

ABSTRACT

Impact of various economic metrics on crime rates in countries in New York and Maryland
Super Awesome Data Scientists Team, cschliem mpohlen1 sshulma3

Goal

The goal of this project is to analyze the impact of gentrification on crime rates across different regions. By leveraging data science techniques, we aim to understand how socio-economic changes associated with gentrification influence crime dynamics in urban areas.

Data

The dataset we use for this project has 29 columns and 1932 rows. It consists of crime statistics and housing value information broken down by county across the last 20 years. Our data the consists of the following columns: County/Region, Year, Population, Index Total Count, Index Total Rate, Violent Total Count, Violent Total Rate, Murder Count, Murder Rate, Rape Count, Rape Rate, Robbery Count, Robbery Rate, Agg Assault Count, Agg Assault Rate, Property Total Count, Property Total Rate, Burglary Count, Burglary Rate, Larceny Count, Larceny Rate, MV Theft Count, MV Theft Rate, State, Average Value, FIPS_Code, Employed, Unemployed, Unemployment_rate

Results and Analysis

Claim #1: The unemployment rate and the average housing price of a given county are not independent.

Support for Claim #1: Using a chi square test to test for independence we reject the first null hypothesis since the p-value is $4.376e-42$, which is less than 0.05. This means we cannot conclude that average housing prices and unemployment rates are independent. This result makes sense because it is logical for these two variables to have some correlation in any given county.

Claim #2: The observed difference in aggressive assault rates between New York and Maryland is statistically significant.

Support for Claim #2: By using two sample p-tests in order to test this hypothesis we discovered that the P-value is $1.3305636493005464e-50$, giving us very strong evidence to reject the H_0 . Therefore, there is a significant difference in the mean murder rates between the two states. Again, this intuitively makes sense as NY contains the city of New York which contributes significantly to its higher crime rates.

Claim #3: There is a significant difference in average unemployment rates between the two states.

Support for Claim #3: the p-value is 0.00126390 which is less than 0.05. So we reject the null hypothesis and conclude that there is a significant difference in the average unemployment rates between the two states. This makes sense intuitively as NY is generally considered a more prosperous and economically active state than MD.

Claim #4: There is no meaningful correlation between violent crime rate and average home value.

Support for Claim #4: Using a linear regression model we found that analysis between average home value and violent crime rate yielded a coefficient of 0.0002617, indicating a positive but small relationship—suggesting that as average home value increases, so does the violent crime rate, though minimally. The low R-squared value of 0.027 suggests the model explains only 2.72% of the variation in violent crime rates, indicating limitations in its predictive power. Ultimately linear regression failed to find much meaningful correlation between average home value and violent crime rates which looks to disprove our question about whether the two are related in any way.

Claim #5: We are able to predict unemployment rates based on our data on a better than average rate

Support for Claim #5: The SVM classifier had an overall accuracy of 48%, which is higher than chance (25%) when predicting one of four classes. The model is better at recall for ‘very high’ and ‘low’ unemployment rates, with recall rates of 66% and 64% respectively. This accuracy discrepancy may be due to unequally distributed data between the four classes or potentially crime rate statistics are just more predictive of extreme values (very high and low) rather than mild values (medium and high).