Loan Campaign Modelling Project

Contents

- Background & Problem Statement
- Data Dictionary
- Exploratory Data Analysis Highlights
- Model Performance Summary
- Recommendations

Background & Problem Statement

There is a growing customer base for AllLife Bank, especially in liability customers.

The management wants to explore ways of converting its liability customers to personal loan customers.

while at the same time, retaining them as depositors

We require a model that can predict potential customers who have a higher probability of purchasing the loan.

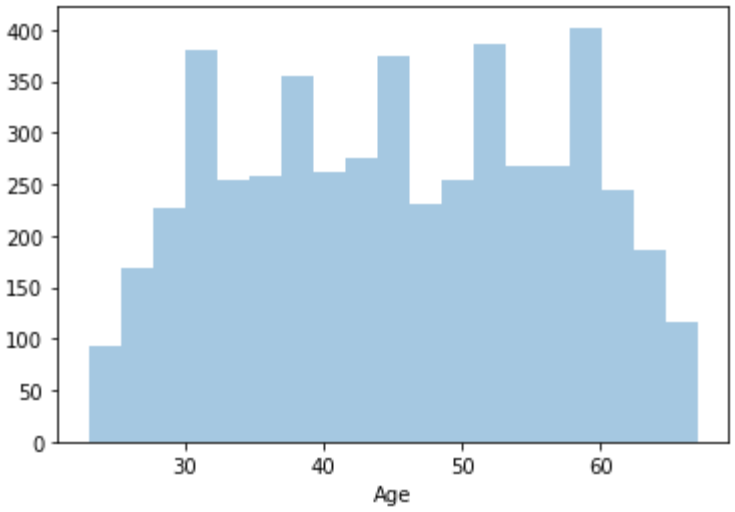
Here, logistic regressions and decision trees will be used.

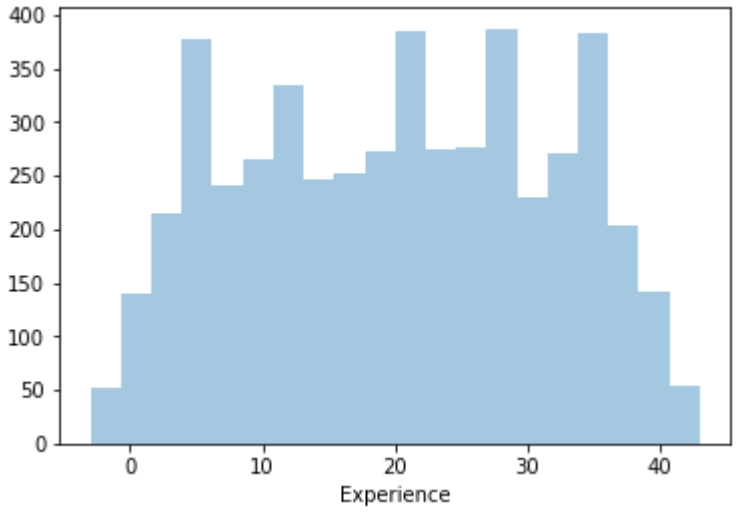
Data Dictionary

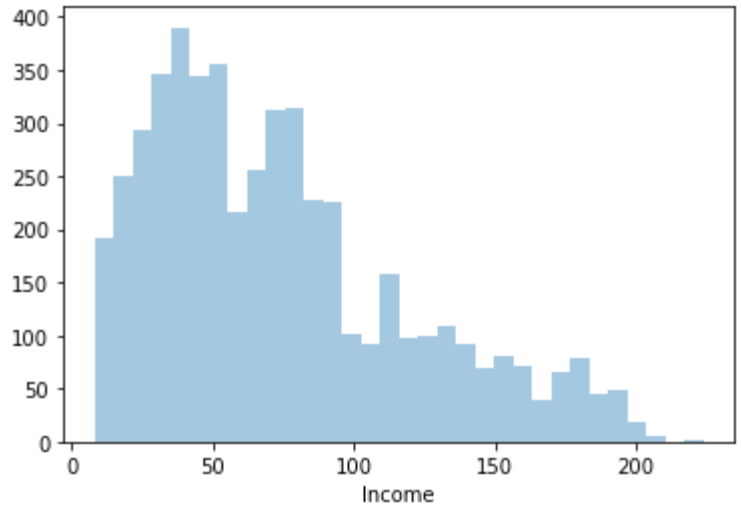
The data contains information about 5000 customers and their characteristics. No further pre-processing is needed to the data, other than erroneous negative values in work experience.

