

Mitigating Effects of Future Pandemics on Disproportionately Affected Populations

By Michael Hartnett, Yunus Herman, Josh Slizinov

Agenda

Problem Statement

Solution Proposition

Gathering Data & EDA

Modeling

Model Performance

Improvements

Recommendation

Questions

Problem

We've all been affected by COVID this year. However, some groups have been affected more than others. We set out to discover what factors in the socioeconomic and demographic realm may have led this pandemic to disproportionately affect certain groups.

Problem Statement:

Can we use socioeconomic, demographic, and employment data to identify groups or clusters most adversely affected by the pandemic in order to properly allocate resources to those areas or groups for both current and future disproportionate effect mitigation?

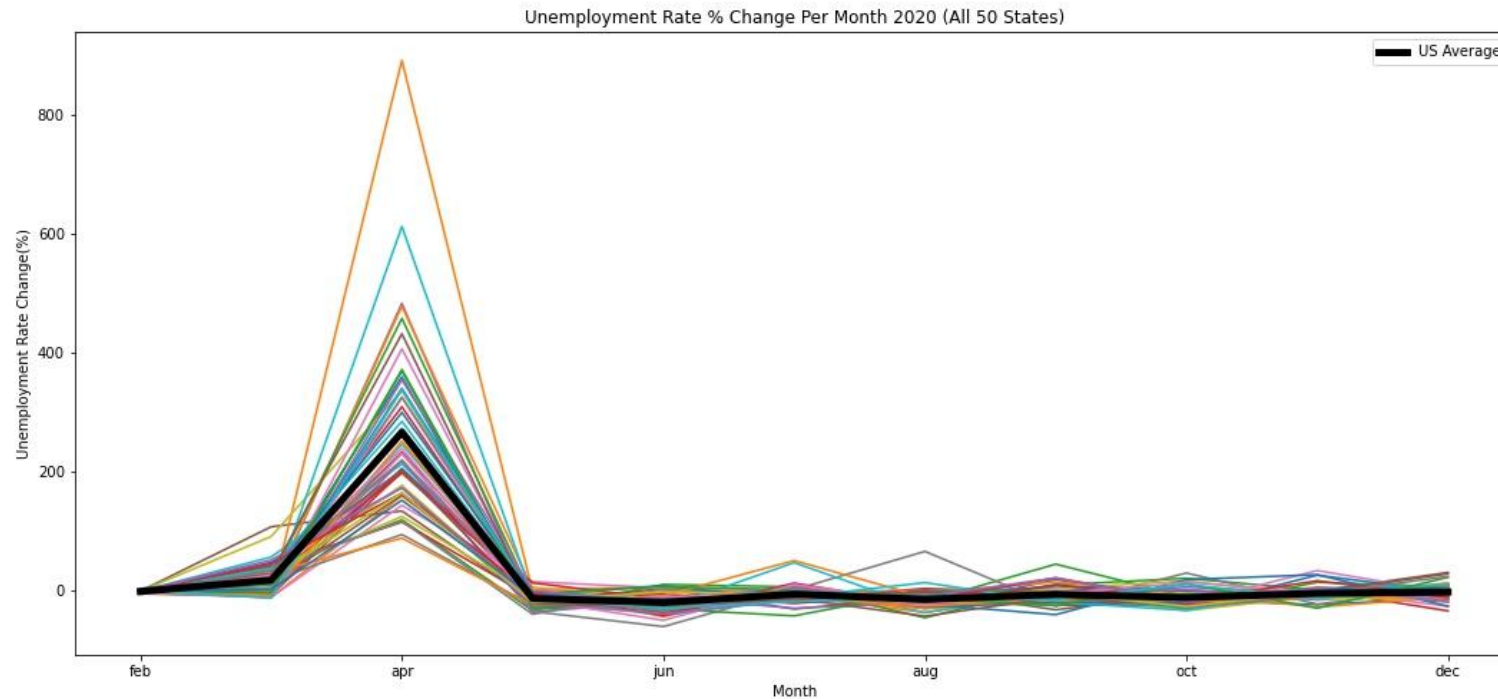
Solution Proposition

Attempt to cluster states based on demographic, socioeconomic, and COVID-19 data to help identify factors that may lead to disproportionate pandemic effects.

