CREDIT RISK ANALYSIS

EDA CASE STUDY BY

-Bhargav Ram Bandi

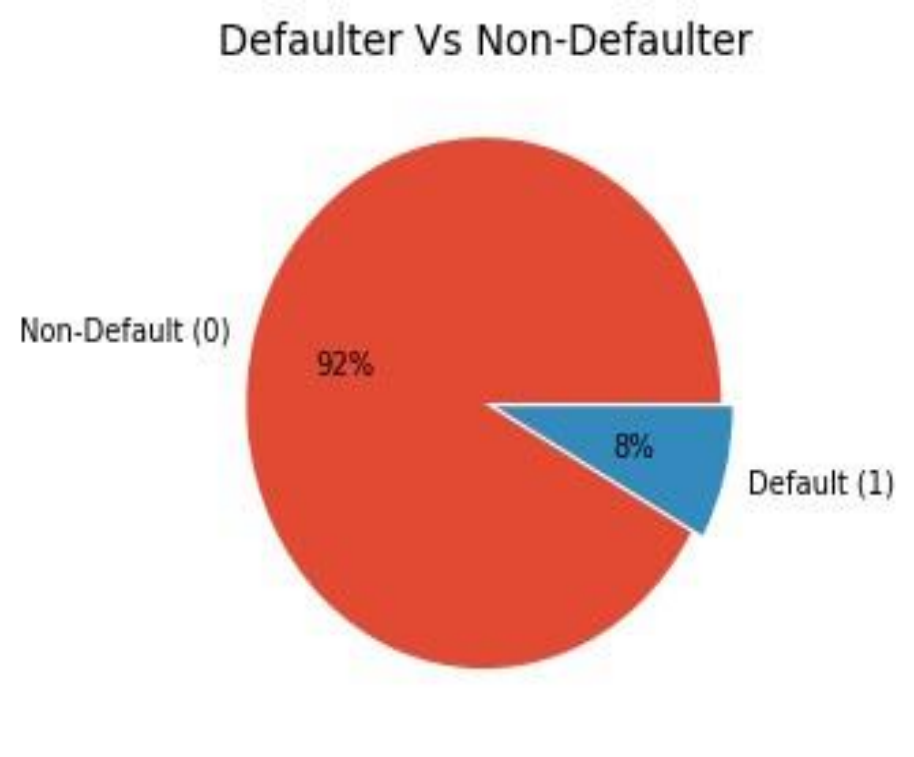
PROBLEM STATEMENT -

- Understand the working structure of Banking finance service, how bank approve, reject the loan to the customer by applying EDA techniques. Find out patterns, graphs and observe how banks minimize the risk of losing money while lending to customers.

ANALYSIS DONE ON APPLICATION DATA -

- Data Inspection
- Data Type Errors
- Check Null Values
- Check Outliers
- Check Data Imbalance
- Univariate Analysis
- Byvariate Analysis
- Multivariate Analysis

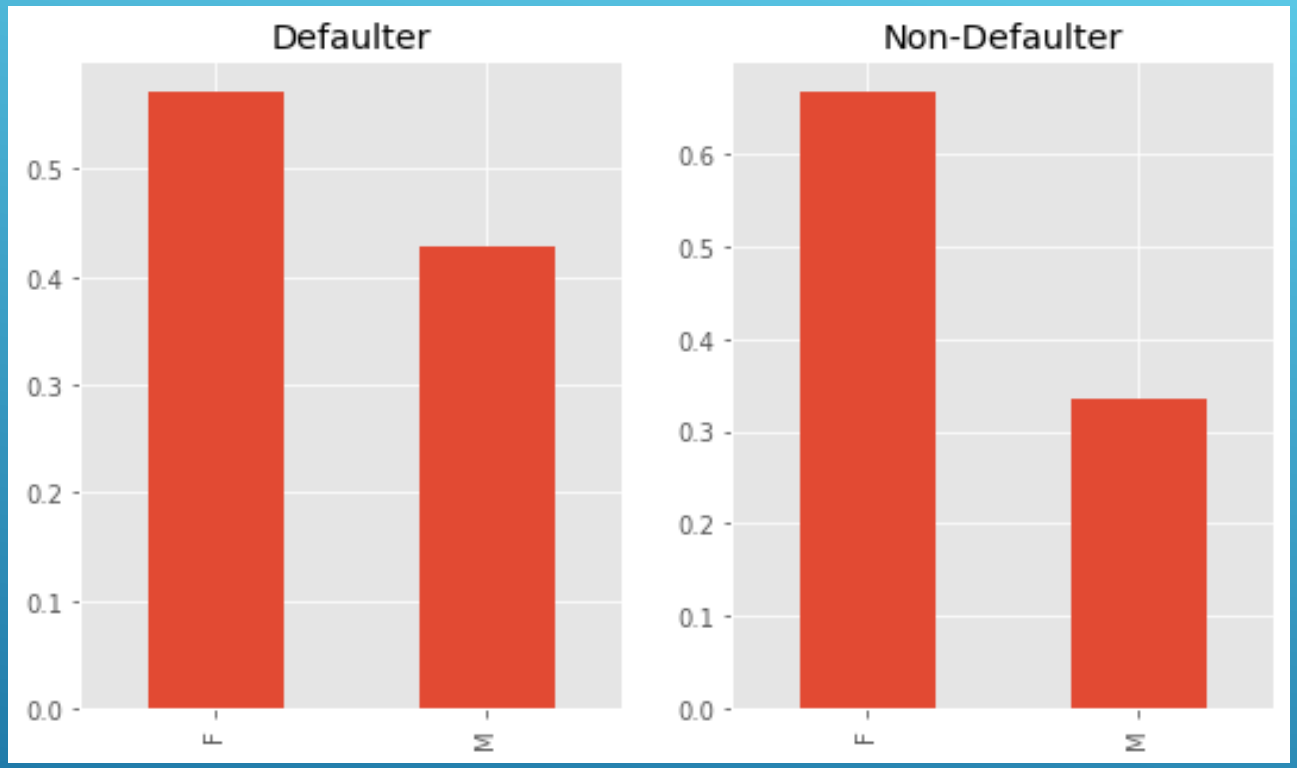
DATA IMBALANCE %



- The above pie chart clearly shows that, there is high imbalance between two TARGET variables, for Non-Defaulters it is 92% & for Defaulters it is 8%.

