Business Case.

The target marketing of loans is a major issue in banks. This can be solved by using our classification models which may help us find the best fit to market the loans.

Data with 5000 records has been used for the model.

Q1. Data Preprocessing.

The data after has different columns which are either numeric or categorical. Here, one of the columns- Education has been represented by different numbers due to which it was read as numeric by DataRobot but it actually is a categorical variable where the digits represent different levels of education. Thus, its type has been changed to categorical.

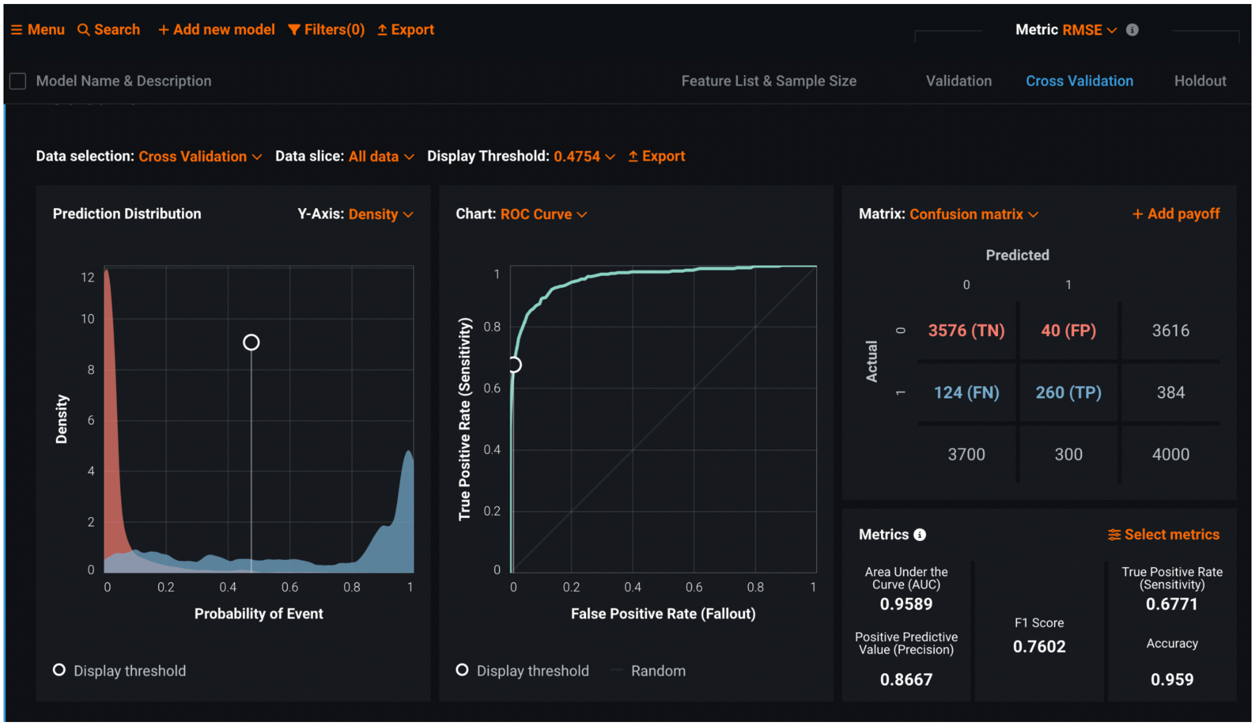
Education ...rical Int)	8		Categorical	3	0					
Education	8		Numeric	3	0	1.88	0.84	2	1	3

- Here, we have created a new feature list called LoanPreprocessedData. It has all the columns except Id, and the Education column which is numeric, but we have considered the categorical one.
- Zip Code could be excluded in this case, but it is best to keep the features unless you have a strong reason to do so.
- The new feature list contains:
    - Age, Income, ZIP Code, CreditCard, Securities Account, Education, Online, Family, Experience, Mortgage, CD Account, CCAvg, Personal Loan

Menu Search Feature List: LoanPreprocessedData View Raw Data + Create feature list 1-13 of 13

