International Journal of Computer Applications (0975 – 8887)
Volume 187 – No.54, November 2025

Auditing Location-Linked Bias in Credit Scoring

Joseph Issa
Notre Dame University

Justin Issa
Arizona State University

ABSTRACT
Credit scores are the gatekeepers to various benefits, including housing options, career opportunities, insurance costs, retirement funding, low-interest rates, loan obtainment, and more. Before the development of credit scores, prejudiced loan officers conducted face-to-face applications and used factors such as income, referrals, reputation, and character judgment to determine who received a loan. Then, in the 1950s, engineers William Fair and Earl Isaac invented the credit scoring models, in hopes of removing human bias from the equation for determining someone's creditworthiness. These models, in the hands of credit companies that began collecting massive amounts of personal information from their customers, sparked public backlash over data privacy and discrimination. These public outcries eventually prompted the US government to intervene and enact laws that protected consumer rights. Specifically, the 1974 Equal Credit Opportunity Act barred credit companies and their models from using information like race, sex, marital status, religion, and national origin . Although the invention of credit scores was originally an attempt to eliminate bias, other factors can be included or excluded from these models that still put marginalized communities at a disadvantage.

Keywords
Credit scoring bias, Redlining, Ghost variables, Location-based discrimination, Algorithmic fairness & data transparency

1. INTRODUCTION
Today, FICO[1] and VantageScore[2] models use factors such as payment history, amounts owed, length of credit history, and credit mix, yet there likely exists a ghost variable in the underlying data set that powers these models, resulting in biased credit scores. The research group's theory is that a ghost variable representing the historical injustices, inequities, and blatant racism of marginalized communities in redlined neighborhoods correlates with credit scores. Although location is not explicitly listed as a factor in these models, existing research shows a clear correlation between credit scores and location. How might this occur?

In the 1960s, the term "redlining" emerged to describe the racist practice in which "the federal government and lenders literally drew a red line on a map around neighborhoods they would not invest in based on demographics alone"[5]. Because residents in redlined neighborhoods hoped to own a house but were unable to obtain a traditional loan with reasonable interest rates and fees, they were forced into predatory subprime loans designed to exploit their financial need and low incomes. These predatory lending practices targeting certain geographies drive down the area's average credit score as locals take on more debt they can't pay back. This is one possible explanation to explore further to determine why areas affected by redlining might have lower credit scores.

This calculation between 300 (poor) and 850 (excellent) critically impacts one's life, livelihood, and children for better or worse (and more often worse). With such a significant impact on real lives, there must be accountability and transparency to ensure that the data and models used to generate these scores accurately reflect a person's actual creditworthiness. It is recommended to hold these private, for-profit companies accountable by conducting an external audit, as the data used in their models likely reflects the past inequalities and amplifies them today.

2. DEFINITIONS & TERMINOLOGY
Ghost variables: data omitted from models that might introduce bias if not accounted for.

Marginalized communities: individuals or families that live in zip codes included in redlined geographic areas.

Redlining: "a discriminatory practice that puts services (financial and otherwise) out of reach for residents of certain areas based on race or ethnicity. It can be seen in the systematic denial of mortgages, insurance, loans, and other financial services based on location (and that area's default history) rather than on an individual's qualifications and creditworthiness. Notably, the policy of redlining is felt the most by residents of minority neighborhoods." 7

A. Research Question
Do the current models and algorithms used to determine a person's credit score disproportionately and negatively impact individuals in historically marginalized communities? If so, how can this technology be used to address and mitigate the effects of these biases?

Hypothesis: The current algorithms to determine a person's credit score disproportionately and negatively impact individuals in historically marginalized communities.

Null Hypothesis: The current algorithms to determine a person's credit score do not disproportionately and negatively impact individuals in historically marginalized communities.

B. Initial Evidence and Working Conclusions
Alternative (Initial Working) Hypothesis: The current algorithms used to determine a person's credit score disproportionately negatively impact individuals from historically marginalized communities, but these effects can be mitigated by manipulating and sanitizing the ghost variable.

