**To:** Director of Smart Cities

**From:** Bailey Bradford, Frances Murray, Jonathan Zisk

**Date:** May 13, 2024

**Subject:** A Model for Neighborhood Gentrification Risk

**Executive Summary**

We have developed a logistic regression model that offers a comprehensive solution to risks of gentrification by analyzing demographic shifts and socioeconomic trends over time in Nashville, TN. The validity of the model is tested by predicting gentrification in Chattanooga, TN, which has been identified as a comparable region to Nashville. By leveraging advanced modeling techniques, our model anticipates neighborhood changes that can be indicative of gentrification, ultimately providing actionable insights for policymakers and community stakeholders to address risky trends and curb displacement before it happens. Though the model is limited in its ability to predict with total accuracy the neighborhoods that will indeed gentrify, it gleams insight into a neighborhood’s change over time. These changes should be considered and used to inform the promotion of practices aimed at retaining residents and maintaining affordability for all. Our model predicts neighborhood change using a combination of median household income, Black population rate, Hispanic/Latino populations, and the rate of graduate degrees variables.

**Operationalization of Terms**

In this report, our operationalization of gentrification is informed by the Urban Displacement Project’s definition: A process of neighborhood change that includes economic change through real estate investment and new higher-income residents moving in, as well as demographic change in income level, education level, or the racial make-up of residents (Urban Displacement Project). The risk of such changes manifests as displacement, pushing out long-term residents and most often minority groups.

**Methodology**

1. **Data Collection & Feature Selection:** Demographic data is sourced from the US Census Bureau 5-year American Community Survey in 2012 and 2019 by census tract for Nashville and Chattanooga. To prepare modeling, we analyzed variables that, according to our operationalized definition, could be indicators of gentrification over time. These variables include Median Home Value, Median Household Income, Renter-Occupied Rate, White Population Rate, Black Population Rate, Hispanic/Latino Population Rate, Graduate Degree Rate, and SNAP Recipient Rate. A correlation matrix was created to examine collinearity between variables. Variables with strong negative or positive relationships indicated relative correlation and were used to develop the model.
2. **Clustering Analysis:** Using k-means clustering, we identified groups of neighborhoods demonstrating similar trends based on the listed variables. The optimal number of clusters was determined using the elbow method. By assigning census tracts to trend-specific clusters, we can create a binomial variable model to anticipate changes in neighborhoods of concern. The clusters are outlined below and visualized in Figure 1.

A map of different colored states

Description automatically generatedA map of different colored states

Description automatically generated

Figure 1: Cluster Analysis of Nashville and Chattanooga

1. Increasingly Affluent: high household income; high median home value; majority white.
2. Potentially Gentrifying: moderate increase in income and home value over time.
3. Rapidly Gentrifying: increase in household income and median home value; increase in white population and decrease in Black and Hispanic/Latino populations; increase in graduate degrees.
4. Resisting Gentrification: stable home values; low-moderate change in median income; stable change in Black and Hispanic/Latino populations.
5. **Spatial Analysis:** To consider multiple factors, we included the evaluation of spatial data that could be indicative of gentrification trends. These include the rate of change in crimes over time, the number of public art pieces in each neighborhood, 311 calls (all sourced from municipal databases), and proximity to bus stations (sourced from the Chattanooga Area Regional Transportation Authority (CARTA) and Nashville’s WeGo Public Transit).
6. **Model Selection & Validation:** We created four logistic regression models using Nashville data to predict the change in clusters based on different sets of independent variables. Our model combines demographic and socioeconomic variables to make predictions about a neighborhood’s likelihood to gentrify in the future using 2012 data.Specifically, we examine the neighborhoods categorized as Rapidly Gentrifying to test if the model accurately predicts 2019 gentrification with the 2012 data. The Nashville model is then cross-validated with Chattanooga data do test for replicability and evaluated with confusion matrices and Receiving Operating Characteristic (ROC) Curves.

**Results of Model Testing**

1. **Training Data (Nashville):** The model was developed using Nashville data. As seen in Figure 2, it has a high frequency of predicting true negative values, but a moderate frequency of predicting false negatives. Evaluation shows a recall rate of 32%, suggesting it is low in sensitivity and therefore not good for predicting true positives. The specificity rate is 95%, demonstrating high accuracy in predicting true negative rates. The rate of accuracy is 88%, supporting that this model is highly likely to make accurate predictions. The model’s precision is 50%, which is relatively low. This means only half of the neighborhoods predicted to undergo gentrification will be accurately predicted.
2. **Testing Data (Chattanooga):** The model is tested by cross-validating with Chattanooga data. Similar to the training data, the testing data results show a high frequency of predicting true negatives, a moderate frequency of predicting false negatives, and a low frequency of predicting false positives, as shown in Figure 3. It has no instances of predicting true positives, which suggests the model is overfit to Nashville data. However, it shows an accuracy rate of 85% and a specificity rate of 97%.

A table with numbers and words

Description automatically generatedA table with numbers and text

Description automatically generated

Figures 2 & 3: Confusion Matrices

1. **ROC Curve:** The tradeoffs between specificity and sensitivity can be examined in the ROC Curves depicted in Figures 4 and 5 below. For both cities, the ROC Curve shows high sensitivity and specificity greater than random chance, as shown in the space between the curve and the diagonal line. This supports future use of this model.

**A graph of a graph showing a curve

Description automatically generated with medium confidence**

Figures 4 & 5: ROC Curves

**4. Overall Model Performance:** The model’s trade-off between sensitivity and specificity suggests it can effectively identify neighborhoods at risk of gentrification. While it is over-fitted to Nashville neighborhoods, its use can still benefit the allocation of time and resources to mitigating gentrification.

**Recommendations for Implementation**

Based on the insights gleaned from our model’s analysis of neighborhood gentrification risk, we recommend a multifaceted approach to planning and policy-making. We recommend interventions prioritize neighborhoods identified as rapidly or potentially gentrifying, focusing on preserving affordability and inclusive development. This may include promoting affordable housing initiatives, rent stabilization programs, or community land trusts. Incentivizing mixed-income housing developments with subsidies and credits can also help to retain current residents and discourage future displacement. Fostering community engagement and collaborating with local stakeholders should be at the forefront of gentrification-mitigation efforts. This model indicates which neighborhoods are at risk of gentrification, but the categorical nature prevents it from specifying characteristic differences between cities. Understanding the complexity of a city’s population and economy is integral to successful intervention.