Variable	Description
ID	Customer’s ID
Age	Customer’s age in completed years
Experience	Number in years of professional experience
Income	Annual income of the customer (in thousand dollars)
ZIP Code	Home Address ZIP code
Family	Family size of the customer
CCAvg	Average spending on credit cards per month (in thousand dollars)
Education	Education Level (1: Undergrad; 2: Graduate; 3: Advanced/Professional)
Mortgage	Value of house mortgage if any (in thousand dollars)
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_Account	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 Highlights

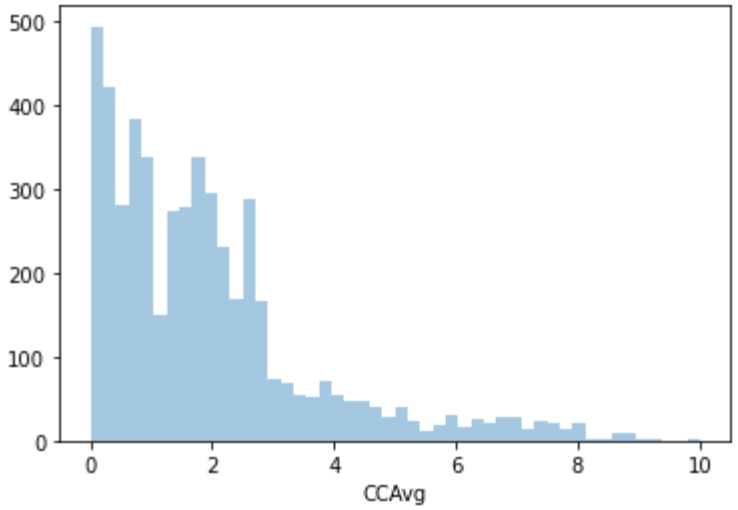


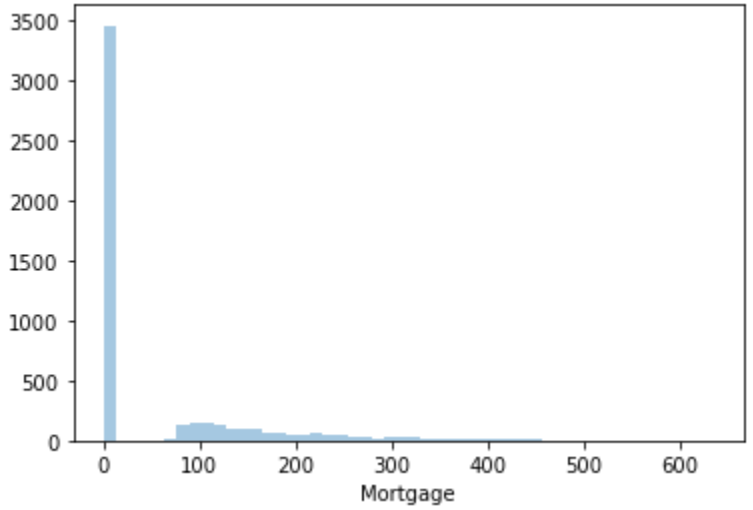


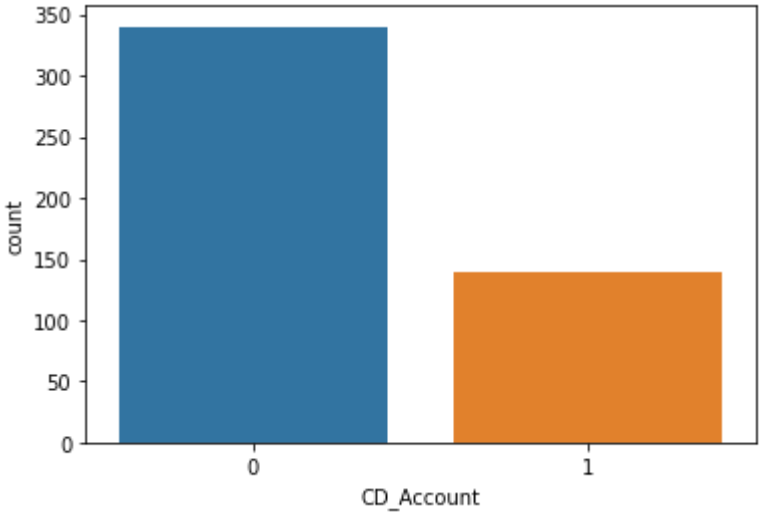


- Age and Experience have a similar distribution
- Income is right skewed

Exploratory Data Analysis Highlights

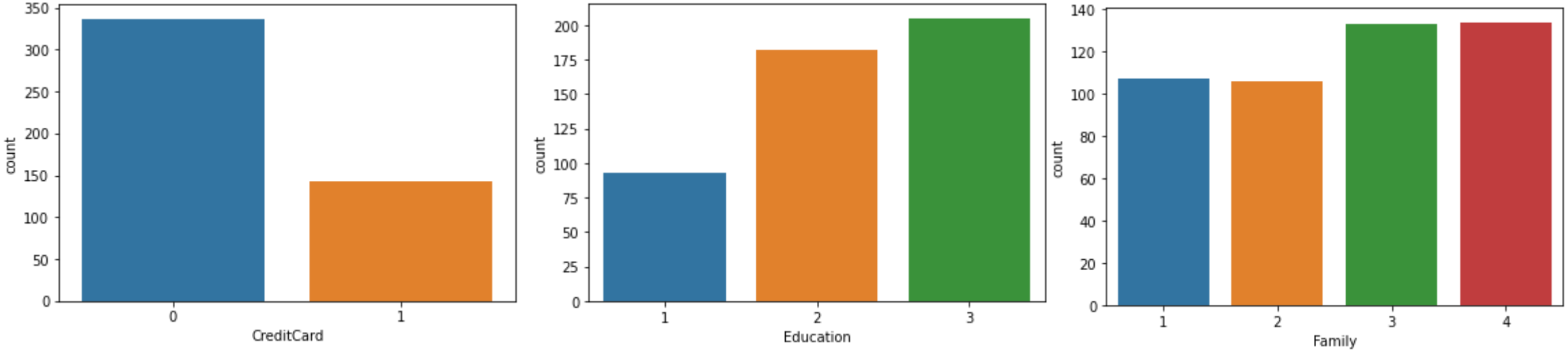






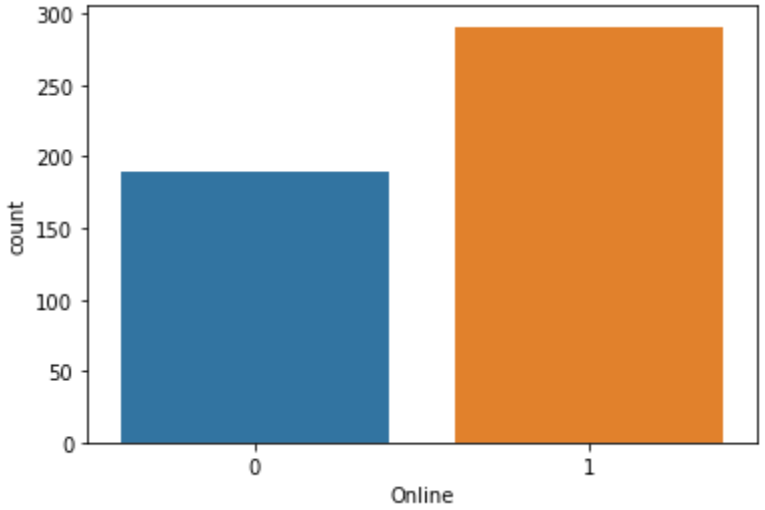
- CCAvg is right skewed
- Most people have 0 in mortgage amount
- Most of the customers don't have a certificate of deposit (CD) account with the bank

