Crowdsourcing Analysis of Bank Application Reviews

Class: Text Analytics
Professor: Yilu Zhou

Group Members: Annemarie Donohue,
Vanessa Asaro, Wanshan Mao, Yan Li, and Xinyi Xu

May 12, 2021

Executive Summary

Consumer and commercial mobile banking applications use cloud platforms to provide a gateway to accessible financial services for upper middle and upper income countries, like the United States. These applications provide financial services such as depositing a check, checking your balance or paying a bill, all from a mobile device. Fintech and Big Tech companies, like Facebook, Google and Amazon, have created immense competition for structured, regulated banking. Cloud digital platforms consisting of artificial intelligence are predicted for continuous, exponential growth. Although there are more regulations on banks due to the fact they have been around longer, banks need continous growth of their artificial intelligence and cloud presence to stay relevant. This analysis will provide the correlation between the reviews of the Chase mobile banking app and variables of the consumer and commercial banking sector to provide growth in net revenue for this sector.

Chase must understand the demands of their consumer in order to fulfil their needs. What attributes of these consumer and commercial banking apps benefit the consumer and producer? How can Chase Bank accelerate growth of their mobile banking to stay competitive with the fintech adoption? These are the questions this project will be focusing on.

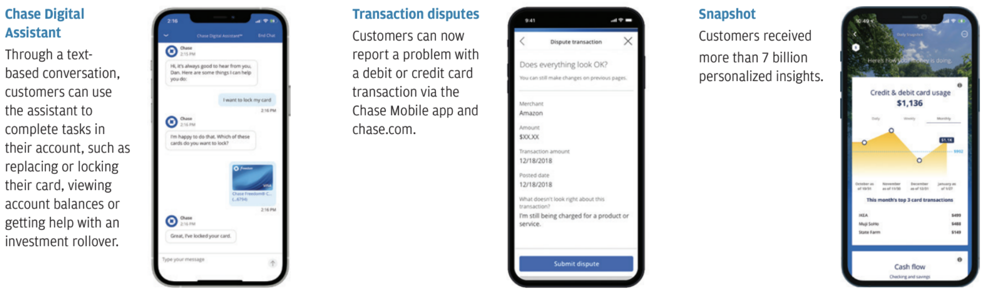
The methodology consists of data collection, data preprocessing, system building and implementation, analysis and conclusion. A random 60,000 reviews and ratings from the Apple Store for the Chase Bank Mobile Application will be scrapped for analysis from 2008-2021. Years were condensed to 2015-2021, with variables of year, month, rating and review text. The approaches of Bing Liu, Language Model, TextBlob and Vader were performed and measured through precision, recall and f measure, with Vader having the highest measures. From here, multiple methods were used to find conclusions. The most helpful ended up being

The reviews were then compared to collect data from the Chase Annual Reports, consisting of various descriptive and predictive indicators of revenue for this sector. There is a direct correlation between revenue of this sector and usage of mobile applications, therefore increasing the evidence that mobile application growth will lead to net revenue growth for Chase.

Business Goal Analysis

Mobile banking is a way for banks to compete with emerging competitors in financial technology. Chase bank is the highest visited banking portal in the US, primary bank within Chase footprint, US deposit share for retail banking and US total credit card sales outstandings and volume [1]. From launching the app in 2008, Chase had 41 million active mobile customers recorded in 2020. Competitors are also aware of the revenue from mobile banking, constantly updating their similar versions of credit cards, banking services, stock and retirement investing and money sharing [1]. Evolving competitors (Google, Facebook, Amazon, Apple) went from $0.5 trillion in 2010 to $5.6 trillion in 2020 in market capitalization, with private and public fintech companies holding $0.8 trillion in 2020 as well. US banks held $1.3 to $2.2 trillion from 2010 to 2020, which is a much slower growth than their evolving competitors [1]. As a result, banks must update apps to the likings of their consumers to outbeat competition, therefore needing textual analysis on application reviews. Growth of this sector of banking at Chase will produce net revenue.

In 2020 Chase had implemented new features to their application, which was especially helpful considering the period of branch closings and stay at home orders from Covid-19. Chase digital assistant is a chat to hold conversations about questions for their account. This is faster than calling customer service if your question is direct [1]. Transaction disputes can easily be reported via app, and Snapshot created analytic trends of their financial behaviors for financial literacy. More information on these features from the Chase Annual Report 2020 are stated below (Figure 1).



Chase Digital Assistant

Through a text-based conversation, customers can use the assistant to complete tasks in their account, such as replacing or locking their card, viewing account balances or getting help with an investment rollover.