Feature Name	Data Quality	Index	Importance	Var Type	Unique	Missing	Mean	Std Dev	Median	Min	Max
Personal Loan		10	Target	Numeric	2	0	0.10	0.29	0	0	1
Income		4		Numeric	162	0	74.08	46.34	64	8	224
CCAvg		7		Numeric	104	0	1.94	1.75	1.50	0	10
CD Account		12		Numeric	2	0	0.06	0.24	0	0	1
Mortgage	i	9		Numeric	319	0	55.77	102	0	0	635
Education ...rical Int)		8		Categorical	3	0					
Family		6		Numeric	4	0	2.40	1.15	2	1	4
Experience		3		Numeric	47	0	20.11	11.44	20	-3	43
CreditCard		14		Numeric	2	0	0.30	0.46	0	0	1
Securities Account		11		Numeric	2	0	0.10	0.30	0	0	1
Online		13		Numeric	2	0	0.60	0.49	1	0	1
Age		2		Numeric	45	0	45.34	11.43	45	23	67
ZIP Code	i	5		Numeric	460	0	93,177	1,756	93,437	90,005	96,651

Q2. Model Performance Evaluation.



Recall	0.6771
Precision	0.8667
F1	0.7602
Accuracy	0.959
Error Rate	0.041
ROC AUC	0.9589

The values are noted at a threshold value of 0.4754 which maximizes the F1 value.

Naive model would predict that no one accepts the loan offered because only 10% will accept a loan offer from the bank. ROC AUC of naive would be 0.5. Accuracy of the naive model would be 90%.

Logistic regression is better performing in this case than the naive model.

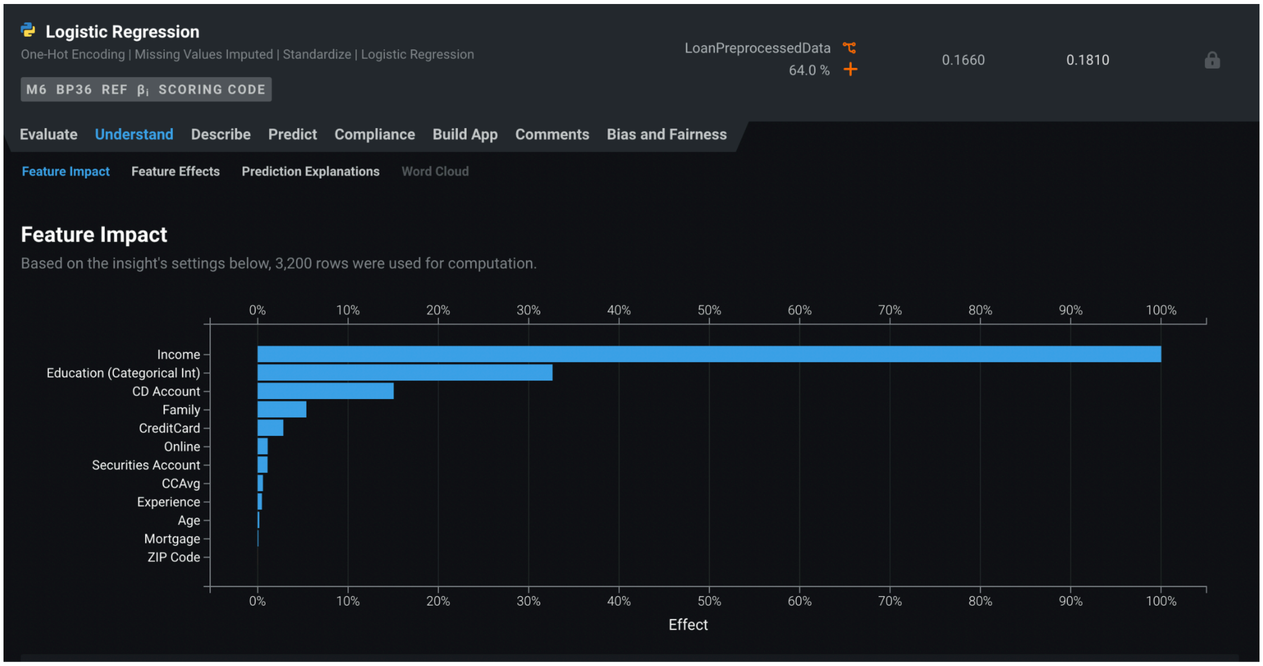
Also, the values for the hold out will be as shown below:

Recall	0.6979
Precision	0.9054
F1	0.7882
Accuracy	0.964
Error Rate	0.036
ROC AUC	0.967

The metrics for the hold out are close to the cross validation data's metrics. There is not much difference between them. Hence, the model is reliable and has good generalization.

