EDA Case Study
-Bank Loan
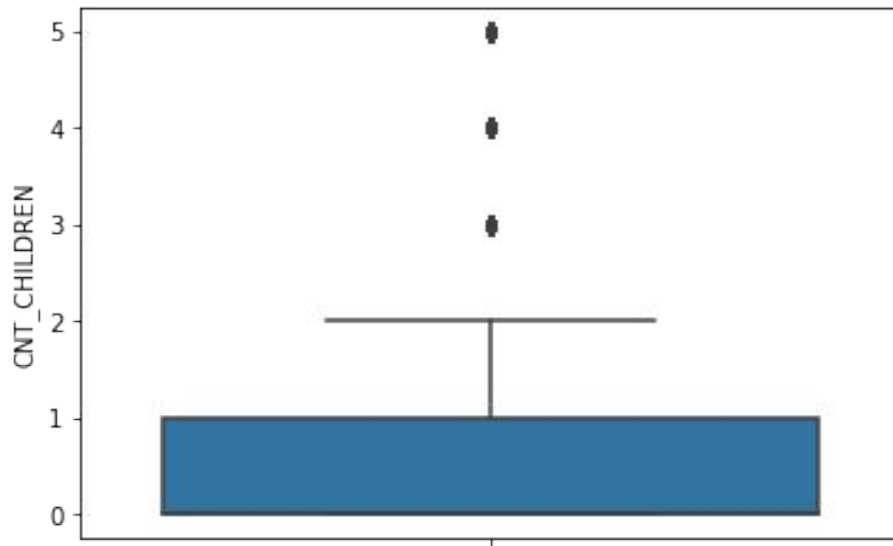
By:- Amit Chawla and Harnoor Singh

Objective -

- Analyze the type of clients those apply for loans with the given financial institution
- Analyze patterns for determining clients who face payment difficulties
- Based on the patterns identified above decide whether a new applicant should be given a loan.
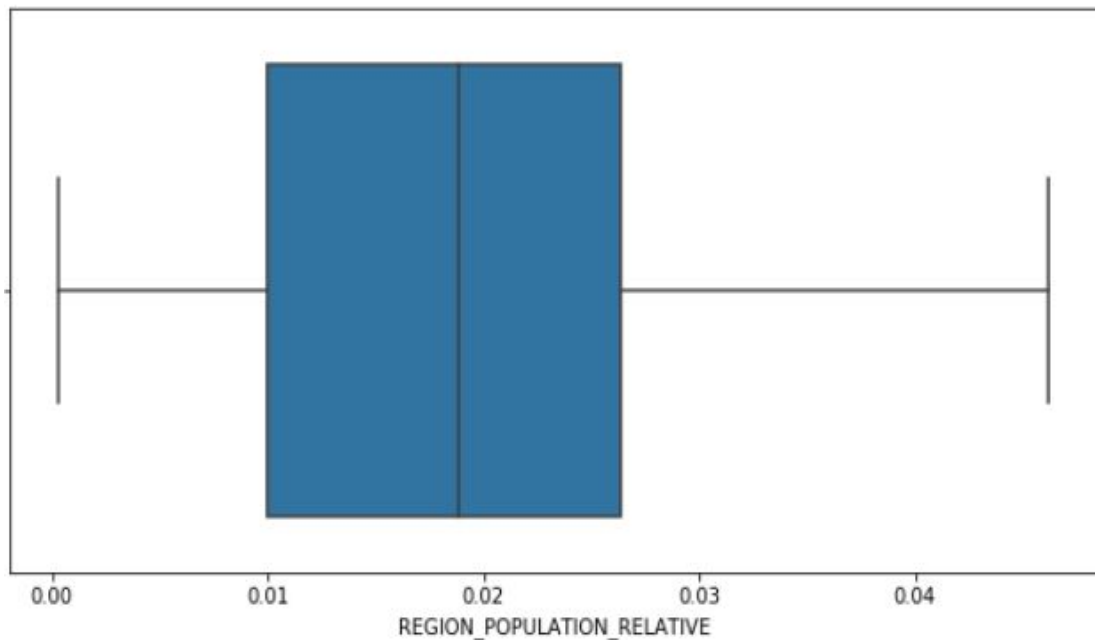
Loan application - CNT_CHILDREN

As we can observe, loan are mostly applied by clients with 0 or 1 children.