Chase 2:15 PM: Hi, it's always good to hear from you, Dan. Here are some things I can help you do:

I want to lock my card 2:16 PM

Chase 2:16 PM: I'm happy to do that. Which of these cards do you want to lock?

Chase Freedom® C... (...6794) 2:16 PM

Chase 2:16 PM: Great, I've locked your card.

Transaction disputes

Customers can now report a problem with a debit or credit card transaction via the Chase Mobile app and chase.com.

Dispute transaction

Does everything look OK?

You can still make changes on previous pages.

Merchant: Amazon

Amount: $XX.XX

Transaction amount: 12/18/2018

Posted date: 12/18/2018

What doesn't look right about this transaction? I'm still being charged for a product or service.

Snapshot

Customers received more than 7 billion personalized insights.

Here's how your money is doing.

Credit & debit card usage $1,136

This month's top 3 card transactions

IKEA	$499
Muji SoHo	$488
State Farm	$149

Figure 1: Chase Annual Report 2020 Feature Updates [1]

Crowdsourcing has been around for hundreds of years. Professor Natalia Levina from

New York University explains in a video how the ideology can be split into “crowd of solvers”, “crowd of evaluators”, “decompose the problem”, “motivation of the crowd”, “ protect the proprietary” [2]. Crowd of solvers refers to people who have knowledge of the problem and can identify it, while the crowd of evaluators might understand how to interpret the values of solutions to issues companies have produced. An example issue would be the camera for a check deposit image producing blurry images, therefore not getting a clear enough image for check validation. The crowd of solvers would identify the issue, and the crowd of evaluators would understand why the pixelation was not as accurate. Decomposition of the issue will lead to validating if the problem is worth investing in understanding. Chase would have to determine if depositing checks via mobile application is worth the investment of paying engineers to fix this issue. Motivation of the crowd is based on money, glory and love. Since this is a banking app reviews would be based on money. Finally, protection of proprietary information would be Chase understanding that in order to get these reviews, the consumer will also know the issues of the application, which can be a deterrent. Chase would have to decide if this is worth the reward of understanding consumer issues [2].

Crowdsourcing via reviews helps the supplier to understand the issues of the consumer in order to reach a higher net revenue. Textual analysis is needed to understand the reviews in order to improve the application, especially regarding ambiguity, variability, quality and authority of text. Machine learning methods create a correct contextual understanding, no matter the tense. Even synonyms must be accurately accounted for, with a direct correlation between quality and relevance. Understanding the text within context is imperative to conceptualizing what the consumer was trying to get across via review. With the rising competition of financial technology and other shadow competitors, textual analysis will increase profitability with application growth.

Dataset Description

a. Chase Mobile Application Dataset

The uncleaned dataset consists of Chase Bank Mobile Application Reviews from the

Apple's Application Store. The variables include user, date, title, text, rating, year, month. The years scrapped were from 2015-2021, resulting in 53,668 rows. The scraped data is meant to be a sample of the 41 million mobile app users recorded in 2020.

a. **Chase Consumer and Commercial Dataset:**

This data was collected over multiple Chase Annual Reports in excel, spanning from 2015-2020. Since 2021 is the current year, the annual data is not available. Due to the reports being 300+ pages over 6 years and only a few numbers needed, the data was collected by hand for ease. Variables included percentage of average deposits growth rate, percentage of average mobile customers growth rate, number of branches, client investment assets ($ billion), consumer clients new and renewed credit capital ($ Billion), total credit capital, percentage of new and renewed credit per consumer, consumer deposits ($ billion), active mobile customers (thousands), active digital customers (thousands). There are 11 columns of variables with 6 rows of years for a total of 66 cells.

System Design

Figure 2 below shows a workflow chart of our overall system design and methodology.

