Do changes in residential stability predict changes in burglary rates?

Caleb Collins, Hiza Mvuendy, Meghana Sathi, and Bethany Schweitzer

University of North Carolina at Charlotte

DTSC 3602: Predictive Analytics and Their Implications

Professor Nadia Najjar and Professor Shannon Reid

December 9, 2024

Abstract

Keywords: social disorganization, burglary, income, graduation rate, household size, location of community, family dynamics

Introduction

On average, someone in the United States is victimized by burglary every 30 seconds (FBI, 2019). The FBI categorizes burglary as "the unlawful entry of a structure to commit a felony or theft", and in 2018 it accounted for 17.1% of all property crimes nationwide (FBI, 2019, p.1). This crime not only disrupts individuals' lives but also leaves a huge financial impact on its victims. In 2019 alone the estimated losses for those who were victims of burglary were approximately 3.4 billion dollars (FBI, 2019). With the effects of the average burglary costing its victims $2,799, many families simply cannot financially afford to be burglarized without it causing them extreme financial strain (FBI, 2019; Chen 2019). Despite technological advancements, burglary seems to be persistent, especially affecting urban and rural communities. Furthermore, the psychological toll endured by victims of burglaries is both significant and long-lasting for many families (Beaton et al., 2008). For families and communities, the impact of burglary is not only about replacing stolen property but also about rebuilding the sense of security within their homes which is difficult to restore. Considering the impact of burglary on people, it is imperative to leverage research to find practical, evidence-based solutions that can focus on making a real change.

Understanding the damages caused by burglaries helps to determine what factors are common in communities with higher burglary rates, and how these factors can be mitigated to protect vulnerable communities. One credible theory that addresses these questions is Social Disorganization Theory. This theory posits that a community's social cohesion and the strength of its interpersonal ties directly impacts its susceptibility to crime (Bellair, 2017). Disadvantaged communities with high turnover, economic instability, and weakened social structures often correlate with higher crime rates, including burglary. In light of these observations, the continuation of unchecked burglary rates affects not only the financial stability of victims but also the social health and resilience of entire communities. Understanding the factors that contribute to an increase in burglary rates is important to develop effective prevention strategies. While reactive measures like technology or police monitoring can help, they are not long-term solutions. Without an understanding of the root of the problem, many communities will continue to be at risk.

Therefore, this project aims to explore how changes in residential stability predict changes in burglary rates, based on the social disorganization theory, with the hypothesis that increasing residential stability decreases burglary rates. By identifying turning points where changes in stability significantly impact burglary rates, this research focuses on building on top of the Social Disorganization Theory to contribute insights on how communities can strengthen their social environment and reduce their vulnerability to crime.

Background

The relationship between residential stability and crime is well-supported in criminological literature, particularly through Social Disorganization Theory, which emphasizes how weakened social cohesion in communities with high turnover increases vulnerability to crime. Shaw and McKay (1942) initially proposed that neighborhoods with strong interpersonal ties and a collective ability to regulate behavior are more resistant to crime. In contrast, communities characterized by frequent residential mobility and socio-economic instability often lack the trust and social networks necessary to deter criminal behavior. Bellair (2017) reinforced these concepts by demonstrating that residential mobility disrupts social cohesion, creating conditions where crime can flourish. Bellair's study found that residential instability directly correlates with weaker social ties, which in turn increases the likelihood of various types of crime, including property crimes like burglary.

Recent research has delved deeper into the protective role of long-term residency in fostering collective efficacy. Studies by Sampson et al. (1997) revealed that neighborhoods with higher percentages of long-term residents develop stronger mutual trust and shared responsibilities, reducing crime rates. They argue that long-term residents are more likely to care for their neighborhood, become involved in community activities, and help deter crime. In contrast, communities with a high turnover rate face challenges in forming these networks, leading to weakened informal social controls, which increases crime rates. This finding is supported by Sampson and Groves (1989), who argued that a lack of social networks and informal control within neighborhoods significantly contributes to the vulnerability of communities to crime. Additionally, Sampson and Raudenbush (1999) emphasized the role of social cohesion and how communities that lack it face elevated crime risks, particularly burglary, due to the inability of residents to collectively oversee public spaces.

Hipp (2010) added to this by finding that neighborhoods with high turnover and limited socio-economic resources experience compounded social fragmentation, leaving them more vulnerable to crimes like burglary. His research suggests that neighborhoods with greater mobility struggle to maintain shared norms and a sense of belonging, which erodes the informal mechanisms that prevent crime. Hipp's work illustrates that socio-economic status and mobility have a synergistic effect—when combined with poor economic conditions, residential instability exacerbates crime. He also found that communities with higher turnover rates and fewer resources were at a greater risk of crime because new residents were less likely to feel a sense of ownership or responsibility toward the community, weakening collective efficacy.

Despite these advances, the literature reveals several key gaps in understanding how residential stability impacts burglary specifically. While Social Disorganization Theory broadly explains the connection between community instability and crime, fewer studies explore specific predictors of burglary risk. Warner and Rountree (1997) identified that communities with strong family structures and higher neighborhood investment were more resilient to crime, yet their study didn’t delve deeply into how these factors might directly mitigate burglary. Kubrin and Weitzer (2003) further emphasized that economic investment and stable family structures help resist crime, but they also noted a gap in understanding exactly how changes in family dynamics and neighborhood investment influence specific types of crime, particularly property crime. They proposed that while stable families and investment foster resilience, these factors need to be understood in relation to specific criminal behaviors, like burglary, which can often be more opportunistic and less related to interpersonal crime.

