# Presentation on CREDIT EDA Case Study

# PURPOSE

- Credit risk analysis will help the company to make a decision for loan approval, refusal, cancellation based on the applicant's profile analysis.
- Helps banks to avoid financial loss by approving loan to clients that doesn’t have to repay loan later on.

## STEPS TO FOLLOW FOR BETTER ANALYSIS

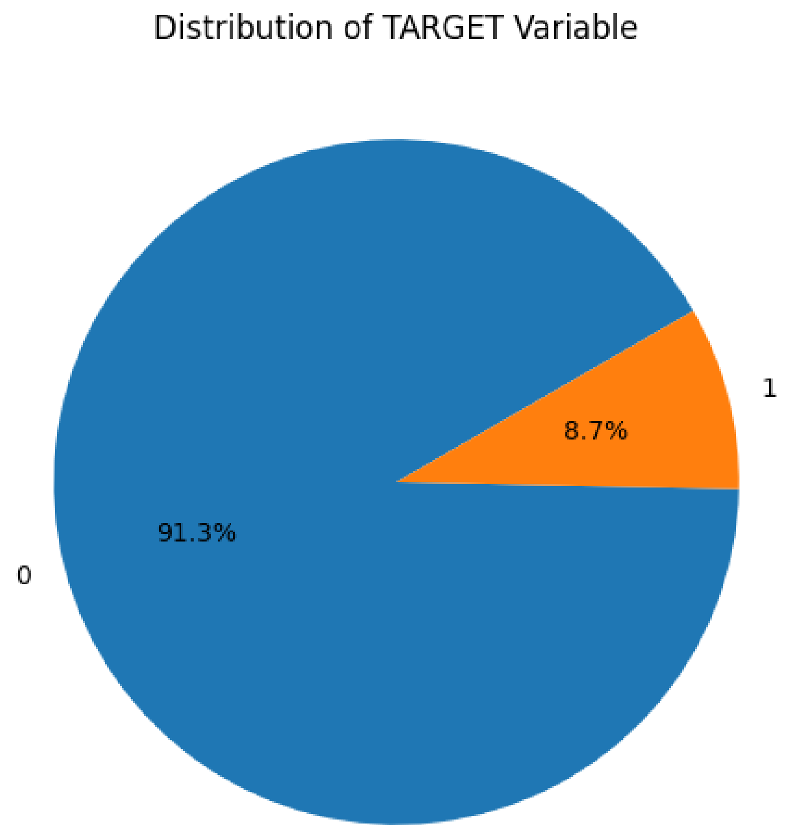
1. **Understanding the data (Data info, dimensions, shape, statistics, dtypes, description about columns)**
2. **Handling missing data**
3. **Data cleaning**
4. **Analysis – Univariate, Bivariate, Multivariate (for both Categorical and Numercial variables)**
5. **Feature Engineering (Binning, scaling)**
6. **Visualization and insights**

# Regular check

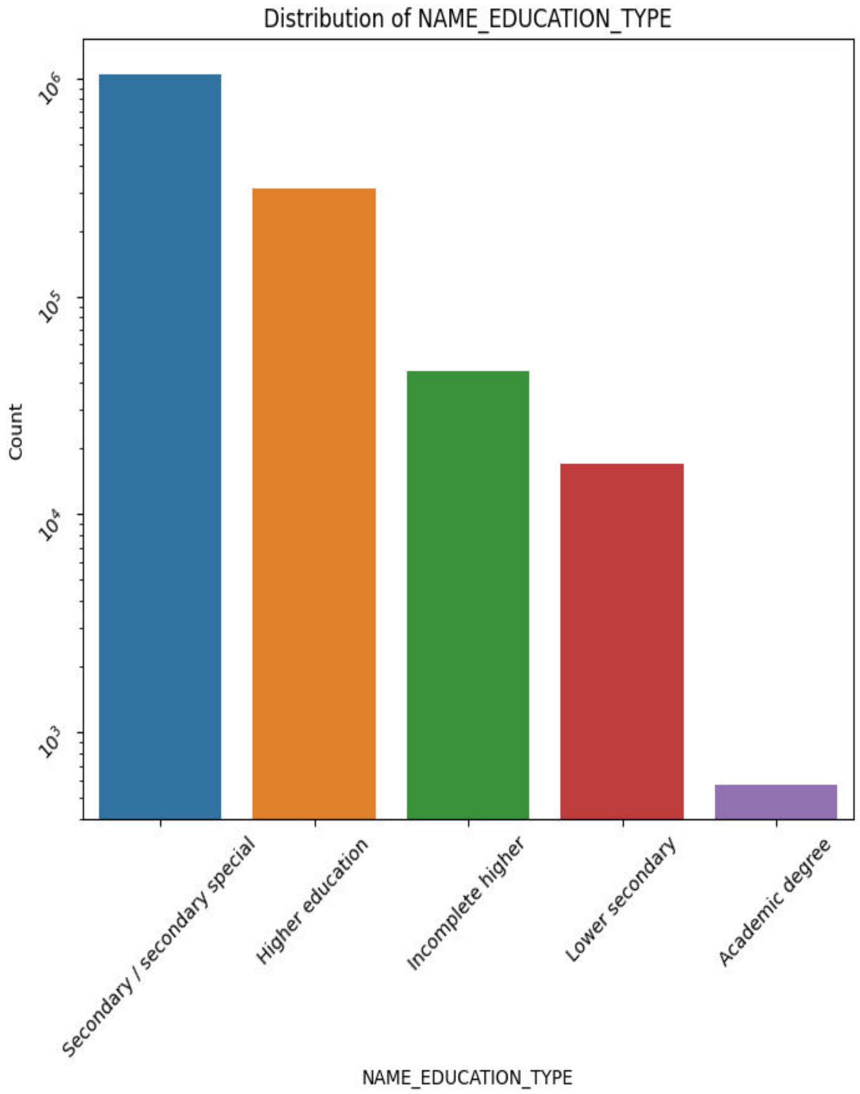
- Checking data info, dimensions, shape, statistics, dtypes, description about columns.
- Dropped the columns with missing values more than 50% .
- Dropped the columns found to be of no importance in analysis.
- Suggested imputations for columns having missing values less than 50% .
- Fixing data containing negative values and which can't be negative (like Days, Age).
- Here in this Data set includes Days data. Let's fix using abs() function to covert any negative value to positive .
- Conversion of Day in birth or any other Days data to Year format for better analysis later on.

