Bigram Collocations and Latent Dirichlet Allocation to Extract Topic and Substance from Social Media Messages

Paul Glenn
Vijay Velagapudi

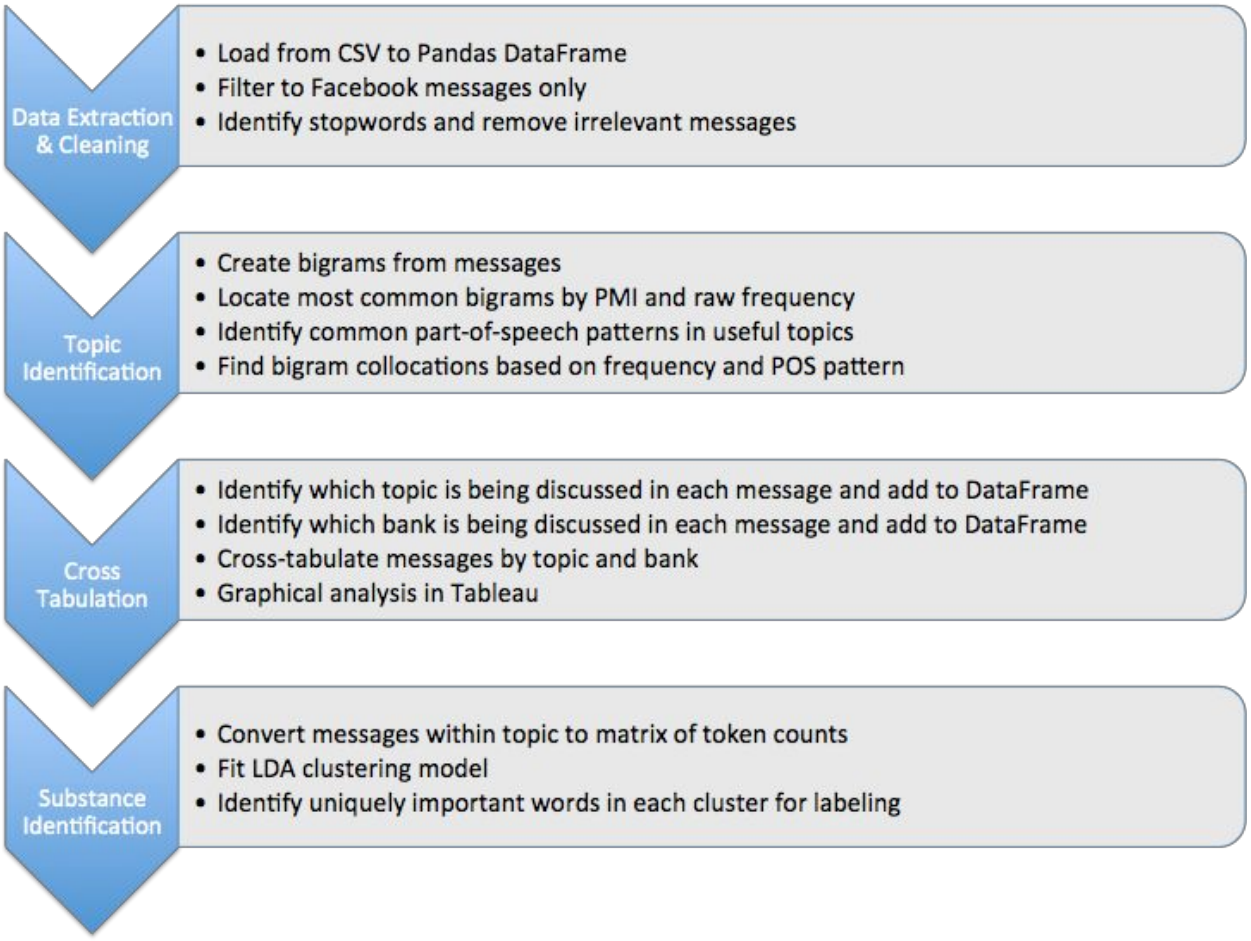
Masters Candidates
UC Berkeley School of Information

Presented to Wells Fargo Campus Analytics Challenge

What financial topics do consumers discuss on social media and what caused the consumers to post about this topic?

Approach and Methodology

Figure 1.1 Visual representation of our approach to the problem



Our approach to the problem was a four-step process:

1. Clean data by removing irrelevant messages and common words resulting from data preprocessing (i.e. NAME, ADDRESS)
2. Identify main topics being discussed with bigram collocations
3. Cross-tabulate messages by topic and bank
4. Use Latent Dirichlet Allocation (LDA) clustering to further separate and identify substance of messages

We opted for this analysis to only focus on Facebook messages for several reasons:

1. Facebook messages gave us a longer timeframe to analyze (one year, versus one month of Twitter messages)
2. There is no character limit on Facebook messages
3. Twitter messages did not have any user information associated with them, which would have allowed us to stitch together multiple Tweets constituting one message

Our preliminary analysis of the messages found that a significant portion of the messages were politically motivated. We analyzed the most frequently occurring tokens in the messages, and were able to remove a significant portion of these messages by excluding anything containing one or more of the following words:

```
{'â', 'giannis', 'banksters', 'classwarfare',
financialterrorists', 'morganstanley', 'vote',
'banke', 'BankE', 'bankE'}
```

We also removed messages regarding BankE. In total, this action reduced the size of the overall dataset by about 17%.

