CFPB Consumer Complaint Database

Boon or Bane?

Dipika Pushpakaran, 11810132

Shrey Nayak, 11810033

Pooja Tanaji Patil, 11810123

Executive Summary

The Consumer Complaint Database maintained by the Consumer Finance Protection Bureau (CFPB) has been used to log and track the complaints against various financial companies, banks, debt collectors and more. The new leadership at CFPB is reviewing whether the public database will be continued. This report provides an analysis and evaluation of the Consumers Complaint Database maintained by the CFPB since July 2011 and evidence to substantiate recommendations to allow continued public access to the database.

Exploratory data analysis performed on the Hadoop ecosystem using Cloudera and Apache Pyspark include trend, horizontal and vertical analysis as well as ratios such as Company Response Time and Consumer Dispute Rate ratio. The report also takes a closer look at the observations during the 2017 Equifax Data Breach incident for valuable insights into the company's response as well as the impact of the breach incident. All SQL queries and calculations can be found in the appendices.

Results of the analysis show that the Consumer Complaint Database provide a rich trove of information not only for the consumers and CFPB to highlight problems and regulate the financial marketplace but also for the companies who can use this information to improve their products and services.

Recommendations include

- Continue to provide public access to the consumer complaint database so that the financial institutions and CFPB can be held accountable for fair practices in the market place.
- Continued access will also benefit the customer who can be more aware of their rights and make informed decisions while availing products and services from various financial institutions.
- Extending the analysis and research to include forecasting of company or product behavior as well as a means of identifying issues in the financial market place.

The report also accepts the fact that the analysis conducted has limitations. Some of the limitations include: data limitations as not enough information is provided if the company has received equal or more complaints for the products and services. Also, the research assumes that complaints listed in the database are all genuine and verifiable complaints.

CONTENTS

Consumer Complaint Database Analysis

The Consumer Financial Protection Bureau (CFPB) in United States regulates the offering and provision of consumer financial products or services under the federal consumer financial laws and empowers consumers to make better informed financial decisions. They collate and send thousands of consumers' complaints about financial products and services to companies for response. Those complaints are published in the **Consumer Complaints Database** after the company responds or after 15 days, whichever comes first.

The Consumer Complaints Database has been made public for almost 7 years now. Financial institutions have been against the database because they believe it to be misleading as the agency does not vet the complaints it receives. Consumer advocate groups claim that the database can be used to identify bad behavior and coerce companies to provide better service. Early this year, the database was in the news as the bureau considers closing the public database.

Our exploration of the database was aimed at checking the claims of the two groups along with analysis of the complaints. Our analysis has produced several valuable insights about the nature of the complaints as well as the company's response to the consumer grievances. The insights show how retaining the Consumer Complaints Database and allowing open access to the general public not only benefit the consumers, but also the financial institutions and the CFPB.
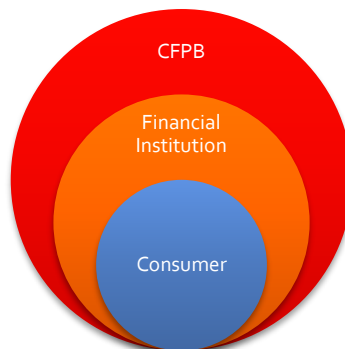
Introduction

The analysis aims at exploring the following ideas for the three main stakeholders for the Consumer Complaint Database.

Consumer: Can the database provide the consumers with information to make better product and company choices and be aware of the consumer rights.

Financial Institutions: Can the database equip the financial service providers with customer experiences and reviews to better their services.

CFPB: Can the database provide information to the CFPB to make responsible laws and regulate the financial marketplace.

CFPB

Financial Institution

Consumer

About the Database

The Consumer Complaint Database is a collection of complaints on a range of customer financial products and services, sent to the company for a response. Since it was introduced in 2011, more than 1 million complaints have been logged in the Consumer Complaint Database. The consumer complaint process includes 11 popular financial products and services like mortgages, credit cards, student loans, bank products and services amongst others. The database can be downloaded from the consumerfinance.gov website as a CSV or JSON file.

HOW THE DATABASE WORKS

The CFPB started accepting consumer's complaints in July 2011. Consumers can submit their complaints online, over the phone, fax or email in the format outlined in the submit form.

The database generally updates daily and contains information for each client in addition to the actions taken by the company.

Each week the CFPB reviews the customer complaints and publishes the complaints on their website in 15 days or after the company responds whichever is earlier. The bureau does not verify the complaints for accuracy or intervene in the resolution of the complaints but does verify there is a relationship between the consumer and financial institution.

The CFPB also analyzes the data in the database and produces periodic reports depicting their findings and recommendations. They also share the complaint data with the state and federal agencies. The data is also used to enforce federal consumer financial laws, and to write better rules and regulation.

DATASET LAYOUT

Field	Description
DATE RECEIVED	The date the CFPB received the complaint
PRODUCT	The type of product the customer identified in complaint
SUB-PRODUCT	The type of sub-product the customer identified in the complaint
ISSUE	The issue the consumer identified in complaint
SUB-ISSUE	The sub-issue the consumer identified in the complaint
COMPANY PUBLIC RESPONSE	The company's optional public-facing response to consumer
COMPANY	The complaint is about this company
STATE	The state of the mailing address of the consumer
ZIP CODE	The mailing ZIP code provided by the consumer
TAGS	Data that supports easier sorting and searching of complaints
CONSUMER CONSENT PROVIDED	Identifies whether consumer has identified to publish the complain
SUBMITTED VIA	How the complaint was submitted
DATE SENT TO COMPANY	The date CFPB sent to the company
COMPANY RESPONSE TO CONSUMER	This is how the company responded
TIMELY RESPONSE	Whether the company gave a timely response
CONSUMER DISPUTED	Whether the consumer disputed the company's response
COMPLAINT ID	The unique identification number of a complaint

DATASET USED IN ANALYSIS

Size : 262MB

Format : CSV

Source : Consumerfinance.gov

Download date: 21st January 2019

Complaints : 1,048,575

Start Date : January 2013

End Date : January 2019

Exploration Questions of Interest

To edit the data for either chart in this document, select the chart and then, on the Chart Tools Design tab, in the Data group, click Edit Data.

1. What are the top ten companies by the volume of complaints, and what is the distribution of products for which the complaints have been raised for each company?
2. What are the top ten products and issues by the volume of complaints?
3. Has the public database forced companies to take complaints seriously? – For the top ten companies by volume of complaints, check the distribution of the complaints over the past five years - has it decreased, increased or remained the same.
4. Can the database be used to detect bad behavior of the companies?
    a. What are the top five companies which do not give a timely response to the consumers?
    b. What are the top five companies for whose complaints the consumers have disputed the resolution?
    c. What is the distribution of complaints with timely response and consumer disputes from 2013 – 2019?
5. Are complaints misleading or misinterpreted – what is the share of complaints that were closed in the favor of the consumer?
6. Are the complaints misleading or misinterpreted – of the complaints that were filed, how many are misunderstood by the consumer?
7. What is the share of the complaint resolutions that were disputed by the consumer?
8. Is the Database easily accessible - what is the most popular method for consumers to submit the complaints to the CFPB?

9. How can the CFPB use the consumer complaint database to identify problems in the financial services industry by the trends of the complaint database. Take a close look at the observations during the Equifax Data Breach.
   a. What were the products and issues which were impacted by the Equifax Data Breach?
   b. What are the top ten states which were impacted by the data breach?
   c. How was the company response during the crisis – did they respond to the consumers in a timely manner? Were consumers satisfied with the resolutions provided? How many complaints were closed with monetary relief?

