**OPTIME.**
optime.software

# Document

May 2020

Any Questions? Don’t hesitate to contact us:
**Vishakh Arora -** vishakh.arora29@gmail.com
**Chaitanya Nookala -** chaitanya.nookala@gmail.com
**Shlok Mehrotra -** shlokmehrotra@gmail.com
**Shubham Kumar -** kumar.shubham5504@gmail.com

# Problem

With the COVID-19 pandemic maturing well beyond the mere infancy stages of what was anticipated, it is of utmost importance to follow guidelines set forth by health organizations in this time of uncertainty and misinformation. Words simply cannot put into perspective the ramifications of this outbreak. There are currently over three million cases internationally and over two hundred thousand deaths attributed to the virus[1].

While the officials in healthcare organizations are indeed doing everything in their power to ensure the welfare of people, the virus continues to propagate through communities, yielding unprecedented effects. Grocery stores and pharmacies are running low on previously accessible and essential products, such as toilet paper and hand sanitizer, and therefore cannot meet the demands of distressed shoppers. Citizens are simply unaware of how to obtain necessary resources while limiting the spread of this deadly virus.

Currently, all guidelines revolve around the singular idea of social distancing and there has been minimal research with the objective of pinpointing other factors, such as those pertaining to environmental conditions. Although social distancing is undoubtedly among some of the most effective preventative measures being taken to slow the spread, it is not entirely a reasonable expectation. People still need to leave their homes to shop for essential items, get in some physical exercise and complete other tasks, and it is amidst these errands that one is most susceptible to the virus . Unfortunately, there are no concrete specifications for this. While online shopping is indeed a comfortable alternative, services provided by Amazon and Instacart among others are unable to fulfill orders at this time because of the high demand.

# Proposed Solution

## Motivation

Seeing the infeasibility of enforcing social distancing at all times, we have decided to analyze epidemic data[2] and patient data[3] to find other factors that may influence the spread of disease. Through the use of geographical information systems, simulations of the spread, and analysis of environmental conditions, we have been able to examine the intricate interplay of different factors and ultimately identify some of the most influential parameters that could effectively be optimized to reduce spread. The two factors that we have identified and decided to form as the basis for our solution are climate and the rate of reproduction[4] metric.

---

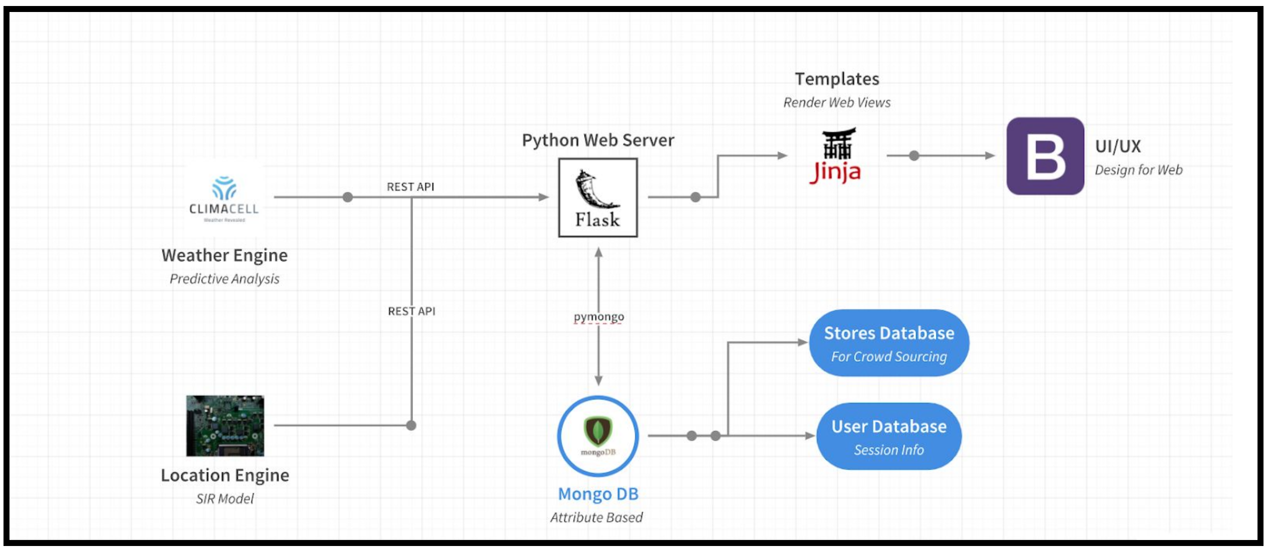
[1] https://www.worldometers.info/coronavirus/about/#sources
[2] https://github.com/CSSEGISandData/COVID-19
[3] https://github.com/beoutbreakprepared/nCoV2019/blob/master/latest_data/latestdata.csv
[4] https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf

## Web Application

The proposed solution to the stated problem is a web application that caters to the needs of users through two main features: scheduling tasks (powered by the weather engine) and planning shopping trips (powered by location and weather engine as well as crowdsourcing).

