CREDIT EDA CASE STUDY

PROBLEM STATEMENT

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history.

- So we have to use EDA to analyse the patterns present in the data

Objectives of the study

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

- To understand the driving factors (or driver variables) behind loan default

- To understanding the types of variables and their significance

Steps In The Analysis

- Reading & Understanding the data

- Data Cleaning & Manipulation

    - Null Value Calculation

    - Analyze & Delete Unnecessary Columns in application data

    - Analyze & Delete Unnecessary Columns in previous application data

    - Standardize Values

    - Data Type Conversion

    - Null Value Data Imputation

    - Identifying the outliers

- Data Analysis

    - Categorical Variables Analysis

    - Numeric Variables Analysis

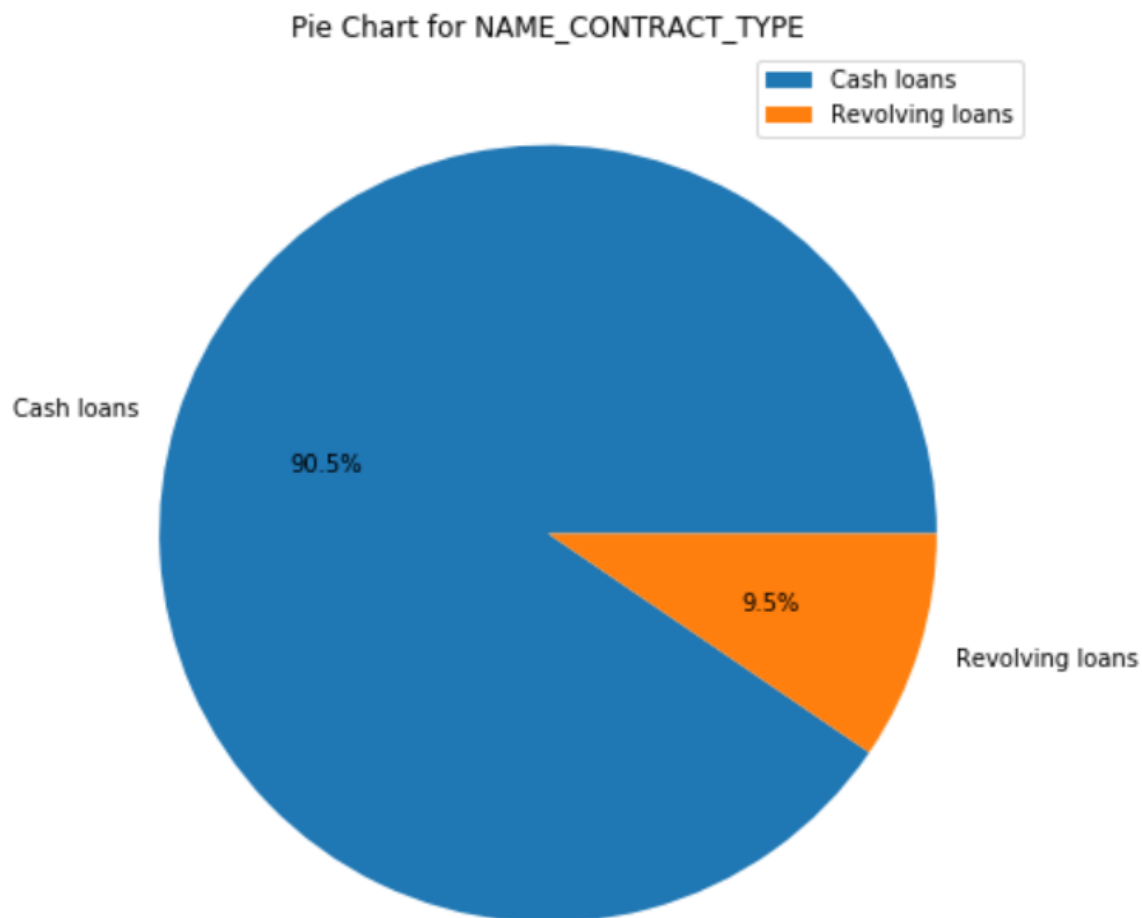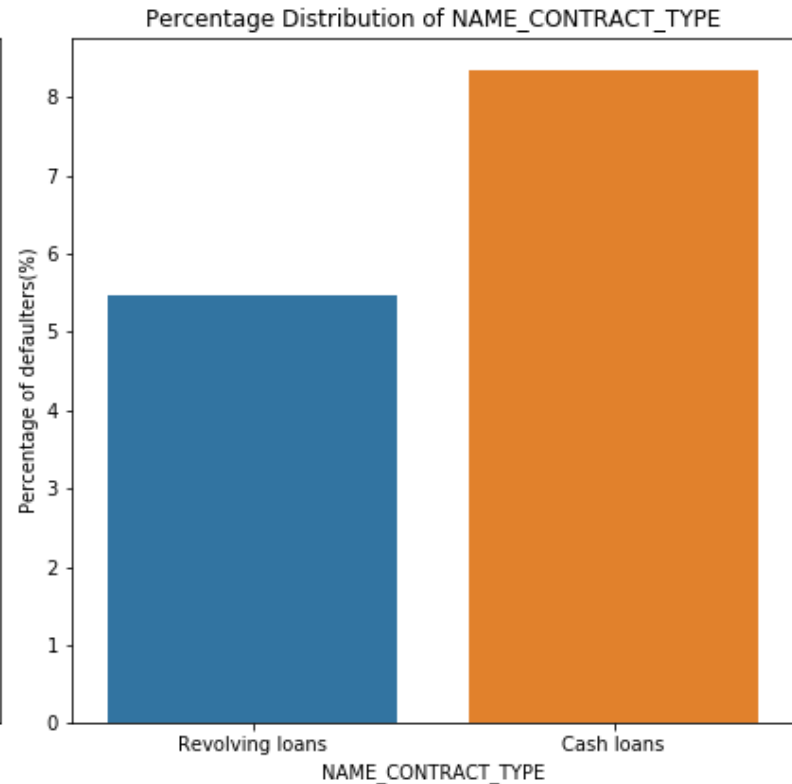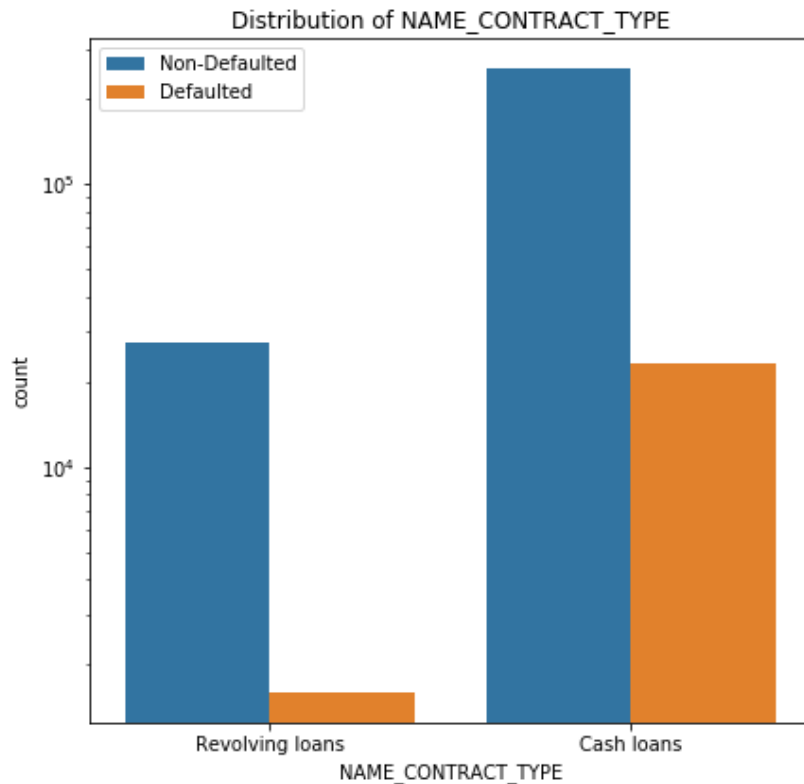- Merged Dataframe Analysis

Checking Data Imbalance
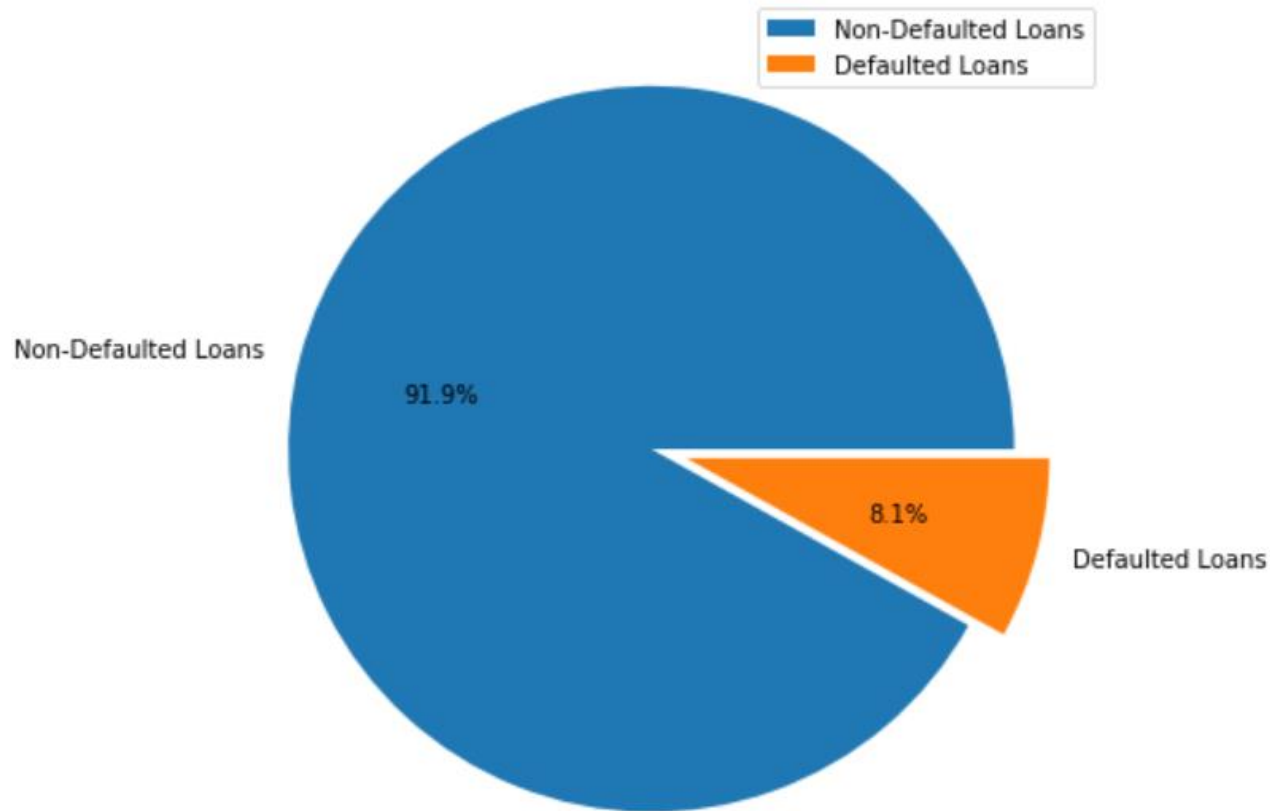
- The dataset is imbalanced with loans being Non-Defaulted around 92% and Defaulted around 8%

ANALYSIS OF APPLICATION DATA

UNIVARIATE ANALYSIS

Distribution based on Contract Type

Pie Chart for NAME_CONTRACT_TYPE

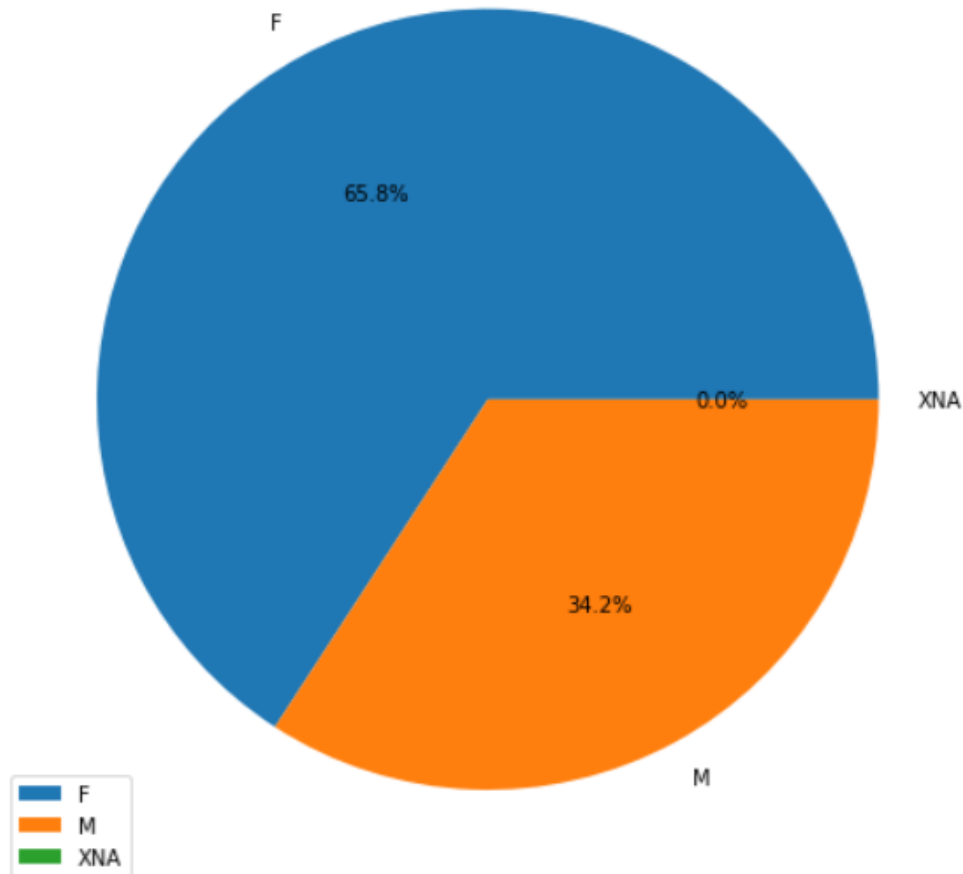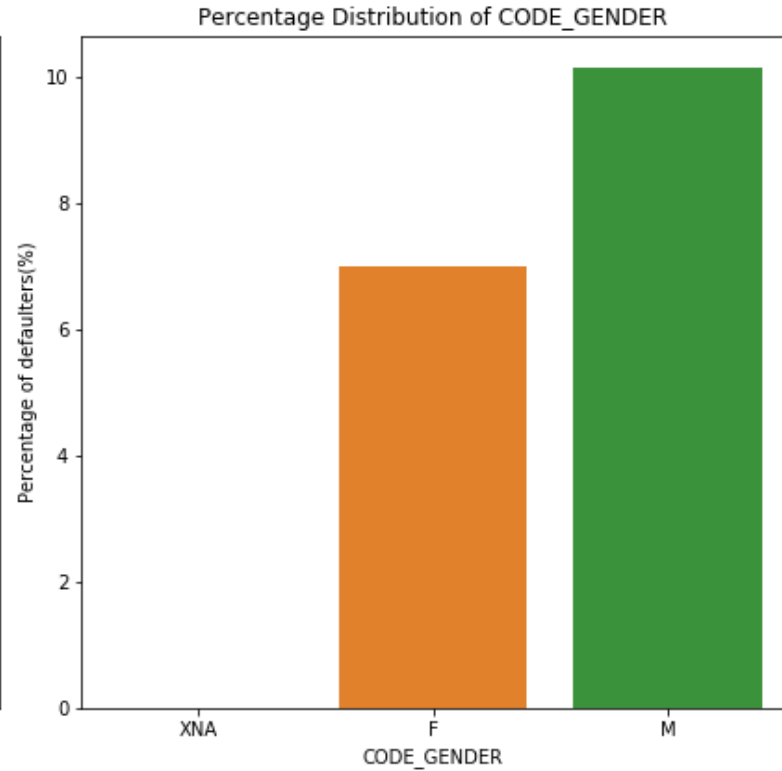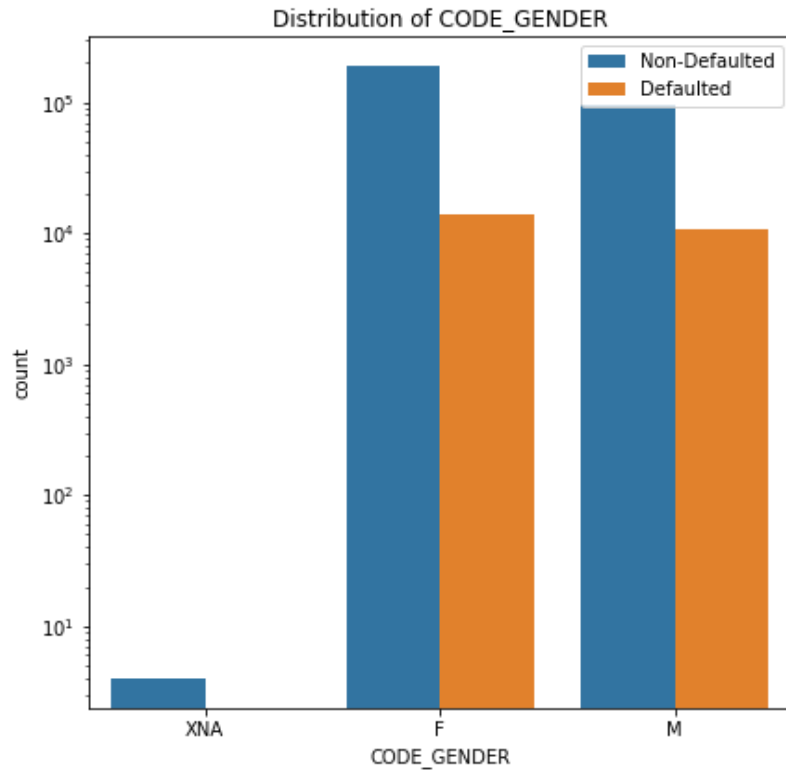Distribution of NAME_CONTRACT_TYPE / Percentage Distribution of NAME_CONTRACT_TYPE

•Cash Loans dominate the Loan category with over 90%
•Contract type - Revolving are smaller fraction in overall category, but have a large percentage of Defaulters
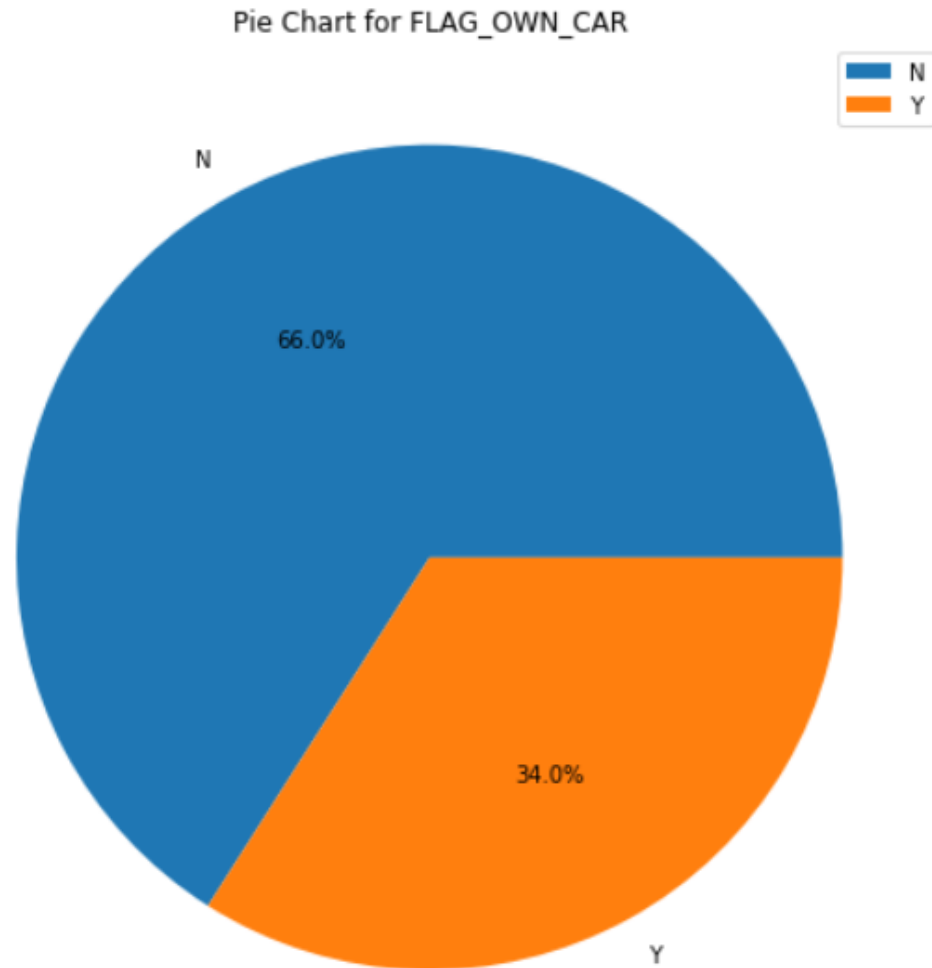
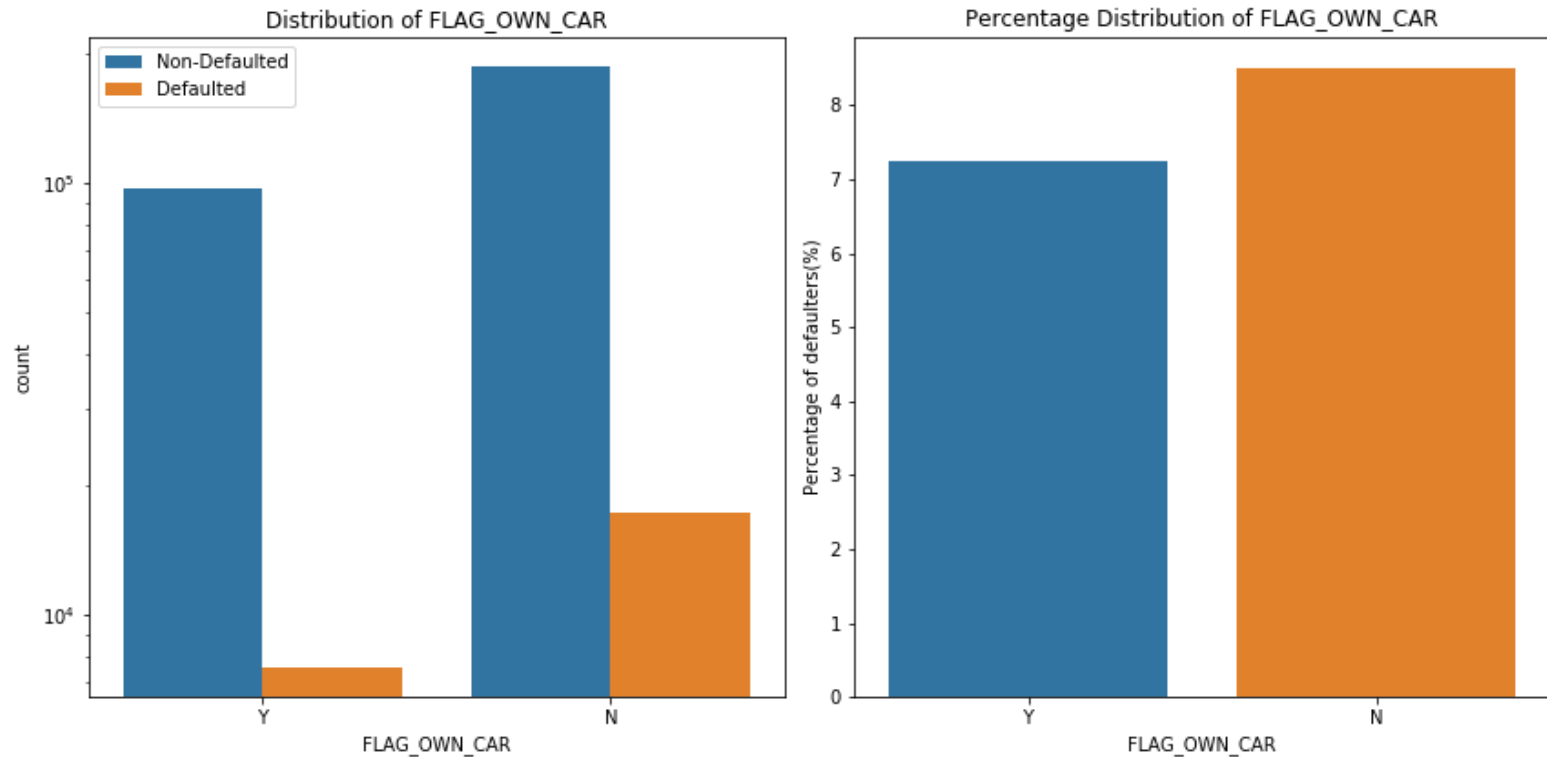# Distribution based on Gender

