

EDA Case Study



Steps for EDA

- Sourcing the CSV data
- Understanding the data
- Finding & Deleting some un-necessary features
- Checking missing values
- Checking outliers and imputing it wrt index data
- Converting the data to Categorical datatype
- Analysis
 - Univariate analysis
 - Bivariate Analysis

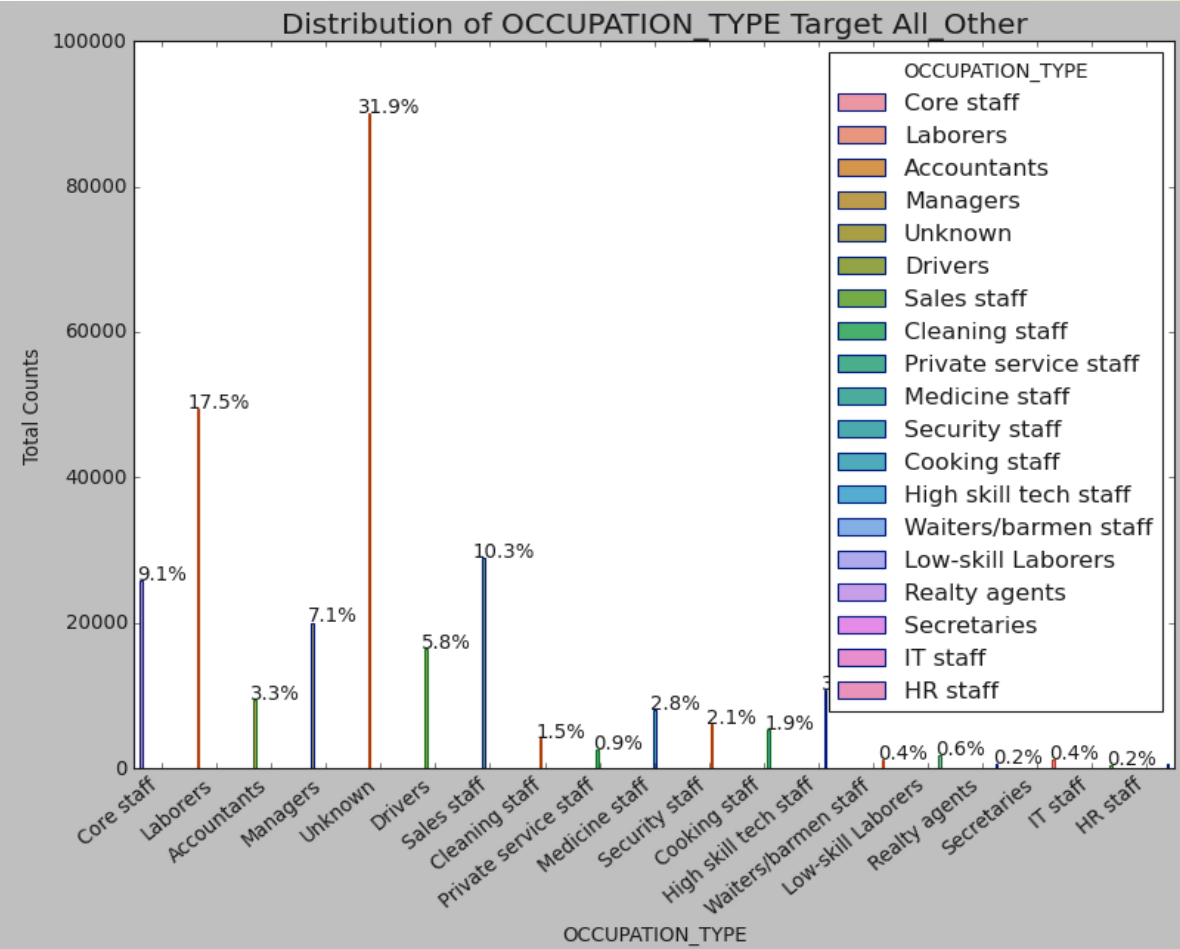
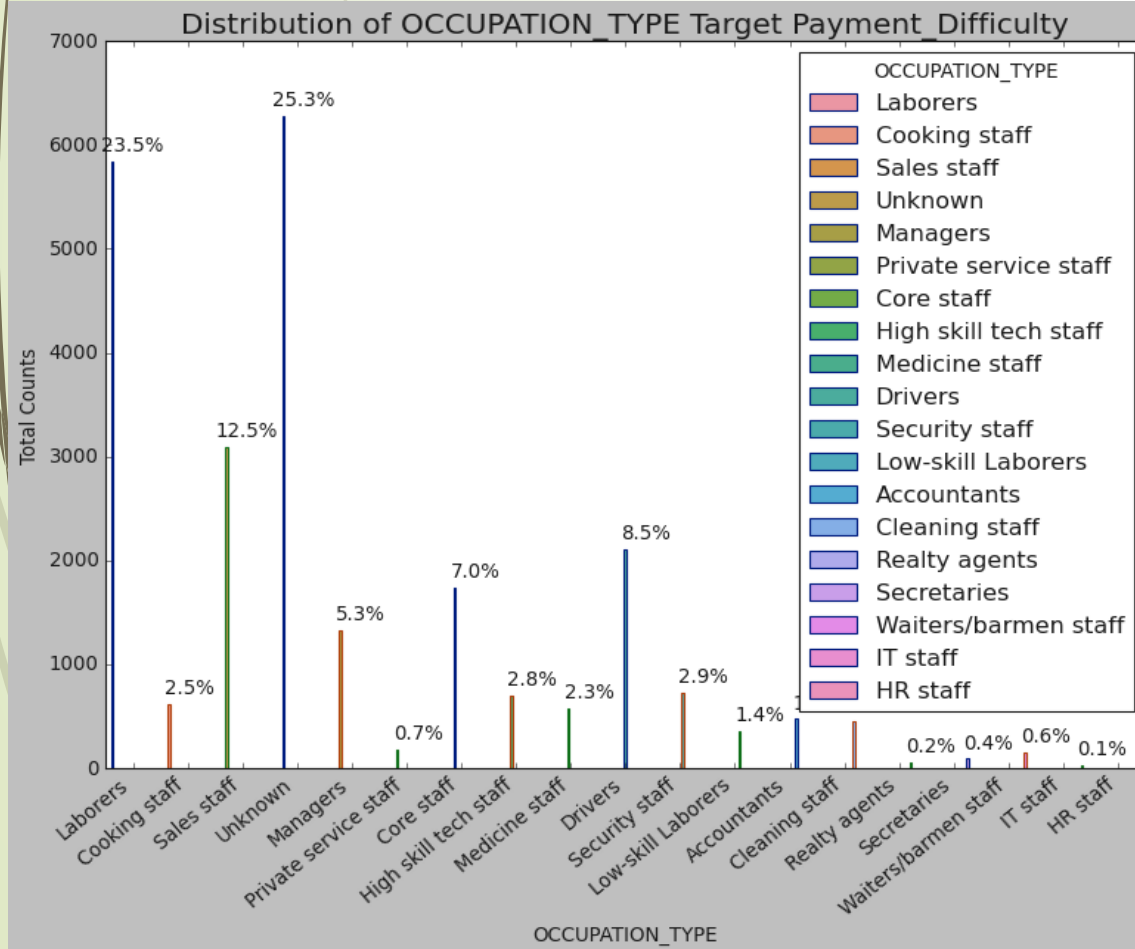