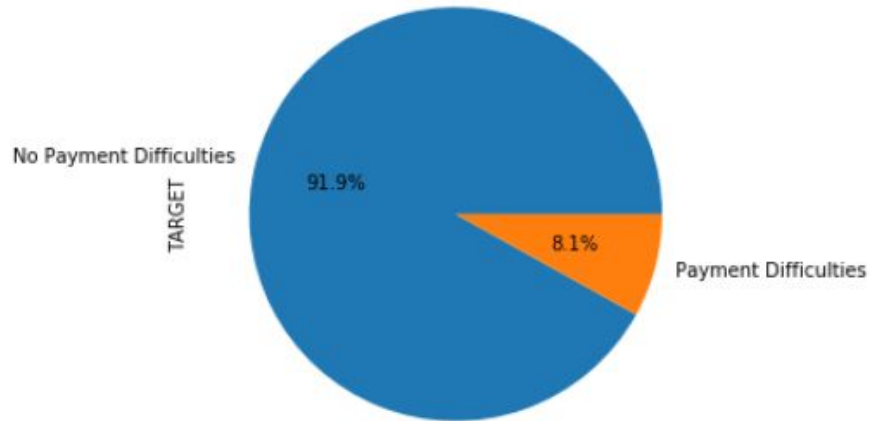There are also some data points where clients have 2-5 children.

Presence of institution
- based on population of region

97.5% of the applicants are from low to medium population density region.
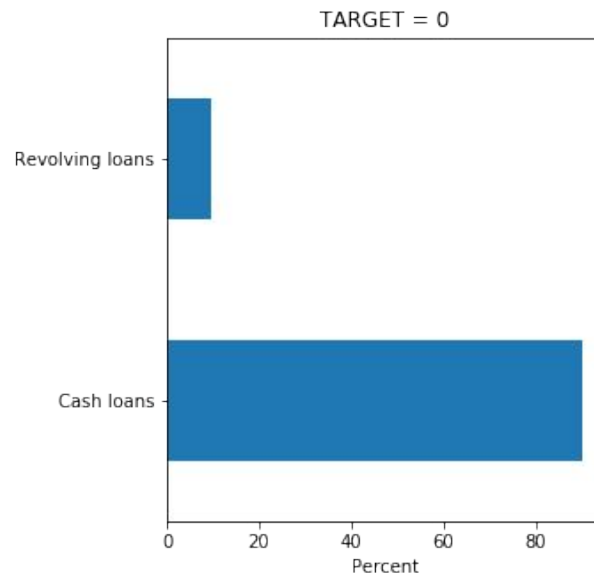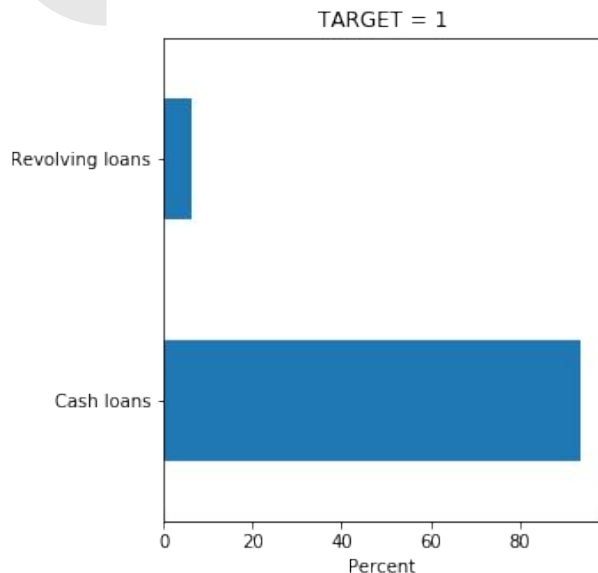
Ratio of payment defaults

- 91.9% people did not face any payment difficulties while 8.1 % did.

