CREDIT EDA CASE STUDY

Group Members:
Vigneswara Ihita Gangavarapu(gihita@gmail.com)
Jaswanth NA(jaswanthna@gmail.com)

The Problem Statement

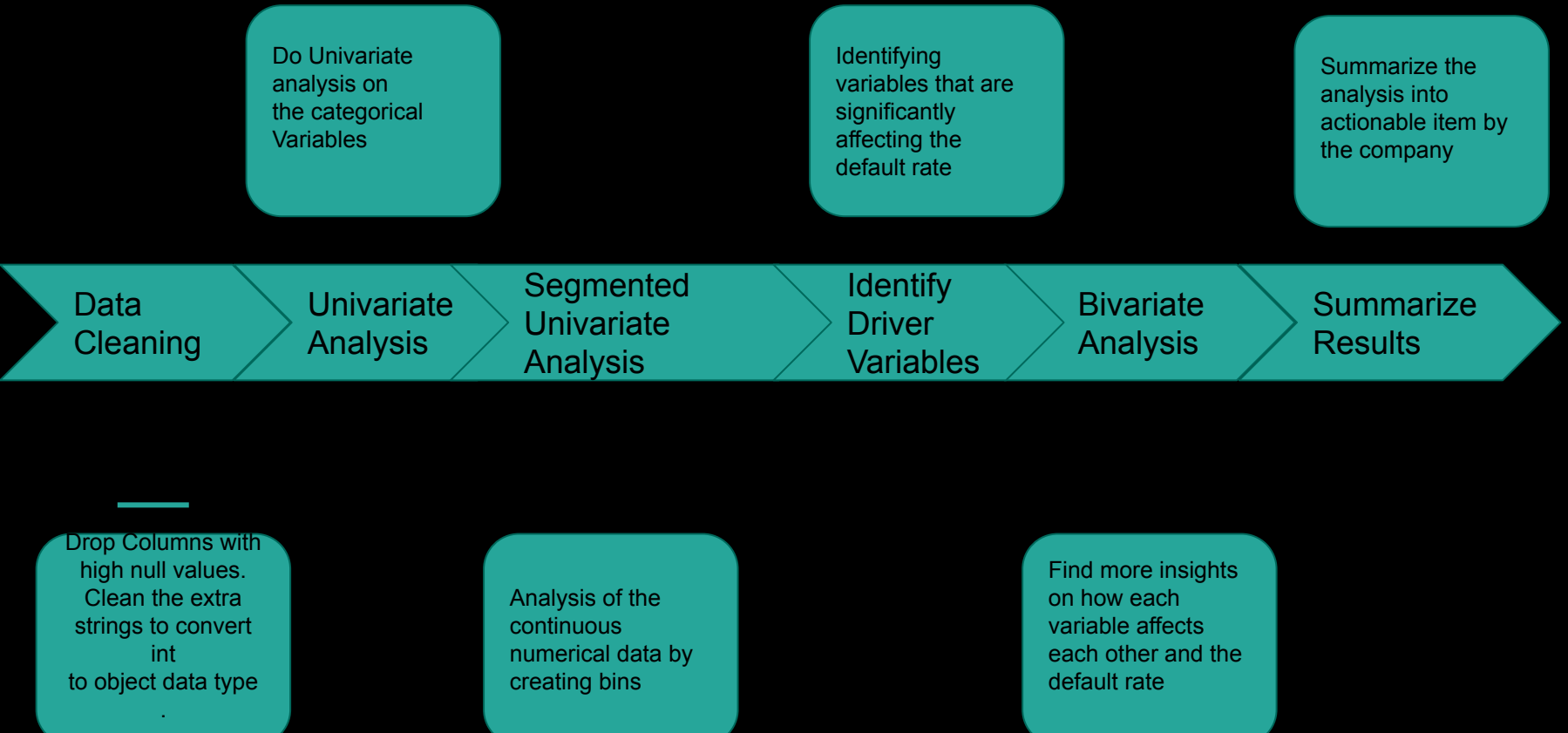
- It is challenging for a loan company to evaluate loan applications and provide loans to customers due to their insufficient or non existent credit history
- This results in two scenarios:
  - Not approving the loan results in a loss of business to the company if the applicant is likely to repay the loan
  - Approving the loan may lead to a financial loss for the company, if applicant is not likely to repay the loan
- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

We **aim to identify patterns and the driving factors** which indicate if a person is likely to default,which may be used for taking the following actions:

1. Approved the loan application
2. Refused the loan application

The company can utilise this knowledge for its portfolio and risk assessment.

