Analyzing Socioeconomic and Environmental Factors Affecting Crime Patterns in Tucson

Imronbek Abduvaliev

Javokhir Kakhkhorov

Tokhirjon Vokhidov

University of Arizona

CSC380 Principles of Data Science

Cesim Erten

12/10/24

Introduction

Crime significantly impacts the social and economic fabric of communities, influencing both the quality of life and the distribution of resources. Understanding the factors associated with crime occurrences is therefore critical for developing effective prevention strategies. In this project, we explore the relationship between environmental and socioeconomic variables—such as income levels and streetlight presence—and the frequency and types of crimes reported in Tucson, Arizona. By integrating multiple publicly available datasets and performing advanced analyses, we aim to identify patterns and correlations that can guide public policy, inform resource allocation, and ultimately help reduce crime rates in the community.

Objective

Our primary research question focuses on identifying the key factors associated with crime frequency in Tucson. To complement this main inquiry, we also consider several secondary questions. Are certain types of crimes more prevalent in specific neighborhoods? What roles do socioeconomic factors, like income levels, play in influencing crime rates? How does the presence of streetlights and other forms of public infrastructure correlate with criminal activity? By investigating these aspects, we can shed light on complex and interrelated influences on crime, offering a foundation for data-driven strategies to improve community safety.

Data Overview

To address these questions, we have integrated six core datasets that cover a range of relevant dimensions:

- Neighborhood_Income.csv: Socioeconomic data, including income levels across neighborhoods, providing insight into the correlation between economic factors and crime rates.
- Streetlights_-City_of_Tucson-_Open_Data.csv: Details about the distribution and density of streetlights within Tucson, offering a lens into how infrastructure may influence safety.
- Tucson_Police_Reported_Crimes.csv: Reported crime data categorized by type, location, and time, forming the centerpiece of our analysis.
- City_of_Tucson_Ward_Boundaries.csv & City_of_Tucson_Ward_Boundaries.geojson: Geographical and administrative boundaries to map crimes and link them with infrastructure and socioeconomic indicators at the ward level.

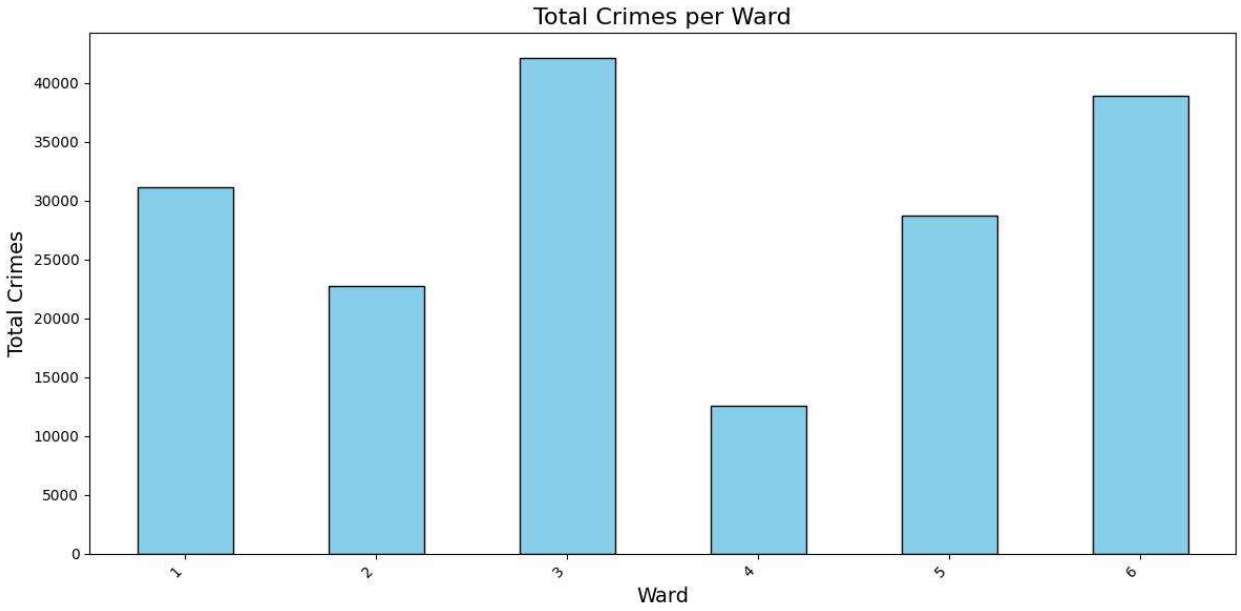
By combining these datasets, we are able to examine crime frequencies, types, and their associations with both structural and social factors.

