

The background features a series of concentric circles in light gray, some solid and some dashed, creating a ripple effect. A large, solid red oval is centered on the page, containing the main text. A dark gray, curved shape is positioned to the left of the red oval, partially overlapping it.

# EDA Credit Assignment

Analyzed by: Gursewak Singh

Batch code: 3548

Program: MS in Data Science



# Agenda / Topics

- Problem Statement
- Project Methodology/approach
- Key Findings/Results
- Outcomes

# Problem Statement

- Loan providing company hesitates to provide loan to people who have either no or poor credit history. And there are two risk factors involved here:
  - - Company will lose the business if they reject the potential customers.
  - - The company will face financial losses if they sanction the loan to applicants who may default
- Given the data, target is to perform end to end EDA to ensure that applicants who are capable of repaying the loan are not rejected. The inferred knowledge from this EDA would help the loan company for its portfolio and risk assessment. Moreover to lower the potential defaults.
- Scenarios: 1) Client with payment difficulties, 2) All other cases
- Types of Decision:
  - 1. Approved,
  - 2. Cancelled,
  - 3. Refused,
  - 4. Unused Offer

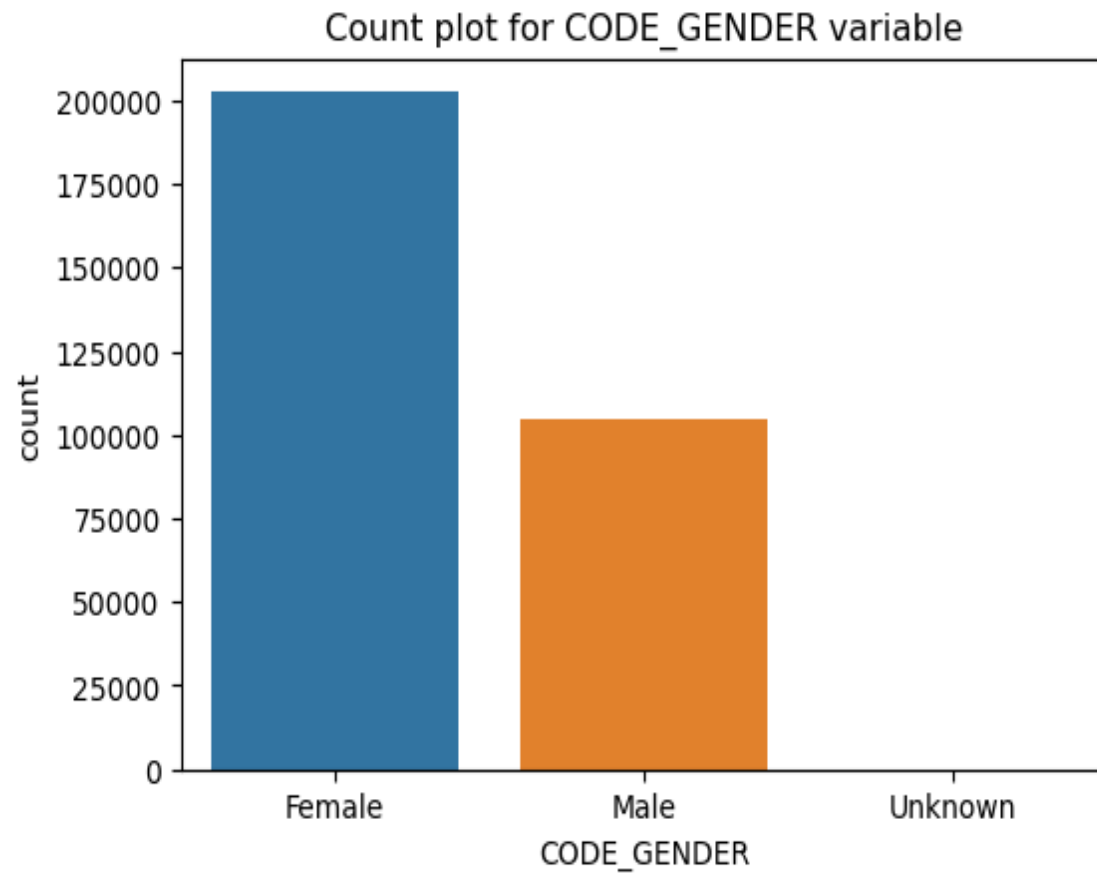
# Procedure/Methodology

- **Section-1 : Data Sourcing and Understanding**
  - Current application data set
  - Previous application data set
- **Section-2: Data Cleaning**
  - **Current application**
    - Data types and null values calculation
    - Fixing rows and columns
    - Impute or remove missing values
    - Handling outliers
    - Standardizing values
    - Fixing invalid values
  - **Previous application**
    - Data types and null values calculation
    - Fixing rows and columns
    - Impute or remove missing values
    - Handling outliers
    - Standardizing values
    - Fixing invalid values

# Procedure/Methodology cont..

- **Section-3 Analysis**
  - **Current application**
    - Univariate Analysis
    - Bivariate/multivariate Analysis
  - **Merged Data Set (current application + previous application)**
    - Univariate Analysis
    - Bivariate Analysis
- **Section 4: Outcomes**

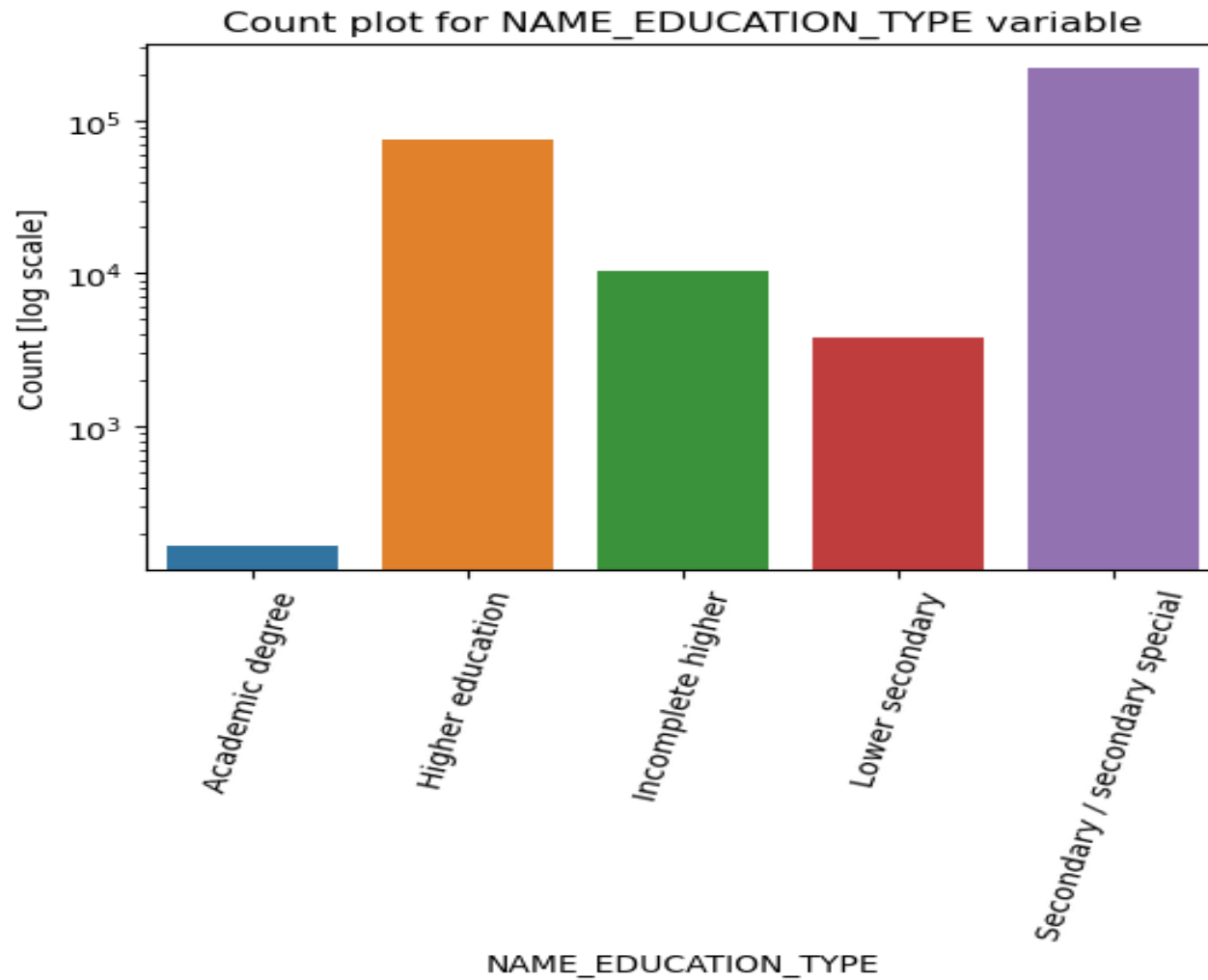
## Key Finding and Results [CODE\_GENDER]



```
F      65.834710
M      34.163989
Unknown  0.001301
Name: CODE_GENDER, dtype: float64
```

Females are most interested in loans as compared to males, which is almost 66%.

