Project “Virus Recovery” will be led and completed by a single member, Angela Tseng. The research question examined is, “How does the location of an individual affect their chance of recovering from the coronavirus?” I find this question interesting due to its current impact on society; my family, friends, and myself have all been affected either directly or through social stigmatization. This project is intended to direct attention away from the panic surrounding the virus. Instead, it will focus on the spread and recoverability of the illness in order to help contain or provide the best resources for those affected. Ideally, through analysis of the data, insight will be provided about which countries are handling the virus the most effectively. From this, other countries can observe their precautions and methods and hopefully provide better resources for their people. For my overall plan for analysis, I intend to use data munging and data visualization methods on the dataset John Hopkins University has published to discover geographical trends.

Features used include working with multiple DataFrames, data munging, and information visualization to help predict trends and create machine learning methods as a response. Predictors include the Country/Region, and the sum of the deaths, confirmed cases, and recovered. The primary visualization method will be through density plots.

A close up of a piece of paper

Description automatically generated

All the data was in order and identified by a common index among the three files, but since I am only interested in the cumulative values for each location rather than individual values per date, I added up the values for each file first.

A screenshot of a cell phone

Description automatically generated

There are a significant amount of values missing in the Province/State column. Since there is no way to specify the values by taking the average or filling with the most common value, I dropped that column and latitudes and longitudes column as they are no longer relevant.

Upon examining the data further, duplicate entries in the Country/Region column become prevalent due to the deletion of the Province/State column.

A screenshot of a cell phone

Description automatically generated

To fix this, I add up the confirmed, deaths, and recovered columns for each country in separate columns. Afterwards, I delete the unnecessary columns and reset the index, leaving 162 rows.

A screenshot of a cell phone

Description automatically generated