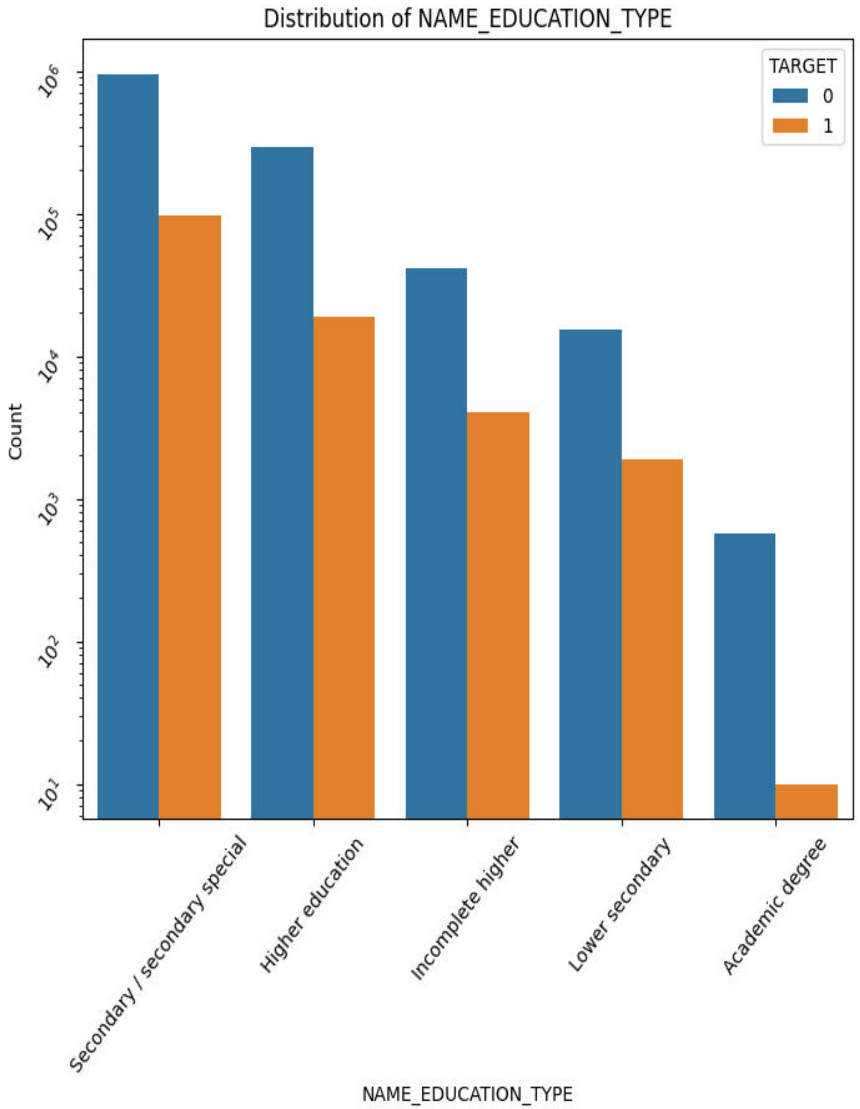
# DATA IMBALANCE RATIO

- **91.3%** are having non- defaulter (Target = 0) category for merged dataframe.
- **8.7%** are having defaulter (Target = 1) category for merged dataframe.
- Finding percentage of client with outstanding dues/payment difficulties and no outstanding dues.(In merged data set)
- Imbalance ratio for merged dataframe is **10.55.**
- Imbalance ratio for primary dataframe (application.csv) is **11.39.**



