**Loan Tap Business Case Study**

**Insights:**

1. Annual income range between 0-10,00,000 for the current customers.
2. DTI ranges from 0-1000 for current customers.
3. There is a variety of different employment length for the customers, highest being the 10+ years. It is distributed almost equally between 0-10 years.
4. Installment amount follows skewed normal distribution with peak between 200-400.
5. There are higher chances for customers to get interest rate between 10-15%, but it ranges from 5-25%
6. Most number of loan amounts are between the range 5-10k
7. Very high percentage of loans are being lend for 36 months
8. 29% customers are put into B-Grade Category
9. 0.7% customers are put into G-Grade Category.
10. 50% customers have mortgage and 40% are on rent.
11. 80% loan tap customers have fully paid the loans, which is very high in comparison to the charged off loans.
12. 99% applications for loans are individual type.
13. Teachers and Managers are the two most common professions seeking for loans.
14. Loan amount and Interest rate, installment amount is higher in the case of Charged off loans than fully paid loans.
15. There is high risk of NPA cases in F and G category customers.
16. There is low risk on NPA cases when loans are taken with debt consolidation, credit card and home improvement purpose.
17. Customers from the following Pin code/Geographical Locations are likely to be defaulters: 05611,03668,00739.
18. As the loan amount is increased, installment amount increases linearly.
19. Risk for F and G category is high, as they have high loan amounts and higher installments.
20. Loan amount and Annual Income does not have high correlation.
21. Loan amount is not dependent on employment length.
22. There is low risk in 36-month installment scheme.
23. It has been observed that geographical location could be a critical measure for distinguishing between Low risk and High-risk clients.
24. Metric Scores:

* Precision: 0.89
* Recall: 0.99
* F1 Score: 0.93
* AUC Score:0.92

**Recommendations:**

1. The data presented is highly imbalanced. It is recommended to gather more information for the NPA customers.
2. High risk analysis should be done for F and G category customers as they induce high chances of NPA’s for the organization.
3. Less chances are for NPA’s for customers with Credit card, Debt Consolidation and Home Improvement purposes. Hence, loans can be provided with due diligence.
4. Customers from the following Pin code/Geographical Locations are likely to be defaulters: 05611,03668,00739. Hence, High risk analysis is recommended for these locations.
5. Higher the Loan amount and interest rate, higher are the chances for NPA’s. High loan amount should only be provided to customers with high Annual income and Employment length.
6. So far, we have achieved great results with model optimisation. As all the key metrics are nearly 90. Precision should be high to prevent loans disbursement to defaulters. Recall should be high to prevent denying loans to genuine customers which adds value to our revenue. Hence, F1 Score is recommended.
7. As the AUC score is high, it suggests that the model is working well and can be used in deployment.
8. Precision can be optimized by 2-3% by bringing down recall from 99% to 92.5%.

**Trade-off Questions:**

1. **How can we make sure that our model can detect real defaulters and there are less false positives? This is important as we can lose out on an opportunity to finance more individuals and earn interest on it.**

**Ans: According to our labelling, Recall should be high in this case.**

1. **Since NPA (non-performing asset) is a real problem in this industry, it’s important we play safe and shouldn’t disburse loans to anyone.**

**Ans: According to our labelling Precision should be high in this case, as we need to minimize the risk for giving loans to defaulters.**

**Questions**

1. What percentage of customers have fully paid their Loan Amount?

80%

1. Comment about the correlation between Loan Amount and Installment features.

Higher the loan amount, Higher is the installment feature. It is highly correlated.

1. The majority of people have home ownership as Mortgage.
2. People with grades ‘A’ are more likely to fully pay their loan. (T/F)

True

1. Name the top 2 afforded job titles.

Teachers and Managers

1. Thinking from a bank's perspective, which metric should our primary focus be on..
   1. ROC AUC
   2. Precision
   3. Recall
   4. F1 Score

F1 score should be key metric.

1. How does the gap in precision and recall affect the bank?

This could induce loans given defaulters and could prevent the bank to give loans to genuine customers.

1. Which were the features that heavily affected the outcome?

Employment title, Geography ,Loan amount, Installment Features.

1. Will the results be affected by geographical location? (Yes/No)

Yes