

# EDA CASE STUDY CREDIT RISK ANALYSIS

BY EPSITA  
BOSE(EPSITABOSE955@GMAIL.  
COM) AND SHASHANT  
RATHORE(NANICKS45@GMAIL  
.COM)

## PROBLEM STATEMENT

1. This case study aims 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.
2. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study



# Steps to Follow

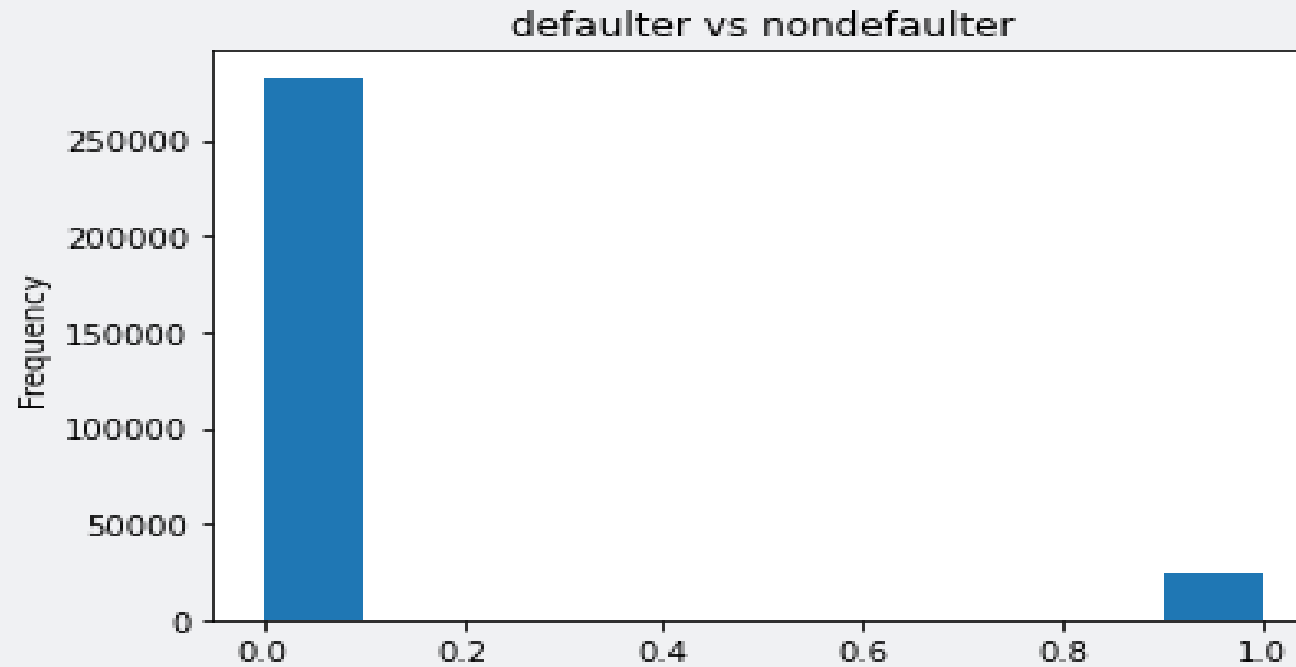
- ♦ Import the Data
- ♦ Read the Data
- ♦ Check for the data quality issues and start binning into groups for easy analysis.
- ♦ Check for data imbalance for univariate, bivariate analysis and correlation.
- ♦ Now merge the application data with previous data.
- ♦ Do data analysis univariate, bivariate analysis , multivariate and correlation.
- ♦ Making inferences by using data



# Defaulter Vs Non Defaulter Graph

.Identify the Defaulter

Identify the Non Defaulter



#its concluded that there is only few defaulters

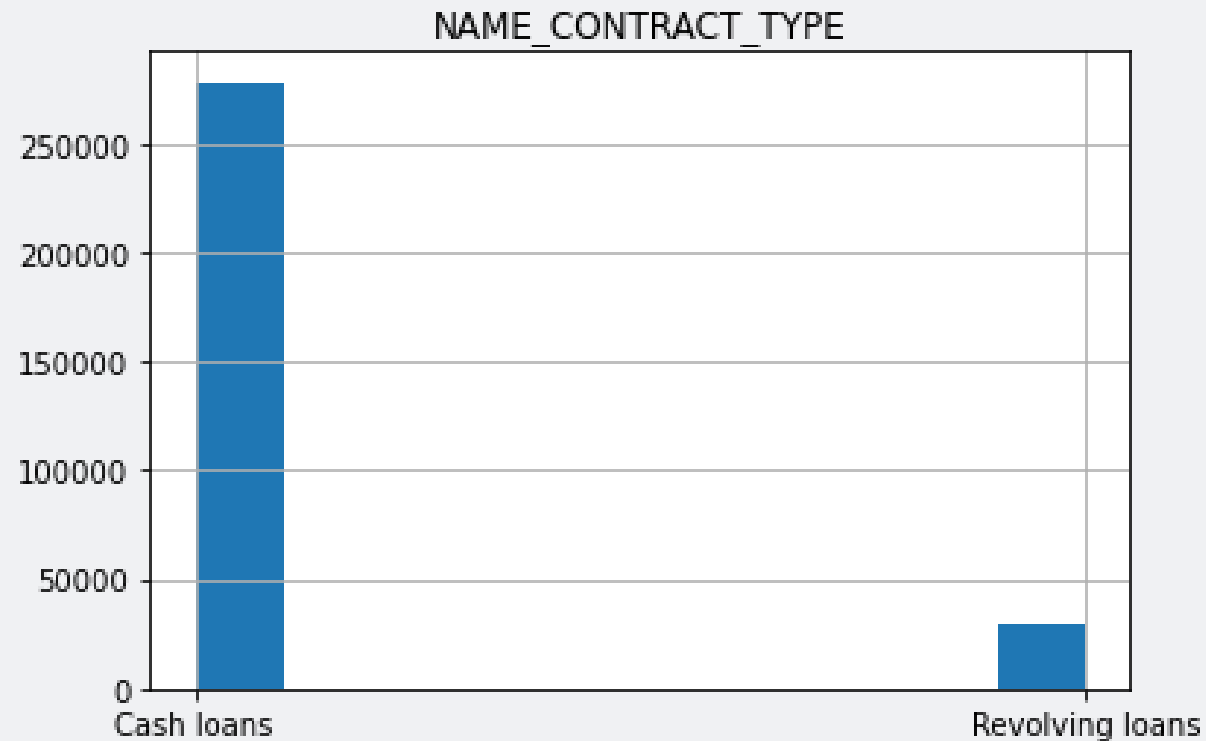
# Identify the how many contract type are there

