

# New York Versus Los Angeles

---

# Introduction/Business Problem

---

Many people have aspirations to live in some of the biggest cities in world. Some countries have multiple amazing modern cities with amazing amenities and opportunities, so it can be hard to decide which of cities to choose. We're going to look at two of major metropolises in the United States: New York City and Los Angeles. We are going to compare the different aspects of these two cities to find out which one is right for you. We are going to compare the different neighborhoods in both cities in terms of the available amenities and places you can go to in the different cities. We will find out if these two cities are similar or actually quite different. This will be for anyone finding it difficult to choose between living in any of the two cities.

# Data

---

The data we will be using for this report will be a dataset of the different neighbourhoods in the two cities as well as the different places in the two cities which we will acquire from the Wikipedia page: and the foursquare API for collecting location data. Using the data, we will group the neighbourhood into similar clusters and we will see if the neighbourhoods from the different cities fall into similar clusters or different clusters. This should give a sense of the similarities between the two cities.

Data Cleaning: The New York neighborhood data was retrieved from an IBM server so it didn't need much data cleaning. The Los Angeles data was retrieved from the Wikipedia page below: [https://en.wikipedia.org/wiki/List\\_of\\_districts\\_and\\_neighborhoods\\_in\\_Los\\_Angeles](https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_in_Los_Angeles). Some of the neighbourhood retrieved couldn't be match with coordinates using the Geopy API so those neighborhoods were dropped. We also got the wrong coordinates for the Angelus Vista neighborhood so we drop that as well.

# Methodology

---

## **Explanatory Data Analysis:**

All our neighborhoods along with their coordinates are put into a single data frame. Using the Foursquare API we gather all the venues that we can for all the neighborhoods. We find the top 10 most common venue-categories for each neighbourhood and we put that into a dataframe.

	Neighborhood	Latitude	Longitude	City	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
0	Wakefield	40.894705	-73.847201	New York	8.0	Pharmacy	Caribbean Restaurant	Donut Shop	Sandwich Place	Dessert Shop	Deli / Bodega	Ice Cream Shop	Laundromat	Escape Room	Factory
1	Co-op City	40.874294	-73.829939	New York	8.0	Bus Station	Fast Food Restaurant	Pizza Place	Bagel Shop	Basketball Court	Park	Donut Shop	Liquor Store	Pharmacy	Discount Store
2	Eastchester	40.887556	-73.827806	New York	8.0	Bus Station	Caribbean Restaurant	Deli / Bodega	Diner	Cosmetics Shop	Bowling Alley	Bus Stop	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant
3	Fieldston	40.895437	-73.905643	New York	8.0	Plaza	Medical Supply Store	Bus Station	River	Empanada Restaurant	English Restaurant	Entertainment Service	Escape Room	Farm	Ethiopian Restaurant
4	Riverdale	40.890834	-73.912585	New York	8.0	Bus Station	Park	Home Service	Bank	Medical Supply Store	Baseball Field	Playground	Plaza	Gym	Food Truck