Data Sources

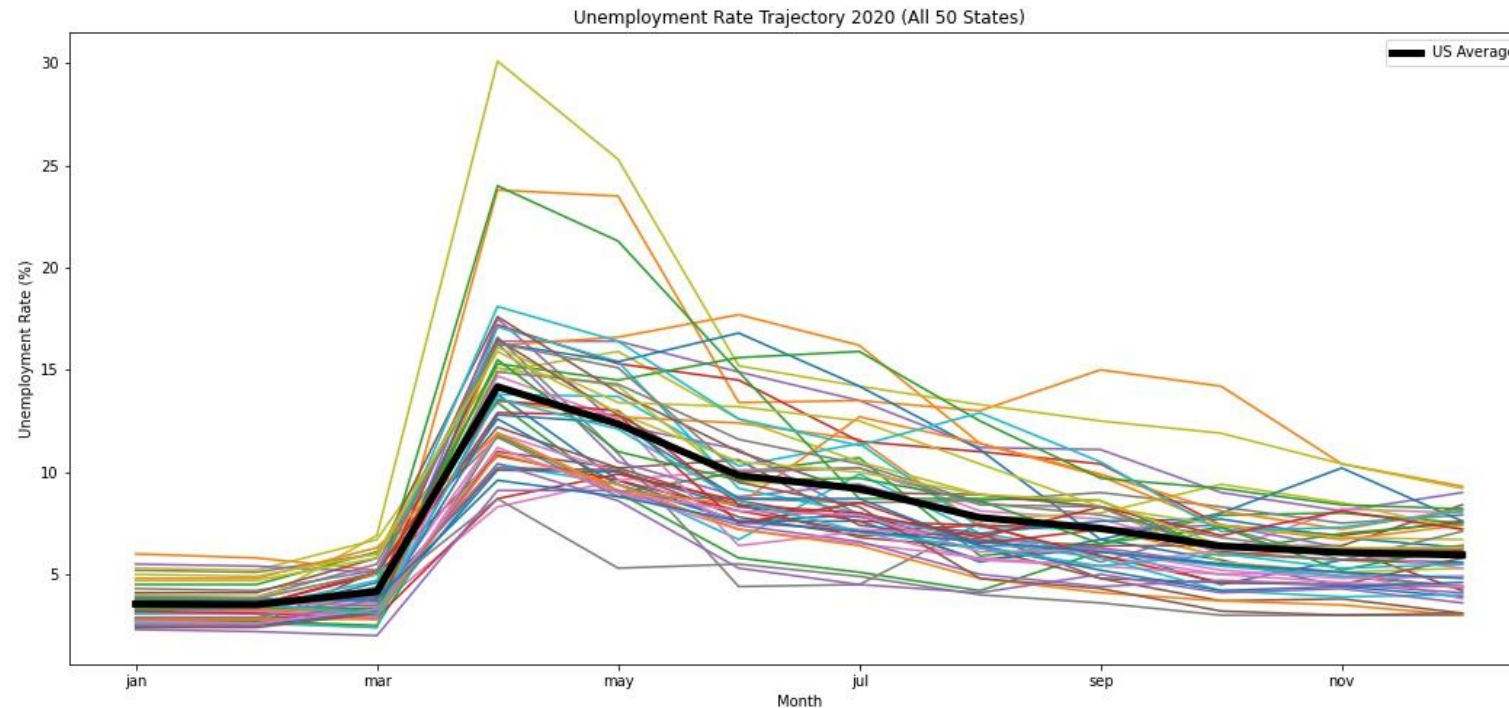
1. Bureau of Labor Statistics
2. John Hopkins University center
3. CDC.gov
4. USAFACT

Unemployment EDA



