

Crime Rate Analysis for Charlottesville

Group B - Ajay Sanjeevan, Anushruti Huilgol, Sarvin Motamen

May 10, 2023

1 Introduction

Crime and violation are the threat to justice and meant to be controlled. In recent months, there has been a surge in crime rate, especially gun violence, in Charlottesville. In this project we will try to determine which variables affect crime rate by looking at trends. We will also aim to predict future criminal activities by training linear regression models using data from the past, as well as help police distribute forces optimally.

2 Dataset

We have collected the dataset from <https://opendata.charlottesville.org/>. The dataset was created by a private members whose details are unknown due to security reasons. This dataset contains data of the initial information given by an individual about an incident to the police. This is a live dataset which gets updated almost everyday. The data for the last five years is available. Currently, The dataset contains around 25,000 observations, and it contains information such as: type of offense, name of the street, date and time of the report.

RecordID	Offense	IncidentID	BlockNumber	StreetName	Agency	DateReported	HourReported	ReportingOfficer
1	Assault Simple	2.023e+11	800	HARDY DR, A	CPD	2023/04/13 22:24:09+00	1824	Middleton, Matthew
2	Hit and Run	2.023e+11	1900	EMMET ST N	CPD	2023/04/13 21:50:40+00	1750	Lowry, Ryan
3	Assault Intimidation	2.023e+11	1600	MERIDIAN ST	CPD	2023/04/13 19:20:45+00	1520	Burnett, James
4	Larceny - All Other	2.023e+11	100	2ND ST NE	CPD	2023/04/13 17:50:00+00	1350	Nash, Darius
5	Hit and Run	2.023e+11	1	EMMET ST S / IVY RD	CPD	2023/04/13 03:36:48+00	2336	Payne, Katherine
6	Hit and Run	2.023e+11	1200	GRADY AVE	CPD	2023/04/13 03:14:22+00	2314	Payne, Katherine

Figure 1: Crime rate dataset

3 Statistical Analysis

Data, on its own, is not capable enough to show the valuable insights that can help solve many business problems. To unravel the data patterns and correlations, performing a statistical analysis is an important step for any dataset. This helps us understand the data trends and gather more relevant information from the raw data. In our work, we have brought out various trends in our dataset to get in-depth understanding of the crime trends in Charlottesville. The intention is to understand the pattern to bring down the crimes and make Charlottesville a safe place.

3.1 Understanding the overall crime pattern over the years

The first step for our statistical analysis would be to understand the overall trend of crime over the last five years. From Figure 1, we can see that the trend of crime shows an increasing pattern. From 2018 to 2019 and from 2021 to 2023, we see there is a linear increase. However, we do observe a drop in the year 2020 to 2021. We can assume here that the reason could be covid and lockdowns.

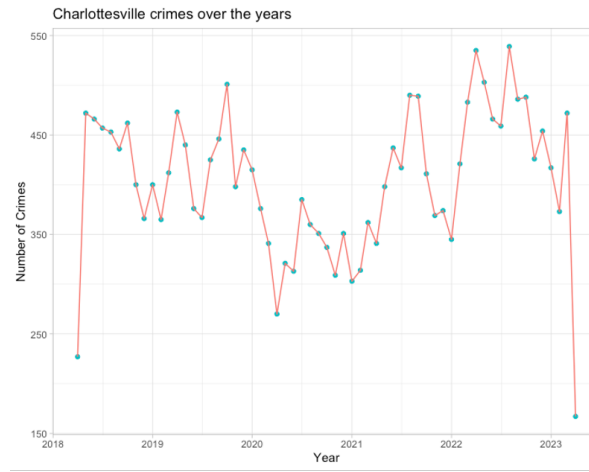


Figure 2: Year Wise Trend

To understand the further pattern of crimes, we plotted the graph individually for every year to see how the crime rate was for the entire year, and we could see that for some of the months, there was an increase and decrease. However, we failed to gather a clear indication. Hence, we did more analysis on monthly data to see if there was any month with high crime rate. From Figure 2, we can see that year 2022, has the highest crime trend compared to other years thus confirming our initial analysis of a linear increase in crime rate. From the monthly average trend, we can see that for all the years, August month tends to have a higher crime rate. Though, we can see that there is no clear pattern and could not gather or find a linear increase or decrease trend in the monthly analysis

