AllLife Bank

PERSONAL LOAN CAMPAIGN MODEL

by **JAKE EIDE**

BUSINESS PROBLEM OVERVIEW

How does AllLife Bank bring in more loan business? This is a key business problem that AllLife Bank currently faces, and one I will focus on in this presentation. We know that the number of AllLife's asset customers is small in comparison to its liability customers. Recognition of this fact provides AllLife with an incredible opportunity to expand the size of its asset customers, and in turn grow its profits through interest earnings on loans.

SOLUTION APPROACH

How do we begin to expand the number of asset customers? To best solve this problem, we need to identify the potential customers who are most likely to purchase a loan. I will build a series of models to identify these customers, and then from these models, choose one model that is most likely to provide the most useful prediction. The models will be built according to the following objectives:

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

2) To identify the variables that are the most significant in this prediction.

3) Identify the customer segments that should be further targeted for loans.

DATA INFORMATION

The data set contains information about 5000 AllLife Bank customers.

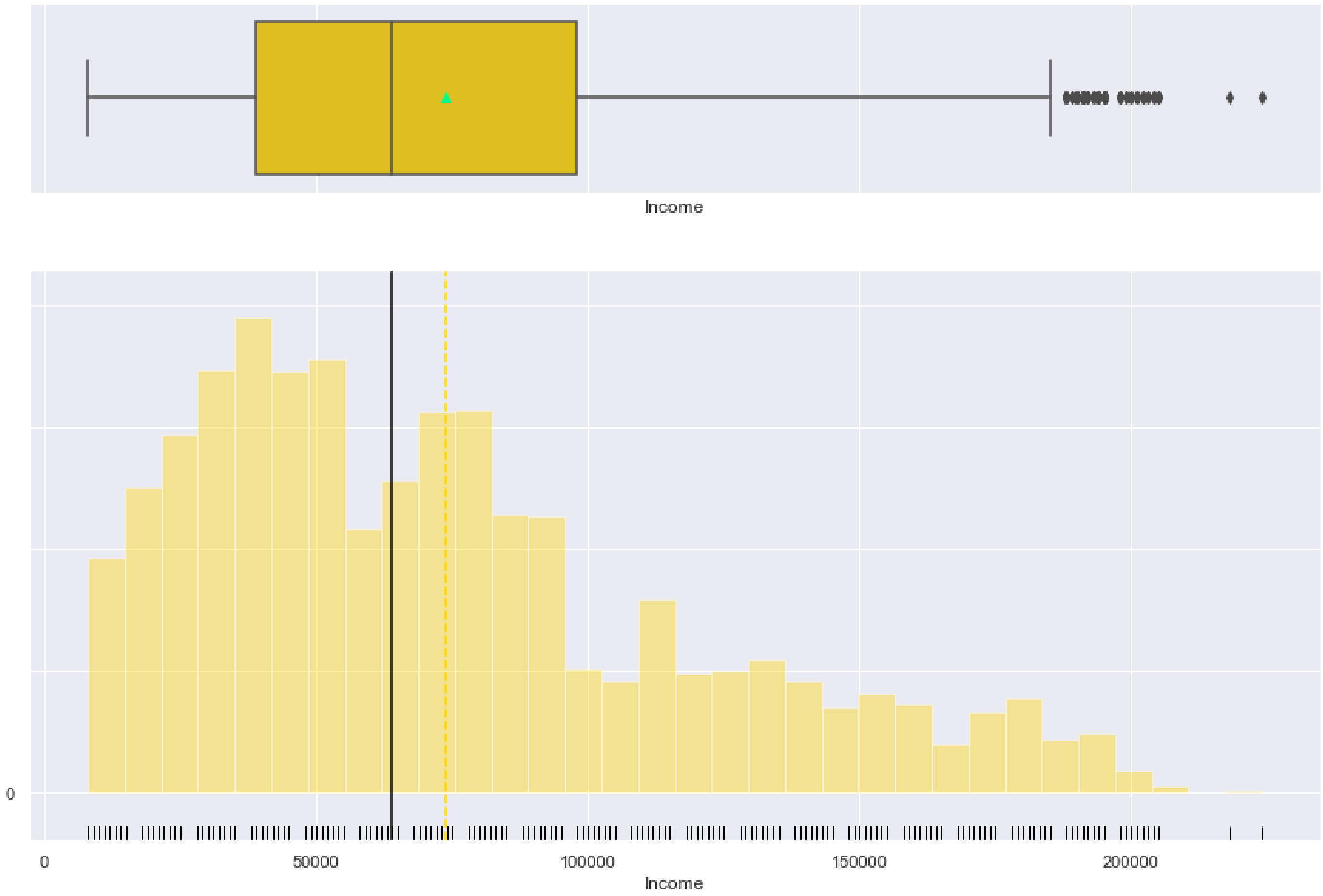
VARIABLE	DESCRIPTION
ID	Customer ID
Age	Customer's age in completed years
Experience	Number of years of professional experience
Income	Annual income of the customer
ZIP Code	Home Address ZIP code
Family	The Family size of the customer
CCAvg	Average spending on credit cards per month
Education	Education Level. 1: Undergrad; 2: Graduate;3: Advanced/Professional
Mortgage	Value of house mortgage if any
Personal_Loan	Did this customer accept the personal loan offered in the last campaign?
Securities_Account	Does the customer have securities account with the bank?
CD_Accoun	Does the customer have a certificate of deposit (CD) account with the bank?
Online	Do customers use internet banking facilities?
CreditCard	Does the customer use a credit card issued by any other Bank (excluding All life Bank)?