It will tell us the different types of loan and the number of loans running

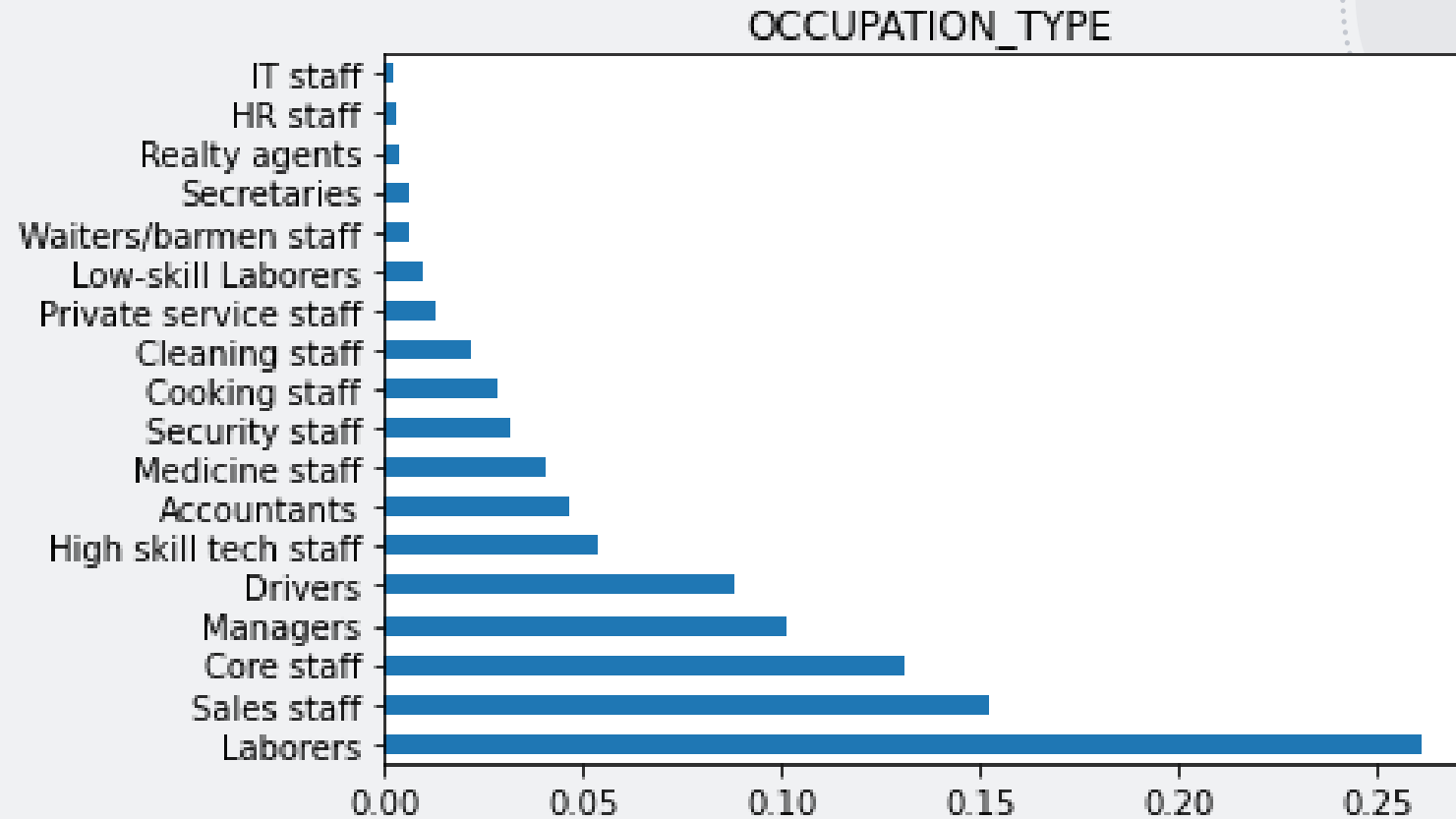
Cash loans      278232

Revolving loans    29279

Inference – Cash Loans are  
Very Large as compared to  
Revolving Loans



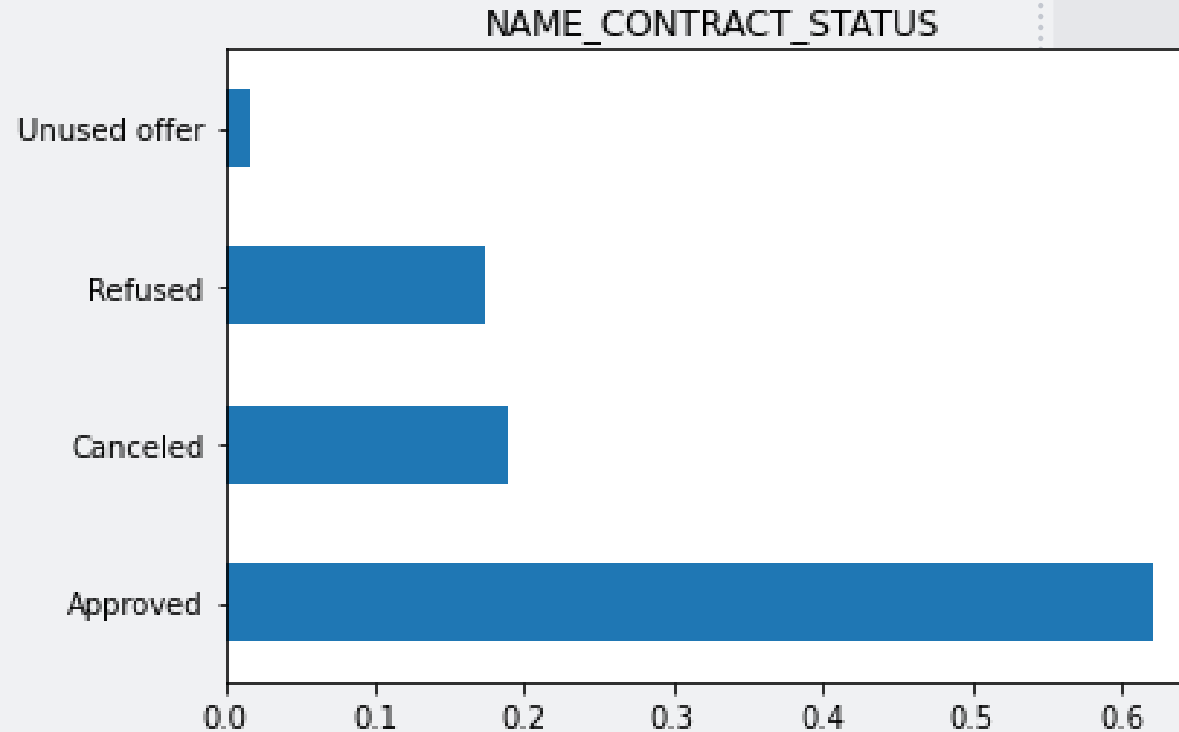
# Identify the how many occupation type are there



Inference: Maximum clients are of laborer types and minimum is IT staff.

