**Proposed Study: Using Nextdoor Posts to Measure Gentrification in Tacoma, WA**

**August Posch**

**Northeastern University**

**posch.au@northeastern.edu**

**Proposed Study: Using Nextdoor Posts to Measure Gentrification in Tacoma, WA**

**August Posch**

**Northeastern University**

**posch.au@northeastern.edu**

**Abstract**

This project explored gentrification in Tacoma, Washington and proposed a novel way to measure the emotional effects of gentrification. Posts from the Nextdoor social network will be scraped and added to a database, and natural language processing (NLP) will be used to describe the sentiments, emotions, and topics of the posts. These outputs will be compared with existing census data on racial changes as well as rent and property price appreciation, and also will be compared with resident interviews. The intent is to establish a useful database that may be used by future researchers to further understand all the dimensions of gentrification.

**Introduction**

This study develops a novel way to measure gentrification. In the past people have often equated gentrification with rising property prices – while this is one way to measure gentrification, many scholars call for gentrification to be measured by the pain it causes existing residents of a neighborhood. We aim to determine how we can utilize Nextdoor posts and state-of-the-art natural language processing techniques to measure the human dimensions of gentrification. We use Tacoma as our area of study, as Tacoma is, by some measures, the most quickly gentrifying city in the United States. Our new Nextdoor-based processing techniques provide a way to measure displacement within Tacoma (switching neighborhoods) as well as displacement out of Tacoma altogether. In addition to displacement, the new system will also measure neighborhood change with regard to the types of interactions that residents of each neighborhood engage in, and the sentiments and emotions associated with these interactions.

**Background**

**Rising home prices and displacement in Tacoma**

Since 2011, home prices have increased an average of 17% per year. Properties regularly see bidding wars. As of February 2020, Tacoma was the fastest rising housing market in the United States (Long 2020). Although locations like Boise, Idaho have taken the “fastest rising” title due to unique conditions during the pandemic, the Tacoma market hasn’t slowed down either, remaining consistent at an average of 17.7% increase in home value in mid-2021 versus mid-2020 (Peterson 2021). Recent reports from November say there’s no slowdown in sight, as the average home value increase is 21% and supply will be constrained until at least March of 2022 (Cockrell 2021).

The booming market presents a challenge to Tacoma residents looking to buy homes, and a financial boon to those who have held onto homes they already own during the rising market. Now we turn to renters – for some of them, the situation is dire.

According to 2021 work by researcher Tim Thomas, the neighborhoods in Washington where people are at greatest risk of losing their housing are in Kent, Federal Way, and Tacoma (Brownstone 2021). An evictions study covering 2013-2017 found that evictions are pervasive in Washington in general, with 1 in 55 adults facing an eviction during that time frame. The evictions problem is colored by geographic and racial factors as well: In Pierce County, where Tacoma sits, 1 in 6 black adults faced an eviction during the 2013-2017 timeframe. Furthermore, evicted people were found to be legally underrepresented, with only 8% of them having counsel for their trial. Finally, the eviction problem is even more widespread if we factor in informal evictions, in which the landlord might stop fixing appliances or performing other landlord duties, or the landlord might raise rent to unpayable levels; in cases where the tenant chooses to move out before the case goes to court, it is not a formal eviction and therefore not counted in the above study (Thomas et al. 2019).

Why is Tacoma experiencing rising prices and evictions? Seattle and regional dynamics are the main cause. Seattle is located 30 miles north of Tacoma and is one of the most expensive real estate and rental markets in the country. As prices continue to rise in Seattle and people can no longer afford to live in the Seattle neighborhoods they grew up in, they look to Tacoma. Even restaurants that can no longer afford Seattle rent move to Tacoma, so many of their regular customers who have moved can feel right at home in Tacoma. For many Seattleites, Tacoma feels like home, but with lower rent: Rachel Collins grew up in Central District, Seattle and appreciates being surrounded by black people who are from her new home of Hilltop, Tacoma. For others from Seattle and elsewhere, Tacoma is a convenient place to live while working a high-paying job in Seattle (Long 2020).