Data Visualizations

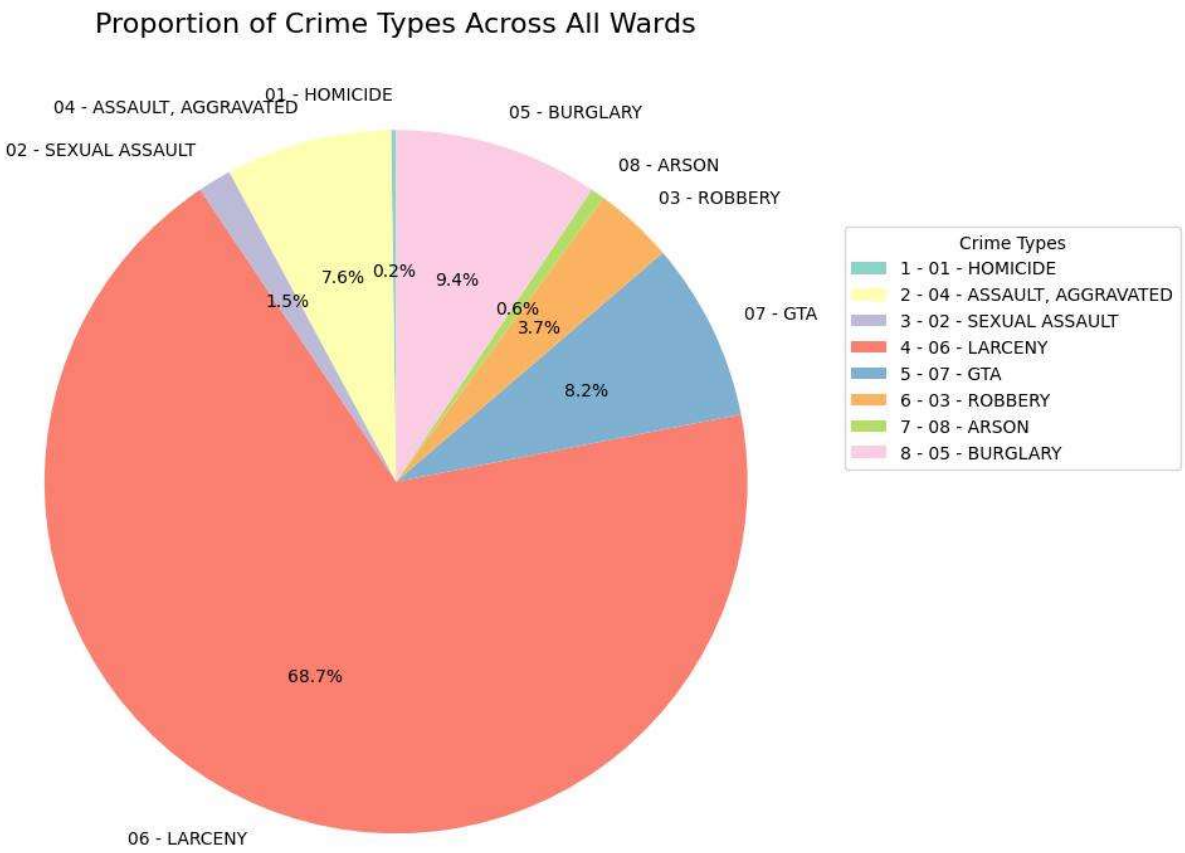
To bring our analysis to life, we employed a range of graphs and charts. These visual tools present key trends, highlight differences among wards, and help us identify relationships between crime and various predictors. For instance, we compare crime frequencies and types across Tucson’s wards, analyze how streetlight presence might correlate with crime levels, and

investigate whether income levels play a role in shaping criminal activity. Through these visual representations, patterns and potential correlations become clearer, enabling more informed discussions about policy interventions and resource allocation.

