DATA SCIENCE PROJECT : COVID IMPACT TRACKING AND CLUSTERING

PROBLEM STATEMENT

- With the increase of COVID disease, it is critical to impose strict isolation rules in severely impacted regions. However, as the disease is not expected to go away soon, it is also very important to avoid adverse impact on economy due to lockdown for a longer period.
- Regions with severe impact also cannot follow complete shutdown and hence it is crucial to understand which business/venues/places are more prone to spread of the disease and only those needs to follow strict rules, while the rest can observe little lighter rules to keep economy running and help reducing impact on day to day life of local residents
- At present we have several dashboards which shows location wise impact of the disease, but none of them provides view on what are the most common business/venues in those locations, which plays vital role in determining strategy to impose isolation and/or lockdown rules.
- Getting this full view on most common business/services around various regions with severity of the disease in those regions is challenging in absence of such dashboard

GOAL STATEMENT

In this project, I primarily aim to -

- Identify one of the most COVID impacted place of the world
- Explore various regions of that place
- Cluster them together based on type of most common venues/businesses in surrounding area
- Create visualizations to display region wise severity of the disease along with most common venues/businesses in those regions

Achieving above goal will help our target audience to determine strategy to impose isolation and/or lockdown rules in each region and thus control the spread of the disease and minimize impact on economy

TARGET AUDIENCE

Who can benefit from this project outcome?

- Local authorities and government of the severely impacted county/city to determine the best strategy to impose isolation/lockdown rules
- Local Police department to get strict isolation rules followed by public
- Local residents to understand severity of the disease by regions and venues/types of business

DATA SOURCES AND API/LIBRARIES

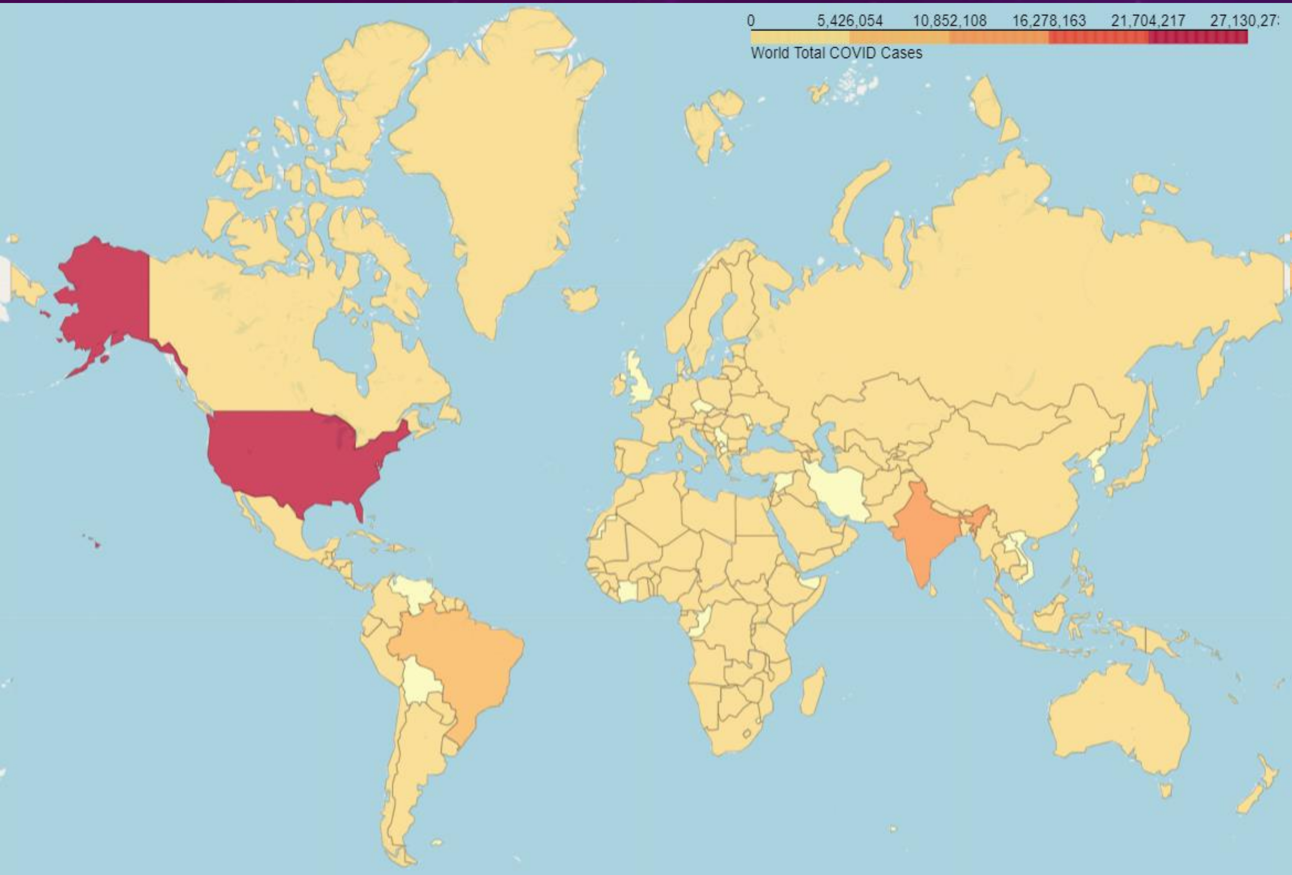
- WHO : Country wise Total # of COVID cases across the world
- Worldometer - USA : State wise Total # of COVID cases across USA
- Worldometer-California : County wise Toal # of COVID cases across state of California
- LA county public dashboard : Region wise Total # of COVID cases across Los Angeles county
- **The Geocoder Python library :** Location Coordinates of various regions of Los Angeles county
- **Folium Python library :** Visualize geographic details of region or place
- **Foursquare API :** Explore various regions and find out most common venues around those regions to cluster them together
- **Choropleth Map :** Visualize COVID severity across various regions and display clusters of most common venues superimposed on covid severity map

