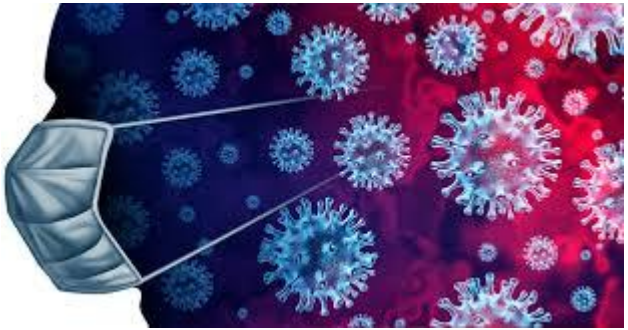
Final Presentation
IND ENG 142

COVID-19 Death and Vaccination Analysis

Dylan Hungate
Ronel Solomon
Tiffany Natasha
Eeshan Sharma

Problem

COVID-19 has caused global catastrophe in the form of economic collapse, social unrest, mental health crises, and many other unchronicled problems. Many people got laid off from their jobs, lost their family members, and became despondent due to the continuous quarantine, lack of job opportunities, and resulting crime and poverty increases.



Vaccination and effective public health policy is one the solutions of curbing the spread of the virus. However, there are variations in vaccination preference and available supply that places certain constraints on the objective function of maximizing vaccination rates!

Goal and Motivation

GOAL

- To stop the spread of COVID-19 virus
- To give awareness to the some certain groups of people who have low number of vaccinations
- To give more attention to a certain county who has high number of death cases

MOTIVATION

- Decrease the number of covid cases in United States
- Increase the number of vaccinations in United States

Data / Source

1. **Death Cases of COVID-19**
   Source:https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4
   Number of observations: 92431
   Number of variables: 17
   Variables: case_month, state_fips_code, county_fips_code, age_group, gender, race, hispanic/latino, case_positive_specimen_interval, case_onset_interval, process, exposure_yn, current_status, symptom_status, hosp_yn, icu_yn, death_yn, underlying_conditions_yn

2. **Vaccination by State**
   Source:https://github.com/CHIDS-UMD/Covid19-Vaccination-Race-Disparity
   Number of observations: 1141
   Number of variables: 33
   Variables: State, County, GEOID, WholeNum, WhiteNum, BlackNum, CvdVax_WhiteRate, CvdVax_NHWhiteRate, CvdVax_MBlackRate, HighSchool_WholeRate, Bachelor_WholeRate, HighSchool_WhiteRate, Bachelor_WhiteRate, HighSchool_NHWhiteRate, Bachelor_NHWhiteRate, HighSchool_BlackRate, Bachelor_BlackRate, Unempl_WholeRate, Unempl_WhiteRate, Unempl_NHWhiteRate, Unempl_BlackRate, MeanInc_WholeAvg, MeanInc_WhiteAvg, MeanInc_NHWhiteAvg, MeanInc_BlackAvg, poorhealth_WholeRate, AvgTrafficVolum, inaccessibility, urban, republican_rate, hesitancy

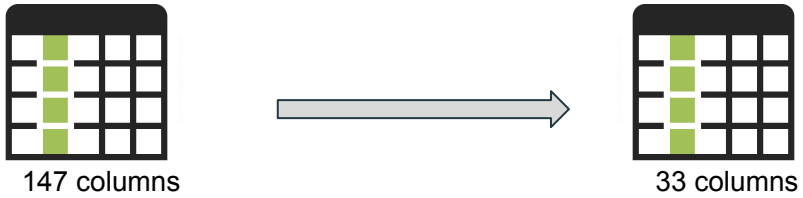
Data Cleaning Process

Covid Cases:

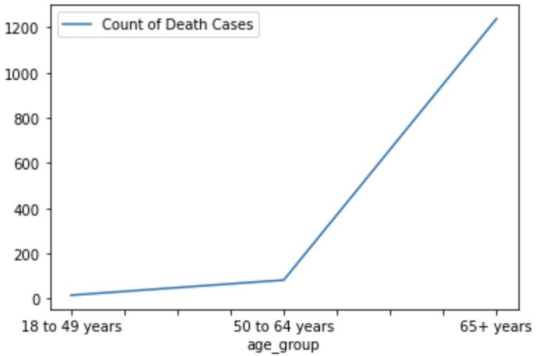
1. Changed missing values, such as “Missing”, “Unknown”, and “Other” to null values.
2. Drop null values
3. Changed the value into binary values (0 and 1) if it is possible
4. Change the type of the values

Vaccination rate:

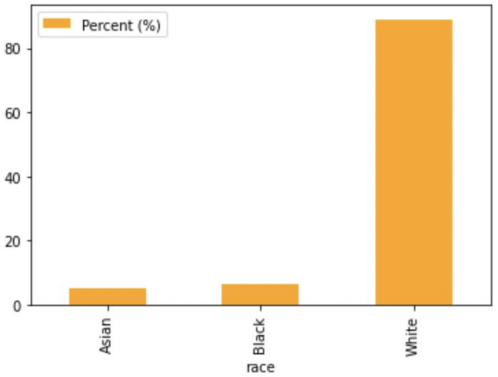
1. Remove unnecessary columns



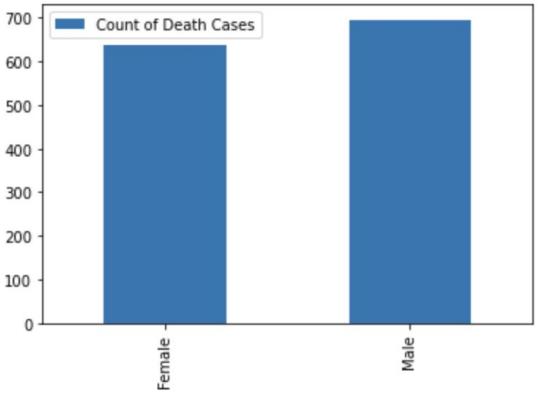
Data Visualization for Death Cases



The most number of deaths caused by COVID-19 comes from 65+ years old group.

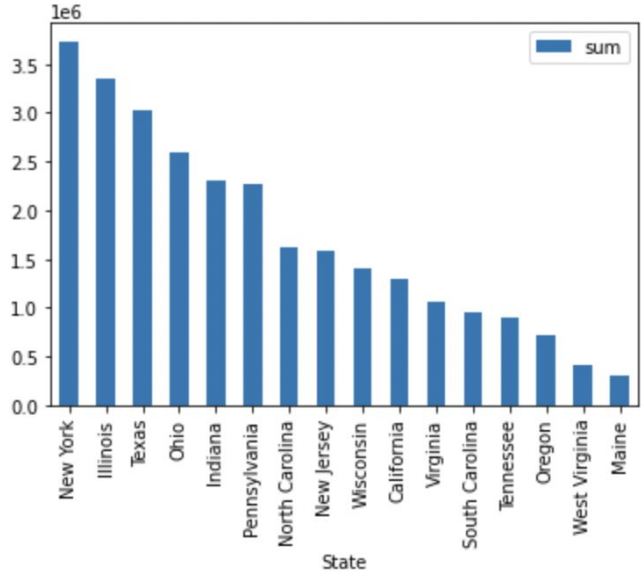


The % of COVID-19 deaths are dominated by white people

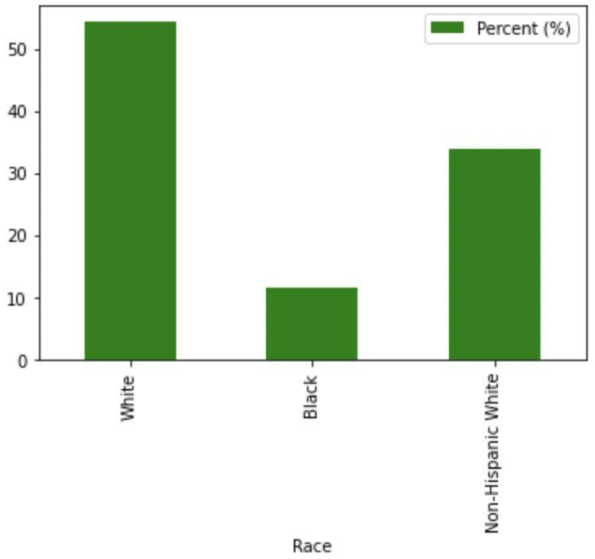


Males are more likely to die because of COVID-19, albeit by a small amount

Data Visualization for Vaccination Rate



The cumulative sum of vaccinations based on each state



Around 55% of white people have been vaccinated

Analysis

- White is the race with the highest number of vaccination, but they also have the highest number of deaths.
- Most of the positive cases come from 18-49 years age group with around 5000 cases, but most of the death cases come from 65+ age group with around 1200 cases.
- The number of people who died from COVID-19 comes from those who are hospitalized, but was not go through the ICU.
- During March 2020 until August 2021, there is a significant increase between September 2020 and October 2020 with 5000 death cases differences.
- Female has a higher number of the positive cases, but male has a higher number of the death cases.

