Creating an Interactive Machine Learning Model to Track and Predict Gentrification within Major Cities Using Census Data

(Technical Topic)

Using Actor Network Theory to Understand Covid-19's Effect on the Impact and Spread of Gentrification within Poor and At Risk Communities

(STS Topic)

A Thesis Project Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia, Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Eric Guan

Fall, 2020

Technical Project Team Members: Jonathan Wen

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signature _____

Approved _____ Date _____
Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Approved _____ Date _____
Seongkook Heo, Assistant Professor, Department of Computer Science

Background on Gentrification and our Machine Learning

In 2020, the world has been swept up by a global pandemic. Organizationally, the pandemic has magnified glaring issues within the system that prevented proper responses, such as not having enough accurate data to act, in addition to lack of crisis management plans to deal with such a large scale event. (Mehta 20) At the heart of the problem, however, is an extreme disparity in how different demographics were affected by the crisis. Looking at the statistics, Latinos and African Americans are 4.6-4.7 and 1.1-2.1 times respectively more likely to be hospitalized or die from Covid. (CDC 20) This was exacerbated by a historic decline in the economy. (Solomon 20) Interestingly though, while the economy was experiencing this recession the housing market reached record highs. (O'Donnell 20, Thorsby 20)) While the market does fluctuate normally, in general when the prices of houses begin increasing in areas it is an indication that particular neighborhood is "gentrifying", (Olito 19) meaning that the area is changed economically through real estate investment and new higher-income residents moving in. This usually means a change in the demographic level - as low-income minority residents with little education are displaced and replaced with higher income residents who can afford the rising cost of living. (Maciag 2015) While gentrification is hailed by some as key to "revitalizing" cities and impoverished areas through an influx of money and new real estate developments(Buntin 15), there are several key problems that it causes. These include cultural displacement, homelessness, and increased crime. (Atkinson 04, Chong 17, Murdie 11)

Outside of the aforementioned systemic issues that spur gentrification, it is important to urban researchers to track and predict gentrification so that they can prepare for it and aid these neighborhoods. (Chong 17) However, the quality and categories of data vary greatly, ranging from zip code specific information on property owner's education level and income, to bare

bones information about property characteristics such as bed/baths. (Reades et al 19) It is crucial to select the correct data set when conducting gentrification analysis. (Stewart 20) A common problem with smaller studies is response bias, so my group opted for a more national effort that encourages more participation: The Census. From this data which is collected every decade, we can look deeply into zip code information about demographics, and cross reference these with the housing data from the past 10 years, to determine if a community is at risk for gentrification and form predictive models if it is. Our model will look at how demographics change in specific areas, with special focus on education level, median income, and race. Analyzing the changes over 10 years in areas will reveal how much each community has changed.

Creating an Interactive Machine Learning Model to Track and Predict Gentrification within Major Cities Using Census Data

The field of Machine Learning, although relatively recent, is already quite large. Machine Learning is an application of Artificial Intelligence that takes in large sets of data, and identifies patterns and draws inferences. The premise is that Machine Learning models are thoroughly built and trained, such that they grow more accurate as time goes on. As the intersection between statistics, data science, and software engineering, there are a multitude of models and algorithms already developed that account for differing data sets. (Wakefield n.d.) As previously stated, the data being utilized for this study is the Census data as well as information about housing prices from the National Association of Realtors. The goal is to localize information to specific zip codes that have previously been labeled at-risk, and formulate predictions using these localized clusters. Machine Learning is advantageous in this context compared to other regression techniques because it has shown to be better where there is a lot of "noise" in the data. (Harrell 20) Noisy data means data which has a lot of irrelevant categories which would confuse typical

vanilla regression models. Since the census has over 60 checkboxes, it was important for our team to be able to deal with the unimportant columns easily. In terms of specifically using machine learning to analyze urban centers, past studies have used Random Forest modeling as well as simple regressions and deep learning. (Illic 19, Reades et al 19)