APPROACH

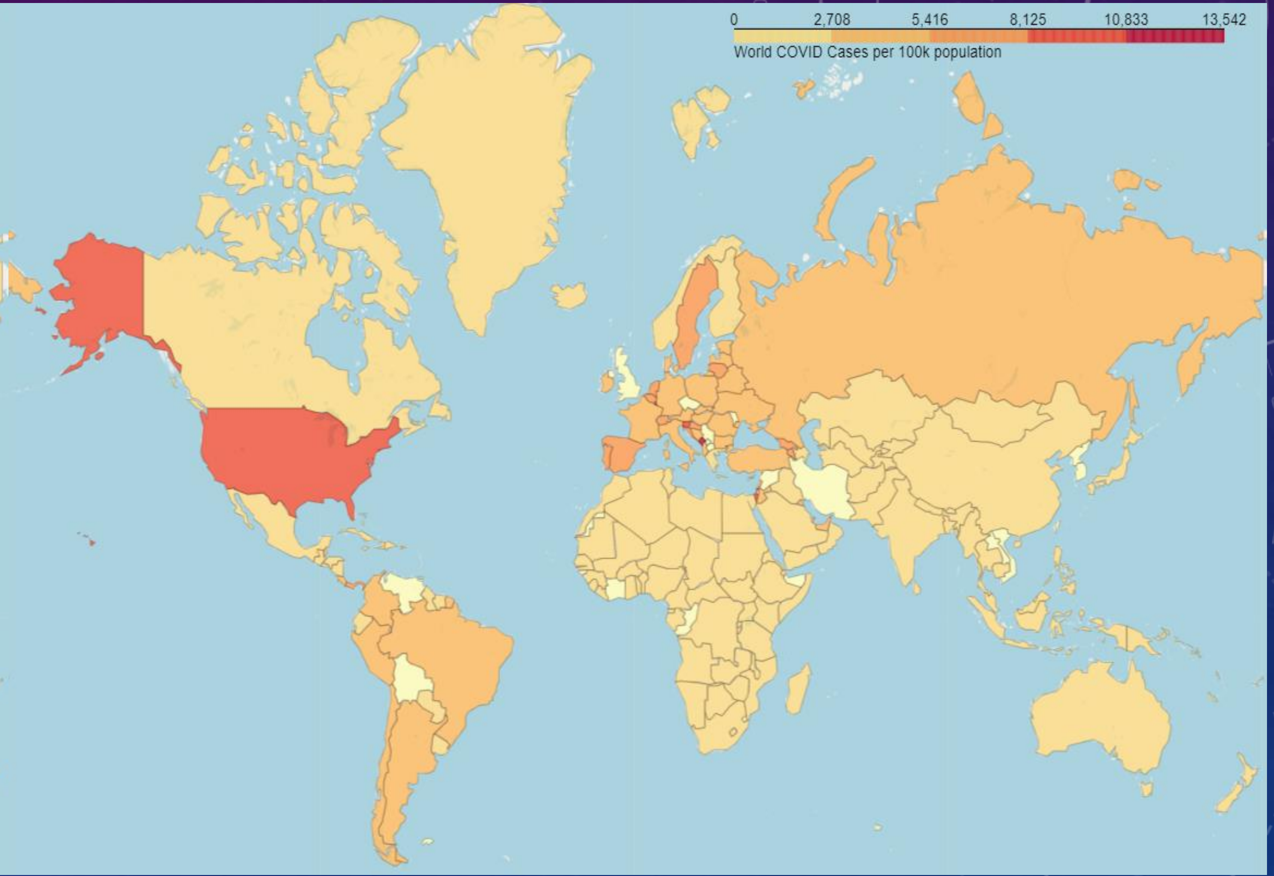
- Collect required data from various sources and preprocess
- Visualize COVID impact on the countries on world map using Choropleth
- visualize state wise impact of the country that is severly impacted in most aspects
- Pick the most impacted state and visualize county wise impact
- Pick the most adversly impacted county and visualize various regions of that county to understand severity of the disease
- Get coordinates of county regions using Geocoder Python library
- Explore nearby venues of each the region using Foursquare API
- Analyze data using K-means clustering an d superimpose clusters on the choropleth map showing COVID impact by various regions of the county
- Categorize each cluster based on top 3 most common venue and determine cluster wise COVID impact
- From the results, conclude severly, moderately and lighly impacted clusters and corresponding venue types
- Recommend lockdown strategy from conclusion of analysis

World COVID CASES

Country wise Total number of cases



Country wise Total number of cases per 100k population

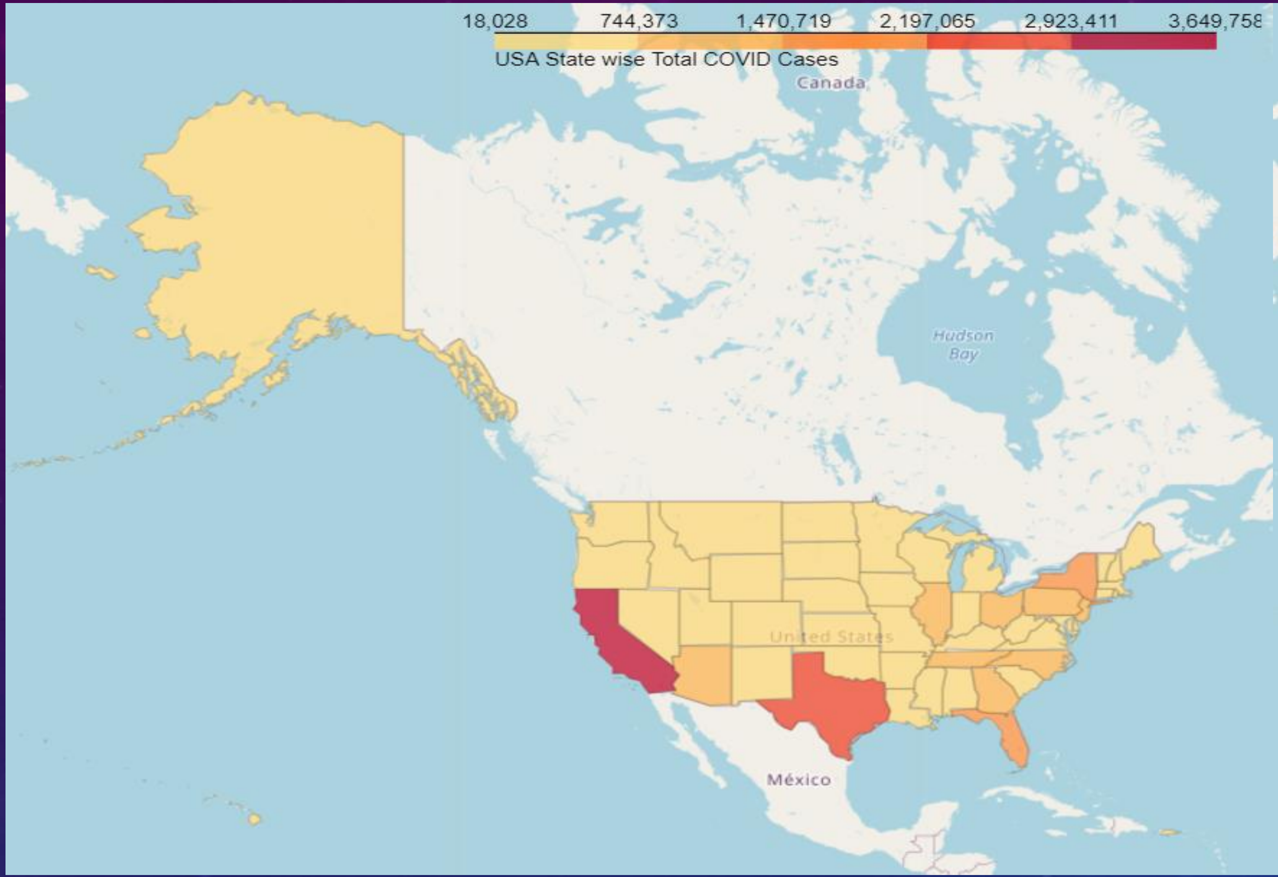


Conclusion :