# Identify the contract status

Approved	1036781
Canceled	316319
Refused	290678
Unused offer	26436

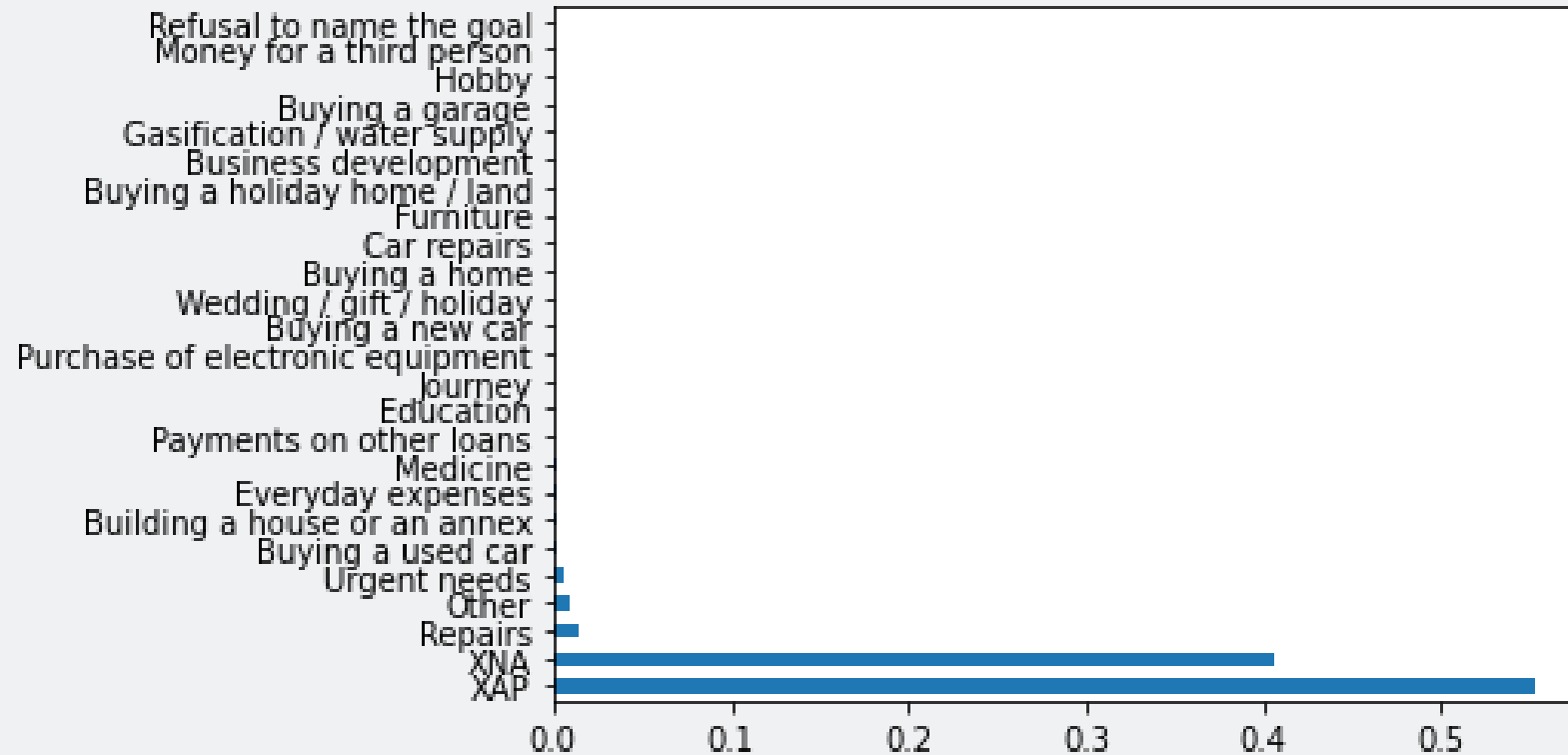


Inference – Maximum loan applications are approved



# Identify the CASH\_LOAN\_PURPOSE

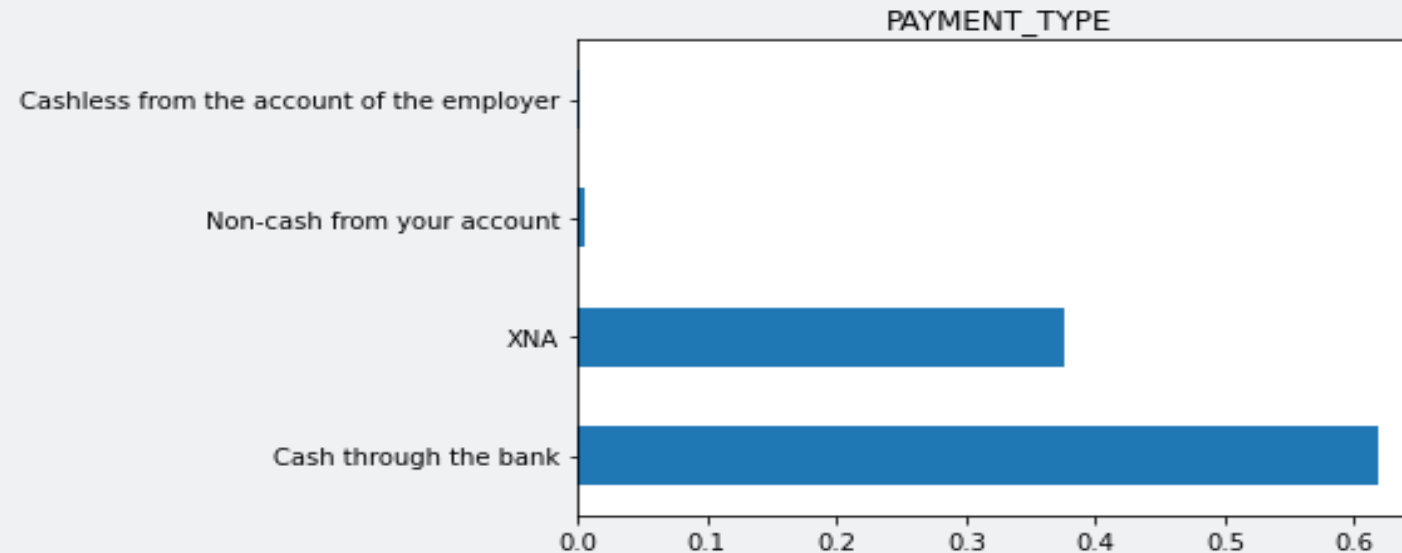
Inference- least is for Urgent Needs and Max for XAP