**Live Demo** @ optime.software
**Project UI @** https://www.beautiful.ai/player/-M6g6syKFuwmpT_qVHe3

### App Schema




### Weather Engine

After extrapolating from the epidemic case data, we were able to form a connection between the climate and the rate of spread. More specifically, we discovered that a high humidity and high temperature reduced the rate of reproduction significantly. After comparing our observations with the findings[5] of epidemiologists, we found that our results were similar. Our findings were also consistent with those of other researchers.

In order to obtain weather information for any given location, we utilized the Climacell API[6], which supplies accurate weather data on a minutely, hourly and daily basis. We parsed the humidity and temperature attributes to determine a "weather index" which is an adjustable weighted quantity. We then optimized this index to determine the safest time to leave the house and run errands. Users can interact with this feature using the "scheduler" or "shopping"

---

[5] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3537084
https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3556998
https://www.hindawi.com/journals/av/2011/734690/
https://www.medrxiv.org/content/10.1101/2020.03.26.20044727v1.full.pdf
[6] https://developer.climacell.co/

page in our web application. They will be prompted to input a time window to go out (ex: within the next week) as well as a duration (ex: thirty minutes) and our app will return the safest time for them to leave the house.

**Location Engine**

A common metric used to measure the infectivity of a virus in a pandemic is the average number of people an infected individual would likely infect before recovery or realization that they have the disease. Another way to think of this is a quantity to describe the rate of reproduction of the virus. This metric, henceforth referred to as $R_0$, is what we calculated in our application for counties near any user. In this context, a high $R_0$ implied a higher chance of getting infected and a low $R_0$ implied a lower likelihood of contracting the virus which are the areas that we recommended in our location analysis.

To calculate the $R_0$, we developed an SIR model to analyze data for every county in the United States from information that was provided to us by New York Times[7]. The data contained the number of deaths and positive cases of each county and is updated daily. The infectivity quotient that we calculated was relative to the population size of each county provided by US Census API[8].

In our model, we started by scraping the infected population on a daily basis for all counties in the United States. Having cleaned the data, we created a system of differential equations representing the change in subpopulations for the susceptible (people who can get infected), infected (people who are infected) and removed (deceased or recovered individuals) as shown below.

$$\frac{ds}{dt} = -b\, s(t)\, i(t),$$

$$\frac{di}{dt} = b\, s(t)\, i(t) - k\, i(t),$$

$$\frac{dr}{dt} = k\, i(t),$$

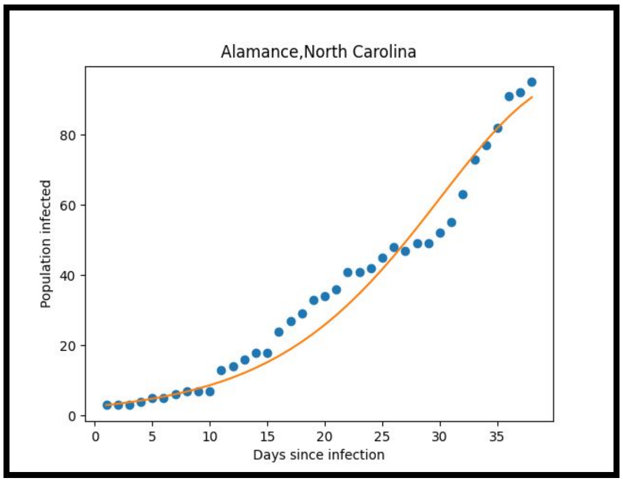
To solve this system of differential equations it is assumed that the initial parameters taken by any arbitrary county was that $s(t) = 1,\ r(t) = 0 @ t = 0$, meaning there is exactly one person infected in the beginning and no one has been removed from the population (passed away or recovered). The system was solved for every county, effectively calculating the parameters $b$ and $k$, which correspond to ɑ and **β** constants in a standard SIR.

