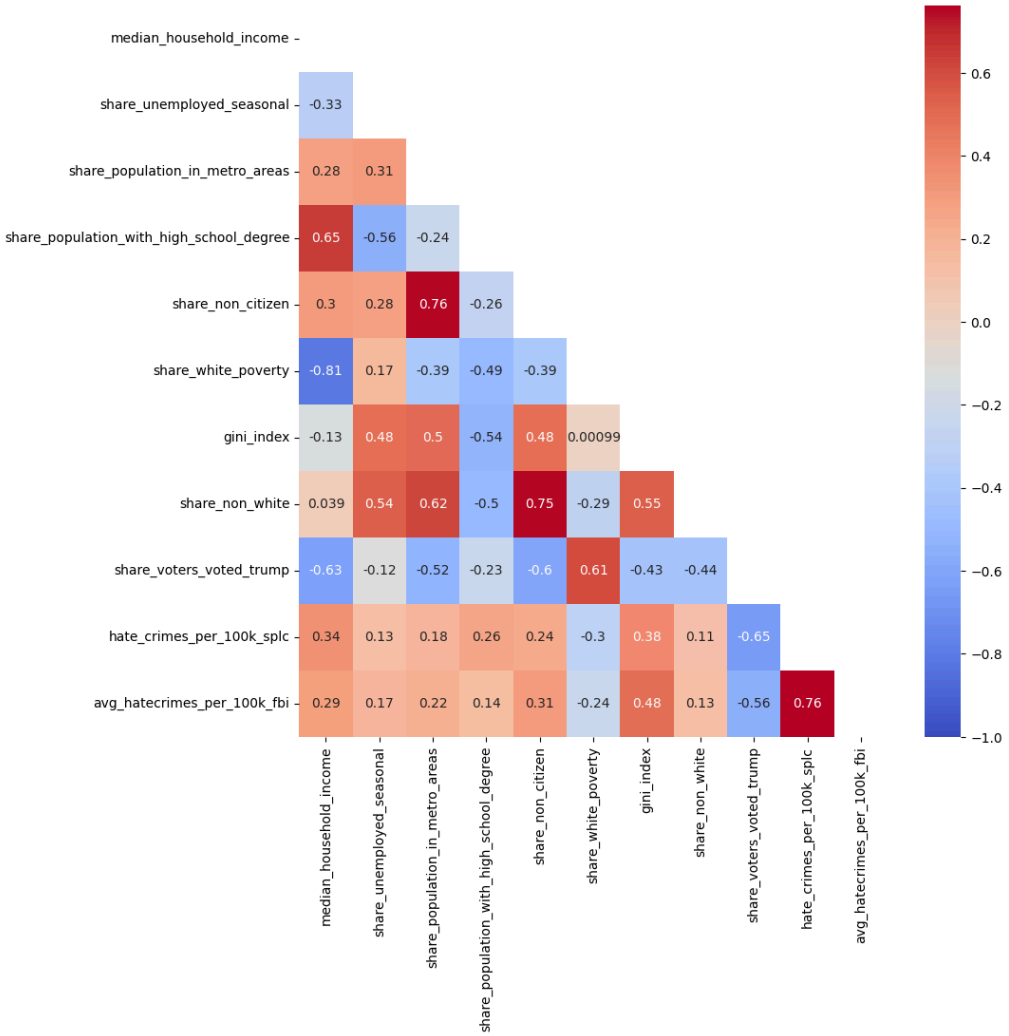
INSIGHTS:

Upon running a small correlation code on the parameters, it was observed that the gini index was correlated to both the number of hate crimes from SPLC as well as the FBI. Hence I chose these parameters for testing the relationship between hate crime and income inequality.

Secondly, the number of hate crimes was related to the columns “share_non_citizen” and “share_non_white”, which is why for the second regression model I chose these parameters to demonstrate how to predict the number of hate crimes and how it is related to these said parameters.



2.1. How does income inequality relate to the number of hate crimes and hate incidents?

To understand the relation between income inequality and the number of hate crimes in the various states, we will have to derive a linear regression model based on these parameters. Consequently, we built a linear regression model where the independent variable was the number of hate crimes, while the response variable was the gini index. Upon using OLS to uid the model, these are the interpretations from it:

1. R-squared : This typically measures the goodness of the model. The model gives an understanding that about 23.5% of the variability in the gini index can be explained by this model.
2. The constant coefficient was 0.4417. This is indicative of the gini index when the hate crime is 0. The coefficient of the number of hate crimes as per splc is 0.0024 and as per fbi was 0.0056. This means that when there is a one unit change in the number of hate crimes as per splc, the gini index increases by 0.0024. Similarly, when there is a one unit change in the number of hate crimes as per fbi, the gini index increases by 0.0056. The positive correlation between these two parameters tells us that the higher the number of hate crimes, the more is the income inequality.
3. We can rely on this understanding because the p-value is 0.033 for our independent variable suggesting that this is statistically significant in predicting the Gini index.