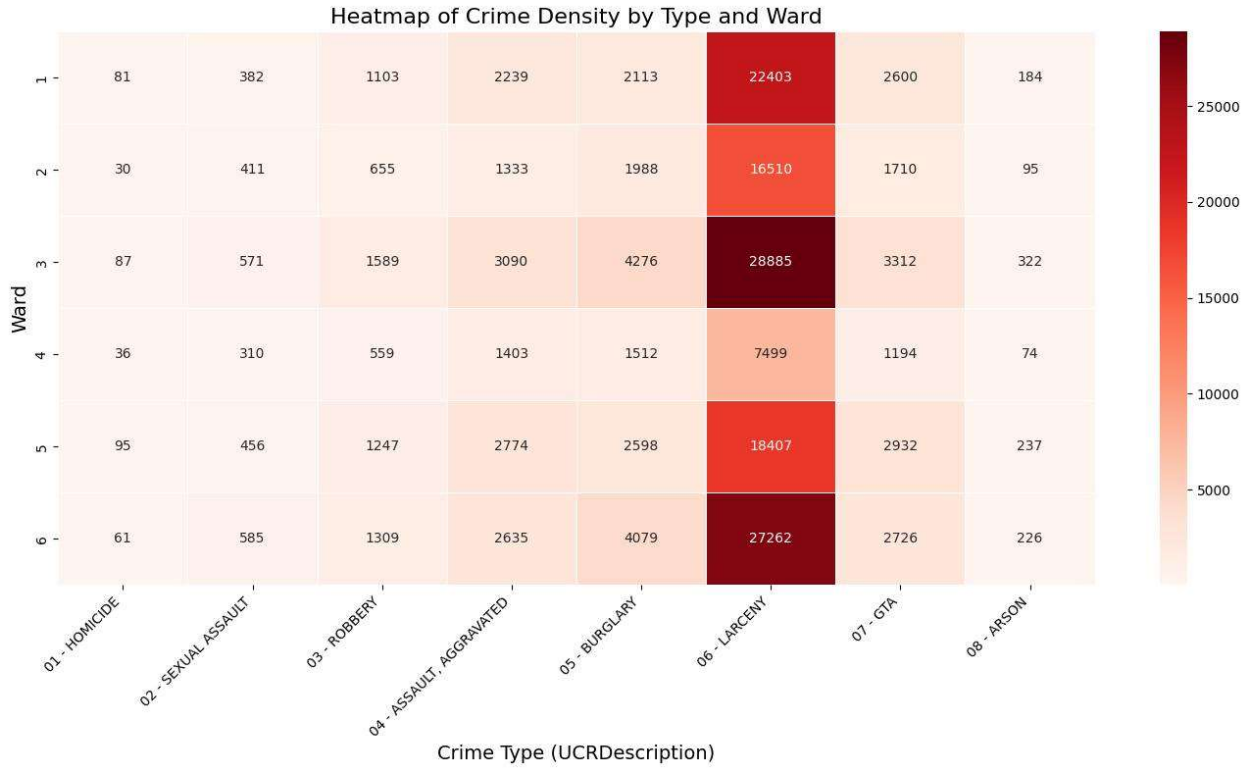
Crime Patterns and Distribution in Tucson



The bar chart illustrates the total number of crimes reported across different wards in Tucson, providing a clear view of crime distribution across the city. Ward 3 and Ward 6 have the highest crime totals, identifying them as significant hotspots for criminal activity, which may require prioritized intervention and resource allocation. Ward 1 and Ward 5 also exhibit notable crime levels, although they are lower compared to the most affected wards. In contrast, Ward 4 stands out with the lowest crime rate, reflecting a comparatively safer environment. This disparity highlights the uneven distribution of criminal activity across Tucson, emphasizing the need for targeted strategies to address high-crime areas while understanding the factors contributing to safety in lower-crime wards.



Delving deeper into the nature of these crimes, a pie chart illustrates the distribution of crime types across all wards in Tucson. Larceny overwhelmingly dominates, accounting for 68.7% of all crimes, indicating it is the most common issue citywide. Burglary, GTA (grand theft auto), and robbery collectively form a significant portion of the remaining crimes, while violent crimes such as homicide, sexual assault, and aggravated assault represent smaller proportions. Arson contributes the least to the overall crime distribution. This breakdown underscores the need for focused strategies to mitigate property crimes, particularly larceny, which constitutes the majority of criminal activity in Tucson.