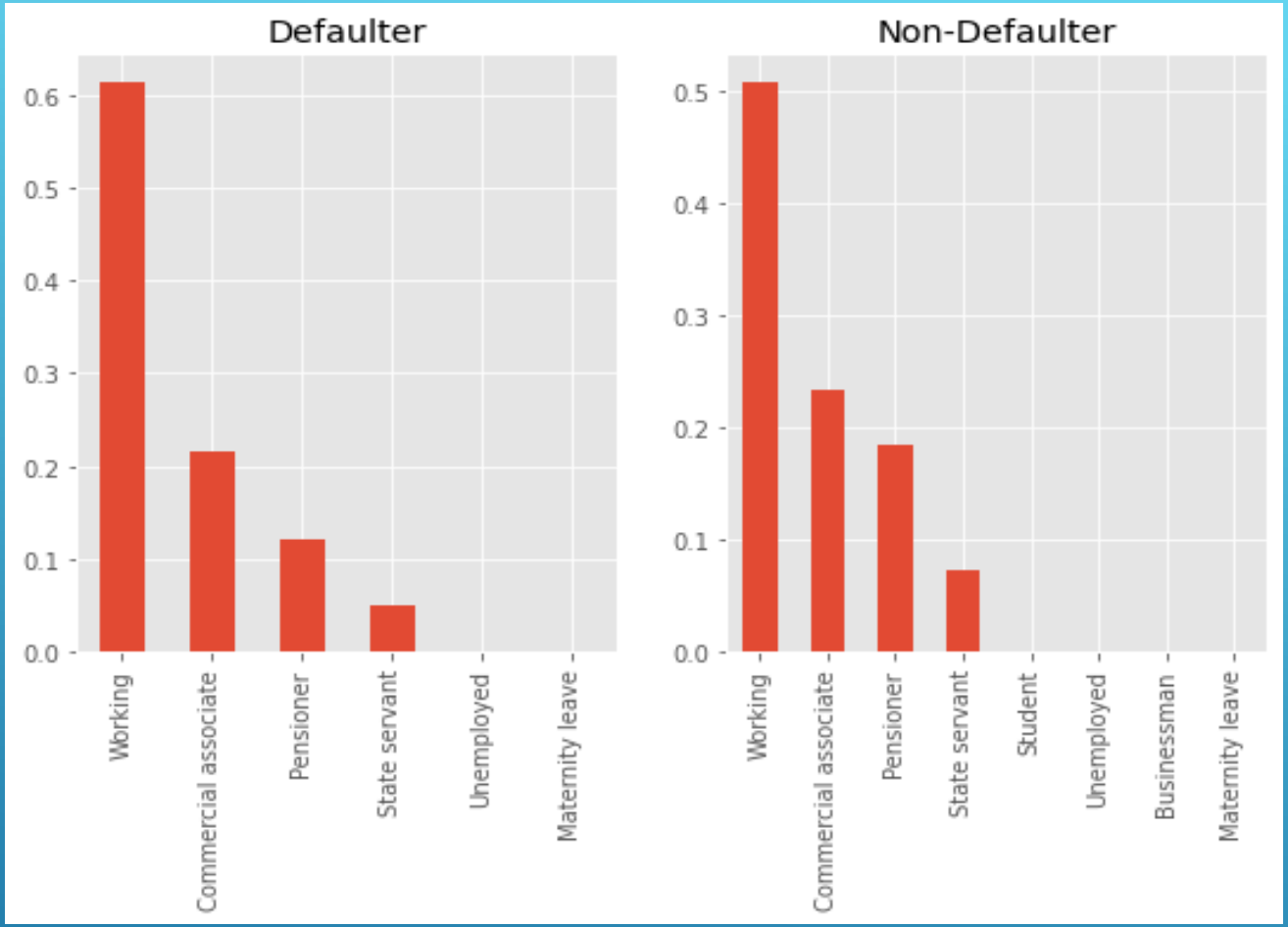
UNIVARIATE ANALYSIS

- Categorical Variable



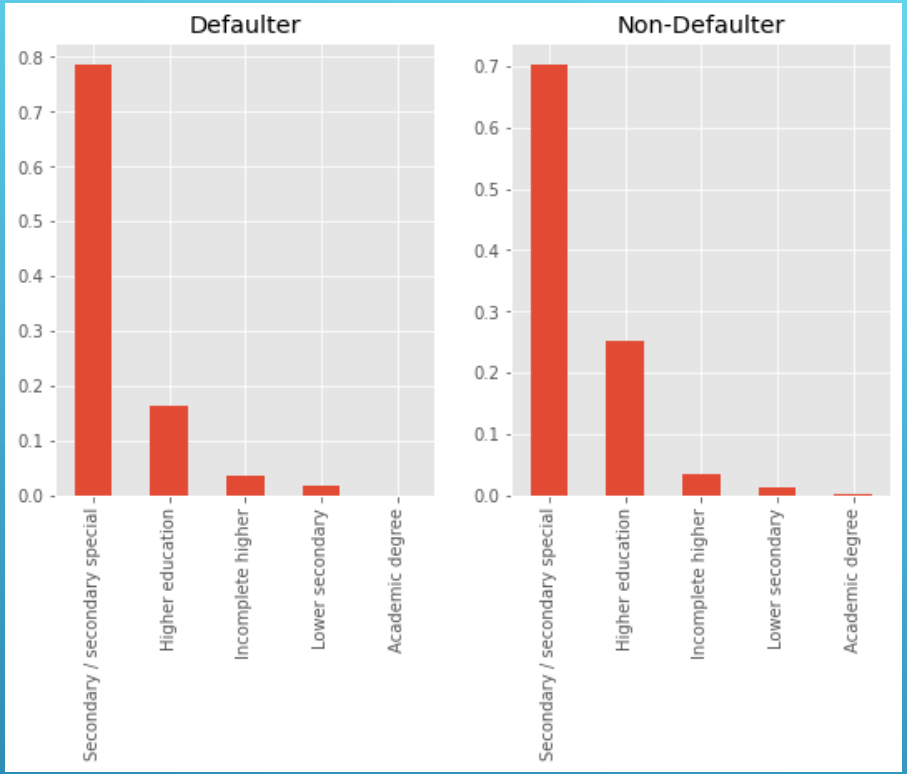
- GENDER -

  By comparing both the plots on the basis of gender, we can say Female are in majority in both the cases, Also Male % increases in Defaulter plot than Male % in Non-Defaulter plot.



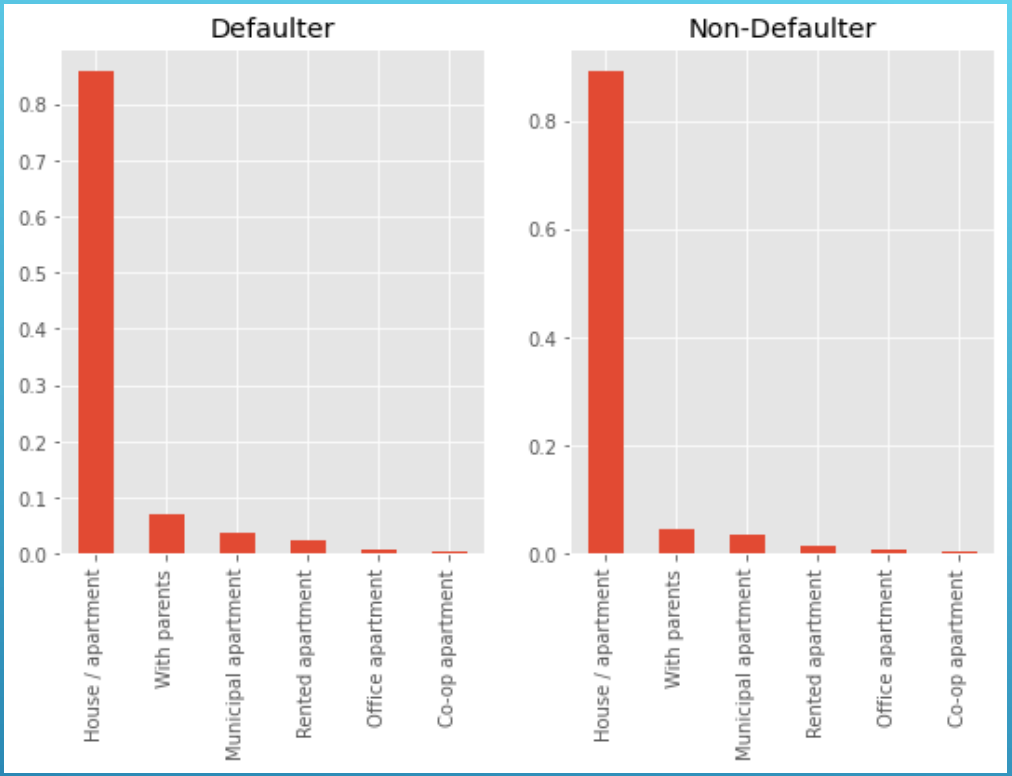
- INCOME TYPE –

  By comparing both the plots on the basis of Income Type, we can say % Non-Defaulter decrease for who are working also the % of Non-Defaulter increases for Pensioners.