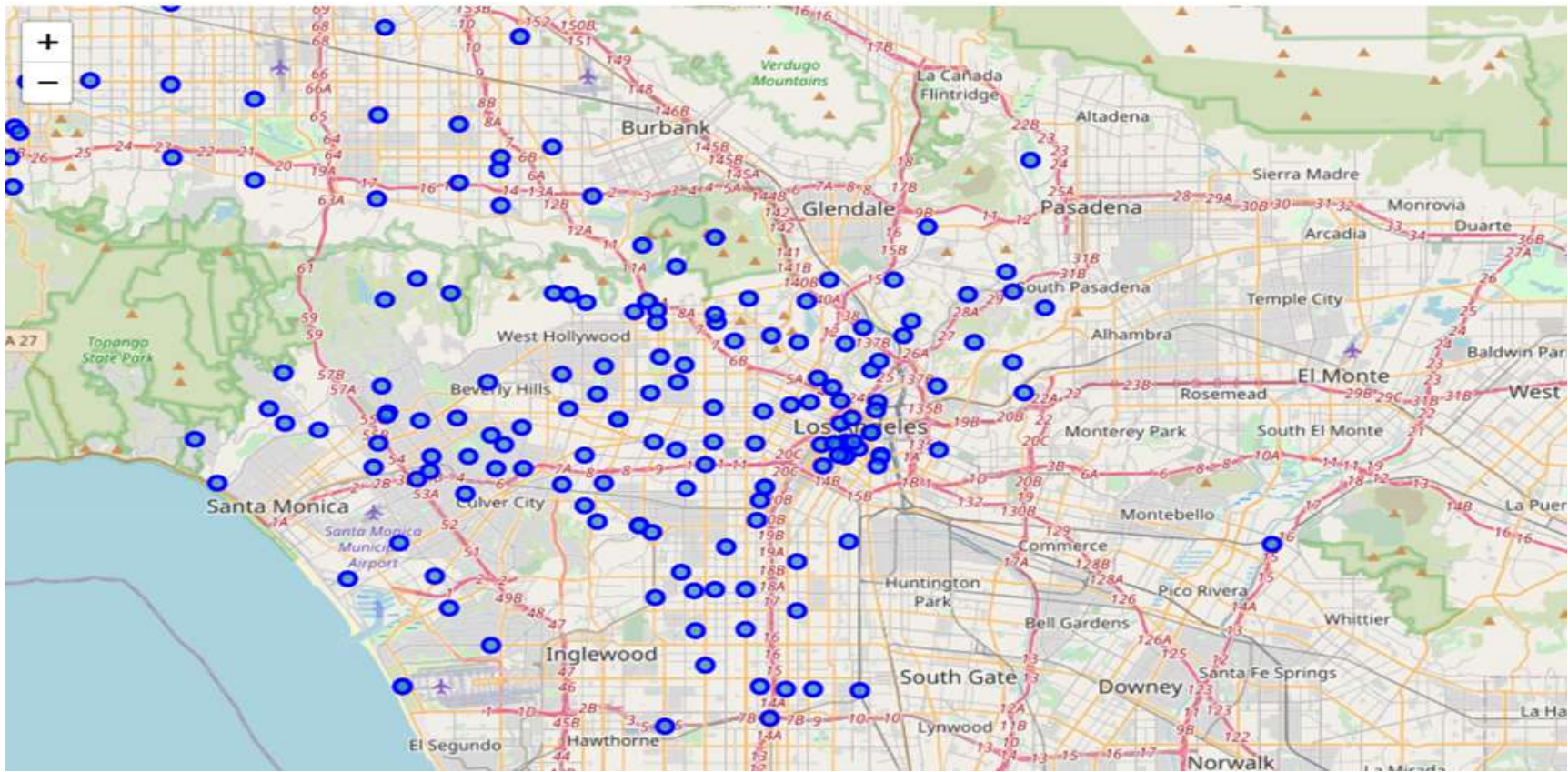
We also have maps of the different neighbourhoods in the two cities. The blue dots represent each neighbourhood.