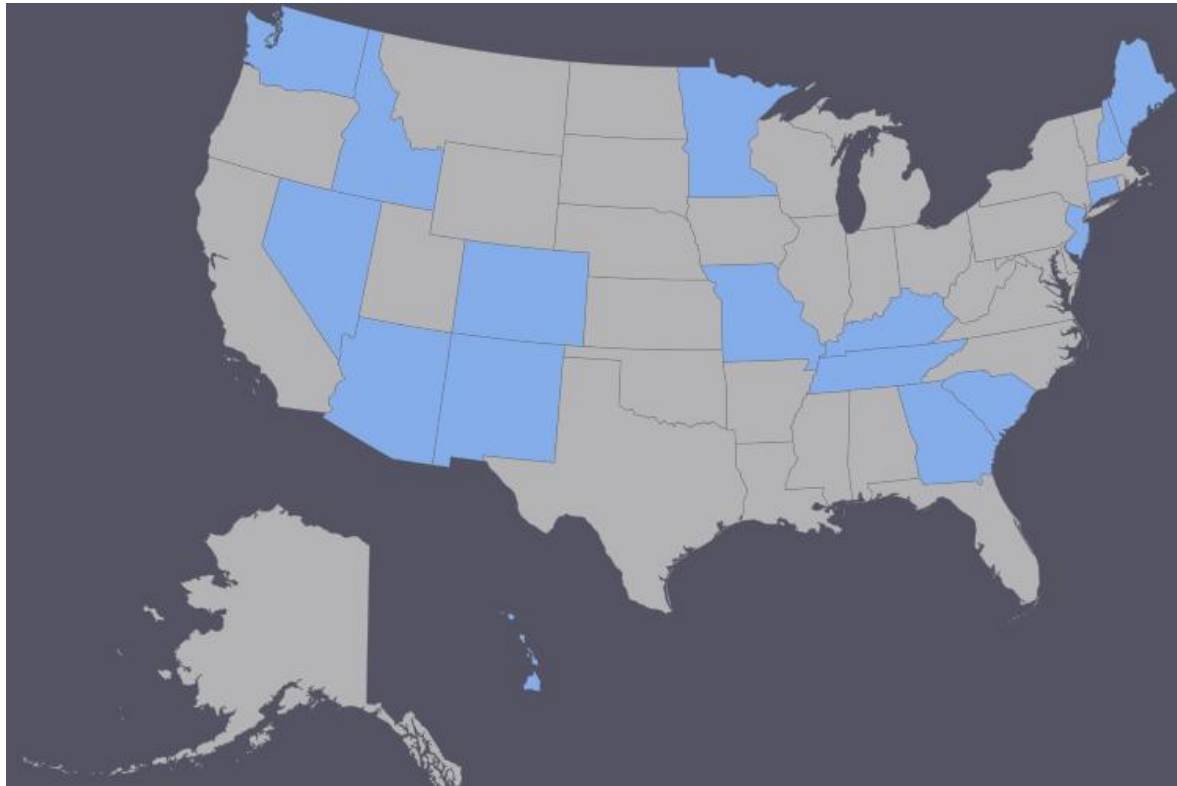
- Percent change MoM better compares states to one another
- Use this to better interpret abnormal unemployment rates