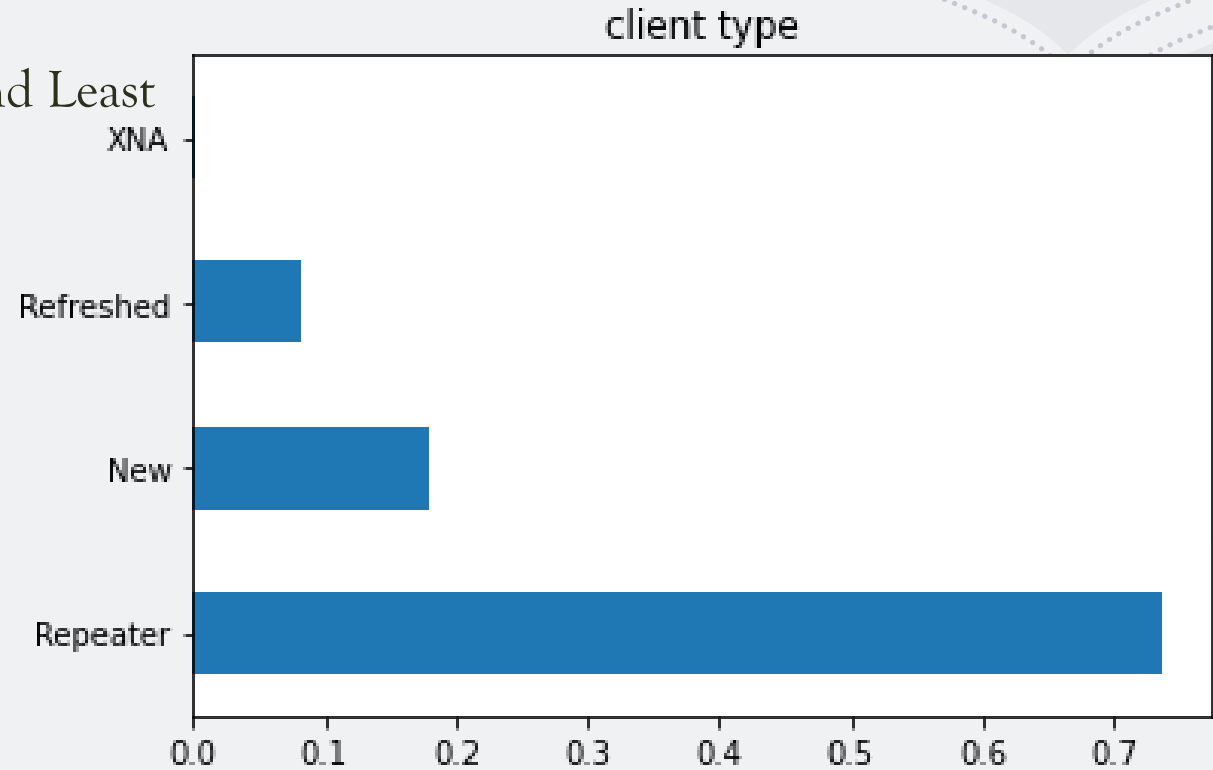
# Identify the PAYMENT TYPE status

Inference - CASH through bank is maximum



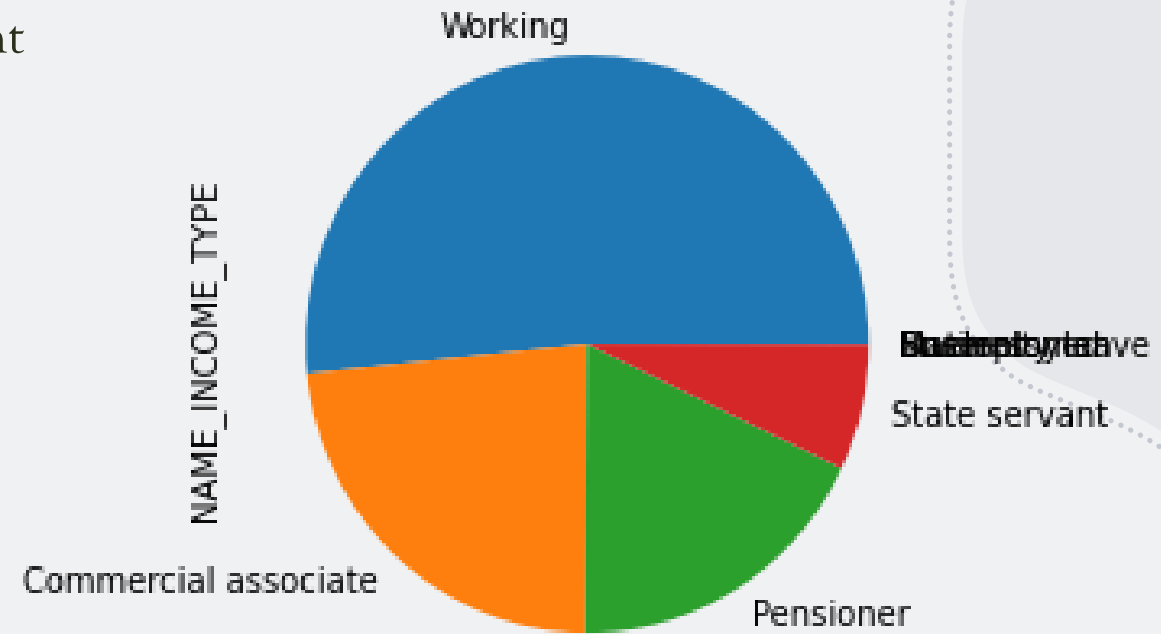
# Identify the client type

Inference- Maximum clients are repeater and Least are refreshed



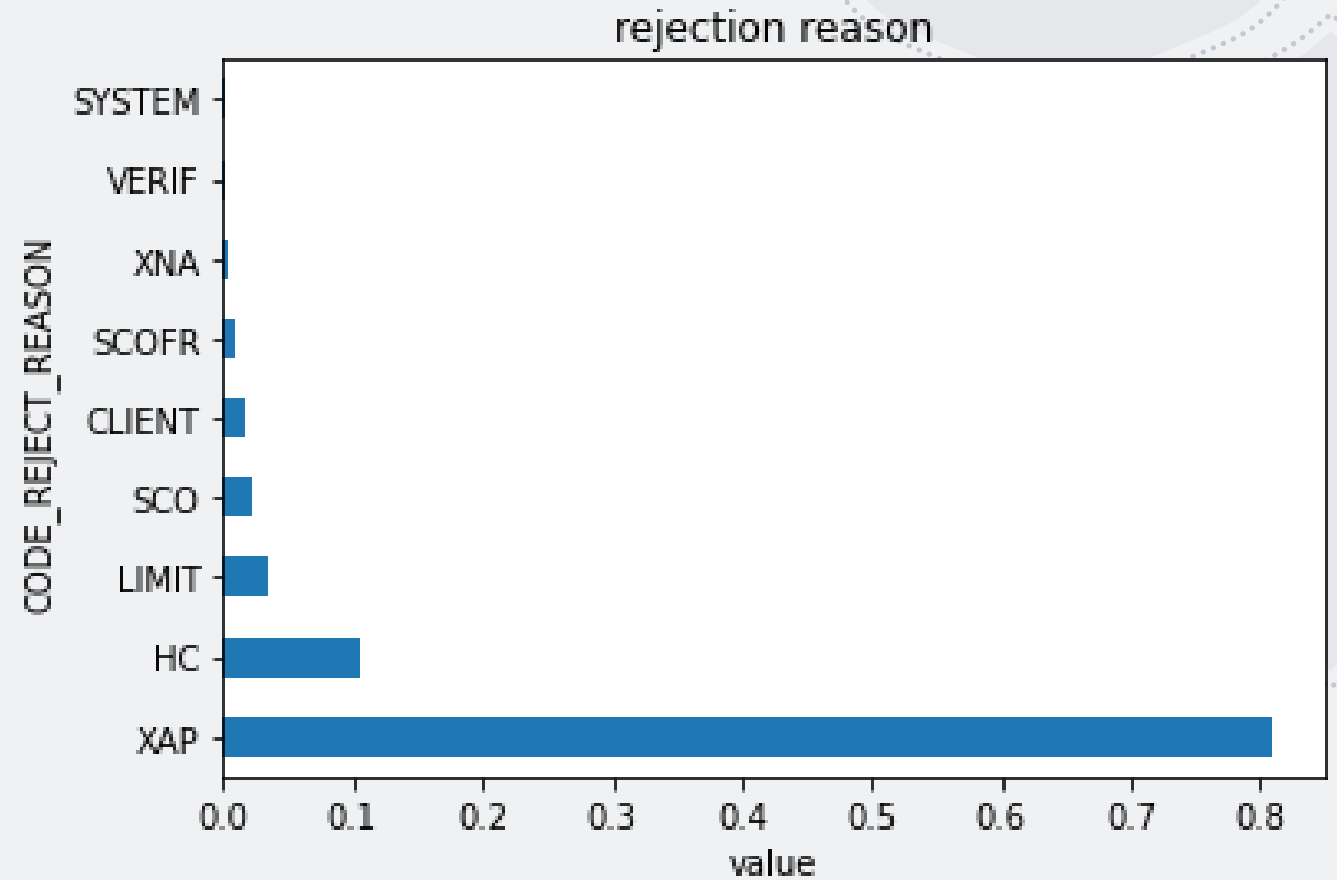
# Identify the income type

Inference- Income of Maximum people is through Working  
and Minimum number of people are State Servant



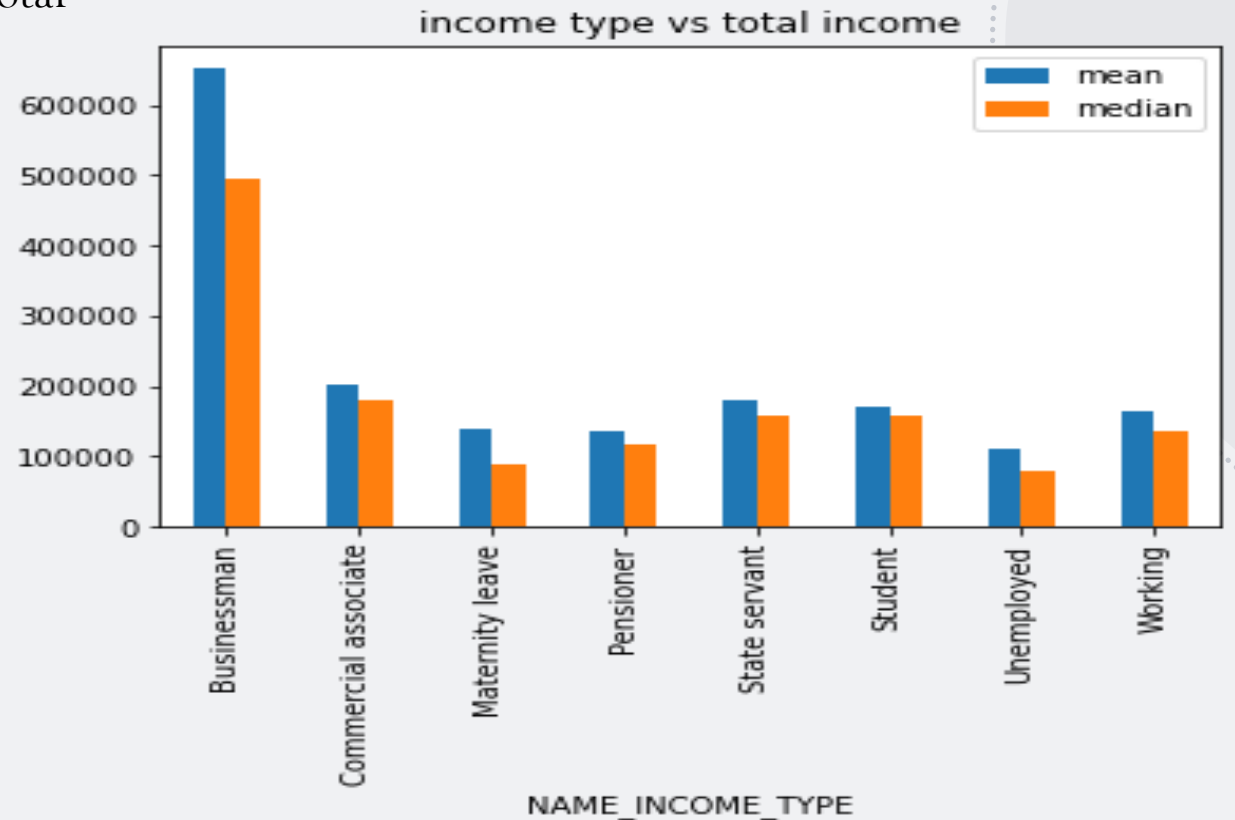
# Why the previous loan was rejected

Inference-Maximum loan were rejected  
Because of XAP and minimum because of  
XNA



