

Crime and Place: A Spatial Analysis of Crime

Spencer Epley, Gavin Pilato, John Byiringiro, Jackson Stevens, Agyei Sraha

Department of Data Science, University of North Carolina at Charlotte

DTSC 3601/3602: Predictive Analytics and Their Implications

Dr. Nadia Najjar & Dr. Shannon Reid

December 8, 2023

INTRODUCTION

The geographical context of crime has long been recognized as a crucial factor in understanding and addressing criminal activities (Brantingham & Brantingham, 1995). This paper provides a comprehensive review of the fundamental concepts related to the role of location in crime analysis, with a particular focus on Crime Generators, Crime Attractors, and Offender Anchor Points. It explores the significance of these concepts in the fields of crime mapping, predictive policing, community safety, and crime prevention. This report also emphasizes the importance of tailored interventions and the evaluation of intervention effectiveness. Furthermore, it calls for continued research in this area to develop more optimal prevention methods as well as a deeper understanding of the interaction between location and criminal activities. Unique scholarly articles are integrated to enrich the discussion of our topic; this allows us to contextualize our problem statement, research, and eventual findings based on previous work.

The spatial dimension of criminal activities has been a subject of increasing interest in recent years. A deeper understanding of how crime patterns are influenced by the locations in which they occur is essential for law enforcement agencies, policymakers, and communities. The present analysis provides an in-depth exploration of the concepts of Crime Generators, Crime Attractors, and Offender Anchor Points and their implications in crime analysis. By drawing on a range of scholarly articles, both methodological and subject matter-oriented, we seek to elucidate the key concepts and their significance.

One of the foundational concepts we explore is the differentiation between Crime Generators and Crime Attractors. According to Patricia and Paul Brantingham (1995), Crime

Generators are specific areas that draw large numbers of people for reasons unrelated to any particular level of criminal motivation they might have or to any specific crime they might end up committing. These are places that are generally bustling with foot traffic (i.e. popular shopping malls). Conversely, Crime Attractors are places that create opportunities for criminal activity, which, in turn, attracts criminals due to the location's association with a particular type of crime (i.e. drug markets, prostitution areas). It's almost like a vicious cycle—people are drawn to popular places, turning them into Crime Generators, and once crime takes root in these areas, they become Crime Attractors. For instance, think of a bar in a bustling shopping district known for its rowdy patrons who engage in fights. The shopping district consistently attracts a crowd, and the bar – with its reputation for rambunctious behavior – draws people in who eventually get into altercations (Brantingham & Brantingham, 1995)

Taking this a step further, Wim Bernasco and Richard Block in their study “Robberies in Chicago: A Block-Level Analysis of the Influence of Crime Generators, Crime Attractors, and Offender Anchor Points” introduce the concept of Offender Anchor Points. This concept underscores the relevance of where offenders reside in relation to the locations where crimes occur. Based on the principle of least effort, we can infer that offenders prefer shorter trips over longer ones. Unsurprisingly, their analysis showed that a place is more likely to experience crime if it is located near offenders' homes. This finding is intriguing as it underscores not only the impact of the place itself, but also the proximity of potential offenders (Bernasco & Block, 2010).

In their work published in 2000, Anselin and colleagues discuss various factors related to crime and place. They break down place-based theories of crime, including the Routine Activities Theory, which posits that at certain location types, crime is influenced by the routine

activities taking place there. In addition, they explore the concept of crime hot spots, which are areas where higher levels of crime are concentrated. These hot spots can be influenced by factors such as the number of taverns, social disorder, and environmental conditions, underscoring the relationship between specific facilities and elevated crime levels (Anselin et al., 2000).

Chapter 8 of the Oxford Handbook of Crime Prevention delves into the theories surrounding crime and place. This chapter provides practical advice and insights on how to prevent crimes at specific locations. It suggests that interventions, in most cases, can effectively reduce crime in particular places. It also evaluates the effectiveness of various intervention measures, such as the use of CCTV cameras in public places. Interestingly, researchers Eck and Guerette note that while there is not a large sample size for this study, there is still a positive correlation between most forms of intervention and their corresponding reduction in crime. Surprisingly, CCTV does not appear to be very effective in deterring or preventing crime in places; it consistently ranks as one of the least effective methods (Eck & Guerette, 2012).

However, there is a theoretical and practical complexity when it comes to location-based crime prevention. One dominant theory suggests that different prevention measures tailored to different places should reduce crime. The idea is that by customizing interventions based on the unique characteristics of a location, crime rates can be lowered. As Eck & Guerette state, "While place-based prevention is effective, we cannot say precisely how useful it is. There are a number of reasons for this, including the methods used to evaluate these interventions. But a more important reason is that places are quite varied." (87) In essence, it is challenging to implement personalized interventions because places are brimming with diverse populations and a wide array of criminal activities. What works for one park is very likely to not work at another park (Eck & Guerette, 2012).

