Battle of the Neighborhoods – A study of Los Angeles

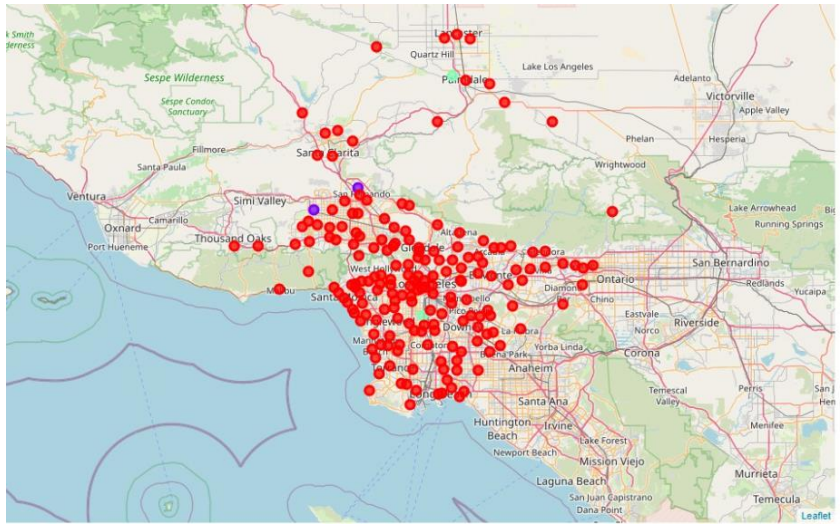
Introduction

- Focus of Study – Los Angeles and its neighborhoods
- Business Problem – Businesses with a physical store has location as one of the things to consider when deciding where to open the store in order to engage the most number of the potential target audience. Certain businesses want to avoid competition (eg. a specific cuisine restaurant) while certain businesses want to be placed near similar stores to engage more people (eg. Clothing store in shopping district).
- Target Audience – Business owners who need suggestions on where they could potentially open their store.
- This presentation serves to present the results of the study as well as a gist of the methodology.

Data

- Neighborhood-Zipcode
  Data: http://www.laalmanac.com/communications/cm02_communities.php
- Region-Zipcode
  Data: https://www.usmapguide.com/california/los-angeles-zip-code-map/
- Data of Nearby Venues: Foursquare
- Geographical Coordinates: Longitude and Latitude values from arcgis

gion	Neighborhood	Coord	Latitude	Lo
(Los eles)	Chatsworth (Los Angeles)	34.24428500000005,-118.58597999999995	34.24428500000005	-118.5859799
(Los eles)	Newhall (Santa Clarita)	34.381335000000036,-118.53003999999999	34.381335000000036	-118.5300399
(Los eles)	Los Angeles (Northridge)	34.269265000000075,-118.54840999999999	34.269265000000075	-118.5484099
(Los eles)	Bouquet Canyon	34.44731060000004,-118.51256279999996	34.44731060000004	-118.5125627
(Los eles)	Canyon Country (Santa Clarita)	34.42006010000006,-118.46667029999998	34.42006010000006	-118.4666702
...	...	...	...	
eles)	Long Beach (North Long Beach)	33.85147000000006,-118.18981499999995	33.85147000000006	-118.1898149
eles)	Bixby Knolls (Long Beach)	33.832610000000045,-118.18281499999995	33.832610000000045	-118.1828149
eles)	Long Beach (World Trade Center)	33.767660000000035,-118.19939	33.767660000000035	-1
eles)	California State Univ Long Beach (Long Beach)	33.77532500600006,-118.11814576899997	33.77532500600006	-118.1181457
eles)	Long Beach (Boeing)	33.830085000000054,-118.14268499999997	33.830085000000054	-118.1426849



Methodology

1) Region | Neighborhood | Zipcode Data scraped and extracted using BeautifulSoup into a dataframe.
2) Longitudes & Latitudes obtained with ARCGIS library
3) Visualization done with folium
4) Venues data extracted foursquare with info from previous dataframe
5) K-Means Clustering unsupervised learning model to cluster neighborhoods, silhouette score to determine optimum number of clusters.

Results

- 3 Clusters were successfully obtained.
- Cluster 1 was substantially larger (185 rows) than Cluster 2 and Cluster 3
- Cluster 2 and Cluster 3 were very similar in their own clusters.

Cluster 2

```
In [58]: 1 LA_merged.loc[LA_merged['Cluster Labels'] == 1, LA_merged.columns[[2] + list(range(5, LA_merged.shape[1]))]]
```

Out[58]:

	Neighborhood	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	Chatsworth (Los Angeles)	1	Convenience Store	Yoga Studio	Nail Salon	Optical Shop	Office	Notary	North Indian Restaurant	Noodle House	Nightlife Spot	Nightclub
23	Kagel Canyon	1	Convenience Store	Yoga Studio	Nail Salon	Optical Shop	Office	Notary	North Indian Restaurant	Noodle House	Nightlife Spot	Nightclub

