AllLife Bank: Loan Analysis

Joe Balog
Data Analyst
josephmbalog@gmail.com

table of contents

Business Overview

Why are we here?

Context

Issues

History

Business Problem Overview

context

AllLife Bank is a U.S. bank that has a growing customer base. Most of these customers are liability customers (depositors) of the bank with varying sizes of deposits.

The number of customers who are also asset customers (borrowers) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business and in the process, earn more through the interest on loans.

In particular, the management wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors)..

A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise campaigns with better target marketing to increase the success ratio.

Classification Problem

problem

The ultimate problem is that we do not know the answer to question, “Why did people accept a loan based on the last campaign?” or more generally, “why get accept a loan offer at all?”. What are the incentives for purchasing a loan?

The question is, how likely is it that a current liability customer will want to purchase a personal loan as a result of the ad (accept the loan offer)?

On the flip side, what is the likelihood that asset customers will want to purchase another loan (accept again)?

How do we measure the likelihood of loan-offer acceptance through an ad? Will the same people who bought a loan after the first ad buy another one? If so, who are these people? Will non-loan-offer accepters of the last campaign purchase a loan? If so, why?

Data Science problems to solve:

To predict whether a liability customer will buy a personal loan or not (conversion).

Which variables are most significant in prediction. (ad-content).

Which segment of liability customers should be targeted more by subsequent campaigns. (customer profile).

Business Problem Overview

problem assumptions

Assumptions are baked into every decision. The assumptions we bring to this data exploration and prediction problem as inherent and need to be explicit in order to set the context better for the proposed solution for the marketing team.

When we consider the data set given, we must assume that these are the key features and only the features used in the data gathering process. There may be other and more important factors that were not included. But we must work with what we have.

Success is the conversion of non-to-new personal loan customers for AllBank. We are assuming that this is possible through target marketing and not some other means. As psychology tells us, people make financial decisions based on emotional states rather than rationality (economic logic), so we can assume that an emotional appeal in an ad is more likely to convert a customer. But does the data give us information about customers' psychology?

Business Problem Overview

problem assumptions

We also must make assumption about the campaign. Will the campaign target only the remaining non-loan accepters from the last campaign or target all liability customers? From the objectives we can deduce that the use of segmentation will be needed to create a customer profile of “loan accepters” based on the data’s features.

I think we must assume that the data is the result of the last campaign. Thus, it includes only those customers targeted with the ad.

This also complicates things. Because whose to say that this subset of liability customers were targeted last time represents the type of customers to be targeted next time? How did these criteria come about in the first place? Is it best to target the same types of people again or to target another group and compare the success ratios?

solution approach

The type of supervised learning problem is one of classification. The goal is to predict the likelihood of a liability customer purchasing a loan of the personal kind.

You as a Data scientist at AllLife bank must build a model that will help the marketing department to identify the potential liability customers who have a higher probability of purchasing the loan and becoming asset customers (borrowers).

The goal is to improve the success ratio of future marketing campaigns. The goal is to convert liability customers (depositors) into asset customers (borrowers). Since we are targeting current liability customers with another campaign (internal marketing) then we want to have the same types of people be targeted.

solution approach

To predict whether a liability customer will buy a personal loan or not (conversion).

We will use machine learning models.

Which variables are most significant in prediction. (ad-content).

We will evaluate models and extract statistics.

Which segment of liability customers should be targeted more by subsequent campaigns. (customer profile).

We will use descriptive statistics and model results.

Data Overview

What's the data?

Meta-data

Features

Tasks

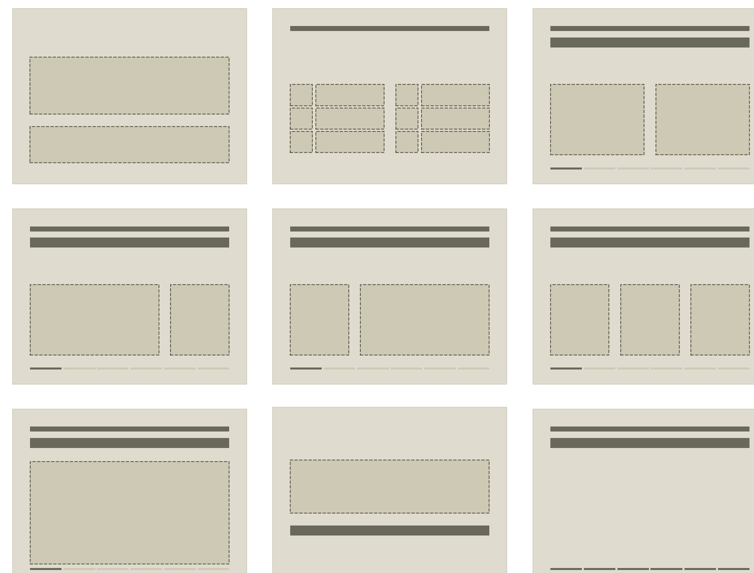
Data Overview

The data contains...