## NEW YORK





# LOS ANGELES



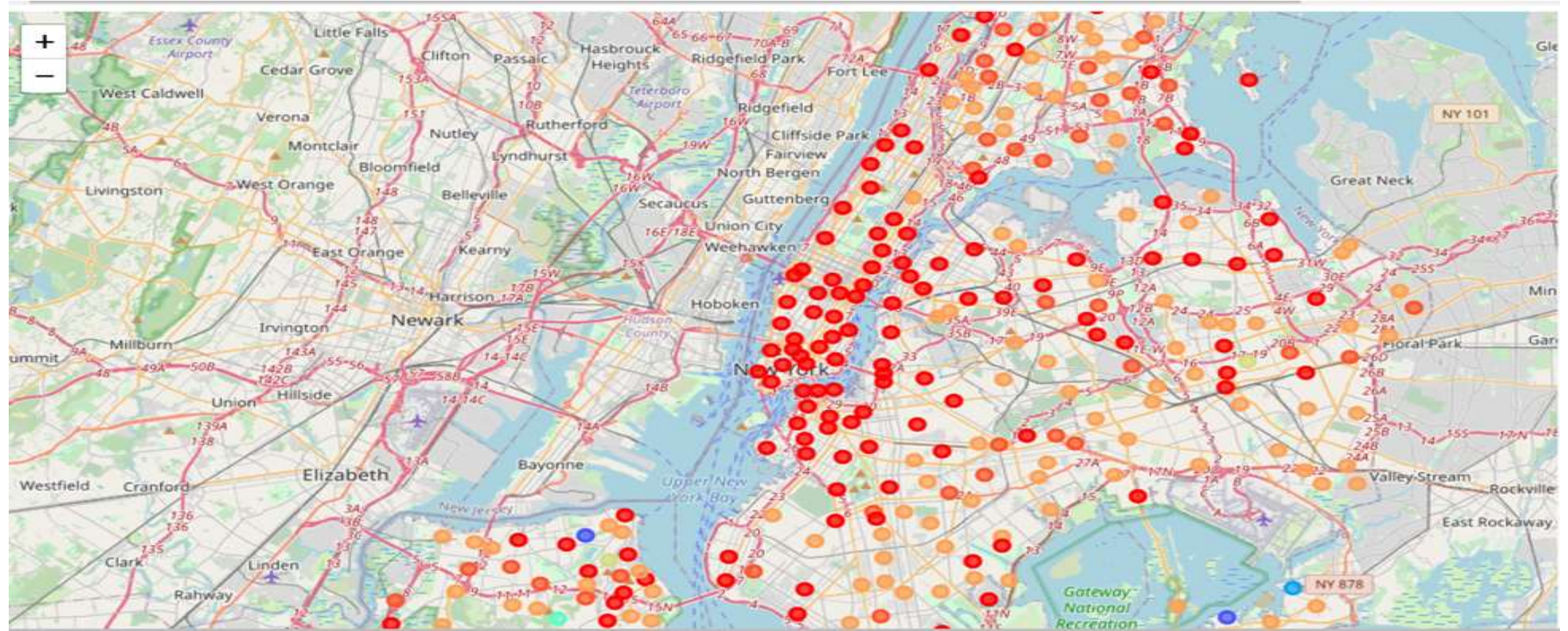
# K-means Clustering

---

In order to find similarities between the neighborhoods in these two cities, we use K-means to group the different neighbourhoods into clusters based on the venues retrieved by the Foursquare API. We group them into 10 clusters. We will visualize this using the maps. Each cluster is differentiated by the color of the neighborhood marker.