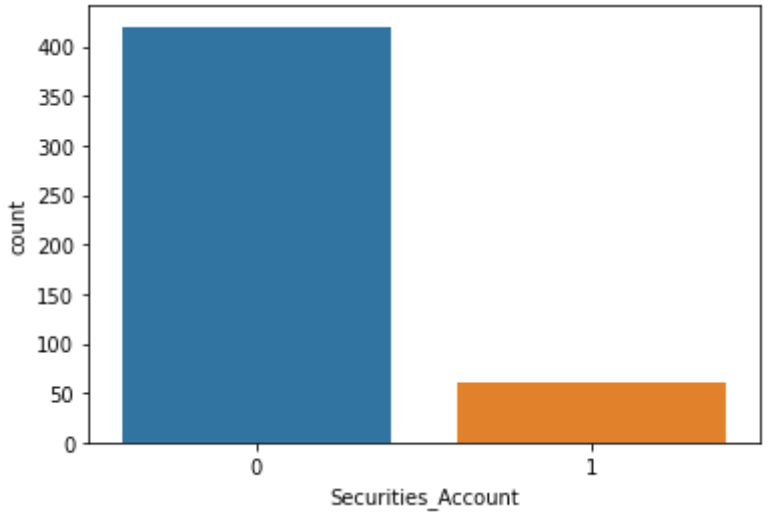
Exploratory Data Analysis Highlights



- Most of the customers do not use a credit card issued by any other Bank (excluding All life Bank)
- Most of the customers' last education are as undergraduates
- Most of the customers have only 1 other family member

