Credit EDA Case Study
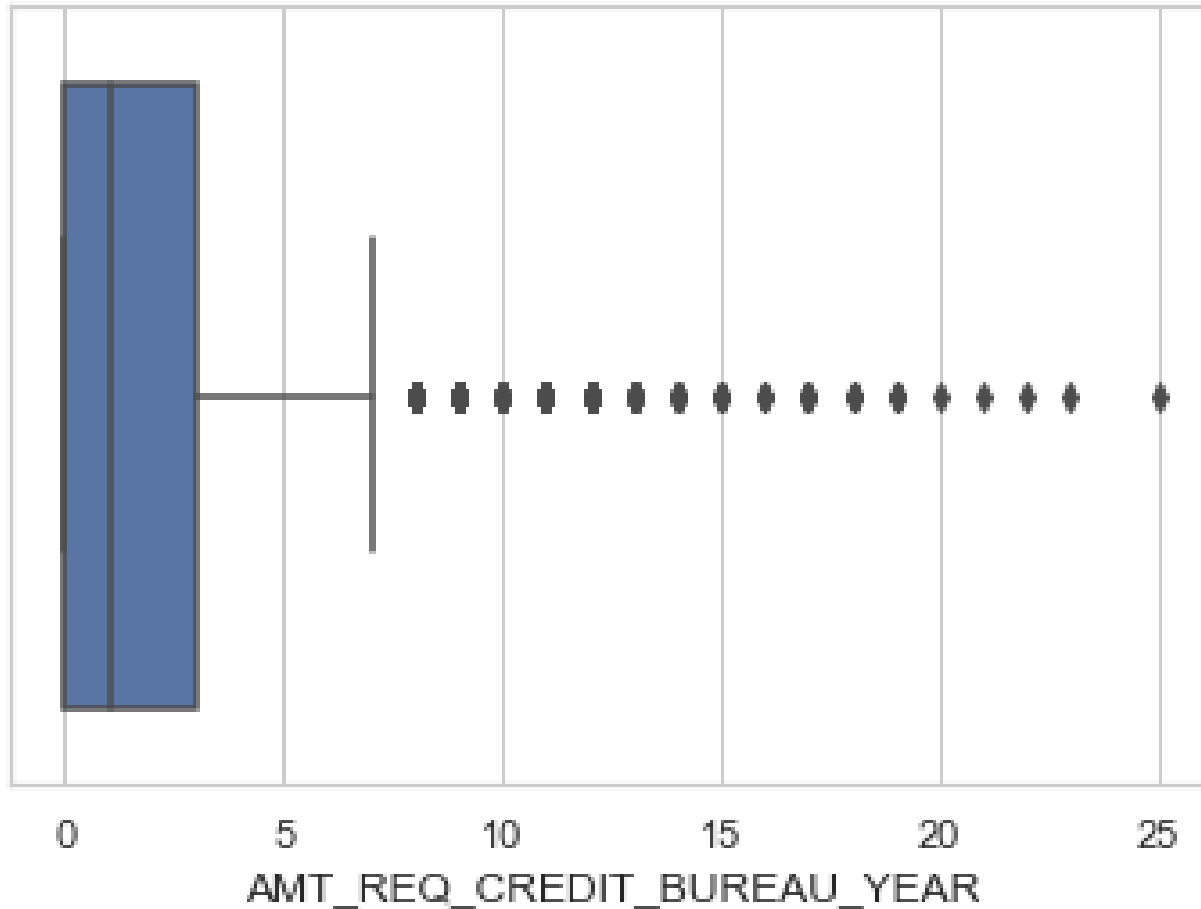## (Group Case Study 1)

Prepared by:-
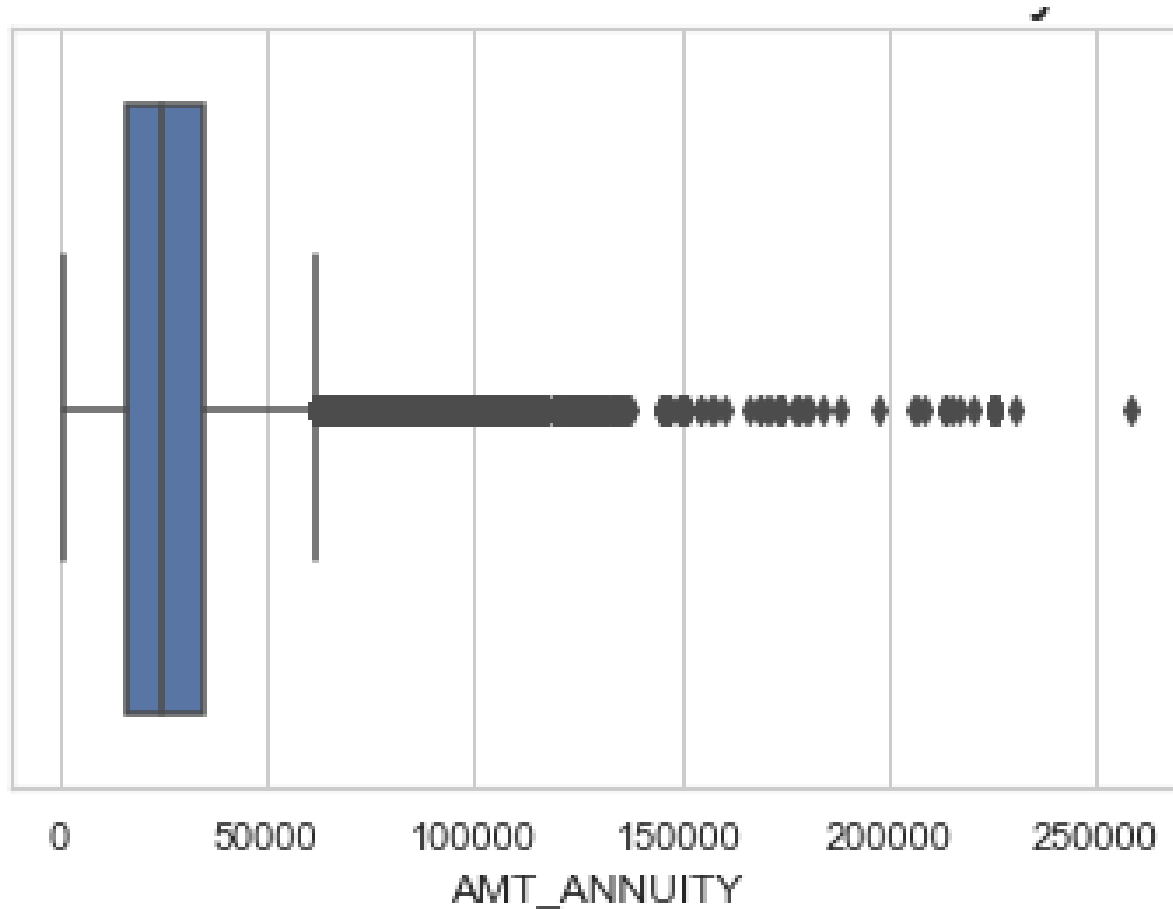
Utkarsh Upadhyay

Mayank Singh Soni

Univariate Analysis

Distribution for AMT_REQ_CRE DIT_BUREAU_ YEAR

- There exist outliers in the AMT_REQ_CREDIT_BUREAU_YEAR variable.

- Median lies around 1, so we use mode to impute the missing values

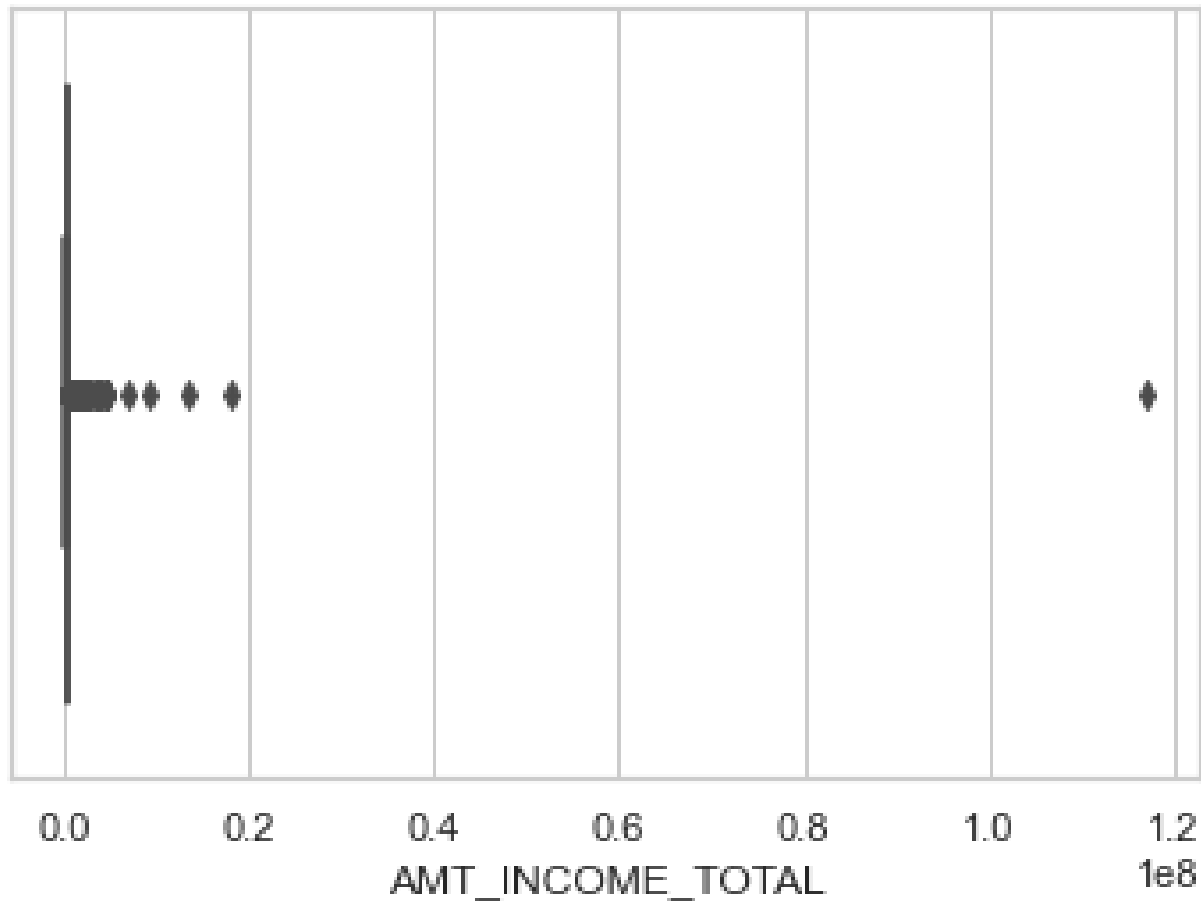Prepared By: Utkarsh Upadhyay & Mayank Singh Soni

