# Insights- Credit EDA Case Study

NEERAJ SAH

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. Use EDA to analyze the patterns present in the data.

## PROBLEM STATEMENT

# Risks associated

When the company receives a loan application, the company must decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company.

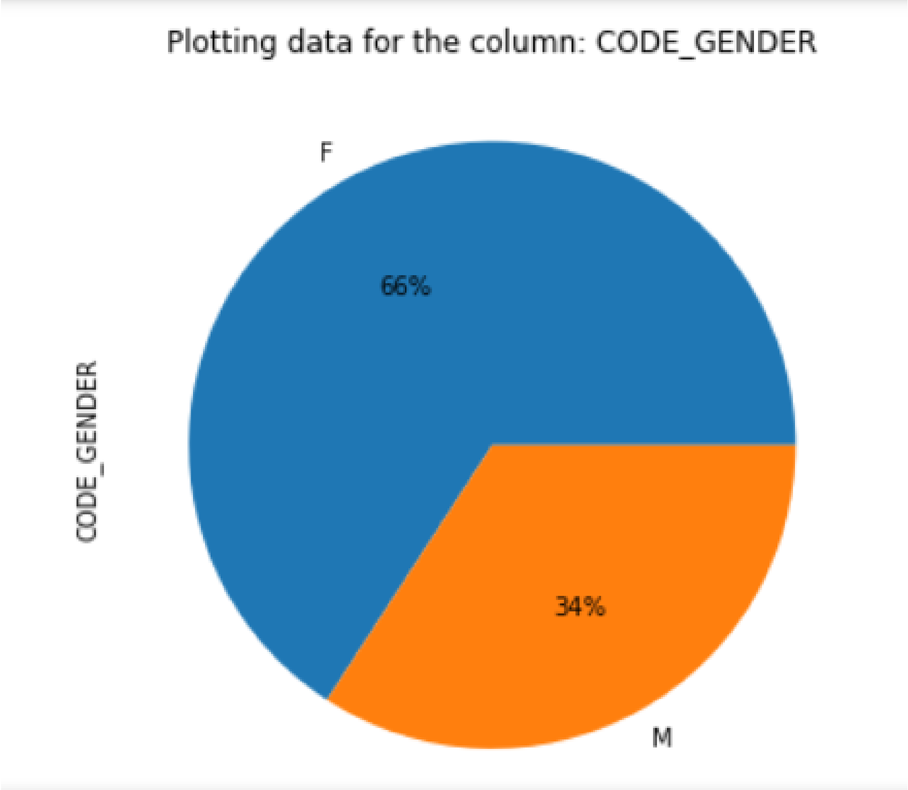
# There are two types of applicants

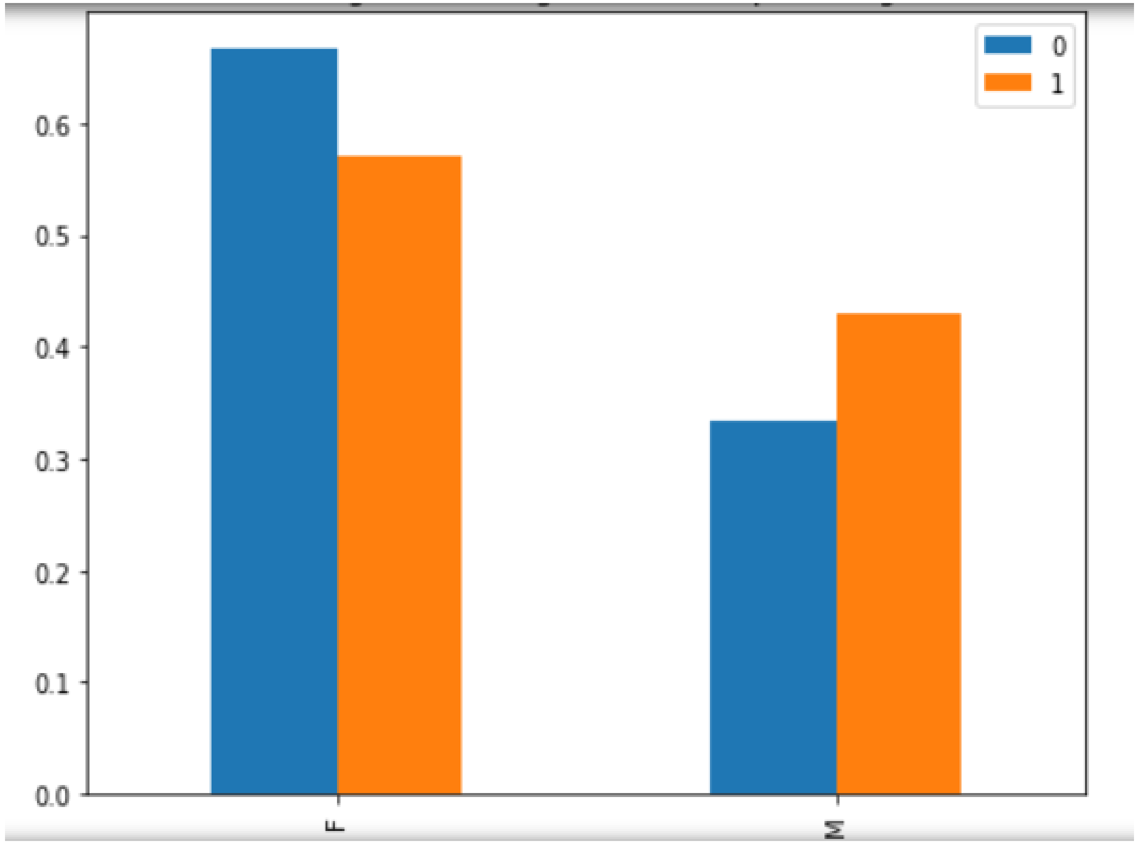
**The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,

**All other cases:** All other cases when the payment is paid on time.

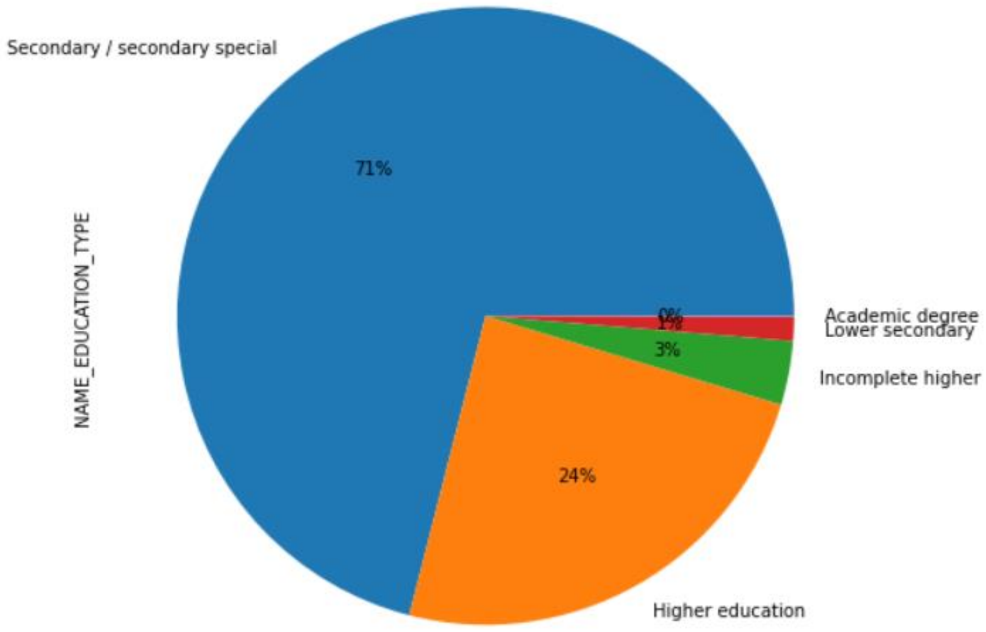
# Univariate Analysis

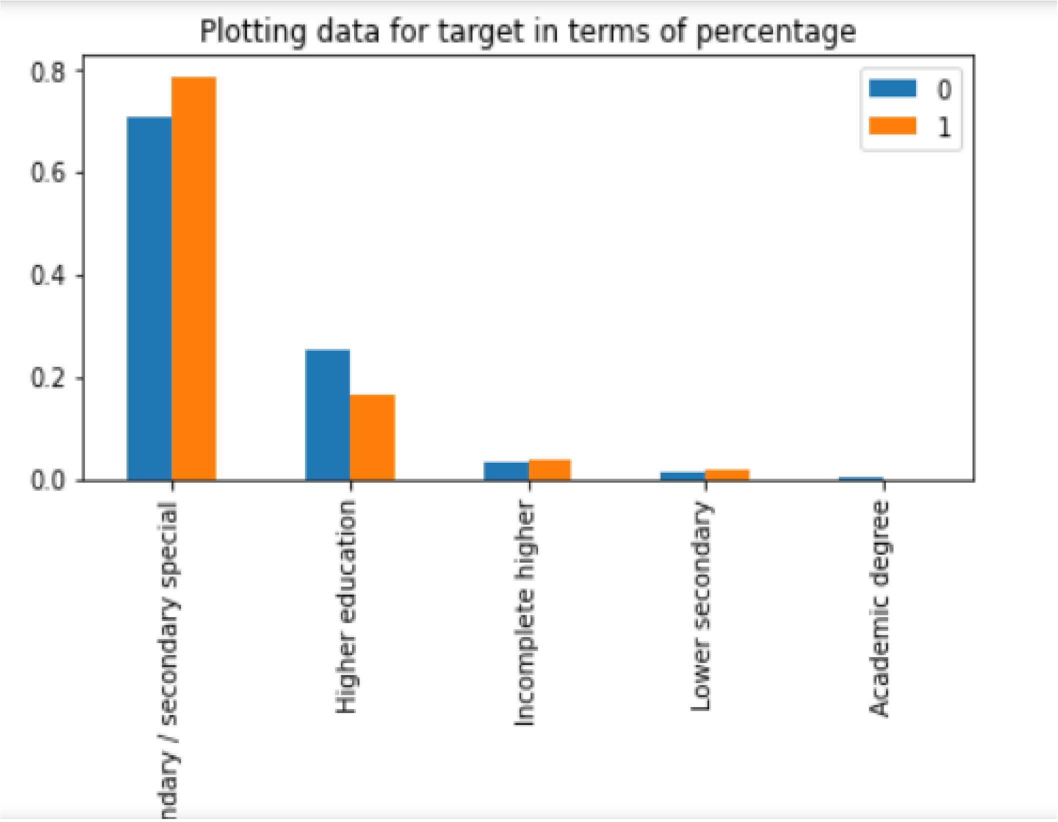
Females are taking more loans as compared to males that can be seen in the pie-chart for CODE_GENDER. But more male are defaulters as compared to female as usual which can be inferred by bar chart of percentages.



