CFPB Complaints Case Study

Dishank Gaglani

New York University

Business Analytics

Professor Mukul Pareek

April 30, 2022

Table of Contents:

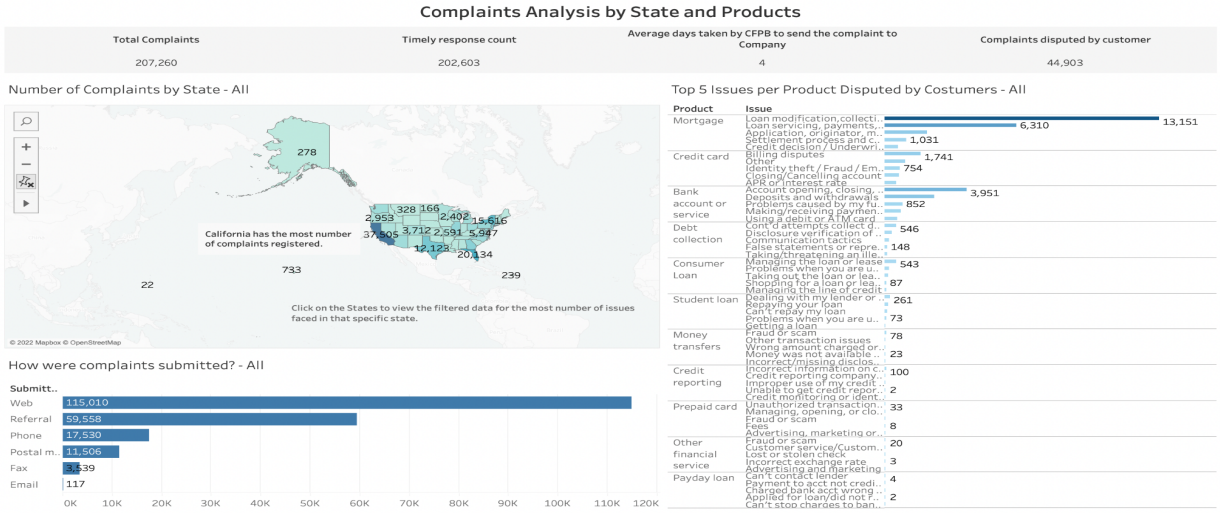
Abstract

The Consumer Financial Protection Bureau (CFPB) is a government agency in the United States that ensures that banks, lenders, and other financial institutions treat their consumers fairly. The Bureau of Consumer Financial Protection's principal role is to collect, investigate, and respond to consumer complaints. Consumers file complaints with the Consumer Financial Protection Bureau (CFPB) via the online, phone, mail, fax, and other methods, describing the financial product or service with which they have a problem, as well as the nature of the problem. This gives data that the Consumer Financial Protection Bureau (CFPB) can use to investigate complaints. The Consumer Financial Protection Bureau (CFPB) directs consumer complaints about financial products and services, as well as any documents they give, to financial companies. Consumers can then opt to challenge the company's resolution. There are 5 major banks, namely, Bank of America, Wells Fargo, JP Morgan Chase, CitiBank, U.S. Bankcorp on whose data we will performing our analysis on.

