John K. Hancock
CUNY MS Data 698 – Spring 2021
CAPSTONE: Senior Research Project – DRAFT Proposal
February 4, 2021

A TIME SERIES FORECAST ANALYSIS OF COVID-19 CASES
First Draft Proposal
and
Proposed Team Plan

Research Topic:

A Python Time series forecast that can predict future covid-19 cases, hospitalizations, and deaths.

Introduction:

The covid-19 pandemic has brought upon the world a global crisis unseen in over 100 years. The number of cases, hospitalizations, and deaths have risen across the globe. Forecasting the potential spread of the virus, hospitalization rates, and even deaths give public health officials and governments the ability to combat the spread, add additional resources to hospitals, and lower death rates.

Time Series forecasting is the study of using historical observations to predict future observations. This project will use a Python Time Series forecast to predict future covid-19 cases, hospitalizations, and deaths.

Question Formulation:

Can a Time Series model using Machine Learning and written using the Python language accurately forecast the spread of covid-19, hospitalizations, and deaths? If so, to what level of accuracy can such a model make a forecast?

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.	2-4 Group members

		This is the most time intensive section of the project.	
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

As we discussed on our call, this is my proposed plan for handling the project. I am open to alternate approaches, e.g. having everyone work on every section individually. However, with that approach, we will still need to coalesce all the disparate parts into one project which would take an inordinate amount of time.