

CST 383 Group Project Report

Group Members: Austin Stewart, Thomas Wooten, & Miguel Nunez

i. Introduction:

a. Why was the project undertaken?

a.i. This project was undertaken to better understand how various crimes, like robbery and property crime for example, can be influenced by the population size of a city. To keep the data more centralized and accurate, we only used cities from a single state, which was California, in our data set.

b. What was the research question, the tested hypothesis, or the purpose of the research?

b.i. Here is the research question that we used to model our project, "Given the population and violent crime of a city, can we predict the amount of Property crime?"

ii. Selection of Data:

a. What is the source of the dataset?

a.i. The source of our data comes from the Uniform Crime Reporting (UCR) program of the Federal Bureau of Investigation (FBI) in 2019. Although it does provide data for every state and city within them, we only chose California to keep our data more centralized.

b. Characteristics of data?

b.i. The characteristics of the data set consists of various crimes committed in each city within California. Although California is home to over 470 cities, our data set provides information on 457 of them. Some of these crimes include robbery, arson, property crime, burglary, and many more. Also, it contains the population size within those cities as well.

c. Any munging or feature engineering?

c.i. As for munging of data and or feature engineering, our data set only required one thing, and that was the removal of commas. If they were left in, then it would be impossible to properly graph our data set into appealing visuals.

iii. Methods:

a. What materials/APIs/tools were used or who was included in answering the research question?

- a.i. The materials/APIs/tools we used for our project was GitHub and Jupyter Notebook through Anaconda. GitHub was used as our repository in which all group members were able to download/upload the latest version of the project. This allowed for quick and easy access to everything related to the project. Also, it made version control simple incase an error occurred on a commit. As for Jupyter Notebook, it was used as our main tool/coding environment since it allowed for coding in block sections, which made it easier to manage and update our code.

iv. Results:

a. What answer was found to the research question; what did the study find?

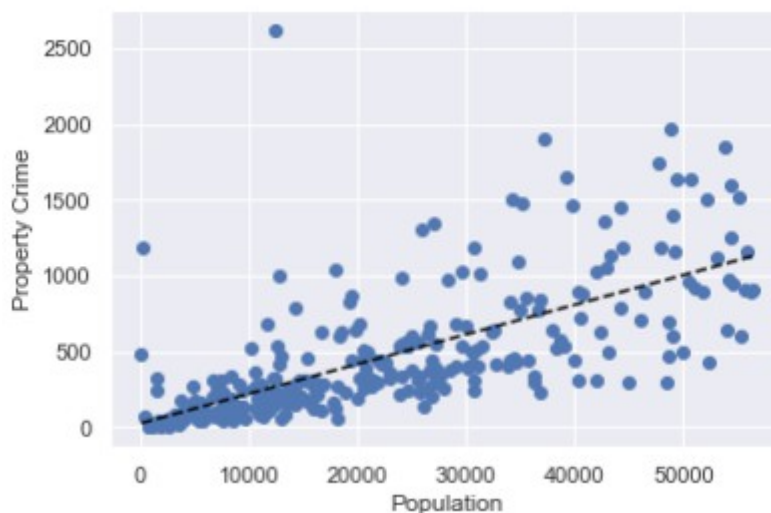
- a.i. The study found that property crime increases linearly in proportion to the size of the population and other various crimes.

b. Was the tested hypothesis true?

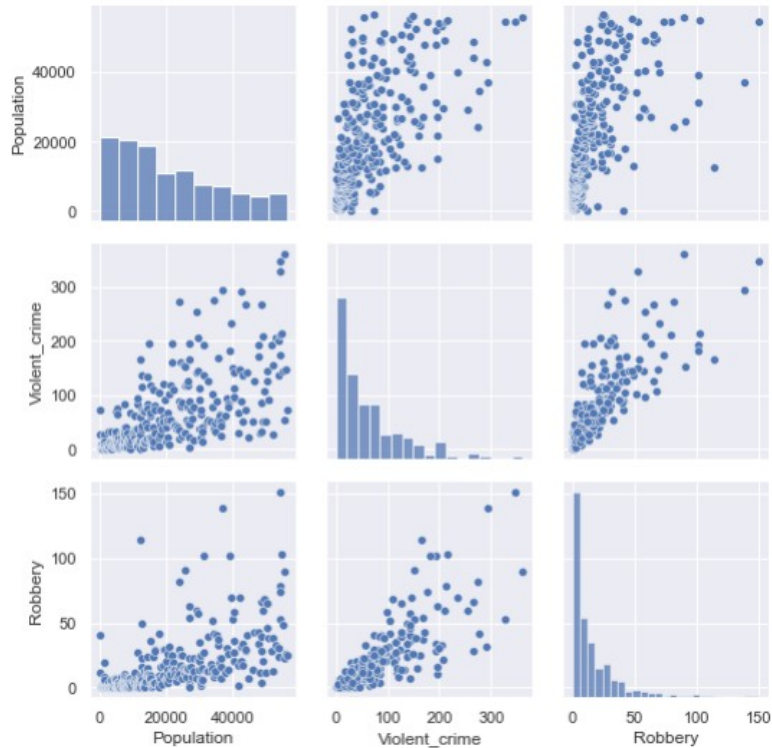
- b.i. Yes, our tested hypothesis was true in the end. We were able to successfully predict the property crime based on population and crimes within a city.

