Credit EDA Assignment

Presented By :- Yathestha Siddh

Yathestha@gmail.com

7506933629

EDA Process Steps

1) Importing Data Source CSV files
2) Checking the Data with help of reference description file
3) Check for Missing Values
3) Imputation of missing value
4) For Application data, Divide the data as per Target value and Compare the target variable across categories of categorical variables.
5) For previous data, Divide the data as per status and Compare the target variable across categories of categorical variables.
6) merge the both dataset
7) univariate and bivariate analysis
8) present the findings

Customer Analysis

a) Raito of Defaulter is very less around 8 % and regular customers are around 92%.
b) As we could see distribution of defaulters and regular customer across different Age groups.
c) Our Product has more penetration for the middle age group, we could target with another group to increase the market base.
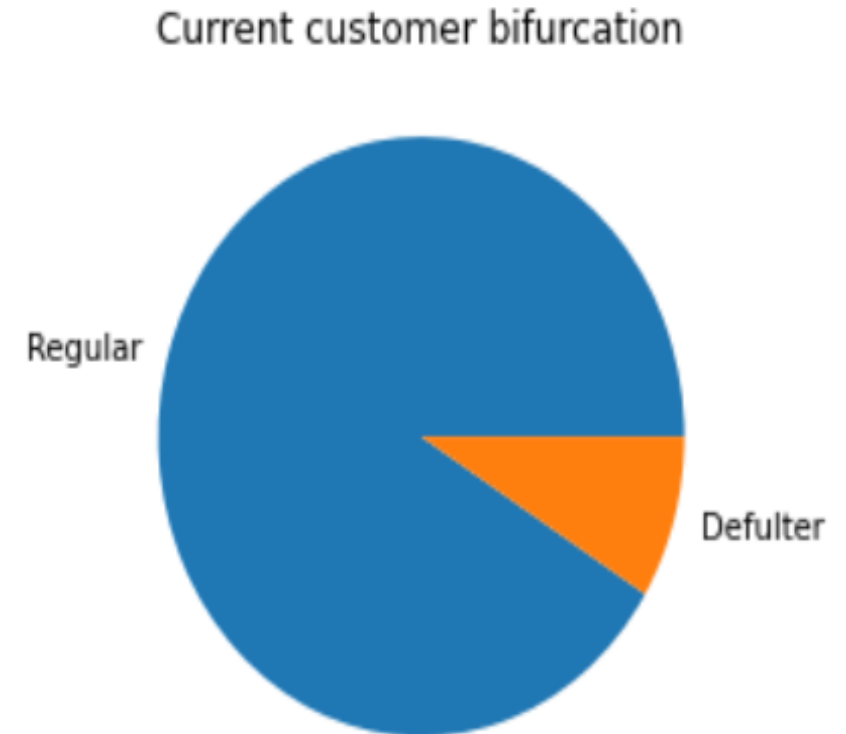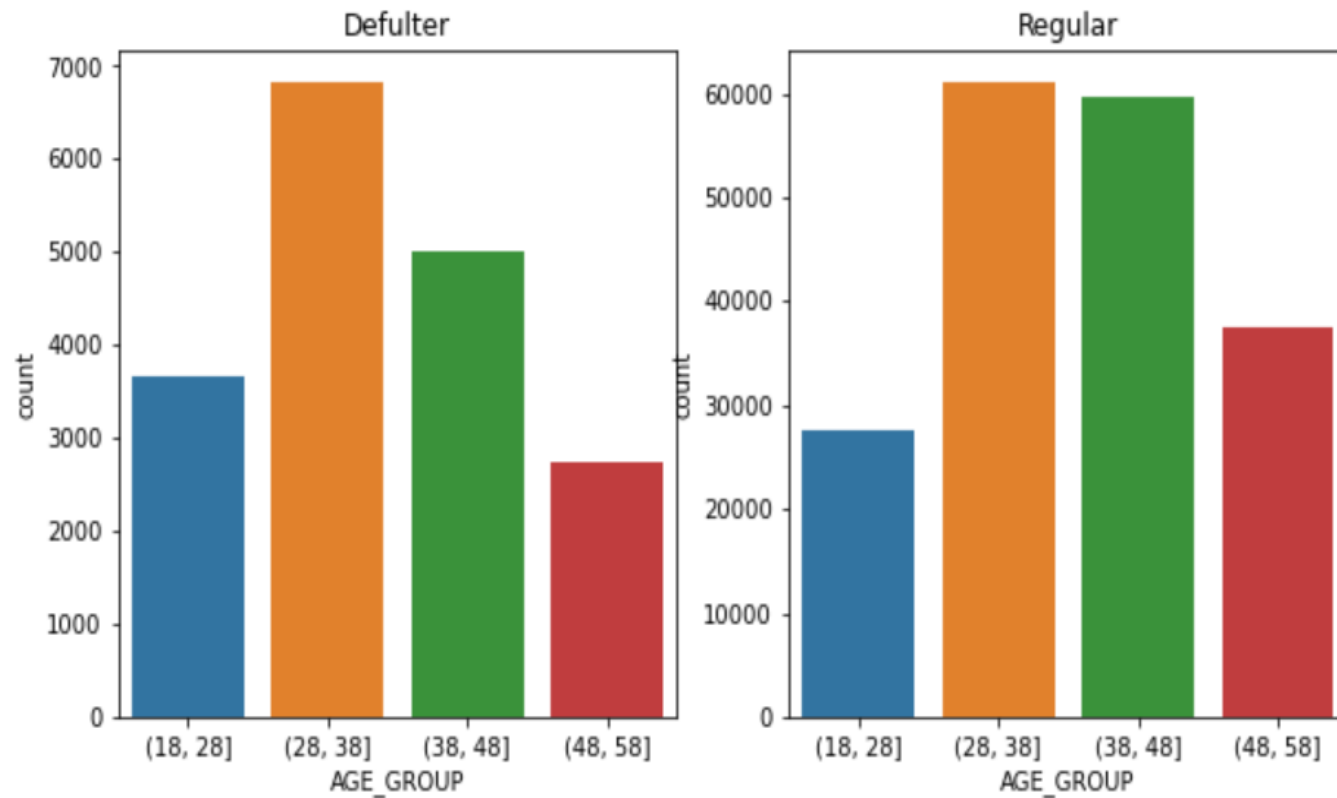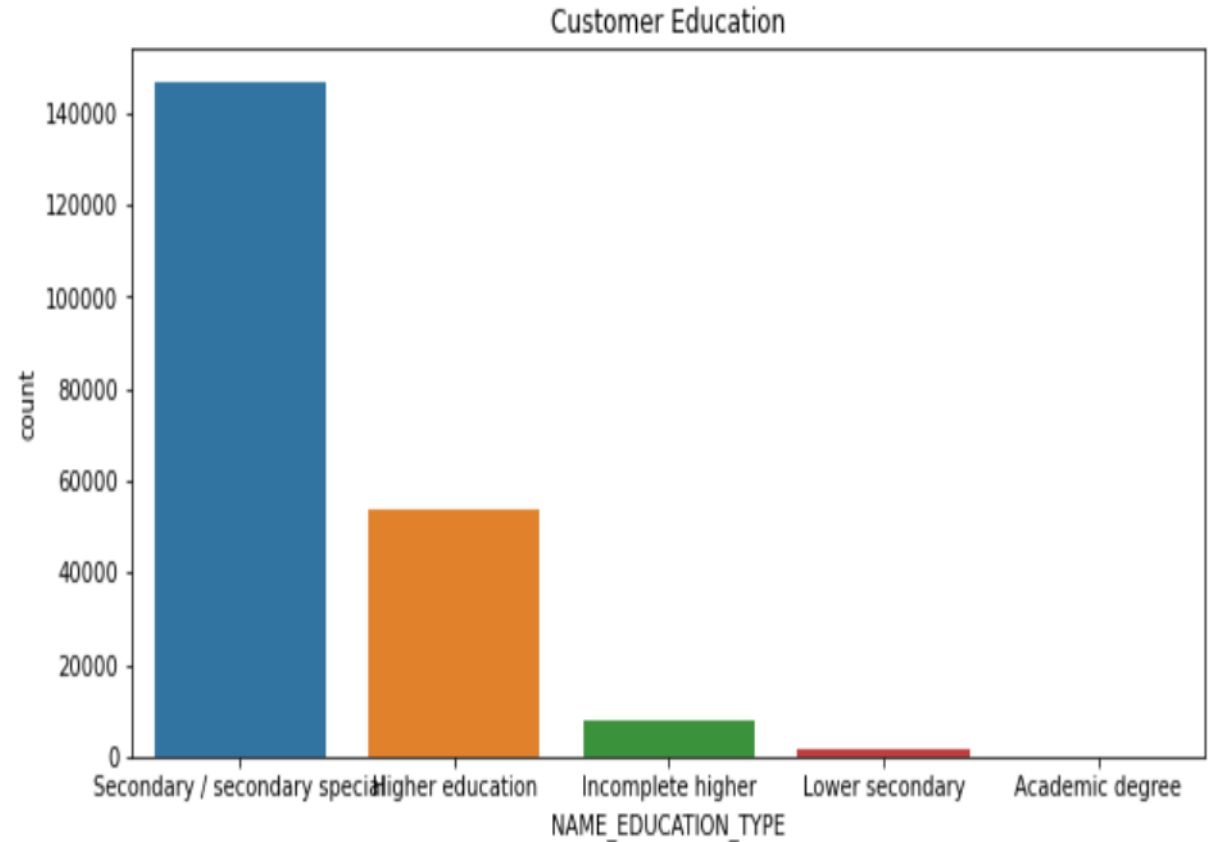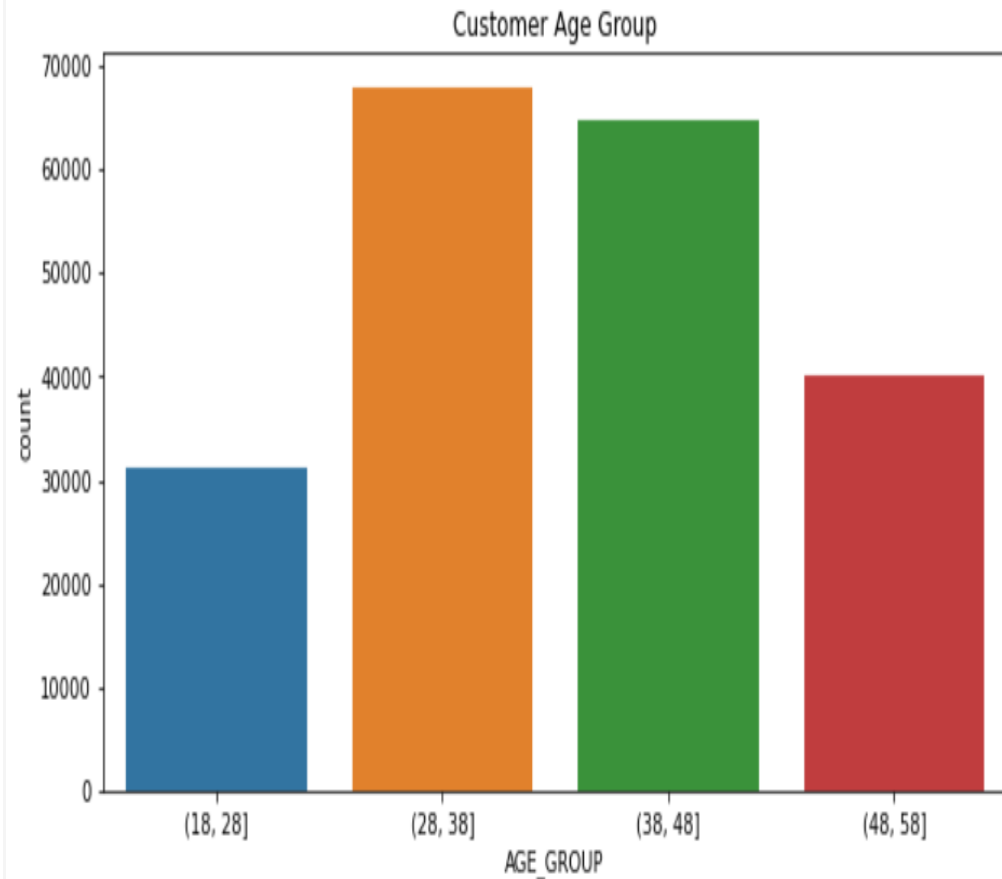
Gender distribution:-
a) Raito of Female customers is higher than male customers, male customer are 38% and female customers are around 61%.
b) As we could see a distribution of defaulters and regular across Gander is identical
c) Our Product has more popular with Females, we could target this group to increase market base and branding.
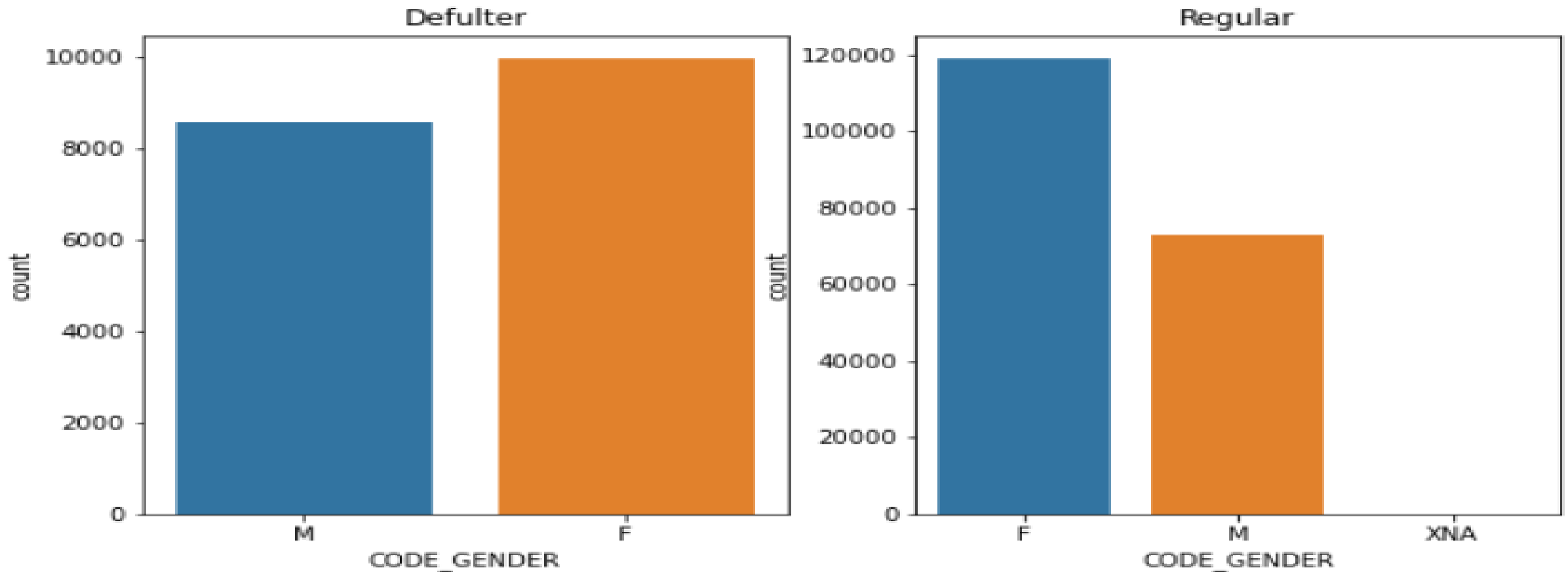
a) As we could see penetration of our product is higher for middle age group and between customers those have secondary and higher education.

