

DSBA 6276 - Consumer Analytics
Dr. Sangkil Moon
Fall 2021

Bank of America®



Consumer Financial Complaint Disputes

Team Bank of America

Adonis Abdullah, Lenka Raslova, Jackie Mease, Akshithaa Suresh

December 10th, 2021

Table of Content

B-1. Introduction.....	2
B-2. Research Problem & Objectives.....	3
B-3. Analysis, Methods, & Results.....	3
Data Description.....	3
Competitors of Bank of America.....	4
Exploratory Data Analysis.....	5
Classification.....	6
Geographic Analysis.....	9
Text Analytics.....	12
B-4. Marketing Strategy Recommendations.....	16
Internal Policy/Process Review.....	16
Internal Feedback Loop.....	16
Invest in the Customer’s Web Experience.....	17
Pilot Customer Assistance Program For Loans.....	17
C. References.....	19
D. Appendix.....	20

Introduction

Bank of America is a multinational, financial services institution headquartered in Charlotte, North Carolina. The bank was initially founded in San Francisco in 1904, by Amadeo Peter Giannini. Since then, it was acquired by NationsBank of Charlotte in 1998, where it now serves over 66 million customers and has an annual revenue of approximately \$85.2 billion (Dolan, 2021). Today, it is the third largest investment bank in the world and part of the “Big Four” - the nation’s top four financial institutions.

Since 2015, Bank of America has been rapidly growing to keep pace with their competitors: JP Morgan Chase, Citigroup, and Wells Fargo. A 2020 SWOT analysis detailed that the company has many strengths, including a strong reputation and stable growth. However, their primary weaknesses are that there are a large number of complaints impacting the brand’s image, and they face a great threat from their competitors (*Bank of America SWOT analysis, competitors & USP*, 2020). With today’s banking landscape, in just the past decade the company has undergone a digital transformation to compete with many of their competitors’ technology-based services. Notably, in 2018 they released Erica, an Artificial Intelligence-driven chatbot - noted as the first widely-available virtual assistant in financial services. They also launched the ability to make peer-to-peer payments via Zelle, and an improved digital banking system through their website and mobile app - together encompassing over 41 million active users (*85% Of Deposits Now Made Digitally at Bank of America as Growth in Client Engagement with Digital Continues*, 2021).

Bank of America serves to make the lives, and finances, of their customers easier. Now, they should focus on their customer’s feedback in order to combat their existing weaknesses. More specifically, Bank of America needs to focus on their main weakness of customer complaints in order to provide a better customer experience and ultimately differentiate from their competitors.

Research Problems and Objectives

Customer complaints are costly for any firm. Not only does it require dollars and time to handle the complaint, but the issue the customer is experiencing can lead to financial and/or reputational harm. With the competition rapidly growing in the financial services industry and the ease with which customers can change banks, it is of the utmost importance to ensure a high quality customer experience and high client satisfaction. This risk to the firm only becomes greater when the company's response to the customer's complaint is disputed. The research objectives of our project are:

- i. Try to identify patterns or relationships in the variables in the complaints in which the customers dispute the outcome of the case using classification algorithms.
- ii. Find any patterns in customers' complaints based on the geographical location.
- iii. Understand the reasons customers complain using text analytics.
- iv. Potentially use this information to handle cases with high likelihood of dispute differently or possibly avoid the complaint altogether. Proactively make our products and services better for our clients.

Analysis Methods and Results

Data Description

Our dataset was sourced from Kaggle.com and is titled US Consumer Finance Complaints. It contains consumer complaint data provided by the Consumer Financial Protection Bureau (CFPB) for various financial institutions from 2011 to 2016 and has 555,958 records and 18 columns. 55,998 records are for Bank of America, which is the company we've chosen to analyze for this project. Table 1, located below, provides a profile of the data we used.

Table 1. Dataset Fields

<u>Date Received</u> - Date complaint was submitted (datetime)	<u>Zip Code</u> - The zip code within the state that the customer lives in (string)
<u>Product</u> - Bank product/service customer is complaining about (string)	<u>Tags</u> - Descriptions about the customer or complaint they are filing. Different across firms (string)
<u>Sub-Product</u> - Elaboration of Bank product/service (string/categorical)	<u>Consumer Consent Provided</u> - Response to whether consumer provided consent for their complaint information to be shared (string)
<u>Issue</u> - The problem a customer is having with the Product (string/categorical)	<u>Submitted Channel</u> - Means for how the complaint was submitted
<u>Sub-Issue</u> - Elaboration of Issue (string)	<u>Date Sent to Company</u> - Date complaint information was sent to the complaint aggregation company (datetime)
<u>Consumer Complaint Narrative</u> - Response customer gave describing their complaint (string/tokens)	<u>Company Response to Consumer</u> - Status of the complaint (string)
<u>Company Public Response</u> - Response from bank to customer about the complaint (string/tokens)	<u>Timely Response</u> - Did the bank respond within a 'timely' matter as defined by the company (boolean)
<u>Company</u> - The customer's bank or financial institution (string/categorical)	<u>Consumer Disputed?</u> - Did the consumer dispute the complaint (boolean)
<u>State</u> - State the customer lives in (string/categorical)	<u>Complaint ID</u> - Unique Identifier (key)