Next we used a bigram collocation analysis to locate the main topics being discussed. We experimented with using a Pointwise Mutual Information (PMI) approach to find the most unique bigrams in the data, but since our goal was to find broad topics, a high-frequency-based approach was more effective.

Among the list of most frequently occurring bigrams, we found those that looked of particular interest for this analysis followed the part of speech pattern adjective-noun, noun-noun and noun-verb. We tagged all of the bigrams by part of speech and extracted collocations based on this pattern to create our topic list.

Once we extracted the list of topics for study, we added a column for each to our Pandas DataFrame which indicated whether a given message discussed that topic. If it did, the value was the content of the message with stopwords removed; if not, it was False. This allowed for topic-level analysis and allowed us to cross-tabulate this information with information about which bank was being discussed in each message. It follows that we were able to explore whether certain topics were isolated to certain banks or if they were applicable to entire industry.

While we were able to identify topics and trends within the dataset, we still faced the far greater challenge of discovering the substance of an individual message -- how a message's content related to the topic. For this, we attempted two approaches: LDA, a clustering algorithm that models latent topics in the messages, and chunking, an NLP technique to extract information by using common patterns of parts of speech in a sentence to show noun and verb phrases. Both approaches proved challenging.

With LDA, the algorithm often struggled with different senses of the same word. For example, the word “check” as a verb generally showed up in the credit card topic in messages about fraud (“check your statements”), but questions about making payments with the noun check (“can I pay by check?”) would fall into the same cluster.

Chunking could distinguish between these senses, but did not have the ability to easily combine similar chunks. Furthermore, chunking rules that we identified as being particularly useful for analyzing one topic created noise for other topics. With more time for analysis, that is an area we would explore further.

Social Media Drivers

Exploring the provided dataset, it was clear that there were some common motivations behind the messages posted on Facebook. There were numerous messages that were focused on fundamental beliefs or were politically motivated, such as messages that included terms like “financial terrorism” and “class warfare”. Other messages were more focused on particular problems or issues that the customers were facing, such as those with ATM problems or rude customer service. Finally, there were many messages that were purely informational, such as those reporting or responding to a news story. The focus of this analysis was to identify topics that could potentially be actionable. Therefore, we focused our efforts, and excluded messages that were purely informational or were politically motivated.

After filtering out the messages that were clearly not useful, and further reducing the set to only those that specifically mentioned one of the topics we uncovered, we still struggled with separating messages that were merely blowing off steam from those that expressed an actionable complaint. Put differently, our methodology easily excludes “BankB is theeeeee worst”, but struggles with filtering out “The customer service at BankB is theeeeee worst,” even though the latter contains only the smallest amount of marginal information.

Our substance analysis clustering algorithm does take the first steps towards dealing with this problem; some clusters are clearly less relevant than others. For example, within the customer_service topic, the cluster defined by: ['guys', 'say', 'sucks', 'hate', 'want', 'im', 'suck'] is obviously less useful than the one defined by: ['closing', 'month', 'reps', 'pay', 'tell', 'think', 'refund', 'error', 'trying'] (see *figure 2.3)*, but this insight requires a level of human intervention.

Code

Please refer to attached IPython Notebook or view this file online at this private gist http://nbviewer.ipython.org/gist/pdglenn/c34fe65b64c7bb2eda1e for our documented code. The section headings within the document closely follow the steps identified in our visual workflow in *figure 1.1*.

Are the topics and “substance” consistent across the industry or are they isolated to individual banks?

Topics

We initially produced the following list of topics from bigram collocations. However, after exploring each of the following topics in closer detail, we discovered that some topics were more insightful than others. For example, “Chicago Marathon” was a topic that was closely related to one of the banks, but not the others. Other bigrams such as “Gon Na” and “Wan Na” were mostly a result of non-English comments. While it is possible to refine the following list further by removing some of these occurrences using some additional NLP tools, we opted to revise the list manually by exploring the comments for each topic. We were also able to remove certain topics from our list by visualizing the topics in Tableau.

Figure 2.1 Initial List of Topics

Financial Advisers	Goldman Sachs	Small Business
Wealth Managers	Data Breach	Close Account
Chicago Marathon	Overweight Rating	Account Bank
Customer Service	Gon Na	Financial Crisis
Bank Account	Good Morning	New Bank
Debit Card	New Photos	Real Estate
Credit Card	Marathon Chicago	Money Account
Neutral Rating	Bank Robbery	New Photo
Checking Account	Wan Na	Tickets See
		Worst Bank

The following table shows the list of topics that we identified as being particularly insightful.

Figure 2.2 Revised List of Topics

Financial Advisers	Credit Card	Real Estate
Wealth Managers	Checking Account	Close Account
Customer Service	Data Breach	Account Bank
Bank Account	Bank Robbery	Financial Crisis
Debit Card	Small Business	New Bank

We used the revised list of topics to perform Latent Dirichlet Allocation (LDA) to cluster the comments to identify the substance within each topic. With LDA, we had to fix the number of clusters that we wanted to find within each topic. After trying various numbers of clusters, we opted for 10 clusters, which produced the best separation for each of the topics. With additional time for analysis, it may be more beneficial to change the number of clusters for each topic, perhaps on an individual basis for each topic.

