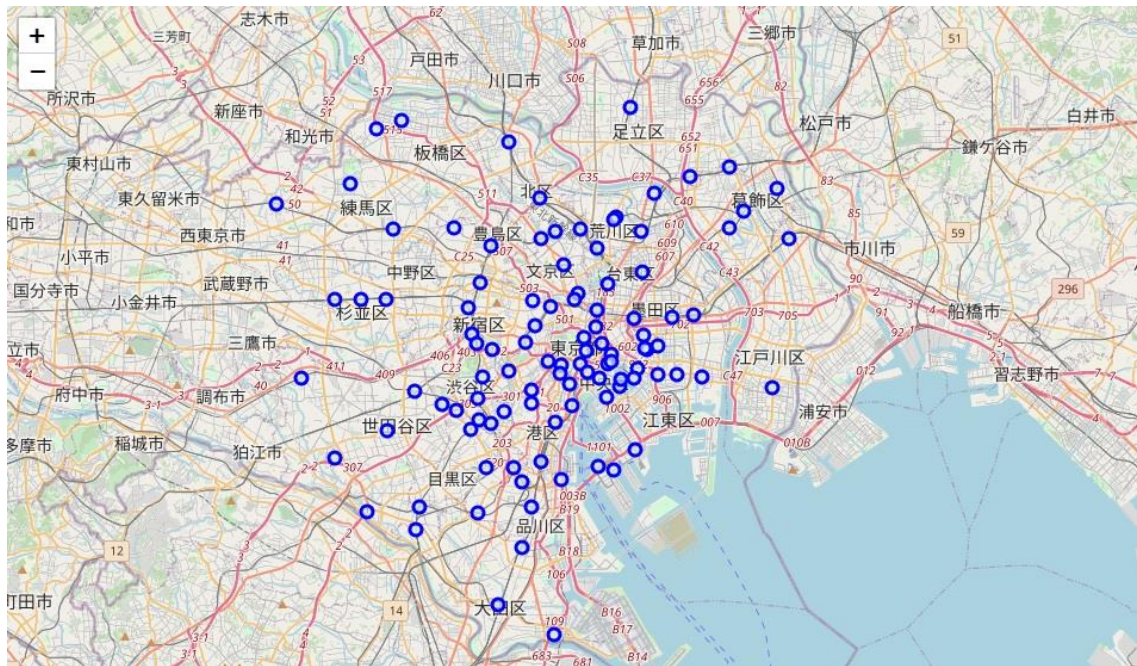


Figure 1 - Neighborhood map with Folium for Tokyo

In total, we use the location data for 106 neighborhoods in Tokyo, 75 in Barcelona and 81 in New York: Queens. With this information we can study the type of venues in these locations.

3. Methodology

After gathering the location data of the cities to explore, we will get information about the most common venues for each neighborhood. This will be done using the Foursquare Api.

In the next step, all the data and information regarding the most common venues for each neighborhood of the studied cities will be compiled in a single dataset, ready for segmenting and clustering.

The clustering step will be done using the K-means machine learning algorithm. The data will be segmented in different clusters according to the type of venues of the neighborhoods. The location data will not be used in the clustering process hence the cluster labeling will not be influenced by the location of the cities.

Lastly, the clusters obtained in the previous step will be associated to the location data of the cities. Then, through data visualization, we will access what is the representation of the clusters in each city.

4. Analysis

After the venue information was returned by Foursquare, a table which shows a list of top 10 venue category for each neighborhood was created. You can see the example for Barcelona neighborhoods below (Table 2):

	Neighborhood	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	Bon Pastor	Grocery Store	Paper / Office Supplies Store	Café	Athletics & Sports	Restaurant	Fabric Shop	Electronics Store	Empanada Restaurant	Ethiopian Restaurant	Exhibit
1	Can Baro	Spanish Restaurant	Scenic Lookout	Grocery Store	Dessert Shop	Chinese Restaurant	Basketball Court	Breakfast Spot	Bar	Bakery	Café
2	Canyelles	Hot Spring	Metro Station	Café	Market	Mediterranean Restaurant	Soccer Field	Flea Market	Flower Shop	Fish & Chips Shop	Food Truck
3	Ciutat Meridiana	Metro Station	Train Station	Plaza	Supermarket	Grocery Store	Park	Fabric Shop	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
4	Diagonal Mar i el Front Marítim del Poblenou	Restaurant	Mediterranean Restaurant	Beach Bar	Athletics & Sports	Beach	Spanish Restaurant	Café	Thai Restaurant	Beer Garden	Lounge

Table 2 - Dataset with most common venues for Barcelona neighborhoods

4.1. K-Means Clustering

To have a dataset that embraces all the neighborhoods to cluster we have combine the three cities datasets in one. Also, the neighborhood information must be dropped, we want to cluster data only based in the kind of venues.