Place-based crime prevention is an open-ended theory encompassing a range of strategies from increased lighting and street closures to enhanced police presence. Each of these methods has varying effectiveness in different location types. Given these considerations, further research is essential. Place-based crime prevention is a promising approach to reducing crime, but its complexity makes it challenging to study. The effectiveness of interventions can vary, and understanding the contextual factors that influence their success is crucial. This prompts us to investigate the combined impact of multiple interventions in addressing crime in specific locations.

PROBLEM

The Charlotte metropolitan area, with its distinctive mix of economic promise and cultural vitality, has risen as a symbol of national growth in the United States. Attracting both individuals and businesses with its opportunities, Charlotte has left a lasting impression by securing the 3rd position in a national survey conducted by the Inspection Support Network of large cities with the most recently-built housing units (Inspection Support Network, 2023). Over a five-year period, from 2018 to 2023, the city saw a remarkable 20% increase in population (Charlotte Metro Area Population 1950-2023, 2023). However, this rapid growth has brought with it a range of challenges. The city is grappling with a significant lack of affordable housing and an infrastructure that is struggling to keep up with the pace of expansion, from traffic congestion to stretched public services. Moreover, the increase in population density has unfortunately been accompanied by a rise in crime, highlighting the complex issues Charlotte is facing (Chemtob, 2021).

In response to the urgent need for affordable housing, developers have demonstrated remarkable adaptability and foresight. Motivated by the situation, they have redoubled their efforts to build new residences, especially apartment complexes, to house the growing population. As a result of these focused efforts, Charlotte now proudly ranks 11th nationwide in terms of new apartment construction, as reported by authoritative sources like RentCafe and Yardi Matrix. Doug Ressler from Yardi Matrix offers a glimpse into the scale of these projects, stating, “You’re talking about 14,000 units that are coming on board this year, over 14,000 that are coming on board next year. The year after that, it’s going to be 11,000” (Dellinger, 2023). As Charlotte works to balance its rapid growth with sustainable development, such initiatives reflect the city’s commitment to creating a harmonious and thriving environment for its diverse residents.

As Charlotte enters a phase of transformation, characterized by a significant increase in its population and a skyline dotted with apartment complexes, our research group is driven to examine the complex relationship between these new residential buildings and the changing crime landscape. While urban growth often leads to economic growth and cultural vibrancy, it can also bring unexpected challenges. Research, such as the study by Sampson, Raudenbush, & Earls (1997) on neighborhood-level factors and crime, suggests that rapid urbanization can sometimes be linked to increases in criminal activity. With the anticipated rise of apartment communities in the Charlotte area, our goal is not only to determine if there is a correlation between these structures and crime, but also to analyze the underlying factors that contribute to such patterns. Additionally, informed by Brantingham & Brantingham (1982) work on crime pattern theory, we are eager to determine if certain criminal acts become more prevalent in areas with high residential density.

Central to our research initiative is the nuanced exploration of the key concept of ‘Place’ and its significant interaction with crime trends. ‘Place’ extends beyond mere physical dimensions to include the socio-cultural, economic, and psychological aspects of a location. Guided by Eck’s (1994) work on crime places in crime theory, we are keenly investigating the possibility of apartment complexes serving as “Crime Generators”. As explained by the Department of Justice and supported by Sherman, Gartin, & Buerger’s (1989) studies on hotspots, a Crime Generator refers to areas that attract large crowds not necessarily predisposed to criminal behavior, but drawn by various attractions. The high concentration of people in these areas might inadvertently increase the likelihood of misdemeanors or crimes due to opportunities arising from crowd dynamics. On the other hand, informed by Brantingham & Brantingham (1995) conceptual framework on crime, we are also interested in exploring the potential impact of high-density residential areas on crime patterns.

