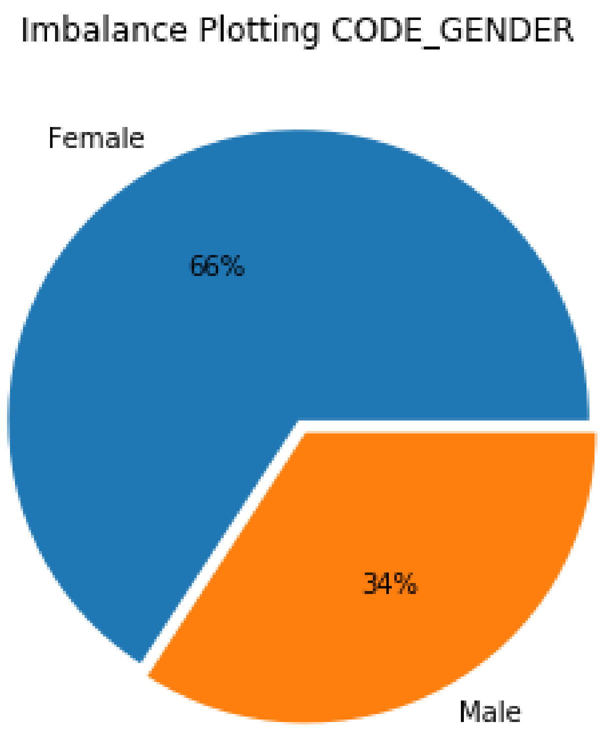
EDA Case Study

LOAN DEFAULTS AND GENERAL LENDING ANALYSIS

Data Imbalance

Given data has higher number of female applicants.

This should be kept in mind through the next slides



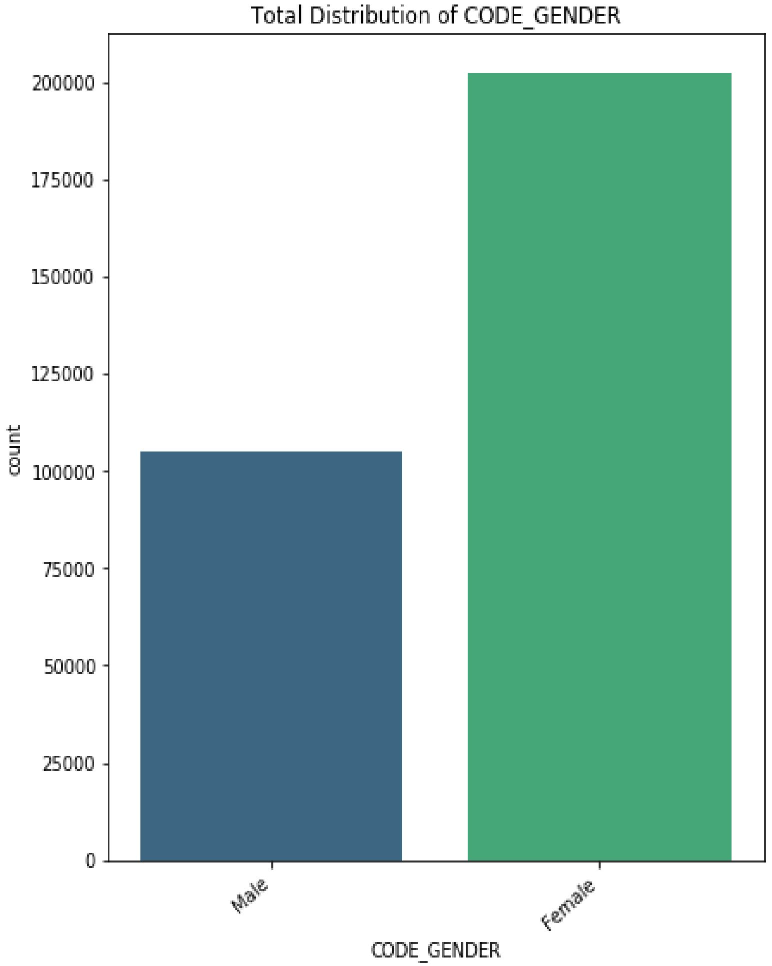
Univariate Analysis on New applications

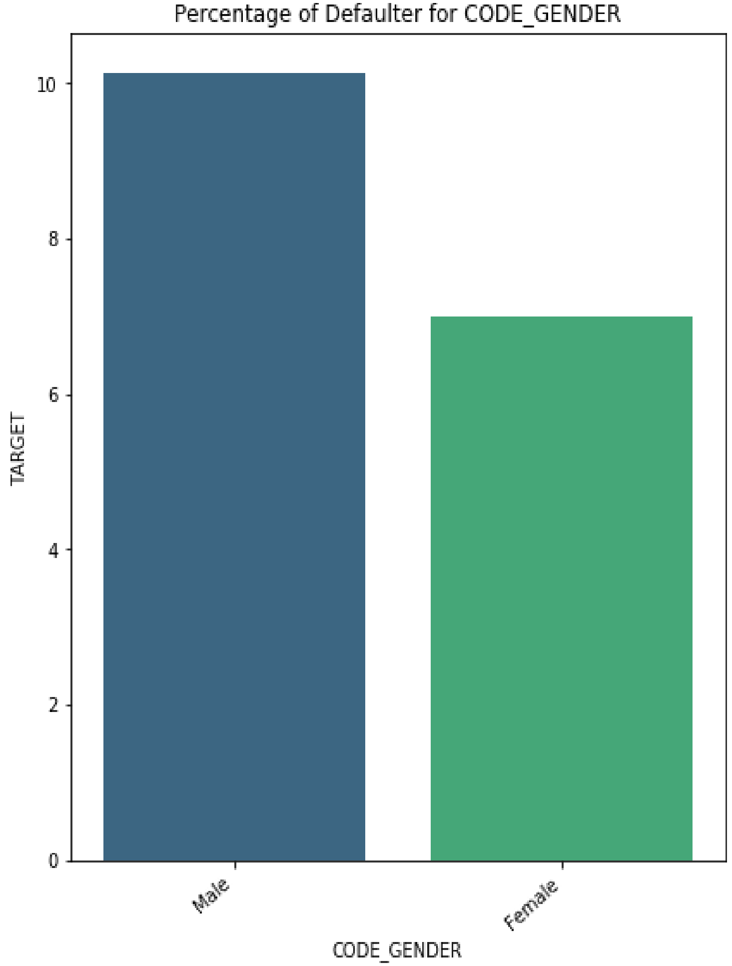
The following few slides would explore various parameters which seem to strongly influence the default rates for new applications in isolation

Gender Vs Defaults

New Apps-

Despite the higher number of female applicants, their percentage default is significantly smaller than males



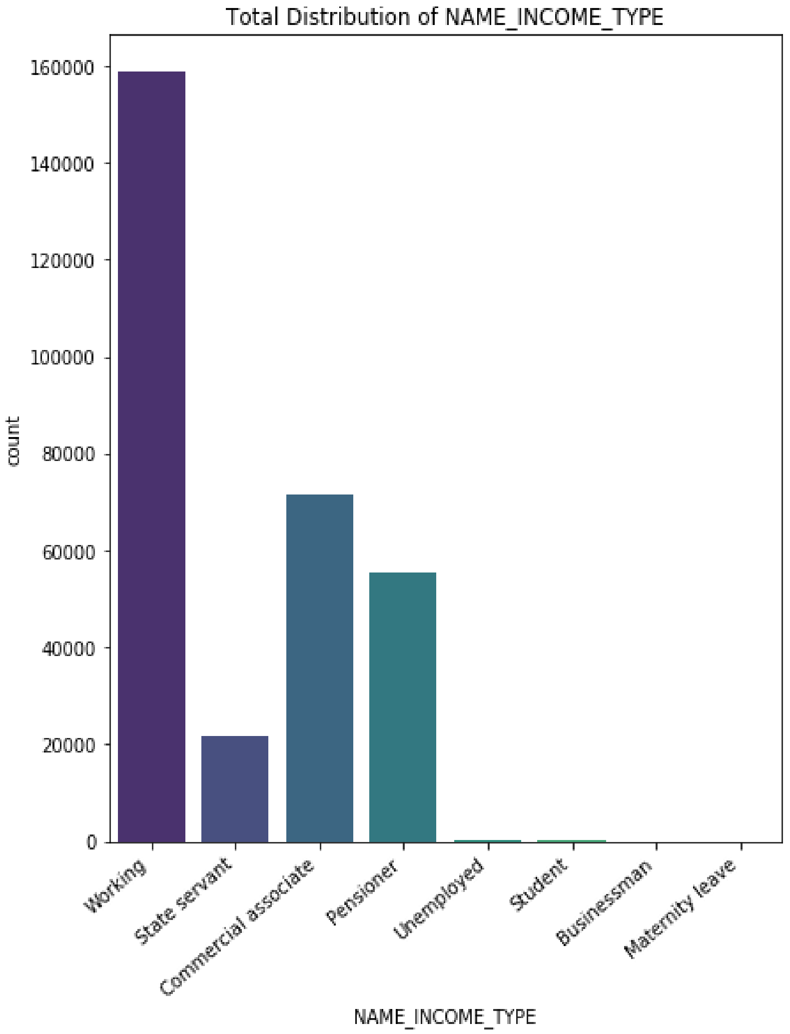


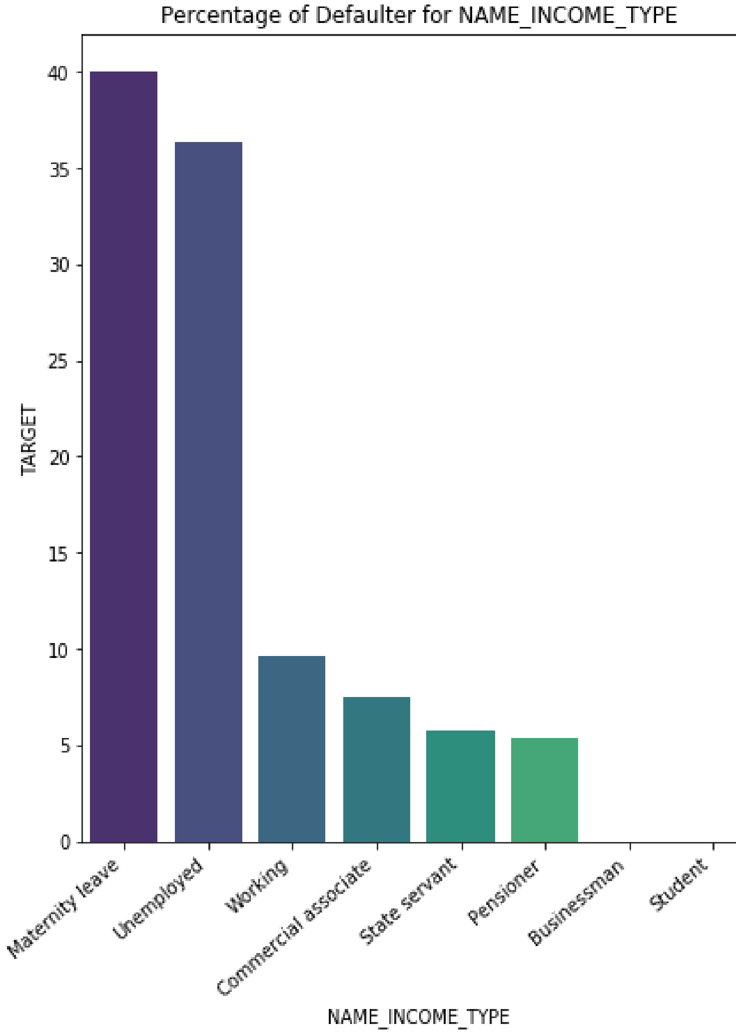
Client Income Type vs Default

New Apps-

Client Income Type vs Default

Clients on maternity leave and unemployed clients default at a much higher rate





Me
Medium 27.414304 High 24.556195 Low 21.848649 VeryHigh 15.322379 VeryLow 10.858473

Income Group vs Defaults

New app data

Very high income groups have significantly lower rate of defaults

All other income groups seem to fair equally worse!