Type of loan and payback

There are slightly more data points for TARGET = 0, where clients had no difficulty in payment for revolving loans. So, clients who opted for revolving loans are slightly more likely to pay back without difficulty

Gender distribution on TARGET
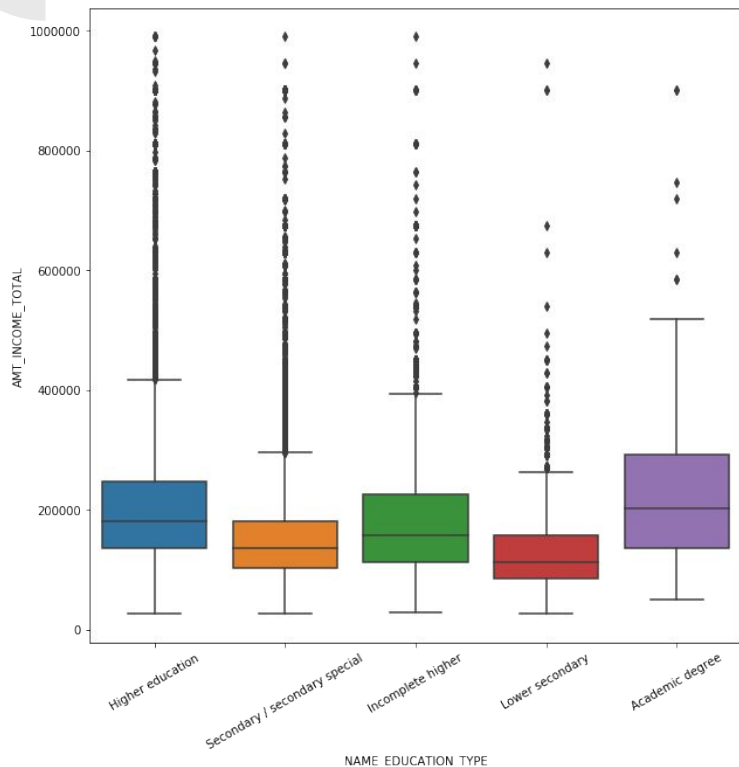
For clients who had payment difficulty, around 57% of them whereas females, where males were around 43%.

TARGET = 0

TARGET = 1
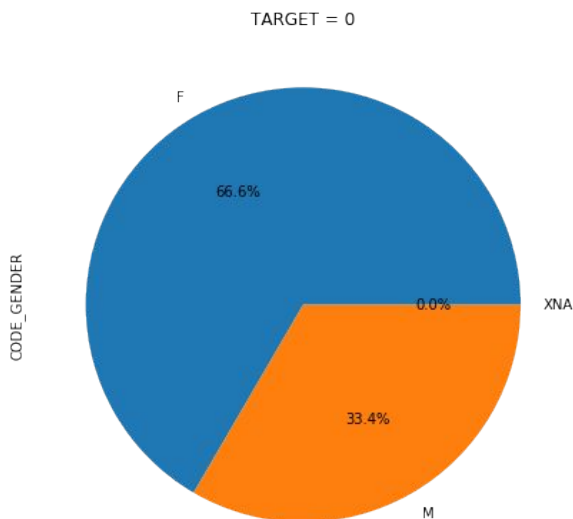
For clients who had no payment difficulty, around 65% of them whereas females, where males were around 33.5%.
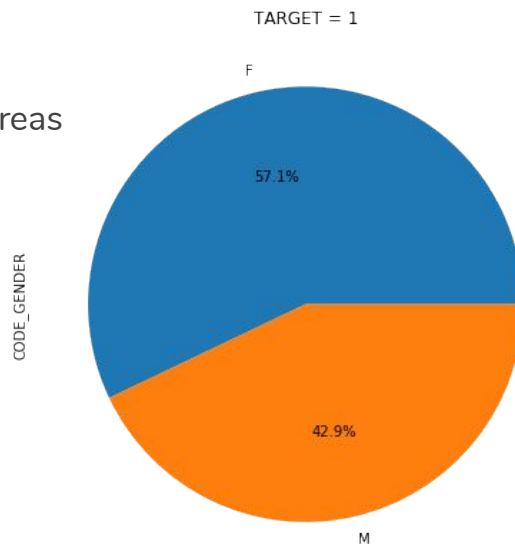
# Total Income vs Education Level Analysis