Environmental factors such as housing conditions also remain underexplored in the context of residential stability and burglary. Spelman (1993) established a clear link between neglected housing, such as vacant or poorly maintained properties, and heightened crime rates. He found that vacant properties act as attractors for criminal behavior, providing opportunities for burglars, and that these conditions often signify the broader neglect of a neighborhood. Vacant houses and properties that are poorly maintained not only reduce the aesthetic value of a community but also signal a lack of oversight and investment, leading to increased crime rates. However, few studies integrate these environmental factors with the social dimensions of residential stability. Spelman’s findings suggest that housing conditions alone may increase burglary, but when combined with high residential turnover, the effects are even more pronounced, yet the social dynamics linking these two factors are not well-studied.

In their work, Sampson, Raudenbush, and Earls (1997) also explored how neighborhood physical characteristics, such as abandoned houses and poor infrastructure, contribute to higher crime rates. Their study suggests that the built environment is a powerful indicator of neighborhood disorder. Vacant homes, in particular, were seen as symbolic markers of urban decay, which leads to increased vulnerability to crimes like burglary. They concluded that environmental neglect directly contributes to crime escalation, as it sends the message that the community does not have the capacity to maintain its public spaces, making them easy targets for criminal activity.

Similarly, Eck and Clarke (2003) also focused on the relationship between physical neglect (such as abandoned properties) and burglary. They proposed that these neglected spaces provide opportunities for criminals, who are attracted to areas that appear weak or uncared for. Their work highlights the importance of environmental design and maintenance as part of the broader social fabric of a neighborhood. While many studies have examined these factors in relation to general crime trends, few have tied them directly to burglary, making it a crucial gap that this research aims to fill.

Moreover, Rosenfeld and Messner (2013) underscore the importance of understanding socio-economic factors in combination with housing conditions. They suggest that both neighborhood disorganization and economic distress contribute to heightened crime, particularly burglary. However, there is a gap in literature on the interaction between socio-economic stability, family structure, housing conditions, and burglary.

This research seeks to address several of the gaps in understanding how residential stability impacts burglary rates by integrating socio-economic drivers, environmental conditions, and predictive modeling into a comprehensive framework. First, it aims to fill the gap in identifying specific predictors of burglary risk rather than broadly examining crime trends, focusing on how factors such as family structure and neighborhood investment shape burglary patterns. These socio-economic variables are explored to clarify their direct and indirect impacts on community resilience against burglary, building on prior research that has largely overlooked their crime-specific implications.

Secondly, this study seeks to expand on the role of housing conditions, addressing a significant gap in how environmental factors interact with residential stability to influence crime. Neglected properties and poor housing maintenance have long been linked to crime risks, but their integration into stability-focused analyses remains sparse. By examining the interplay between physical and social dimensions of neighborhoods, this research provides a more nuanced understanding of the environmental triggers of burglary.

Finally, the incorporation of predictive modeling will fill a methodological gap, offering a robust, data-driven approach to validate and extend Social Disorganization Theory. Predictive modeling enables the analysis of large datasets to uncover hidden patterns and relationships, providing a deeper understanding of how changes in factors such as residential turnover, family structure, and housing conditions influence burglary rates. Additionally, predictive modeling allows for the identification of potential burglary hotspots before they occur, enabling community leaders to act proactively rather than reactively. This capability supports targeted interventions, such as community engagement programs or housing initiatives, to address risks before they escalate. Furthermore, predictive models are highly adaptable, capable of integrating additional variables or updated data to reflect evolving community dynamics. By bridging theoretical analysis with practical applications, predictive modeling supports the development of proactive, evidence-based strategies that address both immediate crime risks and foster long-term community resilience.

Together, these contributions will advance both the theoretical understanding and practical application of crime prevention efforts, offering a pathway for strengthening community stability, reducing burglary risks, and fostering long-term neighborhood resilience.

Dataset Description

The dataset that was used in this project was compiled by UC Irvene's Machine Learning Repository in 2009. The dataset combined socio-economic data from the 1990 United States Census, law enforcement data from the 1990 United States Law Enforcement Management and Administrative Statistics (LEMAS) survey, and crime data from the 1995 FBI Uniform Crime Report (UCR). The United States census is a decennially taken census meant to count the total amount of United States residents, as well as note demographic statistics in an area (United States Census Bureau, 2023). The LEMAS survey is an optional survey of police departments conducted by the Federal Bureau of Investigation (Bureau of Justice Statistics, 2009). This survey asks departments to report their staffing levels, police demographics, and budget. Finally, the UCR aggregates yearly reporting from police departments nationwide, tracking crime statistics and classifying them by type (FBI, 2024). All three of these sources are well-established, primary sources with a national scope. Therefore, this data is very valuable in a data-driven study, as it provides wide-ranging and detailed data that can be tracked and analyzed.

In total, the dataset has 127 features and 2,215 instances. The unit of analysis of this dataset is a community. The features of the dataset vary widely but can be generally organized into eight categories of data. The first type is geographical, and this includes columns such as the name of the state, the percentage of residents living in an urban area, and a community name. The second is population-related variables such as total population and average household size. Demographic statistics covering both race and age are included. Many economic factors are tracked, like median income, percentage of citizens utilizing public assistance, and median income striated by race. Occupational and education factors such as high-school graduation rates, higher education rates, and employment rates are included in the dataset. Family statistics are included, such as divorce rates, percentage of families with two parents, and percentage of families with working moms. Additionally, housing variables, such as rent costs and the percentage of people in homeless shelters were provided. Finally, policing statistics were provided by LEMAS, which included the number of full-time officers, the number of requests for officers, and the racial breakdown of the police force.