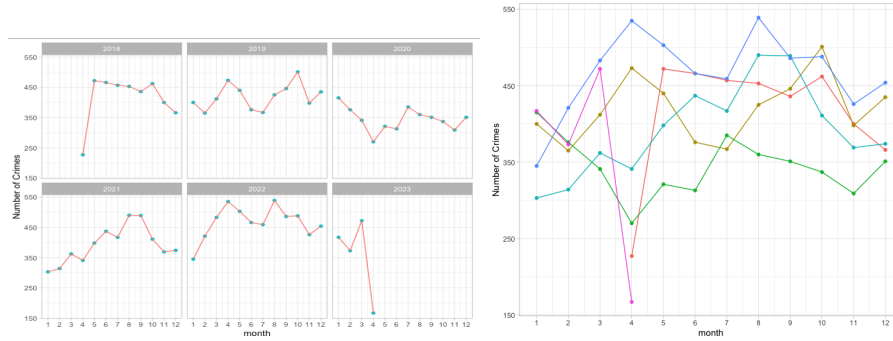


Figure 3: Individual Year and Monthly Average Trend over the years

3.2 Understanding the streets contributing to high crime rate

Once we understood the general trend of crimes in Charlottesville over the years, we gathered more information from the dataset to see which streets were unsafe and had high crime rates. To understand the overall trend, we plotted a graph of the number of crimes over all the years in every street.

From the plot, we saw three major peaks. We observed that for some of the streets, the crime rates were extremely high compared to the others. Some of them, namely, E Market St, Emmet St N, and W Main St had a crime rate above 900. We went on to do further analysis on these few streets to understand how the crime trend has been over the years. For this, we took the top 6 unsafe streets and observed the trend.

From the above plot, what we could analyze is that for all the streets, year by year the crime rate kept on increasing except for University Ave. Among all the trends, Emmet St N seems to be having an extreme increase over the years which is quite alarming.