Methodology

Tools/Software used: Hadoop, Cloudera, Apache Spark, Spark SQL

- The database contains over 1.4 million rows and the size is over 200MB. Given the size of the dataset, it would be prudent to load and process the information on the Hadoop ecosystem. Hadoop is an ecosystem or platform that is used for big data problems. It consists of the following three parts
  - Hadoop Distributed File System : used for file handling.
  - MapReduce : responsible for all data processing tasks
  - Yarn : acts as the resource manager
- Apache Spark can be used on Hadoop to process the big data. Apache Spark is fast and unified analytics computing engine. It has in-built modules for streaming, SQL and machine learning.
- Spark SQL is the inbuilt module of Apache Spark for querying structured data. It overcomes the disadvantages of the Spark RDD to give a dataframe interface that allows us to view and query the structured data.
- We will use the Cloudera VM services. The Cloudera VM comes installed with Hadoop, Spark and other necessary tools and software to conduct the analysis. Any additional packages required as part of the analysis can be downloaded using pip install.
- To begin the process, download the Consumer Complaint Database in the CSV format from the Consumerfinance.gov website.
- Import the data from the CSV file into Spark while removing the Consumer Complaint Narrative coloumn which is not used in the analysis.
- As part of data validation, run a few checks for the count of records and data types.
- As part of data preprocessing, rename the coloumn names to suitable names, reformat the date coloumns to a common format.
- Register the complaints dataframe to a table so that the data can be brought to a structured format and queried.
- Execute the various SQL queries using WHERE, GROUP BY, ORDER BY and JOIN clauses to get the analysis results.
- Use Matplotlib and Seaborn library to plot the various visualizations.
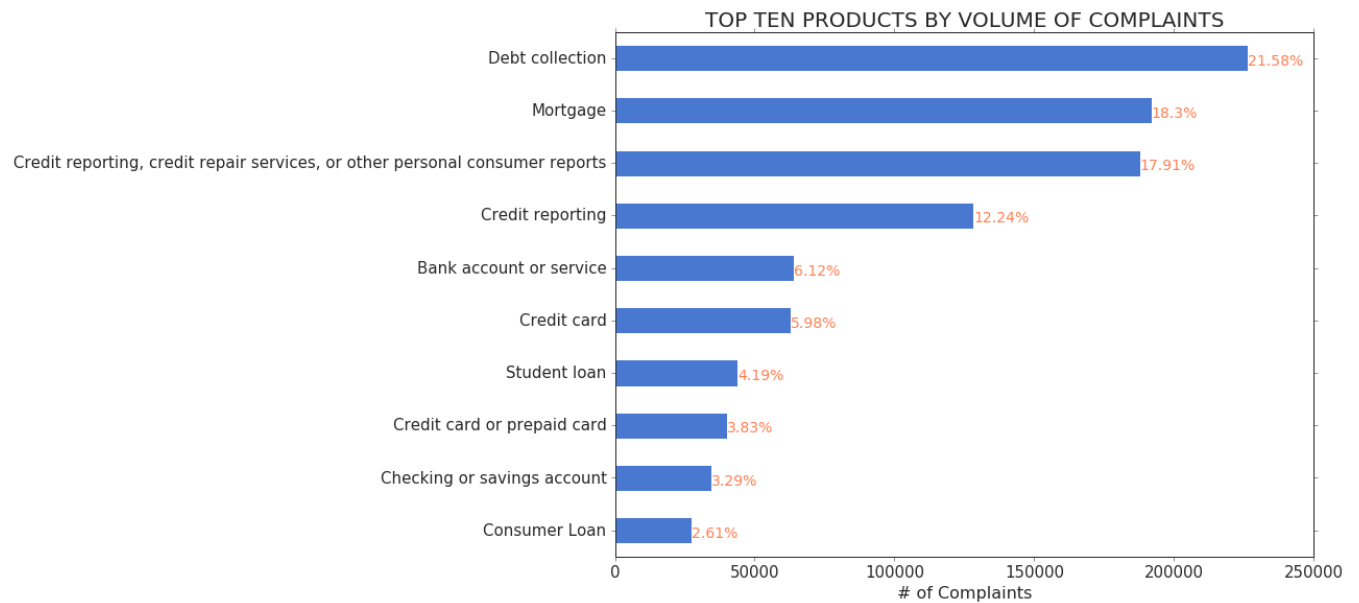
Analysis Reports

