

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.

2. Data

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.

	Sl. No.	District	Positives Reported on 27 May 2020	Total positives	Discharges Reported on 27 May 2020	Total discharges	Total active cases	Deaths	Non-Covid Deaths
0	1	Bengaluru Urban	6	282	2	151	120	10	1
1	2	Mandya	1	255	0	28	227	0	0
2	3	Kalaburagi	28	185	2	75	103	7	0
3	4	Yadagiri	16	156	0	0	155	1	0
4	5	Belagavi	4	146	0	89	56	1	0

2.2 Data Cleaning

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

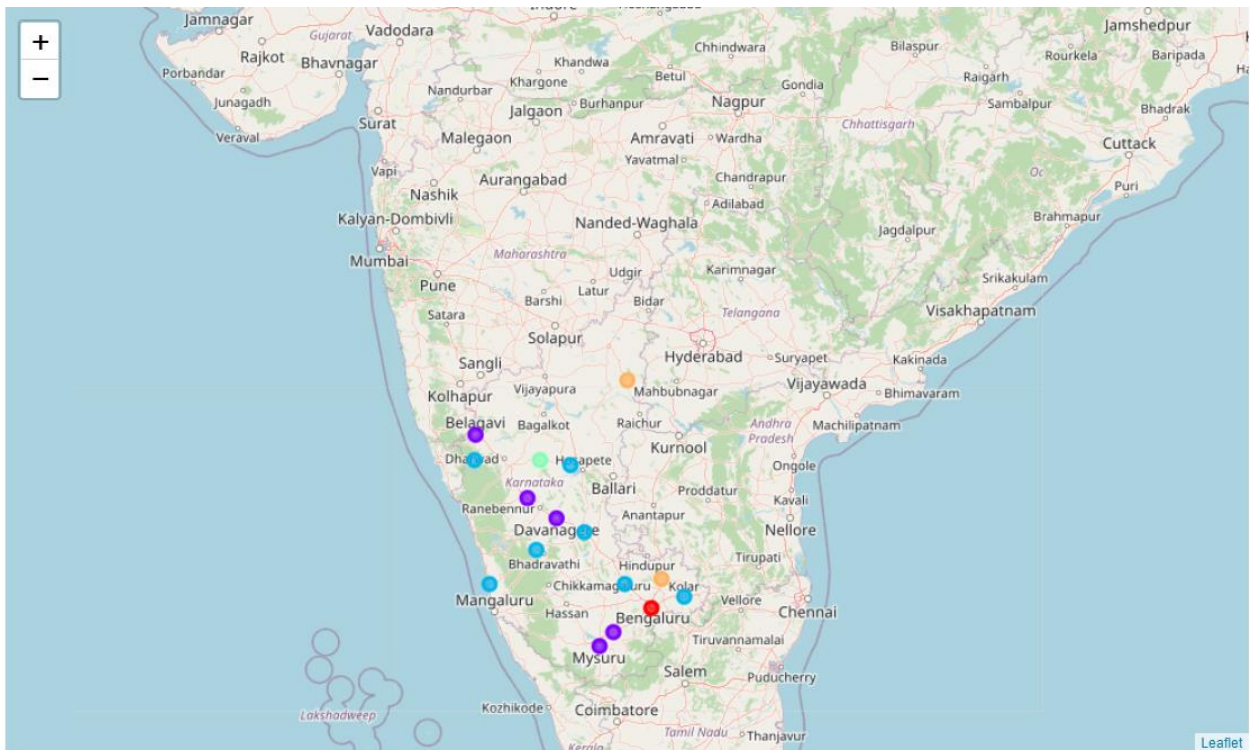
3. Problem Solution using Data

3.1 Venue details

We will use foursquare API to get venue details in each district. Latitude and longitude details for these districts will be obtained using google API.

3.2 Machine Learning

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



3.3 Merging data

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

4. Target Audience

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.

