

Credit EDA Case Study



By :
Vyas Bhaumik Hemantkumar

Understanding Dataset

❑ The dataset has 3 files as explained below:

1. **'application_data.csv'** : contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.
2. **'previous_application.csv'** : contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.
3. **'columns_description.csv'** : is data dictionary which describes the meaning of the variables.

Business Problem

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter.

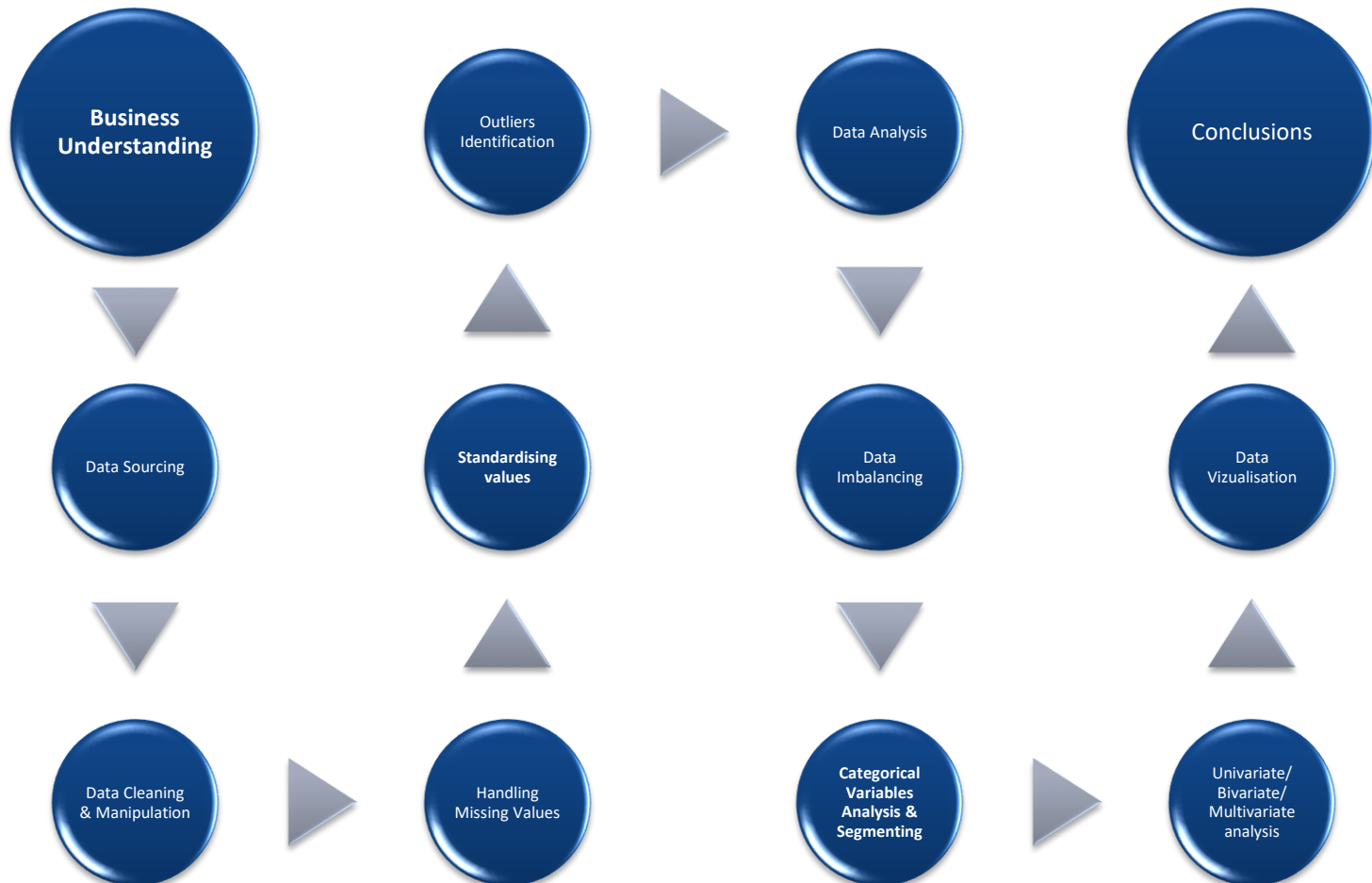
When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

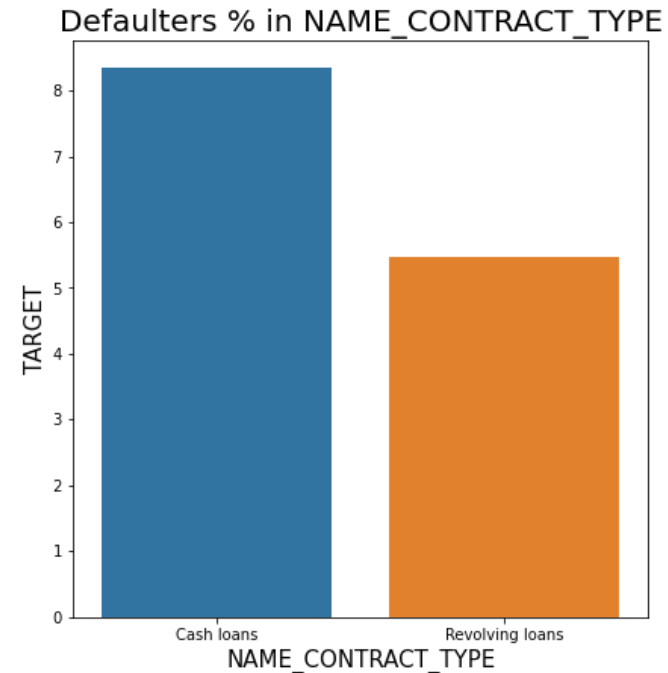
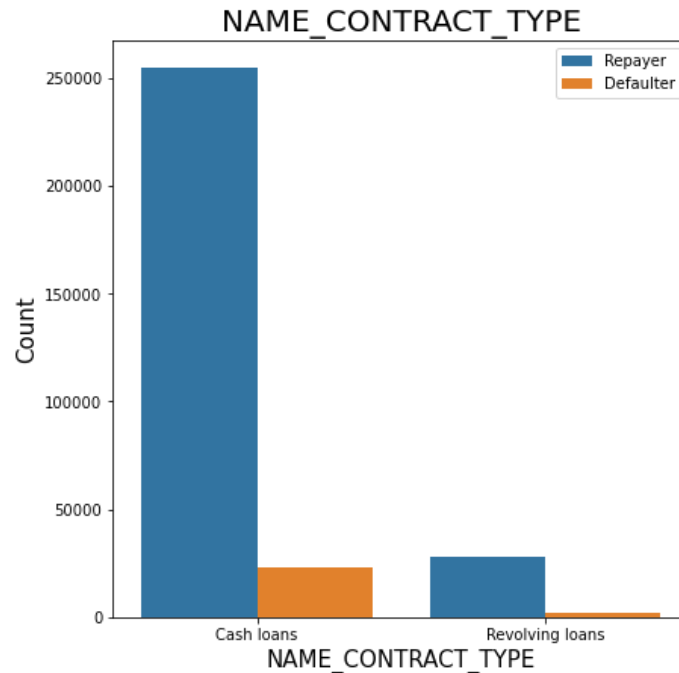
- 1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company**
- 2) If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.**

Business Objective

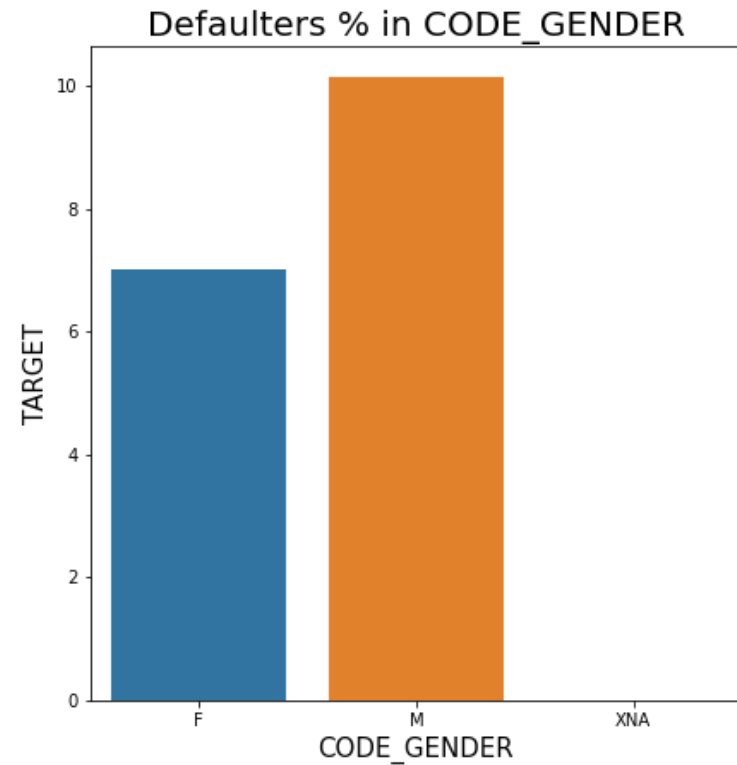
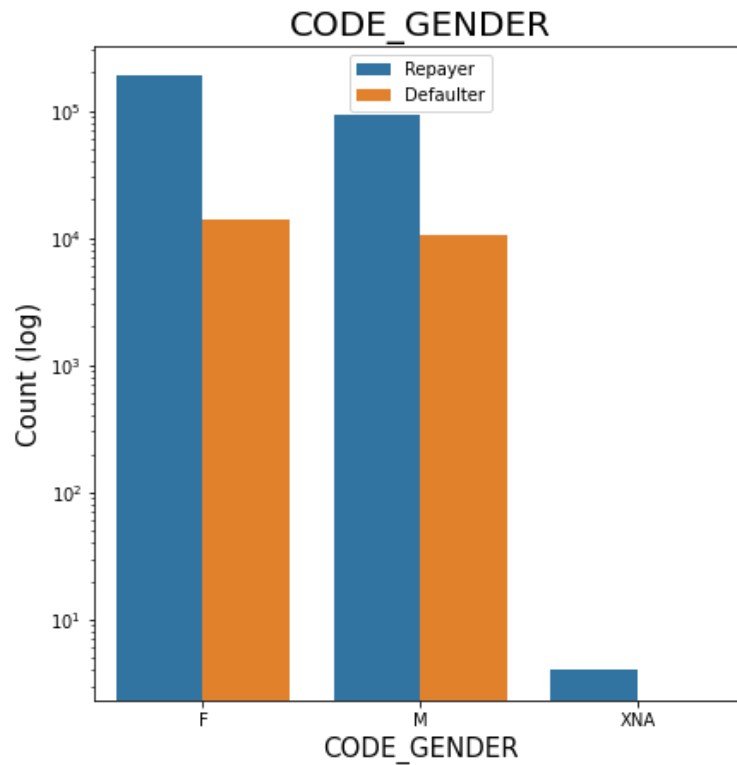
This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected.

