# CREDIT EDA CASE STUDY

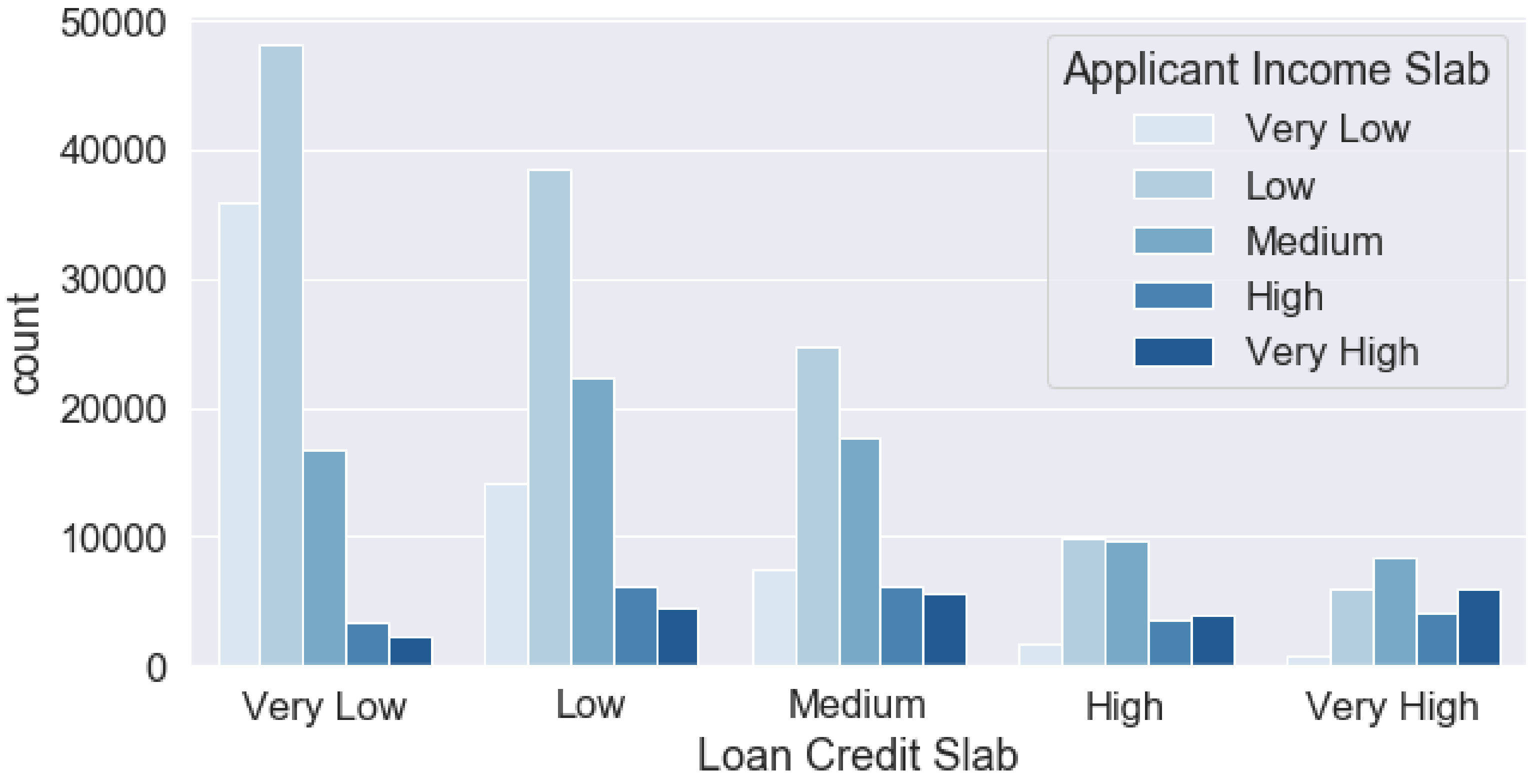
Group Name:
1. Karan Joseph
2. Meghana C

# Business Objectives

- To identify patterns which indicate if a client has difficulty paying their instalments 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.

