A Study on Financial Product Consumer Complaints

-Mid-term Report

Prepared by: Shuojia Shi

Contact: shishuojia1987@gmail.com

June 19, 2018

DEFINE THE PROBLEM

Financial companies rely on their products and services to generate profit and strive. One of the key factors is the satisfaction of their customers. Without data, it is difficult to measure and compare the level of satisfaction and expose areas needed to be improved. The Consumer Financial Protection Bureau (CFPB) has been accepting consumer complaints of financial products and services since 2011. These complaints are sent to companies for response. Consumer complaint data could be a great way to spot apparent issues and understand trends in customers' needs. It could be a supporting evidence when analyzing customers' satisfaction level toward a product or service. It can also give insight in ways to improve their products and customer satisfaction.

The consumer complaints data also accept dispute over the response from company. Whether a complaint was disputed is recorded in the same data set. Can we use machine learning models to predict whether a company response of a complaint will be disputed based on the data set? If the top contributions to disputes can be found, companies can take proactive means to lessen the dispute rate. This adds another perspective into promote customer service and satisfaction.

IDENTIFY THE CLIENT

- **FINANCIAL COMPANIES:**

  Use this report to find insight into problems that people are experiencing, help identify inappropriate practices. The trends in the timing and location of the complaints would be valuable information for companies hoping to proactively improve their performance by keeping customer happy.

- **STATE AND FEDERAL AGENCIES:**

  Take the report data as an indicator of the quality of financial products by the companies. Could provide evidence for some market regulations.

- **CUSTOMERS OF FINANCIAL PRODUCTS:**

  Consider this report as a reference for choosing a company for a certain financial product or service.

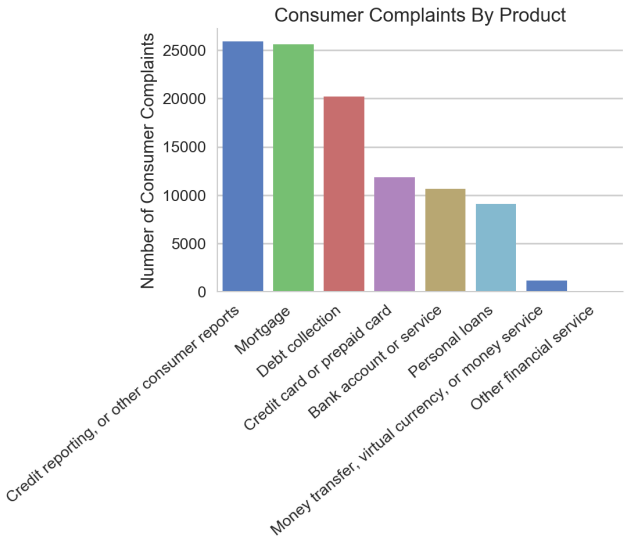
DATA WRANGLING

The financial customer complaint data can be downloaded on the CFPB website, consumer complaint database: https://www.consumerfinance.gov/data-research/consumer-complaints/#download-the-data
The data set contains up-to-date (of this study) record of customer complaints since 2011. Its size is 560.9 MB. The table explaining all the features of the data can be found here: https://cfpb.github.io/api/ccdb/fields.html

After the close examination of the data, I noticed some columns have different input with the same meaning entries. Take the "Product" feature as an example, the top several entries were shown in the table:

Product	Complaints Count
Mortgage	25629
Debt collection	20180
Credit reporting	13947
Credit reporting, credit repair services, or o...	11963
Credit card	9049
Bank account or service	8534
Student loan	4429
Consumer Loan	3004
Credit card or prepaid card	2487
Checking or savings account	2135
Vehicle loan or lease	616

some entries such as credit card, credit card or prepaid card overlap in contents. A close look at the product VS year of complaints revealed that the consumer complaints submission form was changed in 2017 and some product choices have been combined. Both old and new choices of products were put into the final data set and caused the initial confusion. To solve this, I combined the older choices of products using the new sets and stored the new product data in another column. The value counts of the updated product list is shown in the below plot. We



see the most complained products over the years are credit reporting or other consumer reports, mortgage and debt collection.