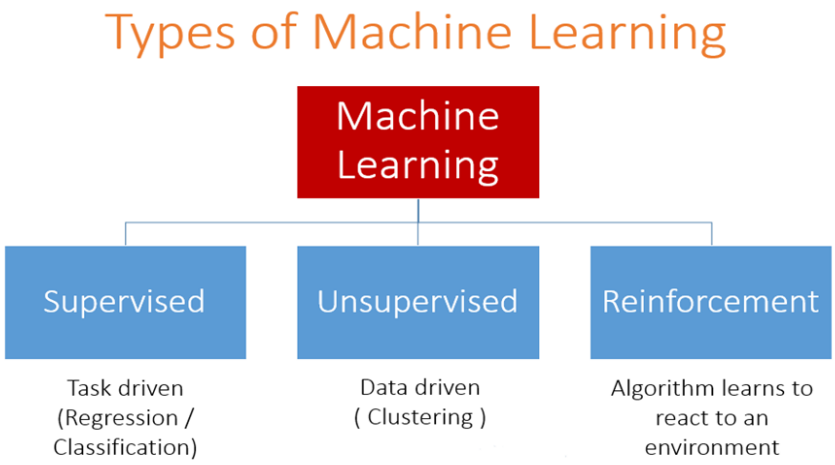

Figure 1: The 3 types of machine learning, with example applications provided under. (Sanjeevi 17) This project's model will focus on kmeans clustering, which falls under the unsupervised category.

From the different types of machine learning displayed above, each one would work for the proposed project, however the end result would change per type. For example, supervised learning is where the machine is taught by example. (Wakefield, n.d.) The operator has a set of data in which the results of an analysis are already known, so as the model takes in this data and returns results, the operator can change the model as necessary until the model can successfully recognize specific patterns. Using supervised learning to create regressions and forecasts was certainly appealing, since the proposed model will require hundreds of thousands of lines of data. However, completely accurate patterns for gentrification haven't been completely understood in many places (Richardson 19) and the numerical data returned would be more difficult to

communicate to those without a strong data science background. With the ease of understanding the final product in our minds, we decided to use an unsupervised kmeans clustering algorithm for our model. The unsupervised simply means that the model will explore data and determine relationships on its own without an operator. (Wakefield, n.d.) This works well for larger data sets as well since the model can run for as long as necessary for it to return these correlations.

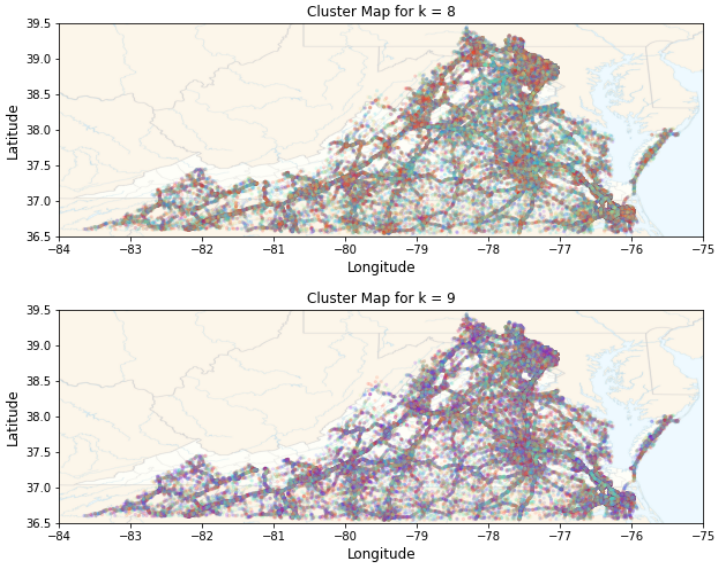

Figure 2: An example of the kmeans clustering algorithm. In this case, these 2 graphs show the results for specific k values on a graph of Virginia. Source: personal past project

A kmeans clustering algorithm means that as the model looks at data, it groups similar sets of data around averages based on preset parameters. As seen in Figure 2, these clusters can be superimposed onto a map, which is ideal because the results from the proposed project could be put onto maps of analyzed areas as an excellent visual aid.

