Credit EDA Assignment

(LOAN CASE STUDY)

Submitted By:

Anumeha Sinha

PROBLEM STATEMENT

- The loan providing companies find it hard to decide if the applicant should be given loan or not?

- Risks involved is to correctly identify the applicants who are capable of repaying and who are defaulters.

- Decision is based on :

  1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.

  2)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.

Through this analysis the company wants to analyze and understand the reasons behind the defaults. Later company can use this analysis to identify the right applicant for loan.

ANALYSIS APPROACH

Understanding Problem Statement and Data → Data Cleaning → Univariate Analysis

Analysis/Conclusions ← Bivariate and Multi Variate Analysis

ANALYSIS APPROACH

1) Understanding Problem Statement and Data :
Understanding the business problem and analyzing it. Going through the dataset and understanding briefly the columns. And rows

2) Data cleaning :
There could be different issues within the datasets like missing values, outliers, incorrect format , inconsistent spelling. Fixing these issues in columns and rows by either dropping or filling relevant data ,changing the format or correcting the spelling are few fixes.
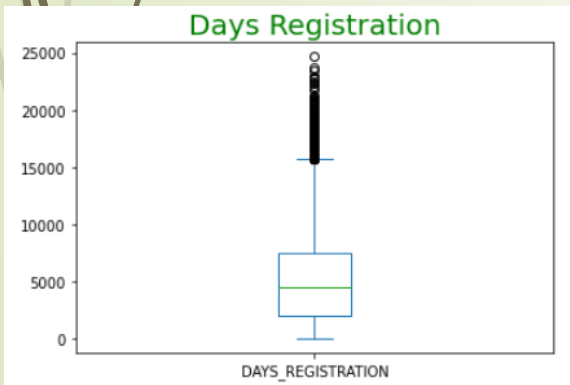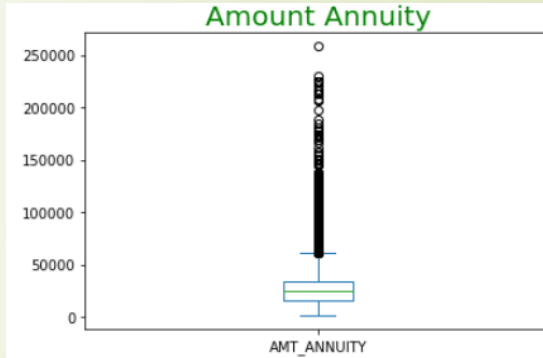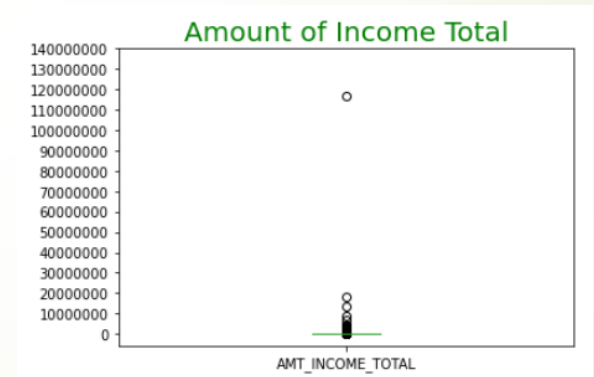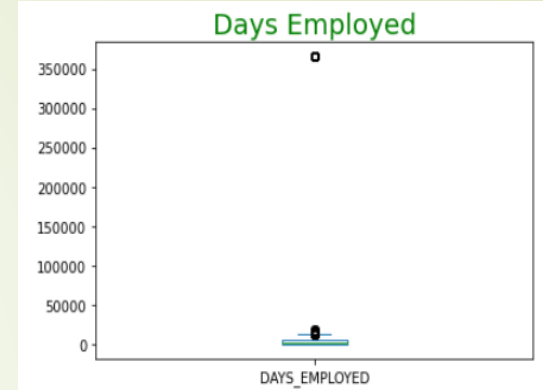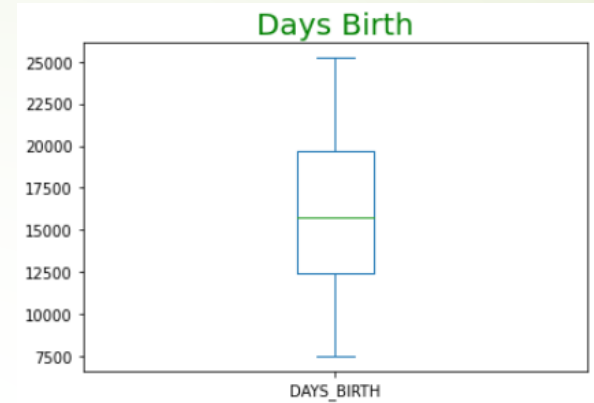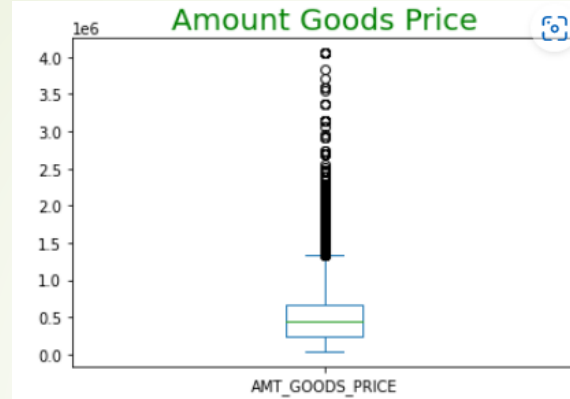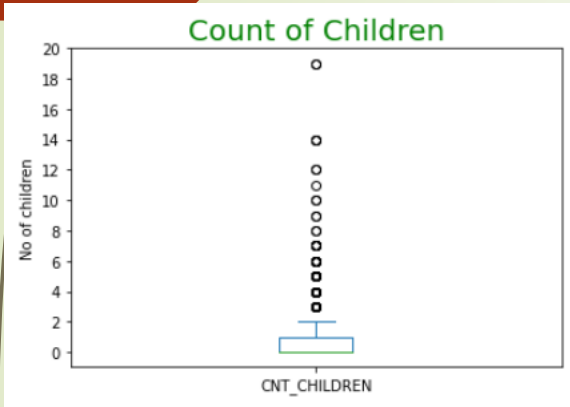
3) Univariate Data Analysis :
Univariate analysis involves the analysis of a single variable at a time.