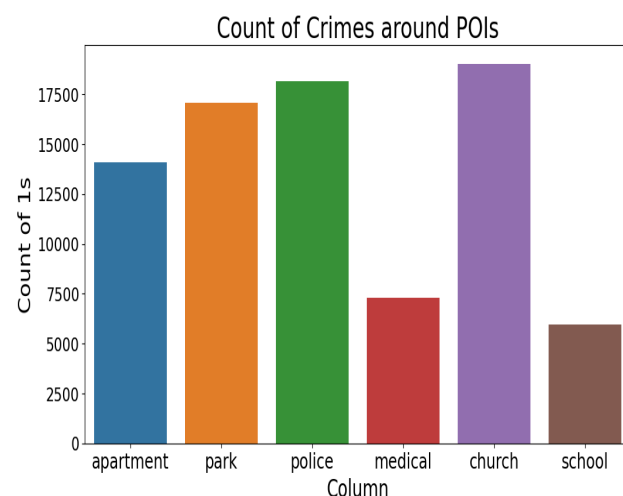
DATA

In our analysis, we used a variety of data sources and went through intensive data cleaning and integration methods. Our primary data source is the Mecklenburg County GIS open source data repository, which provides an abundance of parcel data critical to our investigation of crime and its spatial patterns. This data is instrumental in our research as it outlines the location of points of interest (POIs), allowing us to establish buffer zones. The parcel data predominantly comes with zip codes, which facilitates easy sorting. However, some data points lack zip codes, necessitating careful exclusion of all locations outside of our target zip code. In addition to the parcel data, we are also using crime data from the years 2011-2016, provided in class. This data encompasses a wide range of information, including crime type, location, and coordinates.

Our data cleaning process involved dropping several columns from the original dataset that were not relevant to our analysis. These include zip codes, block number, complaint number, apartment number, city and state, target number, object ID, and several other columns. There have been no missing data values thus far, at least in the columns we are interested in analyzing. Our main target variable is `near_apartment`, a binary variable that indicates whether a crime is located in or around an apartment complex.

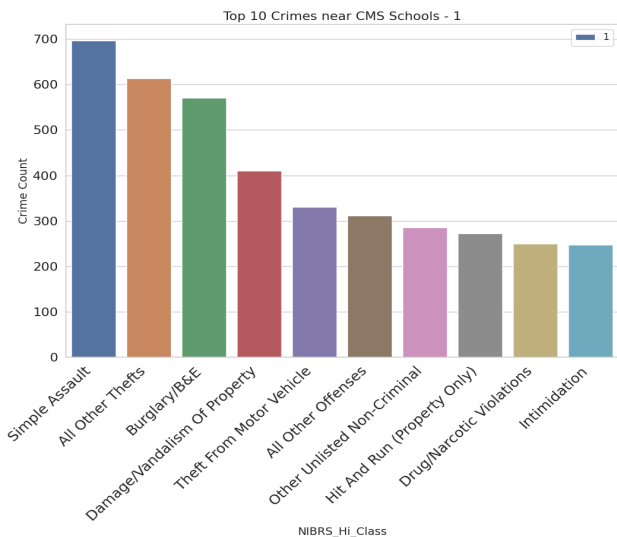
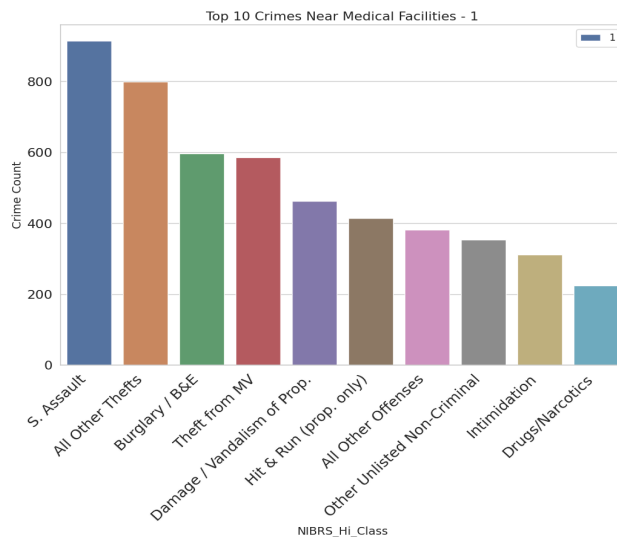
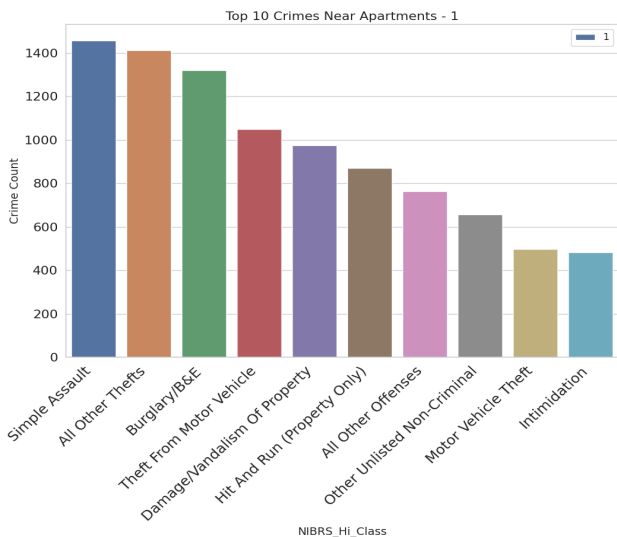
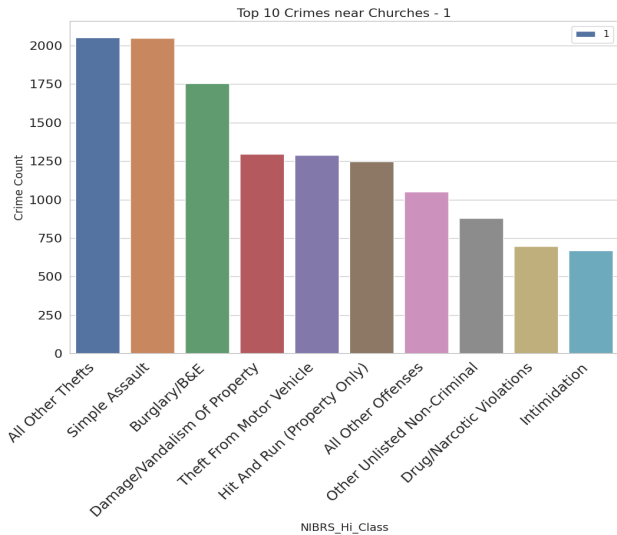
We have multiple other target/predictor variables, including `near_medical_facility`, `near_CMS_school`, `near_church`, `near_park`, `near_police` as location variables. These function similarly to `near_apartment`. We also created dummy variables for our ‘NIBRS_Hi_Class’ variable, which essentially turned each unique crime into its own, binary column. A 1 meant that the specific type of crime had occurred, a 0 meant it hadn’t. Our research involves a comprehensive process of data collection, cleaning, and analysis, with a focus on understanding the relationship between crime and place. We continue to refine our data and methodologies to ensure the accuracy and relevance of our findings.

