



CREDIT ANALYSIS

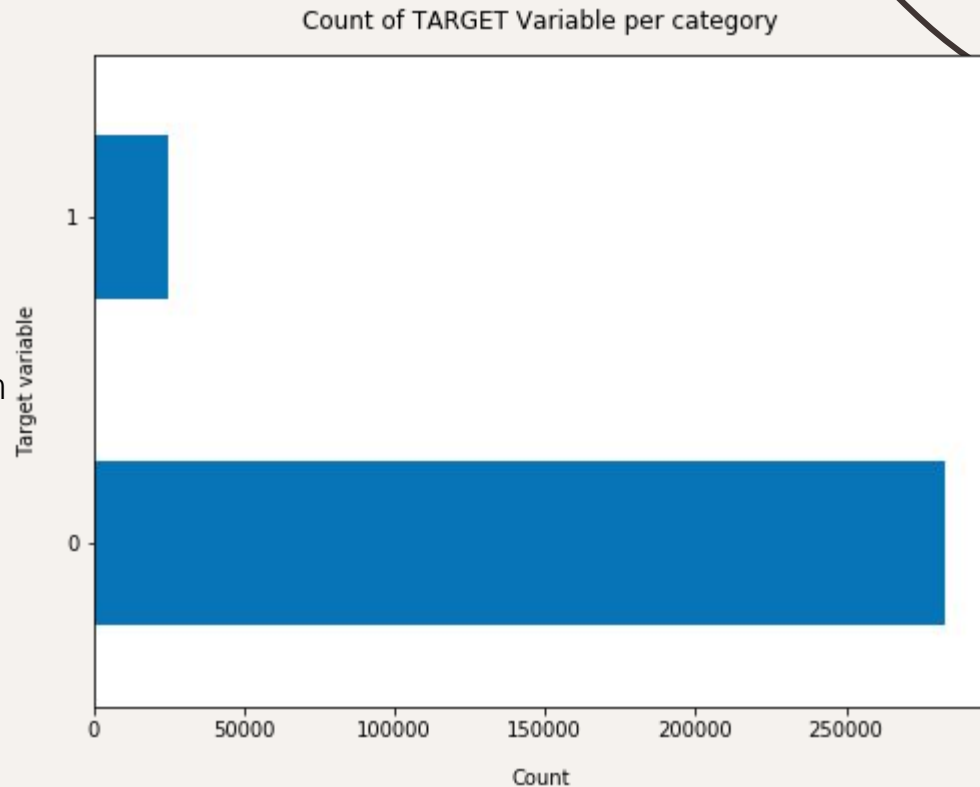
Exploratory Data Analysis

Faaizah Faisal

TARGET VARIABLE

INSIGHT:

- It can be seen that the data is highly imbalanced.
- Almost 92% of the loans were paid back in time(Target 0)