The following table summarizes the substance topics that we identified using this process. A full list of the substance topics is available in the appendix.

Figure 2.3 Substance identified for customer service topic

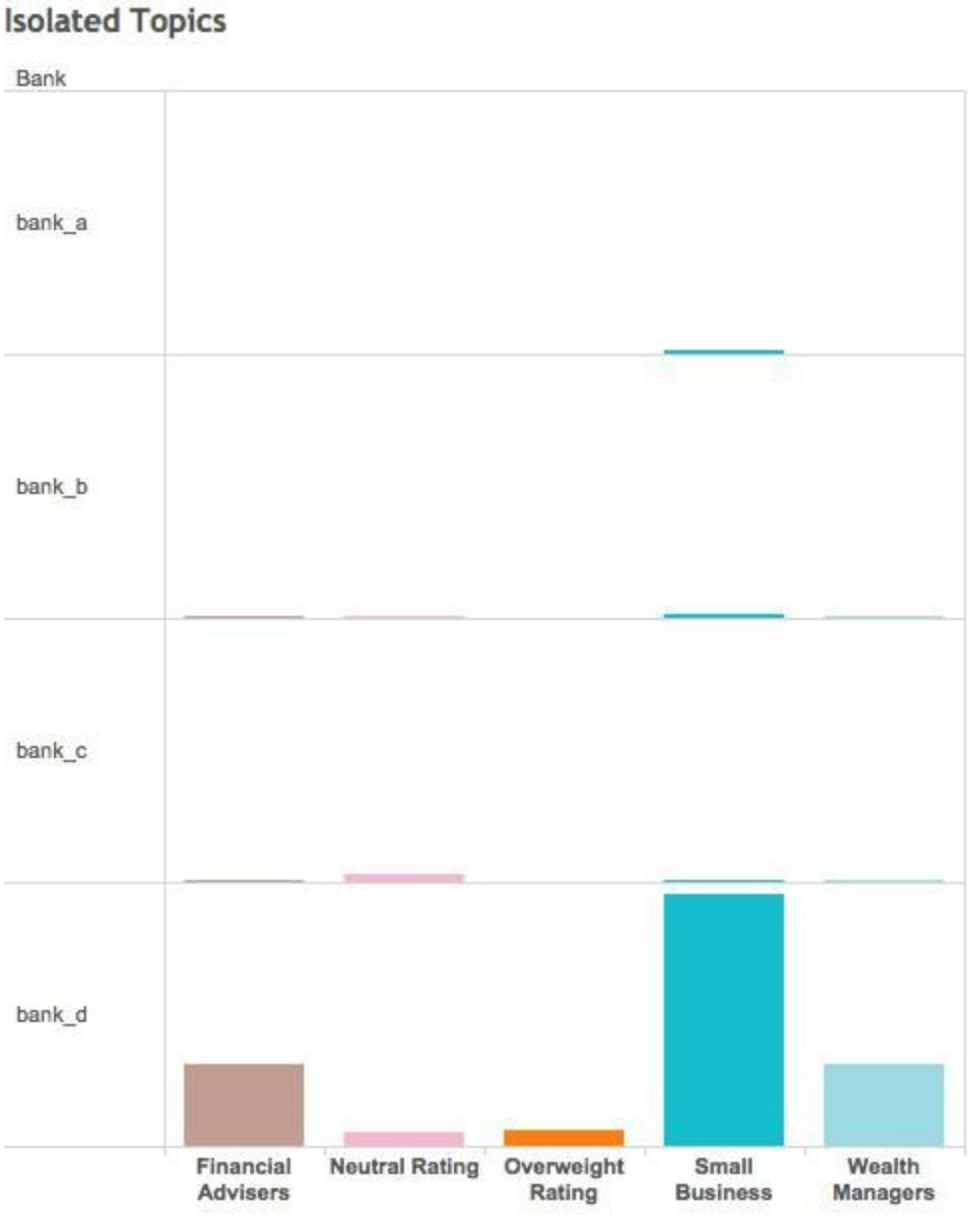
Topic	#	Substance
customer_service	0	['card', 'debit', 'great', 'charges', 'use', 'fraud', 'cards', 'working']
Example Message: 'BankD bank eliminates more jobs. u ask hoe. automated tellees.. instead of customer service with a real person. we deal with more machines that cant answere our questions. another example of how society is becoming more impersonal with each other. society is loosing our basis of human connection. so sad so very sad!!!!'		
customer_service	1	['doesnt', 'real', 'person', 'spoke']
Example Message: 'i spent an hour and 15 minutes on hold with your fraud department this morning and never got through. now i am going on another hour on hold. i need to report fraudulent charges on my debit card. is there another number other than PHONE? this is getting unbelievably epic.'		
customer_service	2	['closing', 'month', 'reps', 'pay', 'tell', 'think', 'refund', 'error', 'trying']
Example Message: 'all the bleeding hearts for red cross dont change BankBs terrible customer service. 20 + years a customer with terminal cancer and i made a deposit error during chemo week. what does bank oif america do? they closed my account.'		
customer_service	3	['guys', 'say', 'sucks', 'hate', 'want', 'im', 'suck']
Example Message: 'Name been a BankC customer for nearly 10 years and sadly considering of leaving due to a mega crap customer service!!!(seeing as i cant swear on here).'		
customer_service	4	['home', 'love', 'job', 'loan', 'representative', 'happy', 'great', 'jobs', 'excellent']
Example Message: 'love BankB excellent customer service service Name making my money work for me saving for my retirement and these kids education'		
customer_service	5	['fees', 'fee', 'charge', 'didnt', 'deposit', 'atm', 'cash', 'closed']
Example Message: 'BankA customer service is one of the worst at getting stuff done i always have to escalate to higher up especibanke they love adding those overdraft fee(Name done with this bank)'		
customer_service	6	['horrible', 'experience', 'poor', 'terrible', 'need']
Example Message: 'have had the worse experience with BankB. the customer service is terrible. thinking that it is time to find a new bank.'		
customer_service	7	['minutes', 'branch', 'going', 'hold', 'close', '20', 'rep', '30', 'answer']
Example Message: 'i am a working person just like the employees of bank of Name i spent 15 minutes of my break on hold and i do not wish to spend any more time. i will be closing my credit card due to the fact of the service or lack of that you give your customers.'		
customer_service	8	['line', 'teller', 'bad', 'work', 'working']
Example Message: 'at BankB at 28th dr and peoria. Name is all the way to front door. clerk told me that there will only be 1 or 2 tellers working from now on cuz BankB wants to keep people out of their banks and to use other ways to do there banking transactions no customer service service!!!'		
customer_service	9	['days', 'issue', 'received', 'payment', 'company', 'weeks', 'department']
Example Message: 'a couple weeks ago i open a new BankB account within a few days they closed my account with no notice notification nothing can i have a direct deposit over$ 800 there i had the time i realize its closed i went to the bank and speak to the customer service live alone day having a stupid conversation their answer was completely rude and unprofessional and one of the ladies on the phone		