4) Bivariate & Multivariate Data Analysis :
It involves analysis of 2(Bivariate)or more than 2(Multivariate) variables.

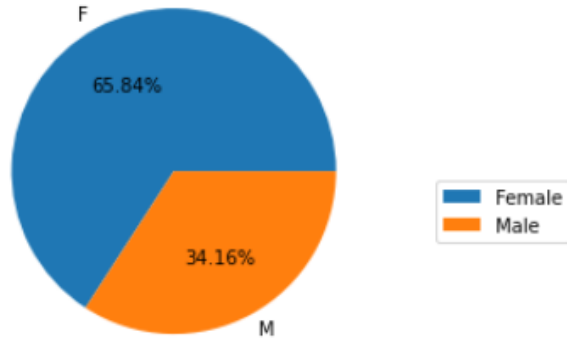# DATA CLEANING - OUTLIER CHECK
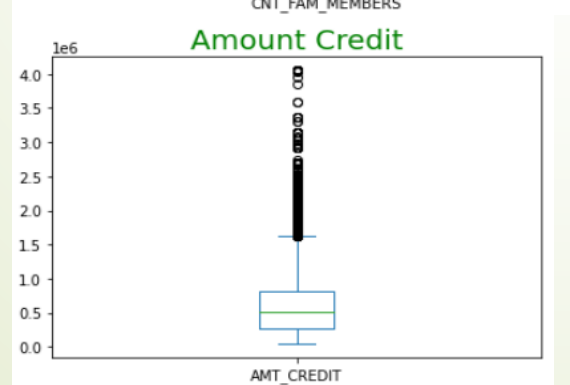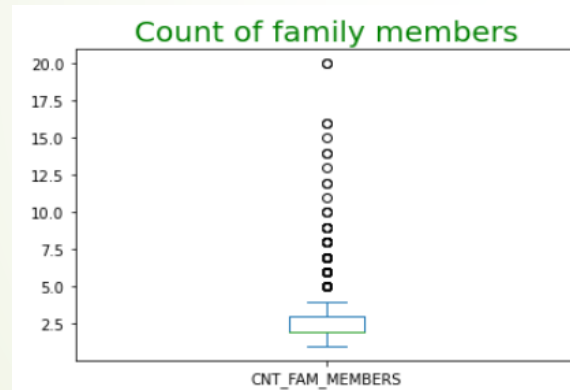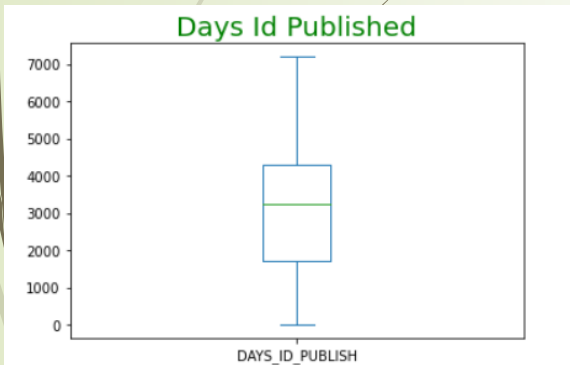
Observation:

Outlier Present :	No Outlier :
DAYS_EMPLOYED	DAYS_BIRTH
CNT_CHILDREN	DAYS_ID_PUBLISH
AMT_INCOME_TOTAL	
AMT_CREDIT	
CNT_FAM_MEMBERS	
AMT_GOODS_PRICE	
DAYS_REGISTRATION	
AMT_ANNUITY	