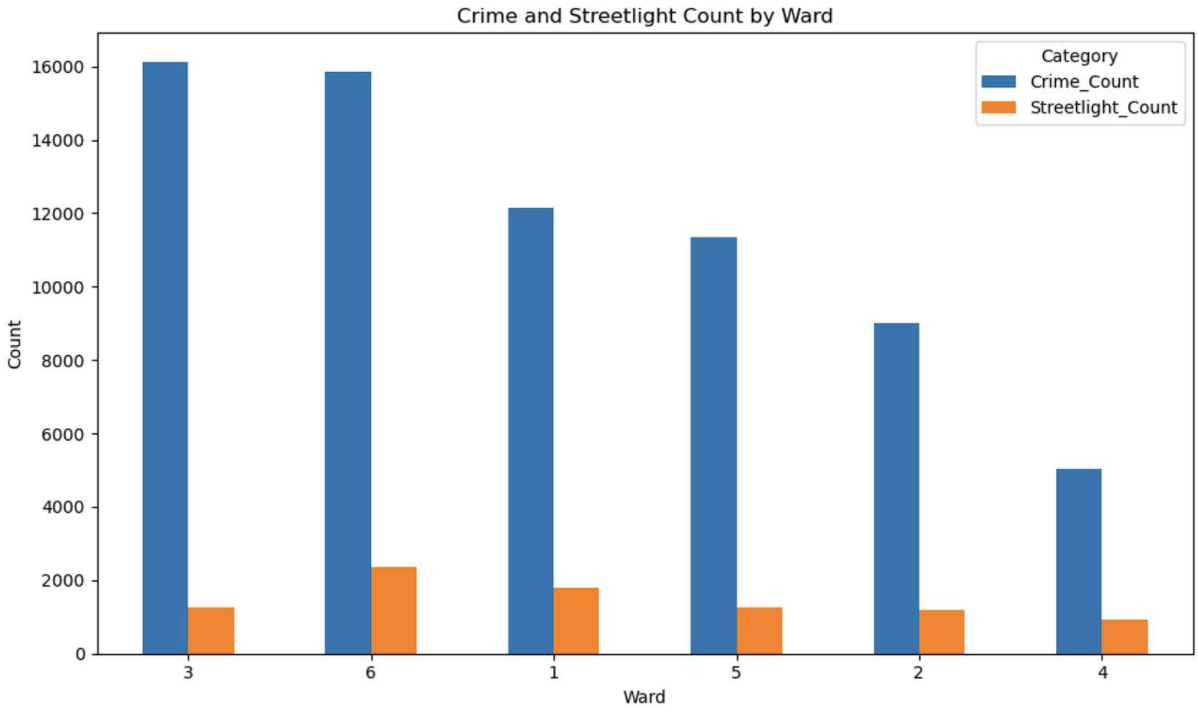
A subsequent heatmap highlights significant variations in crime density by type and ward in Tucson. Larceny emerges as the most prevalent crime across all wards, with Ward 3 and Ward 6 showing the highest concentrations, particularly for property crimes like burglary and aggravated assault. In contrast, homicide and arson remain relatively infrequent citywide. Ward 3 consistently exhibits the highest overall crime density, while Ward 4 stands out as the safest with the lowest crime rates. These patterns suggest the need for targeted interventions in high-crime areas, particularly to address property crimes, while leveraging the factors contributing to safety in lower-crime wards as a model for broader crime prevention strategies.