Credit EDA Case Study

Distribution for Annuity

- Plot shows that there are outliers in the AMT_ANNUITY column which is at the extreme end and can be removed from analysis
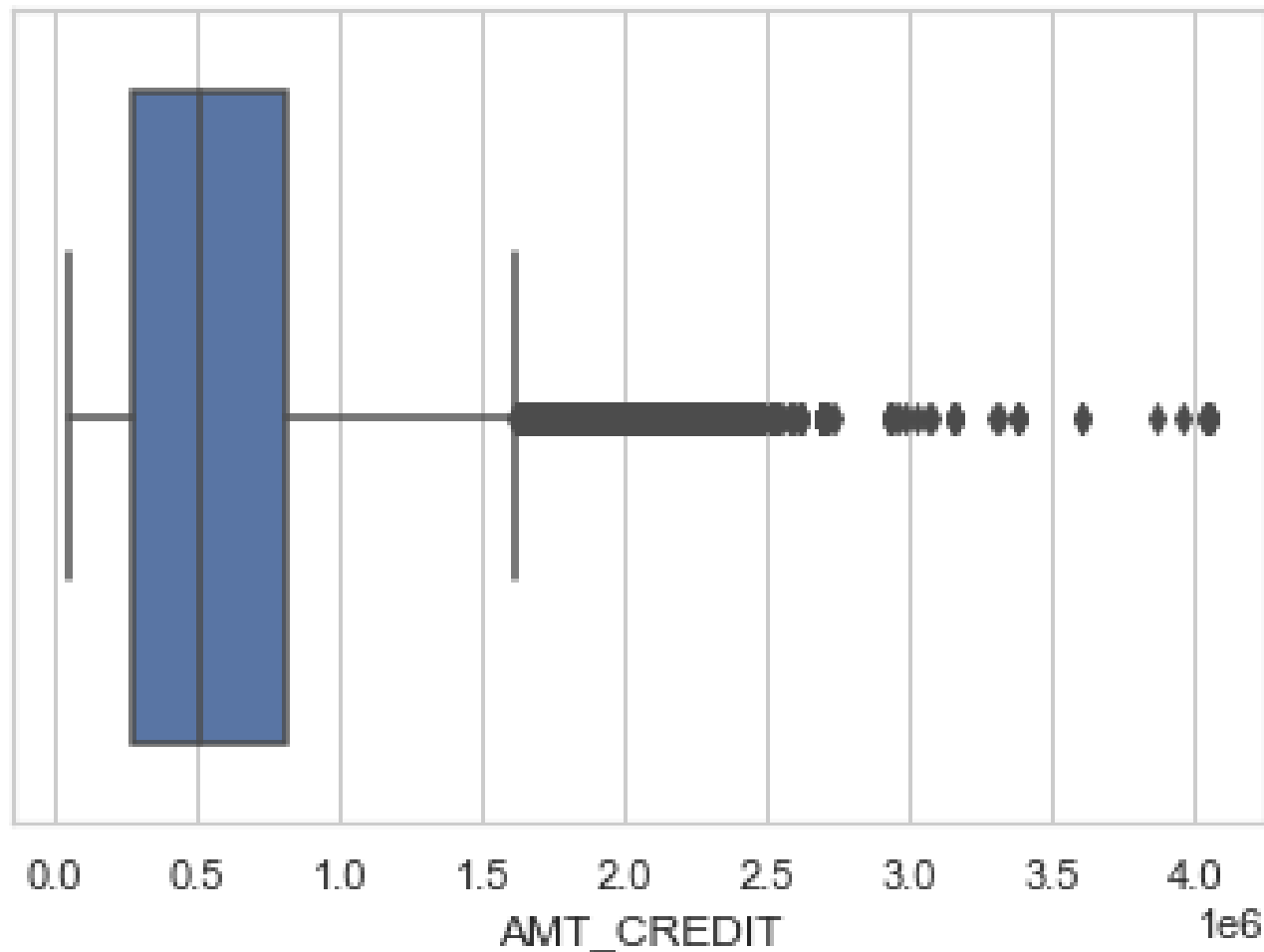
Distribution for Total Income amount

- Total Income amount column shows outliers but the income vary from person to person.

- We cannot remove this outlier.

- It is better to create bins for it.

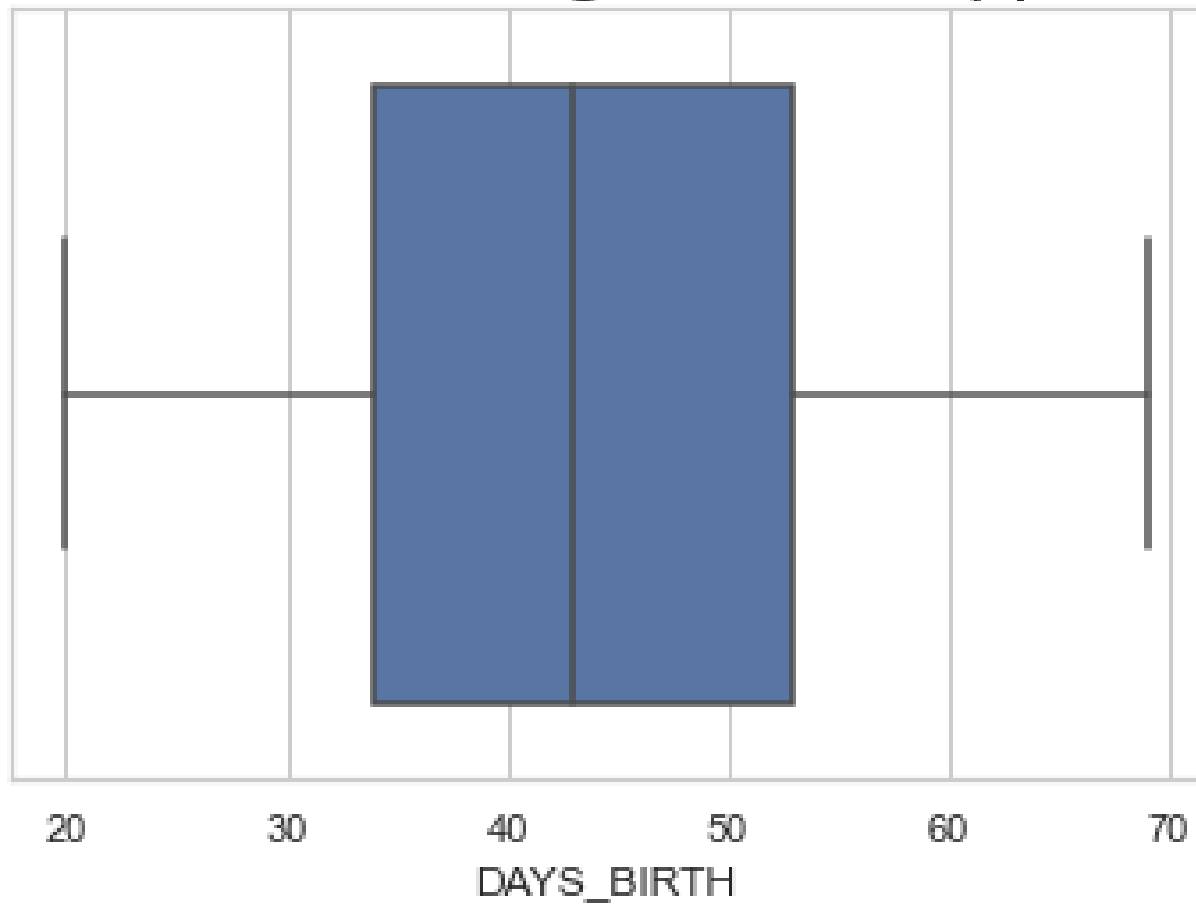Prepared By: Utkarsh Upadhyay & Mayank Singh Soni

Credit EDA Case Study

Distribution for Loan Credit Amount

- Loan Credit amount column shows outliers but the income vary from person to person.

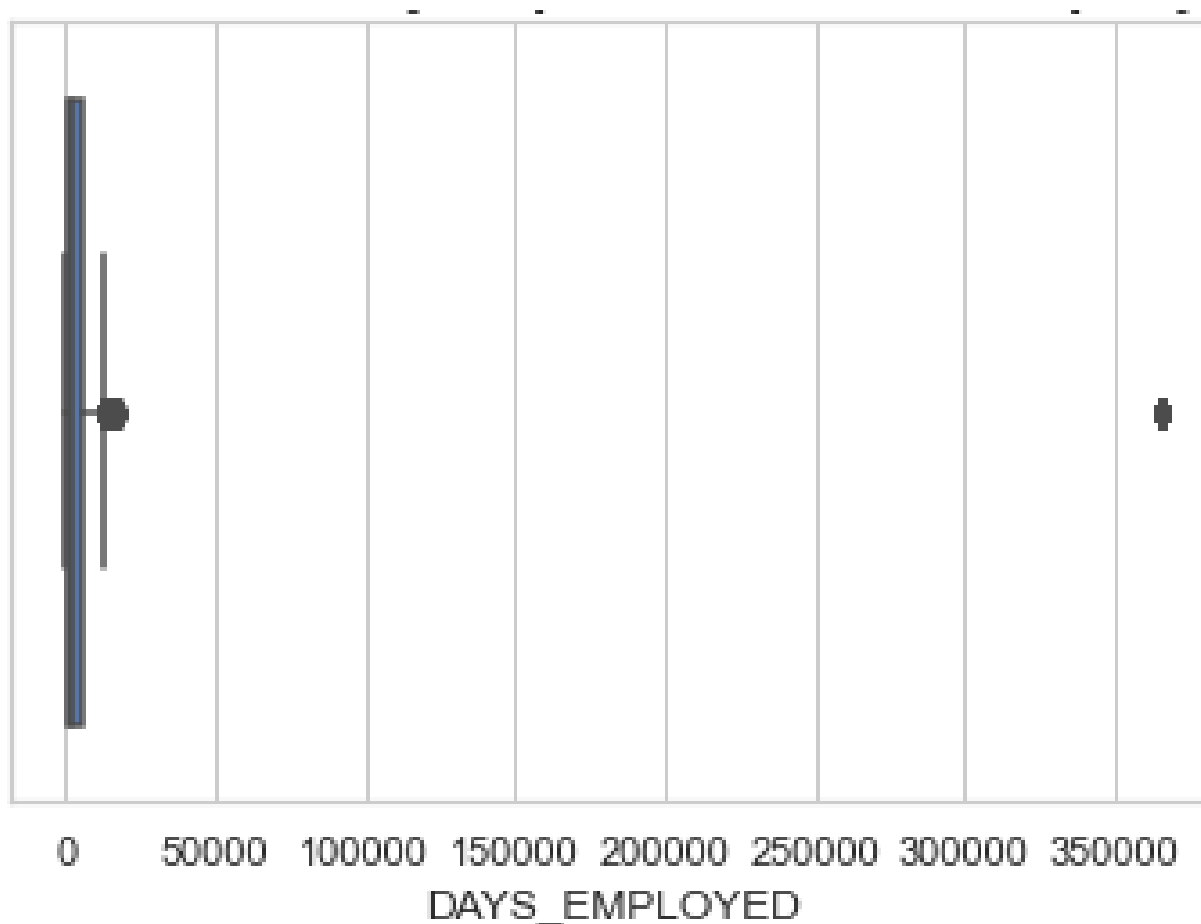- We cannot remove this outlier.

- It is better to create bins for it.

Distribution for Applicant Age

- Applicant age data does not have any outliers.

- It is well distributed from age 20 to age 70.

- Most applicant are fall under age between 35 to 55.
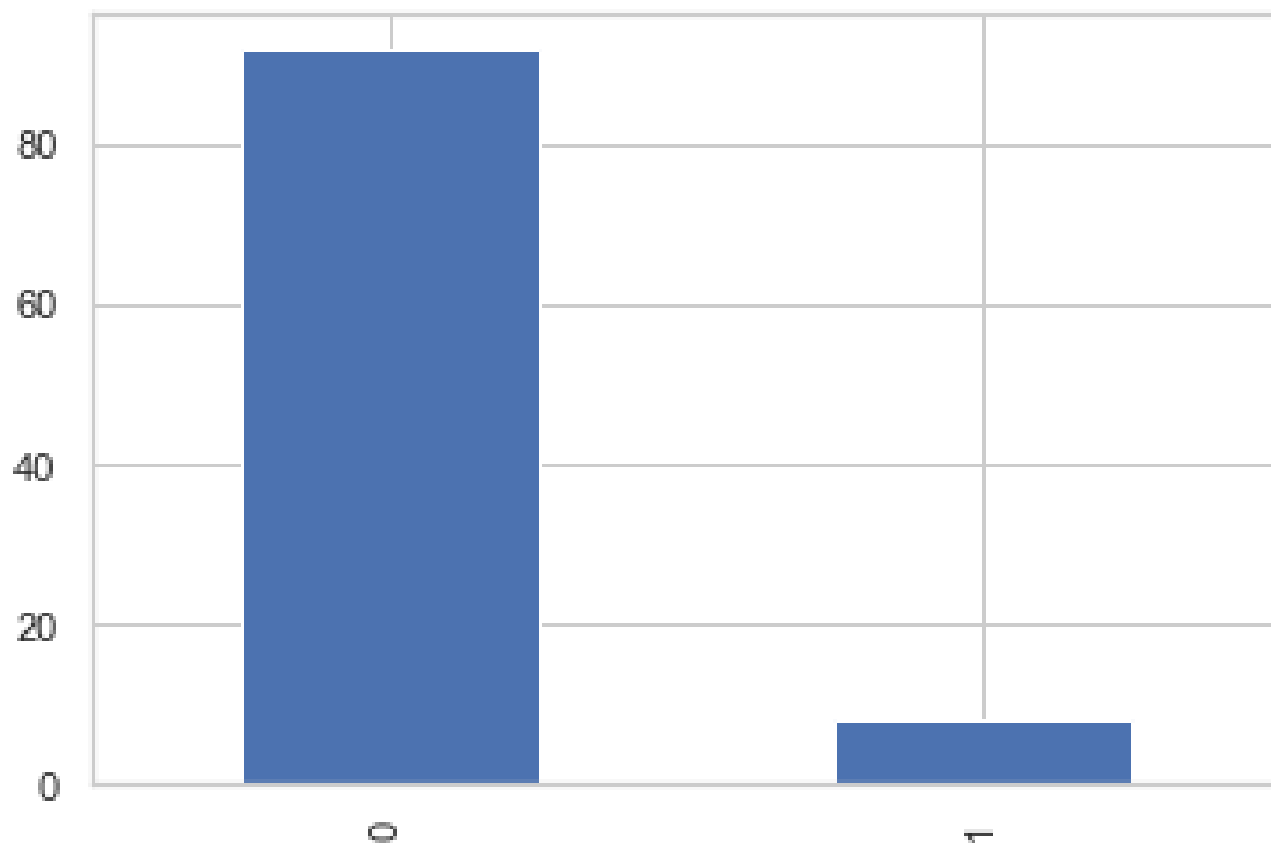
- Median lies around age of 43.

