Los Angeles Neighbourhood analysis for setting up for Indian cuisine

1. Introduction

Los Angles is very vibrant, and it is multi-cultural and famous for Hollywood entertainment industry. Los Angeles is known for its Mediterranean climate and its sprawling metropolis. Los Angeles lies in a basin in Southern California, adjacent to the Pacific Ocean, with mountains as high as 10,000 feet (3,000 m), and deserts.

I had chance to live in Rosemead, Los Angeles for more than year. Hence, I have fair idea of the LA neighbourhoods and food chains in greater LA. This data science project is intended to analyse various communities, population, area income limits, age and ethnic groups and pre-existing restaurants and come up with five possible location recommendations for starting Indian cuisine restaurant.

1.1 Problem statement

This data science project is expected to recommend best possible five locations to start Indian cuisine by studying the greater LA demographics, neighbourhood communities, income and various other factors including existing restaurants and their popularity.

The greater LA is one of the largest cities in united states, and there are different community-based neighbourhoods, it is important to apply data science and arrive at the best possible location for the success of the business.

1.2 Data Acquisition

There are various organizations provides vital information about the Greater Los Angeles demographics including income limits, population, age group and current restaurants with geocodes in neighbourhoods and their corresponding ratings

By applying statistical analysis and ML techniques, it is obvious that meaningful insights can be built and arrive at the best locations for stating Indian cuisine.

1.3 Data Sets

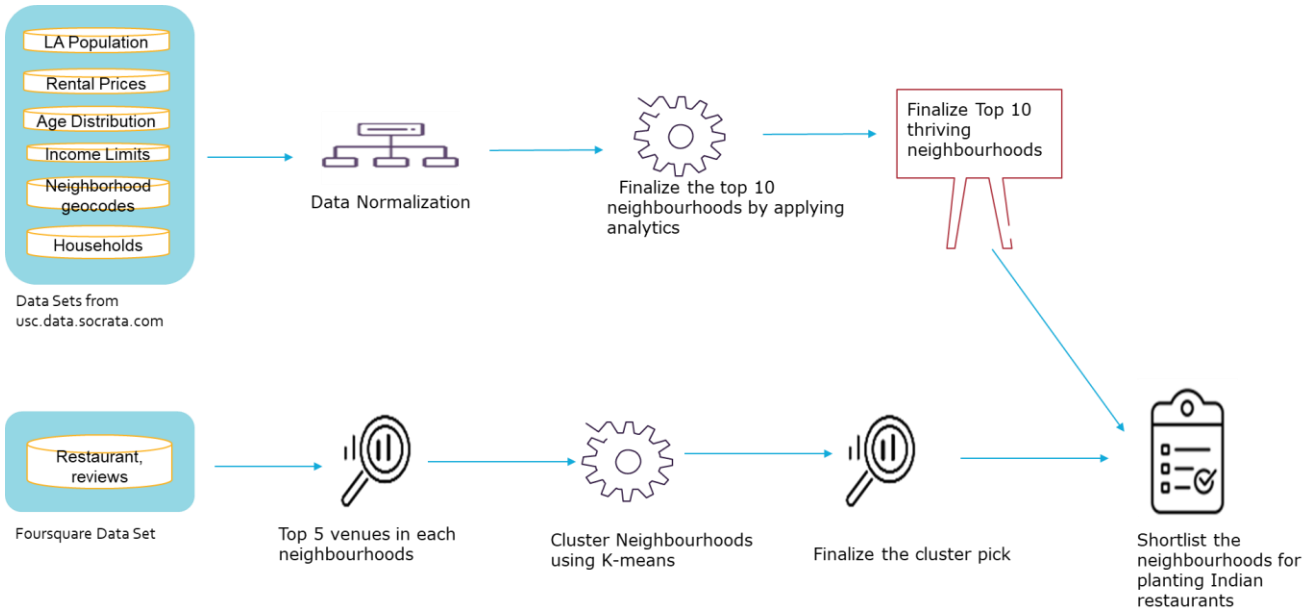
Data Source	Details
https://usc.data.socrata.com/ NEIGHBORHOOD DATA FOR SOCIAL CHANGE	✓ Total population – For each neighbourhood (from 2010 till 2018) ✓ Rental price for each neighbourhood ✓ Age Distribution ✓ Area income limits ✓ Census track location for each neighbourhood ✓ Households (Single/Family/size)
Foursquare FOURSQUARE	✓ Existing restaurants, ratings and reviews
Google Geocode API	Leverage google Geocode API for getting the geolocations of the neighbourhood (only if needed, most of the geocode details are already available provided by https://usc.data.socrata.com/

Table1: Data Sources

1.4 Approach

Following depicted approach details out two stage process,

1. Finalize the Neighbourhoods from the data sets https://usc.data.socrata.com/
2. Obtain the top 5 venues for each neighbourhood and apply K-means cluster and finalize the neighbourhoods for planting Indian restaurants.