Unemployment EDA



- Several states exhibiting abnormal unemployment compared to the national average
- Identify which states exhibited abnormal unemployment

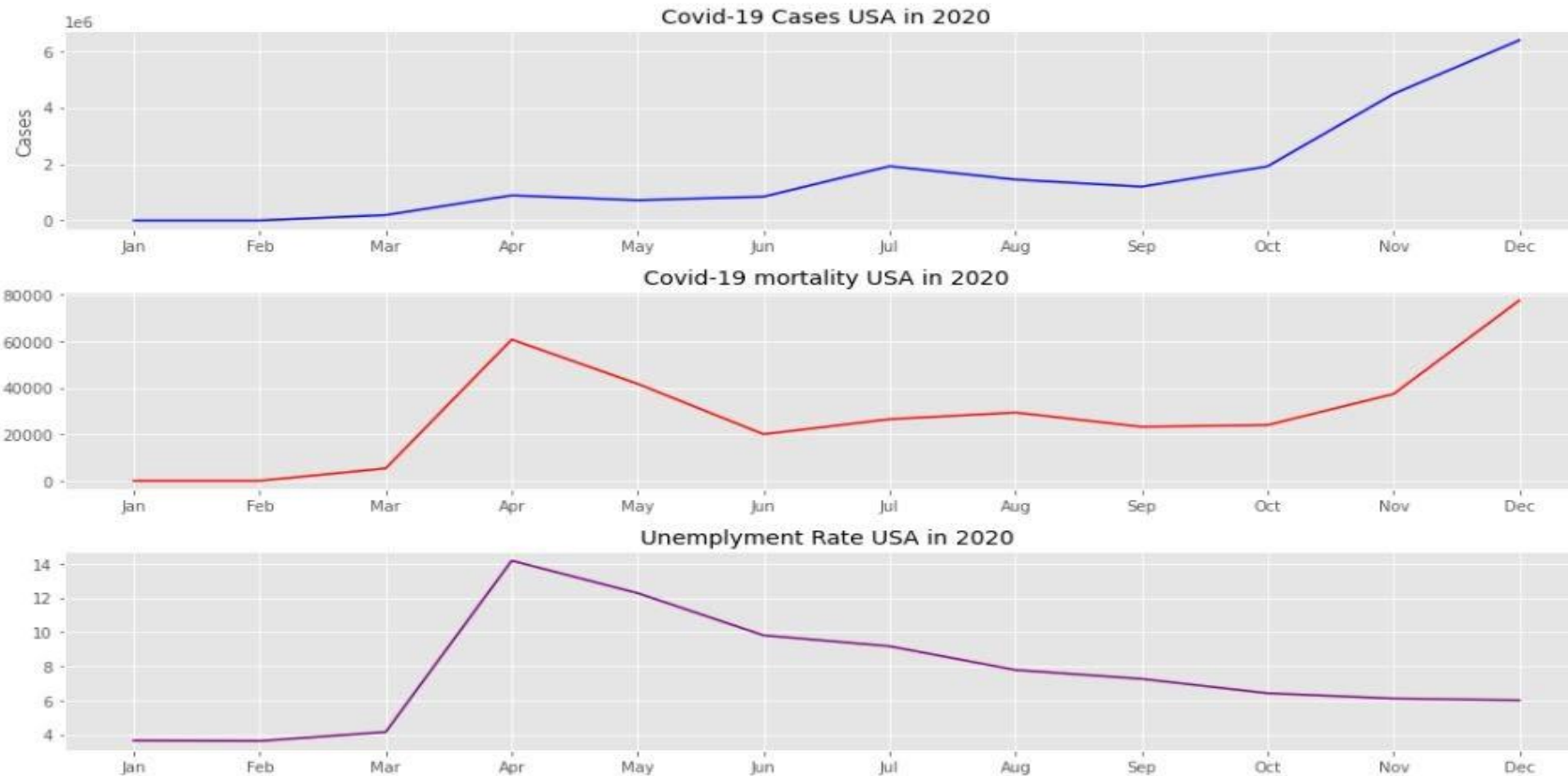
Unemployment EDA



The following states were determined to have abnormally high unemployment rates in 2020 when compared with the national average:

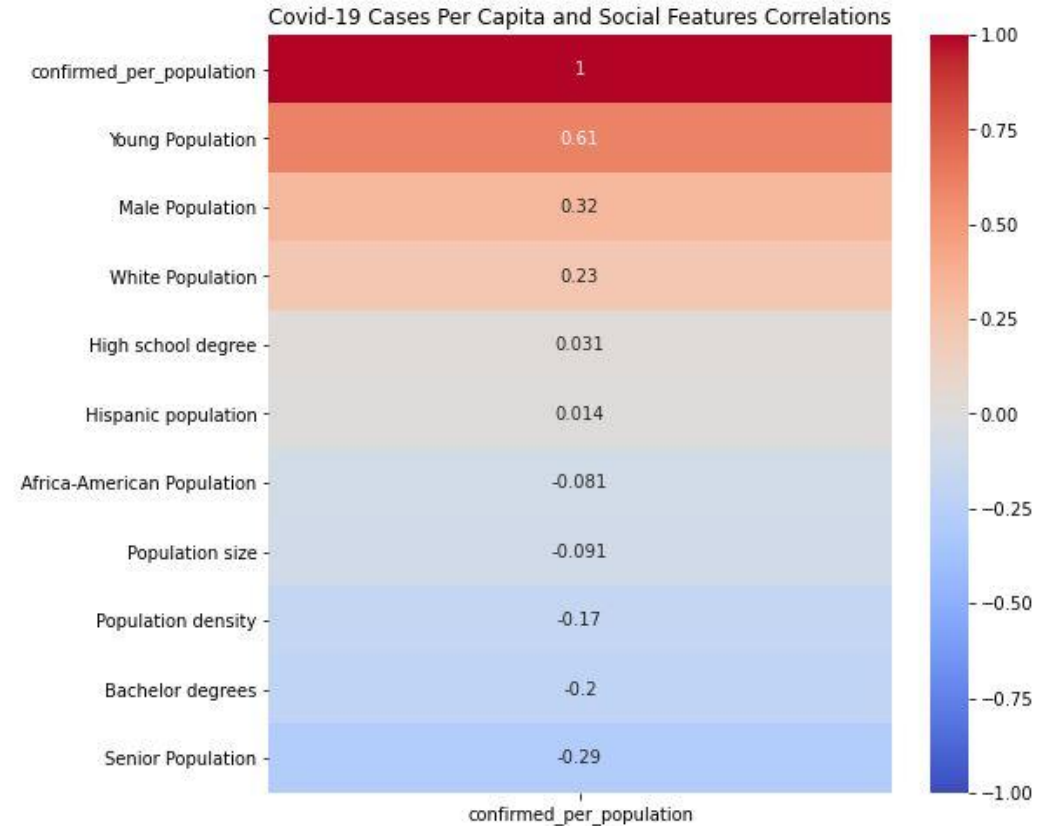
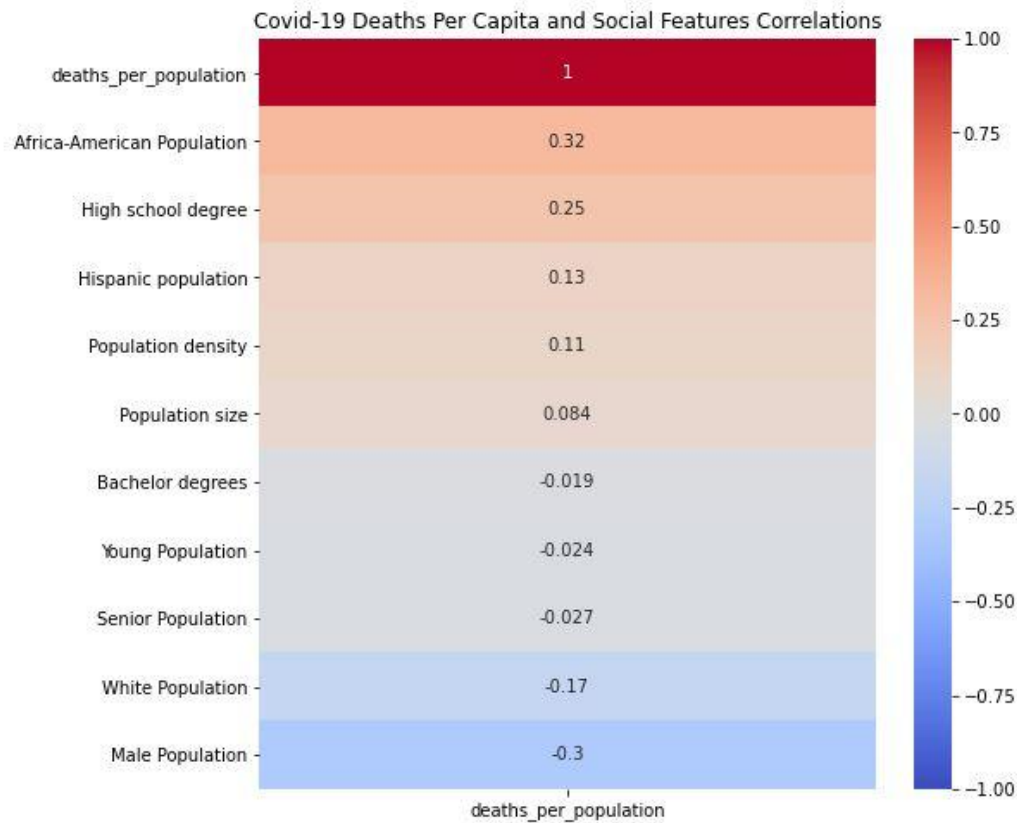
- Arizona
- Colorado
- Connecticut
- Georgia
- Hawaii
- Idaho
- Kentucky
- Maine
- Minnesota
- Missouri
- Nevada
- New Hampshire
- New Jersey
- New Mexico
- South Carolina
- Tennessee
- Washington