Variables in our model were chosen to understand how changes in residential stability predict changes in burglary rates, guided by the Social Disorganization Theory. As a baseline, all of the variables considered had correlations of absolute value 0.3 or higher with the target. Independent features with correlations of absolute value 0.8 or higher were considered to have multicollinearity. Each feature was individually evaluated to decide whether to keep it or not. In general, for the sake of the research question, variables representing households or communities were chosen over those representing families or states.

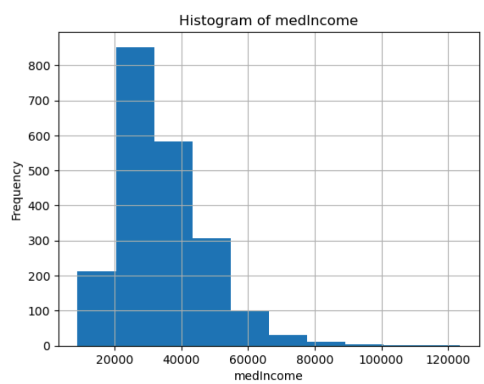
The variable racePctWhite was chosen over racePctBlack due to its stronger correlation with the target variable and its comprehensive representation of racial demographics. MedIncome and MedFamInc were selected for their high correlations with the target and ability to accurately reflect economic conditions. pctWWage was preferred over pctEmploy due to its focus on household income from wages, aligning with our community-level analysis. pctWInvInc and pctWPubAsst were included for their high correlations with the target, representing financial stability and economic distress, respectively. PctPopUnderPov, PctNotHSGrad, and PctUnemployed were added to control for socioeconomic factors that influence both stability and crime rates. TotalPctDiv and PctKids2Par reflect the family structure, while PctPersOwnOccup and PctHousOccup measure homeownership and housing occupancy, both indicative of residential stability. Housing characteristics like PctHousLess3BR, MedNumBR, PctVacantBoarded, and PctWOFullPlumb were included to capture aspects of housing quality and neighborhood decline. MedRentPctHousInc measures housing affordability, impacting residential stability. Finally, PctSameHouse85 is our primary variable of interest, directly measuring residential stability, and allowing us to test our hypothesis that high instability influences higher burglary rates. This comprehensive set of variables allows us to control for various factors and isolate the effect of residential stability on burglary rates.

In addition to their statistical significance, these variables were also carefully chosen to align with this study's Social Disorganization Theory roots. The chosen variables can be sorted into two categories, data types which are known to be important to Social Disorganization Theory and data types which have more ambiguous connections with the theory, which we are testing within the study. Specifically, variables related to the socio-economic and housing factors of a community (medIncome, pctWWage, pctWInvInc, pctWPubAsst, medFamInc, PctPopUnderPov, PctUnemployed, PctPersOwnOccup, PctHousLess3BR, MedNumBR, PctHousOccup, PctVacantBoarded, PctWOFullPlumb, and MedRentPctHousInc) have been shown to be correlatory to the burglary rate of a community, in line with Social Disorganization Theory (Rosenfeld & Messner, 2013). However, what has not been tested as thoroughly is whether introducing further demographic information, such as the racial composition, educational achievements, and family structures of a community, will improve a model's ability to predict a community's burglary rate. So, a mixture of reliable and prospective metrics were chosen, in the hopes of giving the model the most accurate, diverse, and useful information possible.

Once the features of this project were established each one could be investigated in more detail. A function was created which looped through every column, creating a basic histogram for each feature so that the data trends could be investigated. For many of the features, a right-skewed histogram was present, with the vast majority of the values falling within a typical distribution and isolated outliers skewed the graph. An example of this distribution is the histogram for average income, which can be seen in Figure 1.