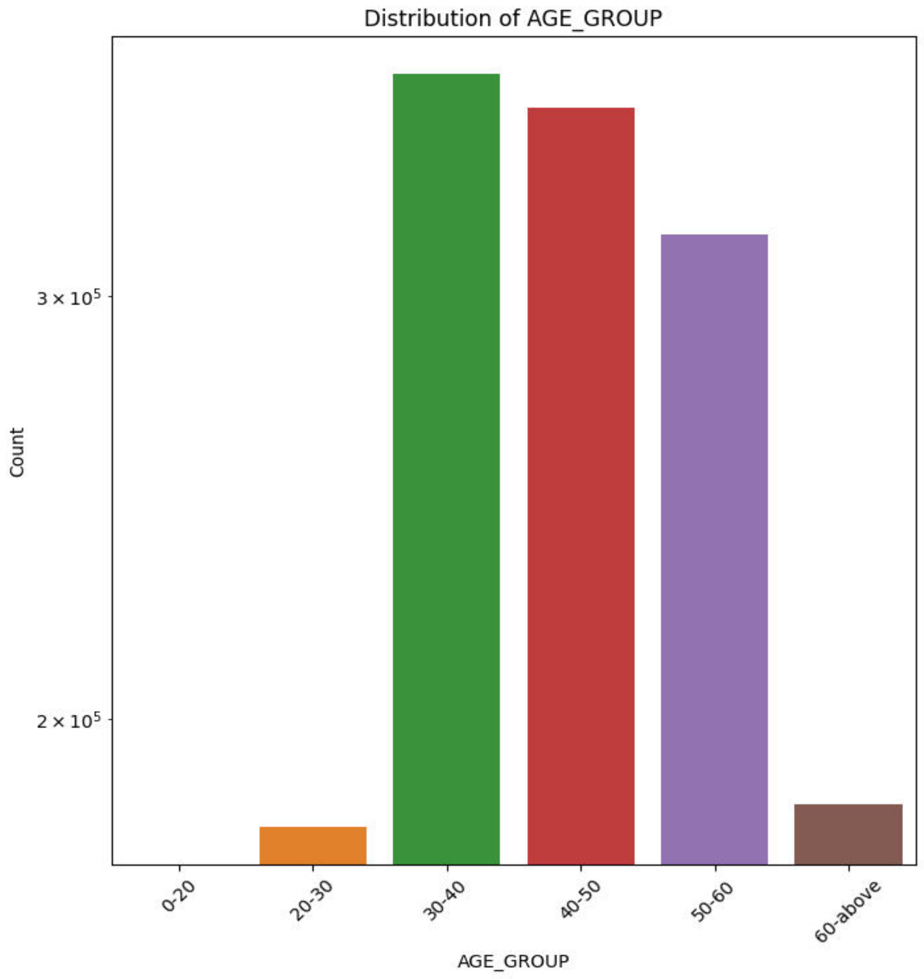
# Defaulter vs Non defaulter analysis for NAME_EDUCATION_TYPE

