CFPB Consumer Complaints Analysis

Drishti Gandhi – Dhg9054

MG-GY 8413 Business Analytics

Prof. Mukul Pareek

Contents

- Summary
- Introduction
- CFPB Background
- Dataset
- Goal of the project
- Execution & Key Insights
  i) Consumer Disputed Analysis
  ii) Timely Response Analysis
  iii) Company Response to Consumer Analysis
  iv) Top 5 Issues & Sub – Issues of CFPB Complaints
  v) Products and Sub – Products of CFPB Complaints
- Predictive Model
- Conclusion

SUMMARY

This report will consist of a detailed analysis of the Consumer Complaints data, which is collected by the Consumer Financial Protection Bureau, for the “Big Bank Association” consisting of the five largest banks in the United States. Goals of this report is to highlight and summarize the significant trends and findings from the submitted complaint data. Additionally, developing and accessing a predictive model that banks can use to anticipate disagreements, resolve complaints right away, and reduce costs by minimizing future disputes.

Introduction

Consumer Financial Protection Bureau (CFPB) is the organization in charge of overseeing consumer protection in the financial industry. Banks, credit unions, securities companies, payday lenders, mortgage-servicing businesses, foreclosure relief services, debt collectors, and other American-based financial businesses are subject to the oversight of the Consumer Financial Protection Bureau (CFPB).

CFPB Background

Since its beginnings, the CFPB has employed technological tools to track how financial institutions target consumers using social media and algorithms.

Mortgages, credit cards, and student loans are the Bureau's top three concerns. The Federal Reserve, the Federal Trade Commission, the Federal Deposit Insurance Corporation, the National Credit Union Administration, and even the Department of Housing and Urban Development all contributed staff and responsibilities to the CFPB, which qualifies as a sizable autonomous agency.

The bureau is a standalone division housed inside the US Federal Reserve and is temporarily affiliated with the US Treasury Department.

Dataset

The CFPB's website has the data we will be working on. The data, which comprises over 2 million current anonymized records and spans 6000+ financial firms of all stripes, is available to the public. It includes a prose narrative of the complaint among other elements.

Goal of the project

The dataset includes details about the customer's interactions with various products and related sub-products. We have information in written text format about the problem they were having and the specifics of their complaint. I have made an effort to illustrate the data and derive some fundamental observations based on the data.