Findings

1. What are the top ten companies by the volume of complaints in the Consumer Complaints Database?

TOP TEN COMPANIES BY VOLUME OF COMPLAINTS
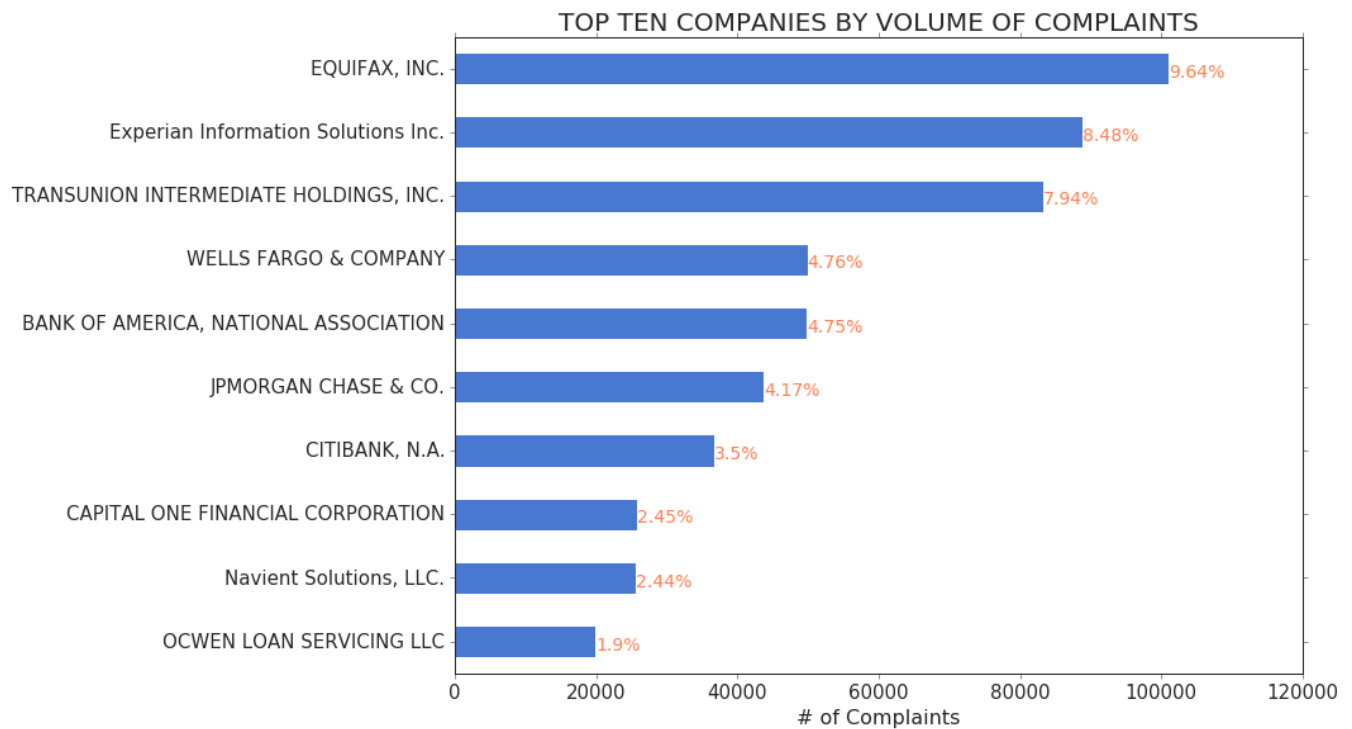
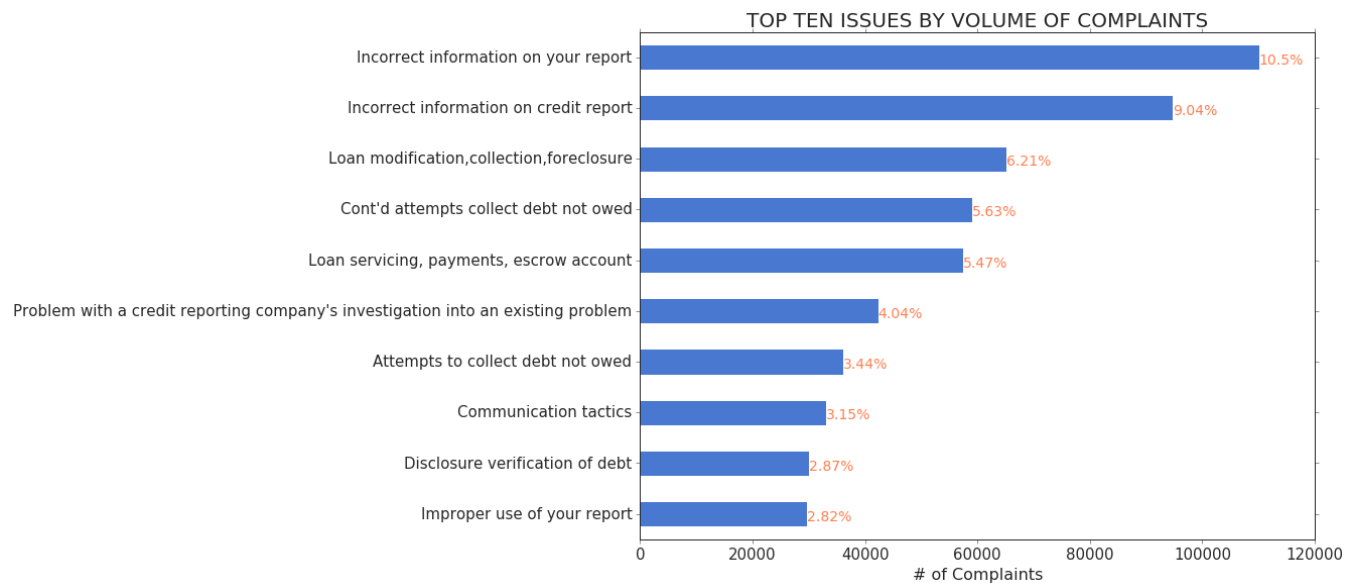- Despite the fact that the number of complaints received for the company depends on the size of the company and it's customer base, the above graph allows the CFPB, academics and watchdog organizations to identify the financial institutions which should be closely monitored in the marketplace.

- Equifax constitute almost 10% of the consumer complaints, which is about 100K complaints. We will analyze this further in our research.

2. What are the top ten products by the volume of complaints in the Consumer Complaints Database?
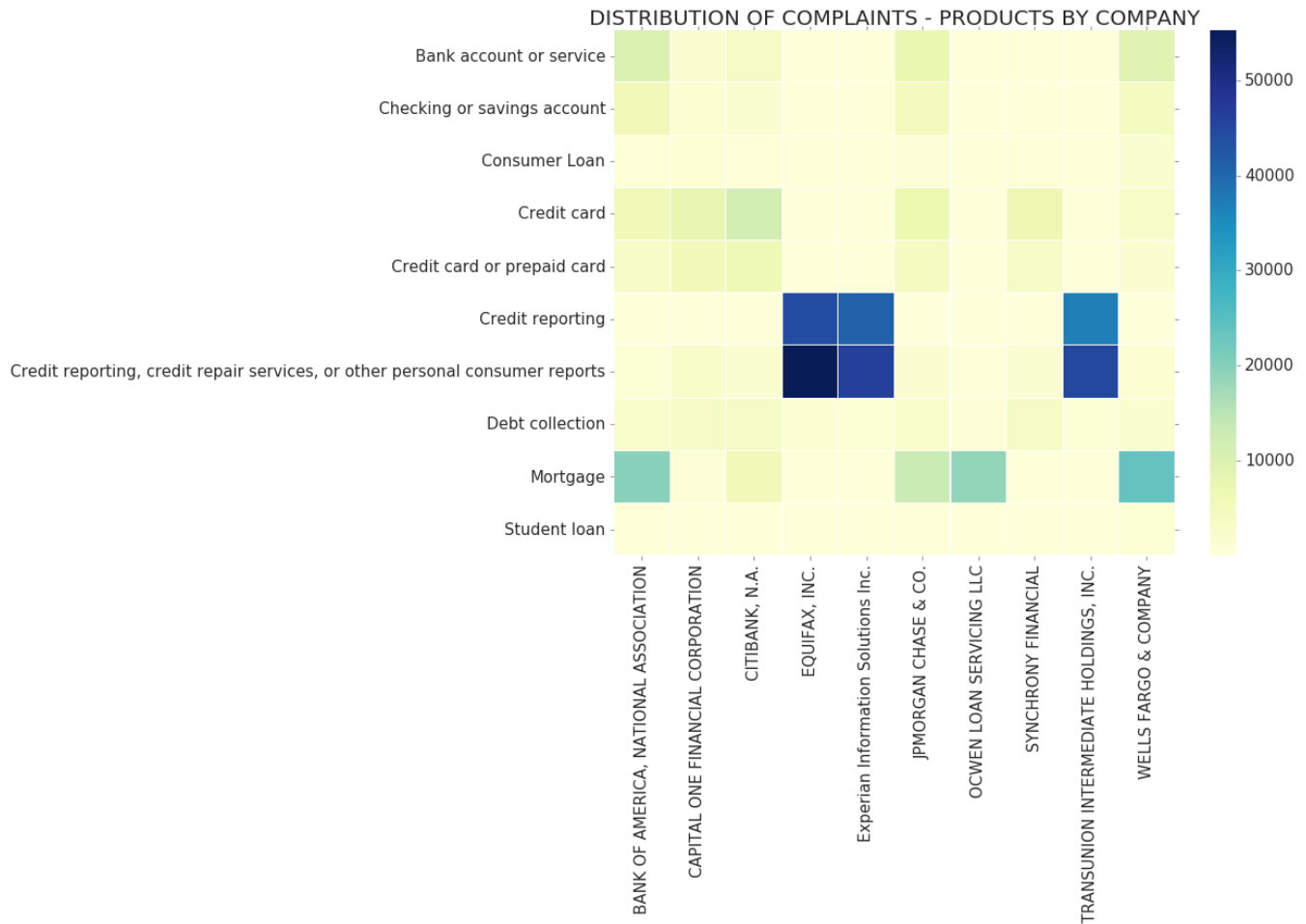
TOP TEN PRODUCTS BY VOLUME OF COMPLAINTS