Problem Statement

- The business problem is related to finding their status of ability to return loan
- Checking their current capability with various factors and previous records of repayment.
- At the end we have to conclude that where is the higher risk if bank giving loan to that category

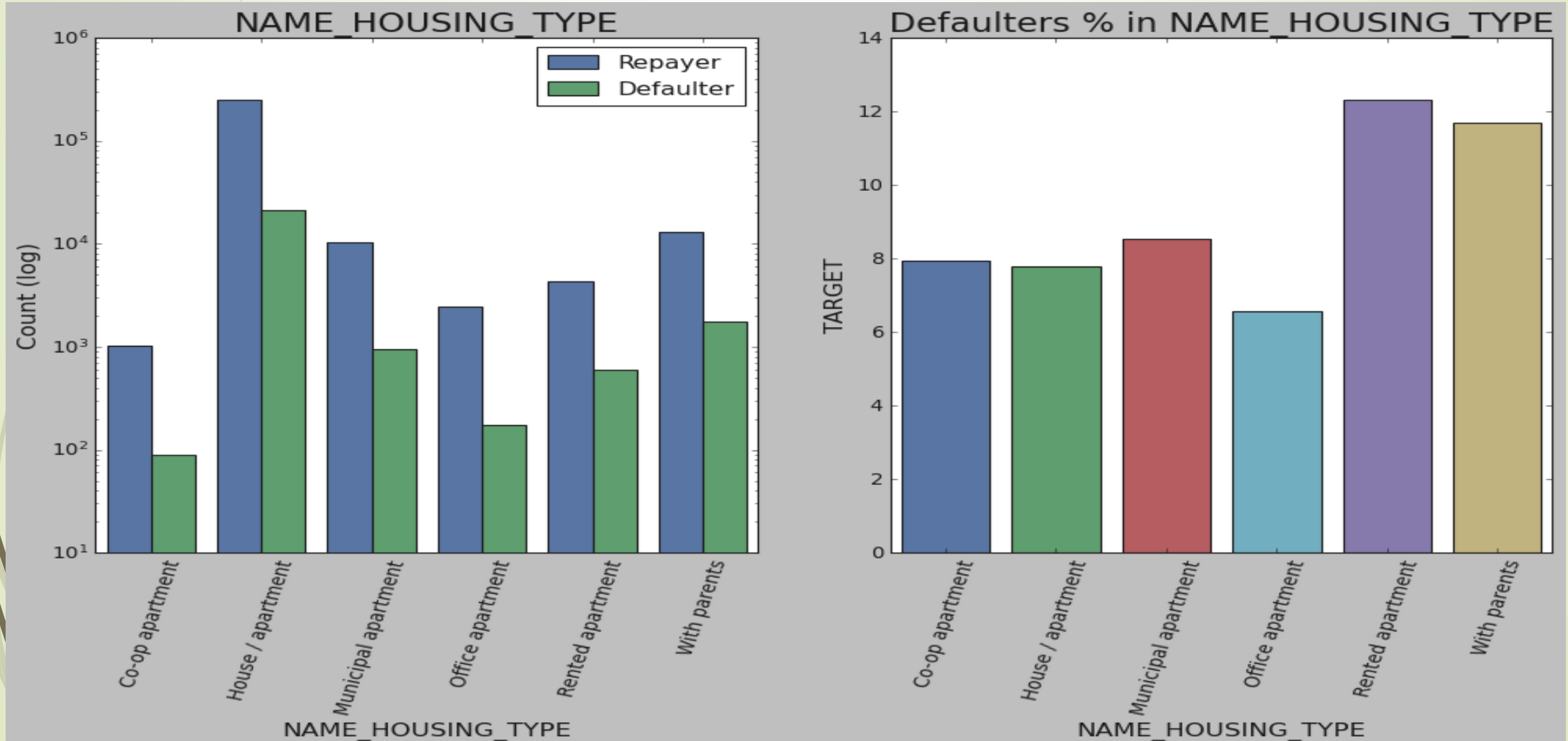
Occupation Type

It seems that the Laborers are taking most of loans followed by Sales Staff and then Drivers



