

PERSONAL LOAN CAMPAIGN



Contents

- ✓ All Life Bank is a US bank that has a growing customer base. The majority of these customers are liability customers (depositors) with varying sizes of deposits. The number of customers who are also borrowers (asset customers) 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.
- ✓ You as a Data scientist at All Life bank have to build a model that will help the marketing department to identify the potential customers who have a higher probability of purchasing the loan.

Objective

- To predict whether a liability customer will buy a personal loan or not.
- Which variables are most significant.
- Which segment of customers should be targeted more.

Business Problem Overview and Solution Approach

- AllLife Bank want to identify potential customers who want to buy a personal loan.
- Logistic Regression and Decision Tree Model we shall use , to predict the price of used cars
- Help the business in identifying potential customer , such that the overall profit of the bank will increase through the interest on the loan
- Logistic regression model will help us to identify the coefficients.

Data Overview

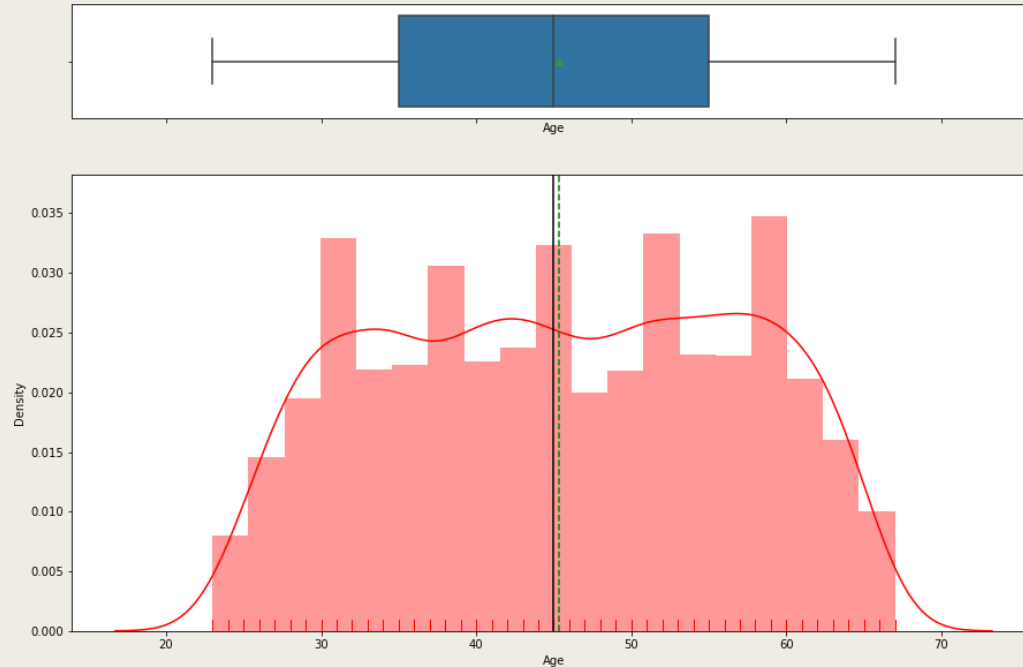
- * ID: Customer ID
- * Age: Customer's age in completed years
- * Experience: #years of professional experience
- * Income: Annual income of the customer (in thousand dollars)
- * ZIP Code: Home Address ZIP code.
- * Family: the Family size of the customer
- * CCAvg: Average spending on credit cards 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)?

Key Points –

- In the data set we have 5000 rows and 13 columns
- In the given dataset we do not have duplicate values
- We do not have any null values in the dataset
- Data Preprocessing is required for couple of variables like experience
- Out layers are supposed to be processed for better Recall and Accuracy values

