# Austin Crime Reports Analysis

Sayesha Aravapalli | Rawini Dias | Ming-Hsiu (Matthew) Peng | Prathik Ullur

## Description of Project Goals

### Description

The dataset that is being explored is the Austin Crime Reports dataset from the Austin Open Data repository. This dataset contains a record of incidents that the Austin Police Department (APD) responded to and wrote a report on. From the master dataset, incident reports from 2010 through 2018 were used for the analysis.

### Importance of the problem

Despite an overall decreasing trend in crime rate through much of the last decade, there was a drastic crime rate increase in 2018 (chart 1).

The goal of this project is to investigate the reasons for this increasing trend in 2018 and to gather actionable insights. As part of the analysis, the team will also look at APD performance by sector to investigate, if any, inefficiencies by sector that may be contributing to overall trend in crime.

The overall objective of this analysis is to help the Austin Police Department make Austin safer.

## Exploratory Analysis

Overall Trends of Crime in Austin

Crime rate of various crime types has fluctuated over the last decade. For 2018, we see a general trend of increasing non-violent crimes (i.e. theft, auto theft, robbery, burglary) and decreasing violent crimes (i.e. aggravated assault, rape). The number of murders increased in 2018 (chart 2).

Further analysis of crime count by type of crime showed that non-violent crime occurs in the highest counts, with the number of theft incidents being five orders of magnitude greater than the next highest crime type count, burglary. From a violent crime standpoint, aggravated assault occurs in the highest numbers.

## Crime Patterns by Time

The crime pattern of 2018 was analyzed based on day, week, and month.

A day was divided into morning(6 am-12pm), afternoon(12pm-10pm), and night (10 pm-6 am). The crime counts for each crime type told us that on average the probability of theft is highest during the afternoon and the probability of violent crimes such as aggravated assault, rape, or murder is highest at night (chart 3).

Furthermore, while investigating patterns by day of the week, we came across some interesting insights. We discovered that crimes like rape and aggravated assault had higher chances of happening on weekends. The total number of aggravated assault crimes on a Wednesday in 2018 was 85 compared to a Sunday, which was roughly 200. This is a nearly 135% increase in aggravated assaults on a weekend day compared to a weekday (chart 4). A similar pattern was seen for rape. It shows a 50% increase in rape cases for the weekend (chart 5). Our hypothesis for this is that people go out and drink more on weekends, which increases the chances of such crimes happening. We conclude that weekends are more dangerous with respect to violent crimes like rape and aggravated assault.

While investigating patterns in the 30 days of a month, we noticed a striking abnormality; the first day of the month had an unusual spike in crime count at an overall level (chart 6). We studied the data to understand if this was due to non-violent crimes. But we observed that both the total non-violent crimes and total violent crimes spiked on the first day of the month (chart 7 and chart 8).


## Future Crime Rate Prediction

The discovery of the above crime patterns leads us to the question of how likely the patterns are to continue in the future. We decided to apply machine learning skills to see if we can use predictive models to catch the patterns and get an idea of how the crime rate will evolve in the next half of 2019.

Data: Our goal is to predict the daily crime rate in the second half-year of 2019 with data collected from January 2010 to July 2019 (see table 1). The features are all derived from dates, including 'day of all data', 'year', 'month', 'day of the month' and 'week day' (see table 2). We use 2010-2018 data as the training set and the first half-year of 2019 as the testing set.

Model: Considering (a) predicting daily crime count is a regression problem and (b) our features are actually categorical variables, the Random Forest

model became our top choice because it can be applied to solve regression and it is very good at processing data with many categorical variables.

Model Performance: RMSE of training data is 12.95; RMSE of testing data is 15.96. Considering the mean of daily crime counts is 113, the test RMSE of 15.96 seems to be acceptable.

Implication: We also used the model to predict the crime counts for the second half-year of 2019 (see chart 9 and chart 10). The result shows repeated peaks on the first day of a month, which is consistent with our exploratory data analysis. Then at the end of the year, crime appears to trend slightly downward, suggesting fewer crime activities towards the end of the year.

Performance by APD Sector

Austin is divided into 10 APD sectors. From an overall crime standpoint, David and Adam had the highest counts of crime in 2018 (chart 11). However, once we normalized the count of crimes by APD sector based on APD sector population, George has the highest incidents of violent and non-violent crime per 100 people (chart 12 and chart 13). This makes sense as George is the downtown Austin APD sector. The population living in downtown is less compared to other sectors, however more people frequent downtown Austin and therefore there are more occurrences of crime.

From a performance standpoint, we would like to measure how well and how efficiently police officers handle reported cases in each APD sector. We will use Clearance Rate (the number of cases solved divided by the number of cases reported) to assess how well APD sectors handle cases. Next, we will use Time Taken to Solve (the number of days between Occurrence Date and Clearance Date) to assess APD sector efficiency.

Clearance Rate: According to chart 14, Baker and Adam sectors have the lowest overall clearance rate, while sector Ida and Edward have the highest overall clearance rate. However, the clearance rate of violent crimes are fairly at a similar level, and the numbers of violent crime are relatively low. Therefore, the violent crime solving efficiency is similar across sectors. The numbers of non-violent crimes, on the other hand, are several times more than violent ones. Therefore, a percent difference actually means a huge difference in the actual number. Based on the analysis, we would suggest APD focus on improving the non-violent clearance rate by applying crime-solving practices utilized by the Ida sector which has the highest non-violent crime clearance rate.

Time Taken to Solve: In [chart 15](#), we can see that the time taken to solve cases for sector Ida is the lowest, which further supports our suggestion to apply Ida's practices to other sectors. Additionally, we can see the time taken to solve violent cases is significantly higher in sector Charlie and George compared to other sectors. Therefore, we would suggest investigating if inefficiencies in Charlie and George are due to limited resources or inefficient practices and resolve accordingly.