5000 rows by 14 columns of numeric data (int64 and float).

The data features include demographic information (income, age...) and customer information (mortgage, securities account, etc.). It also has the response to the last personal loan campaign.

Very few of the liability customers, ~9%, accepted the loan that was offered in the marketing campaign in this California bank.



Never ask for **small loans.**

Onassis, Aristotle

Exploratory Data Analysis

What's in the data?

Statistics

Descriptive Analysis

Relatoinships

Let’s look at the data!

In this data set of 14 features for exploration we have:

Demographic and bank-meta-data:

- Customer ID
- Customer’s age
- Years of professional experience
- Annual income
- Home Address ZIP code
- Family size
- Spending on credit cards per month (credit card debt)
- Education Level
- Value of house mortgage

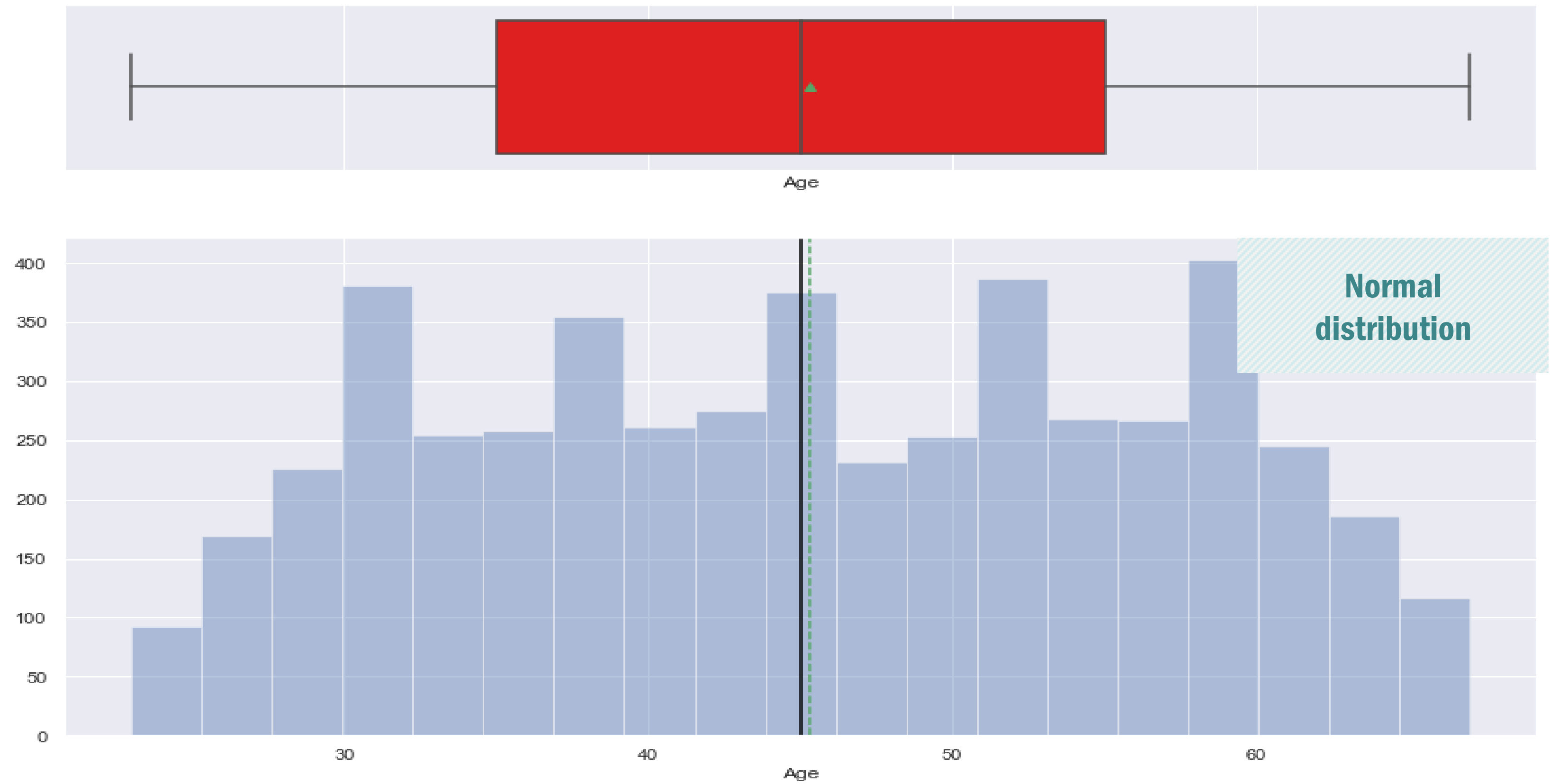
Let’s look at the data!