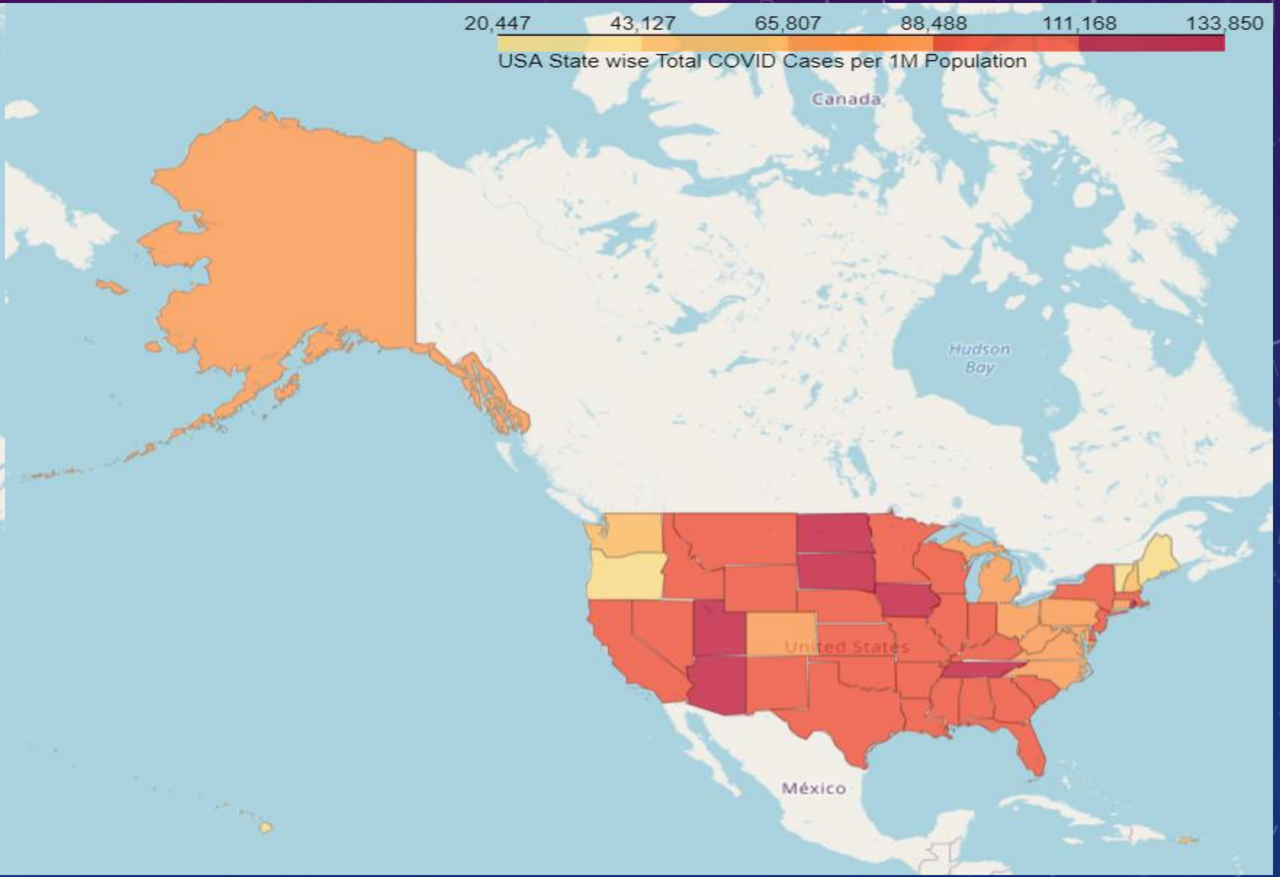
As we can see from these visualizations, some of the countries have more number of total cases due to their higher population, however when we compare countries by total #of cases per 100k population, those are not on the top of the list. However, USA is the only country which is leading the list in every category, and hence is most severely impacted country of the world by COVID

USA COVID CASES

State wise Total number of cases



State wise Total number of cases per 1M population

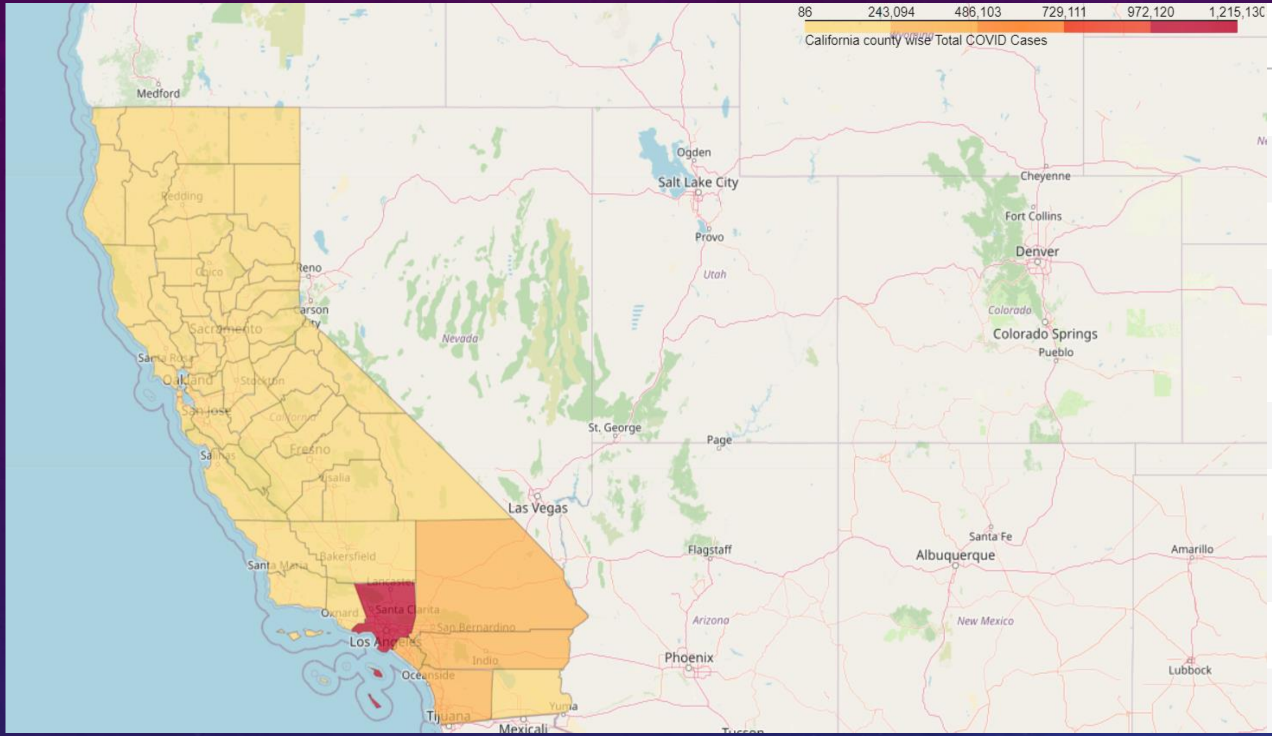


Conclusion :

As we can see from above analysis of USA states, California is leading the show in terms of total# of cases. There are other states which are more impacted when we consider # of cases per 1 million of population. However, California is not far behind in this category as well, and hence being one of the most populated state, my further analysis will be focused on California state