Product	%
Debt collection	21.58%
Mortgage	18.3%
Credit reporting, credit repair services, or other personal consumer reports	17.91%
Credit reporting	12.24%
Bank account or service	6.12%
Credit card	5.98%
Student loan	4.19%
Credit card or prepaid card	3.83%
Checking or savings account	3.29%
Consumer Loan	2.61%

# of Complaints

3. What are the top ten issues by the volume of complaints in the Consumer Complaints Database?

TOP TEN ISSUES BY VOLUME OF COMPLAINTS

Issue	%
Incorrect information on your report	10.5%
Incorrect information on credit report	9.04%
Loan modification,collection,foreclosure	6.21%
Cont'd attempts collect debt not owed	5.63%
Loan servicing, payments, escrow account	5.47%
Problem with a credit reporting company's investigation into an existing problem	4.04%
Attempts to collect debt not owed	3.44%
Communication tactics	3.15%
Disclosure verification of debt	2.87%
Improper use of your report	2.82%

# of Complaints

4. For the top ten financial products by volume of complaints, what is the distribution of complaints across the top ten companies that offer these products?

DISTRIBUTION OF COMPLAINTS - PRODUCTS BY COMPANY

- The above heat map shows the number of complaints raised for the ten products offered by top ten companies by the volume of complaints.
- Bank of America, Wells Fargo, JP Morgan Chase are the top mortgage lenders in the US and consequently also have the largest number of mortgage complaints. Equifax has highest number of complaints for credit reporting. We will review the complaints closely later in the research.
- The heat map can be used by CFPB to easily identify the products which are not serviced to consumer satisfaction by these companies.
- Companies can also use this data to identify their problem areas and provide better services for these products. It also gives a glimpse into the consumer satisfaction for the competitors' products and services.
- Consumers can use this heat map to educate themselves on the offerings by various companies and make better choices for the products and companies. For eg., credit reporting for Equifax has the maximum complaints in the heat map, and a consumer should do well before engaging Equifax for credit reporting.

5. **Has the public database forced companies to take complaints seriously? – For the top ten companies by volume of complaints, check the distribution of the complaints over the past five years - has it decreased, increased or remained the same.**

As you can see, the number of complaints per year has increased from 2013 to 2018. This need not necessarily indicate a downturn in the products and services provided by the company. It could also be due the growing awareness of the consumer about the consumer complaint database and can also be attributed to the growth in the companies itself.
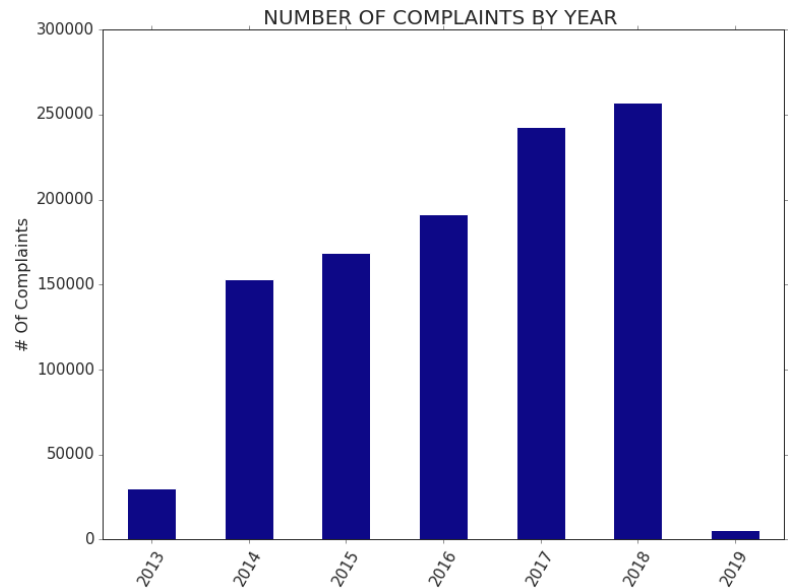
The below heatmap shows the increase in complaints against the top ten companies from 2014 to 2018. Equifax and Transunion Intermediate had a sudden increase in complaints in 2017-2018 and this could be indicative of a service problem in the products offered by the two companies. We will review this later in our research. Ocwen Loan Servicing is the only company for who the volume of complaints has reduced from 2014 to 2018.
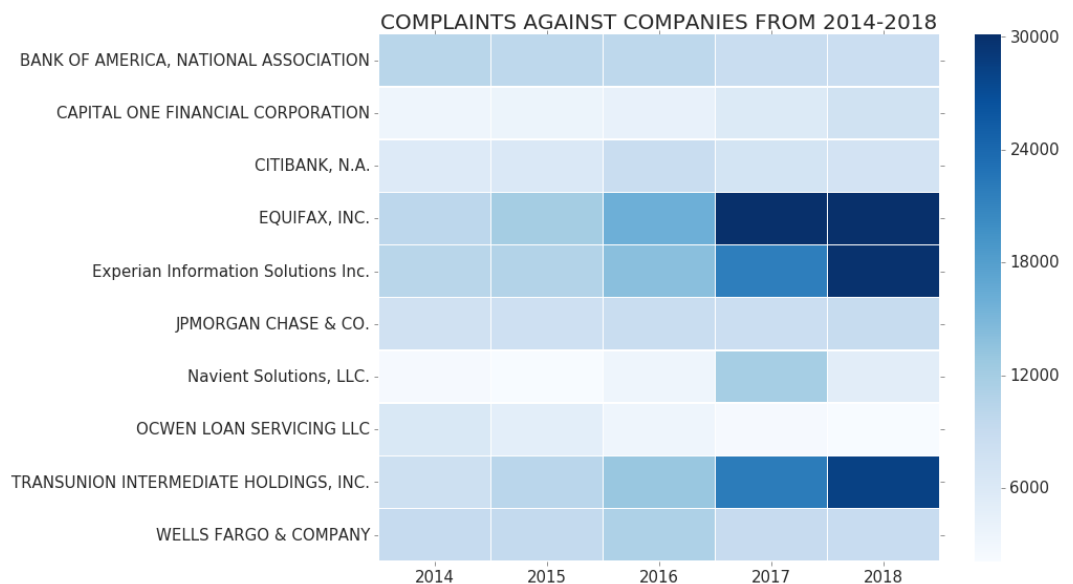
6. Can the database be used to detect bad behavior of the companies?

