Case Study – Personal Loan Marketing for Universal Bank

Business Case

To optimize personal loan marketing and maximize profitability, banks seek to predict customer acceptance of loan offers. By leveraging a model trained on customer demographics and bank relationship data, the bank aims to achieve several key advantages. Targeting the right audience will drive up conversion rates and revenue, while allocating marketing resources efficiently reduces costs. Personalized loan options based on individual profiles enhance customer satisfaction and loyalty, leading to improved long-term value. Additionally, the model uncovers crucial insights into customer behavior, informing impactful product development and marketing strategies. This data-driven approach ultimately grants banks a competitive edge in the market by fostering a deeper understanding of their customer base.

This means targeting the right people, reducing wasted marketing spend, and boosting loan approvals. It's an investment that leads to higher revenue, happier customers, and a competitive edge in the market.

The dataset used here is for Universal Bank and has 5000 customer records.

Q1. Explain data pre-processing steps.

- Selected the following features for modeling – Income, CCAvg, CD account, Mortgage, experience, CreditCard, Family, Securities Account, Online, Age and Education
- Personal Loan is our Target Variable.
- Recoded Education as a categorical feature
- Dropped zip code and ID from the dataset as it is a non-informative feature for forecasting loan marketing.

Menu Search Feature List: PreprocessedDataUni... View Raw Data + Create feature list 1-12 of 12

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		9		Numeric	319	0	55.77	102	0	0	635
Updated_E...tegorical		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

Fig: Selected Features from the Data

Q2. Report Recall, Precision, F1, Error rate, Accuracy, ROC AUC. Does the model have predictive value? Explain (compare to naive).

The below performance metrics are reported for the trained model with the threshold value set for 0.4736 which is used to maximize the F1 score.

Recall	0.677
Precision	0.867
F1 score	0.760
Error Rate	0.041
Accuracy	0.959
ROC AUC	0.959

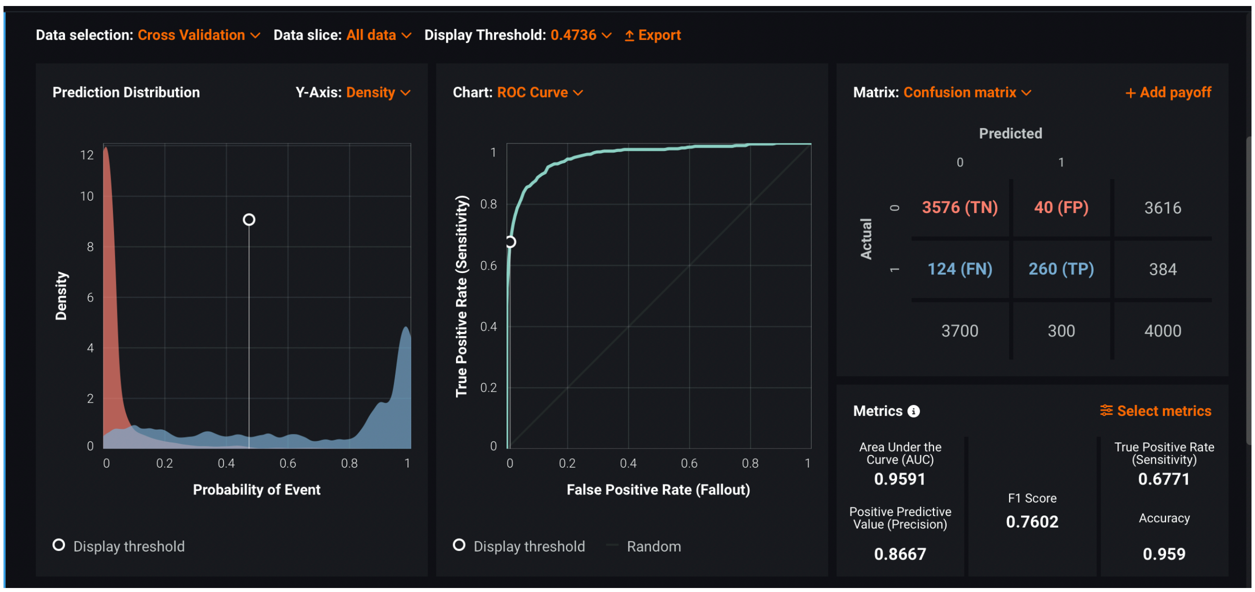


Fig:- Performance Metrics of the Logistic Regression model

Naïve model would predict that no customers took the personal loan because only 10% of the customers in the dataset took the loan. Essentially, the naïve model would be biased towards the majority class (customers not taking the loan), reflecting the imbalanced nature of the data. ROC AUC for the naïve model would be 0.5 and accuracy would be 90%.

Logistic regression model for the above problem statement significantly outperforms the naïve model providing an accuracy of 95.9% and a ROC AUC of 0.96

Q3. Which factors are important in the prediction of loan acceptance in the model? Provide visualizations and 1-sentence summary for the top 5 factor effects.

