EDA – Credit Assignment
Rashi Jain

BUSINESS OBJECTIVE

This case study aims to identify patterns which indicate if a client has difficulty paying their installments 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. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

PROBLEM STATEMENT

To identify patterns which indicate if a client has difficulty paying their installments 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

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.

HANDLING MISSING VALUES

- Columns with more than 40% missing values were dropped in application data.

- For Columns less then 19% missing values:

  - For Numeric columns like EXT_SOURCE_2, values were imputed with median().

  - For Non-Numeric columns like NAME_TYPE_SUITE, values were imputed with mode().

- Rows with less then 0.3% data missing were dropped.

- Previous application data – 50% missing values were dropped and merged to application data on SK_ID_CURR

CLEANING DATA FOR BETTER UNDERSTANDING

- Values like DAYS_BIRTH, DAYS_EMPLOYED had negative values. Hence there values were replaced by its absolute value.

- New columns added:
    - AGE - DAYS_BIRTH was converted into age in years by dividing it by 365
    - AGE_GROUP – age was divided into bins of teens, 20s, 30, 40, 50 and senior to categorise data
    - AMT_INCOME_TOTAL_BIN - Total income was divided in bins of low, medium and high using quantile to handle outliers
    - AMT_CERDIT_BIN - Total amount credit was divided in bins of low, medium and high using quantile to handle outliers

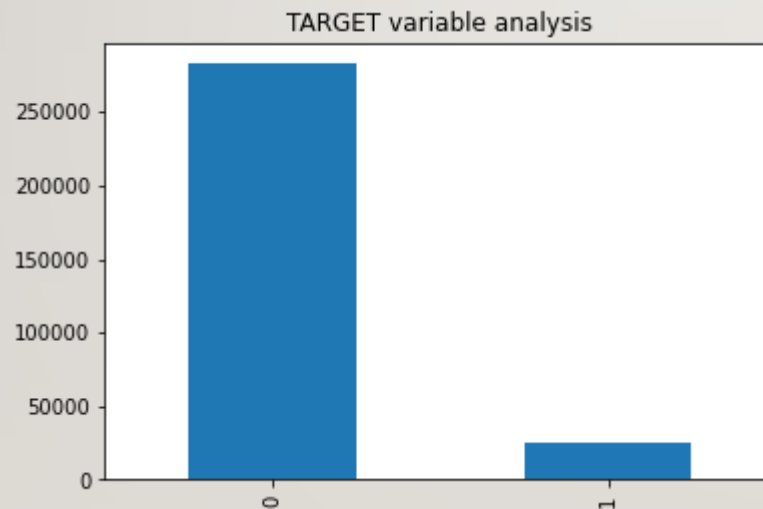# OUTLIERS FOR AMT_TOTAL_INCOME AND AMT_CREDIT IN FORM OF BOX PLOTS
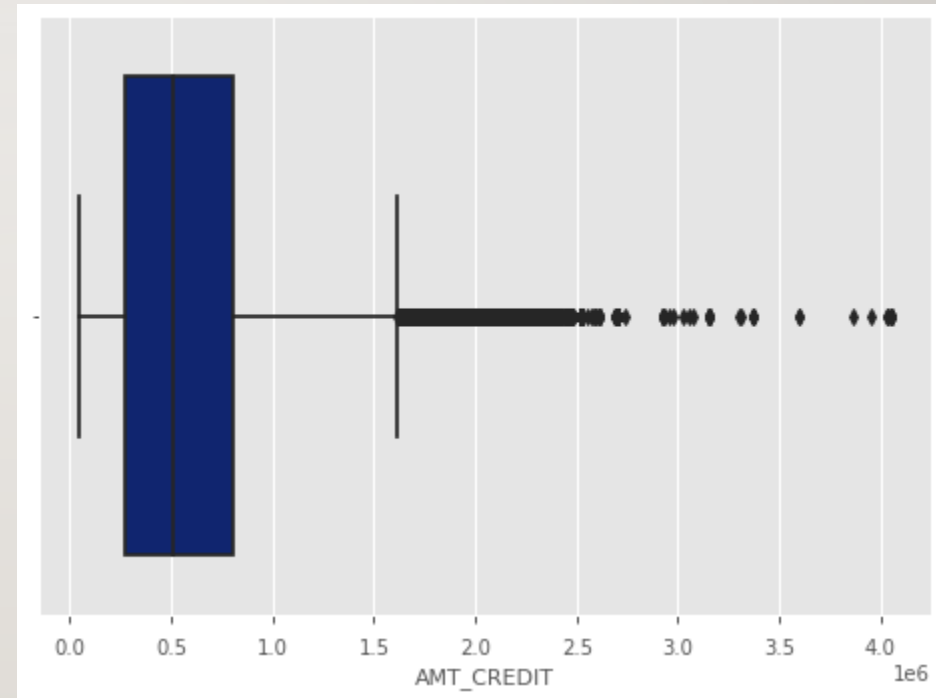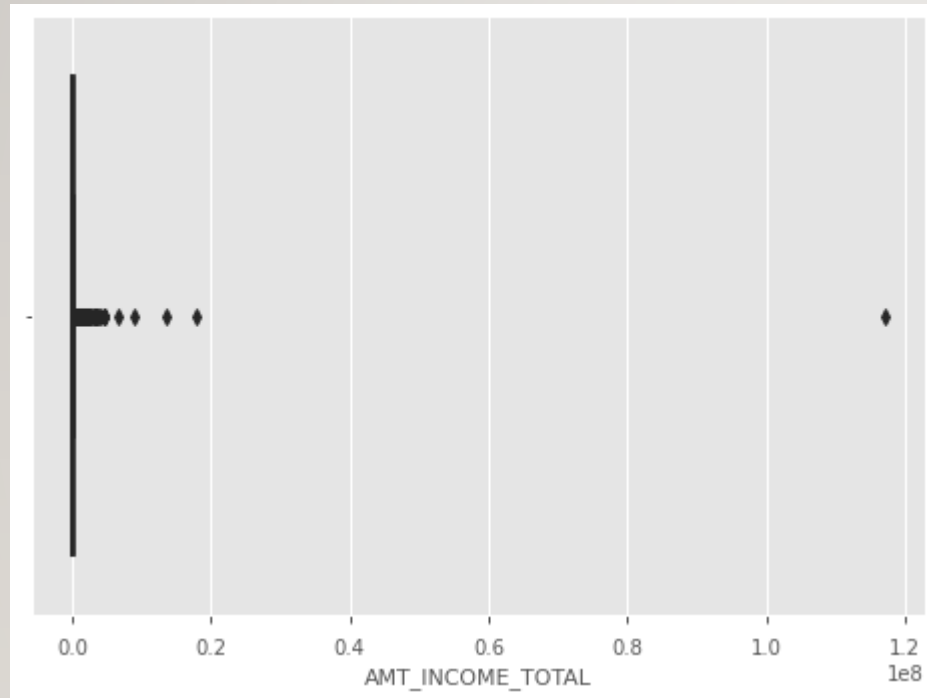
FOCUS COLUMNS