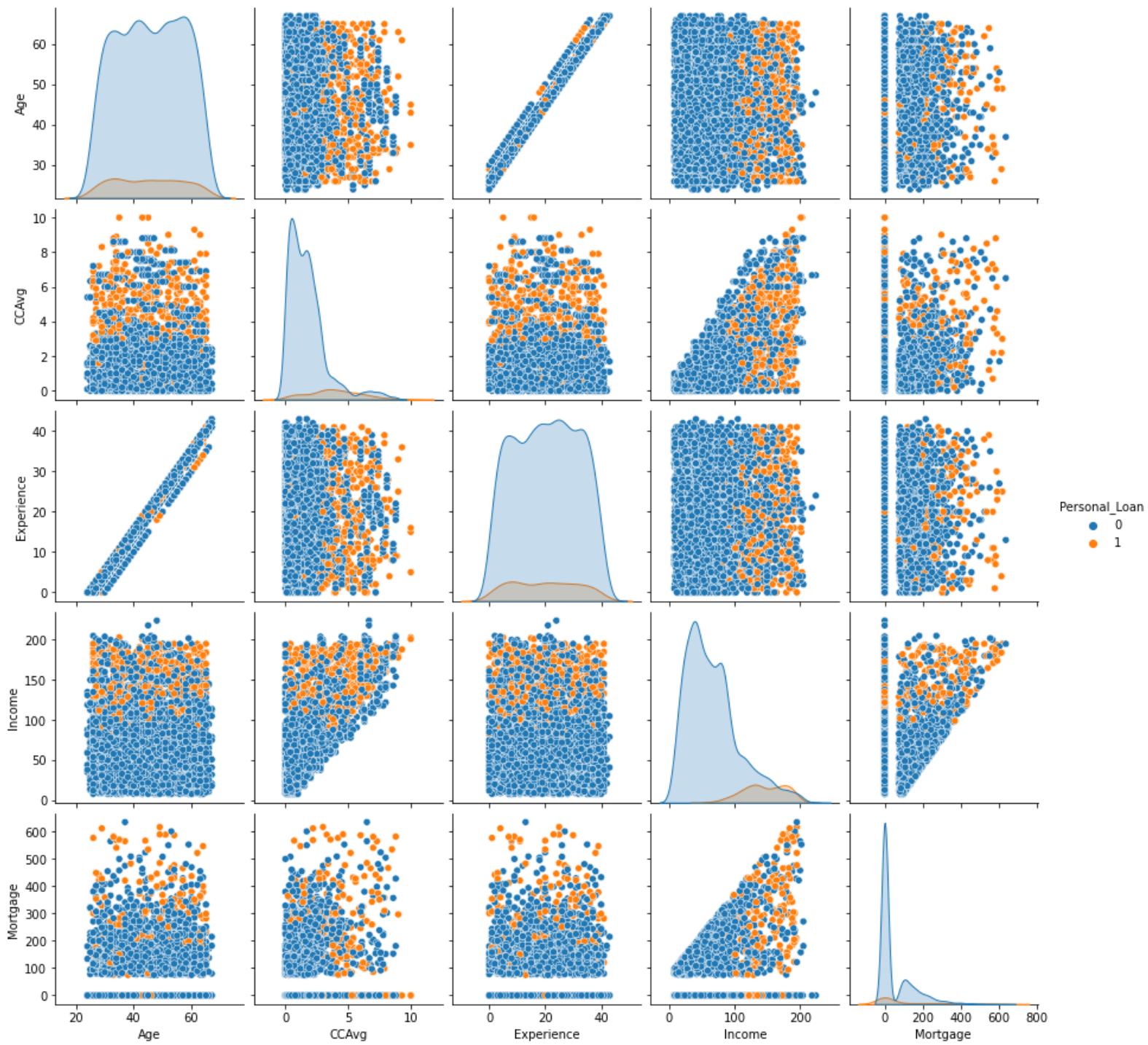
Exploratory Data Analysis Highlights



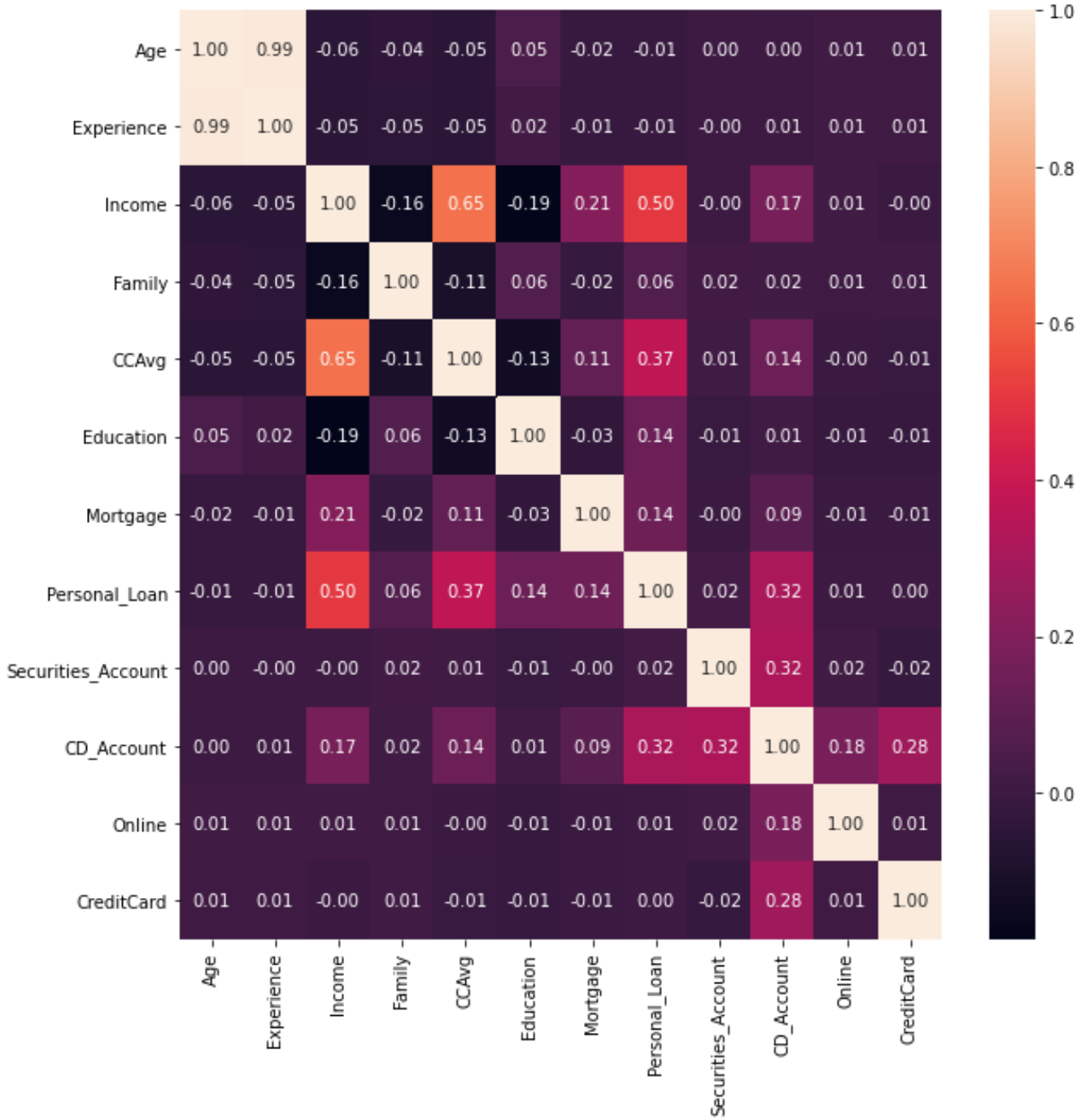


- There are more customers using internet banking facilities
- Most of the customers don't have security account with the bank

Exploratory Data Analysis Pairplot



Exploratory Data Analysis Heatmap



Exploratory Data Analysis Highlights