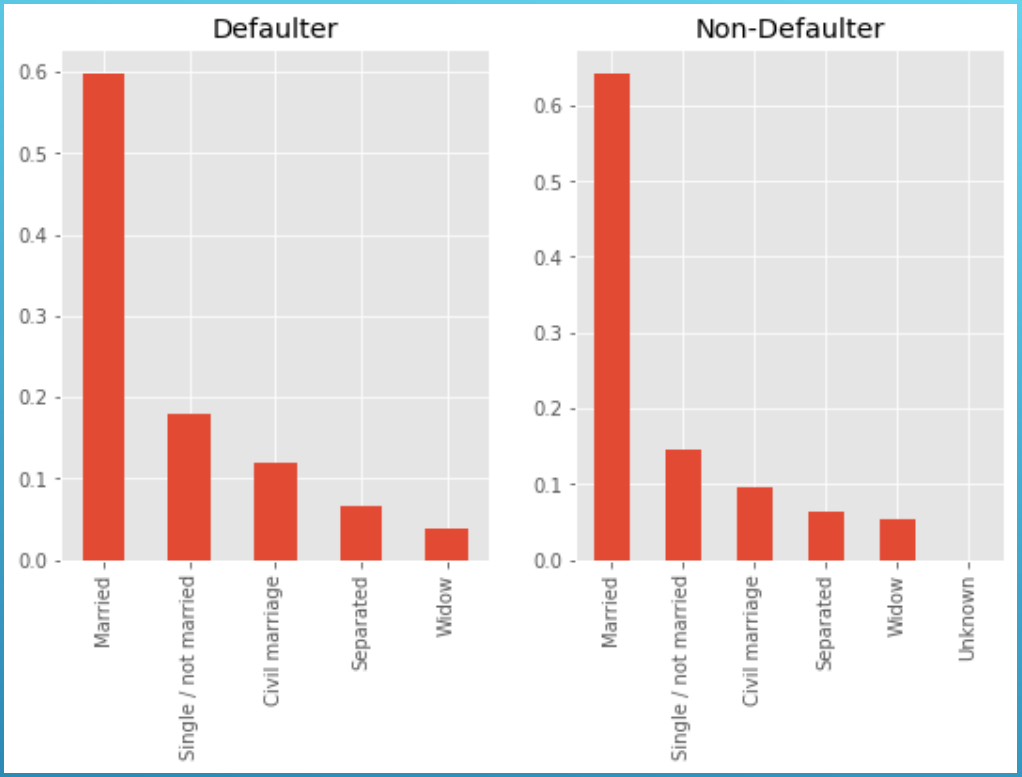
- EDUCATION -

  By comparing both the plots on the basis of Education, we can say % Non-Defaulter slightly decreases for those who are completed their Secondary/Secondary special education also the % of Non-Defaulter increases for those who completed Higher Education.



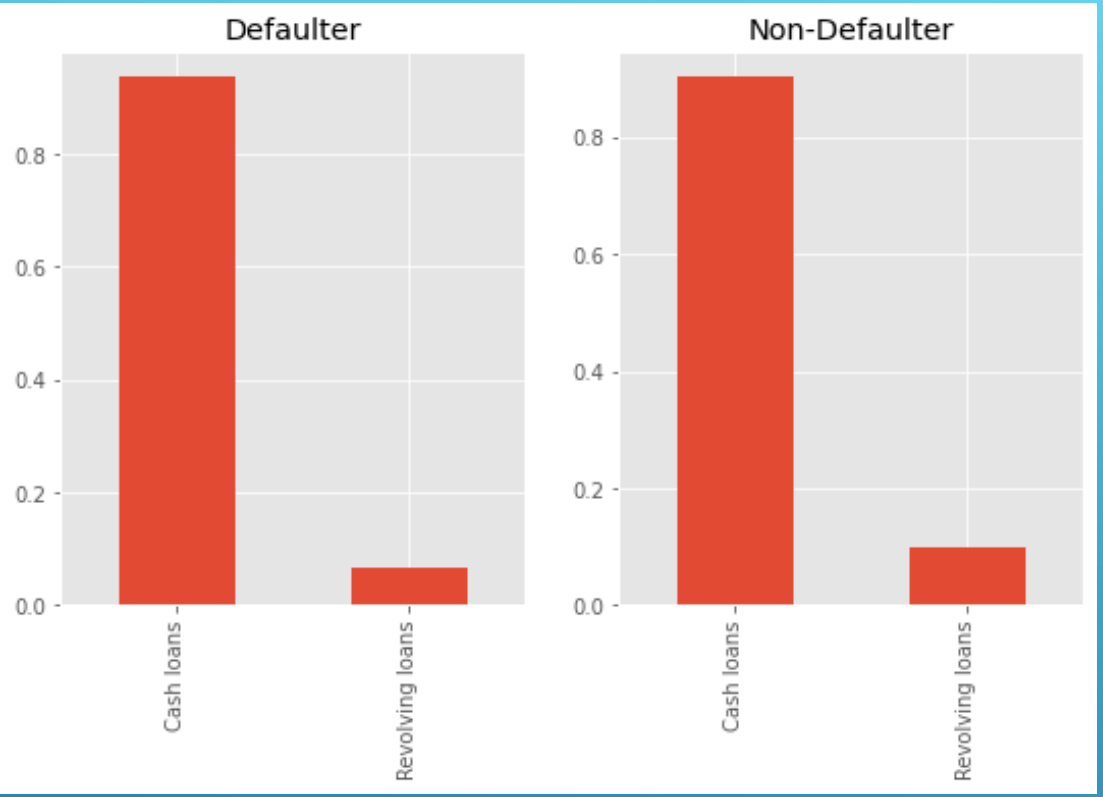
- HOUSING TYPE –

  By comparing both the plots on the basis of Housing Type, we can say % of Non-Defaulter and Defaulter remain same for those who are living in House/Apartment.