- Gender

- Previously refused loans

- Loan amount

- Education

- Applicant's age

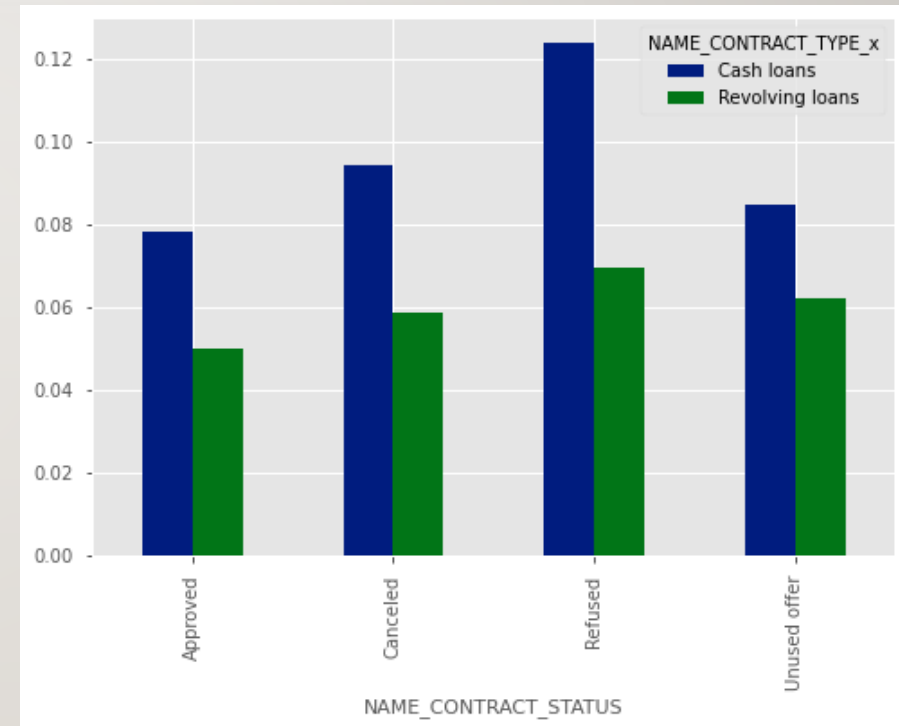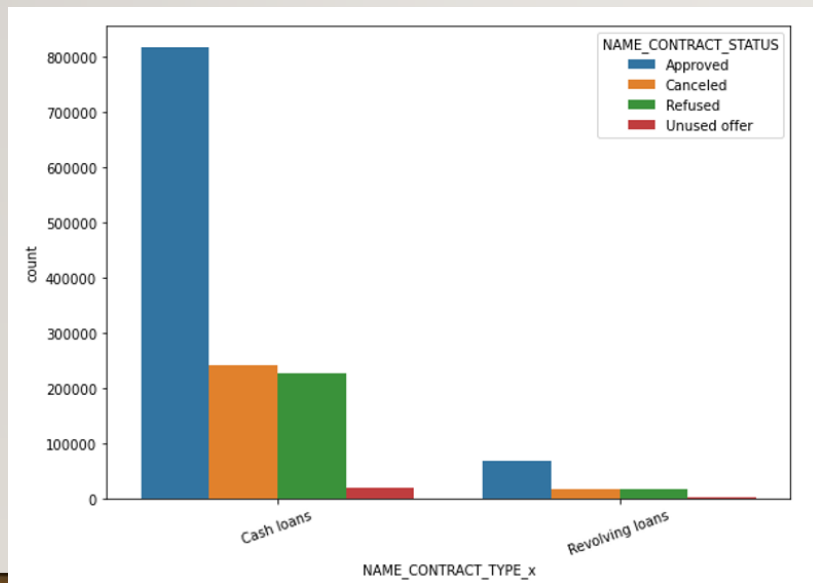- Applicant's family status