Longtime Tacomans worry that the Seattle influx is changing their neighborhoods so as to become unrecognizable. This can come in many forms. One longtime North End resident complains that she always used to go to the grocery store wearing a hoodie and pajama pants, but now she draws looks from well-dressed newcomers. A resident explains that what used to be a lively neighborhood now feels like a bedroom community, as new residents leave before 6am to beat the traffic to Seattle, and get back from their workday at 7pm to walk their dogs in the dark. A Hilltop resident mourns the loss of black-owned small businesses as this historically black neighborhood loses a third of its black residents in recent years (Long 2020).

Changing neighborhood character is not the only effect of the regional system, as regional dynamics determine how many people are displaced and where. In the Washington evictions study, Thomas found that every county had roughly steady numbers of evictions yearly since 2004, except for King County where Seattle sits, which saw a significant decrease in evictions between 2006 and 2012 (Thomas 2019). In Thomas’s analysis, most of the people who were at high risk of eviction already got evicted in King County and moved away, and high rents kept eviction-susceptible tenants from moving to King County since 2012. This line of thought follows an existing Urban Institute study that found a similar dynamic in the District of Columbia (Brennan 2018).

Thus, displaced people of both low incomes and middle incomes are pushed out of Seattle. We have vulnerable people pushed to Kent and Federal Way and we also have middle-income professionals who cannot afford to live in Seattle, who choose to move to Tacoma. When they move to Tacoma, they push longtime Tacoma renters into a more vulnerable position.

We’ve hinted at different dimensions that people associate with gentrification: rising home prices and rents, displacement, and changing neighborhood character. Rising prices are measured quantitatively, and displacement can be measured quantitatively as well; however, when it comes to the life-altering psychological toll of displacement, and when it comes to changing neighborhood character, we typically lean on qualitative tools, such as interviews. These qualitative measures of pain are important, but often resource-intensive to measure of limited in geographic or temporal scope. Later in this paper we will develop a quantitative measure of pain based on Nextdoor posts, and we hope this measure will enrich the discourse around gentrification.

**Definitions of gentrification**

But what is gentrification? The term was originally defined by Glass as a process happening in London, England, whereby the middle classes were overtaking working-class neighborhoods. The emphasis was on the class aspect of gentrification, and it was viewed as a discrete process by which the middle-class newcomers gradually invade and take over working-lass quarters (Smith 2002). Smith sees gentrification as more nuanced, with waves of gentrification happening due to different conditions in different decades, and he identifies five dimensions of gentrification that explain its proliferation since the 1950s. These dimensions are the role of the state (e.g. acting as part of the market by providing subsidies, rather than regulating the market); penetration of global capital (e.g. an Israeli developer funded by a European bank builds a condo in New York City); political opposition (e.g. squatter movements, homeless movements); geographic dispersal (e.g. previously "new" outer areas of the city now face disinvestment as central areas are gentrified); many sectors participate (e.g. governments, private-market finance, combinations of the two). Smith also connects gentrification to Sassen’s New Urbanism, the theory of changing scales and inequalities of urbanism due to globalization (Smith 2002). In Tacoma, we see some of Smith’s insights at play, as rescaling leads well-off residents of the region into the cities and pushes the poor out; Tacoma is simultaneously a desirable city destination for those who newly cannot afford Seattle, and a longtime hub of working-class residents of the region.

Elliot-Cooper et al. emphasize that when talking about gentrification, it is important understand the phenomenological or affective dimensions of displacement, and the anger and despair that is inherent to its experience (Elliot-Cooper et al. 2020). This view of gentrification, and the comparatively sparse scholarly work on the emotional and displacement dimensions of gentrification, led us to focus on these dimensions in our current study.

Overall, the term “gentrification” has been defined variously over the years, and even within single definitions the process can unfold variously in different times and places. However, for now we can safely assume that gentrification is a process by which people of more privileged class and financial means move to neighborhoods previously occupied by less-privileged residents, causing displacement and financial and emotional pain. This study aims to understand part of the emotional pain of gentrification and how our new measures compare to established measures of gentrification.

**Brief history of Tacoma**

The ancestors of today’s Nisqually, Puyallup, Squaxin, Steilacoom and Muckleshoot Indians were the first settlers of the land where Tacoma now sits. In 1792, the English sea captain George Vancouver sailed in as far as present-day Seattle and sent his lieutenant Peter Puget south in search of a Northwest Passage, and these were the first Europeans to arrive at what is now Tacoma. The first permanent settlement was made in 1832 as an outpost for trading beef, butter, and cheese. In 1852 the first sawmill was opened in what became a booming lumber industry. In creating Pierce County in 1852, the territorial legislature scenic beauty, mild and “salubrious” climate, coal and precious metals, and powerful rivers. These visions of possibility would come to fruition in Tacoma over the years.

