**COVID-19 India: Understand Your Neighbourhood**

***Introduction:***

*About COVID-19: -*

*Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus.*

*Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.*

*The best way to prevent and slow down transmission is be well informed about the COVID-19 virus, the disease it causes and how it spreads. Protect yourself and others from infection by washing your hands or using an alcohol-based rub frequently and not touching your face.*

*The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so it’s important that you also practice respiratory etiquette (for example, by coughing into a flexed elbow).*

*At this time, there are no specific vaccines or treatments for COVID-19. However, there are many ongoing clinical trials evaluating potential treatments.*

***Project Overview: -***

*Project gives an overview of COVID-19 cases in entire India till date. The number of confirmed cases, number of cured people and number of deaths due to COVID-19. Also, the information backtracks to March-2020, since the time there was sudden rise in cases. Maps are shown to highlight all the cases across India and sub level to states and to districts too.*

*Project will also take a look to census data of India which was taken in year 2011 and will compare the various factors that might be contributing to the rise in cases in particular state or a district.*

*Each state, further divided to district level, will be clustered using "K-Means Clustering", so that similar states/districts are clustered and any action or decision should impact on the cluster as a whole.*

*People/users of the projects will be able to visualize the impact of COVID-19 on India/each state/ever district within. And they will also be able to get the most common visited places within a state/district using "Foursquare Location Data". So, user can check and decide the place they are visiting and can prepare/act accordingly.*

***User and Stakeholders: -***

*Users can be anyone who would like to know more about COVID-19. It's spread in India, or a particular state or district.*

*People who are visiting a certain place and wants to know the places where can get high volume of crowd due to its popularity and number of people visiting.*

*- Common people of India.*

*- Government of India.*

*- Visitors to any state/district of India.*

***Data***

*Data being used in the project is from multiple sources.*

* *This project will require daily data of COVID-19 in India, sub-level to district. For this we will use website scarping. (Source- https://www.grainmart.in/news/covid-19-coronavirus-india-state-and-district-wise-tally/)*
* *Further, we will use census data 2011 of India to compare the factors they are/or contributing to rise in cases. (Source- https://censusindia.gov.in/2011-Common/CensusData2011.html)*
* *We will also require COVID-19 data since march, for this we will download and use a dataset available online. (Source- https://prsindia.org/covid-19/cases)*
* *Further, to provide the most common places visited in an area and provide location information. (Source- Foursquare Location Data)*

***Methodology***

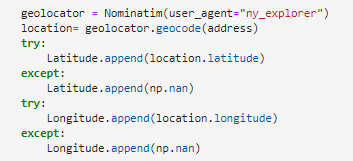
*In this project, many libraries are being used for different purposes, some of them are - pandas, numpy, requests, matplotlib, sklearn, scipy, bs4 and folium. Each helps to attain certain functionality in the project.*

*Data is extracted from sources and then further processed using pandas DataFrames.*

* *"bs4" is used for website scraping.*
* *"requests" to send and receive response to website. Also is used to send request to download files from a website.*

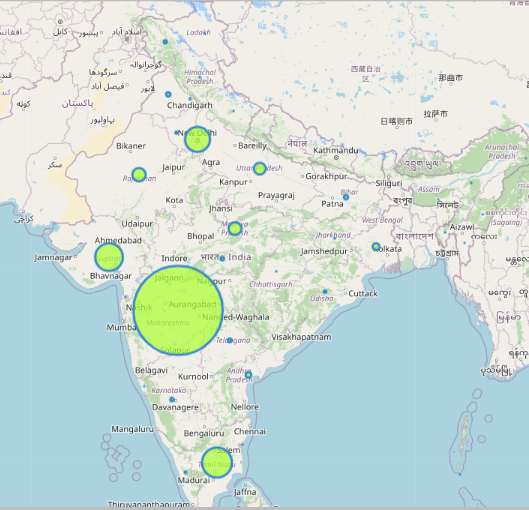


* *"geopy" is used to get geographical latitudes and longitudes of certain state/district.*

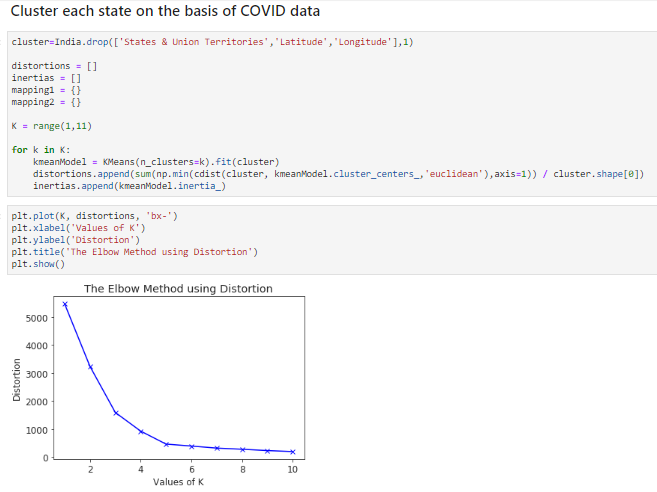


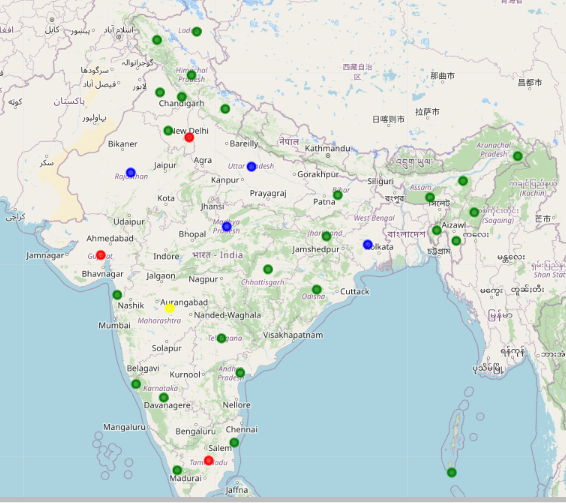
* *"folium" is used to map all the related lat and lng for diff areas and popup label accordingly on the maps.*





* *"K-Means Clustering" is used to cluster states/district on the basis of COVID-19 cases/cured/deaths in an area or on the basis of census data wherein factors like population, literacy rate, workers and non-workers, rural/urban households provide some insight on the rise of these cases in an area.*





***(Number of clusters to be used in "K-Means" is based on elbow method, calculated using Euclidean distance.)***

* *Another feature of the project is, it asks user to input any state he/she wants to explore further to check clusters and venues on that state.*
* *Therein, "Foursquare Location Data" is used to provide all the venue related information.*



***Results***

*Project successfully cluster all the states on the basis of COVID-19 and census data. We can compare both the results and then make policies/actions/discussions to be implemented on a cluster as a whole within the country.*

*Also, users are able to visualize all the data, COVID-19 impact on states and districts and further all the venue in a state/district of their choice.*

***Discussion***

*This infers that project can cluster states/districts and then further provide data related to venues. We can now move forward in examining the impact of factors that were used in census data to know more about the rise in cases and more factors contributing to it.*

***Conclusion***

*This will surely help people to know about their neighbourhood and take precautionary measures. More work can be done by analysing the data and predict impact of COVID on certain cluster of states.*