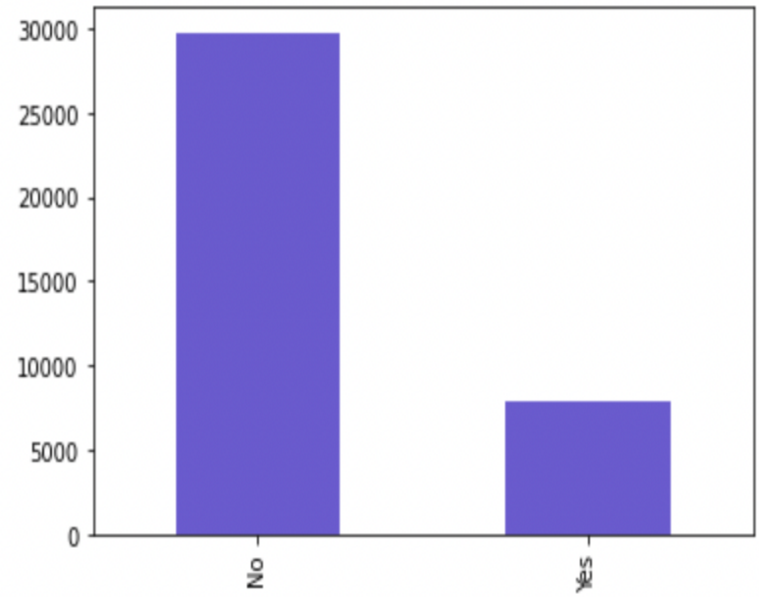
Execution and Key Insights

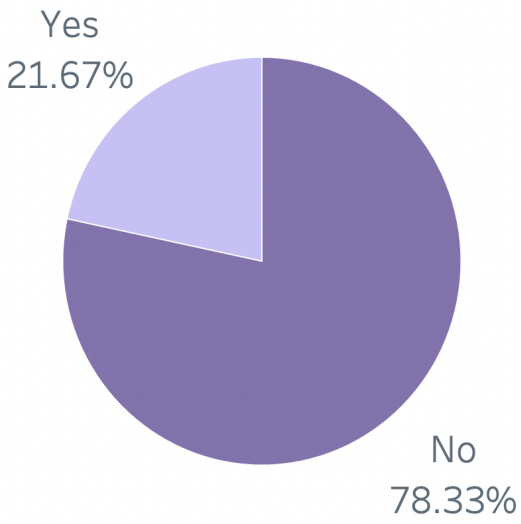
- Consumer Disputed Analysis:

```
No      162357
Yes      44903
Name: Consumer disputed?, dtype: int64
```

```
complaints['Consumer disputed?'].value_counts().plot(kind='bar',color='slateblue')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f461783a0a0>
```

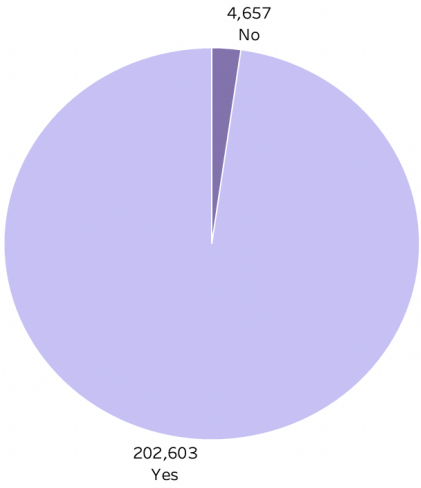




This basic graph shows us the percentage of customers who contested and did not dispute the company's handling of their complaint. As we can see from the chart, 78.3% of the customers (or 162,357 in slateblue) did not dispute the issue again. While the percentage of issues not being disputed is higher than those that were (21.6%), the number of customers disputing the rate is still high. It Even though these numbers appear positive, banks still need to lower their dispute rate (the proportion of customers who were dissatisfied with the company's answer).

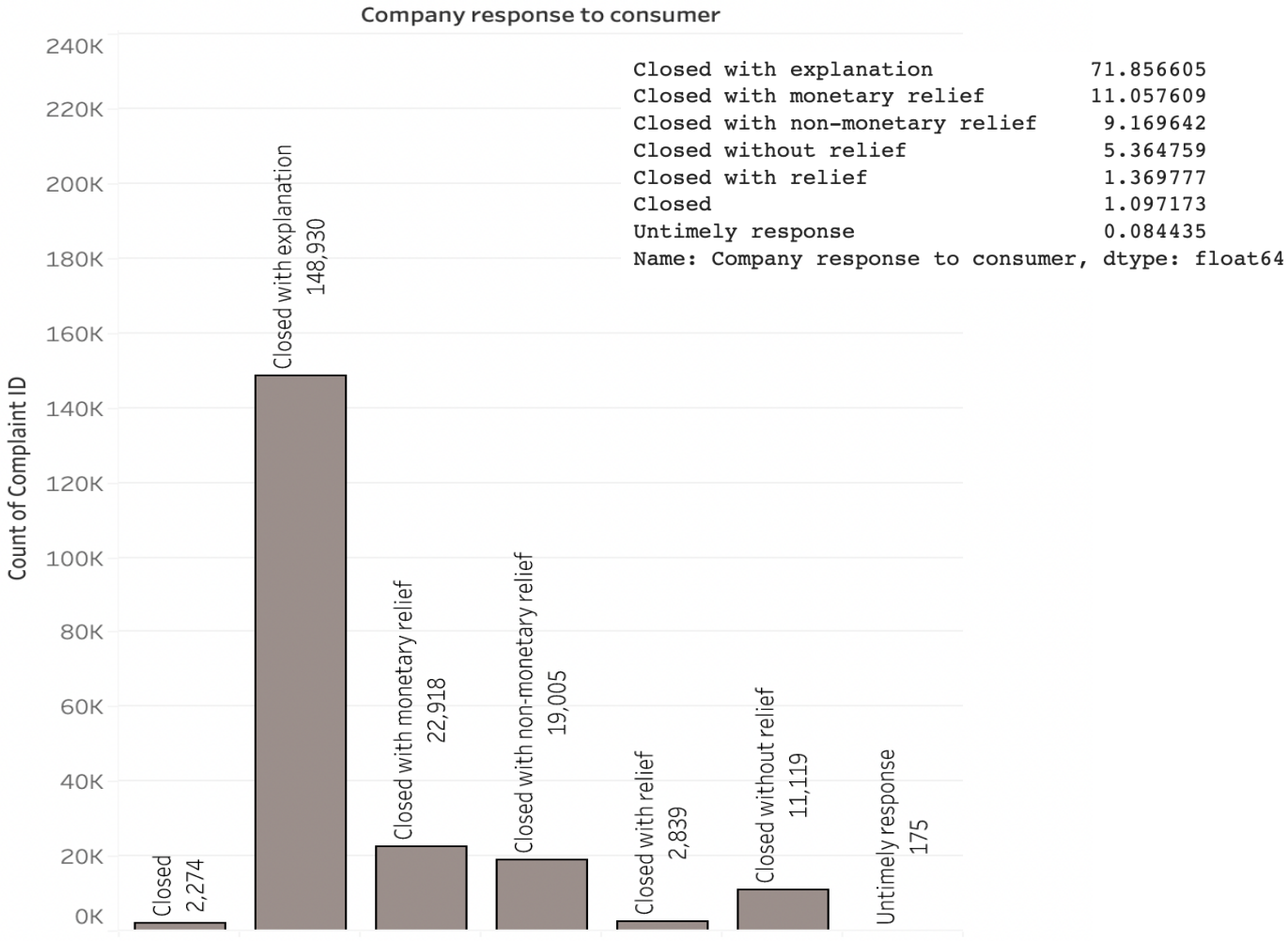
- Timely Response Analysis:

```
Yes     97.753064
No       2.246936
Name: Timely response?, dtype: float64
```