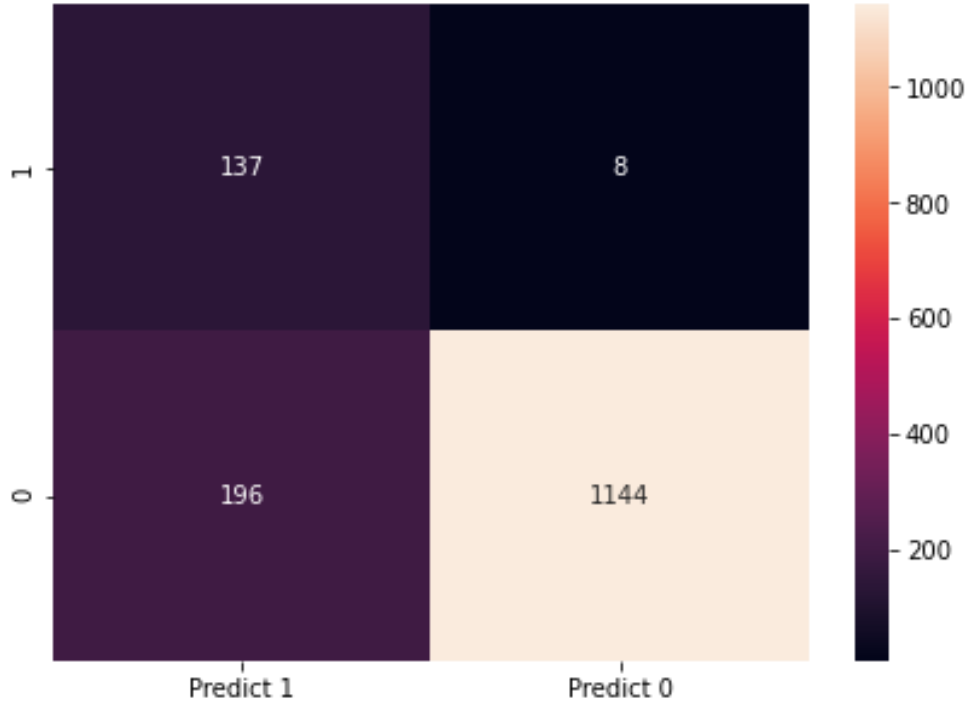
- CCAvg somewhat correlates with Income.
- The target variable Personal_Loan is mostly correlated to Income, followed by CCAvg and then CD_Account.
- Age and experience correlate strongly
- From bivariate analysis,
  - Those who accepted the loan are people who have higher income levels
  - Those who accepted the loan has somewhat higher CCAvg values