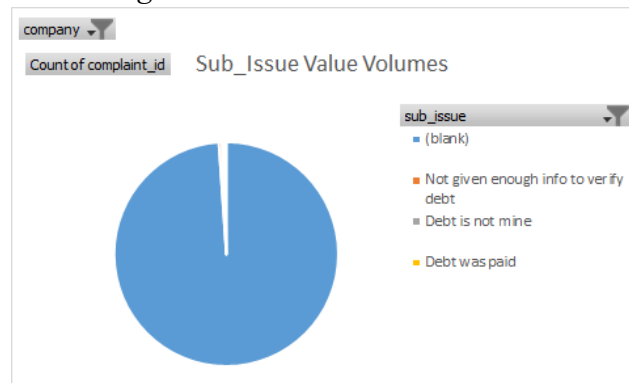
Competitors of Bank of America

The Bank of America compared to its competitors received the highest number of complaints which was 50,190 complaints. The Wells Fargo bank received 36,578 complaints, J.P. Morgan had 29,261 and CitiBank had 21,364. The customers of these banks were most dissatisfied with mortgages, bank accounts or services and credit cards. Compared to other banks, the Bank of America obtained a much higher number of complaints about mortgages than other banks. The numbers of complaints about the remaining products were comparable to the numbers of complaints of the Bank of America competitors. More details are in the Appendix in Figure 10.

Exploratory Data Analysis

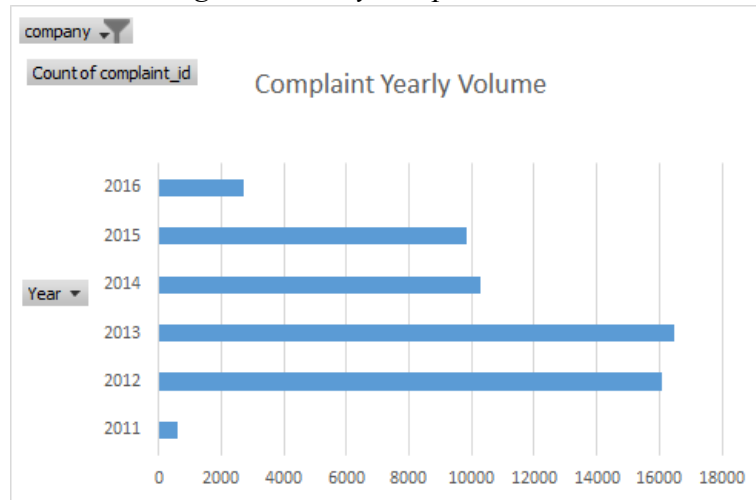
In evaluating our possible input variables, we looked at a few different things. First was to ensure the variable was sufficiently populated and not missing most or all of its values. We found that this was the case with several variables such as Sub-Issue (Figure 1), Company Public Response, and Tags.

Figure 1. Sub-Issue Value Volumes



Next, we checked the frequency of each variable's possible values to look for categories with too many or too few instances to be of value to our model. In many cases, we were able to identify just a few top values that made up most of the complaint volume such as Year of Date Received (Figure 2), Issue, Submitted Channel and Company Response To Consumer. For example, we can see that most of the complaints are between 2012 and 2015.

Figure 2. Yearly complaint volume



We suspected high correlation between the Product and the Issue variables, so we performed a cross tabulation to confirm that they are in fact correlated. Of the top five issues, each is only associated with a single product. Similarly, State and Zipcode have a high correlation. (Figure 3)

Figure 3. Complaint Volume by Issue

Complaint Volume by Issue (Top 5) and Product		
Issue	Bank account or service	Mortgage
Loan modification, collection, foreclosure		21,715
Loan servicing, payments, escrow account		6,855
Account opening, closing, or management	3,501	
Deposits and withdrawals	2,299	
Application, originator, mortgage broker		1,772

Classification

Data Preparation

First, using the data exploration that was done, we can eliminate several variables as inputs to our model, either because they were too sparsely populated, too highly correlated with another variable or their value did not have significant volumes. We can also filter our data down to the top several values for each variable and exclude outliers such as Issue = Fees which only had 2 records.

The final filtered dataset for Bank of America had 34,358 records and 5 input variables. Year, which was extracted from Date Received and filtered to 2012 - 2015, Issue, Region (derived from State), Submitted Channel and Company Response to Consumer.

Our dependent variable will be Customer Disputed?, which is a Yes or No value. As each of our input variables are categorical and are not ordinal, we must dummy code each of them for both of our classification models. We must also encode our Yes/No target variable as 1 and 0 respectively.

Analysis Methods

For our analyses we chose to run a logistic regression model and a discriminant analysis to support our findings from the regression to try to achieve our objectives. These were chosen because they are widely used and accepted for this type of research, results are easily interpreted into possible actions and can be done in numerous technologies such as SAS, R or Python. We chose to use SAS for our regression and discriminant analysis.