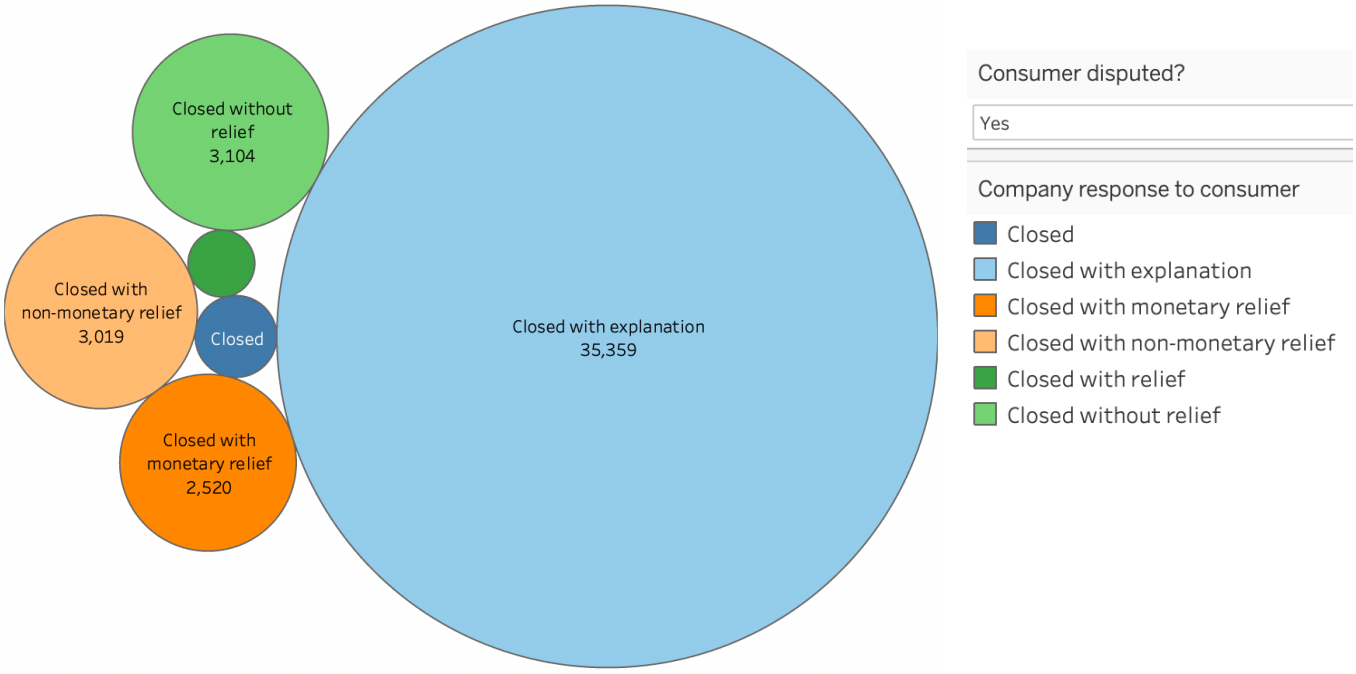
The way a complaint is handled is a crucial part of understanding the problem areas. As we can see in the chart above 97% of the customers said they received a timely response, which is a good sign. If we look at the Company's response to the consumers, 71.8% cases were closed with an explanation and only 0.08% cases were closed with 'Untimely no response', which is again somewhat a good sign.