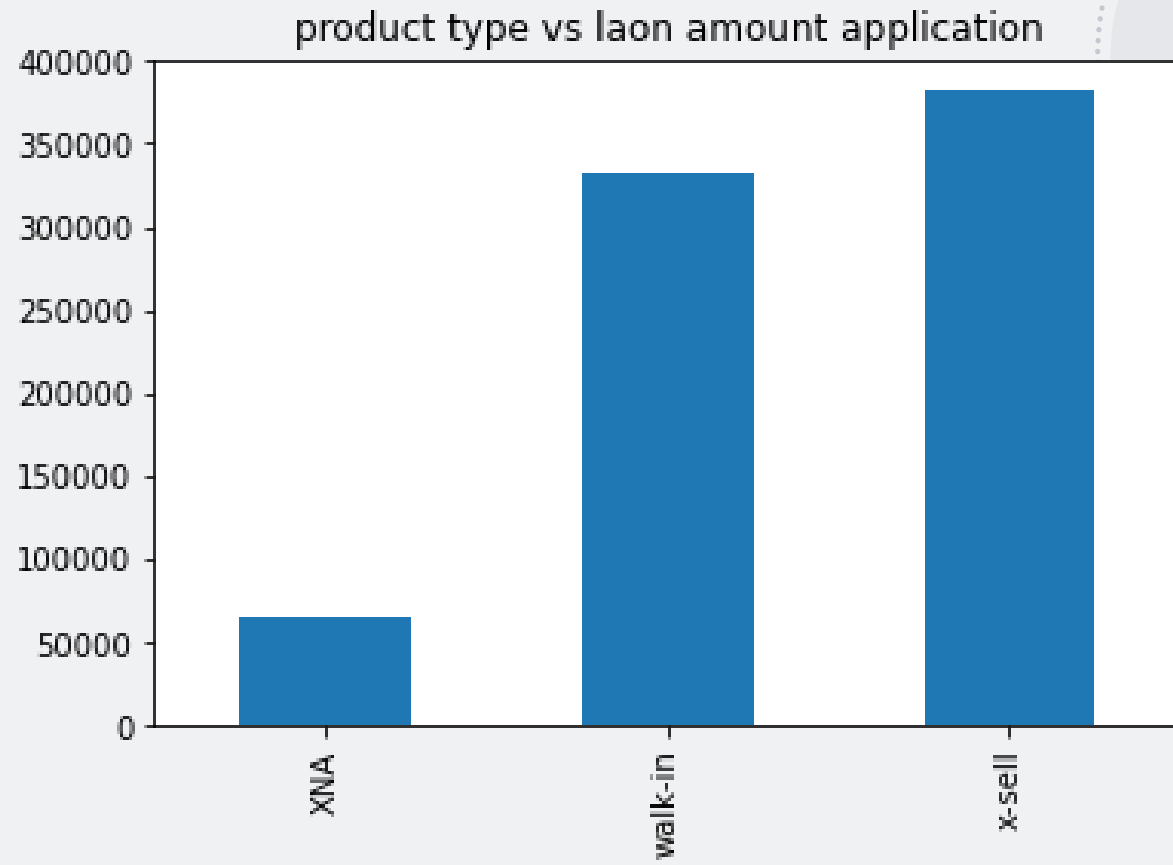
# Analysis income type vs total income

Inference- Businessman have the highest income as as well as total income as compared to unemployed which have the lowest

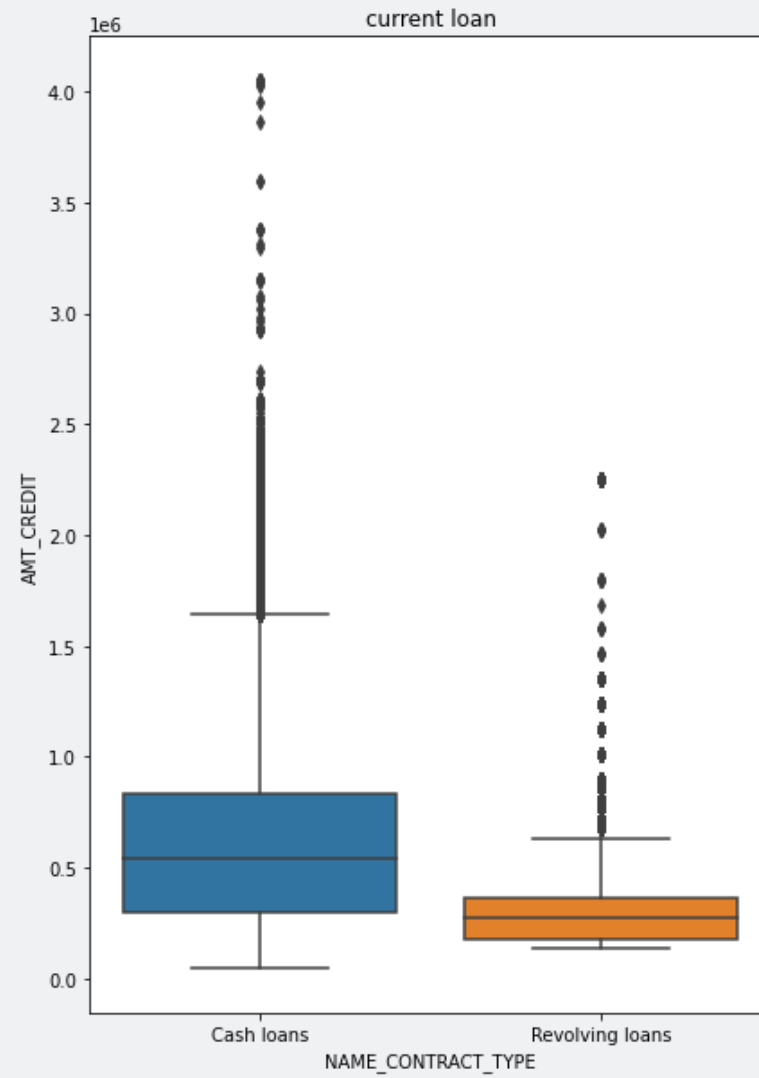
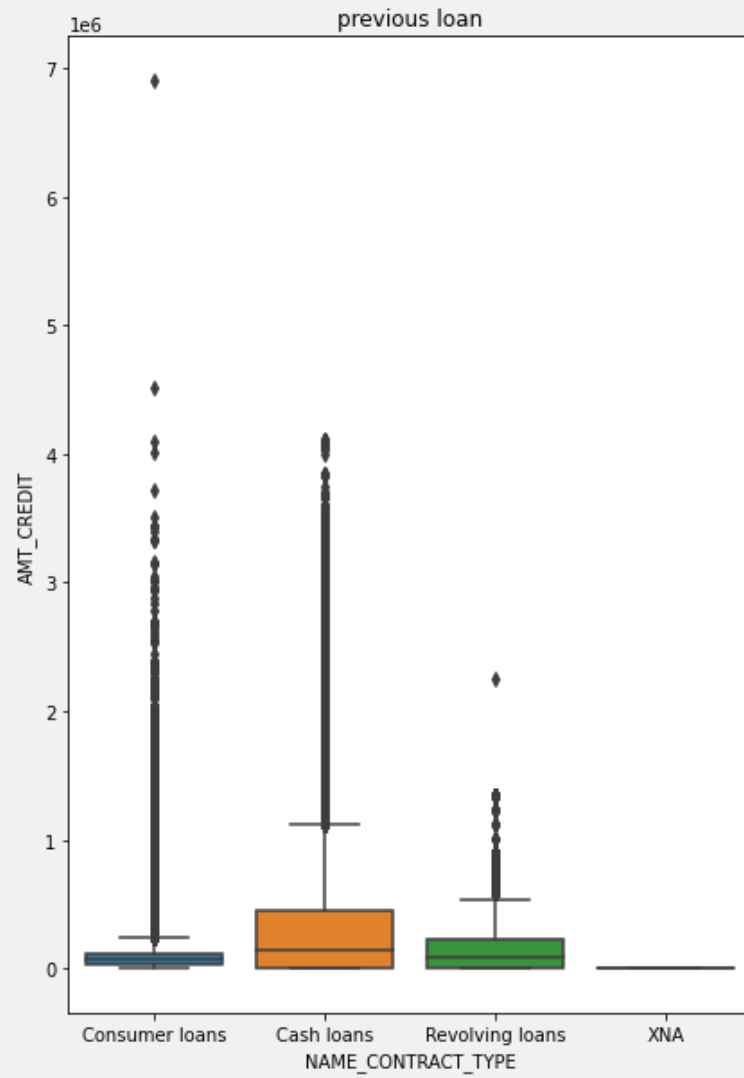


