

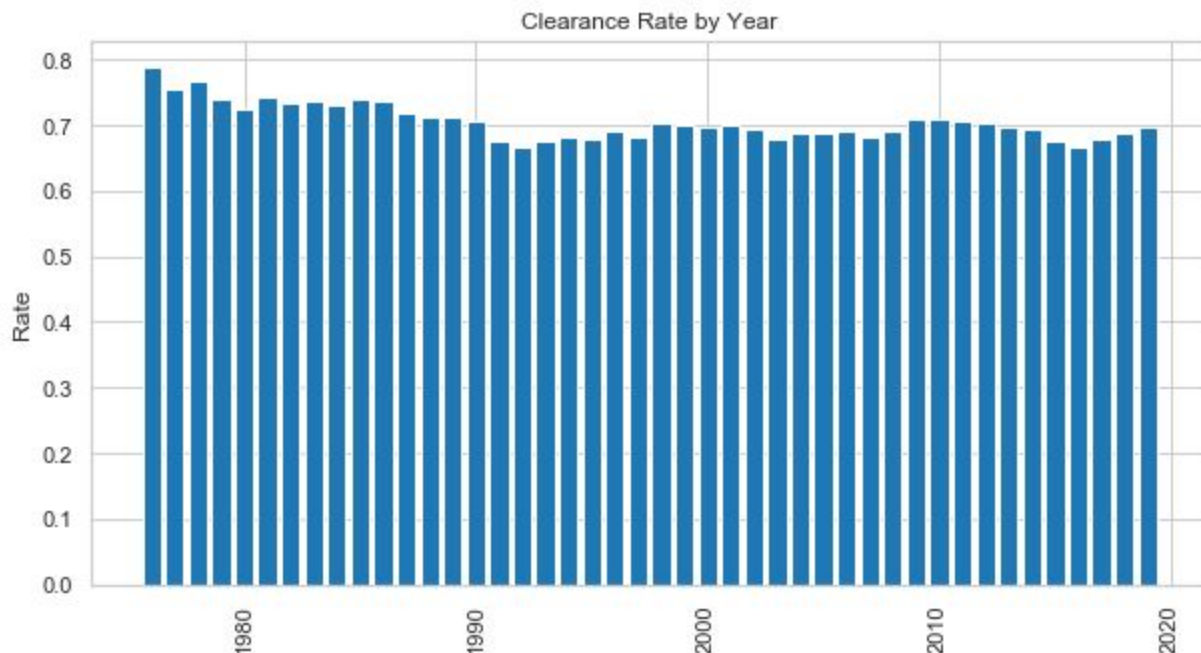
Adam Nunley

Springboard Data Science Student

Capstone Project 1

Murder Stats

In the United States, **30% of murders are unsolved**. And the clearance rate of murders has been getting worse over the past several decades.

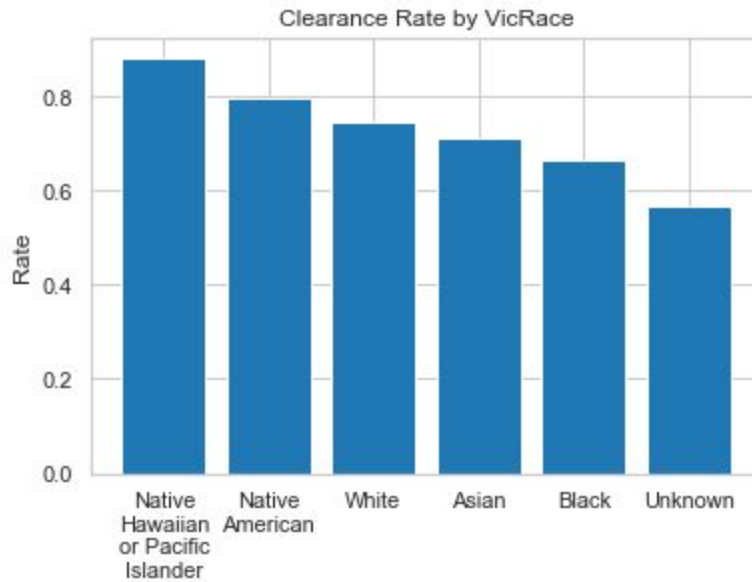


The primary question that I am hoping to answer with this project is:

Is it possible to use the Murder Accountability Project's dataset containing murders in the United States from 1976 to 2019 to predict the likelihood of a murder being solved?