1 *How are FICO scores calculated?* myFICO. (2021, October 27). Retrieved December 9, 2021, from https://www.myfico.com/credit-education/whats-in-your-credit-score.

2 *How it works*. VantageScore. (n.d.). Retrieved December 9, 2021, from https://vantagescore.com/lenders/why-vantagescore/how-it-works.

C. Data Availability

Based on existing and related research, it was initially anticipated that publicly available, anonymized data would be available for a regression analysis to identify causal relationships among characteristics that might impact credit scores. As it turns out, the only publicly available data is either highly aggregated (at the state or national level) or missing variables of interest, like zip code, or unavailable due to an enormous paywall.3, or available only to Federal Reserve System researchers4. As mentioned in the introduction, the aim is to investigate whether a ghost variable contributes to this disparity in credit scores among individuals in marginalized communities.

The variables that are already embedded in the existing models (i.e., payment history) were the only ones included in the single, non-aggregated data set we were able to find online5. This unexpected finding prevents us from properly auditing any data and modeling bias with any level of scrutiny. Immediate questions arose as we discovered this lack of data availability:

- How are the private corporations that are creating these models supposed to be held accountable for the scores they assign to real humans that severely impact their livelihoods?
- Why isn't an anonymized version of this data available to the public? Are they purposefully attempting to avoid scrutiny?
- How can we trust them with such an important task when we see the results of their scores furthering the oppression that exists in America?

When considering the questions above, the research group hypothesized that this data may be unavailable due to privacy-related issues or a lack of capacity to create, maintain, and share it widely. However, there is reason to argue that if these credit companies / "data providers" can decide who has access and how much they are willing to pay for it, as seen by Experian's "Premier Aggregated Credit Statistics" offering6Then they might be deliberately creating this barrier to entry. Regardless, there is clearly both a lack of transparency and accountability for the data they possess and use to power their technology.

By law, these private credit companies cannot include any demographic data in their models, but if they aren't taking an active anti-racist approach in their model building, then we are concerned that the model they have built will only be as good as the data it's built upon, which we already know is racially biased7. Without transparency into this technology that impacts real lives, there is no progress in the fight against racism in America. There must be an anti-racist approach to improve this technology to make more socially-responsible and truly reflective creditworthiness scores that purposefully account for the inequities of marginalized communities.

Thus, we firmly believe that by taking a critical look at the existing research and identifying its limitations, we can then propose an alternative study with an anti-racist lens. In proposing this alternative study, we will also identify the data needed to power it and the possible impacts and limitations it might have as well.

3. EXISTING RESEARCH

This study reviewed existing research and critiqued its approaches.

Existing research conclusions and their data availability issues:

The Wikipedia entry for the topic "Criticism of the Credit Scoring Systems in the United States" is an excellent condensation of all the studies, articles, and critiques that support the entry's overall stance that, "[r]acial discrimination ... results in impacts on the credit scores and economic security of communities of color…"

Though the effect of racial discrimination can be easily derived from the personal accounts of many individuals, these ghost variables are usually absent from the data that is actually used in the research that attempts to address the issues of disparity. The limitations in the data sets released for studies by companies and private/public organizations are usually based on aggregate values that eliminate identifying factors of the underlying research population. However, aggregating and masking data is an excellent way to assure consumer privacy; this very method becomes a gatekeeper into knowing how/when/why the poor narrative began for some of the most vulnerable individuals involved.

Data access and restrictions:

Access to credit score data at the zip code level, but the data the research team analyzed did not specifically look at correlations between credit scores and the impacts on historically redlined geographies.

These studies lead us to believe this research could be conducted:

After analyzing the data collected for this audit, the research team concluded that, while hypothetically this audit could be conducted, the data points needed for accurate analysis and assessment were not readily made available by the companies that collected the data, nor by the institutions that attempted to study it.

---

[3] *Credit score statistics: Consumer credit statistics and credit score by ZIP code: Experian*. Consumer. (n.d.). Retrieved December 9, 2021, from https://www.experian.com/consumer-information/premier-aggregated-credit-statistics.