UNIVARIATE ANALYSIS

Distribution of Gender

F
65.84%

M
34.16%

Legend:
- Female
- Male

Contract type

Cash loans
90.48%

Revolving loans
9.52%

Legend:
- Cash Loan
- Revolving

Applicant owns car

N
65.99%

Y
34.01%

Legend:
- No
- Yes

Applicant owns realty

Y
69.37%

N
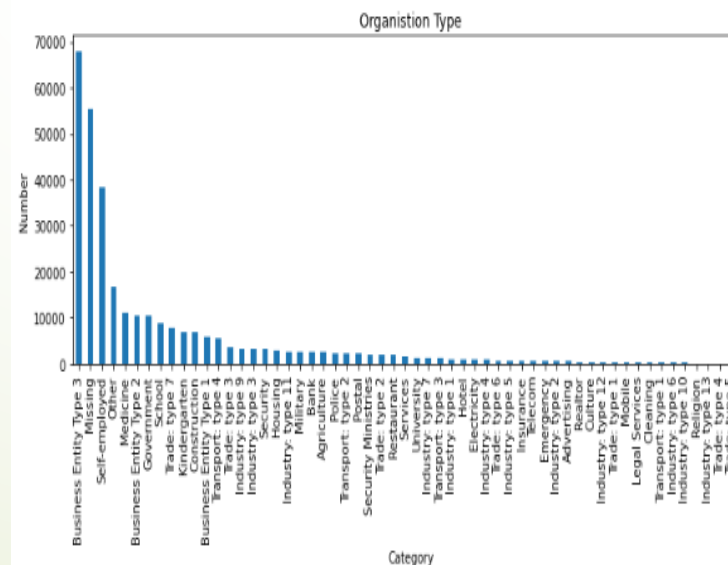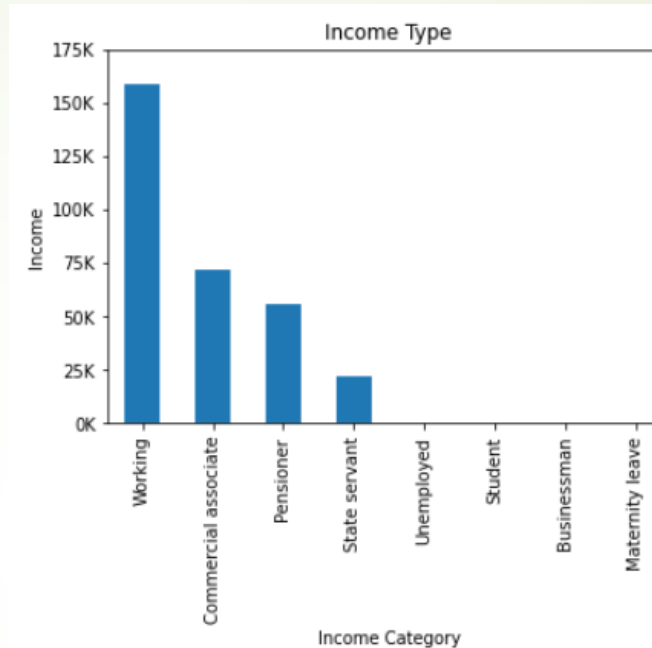30.63%

