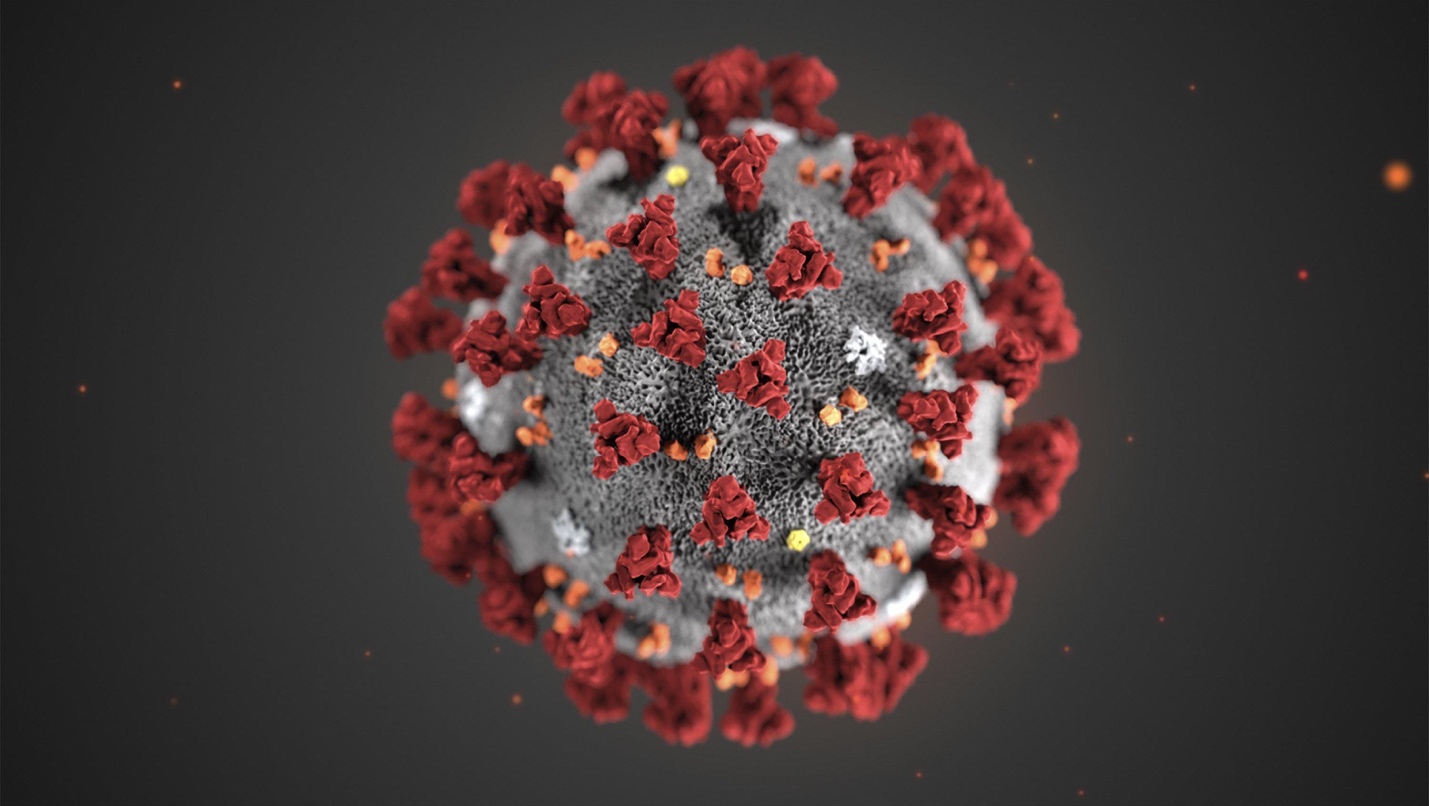
# Factors Impacting COVID

Hedyeh Erfani

DSC 680

Abstract

With COVID being a pandemic that has resulted in the loss of millions of lives, it is important to know whether or not certain factors exacerbate an individual’s risk of being affected by this illness. This kind of information could be used to spread community awareness, find risk mitigation strategies, and help with finding an actionable plan moving forward.



Data

The dataset is from the beginning of the pandemic. I wanted to look at COVID from the earliest perspective possible to then be able to compare it with data from after COVID is no longer considered a pandemic. The data includes the following columns:

State

Race

# of Cases

% of Population

% of Cases

% of Deaths

Poverty %

Cases/Pop

Cases-Over33%

Death/Pop

Deaths-Over33%

The main column from the data that will be focused on is race and whether that has a correlation with the other features of the dataset.

Walkthrough

To begin I first cleaned the data. I removed NaN values and dropped columns that were not pertinent to the analysis. From there I also looked at data.describe() to get a better understanding of the columns and their attributes, such as their mean and count. I created separate dataframes for each race and noted data summaries for each race. This included things such as case to population ratio per race. I also looked at correlation using the Pearson method to see if there was any correlation between the features. From there I created visualizations such as bar plots to gain a better understanding of how each feature played a role.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

To perform my analysis, I attempted to run a train/test/split. Unfortunately, I was unable to perform the train test split accurately. This is where the walkthrough will end since I cannot comment effectively on the accuracy of the metrics that followed, as they were based off of the erroneous train/test/split.

Conclusion

While I was not able to reach a conclusion, I am able to state that some of the visualizations created gave insight regarding COVID and how certain factors may play a role in cases. I look to continue attempting my analysis so I can come up with a conclusion I feel comfortable sharing. Once that happens, I will be awaiting data in the near future from the end of COVID so I could run a comparison between the results from this dataset and overall results from the pandemic.