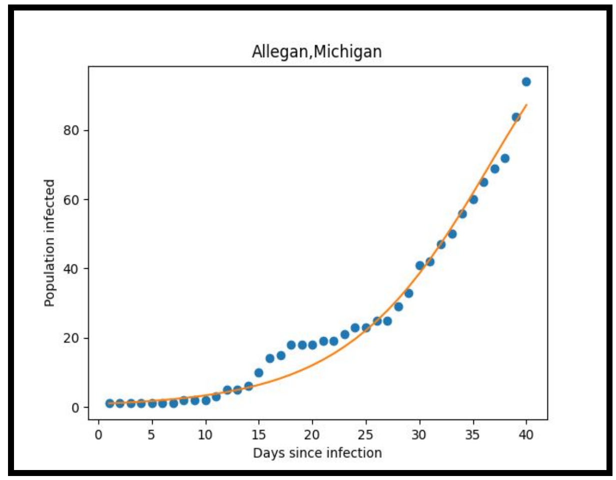
---

7 https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv

8 https://www.census.gov/data/developers/data-sets.html

From here, we designed a simple fitting algorithm to tune the aforementioned constants so that our model reflected the data. Our algorithm was able to apply a curve fit on these data points giving us the most optimal $\alpha$ and $\beta$ values, which represent how fast people moved from being susceptible to infected and from infected to removed respectively. These values gave us the ability to calculate $R_0$, and therefore allowed us to predict how infected populations in the future would behave using historical data. Two of the fits for counties in the United States can be seen on the following page.




**Crowdsourcing**

Due to the inaccurate information regarding store hours and stock during the crisis reported by companies such as Google and Yelp, data is crowdsourced by users. After a user completes a scheduled shopping trip, they are redirected to a form to input:

1. The stock of each product they wanted to buy at that store
2. The store's hours

Each field is pre-filled with the current database entry for the store so the user can easily update the information if it is incorrect. As opposed to a text field, users are prompted with dropdowns containing ranges for each product stock, which allows for greater flexibility in reporting.

When the user submits their updates, the database is automatically updated, and they are returned to the shopping page to continue adding trips for the future. In order to maximize the accuracy of our data, entries which have not been updated for more than 1 week are deleted.

# Impact

Through the testing we have conducted using SEIQR simulations, we have determined that we could certainly lower the current reproduction rate of the virus down substantially. Approximately 3.5 to 2.1 if our application was used by about a quarter of the population. This means that it would be easier for frontline medical personnel to offer treatment to those who are truly in need. Additionally, the amount of time anticipated for the country to return to a relatively normal state would be significantly reduced.

# Further Steps

While our current implementation works, we feel that we can integrate more features that could potentially improve the accuracy of our data and quality of our service.

### Predictive Analysis - A Deep Learning Approach

Seeing the potential of deep learning models in other applications pertaining to weather analysis, it seems reasonable to utilize an artificial neural network trained on climate data as the features and case data as the labels. There exists an abundance of data related to these attributes on the web publicized for developers to use. In our next iteration, we intend on training a predictive model, which could be more accurate than our current weather & location engine.

### Store APIs + Shopkeeper Update Form

One major flaw in crowdsourcing is validating the credibility of the updated information. We are taking appropriate measures to ensure quality of the user inputs, but there is certainly room for improvement. Our idea to mitigate this problem was to introduce a new page for shopkeepers to update their inventory directly, making it easier for users to access the latest & most accurate data.

Works Cited


Bhattacharjee, Soumyabrata. “Statistical Investigation of Relationship between Spread of Coronavirus Disease (COVID- 19) and Environmental Factors Based on Study of Four Mostly Affected Places of China and Five Mostly Affected Places of Italy.” *Arxiv.org*, 2020, arxiv.org/pdf/2003.11277.pdf.

Caspi, Gil, et al. “Climate Effect on COVID-19 Spread Rate: an Online Surveillance Tool .” MedRxiv, 30 Mar. 2020, www.medrxiv.org/content/10.1101/2020.03.26.20044727v1.full.pdf.

Chan, K. H., et al. “The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus.” Advances in Virology, Hindawi, 1 Oct. 2011, www.hindawi.com/journals/av/2011/734690/.

Fisher, Max. “R0, The Messy Metric That May Soon Shape Our Lives, Explained.” The New York Times, The New York Times, 23 Apr. 2020, www.nytimes.com/2020/04/23/world/europe/coronavirus-R0-explainer.html.

Holland, Steve. “Sunlight, Heat and Humidity Weaken Coronavirus, U.S. Official Says.” Reuters, Thomson Reuters, 24 Apr. 2020, www.reuters.com/article/us-health-coronavirus-trump/sunlight-heat-and-humidity-weaken-coronavirus-u-s-official-says-idUSKCN2253SA.

“The SIR Model for Spread of Disease - The Differential Equation Model.” The SIR Model for Spread of Disease - The Differential Equation Model | Mathematical Association of America, www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model.

Slosar, Anze. “Efficient Identification of Infected Sub-Population.” Arxiv.org, 2020, arxiv.org/pdf/2003.14337.pdf.

“Flask – Routing.” Tutorialspoint, www.tutorialspoint.com/flask/flask_routing.htm.

“Advice for Public.” World Health Organization, World Health Organization, www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public.