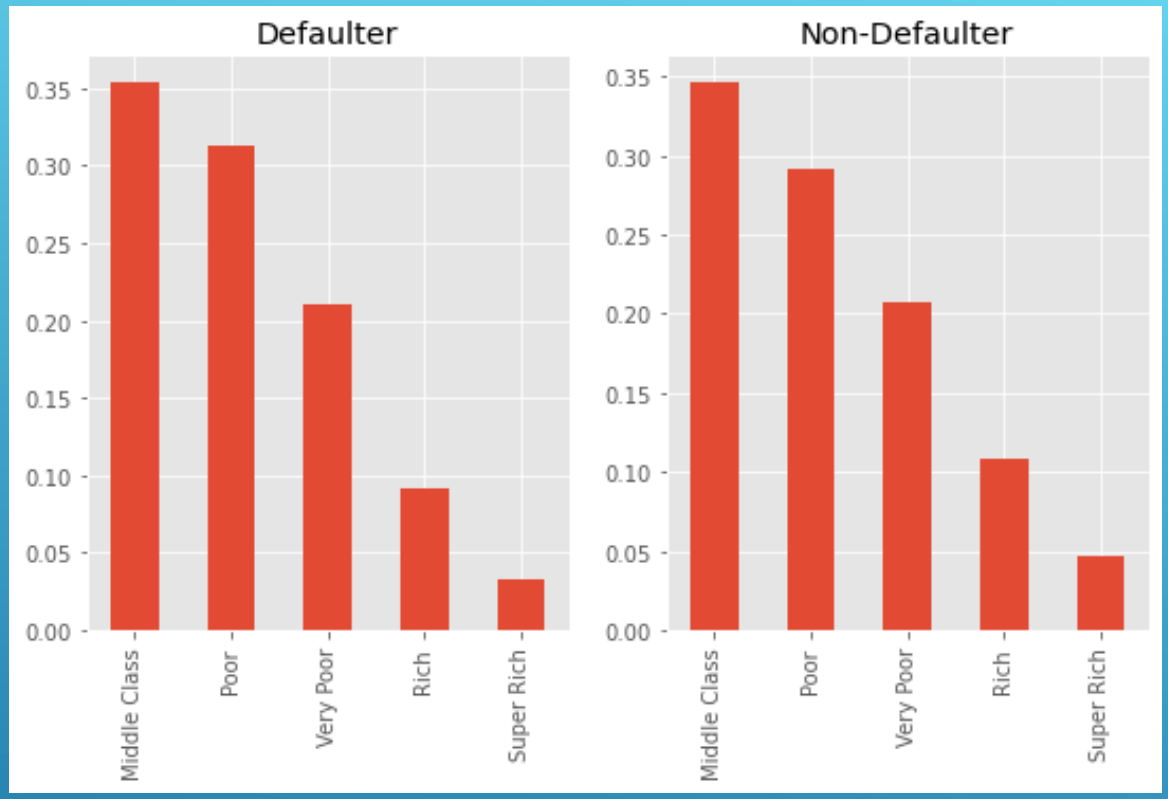


- FAMILY STATUS –

  By comparing both the plots on the basis of Family Status, we can say % of Non-Defaulter decrease for those who are done Civil marriage also the % of Non-Defaulter increases for Married people.

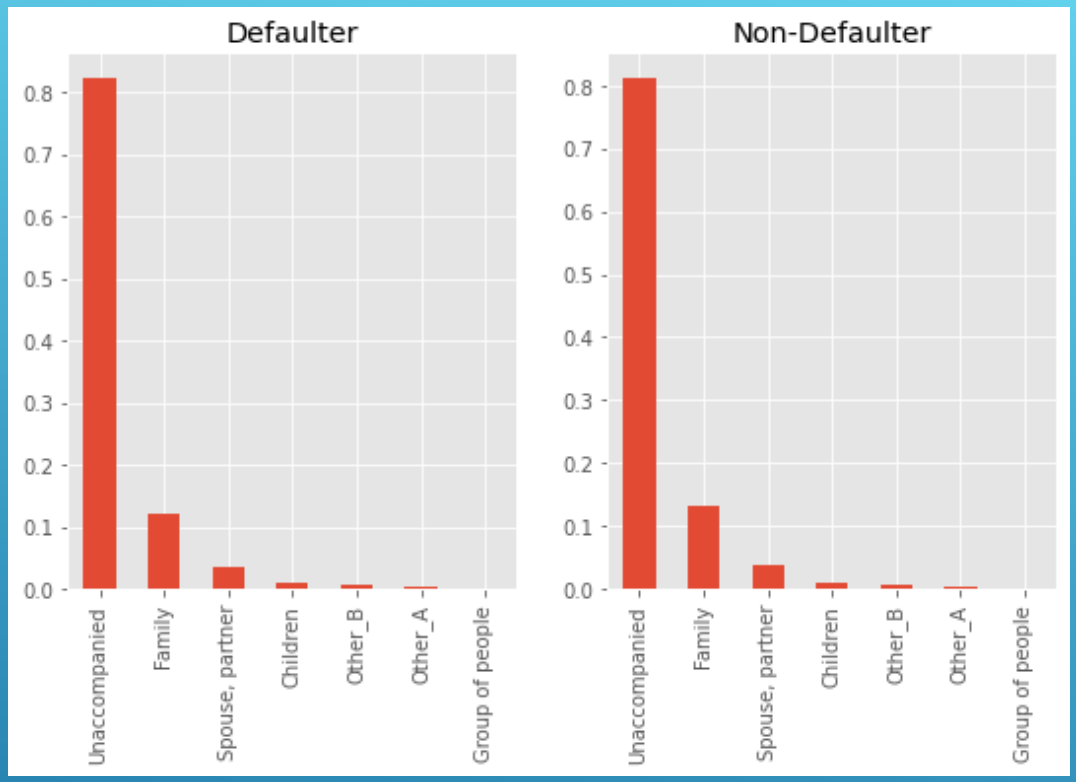


- LOAN CONTRACT TYPE –
  By comparing both the plots on the basis of Loan Contract Type, we can say % of Non-Defaulter and Defaulter remain same for Cash Loan Type also the % of Non-Defaulter increases for Revolving Loan.



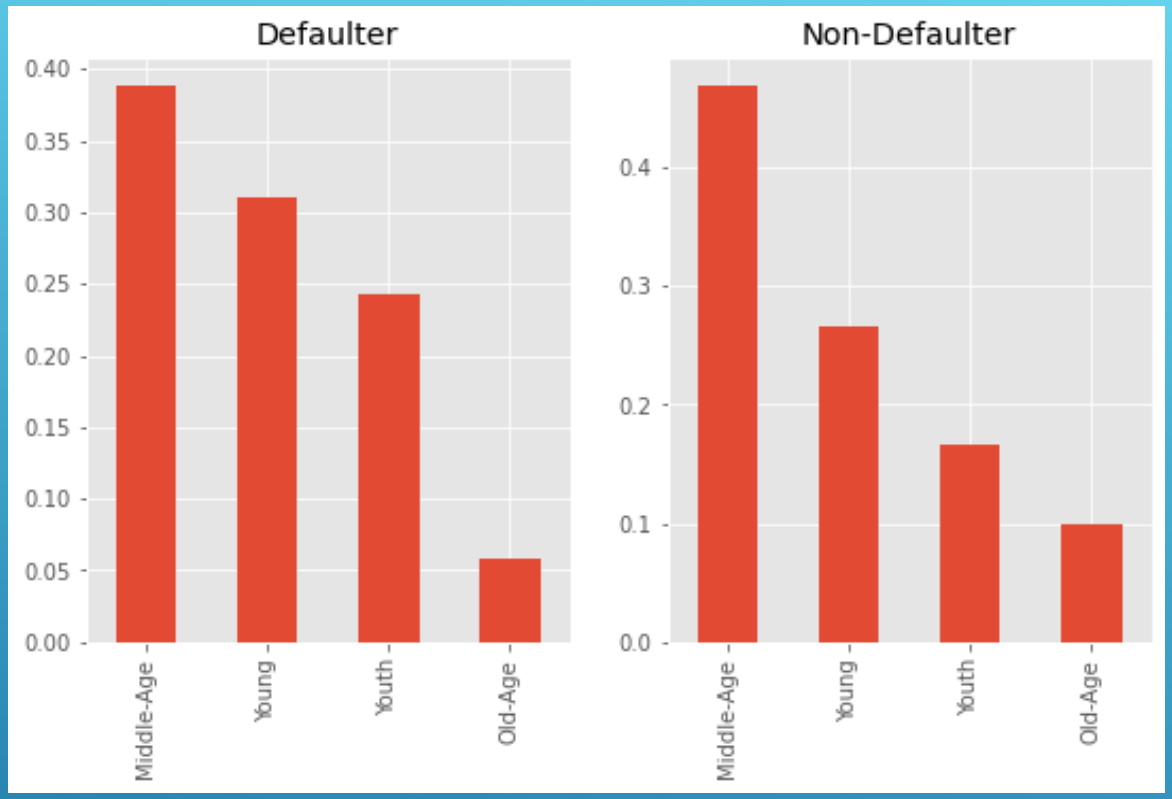
- INCOME RANGE –

  By comparing both the plots on the basis of Income Range, we can say % of Non-Defaulter decrease for those who are Poor people, also the % of Non-Defaulter increases for Super Rich people.