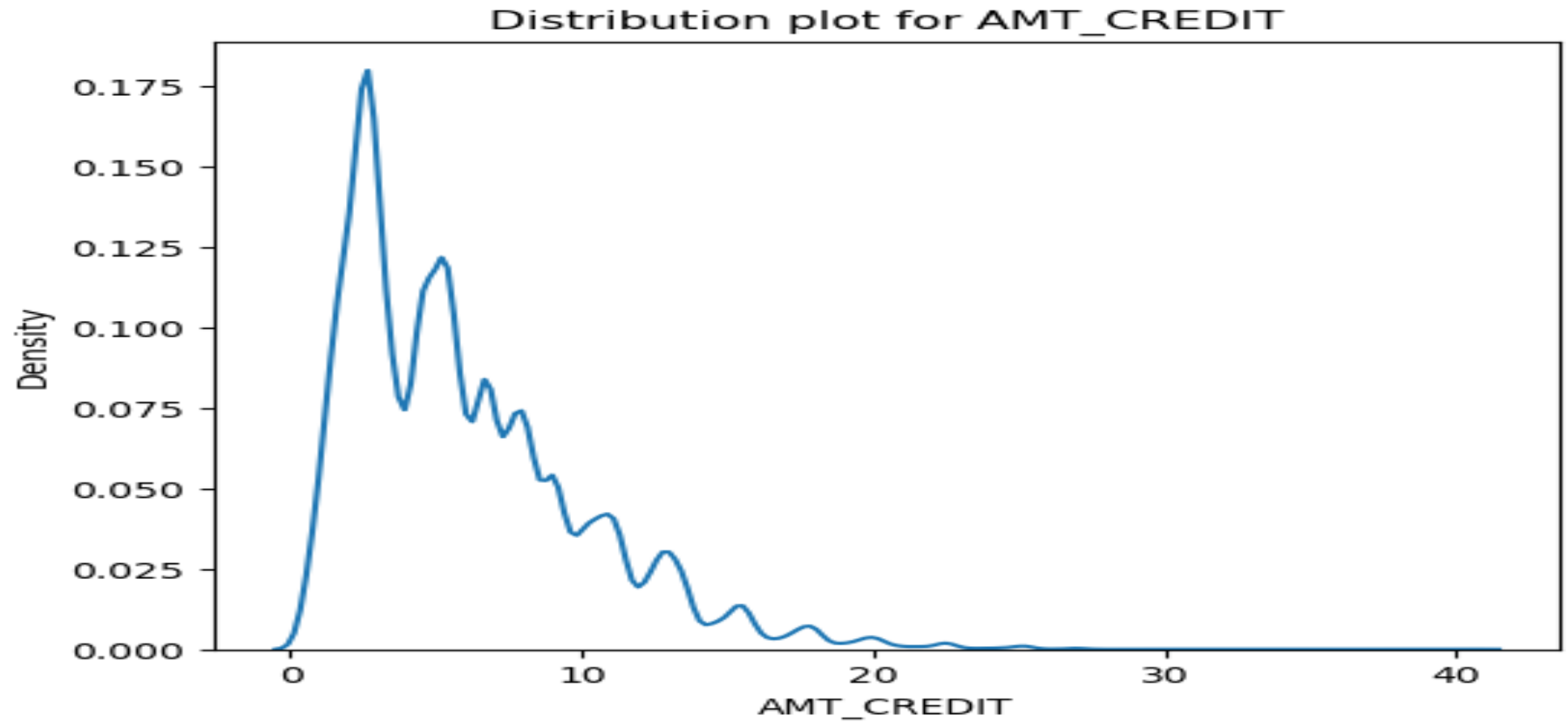
# Key Finding and Results [NAME\_EDUCATION\_TYPE]



```
Secondary / secondary special    71.019616
Higher education                 24.344732
Incomplete higher                3.341702
Lower secondary                 1.240618
Academic degree                 0.053332
Name: NAME_EDUCATION_TYPE, dtype: float64
```

- Vast majority of application are coming from applicants with Secondary/secondary special (71%).
- Academic degree holder applying for loan are few.

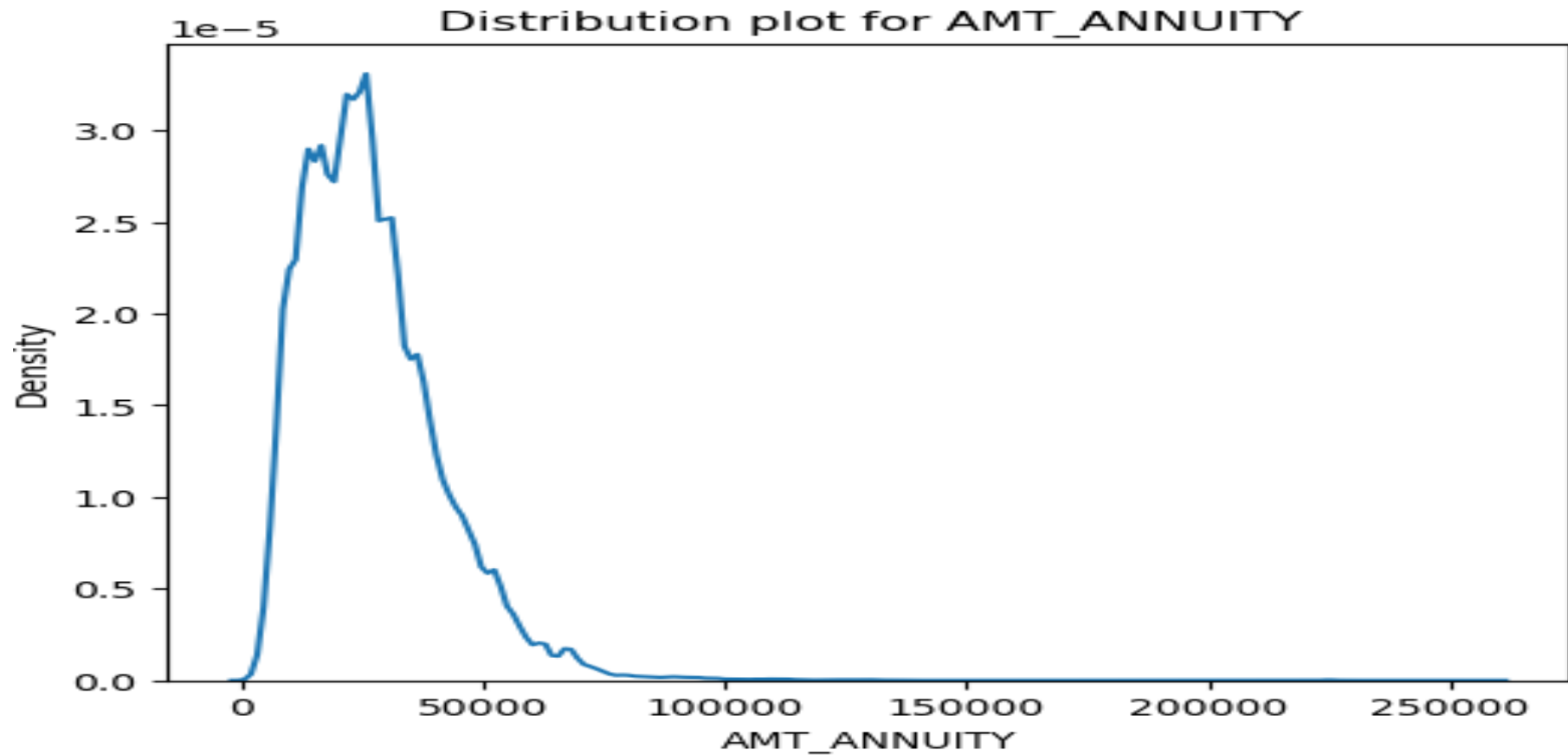
## Key Finding and Results [AMT\_CREDIT]



- Vast majority of loans were given for 10lacs or less.

