**Email:** geetha.vangala@gmail.com

**Name:** Geetha Vangala

**Program:** Code Louisville

**Currently Enrolled Course and Day:** Data Analysis Course 1 (Tuesday)

**Project Objective: (What you Expect the Project to do)**

This Covid-19 data and analysis project aims to provide a statistical and visual representation of potential trends on the death, active, and recovered cases across countries and WAHO regions over time.

**Statement of Work: (what work needs to be done)**

Upon choosing the optimal features to complete this project, an appropriate dataset must be chosen and cleaned. Following that, cleaning and some statistical calculations will be done. Finally, I will visualize the statistics and other relevant graphical information and provide a detailed analysis of the trends observed and the potential causes.

**Feature 1:** Reading in data from a CSV file based on covid-19 data. Importing pandas and using “pd.read\_csv” to correctly complete it.

**Feature 2:** Using pandas, I will perform cleaning on the data. Will identify the columns in the data followed by acquiring a sample of 10. Using the shape, I will also count the number of null values in the dataset. Based on the data acquired, I will make decisions on dropping columns or removing the null values. Finally, I will also look into grouping columns together to acquire some of the data that I want such as “confirmed death according to date.”

**Feature 3:** I will use sort\_values for the active, death, and recovered cases columns to compute the top 10 countries that lead in each category and list them in a table using “head(10).” I will repeat the preceding process for the WHO regions. I will also perform a linear regression on the population and total cases that will allow me to identify any positive or negative correlation between the two variables through a resulting slope and intercept of best fit line.

**Feature 4:** Using seaborn or matplotlib, I will visualize plots as follows, but not limited to: Top 10 countries with most death cases, Top 10 countries with most active cases, active, recovered and death cases based on WHO region, and active, recovered and death stats over time.

**Feature 5:** I will interpret data and plots using markdown cells. I will also discuss why I selected those specific plots.

**Please include a link to a Google/Microsoft Doc/Slides. GitHub Link**

**Please copy the link to your document and paste it in the field below**