In this data set of 14 features for exploration we have:

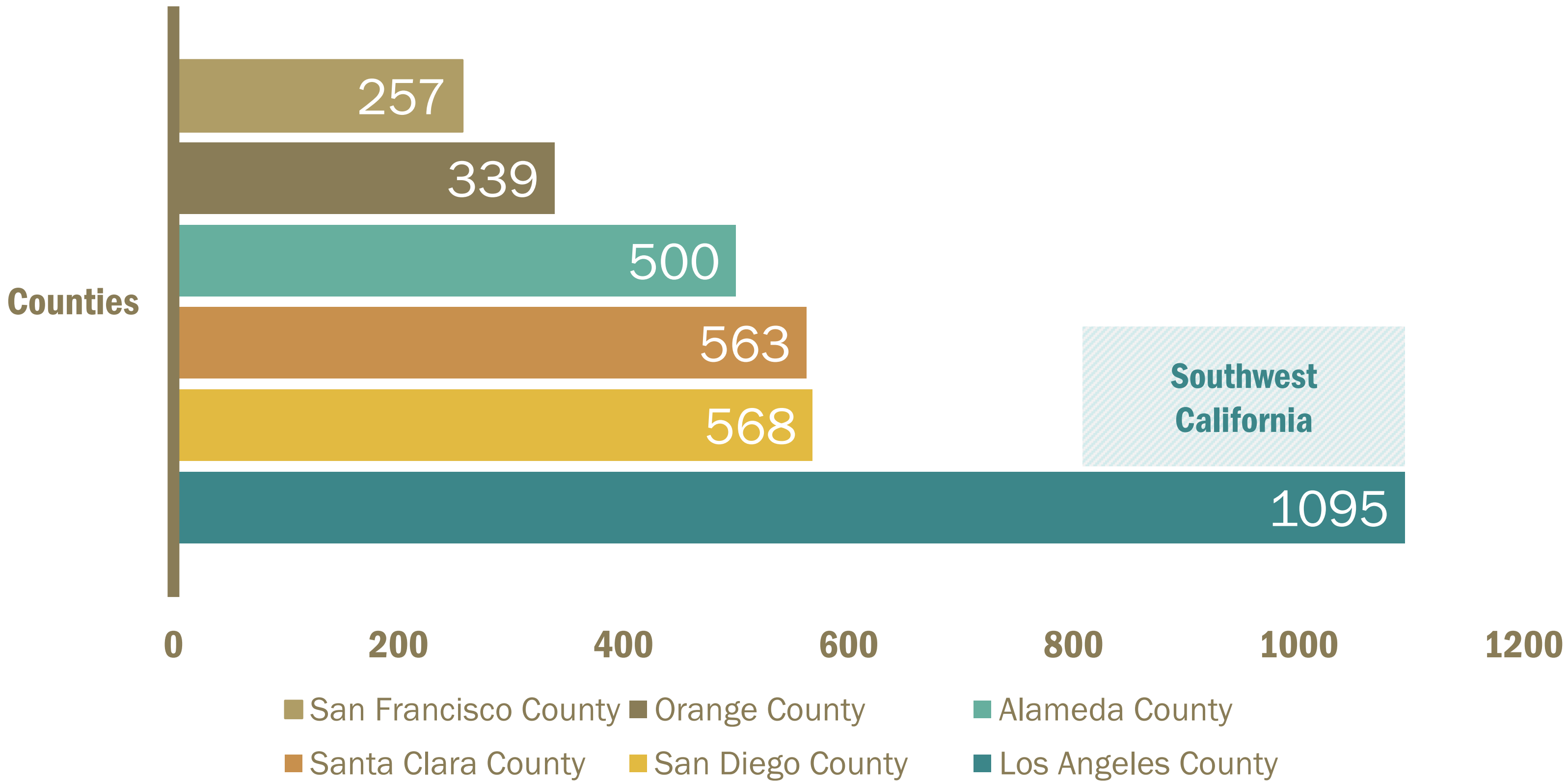
Current liability customer responses to:

- Did this customer accept the personal loan offered in the last campaign?
- Does the customer have securities account with the bank?
- Does the customer have a certificate of deposit (CD) account with the bank?
- Do customers use internet banking facilities?
- Does the customer use a credit card issued by any other Bank (excluding All life Bank)?