Bins used-

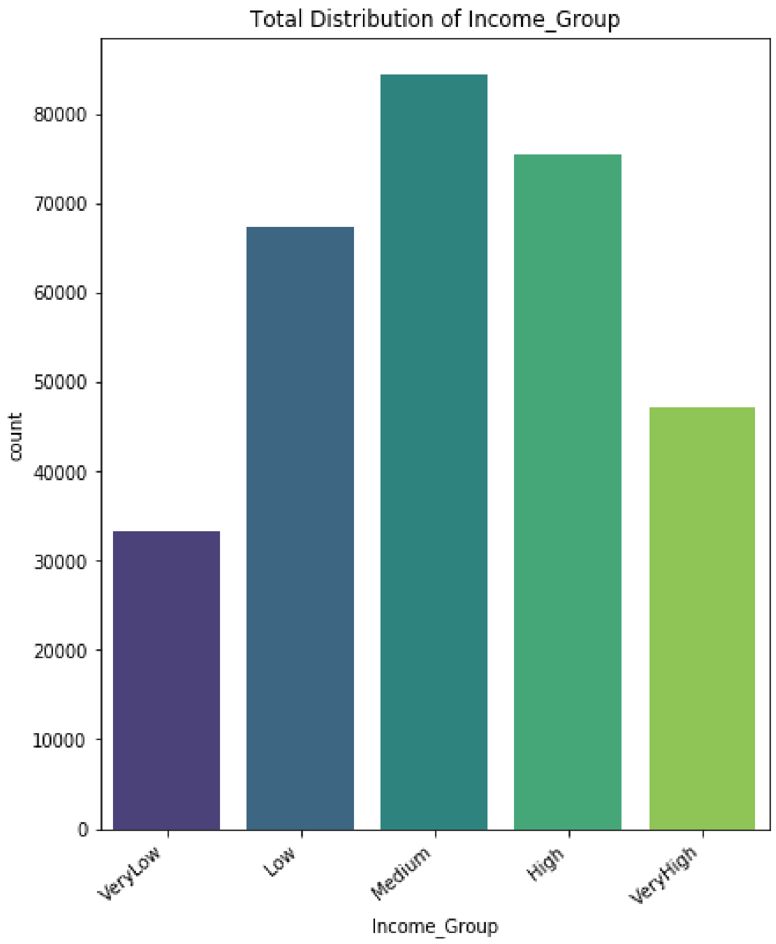
Medium 27.414304

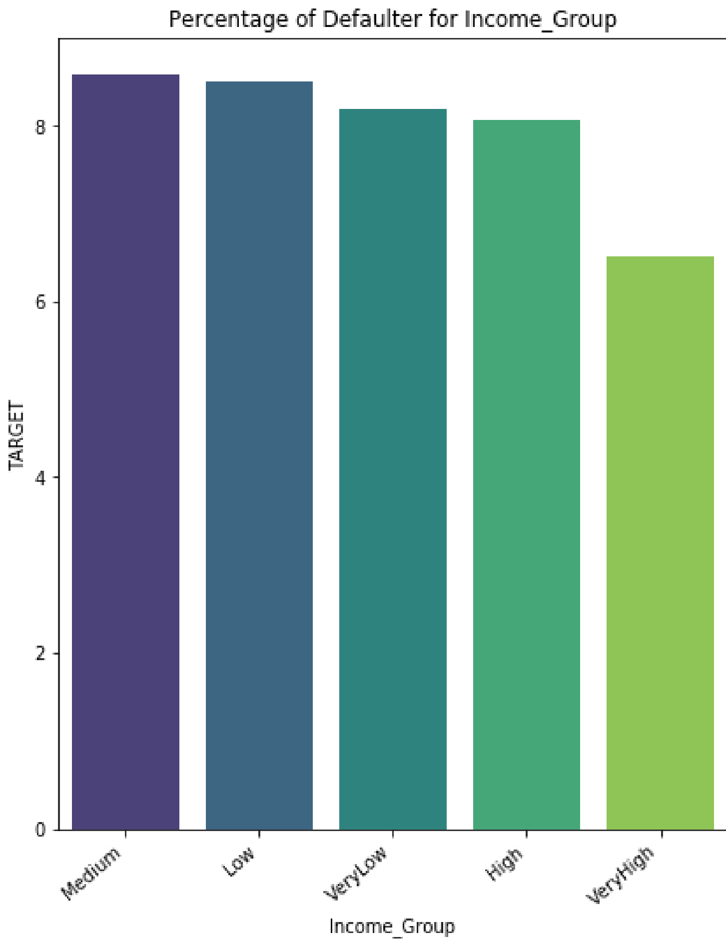
High 24.556195

Low 21.848649

VeryHigh 15.322379

VeryLow 10.858473



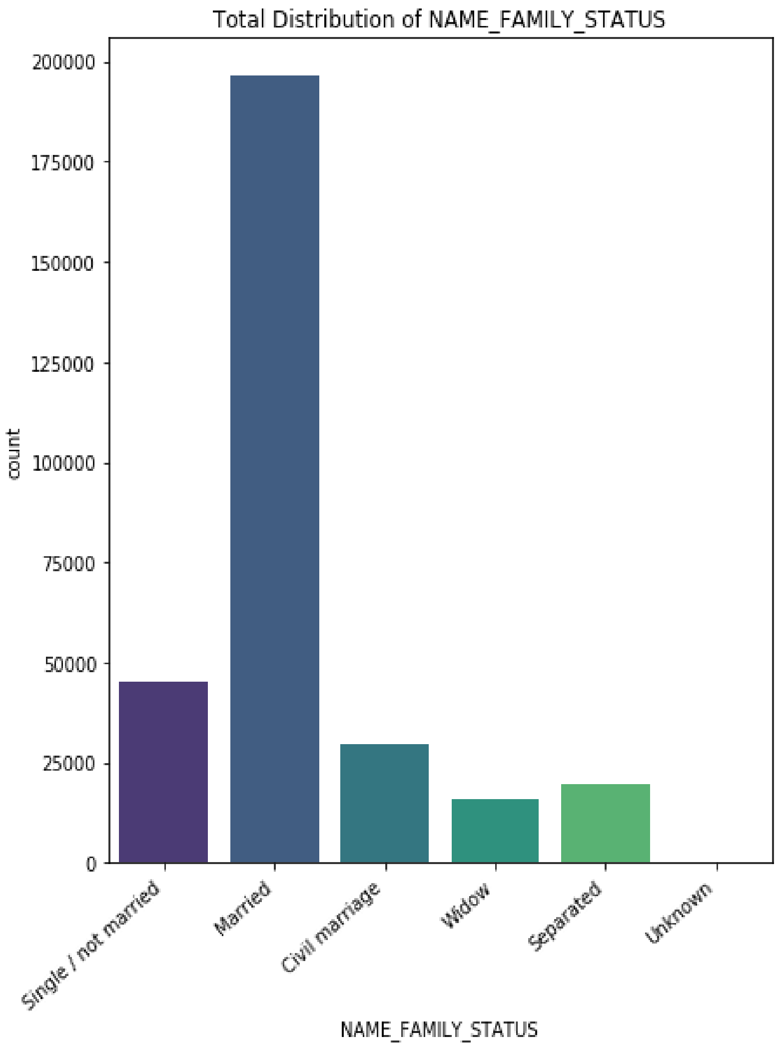


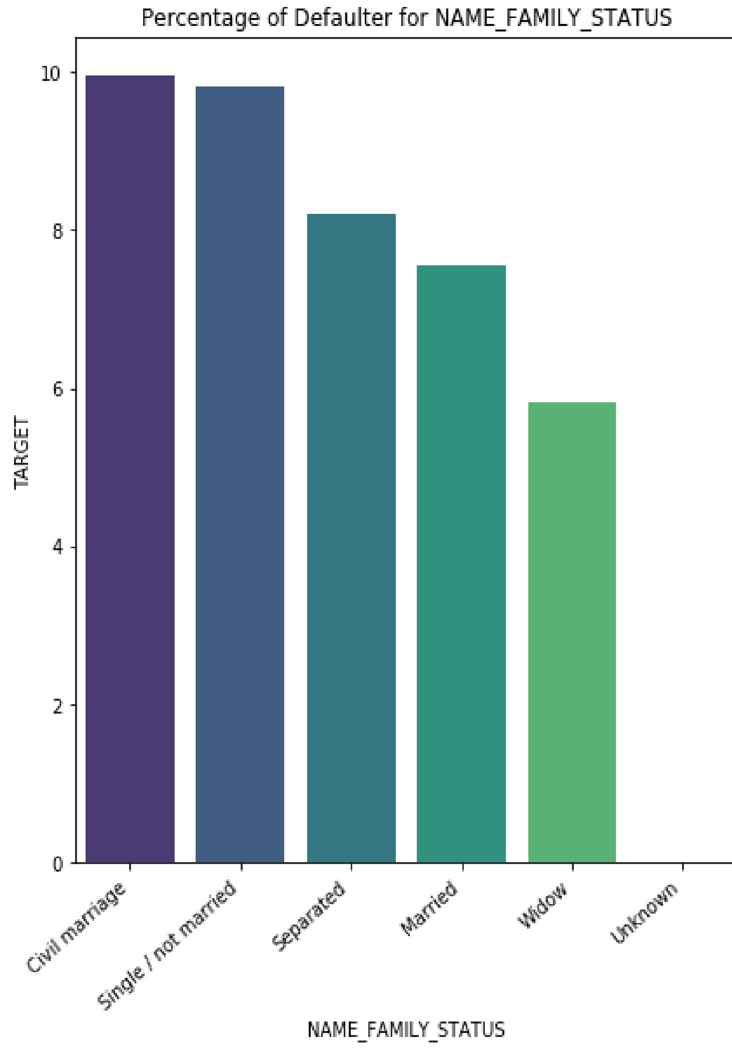
Family Status vs Defaults

New app data-

Civil married and single applicants default significantly more

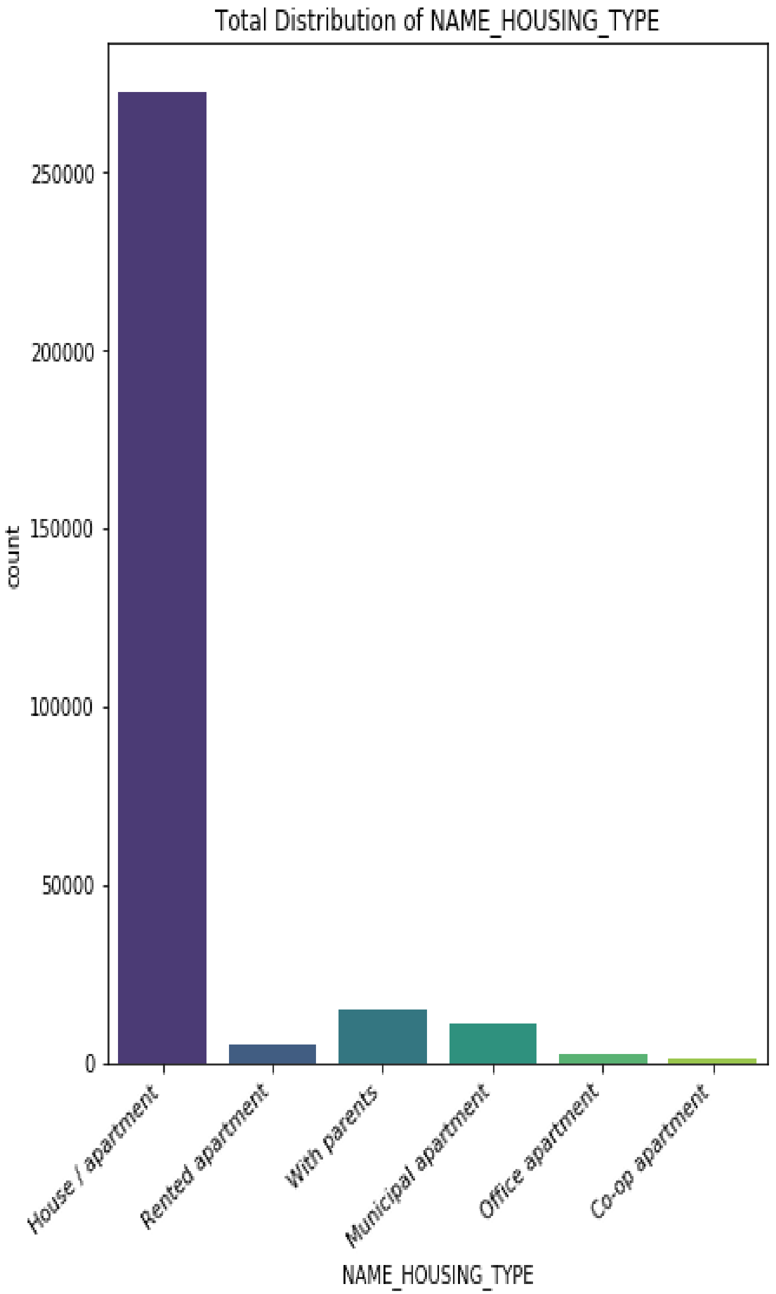
Widows and Married applicants seem relatively safe