Tacoma was founded in 1872 in hopes of riches that a forthcoming railroad would bring. The next 20 years were boom years due to lumber, coal, iron, carrots, potatoes, and shipping. An economic depression followed, but by 1901 industry was helping Tacoma recover. World War I brought an industrial boom related to the lumber and shipyard industries, as well as a new U.S. Army base nearby. While the Great Depression was tough, New Deal programs and military spending helped. Military bases were established and expanded, and World War II increased demand for Tacoma shipyard work. Local Japanese residents were held at an internment camp nearby. The African American population grew as veterans and shipyard workers settled in Tacoma.

After World War II, Tacoma underwent urban renewal and suburbanization. In the 1970s the Port of Tacoma became an important link to Alaska and Asia with the rise of containerization. Tacoma’s metal and farming-related industries connected to the port to keep the economy strong. Tacoma was considered a working-class city, and other residents of Washington often viewed Tacoma with contempt, citing the “Tacoma Aroma” stench coming from its industries. Nevertheless, in recent years people from around the state and around the country are moving to Tacoma as a result of the region’s strong economy and Tacoma’s comparatively lower cost of living as compared to Seattle. (Pierce County 2021; Wilma and Crowley 2003)

**Urban Theory at play in Tacoma**

Globalization plays a role what Tacoma is experiencing today. Sassen broke new ground in the theory of global cities in 1996, as she emphasized the experiences of the workers who enable the global economy to thrive (Sassen 1996). Particular elite cities control the forces of globalization and often have more power than the states where they sit, yet the workers that enable the accumulation of global power must We see this at play in the Tacoma region with the SeaTac international airport. Low-wage workers enable this giant transportation hub to connect the high-tech hub of Seattle to the global economy; however, when external events, like a pandemic, force the airport to reduce its operations, the workers suffer (Long 2020). The case of SeaTac is all the more concerning because it is not built in the middle of an existing city – rather, a city built up around it expressly based on its transportation-related industries. This has put the residents at greater risk due to concentrated exposure to risk in a single sector of the economy.

Another way to conceptualize globalization is in terms of gateway cities (Short et al. 2000). Short et al. use the term “gateway city” to refer to cities that “act as a gateway for the transmission of economic, political and cultural globalization.” This way, they hope to focus on cities below the top echelon. Writing in the year 2000, they focus on Seattle as a gateway city for the U.S., but now 20 years on from their study, we see Tacoma as a gateway city due to its proximity to Seattle, which has become a global center in the high-tech economy.

The state of Massachusetts has adopted the term “Gateway Cities” to refer to cities meeting certain objective characteristics: population between 35,000 and 250,000, educational attainment rate of “bachelor’s degree or higher” at below the state’s rate, and median household income below the state median. Although Massachusetts’ Gateway Cities were the inspiration for this study, we can translate the concept to the state of Washington, using Washington state numbers as our reference point. We find that Tacoma meets these criteria, with a median household income of $70,411 which is less than Washington state’s $78,687, and a rate of bachelor’s degree or higher (among population 25 years and older) at 33.1%, which is less than Washington state’s 37.0%. Tacoma has a population of 219,346 within city limits (U.S. Census Bureau 2021). Washington has no officially-designated Gateway Cities program of its own, but perhaps our work on Tacoma will reveal the need to initiate a similar program in the state of Washington.

So far, we’ve tapped urban theory about globalization and brought it down to the level of gateway cities. However, we can usefully go do a deeper level – the neighborhood level – as we investigate these phenomena. As we saw in our background research, gentrification is often described at the neighborhood level, and it brings about neighborhood change. Galster describes 10 dimensions of neighborhood including social-interactive characteristics that are especially relevant to what we aim to measure: local friend and kin networks, degree of interhousehold familiarity, type and quality of interpersonal associations, residents’ perceived commonality, participation in locally based voluntary associations, strength of socialization and social control forces (Galster 2001). In Tacoma gentrification appears to be impacting these types of attributes, and with our text processing procedure we hope to capture these changes and make them legible. Coulton et al. note that residents’ definitions of neighborhood are often different than state-prescribed or researcher-prescribed conceptions (Coulton et al. 2001); we will take this into account as we identify neighborhoods in our study.