- Contract type

TARGET VARIABLE

About 8% people were defaulting on loans, out if which 12% applicant were refused a loan before.
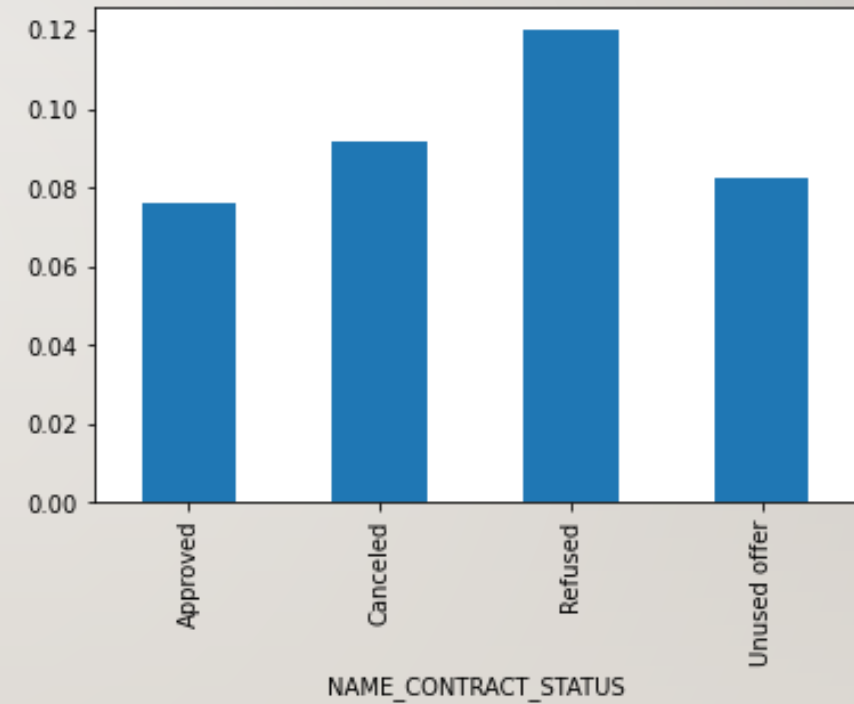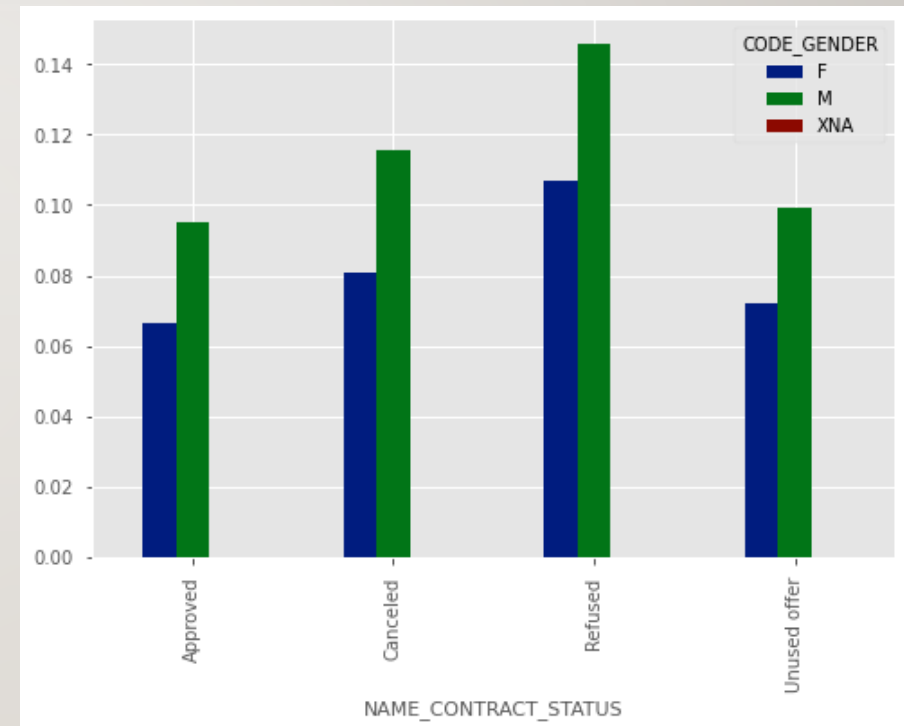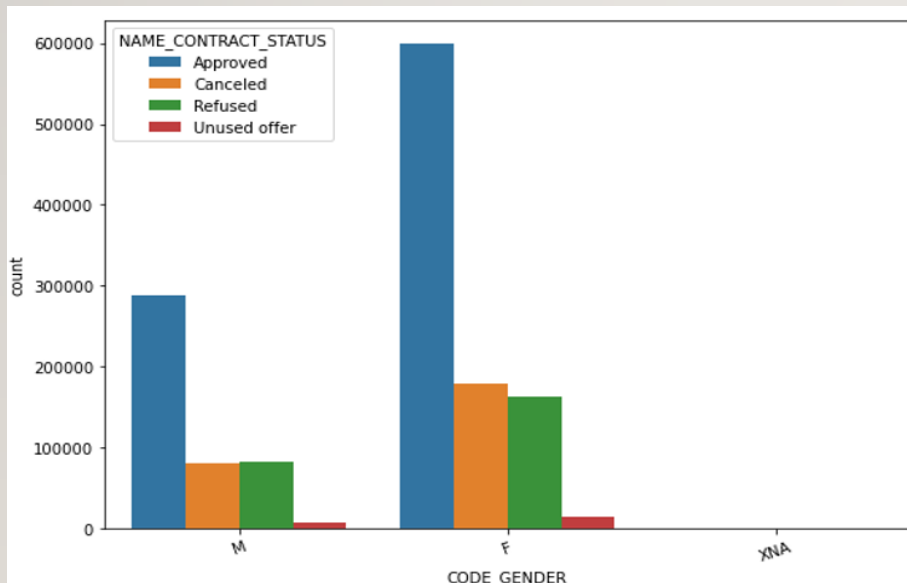
CONTRACT TYPE

