# Coursera – Data Science Capstone Report

# Marge Ogle

March 2021

## **Introduction**

A global pandemic is wreaking our communities! Countries all over the world have instigated mandatory lock-downs on the population or ‘non-essential’ businesses in order to contain or stop the spread of this virus. What makes this specific virus different is that it attacks the respiratory system, causing shortages of vital equipment needed to aid patients in acute respiratory distress.

The current global pandemic referred to as COVID-19 has been compared to the Spanish Flu. The Spanish Flu occurred in 1918 infecting one-third of the world’s population and killing up to 50 million people worldwide (as reported by World Atlas).

Using skills acquired in this Data Science course, let’s examine data obtained from the World Health Organization and the Center for Disease Control to analyze the information being obtained for the COVID-19 pandemic.

## **Data**

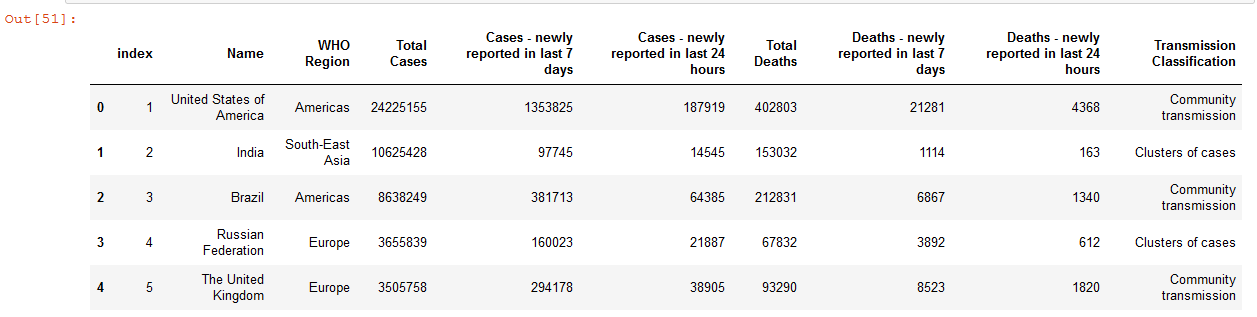
Datasets from the [World Health Organization](https://covid19.who.int/table) and the Center for Disease Control (CDC) will be used in this report. The specific dataset titled [Deaths involving coronavirus disease 2019 (COVID-19),](https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku) will be used because its data can be readily compared to the information obtained from WHO. The CDC dataset also breaks down fatalities resulting from COVID-19 within age groupings and what other maladies were possibly involved.

## **Methodology**

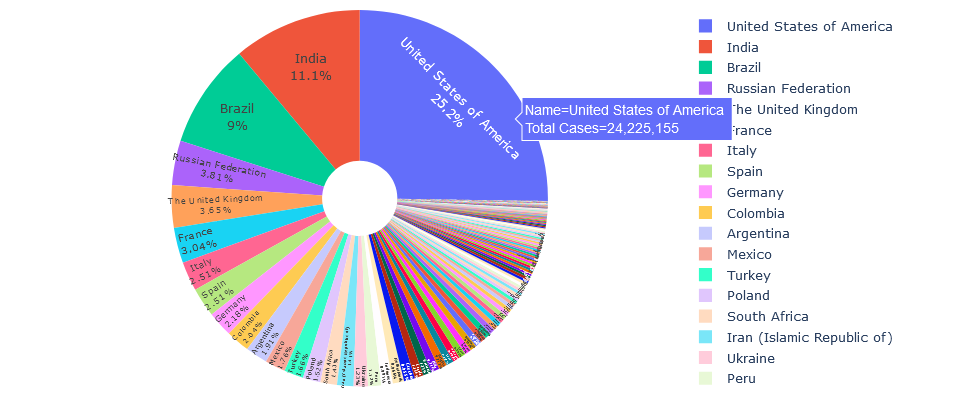
Descriptive methods will be used to display/visualize the information provided on the datasets that has been obtained on the World Health Organization (WHO) and Center for Disease Control (CDC) websites.

