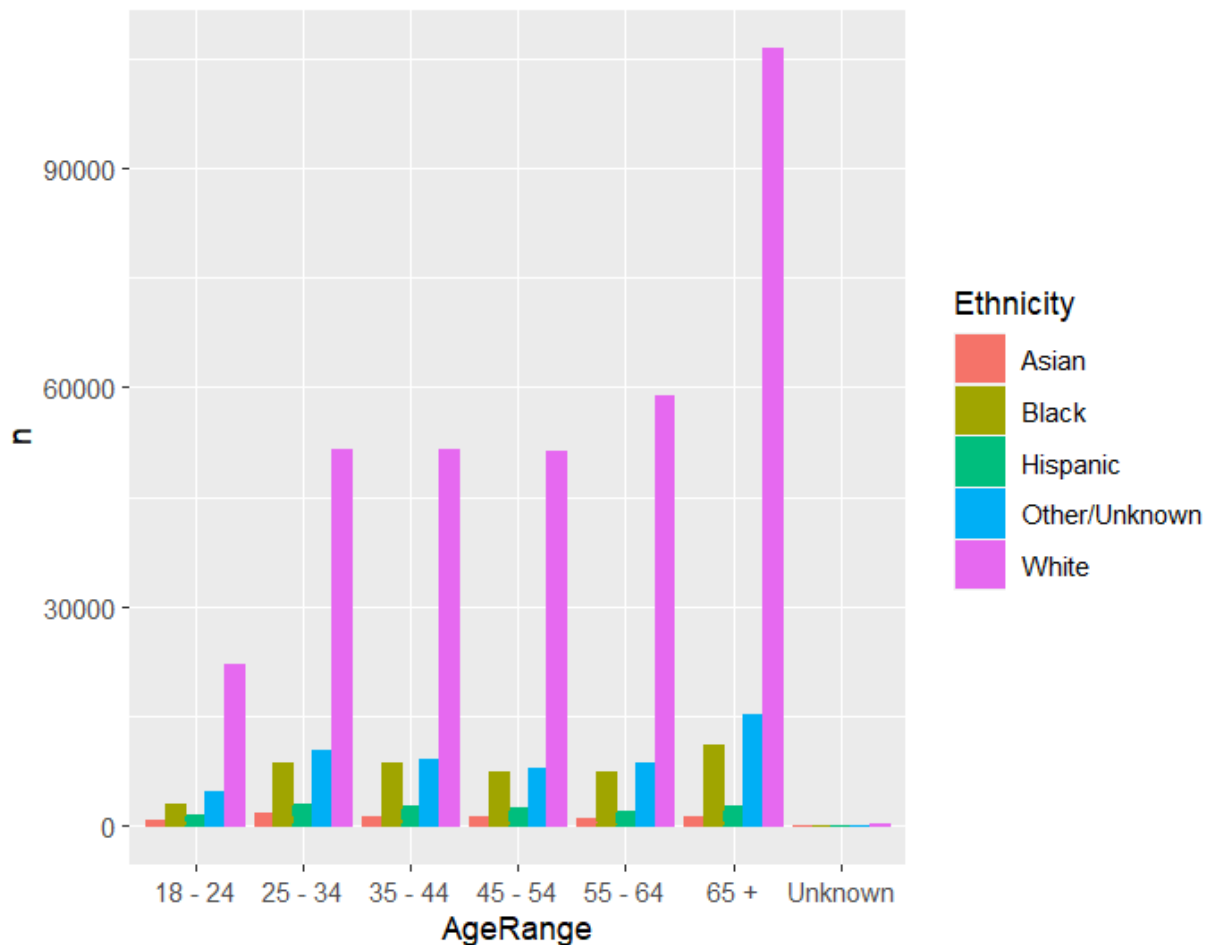


Causeway Solutions Interview Sample Data Set Assessment

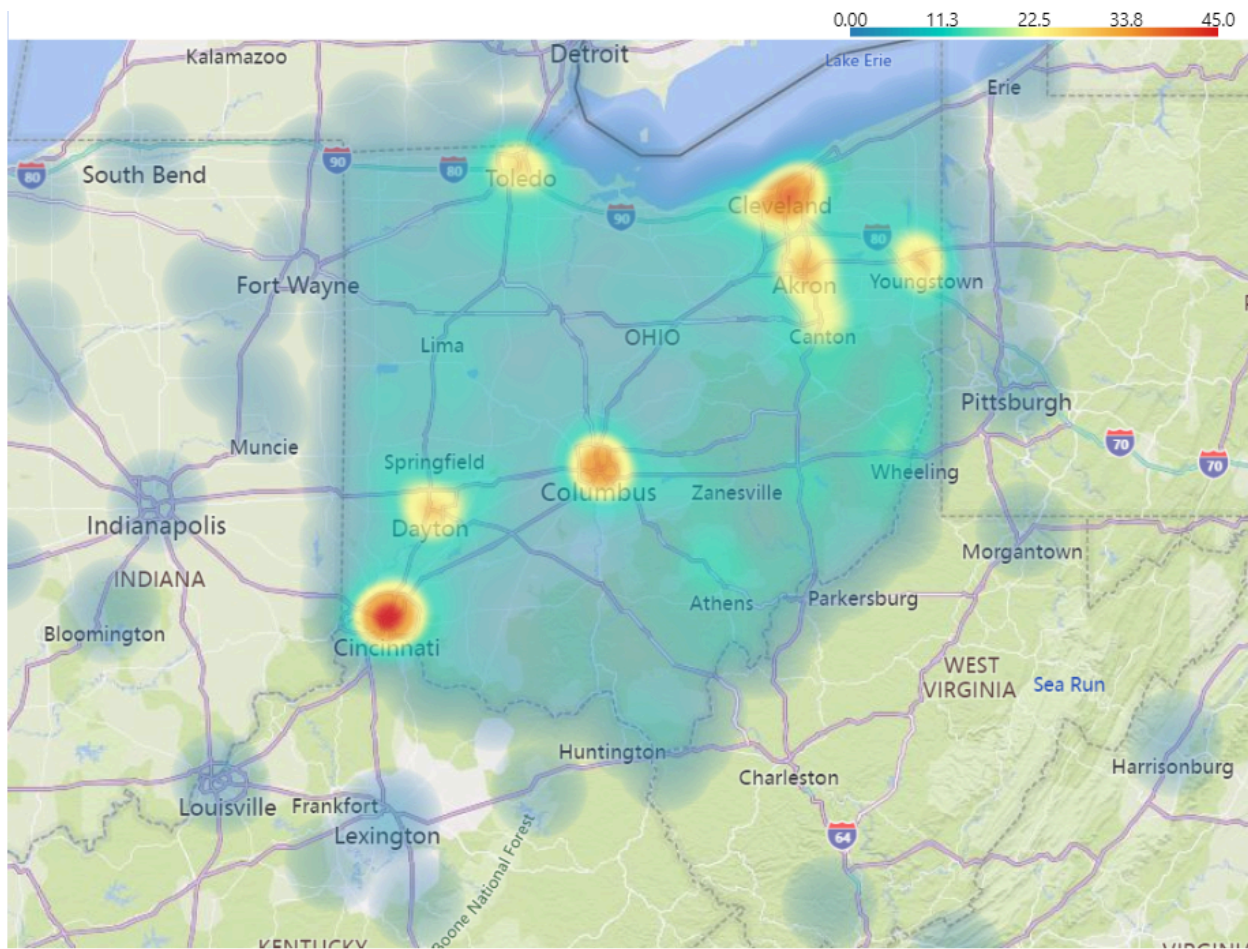
Charles Cantwell



A discrepancy in the data was found by counting the number of White voters in the dataset. According to 2020 US Census data, the percentage of people reporting White alone, not Hispanic or Latino, is 77.3%. Whereas within this dataset, the percentage of White voters is 73.6%. However, this discrepancy can be reduced by reallocating some of those classified as unknown into the White category.

The percent of Black voters in this dataset is 9.5% which is less than the state average of 13.3%. Additionally, the percentage of those who are unknown race is 12.3%.

What is most concerning is the overrepresentation of the elderly (65 years old or more) who represent 27.8% of the dataset when the percent of Ohio that is within that age bracket is 18.4%. Elderly people are more likely to vote than younger people. But according to Census data for the 2020 general election, the 65+ year age group comprised 26% of the vote which is 1.8% less than the dataset. As age increases, the likelihood of voting republican increases. So with the older demographics contributing the most to voting, Biden will be less favorable on average.