- **What are the top five companies which do not give a timely response to the consumers?**

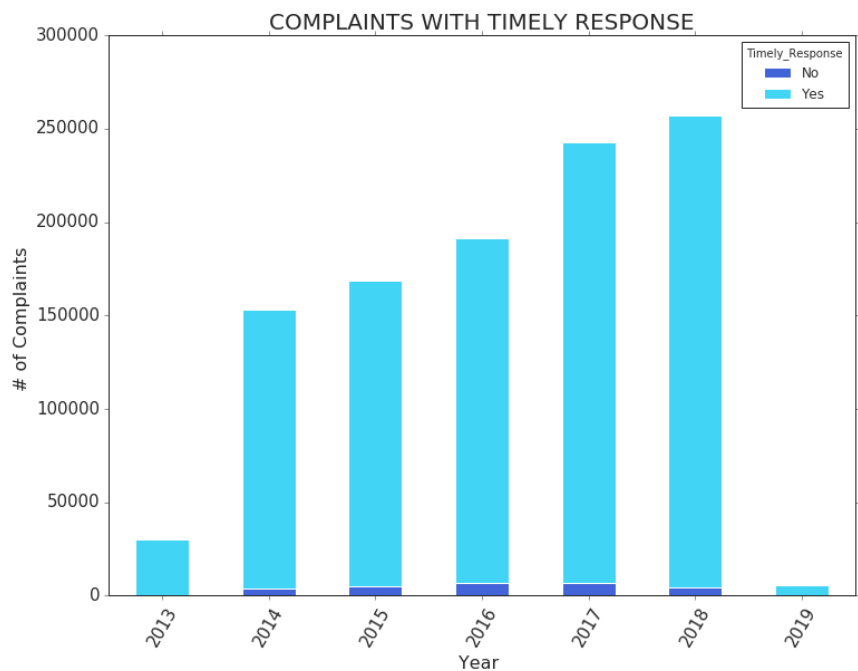The Company No Timely Response Rate was calculated for each company using the below formula

Company No Timely Response Rate = $\dfrac{\text{Number of Complaints with No Timely Response for the Company}}{\text{Total number of Complaints against the Company}}$

The following companies had the highest Company No Timely Response Rate = 1. However, this is not really a good indicator of the company's services since these companies received less than 5 complaints.

a. Medical Services Inc
b. Gillman Interests, Inc
c. Patriot Lending Services Inc
d. WFI Funding Inc
e. Direct Lending Group, Inc

The CFPB claims that more than 97% of the complaints in the database receive a timely response. It would be worthwhile to check if the complaints have been receiving a timely response from the companies. This is a good measure to check if the companies are accountable and taking the complaints in the consumer complaint database seriously.

As you can see from the bar chart on the right, the complaints do receive a timely response. In fact, in 2018 the number of timely responses has drastically improved in comparison to the previous years.

COMPLAINTS WITH TIMELY RESPONSE

- **Have the consumers receiving a good resolution from the companies since the Consumer Complaint Database has been instated?**

The Company Consumer Dispute Rate was calculated for each company using the below formula

Company Consumer Dispute Rate = Number of Complaints with Consumer Dispute = Yes

Total number of Complaints against the Company

The following companies had the highest Company Consumer Dispute Rate = 1. However, this is not really a good indicator of the company's services since these companies received less than 5 complaints.
    a. Midwest Bankers Mortgage Services Inc
    b. First American Financial Corporation
    c. HJ Ventures, LLC
    d. Asset One Mortgage Inc
    e. M H Loans Corporation

In 2017, the indicator for the Consumer Dispute was discontinued. Hence you can see the value as N/A from mid 2017.

However from 2013 – 2016, you can see the number of resolutions which were disputed have gone down. This does not necessarily indicate that the consumer received a monetary relief or was correct. It could also be an indicator that the complaint was closed with a good explanation which satisfied the consumer.

COMPLAINT RESOLUTIONS DISPUTED BY CUSTOMER
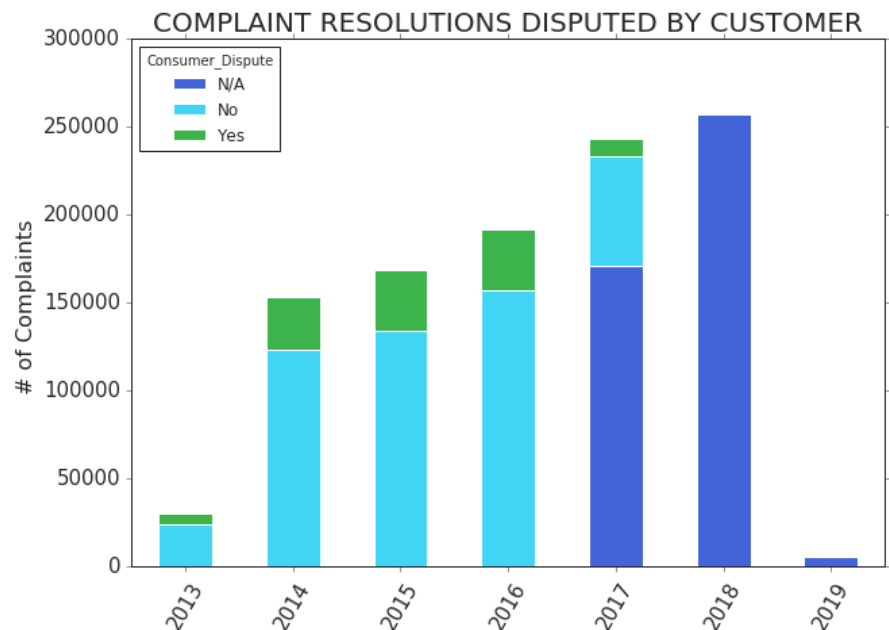
- **Are the complaints misinterpreted or misunderstood – what is the share of the complaints which have received a good resolution in the form of monetary relief from the company?**

Financial institutions can resolve the complaint with monetary relief (in favour of the consumer) or they can close the complaint with explanation.

The number of complaints resolved in favor of the consumer with monetary relief has reduced during the analysis period. This could indicate that a lot of the complaints could be the consumers error vs an actual genuine complaint.

COMPANY RESPONSE TO COMPLAINTS

7. **Are the complaints misinterpreted or misunderstood? Of the complaints that were filed by the consumers and the company gave a public response, how many were misinterpreted or misunderstood as listed in the company's public response?**

COMPANY's PUBLIC RESPONSE TO COMPLAINT	COUNT
Company has responded to the consumer and the CFPB and chooses not to provide a public response	276713
Company believes it acted appropriately as authorized by contract or law	56960
Company chooses not to provide a public response	52472
Company believes the complaint is the result of a misunderstanding	5318
Company disputes the facts presented in the complaint	4961
Company believes complaint caused principally by actions of third party outside the control or direction of the company	3800
Company believes complaint is the result of an isolated error	3509
Company believes complaint represents an opportunity for improvement to better serve consumers	2408
Company can't verify or dispute the facts in the complaint	2213
Company believes complaint relates to a discontinued policy or procedure	61

From the above table, you can see that there only about 1 percent of the complaints that the company responded to are complaints that have been raised because of a misunderstanding or error.

8. Is the database easily accessible - what is the most popular method for consumers to submit the complaints to the CFPB?

The most popular mode of submitting the consumer complaints is via the web form. Less than 1 percent of the complaints were submitted via email and fax.
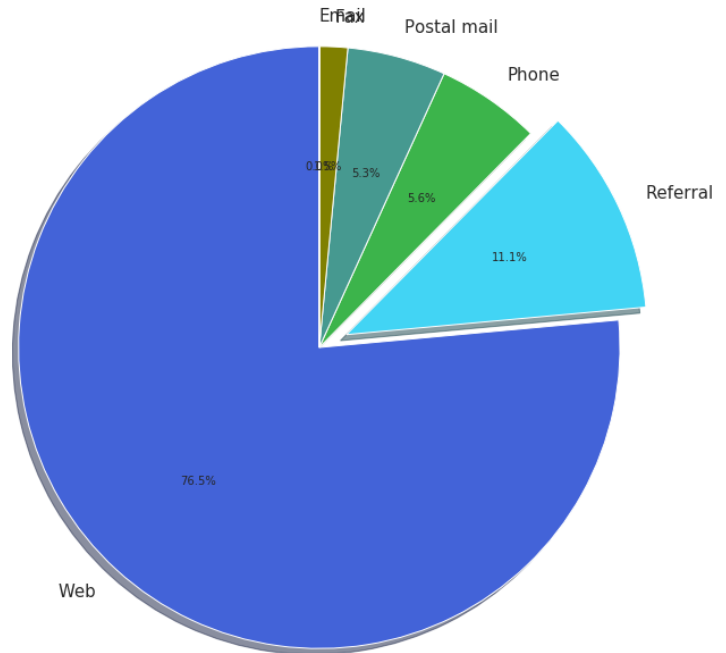
EQUIFAX INC. DATA BREACH - 2017

In September 2017, Equifax, Inc., one of the largest credit reporting companies in the United States of America reported the leakage of personally identifiable information of over 145.5 million US customers due to months-long illegitimate access to the credit-report databases. Even though the company waited for 6 weeks to report the incident, it acted quickly post announcement and sent direct emails to consumers whose credit card details or PII information was accessed and created resources to assist the consumers. Polls conducted by CreditCards.com however showed that Americans were slow to react to the Equifax data breach and only between 1-4 Americans checked the credit scores post the breach.