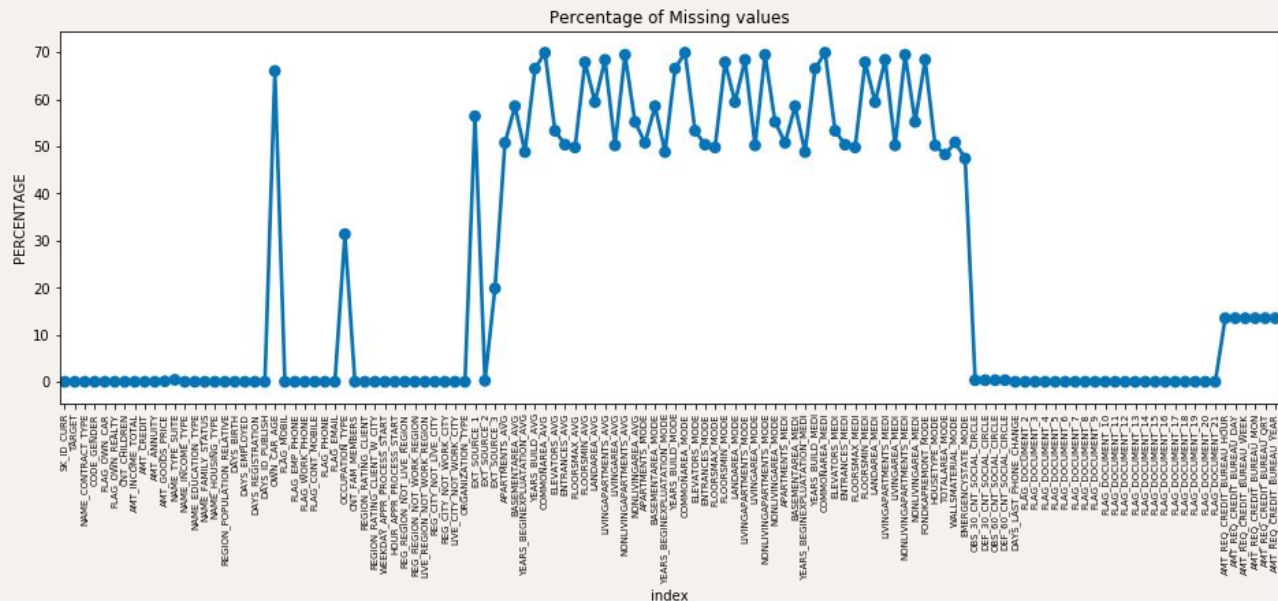


The image features two thin, dark horizontal lines. The top line starts with a curved segment on the left side before becoming straight. The bottom line ends with a curved segment on the right side after becoming straight.

MISSING DATA

INSIGHT:

- The graph represents the percentage of missing values for each column.
- 64 columns have more than 30% of missing values.
- These columns can be deleted if they don't carry much information.



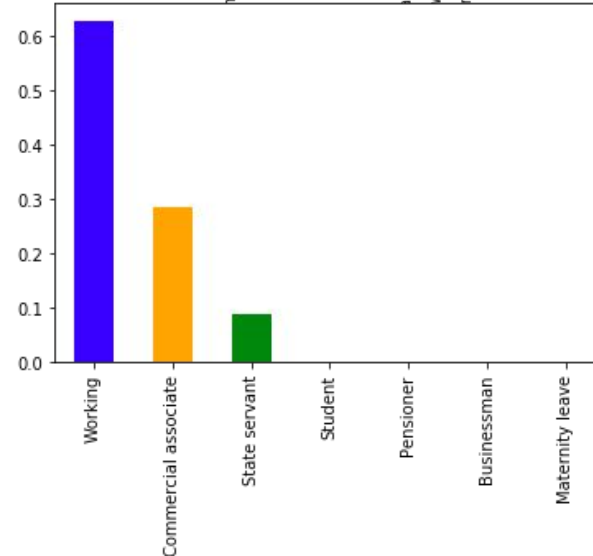
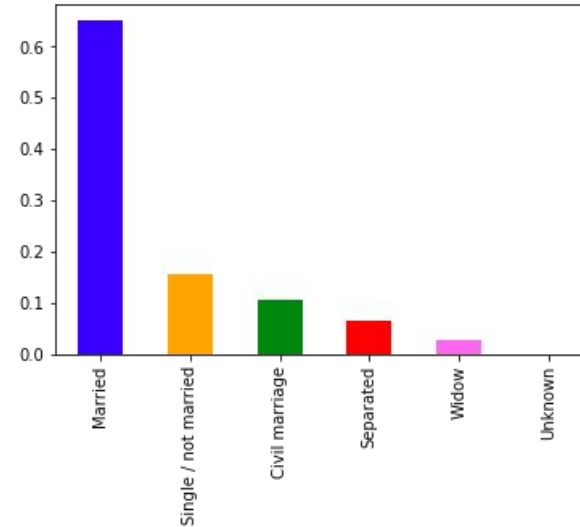
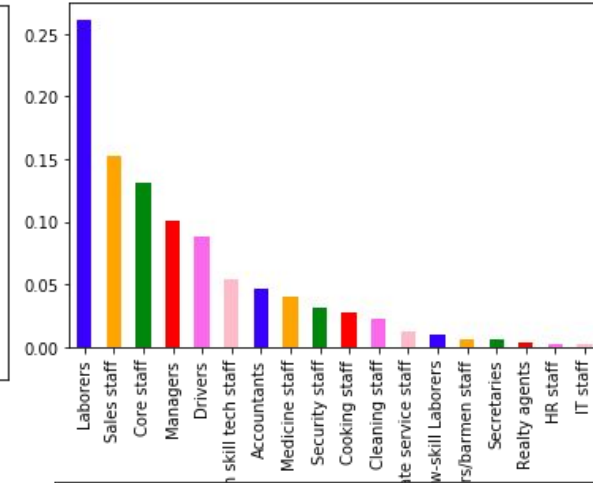
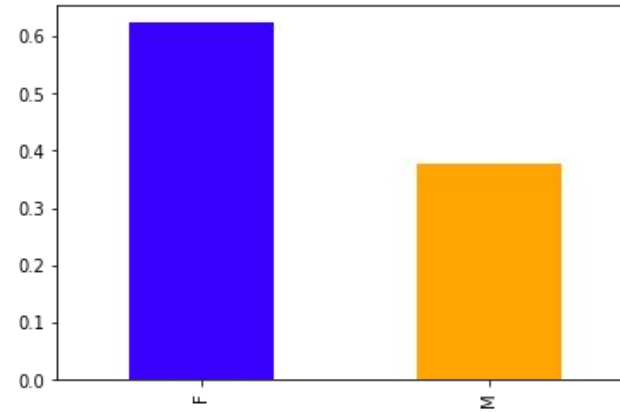
CATEGORICAL ANALYSIS