For our regression model, we used all of our dummy coded variables as inputs (leaving one out as a reference for each variable to avoid the dummy variable trap) and our encoded Customer Disputed? variable as the dependent variable. For our discriminant analysis, we also used the dummy variables, but did not leave out any reference variables, as well as the encoded Customer Disputed? variable as our dependent variable. It is important to note that as our dependent variable is set to 1 = Yes (Disputed), that a positive parameter estimate in the regression indicates a variable has an unwanted effect on our dependent variable in that it skews toward a dispute occurring.

Results

Details for our results can be found in the Appendix in Figures 11 and 12. For the logistic regression, we found that:

- All of the year dummy variable's estimates were negative with 2013 being the most negative and each year getting slightly less negative. This indicates that 2013 - 2015, relative to the reference variable of 2012, had a lower probability of a complaint resulting in a dispute.
 - Our discriminant analysis supports this finding as 2012 had a higher mean in the Yes(Disputed) group, while the other years all had higher means in the No(Not Disputed) group.
- Of our 4 issue dummy variables, the three Loan specific ones were the only to have a positive estimate. So customers reporting these types of complaints had the highest probability of disputing the complaint outcome relative to the other non-loan related issues.
 - We found the same 3 issues skewed towards disputing in our discriminant analysis as well.
- For our region dummy variables, West had the highest positive parameter estimate. So the West is more likely to dispute the output than the other regions.
 - The West region also had a much higher mean in the Yes(Disputed) group in the discriminant analysis. Midwest (which was our reference variable in the regression) also leaned towards Yes, but not by as large of a margin.
- Submitted_Via dummy variable for Web had the highest positive parameter estimate, whereas Referral had a significant negative estimate. Relative to Phone (reference variable), web is more likely to be disputed and referred is much less likely.
 - We found the same indications in our discriminant analysis as well with web having a much higher mean in the Yes(Disputed) group and referred having a much higher mean in the No(Not Disputed) group.

- Company_Response_To_Consumer dummy variable for Closed with Monetary Relief stood out as being a very low negative parameter estimate. It stands to reason that customers that received monetary compensation are less likely to dispute the outcome. Closed with No Relief is the most likely to result in dispute.
 - Our discriminant analysis yielded similar insights as Closed with Monetary Relief had a much higher mean in the No(Not Disputed) group and Closed with no Relief a much higher mean in the Yes(Disputed) group.

In summary, it appears as though the likelihood of a complaint being disputed has dropped since 2012, with the lowest probability occurring in 2013. Loan related issues tended to be the most disputed relative to other issues. Issues originating from the West region had a higher likelihood of disputing relative to other regions. Complaints submitted via the Web are the most likely to dispute and referred complaints are the least likely. And when the company responds to the complaint with monetary relief, the customer is far less likely to dispute. Closed with no relief at all, the most likely to dispute.

Geographic Analysis

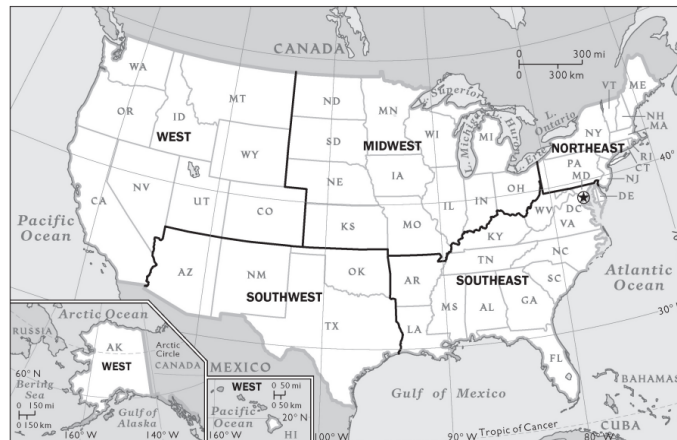
In this section, we tried to identify or find any patterns in customers' complaining based on their geographical locations as a means of supporting the findings from our logistic regression results. Given that some regions held a significant relationship with disputes, we wanted to understand if it was correlated to other variables in the dataset.

Data Preparation

We first divided the US into 5 regions - Northeast, West, Southeast, Southwest, Midwest. The allocation of individual states to regions can be seen in the figure below. Overall, Bank of America received the most complaints in the Southeast and the least in the Southwest. The exact number of complaints in regions are:

- Southwest: 4,251
- Midwest: 6,284
- Northeast: 9,235
- West: 13,321
- Southeast: 17,089

