Crime Data in Los Angeles

By Jai Dave

Report as part of The Springboard Data Science Career Track

Target Audience

POLICE OFFICERS CAN USE THIS DATA TO BE MORE EFFICIENT ON PATROLS

PUBLIC POLICY MAKERS CAN USE THIS DATA TO CREATE LEGISLATION THAT WILL HELP ALLEVIATE CRIME

POLICE CAPTAINS CAN USE THIS DATA TO MAKE DEPARTMENTS MORE EFFECTIVE

Problems and Solutions

My analysis will consist of three main topics: WHERE crimes are occurring, WHEN crimes are occurring, and WHO are most victimized by crimes

Determining where crimes are occurring to aid communities that need it the most

Determining when crimes are occurring to make serving communities more efficient

Determining who are the victims of crime to help populations who are the most vulnerable

Executive Summary

- The area with the highest crime rate is 77th St.
  - Assault with a deadly weapon makes up a larger proportion of the area's crimes with firearms the primary weapon used
  - Creating and promoting campaigns like the anonymous gun buyback in this area may help curb assaults with use of a firearm

- The time with the highest crime rate is Friday Afternoon
  - Friday afternoon has a higher number of thefts than the sample as a whole
  - This could be because individuals have debts due at the end of the week and they turn to crime out of desperation
  - Creating policies to improve income distribution could reduce these types of crime

- The Latino, White, and Black populations make up the most victims
  - The distribution of age and victims appear to follow the demography of Los Angeles with the exception of Black Males who appear to be under represented
  - Black males may be reluctant to report crimes to authorities
  - Promoting community outreach programs may help increase trust with communities