## Solutions and Insights

Based on our exploratory data analysis, we discovered an abundance of insightful information regarding crime trends in Austin. From our analysis, we see that although violent crimes have decreased, non-violent crimes have significantly been on the rise in 2018. This is probably because police officers have focused heavily on curbing violent crimes thereby not spending sufficient resources and/or effort on addressing non-violent crime. A good solution to this would be to allocate additional resources towards handling non-violent crimes.

Additionally, we see that most of the non-violent crimes occur during the day and theft occurs in the largest numbers.. Increasing awareness among the public to be more vigilant of theft during the day would be a step in the right direction. We see that most of the violent crimes occur during the night with a tendency to increase during the weekend nights. We believe this increase is due to the fact that more people tend to go out to bars and parties during the weekend, which increases the number of intoxicated people who may become victims of rape or aggravated assault.  A way of curbing this would be to dispatch extra police patrol to areas of active nightlife and surrounding suburbs.

Among all the sectors, we see that Ida has the highest efficiency in solving crime and Baker, Goerge, and Charlie sectors have the least efficiency. We should investigate what Ida is doing differently that is helping them solve crime faster. In Baker, George, and Charlie, we should investigate if the inefficiency is because of limited resources or police department practices and resolve the inefficiency accordingly.

Finally, based on our exploratory data analysis and prediction modeling, it becomes apparent that time of day, day of the week, and location are the most important predictors of crime occurrence in Austin.

## Appendix

Link to dataset:
https://data.austintexas.gov/Public-Safety/Crime-Reports/fdj4-gpfu

Table 1: Data Occurred Date Summary

| | Occurred Date |
|---|---|
| count | 3488 |
| unique | 3488 |
| top | 2015-10-18 00:00:00 |
| freq | 1 |
| first | 2010-01-01 00:00:00 |
| last | 2019-07-20 00:00:00 |

Table 2: Predicting Variables Summary

| | crime count | day of all data | week of the year | year | month | day of the month | week day |
|---|---|---|---|---|---|---|---|
| count | 3488.0 | 3488.0 | 3488.0 | 3488.0 | 3488.0 | 3488.0 | 3488.0 |
| mean | 113.0 | 1744.0 | 26.0 | 2014.0 | 6.0 | 16.0 | 4.0 |
| std | 21.0 | 1007.0 | 15.0 | 3.0 | 3.0 | 9.0 | 2.0 |
| min | 15.0 | 0.0 | 1.0 | 2010.0 | 1.0 | 1.0 | 1.0 |
| 25% | 99.0 | 872.0 | 13.0 | 2012.0 | 3.0 | 8.0 | 2.0 |
| 50% | 111.0 | 1744.0 | 25.0 | 2014.0 | 6.0 | 16.0 | 4.0 |
| 75% | 126.0 | 2615.0 | 39.0 | 2017.0 | 9.0 | 23.0 | 6.0 |
| max | 213.0 | 3487.0 | 53.0 | 2019.0 | 12.0 | 31.0 | 7.0 |

