

Predict Crime rate in US



Predict number
of crime per
100k
population

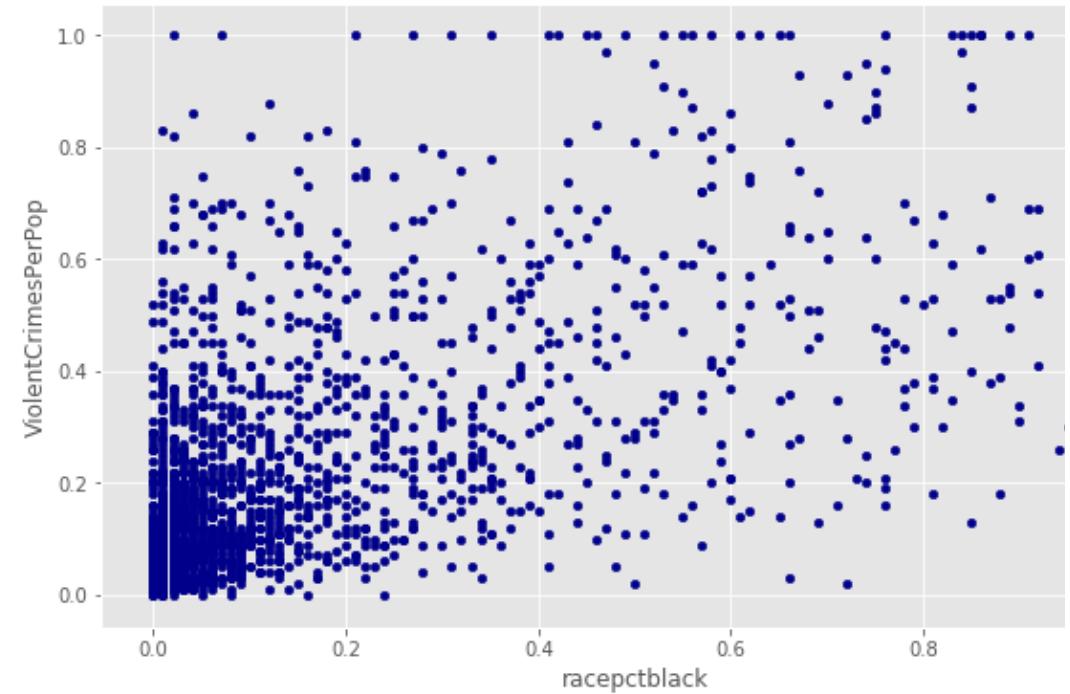
- Normally prediction is based on single dataset But with location data we can adding new feature in predictive model to improve result
- This predict can help governor to understand how to manage city to control crime rate

Data set and data clean

- Data set is from a public data source
 - <https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime>
 - Data set is 1994 rows * 256 columns
- Within this data some of the features contain lots of missing values.
 - First replace these missing values with np.nan
 - Remove all columns with missing values.

