

## Assignment 3: Personal Loan Offer Marketing

Business Case: Personal Loan offer case study signifies the case for Universal Bank whose marketing team wants to evaluate the number of customers who are more likely to accept a personal loan offer. We'll focus our evaluation based on Universal bank data set containing 5000 values. Logistic Regression is used to make the analysis on features such as customer demographic information (age, income, etc.) and the customer's relationship with the bank (mortgage, securities account, etc).

### Question 1:

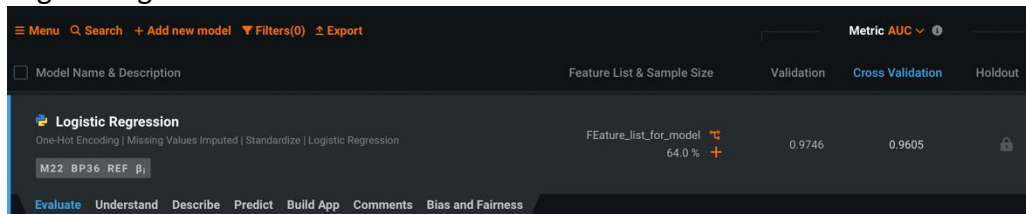
Data preprocessing included the following steps:

1. Excluded the variable due to high cardinality: ID
2. Recoded Education as it had 3 distinct values signifying the education level. 1: Undergrad; 2: Graduate; 3: Advanced/Professional.
3. Recoded Family as it had 4 distinct values, 1,2,3 and 4 signifying the members/size of the family.

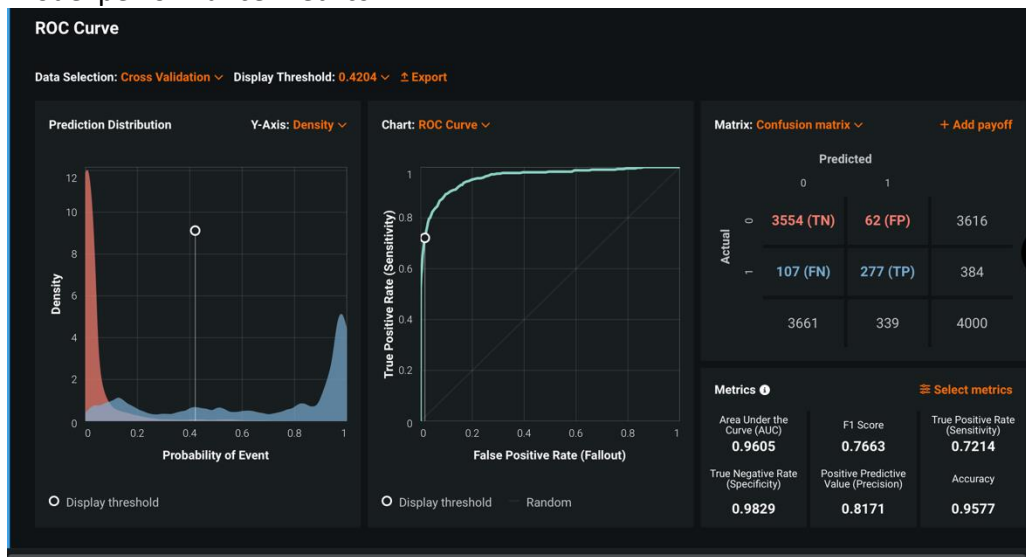
Feature Name	Data Quality	Index	Importance ↑	Var Type	Unique	Missing	Mean	Std Dev	Median	Min	Max
<input type="checkbox"/> Income		4	<div></div>	Numeric	162	0	74.08	46.34	64	8	224
<input type="checkbox"/> CCAvg		7	<div></div>	Numeric	104	0	1.94	1.75	1.50	0	10
<input type="checkbox"/> CD Account		12	<div></div>	Numeric	2	0	0.06	0.24	0	0	1
<input type="checkbox"/> Mortgage	<div></div>	9	<div></div>	Numeric	319	0	55.77	102	0	0	635
<input type="checkbox"/> Updated_E...tegorical		8	<div></div>	Categorical	3	0					
<input type="checkbox"/> Updated_F...tegorical		6	<div></div>	Categorical	4	0					
<input type="checkbox"/> Experience		3	<div></div>	Numeric	47	0	20.11	11.44	20	-3	43
<input type="checkbox"/> CreditCard		14	<div></div>	Numeric	2	0	0.30	0.46	0	0	1
<input type="checkbox"/> Securities Account		11	<div></div>	Numeric	2	0	0.10	0.30	0	0	1
<input type="checkbox"/> Online		13	<div></div>	Numeric	2	0	0.60	0.49	1	0	1
<input type="checkbox"/> Age		2	<div></div>	Numeric	45	0	45.34	11.43	45	23	67
<input type="checkbox"/> ZIP Code	<div></div>	5	<div></div>	Numeric	460	0	93,177	1,756	93,437	90,005	96,651

