



Presentation on EDA Credit

Submitted By: Iqbal Rais

EDA Process steps

- Business Understanding
- Data Understanding
- Data Cleaning
 - Checking NA values
 - Checking for duplicates
 - Correct the data type
 - Binning
 - Finding Outliers
- Data Analysis
 - Define data imbalance percentage
 - Analysis to the business problem answer
- Merging of data
- Data Analysis
- Publishing

Business Objectives

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. This will reduce the financial loss of the company and will know the target customers for loan.

Data Understanding

application_data.csv: This data contains all the information of the client at the time of application. It contains the data whether a client has payment difficulties.

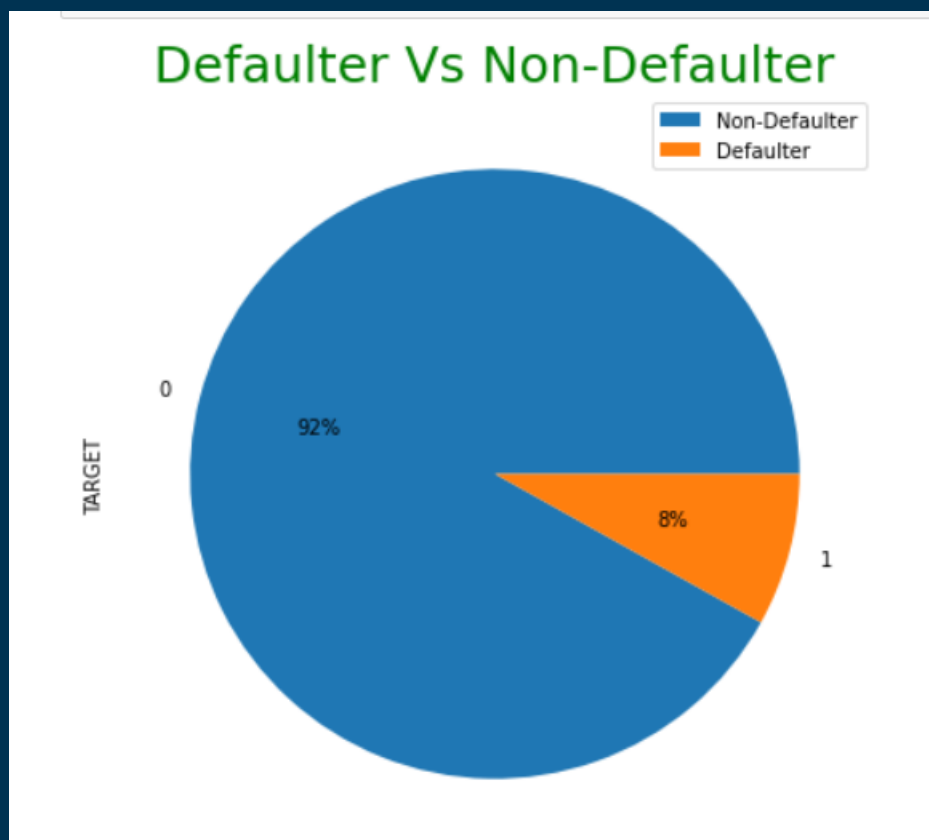
previous_application.csv: This data contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

Data Cleaning

- Dropping the data having missing values more than 40%.
- Filling the data having missing values less than 40% with mean/median for numerical values and mode for categorical values but in **Occupation** column missing values are replaced by missing as there are a lot of missing values.
- Dropping the rows have missing values close to 0%.
- Data Binning on following columns
 - Age Group
 - Income Group
 - Credit Group
- Following are the variables having outliers
 - Children
 - Total Income
 - Goods Price
 - Family Members

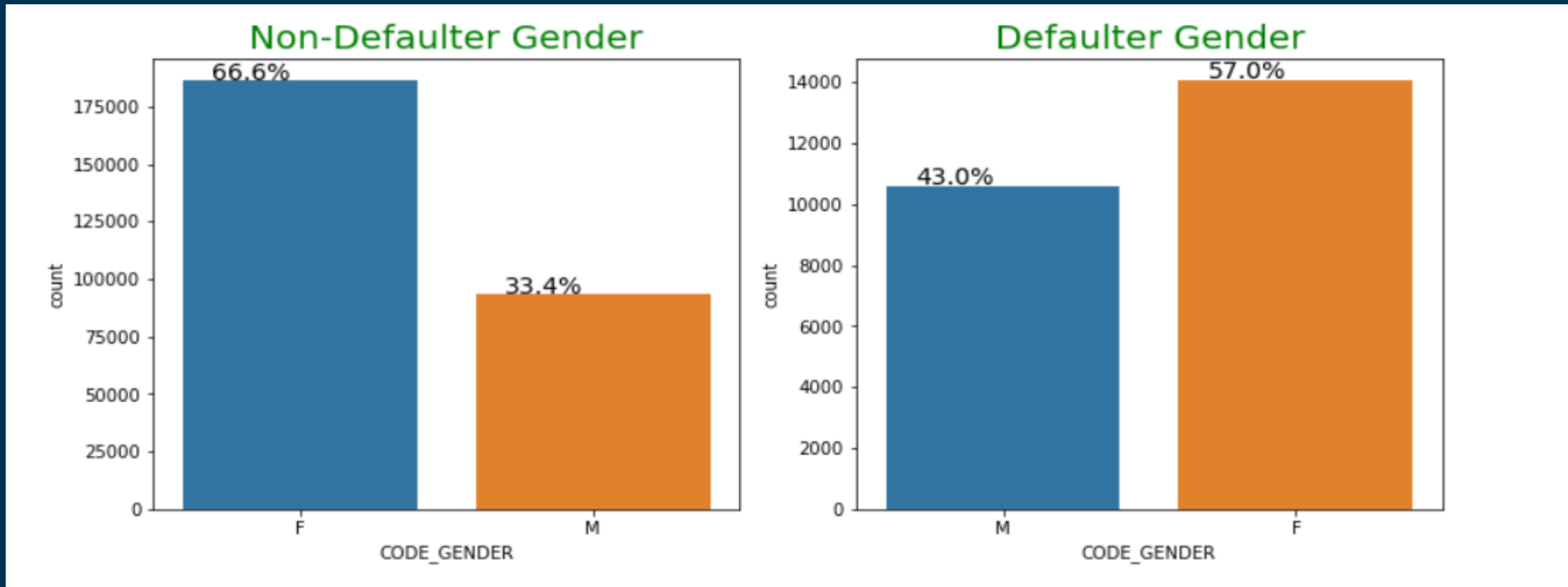
Analysis

Imbalance Percentage



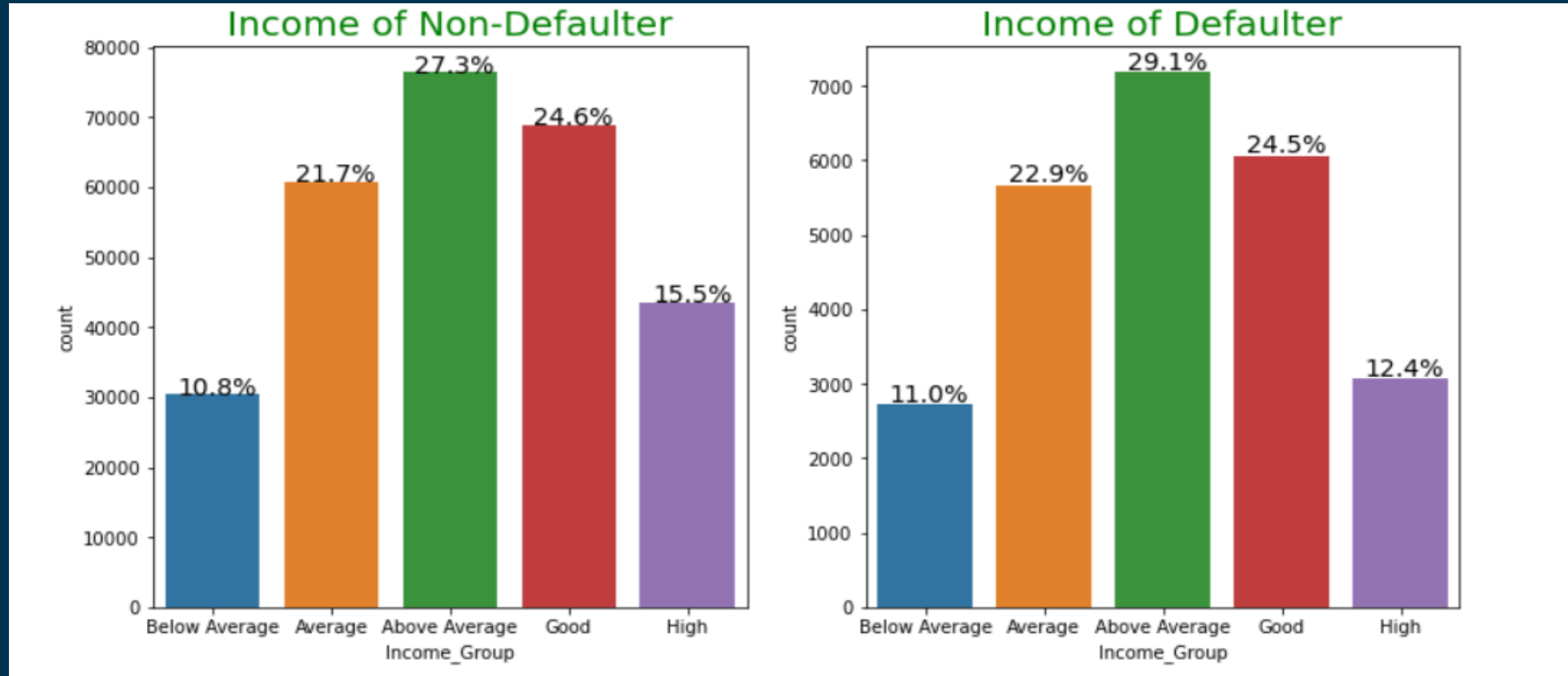
92% of people are Non - Defaulter and only 8% people are Defaulter.