When combing through our data, we noticed that there wasn’t a lot of crime around medical facilities and schools. Conversely, we noticed the highest rate of crime occurred around churches, ironically enough. All columns in our dataset were binary, simply telling us whether or not something had or hadn’t occurred. We then decided to analyze the type of crime based on our different locations. We started with churches, since they had the highest crime count.

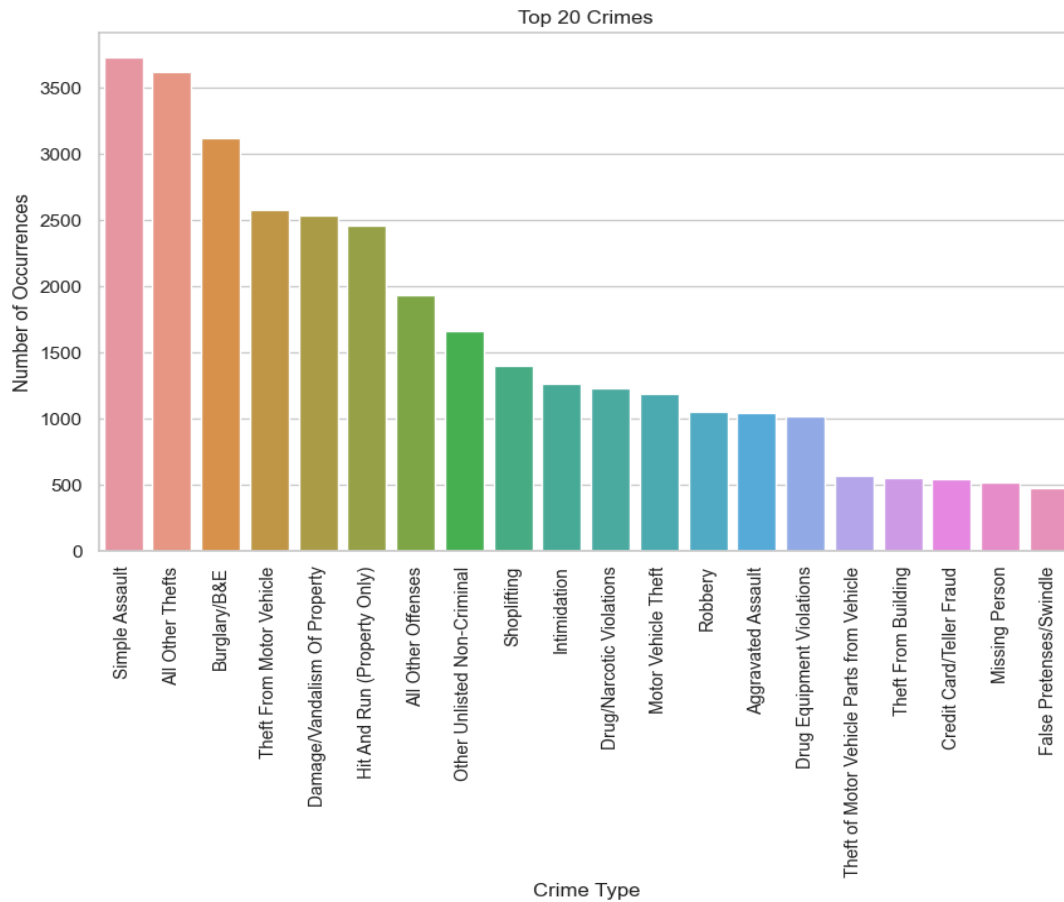


Looking at our church chart, we can see that All Other Thefts, which is a collection of thefts that don't fit into the categories of burglary, theft from motor vehicle, or shoplifting, was our most popular crime. Following that was Simple Assault and Burglary.

Apartments ended up looking very similar to the Church chart. This is where our group started to piece together that crimes like Simple Assault, All Other Thefts, and Burglary were the most popular and simple crimes. We also made charts for our other variables:



As we can see, a lot of our charts ended up looking pretty similar besides a couple of outliers. This is when our group decided to get the crime counts for further analysis:



METHODOLOGY

One of the first steps in our process was collecting necessary data. This consisted of the CMPD data provided in class as well as data used from Mecklenburg County GIS. We used Parcel data from the GIS resource to create our POI variables. Once we collected our data, we dropped some columns we weren't interested in using, such as Street_No, Division, and Apt_No, among others. We ended up dropping around 10 total columns.