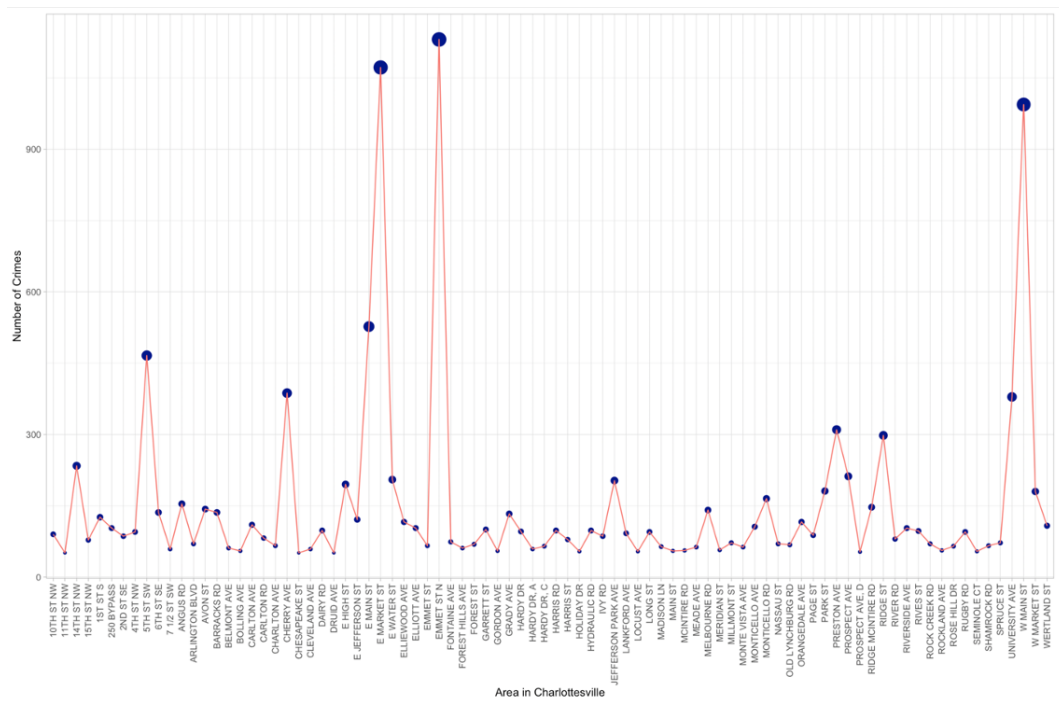


Figure 4: Overall Crime rate in each street

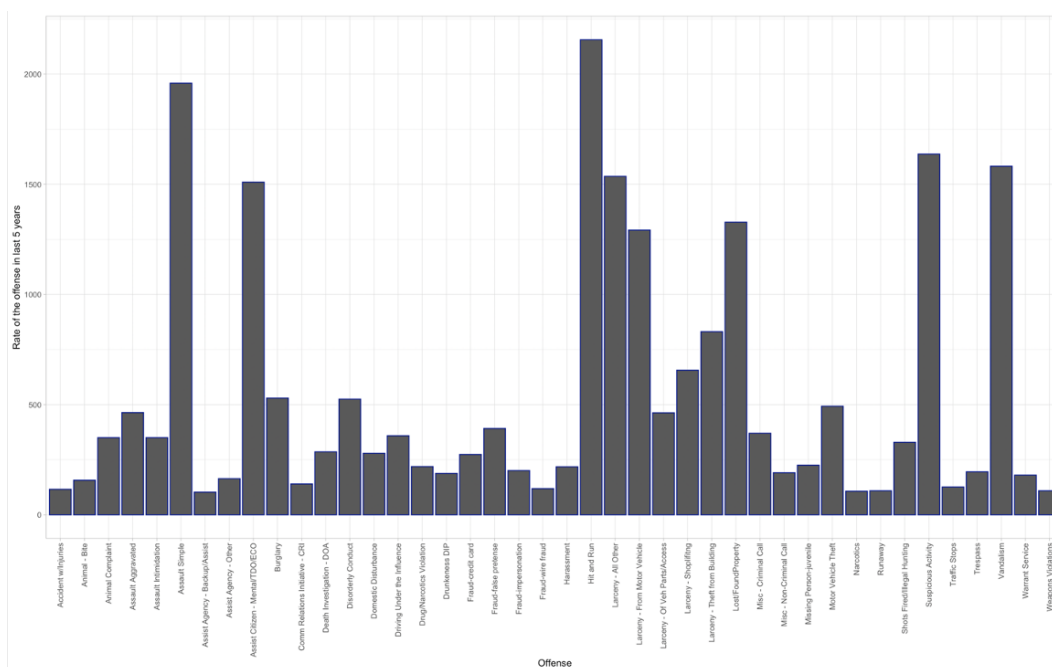


Figure 5: Crime Rate trend in top 6 unsafe streets

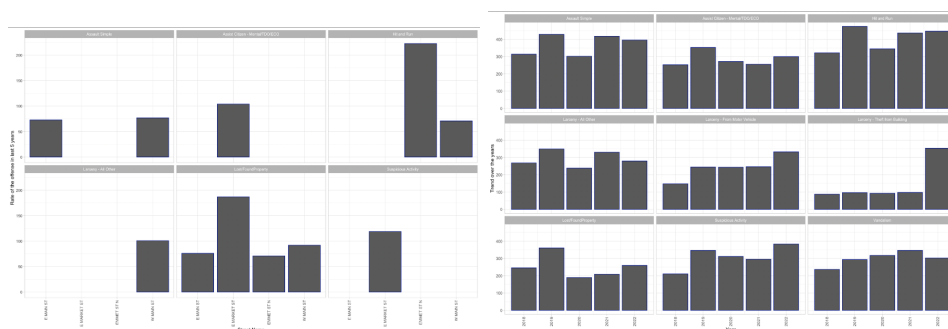
3.3 Understanding the Offense contributing to high crime rate

Based on the streets that were unsafe, we further tried to analyze which crimes were the most committed ones in these streets to understand if there is any pattern. To understand this, we first did an overall analysis of the highest crime that is committed in Charlottesville.

From Figure 5, we saw that the highest crime committed in Charlottesville is "Hit and Run" and "Simple Assault" being the second highest. To further understand what crime was high in the streets



with high crime rates, from the plot Figure 6, we could see which crime was high in which area and also we analyzed the yearly trend of these crimes and we observed a linear increase in all the crimes which is quite bad.



3.4 Understanding the trend of time in which crime is high

Once we observed that over the years, every single crime is having an increasing trend, we see that effective security is needed to curb this and also at the correct time. To see the pattern in the time in which these crimes are mostly occurring, we plotted the crime rate at every hour of the day for the top 9 offenses over the years.

From our plot in Figure 7, it was clear that most of these crimes were highly committed during the early morning hours or in the late evening hours.