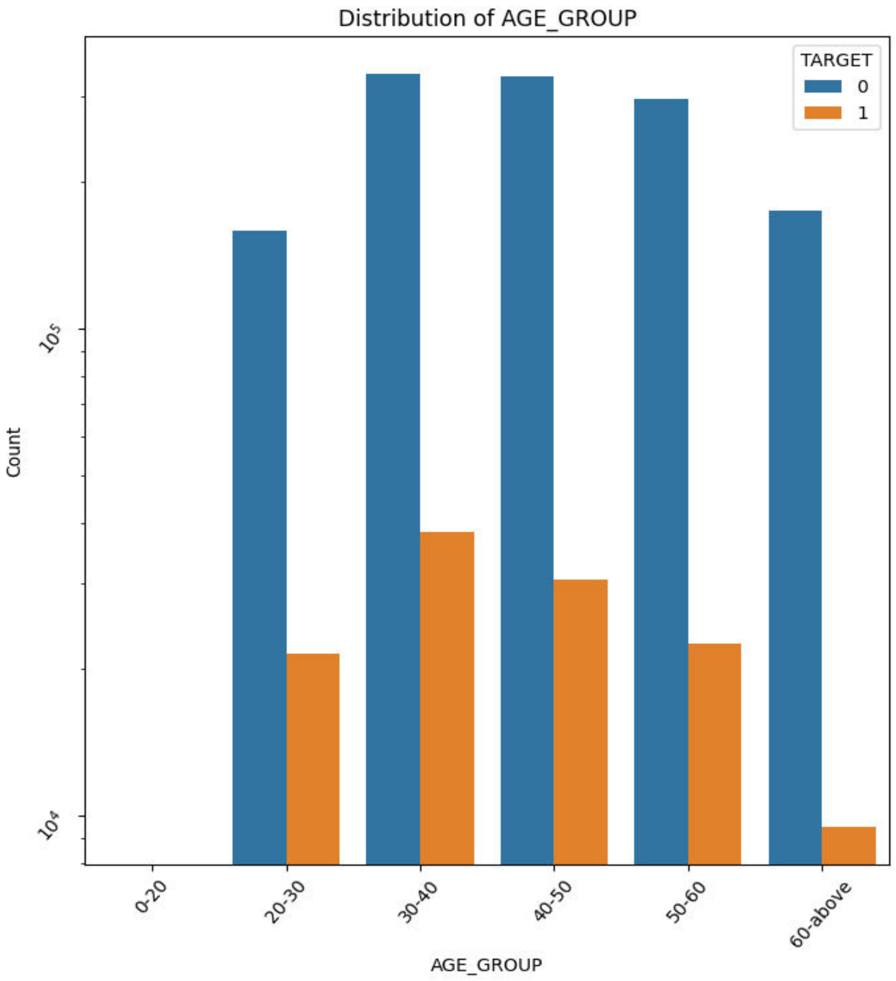



- **Clients or people having education Secondary/ secondary special applied more for the loan than others.**
- **Clients of education Secondary/ secondary special found to have more proportion (target 1) having difficulties to pay laon/ found to have defaulting percentage.**
- **While as education level increases the proportion to default decreases and vice-versa.**

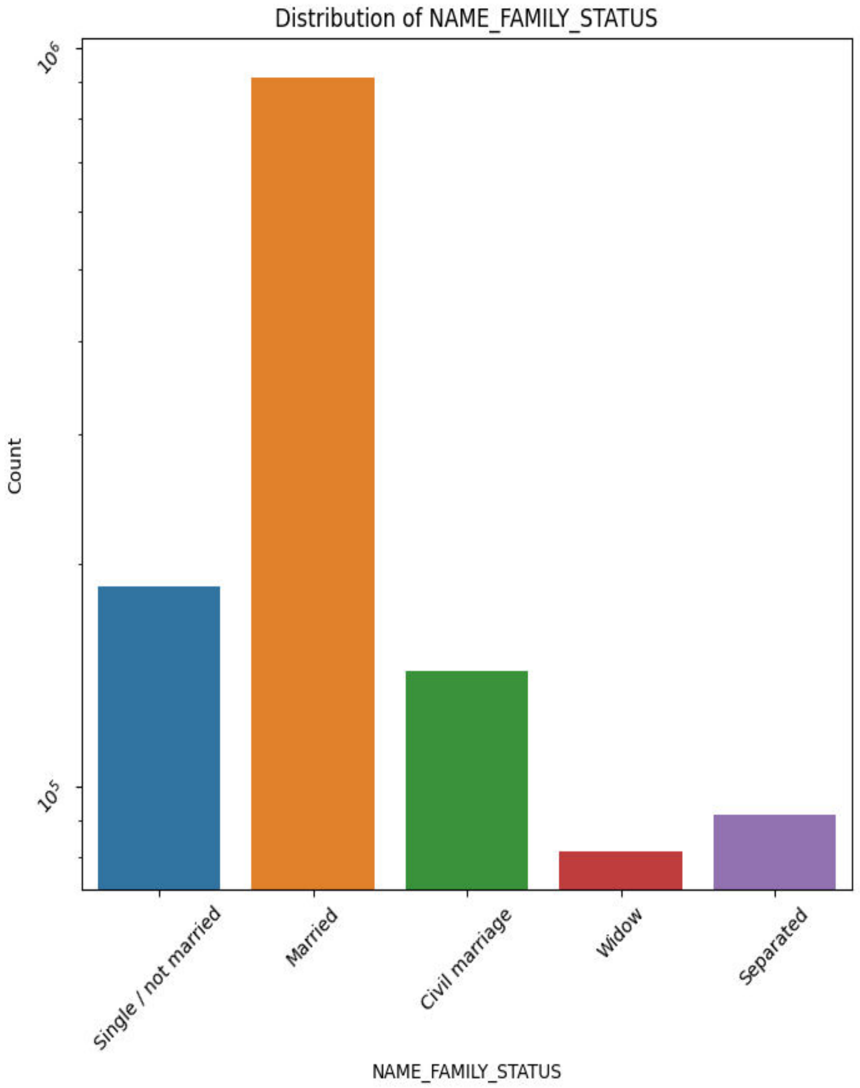
# Defaulter vs Non defaulter analysis for AGE_GROUP

