Denise Dodd

Trends and Predictions in St. Louis Crime

Proposal

TOPIC

This project will explore crime patterns in St. Louis based on 2020 data and predict future St. Louis crime patterns.

BUSINESS PROBLEM

I have lived in St Louis for over 30 years and find there are many positive aspects about the city. However, when I speak to friends and family outside of the city, the conversation always seems to revolve around crime and St. Louis being one of the most dangerous cities in America. I would like to learn more about the volume and trends surrounding crime in my city. Therefore, I will be approaching this project with the mindset of an Analyst within the St. Louis Police Department. By the end of this project, I will be able to identify how crimes vary among the districts, when crimes occur, and what kind of delays occur between when a crime occurs and when the crime is reported. Finally, I hope to be able to predict the volume of future crimes. With this information, the St. Louis police department can better allocate their personnel and preventative resources, educate the public, and curve future crime rates.

DATASETS

I will be concatenating several monthly datasets detailing crime in St. Louis found here: SLMPD Crime Files (St. Louis Metropolitan Police Department, 2021) and creating one dataset encompassing all crimes reported in 2020. The monthly datasets include several variables such as:

Compliant – Each incident is detailed in a complaint. One complaint can have several crimes listed if they are part of the same incident. For example, if in the process of a robbery there is also an assault, both the robbery and the assault crimes will be listed on the same compliant. Each complaint is assigned a unique numeric identifier recorded in the Complaint variable.

CodedMonth – Month the incident was reported. The twelve separate datasets that will be concatenated in this project are originally separated by this variable.

DateOccur – This has a date and timestamp of when the incident actually occurred. At times, an incident is not reported until several months or years after the crime occurred.

FlagCrime/FlagUnfounded - At times, an incident is reported which is later deemed to be not a crime. If an incident is determined to be a crime, there will be a “Y” in the “FlagCrime” column. If an incident is not determined to be a crime, there will be a “Y” in the “FlagUnfounded” column.

Crime/Description - The “Crime” variable is a numerical code which the police department assigns to each crime. The “Description” variable is a description of each numerical code. These variables should have a 1 to 1 correlation. For example, Crime number “152200” should always have the Description “WEAPONS-CITY VIOL/DISCHRGING IN CITY”.

District/Neighborhood - These variables numerically categorize the 7 different districts and 88 different neighborhoods/parks that the police department serves.

There are additional variables such as an address, coordinates, notes, etc., but I anticipate the variables described above will be used the most in this study.

METHODS

I will be using a combination of grouping data for visualizations and predictive analytics to complete this study. Line graphs and bar graphs will be used to show crime across time elements such as day, month, and hour. Heatmaps and stacked graphs will be used to visualize how various categories of crime are spread out across the 7 city districts. Box and whisker or violin plots will be utilized to identify spread and outliers of elements that require further analysis. Finally, a time-series predictive analysis such as SARIMA or Holt-Winters will be deployed to forecast future crime rates.

ETHICAL CONSIDERATIONS

One ethical concern pertaining to this study is that I would caution the police department not to “push” crime around. For example, if larceny rates are a large concern in district 2, they shouldn’t solely focus on preventing larceny in district 2 to the extent that the larceny crimes migrate to district 3. The goal is to minimize and prevent crime, not just move it around.

Another concern is the generalization and stigma around crime. If a neighborhood has high rates of theft, that does not mean that every resident in that neighborhood is a thief. Additionally, a theft in one neighborhood could have been perpetrated by a resident of another neighborhood.

CHALLENGES/ISSUES

The most glaring issue with this dataset is that it represents the year 2020 which is an anomaly in all aspects of life due to the Covid pandemic. Additionally, 2020 was a turbulent year for the policing profession with public incidents causing numbers in police recruitment to decrease. According to the Police Executive Research Forum, a nationwide study found “The total number of hirings at responding agencies decreased by 20.5% from 2019 to 2020. The number of hirings largely rebounded in 2021, increasing 20.8%” (Police Executive Research Forum, 2022). Due to the Covid pandemic and police staffing, a 2020 dataset would not be ideal when exploring an analysis of crime from a policing perspective, but this is what was available. If the police department has access to a 2021, 2022, or 2023 dataset, they can utilize the methods in this study to pursue more reliable findings.

Another challenge stems from data points that would add context to my findings. For example, I do not know the population of the various districts and neighborhoods. If a district has 100 residents and 50 counts of larceny, that is more concerning than a

district with 1,000 residents and 50 counts of larceny. I will search for a supplemental dataset that I can join with my current data that can add this context.

REFERENCES

The St. Louis Metropolitan Police Department has provided a document titled "Crime Data Frequently Asked Questions" (St. Louis Metropolitan Police Department, 2008) which explains how information is collected, what the variables represent, and how to interpret the data. Additionally, the FBI has a UCR Handbook (Federal Bureau of Investigation, 2004) which explains how crime is reported and how the Uniform Crime Reporting (UCR) Codes present in my dataset map to various categories of crime. These supplemental resources will provide needed information and context to complete my study.

CITATIONS

Federal Bureau of Investigation. (2004). *Uniform Crime Reporting Handbook*. https://ucr.fbi.gov/additional-ucr-publications/ucr_handbook.pdf

Police Executive Research Forum. (2022, March). *PERF Survey Shows Steady Staffing Decrease Over the Past Two Years*. https://www.policeforum.org/workforcemarch2022

St. Louis Metropolitan Police Department. (2008). Crime Data Frequently Asked Questions. https://www.slmpd.org/Crime/CrimeDataFrequentlyAskedQuestions.pdf

St. Louis Metropolitan Police Department. (2021). SLMPD Downloadable Crime Files. https://www.slmpd.org/Crimereports.shtml