The top 5 features that have significant impact in the prediction of loan acceptance are -
Income, Education, CD Account, Family and Credit Card

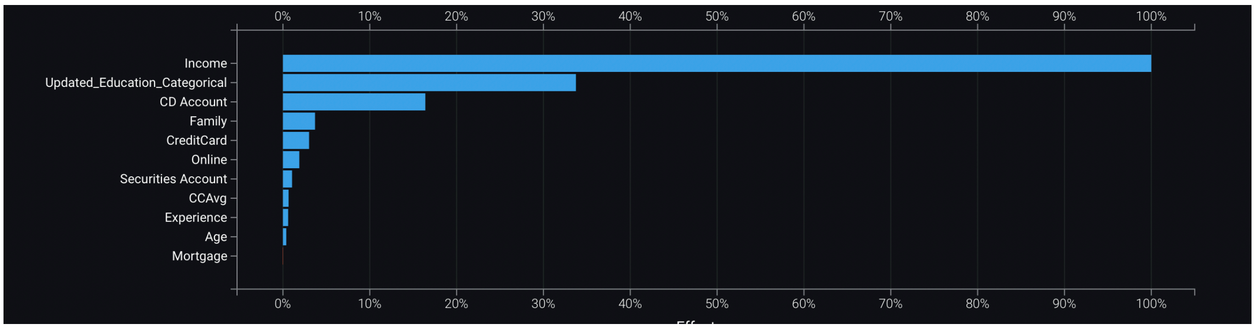
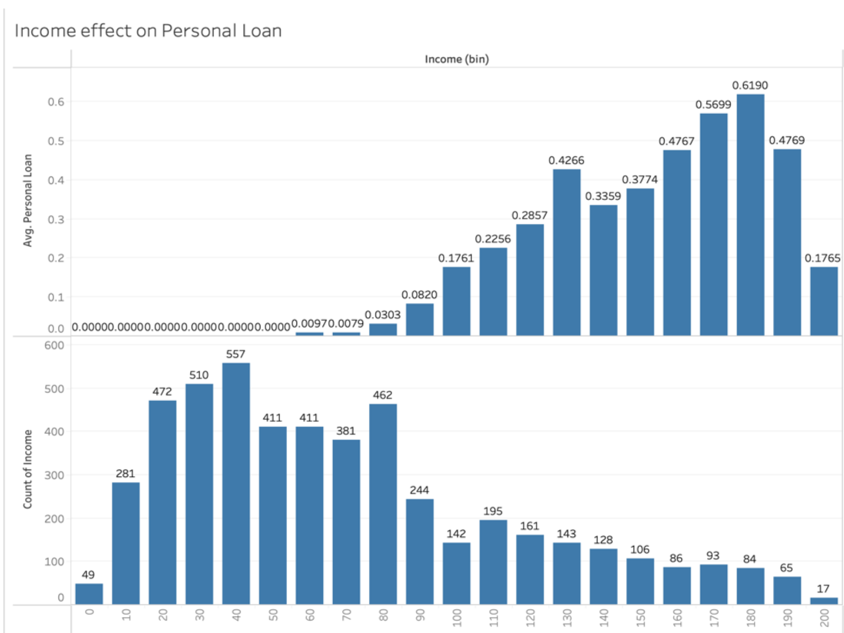


