289.33
XLS BANK MARKETING
LOAN MODEL
By Erik Castro // Flatiron School

Problem Statement

- Improve personal loan conversion rate
  - *Make data driven recommendations*
- Model to predict customer conversion
  - *Create a model that can predict c likelihood of a customer converting to personal loan customers using customer features.*

I will be looking at what factors affect the likelihood of a customer becoming a Personal Loan Customer. This will be answered using data driven recommendations The second part of the problem will be to create a model that can give a percentage of the likelihood of the customer transferring to a Personal Loan Account.

Business Value

UNDERSTAND THE RELATIONSHIP OF DIFFERENT CUSTOMER FEATURES

INVESTIGATE HOW THIS FEATURES CAN HELP TO CONVERSION RATE

GAINING INSIGHT IN HOW LIKELI A CUSTOMER IS TO BECOME A LOAN CUSTOMER CREATING A MARKETING CAMPAIGN

Methodology

Frame the Problem

Collect Raw Data

Process the Data

Explore the Data/ Perform In-Depth Analysis

Train and Evaluate Models

Communicate results

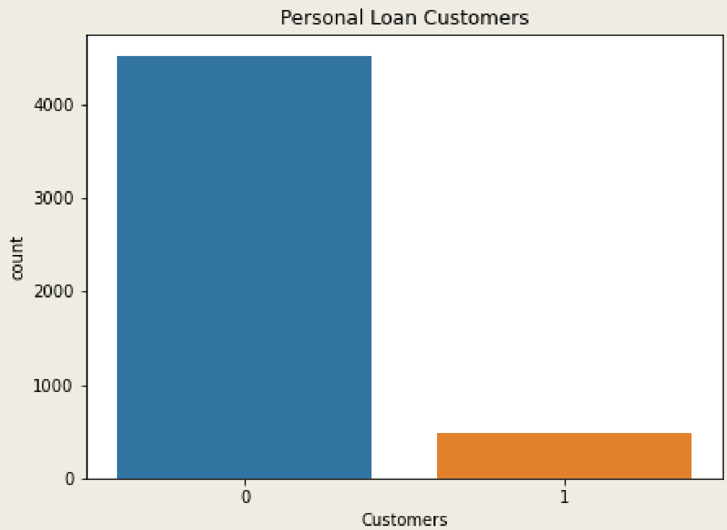
Frame the problem: Identify business priorities and make strategic decisions that will lead my work
Collect raw data: Extract data from the database provided.
Process the Data: Understand the data and proceed to clean it.
Explore the data: Split the data in different ways and use statistics to test and create visualizations to interpret data
Train and Evaluate Models: Create multiple linear models to predict likelihood of a customer transforming to Loan Customer
Communicate Results: Explain findings with visualizations created before. Present findings.

Collect Raw Data and Process the data

- There are 5,000 customer data entries.
- Originally there were 14 features. However, three of them were not necessary for this specific project:
  - *ID*
  - *Zip code*
  - *Experience*
- Clean data: Missing Values, Categorical values, syntax

Looking at the customer data provided from the last campaign, there were 500 data points. I selected 11 features from this dataset as it didn't seem necessary to have ID, Zip Code and Experience for predicting whether a customer would convert to Loan customer. The necessary data features are: Age, Inome, Family, Credit Card Average, Education(Categorical: Undergraduate, Graduate, Professional) , Mortgage, Personal Loan, Securities Account, CD Account, Online and Credit Card
Cleaning the data involves prepping the data for modelling, this includes: addressing missing values, amongst other things.

Explore the data/ Perform In-depth Analysis



In this visualization, we can see the number customers that have a Personal Loan based on our 5,000 customers in the dataset. 9.6% of customers have a personal Loan. This is of course with the initial marketing campaign.