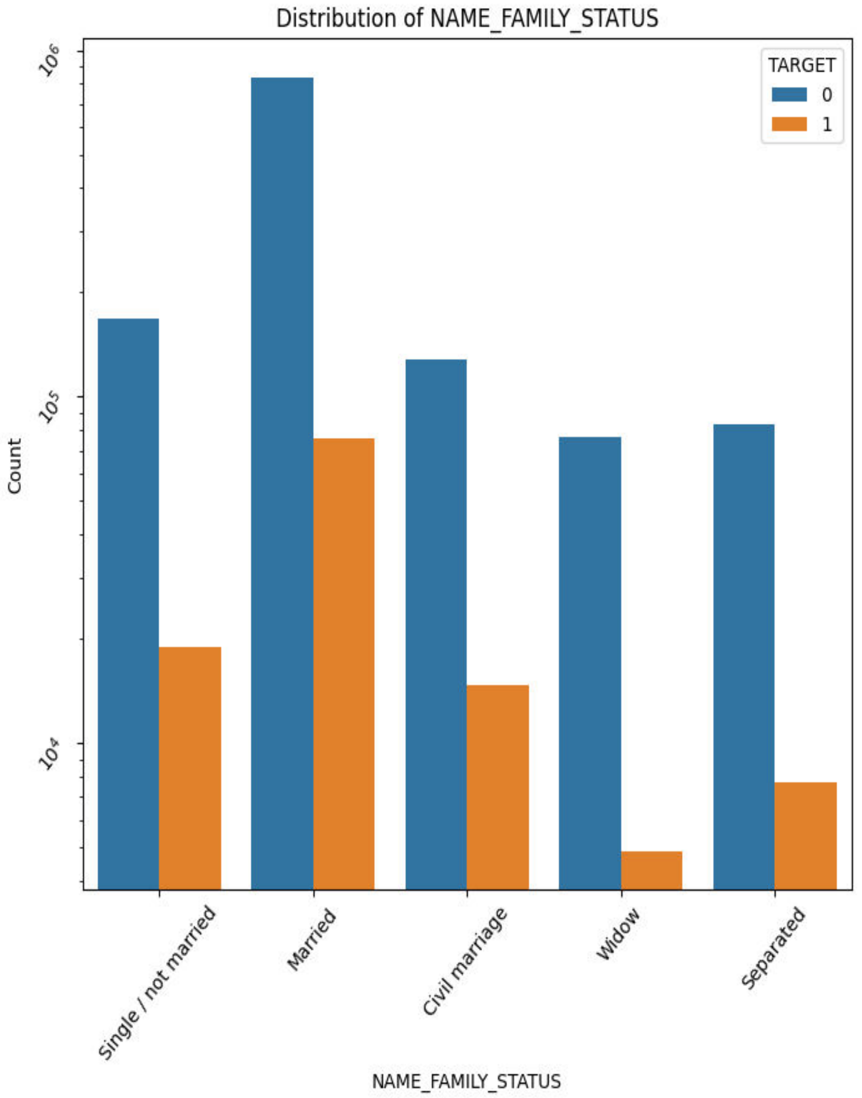



- **Clients with age 30-60 have more defaulting percentage than people with age less than 30 and above 60 age.**
- **People with 60+ age found to have least proportion doing defaults.**
- **So, more analysis to be done before giving loans to people of age 30 to age 60.**

