2020

Predicting Neighborhood Change in Detroit

A DATA AND ETHICAL ANALYSIS OF DATA-DRIVEN POLICYMAKING

Author: Alissa Graff, MPP/MSI 2020

Thesis Committee Members:
Dr. Kentaro Toyama, Chair
Dr. Joy Rohde, Member

Master's Thesis Option Program:
University of Michigan School of Information

Acknowledgements

I would like to acknowledge my thesis committee for their insights and unwavering support and patience throughout this tumultuous year. Your knowledge, wisdom, candor, and sass were critical in helping me achieve this feat.

Similarly, I would like to thank my steadfast partner, Kyle Richardson, for his constant love and encouragement when things got tough. And for taking care of me when I got hit by a tree.

Mom, Dad, Josh, all of The Aunts, Ailey, The Baldwin House, the Detroit Ultimate Frisbee League Board of Directors – thanks for putting up with me and being a steady presence throughout this process. Mister, you're the best cat a gal could ask for.

And finally – the class of 2020 – even though we might not have the end that we wanted, we worked hard to get here, and I hope we take that with us no matter what.

Contents

Introduction

Detroit is currently experiencing an investment renaissance in the 7.2 square miles of the greater downtown area ("the 7.2"). Throughout the remaining 131.6 square miles, some neighborhoods have become "nodes" of real estate and social investment, while many other neighborhoods still struggle to attract the eye of investors (Thompson, 2017b). With a city-wide poverty rate of 35%, insufficient public transportation and limited access to jobs, most residents in the neighborhoods have not benefited at the same rate as the downtown core during this time of resurgence in the city (Jackman, 2014, 2016; Pandey, 2019).

Attempting to jumpstart investment in 10 "tipping point" neighborhoods, the Strategic Neighborhood Fund (SNF) was created as a collaboration between a local community development funding institution and city government (Aguilar, 2018; Jackman, 2016). Soliciting philanthropists, banks, and foundations to invest in real estate and social capital, the fund has attracted much attention yet offers investors little direction of how the money might be invested, and the public minimal information about the government's intentions. Given the scale of Detroit's challenges, it is often difficult for policymakers, real estate investors, and philanthropists to decide how to prioritize policies, grantmaking, or investing.

Detroit is on the brink. The brink of good, the brink of bad, the brink of everything in between. Detroit is a gritty town of hard workers who face relentless struggles living in the city. Detroit's structural decline—physical, financial, and social—has led many to claim that it is "post-apocalyptic," with no chance of survival. Others see the potential for Detroit's residents to rise up and take back the city that has let them down. Still others see the potential to capitalize on the disinvested land, remaking it for the needs of a few, and to the chagrin of many. The changes occurring in Detroit today are setting the scene for Detroit's future—one that resembles an oddly familiar, racially segregated landscape in which many black and brown people struggle to get by, while white people with financial, social, and political capital continue to thrive. As this future plays out, new technologies and analyses are likely to play an increasingly large part; and it is worth investigating the nature of these technologies and their potential impact.

Most Detroiters and decision makers are aware of the potential harms of investing without first understanding the community's needs or planning for future change that may lead to physical or cultural displacement by way of gentrification. With these concerns being realized in the original three SNF neighborhoods, a concerted effort is being made to identify community investments – through policy, grants, or real estate or social investments – that are attuned to the community's needs, investing in a socially impactful and responsible way that does not displace low-income residents. This research develops a technical tool that attempts to predict neighborhood change – as measured by indicators of socioeconomic "wellbeing" – and investigates the ethical challenges inherent in such a process. The technical component utilizes publicly available data to predict changes in socioeconomic status in Detroit neighborhoods from 2012 to 2017 utilizing machine learning techniques. The research investigates how these data can shed light on Detroit's socioeconomic changes since its declaration of municipal bankruptcy, if there is any predictive power to this data, and what the ethical ramifications of such quantitative assessments might be. Can data analysis and algorithms predict neighborhood change – gentrification or decline? Should such processes be utilized in the policymaking realm? This paper also presents an argument against the use of such algorithm alone as a decision-making mechanism, especially without first working within the communities that might be most affected by its implementation in policy or investment decision-making.

Although this research highlights the utility and risk of a machine learning algorithm in a limited scope, it identifies the stakes of what more powerful machinery can do, and lays the groundwork for a discussion of what anticipatory forms of governance might be necessary to ensure that needs of the community are met in the process of urban revitalization. Though beyond the scope of this research, these technologies hold the potential to bypass the democratic process to determine resource allocation and organize the urban landscape in a reflection of the algorithm, replete with the assumptions and biases built into the process. This research identifies these assumptions, biases, and the risks, though a more complete discussion of the risks in relation to democracy and public decision-making is warranted in future work. Methodologies such as this could be used in a variety of government settings, and it is necessary to investigate these processes in each instance to ensure that biases are not engrained in the algorithms, and that the algorithms are not utilized as the primary decision-making tool.

This paper grounds this narrative with a brief history of Detroit, providing an overview of its unequal history and current state of development. I then discuss the various definitions of neighborhood change that are used in topics such as gentrification and urban disinvestment and decline. Chapter Two outlines the methodology of the data processing and analysis process showing how neighborhood change can be predicted through machine learning. Chapter Three focuses on the ethical challenges of machine learning and algorithms in policy and decision making, contextualizing this research within the Strategic Neighborhood Fund initiative, and how a tool such as this could be either helpful or harmful in future endeavors for investing in Detroit’s neighborhoods.

Chapter 1: History and Change

Detroit

Brief history

While I think that it is inappropriate to completely reduce an analysis of the city's reality today to statistics, the data can help to frame the context of the city through comparison with others that are similar in history and current state. Detroit's population shrank from a peak of nearly 2 million in the mid-1950s nearly by half to today's approximately 688,000 (Sugrue, 1996; *U.S. Census Bureau QuickFacts Selected*, 2017). Between 2000 and 2010, Detroit lost 25% of its population (237,500 people), only 4% less than the people who were displaced or lost from New Orleans in the aftermath of Hurricane Katrina in 2005 (140,000 people) (Seelye, 2011). The city's median income is approximately $27,000 (compared to the county's median income of $43,000, which takes into account much wealthier suburbs), has a poverty rate estimated at 36%, and nearly one in two children live below the poverty line (Mack, 2019; Saunders, 2018a).

Detroit also has an important history of social, economic, and geographic inequalities. Consistently ranked as one of the top locations that has the starkest contrast of racial composition between the city and suburbs, Detroit's majority black population has fought through decades of racist federal and local policies, and continue to bear the brunt of the city's structural inequalities. Despite Detroit being nearly 83% black and is the poorest largest city in the country, neighboring Oakland County is nearly 70% white, and is typically one of the wealthiest counties in the country (Schuetz, 2017; William H. Frey, Brookings Insitution, and University of Michigan Social Science Data Analysis Network, 2011). Detroiters often lack access to quality education, good paying jobs, and quality housing (Pandey, 2019; Sands, 2017). While these challenges exist across the city, it should be noted that the downtown area of Detroit – "the 7.2" – has recently received significant investment in its housing and other amenities like restaurants, shops, transportation, and is noted as the hub of investment often not seen in the rest of the city. This part of the city has been experiencing a "rebirth," but it is clear that it may not be inclusive of all races and incomes (Derringer, 2014, 2017; Mondry, 2020; J. Williams, 2020). It is important to consider these inequalities within the context of the research, as the research and its questions are not ahistorical. Detroit has tried many times to "revitalize," though to little avail.

Since the years of Detroit's longest-serving mayor, Coleman Young (led the city from 1973-1993), the city's physical landscape of the downtown area was consistently of mayoral focus while simultaneously the social welfare of the city's residents continued to decline (Rich, 1989). Coleman Young was succeeded by Dennis Archer, a "buddy" with the suburbs who attempted to bring more development and suburban-city partnerships to increase the economic vitality of the city in the 1990s, and Detroit continued to struggle economically and politically (Thompson, 2017a). The election of Kwame Kilpatrick in 2002 was thought to be the dawn of a new era for Detroit and unfortunately indictments on financial and political corruption in 2009 resulted in Kilpatrick being sentenced to 27 years in prison. Kilpatrick's years were highlighted with more borrowing and a debt restructuring plan that ultimately was key in the city's 2013 bankruptcy (Gallagher & Bomey, 2013). During the intermittent years of 2009 and 2013, the city was led for an instant by Kenneth Cockerel Jr., and then for a few years by former Detroit Piston Dave Bing (Thomas, 2013).

Bing was thought to provide a sense of stability and order to city government, although most of his actions were in response to the Great Recession and attempting to keep Detroit from collapsing into bankruptcy (Thompson, 2019). During this time, he acknowledged that the city could not financially upkeep the majority of its city services and discussed the potential of "the shrinking city." Beginning as the Detroit Works Project, Bing commissioned a task force to explore options for the city's "rightsizing"—moving residents in high vacancy areas to create denser neighborhoods—in order to better provide municipal services to residents (Oosting, 2010). This project transformed into the Detroit Future City Framework (DFC), which researched and recommended city-wide changes to the physical landscape for Detroit's next 50 years (Detroit Future City, 2013). Much of this master plan included re-purposing vacant land for new uses, invigorating the economic sector--especially with minority owned businesses--as well as enhancing the city's transportation and general infrastructure. Critics of the plan criticized it for being a neocolonial attempt to colonize Detroit with wealthier outsiders (mostly white), and for not better incorporating resident's thoughts about what they wanted their neighborhoods to look like.

Although not an official city plan, it was heralded by the city's current mayor, Mike Duggan, who has also brought about many controversial developments during his tenure ("Redesigning Detroit: Mayor Mike Duggan's Blueprint Unveiled," 2015). Elected as a write-in candidate with the backing of the Detroit political elite, Duggan has been met with both support and frustrations of residents – the former businessman has fulfilled some of his campaign promises such as making sure all streetlights are operational, but has failed many residents by not finding pathways toward ensuring every resident has running water, or mitigating harm from tax foreclosures and evictions (Florida, 2014; Guyette, 2014; Manick, 2014; Sands, 2017; Schaefer & Walker, 2013; C. Williams, 2013). One of Duggan's primary efforts has been to support a booming commercial sector, finding ways to bring more jobs within the city limits along with other amenities common in big cities (Gallagher, 2014; Hackney, 2014; Hulett, n.d.; Reese et al., 2017; Terek & Guralnick, 2014; The Michigan Citizen, 2014).

A key part of new development, businesses with primarily white-collar jobs have moved into previously-empty office buildings and have offered incentives for employees to take up residence in the city (Felton, 2014; Gallagher, 2014; Live Midtown, n.d.). Unfortunately, many of the permanent jobs created by the newly relocated businesses downtown are white-collar jobs that the average Detroiter does not qualify for, and much of this development is concentrated in the 7.2, raising rents and shifting the demographic makeup of these places in the process (Sands, 2017). Duggan has consistently been explicit about wanting to attract a new tax base that can contribute to the municipal financial wellbeing of the city; and is competing with Detroit's various suburban enclaves for this population. The new residents that he has intentionally attracted are whiter, wealthier, more educated, younger, and newer to Detroit, and are benefitting from these changes that Duggan has helped usher in. Coffee shops, boutique stores, and high-end grocery stores emerge on a monthly basis, filling storefronts and encouraging window-shopping as residents – of Detroit and the suburbs alike – now feel a safety and walkability that these neighborhoods did not have for many decades (Reese et al., 2017).

Simultaneously, longtime Detroiters – "legacy Detroiters," or those who remained in the city limits over the decades despite its challenges – are still encountering daily challenges of paying rent, finding good jobs, having access to quality transportation, and getting running water (Florida, 2014; Foley, 2013b, 2014). During this time, many Detroiters have been frustrated that "the 7.2" has received special attention, while the majority of the city's neighborhoods have not seen similar levels of physical or social investment (Archambault, 2019; Beshouri, 2013; Derringer, 2014; Foley, 2013a). The dichotomy of "two Detroits" has emerged as this new reality, and it is the current environment in which this research takes place.

When Detroit was at its peak population of nearly 2 million people, 20-minute neighborhoods were everywhere throughout the city – residents could walk within 20 minutes of their home to the grocery store, to get their hair cut, to go to dinner, etc. In this vein, Duggan originally released plans for redeveloping the "20-minute neighborhood," which morphed into the current Strategic Neighborhood Fund (SNF) – focusing investment dollars and real estate development into specific neighborhoods. The first identified neighborhoods became the focal neighborhoods of SNF's initial investments – Islandview/The Villages, Southwest/Mexicantown, and Livernois/McNichols (Runyan, 2016).

Mayor Duggan has worked hard to fulfill his 2017 campaign promise of "Every Neighborhood Has a Future," of which the Strategic Neighborhood Fund has become a primary example for his efforts (Boyle, 2016; Runyan, 2016). In his efforts to attract new residents to the city, Duggan partnered with a variety of philanthropies and offered numerous tax subsidies to developers to build and rehabilitate housing and office space throughout the downtown and midtown neighborhoods, part of the 7.2. This approach takes a typical "build it and they will come" narrative – with nearly 4 million people in the Metropolitan Detroit area, there are many people who would likely want to move to Detroit if it could offer similar amenities to some of the inner-ring suburbs that are more attractive, or amenities that are competitive with other big cities (Crawford, 2018; Saunders, 2018b).

Modeling their approach to that of the development in the midtown and downtown areas of the city from the past 10 years, SNF is infrastructure-centric, leaving much of the social and economic components to the non-profit and foundation realm (Invest Detroit, n.d.). However, when everything is built, and this new socioeconomic and demographic class of people start to move into the city, concerns of gentrification and displacement are at the forefront of many Detroiters' minds.

Change in Detroit: What is it, exactly?

Detroit is a city of competing visions. Not just one neighborhood, one block, or one community; one whole city. Exemplary of the competing visions, some choose to describe Detroit as being split into two cities. The common delineation is that one of these cities is white, educated, and relatively wealthy. The other city is black, uneducated, and impoverished. One city has recently arrived or reappeared, while one has stayed the constant, struggling through the decline of the de-industrial age in Detroit. The conversations heard around Detroit today reflect the reality of change in the city, albeit having no concurrence among residents as to the value of this change—positive or negative. Headlines discuss the two extremes of the "two Detroits" vividly: "Is there Room for Black People in the New Detroit?"; "Newcomers move into bankrupt

Detroit"; "A 'Tale of Two Cities' As Detroit Looks to 2014"; "The First Shots in The War Against Detroit Hipsters Have Been Fired"; and, "A Friendly Reminder That The Universe Does Not Revolve Around Midtown," among others (Foley, 2013a, 2013b; Hackney, 2014; Hulett, n.d.). The conversations that these headlines embody take place at all levels of society; among city residents, within non-profits, and during city council sessions.

Addressing the city's challenges – not only those of the competing visions, but also of the deep-seated structural inequalities that require many Detroiters to remain in poverty – is of foremost concern of the current administration. However, as this paper argues, the administration's current process – by which public, philanthropic, and private dollars are channeled through the Strategic Neighborhood Fund – is likely driven not by social or wellbeing indicators, but rather by market indicators, hoping to generate the best return on an investment. As certain neighborhoods are targeted for these investments and begin to change in their demographic and socioeconomic makeup, additional investment without safeguards is likely to lead to gentrification in these neighborhoods. The following section explores change in Detroit currently and the topic of gentrification more broadly.