Train and Evaluate Models

	Accuracy	F1	Precision	Recall
Logistic Regression	0.96	0.76	0.89	0.65
KNN	0.95	0.70	0.92	0.56
Decision Tree	0.98	0.90	0.90	0.91
Random Forest	0.98	0.91	0.98	0.85
AdaBoost	0.97	0.82	0.92	0.75
Gradient Boosting	0.99	0.92	0.97	0.88
XGBoost	0.99	0.93	0.95	0.91

With the data prepared, I tried different parameters with different classification models to predict which customers would become Loan Customers.
The models are on the left and you have the different scores on the other columns.
The metric used is F1, which tells us the percentage of a correct prediction of the likelihood of a customer to transform to a Loan Customer.
Based on this table we can be sure that the model that predicts this dataset the best is XGBoost, with a 93% of correct prediction.

Results

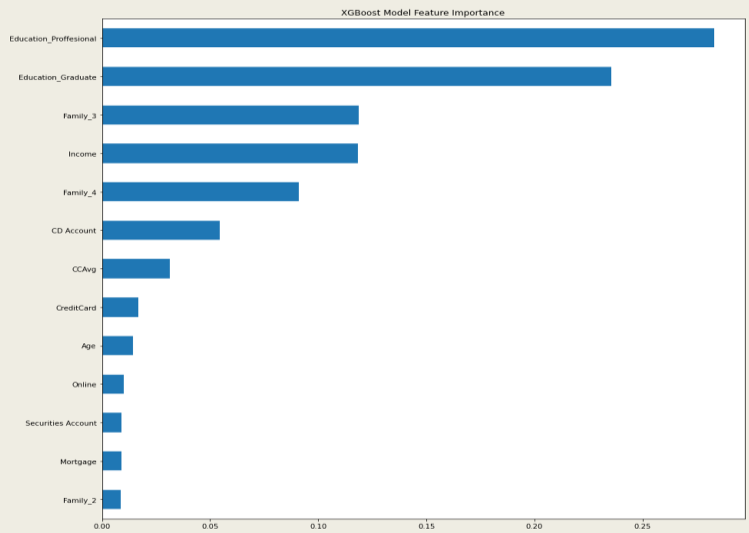
- 400 customers in the test dataset
- Using the normal approach only 9.5% will convert to Personal Loan.
- expensive process,
- marketing strategy is general to all 400 customers

- XGBoost classification selects customers with +50% chance of becoming Personal Loan Customers.
- 38 Customers selected
- 35 actually converted to Personal Loan Customers

400 customers in the test dataset
Using the normal approach only 9.5% will convert to Personal Loan.
This, however, would be an expensive process, as the marketing strategy is general and applied to all 400 customers

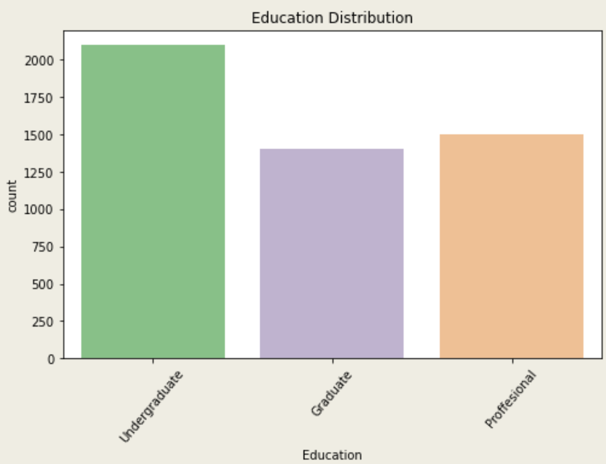
Using XGBoost classification we select only the customers that have a +50% chance of becoming Personal Loan Customers.
The number of customers the marketing strategy is focus on is down to 38, this is a drop of 90.5% of data selection.
35 out of 38 customers became Personal Loan Customers.
This means that there was a 92.1% success rate using the XGBoost Model.

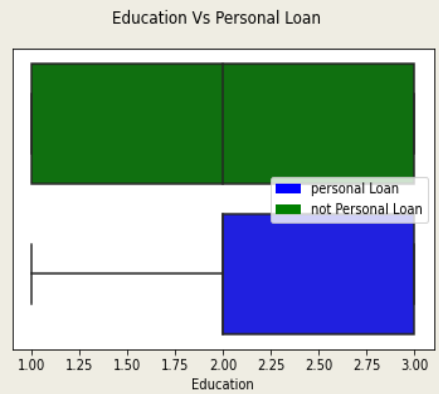
Recommendations – General



Based on the feature importance of XGBoost model , we can conclude that the top 5 features are, Education Professional, Education Graduate, Family of 3, Income and Family of 4.