Figure 4. United States Regions

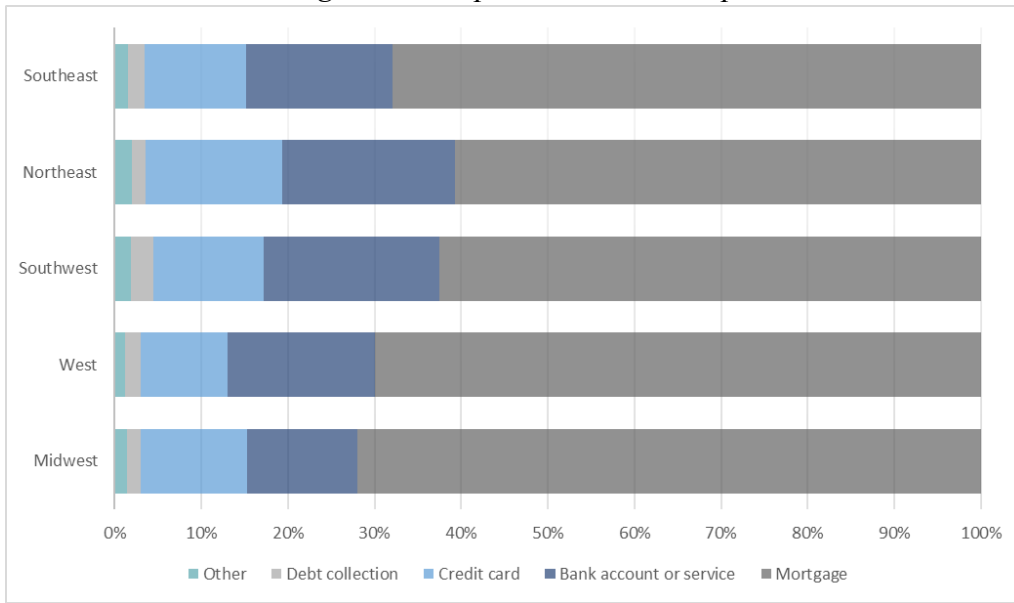


Resource: National Geographic - Education

Analysis Methods

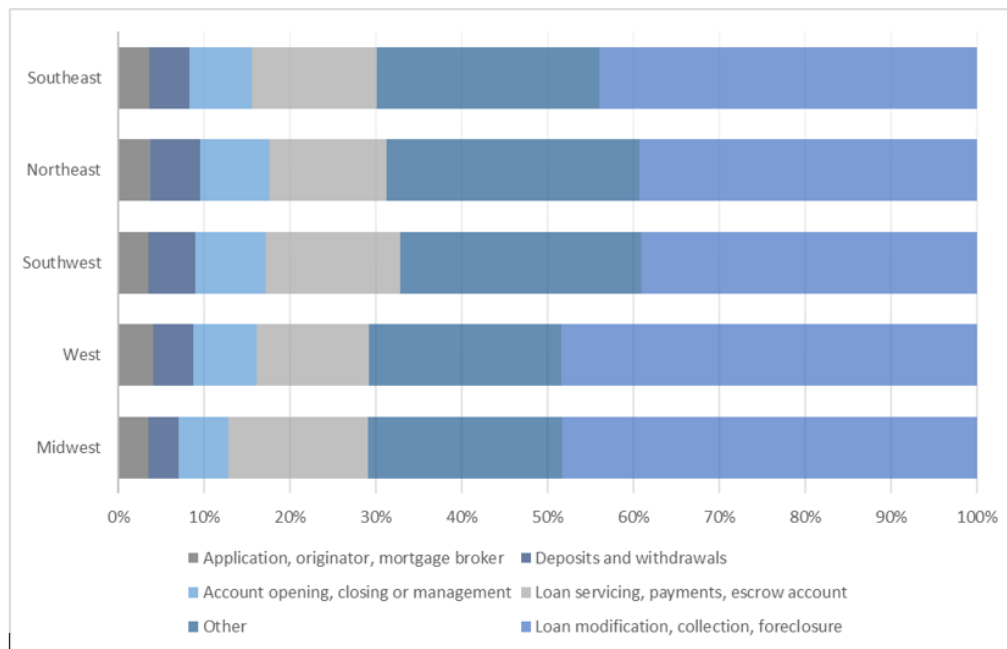
Next, we visualized the complaints in each region. The following figure (Figure 5) shows how complaints on products of Bank of America are represented in the regions as a percentage. Based on the chart, we can assume that customers had very similar behavior in the Southwest and the Midwest. Moreover, it shows that customers had the biggest problem with mortgages in all regions. In the figure the mortgage is represented by gray colour. In each region, customers mostly complained about mortgages, bank accounts or services, and credit cards.

Figure 5. Complaints on the BA's products



The following figure (Figure 6) shows how issues are represented as percentage in the individual regions. As you can see in the chart below, customers were complaining very similarly in individual regions. The customers of Bank of America had the biggest issue with loan modification, collection, foreclosure.

Figure 6. Issues in individual regions



If we look closely on the charts, we can assume that complaints on products and issues are approximately equally represented in individual regions.