Conclusion for Crime Distribution

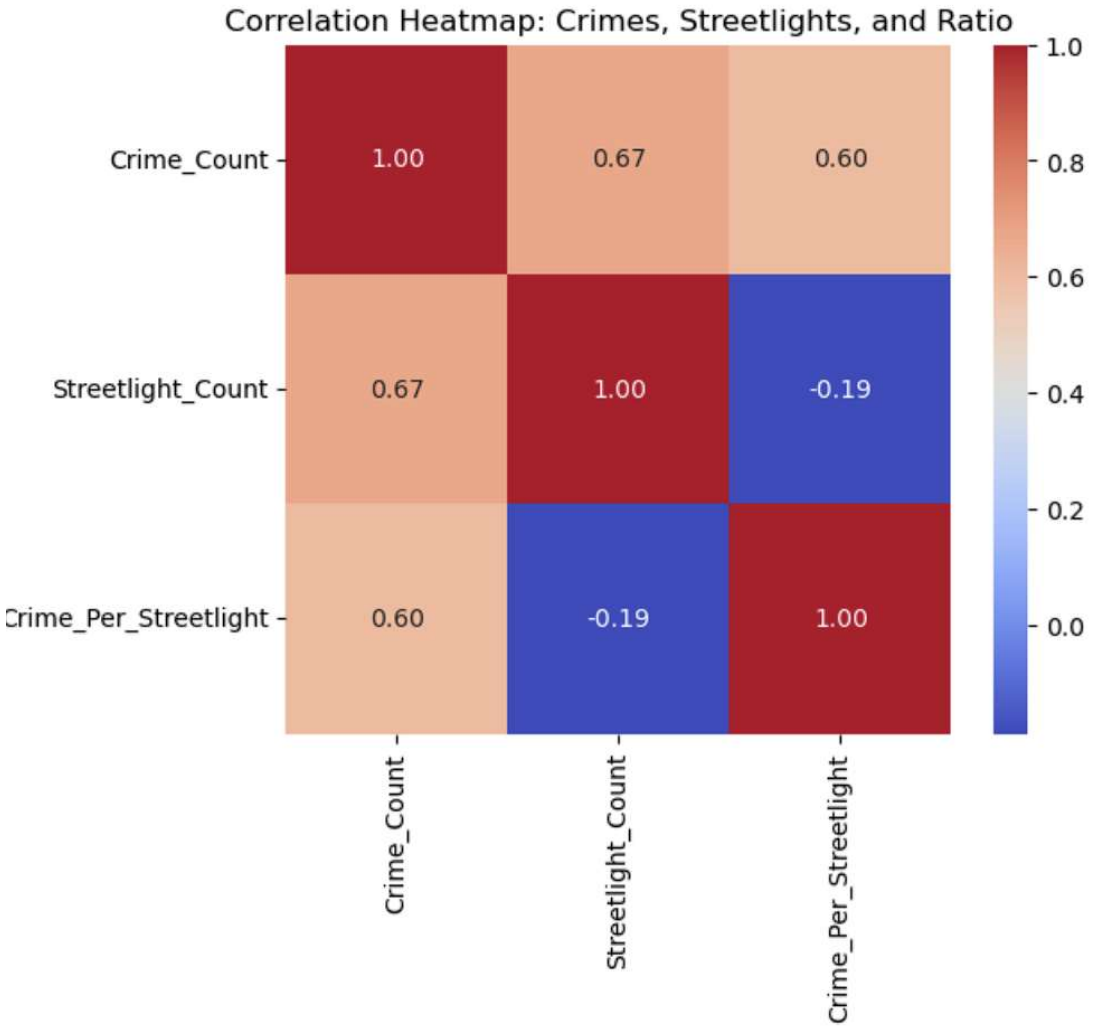
These findings suggest that certain wards—particularly Wards 3 and 6—experience disproportionately high crime rates, and that property crimes dominate Tucson's overall crime profile. Addressing these hotspots through targeted interventions may help reduce citywide crime. Additionally, understanding why Ward 4 maintains relatively low crime levels could offer lessons for broader crime prevention strategies elsewhere in the city.

Crimes and Streetlights Patterns

Next, we examined the relationship between crime frequencies and the presence of streetlights.



The bar chart compares the count of crimes and streetlights across wards in Tucson, highlighting disparities in public infrastructure relative to crime levels. Wards 3 and 6 show the highest crime counts, with over 16,000 incidents each, yet their streetlight counts are relatively modest compared to the crime levels. This contrast raises questions about the adequacy of streetlighting in mitigating crime in these high-incident areas. Conversely, Ward 4, which has the lowest crime count, also exhibits a proportionally lower number of streetlights, suggesting that factors beyond streetlighting may contribute to its lower crime rate. The wards with moderate crime counts, such as Wards 1 and 5, also display modest streetlight coverage, indicating a potential correlation between infrastructure and safety. These observations underscore the need for further investigation into the relationship between streetlighting and crime prevention, particularly in high-crime wards like 3 and 6, where targeted improvements in public infrastructure could play a critical role in enhancing safety.