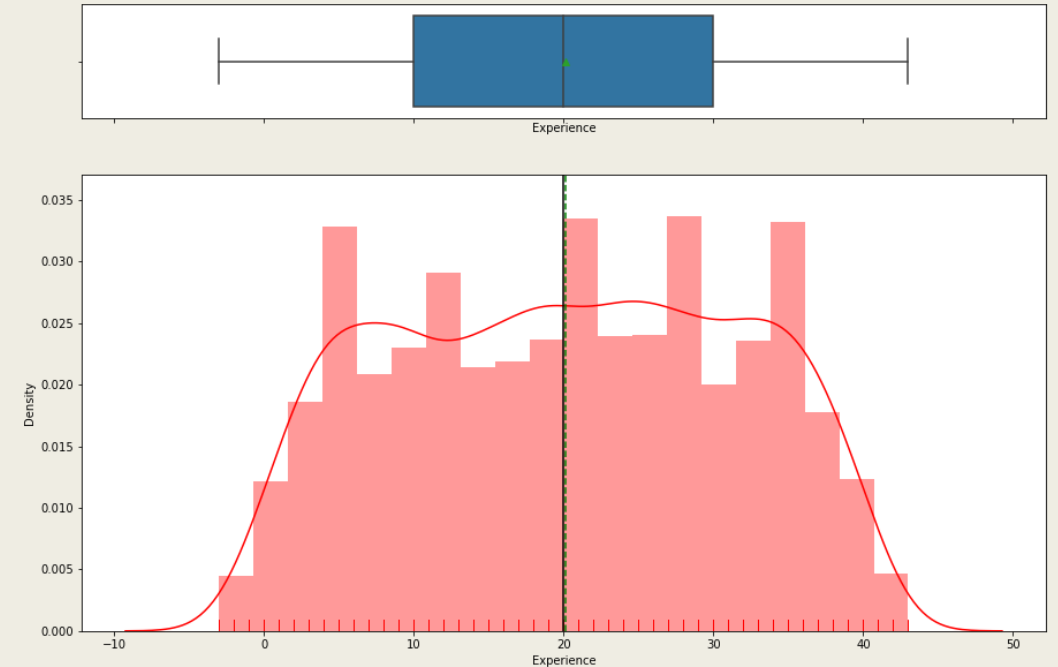
Exploratory Data Analysis

Age



- Minimum Age is 23 , where as maximum is 67
- Distribution of data does not look normal.
- Data transformation is required

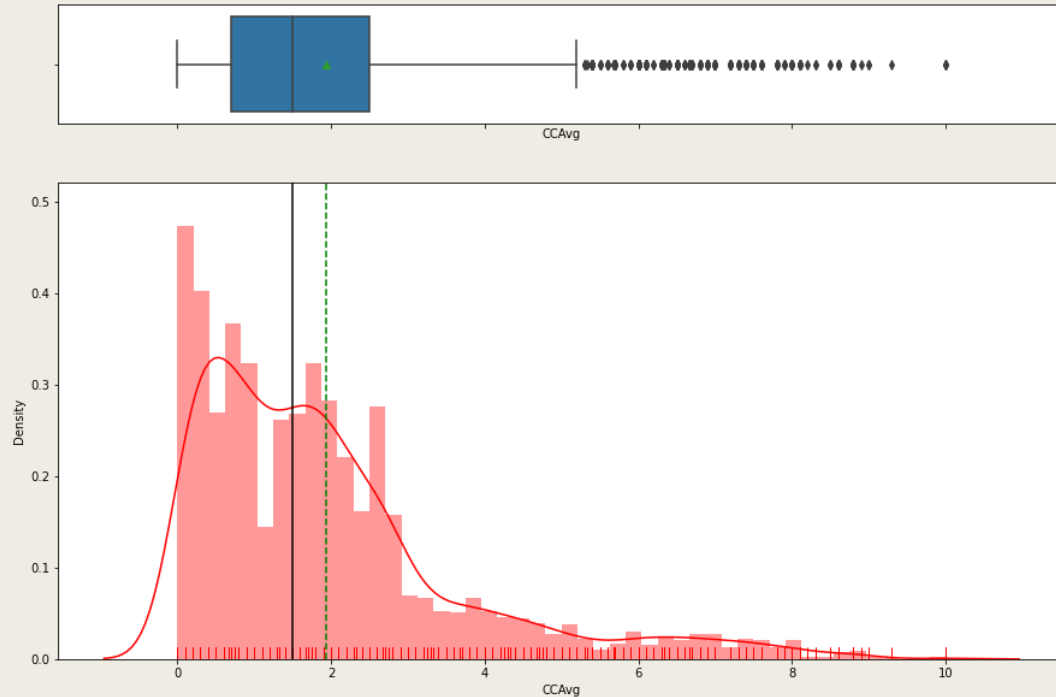
Experience



- Mean experience is 20 , In the given dataset we do have negative value those are supposed to be pre processed
- Max is 43 , date is not normally distributed.

