

Northeastern University

Data Analysis Report

Bank Complaints Analysis

Tanishi Datta, Gianni Diarbi DS 2500 Section 02 Prof. Laney Strange

5 December 2023

Problem Statement and Background

The financial industry's trustworthiness is pivotal for consumers when choosing a bank. Our project delves into the vast pool of consumer complaints against banks, aiming to decipher trust levels through sentiment analysis. This analysis intends to unravel the nuanced customer sentiments, leading to the identification of the most trusted bank based on consumer reviews.

Understanding the impact of trust within the banking sector is multifaceted. A trusted bank not only attracts new customers but also retains existing ones, shaping long-term financial relationships. Moreover, trustworthy banks often boast enhanced brand loyalty and positive word-of-mouth, amplifying their market presence. Therefore, both parties care about the problem.

Our motivation lies in addressing the ambiguity that surrounds banking choices for consumers. Financial decisions are intricate, and a lack of transparent information often leaves individuals uncertain about where to bank. By examining consumer sentiments, we intend to provide clear insights aiding individuals in making informed banking decisions.

Introduction to the Data

The dataset used is the [Consumer Complaint Database](#), obtained from the Consumer Financial Protection Bureau. The subset of the vast database that is used in the analysis contains a comprehensive collection of grievances against banks, from October 24, 2020 to October 15, 2023. There are a total of 376 consumer complaint narratives, all of which pertain to the category “took or threatened to take negative or legal action.” The dataset's substantial size provides a robust foundation for analysis.

Ethical considerations, especially concerning financial privacy, weigh heavily in our data analysis.

Financial information is sensitive, thus it is important to maintain animosity for individuals whose data is included in the dataset. However, the dataset employs redaction practices, ensuring anonymity by

replacing sensitive details with placeholders like "XXXX." This practice safeguards individual identities and secures sensitive financial information.

Though privacy protections are in place, inherent bias may exist in the dataset. The data's bias could stem from several factors, such as underrepresented demographics or the absence of specific complaint categories. For instance, non-tech savvy customers may struggle with reporting online complaints or lack access to complaint resources altogether. Or, there could be cultural or language barriers that prevent certain customers from being able to articulate their complaints clearly. Finally, because the dataset is expansive and this analysis covers just one category of complaints, there could be an overrepresentation of specific legal-related issues that could skew the overall perception of banking issues.

Data Science Approaches

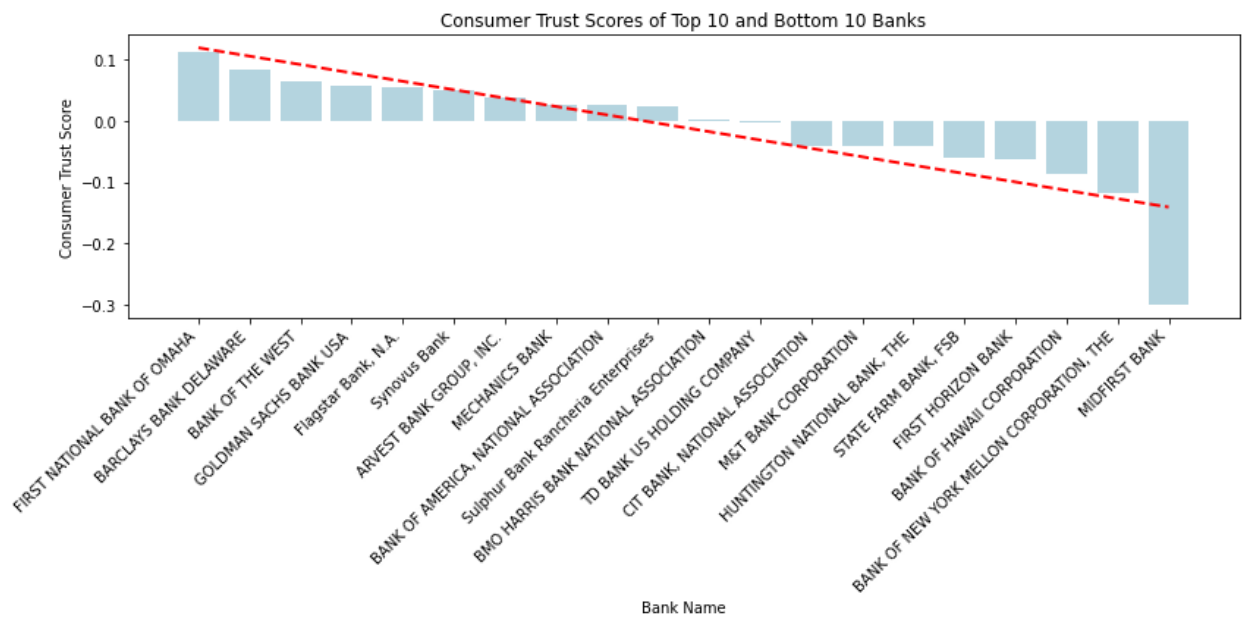
The primary data science technique, sentiment analysis, empowers us to gauge the tone of consumer complaints. This natural language processing (NLP) algorithm aids in discerning the intensity of sentiments expressed in narratives. Humans are able to understand natural language, such as nuance, sarcasm, humor, and context; the goal of sentiment analysis is to allow computers to do the same. The level of emotion contained in a complaint results in a sentiment score that is averaged per institution for comparison purposes.

