

Ethical and Data Quality Issues in the CFPB Complaints Database (2018–2025)

The Consumer Financial Protection Bureau's public Consumer Complaint Database has become a rich resource for identifying patterns of bank misbehavior and for academic research. Over the last five years, studies and analyses of this database have surfaced **ethical issues** – such as evidence of discriminatory practices and poor customer treatment – as well as **data quality and consistency issues** in how complaints are recorded. Researchers have also applied natural language processing (NLP) methods to the complaint narratives to extract insights. Below is a literature-backed review of these findings and conclusions.

Ethical Issues Revealed by CFPB Complaint Data

Racial Disparities and Discrimination: A growing body of research finds that consumers in minority communities are disproportionately affected by poor financial service quality. An analysis of CFPB complaints from 2020–2021 showed that consumers in high-minority neighborhoods file significantly more complaints than those in predominantly white areas ¹ ². In fact, the racial gap in complaint rates *widened* by over 60% during the COVID-19 pandemic ², suggesting that minority consumers bore the brunt of worsening financial service issues. However, there is evidence that **corporate culture can mitigate these disparities** – firms with more inclusive leadership and practices saw smaller racial gaps in complaints and a lower increase in complaints from minority customers during the pandemic ³. This indicates that banks with strong diversity and inclusion commitments tend to treat minority customers more equitably, reducing the need for complaints.

Relatedly, legal scholarship has begun examining specific instances of bias captured in complaints. For example, Greenlee (2025) analyzed CFPB complaints alleging racial discrimination in basic banking (checking/savings) services, documenting numerous cases of consumers of color being unfairly denied service or faced with biased treatment ⁴. These studies underscore that some banks have exhibited racist or discriminatory behavior, as revealed by patterns in the CFPB's complaint data.

"Debanking" and Poor Customer Treatment: Another ethical concern highlighted by complaint data is the practice of **improperly closing customer accounts or refusing to open accounts ("debanking")**, often without explanation. A U.S. Senate Banking Committee staff analysis in 2025 found *thousands* of such complaints in recent years ⁵. Over the 2022–2024 period, more than 8,000 consumers complained to the CFPB about banks **wrongfully closing their accounts**, and nearly 4,000 complained about being unable to open an account ⁵. Notably, the *nation's four largest banks* – JPMorgan Chase, Wells Fargo, Bank of America, and Citigroup – were responsible for over half of these debanking-related complaints ⁶. Each of these big banks had hundreds of cases where customers reported accounts shuttered with no clear cause or recourse (e.g. JPMorgan Chase led with 1,423 closure complaints and 443 account-opening denials) ⁷. Such complaints often describe **lack of communication or justification** from the bank ("no notice and no explanation was given," one Citi customer wrote) and no clear appeals process for the customer ⁸. Consumers report serious hardships resulting from sudden account closures – for instance, frozen funds for 30–60 days, which can lead to missed rent or bill payments ⁹. These patterns in the CFPB data spotlight a troubling ethical issue: **some large banks are failing to appropriately care for customers by cutting off access to banking services without due process**, causing financial distress.

Profit Motives vs. Customer Service: Complaint data also provide evidence that banks may compromise customer care in pursuit of financial targets. A recent study linked banks' earnings-management behavior to surges in customer complaints ¹⁰ ¹¹. Specifically, banks that *narrowly beat* their quarterly earnings forecasts subsequently receive about **13% more complaints** from customers in the following quarter ¹². The likely reason is that, to just meet their targets, these banks often cut costs or reduce services in ways customers notice – for example, slashing customer support or pressuring staff in ways that degrade service quality ¹³. The effect was strongest at banks where CEO pay was heavily incentive-based, indicating that aggressive profit goals can lead to **“shareholder-customer conflicts”**, i.e. practices that please shareholders but harm customers ¹⁴. In short, the CFPB complaints confirm an ethical issue: **when banks put earnings pressure above customer experience, consumers suffer**, as evidenced by spikes in complaints following quarters of intense profit-seeking behavior.

