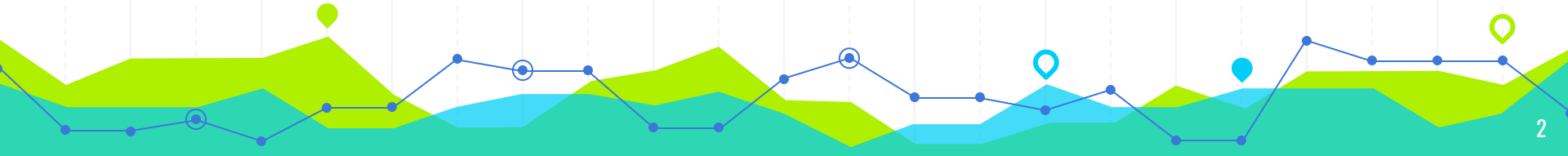
Denver Crime Prediction

Andres Ramirez
Flatiron School - Full Time Data Science Program

Business Problem Statement

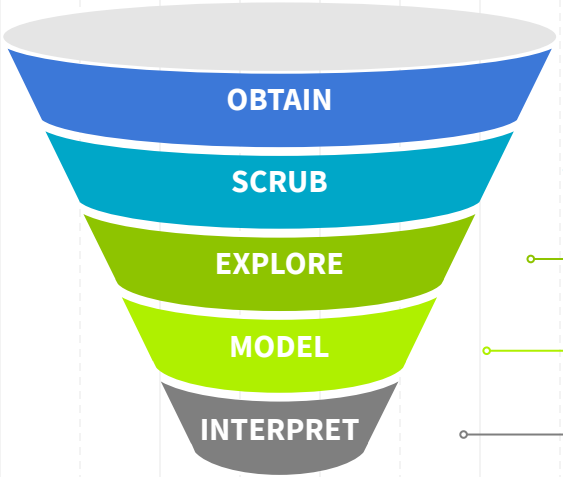
Crime is a challenging social problem across major cities and metropolitan areas across the United States. When we look at the statistics leading up to 2021 the results are astounding. In Denver, business burglaries are up by 143 percent, carjackings are up by 140 percent and homicides are up 81 percent since 2019. As crime increases, it is important to try to find novel ways to implement crime prevention, and this project's goal is to use that through prediction, focused on neighborhoods and time of day.



Business Problem Proposed Solution

- Build a model that can predict the most common crimes occurring in Denver's neighborhoods
- Correlate demographic features to particular crime categories such as theft or assault
- Accomplished using Denver's Open Data Catalog

Data Science OSEMN Model



Multiple Datasets obtained from the Denver Open Data Catalog

Prepared our data to address our project objective, for example removing traffic accidents