Nextdoor is an online social network in which people can only interact with other residents of their own city. All users must use their real name on the site. Users create posts and comment on each other’s posts; they can see their neighbors’ posts in their homepage feed. Users often share local events, trade contractor recommendations, warn each other of crime, and talk about other local topics. Latham and Layton define social infrastructure as the “networks of spaces, facilities, institutions, and groups that create affordances for social connection.” (Latham and Layton 2019). Therefore, we see Nextdoor as social infrastructure. If nothing else, our study can be thought of as discovering the ways residents use this Nextdoor infrastructure to create social connection.

**Current Study**

This is a mixed-methods (qualitative and quantitative) study that aims to create a useful way to measure gentrification. The database will be a new source of knowledge about gentrification patterns that will be accessible to residents and researchers. This lays the groundwork for future studies to dig into other hidden patterns. We will also do an Exploratory Data Analysis (EDA) on the newfound data in this paper, in order to see how it connects to and differs from existing measurements of gentrification, and to lay the groundwork for future studies.

The study will create a database of Nextdoor posts in Tacoma, WA. Posts will have identifying attributes of user, neighborhood, and date. Natural language processing (NLP) techniques will be used to tag each post with sentiments (positive and negative), and emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust). Each post will have an association score for each sentiment and each emotion, allowing us to determine the dominant sentiment and dominant emotion of each post. Once the Nextdoor database is created, we will do an Exploratory Data Analysis to analyze changes in sentiments and emotions over time, by neighborhood. We will also analyze displacement over time. This study will also conduct resident interviews and draw on census data, which will serve to calibrate our natural language processing outputs against existing commonly-used measures of gentrification. We will record which NLP appeared most useful and leave it to future researchers to further leverage that data source.

Several academics have highlighted the importance of local knowledge in urban theory, in different ways. Participatory Action Research is a way of running a study that engages local residents from the outset of the study, so that residents and academics together mold what gets studied and how. This process led residents to be more satisfied with outputs in recent studies of displacement patterns in Massachusetts (Daepp et al. 2021) and neighborhood health measures in the greater Boston area (Binet et al. 2019). The disadvantages of Participatory Action Research are that it requires long time scales and flexible funding, and that its results are not as scientifically robust. The advantages are that it remedies power imbalances between the academics and the subjects of the study - the residents – and provides residents with solutions that are most meaningful to *them*. Our proposed study will not engage local residents the whole way through, but we will conduct resident interviews in order to get an initial calibration of our natural language processing models with respect to gentrification, and we do hope to help academics imagine the lives of residents more complexly by establishing our database of Nextdoor posts. We would also like to make the database accessible to residents in whatever form is easiest for them to use, similar to how Daepp et al. created a mapping tool in the Massachusetts displacement study.

Scott makes the case for caution in well-organized bureaucrats imposing their idea of supremacy of scientific (epistemic) knowledge. In his phrasing, the “combination of the universalist pretensions of epistemic knowledge and authoritarian social engineering” poses a danger to human societies, including cities. He points out that when well-meaning science-oriented people try to impose a utopian organizational scheme on a city, it is in fact the metis, the unplanned practical knowledge, which functions as oil to make the whole thing work. This was the case in Brasilia, a city that was overplanned but saved by the construction workers’ willingness to throw out the plan and use common sense (Scott 1998). Scott’s warnings are relevant in the current study for multiple reasons. On the one hand, our database fundamentally values metis, as it centers a local communication channel, Nextdoor where people are solving practical problems and engaging with their communities. On the other hand, out database attempts to translate this metis into a tool for technocratic urban theorists to scientifically understand the process of gentrification. Thus, we should repeatedly remind our technocratic selves that we cannot impose order on everything. Nextdoor posts are not emotions, and the posts do not reveal everything about the actions themselves: perhaps the neighbor who angrily ranted about your holiday decorations lost their sibling in a holiday-decoration-related accident last year. We cannot know for sure what is going on in any person’s psychology, and we aim not to reduce people to their Nextdoor posts; however, we hope that the Nextdoor posts in our database provide a way for scientists to understand the emotional and practical aspects of people’s lives, and provide a foundation for understanding and resolving practical problems in the city.