- Company Response to Consumer Analysis:

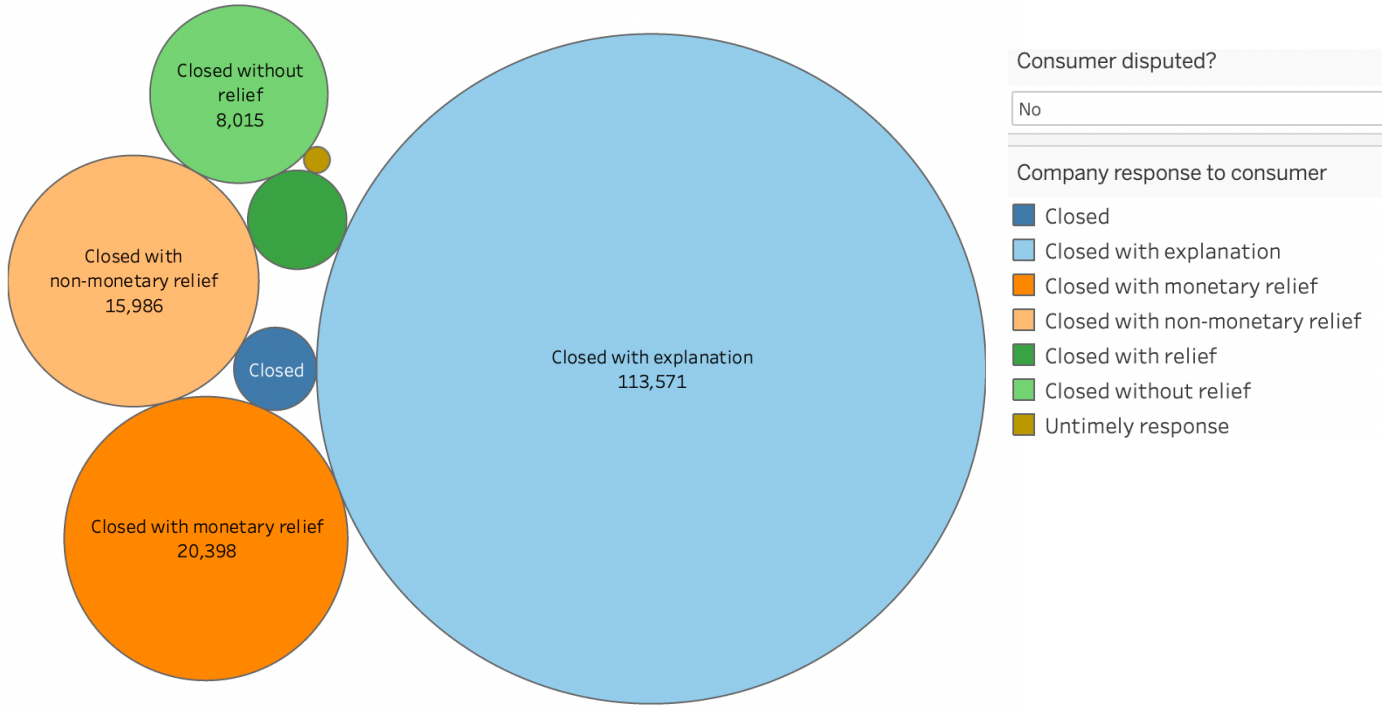



```
Closed with explanation            71.856605
Closed with monetary relief        11.057609
Closed with non-monetary relief     9.169642
Closed without relief               5.364759
Closed with relief                  1.369777
Closed                              1.097173
Untimely response                   0.084435
Name: Company response to consumer, dtype: float64
```

