**The Battle of Neighborhoods, Los Angeles, California**

Wimala Paramastri

April 1st, 2021

1. **Introduction**
   1. Background

California is one of many state located in the Western United States. It is the third most populous state among other states of United States. The most popular county in California is Los Angeles, as it is also being the most populous county in the United States, having around four million residents. Los Angeles is becoming a good place to set up business following the fact that it has the third-largest metropolitan economy in the world. One of the booming business in current trend is gym or sport centre. Moreover, Los Angeles is famous for people who is pursuing career in entertainment industry, which makes it important for them to take care of their body, beauty, and health.

* 1. Problem

Gym location being one of important indicator when people consider and choosing where to live. Apart from that, opening new gym means that we need to consider its preconditioned location too, like its neighborhood characteristic, population, and also nearby competitors. It's causing a big problem to answer, where is the most recommended neighborhood in Los Angeles to open gym?

1. **Data**
   1. Data Sources

To solve this problem it is important to gain proper data that are needed for analyzing characteristic of neighborhoods in Los Angeles. Some data related to Los Angeles which are needed such as:

1. Name of region and neighborhoods

Source: http://maps.latimes.com/neighborhoods/neighborhood/list/

1. Location of neighborhoods (Langitude and Latitude)

Used to derive LA map and get common venues around neighborhoods

Source: geocoders and http://s3-us-west-amazonaws.com/

boundaries.latimes.com/archive/1.0/boundary-set/la-county-neighborhoods-v5.geojson

1. Number of residents (population)

Used to get views and compare the population on every neighborhood

Source: http://maps.latimes.com/neighborhoods/population/total/

neighborhood/list/

1. Common venues around every city

Used to get views and compare characteristic on every neighborhood using venue's characteristic

Source: Foursquare

1. **Methodology**

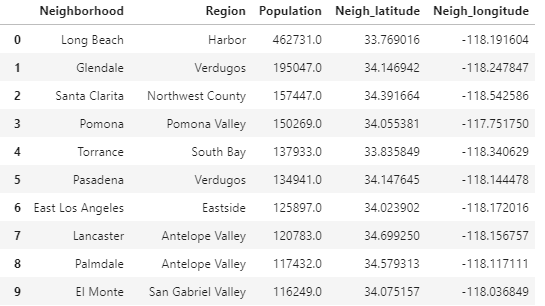
Method we will use are divided into 3 parts. First, importing all data needed and merging into a main dataframe by using beautifulsoup, json request, geocoders, and foursquare API. Second, rank common venues around every city for analyzing neighborhood characteristic using kMeans clustering method. Third, checking potential neighborhood by gym competitor nearby using foursquare API.

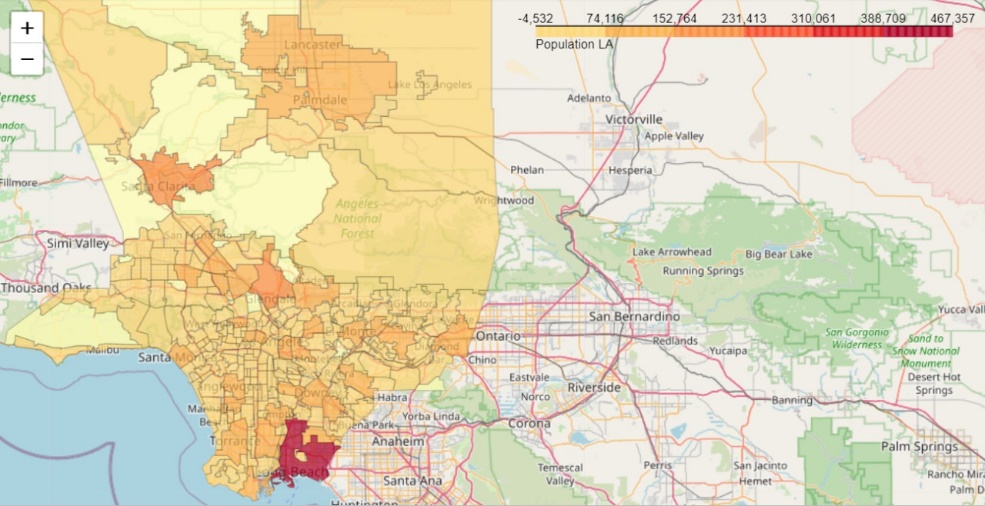
For importing data we use all the source stated before. We will collect all 272 name of neighborhoods around LA, get number of residents data for all neighborhoods, location (by latitude and longitude), and venues around every neighborhoods by foursquare data.