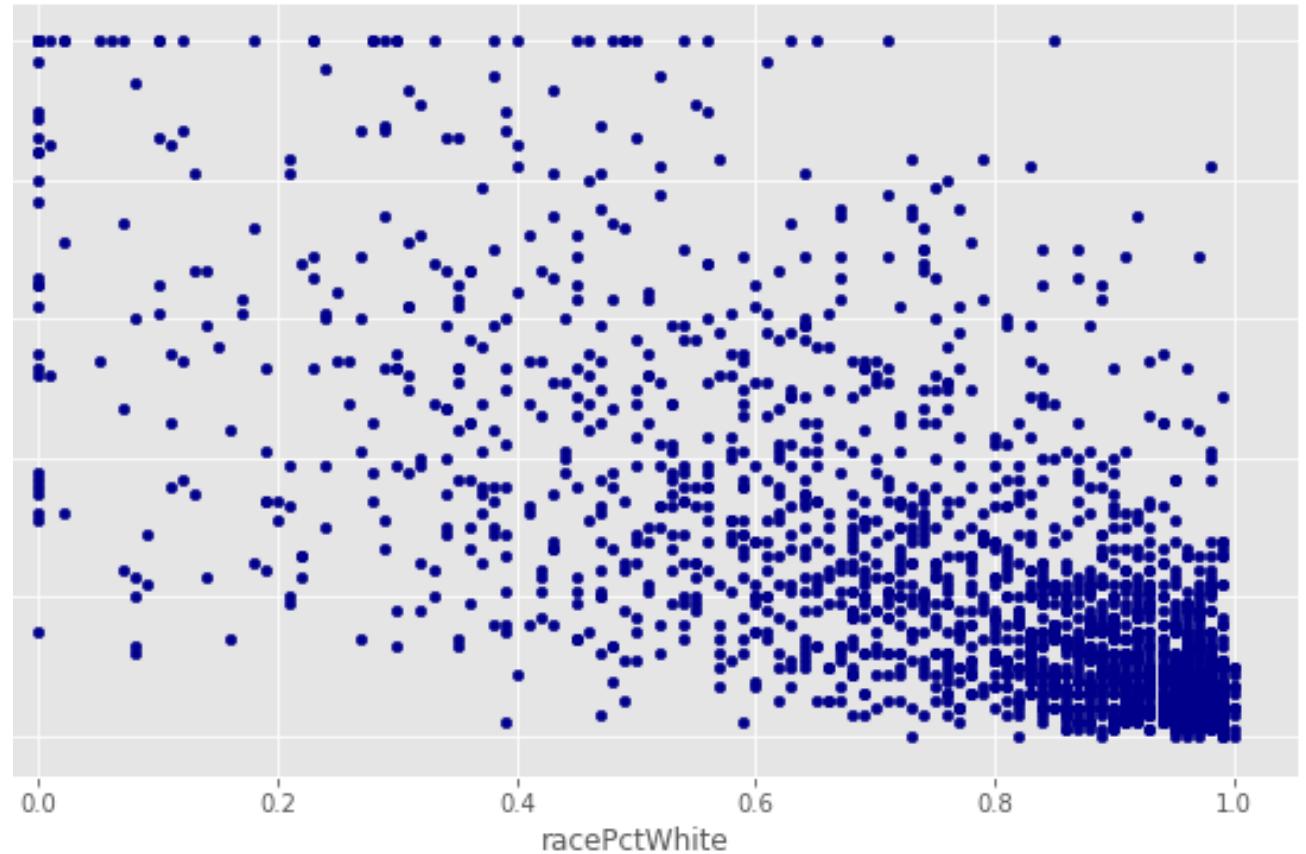
Percentage of population that is
African American has positive
contribution to crime rate

Percentage of
population that is
African American



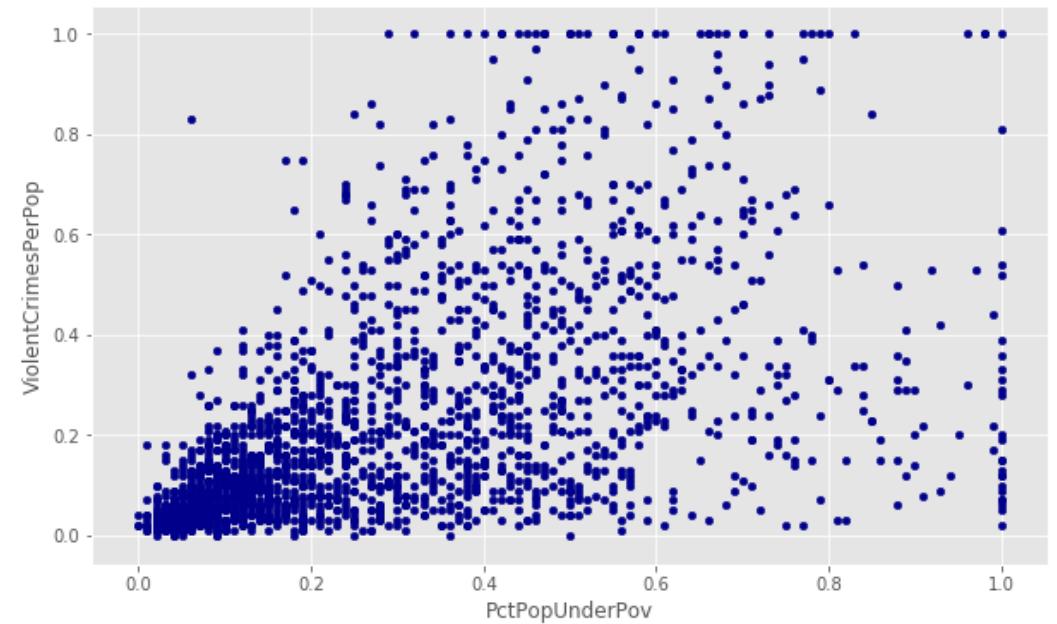
Percentage of population that is Caucasian

- Percentage of population that is Caucasian has negative impact on crime rate.