Definitions

The term gentrification was pioneered by sociologist Ruth Glass in the 1960s to describe the changing demographic and economic makeup in a few London neighborhoods. She noted how a new middle class moved into working class neighborhoods, and the subsequent raise in rent prices and property value displaced many people in the working class (Glass, 1964). In the United States, gentrification in this sense has been happening for a similar amount of time, although it was most poignantly noticed in the 1960s, and then throughout the next six decades to present, typically during times of growth after a recession (Hackworth & Smith, 2001). As well as being a class-conscious topic, gentrification takes on a particularly racial component, as communities of color inhabit the majority of disinvested inner cities (Hanlon, 2011; Jackle & Wilson, 1992; Smith, 1996). When a new middle class moves into these areas, the population is largely composed of a young, white, educated people carrying the associative cultural and social capital that is different from the people who are already living there (Zukin, 1987). As Doucet (2014) describes, gentrification is "…a social process: the upward class transformation of a neighbourhood and the displacement of its low-income population"(Doucet, 2014).

With this change, even if there is not direct displacement of the initial population, the topic of cultural displacement is very pertinent, especially in the case of Detroit (Richardson et al., 2019). Cahill, a scholar of urban studies, discusses how, "Displacement is experienced in this regard as a process of effacement at the neighborhood scale, where the signs of personal and cultural heritages are erased" (Cahill, 2007). For example, in Detroit's new entertainment district, community strongholds such as the corner party store and a local bar, as well as a few historic houses have been torn down to make way for the new stadium. While the party store and local bar are two disparate establishments in the footprint of the new project, they symbolize a larger change that is happening in some neighborhoods, as previously abandoned buildings are becoming coffee shops and art galleries. Ley and Zukin argue that the draw of the "urban kitsch" is something in particular that brings gentrifiers to the city, especially a city with an "idyllic" history (Ley, 1994; Zukin, 1987). This argument is particularly relevant to Detroit, once a symbol of American greatness -- the "Arsenal of Democracy", birth of the automobile and assembly line, Motown -- and now subject to a nostalgic lamenting and fascination with the "ruin porn" of the effects of disinvestment in the city's physical landscape.

While Cahill is accurate in the description of cultural displacement being a consequence (or definition) of gentrification, john a. powell's conceptualization of gentrification in large, poor cities with large minority populations—like Detroit—is that *any* displacement that is felt is "intra-jurisdictional gentrification," where people are forced by economic necessity to move to another place within the city's limits (powell, 2013). He argues that this—intra-jurisdictional gentrification—is an economic necessity for poor cities, and that intra-jurisdictional displacement in poor cities is, "…less likely to be displacement in terms of housing but rather a fear of displacement in terms of power. There is concern that the influx of whites to the city foreshadows white domination." Powell's conceptualization of intra-jurisdictional is an appropriate way to think of gentrification in Detroit.

Often identified through a lens of the gentrified and the gentrifiers, it is helpful to conceptualize the displacement in terms of power as gentrification is discussed by Moskowitz as, "Gentrification is not about individual acts; it's about systemic violence based on decades of racist housing policy in the United States that has denied people of color, especially black people, access to the same kinds of housing, and therefore the same levels of wealth, as white Americans" (Moskowitz, 2017). Although the lived experience of gentrification plays out through these specific actions and movements of people in and out of neighborhoods, it cannot be viewed strictly through a transactional lens – these movements are all part of a system, one that is beset by capital and racism and inequality, and continues to dominate most of how our cities operate today.

Is Everything Gentrifying? An Argument for Neighborhood Change

For the purposes of this research, I would like to utilize a framework that distinguishes between gentrification and neighborhood change. Within this framework, I argue that displacement – either cultural or physical – is a necessary component of gentrification (Shaw, 2001; Slater, 2006). This type of definition is supported by the literature and is helpful in delineating between gentrification and it's encompassing partner, neighborhood change. In part, a neighborhood changing in demographic, socioeconomic, or physical ways contributes to a different lived experience for longtime residents (M. Davidson, 2008). However, neighborhood change is that which can precede gentrification through changes in social, economic, or demographic makeup of a community. Neighborhood change is best defined by Data Driven Detroit as being "…comprised of the economic or physical changes that could substantially affect the composition and culture of a community" (Quesnelle et al., 2019). Although broad, this definition encompasses the range of changes that are happening throughout Detroit today – some which may lead to displacement, and others which may only be the tip of the iceberg.

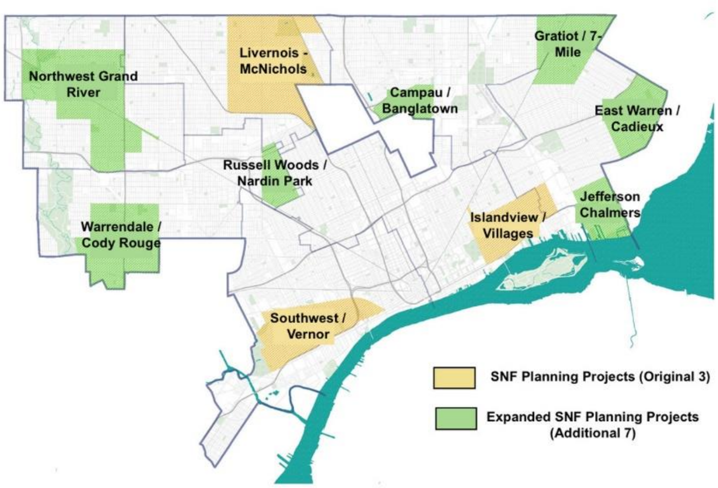
Having undergone an "image makeover," where cities are actively supporting the private sector's housing development actions to encourage gentrification, gentrification has more recently become a tool of urban policy as opposed to a problem with which urban policy has had to concern itself (M. Davidson, 2008; Mark Davidson & Lees, 2005). While every neighborhood has not yet felt the immediate effects of increased investment in their central business district through rising costs and displacement, much of the change identified through this research is more subtle – components of socioeconomic indicators that might tip off residents that the nature of their neighborhood might be changing. This effort is, in part, to encompass the component of gentrification that Moskovitz discusses as the "precursor" – when municipalities start to brand, offer tax incentives, and renew urban planning decisions such as streetscapes and zoning (Moskowitz, 2017).

In Detroit, this is seen through the public-private partnerships such as the Downtown Development Authority and the Downtown Detroit Partnership; all collective cohorts of business and development interests that work with the city government to make the downtown area and greater 7.2 a safer, more attractive place to "live, work, and play"(D:hive, n.d.). It is these types of approaches that Detroit has been pursuing, now through the Strategic Neighborhood Fund, and it is important to frame these changes appropriately. It is true that many changes have already happened in Detroit that could be considered gentrification – specifically, much of the development in the 7.2 meets this criteria.

However, this paper is not meant to be an in depth argument back and forth about the question of gentrification in Detroit, but rather offer a brief investigation into the term while conceptualizing other ways to consider neighborhood change. Commonly used today as a catch-all term for "change" in an urban context, this paper intentionally does not utilize "gentrification" in describing most of the predicted "changes" throughout Detroit. As highlighted by Alan Mallach in "A Divided City," while many forms of investment in the urban landscape in cities like Detroit can be rightfully thought of as impending expensive changes for a neighborhood, there is an importance in also highlighting the changes that are not shiny – identifying neighborhoods that might be considered "declining" in many of the socioeconomic indicators also thought to indicate gentrification (Mallach, 2018). Therefore, in utilizing this broader conceptualization, this research captures both changes that might be precursors to gentrification and their counterparts in disinvestment.

Identifying Neighborhood Change

Arguably, the Strategic Neighborhood Fund fits into this definition, as well, though it is yet unclear how much gentrification will be utilized as a tool of urban policy as compared to a challenge that additional policies might have to address. SNF first directed its efforts to three neighborhoods – Islandview/Villages, Southwest/Mexicantown, and Livernois-McNichols. These efforts were then expanded to include 7 additional neighborhoods – Islandview, Jefferson Chalmers, East Warren/Cadieux, Gratiot/7-mile, Campau/Banglatown, Russell Woods/Nardin Park, Warrendale/Cody Rouge, and Northwest Grand River. Some of these neighborhoods have historically held a strong middle class, despite the years of



disinvestment. The neighborhoods also boast central commercial corridors around which development might focus.

Figure 1: Strategic Neighborhood Fund Neighborhoods (Runyan 2016)

The next chapter will delve deeper into the data processing and machine learning that identifies “changing” census tracts – those that are “ascending” in socioeconomic status, as defined by a collection of data indicators. These findings will be analyzed vis-à-vis the Strategic Neighborhood Fund neighborhoods, and provide an analysis of what the findings might mean in the context of equitable development in Detroit.

Chapter 2: Measuring and Predicting "Change"

There is evidence to suggest that data-driven decision-making (DDD) has become more ubiquitous across both the private and public sectors as sophisticated technologies have risen in popularity and effectiveness (Provost & Fawcett, 2013). It is important to understand the link between the data science methodologies and the decision-making – they are not one and the same. Provost et al outline this relationship nicely as it relates to the private sector in Figure 1(Provost & Fawcett, 2013). This section will outline the process by which data is processed, utilized in machine learning algorithms, and how the results might lead to decisions. Chapter 3 will then investigate how multiple components of this process – from data processing to decision-making – are embedded with certain ethical choices and ramifications that are important to consider, particularly when issues of social justice and equity are at stake (which they usually are).

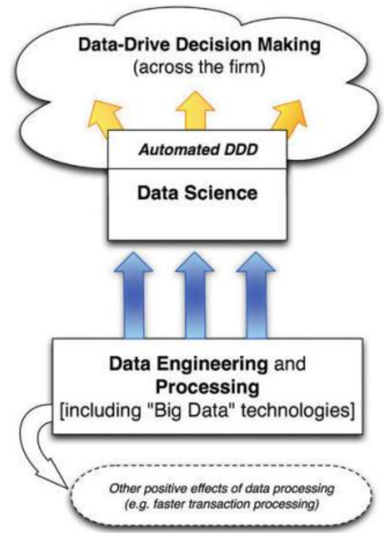


Figure 2: Relationship Between Data Science and Decision-Making

Can Neighborhood Change be Measured? Or Predicted? A Review of the Literature

With the proliferation of data-driven decision-making and rise in computational social sciences, predicting neighborhood change is not necessarily a fringe component of urban studies (Adler, 2017; Helbich et al., 2013; Kontokosta & Jain, 2015; Lepri et al., 2018). More generally, predictive modeling is nothing new to government. Econometric models – regressions, difference in difference methodologies, and others – have been used for decades to predict economic growth and decline. These models have historically operated in a similar way as to machine learning – analyzing a large amount of data to find trends and predict the future. However, though econometric models are the bread and butter of financial policymaking, given the rise of new technologies, machine learning has emerged as a new leader in analytic prowess. Some governments have used predictive technologies to predict crime in cities (Mohler et al., 2015). One of the more high-profile cases of government utilizing machine learning is in sentencing decisions based on likelihood of recidivism as calculated by a number of factors (Angwin et al., 2016). Other agencies use predictive modeling to predict fires, or to predict and prevent human trafficking (Hillenbrand, 2016; Quinn, 2016).

Think tanks and research institutes such as the Urban Institute, the National Neighborhood Indicators Partnership (NNIP), and private entities such as Geolytics and others have aggregated both publicly available and proprietary data of cities to develop indicators and change markers of gentrification, growing this field of study as inequality in cities continues to increase (*NNIP Mission | NNIP*, n.d.; Pettit et al., 2019). Using data to predict neighborhood change, or gentrification, is an attempt to stem the tide of social, economic, and racial marginalization that has enveloped our cities. The theory follows that if we can prevent gentrification before it happens, or if we can mitigate its harms as it happens, maybe we have a chance at maintaining our cities for everyone, without anyone being displaced (Chapple & Zuk, 2016). A lofty goal, it is one worth fighting for. However, these tools must be used within a consideration of the ethical implications of their design and implementation, a topic I will consider in depth in Chapter Three.

Many of the studies that predict neighborhood change focus on individual locations and define "neighborhood change" in a variety of ways (Heidkamp & Lucas, 2006; Lee & Mergenhagen, 1984; Richardson et al., 2019). One of the primary challenges of defining neighborhood change is deciding which data to utilize as indicators and defining neighborhood typologies. There is no set agreement within theories of urban change as to what data points are most informative or relevant in analyzing change. For the purposes of the NNIP project, The Urban Institute reviewed the range of indicators that might be utilized as proxies for a changing neighborhood, identifying resident characteristics, housing markets and conditions, economic activity and investments, and neighborhood conditions as the primary categories by which analysts of neighborhood-level data might identify changes in a neighborhood, separate from a qualitative investigation (Pettit et al., 2019). Other scholars and researchers have utilized other methods and indicators, each of which are valid and important in their own way.

Some of the earliest studies that used a data-driven or machine learning approach to identify revitalization and gentrification utilized a basic statistical approach, calculating housing and population changes in places like Nashville, others utilized a form of factor or discriminant analysis to score census tracts based on their similar indicators of change (Hammel & Wyly, 1996, 1996; Heidkamp & Lucas, 2006; Wyly & Hammel, 1998, p. 00). More recent approaches utilize various forms of unsupervised learning and other machine learning techniques to predict gentrification (Royall & Wortmann, 2015; Winston & Walker, 2012). Owens (2012) utilizes Principal Component Analysis to reduce the dimensionality of data -- such as household income, rent, house values, percentage of residents with a degree, and those with a high status job—to determine socioeconomic status of parts of the city (Owens, 2012). Chapple et al have taken comprehensive approaches to calculating and predicting neighborhood change through logit models and various socioeconomic indicators associated with such changes (Chapple & Zuk, n.d.). Other studies have analyzed gentrification through Google Street View utilizing deep learning methods to assess a changed physical landscape (Ilic et al., 2019).

This research most closely follows Data Driven Detroit's Turning the Corner report (a collaboration with the NNIP project on measuring neighborhood change), and borrows from the theory behind Owens methodology of PCA, and Reades et al. prediction of gentrification in London (Owens, 2012; Quesnelle et al., 2019; Reades et al., 2019). Data Driven Detroit's analysis follows closely George Galster's methodology and analysis of important indicators for neighborhood well-being (Galster et al., 2006). These categories are split into six types of indicators, of which D3 broke down into smaller categories to create an index approach to typologizing neighborhoods. For the purposes of utilizing machine learning, an exact replication of this methodology was not utilized, though referenced when selecting the dependent variables.

Methodology

This research involved three primary components. First, data was selected for its applicability to the context and purpose of the research. Importantly, the data chosen is all publicly available data, meaning that any person who has access to a computer should be able to obtain this data. This was done intentionally to allow for replicability of the study for validation purposes, and in the spirit of open source data. A full list of indicators utilized as independent variables is available in Appendix A.