Once we finished collecting and cleaning our data, we combined all of the CMPD data, by year, into one big dataset. This allowed us to streamline and compact our analysis. After this, we dropped all columns that were not in the 28205 zip code. We then proceeded to add in our location variables. This was done through ArcGIS pro. We created half mile buffer zones around our parcels of interest, and then plotted our crimes onto a map of Mecklenburg county. We spatially joined the crimes to the buffer zones, and created binary variables based on our locations. A 1 would mean that a crime had occurred in this specific location's buffer zone, and a 0 would mean a crime hadn't occurred.

After our data cleaning and transforming, it was time to begin our modeling. Our group first created a baseline, leaving in all variables we hadn't previously dropped, used 'near_apartment' as our target variable, and employed a Random Forest model. Random Forest models are a collection of decision tree models, in which the decision trees are merged together for a more accurate result. Our group decided to use Random Forest, Gradient Boosting, Naive Bayes, and Logistic Regression modeling techniques for our future models. We did this because we wanted a wide array of different model types to compare effectiveness with our dataset.

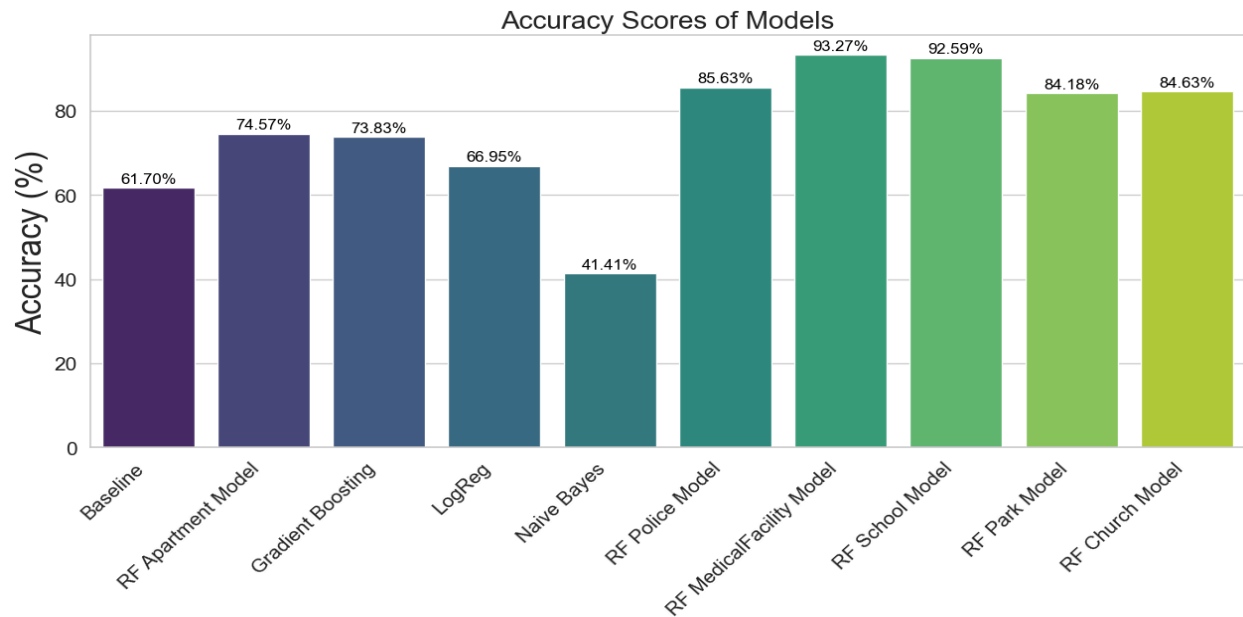
We performed model parameter tuning as our next step. We proceeded to finely tune the Random Forest model and the Gradient Boosting model based on their ideal parameters for this analysis. Logistic Regression and Naive Bayes were not included in this, since these algorithms/models do not contain the same amount of parameters as Random Forest and Gradient Boosting have. Below are the ideal parameters for these models:

<u>Model Type</u>	<u>Best Parameters</u>
Random Forest	• n_estimators: 150 • min_samples_split: 2 • min_samples_leaf: 1 • max_depth: 20
Gradient Boosting	• n_estimators: 100 • max_depth: 5 • learning_rate: 0.2

After parameter tuning, we discovered our updated Random Forest model performed better than our baseline model. We also proceeded to run our Gradient Boosting, Naive Bayes, and Logistic Regression models here. All of these different model types performed worse than our refined Random Forest model, and some even performed worse than the baseline. Gradient Boosting was our most successful model outside of the Random Forest, just slightly below its accuracy score. Logistic Regression and Naive Bayes performed worse, especially our Naive Bayes model. The Naive Bayes model ended up being our worst performing by a large margin.

RESULTS

After we saw how well Random Forest performed with our apartment variable and how it compared to the other model types, we decided to use that model with our other targets. Further model testing showed that the Random Forest was even more useful for the other target variables, as we returned some very high accuracy scores. In our next section, we will break down all the accuracy scores and models even further. Below are the accuracy scores of all our models, organized into a graph:



To obtain these, we used the `accuracy_score` function from Sklearn. `accuracy_score` calculates the accuracy score for a set of predicted labels against true labels.

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i)$$

Formula for the accuracy score function. It compares your Y_{test} set to a set of predicted values, and however many matches result in your accuracy score.

Firstly, you have to create a training and testing set. We used the `test_train_split` function also from Sklearn to do this. We created a `X_train`, `X_test`, `Y_train`, and a `Y_test`. `X` was set to all columns besides our target, and `Y` was the column of our target variable. As explained above, the `accuracy_score` function compares your `Y_test` set (the data set aside from training), and compares it to a `Y_pred` set, which is a set of predicted values based on how you trained your model. For example, our Random Forest School model returned an accuracy score of 92.73%.

This means that our test set labels for the School model matched up with the predicted set about 92.73% of the time.

We also calculated accuracy for our Random Forest models employing a technique known as confusion matrices:

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

A confusion matrix is basically a visual representation of your accuracy score. It shows you True Positives, True Negatives, False Positives, and False Negatives. These terms can be explained as:

- True Positive (TP): It refers to the number of predictions where the classifier correctly predicts the positive class as positive.
- True Negative (TN): It refers to the number of predictions where the classifier correctly predicts the negative class as negative.
- False Positive (FP): It refers to the number of predictions where the classifier incorrectly predicts the negative class as positive.
- False Negative (FN): It refers to the number of predictions where the classifier incorrectly predicts the positive class as negative.

This can be very useful in visualizing model accuracy, as you can physically see more clearly where your model is messing up potentially. It also helps make understanding your data and model very simple. However, with confusion matrices comes new model evaluation techniques. We still employed Accuracy as our main evaluation method, but we later employed Recall, Precision, and F1 score after we made our confusion matrices.

- Recall is the measure of all positive classes, how many ended up being identified correctly.