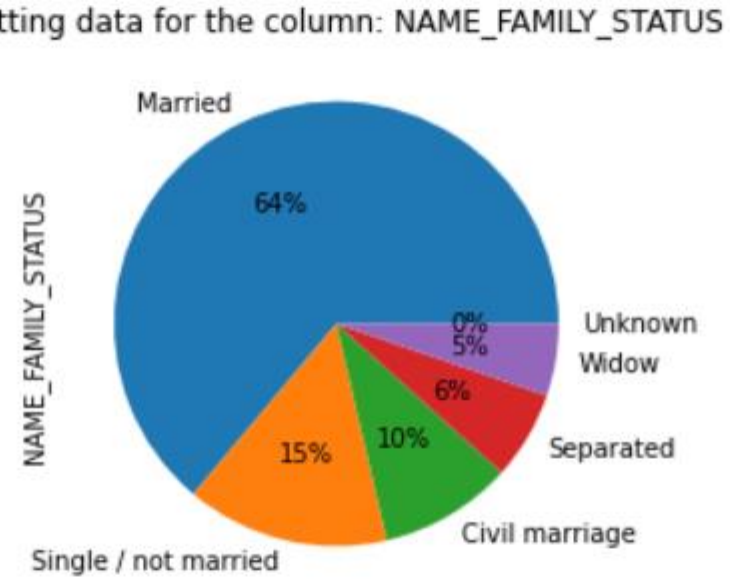



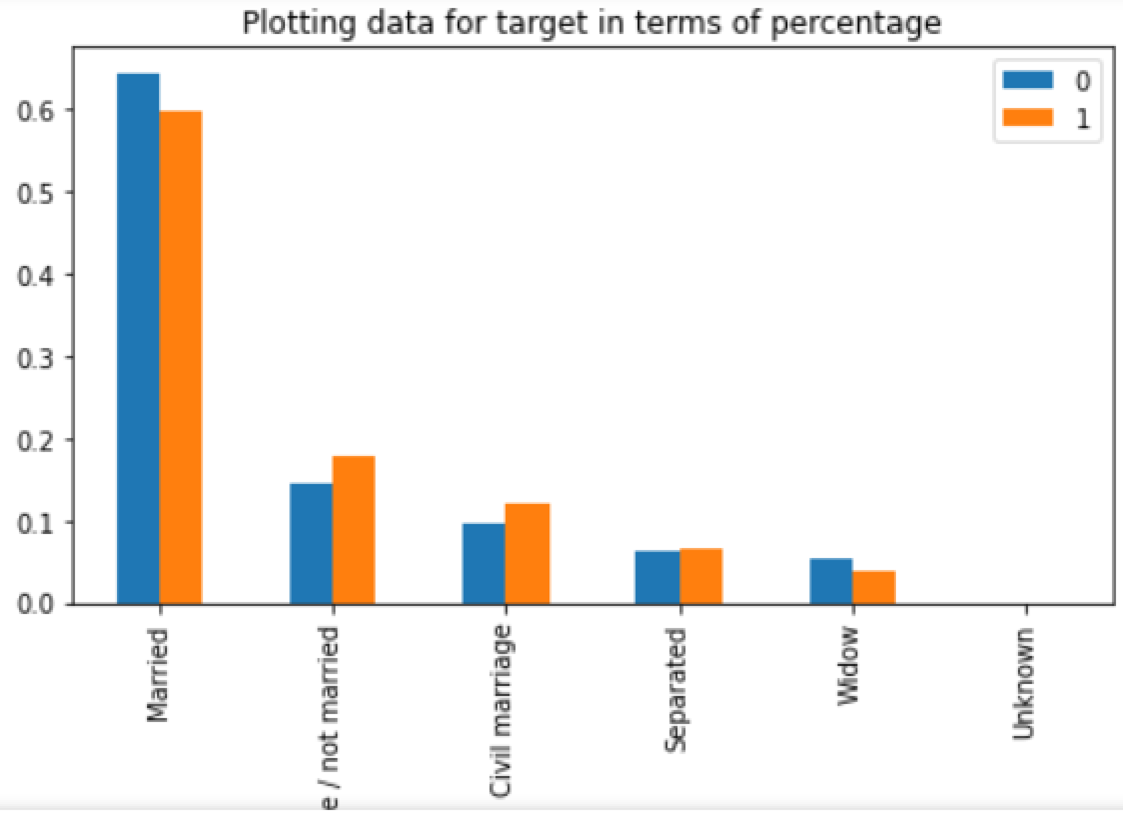
People with secondary education takes more loan and default more as compared to other education levels.



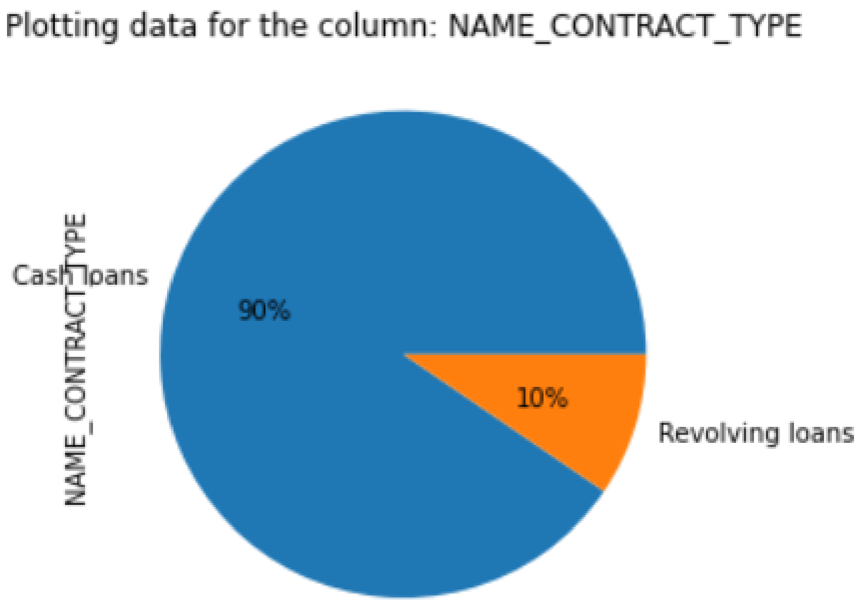



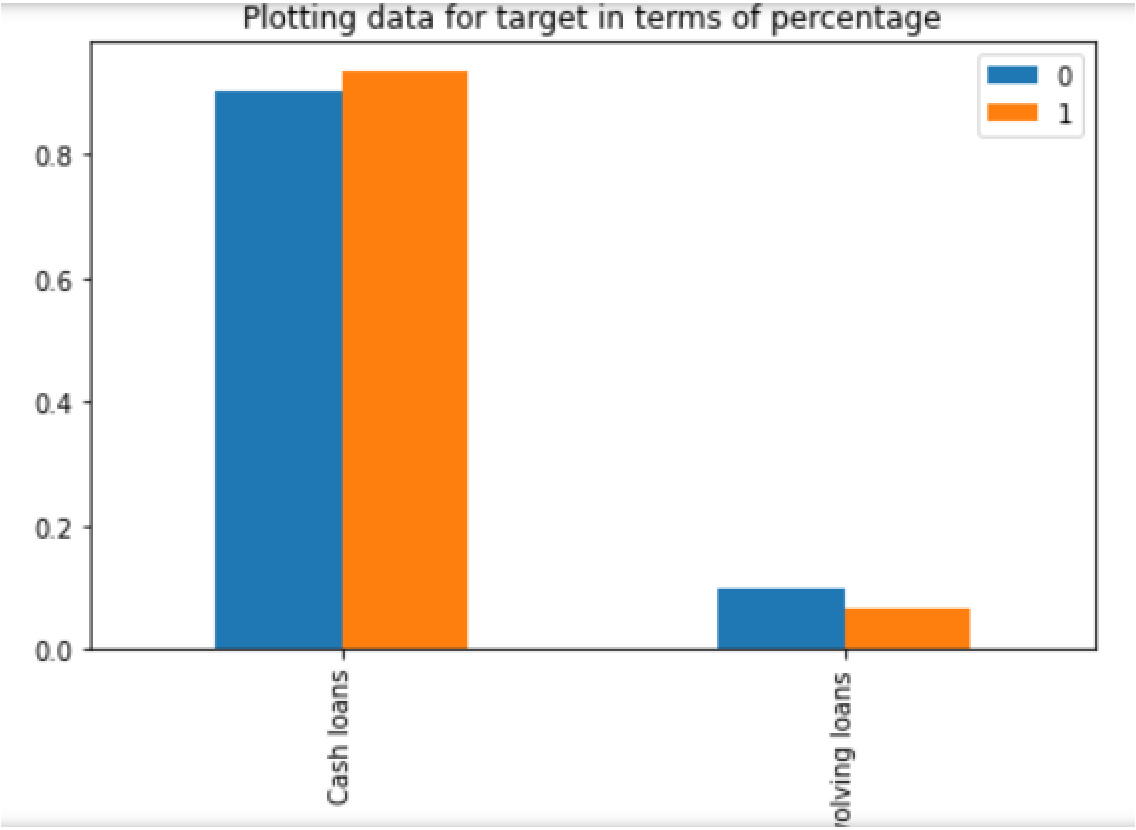
## Married people take more loans but single/not married and civil married have more defaulters.