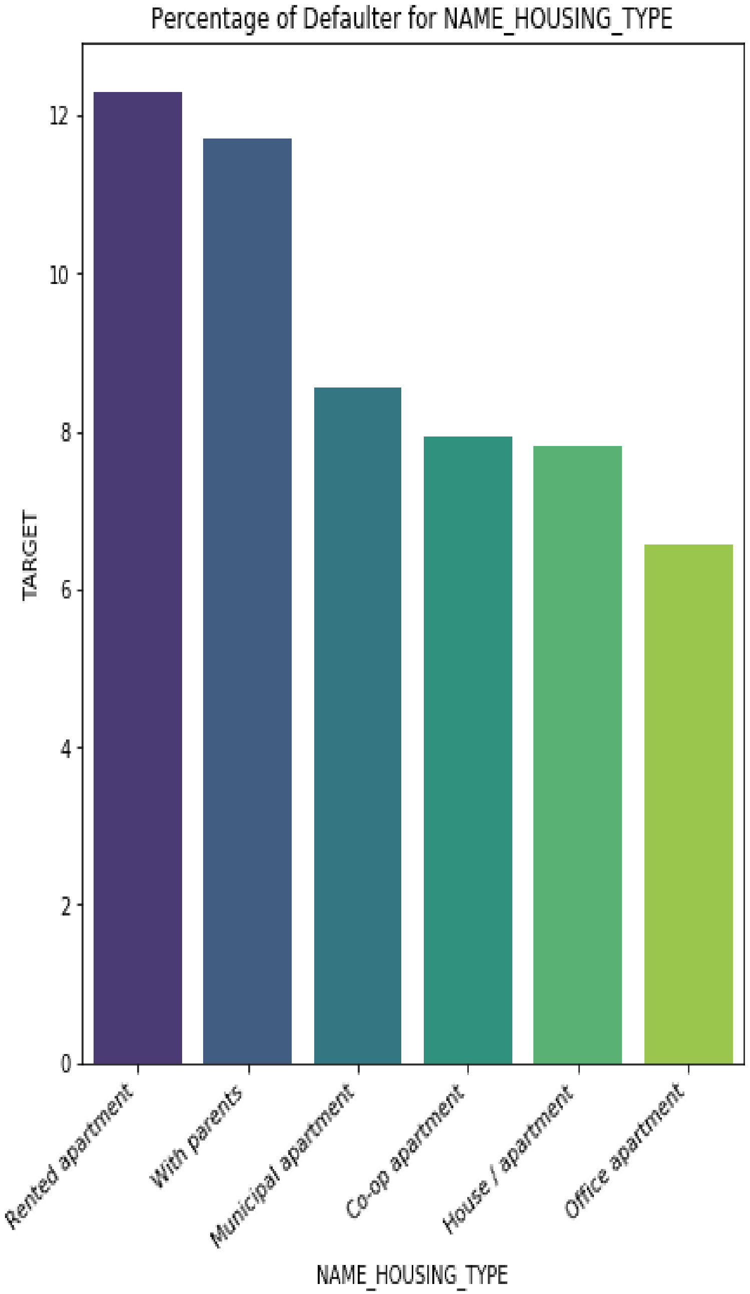




Housing Type vs Defaults

Renters and clients staying with parents default much more than clients living in their own house/apartments

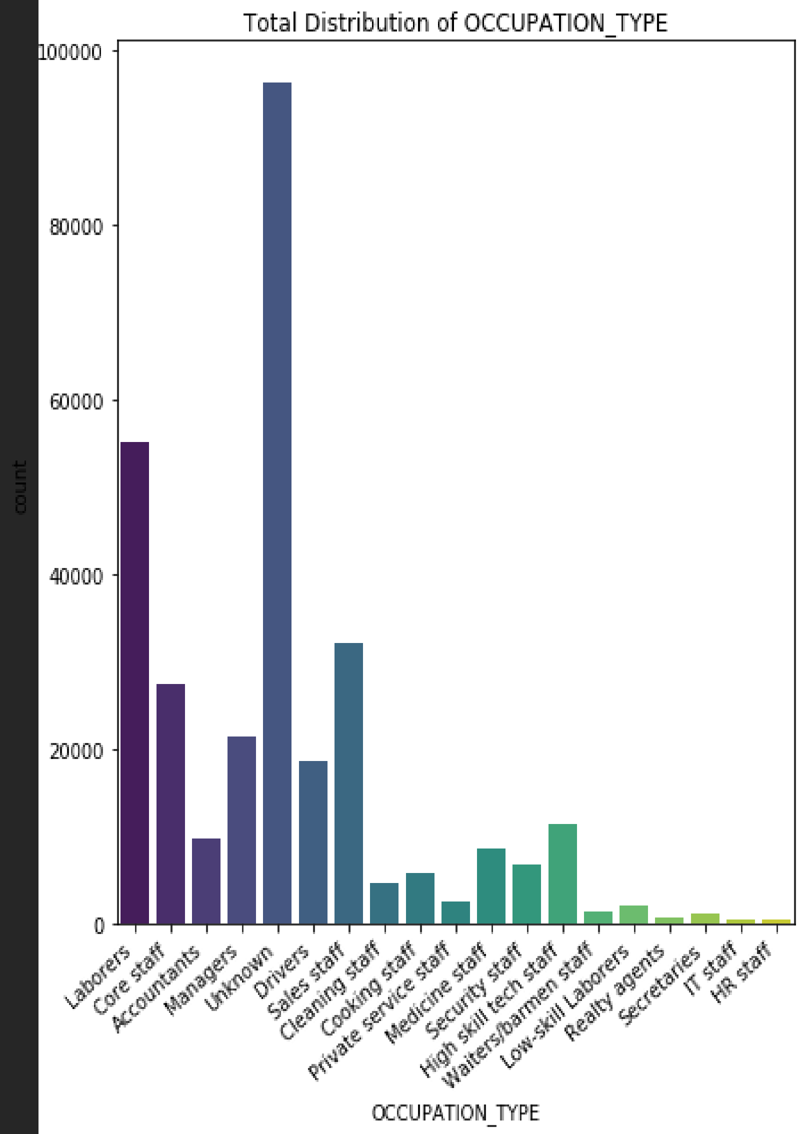


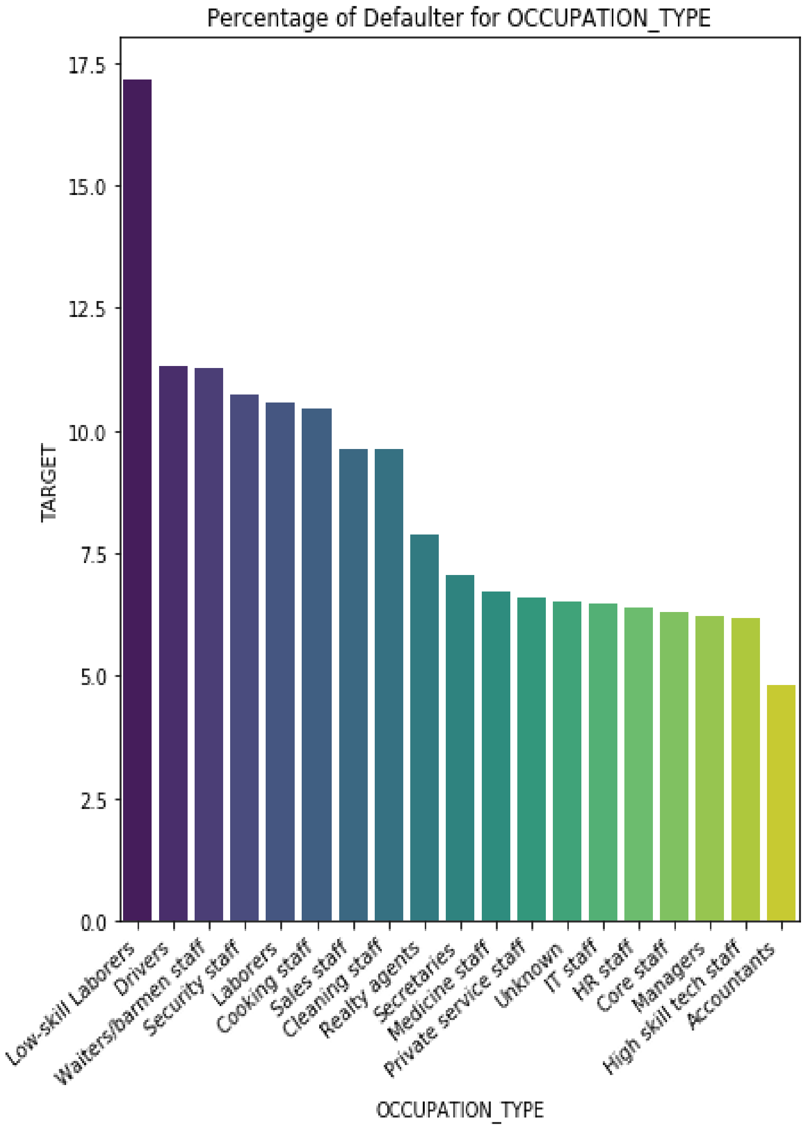


Occupation vs Defaults

Low skilled jobs-which typically pay less- are at a higher risk of defaulting

Higher skilled and managerial post holding applicants default much less, again possibly due to more disposable income due to higher salaries

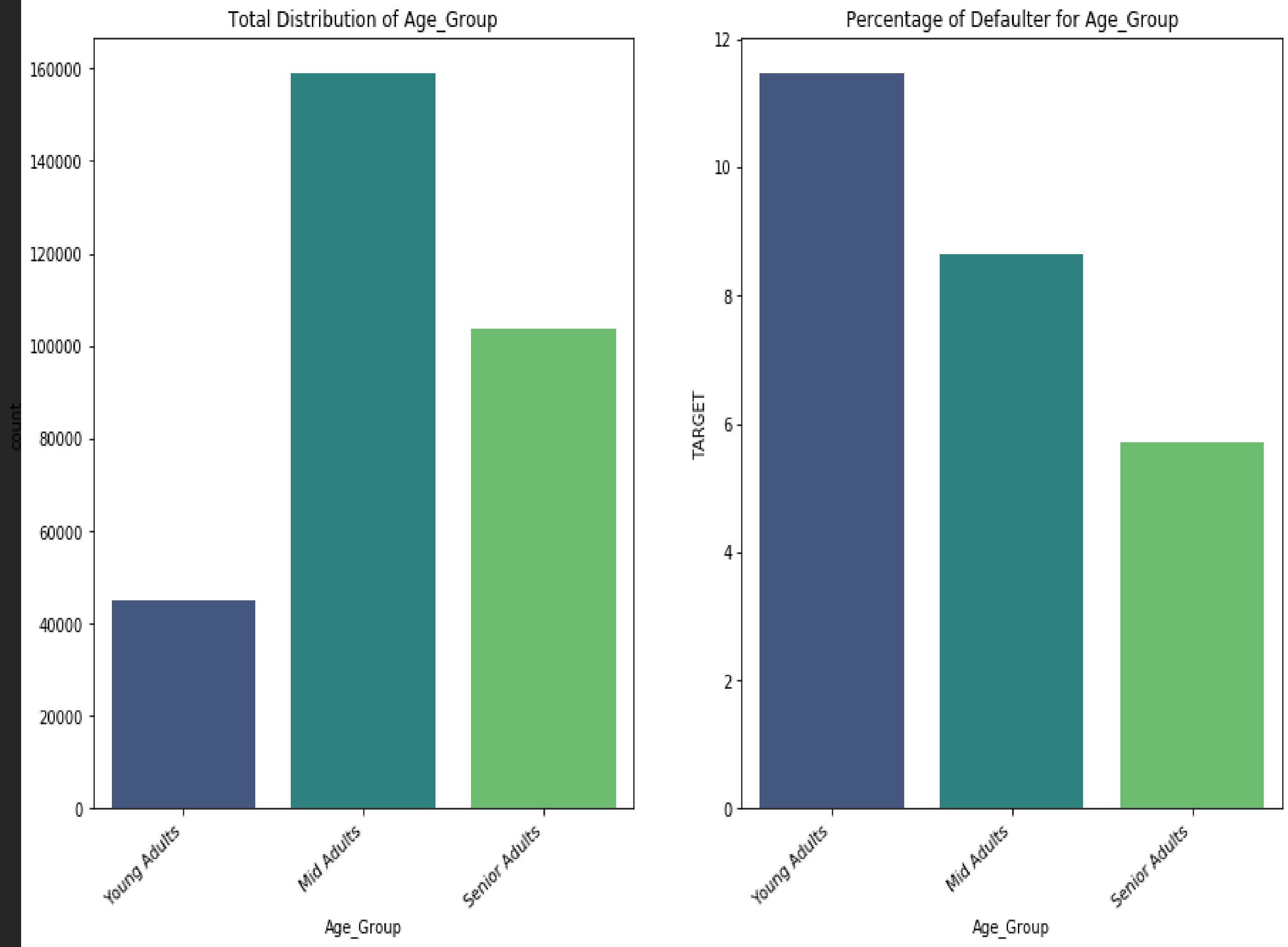




Age Groups vs Defaults

Young clients default at a higher rate. They are high risk group.