In this review of the Consumer Complaint Database, we will look closely at the observations during the Equifax data breach incident to understand nature of the complaints raised and the response of the company to the complaints. We will also review the trends to check if the complaints in the database can be used as an indicator of service issues in the financial marketplace.

9,736
Equifax Inc. Complaints in September 2017
Monthly Average – 1010

99 %
Timely response for complaints

99 %
Complaints closed without monetary relief

9. **How can the CFPB use the consumer complaint database to identify problems in the financial services industry by the trends of the complaint database. Take a close look at the observations during the Equifax Data Breach event of 2017.**

NUMBER OF COMPLAINTS BY DAY -2017
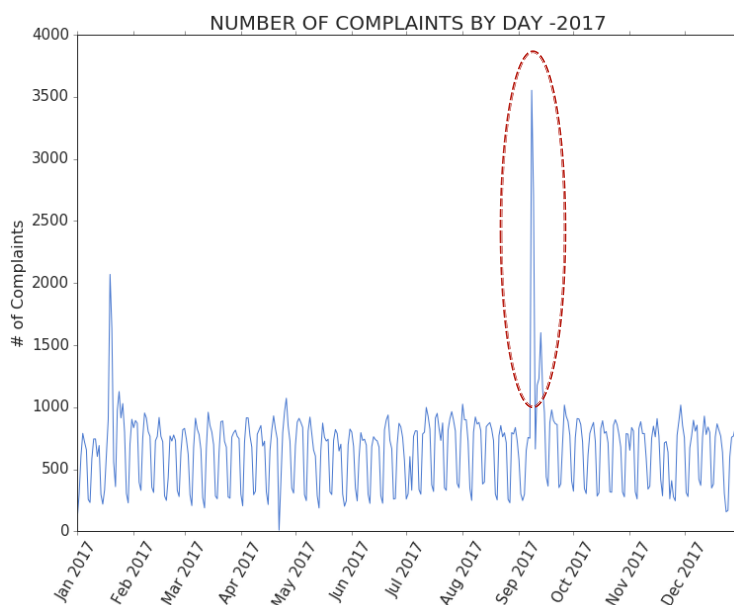
From the heat map listed earlier in the report, we have observed that the Equifax Inc complaints had a sharp increase for all complaints in the year 2017. The line graph on the left shows the sharp increase in the volume of the complaints during September 2017.

Such increase in the volume of complaints can be a good indicator for the CFPB to take a closer look at the observations to identify problems in products and services.

Alternatively, a drop in the complaints for a company can also be an opportunity to review and highlight any good practices taken by the company.

The line graph on the right shows us the steep increase in the complaints on September 7 and after that. This coincides with the announcement that Equifax Inc made in the media about the credit card data breach.
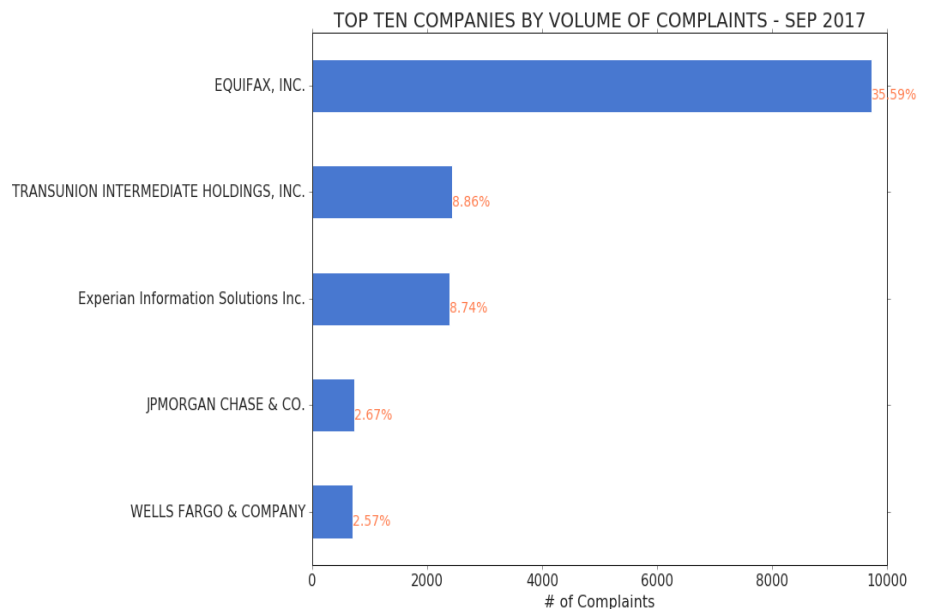

VOLUME OF COMPLAINTS BY DAY - SEPT 2017
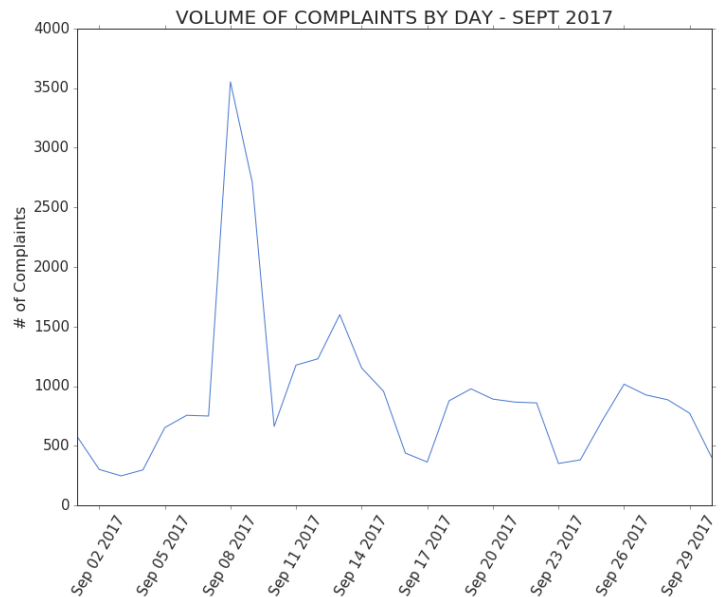
10. Now that we know the dates when there was a spike in the data, let us look at the data to see which companies had highest number of complaints during the period.

As you can see from the graph, a whopping 35% of the complaints in September 2017 were for Equifax Inc.

The monthly average number of complaints received for Equifax Inc is 2342. However in September 2017, the number of complaints for Equifax Inc alone were 2321.