People with NO payment difficulties

- Majority of on time paying clients lies in higher range of AMT_INCOME_TOTAL

- Clients with lower secondary who opted for large sum of loan (945000.0) have paid on time while clients of same education who took loans of moderate size (450000.0) have defaulted.
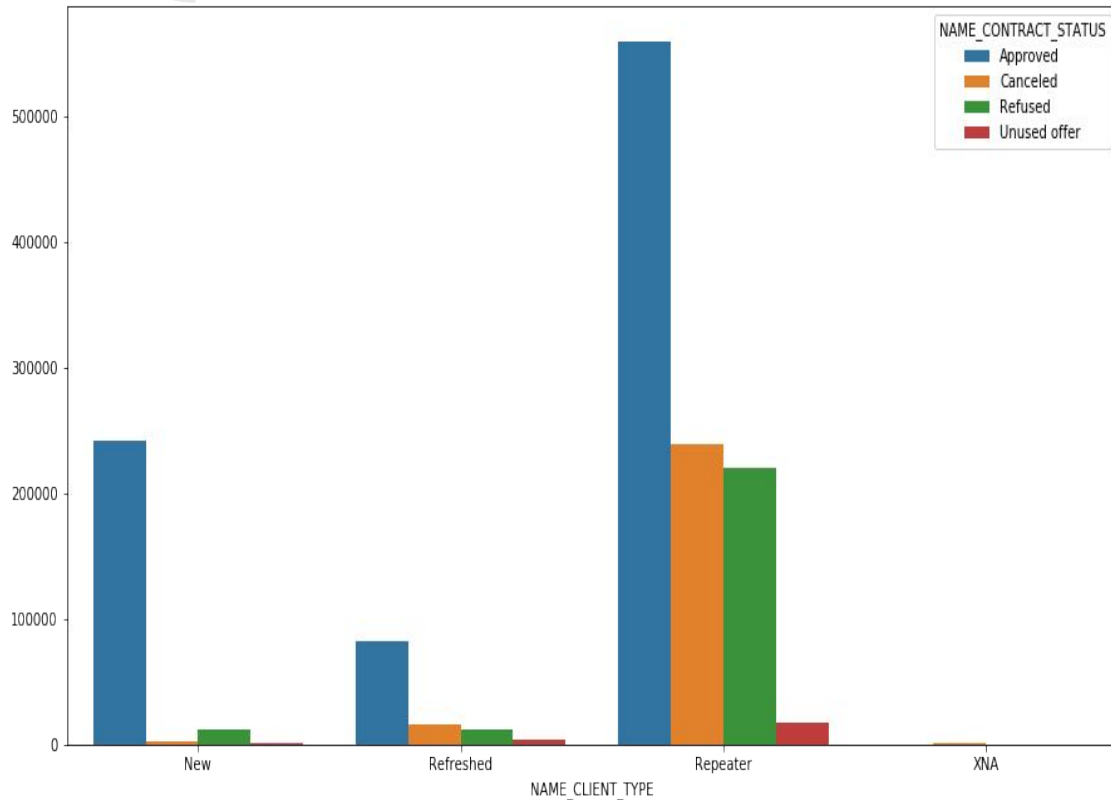
For clients with payment difficulties

- Although academic degree seems to be in higher range, there are only 3 cases of default for clients with academic degree and those are in middle income range.