Distribution for applicant's employed days

- There is a outlier whose value is around 360000 days (i.e. 986 years approx.) which is false data. Thus, it needs to be removed.
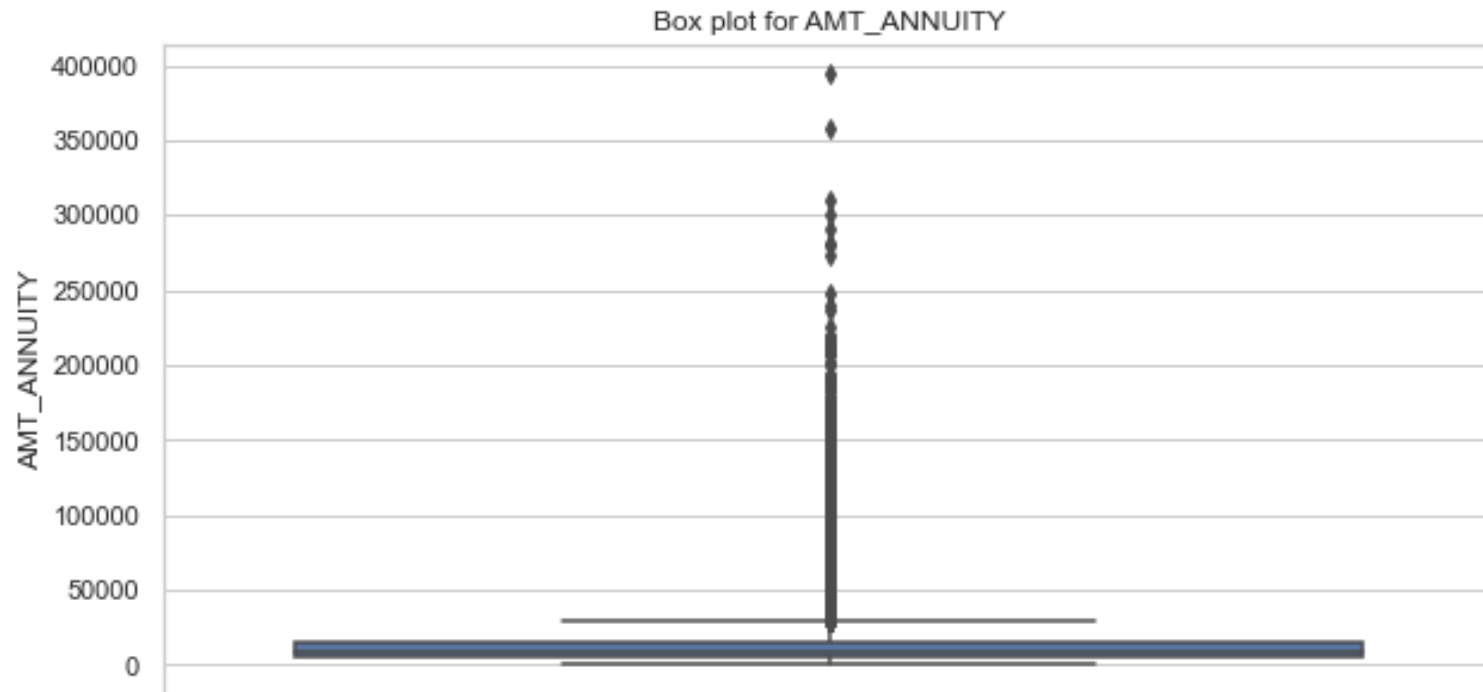
Prepared By: Utkarsh Upadhyay & Mayank Singh Soni

Distribution for target variable

- By means of target variable, we can see that approx. 92% people had no difficulty in loan repayment and only 8% people had difficulty in loan repayment.

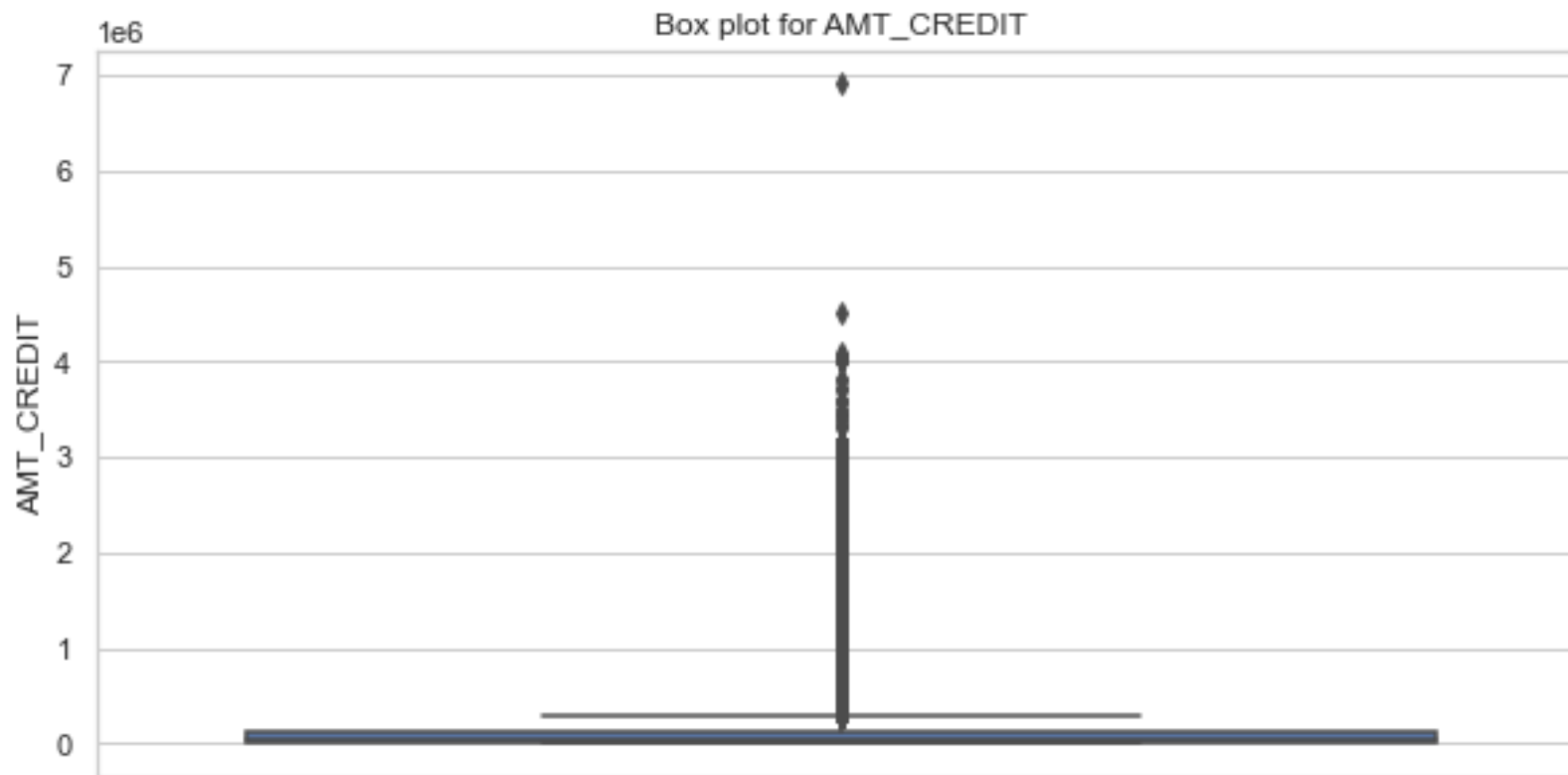Credit EDA Case Study

Box plot for AMT_ANNUITY

Distribution for previous annuity amount

- There are outliers in the previous annuity amount data but annuity amount is also varies from applicant repaying capabilities (High wages/salaried person can afford high annuity amount)

- It is better to be capped or binned
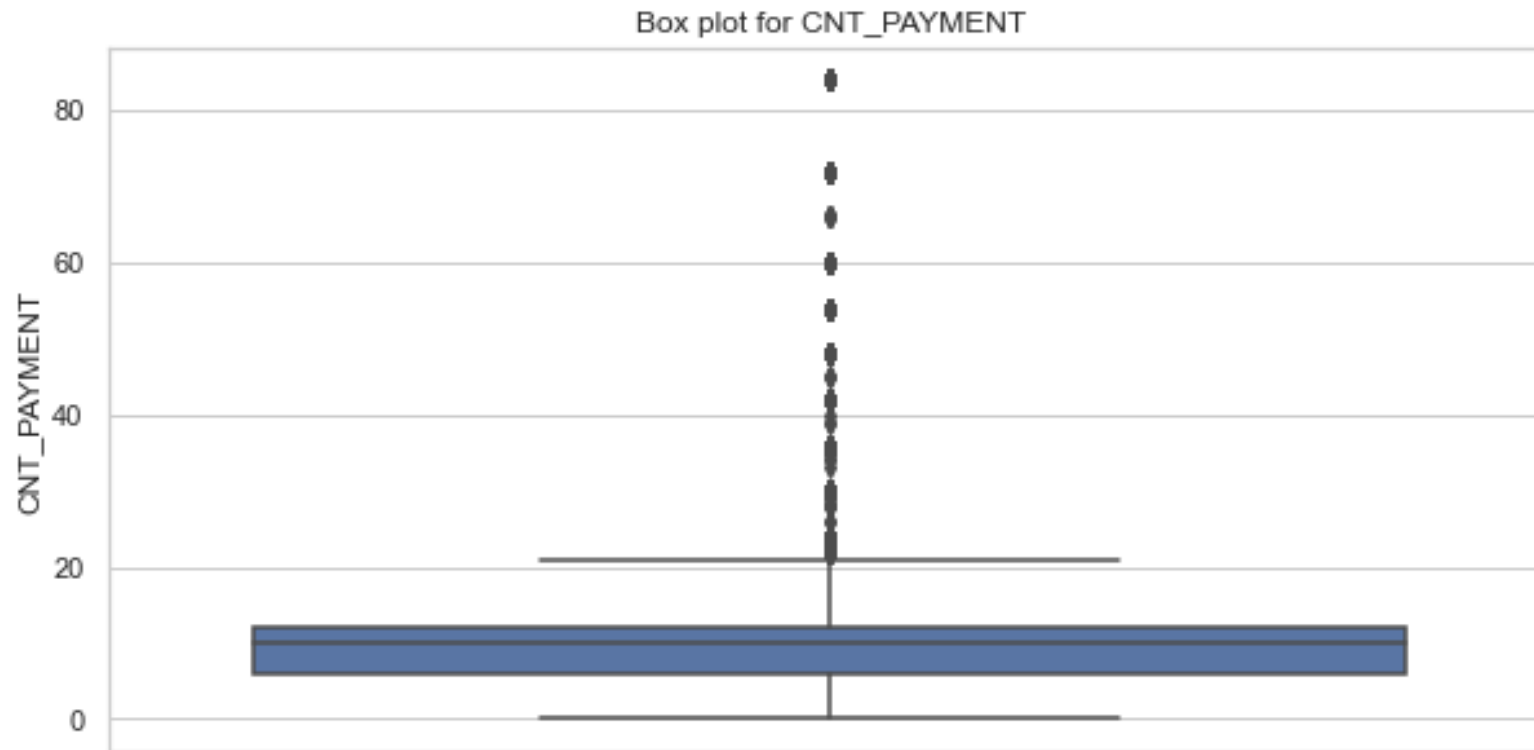
Credit EDA Case Study

Box plot for AMT_CREDIT

Distribution for previous credit amount

- There are outliers in the previous credit amount data but credit amount is also varies from applicant repaying capabilities (High wages/salaried person can afford to high repay amount)