Exploratory Data Analysis is a common technique in the field of data science, and in the context of urban theory, we see it as a Pragmatist approach. Exploratory Data Analysis consists of generating summary statistics, visualizations, and sometimes basic models in order to better understand the data. For example, given an income dataset for a city, the data scientist might create a histogram of incomes by age group or a bar chart of incomes by race, or visualize average incomes by neighborhood on a map. These views are meant to describe the dataset well enough in order to prompt relevant questions that could be answered through modeling. Exploratory Data Analysis is fundamentally about hypothesis generation, not confirmation (Wickham 20xx). The Pragmatist approach-based paradigm of urban science aims to solve real-world problems via using theory generation and theory confirmation, via quantitative and qualitative data. Seeing as Exploratory Data Analysis is quantitative theory generation geared towards solving problems, we see as part of the Pragmatist paradigm.

**Methods**

**Census Data**

We will collect census data from the U.S. Census Bureau’s website. Previous studies that sought to connect theory to empiricism modeled gentrification based on rate of price appreciation of single-family properties (Keenan et al. 2018), so we will focus our empirical study on similar measures. Because our background research revealed many people affected by gentrification were renters, we will collect rent data. Because our background research identified racial change as a potential key factor of gentrification in Tacoma, we will also collect census data on race. Because existing research has highlighted the vulnerability of black residents of Tacoma (Thomas 2019) and specifically the drastic changes in the Hilltop neighborhood for black residents (Long 2020), will focus on black racial statistics in this study. We acknowledge there are problems with how the census measures race (Prewitt 2017), but for now, we will rely on these imperfect data in order to tie our findings to existing research based on the census. We encourage future researchers to pick up the radical work of changing how we measure race in future studies.

The most relevant census data products for us come from the American Community Survey 5-year estimates. These are the only data available at a fine enough geographical granularity for us, but because they come from surveys that only happen every five years, accordingly they suffer in their margin of error – for some products at the census tract level it’s as much as 10% margin of error – but the bureau does release new estimates every year, so as long as we keep in mind the margin of error these are useful to us. Furthermore, we will later combine census tracts when we identify neighborhoods, which will decrease the margin of error. Mathematically our margin of error is halved every time we quadruple the size of our sample; thus if we combine four census tracts that each have a 10% margin of error, we reduce our margin of error to 5%, which may be small enough to read into trends, depending on the data.

For race data, we will use Percent Black or African American, which refers to on the number of people in each census tract in each year. For rent data, we will use Median Gross Rent By Bedrooms, which provides values by year by census tract for one-bedroom, two-bedroom, and three-bedroom renter-occupied housing units. The median is calculated based on housing units, not people. For property values, we will use Median Value (Dollars) which refers to owner-occupied housing units in each census tract in each year. Again, the median is calculated based on housing units (U.S. Census Bureau 2021). We will convert our rent and property value data into rates of annual rates of appreciation using the appropriate mathematical formula, in order to keep our work comparable to the existing research by Keenan et al.

**Resident Interviews**

We will conduct interviews with residents around the city asking about people’s experiences of gentrification. We would ask them to tell stories of how their neighborhood has changed in the last many years, what kinds of people are in the neighborhood, the types of interactions that typically take place in the neighborhood, and how it feels to live in the neighborhood and how that has changed over time. We are trying to get a sense of residents’ direct experiences of gentrification.

Given that the current author team is an expert in data science rather than on-the-ground qualitative data collection techniques such as interviews, we would recruit a social scientist to design and/or conduct the resident interviews.

**Nextdoor Data Collection**

Nextdoor data will be scraped from the Nextdoor website. Here we will address the basics but leave some of the technical details unaddressed, because the process may require problem solving along the way as we discover the underlying design of the Nextdoor website. The overall idea is to log in, scroll through years of posts, and use web scraping code to write the posts into our database. We aim to collect at least four years of posts in order to identify trends over time; if memory and runtime constraints allow, we would like to collect data back through the time when Nextdoor became reasonably popular in Tacoma. In order to log in to the Tacoma Nextdoor page, we would first need to partner with a Tacoma resident who already has an account; otherwise we would need to move to Tacoma, prove our residence to Nextdoor, and use our own account. In scraping the posts from the website, we would keep track of the user, the user’s neighborhood, the date and time of the post, and the text content of the post. Users will be assigned User IDs for the purpose of our database to obviate privacy issues. Depending on our web scraping techniques, we may need to clean the text data to get rid of extraneous HTML tags or other formatting. We would compile all the fields (the user, the user’s neighborhood, the date and time of the post, and the text content) into a database.