## Chart 1: Crime Rate Trend In Austin from 2010 to 2018



Rolling sum of all crimes from 2010 - 2018
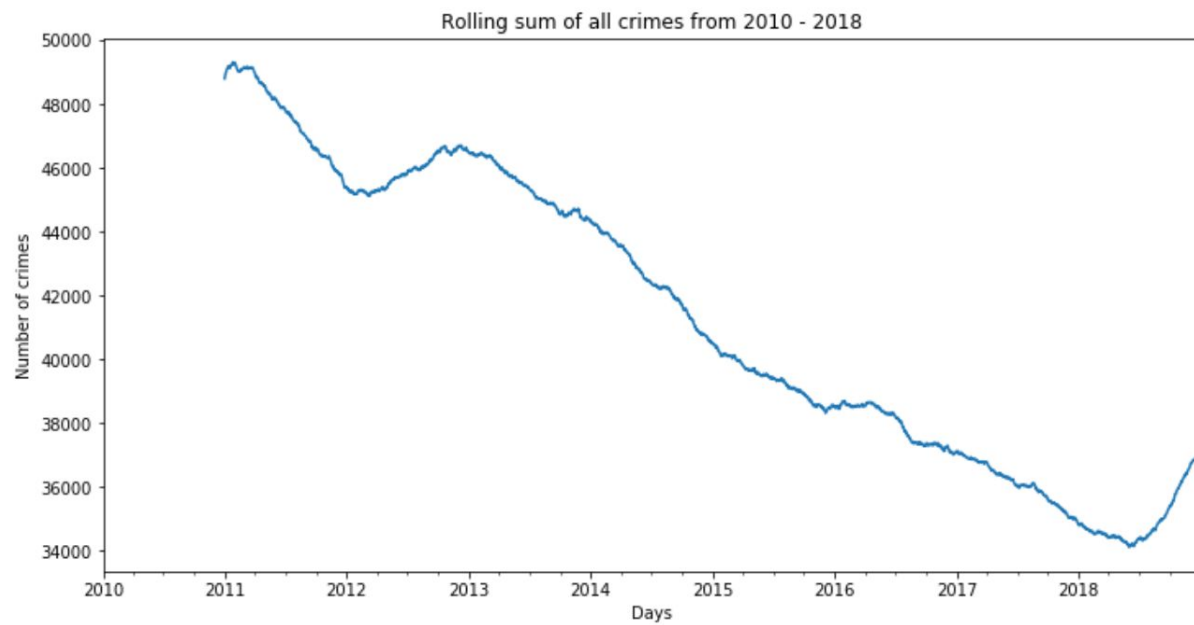
## Chart 2: Time Series of Crime Rate by Offense Type
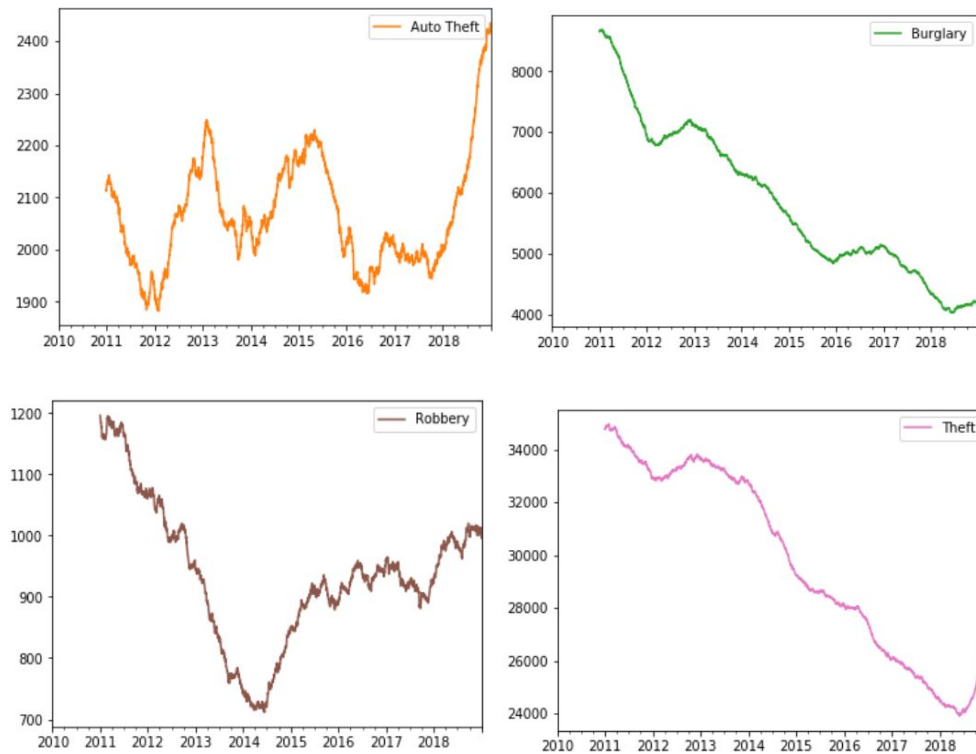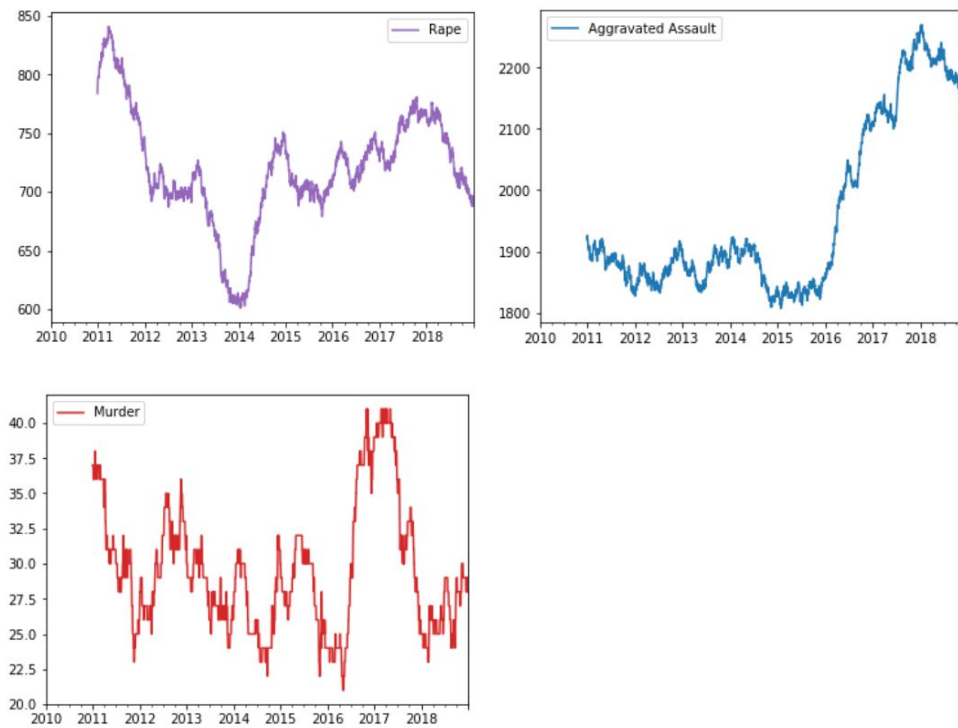
Chart 3: Rate of Crime on an average day
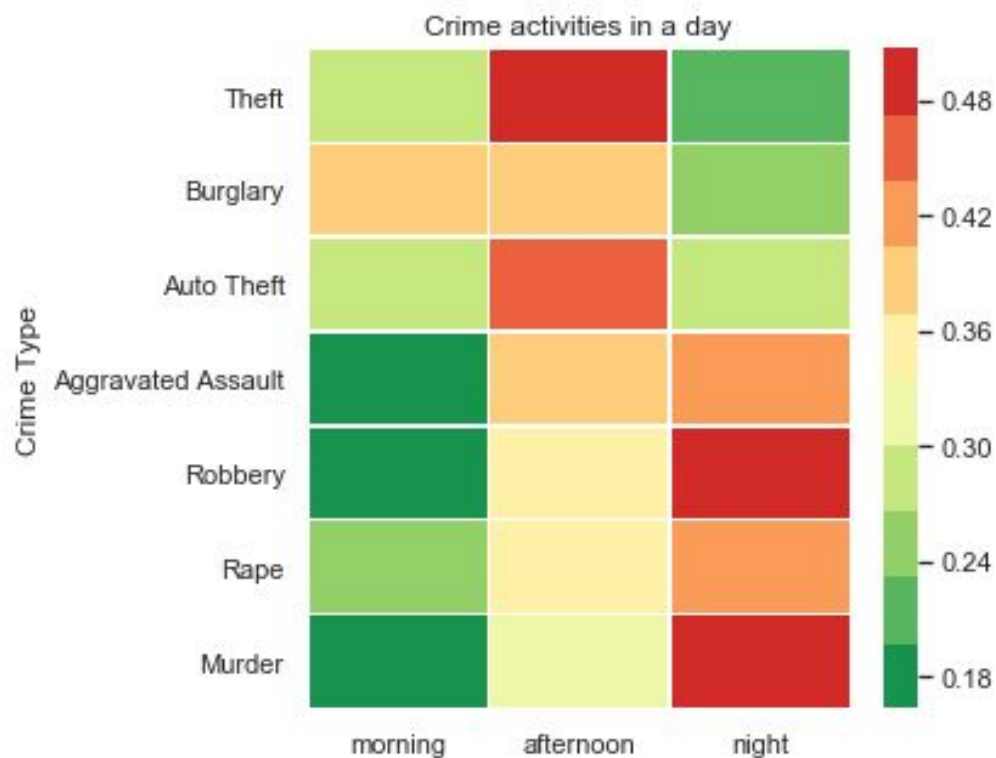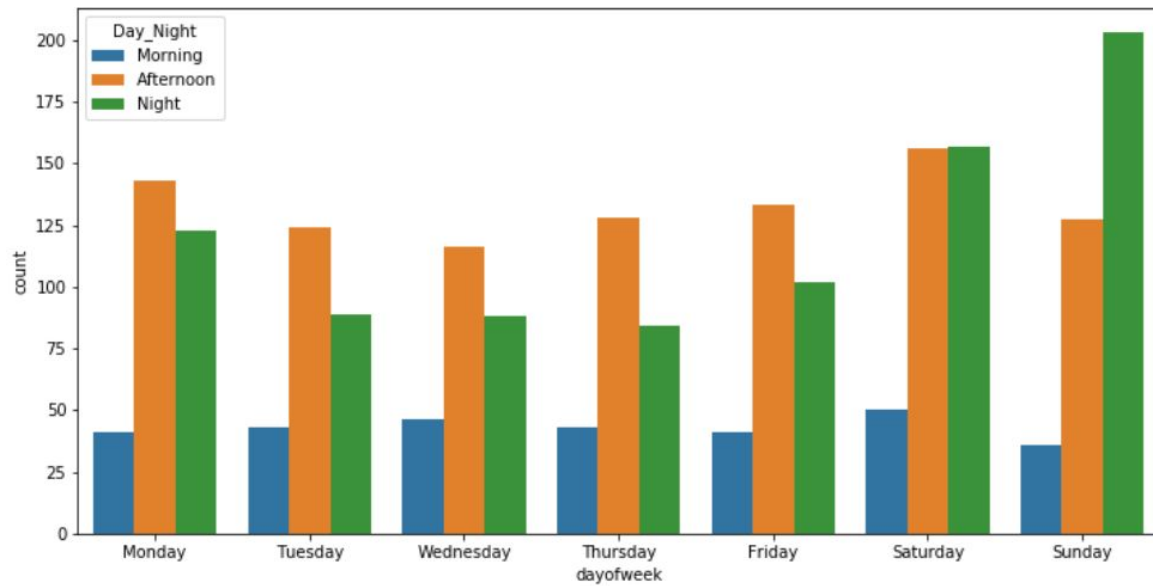
## Chart 4: Aggravated Assault Pattern by Day of Week
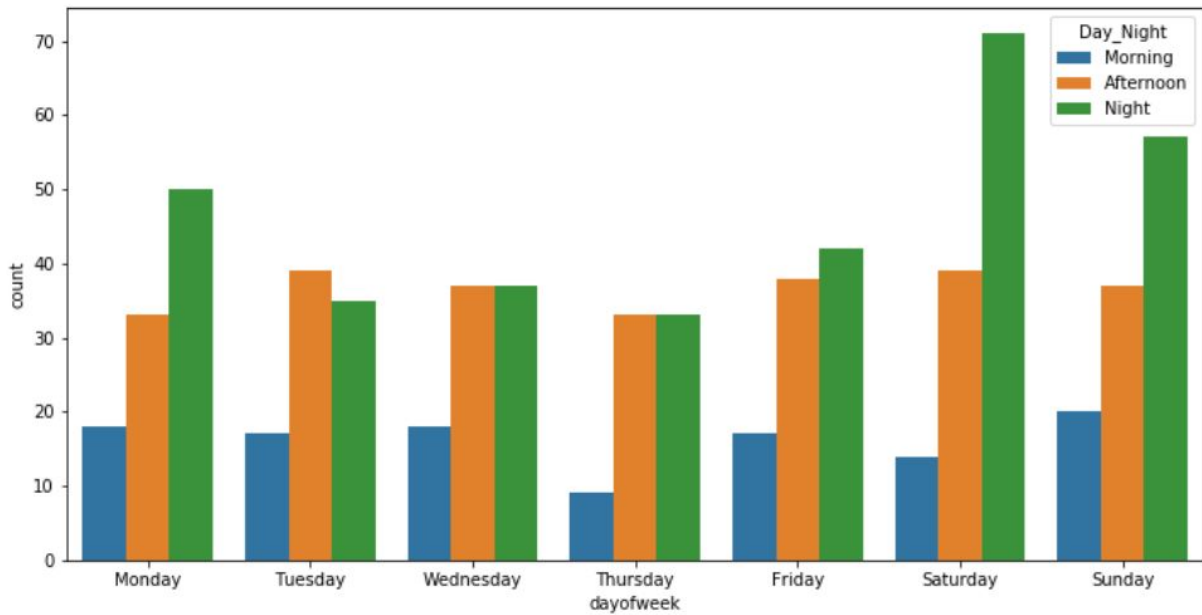