## **Results/Discussion of Observations and Recommendations**

According to data from the [World Health Organization (WHO),](https://covid19.who.int/table) the United States has/had over 24 million total cases resulting a little over 400,000 deaths from this disease. That is 1.66% of the total cases in the United States resulting in fatality. Conversely, this would indicate that 98% of the infected population has recovered or is in the process of recovery. Globally, the fatality percentage is 2.16% and a recovery rate of 97.84%.



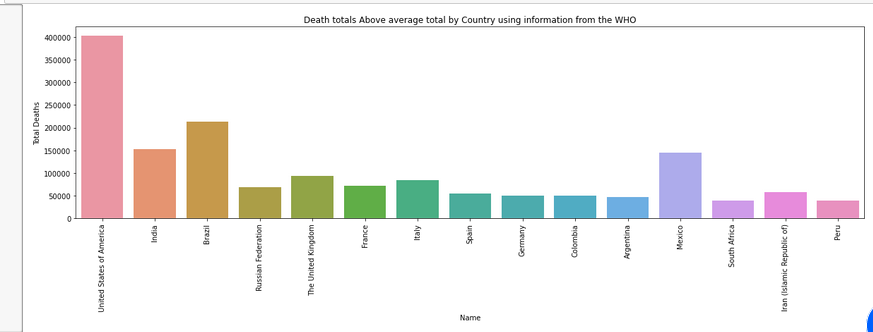




According to this information, the United States leads the rest of the world with the largest number of fatalities (19.4%).

**Total Deaths per Country reported by World Health Organization**

Further analysis of the WHO data, reveals that the ‘average’ number of deaths were 34,575 reported fatalities. Here is a chart representing those countries that reported ‘above average’ fatalities:



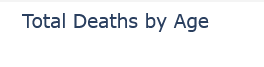
What is interesting in this report is that India is listed second in regards to the total number of cases reported but comes in 3rd when reporting the number of deaths. India, according to [World Population](https://worldpopulationreview.com/countries) census has a population over one billion with the United States reported at 325 million which would indicate that population size does NOT factor in regarding the impact that COVID has on a country/population.

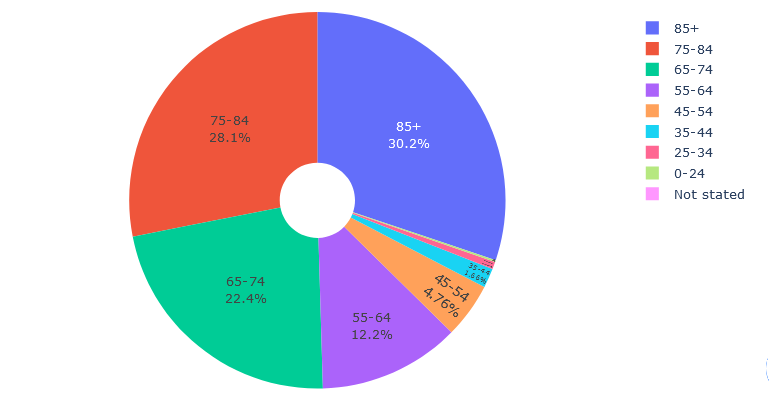


Using information obtained from the Center for Disease Control (CDC), let’s take a deeper look into COVID-19 fatalities/deaths in the United States.

One of the datasets available on the CDC website contains information not only on [deaths reported by state](https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku) but specific age groups and if there were other illnesses that were a contributing factor. Although, the dataset acquired is a subset of the entire file (only 1,000 records); it does contain overall totals which is what we will be using for our review and analysis.