## Question 2:

### Logistic regression model with AUC metric



### Model performance metrics:



Error rate= 1- Accuracy=1-0.958=0.042

Logistic regression has predictive value- ROC AUC =0.9605

The model error rate is 4.2% at the chosen threshold(0.4204).

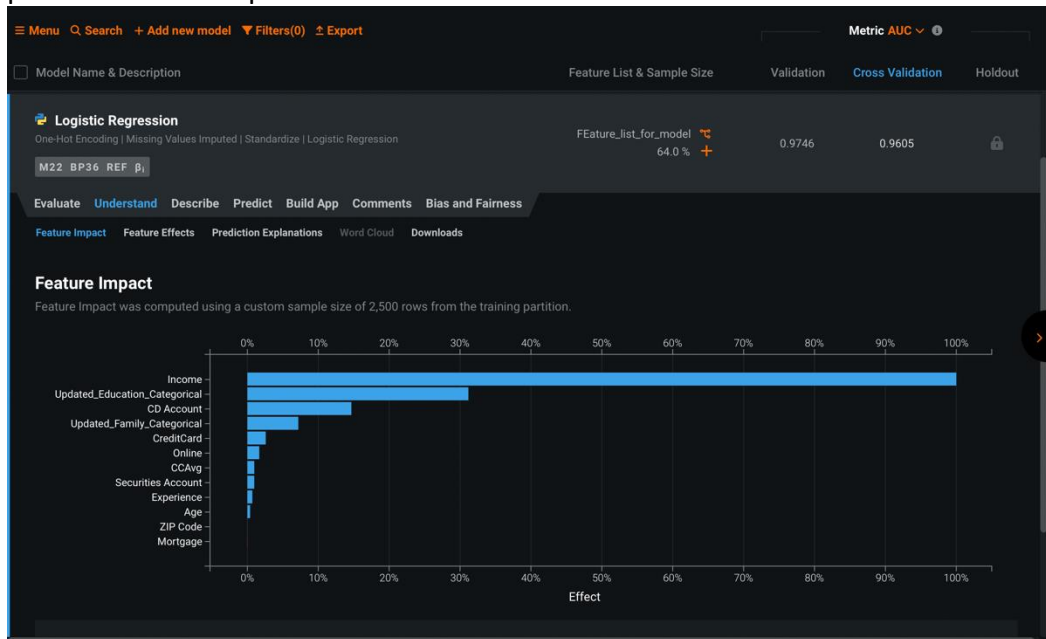
The False positive rate(FPR):  $FP/(FP+TN)$  or  $1- \text{Specificity} = 62/(62+3554)=0.0171$  or  $1-0.9829 = 0.0171$