b) We could come up with new products as per market research for other age group and different education category to increase the market
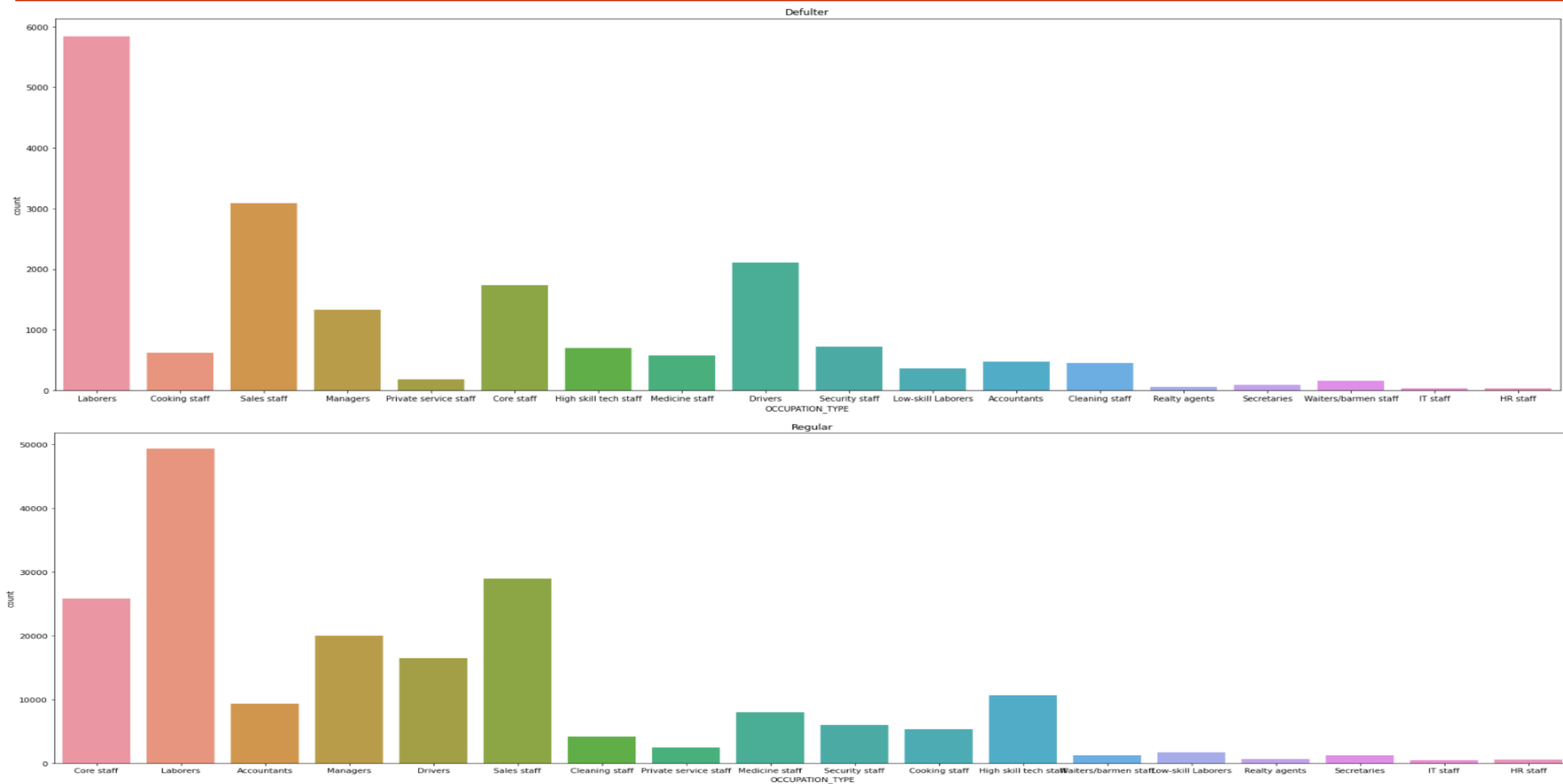
Customer Income distribution
 a) As we could see penetration of our product is higher for middle income customers majority of customer comes under income rage of 100k-500k.
b) We could come up with new products as per market research for lower and higher income customer group to increase the market and focus more in our strong area to become market leader.