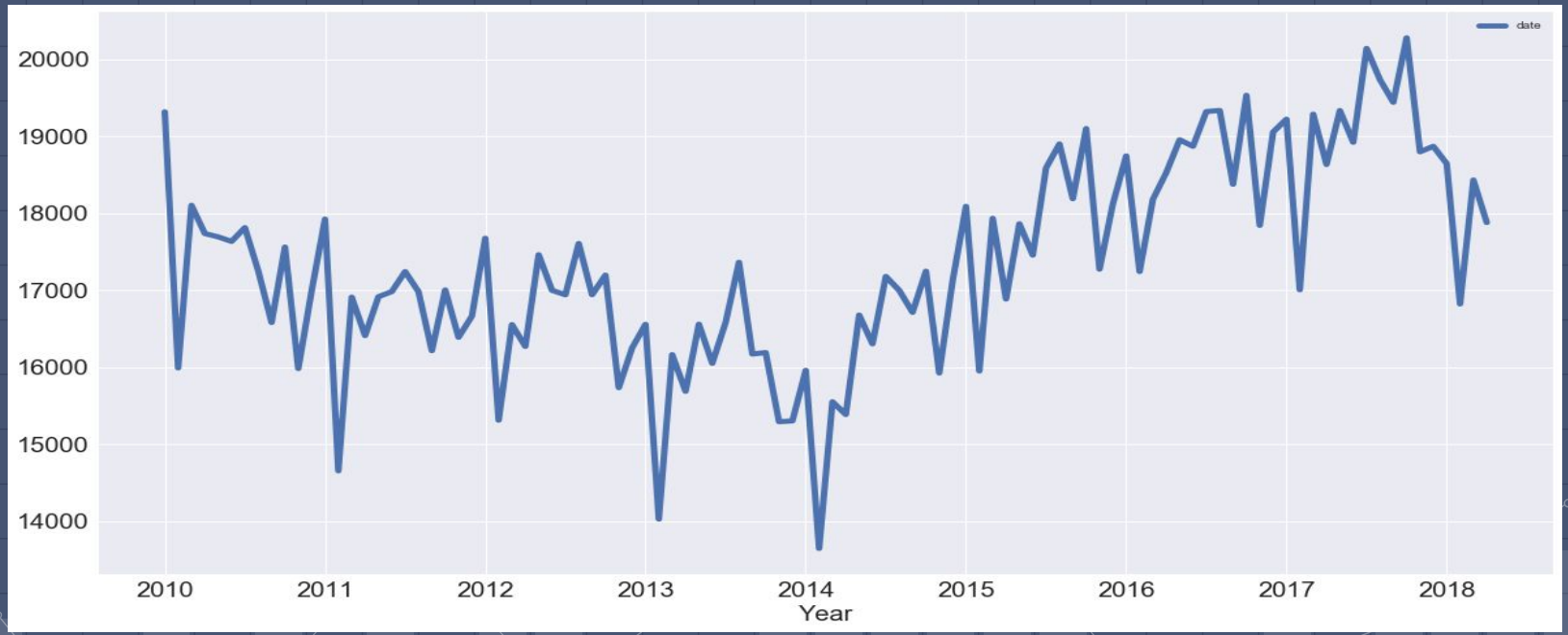
Methodology and Objectives

- Purpose: Find trends in the data that can be acted upon to reduce crime
- Methodology: Explore the data visually, test if differences are significant, and create models to predict where a crime will occur
- Tools: Use Python to get data from the LA city's website and conduct analysis
- Dataset: The dataset reflects incidents of crime dating back from 2010 provided from the Los Angeles Police Department, where each row is a crime incident

233,090

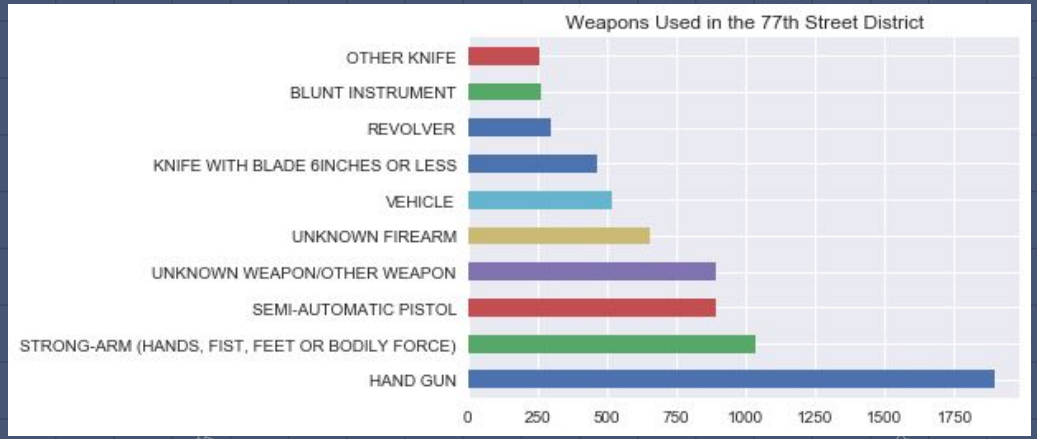
The Number of Crimes Reported in Los Angeles during the year 2017