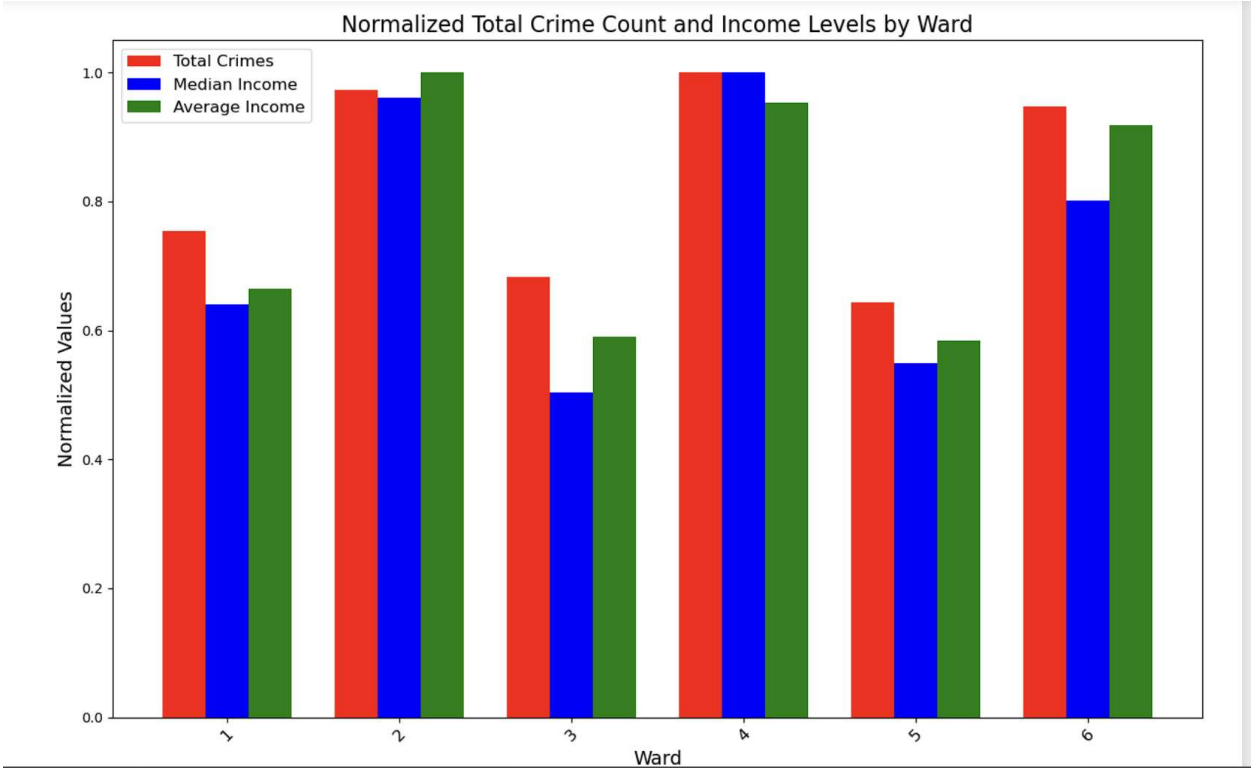


To probe these relationships further, we created correlation heatmap which illustrates the relationship between crime counts, streetlight counts, and the ratio of crimes per streetlight across Tucson's wards. A moderately strong positive correlation (0.67) is observed between crime count and streetlight count, suggesting that wards with more streetlights tend to report higher crime incidents, likely due to these areas being more urbanized or densely populated. However, the negative correlation (-0.19) between streetlight count and crimes per streetlight indicates that a higher number of streetlights might reduce the crime burden per light, potentially reflecting the deterrent effect of improved lighting. Additionally, the positive correlation (0.60) between total crimes and the ratio of crimes per streetlight suggests that areas with high crime counts also experience a higher concentration of crimes relative to their streetlight coverage, underscoring the inadequacy of lighting in mitigating crime effectively in these regions. This analysis highlights the complex interplay between infrastructure and crime, emphasizing the need for strategic placement and optimization of streetlighting in high-crime wards to enhance safety and reduce criminal activity.