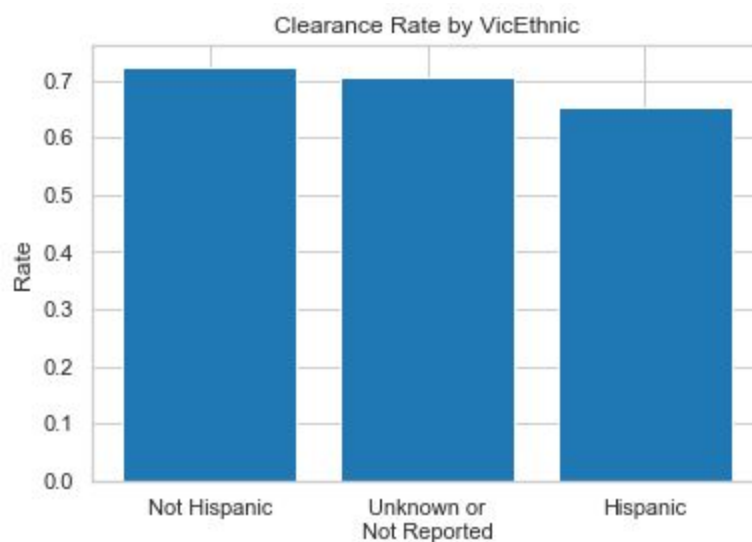
Before fitting the data to various machine learning models, I investigated the relationship that various features have with the predictor variable, a column indicating whether the case was solved or not. Specifically, I calculated the clearance rate (the number of solved cases divided by the total number of cases) for each category within each feature and charted the distribution. I found some results that were surprising and some that were not.

First, I investigated the relationship between the demographics of the victims and the clearance rate. I expected there to be significant differences especially when examining race because of the well-documented institutional and economic racism that exists in the United States.

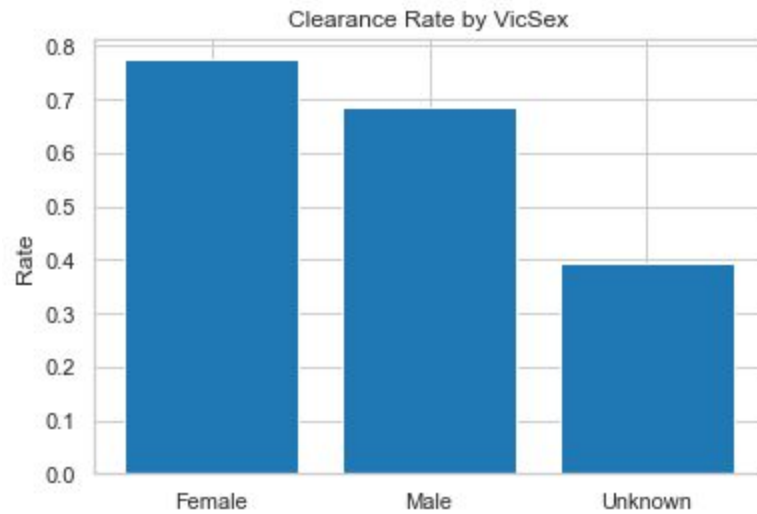


As I expected, murders involving white victims are roughly 10% more likely to be solved than those involving black victims. I was very surprised to see a higher clearance rate for Native Americans - "American Indian or Alaskan Native" in the dataset - and I am skeptical of it. Perhaps this results from the particularities of the specific data that was accessible to the nonprofit group that compiled the dataset. The category for "Native Hawaiian or Pacific Islander" is likely erroneous, because the entire dataset (of 804,751 victims) only contains 92 victims within that racial category.