[4] *Center for Microeconomic Data*. Center for Microeconomic Data - Consumer Credit Panel: Frequently Asked Questions - FEDERAL RESERVE BANK of NEW YORK. (n.d.). Retrieved December 9, 2021, from https://www.newyorkfed.org/microeconomics/faq.

[5] *Give me some credit*. Kaggle. (n.d.). Retrieved December 9, 2021, from https://www.kaggle.com/c/GiveMeSomeCredit/data.

[6] *Credit score statistics: Consumer credit statistics and credit score by ZIP code: Experian*. Consumer. (n.d.). Retrieved December 9, 2021, from https://www.experian.com/consumer-information/premier-aggregated-credit-statistics.

[7] Campisi, N. (2021, July 8). *From inherent racial bias to incorrect data-the problems with current credit scoring models*. Forbes. Retrieved December 9, 2021, from https://www.forbes.com/advisor/credit-cards/from-inherent-racial-bias-to-incorrect-data-the-problems-with-current-credit-scoring-models/.

Gaps or limitations in their research designs:

The Wikipedia article makes the excellent point that, "The outcomes for black Americans because of this bias are higher interest rates on home loans and auto loans; longer loan terms; increased debt collection default lawsuits, and an increase in the use of predatory lenders." The supporting article written by Sarah Ludwig, for theguardian.com states that, "People and communities of color have been disproportionately targeted for high-cost, predatory loans, intrinsically risky financial products that predictably lead to higher delinquency and default rates than non-predatory loans. As a consequence, black people and Latinos are more likely than their white counterparts to have damaged credit."

The gaps in the existing research lie in the data itself: by looking solely at credit report scores, the research overlooks the causes and outcomes of a failed credit scoring system. To really understand the inequality, one would have to go up in the cycle of credit and identify the root cause of the problem: How predatory lenders take advantage of vulnerable populations for profit *and* subsequently keep those same populations from attaining any future wealth - this is akin to financial genocide. Collecting and connecting data from when a specific consumer starts their credit journey, including the products offered to them, and analyzing how these initial options impact their credit score and overall financial future is key to understanding how vulnerable consumers are affected — and when.

4. RESEARCH CONCLUSION

Political implications and reasons:

Most publicly available data was highly sanitized, which is beneficial for individual privacy, but it also means it cannot pinpoint the exact data attributes used to train the algorithms.

When reviewing the datasets released by prior studies, it became apparent that the materials were summarized and highly aggregated; if a ghost variable was present, as suggested by the mapping of ZIP codes to historically redlined neighborhoods, it was obscured by data sanitation, leaving aggregates that had lost meaningful links to underlying demographic variables.

These studies indicate that research should be conducted with access to more granular data and the ability to link that data to the histories of individuals in redlined neighborhoods. The absence of such data introduces gaps and limitations in study design that can further marginalize affected populations by denying them clear pathways out of poor credit cycles and limiting access to financial and credit products needed to achieve financial stability.

Although the research did not address the political implications of the design, the analysis concludes that introducing new financial products and conducting deeper analyses could reduce ghost-variable effects. It became clear that policies surrounding products - to address predatory practices - and data - to provide more granular information without compromising individuals' privacy - are crucial for enhancing study design, data analytics, and study outcomes. And finally, the data is a result of biased societies and the implementation of laws and policies that help support these biases - the data is not the problem, bias is.

Ghost Variable and Underlying Supportive Use Cases

The limitations in the data sets released for studies by companies and private/public organizations are that the data is usually based on aggregate values that eliminate identifying factors of the underlying research population. Though aggregating and masking data is an excellent way to ensure consumer privacy, this very method becomes the actual gatekeeper for knowing how/when/why the poor narrative began for some of the most vulnerable individuals involved. Most importantly, the ghost variable should be assumed to represent a set of underlying data attributes rather than a single, static parameter.

The research team proposes that to eliminate bias from not just how the data is collected but also how it is analyzed and consumed by downstream Artificial Intelligence (AI) systems and data scientists, to create algorithms (i.e., a set of instructions) that focus on real-life use cases that members of marginalized communities experience. These use cases would then inform credit scoring systems in ways that allow them to weigh the importance of certain variables (e.g., age, zip code, current work status, etc.) and incorporate life events in the same manner as the current tax system does. This would allow the collection of additional data points, beyond zip code, to better understand an individual's current life state.