Technical Approach For Solving Business Problem

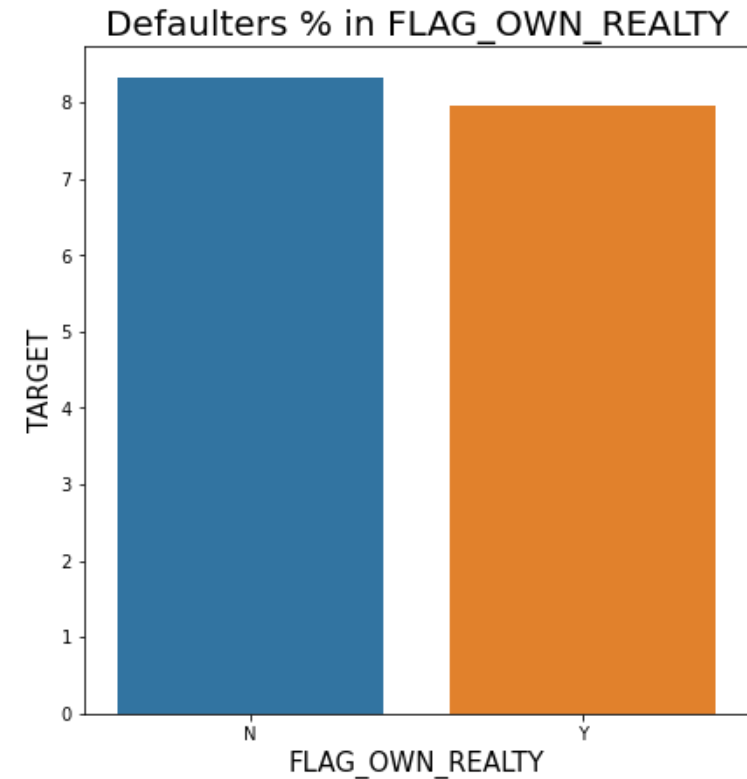
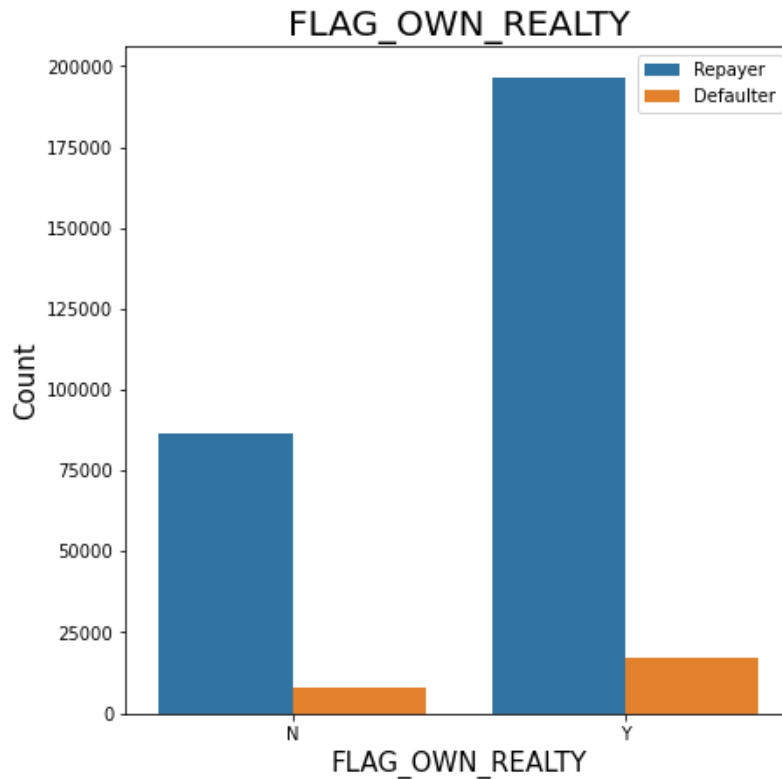




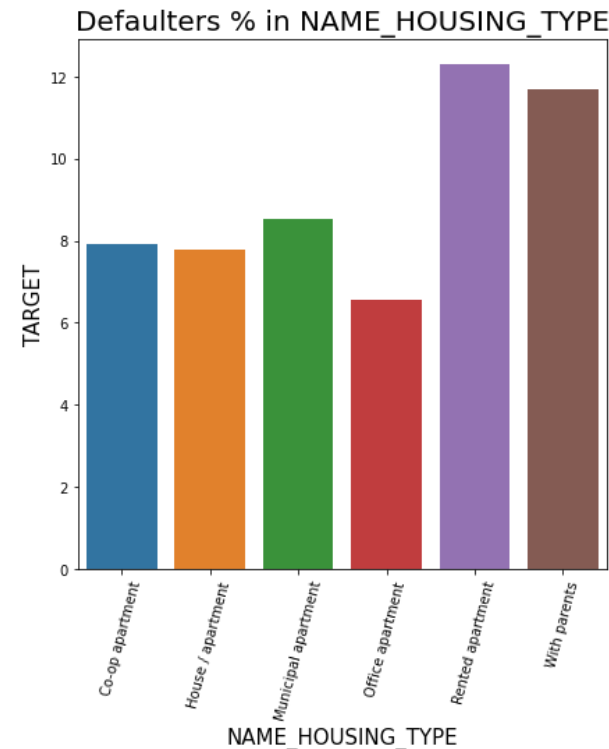
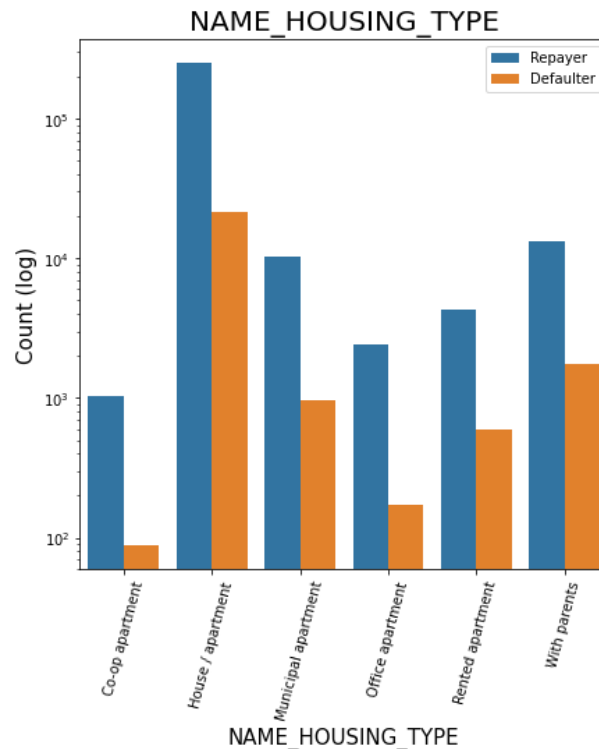
- Revolving loans are just a small fraction (10%) from the total number of loans.
- Around 8-9% Cash loan applicants and 5-6% Revolving loan applicant are in defaulters.



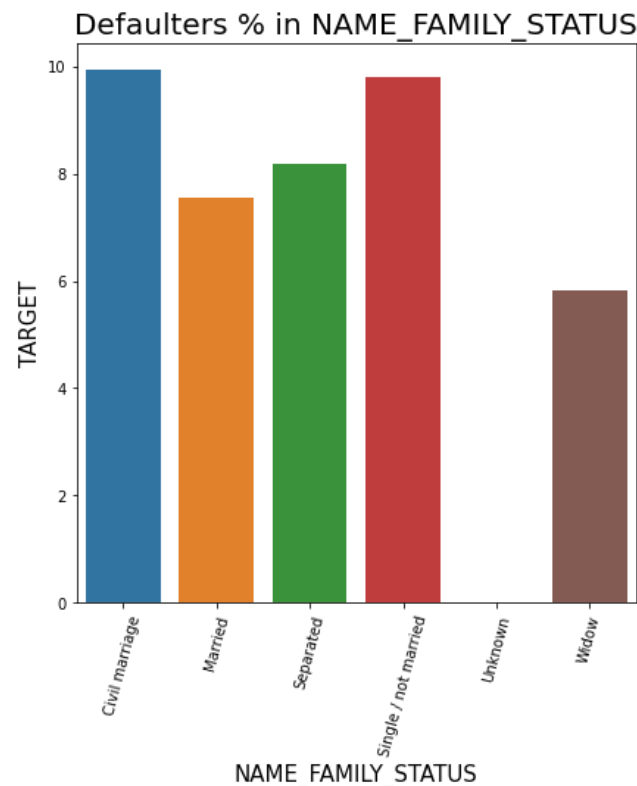
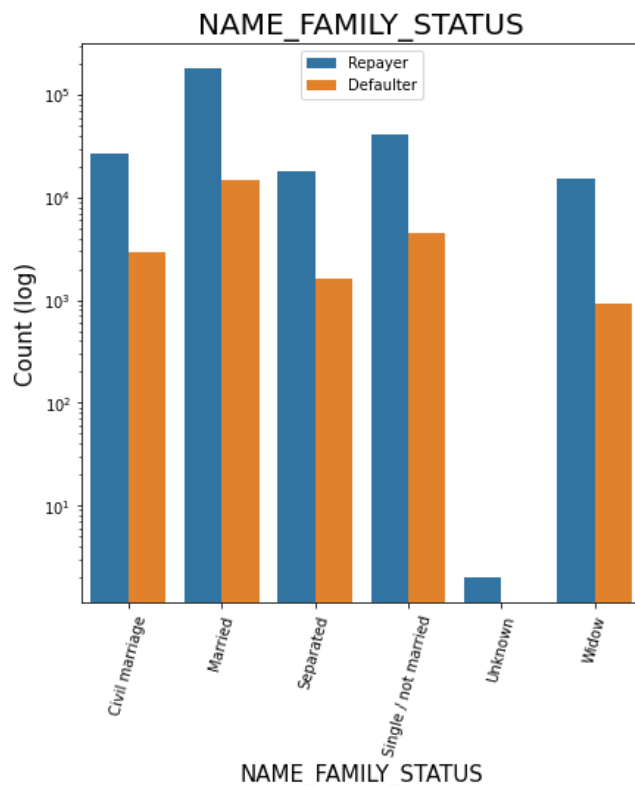
- The number of female clients is almost double the number of male clients.
- Based on the percentage of defaulted credits, males have a higher chance of not returning their loans about 10%, comparing with women about 7%.