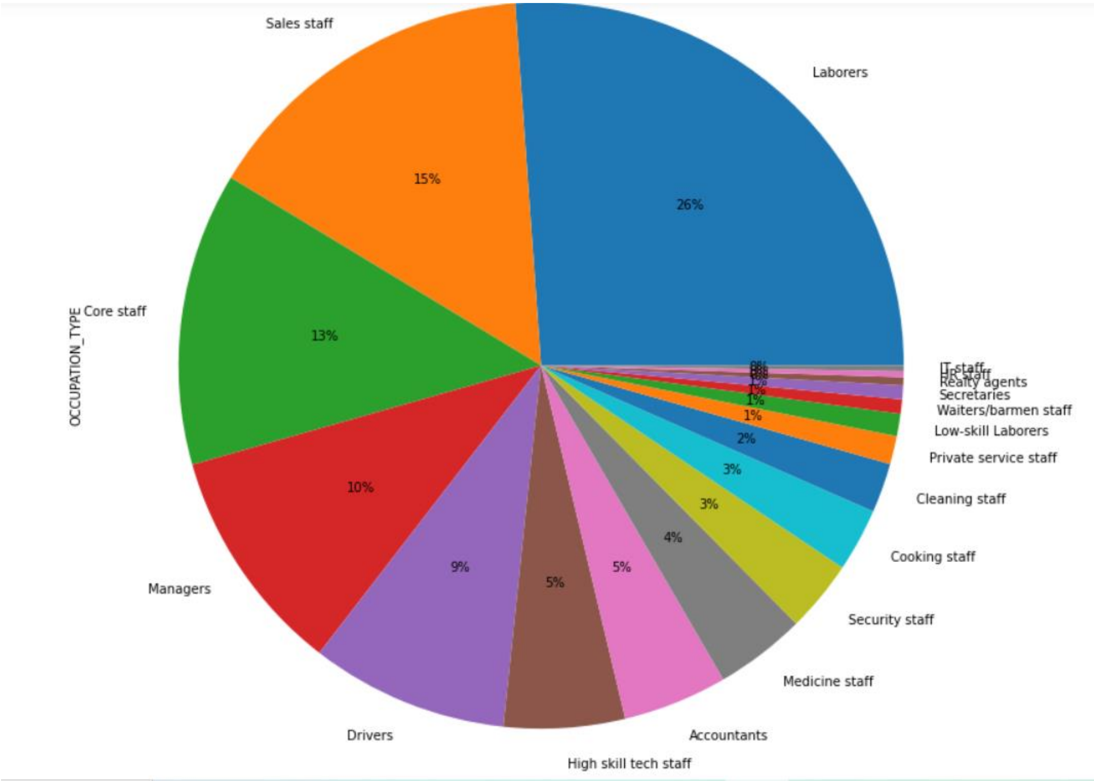



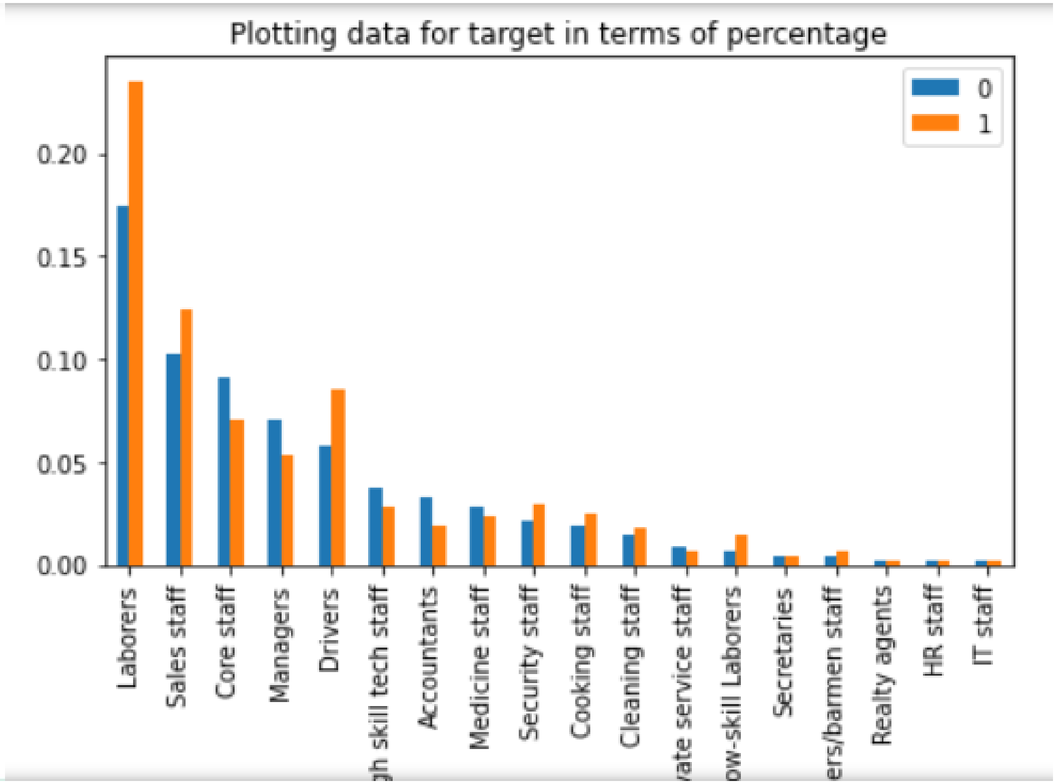
Cash loans are most favourite type of loans and hence sees more defaults.