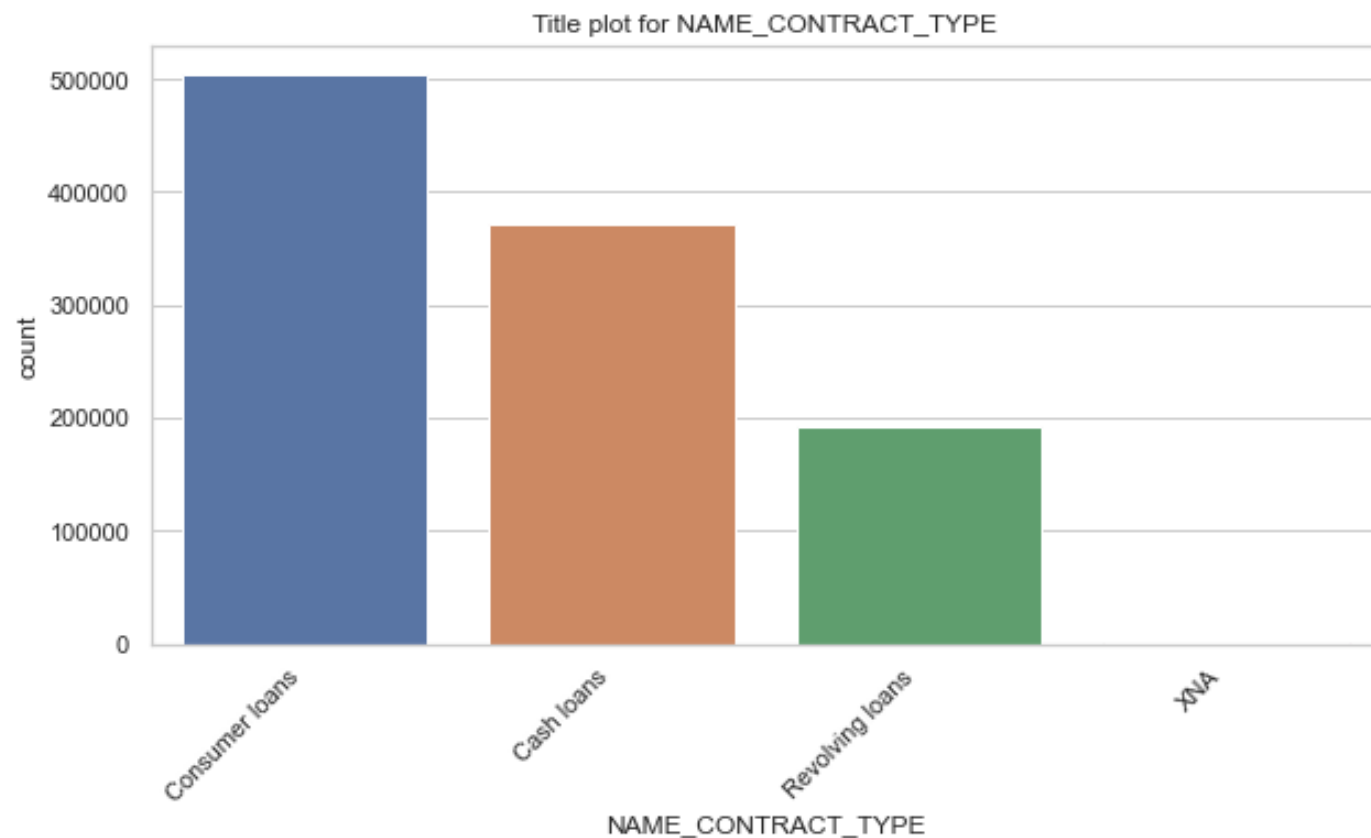- It is better to be capped or binned

Box plot for CNT_PAYMENT

Distribution for previous payment count

- There exist outliers in the data.

- These are outliers are practically possible as a person can have long term EMIs for low EMI cost.
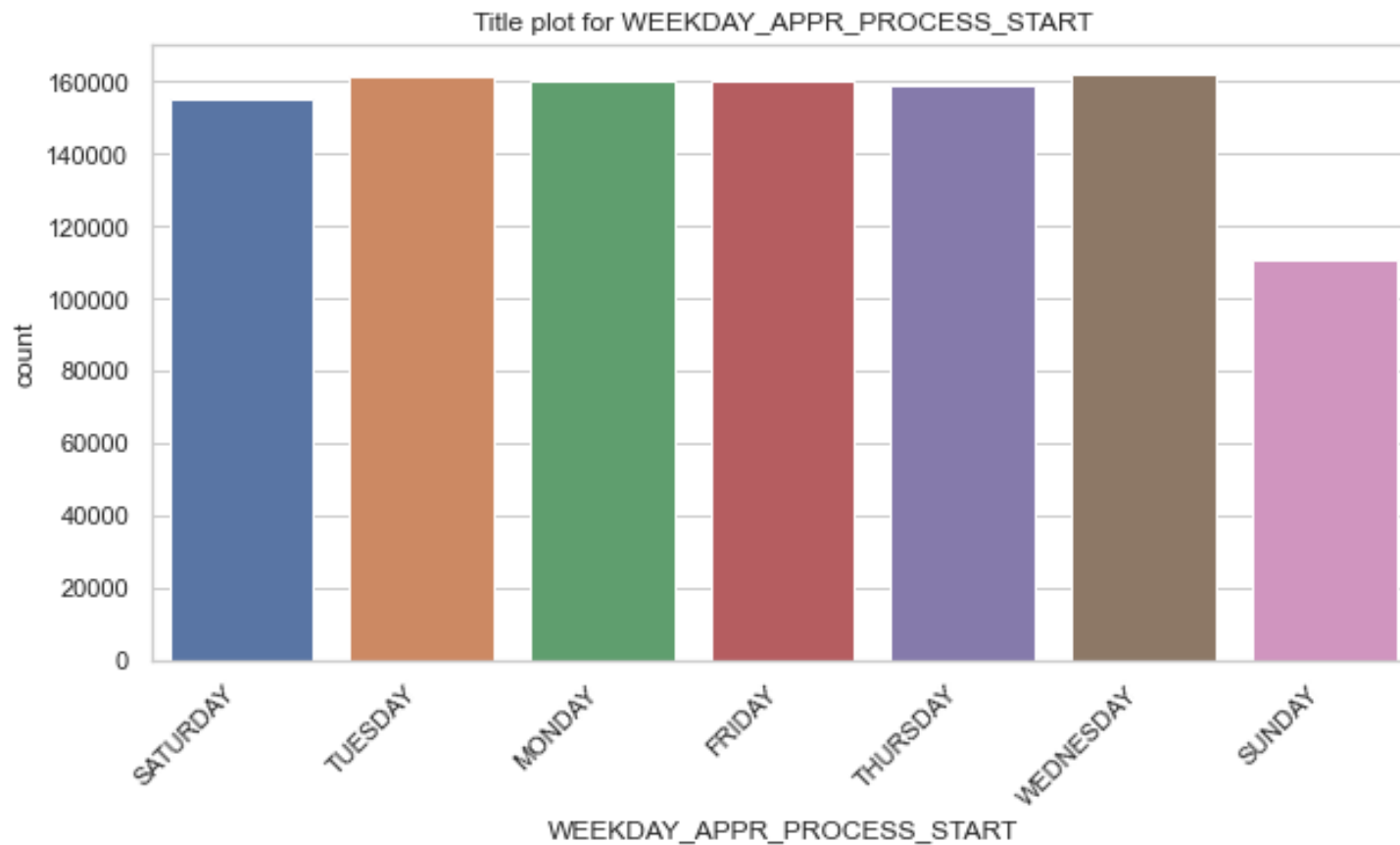
Prepared By: Utkarsh Upadhyay & Mayank Singh Soni

Distribution for Previous Contract Type

- Clearly visible that there exist high number of consumers loans as compared to any other loans.

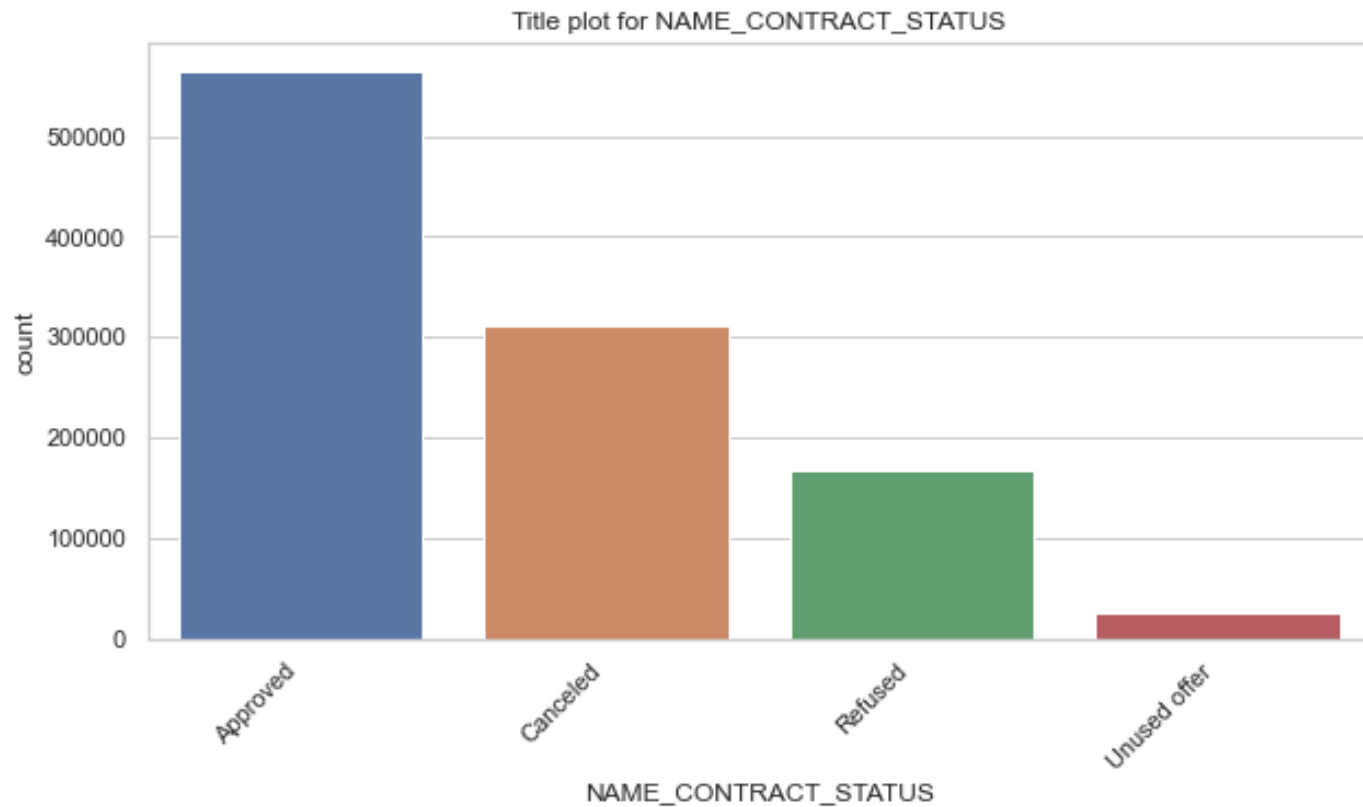Title plot for WEEKDAY_APPR_PROCESS_START

Distribution for

- There are less loan application start on Sunday as compared to any other day.

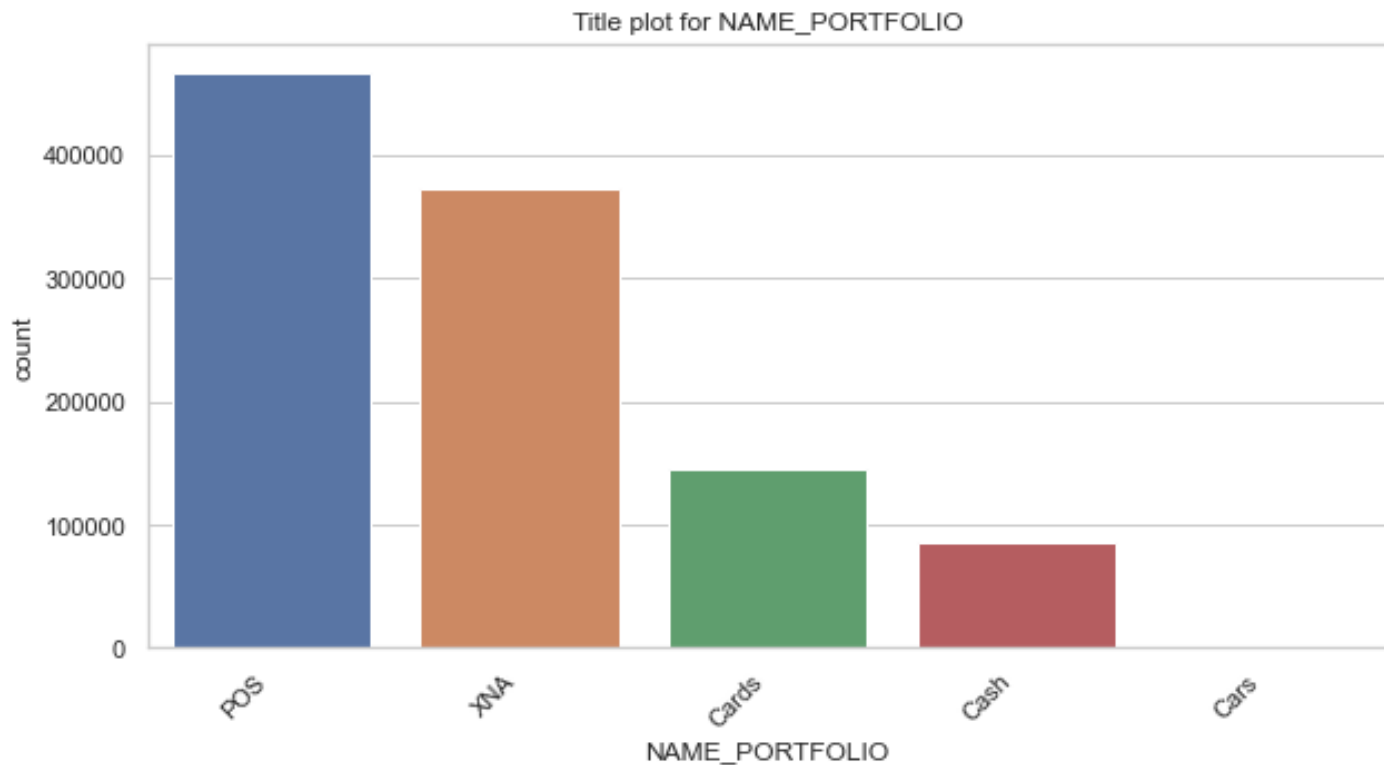- All week days have approximately same numbers of process start.