UNIVARIATE ANALYSIS

INSIGHT:

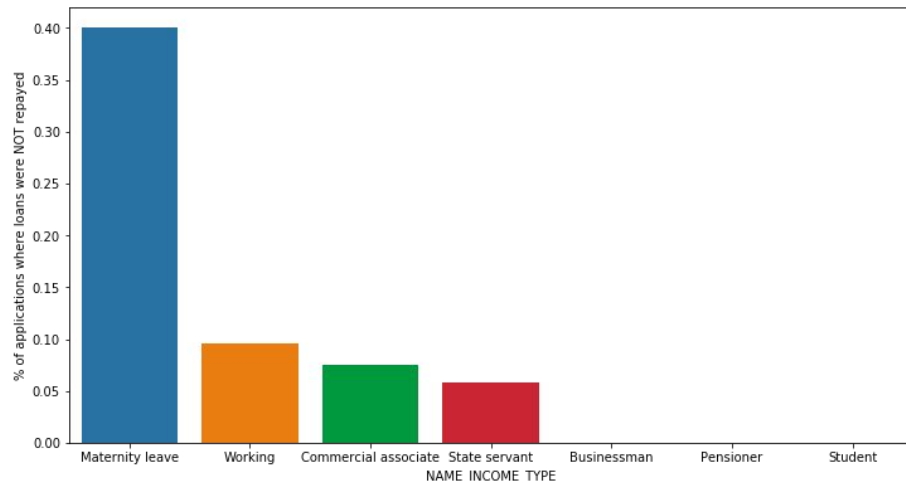
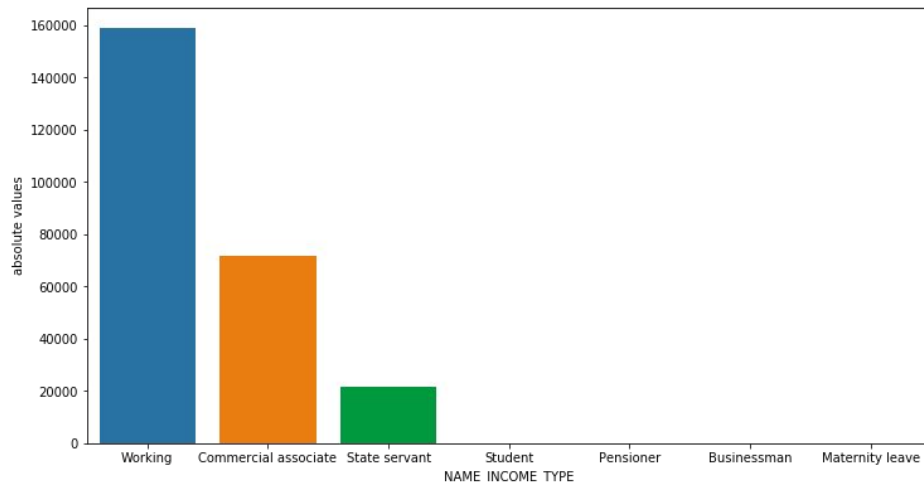
- Univariate analysis does not give much insight, it only shows the maximum number of people in each category.
- It can be seen that most of the customers are females.
- Majority of the customers are labourers.
- Majority of the customers are married.
- Most of the customers are working.