The False Negative rate (FNP):  $FN/(FN+TP)=107/(107+277)=0.279$

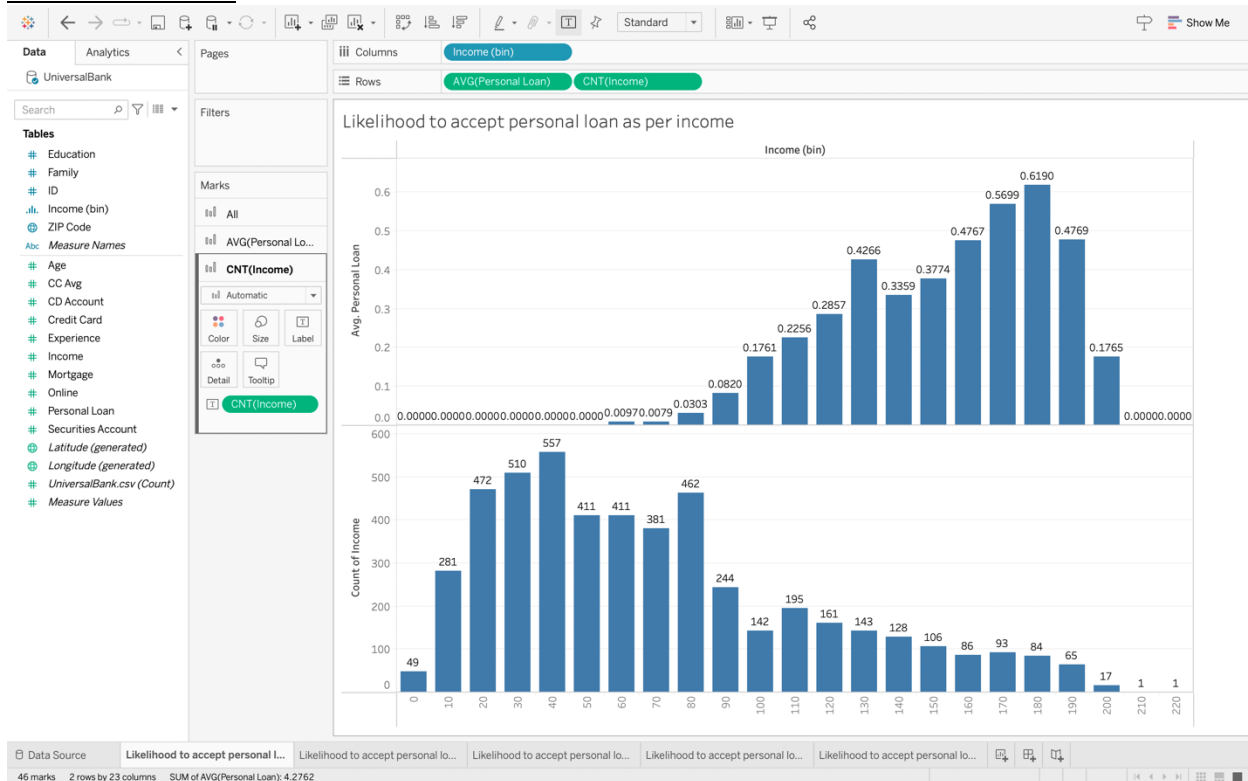
The best metric is ROC AUC, as its score tells us how efficient the model is. The higher the score the better are the chances to distinguish between the positive and negative classes. In our case we are getting a value of 0.9605 which is closer to 1. Further the value of threshold has no impact on the ROC AUC value and hence it remains a constant.

### Question 3:

Income, Education, CD account, Family and Credit card are the key features for likelihood of personal loan acceptance.