The data will be clustered using K-Means algorithm and segmented into 10 different clusters. After clustering, the clustering labels will be added to the most common venues table for interpretation (Table 3).

Cluster Labels		Neighborhood	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	2	Akihabara	Hobby Shop	Café	Electronics Store	Ramen Restaurant	Donburi Restaurant	Arcade	Rock Club	Video Game Store	Toy / Game Store	Comic Shop
1	2	Aobadai	Japanese Restaurant	Convenience Store	Coffee Shop	Chinese Restaurant	Café	Italian Restaurant	Ramen Restaurant	BBQ Joint	Soba Restaurant	Bar
2	2	Aomi	Exhibit	Plaza	Convenience Store	Park	Coffee Shop	Sporting Goods Shop	Museum	Clothing Store	Train Station	Chocolate Shop
3	1	Aoto	Convenience Store	Italian Restaurant	Chinese Restaurant	Mobile Phone Shop	Platform	Intersection	Ramen Restaurant	Sake Bar	Grocery Store	Japanese Restaurant
4	2	Aoyama	Convenience Store	Coffee Shop	Italian Restaurant	Ramen Restaurant	Soba Restaurant	Japanese Restaurant	Sushi Restaurant	Baseball Stadium	Rock Club	French Restaurant

Table 3 - Dataset with most common venues and respective cluster labels

5. Results and Discussion

Once the data is clustered and labeled, we can visualize the cluster distribution and discuss the results obtained.

5.1. Tokyo

Analyzing the map in Figure 2. We can see that most of the neighborhoods in Tokyo are into Cluster 1 (Purple spots) and Cluster 2 (Blue spots).