Gender



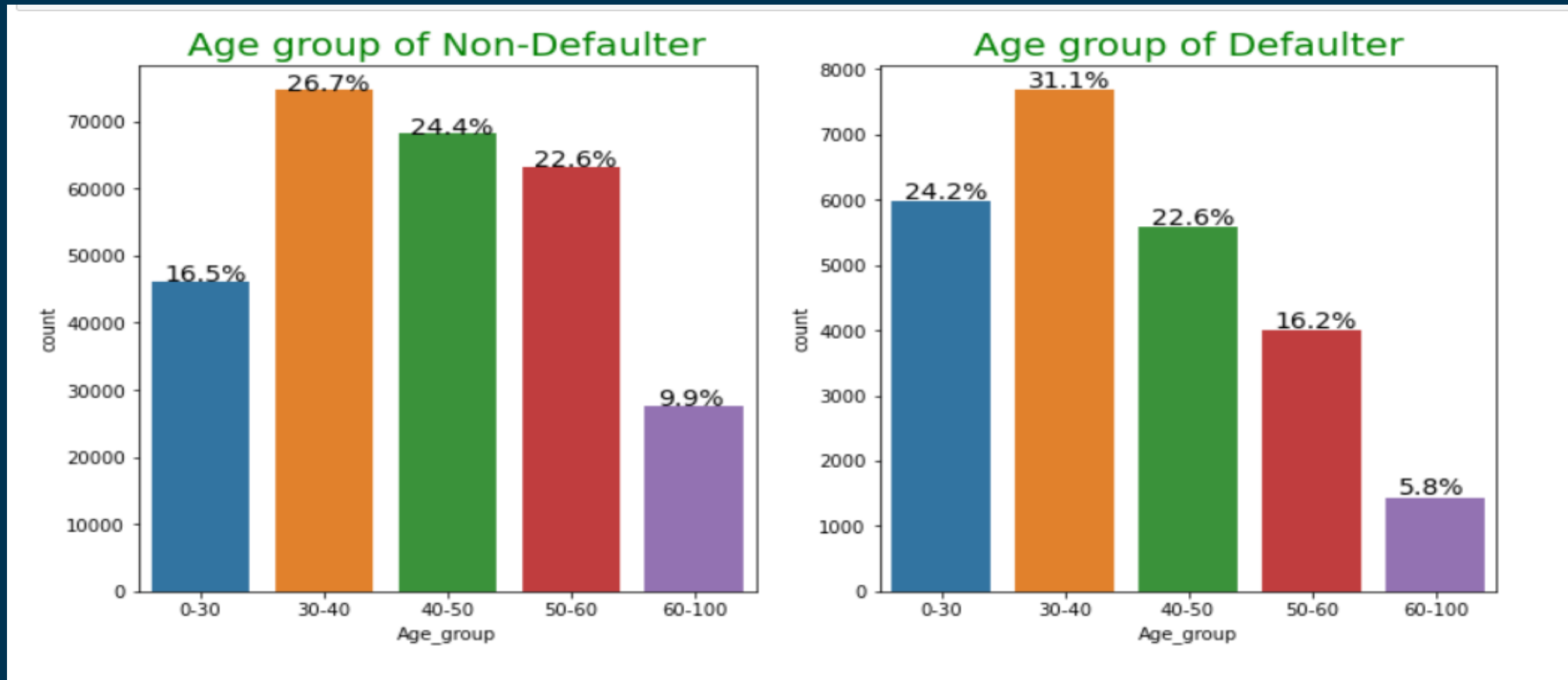
Male are more defaulter as percentage is increased as compared to Non defaulter

Income



People having Above average income are higher as a defaulter and people having High income are less chance to be a Defaulter.