Traditionally, when manipulating Machine Learning models, the code is in a hybrid or cloud environment which will ensure that the appropriate computational resources are allocated. (Dorard 20) Cloud means that all of the computation information is transferred over the internet, to another location where the model will be run. Hybrid means that there is a local cloud

platform in addition to the computation resources being used at another location. (“Public Cloud vs Private Cloud vs Hybrid Cloud”, n.d.) However, these models are displayed inside of the platform in which they are run, and cannot easily be saved and uploaded elsewhere. Thus, this presents a serious challenge when it comes to presenting the analysis results to the general public. In cases where there are models and dashboards embedded into websites, it is often not intuitive and unresponsive. What this means is that instead of letting users see the answers themselves, these websites instead *tell* the user large amounts of information, which goes unretained. (Birkett, 20) In order to tackle this problem, my group started with a web application which contains connectors to the cloud environment in which the model will be constructed. Then, we will conduct user interface surveys and use that to create preliminary wireframes and prototypes of the application. Members of the team have experience with the iterative design cycle, so our hope is that the application will become more and more usable closer to the date of the deliverable.

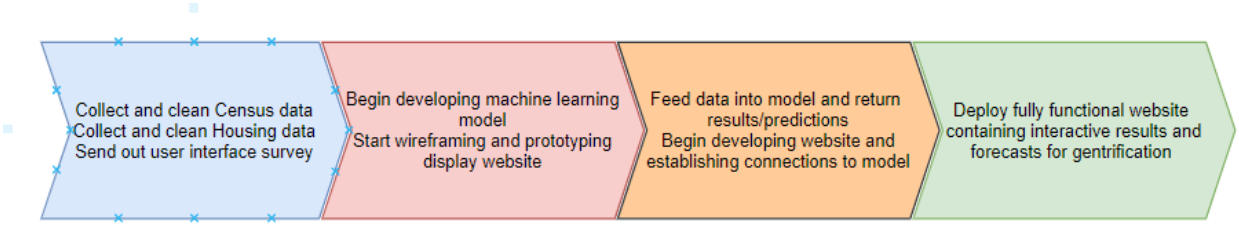

Figure 3: High level process diagram for my Capstone group’s plan to complete the technical portion

Perceptions and Implications of Gentrification within Communities

Ever since the term “gentrification” was coined in the 1960’s, gentrification has been well documented to occur in most major American cities. (Ehrenhalt, 15) The reason that gentrification happens is because city leaders naturally want their cities to have vibrant areas containing fine dining and a plethora of entertainment options, as well as safe neighborhoods and a variety of shops. However, in order to achieve this, cities don’t want to have gross disparities between certain areas within it. So, cities need to invest money into lower income neighborhoods to make them look better. (Maciag 15) In the process of renovating properties so that new tenants can move in, cities also invest and offer tax credits to higher end retailers/grocery stores. Slowly, neighborhoods begin to transform, with new middle class residents moving in, along with brand new stores. With this, the population increases, which means the city generates more revenue from taxes, which it can use in further investments to “improve” the city. (Wogan 15)

Short term, this change does cause somewhat of a resurgence of the city. For the original low income residents of the neighborhood however, the cost of living slowly increases, meaning that they can no longer afford to live there. They are pushed out, and displaced elsewhere. (Twigg 20) In fact, although this gentrification does have certain benefits, objective studies have found that there are significantly more short and long term negative effects associated with it. These include losing local and foreign cultures as immigrants seek cheaper living (Richardson 19), creating areas of extreme poverty (Chong 17), and an emphasis of a toxic power dynamic between the perceived higher and lower class (Fayyad 20). As areas gentrify, the cost of living increases further and further, until the middle class that originally moved in are displaced themselves, pushing the gentrification to other areas. (Ehrenhalt 15) This can cause long term problems, as seen in cities where gentrification has been happening for an extended period of time. (Vock 15) For example, in Austin there is now widespread *suburban* poverty. Suburban

infrastructure struggles to handle the increased low income residents, which causes huge organizational problems for the townships outside these gentrifying major cities.