Analysis Procedure



Summary of Univariate Analysis

The following are the top 7 driver variables:

1.NAME_EDUCATION_TYPE

2. CODE_GENDER

3. NAME_INCOME_TYPE

4. OCCUPATION_TYPE

5. NAME_HOUSING_TYPE

6. DAYS_BIRTH

7. AMT_CREDIT

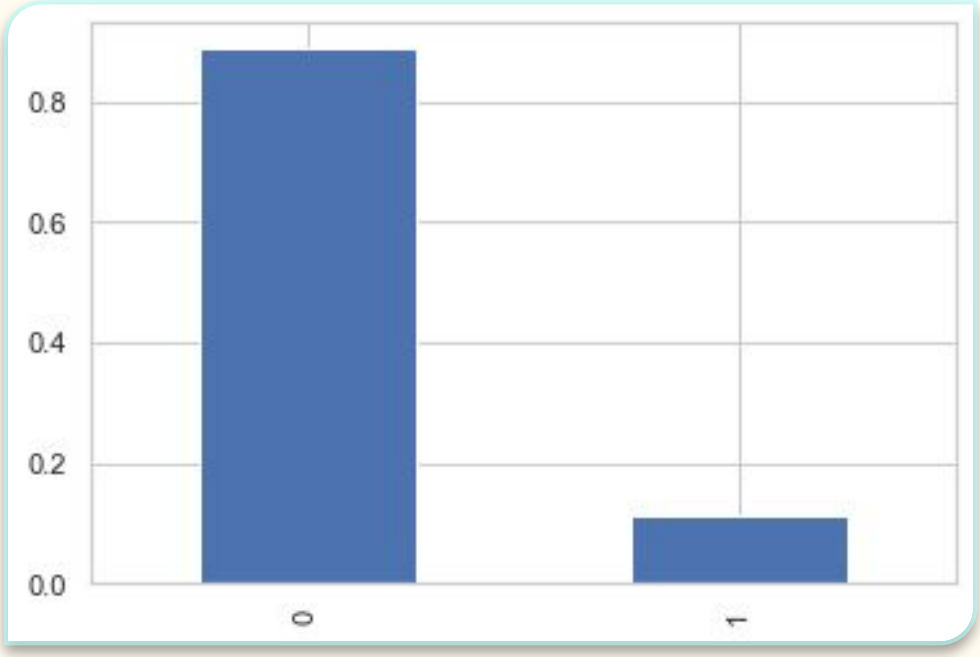
Variable: TARGET

TARGET is our target variable and we want to deduce how other variables affect this variable.

TARGET=1 indicates clients with payment difficulties(Defaulters)

TARGET=0 indicates all other cases (Non- Defaulters)