**Natural Language Processing**

We will supplement our database using natural language processing (NLP) techniques on the text of the posts. We use the NRC Emotion Lexicon to assign sentiments and emotions to the Nextdoor posts (Mohammad and Turney 2013). For each of 14 thousand words in the English language, the lexicon contains sentiments (positive or negative) with association scores of either 0 or 1. For each word it also contains emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) with association scores of not associated, weakly, moderately, or strongly associated. The ethical and methodological issues around creating such a lexicon are interesting and important, and Mohammad’s 2020 paper covers it in detail; for this Tacoma study, it is enough to know the goal of the lexicon was to determine how speakers of a language perceive the emotion associations of words, through crowdsourcing and majority voting (Mohammad 2020). The lexicon reflects the cognitive biases of the crowd that participated in creating it; as such, we must be careful not to use our database in a way that amplifies any harmful biases.

We will apply the lexicon to the Nextdoor posts in a way that is standard in natural language processing. To obtain the sentiment or emotion of a particular post, first we remove stop words (less-meaningful words such as “the” “and”, “with”, etc.) from the text content of the post. Then, each word gets scored for its sentiments and emotions according to the lexicon, and the sentiments and emotions get aggregated to the post level. This leaves us with the two sentiments, the eight emotions, and association scores for all sentiments and emotions, for each post. We would add all these as fields in our database.

We will also use natural language processing to identify the main topic of each Nextdoor post. Based on our prior experience, Nextdoor posts tend to fall into certain topics, such as buy/sell, local recommendations, and local crime. We would start by hand-labeling some training data (say, 100 posts) with the topic that we consider it to fall into. From there, we would develop a Naïve Bayes model to determine the topic of each post, based on our training data. If the Naïve Bayes model cannot achieve high enough accuracy, or if we as researchers determine that the topics cannot easily be labeled for some other reason, then we can abandon the Naïve Bayes model and label the topics by hand, which would simply take more human power to get done. No matter how we produce the topics, we would add the topic as a field in our database, calling it “supervised topic” as this is a supervised machine learning method, meaning it is based on humanly made labels.

We would also use an alternative approach to topics - a Latent Dirichlet Allocation (LDA) model to assign topics to posts. In this model, each post is viewed as a mixture of a small number of topics, and each topic is defined as a probability distribution of words. A human decides on two parameters: the number of topics in the corpus, and a sparsity parameter that determines how many different topics are prominent in each post. Given these parameters and the text data, the algorithm mathematically determines the most likely assignment of topics to posts. In our context, we would probably set the sparsity parameter to a commonly used value, and we would try a few different values for number of topics. We would then look at the resulting topics and see if they make sense; for example, if it looks like two of the topics are made from essentially the same set of words, then we would try reducing the It is valuable to construct LDA topics in addition to, because this is an unsupervised machine learning method, meaning it doesn’t rely on human preconceived notions of what the topics should be – here the human element only comes into play in deciding if the number of topics makes sense. This approach can help eliminate human cognitive bias and help us as researchers see topics that we wouldn’t have even seen or considered otherwise. We will add these topics to our database as well, calling this field “LDA topic”.

**Identifying Neighborhoods**

We strive to carefully identify neighborhoods that align with residents’ experiences but also provide us a great enough geographical area to use statistically significant data – we referenced this problem in the Census Data section. Therefore, we can only decide how to define our neighborhoods of interest *after* we have collected our Nextdoor data. Based on the public-facing side of the Nextdoor website, there are 152 Tacoma neighborhoods on Nextdoor. Our background knowledge tells the number of commonly-understood neighborhoods is much fewer. We expect our Nextdoor data to reveal that certain neighborhood names appear much more commonly as neighborhoods to which users belong; we would like to use these common neighborhoods for our analysis, as greater numbers of people give us more significant results. We acknowledge that this approach marginalizes people living in uncommon areas of the city, but for the purpose of this study we will focus on establishing a useful database for the most common neighborhoods in Tacoma.