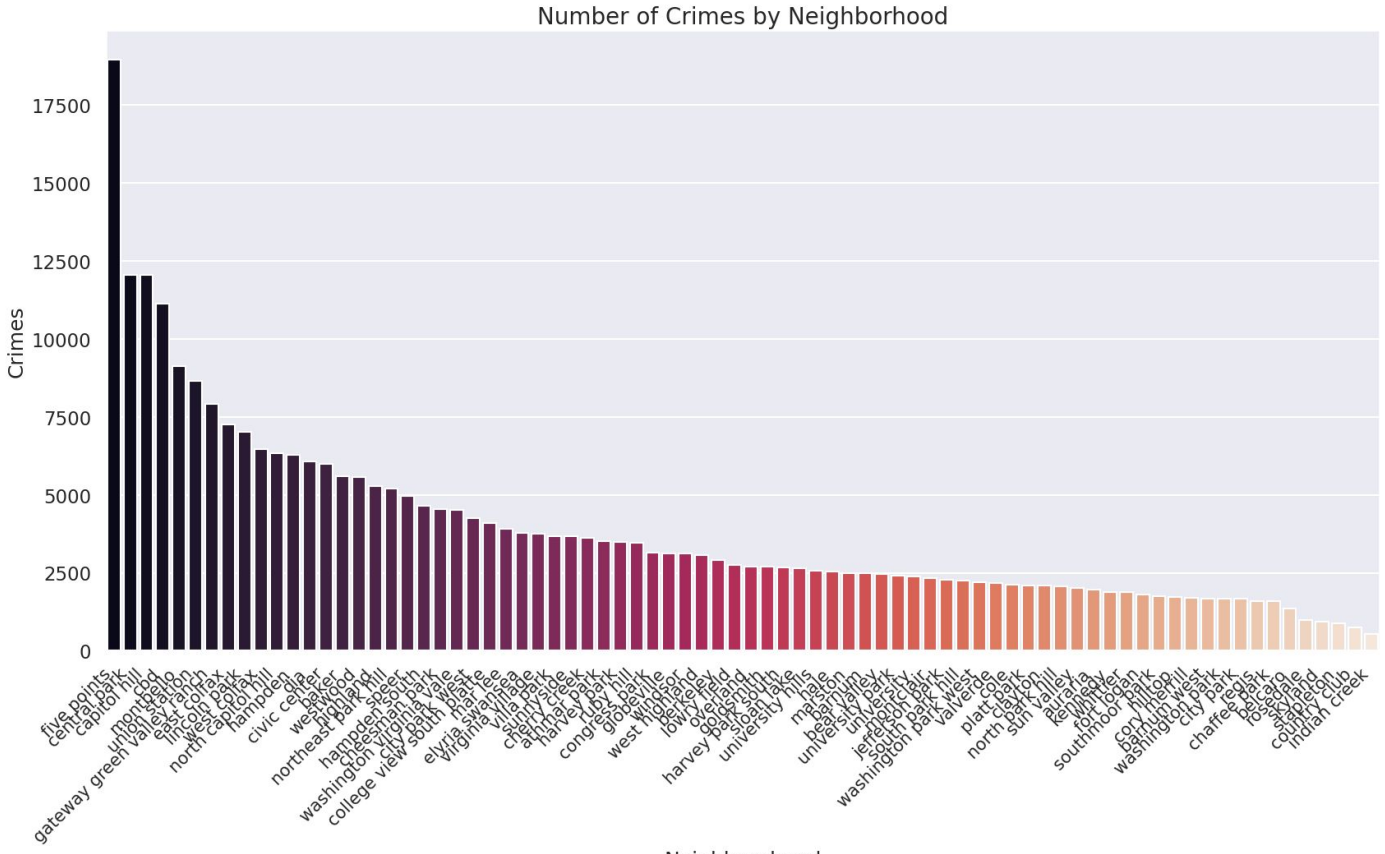
Building visualizations that help outline the basics of our data such as the crime most committed and neighborhoods most affected.

Utilize supervised learning models to be able to predict crime categories in Denver's neighborhoods

Interpret the results in a useful manner that will hopefully increase safety in Denver's neighborhoods.