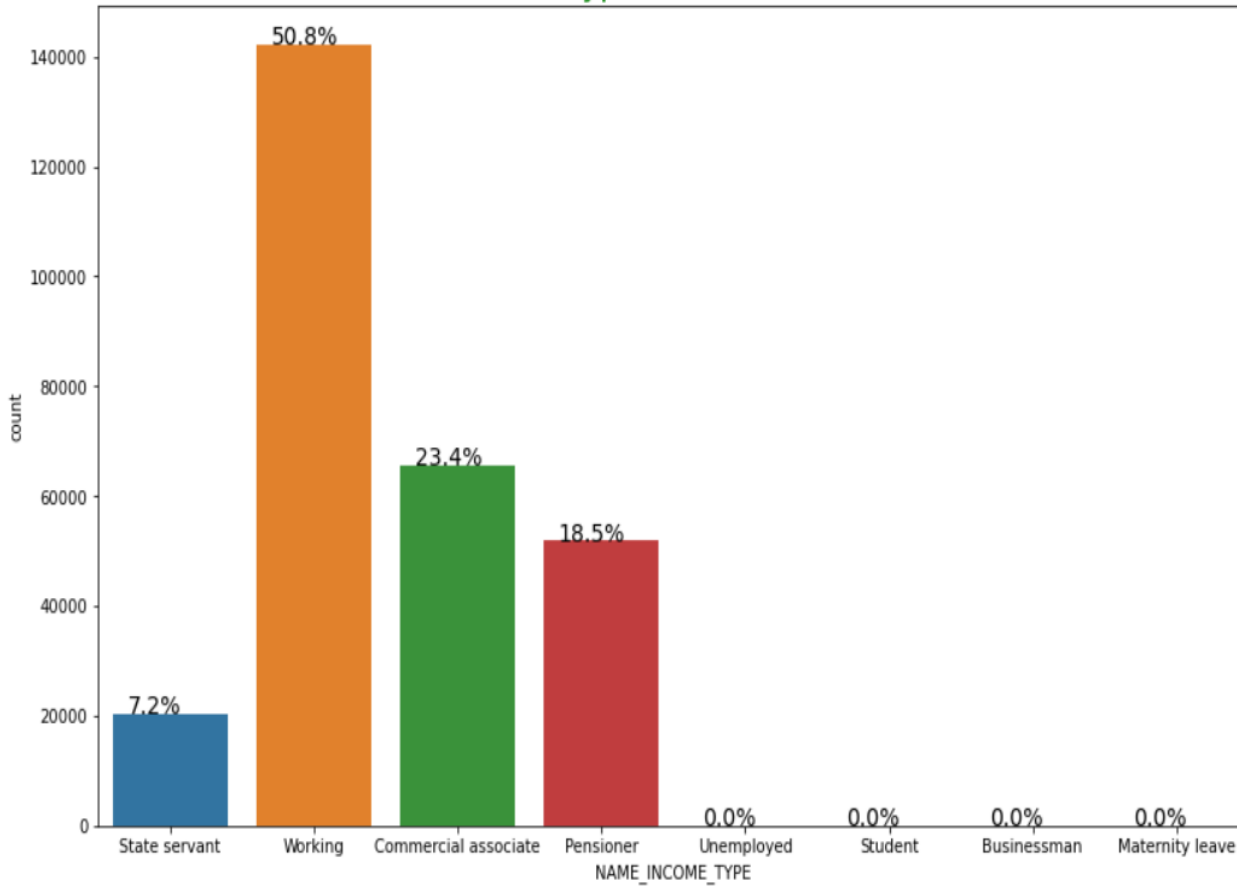
Age Group



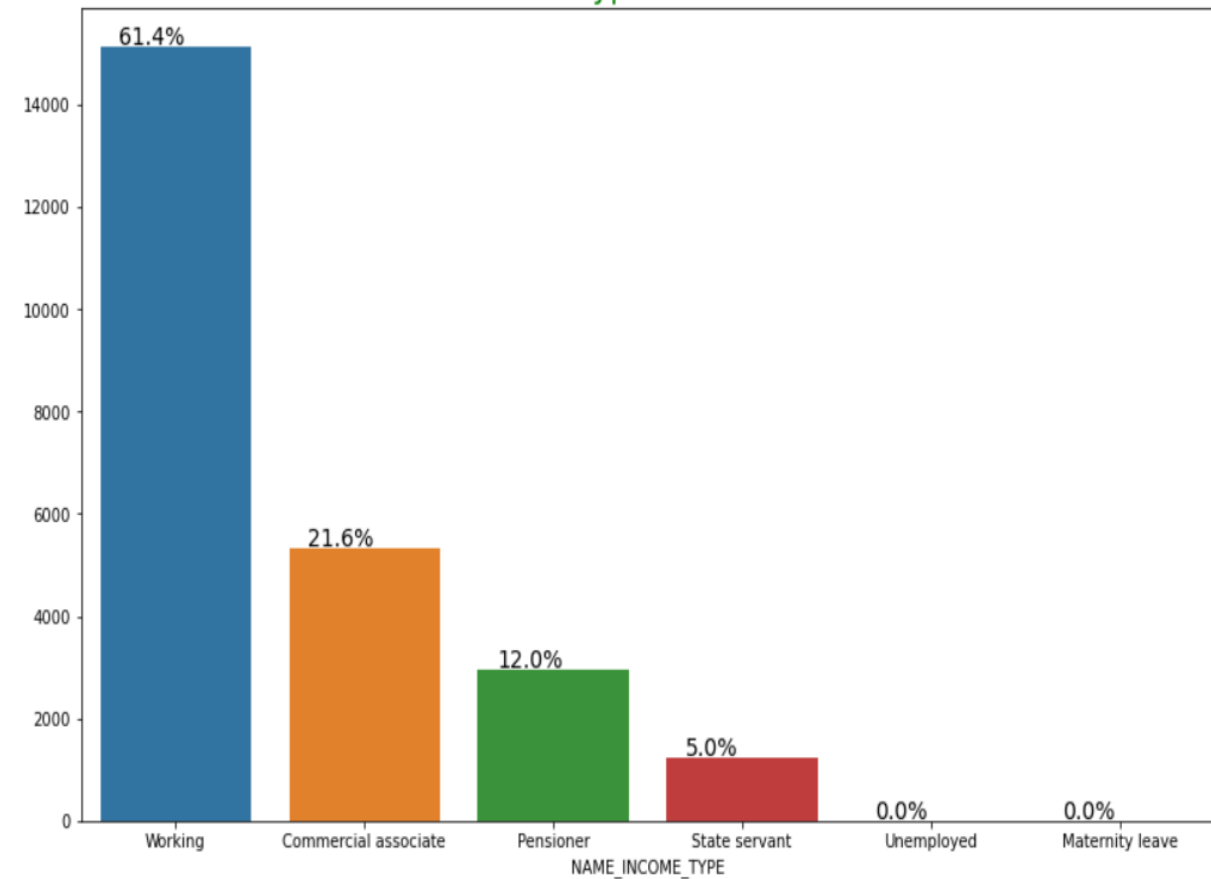
People having Age group of 0-30 and 30-40 years are more defaulter and as the age increases change of defaulter is decrease.