INCOME TYPE vs TARGET VARIABLE

INSIGHT:

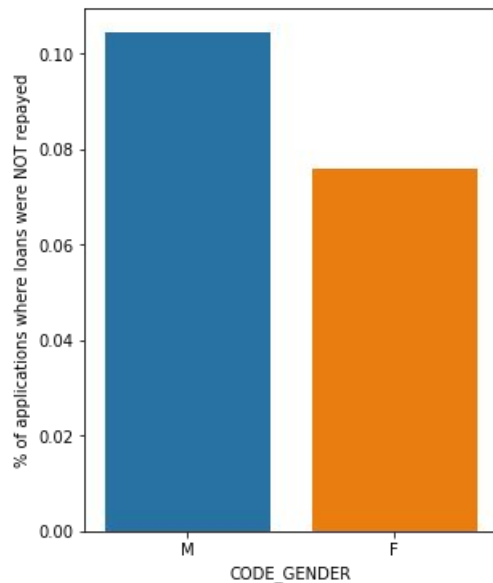
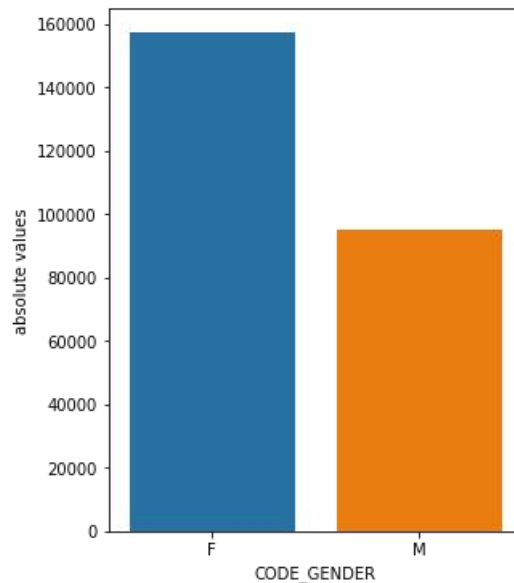
- The Working class applied for most loans and have a lower default rate, hence they are reliable customers.
- Customers on maternity leave have high default rates.
- Customers who are reliable are Commercial associates, State Servant, Businessmen, Pensioners and Students.



GENDER vs TARGET VARIABLE

INSIGHT:

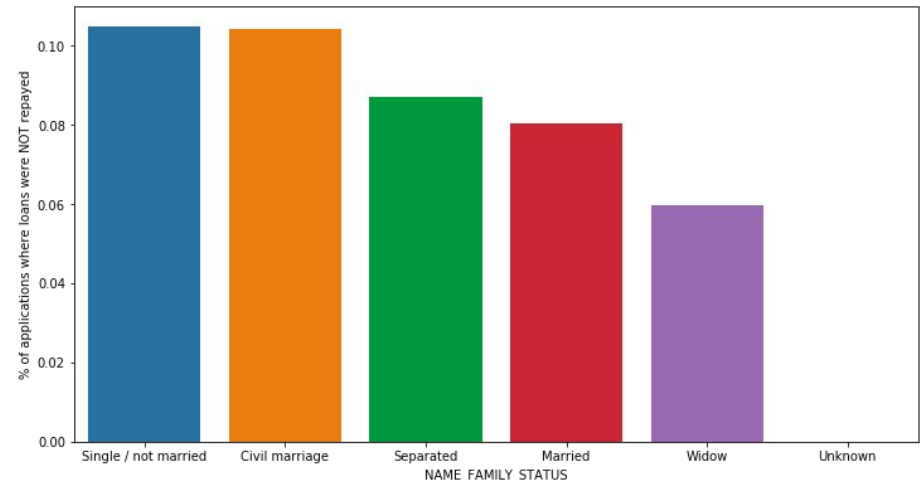
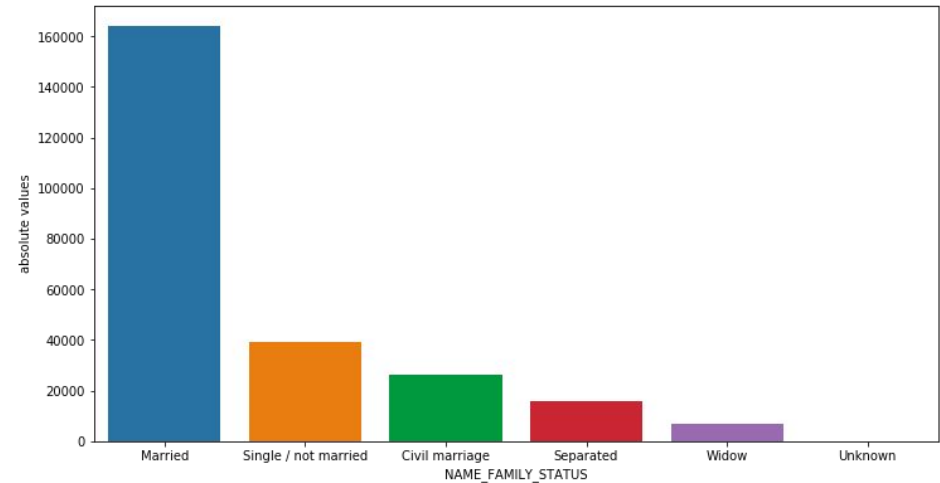
- There are more female clients than male clients.
- However, males have a much higher default rate than women, hence, women are more reliable customers.