CALIFORNIA COVID CASES

County wise Total number of cases



Top 10 county by Toala cases in California

Region	index	Total_cases	population	Case_Percentage	Latitude	Longitude
East Los Angeles	60	24217	125269	19.33	34.03347	-118.159090
Pomona	313	23575	157869	14.93	34.05499	-117.750040
Palmdale	291	23353	159810	14.61	34.57936	-118.116590
Florence-Firestone	147	22051	112150	19.66	33.97475	-118.249932
Lancaster	118	20662	161570	12.79	34.69893	-118.144780
North Hollywood	140	19454	151421	12.85	34.16982	-118.378990
Santa Clarita	174	18543	220424	8.41	34.41389	-118.551180
Glendale	77	18091	206493	8.76	34.14633	-118.248640
South Gate	183	17892	98155	18.23	33.95722	-118.205630
Boyle Heights	24	16928	86884	19.48	34.04004	-118.210500

Conclusion :

As we can see from above, Los Angeles has highest number of cases and being largest and one of the most populated city, I will focus my analysis on Los Angeles county and surroundings from here and find out what all regions of LA county are severely impacted and require strict isolation rules to reduce the spread of the disease

NEXT STEPS

Now, we have details of total # of COVID cases for each neighbourhood region of LA county.

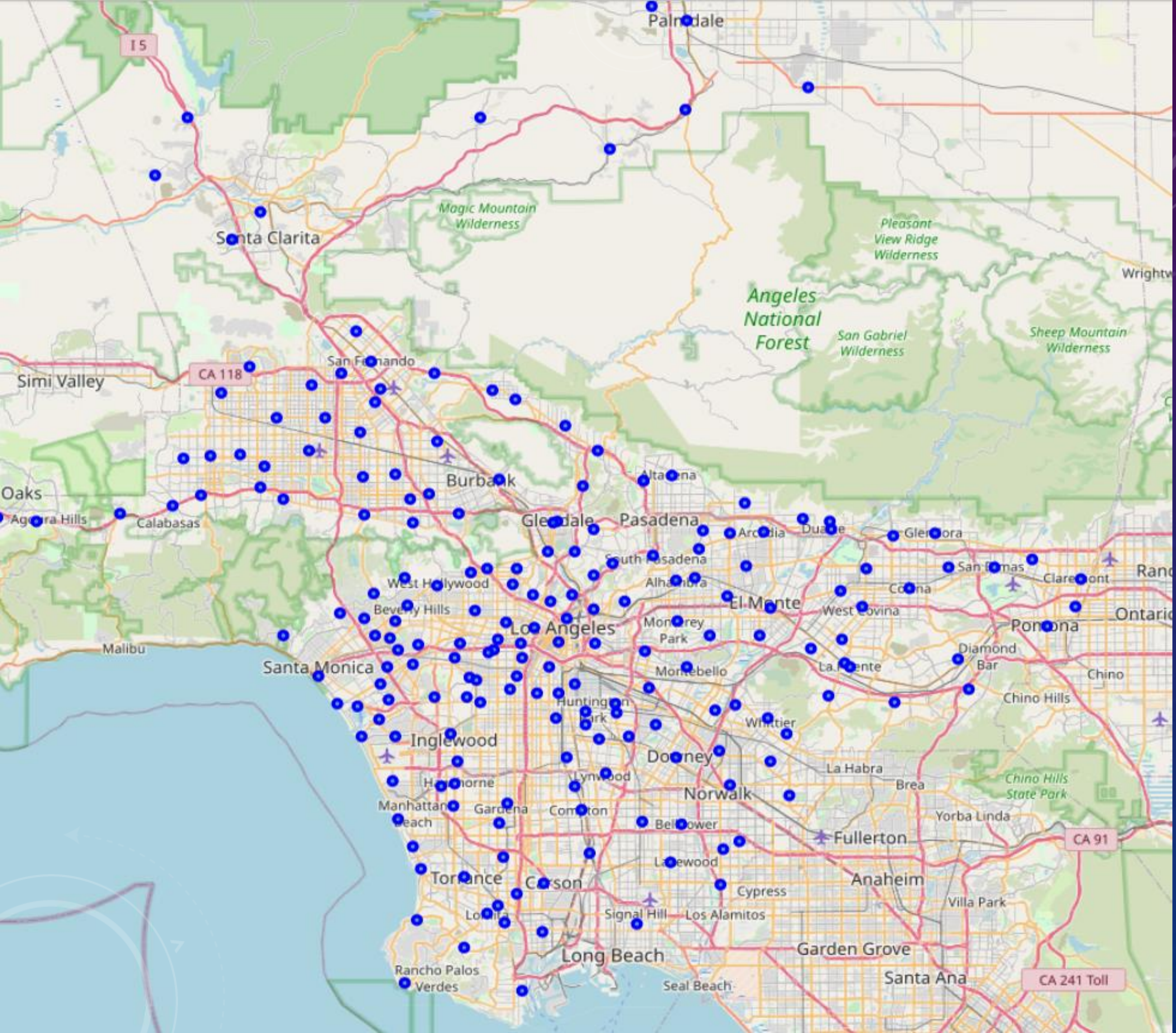
Next I will perform following tasks to achieve the goal to help target audience of this project -

- Find out location co-ordinates of all neighbourhood regions of Los Angeles county
- Show those regions in the map of LA to visualize severity of COVID for each region using choropleth map
- Explore and Analyze nearby venues of each region using Foursquare API
- Create clusters using K-means and assign cluster label to each region
- Find out total %age of COVID infection in each cluster, to determine which which clusters are severely impacted
- Identify most common venues in each cluster and assign venue category lable to each cluster accordingly
- Superimpose clusters on choropleth map showing COVID severity of all the regions of LA county
- Draw conclusion based on the findings from above analysis

LOCATION COORDINATES

Use Geocode Python Library to get location coordinates (latitude and longitudes) of each Neighbour region of LA county

Region	index	Total_cases	population	Case_Percentage	Latitude	Longitude
Acton	0	425	7971	5.33	34.468150	-118.195130
Adams-Normandie	1	1129	8202	13.76	33.901212	-118.299321
Agoura Hills	2	930	20883	4.45	34.146110	-118.778120
Agua Dulce	3	246	4158	5.92	34.495700	-118.326210
Alhambra	4	6544	86724	7.55	34.094420	-118.127780
Altadena	5	3083	43620	7.07	34.185560	-118.131520
Arcadia	13	3125	65735	4.75	34.136410	-118.038620
Arleta	8	6750	34370	19.64	34.249050	-118.433490
Artesia	9	1939	16795	11.55	33.861140	-118.079680
Atwater Village	10	1337	14666	9.12	34.119700	-118.258870



VISUALIZE NEIGHBOURHOOD REGIONS OF LA COUNTY

EXPLORE LA NEIGHBOURHOOD

- Use Foursquare library to explore 100 nearby venues in 1500 meter surroundings of each region of LA county

