

# STAT 390 Weekly Progress Report 1

September 25-29

Cindy Ha, Willie Xie, Erica Zhang

## Progress/Accomplishments

- Set up [Github Repo](#)
- Found [dataset](#) from Kaggle that met requirements
  - 67 variables
  - 341,408 observations
- Added dataset to raw data folder
- Completed quick skim of dataset to check usability
  - Found missingness issues that will need to be addressed

## Plan/Next Steps

- Work on proposal to present on **Thursday, Oct 5th**
  - Develop a prediction question we want to answer regarding Covid-19
    - \* Determine target variable
    - \* Decide between the appropriate regression or classification models
    - \* Complete any necessary data pre-processing steps (e.g. mismatched data types, missingness, etc.)
- Set up timeline to complete project by deadline
- Obtain any necessary sources/references to help answer our question
- After getting approval to move forward, conduct a thorough EDA
  - Split data into training and testing sets

We are optimistic about the current progress of the project and plan to meet regularly.