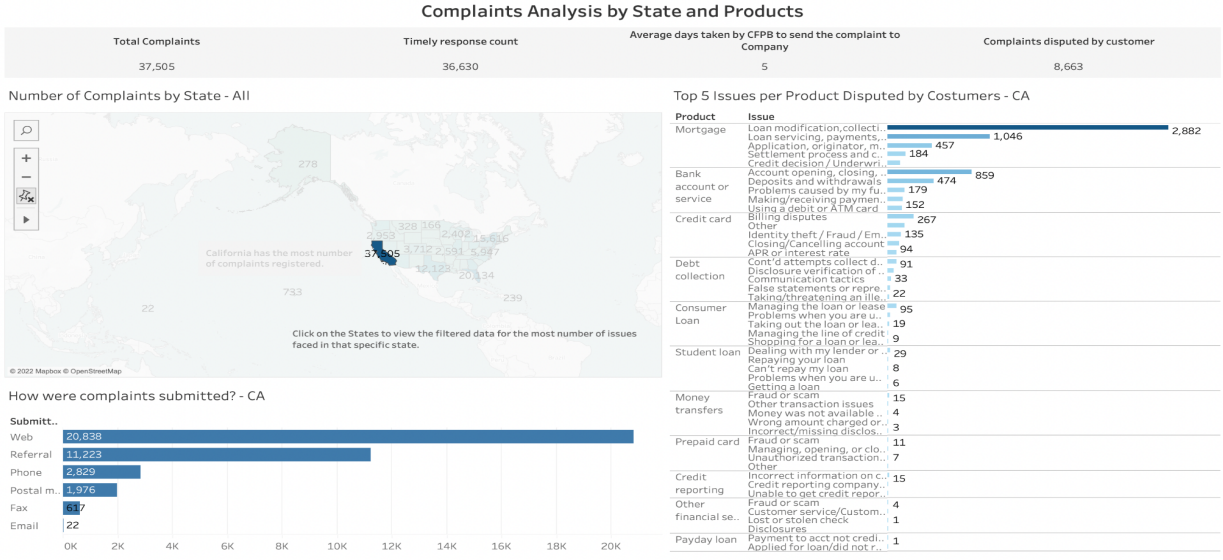
I will be using data from 2010 to 2017, which shows a total of **207,260** complaints filed with the Consumer Financial Protection Bureau (CFPB) through various channels, with the Web being the most popular. There were **44,903** complaints in all that were disputed. My primary goal is to analyze the data and trends in order to develop a model that can assist them in identifying complaints that will result in a dispute, as well as to develop and evaluate a predictive model that can assist the banks in identifying future disputes so that they can perform "**extra diligence**" during the first round of dealing with the complaint in order to avoid future disputes. If a complaint is selected for "extra diligence," the bank will be charged an additional **$90** to complete the further diligence processes. But if they correctly spot a future dispute in a timely way, they can avoid spending the extra **$1500** later to resolve the dispute.

Analysis

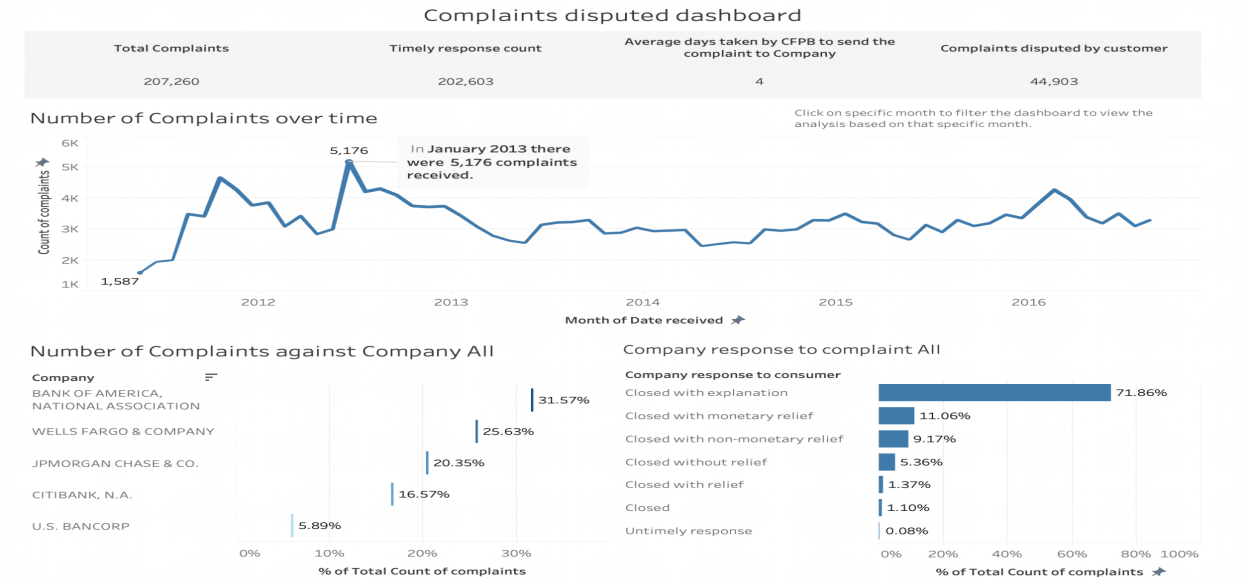
For my analysis, I will be using visualizations and dashboards made on Tableau Software to help understand the situation of complaints like where are they coming from, how have they being dealt with, how many were disputed, how many complaints received timely response, which financial product received most number of complaints and which method did the consumer use to lodge complaints.