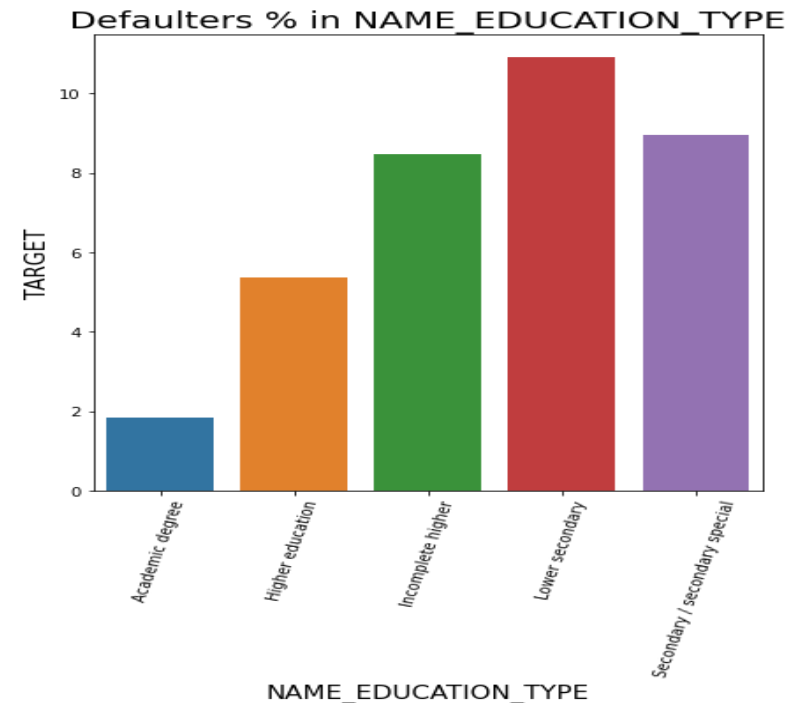
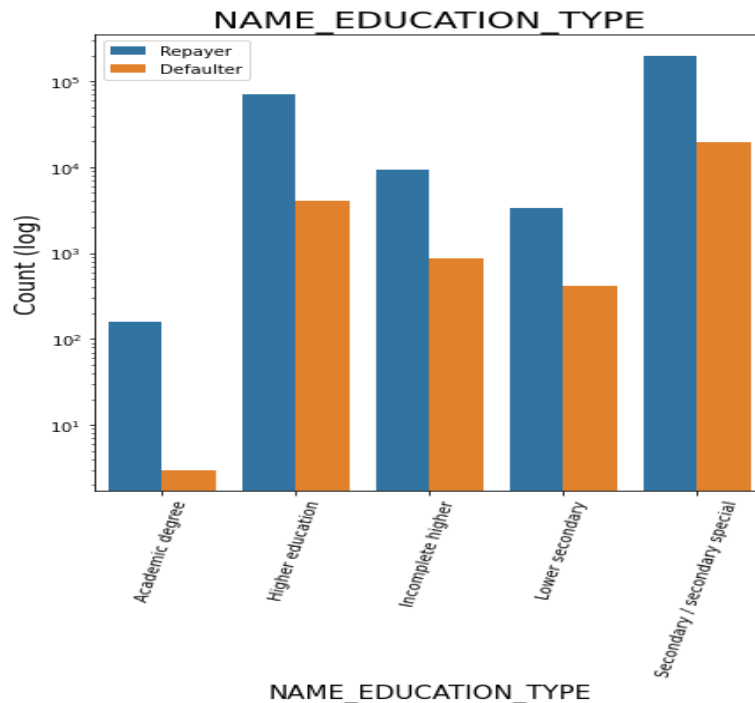
- The clients who own real estate are more than double of the ones that don't own.
- The defaulting rate of both categories are around the same (~8%). Thus we can infer that there is no correlation between owning a reality and defaulting the loan.



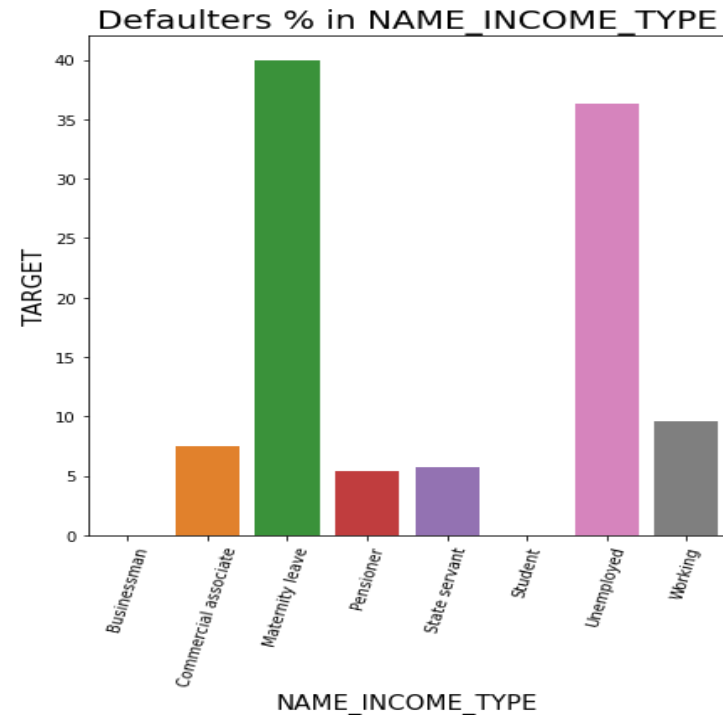
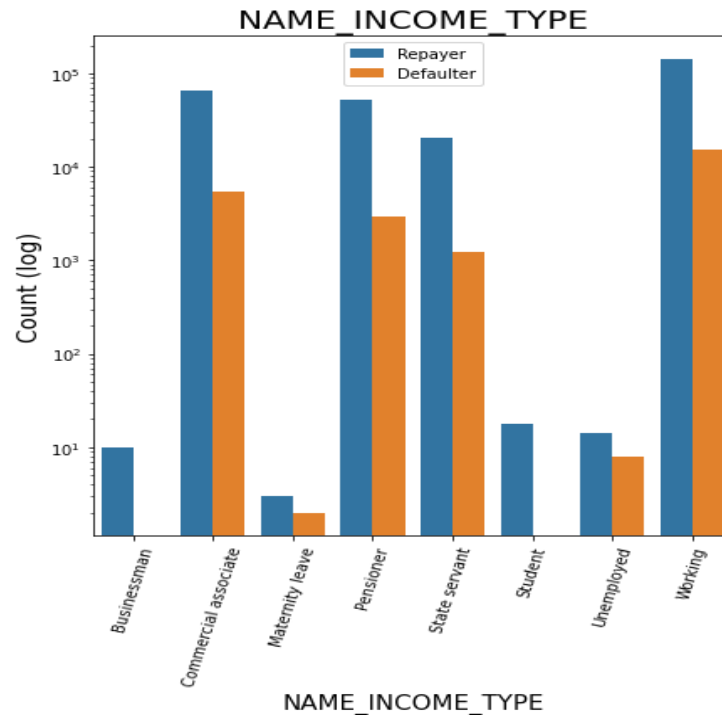
- Majority of people live in House/Apartment.
- People living in office apartments have lowest default rate.
- People living with parents (~11.5%) and living in rented apartments (>12%) have higher probability of defaulting.



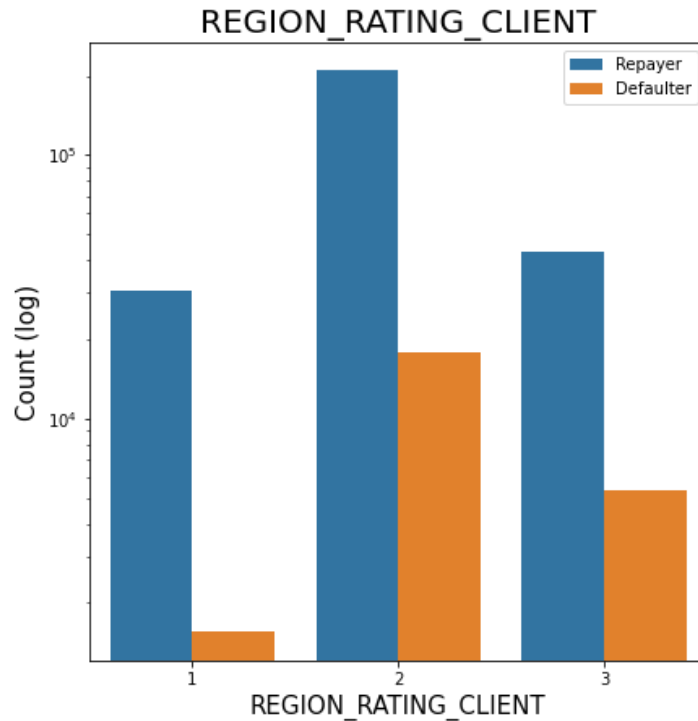
- Most of the people who have taken loan are married, followed by Single/not married and civil marriage
- In Percentage of defaulters, Civil marriage has the highest percent around (10%) and widow has the lowest around 6% (exception being Unknown).



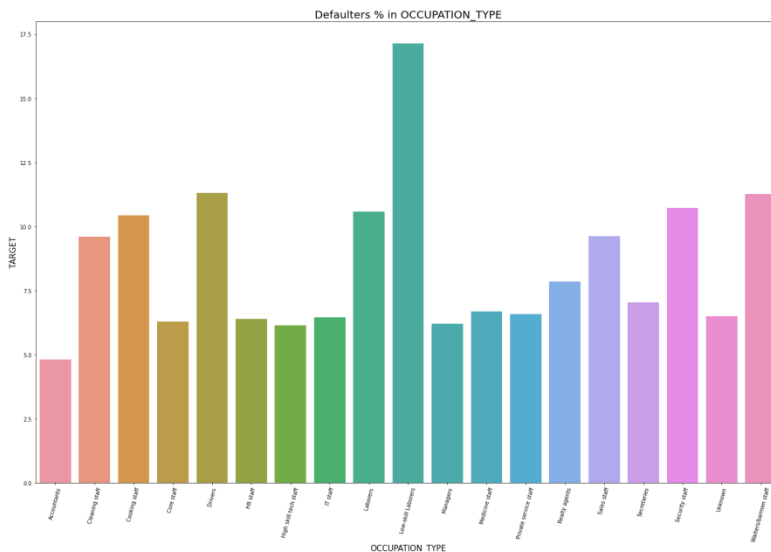
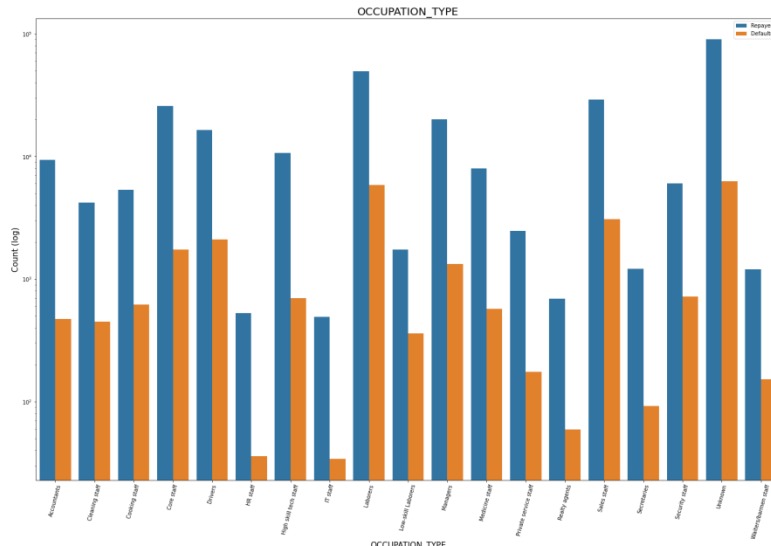
- Majority of clients have Secondary/secondary special education, followed by clients with Higher education.
 - Very few clients have an academic degree.
- Lower secondary category have highest rate of defaulting around 11%.
 - People with Academic degree are least likely to default.