	Region	Region Latitude	Region Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acton	34.46815	-118.19513	Acton Market & Country Store	34.468595	-118.197626	Grocery Store
1	Acton	34.46815	-118.19513	Fox Hay Feed and Grain	34.469565	-118.195481	Pet Store
2	Acton	34.46815	-118.19513	High Mesa	34.467814	-118.196090	Food
3	Acton	34.46815	-118.19513	Acton Market	34.467628	-118.195892	Grocery Store
4	Acton	34.46815	-118.19513	TSW Social Media Marketing	34.470898	-118.192307	Market

- Assign specific venue category type to each venue category to place similar venue category under single bucket

	Region	Region Latitude	Region Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Category Type
0	Acton	34.46815	-118.19513	Acton Market & Country Store	34.468595	-118.197626	Grocery Store	Grocery
1	Acton	34.46815	-118.19513	Fox Hay Feed and Grain	34.469565	-118.195481	Pet Store	Public Service Place
2	Acton	34.46815	-118.19513	High Mesa	34.467814	-118.196090	Food	Food
3	Acton	34.46815	-118.19513	Acton Market	34.467628	-118.195892	Grocery Store	Grocery
4	Acton	34.46815	-118.19513	TSW Social Media Marketing	34.470898	-118.192307	Market	Grocery

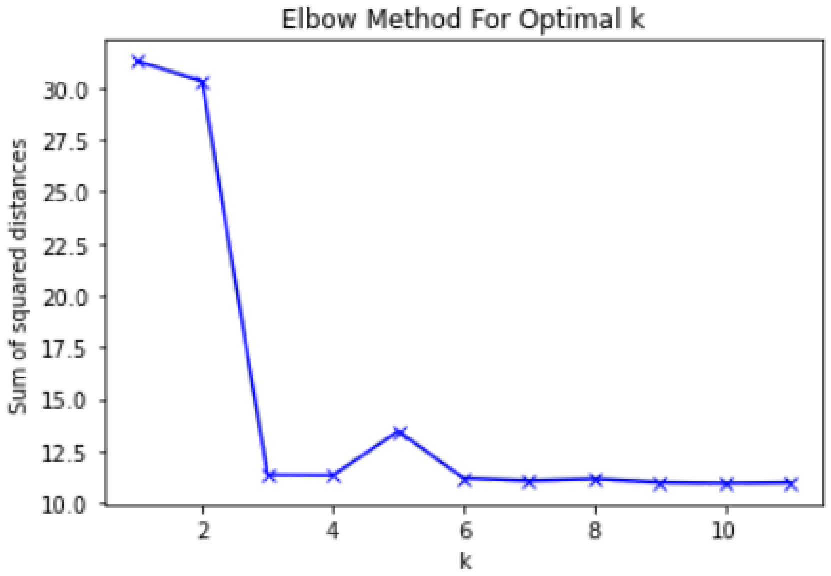
TOP 10 MOST COMMON VENUES AROUND LA REGIONS

	Region	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	Acton	Grocery	Office/Business Place	Public Service Place	Food	Art/Museum	Asian Restaurant	Australian Restaurant	Art/Craft/Flower Shop	Bakery/Breakfast Spot	Bank/ATM
1	Adams-Normandie	Food	American Restaurant	Public Service Place	Gym/Yoga/Spa	Coffee Shop	Office/Business Place	Pizza/Salad Place	Grocery	Sandwich Place	Art/Craft/Flower Shop
2	Agoura Hills	Accomodation	Asian Restaurant	Furniture / Home Service	Bar/Pub/Liquor Store	American Restaurant	Public Service Place	Pizza/Salad Place	Gym/Yoga/Spa	Italian Restaurant	Cafe
3	Agua Dulce	Grocery	American Restaurant	Electronics Store	Mexican Restaurant	Bakery/Breakfast Spot	Pizza/Salad Place	Cafe	Clothing Store	Fast Food Restaurant	European Restaurant
4	Alhambra	Asian Restaurant	Grocery	Dessert/Ice Cream Shop	Bakery/Breakfast Spot	Sandwich Place	Seafood Restaurant	Cafe	Bar/Pub/Liquor Store	Miscellaneous Shopping Place	Food
5	Altadena	Grocery	Miscellaneous Shopping Place	Pizza/Salad Place	Sports	Electronics Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Bank/ATM	Bar/Pub/Liquor Store
6	Arcadia	Sports	Miscellaneous Shopping Place	Park/Garden/Beach	Food	Fast Food Restaurant	European Restaurant	Electronics Store	Dessert/Ice Cream Shop	Coffee Shop	Clothing Store
7	Arleta	Grocery	Art/Craft/Flower Shop	Sandwich Place	Cafe	Fast Food Restaurant	European Restaurant	Electronics Store	Dessert/Ice Cream Shop	Coffee Shop	Clothing Store
8	Artesia	Asian Restaurant	Grocery	Dessert/Ice Cream Shop	Miscellaneous Shopping Place	Salon / Barbershop / Massage	Vegetarian / Vegan Restaurant	Sandwich Place	Bakery/Breakfast Spot	Food	Sports
9	Atwater Village	American Restaurant	Asian Restaurant	Miscellaneous Shopping Place	Bar/Pub/Liquor Store	Coffee Shop	Pizza/Salad Place	Mediterranean Restaurant	Sandwich Place	Art/Craft/Flower Shop	Public Service Place

K-MEANS CLUSTERING -

CREATE CLUSTERS OF LA COUNTY REGIONS BASED ON SURROUNDING VENUES:

Find optimum value of k using Elbow method



From above plot, we can see that plot drops drastically between 2 & 3 and then again rises. At 6 it stabilizes, so I will keep number of clusters = 6 for my analysis

CREATE K CLUSTERS

```
# set number of clusters
kclusters = 6 # this is optimum value found out from Elbow method above

#la_grouped_clustering = la_grouped.drop('Region', 1)

# run k-means clustering
kmeans = KMeans(init="k-means++", n_clusters=kclusters, random_state=10,n_init=50, max_iter=400).fit(la_grouped_clustering)

# check cluster labels generated for each row in the dataframe
labels= kmeans.labels_
labels
#print(kmeans.labels_)
```

```
array([2, 3, 3, 2, 1, 3, 0, 3, 1, 3, 3, 0, 3, 3, 3, 0, 3, 3, 3, 3, 3, 3,
       1, 3, 3, 1, 5, 1, 1, 0, 1, 0, 3, 3, 3, 1, 3, 1, 3, 1, 3, 3, 3, 3,
       3, 3, 2, 3, 1, 3, 1, 3, 1, 3, 1, 3, 3, 2, 3, 0, 5, 0, 2, 3, 3, 2,
       1, 1, 3, 3, 2, 1, 1, 3, 4, 1, 1, 1, 3, 3, 1, 3, 3, 3, 3, 3, 1, 1,
       3, 1, 3, 1, 2, 1, 3, 3, 0, 3, 1, 3, 1, 3, 3, 0, 3, 1, 3, 3, 2, 2,
       3, 2, 3, 3, 2, 3, 1, 3, 3, 3, 2, 1, 3, 1, 3, 1, 3, 1, 1, 1, 3, 1,
       3, 3, 1, 1, 2, 3, 3, 3, 3, 3, 3, 1, 1, 3, 3, 0, 3, 1, 1, 2, 3, 3,
       1, 3, 3, 3, 3, 3, 1, 3, 0, 1, 3, 2, 0, 1, 2, 5, 3, 1, 3, 1, 1, 3,
       1, 1, 1, 3, 3, 5, 3, 2, 3, 2, 1, 3, 3, 3, 3, 4, 2, 1, 2, 0, 3, 2,
       3, 3, 1, 1, 3, 3, 3, 1, 3, 3, 3, 3, 3, 3, 3], dtype=int32)
```