The LDA model we used to identify the substance produces a bag of words result. Therefore, the words associated with each of the clusters above are treated as individual unigrams. From the chart above, we can infer some actionable insights. For example, cluster 7 identifies “waiting” and “teller” at physical bank locations as the cause for these particular messages associated with the customer service topic. Cluster 5, on the other hand, identifies issues with “atms” as the cause.

Insights

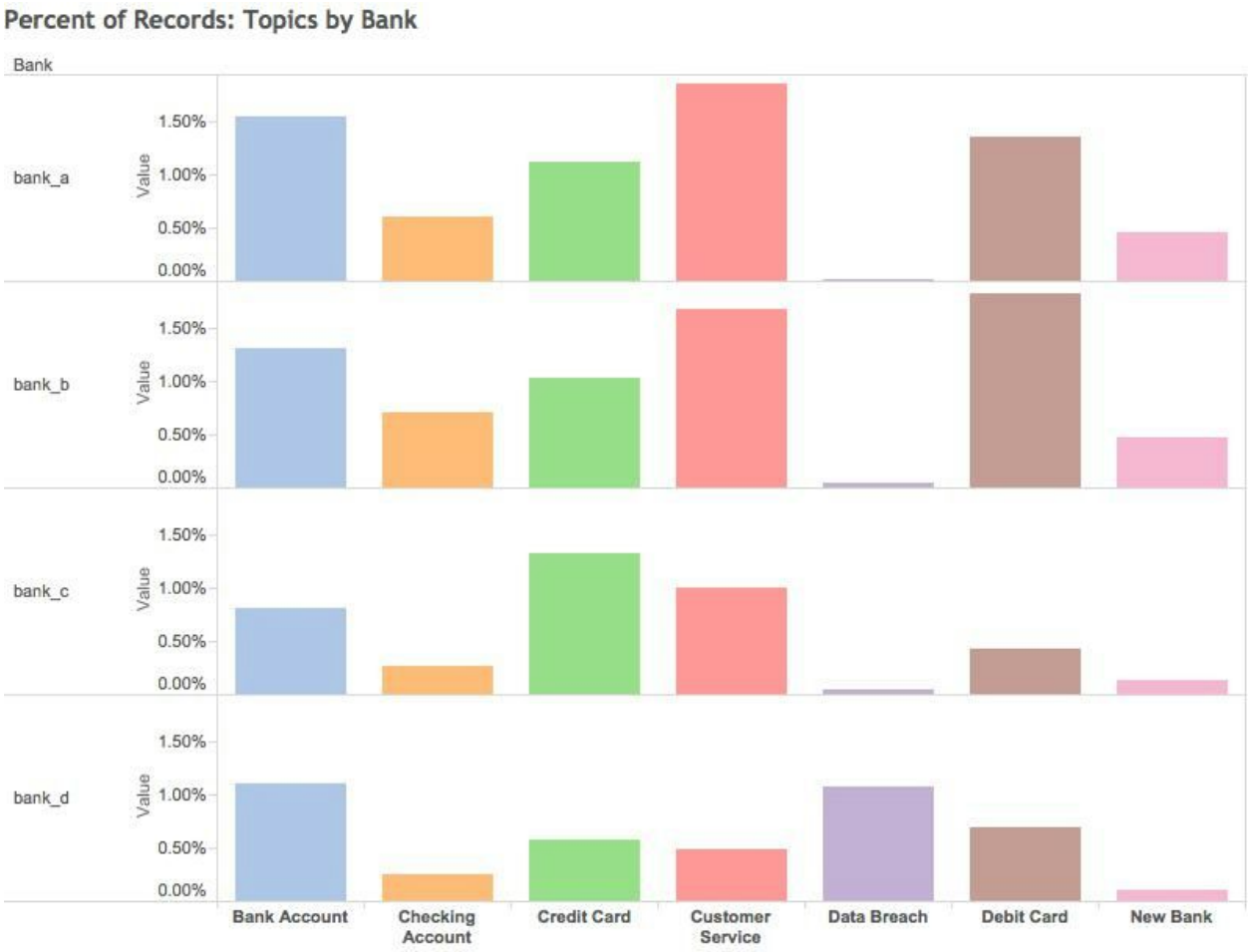
In order to explore whether these topics were applicable to the industry or if they were particular to a specific bank, we decided to visualize the frequency of comments for each topic across the the four banks. Initially, we found that certain topics were isolated to particular banks. For example, comments related to financial advice topics seemed to be heavily skewed towards BankD. While this is initially interesting, it may not be the most actionable finding, since this could be attributed to the fashion in which BankD uses its Facebook account.