5. PROPOSED RESEARCH

Hypothetical research could further investigate and reduce bias in credit scoring and the underlying algorithms. The current scoring dataset includes variables such as payment history, income, amount owed, length of credit history, new credit, and credit mix, each weighted within an assumed linear-regression framework. Because correlations between credit scores and race, gender, or ZIP code could not be established with available data, the proposed ideal dataset would explicitly relate credit scores to race (and other protected characteristics) to identify and quantify potential bias in score generation.

Several studies have found that Black individuals are three times as likely as White individuals to have a FICO score below 620. This disparity raises the question of whether the mathematics or algorithms underlying credit score assignment are fair. According to Investopedia, credit scores are generated from selected variables, while other variables are listed but purportedly not used. The authors propose constructing a dataset that includes variables whose weights would be determined by a regression model (e.g., payment history, income, amounts owed, length of credit history, new credit, types of credit in use, ZIP code, race, color, religion, and number of dependents). The dataset could be assembled using methods similar to those used to manage currently public variables: lender database queries for financial attributes and questionnaires for demographic fields such as ZIP code, race, color, religion, and dependents. By adjusting existing algorithms to be racially informed alongside geographic factors, the approach would seek to account for low-income households, other location-based influences, and racial classifications.

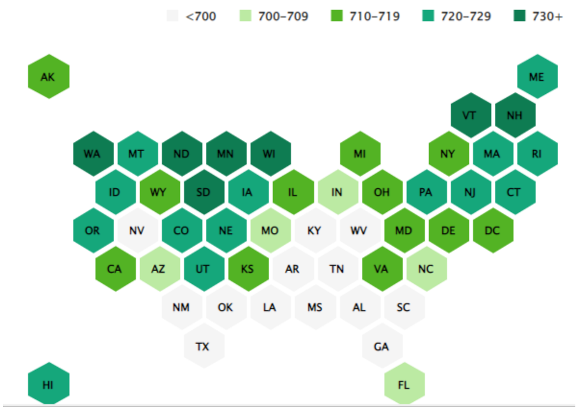
Understanding what affects credit scores is essential for individuals seeking to improve their credit scores. Based on the team's analysis, the ZIP code—often serving as a ghost variable—can influence credit scores alongside other recognized factors such as income and debt. The correlation between ZIP code and credit score can be summarized as follows:

- Areas with high minorities tend to have lower-than-average credit scores.
- Urban vs. rural areas may have little impact on credit scores.

- Low-income neighborhoods may have a higher concentration of predatory lenders.

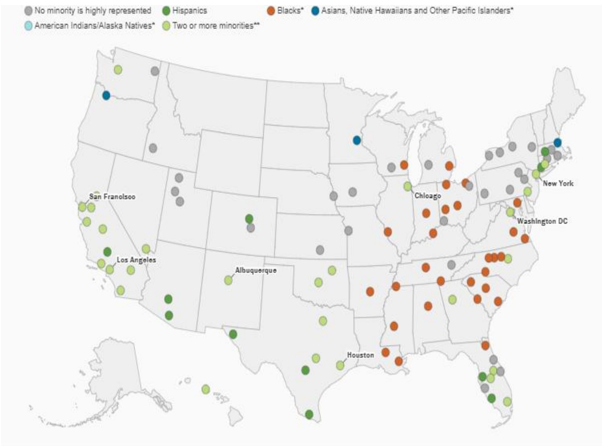
The team did not find a dataset that directly correlates ZIP codes with credit scores; however, it identified data reporting the average credit score by state and mapped these values to the concentration of minority groups within each state.

The following map illustrates the average credit score range for each state.