Legend:
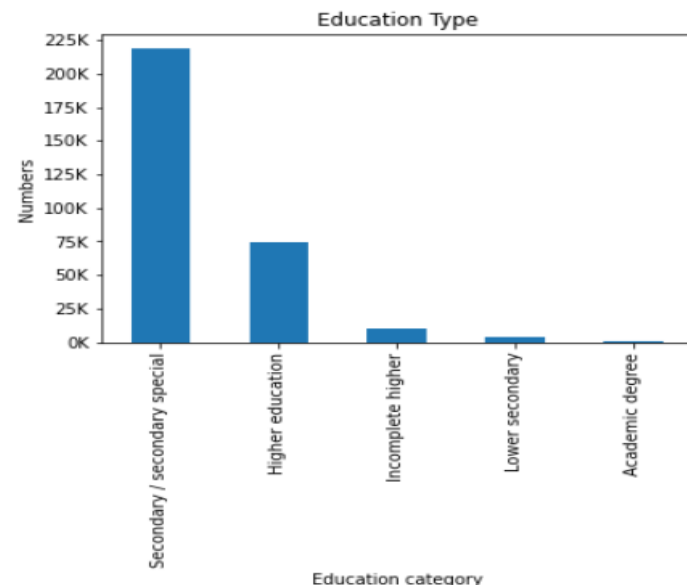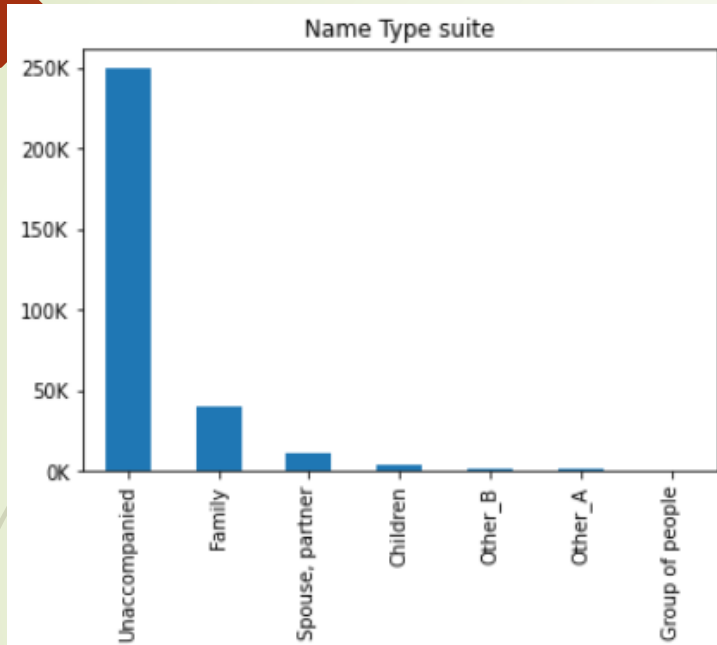- Yes
- No
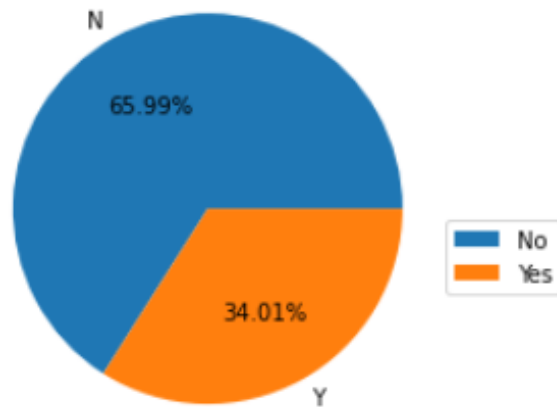
OBSERVATIONS :

Percentage of Female applicants are more as compared to male

Majority are interested in Cash loan and only 9.52% are interested in Revolving loan

Only 34% applicant owns car

Applicants who own realty are almost 70%
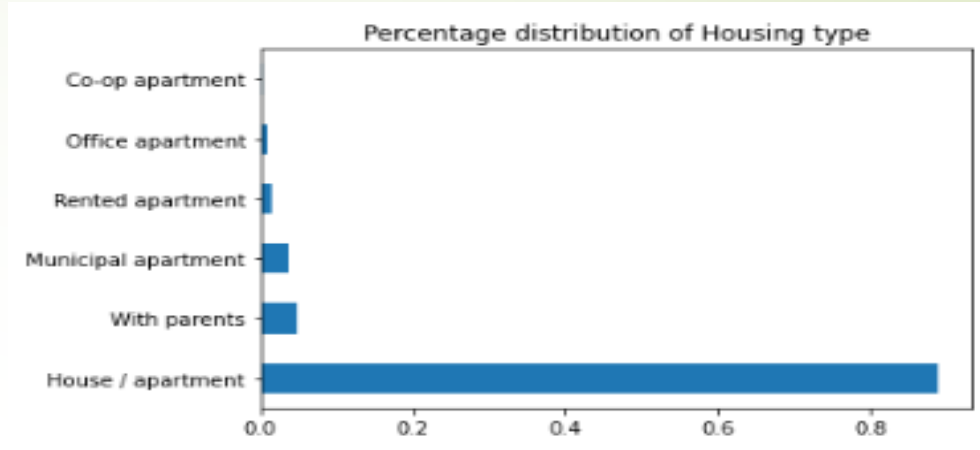
UNIVARIATE ANALYSIS

OBSERVATIONS :

Applicants are mostly Unaccompanied and there are very less applicants taking loans coming in a group

The applicants applying for loans are mostly - Working professional, Commercial Associate and Pensioner

Applicants applying for loan are mostly having Secondary special education type.

Mostly applicants belong to "Business Entity Type 3" category or they have not provided their business type

UNIVARIATE ANALYSIS
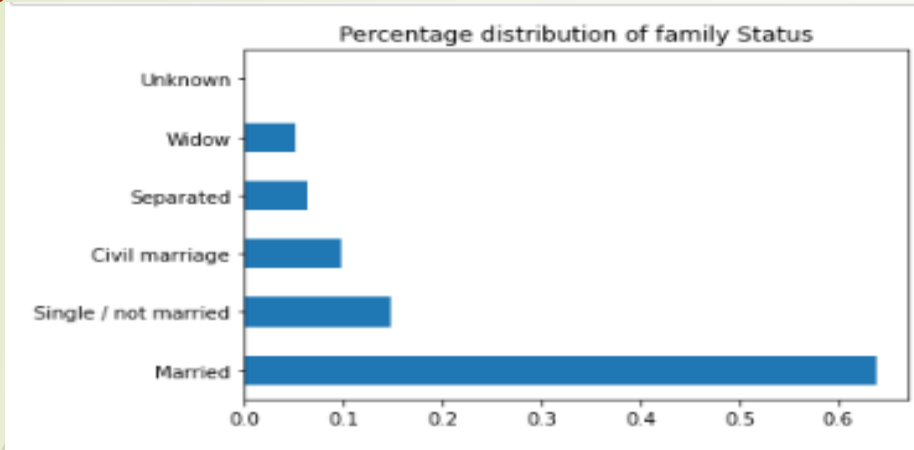
OBSERVATION :