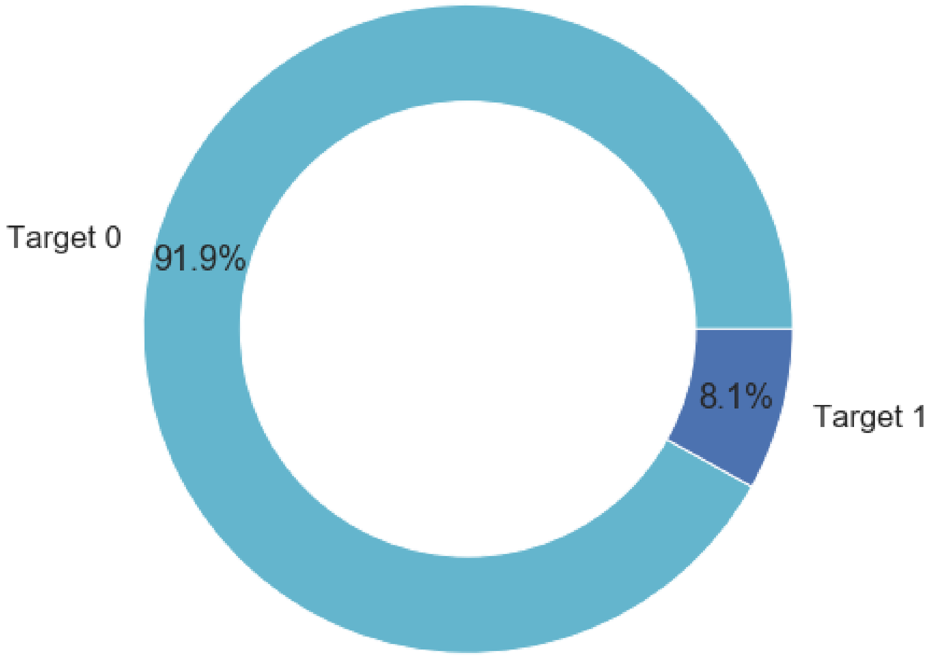
# Applicant Income vs Loan Credit




- Most of loans applied are in the Very Low and Low categories, and the highest applicants in both such cases have low income slabs.
- Most applicants have either low salary or medium salary, followed by people with very low salary.
- Highest applicants in Very High loan category are from the Medium salary slab

# Data Analysis

– Data Imbalance Check:



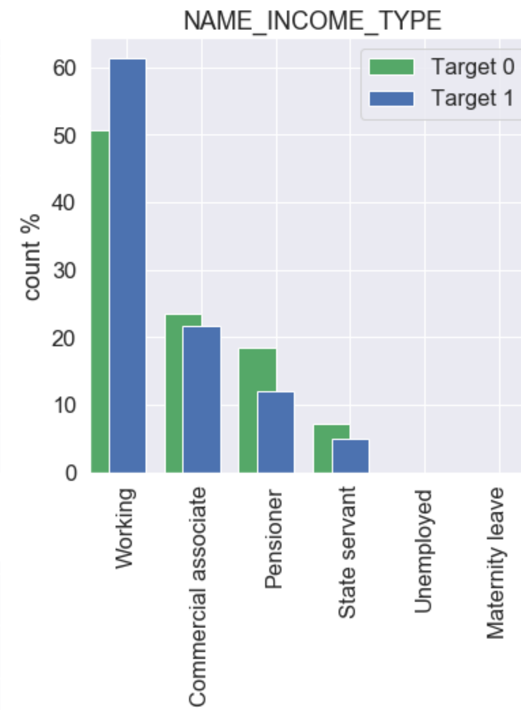

❖ 91.9% of the applicants do have any history late payments.

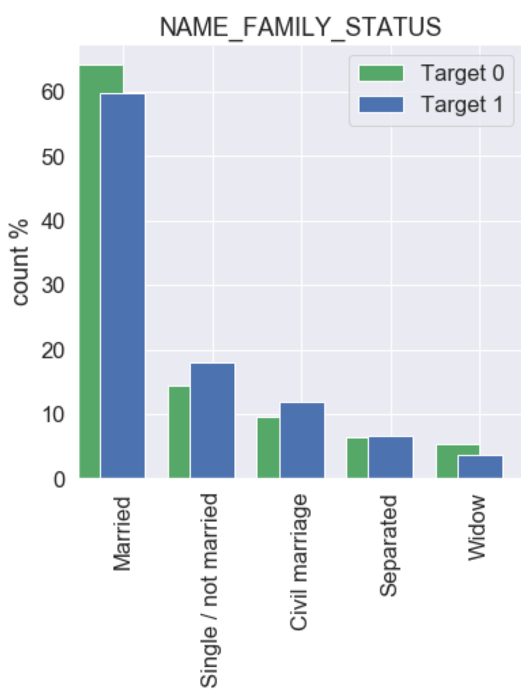
❖ These applicants are potentially capable of repaying the loans.

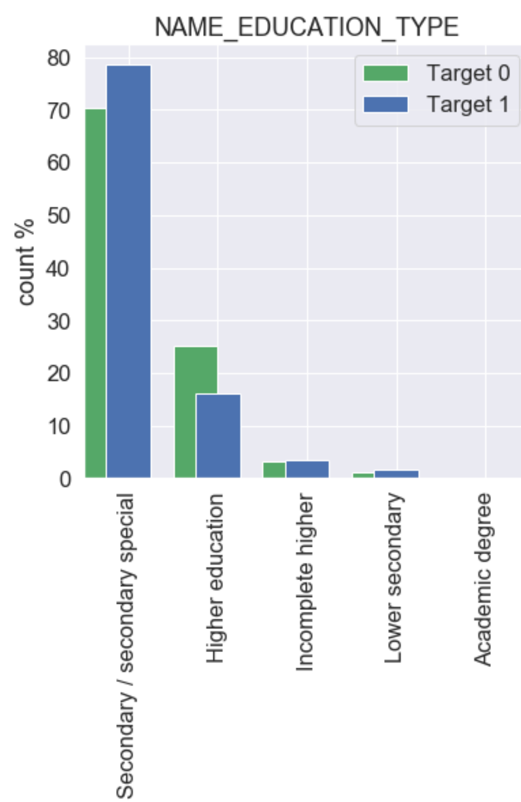
❖ This is **not an indication** of their future repayment capabilities; which spurs the need for further studies on the dataset.

# Univariate analysis for Categorical Variables



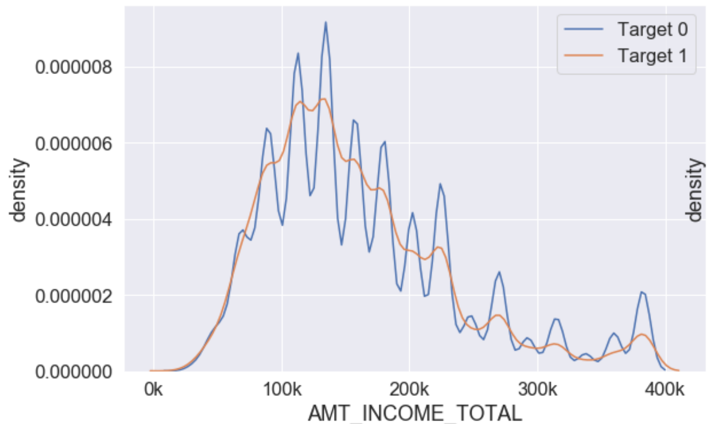


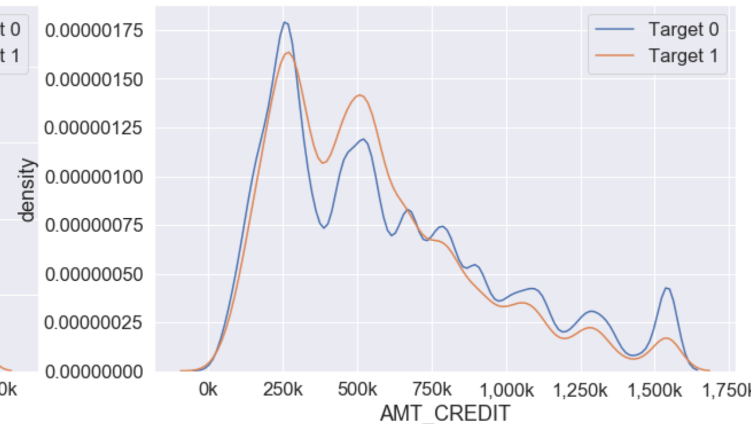



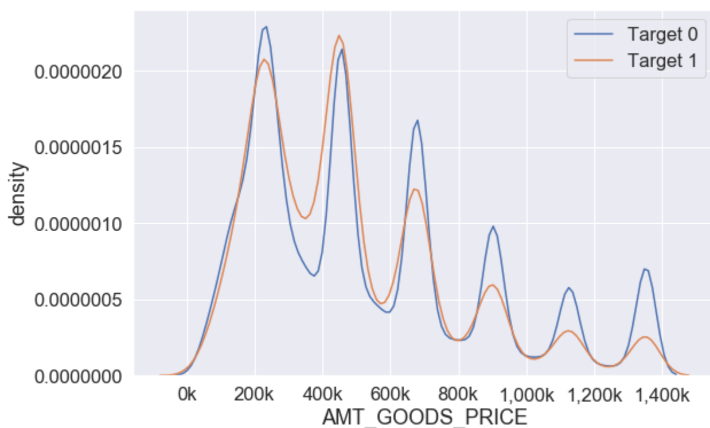



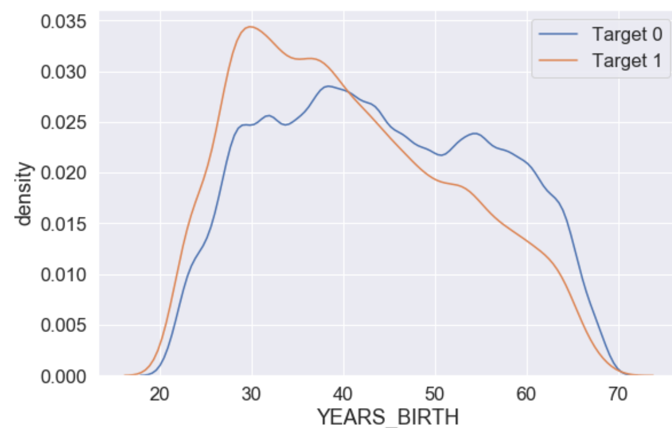

- Majority of the loan applicants are female.
- There are more males having **payment troubles**, even though there are fewer male loan applicants.
- Pensioner defaulters are fewer than non-defaulters.
- Majority of loan applicants are Working class.
- Single applicants have higher percentage of defaulters.
- Married applicants have more non-defaulters than defaulters.
- People Owning Cars have lesser default ratio.
- Majority of the applicants are Married.
- Majority applicants have lesser academic qualifications.
- **Default rate is higher for applicants with just Secondary education.**

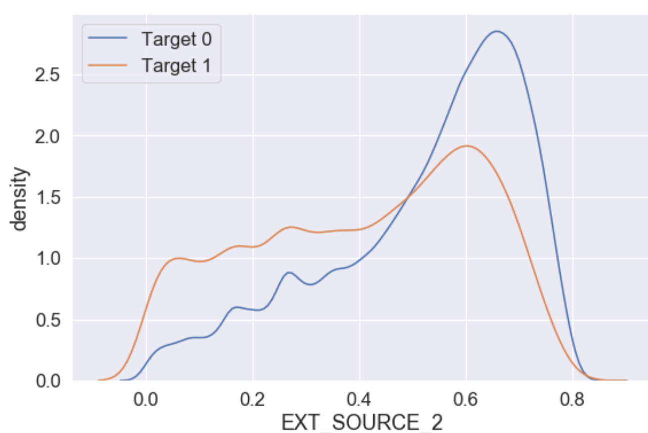
# Univariate analysis for Numerical Variables



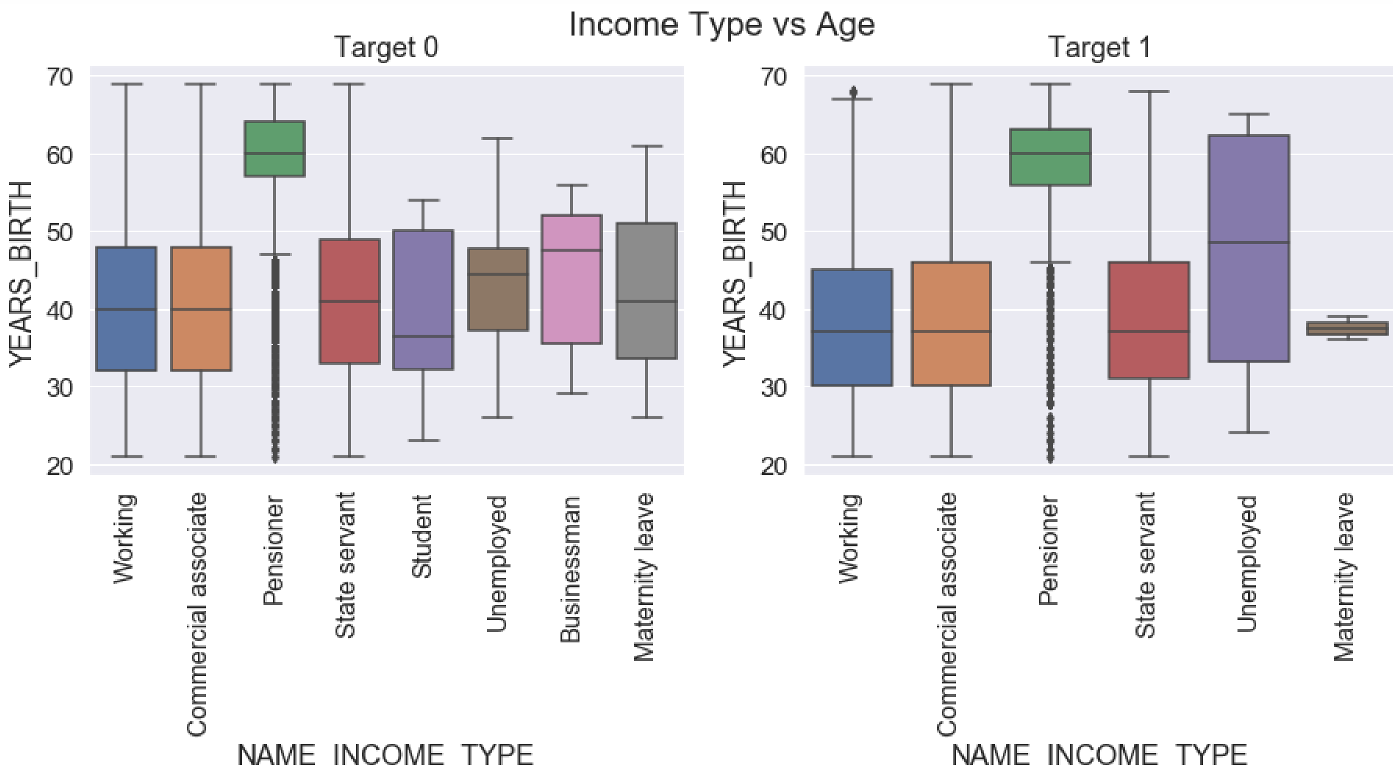



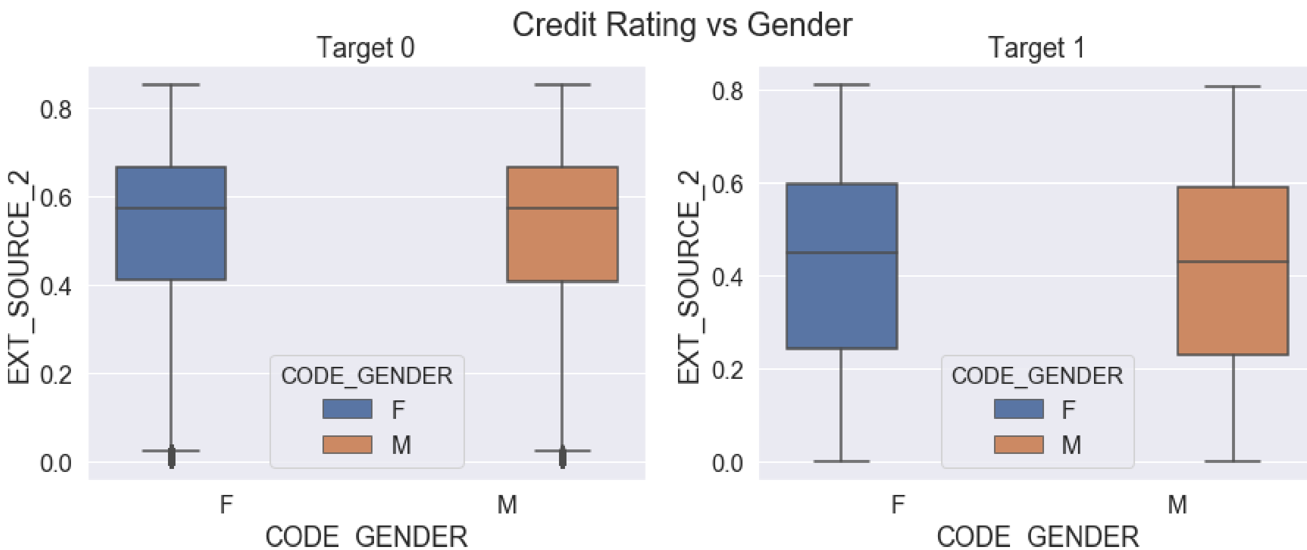

- Defaulters are having **lower credit ratings** than non-defaulters. (This is just as expected)
- Younger applicants have higher loan repayment troubles. This could mean that they have poorer money management skills.
- There are very few higher credit loans.
- Very less applicants with higher income slabs.
- Both defaulters and non-defaulters have almost similar Income-Goods-Credit distribution. But these distributions follow the same trend, which means that they have a strong correlation between them.

# Bivariate Analysis




- From the above plots which shows the bivariate analysis for both Target 0 and Target 1, we can state the following:

- **Businessmen** have highest age cap among all the salaried applicants.
- Giving loans to **Unemployed** applicants with more than 50 years of age is the riskiest( more chance to default).
- Most of the **working** applicants are of the age group 30-50 years.



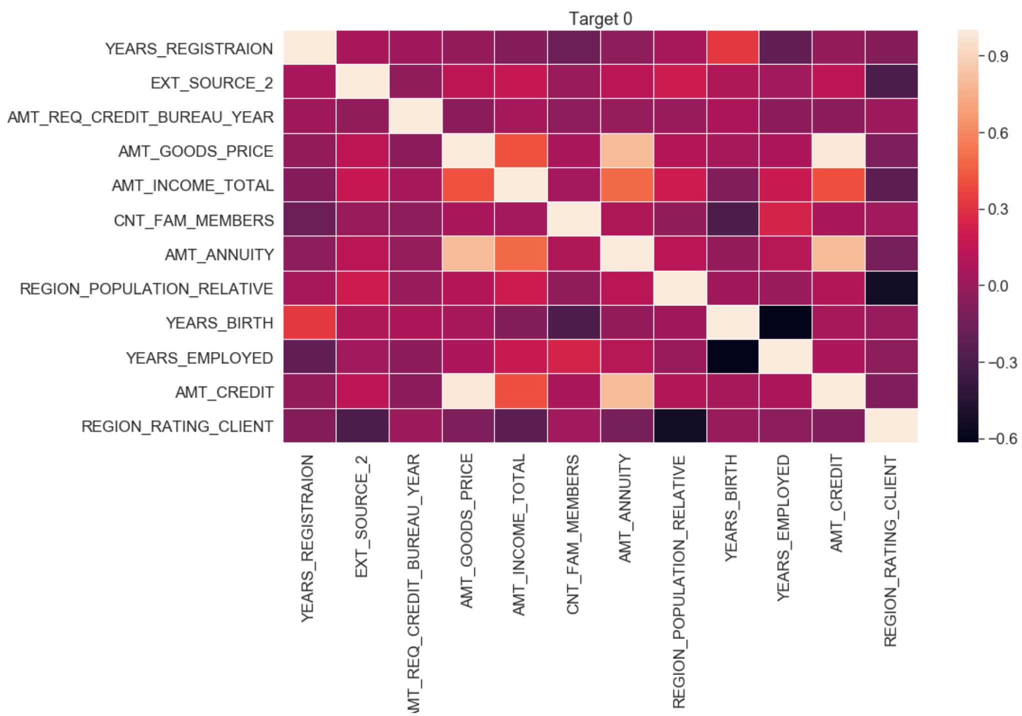

## ❖ Credit Ratings vs Gender

Credit rating does not depend on the gender of the applicant, but on their loan repayment capabilities; as evident from the box plots.


## ❖ Job vs Credit Ratings

- **Businessmen have the highest credit ratings.**
- Unemployed applicants with less than 0.4 credit rating are very risky candidates.
- Women on Maternity leave, but with less credit ratings are risky applicants.

# Correlation Matrix

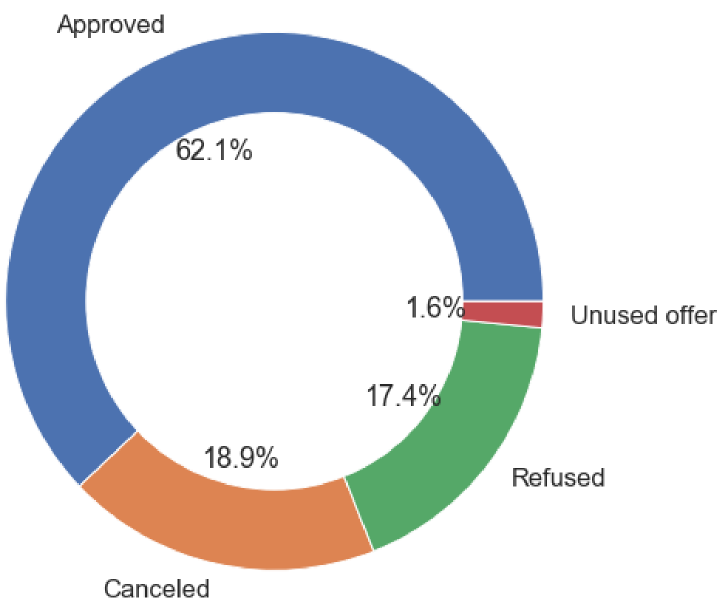



| VAR1 | VAR2 | CORR |
|---|---|---|
| AMT_CREDIT | AMT_GOODS_PRICE | 0.986680 |
| AMT_ANNUITY | AMT_GOODS_PRICE | 0.799405 |
| AMT_CREDIT | AMT_ANNUITY | 0.796758 |
| YEARS_EMPLOYED | YEARS_BIRTH | 0.617763 |
| REGION_RATING_CLIENT | REGION_POPULATION_RELATIVE | 0.539005 |
| AMT_ANNUITY | AMT_INCOME_TOTAL | 0.489586 |
| AMT_INCOME_TOTAL | AMT_GOODS_PRICE | 0.416020 |
| AMT_CREDIT | AMT_INCOME_TOTAL | 0.411285 |
| YEARS_BIRTH | YEARS_REGISTRAION | 0.332980 |
| REGION_RATING_CLIENT | EXT_SOURCE_2 | 0.291624 |

- **Correlation matrix is not the absolute method to distinguish between which variable has the highest impact in identifying who will default or not.**

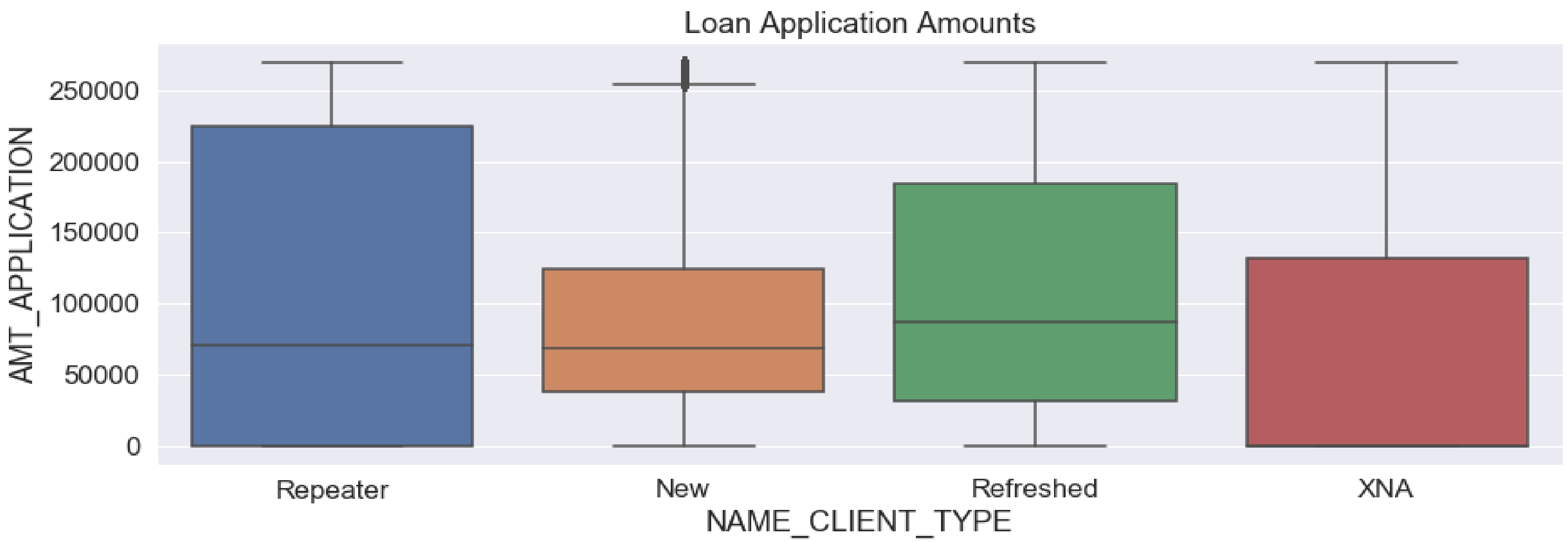
But it can help highlight some interesting interactions:

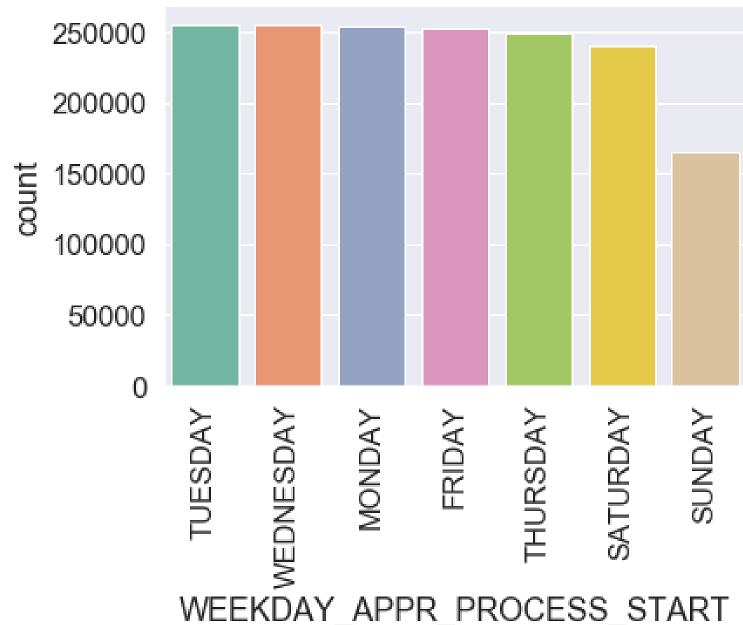
- Loan Credit amount is highly correlated with the value of the collateral goods. The amount of loan you get is directly dependent on the value of the goods one submits.
- Annuity has good correlation with Loan Amount. Hence as loan amount increase, annuity amount will also increase.
- Age of the employee and years employed has a moderate correlation.
- High populous regions have better region rating.
- Annuity amount has a decent correlation with the income of the applicant. Applicants with more income might repay the loan at higher annuity amounts and lesser periods.

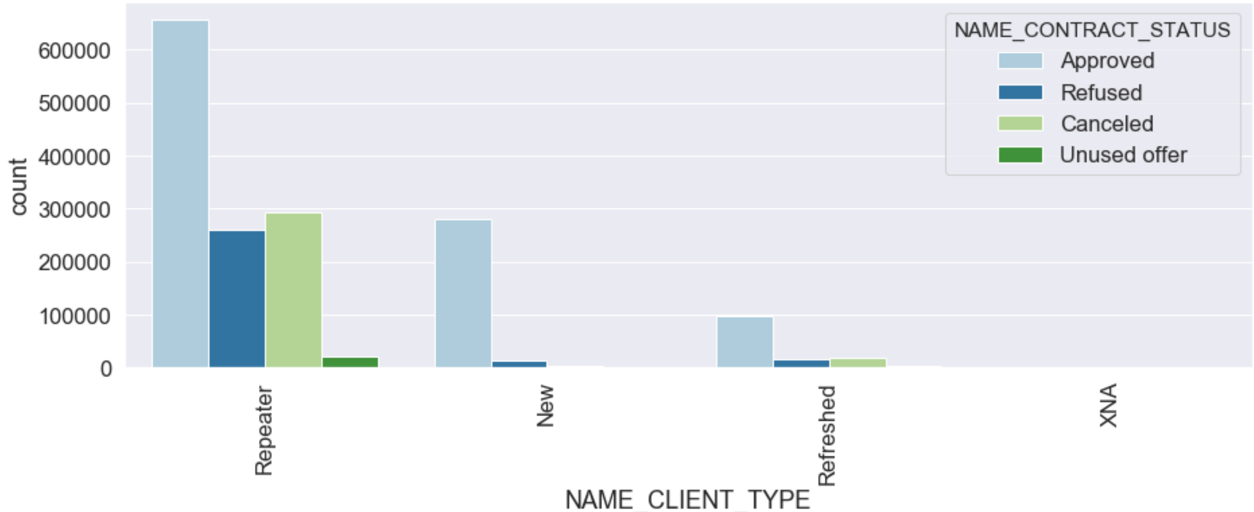



62% of the previous applications were approved and only 17% of the loan applicants were refused.

- **Repeater clients and Refresher clients apply for the biggest loans.**
- New clients generally apply for smaller amounts.




- Most loan applications occur along Tue, Wed, Mon, Fri and Thurs. Hence, weekdays are much busier than weekends.

- **Cash loans have the highest rejection and cancellation chances**. Hence, most of the applied loans are consumer loans. They have the least chance of cancellation; and only a small percentage of consumer loans are refused.

- **New clients have the lowest loan refusal rate**. However, majority of the loan applicants are Repeaters; and they have the highest chance of loan cancellation. Repeater clients have a nearly 40% chance that their application will get rejected.

# Conclusion

- Some of the inferences that can be made based on our analysis are:
- Giving loans to Unemployed applicants with more 50 Years of age is the riskiest.
- There are more youngsters having higher payment issues.
- Default rate is higher for applicants with just Secondary education
- Single/ not married defaulters are more than non-defaulters.
- Repeater and Refresher clients apply for the biggest loans.
- New clients generally apply for smaller amounts.
- Repeater clients has nearly 40% chance of application getting rejected.
- Consumer loans have the least chance of getting cancelled.
- From these inferences, the loan providing companies can minimize the risk of losing money while lending to customers.