Market Discipline and Accountability: On a more positive note, the public availability of complaint data itself has started to drive changes in bank behavior. Research shows that **disclosure of CFPB complaints acts as a disciplinary mechanism on banks**. One large study matched the CFPB database with mortgage and deposit data and found that when banks accumulate more consumer complaints, they face tangible consequences ¹⁵ ¹⁶. Banks under CFPB supervision with higher complaint volumes see **declining stock prices** and abnormal trading activity, as well as **loss of market share in deposits and mortgages** (more customers pulling their money out or choosing other lenders) ¹⁵ ¹⁶. In fact, more complaints correlated with higher deposit withdrawal rates, indicating that consumers actively move their funds away from institutions with poor complaint records ¹⁶ ¹⁷. Interestingly, banks did not significantly raise interest rates to stem these outflows, suggesting the market was truly punishing them for service failings rather than them simply losing business on price ¹⁸. The **content** of complaints also matters: using textual analysis, researchers identified expressions of *customer disappointment* in complaint narratives and showed that higher aggregate “disappointment” predicted larger deposit losses for the bank ¹⁹. All together, these findings illustrate that the CFPB's complaint database shines a light on unethical or subpar behavior, and that transparency enables both consumers and investors to hold banks accountable (by reallocating their business), pressuring banks to improve services.

Moreover, banks are not only disciplined by complaint disclosures – they *learn* from them. A 2024 study in the **Journal of Accounting & Economics** found that after the CFPB began publicizing complaints, rival banks responded in ways that ultimately benefited consumers ²⁰ ²¹. In areas where a given bank had a high volume of mortgage-related complaints, competing banks saw an opportunity to win over disaffected customers. Those competitor banks subsequently **increased their mortgage approval rates** and even **opened more branches and hiring** in those markets ²⁰ ²². This effect was most pronounced for serious complaints about loan underwriting (suggesting the incumbent bank had been denying or mishandling applications) ²⁰ ²³. Essentially, banks used the CFPB data to pinpoint where peers were failing customers and then stepped in to “do better,” increasing credit access in those communities. This outcome demonstrates an ethical upside: **public complaint data spurred competitive pressure to treat customers more fairly**, as rivals strove to avoid the mistakes that drew complaints against their peers ²⁴ ²⁵.

Finally, in the realm of fair lending, transparency via complaints appears to have reduced discrimination in credit. A recent working paper found that once the CFPB began disclosing consumer complaint narratives (in 2015), **racial disparities in mortgage pricing and denials shrank significantly** ²⁶ ²⁷. Before disclosure, minority borrowers often faced higher interest rates or rejection rates even after controlling for credit risk; after disclosure, these racial gaps in rates and loan approvals diminished markedly ²⁶. The study estimates that **making complaint data public saved minority borrowers around \$102 million per year** in interest and enabled **14,000+ additional minority borrowers to get**

approved for mortgages annually (versus what would have happened without the complaint disclosure) ²⁸ ²⁷ . In other words, shining a light on consumer complaints helped root out discriminatory and unfair lending practices, forcing banks to improve equity in their lending. This exemplifies how the CFPB's database not only exposes ethical problems but can also drive reforms that **benefit marginalized consumers** in very concrete ways ²⁷ ²⁹ .

Data Quality and Consistency Challenges

While the CFPB complaint database is a valuable tool, researchers and analysts have noted several **data quality issues and inconsistencies** that must be addressed when using it. First, the CFPB itself cautions that the database is *not* a statistically representative sample of all consumer experiences ³⁰ . Submitting a complaint to the CFPB is voluntary and user-initiated, so the data may be skewed by who chooses to complain. Certain problems (or consumer groups) might be over-represented or under-represented relative to reality. For example, from 2020 onward an overwhelming share of all complaints (nearly 80%) have been about **credit reporting issues**, largely driven by disputes with the big credit bureaus ³¹ ³² . This dominance of credit-report complaints means the raw database isn't evenly distributed across products – banks' own products (loans, accounts, etc.) form a minority of total complaints. Analysts often have to filter out the credit bureau data to focus on banks' behavior ³³ . The takeaway is that the **aggregate complaint statistics require careful interpretation** – a drop in a bank's overall complaint ranking might just reflect industry-wide surges in another category (like credit reporting), rather than an actual improvement in that bank's practices.

