508 Final Project

**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 final project, our team developed a predicating model 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 (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

表格

描述已自动生成

For the independent variable used to predict the 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 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 Philadelphia will be used for the training model, and the data from Chicago will be used for model validation.

So, the data table will show as follows:

**Choose Model:**

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.

**Model Development:**

**Model Testing:**

**Validation:**

**Reference:**

Lees, L., Slater, T., & Wyly, E. (2013). *Gentrification*. Routledge.