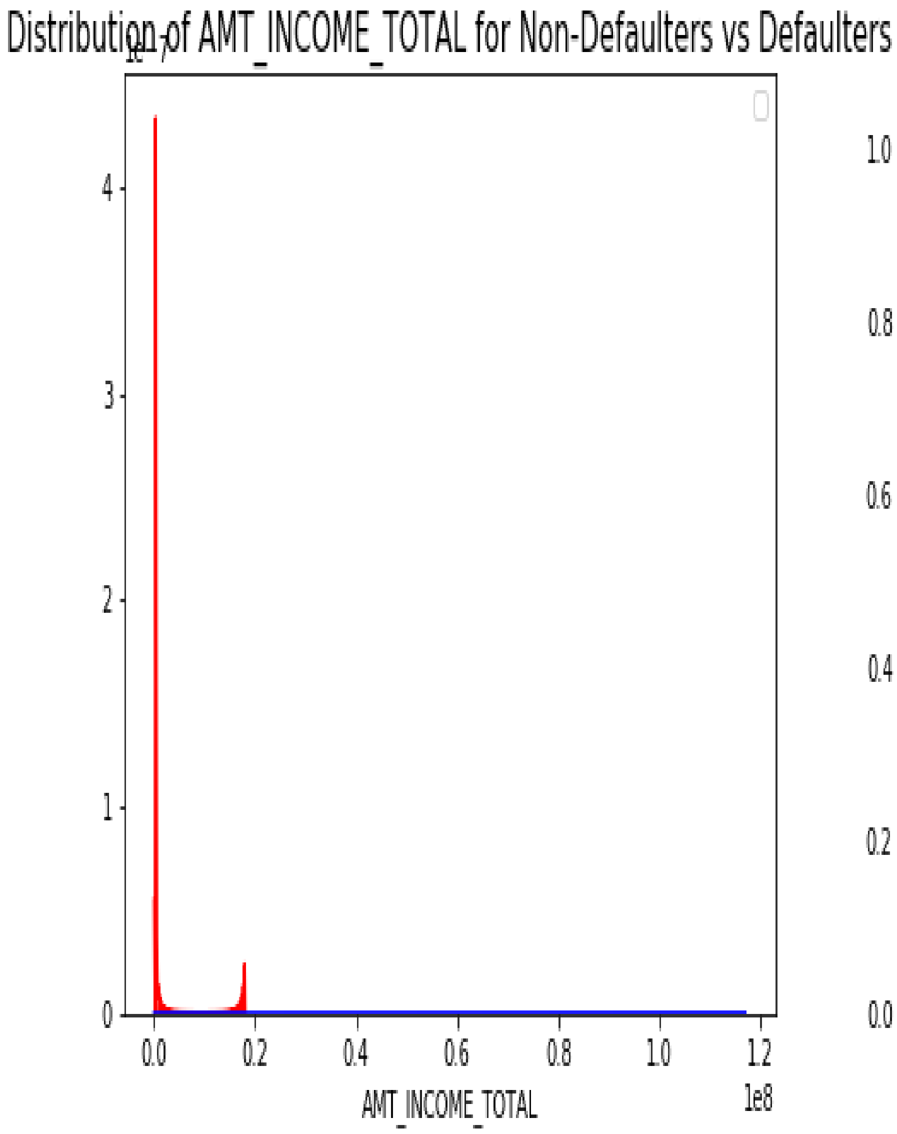
Seniors are safer option especially given they have more loans compared to young adults

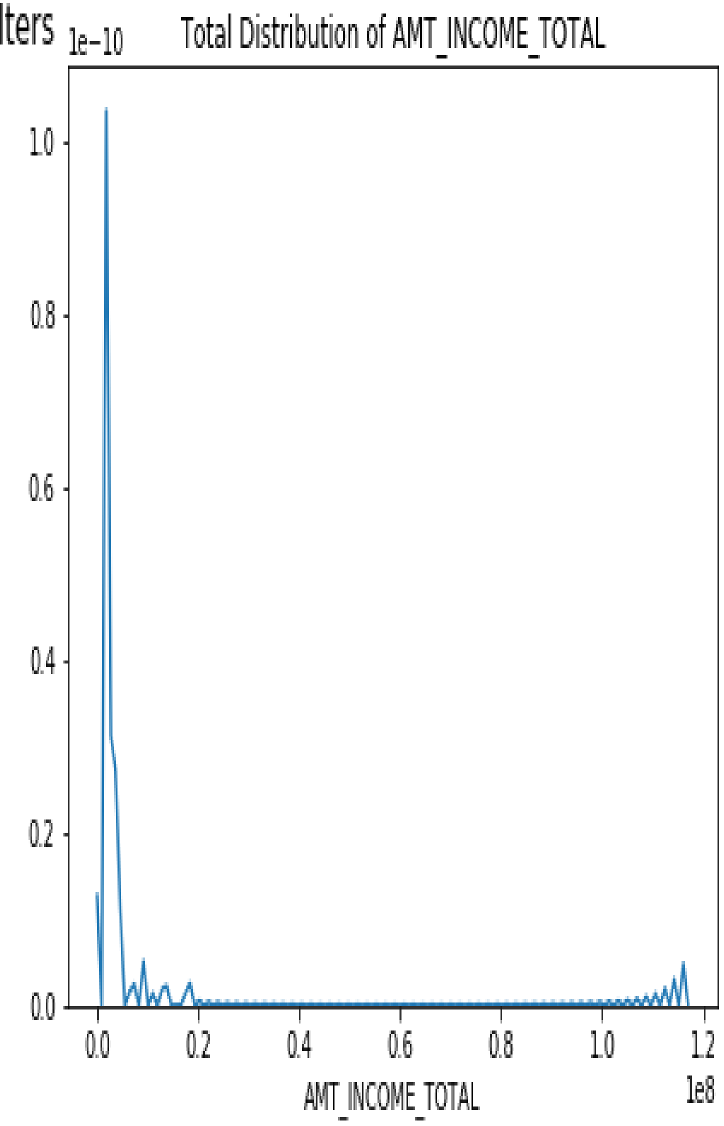


Total Income Vs Defaults

Lower income/unemployed groups

Are higher risk as they default more. This ties with earlier income group plot

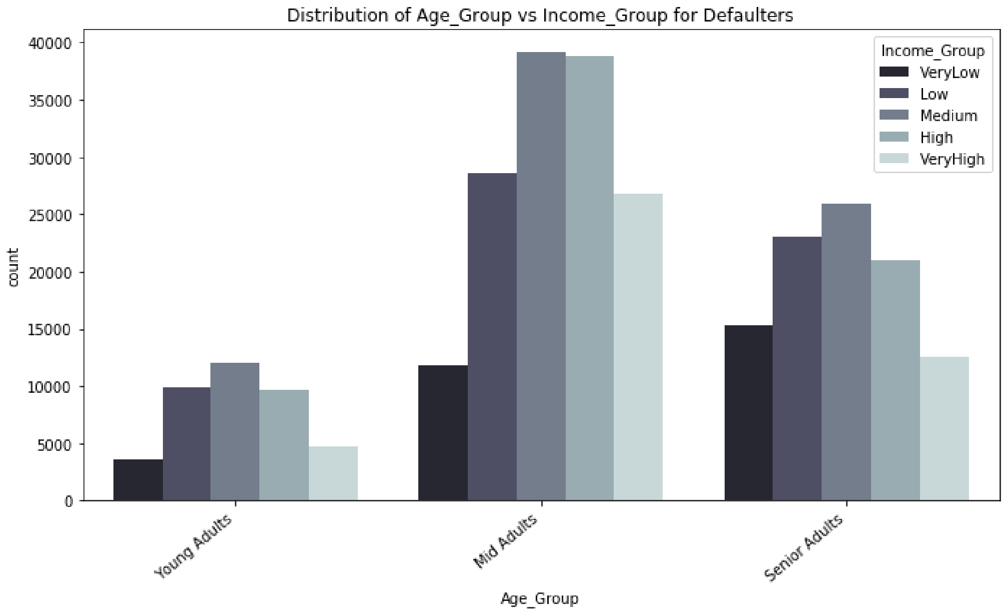


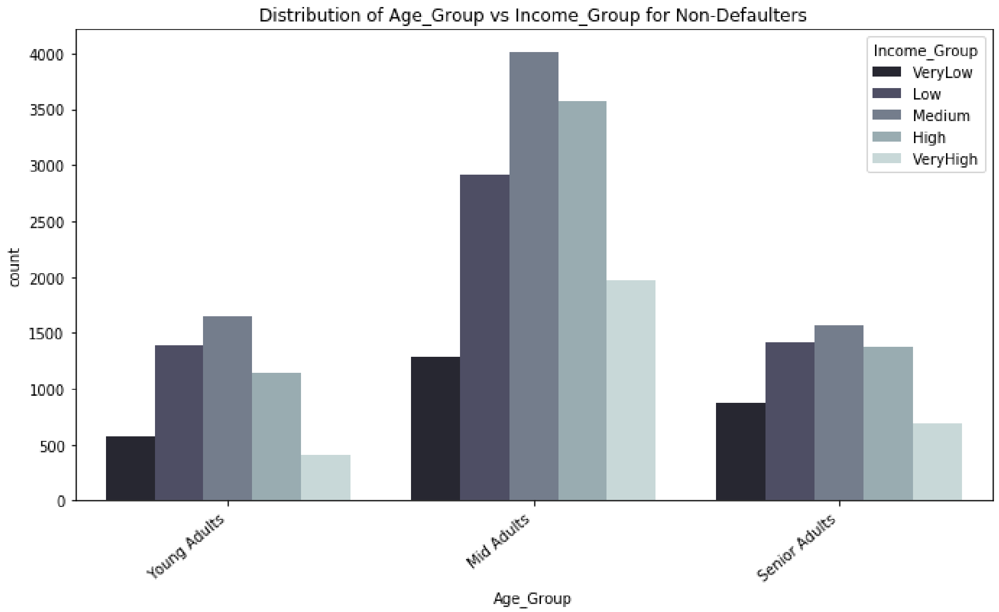


Segmented Analysis

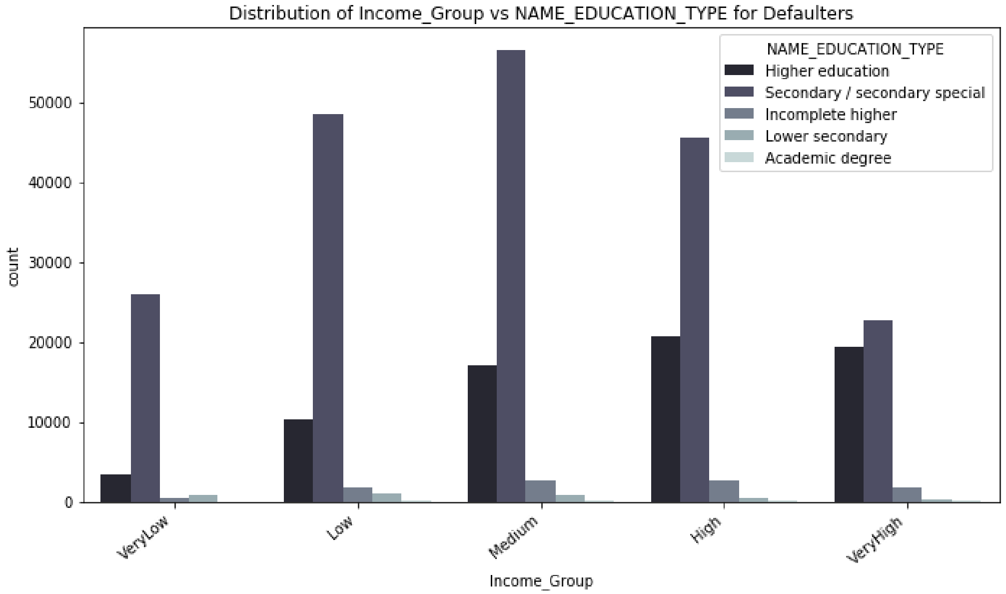
The following few slides show defaulters segmented across multiple attributes, for eg what is the rate of default for senior adults across their income