Credit EDA Case Study

Distribution for previous contract status

- There are high number of approved applications in the previous applications dataset.

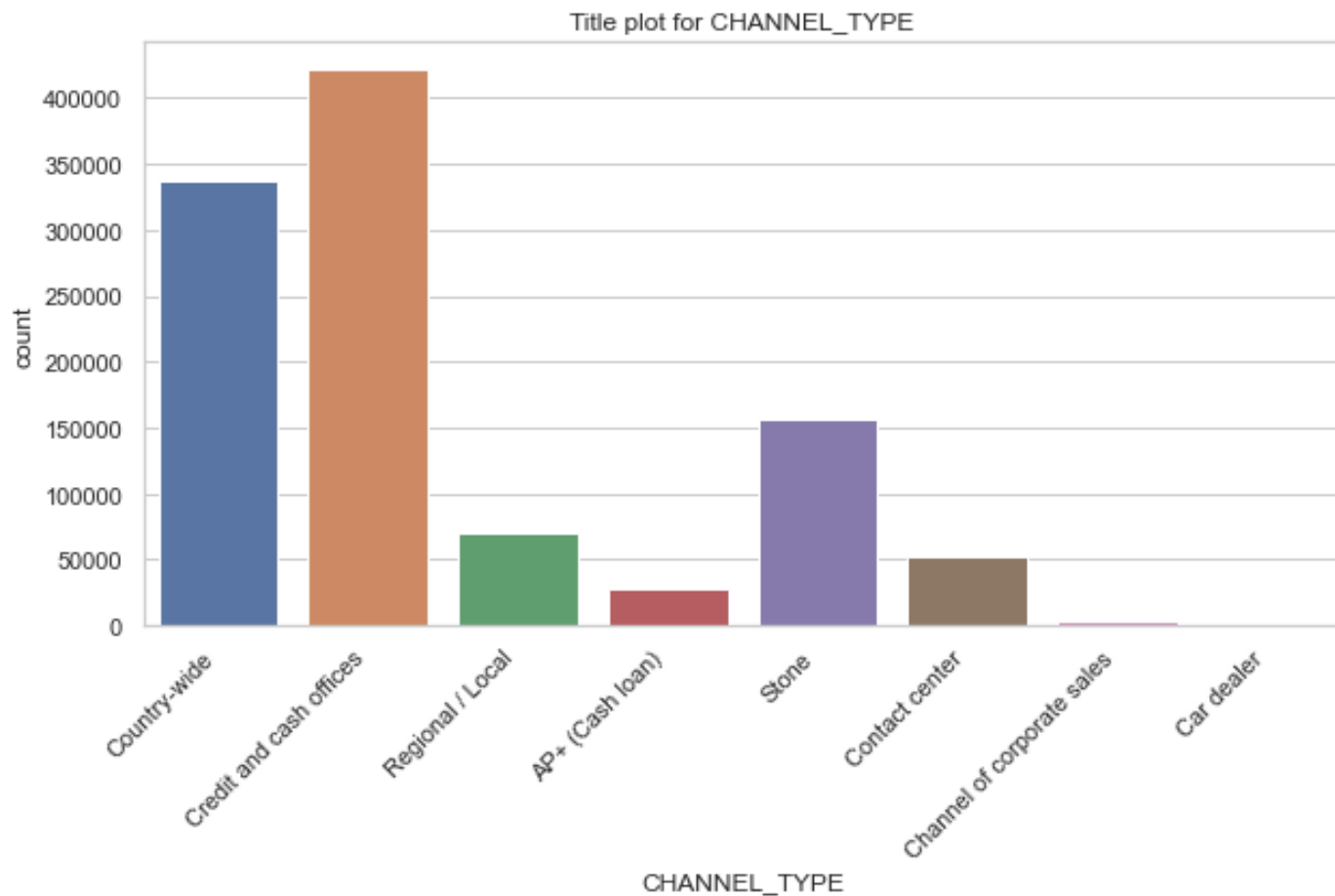# Distribution for previous portfolio

- There are high number of POS based portfolio in the previous application dataset.
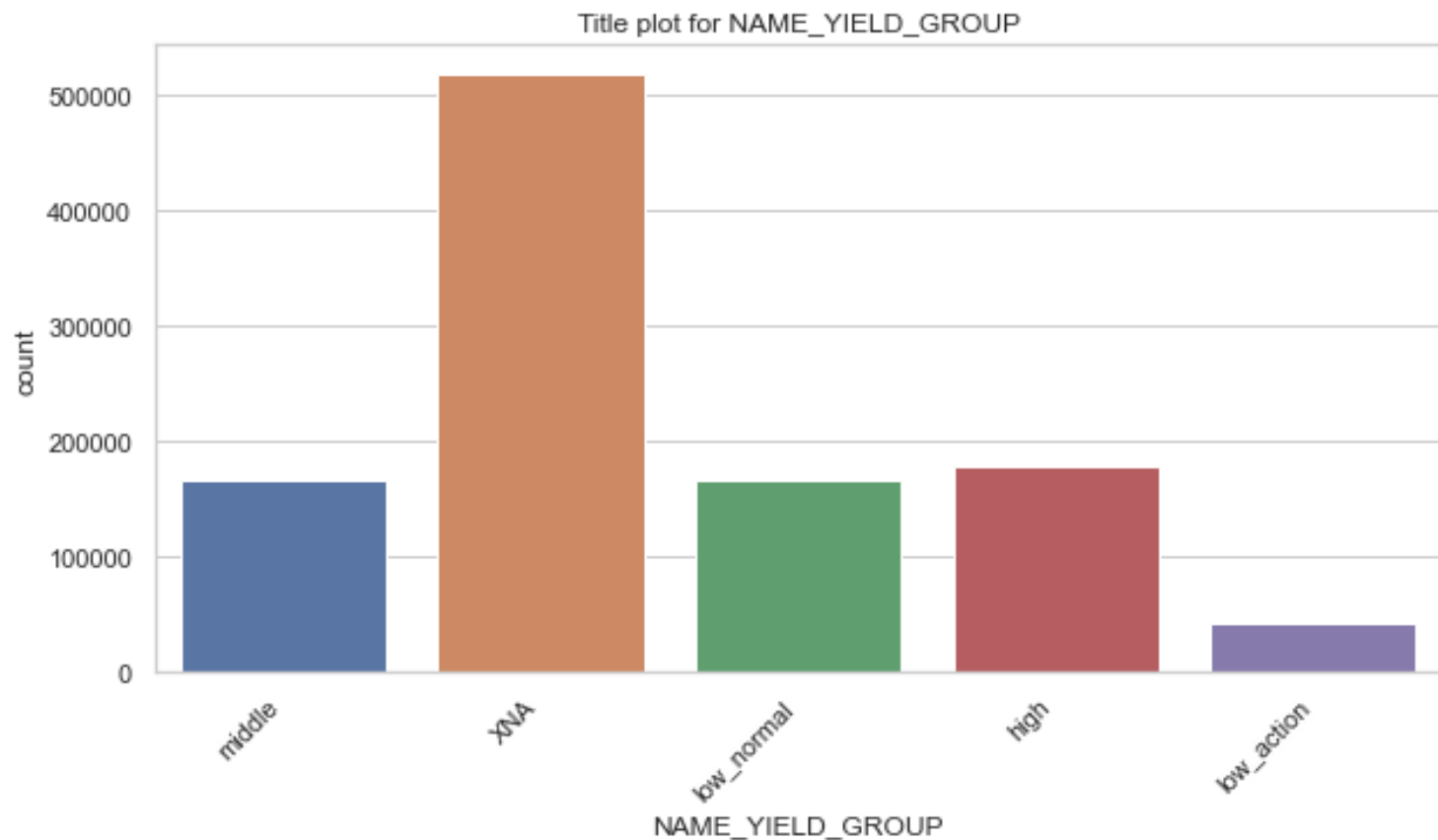
Prepared By: Utkarsh Upadhyay & Mayank Singh Soni

Credit EDA Case Study

Distribution for previous channel type

- There are high number of Credit and cash offices channel type as compared to any other channel.

Distribution for previous yield group

- There are more number of XNA yield group which can be ignored as it is logical not applicable.

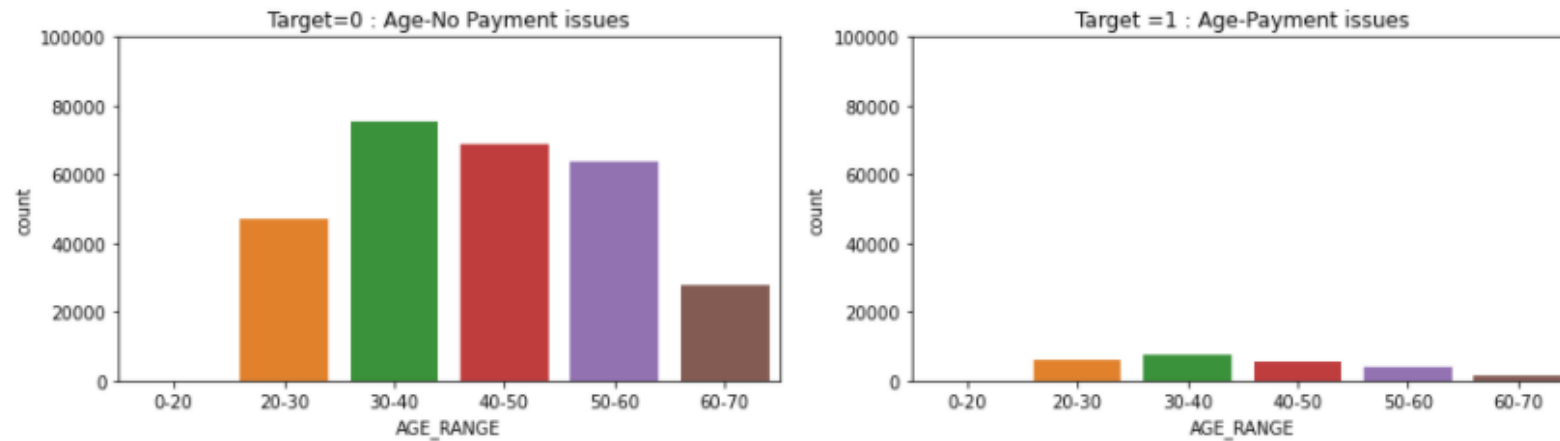- After ignoring XNA, there is high number of high yield group as compared to any other group.