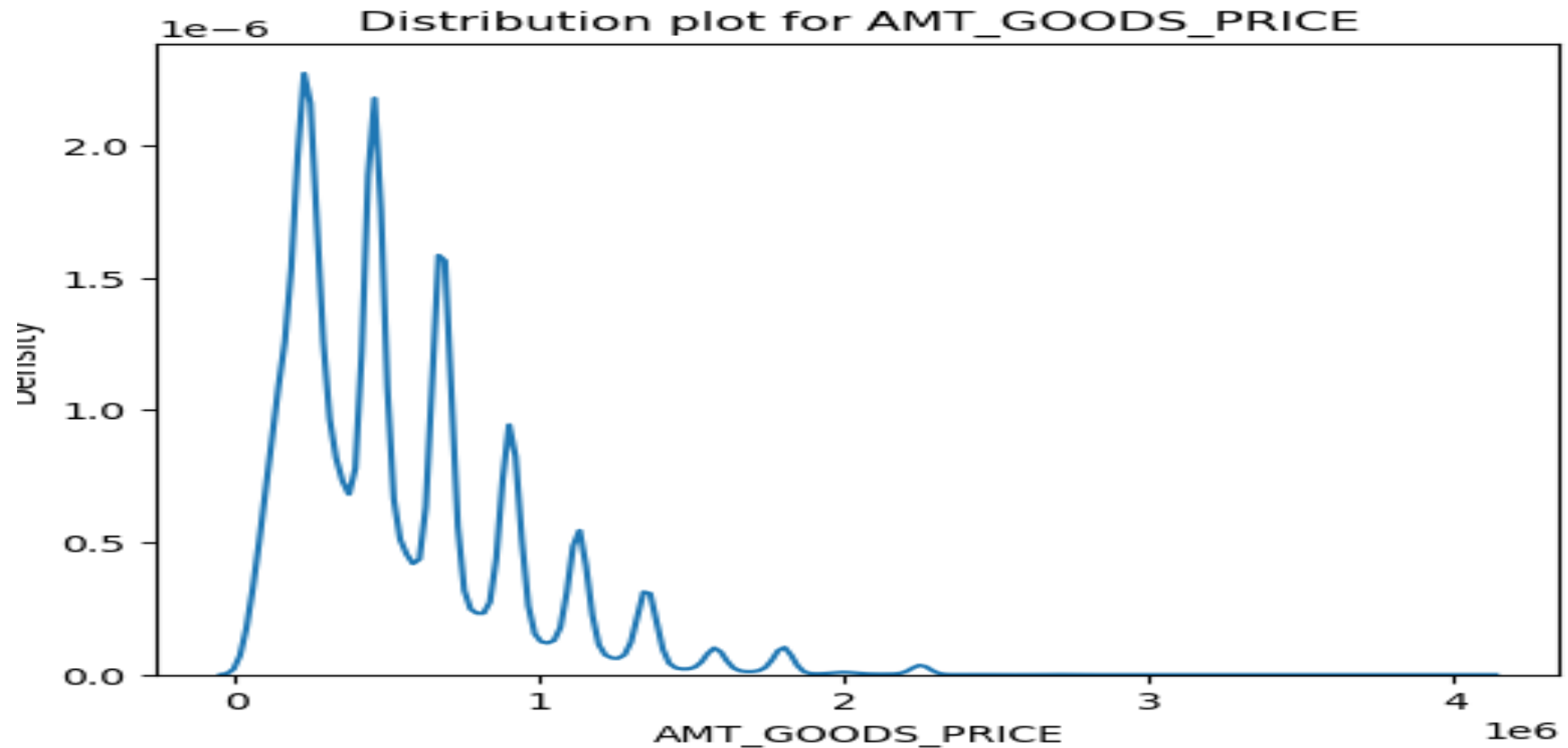


## Key Finding and Results [ AMT\_ANNUITY]



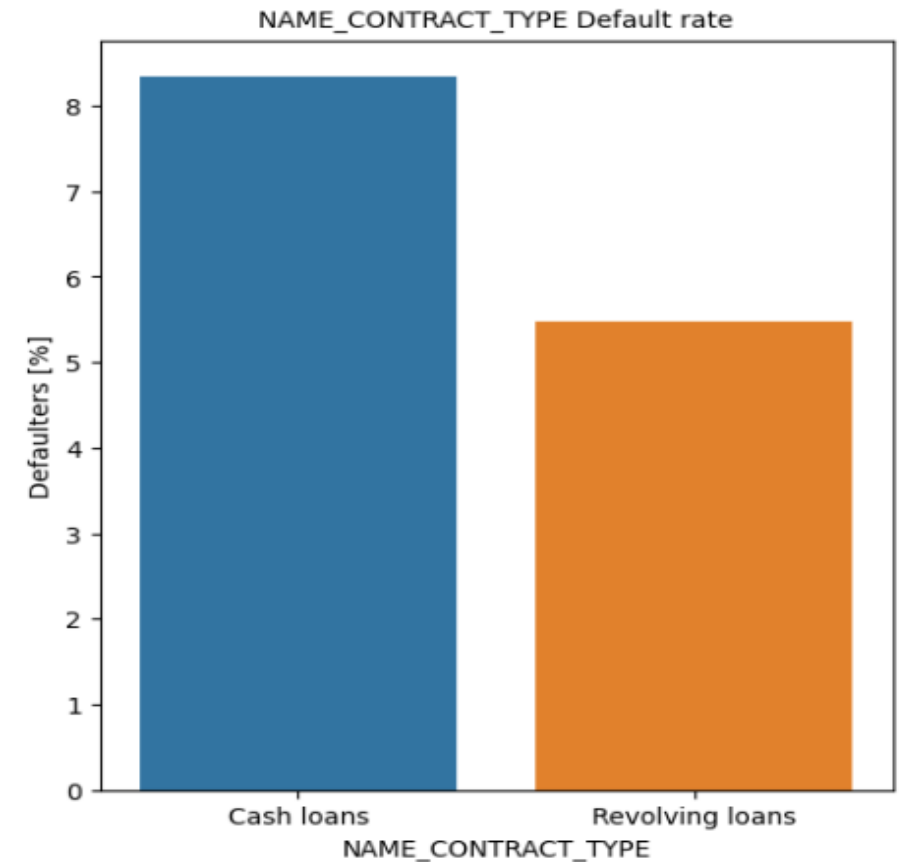
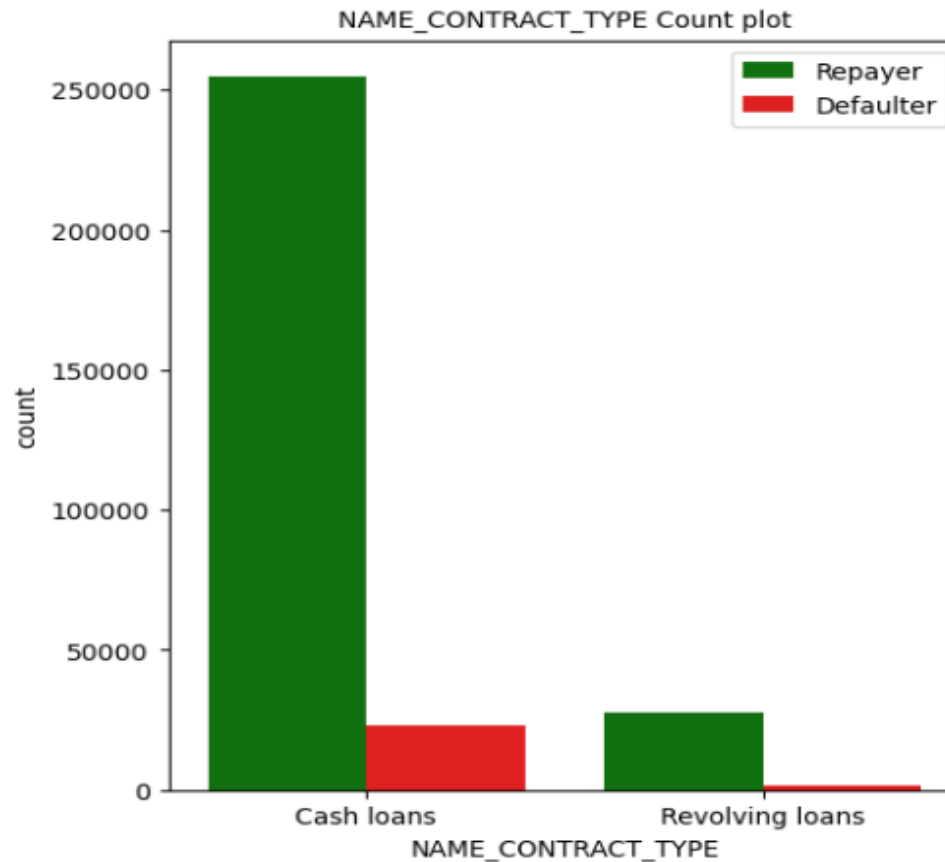
- Applicants mostly applied for loans with annuity of 50K

## Key Finding and Results [AMT\_GOODS\_PRICE]



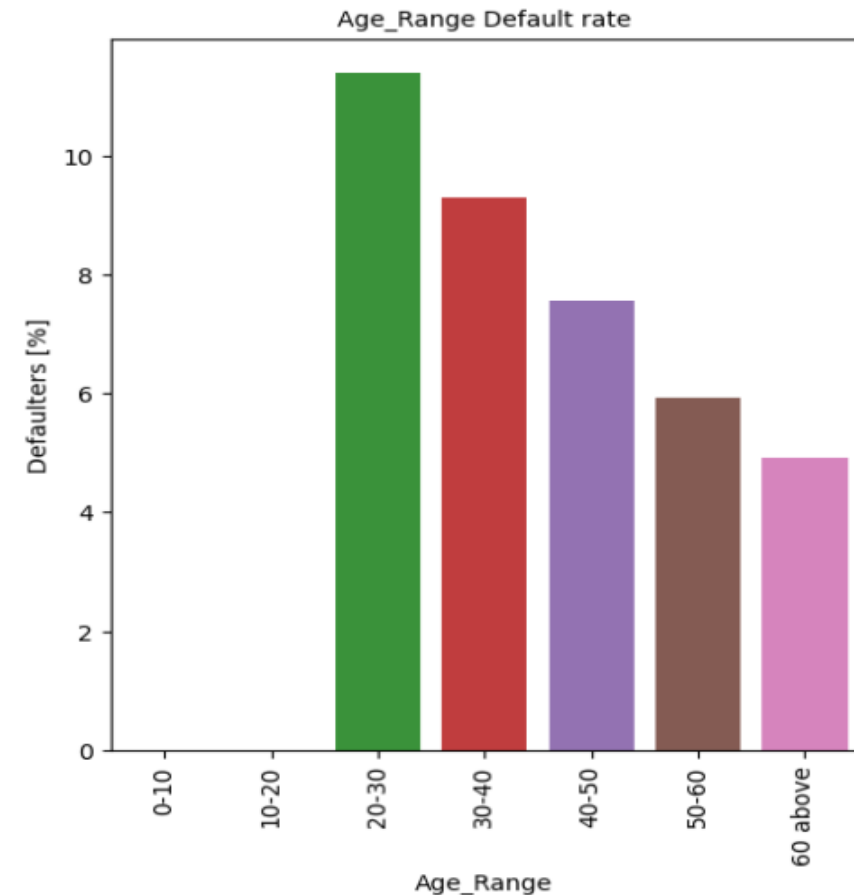
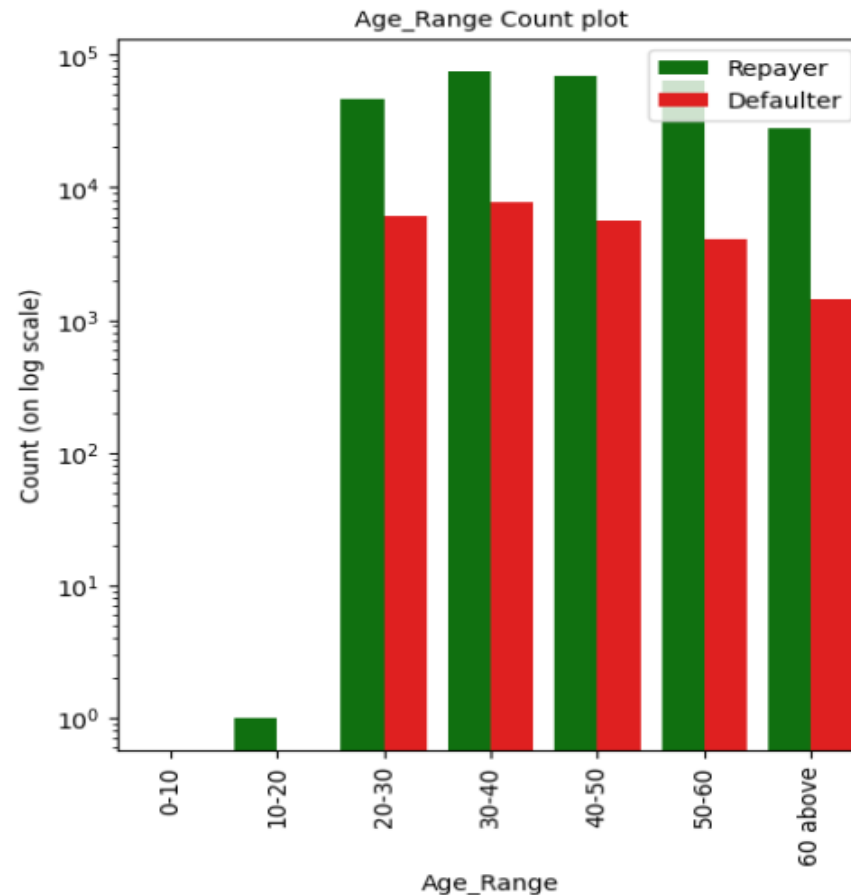
- Majority of applicants applied for loan with a goods price of 10 lacs or less