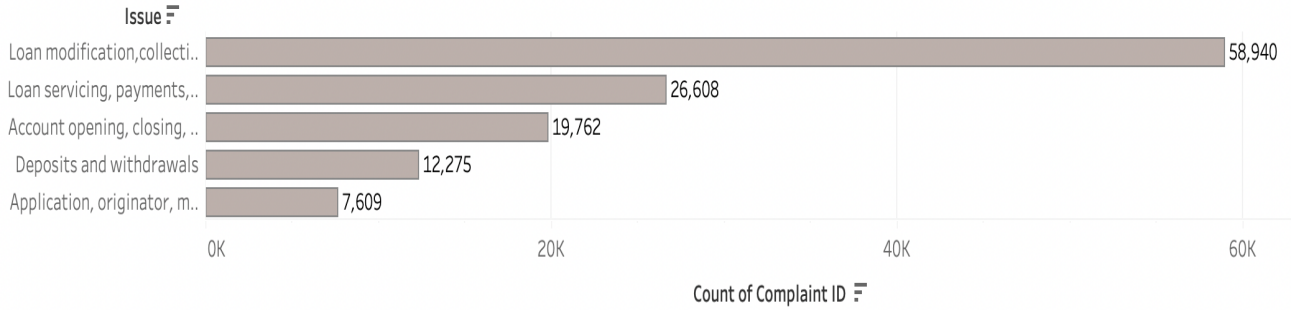
1. Company Response to Consumer when consumer response is YES



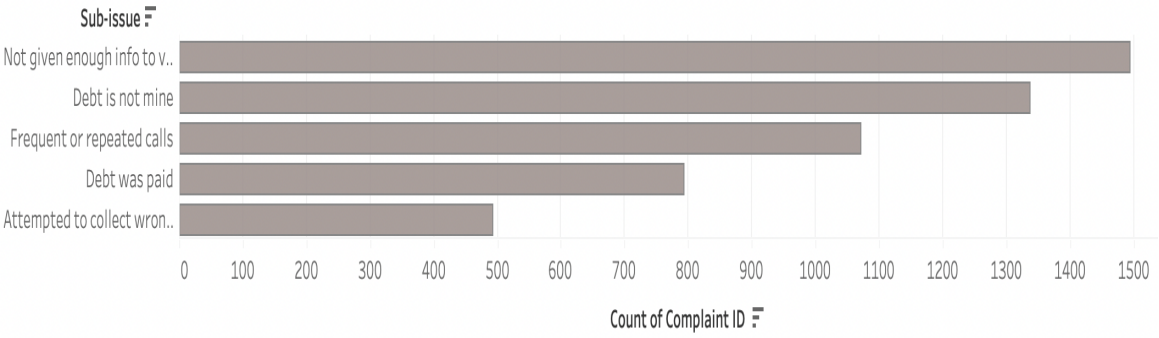
2. Company Response to Consumer when consumer response is NO



