Complaints Analysis Report
(Final Project)

Tae Young Moon

Management of Technology, NYU Tandon School of Engineering

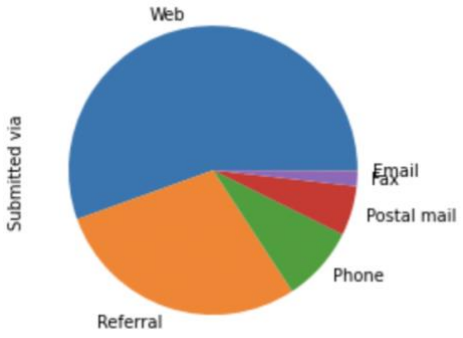
MG-GY: Business Analytics

Mukul Pareek

December 3, 2021

Introduction

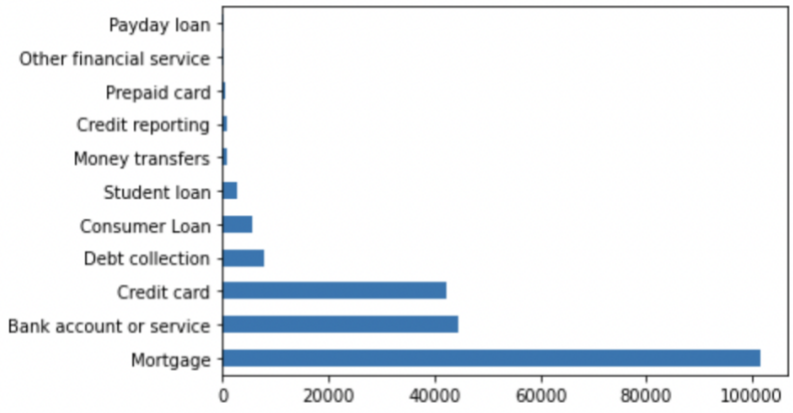
This report includes the analysis of the complaints for “Big Banks’ Association”, which is the five largest banks in the United States. This analysis used the consumers' complaint data provided by the CFPB (Bureau of Consumer Financial Protection). The goal of this analysis is to identify trends and articulate critical issues by showing the validated number and graph. Also, it will include the accuracy of disputed rates from consumers. The pie chart down below indicates that more than 55% of complaints data were submitted via Web, and 28% were submitted via Referral.



Web	0.554907
Referral	0.287359
Phone	0.084580
Postal mail	0.055515
Fax	0.017075
Email	0.000565

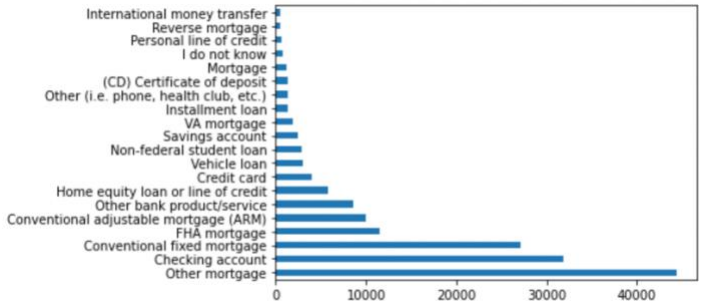
Product Analysis

As you can see in the bar chart down below, the highest number of product use is Mortgage, and it is 49.05% in the overall product. The second-highest number is Bank account or service, which occupies 21.51% of the overall product. As you can see Mortgage is significantly higher than any other product. On the other hand, Money transfers, Credit reporting, Prepaid cards, Other financial services, and Payday loans were less than 1% according to the data, which is very low. Therefore, it would be better to exclude those low-rate products and focus on valuable products to increase revenue.



Mortgage	0.490592
Bank account or service	0.215160
Credit card	0.203566
Debt collection	0.037928
Consumer Loan	0.026575
Student loan	0.013828
Money transfers	0.004193
Credit reporting	0.003575
Prepaid card	0.003064
Other financial service	0.001119
Payday loan	0.000400

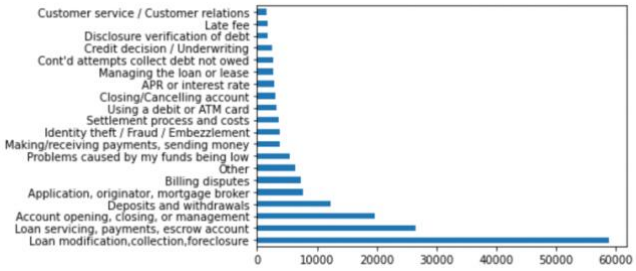
This bar chart down below represents the sub-product, and other mortgages is the highest value at 27%. The checking account occupies 19.41% of the sub-product. Other mortgages, checking account, and Conventional Fixed mortgage are significantly higher than other sub-products. The bar chart down below shows the sub-product.