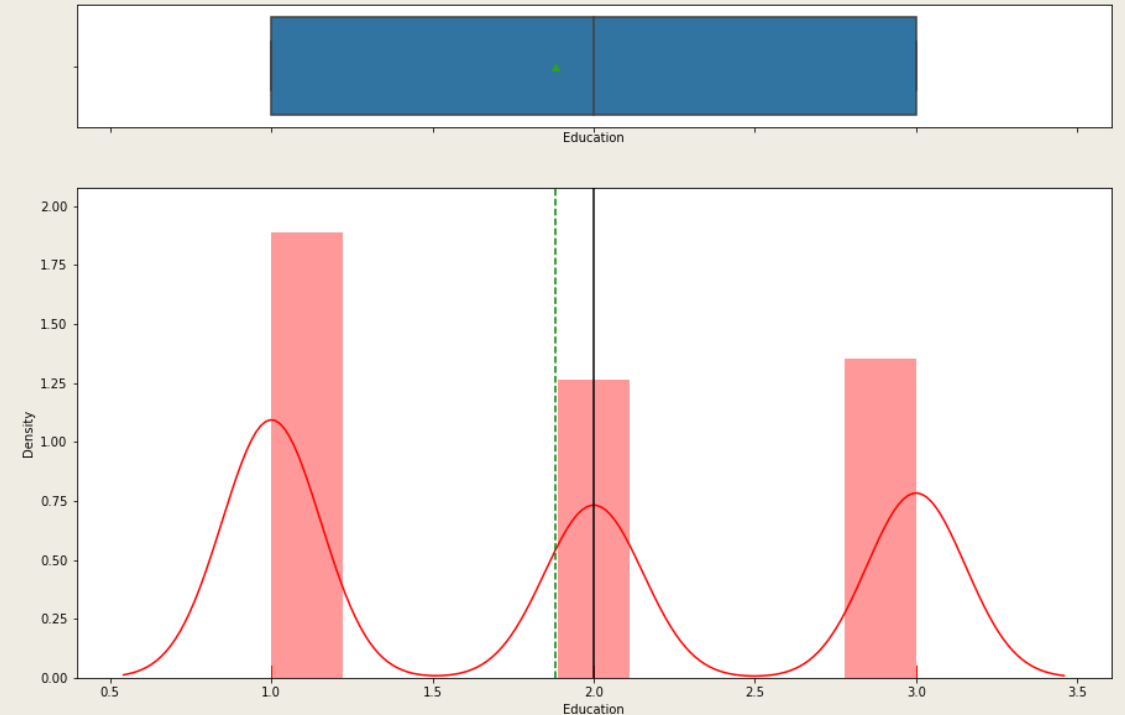
Exploratory Data Analysis

CC Avg



- Average spending on credit card is 1.9 K ,
Max value is 10K
- Data is left skewed