$$\text{Recall} = \frac{TP}{TP + FN}$$

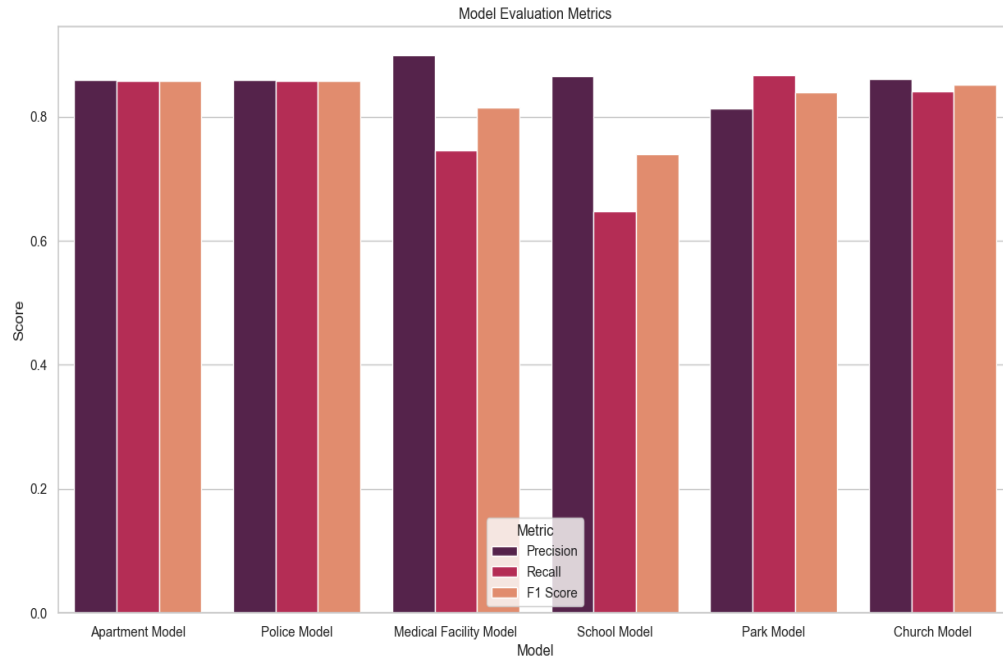
- Precision is the ratio of true positive predictions to total positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- F1 score is a balanced measurement that takes into consideration both Recall and Precision. For example, a high F1 score indicates you have a good, balanced model.

$$F - \text{measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Below is a graphic representation of all our Random Forest model's Recall, Precision, and F1 scores, and a table showcasing our exact scores:



<u>Model</u>	<u>Precision(%)</u>	<u>Recall(%)</u>	<u>F1 Score(%)</u>
Apartment Model	85.84%	85.71%	85.77%
Police Model	85.84%	86.7%	84.36%
Medical Model	89.95%	74.49%	81.50%
School Model	86.45%	64.69%	74.00%
Park Model	81.23%	86.65%	83.86%
Church Model	86.13%	84.10%	85.11%

We can make some interesting insights into our models based on this graph and table.

Firstly, our apartment and police model were the most balanced, sporting the highest F1 scores and also having low variation amongst the tests. We also noticed that our Medical Facility model and School model had very high Precision scores, but lower Recall scores than the rest. These models were highly accurate, but we can see they were very good at matching true positives to total positives, but didn't measure well with just correct positive predictions.

Overall, we were very happy with the results of our Random Forest models. All scored fairly high in accuracy, and all scored higher than our baseline. They also had good recall, precision, and F1-scores. From this, we can conclude that not only were the Random Forest models a good fit for our data, but also that these models were effective at predicting crime at a certain place.

TOOLS

For our modeling and model evaluation, SKlearn was very important. SKlearn, or Sickit learn, is a machine learning library/repository for Python. It contained all of our different modeling techniques, as well as the modules for accuracy score, recall, precision, and F1 score. All coding was handled within Python, and for our graphs we used Seaborn, which is a statistical data visualization package compatible with Python. It is used to create aesthetically pleasing graphs, and provides more ways to manipulate graphs than normal pandas can. For data frame manipulation and cleaning, we used Pandas, a data analysis library for Python that allows you to perform different kinds of operations on dataframes. It can allow you to easily access basic information about your data, and also allows you to easily format your dataframe.

ETHICAL IMPLICATIONS

Towards the start of this research, when we were discussing how exactly we could use information from our study or subsequent ones could be used to assist in policing and lowering crime rates in the area, we suggested that the information from our data exploration and results could be used in the short term to determine which combinations of points of interests to allocate resources toward crime-preventive measures and proactive policing. We also said that in the long term, the information could also be able to inform strategic placements of points of interest to

hopefully reduce the likelihood of violent crime occurring, especially aiming to reduce the chances of vulnerable populations being a victim of one. One of the ideas that can be presented for doing such is placing schools and medical facilities further away from crime-attracting apartments, however an ethical issue with that is that in doing so, you could be moving better quality schools further away from populations who already struggle to get good quality education for their children, or also make it more difficult for disadvantaged families with unreliable transportation to have their children be present for school at all. In a similar vein, moving medical facilities further away can increase medical response time to those communities as well, which can more directly affect their health and wellbeing.

