

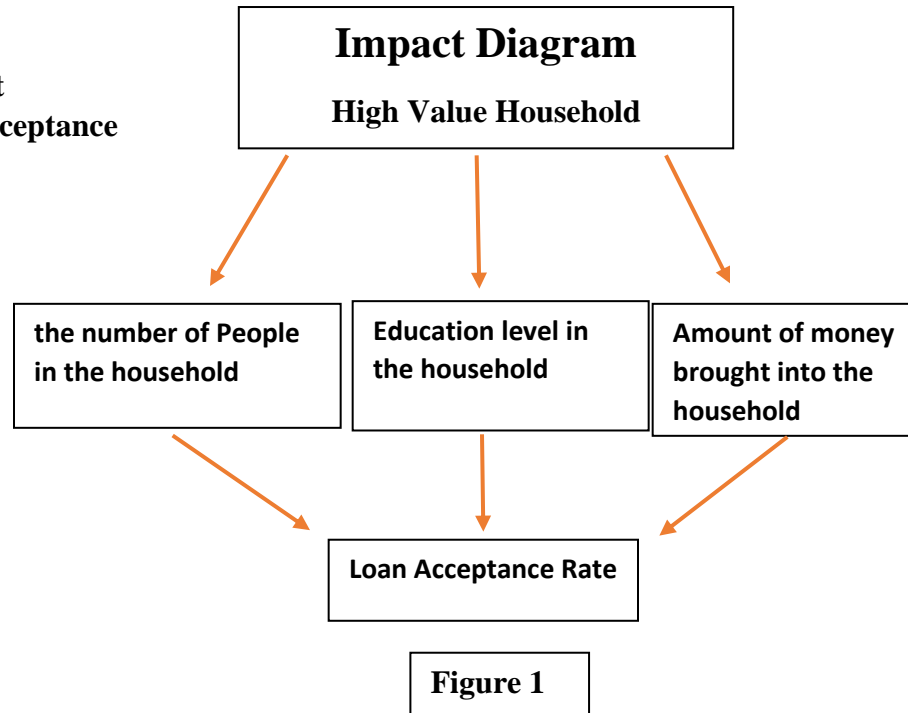
# **Thera Bank**

## **Loan Acceptance: Feature Importance**

**By: Frank Arcuri**

### **Conclusions:**

1. Calculated the most important Features that effect a Loan Acceptance Rate as shown in Figure 1.
2. We are able to predict if a customer will accept the loan based off our model.



### **Business Recommendations:**

1. Based off my results, we recommend that Thera Bank provide lower interest rates first time loaners.
  - a. As show in figure 2, education and income are two main factors with relevance towards accepting a loan. A lower interest rate would allow customers to spend less to get more.
  - b. This in turn will increase the loan department, keep reliable customers, and further expand student loan dependence.
2. Offer families with children to open savings accounts/college funds to increase business and allow for more educated individuals to stick with Thera.
3. Based off models 2-5, Thera should focus on clients that are involved in these features to get the highest chance of accepting a loan.