Income Type

Income type of Non-Defaulter

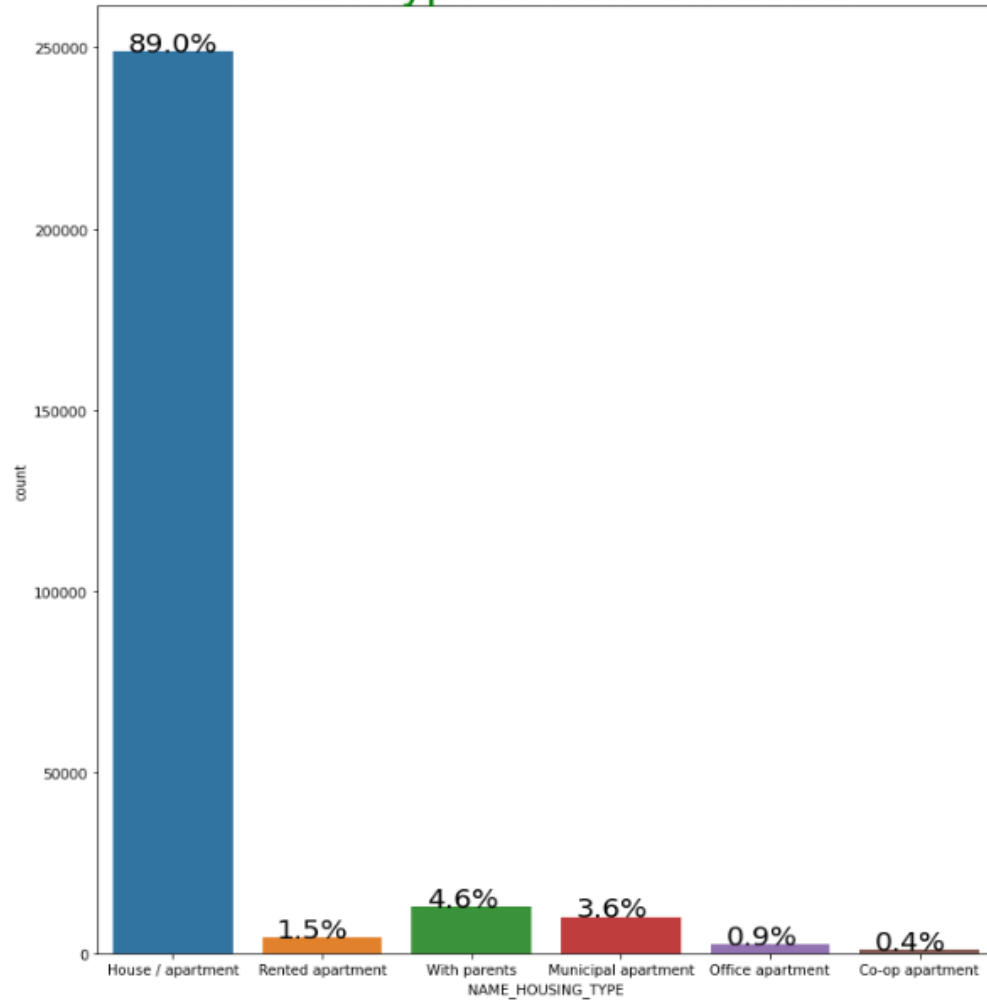


Income type of Defaulter

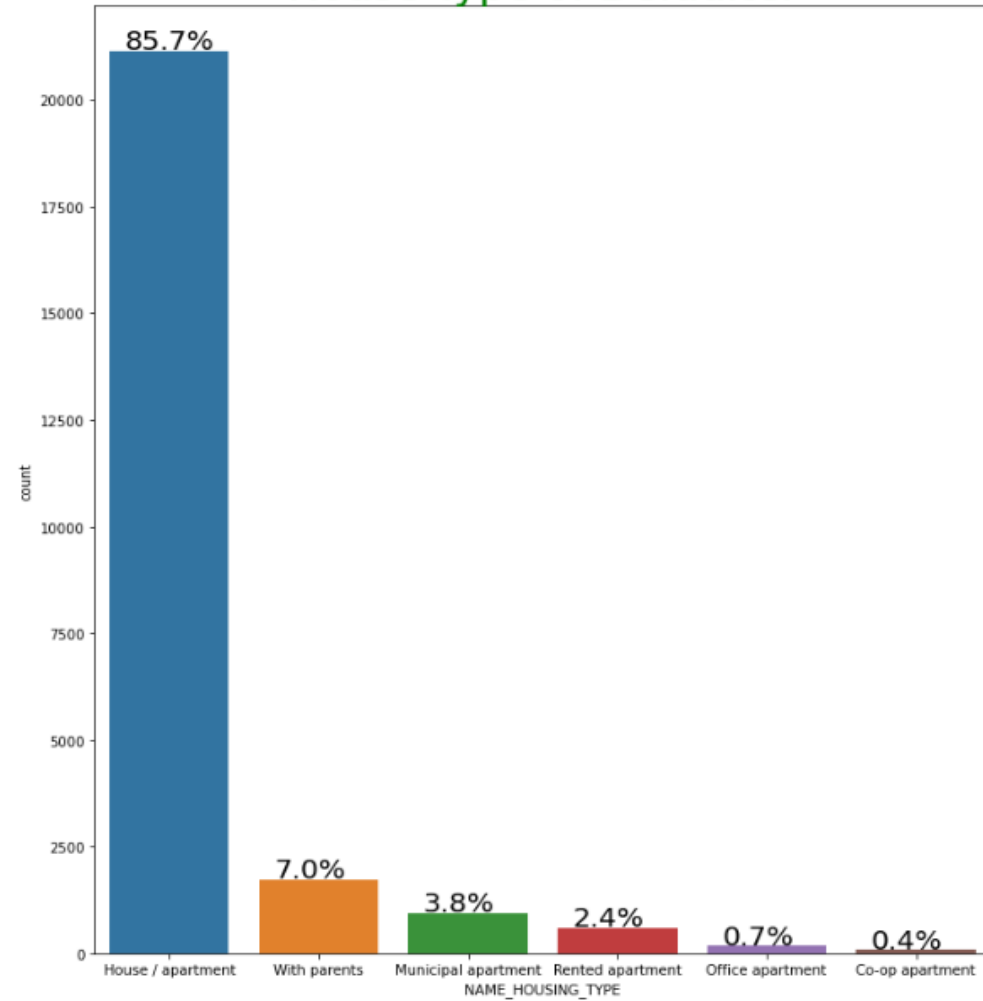


House Type

House Type of Non- Defaulter

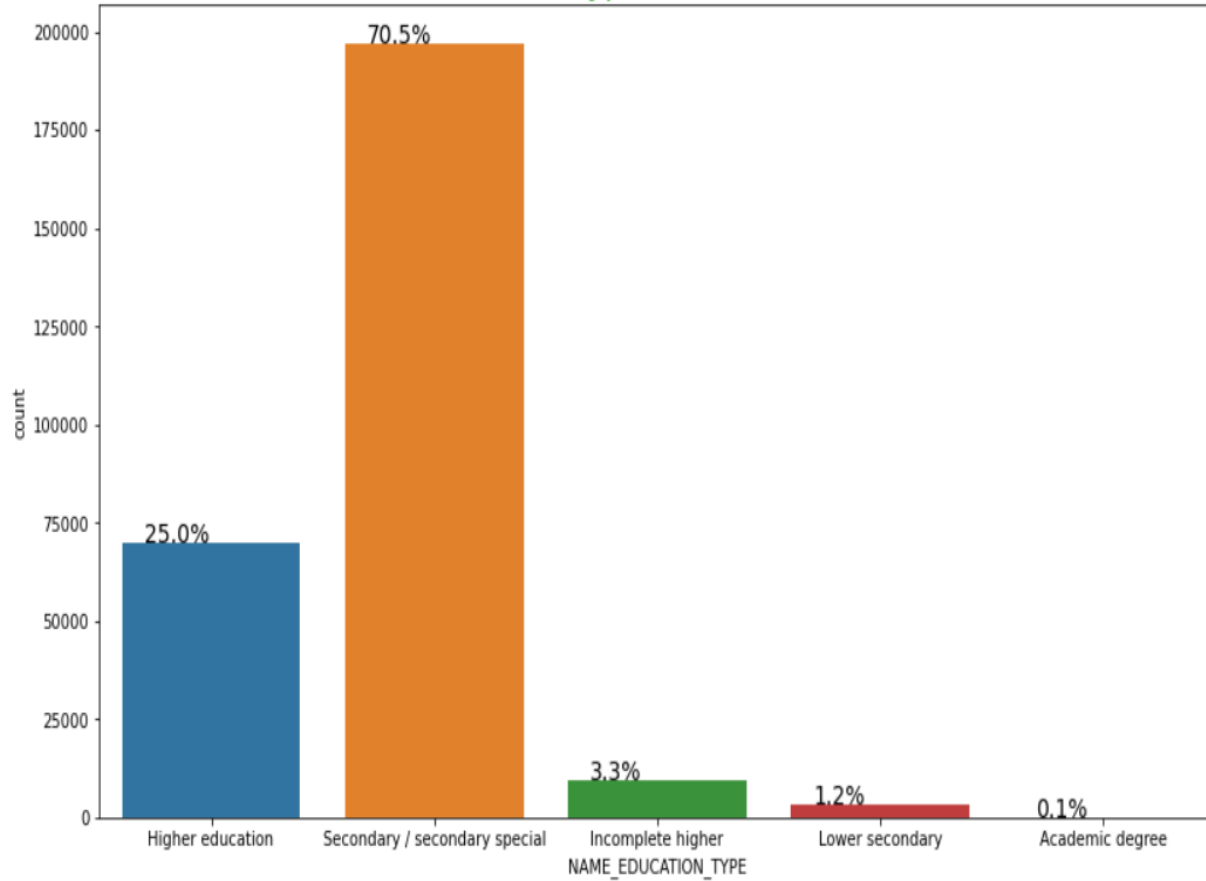


House Type of Defaulter

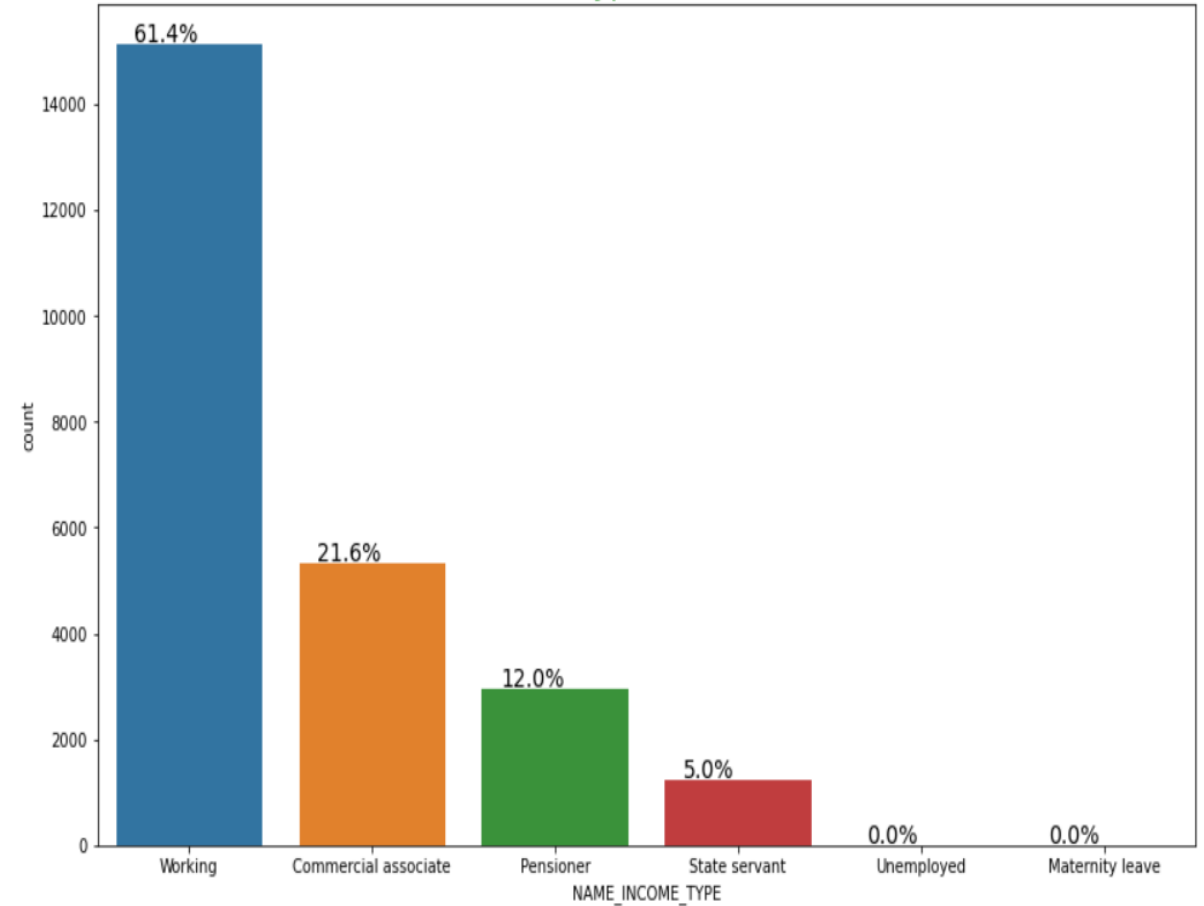