Graph between TARGET variable and count of TARGET Variable values: 1 and 0



TOP 10 CORRELATIONS

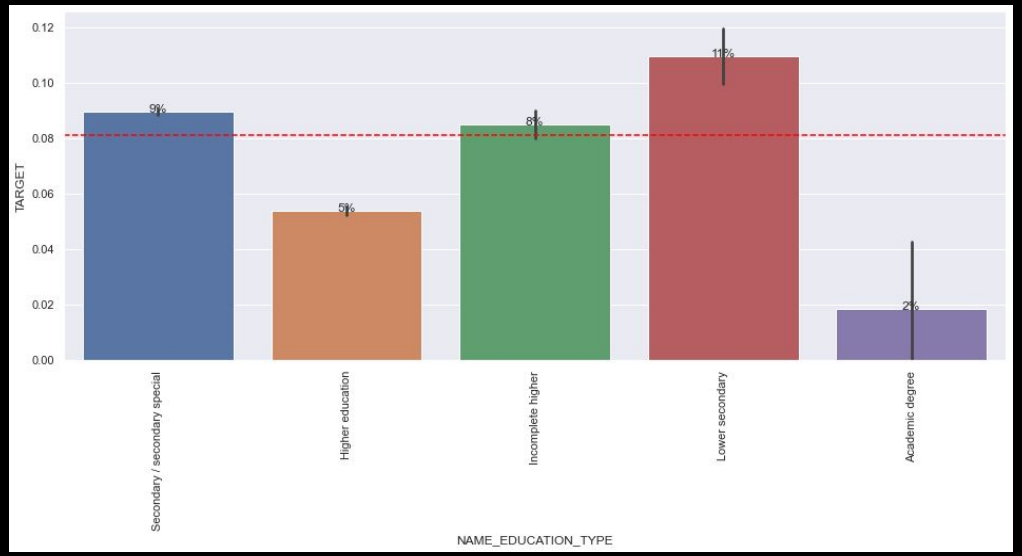
AMT_REQ_CREDIT_BUREAU_YEAR	AMT_REQ_CREDIT_BUREAU_YEAR	1.000000
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.997061
AMT_GOODS_PRICE	AMT_CREDIT	0.976430
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.857446
CNT_FAM_MEMBERS	CNT_CHILDREN	0.823307
AMT_ANNUITY	AMT_CREDIT	0.811443
AMT_GOODS_PRICE	AMT_ANNUITY	0.804099
DAYS_BIRTH	DAYS_EMPLOYED	0.466526
AMT_INCOME_TOTAL	AMT_ANNUITY	0.431617
OBS_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.425990
DEF_30_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.423016

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.997362
AMT_GOODS_PRICE	AMT_CREDIT	0.984995
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.843584
AMT_CREDIT	AMT_ANNUITY	0.831900
AMT_ANNUITY	AMT_GOODS_PRICE	0.829818
CNT_CHILDREN	CNT_FAM_MEMBERS	0.810519
DAYS_EMPLOYED	DAYS_BIRTH	0.583711
AMT_ANNUITY	AMT_INCOME_TOTAL	0.485699
AMT_INCOME_TOTAL	AMT_GOODS_PRICE	0.419720
AMT_CREDIT	AMT_INCOME_TOTAL	0.416017

Top 10 Correlations for Target =0 and Target = 1 respectively

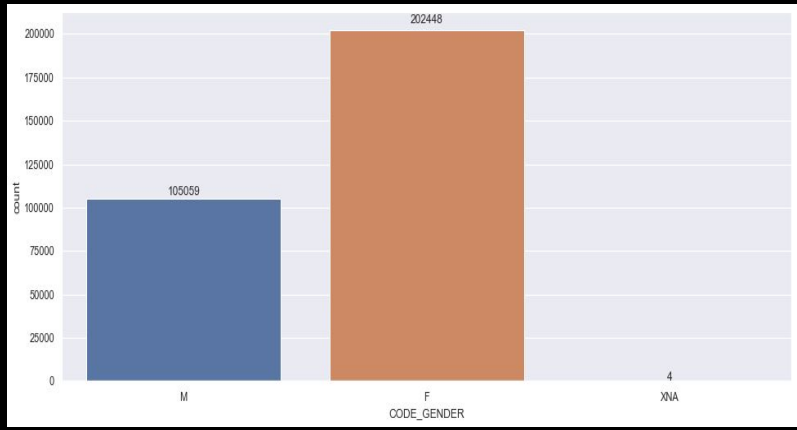
Analysis: Education

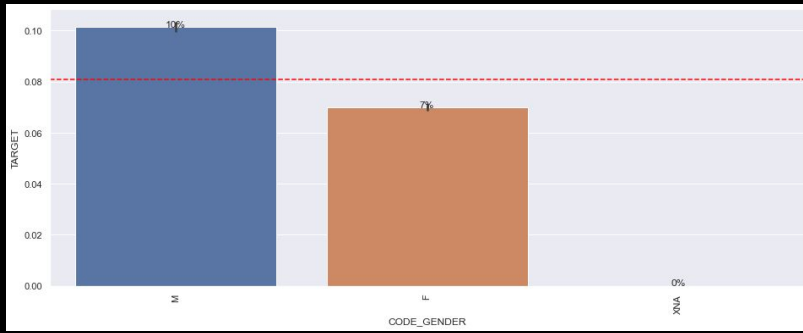
- We can see that those with lower secondary education have a higher chance of defaulting.
- Those with academic degree have much lower default percentage compared to the mean.
- To minimise the risks, we would want to be careful while giving out loans to those with lower secondary education.
- To maximize the returns, we would want to give loans to those with academic degree more easily