Then, "socioeconomic ascent," the dependent variable, was defined using socioeconomic indicators via three different methodologies – a manual method, k-means unsupervised learning, and using dimensionality reduction through principal component analysis. Census tracts were labeled within each methodology as having "ascended" or "not ascended," meaning the indicators stayed static or "declined".[1]

Utilizing these categories, machine learning algorithms were constructed and utilized to "predict" these changes within the realm of the various methods of selecting "ascending" tracts. The language is dichotomous because of an interest to highlight stark inequalities throughout the city, and for ease of data processing and utilization in the machine learning models. While characterizing census tracts into multiple categories may have been possible, it was determined that there were not enough data points to have a statistically significant number of census tracts in each category, which would likely skew the results.

The results are then analyzed and discussed regarding the accuracy of the predictions, false positive and false negative rates of the models predicting correctly the ascending or not-ascending tracts, and contextualizing the findings within the discussion of the Strategic Neighborhood Fund.

Assumptions

It is important to first state the assumptions of this research that underlie the methodology and findings. It is assumed that the Strategic Neighborhood Fund is a gentrifying initiative – the changes that will result from increased investment in the SNF neighborhoods will change the fabric of these communities by becoming more attractive to those of higher socioeconomic status. The purpose of this research is to test the hypothesis that these neighborhoods already have precursors for gentrification as identified by their socioeconomic status – that predicting socioeconomic ascent is possible in these neighborhoods. The precursors to full-fledged gentrification (and, by necessity, displacement) are increases in socioeconomic indicators – these precursors should warn policymakers that these neighborhoods are likely to gentrify – and the machine learning models are hypothesized to predict this.

Data sources and descriptions

American Community Survey Data

Data from the American Community Survey (ACS) from 2012 and 2017 was the primary source of socioeconomic and demographic data. These data set baseline methodology of how to approach this research. The data is collected each year by a sampling method –each year, the Census bureau selects a sampling of households from the Master Address File, with a single household appearing no more than once in a five year span (U.S. Census Bureau, 2008). This research utilizes ACS 5-year estimate data, where the samples taken over a five year span are averaged to arrive at the estimate, and a margin of error is calculated. To eliminate the possibility of overlapping data when

[1] An important note on the use of the language of "ascending" and "not ascending" – this language is used in a normative way to identify census tracts that are "changing" as per metrics of socioeconomic status – where higher incomes, home prices, less vacancy, and other indicators are considered to be "more desirable," and therefore, when a tract has a positive change in these indicators, it is considered to be "ascending." I acknowledge that this inherently separates census tracts by crude measures – something which is an inherent challenge in data-driven analyses such as this. This designation does not derive from lived experiences, and should be understood within this limitation.

comparing the two time points, 2012 and 2017 were utilized, as each years' data are estimates of the previous four years and the immediate year, respectively. Within the research and social science industries, these estimates are often utilized as point in time estimates, which is how I conceptualized my research – how have census tracts changed in their socioeconomic status from 2012 to 2017 based on data that is considered to be a socioeconomic indicator?

Home Mortgage Disclosure Act Data

Mortgage lenders report data per the Home Mortgage Disclosure Act (HMDA) each year – this data theoretically validates fair lending practices (*FFIEC Home Mortgage Disclosure Act*, n.d.). These data are helpful for providing more specific estimates of points in time from a numerical perspective – individual loan amounts and decisions about the applications are provided for each loan application. For the purposes of this research, I took a similar method to the ACS where data from HMDA for each census tract was averaged over the relevant time period. These data include the number of homes that were purchased (as identified using data filtering methodology utilized by the Urban Institute) and a median of the loan amounts applied for (Poethig et al., 2017). Tracts that did not have any mortgages issued but had mortgage loan applications were included – even if mortgages were not issued, the loan amount applied for is useful as a proxy for home valuation. This indicates interest in living in an area, and what willingness to pay to live in an area might be. Theoretically, both these numbers will increase in neighborhoods that are experiencing an ascent in socioeconomic status.

City of Detroit Building Permit Data

Building permit data made available by the City of Detroit's open data portal were utilized to identify parts of the city that have experienced an increase or decrease in building attention during the time frame (*Detroit's Open Data Portal*, n.d.). Since the City of Detroit does not have accessible data from 2008 or 2009, a point in time was utilized as the middle year in each of the 5-year time spans that the ACS data covers – 2010 and 2015 respectively. These years happened to be relatively representative of their respective time-year ranges – 2010 had just 10% more building permits issued than average from 2010-2012, and 2015 had only 5% more issued than average. Due to data limitations, it was determined that the point in time method was sufficient, though not perfect. The data utilized does not include any permit that was pulled for demolition – there have been many demolitions in Detroit during this time frame, with inconclusive evidence as to their effect.

Data Manipulation: Three Different Dependent Variables

The original approach of this research was intended to replicate and validate the methodology of Reades et al. (2019) for predicting neighborhood change. However, when it became clear that there were significant challenges with the data available in Detroit that would make the results difficult to discern, two other methodologies were added that may provide other touchpoints of analysis for predictive purposes. If the general question that is driving this research is about the validity of predicting neighborhood change, it is worthwhile to enquire if different methodologies may work better than others utilizing the same, or similar data.

Manually Selected – "Manual"

One way to conceptualize a change in the socioeconomic makeup of an area is by the movement of people and increases in real estate development, both residential and commercial (M. Davidson, 2008; Mark Davidson & Lees, 2005; He, 2010; Lützeler, 2008). Therefore, this dependent variable – named *Manual* for the purposes of this paper – is a measure of percent change in population growth and number of building permits requested. If a census tract had a percent change increase in population and a building permit percent change increase larger than 0, it was considered an "ascending" tract, and was designated with a 1 (in the binary dependent variable conceptualization). The rest of the tracts were labeled "not-ascending" and were designated with a 0 in the binary conceptualization. This process identified 80 census tracts out of 263 that experienced some amount of ascent in socioeconomic status from 2012 to 2017 per the definition of the *Manual* method. Upon review of these census tracts, they align closely with the neighborhoods of interest with the Strategic Neighborhood Fund, and the downtown/midtown neighborhoods (the 7.2).

K-means Clustered – "K-Means"

Another approach to selecting a dependent variable is that by which the algorithmic and statistical process of k-means clustering identifies similarities between data points – known for the purposes of this paper as the *K-means* method. The percent change in population, percent of homes that are vacant, median loan amount of homes purchased, household income, and number of building permits were utilized as datapoints by which the k-means process would identify clusters. These are data points theorized to indicate components of socioeconomic status – an increase or decrease in any of these indicators can help validate what might be happening in a neighborhood.

After scaling the data with a z-score standard scaler, the data was clustered through a k-means clustering approach. By minimizing the within cluster variance and maximizing the between-cluster variance, this approach was theorized to provide a differentiation of types of census tracts. This approach was tried by Knorr (2019) and was seen to be successful in that case – so this was an attempt to both validate that approach and offer a comparison with the other processes (Knorr, 2019). Once the data is scaled and clustered, the cluster's descriptive statistics are analyzed to identify which cluster has had percent change increases in their socioeconomic indicators – after running this model multiple times, there is consistently one cluster that—by looking at the descriptive statistics and changes—shows growth in population, income, home prices, permits, and a decrease in vacancy, and has approximately the same number of tracts inside the cluster that is similar to the "ascending" tracts of the *Manual* method.

Principal Component Analysis – "SES (Socioeconomic Score)"

This study design was modeled after work by Reades et al. (2019), and in part by Owens (2012). The process by which Reades et al. use dimensionality reduction to create their dependent variable for predicting neighborhood change in London was drawn upon when manipulating and fitting the principal component analysis. The data was gathered for each time frame and loaded into a dataframe for each time period. Each variable's distribution was checked for potentially large ranges or any outliers or skew, as that would potentially limit the explanatory power of the PCA. Null values were dropped, and values of zero were imputed for census tracts that did not have any mortgages issued in either year. Similar to Reades et al.'s findings in London, the median income and mortgage prices had the largest amount of variance and were first scaled before scaling all of the data. Box-Cox and transforming via the natural log were attempted but some census tracts included values of zero which limited the viability of these methods (G.E.P. Box and D.R. Cox, 1964). Eventually, a z-score standard scaling was attempted for these variables, but no significant difference was seen in the results of the PCA (Ciaburro, 2018).

Using the untransformed data, and checking to ensure there were no null values, the two years were concatenated together into a single dataframe, and the values were then loaded into an array. The data was then processed using a RobustScaler technique – one that limits the importance of outliers of the data by scaling according to the interquartile range; a helpful mechanism for data that might have a lot of variance based on outliers (*Scikit-Learn 0.22.2 Documentation*, n.d.). The scaled data is then fit and transformed with the PCA model, and the explained variance ratios of the components are identified. Although similar standardization approaches were attempted (scaling the individual variables before scaling the variables as an entirety) it resulted in no greater explained variance within the first component of the PCA. One important element of the study design and evaluation is that the data was all transformed and fit to the model together – in this way, the scores are more readily comparable between the two years. Because the score itself is not modelling change – it is merely applying dimension reduction to the data, modelling the data projected on the first principal component that was extracted from fitting the model to the data – using this first principal component to achieve the score for each year when using the entire dataset allows for rudimentary comparison between years.

After first assigning the scores to each tract for each year based on the first principal component, the data is split into its respective year ranges and tracts are ranked in order according to these scores and assigned an absolute rank. Then, the tracts are assigned a percentile rank – this helps to identify a census tract's change in relation to the other tracts. Deeper analysis that analyzes the change in ranking by standard deviation – to incorporate magnitude of change relative to the other tracts – is not included in this research, though should be considered in the future.

To select the "ascending" census tracts for the purposes of modeling, I ordered the census tracts by the amount by which they ascended in socioeconomic status absolute ranking and selected the top two quintiles of tracts as the "ascending" census tracts. While is change in rank does not necessarily align with an equal magnitude change in socioeconomic indicators, the proxy is used in order to more easily delineate between ascending and not ascending tracts. The top two quintiles were selected as it was closest in size to the other two identified "changed" census tract groups (manually-selected had 80 tracts, k-means clustered had 94 tracts, this resulted in 105 tracts). The descriptive statistics of this quintile also matched the underlying assumptions about a census tract that has experienced socioeconomic status change, such as an increase in population, home prices, number of mortgages, and number of building permits.

Models Utilized

The prediction of each dependent variable was conducted via three machine learning algorithms to identify the best model. Logistic Regression, Support Vector Machine, and Random Forests were utilized with each dependent variable to understand which model performed best with each dependent variable. The most significant limitation in this research was the unbalanced dataset, with far fewer census tracts identified as having increased in socioeconomic status than not. To address this challenge, Gradient Boosting Machines were utilized in an attempt to manage the unbalanced data, though the findings were insignificantly different from the Random Forest. Other measures to balance the data set – such as upsampling the census tracts identified as changed, and the Synthetic Minority Oversampling Technique –

were attempted though similarly found to be inconclusive in changing the outcomes (Brownlee, 2015, 2020).

Training and Testing Data

Within machine learning, a dataset it typically split into a training and a testing group. The training group is utilized to tweak and tune the parameters of the model, and the testing group is utilized to assess the model's strength. This process limits the potential of data-leakage, a problem that can result when the algorithm, or model, has had insight into the entirety of the dataset – both training and testing groups – which would alter the prediction outcomes. The dataset was randomly split into an 80/20 divide where 80% of the data would be in the training group and the model would test on 20% of the data.

Across the three different methods of selecting a dependent variable, there were only seven tracts that were identified in each method as being "ascending" tracts. However, the *Manual* method and *K-Means* method selected nearly all the same 67 census tracts as being "ascending," and these tracts were primarily in areas that have seen some of the aforementioned development and demographic indicators that can be indicative of a changing neighborhood. Since the overlap of "ascending" census tracts by each method was limited and would have resulted in a more unbalanced dataset, each method utilized its own random split of the data. This limits the comparability between the various methods, so for statistical purposes, the different machine learning methods should be compared only within their respective dependent variable selection method.

Results and Discussion

In measuring the results, the primary process by which I have determined the success of the models are in measuring the precision, recall, and F-1 measurement of the minority class, which is the census tracts that are identified ahead of time as being "ascending." Precision is calculated as the number of true positives divided by the sum of true positives (predicted ascending correctly) and false positives (predicted ascending incorrectly). Recall is calculated as the number of true positives divided by the sum of true positives and false negatives (predicted not-ascending, actual is ascending). F1 measure is the "harmonic mean" of the two metrics, and is widely used as the accuracy metric for the models when the dataset is unbalanced (Brownlee, 2020). The accuracy measure is a simple fraction of correctly predicted outcomes out of the total. Overall, in part due to the unbalanced classes, the models were very good at predicting the tracts that had not experienced socioeconomic status ascent through any of the methods by which the dependent variable was created. However, the models were not as accurate at predicting the "ascending" census tracts. A confusion matrix is shown for the best machine learning method within each dependent variable selection process.

Manual Selection and Machine Learning

This process delivered some of the most promising results of predicting census tracts that have experienced an increase in building permits and population between 2012 and 2017. The logistic regression model showed 83% accuracy – and specifically 77% precision in predicting the changed census tracts with 62% recall. This is a promising result in that it is the most rudimentary and easily replicable among the methodologies utilized. The SVM showed nearly identical results, hinting that it might also be a useful method. Likely due to the small, unbalanced dataset, the random forest performed most poorly.

Logistic Regression Confusion Matrix – *Manual* method	Predicted Values		
Actual Values		**Not Ascending**	**Ascending**
	Not Ascending	35 (True Positives)	3 (False Positives)
	Ascending	6 (False Negatives)	10 (True Negatives)

K-Means Clustering and Machine Learning

This methodology resulted in the best F1 metric utilizing the Support Vector Machine. Again, likely due to an unbalanced dataset, the random forest performed the poorest. However, compared to the random forests utilized in the *Manual* and *SES* approaches, the random forest performed strongest using the k-means clustering approach.

SVM Confusion Matrix – *K-Means Method*	Predicted Values		
Actual Values		**Not Ascending**	**Ascending**
	Not Ascending	36 (True Positives)	1 (False Positives)
	Ascending	10 (False Negatives)	7 (True Negatives)

Principal Component Analysis and Machine Learning

Although the PCA process, when validated with only anecdotal information, proved to be reasonably accurate and helpful in validating many of the changes taking place in the city, it was much more difficult for these machine learning algorithms to predict these changes. Out of the three selected dependent variables, the results of each model reported the worst metrics. While the logistic regression and SVM models had low precision and recall scores, resulting in only a 50% and 63% accuracy metric for each model, respectively. However, the random forest performed much better with this method, with a precision of 58%, recall of 39%, and F1 of 49%, though the overall accuracy was measured at 70%.

Logistic Regression Confusion Matrix – *SES Method*	Predicted Values		
Actual Values		**Not Ascending**	**Ascending**
	Not Ascending	23 (True Positives)	3 (False Positives)
	Ascending	15 (False Negatives)	13 (True Negatives)

Machine Learning Results

The results of the machine learning process are interesting in that it shows the challenge of prediction using real-world data. One of the primary challenges with machine learning methodologies is in achieving a balanced dataset that can effectively train the training model to produce an accurate outcome using the unknown data. To understand the models' strengths in greater depth, it was helpful to identify that most of the models had stronger metrics when predicting the not changed or declining census tracts. The results of the model's performance for the dependent variable in each method is seen below in Table 1.

