

REVIVE KERALA 2020

Rehabilitation and recovery program
Covid -19 pandemic

BY: AMINA FYSEL

JULY 2020



TABLE OF CONTENTS

❖ INTRODUCTION

❖ BUSINESS CASE

❖ WORKING WITH DATA

❖ METHODOLOGY

❖ EXPLORATORY DATA ANALYSIS

❖ MODEL DEPLOYMENT

❖ RESULT AND DISCUSSION

❖ SUGGETIONS FOR FUTURE RESEARCH

❖ GITHUB URL

INTRODUCTION

It all happened in the blink of an eye, to put it metaphorically. Suddenly a virus made its home in the body of thousands of people around the world. The novel Corona virus started its journey from the city of Wuhan, China, according to official sources on December 31st 2019, which is why it is named as covid-19. On March 11, 2020, the World Health Organization declared the COVID-19 outbreak a pandemic. Although for most people COVID-19 causes only mild illness, it can make some people very ill. More rarely, the disease can be fatal. Older people, and those with pre-existing medical conditions (such as high blood pressure, heart problems or diabetes) appear to be more vulnerable. The virus that causes COVID-19 is mainly transmitted through droplets generated when an infected person coughs, sneezes, or exhales. These droplets are too heavy to hang in the air, and quickly fall on floors or surfaces. This project is going to emphasis on the life after covid-19 in Kerala, India. The state has received both national and international recognition for its efforts to contain the Coronavirus pandemic.

A recent study found out some major health concerns of the people who were affected both physically and emotionally. A large number of seriously ill coronavirus patients who survive and leave the hospital are facing a new and difficult challenge: Recovery. They are continuing to struggle with the residual symptoms even after discharge and some may persist for a long period of time. Lying in a hospital bed for such a long time has given rise to muscle weakness, fatigue, confusion and mental foggiess because the experience of being extremely sick and alone amplifies the trauma. It's really common for patients to have PTSD after going through this — nightmares, depression and anxiety because they're having flashbacks and remembering what happened. Another factor that can extend or hamper recovery is a phenomenon called hospital delirium, a condition that can involve paranoid hallucinations and anxious confusion. It is more likely to occur in patients who undergo prolonged sedation, have limited social interaction and are unable to move around — all common among COVID-19 patients.

Also, people are facing anxiety issues due to the social restrictions, many have lost their job and quarantine guidelines has made people emotionally drown. Returned emigrants from different parts of the world especially from gulf countries have even lost their livelihood and are at the verge of despair. So, to deal with such variety of issues everything has to be done just right since the emotional and physical health of our people are equally important.

BUSINESS CASE

The need of a well organised system that can bring back people to their normal life and ensure mental support has now become necessary. Rehabilitation is the process of helping an individual achieve the highest level of function, independence, and quality of life possible. It helps the person to develop and utilize their full potential through a process of learning and support from the environment which can in turn bring back the individual back in to their own community. This does not reverse the damage caused by the disease but rather helps restore the individual to optimal health. There are a lot of rehabilitation centres in our state but we need to especially concentrate on those people who are affected by the pandemic. So, the main objective of this dissertation will be to create a recommender system for a suitable location to set up the main branch of REVIVE in Kerala, India, which is a part of the Rehabilitation and recovery program. It would be suitable to choose a location near hospitals so that supporting facilities can be shared and a multi -disciplinary approach can be adopted for maximum profit.

Problem Statement:

Using Data Science methodology and machine learning algorithms this dissertation comes up with a solution for the business question:

“If a health professional or an investor is looking to open a rehabilitation centre especially due to the covid -19 pandemic, in Kerala, India, where would you recommend them to open it?”

Target Audience:

This project is particularly useful to property developers, investors, medical professionals, government officials etc. After the accomplishment of this project a group or perfect team of physicians, specially trained therapists, and nurses as well as psychologists, nutritionists, biomedical engineers are likely to be increasingly in demand. Since we are taking the hotspots and containment zones declared by the Government of Kerala, as the main matter of concern the recovered patients and also the emotionally affected ones can be benefitted through this.

