PREDICTING COVID-19 INTENSIVE ZONES IN DELHI USING NEIGHBOURHOOD CLUSTERING

ANAND GANGADHARAN
Student, Vellore Institute of Technology, Vellore, India

ABSTRACT
As the world tackles the Covid-19 Pandemic, carefully constructed plans are the need of the hour. Lockdowns are holding back civilians from working and earning a living, and in such a situation, where economies fall and countries will be heading towards recessions, people might be forced to head out and work and live along with the virus. Therefore, mitigation strategies to stop the spread of the virus are an urgent need. This plan shouldn’t be a general one, but should be personalized according to the demographics to be really efficient. To foresee the effects of lifting lock-downs, my research will help people understand which neighbourhoods might see a surge in Covid-19 cases and which ones would show a similar trend. I will be using population density data and location data of popular venues (like popular market places) to estimate the interaction occurring at a particular place. The neighbourhoods will be clustered into groups so that common and effective strategies could be built to handle similar places. I will be focusing on the predictions mainly in my city (Delhi, the capital of India).

KEYWORDS: Covid-19, Delhi, K-Means, Clustering, Location Data, Population Density

INTRODUCTION
Background
Covid-19 is an infectious disease that has caused havoc in the modern world by disrupting the daily life of every civilian on earth. The virus, which primarily affects the lungs, has infected 42,745,971 people globally and 1,152,463 people have lost their lives because of it (as of 24th October, 2020). The pandemic has forced world leaders to come up with stringent measures such as nation-wide lockdowns to curb the spread of the virus. But lockdowns are holding back civilians from working and earning a living, and its effect does not trouble the poor but every section of the society, even the governments themselves. In such a situation, where economies fall and countries will be heading to recessions, people might be forced to head out and work and live along with the virus. This causes a huge risk to countries like mine, India, with an enormous population density.

PROPOSED SYSTEM
To predict the COVID Intensive Zones in Delhi, we shall perform the following steps:

- Find the popular venue categories of each neighbourhood
- Form clusters of the neighbourhoods according to the prevalent venue categories using KMeans Clustering.
- Superimpose a layer of a choropleth map of the population densities in each district to rightly predict the severity of the possible situation

DETAILS OF DATASET

To solve the above problem, I have made use of the following data:

Neighbourhoods of Delhi (and the basis of their segmentation)

There are 9 districts in Delhi and the neighbourhoods are aggregated into one of these districts.

Please note this data is according to Census 2011 which declares 9 districts as opposed to 11 districts as per the current situation. Therefore, I will be adding places in SOUTH EAST DELHI under SOUTH DELHI and places in SHAHDARA under EAST DELHI.

This data would be used to define the neighbourhoods when I will be plotting the results on the map. I used the help of the ‘search’ option in Google Maps to figure out the approximate coordinates of each neighbourhood and created my own dataset.

Distri\text{butions in Delhi}

<table>
<thead>
<tr>
<th>District</th>
<th>Neighborhood</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>North West Delhi</td>
<td>Adarsh Nagar</td>
<td>28.719390</td>
<td>77.173270</td>
</tr>
<tr>
<td>North West Delhi</td>
<td>Ashok Vihar</td>
<td>28.687260</td>
<td>77.177689</td>
</tr>
<tr>
<td>North West Delhi</td>
<td>Azadpur</td>
<td>28.712997</td>
<td>77.177360</td>
</tr>
<tr>
<td>North West Delhi</td>
<td>Bawana</td>
<td>28.797247</td>
<td>77.048331</td>
</tr>
<tr>
<td>North West Delhi</td>
<td>Begum Pur</td>
<td>28.726457</td>
<td>77.064246</td>
</tr>
</tbody>
</table>
Location Data of Popular Venues — Provided by Foursquare API

The Foursquare API provides data related to the frequency of occurrence of different venues at a particular place. I will be using the location data to identify clusters where the proximity of venues with high footfall (like markets) are present. These will be identified as hotspot neighbourhoods.

<table>
<thead>
<tr>
<th>Neighborhood</th>
<th>Neighborhood Latitude</th>
<th>Neighborhood Longitude</th>
<th>Venue</th>
<th>Venue Latitude</th>
<th>Venue Longitude</th>
<th>Venue Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adarsh Nagar</td>
<td>26.71939</td>
<td>77.173270</td>
<td>My idea Store</td>
<td>26.717487</td>
<td>77.170922</td>
<td>Mobile Phone Shop</td>
</tr>
<tr>
<td>Pahar Ganj</td>
<td>26.71930</td>
<td>77.173270</td>
<td>Pahar Ganj</td>
<td>26.719681</td>
<td>77.173468</td>
<td>North Indian Restaurant</td>
</tr>
<tr>
<td>Adarsh Nagar</td>
<td>26.71939</td>
<td>77.173270</td>
<td>Adarsh Nagar Metro Station</td>
<td>26.719586</td>
<td>77.170406</td>
<td>Light Rail Station</td>
</tr>
<tr>
<td>Ashok Vihar</td>
<td>26.59726</td>
<td>77.176808</td>
<td>Mich to Masala</td>
<td>26.587374</td>
<td>77.173744</td>
<td>Mughlai Restaurant</td>
</tr>
</tbody>
</table>

Neighbourhood wise venue data

It is noted that Foursquare API doesn’t have many details regarding venues in Delhi, therefore this is just an approximate picture of the real world. But we can surely say that it is a good approximation and will serve our needs.