Results

Bank of America received the highest number of complaints from the West and Southeast. Complaints from these two regions consist of around 60 % of all complaints. Customers complained very similarly on products of Bank of America in each region, although the volume of complaints slightly differs in each region. Customers had the biggest problems with mortgages, bank accounts or services and credit cards. Similarly, if we look closely at the percentages of complaints on specific issues in regions, we can see that customers had the biggest problems with “loan modification, collection, foreclosure”; “deposit and withdraws”; and “loan servicing, payments, escrow account” in all regions. In short, we can conclude that there are no further relationships to be explored between region and product or issue.

Text Analytics

Data Preparation

To better understand a customer’s disputed complaint, we decided to use text analytics on the consumer complaint narrative field in the dataset. We used a mixture of three methods in order to extract the most out of the information as well as using it as a means of cross validation amongst each other. First, we found the most frequent phrases and terms consumers use in their disputed complaints. To do this, we first needed to further transform the data by converting the entire complaint into small tokens for each word, remove all stop words and other common finance terms, and stem the words to control against different conjugations and/or forms of a single word. This process allowed us to remove much of the noise that comes with text analytics and focus on the meaningful components of the data.

Analysis Methods

We ultimately found these 10 phrases and topics to be the most common amongst Bank of America and their competitors.

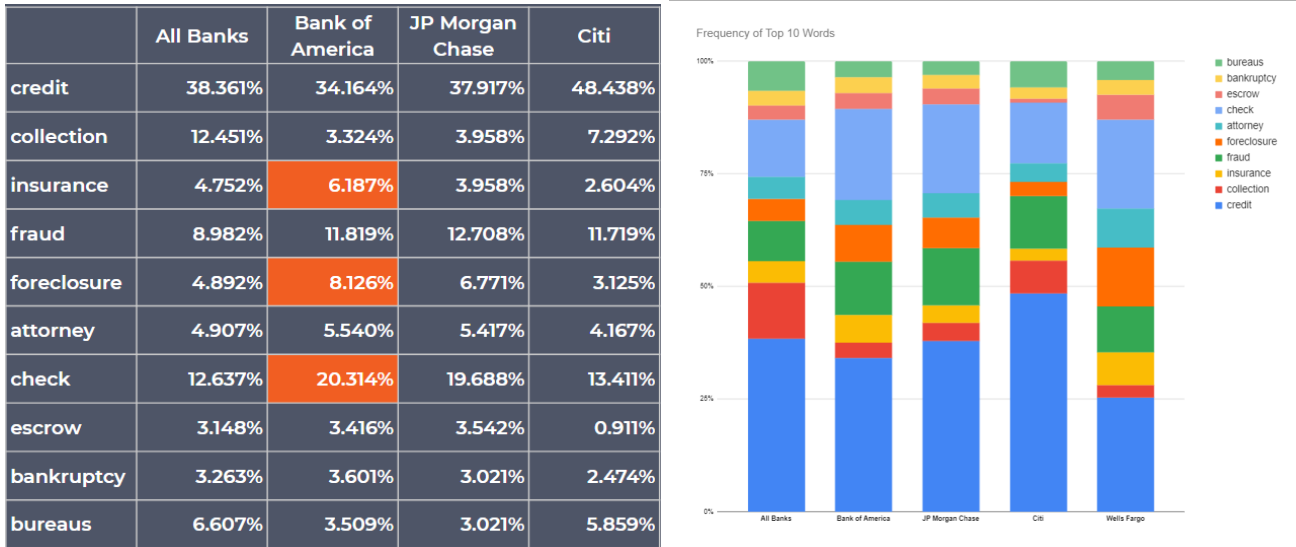
Table 2. Top phrases and topics

Top Phrases in Disputed Complaints		Top Topics in Disputed Complaints	
1	“my credit score”	1	credit
2	“received a letter”	2	collection
3	“a loan modification”	3	insurance
4	“i never received”	4	fraud
5	“they refuse to”	5	foreclosure
6	“fair credit reporting”	6	attorney
7	“a collection agency”	7	court
8	“victim of identity”	8	escrow
9	“a late fee”	9	bankruptcy
10	“account was closed”	10	bureaus

*Phrases are 3-word tokens extracted from Consumer Narrative.
Topics are single word tokens and their synonyms.*

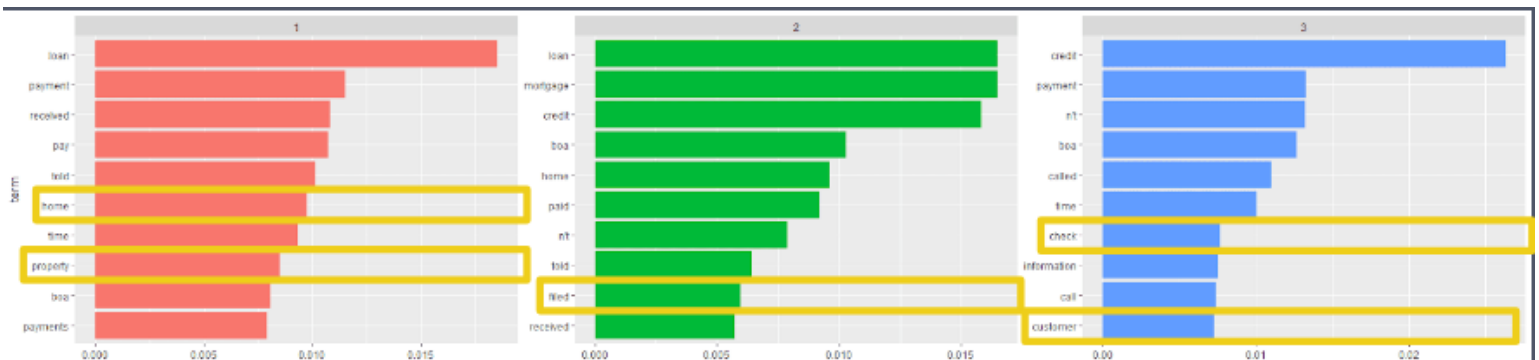
We further expanded upon these findings by looking at the distribution of these topics in Bank of America, their competitors, and the average of all banks. This allowed us to gain an understanding of if Bank of America is witnessing a comparatively higher proportion of disputed complaints in a topic.

