# Los Angeles Neighbourhood analysis for setting up for Indian cuisine

# Neighborhood Prediction analysis is key to the success of setting up food chains

- There are 200+ Indian cosines in in Los Angeles neighbourhoods, it is important that proper  analysis done on the neighbourhoods before planting new restaurants

- Foursquare API's provides various venues in the LA neighbourhoods this can be critical data set for our analysis

- https://usc.data.socrata.com/  site provides vital information about the neighbourhood populations, demographics etc. This data set is also very vital for our analysis

# Data Sets

| Data Source | Details |
| --- | --- |
| https://usc.data.socrata.com/   | ✓ Total population – For each neighbourhood (from 2010 till 2018)<br>✓ Rental price for each neighbourhood<br>✓ Age Distribution<br>✓ Area income limits<br>✓ Census track location for each neighbourhood<br>✓ Households (Single/Family/size) |
| **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/ |

# Approach

LA Population

Rental Prices

Age Distribution

Income Limits

Neighborhood geocodes

Households

Data Sets from usc.data.socrata.com

Data Normalization

Finalize the top 10 neighbourhoods by applying analytics

Finalize Top 10 thriving neighbourhoods

Restaurant, reviews

Foursquare Data Set

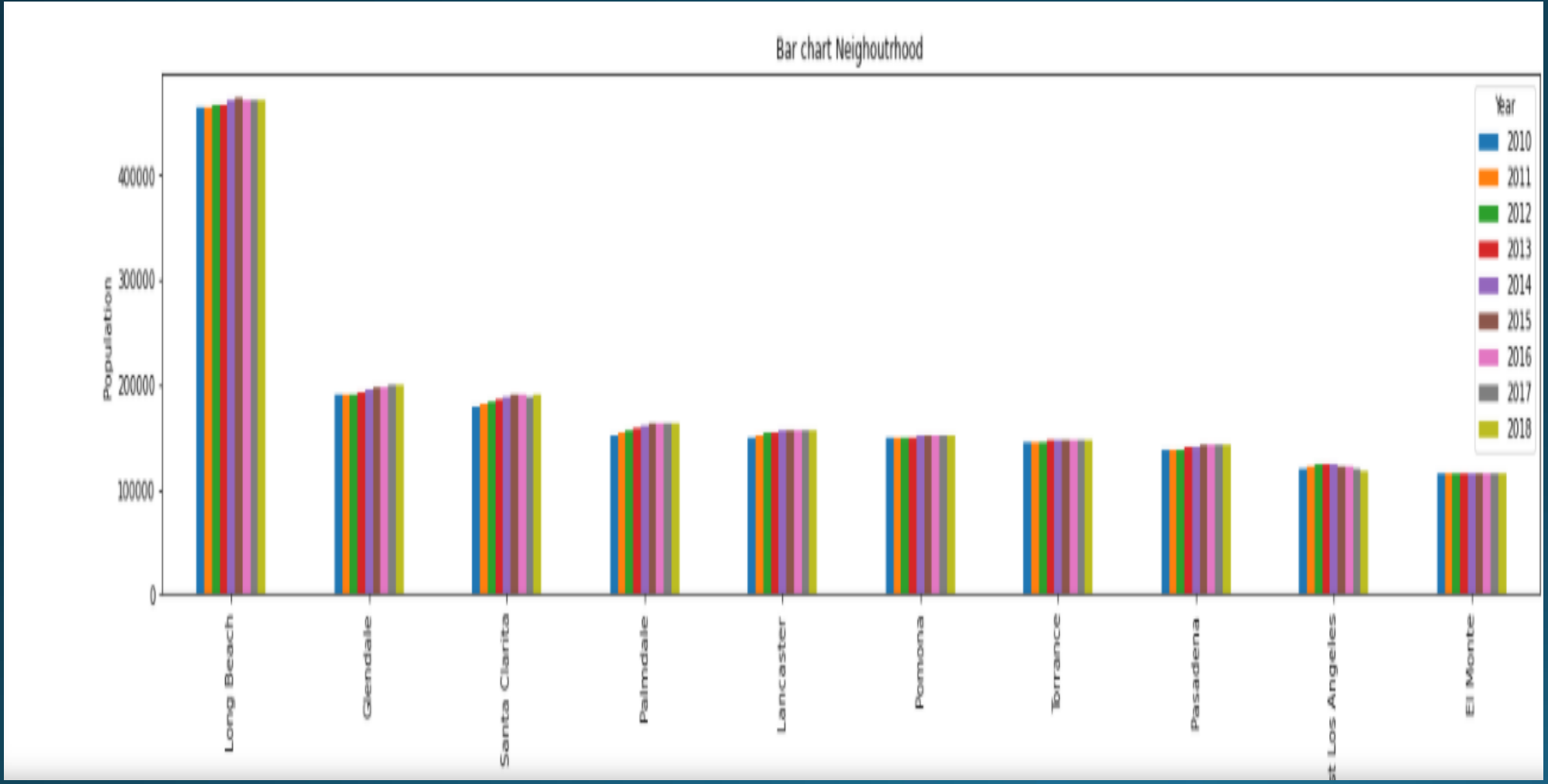
Top 5 venues in each neighbourhoods

Cluster Neighbourhoods using K-means

Finalize the cluster pick

Shortlist the neighbourhoods for planting Indian restaurants

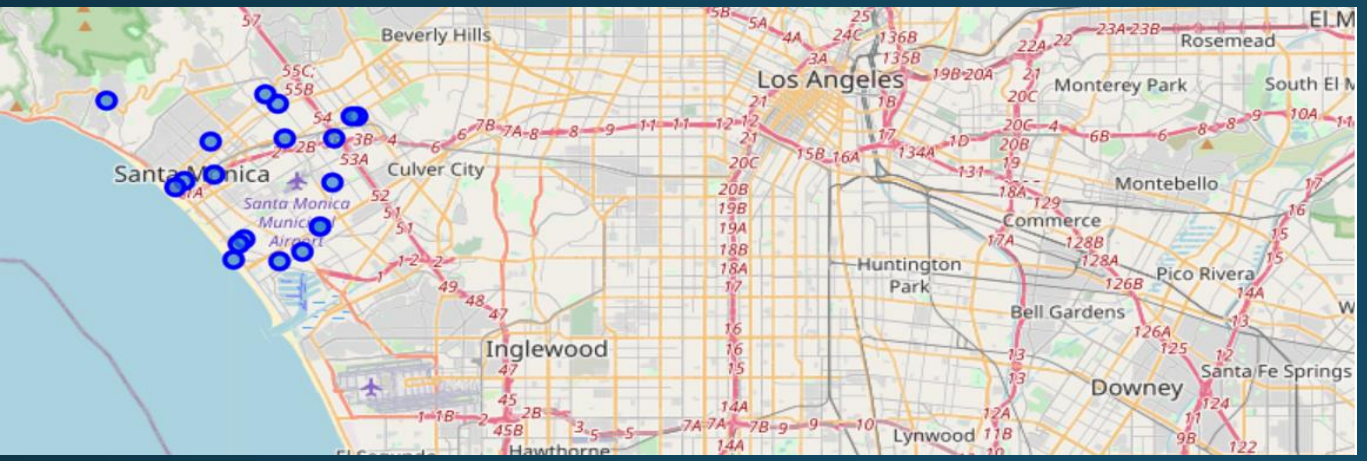
# Top 10 Highly populated LA neighbourhood trends




There are about 265 neighbourhoods in Los Angeles, this bar chart depicts top 10 highly populated neighbourhood yearly trends