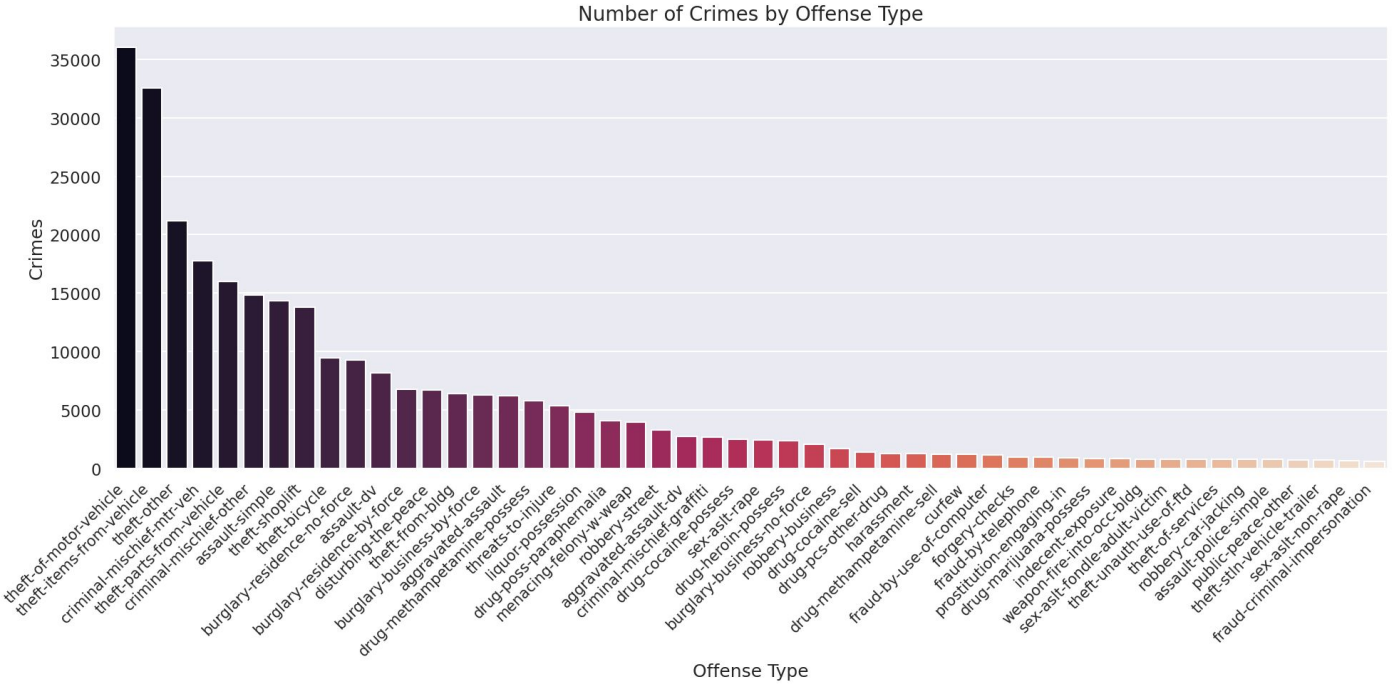
Denver's Neighborhood

This graph outlines the number of total crimes committed from 2016 to 2021 in Denver's 78 neighborhoods. We can see that Five Points, the downtown area of Denver, leads the other neighborhoods by a large margin.