2.2. How can we predict the number of hate crimes and hate incidents from the race/nature of the population?

The prediction of the number of hate crimes and hate incidents based on the race of the population can be done via linear regression. Here we would have the number of non-citizens and the number of non-white population as our independent variables, while the number of hate crimes would be the response/outcome variable. Upon building the linear regression model, it was found the model could explain about 7.1% variability in the number of hate crimes. There are other parameter discussed below that would give us better insight about the prediction:

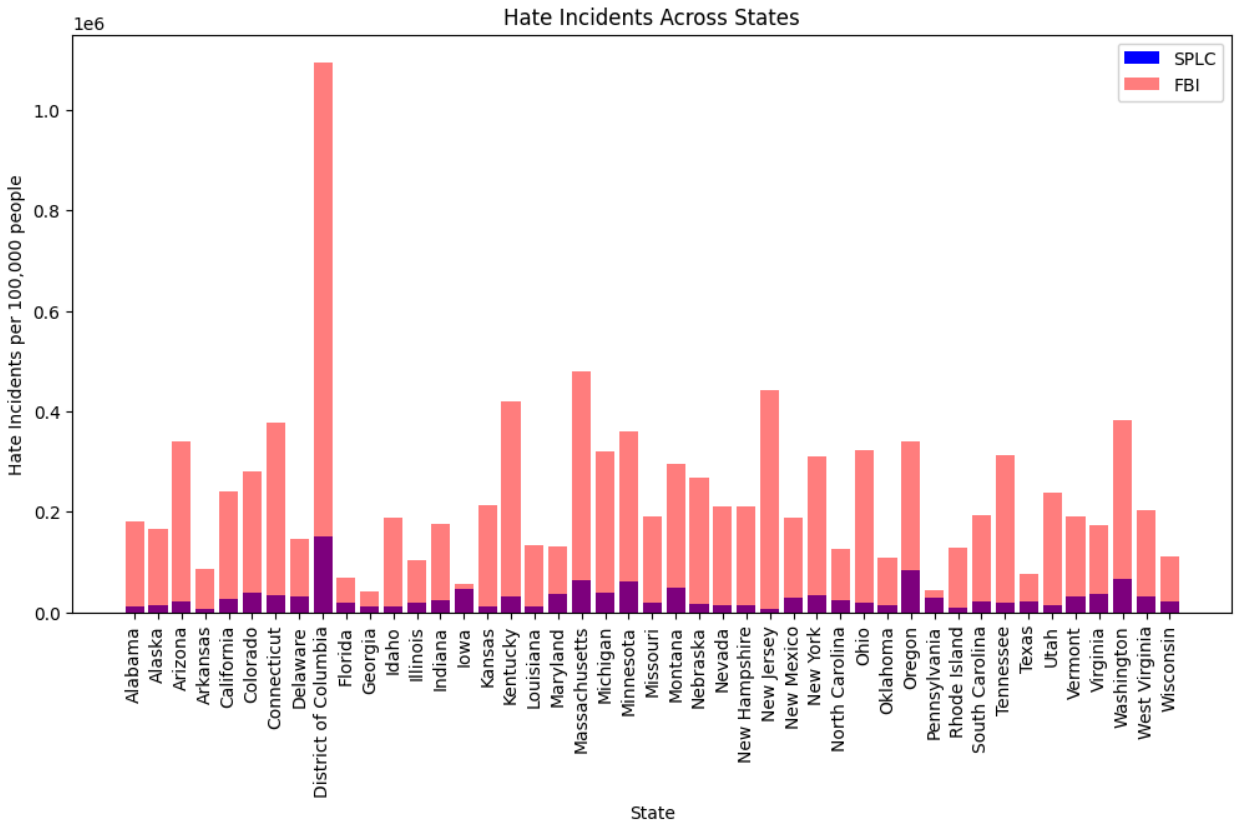
1. The constant coefficient was 0.2275 which tells us that this is the number of hate crimes when the other input variables are 0.

The coefficient for the parameter “share non white” is -0.2813 which indicates that for one unit change in this input variable, the crime rate would decrease by 0.2813. On the other hand, a non citizen of the state is more likely to increase the crime rate by 2.9694.

We understand from this that based on certain parameters such as non white or non-citizen, we could predict the number of crimes.

2.3. How does the number of hate crimes vary across states? Is there any similarity in the number of hate incidents (per 100,000 people) between some states than in others — both according to the SPLC after the election and the FBI before it?

The number of crimes vary tremendously across states with the District of Columbia having the highest number of hate crimes reported by both SPLC and FBI. It seems like a sinusoidal graph with peaks as high as over 100,000 and low as 1000.



To check if there are similarities between states, I used k-means clustering method to understand the behaviour of the dataset based on the states. Taking 3 clusters(i.e. k= 3), we see that states like Arkansas, Georgia, Florida, Texas, Illinois have been grouped in one cluster due to similar hate crime numbers while states such as New York, Kentucky, Connecticut are grouped together. DC is exceptionally high on their hate crimes which is why it seems like its an outlier and its in a cluster of its own.