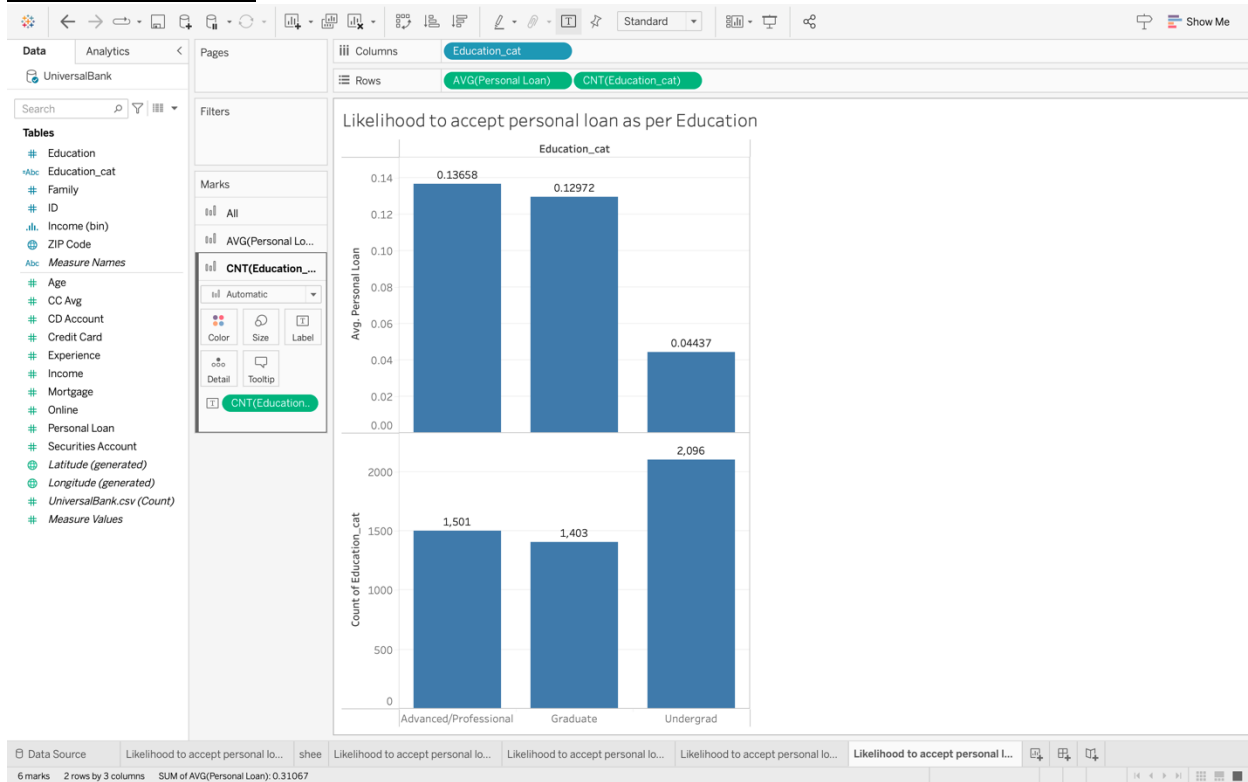
### Factor 1 : Income



Income effect- Customers having an income of \$40,000 (assuming each partition shows a difference of \$10,000 for the given dataset) are the highest in the dataset (having count of 557).

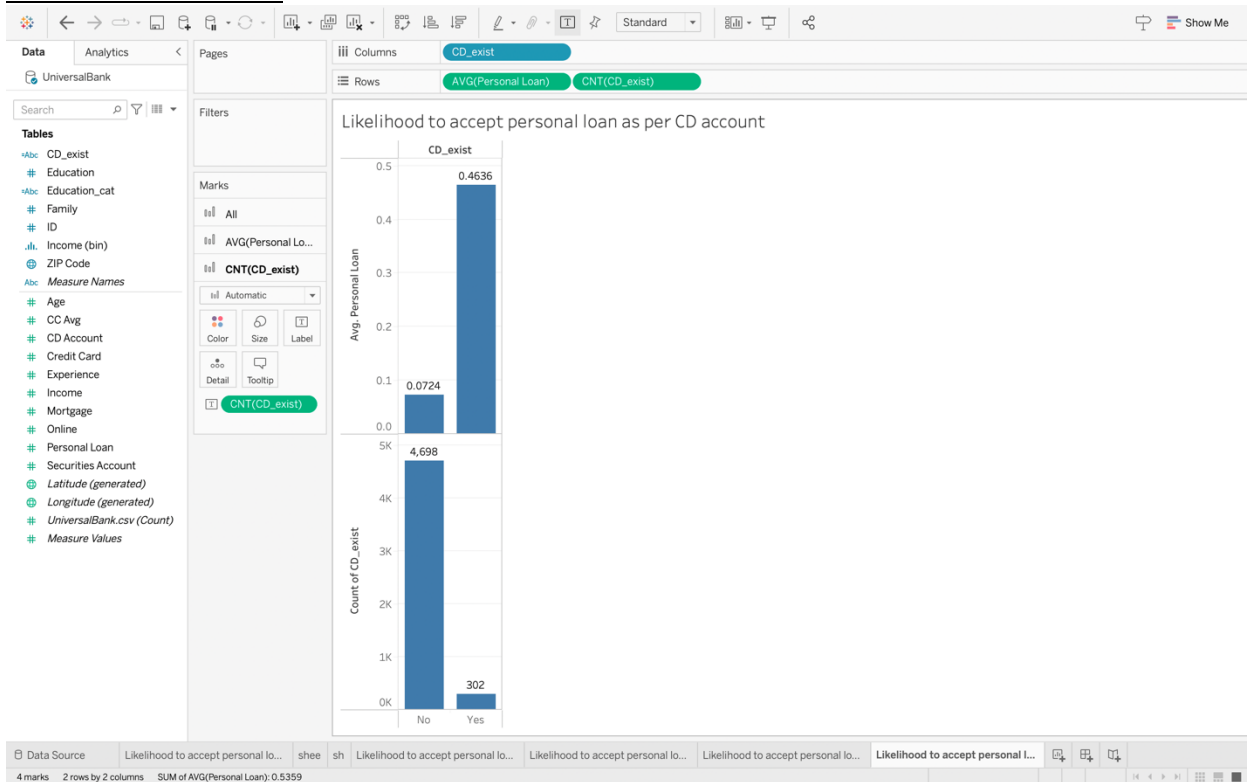
However, customers with less income are less likely to take personal loan. As can be seen in above visualization until \$50,000 where the likelihood for personal loan requirement is 0.97% the personal loan requirement is 0 prior to that. With the increase in income the requirement for personal loan is rising being highest at \$180,000 for 61.90%.