We explore the common venues around every neighborhood by making rank of the 10 most common venues. It can be done by grouping all venues and looking at the means of each venue category appearance in every neighborhood. We then cluster all the neighborhoods by its common venues by using kMeans=5.

Finally, we do a recheck of a potential neighborhood by checking locations of another gym as a business competitor within 3 kms from neighborhood centre.

1. **Analysis**
   1. Importing Main Data and Ranking Common Venues





After getting a total 260 main data of neighborhood, we then find all 4580 venues near each neighborhood with 244 total categories



* 1. Cluster Neighborhood by kMeans

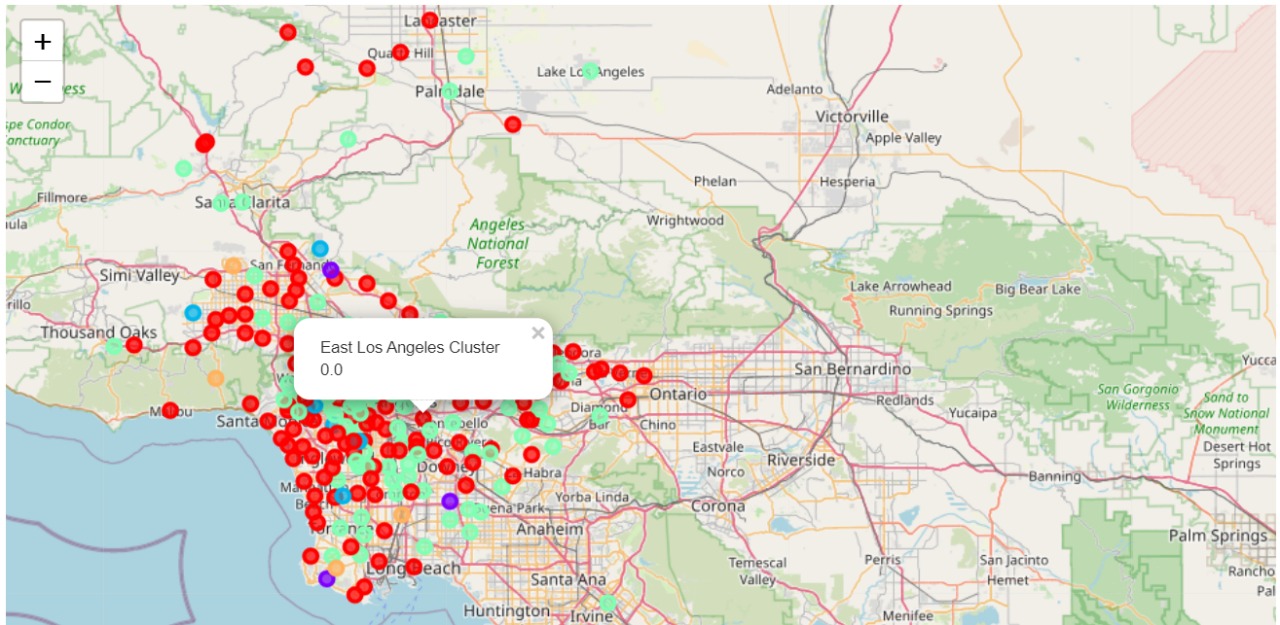
By using one hot encoding, we count the mean of every venue categories appearance in each neighborhood and then sort it from the highest.



After checking several number of kMeans we decided to use kMeans=5 because it has the most appropriate clusters and the possibility to be analyzed. This is resulted in below table.

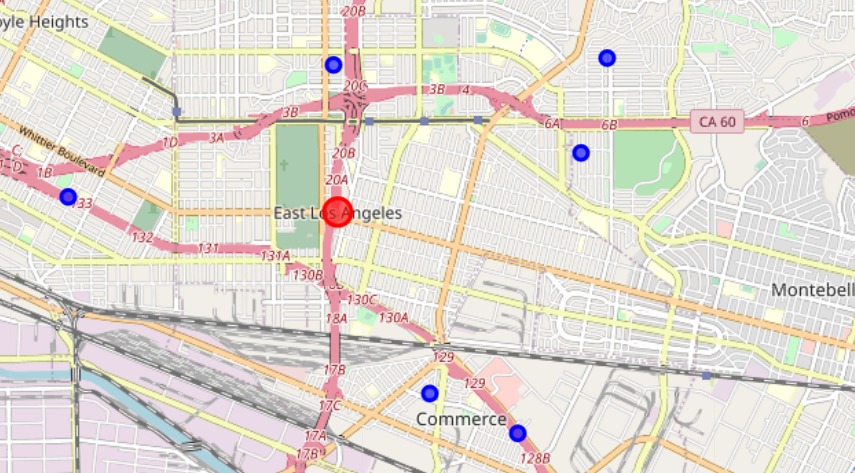
|  |  |
| --- | --- |
| Name of Cluster | Number of Neighborhood |
| Cluster 1 | 141 |
| Cluster 2 | 4 |
| Cluster 3 | 7 |
| Cluster 4 | 88 |
| Cluster 5 | 4 |





