BATTLE OF THE NEIGHBORHOODS

REPORT

By,

Dhanush S Eshwar

1. Introduction/Business Problem:

Background-

Los Angeles is one of the most popular city in United States because of its renowned museums, unique hotels and vibrant multicultural neighborhoods. Los Angeles has a diverse economy and hosts businesses in a broad range of professional and cultural fields. It also has the busiest container port in the entire Americas. A global city, it has been ranked 6th in the Global Cities Index and 9th in the Global Economic Power Index. The Los Angeles metropolitan area also has a gross metropolitan product of $1.0 trillion (as of 2017), making it the third-largest city by GDP in the world, after the Tokyo and New York City metropolitan areas. Los Angeles hosted the 1932 and 1984 Summer Olympics and will host the 2028 Summer Olympics.

Being a popular city means it attracts thousands of tourists every year and thus tourism related business would be successful in L.A.

Problem Description-

A travel agency wants to create travel packages by segregating different neighborhoods of the city so that they can recommend packages to their customers based on their interest and preferences such as –

- Beaches and water sports
- City Night life
- Movie Studios
- Hiking Trail etc

And also create a cluster map so that the customers can visualize it easily.

2. Data:

Dataset of Los Angeles city-

The data of Los Angeles city such as neighborhoods and their GPS co-ordinates was taken from Wikipedia.
https://en.wikipedia.org/wiki/Los_Angeles_County_Board_of_Supervisors
From the corresponding latitude and longitude data I used Foursquare application to get details on venues in and around the neighborhood.

	DISTRICT	NEIGHBOURHOOD	LATITUDE	LONGITUDE
0	FIRST	BOYLE HEIGHTS	34.033888	-118.204444
1	FIRST	CHINATOWN	34.062880	-118.237890
2	FIRST	CIVIC CENTRE	34.054130	-118.244600
3	FIRST	DOWNTOWN L.A	34.057000	-118.238000
4	FIRST	EAGLE ROCK	34.138800	-118.213050

- Information on venues near the neighborhoods will be gathered from Foursquare.
  I used the acquired location data to explore the nearby important and unique venues from the Foursquare API. Using the "explore" option, I look for top 10 venues in 10 km radius for each neighborhood and I get "Venue", "Venue Latitude", "Venue Longitude" and "Venue Category".

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BOYLE HEIGHTS	34.033888	-118.204444	Mariscos El Jato	34.037184	-118.204162	Seafood Restaurant
1	BOYLE HEIGHTS	34.033888	-118.204444	La Mascota Bakery	34.032518	-118.211395	Bakery
2	BOYLE HEIGHTS	34.033888	-118.204444	Yum Yum Donuts	34.027976	-118.201088	Donut Shop
3	BOYLE HEIGHTS	34.033888	-118.204444	Ramirez Liquor & Kegs Delivery	34.034776	-118.215187	Liquor Store
4	BOYLE HEIGHTS	34.033888	-118.204444	Artic Hot Spot	34.040059	-118.210238	Café