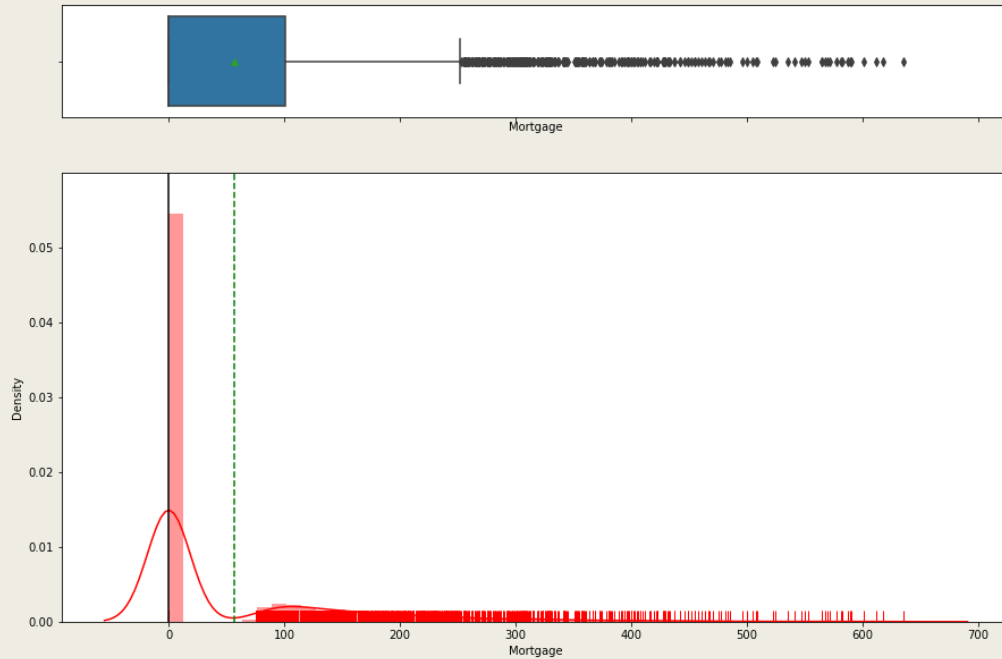
Education



- Minimum education is Undergrad
- People with Advanced/Professional degree are also available in the given dataset , but their volume is less.