## Key Finding and Results [ NAME\_CONTRACT\_TYPE VS TARGET]



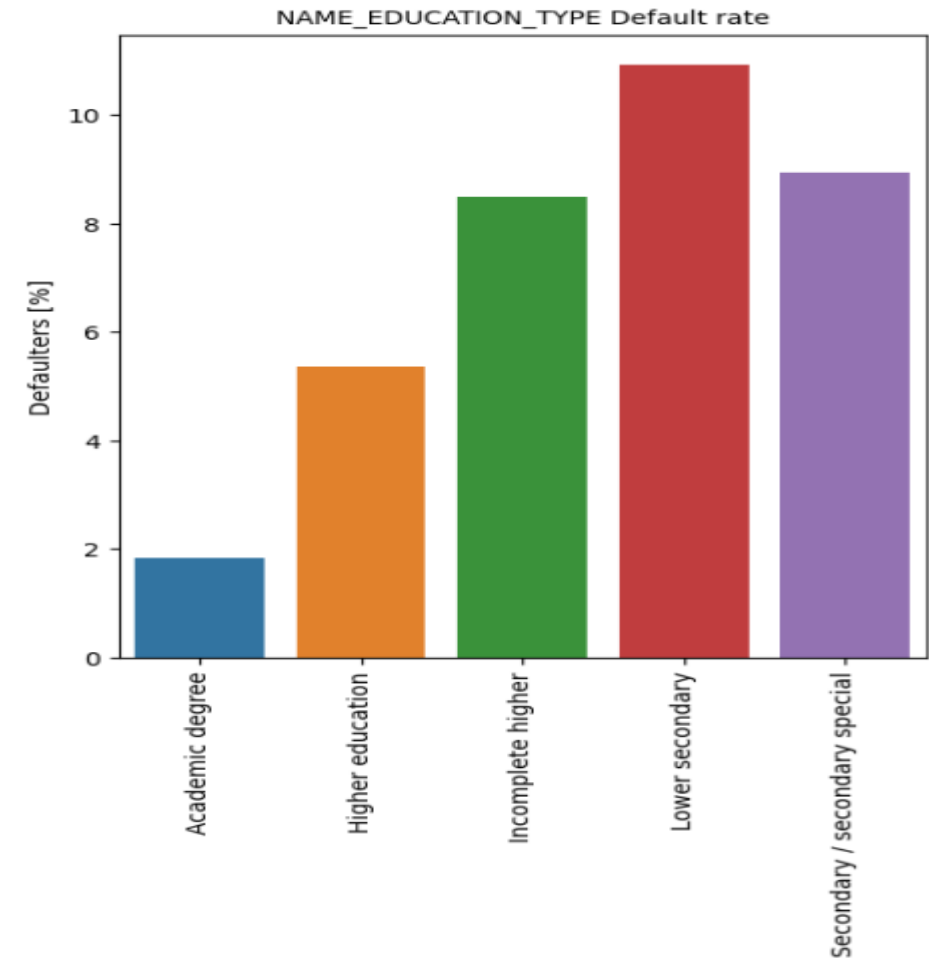
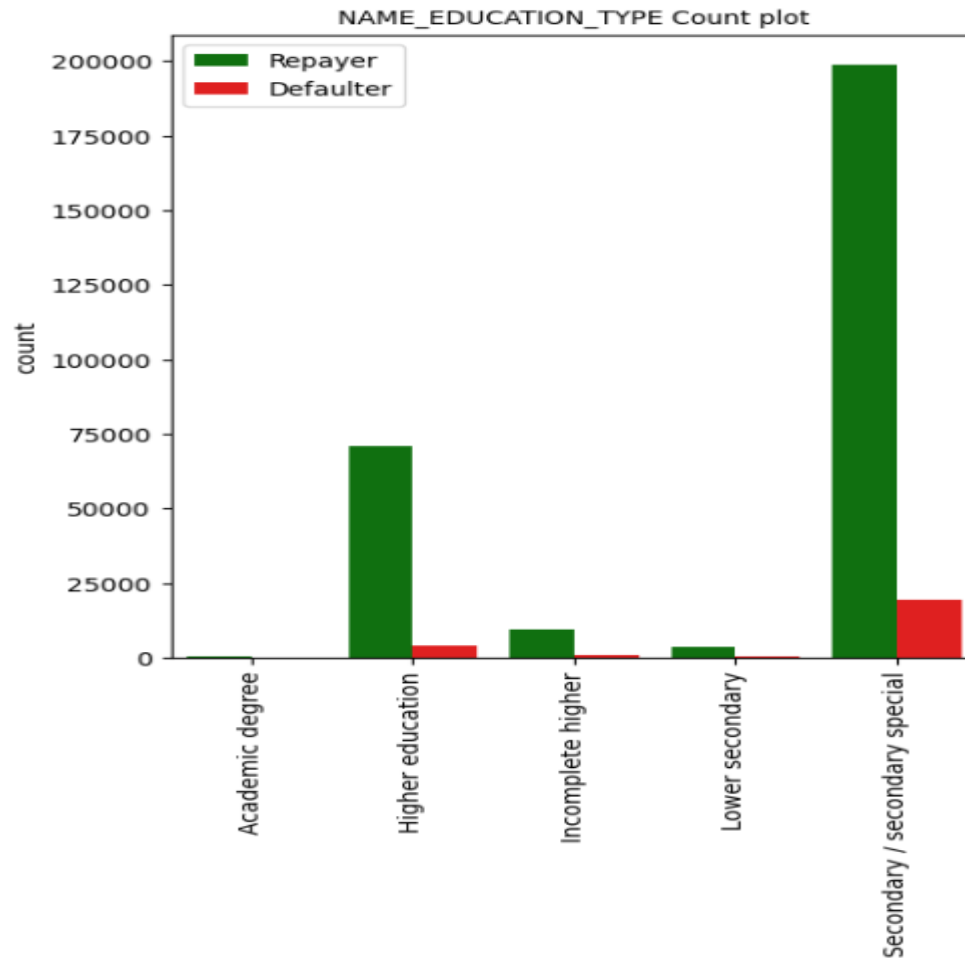
- Repay rate for Revolving loan is more [94%] as compare to Cash loans on a scale of 100 for each.
- Coming to defaulters on a scale of 100, cash loan have more defaulters almost 8 out of 100 cash loans, whereas for revolving loan its its almost 5 percent out of 100 revolving loan.