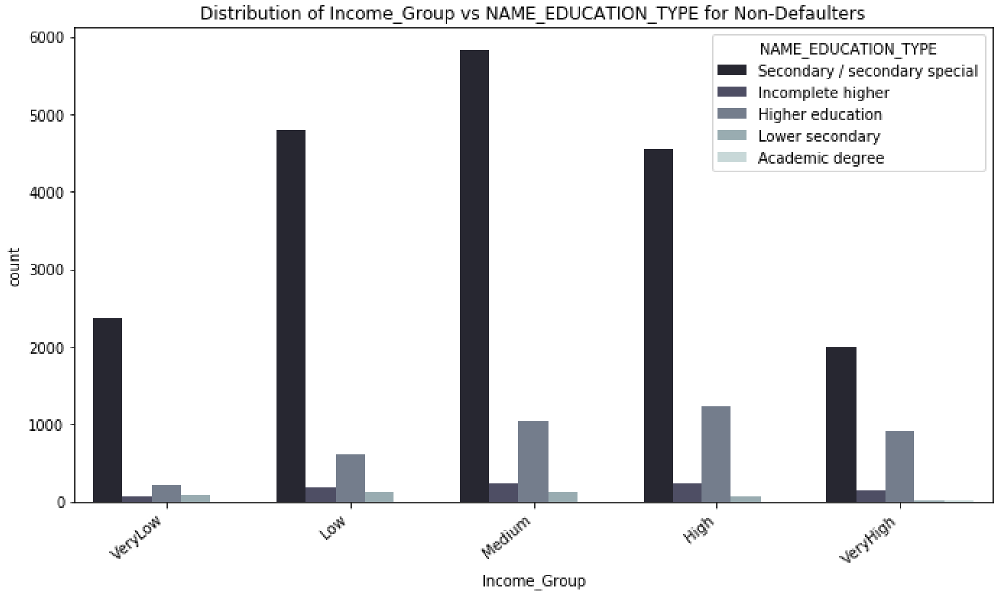
Age vs Income



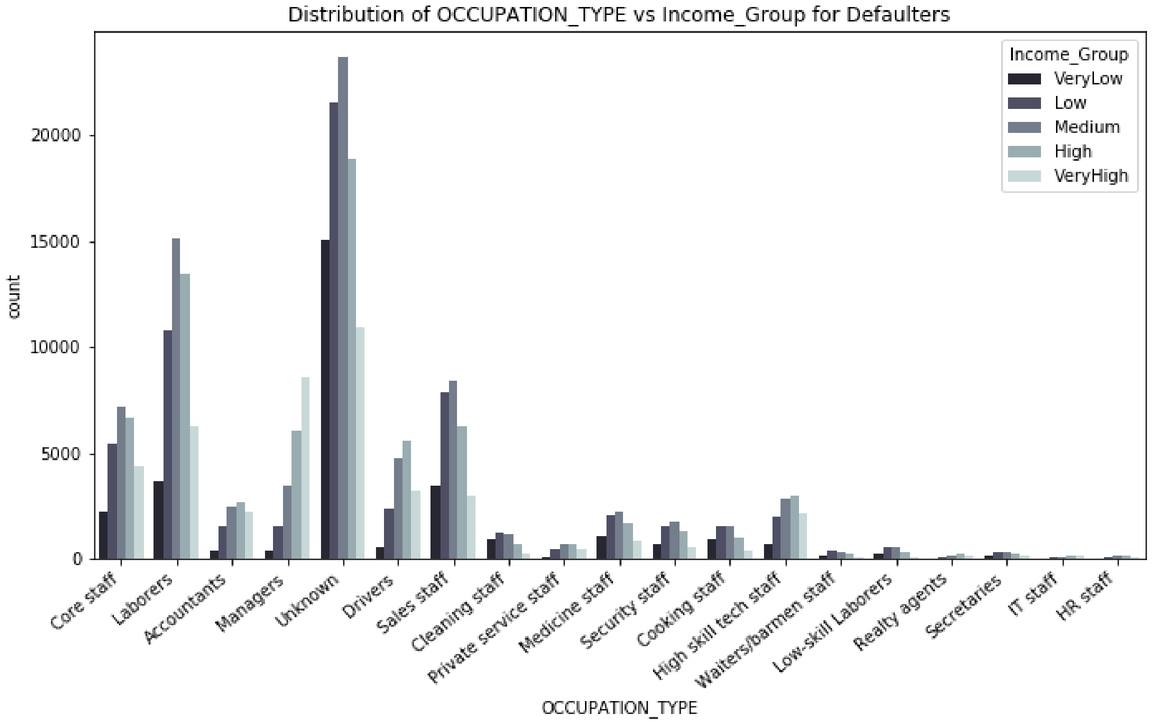


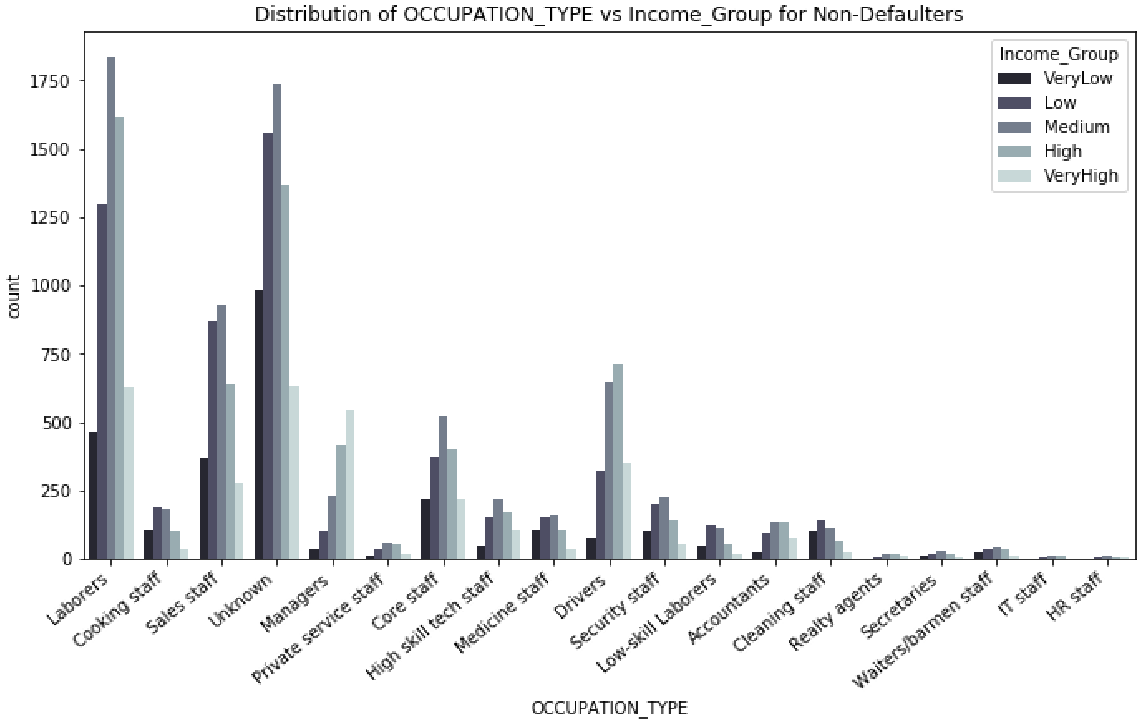
Income vs Education





Income vs Occupation





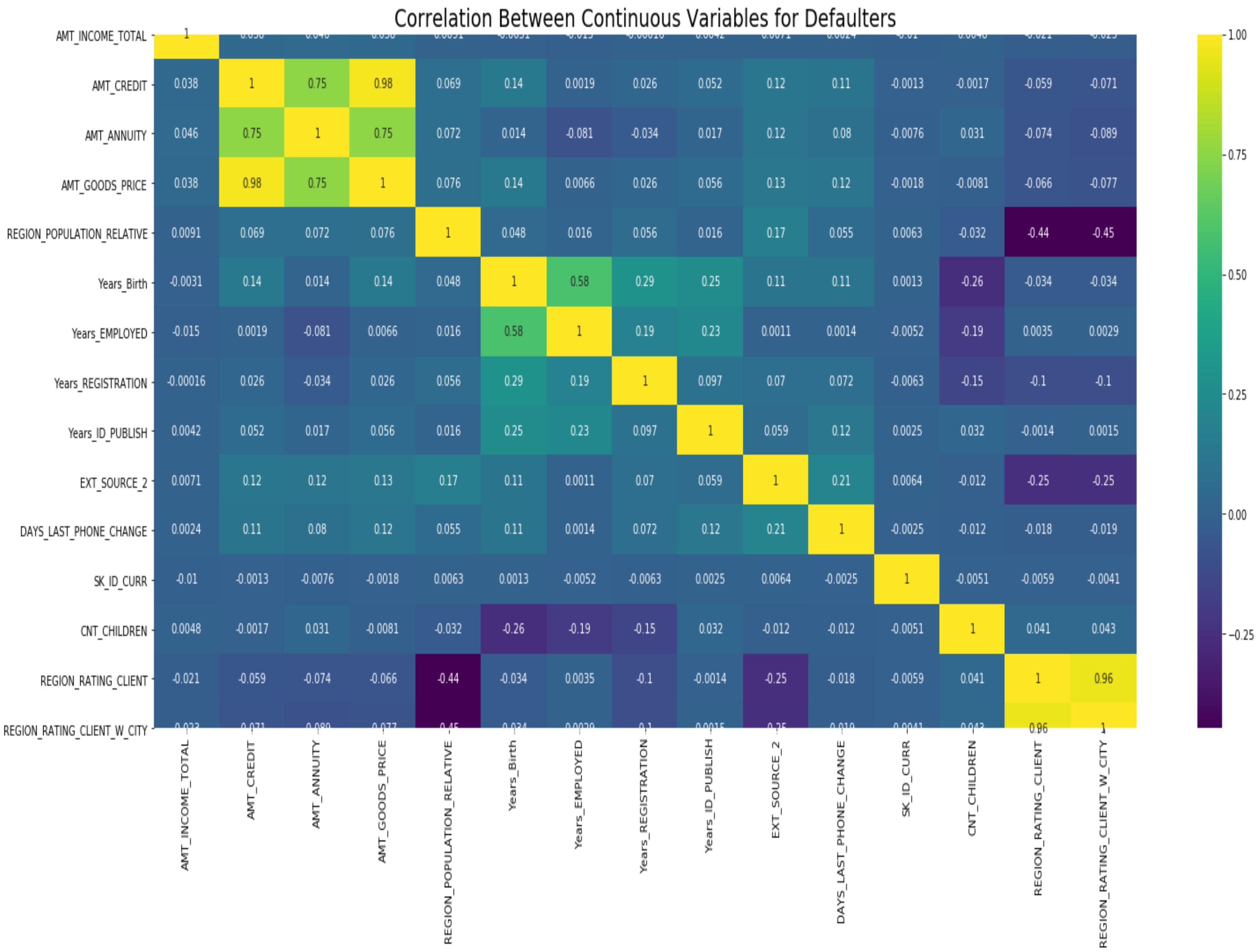
Bivariate Analysis

The following few slides explore the correlation between various attributes for applicants which have defaulted

Numerical attributes Correlation

High correlation between Amt_goods_price and amount_annuity

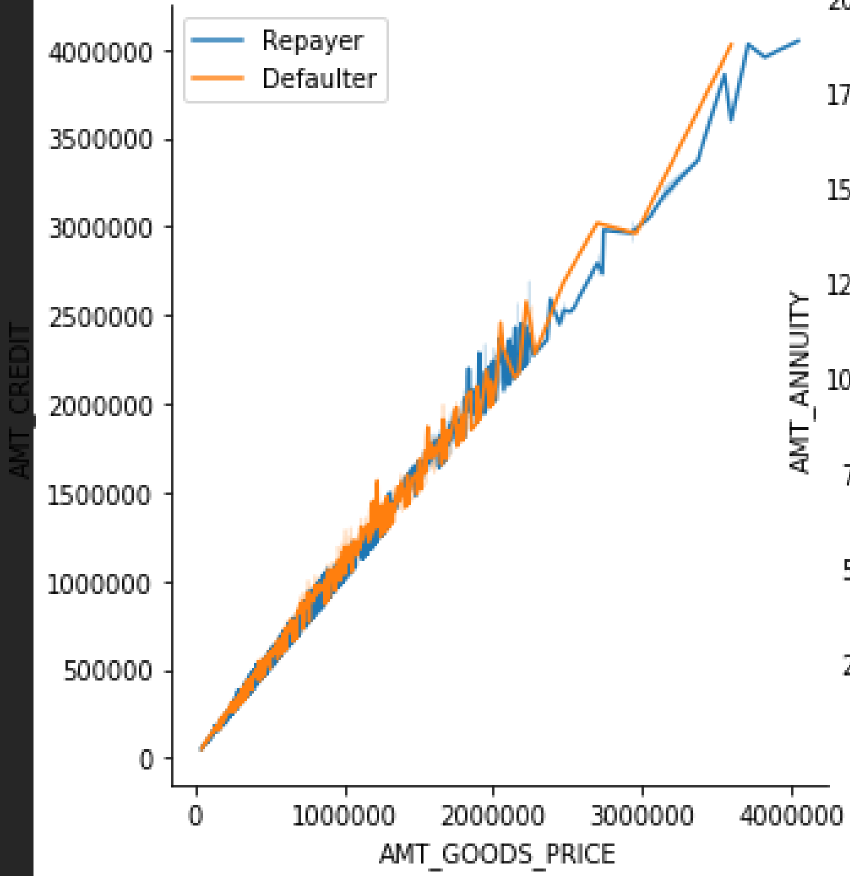
High correlation between Amt_credit and amount_annuity