Figure 2.4 Certain topics within the dataset were isolated within certain banks



We also explored other topics that were distributed more evenly between the banks. We normalized these records to the number of records that we had for each bank, so that we could compare the topics across the banks.

Figure 2.5 Percent of records for each bank that contain that topic



We found that there is a significantly higher proportion of people talking about data breaches with BankD, as compared with the other banks. This particularly piqued our interest due to the fact that it could be indicative of a significant data breach which occurred at a particular bank. While a majority of these messages were informational, there were some particular messages such as “so ... there was some sort of data breach at home depot ... looks like probably be getting another debit card from BankD... i think this will be 3 or 4 this year,” which was not.

One of the surprisingly useful topics is “new_bank”. While we were initially perplexed why this was a commonly occurring bigram, it was obvious after seeing a sample of messages containing the bigram. For example, “okay so i canceled my bank account at BankA cuz stupid shit so Name looking for a new bank ppl let me know which bank has u happy an satisfied with thier service!!!!”, and, “BankA is seriously getting annoying i may be at a hunt for a new bank. Name done with this!”. A majority of the messages included within this topic follow this pattern of people dissatisfied with the bank mentioned in their message and seeking advice to switch to a new bank. We believe that further analysis is warranted on this topic.

Conclusion

Our primary goal for this analysis was to identify various topics that were being discussed on social media and to find potentially actionable insights that were the drivers behind this discussions. Overall, we believe that there is clearly business value to this analysis, especially insofar as removing irrelevant messages and finding preliminary topics of messages. The second-order question of finding the substance of the message remains a challenge, but we believe that the LDA clustering provides a good foundation for a more sophisticated meaning extraction.

Appendix

Figure 3.1 List of revised topics and substances

Topic	#	Substance
bank_account	0	['donate', 'information', 'photo', 'shared', 'share', 'teamkeke', 'wanting', 'town', '5318430526', 'reach', 'felton', 'supporters', 'kenya', '10']
bank_account	1	['union', 'balance', '100', 'hit', 'quick', 'negative', 'brother']
bank_account	2	['dollars', 'told', 'days']
bank_account	3	['inbox', 'banke', 'interested', 'asap', 'making', 'federal', 'navy']
bank_account	4	['shit', 'lol', 'fuck', 'work', 'morning']
bank_account	5	['opening', 'let', 'banking', 'online', 'customers']
bank_account	6	['fees', 'customer', 'year', 'time', 'business', 'years', 'service', 'funds', 'big']
bank_account	7	['family', 'donations', 'does', 'opened', 'friends', 'set', 'funeral']
bank_account	8	['going', 'debit', 'rebanke', 'use', 'hacked', 'charges', 'think']
bank_account	9	['phone', 'scam', 'received', 'info', 'saying', 'text', 'called', 'security', 'email']
bank_robbery	0	['woman', 'robber', 'place', 'miami', 'person', 'alleged', 'new', 'took', 'hotel']
bank_robbery	1	['information', 'department', 'wanted', 'caught', 'phone', 'seek', 'involved', 'update', 'publics', 'asked']
bank_robbery	2	['investigate', 'south', 'near', 'friday', 'charlotte', 'inside', '30', 'taken']
bank_robbery	3	['wednesday', 'following', 'road', 'city', 'investigating', 'shot', 'custody', 'block', 'officers', 'boulevard', 'said']
bank_robbery	4	['searching', 'street', 'county', 'office', 'beach', 'say', 'sought', 'main']
bank_robbery	5	['armed', 'reported', 'attempted', 'avenue', 'north', 'scene', 'rob', 'cops', 'lexington', 'time', 'going', 'right', 'identified']
bank_robbery	6	['today', 'photos', 'occurred', 'video', 'creek', 'vehicle', 'black', 'charges', 'breaking', 'amarillo', '10', 'seen']
bank_robbery	7	['robbing', 'connection', 'monday', 'old', 'township', 'suspected', 'allegedly', 'arrest', '20', 'saturday']
bank_robbery	8	['photo', 'mansfield', 'released', 'surveillance', 'shared', 'suspects', 'yesterday', 'finding', 'downtown']
bank_robbery	9	['twit_hndl', 'incident', 'ave', 'asking', 'identifying', 'teller', 'alert', 'nyc', 'called', 'mapusemergalerts']
checking_account	0	['card', 'debit', 'credit', 'said', 'number', 'atm', 'week', 'use']
checking_account	1	['fee', 'monthly', 'balance', 'charging', '25', 'charged', 'having', 'checking', '12']
checking_account	2	['direct', 'say', 'opening', 'does', 'student', 'bonus']