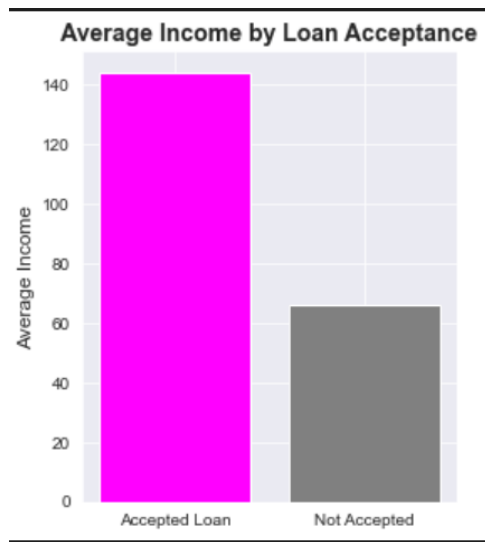


Figure 2

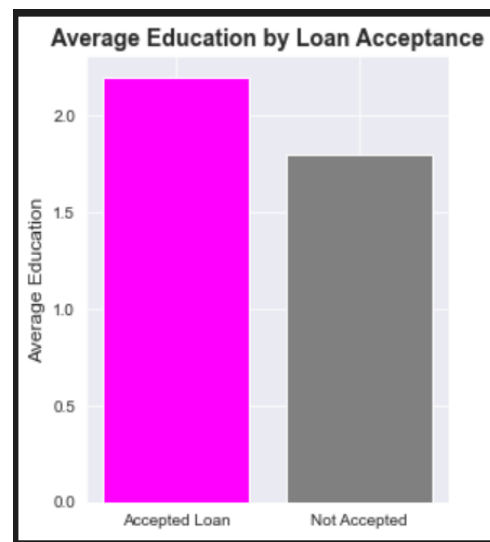


Figure 3

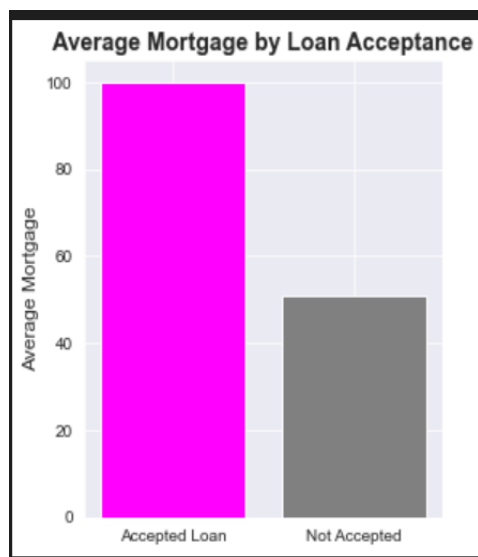


Figure 4

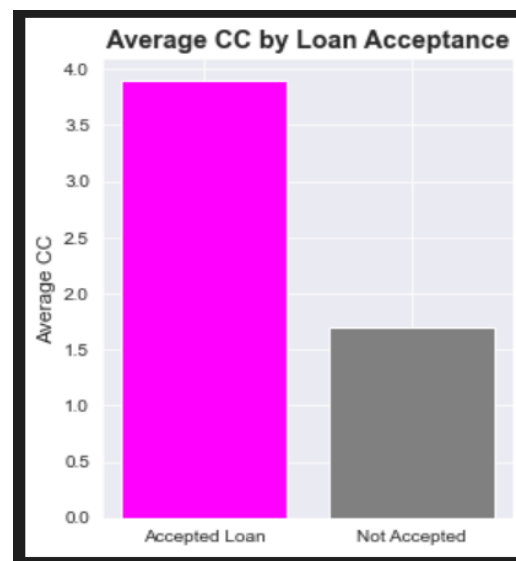


Figure 5

## Loan Acceptance:

- 1) I used bar charts (figures 2-5) to show how relevant certain features are towards accepting a loan or not.
  - a. You can see based off these charts that higher education, more income and people that already have a high mortgage are more likely to accept loans compared to others
- 2) Using XGBoost we ran a classification model to predict if a client will accept or reject a personal loan. As shown in figure 6, you can see the feature importance related to our decision-based model.
  - a. This graph shows that education, income, and family make up the three biggest features that determine loan's being accepted.
  - b. All features are relevant however because you can assume that all features are relatable to one another
    - i. Family relates to income because the more family that works in a household is likely to have more income
    - ii. Income relates to education because the higher education you have, you are more likely to have a higher income
    - iii. Having a credit card relates to income because if you are using credit then you can most likely afford borrowing amounts of money and paying it off later

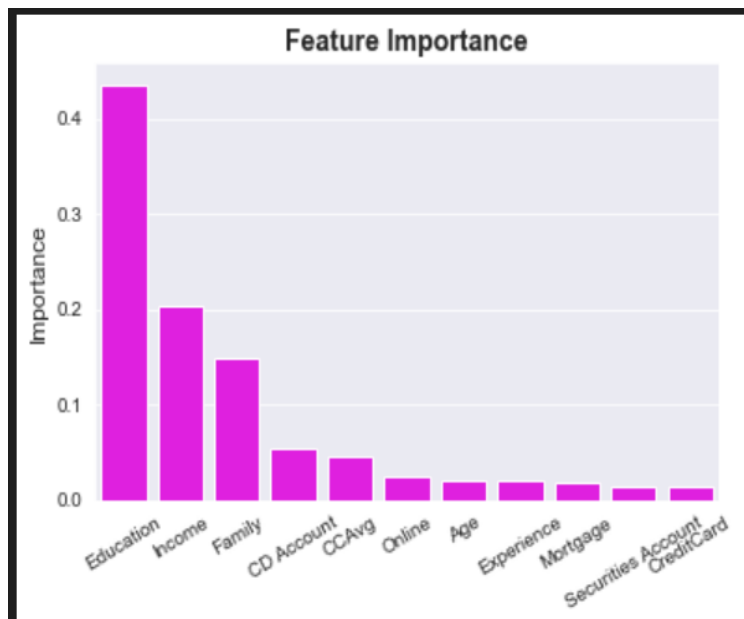


Figure 6

### Problem Statement:

Thera Bank has been looking more closely at expanding their personal loan department, while deducting its marketing budget. They are trying to see how to increase their personal loan customer base with their reliable depositor base. This includes a wide variety of features that they based off their customer's background information that would influence their chances of taking out a loan. In conclusion, Thera Bank is looking for a larger increase in their personal loan department based off the features that they provided from their client's demographics. However, we must find out which features impact the decision of accepting or rejecting the loan the most. This in turn out help the bank figure out which customers are the most likely to accept the loan and increase their interest on loans from borrowers.