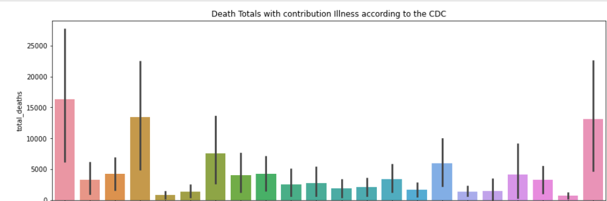
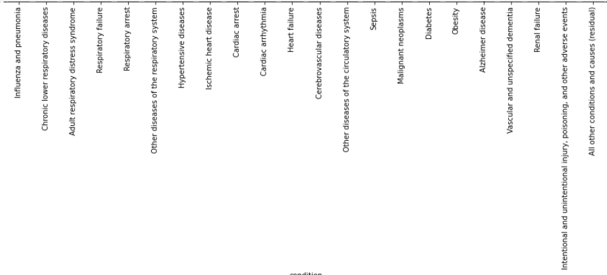
A pie chart was constructed using the total number of deaths per age group.





This clearly shows that ages 65 and older had the greatest number of fatalities nation-wide. That’s more then Three-Fourths of the population being severely impacted.

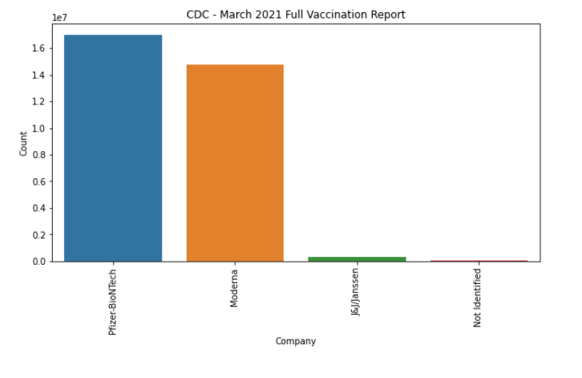
From the same dataset, we can see that pneumonia and influenza were listed as a contributing factor that led to a fatality.

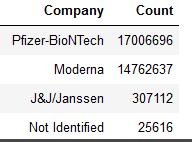


Further analysis, showed that the age group most impacted by COVID-19 AND pneumonia/influenza was the age group of 75-84.

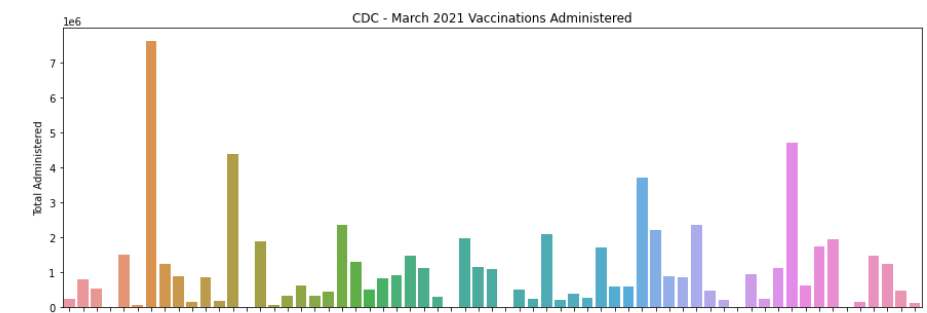
## Vaccinations in the United States

Beginning on December 14, 2020; the United States began mass vaccinations with first responders, caregivers, and citizens in the age group of 65 and older to be the first recipients.