Whatever neighborhoods we choose, we will aggregate our census data up from the census tract level to the neighborhood level if there are multiple census tracts in a neighborhood. Although the borders may not quite match, we will do our best, and we will acknowledge any differences while analyzing our results. We will also reallocate any smaller Nextdoor neighborhoods that fit entirely within larger chosen Nextdoor neighborhoods so that these users count toward the larger neighborhood. Finally, we will tag our resident interviews with neighborhoods according the decided-upon scheme.

**Exploratory Data Analysis**

First, we will do exploratory data analysis on the Nextdoor dataset itself, using our natural language processing outputs. We will look at most-associated emotion of posts by neighborhood by year, using bar charts. For example, one bar chart would show the most-associated emotion of posts for the year 2019 for the Hilltop neighborhood (assuming Hilltop was one of our chosen neighborhoods). Similarly, we will look at most-associated sentiment of posts by neighborhood by year. Similarly again, we will use bar charts to look at supervised topic by neighborhood by year, as well as LDA topic by neighborhood by year. We will aggregate up to the all-neighborhood level and identify city-wide trends over the study years.

Next we will look at displacement. Here, we are using people’s Nextdoor posting patterns as a proxy for who moved where and when. To study moves within the city, we will filter the dataset to users who had two or more neighborhoods over the course of our study period. Note that for this to be possible, Nextdoor needs to have tracked neighborhood changes over time in a way we could scrape from the website – if this is not the case, then, we would need to scrap this displacement analysis. Assuming this data is available, we would record the dates at which we know users departed each neighborhood, as well as the dates at which we know users arrived in each neighborhood. We would aggregate these by year or month – some large enough time interval to give us large-enough samples – and compare these arrival counts and departure counts with the overall number of posts in a given neighborhood in a given timeframe.

Finally, we will bring in the census datasets and the resident interview data. The aim is to find which aspects of the Nextdoor posts might be associated with our other data sources and which are not. For example, we will aggregate our Nextdoor displacement data to the neighborhood-year level, and create a scatterplot of Nextdoor departures versus appreciation of median rents. This would reveal whether residents appear to leave a neighborhood (based on their Nextdoor neighborhood) whenever and wherever rents are appreciating the most. As another example, we will connect back to our resident interviews and summarize the most common emotions and the tenor of neighborhood interactions coming out of each neighborhood. We will then look at our bar charts of NLP emotions for the current year, and see if the emotions match – perhaps people on Nextdoor tend to be more bitter than their non-posting neighbors, or perhaps joy about certain topics on Nextdoor shows up as fear about the same topics in resident interviews. This is a way of calibrating our NLP emotions dataset to the most recent year’s Nextdoor posts, which will set the stage to extend these inferences back in time to earlier eras of Nextdoor posts.

**Results and Discussion**

We produced a database of Nextdoor posts in Tacoma, and performed Exploratory Data Analysis to discover the ways that our database might be useful. Because this study was aimed at establishing a useful tool for researchers, we have already commented on some contingencies based on what we discover in the Methods section above. For example, we may discover that Tacoma Nextdoor users mostly only reside in one of three large neighborhoods, which is an interesting result in itself, but also determines how we must aggregate data when we compare property price appreciation with number of joy-associated Nextdoor posts.

Here is one hypothetical finding. Suppose we find that the emotion of anticipation is strongly associated with decrease in percent black in most neighborhoods. This would prompt further interesting questions about the nature of the association. Do we have theoretical reasons to suspect that one is causing the other? Can we trace both trends to an increase in property price appreciation? Also, who is on Nextdoor – are there comparable racial proportions on Nextdoor as compared with the overall set of people in the neighborhood itself? We would also be interested in what kind of anticipation is occurring – we might turn to our resident interviews to see if people mention a sense of waiting for something good to happen and what that good thing might be, or waiting for impending destruction of their community. Also, if there is a neighborhood where our trend does not apply, we would use our data to ask ourselves what stands out about that neighborhood and what makes it an outlier to the overall trend. Using these lines of reasoning, we will learn about the dynamics of Tacoma neighborhoods with richness, and we will simultaneously get a sense of the urban science capabilities of our Nextdoor dataset.