Pie Chart for CODE_GENDER

•Overall Females have higher percentage of Loans.
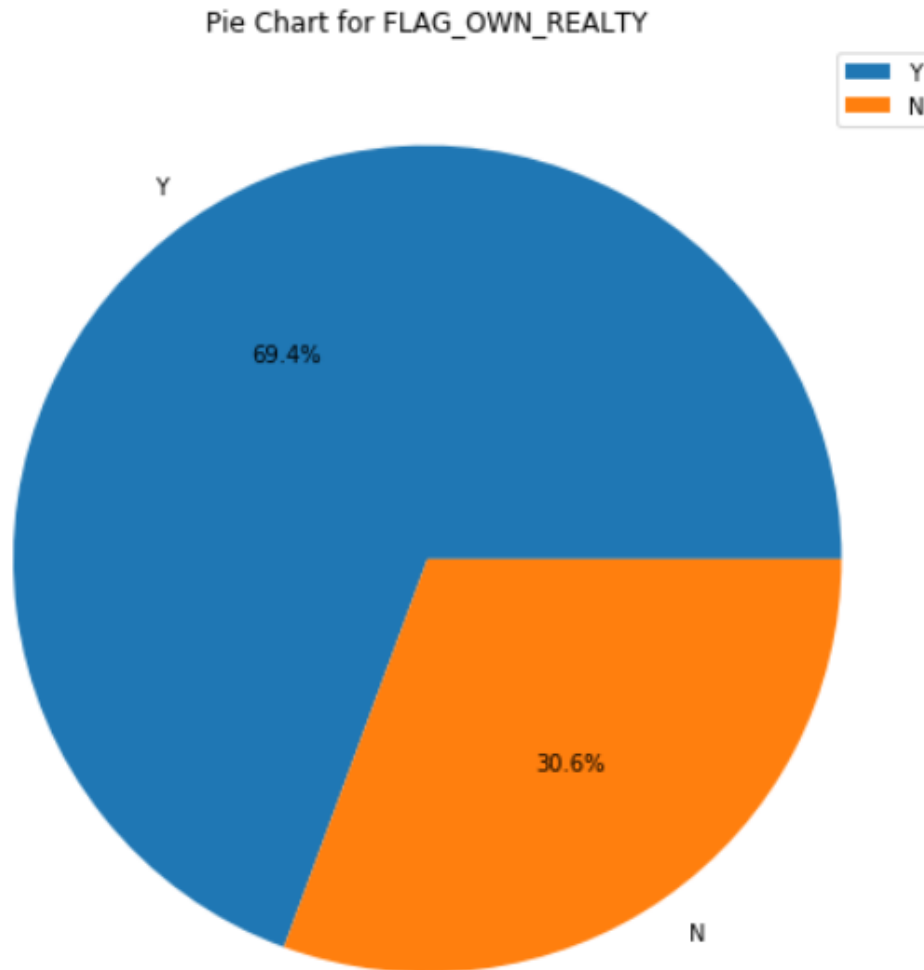•Males are on the higher end in defaulted loans category

Distribution based on Car Ownership

Pie Chart for FLAG_OWN_CAR

N
Y

N

66.0%

34.0%

Y

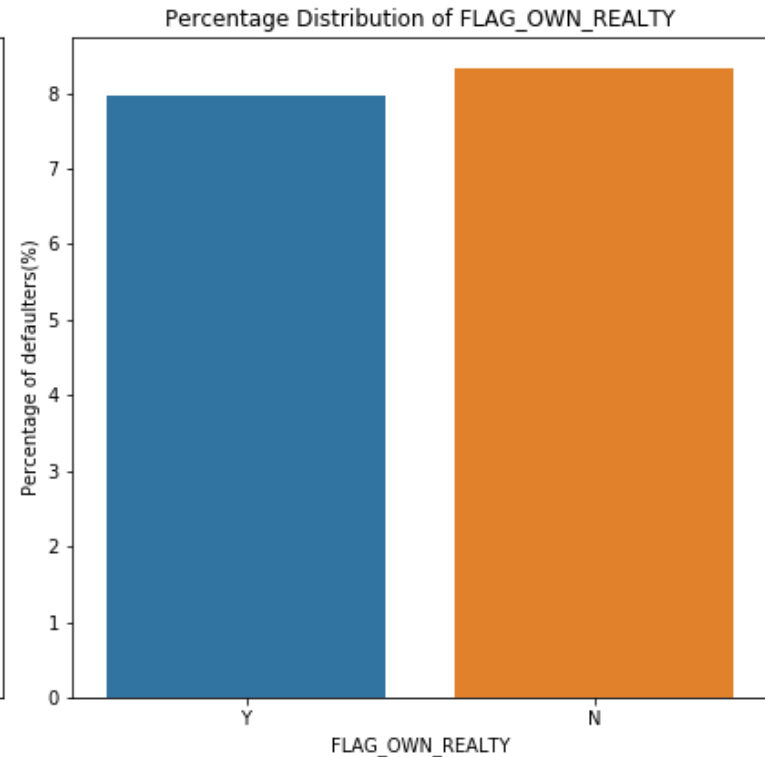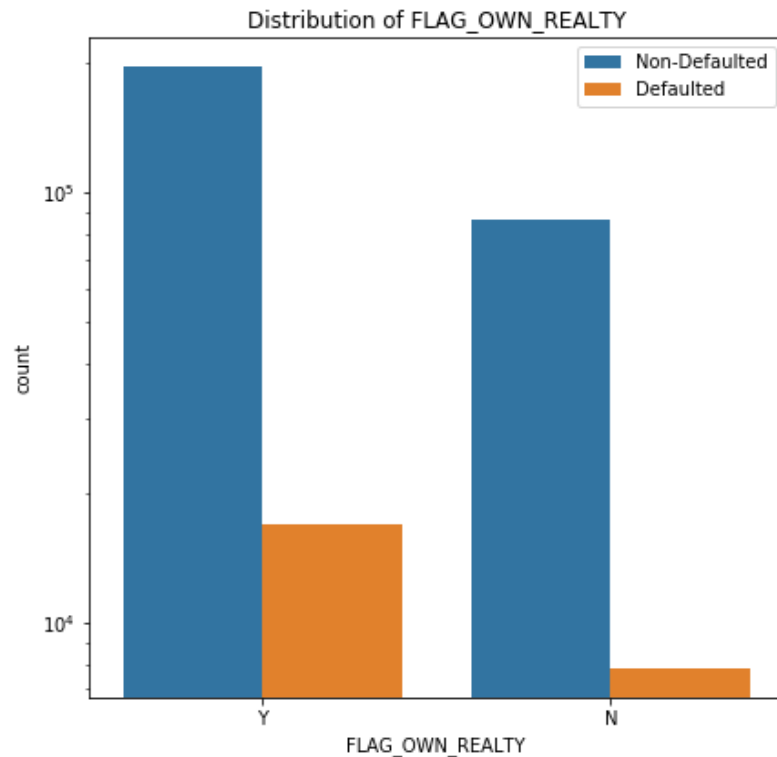Distribution of FLAG_OWN_CAR — Percentage Distribution of FLAG_OWN_CAR

•Majority of applicants do not own a car.
•Applicants who do not own a car are the major defaulters. But both categories have non-repayment above 70%

Distribution based on House Ownership

Pie Chart for FLAG_OWN_REALTY

Y — 69.4%

N — 30.6%

Distribution of FLAG_OWN_REALTY / Percentage Distribution of FLAG_OWN_REALTY
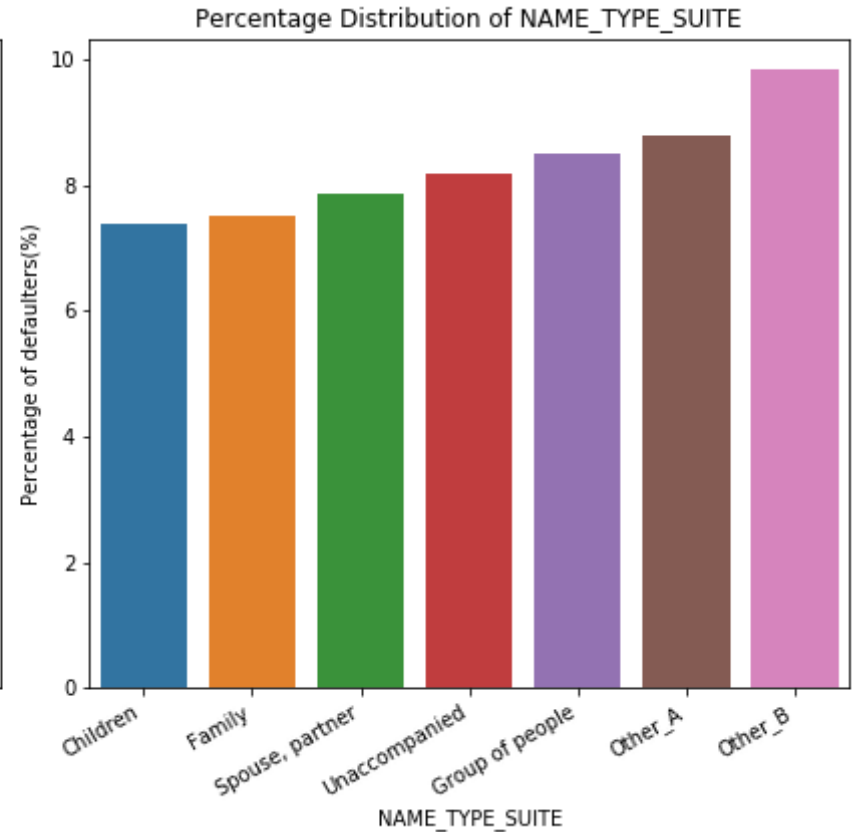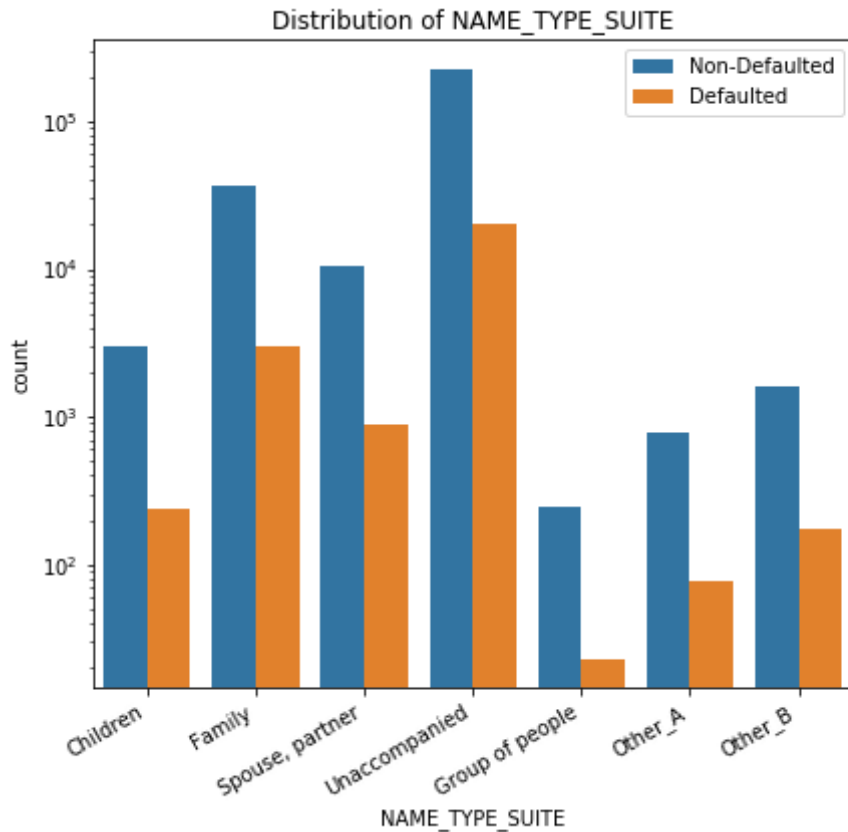
- Majority of applicants own a house or flat.
- Both categories have non-repayment rates around 80%

Distribution based on Name Type Suite

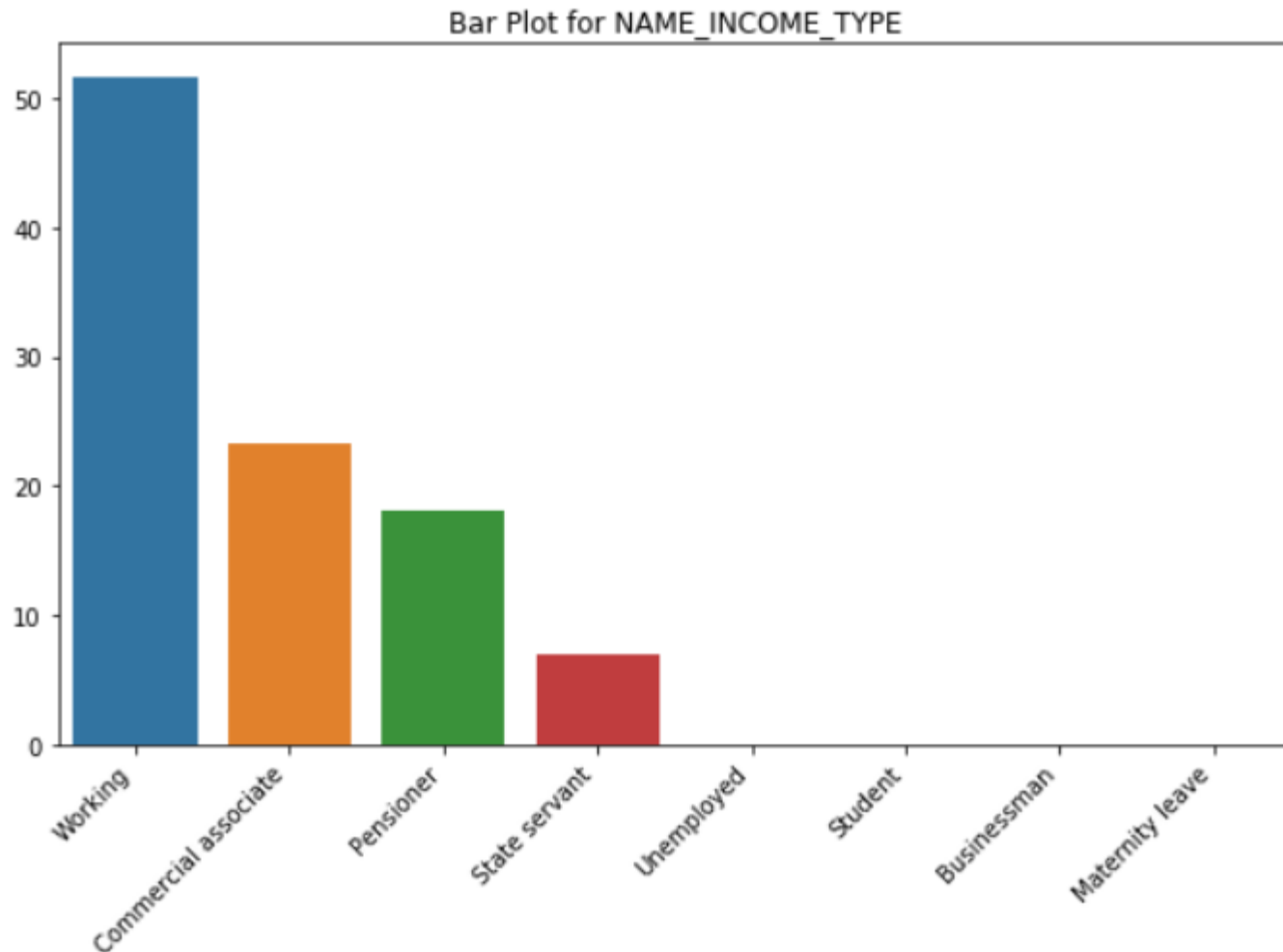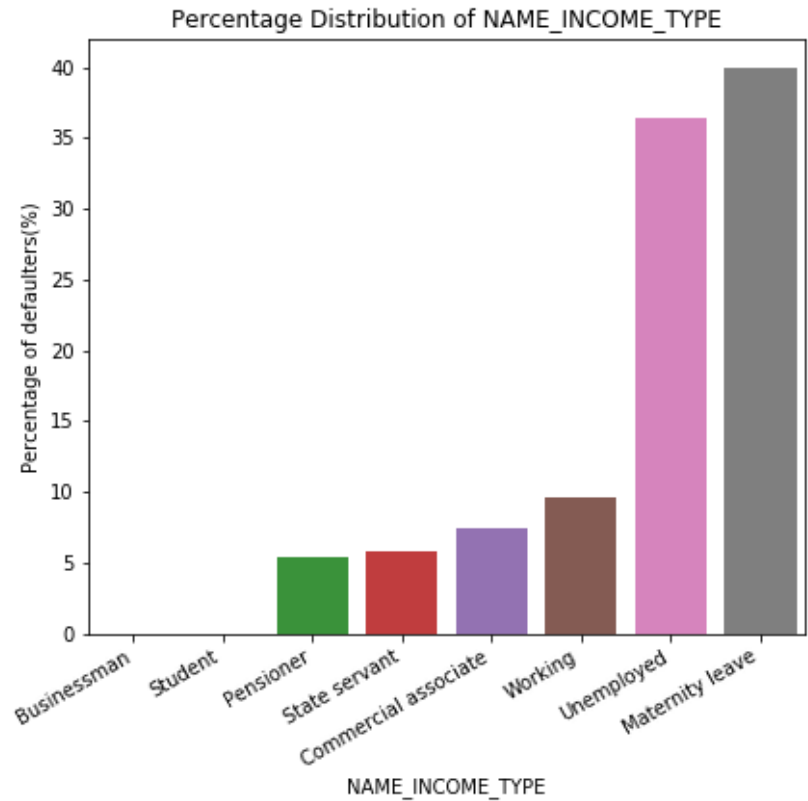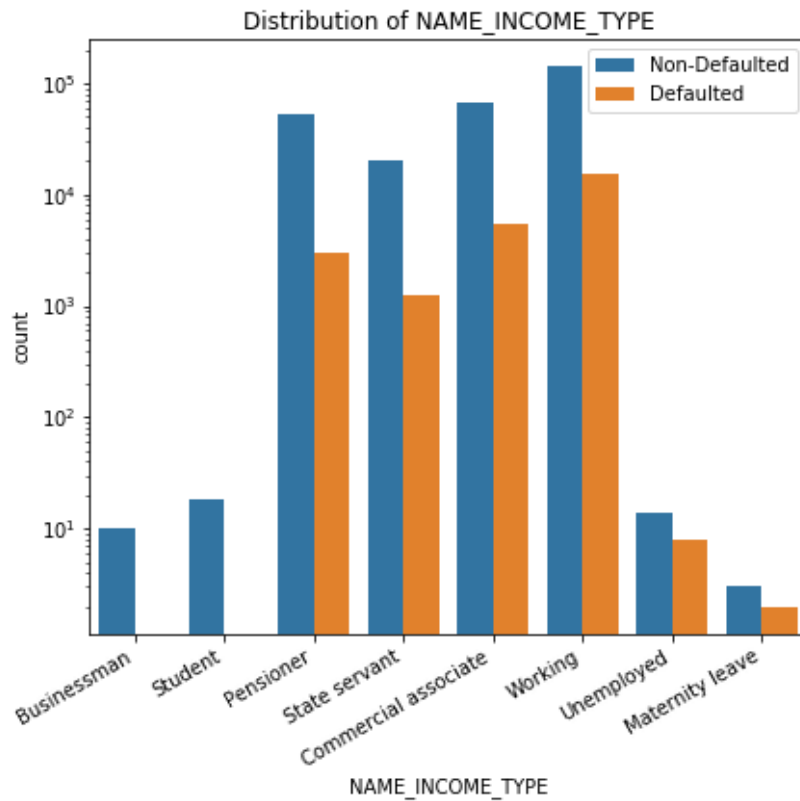Bar Plot for NAME_TYPE_SUITE
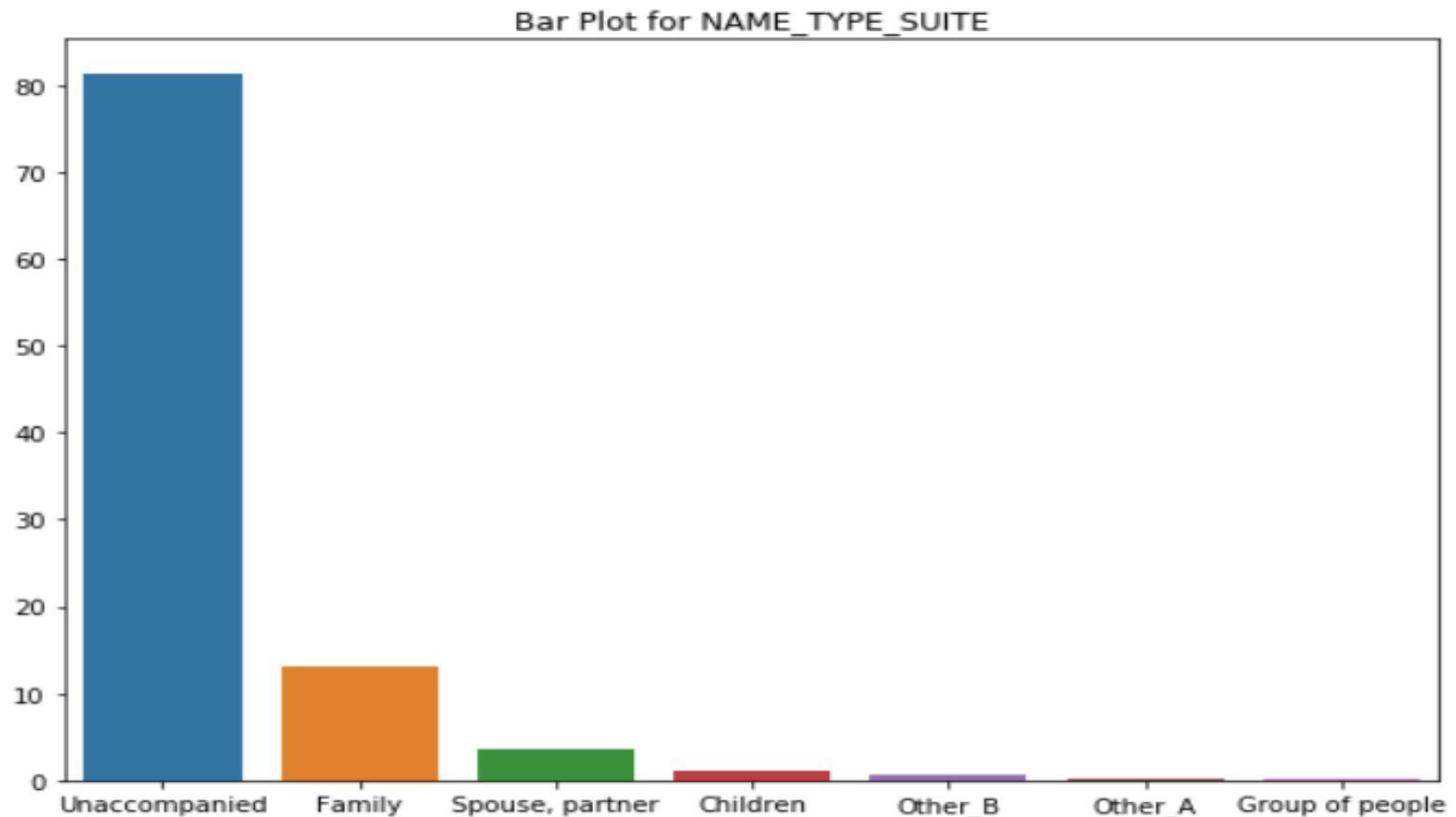
- Majority of the Loans are of Individual applicants
- Group Applications are the lowest
- Children and Family Co-Applicants have lower percentage of repayment difficulty

Distribution based on Income

Bar Plot for NAME_INCOME_TYPE

Distribution of NAME_INCOME_TYPE / Percentage Distribution of NAME_INCOME_TYPE

•Working people, Commercial associates, Pensioners and State servants are the major applicants for the loan (Credit).
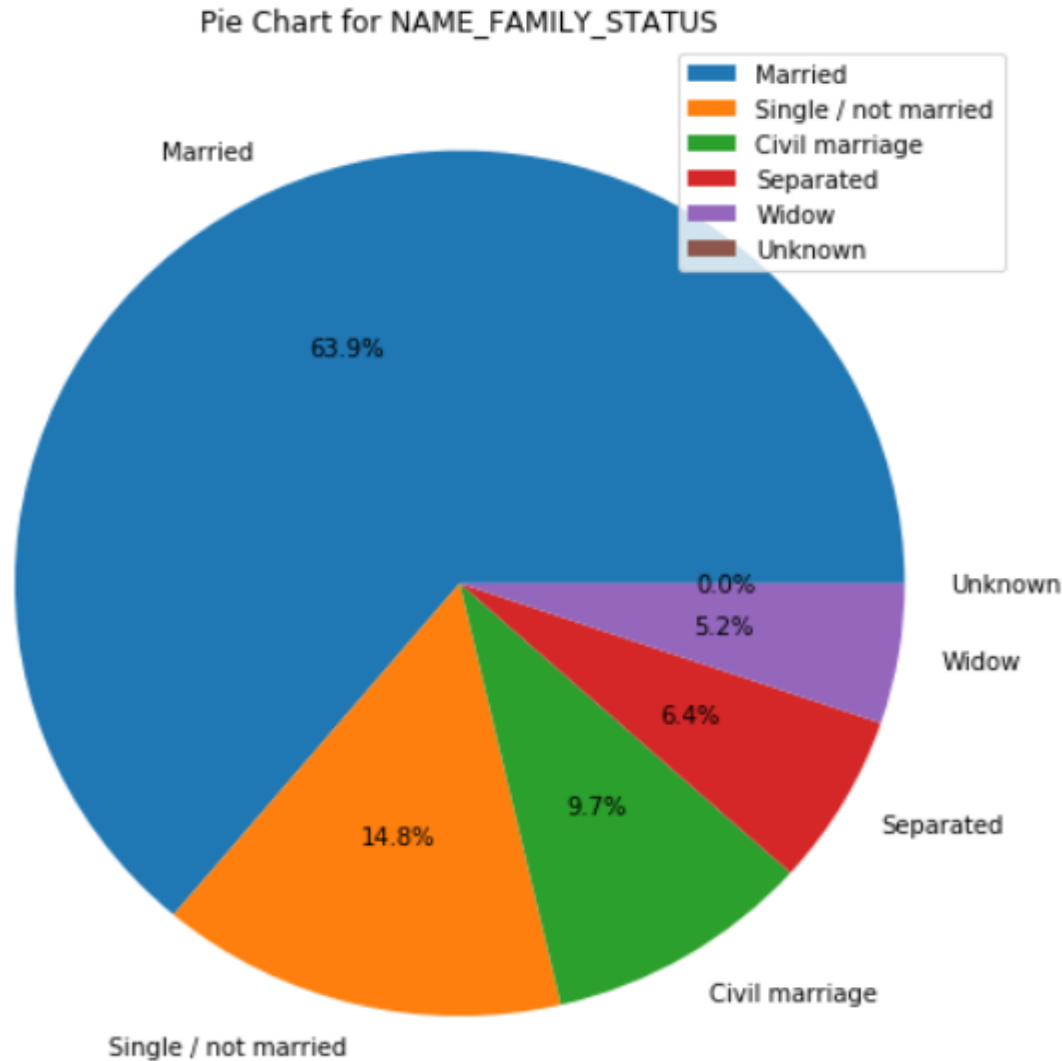•Applicants with 'Maternity Leave' Income Type have maximum % of Loan-Paymend difficulties (40%)
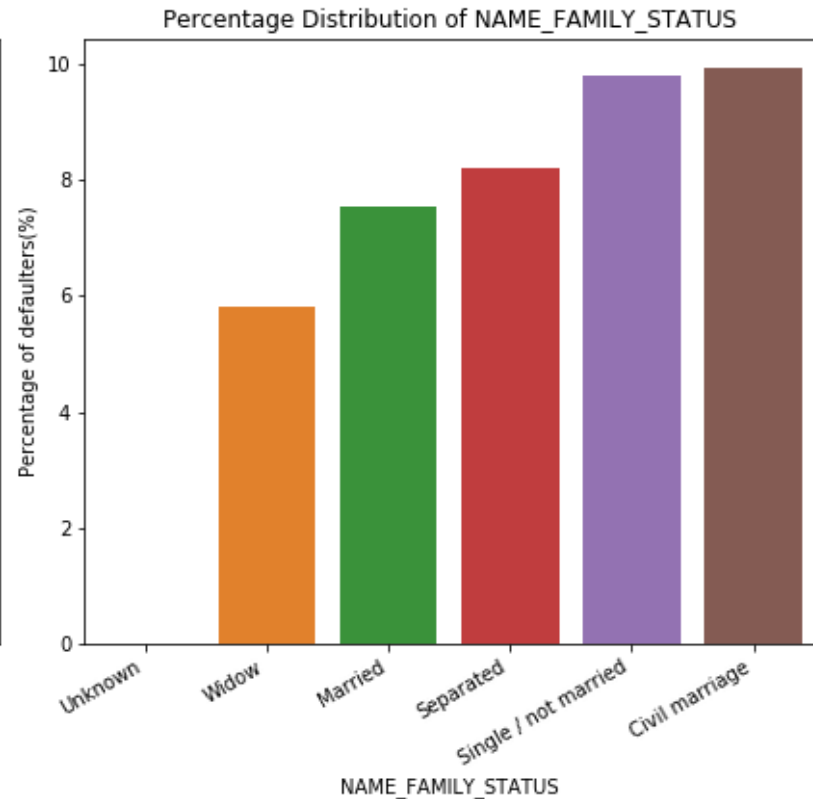
Distribution based on Education

