

# Covid-19 Infection Analysis

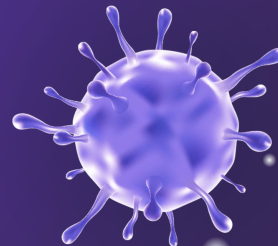


*Tonirose Babasoro, Elsa Bustos, Utsav Chaudhary, India Scott*

Program: Data Analytics

Group Name: Endless Knot

Presentation Date: April 7, 2022

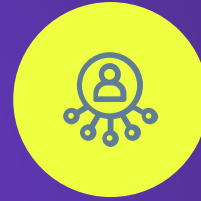


# INTRODUCTION



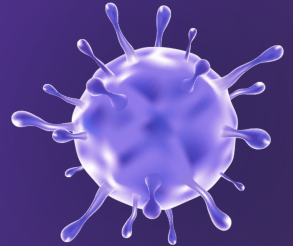
## TOPIC

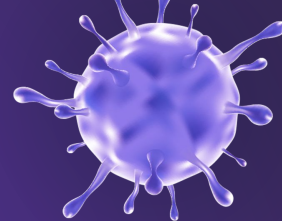
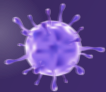
What is the likelihood of being infected by Covid-19? How is infection affected by factors such as vaccination rates, gender, and ethnicity?



## PURPOSE

It is important to analyze future trends following the Covid-19 pandemic to understand the prevalence of infection within the American population.





## DATA SOURCES

We gathered data from reliable organizations such as New York Times and the Center for Disease Control (CDC) which provide csv files on their findings. Our data identifies:

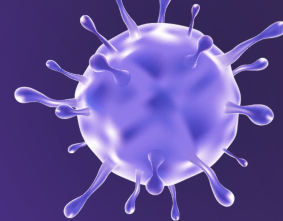
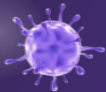
- vaccination rates
- gender ratios
- ethnicity statistics



## QUESTIONS ANSWERED

- Are certain populations more likely to be infected than others?
- How do these factors affect the other?
- What other factors should be considered in identifying risks of infection?





## DATA EXPLORATION

- Post-manual review of multiple csvs, we were able to identify a primary key: "States"
- Data is filtered and aggregated via SQL server and filter null values with pandas (Python)



## DATA ANALYSIS

- Machine learning
  - Supervised learning
    - Neural Networks
    - Linear regression
    - Support Vector Machine (SVM)

```
1
2 import matplotlib.pyplot as plt
3 import numpy as np
4 import pandas as pd
5 import statsmodels.api as sm
```


```
1 model = sm.OLS(y, X)
2 results = model.fit()
3 print(results.summary())
```

### OLS Regression Results

```
=====
Dep. Variable:                Women    R-squared:                1.000
Model:                        OLS      Adj. R-squared:           nan
Method:                       Least Squares    F-statistic:             nan
Date:                         Sun, 27 Mar 2022    Prob (F-statistic):       nan
Time:                         21:29:57    Log-Likelihood:          498.58
No. Observations:              51    AIC:                     -895.2
Df Residuals:                  0    BIC:                     -796.6
Df Model:                      50
Covariance Type:               nonrobust
=====
```



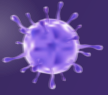
## TOOLS USED TO CREATE

- Data Visualization
    - Tableau
      - Correlation graphs
  - Data Presentation
    - Google Slides
      - Summarization of project
- 