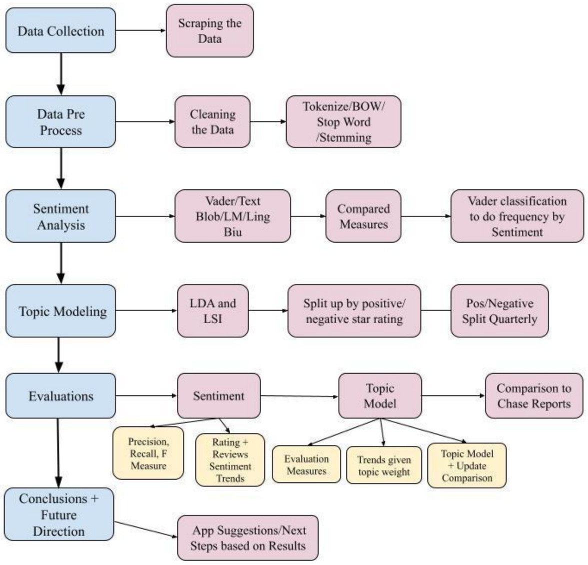

Data Collection
Scraping the Data
Data Pre Process
Cleaning the Data
Tokenize/BOW/ Stop Word /Stemming
Sentiment Analysis
Vader/Text Blob/LM/Ling Biu
Compared Measures
Vader classification to do frequency by Sentiment
Topic Modeling
LDA and LSI
Split up by positive/ negative star rating
Pos/Negative Split Quarterly
Evaluations
Sentiment
Topic Model
Comparison to Chase Reports
Precision, Recall, F Measure
Rating + Reviews Sentiment Trends
Evaluation Measures
Trends given topic weight
Topic Model + Update Comparison
Conclusions + Future Direction
App Suggestions/Next Steps based on Results

Figure 2: Workflow Chart of System Design

System implementation

a. Data Collection

Chase mobile application reviews and ratings were scrapped from the Apple Application Store of region US through python package app_store_scraper. 60,000 randomly selected feedback per person was taken as a sample of the 3.4 million ratings. The user, date, title, text and rating were imported into a csv for further analysis. Chase Annual Report data was collected for a comparison analysis against review data. The variables of average deposit growth rate (%), average mobile customers growth rate (%), number of branches, client investment asset ($ billion), consumer clients new and renewed credit capital ($ billion), total credit capital, % of new and renewed credit per consumer, consumer deposits ($ billions), active mobile customers (thousands), and active digital customers (thousands).

b. Data Preprocessing

Data preprocessing consisted of cleaning and condensing the scraped dataset for the most accurate results. The years shortened from 2015 to 2021, resulting in 53,668 rows of data. The date was broken up into two rows consisting of year and month. Quantifying the significance of each word was the first step in exploring the scraped Chase app review data. Here, basic approaches like simple bag of words, bag of words (stemming and stop word removal), part of speech for all noun forms and part of speech for proper nouns were used for the years 2015-2021. The top 30 terms were found, with top words of app, use, chase, bank and account. After getting a general idea of the text in the reviews, more advanced approaches were taken to dig deeper for conclusions.

c. Sentiment Analysis

Sentiment analysis studies the attitude, opinions or emotions in a computational manner, showing textual information as opinions over facts here. This analysis uses a combination of the ratings and reviews. Frequency by sentiment analysis per year, using Bing Liu, LM Dictionary, TextBlob and Vader approaches, were performed to compare trends of the positive vs negative reviews in a time series. In order to obtain a gold standard to evaluate the performance of the 4 approaches, the ratings of each review are divided into positive reviews (3-5 stars) and negative reviews (1-2 stars). A boolean range was used, with positive reviews as 1 and negative as 0. Standard measurement was assigned after precision, recall and F1 were utilized to compare the performance of the 4 approaches. Frequency of positive and negative sentiment by year were analyzed for targeted likes and dislikes. Chase will therefore know what to keep and change. Although the dataset we are looking at has labeled ratings, the sentiment analysis and evaluation gives us a template of what to run future unlabeled data about the mobile app through not on the rated app interface through reviews or comments made in social media posts, etc., It provides a template to run this data through to most accurately represent the results when split up and put through the topic modeling process.

d. Topic Modeling

From here, the goal was to further contextualize popular words seen in sentiment analysis through target clusters to redeem quality ideas for app improvement. The scraped reviews were compiled into a list format, and condensed to 2020-2021. Since the bank branches were closed temporarily and people have not been leaving their houses much due to the Covid-19 pandemic, these years were targeted to understand the opinions of users during the dependency surge of cloud platforms. The years were further condensed into yearly quarters for a further drill down, and a direct idea of what the consumer liked and disliked about the updates from the last 1.5 years.

A function was defined so LDA and LSI topic models could be used for different data frames. The Latent Dirichlet Allocation (LDA) is an unsupervised, generative model, known for the ability to drill down layers of aggregation. This model will make predictions of affiliation regarding word inputs per each cluster or class using the machine learning Bayes Theorem. The function is first defined by making the words lowercase and tokenizing. Dimension reduction consisted of removing all but words, discarding stop words, dropping single character words, and lemmatization. Bigrams and trigrams, or words that appear frequently together in the form of a phrase, are also added. A dictionary was created to hold all the reduced text, which included an outlier reduction as a filter on extremities, bag of word representation and token generation. With parameters, like number of words per topic, chunksize and iterations, the parameters were added to the LDA model to print the top 5 topics with 10 associated words. The average topic coherence measures the score of a single topic through measurement of semantic similarity associated with other words in that cluster. Measurements like coherence grant a quantifiable distinction between interpretable or irrelevant topics and terms. The score, or corpus, is a representative quantitative measure of the bag of words within the review. The terms were then clustered to get the final dimension reduction of 4155 unique tokens from the 53668 total documents.