Distribution of NAME_EDUCATION_TYPE
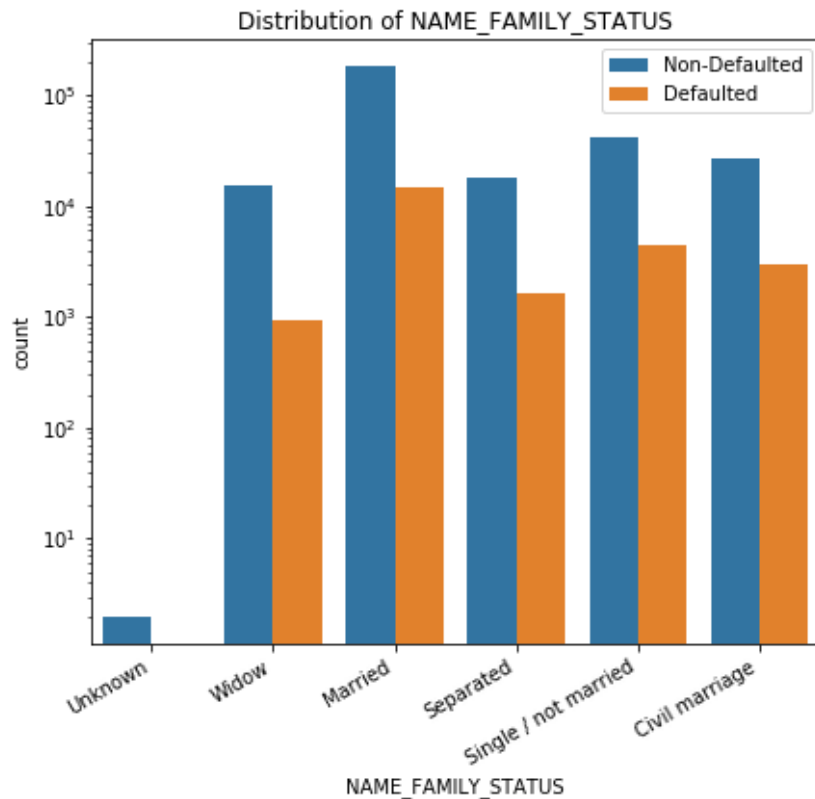
Percentage Distribution of NAME_EDUCATION_TYPE
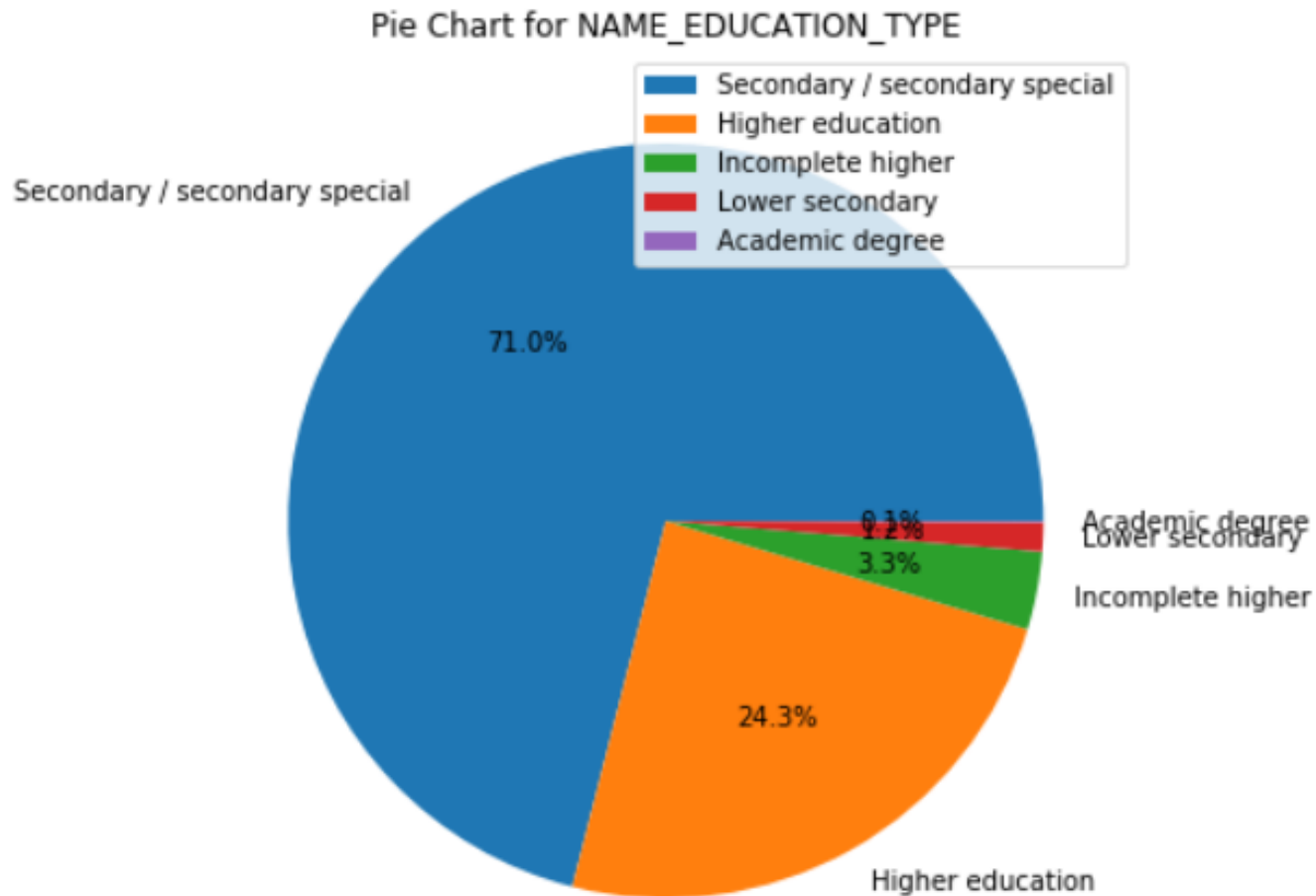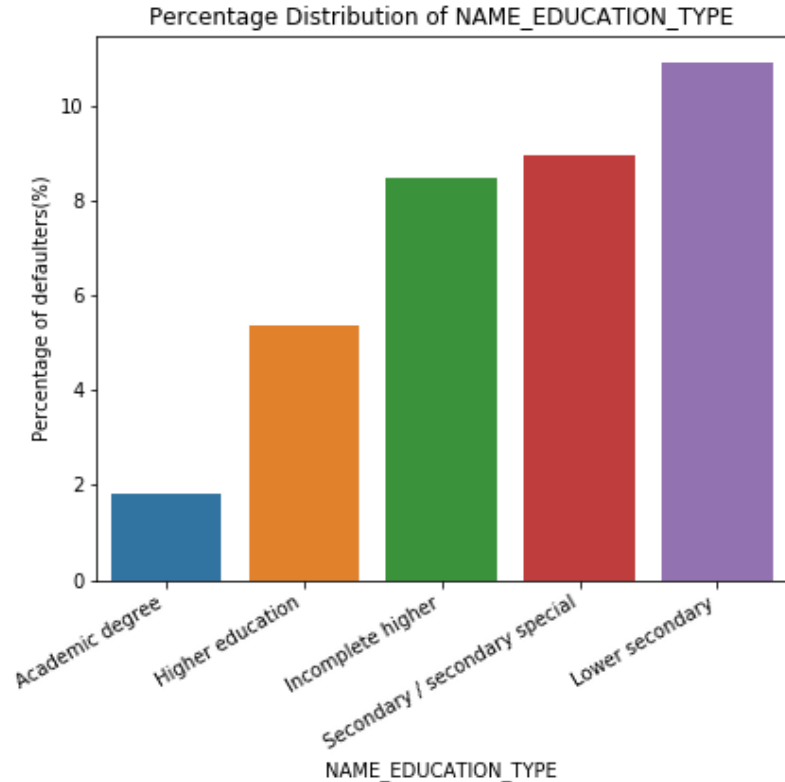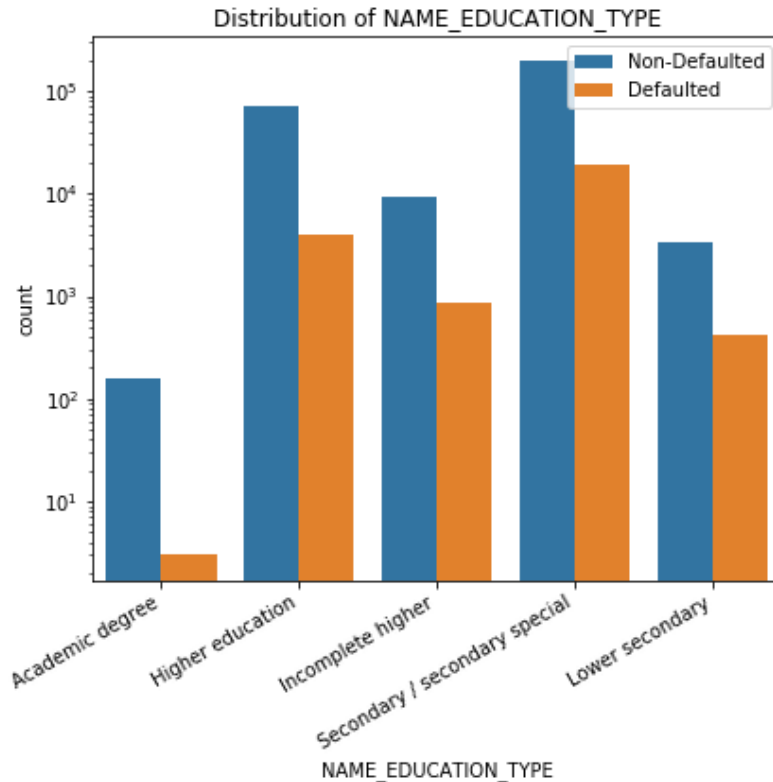
•Most applicants are with Secondary/secondary special education.
•Lower secondary category applicants are lower borrowers but top the Defulters. Applicants with academic degree have lowest Defaulters % compared to other categories.
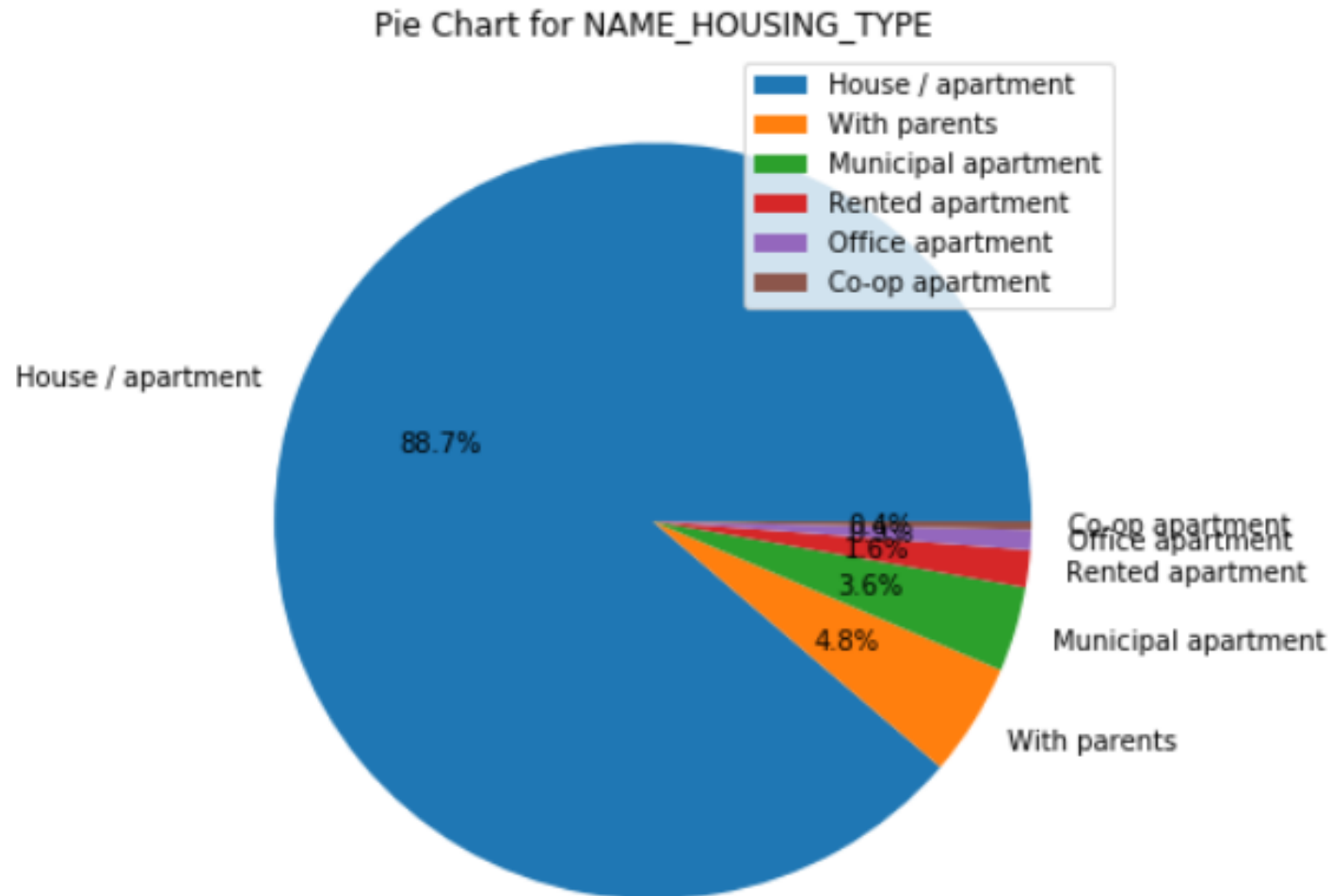
Distribution based on Family Status
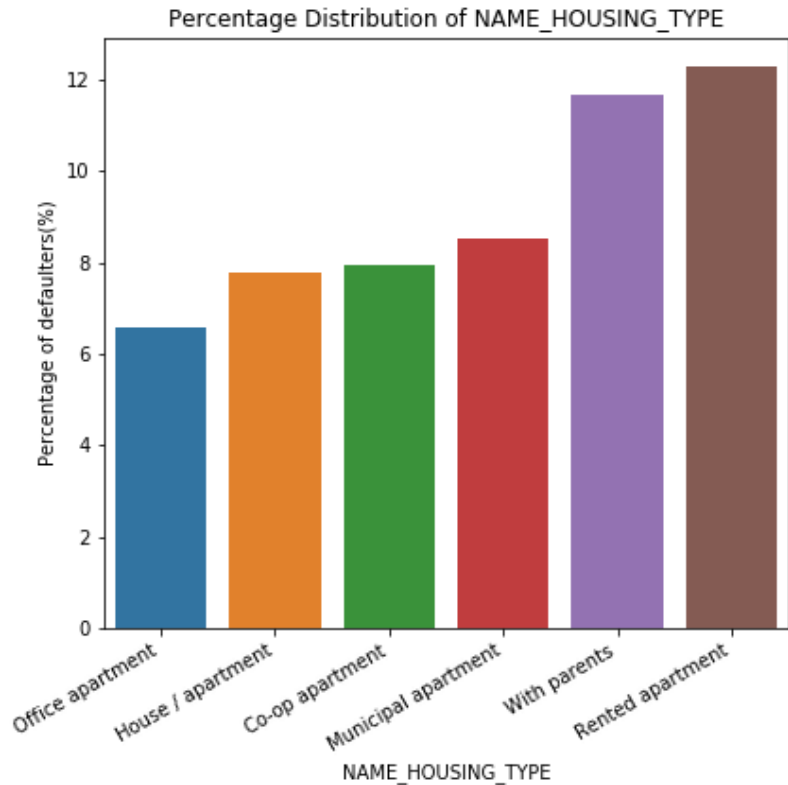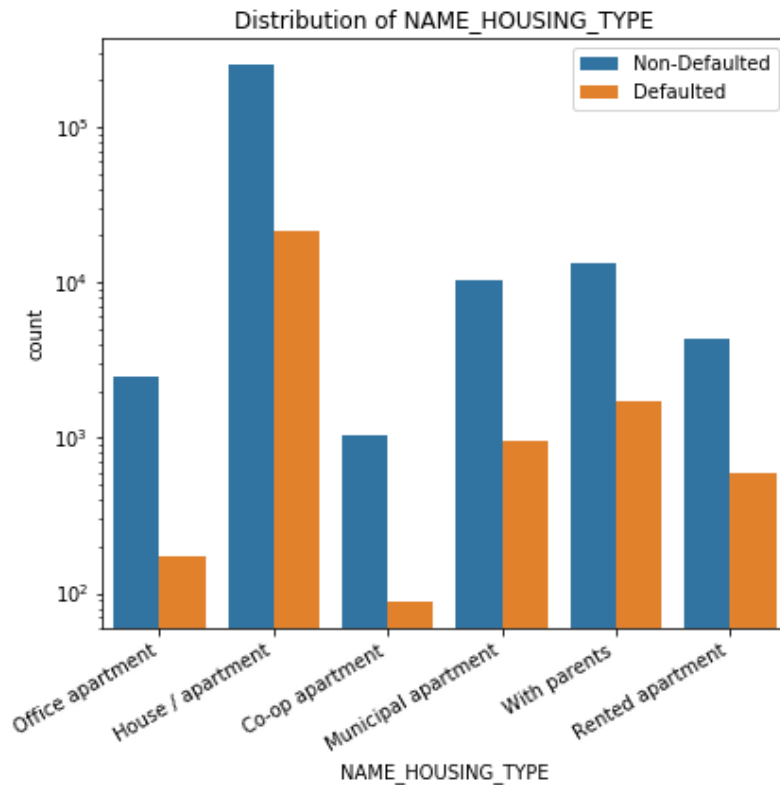

Pie Chart for NAME_FAMILY_STATUS

Legend:
- Married
- Single / not married
- Civil marriage
- Separated
- Widow
- Unknown

- Married: 63.9%
- Single / not married: 14.8%
- Civil marriage: 9.7%
- Separated: 6.4%
- Widow: 5.2%
- Unknown: 0.0%

Distribution of NAME_FAMILY_STATUS / Percentage Distribution of NAME_FAMILY_STATUS

•The family status shows same trend for non defaulters and defaulters.
•Civil Married people are the main defaulters compared to other sections.
•Widows are lowest in number in the defaulters category.
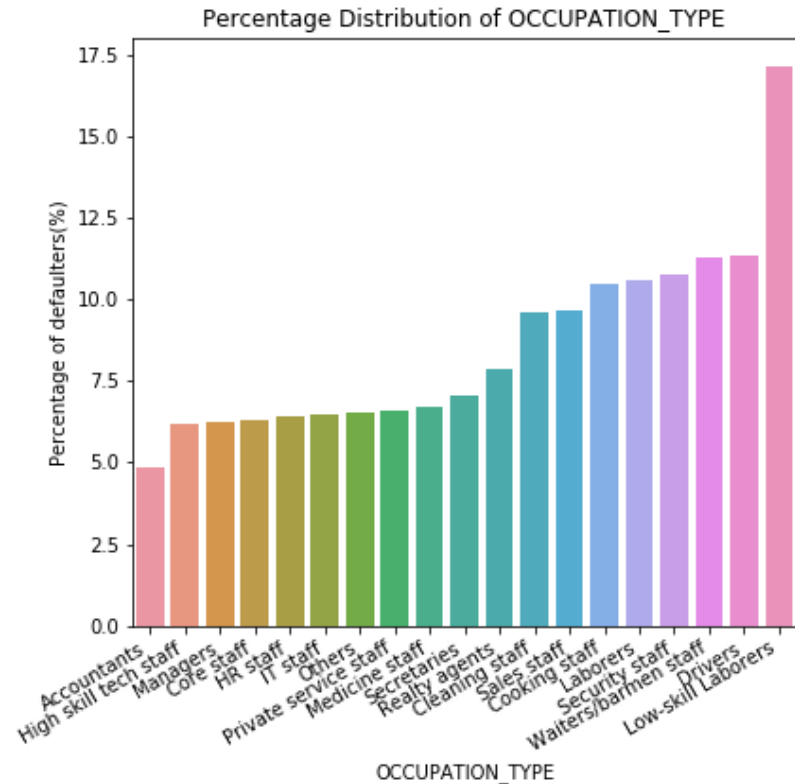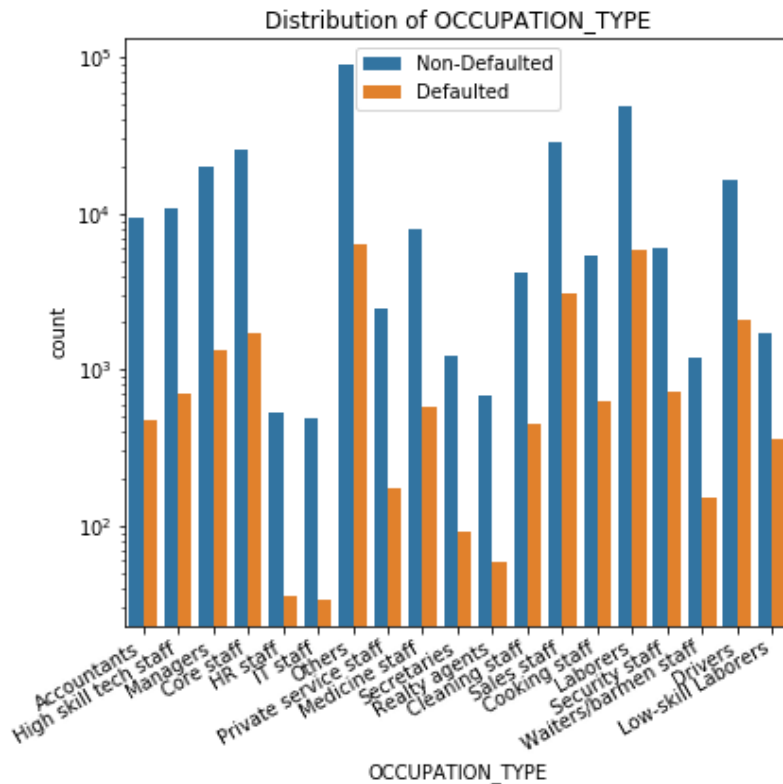
Distribution based on Housing Type

Pie Chart for NAME_HOUSING_TYPE

Legend:
- House / apartment
- With parents
- Municipal apartment
- Rented apartment
- Office apartment
- Co-op apartment

House / apartment — 88.7%
With parents — 4.8%
Municipal apartment — 3.6%
Rented apartment — 1.6%
Office apartment
Co-op apartment — 0.4%

Distribution of NAME_HOUSING_TYPE

Percentage Distribution of NAME_HOUSING_TYPE

•Majority of the applicants possess a house/apartment.
•Applicants in Rented apartment or Staying with parents top the Defaulters list