- Top 5 Issues & Sub – Issues of CFPB complaints



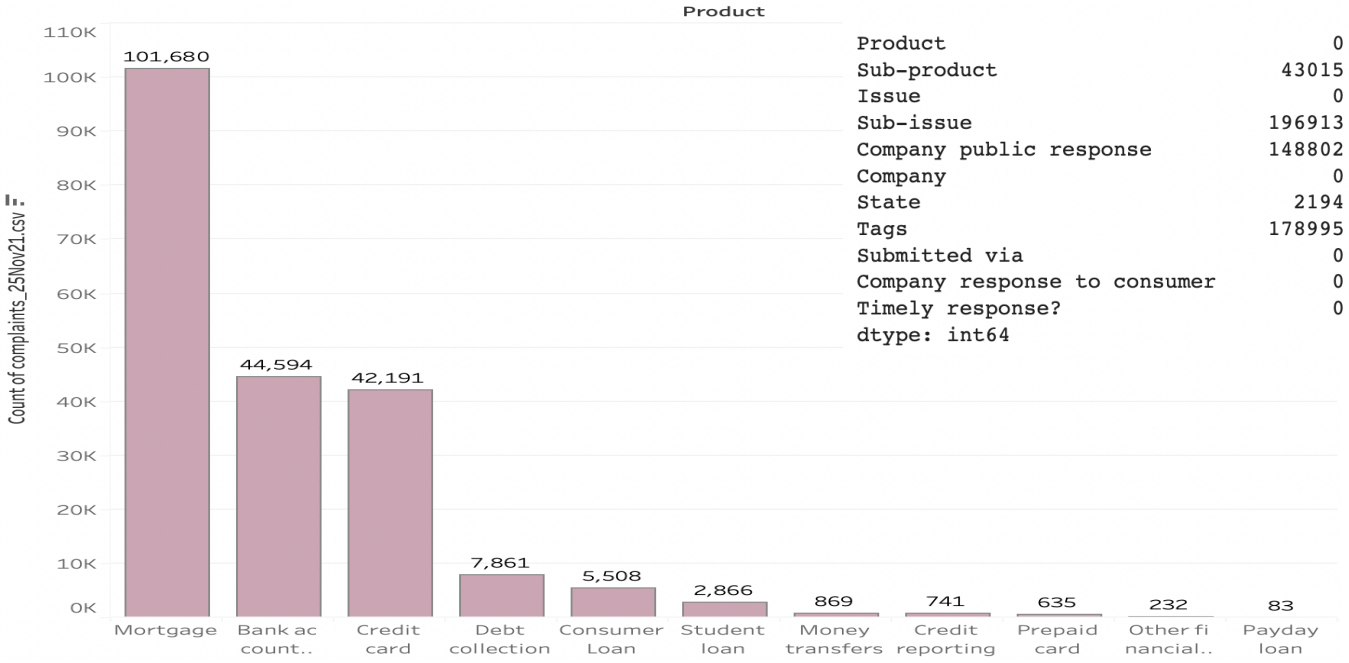

```
Loan modification,collection,foreclosure    28.437711
Loan servicing, payments, escrow account    12.837981
Account opening, closing, or management      9.534884
Deposits and withdrawals                     5.922513
Application, originator, mortgage broker     3.671234
Name: Issue, dtype: float64
```

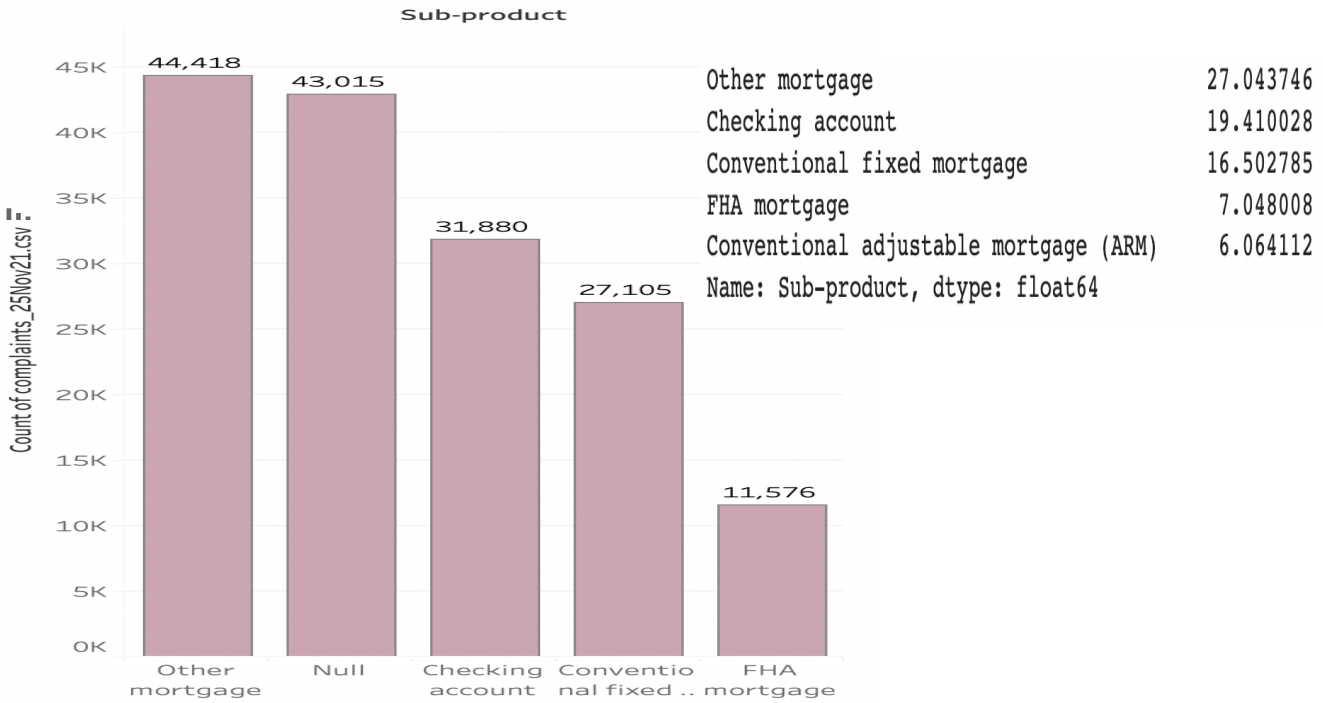



```
Not given enough info to verify debt    14.429303
Debt is not mine                        12.921620
Frequent or repeated calls              10.350826
Debt was paid                            7.664057
Attempted to collect wrong amount        4.764666
Name: Sub-issue, dtype: float64
```