Prepared By: Utkarsh Upadhyay & Mayank Singh Soni
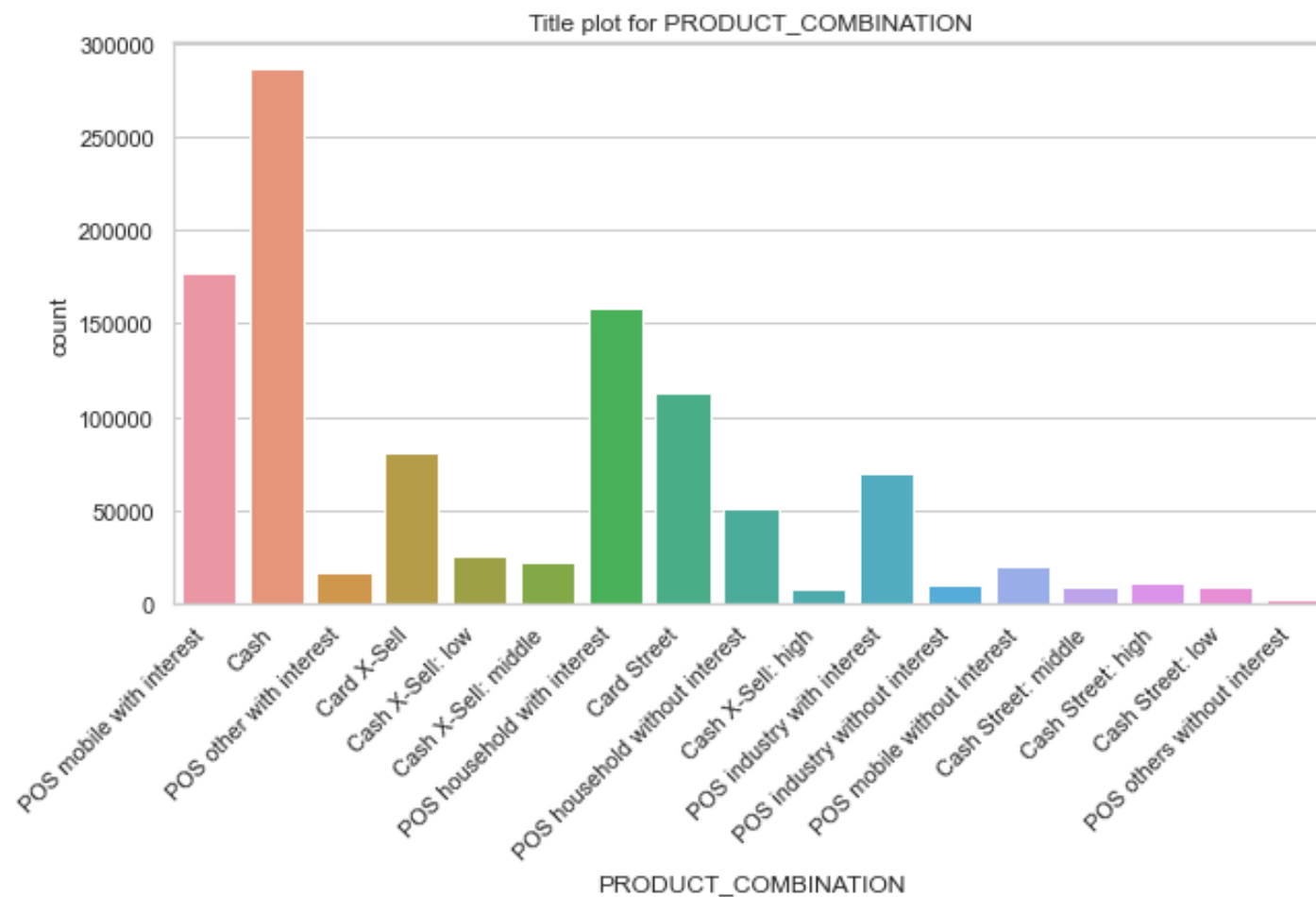
Distribution for

- **There is**

Bivariate & Multivariate Analysis
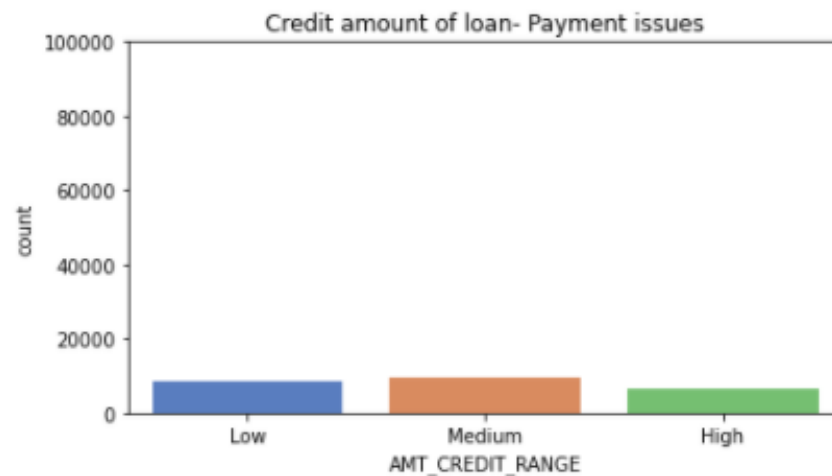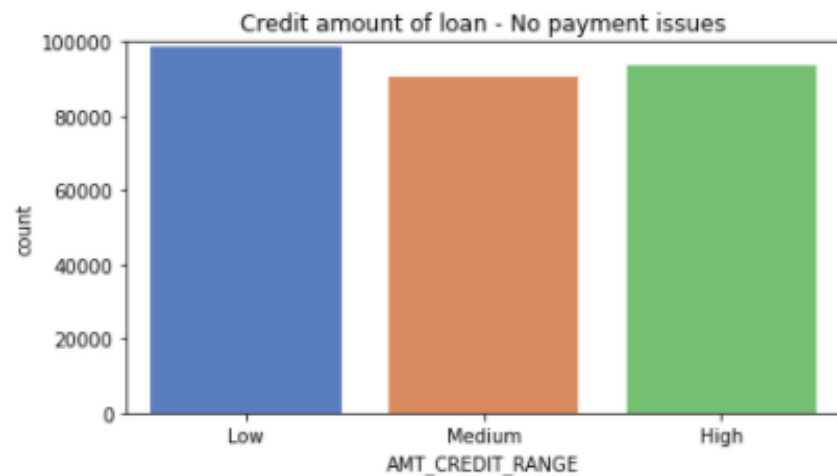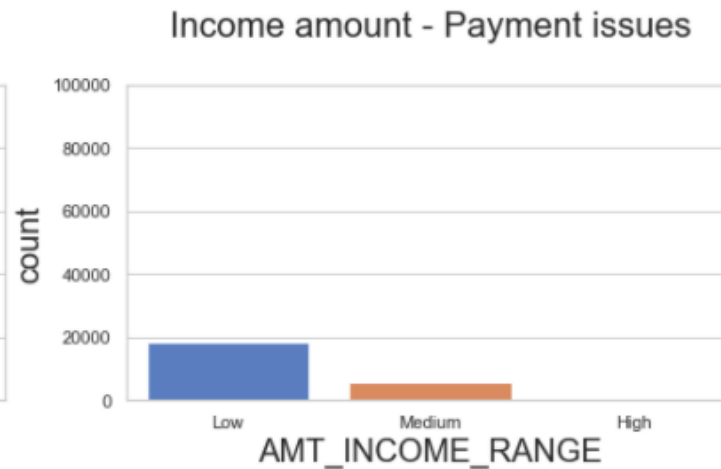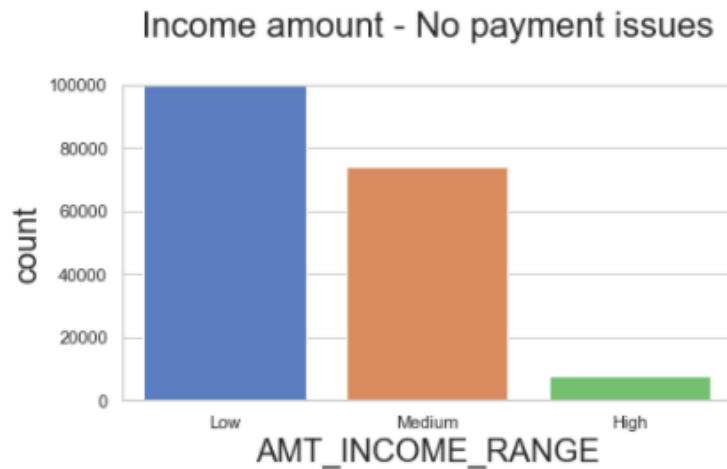
Age Vs Payment Issue

- We can observe that customers belonging to age group 30-40 & 40-50 are able to make payment on time and can be considered while lending loan

Credit EDA Case Study

Credit Amount Vs Payment Issue

- **Customers with less credit and most likely to make payment. Customers having medium and high credit can also be considered while lending the loan.**
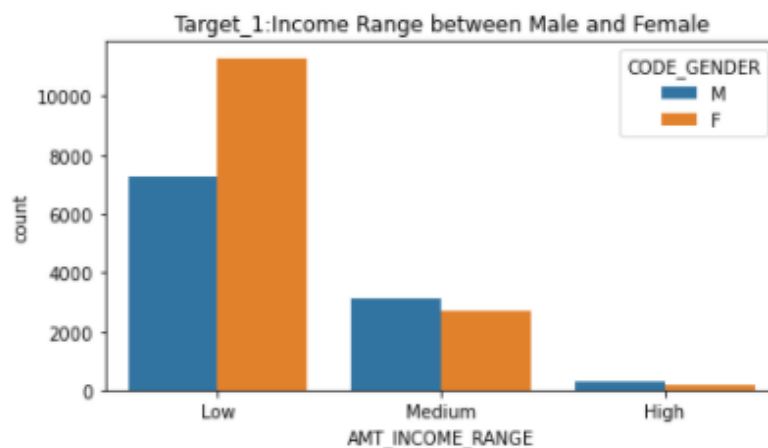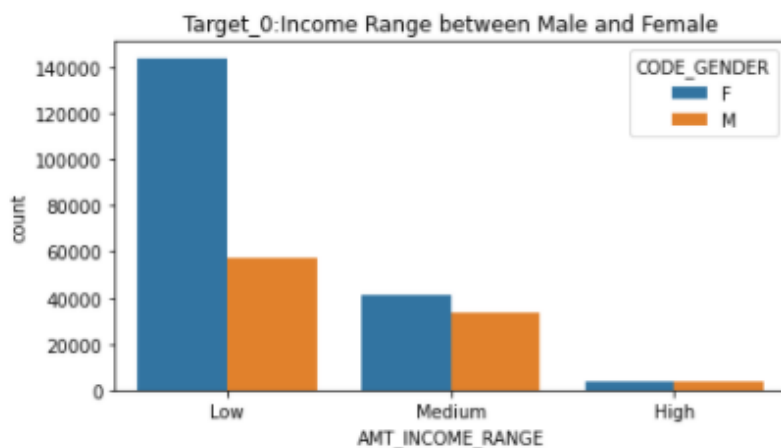
Income Amount Vs Payment Issue

- as majority lies in low income range the stats shows that people from low-income range are able to repay the loan as compared to mid/high
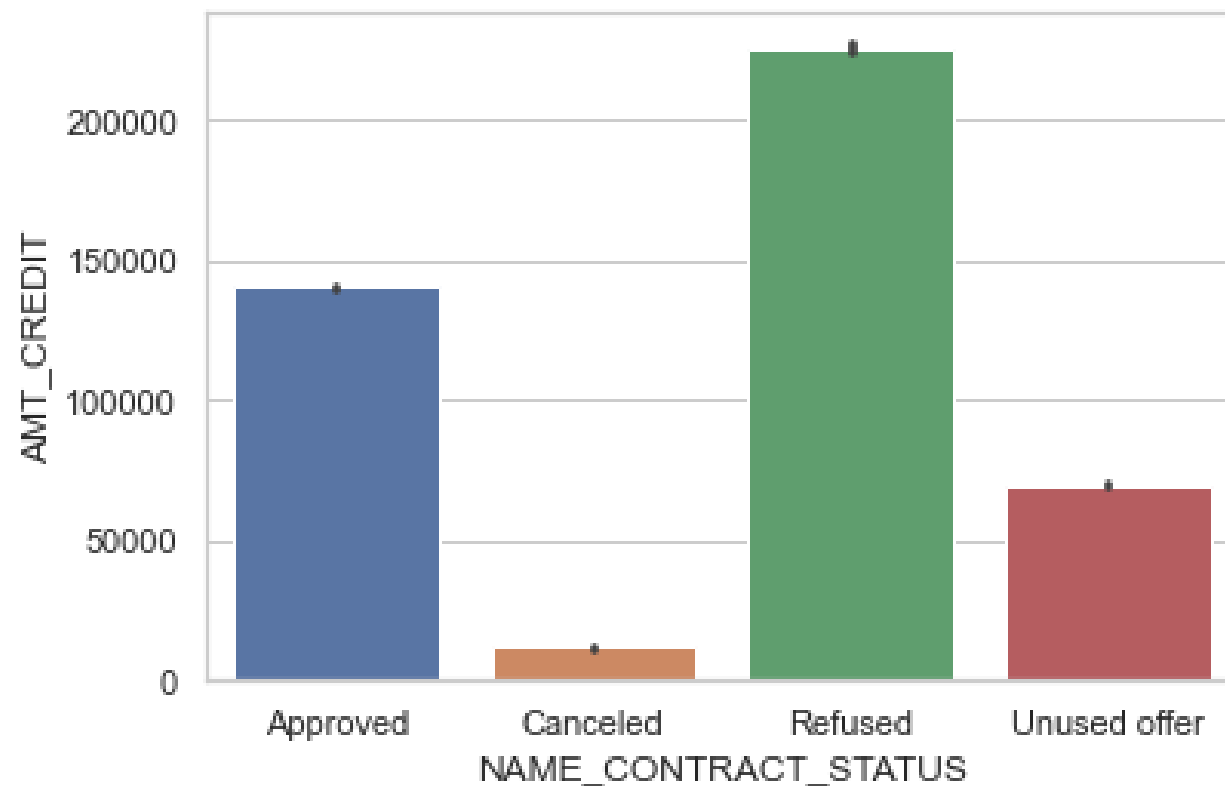
Credit EDA Case Study

