CREDIT EDA

CASE STUDY

BY SHRISTY DADRIWAL

PROBLEM STATEMENT

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. Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers. You have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected

BUSINESS OBJECTIVES

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. Identification of such applicants using EDA is the aim of this case study

APPROACH

There were two data sets given with the problem statemen and EDA is done in below steps:

1. Data Cleaning : Reading the data and analyzing it by identifying the missing values, dropping columns with more than 50% of missing data and imputing the missing values with the mean ,median etc. values by analyzing the data.

2. Data types: checking for the Data types of the columns and analyzing if the changing of data type is required.

3. Binning: Analyzing the categorical data and binning the data into suitable interval based on the analysis. This also includes creating additional field if required for better understanding of Data.

4. Outliers: Identifying the Outliers if any and using suitable method to handle them byplotting.

APPROACH

5. Data visualization: This includes Categorical Univariate Analysis and Univariate Categorical Ordered Analysis by creating the reusable function and plotting the Graph. We split the data into two frames one with Target =0(Non –Defaulter) and other with Target =1(Defaulter).

6. Correlation Matrix: Identifying the columns for creating the correlation matrix and plotting the heat map for checking the linear correlation between different variables. This is done for two groups one for Defaulter(Target =1) and for Non defaulter(Target =1)

7. Merging Dataframes: The two data frames are then merged based on the Key column and the details are checked using the pandas Library and performing univariate categorical Analysis on the merged dataframes

CONCLUSION

Dividing the Conclusion into three Categories:

1. Factors that decides whether or not Applicant will pay the loan
2. Factor that decide whether or not an applicant will be Defaulter
3. Suggestions

1. Factors that decides whether or not Applicant will pay the loan

NAME_EDUCATION_TYPE: The one with Academic Degrees are less likely to be defaulters

NAME_INCOME_TYPE: There are no defaulters in case of Students and Businessman

ORGANIZATION_TYPE: Less than 3% of Clients with Trade type 4 and 5 and Industry type 8 have defaulted.

DAYS_BIRTH/AGE_GROUP: prople of age group 50+ are less likely to default.

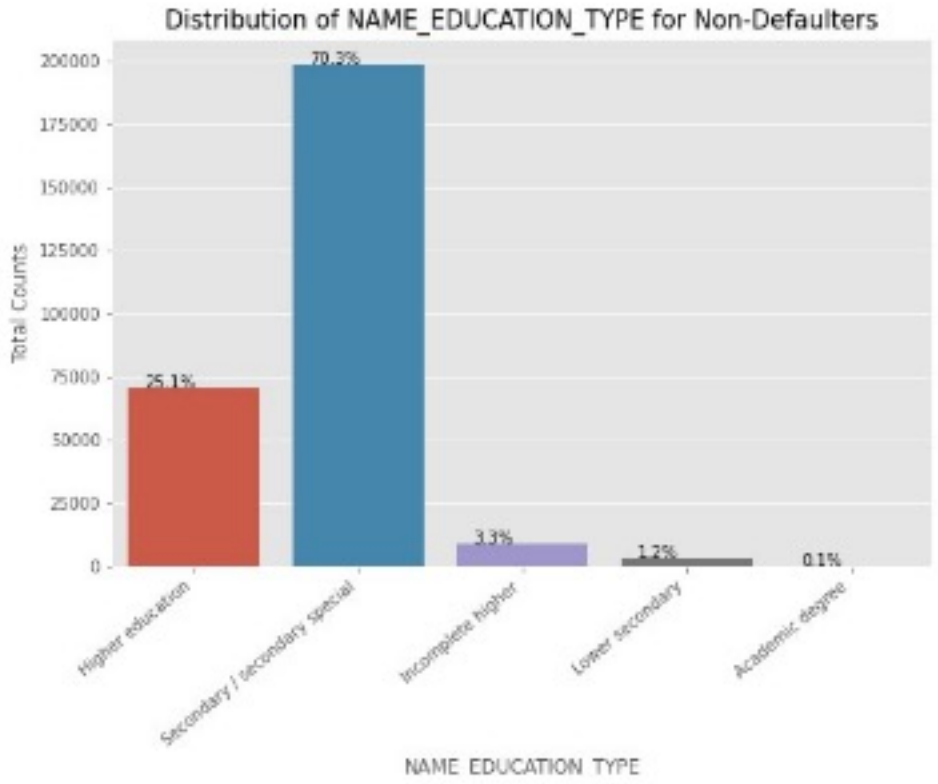
DAYS_EMPLOYED: 40+ year experience Clients are having less than 1% Default rate

