BUSINESS ANALYTICS

Customers' Complaint Analysis of Top 5 Banks

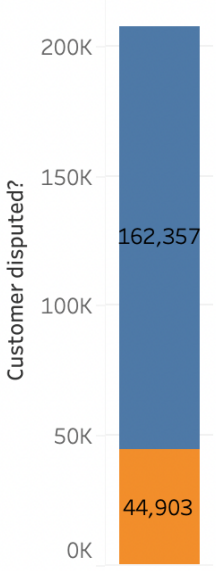
Sunny Advani

Sa6349

Introduction

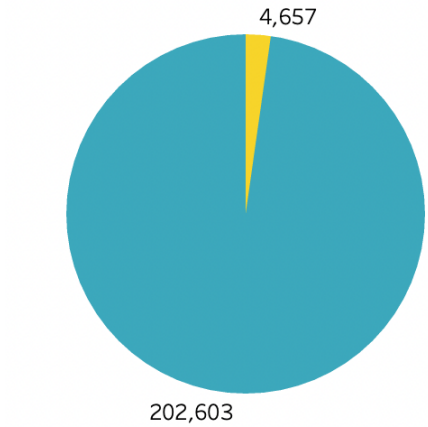
The report has the analysis of the Customer Complaints data, which is collected by the Consumer Financial Protection Bureau, for the “Big Bank Association” consisting of the five largest banks in the United States. Goals of this report are to identify and describe the trends and key insights from the complaints data provided. Also, creating and evaluating a predictive model that can be used bythe banks to identify the future disputes, solve the complaint in the first round and cut back on their expenses by avoiding eventual disputes.

Key Insights

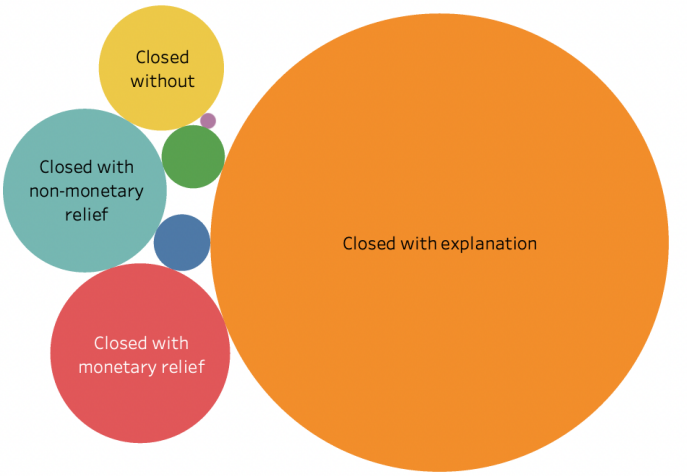


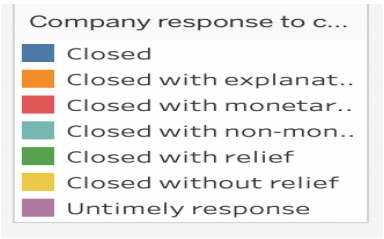

```
No     78.334942
Yes    21.665058
Name: Consumer disputed?, dtype: float64
```

As we can see the chart, out of 207260, 78.3% of them did not dispute the issue again (which is 162,357 customers in the blue). Even through the rate of the issue not being disputed is higher than the ones that were disputed (21.6%), the amount of the customer who disputed the rate is still high. This proves that they need to find a way to cut down on their disputed rate in order to cut down their expenses and better customer satisfaction and retention.