# Key Finding and Results [Age\_Range vs TARGET]



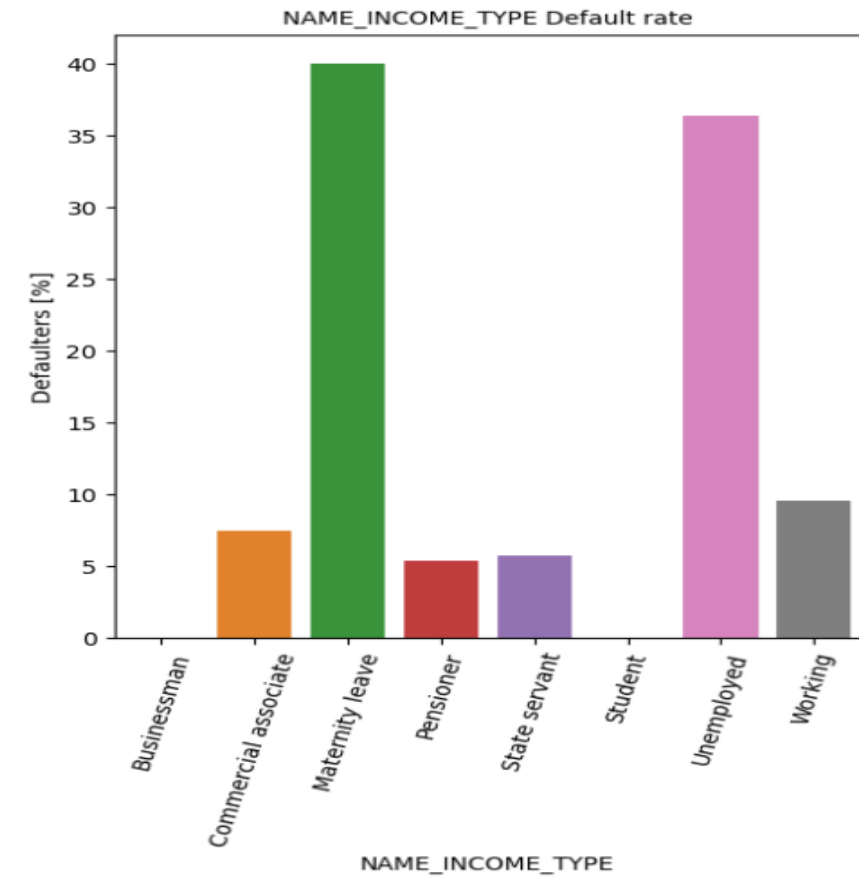
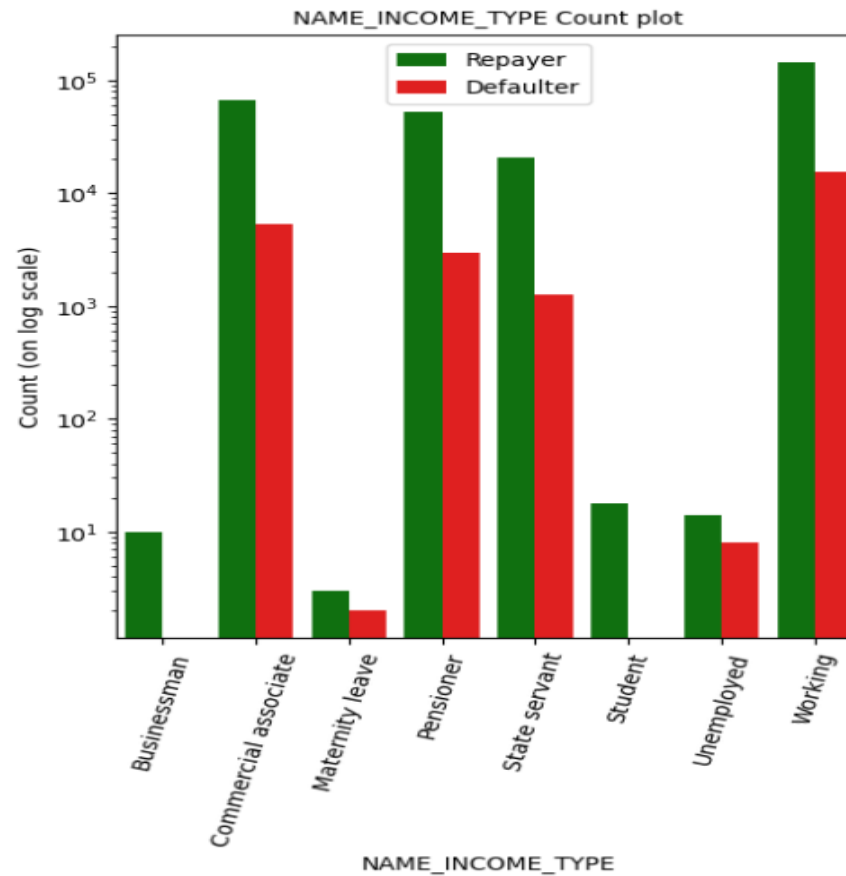
- Applicants with age about 60 and above tends to make few default (as their default rate is about 5 %)
- Younger section (10-20) have no default record (although very few in number)

# Key Finding and Results [NAME\_EDUCATION\_TYPE VS TARGET]



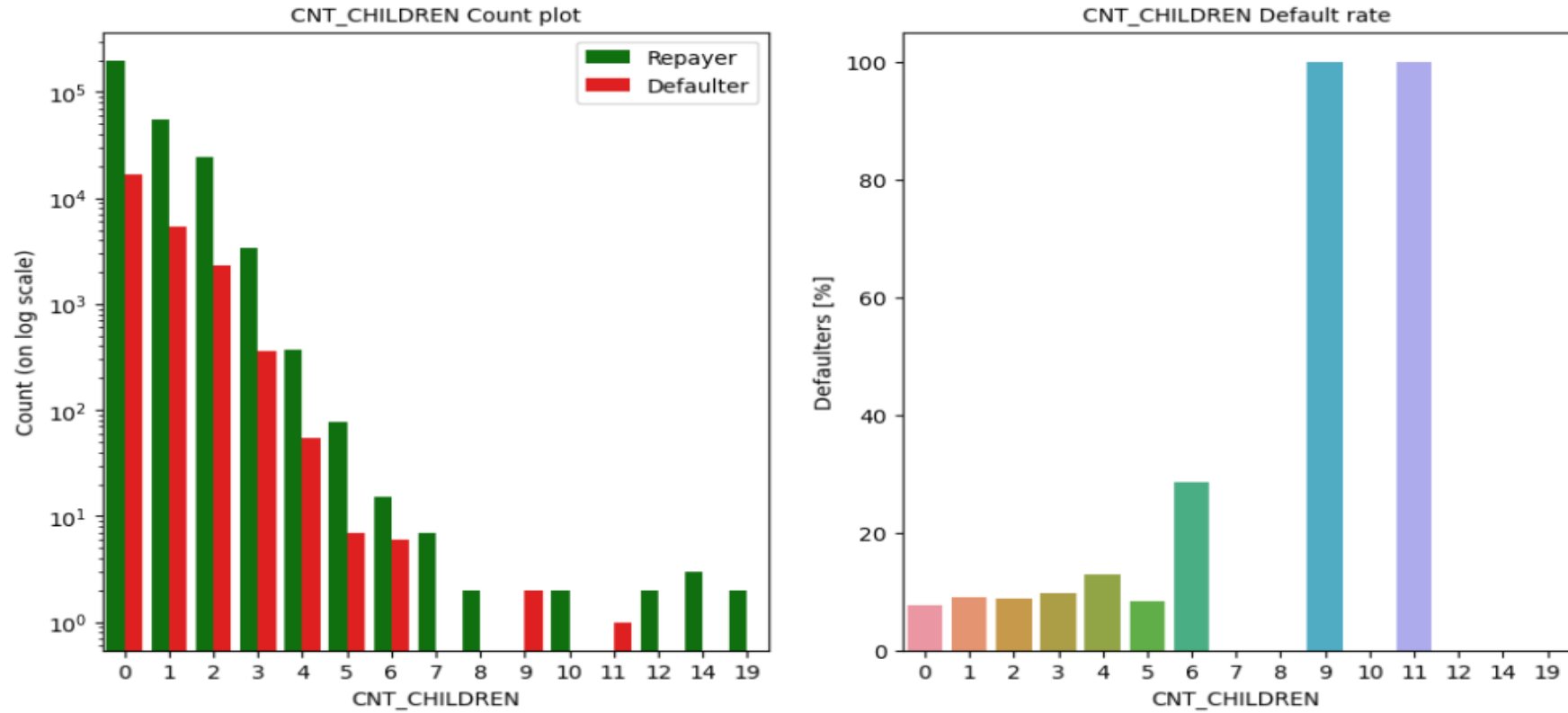
- Applicant with academic degree (although comparatively few counts) have very less default rate (2%)