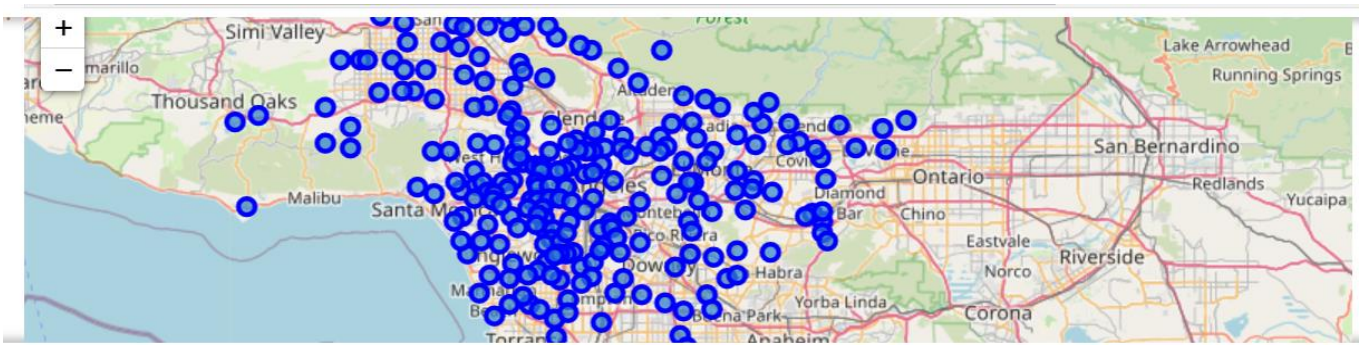
1.5 Data Preparation and Analysis

1.5.1. Census Analysis

Neighbourhood Data For Social Change Organization has provided Los Angeles neighbourhoods and Geo locations. This data set contained 2345 rows for 265 LA neighbourhoods. There were multiple geo codes for each neighbourhood at each street is present in the data set. One geo code per neighbourhood is picked up by removing duplicates for analysis

	GEOID	Tract	Tract Number	Neighborhood	Location	Latitude	Longitude
0	1400000US06037101110	Census Tract 1011.10, Los Angeles County, Cali...	101110	Tujunga	(34.2595555, -118.293602)	34.259555	-118.293602
6	1400000US06037102103	Census Tract 1021.03, Los Angeles County, Cali...	102103	Shadow Hills	(34.224155, -118.354339)	34.224155	-118.354339
8	1400000US06037102105	Census Tract 1021.05, Los Angeles County, Cali...	102105	Sun Valley	(34.210852, -118.3480495)	34.210852	-118.348050
10	1400000US06037103101	Census Tract 1031.01, Los Angeles County, Cali...	103101	Sunland	(34.274431, -118.30714)	34.274431	-118.307140
12	1400000US06037103200	Census Tract 1032, Los Angeles County, California	103200	Lake View Terrace	(34.2745565, -118.3439025)	34.274557	-118.343902

Table 2: LA neighbourhoods and geo codes



Map 1: LA neighbourhood geo locations in google maps using Folium

1.5.2 Total Population

Total population data set contained the population for each neighbourhood from 2010 to 2018. This data set is very vital for our analysis. Pandas pivot table function is leveraged to build yearly population trend analysis for each neighbourhood.

The below population trends display Long Beach has the highest population among all neighbourhoods. For this analysis top10 populated neighbourhoods are considered for future analysis.