Table 1: Results of Machine Learning Models

	Precision	Recall	F1	Train Accuracy	Test Accuracy
Manual					
Ascending					
Logistic	0.77	0.62	0.69		
SVM	0.77	0.62	0.69		
RF	0.8	0.25	0.38		
Not Ascending					
Logistic	0.85	0.92	0.89		
SVM	0.85	0.92	0.89		
RF	0.76	0.97	0.85		
Overall Accuracy					
Logistic				0.84	0.83
SVM				0.84	0.83
RF				0.70	0.80
K-Means					
Ascending					
Logistic	0.75	0.53	0.62		
SVM	0.88	0.41	0.56		
RF	0.78	0.41	0.54		
Not Ascending					
Logistic	0.81	0.92	0.86		
SVM	0.78	0.97	0.87		
RF	0.78	0.95	0.85		
Overall Accuracy					
Logistic				0.78	0.79
SVM				0.86	0.80
RF				1.00	0.78
SES					
Ascending					
Logistic	0.81	0.46	0.59		
SVM	0.67	0.36	0.47		
RF	0.67	0.21	0.32		
Not Ascending					
Logistic	0.61	0.88	0.72		
SVM	0.54	0.81	0.65		
RF	0.51	0.88	0.65		
Overall Accuracy					
Logistic				0.71	0.66
SVM				0.77	0.57
RF				1.00	0.53

The models showed the best results when predicting the census tracts that were classified originally as "not ascending". While this may be due in part to the unbalanced data, it is possible that there are closer similarities within the majority of the census tracts that have "not ascended" in socioeconomic status compared to those that do. Utilizing the *Manual* method, all three methods recorded F1 values between .85 and .89. Similarly, the *K-Means* method also recorded F1 scores between .85 and .87. The *SES* method still performed most poorly with the best F1 score of .72 utilizing the logistic regression. These findings suggest that the separation of census tracts utilizing these various methods of determining "socioeconomic ascent or decline" may be useful in machine learning.

One of the primary reasons that the SES method does not perform very well is due to the nature of how census tracts were classified as ascending or not ascending. Because a change in ranking – absolute or percentile-ranking – is not perfectly correlated with the relative change in indicators, using the score as a proxy for predicting the change is not necessarily a best practice. Although the SES score should be a proxy for a tract's socioeconomic status, predicting a tract's change in rank between years does not account for the entire city also changing at once, or for the specific magnitude in change of the indicators that create the score and subsequent rank. Although the dependent variable selected tracts as ascending that had increased in rank by more than 1 ranking spot between years, since all the other tracts were also changing, the meaning associated with this change is difficult to discern, especially when utilizing machine learning models.

Reades et al. (2019) use a more sophisticated approach, in part due to the amount of data accessible to them in London, compared to what is available in this context in Detroit. While a closer following of Reades et al. was attempted for this research, closely following their methodology with such a small dataset resulted in poor results. Although using the ranking as a proxy was explored to see if the results were meaningful, this process should not be utilized to classify and predict a tract's socioeconomic ascent.

Analysis of Census Tracts: Manual and K-Means selection methods

While the results of the machine learning methodologies are interesting, and suggest that there are some models that are effective at predicting these census tracts, it is helpful to do a deeper analysis of the census tracts that are identified through the simplest analysis – the *Manual* method. This method identified any census tract that had both a percent change increase in building permits and population to be "ascending," and the remaining census tracts to be considered "not ascending". Future research should compare these results with the *K-Means* and *SES* methods to determine any significant differences.

Overall Changes between 2012-2017

Overall, the median percent change in population for each census tract in Detroit was a decrease of 6.2% -- half of the city's census tracts experienced larger losses, the other half experienced fewer losses and even some gains. When looking at the entire city, Detroit lost population during this time – across the whole city, there was an overall 5% population loss, or 36,553 people. These data suggest that a few census tracts had population surges, while overall, the majority of the city's census tracts lost people. It is likely that these tracts correlate with the identified "ascending" tracts in both the *Manual* and *K-means* methods of selecting "ascending" tracts.

Manually Selected Tracts: Ascending vs. Not Ascending

To recall, 76 census tracts were identified as "ascending" and 192 tracts were identified as "not ascending" using this method. One of the more interesting findings of this analysis is identifying the census tracts that had both a population and building permit percent increase between the two years. Compared to the city's loss of 5% of its population, these census tracts experienced a 15% increase in population. In comparison, the rest of the census tracts – "not ascending"—had a 13% decrease in population. The other indicators show similar trends of the "ascending" census tracts faring better than the city during this time period – home prices decreased less overall, incomes grew by nearly 6% more than the rest of the city, and there was nearly double the percent change in number of mortgages issued. However, even these statistics are not evenly distributed among the "ascending" census tracts – when looking at the difference between median percent change and mean percent change of all the census tracts, it is clear that some census tracts experienced more population, more mortgages, and larger increases in home prices, permits, and incomes. It is important to note that these changes were not checked for statistical significance, and deeper statistical analysis is warranted.

Comparing these measures is helpful to understand that the revitalization efforts driving population increases and a stronger housing market are not evenly distributed across the city, and are likely concentrated in only a few census tracts. This helps to show the stark inequalities that are taking place during these years, located primarily in the neighborhoods of downtown and midtown where the city's first investments were focused in the years after bankruptcy, and also concentrated in the neighborhoods selected as part of the Strategic Neighborhood Fund. Particularly, when looking at the summary statistics of the "ascending" census tracts, although only half of these tracts had no percent increase in white population, the average percent change in white population was 99% between the two years. This is further supported by the "not ascending" census tracts experiencing the opposite – though its median was also 0% change, the not-ascending census tracts had a lower mean, and a lower maximum percent change. This suggests that there are a few census tracts that had a significant influx of white population, leading to cultural displacement, and likely some amount of subsequent displacement from gentrification (Richardson et al., 2019). However, a difference of means test suggests that the difference in means is not statistically significant – that the underlying distributions of the populations within these census tracts are not different. Therefore, this should be investigated in further depth.

Figures 3 and 4: Percent changes of Socioeconomic Indicators 2012-2017[2]

Ascending – Manual method	Median % Change	Mean % change	Max % Change	Min % Change
Population	13%	19%	181%	0%
Mortgages (raw change)	0	1.87	25	-6.8
Home Prices	-7%	10%	286%	-100%
Incomes	2%	9%	188%	-31%
Building Permits	85%	166%	1178%	0%
% Black	-3%	-6%	71%	-92%
% White	0%	99%	6150%	-100%
% HS Degree	-3%	-2%	135%	-83%
% College Degree	4%	9%	220%	-28%
% Poverty	4%	12%	183%	-73%
% UE	-29%	-21%	97%	-87%
% renter	7%	30%	554%	-99%

Not Ascending - Manual method	Median % Change	Mean % change	Max % Change	Min % Change
Population	-14%	-14%	41%	-57%
Mortgages (raw change)	-0.2	-.02	14.6	-8
Home Prices	-19%	-10%	395%	-100%
Incomes	1%	6%	101%	-73%
Building Permits	0%	40%	1000%	-92%
% Black	-3%	-4%	184%	-80%
% White	0%	65%	3000%	-100%
% HS Degree	-1%	0%	71%	-44%
% College Degree	6%	10%	201%	-39%
% Poverty	-3%	3%	142%	-72%
% UE	-31%	-22%	162%	-92%
% renter	17%	39%	835%	-87%

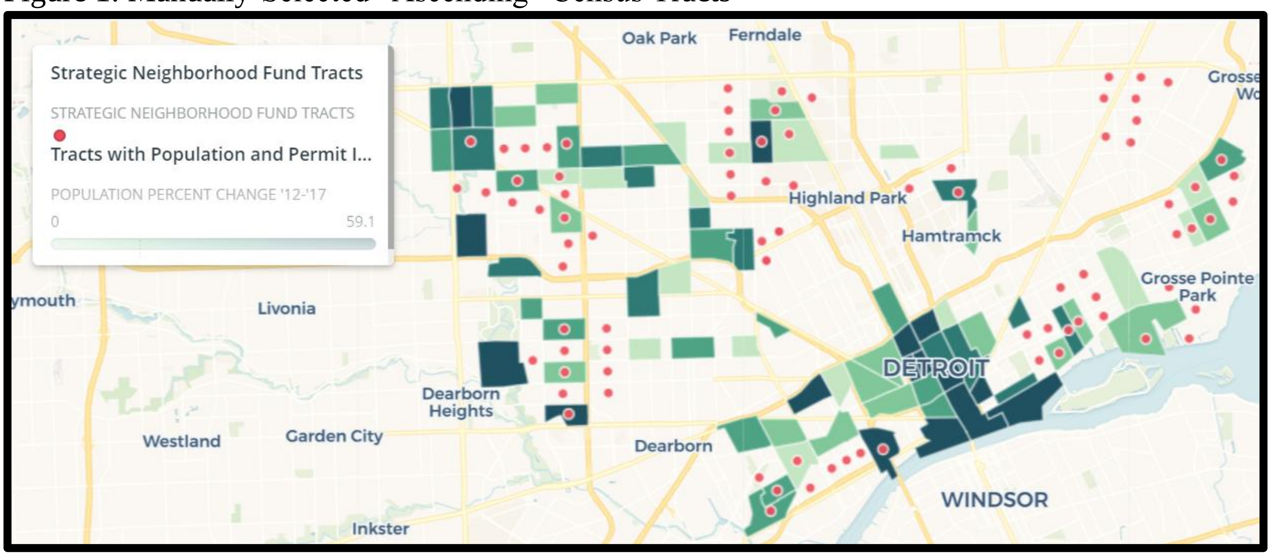
Looking at the selected census tracts, it is interesting to see the overlap in amount of percent change increase in population with the Strategic Neighborhood Fund areas. The 7.2 – the area that experienced the first round of investment before SNF was created – is clearly outlined by the highest increases in population during this time period. Since the time periods measured in this data analysis precede the expansion of SNF to the additional neighborhoods, it is interesting to consider what components of these neighborhoods might have led to their selection. The census tracts identified in this map as "ascending" overlap with the SNF neighborhoods, except for the Gratiot/7-mile neighborhood. The logistic regression model predicts these changes with

[2] Because some census tracts had no mortgages issued in 2012, the raw change in mortgages is utilized as a measure.

80% accuracy, a decent metric when utilizing real-world data. What these findings suggest are that these census tracts are likely to continue to "ascend" in their socioeconomic status and that, when paired with the Strategic Neighborhood Fund investments, this ascent may be costly for those who cannot afford changes in the environment's socioeconomic status, such as rent and access to affordable food. If the neighborhoods are changing – potentially gentrifying – and there is a program that is intended to accelerate these changes focused on these neighborhoods – I would argue that the likelihood for gentrification and displacement is higher.

Comparatively, the rest of the census tracts that are "not ascending" differ impressively in their descriptive statistics of measures of socioeconomic status – and likely, other measures of wellbeing. While the "ascending" tracts gained population (tied, in part to the definition of "ascent" in this case), the "not ascending" tracts experienced a median percent change of their population of minus 14%. Similarly, while there was an overall increase in mortgages issued across the "ascending" tracts by 68%, there was only a 1% increase in mortgages issued in the "not ascending" tracts, and home prices decreased by 20% in the "not ascending" parts of the city compared to only a 14% decrease in the "ascending" areas. Not only did the real estate market change in the "ascending" tracts, but so did peoples' incomes – "ascending tracts had an increase in income by 7%, while the "not ascending" tracts' incomes only increased by 1%.

Figure 1: Manually-Selected "Ascending" Census Tracts



Challenges

The primary challenges with this research were the lack of data (only analyzing 283 census tracts), and the estimation of the data over the two five-year periods. Additionally, the data is not perfect as it comes from ACS estimates (samplings averaged over five years) and mortgage data that are reported only by lenders who write more than 50 loans, leaving out potential credit union data. The housing data also does not capture the amount of homes that are bought via cash sale – something that is known to be common in Detroit especially (the city's average credit rating is far below the nation's, limiting access to mortgages for many residents) (Poethig et al., 2017). This research would be valuable to replicate after the next census, as a comparison between 2010 and 2020 would provide a much clearer and accurate picture of what changes might be happening demographically and spatially. Additionally, the primary reason why block groups were not utilized is because housing data at this level is often proprietary – instead, with the intention to have 100% publicly available data, this research could only conduct analyses at the census tract level. Block group level analysis might produce stronger results as it would provide more data points for analysis.

Another primary challenge with this analysis is the process of selecting what might be classified as “ascent” in socioeconomic indicators. Though the best use of theory in the *Manual* method proved to be relevant to lived experience, the *K-Means* method provides similar results, suggesting that there is something that should be understood about the similarities between these “ascending” census tracts. This would require significantly more in-depth statistical work about the data, though would be helpful for understanding how these constructions are relevant to future work in identification and prediction of gentrifying areas.

Key Takeaways

One of the primary questions that I sought out to answer was to figure out which of the methodologies of selecting a dependent variable is the “best,” and which machine learning algorithm is the “best.” In analyzing the results of the machine learning models, the *Manual* approach appears to work the best with all three models, predicting with the strongest accuracy. It is also the approach that is most easily replicable by those with beginner to moderate data analysis and machine learning knowledge – a promising finding for the purpose of members of the general public being able to utilize these findings for their own advocacy. Each method has its own merits, and all would be strengthened with more data to utilize – potentially conducting the analysis using census block groups.

However, the interesting results from the *SES* method that shows that the declining or static census tracts are easier to predict than the changed census tracts provokes an interesting question of the purpose of the algorithm. Specifically, these findings suggest that there are greater similarities among the majority of the declined or static census tracts that could suggest that policymakers take this into stronger consideration than only looking at the census tracts that have experienced some increase in socioeconomic status. What is clear from this rudimentary process utilized in this project is that basic machine learning methodology is limited in its ability to predict with confidence the changing nature of society. In part, this is due to limited data; Reades et al. were able to predict with better confidence in part due to the abundance of data that was available for the study in London. Although each method of selecting a dependent variable and creation of the machine learning algorithm can be considered important and raise interesting questions, it is also important to address the broader questions about the attempt to predict neighborhood change in the first place.

Related to SNF

The implications of these findings as it relates to SNF is that these models identify parts of the city that are experiencing socioeconomic ascent – a precursor to gentrification and potential displacement. Additionally, the identified “ascending” tracts largely overlap with the SNF neighborhoods that either have recently received or are expected to receive increased investments through the fund in commercial, real estate, and street scaping initiatives. Given the socioeconomic status of these neighborhoods, additional investments could lead to gentrification and its negative effects in these neighborhoods as the neighborhoods become more attractive to those with the social, economic, and cultural capital to move. Should elected or appointed officials use tools such as these to help direct policy or investments, it is arguable that stronger policies are necessary to mitigate what could be additionally harmful results of continued changes.

Arguably, these components of neighborhood change identified in these data are certainly felt among the residents in neighborhoods that are already experiencing these changes – some of which are SNF neighborhoods. As Data Driven Detroit found through their qualitative analysis of two neighborhoods in Detroit, there were often intangible components to their neighborhoods changing that came along with the more tangible – houses being sold, new businesses opening up (Quesnelle et al., 2019). These intangible components were elements of community that are difficult to quantify – neighbors saying hello to one another, what police might be called for to address, or the simple act of showing up to community meetings and connecting with neighbors. The more tangible components such as new streetscapes, homes being rehabilitated, or demographic changes, are all elements that cities can theoretically have an easier time understanding. Chapple and Zuk found that since the rise of early warning systems for gentrification in the 1980s, public officials have increasingly utilized newer analyses attempting to quantify risk of gentrification and displacement (Chapple & Zuk, 2016). The basic findings of this research highlight the importance of considering what changes are already taking place in the SNF neighborhoods, and what that might mean for longtime residents as new investments begin to pour into their streets. Importantly, it is important to consider what types of investments that are not physical could be critical to help stabilize the neighborhood's wellbeing while physical investments are made to revitalize.