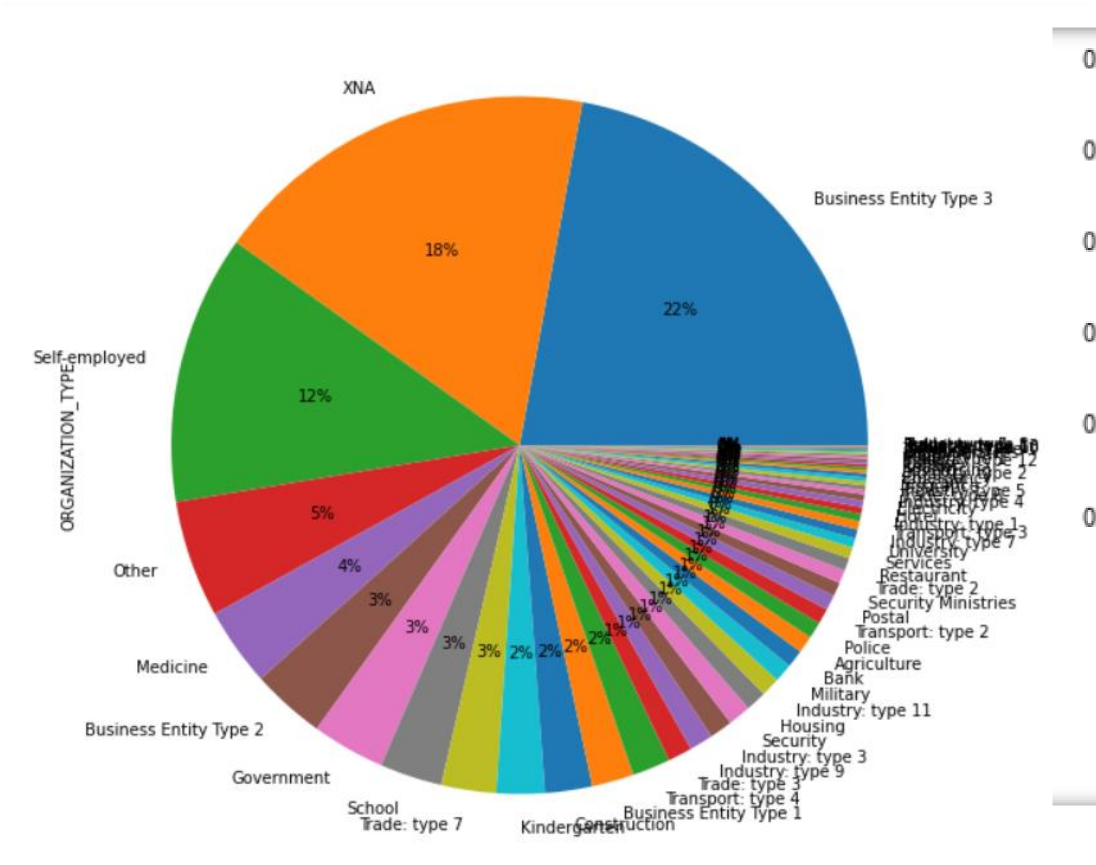



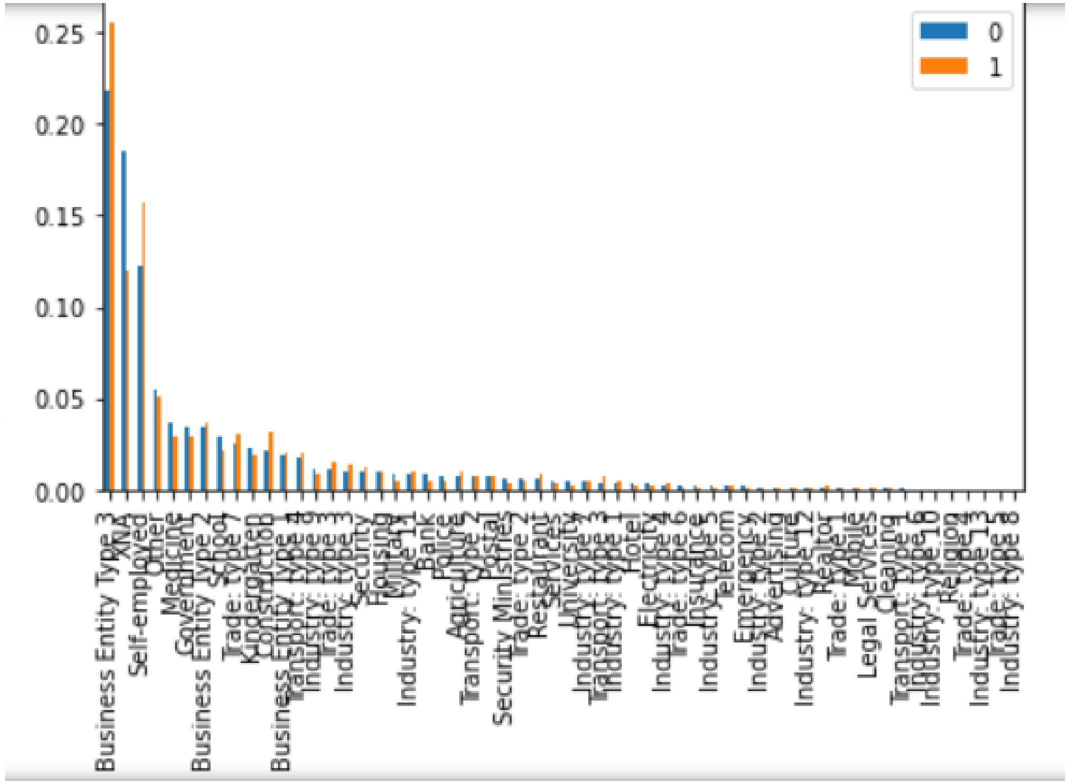
## Laborers are the highest and IT staffs are the least in terms of taking loan. And laborers, drivers and sales staff have highest number of defaulters



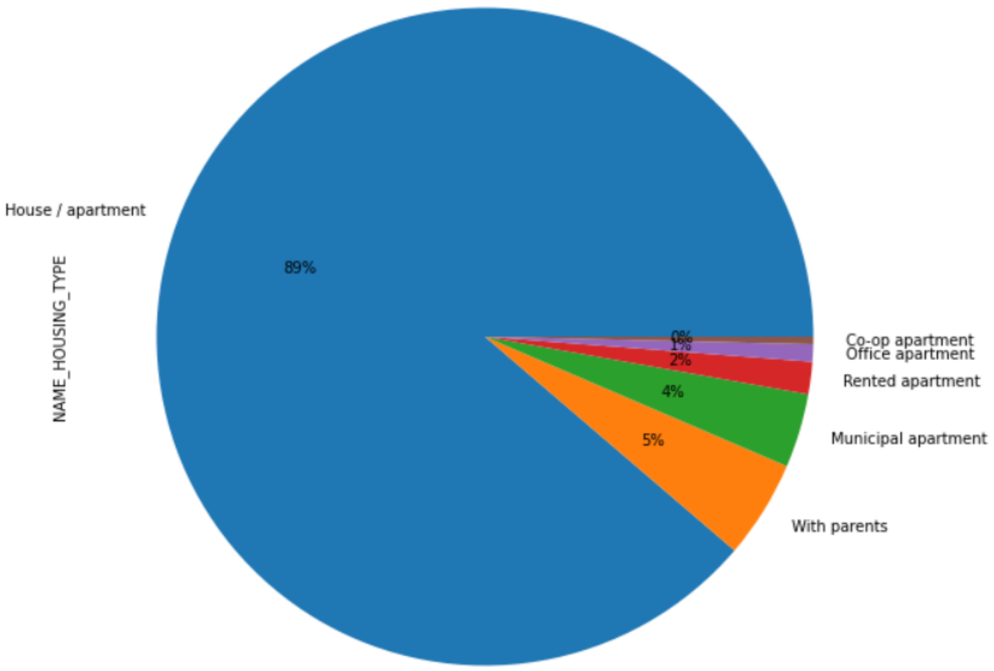



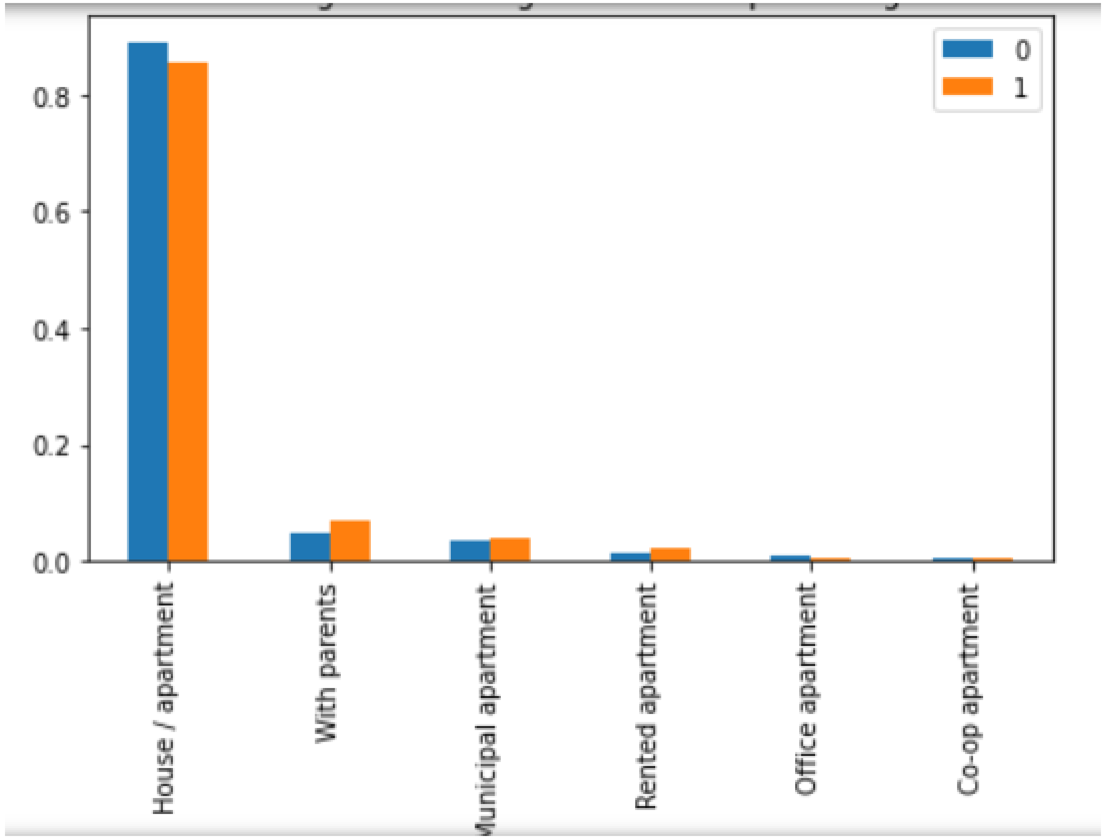
# Business entity type 3 and self-employed people take more loans and defaults more.