Education Type

Education type of Non-Defaulter

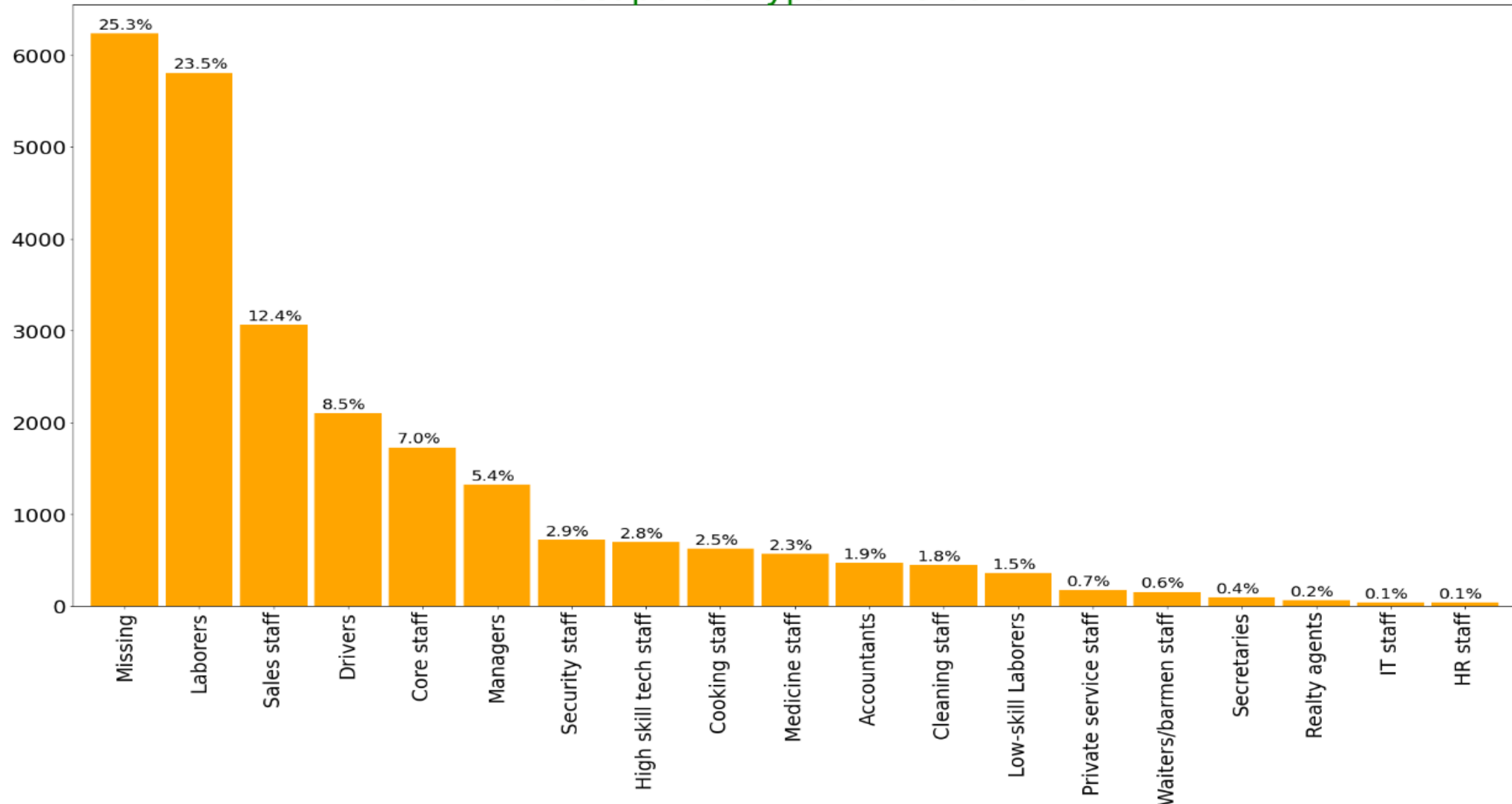


Income type of Defaulter

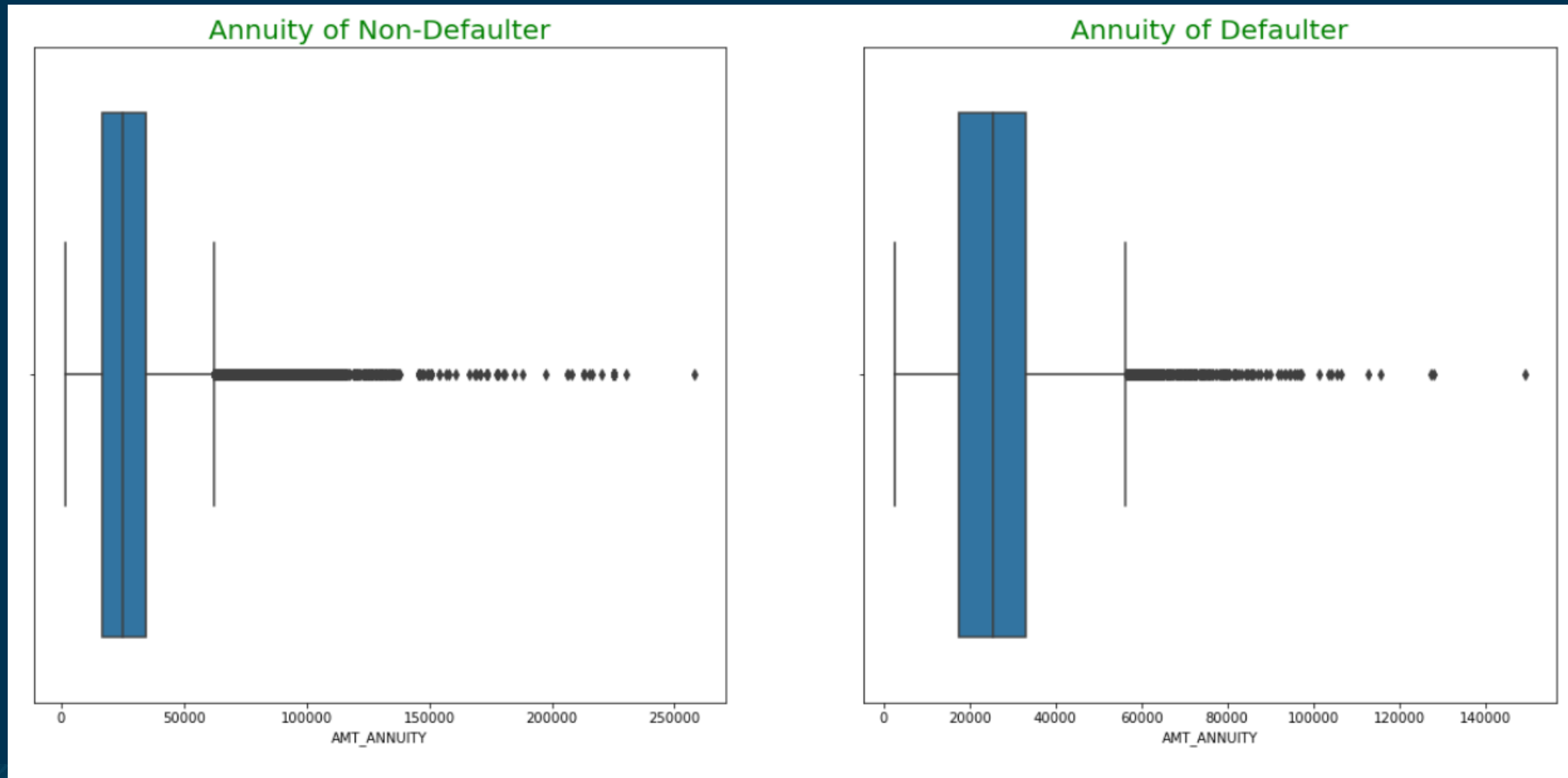


Occupation Type

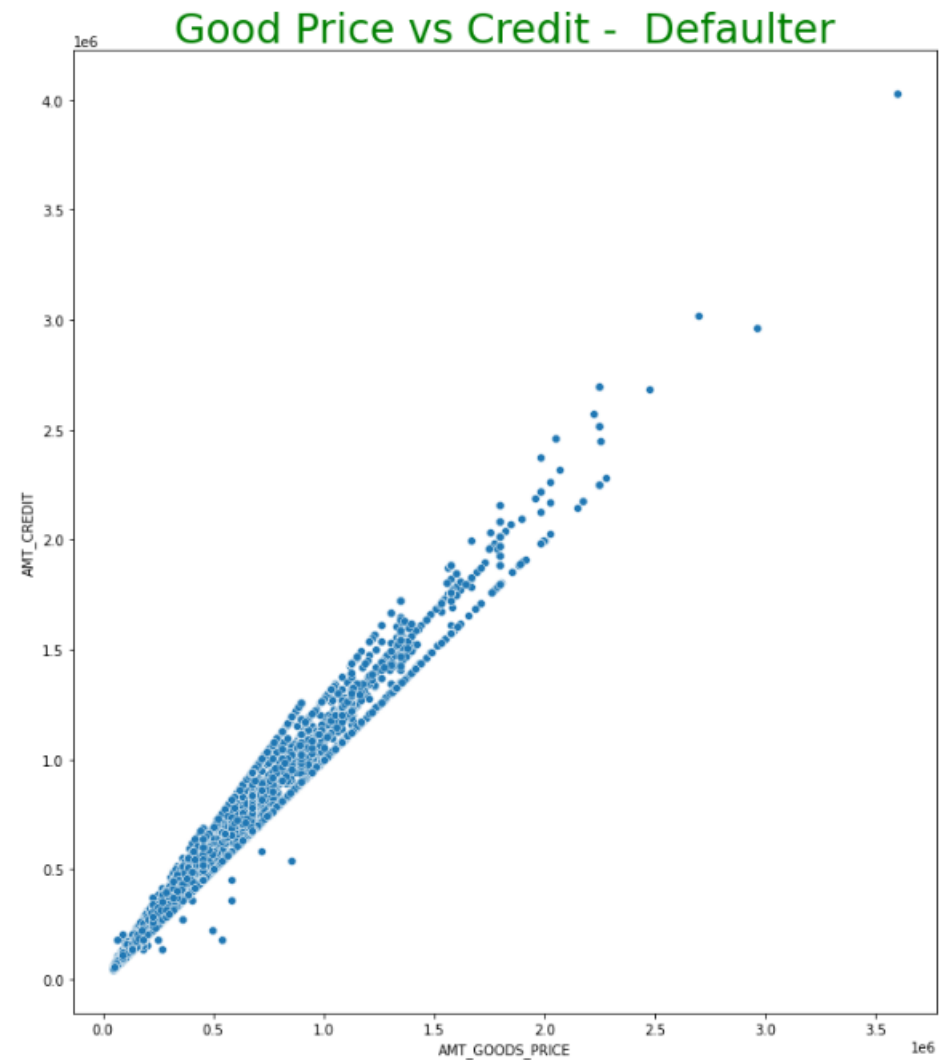
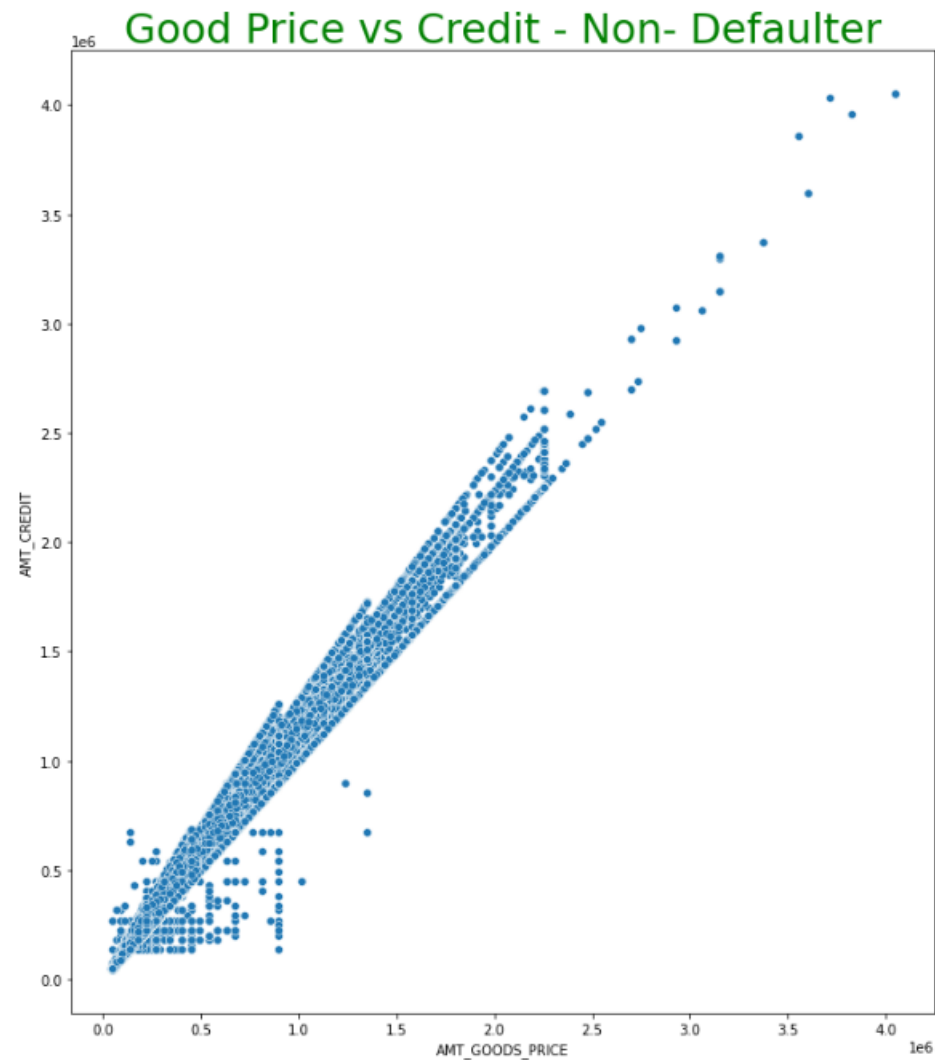
Occupation type of Defaulter