Distribution based on Occupation Type

Bar Plot for OCCUPATION_TYPE

Distribution of OCCUPATION_TYPE / Percentage Distribution of OCCUPATION_TYPE
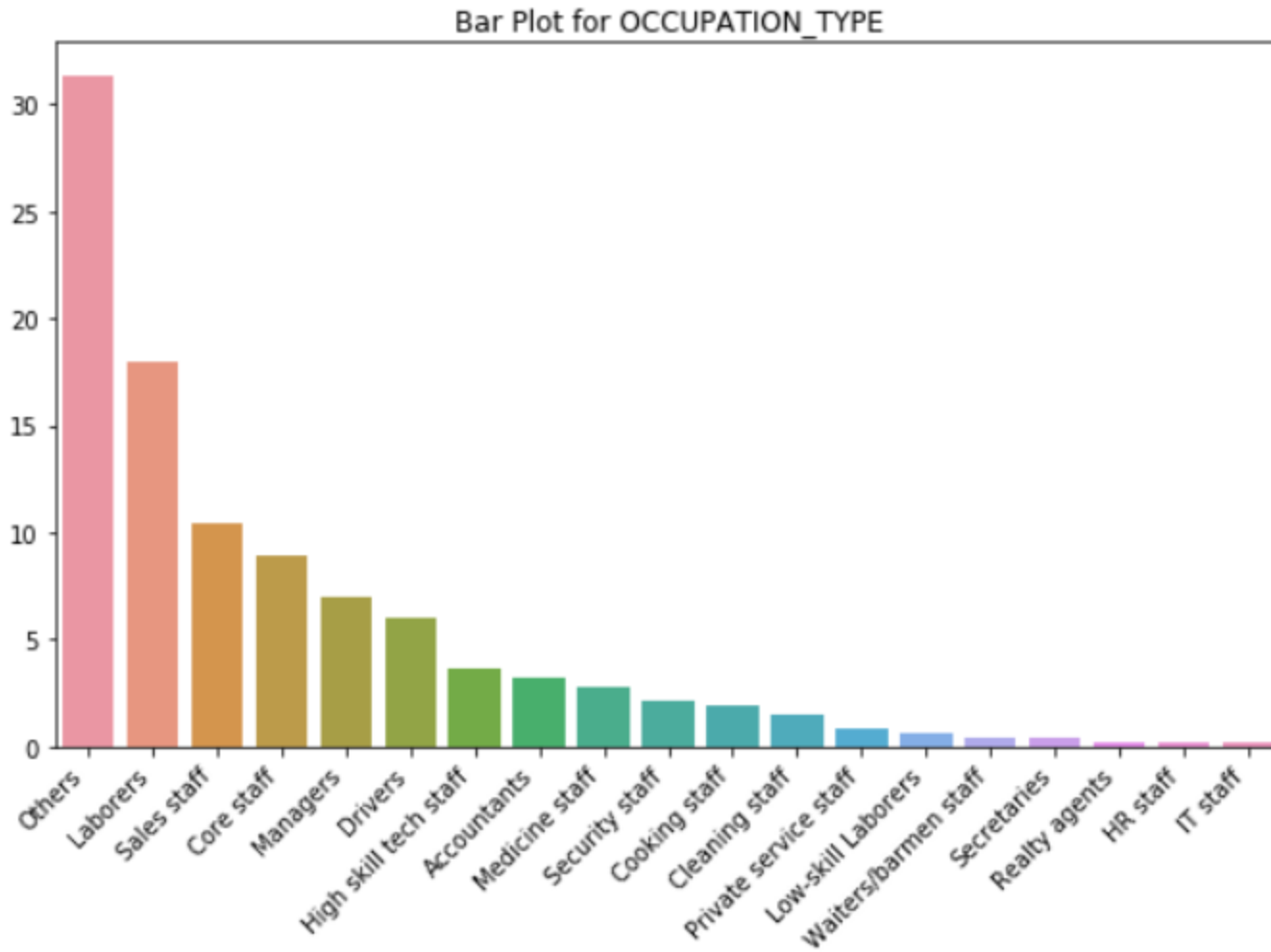
•Most of the loans are taken by Laborers followed by Sales Staff.
•Low-skill/Laborers top the Defaulters category with 17% followed by rivers and Waiters/barmen staff

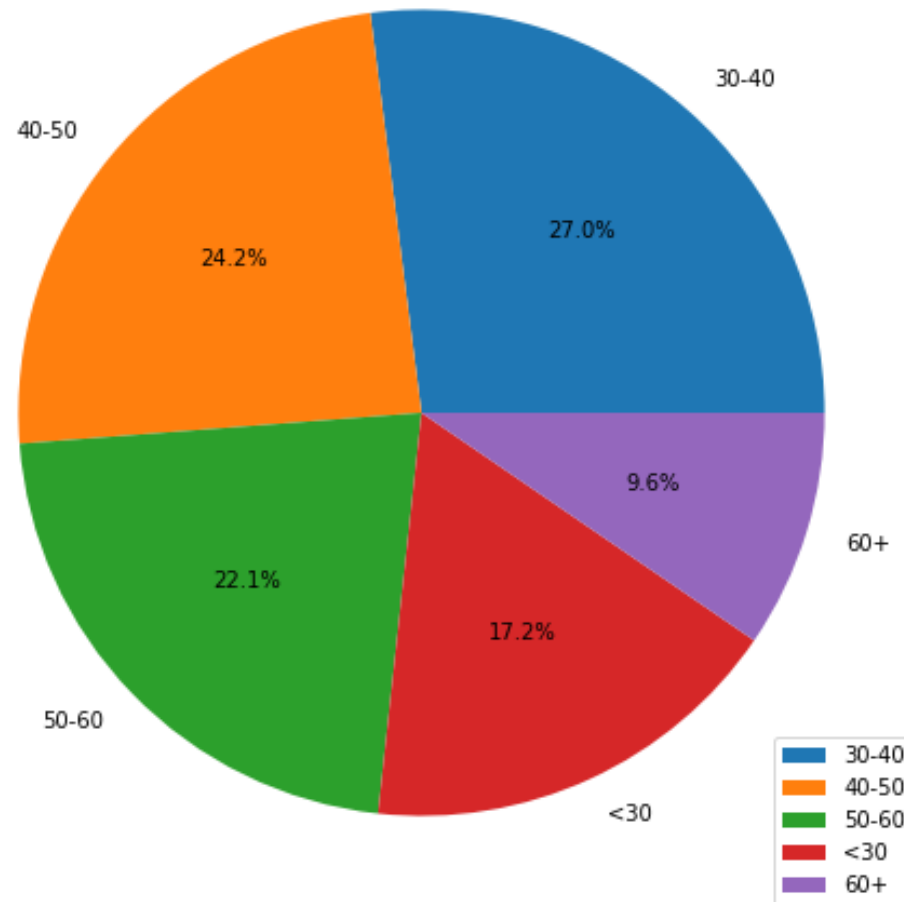# Distribution based on Organization Type
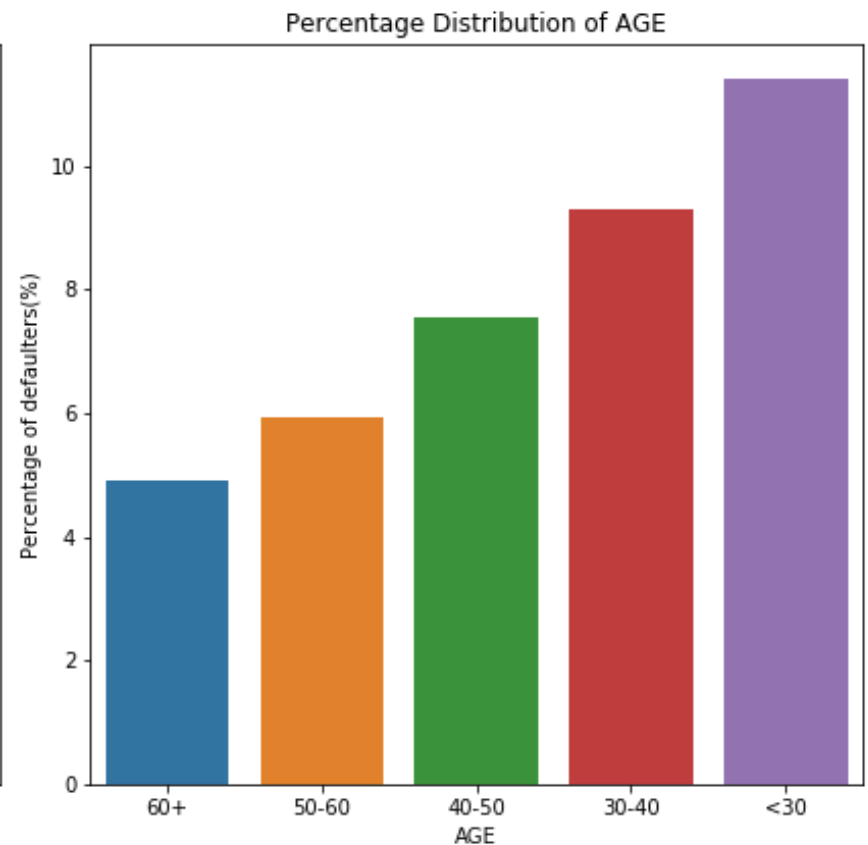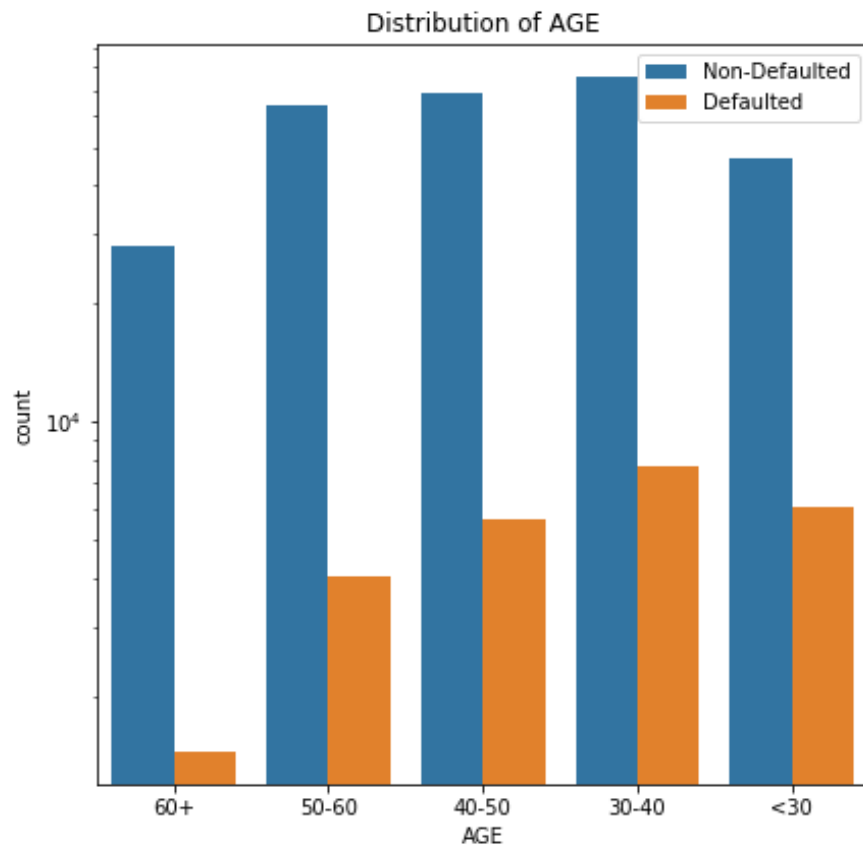

Bar Plot for ORGANIZATION_TYPE

Distribution of ORGANIZATION_TYPE

Percentage Distribution of ORGANIZATION_TYPE
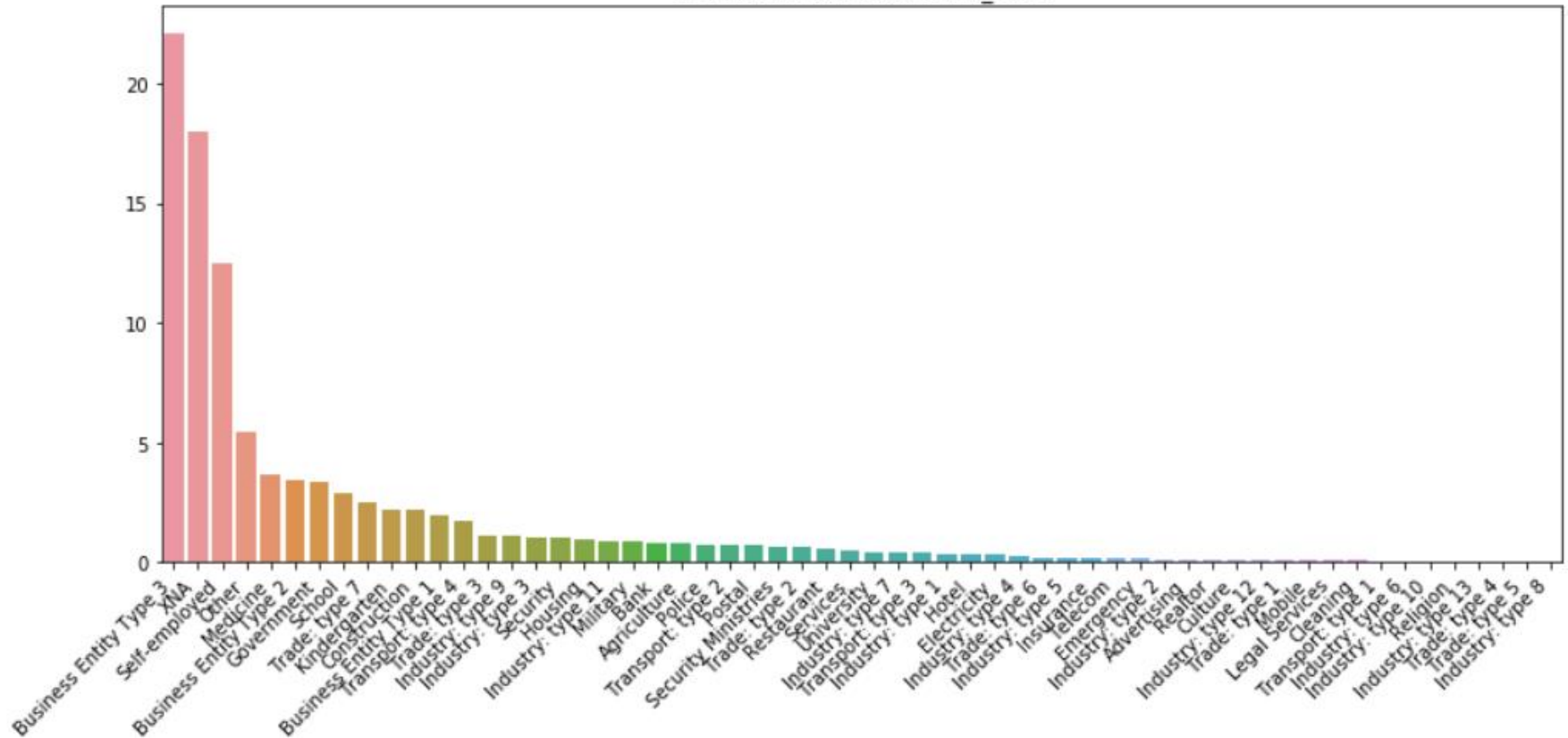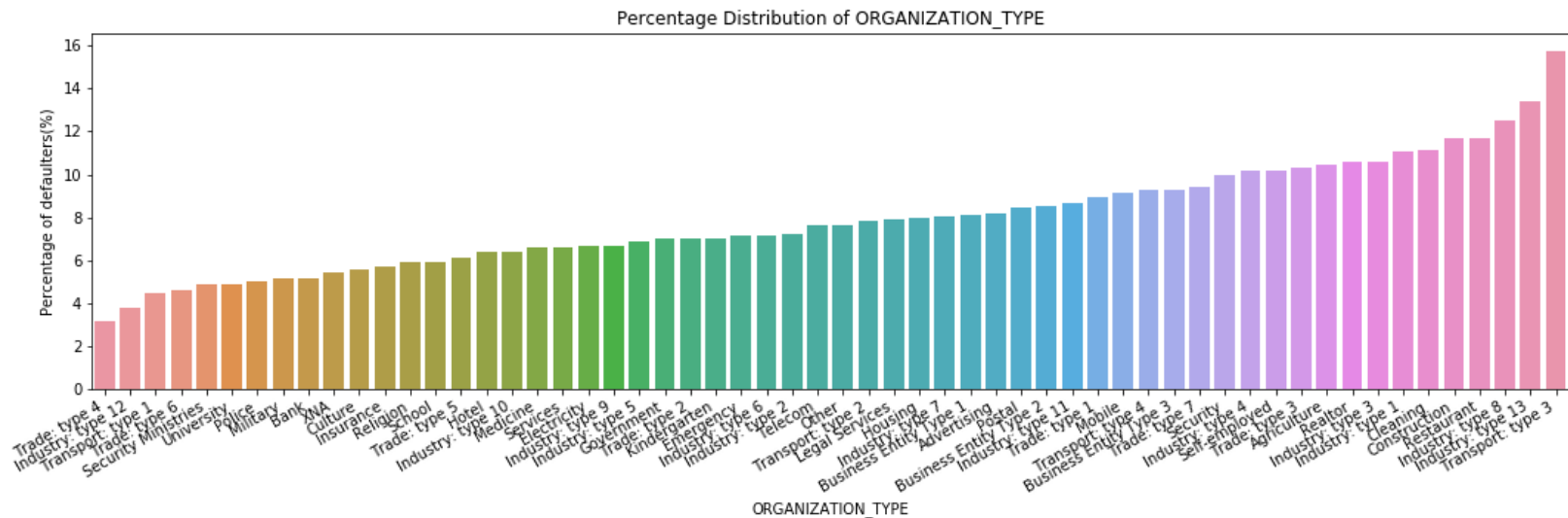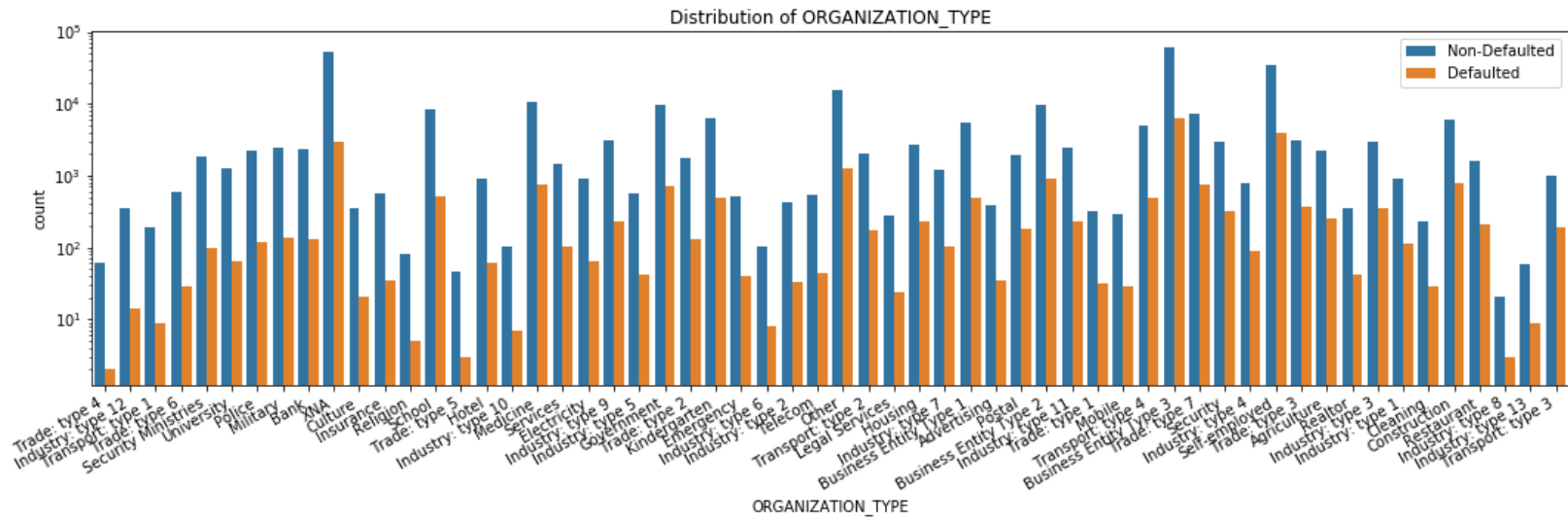
•Among the Organisation type the major defaulters are - Transport: type 3 (15%), Industry: type 13(13%)
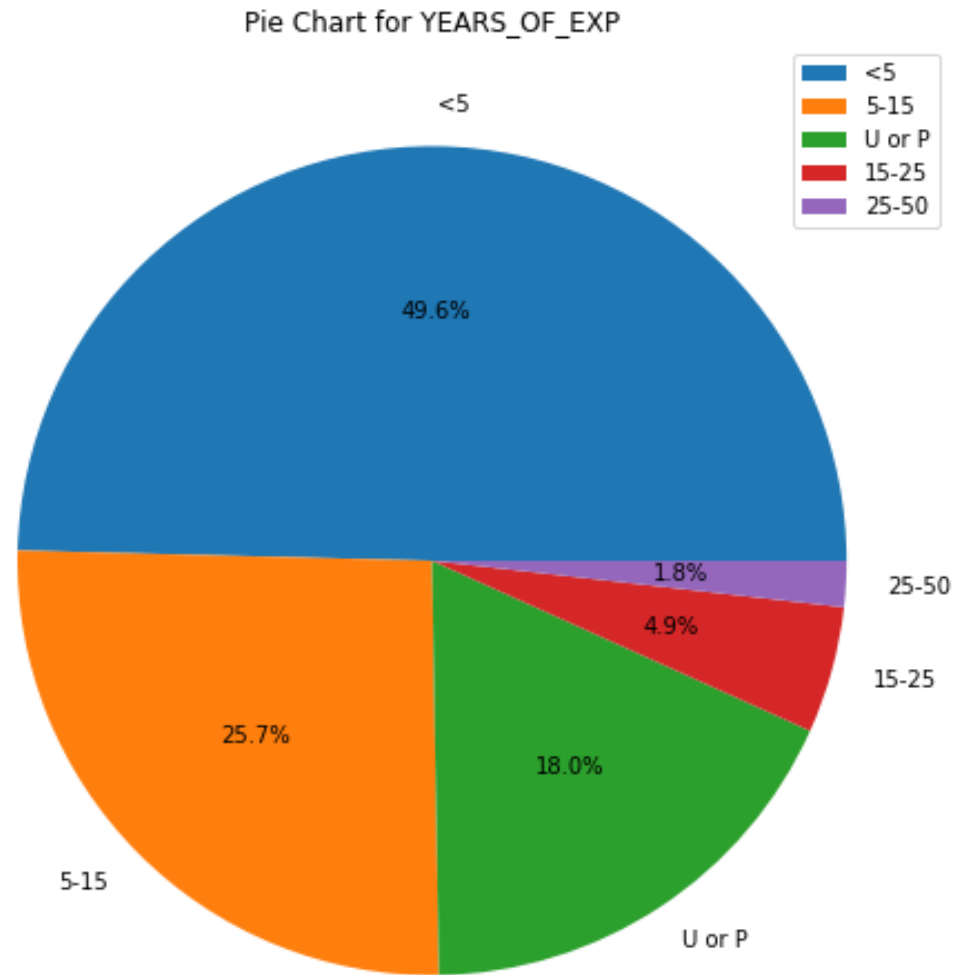
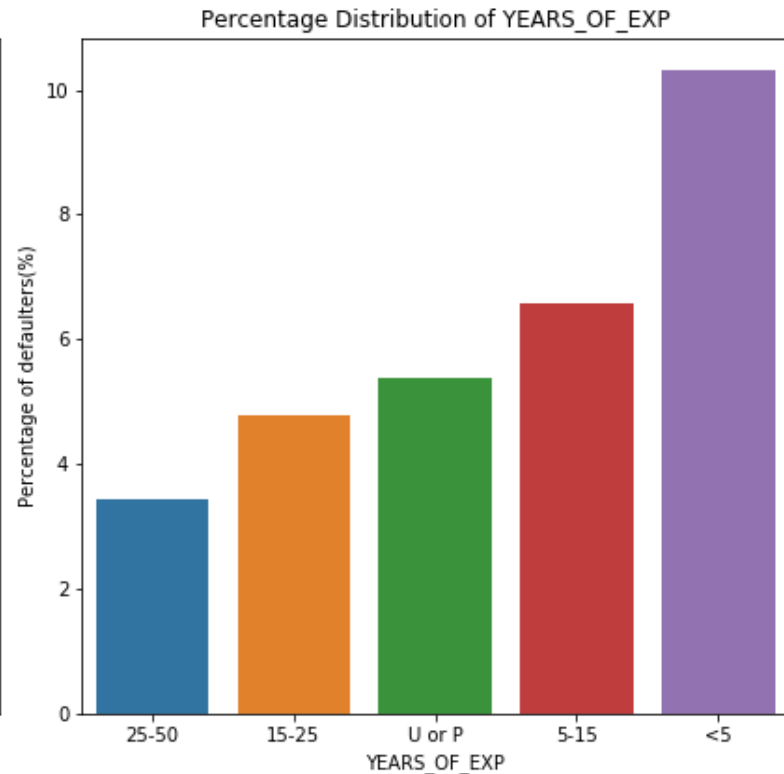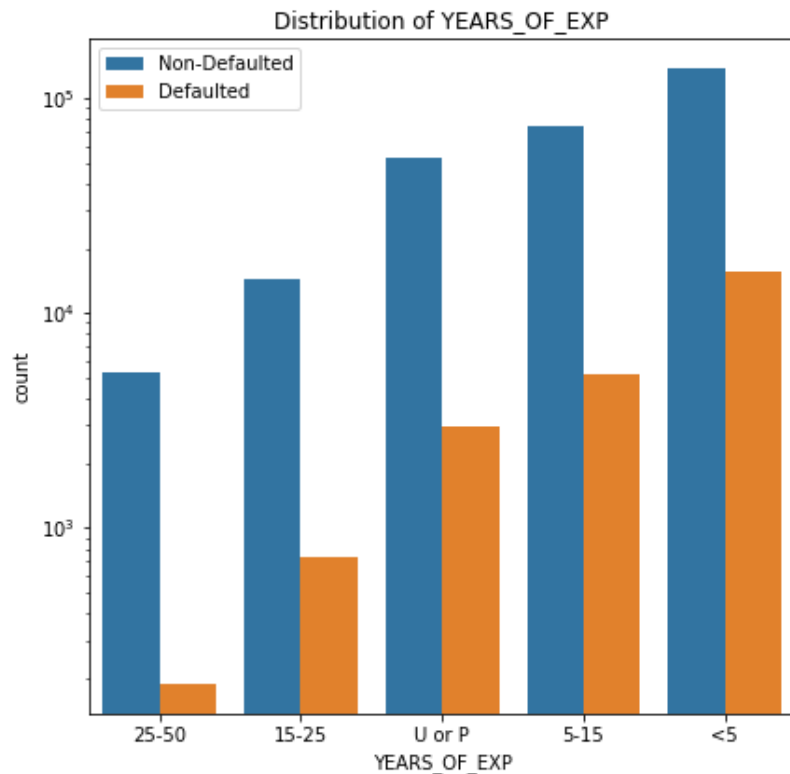# Distribution based on Age

Pie Chart for AGE

- Maximum number of defaulters are in the age group <30
- Minimum number of defaulters are in the group 60+