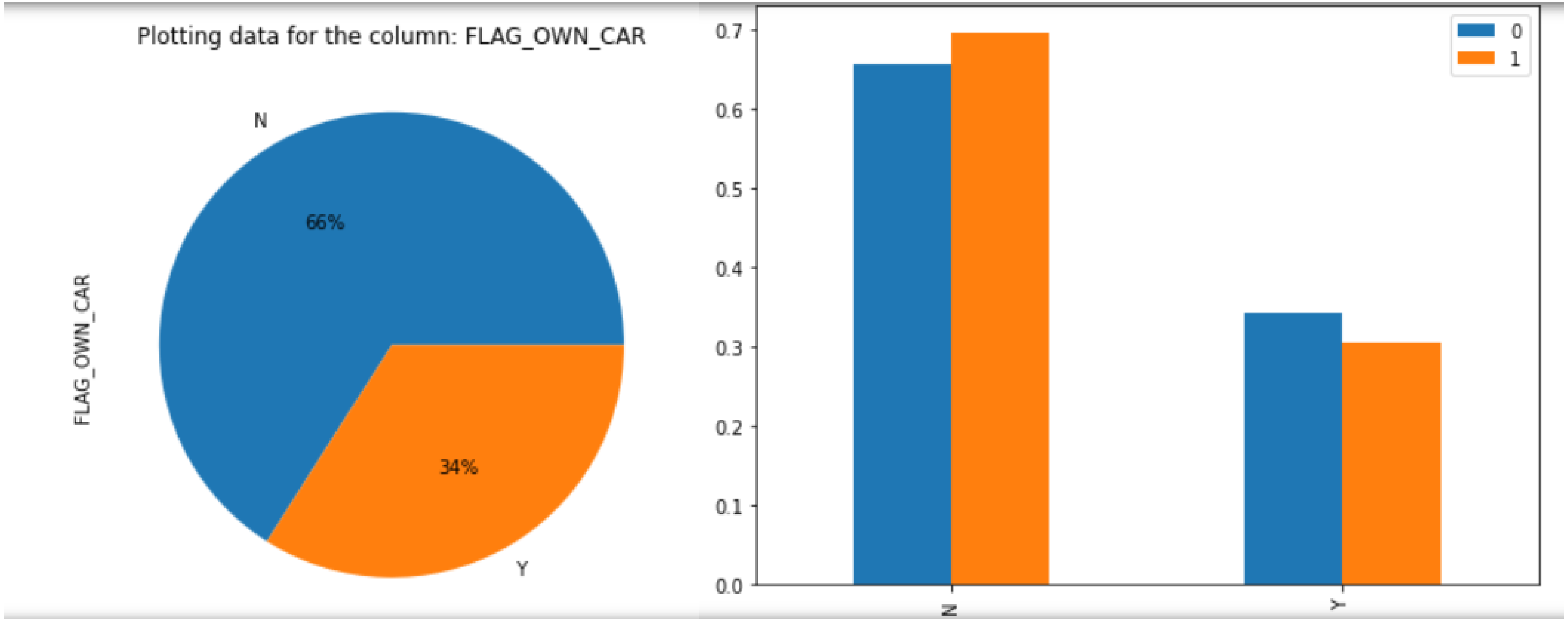



People who own house/apartment are more prominent to take loans but people who live with parents or in rented apartments have more defaulters.

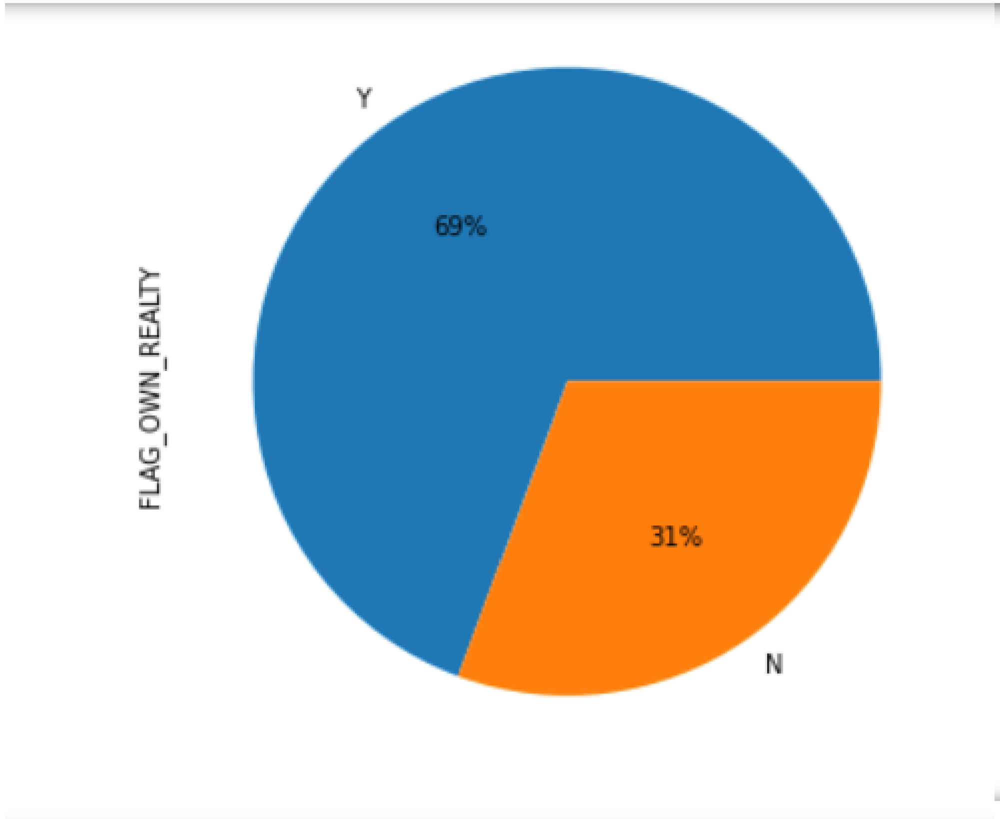


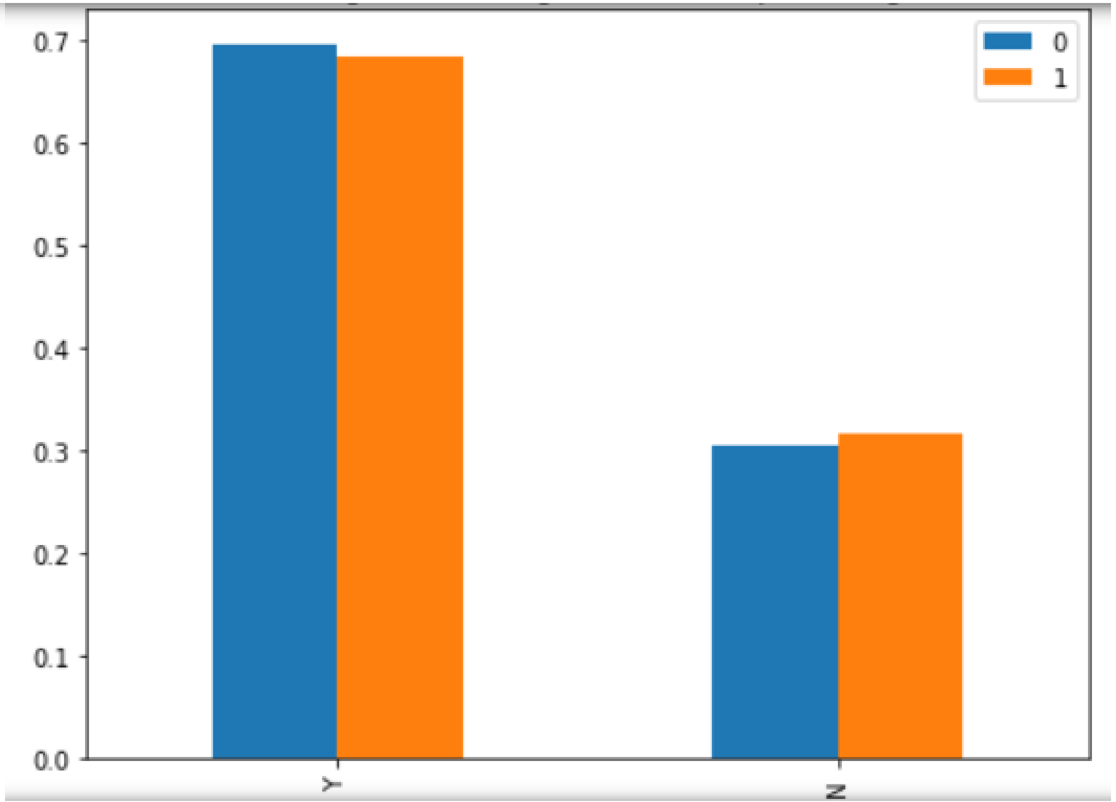



People who do not own cars are more prone to take loans and default.



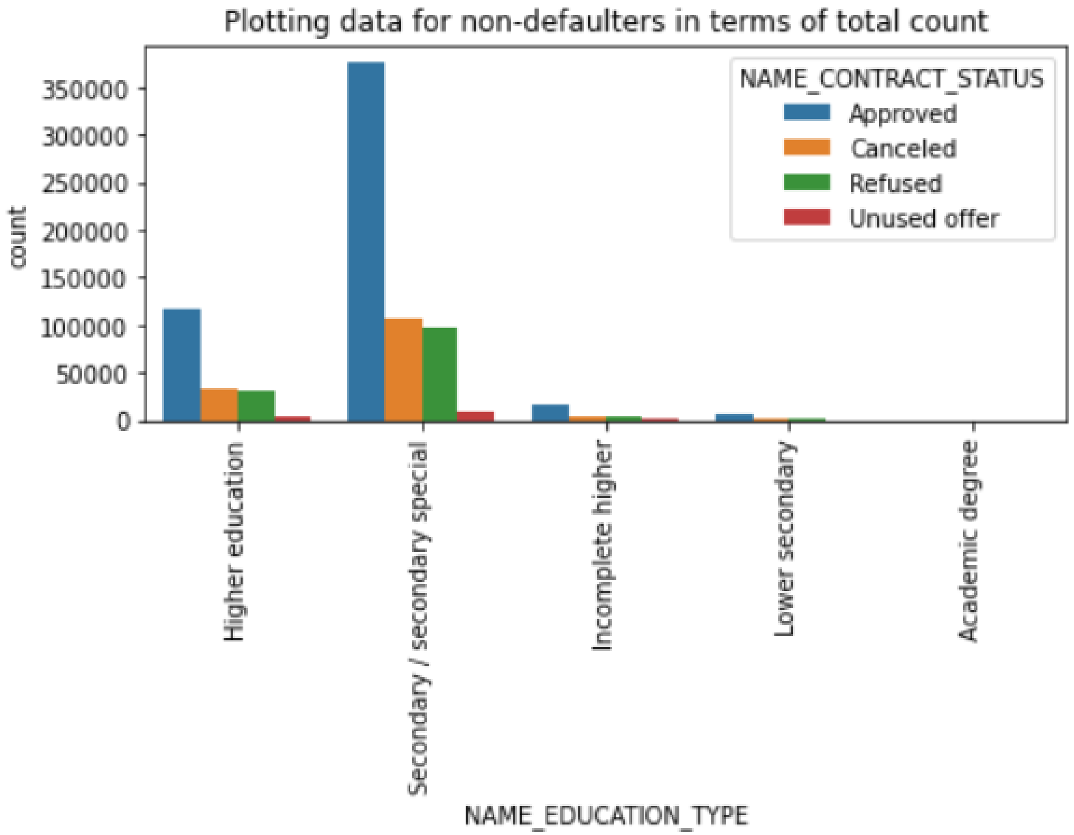
People who do not own property are more prone to take loans and default.