checking_account	3	['savings', 'customers', 'didnt', 'change', 'think', 'rebanke', 'people']
checking_account	4	['service', 'payment', 'phone', 'loan', 'person', 'deal', 'car', 'cents', 'later']
checking_account	5	['thank', 'years', 'fraud', 'called', 'work', 'days', 'love', 'getting', 'taken']
checking_account	6	['cash', 'time', 'good', 'went', 'transfer', 'taking', 'closing', 'personal', 'mortgage', 'banks']
checking_account	7	['help', 'know', 'let', 'come', 'look', 'sure', 'way', 'right', 'youre', 'cards']
checking_account	8	['year', 'old', '30', 'hard', 'set', 'paid', 'ad', 'things', 'favorite', 'problem']
checking_account	9	['check', 'morning', 'closed', 'did', 'online', 'checks', 'banke', 'hold', 'yesterday', 'tell']
credit_card	0	['old', 'introduce', 'india', '10k', 'banking', 'instant', 'indian', 'federal', 'branch', 'usa', 'foreign', 'mr', 'limit']
credit_card	1	['payment', 'years', 'paid', 'mail', 'days', 'didnt', 'paying', 'late']
credit_card	2	['free', 'accept', 'small', 'apply', 'major', 'financial', 'weekend', 'million']
credit_card	3	['thank', 'fraud', 'week', 'order', 'time', 'trying', 'person', 'people']
credit_card	4	['number', 'phone', 'said', 'called', 'security', 'saying', 'scam']
credit_card	5	['debit', 'visa', 'rewards', 'chip', 'program', 'better', 'points', 'issue', 'mobile']
credit_card	6	['charge', 'fees', 'balance', 'going', 'closed', 'checking', 'fee']
credit_card	7	['great', 'amazon', 'think', 'past', 'did', 'banks', 'night', 'offer', 'visit', 'id']
credit_card	8	['company', 'home', '20', 'apple', 'wont', 'youre']
credit_card	9	['make', 'need', 'say', 'work', 'shit', 'bad', 'dollars']
customer_service	0	['card', 'debit', 'great', 'charges', 'use', 'fraud', 'cards', 'working']
customer_service	1	['doesnt', 'real', 'person', 'spoke']
customer_service	2	['closing', 'month', 'reps', 'pay', 'tell', 'think', 'refund', 'error', 'trying']
customer_service	3	['guys', 'say', 'sucks', 'hate', 'want', 'im', 'suck']
customer_service	4	['home', 'love', 'job', 'loan', 'representative', 'happy', 'great', 'jobs', 'excellent']
customer_service	5	['fees', 'fee', 'charge', 'didnt', 'deposit', 'atm', 'cash', 'closed']
customer_service	6	['horrible', 'experience', 'poor', 'terrible', 'need']
customer_service	7	['minutes', 'branch', 'going', 'hold', 'close', '20', 'rep', '30', 'answer']
customer_service	8	['line', 'teller', 'bad', 'work', 'working']
customer_service	9	['days', 'issue', 'received', 'payment', 'company', 'weeks', 'department']
data_breach	0	['cyberattack', 'recent', 'stolen', 'details', 'fbi', 'bankes', 'federal', 'huge']
data_breach	1	['personal', 'hit', 'household', 'names']
data_breach	2	['largest', 'hacked', 'biggest', 'history', 'world', 'latest', 'cybersecurity', 'suffered', 'recently']
data_breach	3	['cyber', 'financial', 'today', 'revealed', 'identity', 'major', 'hacking', 'good', 'video', 'attack', 'doing', 'services']