Figure 7. Disputed complaint topic distribution



The second method we implemented was topic modeling. This differs from our previous method as this one defined its topics using the R functions tidytext and topicmodels. With their topics, we confined the number of topics to three in order to reduce overlap and have clearly defined subjects. The first topic we found was pertaining to loans and loan management. The second topic was for home mortgages. The final topic was for credit and credit cards. We took these three topics and the word frequency of each word in them and compared the frequencies between banks. The goal here was to find any words being found within Bank of America's topical top ten that weren't in the others.

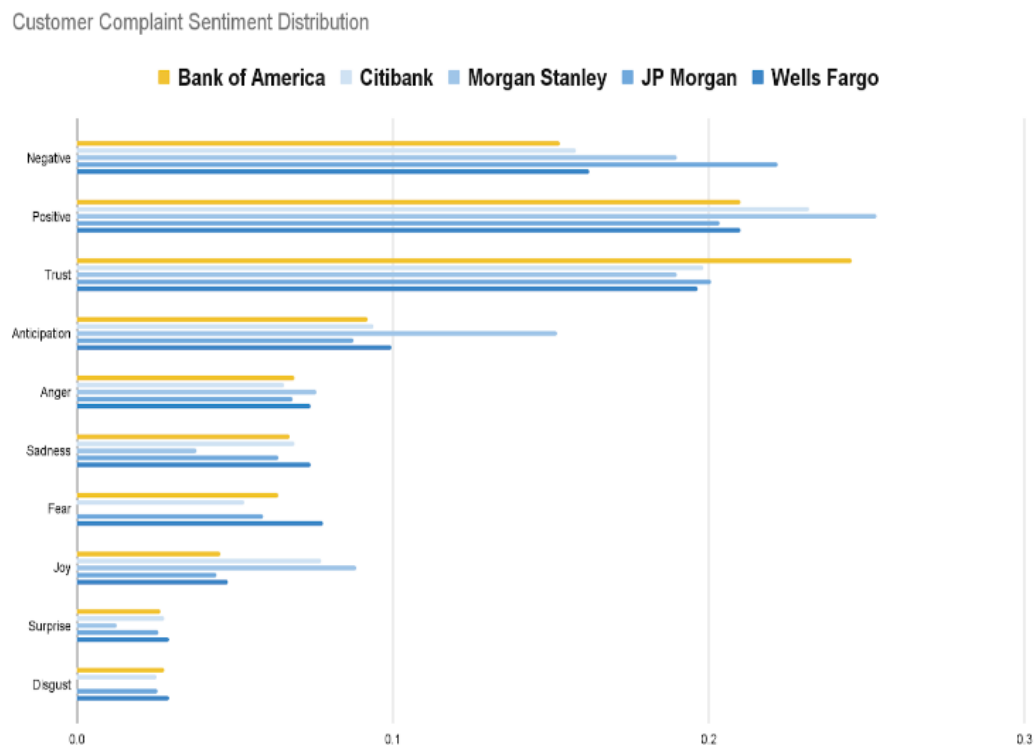
Figure 8. Unique key words



Bank of America's Top 10 within the three topics. Words boxed in Yellow are unique. (Topic 1: "home", "property") (Topic 2: "filed") (Topic 3: "check", "customer")

Finally, we decided to run a sentiment analysis on the data to better understand the type of emotions being used from customers in their complaints. Each word was given a score and association from the R packages, and we aggregated and compared those scores to gain insight into how customers feel.

Figure 9. Customer complaint sentiment distribution



Results

From our text analytics analyses, we found that there are identifiable common topics and phrases that can be found in many disputed complaints. Second, we learned that Bank of America has a higher than average proportion of disputed complaints regarding insurance, foreclosures, and checks compared to their competitors, making it a troubled area. Finally, we found that Bank of America elicits more feelings of trust in their complaints than their competitors, but their complaints also generate higher levels of disgust, fear, and sadness.

Marketing Strategy Recommendations

Based on the combined findings of our various analyses, we have four actionable recommendations for Bank of America.