WORKING WITH DATA

In order to solve the business problem, we have to come up with the best suited location to open the medical rehabilitation centre so that it can function with the optimal results. We need the following data to form insights accordingly.

Data Requirements	Source	Relevance
Cumulative summary district wise	Directorate of health service, Kerala	Visualizing most affected regions
Phsyco-social support data	Directorate of health service, Kerala	Find order of districts needing mental support
LSG's needing special attention	Directorate of health service, Kerala	Identifying hotspots on a map
Venues nearby hotspots	Foursquare API	Spotting best location for Rehab

Method of extraction:

All the necessary data regarding the covid-19 pandemic was available on Kerala government official website (<http://dashboard.kerala.gov.in/>). I prepared a few excel sheets out of the data extracted from the same. Namely “Kerala district wise breakdown” giving the current statistics of the ones who were physically affected, “Psychosocial data” giving an insight on the emotionally affected ones and finally the count of LSGs needing special attention which were defined to be the hotspots.

Then the geographical coordinates of these hotspots were found using the python geocoder package. Using the latitude and longitude coordinates we can visualize the hotspots on the map to easily find out the badly affected regions.

For the venue data, we use the Foursquare API. Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues in a radius and also corresponding coordinates, venue location and names. Since we are particularly interested in the hospital category, we will extract the names of nearby hospitals in order to solve the business problem put forward. We are mainly concentrating on the hospital category because after acute-stage care in hospitals, an intensive program of treatment and rehabilitation is necessary to assist the patient to recover quickly as possible. We are also giving equal importance for psychosocial support needed and planning to start certain state wide volunteer programs on the regions of most concern.

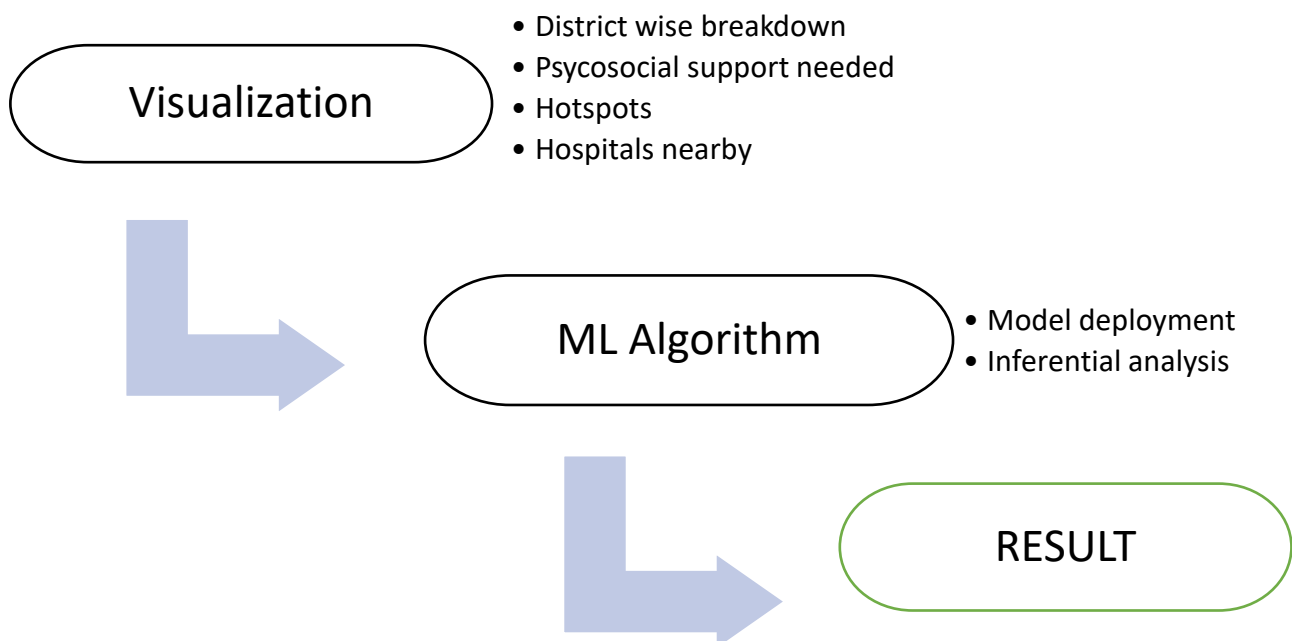
METHODOLOGY

The first and foremost thing to start with is to carefully analyse the data collected because it gives us initial insights and may help to get partial idea of the answers that we are looking to find out from the data.