# Analysis product type vs loan amount application

X-sell has the highest loan application



# Bivariate analysis





# Plotting Client's Education Type

Secondary / secondary special 218391

Higher education 74863

Incomplete higher 10277

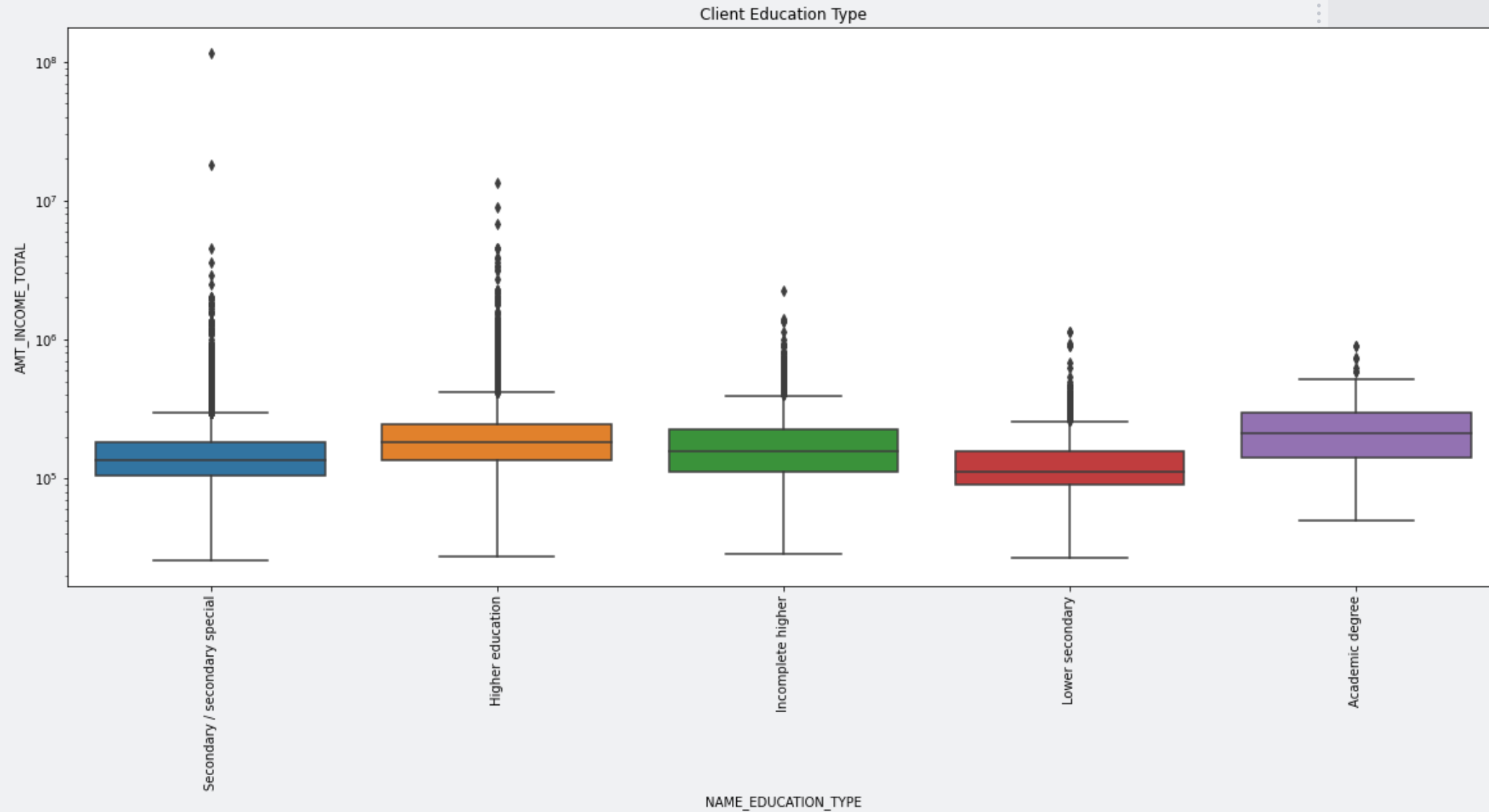
Lower secondary 3816

Academic degree 164

Inference- People applying for loan are maximum from secondary education and least from Academic degree