Conclusion for Streetlight Analysis

These results highlight the complexity of infrastructure's role in crime prevention. While more urbanized areas tend to have both higher crime and more streetlights, improved lighting alone may not guarantee lower crime rates. Careful placement and possibly pairing streetlight improvements with other interventions—such as community policing or economic development initiatives—may be necessary to effectively enhance public safety.

Crime and Income Pattern



Our analysis also explored the relationship between crime levels and socioeconomic factors. A bar chart comparing normalized total crime counts with median and average income levels reveals that wards with higher crime counts, such as Wards 3 and 6, generally have lower median and average incomes. Conversely, Ward 4, with its lower crime rate, enjoys higher income levels. This pattern supports the idea that economic disadvantage may contribute to elevated crime rates, while greater financial stability could foster safer communities.

However, the relationship between income and crime is not uniformly straightforward. For instance, Ward 1 shows a relatively balanced dynamic, where income levels do not clearly predict the crime count. These nuances suggest that while socioeconomic indicators are important, they do not operate in isolation. Other factors, such as population density, urban design, cultural elements, community engagement, and public infrastructure quality, may also influence crime patterns.

Conclusion for Income Analysis

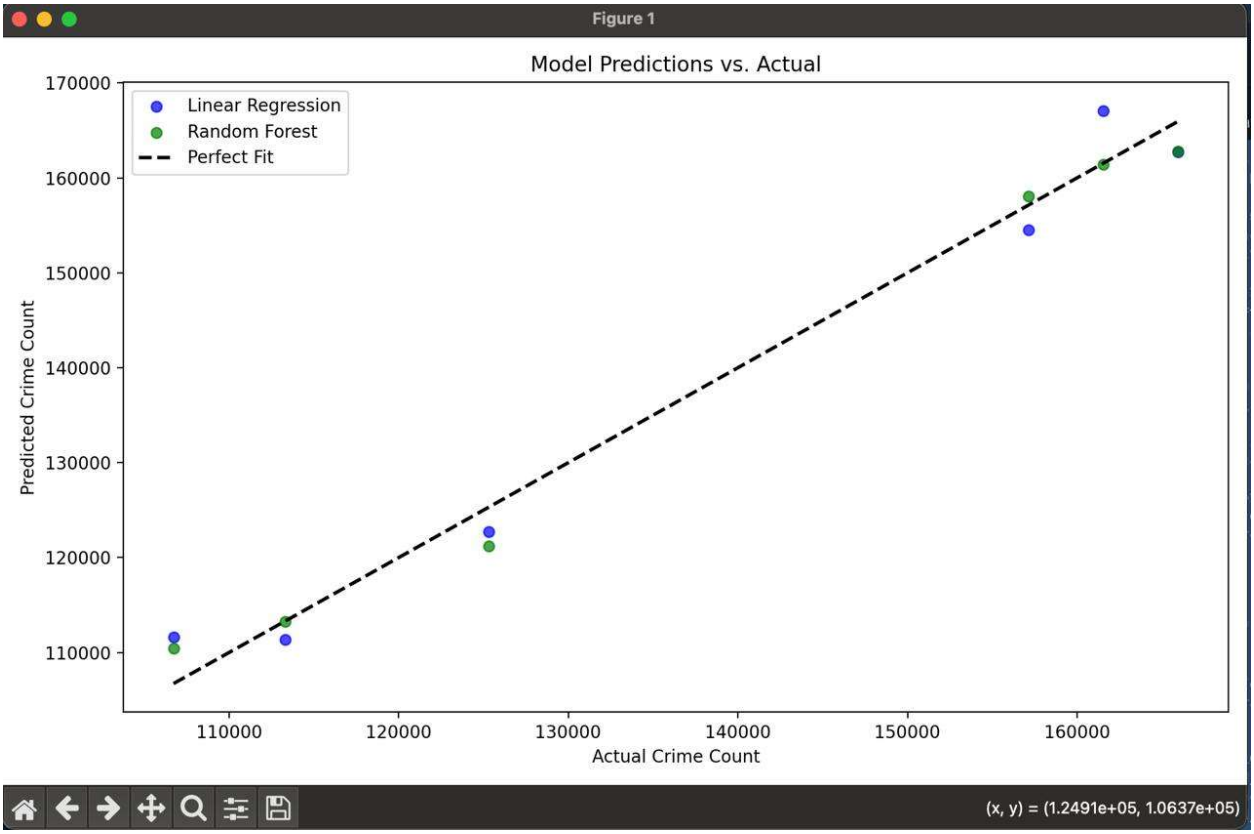
The correlation between income and crime suggests that addressing economic disparities should be part of any holistic crime reduction strategy. Policymakers might consider targeted socioeconomic interventions—such as job training, educational opportunities, and community development programs—in wards struggling with both low income and high crime. While income is not the sole determinant of crime rates, investing in economic upliftment could complement other crime prevention measures, ultimately reducing criminal activity and improving the overall quality of life.