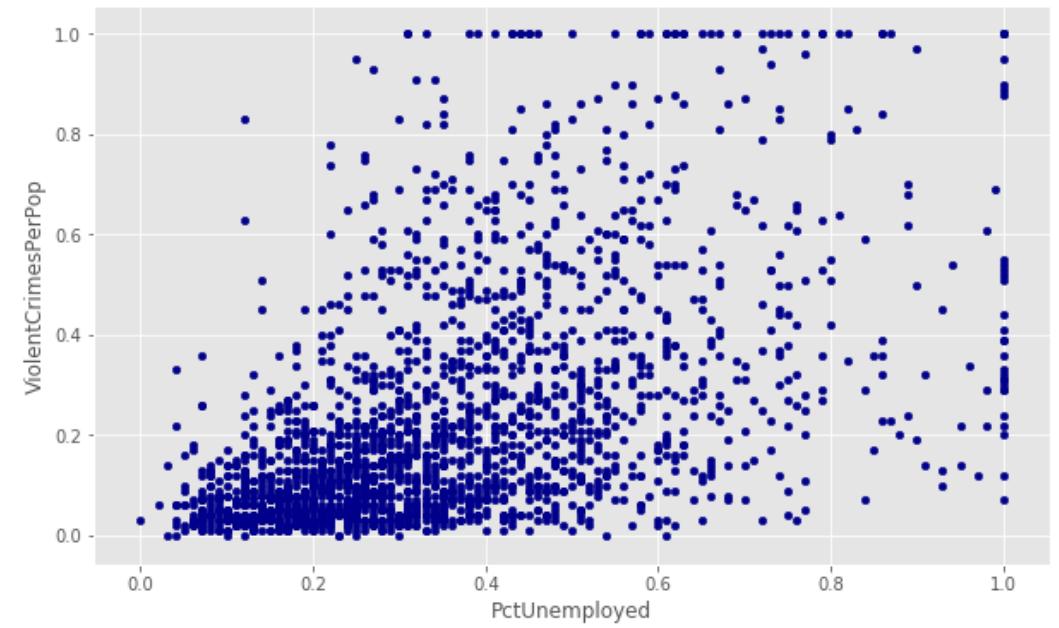
Increase people's income will help to lower crime rate

Percentage of people under the poverty level

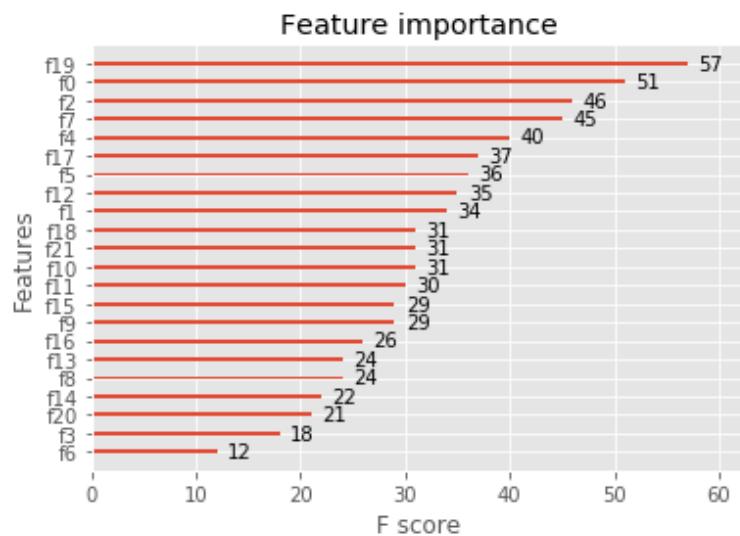


Increase job position and lower unemployed rate will help on decrease crime rate

Percentage of people 16 and over, in the labor force, and unemployed



Regression models performance



- Regression model runs on 108 combination and pick up as final combination
 - Best model can achieve 0.51 for R2-score

Conclusion

- Predict model itself is not good enough
- But the predict importance provide insights of factor impact on crime rate and it matches what we examined during data analysis step