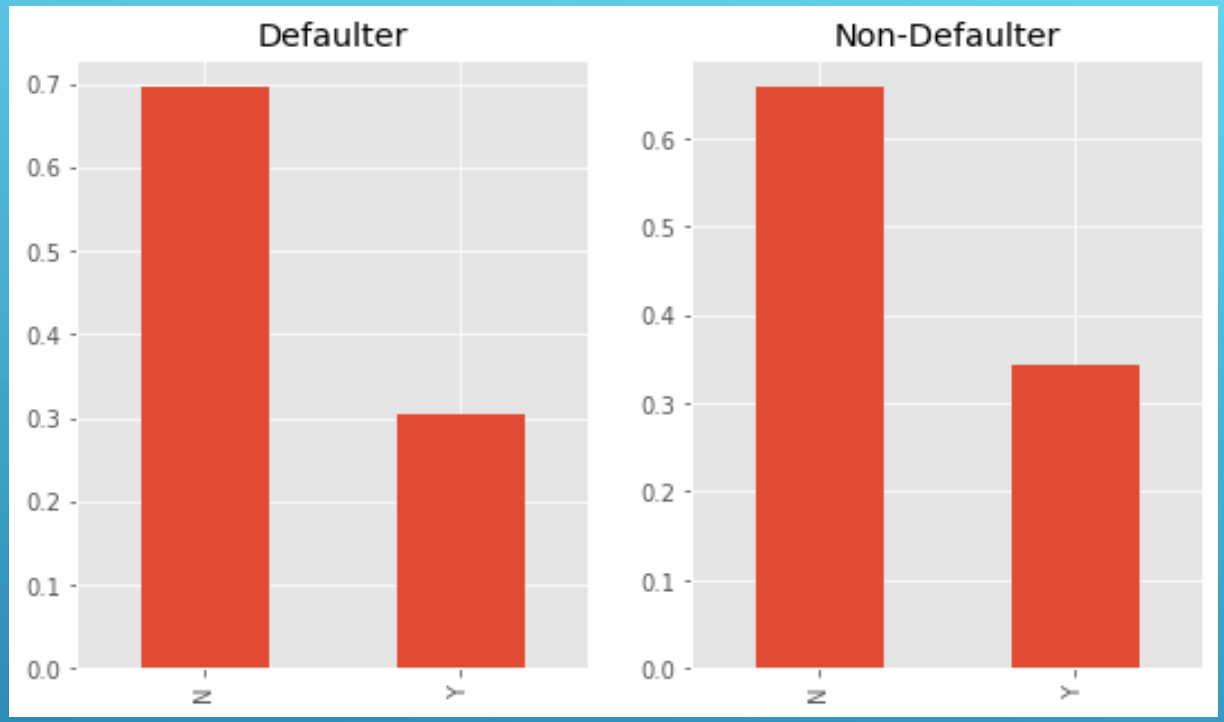
- NAME SUITE TYPE –

  By comparing both the plots on the basis of Suite Type, we can say % of Non-Defaulter and Defaulter for all type remain the same.



- AGE RANGE –

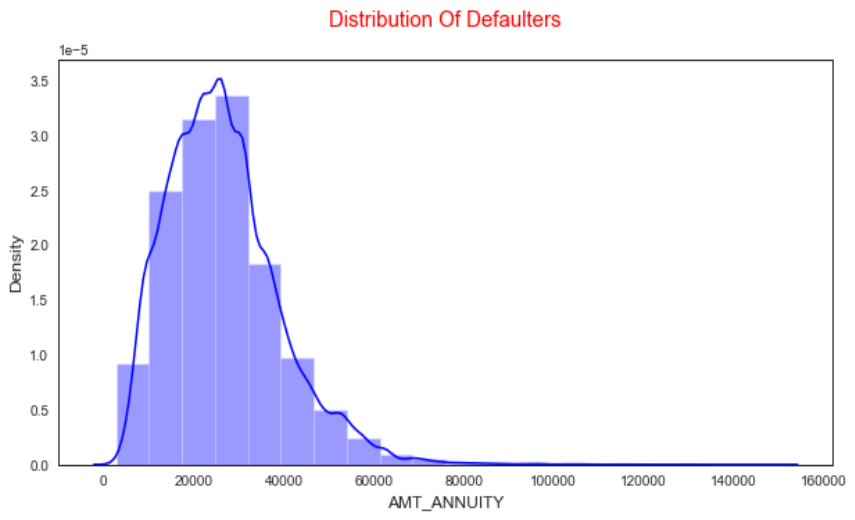
  By comparing both the plots on the basis of Age, we can say % of Non-Defaulter decreases for Young people, also the % of Non-Defaulter increases for Middle Age people & Old age people.

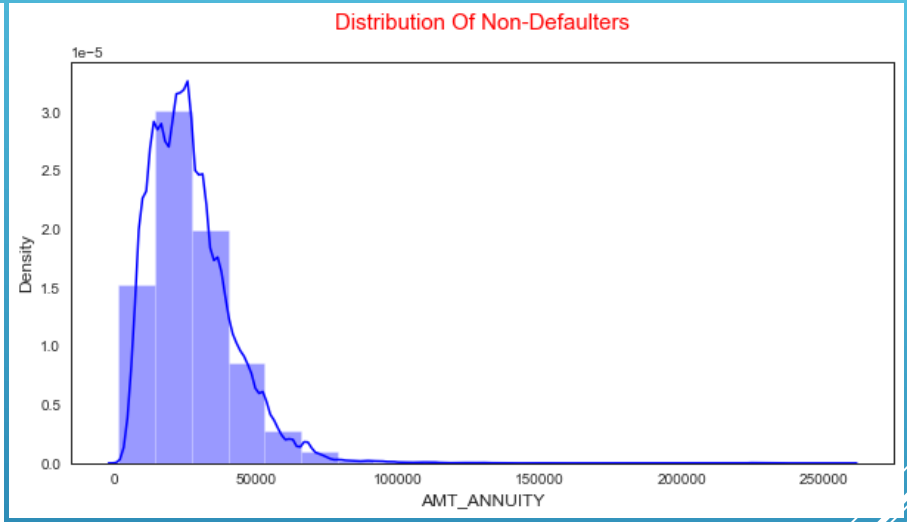


- FLAG OWN CAR –

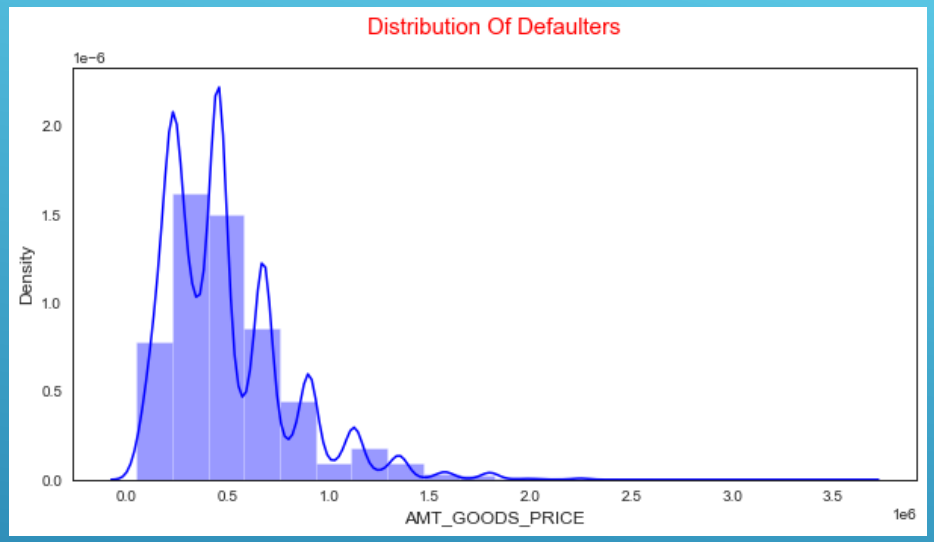
  By comparing both the plots on the basis of Own Car, we can say % of Non-Defaulter slightly decreases for those who have No Car, also the % of Non-Defaulter increases for those people who have Own Car.