### Data/Approach

Pertaining to the data, I got the data off Kaggle I which includes 5000 rows of bank cliental demographics that relate to their lives. The data is for Thera Bank that is trying to expand their personal loan department. I did the following to formulate the dataset into what features are the most impactful for the acceptance or rejection of the loan. First, I split the data up into two separate tables, one being with the personal loan being accepted and the other being if the loan was rejected based off all the features provided. I then used the describe function () for both data frames to find the count, mean, std, min, 25%, 50%, 75% and the max for each. This allowed me to create bar charts based off the features with the largest difference in means for each. Figures 2-5 above, are the charts that had the largest variations between accepting and rejecting a loan from Thera. Next, I created a feature set (X) and output variables (Y) to represent the personal

loan and all the features involved. I did drop personal loan, zip code and ID from the x values as they were structured/standardized numbers that had no impact on the loan. From there I split the data using `train_Test_Split` for training and evaluating the model. I used a test size of 30% which sounds about a good proportion of the data and when randomized, interpreted positive results. Following the split test, I ran a `XGBclassifier` as binary classification because the only two outputs there can be is if the loan is accepted or rejected. The training set results reported at a 1.00 and the test set was at .99 which is very close. The results we had were 1.00 on the training data and .99 on our test data, which is significant evidence of overfitting. This is most likely because of the lack of data. So, we adjusted multiple hyper parameters such as: max depth, early stopping rounds and number of boosted rounds to reduce overfitting. However, this had little to no affect on reducing our original results, so to have increased confidence in our evaluation results we implemented the Kfold cross-validation technique. I then used the built-in function from `XGBoost` to grab the importance of each feature that was found while training the model. This shows that clients that have a higher education, income and a larger family have a higher chance of accepting a loan compared to the other features. This showed us that our method once ran, has a 98% accuracy from the test and training set. I decided on this because it seemed like the best way to show the most important features based off their differences of the two data frames. Then, being able to show the importance between all the features and the loan, which should give Thera an idea of what to look for in their cliental and if they meet the credentials from the information I have provided.

## Analysis

The results show us that personal loans are accepted based off all features but more heavily on certain ones. Education, income and family all in some way relate to one another. If you have a higher education, you will get a better job and have more income. If you have a family and both parents work, then you will increase your income as well. Also, this provides a look forward towards people who are educated and have money, will be paving the path for their children as well. The other features that still have an impact on the acceptance of a loan relate to the main factors as well. If you already have a mortgage, then you likely to take out another loan and can pay it back. People already having bank accounts and CD accounts with Thera are more likely to take out a loan with them since they are used to the system and familiar with the policies they have. Clients that are older are usually more established individuals and have a family, high income, are educated, etc. This analysis showed that all features are related to one another and can have an impact on the result of a loan being accepted or reject. Also, it shows that certain factors such as education, family and income make a large portion of the clients who have accepted loans already. So, Thera should focus their attention to clients with an higher education, more family members and report higher incomes than average. This will help Thera increase their loan department and their payments through interest payments. I was able to see that our model is 98% accurate based off the test and training methods used. Proving that figure 6 is a heavily reliable graph that shows accurate influence for personal loans being accepted and what criteria Thera Bank should look at when offering loans to their clients.

## Conclusion

In conclusion, I discovered that there are certain features that are more relative to loan acceptances than others. Based off my tests, I was able to see the importance of each feature and

determine how relative they are to loan acceptance. While all that were included in my tests had relevance; education, income and family had the most influence towards accepting a loan from Thera Bank. Based on our data if Thera focuses on clients that are more than average in those three categories, they will have a higher success rate of clients to go after the loan. This helps Thera increase their loan department by focusing their resources towards more fitting cliental. Based off figures 2-5, you can see the drastic differences between clients who accepted the loan and did not accept one. Also, we saw that even in small, randomized test group of the data that you can likely view clients based off these features and determine if they are worth talking or not talking to about a loan.