- More than 60% applicants applying for loan are married

- More than 80% applicants applying for loan have House/Apartment
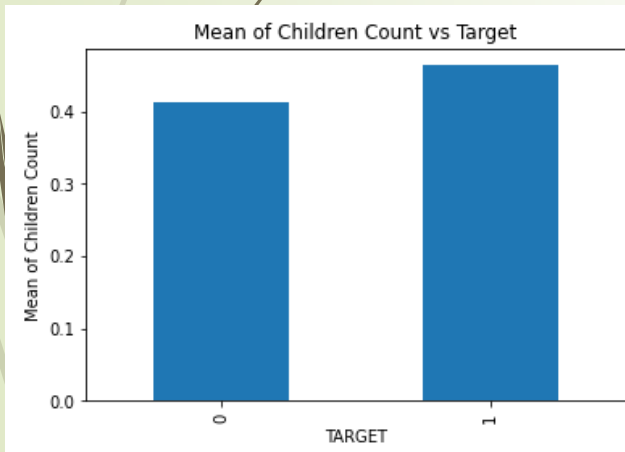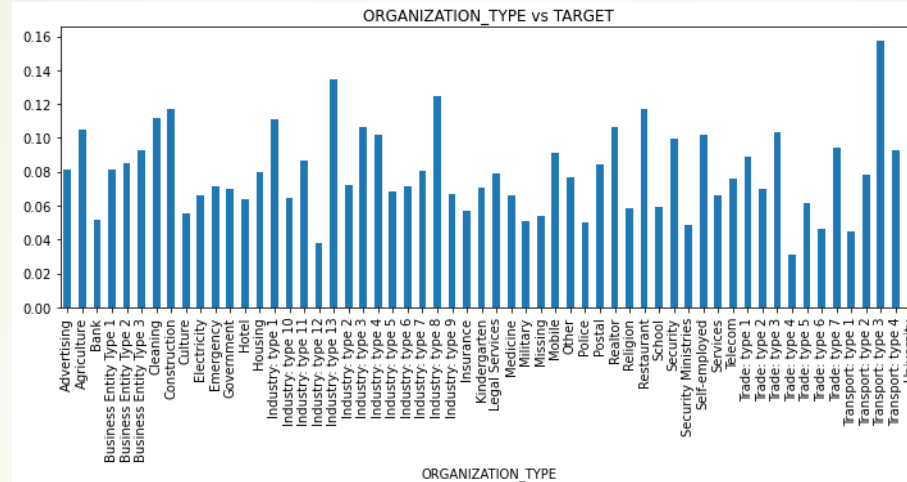
BIVARIATE & MULTIVARIATE ANALYSIS

OBSERVATIONS :