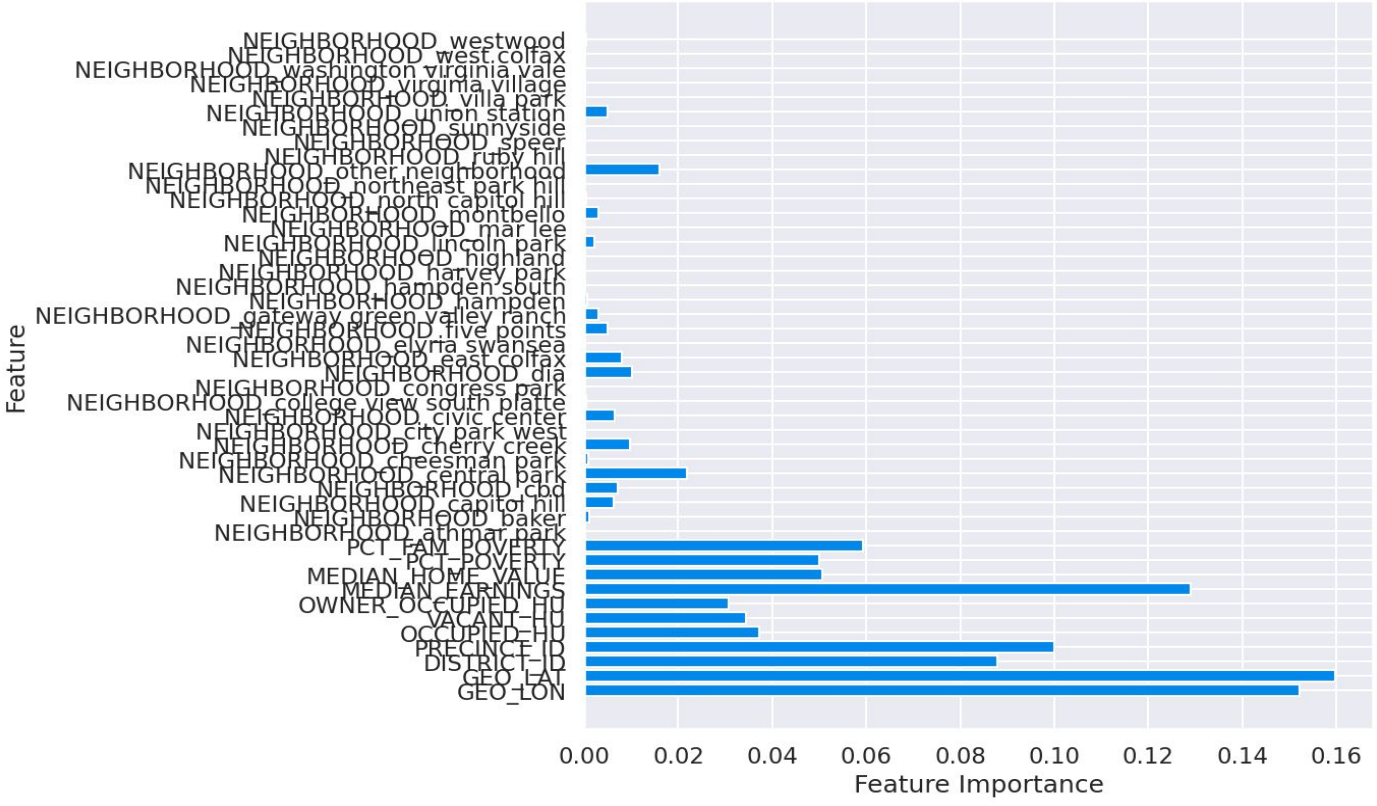
Types of Offenses

From this graph of the types of offenses committed in Denver that burglary and theft are the most frequently committed crimes. The most common violent crime committed is simple assault.