People with cash loans have more difficulty in payment and defaults more.
Revolving loans should be preferred more.
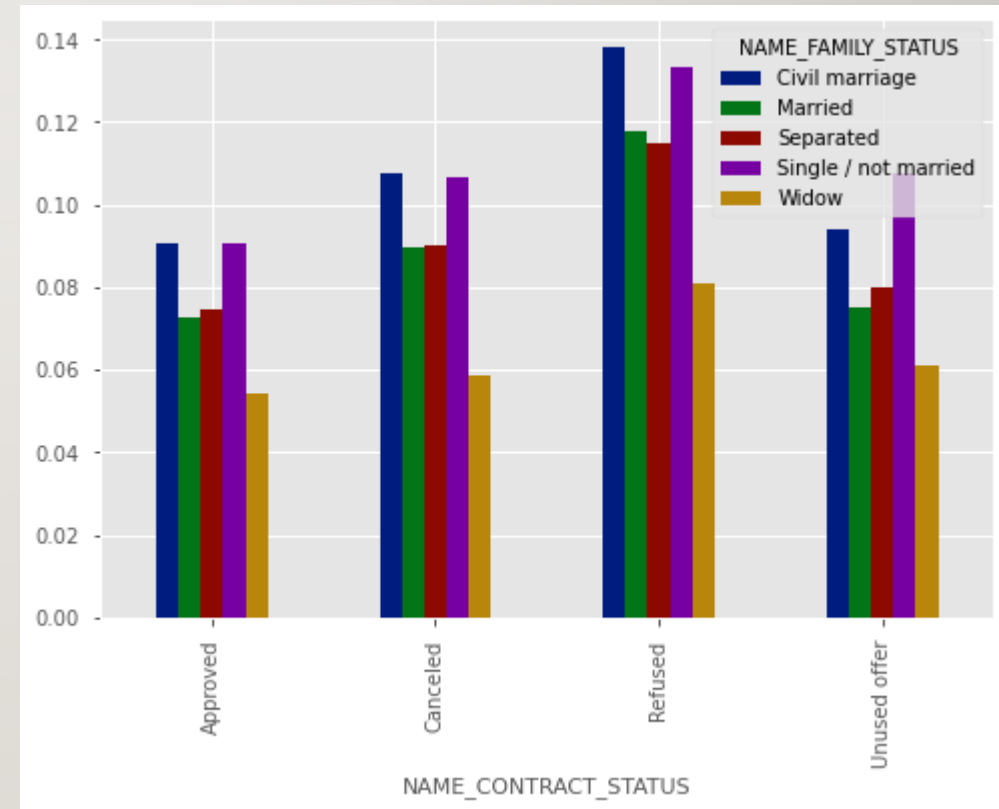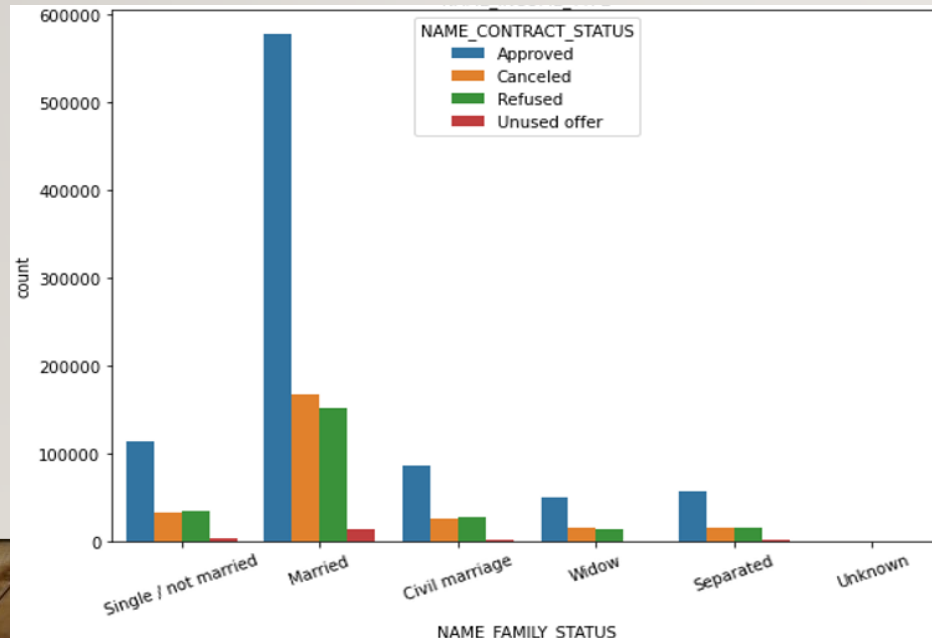
APPLICANT'S GENDER