# Key Finding and Results [NAME\_INCOME\_TYPE VS TARGET]



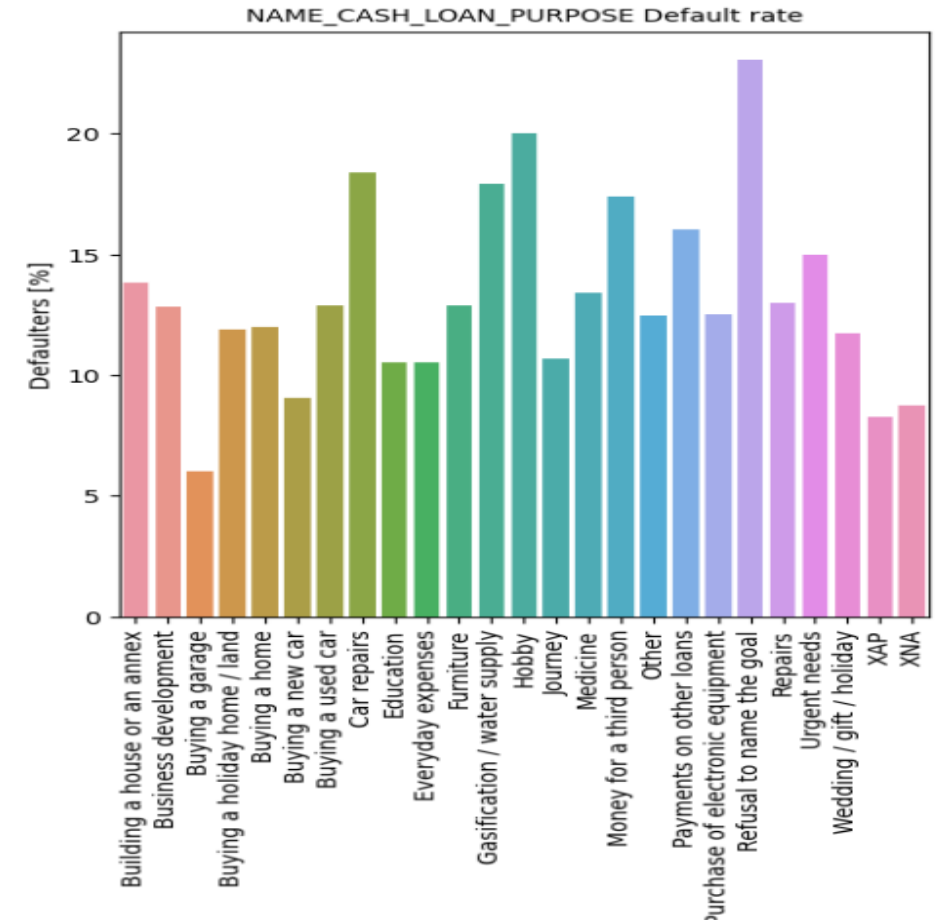
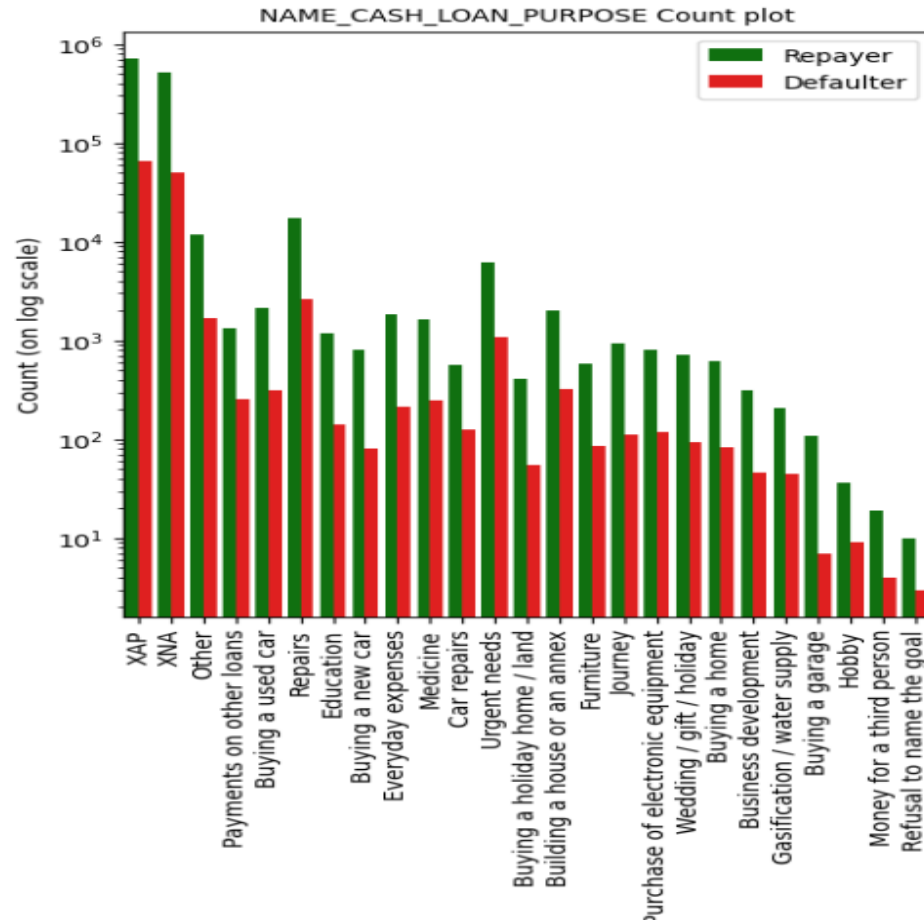
- Applicants having income type of Businessman and Student are making no default.
- Applicants with Maternity Leaves are among the highest with default rate of 40

# Key Finding and Results [ CNT\_CHILDREN VS TARGET]



- Seeing applicants with no children, they have high repay rate and less default rate as compared to applicants with more than 0 but less the 6 children count.
- its interesting to see, applicant with children count of 12, 14, 19, 7 and 8 are all repayers and with 9 and 11 are all defaulters.

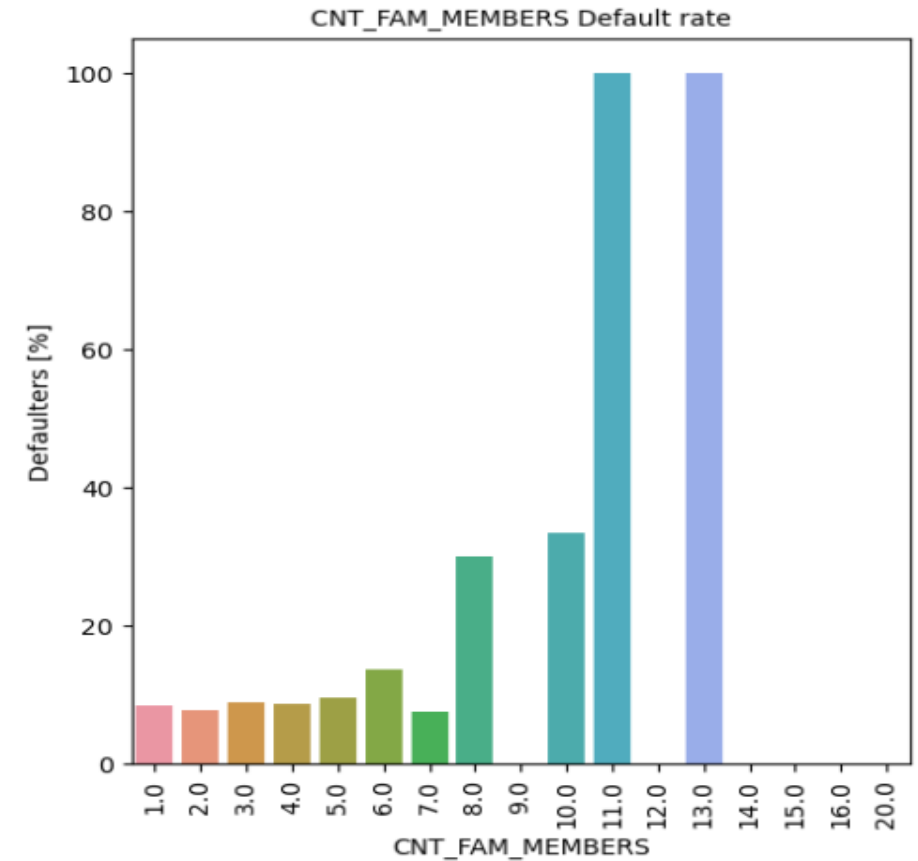
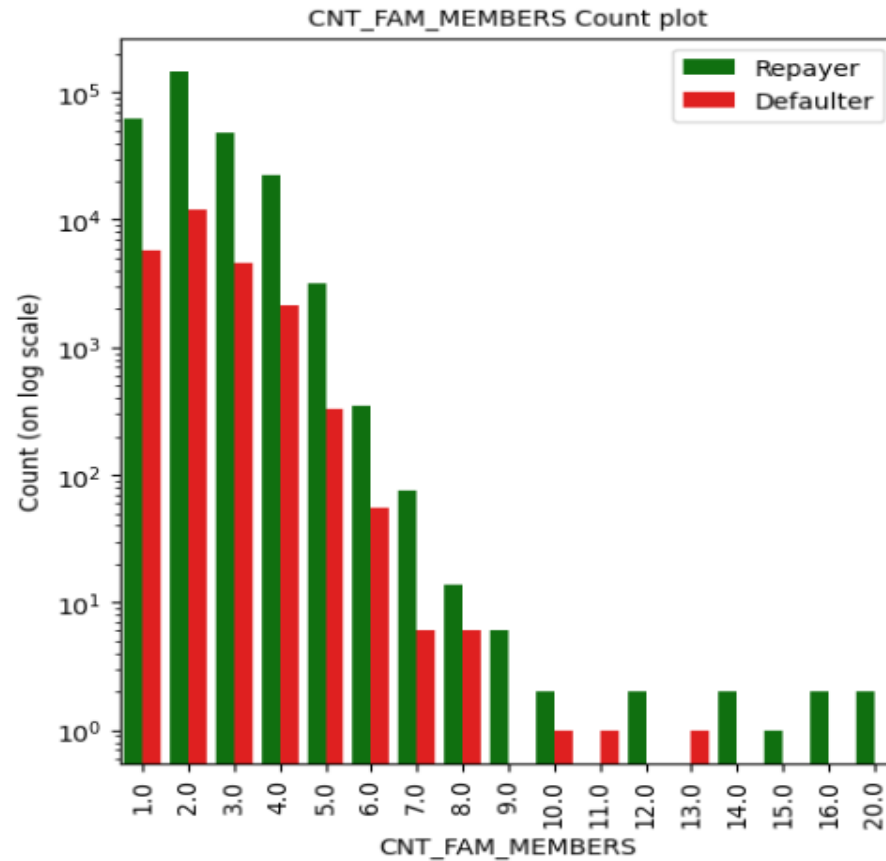
# Key Finding and Results [NAME\_CASH\_LOAN\_PURPOSE VS TARGET]