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018
Neighborhood									
Long Beach	464491.0	465074.0	466201.0	468083.0	471080.0	473084.0	471985.0	472459.0	470990.0
Glendale	191089.0	190718.0	191166.0	193381.0	195380.0	196984.0	197895.0	199750.0	200372.0
Santa Clarita	179009.0	181304.0	184743.0	186384.0	188369.0	190484.0	191708.0	188889.0	190304.0
Palmdale	151700.0	154477.0	157437.0	159248.0	161374.0	162318.0	162126.0	163077.0	162358.0
Lancaster	149533.0	150851.0	153858.0	154872.0	155839.0	156540.0	156585.0	156735.0	155605.0
Pomona	149201.0	149368.0	149928.0	150448.0	151825.0	152022.0	152353.0	152802.0	152823.0
Torrance	143951.0	144622.0	145443.0	146187.0	147181.0	147589.0	147307.0	147190.0	146392.0
Pasadena	138295.0	138719.0	138961.0	139878.0	140781.0	141632.0	141931.0	143113.0	143173.0
East Los Angeles	120507.0	121422.0	124140.0	124858.0	124869.0	122557.0	121361.0	120896.0	117348.0
El Monte	114293.0	114203.0	114161.0	114714.0	115485.0	115948.0	115547.0	116117.0	116308.0

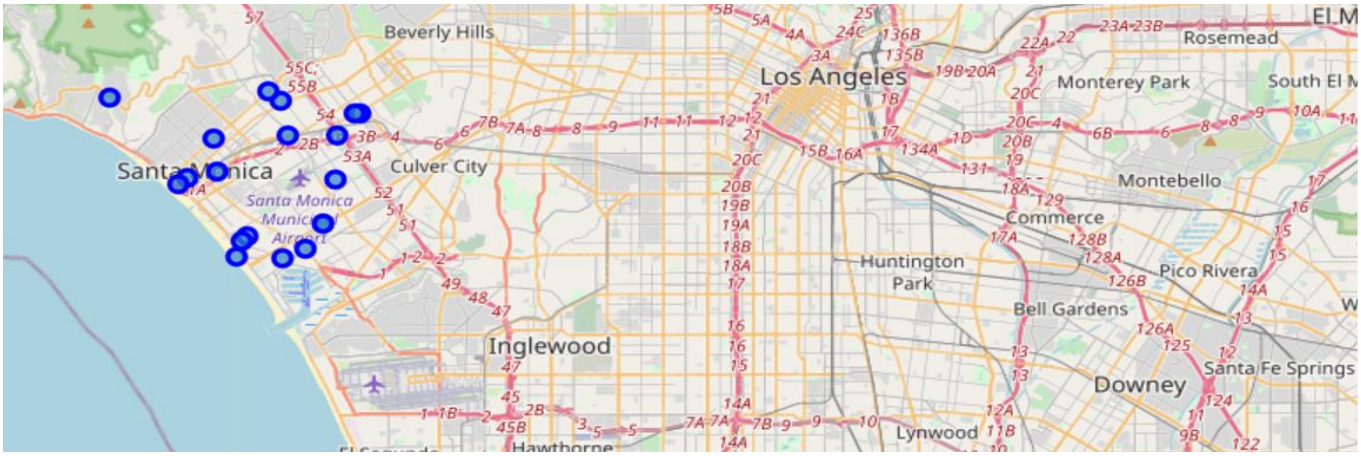
Fig 1: Highly populated LA neighbourhood trends

1.5.3 Foursquare – Existing Indian restaurants

Foursquare API was leveraged to find out how many Indian cuisines exists in LA neighbourhoods. We started our analysis in Santa Monica, Los Angeles.

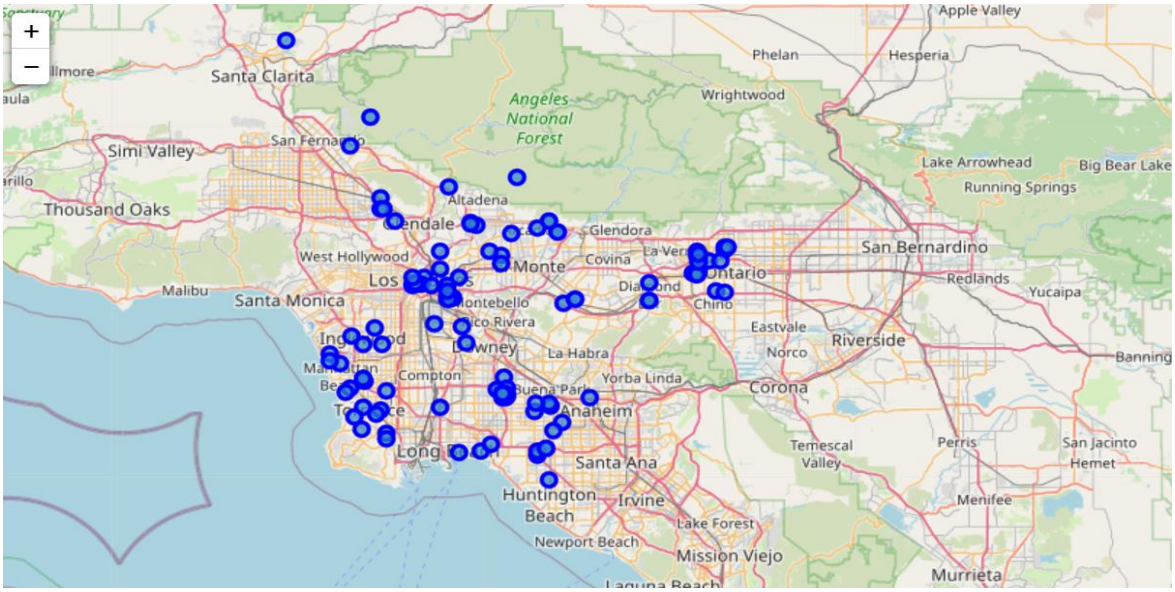
Following key paraments were applied to fetch the Indian restaurants in Santa Monica.

```
address = 'Santa Monica, Los Angeles'
search_query = 'Indian'
radius = 5000
```