Crime Over Time

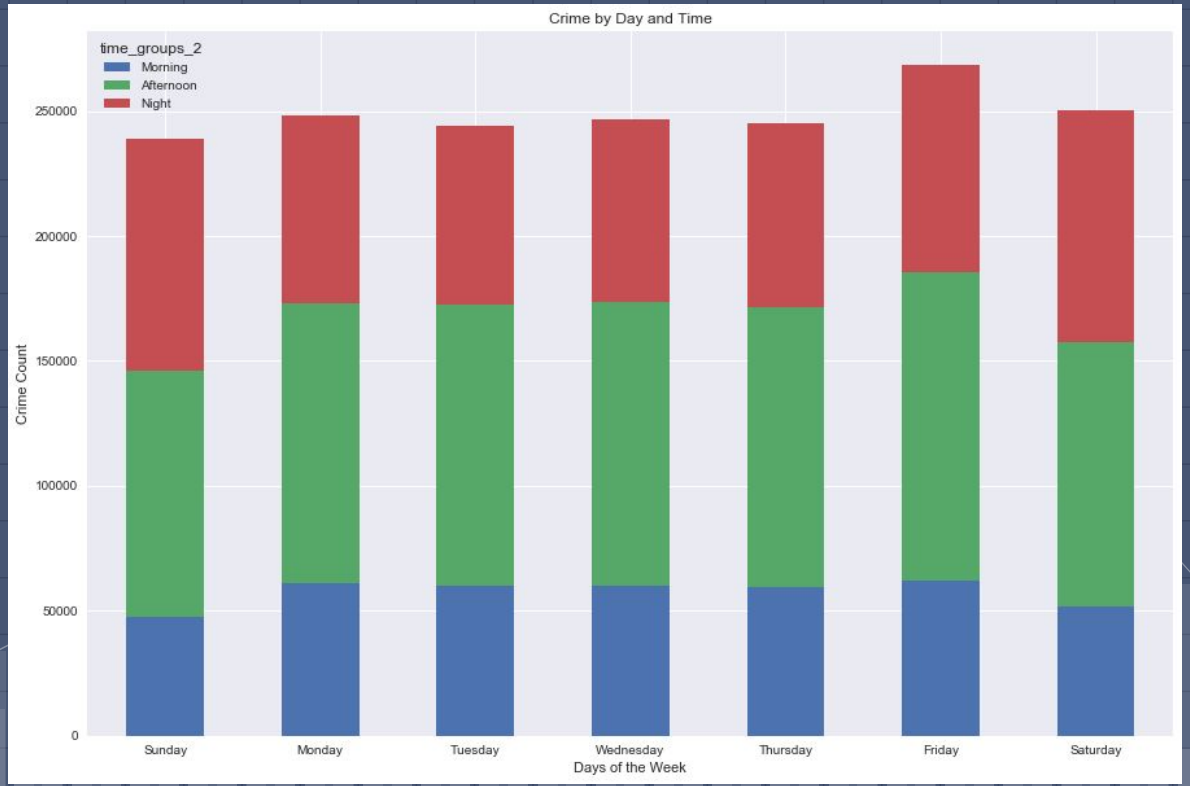


Where are Crimes Occuring?

- My analysis found that the 77th St. Area had the highest crime rate
- Approximate Population: 175,000
- Area: 11.9 Square Miles
- 77th St. has a significantly higher number of “Assaults with a Deadly Weapon” than the sample as a whole
- What weapons are being used for assaults?