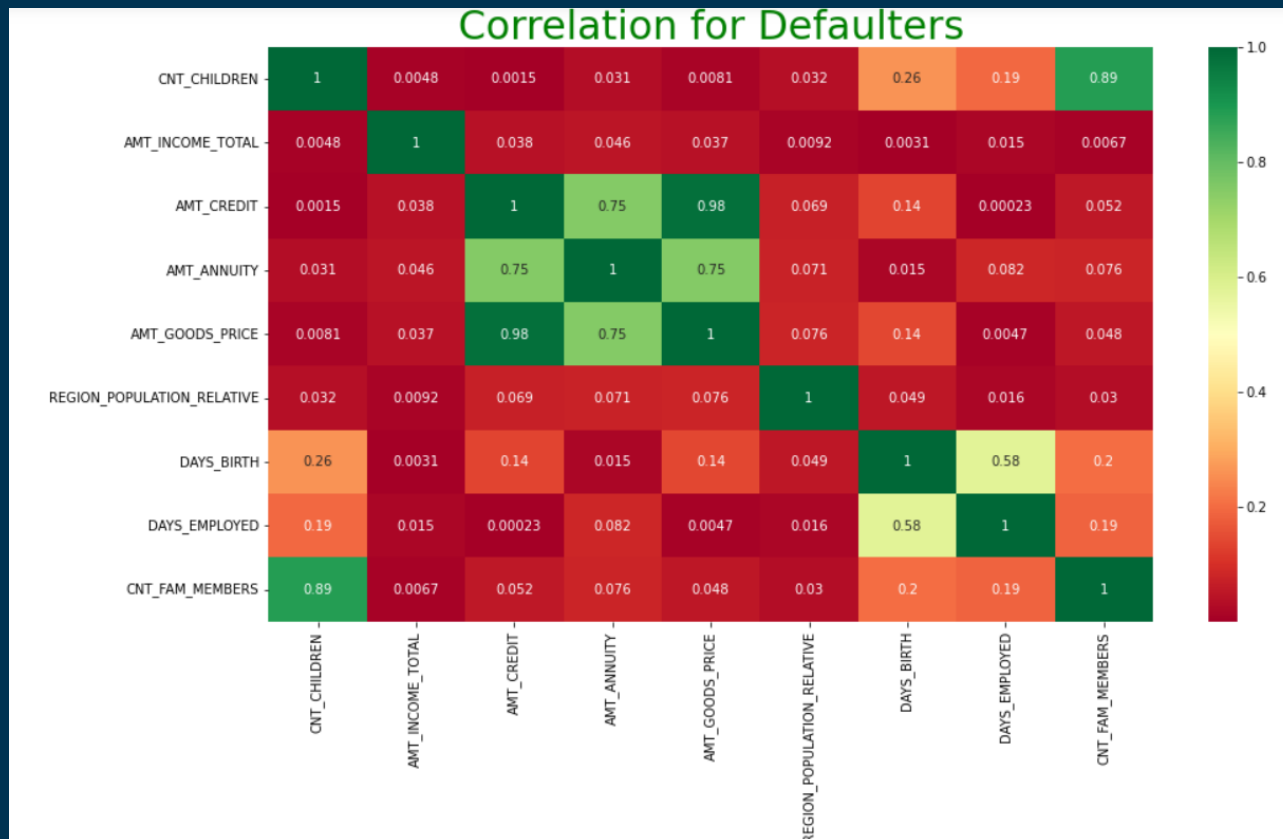
Annuity



Goods Price vs Credit



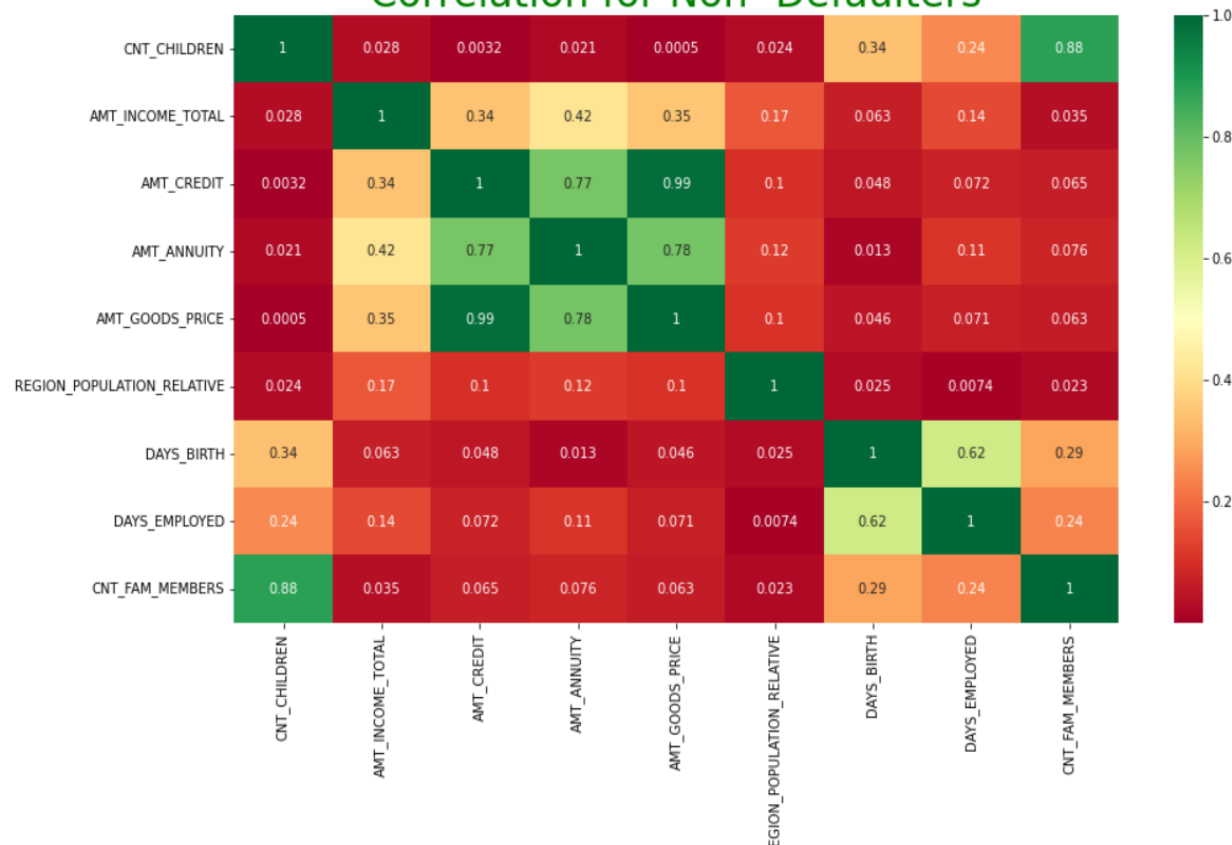
Top 10 Correlation for Defaulters



AMT_CREDIT	AMT_GOODS_PRICE	0.983065
CNT_FAM_MEMBERS	CNT_CHILDREN	0.885556
AMT_GOODS_PRICE	AMT_ANNUITY	0.752206
AMT_ANNUITY	AMT_CREDIT	0.751400
DAYS_BIRTH	DAYS_EMPLOYED	0.575949
	CNT_CHILDREN	0.258889
CNT_FAM_MEMBERS	DAYS_BIRTH	0.203105
CNT_CHILDREN	DAYS_EMPLOYED	0.192111
CNT_FAM_MEMBERS	DAYS_EMPLOYED	0.186342
AMT_GOODS_PRICE	DAYS_BIRTH	0.137117