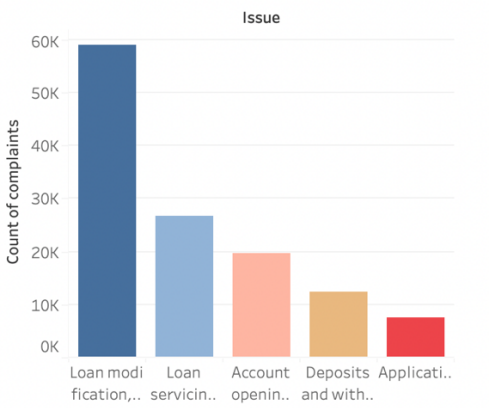

```
Yes    97.753064
No      2.246936
Name: Timely response?, dtype: float64
```



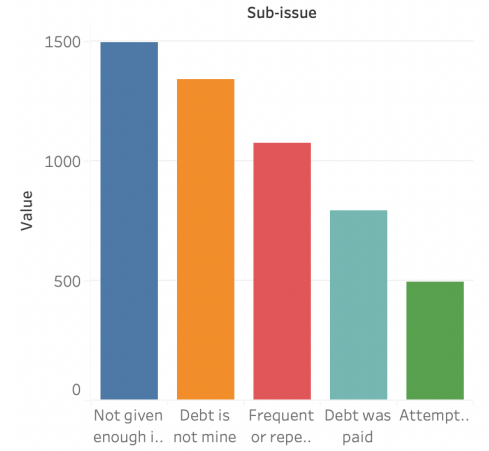



```
Closed with explanation            71.856605
Closed with monetary relief        11.057609
Closed with non-monetary relief     9.169642
Closed without relief               5.364759
Closed with relief                  1.369777
Closed                              1.097173
Untimely response         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

The way a complaint is handles 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.

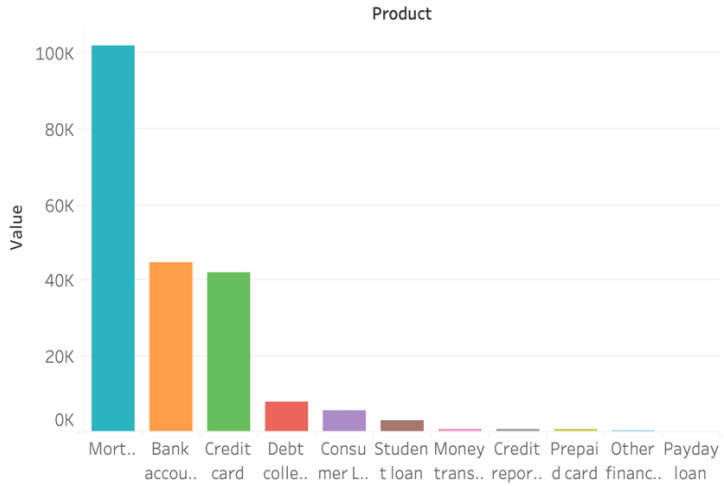



```
Issues
Loan modification,collection,foreclosure    28.437711
Loan servicing, payments, escrow account    12.837981
Account opening, closing, or management      9.534884
Deposits and withdrawals                     5.922513
Application, originator, mortgage broker     3.671234
Name: Issue, dtype: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

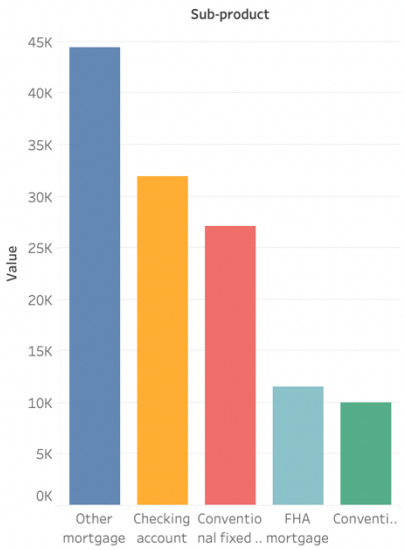



```
Sub-issues
Not given enough info to verify debt    14.429303
Debt is not mine                        12.921620
Frequent or repeated calls              10.350826
Debt was paid                            7.664057
Attempted to collect wrong amount        4.764666
Name: Sub-issue, dtype: float64
```