Age 60+ has less chances of defaulting

Applicants with Organization type Transport type 3 are defaulting the most

With increase in count of children default rate goes up

Majority of the applicants are having children count below 6 and amount credit is not more than 30lacs but there are few exceptions
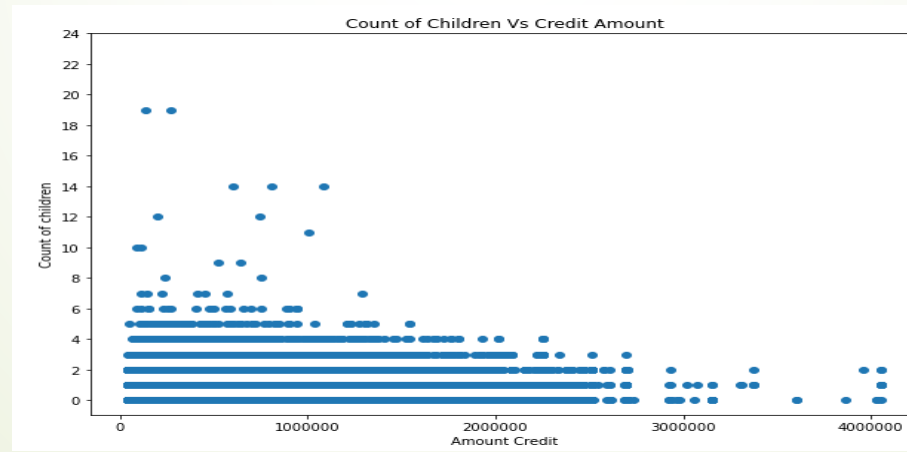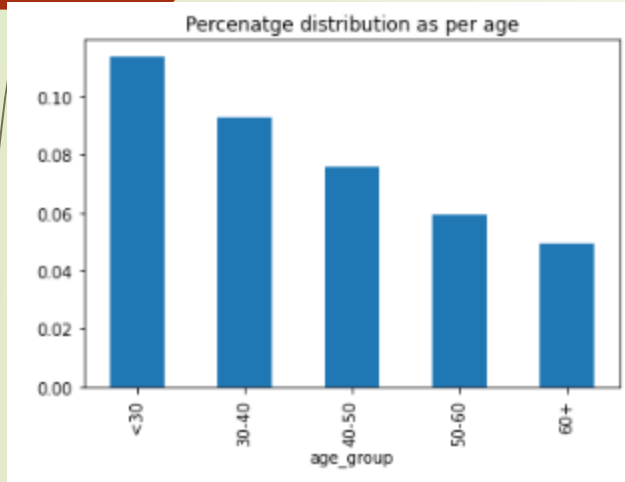
BIVARIATE & MULTIVARIATE ANALYSIS
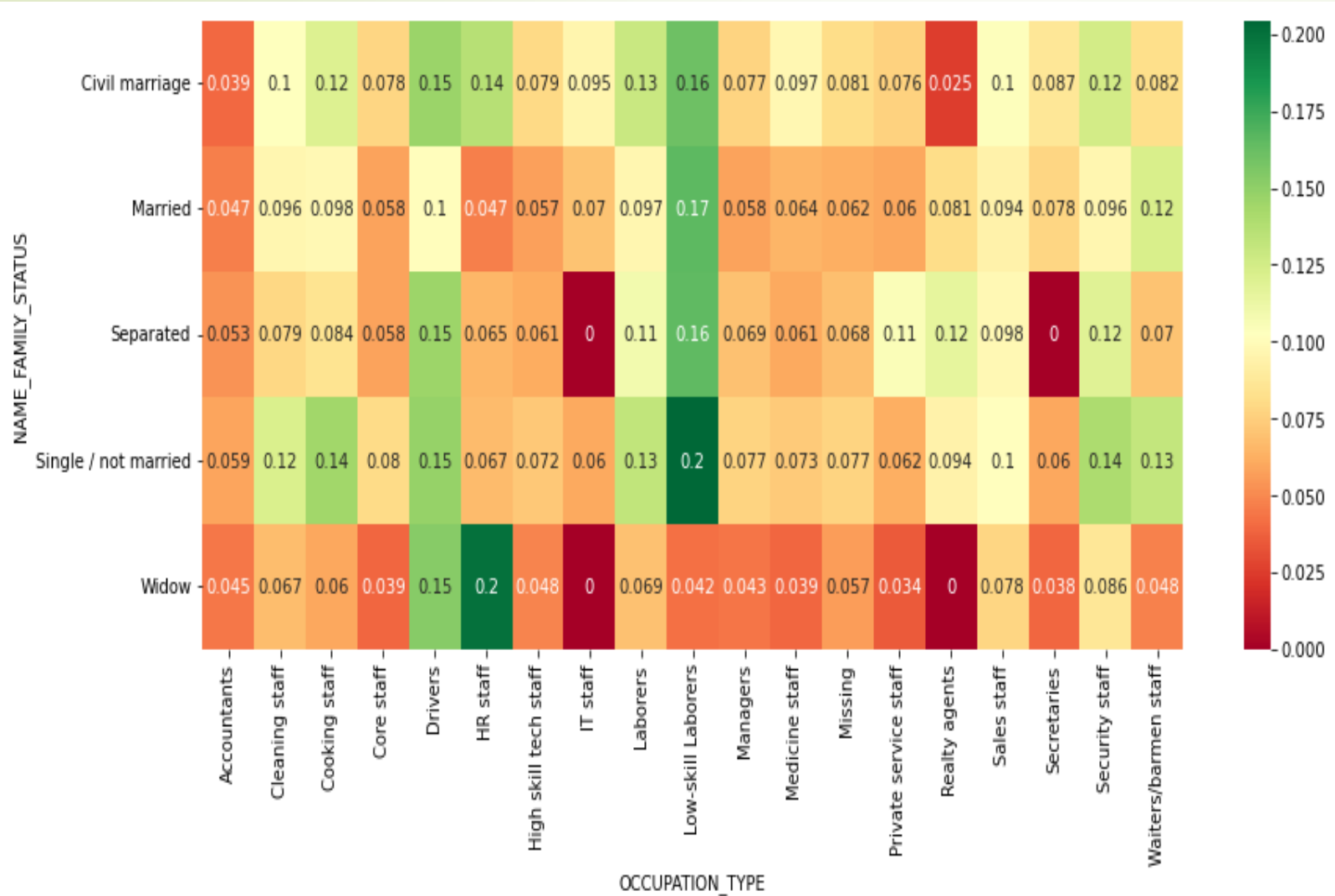
OBSERVATION :