Age

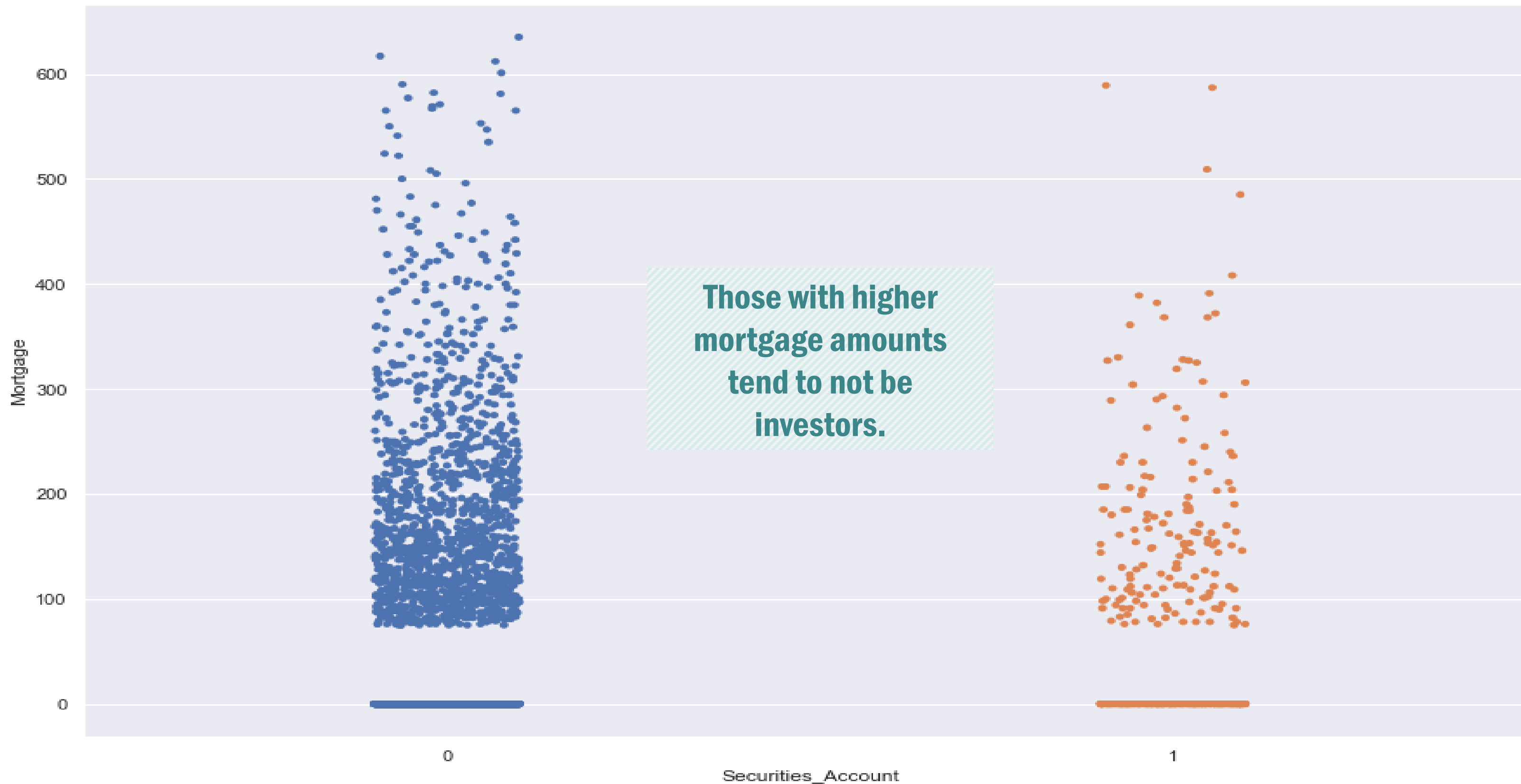

Age
400
350
300
250
200
150
100
50
0
30
40
50
60
Age
Normal distribution

Counties


Counties
257
339
500
563
568
1095
Southwest California
0
200
400
600
800
1000
1200
San Francisco County
Orange County
Alameda County
Santa Clara County
San Diego County
Los Angeles County