## Factor 2: Education



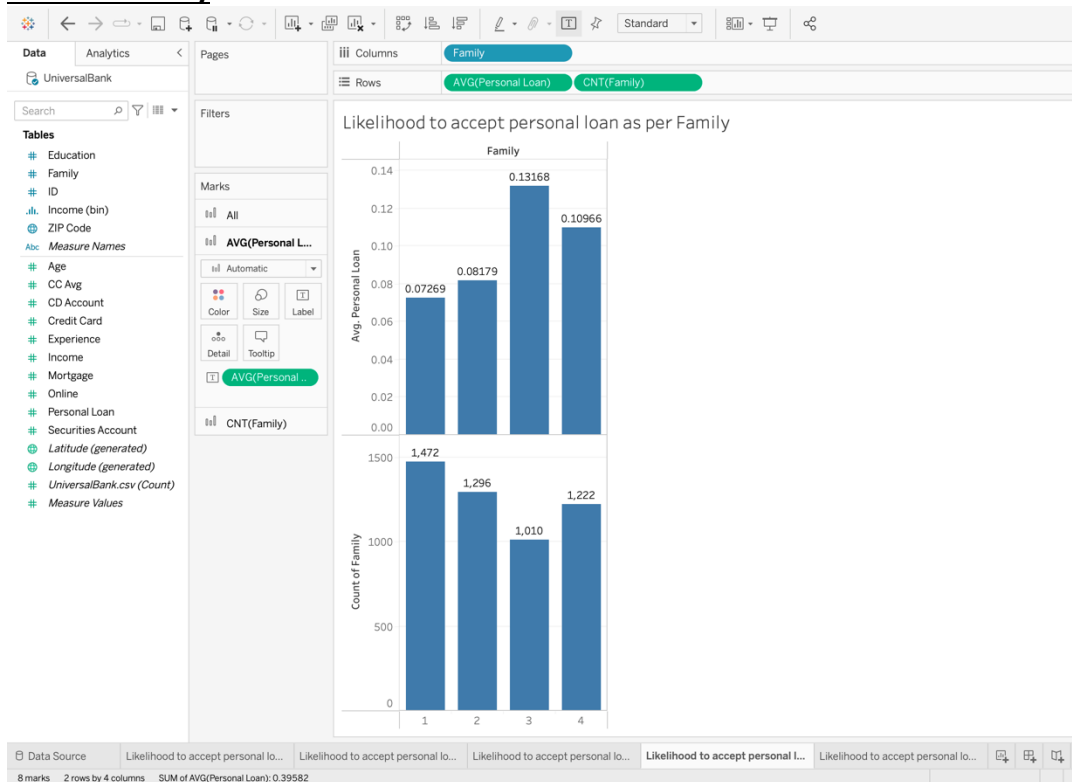
Education Effect- Although the customers are more in Undergrad category (having count of 2,096) when compared to graduate( having count of 1,403) and advance/professional( having count of 1,501) , the advance/professional qualified customers are more likely to have personal loan requirement that is 13.66% when compared to graduate 12.97% and Undergrad 4.44% category customers.