Modeling, Prediction and Evaluation

Modeling, Prediction, and Evaluation

To further understand the predictive power of socioeconomic indicators on crime frequencies, we employed two machine learning models—Linear Regression and a Random Forest Regressor. The Linear Regression model served as a baseline, capturing direct linear relationships, while the Random Forest model offered a more complex, non-linear approach that can better handle intricate interactions between variables.

We carried out cross-validation, tuning, and comparative analysis to ensure robust and reliable predictions. The Random Forest model, after hyperparameter optimization, performed slightly better than the Linear Regression model in predicting crime counts, as indicated by its higher R^2 and lower Mean Squared Error (MSE) values. The scatter plot comparing predicted versus actual crime counts shows that both models track closely with the observed data, though the Random Forest is somewhat more accurate, especially at higher crime levels.



Output:
Linear Regression - Cross-validated R^2: 0.82 ± 0.09

Linear Regression - Cross-validated MSE: 92362722.35

Linear Regression - Final MSE: 13841243.97, Final R^2: 0.98

Fitting 3 folds for each of 18 candidates, totalling 54 fits

Random Forest - Cross-validated R^2: 0.87 ± 0.10

Random Forest - Cross-validated MSE: 66834073.16

Random Forest - Final MSE: 6851218.75, Final R^2: 0.99

Best Random Forest Parameters: {'max_depth': None, 'min_samples_split': 2, 'n_estimators': 100}

The scatter plot compares the predicted crime counts from two models—Linear Regression and Random Forest—against the actual crime counts, with a dashed line representing a perfect fit. Both models demonstrate strong predictive performance, as most data points lie close to the perfect fit line. The Random Forest model (green points) appears slightly more accurate for certain data points, staying closer to the ideal fit than the Linear Regression model (blue points), which shows more deviation at higher crime counts. This indicates that the Random Forest model may capture nonlinear relationships in the data more effectively, making it

better suited for predicting complex patterns in crime counts. However, the overall alignment of both models with the perfect fit line suggests that either can be useful depending on the level of precision required. This analysis highlights the effectiveness of using machine learning techniques to model and predict crime trends, enabling data-driven decision-making for crime prevention and resource allocation.

Conclusion for Modeling and Prediction

Our modeling results suggest that income indicators can meaningfully predict crime occurrences, and that more sophisticated models like Random Forests may capture the complex, non-linear nature of these relationships more effectively. Policymakers and law enforcement agencies could leverage these predictive insights to identify high-risk areas before crime spikes occur.

At the end, we came to conclusion that integrating crime, infrastructure, and socioeconomic data from Tucson allowed us to identify patterns and correlations that can inform more targeted and effective crime prevention strategies. While high-crime wards like Wards 3 and 6 struggle with property offenses and may benefit from a combination of infrastructure improvements and economic development initiatives, lower-crime wards like Ward 4 may serve as models for best practices. Streetlights alone do not appear to guarantee lower crime, but careful urban planning and community-focused interventions could make them more effective. The consistent link between socioeconomic disadvantage and higher crime rates emphasizes the need for policies that address root causes of criminal behavior, such as poverty and economic instability.

By employing machine learning models, we confirmed that socioeconomic indicators can predict crime trends with high accuracy. These predictive capabilities enhance our ability to proactively allocate resources and implement interventions in a timely manner. Taken together, this integrated approach—rooted in data-driven analysis—provides a solid foundation for policymakers, law enforcement, and community stakeholders to collaborate on reducing crime and building safer, more equitable neighborhoods across Tucson.