Label each region with corresponding cluster

	Region	index	Total_cases	population	Case_Percentage	Latitude	Longitude	Labels
0	Acton	0	425	7971	5.33	34.468150	-118.195130	2
1	Adams-Normandie	1	1129	8202	13.76	33.901212	-118.299321	3
2	Agoura Hills	2	930	20883	4.45	34.146110	-118.778120	3
3	Agua Dulce	3	246	4158	5.92	34.495700	-118.326210	2
4	Alhambra	4	6544	86724	7.55	34.094420	-118.127780	1
5	Altadena	5	3083	43620	7.07	34.185560	-118.131520	3
6	Arcadia	13	3125	65735	4.75	34.136410	-118.038620	0

CLUSTER RESULTS

Labels	0	1	2	3	4	5
Total_cases	3966.461538	5440.966102	4736.363636	4390.061947	537.00	2413.25
population	31815.000000	49211.305085	30971.590909	41132.902655	8624.00	25956.75
Case_Percentage	11.961538	10.645932	12.121364	9.648230	32.69	6.29

Identify total %age of infected population in each cluster

Add clustering labels showing most common venue categories

	Cluster Labels	Region	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	2	Acton	Grocery	Public Service Place	Office/Business Place	Food	Art/Museum	Asian Restaurant	Australian Restaurant	Art/Craft/Flower Shop	Bakery/Breakfast Spot	Bank/ATM
1	3	Adams-Normandie	Food	American Restaurant	Pizza/Salad Place	Sports	Gym/Yoga/Spa	Coffee Shop	Grocery	Sandwich Place	Art/Craft/Flower Shop	Park/Garden/Beach
2	3	Agoura Hills	Furniture / Home Service	Accomodation	Bar/Pub/Liquor Store	Art/Craft/Flower Shop	Fast Food Restaurant	American Restaurant	Gym/Yoga/Spa	Italian Restaurant	Miscellaneous Shopping Place	Cafe
3	2	Agua Dulce	Grocery	American Restaurant	Electronics Store	Mexican Restaurant	Bakery/Breakfast Spot	Pizza/Salad Place	Cafe	Clothing Store	Fast Food Restaurant	European Restaurant
4	1	Alhambra	Asian Restaurant	Grocery	Dessert/Ice Cream Shop	Bakery/Breakfast Spot	Sandwich Place	Bar/Pub/Liquor Store	Public Service Place	Seafood Restaurant	Cafe	American Restaurant

Create a new dataframe which contains both clusters as well as top 10 venues of each neighbourhood Region

	Cluster Labels	Region	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	2	Acton	Grocery	Public Service Place	Office/Business Place	Food	Art/Museum	Asian Restaurant	Australian Restaurant	Art/Craft/Flower Shop	Bakery/Breakfast Spot	Bank/ATM
1	3	Adams-Normandie	Food	American Restaurant	Pizza/Salad Place	Sports	Gym/Yoga/Spa	Coffee Shop	Grocery	Sandwich Place	Art/Craft/Flower Shop	Park/Garden/Beach
2	3	Agoura Hills	Furniture / Home Service	Accomodation	Bar/Pub/Liquor Store	Art/Craft/Flower Shop	Fast Food Restaurant	American Restaurant	Gym/Yoga/Spa	Italian Restaurant	Miscellaneous Shopping Place	Cafe
3	2	Agua Dulce	Grocery	American Restaurant	Electronics Store	Mexican Restaurant	Bakery/Breakfast Spot	Pizza/Salad Place	Cafe	Clothing Store	Fast Food Restaurant	European Restaurant
4	1	Alhambra	Asian Restaurant	Grocery	Dessert/Ice Cream Shop	Bakery/Breakfast Spot	Sandwich Place	Bar/Pub/Liquor Store	Public Service Place	Seafood Restaurant	Cafe	American Restaurant

ANALYZE LA REGIONS AND CLUSTERS

Region	Venue Category Type	Counts
Agua Dulce	Bakery/Breakfast Spot	1
Agoura Hills	Furniture / Home Service	3
Agoura Hills	Accomodation	2
Agoura Hills	Bar/Pub/Liquor Store	2
Adams-Normandie	American Restaurant	2
Adams-Normandie	Food	2
Adams-Normandie	Art/Craft/Flower Shop	1
Acton	Grocery	3
Acton	Food	1
Acton	Office/Business Place	1

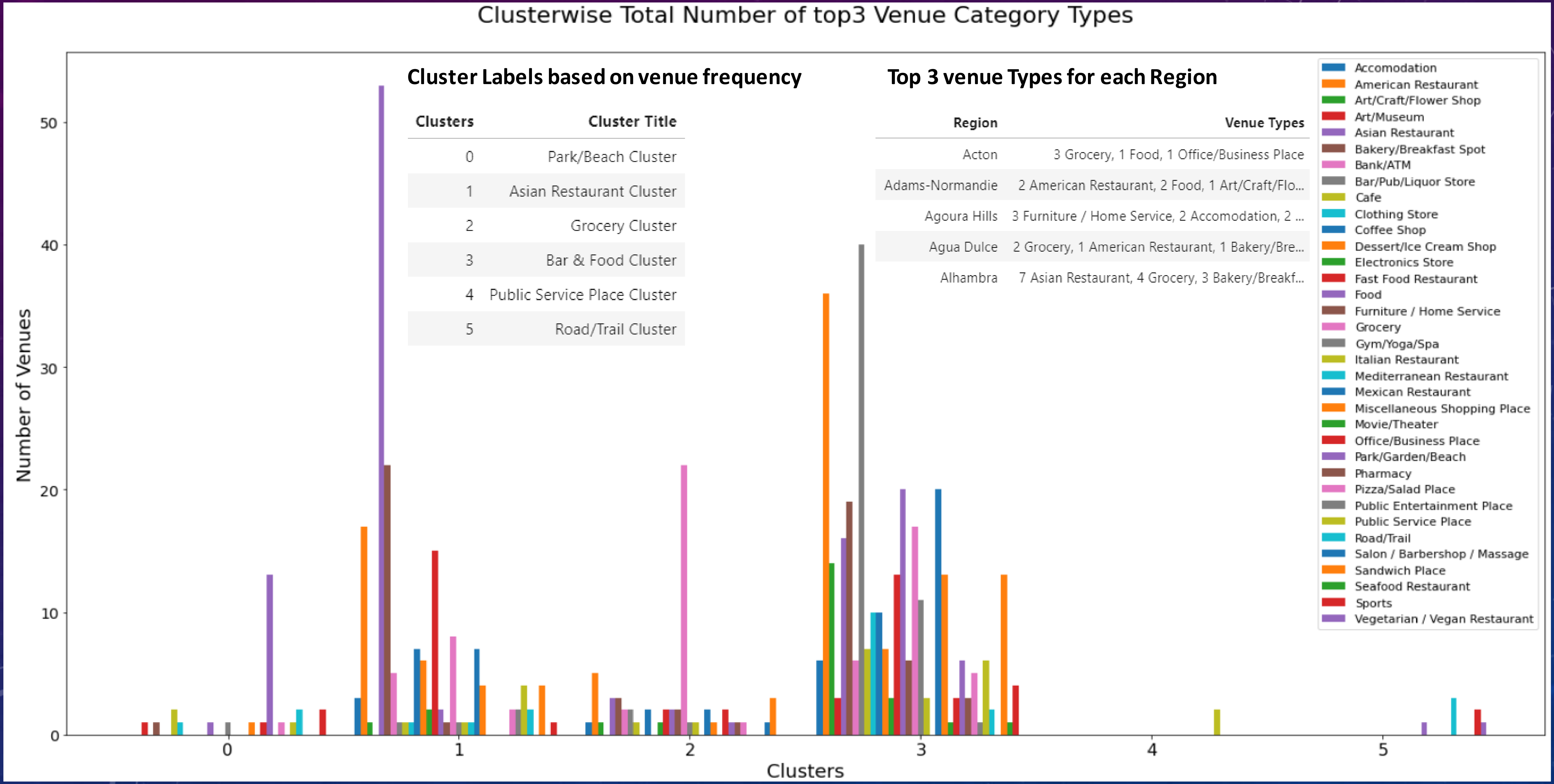
Frequency of top3 most common venue for each region