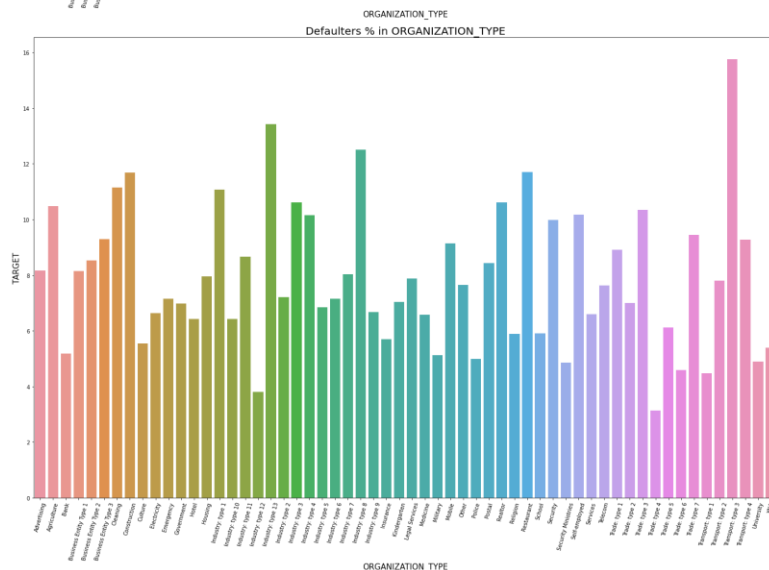
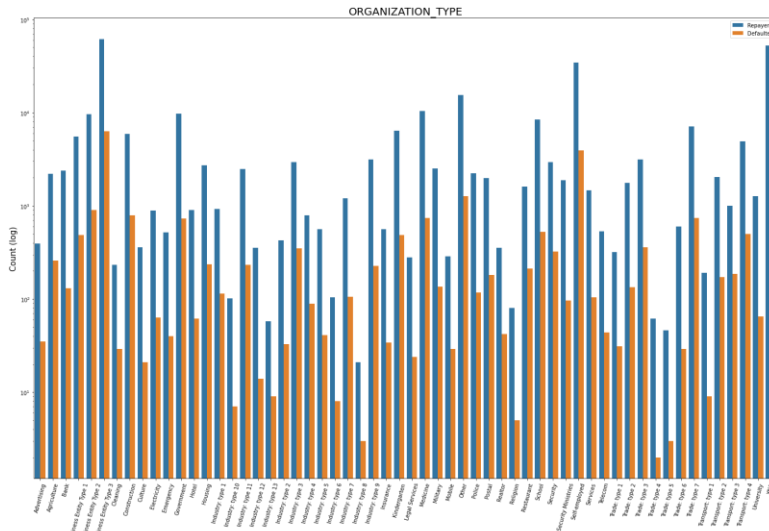
- Most of applicants for loans income type is Working, followed by Commercial associate, Pensioner and State servant.
- The applicants who are on Maternity leave have defaulting percentage of 40% which is the highest, followed by Unemployed (37%). The rest under average around 10% defaulters.
- Student and Businessmen though less in numbers, do not have default record. Safest two categories for providing loan.



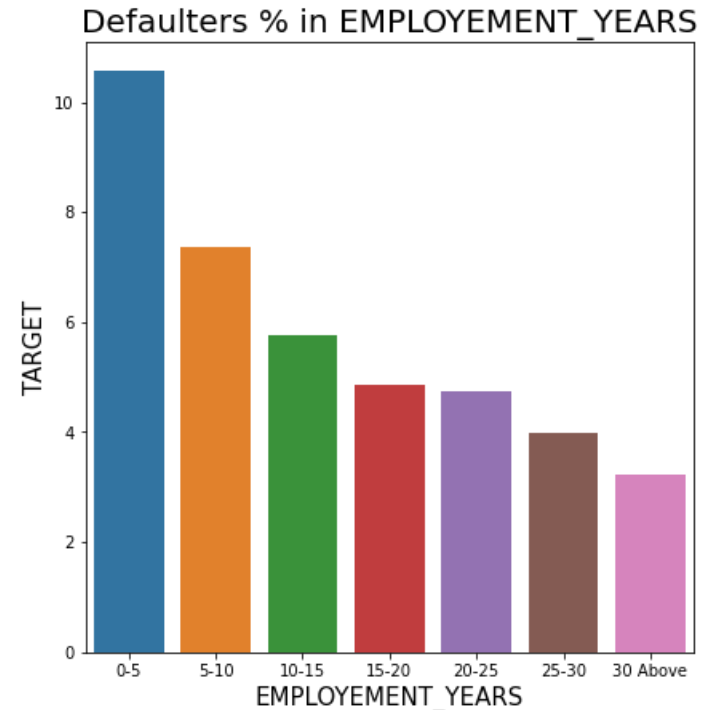
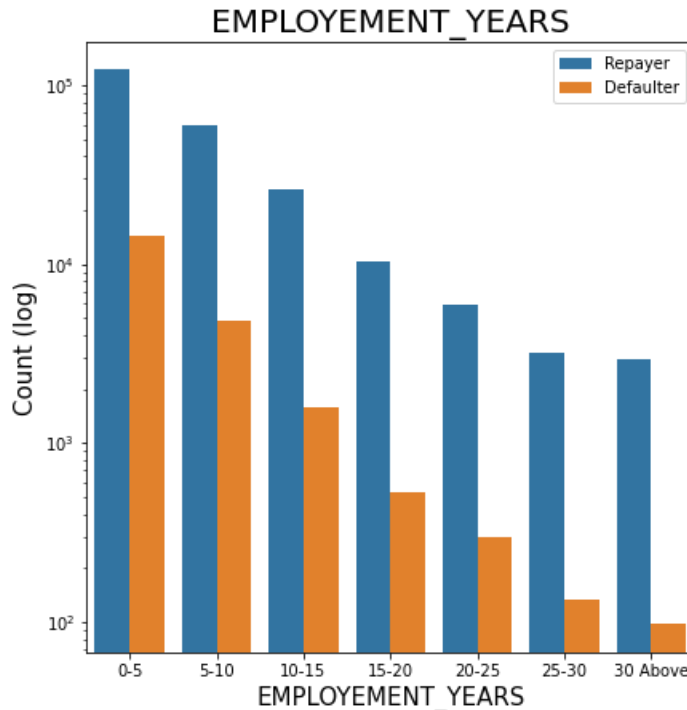
- Most of the applicants are living in Region with Rating 2 place.
 - Region Rating 3 has the highest default rate (11%).
- Applicant living in Region Rating 1 has the lowest probability of defaulting, thus safer for approving loans.



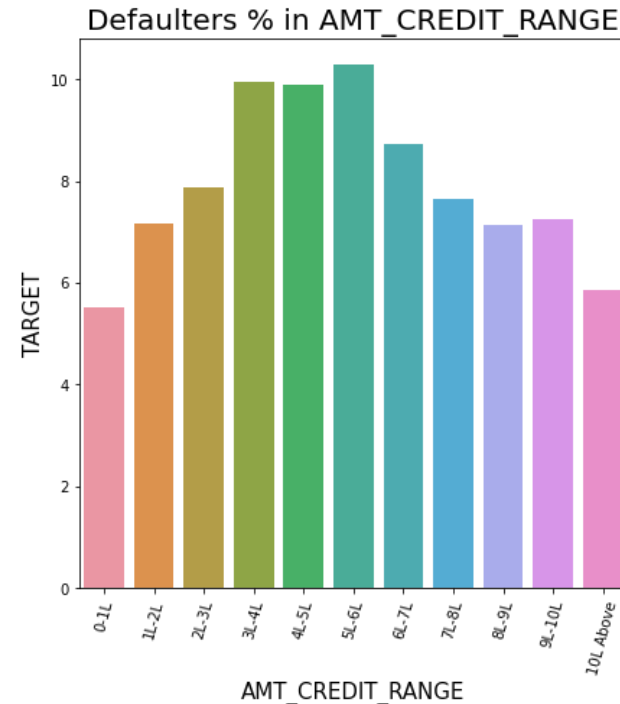
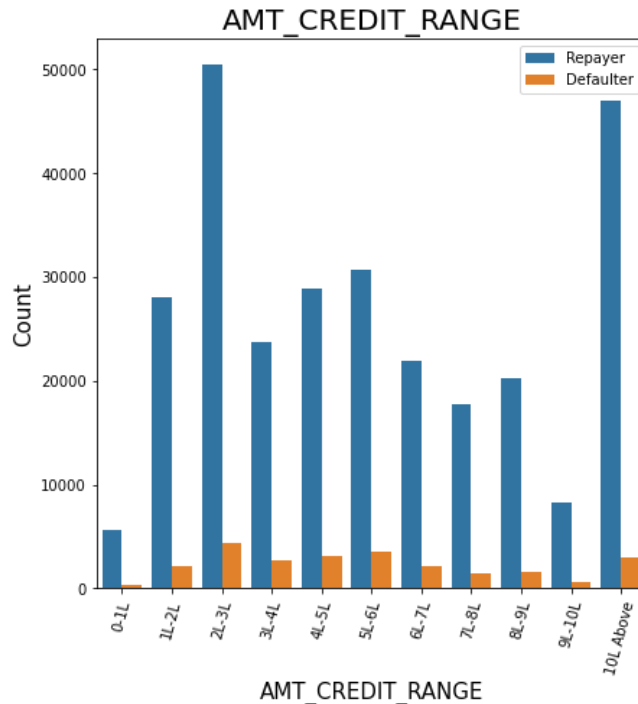
- Most of the loans are taken by Laborers, followed by Sales staff.
- IT staff are less likely to apply for Loan.
- Category with highest percent of defaulters are Low-skill Laborers (above 17%), followed by Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff.



