

1. Introduction

The 2020 COVID pandemic is an event of historical proportions. In addition to its life toll, with more than 2.7 million people infected globally and over 190'000 fatalities (as of 04/24/2020)¹, the pandemic has imposed an extreme financial toll resulting from the necessary social distancing and consumer behavior changes. The United States is one of the hardest hit countries², with an economic impact that exceeds trillions of dollars³, millions of job losses⁴, and expected bankruptcies⁵. The continental proportion of the US, its large population, and its heterogeneity, led to multiple shelter-in-place orders and the complete stagnation of some states, while others were less affected⁶.

Given that most of the economic activity had to be shut off, and the cost to keep it that away, the United States Federal government released a plan to reopen the country in three stages, going from lower risk activities to those that carry a larger risk to public health⁷. Despite these guidelines, the decision to and how to reopen was largely left to the states⁸. Not only it may occur in phases, but states have the power to also open specific counties and cities at different times. The question of how to restart the economy, therefore, is one of the most important decision each state will have to make.

A complicating factor in making reopening decisions is the lack of uniformity in the data. States and counties within states are not testing people at the same rate or following the same guidelines, and there is no guarantee that cases are being efficiently reported by the local health providers to the government, creating substantial heterogeneity in the number of cases and epidemiological data obtained⁹. This implicates that economical decisions made uniquely based in number of cases and other simple epidemiological data are likely to be biased, imperfect, and therefore, dangerous.

Given the above, finding auxiliary data that can help us understand the spread of COVID-19 would help feed models of reopening, allowing states to make more effective decisions in terms of what business and areas can be allowed to return to operation sooner. One possible data that could feed these models is venue data, here understood as parks, open areas, commerce, services, and buildings found in a determined location. Therefore, in this project, we will test the hypothesis that venue data can be used to predict the susceptibility of a county to COVID-19 spread regardless of the testing performed by that state.

¹ Source: https://en.wikipedia.org/wiki/Template:2019–20_coronavirus_pandemic_data

² Source: <https://www.sciencemag.org/news/2020/04/united-states-leads-coronavirus-cases-not-pandemic-response>

³ Source: <https://www.washingtonpost.com/business/2020/03/25/trump-senate-coronavirus-economic-stimulus-2-trillion/>

⁴ Source: <https://www.washingtonpost.com/business/2020/04/16/unemployment-claims-coronavirus/>

⁵ Source: <https://www.businessinsider.com/coronavirus-could-trigger-retail-bankruptcies-and-mass-store-closings-2020-4>

⁶ Source: <https://www.bbc.com/news/world-us-canada-52103066>

⁷ Source: <https://www.politico.com/news/2020/04/16/trump-reopening-plan-191959>

⁸ Source: <https://www.politico.com/news/2020/04/16/trump-plan-for-reopening-economy-191073>

⁹ Source: <https://www.washingtonpost.com/health/2020/04/21/kentucky-rhode-island-coronavirus-testing/>