- people taking loan for repair or other purpose is among the most applied loans
- loan took for Buying garage have lesser default rate (6%)

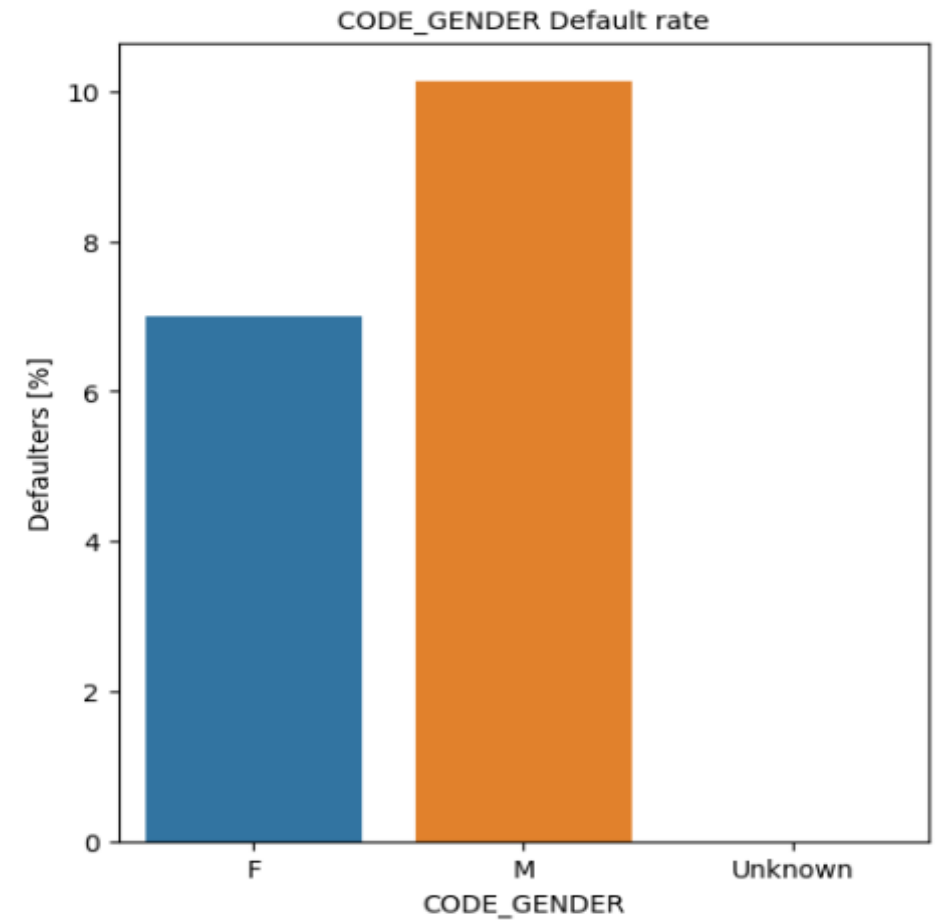
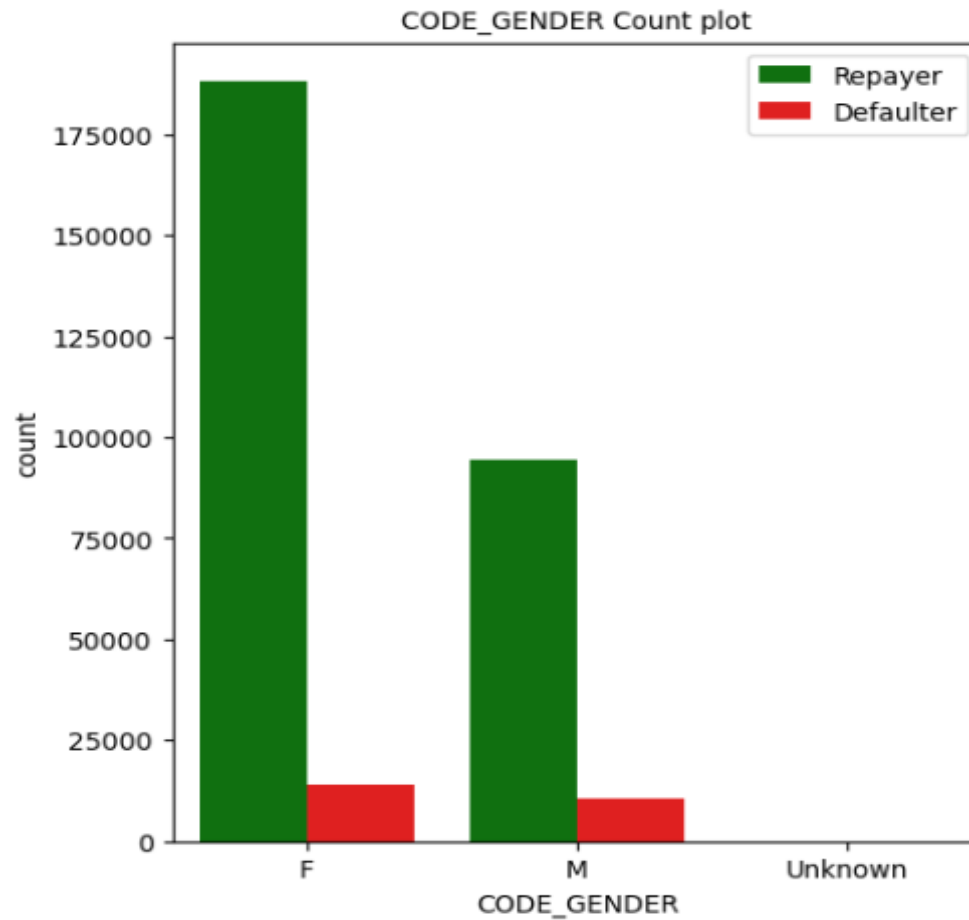


# Key Finding and Results [CNT\_FAM\_MEMBERS VS TARGET]



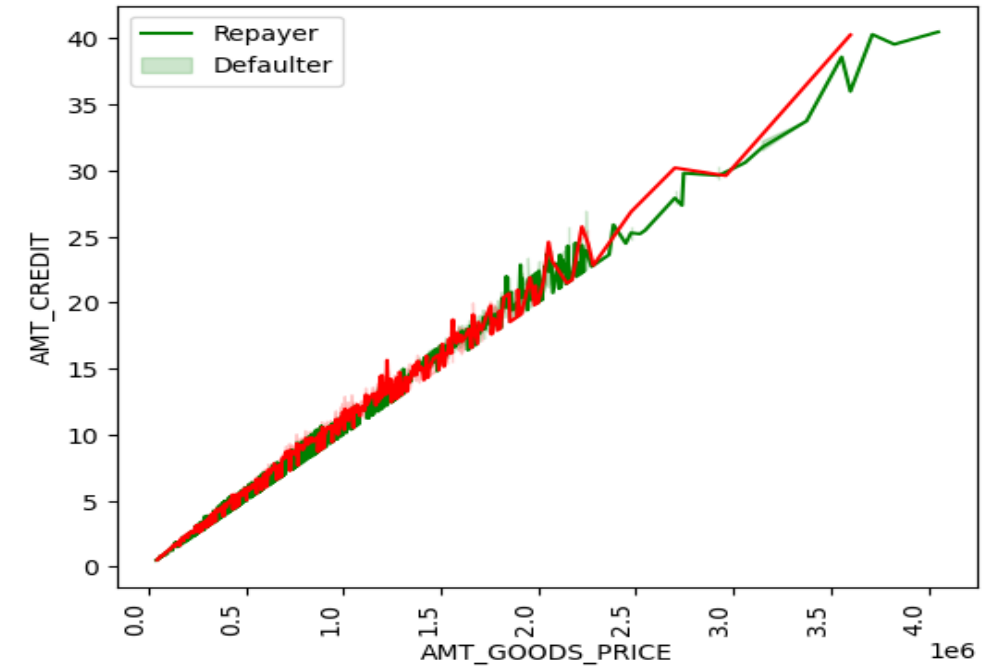
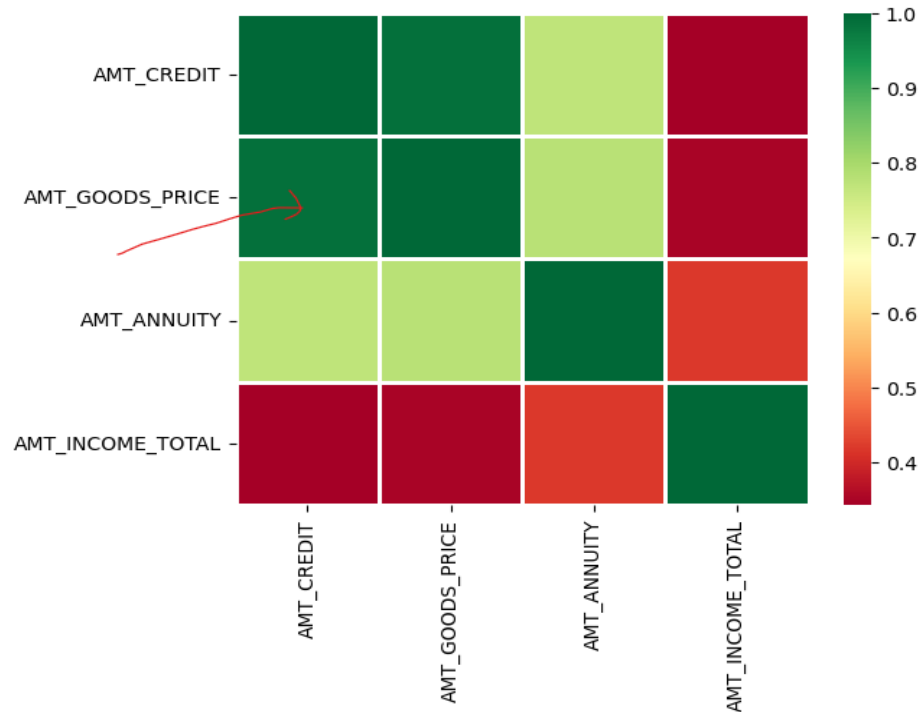
- applicants having family member around 11 or 13, have high default rate.
- default rate is less among people with family members between 1 to 7