Figure 1

Right-Skewed Histogram for Median Income

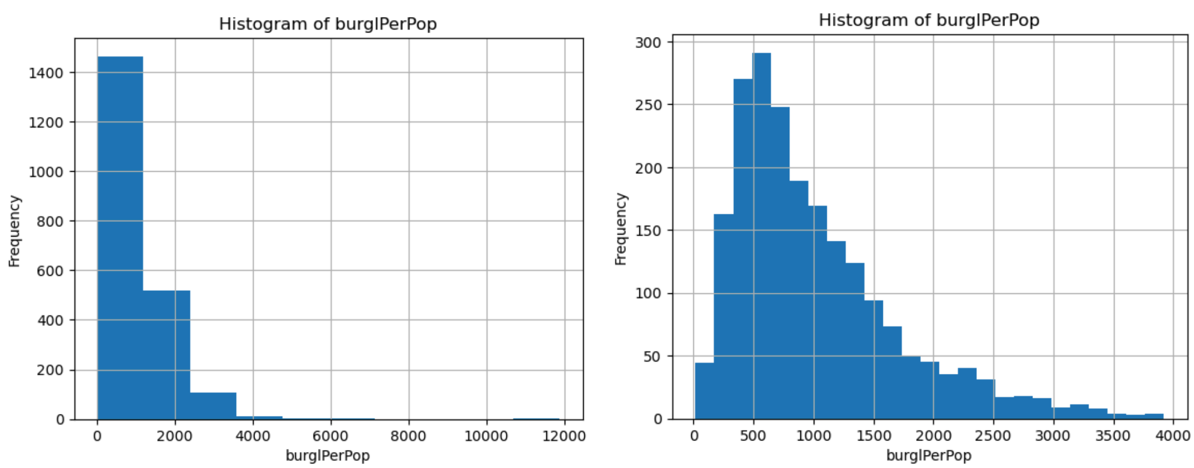


This kind of distribution was present with the columns medIncome, pctWPubAsst, medFamInc, PctPopUnderPov, PctNotHSGrad, PctUnemployed, PctHousOccup, PctVacantBoarded, PctWOFullPlumb, burglPerPop. Additionally, PctHousOccup, pctWWage, and racePctWhite had the opposite problem, being left-skewed. The presence of all of these outliers suggested that categorical models, which would help normalize the data against extreme outliers, would likely be useful in this project.

Inspection of the histogram for the target variable, burglPerPop, showed that a few outliers were considerably lessening its efficacy, as can be seen in Figure 2. Only ten very burglarized cities, whose burglPerPop value ranged from 4,000 to 12,000, caused the histogram to be widely skewed right, reducing its usability. Therefore, a new data frame that excluded those ten rows was created, and a new histogram was created, which can be seen in Figure 3.

Figure 2 and Figure 3

Original and Corrected Histogram for Burglary Per Population

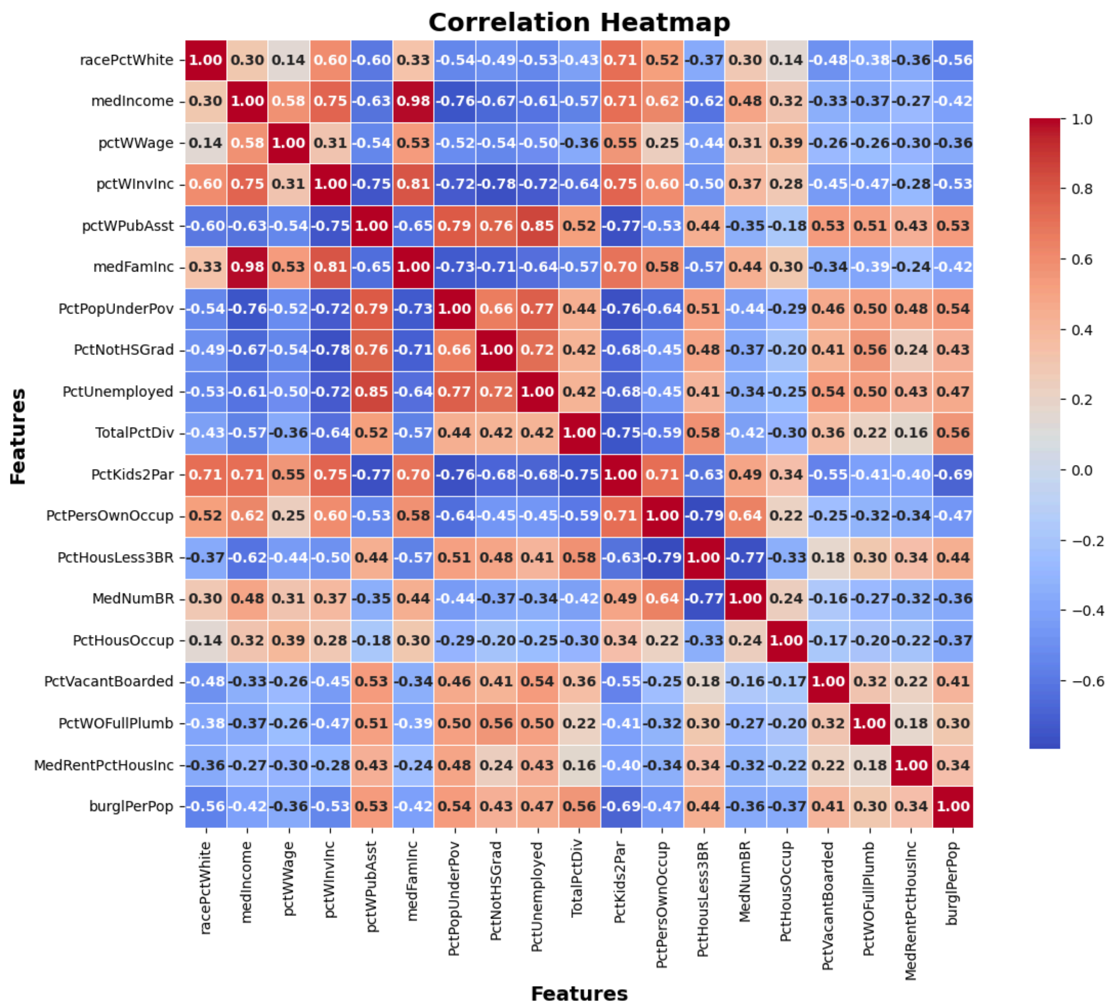


The new burglPerPop histogram showed that the majority of cities had between 250 and 1,400 burglaries per 100,000 people. Most importantly, this further shows that this dataset is very susceptible to outliers, which will need to be noted during the modeling portion to ensure it does not negatively impact the findings.