The findings of this research suggest that there is reason to investigate how this research might be expanded to include other, nonconventional datasets. While this research was intentionally conducted with a limited dataset that is publicly available (so that the process could be replicated), it is reasonable to suggest that results would be more compelling with an expanded dataset, including sources that are proprietary. Such sources may be from mobile applications, business listings, real estate marketing or construction companies, or other companies that collect demographic or socioeconomic data. However, there are important ethical concerns that should be addressed with utilizing an expanded dataset – not only is the component of replicability of concern, but also the broader epistemological understanding of why and for what purpose the prediction is taking place. Without a conversation that investigates why we, as a society, might want to predict where socioeconomic changes are happening in our cities – and why we might not – utilizing machine learning for predicting our social and demographic changes runs a risk of being ethically murky at best, and consequentially harmful at worst.

Of the studies discussed earlier in this chapter that have utilized forecasting or predictive methodologies within this realm, few have discussed the ethical and policy implications of such work, and what it means for these technologies to be available in the public sector. While this research is limited in its scope, it is the hope that it sparks further discussion of ways in which these tools can be utilized for good and continue to investigate the built-in assumptions upon which the methods rest. The next section will focus on the ethical and policy implications of utilizing such technologies within the urban planning and development realm, and the ethical implications of use in cities like Detroit.

Chapter 3: Data, Machine Learning, and the Strategic Neighborhood Fund

This research both utilizes data analysis and machine learning methodologies to identify and predict neighborhood change, and in the following section investigates the ethical dimensions of using such processes in the policy arena from a perspective of socioeconomic, racial, and geographic equity. As discussed earlier, the relevant context of this research is that of the Strategic Neighborhood Fund, the City of Detroit's neighborhood revitalization investment mechanism, intended to draw financial investments from philanthropies, private sector developers, and other government entities and invest in neighborhood amenities in a select handful of neighborhoods throughout the city. A question of this research is to understand potential ethical challenges in utilizing a machine learning prediction of neighborhood change in approaching decisions related to the Strategic Neighborhood Fund. As a policy approach to "develop" some of Detroit's neighborhoods, I investigate if technical tools such as these still serve the normative ends of equity by encouraging a normative ends of equity when applied to policy approaches that may have unclear outcomes regarding social and racial equity. In doing so, I explore the inherent danger of embedding – and hiding—politics and normative ways of thinking behind and within these tools. The following chapter will first include a discussion of the ethics of algorithms and machine learning, situate the ethical landscape vis a vis the Strategic Neighborhood Fund and a cautionary tale of Market Value Analysis, and assess the potential concerns of using this research's process or results in any decision-making related to SNF.

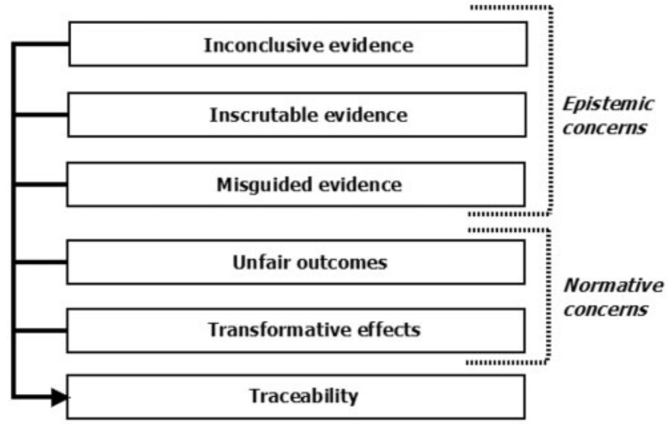
Data justice and data ethics: Ethical concerns of using data for decision-making and urban planning

There are three primary ethical considerations relevant to this research, though this next chapter primarily focuses on one – the ethical concern of using technical tools to make political decisions that may have ramifications on the socioeconomic, racial, or geographic marginalization of the population. There are additional ethical concerns around the creation and use of quantitative data and metrics in comparison to more qualitative measures of peoples' lived experiences, as well as the ethical considerations of equitable development decisions. However, these additional ethical considerations do not fall within the scope of this research and should be addressed in future work.

Mapping the Ethics of Algorithmic Decision-making

Increasingly, algorithms are being implemented in decision-making spaces – private and public – and the ethics of how these data are collected, analyzed, and drive outcomes are of particular concern. Mittelstadt et al (2016) map the current landscape of the ethics of algorithms in decision-making, and note that algorithms today are used to,

"…turn data into evidence for a given outcome…this outcome is then used to trigger and motivate an action that (on its own, or when combined with other actions) may not be ethically neutral. This work is performed in ways that are complex and (semi-)autonomous, which complicates apportionment of responsibility for effects of actions driven by algorithms" (Mittelstadt et al., 2016).



The authors importantly note that it is difficult to assign blame that is caused by algorithms and subsequent decisions made from the outcomes. Many ethicists of algorithms argue that harm is inherent in algorithms – a feature of the system, as the algorithm incorporates any biases in the data into its final output and outcome (Eubanks, 2017; Gangadharan et al., 2014). When creating a machine learning algorithm, various decisions are made about the type and form of data being utilized, and the outcomes depend on a multitude of decisions in the data's cleaning and processing and in construction of the algorithm's form and parameters.

As Mittelstadt et al. (2016) note, inconclusive, inscrutable, and misguided evidence are all potential ethical complications in algorithm creation and use, which may lead to unfair outcomes and transformative effects, such as changing our conceptualization of the world and its sociopolitical organization. A discussion of such outcomes in this present research about neighborhood change is relevant – much of the evidence presented in the previous chapter might point to certain policy decisions that could be considered inconclusive, inscrutable, or misguided, and could lead to unfair outcomes or transformative effects that reconceptualize our urban framework in cities like Detroit.

Encompassing all these challenges is that of traceability – a component that is often unattainable in machine learning algorithms – how can one trace where a specific outcome was reached based on the data input? This challenge leaves little agency for any member of the public to address some of the algorithmic violence and harms that are caused through the data manipulation and analysis process. Though it is not the focus of this paper, it is important to consider the ways in which data analysts and researchers might mitigate this harm, such as by creating research that is replicable and is traceable. Researchers must mitigate the risk that their process could be used for improper decision-making, and evaluate on a case by case basis how to make the information open and accessible – as an accountability measure for the public that may be harmed by the results of the research (Kenneally & Dittrich, 2012).

The chapter will utilize Mittelstadt et al.'s framework for understanding the ethical implications associated with the development of the tool for predicting neighborhood change. Overarching this discussion is one about the concern of traceability and equity within algorithms and data-driven decisionmaking – something which is initially addressed as perhaps the most complex component of ethical challenges in employing algorithms, and deserves further discussion beyond the scope of this specific research. I briefly discuss how the process of building the algorithms in this research could arguably constitute inconclusive evidence, and the risks associated with making decisions based on these findings. Second, I investigate how these algorithms' opaque nature may limit the ability for the data subjects upon which the utilized data is based to know how their data is being used – in part due to the nature of the algorithm, and in part due to the risk of opening the algorithm's creation to the public, risking bad actors utilizing it to "game the system". Third, and perhaps most importantly, I discuss how the data's initial processing and analysis is arguably the source of misguided evidence, as values are embedded within the code, and biases may result as a reflection of these values. Fourth, I discuss how actions, or policies, driven by algorithms can lead to unfair outcomes and further marginalization within Detroit, per the context of revitalization and gentrification that was discussed in the previous chapters. I will utilize a cautionary tale of Market Value Analysis – a form of data analysis that embeds politically-relevant norms and values into its process and results. Although not an automated algorithm, this is a complex process that involves similar opacity to machine learning algorithms, and machine learning algorithms are a likely next step or extension of this methodology. The final discussion mentions the potential risks of re-ontologizing the urban environment within a data-driven and algorithmic epistemology that embeds values into algorithmic biases, leading to questionable policymaking. I conclude with a discussion of policymaking within a data-driven context and argue the need for anticipatory governance of the use of machine learning technologies within the realm of public policy.

Traceability: Accountability Challenges

Some ethicists argue that harm is inherent in the creation of any algorithm that may have an impact on human life. The unknown decisions that are made in the algorithm, the biases that may be present in creating the algorithm, and the use of the algorithm's results are all disconcerting factors that may impact equity considerations when the algorithm is utilized. Sandvig et al note that complex computer code, the unknowns that occur during the algorithm's execution, and the intent of the creator all contribute to the need for a "Consumer Reports" of algorithms, so that the public can know the essential components of an algorithm and how it might be affecting their daily life (Gangadharan et al., 2014). Similarly, Martin argues that algorithms are not neutral, and are in fact value-laden, and that algorithm's creators should be held responsible for any harms that occur from the result of the algorithm (Martin, 2019).

As Sandvig et al argue that the intent of a private sector corporation through the use of algorithms for decisionmaking is unlikely to be in the citizen's best interest, within the context of this research, I argue that this concern can be extended to the public sector in our current reality of the neoliberal state and rise in public-private partnerships. Importantly, within the context of the Strategic Neighborhood Fund, it is exactly the private sector that is seeking to benefit the most from any sort of investment into the fund, and it will likely seek to do so at the expense of a group of people – likely a socioeconomically, racially, or geographically marginalized population. In fact, it may be an implicit goal of such algorithms in the context of neighborhood change to continue to accelerate change, and utilize the algorithms as both a tool and justification toward these ends. In doing so, accountability and traceability of the algorithms origins, purpose, and construction is difficult to ascertain and often goes unquestioned, as policy and politics obscure the roots of decisionmaking.

Although the government should be seeking to act in the best interest of all its citizens, the calculus of who might be harmed utilizing such an algorithm may not be clear or intentionally considered by policymakers – further embedding systemic or structural racism within policymaking. Without a clear method of holding the algorithm or its creators accountable – due to the complexity of the code, unknown decisions in creating and built into the algorithm – the ethical concerns go beyond the creation of the algorithm to its potential unequal impact, and who should be held to answer for its potentially disparate impacts when driving policy. Sandvig et al call for an "audit" of various online and social media platforms to understand the inequalities and biases inherent in their respective algorithms, and I argue should be a point of future research and discussion in the public sector as well. Unfortunately, doing so for algorithms that drive public policy is highly unlikely, given the bureaucracy and nature of many local governments.

This risk of marginalization without accountability lies not only with the proposed algorithms from this research, but also within the greater set of algorithms that are increasingly being used in the public and private sectors related to the real estate market, health outcomes, and education. Therefore, it is imperative to discuss the broader challenges of such data manipulation and analysis within this context, and what it might mean for data-driven decision-making moving forward.

Inconclusive Evidence: "Evidence"-based Decision-making

Uncertainty abounds when the algorithm is not predicting with a high-enough level of accuracy, and false positives and false negatives are frequent. This concern is part of a broader discussion of the epistemic issues of data analysis in policymaking and should be discussed more specifically in future research. Within the present data and model, the results outlined in Chapter 2 show that accuracy metrics can be good – yet still not show the whole picture of what an algorithm might be missing. Understanding where the machine learning model might be inaccurate is challenging, though the final accuracy metric is an appealing measure by which people might like to make decisions, despite its uncertainty.

There is limited ability within policymaking to handle uncertainties such as these, and policies driven by analyses such as this can be significant contributors to unfair outcomes. Mittelstadt et al.'s review of the literature on what amounts of data and correlations are credible for prompting action shows that there is typically only need for a "sufficient" amount of evidence from an algorithm to drive any sort of action, even when correlations or predictions in algorithms are particularly uncertain (Mittelstadt et al., 2016). As Ananny (2016) notes, " algorithmic categories…signal certainty, discourage alternative explorations, and create coherence among disparate objects" (Ananny, 2016). These classifications can then result in an oversimplification that might result in decisions made from inaccurate or uncomplex data (Barocas, 2014). In the context of this research, processes such as the unsupervised learning of k-means clustering is a perfect example of categories being determined by the algorithm, validated in part by the descriptive statistics and a normative understanding of "neighborhood change," and presumed, for the purposes of the subsequent machine learning models, to be a way of "cohering" what might have seemed like disparate data points. Without engaging residents of the census tracts themselves to understand the lived experience as it relates to these data, it is impossible to truly assign these categories for the purposes of decisionmaking – an inherent flaw in much public policy within the realm of science and technology (Joss & Brownlea, 1999).

Inscrutable Evidence: Tension Between Evidence, Knowledge, and Autonomy

When ground-truthing does not accompany the results or processes of complex data analytics and machine learning, even transparency of the data processing and machine learning algorithm itself do not sufficiently simplify how classifications are made. Oftentimes, this is because the necessary components of being "transparent" includes being both comprehensible and accessible. Cloaked in predictive measures such as machine learning algorithms, additional layers of complexity are added. By utilizing data in opaque processes such as machine learning algorithms, government becomes less accessible, trading off transparency for efficiency and limiting potential for oversight (Burrell, 2016; Schermer, 2011).Within this scenario, the power differential – between those who are familiar with the data and its processing and an average citizen with limited knowledge of sophisticated data analytics—is stark. Even when algorithms or their creations are made public, it can be difficult to understand not only the "black box" of the algorithm that delivers results without explanation. Again, though this is not within the full scope of this research, it deserves additional conversation as it relates to policymaking in the context of social, racial, and economic justice issues.

Zarsky (2016) outlines the challenges of the rise of algorithms in decisionmaking as those of efficiency and fairness-based concerns, specifically related to the opaque and automated natures of algorithms. These tools do not necessarily change any sort of decision a public entity might make, rather, they can serve as a tool behind which public officials can hide in the name of "fairness," with the "objective" algorithm driving the decisions. Within in their creation and use, the algorithms can hide a political agenda, either intentionally or unintentionally – in this case, one in which rising socioeconomic status and population numbers are considered indicators of wellbeing, even though lived experiences might suggest other types of indicators. This design upholds a political status quo despite the appearance of objectivity or fairness. Specifically, within the context of this research, indicators of increasing population and other measure of "wellbeing" as determined by the quantitative indicators are seen within policy spaces as successful components of increasing wellbeing, without knowing if residents' lived experiences align with these measures. In fact, residents may have different conceptions of what constitutes their "wellbeing" that may not be captured in the data or algorithms, and this is not typically considered by policymakers, as instead there are "data-driven" decisions to hide behind.

There is an important role for disclosure of the data and the algorithms' structures, though this can also result in bad actors attempting to "game the system" (Zarsky, 2016). Machine learning algorithms such as the one proposed in this paper have the potential to be similar tools of discrimination. Algorithms and other data mining projects could lead to de facto redlining of a new type – where development and changes to the landscape are confined to only select neighborhoods because they might offer a higher return on investment. Similarly, if analyzed incorrectly, one might improperly infer a certain set of attributes about residents of certain parts of the city, and make decisions accordingly, which is another form of bias (Barocas 2014). Still, it is within the interest of citizens – who are considered the 'data subjects', or those from which the data was created' – to know how data about them is being utilized, especially when there might be harmful ramifications for some of these citizens.