Income wise Distibution
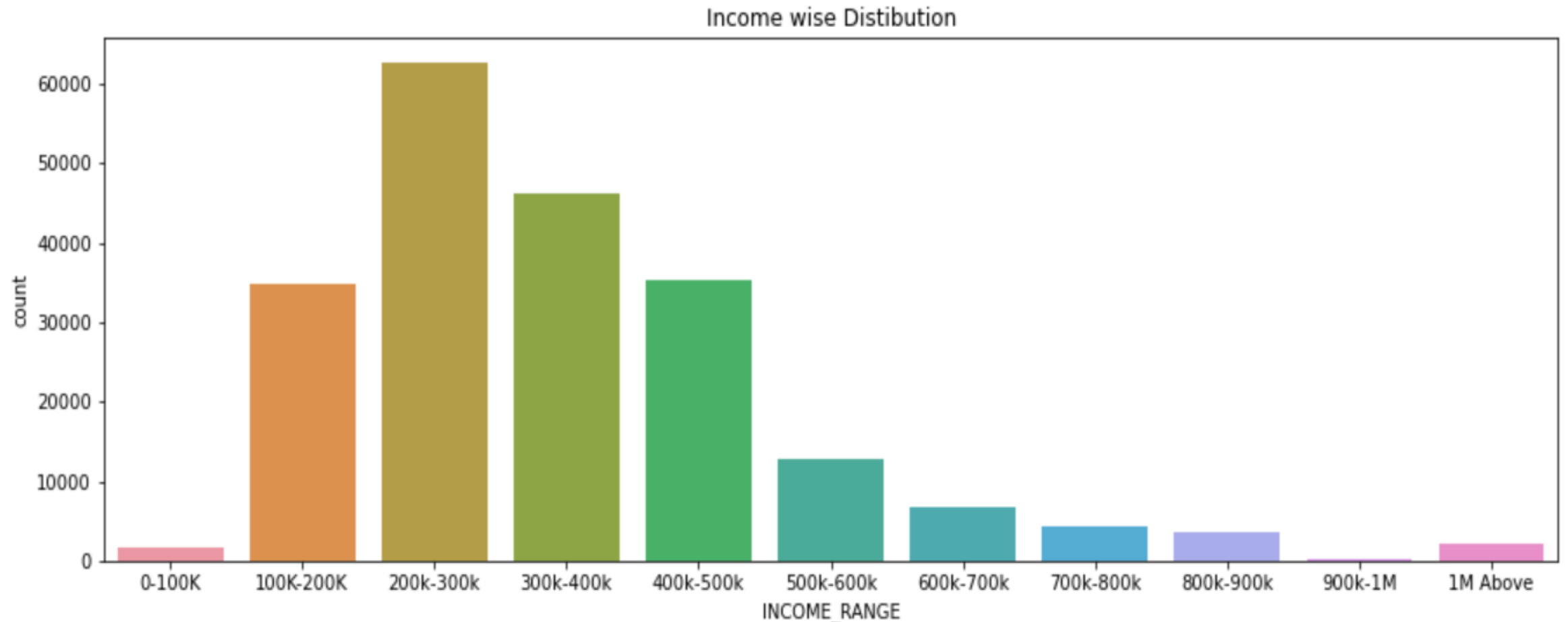
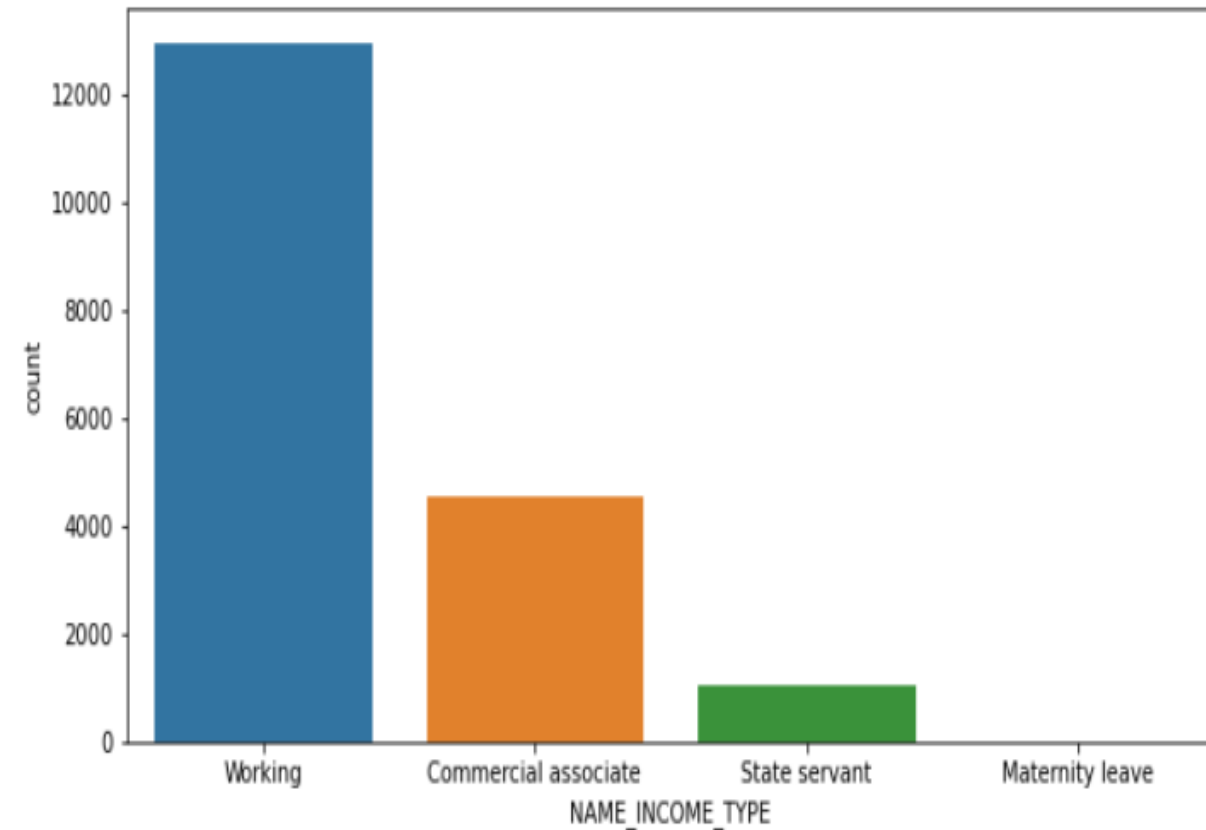# Distribution Customer Occupation as per defaulter and regular category

Customer Source of Income distribution graph as per defaulter and regular category