Female got more approved loans
Male clients have more difficulty in paying
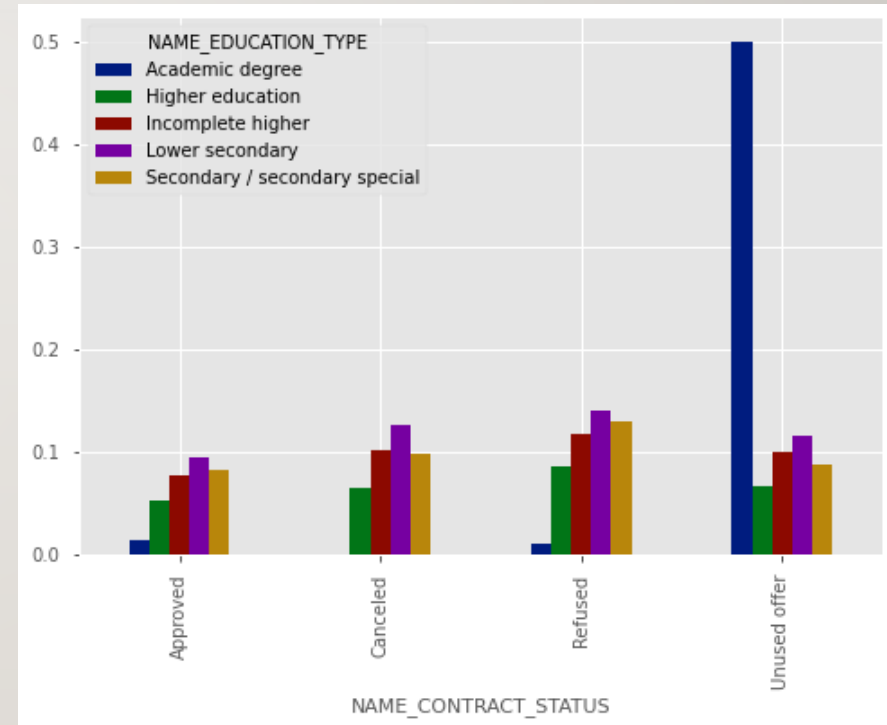loans, females are at less risk.
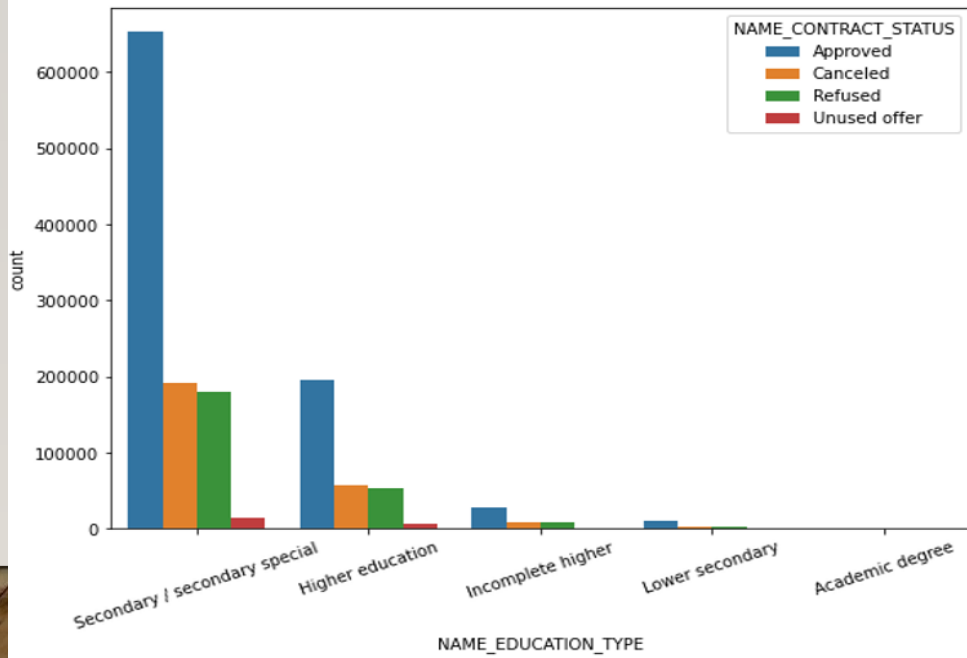
FAMILY STATUS OF APPLICANT