Map 2: Foursquare responded **19 Indian restaurants** in Santa Monica neighbourhood

The same analysis was extended to fetch all the Indian restaurants in top 10 highly populated neighbourhoods. Foursquare API reported 123 Indian restaurants in top 10 highly populated neighbourhoods.



Map 3: Foursquare responded **123 Indian restaurants** in in top 10 highly populated neighbourhoods.

1.5.4 Fetch all the Los Angeles venues from Foursquare

Within 500 meters of radius all the venues in LA neighbourhoods are obtained using Foursquare API. Foursquare returned 2151 venues in entire LA neighbourhoods.

	Unnamed: 0	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	0	Tujunga	34.259555	-118.293602	Commercial Cool Roof Systems	34.261556	-118.291594	Construction & Landscaping
1	1	Tujunga	34.259555	-118.293602	Canyon Market & Liquor	34.263107	-118.294111	Liquor Store
2	2	Tujunga	34.259555	-118.293602	Burlap and Silk	34.256287	-118.290756	Bridal Shop
3	3	Sun Valley	34.210852	-118.348050	U-Haul Moving & Storage of Sun Valley	34.211351	-118.349032	Storage Facility
4	4	Sun Valley	34.210852	-118.348050	Shell	34.210899	-118.348384	Gas Station

Table 3: Los Angeles venues for each neighbourhood

1.5.5 Data Pre-processing

one hot encoding is applied on venue category, next group rows by neighbourhood and by taking the mean of the frequency of occurrence of each category

[42]:

	Neighborhood	ATM	Adult Boutique	Airport Terminal	American Restaurant	Andhra Restaurant	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Entertai
0	Adams-Normandie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
1	Agoura Hills	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
2	Alhambra	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
3	Alondra Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
4	Altadena	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.(
5	Arleta	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.(

1.5.6 Find out neighbourhood along with the top 5 most common venues

[47]:

	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	Adams-Normandie	Playground	Taco Place	Gas Station	Indie Theater	Escape Room	Ethiopian Restaurant	Event Service	Event Space	Fabric Shop	Farm
1	Agoura Hills	Fast Food Restaurant	Pizza Place	Dry Cleaner	Ramen Restaurant	French Restaurant	Car Wash	Park	Farm	Escape Room	Ethiopian Restaurant
2	Alhambra	Bank	Fast Food Restaurant	Park	Grocery Store	Bubble Tea Shop	Market	Shoe Store	Tea Room	Optical Shop	Automotive Shop
3	Alondra Park	Football Stadium	Baseball Field	Park	Yoga Studio	Fast Food Restaurant	Event Service	Event Space	Fabric Shop	Farm	Farmers Market
4	Altadena	Food	Basketball Court	Coffee Shop	Yoga Studio	Fast Food Restaurant	Event Service	Event Space	Fabric Shop	Farm	Farmers Market

1.6 Unsupervised ML – K Means Clustering

I applied **K-means clustering** algorithm to cluster the neighbourhoods

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.

Elbow curve method helped me to determine number of clusters, finalized the number of clusters to 4.