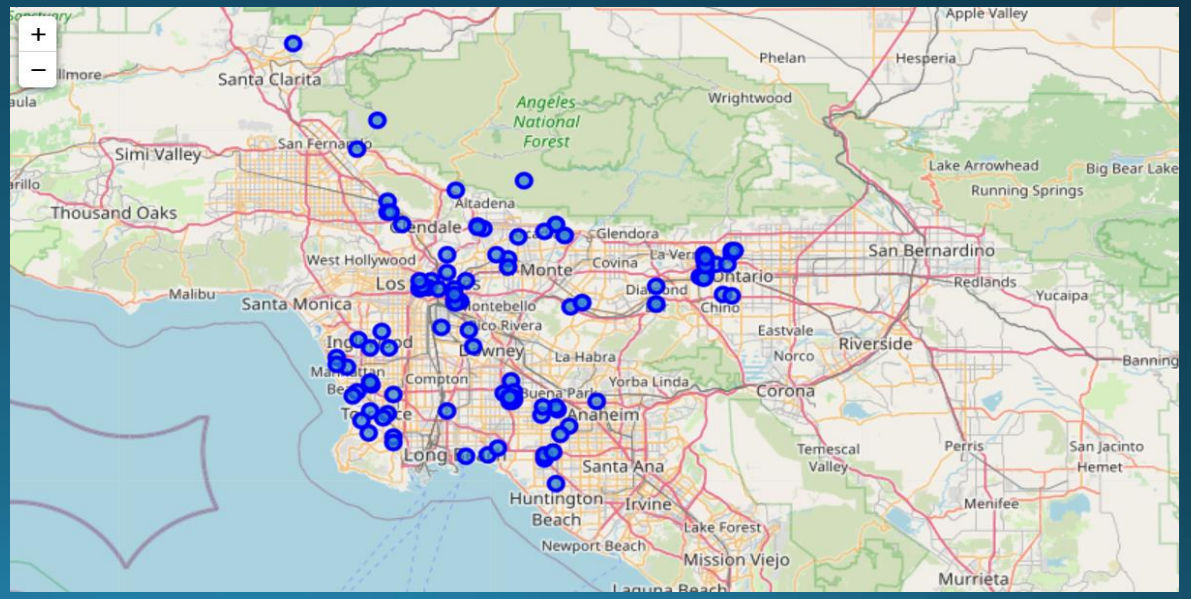
# Existing Indian Restaurants

Foursquare API was leveraged to find out how many Indian cuisines exists in LA neighbourhoods. We started our analysis in Santa Monica, Los Angeles. Gradually extended the search in top 10 highly populated neighbourhoods




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

Indian venues in LA top 10 highly populated neighbourhoods



# Unsupervised ML – K Means Clustering

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

- 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

- Data set prepared with neighbourhood along with the top 5 most common venues

# Unsupervised ML – K Means Clustering (Number of Clusters)

K Means clustering algorithm is applied to cluster neighbourhoods based on top 10 common venue data set.