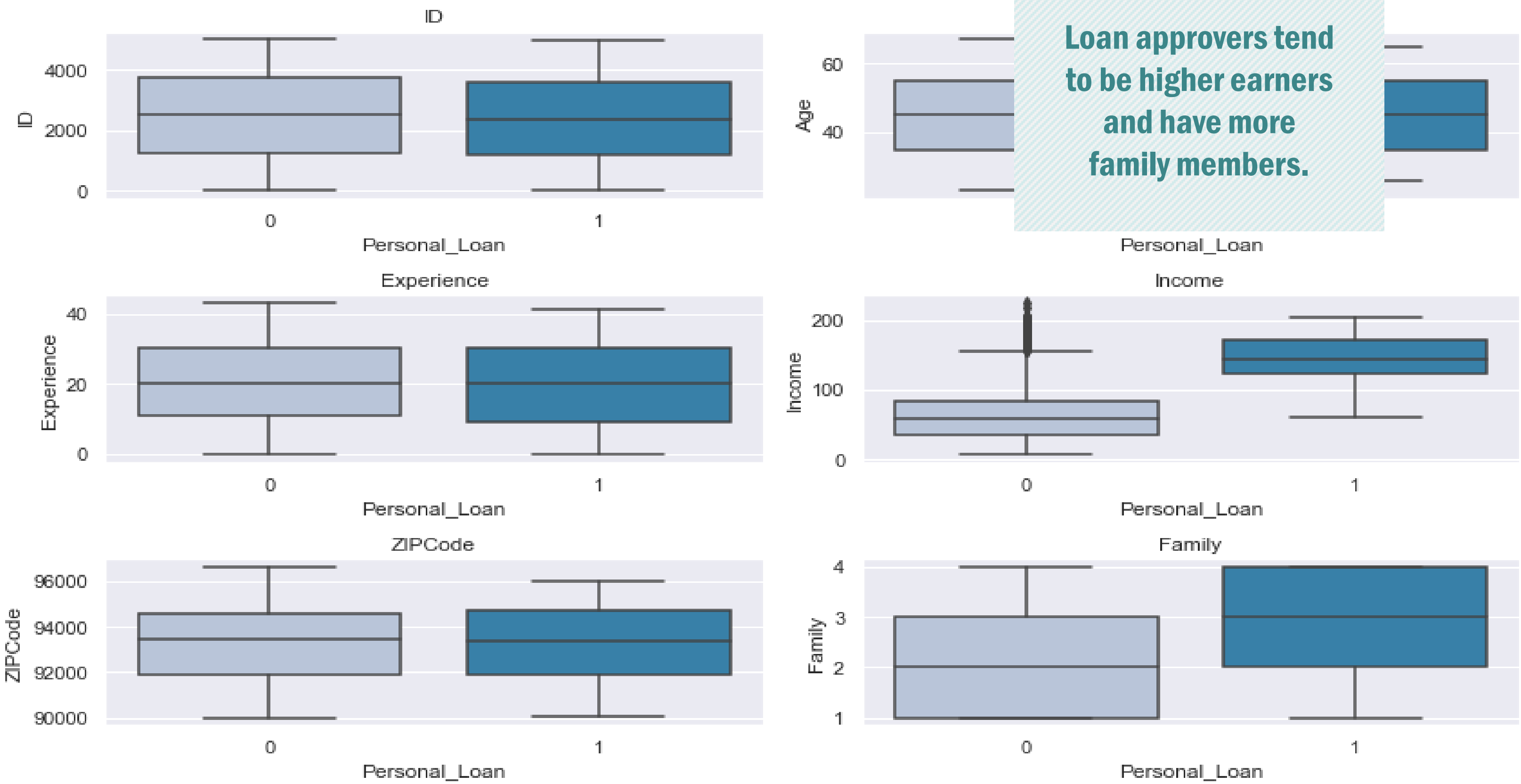
Credit holders and average spending

Mortgage (1000's) and investors.