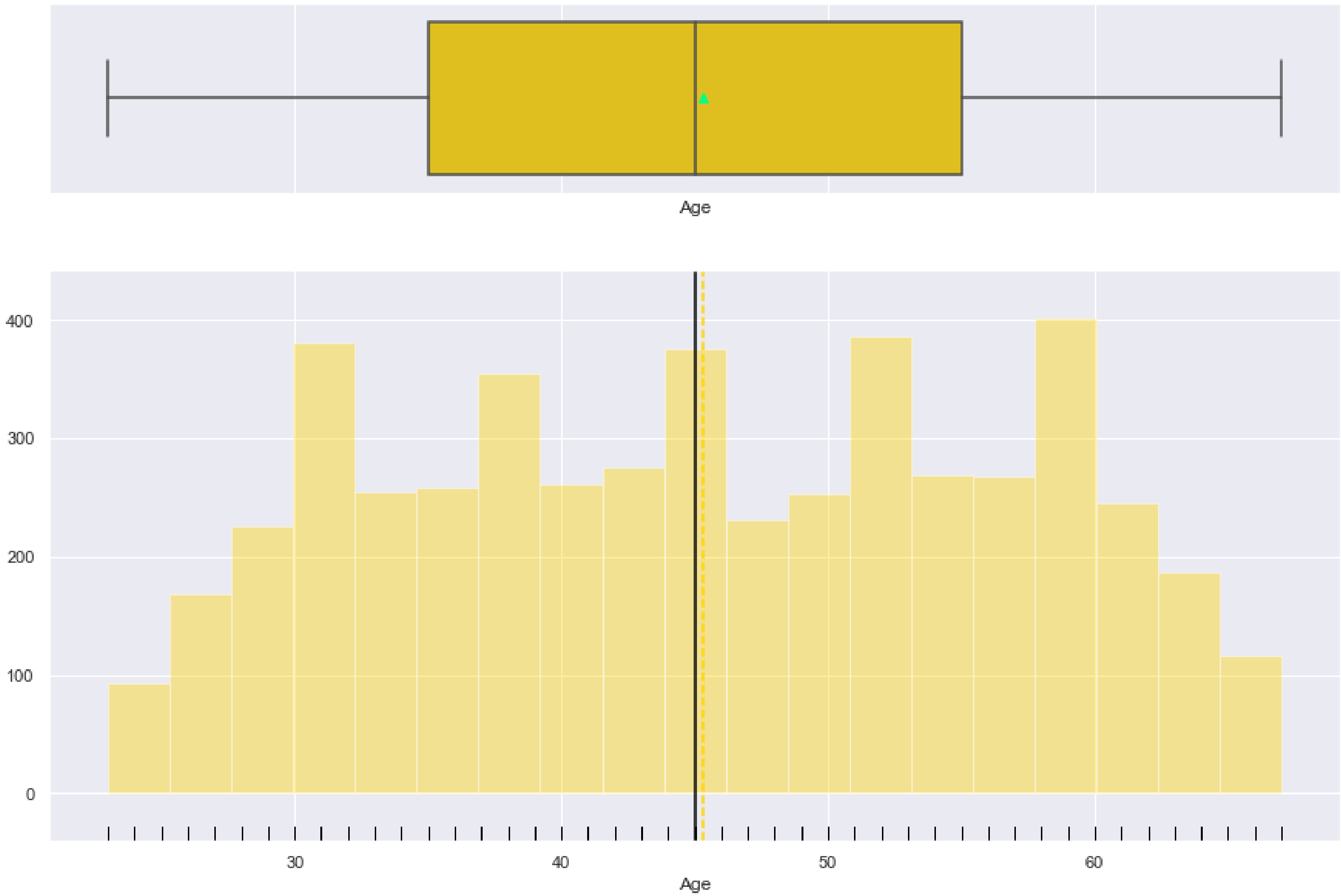
EXPLORATORY DATA ANALYSIS: **DATA SET OVERVIEW**

The purpose of this presentation is to help the bank increase the number of loan customers. For this reason, I won't spend a lot of time going through all of the individual variables in the data set. That said, it is useful to get an overview of AllLife's customers.

Key Observations:

- The income distribution of customers is right-skewed, with some outliers having high incomes
- The median income is around $64K per year
- Age is distributed fairly evenly between 23 and 67 years old
- The average AllLife customer is around 45 years old





EXPLORATORY DATA ANALYSIS: **DATA SET OVERVIEW**

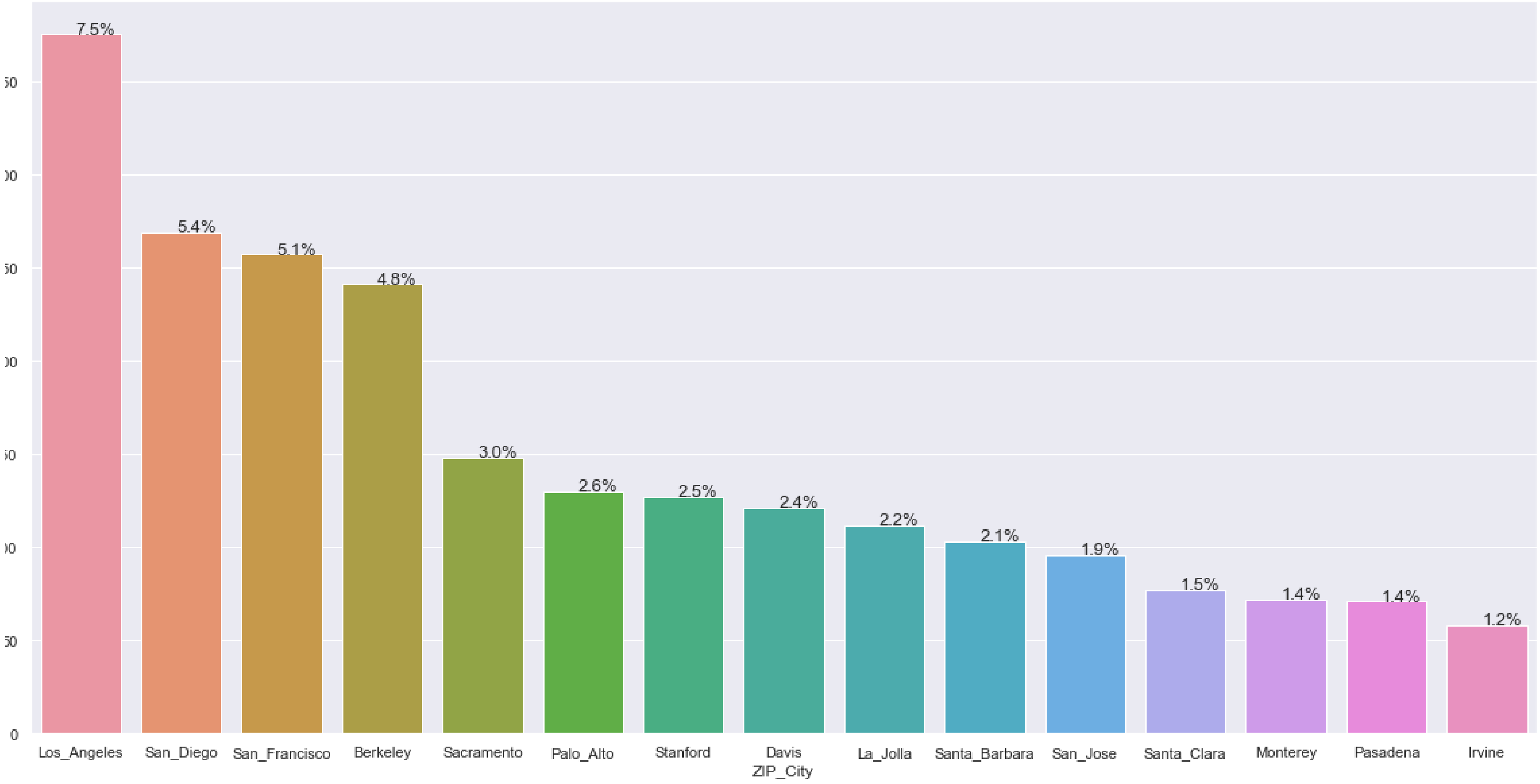
Key Observations:

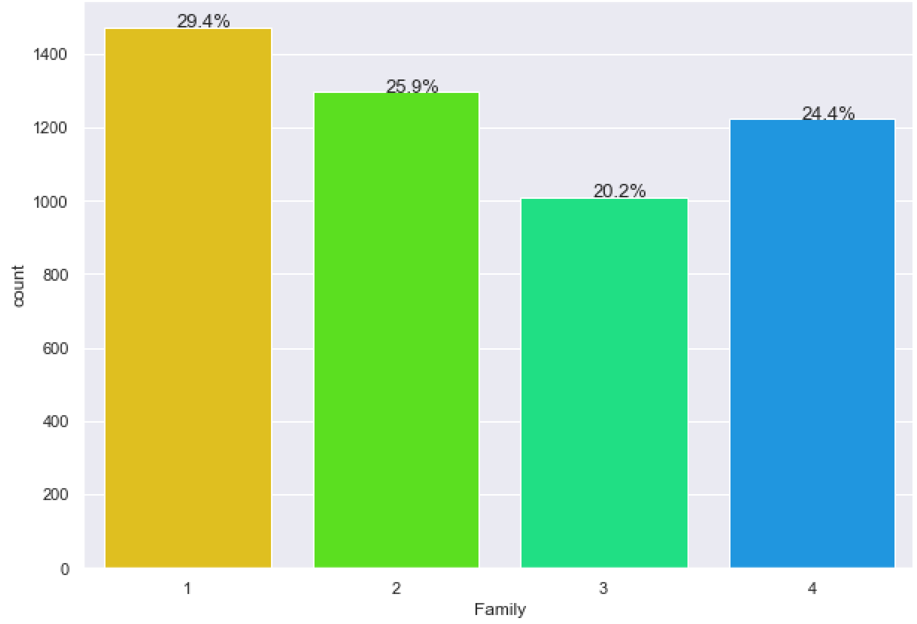
• Los Angeles has the most customers compared with other cities

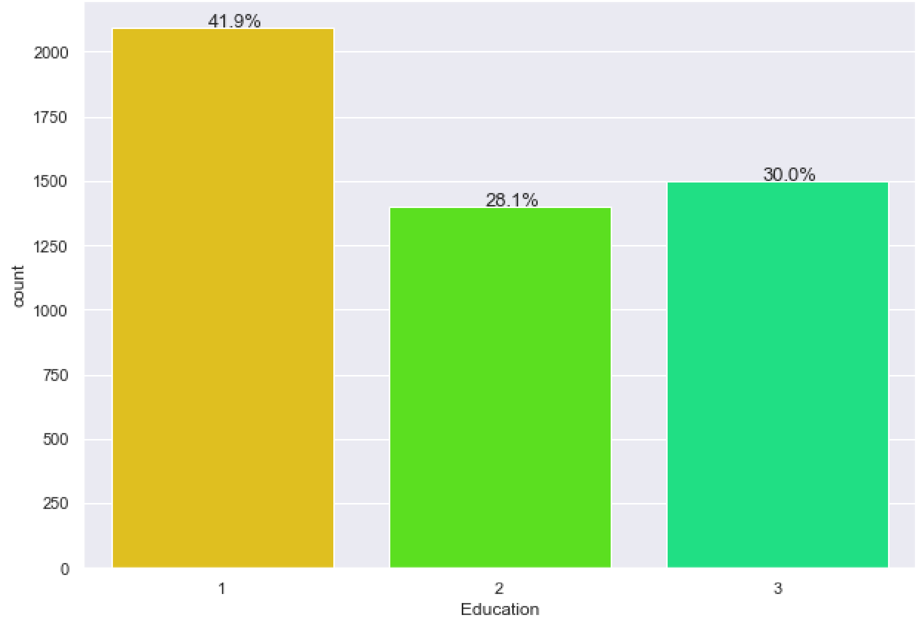
• Family size ranges from 1 to 4 people

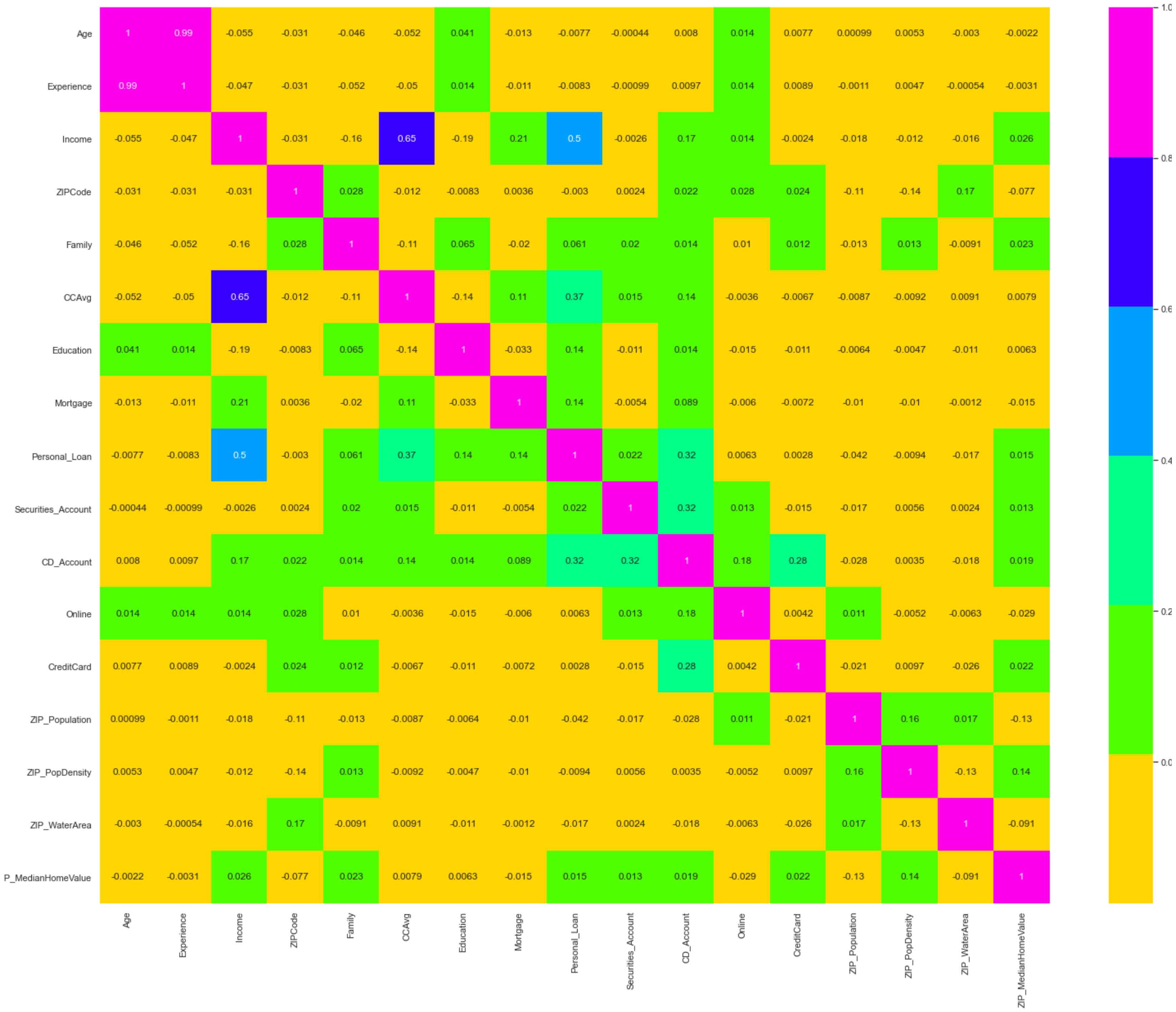
• A family size of 1 is most common

• Most customers have only an undergraduate education









EXPLORATORY DATA ANALYSIS:

BIVARIATE CORRELATION

Observations:

- Age and Experience are very highly correlated
- This means that only one of these categories (either Age or Experience) will be of much use to us in our predictions
- The next highest correlation (0.65) is between Income and CCAvg, indicating that people who make more money spend more on their credit cards each month
- After that, the next highest correlation is between Personal_Loan and Income (0.5)
- The correlation between Personal_Loan and Income is significant, as Personal_Loan is exactly the category that we are trying to predict; Keep an eye out for Income, as this will probably come back
- Other correlations to notice are Personal_Loan vs. CCAvg (0.37) and Personal_Loan vs. CD_Account (0.28)

EXPLORATORY DATA ANALYSIS: **PERSONAL LOAN**

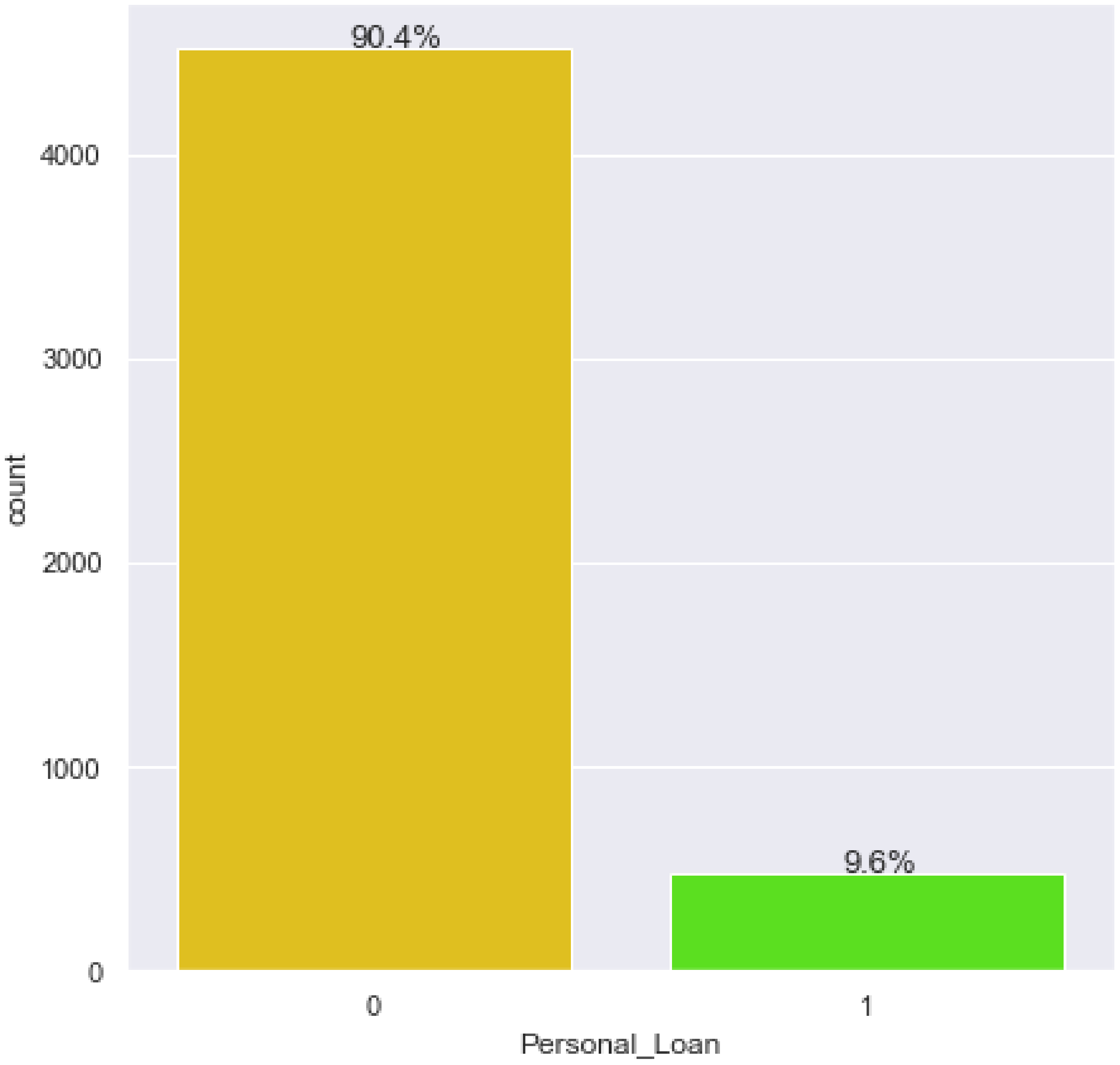
Personal Loan is the category that shows AllLife customers that accepted the loan offered in the last campaign. This is the most important variable in the data, because this is what we wish to predict.

Observations:

• We can see that 9.6% of customers accepted the loan, compared with 90.4% of customers that did not accept the loan.

Coming Up:

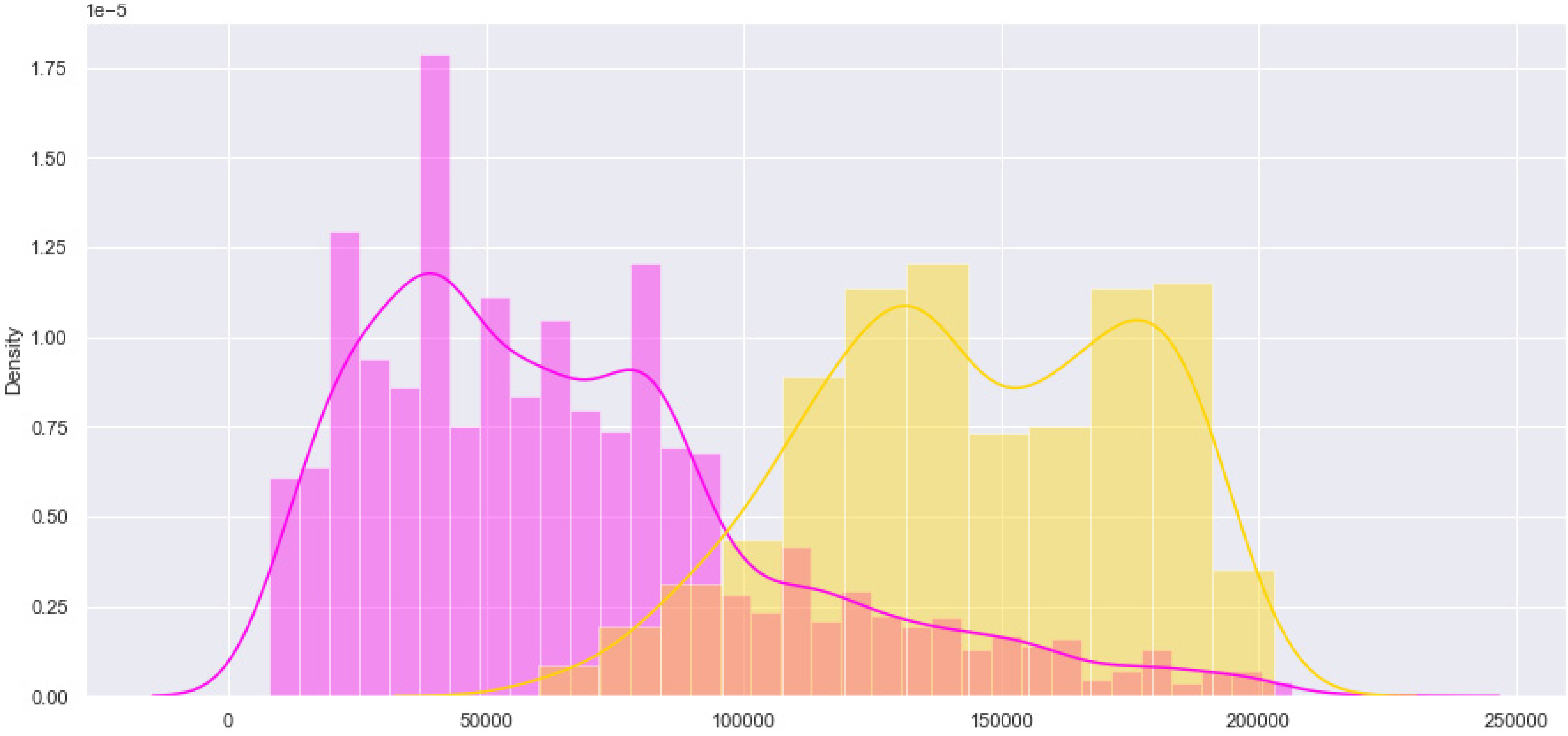
• In the coming pages, I will show a bivariate analysis showing the relationship between Personal Loan and the other variables.