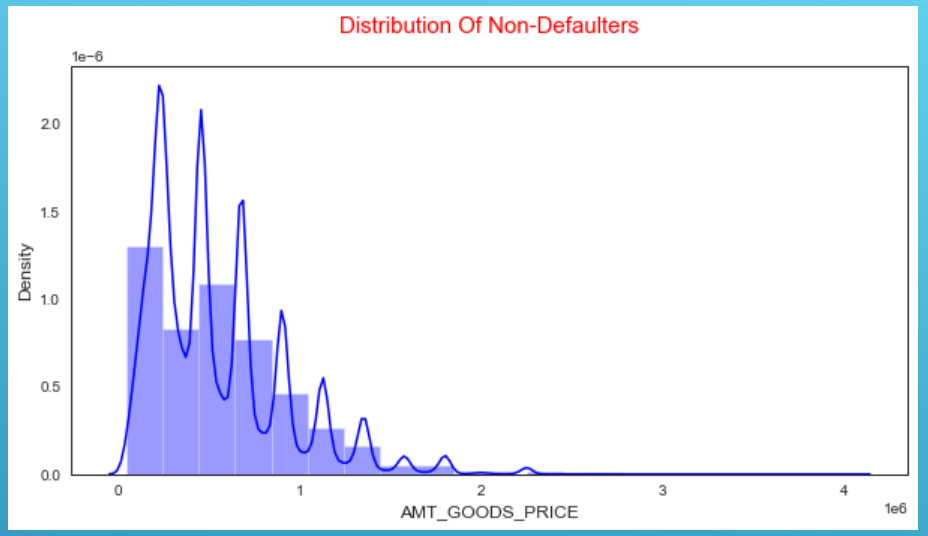
➢ Numerical Variable



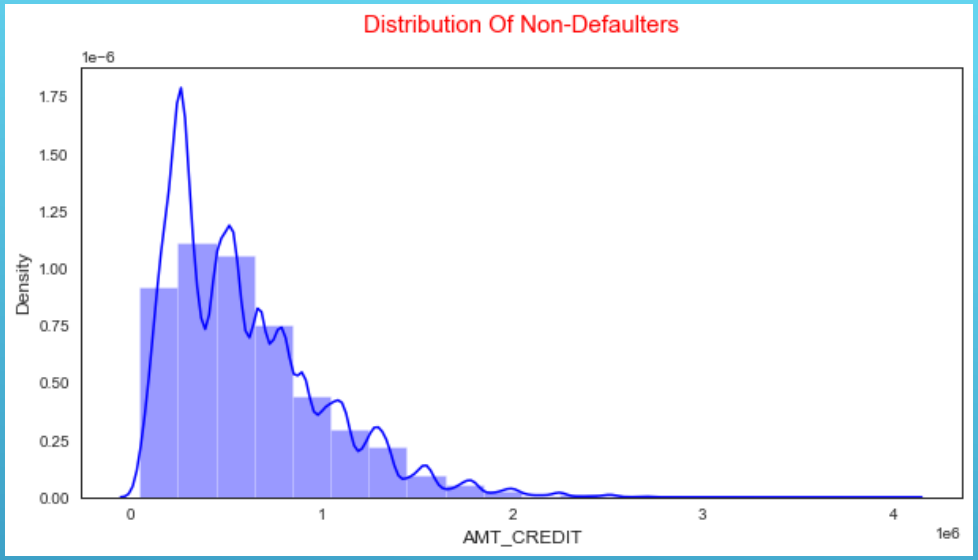


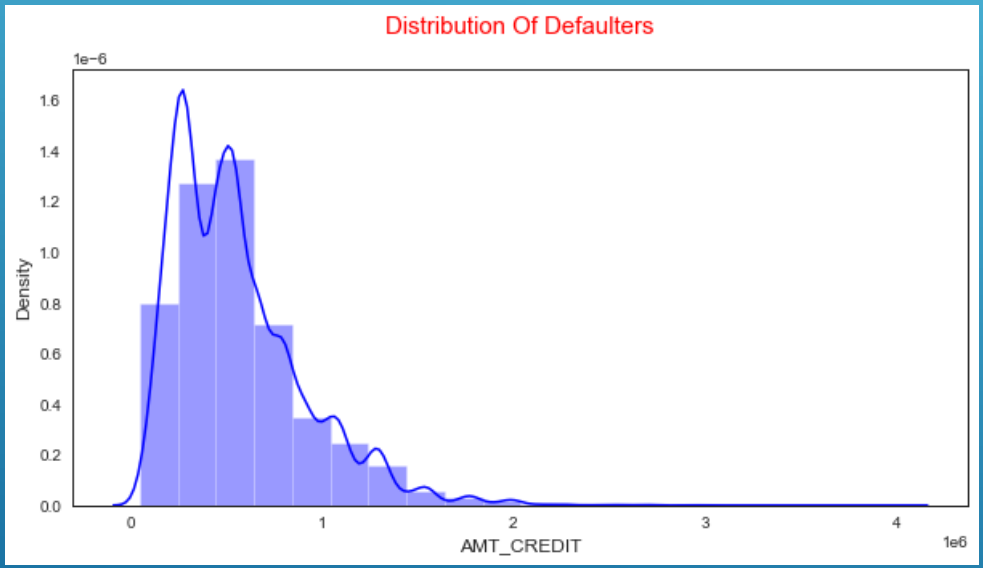
- LOAN ANNUITY –
  We can observe from both the plots that most of the clients are from first quartile than third quartile.





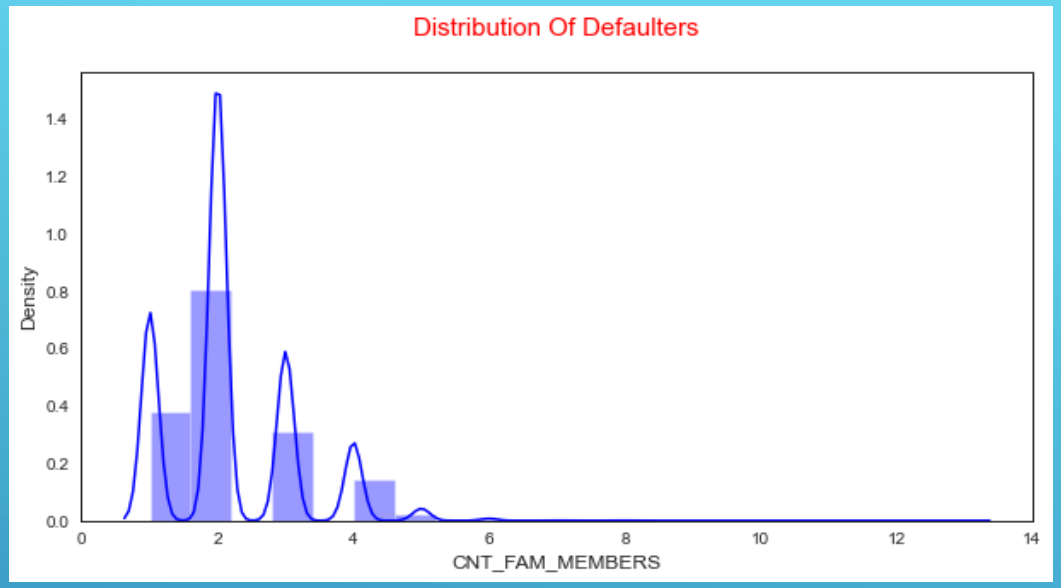
- GOODS PRICE AMOUNT –
  We can observe from both the plots that most of the clients are from first quartile than third quartile.