Covid-19 Cases, mortality and Unemployment Rate



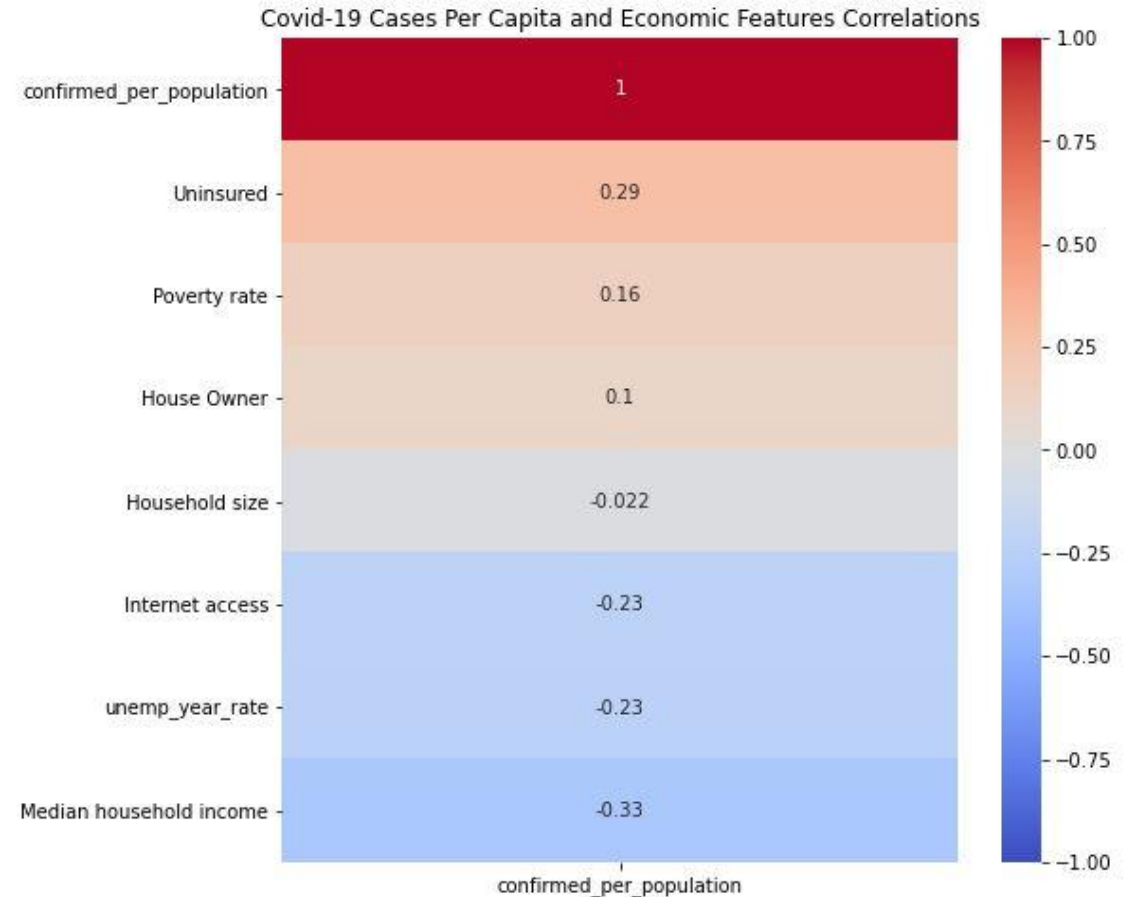
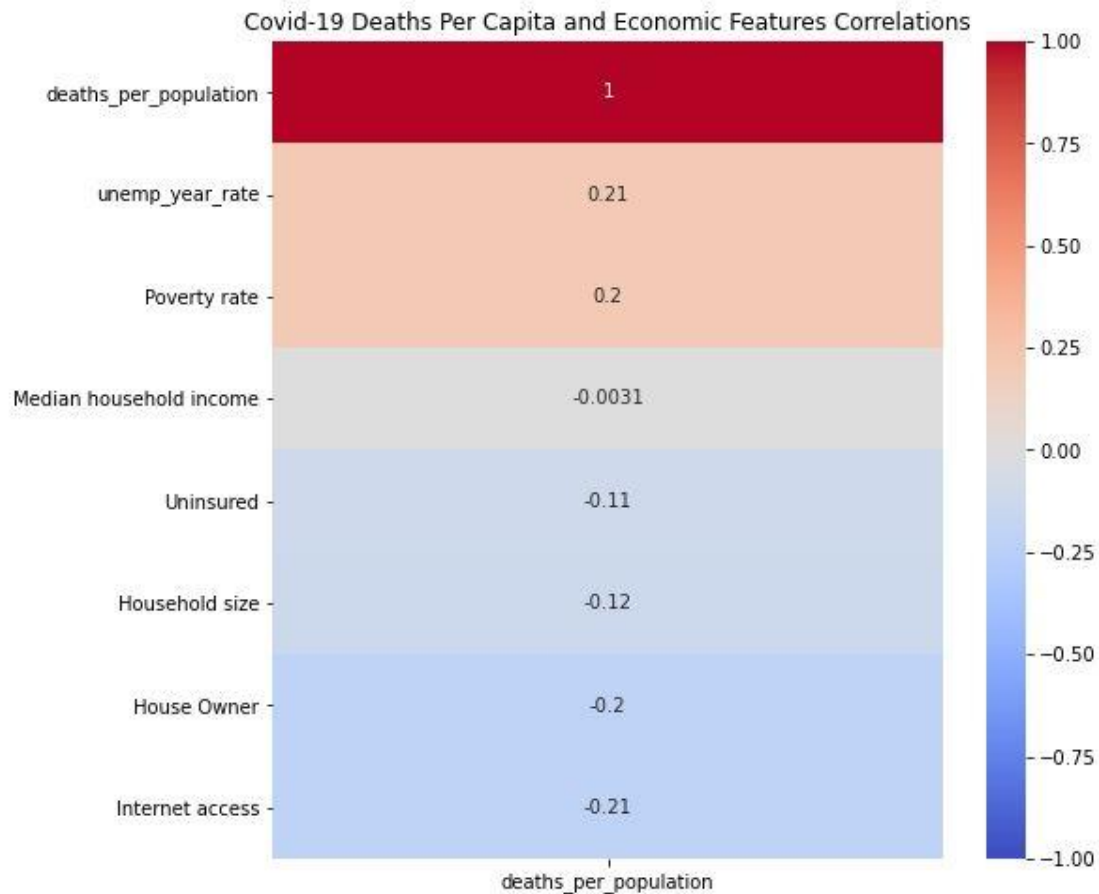
- A big spike in both death and unemployment in April
- Unemployment line follow mortality covid-19 line until July, then went down when mortality went up