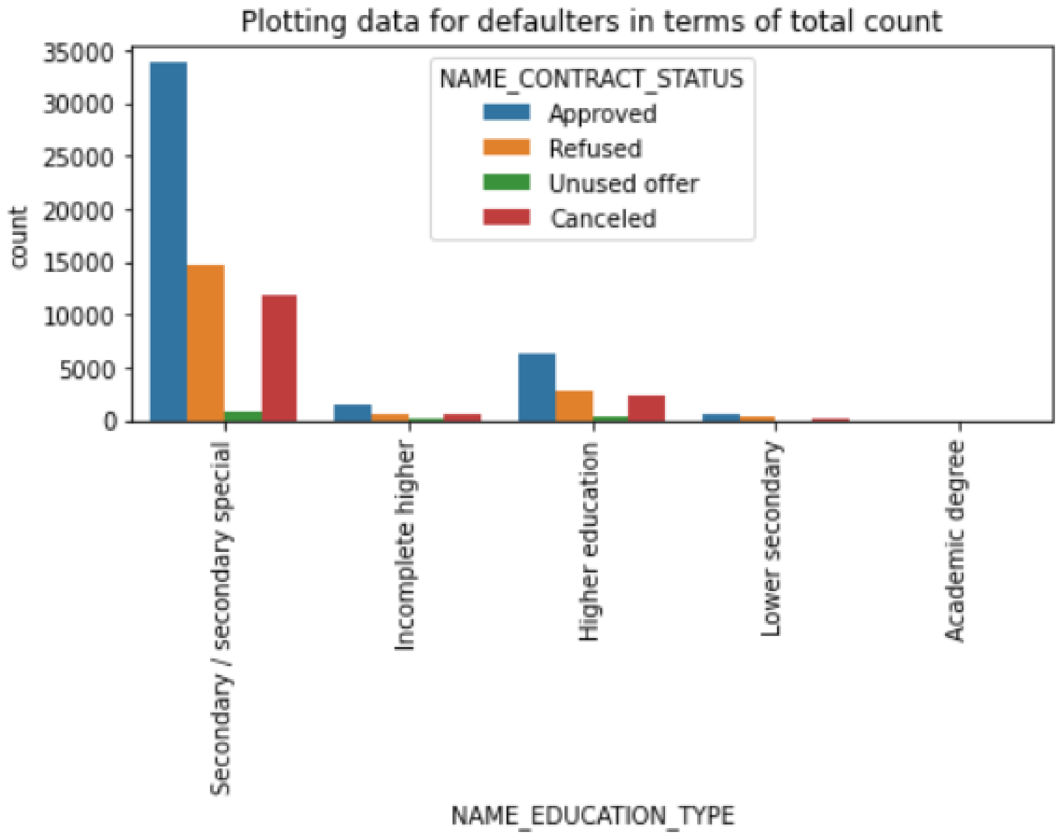


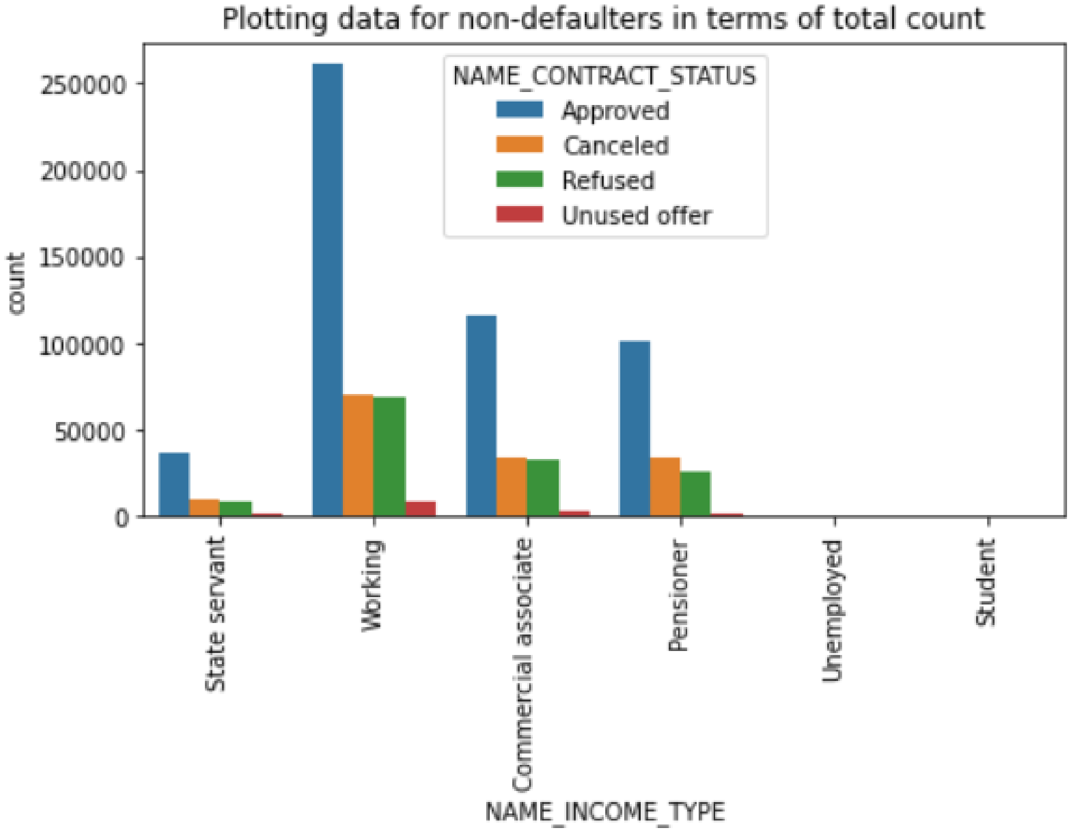
# Bivariate Analysis

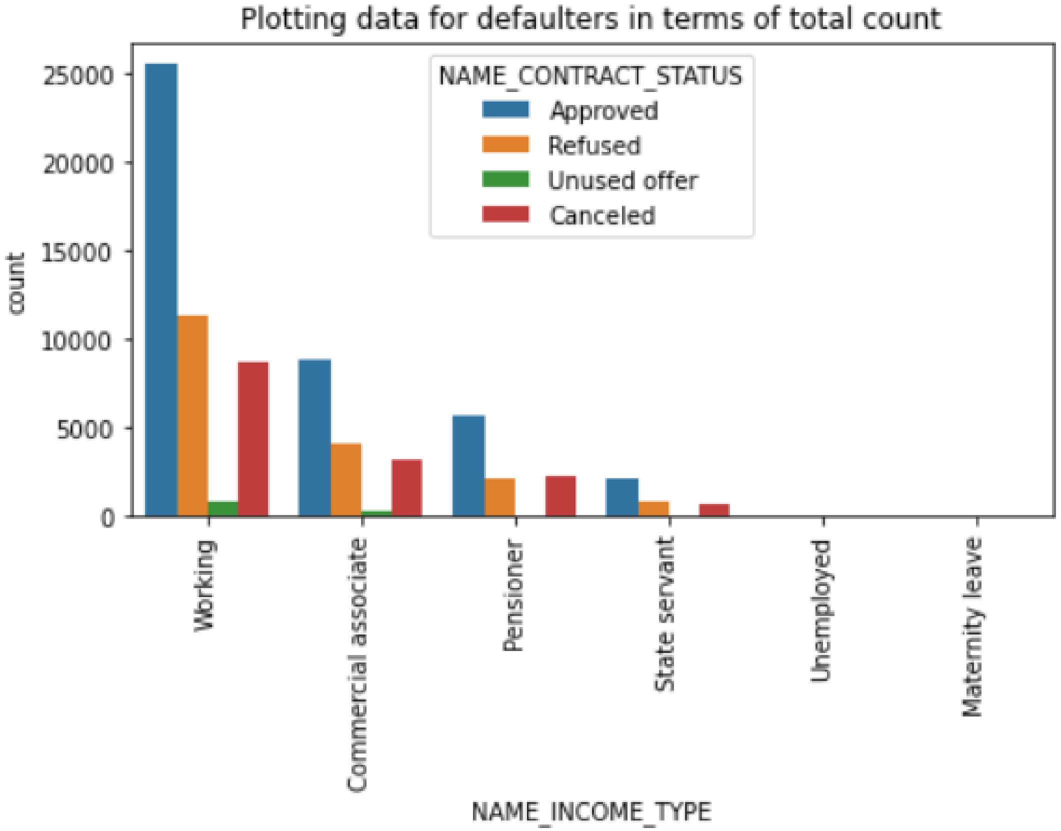
# EDUCATION_TYPE - CONTRACT_STATUS: In defaulter category, secondary and higher educated people are cancelling more applications while in non-defaulter category, more application are refused for higher and secondary education.