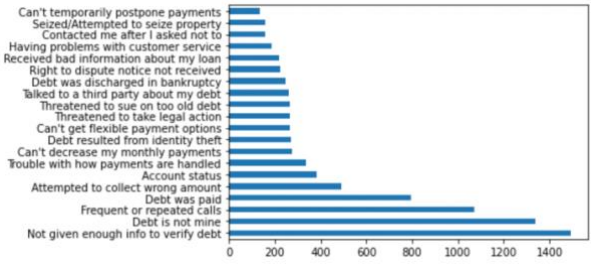


Other mortgage	0.270437
Checking account	0.194100
Conventional fixed mortgage	0.165028
FHA mortgage	0.070480
Conventional adjustable mortgage (ARM)	0.060641
Other bank product/service	0.051904
Home equity loan or line of credit	0.035630
Credit card	0.024713

Issues and Sub-issue analysis

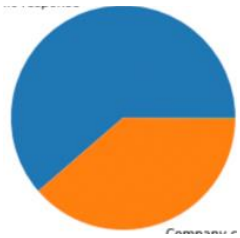
These two bar charts down below show the highest 20 issues and sub-issue from the consumers. The highest issue from the consumers is “loan modification, collection, foreclosure”. And the second-highest issue is the “Loan servicing, payments, escrow account”. For the sub-issues, “Not given enough info to verify debt” & “Debt is not mine” are the highest complaint. The Big banks need to focus on reducing and solving these issues to satisfy customers because those issues are significantly higher than any other issues.





The chart down below indicates the number of company public responses to issues. As you can see “Company has responded to the consumer and the CFPB and chooses not to provide a public response” and “Company Chooses not to provide a public response” are 99% of public response. However, “Company Chooses not to provide a public response” is too high (38%) as a service for the consumer. It is recommended to reduce those numbers.

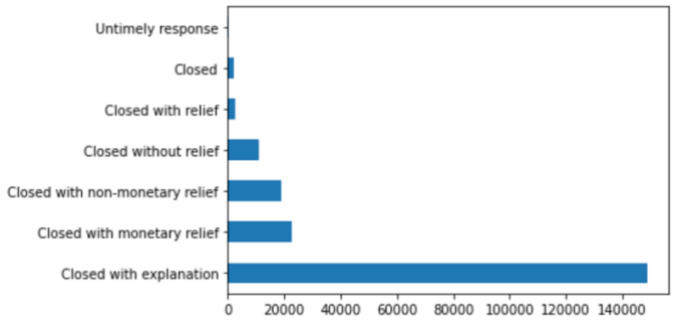
```
Company has responded to the consumer and the CFPB and chooses not to provide a public response                    35858
Company chooses not to provide a public response                                                                   22535
Company believes it acted appropriately as authorized by contract or law                                             58
Company believes complaint caused principally by actions of third party outside the control or direction of the company      3
Company believes complaint represents an opportunity for improvement to better serve consumers                        1
Company believes complaint is the result of an isolated error                                                         1
Company believes complaint relates to a discontinued policy or procedure                                              1
Company believes the complaint is the result of a misunderstanding                                                    1
```



Company has responded to the consumer and the CFPB … =	0.613398
Company chooses not to provide a public response. =	0.385490