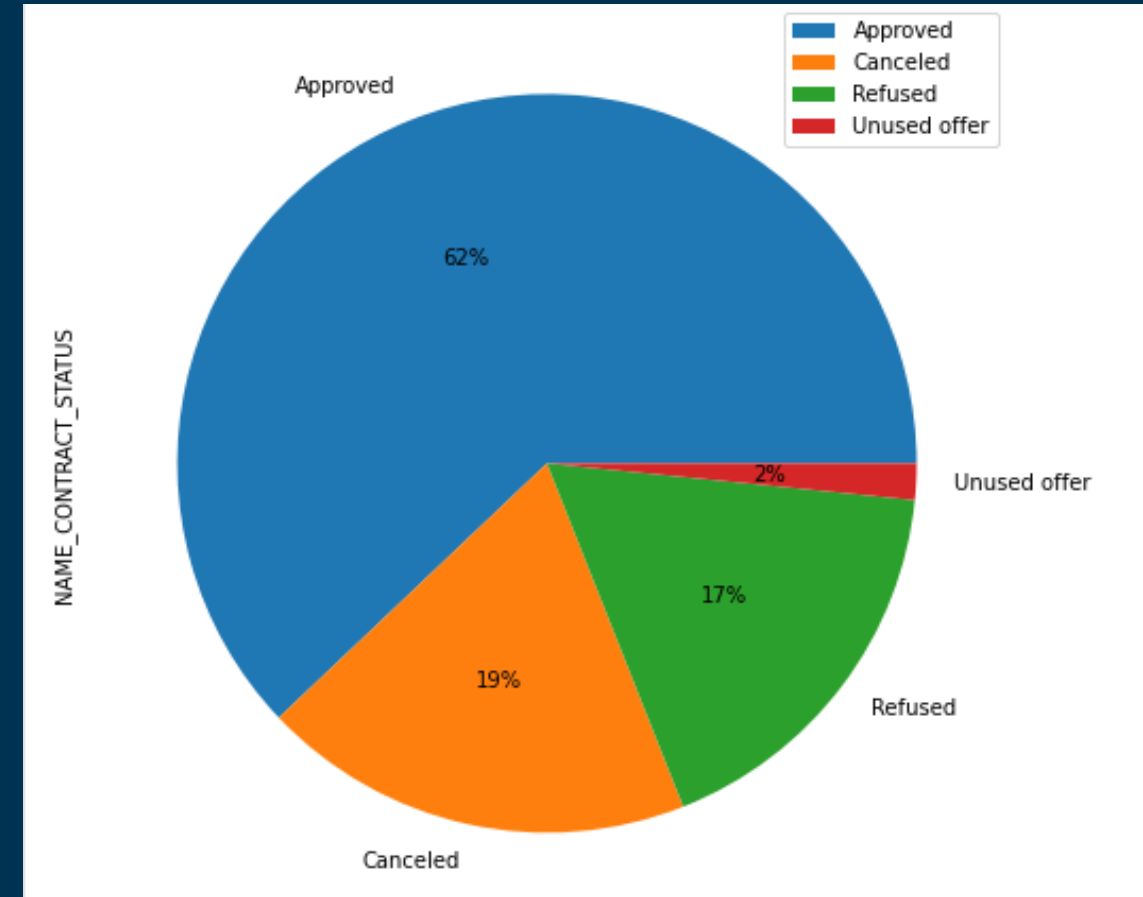
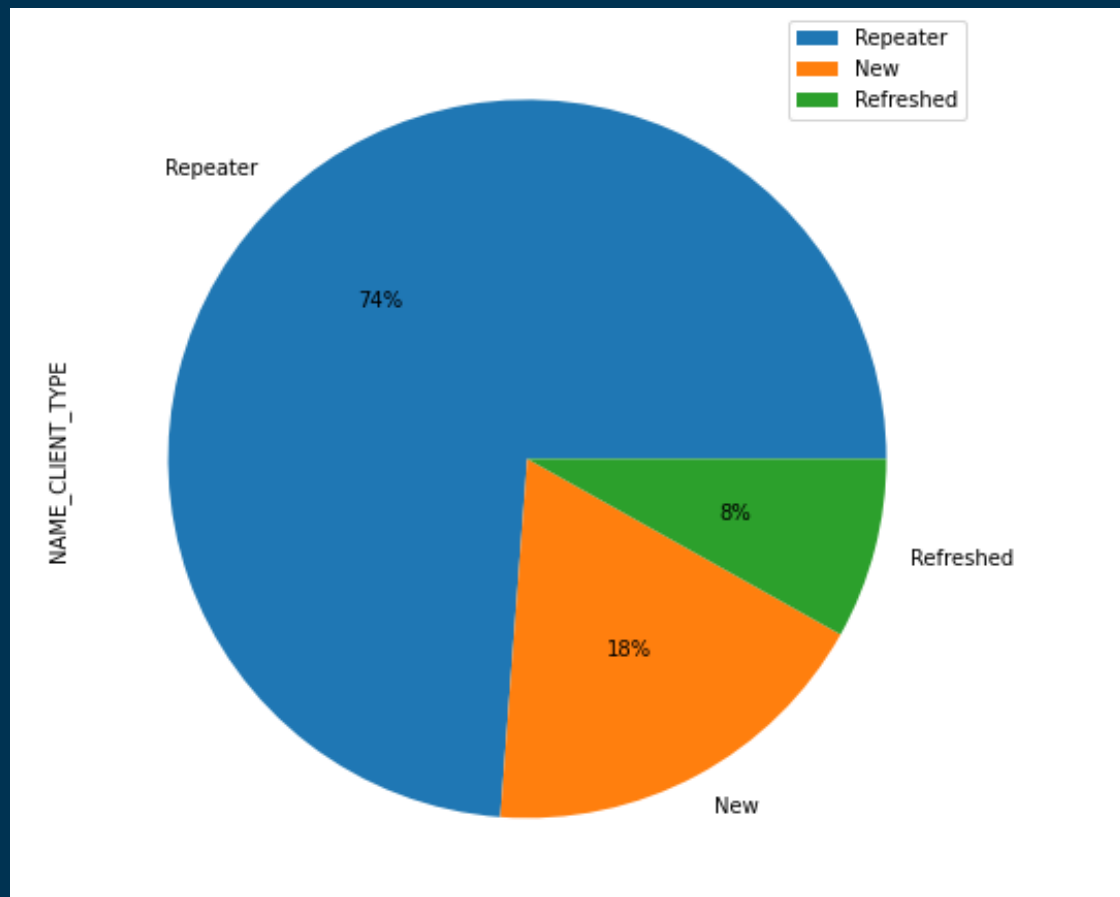
Top 10 Correlation for Non- Defaulters

Correlation for Non- Defaulters



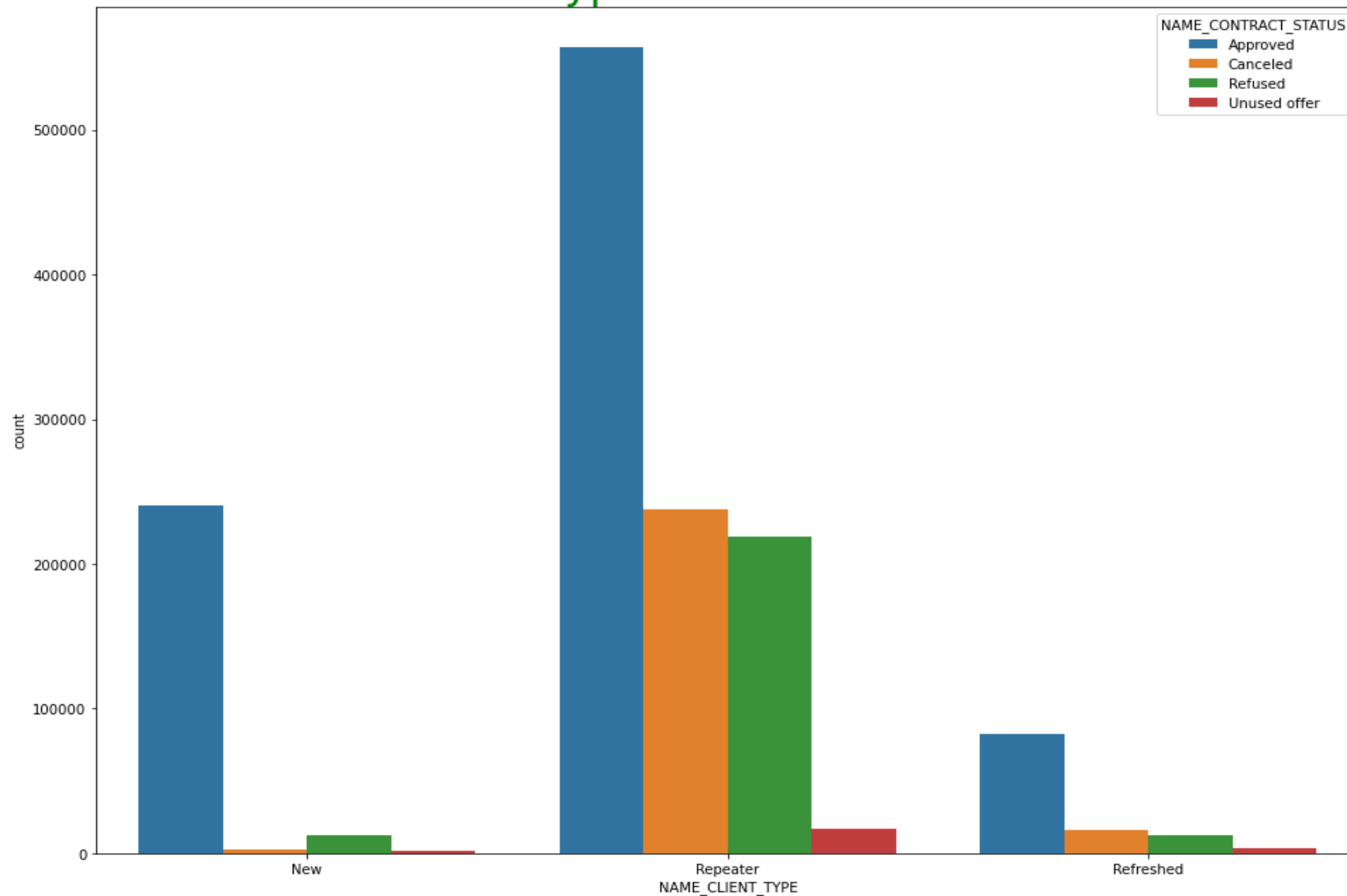
AMT_CREDIT	AMT_GOODS_PRICE	0.987260
CNT_FAM_MEMBERS	CNT_CHILDREN	0.878681
AMT_GOODS_PRICE	AMT_ANNUITY	0.775838
AMT_ANNUITY	AMT_CREDIT	0.770379
DAYS_BIRTH	DAYS_EMPLOYED	0.618068
AMT_ANNUITY	AMT_INCOME_TOTAL	0.417676
AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.347984
AMT_INCOME_TOTAL	AMT_CREDIT	0.341478
CNT_CHILDREN	DAYS_BIRTH	0.337664
CNT_FAM_MEMBERS	DAYS_BIRTH	0.286655