FAMILY STATUS vs TARGET VARIABLE

INSIGHT:

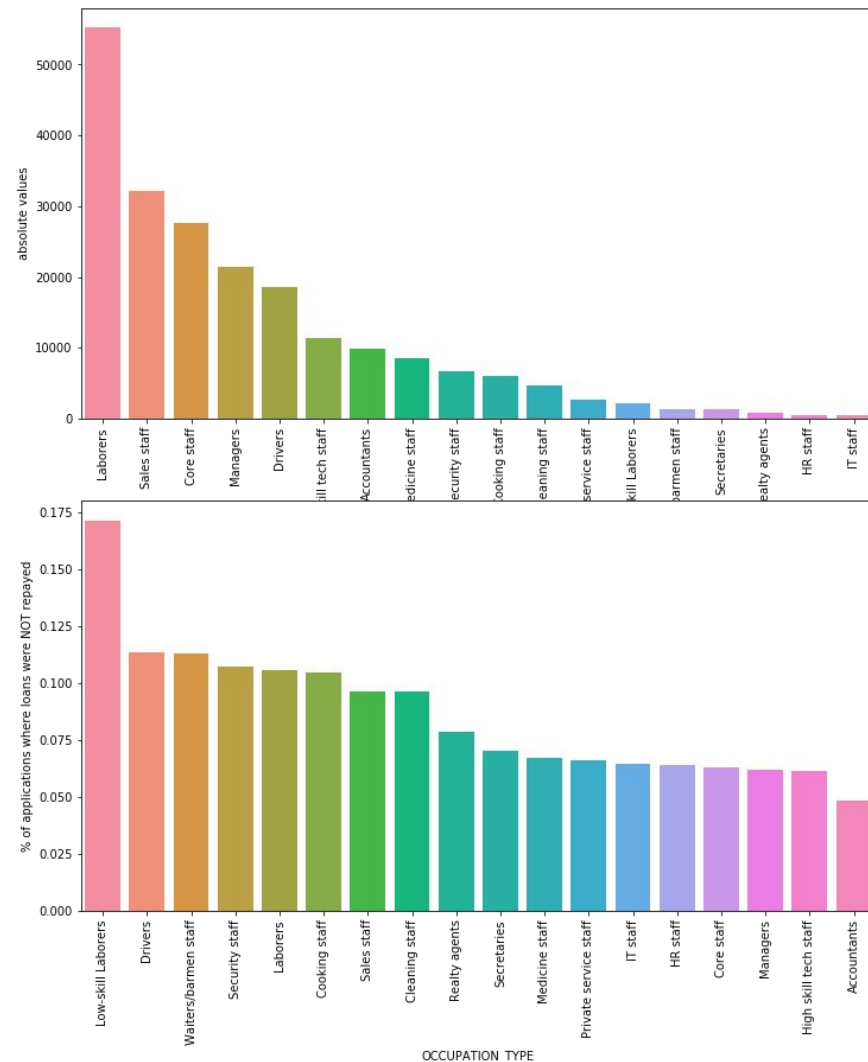
- Married customers were the ones who applied for most loans.
- Single or unmarried customers and civil marriages have higher default rates compared to other categories, hence they are not reliable customers.