The final step of the data investigation was to create a correlation matrix of the chosen features. This can be seen in Figure 4. Due to the previous evaluation done on the features, all of them had a correlation of at least 0.3 with the target. Interestingly, exactly half of the dataset had a negative correlation with the target, and half had a positive correlation with it. The most strongly positively correlated features were TotalPctDiv and PctPopUnderPov, with correlation scores of 0.56 and 0.54, respectively. This showed that marriage statistics and economic features would likely be important in this project. Similarly, PctKids2Par was by far the most absolutely correlated with burglPerPop, with a correlation of -0.69. Clearly, the presence of two-parent households, unaffected by divorce, was a very important factor in predicting burglary rates.

Figure 4

Correlation Matrix of Features Being Used in Modeling



Methodology

The primary piece of preprocessing which was done was to handle null values within the dataset. While the dataset was mostly null-free, a few columns were found to be problematic. In particular, the data sourced from LEMAS all had question marks, which were presumably stand-ins for null values, in the same 1,872 rows. Outside of the LEMAS values, the columns communityCode and countyCode both had 1,224 and 1,221 null values respectively. ViolentCrimesPerPop, rapesPerPop, and rapes all had between 200 and 225 null values. Additionally, nonViolPerPop, arsonsPerPop, and arsons all had between 90 and 100 null values. Finally, eleven other columns had a total of forty-seven nulls between them. In all, 27.9% of columns had any null values, though many of them stemmed from the LEMAS data, which was not used in this study. When excluding LEMAS data, only 1.4% of columns had null values accounting for more than ten percent of their data. As this was a large dataset with many rows, the null values could be dropped from the dataset without comprising its predicting power

The first step after completing data preprocessing was to create a baseline model. As the target for this study is continuous, the average of burglPerPop was taken and compared to the actual burglPerPop values. The purpose of the baseline model was to establish a control in this study, providing a reference point to use against the results of future models. If the future models had results better than the baseline model, then they could be considered successes.

We identified three models for testing that met our requirements for interpretability and accuracy: Decision Tree, Random Forest, and Gradient Boosting. Decision Tree models provide quick glances at the most important parts of the data, as well as a peek at the modeling process. Random Forest models offer a very stable output that is less prone to overfitting or biases while still being generally accurate. Finally, Gradient Boosting models offer potentially the most accurate output, though it can fall victim to overfitting. The hope was that by taking the results from all three model types each of their strengths could be highlighted, and their weaknesses minimalized.

The metrics chosen to be used to measure the regressor model's accuracy were MAE, MSE, RMSE, R-squared, and Cross-Validation Mean. Mean Average Error measures how much difference there is between the predicted and actual values. Mean Squared Error does the same thing but squared so that values which are extremely incorrect are more highly valued. R-squared measures how much of the variance in a dataset can be explained by the model. Finally, the Cross-Validation Mean is the result of 5-fold analysis and a more accurate version of MAE. It is created by splitting the dataset into five subsets, treating each as the testing portion, and averaging the results, giving an output which is less affected by random biases within the data. When used together, these variables give a clear and thorough picture of the mode's efficacy.

One advantage of the chosen dataset was its abundance of variables. Within the dataset, there were columns representing both the median income of families, as well as the median income of individuals. To ensure the robustness of the model, it was decided that both types of income would be tested separately. Because both of these variables had a similar correlation with the target variable, and considerable overlap in their function, testing them independently helped ensure the most effective model was created.

After every model had been tested with both the individual median income and household median income it was decided that the process would be repeated with binned categorical data. One benefit of binning the data was that the negative impact of outliers, which were seen to be a potential issue within the preprocessing stage, would be negated. Additionally, having a categorical target can oftentimes improve model performance, as it allows for the model to focus on general accuracy, rather than fixating on predicting the target's exact value.

In order to keep the binning process simple and easily understandable every non-target column was binned into quartiles. These quartiles were labeled as Low, Medium-Low, Medium-High, and High. The only column which was treated differently was burglPerPop, which was divided into two bins. As the focus of this study was to investigate the highest crime areas in comparison to others, only the top quartile of communities were sorted into High Burglaries, with the other seventy-five percent of the data all being categorized as Low Burglaries.

For the classification models, new accuracy metrics had to be used. The metrics chosen were Accuracy, Precision, Recall, F1-Score, and Cross-Validation Mean. Accuracy measures the total accuracy of the model, by dividing the amount of correct predictions the model made by the total amount of predictions. A more specific measurement is Precision, which measures how often the predicted true values were actually true. Similarly, Recall measures how many of the total true values were correctly predicted as true. The F1-Score synthesizes Recall and Precision, giving a more accurate measure of the total accuracy of the model. Finally, the Cross Validation Mean once again used the data from the 5-Fold Analysis.

At this point, the categorized Gradient Boosting model which used medFamInc was adjusted slightly. This model proved to be the most accurate, so it was selected as the prospective final model. Therefore, the model was run again, only this time also including a column related to the amount of residential mobility in a community, PctSameHouse85. This feature measures what percentage of a community was living in the same house as they were five years previously. This was done as a check to ensure that Social Disorganization Theory was being correctly applied to this study, as residential stability is an essential part of Social disorganization Theory. If the model regressed or remained stagnant then the principles of Social Disorganization Theory may not be applicable to this study. However, if the model substantially improved then it could be reasoned that Social Disorganization Theory was a guiding force within this study.

Analysis

The first set of models that had to be analyzed were the uncategorized Gradient Boosting and Random Forest Regressors. Each model was run twice, once using medIncome and once using medFamInc. As these were regression models using uncategorized data, the evaluation metrics MAE, MSE, RMSE, R^2, and Cross-Validation Mean were used to evaluate the models. Their accuracy metrics, compared against the baseline model, are shown in Table 1.