EXPLORATORY DATA ANALYSIS: **PERSONAL LOAN VS. INCOME**

Observations:

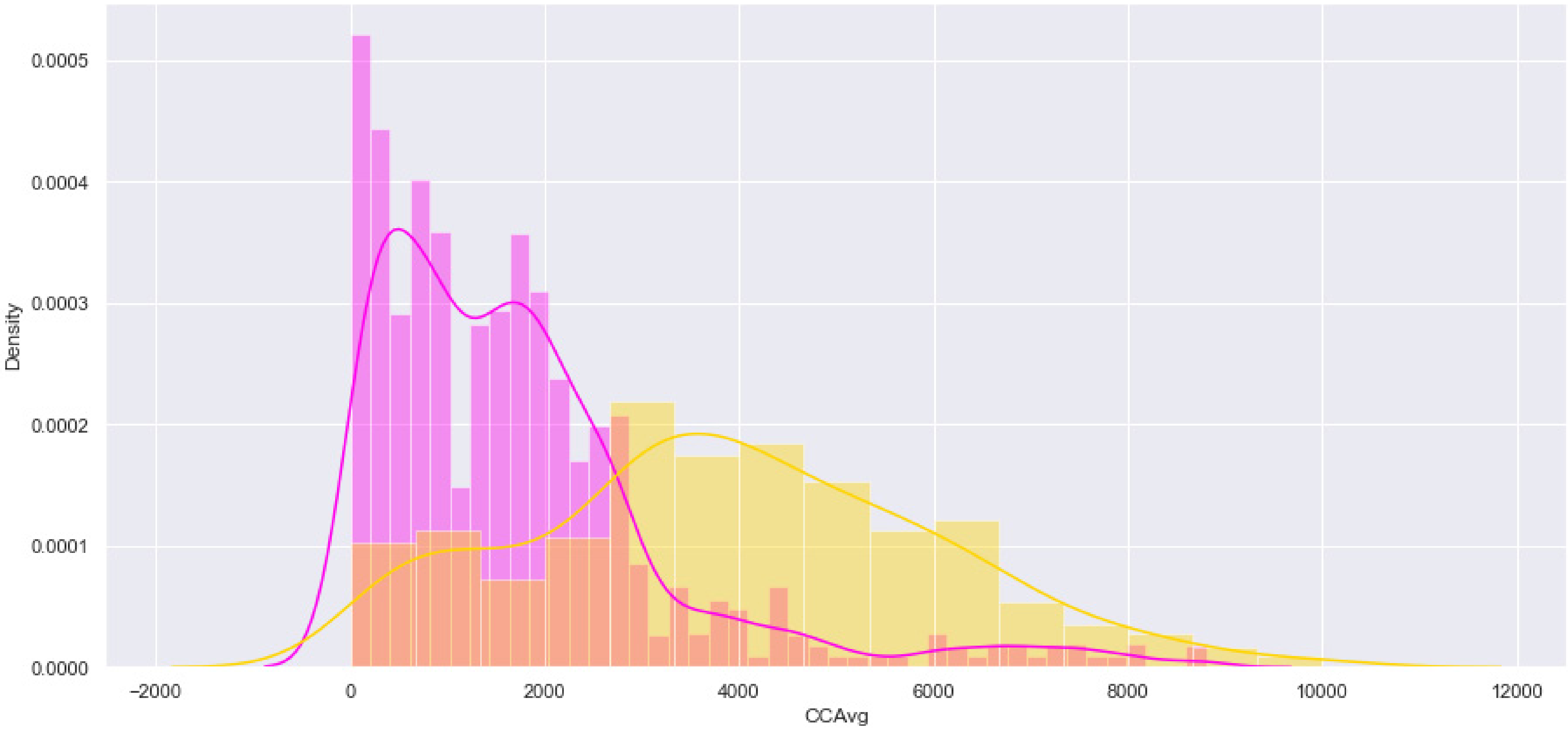
• The bars in pink are customers who did not take the loan offer, and the bars in yellow are those who took the loan offer are represented by the yellow bars
• This graph shows that Income is a significant variable
• Most of the customers who took the loan have an income over $100,000 per year



EXPLORATORY DATA ANALYSIS: **PERSONAL LOAN VS. CREDIT CARD AVG**

Observations:

- The bars in pink are customers who did not take the loan offer, and the bars in yellow are those who took the loan
- This graph shows that CCAvg is another significant variable
- A large amount of those that took the loan had higher amounts of monthly credit card spending