c. Any visualizations?

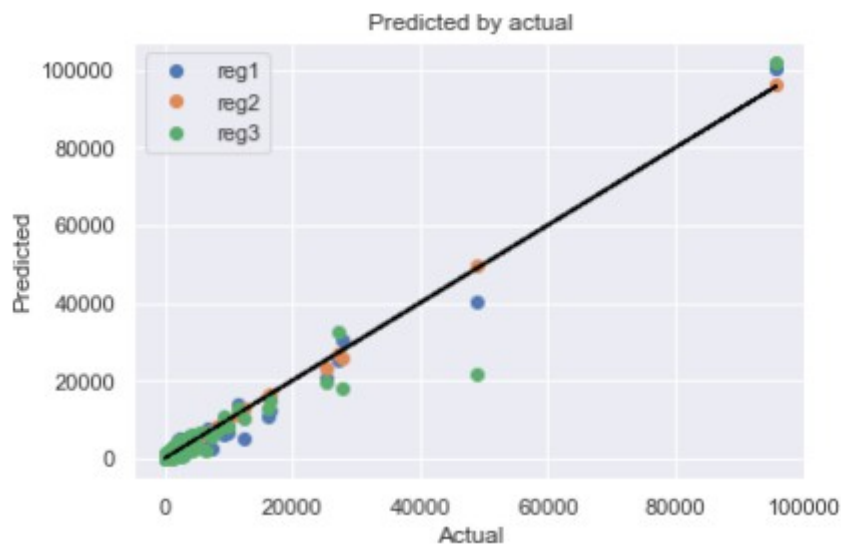
- c.i. Here are the visualizations we generated based on our hypothesis and data set. In each of these graphs' property crime is on the x-axis while our predicted, which consists of population, robbery, and violent crime, is on the y-axis. There are also other visualizations in our project, however, these showed off our findings the best.



Graph #1:



Graph #2:



Graph #3:

v. Discussion:

a. What might the answer imply and why does it matter?

- a.i. Based on what we have found, the higher the population size of a city, the more property crime occurs. This is very important as it shows that police

officers have a harder time preventing this type of crime within a large populace city.

b. How does it fit in with what other researchers have found?

- b.i. Our findings accurately fit what other researchers have found about property crime in California. When compared to violent crime, that averages 379.4 crimes per 100,000 people, property crime fluctuates around 2,109.9 crimes per 100,000 people (Gramlich, 2020).

c. What are the perspectives for future research?

- c.i. Some perspectives of future research should include more variables when collecting information rather than just the city, population size, and the type of crime. Some of these new variables could include gender, occupation, ethnicity, religion, and much more.

d. Survey about the tools investigated for this assignment.

- d.i. The tools that were used to investigate the data in this project was from two years ago. Even though it was from a reputable source like the FBI, having data that is more up to date will always be better. For example, having the UCR data from the previous year will provide more accurate data and will be more beneficial to those who view it.

vi. Summary:

a. Most important findings.

- a.i. One of the most important findings within our project is that the more variables that were used to predict the property crime, the more accurate it became. Another interesting discovery is how cities with population sizes large than 40,000 tend to have a 50-50 split between the linear model (slope intercept line). Also, cities with populace sizes lower than 30,000 are mostly below that linear model line.

References

Gramlich, J. (2020, November 23). *What the data says (and doesn't say) about crime in the United States*. Pew Research Center. Retrieved February 24, 2022, from <https://www.pewresearch.org/fact-tank/2020/11/20/facts-about-crime-in-the-u-s/>