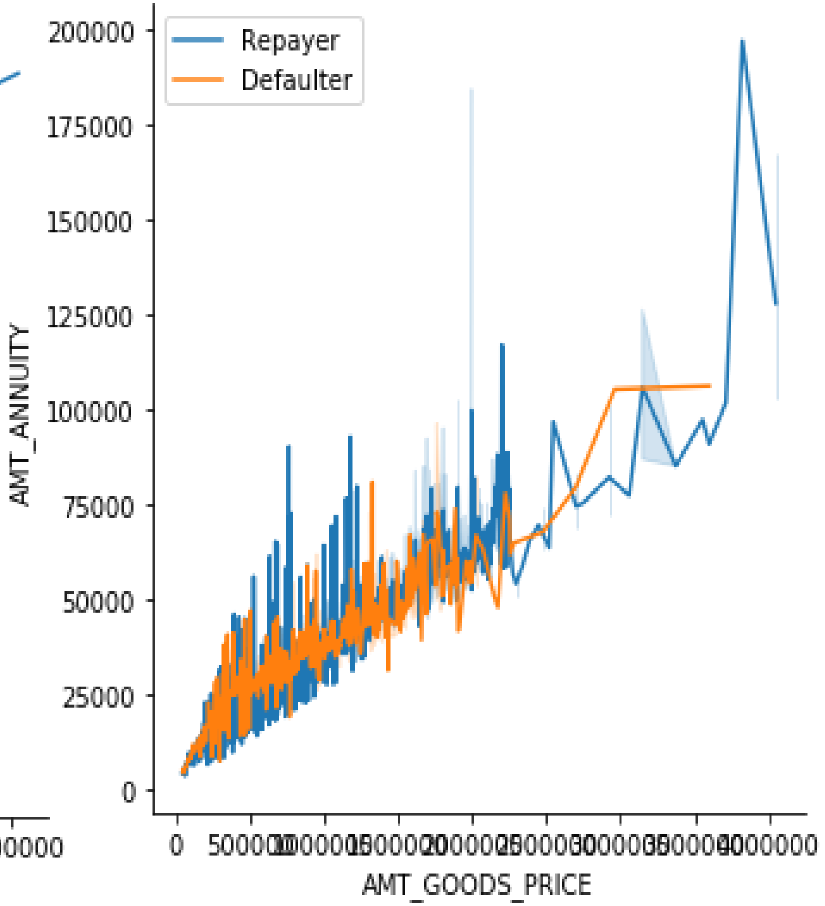


Defaulter vs Repayer for the correlation noted in previous slide

Same correlation seen for Amt_goods_price vs Amt_credit for both repayer and defaulter so not much can be concluded

Same correlation seen for Amt_annuity vs Amt_goods_price for both repayer and defaulter so not much can be concluded

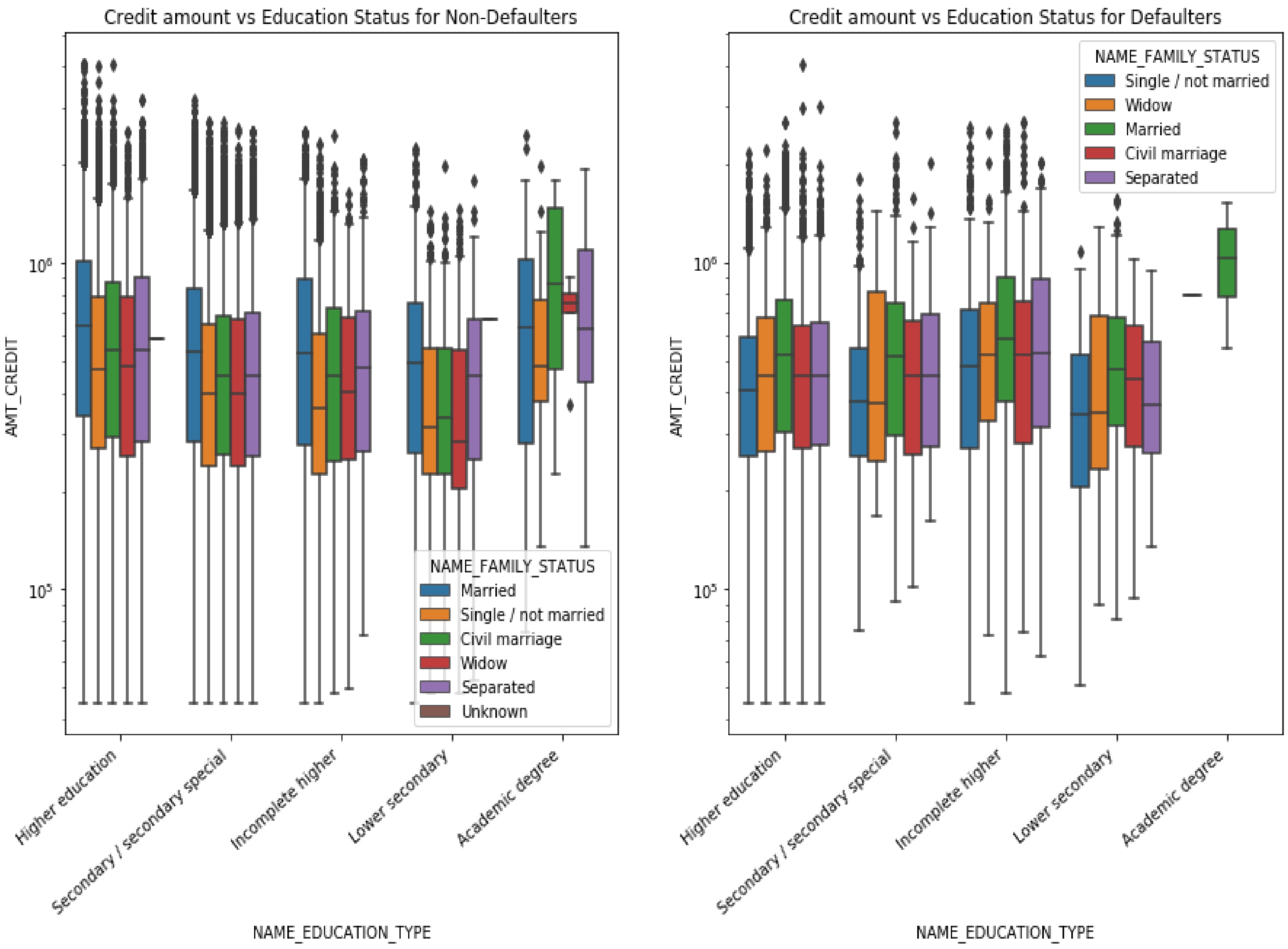




Family Status vs Education vs Credit Amount

Civil marriage, marriage and separated holding a degree get higher credit

Highly educated clients who ate married/single have high outliers

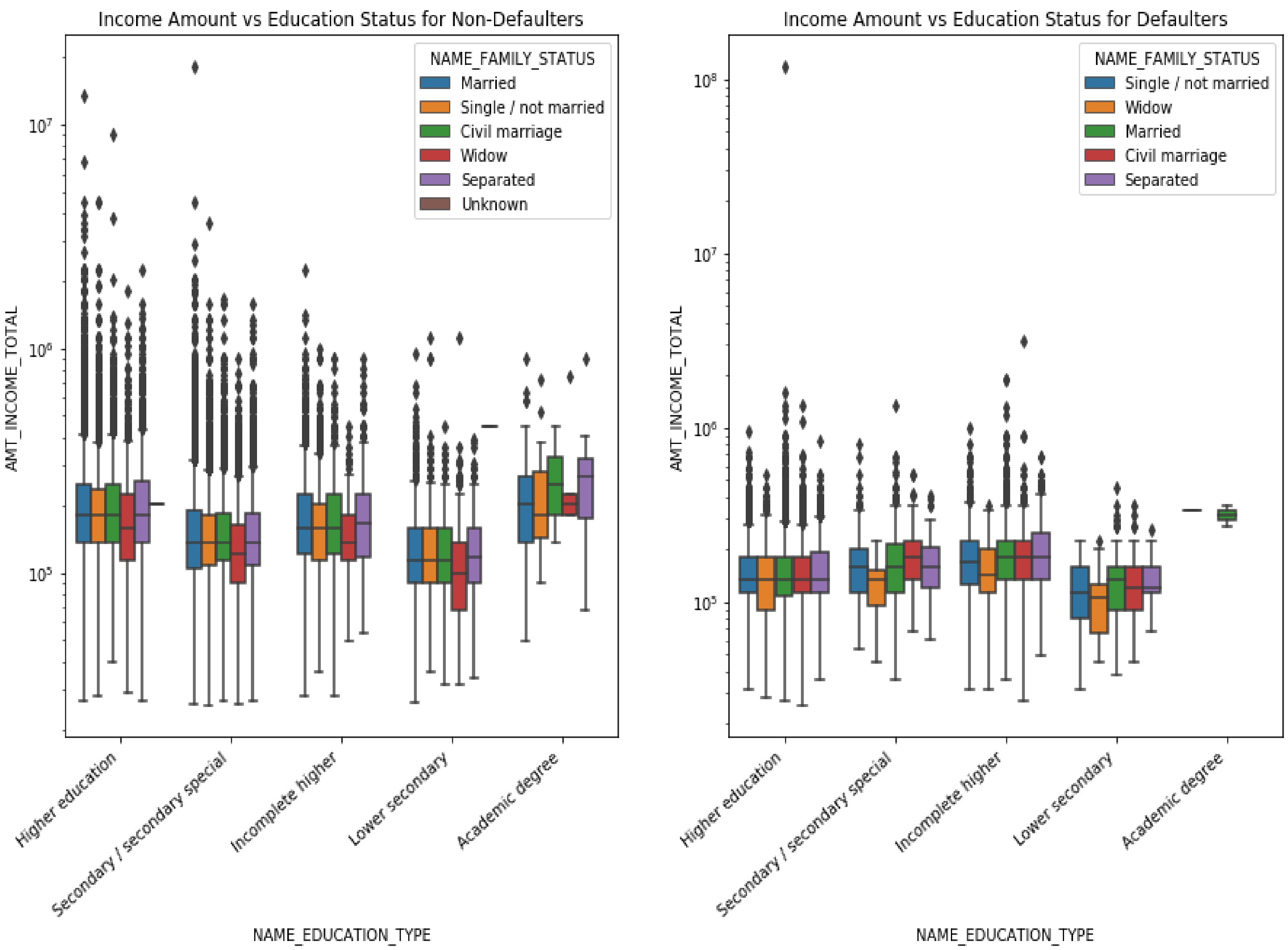


Income vs Education vs Family Status

Lower secondary educated group has lower income which is expected

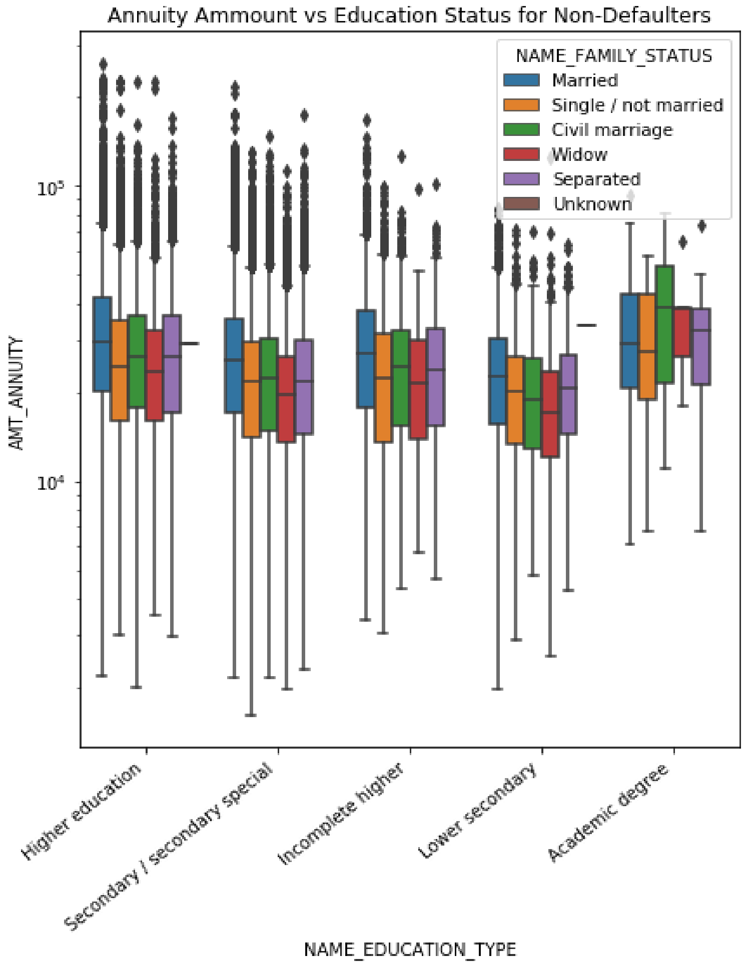
Higher education group has more outliers which is expected as many high earners would be highly educated

