

Analyzing the factors causing Rejection Rate

BY PARTHIBAN GOWTHAMARAJ



Contents

Background

Business Objective

About Data

Data Understanding

Data Preparation

Modeling (or Approach)

Business Insights

Conclusions



Background & Business objective

- ✓ Client is one of the major credit card provider in North America.
 - ✓ Launched a new home network private label credit card which is related to home product purchases.
 - ✓ New applicants are accepted online. However, they are facing issue with higher rejection rate.
- Client is interested in understanding what factors led to higher applicant rejection rate.
 - So, this study attempts to understand factors contributed to the rejection rate and provide analytical insights around those factors for business implications.

Data

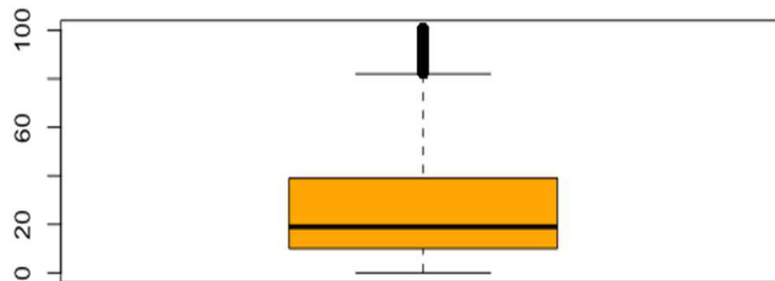
- Client has provided database resources pointing to a million applicants.
- Cust ID, Gender, Income, Applicant Type, State are pulled from FDR.
- Number of Transactions, Number of Intl Transactions are pulled from Consumer Center database.
- Status is pulled from Surveyor database.
- Altogether we received around 20+ columns / variables.
- However, few columns i.e. around 10 are of no use as they are duplicates for the existing, and kept for record purposes.
- And left out with 9 variables for our data analysis.
- Data understanding of the same has been provided in the next slide.

Data Understanding.

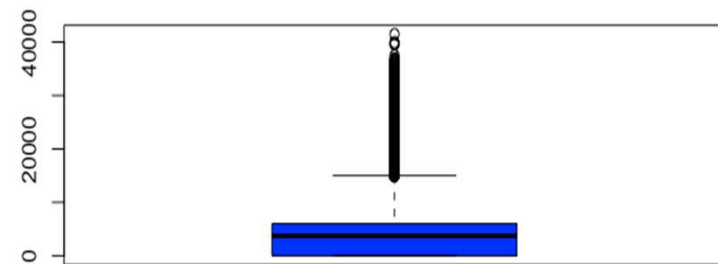
- We have received around 1,000,000 unique applicants data.
- Gender percentage of the applicants is 61.78% belonging to male, against 38.22% belonging to female.
- There are 51 states. Top 5 states contributing more. The states are: 5, 44, 10, 35 and 39.
- 96% of the applications came from Primary cardholders and 4% is from the secondary cardholders.
- Also, understood most of the applications, mentioned their income as zero.
- Also, we understood when integrated with transactions data, Number of transactions and Number of intl transactions are zero then they are tend to get rejected.

EDA

Number of transactions (Box Plot)



Applications Income (Box Plot)



Above, two plots, one for number of transaction coming from transactions data, informs us that on and average customers have done around 20 transactions (median).

Other box plot is related to income, whereas, one can observe that most of the applications didn't mentioned their real income and quoted it as zero.