Covid-19 and Demographics



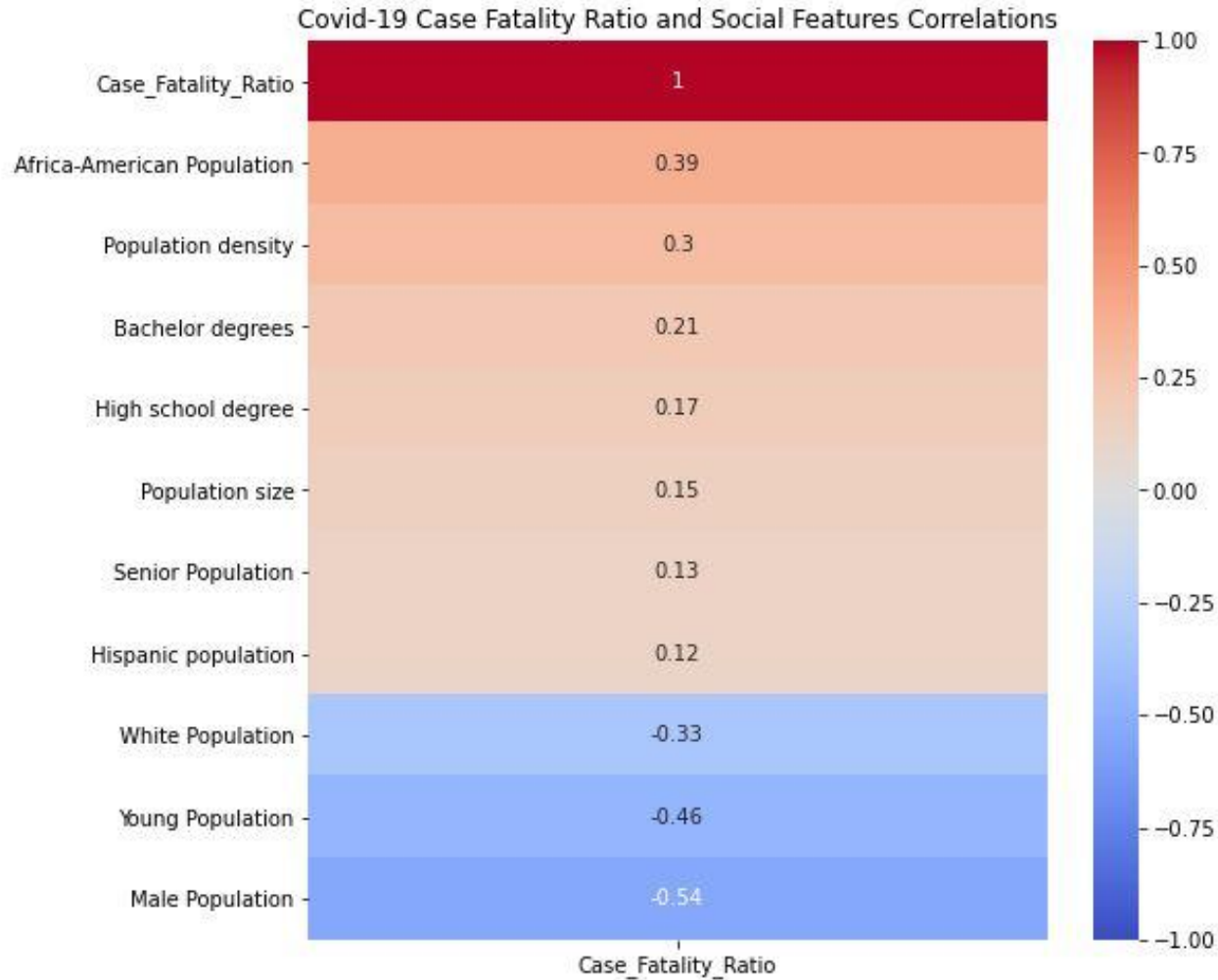
- The deaths per capita heatmap the African-American population measures the highest correlational value of 0.32
- The white population correlates negatively with deaths per capita at -0.17

Covid-19 and Economic



- As we can see poverty rate correlates at a rate of 0.2 and 0.16 with deaths and cases respectively
- Unemployment rate we see it correlates negatively with confirmed cases (-0.23), but positively with deaths (0.21) - unemployment leads to less cases

CFR: Case Fatality Ratio



- Number of Deaths from Disease / Number of Confirmed Cases of Disease) * 100
- African-American population as the highest correlation amongst features, Hispanic as positive but more moderate, but White population again as negatively correlated
- Senior population has positive but young population has significant negative correlation

Modeling

- **DBSCAN** is somewhat limited when it comes to clustering observations that may not be overly distinct in separation.
- **Kmeans** is one of the easier methods to understand and other clustering techniques
- Using Demographic, Social and Economic features

DBSCAN Model Performance

-Consistent with our EDA we notice that across each cluster, as CFR rises, so do the minority populations