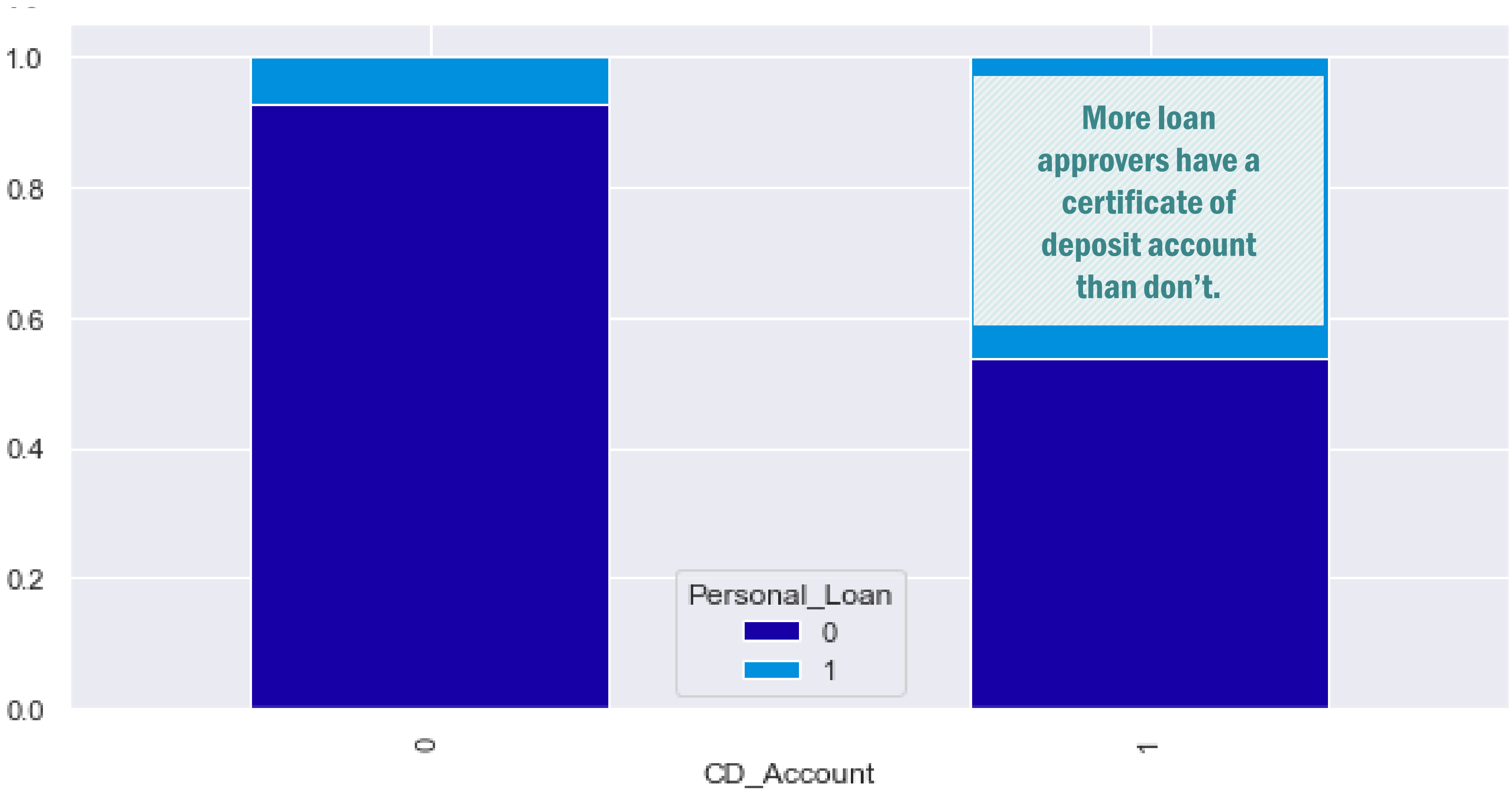
Exploratory Data Analysis

Demographics and loan offer approval.



Loan approvers tend to be higher earners and have more family members.

Do loan approvers have a deposit account?

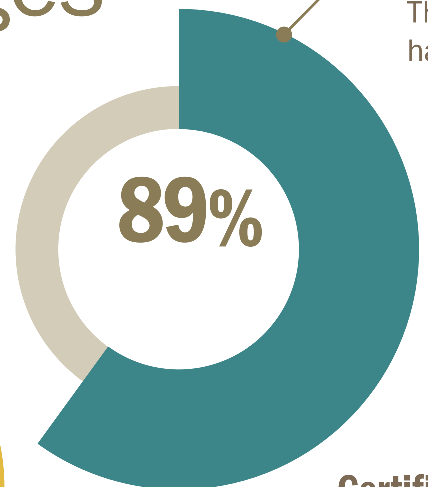


Exploratory Data Analysis

High-level percentages

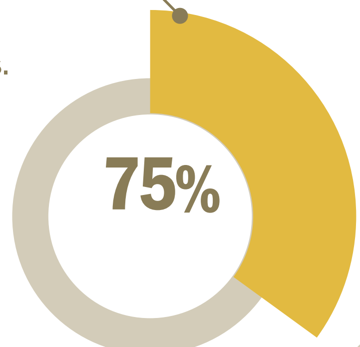
Securities Accounts

The majority *do not* have these accounts.



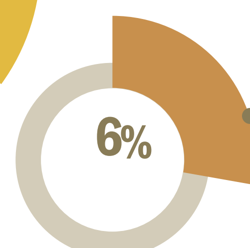
Education

Liability customers tend to lie below Advanced Professional status.



Certificate of Deposit Account

Of the 5000 data points only a very few have CD accounts.



Who are we looking at?

We can conclude that liability customers targeted in the last campaign include folks who are 35 - 43, have 9-32 years of professional experience, make around $38-44k a year, live near Palo Alto or Berkeley, have 1 -2 children, have credit debt of $300-$1000/month, have an undergraduate degree, no mortgage, non-investors, non-CD holders, online feature-users, and credit card holders with other banks.

If you torture the data long enough, it will confess.

H. Coase, Ronald

Essays on Economics and Economists

Model Performance Summary

What algorithm?

Logistic Regression

Decision Tree

Performance Measures

Prediction

We want to predict whether a liability customer will buy a personal loan or not based on whether they did before. **Put simply, is a current liability customers (depositor) likely to become an asset customer (borrower) after a campaign ad?**

We will use **Logistic Regression** and **Decision Tree** models to predict outcomes and compare results between them.