Population density of various districts — Census 2011 data

The population density data would be used to mark districts with a potentially high risk of community transmission due to the high proximity in living conditions.

District Boundaries Data

I created a data set of my own having the polygon coordinates of each district in Delhi.

I made use of two websites -http://nominatim.openstreetmap.org/ and http://polygons.openstreetmap.fr/index.pyto first extract the location of a place of the map and then extract its coordinates respectively.

The two factors, population density and location venue data help us estimate the proximity of humans in a particular region and the amount of interaction between them. Thus, we will be able to understand the degree of spread of the virus at a particular region and therefore devise efficient strategies.
IMPLEMENTATION

I will be using population density data and location data of popular venues (like popular market places) to estimate the interaction occurring at a particular place and formulate and analyse a network. The neighbourhoods will be clustered into groups so that common and effective strategies could be built to handle similar places.

Now, I created a table containing the Top 2 venue categories found in each neighbourhood.
Neighbourhood wise top 2 venues in Delhi

According to the top venues found, I used KMeans Clustering to group similar neighbourhoods. I made use of the Silhouette Score Comparison to find the optimum value of K (i.e. the number of clusters).

Calculating the Silhouette Scores

The optimum value of K was found out to be 6, following which neighbourhoods were clustered into 6 clusters using KMeans Clustering. The result was then merged into a single Data Frame.

Neighbourhoods clustered according to the top 2 venues

Upon creating a Choropleth map using the district-wise population density data, I arrived at this.
RESULTS

In my project we look at a technique to classify areas into zones to facilitate analysis and development of strategies. The categorization occurs on the basis of two factors namely population density and location venue data. The neighbourhoods of Delhi will be categorized into 6 zones, each indicating a particular area according to the activities happening in them and intensity of spread in that area. The zones are colour coded to help us easily differentiate.
Predicting Covid-19 Intensive Zones in Delhi Using Neighbourhood Clustering

- Most Concerning — RED, YELLOW
- Moderately Concerning — PURPLE, BLUE
- Least Concerning — TURQUOISE
- Not Concerning — GREEN

Neighbourhood Clustering Attributes Were:
- Neighbourhoods predominantly having ‘Fast Food Restaurants’ and other socially active places. (Most Concerning) — Yellow Marker
- Neighbourhoods with housing complexes and day-to-day civilian activity zones. (Moderately Concerning) — Blue Marker
- Neighbourhoods with many daily life associated venues. (Moderately Concerning) — Purple Marker
- Neighbourhoods predominantly having ‘Indian Restaurants’. (Most Concerning) — Red Marker
- Neighbourhoods with venues of essential needs only. (Least Concerning) — Turquoise Marker
- Neighbourhoods with very few venues. (Not Concerning) — Green Marker

DISCUSSION

With the help of the final map, we can see which all districts are vulnerable to the spread of the virus and also which neighbourhoods in each district need to be taken special care of. Places with Red and Yellow markers (Most Concerning zones) especially in districts like North East Delhi, North Delhi, and Central can aggravate the situation if timely action is not taken. South West Delhi and North Delhi being a sparsely populated region should be the least concerning area in Delhi.

The following improvements could be worked upon in future studies:
- As mentioned earlier, the Foursquare API doesn’t provide very detailed results when used for collecting data about Delhi. Other Location Data providers could be utilized in further studies to result in more accuracy.
- Furthermore, rather than working with coordinates of neighbourhoods, if neighbourhood boundaries could be utilized and all venues within the boundaries are extracted, it would produce near-perfect results.
- Census 2011’s data was used in this project. The current year is 2020 which we could mean the statistics could be changed a lot. So working with the latest Census, as and when it is released could better the results.

CONCLUSIONS

Covid-19 has wreaked havoc all around the globe and almost half of the year has now been spent in just fighting off the pandemic, by all the people on earth. With better strategies and more efficient systems, we should be able to tackle this problem in a better way so as to be done with it as soon as possible, for not just the welfare of all the people but also to safeguard the economies. I hope this project of mine aids in the formation of mitigation strategies so that we all could win this battle and return to our normal lives at the earliest.
REFERENCES


Links that aided in dataset creation:


