

Presentation on Credit EDA Assignment

- Vasundhara Kadekodi

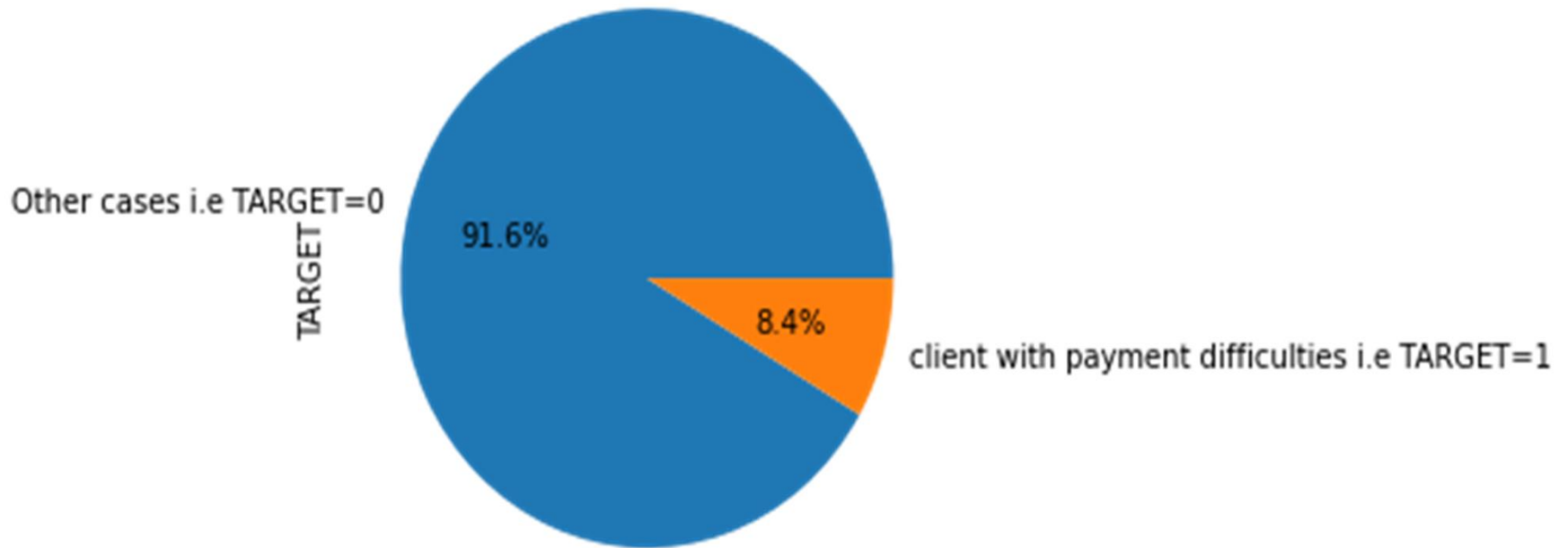
Importance of the study

- Credit risk analysis will help the company to make a decision for loan approval based on the applicant's profile. Which controls loss of business to the company and avoid financial loss for the company.

Steps involved in handling the data

- Data Understanding and Preparation.
- Check for Data quality issues.
- Handling missing values and outliers.
- Check for Data imbalance.
- Univariate, Segmented Univariate, Bivariate and Correlation analysis.
- Merging of Application data with Previous application data.
- Bivariate and Multivariate Analysis on merged data

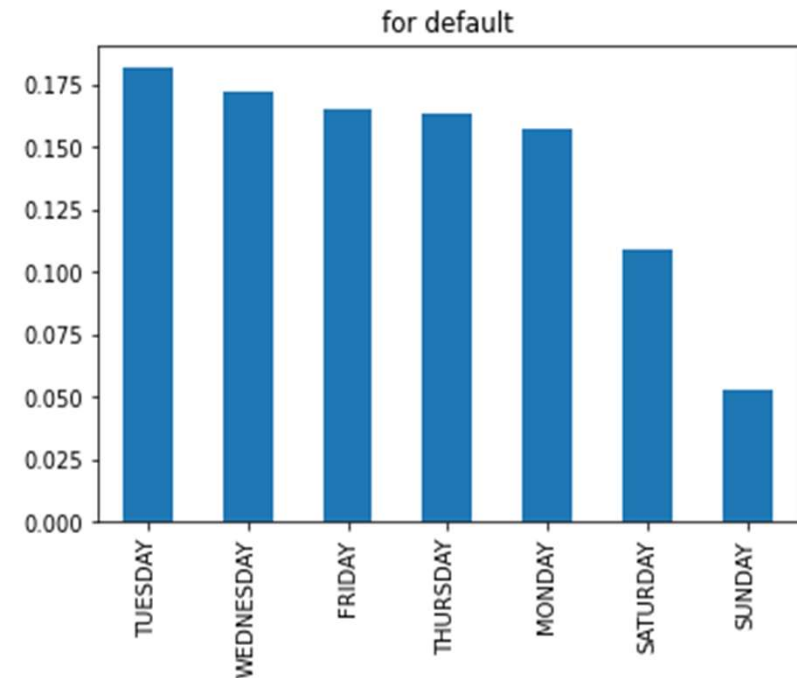
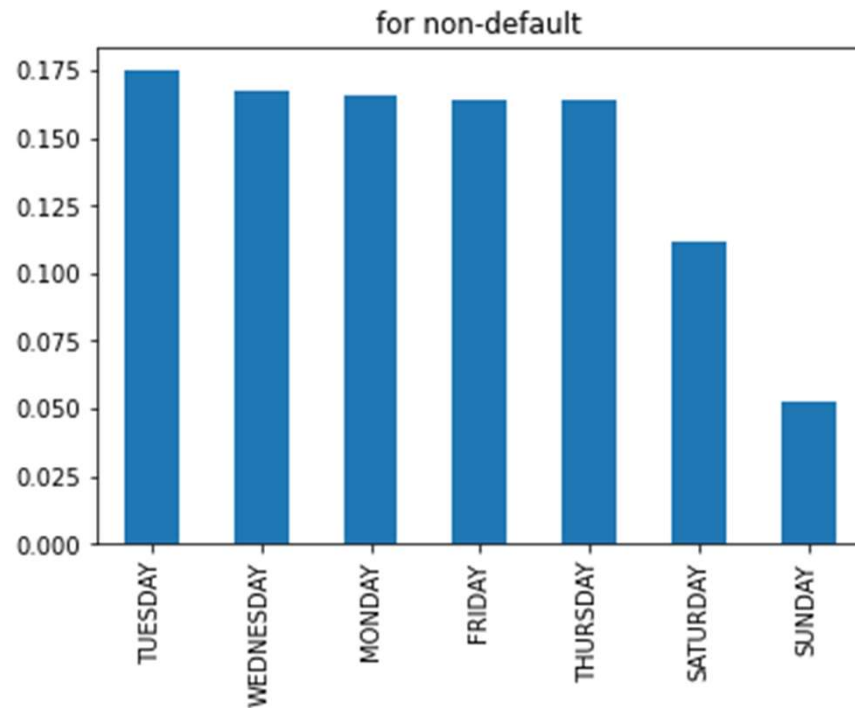
Data Imbalance



Application Data having high data imbalance between client with payment difficulties=8.4% and other cases=91.6%,

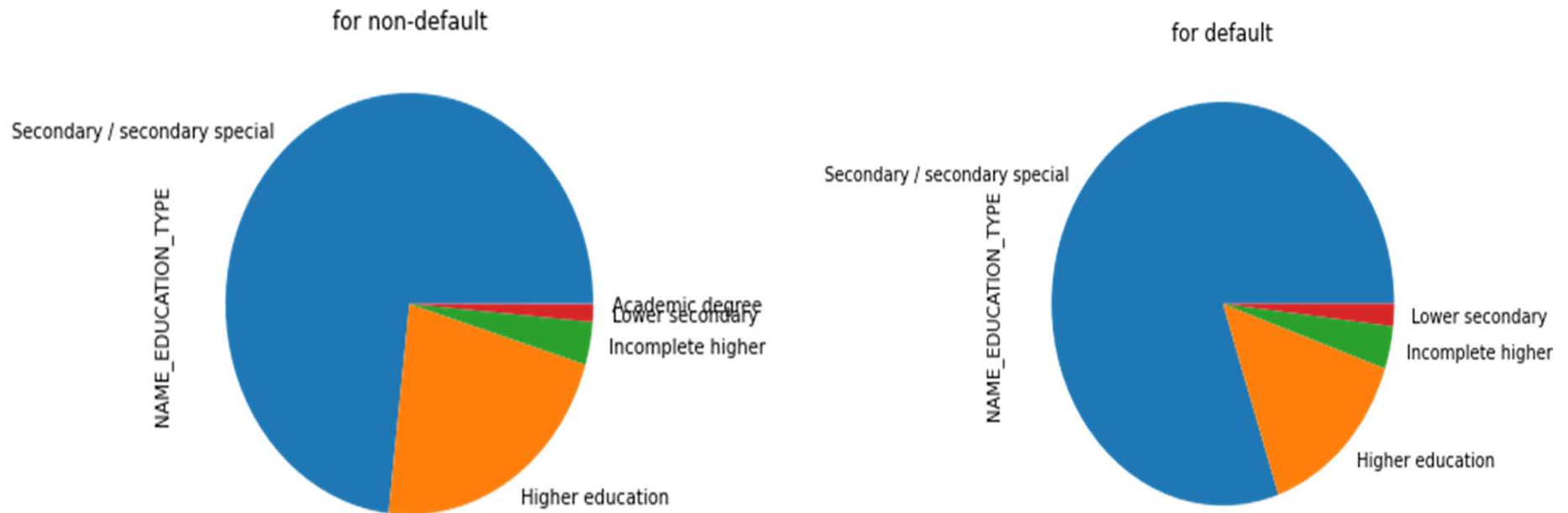
**Segmented Univariate analysis
with respect to target column**

Distribution of Bank working days on loan application process



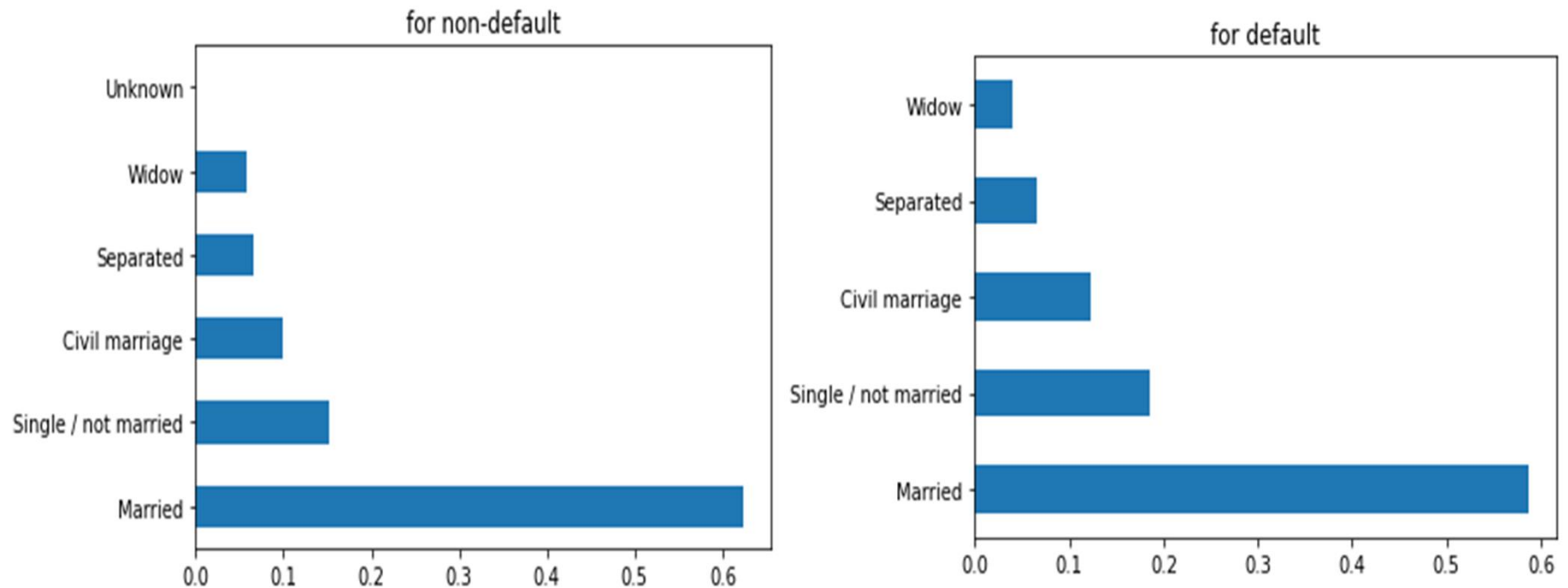
For both non-default and default, application starting processes will be less in saturday and sunday.

Distribution on Education background



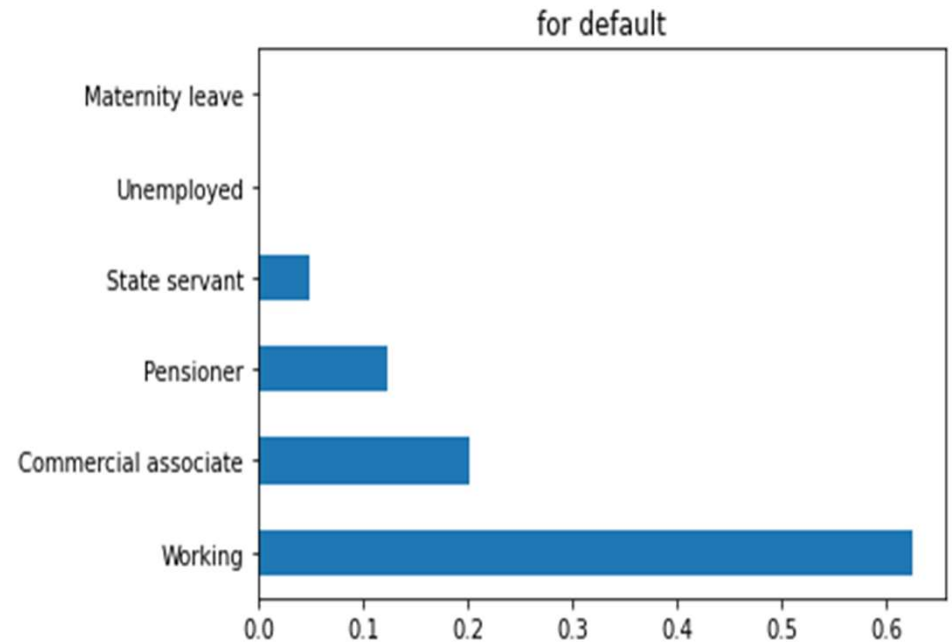
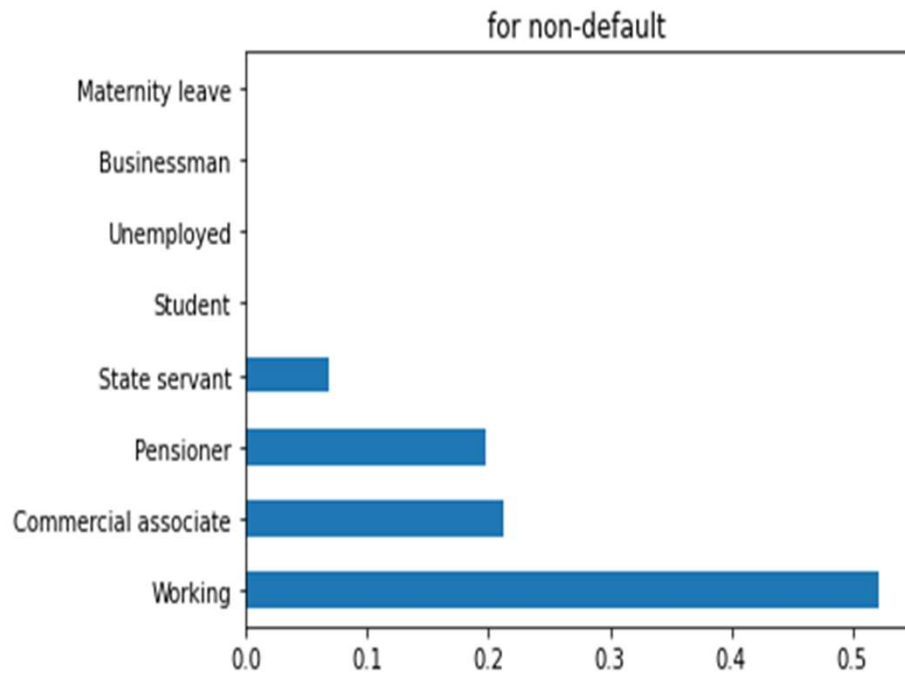
Secondary/Special type of people have high chance of applying loans as compare to other educated people

Distribution on Family background



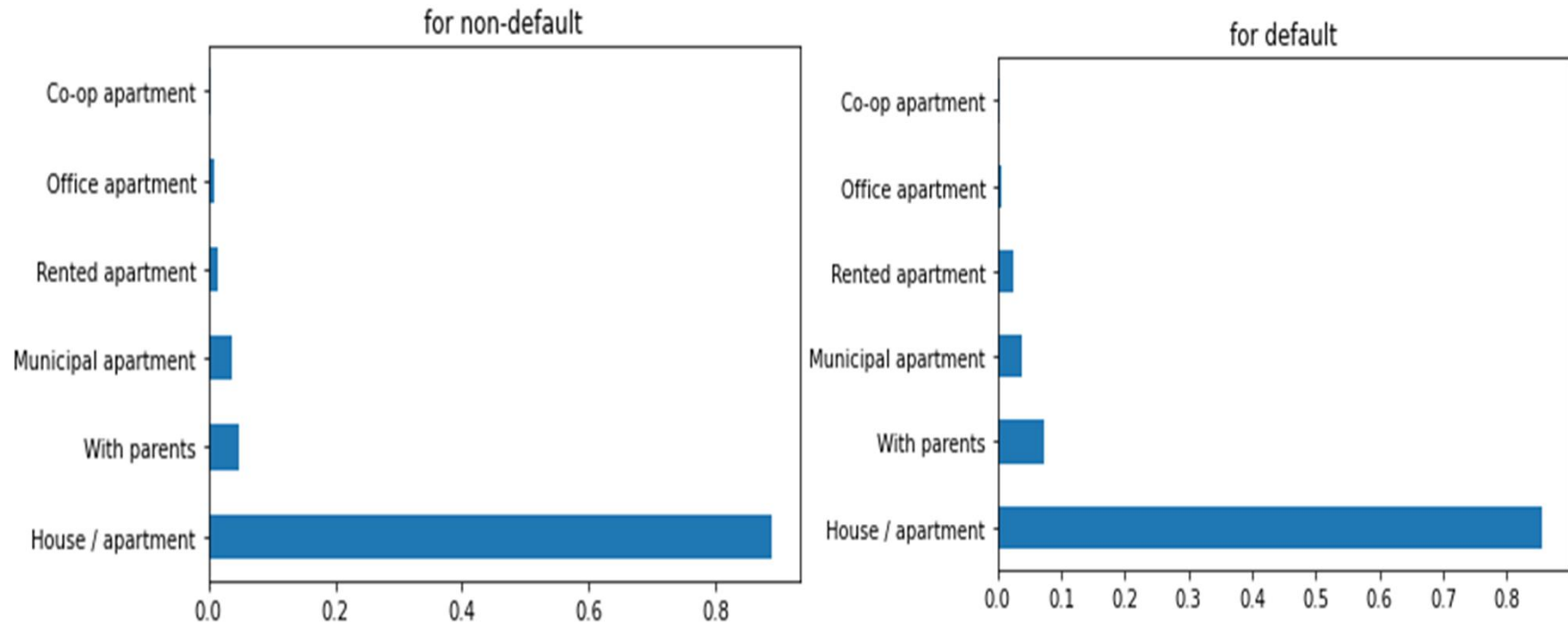
Married people have high chance to take Loan as compared to others

Distribution on Income



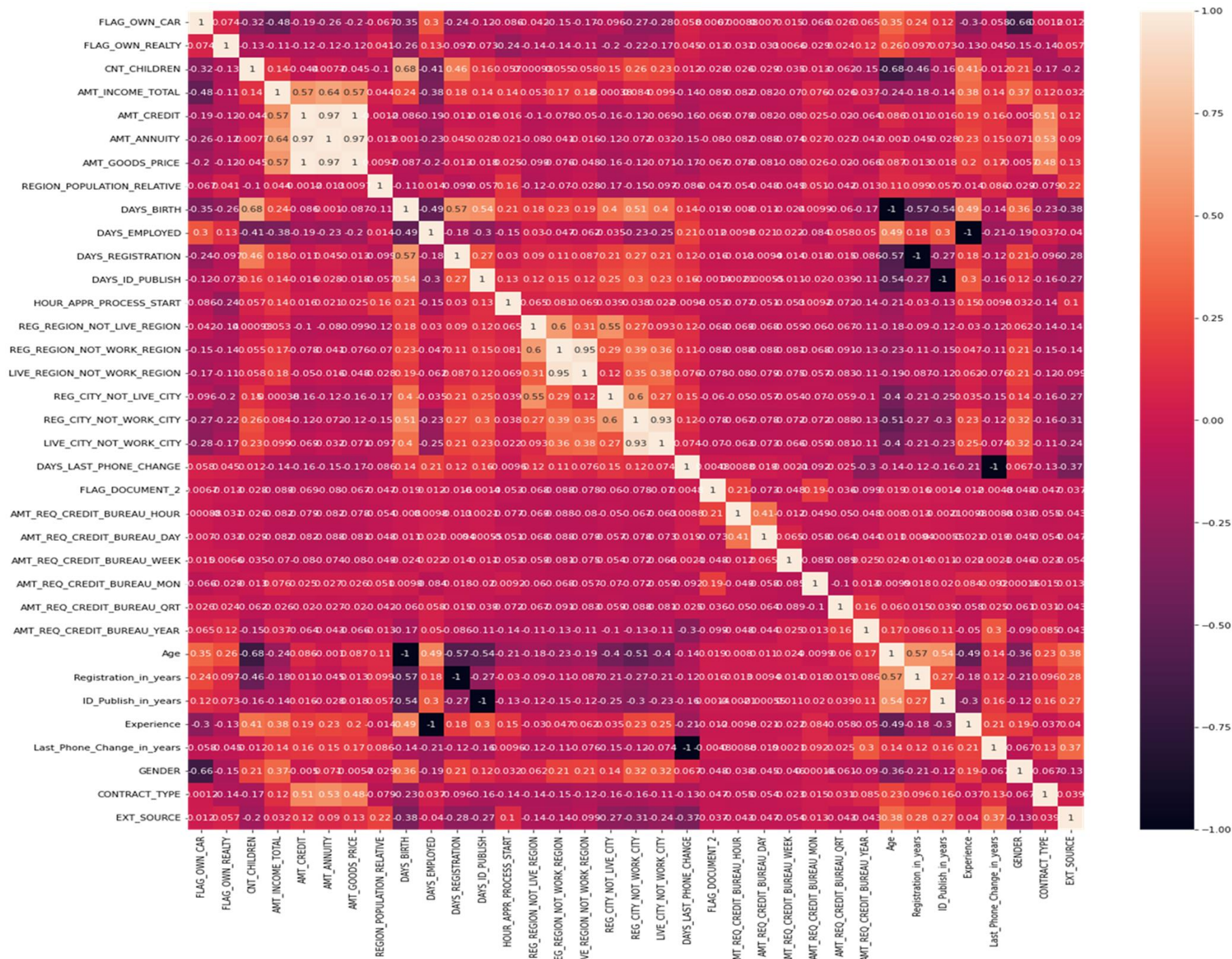
Working people have high chance of getting loan