Data Preparation

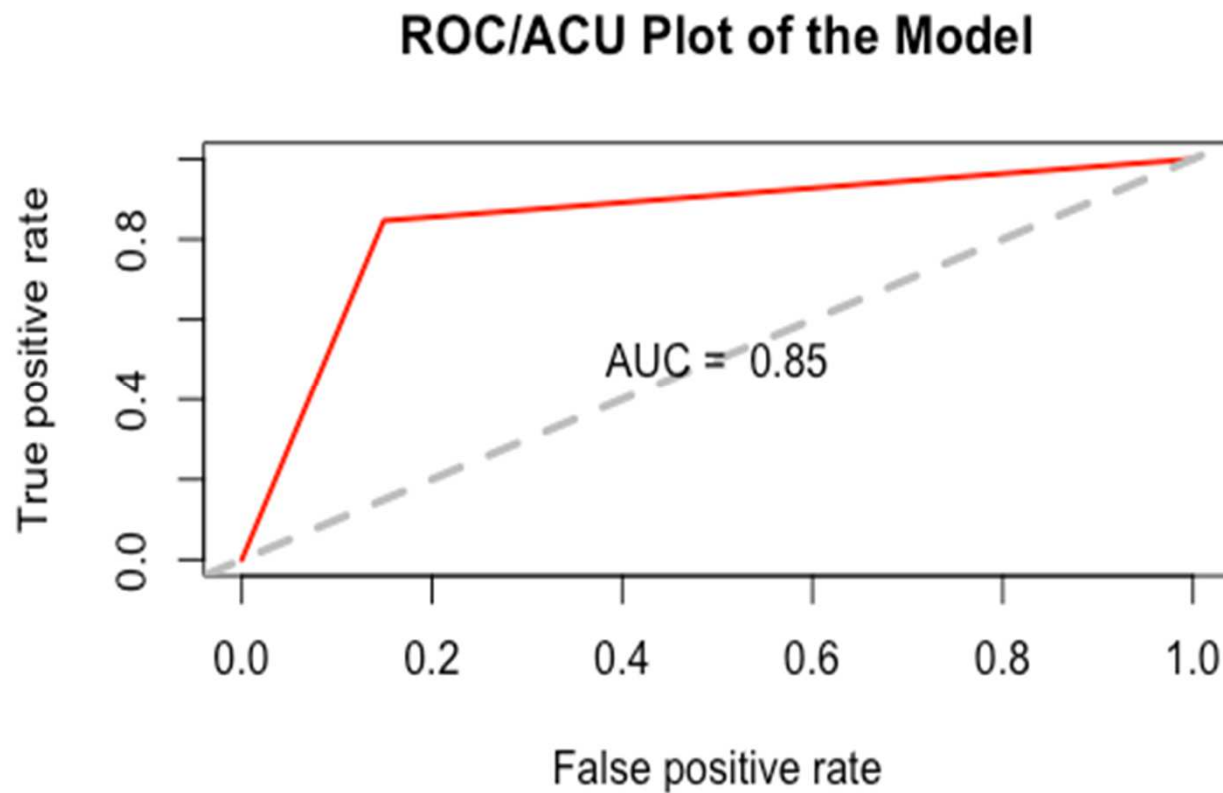
- General data preparation like converting binary nominal characters to binary numeric, for instance, gender and applicant type.
- Further, as we have 51 state values, we retained top states contributing 80% to 90% of the data and converted them to state wise binary dummies for our analysis.
- Next, we have generated binary nominal numeric for income equals to zero for further analysis.
- Also, checked data for outliers, however, found to be prepared data to be good for modelling purposes.
- Also, made sure that target or dependent variable to be binary factor for modelling purposes.



Modeling Approach

- Logistic regression (a variation of linear regression) is used when the dependent variable is a dichotomous variable (i.e. it takes only two values, which usually represent the occurrence or non-occurrence of some outcome event, in our case rejected and accepted) and the independent variables are continuous, categorical, or both.
- Dissimilar to linear regression, logistic regression neither assume existence of linear relationship between the independent variables and the dependent variable nor does it assume that the error terms are distributed normally.
- Thus, logistic regression forms a predictor variable ($\log(p/(1-p))$) which is a linear combination of the independent variables. The values of this predictor variable are then transformed into probabilities by a logistic function. Such a function has the shape of an S as shown to extreme left-below. On the horizontal axis we have the values of the predictor variable, and on the vertical axis we have the probabilities.
- Logistic Regression explain the rate of change of each independent variable (similar to linear regression) , wherein, β (slope of variable) determines the rate of changes of such that if $\beta > 0$ then the predictor increases with increases in x , and if $\beta < 0$ then the predictor decreases with decreases in x . However, obtained betas are converted into exponential form where the values provide more intending explanation.

Modeling Accuracy (80:20)



Results & Conclusion

- We found from our logistic regression that following variables are insignificant, i.e. gender, applicant type and top state 4.
- Variables that are significant are, namely., top state 1, 2, 3 and 5 along with number of transactions, number of internal transactions and income equaling zero.
- Coming to positive contributors i.e. rejection rate, we have state 1, 2, income and number of national and internal transactions.
- So what is driving rejection rate, first the triggering of zero income, and next because most of the applicants came from top 1 state where client was not interested to sell in that state as it already have good amount of customer base for similar type of cards.
- Similarly, we have provided detailed decile analysis on each dependent variable against target that are significant in our analysis.