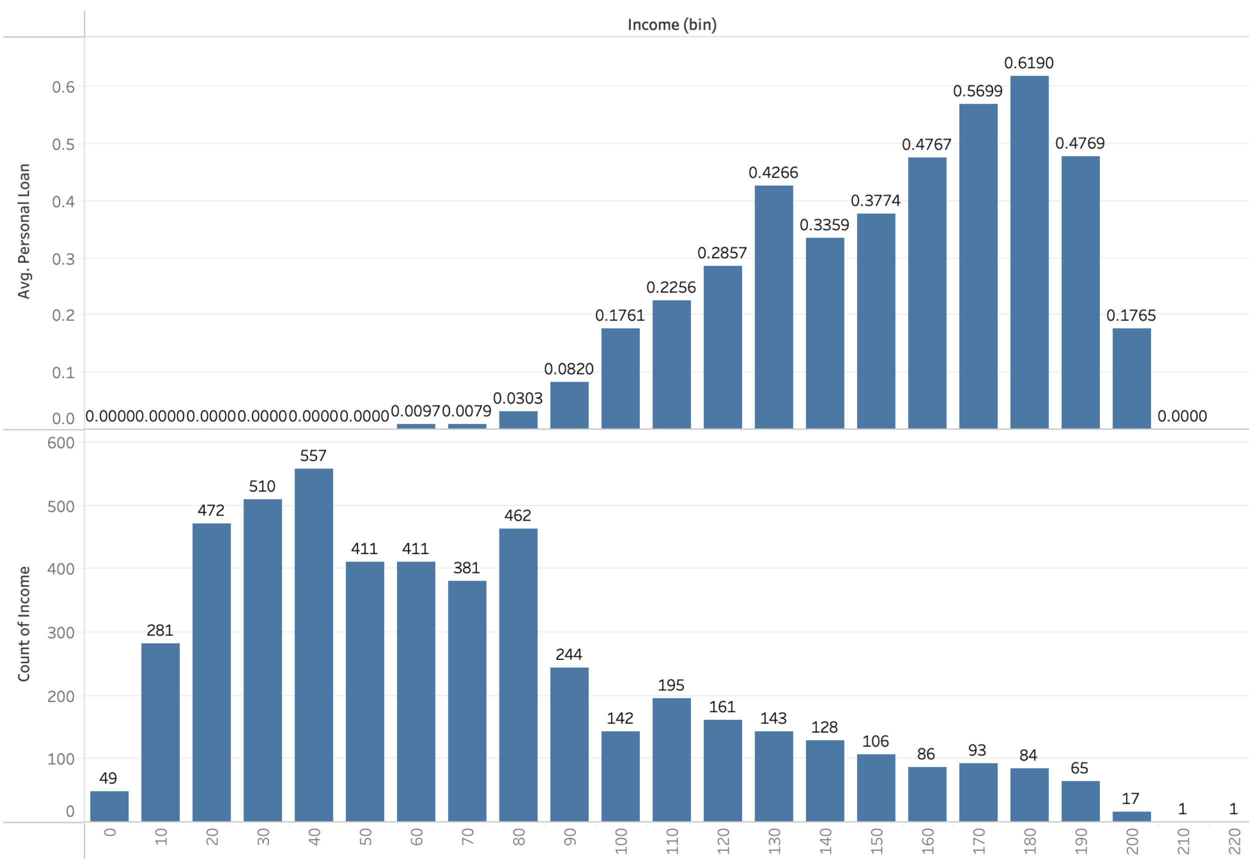
Q3. Feature impact.

The top 5 impactful features are Income, Education, CD Account, Family and Credit Card.