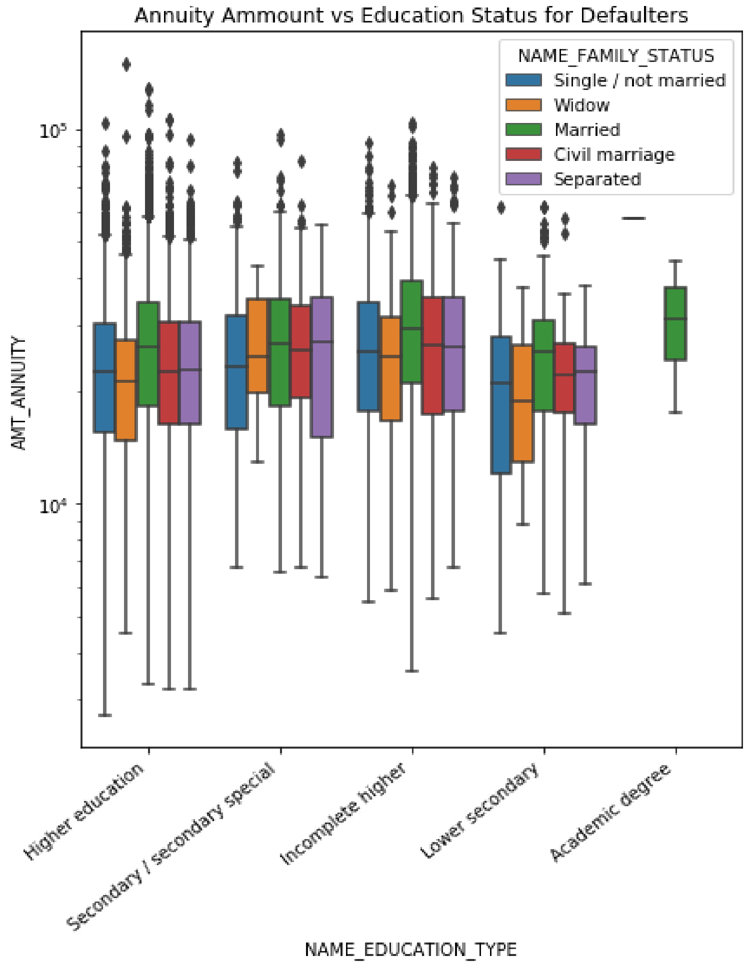
Few outliers for clients holding an academic degree



Annuity Amount vs Education vs Family Status

Similar tends as seen in previous slide



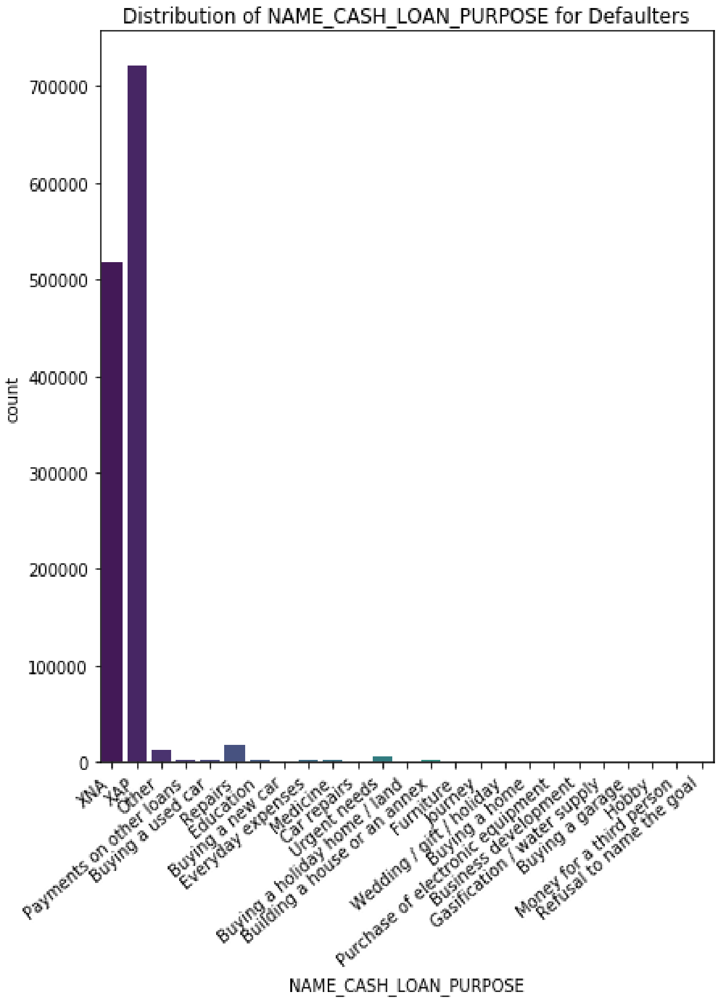


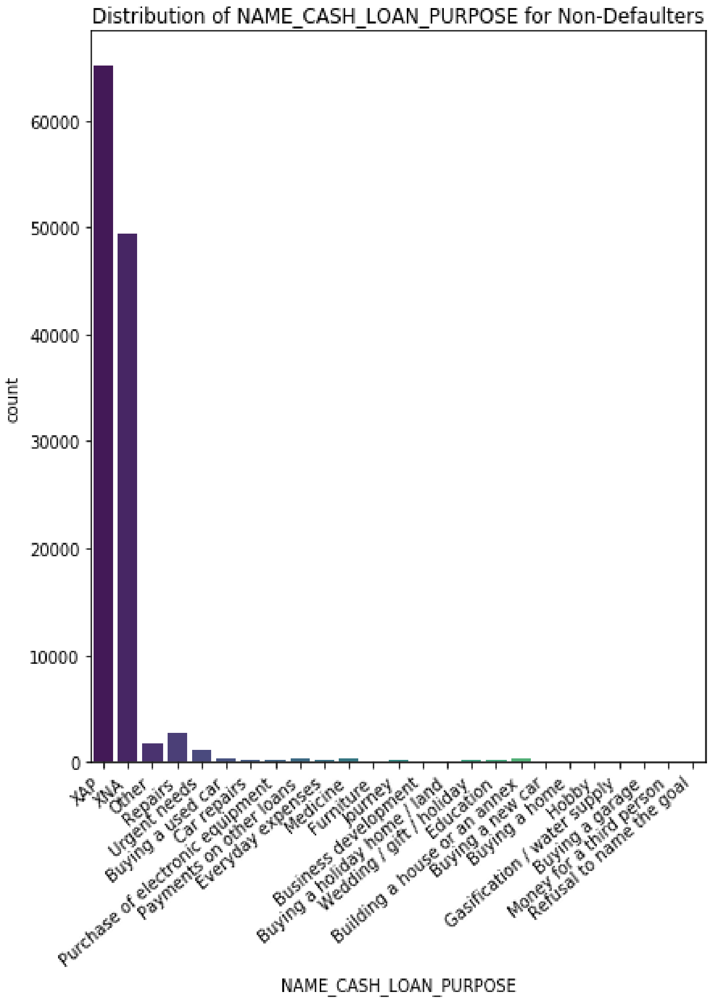
Univariate analysis on merged Data

Following few slides explore attributes after merging new and previous application data for each client

Loan Purpose vs Defaults

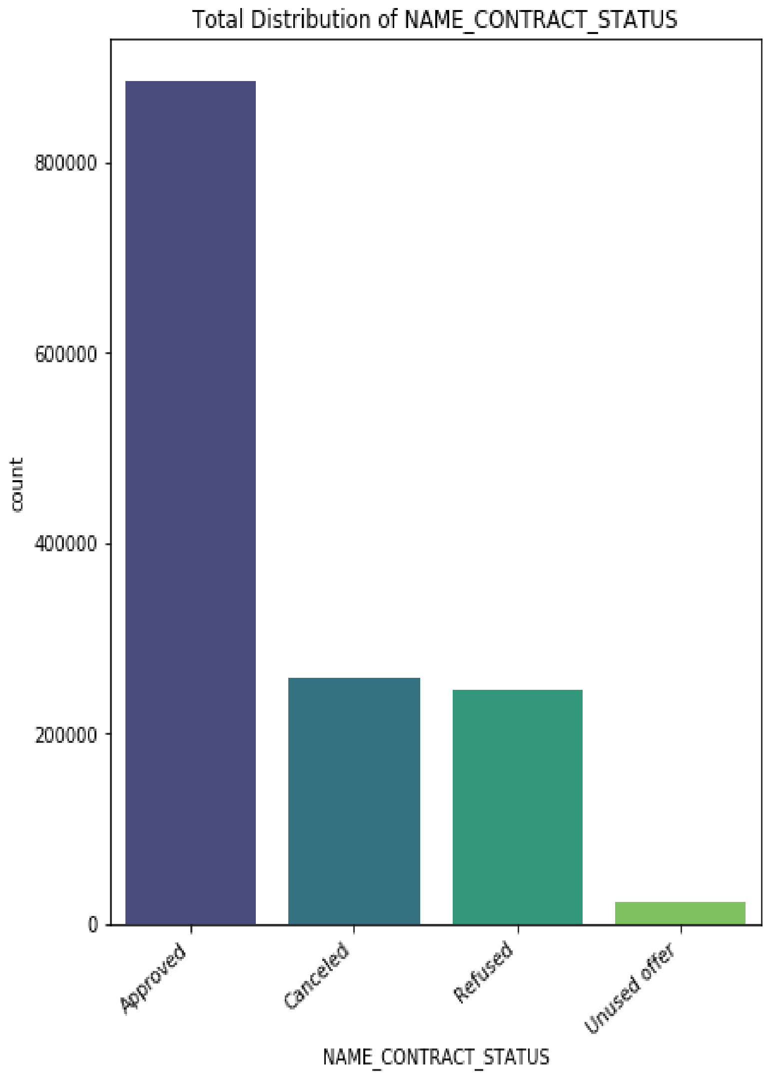
Large number of loans in the data set donot specify a purpose and are marked as XNA/XAP