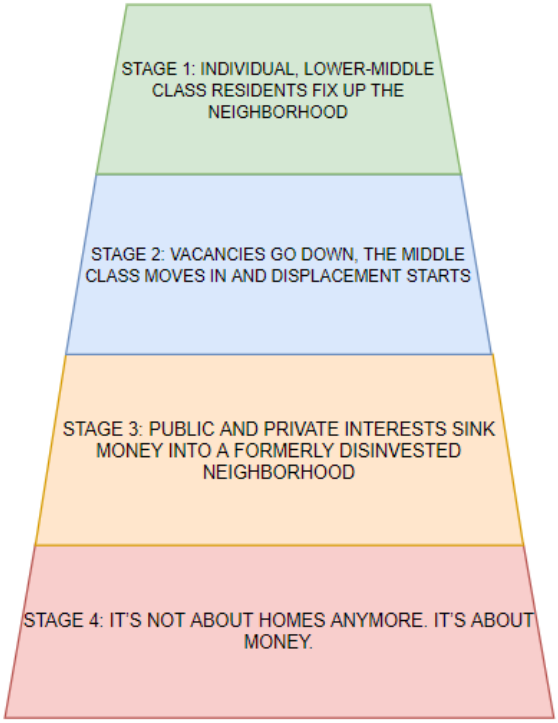

Figure 4: 4 stages of Gentrification (Twigg 20)

As seen in the above diagram, there are 4 major stages of gentrification. (Twigg 20) In the last step, parts of the city that have been gentrified have now become luxury goods, and as such these places stop being somewhere to live normal lives and instead become a place where luxury defeats the original purpose of revitalization. However, cities undergoing gentrification still need the increased revenue in order to accomplish their own planning goals. Therefore, it's important to balance economic and social interests when looking at urban planning within established communities. For our Capstone project, conveying the information found can prove difficult when explaining to those who don't have a problem with gentrification that it is in fact

occurring, and displacing residents. Likewise, in areas that we find insignificant displacement, explaining that to residents who think that the cost of living is increasing too fast will be hard. Thus, we propose an approach using ANT that will balance the needs of the actors and display the findings in relation to the data findings, instead of swaying towards any one side. Our analysis will look at a variety of factors for all actors involved, and rely on faculty feedback on the best ways to communicate our results.

Conclusion

The anticipated deliverable will be the machine learning model that can be adapted to future sets of census data, in addition to statistical insights gained from the model. This will prove extremely useful, as it can continue to be used in the future with increasing accuracy in drawing inferences. While small, the results from this project could help inform citizens who wish to see how the area they are living in has been affected over the years by gentrification. In addition, this project also provides insights into how gentrification has happened from decade to decade using census data, which will be extremely useful to urban planners when understanding where next to build projects.

References

Atkinson, Rowland (2004) The evidence on the impact of gentrification: new lessons for the urban renaissance?, European Journal of Housing Policy, 4:1, 107-131, DOI: 10.1080/1461671042000215479

Birkett, A. (2020, July 31). How To Use Interactivity to Increase Engagement (and Conversions). Retrieved November 02, 2020, from https://cxl.com/blog/interactivity-user-engagement/

Buntin, J. (2015, January 15). Gentrification Is a Myth. Retrieved November 01, 2020, from https://slate.com/news-and-politics/2015/01/the-gentrification-myth-its-rare-and-not-as-bad-for-the-poor-as-people-think.html

Chong, Emily. "Examining the Negative Impacts of Gentrification." Georgetown Law, 17 Sept. 2017, www.law.georgetown.edu/poverty-journal/blog/examining-the-negative-impacts-of-gentrification/.