Misguided Evidence: Value-Laden Processes leading to Biased Decision-making

The bias that is written into this algorithm is not that of assigning positive or negative sentiment to any of the data points, but rather, a manufactured conception of socioeconomic status. This manufacturing – in this case, aligned with market-related conditions of the urban environment – is a reflection of the Mayor's political and economic objectives. Utilizing Friedman and Nissenbaum's (1996) conceptualization of how bias in algorithms can arise from social values found in the institutions and attitudes where the technology is produced, Mittelstadt (2016) et al highlight how, "Algorithms inherently make biased decisions. An algorithm's design and functionality reflects the values of its designer and intended users…development is not a neutral, linear path." This section explores how social biases – those which are embedded within a system or structure that reflects its overarching values – and how technical biases, or those related to the technical limitations of the data or algorithms, can create misguided evidence which leads to bias in decisionmaking (Friedman & Nissenbaum, 1996).

Inherently, predicting neighborhood change by socioeconomic status could be thought of as a discriminatory act – identifying a class of people based on their class, utilizing proxies of wealth, homeownership, and population changes (Sandvig et al., 2014). Much of data mining and machine learning rely on data sources that reflect discriminatory policies and economic systems – and the decisions that occur after the data collection and analysis are the typical outcomes of such processes. Discriminatory acts of the mid-20th century such as redlining were in part due to data collection that utilized data as proxies for race and appeared on the surface to be non-discriminatory by such measures (Croll, 2012; Diakopoulos, 2014; Safransky, 2019; Sandvig, 2016).

Social Bias

It is critical to understand that these data are not inherently "objective," and that each choice made when designing research or constructing an algorithm is filled with the values of the human behind the data (Barocas et al., 2019). In the context of neighborhood change, each of these decisions to select variables or proxies of lived experience using data makes a variety of assumptions about what "change" can be defined as, and calculates these changes in an abstracted way from the lived experience – giving power to these data at an arms length from the conditions that underlie the numbers. In the case of this research project, I utilized a variety of qualitative research that discusses how "ascent" is measured in urban environments and what data can be analyzed to understand "socioeconomic status". Though I attempted to ensure my research was thorough, unbiased, and objective, there are limitations to such an approach, and some assumptions were made for each step.

Furthermore, my research uses this concept of "misguided evidence" by the nature of the data. The history of structural racism is built into the data that I used – as noted in Chapter one, Detroit developed unequally through systemic structures that marginalized the black population through racist local and federal policies. Each of these indicators – of housing, population mobility, real estate, race, and others – bears with it this history of Detroit being shaped through racist housing, policing, and economic policies. Policy choices of the past have led to Detroit's spatial, socioeconomic, and racial inequalities today, which is baked into the contemporary data which was analyzed in this research, and is used in policymaking on a daily basis. A deeper discussion of this point is warranted, though is beyond the scope of this research.

Technical Bias

Within the context of this research, the limitations of the chosen data are inherently ethical challenges. Utilizing a small dataset such as the one in this research is one of the reasons why the accuracy metrics were somewhat limited it their explanatory power. Similarly, should any of the steps in creating the algorithm – from determining the dependent and independent variables to the model's actual structure – stray from statistically or computationally sound practices, the results could be skewed without an overt connection to the errors made in the creation of the algorithm. These characteristics are not uncommon for machine learning algorithms, and can result in unfair outcomes based on this misguided evidence.

Unfair Outcomes: Marginalization and Discrimination

Undergirding this discussion is the definition of harm – that the potential for harm exists at every step in a process that attempts to utilize machine learning to impact a societal outcome. Violence and harm exist in a multitude of forms throughout our society, and the use of algorithms, data, and information in decision-making where the outcomes may not be equal is one of the ways in which this violence and harm can manifest. Structural and systemic racism have embedded inequalities within daily life, and the data that is utilized to make decisions draws from an already unequal world, and often reinscribes this inequality. This research utilizes the framework and definition of algorithmic violence, defined by Safransky (2019) as " … a repetitive and standardized form of violence that contributes to the racialization of space and spatialization of poverty." In Safransky's analysis of the use of Market Value Analysis in Detroit in the early 2010s, she highlights the importance of understanding the use of algorithms such as this within the context of algorithmic violence, given that the increasing use of data to drive decision-making has the potential to diffuse the blame of unpopular decision-making – and having it rest on the "objective" nature of the data.

One of the primary questions that initiated this research was a curiosity to understand what data can help policy and decision makers in their daily work, and to investigate the ethical dilemma faced when making decisions based on data, and not necessarily lived experiences. Specifically, I was curious to know, within the context of the SNF investment vessel, is there a way in which data can help show policymakers that such an initiative might, in fact, be increasing inequality in the city? And could these results direct resources toward the initiatives and areas of highest need, instead of trying to get the best return on real estate investment? This research shows that while data can show the inequalities that are present – and predict a continuance of such inequalities with some reliability – there is plenty of fallibility within the data analytics and data science processes and the decisions that are made from these results that can result in greater harm than good. This section will investigate these questions, and engage with the question of what parameters need to be in place to ensure ethical data analysis and policymaking.

With an intent to implement policy that aligns with political objectives, data is often used to validate policy and programs – utilizing peoples' data – data about their lived experiences – to direct policy. This is not to say that all data and its use in policymaking is bad. These same data can often be used to direct the most critical services to areas of highest need, or can validate the effectiveness of a current program or process that similarly helps those of most need. The costs and benefits of each use of data should be deliberated in the context of the normative policy goals – i.e., why the data is being used in the first place – and the integral ethical questions of using data and machine learning algorithms in the public realm similarly deserve investigation. While data points of a population can be helpful for policy and decision makers to sort through the massive data arrays and attempt to limit biases present in testimonies or anecdotes, the utility of these data points is a double-edged sword. There are ways in which these data can be used to vastly improve peoples' lives – individually and collectively – and there are ways in which these data can be used to further engrain the capitalistic monopolization of real estate by way of gentrification. Arguably, that the results of data analysis is likely to be sculpted to fit whatever political or policy priorities exist – and the use of data to drive policymaking is ultimately a policy decision in and of itself.

Detroit is currently experiencing a wave of gentrification that is closely linked to large-scale capital, as has been seen in cities like New York in the late 1990s – where the public sector is recruiting large developers to address challenges at the neighborhood-level (Hackworth, 2019; Hackworth & Smith, 2001). This wave most recently happened immediately after the recession in the early 1990s in other cities, and Detroit is most familiar with the re-making of the Downtown and Midtown neighborhoods during the years following the Great Recession, and is continuing now into the neighborhoods beyond the core. These data of peoples' socioeconomic status is most useful to the public and private sectors to identify where developers might be able to purchase land at a lower cost and encourage a wealthier class to support their return on investment through the purchasing of homes and businesses (Mark Davidson & Lees, 2005). Our present societal reality is such that socioeconomic data – and beyond – is obtainable and usable by governments, private companies, and other large institutions to make similar choices: who should be targeted for a policy, a product, a store, and who should be avoided (Croll, 2012). Increasingly sophisticated measures of collecting data through cell phones, cars, and other sensors in the urban environment feed data to these entities – consent obtained through hard to understand legal waivers in applications, websites, and by simply being a citizen (Catlett et al., 2017; Shelton et al., 2015).

Urban investments in the real estate and social sectors exists within this context, and within the current neoliberal configuration that encourages the public-private partnerships of government and capital to find the best return on investment. To accomplish this, increasingly complex data analysis and manipulation is occurring, with few options for the public to understand how decisions are made and how much of their own data is being used within these analyses (Zarsky, 2016). As such, predictive measures could also be utilized by the public sector to identify which neighborhoods might need anticipatory governance measures to slow the potential displacement of residents and encourage inclusionary development. Similarly, the predictive measures could help decisionmakers address the areas of the city that are predicted to either remain static or decline in the future – where policies might approach those challenges differently. Generally, it is important to consider the flawed reliance on big data and sophisticated analysis techniques for policy decision-making, as outlined in depth in the following section.

Market Value Analysis – A Cautionary Tale

As of this writing, there is no publicly available knowledge about specific calculations or reasoning behind real estate and social investments in the City of Detroit. However, the following section outlines how the city has tried to use a data-oriented approach to investment before, and how this story serves as a cautionary tale for what a more advanced data analytics process like machine learning could do in the city. This narrative cautions that the use of data – either through machine learning algorithms or otherwise – can be problematic when used to achieve urban planning goals that are already rife with disagreement. The addition of my research on using machine learning to achieve urban policy and planning goals serves as another example of the ethical questions that arise within this context.

SNF and MVA

One of the ways in which data is utilized in urban planning and development is by tying social, demographic, and economic data to the real estate market. Known as Market Value Analysis (MVA) -- identifying neighborhoods based on their economic vitality as per the conditions of the real estate market -- this process has been criticized for being reminiscent of the redlining practices of the early part of the 20^{th} century. Developed and promoted by The Reinvestment Fund, a CDFI that partners with investors, governments, and other sectors to direct financial investments in cities, MVA analyzes a variety of metrics such as vacancy rates, housing prices, and commercial data to categorize neighborhoods based on their market vitality (Goldstein et al., n.d.). One of the key categories developed by the Reinvestment Fund through MVA is the "middle neighborhoods" – neighborhoods that have maintained a relatively stable housing stock and a predominantly middle class yet have seen some decline in various social and economic indicators. These neighborhoods are understood to be "tipping point" areas where cities can then direct investment in order to further "stabilize" the market. In a critique of MVA, Joshua Akers calls this process a shift toward an "emerging market city," a city which is managed for the market and micro-market geographies rather than for the population itself (Akers, 2015).

Admirably, SNF appears to be an asset-based community development approach to revitalization. In its brochure, SNF states that neighborhoods were selected based on criteria such as strong local leaders, neighborhood density, and proximity to historic areas and other neighborhood assets – components that do not appear to be strictly market oriented. However, by implementing "tools" such as neighborhood planning, improving streetscapes, redeveloping parks, rehabilitating homes, and strengthening commercial corridors – primarily infrastructure-based projects – the approach appears to be more market-oriented than socially-driven by other community needs.

The SNF approach appears to have a similar method as MVA, the method of choice for the approach of the Detroit Works Project and Detroit Future City plan of 2011. DWP encountered significant resident pushback and broader political problems when it was made clear that the process selected “winners and losers” to receive increased investment and attention, and other areas would be repurposed for green or industrial spaces (Safransky, 2014, 2019). From present analysis there has been more limited pushback during the implementation of SNF, though a deeper analysis of the neighborhood planning component and the broader public engagement process is warranted to validate resident perceptions.

SNF’s presentation of its investment plan in its primary public brochure has some of the components that were noted in the original DWP project as problematic for residents. As Akers (2015) notes, the shift toward a “trickle-down model for the allocation of public resources” is one which prioritizes the market components of a neighborhood rather than any of the present social, economic, or other realities which the neighborhood might be facing. The similarities in approach are highlighted in SNF’s “emergence” from the Woodward Corridor initiative, a “transit-oriented development plan.” Like DWP’s focus on transit plans and long-term development in various neighborhoods, SNF is similarly focusing on these long-term development initiatives, perhaps at the expense of some of the neighborhoods’ immediate needs. The fund is actively utilizing the learnings from the revitalization of Midtown and Downtown in the target neighborhoods, focusing on stabilizing the housing and commercial sectors, driven by the idea that this will maintain and increase the attractiveness of the neighborhood. In this sense, market data is the driving force of revitalization without much transparency as to this is the strategy and if there is a potential component of people-driven decision making. While SNF is explicitly attempting to revitalize neighborhoods and increase wellbeing for all residents, it is implicitly seeking to attract higher income residents, and change the socioeconomic status of many of these neighborhoods. Many residents are concerned that with this market-based focus, long-time residents may no longer be able to afford to live in these neighborhoods, or long-time Detroiters will not be able to afford to move to these neighborhoods once the amenities are in place (Carlisle, 2020).

This is not to say that the community does not need these amenities – these neighborhoods can certainly benefit from improvements in each component of SNF. Most people would like to live in neighborhoods that are aesthetically pleasing and have a range of housing, shopping, and quality of life options. However, it is unclear how these improvements, driven by the private sector, are likely to address some of the immediate social and economic needs of residents in these neighborhoods such as income or health disparities – or that these concerns were considered initially.

Specifically, it is unclear how an SNF objective such as “neighborhood planning to understand community needs” necessitates a focus on single family housing and strengthening the commercial corridors – what are some of the other community needs that might exist, and could SNF also address those needs? SNF does not provide any data or narratives that support the idea that single family housing and commercial corridors are, in fact, the primary needs of neighborhood residents. Though SNF identifies each neighborhood’s already existent assets, there is little public information about what components of SNF will address some of the social and economic challenges facing the neighborhoods as these changes are taking place. As is understood from the literature on machine learning and data ethics, market data – or data utilized to proxy a changing market – is one mechanism by which a program such as SNF is likely designed and can be assessed, reflecting and extending the broader political values that brought the initiative to life. The quantitative part of this research highlights these market-related data as likely helpful components of predicting neighborhood change, and align with this idea that cities could use these market-based data for reimagining the urban landscape behind the argument of seeing increases in "wellbeing" indicators".

It is currently unclear if SNF and those involved are interested in community well-being data – data related to poverty levels, employment, education, among others that might be identified by residents not captured in the market-based data. As a tool for infrastructure development and creating "emerging markets" in the neighborhoods, SNF is likely to be successful. This lack of reflexivity points to a focus on the more market-led components of their toolkit – single family housing and commercial corridors.

Transformative Effects: Reontologizing the Urban Fabric

The methodology and results of my research supports Safransky's claim that these data tools are contributing to an unequal development of the urban environment through policies designed in the image of the data analysis. Particularly, the indicators that are utilized in my methodology are most closely related to real estate markets and a sense of financial wellbeing – number of building permits, mortgages, housing prices, incomes, and population changes. The differences in these indicators in census tracts that are considered to be "ascending" in socioeconomic status highlight the inequalities present in the city's fabric – data which could also lead to policy decisions that continue this kind of investment, rather than addressing the challenges of socioeconomic and racial inequality associated with these data. Specifically, within the framework of the Strategic Neighborhood Fund, an analysis such as this one could arm policymakers with evidence that the investments are working, even though the data show that to be the case only in certain parts of the city. With the machine learning algorithms predicting these changes relatively accurately, policymakers could utilize this methodology to drive future decision-making, embedding the assumptions of what constitutes as "socioeconomic indicators or status" identified in the algorithmic process – into the policies.

This embedding of assumptions relates most closely to the transformative effects component of Mittelstadt et al.'s (2016) normative framework – how unfair the evidentiary problems of algorithms can lead to transformative ways in which we think about our environment. With inconclusive and misguided evidence, using any of the processes or algorithms in this research might lead to policy decisions that label parts of the city as "ascending" or "not ascending," potentially embedding these labels into the fabric of the city through these normative policy measures that direct investments to certain neighborhoods and not others. Using the results of an algorithm to determine what neighborhoods are "good," sounds like a far-fetched possibility, though is actively being pursued in public and private sectors alike, and results of these processes might be changing how we conceptualize and talk about neighborhood change (Chapple & Zuk, 2016; *NNIP Mission | NNIP*, n.d.; Pettit et al., 2019; Reades et al., 2019; Royall & Wortmann, 2015; Winston & Walker, 2012).