**Conclusion**

Given that this is only a proposed study, we cannot draw any conclusions at this time. However, we anticipate future directions to go with this study. One direction is to do more thorough research on Nextdoor specifically and incorporate fields in our database that are relevant to existing scholarship. Another useful direction is to draw on more census data, continuing to perform Exploratory Data Analysis to identify which existing census products are most closely associated with our NLP data. Lastly, the purpose of this study was to establish a database that would be useful for future research on gentrification, so we hope that other researchers are inspired to connect other methods, theories, and data sources to our new database.

**References**

Brennan, Maya. 2018. “DC’s Eviction Filings Are Down. Are High Rents to Blame?” *Urban Institute.*

Brownstone, Sydney, 2021. “To predict who’s most likely to become homeless, look to South King County, researchers say.” *The Seattle Times* July 14, 2021.

Cockrell, Debbie. 2021. “No slowdown in sight for hot real estate market. Lack of inventory forecast into ’22.” *The News Tribune.* Nov 8, 2021.

Coulton, Claudia J., Jill Korbin, Tsui Chan, and Marilyn Su. 2001. “Mapping Residents’ Perceptions of Neighborhood Boundaries: A Methodological Note.” *American Journal of Community Psychology* Vol. 29, No. 2: 371-83.

Daepp, Madeleine I. G., Andrew Binet, Vedette Gavin, Mariana C. Arcaya, and The Healthy Neighborhoods Research Consortium. 2021. “The Moving Mapper: Participatory Action Research With Big Data.” *Journal of the American Planning Association,* August, 1–13.

Elliott-Cooper, Adam, Phil Hubbard, and Loretta Lees. 2020. “Moving beyond Marcuse: Gentrification, Displacement and the Violence of Un-Homing.” *Progress in Human Geography* 44 (3): 492–509.

Galster, 2001. “On the Nature of Neighbourhood.” *Urban Studies*, Vol. 38, No. 12, 2111–2124, 2001

Keenan, Jesse M, Thomas Hill, and Anurag Gumber. 2018. “Climate Gentrification: From Theory to Empiricism in Miami-Dade County, Florida.” *Environmental Research Letters* 13 (5): 054001.

Latham, Alan, and Jack Layton. 2019. “Social Infrastructure and the Public Life of Cities: Studying Urban Sociality and Public Spaces.” *Geography Compass* 13 (7): e12444-n/a.

Long, Katherine Anne. 2020. “As Seattleites and their money flow south, Tacoma residents grapple with changing neighborhoods.” *The Seattle Times.* Feb 17, 2020.

Mohammad, Saif. 2020. “Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons.” *arXiv preprint arXiv*:2011.03492.

Mohammad, Saif and Peter Turney. 2013. “Crowdsourcing a Word-Emotion Association Lexicon.” *Computational Intelligence*, 29 (3), 436-465, 2013.

Nextdoor. 2021. “Tacoma, WA.” *Nextdoor.*

Peterson, Josephine. “Pierce County average home value continues to soar in 2021, according to assessor.” *The News Tribune.* June 27, 2021.

Pierce County. 2021. “History.” *Pierce County.*

Prewitt, Kenneth. 2013. “What is Your Race? The Census and Our Flawed Efforts to Classify Americans.” Princeton University Press.

Sassen, Saskia. 1996. “Whose City Is It? Globalization and the Formation of New Claims.” *Public Culture* 8 (2): 205–23.

Scott, James C. 2020. "Thin Simplifications and Practical Knowledge: Metis." In *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, 309-341. Yale University Press.

Short, John Rennie, Carrie Breitbach, Steven Buckman, and Jamey Essex. 2000. “From World Cities to Gateway Cities: Extending the Boundaries of Globalization Theory.” *City* 4 (3): 317–40.

Smith, Neil. 2002. “New Globalism, New Urbanism: Gentrification as Global Urban Strategy.” Antipode 34 (3): 427–50.

Thomas, Timothy, Ott Toomet, Ian Kennedy, and Alex Ramiller. 2019. “The State of Evictions: Results from the University of Washington Evictions Project” *University of Washington* Feb 7, 2019.

U.S. Census Bureau. 2021. *The U.S. Census*.

Wickham, Hadley and Garrett Grolemund. 2017. *R for Data Science.* O’Rielly.

Wilma, David and Walt Crowley. 2003. “Tacoma – Thumbnail History.” *HistoryLink.org.*