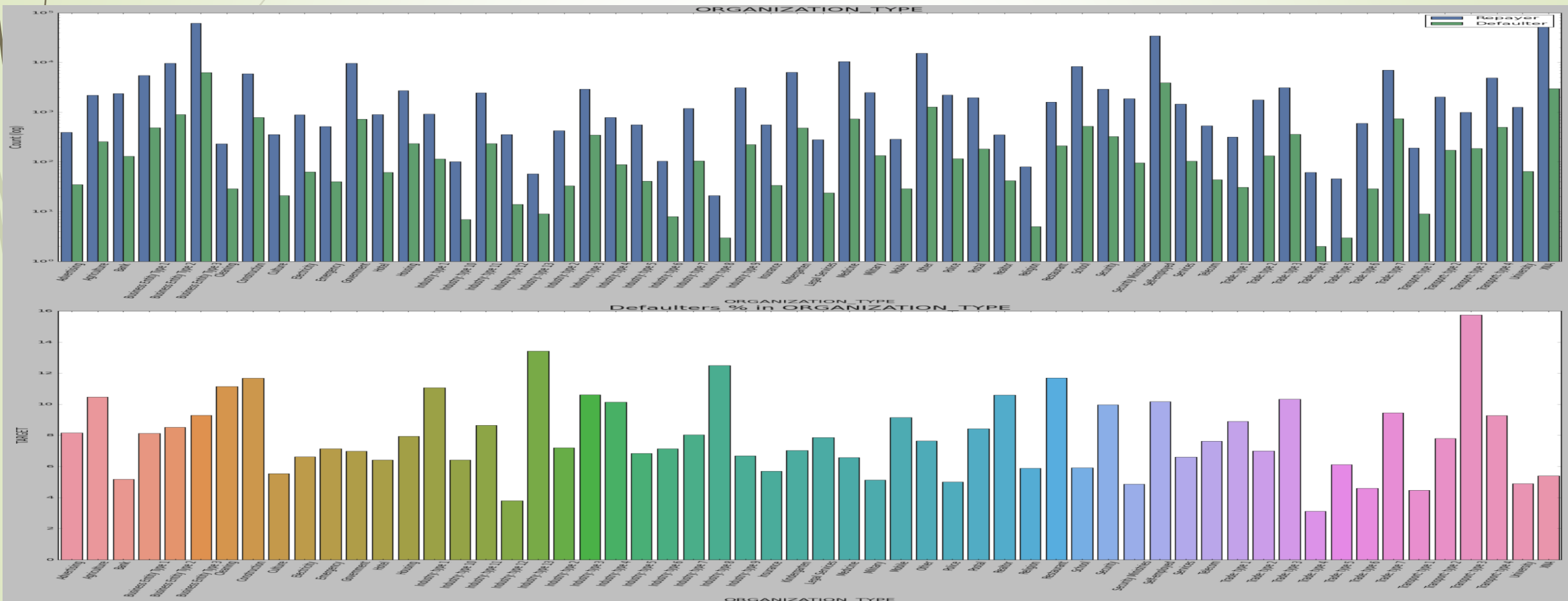
Housing Type

It is observed that Most of peoples living with Family
And the peoples living in rented apartments and with parents having
higher defaulter percentage