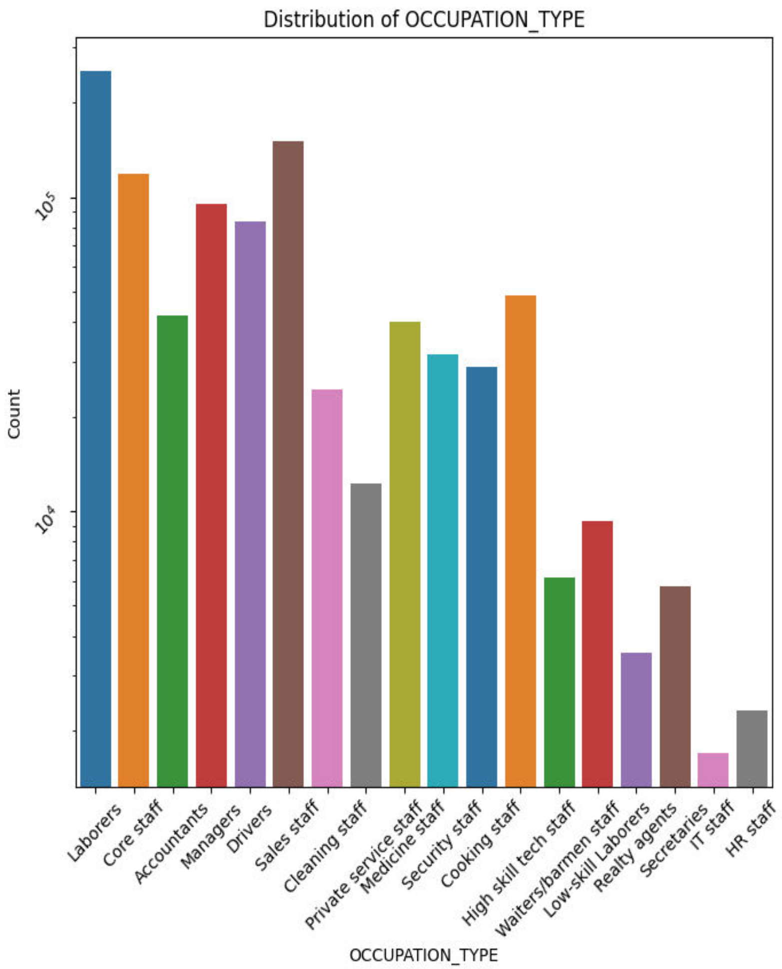
# Defaulter vs Non defaulter analysis for NAME_FAMILY_STATUS

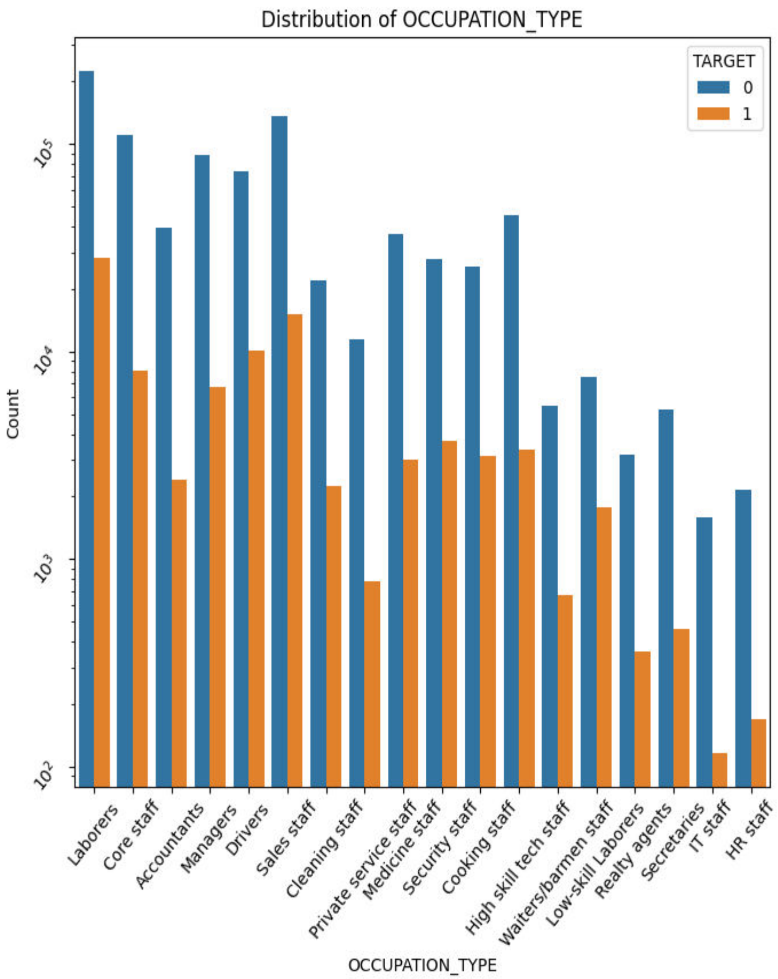



- **Married Clients applied more for the loans.**
- **Married clients found to have more default proportion than other family status.**
- **Default ratio is less in case of widows and separated family status.**
- **In case of Defaulters, Widows shows Minimum proportion for doing a default out of all.**