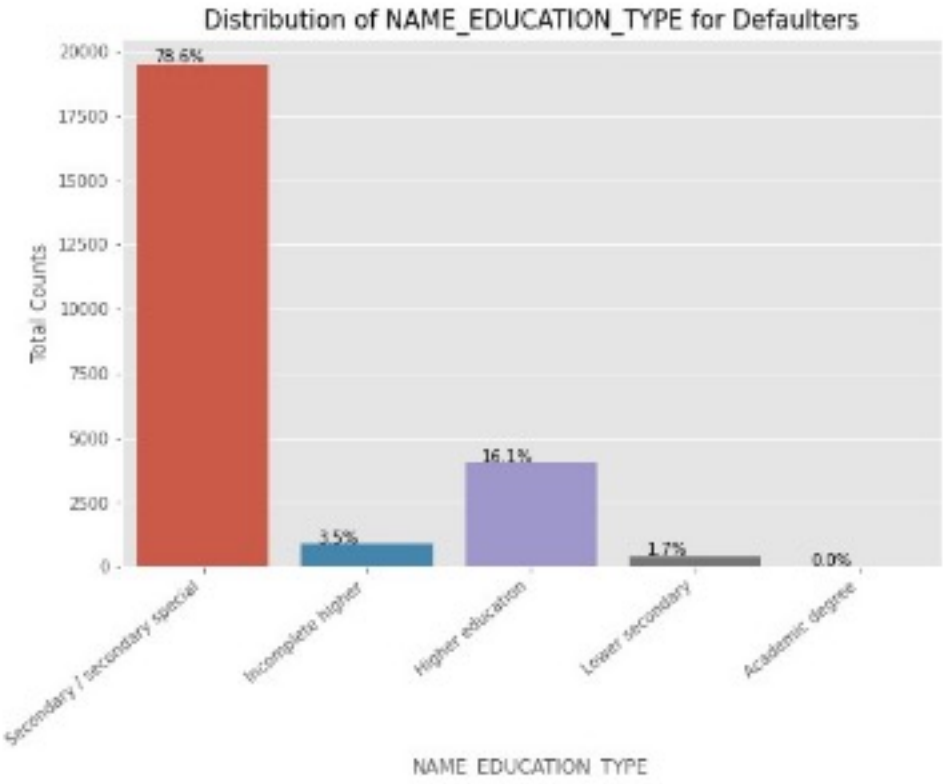
AMT_INCOME_TOTAL: Income more than 700,000 are less likely to default

NAME_CONTRACT_STATUS:people who were approved for a loan earlier, defaulted less often

NAME_EDUCATION_TYPE :

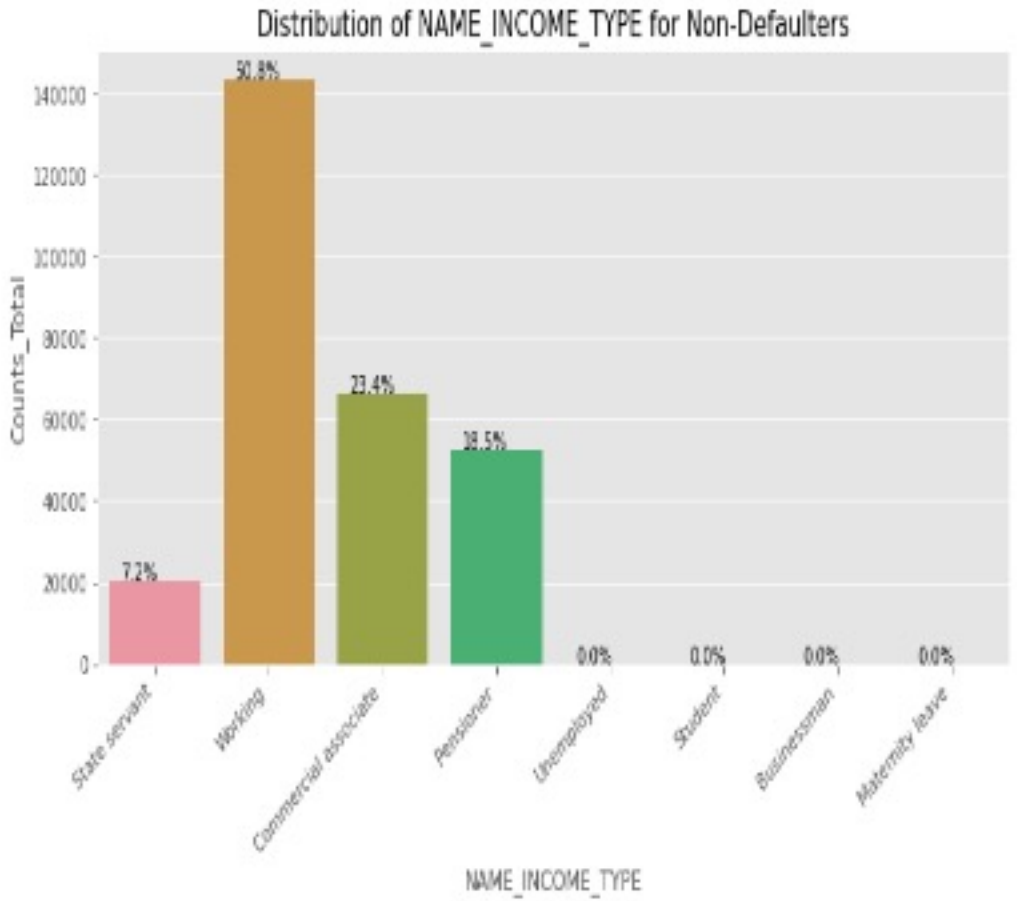
The one with Academic Degrees are less likely to be defaulters