- Finalizing the number of clusters is the key
- Multiple k-means algorithms implemented
- Elbow curve approach used to finalize the number of clusters

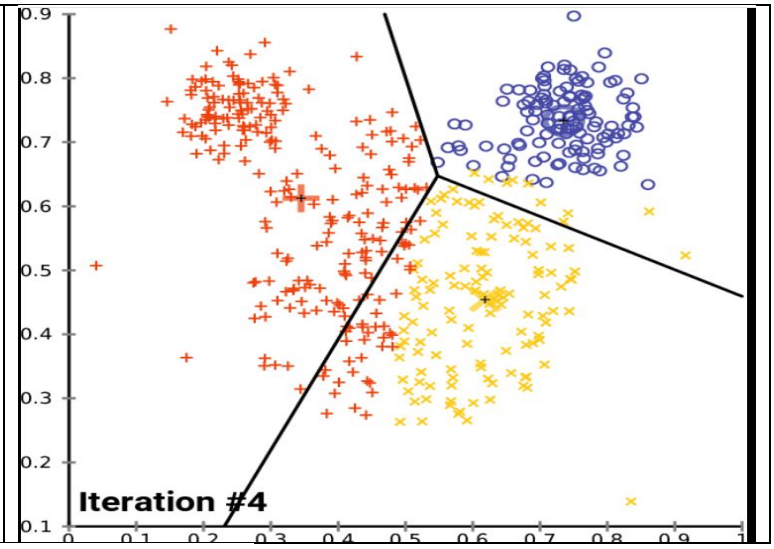


Fig2: K-Means clustering approach

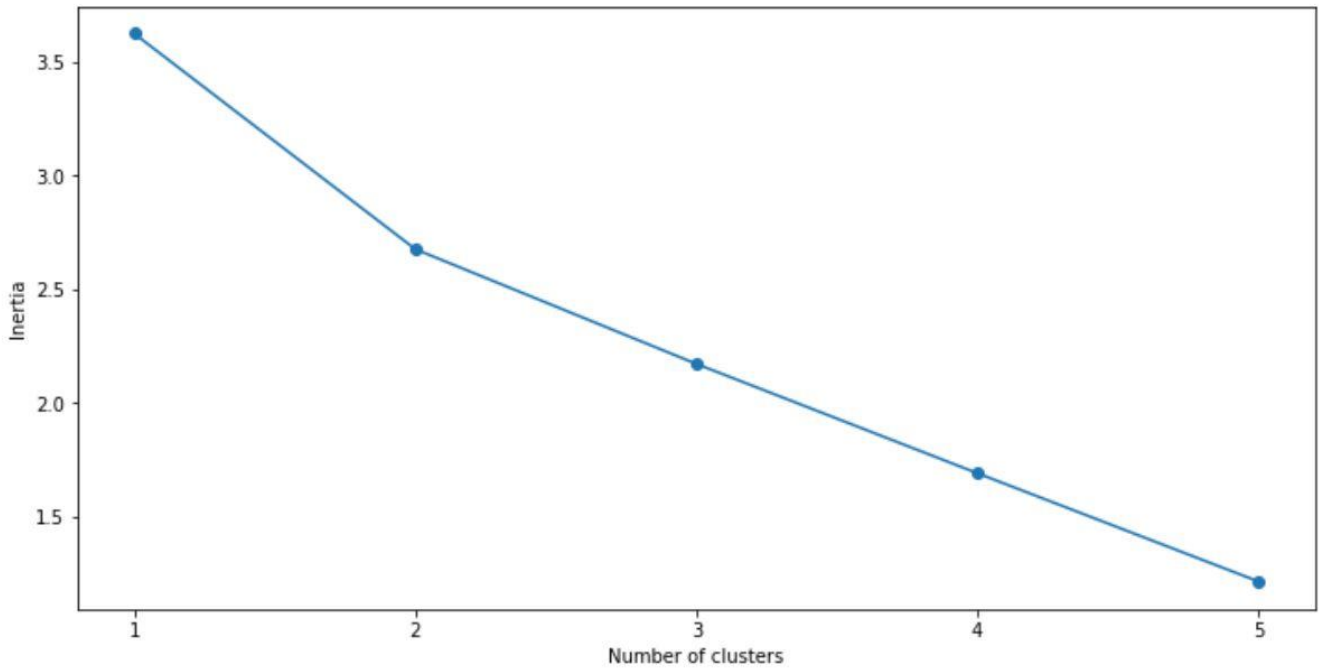
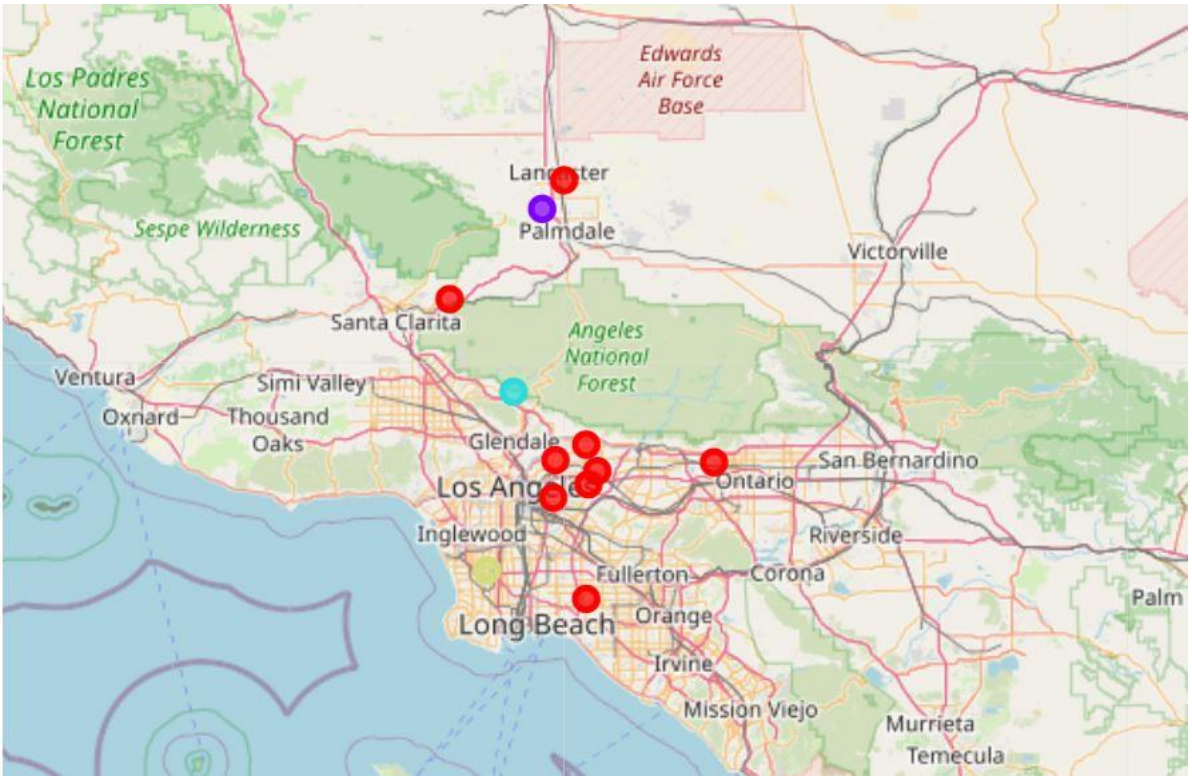