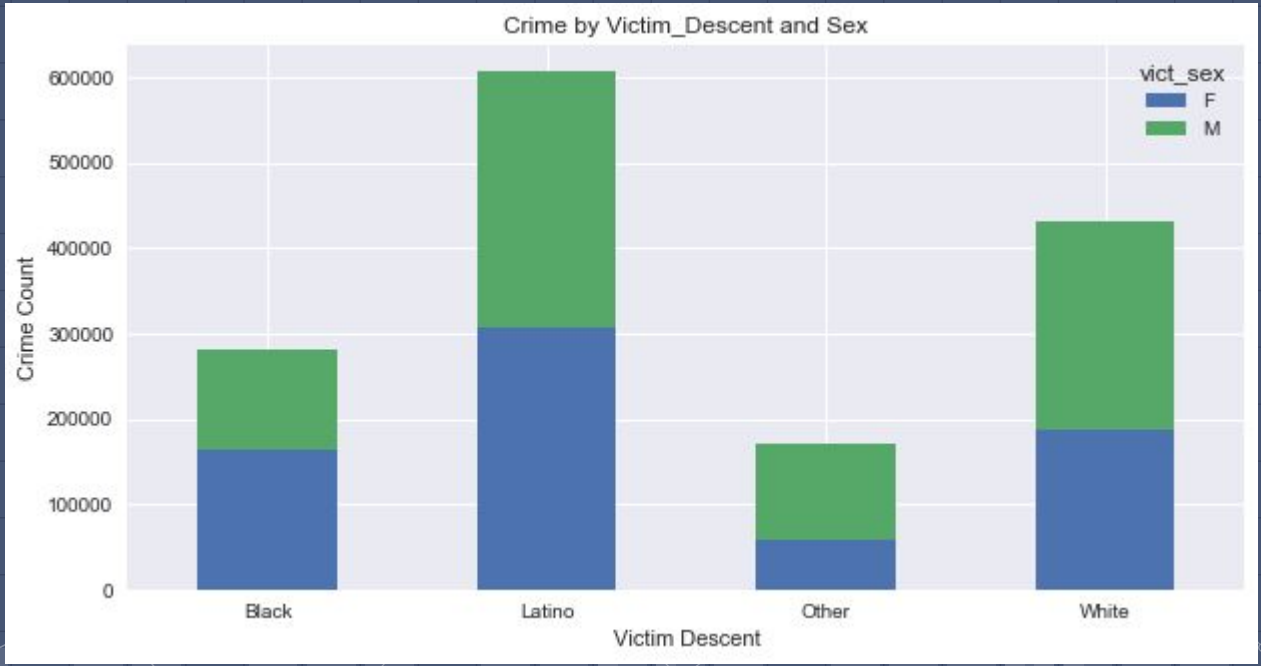


When are Crimes Occurring?



- Friday afternoons show a spike in crime rate
- On closer inspection Thefts appear to be come common during this time
- Perhaps as a result of people with debts owed at the end of the week or more money in circulation at the end of the week

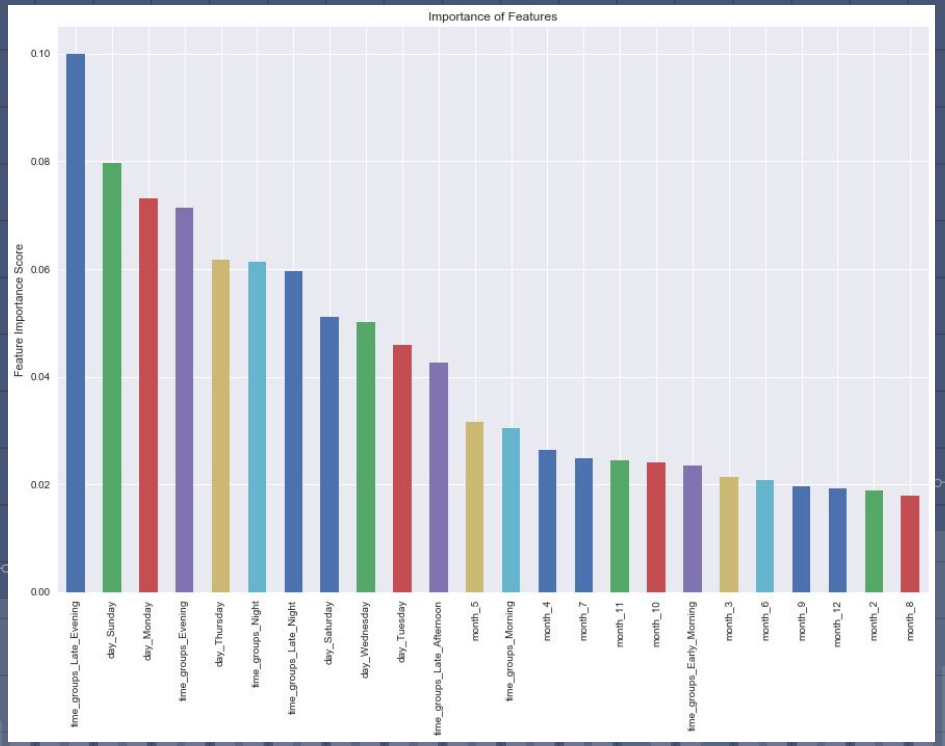
Who are the Victims of Crimes?



