Fighting COVID-19 using Data Science.

1. Introduction:

1.1. Background:

Coronavirus has now infected more than four million people globally, according to data collected by Johns Hopkins University. Many countries have been in lockdown since at least March, but some are beginning to ease restrictions. That must be done carefully, the World Health Organization warns, otherwise it risks a resurgence of infections. Nonetheless, billions remain largely at home, and many are struggling with the economic and social consequences. India remains in lockdown, first introduced on 24 March, but some restrictions have been eased. Around 122 million are believed to have lost their jobs in April, and many say they will starve if they cannot work. India's worst affected areas include the financial hub of Mumbai, capital New Delhi, southern state of Tamil Nadu and the western state of Gujarat. The nationwide lockdown, the world's largest affecting 1.3 billion people, was on Monday extended until the end of May.

1.2. Business Understanding/Problem Description:

India says the number of total recoveries has outstripped active Covid infections for the first time. The health ministry said data showed that more people had been discharged than new infections recorded. The news is being cautiously welcomed in local media, but it comes amid concerns that picture is bleaker. A significant spike in infections in recent weeks has begun taking a toll on the healthcare system - and though India has ramped up testing, it is not uniform across the country, with some states testing much more than others.

To reduce the load on hospitals, the government of India intends to make COVID-19 tests easily available to the public, by opening temporary clinics for testing. For this project, we will be restricting ourselves to the worst affected region of Maharashtra: Mumbai. Our goal will be to identify the perfect locations for these test clinics, to ensure the ease of accessibility for the citizens of Mumbai. These clinics will be set up in locations which are devoid of Hospitals nearby.

Note: Certain assumptions have been made in this project. One of which being that every hospital in Mumbai provides the test for COVID-19.

1.3. Target Audience:

This project tends to serve the citizens of Mumbai, but the project can be extended to the entirety of India.

2. Data:

- Mumbai Neighbourhood Data: (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai)
- WHO COVID-19 Data: (https://covid19.who.int/info)
- India COIVID-19 Data: (https://github.com/imdevskp/covid-19-india-data/blob/master/complete.csv)
- Hospitals in Mumbai: (using Foursquare API)

2.1. Mumbai Neighbourhood Data:

This dataset contains all the neighbourhoods in Mumbai. It was scraped from the mentioned Wikipedia page. It has the following columns:

- Area
- Location
- Latitude
- Longitude

Below is a snapshot of the dataset:

	Area	Location	Latitude	Longitude
0	Amboli	Andheri,Western Suburbs	19.129300	72.843400
1	Chakala Andheri,	Western Suburbs	19.111388	72.860833
2	D.N. Nagar	Andheri,Western Suburbs	19.124085	72.831373
3	Four Bungalows	Andheri,Western Suburbs	19.124714	72.827210
4	Lokhandwala	Andheri,Western Suburbs	19.130815	72.829270

2.2. WHO COVID-19 Data:

Dataset Description:

WHO Coronavirus Disease (COVID-19) Dashboard
Data last updated: 2020/6/27, 11:25am CEST

Back to top

Data Download

Daily aggregate case and death count data is available for download as a comma-separated values (CSV) file: Download Data

Users should note that, in addition to capturing new cases and deaths reported on any given day, updates are made retrospectively to correct counts on previous days as needed based on subsequent information received.

Table 1. Daily aggregate case and death count data dictionary

Short field name	Type	Description
Date_reported	Date	Date of reporting to WHO
Country_code	String	ISO Alpha-2 country code
Country	String	Country, territory, area
WHO_region	String	WHO regional offices: WHO Member States are grouped into six WHO regions -- Regional Office for Africa (AFRO), Regional Office for the Americas (AMRO), Regional Office for South-East Asia (SEARO), Regional Office for Europe (EURO), Regional Office for the Eastern Mediterranean (EMRO), and Regional Office for the Western Pacific (WPRO).
New_cases	Integer	New confirmed cases. Calculated by subtracting previous cumulative case count from current cumulative cases count.*
Cumulative_cases	Integer	Cumulative confirmed cases reported to WHO to date.
New_deaths	Integer	New confirmed deaths. Calculated by subtracting previous cumulative deaths from current cumulative deaths.*
Cumulative_deaths	Integer	Cumulative confirmed deaths reported to WHO to date.

** See "Daily aggregate case and death count data" above for further details on the calculation of new cases/deaths.*

After pre-processing, the result is:

	Date	Month	Country	Country Name	Region	Confirmed	Cumulative Confirmed	Deaths	Cumulative Deaths
0	2020-02-24	February	AF	Afghanistan	EMRO	1	1	0	0
1	2020-02-25	February	AF	Afghanistan	EMRO	0	1	0	0
2	2020-02-26	February	AF	Afghanistan	EMRO	0	1	0	0
3	2020-02-27	February	AF	Afghanistan	EMRO	0	1	0	0
4	2020-02-28	February	AF	Afghanistan	EMRO	0	1	0	0

2.3. India COIVID-19 Data:

This dataset was compiled after scrapping and cleaning data on covid-19 in India from https://www.mohfw.gov.in/ website and API provided by https://www.covid19india.org/.

Note: This is pre-prepared data hosted on GitHub by imdevskp: (https://github.com/imdevskp)

Below is a snapshot of the dataset:

	Date	Name of State / UT	Latitude	Longitude	Total Confirmed cases	Death	Cured/Discharged/Migrated	New cases	New deaths	New recovered
76	2020-03-09	Maharashtra	19.7515	75.7139	2	0	0	0	0	0
88	2020-03-10	Maharashtra	19.7515	75.7139	5	0	0	3	0	0
100	2020-03-11	Maharashtra	19.7515	75.7139	10	0	0	5	0	0
113	2020-03-12	Maharashtra	19.7515	75.7139	11	0	0	1	0	0
126	2020-03-13	Maharashtra	19.7515	75.7139	11	0	0	0	0	0

2.4. Hospitals in Mumbai:

Foursquare API was employed to generate a dataset consisting of a list of hospitals in a 5km radius of every neighbourhood in Mumbai.

Below is a snapshot of the dataset:

	Name	Latitude	Longitude
0	Shivam Netram Eye Hospital	19.127722	72.843875
1	Ladies Clinic & Hospital	19.131188	72.841778
2	Ananda hospital	19.127574	72.847758
3	RG hospital	19.126204	72.839954
4	Shushurut Hospital	19.123895	72.842927