Analysis: Gender

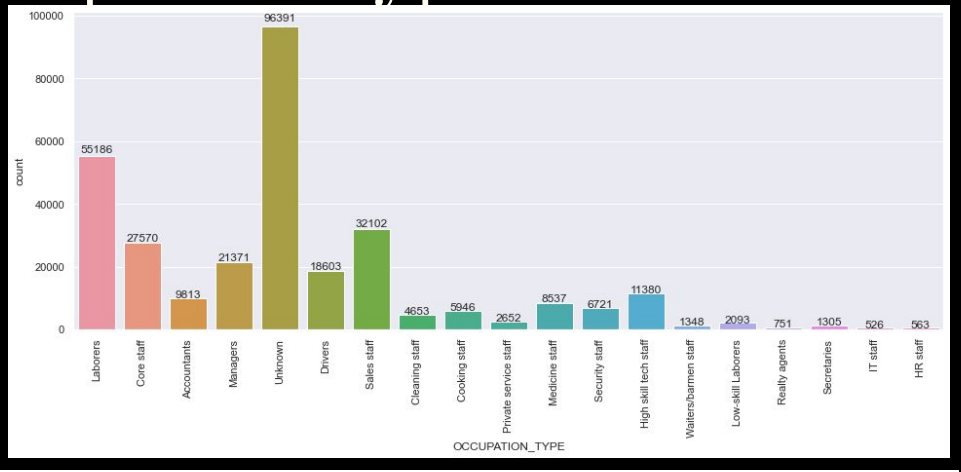
- We can see that even though females take more loans that males, they have a less default rate.
- Thus going by this analysis we can say that loans can be given to women much more easily and at a lower interest rate