* 1. Nearby Competitors

We check numbers of gym competitors within 3kms and getting 6 Gym venues as below.





1. **Results and Discussion**

From clustering above, we can see that there are 3 big cluster (Cluster 1,3,4) and 2 smaller cluster (Cluster 2,5). In this case we will take a look at 3 biggest cluster that are Cluster 1, 3, and 4. Los Angeles is a big county, venues are almost fairly distributed causing some cluster looks a bit similar. Although their similarities, we can derive some differences too. Here are description for cluster analysis:

* Cluster1 : Big neighborhood with the most population among other cluster. Having various choice of restaurants as its common venue. Having lots of good place for socializing, for fun and entertainment such as, music venue, nightclub, event space, stadium, bowling alley, escape room and recreation center. Having various kind of place to enjoy food/drink/snack such as, restaurant, cafe, bakery, dessert, coffee/tea shops, and bar. Having various places to shop like mobile phone, grocery, cosmetic, clothing, hardware, and furniture. Overall it is a typical urban core neighborhood with a complete and diverse facilities in it.
* Cluster 3: Place for various activities outdoor. Main venues are parks. Having only some type of restaurants like english, ethiopian, or falafel restaurant. Good place for relaxing by doing yoga. Also having some entertainment venues and farm. Generally it looks like a suburban neighborhood.
* Cluster 4: Similar to cluster 1, which has a lot of various restaurant except dominated with fastfood. Lot of venues for activity as well such as bowling, pool, theatre, yoga, and event space. There are lots of place to shop daily needs and they are distributed everywhere such as grocery store, convenience store, supermarket and farmer's market. Overall, it's a type of residential place that is good for living.

While it is important to open Gym at a good residential type of neighborhood, it is also important to target correct market that will going to use the gym service, like Cluster 1 type of neighborhood. It is a type of urban core neighborhood which has a lot of people reside and fit our target which is people who would like to pursue career in entertainment industry. Gym members are likely belong to a facility that is located near where they live. In that case, we need to find one best neighborhood in the cluster 1 which has big enough population and also having supportive common venues, such as health related things. In this case, we choose East Los Angeles as the most recommended place to open Gym. It's because it has big enough population among its cluster. Also, looking at its characteristic of neighborhood, East Los Angeles has a lot of common facilities/venues which is in line with health caring such as Bakery, Sporting Goods Shop, Track, Sandwich Place, and Farm. It also doesn't have unhealthy venues as its common venues such as Bar, Liquor Store, and Fast Food Restaurant, even though other have them. Lastly we check about the competitors and Gym doesn't appear in the 10th most common venues list. This means that Gym is not categorized as common facilities in this neighborhood yet. We did a re-check to make sure that there are only few competitors nearby and it shows that there are only 6 gym within 3kms from the neighborhood centre which is not too much considering its population number. From this, we can make a conclusion that East Los Angeles is a good place to open Gym.

1. **Conclusion**

Purpose of this project was to make a recommendation of a location in Los Angeles County to open gym. In order to narrow down the location we pick a neighborhood which has a specific characteristic, such as number of population and the neighborhood type itself. By collecting data from all venues and checking their location, we can conclude type and characteristic of the neighborhood which has some venues close to it. In this case, we got 3 main clusters derived from its most common venue nearby. Surprisingly, it is hard to distinguish one to another cluster because of its similarity, that might be a typical big city character as facilities are fairly distributed. Cluster analysis shows that cluster 1 is a good potential location for opening new business because of its characteristic of urban core neighborhood that might highly attract more people within the project target to live there. East Los Angeles is one of a neighborhood on cluster 1 that has big population but not considered too crowded. By checking its most common venues, we can see that East Los Angeles has supportive facilities for keeping up with health related things, such as sport shop, healthy food, and healthy market. After did a recheck, it can also be seen that there are only few gym competitors nearby. Finally, we can recommend that East Los Angeles is a good location to open gym based on current environment, also for future, considering the good characteristic of neighborhood itself.