Based on the bar chart above, the Top 5 issues and sub-issues were reported based on feedback from customers. In order for these banks to succeed, they need to focus on these issues as soon as possible, to minimize the possibility of a dispute by the customer. We can see in the bar above, 'Loan modification, collection, foreclosure' is the top issue (28.4% of all complaints) and 'Not given enough information to verify debt 'is the top sub-issue (14.4% of all sub-issue complaint reasons reported) that these banks need to work on.

- Products & Sub-Pproducts of CFPB Complaints





The sort of products and sub-products the customer mentioned in the complaint are depicted in the bar graph above. Mortgage, followed by bank accounts or other services, is the product that had the most problems when it first came out, as can be seen in the graph above. In order to decrease the amount of complaints received, these institutions should investigate the problems that consumers are having with these products (most particularly mortgages) and strive to fix them. Interestingly, four out of the five sub-product categories are found under the product "Mortgage" in the sub-products bar chart.

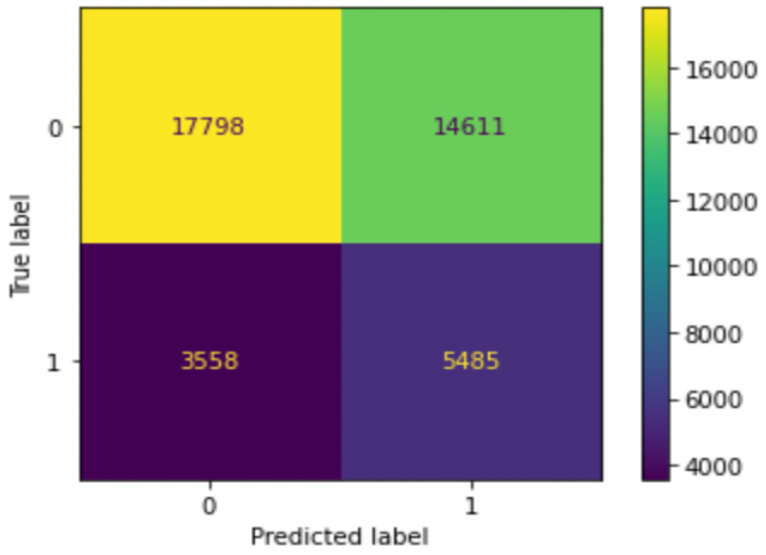
Predictive Model - Fitting the Model to XGBoost

One of the goals of this research was to develop and assess a predictive model that can be used to assist banks in foreseeing future conflicts, allowing them to take all necessary actions in the initial round of responding to customer complaints and so preventing future disputes. It cost the banks $100 to settle a grievance in the first round. If the complaint is disputed, the bank must pay $90 to initiate the case and an extra $1500 to settle this newly disputed complaint.

By forecasting dispute rates using Python's XGBoost Machine Learning, the following prediction model can assist banks in controlling their complaint-related expenditures. The banks will be able to identify the concerns with the highest likelihood of being disputed with the use of this model.

