The Battles of Neighborhoods

1. Introduction

1.1 Background

The current covid 19 pandemic has affected almost all 200 countries. Several countries have been in lockdown since March 2020 and are now planning to ease the lockdown. In this project we will focus on India. India has been in a lockdown since March 25th and is planning to lift the lockdown in zones where the number of cases is less. Lockdown remains in place in containment zones (these zones are decided by each individual state based on the number of active cases).

1.2 Problem

We look at the problem of deciding which zones to open based on the number of cases. We also take into consideration the number of different venues (restaurant, shops, malls, etc.). This problem needs to be solved carefully for a state to design a model which can then be used across other states and help lifting the lockdown effectively and deciding which venues to open.

1.3 Interest

In general, people might be interested in analyzing the visualization of active cases in each district. Shop owners might find this data useful for deciding whether it is safe to open their shops/businesses. And finally, the data analysis might be helpful for the government in making a decision.

2. Data Acquisition and Cleaning

2.1 Data Sources

We will scrape data from Wikipedia for district wise coronavirus cases in Karnataka (https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Karnataka). We are interested in the district and total active cases column. We use google API to get the latitude and longitude of each district. We use foursquare API to get venue details in each district.

2.2 Data Cleaning

Since the sample size of the data from Wikipedia is small (30 rows), there wasn't much data to perform univariate or bivariate analysis. We had to drop 'sl. No.' column and a row containing total values for each column.

Foursquare API did not return venue details for all 30 districts, venue details for 18 districts were returned so we had to filter/drop 12 districts.

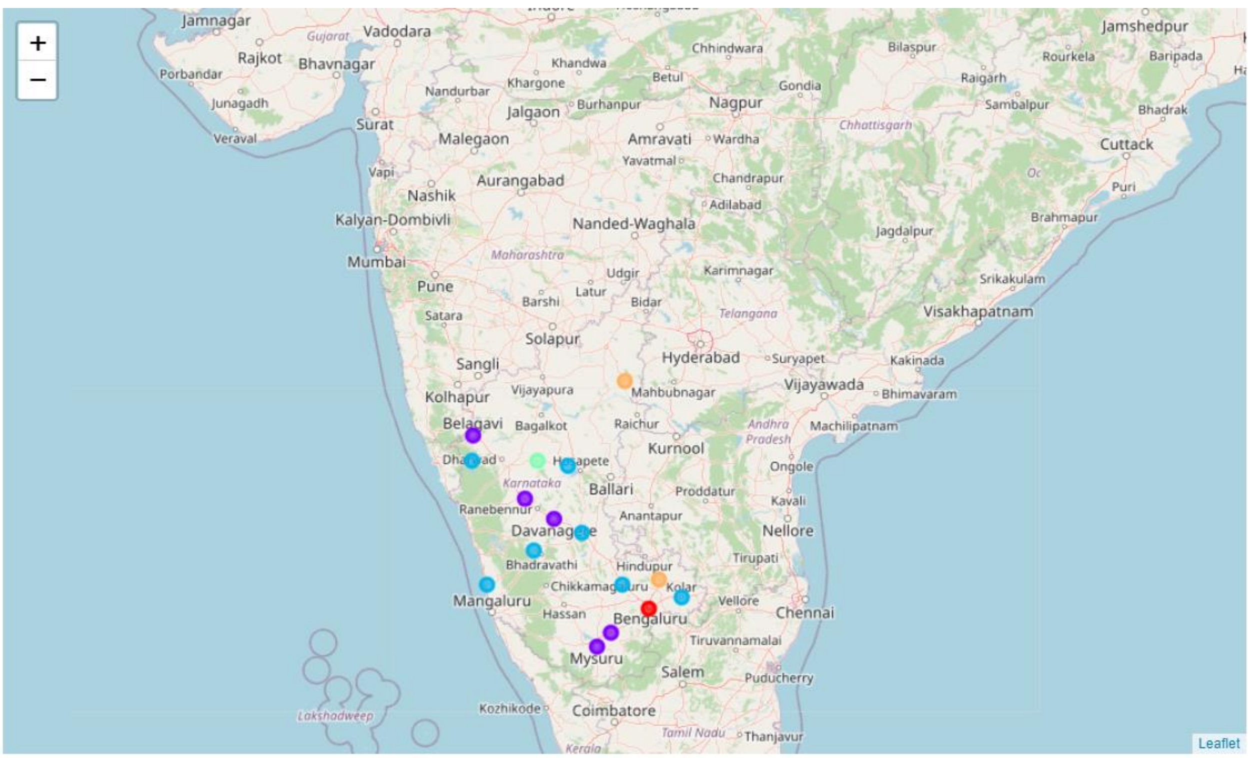
3. Data Analysis

3.1 Venue details

We used foursquare API to get venue details. As mentioned earlier, venue details were returned for 18 out of 12 districts. Latitude and longitude details for these districts were then obtained using google API.

3.2 Machine Learning

We used K means to cluster districts based on the number of active cases and venues. Color coding was used to distinguish clusters based on the number of active cases.



3.3 Merging data

Latitude, longitude, cluster labels and venue details data were merged based on each district to obtain final dataset for analyzing clusters.

4. Results

Districts were divided into 5 clusters and the number of venues in each cluster were analyzed. In each cluster we looked at 6 most common venues.

```
#Cluster 1
kar_merged.loc[kar_merged['Cluster Labels'] == 0, kar_merged.columns[[1] + list(range(5, kar_merged.shape[1]))]]
```

	Latitude	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
1	12.945142	Fast Food Restaurant	Coffee Shop	Café	Wings Joint	Train Station	Grocery Store

```
#Cluster 2
kar_merged.loc[kar_merged['Cluster Labels'] == 1, kar_merged.columns[[1] + list(range(5, kar_merged.shape[1]))]]
```

	Latitude	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	15.857267	Shopping Mall	Italian Restaurant	Bakery	Multiplex	Wings Joint	Café
2	43.439170	Hotel	French Restaurant	Tapas Restaurant	Surf Spot	Stadium	Pizza Place
5	14.466127	Hotel	Residential Building (Apartment / Condo)	Pizza Place	Train Station	Art Museum	Café
7	14.787482	Hotel	Clothing Store	Bus Stop	Andhra Restaurant	Chinese Restaurant	Grocery Store
12	14.168827	Filipino Restaurant	Convenience Store	Coffee Shop	Bar	Tea Room	Italian Restaurant
11	12.305183	Palace	Chinese Restaurant	Andhra Restaurant	Art Museum	Historic Site	Shopping Mall
10	12.523889	Bus Station	Train Station	Indian Restaurant	Indie Movie Theater	Pharmacy	Snack Place

```
#Cluster 3
kar_merged.loc[kar_merged['Cluster Labels'] == 3, kar_merged.columns[[1] + list(range(5, kar_merged.shape[1]))]]
```

	Latitude	1th Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
6	15.426365	Other Repair Shop	Wings Joint	Historic Site	French Restaurant	Food	Filipino Restaurant

5. Observation

According to the analysis cluster 2 has the greatest number of venues and least number of active cases. These are the cluster marked in blue color depicting safe zones. Based on the data we can say that it safe to open businesses in cluster 2.

6. Conclusion

In this analysis we have looked at district wise active cases and analyzed whether the venues in these districts should be opened or not. In future work we can look at ward (each district is divided into smaller units called wards). Foursquare API did not return a lot of venues and did not have any details for 12 districts. In a follow up analysis, we can get venue details from a different database (for e.g. Google).