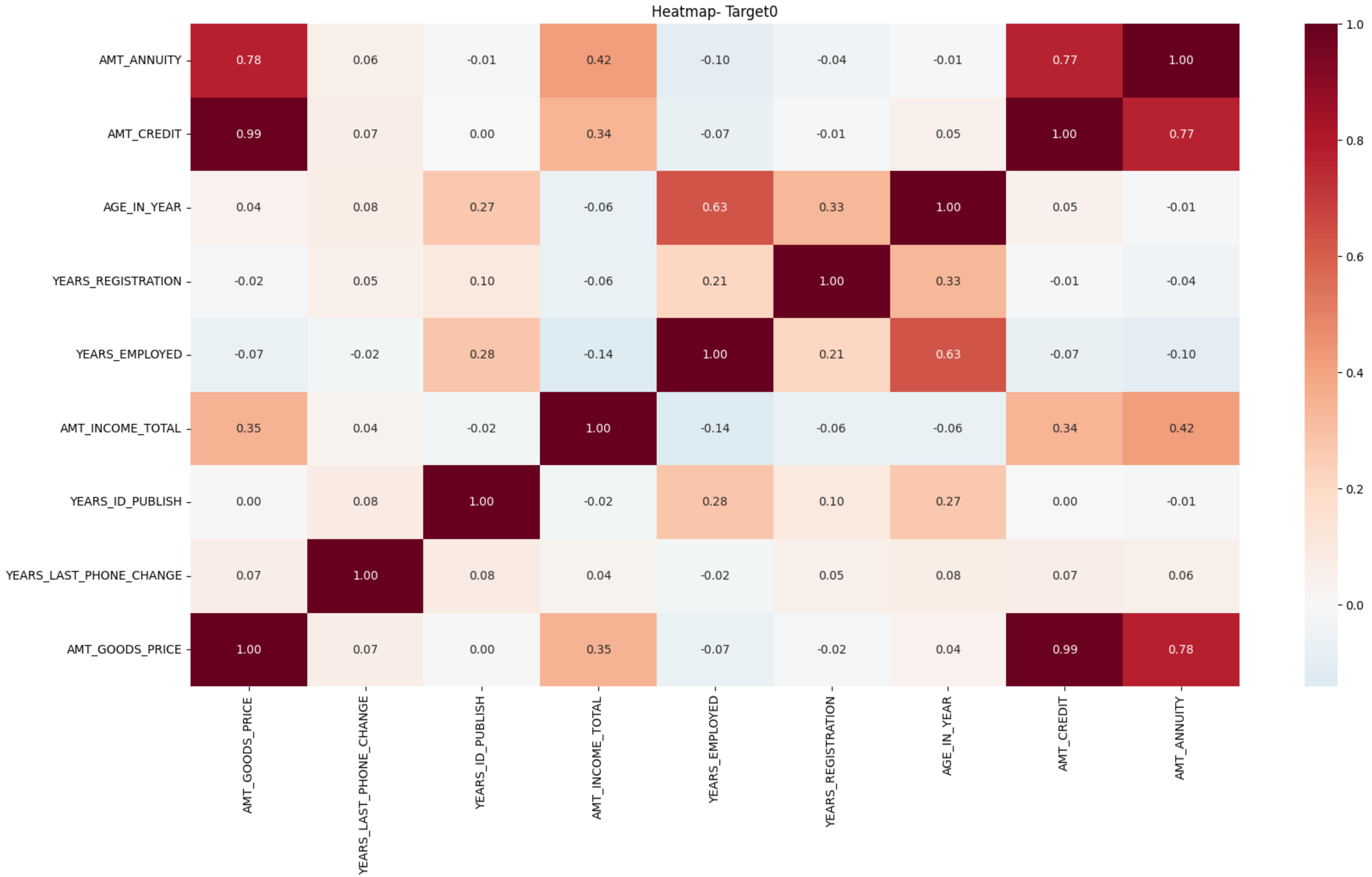
# Defaulter vs Non defaulter analysis for OCCUPATION_TYPE




- **Laborers found to have applied more for the loan.**
- **Laborers also found to have more default proportion than all other occupation type.**
- **Driver also founds to have more default proportion to non default ratio.**