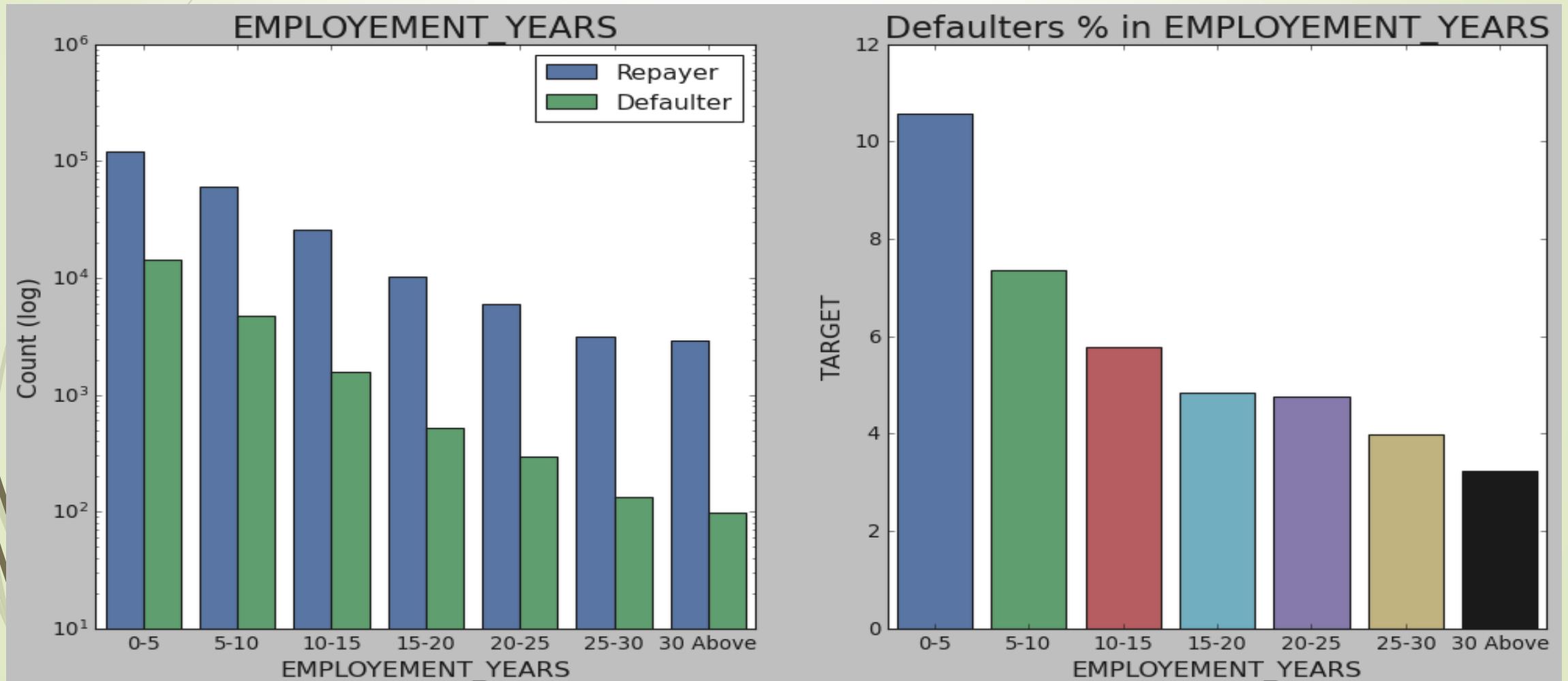
Organization Type

Organization type of like Transportation & Industry peoples having more numbers of defaulters



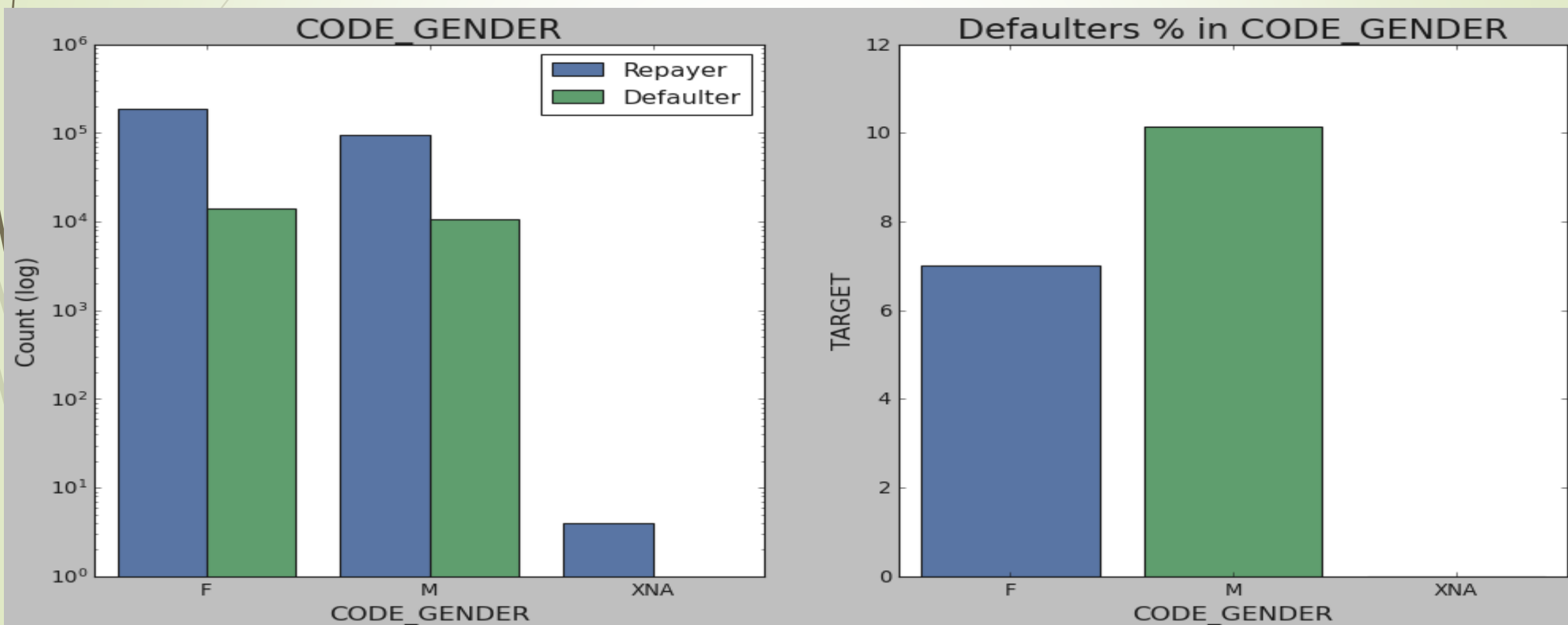
Employment Years

More numbers of defaulters having upto 5 years of experience and their employment is also high