OTHER POTENTIAL DATASETS

The zip code data in the consumer complaints could be combined with the Rural-Urban Continuum Codes data here:

https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx

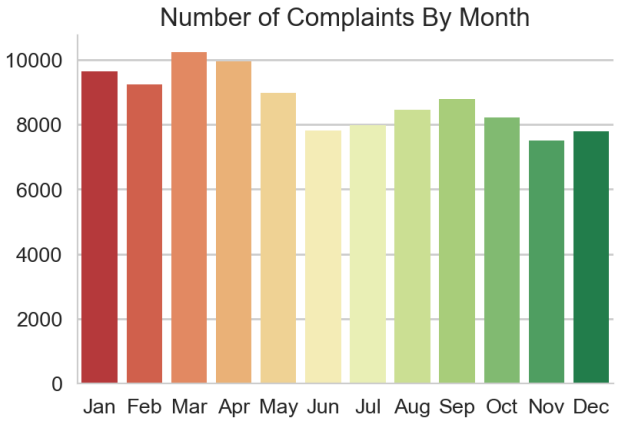
There could be some pattern to be found when the metro-nonmetro of the complaints are identified.

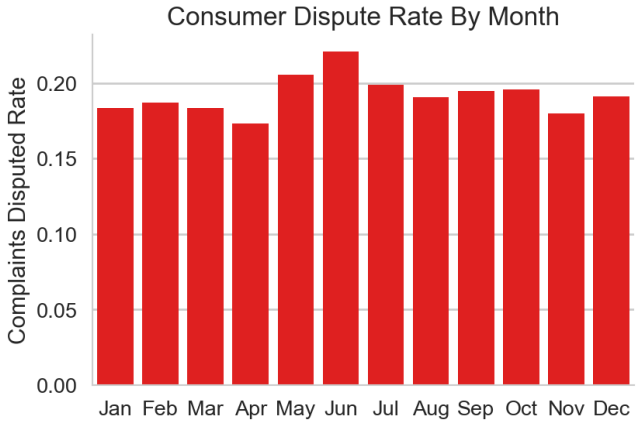
KEY FINDINGS

TIME EFFECT

Here we take a look at how does month affect the number of consumer complaints received by CFPB and the rate of the company response being disputed by consumer:

We notice that March, April, January are the months with most complaints while less complaints were received in June, November

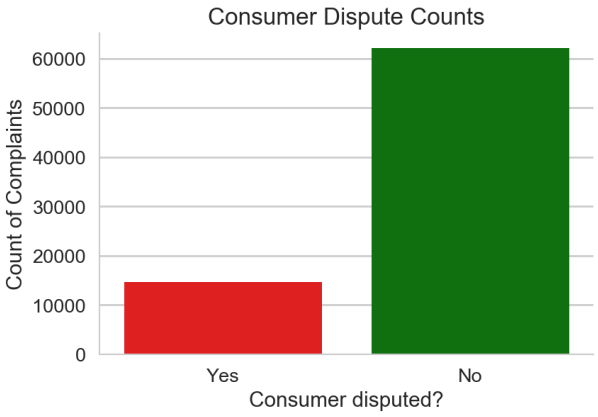


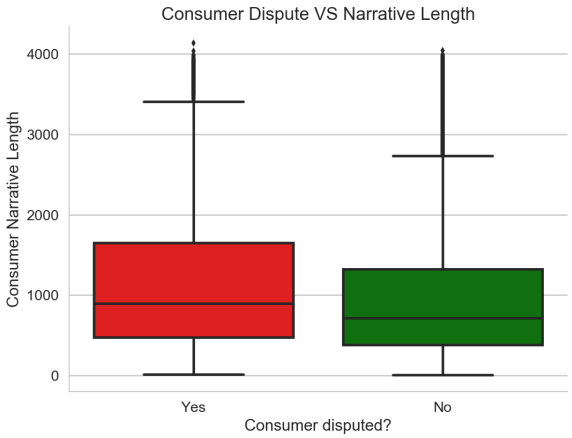


and December. On the contrary, the complaints received in June is mostly likely to be disputed by almost 20%. This is very interesting and require some more study in explaining why.