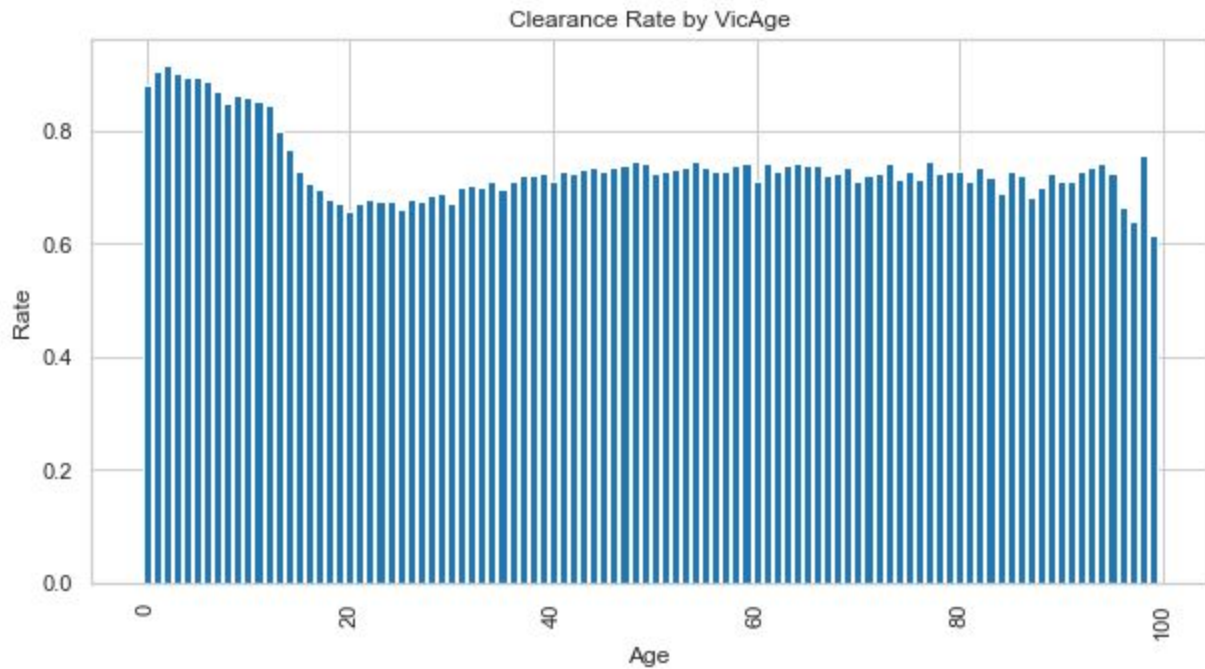
There was a similar relationship suggesting some kind of institutional and/or economic bias for the category of Hispanic vs non-Hispanic origin, which is (correctly) organized as a feature separate from the victim's race.