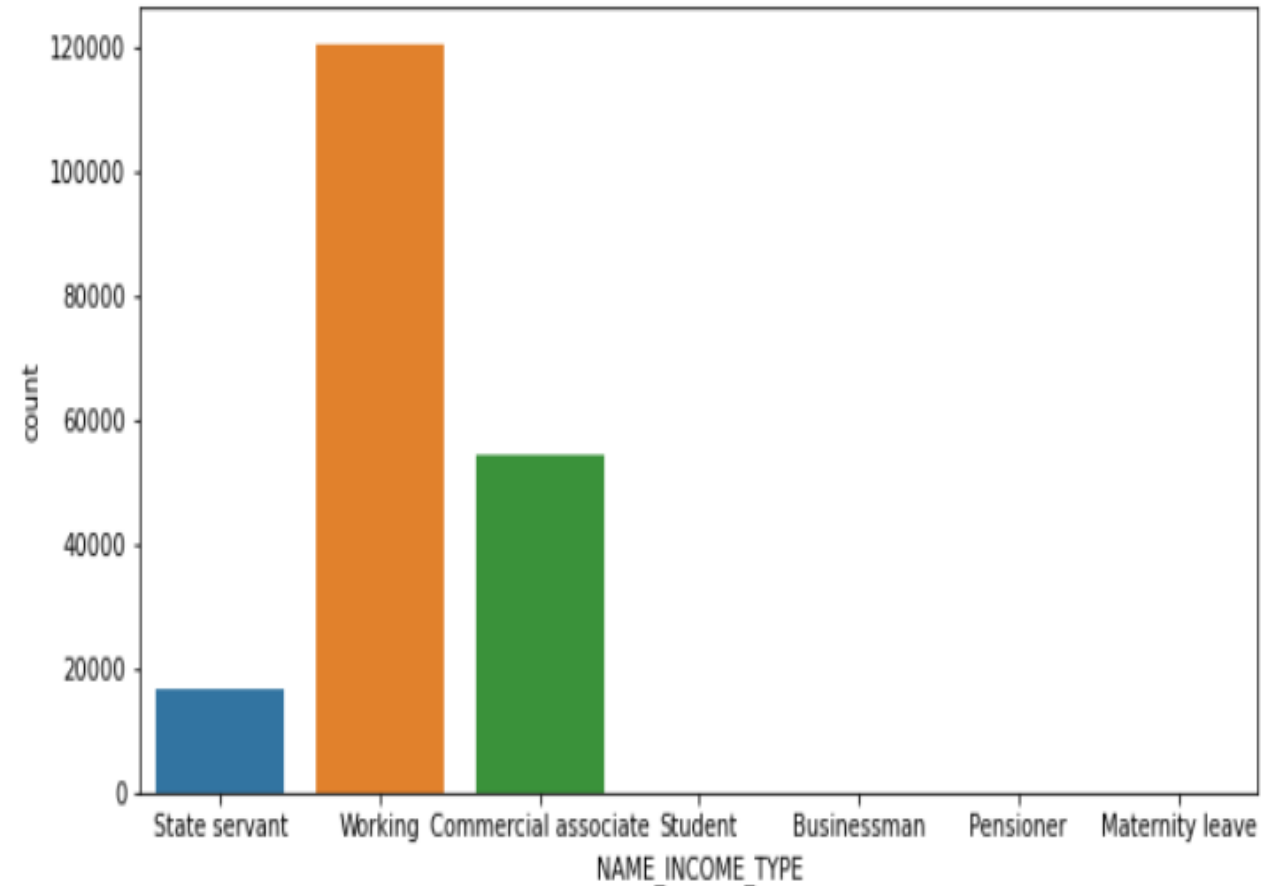a) This identical for defaulter and regular category
b) as most of customer source of income (could also used as sub category for **Customer Occupation**) is working, commercial associate and government employees
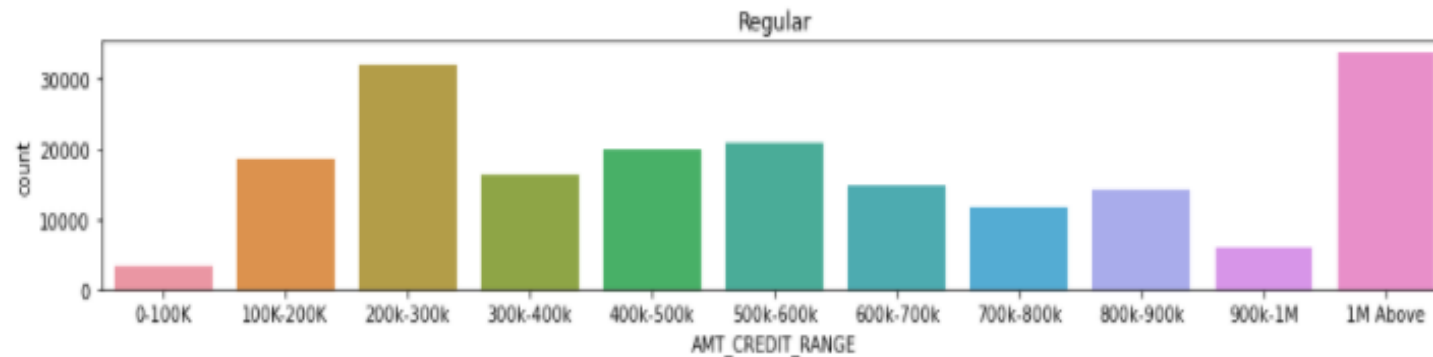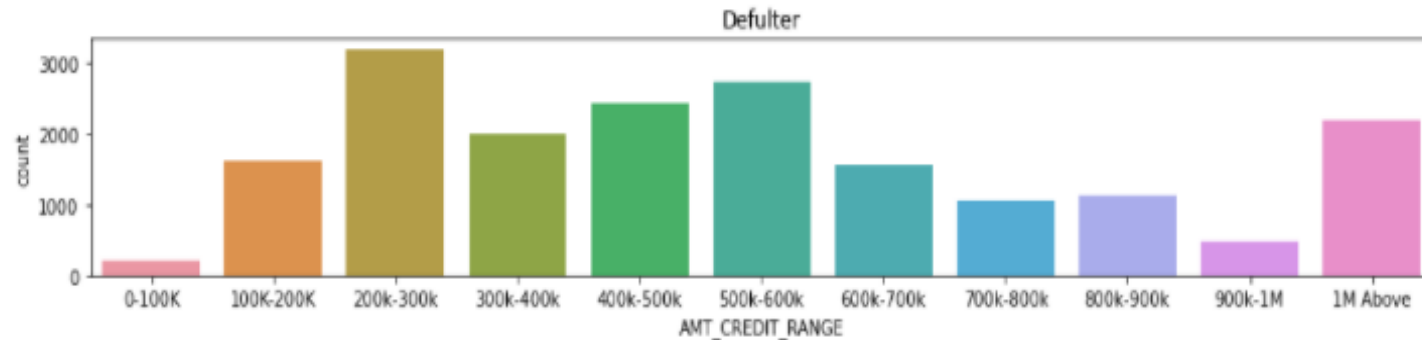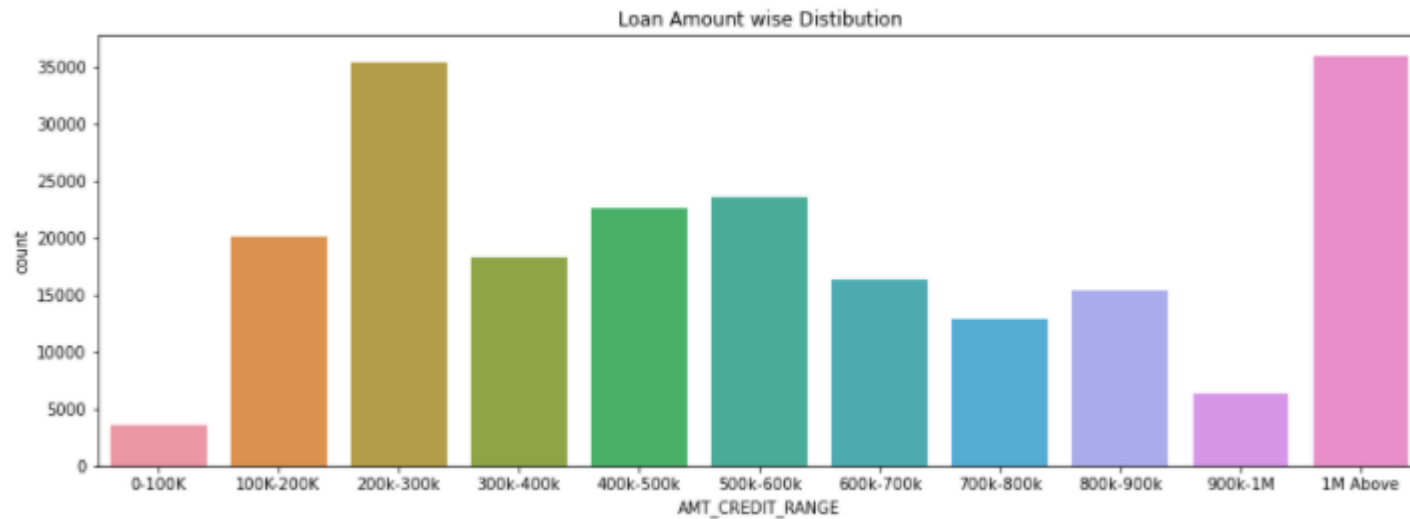C) We could focus more on these groups
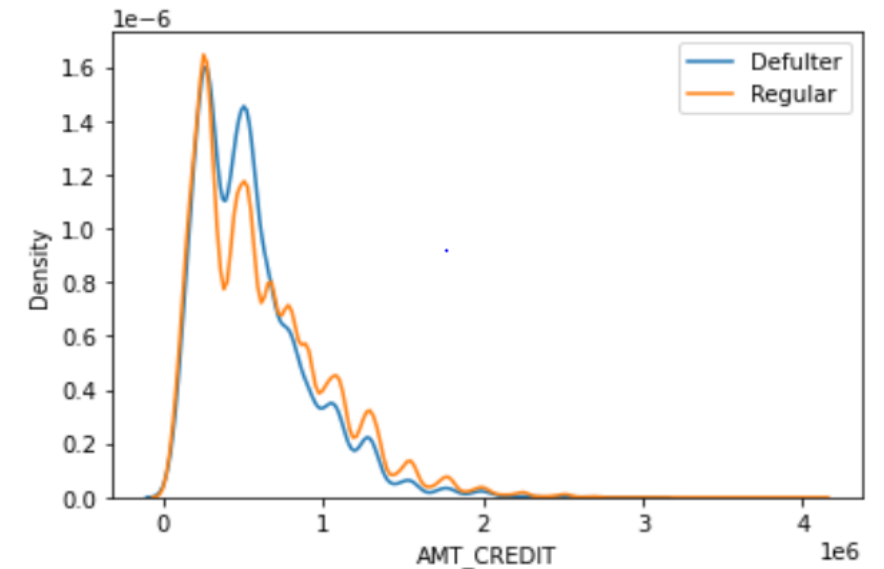
Loan Amount wise Distribution