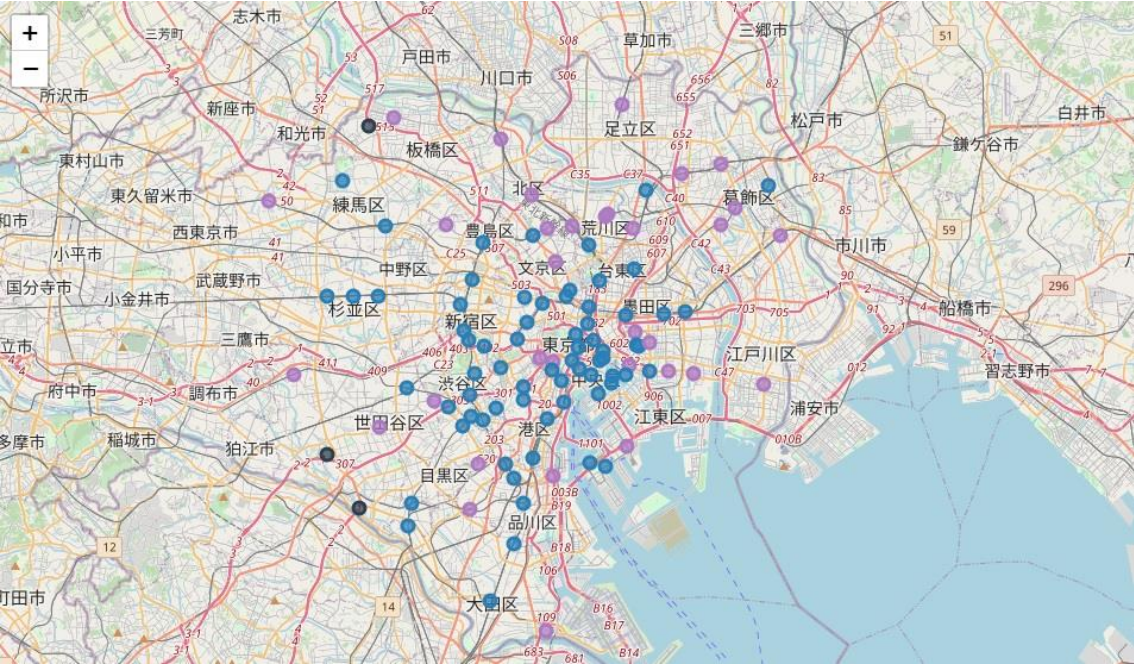


Figure 2 - Map of Tokyo with clustered neighborhoods

5.2. Barcelona

In Barcelona (Figure 3), we can see that most of the neighborhoods a are into Cluster 5 (Green spots).

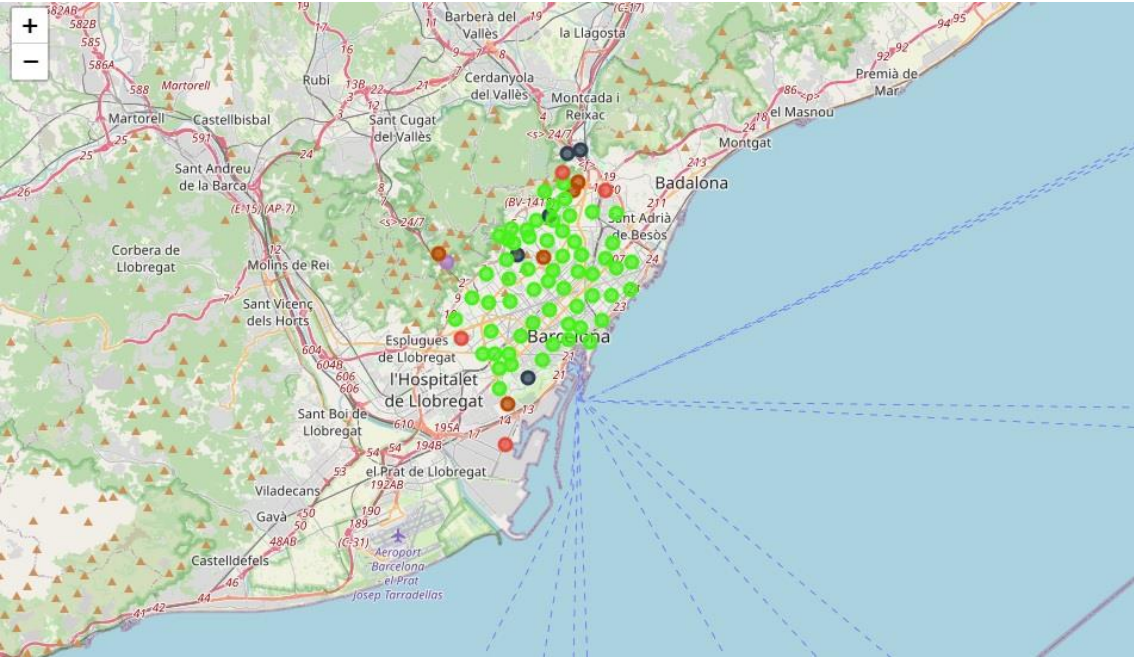


Figure 3 - Map of Barcelona with clustered neighborhoods

5.3. New York: Queens

Lastly, we can see that most of the neighborhoods in Queens (Figure 4) are into Cluster 7 (Red spots).