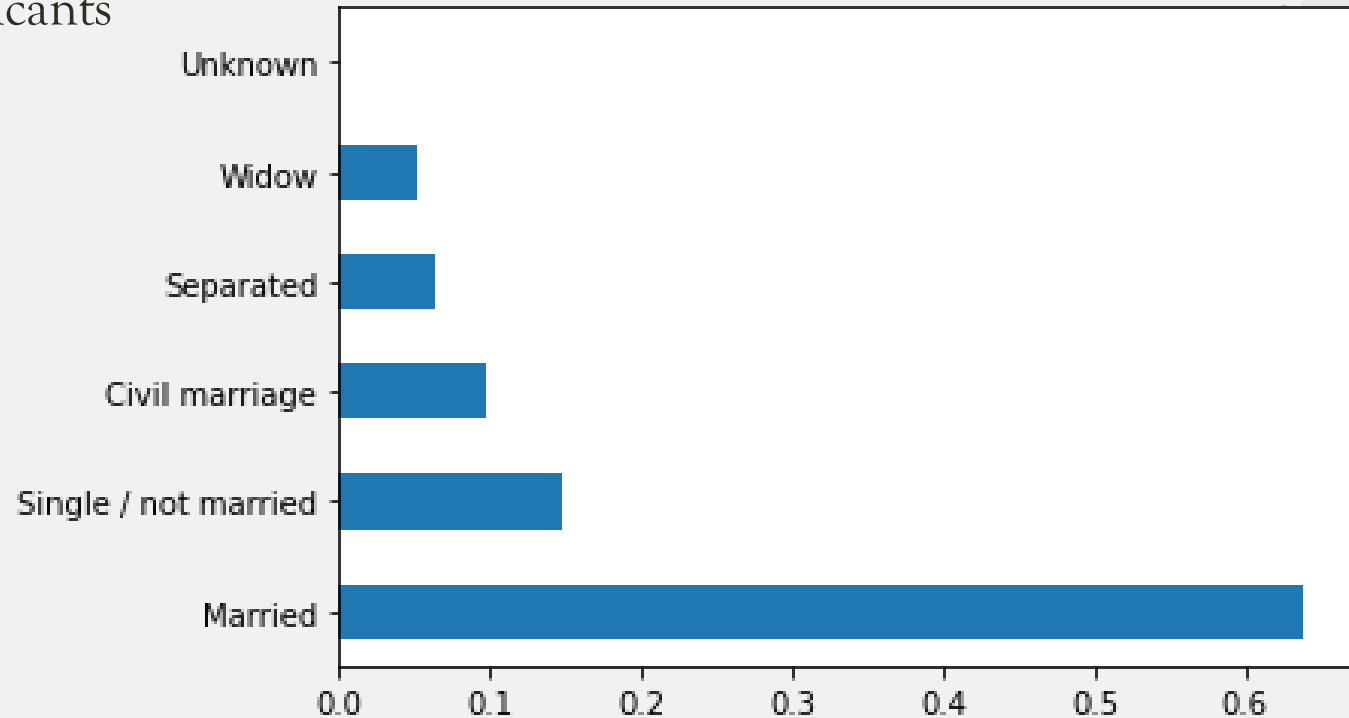


Plot against client education type



# Identifying the Client's Family status

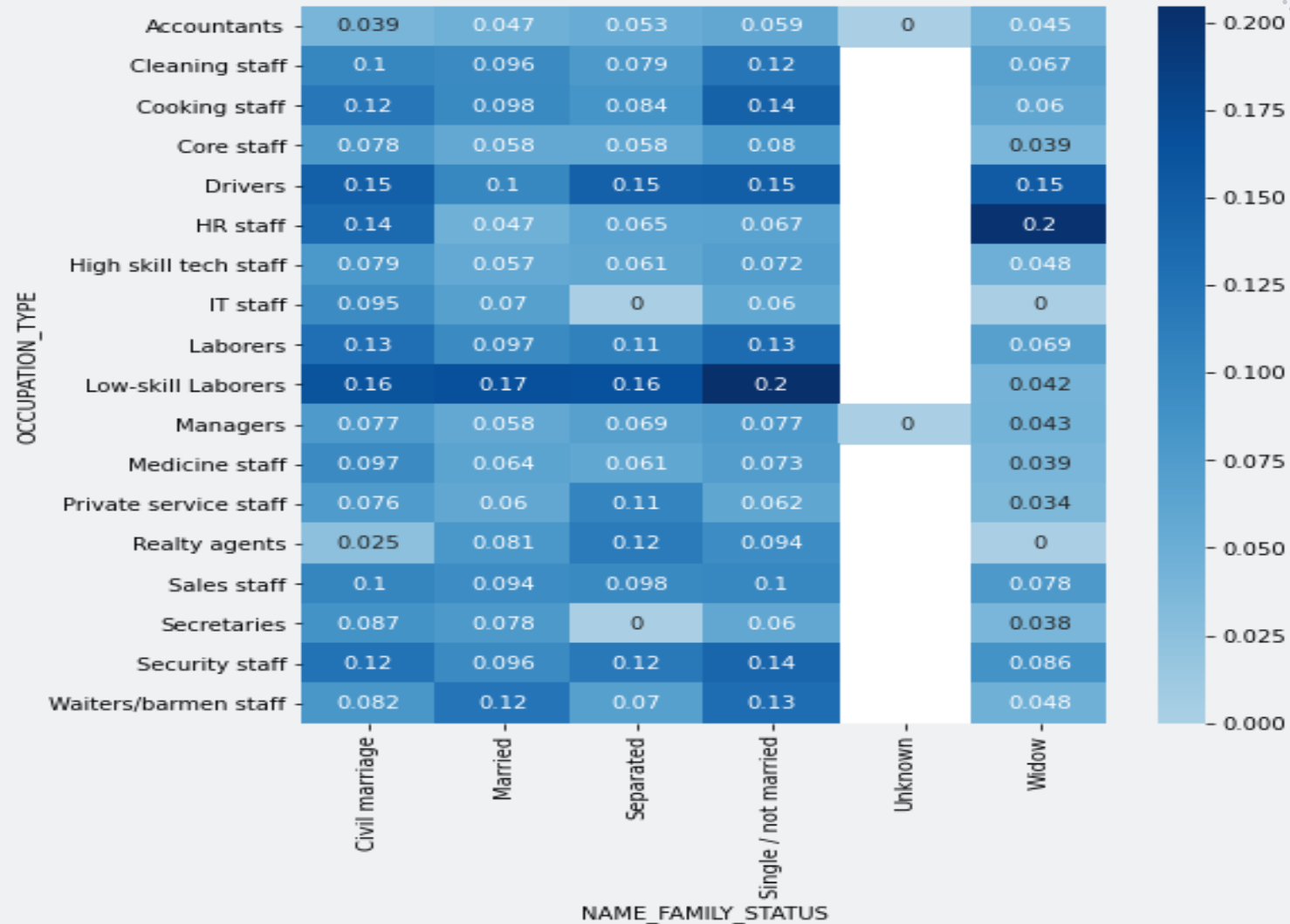
Inference- Maximum applicants are  
Married and least number of applicants  
are Widow



# Multivariate Analysis

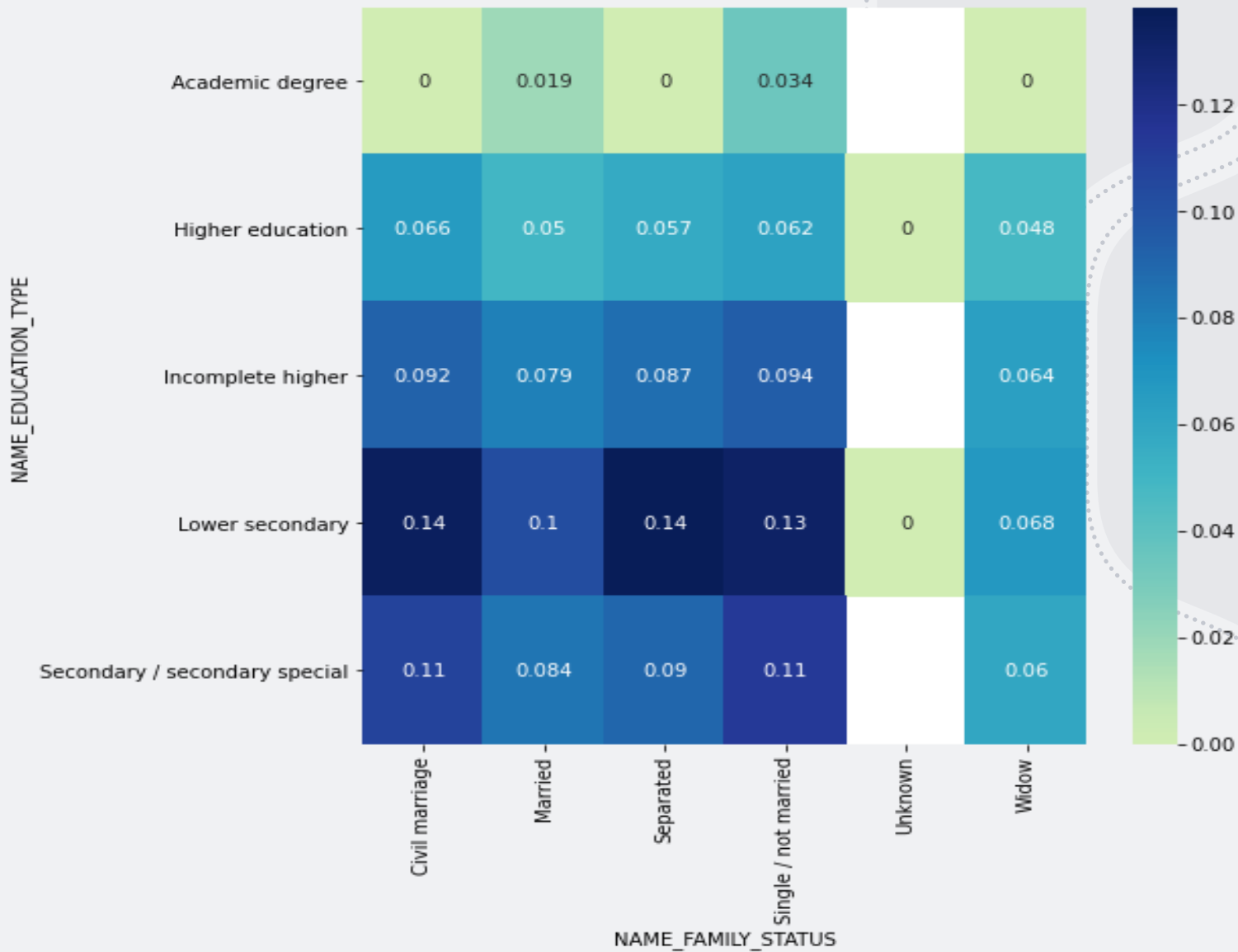
NAME\_FAMILY\_STATUS vs OCCUPATION\_TYPE vs TARGET