Interestingly, the monthly average of the Transunion Intermediate Holdings Inc company, which is also a big credit reporting company also went up in September 2017. This shows how the consumers all over the financial market place may have worried about the data breach and would have checked their credit reports.

TOP TEN COMPANIES BY VOLUME OF COMPLAINTS - SEP 2017

EQUIFAX, INC. — 35.59%
TRANSUNION INTERMEDIATE HOLDINGS, INC. — 8.86%
Experian Information Solutions Inc. — 8.74%
JPMORGAN CHASE & CO. — 2.67%
WELLS FARGO & COMPANY — 2.57%

11. Now that we know the dates when there was a spike in the data, let us look at the data to see which products had highest number of complaints and the highest occurring issues during the period.

TOP TEN PRODUCTS BY VOLUME OF COMPLAINTS - SEP 2017

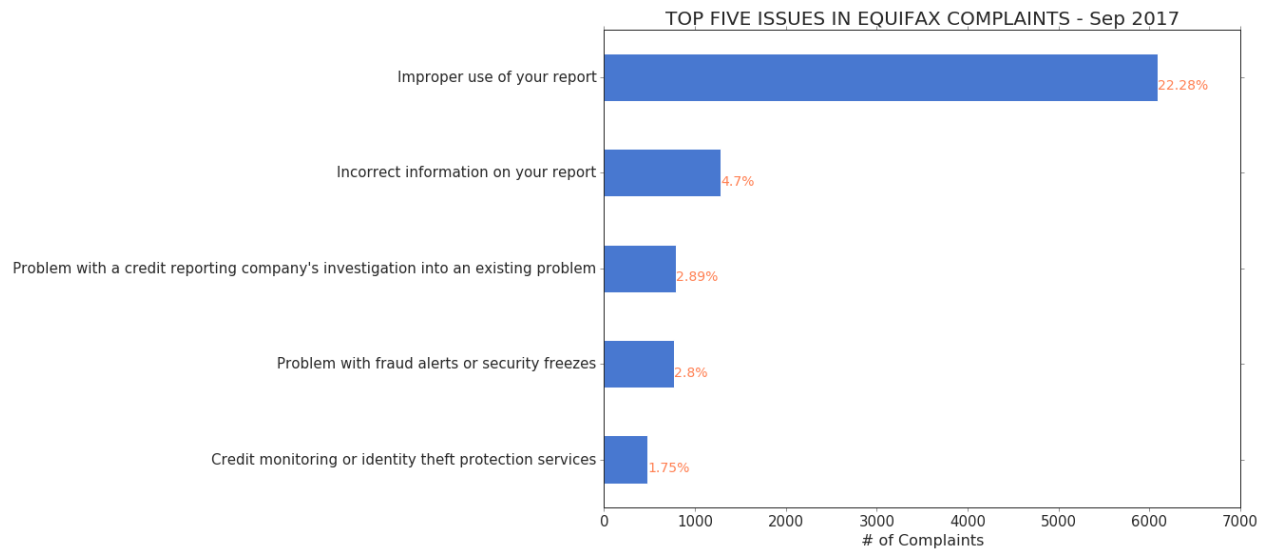| Product	Percentage
Credit reporting, credit repair services, or other personal consumer reports	58.83%
Debt collection	12.94%
Mortgage	8.28%
Credit card or prepaid card	6.84%
Checking or savings account	5.75%
Student loan	3.16%
Vehicle loan or lease	1.61%
Payday loan, title loan, or personal loan	1.29%
Money transfer, virtual currency, or money service	1.28%

# of Complaints

TOP TEN ISSUES BY VOLUME OF COMPLAINTS - Sep 2017

Issue	Percentage
Improper use of your report	26.27%
Incorrect information on your report	16.42%
Problem with a credit reporting company's investigation into an existing problem	7.41%
Attempts to collect debt not owed	5.33%
Problem with fraud alerts or security freezes	5.24%
Managing an account	3.49%
Struggling to pay mortgage	3.4%
Trouble during payment process	3.25%
Written notification about debt	2.96%
Credit monitoring or identity theft protection services	2.32%

# of Complaints

12. Check the top issues listed in the Equifax Inc data breach in September 2017.

TOP FIVE ISSUES IN EQUIFAX COMPLAINTS - Sep 2017

As is expected, the top five issues listed in the Equifax data breach coincide with the top ten issues reported in the consumer complaint database that month.

13. What were the top ten states that were impacted by the Equifax Inc data breach incident? This is estimated by the total number of complaints.

Almost 20% of the complaints are from the California, which make it the top state to be impacted by the credit data breach. This can also be considered an indicator of the awareness of the California consumers who raised complaints after checking their credit reports.

```
+-----+--------------+
|State|TotComplaints|
+-----+--------------+
|   CA|         1791|
|   TX|          773|
|   NY|          578|
|   IL|          456|
|   GA|          399|
|   WA|          395|
|   FL|          385|
|   PA|          325|
|   NC|          309|
|   VA|          293|
+-----+--------------+
```

In line with the financial market issue ie the Equifax Inc credit data breach during the September 2017, the majority of the complaints in the period are for the credit reporting product and the various issues in the credit card reporting such as improper use of the report, incorrect information on the report and problems with the credit reporting company's investigation into an existing problem. This goes to prove how a closer analysis of the consumer complaint database can be used to identify the real financial marketplace problems.

Evaluation

The detailed research into the Consumer Complaints Database to identify trends and risk signals was done to understand the various ways in which the public access to the database can be useful to the three primary stakeholders – Consumers, Financial Institutions and the CFPB and other watchdog organizations.

In the text below, we map the various findings as benefits to the consumers, financial institutions and watchdog agencies

1. Consumers
    a. Findings about the top ten companies with complaints as well as the products offered by these companies help the customer be aware of the 'good' and 'bad' service providers in the industry.
    b. Data about the resolution of the complaints (with monetary relief, explanation etc.), company's response to the complaint as well as the timely manner of the resolution can be used to benefit the consumer when he/she is filing their own complaint.

2. Financial Institutions
    a. Data in the consumer database does prove that most of the complaints are resolved with an explanation or non-monetary relief, which is indicative that the consumers often misinterpreted or misunderstood the services. This is a winning point for the financial institutions who have been using this argument to shut the public access to the complaint database.
    b. Most companies nowadays are looking at consumer reviews to improve their products and services. The consumer narratives in the Consumer Complaint Database can be used by companies to better their services and products.
    c. By mapping the products complaints to the companies, the financial institutions can also observe the weaknesses of their competitors and capitalize to increase the market share.

3. Consumer Financial Protection Bureau & other watchdog organizations
    a. We cannot use the growing trend of the complaints to judge if the public access of the Consumer Complaint Database has improved the products and services provided by the financial institutions. This is because the complaints could be increasing due to various reasons
        i. Increasing awareness of the Consumer Complaint Database
        ii. Growth in the customer base of the financial institutions
    b. Information on the top ten companies, products and issues in the Consumer Complaint Database allow the CFPB and other watchdog organizations identify which areas need additional focus.
    c. Analysis of the complaint trends across the years also allow the CFPB to identify some of the good practices of the companies. For eg. Ocwen Loan Serving LLC is the only company amongst the top ten companies by complaint volume which has reduced the number of complaints. This could be a result of steps taken by the company to reduce the consumer grievances and provide better services.
    d. Frequently occurring problems like the mortgage complaints can help the CFPB make better laws and regulations to reduce the grievances of the consumers.
    e. Closer look at the observations during financial market incidents such as the Equifax Inc Data Breach incident allow the CFPB to gauge the response of the company to such incidents. For eg did the company respond to complaints in a timely manner, how did the company close the complaints – with or without monetary relief. It can also be used to gauge the impact of the complaints across the state and zipcodes and different sectors of the demographic like Elderly consumers or army veterans (by Tags in the consumer complaint database).