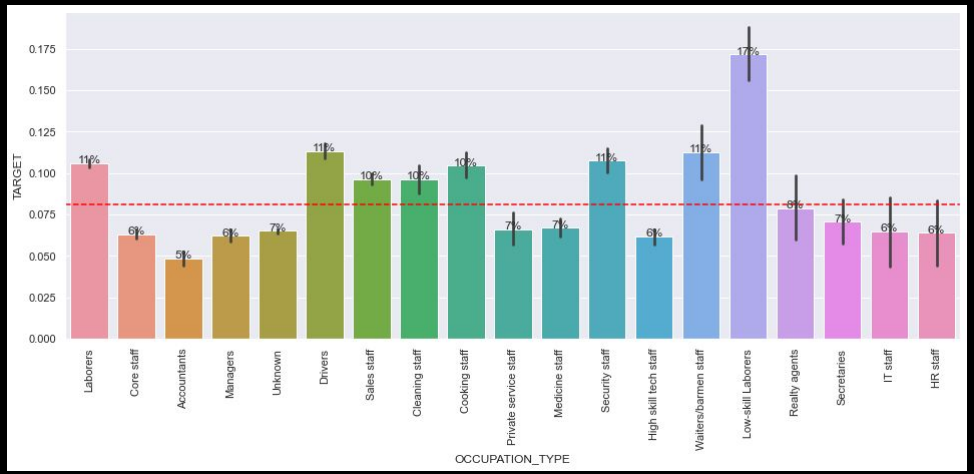




Analysis: Occupation Type

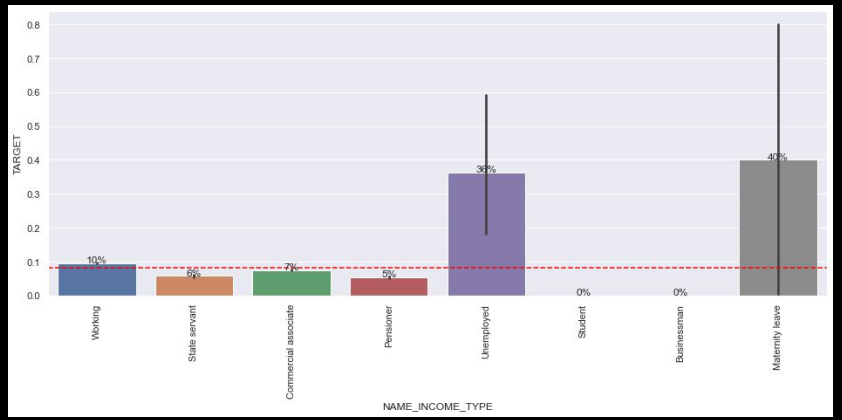
- We can see that laborers take the most amount of loans after Unknown.
- Though the amount of loans taken by low skill laborers is low they still have a high default rate.
- Accountants have the lowest default rates of all.





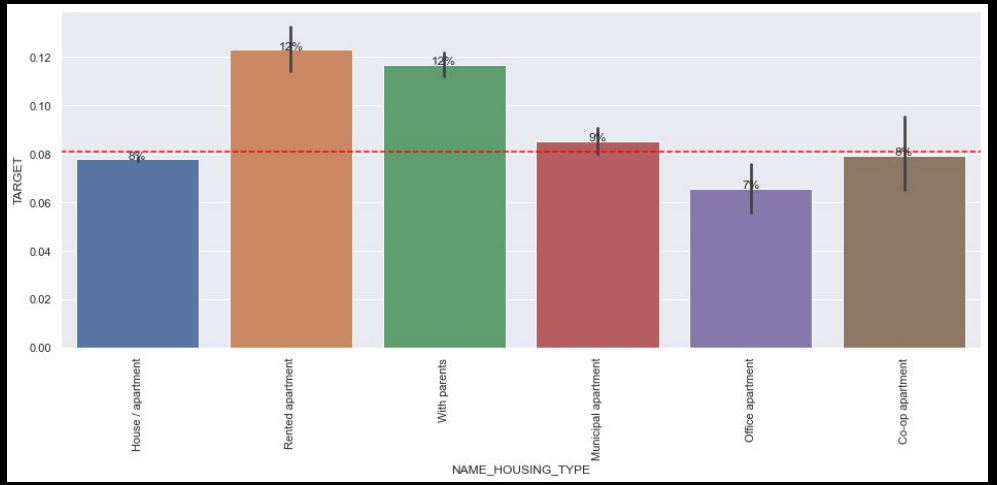
Analysis: Income Type

- Unemployed people have the most amount of defaults as is expected.
- Thus loans to unemployed people must be given after greater scrutiny.
- Giving loans to employed and state servants is a good idea and way to maximise profits for the company.
- But employed also have a higher than usual default rate and as they take the most amount of loans, there should be scrutiny if loan amount is large.



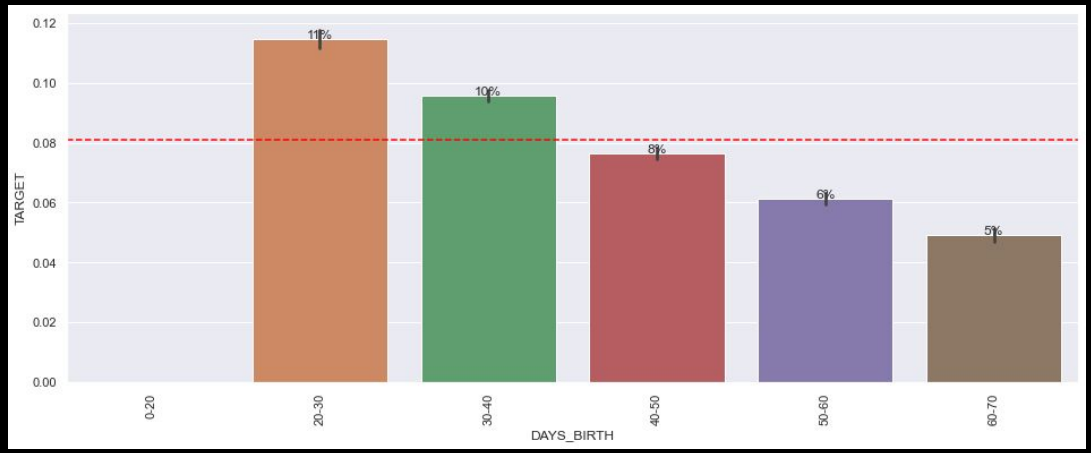
Analysis: Housing Type

- Those staying in rented apartments have the highest default rates
- Office apartments have the lowest.