Model Performance Summary

- Both logistic regression and decision tree models are applied, with 70% of data used as training data and 30% of data used as testing data
- Recall is used as the main indicator of the models, as we want to avoid having false negatives, which severely impact the company
- The experience column is removed due to its very high correlation to the age column, leading to multicollinearity

Logistic Regression

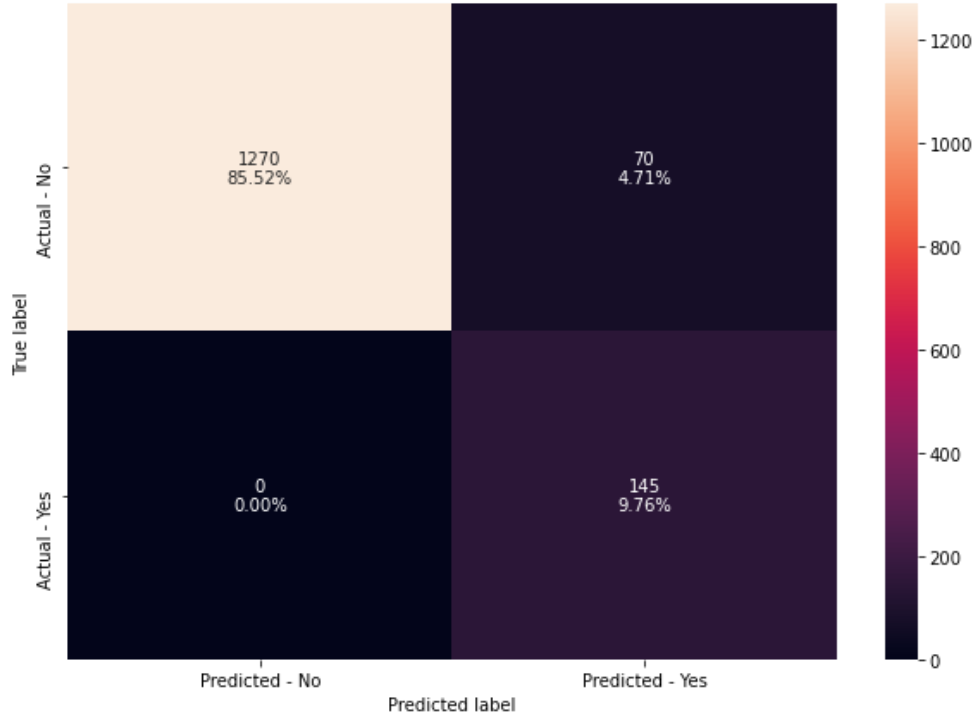
- Among various logistic regression models applied, logistic regression model with threshold value adjustment is the most ideal model that can be used due to a high value in recall.



Model	Recall
Basic Logistic Regression Model	0.58
Logistic Regression Model with Threshold Value Adjustment	0.95
Logistic Regression with Statsmodel	0.68

