Weekly Progress Report: Sep 25 - Sep 29

Contributors: Ryan Choe, Rakin Hussain, Sid Taneja

Progress/Accomplishments:

- Made Github Repository on schedule and added contributors.
- Found a dataset for project use and confirmed that it met the requirements.
  - Dataset Link (Kaggle)
- Analyzed metadata of dataset (number of total countries accounted for, column/variable names, any missing data, etc.)

Problems/Challenges:

- Had difficulty choosing a dataset that meant the requirements, but were eventually able to do so
- Encountered some Github administrative issues (i.e. adding contributors), but were able to eventually resolve
- Looking at the dataset, there is a lot of missing data for multiple countries for many variables. For example, Albania does not have any data reported from January 25 to March 8 in 2020. Down the road, we will need to consider how to handle the missing data when forming our predictive model (ie. Do we drop the data, fill it somehow, etc.).

Plan/Next Steps:

- Complete Project Proposal and Slides by October 4th
  - Outline project objectives and goals
- Begin work on Exploratory Data Analysis and Data Cleaning to stay on schedule
- Plan to divide data into training dataset (2020-2021 data) and testing dataset (2022 January - March data)
- Do individual initial research on time series modeling, just to gain some background on the topic

Overall, the team has shared collaboration access on their GitHub repo and has started to formulate ideas for the project proposal. We will aim to look at the dataset more in-depth and start to consider which variables to use for a potential predictive model, as well as experimenting with cleaning the dataset.