## Chart 5: Rape Pattern by Day of Week
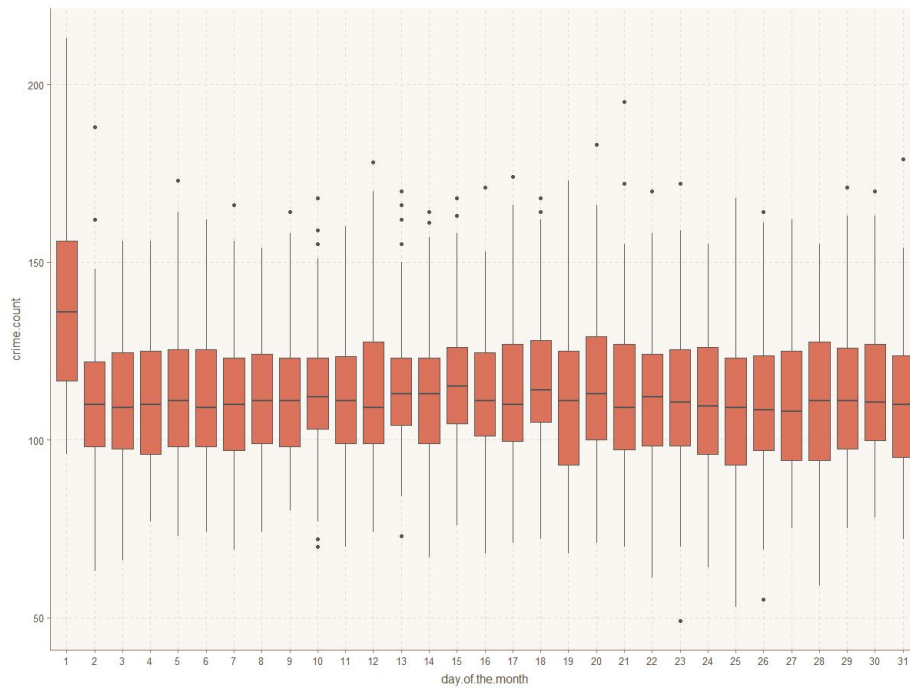
## Chart 6: Crime Distribution by Day of Month