- We will start with the main data and visualize the order of districts affected in terms of quarantined, active, recovery and death cases. From which we can confirm the regions which has to be noted with at most concern.
- Using the mental health data, we will categorize the types of discomfort people are facing. There was an increasing number of suicides reported due to unavailability of alcohol during the lockdown days. Hence this an important matter of concern.
- Using the count of hotspots per district we will use the folium package and explore them on the map. Python's geocoder package allows us to convert address into geographical coordinates in the form of latitude and longitude. This can provide us the areas to be considered while building the recommender model that can be of maximum benefit. Plotting them on a map allows us to perform a sanity check to make sure that the geographical coordinates data returned by geocoder are correctly identified.
- With the help of Foursquare API, we can get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the hotspots in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each hotspot and examine how many unique categories can be curated from all the returned venues.
- Then, we will analyse by grouping the rows by hotspots and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since, we are analysing the "Hospital" data, we will filter the "Hospital" as venue category.
- Finally, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning

algorithms and is particularly suited to solve the problem for this project. We will cluster the hotspots into 3 clusters based on their frequency of occurrence for “Hospital”. The results will allow us to identify those regions with least, moderate and highest number of hospitals including the containment zones, that can be incorporated with the insight we drew from the exploratory data analysis of the districts highly in need of a rehab.

FLOW CHART:



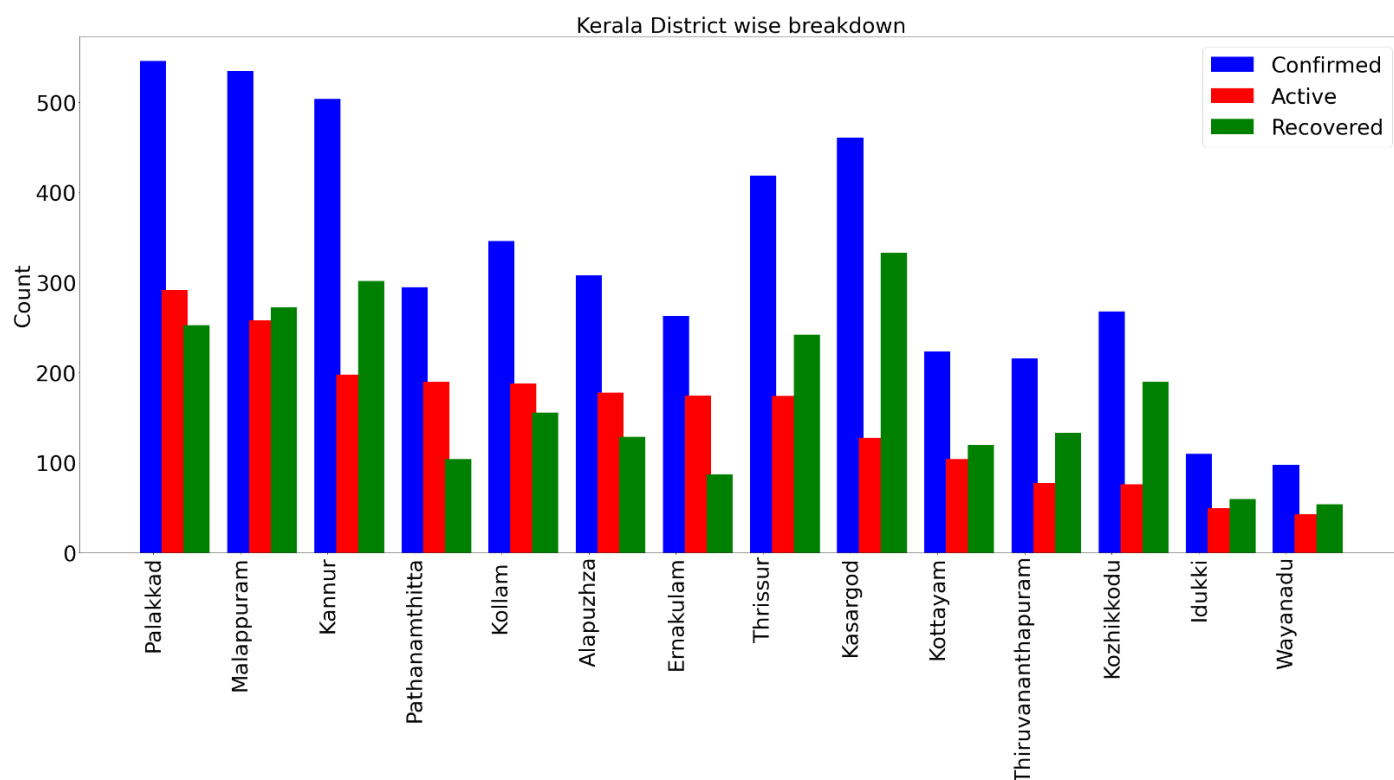
EXPLORATORY DATA ANALYSIS

Summary of Kerala:

After carefully analysing the official details provided by the Kerala Government, an excel worksheet in CSV format was prepared. Using python codes, the data was read into a panda's data frame and using matplotlib package a visual representation of “Kerala district wise breakdown “was obtained. The data frame and the bar chart look as follows;



	District	Quarantined	Confirmed	Active	Rcvrd	Death
9	Kottayam	7585.0	224.0	104.0	120.0	0.0
10	Thiruvananthapuram	13504.0	216.0	78.0	133.0	5.0
11	Kozhikkodu	17898.0	268.0	76.0	190.0	2.0
12	Idukki	4411.0	110.0	50.0	60.0	0.0
13	Wayanadu	3645.0	98.0	43.0	54.0	1.0



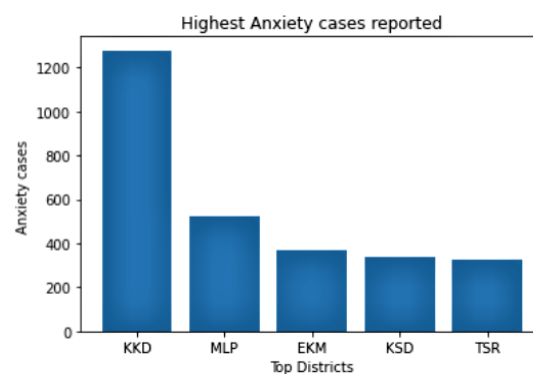
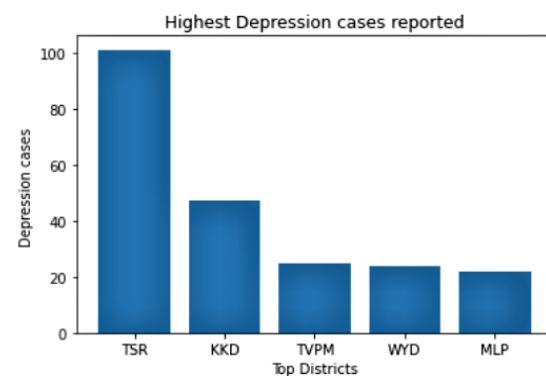
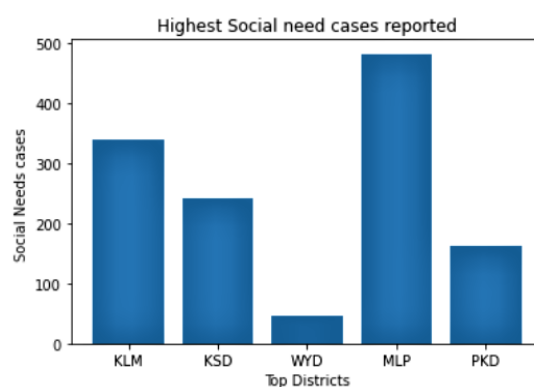
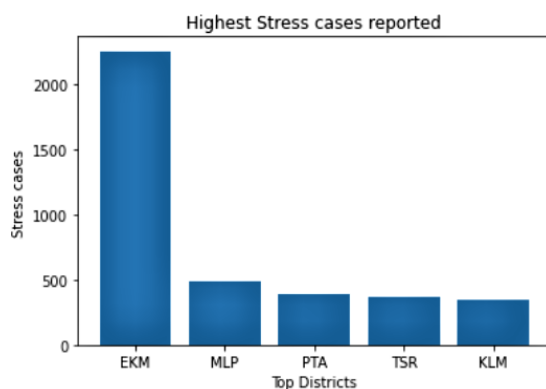
As of now, reflecting the migratory nature of districts too, confirmed cases are accumulating in northern districts. The spike in Palakkad is mostly on account of the influx from Tamil Nadu which is at present among the states having highest count of covid patients. From the southern districts Kollam, Alappuzha and Thiruvananthapuram are also gradually moving out of their comfort zones.