- Clients with lower secondary education level are more likely to default for Total Income< 400000
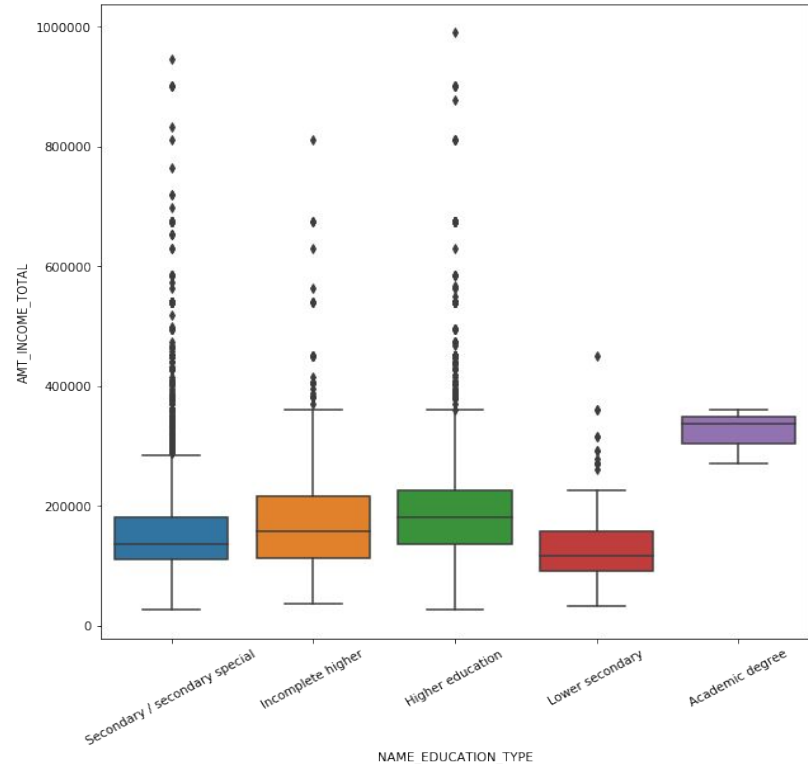
People with payment difficulties

CONTRACT STATUS VS CLIENT TYPE

- The count of repeaters is high, clients are loyal to the bank.
- The Approved percentage is high for New, Refresher and Repeater. It is significantly higher for Repeater.
- Likelihood of new applicants cancelling the application is vey low.
- The percentage for Canceled and Refused is also significantly higher for Repeater.
- Refreshed clients have slightly higher Canceled percentage as compared to New clients
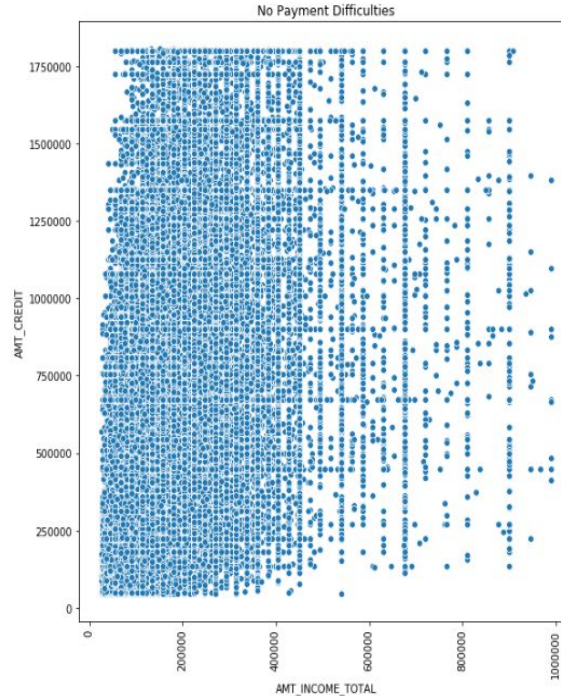
Impact of Income on payment difficulties
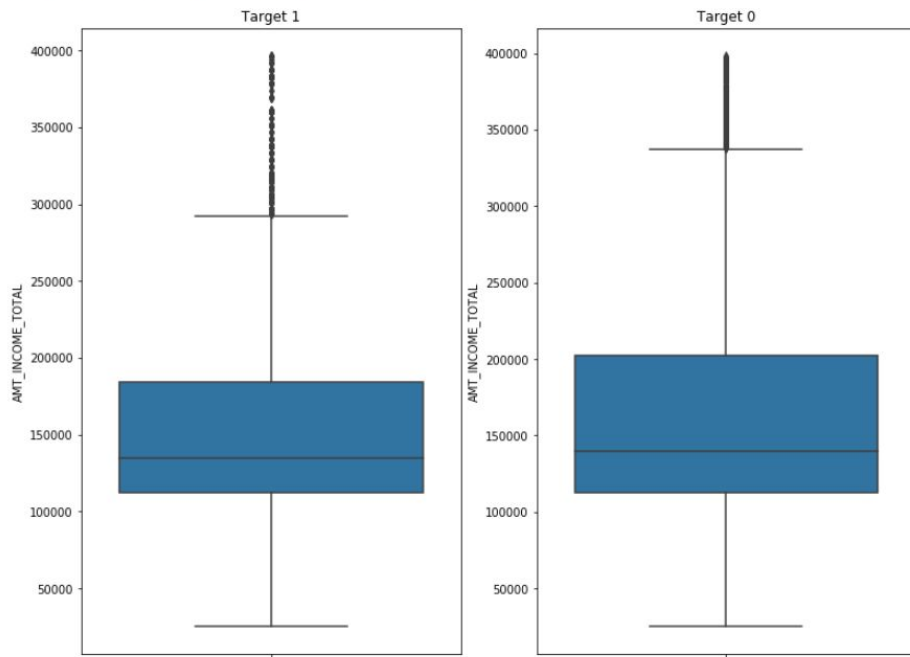
- For people with income more than 10 lac there is very less risk of default
- People with income less than 3 lacs are most likely to face difficulties in payments
- More specifically, majority of people who faced difficulties had income between 1 lac and 2 lacs approximately