- **Applicants doing Civil Marriage or Single Married have higher rates of defaults**

- **Laborers, Drivers, Waiters and other Low skilled Laborers are defaulting the most**

BIVARIATE & MULTIVARIATE ANALYSIS
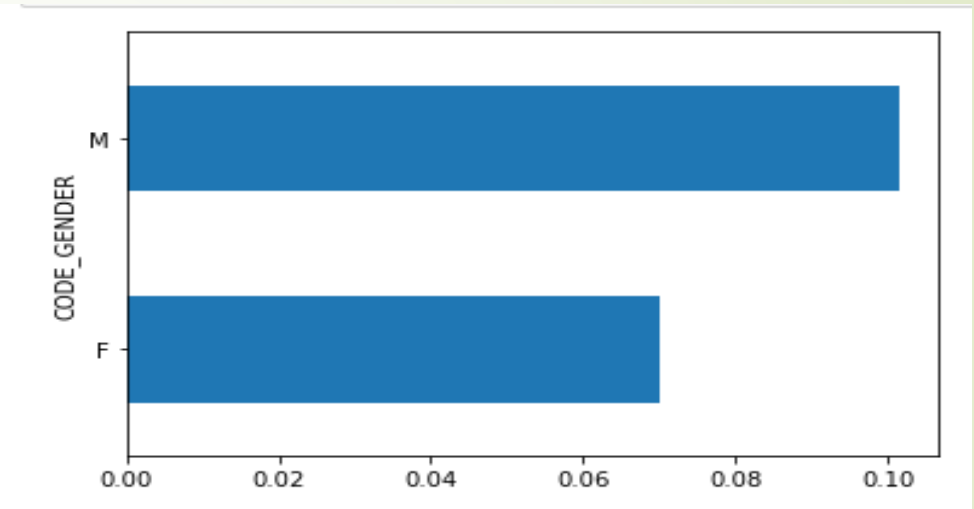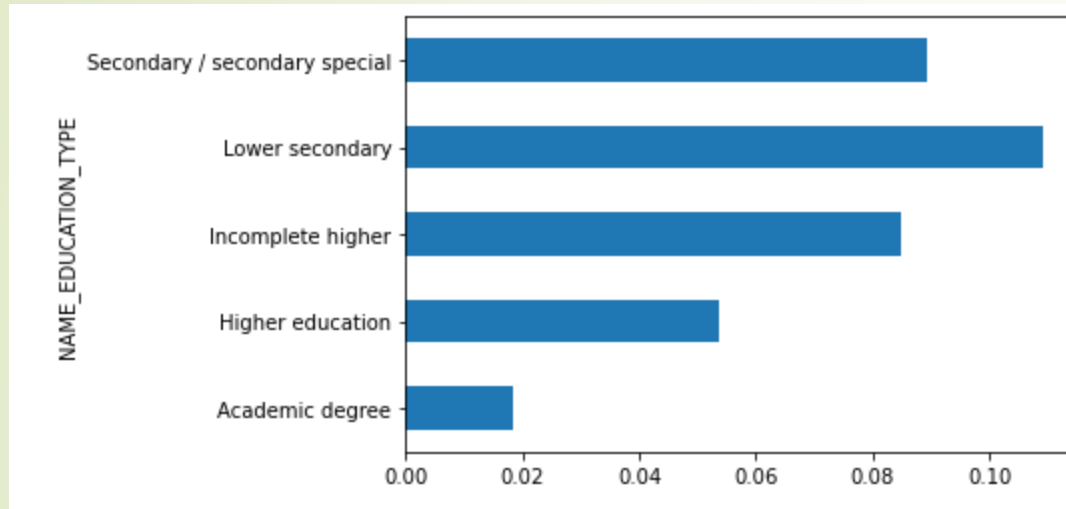
OBSERVATION :

- Male have higher rates of defaults as compared to Female
- Academic Degree/Higher Education have less defaults as compared to others