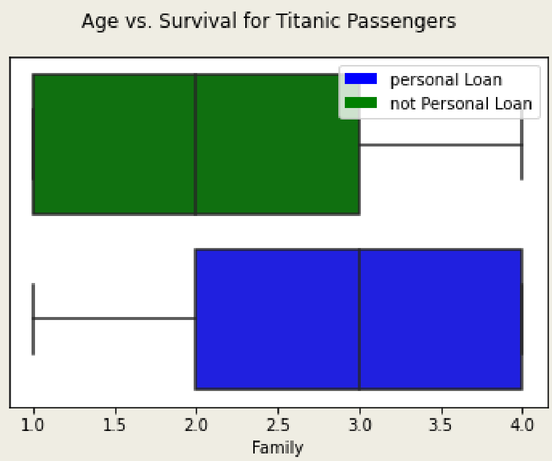
Recommendations – Education





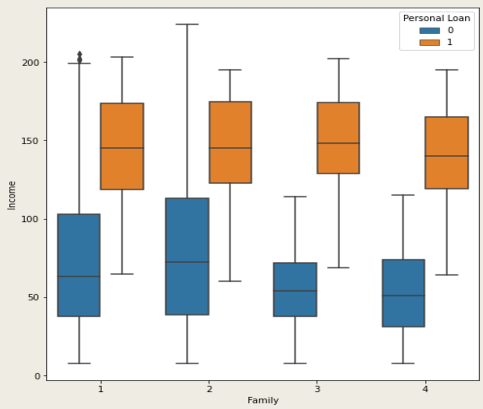
Based on the findings Education is the first Recommendation, from this chart it can be seen that within Education, Graduate and professional form 58.08% of the total data gaining room for targeting with the marketing strategy.
On the chart on the right we can see that customers that have a Personal Loan account are from Graduate and Professional. However, the data also shows that those that have no personal Loan are from all Education classes. This means that there is a lot of chances of getting Graduate and Professionals account holders to move to a Personal Loan account.

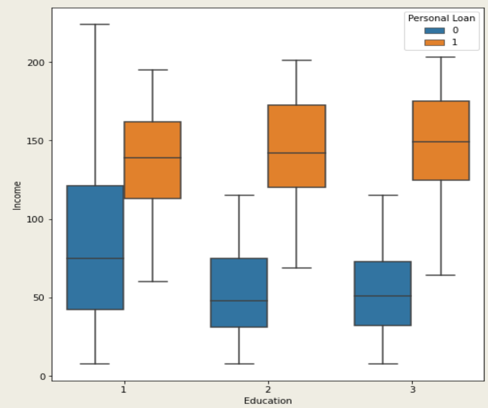
Recommendations – Family



Based on the previous recommendation we are now looking at family. Based on the chart above we can appreciate that Families that are compose of +3 members are more likely to get a Loan.
It is recommended that customers with more than 3 family members are targeted in this campaign.

Recommendations – Income VS Education and Family





We are now looking at a more complex chart. This recommendation, is based on the previous recommendations. On the boxplot on the left, we can appreciate that the higher the income and the bigger the family the higher the chances of having a Personal Loan customer.

On the chart on the right the higher the income and the higher the Education, from Graduate to Professional , the higher the probability of having a Personal Loan Customer.

Future Work

- More data
- Testing on new dataset.
- Zipcode

More data: This dataset is relatively small, with only 5,000 entries. Perhaps, if the dataset had more entries the models would have behave differently.
Testing on new dataset: Once this marketing strategy has been tested in real life, with the new dataset I can further tune the models to find out more specific target customers.
Zip Code: Introducing Zipcode could change a lot in the models, specially looking into rich vs accommodate areas. This could be included in the models.

Thank you

Erik Castro – Flatiron School

Any Questions?