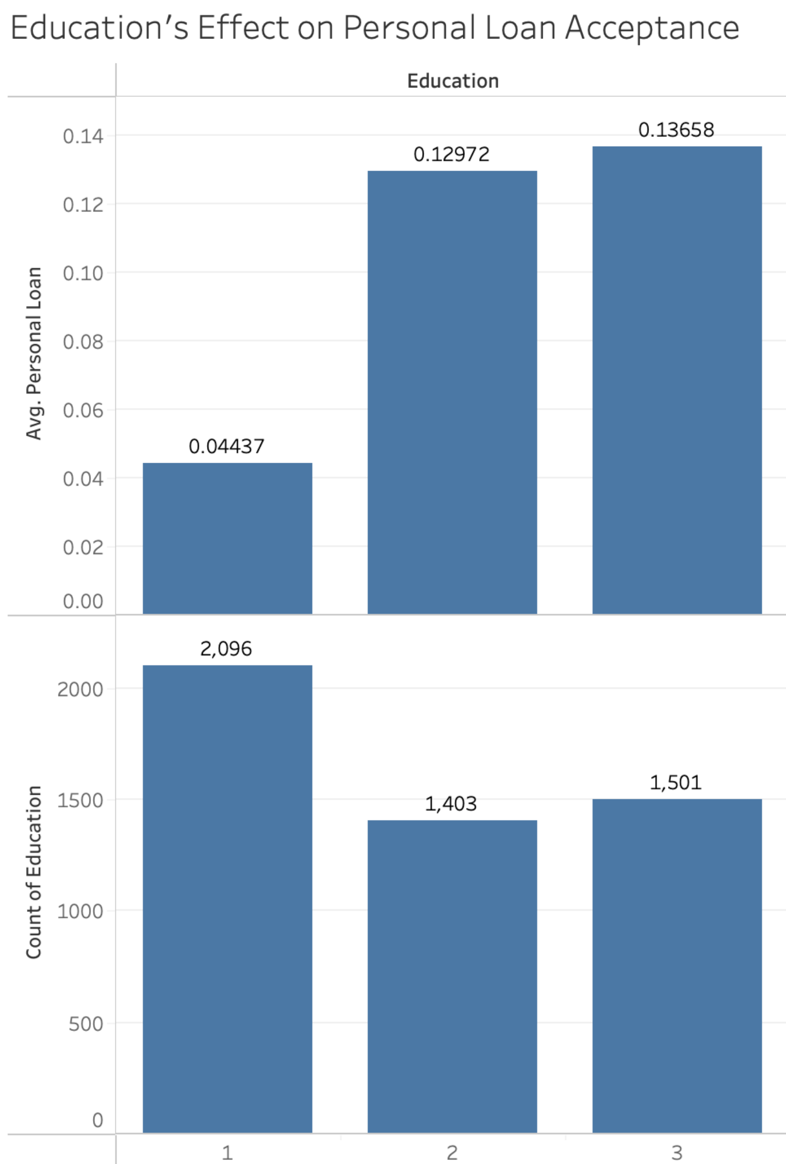
1. Income

Income's Effect on Personal Loan Acceptance



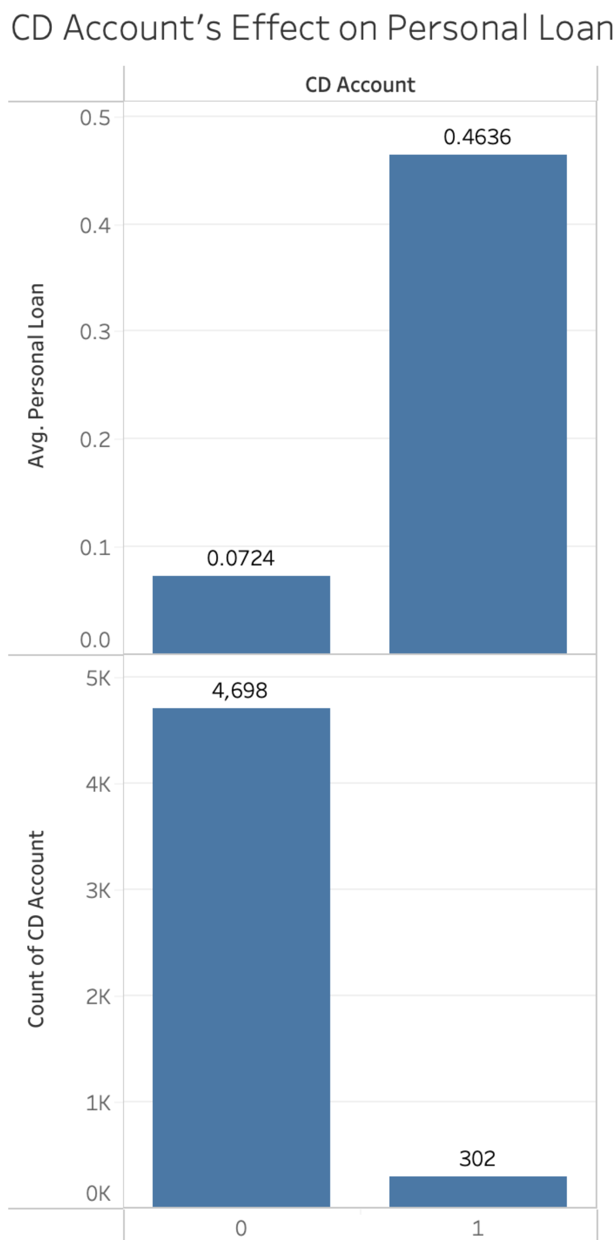
The overall trend indicates a rise in loan acceptance rates from 60k to 180k, with a notable upward spike at 130k. Beyond 180k, there is a subsequent decline, suggesting that individuals earning higher incomes may have reduced demand for personal loans, resulting in a gradual decrease in acceptance rates.