Inference- Single and Low-Skill laborers have the highest impact

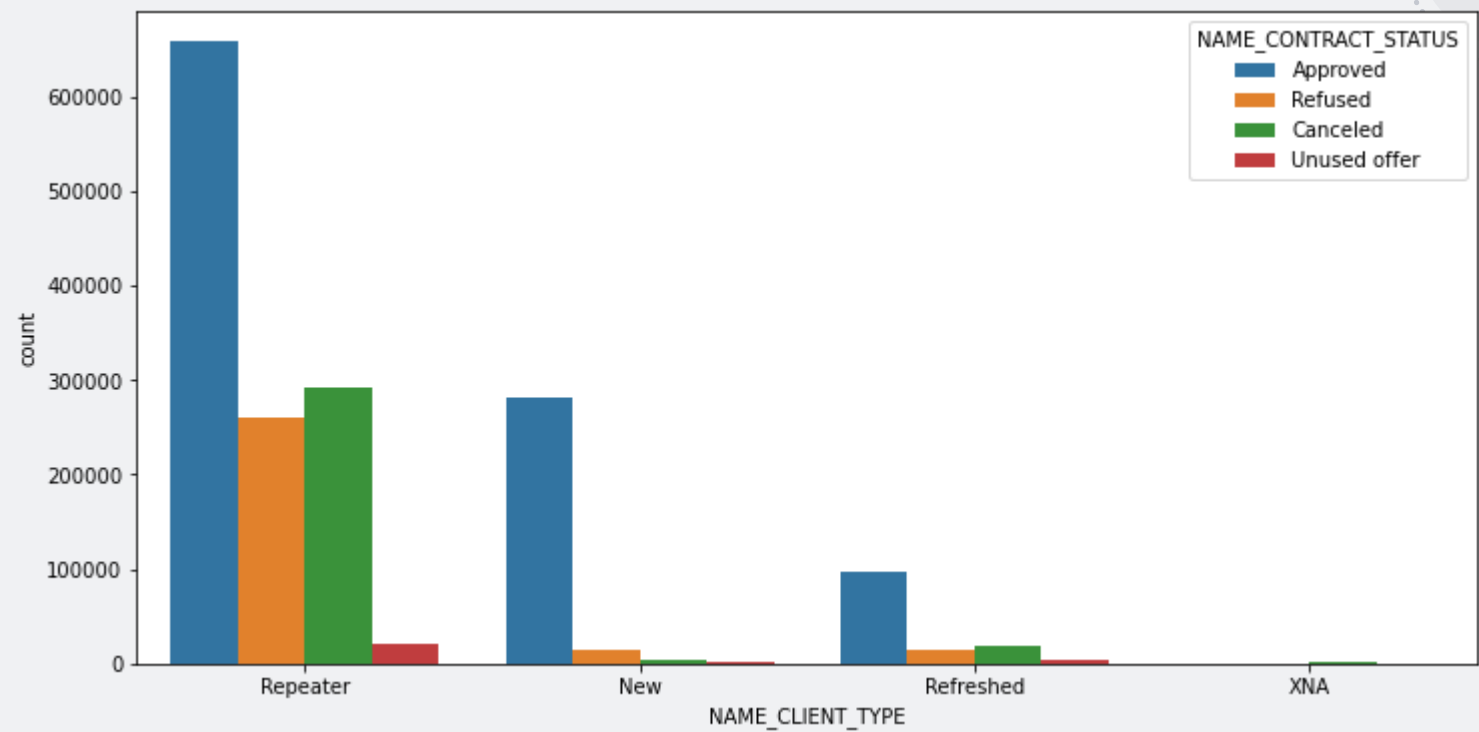


NAME\_FAMILY\_STATUS vs NAME\_EDUCATION\_TYPE vs TARGET

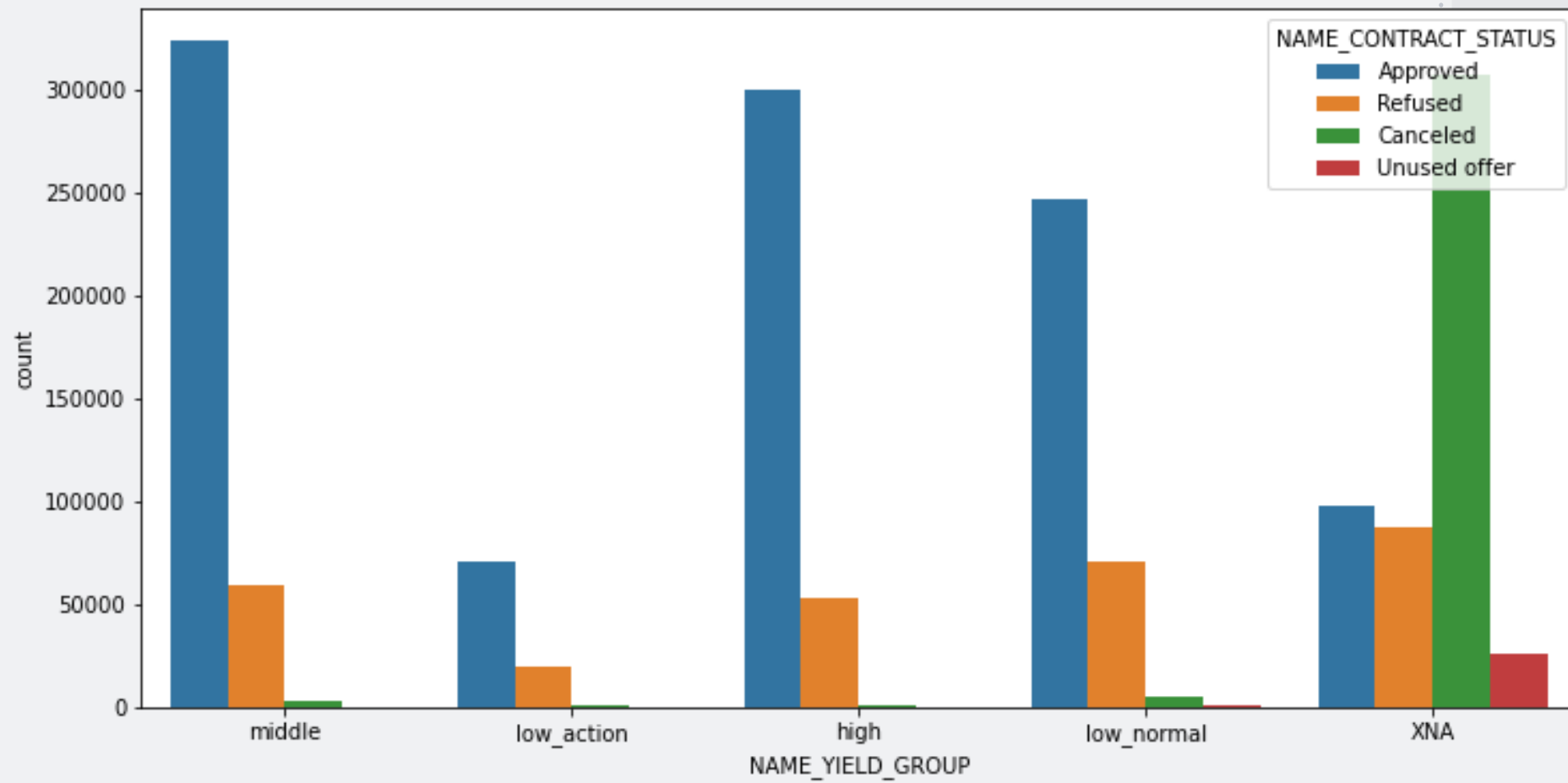
Inference – From the Heat Map it can be easily seen that lower secondary and civil marriage as well as separated have high impact



# Positive correlation between application amount and credit amount



Inference-Both Approval and refused rate is higher for repeater clients





Factors to decide whether an applicant will be

Defaulter:

- NAME\_FAMILY\_STATUS : People who are married are maximum number of loan applicants and widow are least
- NAME\_EDUCATION\_TYPE: People with Lower Secondary education have higher default rate
- NAME\_INCOME\_TYPE: Clients who are Unemployed default a lot.

Thank You