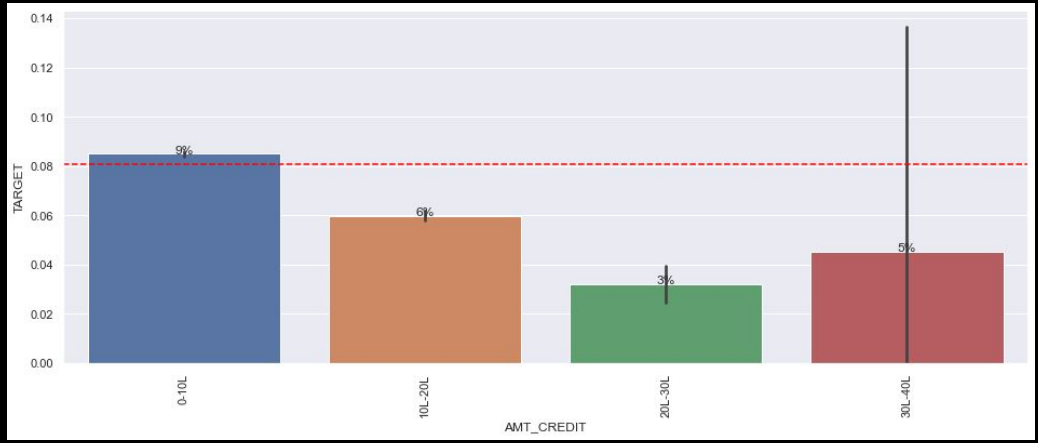
Analysis: Days Birth

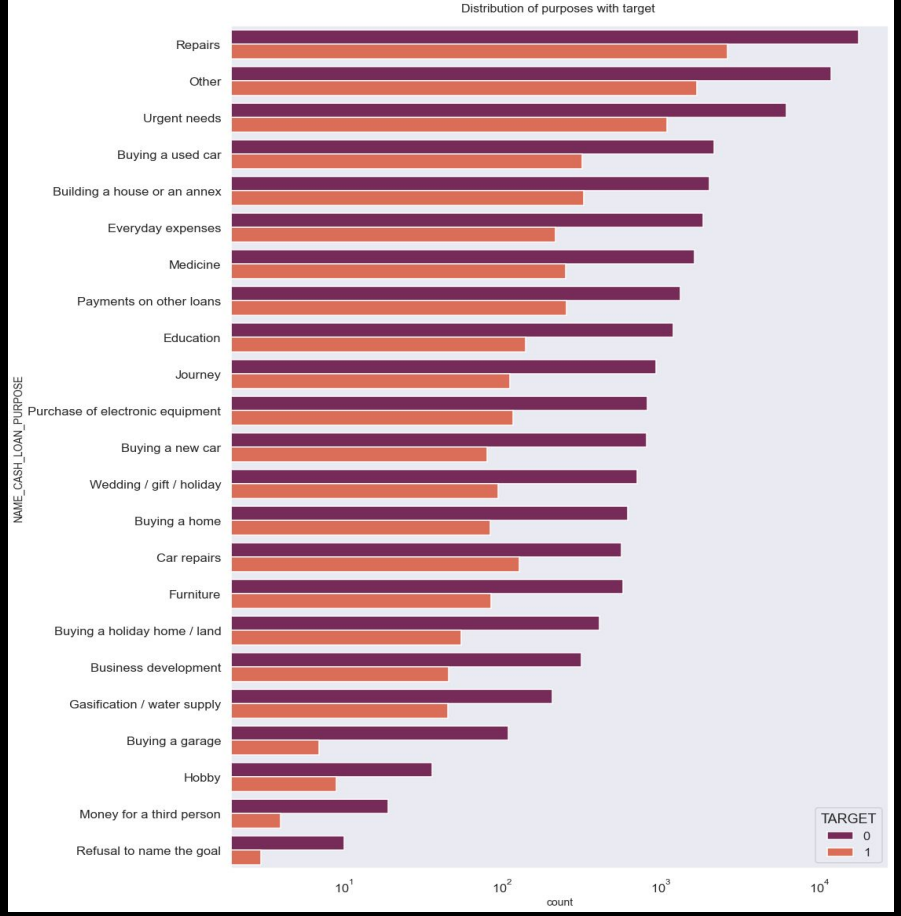
- We can see that with increasing age the defaults rate is decreasing.
- Older people can be given loans at a lower interest rate and far less scrutiny.