# Correlation between numerical columns for target = 0, (Non-defaulters data)

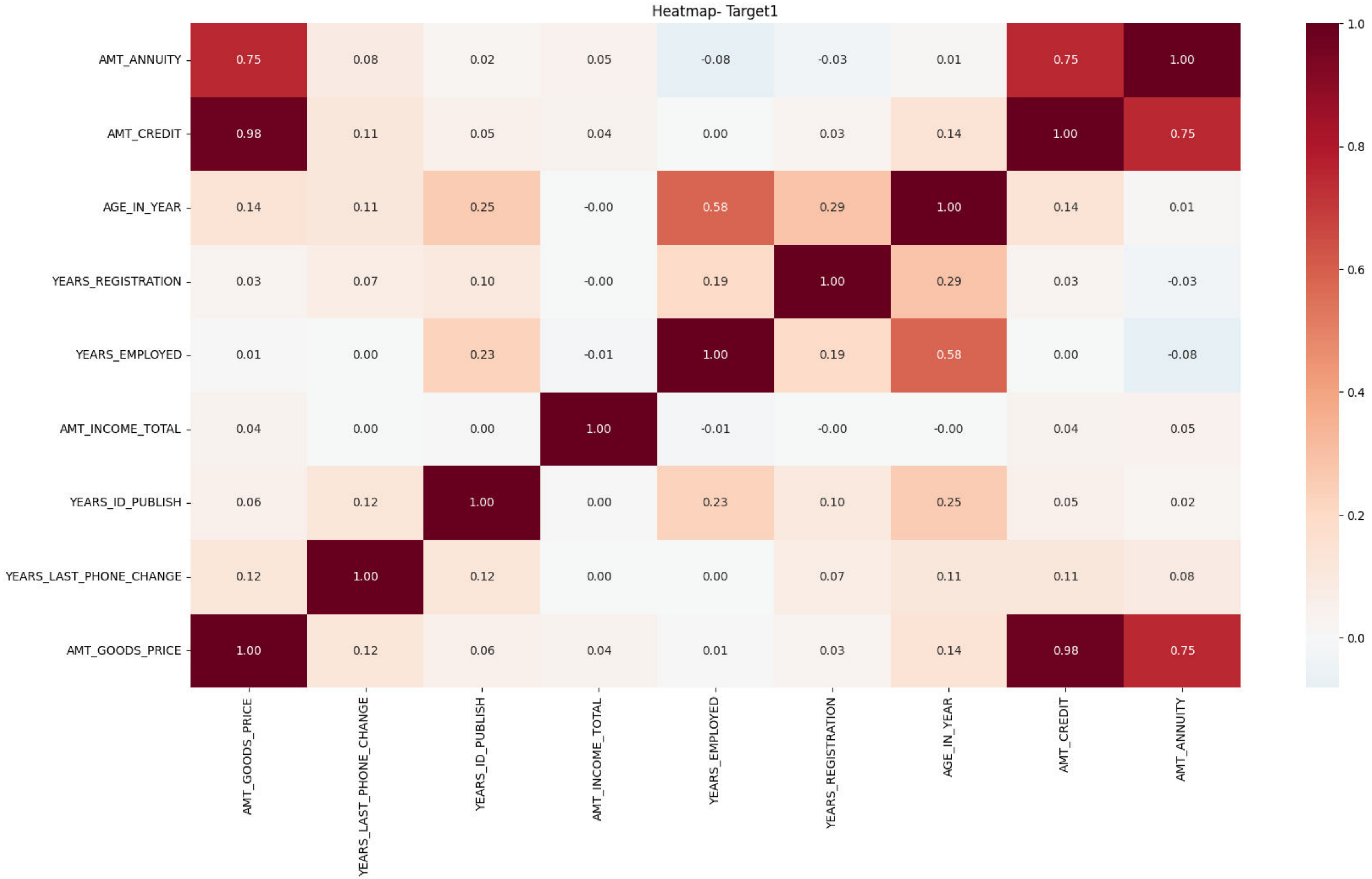



Target = 0

- Area with darker color zone shows more correlaton between variables.

- Showing that a positive high correlation  is seen between all three (AMT_ANNUITY , AMT_CREDIT and AMT_GOODS_PRICE).

# Correlation between numerical columns for target = 1, (defaulters data)




Target = 1

- Area with darker color zone shows more correlaton between variables.

- Showing that a positive high correlation is seen between all three (AMT_ANNUITY , AMT_CREDIT and AMT_GOODS_PRICE).

# Insights

- The number of loans approved to females are more than that of males.
- The inclination of number of loan approvals is high towards the secondary special education.
- Repairs have the maximum number of defaulters .
- As compared to the middle age group and senior citizens, the younger age group got less amount of credited loan.
- The loan amount credited to higher income group is more.
- The middle age group got more amount of loan credited as compared to the younger age group and senior citizens.
- Higher income group have more loan amount credited and lower the lowest.
- Lower secondary educated clients are more defaulted followed by Secondary and Incomplete higher educated clients.
- The Higher educated group are less defaulted.
- Females are less defaulted than male across all educated levels .
- Young clients with medium and low credit amount group are highly defaulted.
- Senior citizens across all credit amount groups are less likely defaulted.
- Young clients are more defaulted than Mid age and senior.
- Young low income people are more defaulted.
- For Mid age and senior people the default rate is almost same in all income group.