Region	Venue Category Type	Counts
Woodland Hills	Bar/Pub/Liquor Store	2
Woodland Hills	Grocery	2
Woodland Hills	American Restaurant	1
Winnetka	American Restaurant	1
Winnetka	Art/Craft/Flower Shop	1
Winnetka	Coffee Shop	1
Wilmington	Grocery	4
Wilmington	Fast Food Restaurant	2
Wilmington	Mexican Restaurant	2
Willowbrook	Bar/Pub/Liquor Store	1

Pick top3 venue for each region to estimate frequency of most common venues in each cluster

Venue Category Type	Accomodation	American Restaurant	Art/Craft/Flower Shop	Art/Museum	Asian Restaurant	Bakery/Breakfast Spot	Bank/ATM	Bar/Pub/Liquor Store	Cafe	Clothing Store	...	Pharmacy	Pizza/Salad Place	Public Entertainment Place	Public Service Place	Road/Trail	Salon / Barbershop / Massage
0	0	0	0	1	0	1	0	0	2	1	...	0	1	0	1	2	0
1	3	17	1	0	53	22	5	1	1	1	...	0	2	2	4	2	0
2	1	5	1	0	3	3	2	2	1	0	...	1	1	0	0	0	1
3	6	36	14	3	16	19	6	40	7	10	...	3	5	1	6	2	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	2	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	3	0

VISUALIZE VENUE CATEGORY FREQUENCY IN EACH CLUSTER

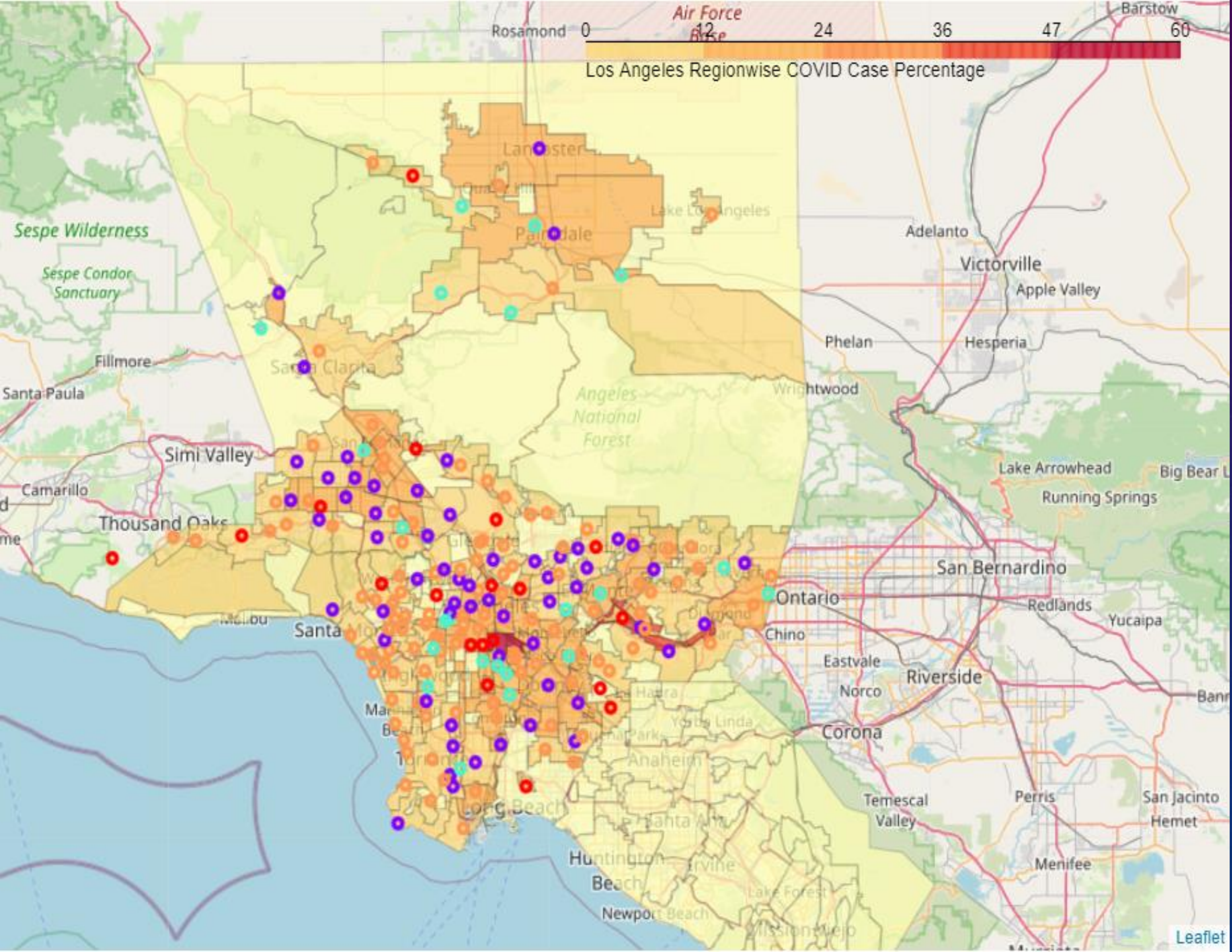


Cluster Labels based on venue frequency

Clusters	Cluster Title
0	Park/Beach Cluster
1	Asian Restaurant Cluster
2	Grocery Cluster
3	Bar & Food Cluster
4	Public Service Place Cluster
5	Road/Trail Cluster

Top 3 venue Types for each Region

Region	Venue Types
Acton	3 Grocery, 1 Food, 1 Office/Business Place
Adams-Normandie	2 American Restaurant, 2 Food, 1 Art/Craft/Flo...
Agoura Hills	3 Furniture / Home Service, 2 Accomodation, 2 ...
Agua Dulce	2 Grocery, 1 American Restaurant, 1 Bakery/Bre...
Alhambra	7 Asian Restaurant, 4 Grocery, 3 Bakery/Breakf...