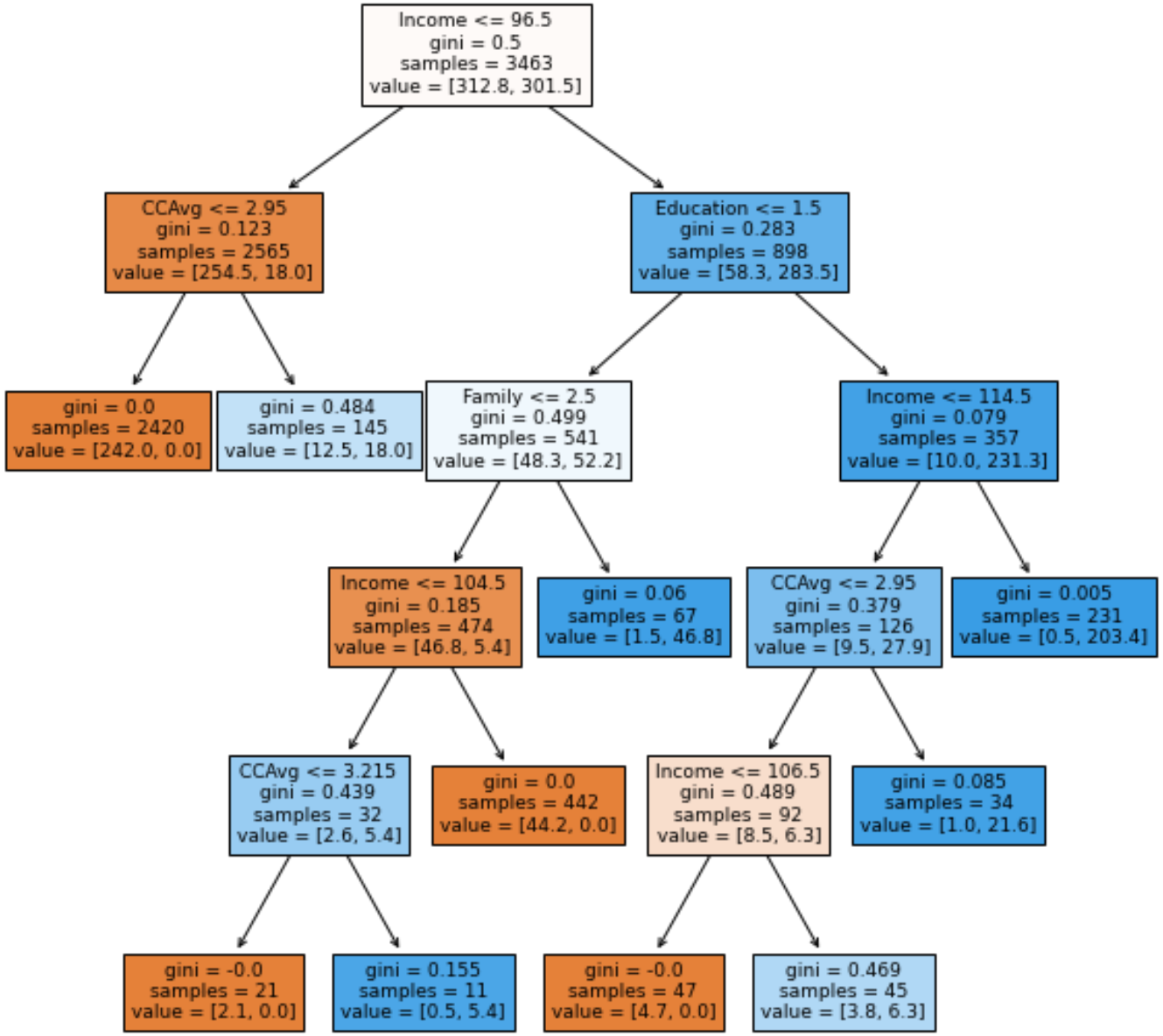
Decision Tree

- Among various decision tree models applied, decision tree with post-pruning has the highest recall and hence is the most ideal model that can be used for prediction.



Model	Train_Recall	Test_Recall
Basic Decision Tree Model	1.00	0.87
Pre-pruned Decision Tree with max depth of 3	0.84	0.83
Pre-pruned Decision Tree with Hyperparameter Tuning	0.98	0.96
Decision Tree with Post-pruning	1.00	1.00

Best Decision Tree Model



Most Important Predictors

1st : Income

2nd : Family

3rd : Education

4th : CCAvg

Recommendations to Marketing Team (ordered by priority)

1. To target people with higher income, with at least around 130
2. To target people with another family member (married couples, with or without children)
3. To target people with the education level as a graduate or advanced/professional
4. To target people with higher level of average credit card spending/month, at least 2.5 or above
5. To target people with no mortgage value