Policy implications and recommendations

Designing algorithms to predict neighborhood change – though potentially accurate – is likely not the necessary tool to affect a radical rethinking of community and economic development practice. Although it can serve as an important validation tool for community organizations to communicate with policymakers, the likelihood of such a tool being utilized to maintain the status quo is high. Safransky (2019) argues that this is an intentional component in the rise of data-driven decision-making – utilizing data to absolve humans of potentially controversial ethical decisions. From a policy perspective, this condition must be investigated with greater depth – is there a way in which policy decisionmaking can avoid this ethical trap and limit the harms that may result from such actions?

In fact, the findings of my research could support an argument in defense of the SNF initiative, and that argument could rest on the data at hand – that investments in these neighborhoods are working, and are achieving the policy outcomes that were intended, such as increased "wellbeing" by nature of the physical environment. Alternatively, these data and methodologies can also be utilized to advocate for a different kind of policy approach – one that might encourage investment or policies that are directed at different well-being measures, likely to be more socially-related, or defined by residents, or that are directed to neighborhoods that might not be attracting a population of increasingly higher socioeconomic status. However, the potential to use these data for "good" purposes – in alignment with community needs and desires – cannot be done in the vacuum of following these specific data. Rather, a more comprehensive approach that centers the community's needs, and may not rely on these data as much, is likely to result in better policy outcomes, addressing a more comprehensive set of needs.

For the sake of academic writing, the tool created by this research serves its purpose --- it predicts with 80% accuracy the likelihood of a neighborhood's socioeconomic ascent or decline. For the sake of implementing in a real-world setting, this tool does nothing more than pick winners and losers – what the current administration and structure already accomplishes. The data clearly shows parts of the city that have experienced stasis or decline on a variety of indicators, but investments are concentrated in the neighborhoods that have higher socioeconomic indicators. These indicators – housing and population related – are primarily tied to real estate and investment markets, not to basic needs and other components of wellbeing. Most importantly, qualitative evidence from residents would be critical for complementing this tool and offering a ground-truth that the data can validate. This is a space in which further research should be directed – seeking to understand how to best encapsulate residents needs and definitions of wellbeing, likely through some form of anticipatory governance that develops methodologies for understanding these concerns.

Policymakers should consider the ethical challenges of algorithms when designing policy, seeking to understand how much of the evidence is possibly misguided, inscrutable, or inconclusive and what the transformations of society might occur based on this evidence. As the public sector's expertise in advanced analytics continues to grow, there will be a need for an accountability measure, identifying the underlying assumptions of algorithms and analyses alike. Additionally, elected officials and appointed bureaucrats alike should take caution in the ability of data to direct policies and responses to a changing environment. The potential for harm from this process is significant, and there are few ways in which the public can hold accountable those who might be responsible.

While this is, arguably, a feature of data-driven decisionmaking today – not a bug – it is imperative that research institutions such as the University of Michigan's Center for Poverty

Solutions continue to investigate the ground-truth lived experiences of residents through surveys and focus groups, such as the Detroit Metropolitan Community Survey (DMACS). Organizations like Data Driven Detroit are equally critical non-partisan components of the system – although their primary purpose is to utilize data to inform policy and decisionmaking, within the context of neighborhood change, D3 takes the appropriate measures of ground-truthing the results with qualitative analysis through focus groups.

However, what is lost in this broader process are the political choices that are likely to take shape that are masked behind the veil of the data manipulation and analysis. Even if the final product is accessible and tangible to the citizen public, it does not mean that decisions based on the data manipulation and analysis have any concern for what the public think should be done with the data. Oftentimes, decisions based on data crunching are made by non-elected bureaucrats – people who are hired into a government. With little accountability, these decisionmakers – policymakers – craft and shape a city to obtain certain goals. These goals are often purported to be in the public interest – helping the worst off, maintaining a competitive economic and regulatory environment, creating a better city for the future. However, it is necessary to investigate the methods and processes by which these goals are achieved.

As noted above, one of the important components of future research would be to focus on data that is not strictly market oriented. What does it mean to analyze employment, education, or health data in these neighborhoods and how could it contribute to a more comprehensive approach to SNF that also centers some of the more immediate lived experiences of its residents? Similarly, what would it mean for policy priorities to reflect resident well-being over market-based data or understandings? This is not to say that a failing housing market is not central to many residents' current lived experiences – it most certainly affects people's mobility and housing choices, safety, and other components of a healthy urban environment. However, with SNF's primary focus on the "emerging markets" of these neighborhoods it is likely that the other components of wellbeing may get lost in the shuffle. Additional data analysis can offer a more comprehensive understanding of the neighborhoods when combined with the qualitative data and other survey-based research. It is important to collaborate these two research processes – by reducing neighborhoods to market data there are important nuances that are often lost in the process.

Conclusion

So what can the "analyst" do to counteract the preeminence and power of data in decision-making spaces? One of the key components to breaking down this barrier is by ensuring that complex predictive tools are made accessible to everyone. Not only should a technical expert be able to manipulate the tool, but so should anyone with a basic level of technology understanding (i.e. using a computer or a phone). If the algorithm itself is not easily manipulatable, then the variety of analyses of the findings should be made accessible. Through this accessibility, it is possible that residents could use the findings of a predictive algorithm to self-advocate for changes in their own neighborhoods. Not only can public officials make data-informed decisions, but so, too, can the public – and decide to advocate for the resources and policies they need to thrive.

Although data can be utilized as a measure of socioeconomic status and to predict future status in city neighborhoods, it is imperative that scholars, policymakers, and community organizations think critically about the use of data in such circumstances. The commodification of a neighborhood through data is inherent in most societal and market structures and solidifying the use of these data to further policymaking goals should be done with caution. Processes and policies should be established to center resident needs, and investigate the multitude of social, political, and economic justice issues that are present to understand if machine learning can be used as a mechanism to address these needs. Without this, we are likely to hide behind the numbers as unequal and unjust structures of our society remain in place.

Appendix A

Independent Variables	Source
Total Population Estimate	ACS
Percent estimate of residents under age 5	ACS
Percent estimate of white residents	ACS
Total number of housing units	ACS
Percent Estimate of vacant housing units	ACS
Percent Estimate of Renters	ACS
Median Rent	ACS
Percent Estimate of Renters who are housing burdened (housing costs between 31-35% of income)	ACS
Percent Estimate of residents who are unemployed	ACS
median household income	ACS
Percent Estimate of residents receiving social security income	ACS
Percent Estimate of residents receiving SNAP benefits	ACS
Percent estimate of residents living below poverty line	ACS
Percent estimate of residents over age 65	ACS
Percent estimate of residents with a High School degree or less	ACS
Percent estimate of residents with a Bachelor's degree or higher	ACS
Number of building permits issued by City of Detroit	City of Detroit
Number of mortgage applicants denied	HMDA
Number of mortgages issued	HMDA
Median mortgage amount	HMDA
Total number of mortgage applications	HMDA

Bibliography

Adler, L. (2017, March 20). Planning the Data-Driven City. *Data-Smart City Solutions*. https://datasmart.ash.harvard.edu/news/article/planning-the-data-driven-city-1003

Aguilar, L. (2018, December 10). *Seven Detroit neighborhoods to get $35M boost for development*. Detroit News. https://www.detroitnews.com/story/news/local/detroit-city/2018/12/10/seven-detroit-neighborhoods-get-35-million-dollars-boost-development/2262979002/

Akers, J. (2015). Emerging market city. *Environment and Planning A: Economy and Space*, *47*(9), 1842–1858. https://doi.org/10.1177/0308518X15604969

Ananny, M. (2016). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology, & Human Values*, *41*(1), 93–117. JSTOR.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias—ProPublica. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Archambault, D. (2019, October 10). *Can anything be built in Detroit without subsidies?* Curbed Detroit. https://detroit.curbed.com/2019/10/10/20907932/detroit-development-subsidies-ford-station-fca-plant

Barocas, S. (2014). *MINING AND THE DISCOURSE ON DISCRIMINATION*.

Barocas, S., Hardt, moritz, & Narayanan, A. (2019). *Fairness and machine learning*. fairmlbook.org. https://fairmlbook.org/

Beshouri, P. (2013, February 20). *7.2 Square Miles*. Curbed Detroit. https://detroit.curbed.com/2013/2/20/10271684/72-square-miles

Boyle, R. (2016, June 15). Could the 20-minute neighborhood work in Detroit? *Detroit Free Press*. https://www.freep.com/story/opinion/contributors/2016/06/14/could-20-minute-neighborhood-work-detroit/85847554/

Brownlee, J. (2015, August 18). 8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset. *Machine Learning Mastery*. https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/

Brownlee, J. (2020, January 2). How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification. *Machine Learning Mastery*. https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 2053951715622512. https://doi.org/10.1177/2053951715622512

Cahill, C. (2007). Negotiating Grit and Glamour: Young Women of Color and the Gentrification of the Lower East Side. *City & Society*, *19*(2), 202–231. https://doi.org/10.1525/city.2007.19.2.202

Carlisle, J. (2020, May 24). Detroit neighborhood group sees gentrification as the enemy. *Detroit Free Press*. https://www.freep.com/in-depth/news/columnists/john-carlisle/2020/05/24/detroit-neighborhood-gentrification-protest-carlisle/4954702002/

Catlett, C. E., Beckman, P. H., Sankaran, R., & Galvin, K. K. (2017). Array of things: A scientific research instrument in the public way: platform design and early lessons learned. *Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering*, 26–33. https://doi.org/10.1145/3063386.3063771

Chapple, K., & Zuk, M. (n.d.). *Systems for Gentrification*. 23.

Chapple, K., & Zuk, M. (2016). Forewarned: The Use of Neighborhood Early Warning Systems for Gentrification and Displacement. *Cityscape*, *18*(3), 23.

Ciaburro, G. (2018, January). *z score standardization—Regression Analysis with R*. https://learning.oreilly.com/library/view/regression-analysis-with/9781788627306/c0fdcc36-0503-45dd-8f6d-3201cd1fb857.xhtml

Compare the effect of different scalers on data with outliers—Scikit-learn 0.22.2 documentation. (n.d.). [Scikit learn]. Retrieved April 16, 2020, from https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html

Crawford, A. (2018, April 25). *What Will Become of Detroit's Suburban Malls, Stadiums, and Office Parks?* CityLab. https://www.citylab.com/design/2018/04/can-detroits-suburbs-survive-a-downtown-revival/558764/

Croll, A. (2012). *Big Data is our generation's civil rights issue, and we don't know it – Solve for Interesting*. http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it/

Davidson, M. (2008). Spoiled Mixture: Where Does State-led `Positive' Gentrification End? *Urban Studies*, *45*(12), 2385–2405. https://doi.org/10.1177/0042098008097105

Davidson, Mark, & Lees, L. (2005). New-build 'gentrification' and London's riverside renaissance. *Environment and Planning A*, *37*(7), 1165 – 1190. https://doi.org/10.1068/a3739

Derringer, N. (2014, August 21). *Bridge: In a gentrifying Detroit, an uneasy migration of urban millennials*. Bridge Magazine. http://bridgemi.com/2014/08/in-a-gentrifying-detroit-an-uneasy-migration-of-urban-millennials/

Derringer, N. (2017, April 20). Welcome to the New Detroit, white people. So long, poor folks. *Bridge Magazine*. https://www.bridgemi.com/detroit-journalism-cooperative/welcome-new-detroit-white-people-so-long-poor-folks

Detroit Future City. (2013). *Detroit Future City: 2012 Detroit Straategic Framework Plan*. Inland Press.

Detroit's Open Data Portal. (n.d.). Retrieved December 9, 2019, from https://data.detroitmi.gov/

D:hive. (n.d.). *D:hive: Detroit's Welcome Center*. D:Hive. Retrieved December 5, 2014, from http://dhivedetroit.org/

Diakopoulos, N. (2014). *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*. https://doi.org/10.7916/D8ZK5TW2

Doucet, B. (2014). A Process of Change and a Changing Process: Introduction to the Special Issue on Contemporary Gentrification: Introduction to the Special Issue on Contemporary Gentrification. *Tijdschrift Voor Economische En Sociale Geografie*, *105*(2), 125–139. https://doi.org/10.1111/tesg.12075

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First edition.). St. Martin's Press.

Felton, R. (2014, November 12). *Dan Gilbert, downtown Detroit's demigod*. Detroit Metro Times. http://www.metrotimes.com/detroit/what-kind-of-track-record-does-quicken-loans-have-in-detroit-does-anyone-really-care/Content?oid=2266383

FFIEC Home Mortgage Disclosure Act. (n.d.). Retrieved April 26, 2020, from https://www.ffiec.gov/hmda/

Florida, R. (2014, September 30). *A Conversation With Detroit Mayor Mike Duggan*. CityLab. http://www.citylab.com/politics/2014/09/a-conversation-with-detroit-mayor-mike-duggan/380900/

Foley, A. (2013a, June 19). *A Friendly Reminder That The Universe Does Not Revolve Around Midtown*. Jalopnik Detroit. http://detroit.jalopnik.com/a-friendly-reminder-that-the-universe-does-not-revolve-514165837

Foley, A. (2013b, July 29). *The First Shots In The War Against Detroit Hipsters Have Been Fired*. Jalopnik Detroit. http://detroit.jalopnik.com/the-first-shots-in-the-war-against-detroit-hipsters-hav-950892545

Foley, A. (2014, April 1). *Opinion: What we really mean when we say "The Neighborhoods."* Model D. http://www.modeldmedia.com/features/theneighborhoods4114.aspx

Friedman, B., & Nissenbaum, H. (1996). *Bias in computer systems*. Association for Computing Machinery. https://doi.org/10.1145/230538.230561

Gallagher, J. (2014, July 27). *One downtown, two empires: Mike Ilitch and Dan Gilbert reshape Detroit*. Detroit Free Press. http://archive.freep.com/article/20140727/NEWS01/307270087/Gilbert-Ilitch-Red-Wings-Tigers-Quicken-Detroit

Gallagher, J., & Bomey, N. (2013, September 15). *How Detroit went broke: The answers may surprise you — and don't blame Coleman Young*. Detroit Free Press. http://archive.freep.com/interactive/article/20130915/NEWS01/130801004/Detroit-Bankruptcy-history-1950-debt-pension-revenue

Galster, G., Tatian, P., & Accordino, J. (2006). Targeting investments for neighborhood revitalization. *Journal of the American Planning Association*, *72*(4), 457–474.

Gangadharan, E. S. P., Eubanks, W. V., & Barocas, S. (2014). DATA AND DISCRIMINATION: COLLECTED ESSAYS. *NEW AMERICA*, 6.

G.E.P. Box and D.R. Cox,. (1964). An Analysis of Transformations, Journal of the Royal Statistical Society B. *Journal of the Royal Statistical Society B*, *26*, 211–252.

Glass, R. (1964). London: Aspects of Change. In L. Lees, T. Slater, & E. Wyly (Eds.), *The Gentrification Reader*. Routledge.