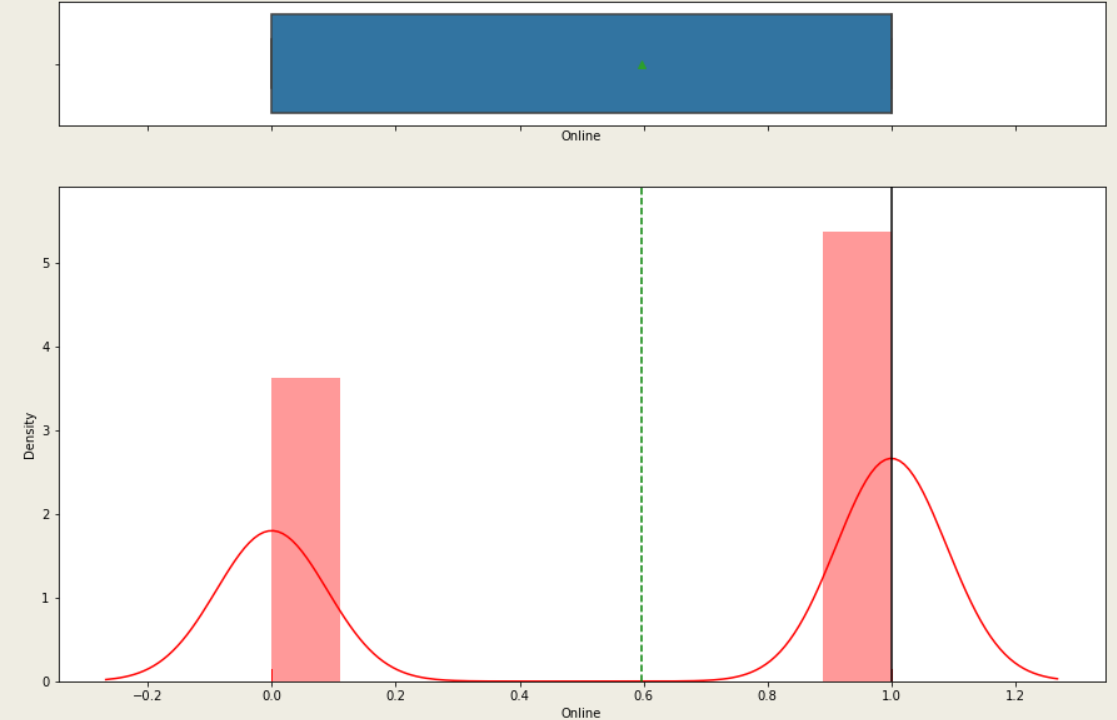
Exploratory Data Analysis

Mortgage



- Mean mortgage is 56.4K , Max is 635K
- We do have outliers in the dataset

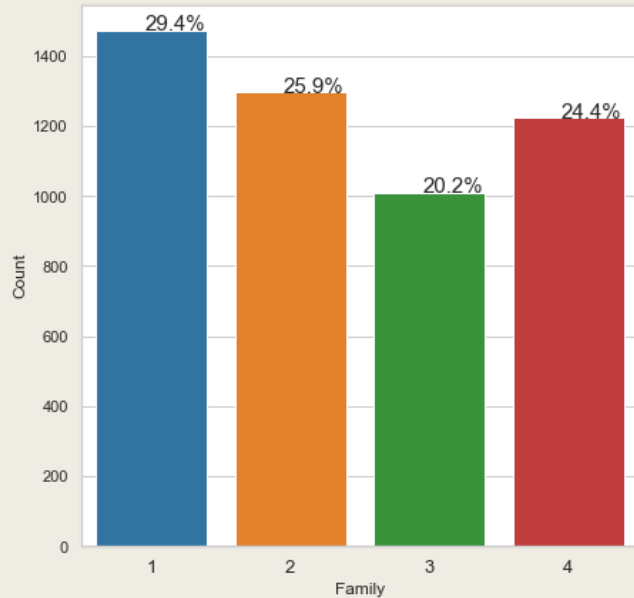
Online



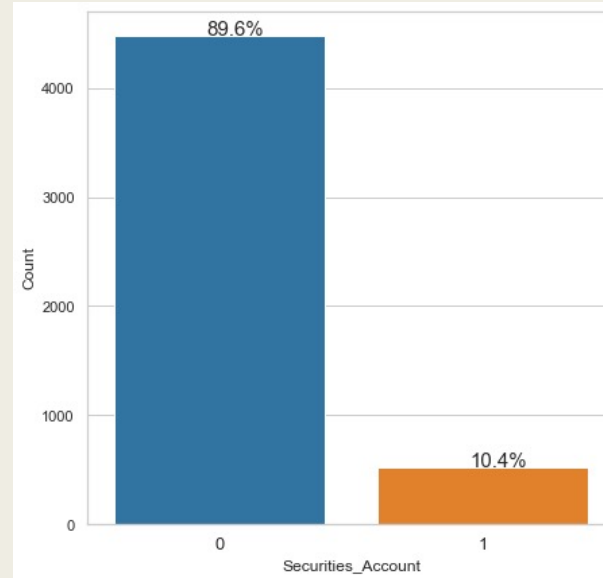
- Minimum education is Undergrad
- People with Advanced/Professional degree are also available in the given dataset , but their volume is less.

