**Predicting Housing Price Fluctuation in Urban Neighborhoods Using Yelp and Zillow Data**

Christian Conroy and Lawrence Doppelt

By making public the data behind the online and app-based platform that publishes crowd-sourced reviews and ratings about local businesses, Yelp has created an opportunity for research minded academics to develop unique insights into the relationships between local restaurants and the greater fabric of the neighborhoods around them. Our goal is to use geocoded Yelp data to develop a model that can predict changes in housing prices in a neighborhood based on Yelp rankings and text reviews.

While many have developed models using Yelp data in order to improve the efficiency or profitability of the local businesses being reviewed, few have looked at the relationship between the rankings and reviews and broader development within specific neighborhoods. Our hypothesis is that a rise in average restaurant ratings, advent of key positive words in reviews, and increase in check-ins will lead to a detectable rise in housing prices and rents. The logic behind the hypothesis is that more and better restaurants rising up in a neighborhood will prompt other local businesses to move into the neighborhood and eventually property values will rise due to the increased access to amenities and overall development of the neighborhood.

There are several interrelated research questions we will explore in testing this hypothesis. What words signal an improvement in a restaurant’s overall rating? What kind of words in reviews for early entrants into a community signal coming increases in restaurant openings and overall investment into an area? What kind of words in reviews signal that property values and rents will increase in a neighborhood? What is the relationship between the number of check-ins in restaurants and the property values in that neighborhood? The model that we develop to answer these questions can ultimately be used by real estate investors seeking to value invest through a classic “buy low, sell high” strategy. At the same time, policymakers can use the model to better detect when development might spur the inequality effects often produced as neighborhoods gentrify and residents are pushed out by rising housing prices. By better understanding the timing of gentrification, policymakers can better target inclusive zoning, low-income development credits, and other policies aiming at mitigating the negative inequality effects of urban gentrification.

The dataset that we will use in developing the predictive model will come from Yelp and Zillow and include zip-codes and geographic coordinates, review dates, the text of reviews, the star ratings provided in a review, check-ins, median listing prices for different housing types, and median rental prices for different housing types, among other supplementary metrics.