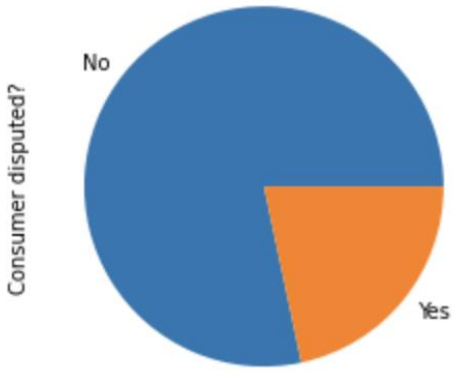
Consumer Analysis

This consumer analysis includes how the company response to consumer, timely response, dispute rate, and accuracy. First, as you can see on the bar chart down below, the highest rate of response to consumer is “closed with explanation” at about 71%. On the other hand, “Untimely response” rate is very close to 0%, which is great.




```
Closed with explanation            0.718566
Closed with monetary relief        0.110576
Closed with non-monetary relief    0.091696
Closed without relief              0.053648
Closed with relief                 0.013698
Closed                             0.010972
Untimely response                  0.000844
```

This pie chart down below shows the consumer disputed rate. “No” is 78.33%, and “Yes” is 21.66%. The rate of not disputed is significantly higher than disputed rate. However, 21% of disputed rate is very high since it represents 2 consumers out of 10 disputed about the service. The cost of dealing with disputes can be expensive. Therefore, it is necessary to minimize the disputed rate.



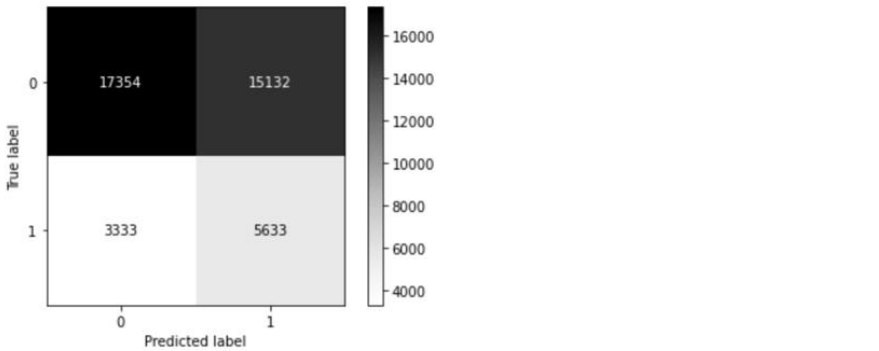

```
No     0.783349
Yes    0.216651
```

Predictive Model

This predictive model will help the banks keep their complaint-related costs low. It is possible to use a predictive model for predicting disputed rates by using XGBoost machine learning algorithm in Python. The data is split into 80% and 20% split for training and testing sets. The picture down below represents the classification report and confusion matrix. The label mapping 0 means 'No' and 1 means 'Yes'. According to the classification report and the confusion matrix, the recall rate for 1 (True Positive Rate), meaning 'Yes, disputed', is 0.66, which means it can help identify 63% of the disputed complain. However, as you can see the recall rate for 0 (False positive Rate), meaning 'No, not disputed' is 0.53, which means it can help identify 53% of the not disputed complain. By using this model, the banks can identify complaints that will end in a dispute. This model sets the 'Consumer disputed?' column as a target, y variable, and the rest of the columns as an X variable. It will be possible to predict the disputed rate and reduce the extra cost for resolving a complaint if it has been disputed.

	precision	recall	f1-score	support
0	0.84	0.53	0.65	32486
1	0.27	0.63	0.38	8966
accuracy			0.55	41452
macro avg	0.56	0.58	0.52	41452
weighted avg	0.72	0.55	0.59	41452

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f6f31eec760>
```



The model was tried to improve the true positive recall rate by changing threshold. As I changed the threshold value from 0.5 to less, I could observe that the true positive rate increases, but it pushes down the false positive number. As the balance given the costs are symmetric $1500 and $90, the true positive rate of 0.94 will be ideal number due to the ratio.

```
# Set threshold for identifying class 1
threshold = 0.34
```

```
pred_prob = model_xgb.predict_proba(X_test)
pred_prob = pred_prob[:,1]
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 :
 [[ 5148 27282]
 [  539  8483]]
Test accuracy =  0.3288381742738589
```

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

	precision	recall	f1-score	support
0	0.91	0.16	0.27	32430
1	0.24	0.94	0.38	9022
accuracy			0.33	41452
macro avg	0.57	0.55	0.32	41452
weighted avg	0.76	0.33	0.29	41452

Lastly, using this true positive rate and false positive rate, we can get the 0.61 AUC (Area Under the Curve) score.