The LSA, or Latent Semantic Analysis, is a natural language processing technique for distributional semantics. Correlating relationships are analyzed between document and term sets, producing a related set of concepts. TF-IDF, or term frequency- inverse document frequency, is a statistical measure combination representing the importance of a term per document within a corpus or class. Once again, the top 5 clusters with the top 10 associated terms are printed.

The function is split into different time frames per star as stated above, which the function will be assigned to for target results. The function will be applied to 1 and 2 star, or negative ratings, and 4 and 5 star, or positive ratings. There will also be a time series application, where the months of 2020-2021 are split into yearly quarters, providing an targeted analysis of reviews per time of year.

Evaluation

a. Sentiment

i. Precision, Accuracy, F MeasureRatings and Reviews: Frequency by Sentiment

Using accuracy, precision, recall and f score to measure which approach would be best, Vader was the highest. F score encompasses precision and recall, making this measure the most important. Vader had the highest average of 0.81 in f score, and the highest average in all the other 3 approaches as well. Additionally, Vader has the highest negative recall measure which is also a good evaluation, given the project concept of capturing reviews for crowdsourcing purposes, bad reviews aren't costing us anything we just want to capture the most we can to put into our model and if there are a few accidental good ratings it will not hold much weight when put through the topic model.

ii. Rating and Review Sentiment Frequency Trends

The sentiment frequency chart by year (Figure 3) consisted of actual ratings and the vader sentiment ratings using the top 30 words with POS (part of speech, noun forms) approach.

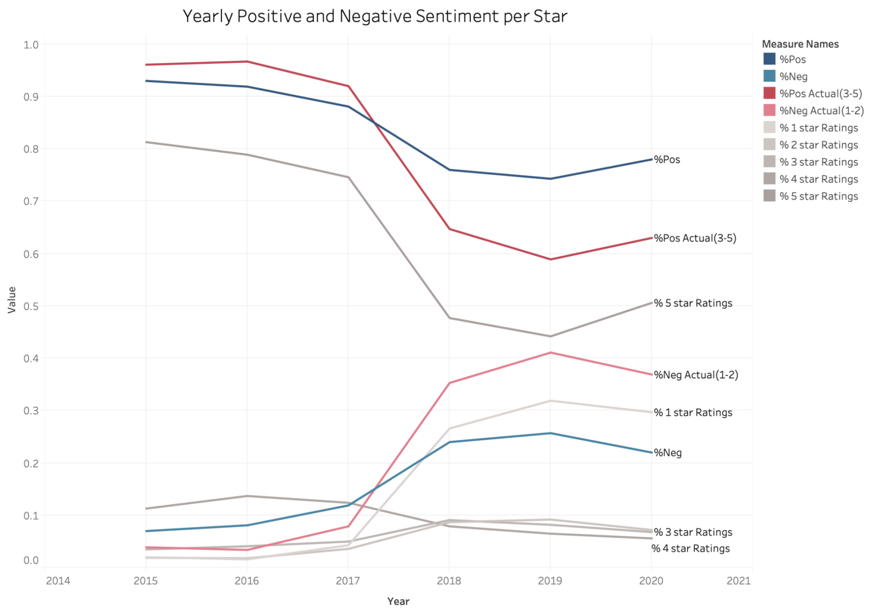


Figure 3: Time Plot of Star rating, Sentiment, and Predicted sentiment

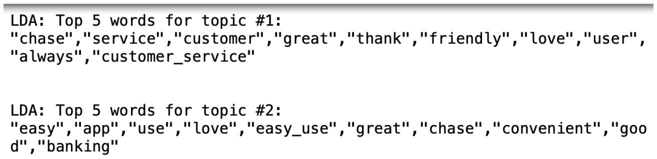
From 2015 to 2020, there has been a drop in percentage of positive and 5 ratings, with an increase in negative and 1 star ratings. The first word is three times the quantity of the second word, with app frequency of ~29,000 dropping to Chase at ~9,000. Words that are affiliated with positivity are “easy” and “love.” Word frequency peaked and declined after 2016, detailing the quantity of app reviews have decreased. The negative showed similar results, with exponential decay and the first word app being almost three times the frequency of the second word “chase.” Time was referenced twice, indicating that customers are not happy wastefully spending time. iOS was mentioned twice, meaning the updates were concerning to customers.