Inconsistent Categorization and Data Entry: There is evidence that not all companies use the CFPB complaint fields in a consistent manner, which can lead to data quality problems. Complaints in the database have both structured fields (e.g. product type, issue category, company response) and unstructured text (consumer narrative). In theory, a complaint about fraud or scams should be labeled accordingly, but an industry analysis found that many complaints with narratives describing fraud were *not* tagged as such in the structured data ³⁴ ³⁵ . Specifically, an analysis by the Consumer Bankers Association (CBA) (covering 2020–2024 data) showed that **fraud and scam incidents are likely underreported in the official categories** – numerous consumers wrote in their narratives about being scammed or defrauded, yet their complaints were filed under other issue codes ³⁴ . This mismatch suggests **inconsistent classification practices**, either by consumers selecting categories or by how companies report the issues. It points to a data quality challenge: important problems might be “hidden” unless one analyzes the free-text narrative. The CBA noted that better standards for labeling fraud-related complaints could improve the accuracy of the database and help regulators and banks spot emerging scam trends ³⁴ ³⁵ .

Similarly, banks differ in how they log and respond to complaints, which affects the data. Not all banks have equally robust internal complaint tracking, so the information they feed into the CFPB system (or the outcomes recorded) may vary. Compliance experts emphasize that each bank should establish a **uniform complaint management program** with standard definitions and data fields, to ensure every complaint is documented consistently ³⁶ ³⁷ . Key details like the product, issue, complaint channel, dates, and customer segment should be recorded in standard formats across the organization ³⁷ . In practice, if one bank treats a certain customer gripe as a “complaint” but another bank's policy does not log the same scenario as a formal complaint, the data will be inconsistent across institutions. A well-defined program (including a clear definition of what counts as a complaint, and a common taxonomy for issues) is needed to **produce high-quality, comparable complaint data** ³⁶ ³⁸ . The CFPB does provide some standardized categories (as seen in the data schema for product, issue, etc.), but how companies interpret and use those fields can differ. Without internal controls, one bank might close many complaints with a vague “closed with explanation” response, while another might more often provide monetary relief – making direct comparison tricky. The **Office of Consumer Response** at CFPB

has put controls in place to monitor data accuracy and consistency, but some variability remains in the way companies respond and what they report ³⁹ ³⁶ .

Unstructured Text and Missing Data: The complaint narratives (free-form text descriptions of the issue) are a double-edged sword for data quality. On one hand, they contain rich detail; on the other, they introduce challenges in analysis. The narratives often require cleaning and NLP techniques to interpret at scale. They may include personally identifiable information, use colloquial or industry-specific language, or vary greatly in length and clarity. Additionally, not every complaint has a narrative (it's optional for consumers to consent to publish their story), so there is **missing data for many entries** ⁴⁰ ⁴¹ . A 2023 data science study noted that dealing with unstructured text and incomplete fields are key obstacles in leveraging the CFPB data ⁴² . For example, consumers sometimes leave certain fields blank or select "Other" as a category, and some companies do not provide a public response – these gaps can hinder analyses. Overall, **data quality issues like missing values, inconsistent field usage, and the need to interpret free text are known challenges** when working with the CFPB complaint database ⁴² ⁴³ . Researchers must apply extensive data preprocessing (cleaning text, handling nulls, standardizing categories) before reliable insights can be drawn.

Analysis of Complaint Text and NLP Applications

One notable development in recent years is the use of **natural language processing (NLP)** and machine learning techniques on the CFPB complaint narratives to extract trends or to automate classification. Given the huge volume of text (millions of complaints and counting), manual review is infeasible, and so scholars have turned to computational methods:

- **Topic Modeling:** Researchers have applied topic modeling to identify common themes in the complaint corpus. Early efforts using Latent Dirichlet Allocation (LDA) demonstrated that topic models can successfully surface the main issues consumers are facing ⁴⁴ . For instance, *Bastani et al. (2019)* used LDA on CFPB complaints to automatically group complaints by themes, which helped reveal the **root causes of low customer satisfaction** in different product areas ⁴⁴ ⁴⁵ . More recent work has experimented with advanced models like BERTopic (which leverages BERT sentence embeddings) to get more coherent topics from the narratives ⁴⁶ ⁴⁷ . These topic modeling approaches allow analysts to summarize the unstructured complaint text into clusters such as "billing errors," "loan modification issues," "fraudulent charges," etc., without predefined labels. The result is valuable to both regulators and banks – for example, identifying an emerging topic of "Zelle fraud scams" could prompt a bank to take proactive steps before the issue becomes widespread.
- **Sentiment and "Disappointment" Analysis:** Beyond topic themes, the tone of complaint narratives has been analyzed as well. The FDIC-affiliated study mentioned earlier introduced an approach to quantify **consumer disappointment** by processing the text of narratives ¹⁹ . Although details are scant in public summaries, this likely involved sentiment analysis or keyword analysis to gauge how severely negative the narratives were. The finding that "disappointed" language in complaints predicts customer attrition (deposit withdrawals) suggests that sentiment scoring of complaints could serve as an early warning indicator for banks ¹⁹ . In general, sentiment analysis can label each narrative as positive, neutral, or (overwhelmingly in this case) negative ⁴⁸ . Consistent highly negative sentiment around a particular product or practice can flag an ethical or customer-service problem that needs attention.

- **Predictive Modeling and Classification:** Several studies have treated the CFPB dataset as a text classification task – for example, predicting the **complaint category or outcome** from the narrative. One effort built a machine learning model (random forests) to predict whether a complaint would result in monetary relief, non-monetary relief, or just an explanation, based on the text and other features ⁴⁹. More recently, deep learning models have been deployed. Oyewola et al. (2023) introduced a *Two-Stage Residual 1D Convolutional Neural Network (TSR-1DCNN)* tailored to the CFPB complaint data ⁵⁰ ⁵¹. Their model was trained on over 555,000 complaint narratives to classify the complaints into the correct product/service categories. By combining convolutional neural networks and residual connections, they achieved about **76.5% accuracy on the test set**, outperforming baseline models like a standard 1D-CNN or LSTM recurrent network ⁵² ⁵³. This is an impressive result given the multi-class nature of the problem and the noisy text input. It demonstrates the promise of NLP in **automating complaint triage** – regulators (or the banks themselves) could use such models to quickly route complaints to the appropriate department or identify the key issue without manually reading each one. Other NLP techniques noted in the literature include using domain-specific language models (e.g. FinBERT) to better capture financial terminology in complaints ⁴⁶ ⁴⁷, and hierarchical models to account for sub-issues within complaints.

Overall, the application of NLP methods has become a “**notable exception**” in the otherwise finance-focused research on CFPB complaints – bringing computer science tools to a traditionally policy and compliance dataset. These methods enrich our understanding by uncovering patterns that aren’t readily apparent from the structured fields alone. They also help surmount the data quality issues (like inconsistent labeling) by **directly learning from the raw consumer narratives**. For example, topic modeling can highlight a problem that many consumers describe in text even if they all chose varying official categories for it. Likewise, classification models can standardize how complaints are categorized, potentially flagging when a complaint’s narrative doesn’t match the company-selected category (which itself could reveal a data entry inconsistency).

Recent Research Using the CFPB Complaint Database

In addition to the specific studies mentioned above, it’s worth noting the breadth of **academic and investigative work (2020–2025)** that has leveraged the CFPB’s complaint data:

- **Fair Lending and Corporate Responsibility:** Multiple studies (Hayes *et al.* 2021; Greenlee 2025) have used the data to examine racial fairness in banking ¹ ⁴. These works have concluded that complaint frequency is higher in communities of color, pointing to service gaps, and have explored how banks’ internal cultures or policies correlate with those gaps ². Such research informs the policy debate on how to ensure banks serve all communities equitably.
- **Consumer Outcomes and Market Discipline:** Researchers like Dou *et al.* (2024) and Jou *et al.* (2024) have studied how the **public disclosure** of complaints influences market outcomes. Their findings – from rivals increasing lending in complaint-heavy markets ²⁰ to banks with many complaints suffering deposit run-off and stock declines ¹⁵ – provide evidence that the CFPB database is an effective tool for market discipline and that transparency can drive better behavior without direct regulatory action.
- **Behavioral Responses to Regulation:** Shi *et al.* (2023) used complaint data to link banks’ internal decisions (like cost-cutting to meet earnings forecasts) with immediate customer fallout ¹². This kind of study reveals the unintended consequences of bank management strategies,

reinforcing the idea that **customer experience metrics (like complaints) should be tracked alongside financial metrics** as indicators of a bank's health and ethics.