Table 1

Metrics for the first set of models

Model	MAE	MSE	RMSE	R^2	Cross Validation Mean
Baseline	554.2	8587785.9	766.7	NA	NA
RF (medIncome)	328.9	208187.5	456.3	0.58	329.9
GRB (medIncome)	322.8	211449.1	459.8	0.57	325.8
RF (medFamInc)	323.2	222359.3	471.6	0.59	331.4
GRB (medFamInc)	319.9	217928.3	466.8	0.60	326.2

The results from the first round of modeling showed that the baseline model, unsurprisingly, was the least effective at predicting highly burglarized communities. This means that at the very least, every model does better at predicting the target than simply looking at the burglPerPop mean. More specifically, medFamInc seemed to consistently outperform medIncone, with both of its R-squared values higher. Similarly, Gradient Boosting models generally performed better than their Random Forest counterparts. The GRB models had the two lowest MAEs and Cross-Validation Means. Along those lines, the best performing model of all was the Gradient Boosting model which was trained on medFamInc.

The Decision Trees for this round of models showed that the most important variable may be PctKids2Par. PctKids2Par appeared at the top of the medIncome decision tree and as both of the second layer nodes in the medFamInc decision tree. In every case, areas with more two-parent households tended to have fewer burglaries.

The next set of models used categorical data, which meant that new accuracy metrics had to be used. The metrics chosen were Accuracy, Precision, Recall, F1-Score, and Cross-Validation Mean. Once again, both Random Forest and Gradient Boosting models were used. Additionally, both models were again run twice, once using medIncome and once using medFamInc. The data from the second set of models can be seen in Table 2.

Table 2

Metrics for the second set of models

Model	Accuracy	Precision	Recall	F1-Score	Cross Validation Mean
RF (medIncome)	0.75	0.25	0.04	0.07	0.75
GRB (medIncome)	0.56	0.23	0.39	0.29	0.76
RF (medFamInc)	0.77	0.29	0.03	0.06	0.74
GRB (medFamInc)	0.57	0.21	0.33	0.25	0.75

The results from the second wave of models were very interesting, with a few notable trends. Firstly, the models struggled to predict the Precision, Recall, and F1-Score of true values, as can be seen in the table. However, the models were very effective at predicting false values. For instance, the Random Forest model which used medIncome had a true precision and recall of

0.25 and 0.04. However, its false precision and false recall were 0.77 and 0.96. As this project was meant to predict highly burglarized areas and not safe ones, these values led to deceptively high accuracy and Cross-Validation scores. This also led to Random Forest models having higher accuracy scores, despite the gradient-boosting models actually having significantly better recall and F1 scores. Overall, these models struggled to accurately predict the highly burglarized communities, though they clearly had predicting power.

The second round of Decision Trees, now using binned data, showed that income level was found to be significantly correlated with highly burglarized areas. Unsurprisingly, higher-income areas tended to have lower rates of burglaries, whether it was medFamInc or medIncome. Similarly, areas with high rates of vacant homes were generally more prone to burglaries.

The final models created hoped to more effectively capture the categorical model's predicting ability, only used more effectively. This was done by introducing a new variable, PctSameHouse85, which tracks how much of a community is still living in the same house that they were five years previously.

Table 3

Metrics for the final model

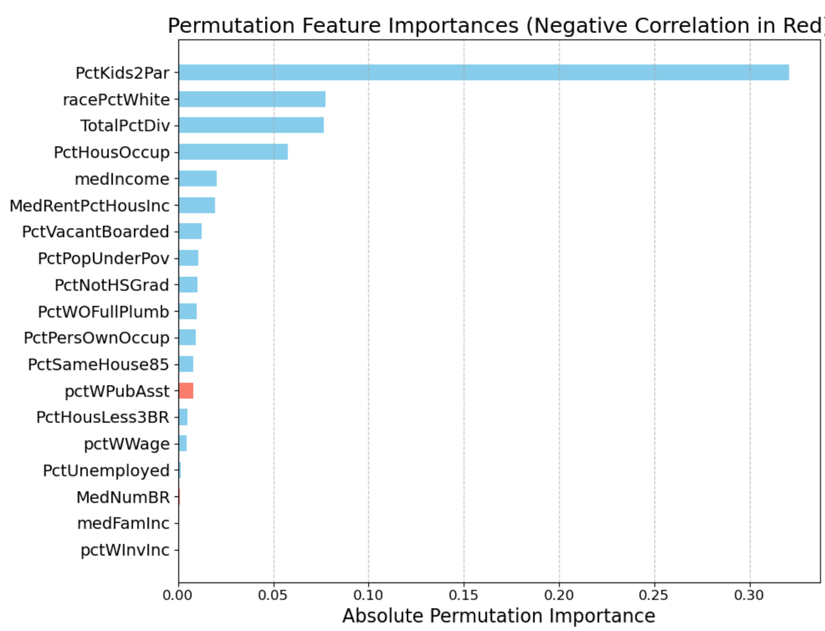
Model	Accuracy	Precision	Recall	F1-Score	Cross Validation Mean
RF (medFamInc)	0.81	0.86	0.89	0.88	0.85
GRB (medFamInc)	0.82	0.92	0.82	0.87	0.84

Introducing the new variable into the models had a dramatic, positive effect on it. While the overall accuracy and Cross-Validation Mean did improve, it was the precision and recall that had enormous changes. In total, the accuracy and cross-validation mean increased by about 0.1, while the precision and recall both increased by closer to 0.5. Clearly, PctSameHouse85 was the missing variable that the models needed in order to accurately understand burglarized communities.