Order of districts in terms of patient count:

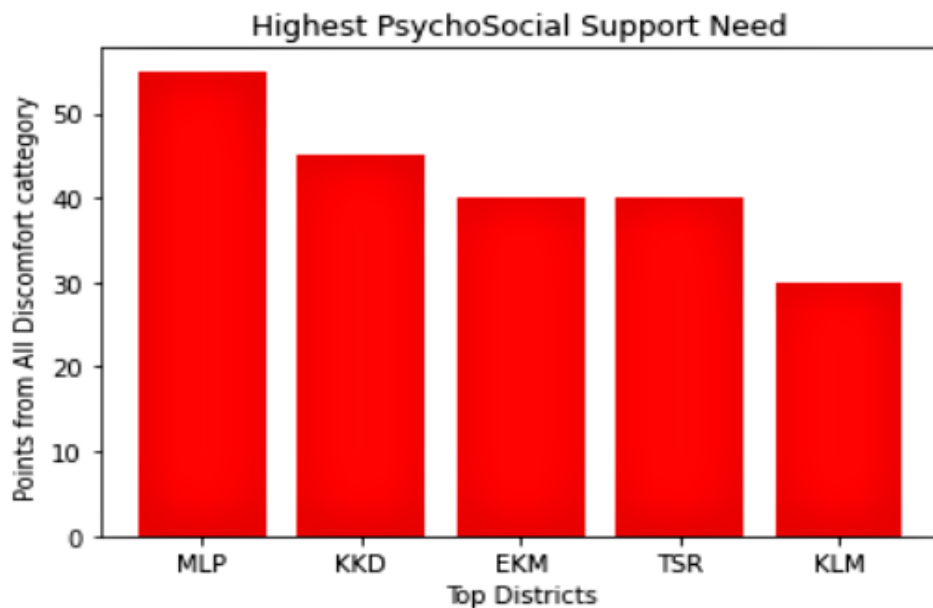
1. Palakkad
2. Malappuram
3. Kannur
4. Kasaragod
5. Thrissur

Phsyco - social support:

We were able to collect the total count of emotionally affected ones and with the help of some algorithms and python codes we were able to bring the types of discomfort and the count of phsyco -social support needed by the people district wise on different charts as follows.