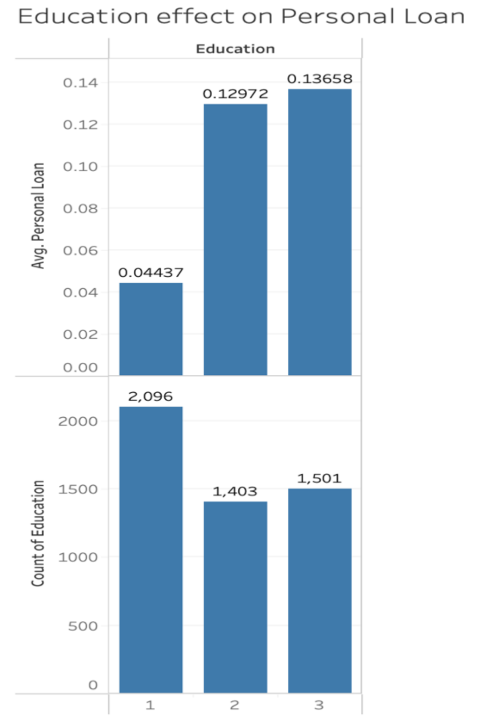
Fig: - Feature Importance in the model

1. Effect of income on the model predictions



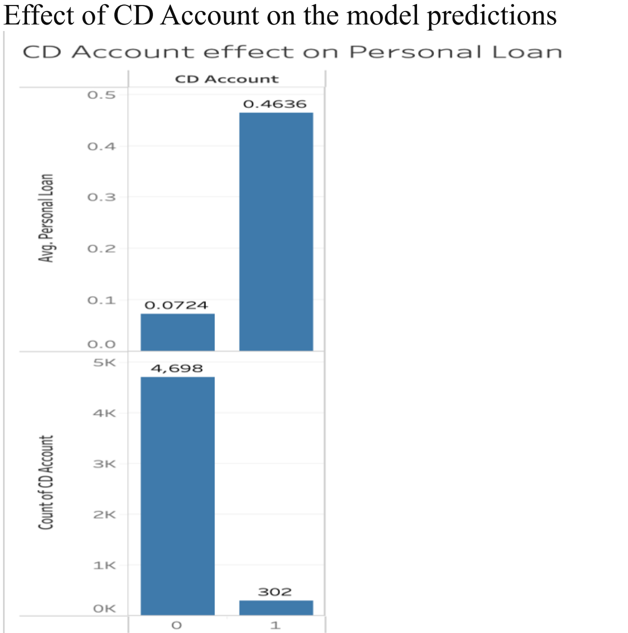
Summary - It is seen that for customers with income lower than 60k the chance of taking a personal loan is almost null and it increases as the income increases. But we can again see a significant decline in the number of accepted offers after 190k. Visualization excludes records where income>200k as the chances of loan are too low and such data points are minimal.

2. Effect of education on the model predictions



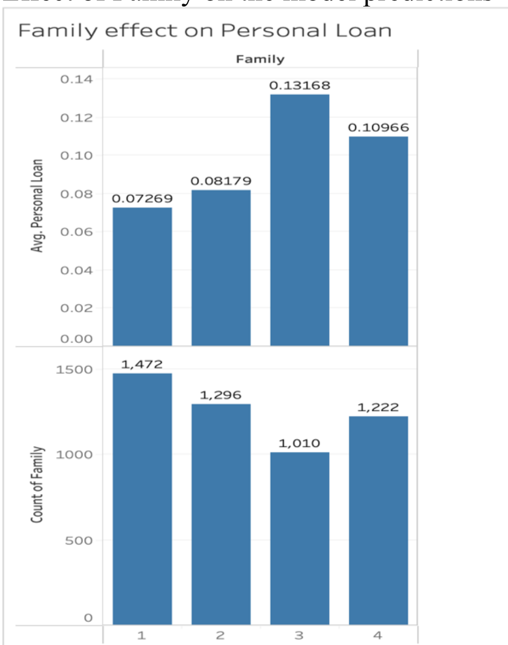
Summary – Customers with higher level of education are more likely to accept a personal loan offer. Even though the number of customers with education level as 1 is almost 1.5 times more than customers with education level 2 or 3, the acceptance rate for the loan is almost 3 times more for people with education level 3(13.65%) as compared to 1(4.43%).

3. Effect of CD Account on the model predictions



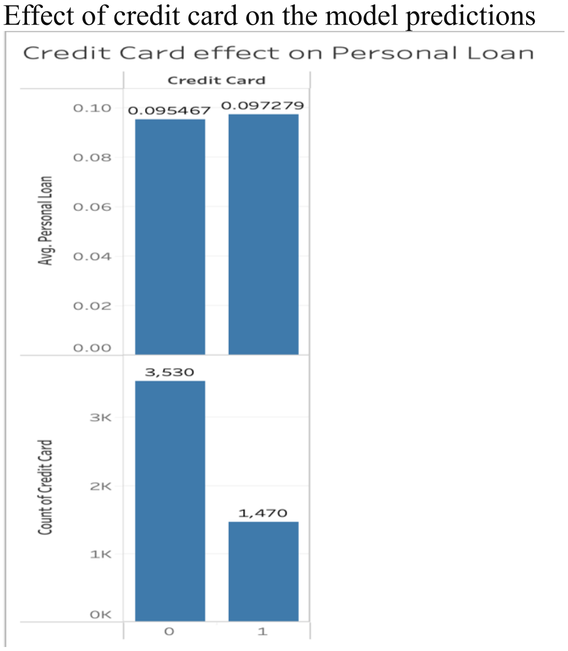
Summary – Count of customers that possess a CD Account is almost 15 times less than those that don't but still their chances of accepting a loan offer are 6.5 times more than those who do not possess a CD account.

4. Effect of Family on the model predictions



Summary – Customers with bigger family sizes are more likely to take a personal loan. Acceptance rate for customers with family size of 3(13.16%) is almost 1.8 times more likely as compared to individuals or people with family size as 1(7.3%).

5. Effect of credit card on the model predictions



Summary – Though customers without credit card are almost 2.4 times the customers with credit card, the percentage of converts for personal loan is almost same, with customers having credit card as 97.2% and those without as 95.4%.

Q4. Does the model make sense?

The model has predictive value. ROC AUC is 0. 959 and F1 score at optimal threshold is 0.760.

The model's logic is grounded in the observation that individuals falling within a specific income range are more inclined to accept a personal loan, a pivotal factor in the model. Those with insufficient income are less likely to secure a loan, while those surpassing a certain income threshold typically do not require one. Additionally, education level, possession of a CD account, and family size are significant contributors. Larger families are more prone to loan acceptance, given their potentially greater need for personal loans. Similarly, individuals with higher education levels exhibit an increased likelihood of accepting a personal loan.
Another factor considered is the presence of credit cards, constituting approximately 3% of the feature importance. This suggests that possessing or lacking a credit card has a minimal impact on the model. However, further investigation is warranted to gain a deeper understanding of this aspect.