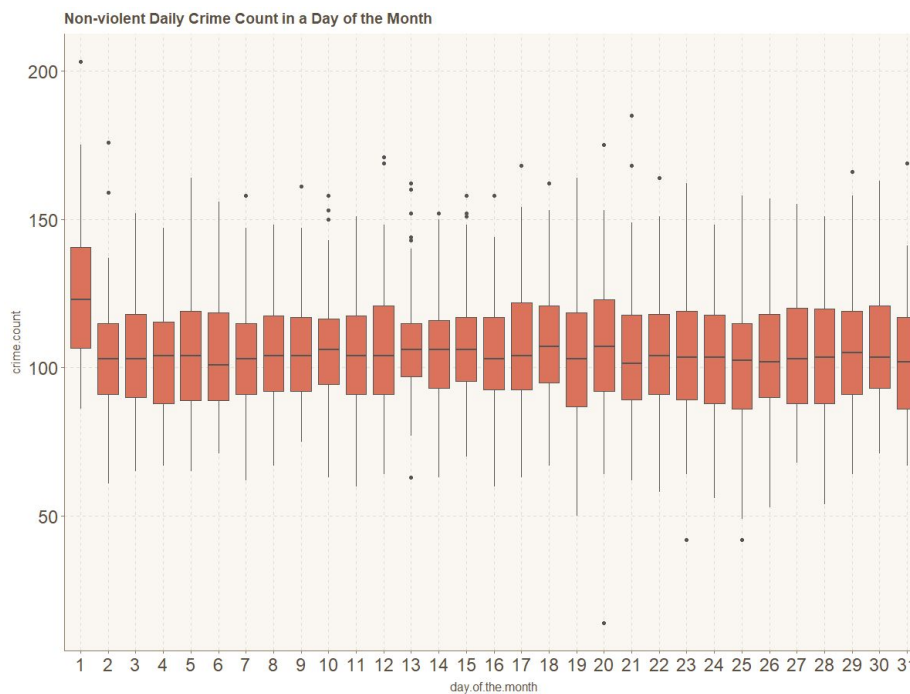## Chart 7: Crime Distribution by Day of Month - non-violent crime
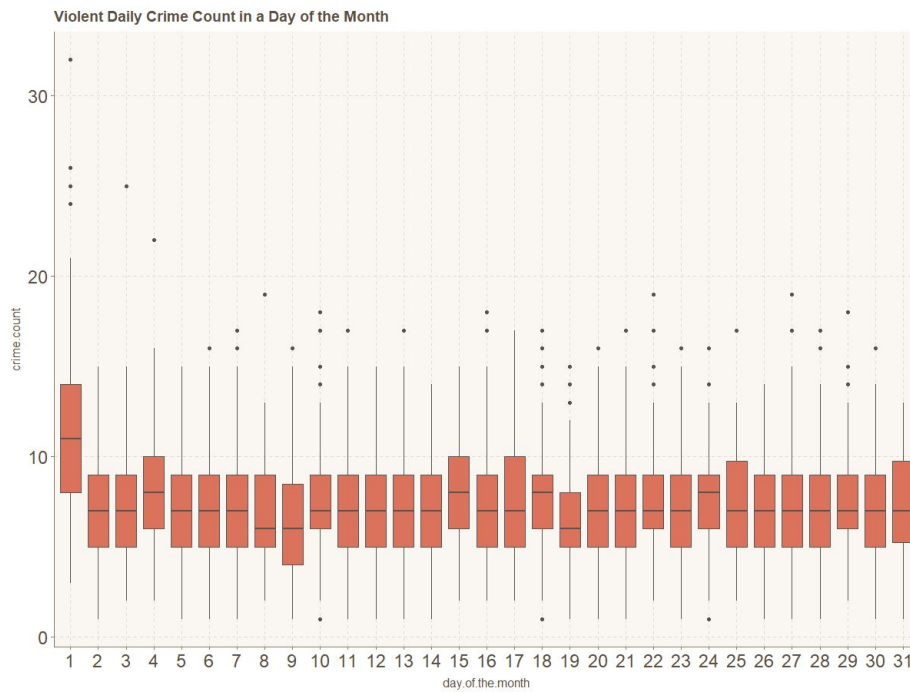
## Chart 8: Crime Distribution by Day of Month - violent crime