The y variable in this model is "Consumer disputed?" products, sub-products, issues, sub-issues, company public reaction, company, State, tags, customer consent given, submitted via, company response to consumers, and prompt response are the x factors.

	precision	recall	f1-score	support
0	0.83	0.55	0.66	32409
1	0.27	0.61	0.38	9043
accuracy			0.56	41452
macro avg	0.55	0.58	0.52	41452
weighted avg	0.71	0.56	0.60	41452



Since the cost of dealing with false negatives is the largest, I changed the threshold value to have a lower false negative from our predictive model in order to have a better recall for category 1. With a threshold of 0.3, we have a higher true positive rate and a lower false negative rate.

```
threshold = 0.2
```

```
pred_prob = model.predict(X_test)
pred = (pred_prob>threshold).astype(int)
cm = confusion_matrix(y_test, pred)
print ("Confusion Matrix : \n", cm)
print('Test accuracy = ', accuracy_score(y_test, pred))
```

```
Confusion Matrix :
 [[   96 32313]
 [    9  9034]]
Test accuracy =  0.2202547524848017
```

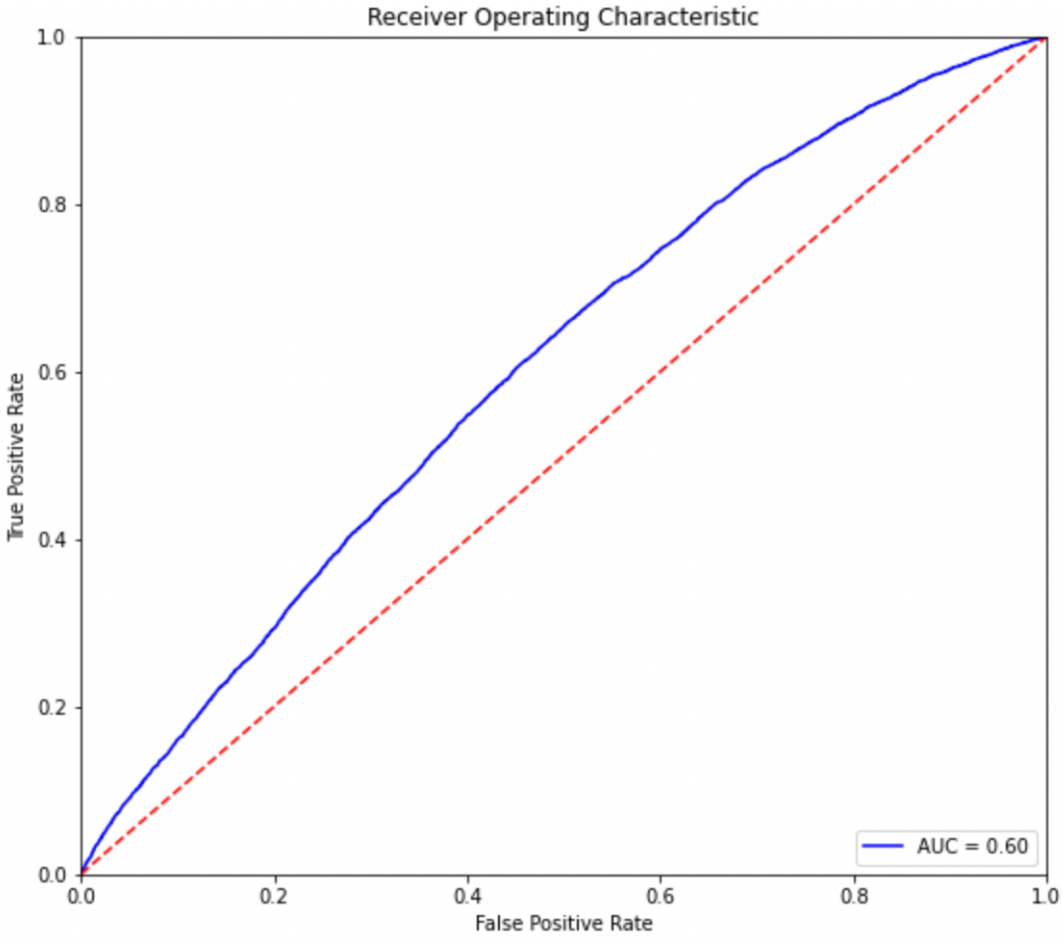
```
print(classification_report(y_true = y_test, y_pred = pred))
```

```
              precision    recall  f1-score   support

           0       0.91      0.00      0.01     32409
           1       0.22      1.00      0.36      9043

    accuracy                           0.22     41452
   macro avg       0.57      0.50      0.18     41452
weighted avg       0.76      0.22      0.08     41452
```

Area Under the Curve is,



Conclusion

By varying the threshold value I was able to obtain a relatively better recall value. Since the cost of addressing false negative complaints is the highest, a better recall value is essential.