Crime, Housing, and Employment in New York and Maryland

Sam Shulman, Martin Pohlen, Caspar Schliemann

BROWN

Background

We aim to investigate the relationship between **crime rates and key economic factors** such as unemployment and housing prices in the states of **New York and Maryland.** By delving into our data and employing advanced data science techniques, our project is seeking to analyze the relationship between socio-economic factors and crime data and how this data differs in New York and Maryland. This exploration not only contributes to our understanding of the underlying drivers of crime but also informs evidence-based strategies for promoting community safety and economic well-being.

Hypothesis / Problem Statement

To understand the relationship between crime and economic data across NY and Mayland, we looked at these hypotheses:

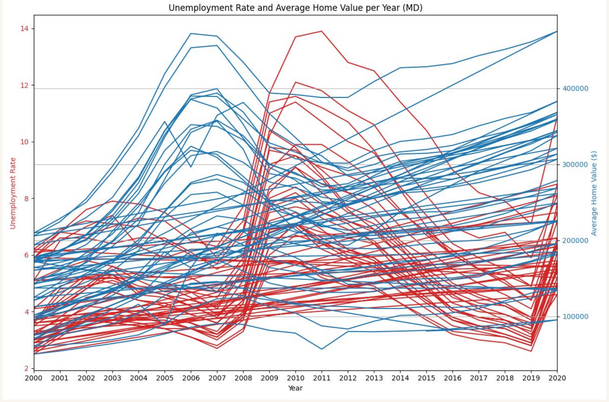
1. The unemployment rate and the average housing price of a given county are independent
2. There is a significant difference in the aggressive assault rates between New York and Maryland
3. There is a significant difference in the unemployment rate between New York and Maryland

Data & Sources

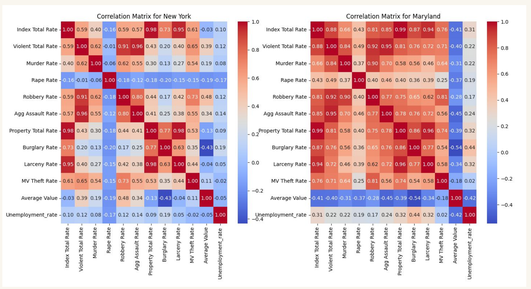
To build our dataset, we had to combine tables of crime statistics and economic indicators from different sources. As we were looking at two different states for crime statistics by county we went to each states respective crime databases and retrieved tables from the last 20 years. After cleaning the data and combining it, we found average home values through Zillow and unemployment rates through another government website and added that to our database. The table we were left with not only contains total crime rates but also breaks it down by type of crime allowing us to analyze each crime type individually with various economic factors.

Hypotheses, Results, and Visualizations

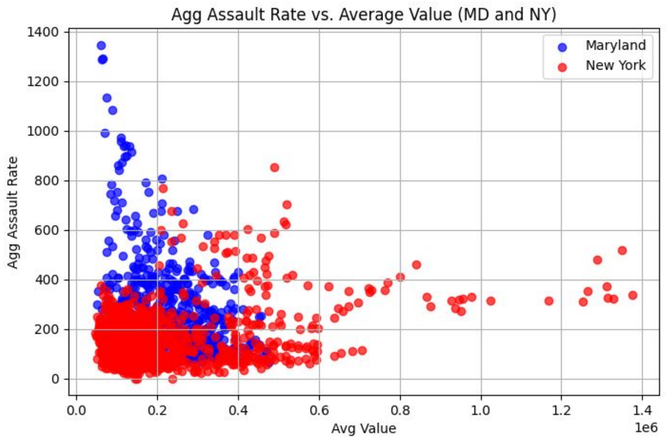
There is statistically significant evidence to support that unemployment rate and the average housing of a given county are not independent. (p-value = 4.38e-42)



Correlation matrix to see which variables in the dataset have a correlation.



The difference in aggressive assault rates between New York and Maryland is statistically significant. (p-value = 1.33e-50)



The difference in average home values between New York and Maryland is statistically significant. (p-value = 3.83e-06)

SVM classifier predicting unemployment rates (low, medium, high, very high) based on crime data:

Best at recall for ‘very high’ and ‘low’ unemployment rates, with recall rates of 66% and 64% respectively

Classification Report:	precision	recall
high	0.43	0.23
low	0.55	0.64
medium	0.39	0.31
very high	0.47	0.66

The linear regression analysis between average home value and violent crime rate yielded coefficient of 0.0002617, indicating a positive but very minimal relationship.

Methodology

- **Statistical tests (with 0.05 significance level):**
  - Chi-squared test to examine independence of unemployment and violent crime rates
  - Two-sample t-test to examine disparities in murder rates and average home values
- **Machine learning techniques:**
  - Support Vector Machines (SVM) for classification of unemployment rates based on crime rates
  - Linear regression to predict average home values based on violent crime rates

Limitations/Challenges

A key limitation of our methodology is that we did not consider how variations in policing strategies and sentencing guidelines influence reported crime statistics. For example, we did not account for factors such as differential policing intensity, variations in reporting practices, and the impact of predictive policing technologies. NYC in particular tends to follow a “tough on crime” approach, which may skew our data in comparison with Maryland.

Significance

Government agencies responsible for economic development and public safety, could benefit from our findings of this project. The results could be significant for these policymakers to formulate to better decisions regarding policy formulation aimed at addressing unemployment, crime, income disparities, and housing issues. Our data-driven approach could be utilised to allocate resources in NY and Maryland more efficiently and implement better programs to address the needs of vulnerable communities.

Sources

https://www.census.gov/library/visualizations/time-series/demo/census-poverty-tool.html https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas https://www.nyc.gov/site/nypd/stats/crime-statistics/historical.page https://data.ers.usda.gov/reports.aspx?ID=17829#Pf7081e5b1f514b5fb1a0027a9fe44848_5_47iT1

https://data.ers.usda.gov/reports.aspx?ID=17828 https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/ https://www.zillow.com/research/data/

Acknowledgements

Huge thank you to our wonderful TA Omar!