Cluster 3

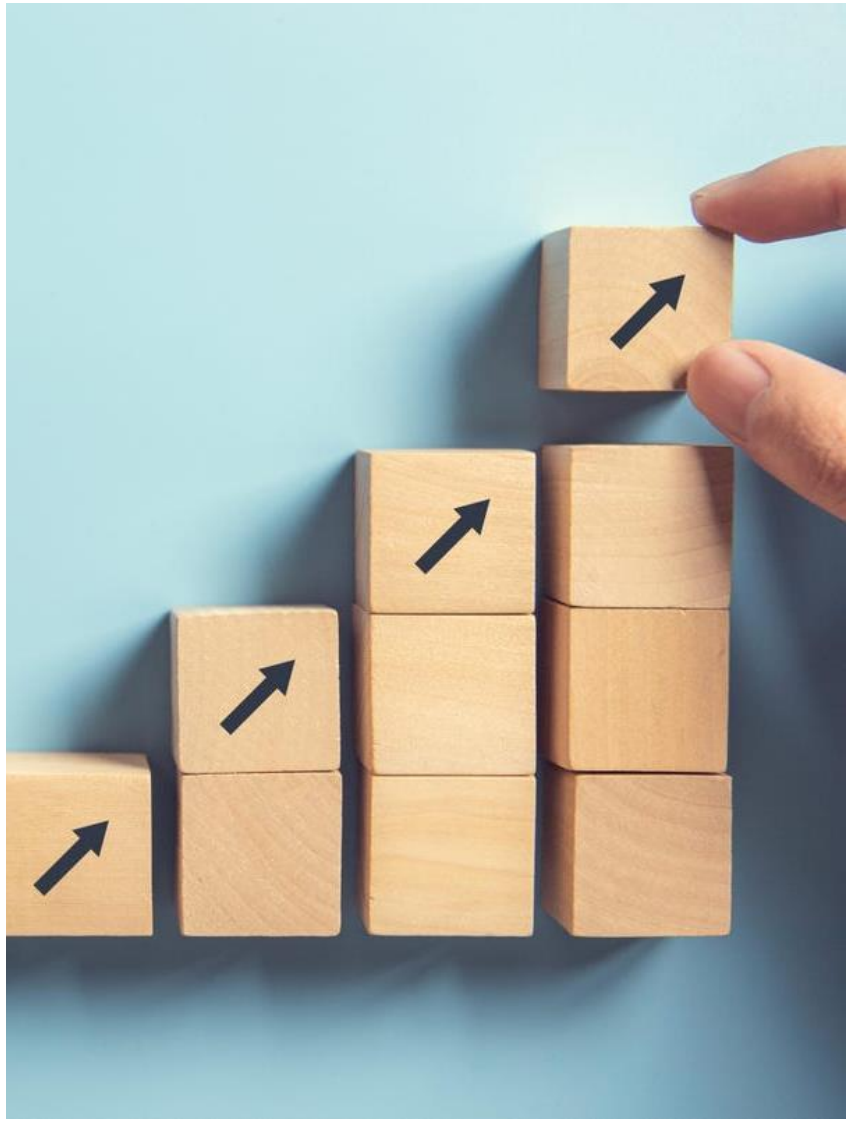
```
In [59]: 1 LA_merged.loc[LA_merged['Cluster Labels'] == 2, LA_merged.columns[[2] + list(range(5, LA_merged.shape[1]))]]
```

Out[59]:

	Neighborhood	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
16	City Ranch	2	Grocery Store	Optical Shop	Office	Notary	North Indian Restaurant	Noodle House	Nightlife Spot	Nightclub	New American Restaurant	Nail Salon
126	Florence-Graham, South Los Angeles (Los Angeles)	2	Grocery Store	Optical Shop	Office	Notary	North Indian Restaurant	Noodle House	Nightlife Spot	Nightclub	New American Restaurant	Nail Salon

Discussion

- With the data from the clustering done, a recommendation that business owners who are considering opening eateries can consider opening in Cluster 2 instead of Cluster 3, as amongst the top 10 most common venues, Cluster 2 has less eateries and variety would definitely be welcome.
- Also, opening a grocery store/convenience store would not be recommended in Cluster 2 or 3 as there seems to be too many of them already and these types of stores rarely want competition as goods sold are similar.
- Cluster 1 is too huge with too many varieties to make any conclusions about the cluster. Further EDA must be done to cluster 1 or the data gathering and clustering process must be refined.
- LA is a bustling city with high density, thus this overlap was not too out of the blue. Further work on future refinement could be done.



Future Expansions

- It is recognized that further refining could be used in order to increase the accuracy and usefulness of the results.

- Suggestions include:
- 326 venue categories were found, could be reduced by joining similar categories; eg. 'Nightclub' and 'Nightlife'
- Reducing search radius of each Neighborhood point as there could be an overlap where the same venue was fetched for many neighborhoods due to the density
- Foursquare is user-supported and may have inaccuracies due to the inactivity of the users to update or less users in the area; perhaps another location data provided could be used

Conclusion

- This project has the potential to help business owners choose a good location for their business in order to thrive.
- Further improvements and perhaps other learning algorithms could be used to help increase accuracy in order to be truly useful.

Thank You

For Coursera – Applied Data Science Capstone Project