Project Proposal

**Research Topic**

We have decided to research COVID-19 data to create trends, models and visualizations that analyze the spreading global pandemic. COVID-19 has negatively affected almost everyone in the world in some way. Therefore, it would be interesting to observe any statistical differences that compare the US spread to other countries. Not only will we look at confirmed cases and deaths across multiple countries and states within the US, but we will also analyze confirmed cases as well. This will allow us to have a broader view of how the virus is being handled and the factors that go into this effort.

**SMART Questions**

* Are trends of COVID-19 cases different across countries?
* Does there seem to be a density of recoveries of COVID-19 in specific countries?
* Looking at confirmed cases, deaths and recoveries, is there any way to predict when the pandemic might slow down? (Flatten the curve)
* Are there certain pre-existing conditions that influence the contraction of COVID-19?

**Dataset Source**

<https://data.cdc.gov/NCHS/Conditions-contributing-to-deaths-involving-corona/hk9y-quqm>

This data repository was provided by the National Center of Health Statistics (NCHS) and is held on the Center of Disease Control and Prevention’s (CDC) website. The actual data is updated every week and contains COVID-19 data from February, 2020 to November, 2020. The actual dataset contains more than 12,000 rows and 10 columns of significant variables that can be used for testing.

**Github Link**

<https://github.com/HarishRam10/DATS6103TeamProject>

**Modeling Methods**

We plan to use predictive analysis algorithms such as Linear and Logistic Regression, Correlation, KNN and other statistical tests to find significant trends in our data. From this, we hope to come up with several conclusions that will detail the rise and fall of the virus and other factors that might influence its spread.