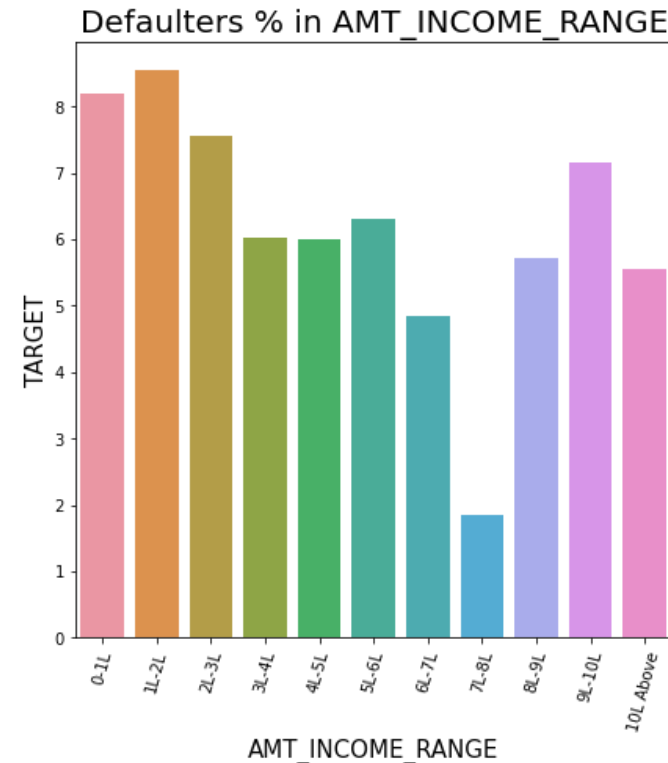
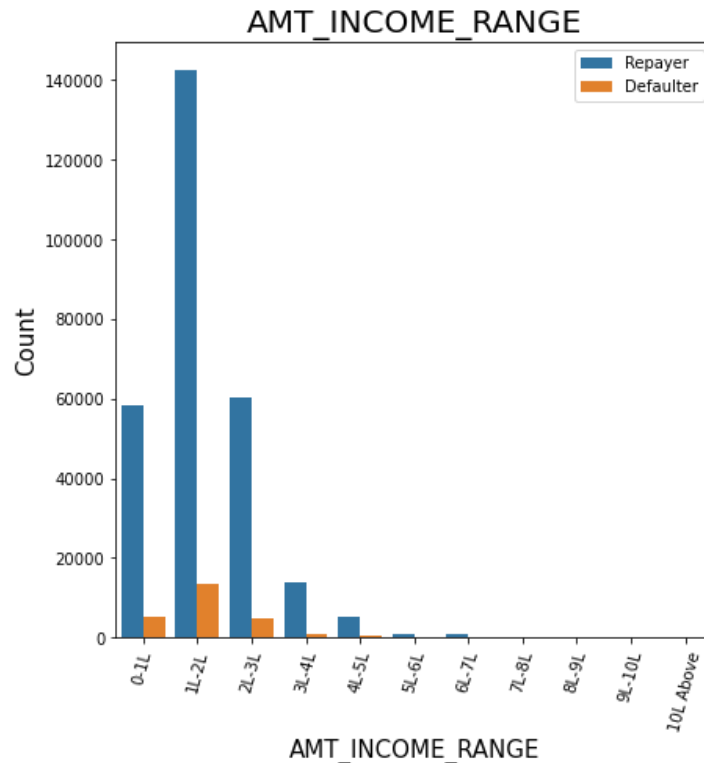
- Organizations with highest percent of defaulters are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%).
- Self employed people have relative high defaulting rate, to be safer side loan disbursement should be avoided or provide loan with higher interest rate to mitigate the risk of defaulting.
- Most of the people application for loan are from Business Entity Type 3.
- For a very high number of applications, Organization type information is unavailable(XNA).
- It can be seen that following category of organization type has lesser defaulters thus safer for providing loans: Trade Type 4 and 5, Industry Type 8.



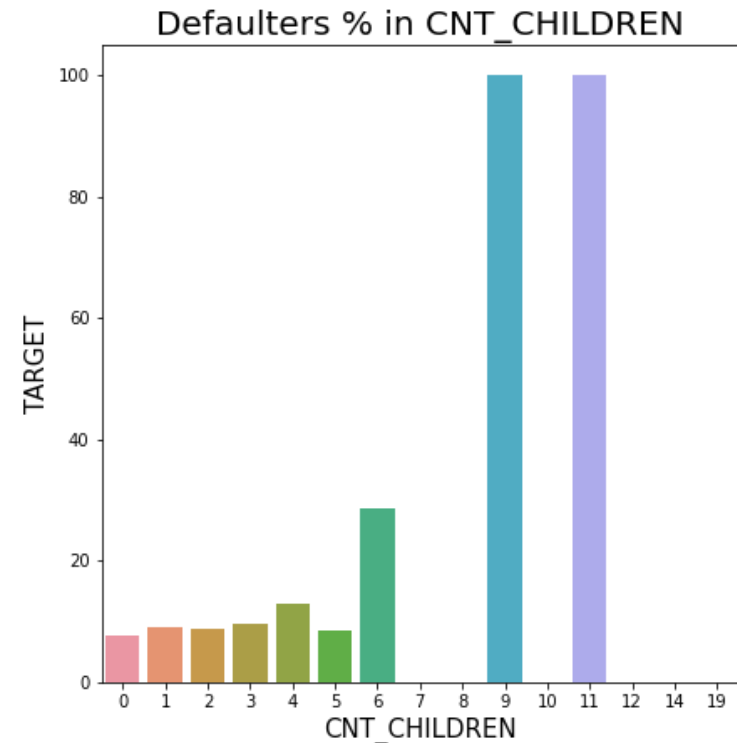
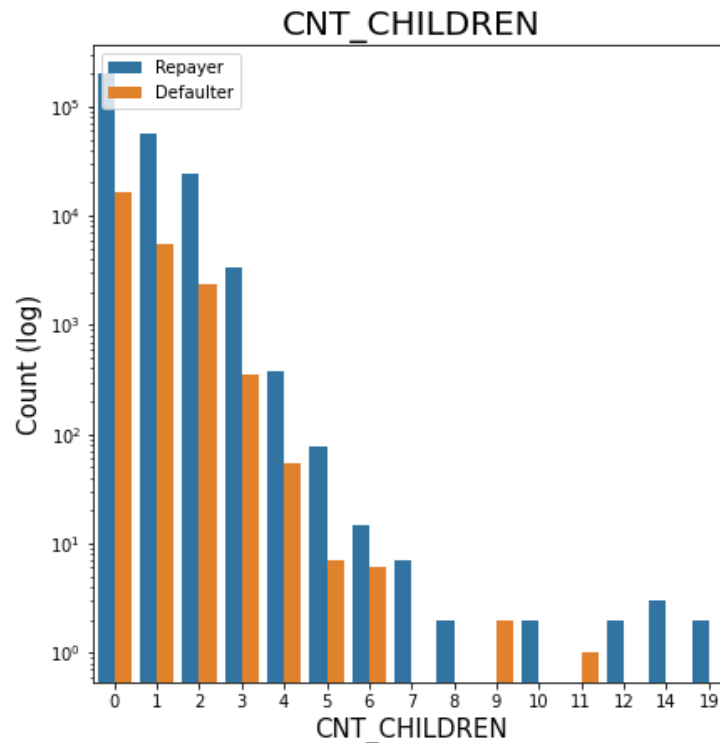
- Majority of the applicants having working experience between 0-5 years are defaulters. The defaulting rating of this group is also the highest which is around 10%.
- With increase of employment year, defaulting rate is gradually decreasing.
- With people having 40+ year experience have less than 1% default rate.



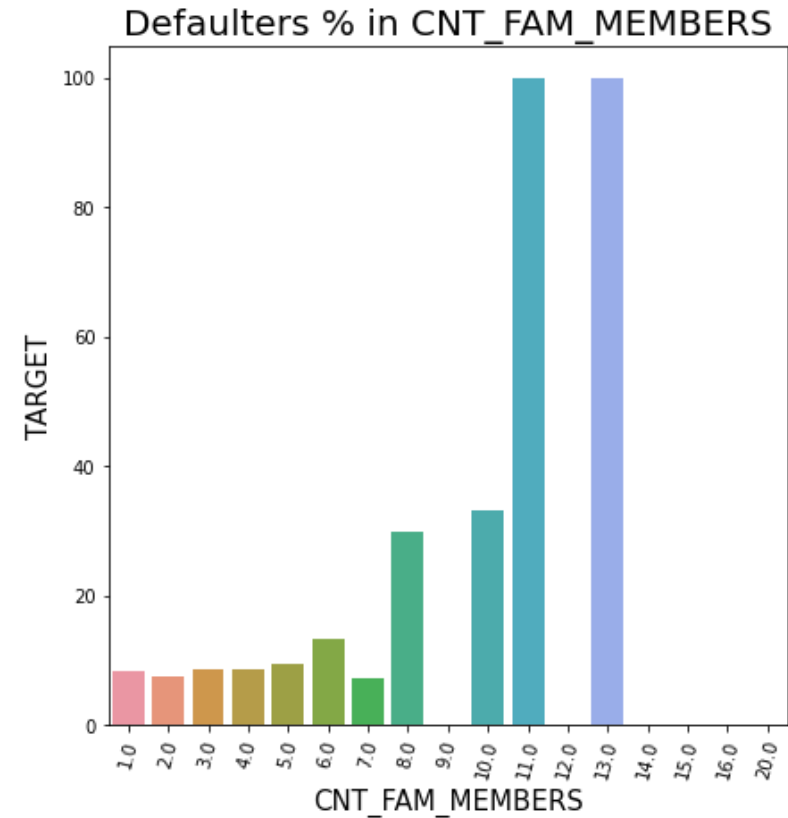
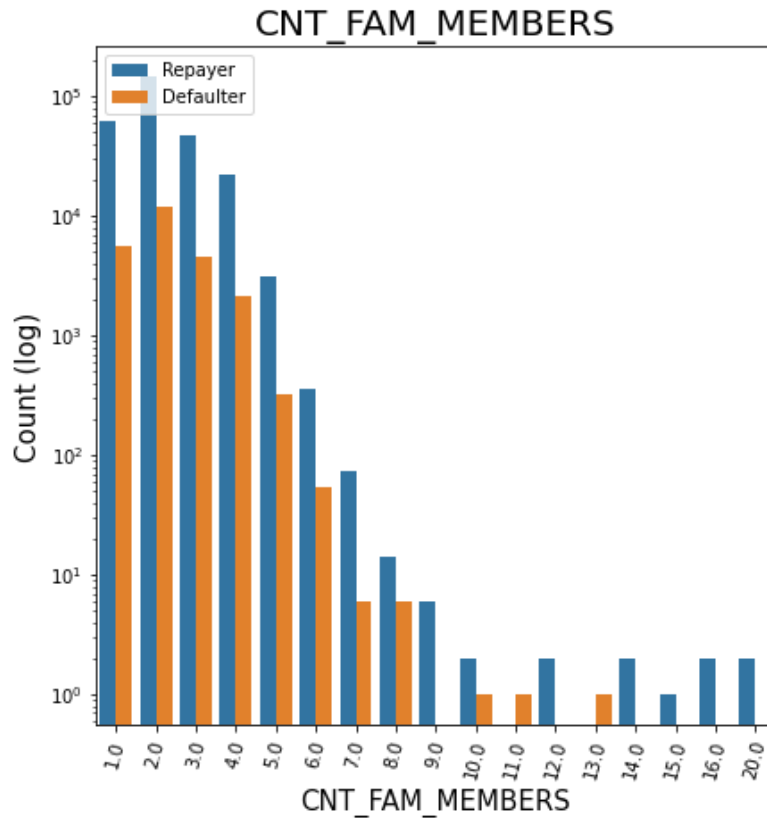
- There are high number of applicants have loan in range of 2-3 Lakhs followed by 10 Lakhs above range
- People who get loan for 3-6 Lakhs have most number of defaulters than other loan range.