BIVARIATE & MULTIVARIATE ANALYSIS
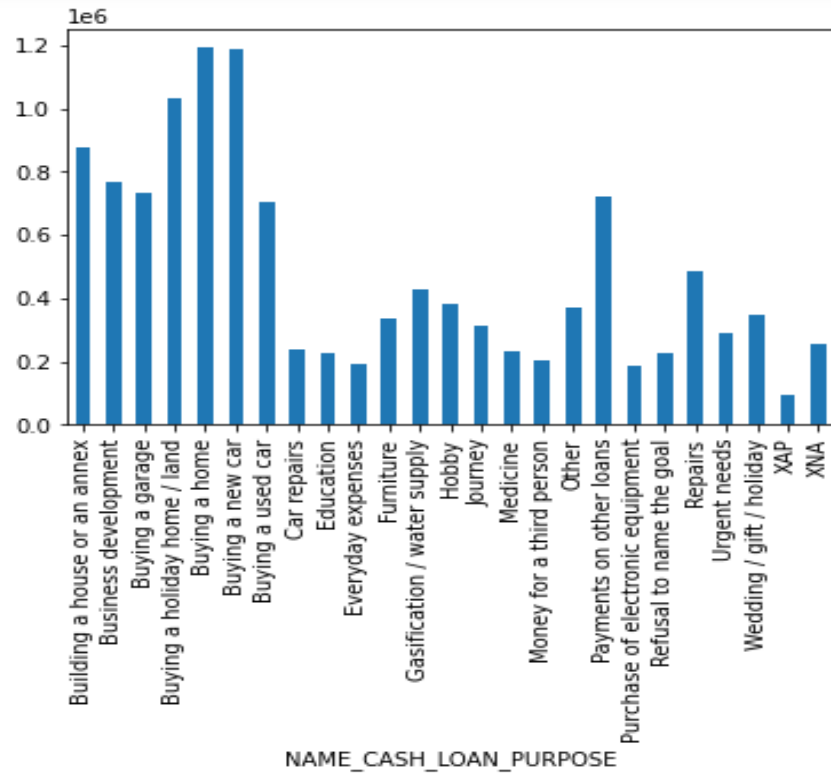
OBSERVATION :

- Applicants are taking loans mostly to buy a home or car
- Male defaulters are more as compared to Female defaulters

Target data balance check

TARGET data is imbalanced as Non Defaulters are dominating against Defaulters

CONCLUSIONS

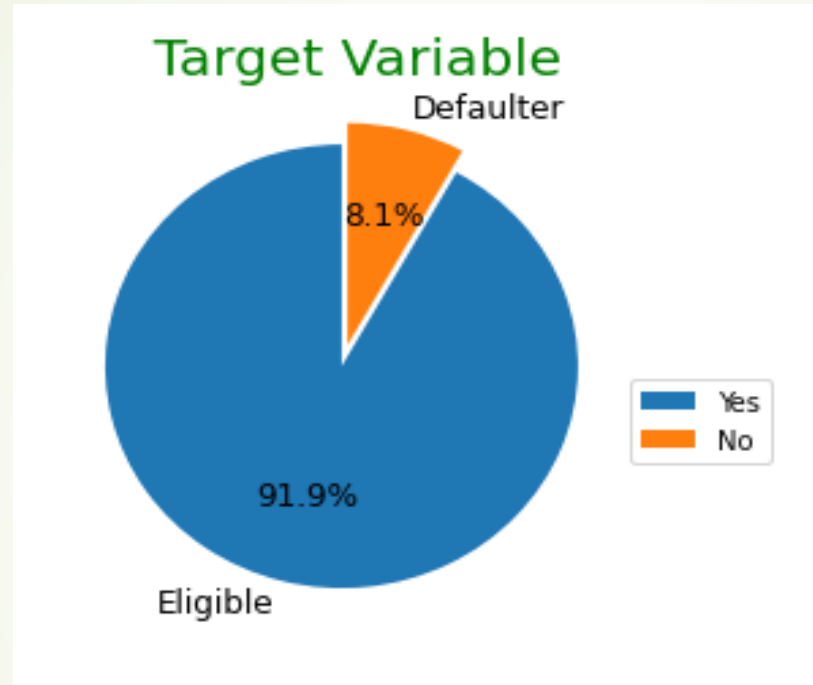| NON DEFAULTERS	DEFAULTERS
1. NAME_EDUCATION_TYPE : Applicants with higher education and academic degree are less likely to be defaulters	1. NAME_EDUCATION_TYPE - Applicants with incomplete education, secondary education or lower secondary education are high default rates
2. CODE_GENDER : Males are more likely to be defaulters as compared to Females	2. CNT_CHILDREN: With higher count of children default rate goes up.
3. AMT_INCOME_TOTAL : Applicants with higher salary are less defaulting	3. NAME_CASH_LOAN_PURPOSE : Applicants refusing to share the reason of loan or taking loan for Hobby are most defaulters
4. OCCUPATION_TYPE : Accounts, Managers , Core Staff, Secretaries, High Skilled Tech Staff, IT/Medicine/Private Staff are non-defaulter	4. OCCUPATION_TYPE : Laborers, Drivers, Waiters and other Low skilled Laborers are defaulting the most
5. NAME_FAMILY_STATUS : Married applicants are mostly non defaulters	5. NAME_FAMILY_STATUS : Single/not married , Civil Marriage and Widow have high default rates
6. ORGANIZATION_TYPE : Applicants in Industry 12 and Trade 4 are least defaulted among all	6. ORGANIZATION_TYPE :Applicants in Transport Type 3, Industry Type 13 and 8 have more than 10% rate of defaulting
7. DAYS_BIRTH : With increasing age chances of defaulting decreases	7. DAYS_BIRTH: Below 30 applicants have highest rate of defaulting