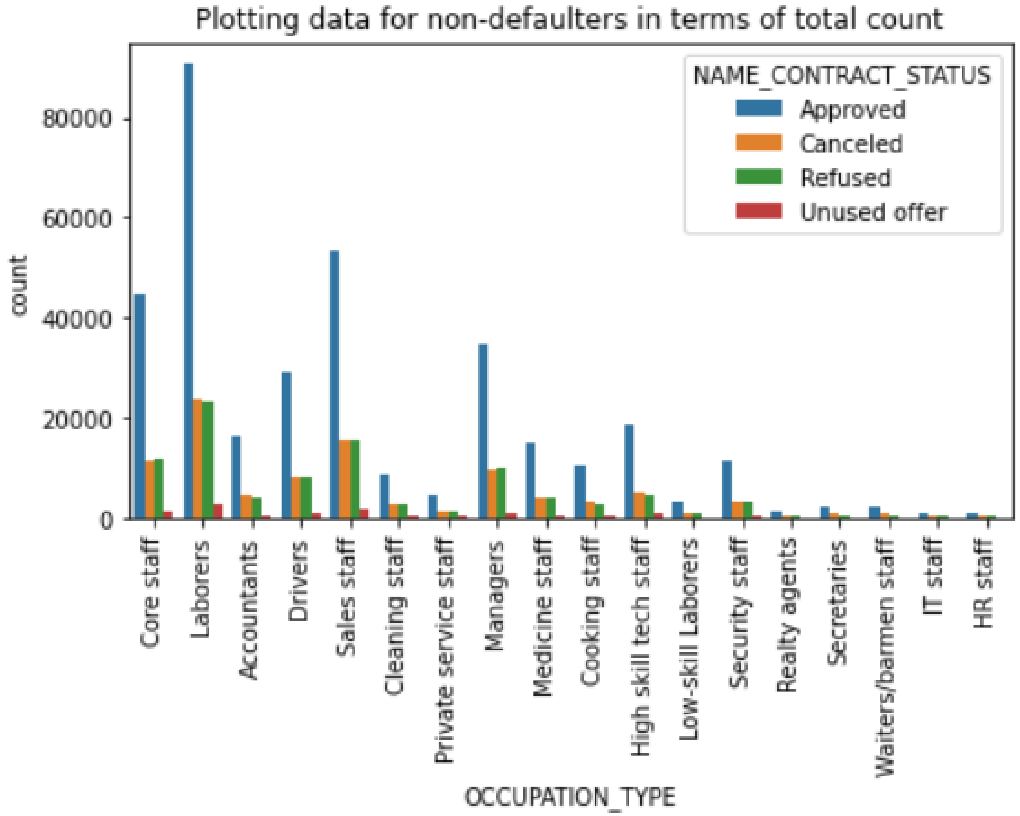



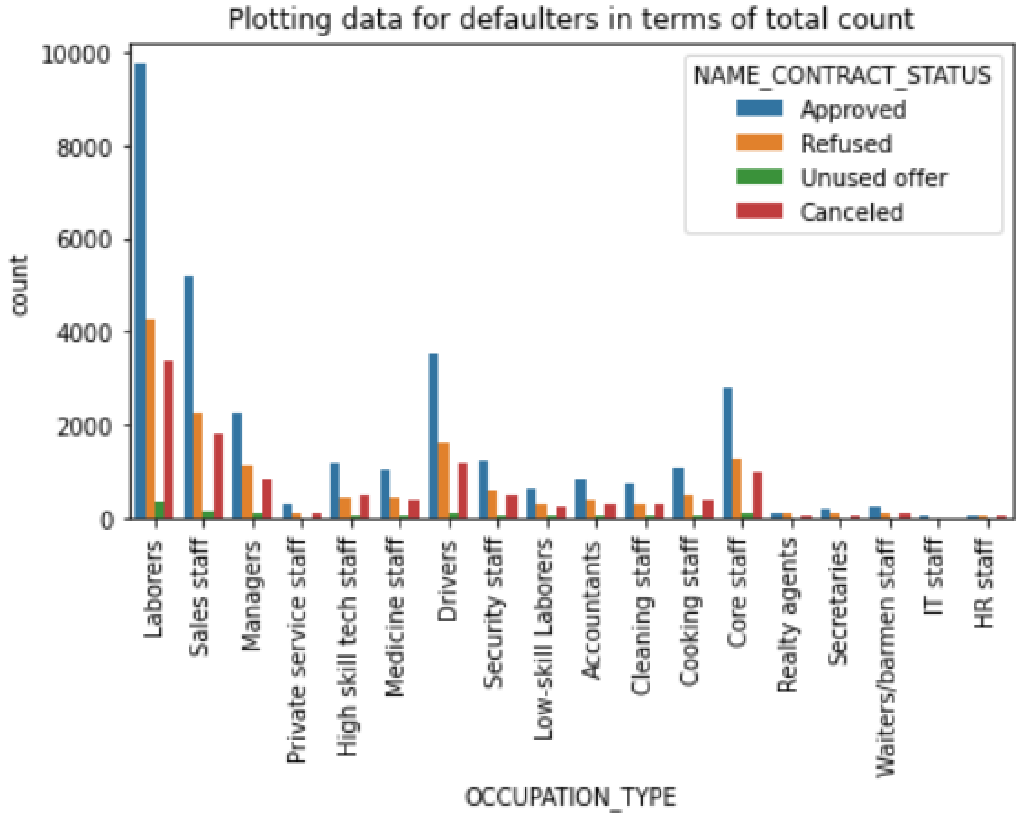
# INCOME_TYPE - CONTRACT_STATUS: Working professionals have more approved and refused applications in both defaulter and non-defaulter category.



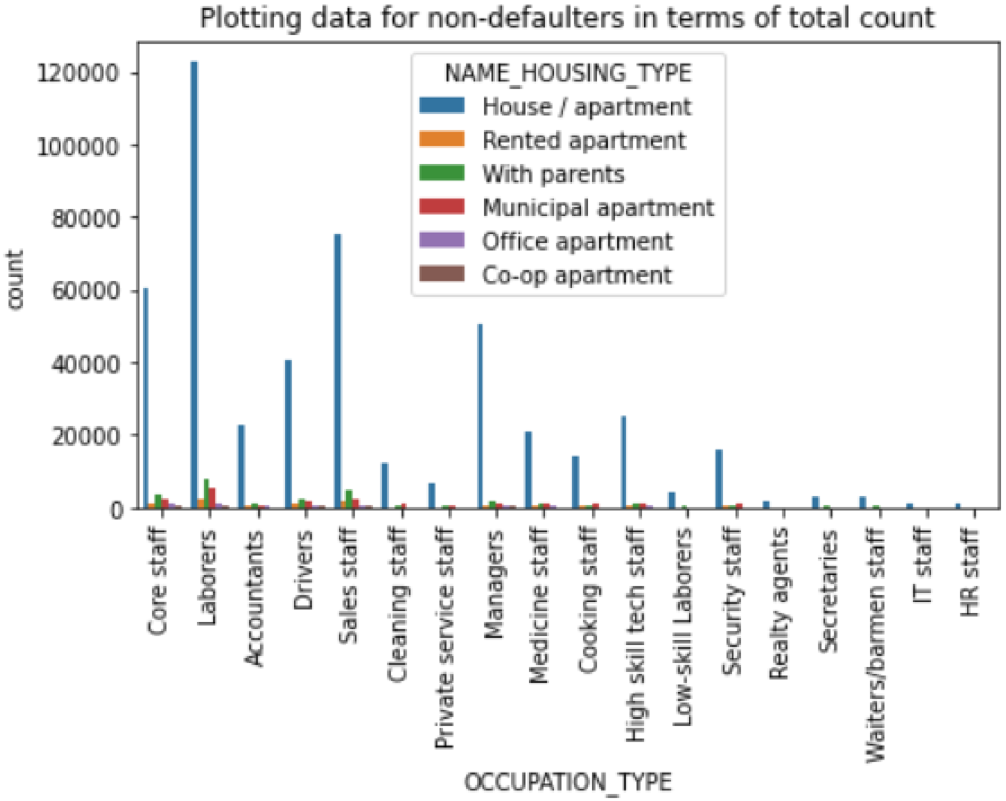



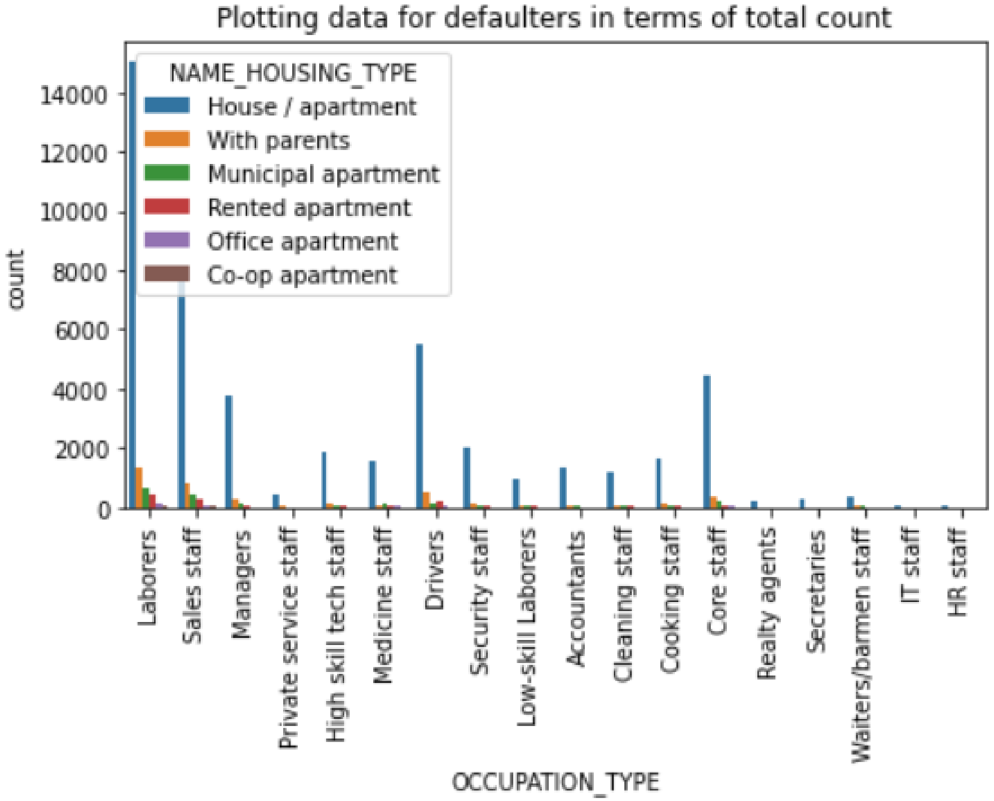
# OCCUPATION_TYPE- CONTRACT_STATUS: Laborers are applying for more applications and have accordingly have approval and refused.