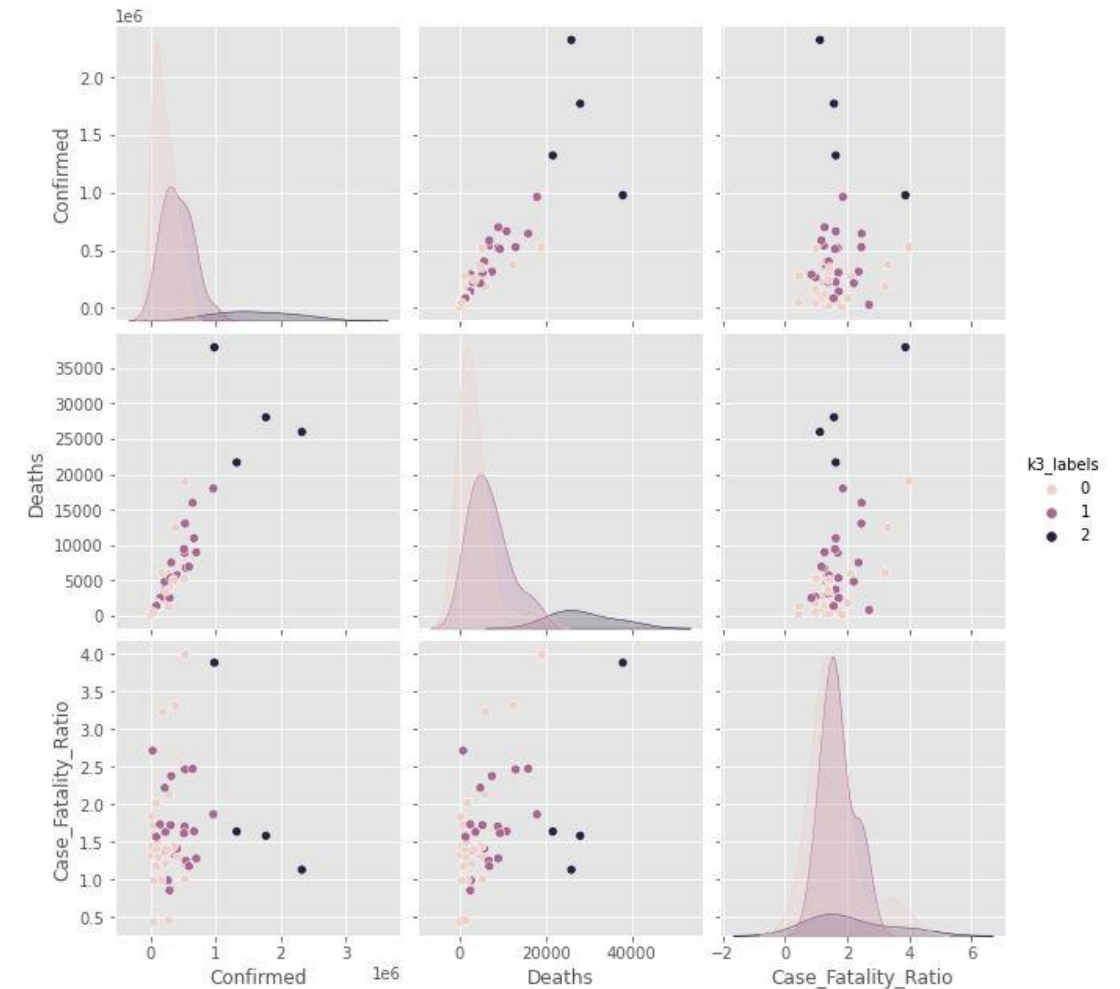
-inversely the white population percentage decreases

-Covid disproportionately affect minority communities at a concerning rate

cluster	0	1	2	3
Median household income	62974.718750	83475.666667	62865.250000	48432.500000
Poverty rate	11.846875	9.533333	12.300000	19.300000
hospital	101.375000	83.000000	223.500000	164.500000
White Population	83.756250	75.833333	78.450000	61.450000
Africa-American Population	10.243750	12.733333	15.300000	36.150000
Hispanic population	7.778125	16.733333	14.250000	4.200000
Deaths	3497.593750	12486.666667	17154.000000	6137.500000
Case_Fatality_Ratio	1.297728	3.509479	2.109361	2.296607
deaths_per_population	0.000834	0.001876	0.001244	0.001610
confirmed_per_population	0.065681	0.055360	0.060246	0.070166
unemp_year_rate	6.655312	8.963333	9.042500	8.420000

K-Means Model Performance

- 3 clusters is the best among 2, 4, 6, 8 and 10 clusters
- 3 clusters containing 26 states, 21 states, and then 4 states



Improvements

1. Examine data on a lower level (county/city data)
2. Use more data (data from other pandemics)

Recommendation

- We recommend that our audience pay extra attention to the minority demographics when providing aid and resources as they have shown through our data to be the most concerning categories.
- States to closer examine based off models:
Connecticut, Florida, Illinois, Louisiana, Massachusetts, Michigan, Mississippi, New Jersey, Pennsylvania
- For more complete analysis we request access to same data but on a county level, more precise, more defined clusters

Questions?