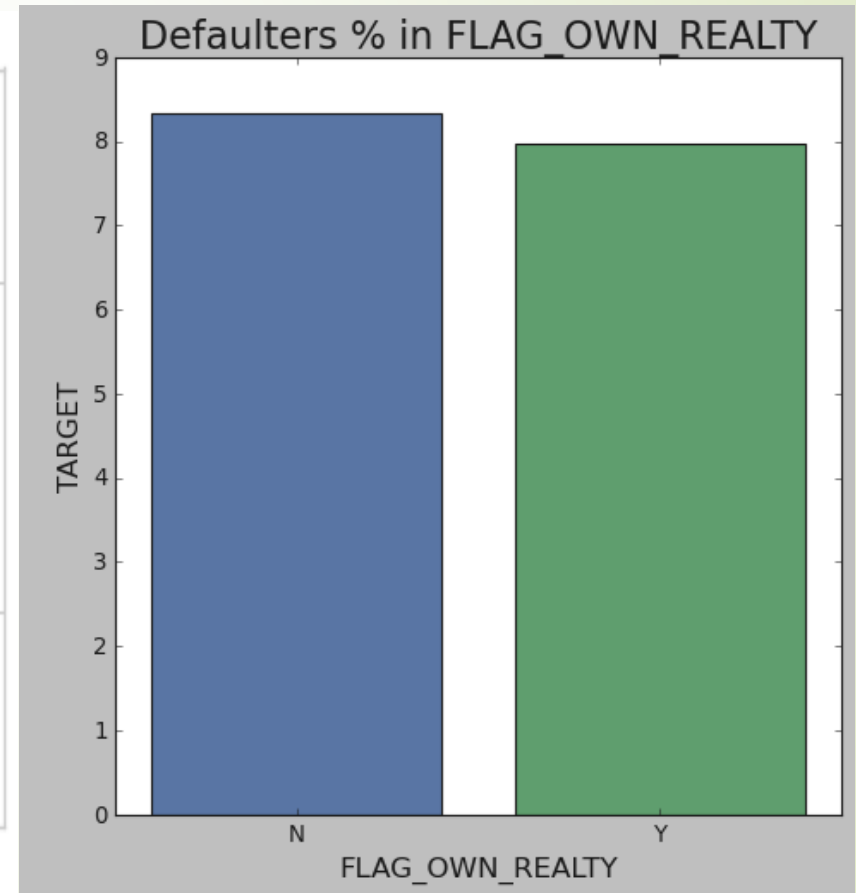
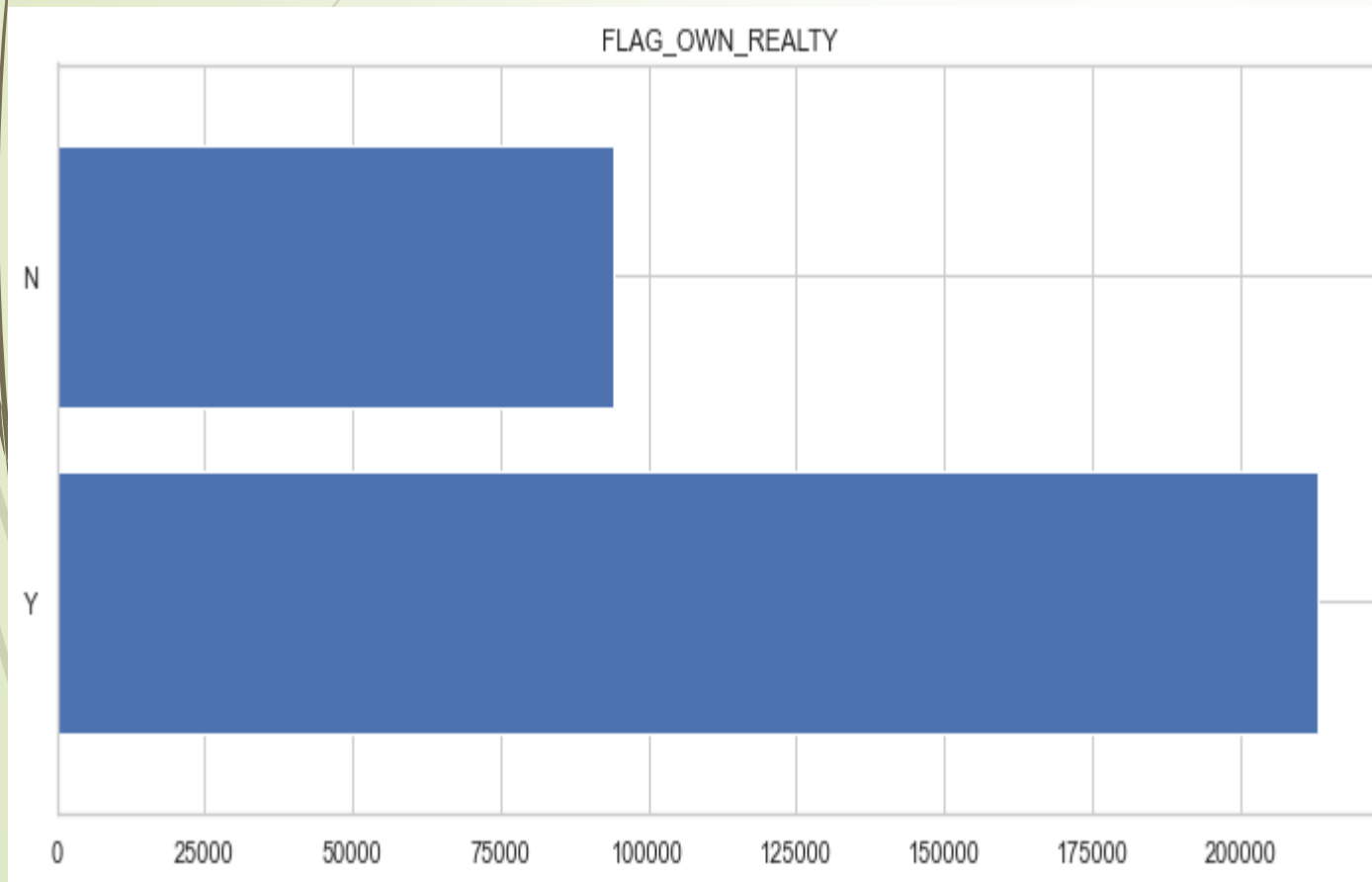
Code Gender