Distribution of Years of Experience

Pie Chart for YEARS_OF_EXP

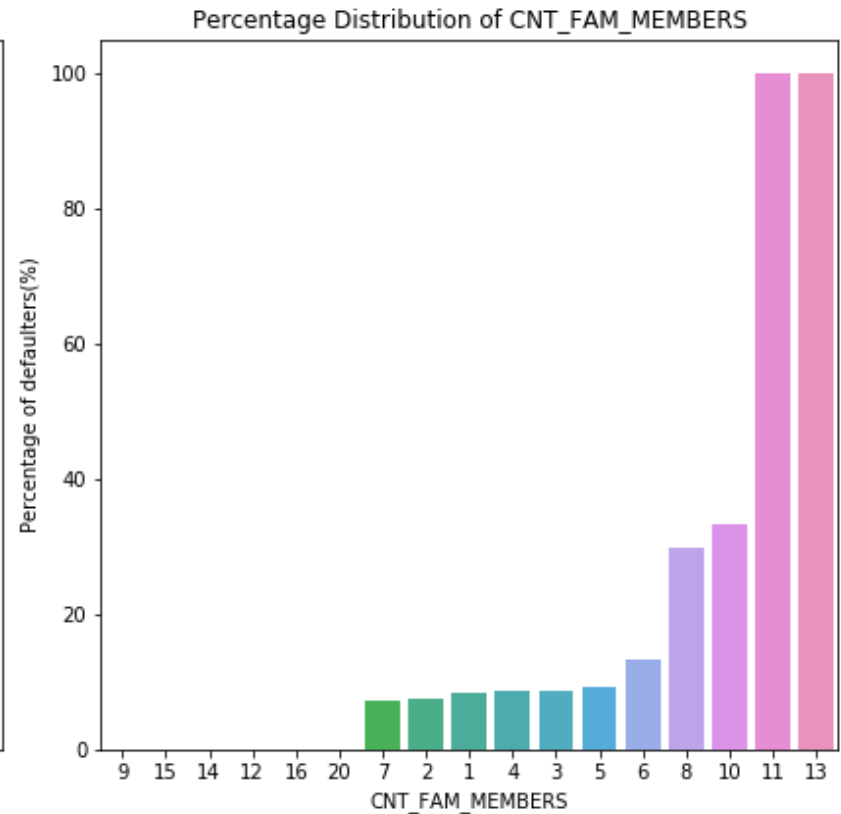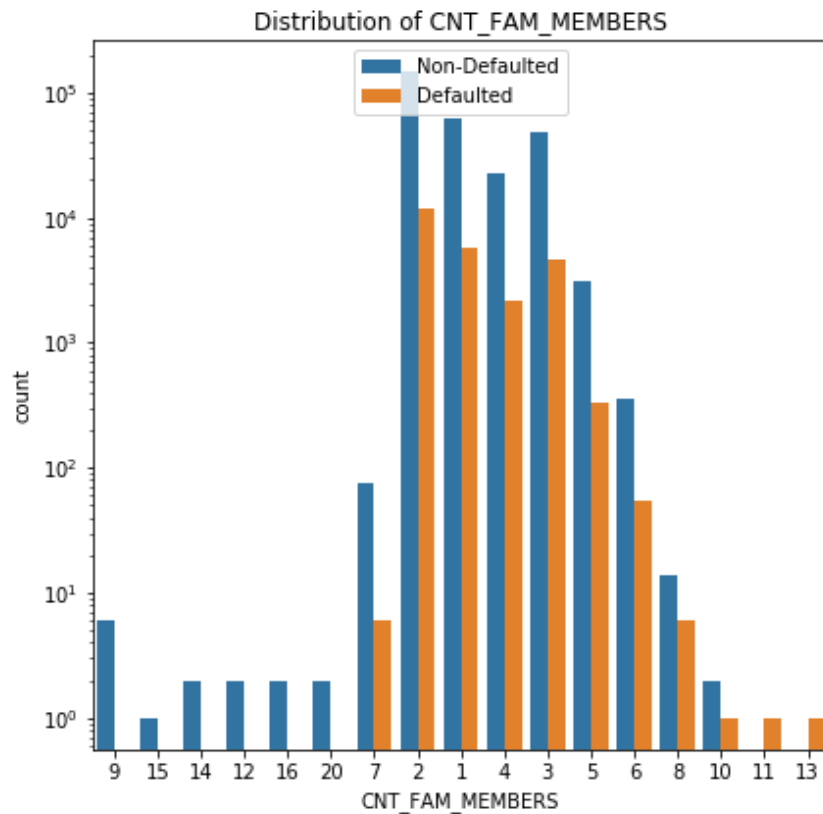Distribution of YEARS_OF_EXP / Percentage Distribution of YEARS_OF_EXP

•The major defaulters are having an experience of less than 5 years in the present employment.
•The number of defaulters decrease with increase in years of employment.

Distribution of Family Members

Distribution of CNT_FAM_MEMBERS — Percentage Distribution of CNT_FAM_MEMBERS
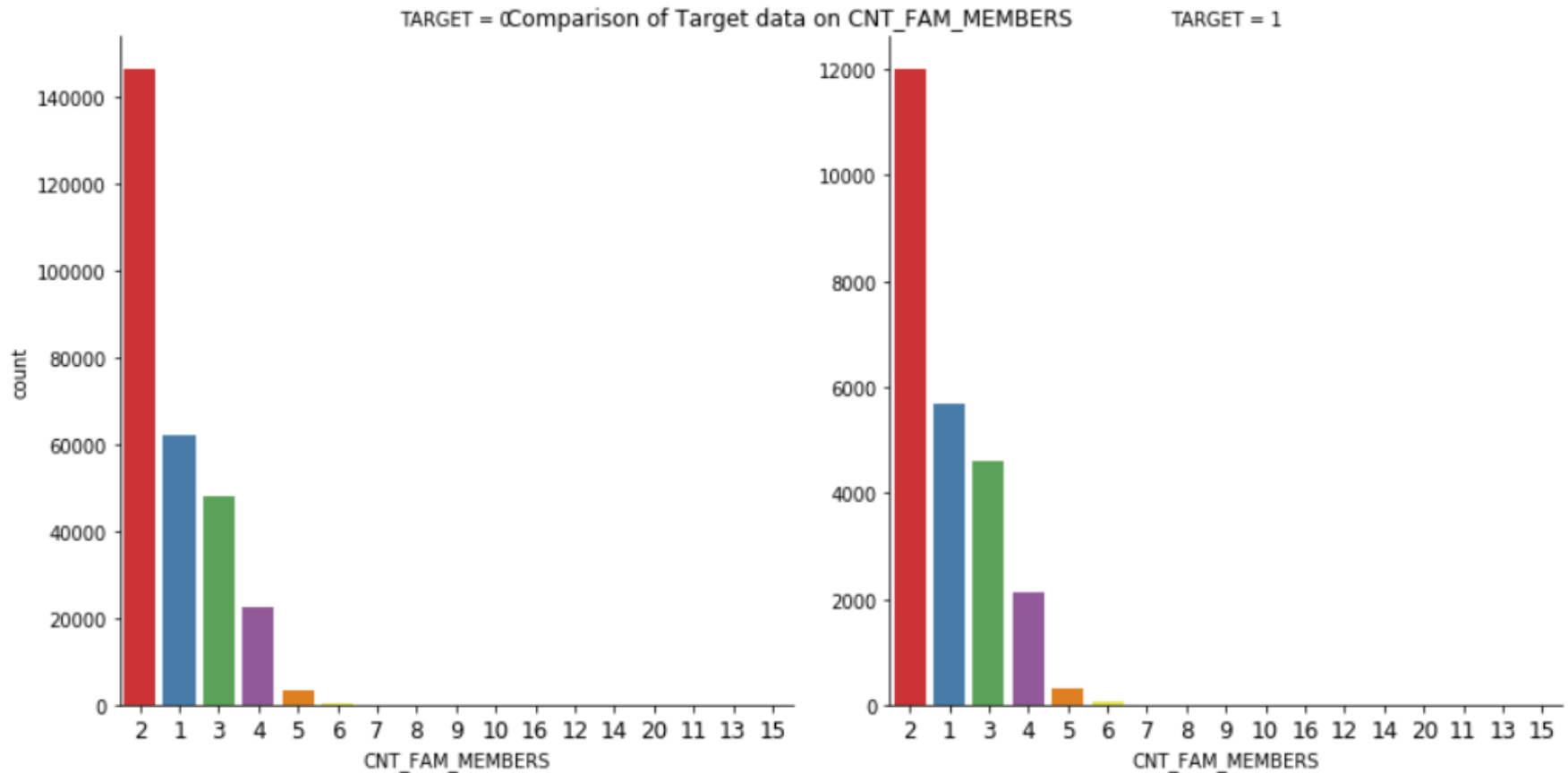
- Families with 2 members have highest percentage of loans.
- Applicants with family size > 10 have highest percentage of defaulted loans
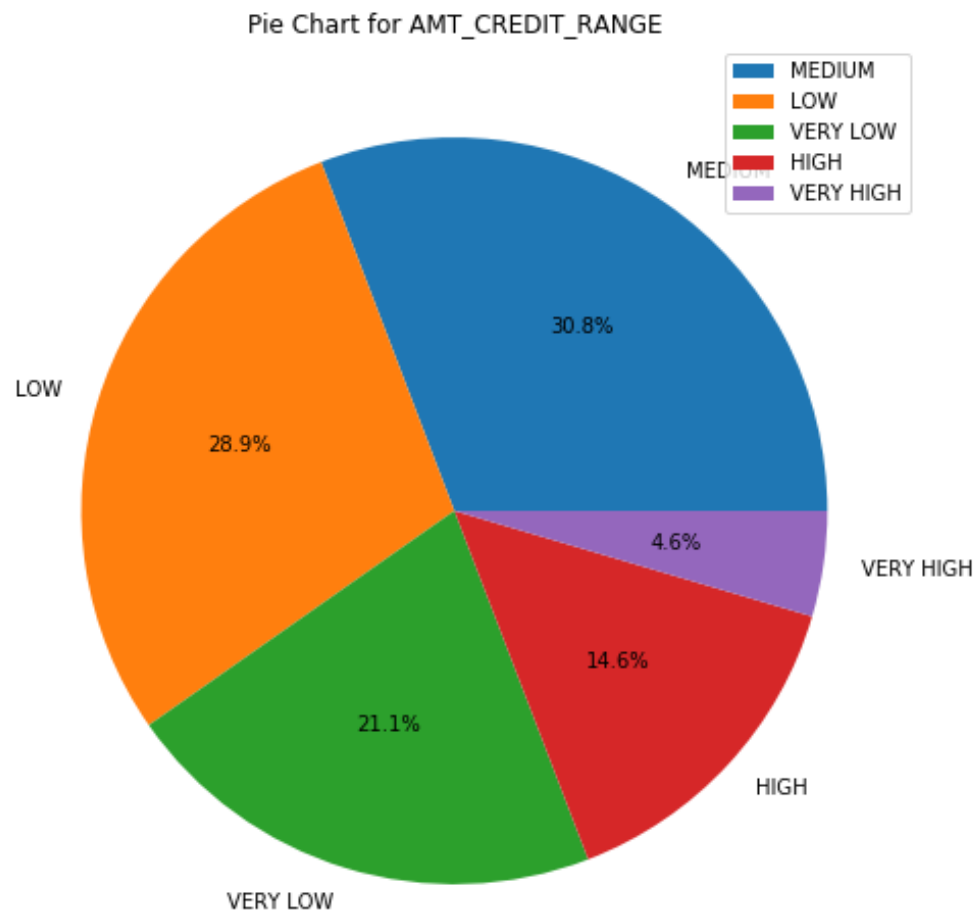
Distribution of Income

Pie Chart for INCOME_RANGE
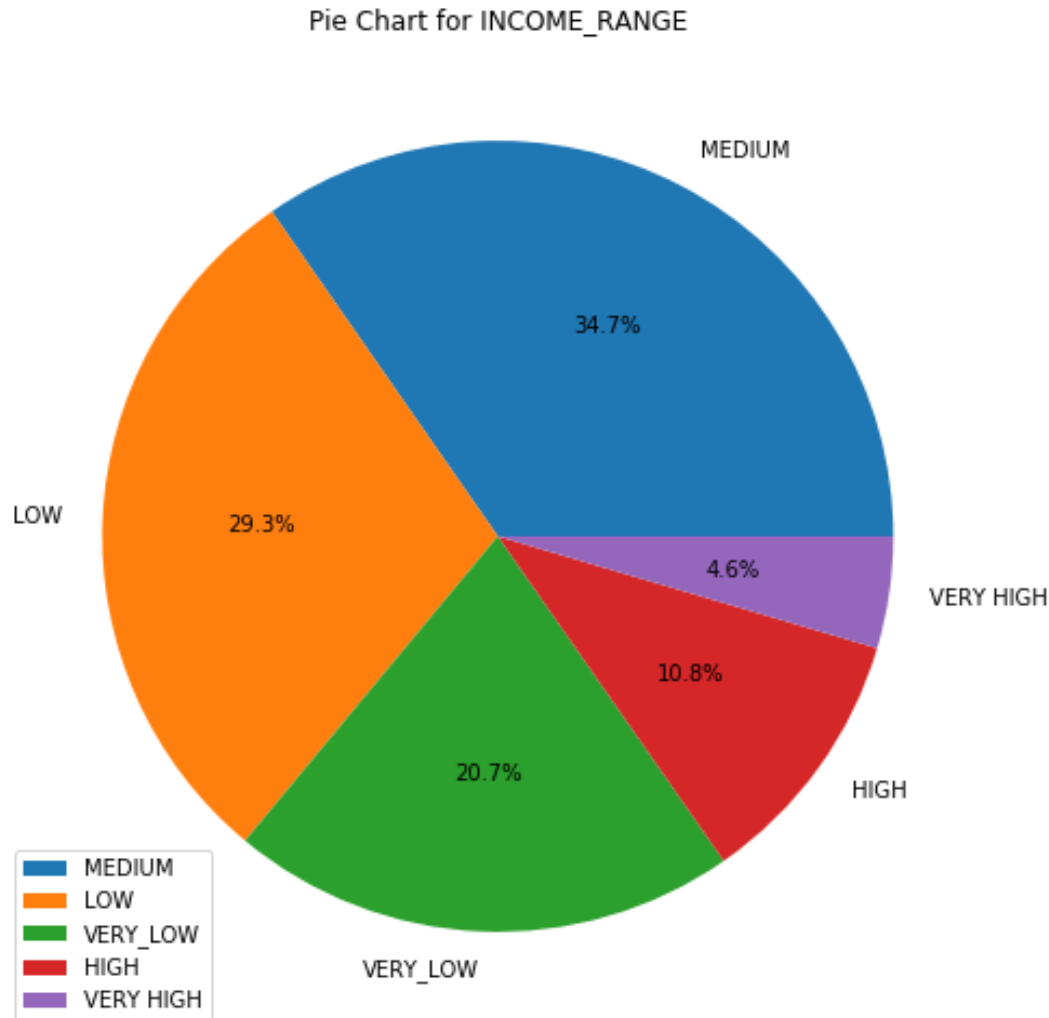
Low income group top the Defaulters list and Very High income group are the lowest defaulters.
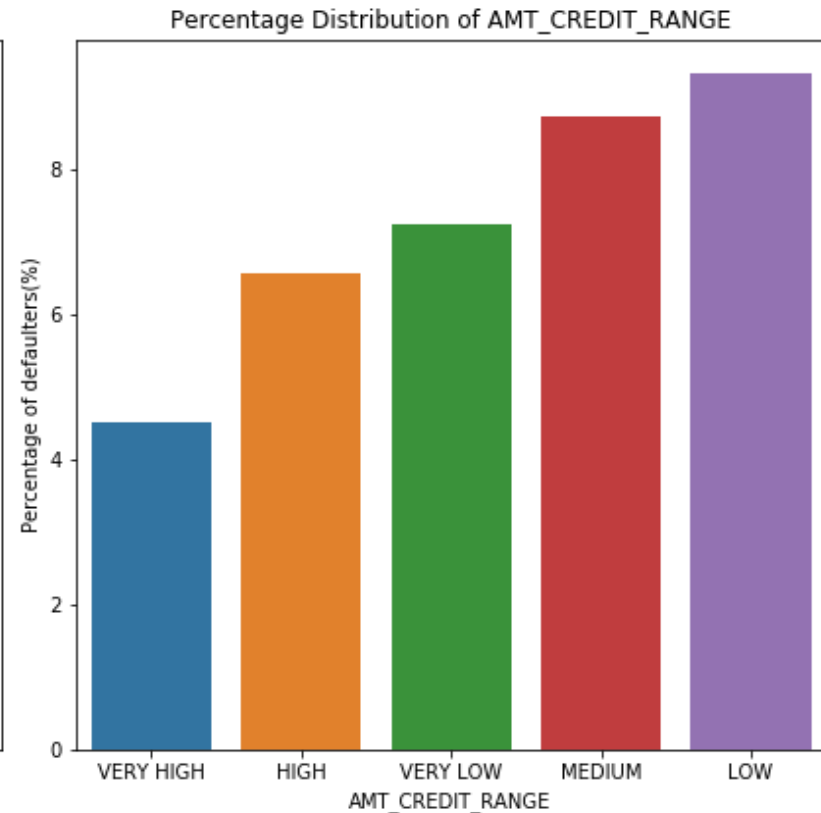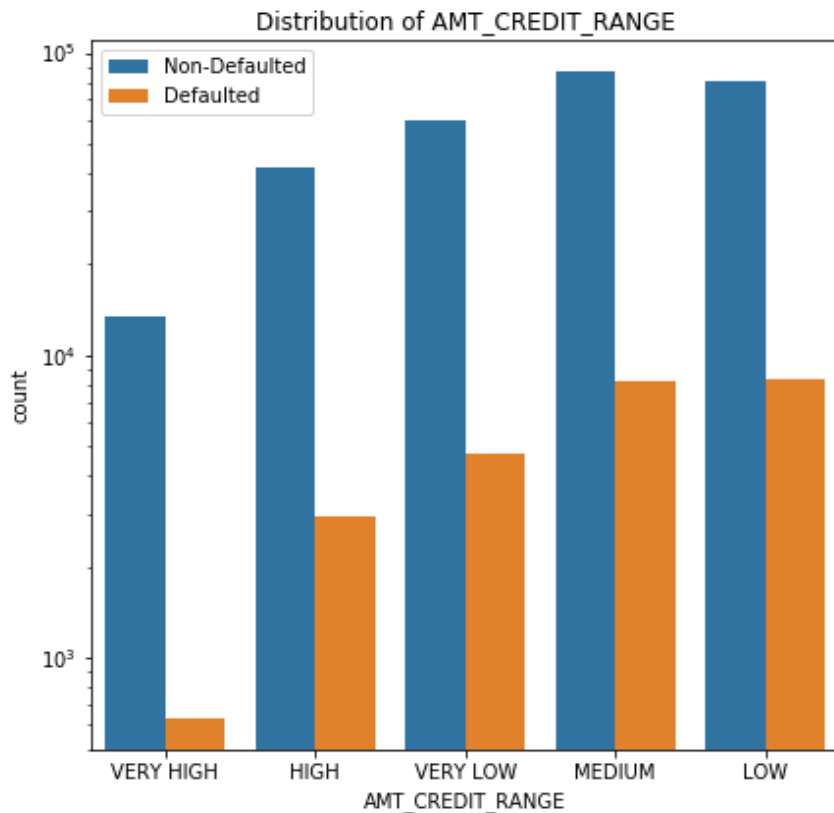•The Medium income group has the highest percentage of loans.
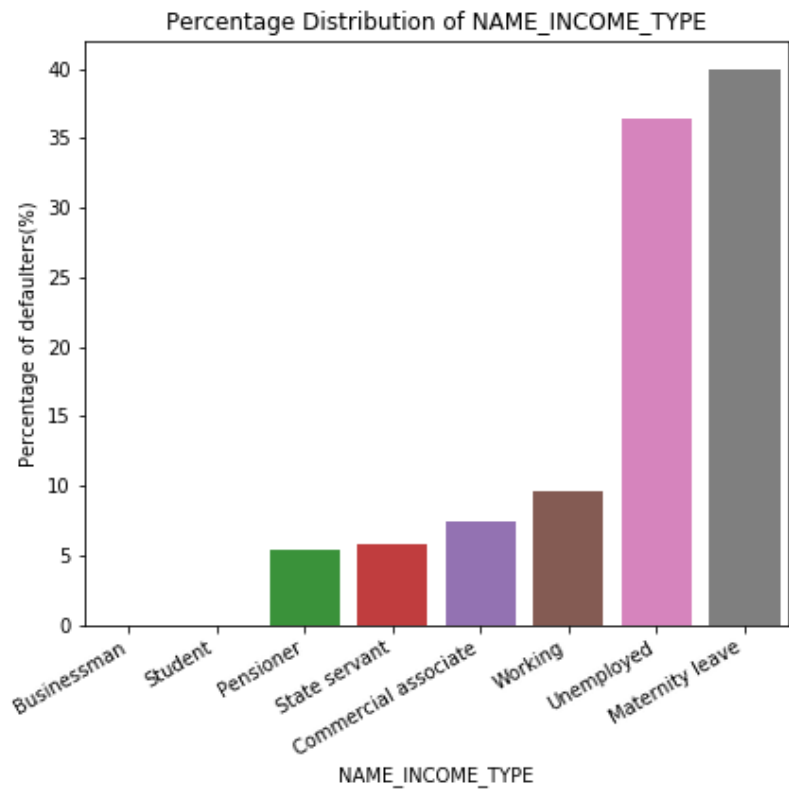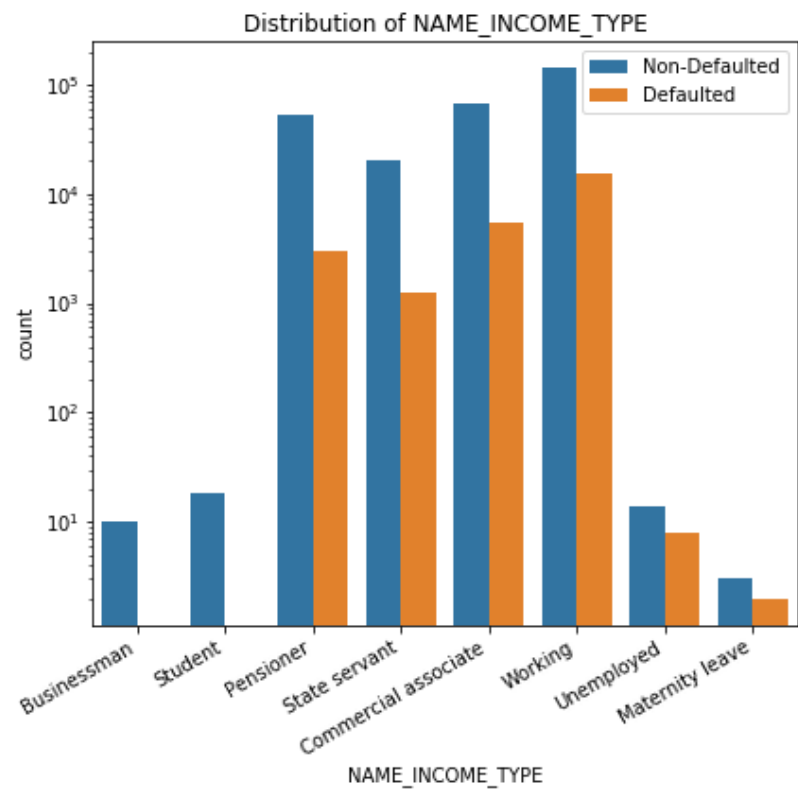
- Defaulters are less among the very high income group compared to non defaulters.
- The medium income people are the major defaulters.