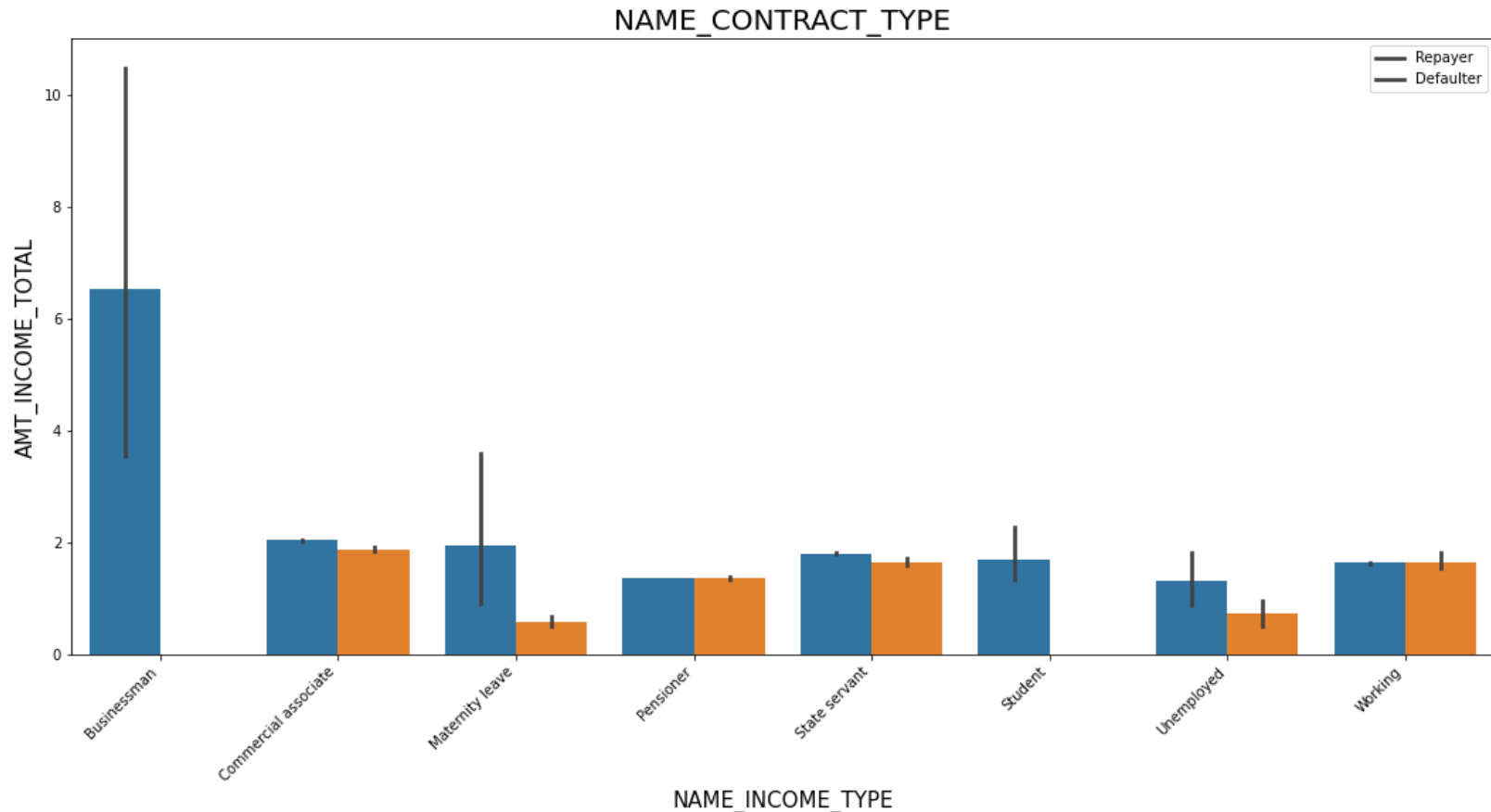
- Majority of the applications have Income total less than 3 Lakhs.
- Application with Income less than 3 Lakhs has high probability of defaulting.
- Applicant with Income 7-8 Lakhs are less likely to default.



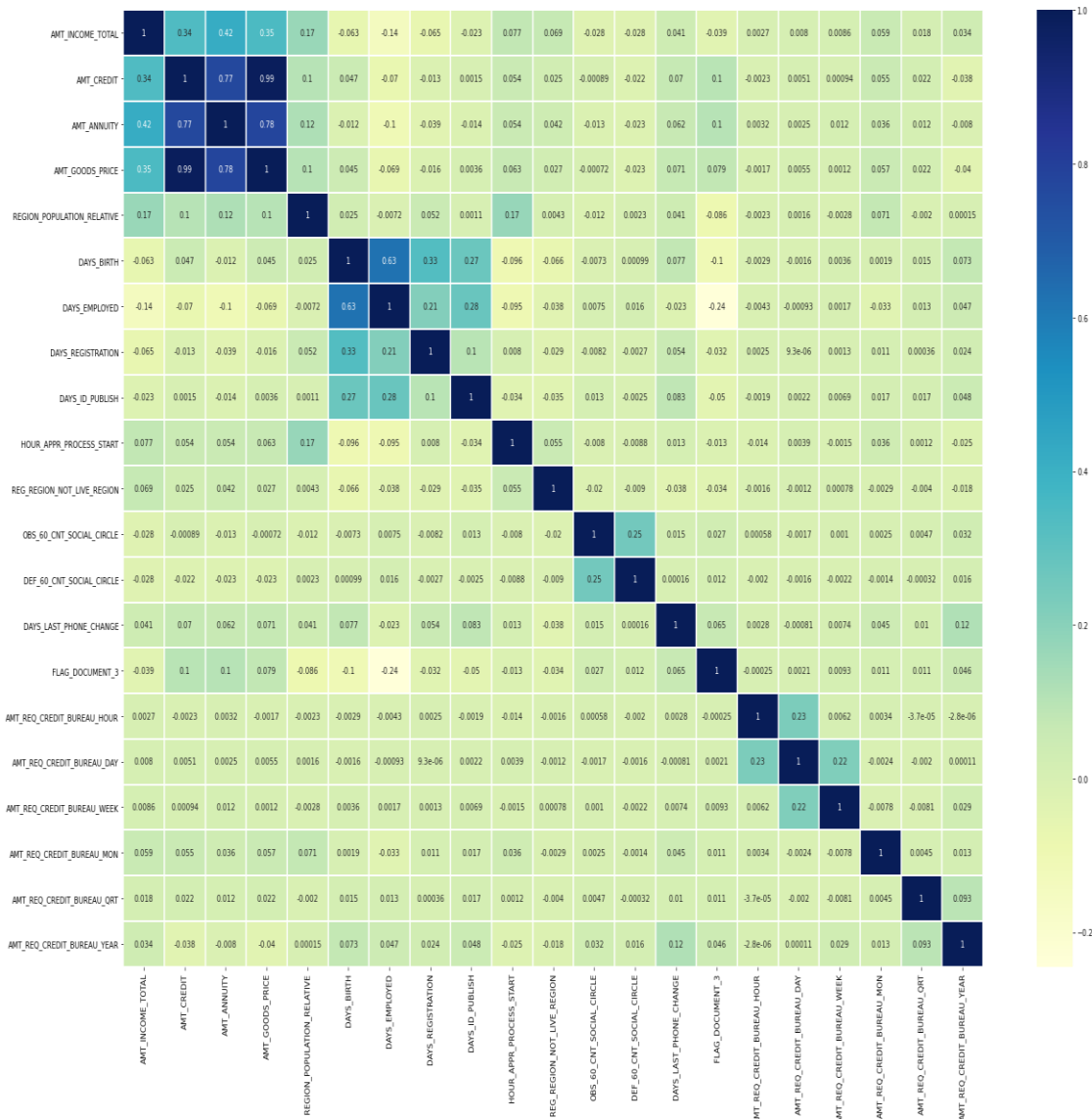
- Most of the applicants do not have children.
- Very few clients have more than 3 children.
- Client who have more than 4 children has a very high default rate with child count 9 and 11 showing 100% default rate.



- Family member follows the same trend as children where having more family members increases the risk of defaulting.

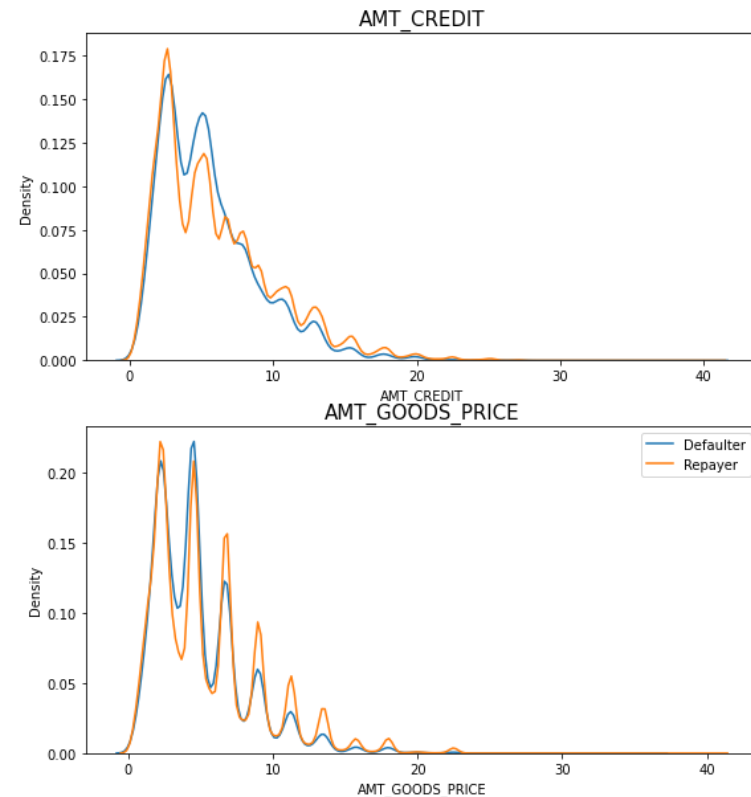
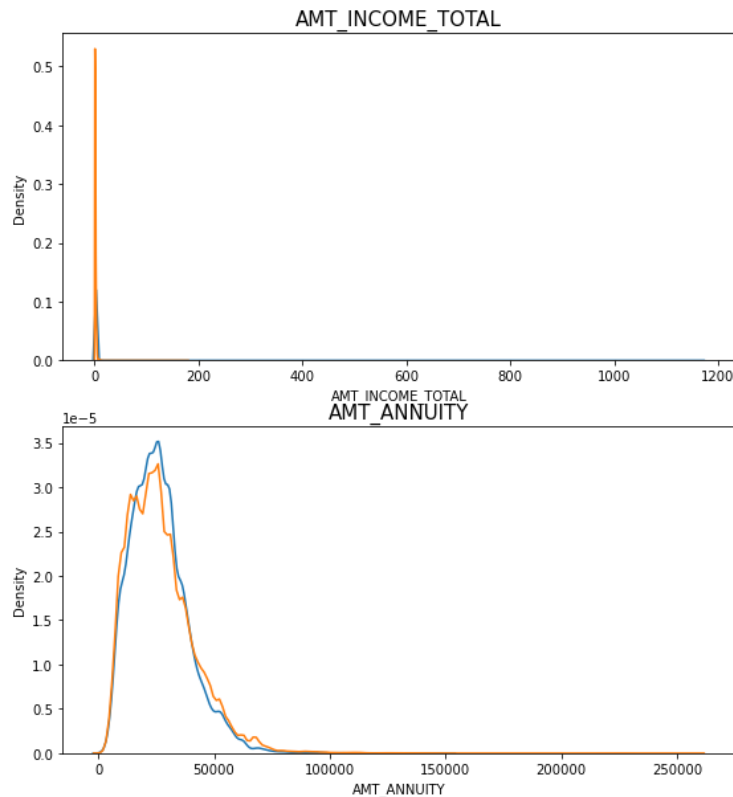


- It can be seen that Businessman income is the highest and the estimated range with default 95% confidence level seem to indicate that the income of a Businessman could be in the range of slightly close to 4 lakhs and slightly above 10 lakhs.