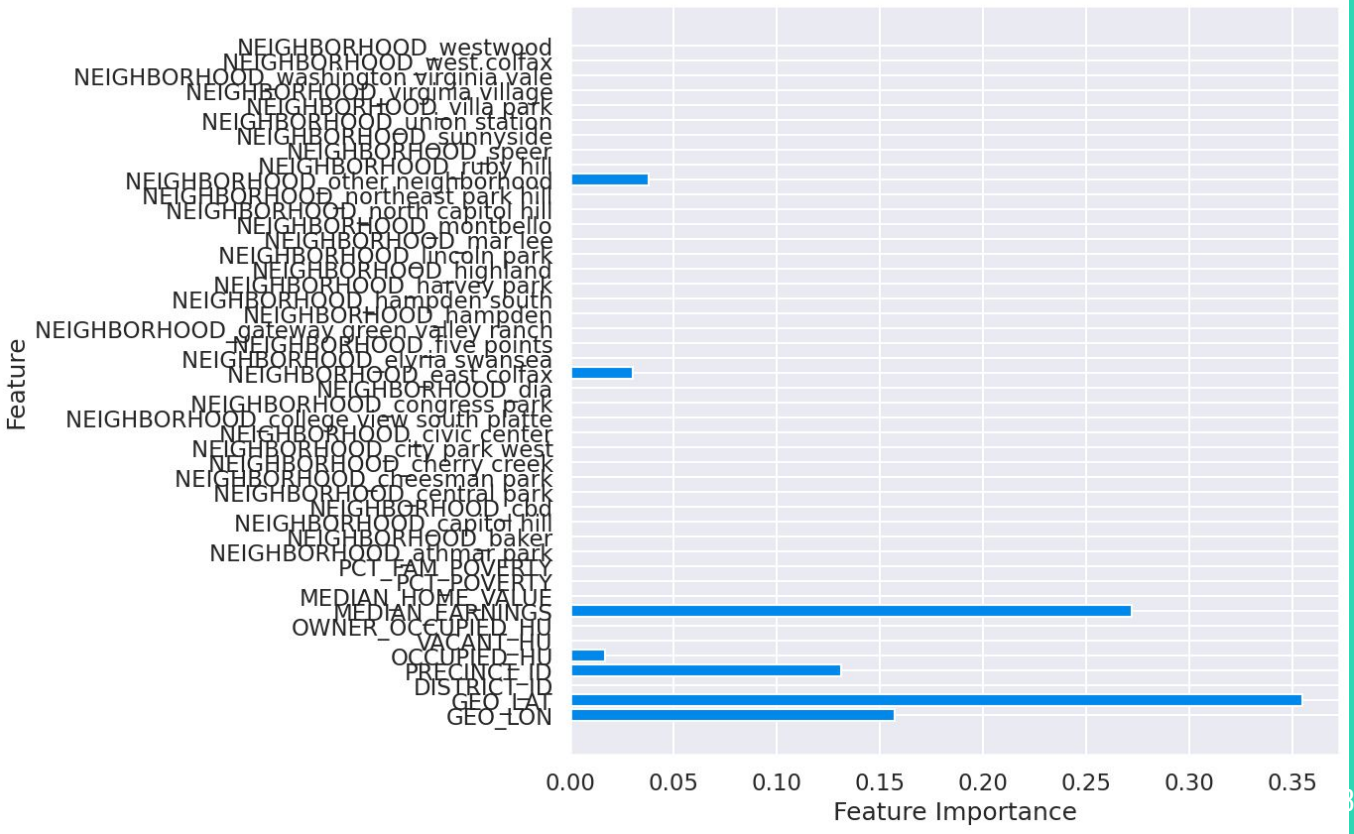
Random Forest Feature Importance

From this graph of the different features we were able to look at we can see that latitude and longitude play an important role in crime prediction. Interestingly, median earnings also plays a large role in our model



