NYC vs LA – defining similar neighborhoods

1.Introduction

NYC and LA - No two cities in the states are compared as much as these two. One might prefer living in LA for many reasons but others will choose to live in NYC for almost same amount of reasons. Millions of articles have been published comparing many different aspects of these two Cities. But for this project, we are going to focus on similarities rather than differences. Our goal is to explain a New Yorker, who has never been in LA, about neighborhoods in LA by highlighting their similarities with neighborhoods in in NYC, or vice versa.

2.Data

a. List of NY neighborhoods and zip codes is retrieved from
https://geo.nyu.edu/catalog/nyu_2451_34572 as a form of Json.

	City	Neighborhood	Latitude	Longitude
0	NY	Wakefield	40.894705	-73.847201
1	NY	Co-op City	40.874294	-73.829939
2	NY	Eastchester	40.887556	-73.827806
3	NY	Fieldston	40.895437	-73.905643
4	NY	Riverdale	40.890834	-73.912585

b. List of LA cities and counties within LA (we will name them as 'neighborhoods' in this project) and zip codes is retrieved from http://www.laalmanac.com/communications/cm02a90001-90899.php
Data is scraped from above link using BeautifulSoup library

	Zipcode	Neighborhood
0	90001	Los Angeles(South Los Angeles), Florence-Graham
1	90002	Los Angeles(Southeast Los Angeles, Watts)
2	90003	Los Angeles(South Los Angeles, Southeast Los A...
3	90004	Los Angeles(Hancock Park, Rampart Village, Vir...

c. latitude, longitude information of neighborhoods in LA is downloaded from
https://public.opendatasoft.com/ In xlsx format

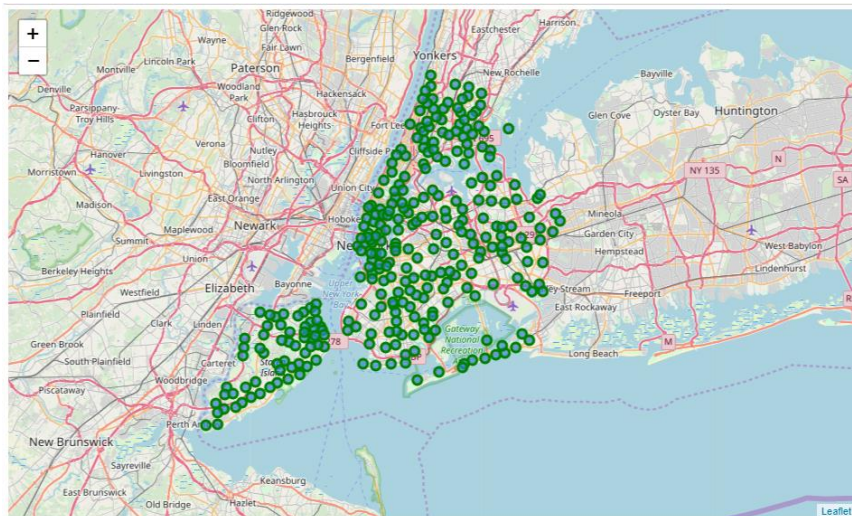|	Zip	City	State	Latitude	Longitude	Timezone	Daylight savings time flag	geopoint
0	92232	Calexico	CA	33.026203	-115.284581	-8	1	33.026203, -115.284581
1	93227	Goshen	CA	36.357151	-119.425371	-8	1	36.357151, -119.425371
2	93234	Huron	CA	36.209815	-120.084700	-8	1	36.209815, -120.0847
3	93529	June Lake	CA	37.765218	-119.077690	-8	1	37.765218, -119.07769
4	93761	Fresno	CA	36.746375	-119.639658	-8	1	36.746375, -119.639658

d. Foursquare API geolocation data for venue information of neighborhoods two cities.

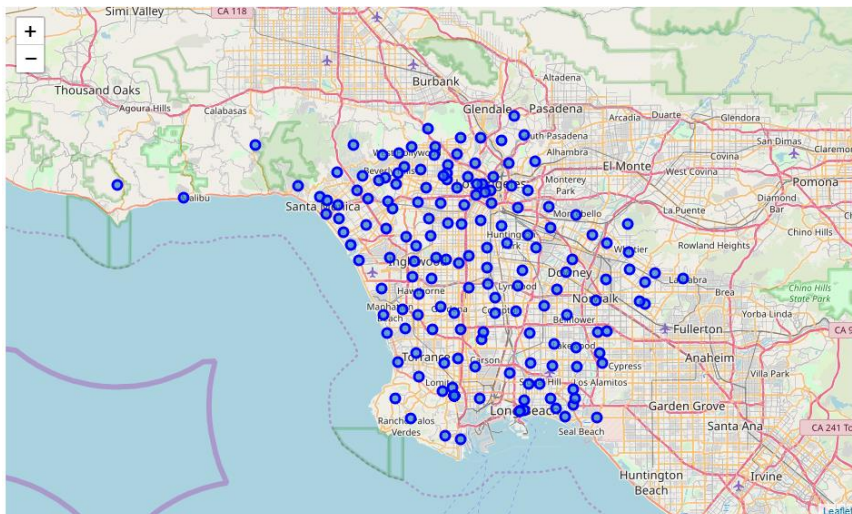# 3. Methodology

a. explanatory analysis

NYC is consisting of 5 boroughs and 306 neighborhoods.

Average distance of two closest points : 533.1m

There are 88 cities in Los Angeles, plus dozens of unincorporated county areas. In total, I was able to retrieve 178 zip codes for all LA neighborhoods. (zip codes for PObox are excluded)

Average distance of two closest points : 1023.3m

b. Clustering

Venue information is downloaded from foursquare API for all 484 neighborhoods from two Cities (NY – 306, LA -178). Then all 484 neighborhoods are categorized into 15 Clusters, using K-mean clustering algorithm.

