

CS5830/6830 - Project 2

Crime Prevention

Elise Francis, Preethi Tera

GitHub: [Git Repository](#)

Presentation: [CS5830-6830-Project2](#)

Introduction

This project is mostly based on the Austin Crime Report 2015 dataset as well as the population densities for Austin, Texas. Our analysis mainly focuses on the number of crimes occurring in an area by considering its zip code and location. We analyzed the factors affecting the crimes in those areas by considering the unemployment rate, median household income, as well as location. We employ scatterplots, a linear regression plot, and a bar chart to visualize our analysis. The result of the analysis provides us with more information about the crime around Austin and with the help of this analysis we can have a clear idea for any future crime prevention efforts as we determine what factors contribute to crime.

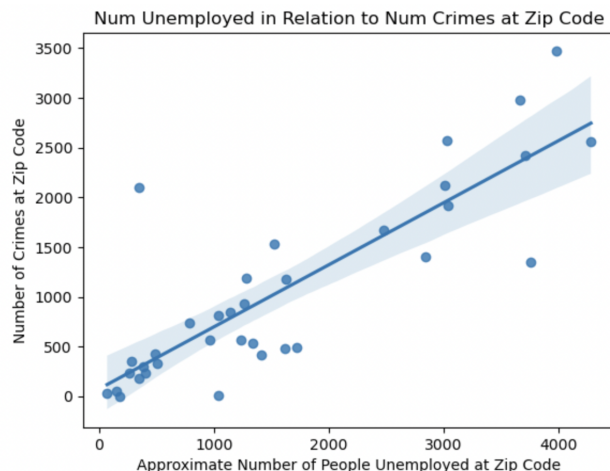
Dataset

There are two datasets used for the analysis in this project. One of the datasets contains the crime and housing statistics for Austin, Texas in 2015. This dataset included important attributes for analysis such as zip codes and median household income associated with the zip code. The other dataset used contains the population densities for Austin, Texas. This dataset contained the zip code attribute which can be used to correlate the information with the matching zip code using the previous dataset. Other important attributes for analysis included coordinating location values and population.

Analysis Technique

In this analysis, we used a bar chart, a regression plot, and two scatterplots. For our first analysis, we used a regression plot to describe the correlation between the number of crime occurrences and the unemployment rate in a specific region. For the second analysis, we used a scatter plot to display the median household income versus zip code. For the third analysis, we used the bar graph to display the relationship between crime rate and median income. In the final analysis, we used a scatter plot to present the information on the number of crimes committed in neighboring areas.

Result



A regression plot was made to show the correlation between the number of unemployed individuals and the number of crimes committed in that zip code. The positive coefficient signifies that there is a relationship between the number of crimes and the

number of unemployed individuals in the same zip code.

The result of the Pearson Correlation for the Number of Unemployed Individuals in Relation to the Number of Crimes in the Zip Code is as follows:

- $r = 0.8456052900146014$
- $p = 1.626042667544976e-10$

The p-value is less than 5%, therefore this analysis is statistically significant.

The result of the T-Test for the Number of Unemployed Individuals in Relation to the Number of Crimes in the Zip Code is as follows:

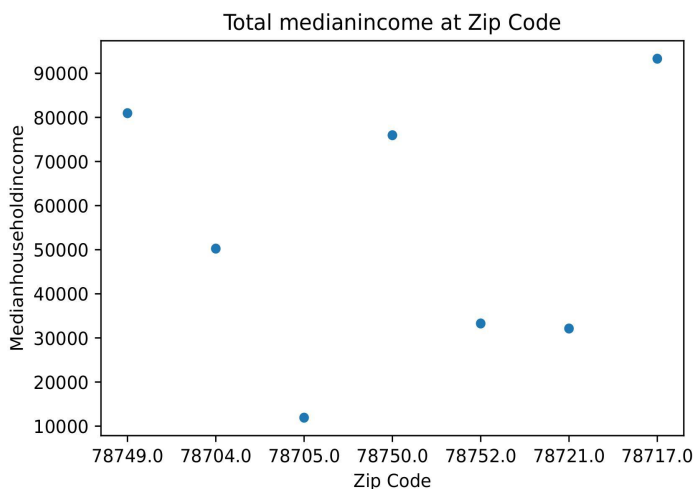
- statistic = 1.9350765687296634
- pvalue = 0.05714123929706782

For our analysis, we got a p-value of 0.057 which is close enough to the ideal value of 0.05. This means the relationship between unemployment and the number of crimes is not strongly significant, i.e we do not have any proof that both of them are co-related to each other.