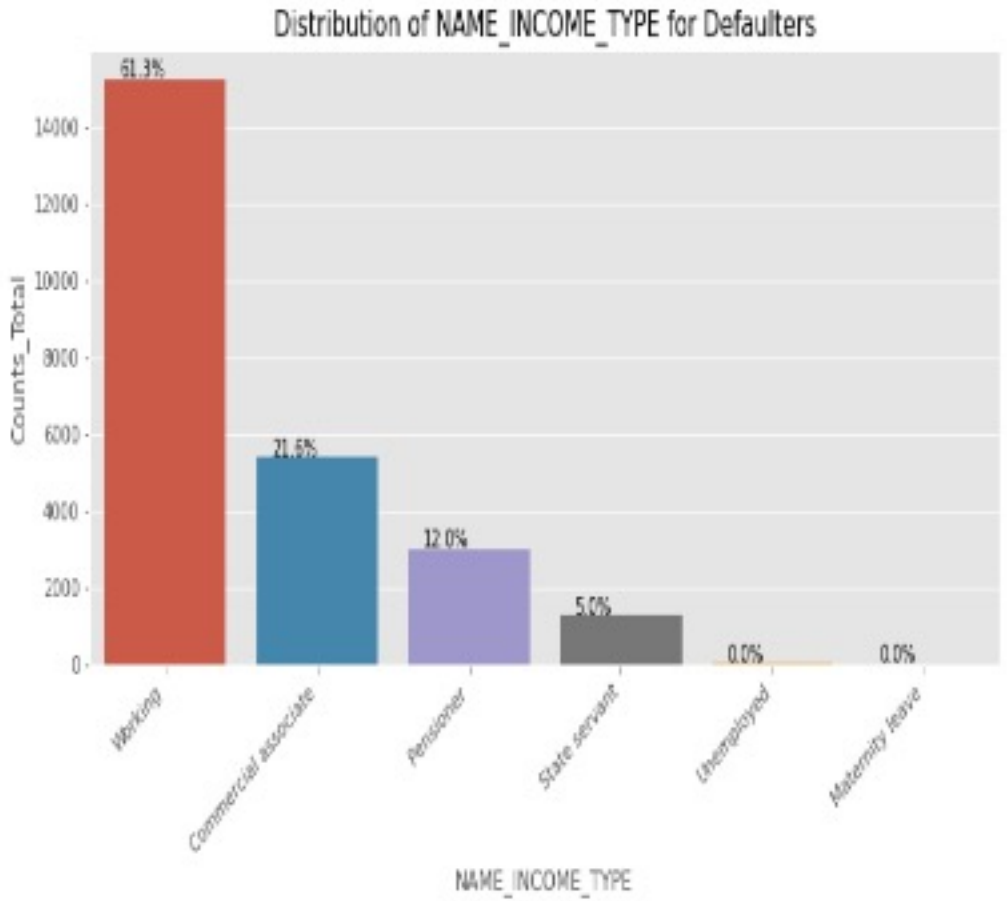




NAME_INCOME_TYPE:

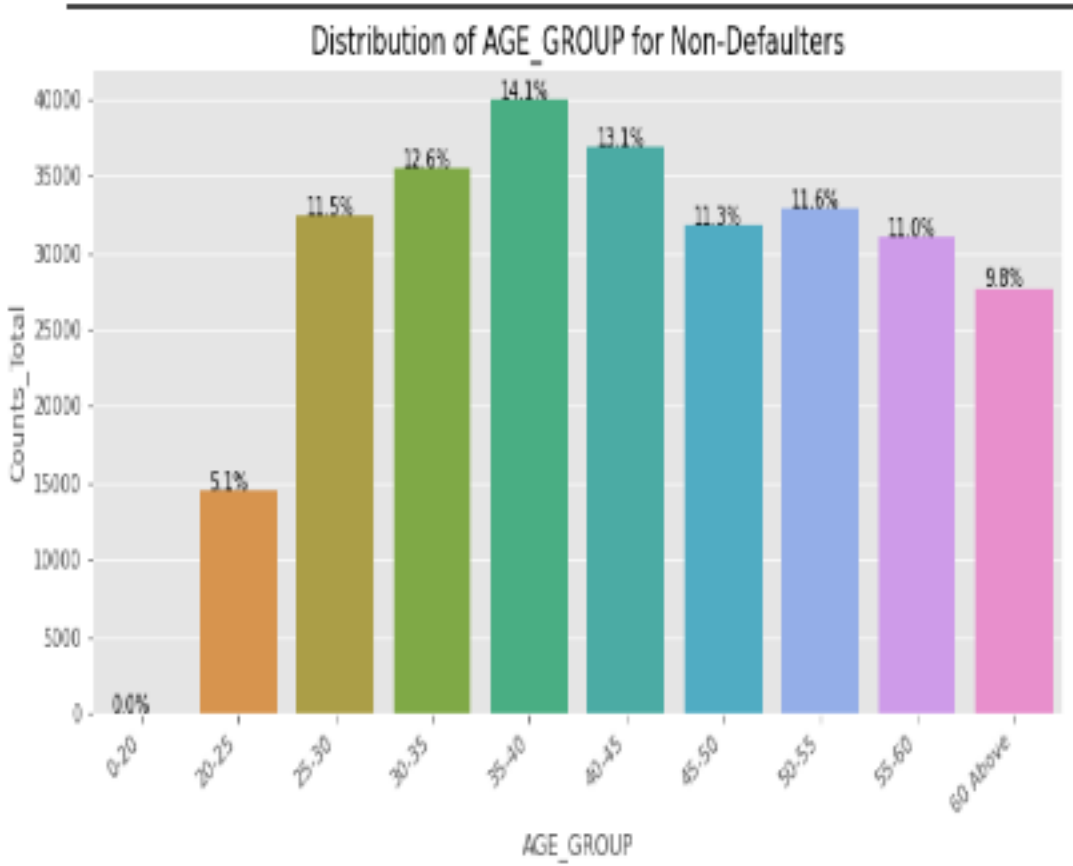
There are no defaulters in case of Students and Businessman

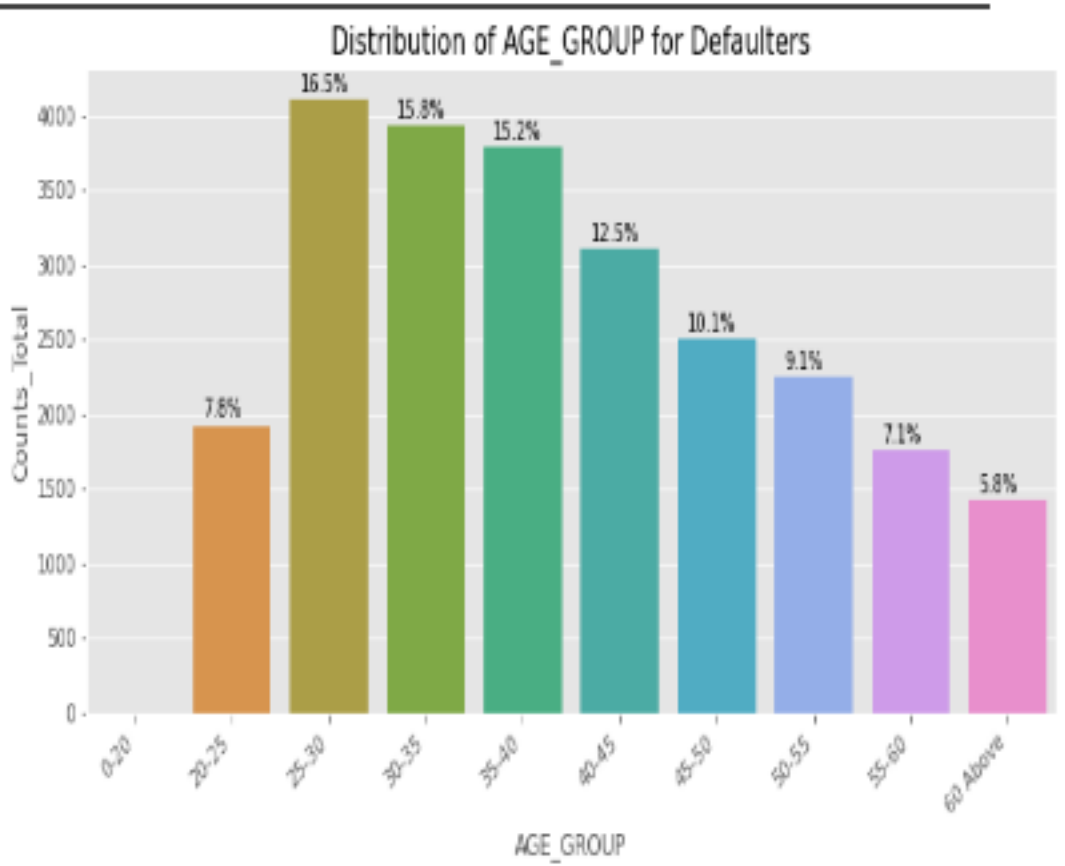




AGE_GROUP:

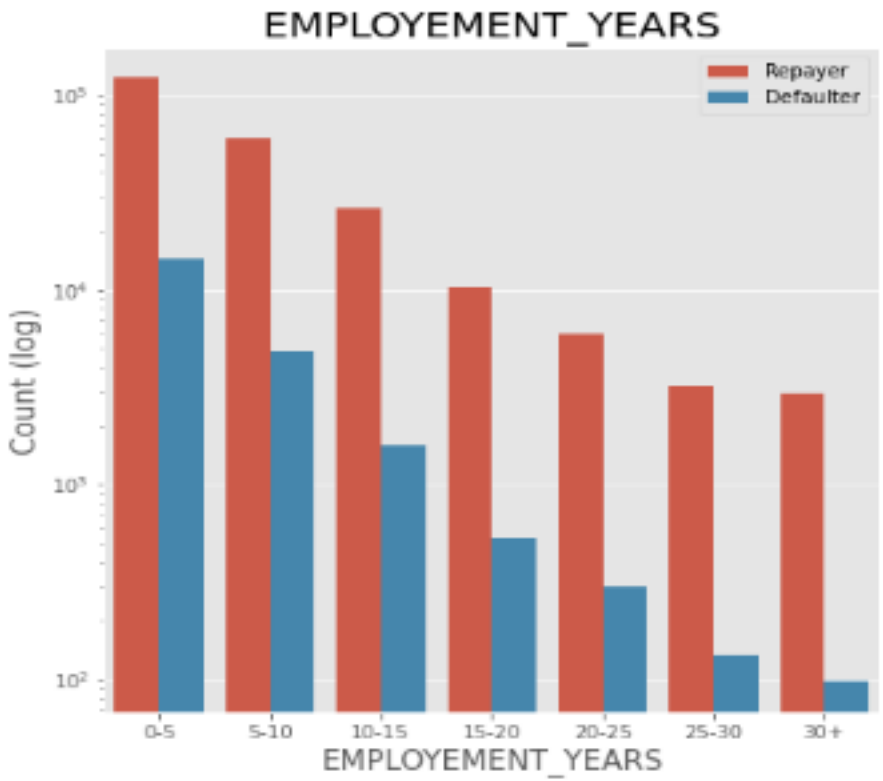
People with higher age group 50+ are less likely to default

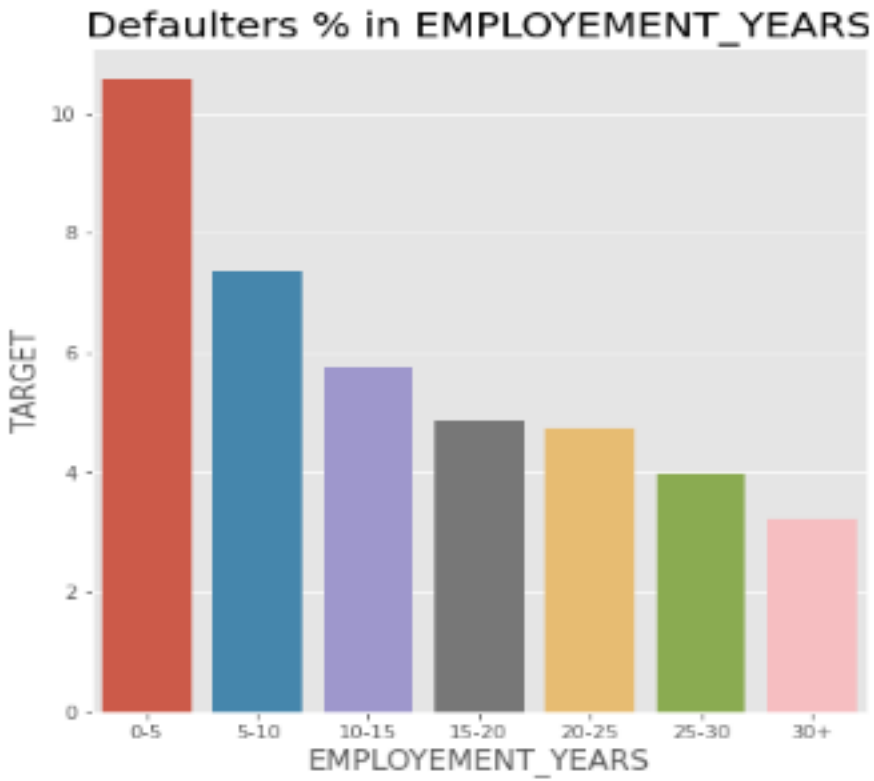




DAYS_EMPLOYED:

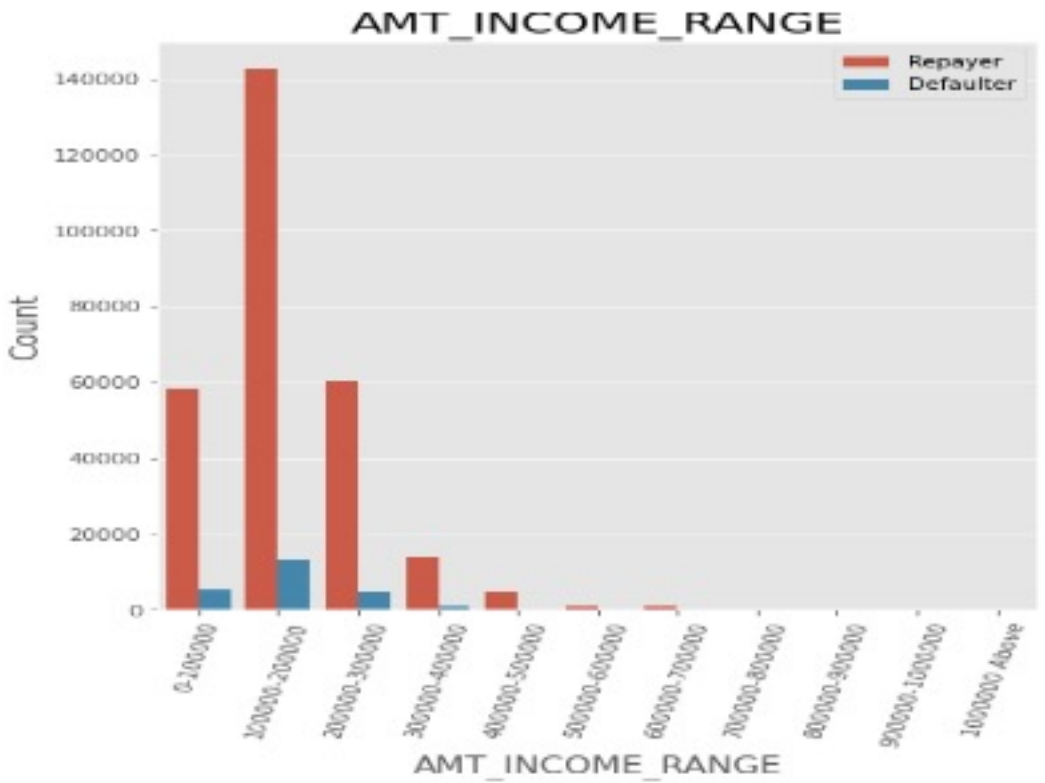
There are less defaulters in case of employees having more experience(40+)

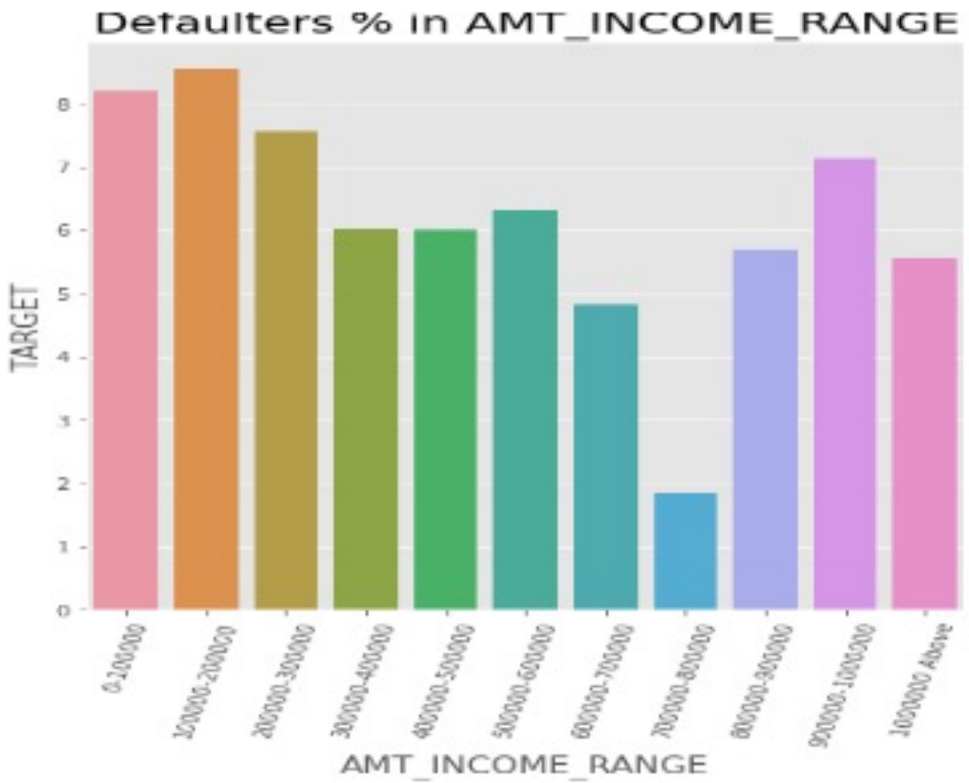




AMT_INCOME_TOTAL:

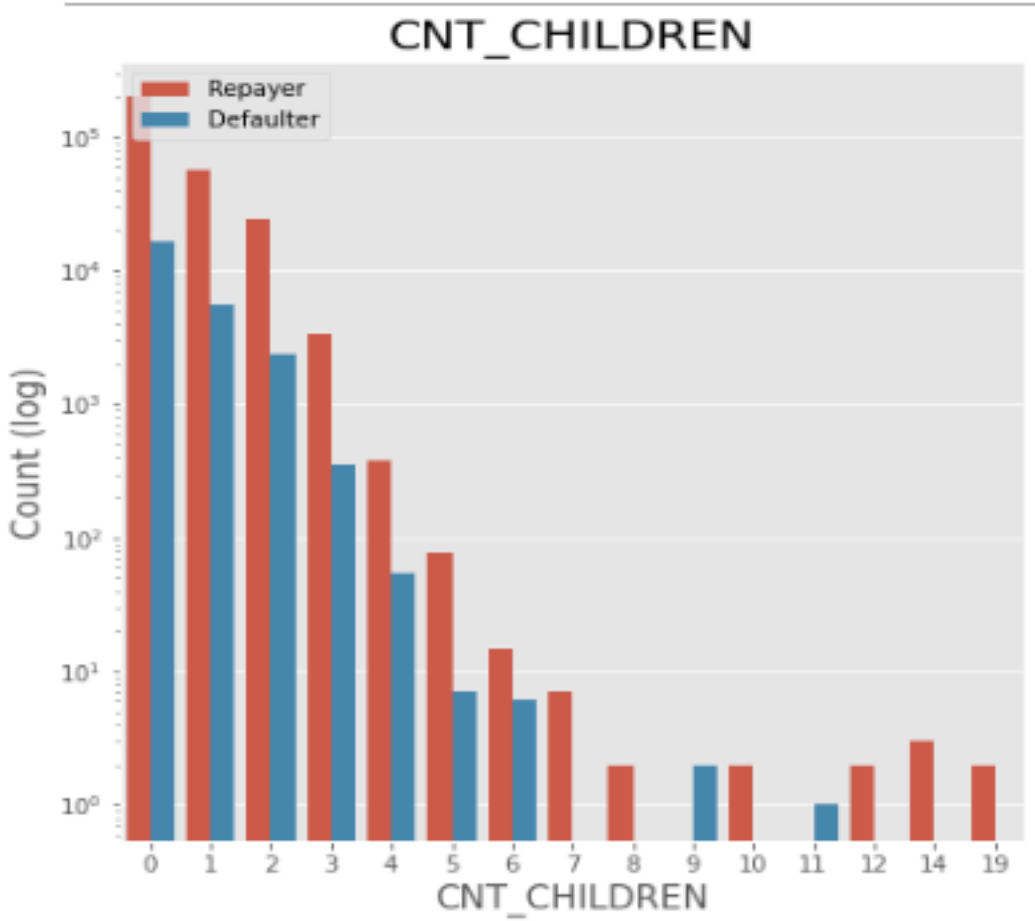
There are less defaulters in case of employees having high Salary (7-8lakh)