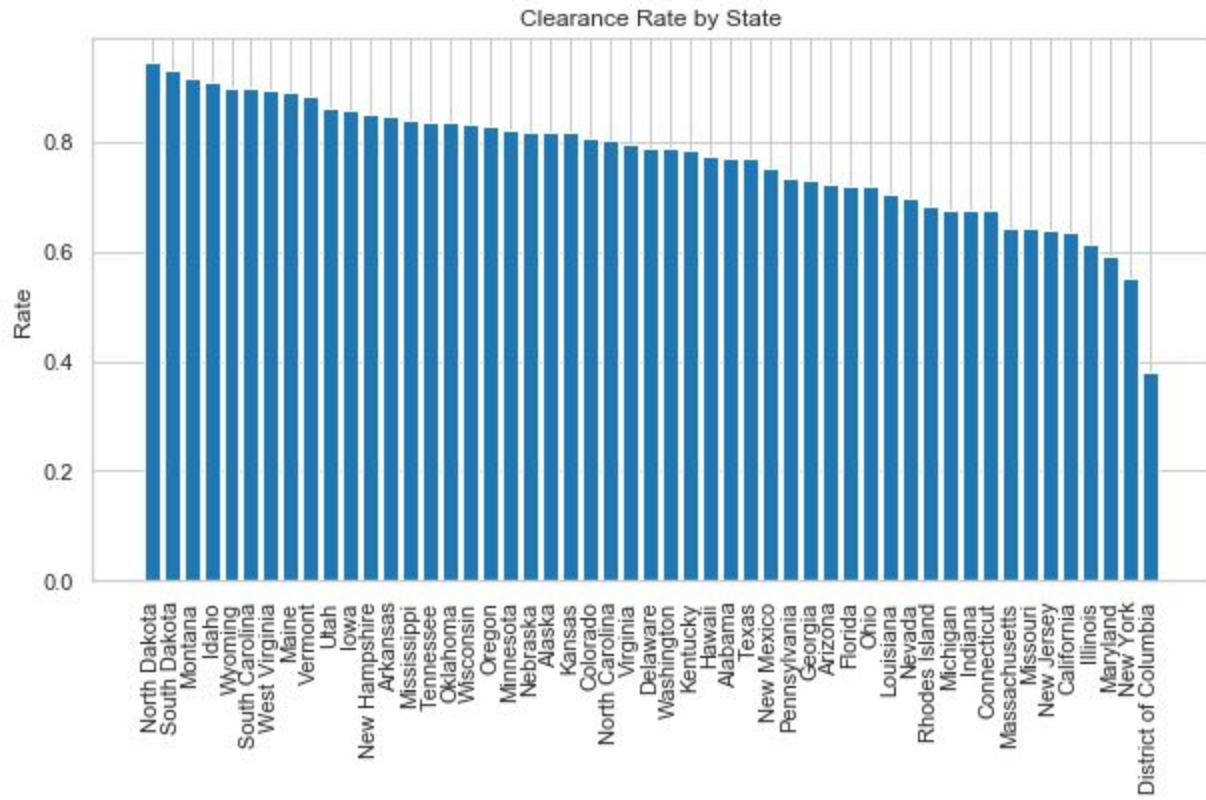
Continuing with demographics, I was surprised to find that the gender of the victim appears to have a significant effect on clearance rate.



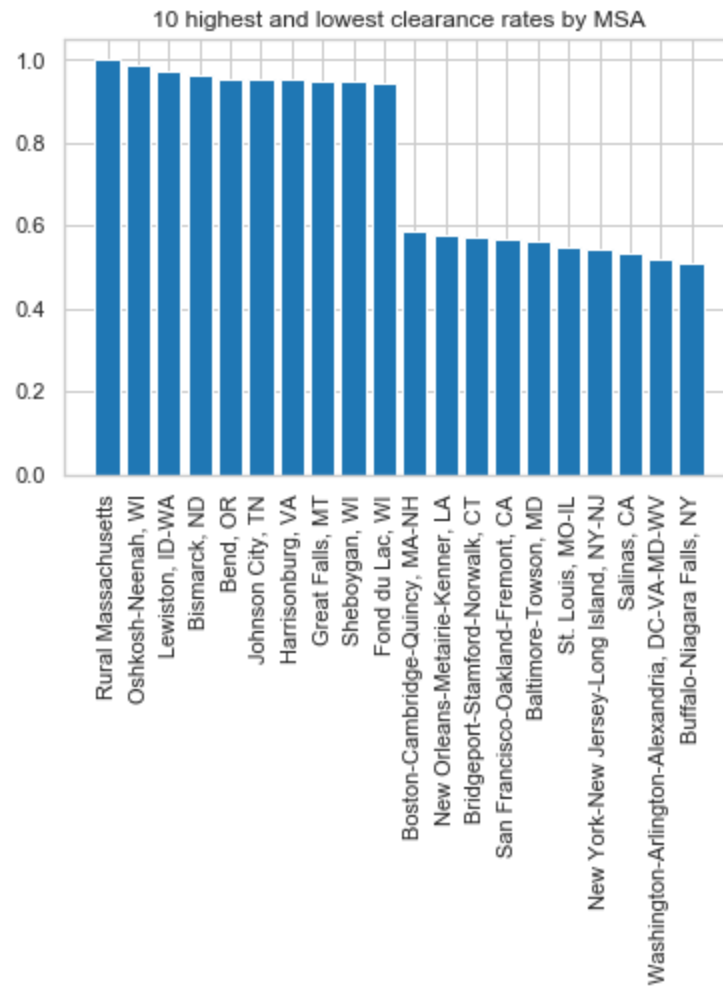
Age also has a clear effect, with the murders of young children being the most likely to be solved while teens and young adults are the least likely.