### Factor 3: CD Account



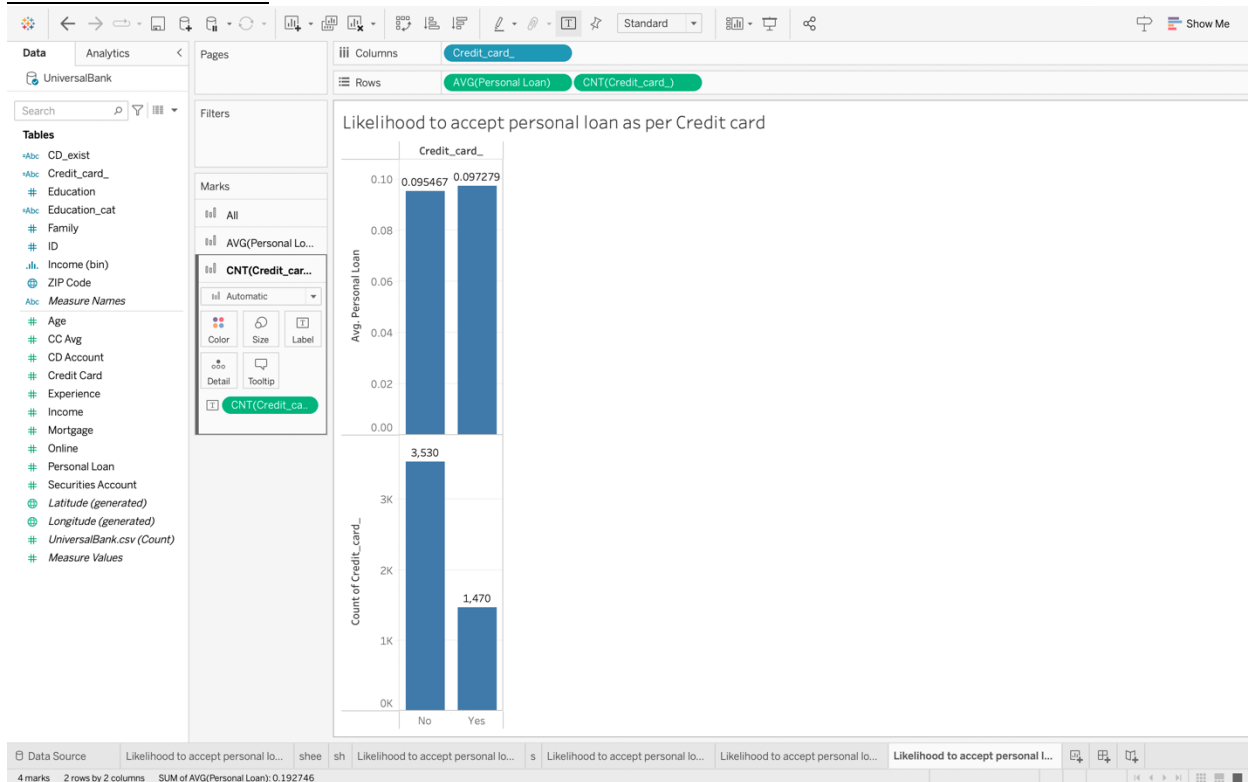
CD account effect- Population of Customers not having certificate of deposit (CD) account with bank are higher(having count of 4,698) than customer having the CD(having count of 302). The customers who have CD are more likely to have personal loan requirement 46.36 % when compared to customer not having CD account 7.24%.

## Factor 4: Family



Family effect- As can be seen that family having 1 member has the highest count of 1,472, however, family of 3( having count of 1,010) seems to have the highest personal loan requirement of 13.17% when compared to family having 1 member have requirement of 7.27% ; 2 members( having count of 1,296) have requirement of 8.18% ; and 4 members (having count 1,222) have requirement of 10.97%.

## Factor 5:Credit card



Credit card effect- The customers not having credit card for the bank are higher (having count of 3,530) when compared to customers having the bank credit card (having count of 1,470). The likelihood of personal loan requirement is higher for people having credit card by 9.73% when compared to customer not having the credit card 9.55%.

### Question 4

Yes, the model does make sense, as the model have predictive value with ROC AUC curve being the best metric for its evaluation (0.9605 in our case) . Examination of feature effects shows that Income, Education, CD account, Family and Credit card are the key features for likelihood of personal loan acceptance. The observed effect shows that people with high income (\$180,000 being highest in our case), with higher level of qualification (Advance/professional qualified people in our case) , Customers having certificate of deposit (CD) with the bank, family size of 3 members and customers having credit cards are more likely to accept personal loans from the bank.