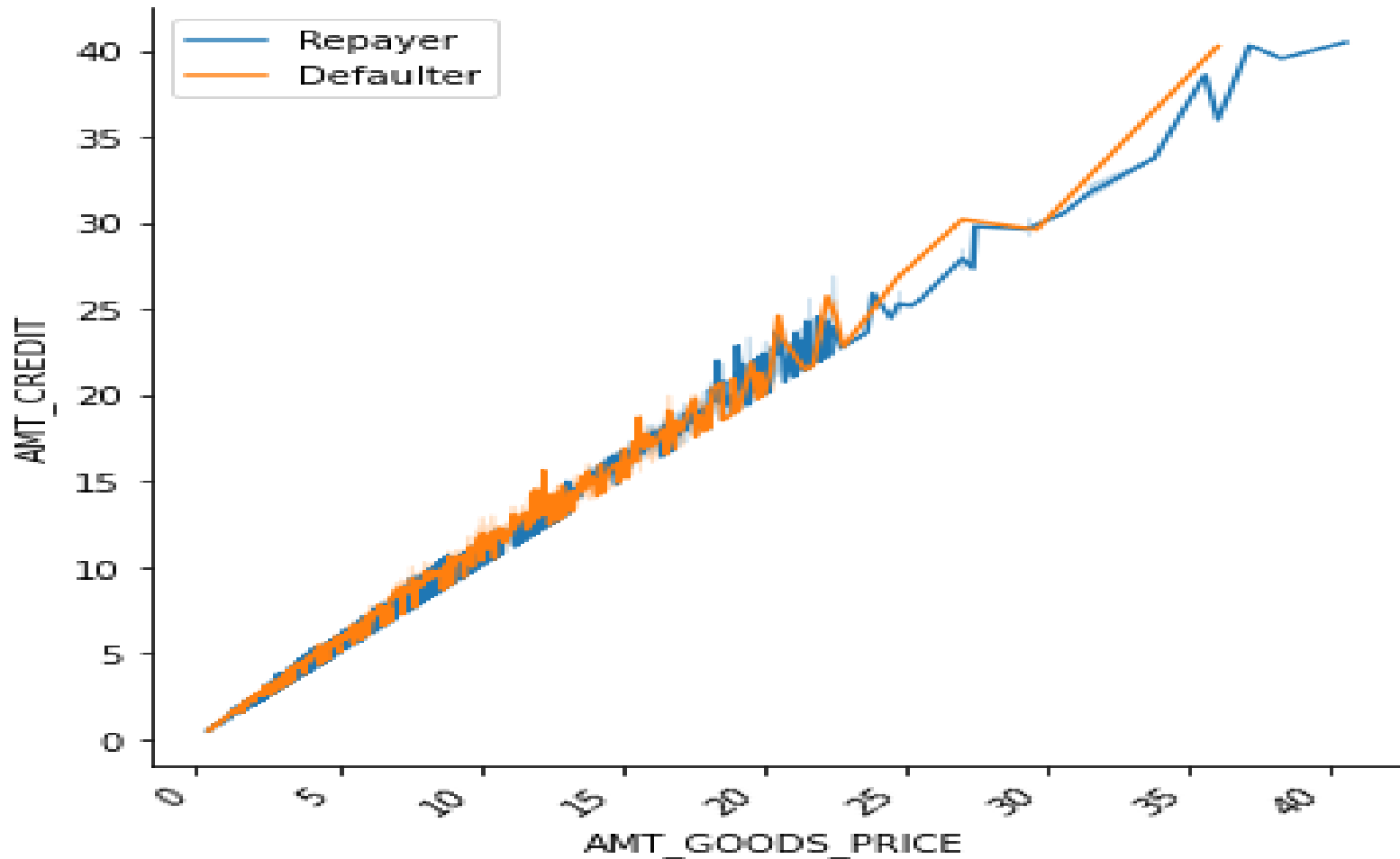


Correlating factors amongst repayers :-

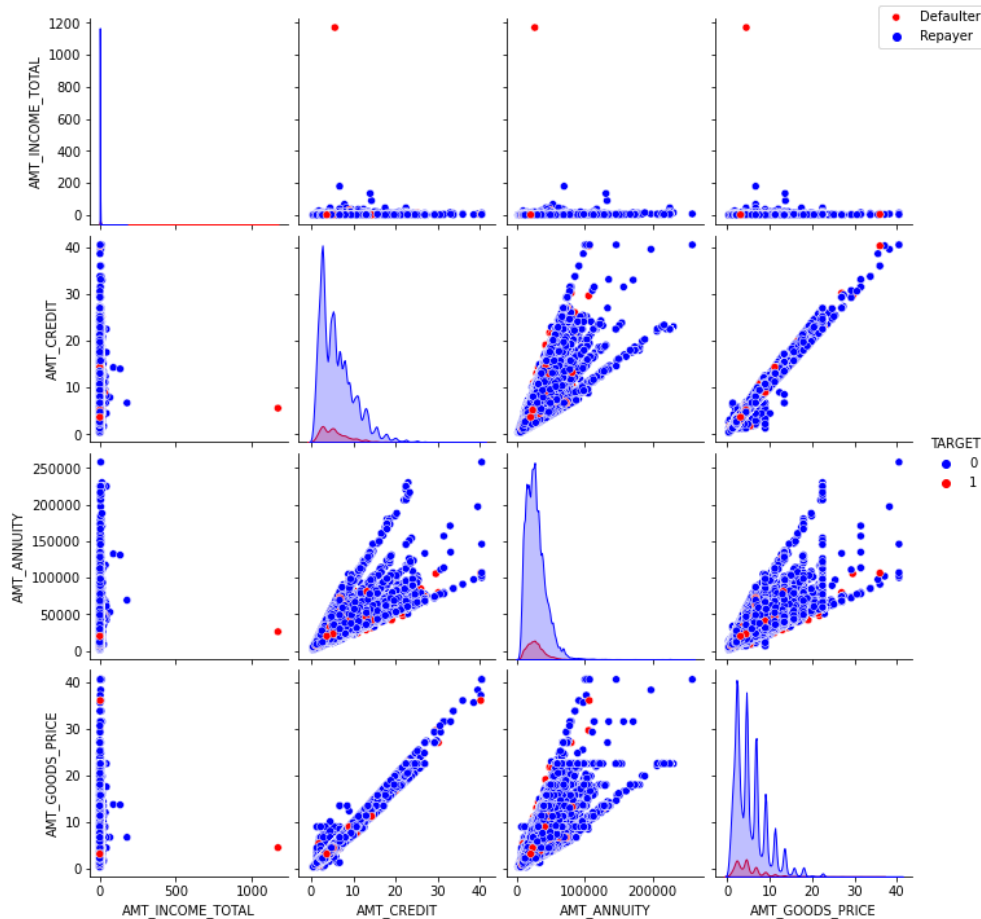
- Credit amount is highly correlated with good price amount which is same as repayers.
- Loan annuity correlation with credit amount has slightly reduced in defaulters(0.75) when compared to repayers(0.77).
- We can also see that repayers have high correlation in number of days employed(0.62) when compared to defaulters(0.58).
- There is a severe drop in the correlation between total income of the client and the credit amount(0.038) amongst defaulters whereas it is 0.342 among repayers.
- Days_birth and number of children correlation has reduced to 0.259 in defaulters when compared to 0.337 in repayers.
- There is a slight increase in defaulted to observed count in social circle among defaulters(0.264) when compared to repayers(0.254).



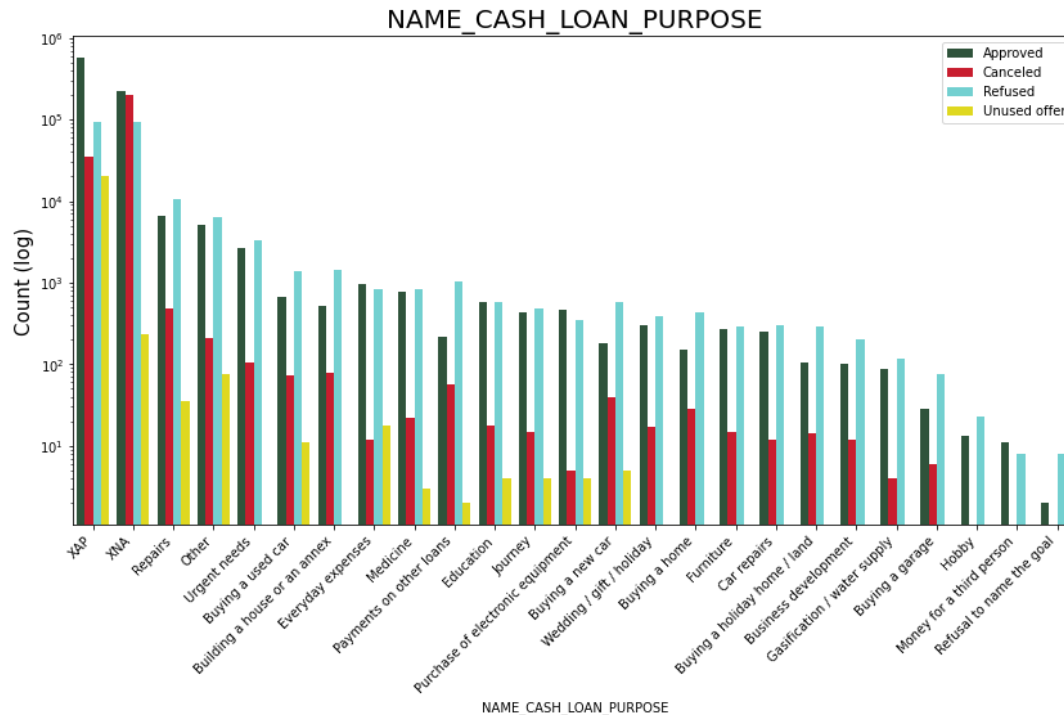
- Most no of loans are given for goods price below 10 lakhs.
- Most people pay annuity below 50K for the credit loan.
- Credit amount of the loan is mostly less then 10 lakhs.
- The repayers and defaulters distribution overlap in all the plots and hence we cannot use any of these variables in isolation to make a decision.