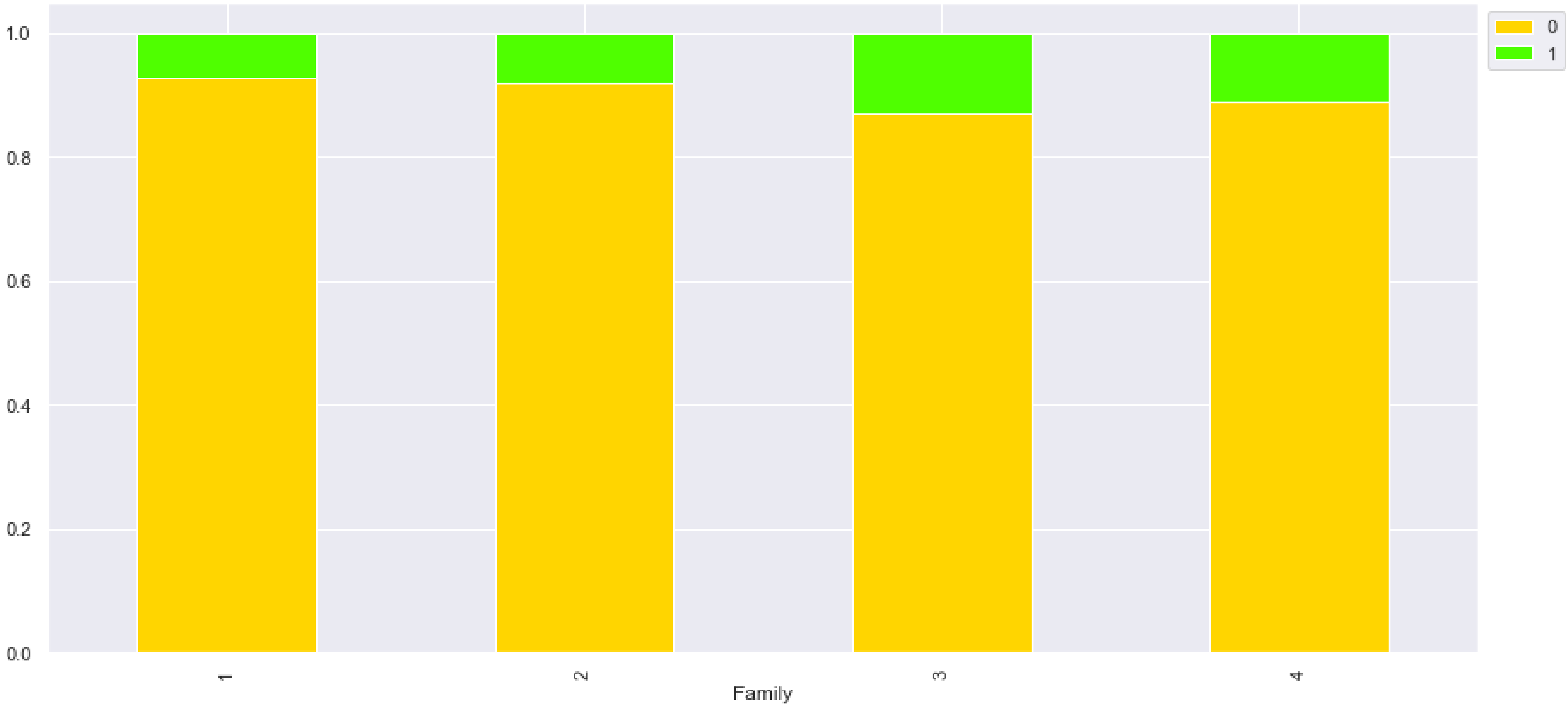
EXPLORATORY DATA ANALYSIS: **PERSONAL LOAN VS. FAMILY SIZE**

Observations:

• The yellow portion of the bars indicate the percentage of customers who did not take the loan offer, and the bars in green are those who took the loan offer

• While it may not appear to be a huge difference between the bars, the difference is large enough to be significant

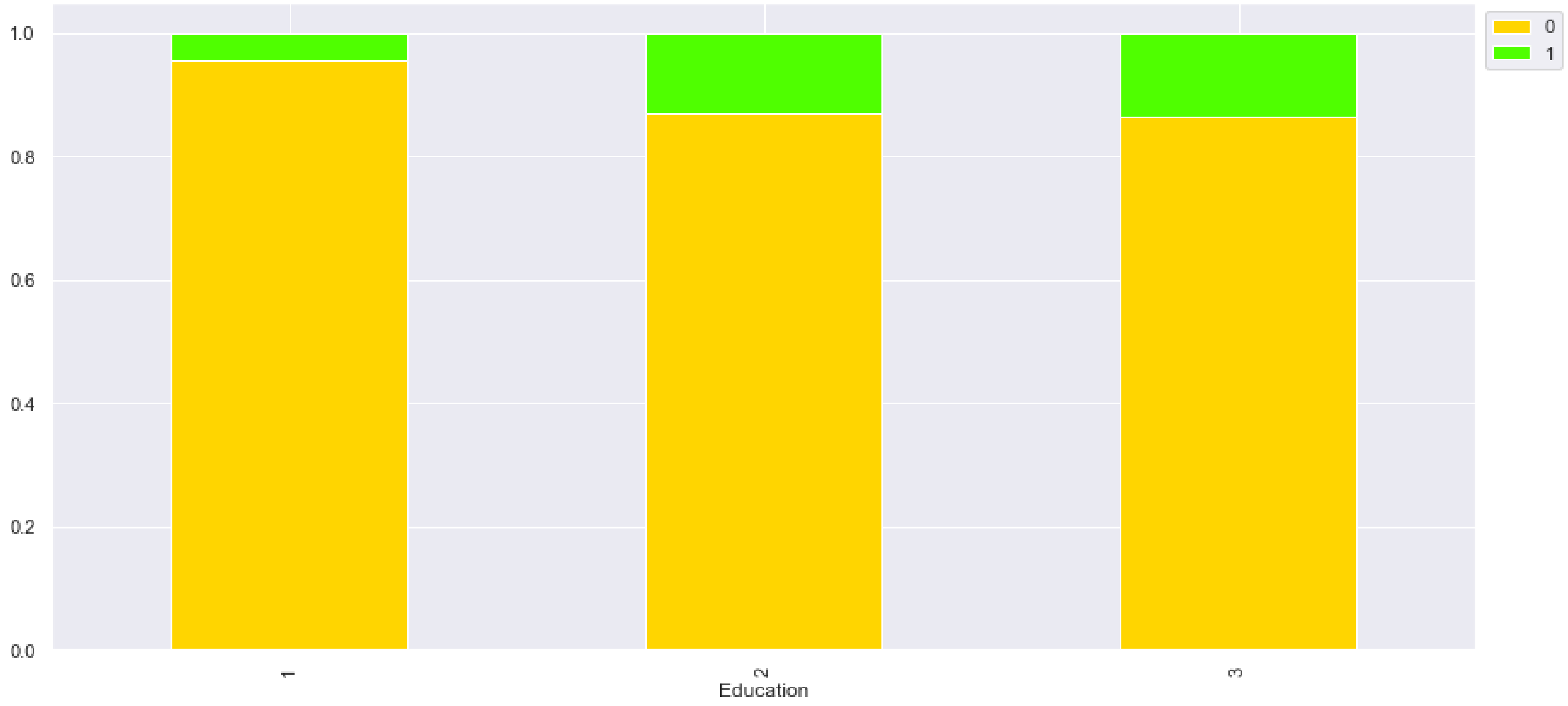
• Customers with larger families (3 or 4 people) accepted the loan more often than smaller families (1 or 2 people)



EXPLORATORY DATA ANALYSIS: PERSONAL LOAN VS. EDUCATION

Observations:

• The yellow portion of the bars indicate the percentage of customers who did not take the loan offer, and the bars in green are those who took the loan offer
• Again, while the difference between the bars may not look very big, it is large enough to be significant
• Customers with more education accepted the loan more often than those with only an undergrad degree

