Monica Flores – Group 1

Assignment 3
Applications of Machine Learning Modeling to our Project

Research Question Update:

Are Business Improvements Districts (BIDs) associated with residential gentrification? The hypothesis is when comparing that areas that were economically and demographically similar 15 years ago--some becoming BIDs some others not--the BID areas are now more gentrified areas than non-BID areas. Moreover, a second question arise: are BIDs likely to regenerate blighted areas? A final question will be if there are any cases in which economic benefits of BIDs are actually accruing long time residents instead of displacing them?, and if so what initial characteristics of the neighborhood are good predictors for a non-gentrifying BID?

The hypothesis is that depending on the initial characteristics of the neighborhood, the creation of a BID might drive either gentrification or economic benefits for the original residents. Several models will allow us to examine what predictor variables are associated positively or negatively with a desired social outcome.

Methodology:

Firstly, having groups of Census tracts that belong to a BID and all Census Tract that do not, we will use GIS to define non-BIDs Cenus Tract clustering (neighborhoods). We will have several groups of Census Tracts (neighborhoods), some of them being BIDs other not.

Then we will use an unsupervised model (K-means) for creating clusters of neighborhoods, using housing and business economic characteristics in 2000.

Next we will compare this economically-similar-in-2000 groups using time series data: 2000, 2010, and 2015[1]. For this we will use a supervised machine learning model that will use initial housing market and demographic characteristics (2000) and the formation of a BID (or not), to predict gentrification, the label(s) variable (LHS) will be: % of population that moved in the last year and in the last 5 years; changes in racial composition of the neighborhood (difference in % of white alone householder); changes in educational level of the population older than 25 years-old (%change in college educated 25+ population); changes in asking rent; and changes in household income.

Overall, we want to focus in accuracy over precision, given that the prediction of "non-gentrifying BIDs" is important for policy making: what kind of neighborhoods will actually see benefits for their long-time residents rather than generate displacement by attracting new comers?

[1] The selection of years was based on census tract level data availability. For the household and housing stock data data is available for decenial census years and ACS 5yr estimates (the first year being 2010). As we were using the 5y estimates for a time-line comparison, it made more sense to have a 5yr difference rather than retrieve data for all years between 2010 and 2015.