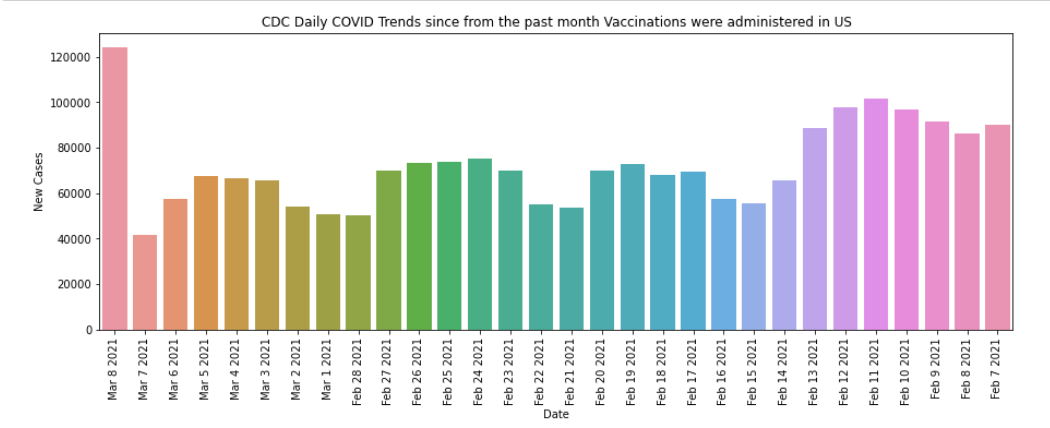
As of March 2021, two of the [vaccination companies](https://covid.cdc.gov/covid-data-tracker/#vaccinations) have over 14 million vaccinations. In this particular case, Pfizer-BioNTech delivered over 17 million vaccinations throughout the US and its territories as well as Moderna delivering over 14 million. A new company from Johnson & Johnson is just beginning its delivery schedule.

A display of [States & Territories](https://covid.cdc.gov/covid-data-tracker/#vaccinations) within the United States is shown below. This chart shows the total doses that were administered as of March 2021. Both data sets were obtained from the CDC’s COVID Data Tracker page regarding vaccinations.

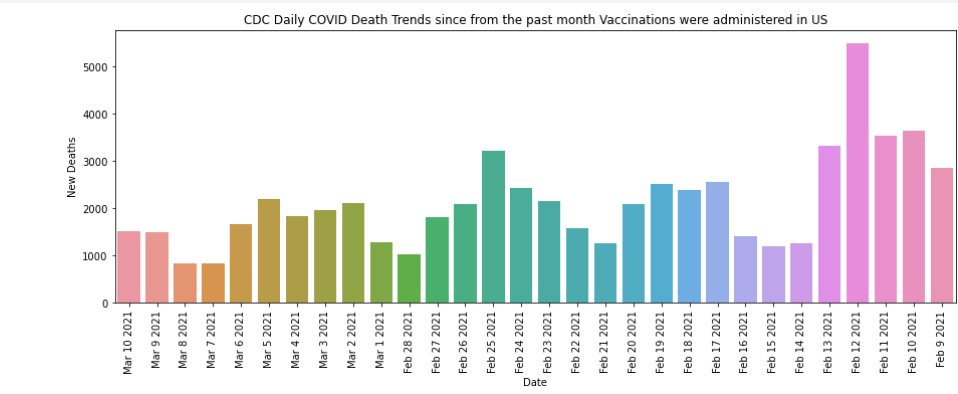




Now that the vaccinations have been delivered in such large numbers, lets see once again the effectiveness.

Going back to the [CDC data site](https://covid.cdc.gov/covid-data-tracker/#trends_dailytrendsdeaths), here is the last months’ worth of cases that have been reported for Deaths in the United States:

There was a spike in reported cases throughout the United States on March 8th. That is alarming but we need to compare that with the [daily tracking of fatalities](https://covid.cdc.gov/covid-data-tracker/#trends_dailytrendscases) within the same time period.



The reported deaths (according to the CDC) have greatly decreased! On the same date of March 8th, the count was a low 844. Reported death counts are still devastating for families, caregivers, hospital staff; but seeing the numbers in decline is encouraging especially since the individual states are ‘re-opening’ their economies.

## **Conclusion**

Through the use of Data Science Visualization tools, we are able to take information supplied by various health organizations and determine for ourselves the progress that is being made against this historic Global Pandemic known as COVID-19. “The vaccinations do NOT prevent you from contracting COVID-19 but it does help keep you from getting seriously ill”. This quote was taken directly from the CDC [website](https://www.cdc.gov/coronavirus/2019-ncov/vaccines/keythingstoknow.html) and our analysis of the data provided does show that to be accurate.