**Violent Daily Crime Count in a Day of the Month**

## Chart 9: Prediction Visualization

**Daily crime count: actual v.s. prediction**
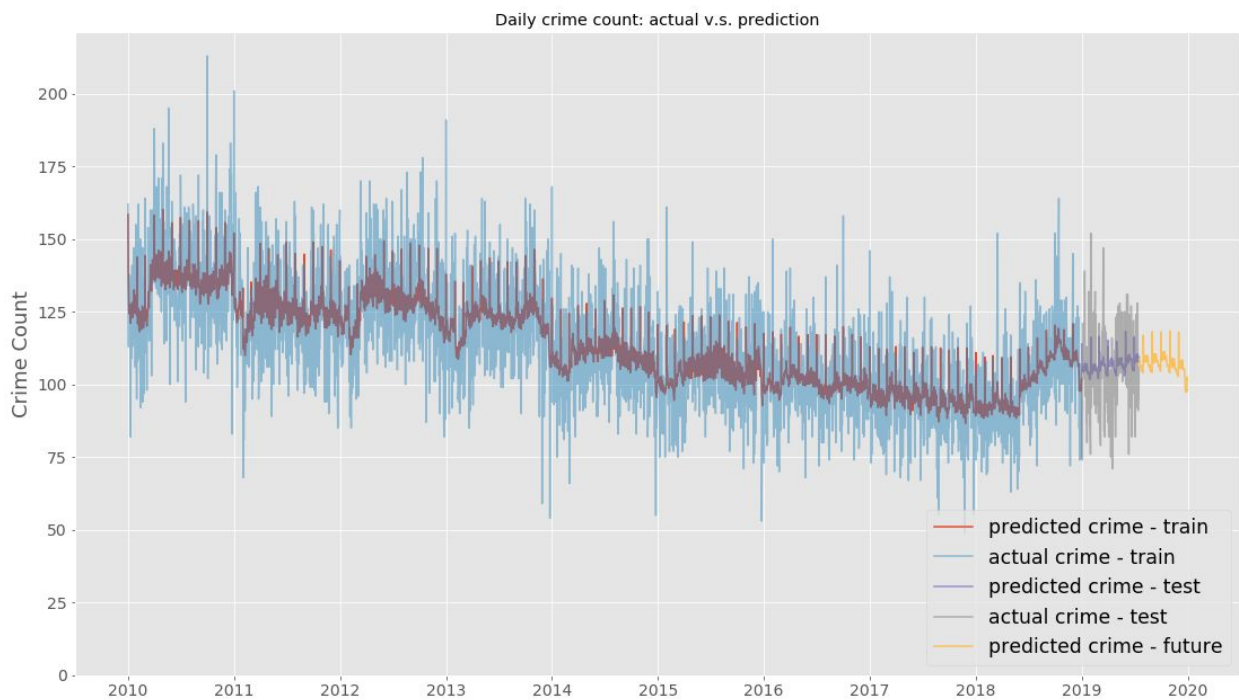
# Chart 10: Prediction Visualization - a closer look