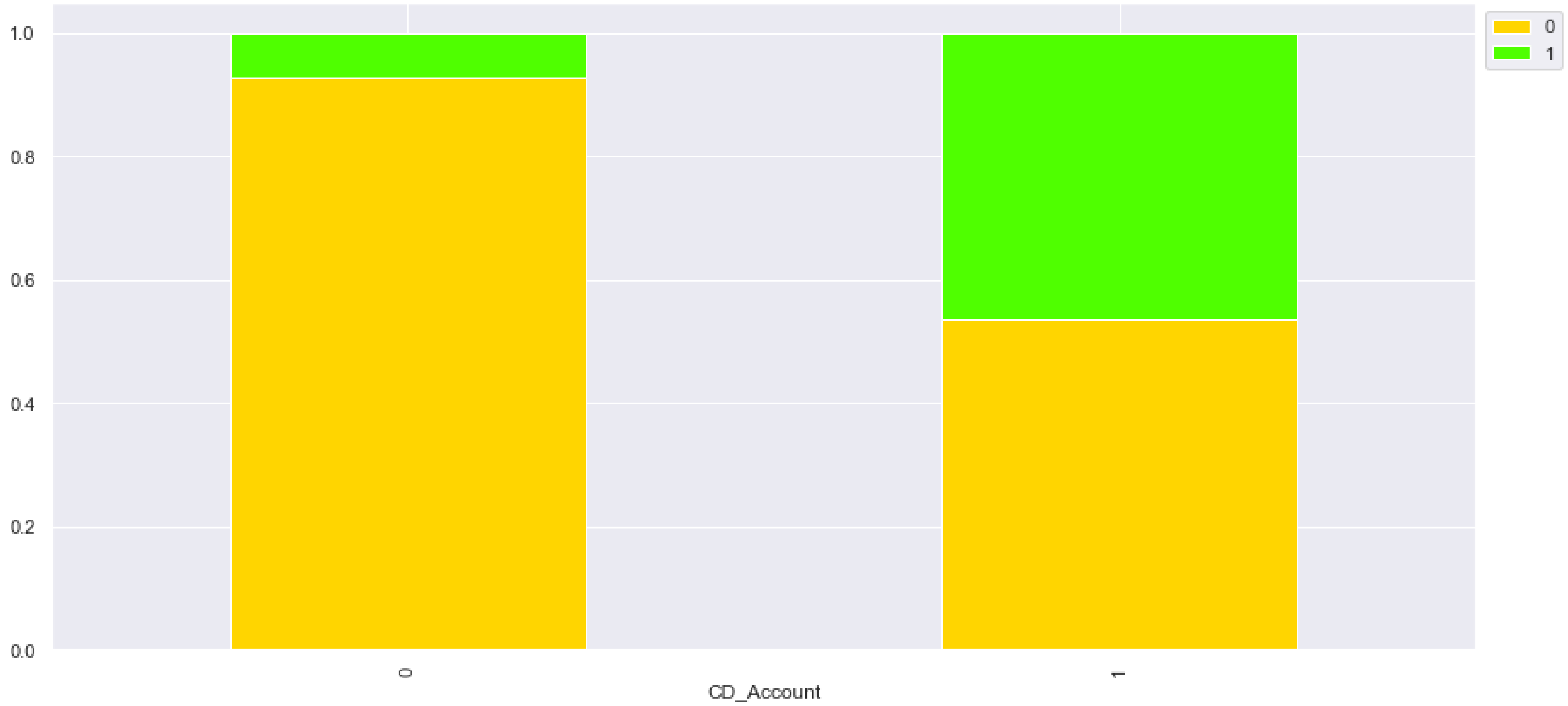


EXPLORATORY DATA ANALYSIS: **PERSONAL LOAN VS. CD ACCOUNT**

Observations:

• The yellow portion of the bars indicate the percentage of customers who did not take the loan offer, and the bars in green are those who took the loan offer

• Customers with a CD account accepted the loan more often than those without a CD account



SOLUTION: **MODEL OVERVIEW**

Approach:
• I built seven different models, looking for one that best solved our particular business problem
• The model was built to predict whether a liability customer will buy a personal loan or not
• The metric that I am judging the models on is Recall
• The model with the recall number closest to 1.00 is what we are looking for in this situation

MODEL NAME	RECALL ON TEST DATA
1. Logistic Regression Model - Sklearn	0.55
2. Logistic Regression Model - Statsmodels	0.58
3. Logistic Regression Model - Optimal Threshold .05	0.93
4. Logistic Regression Model - Optimal Threshold .34	0.65
5. Decision Tree Model - No Pruning	0.83
6. Decision Tree Model - with Pre-Pruning	0.86
7. Decision Tree Model - with Post-Pruning	0.99

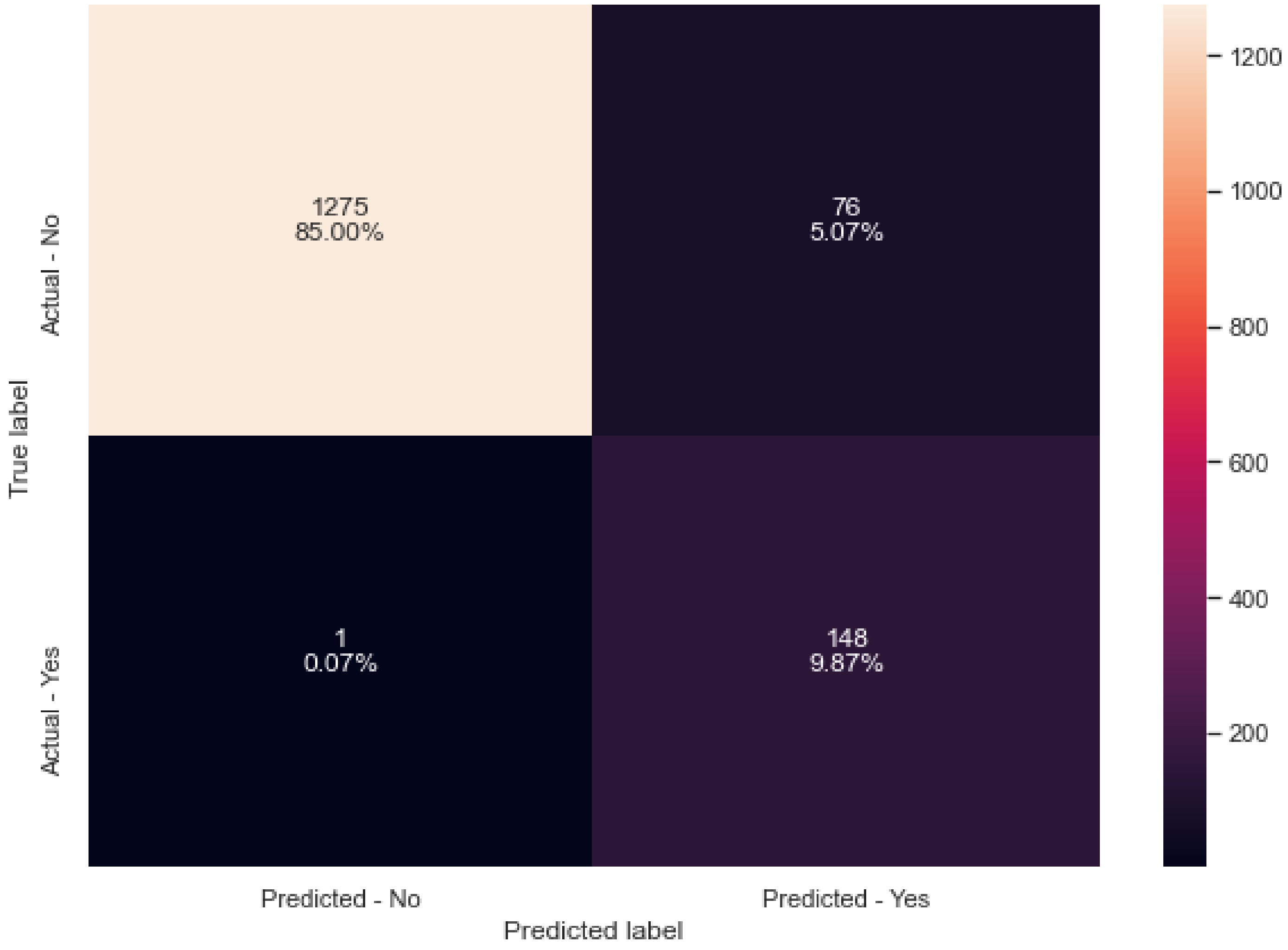
Solution:
• Model 7, Decision Tree Model with Post-Pruning performed very well on the recall metric (0.99)
• This is the model that I recommend AllLife Bank uses
• I will be present this model in more detail in the coming pages