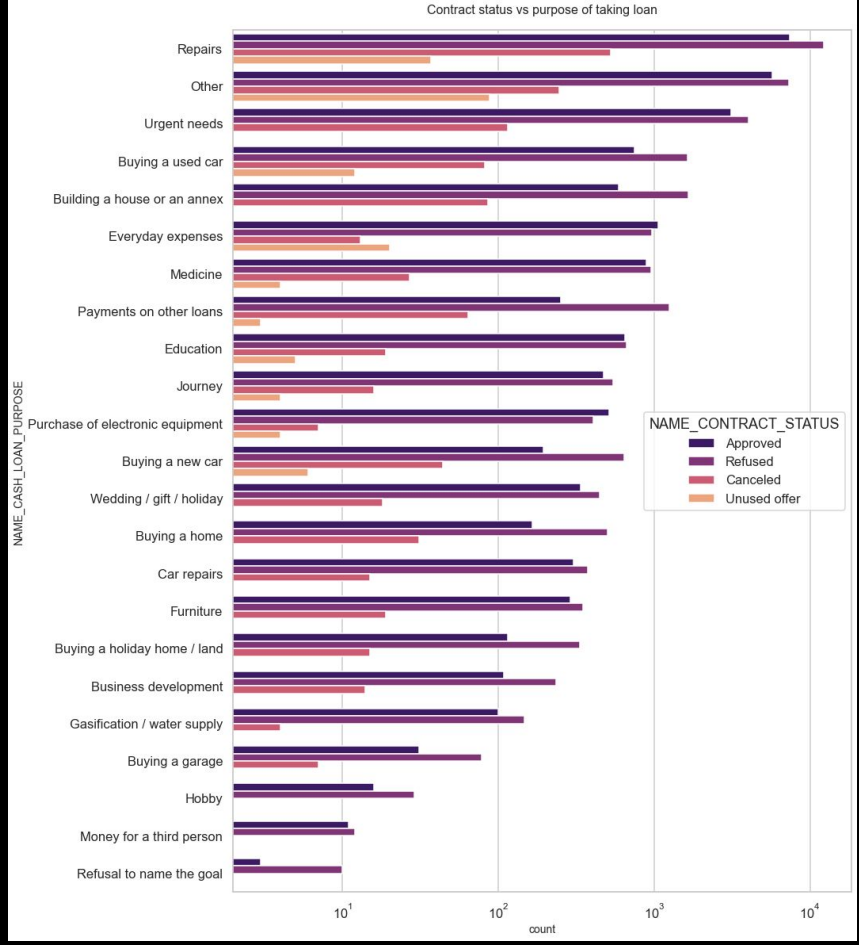


Analysis: Amt Credit

- Amounts which are lower in nature have more default rates than the higher one's
- The 20-30L range has a very good default rate for the company and should be something that it should look to maximise.







- Loan purposes with 'Repairs' are facing more difficulties in payment on time.

- We can see that most rejections of loans come from purpose Repairs

Recommendation

1. The 20-30L range should be loaned more as they must be giving higher returns to the bank and they have the lowest default rate.
2. Elderly people can be trusted with loans as they have a lower default rate and can be given loans with less scrutiny.
3. Females take more loans than male but still have a lower default rate, thus females can be given preference when giving loans.
4. Low skilled labourers have a high default rate and should be given loans with more scrutiny.
5. Loans for people with only lower secondary education should be given out with more scrutiny.