*678 Mid Term Project*

*Handing Zhang*

***Personal Statement:***

My career goal is to work as a data analyst in a bio related technology company. To align my ability with that position I need to exercise and showcase my ability in wrangling and cleaning large sized datasets related to space such as the covid vaccination process, test process, evolution and the worldwide trend of the pandemic. I found several datasets from Kaggle and Our World In Data, including the vaccination process and the world wide cases and deaths of covid 19. I intend to combine the datasets by area in seeking more insights from them. First I would perform data wrangling and cleaning, where I intend to subset the data into certain areas, possibly ten countries, and take into consideration different manufacturer’s vaccines to do comparisons in a multilevel analysis. I would also try to explore the relationship between vaccination, testing and cases of covid 19.

***Question:***

Do number of vaccinations, testing and geographic factors influence the number of cases, evolution and deaths of covid 19?

***Data Sources:***

Datasets: country\_vaccinations\_by\_manufacturer.csv country\_vaccinations.csv

https://www.kaggle.com/gpreda/covid-world-vaccination-progress?select=country\_vaccinations\_by\_manufacturer.csv

Dataset: covid-variants.csv

https://www.kaggle.com/gpreda/covid19-variants

Dataset: covid-testing.csv

https://www.kaggle.com/gpreda/covid19-world-testing-progress

Dataset: covid\_cases.csv

<https://ourworldindata.org/covid-deaths>

***Proposed Timeline:***

EDA: Nov 8th

Data Processing: Nov 12th

Modeling and Validation: Nov 19th

Write up: Nov 27th