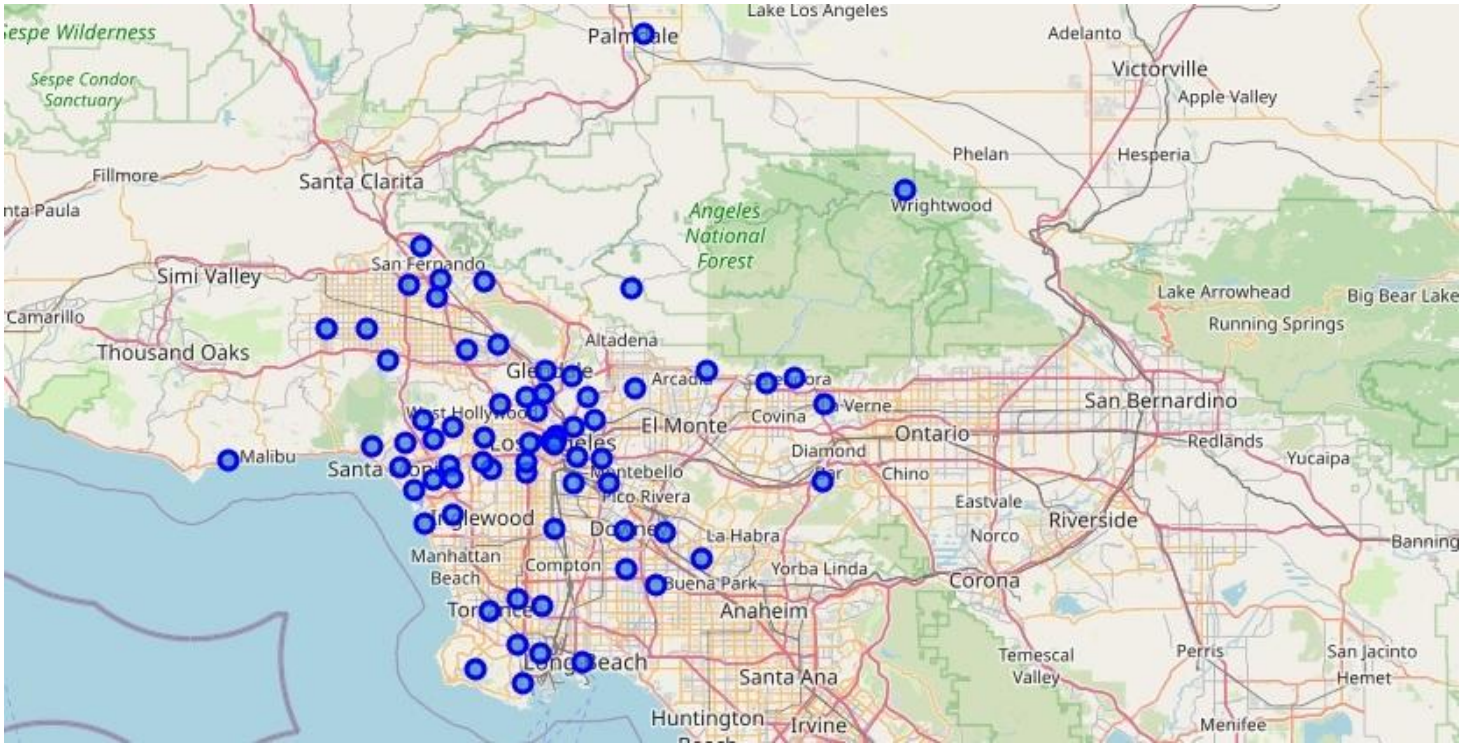
3. Methodology:

Business Understanding-

Our main objective is to segregate different neighborhoods of Los Angeles city with similar venues.

Exploratory Data Analysis-

- **Los Angeles City Neighborhood Visualization**
  Using the geopy and folium libraries we can create Los Angeles City Map with the neighborhoods superimposed on top as shown in the below figure.