the Loan amount distribution

This Graph shows us the Loan amount distribution rage

a) most of the disbursed loan are for middle rage loan

b) the distribution graph is almost identical for defaulter and regular customer

Loan Product distribution

a) Loan Product distribution is identical for defaulted and regular category
b) majority of loans are Cash loans and these are most popular, we could produce more verity of these loan to increase market share
c) there lot of scope to increase market share in revolving loan as currently its ratio is less in compare to cash loan , we could redefine new product in this category to capture more customers
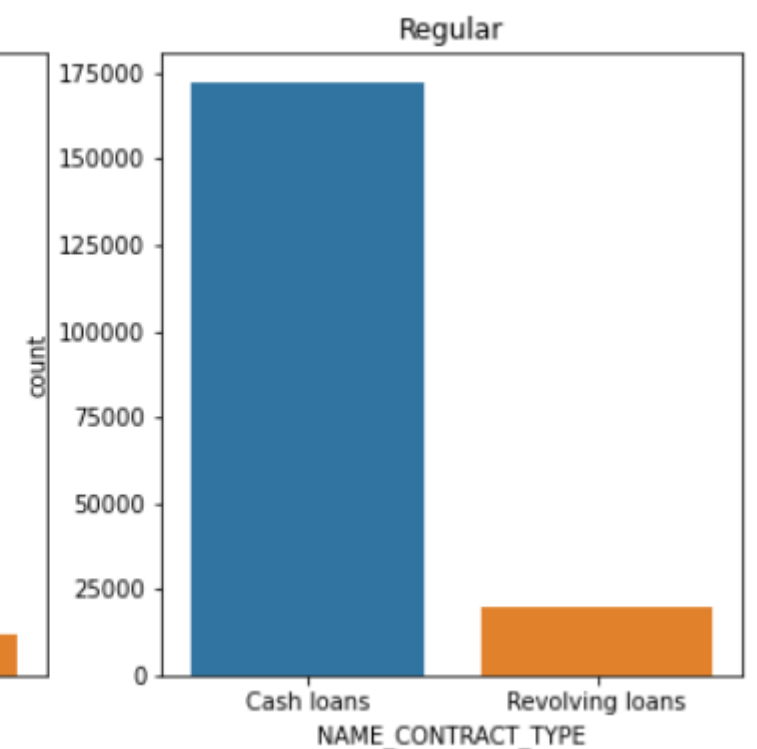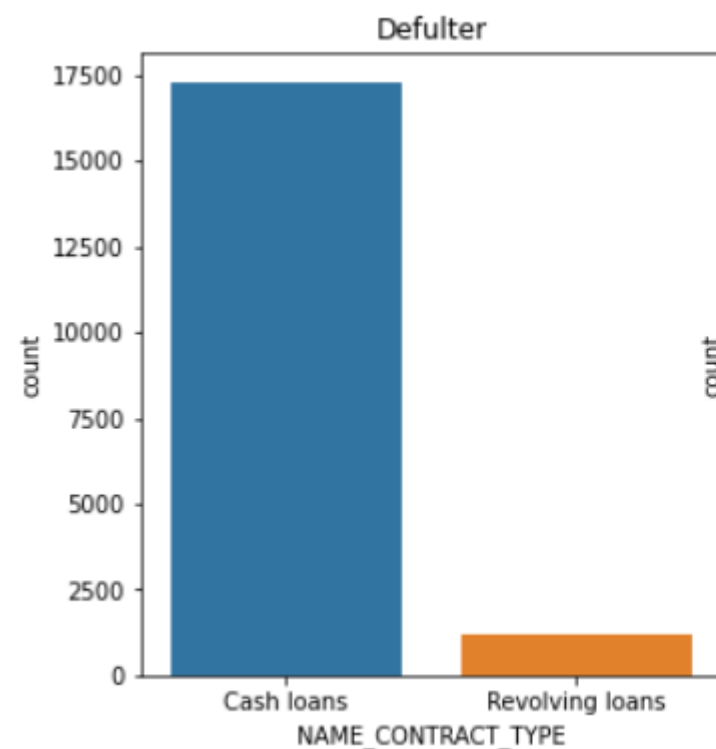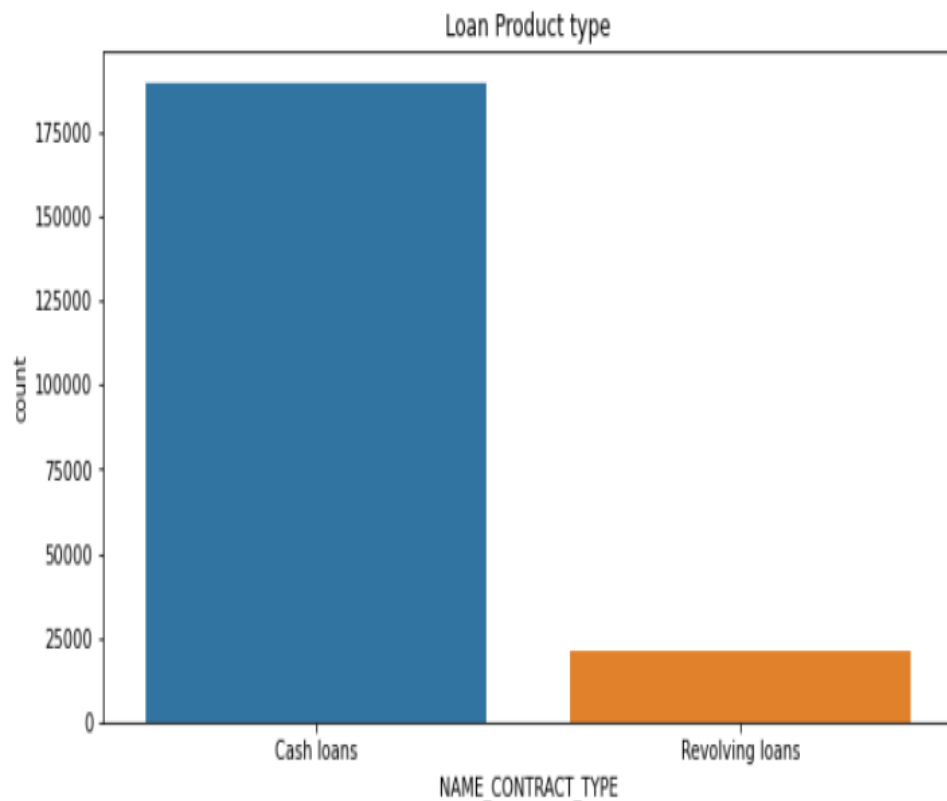
Loan status Distribution graph

As we could see the approval rate is much higher than cancel and refuse and unused loans for middle value loans
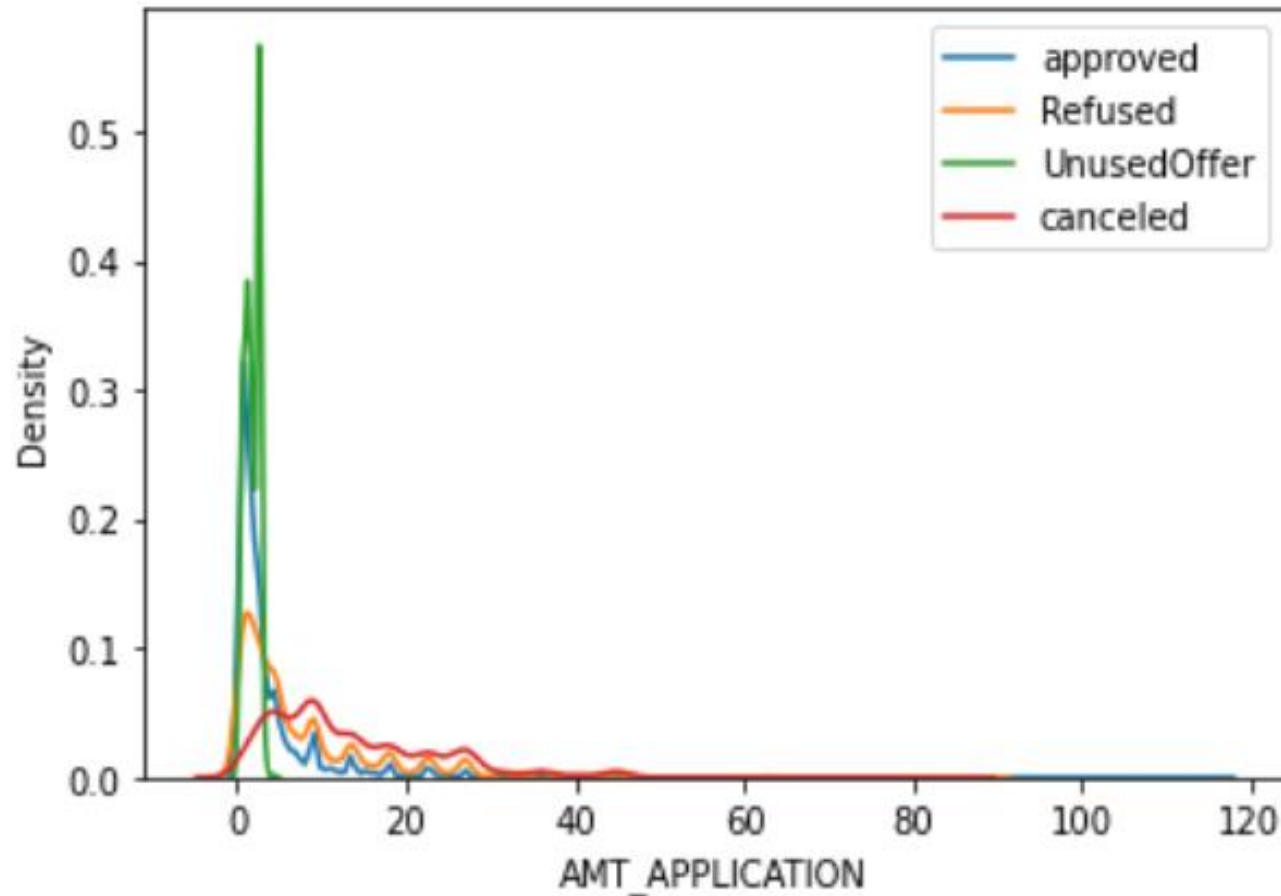

Approved: The Company has approved loan Application

Canceled: The client canceled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.