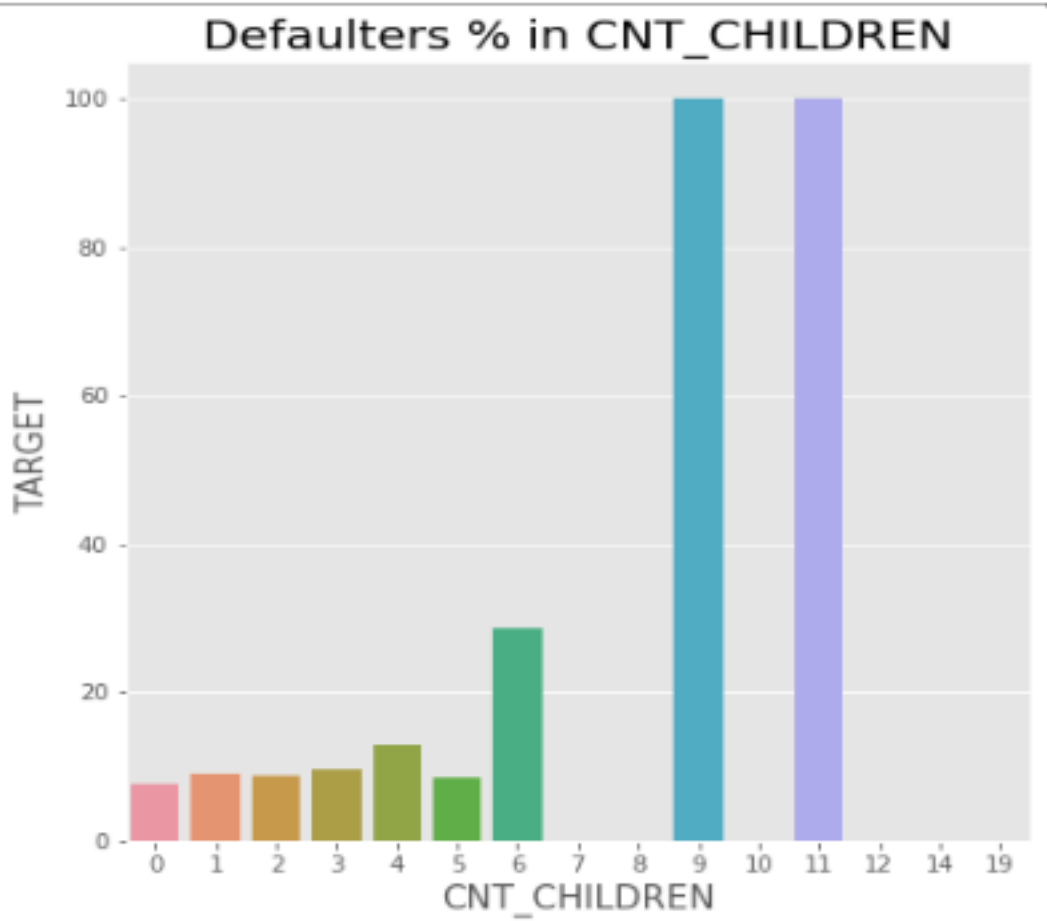




CNT_CHILDREN:

People with 0-2 Children's repays the loan on Time and defaults less





1. Factors that decide whether or not an applicant will be Defaulter

Factor that decide whether or not an applicant will be Defaulter:

CODE_GENDER: Men's have more defaulter rate comapred to women NAME_EDUCATION_TYPE:

People with Lower Secondary & Secondary education default alot NAME_INCOME_TYPE:

Unemployed People Default more

REGION_RATING_CLIENT: People who live in Rating 3 has highest defaults.

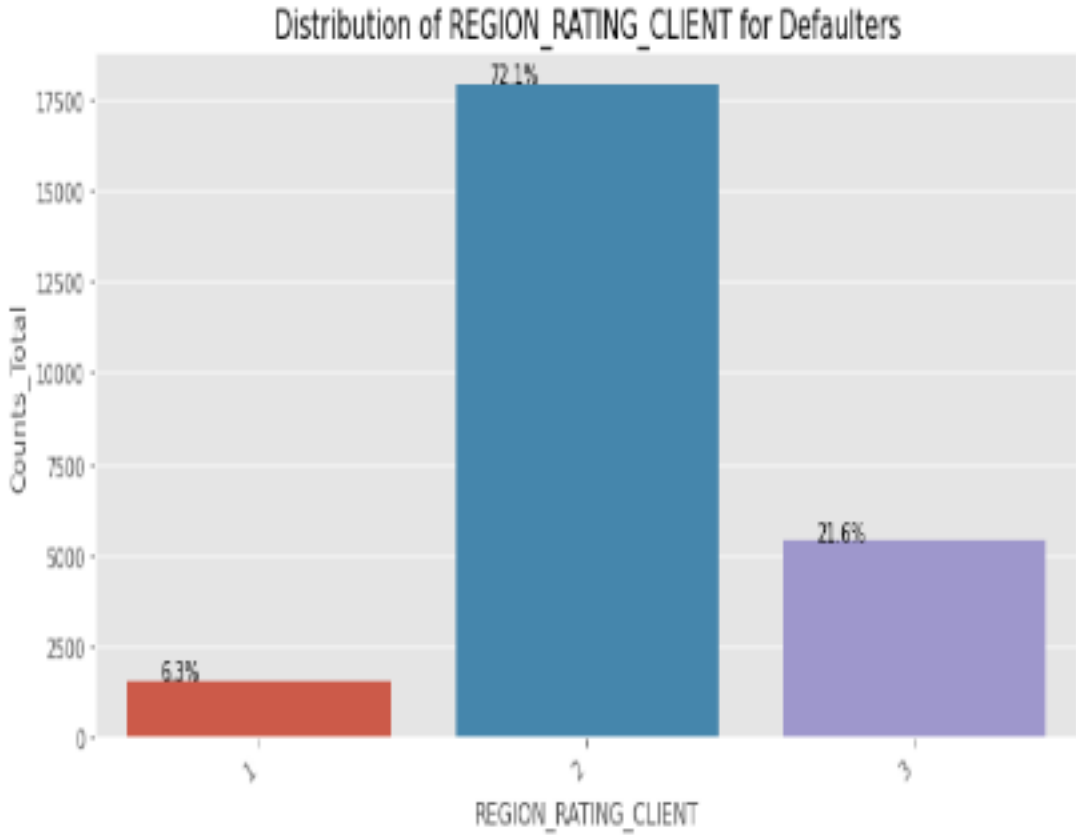
OCCUPATION_TYPE: Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as their default rate is huge.

AGE_GROUP: 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.

REGION_RATING_CLIENT:

Rating 3 has highest default rate



SUGGESTIONS:

Recording the reasons for previously cancelled client can help in identifying the terms for negotiation with the clients as previously cancelled clients have actually repayed the loan.

Female candidates are more likely to pay the Loans on Time so there can be better schemes introduced for them like when on maternity without salary the payment dates could be made more relaxed.

Target the Audience with higher income and make them to take loan of higher amount with lower interest rate schemes.