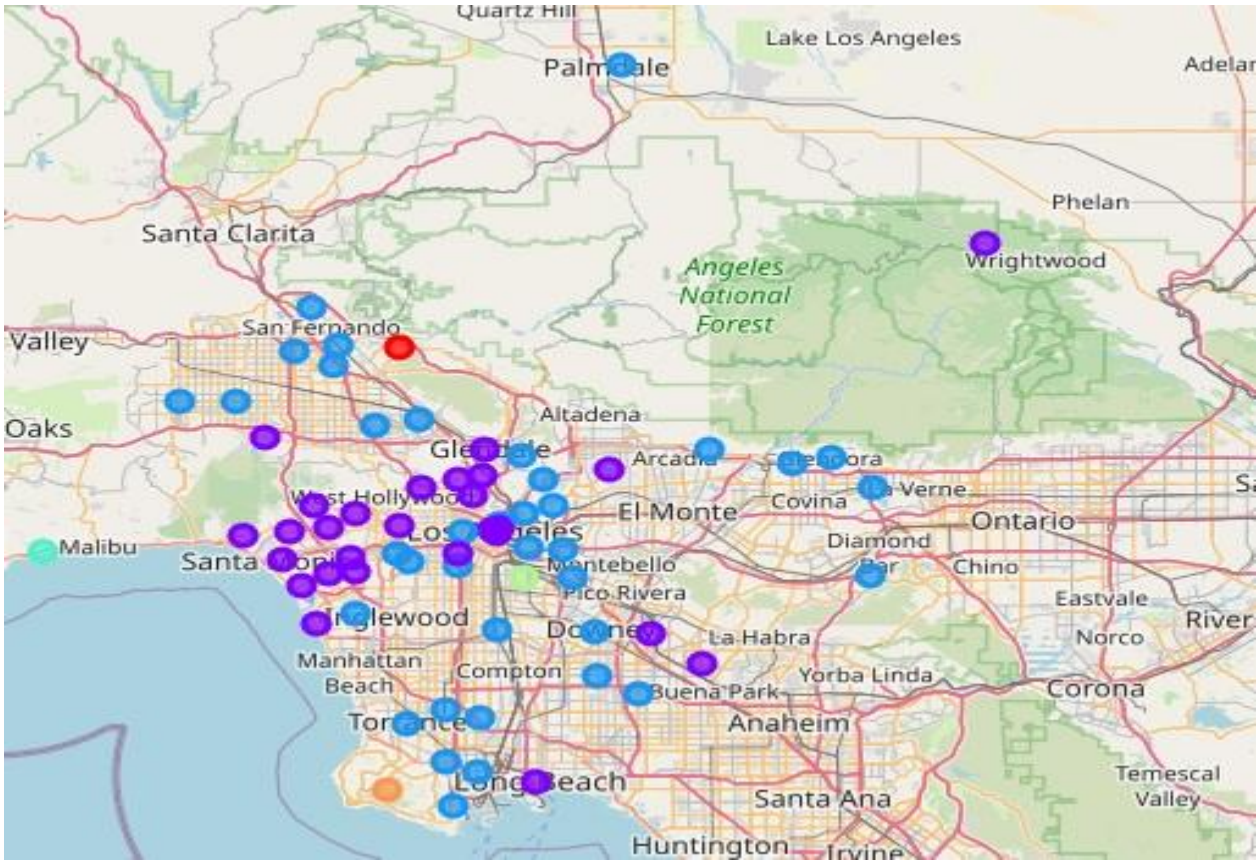
- **Foursquare API-**

  Using Foursquare we get all the venue details near the neighborhood and then we get a total of 3964 venues by setting the limits as 100 venues maximum per neighborhood and radius of 10 km.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
ARLETA	25	25	25	25	25	25
ARTESIA	93	93	93	93	93	93
ATWATER VILLAGE	78	78	78	78	78	78
AZUSA	41	41	41	41	41	41
BELL AIR	9	9	9	9	9	9
BELL FLOWER	55	55	55	55	55	55
BEVERLY HILLS	100	100	100	100	100	100
BIG PINES	5	5	5	5	5	5
BOYLE HEIGHTS	42	42	42	42	42	42
BRENTWOOD	78	78	78	78	78	78
BURBANK	76	76	76	76	76	76
CANOGA PARK	85	85	85	85	85	85
CARSON	66	66	66	66	66	66
CHINATOWN	71	71	71	71	71	71

- **Neighborhoods K-Means clustering-**

This Machine learning technique is used for segregating different neighborhoods based on their unique venues in their neighborhood. The below figure is the cluster map –



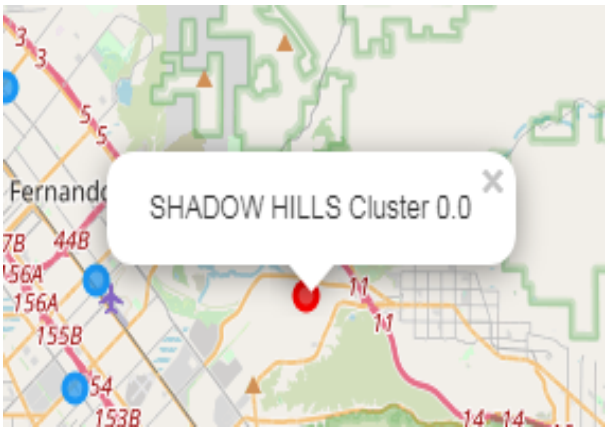
4. Results:

By using the methodology discussed in the above section we get satisfactory results when there is six ‘’K” centroids or mean centers. The below parts show the different clusters obtained and inferences made by the obtained result.