- Victims are more likely to be Hispanic, White, or Black
- Among black victims females outnumber males
- This could possibly be because Black Victims are less likely to report crimes

Predicting Where Crimes will Occur

- Finally I used various modelling techniques to predict the zipcode a crime will occur based on the day of the week, the time of day, and the month a crime occurred.
- The most successful model a Gradient Boosting Machine Model which only had an accuracy score of 37.42%
- This model was created using a subset of the data which contained all assaults by handgun in the 77th Street district. The importance of the features is shown to the right



Final Remarks and Next Steps

- Through this project I learned how to process data, explore it, perform various tests, and create predictive models
- I was able to identify where and when crimes most often occur and what subpopulations are the most victimized
- Ways to improve the model would be to further optimize the hyperparameters and try other modelling techniques, such as XGBoost
- Possible next steps to take would be to examine which subpopulations are more susceptible to particular types of crime, see what leads to certain areas having higher crime rates, and to further examine crime targeting the homeless

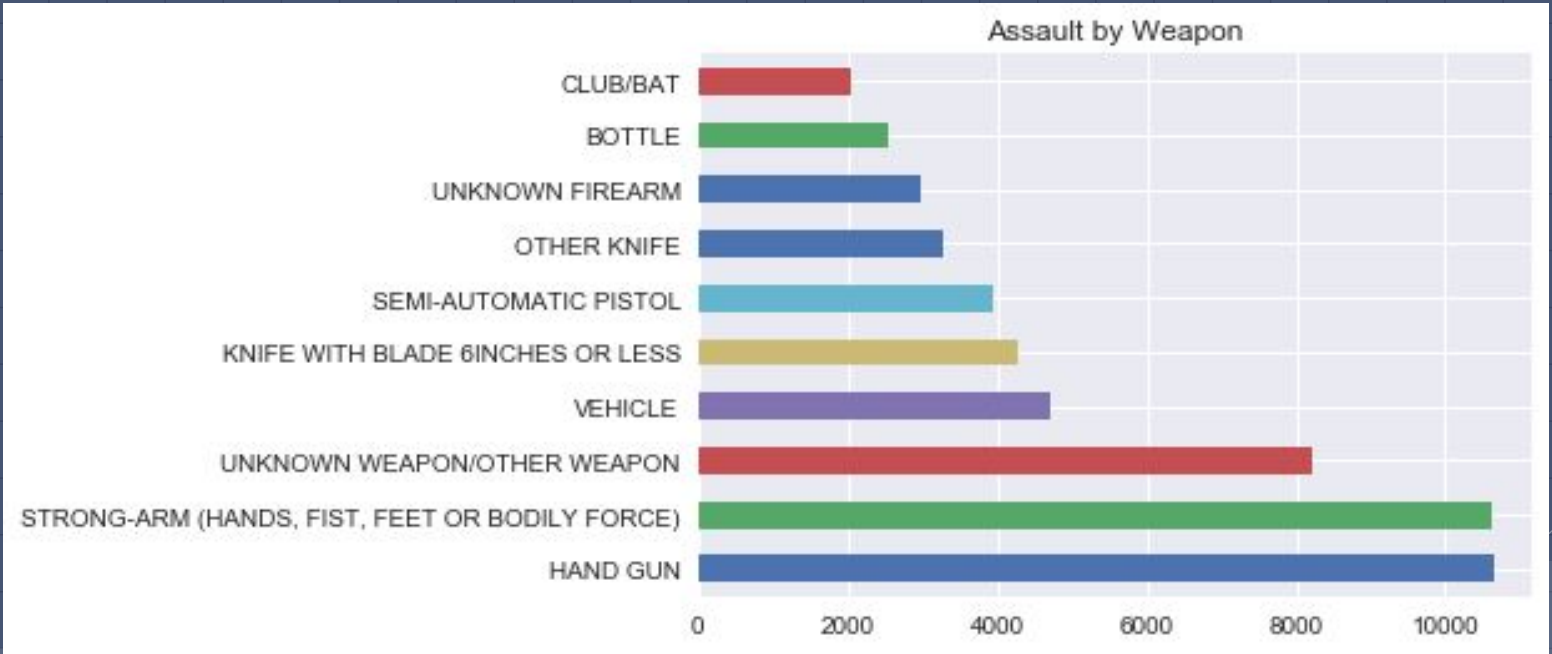
THANKS!

Any questions?

You can find me at

- jbdave112@gmail.com

Appendix A: Weapons used for Assault



Appendix B: Distribution of the Age of Victims