2. **Education**



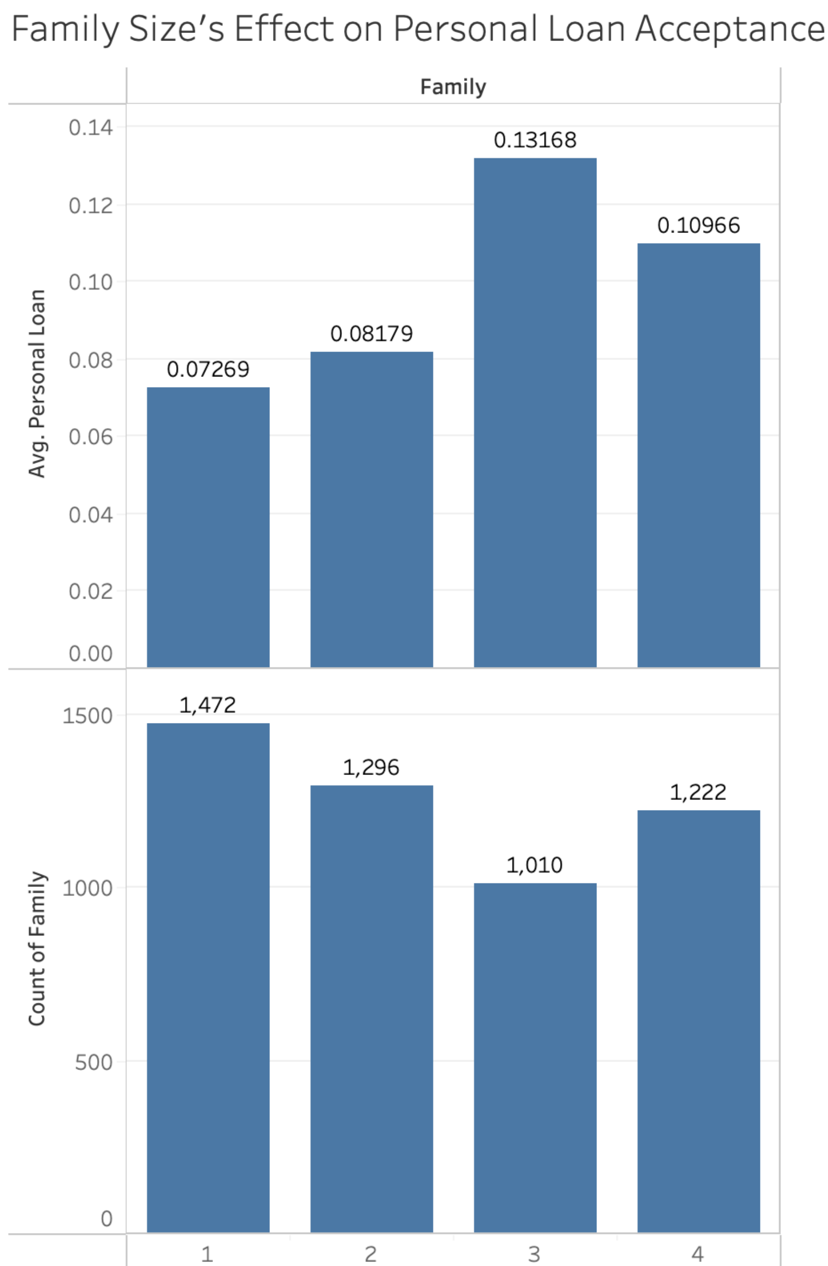
People who have an educational level of either 2 (12.9%) or 3 (13.6%) are more likely to accept personal loans. The least accepted loans are by people with level 1 (4.4%) education even though they are more populous in comparison to each level 2 and 3.