Daily crime count: actual v.s. prediction
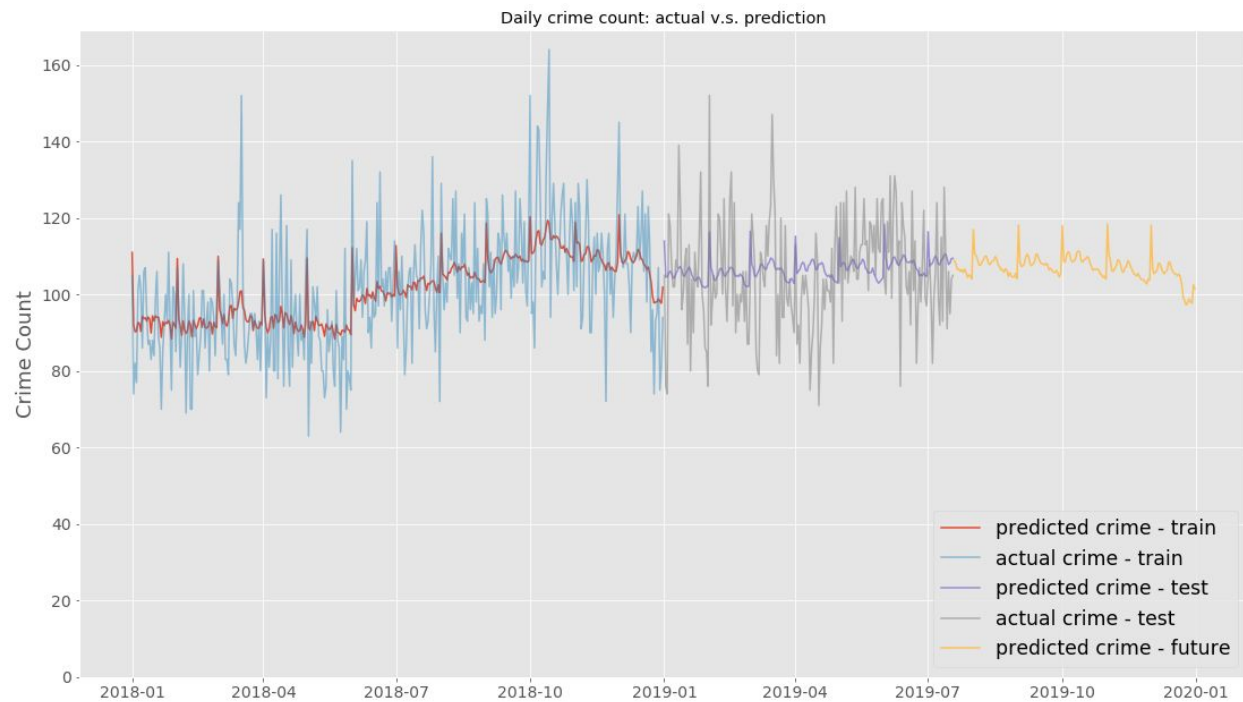
# Chart 11: Crime Distribution per APD Sector in 2018
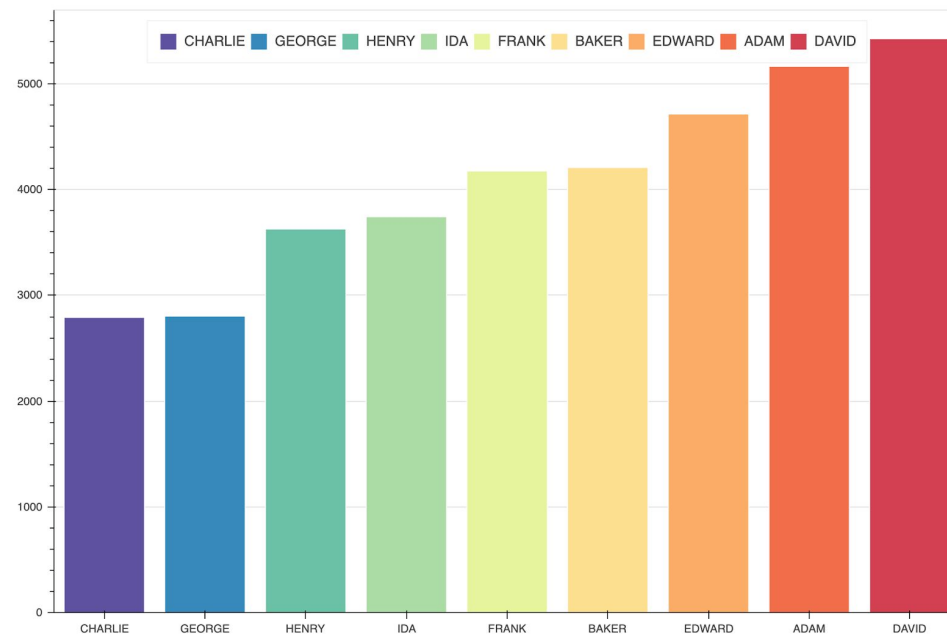
## Chart 12: Crime by APD Sector per 100 Austinites in 2018 - violent crime



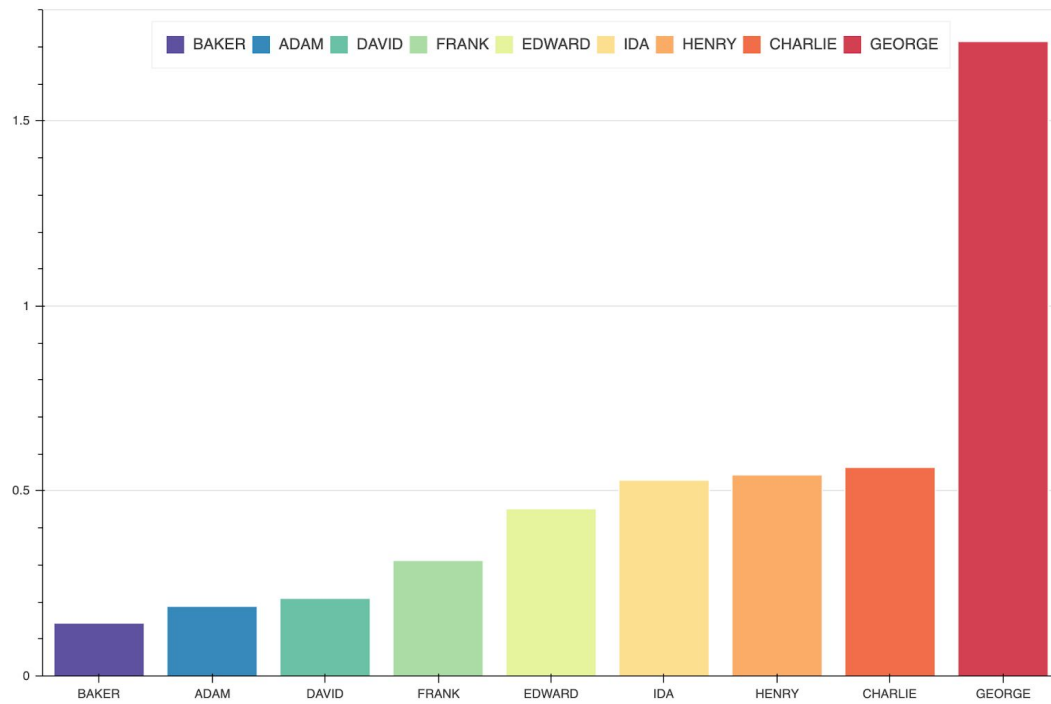Legend: BAKER, ADAM, DAVID, FRANK, EDWARD, IDA, HENRY, CHARLIE, GEORGE

