Draft 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 – a disease caused by the virus Severe Acute Respiratory Syndrome Coronavirus 2 or "SARS-CoV-2"—was discovered in Wuhan, China. Covid-19 is an aerosol transmission which means that it can spread through 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 and public health officials were not prepared early for this outbreak. For example, there were equipment shortages at hospitals. 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.

Question Formulation:

Given historical data from the early months of the covid-19 pandemic, is there a Time Series model that could have accurately forecast, within a reasonable margin of error, the number of cases, rate of hospitalizations, and deaths over the immediately succeeding months. In other words, given data from March through May of 2020 forecast cases, hospitalizations, and deaths for June through August 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 areas 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.

The seven sections are:

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. Once data is collected, it will have to be cleaned and prepared for analysis. Once complete, the cleaned data will be presented to the group.	2-3 Group members
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. If 2-3 Group members want to combine sections 3 and 4 to work on it as one unit, that would be acceptable.	1-2 Group members
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. How are these other works different from what we are doing?	1-2 Group members
7.	Model Building	Build Python Time Series models using Machine Learning and Deep Learning neural networks. This is the most time intensive section of the project.	2-4 Group members
8.	Results Analysis	A write up of findings from the project. Accept or reject the null hypothesis.	1-2 Group members
9.	Journal Paper Presentation	A final presentation of the project	1-2 Group members