

Credit EDA: Case Study

By Krunal Nagda

Problem Statement

- To identify if a client has difficulty paying their dues so that no loan is approved which could default and no clients who are capable of paying the loan are rejected
- Identifying variables which are strong indicators of default

Missing Values

- Columns with more than 50% missing values were dropped
- Rows with majority missing values were dropped which was around 0.3% of the data
- Some missing values could be imputed, for example, missing values of NAME_TYPE_SUITE variable having 0.4% missing values could be imputed using “Unaccompanied” suite since about 81% of the variable falls in this category
- Another example could be, imputing missing values of AMT_ANNUITY variable using the median of the field since there are outliers present

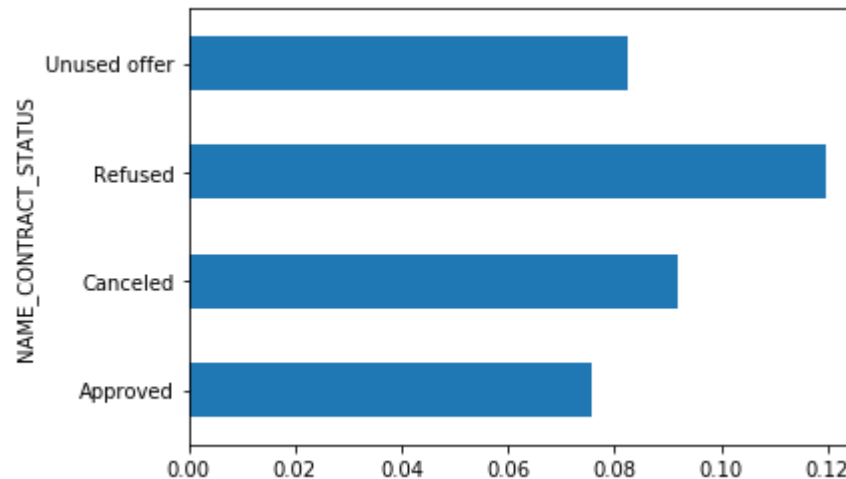
Outliers

- Outliers lie in the following fields:
 - One outlier lying at 11.7 crore in AMT_INCOME_TOTAL
 - Many above 15 lacs in AMT_CREDIT
 - Many above 60,000 in AMT_ANNUITY
 - Many above 13 lacs in AMT_GOODS_PRICE
 - A few around 1000 years in EMPLOYED_YEARS which is not possible so can be removed
- These outliers should be handled separately or removed if it is an incorrect value

Key Indicators of Default

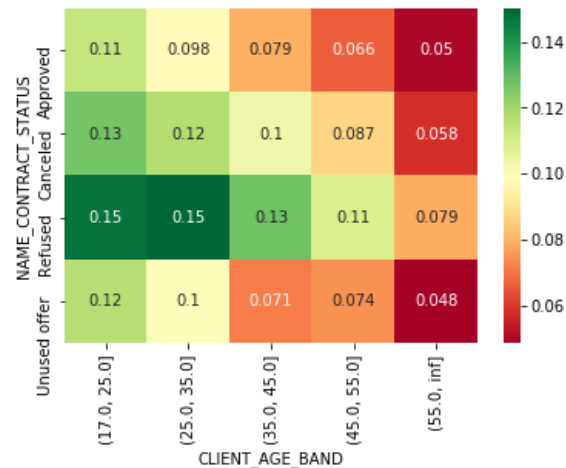
- ◉ Previously refused loans
- ◉ Client's age
- ◉ Client's family status
- ◉ Contract type
- ◉ Client's gender
- ◉ Client's education level
- ◉ Loan amount

Key Indicator: Previously Refused Loans



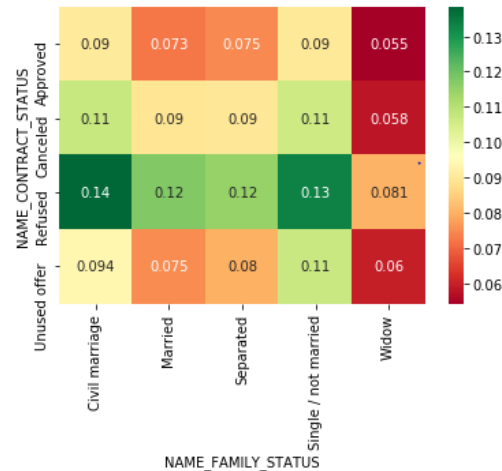
- The clients whose loans were previously refused, larger proportion of them have payment difficulties currently more than other cases
- The clients whose loans were previously approved, least proportion of them have payment difficulties currently
- As seen above, it is almost 12% for refused loans while 7.5% for the approved loans