data_breach	4	['general', 'state', 'attorneys', 'illinois', 'attorney', 'connecticut', 'journal', 'street', 'read', 'look', 'servers', 'business']
data_breach	5	['card', 'cards', 'debit', 'credit', 'government', 'banks', 'issued', 'number', 'just']
data_breach	6	['76m', 'disclosed', 'affects', 'previously', 'systems', 'thursday']
data_breach	7	['83', 'reveals', 'affecting', 'millions', 'twitter', 'exposed', 'breaking']
data_breach	8	['simple', 'hack', 'fix', 'know', 'times', 'admits', 'need', 'avoided', 'tags', 'american']
data_breach	9	['year', 'depot', 'home', 'report', 'company', 'like', 'reported', 'identified', 'entry', 'target', 'high', 'cybercrime']
debit_card	0	['number', 'scam', 'text', 'saying', 'received', 'security', 'locked']
debit_card	1	['thank', 'info', 'tried', 'charges', 'shopping']
debit_card	2	['charge', 'business', 'fee', 'fees', 'pay']
debit_card	3	['cash', 'tell', 'deposit', 'went', 'did', 'day', 'say']
debit_card	4	['service', 'rebanke', 'having', 'great', 'night', 'think', 'problem']
debit_card	5	['depot', 'home', 'hacked', 'breach', 'information', 'data']
debit_card	6	['help', 'right', 'getting']
debit_card	7	['want', 'person', 'hope', 'took', 'lot', 'police', 'transactions']
debit_card	8	['credit', 'free', 'weekend', 'customers', 'chip', 'chips', 'museum', 'museums', 'debit', 'issue', 'holders']
debit_card	9	['sent', 'days', 'ago', 'years', 'hold', 'months', 'past']
financial_advisers	0	['uk', 'independent', 'growth', 'economy', 'debt', 'talks', 'forecast', 'brief', 'slows']
financial_advisers	1	['pension', '2015', 'tax', 'share', 'insurance', 'giant', 'car', 'revealed', 'help']
financial_advisers	2	['manchester', 'business', 'boss', 'boost', 'open', '2014', 'launches', 'sector']
financial_advisers	3	['rate', 'best', 'small', 'account', 'free', 'deals', 'businesses', 'fixed']
financial_advisers	4	['sales', 'prices', 'record', 'price', 'oil', 'years', 'low', 'energy']
financial_advisers	5	['bank', 'tesco', 'pay', 'home', 'hsbc', 'faces', 'bn', 'chief']
financial_advisers	6	['city', 'firm', 'building', 'hotel', 'jobs', 'plans', 'rail', 'salford', 'plan']
financial_advisers	7	['deal', 'firms', 'service', 'power', 'virgin', 'capital', 'loan']
financial_advisers	8	['profits', 'market', 'profit', 'stock', 'despite', 'hit', 'bid', 'warning']
financial_advisers	9	['ftse', 'live', 'markets', 'china', 'footsie', 'data', 'global', 'close', 'ahead']
real_estate	0	['2015', 'today', 'house', 'loan', 'got', 'did', 'money', 'day', 'housing', 'contact', 'phone', 'management', 'closing', 'time', 'april']
real_estate	1	['blackstone', 'deal', 'billion', 'buy', 'real', 'group', 'selling', 'portfolio', 'sell', 'assets', 'electric', 'general', '30']
real_estate	2	['friends', 'want', 'just', 'deck', 'need', 'originbanke', 'paper', 'song', 'feat', 'views', 'work', 'big', 'wait']

real_estate	3	['street', 'main', 'vote', 'https', 'mission', 'tax', 'internet', 'program', 'learn', 'rating', 'global', 'building']
real_estate	4	['agent', 'center', 'help', 'resource', 'social', 'media', 'education', 'professionals', 'grow', 'information', 'manage', 'broker', 'love', 'buyer']
real_estate	5	['commercial', 'market', 'office', 'property', 'buildings', 'firm', 'owned', 'sale', 'deals', 'downtown', 'current', 'journal', 'realestate']
real_estate	6	['news', 'development', 'thanks', 'report', 'video', 'court', 'membership', 'second', 'panel', 'case', 'series', 'foreclosure', 'training']
real_estate	7	['international', 'article', 'irvine', 'settlement', 'national', 'list', 'lending', 'pay', 'view', 'lender', 'paid']
real_estate	8	['home', 'great', 'community', 'buyers', 'good', 'job', 'homes', 'making', 'million', 'like', 'purbankd']
real_estate	9	['nyc', 'photo', 'tigho', 'shared', 'bike', 'area', 'manhattan', 'residential', 'team', 'party', 'family', 'florida', 'looking', 'north', 'tonight']
small_business	0	['llc', 'photography', 'farm', 'consulting']
small_business	1	['salon', 'care', 'home', 'spa']
small_business	2	['studio', 'bakery', 'fitness', 'dance']
small_business	3	['million', 'es', 'business', '76', 'households', 'bank', 'account', 'breach', 'data', 'grant', 'new', 'information', 'customer', 'compromised', 'owners']
small_business	4	['center', 'house', 'yoga', 'coffee']
small_business	5	['design', 'solutions', 'club', 'corporation']
small_business	6	['group', 'cafe', 'music', 'service']
small_business	7	['academy', 'institute', 'therapy', 'arts']
small_business	8	['company', 'services', 'restaurant', 'wellness']
small_business	9	['grants', 'entertainment', 'big', 'blue']
wealth_managers	0	['rate', 'tax', 'rates', 'low', 'account', 'mortgage', 'free', 'deals']
wealth_managers	1	['deal', 'profit', 'warning', 'warns', 'greece', 'tesco', 'government']
wealth_managers	2	['boss', 'firm', 'pay', 'service', 'chief', 'power', 'mobile']
wealth_managers	3	['uk', 'sales', 'profits', 'fall', 'growth', 'hit', 'record', 'rise', 'despite']
wealth_managers	4	['market', 'money', 'stock', 'london', 'virgin', 'firms', 'best', 'quarter', 'plan']
wealth_managers	5	['manchester', '2015', 'city', 'building', 'open', 'hotel', 'plans', 'revealed']
wealth_managers	6	['years', 'share', 'house', 'property', 'group', 'christmas', 'prices', 'launch', 'cap']
wealth_managers	7	['ftse', 'live', 'oil', 'markets', 'china', 'footsie', 'data', 'global']
wealth_managers	8	['set', 'energy', 'cut', 'says', 'bid', 'customers', 'bn', 'bills', 'months', 'gas']
wealth_managers	9	['pension', 'boost', 'cash', 'big', 'banks', 'businesses', 'world', 'car']