Distribution on Income



The majority of people who have loans have their own house

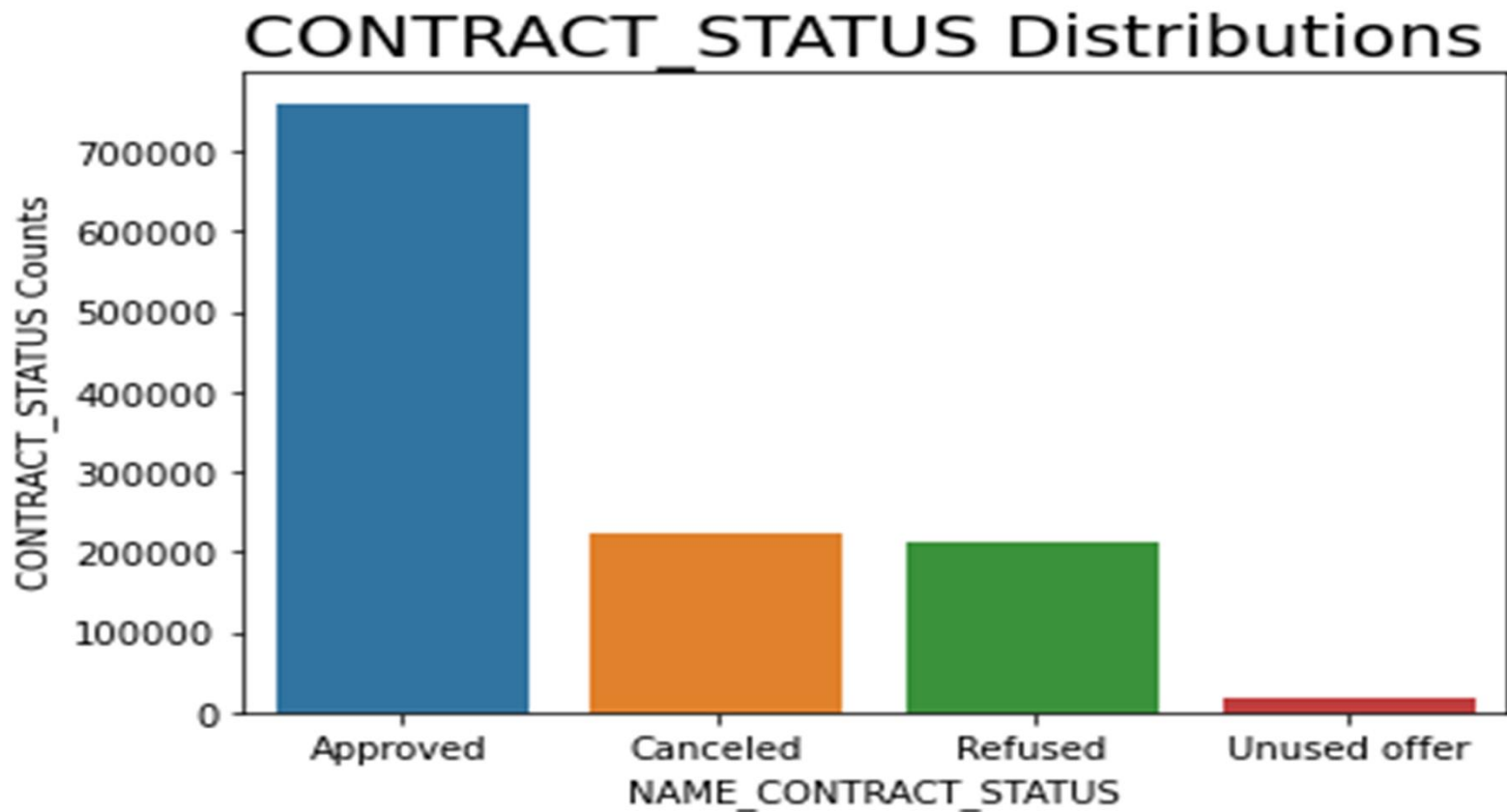
Correlation for Defaulter data frame



Result

1. AMT_GOODS_PRICE and AMT_CREDIT, AMT_ANNUTY and AMT_AMT_CREDIR have perfect positive correlation.
2. AMT_ANNUTY, AMT_CREDIT, and AMT_GOODS_PRICE are highly correlated to each other.

DISTRIBUTION OF CONTRACT STATUS ON MERGED DATASET



There are more approved loans and very less unused offers

Conclusion

Recommended groups where loan can be credited:

- Clients who are working.
- Clients with high income category.
- Clients with higher education.
- Previously accepted loan status group.

Risky groups:

- Lower secondary educated clients.
- Clients with no working background.
- Previously refused loan status group.