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

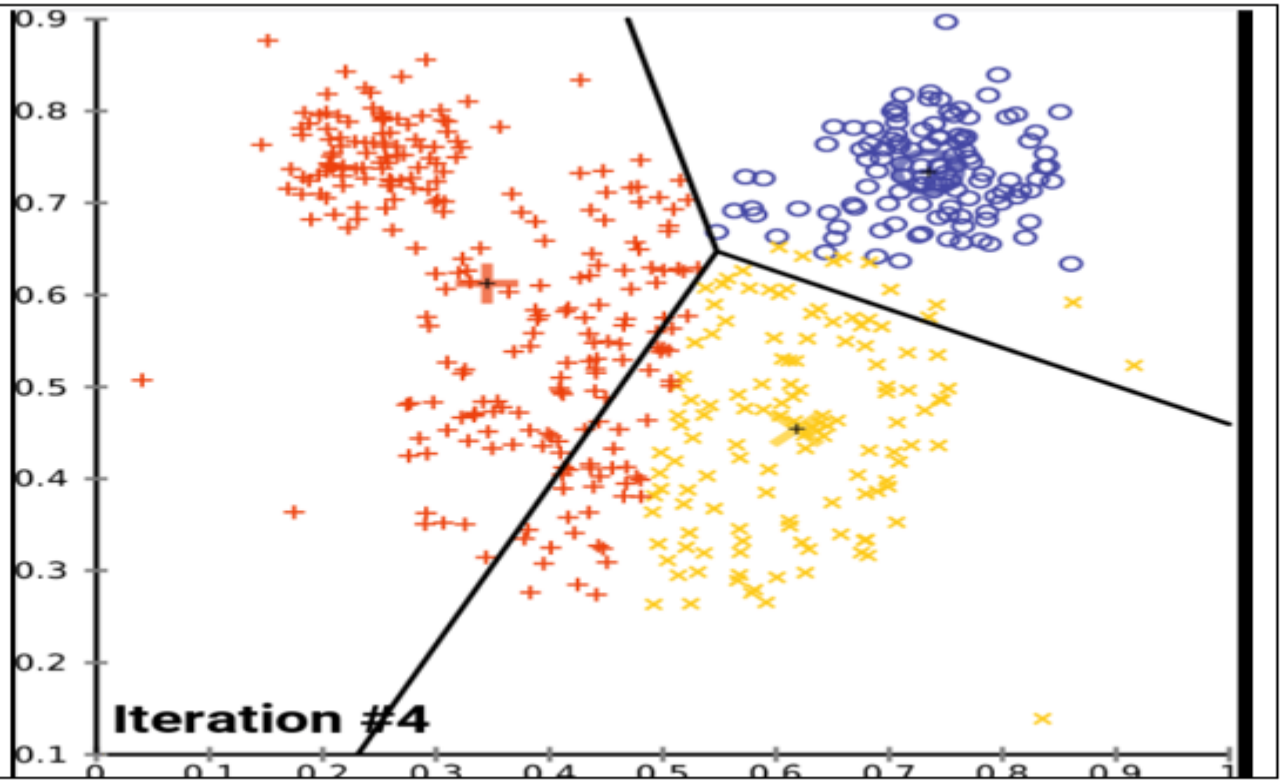
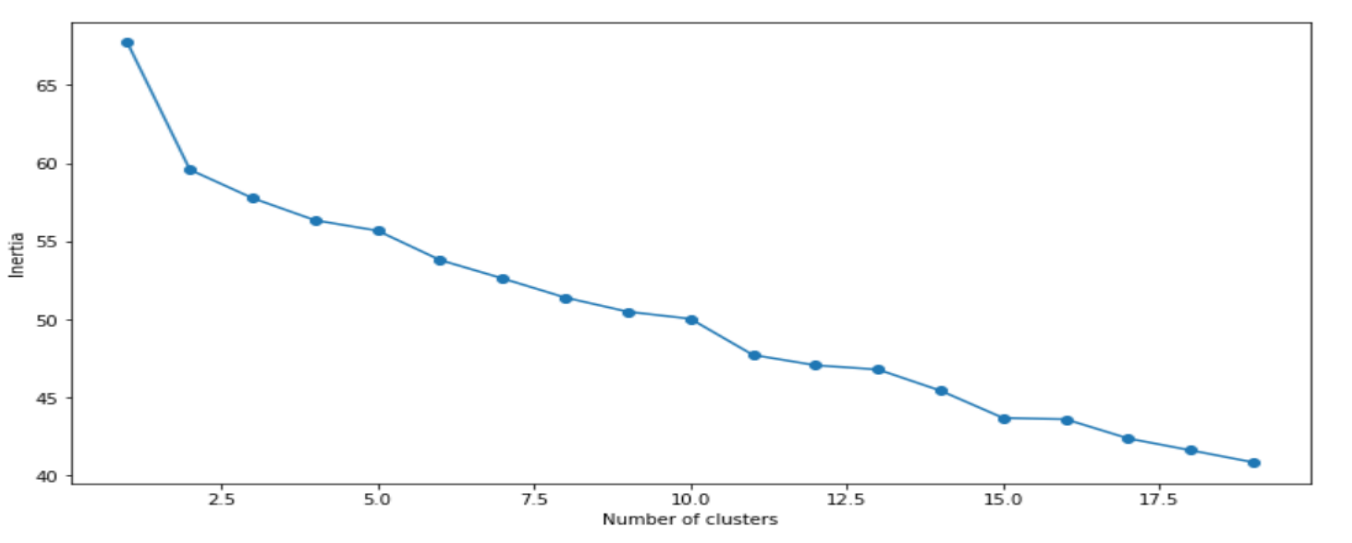



Fig2: K-Means clustering approach

# Elbow curve approach to finalize the number of clusters




Total number of clusters finalized after elbow curve approach applied

Clustered neighbourhoods

```
# venues in clusters
LA_merged["Cluster Labels"].value_counts()
```

```
0    2095
1      28
2      28
Name: Cluster Labels, dtype: int64
```

# Conclusion

As per the analysis cluster o Neighbourhood is picked up after deep analysis there were 49 neighbourhoods, in which based on top populate neighbourhoods we finalized following six Neighbourhoods for planning new Indian restaurants

'East Los Angeles', 'East Pasadena', 'El Monte', 'Long Beach', "Northwest Palmdale", "Pomona", "Torrance"