OCCUPATION TYPE vs TARGET VARIABLE

INSIGHT:

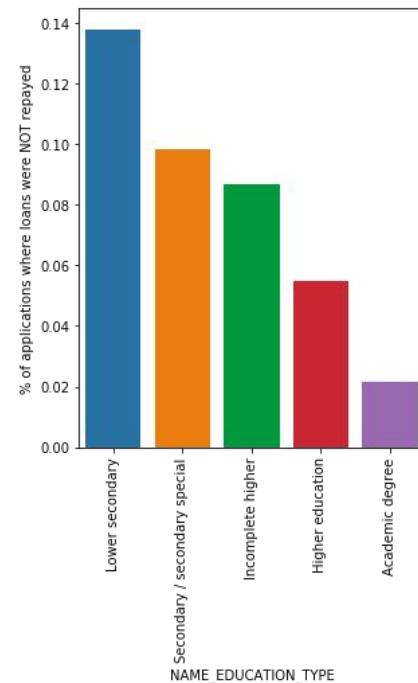
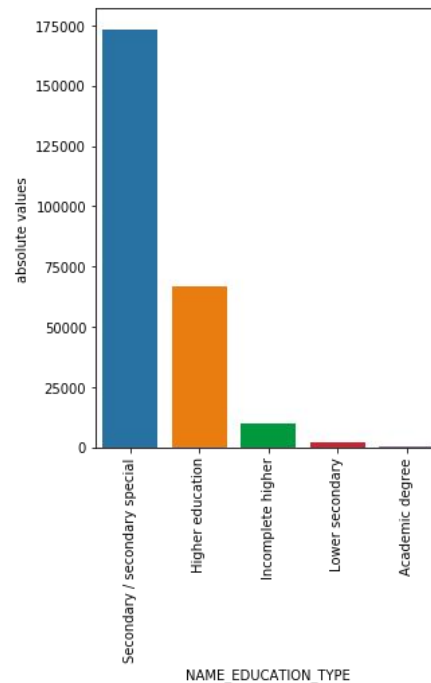
- It can be seen that customers in the lower income groups like low skilled Labourers, Drivers, waiters, Security Staff, Cooking Staff have high default rates hence they are not reliable customers.
- However, higher paid customers like Accountants, tech staff, HR staff, Managers have lower default rates and are more reliable customers.



EDUCATION vs TARGET VARIABLE

INSIGHT:

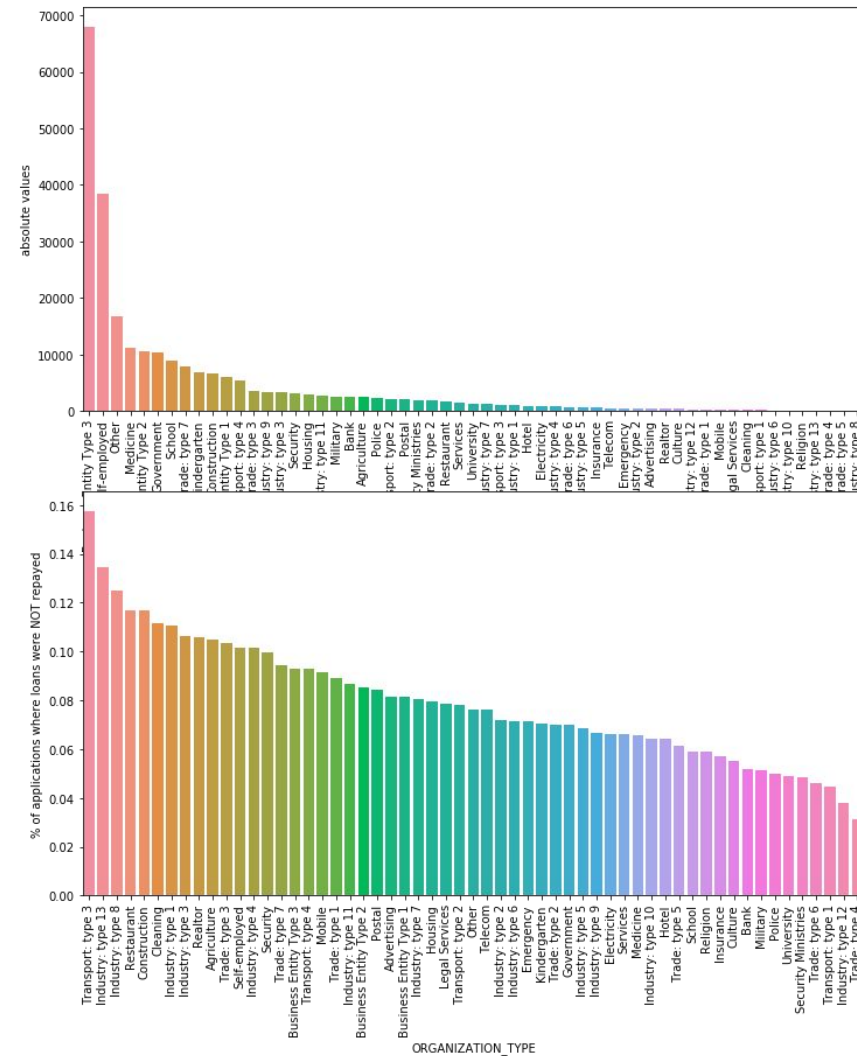
- Most loans are taken for Secondary and Higher education.
- Lower secondary education and Secondary special are the risky categories due to high default rates.