3. CD Account



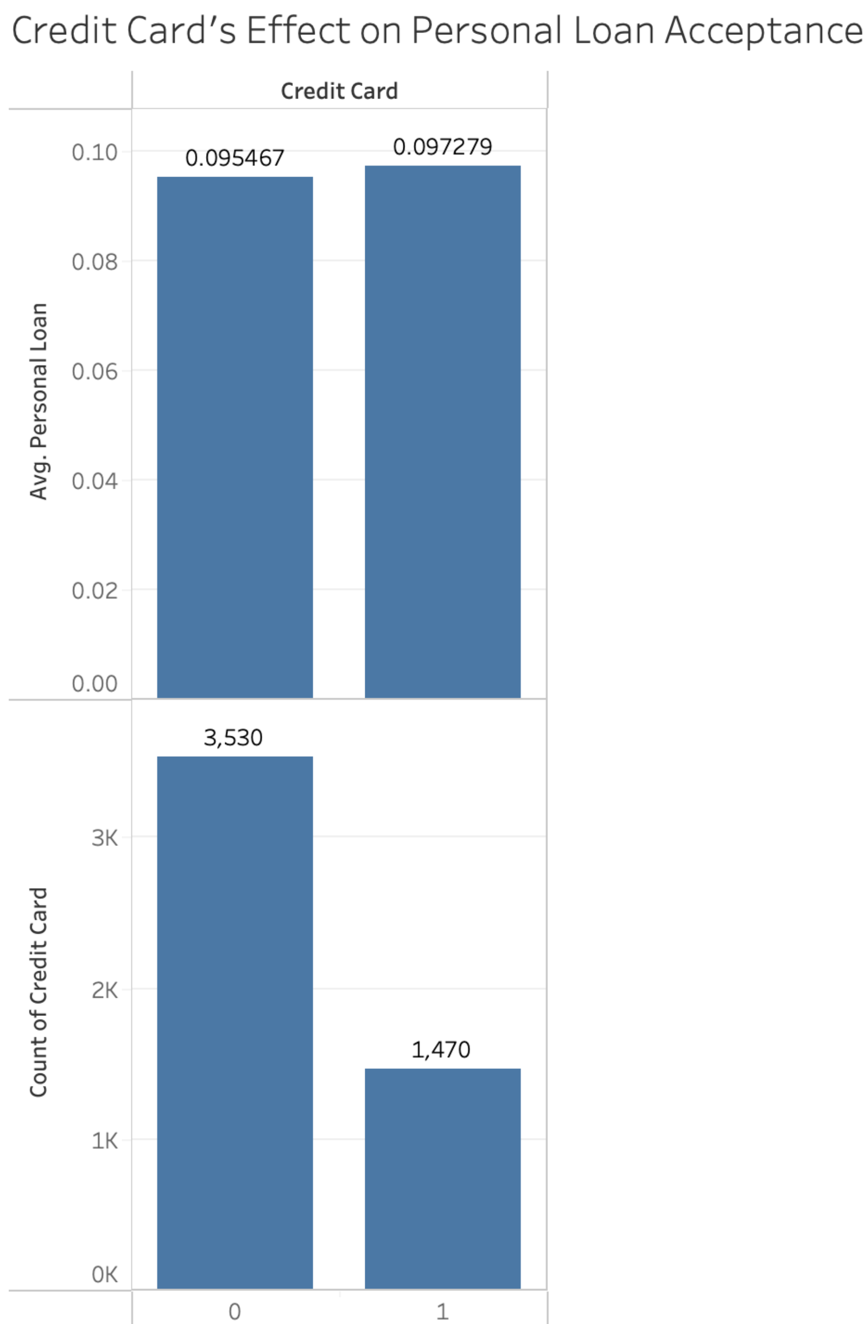
People with a CD account were ~6 times more likely to accept a personal loan than those without a CD account, even though people without a CD account are ~15.5 times more populous.

4. **Family Size**



It can be observed that larger family sizes are more likely to accept a personal loan offer than the ones with smaller families. Even though the families with a single member are maximum in numbers, they are less likely to accept a loan offer than people in larger family sizes.

5. **Credit Card**



Even though the people not owning credit cards are almost twice in number as the people owning credit cards, they are as likely to accept the loan offer as someone holding a credit card would be.

Q4. Model Sanity Check.

The model has predictive value as the ROC AUC = 0.96. F1 at optimal threshold is 0.76.

The key features predicted by the outcome make sense. It is reasonable that the people who have a high income are more likely to accept a loan. Also, people with more education are likely to accept more personalized loans. People with larger family sizes need personalized loans given their expenditure as a large family. A person with CD account ownership is likely to accept a loan as they have their fixed deposits in the banks already which may be one of the reasons their acceptance ratio is much higher than the people without CD accounts. One of the ambiguous areas is the credit card ownership as there isn't much difference in loan acceptance between people who own and who don't own a credit card.