## Chart 13: Crime by APD Sector per 100 Austinites in 2018 - non-violent



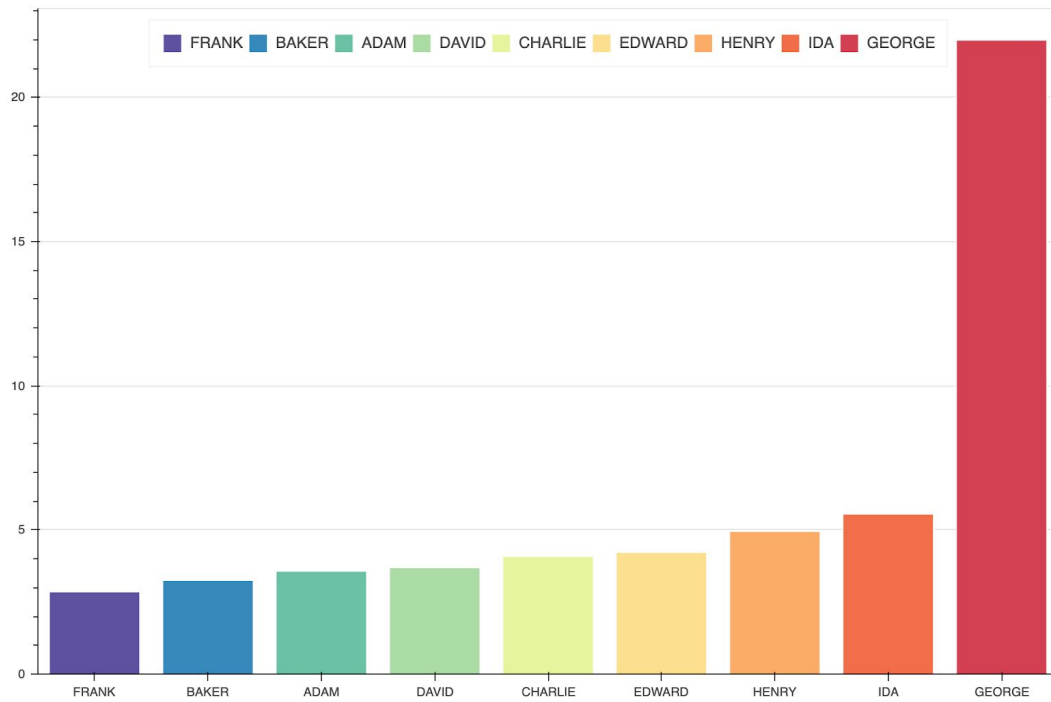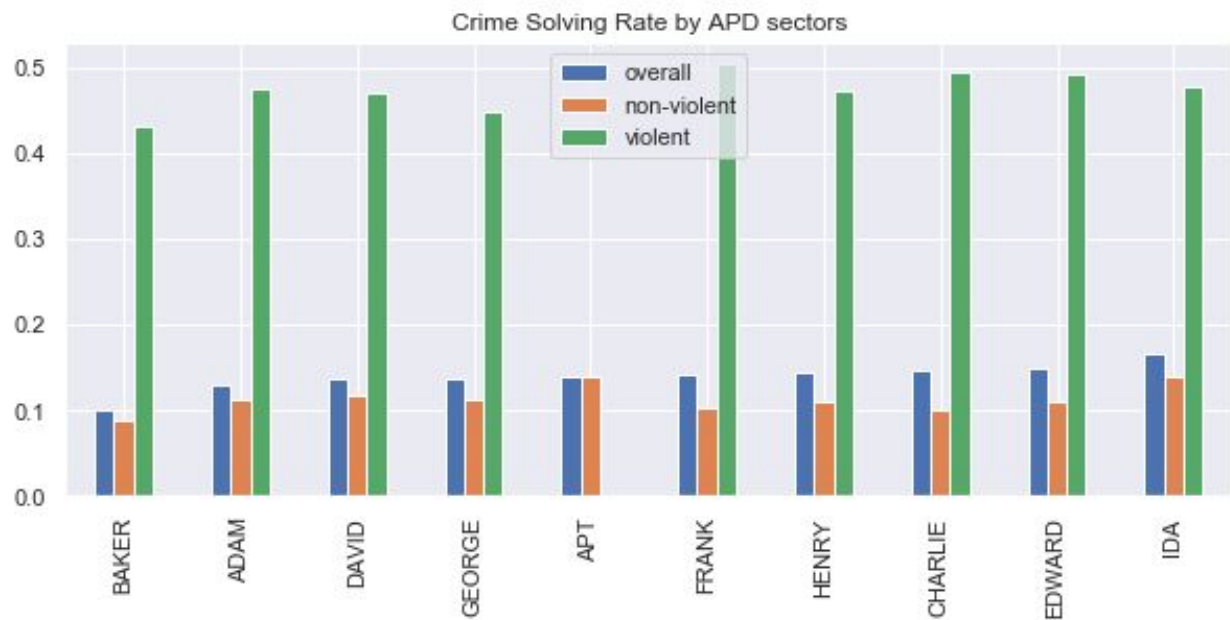Legend: FRANK, BAKER, ADAM, DAVID, CHARLIE, EDWARD, HENRY, IDA, GEORGE

Chart 14: Crime Clearance Rate by APD sectors



Chart 15: Time Taken to Solve by APD sectors