Internal Policy/Process Review

Our exploratory analysis showed that the volume of complaints for Bank of America began decreasing after 2013. Our Logistic Regression and Discriminant Analysis results also revealed that 2012 was the year where complaints were most likely to result in a dispute and the following 3 years were less likely. The change in customer complaint behavior could have resulted from changes that Bank of America made to its policy, modifying its products/services, enhancing its digital experience via apps and web, or some other action taken by the bank. We recommend that Bank of America internally review changes that may have occurred during this period of time and try to identify a root cause of the trends found in our analyses. Results of this review may lead to further recommendations on specific actions that can be taken to continue the decline of disputed complaints.

Internal Feedback Loop

There were a number of patterns that appeared consistently across our analyses. For example, the West region not only complains more often, but the likelihood of the complaint outcome being disputed is the highest in that region relative to the others. Similarly, all of our analyses showed lending products, specifically mortgages, as pain points that customers are consistently experiencing. These are the most frequently complained about products with the highest likelihood of disputing the complaint outcome, and terms such as “foreclosure” and “collections” appear often in the complaint narrative. We recommend using these insights to conduct a targeted feedback loop with employees that interact directly with customers to see if they can offer any explanation as to why these things in particular are causing customers to be dissatisfied. Getting feedback from the people who are selling these products

and hearing from customers the most could be an easy and inexpensive way to glean valuable information as to what is so problematic in these areas.

Invest in the Customer's Web Experience

Our work showed that complaints submitted via the web are not only happening more than other channels, but they are the most likely to result in a customer disputing the outcome. Customers using the internet for this purpose is only likely to continue to increase over time as more and more people gain access to the web and grow comfortable interacting this way. While submitting a complaint via the web is more convenient for the client, it does not offer the personal experience they would receive with an employee, nor would the client receive immediate feedback about their complaint. We recommend Bank of America make investments in its web platform in which customers submit complaints to make it feel more like they are interacting with a live person. With innovations such as chatbots, or even live agents reaching out to the client virtually while they are on the page to attempt to resolve the complaint successfully, it may be possible with these mechanisms to resolve the customer's concern without them ultimately submitting a complaint, all while they are still conveniently interacting via the web on their own time.

Pilot Customer Assistance Program For Loans

As mentioned earlier, loan products tended to be the largest pain point for Bank of America's customers. Specifically, topics such as foreclosure, collections and loan payments showed up in all of our analyses as being the most frequent issues and the most likely to result in a complaint outcome being disputed. The cost of going through the collections process and foreclosure can be high for a lender. A 2009 report by the Joint Economic Committee estimated collections could cost the lender anywhere from 20 to 60 cents per dollar of the loan (Harms, 2017). It would be in Bank of America and the client's best interest to work to find a solution to avoid these things with assistance programs that

provide relief like lowering monthly payments, late payment forgiveness or even deferring payment during times such as job loss, death of a significant other, etcetera. Receiving a partial, late or lower payment is better than receiving no payment at all, or worse, having to go to collections and foreclosure. As this approach would be a large investment, we recommend a targeted trial of the program to mortgage customers in the West region to see if there is a measurable decrease in complaints/disputes and ultimately a decrease in the number of foreclosures and loans sent to collections.

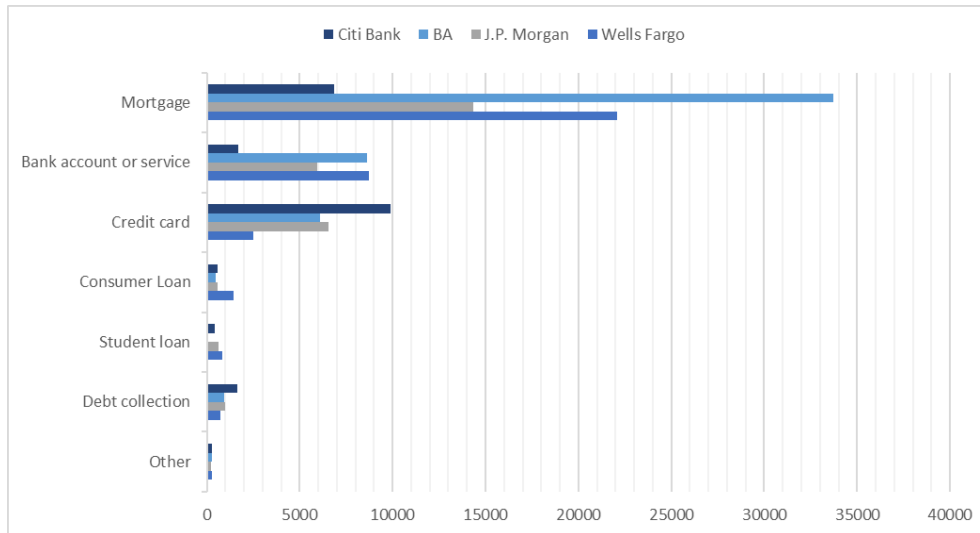
Our analyses encountered several limitations over the course of our research. Primarily, there were not many usable inputs due to not only missing values in columns that were present, but an overall lack of features in the dataset. Additionally, the data in the dataset is several years old and not recent. To continue our analysis, we would need to find additional datasets that contained not only more recent and relevant data, but also more information about the customer and the complaint to better understand who is complaining and why.