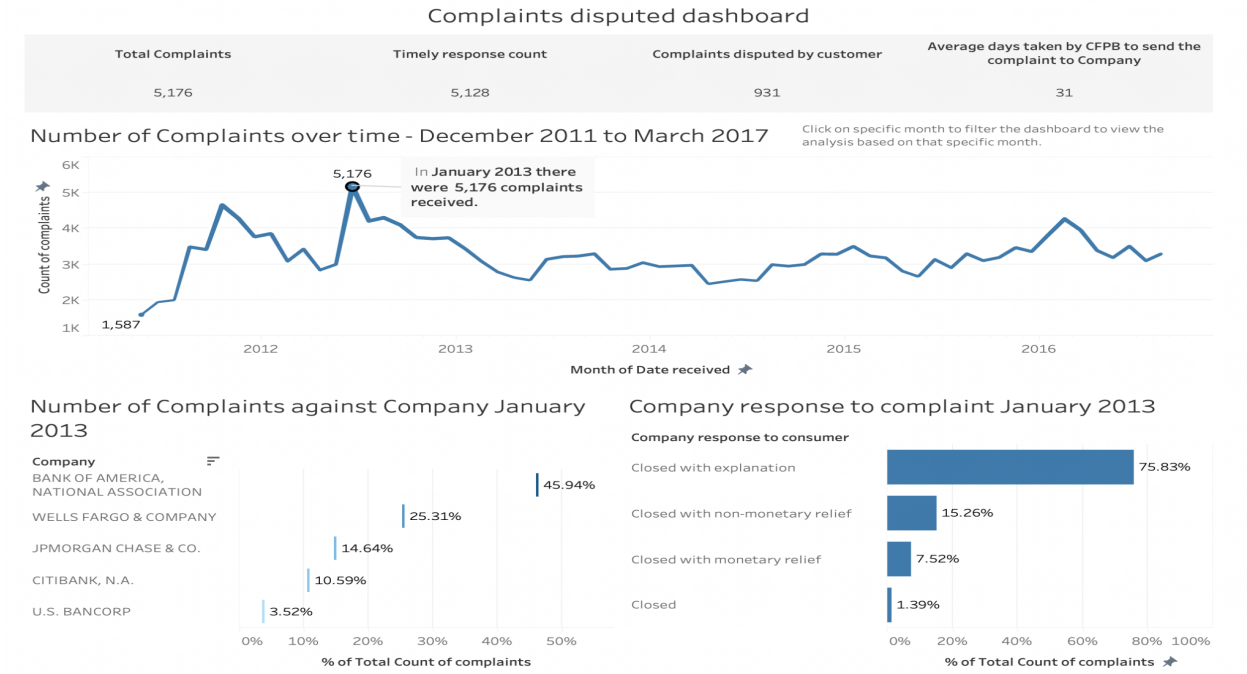
From this dashboard, I wanted to show the total number of complaints per state, financial product and what was the issue that the consumer wanted to report and how did they report it. I have filtered the data to show only Top 5 issues for each financial product to drive attention towards the most important issues faced. I have also added filter to the Map which showed that California had the most number of complaints and analyzing the states situatuion.