Below is an example of 5 Clusters and their 5 most common venues.

	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	0	Coffee Shop	Pizza Place	Bar	Italian Restaurant	American Restaurant
1	1	Park	Water Park	Gay Bar	Food Truck	Fountain
2	2	Scenic Lookout	Rock Climbing Spot	Water Park	Gay Bar	Fountain
3	3	Deli / Bodega	Beach	Italian Restaurant	Bus Stop	Bagel Shop
4	4	Pizza Place	Deli / Bodega	Chinese Restaurant	Pharmacy	Grocery Store
5	5	Sculpture Garden	Water Park	Gay Bar	Food Truck	Fountain

4. Results

Cluster Labels	NY ct Neighborhoods	LA ct Neighborhoods
0	101	86
1	2	
2		1
3	28	1
4	72	2
5	1	
6	2	
7	1	
8	76	7
9	1	15
10	1	
11	11	64
12	8	
13	1	2
14	1	
Total	306	178

a. 101 Neighbor hoods from NY/ 86 from LA are Cluster 0. Based on top 5 venues, it looks like Cluster 0 is food district with lots of Coffee shop, Pizza Place, Bars and Italian Restaurants.

Example: Bayside, Dumbo, Flatiron, Gramercy, etc. (from NY) Beverly Hills, Gardena, Long Beach, Signal Hill, etc. (from LA

)

b. 72 Neighbor hoods from NY/ 2 from LA are Cluster 4. Based on top 5 venues, it looks like Cluster 4 is also food district with lots of Pizza Place, Deli / Bodega and Chinese Restaurant

Example: Crown Heights ,Norwood, Queensboro Hill, Sunnyside, etc. (from NY) South Los Angeles, City of Industry , etc. (from LA)

c. 76 Neighbor hoods from NY/ 7 from LA are Cluster 8. Based on top 5 venues, it looks like Cluster 8 is popular place for commuters since the district is full of quick-prepared food (Donut Shop/Sandwich Place), Pharmacy and   Grocery Store
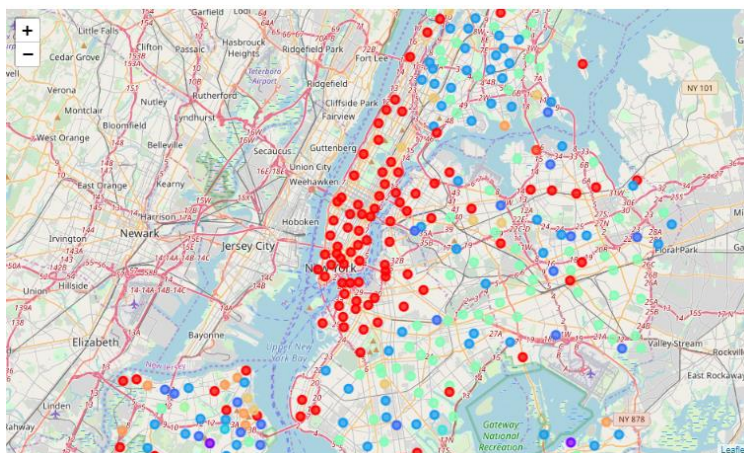
Example: Brownsville, Coney Island, Forest Hills, Prospect Park South, etc. (from NY) , Compton East Rancho Dominguez, La Mirada , Hyde Park , etc. (from LA)

d. 11 Neighbor hoods from NY/ 64 from LA are Cluster 11. Based on top 5 venues, it looks like Cluster 8 is popular place for commuters since the district is full of quick-prepared food (Mexican Restaurant, Fast Food Restaurant, Pizza Place , Sandwich Place) and Convenience Store.
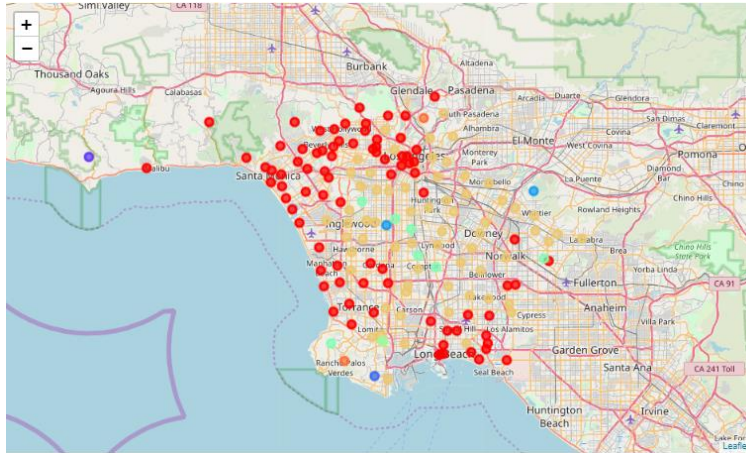
Example: East Harlem, Flatbush, Kingsbridge, Stapleton, etc. (from NY) Carson, Downey, Inglewood, Montebello, etc. (from LA)

Below are map of two cities, with cluster colored by labels.

NY

LA

5. Discussion / Conclusion

It is noticeable that both LA/NY have neighborhoods who are categorized into a same Cluster, based on it's venue information. If you are a New Yorker and wants to find a new place in LA (or vice versa) you can refer to this project to find a neighborhood which is similar to your comfort zone. I believe application for this project is bondless. Beyond LA vs. NY, I hope all people who are planning to settle down in a new place to have a soft-landing with the help from this project.