- **NLP/Data Science Innovations:** Data scientists have treated the CFPB database as a case study for text analytics in finance. Beyond the deep learning classifier noted above ⁵², others have proposed **topic analysis dashboards** and risk monitoring systems based on the complaint narratives ⁴⁵. Such interdisciplinary research (mixing finance, NLP, and consumer protection) is on the rise, underlining the CFPB data's value beyond traditional economics or law – it's now a benchmark dataset for text-driven insights into consumer finance.

In summary, the CFPB's consumer complaint database over the past five years has shed light on several **ethical issues** in banking, from racial bias to customer-unfriendly practices, while also exposing some **data quality challenges** that need careful handling. The literature consistently finds that higher complaint volumes signal real problems – often disproportionately affecting vulnerable groups – and that banks have room to improve how they record and respond to these grievances. At the same time, the very existence of the database has empowered external stakeholders (consumers, competitors, researchers) to identify and correct those problems. Going forward, continued analysis – especially with advanced NLP – will likely further harness the CFPB data to improve transparency, accountability, and fairness in the financial industry.

Sources:

- Hayes, R.M., Jiang, F., Pan, Y., & Tang, H. (2021). *"Racial Disparities in Financial Complaints and the Role of Corporate Social Attitudes."* **SSRN Working Paper** – Evidence that minority communities file more CFPB complaints and that inclusive bank culture narrows the complaint gap ¹ ².
- Greenlee, M.B. (2025). *"Color in Line: Discrimination Against People of Color at the Deposit Window."* **North Carolina Banking Institute Journal 29(1)** – Analysis of CFPB complaints alleging racial bias in banking services ⁴.
- U.S. Senate Committee on Banking (2025). *"Investigating the Real Impacts of Debanking in America – CFPB Complaints Analysis."* **Minority Staff Memo** – Finds thousands of complaints about improper account closures, primarily at big banks, causing consumer harm ⁵ ⁶.
- Shi, X., Nguyen, D.D., & Wang, M. (2023). *"Earnings expectations and the quality of financial services."* **Journal of Accounting and Public Policy 42** – Shows a 13.3% spike in complaints after banks narrowly beat earnings forecasts, linking earnings pressure to weaker customer service ¹² ⁵⁴.
- Jou, J., Kleymenova, A., Passalacqua, A., Sándor, L., & Vijayaraghavan, R. (2024). *"Disciplining Banks through Disclosure: Evidence from CFPB Consumer Complaints."* – Research (FDIC working paper) demonstrating that banks with more complaints see stock price drops and lose deposits and loans, and using topic modeling to connect complaint content ("disappointment") with deposit outflows ¹⁵ ¹⁹.
- Dou, Y., Hung, M., She, G., & Wang, L. (2024). *"Learning from Peers: Evidence from Disclosure of Consumer Complaints."* **Journal of Accounting and Economics 77(2-3)** – Finds that competing banks increased mortgage approvals and expansion in markets where CFPB-supervised banks had many complaints, especially serious underwriting complaints ²⁰ ²⁴.