Key Indicator: Age



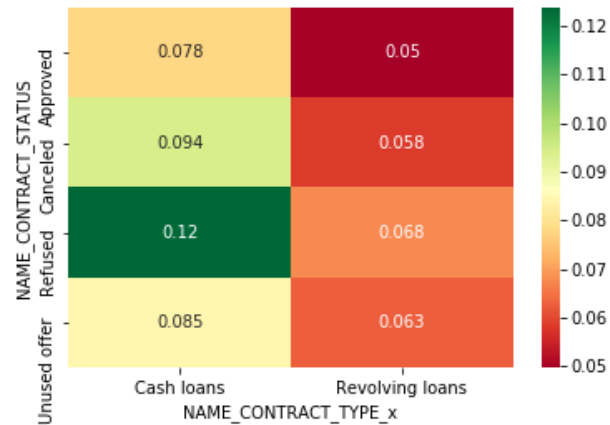
- Age group 17-35 show the highest proportions of clients who have payment difficulties so the riskiest for the bank
- Age group 55 and above shows the lowest proportions of clients who have payment difficulties and so are the least risky for the bank

Key Indicator: Family Status



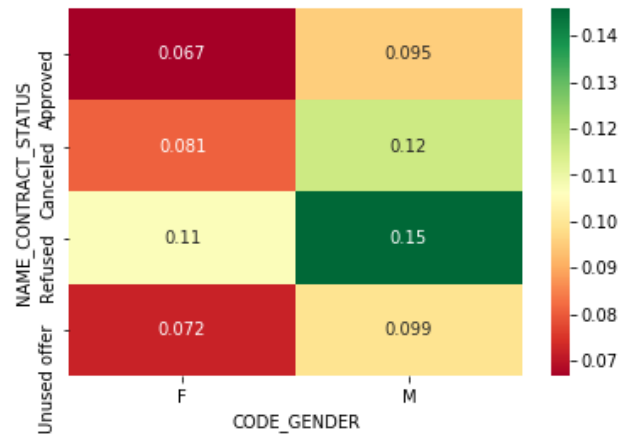
- Widows show the lowest proportion of clients who have payment difficulties so are the least risky for the bank
- Unmarried and civil married clients show the highest proportion who have payment difficulties and so are the riskiest for the bank

Key Indicator: Contract Type



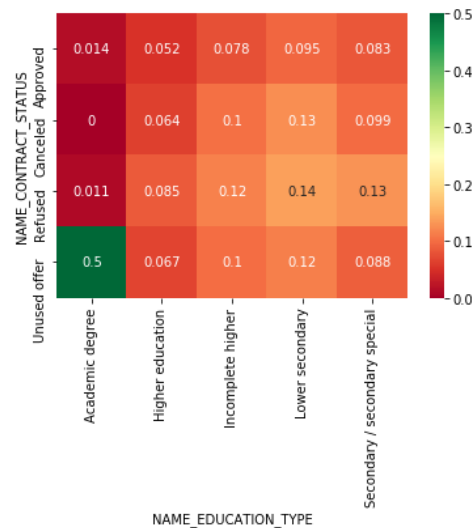
- Cash loans have a significantly higher proportion of clients who have payment difficulties compared to revolving loans
- On an average, proportion of clients with payment difficulties is ~9.5% for cash loans while only ~6% for revolving loans

Key Indicator: Gender



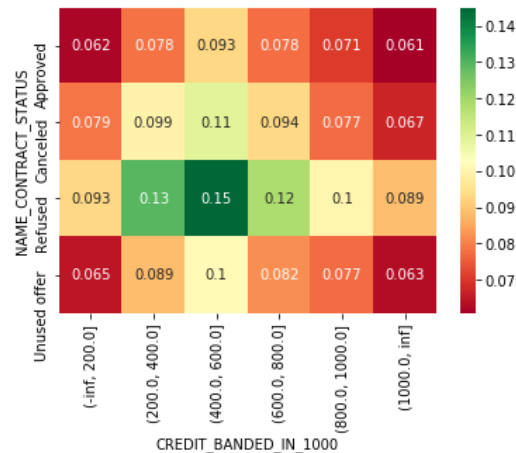
- Female clients have a lower proportion of clients who have payment difficulties compared to males and so are less riskier for the bank
- On an average, the proportion of clients who have payment difficulties is ~11.6% for Males while only ~8.2% for Females

Key Indicator: Education Level



- Clients obtained education only till lower secondary have a higher proportion of ~12% compared to rest
- Clients who have an academic degree and previously did not use the offer have a very high proportion of ~50% of having payment difficulties

Key Indicator: Loan Amount



- Loan credit amount in the range of 4 lacs to 6 lacs has the highest proportion of ~11.3% of clients having payment difficulties

Recommendations

- Avoid sanctioning loans to clients whose applications were previously rejected
- Avoid sanctioning loans to the age group of 17 to 35 and to sanction more loans to age group of 55 and above
- Prefer sanctioning loans to widows and avoid to sanction loans to unmarried and civil married clients
- Female clients should be preferred more over males since lower risk
- Clients who have completed education till lower secondary only should be avoided because they are bad risks
- Sanctioning revolving loans is a better option than cash loan since lower risk for the bank and hence, it could also be used to hedge the risk of cash loans
- Loan credit amount of 4 lacs to 6 lacs is the riskiest range and so should be avoided

Conclusion

After all the exploratory data analysis done for the bank, it should keep a close look on these key factors to not miss a good business opportunity and also not to take a decision which could result in a default.



Thank You.