Married people were approved more loans previously
Widows have less defaults payments, single and civil married are higher risk candidates.
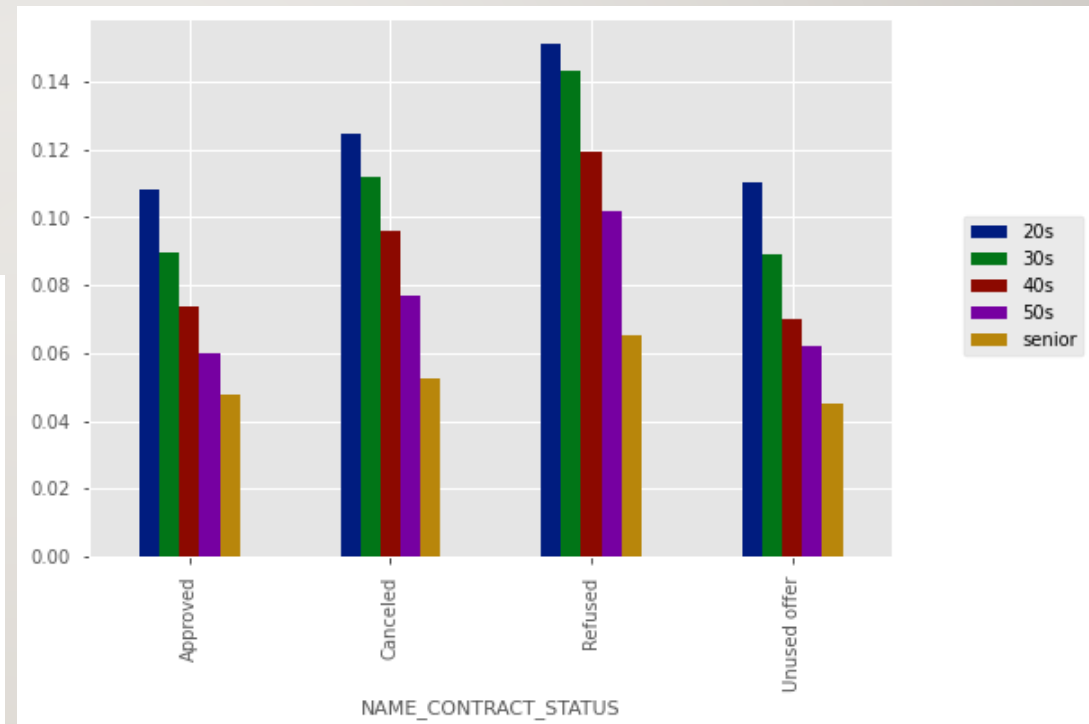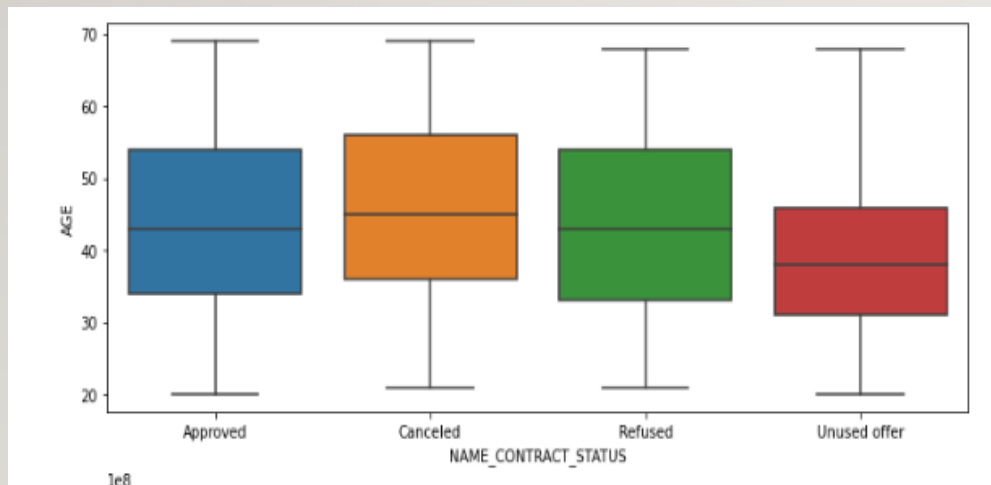
EDUCATION STATUS