The following dashboard has a time series line chart to show the number of complaints received during a certain time period (month). The second graph shows the total percentage of complaints received against each bank compared to the overall percentage of complaints received. The time series graph also acts as a filter to view the other charts filtered to the specific month selected. This dashboard also shows the banks dealt with the complaints and what kind of responses were mostly used for complaints.



The next dashboard has been filtered by January 2013, which was also the month when most of the complaints were lodged. It is evident that the number of complaints against the Bank of America was the highest. Complaints disputed were **931** in that month.



Key insights from the dashboards:

- Number of Complaints peaked in the month of January of 2013 with complaint count being **5176** for that month.
- Out of 5 banks, from 2010 to 2017, Bank of America has the highest number of customer complaints with almost **32%** of total complaints lodged. This number increased to **46%** of total complaints during Januray of 2013.
- Analyzing the complaints, California has highest number of complaints with most complaints coming for the Mortgage financial product.
- The Web has always been used as a major platform by customers to lodge their complaints.

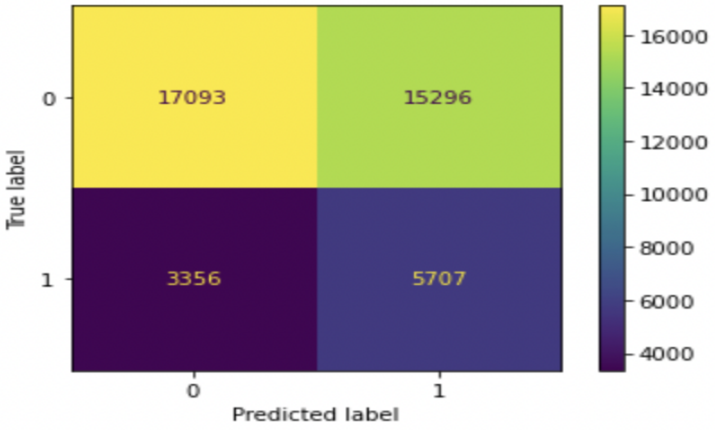
These insights derived from the dashboards can help the management to figure out where to focus their focus such as which states need more attention, which product department needs to work on their issues more and figure out how to get the complaints resolved without monetary relief and less disputes.

Predictive Model:

The Data from the CFPB, is being cleaned and preprocessed for the model. The XGBoost model from XGBoost classifier is used to model the data. Cost of extra diligence is very low ($90), but if a complaint reaches the dispute stage, the cost can get quite high ($1500).

The below confusion matrix is from before adjusting the threshold and the total cost from the model is **$6,410,640**. Also, the first model has low recall and good precision. So the goal is to reduce false negatives from our predictive model, even if it means we push up false positives. That is because the cost of false positives is quite low compared to the cost of false negatives.

	precision	recall	f1-score	support
0	0.84	0.53	0.65	32389
1	0.27	0.63	0.38	9063
accuracy			0.55	41452
macro avg	0.55	0.58	0.51	41452
weighted avg	0.71	0.55	0.59	41452

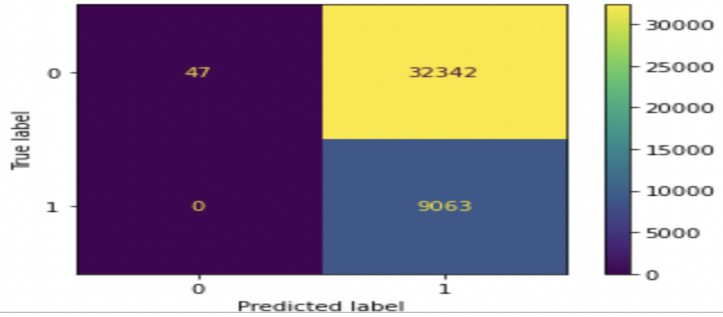


After adjusting the threshold to 0.1, we find out the total cost from analysis to go downby more than 50% to $2,910,780. After tuning the model, we also achieve high recall and low precision which proves that model is optimal for the business problem at hand and can be helpful in minimizing loses.