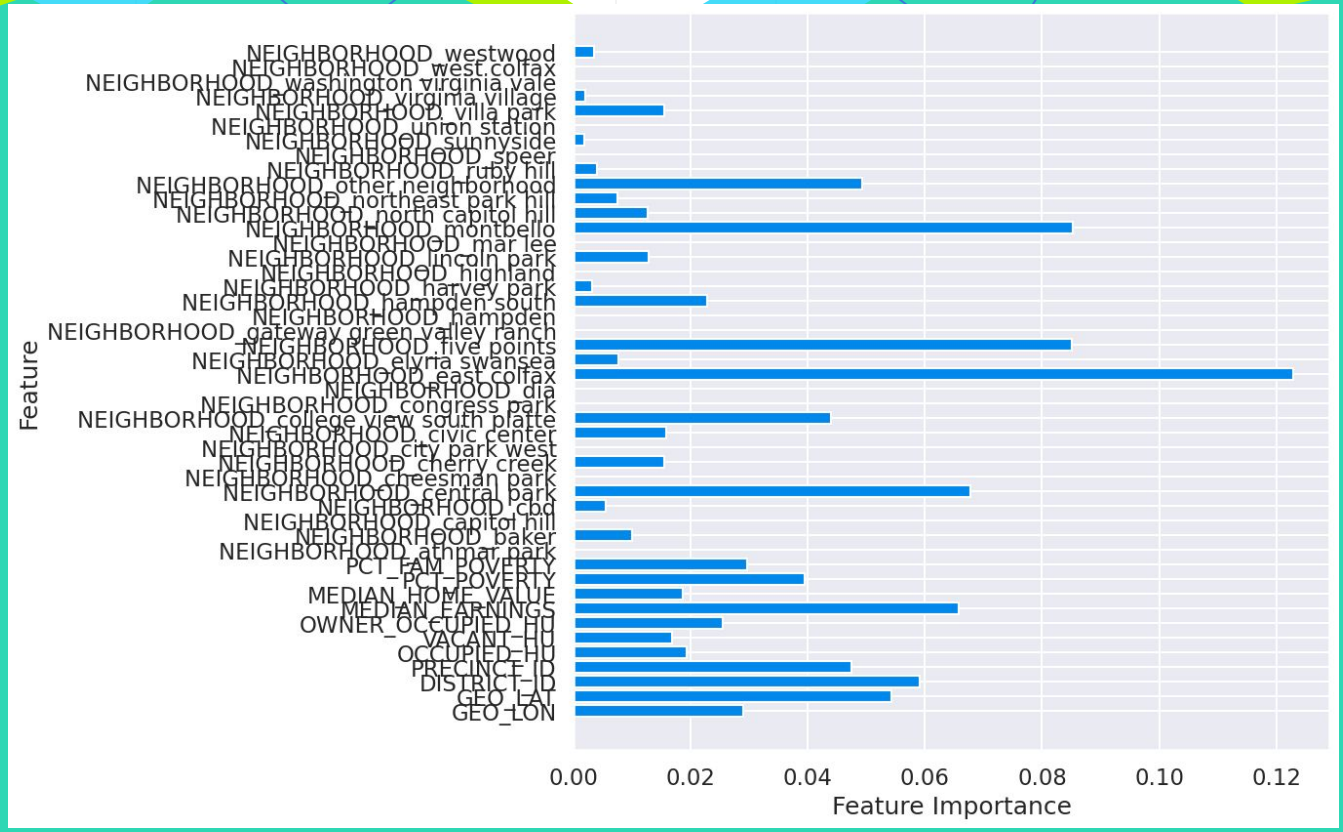
Decision Trees Feature Importance

From this graph of the different features we were able to look at we can see, similar to our Random Forest model, that latitude and longitude play an important role in crime prediction. Similar to our previous model, median earnings also plays a large role in our model.



XGBoost Feature Importance

From this graph of the different features we were able to look at in our XGBoost model, we can see that latitude and longitude do not play as an important role in crime prediction compared to the other models. In this instance, the neighborhoods East Colfax, Montbello, and Five Points play a role, which is understandable as they have some of the highest levels of crime.



Recommendations

Although our models did not produce as robust as results as we would have liked, 32% (Decision Trees), 27% (XGBoost), and 32% (Random Forest) there are some recommendations we can make.

- Median earnings was an important feature, and this can in both directions. More crime prevention and community services should be offered in low median earning locations and more patrolling should be done in high median earning areas as much of the crime in these areas involve larceny.

- Geolocation plays a big role, and crime tends to occur in an area where criminals are familiar with the location. This should lead to greater patrolling in areas where there is geographically more crime

- As highlighted by our exploratory analysis and our XGBoost model, neighborhood plays an important role in identifying and predicting where crime might happen. Extra patrolling and community outreach services should be implemented in the areas of Five Points, Montbello, and East Colfax. This can also include community education in term of how to keep oneself, our neighbors and communities safe. Also using apps like Nextdoor can help community members stay informed of what is going on in their neighborhood.

Future Work

In an extension to the work that has been completed, it is planned to add more classification models to hopefully increase the crime prediction accuracy and hopefully enhance overall performance. It is also the hope to include previously excluded features from the demographic portion of our dataset. We have the option to add gender, education level, earnings levels and race into our model. It will be interesting to see if any of these demographics make a sizable impact on the performance of our model.

Also, it will be helpful in the future to continue including more data from the Denver crime dataset to see if there are more noticeable trends in the future. This framework can also hopefully be extended to other major metropolitan cities such as Chicago and Los Angeles to see if there may be other pivotal features that can increase model performance. The goal is to hopefully increase the audience to the work of using Machine Learning programs to help inform local law enforcement and community leaders to keep neighborhoods safe.

THANKS!

Any questions?

You can find me at:

github.com/6cats1dog

medium.com/@aj.ramirez23

aj.ramirez23@gmail.com