Refused: The company had rejected the loan (because the client does not meet their requirements etc.).

Unused offer: The loan has been canceled by the client but at different stages of the process.

Bifurcation of Loan as per Status
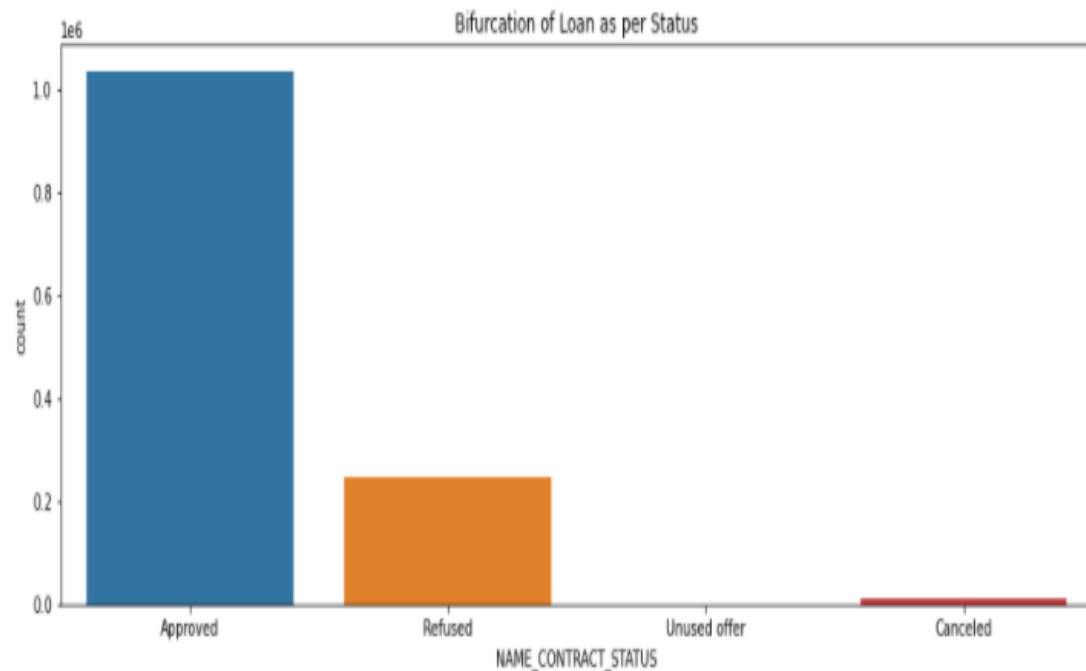
Loan status distribution across defaulter and regular Loans

As per Common understanding Loan could defaulted only when its discussed, here we could see in defaulted also we have refused, canceled, and unused offers.

B) These could be part of data correction and rechecking as it doesn't match the status definition shared.

*C) This could lead to misleading representation and analysis*

Defulter
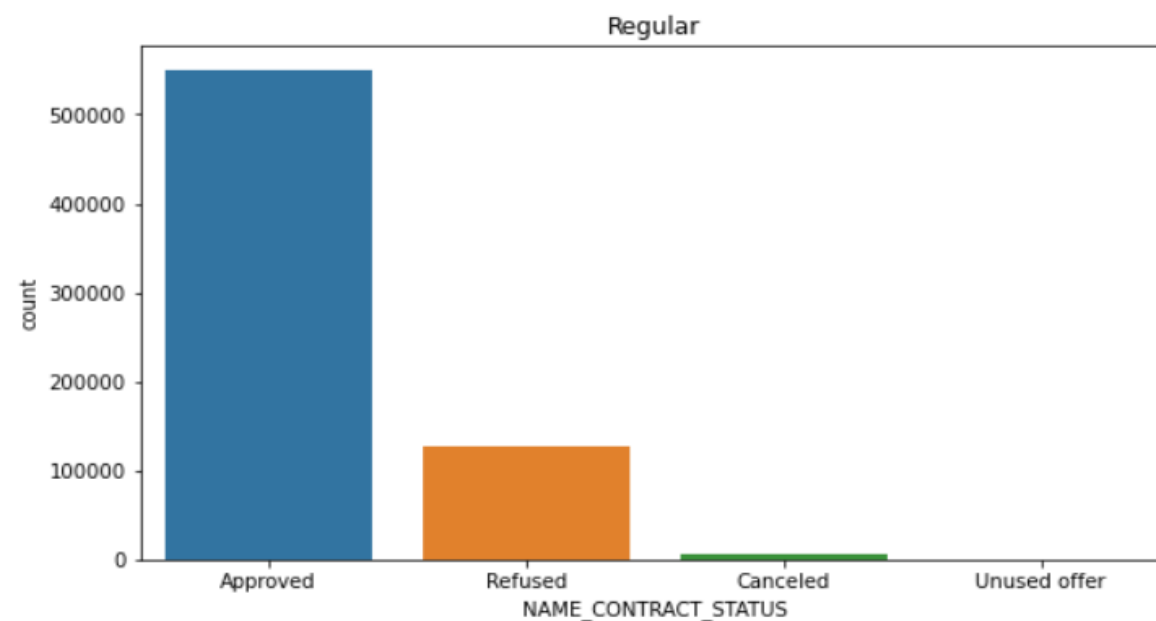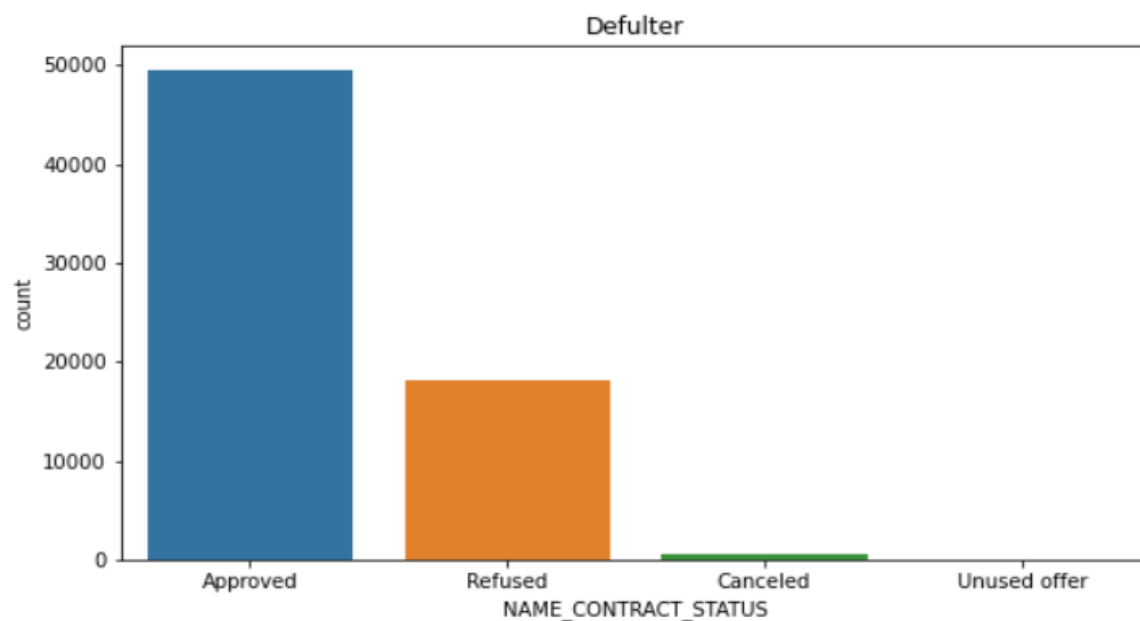
Regular