b. Topic Modeling

i. Evaluation Measures

______The LDA model was split into 4-5 star and 1-2 star ratings from 2020-2021. The four and five star ratings were grouped and examined as positive. The LDA had a coherence score of -2.82, and a TF-IDF overall score of 0.8. The LDA had the first cluster as "easy", "app", "use", "love", "easy_use", "chase","good", "convenient", "great", "banking.” Conclusions were about the bank app being user friendly and easy to use. Broad terminology was used, like “love” or “good.” The LSI had an overall TF-IDF score of 0.10, with the first cluster being "love", "easy", "app", "easy_use", "use", "chase", "great", "bank","best","banking." The score is the TF-IDF, which is based on a bag of words corpus frequency for LDA and TF-IDF corpus frequency for LSA. This means that the LDA had the more accurate word associations because there is a high frequency of terms in the document, therefore holding a higher weight. There is a constant highlight of simplicity, which should be a goal for Chase. Results for topic modeling outputs are shown below.

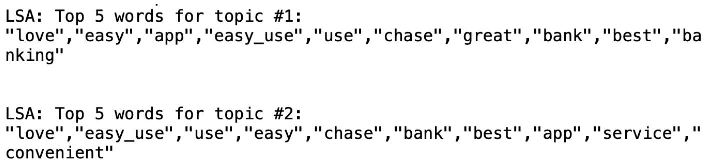
Output for 4-5 Stars LDA:



```
LDA: Top 5 words for topic #1:
"chase","service","customer","great","thank","friendly","love","user",
"always","customer_service"


LDA: Top 5 words for topic #2:
"easy","app","use","love","easy_use","great","chase","convenient","goo
d","banking"
```

LSA:

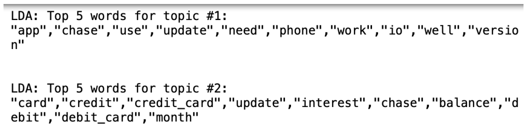


```
LSA: Top 5 words for topic #1:
"love","easy","app","easy_use","use","chase","great","bank","best","ba
nking"


LSA: Top 5 words for topic #2:
"love","easy_use","use","easy","chase","bank","best","app","service","
convenient"
```

The lower star ratings will indicate what and how to improve from the last few years of updates by putting terms into clusters of affiliated words, therefore putting an object into context. Common complaints include consumers logging in many times to retrieve any information, constant updates that correlate with the app crashing, problems with simple banking services like depositing a check, time consuming customer service, and dislike of exclusive face ID use instead of fingerprint alternatives for login. As shown in the last figure of the presentation, deposits are directly correlated to mobile application growth, along with branch numbers. The LSA model had extremely similar results, with the heaviest focus on too many updates and the service being slow. The LDA had an average topic coherence of -3.15, with the first TF-IDF topic score of 0.27 consisting of "app", "update", "time", "every", "fix", "please", "log", "every_time" ,"password", "crash." The LSA had a TF-IDF score of 0.092, having "app", "chase", "update", "account", "time", "use", "bank", "phone", "new", "work." Overall LDA had 0.5 and LSA had 0.3, This means that the LDA was more accurate once again.

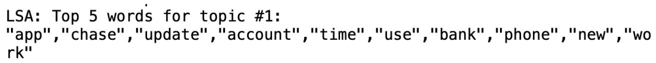
1-2 Stars LDA:



```
LDA: Top 5 words for topic #1:
"app","chase","use","update","need","phone","work","io","well","versio
n"


LDA: Top 5 words for topic #2:
"card","credit","credit_card","update","interest","chase","balance","d
ebit","debit_card","month"
```

LSA:



```
LSA: Top 5 words for topic #1:
"app","chase","update","account","time","use","bank","phone","new","wo
rk"
```