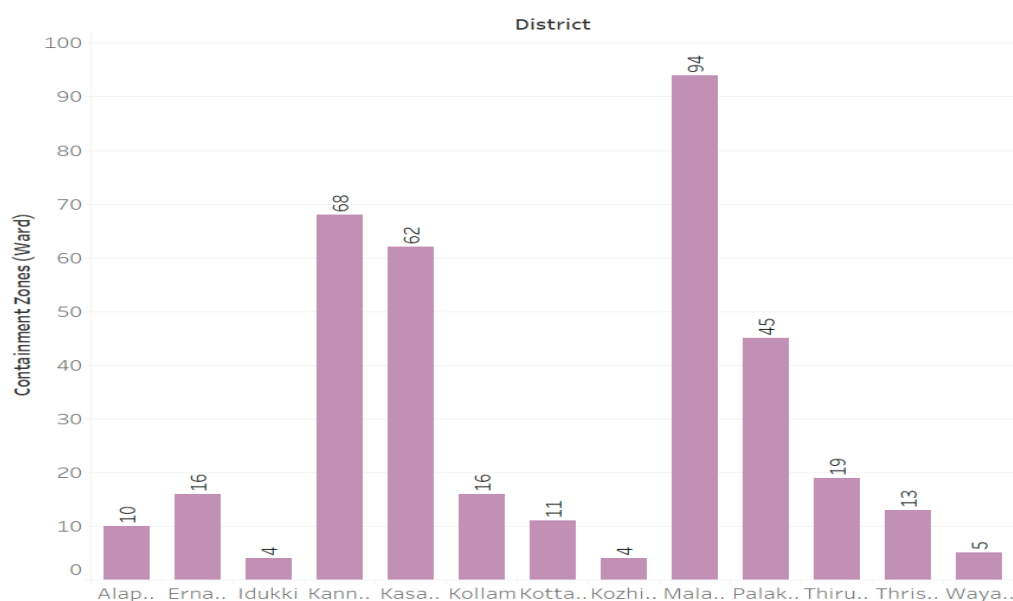
Analysing all the above charts we rank the districts according to their needs of a mental support and plot a graph as below;



We can now see that Malappuram (MLP) district has the highest number of people who can be benefitted through our recovery program.

Count of Hotspots:

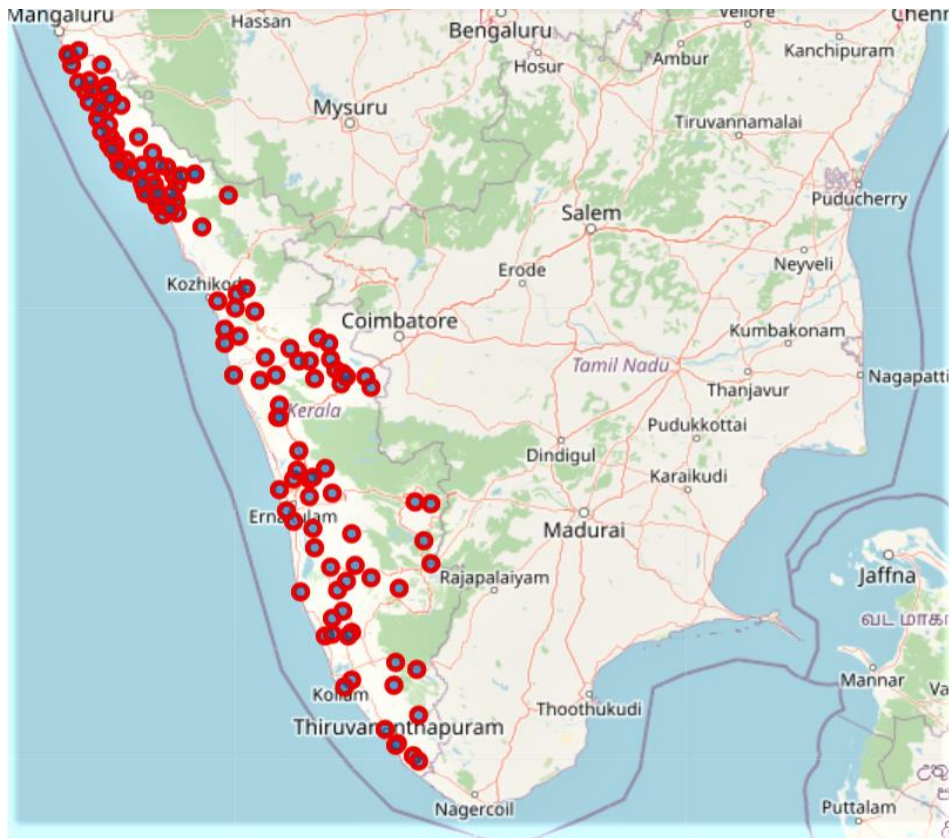
On 16th of April, 2020 the state has expressed reservations over the Centre's criteria for designating districts as COVID-19 hotspots and has recommended earmarking of worst-affected districts into zones. Curfew-like measures will be implemented in these hotspot areas while normal lockdown measures will be in place in other areas. We took the district wise count of hotspots and formed a bar chart using tableau,



We can again rank them depending on the count as follows;

1. Malappuram
2. Kannur
3. Kasaragod
4. Palakkad
5. Thiruvananthapuram

Let us now bring these hotspots on the map using python folium library. Given below are the hotspots identified on a map.