Impact of Income on payment difficulties
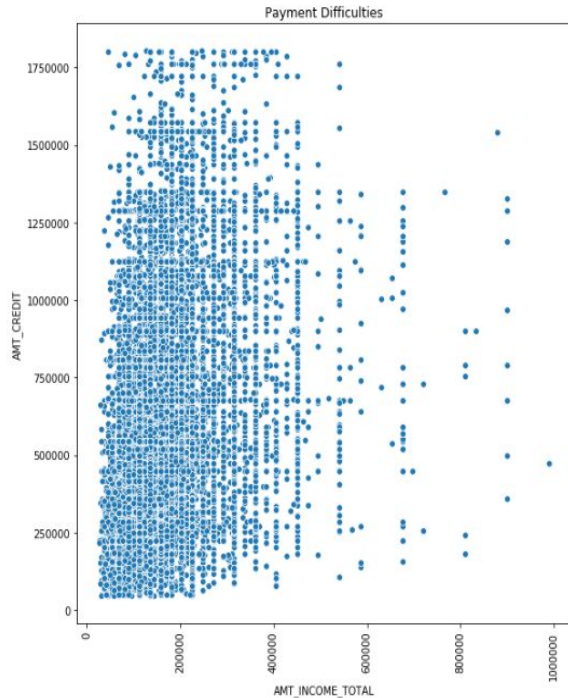
- Upper limits of the credits in case of people with payment difficulties is less.
- Most of the payment difficulties are associated with the people with incomes less than 5 lacs

Applicants relation with bank
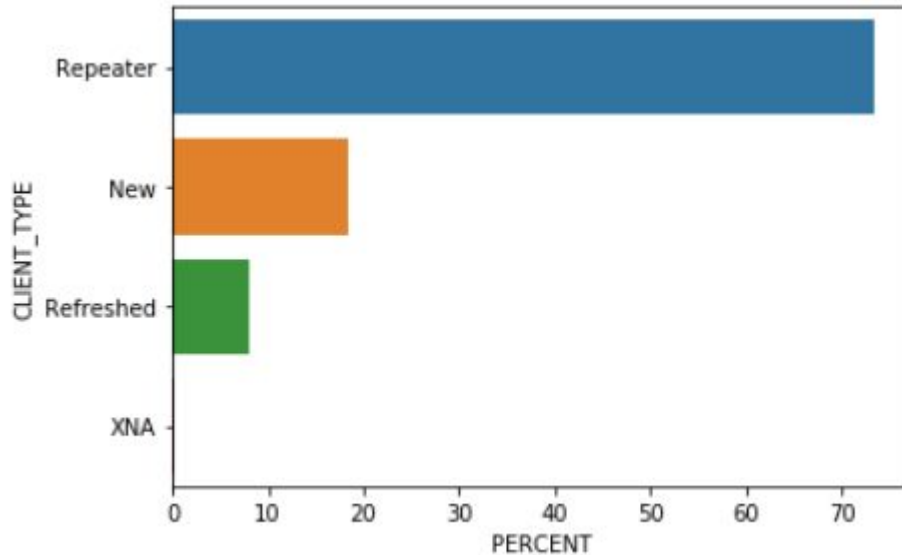
- Most of the applications are from Repeaters.
- Clients are loyal to the bank
- There are decent New applications too. Hence, the banks might have some attractive offers since it is attracting new applications too

