To: Director of Smart Cities
From: Yuanhao Zhai, Jiewen Hu
Date: 2024/05/13
Subject: Predictive Model for Gentrification Risk Recommendation

Introduction:

Gentrification is a common process in urban development. Although the accompanying effect of gentrification may benefit the residents by improving amenities and providing job opportunities, it may also carry steep costs for residents who cannot afford to stay in the neighborhood due to increased rent or other reasons. In this context, identifying gentrified areas allows the city to get ahead and allocate additional funds to save the living environment of neighborhood residents.

In this program, our team developed a predicating model software to help the city government identify areas at risk of gentrification using the gentrification index. In the following article, I will introduce three parts of our project including the measure of defining gentrification, the model development process, and the model validation on other major cities of the US.

Literature Review:

Since the early 2000s, a large number of literature has examined the definition and potential causes of gentrification. The U.S. Department of Housing and Urban Development (HUD) defines gentrification as "a form of neighborhood change that occurs when high-income groups move to low-income areas, potentially altering the cultural and financial landscape of the original neighborhood". Previous researches emphasize the changing neighborhood's socioeconomic characteristics in terms of demographics, land use, and housing affordability, which is caused by the inflow of upper-class migration (Lees et al., 2008). Some literature also suggests that property renovation can be a major indicator of gentrification (Mayer, 1981). So, in this project, in order to determine the risk of gentrification, we will focus on the changing of neighborhoods' demographics, housing market, and the number of amenities.

Measure Gentrification & Data Selecting:

With the rising attention to gentrification, the quantitative measurements of gentrification become an important issue in the research field. The **Gentrification Index** which contains multi-related variables was developed as an important determinant of a neighborhood's socioeconomic status with regard to gentrification (*Gentrification Index | Nathalie P. Voorhees Center for Neighborhood and Community Improvement | University of Illinois Chicago*, n.d.). The literature for predicting which residents and neighborhoods will be affected by gentrification uses this index as an indicator variable for model building (Bureau, n.d.). So, in this project, to better measure the gentrification of areas, we will use the Gentrification Index as an essential indicator of the gentrification areas

(Figure 1). The model dependency will be the changing level of the index. The areas with a higher increase in the gentrification index are facing more risk of being gentrified.

Figure 1: Variable Score Assignments

Variables	Type of Association
% White (Non-Hispanic)	Above City Average, Positive (+1)
% Black	Above City Average, Negative (-1)
% Latino	Above City Average, Negative (-1)
% Elderly (Age 65+)	Above City Average, Negative (-1)
% Children (Age 5-19)	Above City Average, Negative (-1)
% College Education (Bachelor's degree or higher)	Above City Average, Positive (+1)
Median Family Income (Adjusted for inflation)	Above City Average, Positive (+1)
% Owner Occupied	Above City Average, Positive (+1)
Median House Value (Adjusted for inflation)	Above City Average, Positive (+1)
% Families Below Poverty	Above City Average, Negative (-1)
% Manager Occupations	Above City Average, Positive (+1)
% Female Households with Children	Above City Average, Negative (-1)
% Private School Attendance (Pre-K through 12)	Above City Average, Positive (+1)

For the independent variable used to predict future risk, we use the following variables:

Socioeconomic indicators: median income changes, education level composition changes, age composition changes, family type changes, marriage status, tenure status, median housing price, ethnicity composition

Housing Characteristics: building types, building year, housing condition.

Local Amenity: Coffee shop numbers, Number of restaurants, grocery stores, parks nearby, education facilities numbers nearby,

The data we use here will be collected from the Census Bureau, and the Longitudinal Tract Database (LTDB) will be our database since our model will develop under census tracts scale. The amenity data will be from the Commercial real estate. American Housing Survey will provide housing characteristics data. The data from Chicago will be used for the training model, and the data from Philadelphia will be used for model validation.