The bar chart above shows the Top 5 issues and sub-issues, that were reported by the customers. These are the main issues that these banks need to focus on instantly, in order to reduce the chances of a customer to dispute the solution provided to them. We can see in the bar above, ‘Loan Modification, collection, forecast’ 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.




```
Mortgage                   49.059153
Bank account or service    21.515970
Credit card                20.356557
Debt collection             3.792821
Consumer Loan               2.657532
Student loan                1.382804
Money transfers             0.419280
Credit reporting            0.357522
Prepaid card                0.306378
Other financial service     0.111937
Payday loan                 0.040046
Name: Product, dtype: float64
```




```
Other mortgage                          27.043746
Checking account                        19.410028
Conventional fixed mortgage             16.502785
FHA mortgage                             7.048008
Conventional adjustable mortgage (ARM)   6.064112
Name: Sub-product, dtype: float64
```

The bar chart above shows the type of products and sub-products the consumer identified in the complaint. As seen in the graph above, the product that faced the most issues to begin with is ‘Mortgage’, followed by ‘Bank Account or services. These banks should take a deep dive in the issues faced by the customers in these products (most importantly Mortgages) and try to rectify them in order to reduce the amount of complaints inflow. Interestingly, if we see the bar chart of the sub-products, four sub-product categories (out of five) lie under the product ‘Mortgage.’

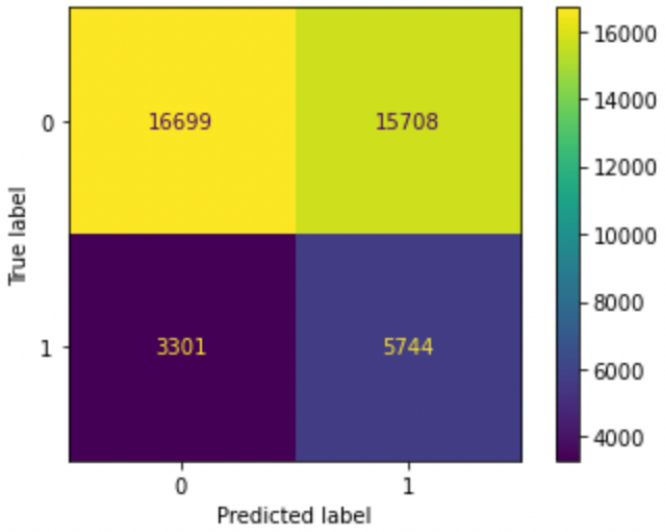
Predictive Model

One of the goals of this report was to create and evaluate a predictive model that can be used to help the banks to identify future disputes so that they can perform all the necessary steps in the first round of addressing to their customer's complaints and in turn prevent eventual disputes. In order to resolve a complaint in the first round, it cost the banks $100. If the complaint is disputed then to open that complain it costs the bank $90 and to resolve this re-disputed complaint, it costs the bank an additional $1500.

The following predictive model can be used to help the banks keep their complaint-related costs low by predicting dispute rates with the means of XGBoost Machine Learning in python. This model will help the banks to pinpoint the issues which has more probability of getting disputed.

In this model, y variable is 'Consumer disputed?' and the x variables are products, sub-products, issue, sub-issue, company public response, company, State, tags, Consumer consent provided, submitted via, company response to consumers and timely response.

	precision	recall	f1-score	support
0	0.83	0.52	0.64	32407
1	0.27	0.64	0.38	9045
accuracy			0.54	41452
macro avg	0.55	0.58	0.51	41452
weighted avg	0.71	0.54	0.58	41452



In order to have a better recall for category 1, I adjusted the threshold value in order to have a reduced false negative from our predictive model since the cost of dealing with false negatives is the highest. The threshold set is 0.3, which gives us an increased true positive rate and a reduced false negative.

```
[104] threshold = 0.30
```

```
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 :
 [[ 2292 30156]
 [  184  8820]]
Test accuracy =  0.26806909196178713
```

+ Code + Text

```
[107] print(classification_report(y_true = y_test, y_pred = pred))
```

	precision	recall	f1-score	support
0	0.93	0.07	0.13	32448
1	0.23	0.98	0.37	9004
accuracy			0.27	41452
macro avg	0.58	0.53	0.25	41452
weighted avg	0.77	0.27	0.18	41452

Area under the curve (AUC), which is calculated by using true positive rate and false positive rate, is 0.62.