Previous application processing record
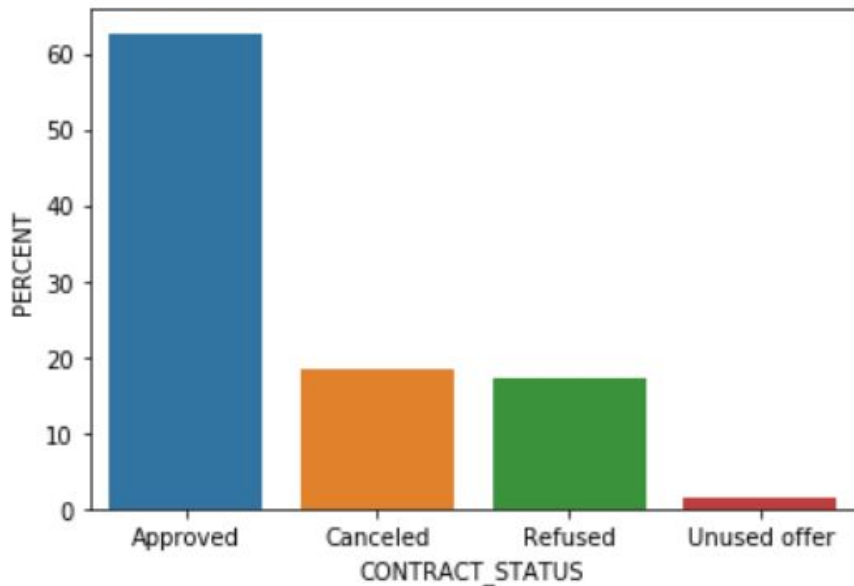
- 62.6% of the previous loan applications were approved.
- 18.35 % applications were cancelled
- Only 17.35% applications were refused.
- 1.6 % applications were cancelled at some different stages.

Remarks - Bank may want to review the reasons for cancellations of the applications

Number of days since last application

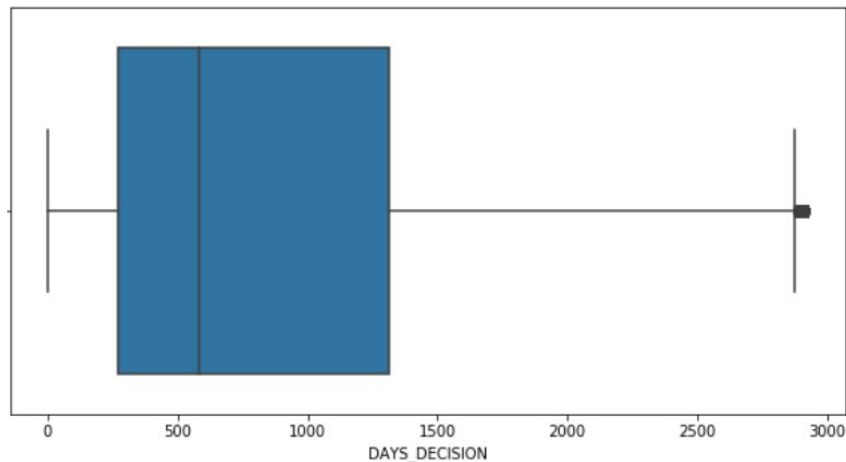- Clients take an average 2.41 years (880 days) to apply a loan again with the bank
- There are a few cases where clients apply a loan instantly relative to decision made to previous application.
- The median lies around 1.64 years (600 days) and the 99 percentile lies around 7.94 years (2900 days). Hence, half of the loans are applied much sooner than the others.