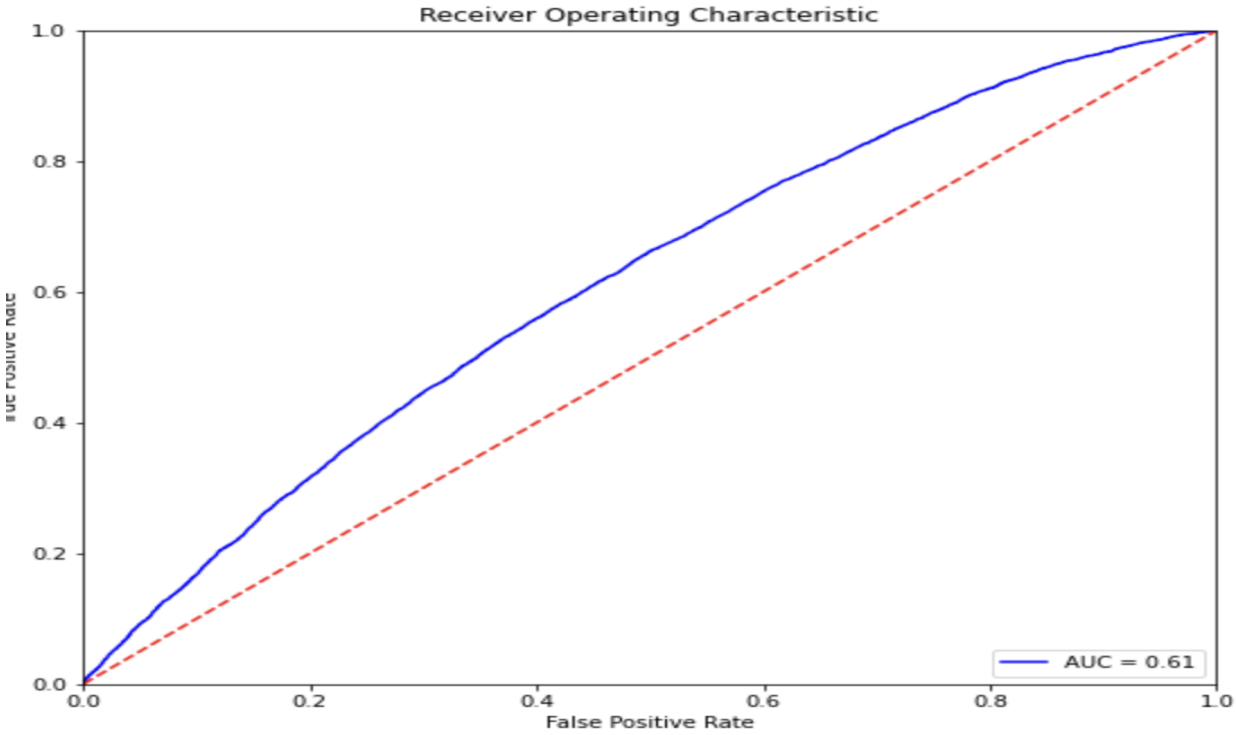
```
Confusion Matrix :
 [[    47 32342]
 [     0  9063]]
Test accuracy =  0.21977226671813183
                precision    recall  f1-score   support

           0       1.00      0.00      0.00     32389
           1       0.22      1.00      0.36      9063

    accuracy                           0.22     41452
   macro avg       0.61      0.50      0.18     41452
weighted avg       0.83      0.22      0.08     41452
```



AUC is used to measure how well the model performs and at the different different thresholds we have different AUC curve values. So, to improve the performance of the model we use different thresholds to classify more positives or negatives. As visible from the graph, we have AUC at 61 which proves that the model is performing well.



Result:

This algorithm would be extremely accurate in predicting likely disputed customers, and it would save the Big Banks' Association millions of dollars each year by offering extra diligence to the likely disputed customers in order to avoid disputes.