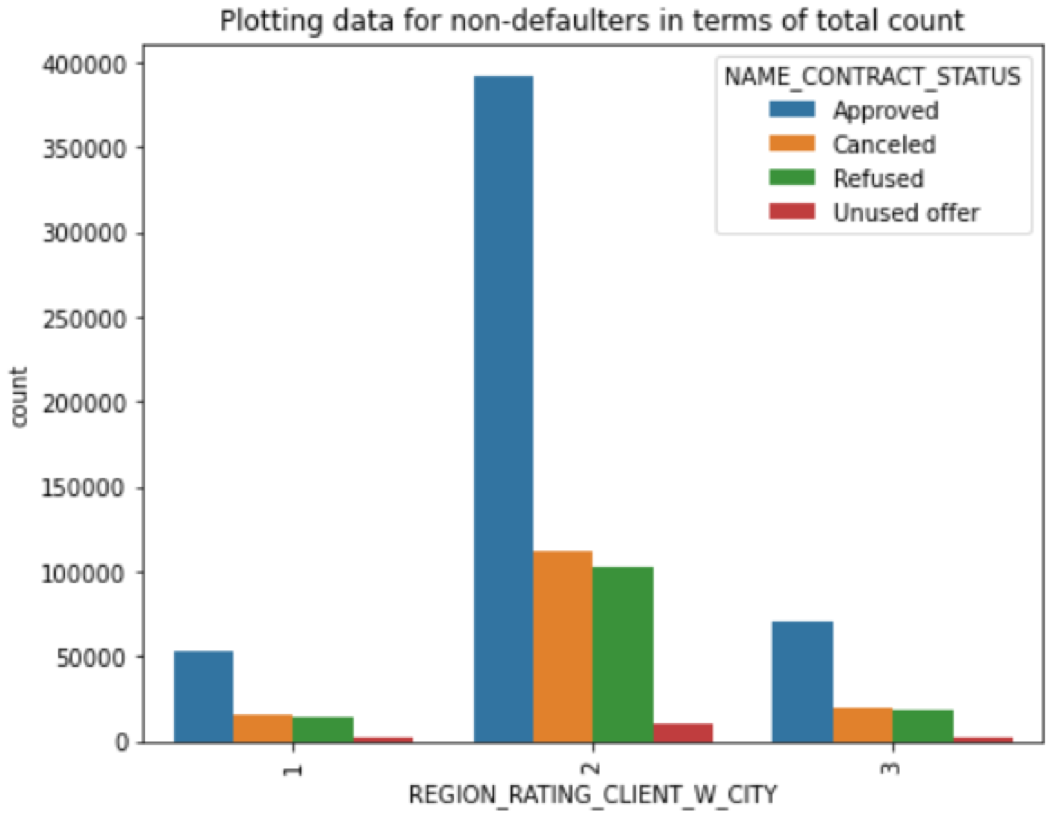



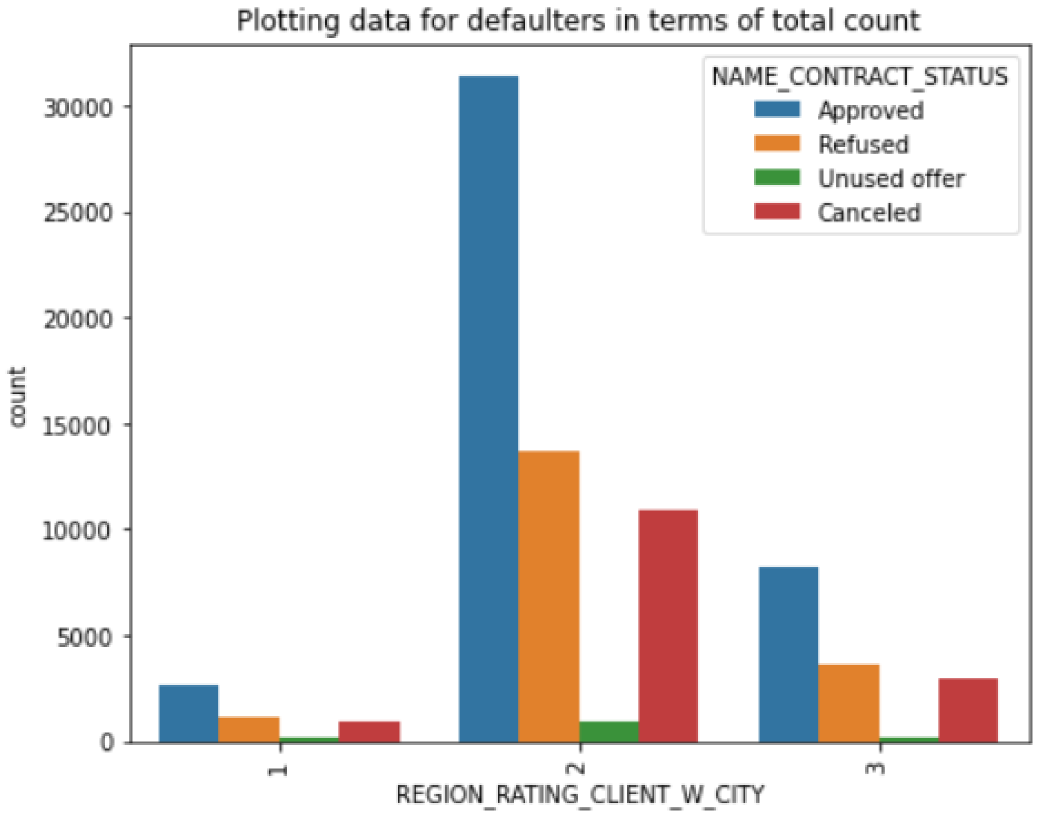
OCCUPATION_TYPE- HOUSING_TYPE: Laborers get approved their loans based on their housing type. Meaning most of the laborers own house/apartment which helps them in approval of application.



REGION_RATING_CLIENT_W_CITY- CONTRACT_STATUS: People in city type 2 are most likely to apply for loan.



# Top 10 correlations

# Client's with cases other than defaults

| | | |
|---|---|---|
| FLOORSMAX_MEDI | FLOORSMAX_MODE | 0.988157 |
| FLOORSMAX_MODE | FLOORSMAX_MEDI | 0.988157 |
| YEARS_BEGINEXPLUATATION_MEDI | YEARS_BEGINEXPLUATATION_AVG | 0.993582 |
| YEARS_BEGINEXPLUATATION_AVG | YEARS_BEGINEXPLUATATION_MEDI | 0.993582 |
| FLOORSMAX_AVG | FLOORSMAX_MEDI | 0.997019 |
| FLOORSMAX_MEDI | FLOORSMAX_AVG | 0.997019 |
| OBS_30_CNT_SOCIAL_CIRCLE | OBS_60_CNT_SOCIAL_CIRCLE | 0.998393 |
| OBS_60_CNT_SOCIAL_CIRCLE | OBS_30_CNT_SOCIAL_CIRCLE | 0.998393 |
| FLAG_EMP_PHONE | DAYS_EMPLOYED | 0.999758 |
| DAYS_EMPLOYED | FLAG_EMP_PHONE | 0.999758 |

These are the columns which are highly related to each other either positively or negatively for applicants who pay installments/loans on time.

# Client’s with payment difficulties

| | | |
|---|---|---|
| FLOORSMAX_MEDI | FLOORSMAX_MODE | 0.989369 |
| FLOORSMAX_MODE | FLOORSMAX_MEDI | 0.989369 |
| YEARS_BEGINEXPLUATATION_MEDI | YEARS_BEGINEXPLUATATION_AVG | 0.996125 |
| YEARS_BEGINEXPLUATATION_AVG | YEARS_BEGINEXPLUATATION_MEDI | 0.996125 |
| FLOORSMAX_AVG | FLOORSMAX_MEDI | 0.997233 |
| FLOORSMAX_MEDI | FLOORSMAX_AVG | 0.997233 |
| OBS_60_CNT_SOCIAL_CIRCLE | OBS_30_CNT_SOCIAL_CIRCLE | 0.998270 |
| OBS_30_CNT_SOCIAL_CIRCLE | OBS_60_CNT_SOCIAL_CIRCLE | 0.998270 |
| FLAG_EMP_PHONE | DAYS_EMPLOYED | 0.999702 |
| DAYS_EMPLOYED | FLAG_EMP_PHONE | 0.999702 |

These are the columns which are highly related to each other either positively or negatively for applicants who do not pay installments/loans on time.

# What is ratio of imbalance?

**11.38863819500403**

Since there is a huge difference in counts of 0 and 1, therefore data is imbalanced. Means data contains more number of people's information who are paying installments on time and a smaller number of people’s information with difficulties in payments.