Exploratory Data Analysis

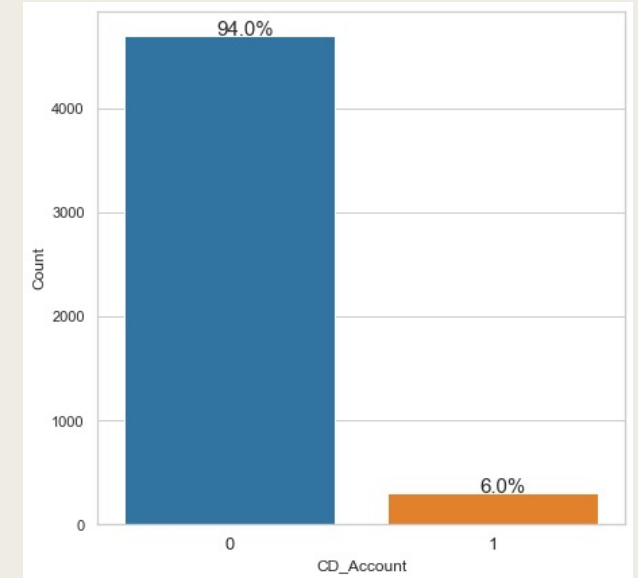
Family



Securities Account

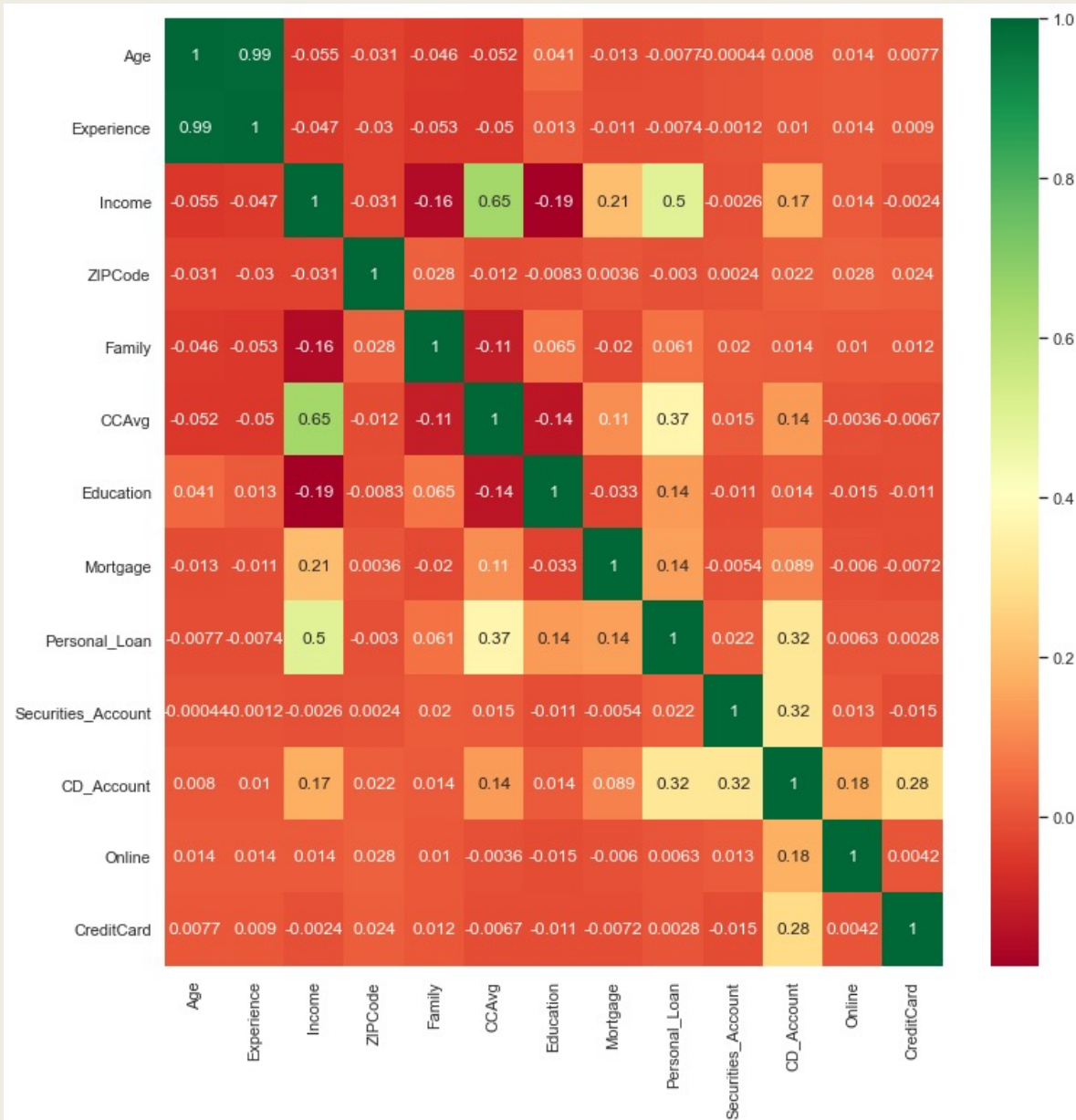


CD Account



- 70% of the people do use credit card issues by other bank and only 29.4% use the credit card of All Life Bank
- Internet banking is used by 60% of the people and 40% does not use internet banking
- 94% of the people does not have certificate of deposit and 6% only has CD.
- Almost 90% does not have securities account and only 10% has.
- Lot of scope on Personal Loan section as well , 90% did not take personal loan.