The outcome we want to avoid is predicting (and therefore targeting) non-loan purchasers as loan purchasers.

In order to understand the model we will consider:

- Reducing Multicollinearity in a regression model
- Pruning a Decision Tree for Gini-coefficient (purity)
- Tuning hyper-parameters of Decision Trees

We will also utilize a **Confusion Matrix** to grasp the predictions strength and validity.

0.
0.009

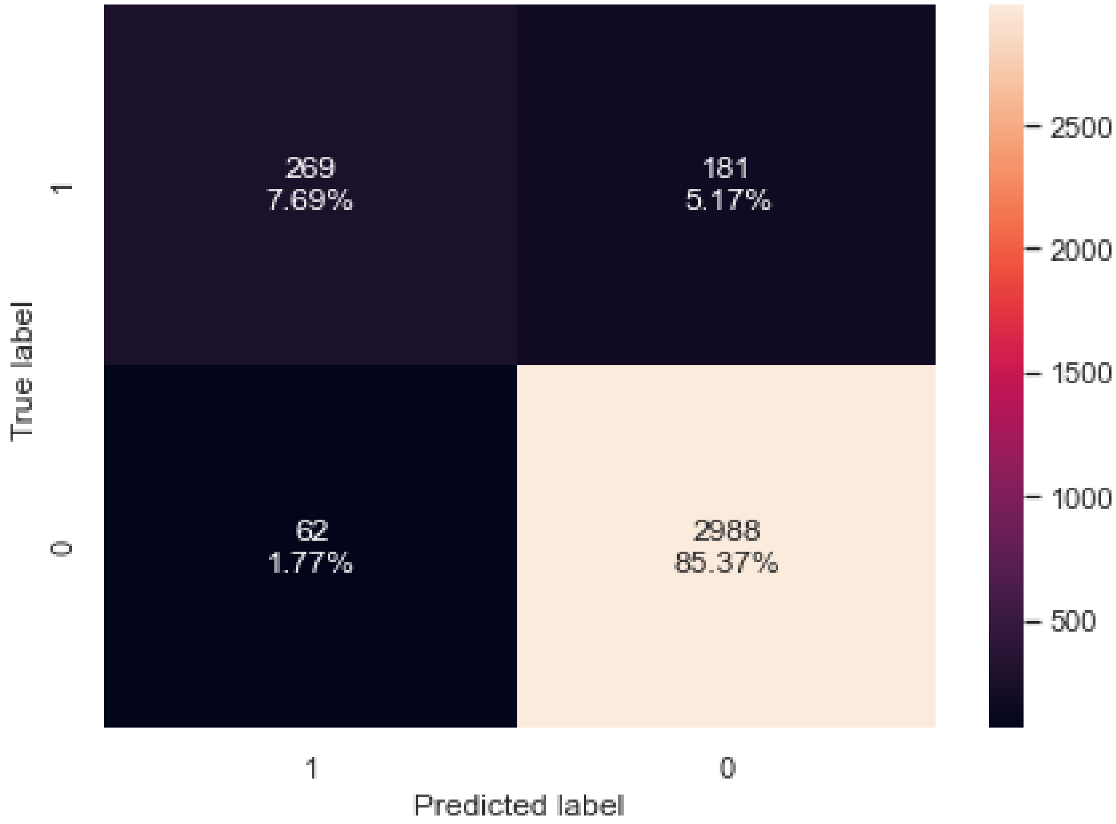
Performance Metrics

Model	Recall	Precision	Accuracy	F1 score
Logistic Regression T= .56	Training: 92% Test: 91%	Training: 38% Test: 43%	Training: 85% Test: 87%	Training: 68% Test: 69%
Decision Tree Classifier ccp_alpha – 0.009	Training: 99% Test: 98%	Training: 59% Test: 62%	Training: 93% Test: 94%	Training: 74% Test: 76%

Recall was chosen to reduce the effect of False Negatives (predictions of 0 where 1). F1 is used to compare the two models. The Decision Tree performs better and is thus more effective as a model for prediction loan purchases.