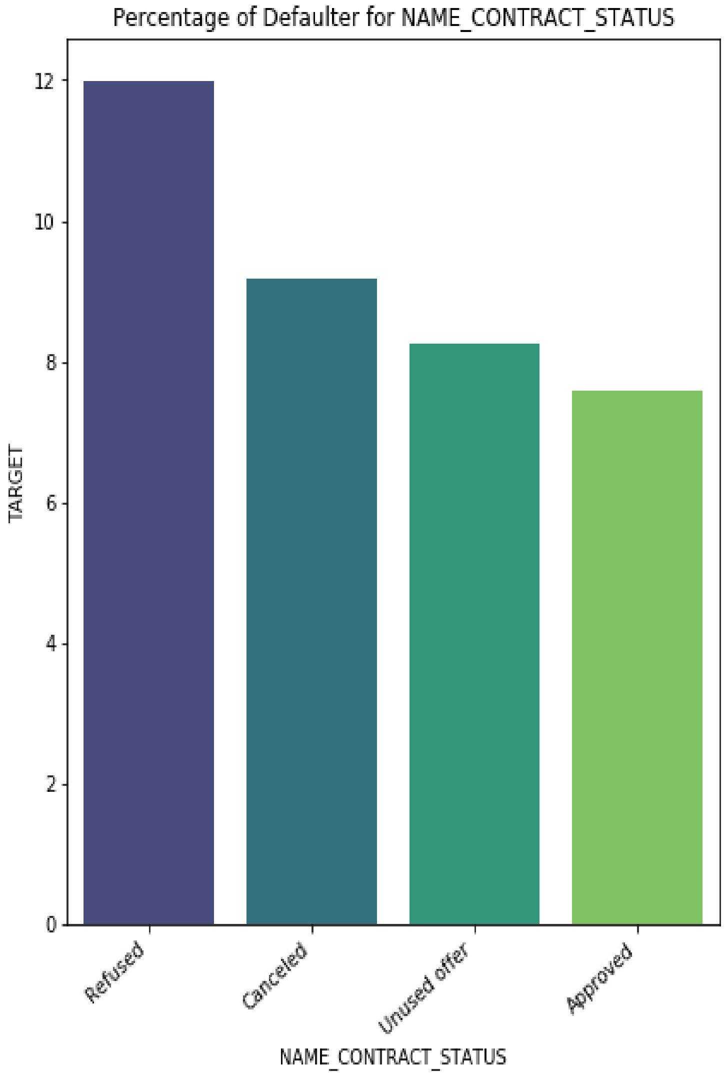




Previous loan status vs Defaults

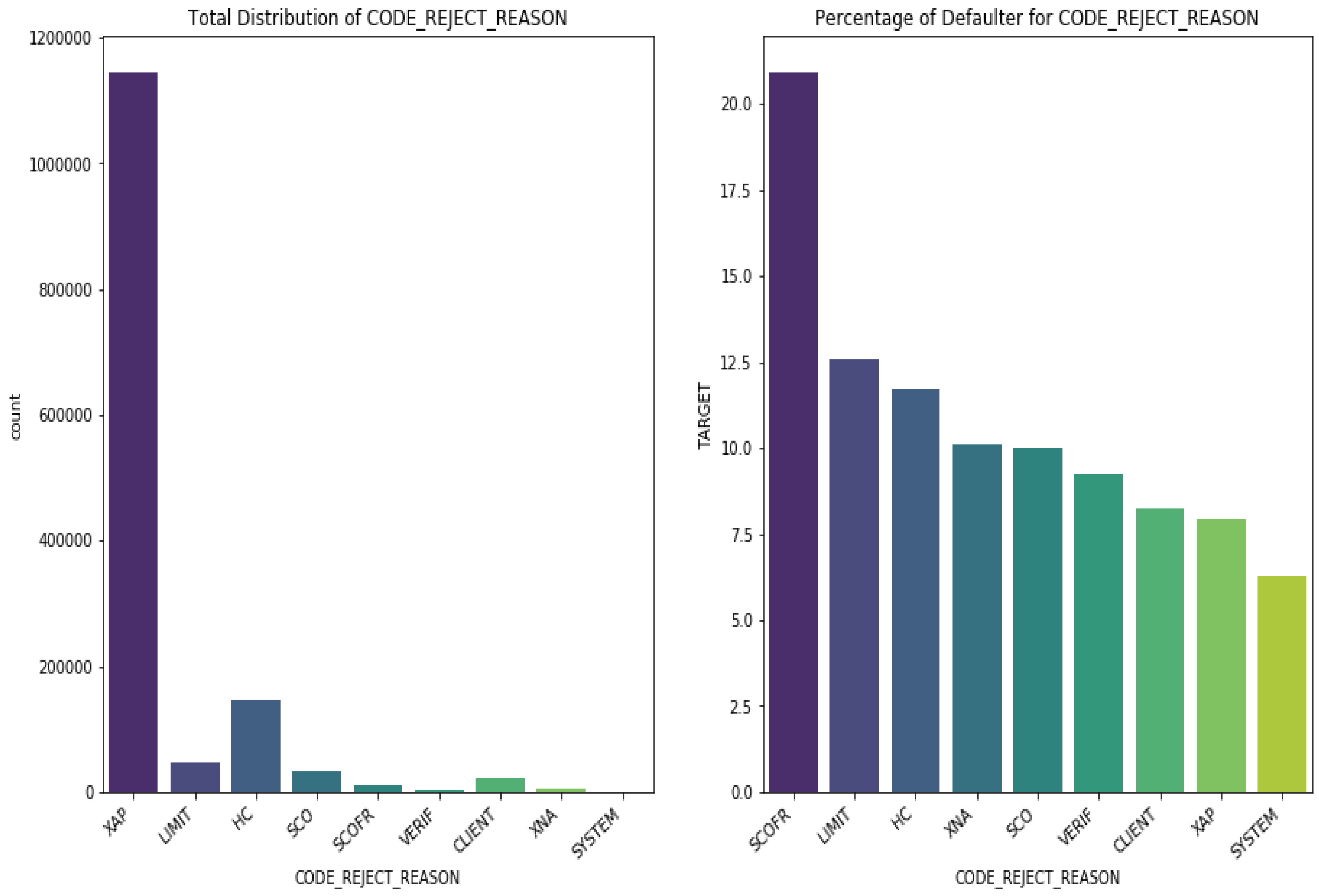
Lot of defaulters have had a previous application rejected





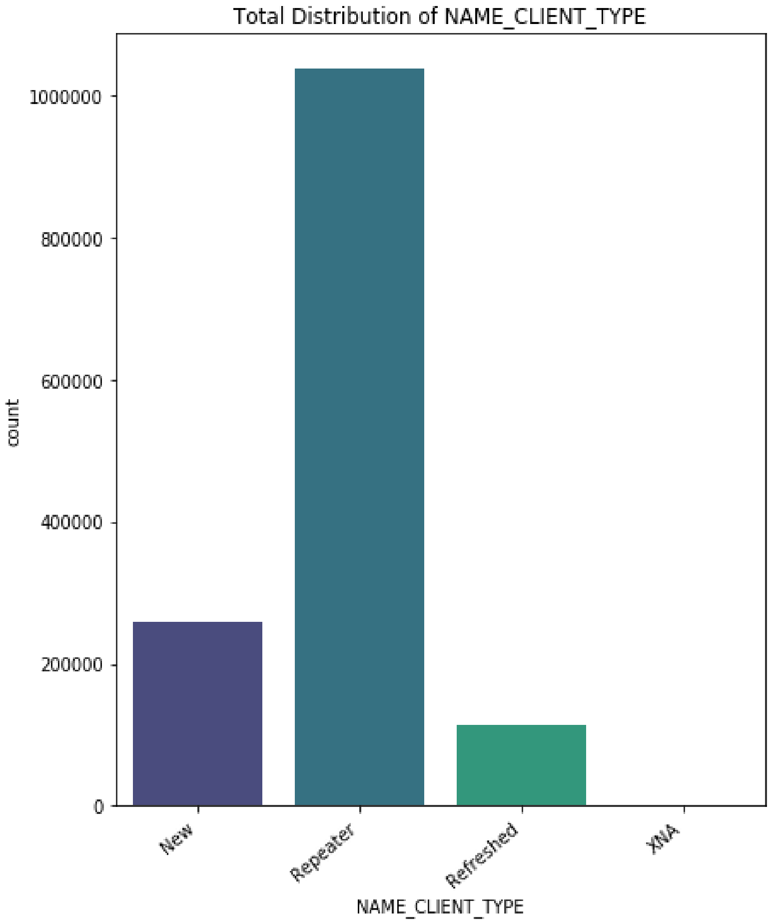
Reject reason vs defaulters

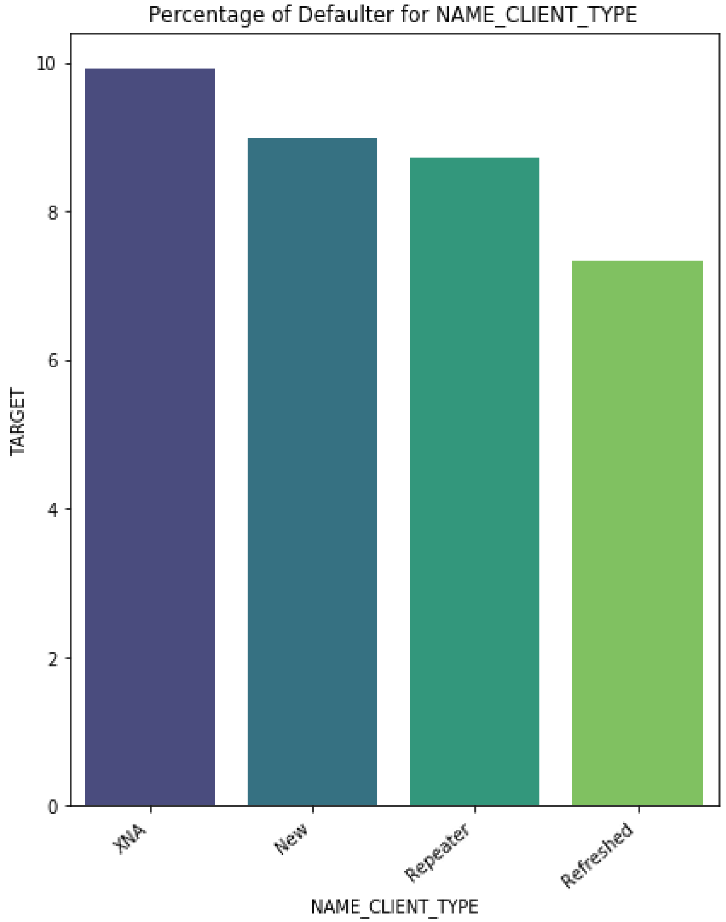
Most defaulters have had a previous rejection with code SCOTR



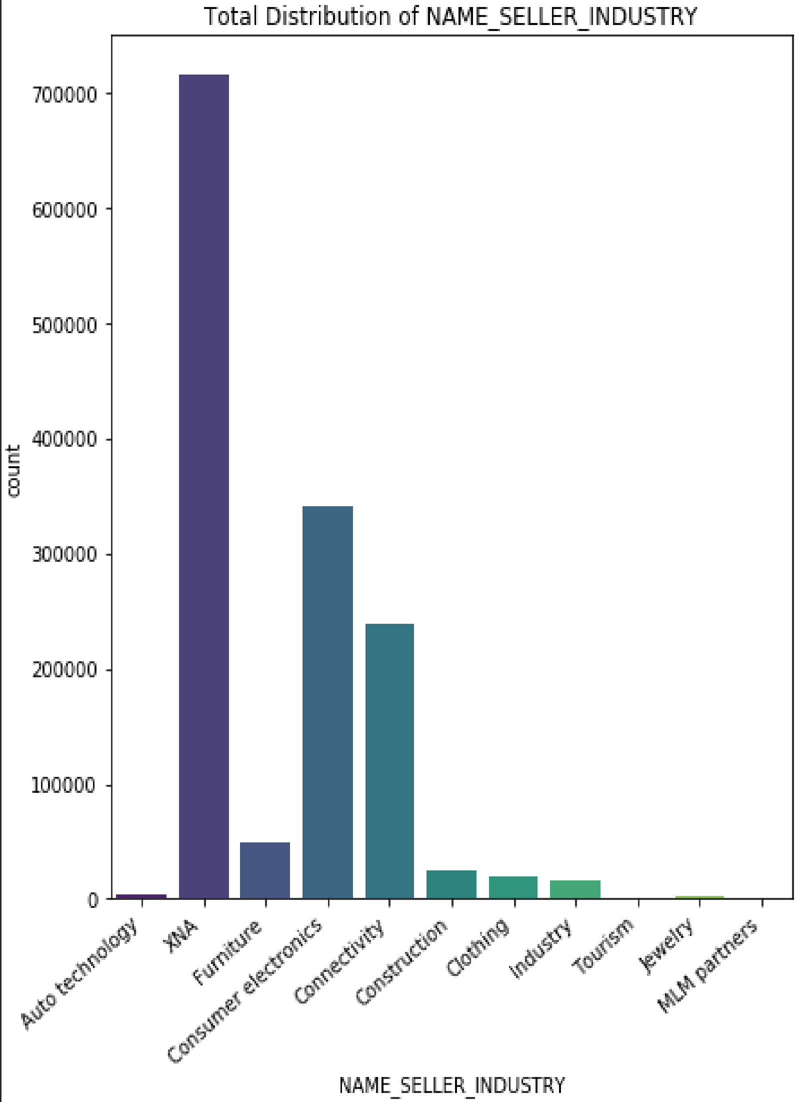
Client type vs Defaults

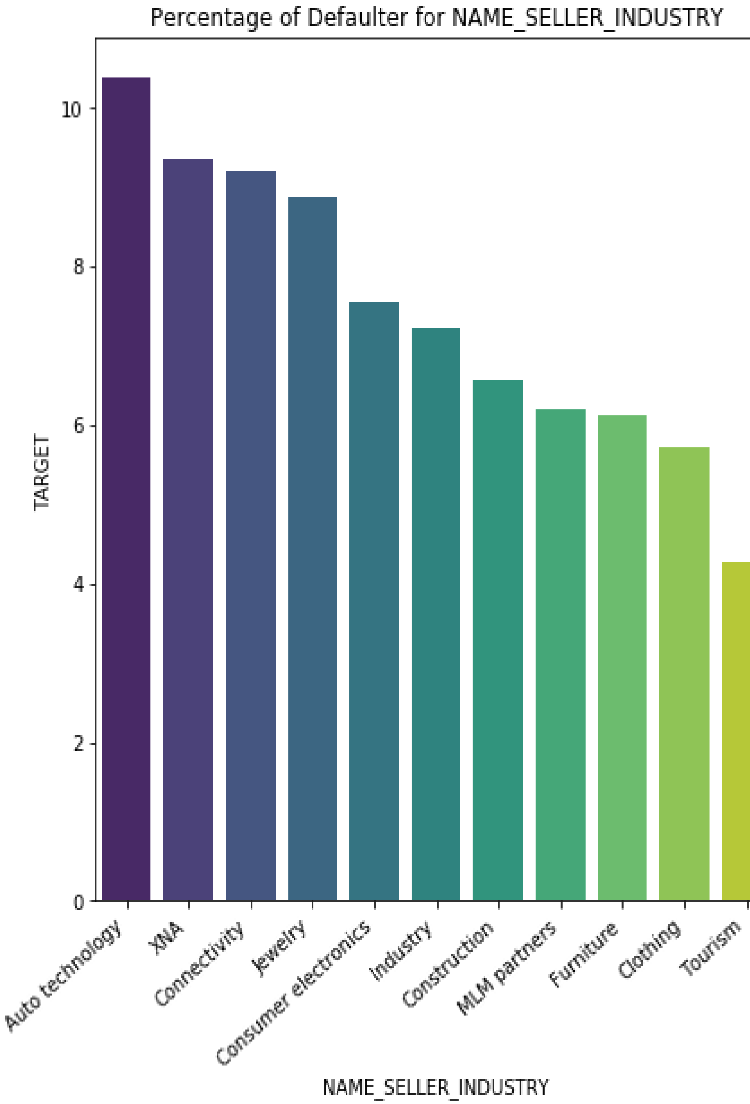
Repeat applications have slightly less rate of defaults





Industry vs Defaults





Conclusions

1) Female loan applicants are lower risk, so they should be given higher priority

2)low income group applicants should be treated with caution as they default more. Typically a lower amount may be sanctioned /or and with higher rates after due diligence

3)Applicants lacking higher education are at a high risk of defaulting, so banks should prioritize applicants with a degree.

4)Young adults default at a higher rate so they should be treated with caution.

5) Banks should invest more in marketing loans to seniors as they default less and currently are a bit under served in the industry.

6)Applicants working in low skill industry/blue collar jobs default more so banks should be careful in doling out loans to this group

Conclusions (Cntd.)

7) A previous loan rejection is a red flag for banks as they lead to higher number of defaults

8)Repeat customers(previous good loan history) are somewhat safer