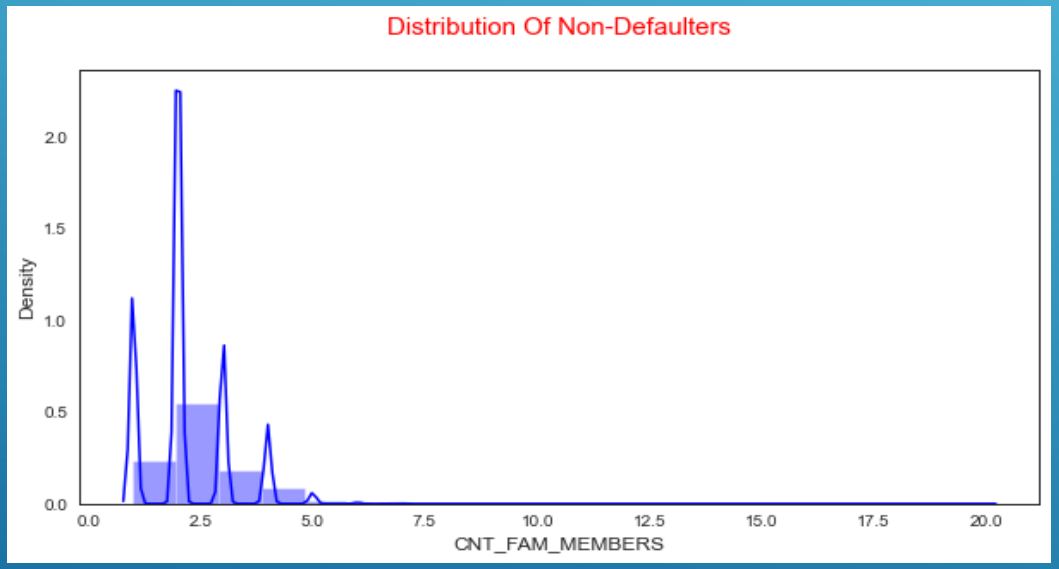




- CREDIT AMOUNT –

  We can observe from both the plots that most of the clients are from first quartile than third quartile.



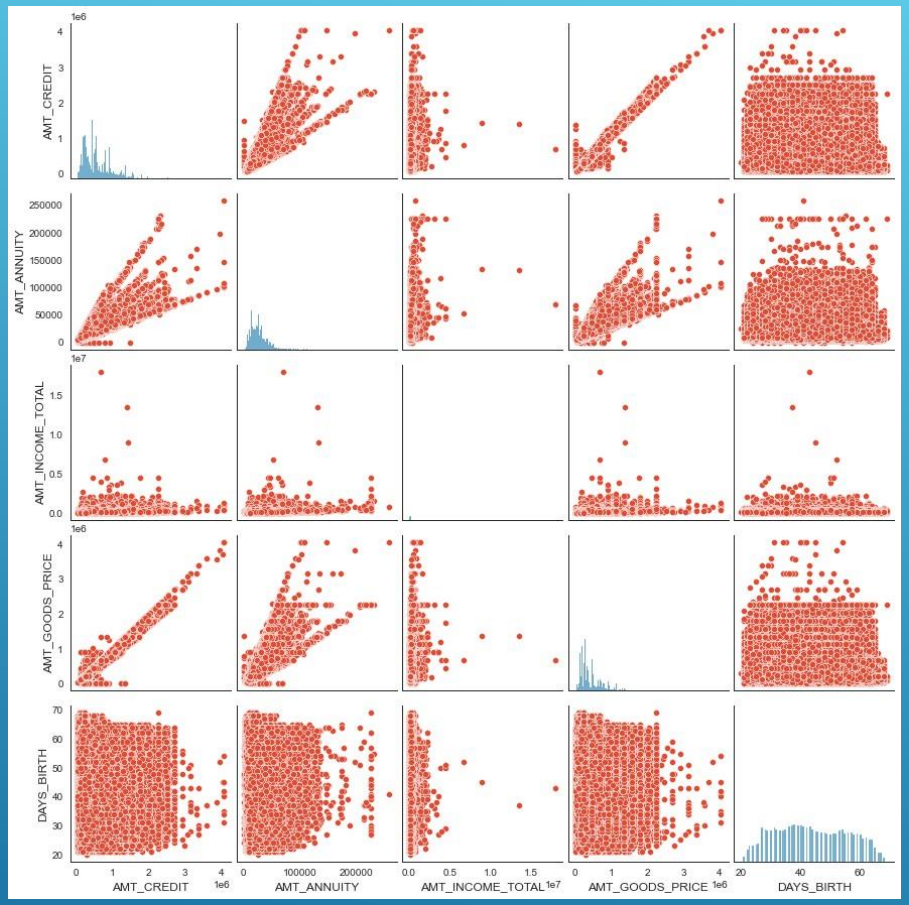


- FAMILY MEMBERS –

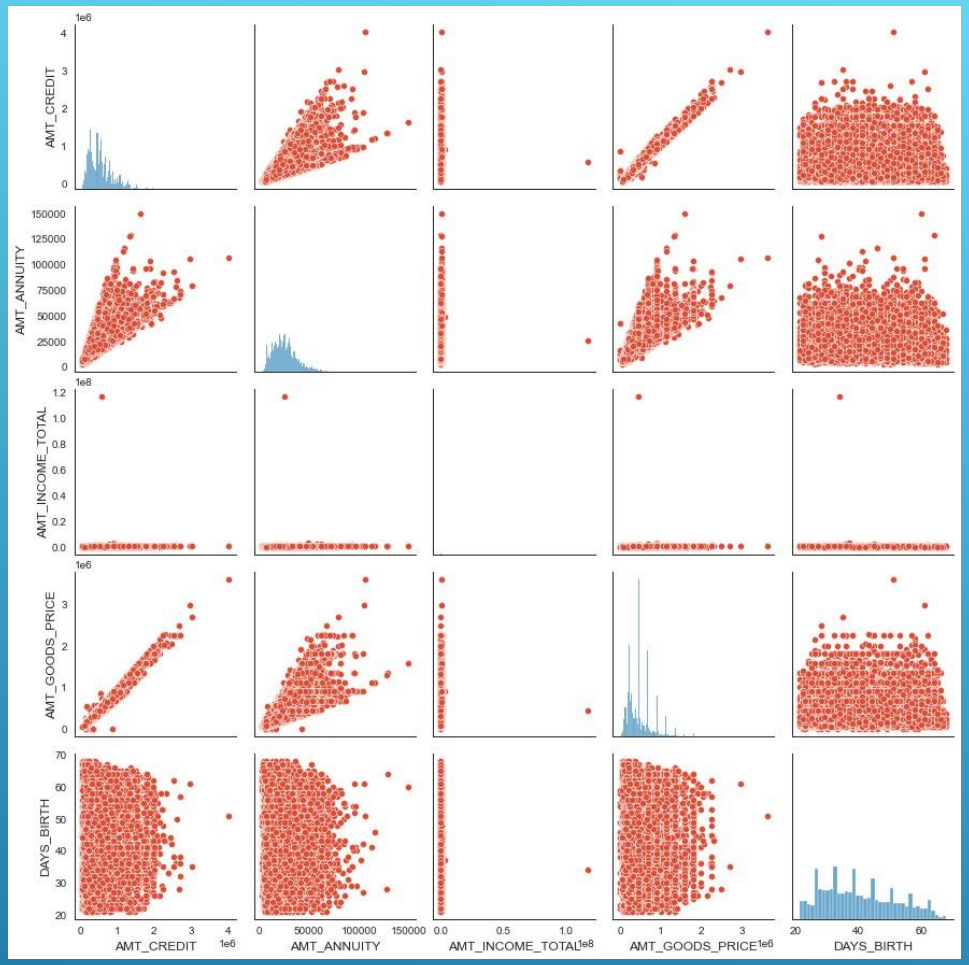
  We can observe from both the plots that most of the clients are from first quartile than third quartile.

