DATA SCIENCE REPORT:
U.S. State Prioritization for Covid-19 Resource Allocation

CONTENTS

Report Compiled by:

Marceline Chiedza Manyika

GLOSSARY OF TERMS

COVID-19 Corona Virus Disease of 2019

WHO World Health Organization

JHU CCSE Johns Hopkins University Center for Systems Science and Engineering

NHC National Health Commission of the People's Republic of China

CCDC China Centre for Disease Control

ECDC European Centre for Disease Prevention and Control

MOH Ministry of Health Singapore

CDC Centre for Disease Control

U.S.A. United States of America

1 INTRODUCTION

A rapidly spreading global pandemic, COVID-19, has vexed the world from December 2019. According to the World Health Organization (WHO), "Coronavirus disease (COVID-19) is an infectious disease caused by a new virus. The disease causes respiratory illness (like the flu) with symptoms such as a cough, fever, and in more severe cases, difficulty in breathing. Coronavirus disease spreads primarily through contact with an infected person when they cough or sneeze. It also spreads when a person touches a surface or object that has the virus on it, then touches their eyes, nose, or mouth."

There is currently no vaccine to prevent coronavirus disease (COVID-19). To date more than 3,600,000 people have been affected globally and more than 250,000 have died globally. This study will focus on the period 22 January 2020-30 April 2020. *(data source: John Hopkins).*

Each country now has the mandate to allocate available resources to mitigate effects of the pandemic at different levels which include prevention and treatment measures with the primary objective of saving lives.

When a pandemic hits the world, there are seldom enough resources to minimize the negative effects in affected locations. There is critical need to prioritize resource allocation based on effects emerging in different communities battling the pandemic.

This study aims to produce a list of top 5 US states to be prioritized for each measure of response in dealing with the Covid-19 pandemic in the United States of America. This will enable the U.S. CDC and COVID-19 response teams to focus on and fully support states which are struggling to mitigate effects of the disease and successfully slow down its spread.

Study Objectives

This study was carried out to investigate which U.S. states must be prioritized in the allocation of COVID-19 resources. Specific focus was made to establish the following:

1. What is the extent of COVID-19 cases globally?

2. What is the extent of COVID-19 cases in the U.S?

3. What other factors are contributing to the spreading and death from COVID-19 in the US?

4. What is working elsewhere in mitigating effects of the COVID-19 pandemic?

5. Which U.S. states must be prioritized for COVID-19 resource allocation and why?

2 APPROACH, SCOPE AND METHODOLOGY

2.1 APPROACH

2.1.1 Quantitative data collection and assessment

Desk study. Secondary data collection: Data for this project was scraped from available COVID-19 datasets on the internet. Specifically, data on death, recovered and confirmed cases was sourced from the epidemiological data being compiled by the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) from various sources including the World Health Organization (WHO), DXY.cn. Pneumonia. 2020, BNO News, National Health Commission of the People's Republic of China (NHC), China CDC (CCDC), Hong Kong Department of Health, Macau Government, Taiwan CDC, US CDC, Government of Canada, Australia Government Department of Health, European Centre for Disease Prevention and Control (ECDC), Ministry of Health Singapore (MOH). (csv downloaded from https://data.humdata.org/event/covid-19)

U.S. Population data was obtained from the United States Census Bureau (https://www2.census.gov/programs-surveys/popest/datasets/). The population data is disaggregated by state.

To determine the typical social activities in the state with the highest loss of lives to COVID-19, data for most popular venues in the State with highest deaths, was called from the Foursquare, social location service.

Quantitative methods will be used to summarize data gathered from this study. The focus will be on analyzing secondary data in order to construct pertinent indicators. The assessment will require:-

i- Variable Identification from web scraped datasets (i.e. available variable types and categories) and Variable transformation.
ii- Univariate analysis which will involve gathering summary statistics from individual variables in the datasets i.e. mean, minimum, maximum, quartiles.
iii- Bivariate and Multi-variate analysis which will involve establishing correlation between two or more variables of interest.