Client Type and Contract Status

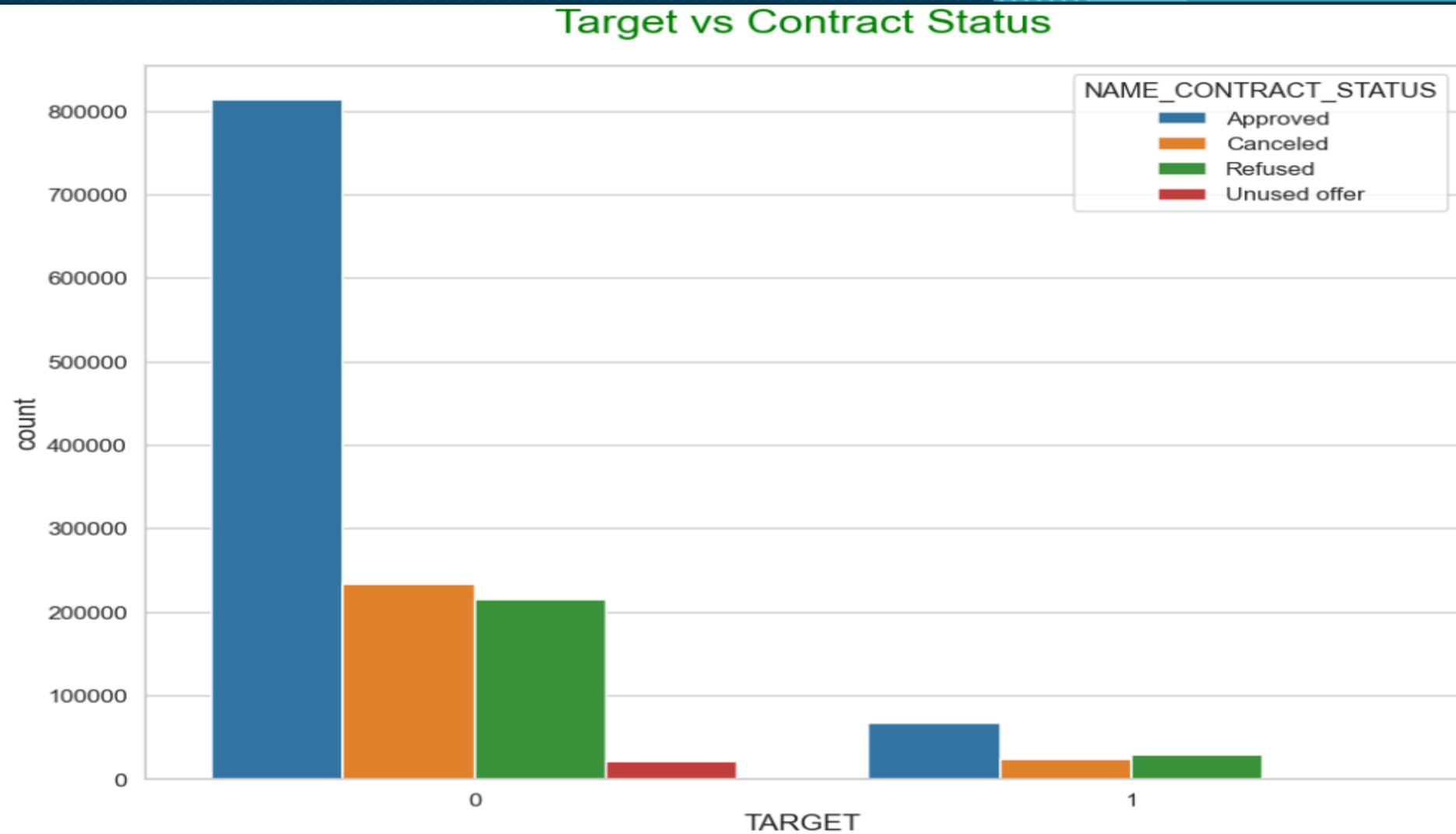


Client Type vs Contract Status

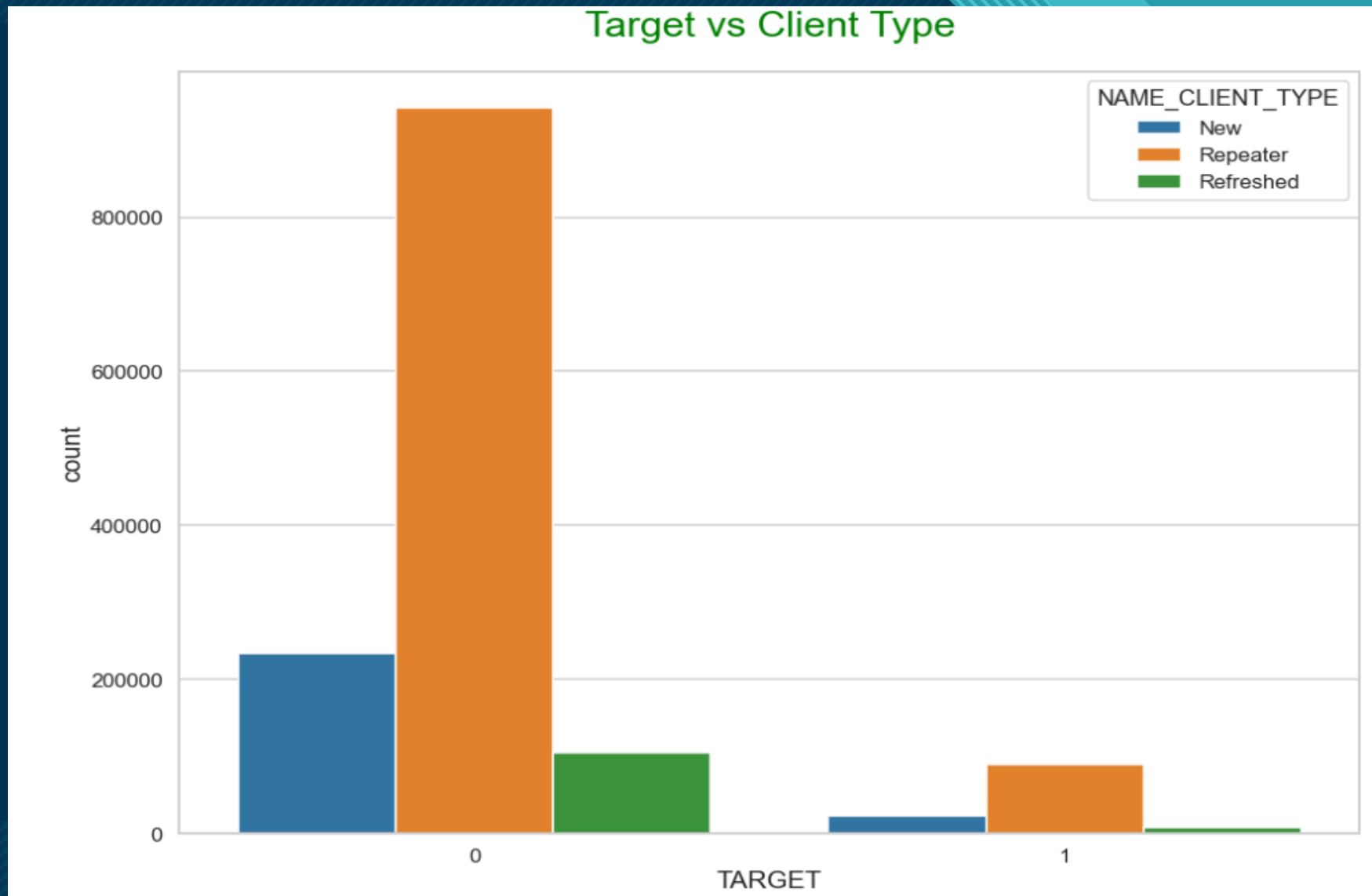
Client Type vs Contract Status



Target vs Contract Status



Target vs Client type



Observation



- Bank should focus on females as male are more defaulters.
- Age group of 30-40 Years are more risky as they are more defaulters.
- Bank should focus on Business men and students as they never defaulter and should not focus on people who are working as they are most defaulter.
- Bank should focus on people having Academic degree and Higher Education. And they should not focus on people having Secondary education as they are the most who are defaulter.
- Bank should not focus on the people who leave with Parents.
- Bank should focus on Managers and should not focus on Laborers and Drivers.
- Business Entity Type 3 are more defaulter.
- People are on maternity leave are more defaulter.
- Bank should focus on repeated customers as chance of Defaulter is less.
- Mobile is the category on which loan is applied mostly and mobile category people are the most who are Non - Defaulter.
- Mostly People use payment mode is cash through bank.
- Main reason of loan Rejection is HC.



Thank You