ORGANISATION TYPE vs TARGET VARIABLE

INSIGHT:

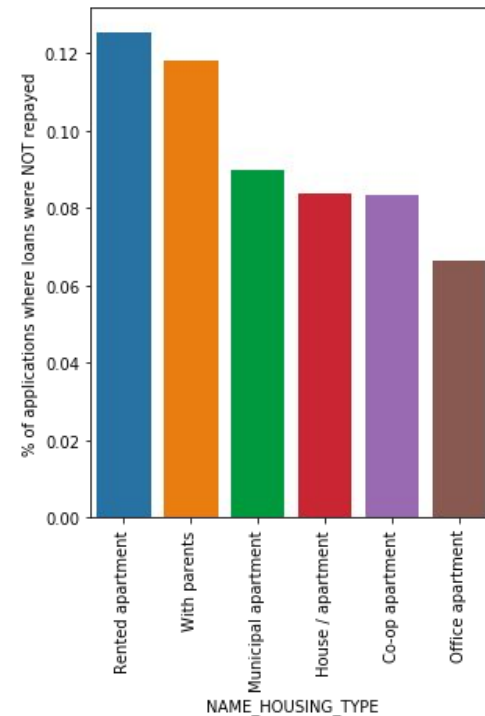
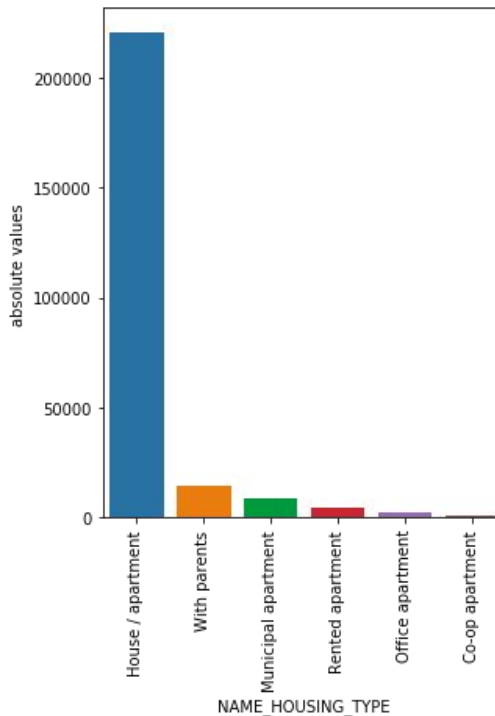
- Organisation types that are risky with high default rates are Transport type 3 (16%) , Industry type 13 (13.5%), Industry: type 8 (12.5%)and Restaurant with almost 12%.
- Trade: type 4, Industry: type 12, Transport type 12 have the lowest default rates and are reliable customers.



HOUSING TYPE vs TARGET VARIABLE

INSIGHT:

- Customers who live in rented apartments and with parents have a high default rate, hence they are unreliable customers.
- Clients belonging to the office apartment category have low default rates (around 6%), hence they are reliable customers.



CONTRACT TYPE vs TARGET VARIABLE

INSIGHT:

- Most of the loans taken by clients are Cash Loans.
- However, cash loans are not repaid in comparison with revolving loans.