Methods

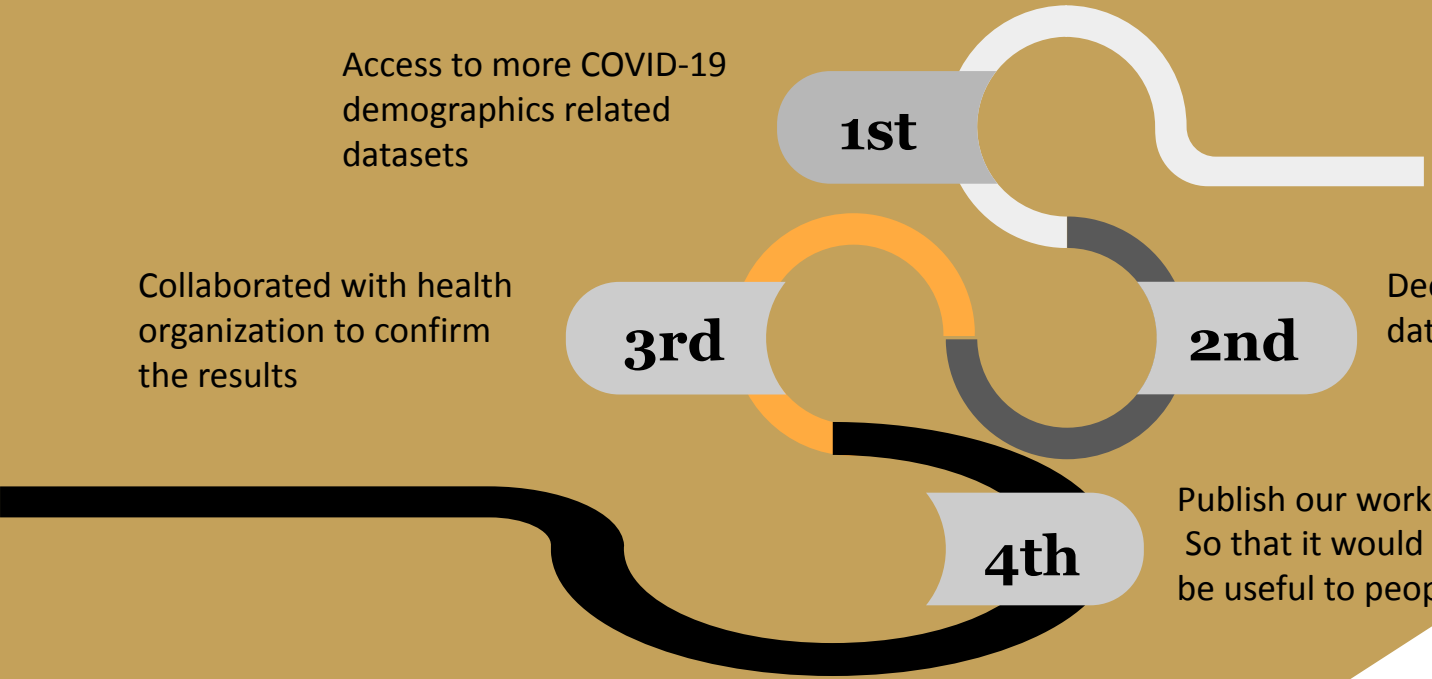
Death Cases

1. Baseline model (We are using death_yn = 0 as the most frequent class and the accuracy is 98.64%)
2. Logistic regression (after removing some insignificant variables, the accuracy is 98.63%
3. CART (using the best ccp_alpha = 0, we obtain 98.63% as the result)
4. Random Forest (we got the accuracy of 98.68%)
5. LDA (97.48% accuracy)

Vaccination rate

1. Linear regression (we separate the linear regression based on the race)
   a. Vaccination rate for White
      R^2: 0.932, OSR^2: 0.928
   b. Vaccination rate for Black
      R^2: 0.901, OSR^2: 0.896
   c. Vaccination rate for Non-Hispanic White
      R^2: 0.391, OSR^2: 0.357
2. CART (Accuracy: White 74.05%, Black 95.91%, Non-Hispanic White 94.75%)
3. LDA (Accuracy: White 48.98%, Black 8.75%, Non-Hispanic White 55.68%)

Next Steps


1st
Access to more COVID-19 demographics related datasets
2nd
Deepen the analysis of the datasets
3rd
Collaborated with health organization to confirm the results
4th
Publish our work
So that it would be useful to people

Impact

- Confirm that vaccination helps to decrease covid cases
- Recommend any future action that need to be taken in order to stop the spread of COVID-19 virus

THANK YOU