Proposal Title

# **Name Semester GitHub Portfolio URL**

# Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you’re doing with these data.

It is not a secret that the COVID-19 pandemic has taken a huge toll on human lives throughout the world. This pandemic strikes at an unprecedented rate, and it does not discriminate against age, race, gender, and ethnicity. Therefore, COVID-19 kills folks off all ages whether they are white, Asian, black, Hispanic, and other. When it comes to study human deaths associated with the Coronavirus, it appears that the virus attacks certain ethnic group at a higher proportion. To understand the logic behind COVID-19 age disparity, I collected a dataset from data.gov that recorded COVID-19 confirmed death cases by race and ethnicity distribution Maryland residents. The data will be analyzed to investigate what role race plays COVID-19 deaths. Also, I investigate which ethnic group is more vulnerable to covid-19 deaths, and which ethnic group that is not. Further, I will perform correlation tests among the different ethnic groups.

# Which Data?

What is the dataset you’ll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

This part of any project has never been easy. The dataset I will be examining is called MD COVID-19 - Confirmed Deaths by Race and Ethnicity Distribution, and it was retrieved from data.gov. It is a collection of the statewide confirmed and probable COVID-19 related deaths that have been reported each day by the Vital Statistics Administration by categories of race and ethnicity. The collection classifies a confirmed death when a person had a laboratory-confirmed positive COVID-19 test result. It lists confirmed COVID-19 deaths among Maryland residents by race. Below is a link to the data source.

https://catalog.data.gov/dataset/md-covid-19-confirmed-deaths-by-race-and-ethnicity-distribution

# Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you’ll be proposing your research questions as well as justifications for why you’d offer these data in this way.

COVID-19 has been known to have a higher mortality rate in certain ethnic group. I will analyze the data because I think it is critical to understand COVID-19 deaths in race disparity. A thorough investigation of COVID-19 deaths rates in different race can be very beneficial when it comes to make informed decisions. In this project, I will make attempts to understand COVID-19 deaths disparity in race. I will define a few questions and set clear measurement priorities. Also, I will analyze the data by plotting it out and finding correlations. The plots will be of types bars, lines, histograms, and so forth to display the results of my analysis. Also, I will calculate the mean, maximum, minimum, and standard deviation of the data. After analyzing the data, I will interpret the results with these key questions in mind:

• Does the data answer my original question? How?

• Does the data help me defend against any objections? How?

• Are there any limitations on my conclusions, any angles I have not considered?

# What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

To study if there is race disparity in COVID-19 deaths, I will use a dataset called MD COVID-19 - Confirmed Deaths by Race and Ethnicity Distribution. The data recorded positive COVID-19 deaths among Maryland residents by ethnicity. I will conduct exploratory data analysis, investigate whether race disparity exists in deaths that are associated with COVID-19 infections or not, perform correlation tests between different ethnic groups. The presence of correlation will imply disparity in COVID-19 related deaths. Also, I will include all deaths, and I will display them by different ethnic groups.

# Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

I am sure that this project will be challenging like any other project that I worked on. My worries for this project are to be able to predict the outcomes that I intend to through predictive analysis. I have not yet done any analysis of the data, and I hope that the data set is a good fit to solve the business problems. It can sometime be a challenge to come up with expected results base on the data set selected. I am moving forward with, and I hope that it does not come with risks that will challenge the algorithms I will be working with.

# Concluding Remarks

Tie it all together. Think of this section as your final report’s abstract.

The mysterious appearance of COVID-19 in early winter of 2019 that accompanied millions of infections around the world, which resulted in millions of innocent lives being lost in a short interval of time has raised concerns that left authorities stranded as to how to contain the spread of this deadly disease. Early in the pandemic, lockdown measures were adapted in lieu to understand the mechanism of the deadly virus, so appropriate measures could be taken to minimize the death toll. COVID-19 does not discriminate against ethnic. The virus kills folks of all ages, gender, and ethnicity. As universal as it can be, it seems to cause more deaths in certain ethnic group compared to others. Thus, does COVID-19 discriminate against race as far as deaths is concerned? Or are COVID-19 deaths essentially random events that occur at about the same rate in all ethnic group over the long run.

These questions can be answered by studying COVID-19 confirmed death cases by race and ethnicity distribution among Maryland residents, as based on the data.gov database.