VISUALIZE CLUSTERS SUPERIMPOSED ON COUNTY WISE SEVERITY MAP OF LOS ANGELES

MOST IMPACTED REGIONS WITH CASE %AGE, CLUSTER TITLE AND 5 MOST COMMON VENUE CATEGORY TYPES

Region	Case_Percentage	Clusters	Cluster Title	Venue Types	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Vernon	59.81	4	Public Service Place Cluster	1 Public Service Place	Public Service Place	Vegetarian / Vegan Restaurant	Bar/Pub/Liquor Store	Fast Food Restaurant	European Restaurant
Industry	37.07	1	Asian Restaurant Cluster	3 Asian Restaurant, 2 Food, 2 Mexican Restaurant	Asian Restaurant	Food	Sandwich Place	Mexican Restaurant	Fast Food Restaurant
Pacoima	21.42	3	Bar & Food Cluster	2 Fast Food Restaurant, 1 Clothing Store, 1 Food	Fast Food Restaurant	Food	Clothing Store	Furniture / Home Service	Sandwich Place
San Fernando	20.32	3	Bar & Food Cluster	7 Mexican Restaurant, 4 Dessert/Ice Cream Shop...	Mexican Restaurant	Dessert/Ice Cream Shop	Fast Food Restaurant	Food	Bakery/Breakfast Spot
Florence-Firestone	19.66	2	Grocery Cluster	2 Grocery, 1 Bakery/Breakfast Spot, 1 Bank/ATM	Grocery	Bakery/Breakfast Spot	Mexican Restaurant	Miscellaneous Shopping Place	Bank/ATM
Arleta	19.64	3	Bar & Food Cluster	1 Art/Craft/Flower Shop, 1 Sandwich Place, 1 S...	Sports	Art/Craft/Flower Shop	Sandwich Place	Cafe	Food
Boyle Heights	19.48	1	Asian Restaurant Cluster	2 Asian Restaurant, 2 Bank/ATM, 2 Grocery	Grocery	Bank/ATM	Asian Restaurant	Fast Food Restaurant	Electronics Store
East Los Angeles	19.33	3	Bar & Food Cluster	3 Mexican Restaurant, 2 Dessert/Ice Cream Shop...	Mexican Restaurant	Dessert/Ice Cream Shop	Public Service Place	Pharmacy	Park/Garden/Beach
Cudahy	18.96	3	Bar & Food Cluster	1 Bar/Pub/Liquor Store	Bar/Pub/Liquor Store	Cafe	Food	Fast Food Restaurant	European Restaurant
South Park	18.67	0	Park/Beach Cluster	1 Park/Garden/Beach	Park/Garden/Beach	Vegetarian / Vegan Restaurant	Cafe	Fast Food Restaurant	European Restaurant

RESULT TABLE – CLUSTERS WITH COVID SEVERITY

Clusters	Cluster Title	Total_cases	population	Case_Percentage
0	Park/Beach Cluster	3966.461538	31815.000000	11.961538
1	Asian Restaurant Cluster	5440.966102	49211.305085	10.645932
2	Grocery Cluster	4736.363636	30971.590909	12.121364
3	Bar & Food Cluster	4390.061947	41132.902655	9.648230
4	Public Service Place Cluster	537.000000	8624.000000	32.690000
5	Road/Trail Cluster	2413.250000	25956.750000	6.290000

DISCUSSIONS AND RECOMMENDATIONS

- I have clustered LA county neighborhood regions based upon top 3 most common surrounding venues and superimposed those clusters on LA county map showing impacted population across various regions. This is very helpful to understand severely impacted regions and most common venues around those regions. Similar study can be extended to other locations of the world

- In my study, we could see that elbow point for k values is at 3, which again rises and then stabilizes at the value of 6. Hence I have used k=6 (number of clusters) for this analysis. In study above, I have used canberra method to find the distance between two points. This data can be analyzed using other methods as well

- Additionally, I have used static data to pull COVID cases information in the world, LA county list as well as COVID cases information from LA county portal. This data can be fetched dynamically in future studies to retrieve up to date information at any point of time. Also, due to higher size of value categories, I defined venue category type and placed similar venues together in a single bucket in my analysis. In order to get more detailed view of analysis, this analysis can be extended at more granular level (without bucketing multiple venue into one) which will give more detailed results of COVID impact

SUMMARY

- My aim for this study was to identify most common venues around one of the most severely impacted region by COVID disease in the world. In my initial study, I identified USA as the most impacted country in all aspects. After that, I identified California as one of the most severely impacted and most populated state of USA. Hence, I focused my analysis on this state and further identified Los Angeles as most impacted county of California state
- After that, I explored Los Anageles county further and identified most common venuse across various neighborhood regions of this county. Later I used K-means clustering to create clusters of top three most common venues category types around each region. Finally, I superimposed those clusters on LA county choropleth map, reflecting severity of COVID in terms of percentage of impacted population

CONCLUSION

- USA is the most impacted country of the world by COVID disease in all aspects (total cases, Number of cases by 100k population, cases in last 7 days..etc)
- California, being the most populated state, is one of the most adversely impacted state of USA by COVID
- Los Angeles is the most impacted county of California, and hence should be focused for further analysis
- Most parts of Los Angeles have mixture of various social venues, mostly led by various continental restaurants (Asian, American & Mexican restaurants to be precise)
- Cluster 4 is the least populated cluster and unlike other clusters, this doesn't have many social venues. Here most venues are primarily public service places . Although total number of cases are less here due to less population, in terms of impacted population percentage, it is the worst impacted cluster and hence needs to follow very strict isolation rules
- Cluster 2 with most grocery places and cluster 1 with Asian and other Restaurants are in populated regions and also requires strict isolation rules to reduce the spread
- Cluster 0 has more public gathering places like parks and beaches and cluster 3 has more bars/pubs and food places. These two clusters are also candidates for moderately strict regulation rules but can have necessary businesses open with precautions
- Cluster 5 is the least impacted cluster and even though being crowded place, it does not have many social venues nearby except a few Trails and Sport shops and hence least impacted by COVID. This can follow comparatively lighter rules

This study will help our target audience to review the impact of COVID across various regions of LA and decide the best isolation strategy

Clusters	Cluster Title	Total_cases	population	Case_Percentage
0	Park/Beach Cluster	3966.461538	31815.000000	11.961538
1	Asian Restaurant Cluster	5440.966102	49211.305085	10.645932
2	Grocery Cluster	4736.363636	30971.590909	12.121364
3	Bar & Food Cluster	4390.061947	41132.902655	9.648230
4	Public Service Place Cluster	537.000000	8624.000000	32.690000
5	Road/Trail Cluster	2413.250000	25956.750000	6.290000

REFERENCES

Some of the key references used for this study are mentioned below -

[1]. WHO COVID Dashboard

[2]. Worldometer COVID Dashboard

[3]. LA County COVID Dashboard

[4]. LA County Regions Location Coordinates

[5]. Foursquare API