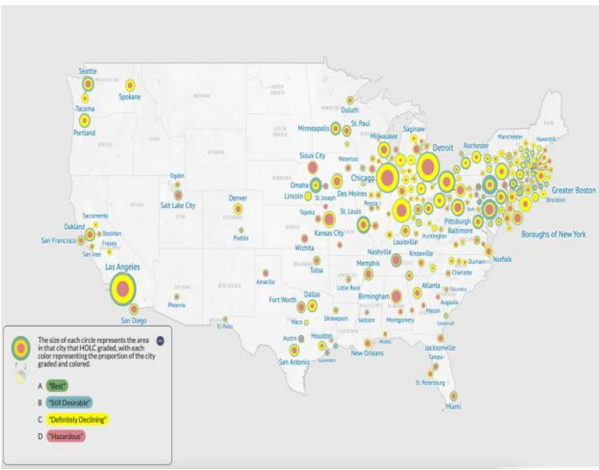


Source: https://grow.acorns.com/average-credit-score-state-map/

The following map shows that in 28 of the nation's 100 largest metro areas, two or more minority groups are highly represented.



Source: https://www.brookings.edu/research/americas-racial-diversity-in-six-maps/



Source: https://dsl.richmond.edu/panorama/redlining/

Overlapping the two maps suggests a correlation across several states between lower credit scores and the concentration of minority groups in different cities. For example, southern states such as GA, AL, SC, LA, and MS are good examples. Unzipping the zip code is key to understanding the ghost variable in determining the credit score. There is a clear relationship between credit score and location. Living in areas with high foreclosure rates is a problem, as lenders use zip codes to cancel credit cards or reduce credit limits. Several credit card issuers may use zip codes as ghost variables to profile credit card holders.

Technological Improvements

Let's think further about technological improvements. It is possible to design credit models that intentionally nullify bias and distribute credit lines to audiences with lower credit scores. Tools such as Google's 'What If' can help explain credit models for marginalized groups and pinpoint profiles with lower scores that may be due to bias. Automated systems can be used to notify audiences with low scores and guide them through a step-by-step process to improve their scores. Lending systems can be connected to the government to obtain additional lines of credit, enabling payment for government services and improving credit scores for marginalized groups. Lastly, the government should make it mandatory for credit agencies to expose anonymized data using APIs, allowing regulators and other research institutions to evaluate the effectiveness of credit scores and eliminate bias against marginalized groups.

The impact of such improvements can be profound. To begin with, introducing transparency and government oversight into credit models can enhance economic opportunities for marginalized communities. The resulting opportunities will have a cascading effect over the next few decades as more data reveal positive lending outcomes, thereby reducing bias against marginalized groups. As a result of reduced bias and increased credit, redlined neighborhoods will evolve in a positive direction with increased access to education, healthcare, public services, internet, infrastructure, and investments from private and public sectors, all of which will lead to reduced inequality that economists predict will be a massive issue for the USA in the next decade. While the positives are numerous, there are also some limitations. The state of available data used in credit models reflects the current marginalization of communities. Local and state leaders need to demonstrate a commitment and drive to regulate these credit agencies and reverse decades of marginalization.

6. CONCLUSION

Existing work linking location and credit scores broadly describes correlations and shifts responsibility to individual behavior, overlooking how historical segregation and discrimination can re-enter models through "ghost variables" (e.g., ZIP code proxies, housing and pricing signals). Framed through an anti-racist lens, these patterns indicate structural bias embedded in data generation and feature pipelines, not merely borrower choices. A rigorous contribution is to elevate transparency and accountability: document data limitations, quantify proxy effects, and replace reductive "improve your habits" advice with evidence on how place-based disadvantage shapes model outputs. This reframing positions equity as a primary model requirement alongside accuracy and calibration.

Future work should advance from diagnosis to engineered remedies. Priorities include causal and counterfactual tests to isolate location-driven disparities; scalable proxy/leakage detection (e.g., adversarial prediction, SHAP-based scans); and fairness-constrained training that bounds or removes proxy influence while monitoring performance, error parity, and calibration. Complementary interventions—responsible alternative data (e.g., verified rent/utility histories), robust checks across geographies and cycles, and privacy-preserving methods—can harden results. Public-facing model cards and data sheets, along with co-design with impacted communities and regulators, should institutionalize transparency, establish redress pathways, and translate findings into durable governance improvements.