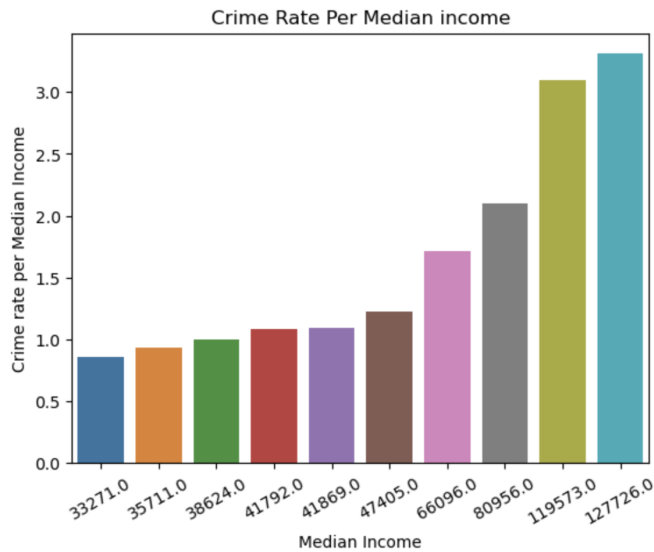
The Standard Deviation for the Number of Unemployed Individuals in Relation to the Number of Crimes in the Zip Code is as follows:

- Percentage Unemployed: 0.030131
- Num Crimes at Zip Code: 941.781734
- Population: 13843.430232
- Approx Num Unemployed: 1276.795643

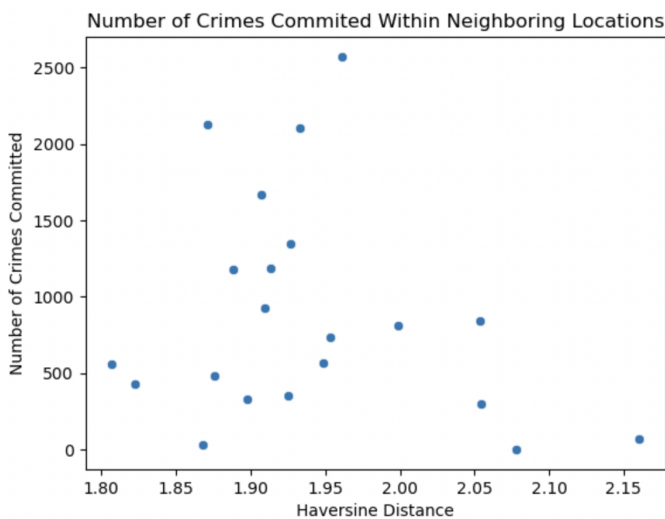
Standard deviation mainly gives us information about how the data is clustered together. A lower standard deviation indicates that data is clustered and the values in the data are closer to the mean whereas a higher standard deviation indicates that data is widely spread and are not clustered around the mean value.



In this analysis, we used a scatter plot to visualize the relationship between total median income in particular areas with respect to their zip codes. In doing this analysis, we are able to see the income in relation to the area.



In this analysis, we used a bar chart to show the relationship between crime rate per median income. It gives us a detailed analysis of the crime rate with respect to the median income. As the population increases, the median income increases. The increase in median income can also be in relation to the wealth of a particular area. From this analysis, we can see as the median income increases, crime increases. This may be because of an increase in thefts or population.



Another analysis we were interested in was seeing the relationship between the number of crimes committed in a particular area. The location at 30.239436, -97.671420 had the nearest neighbors. The haversine distance is the great-circle distance between points. As the locations get further away from the location at 30.239436, -97.671420 the number of crimes committed appears to decrease. There are more crimes committed in the haversine distance of 1.85-2.00 from the original location. Further investigation as to what is in the area or the types of crimes committed may provide further information.

Technical

Our analyses in this project made use of several statistical methods. The Pearson Correlation was used to determine how statistically significant the relationship is between the number of crimes committed and the unemployment rate. The standard deviation and T-Test methods were also used for this analysis to determine how statistically significant our results were. The regression plot allowed us to visualize the positive relationship between the unemployment rate and the number of crimes committed. The bar chart allowed us to visualize the relationship between crime rate and median income. The scatterplot allowed us to see some clustering of crimes within a certain location.

Our overall goal of this project was to determine what factors affect crime such as unemployment rate, median income, and location. From this analysis, we were able to see that the unemployment rate and median

income are contributing factors to crime. This may be because individuals are trying to survive economically and these factors may lead to more crime in a particular location.