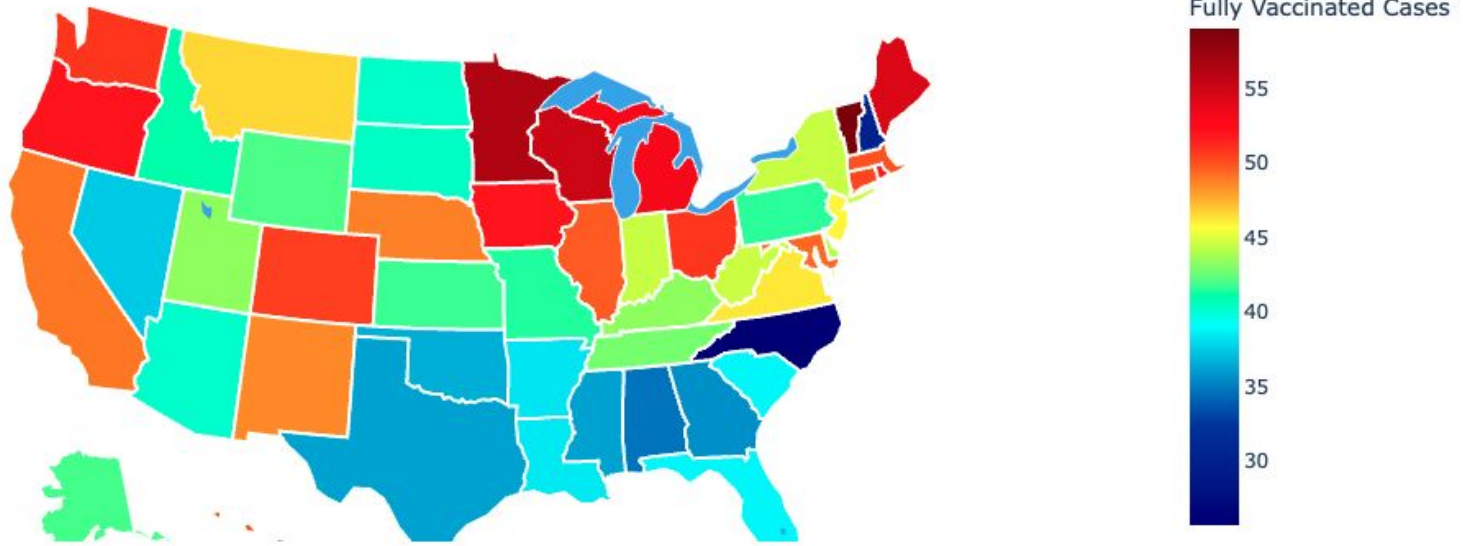
## INTERACTIVE ELEMENTS

- Plotly
  - Map with options to toggle different factors of each state such as:
    - Vaccination rates
    - Infections
    - Gender data



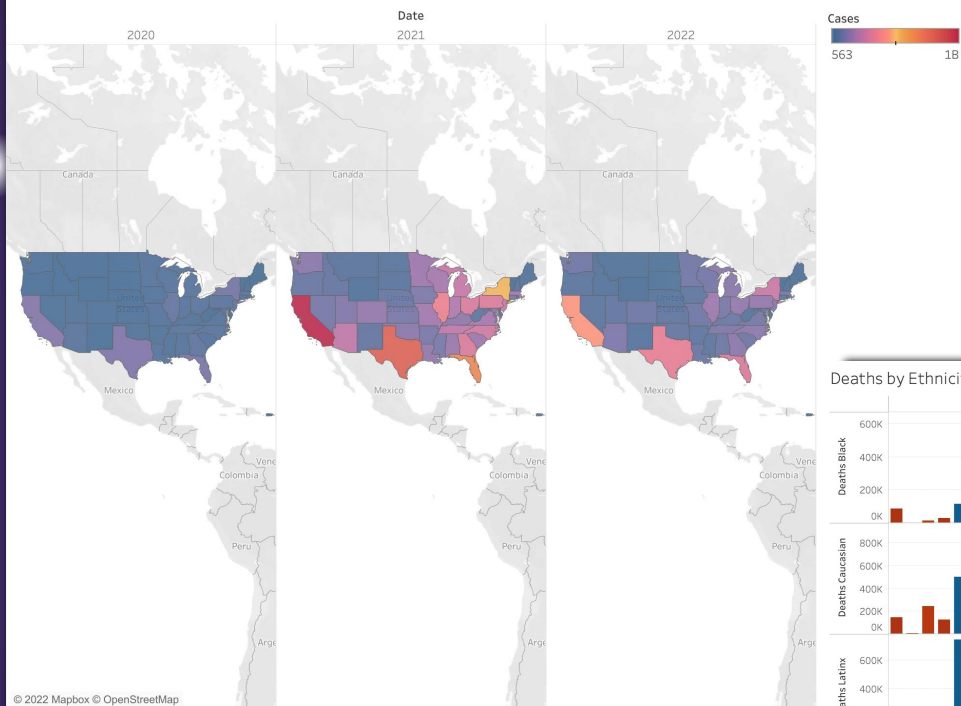
# Interactive Element

US Covid Cases & Vaccination Rates





Cases by Year Map



# Visualization

Deaths by Ethnicity