# Key Finding and Results [CODE\_GENDER VS TARGET]



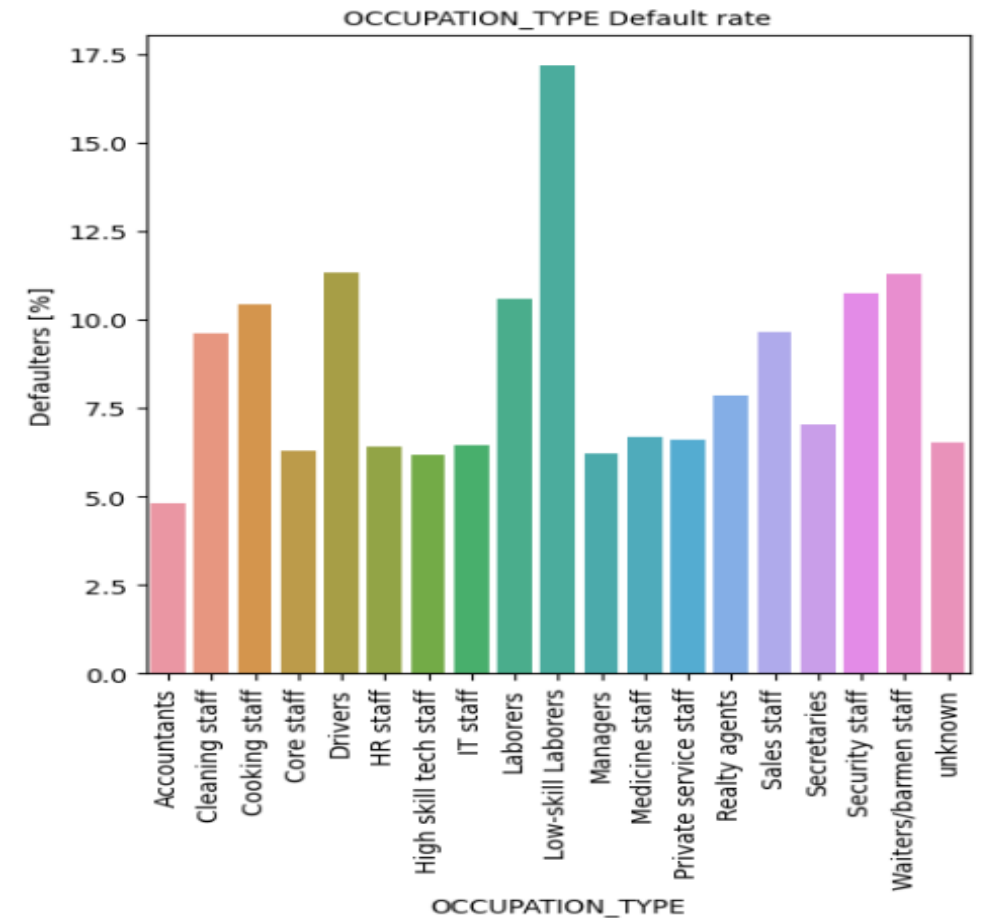
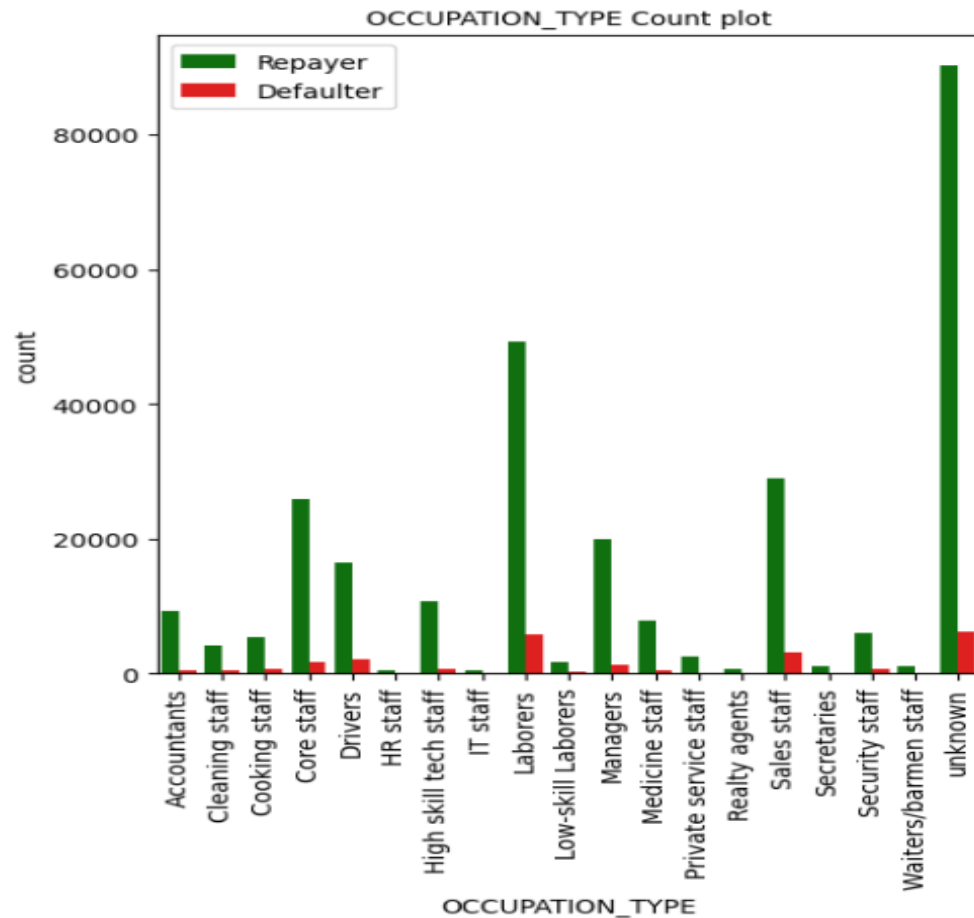
- Female applicants overall performing well as a repayer (93%), whereas males tends to be more defaulters (10%)

# Key Finding and Results [AMT\_GOOD\_PRICE VS AMT\_CREDIT FOR TARGET]



- There is high correlation between AMT\_CREDIT and AMT\_GOOD\_PRICE, hence creating lineplot might give us some insights **NOTE:** Scatter plot not used because of overlapping issue with it.
- Good price with value of 30 lacs or above, people are defaulting more as compared to repayers.

# Key Finding and Results [OCCUPATION\_TYPE VS TARGET]



- low skill laborers tends to default more as compared to other occupation type (almost 17%)
- accountants have high repay rate and less default rate

# Outcomes

## ■ Driving variable or feature for potential Repayer

- **Age\_Range**: (derived from **DAYS\_BIRTH**) 60 and above aged applicant have less default rate (4%)
- **NAME\_CONTRACT\_TYPE**: Kind of loan provided, plays an important role among applicants. Revolving (although less popular among applicants) repay rate is high and default rate is less as compared to cash loan.
- **NAME\_EDUCATION\_TYPE**: Applicants having academic degree makes less default (default rate is close 2%)
- **NAME\_INCOME\_TYPE**: Applicants having income type of Businessman and Student are making no default.
- **CNT\_CHILDREN**: Applicants with no children are most interested section of applicants when it comes to loans. With high repay rate and and less default rate
- **NAME\_CASH\_LOAN\_PURPOSE**: loan took for Buying garage have lesser default rate (6%)
- **CNT\_FAM\_MEMBERS**: small family applicants (1-7 members) tends to default less.

## Outcomes cont.

### ■ **Driving variable or feature for potential Defaulters**

- **CODE\_GENDER**: males rate of default is quite high and tends to default more
- **NAME\_EDUCATION\_TYPE**: applicants with Lower secondary education default rate is significantly high.
- **NAME\_CONTRACT\_TYPE**: applicants applying for Cash loans tends to default more as compared to revolving loan
- **OCCUPATION\_TYPE**: low skill laborers have high default rate
- **Age\_Range**: (derived from DAYS\_BIRTH) applicant with age between 20 -30
- **CNT\_FAM\_MEMBERS**: applicants having huge family (11, 13 members) default rate is noticeable high.

# Outcomes cont.

## ■ Suggestions

- Applicants having big family tends to default more, therefore high interest rate could mitigate the risk of defaults
- Applicant applying for loan with good price of 30 lacs or more, bank should impose high interest as they tends to default more.
- Student and Business makes no defaults so far, but also few in number. Providing better offers, ease of applying and better interest might attract more students and businessman to apply for loan.
- Applicant who were refused in previous applicant, came out to be repayer in current application. Hence a proper refusal reason is need to increase the loan company business
- Applicants with revolving loan have better profile with high repay rate and less default rate. A proper understanding of client need and meets might interest them to opt for revolving loan. Which further might decrease the default rate and more potential applicants
- applicant children count above 14 seems to be having no defaults. Banks should looks those people more closely. Which might increase the business.



Thank you