Additionally, regression analysis complements our approach by revealing trends and patterns within complaint data. This particular algorithm is used to plot each financial institution's sentiment scores against each other, allowing for comparison; this has practicality for consumers deciding between several choices. In addition, it provides a very simple identification of the most and least trustworthy banks, for those with limited background knowledge or time to compare between options, and want to make a simple choice based on one factor.

Results and Conclusions

In our goal to provide actionable insights for individuals navigating complex banking decisions, our analysis revealed a vast diversity of emotions contained within consumer complaints against banks. Through this exploration, we gained a nuanced understanding of varying sentiments, allowing us to compute average sentiment scores per bank. This calculated metric became an important indicator in delegating levels of consumer trust among these financial institutions. Notably, our analysis found First National Bank of Omaha as the most trusted firm based on sentiment analysis.

This identification of First National Bank of Omaha as a trusted institution illustrates the importance of understanding consumer sentiments in shaping trust perceptions within the banking sector. It emphasizes the significance of sentiment analysis in revealing the intricacies of consumer trust dynamics. The pivotal role played by NLP algorithms can effectively capture and interpret consumer sentiments, enabling us to offer valuable insights for informed decision-making in banking.



In terms of understanding the comparison that First National Bank of Omaha has with other financial institutions, we can look at sentiment scores. In the diagram above, the 10 highest and lowest-ranked banks are represented, with a regression line that emphasizes the zero-point and relativity between the bars.

A shocking discovery is that First National Bank of Omaha performed better than other well-known banks, such as CitiBank, TD Bank, and the Bank of America. According to Adam McCann, these institutions have market shares of 5.8%, 2.9%, and 14.8%, respectively (based on total domestic deposits). Though these numbers seem quite small, it is important to remember that First National Bank of Omaha is not in the top 15 banks by market share and has a percentage much less significant. One particular explanation for this finding is that, because First National Bank of Omaha has fewer clients than competitors, there may be significantly lower volumes of complaints associated with the entity. However, one could argue that larger firms must also have higher volumes of positive reviews, which could push more popular institutions higher up the rank.

Moreover, the identification of First National Bank of Omaha as a trusted entity carries implications beyond the actual identification of a trusted bank. It highlights the importance of continuous analysis and comprehension of consumer emotions. This understanding leads to reliable and timely recommendations that align with individual banking preferences. In other words, the emphasis lies not just in the identification of a trusted entity, but in the broader lesson of continual adaptability in understanding consumer sentiments for banking decisions.

Future Work

Though our analysis provides a robust framework for banking decisions, digging deeper into the issue may present results that favor different banks. As previously mentioned, there are implicit biases in the data, and enhancing the analysis may paint a more accurate, full picture. For example, a vigorous analysis

combining demographic data with sentiment analysis could reveal trust patterns associated with diverse customer segments. It has been acknowledged that non-native English speakers may face difficulty expressing their sentiments towards the institutions in the study; thus, using demographic information in the analysis may ensure that the voice of all customers holds an equal weight. However, there are ethical concerns with this, as privacy could be compromised when looking into aspects such as age, race, and gender.

A potential exploration beyond this could be geographic trust variances, where one investigates regional variations in trust levels. Again, privacy concerns are at the forefront of the analysis. However, if successful, this could serve as a strong indicator of trusted financial firms. It could provide insights into where specific bank branches are most successful, especially for nationwide institutions such as the Bank of America. This not only benefits customers in their respective geographic locations, but also corporate offices as they decide which branches to provide funding and improvement efforts to.

Finally, a method to improve the analysis is incorporating a second dataset, specifically from a separate source, to cross-validate the information and eliminate any biases that either set may have. It is ideal for the second dataset to span a longer timeline in order to provide a holistic and comprehensive perspective of long-term trust trends in the industry. Due to the nature of the project and level of complexity associated, this had not been incorporated in our analysis.

Works Cited

“Consumer Complaint Database.” *Consumer Financial Protection Bureau. Consumer Complaint Database.*

McCann, Adam. “Bank Market Share by Deposits and Assets.” *WalletHub*, 6 Sept. 2023, wallethub.com/edu/sa/bank-market-share-by-deposits/25587.