Distribution of loan status with age grouping respective of defaulter and regular customers
a) This also matches our initial analysis that majority of our customer are middle age customer.
c) As we have majority of customer are female and middle age persons , so our target customer will be female in mid age belongs to middle income group this biggest group and if we add family persons this we have a majority focus group as ideal target customer and for which we could deign special offerings and new products and services and give priority in approval .
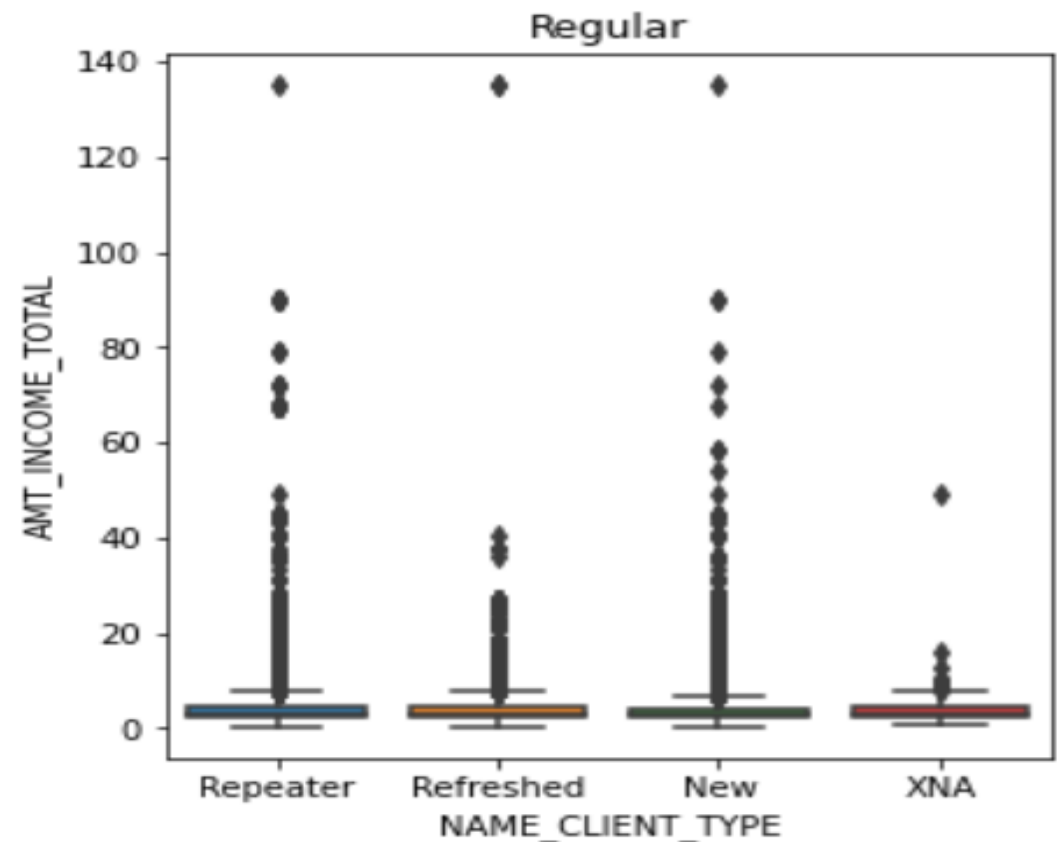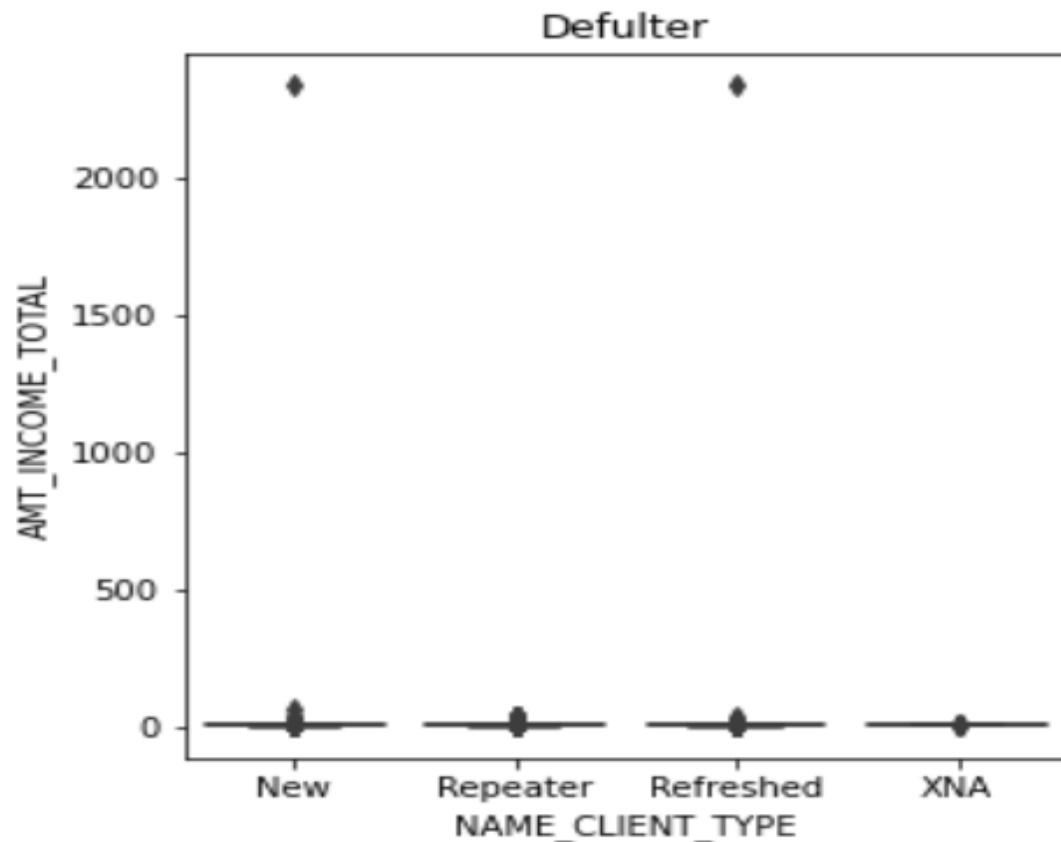c) age relation with loan status across defaulted and regular customer is identical