LESSONS LEARNED

Our investigation into the spatial dynamics of crime in Charlotte has yielded valuable insights, particularly in the realms of location-based criminal behavior and preventive measures. Clearly, locationally-based crime prevention, while promising, presents complexities. Customizing interventions is crucial, but their effectiveness can vary due to diverse populations and criminal activities. Further research is essential to understand contextual factors influencing intervention success (Eck & Guerette, 2012).

Our analysis of Charlotte's growth reveals a nuanced relationship between urban development, population growth, and crime patterns. As the city undergoes rapid changes, understanding the impact of these transformations on crime becomes imperative. Our models provide some intriguing perspectives into the ethical concerns that could arise from changes in the geographical patterns in which cities are planned. The relationships between crime and our

respective Points of Interest show that these questions cannot be treated as trivial or straightforward.

Many ethical considerations arise in using research findings to inform policing and preventive measures. Balancing crime prevention with potential impacts on vulnerable populations, such as school and medical facility placements, requires careful consideration of social interaction and access to essential services. These ethical questions should be examined relentlessly to ensure that the most good and the least harm is done for the citizens of that locale, and that the reduction of crime can yield a flourishing society.

In conclusion, our research highlights the complexity of the spatial dimension in understanding and preventing crime. The interplay between location, social dynamics, and criminal activities necessitates ongoing interdisciplinary research and ethical considerations for the betterment of communities. Our results not only contribute to the academic discourse on crime analysis but also offer practical insights for informed decision-making in crime prevention and mitigation strategies in Charlotte and potentially in other urban settings.

References

- Anselin, L., Varga, D., Acs, G., & Rey, S. J. (2000). *USA County Level Crime* (pp. 119–131). Springer.
- Bernasco, W., & Block, R. (2010). Robberies in Chicago: A Block-Level Analysis of the Influence of Crime Generators, Crime Attractors, and Offender Anchor Points. *Journal of Research in Crime and Delinquency*, 48(1), 33–57.
<https://doi.org/10.1177/0022427810384135>
- Brantingham, P., & Brantingham, P. (1995). Criminality of place. *European Journal on Criminal Policy and Research*, 3(3), 5–26. <https://doi.org/10.1007/bf02242925>
- Charlotte Metro Area Population 1950-2023*. (2023). Macrotrends.net.
<https://www.macrotrends.net/cities/22954/charlotte/population>
- Chemtob, D. (2021, October 15). *Our affordable housing shortage, in numbers*. Axios Charlotte.
<https://charlotte.axios.com/276572/our-affordable-housing-shortage-in-numbers/>
- CMPD incidents*. (n.d.-a). <https://data.charlottenc.gov/datasets/charlotte::cmpd-incidents-1/about>
- Dellinger, D. (2023, August 31). *Boom in Charlotte apartments could be good news for renters*. Queen City News; Queen City News.
<https://www.qcnews.com/news/local-news/boom-in-charlotte-apartments-could-be-good-news-for-renters/>
- Eck, J., Clarke, R., & Guerette, R. (2007). *Risky Facilities: Crime Concentration in Homogeneous Sets of Establishments and Facilities*.
https://popcenter.asu.edu/sites/default/files/tools/risky_facilities/PDFs/Eck_etal_press.pdf
- Felson, M. (1987). Routine activities and crime prevention in the developing metropolis. *Criminology*. *Criminology*, 25(4)(911-932).

https://heinonline.org/HOL/Page?lname=&public=false&collection=journals&handle=hein.journals/crim25&men_hide=false&men_tab=toc&kind=&page=911

Geospatial Information Services (GIS) | Geospatial Information Services. (n.d.).

<https://gis.mecknc.gov/>

Inspection Support Network. (2023, February 20). *Cities With the Most Recently-Built Homes - ISN*. Inspection Support Network.

<https://www.inspectionsupport.com/cities-with-the-most-recently-built-homes/>

pandas - Python Data Analysis Library. (n.d.). <https://pandas.pydata.org/>

scikit-learn: machine learning in Python — scikit-learn 1.3.2 documentation. (n.d.).

<https://scikit-learn.org/stable/>

Sampson, R. J., Raudenbush, S., & Earls, F. (1997). Neighborhoods and Violent Crime: a Multilevel Study of Collective Efficacy. *Science*, 277(5328), 918–924.

<https://doi.org/10.1126/science.277.5328.918>

seaborn: statistical data visualization — seaborn 0.13.0 documentation. (n.d.).

<https://seaborn.pydata.org/>

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot Spots Of Predatory Crime: Routine Activities And The Criminology Of Place. *Criminology*, 27(1), 27–56.

<https://doi.org/10.1111/j.1745-9125.1989.tb00862.x>