- CFPB Consumer Response Annual Reports (2020–2023) – Annual summaries of complaint trends; note that ~80% of 2020–2024 complaints were about credit reporting, dwarfing other categories ³¹ ³² .
- Mulholland, J. (2024). *"Refocusing the Regulatory Lens Using the CFPB's Complaint Database."* **CBA "Data Desk" Blog** – Industry analysis highlighting under-labeling of fraud/scam issues in complaint data and the need to improve how complaints are categorized ³⁴ .
- Warner, L. (2020). *"Analyzing Complaints – How to Understand Your Bank's Complaint Data."* **CrossCheck Compliance White Paper** – Advises banks on standardizing complaint definitions, data fields, and internal processes for consistency ³⁶ ³⁷ .
- Oyewola, D.O., Omotehinwa, T.O., & Dada, E.G. (2023). *"Consumer complaints of CFPB via two-stage residual 1D CNN."* **Data and Information Management 7(4)** – Developed a deep learning model that classifies complaint narratives by product type with ~76% accuracy, illustrating NLP applications on the CFPB data ⁵² .
- Vasudeva Raju, S. et al. (2022). *"Topic Modelling on CFPB Data: An Approach Using BERT-Based Embeddings."* **arXiv Preprint 2205.07259** – Experimented with BERTopic vs. LDA on complaint narratives, finding that FinBERT (financial domain BERT) produced more coherent topics than standard LDA ⁴⁶ ⁴⁷ .

¹ ² ³ Racial Disparities in Financial Complaints and the Role of Corporate Social Attitudes by Rachel M. Hayes, Feng Jiang, Yihui Pan, Huayi Tang :: SSRN
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3952885

⁴ "Color in Line: Discrimination Against People of Color at the Deposit W" by Mark B. Greenlee
<https://scholarship.law.unc.edu/ncbi/vol29/iss1/7/>

⁵ ⁶ ⁷ ⁸ ⁹ report
https://www.banking.senate.gov/imo/media/doc/debanking_complaints_analysis.pdf

¹⁰ ¹¹ ¹² ¹³ ¹⁴ ⁵⁴ Earnings expectations and the quality of financial services
<https://durham-repository.worktribe.com/OutputFile/1719202>

¹⁵ ¹⁶ ¹⁷ ¹⁸ ¹⁹ Disciplining Banks through Disclosure: Evidence from CFPB Consumer Complaints
<https://www.fdic.gov/system/files/2024-09/kleyменова-paper-090324.pdf>

²⁰ ²¹ ²² ²³ ²⁴ ²⁵ Learning from Peers: Evidence from Disclosure of Consumer Complaints by Yiwei Dou, Mingyi Hung, Guoman She, Lynn Linghuan Wang :: SSRN
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4506489

²⁶ ²⁷ ²⁸ ²⁹ ecgi.global
<https://www.ecgi.global/sites/default/files/Paper%3A%20Does%20the%20Disclosure%20of%20Consumer%20Complaints%20Reduce%20Racial%20Disparities%20in%20the%20Mortgage%20Lending%20Market%3F>

³⁰ ³¹ ³² ³³ ³⁴ ³⁵ The Data Desk — Complaints Part I: "I'm Not Complaining But..." Refocusing the Regulatory Lens Using the CFPB's Complaint Database - consumerbankers.com
<https://consumerbankers.com/blog/the-data-desk-complaints-part-i-im-not-complaining-but-refocusing-the-regulatory-lens-using-the-cfpbs-complaint-database/>

³⁶ ³⁷ ³⁸ ³⁹ How to Understand Your Bank's Complaint Data
<https://crosscheckcompliance.com/resources/articles/analyzing-complaints-how-to-understand-your-banks-complaint-data/>

40 41 42 43 48 50 51 52 53 (PDF) Consumer complaints of consumer financial protection bureau via two-stage residual one-dimensional convolutional neural network (TSR1DCNN)

[https://www.researchgate.net/publication/](https://www.researchgate.net/publication/372822603_Consumer_complaints_of_consumer_financial_protection_bureau_via_two-stage_residual_one-dimensional_convolutional_neural_network_TSR1DCNN)

372822603_Consumer_complaints_of_consumer_financial_protection_bureau_via_two-stage_residual_one-dimensional_convolutional_neural_network_TSR1DCNN

44 Consumer complaints of consumer financial protection bureau via ...

<https://www.sciencedirect.com/science/article/pii/S2543925123000207>

45 46 47 Paper Title (use style: paper title)

<https://arxiv.org/pdf/2205.07259>

49 [PDF] PREDICTIVE MODEL FOR CFPB CONSUMER COMPLAINTS

<https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=2995&context=etd>