Exploratory Data Analysis - Correlation



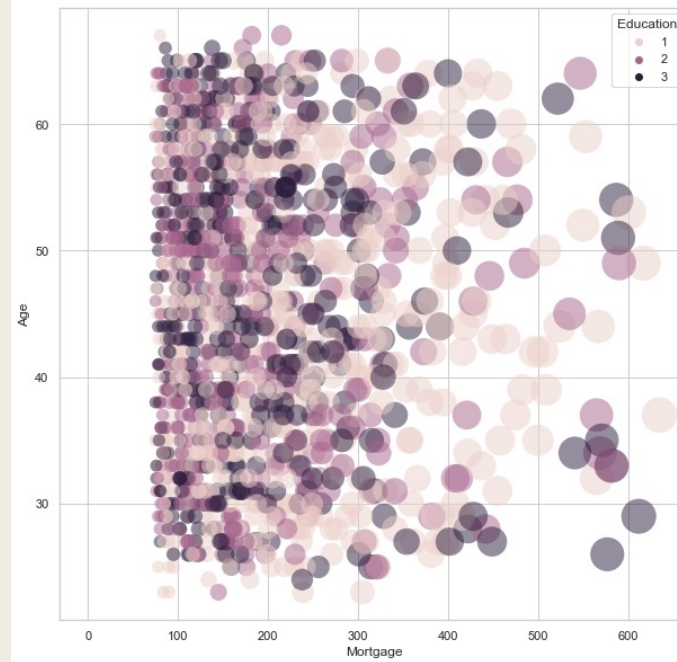
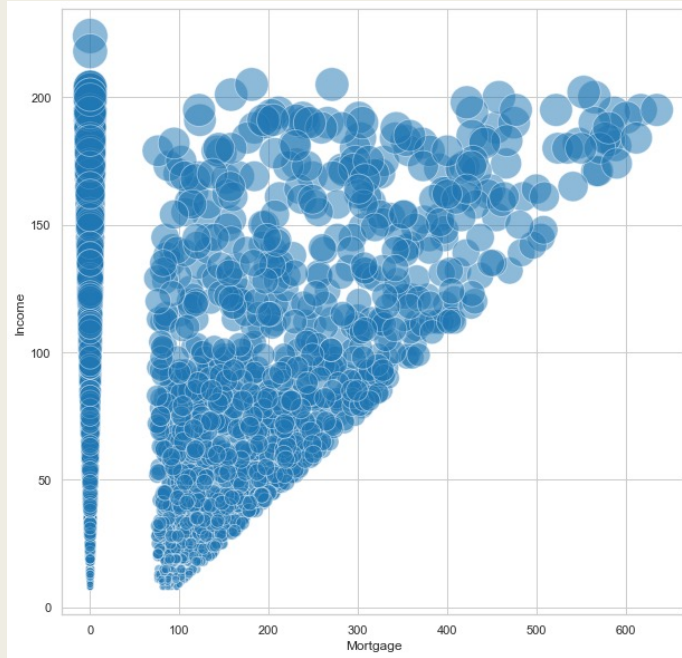
- ✓ Majority of the variables are not correlated.
- ✓ Income and Credit Card Average are correlated with a value of 0.65
- ✓ No need to remove any variable from the model , we do not have any variable with value more than 0.8.