The final step of analysis taken was to create a feature importance graph, which shows which features of the model had the most impact on it. This graphic can be seen in Figure 5. The graphic was developed so that the absolute value of the importance was taken, so that both positively and negatively correlated factors could be compared easily. Therefore, in order to accurately track the difference between negative and positively correlated factors, positively correlated factors were shown in blue, while negatively correlated ones were displayed in red.

Figure 5

Feature Importance for Final Model



The feature importance graphic's most apparent result was to show that PctKids2Par had an enormous impact on the model, with an importance of 0.34, and no other feature having an importance of more than 0.15. This was also in line with the discoveries of the first set of Decision Trees. The second most important feature was one very similar in nature, TotalPctDiv. PctHousOccup and racePctWhite were the only other columns with an importance greater than 0.03.

Discussion and Conclusion

This research addresses a critical gap in criminological literature by examining how residential stability and related socio-economic factors impact burglary rates. While Social Disorganization Theory had established a general link between weakened social cohesion and increased crime, prior research had often overlooked specific predictors of burglary risk at the community level. By incorporating variables such as long-term residency, two-parent households, and vacant housing, this study provides nuanced insights into how stability-related factors shape crime patterns. These findings highlight that residential stability is not only a significant predictor of burglary but also a tangible factor that communities can target to develop effective interventions.

The variable measuring long-term residency emerged as an influential predictor of burglary rates. Long-term residency fosters stronger social networks and enhances collective cohesion, a community's ability to maintain order and deter criminal behavior. This variable underscores the protective role of stable neighborhoods in reducing crime, aligning with Social Disorganization Theory. Policies aimed at increasing long-term residency, such as affordable housing programs and incentives for homeownership, can directly strengthen neighborhood cohesion and indirectly reduce burglary rates.

Family structure variables, such as the percentage of children in two-parent households and divorce rates, were also found to significantly impact burglary risk. Communities with higher proportions of two-parent households tend to exhibit greater stability and social oversight, while areas with elevated divorce rates may suffer from weakened family structures and diminished community ties. These findings emphasize the importance of family stability in reducing crime. Programs such as subsidized child care, parenting education, and marital counseling can support family units, fostering environments where strong social bonds deter criminal activity.

Housing-related variables, including housing occupancy rates and the percentage of vacant properties, provide further depth to the understanding of how neighborhood dynamics influence burglary rates. High housing occupancy reflects a community's vibrancy and resilience, whereas vacant properties often signal neglect and attract criminal activity. Interventions that promote housing stability, such as rehabilitating vacant properties or providing financial assistance to prevent evictions, can reduce opportunities for crime while bolstering the social fabric of neighborhoods.

These findings validate our hypothesis that increasing residential stability reduces burglary rates and emphasize that crime is deeply interconnected with broader social, economic, and demographic conditions. This research advances the field by pinpointing specific socio-economic and environmental factors that significantly impact burglary risk. Identifying how changes in residential stability, family structure, and housing conditions influence burglary rates fills a notable gap in the literature, providing actionable guidance for policymakers and community organizations.

For policymakers, this research underscores the need for funding and implementing evidence-based programs that strengthen families and communities. Practical policies such as accessible childcare and pre-parenting programs are particularly impactful. Subsidized childcare and parenting education create stable environments for children, reducing risks of family disorganization and fostering long-term community engagement. Additionally, investments in school transportation and after-school programs relieve logistical and financial burdens on families, encouraging them to remain in their neighborhoods and deepen community ties. Expanding community resource hubs—offering services like financial counseling, job training, and mental health support—can further stabilize households, enabling families to overcome challenges and remain rooted in their communities.

Community members themselves are central to fostering residential stability. Policies that prioritize community engagement—such as resource hubs and voluntary participation in educational programs—empower residents to take an active role in shaping their neighborhoods. Transparent communication about data use and ensuring equitable access to resources can mitigate concerns about privacy and stigmatization, building trust and encouraging participation.

The predictive modeling approach used in this study not only validates theoretical frameworks but also can be used to anticipate burglary hotspots before they occur, enabling proactive interventions. Our predictive modeling allows for the identification of areas at higher risk of burglary based on changes in residential stability, providing community leaders with insights to act ahead of time rather than relying solely on reactive measures. This strengthens the overall strategy for crime prevention by allowing policymakers to implement targeted, preventative actions.

By investing in long-term strategies that stabilize families and neighborhoods, communities can effectively reduce burglary rates and associated financial and psychological burdens. Stabilizing neighborhoods also yields broader economic benefits by reducing costs associated with burglary, including property damage, lost productivity, and emotional trauma. Programs that promote stability improve educational outcomes, workforce readiness, and overall community well-being. At the same time, targeted interventions must be implemented equitably to avoid exacerbating existing inequalities. Location-based assessments can help prioritize underserved areas, ensuring that resources are distributed fairly. Programs that emphasize voluntary participation and community-driven decision-making further enhance trust and inclusivity, contributing to safer and more resilient communities.