Distribution of Credit

Pie Chart for AMT_CREDIT_RANGE

Legend:
- MEDIUM
- LOW
- VERY LOW
- HIGH
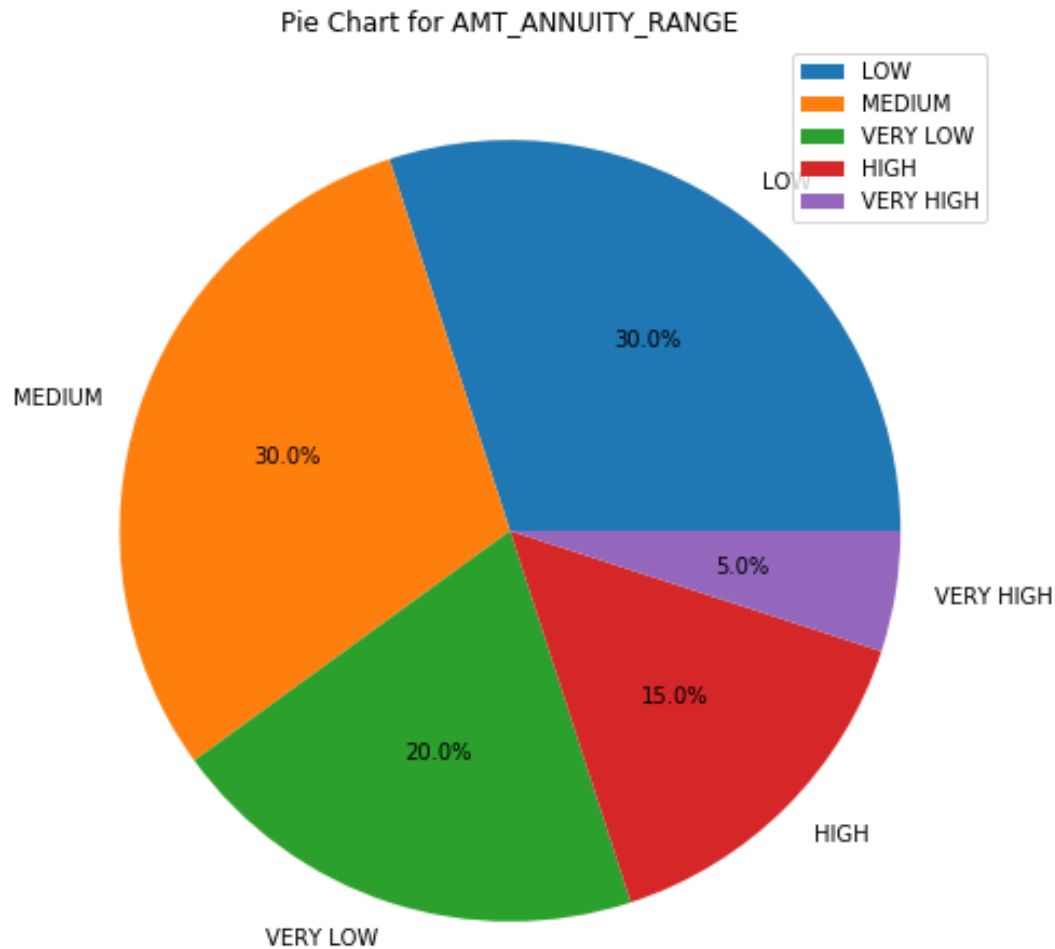- VERY HIGH

MEDIUM 30.8%
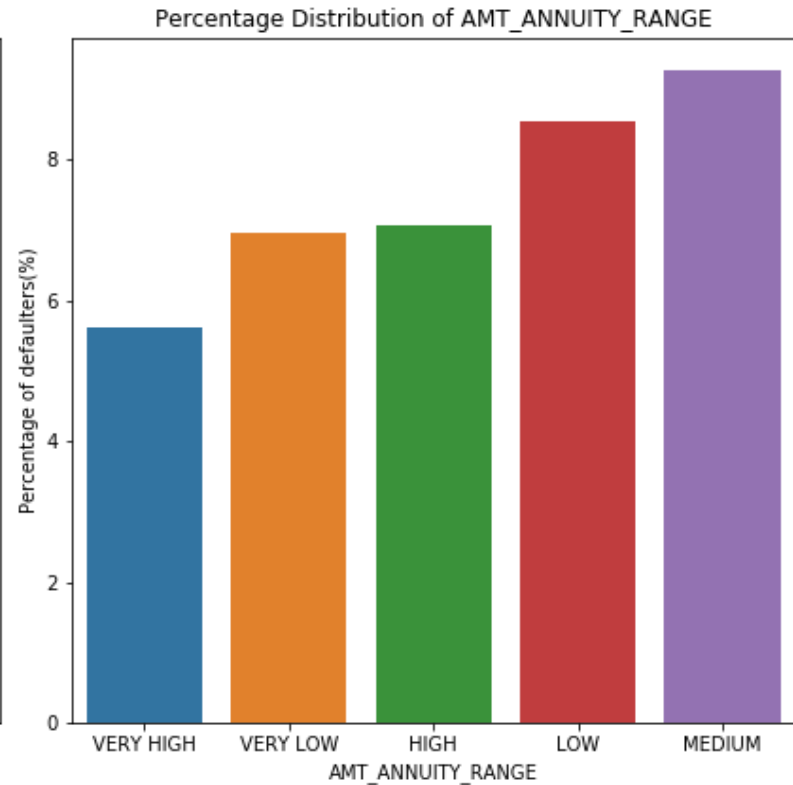LOW 28.9%
VERY LOW 21.1%
HIGH 14.6%
VERY HIGH 4.6%

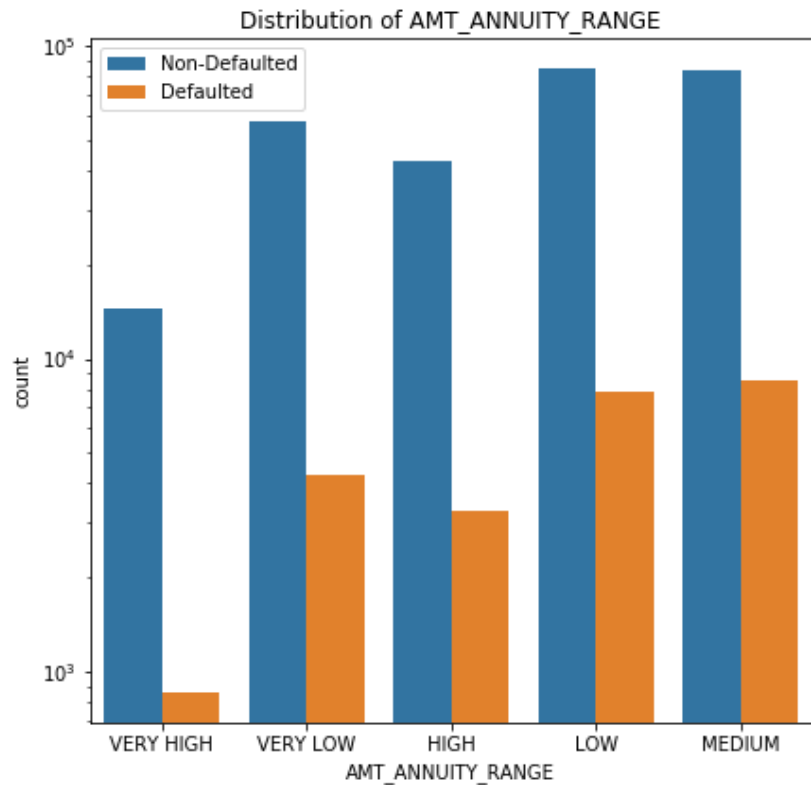- Defaulters are less in very high credit group compared to other defaulters.
- Defaulters are more among low and medium credit group.

Distribution of Annuity
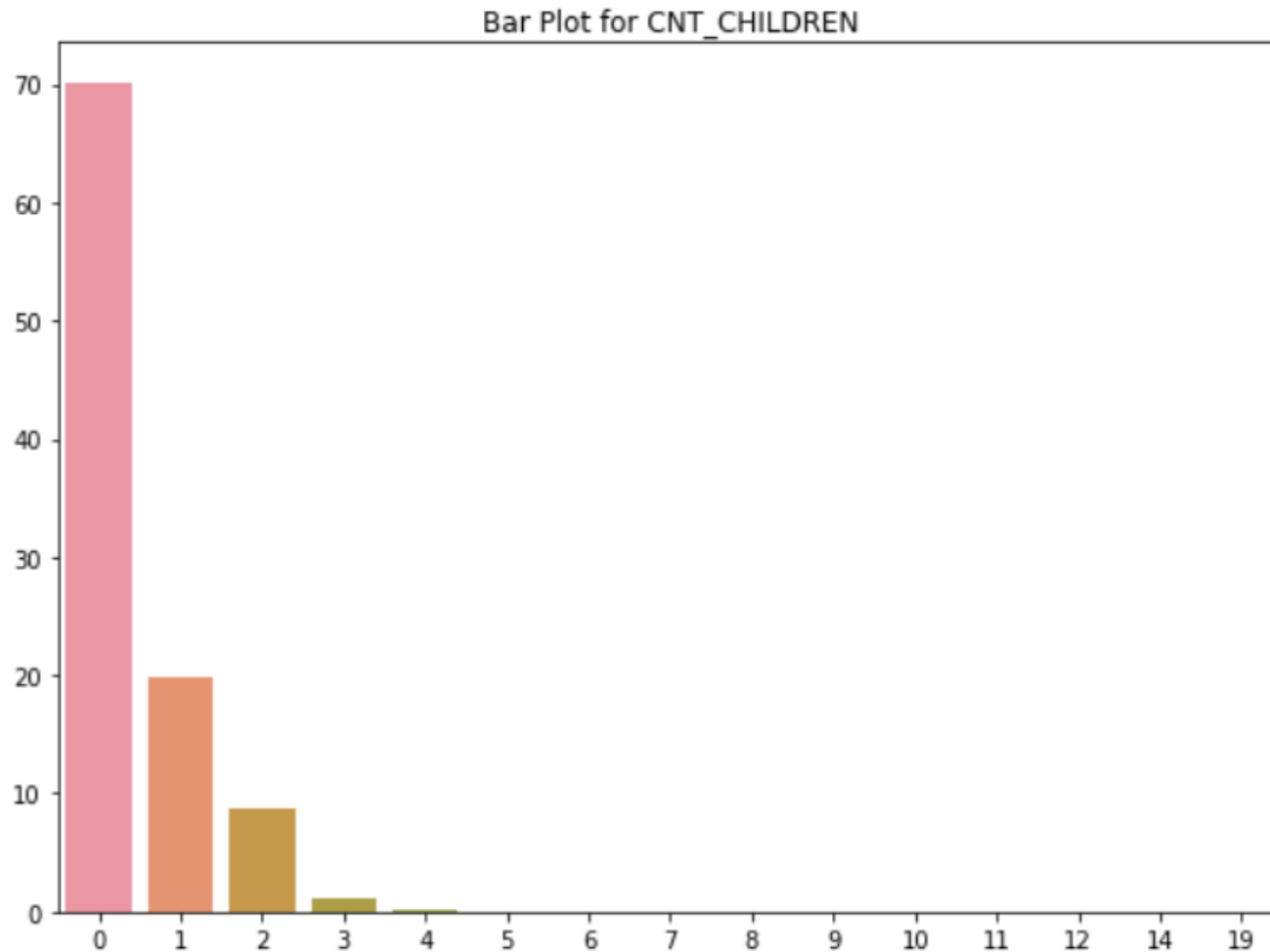

Pie Chart for AMT_ANNUITY_RANGE

- Defaulters are less among the very high annuity group compared to other defaulters.
- Defaulters are more amount Low and Medium Amt Annuity group.

Distribution of Children

Bar Plot for CNT_CHILDREN

Distribution of CNT_CHILDREN — Percentage Distribution of CNT_CHILDREN
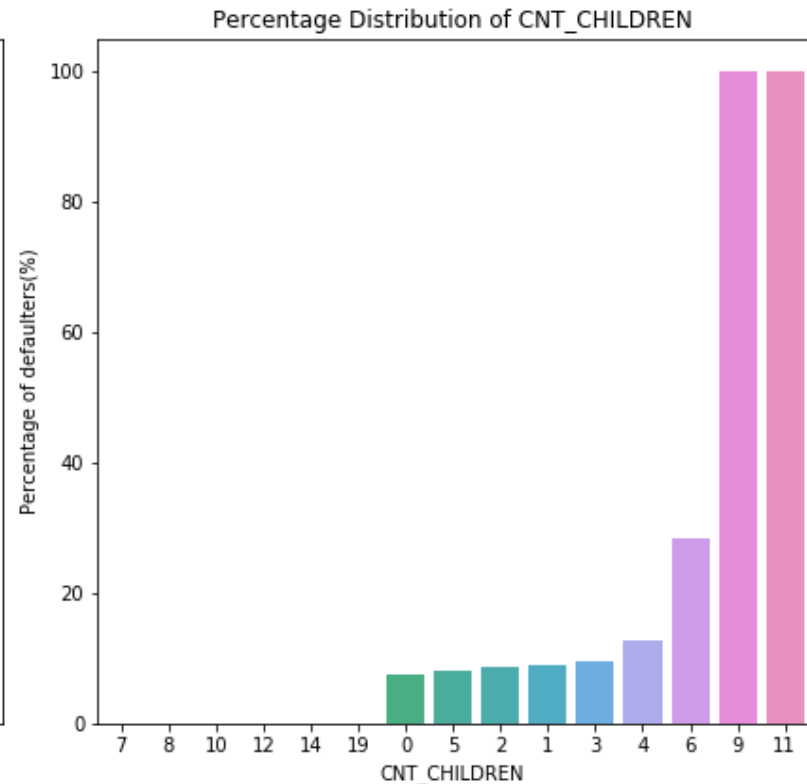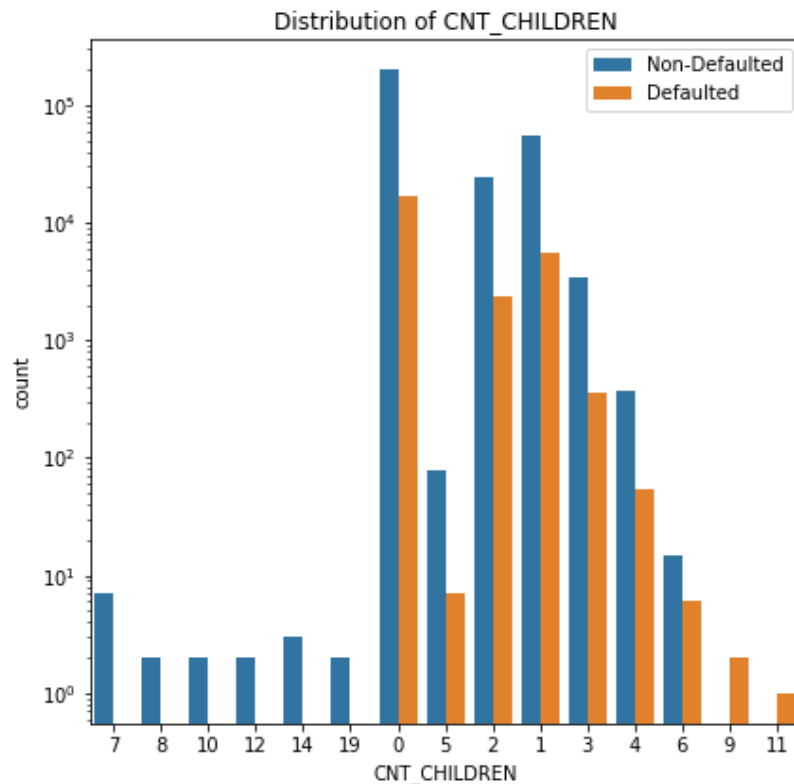
•Most of the applicants have no children. There are negligible loans for applicants with more than 3 children
•Applicants with number of children > 9, have defaulted loans

BIVARIATE ANALYSIS

Credit Amount Vs Education For Target 0 and 1

Credit Amount vs Education of Non-Defaulted Loans

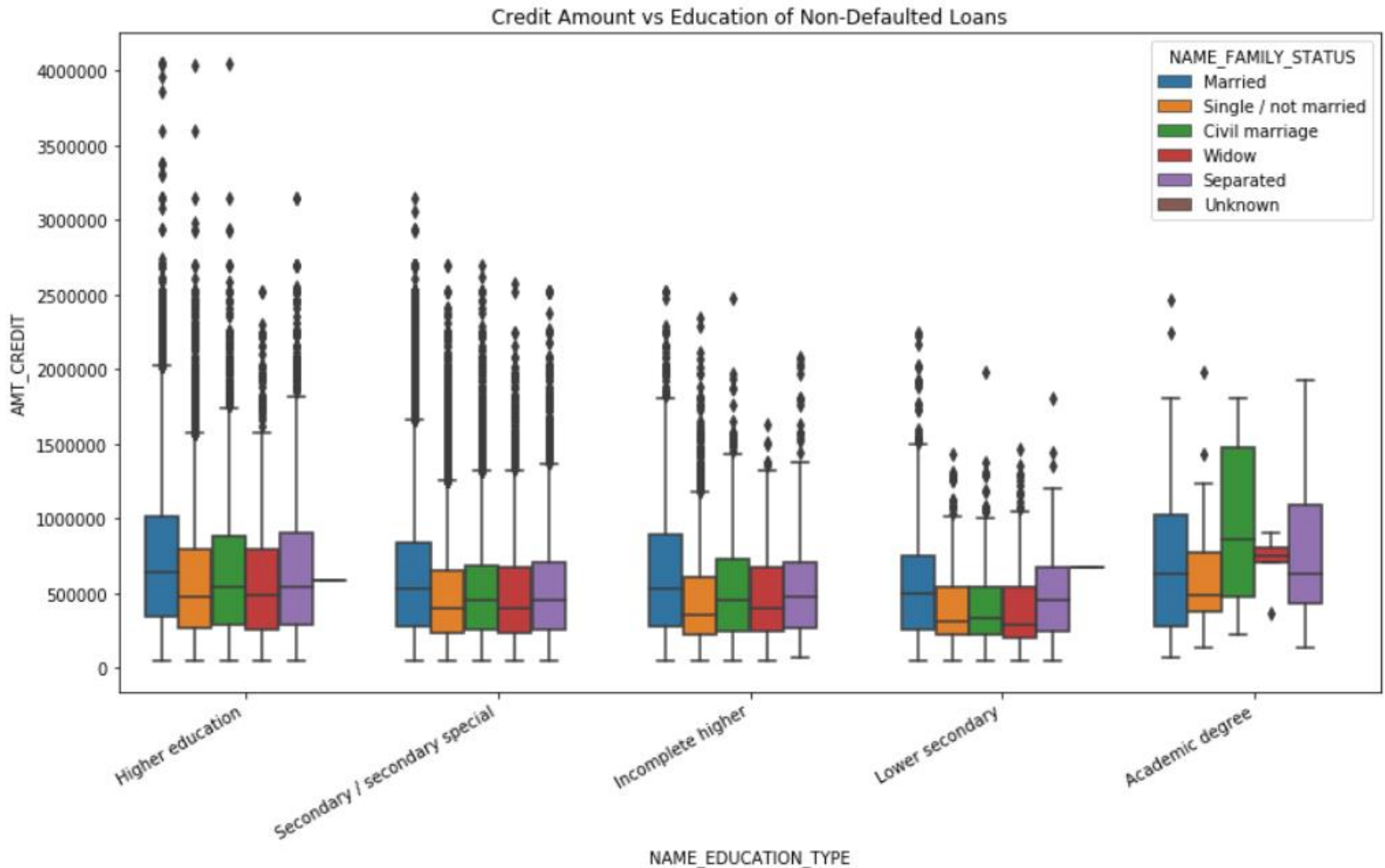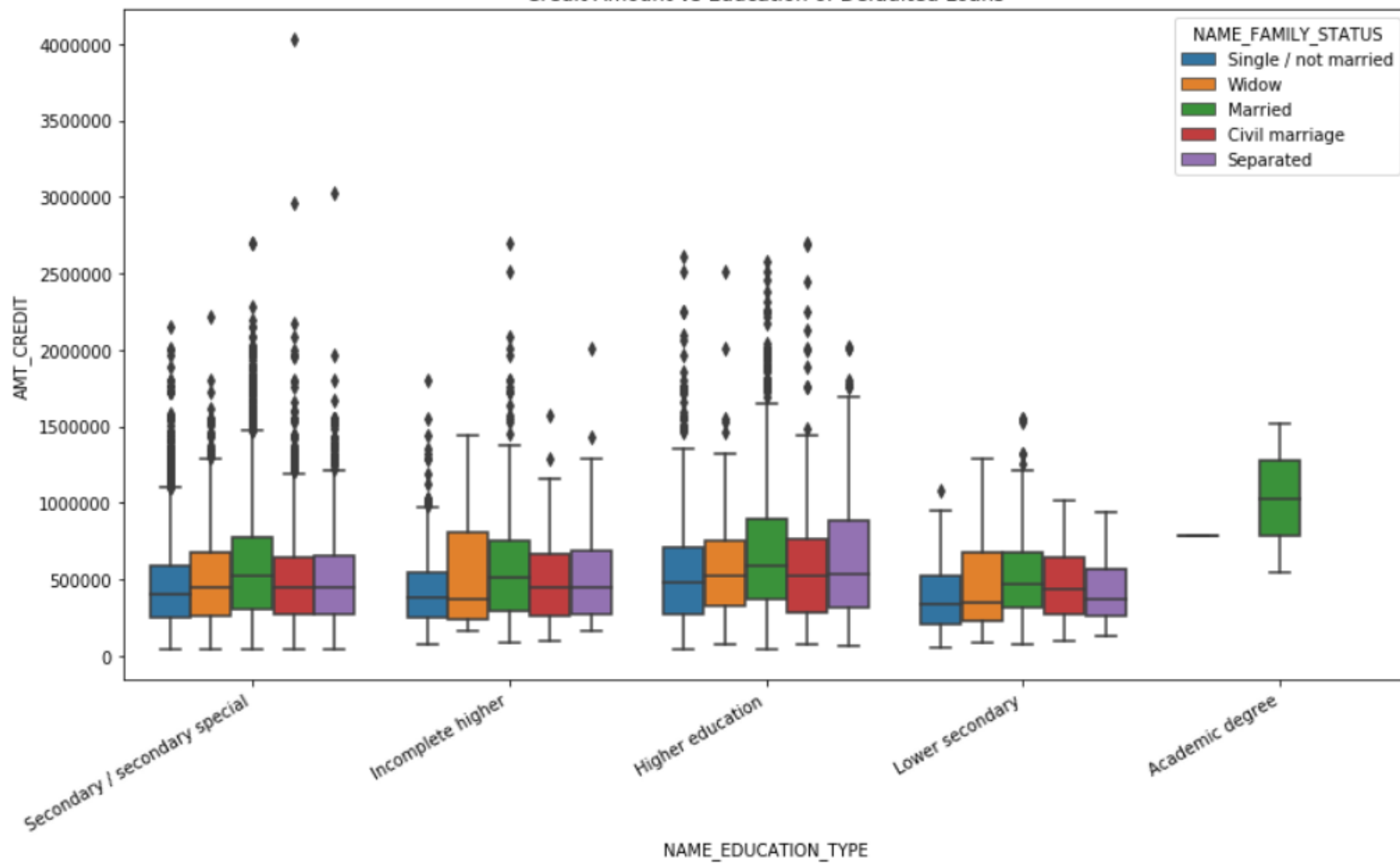Credit Amount vs Education of Defaulted Loans

From the above charts, we can conclude the following:

- Married Applicants holding Academic degree have a higher Credit Amount (Defaulted and Non-Defaulted Loans)
- Overall Academic degree Applicants have highest Credit Amount
- Defaulted Loans are mostly of Married categories across Education type

Income Range Vs Credit Amount For Target 0 and 1

Income Range vs Credit Amount of Non-Defaulted Loans
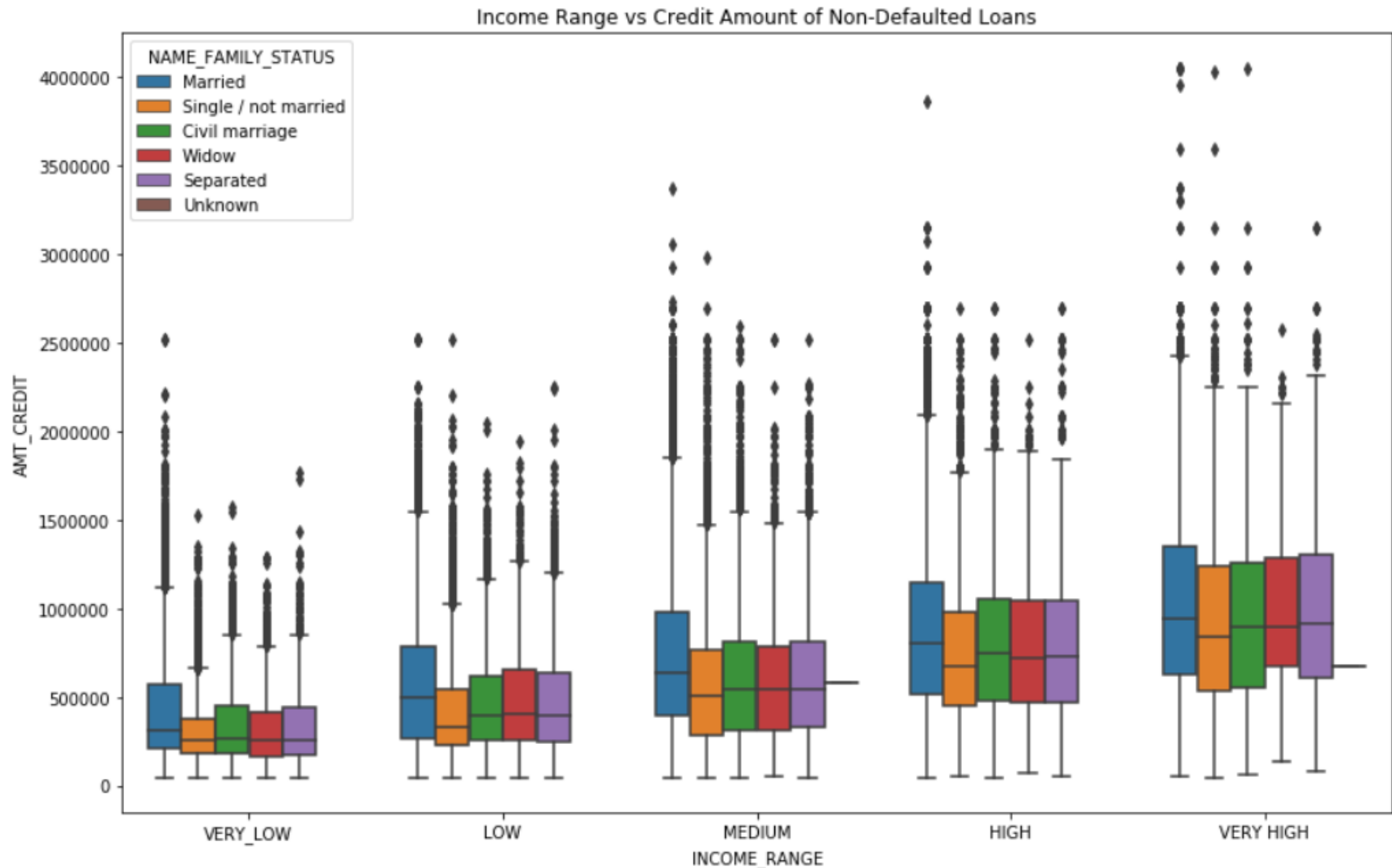
Income Range vs Credit Amount of Defaulted Loans
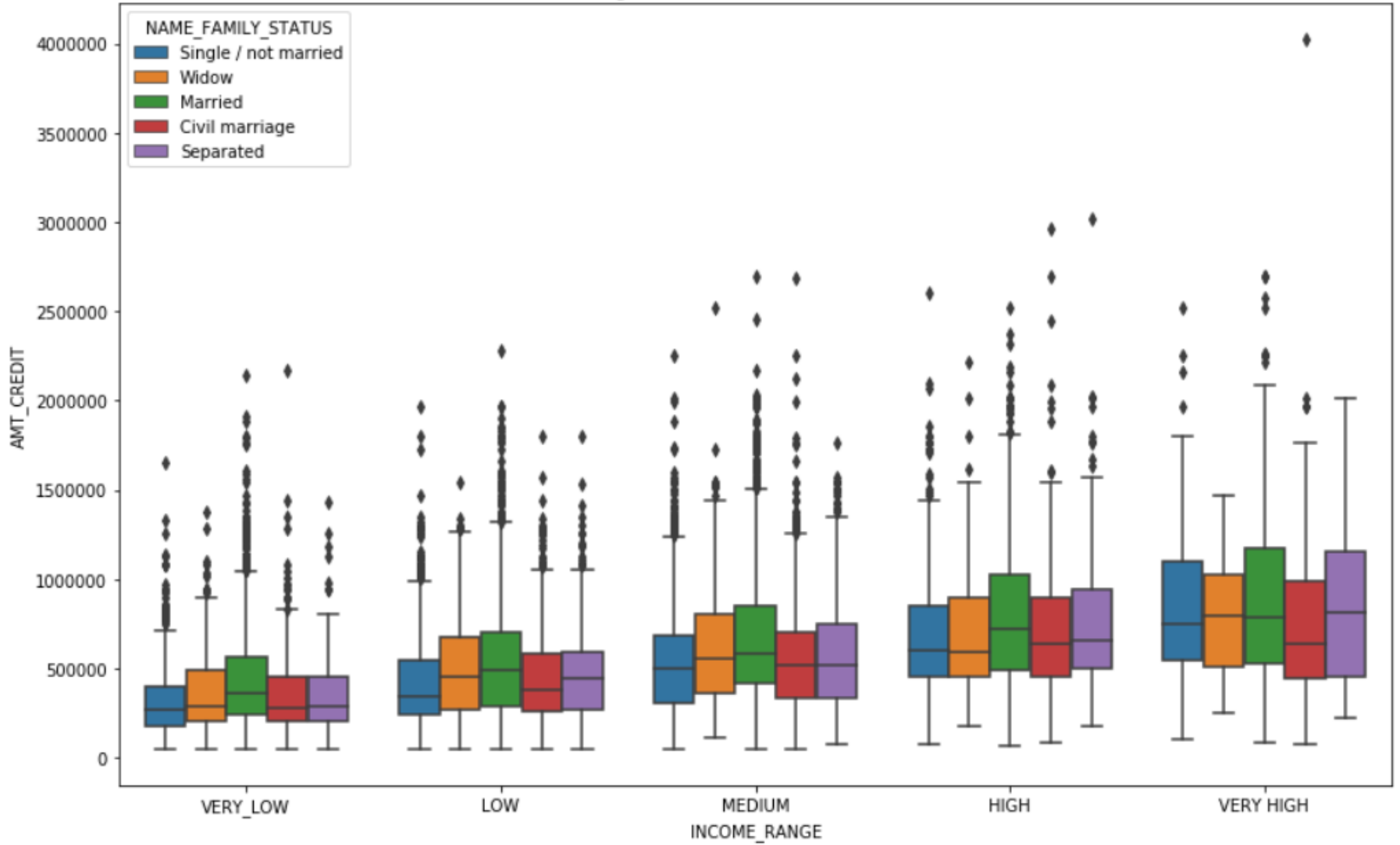
From the above charts, we can conclude the following:
- There is a correlation between Income Ranges and Amount Credited. Higher the Income, higher the amount credited.
- Defaulted Loans are mostly of Married categories across Income Ranges