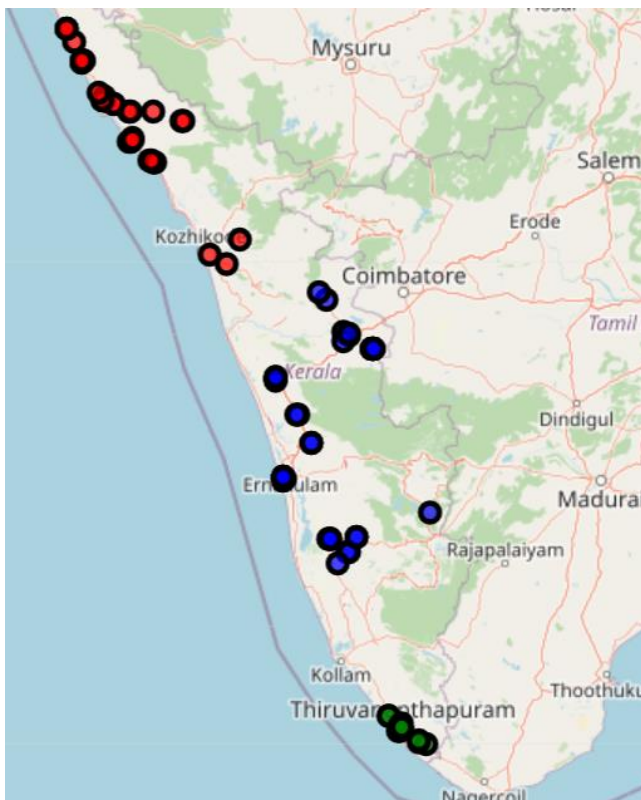


MODEL DEPLOYMENT

KMeans clustering algorithm:

KMeans algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

After performing the algorithm, we were able to categorize the hotspots into 3 clusters based on the frequency of occurrence of a hospital nearby.



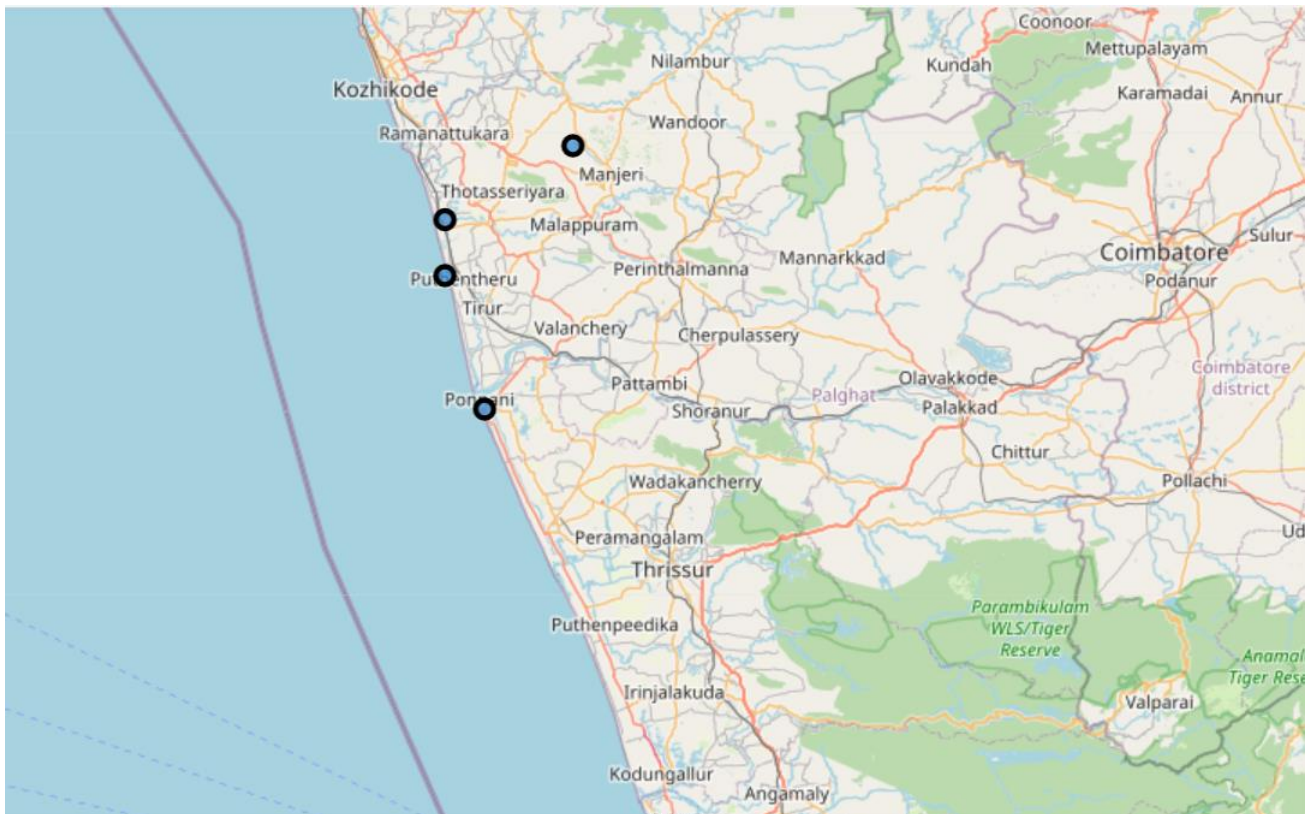
- Cluster 0: highest number of hospitals in the neighbourhoods
- Cluster 1: lowest number of hospitals in the neighbourhoods
- Cluster 2: moderate number of hospitals in the neighbourhoods