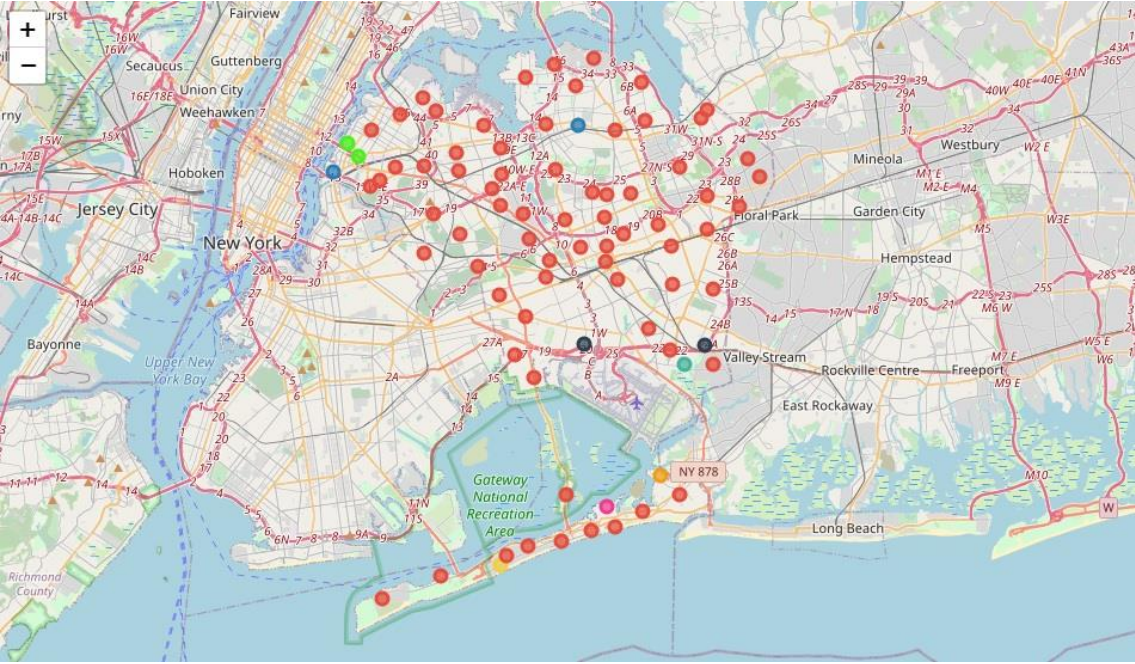


Figure 4 - Map of Queens with clustered neighborhoods

5.4. Cluster Distribution

We can access neighborhood distribution for each city in an aggregated way. First, we insert one column with the city name to the dataset containing the venues and cluster information. Then, the dataset is transformed in a pivot table for better interpretation (Table 4).

	0	1	2	3	4	5	6	7	8	9
Barcelona	5	1	0	0	0	60	0	4	5	0
Queens	0	0	2	1	1	2	1	71	2	1
Tokyo	0	34	69	0	0	0	0	0	3	0

Table 4 - Pivot Table with the number of neighborhoods in each cluster (column index) by City (row index)

Using the pivot table, we can plot the number of neighborhoods in each cluster by city using matplotlib (Figure 5).

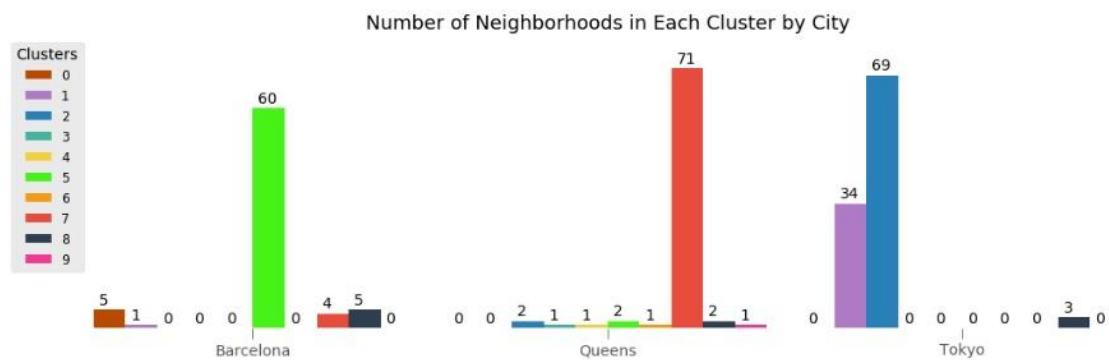


Figure 5 - Bar plot with number of neighborhoods in each cluster by City

Interpreting the graphic above we can observe:

- Barcelona's neighborhoods are mostly inside Cluster 5 (60 neighborhoods).
- New York represented by Queens is the city with most varied clusters. Still Cluster 7 (71 neighborhoods) is the most frequent.
- Tokyo has high representation in Clusters 1 and 2 (34 and 69 neighborhoods respectively). Cluster 1 is the most frequent.

5.5. Cluster Analysis

In Tokyo, Clusters 1 and 2 were the main clusters. Cluster 1 can be described by the Table 5 below.