2.1.2 Qualitative data collection.

Qualitative methods will be used to gather data on approaches used by countries which have had the highest proportion of recoveries in the first quarter of 2020.

2.2 PLANNING FOR SURVEY QUALITY

Effort will be made to acquire COVID-19 data from a source who has consolidated from many global sources to ensure coverage of as most affected countries as possible.

2.3 INDICATORS AND DATA COLLECTION PLAN

Task Description	Investigating questions	Possible Data Source	Tool/Method	Summary Technique
1) How many were affected globally by COVID-19 and where?	▪ How many COVID-19 cases were confirmed globally (Jan-Apr 2020), disaggregated by location?	- John Hopkins CSSE	-Desk study, Quantitative analysis	Historical maps, Time trend charts, Coverage maps, Bar graphs, Scatter plots bar charts, Line graphs, Flow diagrams
	▪ How many of the confirmed COVID-19 cases have recovered globally Jan-Apr 2020?	- John Hopkins CSSE	-Desk study, Quantitative analysis	
	▪ How many have died globally from COVID-19 Jan-Apr 2020?	- John Hopkins CSSE	-Desk study, Quantitative analysis	
2) How many were affected in the United States by COVID-19 and in which areas?	▪ How many confirmed COVID-19 cases are there in the U.S. for the period Jan-Apr 2020 ,disaggregated by State/Territory?	- John Hopkins CSSE	-Desk study, Quantitative analysis	
	˙ Which U.S. States have been most affected by COVID-19 Jan-Apr 2020?	- John Hopkins CSSE	-Desk study, Quantitative analysis	
	˙ How many have died in the U.S. from COVID-19 Jan-Apr 2020?	- John Hopkins CSSE	-Desk study, Quantitative analysis	
	▪ Which U.S. States have lost the greatest proportion of those affected by COVID-19 between Jan-Apr 2020?	- John Hopkins CSSE	-Desk study, Quantitative analysis	
3) What other factors are contributing to the spreading and death from COVID-19 in the US?	▪ Explore which factors may be contributing to the spread of COVID-19 (temperature, population density, state median age, average temperature)?	- Web scrape from various websites	-Desk study, Quantitative analysis	Regression analysis
	˙ What are the most common venues in the State with the highest COVID-19 deaths in the Jan-Apr 2020 period? This may indicate social behaviors which may be contributing to the rapid spread of the disease in this State.	- Foursquare	-Desk study, Quantitative analysis	Word cloud
4) How are other nations dealing with the crisis?	▪ Which countries have recorded the highest proportion of recoveries and lowest deaths?	- Internet search	-Desk study, Qualitative analysis	Case studies, Flow diagram
	˙ Which strategies are they using to minimize negative impact of COVID-19?			
5) US State prioritization	˙ Which 5 US States need the most urgent support in managing the health crisis resulting from the COVID-19 pandemic?	- John Hopkins CSSE	Quantitative analysis	Table

2.4 DATA VISUALIZATION AND RESEARCH TECHNIQUES

Data collected through secondary research will be presented using the following visualization techniques in a specified format :-

VISUAL TECHNIQUES		
Diagrams	**Venn**	
	To illustrate which individuals, organizations, projects or services are interacting with each other or overlap to deal with the COVID-19 crisis.	
Timelines	**Time trends charts**	
	This tool will be used to visualize changes over time (confirmed, recovered, deaths)	
Mapping	**Coverage maps**	
	Maps to visualize COVID-19 confirmed, recovery and death cases.	
Charts	**Bar graphs**	**Scatter plots**
	Bar graphs to depict case counts by Region	Scatter plots will be used to establish whether there is correlation between deaths vs average temperature, median age, population density
COMPLEMENTARY TECHNIQUES		
Secondary data review	**Case study**	
This will involve review of Information which is part of a publicly available set of data or other forms of official statistics not generated by the researcher.	This technique will be used to explore approaches implemented by nations whose COVID-19 negative effects are being successfully and significantly reduced since January 2020	