Exploratory Data Analysis

Income vs Mortgage

Age vs Mortgage

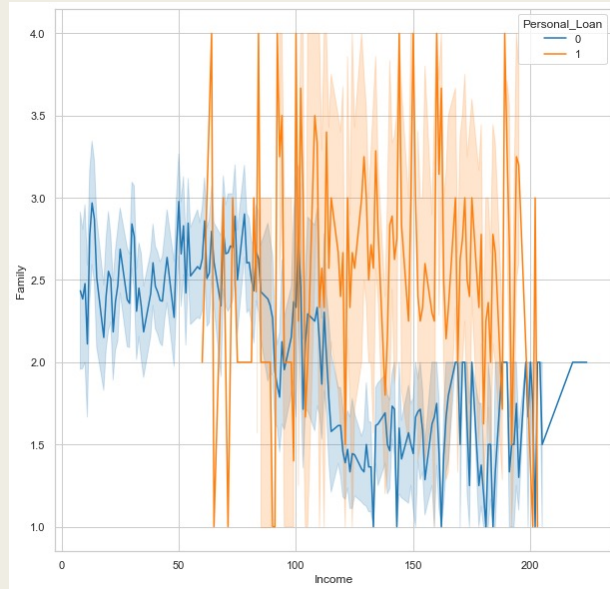
Experience vs Income



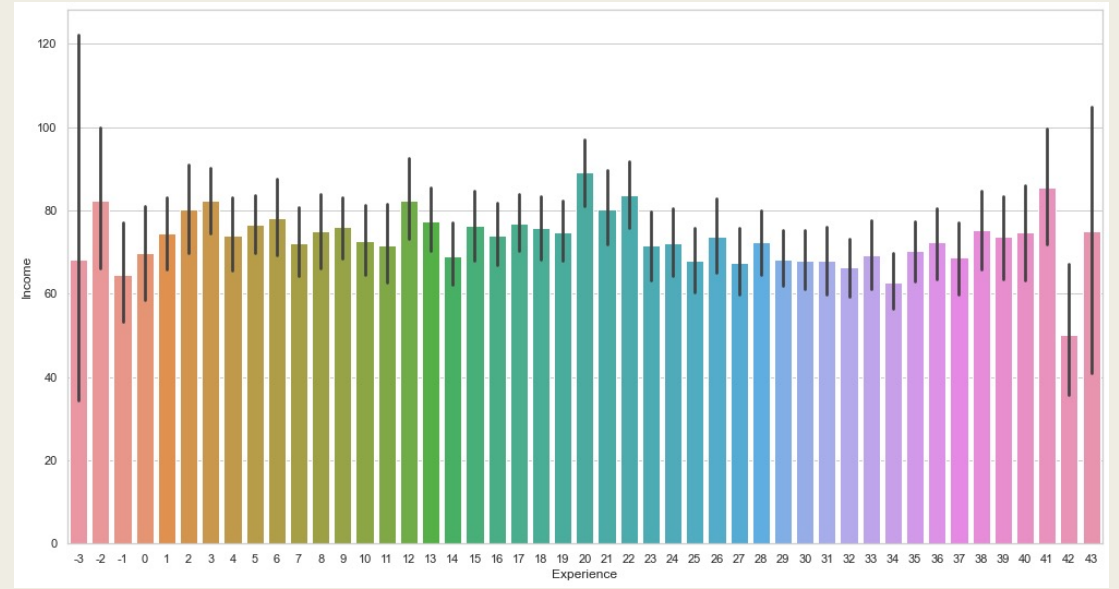
- As income increases , Mortgage also increased.
- Loan distributed across all the ages.
- people with income more than 140K took personal loan
- Family with 1 and 2 people did not take any personal loan , So they are potential customers
- Majority of the people with income less than 100K did not take personal loan , So they are potential customers as well.

Exploratory Data Analysis

Income vs Family



Income vs Experience



✓ Family with income less than 50K , did not take any personal loan.

Conclusion

- - We analyzed the "AllLife Bank" using different techniques , Logistic Regression and Decision Tree Classifier were used to build a predictive model for the same.
- - The model built can be used to predict if a customer is going to buy a personal loan or not.
- - We visualized different trees and their confusion matrix to get a better understanding of the model. Easy interpretation is one of the key benefits of Decision Trees.
- - We verified the fact that how much less data preparation is needed for Decision Trees and such a simple model gave good results even with outliers and imbalanced classes which shows the robustness of Decision Trees.
- - Income , Education , Family , CC Avg are the most important variable in predicting the customers that will buy a personal loan
- - We established the importance of hyper-parameters/ pruning to reduce overfitting.

Model Performance Summary

Logistic Regression

	Model	Train_Accuracy	Test_Accuracy	Train Recall	Test Recall	Train Precision	Test Precision	Train F1	Test F1
0	Logistic Regression Model - Statsmodels	0.952	0.944	0.521	0.488	0.934	0.940	0.669	0.643
1	Logistic Regression - Optimal threshold = 0.12	0.911	0.909	0.872	0.860	0.516	0.536	0.648	0.661
2	Logistic Regression - Optimal threshold = 0.36	0.953	0.941	0.715	0.682	0.763	0.727	0.738	0.704

Decision Tree

	Model	Train_Recall	Test_Recall
0	Initial decision tree model	1.000	0.890
1	Decision treee with hyperparameter tuning	0.930	0.860
2	Decision tree with post-pruning	0.920	0.860

- Decision tree model with Hyper Parameter has given best recall value

Business Insights and Recommendations

- * According to both the models it looks like making customer to buy a product is not an easy task.

- * People who does not have mortgage can buy a personal loan.

- * Family with 1 and 2 people did not take any personal loan , So they are potential customers

- * Customer retention - Member Loyalty programs initiatives like special discounts, coupons, etc can be provided.

- * Majority of the people with income less than 100K did not take personal loan , So they are potential customers as well.

- * Family with income less than 50K , did not take any personal loan. Those people can be targeted too to buy personal loan