Recommendations

The Consumer Complaint Database is a young database, and we may not have realized its full potential given that it has data for only a few years. However, the insights derived from the few years data indicate what a powerful tool this can be not just for the consumers and the watchdog organizations like CFPB but also for the financial institutions.

The main purpose of the CFPB is to make the consumer financial market work for the consumers, responsible providers and the economy as a whole. The Consumer Complaint Database allows the CFPB to accomplish its goals. It not only empowers the consumers and financial institutions to navigate the financial market place, but also educates them with the best practices and performance in the market. It also makes it possible for the agency to identify, correct and prevent malpractices in the industry.

Our recommendation is for the Consumer Financial Protection Bureau to continue to maintain an open and accessible Consumer Complaint Database so that the agency can serve the public and make the financial institutions accountable for their activities in the market place.

Assumptions and Limitations

We understand and accept the following assumptions and limitations of the analysis

1. The analysis does not consider any other data source to verify the quality of the products and services provided by the various companies as indicated in the findings of the analysis. For eg if the findings highlight the mortgage services offered by JP Morgan Chase as having the highest number of complaints, it does not compare with the findings from any other data source to confirm the truth in the finding from the consumer complaint database analysis.
2. In assessing the quality of the products and services provided by the various companies, the analysis assumes that the consumers will only use the Consumer Complaint Database to log their complaints. It does not consider that the consumers may directly approach the financial institutions to log their complaints.
3. The data is limited to less than 10 years and may not necessarily be a good indicator to use the database for further analysis and forecasting.
4. The CFPB does not verify any of the complaints logged in the database, and for all the complaints that do not have a public response from the company – we will never know if the complaint was correct or just misunderstood or misinterpreted services by the consumer.

References

1. www.consumerfinance.gov
2. www.stackoverflow.com
3. www.matplotlib.org
4. Consumer Bureau looks to end public view to the Complaints Database - https://www.nytimes.com/2018/04/25/business/cfpb-complaints-database-mulvaney.html
5. The Consumer Complaints Database that could disappear from view - https://www.npr.org/2018/04/25/605835307/the-consumer-complaints-database-that-could-disappear-from-view
6. Equifax Data Breach affects millions of customers. Here's what to do - https://www.lifelock.com/learn-data-breaches-equifax-data-breach-2017.html

Appendix

SQL Spark queries used as part of the analysis

Refer to the attached ipynb file for all the SQL queries.

- SELECT Company, count(1) AS TotComplaints FROM Complaint_df

  GROUP BY Company

  ORDER BY count(1) DESC LIMIT 10

- SELECT Product, count(1) AS TotComplaints FROM Complaint_df

  GROUP BY Product

  ORDER BY count(1) DESC LIMIT 10

- SELECT Issue, count(1) AS TotComplaints FROM Complaint_df

  GROUP BY Issue

  ORDER BY count(1) DESC LIMIT 10

- SELECT Company, Product, count(1) as TotComplaints FROM Complaint_df

  WHERE Company IN

  (SELECT Company FROM Complaint_df

  WHERE Company NOT LIKE '%Navient%'

  GROUP BY COMPANY ORDER BY count(1) DESC LIMIT 10)

  AND Product IN

  (SELECT Product FROM Complaint_df

  GROUP BY Product ORDER BY count(1) DESC LIMIT 10)

  GROUP BY Company, Product ORDER BY Company

- SELECT count(1) AS TotComplaints,

  year(CAST(UNIX_TIMESTAMP(Complaint_df.Date_Rxd, 'MM/dd/yyyy') AS TIMESTAMP)) AS Year_Rxd FROM Complaint_df GROUP BY Year_Rxd ORDER BY Year_Rxd

- SELECT Company,

  year(CAST(UNIX_TIMESTAMP(Complaint_df.Date_Rxd, 'MM/dd/yyyy') AS TIMESTAMP)) AS Year_Rxd,

  count(1) as TotComplaints FROM Complaint_df

  WHERE Company IN (SELECT Company FROM Complaint_df GROUP BY COMPANY ORDER BY count(1) DESC LIMIT 10)

AND year(CAST(UNIX_TIMESTAMP(Complaint_df.Date_Rxd, 'MM/dd/yyyy') AS TIMESTAMP)) <> 2013

AND year(CAST(UNIX_TIMESTAMP(Complaint_df.Date_Rxd, 'MM/dd/yyyy') AS TIMESTAMP)) <> 2019

GROUP BY Company, Year_Rxd ORDER BY Company

- SELECT Timely_Resp as Timely_Response,

  year(CAST(UNIX_TIMESTAMP(Complaint_df.Date_Rxd, 'MM/dd/yyyy') AS TIMESTAMP)) AS Year_Rxd,

  count(1) AS TotComplaints FROM Complaint_df

  GROUP BY Timely_Resp, Year_Rxd

  ORDER BY Timely_Resp, Year_Rxd

- SELECT Company, count(1) AS DispCnt FROM Complaint_df WHERE Cnsmr_Disp = 'Yes' GROUP BY Company ORDER BY count(1) DESC

- SELECT DF1.Company, (DF1.DispCnt/DF2.AllCompl) AS Dispute_Rate FROM DF1 INNER JOIN DF2 ON DF1.Company = DF2.Company ORDER BY Dispute_Rate DESC

- SELECT Cnsmr_Disp AS Consumer_Dispute,
  year(CAST(UNIX_TIMESTAMP(Complaint_df.Date_Rxd, 'MM/dd/yyyy') AS TIMESTAMP)) AS Year_Rxd,
  count(1) AS TotComplaints FROM Complaint_df
  GROUP BY Cnsmr_Disp, Year_Rxd
  ORDER BY Cnsmr_Disp, Year_Rxd

- SELECT year(CAST(UNIX_TIMESTAMP(Complaint_df.Date_Rxd, 'MM/dd/yyyy') AS TIMESTAMP)) AS Year_Rxd,
  count(1) AS Monetary_Relief FROM Complaint_df
  WHERE Comp_Resp_Cnsmr = 'Closed with monetary relief'
  GROUP BY Year_Rxd
  ORDER BY Year_Rxd

- SELECT Company_Pub_Resp, count(1) AS TotComplaints from Complaint_df
  WHERE Company_Pub_Resp <> ' '
  GROUP BY Company_Pub_Resp ORDER BY count(1) DESC

- SELECT Submit_Via, count(1) as TotComplaints FROM Complaint_df
  GROUP BY Submit_Via
  ORDER BY count(1) DESC

- SELECT Product, Company, count(1) AS Complaints FROM Complaint_df
  WHERE Product IN
  (SELECT Product FROM Complaint_df GROUP BY Product ORDER BY count(1) LIMIT 1) GROUP BY
  Product, Company ORDER BY Complaints DESC
- SELECT Product, Company, count(1) AS Complaints FROM Complaint_df
  WHERE Product IN
  (SELECT Product FROM Complaint_df GROUP BY Product ORDER BY count(1) DESC LIMIT 1) GROUP BY
  Product, Company ORDER BY Complaints DESC