In red, green and blue respectively. The above map clearly indicates a high concentration of hospitals in the northern part of the state.

RESULT AND DISCUSSION

After deployment of the model, we found that cluster 0 had a high concentration of hospitals as compared to the rest of the clusters. So, the neighbourhoods in this cluster would be the most preferred location to set up a Rehab as choosing a location near hospital zones would provide maximum profit. Now to select a specific city from the cluster we can make use of the Exploratory Data Analysis, from which after visualizing the summary of the patient count, count of social support needed, and the count of hotspots, **Malappuram** district was in the top of the list. And fortunately, the hotspots clustered in Malappuram district happens to be in cluster 0.

Incorporating all these insights, this location will be a boon for the establishment of the project.



Since the other districts are also gradually growing with number of patients, setting up a sub branch can also be considered. The hotspots identified in the capital city; “Thiruvananthapuram” was found to have comparatively low number of hospitals in the neighbourhood irrespective of the growing positive cases. So, this can be a better option for developing a sub branch.

CONCLUSION

The recommender system for establishing a covid special rehabilitation centre in Kerala, India was successful in predicting an excellent location from where it can make maximum profit. we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data. The deployed system used information about the count of hospitalized and quarantined individuals, along with the help of Foursquare API in leveraging the location data regarding the hotspots. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding inappropriate areas in their decisions to open a new medical rehab.

SUGGESTIONS FOR FUTURE RESEARCH

The research for this project was done during the ongoing pandemic in the state. All the datasets were collected till 5th of July 2020. We have only considered the frequency of occurrence of hospitals nearby the covid hotspots. Unfortunately, data regarding covid special treatment hospitals was unable to find on the web and to extract the district wise count of those hospitals was a much difficult task. Also, the count of foreign return emigrants would have played a major role in deciding location for the rehab centres, as their return caused a major hike in the current scenario. Future research could devise a methodology to estimate such data to be used. The results of such a study can throw light on the intensity of the epidemic in the future and will help the government administrations in Kerala to formulate effective measures and policy interventions to curb the virus in the coming days.

GITHUB URL

Link to the project code is given below

<https://github.com/aminafysel/IBM-COURSERA-CAPSTONE-PROJECT/blob/master/IBM%20Capstone%20final%20project.ipynb>