Despite its strengths, this study has limitations. The dataset, based on data from the 1990s, may not fully capture contemporary crime trends or the influence of modern socio-economic dynamics. Future research should incorporate more recent data to validate these findings and explore new variables, such as the impact of technological advancements or urbanization patterns. Additionally, while this study focuses on U.S. communities, its findings may not generalize to other countries or account for regional variations. Expanding the analysis to include diverse geographic and cultural contexts would enhance its applicability and relevance. Finally, unmeasured factors, such as specific community programs or local cultural attitudes, may also play a role in shaping burglary rates and should be investigated in future studies.

Overall, this research emphasizes the interconnectedness of family, housing, and community dynamics in shaping crime patterns, and particularly the significance of residential stability on burglary rates. Proactive, stability-focused strategies—such as parenting education, housing policies promoting homeownership, and community engagement initiatives—offer a sustainable approach to reducing burglary while fostering safer and more cohesive communities. Beyond crime prevention, these interventions have far-reaching benefits, creating neighborhoods where families and individuals can thrive. By addressing the root causes of burglary through evidence-based policies, this study provides a roadmap for future research and actionable solutions for policymakers and practitioners aiming to build resilient, thriving communities.

References

Beaton, A., Cook, M., Kavanagh, M., & Herrington, C. (2008). The psychological impact of burglary. *Psychology, Crime & Law, 6*(1), 33–43. https/doi/pdf/10.1080/10683160008410830

Bellair, P. (2017). Social Disorganization Theory. In *Oxford Research Encyclopedias.* Retrieved from https://oxfordre.com/criminology/display/10.1093/acrefore/9780190264079.001.0001/acrefore-9780190264079-e-253?print=pdf

Bureau of Justice Statistics. (2009). *Law Enforcement Management and Administrative Statistics (LEMAS).* https://bjs.ojp.gov/data-collection/law-enforcement-management-and-administrative-statistics-lemas

Chen, A. (2019). Why are so many households unable to cover a $400 unexpected expense? *Center for Retirement Research at Boston College, 19*(11), 1-10. https://crr.bc.edu/wp-content/uploads/2019/07/IB_19-11.pdf

Eck, J. E., & Clarke, R. V. (2003). Classifying burglaries for prevention. *Problem-Oriented Guides for Police Problem-Specific Guides Series, No. 18*. U.S. Department of Justice. https://popcenter.asu.edu/sites/default/files/problems/burglary_apartmentcomplexes/PDFs/Eck_Clarke_2003.pdf

Federal Bureau of Investigation. (2019). *Burglary*. https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/burglary.pdf

Federal Bureau of Investigation. (2024). *Crime/Law Enforcement Stats (Uniform Crime Reporting Program)*. https://www.fbi.gov/how-we-can-help-you/more-fbi-services-and-information/ucr

Hipp, J. R. (2010). A dynamic view of neighborhoods: The reciprocal relationship between crime and neighborhood structural characteristics. *Social Problems, 57*(2), 205–230. https://doi.org/10.1525/sp.2010.57.2.205

Hipp, J. R., & Yates, D. K. (2011). Ghettos, thresholds, and crime: Does concentrated poverty really have an accelerating increasing effect on crime? *Criminology, 49*(4), 955–990. https://doi.org/10.1111/j.1745-9125.2011.00252.x

Rosenfeld, R., & Messner, S. F. (2013). The social sources of homicide in different types of societies. In M. A. Zahn, H. H. Brownstein, & S. L. Jackson (Eds.), *Violence: From theory to research* (pp. 59–74). Routledge. https://doi.org/10.4324/9780203940672

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology, 94*(4), 774–802. https://doi.org/10.1086/229068

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328), 918–924. https://doi.org/10.1126/science.277.5328.918

Sampson, R. J., & Raudenbush, S. W. (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology, 105*(3), 603–651. https://doi.org/10.1086/210356

Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas: A study of rates of delinquency in relation to differential characteristics of local communities in American* *cities*. University of Chicago Press. https://www.taylorfrancis.com/chapters/edit/10.4324/9781439817803-9/juvenile-delinquency-urban-areas-study-rates-delinquency-relation-differential-characteristics-local-communities-american-cities-1969-shaw-mckay

Shover, N. (1991). Burglary. *Crime and Justice: Review of Research, 14*(1)*,* 73-114. https://www.ojp.gov/ncjrs/virtual-library/abstracts/crime-and-justice-review-research-volume-14

Spelman, W. (1993). Abandoned buildings: Magnets for crime? *Journal of Criminal Justice, 21*(5), 481–495. https://doi.org/10.1016/0047-2352(93)90034-X

United States Census Bureau. (2023). *How the Data Are collected. https://www.census.gov/programs-surveys/gov-finances/technical-documentation/methodology/how-the-data-are-collected.html*

Vargas, J. (2023). The impact of socioeconomic factors on crime rates. *Addiction & Criminology, 6*(4), 1-2. https://doi.org/10.35841/aara-6.4.161

Warner, B. D., & Rountree, P. W. (1997). Local social ties in a community and crime model: Questioning the systemic nature of informal social control. *Social Problems, 44*(4), 520–536. https://doi.org/10.1525/sp.1997.44.4.03x0219h

Willibald, O., Mukiibi, S., & Limbumba, T. (2018).Understanding residential mobility. *American Journal of Engineering Research, 7*(5), 503-507. https://www.researchgate.net/profile/Stephen-Mukiibi-2/publication/372953265_Understanding_Residential_Mobility/links/64d0dd9cd394182ab3b13259/Understanding-Residential-Mobility.pdf