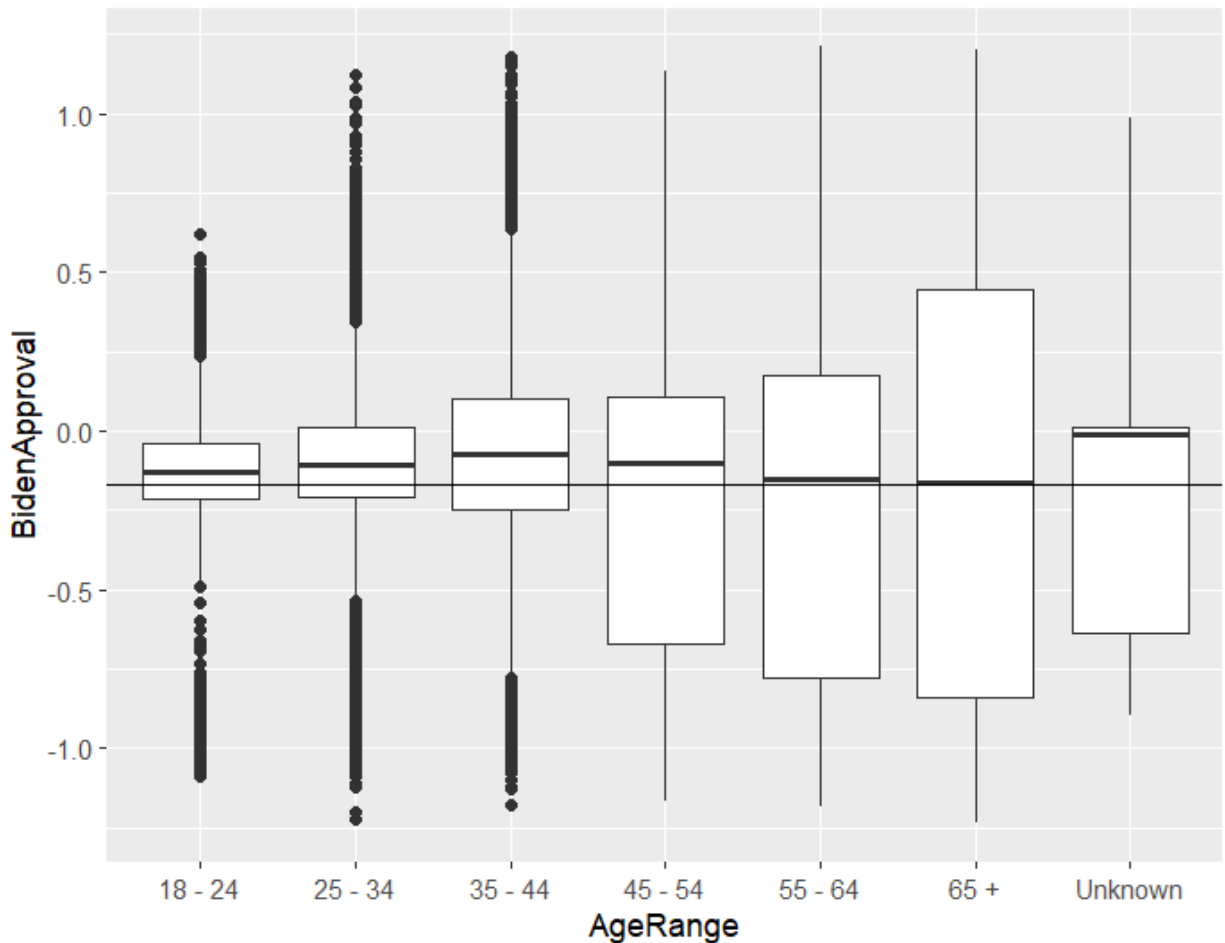


The above heatmap is generated using the average value of BidenApproval which is a metric within the dataset that ranges from about -1 to 1 expressing each voter's approval. The approval rating is averaged for each zip code. It is common for larger cities and more affluent areas to approve of democratic presidents, and this is seen in warmer areas of the heatmap.

The original data was rescaled from the interval $(-1, 1)$ to the interval $(0, 2)$ to accommodate the parameters within Power BI. Additionally, publicly available longitude and latitude data was combined with the zip codes to generate this heatmap.

The heatmap is largely contained to the Ohio state borders, but the data points outside of the state borders are either incorrect zip codes or the current location of the voter. It is also possible that these voters are filling out absentee ballots from outside the state.

Furthermore, the Electoral College of Ohio is a winner-takes-all state meaning that whoever gets the majority vote of Ohio for president, all eighteen of Ohio's presidential votes go to the winner. Thus regional democratic wins in urban areas are not effective in winning the votes of Ohio.



The average BidenApproval rating is largely skewed by the older population. The average approval rating for Biden in Ohio is -0.1662765. The above box plot shows that all of the medians are below zero, but the mean is further below than the median for each. The third quartile and minimum extend more below the average than the first quartile and maximum causing the mean to be weighted to the negative values within the older age groups. The surplus of outliers for the younger age group causes further decrease in the mean approval rating.

Cramer's V test which is based upon a Chi-squared test was used to assess the correlation between the two categorical variables AgeRange and Ethnicity and found no statistical significance with a value of $V = 0.03$.

No predictive modeling was done as it is out of scope of this project.