Goldstein, I., Schrecker, W., & Rosch, J. L. (n.d.). *Demographics and Characteristics of Middle Neighborhoods in Select Legacy Cities* (Community Development Investment Review, p. 25). Federal Reserve Bank of San Francisco.

Guyette, C. (2014, August 29). *Detroit's Water Woes: A Payment Plan Is Not An Affordability Plan*. ACLU of Michigan. http://www.aclumich.org/democracywatch/index.php/entry/detroit-s-water-woes-a-payment-plan-is-not-an-affordability-plan

Hackney, S. (2014, September 28). *Is There Room for Black People in the New Detroit?* POLITICO Magazine. http://www.politico.com/magazine/story/2014/09/is-there-room-for-black-people-in-the-new-detroit-111396.html

Hackworth, J. (2019). Gentrification As A Politico-Economic Window: Reflections On The Changing State Of Gentrification: Gentrification As A Politico-Economic Window. *Tijdschrift Voor Economische En Sociale Geografie*, *110*(1), 47–53. https://doi.org/10.1111/tesg.12330

Hackworth, J., & Smith, N. (2001). The Changing State of Gentrification. *Journal of Economic & Social Geography*, *92*(4), 464–478.

Hammel, D. J., & Wyly, E. K. (1996). A Model for Identifying Gentrified Areas with Census Data. *Urban Geography*, *17*(3), 248–268. https://doi.org/10.2747/0272-3638.17.3.248

Hanlon, J. (2011). Unsightly Urban Menaces and the Rescaling of Residential Segregation in the United States. *Journal of Urban History*, *37*(5), 732–756. https://doi.org/10.1177/0096144211407744

He, S. (2010). New-build gentrification in Central Shanghai: Demographic changes and socioeconomic implications. *Population, Space and Place*, *16*(5), 345–361. https://doi.org/10.1002/psp.548

Heidkamp, C. P., & Lucas, S. (2006). Finding the Gentrification Frontier Using Census Data: The Case of Portland, Maine. *Urban Geography*, *27*(2), 101–125. https://doi.org/10.2747/0272-3638.27.2.101

Helbich, M., Brunauer, W., Hagenauer, J., & Leitner, M. (2013). Data-Driven Regionalization of Housing Markets. *Annals of the Association of American Geographers*, *103*(4), 871–889. https://doi.org/10.1080/00045608.2012.707587

Hillenbrand, K. (2016, June 9). Predicting Fire Risk: From New Orleans to a Nationwide Tool. *Data-Smart City Solutions*. https://datasmart.ash.harvard.edu/news/article/predicting-fire-risk-from-new-orleans-to-a-nationwide-tool-846

Hulett, S. (n.d.). *A "Tale Of Two Cities" As Detroit Looks To 2014*. NPR.Org. Retrieved November 19, 2014, from http://www.npr.org/2013/12/18/252021693/a-tale-of-two-cities-as-detroit-looks-to-2014

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). https://doi.org/10.1371/journal.pone.0212814

Invest Detroit. (n.d.). *Strategic Neighborhood Fund 2.0: One City. For all of us.* Invest Detroit. Retrieved April 26, 2020, from https://www.dropbox.com/s/n0r7xjn9p9x6edr/SNF2.0%20book%20-%20single%20page.pdf?dl=0

Jackle, J. A., & Wilson, D. (1992). *Derelict Landscapes: The Wasting of America's Built Environment*. Rowman & Littlefield Publishers, Inc.

Jackman, M. (2014, October 23). *The problem with Detroit: Too many poor homeowners*. Detroit Metro Times. http://www.metrotimes.com/Blogs/archives/2014/10/23/the-problem-with-detroit-too-many-poor-homeowners

Jackman, M. (2016, November 23). In the "other Detroit," inequality and poverty are increasing | News Hits. *MetroTimes*. https://www.metrotimes.com/news-hits/archives/2016/11/23/in-the-other-detroit-inequality-and-poverty-in-detroit-are-increasing

Joss, S., & Brownlea, A. (1999). Considering the concept of procedural justice for public policy- and decision-making in science and technology. *Science and Public Policy*, *26*(5), 321–330. https://doi.org/10.3152/147154399781782347

Kenneally, E., & Dittrich, D. (2012). The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2445102

Knorr, D. (2019). *Using Machine Learning to Identify and Predict Gentrification in Nashville, Tennessee.* [Vanderbilt University]. https://etd.library.vanderbilt.edu//available/etd-07192019-110344/unrestricted/07242019_Dknorr_Thesis_Final2.pdf

Kontokosta, C. E., & Jain, R. K. (2015). Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy. *Sustainable Cities and Society*, *18*, 44–55. https://doi.org/10.1016/j.scs.2015.05.007

Lee, B. A., & Mergenhagen, P. M. (1984). Is Revitalization Detectable?: Evidence from Five Nashville Neighborhoods. *Urban Affairs Quarterly*, *19*(4), 511–538. https://doi.org/10.1177/004208168401900407

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes. *Philosophy & Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

Ley, D. (1994). Gentrification and the Politics of the New Middle Class. *Environment and Planning*.

Live Midtown. (n.d.). *Incentives*. Livemidtown. http://www.livemidtown.org/incentives

Lützeler, R. (2008). Population Increase and “New-Build Gentrification” in Central Tōkyō. *Erdkunde*, *62*(4), 287–299. JSTOR.

Mack, J. (2019, September 26). Flint and Detroit among nation’s top 5 poorest cities, new census data shows—Mlive.com. *MLive*. https://www.mlive.com/news/2019/09/flint-and-detroit-among-nations-top-5-poorest-cities-new-census-data-shows.html

Mallach, A. (2018). *The Divided City: Poverty and Prosperity in Urban America*. Island Press. http://ebookcentral.proquest.com/lib/umichigan/detail.action?docID=5406162

Manick, B. (2014, March 16). Feedback: Let established groups, not new startups, help minority youths. *Detroit Free Press*. http://archive.freep.com/article/20140316/OPINION04/303160032/Feedback-Let-established-groups-not-new-startups-help-minority-youths

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association*, *110*(512), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710

Mondry, A. (2020, January 23). New report shows Detroit diversifying, but anxieties around race remain. *Curbed Detroit*. https://detroit.curbed.com/2020/1/23/21078956/detroit-demographics-black-white-diversity

Moskowitz, P. (2017). *How to Kill a City: Gentrification, Inequality, and the Fight for the Neighborhood*. PublicAffairs. http://ebookcentral.proquest.com/lib/umichigan/detail.action?docID=5369164

NNIP Mission | NNIP. (n.d.). National Neighborhood Indicators Partnership. Retrieved April 26, 2020, from https://www.neighborhoodindicators.org/about-nnip/nnip-mission

Oosting, J. (2010, August 17). Mayor Dave Bing seeks input on plan to rightsize Detroit—Without eminent domain—Mlive.com. *MLive*. https://www.mlive.com/news/detroit/2010/08/detroit_mayor_dave_bing_sets_f.html

Owens, A. (2012). Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent: CITY & COMMUNITY. *City & Community*, *11*(4), 345–369. https://doi.org/10.1111/j.1540-6040.2012.01412.x

Pandey, E. (2019, November 20). *Detroit making a comeback—But the progress is uneven*. Axios. https://www.axios.com/comeback-detroit-downtown-rust-belt-michigan-84ddf203-98a4-4033-aceb-2c48c5d61220.html

Pettit, K. L. S., Cohen, M., & Levy, D. K. (2019, April 19). *Turning the Corner: Lessons from Five Cities on Displacement Risk in Changing Neighborhoods*. Urban Institute. https://www.urban.org/research/publication/turning-corner-lessons-five-cities-displacement-risk-changing-neighborhoods

Poethig, E. C., Schilling, J., Goodman, L., Bai, B., Gastner, J., Pendall, R., & Fazili, S. (2017, March 3). *The Detroit Housing Market*. Urban Institute. https://www.urban.org/research/publication/detroit-housing-market

powell, john a. (2013). Structural Marginalization john a. Powell.pdf. *Poverty & Race*, *22*(5).

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51–59. https://doi.org/10.1089/big.2013.1508

Quesnelle, S., Urban, N., & Rubio, A. (2019). *Turning The Corner*. Data Driven Detroit for The Urban Institute. https://datadrivendetroit.org/blog/2018/03/22/turning-the-corner/

Quinn. (2016, November 1). Modern Slavery. *Trajectory Magazine*. https://trajectorymagazine.com/modern-slavery/

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

Redesigning Detroit: Mayor Mike Duggan's blueprint unveiled. (2015, August 18). *Detroit Future City: News*. https://detroitfuturecity.com/2015/08/18/redesigning-detroit-mayor-mike-duggans-blueprint-unveiled/

Reese, L. A., Eckert, J., Sands, G., & Vojnovic, I. (2017). "It's safe to come, we've got lattes": Development disparities in Detroit. *Cities*, *60*, 367–377. https://doi.org/10.1016/j.cities.2016.10.014

Rich, W. C. (1989). *Coleman Young and Detroit Politics: From Social Activist to Power Broker*. Wayne State University Press.

Richardson, J., Mitchell, B., & Franco, J. (2019). *Shifting neighborhoods: Gentrification and cultural displacement in American cities*. NCRC. https://ncrc.org/gentrification/

Royall, E., & Wortmann, T. (2015). *Finding the State Space of Urban Regeneration: Modeling Gentrification as a Probabilistic Process using k-Means Clustering and Markov Models*. 24.

Runyan, R. (2016, June 15). *The Mayor's vision for 20-minute neighborhoods*. Curbed Detroit. https://detroit.curbed.com/2016/6/15/11946166/mayor-detroit-neighborhoods-walk-bike

Safransky, S. (2014). Greening the urban frontier: Race, property, and resettlement in Detroit. *Geoforum*, *56*, 237–248. https://doi.org/10.1016/j.geoforum.2014.06.003

Safransky, S. (2019). Geographies of Algorithmic Violence: Redlining the Smart City. *International Journal of Urban and Regional Research*, *n/a*(n/a). https://doi.org/10.1111/1468-2427.12833

Sands, L. A. R., Gary. (2017, February 19). *Is Detroit Really Making a Comeback?* CityLab. https://www.citylab.com/housing/2017/02/detroits-recovery-the-lass-is-half-full-at-most/517194/

Sandvig, C. (2016). *When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software*. 19.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*. 23.

Saunders, P. (2018a, July 19). *Detroit, Five Years After Bankruptcy*. Forbes. https://www.forbes.com/sites/petesaunders1/2018/07/19/detroit-five-years-after-bankruptcy/

Saunders, P. (2018b, October 30). *Can Detroit's Suburbs Survive The City's Rebirth?* Forbes. https://www.forbes.com/sites/petesaunders1/2018/10/30/can-detroits-suburbs-survive-the-citys-rebirth/

Schaefer, J., & Walker, M. (2013, November 6). *Detroit elects first white mayor in years—And reasons go well beyond race*. Detroit Free Press. http://www.freep.com/article/20131106/NEWS01/311060037/mike-duggan-white-mayor-predominantly-black-Detroit-race

Schermer, B. W. (2011). The limits of privacy in automated profiling and data mining. *Computer Law & Security Review*, *27*(1), 45–52. https://doi.org/10.1016/j.clsr.2010.11.009

Schuetz, J. (2017, December 8). Metro areas are still racially segregated. *Brookings*. https://www.brookings.edu/blog/the-avenue/2017/12/08/metro-areas-are-still-racially-segregated/

Seelye, K. Q. (2011, March 22). Detroit Population Down 25 Percent, Census Finds. *The New York Times*. http://www.nytimes.com/2011/03/23/us/23detroit.html

Shaw, D. (2001). The Post-Industrial City. In R. Paddison (Ed.), *Handbook of Urban Studies*. SAGE.

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, *142*, 198–211. https://doi.org/10.1016/j.landurbplan.2015.02.020

Slater, T. (2006). Eviction of critical persectives from gentrification research Slater 2006.pdf. *International Journal of Urban and Regional Research*, *30*(4), 737–757.

Smith, N. (1996). *The New Urban Frontier*. Routledge.

Sugrue, T. (1996). *The Origins of the Urban Crisis: Race and Inequality in POstwar Detroit*. Princeton University Press.

Terek, D., & Guralnick, D. (2014). *Newcomers move into bankrupt Detroit*. http://www.detroitnews.com/longform/news/local/wayne-county/2014/09/16/against-the-tide-newcomers-move-into-bankrupt-detroit/15736435/

The Michigan Citizen. (2014, May 1). *Two Detroits? Gentrification. talk gets real – The Michigan Citizen*. http://michigancitizen.com/two-detroits-gentrification-talk-gets-real/?utm_content=buffer27413&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer

Thomas, J. M. (2013). *Redevelopment and Race: Planning a Finer City in Postwar Detroit*. Wayne State University Press.

Thompson, B. (2017a, June 21). Bankole: Archer, Bing differ widely on today's Detroit. *Detroit News*. http://www.detroitnews.com/story/opinion/columnists/bankole-thompson/2017/06/21/bankole-bing-archer-detroit-resurgence/103087562/

Thompson, B. (2017b, July 9). Detroit is "booming" again. You have to be rich and powerful to notice, though | Bankole Thompson. *The Guardian*. http://www.theguardian.com/commentisfree/2017/jul/09/detroit-economic-recovery-poverty-mike-duggan

Thompson, B. (2019, June 5). Bankole: Ex-mayor Bing's black leadership message fails. *Detroit News*. https://www.detroitnews.com/story/opinion/columnists/bankole-thompson/2019/06/06/bankole-ex-mayor-bings-black-leadership-message-fails/1351032001/

U.S. Census Bureau. (2008). *A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know* (p. 68).

U.S. Census Bureau QuickFacts selected: Detroit city, Michigan. (2017). https://www.census.gov/quickfacts/fact/table/detroitcitymichigan/PST045216#qf-headnote-a

William H. Frey, Brookings Insitution, and University of Michigan Social Science Data Analysis Network. (2011). *Dissimilarity-Metros-Black*. http://censusscope.org/2010Census/PDFs/Dissimilarity-Metros-Black.pdf

Williams, C. (2013, November 6). *Detroit mayor-elect says he resents focus on race*. Yahoo News. http://news.yahoo.com/detroit-mayor-elect-says-resents-focus-race-220639912.html

Williams, J. (2020, January 22). A Tale of Two Motor Cities: Amid Detroit's Rebirth, Many African Americans Feel Left Behind | Cities | US News. *US News*. https://www.usnews.com/news/cities/articles/2020-01-22/amid-detroits-rebirth-many-african-americans-feel-left-behind

Winston, F., & Walker, C. (2012). *Predicting Gentrification in Houston's Low- and Moderate-Income Neighborhoods*. Local Initiatives Support Corporation. https://www.lisc.org/our-resources/resource/predicting-gentrification-houstons-low-and-moderate-income-neighborhoods

Wyly, E. K., & Hammel, D. J. (1998). Modeling the Context and Contingency of Gentrification. *Journal of Urban Affairs*, *20*(3), 303–326. https://doi.org/10.1111/j.1467-9906.1998.tb00424.x

Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology, & Human Values*, *41*(1), 118–132. JSTOR.

Zukin, S. (1987). Gentrification: Culture and Capital in the Urban Core. *Annual Review of Sociology*.