

Disclaimer – the tone of the following report is informal; neither this project's scope nor its incompetence warrants a formal report.

INTRODUCTION

The safety of the public is possibly the most visible aspect of the law and order that our government has the responsibility for. Homicide, specifically, is an abrupt punctuation to our otherwise safe lives, far removed from violence in the day to day. Charlotte, as a relatively large US city with one of the highest rates of growth, is a good case to study crime in that it is undergoing rapid socioeconomic transformation and is slightly above average in overall crime rate.

To tackle the topic of crime rate and data is no small task, so I decided to take a very small chunk and just rank what sort of venues are closest to the recent homicide cases in the city. As will be discussed below, this analysis is so limited that the only insights it should yield are that I learned how to use the Foursquare API and applied a few concepts from this course in its completion.

DATA

The primary sources of data for this project are the Charlotte Open Data Portal (<https://opencharlotte-charlotte.opendata.arcgis.com/>) and the Foursquare API.

The data for the homicides from the Charlotte Open Data Portal contains details about each case, including racial backgrounds and potential motivation, but most importantly for this analysis was simply the location of the crimes as coordinates.

METHODOLOGY

The biggest task was using the Foursquare API to get nearby venues for each case and categorize them appropriately. Within the API response, the categories listed for each venue are too specific for any meaningful analysis (e.g. 'Shanghai Restaurant' and 'Beijing Restaurant' should both be classified as 'Chinese Restaurant'), so a significant amount of my time was spent trying to figure out how to meaningfully use Foursquare's category hierarchy to recategorize all the venues.

I ended up just creating the most broad categories - 'Food,' 'Professional & Other Services,' 'Arts and Entertainment,' etc. - in the interest of having meaningful insight with the size of this dataset and not pulling my hair out too much in doing so.

With the nearby venues' categories for each case, I used very cursory data analysis methods to find which kinds of venues you would be more likely to get murdered near. I employed one-hot encoding and averaged out the cases near each kind of venue, and then created a matrix showing the relative percent chance of murder near two different kinds of venues. Importantly, however, is that these figures are only measured relative to each other and do not take into account the density of different types of venues.

To give a concrete example, I found that you are over five times as likely to be killed near a restaurant and a store compared with near an office and a college; a very easy criticism of this stat would be that there may be five times as many restaurants and shops as offices and colleges, so

the statistic may be meaningless or even harmful in a person's intuition about their safety. On top of this, with only around 500 homicide cases to choose from, several of which happening in the same exact incident, there is little potential for any statistical significance across such a large city no matter what.

RESULTS

As could be guessed, the venue categories for “Shops & Services” as well as “Food” were far and away the most likely types of venues to be located near a homicide. “Colleges & Universities” and “Professional & Other Places” were both hardly near any homicides. There was a significant disparity between the highest and lowest averages, with “Shop & Services” being within 200 meters of about 38% of crimes where as “Colleges & Universities” were within 200 meters of under 1% of homicides.

DISCUSSION

I don't think anyone would be surprised by the statistics generated in this project. A cursory Google search would yield such a volume of shops and restaurants that it would be rather foolish to assume more murders take place near the few-in-numbers parks of Charlotte compared to the thousands of the smaller venues. Further, when you start considering what sorts of activities involved at places, what time they are trafficked, what demographics move through there, etc. then you start to dig at the complexity that informs a murderous situation.

It is almost insulting how many hours I spent trying to work with the data to yield a result that I could have predicted in under five seconds of thoughtful reasoning.

CONCLUSION

In the course of working on this project, I primarily learned about the volume and variety of problems one can run into in the pursuit of creating meaningful data analysis. For example, if I actually wanted to create a more universal metric for murder rate in relation to venue proximity, I would need every single venue registered in Charlotte as a data point so as to create a normalized measure, conceptually something like:

$$\text{Murders within 200 meters of restaurant} / \text{Total restaurants in Charlotte}$$

Truly, the tools of data analysis are endless in the complexity of their use and implications; I know nothing yet, but I'm on my way to knowing something.