Figure 3.2 Number of Records by Bank

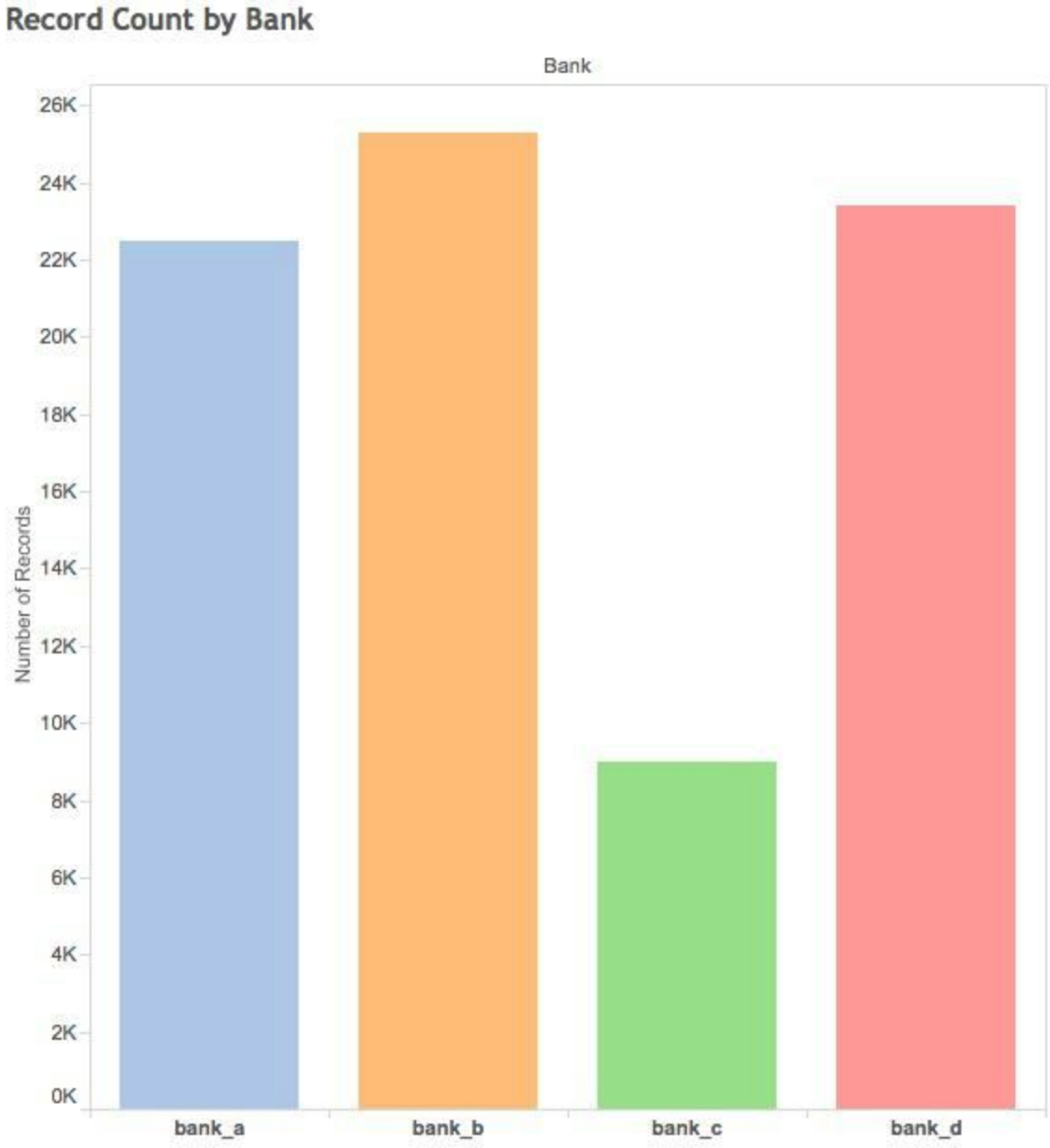


Figure 3.3 Number of Records by Topic and Bank

Bank	Bank Account	Checking Accour	Credit Card	Customer Servic	Data Breach	Debit Card	New Bank
bank_a	348.0	136.0	253.0	416.0	3.0	303.0	101.0
bank_b	330.0	177.0	259.0	425.0	11.0	463.0	120.0
bank_c	73.0	24.0	119.0	90.0	4.0	39.0	12.0
bank_d	260.0	59.0	133.0	114.0	251.0	161.0	24.0