References

- “85% Of Deposits Now Made Digitally at Bank of America as Growth in Client Engagement with Digital Continues.” Bank of America, 5 Aug. 2021, <https://newsroom.bankofamerica.com/content/newsroom/press-releases/2021/08/85--of-deposits-now-made-digitally-at-bank-of-america>
- Dolan, S. (2021, July 22). Bank of America Revenue Breakdown - annual revenue by segment, region, and Year. Insider Intelligence. Retrieved December 5, 2021, from <https://www.insiderintelligence.com/insights/bank-of-america-revenue-breakdowns/>.
- MBA Skool. (2020, April 26). Bank of America SWOT analysis, competitors & USP. MBA Skool. Retrieved December 5, 2021, from <https://www.mbaskool.com/brandguide/banking-and-financial-services/597-bank-of-america.html>.
- National Geographic Society. (2012, November 9). *United States regions*. National Geographic Society. Retrieved November 16, 2021, from <https://www.nationalgeographic.org/maps/united-states-regions/>.
- Nicole Harms (2017, July 27) “The Average Cost of Foreclosure”, Retrieved December 5, 2021 from <https://pocketsense.com/time-between-foreclosure-eviction-illinois-7758021.html>
- Sharpe, Donald (2015) "Chi-Square Test is Statistically Significant: Now What?," Practical Assessment, Research, and Evaluation: Vol. 20 , Article 8. DOI: <https://doi.org/10.7275/tbfa-x148> Available at: <https://scholarworks.umass.edu/pare/vol20/iss1/8>
- Team, MBA Skool. “Bank of America SWOT Analysis, Competitors & USP.” MBA Skool, MBA Skool, 26 Apr. 2020, <https://www.mbaskool.com/brandguide/banking-and-financial-services/597-bank-of-america.html>.

Appendix

Figure 10. Bar chart of Bank of America and competitors, and the number of complaints sorted by product category



Note: Total numbers of complaints: Bank of America - 50,190, Wells Fargo - 36,578, J. P. Morgan - 29,261, CitiBank - 21,364

Figure 11. Logistic Regression Results

Logistic Regression			
<i>Original Variable</i>	<i>Dummy Variable</i>	<i>Parameter Estimate</i>	<i>p-Value</i>
Year	2013	-0.1663	<.0001
	2014	-0.1034	0.0105
	2015	-0.0927	0.0702
Issue	Loan servicing, payments, escrow account	0.3552	<.0001
	Loan modification, collection, foreclosure	0.2713	<.0001
	Deposits and withdrawals	-0.0773	0.312
	Application, originator, mortgage broker	0.3842	<.0001
Region	Northeast	-0.0957	0.0455
	Southeast	-0.0895	0.0341
	Southwest	0.00756	0.895
	West	0.0522	0.2267
Submitted_Via	Mail	-0.0427	0.5585
	Referral	-0.2899	<.0001
	Web	0.0791	0.1093
Company_Response	Closed with monetary relief	-0.7473	<.0001
	Closed with non-monetary relief	-0.1023	0.0057
	Closed with no relief	0.1289	0.0072

Note: Green indicates negative parameter estimates which have a desired impact on the dependent variable. Yellow indicates positive parameter estimates.

Figure 12. Discriminant Analysis Results

Discriminant Analysis			
<i><u>Original Variable</u></i>	<i><u>Dummy Variable</u></i>	<i><u>"No"</u> <u>Cluster</u> <u>Mean</u></i>	<i><u>"Yes"</u> <u>Cluster</u> <u>Mean</u></i>
Year	2012	0.3253417	0.362043
	2013	0.3750283	0.354167
	2014	0.2007853	0.194741
	2015	0.0988447	0.08905
Issue	Account opening, closing or management	0.1006947	0.07063
	Application, originator, mortgage broker	0.0475723	0.058943
	Deposits and withdrawals	0.0680359	0.042175
	Loan modification, collection, foreclosure	0.5978253	0.617759
	Loan servicing, payments, escrow account	0.1858718	0.210493
Region	Midwest	0.125689	0.132622
	Northeast	0.1788492	0.162729
	Southeast	0.3446727	0.32622
	Southwest	0.0820056	0.084477
	West	0.2687835	0.293953
Submitted_Via	Phone	0.0815903	0.081936
	Mail	0.050253	0.051194
	Referral	0.4289058	0.350991
	Web	0.4392509	0.515879
Company_Response	Closed with explanation	0.6896474	0.706555
	Closed with monetary relief	0.0564827	0.022993
	Closed with non-monetary relief	0.1617081	0.150788
	Closed with no relief	0.0921619	0.119665

Note: Green indicates the variable had a higher mean in the No group. Yellow indicates a higher mean in the Yes group.