From below plot it is seen that Males are more numbers in defaulters than Females



Flag Own Realty

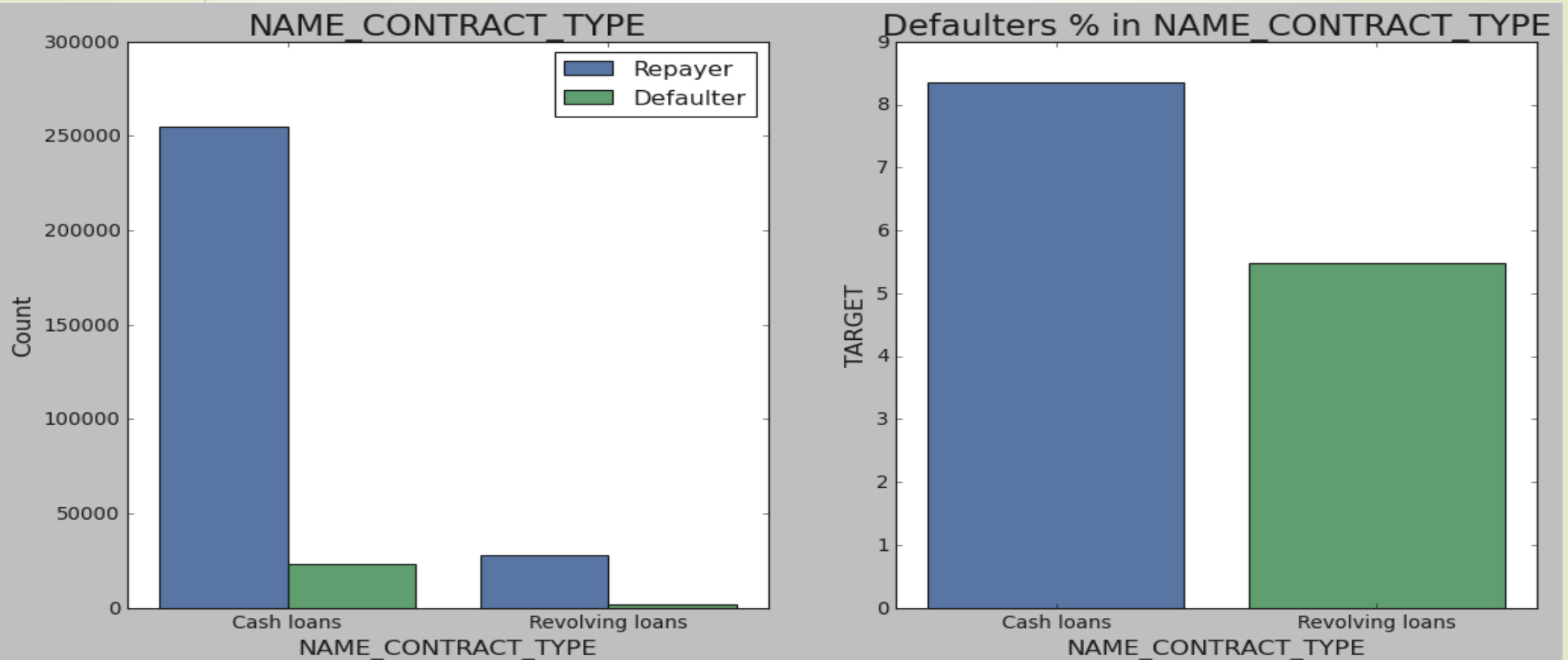
The real Estate invested are around double in numbers than not.
The plot showing around same percentage so it can't be
conclude to a particular decision