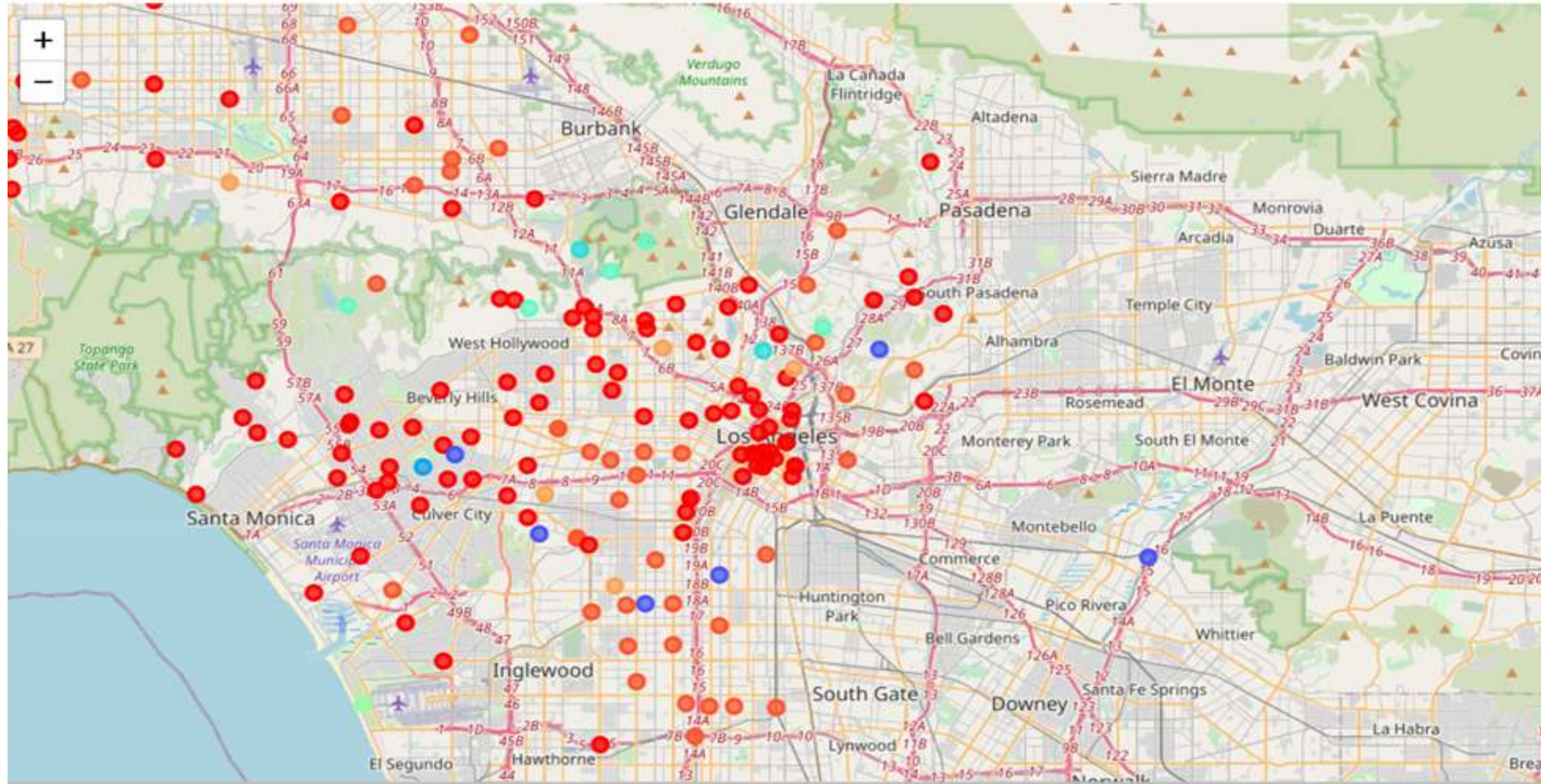


# NEW YORK





# LOS ANGELES



# Results

---

From our maps above, we see that majority of the neighborhoods in the both cities fall into the red and orange clusters. There are some outliers in both cities they fall under the same clusters too.

# Discussion

---

From our results above, we can determine that since the neighborhoods in these two cities fall under the same clusters, they aren't that different. Majority of the neighborhoods New York seem to have a similar brother in Los Angeles based on the venues around them. Hence, these two cities do not differ as much as people think. So if you want to choose between living in New York and Los Angeles, based on the venues, there isn't much difference, if you prefer the East Coast and some winter, pick New York, if you prefer the West Coast and warmer all year round, then Los Angeles is for you.

# Conclusion

---

In this study, I analyzed the similarities between the neighborhoods in New York City and the City of Los Angeles based on the venues in these neighborhoods retrieved using the Foursquare API. I used K-means to cluster these neighborhoods based on the venues into different clusters. I saw that majority of the neighborhoods fell into the same clusters; hence, this means that the neighborhoods in both cities aren't that different based on the venues in them. Hence, the two cities are actually quite similar.