ii. Trends given Topic Weight

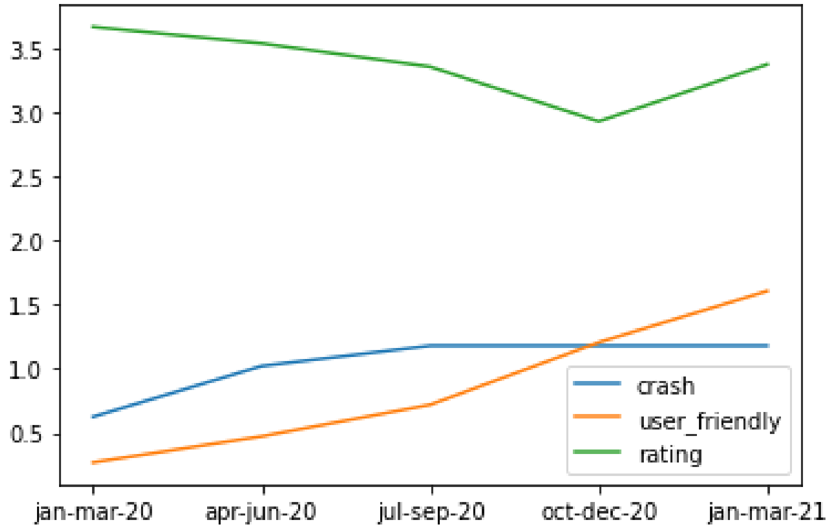


Figure 4: Topic Weight Time Plot

Figure 4 shows the changes in topic weights and ratings from January 2020 to March 2021, with the scores referring to the tf-idf. “Crash” and “user friendly” are two important and frequently brought up topics that were extracted from reviews in the time period 2020-2021. October - December 2020 is a turning point for the ratings and topics, as the positive topic weight exceeds the negative topic weight after that time. With an increase in topic user friendly and decrease in topic crash, the overall ratings grow.

iii. Topic Model and Update Comparison

Figure 5 Below outlines a timeline of Prominent Updates from 2020 - 2021, and it can be used to evaluate the context of the topic modeling results.

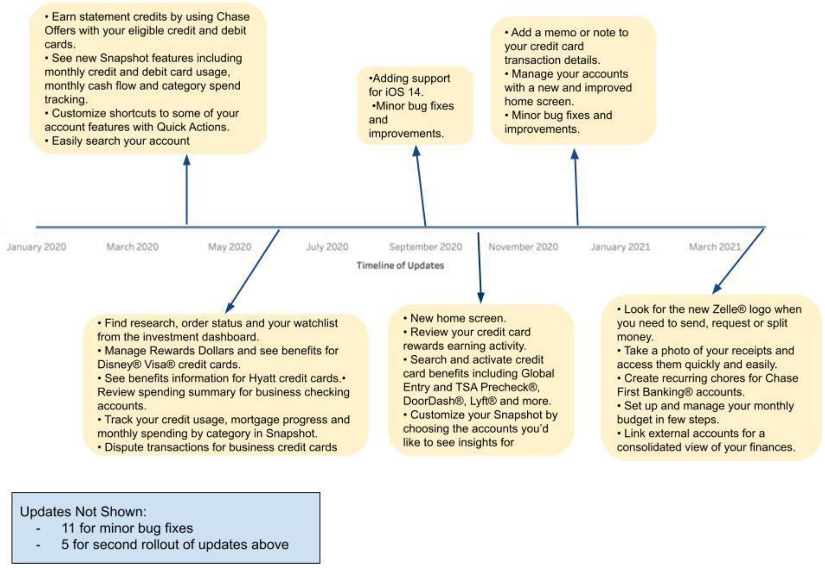

• Earn statement credits by using Chase Offers with your eligible credit and debit cards.
• See new Snapshot features including monthly credit and debit card usage, monthly cash flow and category spend tracking.
• Customize shortcuts to some of your account features with Quick Actions.
• Easily search your account
•Adding support for iOS 14.
•Minor bug fixes and improvements.
• Add a memo or note to your credit card transaction details.
• Manage your accounts with a new and improved home screen.
• Minor bug fixes and improvements.
January 2020
March 2020
May 2020
July 2020
September 2020
November 2020
January 2021
March 2021
Timeline of Updates
• Find research, order status and your watchlist from the investment dashboard.
• Manage Rewards Dollars and see benefits for Disney® Visa® credit cards.
• See benefits information for Hyatt credit cards.• Review spending summary for business checking accounts.
• Track your credit usage, mortgage progress and monthly spending by category in Snapshot.
• Dispute transactions for business credit cards
• New home screen.
• Review your credit card rewards earning activity.
• Search and activate credit card benefits including Global Entry and TSA Precheck®, DoorDash®, Lyft® and more.
• Customize your Snapshot by choosing the accounts you'd like to see insights for
• Look for the new Zelle® logo when you need to send, request or split money.
• Take a photo of your receipts and access them quickly and easily.
• Create recurring chores for Chase First Banking® accounts.
• Set up and manage your monthly budget in few steps.
• Link external accounts for a consolidated view of your finances.
Updates Not Shown:
- 11 for minor bug fixes
- 5 for second rollout of updates above

Figure 5: Timeline of Prominent Updates from 2020 - 2021

The topics we see through the topic model correlate with new ones rolled out this year. The iOS version and update came up within the overall negative reviews although they started coming up in October topic models, meaning this topic was very relevant in the reviews. This has to do with the new iOS operating system that needs to be used with the app and people with older phones not being able to get this system, essentially rendering the app useless to a large customer base. There is also a crash mentioned in conjunction with the update, meaning the minor fixes and bugs happening with the updates are still not solving the problems. In October, an update included a new home screen and the negative topic model for october included the word 'screen' and 'update' clearly expressing dissatisfaction with the new homescreen update. Additionally, in Chases 2020 report [1] it is mentioned that there is a new digital assistant that was introduced. In the overall negative rating topic model and within every 3 months, bad customer service comes up as a prominent feature.