Cluster Labels	NEIGHBOURHOOD
0.0	1
1.0	27
2.0	37
3.0	1
4.0	1
5.0	1

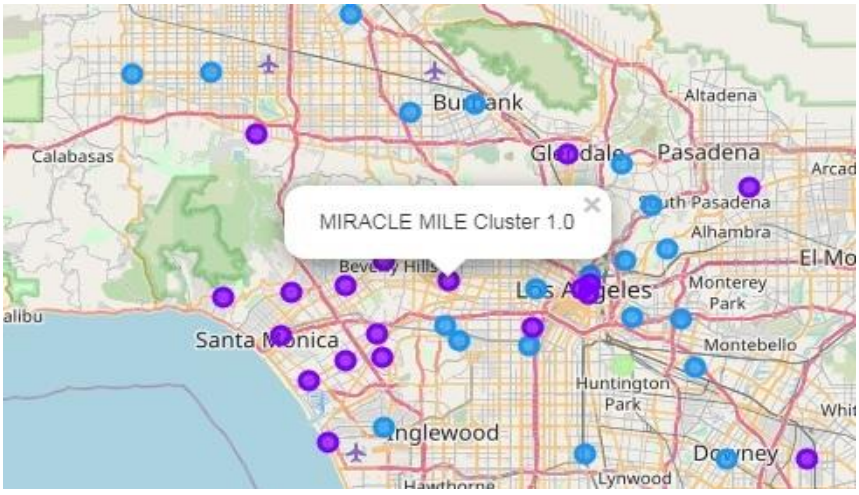
CLUSTER 0-

This cluster has only one entry that is shadow hills, and the main venues are stables and movie studios.