Name Contract Type

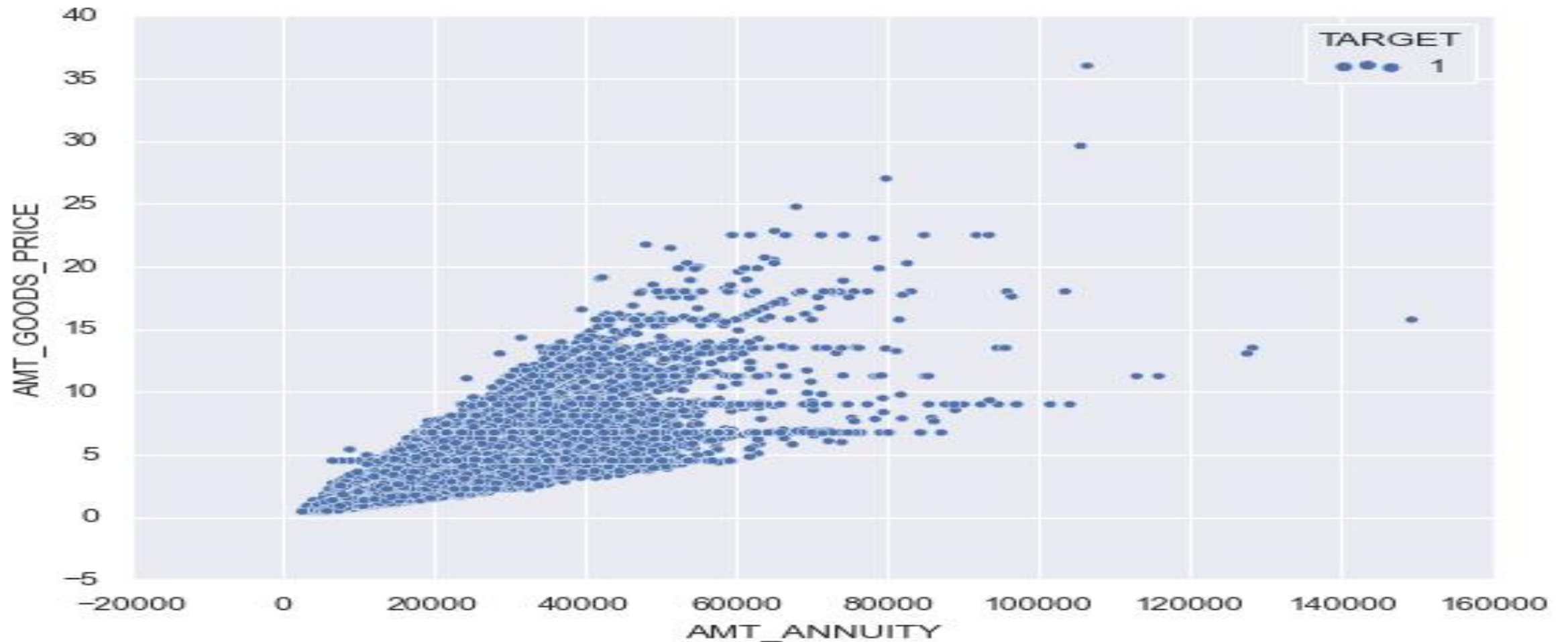
The Revolving loans are very less

Defaulters in Cash loans are around 8-9% Whereas Revolving loans having around 5-6% defaulters



Bivariate Analysis

AMT Annuty VS AMT Goods Price



Imbalance

Around 90% of peoples are repaying the loan amount and roughly 10% are may leading deaefaults at the time of repayment of loan

Imbalance data (Repayer Vs Defaulter)