[4] https://www.fiscaltiger.com/zip-code-affects-credit-score/

[5] Hayes, A. (2021, December 7). *What is redlining?* Investopedia. Retrieved December 9, 2021, from https://www.investopedia.com/terms/r/redlining.asp.

[6] *Fraser | Discover Economic History | St. Louis Fed*. (n.d.). Retrieved December 9, 2021, from https://fraser.stlouisfed.org/files/docs/publications/FRB/1990s/frb_111991.pdf.

[7] Hayes, A. (2021, December 7). *What is redlining?* Investopedia. Retrieved December 9, 2021, from https://www.investopedia.com/terms/r/redlining.asp.

7. REFERENCES

[1] Campisi, N. (2021, July 8). From inherent racial bias to incorrect data-the problems with current credit scoring models. Forbes. Retrieved December 9, 2021, from https://www.forbes.com/advisor/credit-cards/from-inherent-racial-bias-to-incorrect-data-the-problems-with-current-credit-scoring-models/.

[2] Center for Microeconomic Data. Center for Microeconomic Data - Consumer Credit Panel: Frequently Asked Questions - FEDERAL RESERVE BANK of NEW YORK. (n.d.). Retrieved December 9, 2021, from https://www.newyorkfed.org/microeconomics/faq.

[3] Credit score statistics: Consumer credit statistics and credit score by ZIP code: Experian. Consumer. (n.d.). Retrieved December 9, 2021, from https://www.experian.com/consumer-information/premier-aggregated-credit-statistics.

[4] Fraser | Discover Economic History | St. Louis Fed. (n.d.). Retrieved December 9, 2021, from https://fraser.stlouisfed.org/files/docs/publications/FRB/1990s/frb_111991.pdf.

[5] Frey, W. H. (2019, September 9). Six maps that reveal America's expanding racial diversity. Brookings. Retrieved December 9, 2021, from https://www.brookings.edu/research/americas-racial-diversity-in-six-maps/.

[6] Gabecortes. (2021, January 25). This map shows the average credit score in every U.S. state. Grow from Acorns + CNBC. Retrieved December 9, 2021, from https://grow.acorns.com/average-credit-score-state-map/.

[7] Give me some credit. Kaggle. (n.d.). Retrieved December 9, 2021, from https://www.kaggle.com/c/GiveMeSomeCredit/data.

[8] Hayes, A. (2021, December 7). What is redlining? Investopedia. Retrieved December 9, 2021, from https://www.investopedia.com/terms/r/redlining.asp.

[9] How are FICO scores calculated? myFICO. (2021, October 27). Retrieved December 9, 2021, from https://www.myfico.com/credit-education/whats-in-your-credit-score.

[10] How it works. VantageScore. (n.d.). Retrieved December 9, 2021, from https://vantagescore.com/lenders/why-vantagescore/how-it-works.

[11] Mapping Inequality. Digital Scholarship Lab. (n.d.). Retrieved December 9, 2021, from https://dsl.richmond.edu/panorama/redlining/#loc=5/37.8/-97.9&text=intro.

[12] Singletary, M. (2020, October 16). Perspective | credit scores are supposed to be race-neutral. that's impossible. The Washington Post. Retrieved December 9, 2021, from https://www.washingtonpost.com/business/2020/10/16/how-race-affects-your-credit-score/.

[13] Team, T. S. E. (2021, September 22). 7 major life events that affect your taxes. The Official Blog of TaxSlayer. Retrieved December 11, 2021, from https://www.taxslayer.com/blog/life-events-affect-tax-return/.

[14] What your ZIP code says about your credit score: Fiscal tiger. Fiscal Tiger | Better Information. Better Finances. Better You. (2021, December 9). Retrieved December 9, 2021, from https://www.fiscaltiger.com/zip-code-affects-credit-score/.

[15] Wikimedia Foundation. (2021, December 4). Criticism of credit scoring systems in the United States. Wikipedia. Retrieved December 11, 2021, from https://en.wikipedia.org/wiki/Criticism_of_credit_scoring_systems_in_the_United_States.