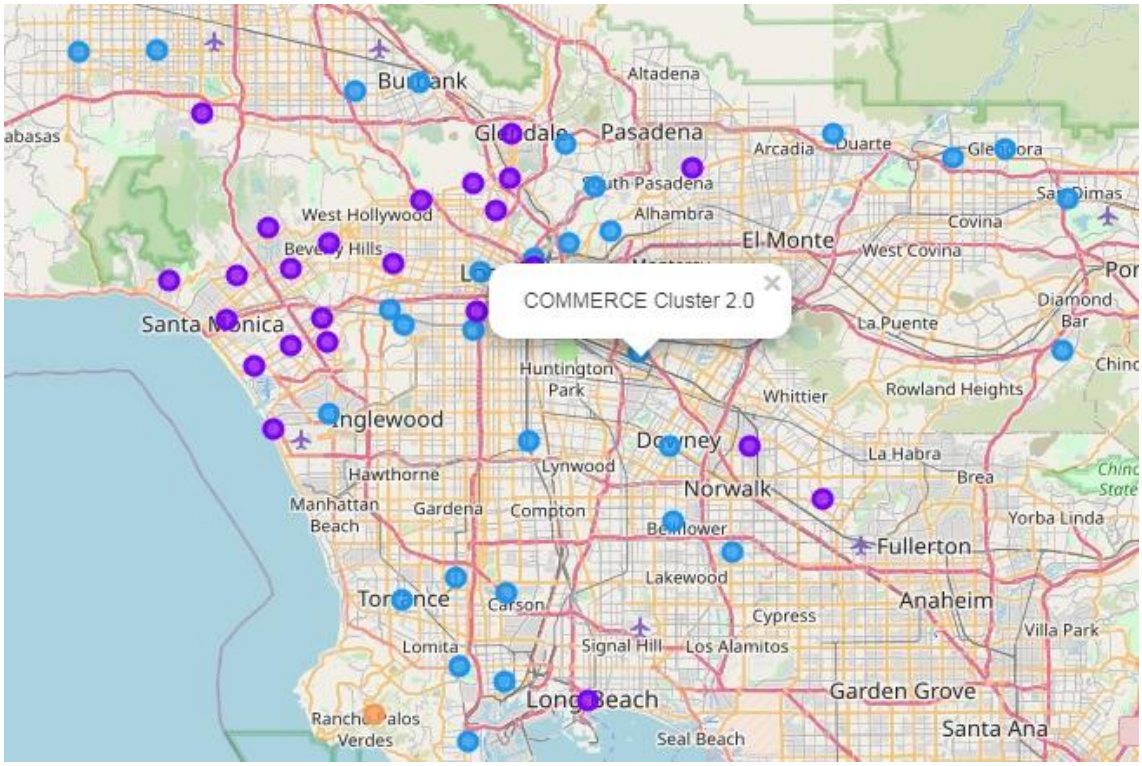
CLUSTER 1-

This cluster has 27 neighborhoods with common venues as beaches and scenic spots along with glamorous part Los Angeles such as Downtown L.A, Long beach, and Santa Monica.



CLUSTER 2-

This cluster has 37 neighborhoods; most of the common venues are restaurants, unique hotels, art gallery, museums etc and all of them are in the city centre.

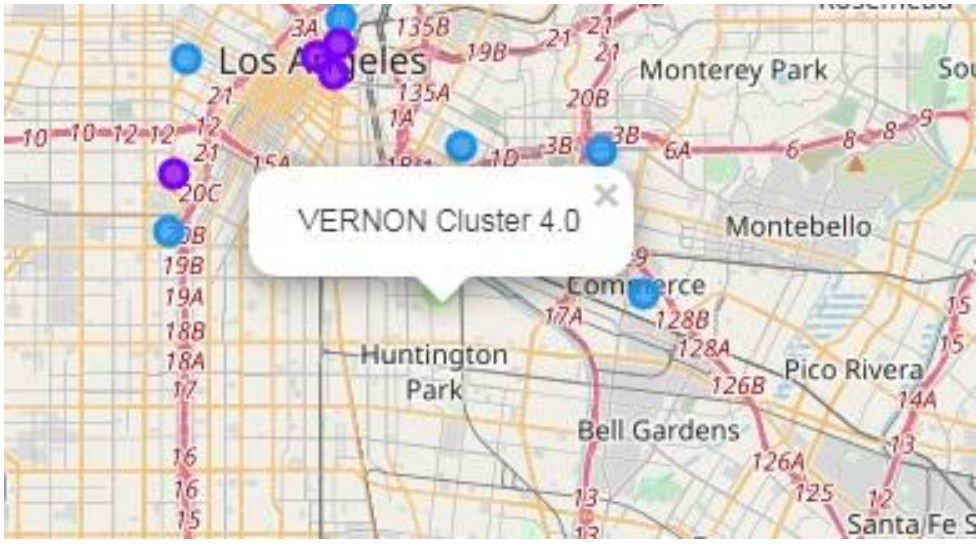


CLUSTER 3-

Malibu is the only neighborhood in the cluster 3 because it's the iconic destination for surfing, water sports and scenic beach.

CLUSTER 4-

Vernon is the only neighborhood in this cluster and it's in city centre with no unique attractive spots for tourists however this place is good scouting for beginning a business venture or looking for an office.



CLUSTER 5-

Rancho Palos Verdes is the only neighborhood in cluster 5 and it is known for its nature park and hiking trails also located on a scenic coastline.

5. Discussions:

There is further scope for this project if price or cost data for transportation, park entry fees etc is available and using that data, different economical packages can be made based on customer interests. However I have done clustering only based on tourist choice of spots or common venues between different neighborhoods.

6. Conclusions:

I would like to conclude by saying that people who have interests in movie studios, beaches and stables can go to cluster 0 and 1, people who have interests in art,music, night life will prefer cluster 2.
People who like water sports like surfing can visit the malibu which is the only neighborhood in cluster 3 and also people who love nature trails along a coast of a beatiful beach can visit the neighborhood in cluster 5.