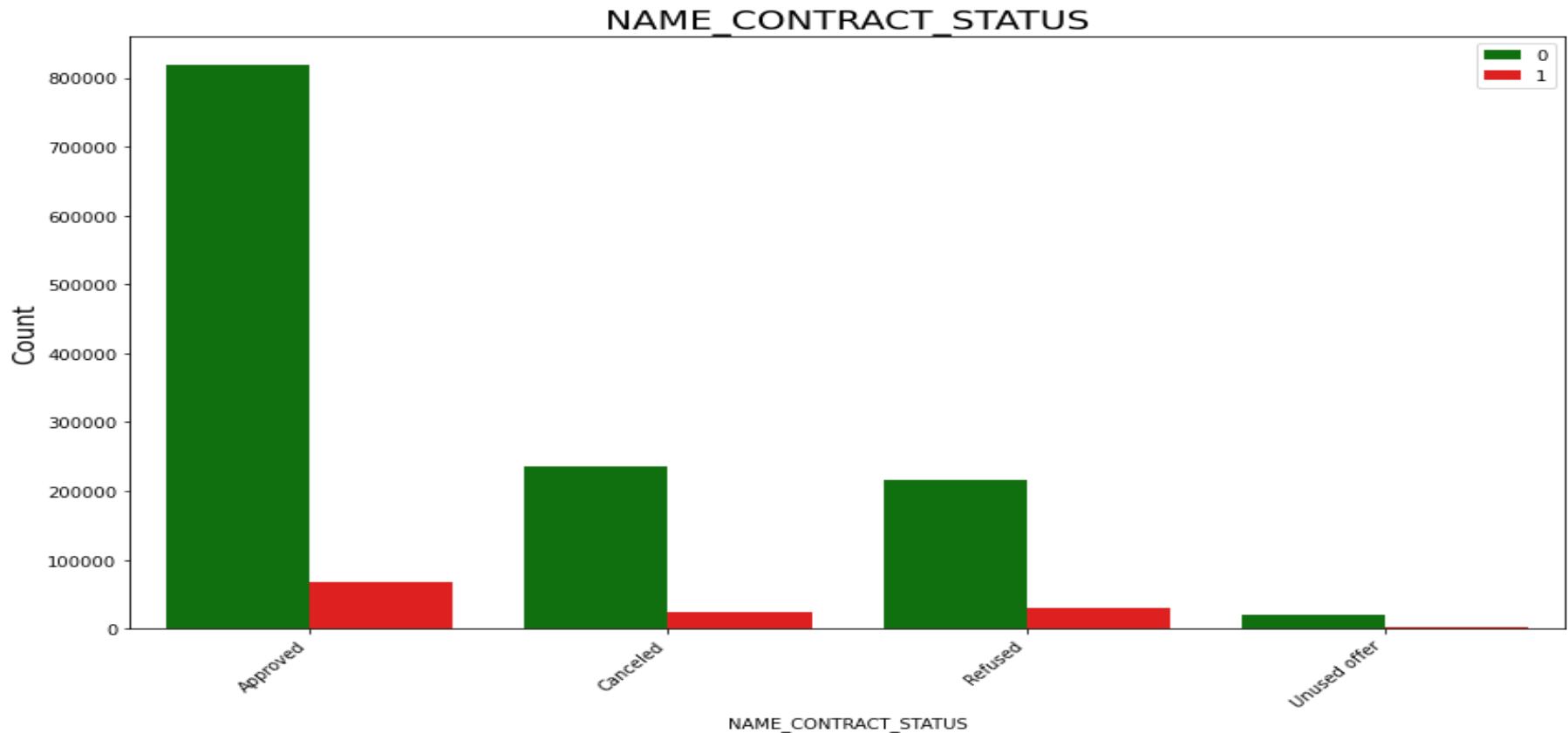
- When the credit amount goes beyond 30 Lakhs, there is an increase in defaulters.



- When Annuity Amount > 15K and Good Price Amount > 20 Lakhs, there is a lesser chance of defaulters.
- Loan Amount(AMT_CREDIT) and Goods price(AMT_GOODS_PRICE) are highly correlated as based on the scatterplot where most of the data are consolidated in form of a line.
- There are very less defaulters for AMT_CREDIT > 20 Lakhs.

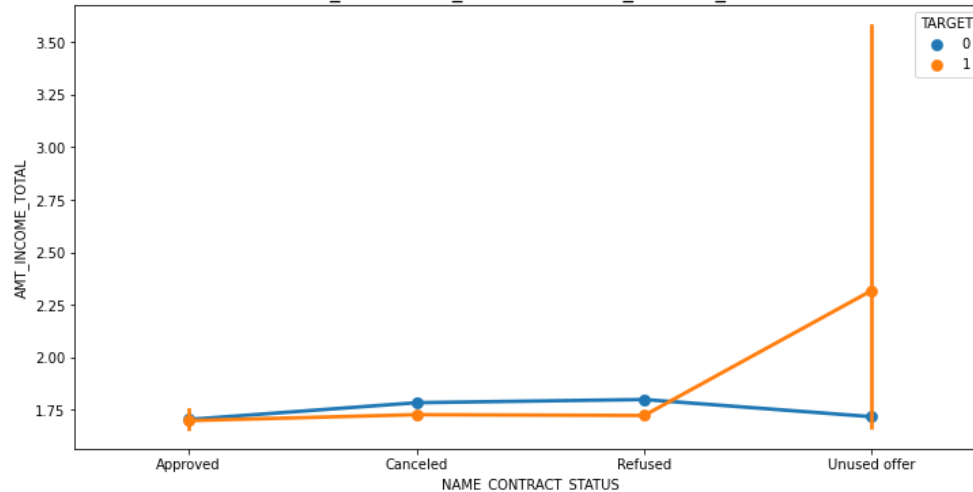


- Loan purpose has high number of unknown values (XAP, XNA).
- Loan taken for the purpose of Repairs looks to have highest default rate.
- Huge number application have been rejected by bank or refused by client which are applied for Repair or Other. from this we can infer that repair is considered high risk by bank. Also, either they are rejected or bank offers loan on high interest rate which is not feasible by the clients and they refuse the loan.



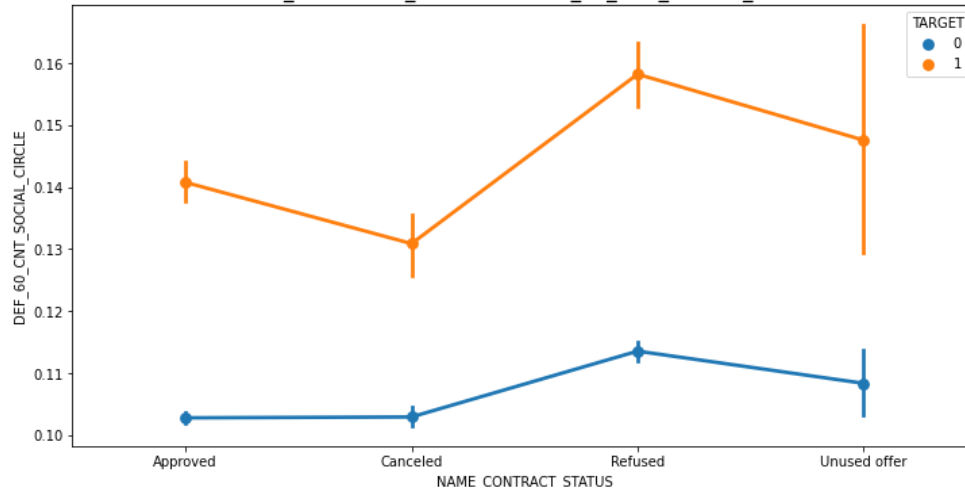
- 90% of the previously cancelled client have actually repayed the loan. Revising the interest rates would increase business opportunity for these clients.
 - 88% of the clients who have been previously refused a loan has paid back the loan in current case.
- Refusal reason should be recorded for further analysis as these clients could turn into potential repaying customer.

NAME_CONTRACT_STATUS VS AMT_INCOME_TOTAL



- The point plot show that the people who have not used offer earlier have defaulted even when there average income is higher than others.

NAME_CONTRACT_STATUS VS DEF_60_CNT_SOCIAL_CIRCLE



- Clients who have average of 0.13 or higher their DEF_60_CNT_SOCIAL_CIRCLE score tend to default more and thus analysing client's social circle could help in disbursement of the loan.

A. Decisive Factor whether an applicant will be Repayer:

- NAME_EDUCATION_TYPE:** Academic degree has less defaults.
- NAME_INCOME_TYPE:** Student and Businessmen have no defaults.
- REGION_RATING_CLIENT:** RATING 1 is safer.
- ORGANIZATION_TYPE:** Clients with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%.
- DAYS_BIRTH:** People above age of 50 have low probability of defaulting.
- DAYS_EMPLOYED:** Clients with 40+ year experience having less than 1% default rate.
- AMT_INCOME_TOTAL:** Applicant with Income more than 700,000 are less likely to default.
- NAME_CASH_LOAN_PURPOSE:** Loans bought for Hobby, Buying garage are being repayed mostly.
- CNT_CHILDREN:** People with zero to two children tend to repay the loans.

B. Decisive Factor whether an applicant will be Defaulter:

- CODE_GENDER:** Men are at relatively higher default rate.
- NAME_FAMILY_STATUS :** People who have civil marriage or who are single default a lot.
- NAME_EDUCATION_TYPE:** People with Lower Secondary & Secondary education.
- NAME_INCOME_TYPE:** Clients who are either at Maternity leave OR Unemployed default a lot.
- REGION_RATING_CLIENT:** People who live in Rating 3 has highest defaults.
- OCCUPATION_TYPE:** Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as their default rate is huge.
- ORGANIZATION_TYPE:** Organizations with highest percent of loans not repaid are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%). Self-employed people have relative high defaulting rate, and thus should be avoided to be approved for loan or provide loan with higher interest rate to mitigate the risk of defaulting.
- DAYS_BIRTH:** Avoid young people who are in age group of 20-40 as they have higher probability of defaulting
- DAYS_EMPLOYED:** People who have less than 5 years of employment have high default rate.
- CNT_CHILDREN & CNT_FAM_MEMBERS:** Client who have children equal to or more than 9 default 100% and hence their applications are to be rejected.
- AMT_GOODS_PRICE:** When the credit amount goes beyond 3lakhs, there is an increase in defaulters.

C. Factors that Loan can be given on Condition of High Interest rate to mitigate any default risk leading to business loss:¶

- NAME_HOUSING_TYPE:** High number of loan applications are from the category of people who live in Rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default.
- AMT_CREDIT:** People who get loan for 3-6 Lakhs tend to default more than others and hence having higher interest specifically for this credit range would be ideal.
- AMT_INCOME:** Since 90% of the applications have Income total less than 3Lakhs and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.
- CNT_CHILDREN & CNT_FAM_MEMBERS:** Clients who have 4 to 8 children has a very high default rate and hence higher interest should be imposed on their loans.
- NAME_CASH_LOAN_PURPOSE:** Loan taken for the purpose of Repairs seems to have highest default rate. A very high number applications have been rejected by bank or refused by client in previous applications as well which has purpose as repair or other. This shows that purpose repair is taken as high risk by bank and either they are rejected, or bank offers very high loan interest rate which is not feasible by the clients, thus they refuse the loan. The same approach could be followed in future as well.

OTHER SUGGESTIONS

- **90% of the previously cancelled client have actually repayed the loan. Record the reason for cancellation which might help the bank to determine and negotiate terms with these repaying customers in future for increase business opportunity.**
- **88% of the clients who were refused by bank for loan earlier have now turned into a repaying client. Hence documenting the reason for rejection could mitigate the business loss and these clients could be contacted for further loans.**



International
Institute of Information
Technology Bangalore

upGrad

THANK YOU