Gender Vs Income Range

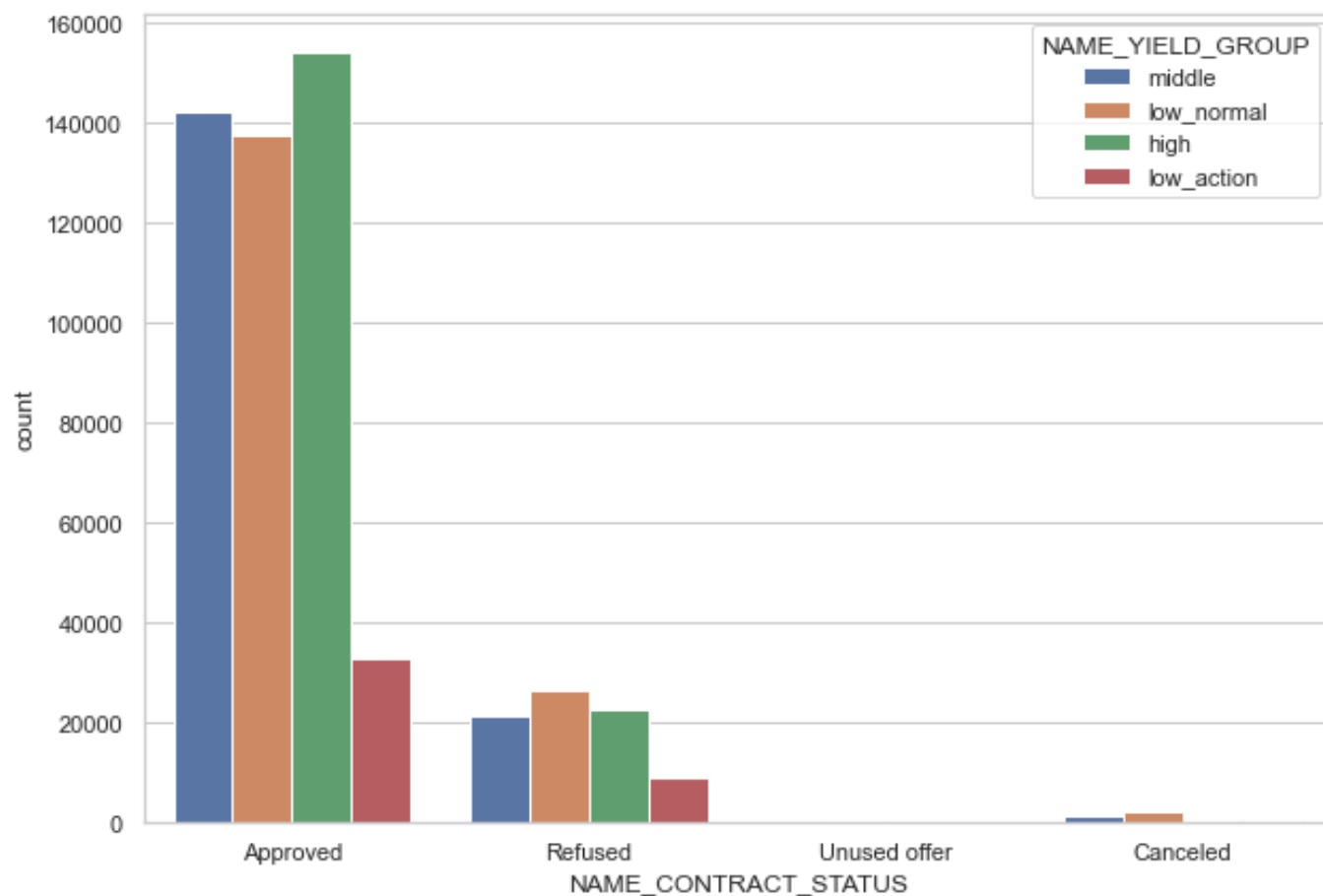- **Females with low income does not have payment issues**

Distribution for Contract Status and Credit Amount

- There is high change that loan application got refused or reject if the previous loan credit amount is higher.

Distribution for Contract Status and Yield Group

- **High grouped interest rate has maximum number of approved loans**

- **Low action grouped interest rate has least number of approved loans**

- **In medium and low normal grouped interest rate, there are records with cancelled status which means if interest rate falls under medium or low normal group then there is chance that user cancel the loan application**

Prepared By: Utkarsh Upadhyay & Mayank Singh Soni
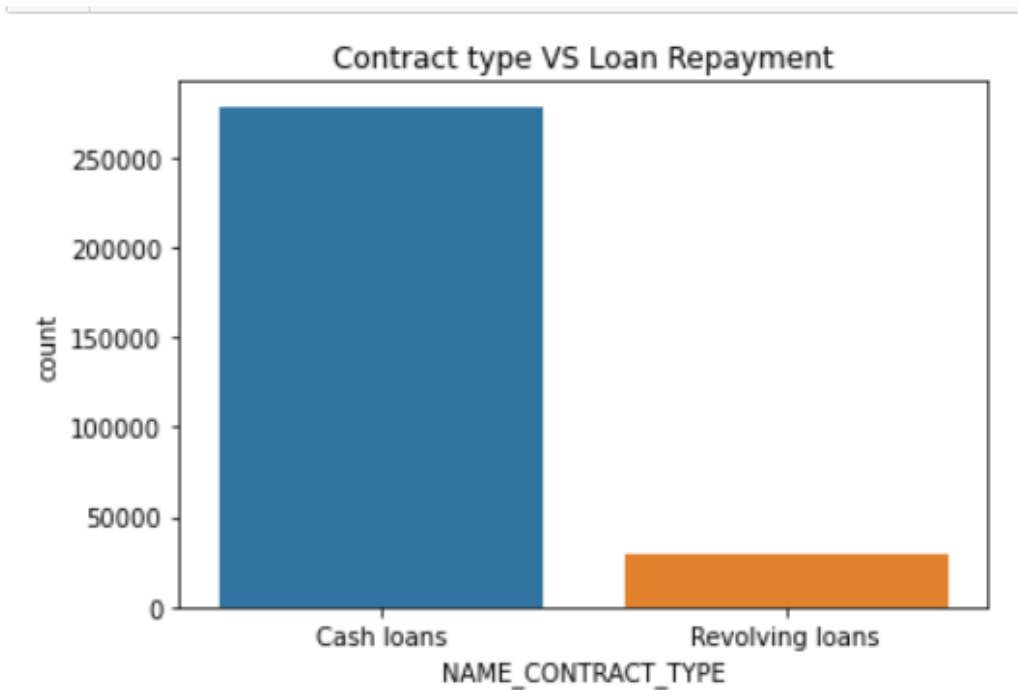
Credit EDA Case Study

Distribution for Cash loan Purpose and Contract Status
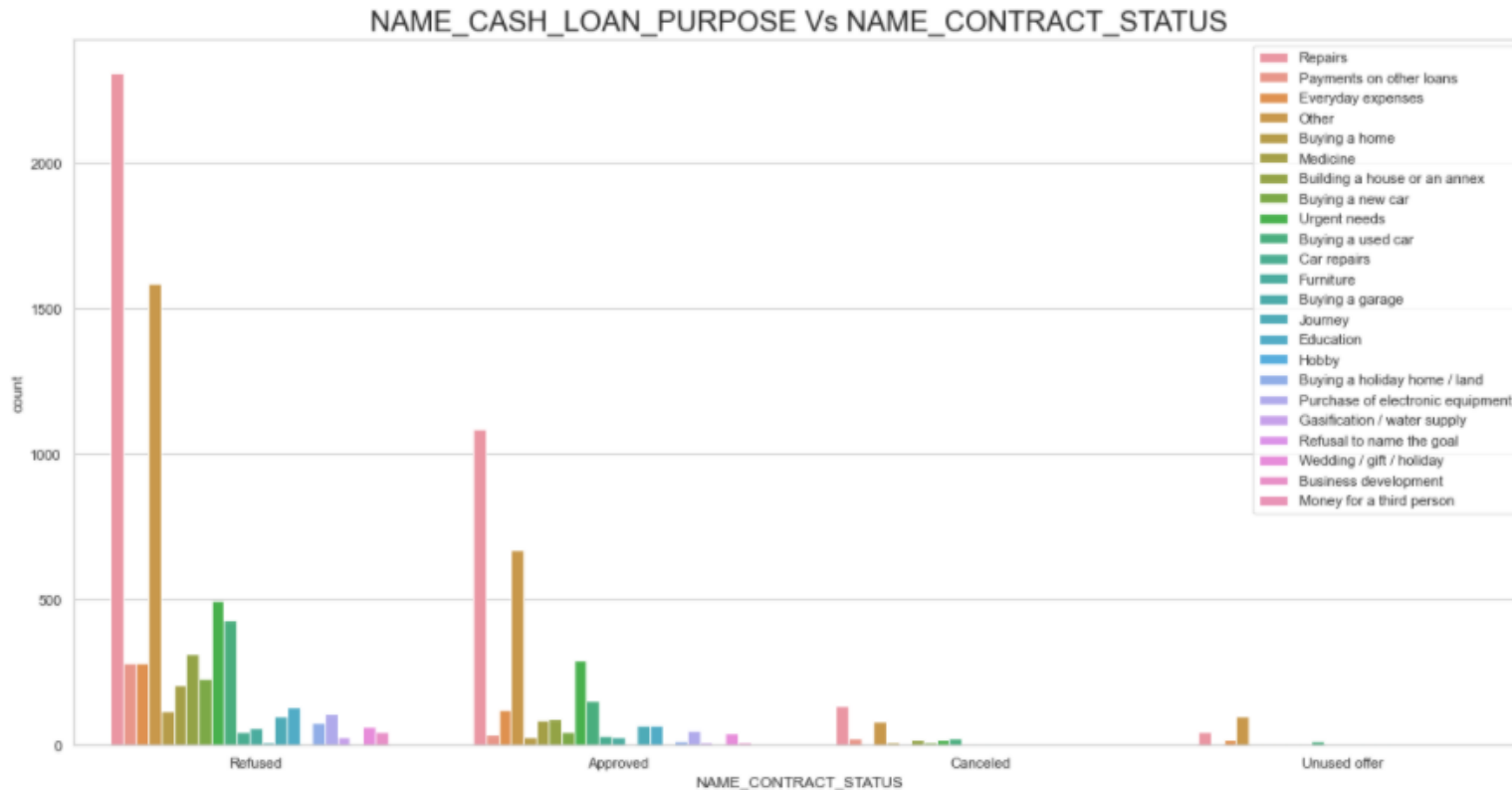
- There are high refused loan application for cash loan purpose of Repair and Everyday expenses.

Prepared By: Utkarsh Upadhyay & Mayank Singh Soni
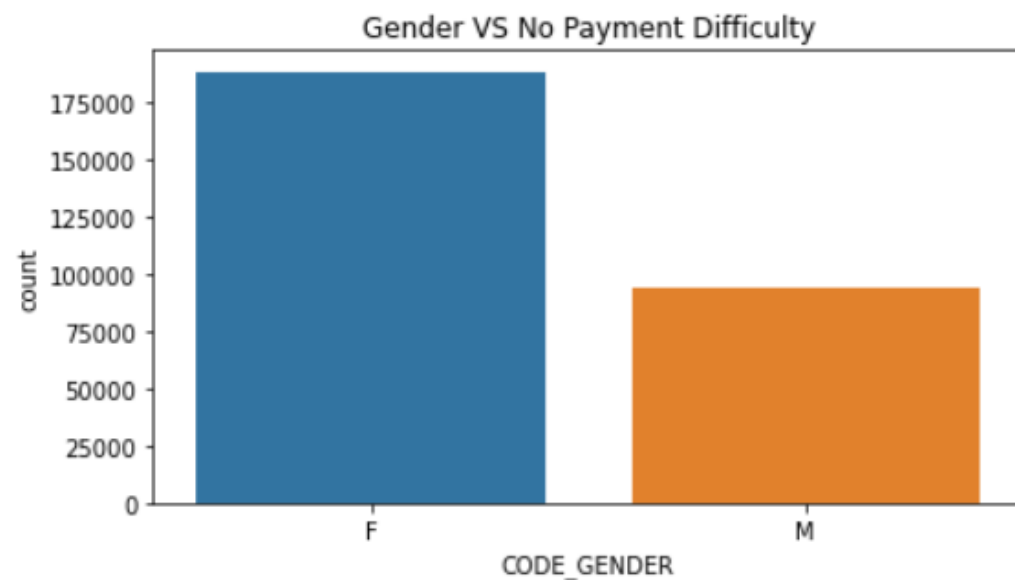
Credit EDA Case Study