People with secondary/special education were given more loans
Clients with academic degree and have unused offer have more defaults.
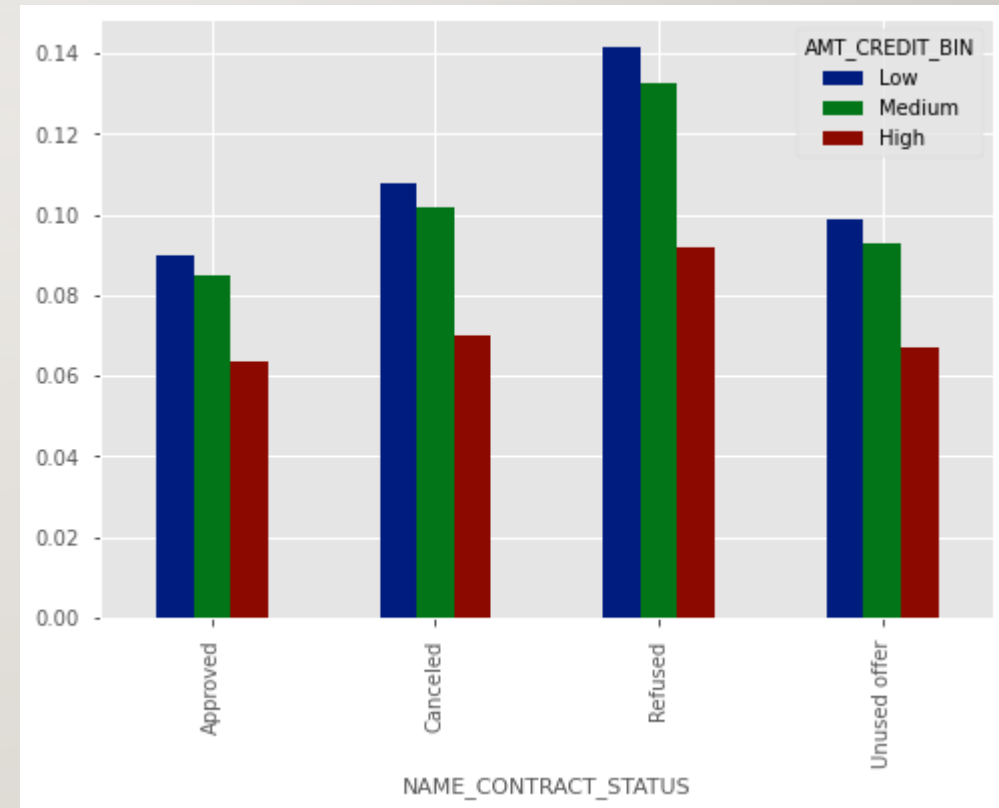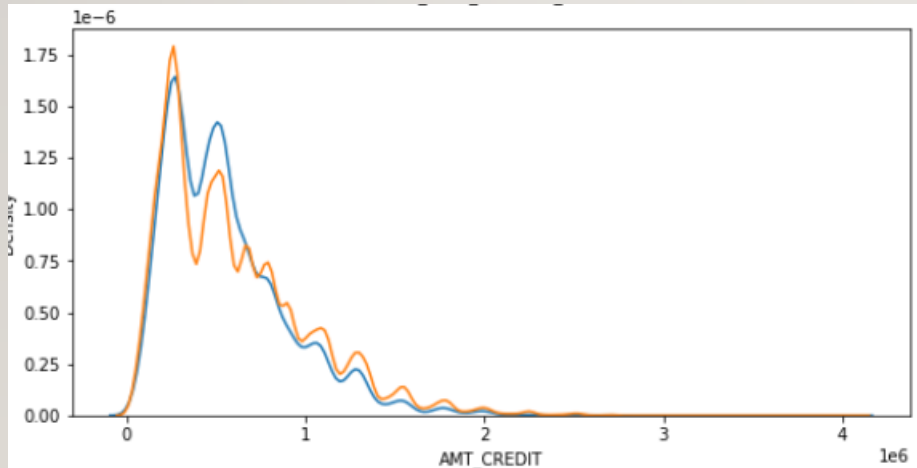
APPLICANT'S AGE

People in there 20s and 30s are more at risk, senior citizens on the other hand are at least.

LOAN AMOUNT

Applicant with higher Credit amount are less likely to default.
Applicants with lower amount loans have who are defaulting were refused loans previously.

CONCLUSIONS

The factors that most affects the loan approvals are:

- Gender - Females are more likely to pay the installments and default less and should be considered to give loans

- Previously refused loans - Applicants who were refused loans previously should be avoided since they default more.

- Loan amount - Loans for lower amount between 270000-513531 is much risker range and sould be avoided, higher range are more likely options.

- Education- Applicants who completed education till lower secondary should be avoided.

- Applicant's age - People in there 20s should be avoided and senior citizens can be preffered more.

- Applicant's family status - Widows should be preferred rather then single and civil married folks.

- Contract type - Revolving loans are much better option then cash loans.