Correlation Of Parameters For Non-Defaulted Loans

Correlation Heatmap for Non-Deafulted Loans
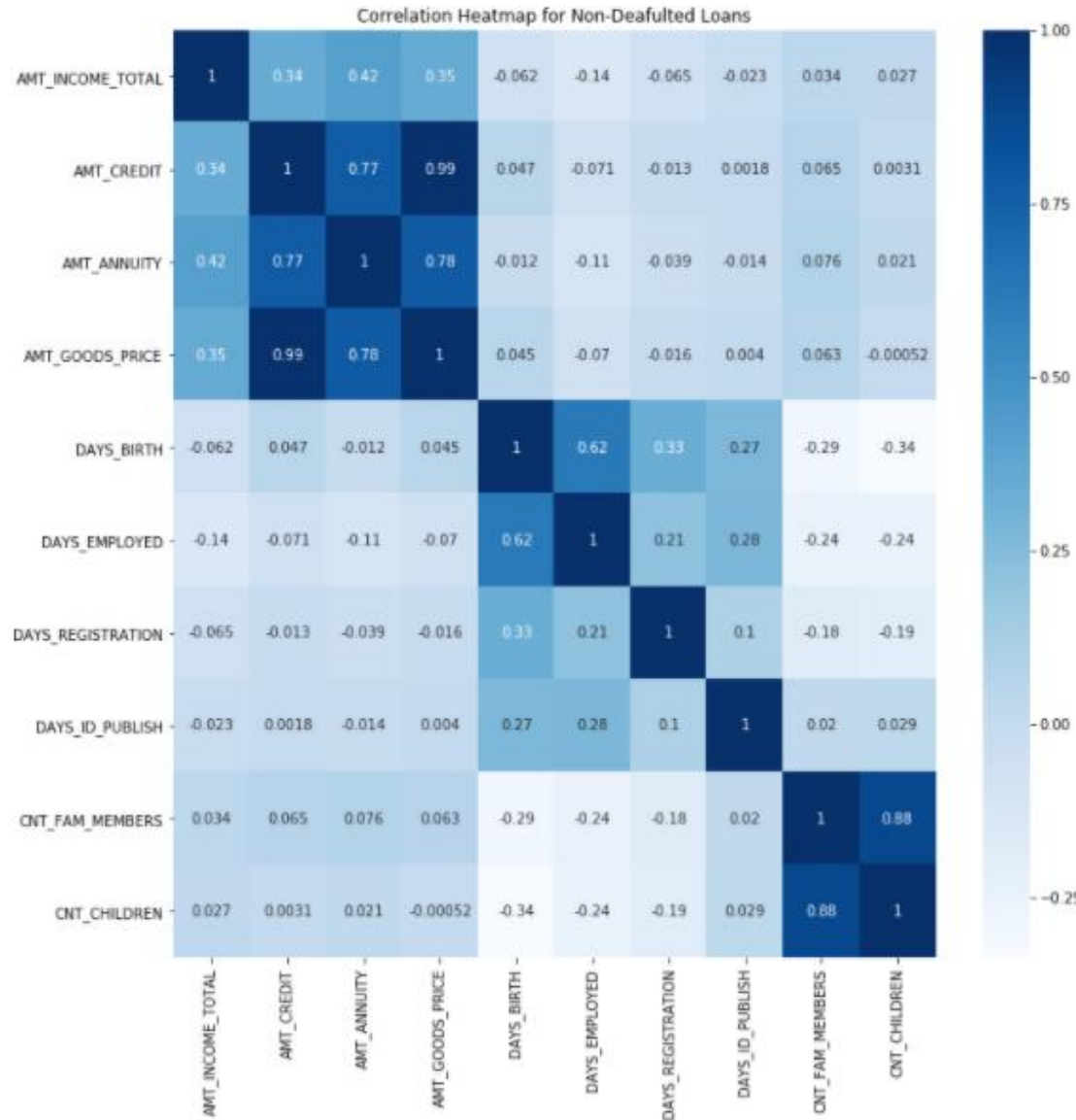
Top 10 Correlations for Non Defaulted loans

1. AMT_CREDIT & AMT_GOODS_PRICE
2. CNT_FAM_MEMBERS & CNT_CHILDREN
3. AMT_GOODS_PRICE & AMT_ANNUITY
4. AMT_ANNUITY & AMT_CREDIT
5. DAYS_EMPLOYED & DAYS_BIRTH
6. AMT_INCOME_TOTAL & AMT_ANNUITY
7. AMT_INCOME_TOTAL & AMT_GOODS_PRICE
8. AMT_INCOME_TOTAL & AMT_CREDIT
9. DAYS_REGISTRATION & DAYS_BIRTH
10. DAYS_EMPLOYED & DAYS_ID_PUBLISH

Correlation Of Parameters For Defaulted Loans

Correlation Heatmap for Deafulted Loans
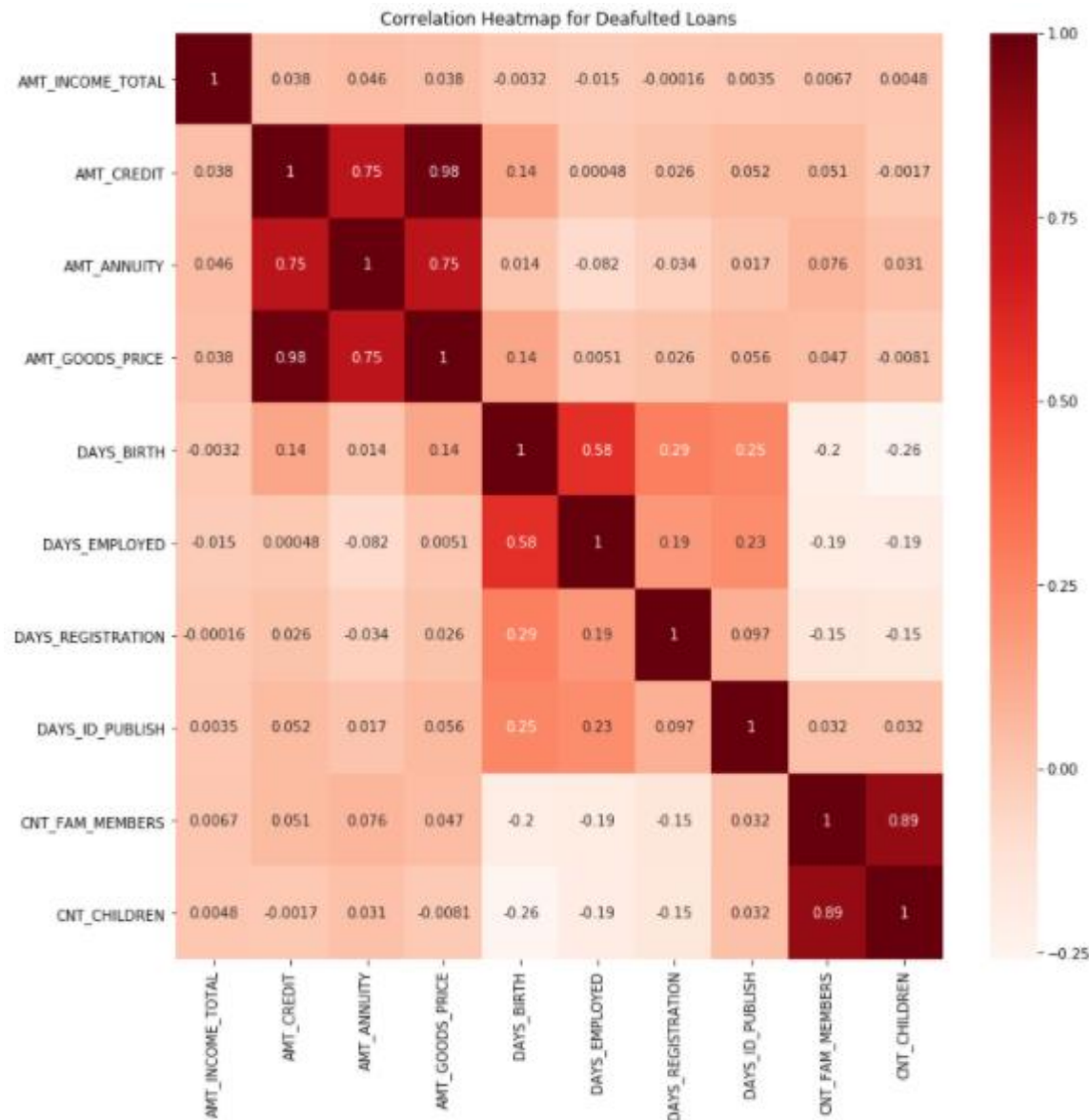
Top 10 Correlations for Defaulted loans

1. AMT_CREDIT & AMT_GOODS_PRICE
2. CNT_FAM_MEMBERS & CNT_CHILDREN
3. AMT_GOODS_PRICE &   AMT_ANNUITY
4. AMT_ANNUITY & AMT_CREDIT
5. DAYS_EMPLOYED & DAYS_BIRTH
6. DAYS_REGISTRATION & DAYS_BIRTH
7. DAYS_ID_PUBLISH & DAYS_BIRTH
8. DAYS_ID_PUBLISH & DAYS_EMPLOYED
9. DAYS_REGISTRATION & DAYS_EMPLOYED
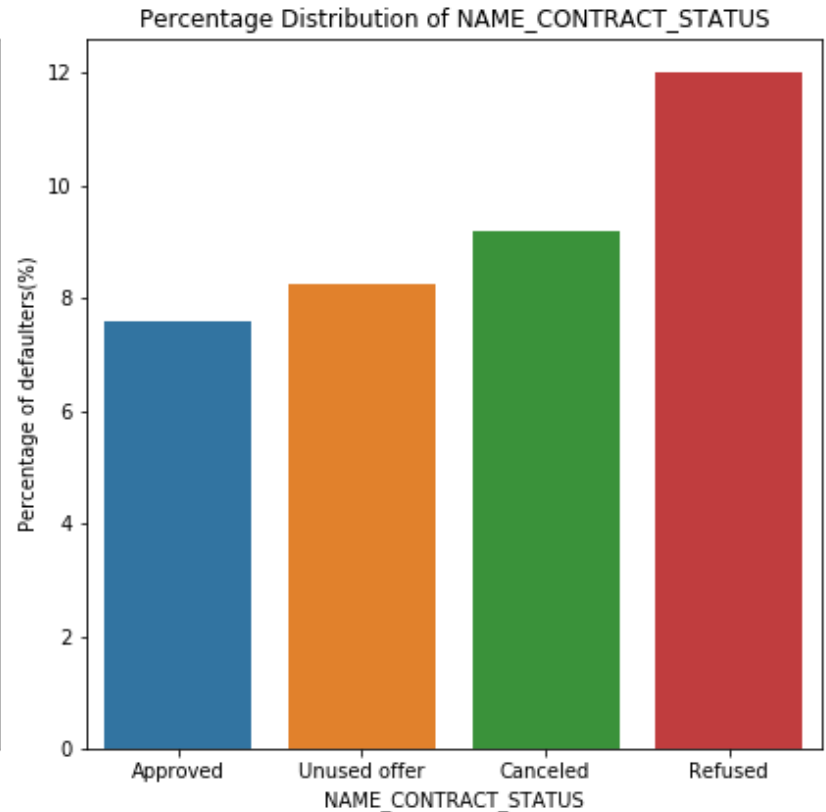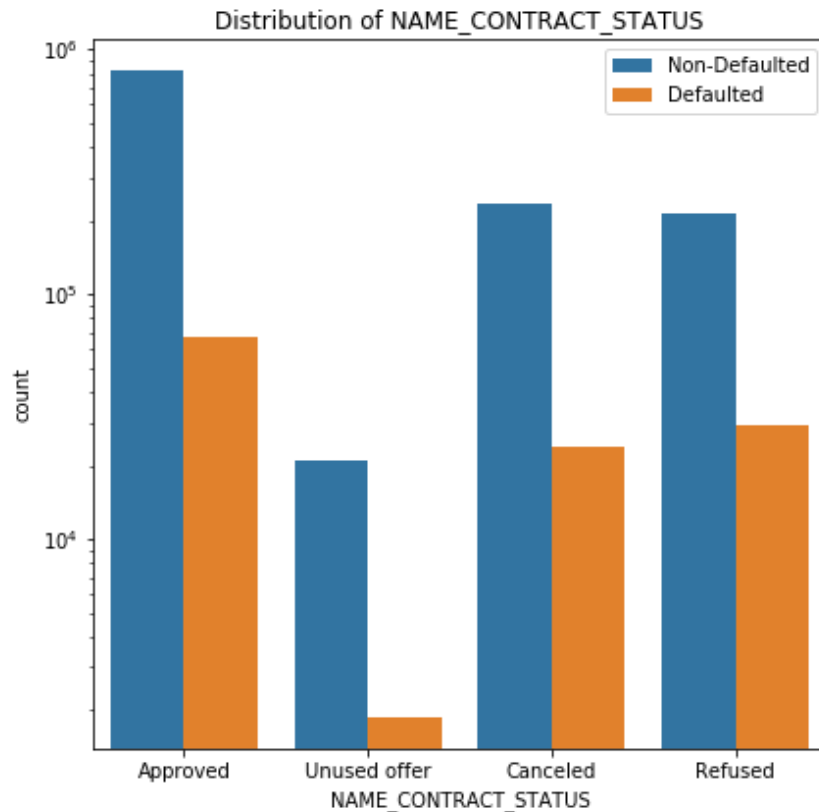10. DAYS_ID_PUBLISH & DAYS_REGISTRATION

Strong Correlation Findings:

- Amount Credited and Goods Price

- Count Family Members and Count Children

- Amount Annuity and Amount Credit

ANALYSIS OF PREVIOUS APPLICATION DATA

Distribution of Contract Status
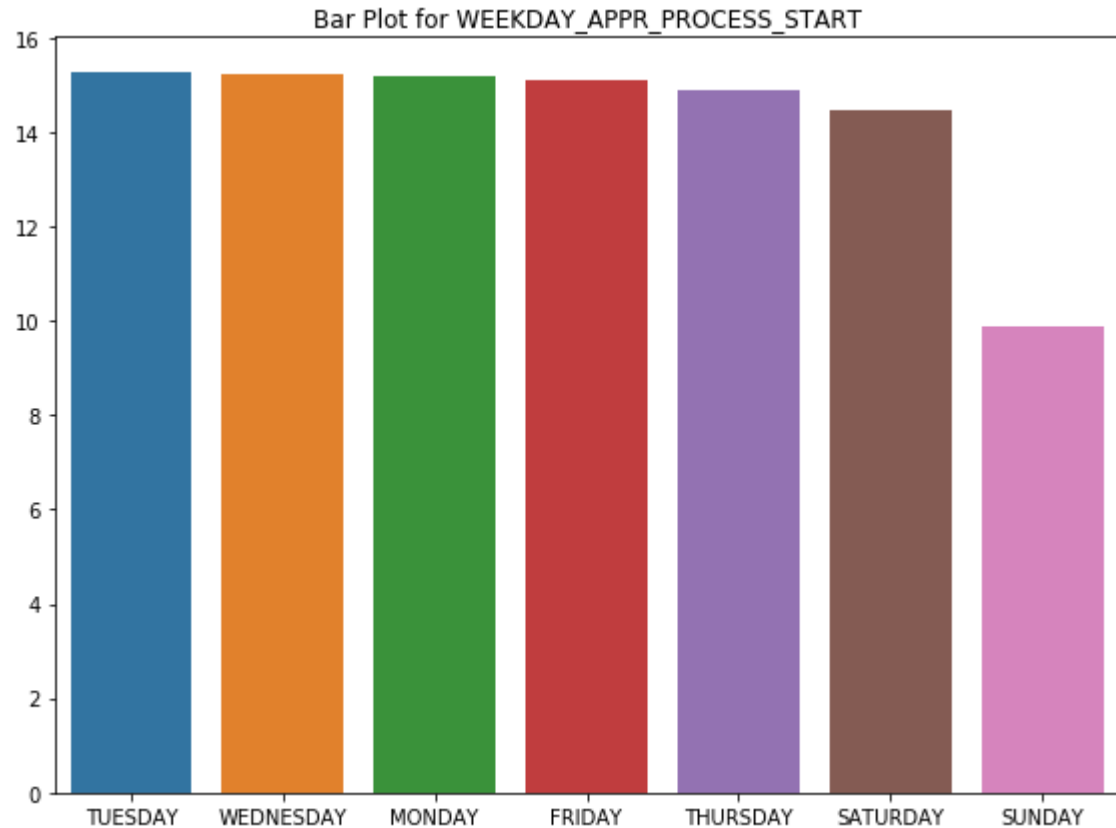

Pie Chart for NAME_CONTRACT_STATUS
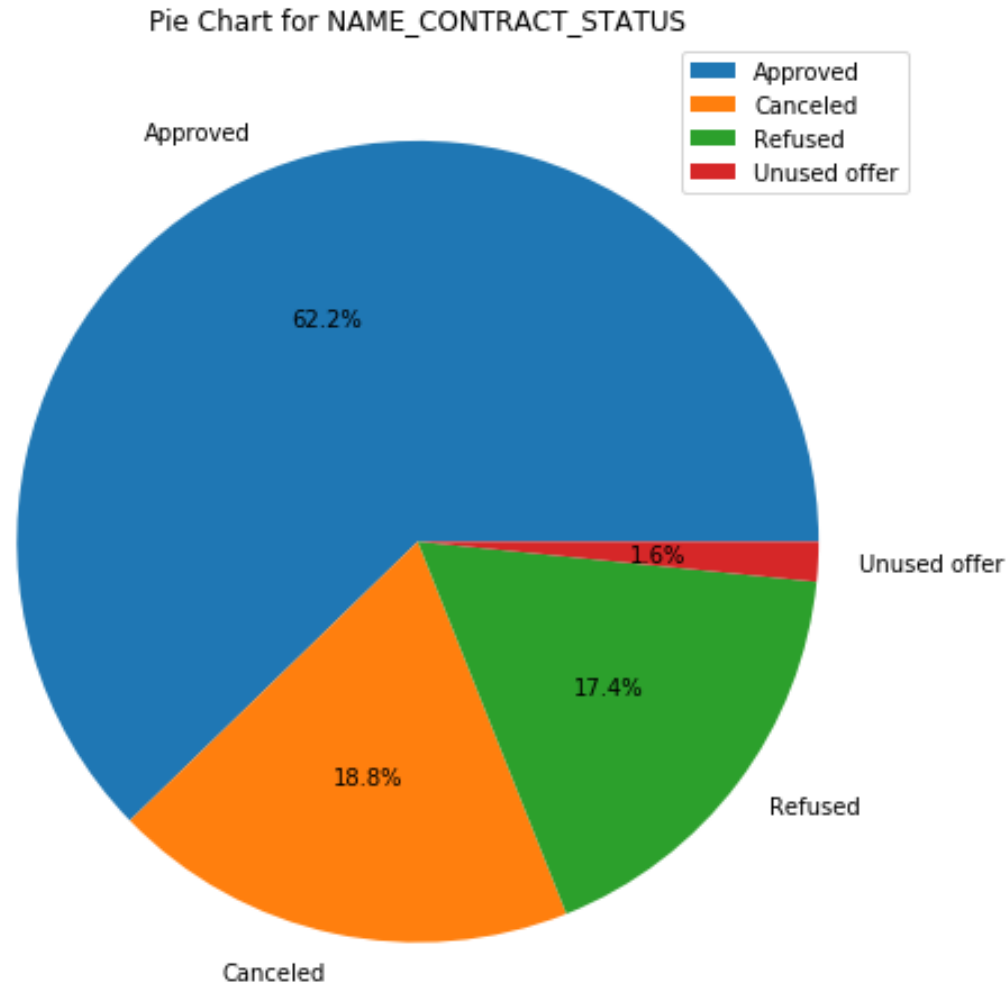
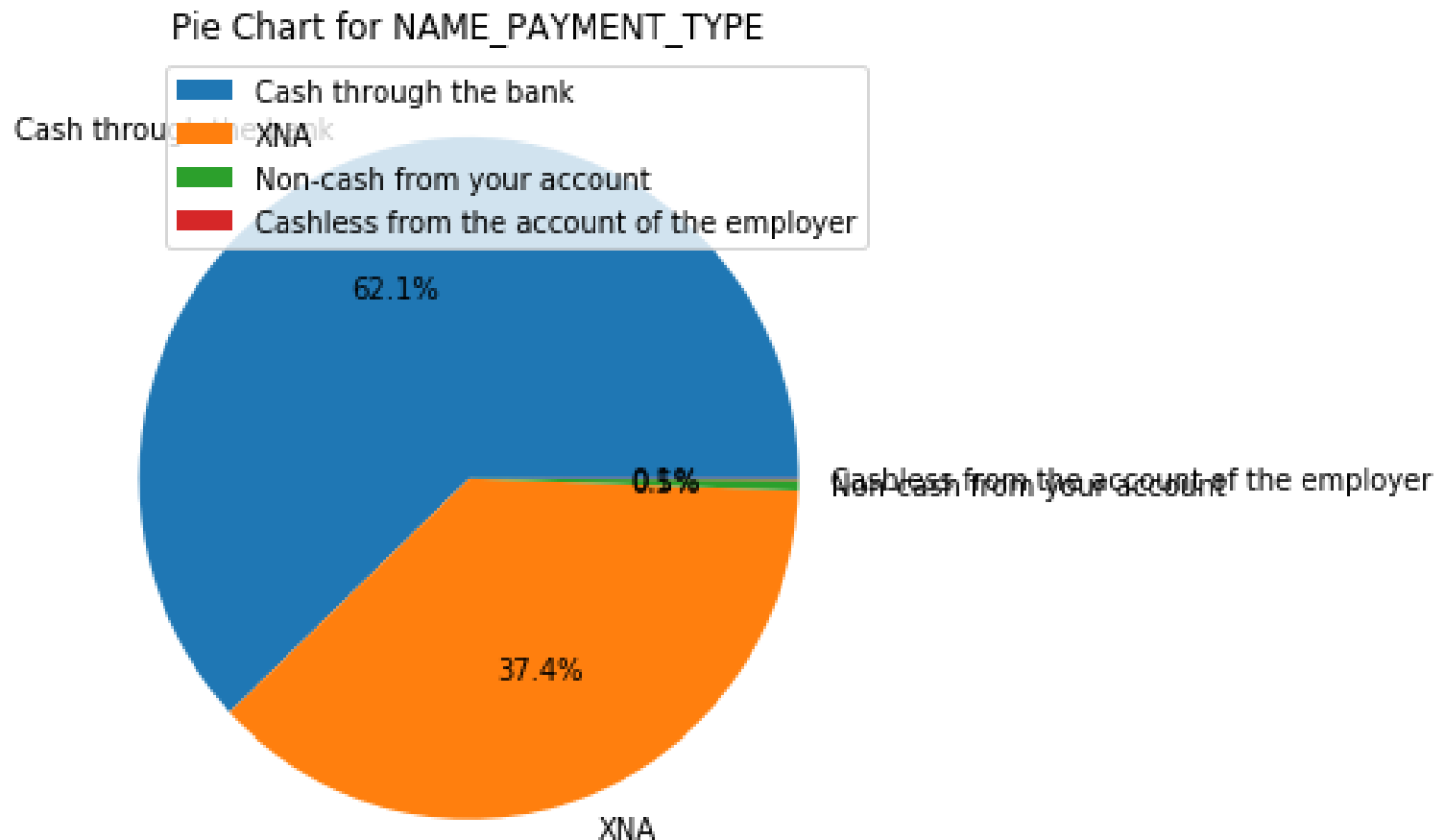1.Majority of loans are approved
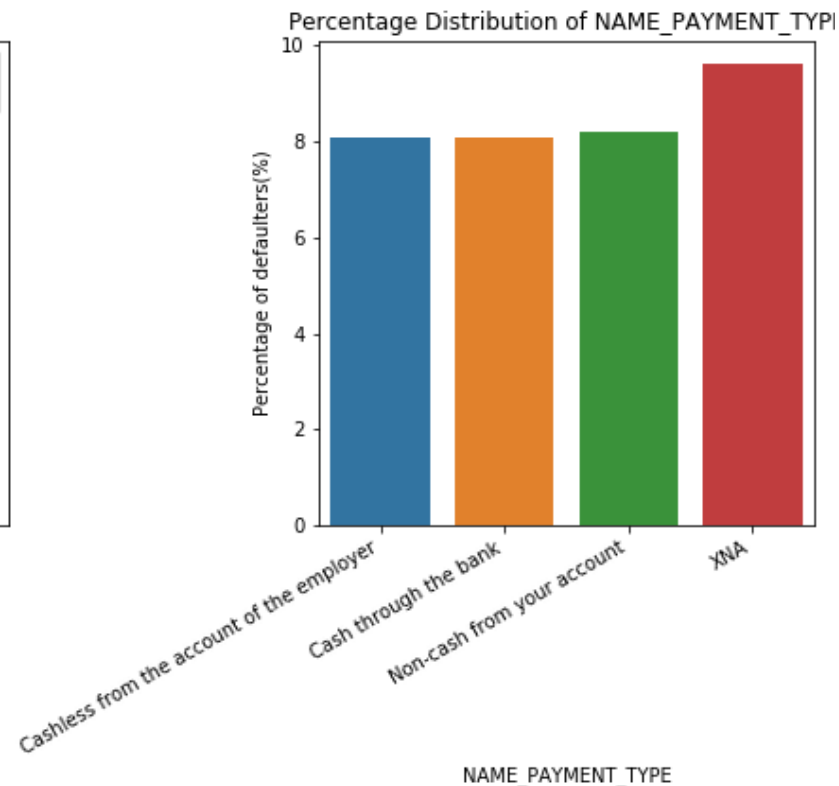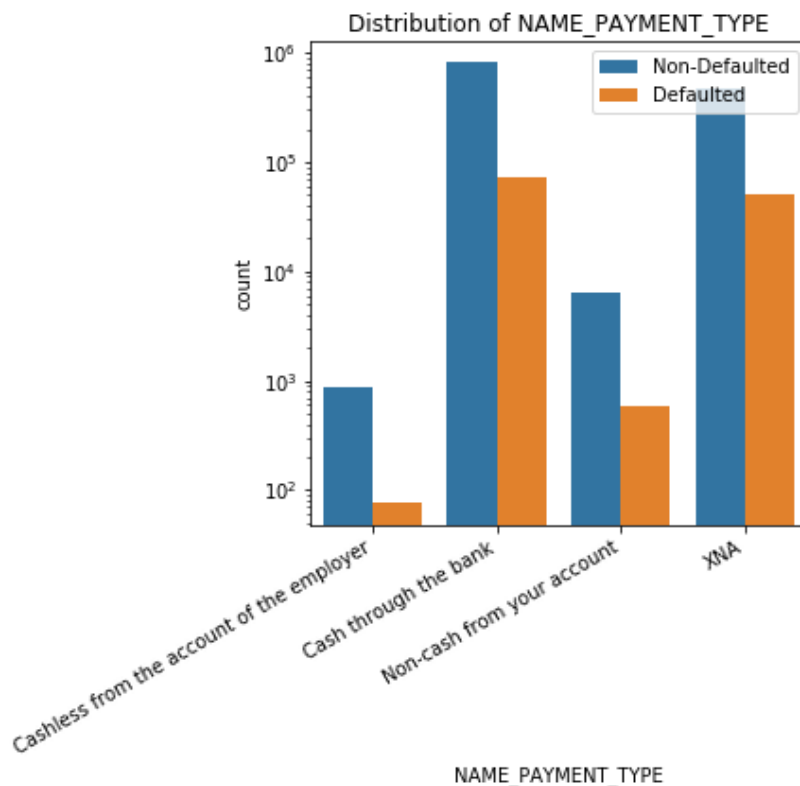2.Canceled and Refused loans status have more Defaulters