MODEL: PERFORMANCE SUMMARY

• As mentioned, the chosen model had a testing recall of 0.99
• The confusion matrix to the right shows the models performance on Type I and Type II errors

Error Summary:
• False Positive (FP) = 5.07%: The model predicted that a customer would take a Personal Loan when they did not
• False Negative (FN) = 0.07%: The model predicted that a customer would not take a Personal Loan when they did in fact take the loan

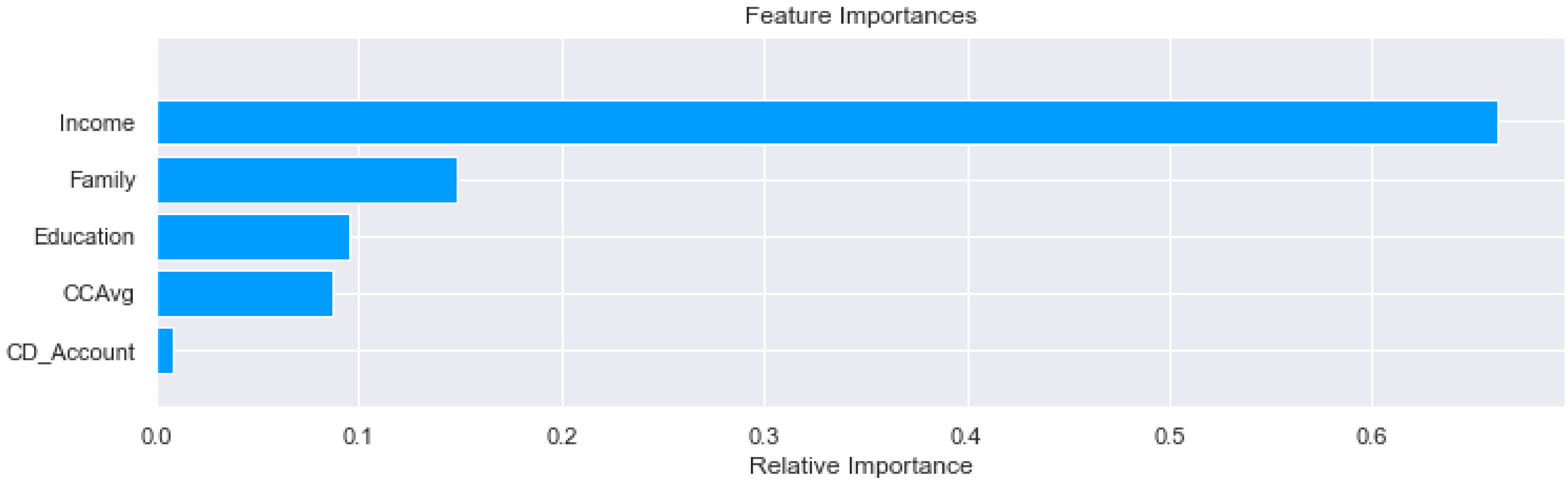
What kind of losses do we face if we make the wrong prediction?
• A false negative would means that we predict a customer would not take a Personal_Loan when in fact they would. This would result in loss of revenue for the bank.
• A false positive would means that we predict a customer would take a Personal_Loan when in fact they would not. This would result in a loss of marketing resources, as the bank would market a loan to a customer who is unlikely to accept the loan.
• A loss of revenue is a bigger loss to the bank than a loss in marketing resources.
• In this case, we'd rather have some false positives than false negatives



MODEL: SIGNIFICANT FEATURES

Key Takeaways:

• Income is the most important feature that our model uses to predict whether a customer will accept a personal loan or not
• The other features that are important to the model are Family, Education, Avg Credit Card Spending, and CD Account
• Features that do not help in our prediction are Age, Experience, ZIP Code, Mortgage, Securities Account, Online, and Credit Card from a different bank



MODEL: IDENTIFYING TARGET SEGMENTS

Which segment of customers should be targeted?
• Using the model will be more precise in targeting customers, but here are some general guidelines of the types of customers that are more likely to take a loan:

Income
• Primarily target customers with a yearly income over $92,500

Secondary factors
• Family Size: target customers with a family size greater than 2

• Credit card Average: target customers that spend more than $2850 a month on their credit cards

• Education: target customers with graduate or advanced/ professional levels of education

• CD Account: target customers with a CD Account

KEY TAKEAWAYS FOR THE MARKETING TEAM

• The model presented here was created based on the types of customers who accepted the loan last time around. With that in mind, the upcoming campaign should be similar to the past campaign because it was proven to work for this demographic in the past.

• Again, we can put this model into production to predict if a specific customer will accept a loan or not. The potential benefits of using this model is that the results apply specifically to AllLife's customer data. The marketing message will be more likely to reach potential personal loan customers. The true potential benefit is business growth.

• When designing messaging and marketing materials around specific demographics, refer to the target segments on the previous page for guidance. To reiterate, the target markets are customers with yearly incomes over $92,500, with family sizes greater than 2, those that spend more than $2850 a month on their credit cards, have a graduate or advanced degree, and/or have a CD Account.

• In addition to creating marketing materials that appeal to these customers, we will want to make sure that the marketing message reaches the target audience. I know that this is what the marketing team does best, so I don't want to step on anyone's toes, but here are a couple ideas for marketing placement.

• If the marketing campaign has a television advertising component, buy ad time on shows with where the viewership is known to be of higher income, geared towards families, or known to appeal to those with a higher level of education.

• Another idea would be to mail personal loan offers to CD account customers and to credit card customers who spend over $2850 per month.

• Perhaps the loan offers themselves (loan types, rates, or terms) can be tailored even further to the target segments,

OTHER ADVICE ON HOW TO GROW THE BUSINESS

• Again, the model presented here was created based on the types of customers who accepted the loan last time around. This model should be able to help target these types of customers, aiding in the expansion of its asset customers. I believe that this is a good strategy to grow business profits through interest on loans.

• That said, there is potential economic benefit in targeting the inverse segment as well. This model has also identified the types of customers who DID NOT take the loan last time. My advice to the banking team would be to create alternative types of loans that may be more beneficial to these other segments of customers. Here are some further ideas on this concept:

• We saw that customers that made higher amounts of money were more likely to take the loan offer. Perhaps there are smaller, more specialized personal loans that could be marketed to people who make less that $90,000 per year? For example, maybe you could offer loans for computers, electronics, home appliances, or cars to those customers.

• We saw that customers that had larger families were more likely to take the loan offer. Perhaps there are loan types that are more applicable to single customers?

• We saw that customers that had advanced degrees were more likely to take the loan offer. Maybe the bank could offer educational loans to those with only a bachelor's degree?

• One other consideration is that the past marketing campaign itself did not appeal to lower income customers. It could be worth exploring how to tailor a campaign that is directed towards this other demographic.

THANK YOU

PERSONAL LOAN CAMPAIGN MODEL

by **JAKE EIDE**