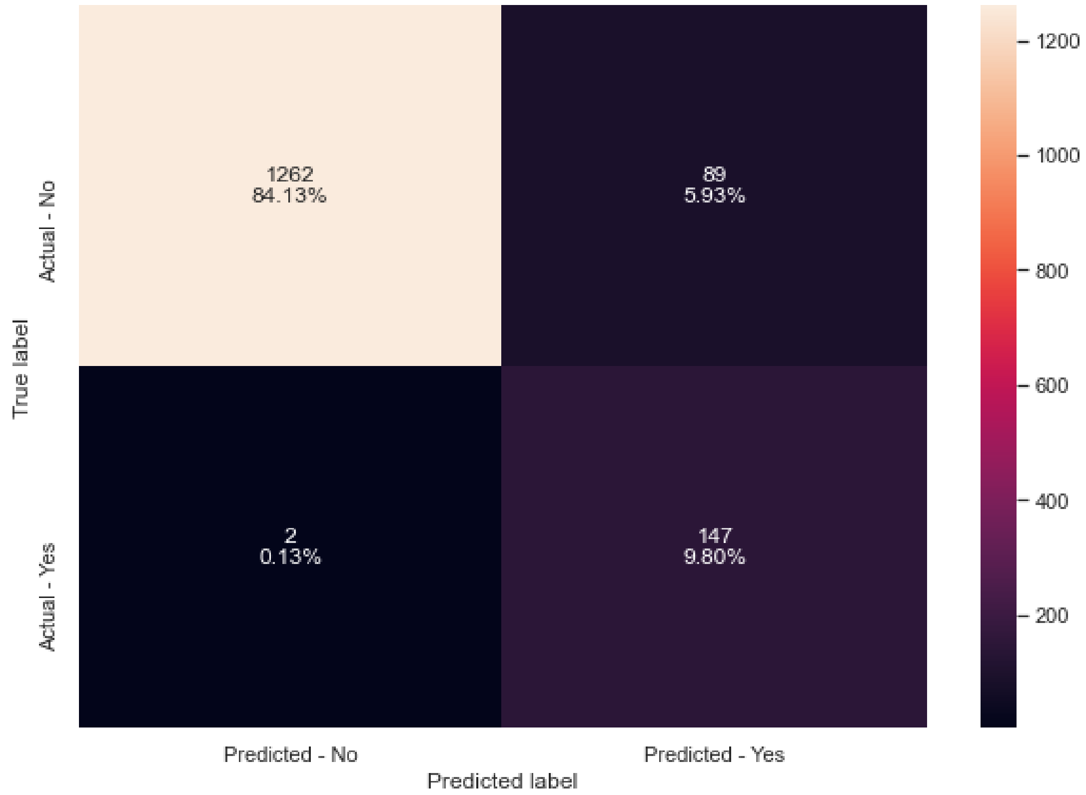
We see the recall was higher for the decision tree model. In fact, all performance measures were improved from the logistic model to the decision tree model. The final decision tree model used a very low alpha (0.009) and weighting 85/15. The other decision tree models did not perform as well on recall.

Confusion Matrix: Logistic Regression



Model Performance Summary

Confusion Matrix: Decision Tree



Prediction: Conclusion

We used **Logistic Regression** and **Decision Tree** models to predict outcomes and compare results between them.

The Decision Tree Model outperformed the Logistic Regression model on every measure. It is useful for predicting the likelihood that a liability customer will convert to an asset customer.

It is also useful for creating our customer profile of which target segment to market to.

In order to understand the model we will consider:

- True Positives (TP): we correctly predicted that they do accept the offered loan: 147.
- True Negatives (TN): we correctly predicted that they don't accept the offer. 1262.
- False Positives (FP): we incorrectly predicted that they do accept the offered (a "Type I error") 89. Falsely predict positive Type I error
- False Negatives (FN): we incorrectly predicted that they don't accept the offered (a "Type II error") 2. Falsely predict negative Type II error

Business Insights and Recommendations

What now?

Key Findings

Decisions

Next Steps

Key Findings

To predict whether a liability customer will buy a personal loan or not.

Those who are already liability customers (borrowers) are likely to buy a personal loan if they are

Which variables are most significant.

Those who are likely to accept future loan offers are most influenced by:

- Income
- Experience
- CD Account (already a depositor).

Exploration looks back.
Prediction looks forward.

Decisions

Marketing is the art of who, what, when, where and how. But business starts with why.

Which segment of customers should be targeted more.

To help the marketing department to identify the potential liability customers who have a higher probability of purchasing the loan and becoming asset customers (borrowers) and thus to improve the success ratio of future marketing campaigns, I recommend the following profile:

1. Income less than or equal to $98,500
2. Credit card average monthly debit of greater than $2,950
3. Undergraduate degree holders

Next Steps

Target Marketing

Now that we know who to target it's up to the marketing team to put together creative collateral.

After the next campaign pull new data with the following to boost our predictive power:

- Credit score
- Car ownership
- Equity measure

As the goal is to earn more money through the interest on loans it is worth considering alternative avenues in gaining profit aside interest returns.

For example, experiment with increasing fees in the following areas:

- Account fees
- ATM fees
- Penalty charges

As the goal is to earn more money through interest it might be interesting to investigate alternative methods in persuading customers to purchase personal loans.

Consider the following "nudges":

- "Anchoring" numbers on loan pre-approvals
- "Low-Balling" 1-3 year interest rates

Learn more about nudges here.

thank you

Joe Balog
For more info, please contact us at www.josephbalog.com

josephmbalog@gmail.com
231-313-7225