AgeGroup vs CODE_GENDER vs NAME_CONTRACT_STATUS

- Females of the age 40-50 have slightly higher count for loan getting approved than females of the age 30-40 and 50-60

- Males of the age 30-40 have slightly higher count for loan getting approved than the males of age 40-50 and 50-60

CODE_GENDER distribution with CONTRACT_STATUS and AgeGroup
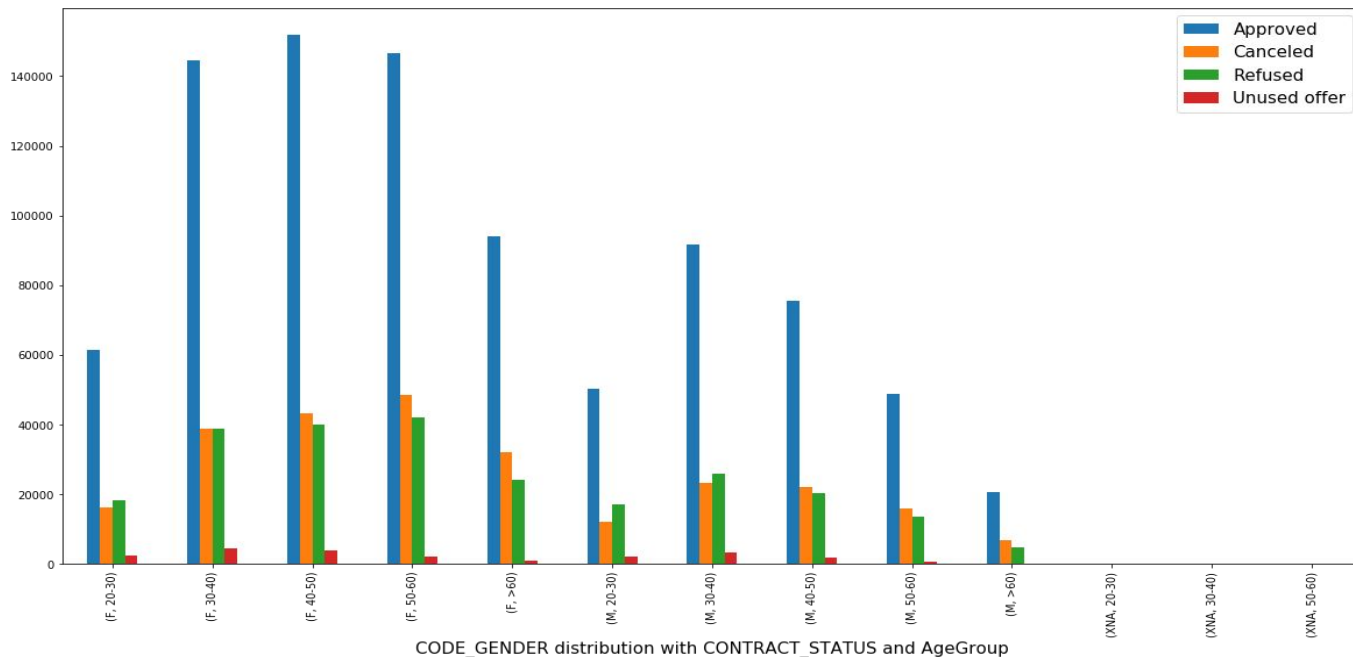
Highest AMT_APPLICATION loans

SK_ID_CURR	AMT_APPLICATION	NAME_CONTRACT_STATUS	TARGET
975055	346243	5850000.0	Approved	0

- We observed that highest AMT_APPLICATION for loan amounted to 58.5 Lacs.

- The loan was approved.

- Client who applied for this loan has successfully paid back on time without default.

Final Words...

- Loan applications are mostly from clients having 0 or 1 children and the loan amount is mostly less than 5 Lakhs.

- Females have a high percentage in both, defaulting a loan and paying back on time.

- Clients with Lower secondary education level are more likely to default for AMT_INCOME_TOTAL < 400000

- The count of Repeaters is high, clients are loyal to the bank

- Highest number of people who have applied for loan and defaulted are in age of 25-40 years.

- Mostly payments are done by cash through the bank by clients.

...contd Final Words

- Mostly payments are done by cash through the bank by clients.

- Very high AMT_APPLICATION (greater than 50 Lac) was approved and the client successfully paid back on time.

- For people with income more than 10 lac there is very less risk of default

- AMT_CREDIT And AMT_ANNUITY have the highest correlation

- Mid-age females have the highest count for loan getting approved (40-50 years)

- Males of the age 30-40 have slightly higher count for loan getting approved

Thank you!

By:- Harnoor Singh and Amit Chawla