Within all the topic models, the type of activity recorded does not have to do with these new features like rewards dollars, sending money with Zelle, or Snapshot. Most of the topics have to do with "deposit", "transaction", "balance". This is reflective of what people actually do and care about when they are using a mobile banking app. They are doing simple checking balances, making deposits, transactions. However, these are not things that are being mentioned in any of the updates. Instead, these updates are actually adding noise to the app and having to update before being able to use the app restricts the quick nature aim that a mobile banking app promises.

c. Comparison to Chase Annual Reports

The Chase Annual Reports data was used to uncover trends visually through Tableau. Figure 6 shows the yearly growth rate percentages of the Chase Consumer and Community Bank sector. The number of branches are in a size detail, with percentage of growth rate variables and their correlating trend line in color. The new and renewed consumer credit stayed stagnant, but the deposits and mobile customers declined at a very similar rate. This shows correlation between mobile customers, deposits and branch growth over time.

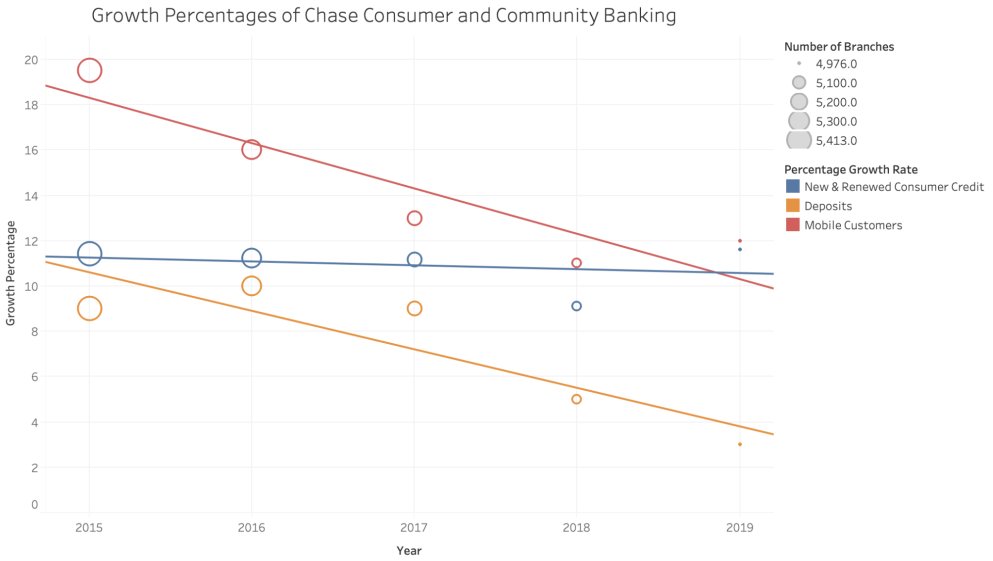


Figure 6: Growth Percentages of Chase Consumer and Community Banking

Another time series was constructed based on year over year total growth in Figure 7. This graph has consumer new and renewed credit per capita, client investment assets and consumer deposits in a stacked area bar chart. Growth is shown overtime, with 2020 being the highest year. There is a consistent variance between the variables year over year, showing the correlation between them.