Revolving loans are just a small fraction from the total number of loans and are less repaid as compared to Cash Loans
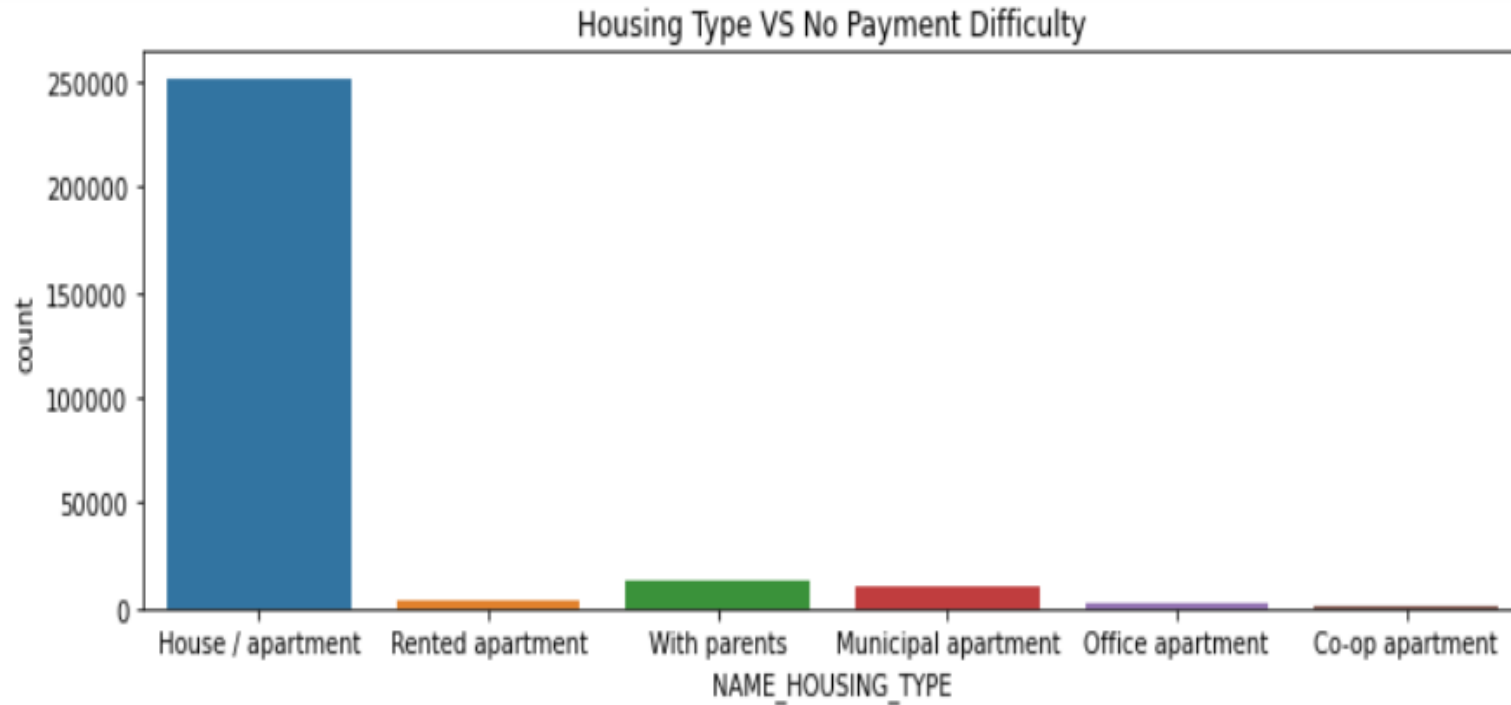
The number of female clients is almost double the number of male clients.

- Based on the percentage of defaulted credits, males have a higher chance of not returning their loans as compared with women
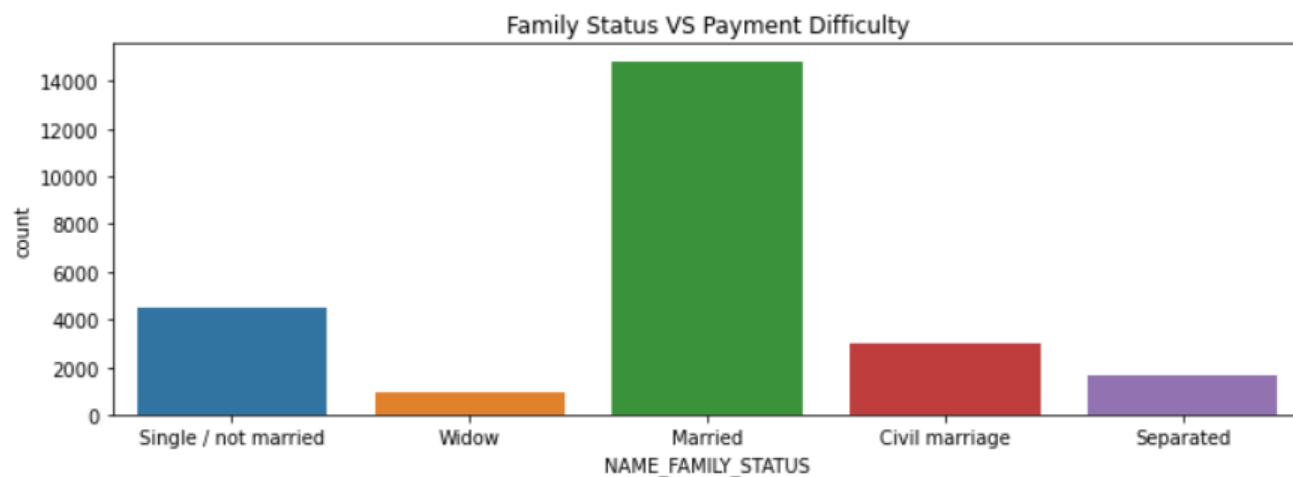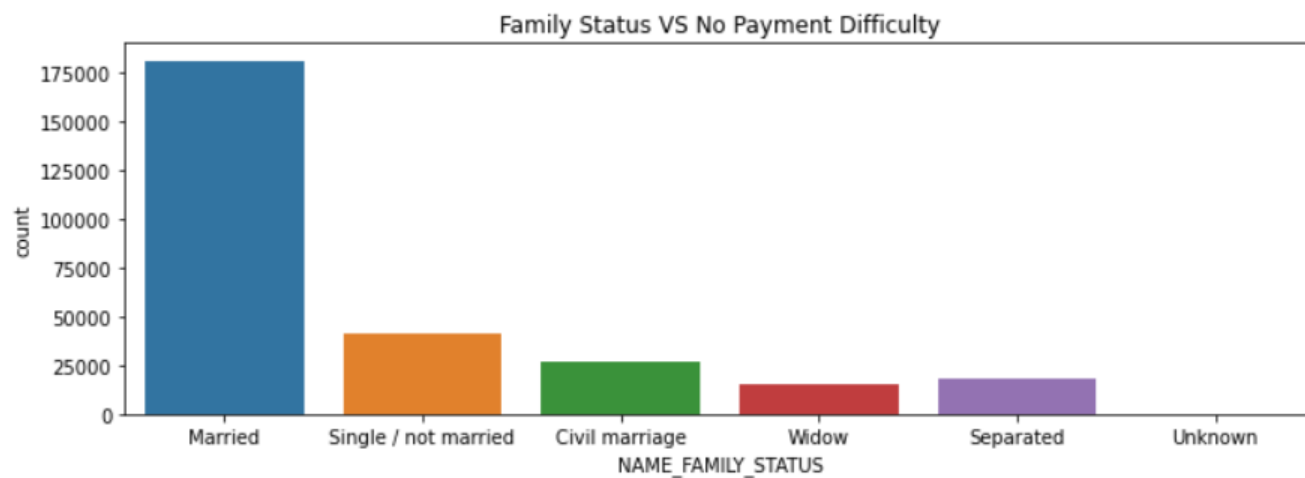
Housing Type VS No Payment Difficulty

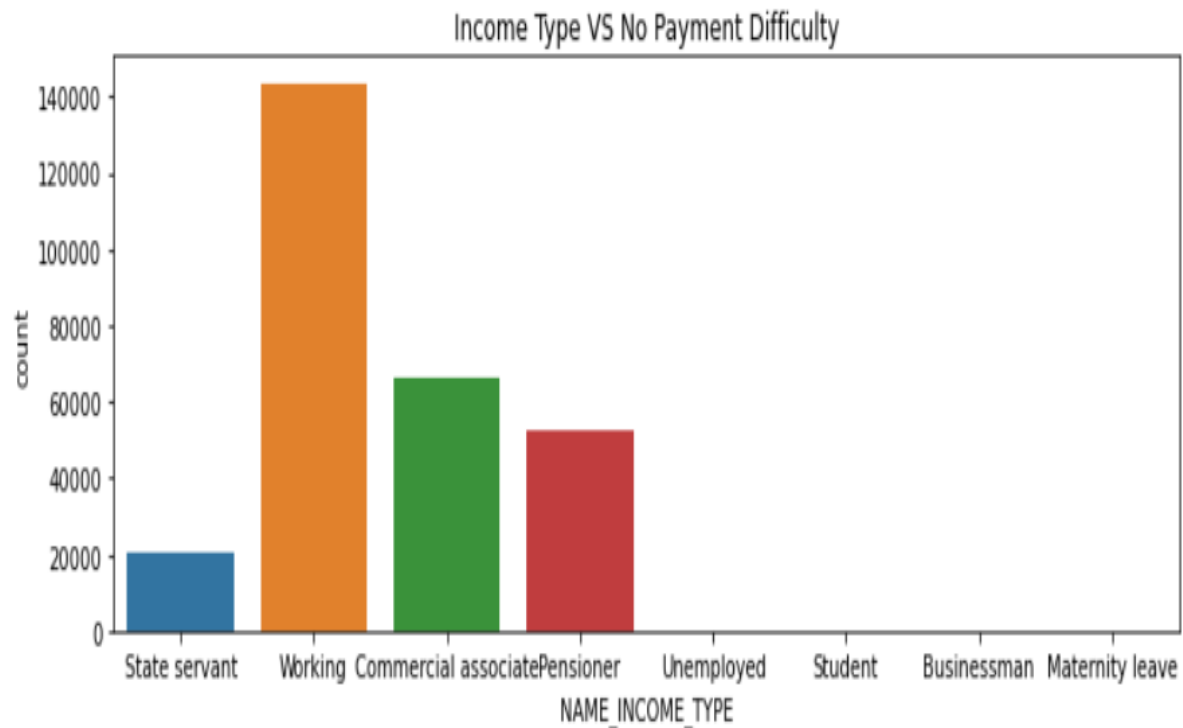**Majority of people live in House/apartment**

People living in office apartments have lowest default rate

Credit EDA Case Study

Family Status vs Loan Payment

Most of the people who have taken loan are married, followed by Single/not married and civil marriage

Income Type Vs Payment Difficulty

Most of applicants for loans have income type as Working, followed by Commercial associate, Pensioner and State servant.

Student and Businessmen, though less in numbers do not have any default record.

Prepared By: Utkarsh Upadhyay & Mayank Singh Soni

Credit EDA Case Study

% of Loan Payment Difficulties for NAME_CONTRACT_STATUS and NAME_CONTRACT_TYPE
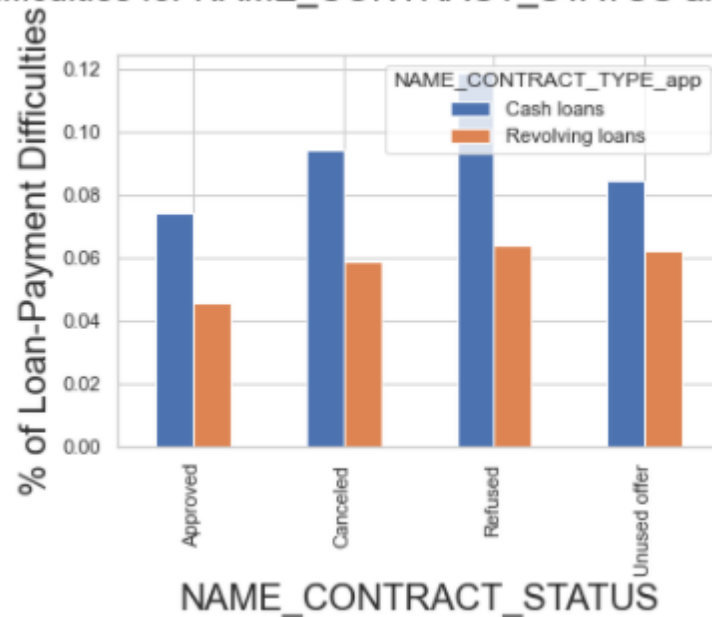
% of Loan Payment Difficulties for NAME_CONTRACT_STATUS and NAME_CONTRACT_TYPE

- Percentage of rejection is high in cash loans

- Percentage of approved loan for revolving loan is comparatively less.

Prepared By: Utkarsh Upadhyay & Mayank Singh Soni

Thank-You