Culture & COVID-19: Impact and Response Tracker. (2020, July 24). UNESCO. https://en.unesco.org/news/culture-covid-19-impact-and-response-tracker

Dorard, L. (2020, April 09). An overview of ML development platforms. Retrieved November 02, 2020, from https://medium.com/louis-dorard/an-overview-of-ml-development-platforms-df953060b9a9

Ehrenhalt, A. (2015, February). What, Exactly, Is Gentrification? Retrieved November 02, 2020, from https://www.governing.com/topics/urban/gov-gentrification-definition-series.html

Fayyad, A. (2020, June 16). The Criminalization of Gentrifying Neighborhoods. Retrieved November 02, 2020, from

https://www.theatlantic.com/politics/archive/2017/12/the-criminalization-of-gentrifying-neighborhoods/548837/

Harrell, F. (2020, September 15). Road Map for Choosing Between Statistical Modeling and Machine Learning. Retrieved November 15, 2020, from https://www.fharrell.com/post/stat-ml/

"Health Equity Considerations and Racial and Ethnic Minority Groups." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 2020, www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html.

"Healthy Places." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, www.cdc.gov/healthyplaces/healthtopics/gentrification.htm.

Ilic L, Sawada M, Zarzelli A (2019) Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. PLoS ONE 14(3): e0212814. https://doi.org/10.1371/journal.pone.0212814

Maciag, M. (2015, February). Gentrification in America Report. Retrieved November 01, 2020, from https://www.governing.com/gov-data/census/gentrification-in-cities-governing-report.html

Mehta, A. (2020, August 11). Conquering the organizational challenges of Covid-19. Retrieved October 11, 2020, from https://inform.tmforum.org/insights/2020/05/conquering-the-organizational-challenges-of-covid-19/

Murdie, R., & Teixeira, C. (2011). The Impact of Gentrification on Ethnic Neighbourhoods in Toronto: A Case Study of Little Portugal. Urban Studies, 48(1), 61–83. https://doi.org/10.1177/0042098009360227

Public Cloud vs Private Cloud vs Hybrid Cloud: Microsoft Azure. (n.d.). Retrieved November 02, 2020, from https://azure.microsoft.com/en-us/overview/what-are-private-public-hybrid-clouds/

O'Donnell, K. (2020, July 23). Housing market defies expectations amid economic turmoil. Retrieved November 01, 2020, from https://www.politico.com/news/2020/07/22/housing-market-boom-coronavirus-millennials-379084

Olito, F. (2019, September 04). 7 signs your neighborhood is gentrifying. Retrieved November 01, 2020, from https://www.businessinsider.com/signs-your-neighborhood-is-gentrifying

Reades, J., De Souza, J., & Hubbard, P. (2019). Understanding urban gentrification through machine learning. Urban Studies, 56(5), 922–942. https://doi.org/10.1177/0042098018789054

Richardson, J. (2019, October 18). Shifting neighborhoods: Gentrification and cultural displacement in American cities " NCRC. Retrieved November 02, 2020, from https://ncrc.org/gentrification/

Solomon, C. (2020, May 08). The Economic Fallout of the Coronavirus for People of Color. Retrieved November 01, 2020, from https://www.americanprogress.org/issues/race/news/2020/04/14/483125/economic-fallout-coronavirus-people-color/

Stewart, M. P. R. (2020, July 29). The Limitations of Machine Learning - Towards Data Science. Medium. https://towardsdatascience.com/the-limitations-of-machine-learning-a00e0c3040c6

Thorsby, Devon. “How Will a Recession Affect the Housing Market?” U.S. News & World Report, U.S. News & World Report, 28 May 2020, realestate.usnews.com/real-estate/articles/what-will-the-housing-market-look-like-in-the-next-recession.

Twigg, M. (2020, August 11). Revitalization or displacement: What is gentrification really? Retrieved November 02, 2020, from https://www.matternews.org/developus/gentrification-explained

Vock, D. (2015, February). Suburbs Struggle to Aid the Sprawling Poor. Retrieved November 02, 2020, from https://www.governing.com/topics/health-human-services/gov-suburban-poverty-gentrification-series.html

Wakefield, Katrina. “A Guide to Machine Learning Algorithms and Their Applications.” SAS UK, www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html.

Wogan, J. (2015, February). Why D.C.'s Affordable Housing Protections Are Losing a War with Economics. Retrieved November 02, 2020, from https://www.governing.com/topics/urban/gov-washington-affordable-housing-protections-gentrification-series.html