INITIAL FINDING ON THE CONSUMER NARRATIVES

On the left plot below, we see about 19% of the consumer complaints were disputed based on the response from the companies. The box plot on the right shows the distribution of the length of the consumer narrative for both disputed and not disputed complaints. The median narrative length for the “Yes” group is 28% longer than the “No” group! This insightful finding implies the longer the consumer narratives are, the more possible that the narratives will be disputed. Consumer narratives could help classify the consumer dispute. It gives some confident in the natural language processing (NLP) that I will proceed to explore in the next phase of this study.

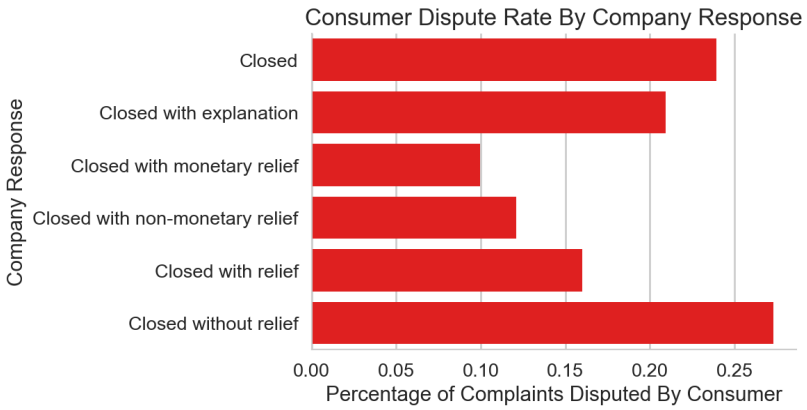




CONSUMER DISPUTE RATE VS COMPANY RESPONSE

There are several responses companies can choose from to report the outcome of the response to the customers. They are listed in the plot below together with the rate of the reply being disputed. It is not difficult to notice that the response of “closed without relief” receives the highest dispute rate while “closed with monetary relief” gets less than half of the dispute rate. This makes perfect sense since consumers tend to dispute when their complaint could not be relieved by the company. In the mean time, when the consumer is

monetarily compensated, they tend to be happier and less likely to dispute. It is also noticeable that “close with non-monetary relief” gives the second to last probability to be disputed. And all top three most disputed replies are non-relief replies. It seems being able to receive some relief on the complaints from the company, monetary or not, keeps consumers happier.



SUMMARY

This is the mid-term report of my capstone project 2: A Study on Financial Product Consumer Complaints. I have re-iterate the problem (motivation) and suggested the clients who could be interested in this project. The steps taken for data wrangling in this study was explained. Among the results, some keys EDA findings were listed: there is a trend in the time and the number of complaints received by the CFPB. It seems that the months when less complained were filed get the higher dispute rate. We also see a clear difference in the quartiles of the length of the consumer narratives for disputed or not disputed groups. The consumer is more likely to dispute when the narrative is long. Many other trends were found at this stage, such as how the company response, product, issue affects affect whether a dispute will follow. It gives a deep understanding of the data set and encourages the future machine learning work.

There are still aspects left to study: combining the metro data and the zip code column for feature engineering, feature selections, machine learning, NLP and problems solving and etc. They will be included in the final report of this project.
Find the code on my Github: https://github.com/shuojiashi/capstone_project_2/blob/master/consumer_complaints.ipynb