Distribution of Children

Observations:

- Weekends have the lowest applications

Bar Plot for WEEKDAY_APPR_PROCESS_START

Distribution of Payment Type

Pie Chart for NAME_PAYMENT_TYPE

Legend:
- Cash through the bank
- XNA
- Non-cash from your account
- Cashless from the account of the employer

Cash through the bank — 62.1%

XNA — 37.4%

0.5%

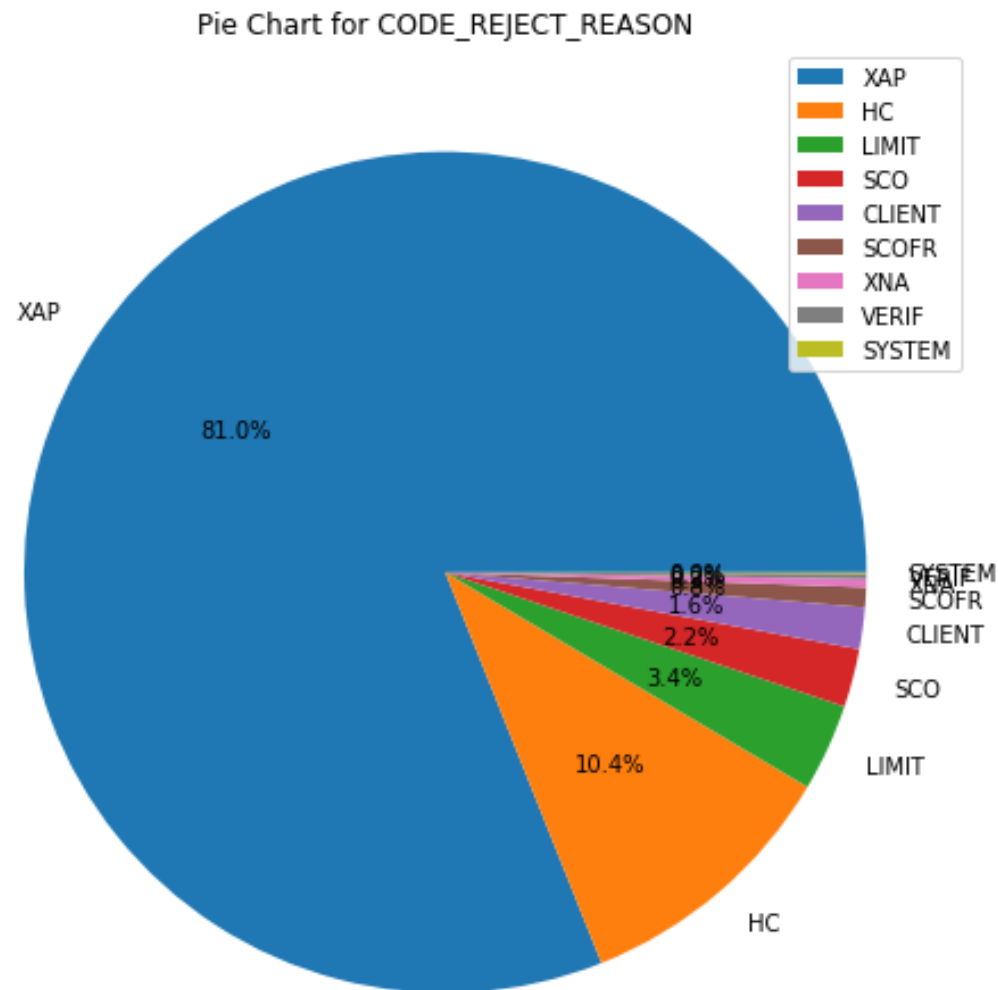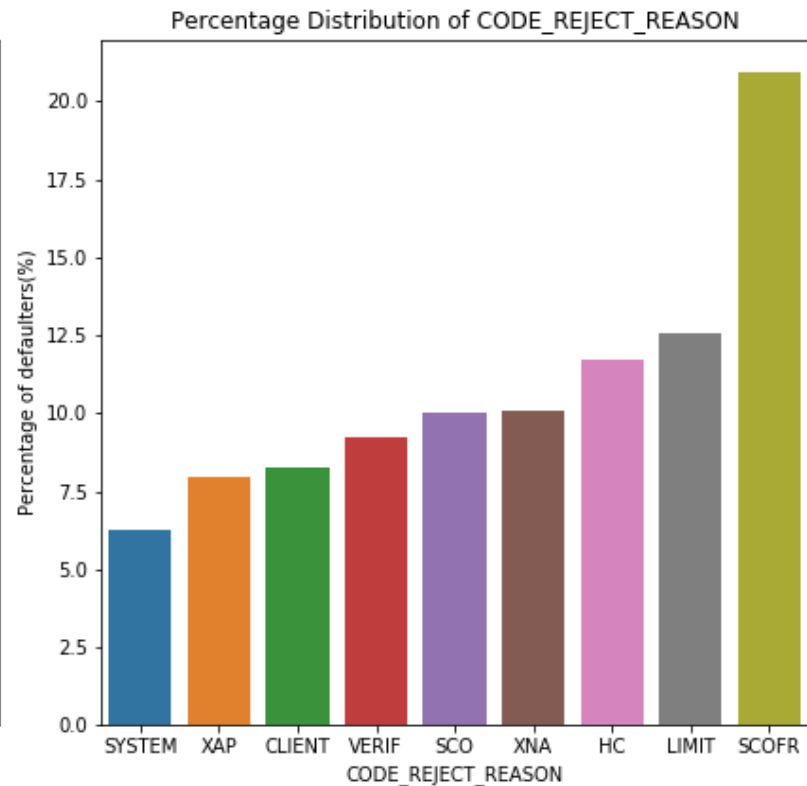Cashless from the account of the employer / Non-cash from your account
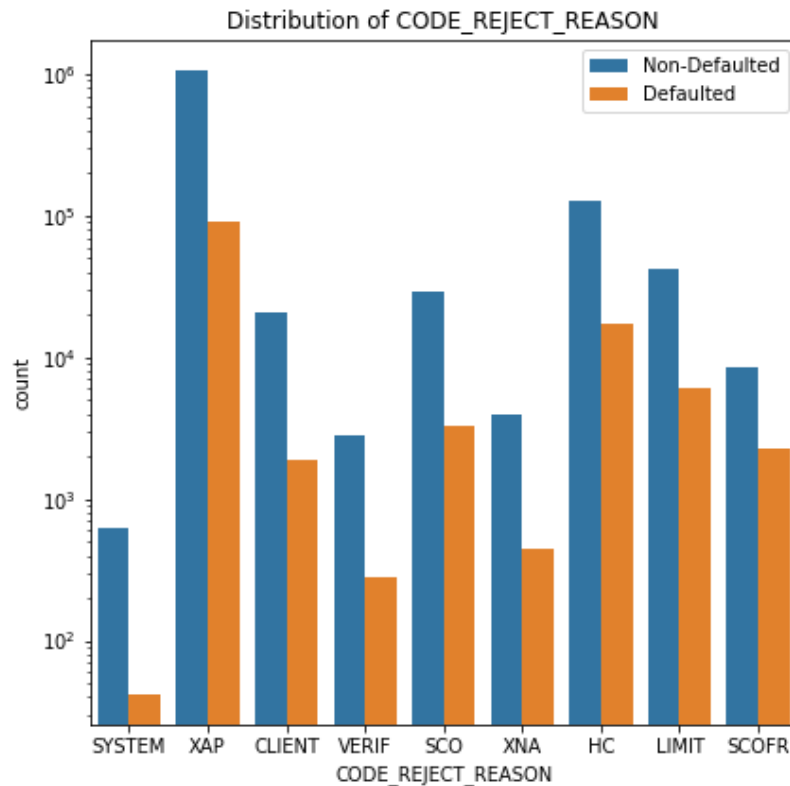
Observations:

- Majority of the loans have Payment Type - Cash through the
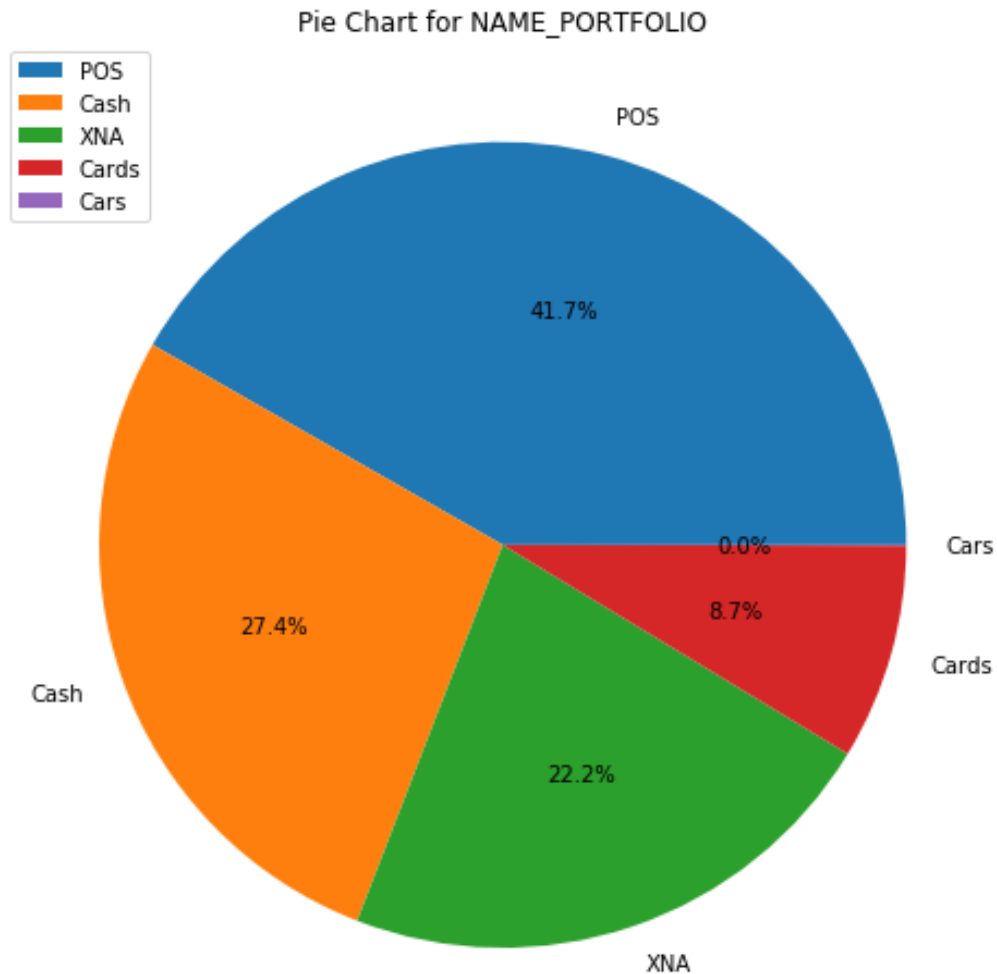 bank

Distribution of Code Reject Reason

Pie Chart for CODE_REJECT_REASON

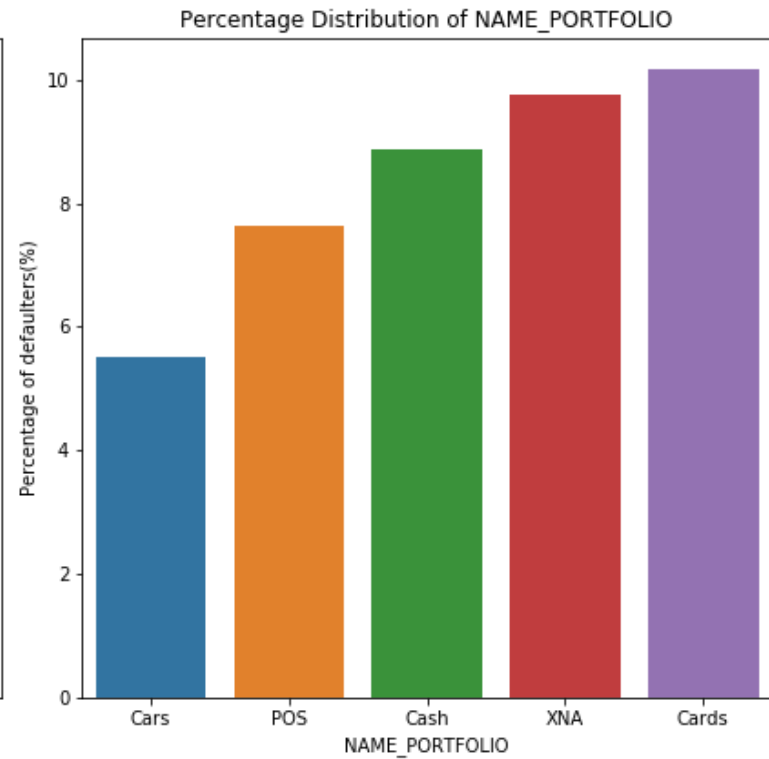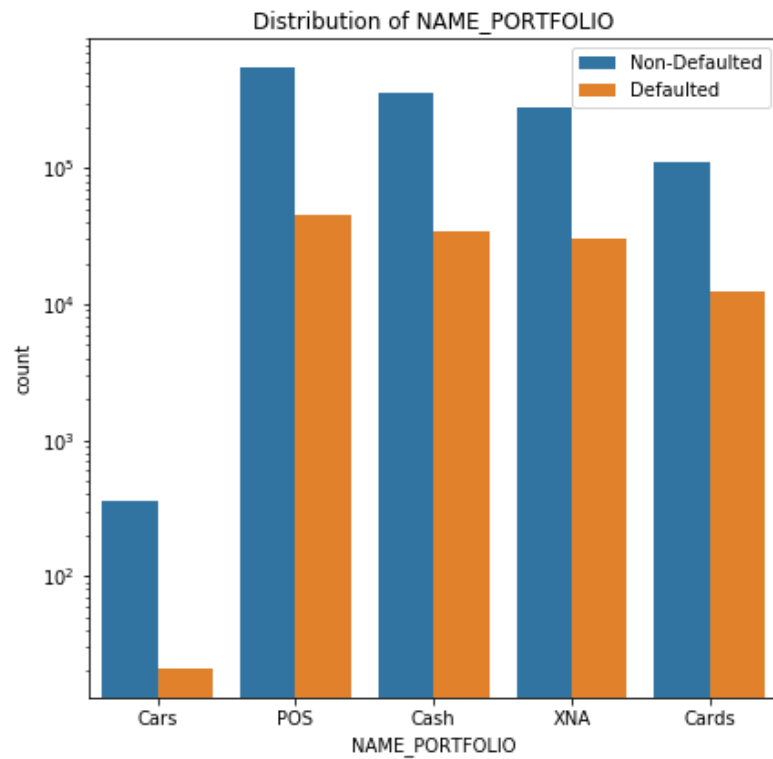Distribution of CODE_REJECT_REASON / Percentage Distribution of CODE_REJECT_REASON

1.'SCOFR' has the highest percentage for Reject Reason
2.'XAP' has the highest overall percentage

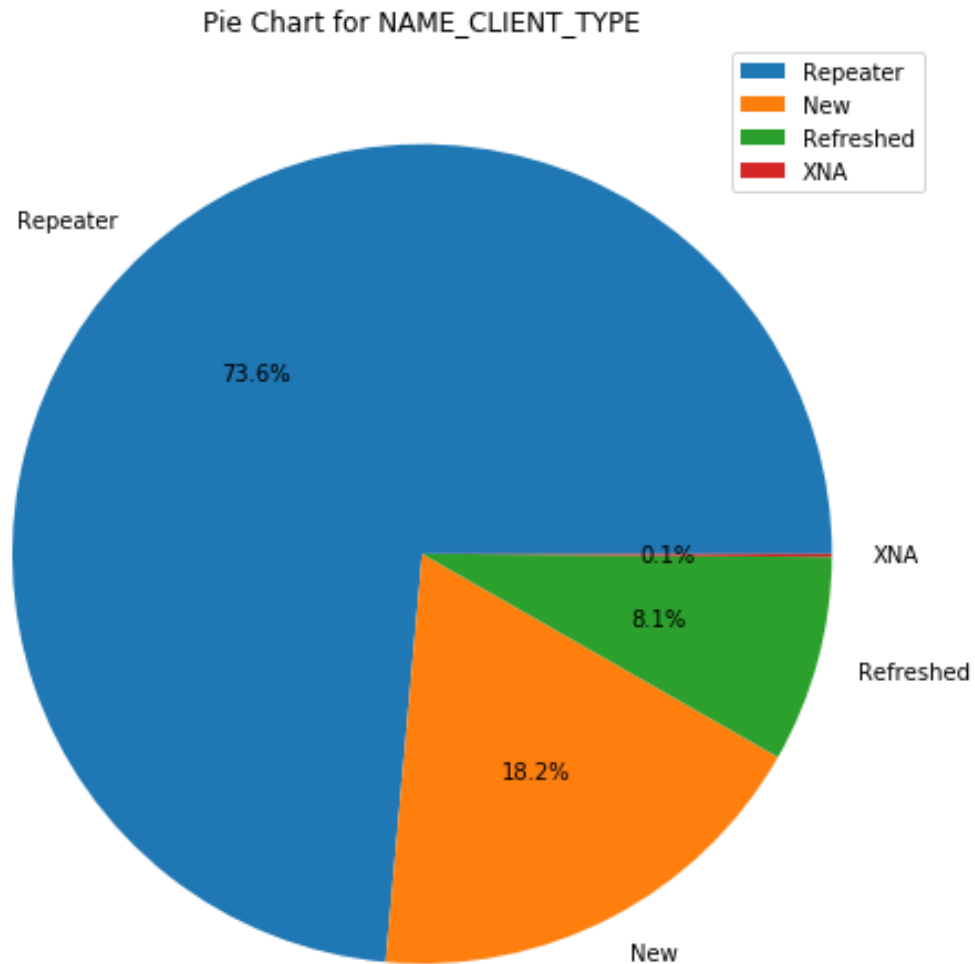# Distribution of Portfolio

Pie Chart for NAME_PORTFOLIO

- POS
- Cash
- XNA
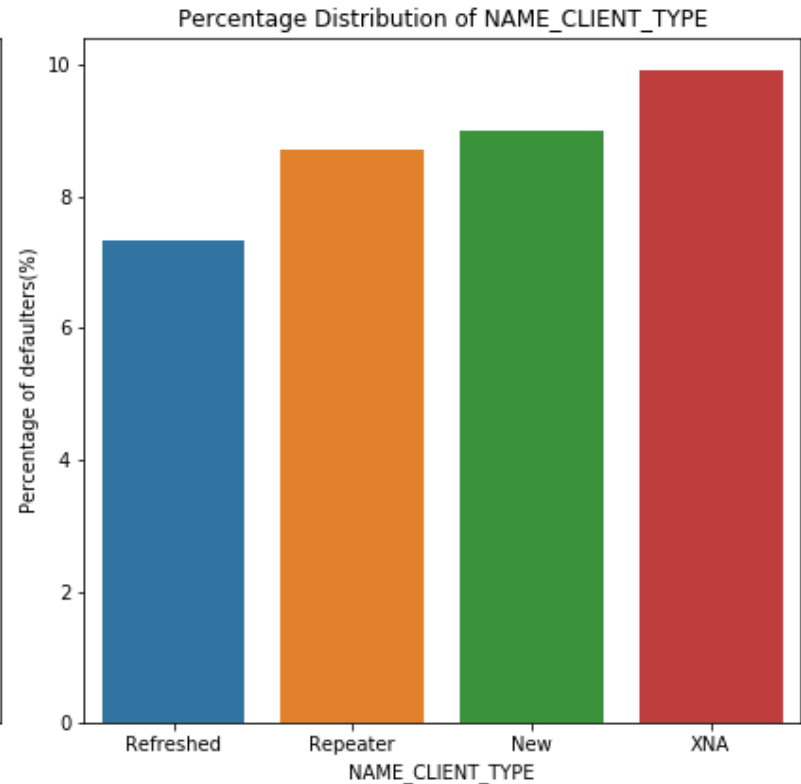- Cards
- Cars

POS 41.7%

Cars 0.0%

Cards 8.7%

Cash 27.4%

XNA 22.2%
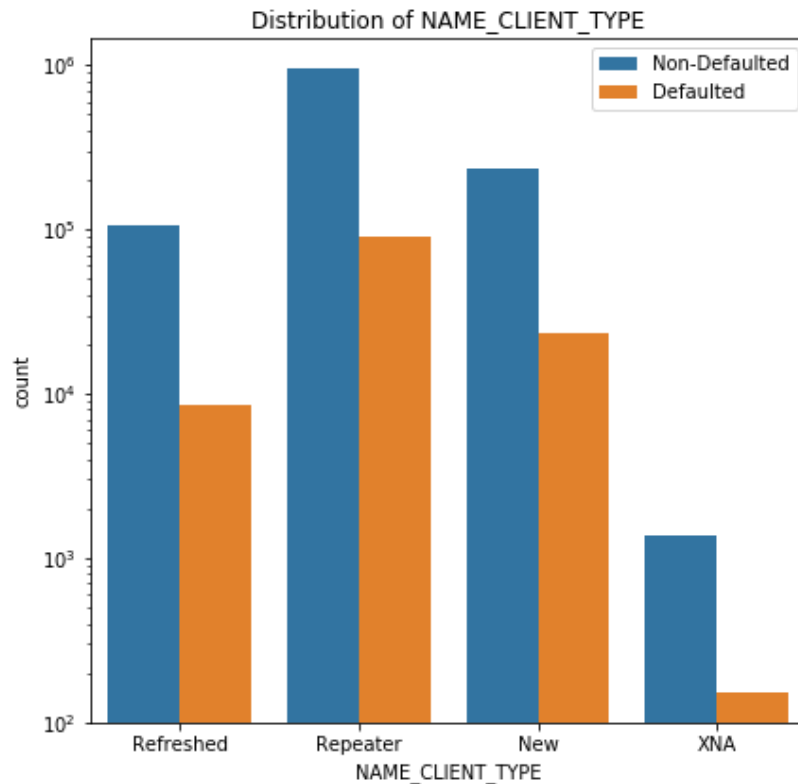
1.'POS' and 'Cash' lead the Portfolio metric
2.'Cards' category has the highest Rejection percentage

Distribution of Client Type

Pie Chart for NAME_CLIENT_TYPE

Observations:

- Majority of the Clients are Repeater.

- Refreshed client types have lowest rejection percentage

CONCLUSIONS

- Banks should focus more on giving loans to students, businessmen than working professionals and commercial associates.

- Banks must focus more on people with academic degree ,experience of more than 5 years in their job mainly HR and IT staff other than labourers, sales staff and drivers

- Single, widow/separated and persons with age more than 60 years should be given more preference

- Get more clients from high income, high credit and high annuity group

- Give more preference to people living in co op apartments and office apartments