Regression Model Development

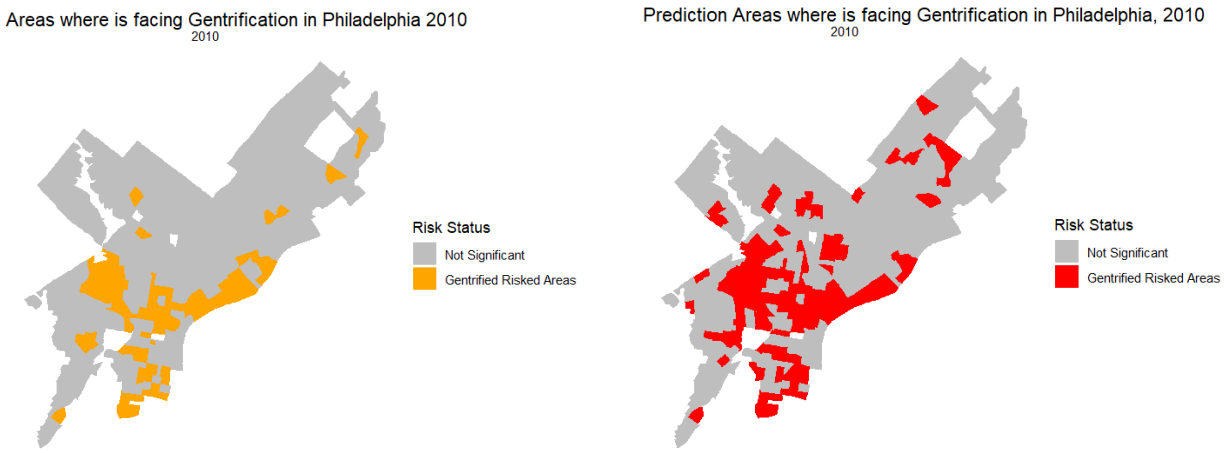
Since the goal of this model is to predict the high-risk areas of gentrifying, we can use the Logistic regression as our model. Logistic regression is a parametric classification model that uses a logistic function to estimate binary output models. So it's very useful to predict the binary result, which is risked or not risked.

We obtained the best model for the prediction of the risk of gentrification through a series of model tests during the construction of the model. According to the results, the McFadden R^2 of this model is **0.3588**, which is more

accurate compared with the same type of regression model. Meanwhile, according to the results of the Confusion Matrix, the accuracy of this regression prediction model is **84.25%**, and the result of the ROC curve is **0.8847**. It can predict most of the gentrification risks.

Validation:

To validate the model's functionality, we applied the model to the Philadelphia region. The results are shown below, with Philadelphia's 2010 gentrification risk areas on the left and the model's predictions on the right. From the results, the model was largely successful in predicting high-risk areas in Philadelphia. You can also get more detailed information from the technical appendix in Github repository.



Conclusion

Overall, the prediction software we built can accurately predict the high-risk areas for the current year. The adoption of this software can greatly reduce the cost to the government in identifying such areas and help the government to quickly adopt relevant policies. Therefore, it should be adopted. However, the model also has some errors and should be adjusted with experience and field research to increase the accuracy of the results in practical application.

Reference:

Bureau, U. C. (n.d.). Identifying gentrification using machine learning. Census.Gov. Retrieved May 9, 2024, from https://www.census.gov/library/working-papers/2023/demo/SEHSD-WP-2023-15.html

Gentrification Index | Nathalie P. Voorhees Center for Neighborhood and Community Improvement | University of Illinois Chicago. (n.d.). Retrieved May 9, 2024, from https://voorheescenter.uic.edu/what-we-do/areas-of-research/gentrification-index/

Lees, L., Slater, T., & Wyly, E. K. (2008). Gentrification. Routledge.

Mayer, N. S. (1981). Rehabilitation decisions in rental housing: An empirical analysis. Journal of Urban Economics, 10(1), 76–94.