Fig3: • Elbow curve approach to finalize the number of clusters

K-means clustering was applied on pandas data frame which has highly populated neighbourhoods and their corresponding top 10 most popular venues. Cluster wise neighbourhoods are depicted in the below google maps



Map 4: Neighbourhood clusters in LA

Cluster Analysis

Cluster 0 has, totally nine neighbourhoods, all the cluster 0 neighbourhoods have at least 3 "restaurants" as most common venues.

Clusters 3,2,1 having one neighbourhood each, none of these neighbourhoods having restaurants as common venues. Most of common venues are formers market, parks and Yoga studio as common venues. Hence we need to drop these clusters. People are not preferring to go to restaurants in these neighbourhoods.

1.4 Conclusion

Our first part of the analysis revealed there are 14 neighbourhoods, which are highly populated.

once we cluster these 14 neighbourhoods, we are able to narrow down to 9 neighbourhoods which are having most common venue as "Restaurants" and are highly populated.

Highly populated neighbourhoods and having "Restaurants" as common venues are

East Los Angeles, East Pasadena, El Monte, Glendale, Lancaster, Long Beach, Northwest Palmdale, Pomona, Santa Clarita, South El Monte, South Pasadena, Torrance

Since we are interested in setting up Indian restaurants, our Foursquare, API revealed following top 4 neighbourhoods having highest number of Indian restaurants

Long Beach , Pomona , Torrance , Pasadena

Hence, we recommend these are the best suited top 4 neighbourhoods which are highly populated, and having restaurants as top venues and having most number of Indian Restaurants

1.5 References

1. Foursquare
2. Google geolocation API
3. https://usc.data.socrata.com/ for Lose Angeles Census data