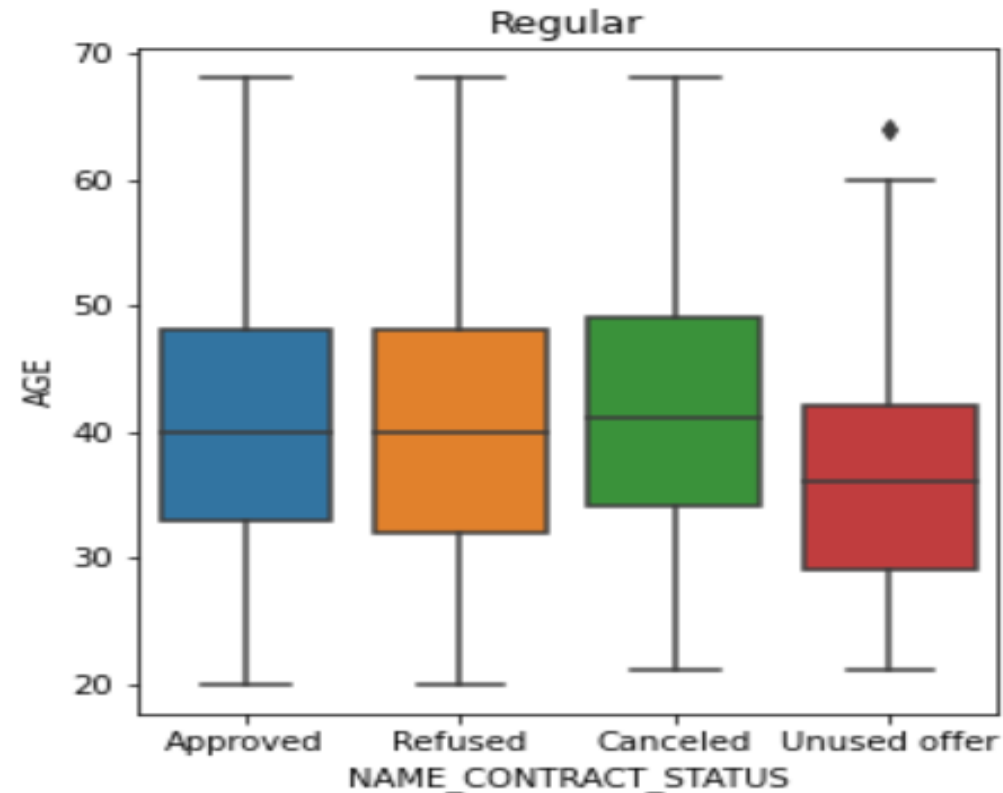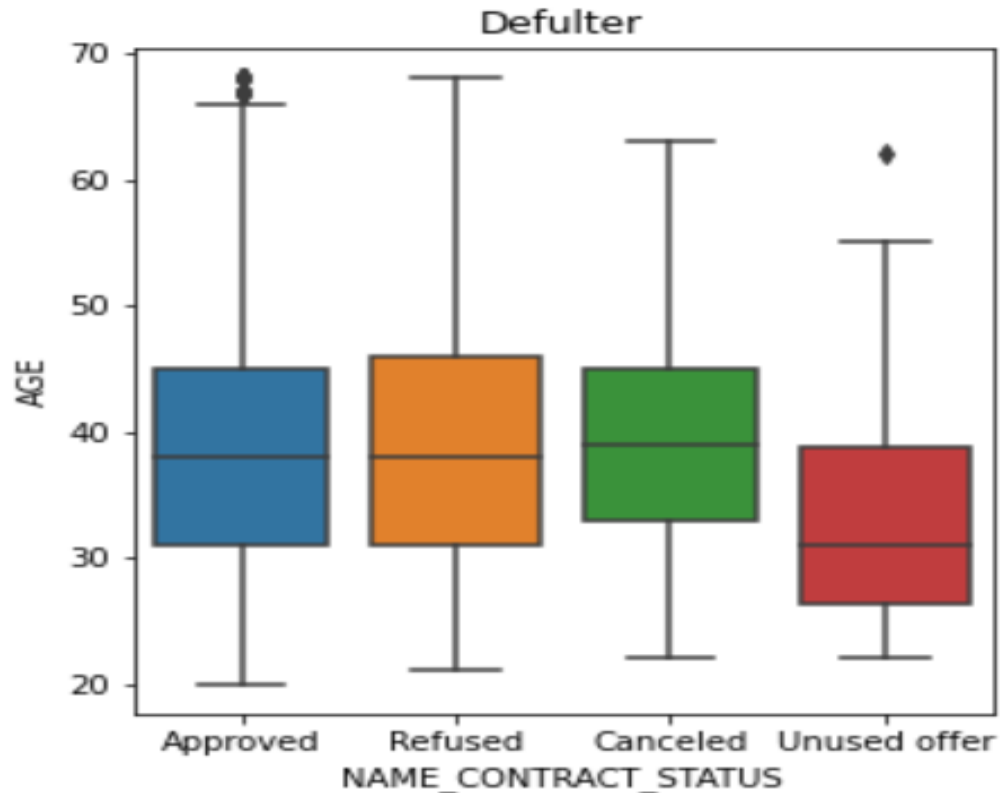
Product Type distribution across Loan status
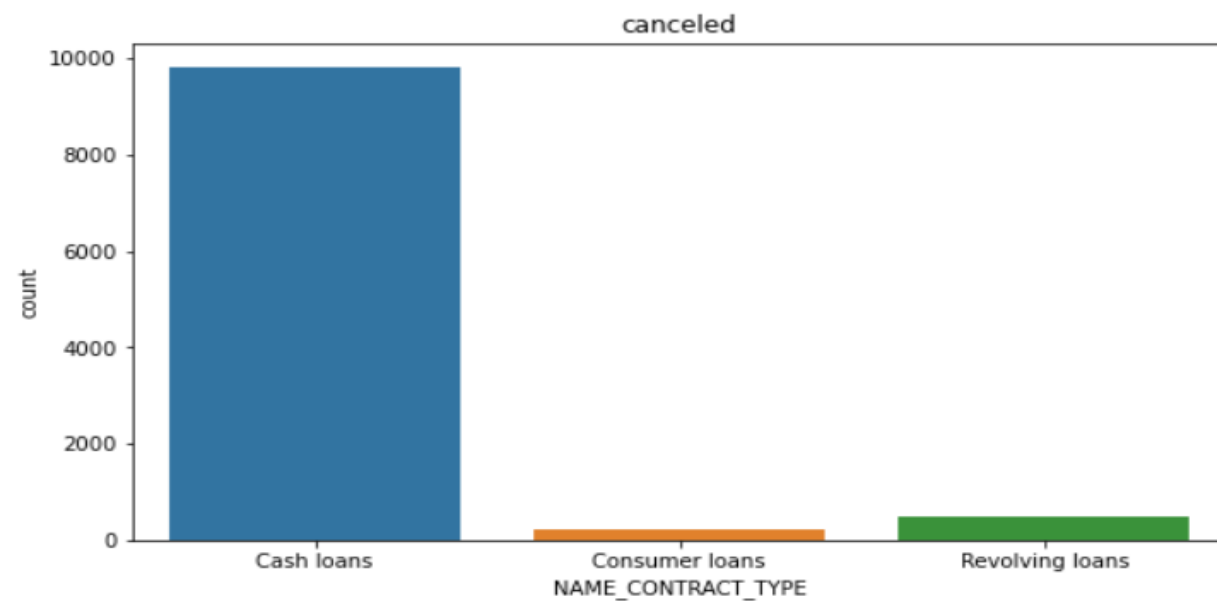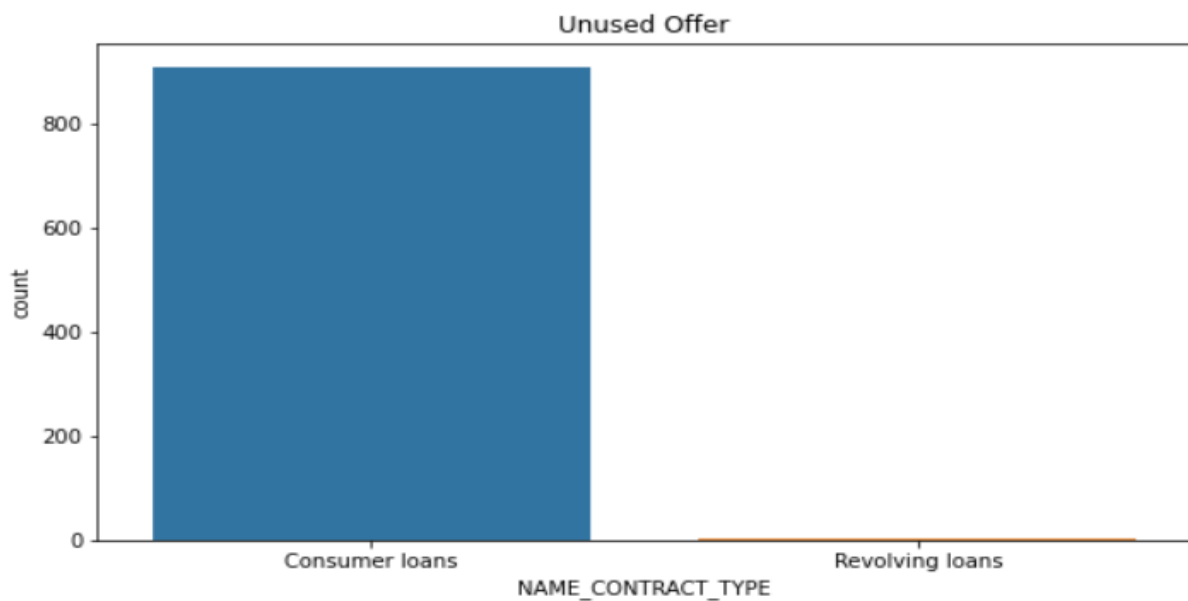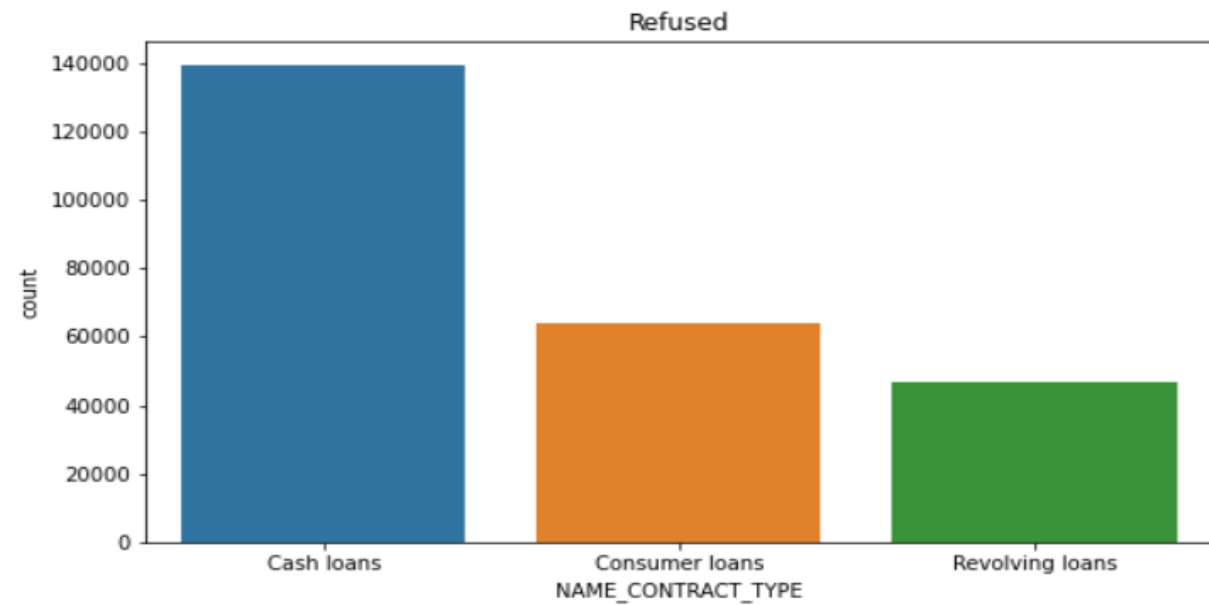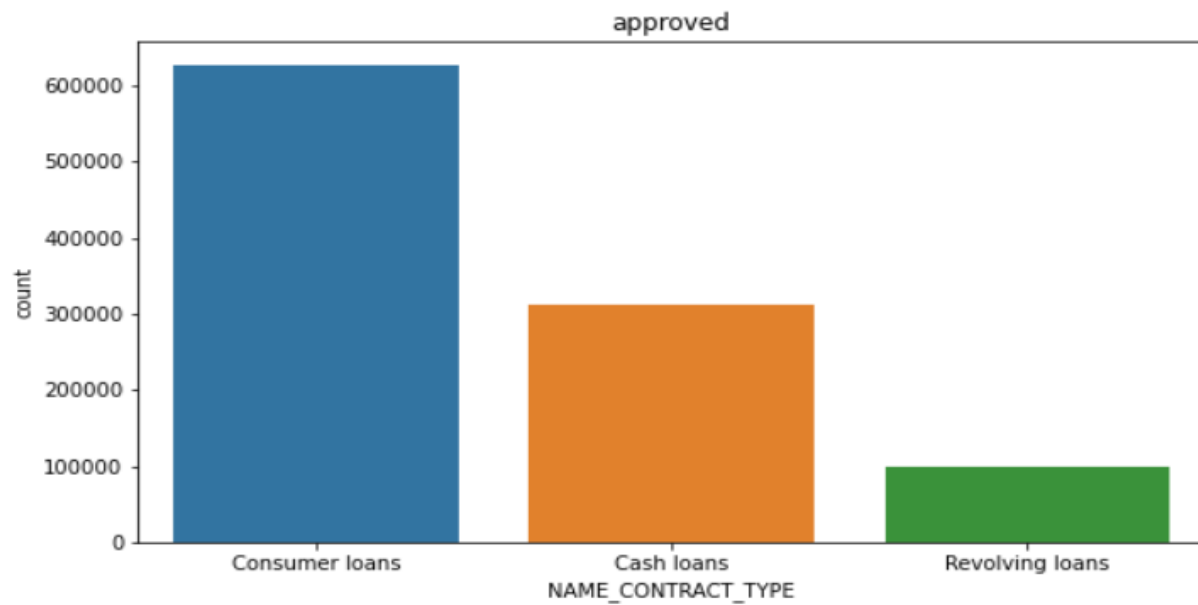
This to identify the distribution of Type of customer with defaulter and regular category with respect customer income

a) Its how's there are a wide rage of outliers
b) Majority of income are lower and middle rage
c) The distribution of income with defaulter and regular customer different.

Thank You