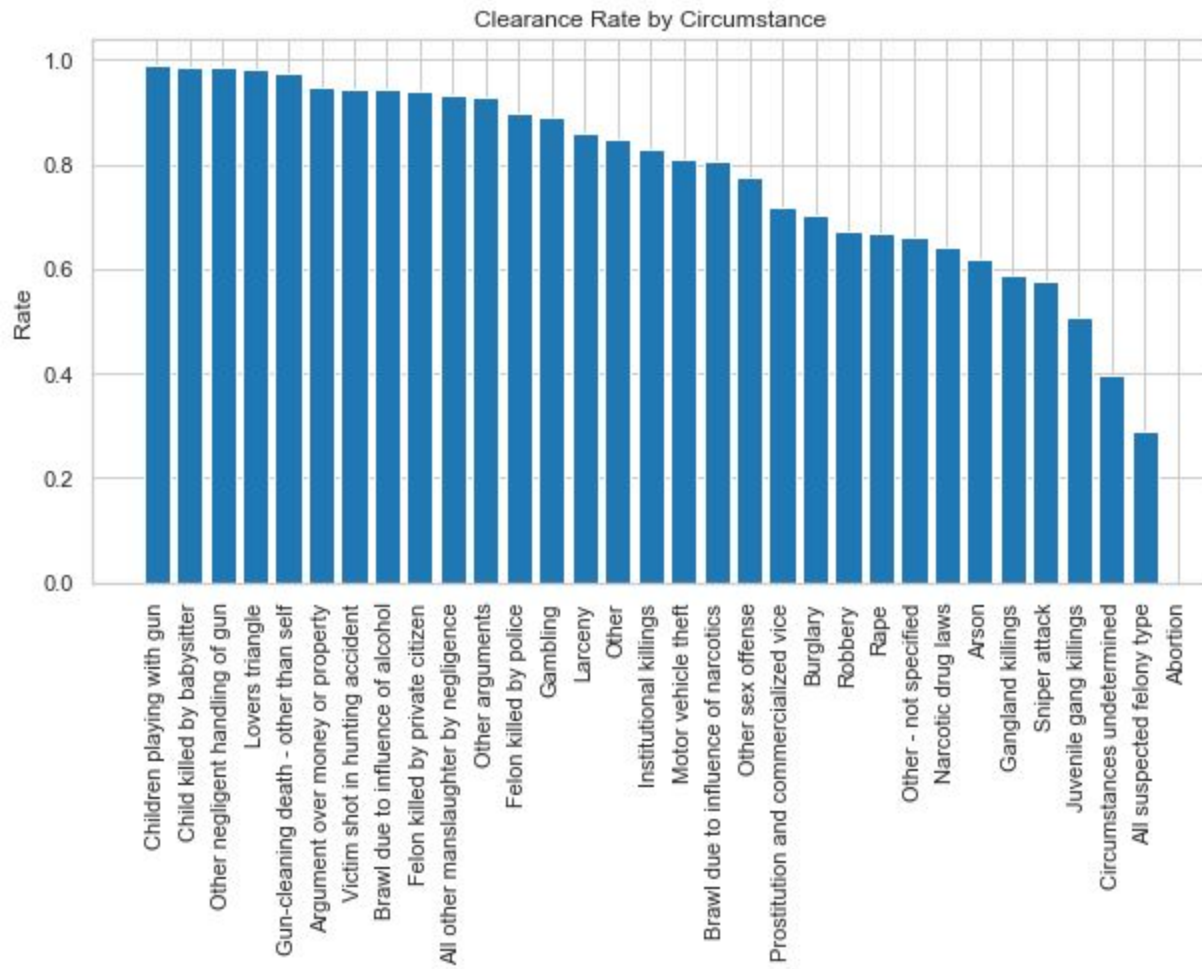
Beyond demographics, I found interesting relationships between the clearance rate and the location of the murder:



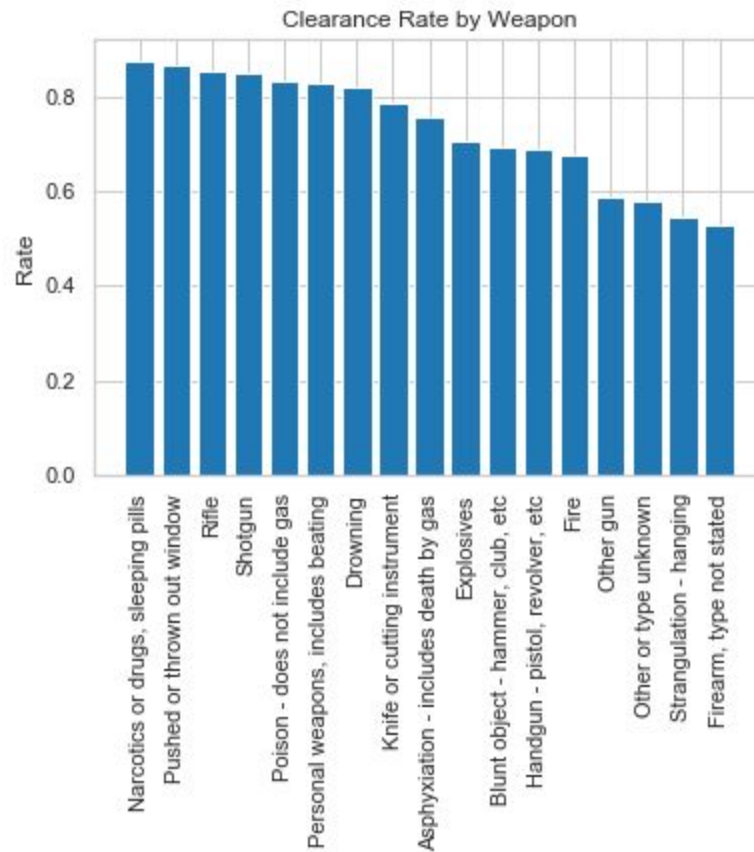
MSA = Metropolitan Statistical Area



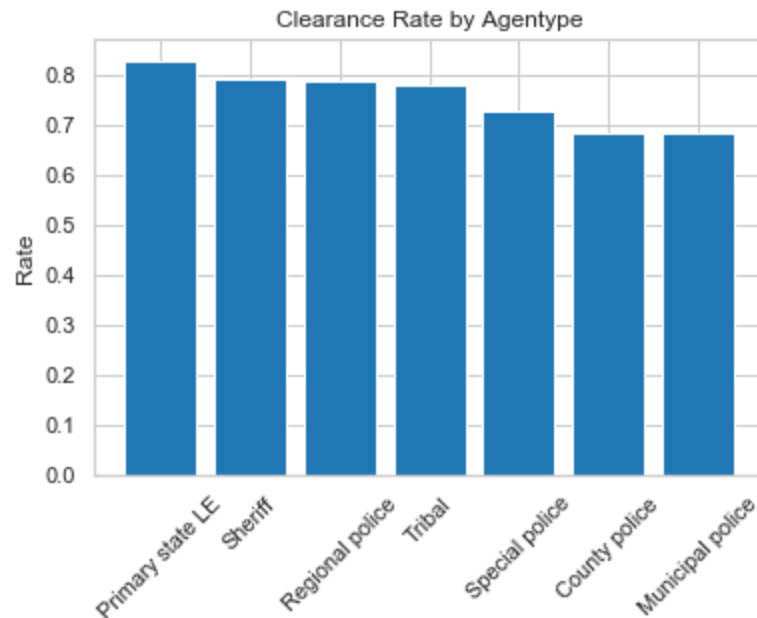
The circumstances surrounding the murder also have a major effect:



And, interestingly, the murder weapon also seems to have a large effect on the likelihood of a case being solved. For example, strangulations have around a 50% clearance rate, whereas beatings, poisonings, and drownings all have clearance rates greater than 80%.



Also of note is that the type of police agency investigating the crime seems to have a significant effect:



Similarly, the range of clearance rates among specific police agencies, of which the dataset contains 9,606 unique entries, varies widely.

