# Master's Research Project Proposal
# BUILDING A TIME SERIES FORECAST
# COVID-19 CASES, HOSPITALIZATIONS, AND DEATHS

## Problem Statement

In December of 2019 a new highly contagious disease Covid-19 was discovered in Wuhan, China. This disease is caused by the virus Severe Acute Respiratory Syndrome Coronavirus 2 or "SARS-CoV-2" and is spread through aerosol transmission of human respiratory droplets. By March of 2020 the World Health Organization declared the virus a pandemic.

Given its rapid transmission rate, high hospitalization rates, and deaths, governments were not prepared early for this outbreak. This lack of preparation saw shortages of Personal Protection Equipment ("PPE") for frontline health care workers and using shipping containers for the over flow of deceased bodies.

The goal of this project is to build a time series forecast model written in python that can accurately predict cases, hospitalizations, and deaths from this pandemic. A forecast model could be used by public health officials to better prepare for consequences from a pandemic.

## Question Formulation

Can a time series model created from data taken from different periods of 2020 forecast a future period in 2020? As an example, could time series data taken from March through May of 2020 predict cases, hospitalizations, and deaths in June 2020? Could that same model provide a reliable forecast for July 2020?

## Draft Hypothesis Formulation

**$H_0$** A time series Forecast model could not have reasonably and accurately predicted cases, hospitalizations, and deaths in the early months of the covid-19 pandemic.

**$H_A$** A time series Forecast model could have predicted cases, within a reasonable amount of error, hospitalizations, and deaths in the early months of the covid-19 pandemic.

## Data Sources

Johns Hopkins University
https://coronavirus.jhu.edu/

Center for Disease Control
https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/index.html

The Atlantic Covid-19 Data Tracking
https://covidtracking.com/data/download

## Team Members

Santosh Cheruku
Angel Claudio
John K. Hancock
John Suh
Subhalaxmi Rout

## Proposed Team Plan

The best strategy to manage this project is a divide and conquer strategy. There are 7 sections of the project that can be assigned to one or more members of the group. If someone wants to work on more than one section, they are allowed to do so.

| Section Number | Project Section | Description | Assigned to |
|---|---|---|---|
| 1. | Research Topic | A formulation of the topic that will be the subject of the research. | ALL |
| 2. | Question Formulation | The question that the project is trying to answer. | ALL |
| 3. | Data Collection and Cleaning | Review, collect, and clean data sources from sites such as the CDC, Johns Hopkins, etc. for covid-19 cases.<br><br>Once data is collected, it will have to be cleaned and prepared for analysis. Once completed, the clean data will be presented to the group. | Angel Claudio, Santosh Cheruku, Subhalaxmi Rout |
| 4. | Exploratory Data Analysis | An exploratory analysis and write-up of the data, e.g. missing values, outliers, incorrect entries (e.g. negative values), duplicate entries, etc.<br><br>If 2-3 Group members want to combine sections 3 and 4 to work on it as one unit, that would be acceptable. | Angel Claudio, Santosh Cheruku, Subhalaxmi Rout |
| 5. | Hypothesis Generation | Define the null, $H_0$, and the alternative hypotheses, $H_A$ | ALL |
| 6. | Literature Review | An analysis and write up of other online articles on time series, python models, and Kaggle projects on the topic.<br><br>How are these other works different from what we are doing? | John Hancock and John Suh |

| | | | |
|---|---|---|---|
| 7. | Model Building | Build python time series models using different machine learning algorithms for performance comparison purposes.<br><br>This is the most time intensive section of the project. | Subhalaxmi Rout, Santosh Cheruku, John Hancock, and John Suh |
| 8. | Results Analysis | A write up of findings from the project. Accept or reject the null hypothesis. | All |
| 9. | Journal Paper Presentation | A final presentation of the project. | All |