	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
count	35	35	35	35	35	35	35	35	35	35
unique	3	11	17	27	22	23	28	28	25	26
top	Convenience Store	Ramen Restaurant	Japanese Restaurant	Intersection	Donburi Restaurant	Bus Stop	Chinese Restaurant	Sake Bar	Chinese Restaurant	Supermarket
freq	31	9	7	3	3	4	3	3	4	3

Table 5 - Most common venues for cluster 1

We can see that cluster 1 has a lot of Convenience Stores, Ramen and Japanese Restaurants, indicating great culture influence. Another example, Sake Bar is the 8th most common type of venue in this cluster.

For Cluster 2 we have the following common venues (Table 6).

	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
count	71	71	71	71	71	71	71	71	71	71
unique	20	15	29	28	31	39	45	47	45	50
top	Japanese Restaurant	Japanese Restaurant	Convenience Store	Japanese Restaurant	Coffee Shop	Chinese Restaurant	Hotel	BBQ Joint	Chinese Restaurant	Chinese Restaurant
freq	16	14	10	9	7	7	5	7	5	4

Table 6 - Most common venues for Cluster 2

Cluster 2 is dominated by Japanese Restaurants. Also, there is a lot of Chinese Restaurants in these areas. Once again, culture plays a pivotal role in this outcome.

In Barcelona, we see that Cluster 5 is the most common among neighborhoods (Table 7).

	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
count	62	62	62	62	62	62	62	62	62	62
unique	24	26	35	30	35	37	39	43	45	42
top	Tapas Restaurant	Hotel	Mediterranean Restaurant	Spanish Restaurant	Spanish Restaurant	Restaurant	Café	Supermarket	Tapas Restaurant	Supermarket
freq	9	6	8	6	5	4	4	5	3	4

Table 7 - Most common venues in Cluster 5

In Cluster 5, culture has a high impact in the cluster with a lot of Tapas, Mediterranean and Spanish Restaurants.

Hotel is the 2nd most common venue for many neighborhoods, this is explained by the high level of tourism in Barcelona.

New York: Queens neighborhoods are mostly into Cluster 7 (Table 8).

	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
count	75	75	75	75	75	75	75	75	75	75
unique	43	46	52	49	46	51	53	55	52	56
top	Deli / Bodega	Pizza Place	Deli / Bodega	Donut Shop	Pizza Place	Fast Food Restaurant	Donut Shop	Fast Food Restaurant	Bakery	Sandwich Place
freq	9	6	4	5	9	4	4	4	4	4

Table 8 - Most common venues in Cluster 7

Cluster 7 is characterized by the Deli /Bodega venues (kind of convenience store), Pizza, Fast Food and Donut Shops show the culture influence in the type of venues.

6. Conclusion

The purpose of this project was to access the culture impact in the type of venues. After segmenting the information on the venues of neighborhoods in each city we can conclude that culture influences venues.

The data was segmented in 10 distinct clusters using the K-Means algorithm. Each of the cities had shown a trend of specific clusters that had some common features. New York (Queens in particular) where the city with the most varied set of clusters, with neighborhoods placed within a range of 8 distinct clusters.

Cities were structured as follows:

- **Tokyo** - Tokyo neighborhoods were mainly into Clusters 1 and 2 in which among the most common venues are Oriental Restaurants, Convenience Stores and Sake Bars.
- **Barcelona** - Barcelona neighborhoods were mainly into Cluster 5 in which among the most common venues were Tapas Restaurants (typical south Spanish restaurants), Mediterranean and Spanish Restaurants.

- **New York: Queens** - Queens neighborhoods were mainly into Cluster 7 in which among the most common venues were Deli / Bodegas (typical New York stores), Pizza Places and Fast Food.

Attending this information, we can confirm that *Culture has a high influence in the kind of venues in the cities.*

7. References

- [1] - https://en.wikipedia.org/wiki/Special_wards_of_Tokyo
- [2] - https://en.wikipedia.org/wiki/Districts_of_Barcelona
- [3] - https://cocl.us/new_york_dataset
- [4] - <https://nbviewer.jupyter.org/github/rafa2024/github-coursera-IBM/blob/master/TheBattleOfNeighs2.ipynb>