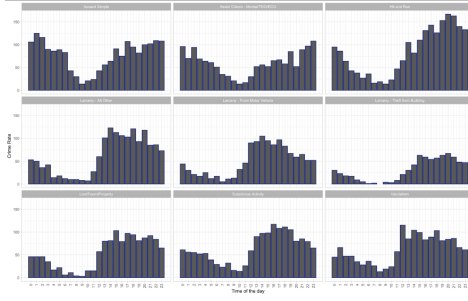


Figure 8: Time of Crime Trends

Overall, the statistical analysis that we performed on the dataset helped us get a better insight into the overall trend of crimes over the years. We saw an increasing trend over the years. With our further in-detail analysis, we uncovered data patterns such as streets that were contributing to high crime rates, the offenses that were mostly committed in those streets, and the time of the day the crimes were high. With these insights, the security in the city can be more intelligently distributed in the streets with high crime rates and around the time when the crime rate is high. This will indeed help in reducing the crime rate.

4 Model Analysis 1

In our work, we used Linear Regression for crime prediction. The main reason for choosing Linear regression is because:

1. Most of the statistical analysis graphs on our dataset showed a linear trend.
2. Linear Regression is one the most popular model for Time Series data Prediction.

For our Dataset, we used Linear regression for predicting the number of crimes in a particular area for a given month and year. We tried multiple degrees of linear fit and evaluated the results.

Degree/ Statistic	1st degree linear model	2nd degree linear model	3rd degree linear model
MSE	134.18	133.927125	133.870691
p-val	< 2.2e-16	< 2.2e-16	< 2.2e-16
R-squared	0.7711	0.772	0.7722

Figure 9: Model Analysis 1

As shown in Figure 9, the MSE value kept decreasing and the R-squared value kept increasing with the increase in the complexity of the linear model. From this, we can conclude that a 3rd-degree linear model is the best fit compared to a 1st-degree and 2nd-degree linear model for this dataset and this scenario. Also, the p-value was less than 0.05 in all three cases which further asserts the significance of our model fit.

5 Model Analysis 2

In Model Analysis 2, we fit a Linear regression model for predicting the number of crimes for a particular offense for a given month and year. The results are as follows.

Similar to Model 1 analysis, the 3rd-degree Linear Model was again the best fit for this scenario (from Figure 10).

Degree/ Statistic	1st degree linear model	2nd degree linear model	3rd degree linear model
MSE	236.7818	235.8914	235.1656
p-val	< 2.2e-16	< 2.2e-16	2.2e-16
R-squared	0.8184	0.8198	0.8209

Figure 10: Model Analysis 2

6 Hypothesis Testing

We performed hypothesis testing on the Linear Regression fit with The null hypothesis stating that there is no correlation between predictor variables and the response and the alternate hypothesis stating that there exists a relationship between the predictor variables and the response.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.233e-01	1.311e+00	0.704	0.481185
poly(year, 3)1	9.358e-01	1.611e+00	0.581	0.561373
poly(year, 3)2	5.084e+00	1.533e+00	3.316	0.000916 ***
poly(year, 3)3	-3.349e-02	1.596e+00	-0.021	0.983252
poly(month, 3)1	4.661e+00	1.610e+00	2.895	0.003803 **
poly(month, 3)2	-8.363e+00	1.535e+00	-5.449	5.17e-08 ***
poly(month, 3)3	-4.459e+00	1.503e+00	-2.967	0.003019 **

Figure 11: Summary Model 1

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.696e-02	4.027e+00	-0.014	0.988715
poly(Year, 3)1	1.095e+01	4.261e+00	2.571	0.010193 *
poly(Year, 3)2	1.261e+01	4.099e+00	3.077	0.002106 **
poly(Year, 3)3	-1.218e+01	4.230e+00	-2.879	0.004016 **
poly(Month, 3)1	1.544e+01	4.322e+00	3.573	0.000358 ***
poly(Month, 3)2	-1.507e+01	4.142e+00	-3.639	0.000278 ***
poly(Month, 3)3	-1.437e+01	4.061e+00	-3.538	0.000409 ***

Figure 12: Summary Model 2

We collected summary of the first model fit and the second model fit for result analysis. As shown in the summary of the first model (Figure 11) and the second model (Figure 12), we could see that year and month have p-val less than 0.05 which supports the rejection of null hypothesis. Hence, we can state that there exists a correlation between the predictor variables and the response variable.

7 Conclusion

Safety has by far been the most important goal for any city or state or country. The goal of the government or local security is to make its people feel safe and have the freedom to explore the town fearlessly. Our analysis intends to reduce the crime rate through an effective distribution of the security forces. Through our statistical analysis, we tried to bring out patterns to see which streets do need higher security and around what time of the day, and based on our model predictions of crime rate for a particular month and year, our suggestion would be to recruit more security in the areas where the prediction of crime is high and also to enforce more specialized security in areas to curb the particular offense which is predicted to be high.

8 References

1. Dataset - <https://opendata.charlottesville.org/>