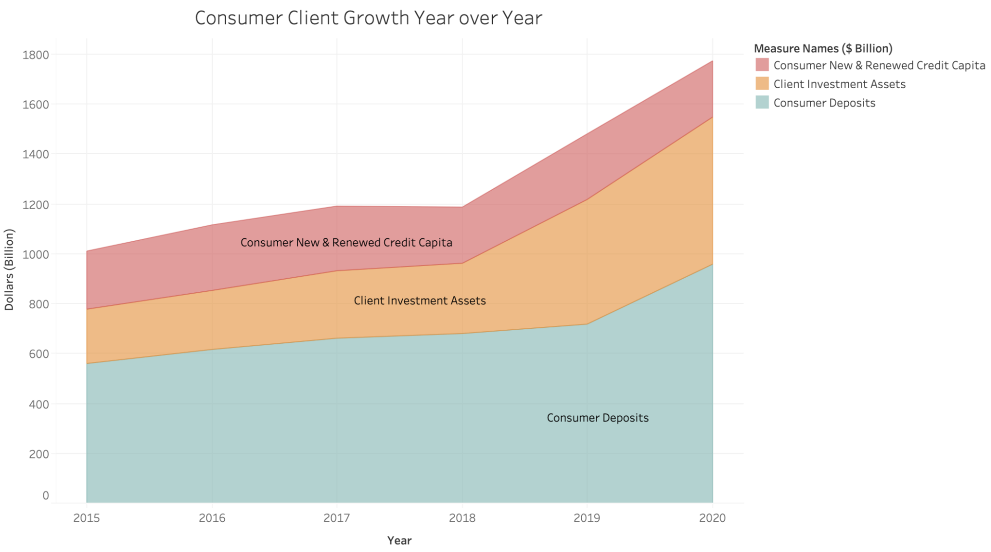


Figure 7: Consumer Client Growth Time Plot

Conclusion and Future Direction

Chase has continued to try to stay relevant as financial technology companies have quadrupled growth in the last decade. The results from this analysis conclude that mobile banking directly correlates with growth of the consumer and commercial banking sector of Chase Bank. The result concludes to Chase lacking is communication with their customers. Informative update descriptions would ease the frustration of the users who are annoyed with updates, not knowing what is even being altered. An example of this is the chase digital assistant. The chase digital assistant answers more direct, less complex questions so people need to know what to do if their question cannot be answered by the digital assistant. If the audience knows that Chase is trying to improve common issues instead of ambiguous bug fix descriptions or trying to add more features that the typical mobile banking app consumer will not be using, the customer will feel heard.

From the negative review topic model, while Chase is using the new iOS to update their app, customers with older phones and iOS need to have access. People are not going to pay for a new phone in order to access iOS updates just for this app. Instead there could be optional updates, but can still access the app and go on your phone quickly without updating (which is the point of mobile banking). This could also help with people who still want to use the fingerprint feature and do not have the face id available to them. The fingerprint login is a faster way to access your account, not having to remember passcodes. Although face ID is a similar idea, older iPhones do not have this feature, therefore making them type in their password for every log in. While there are more and more features with every update, “transactions” and “deposits” come up alot in the topic modeling, seeming to be what people care about and access their mobile banking app to look at. The new distracting home screen and the many options can annoy customers whose mobile banking experience is meant to be a quick and easy user friendly one. A suggestion to this is an option in settings for add ons and a customizable homepage where the user picks what is on it. Additionally, customers are dissatisfied with problems in the app and then not getting adequate customer service as a result. Instead of a digital assistant, there should be a real person in the mobile chat. Another option is having a customer service team who deals with the app and specializes in problems pertaining to mobile banking - especially since mobile banking is taking over physical banks and many times there can be glitches causing churches and fees that normally wouldn’t happen at a physical branch. To gather more feedback from customer text specifically pertaining to the app and its features, Chase could use the sentiment classification on the digital assistant feature and then put that through the topic model. This way they dont have to ask for a rating, which is many times not filled out after the chat. This could additionally help with their customer service negative reviews on the app store and they can gather what the issues are about and how to better respond to that.

The Chase Annual Report analysis depicts a correlation between deposits, mobile customers and number of branches. This information is vital for Chase, as the mobile platform declining is correlated with revenue falling in this sector. Consumer new and renewed credit per capita, client investment assets and customer deposits all increased year-over-year with a similar variance, meaning these variables are highly correlated as well.

With these results comes limits, possibly skewing the conclusions. The reviews scraped from the Application Store only shows the feedback from customers using Apple products, which might be limited. Another study could include Google and Samsung phones, with the ratings and reviews on their applications. These reviews would have to be segmented by phone, as each phone will have different complications. There is also the idea that reviews are only submitted by people who have strong opinions, either being negative or positive. This relates back to the ideas in crowdsourcing, where people only act for money, power or glory. Part would be money, considering this is a mobile banking application, but part could stem from emotions (glory). Otherwise, Chase should focus on consolidating updates, including the option for face ID, and other improvements listed in the findings of this project. Chase should continue to listen to their customers instead of continuing to add pieces to the app that aren't needed, like multiple updates that are imperative to download in order to get into the app. With improvements, and more importantly attention to the feedback of their customers, Chase can continue to grow in the cloud platform space to compete with other growth competitors using crowdsourcing analysis.

References

[1] Chase Annual Report 2020:
https://www.jpmorganchase.com/content/dam/jpmc/jpmorgan-chase-and-co/investor-relations/documents/annualreport-2020.pdf

[2] Crowdsourcing:
https://www.stern.nyu.edu/experience-stern/faculty-research/identifying-crowdsourcing-sweet-spot