BIVARIATE ANALYSIS

- Numerical Vs Numerical

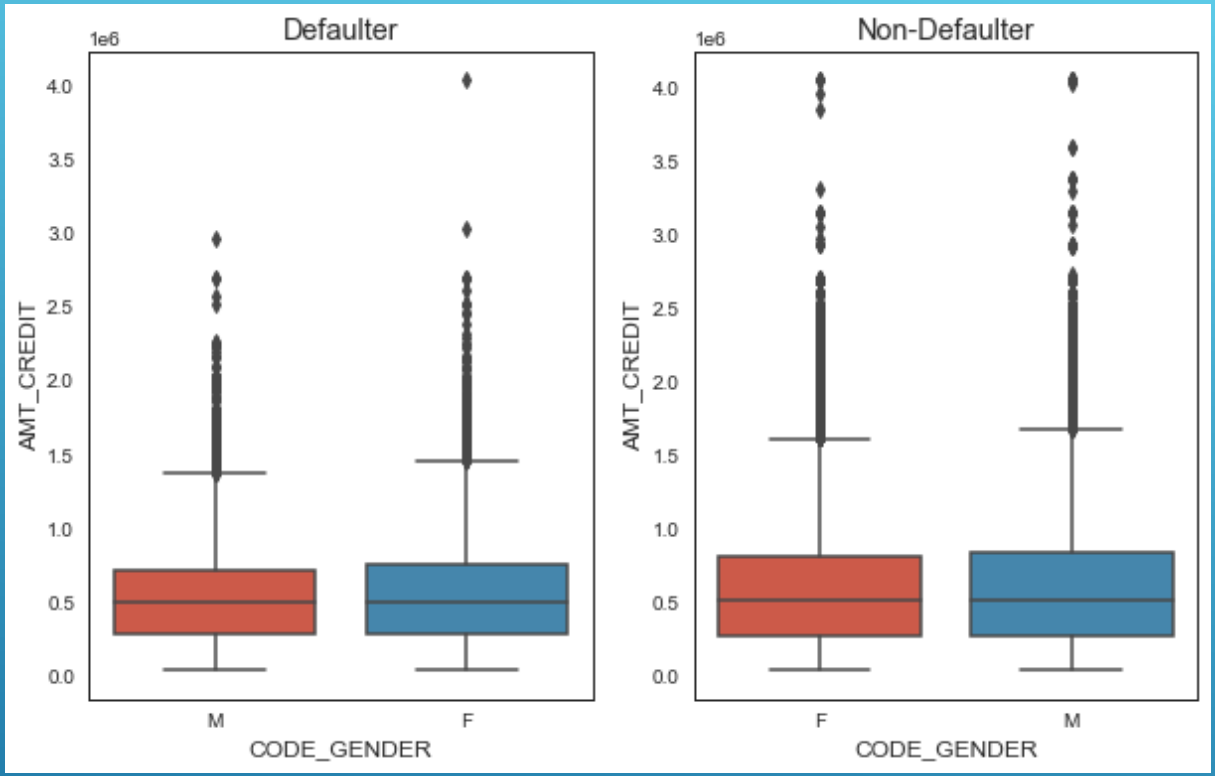


- We can say by observing above pair plot for Non-Defaulters that the density in the lower left corner is similar in all the cases, so the people are equally likely to Non-Default if the AMT_INCOME_TOTAL is small and the AMT_CREDIT is low.



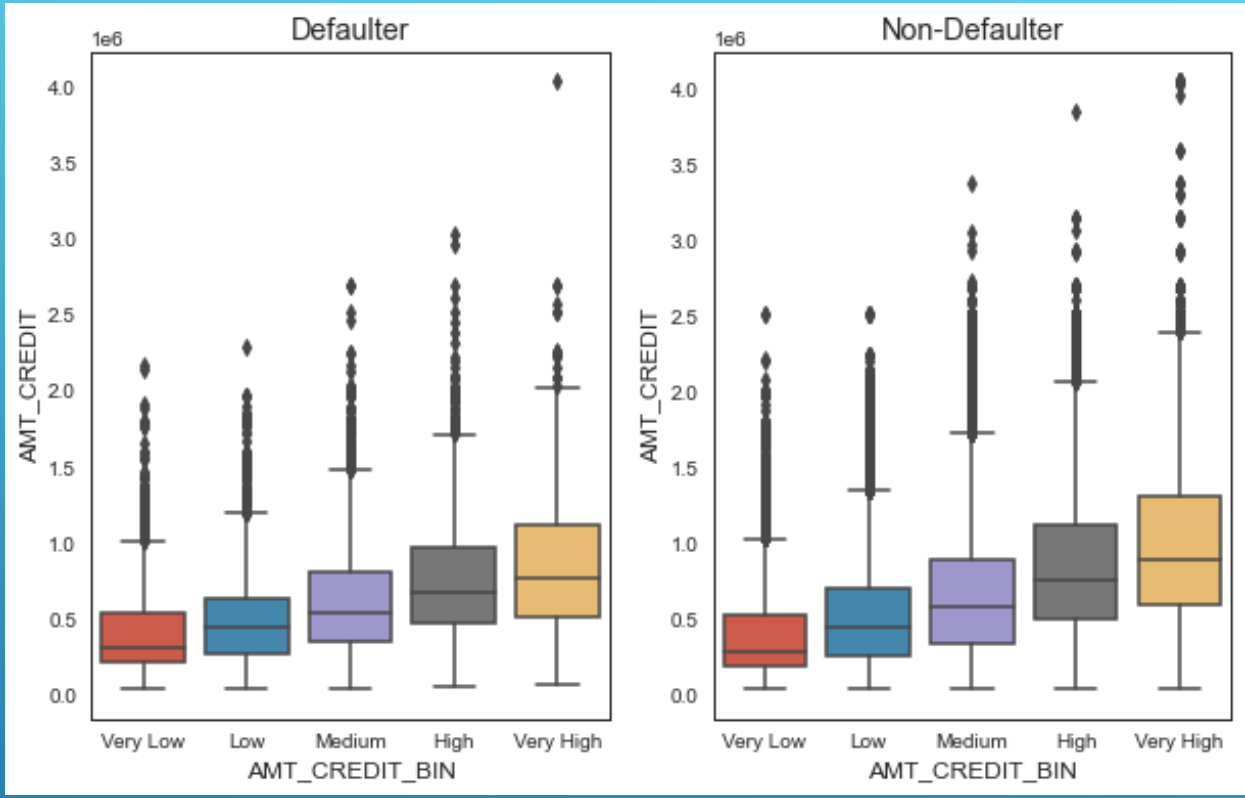
- We can say by observing above pair plot for Defaulters that the density in the lower left corner is almost similar in all the cases, so the people are equally likely to default if the AMT_INCOME is small also the AMT_CREDIT is low.

➢ Categorical Vs Numerical

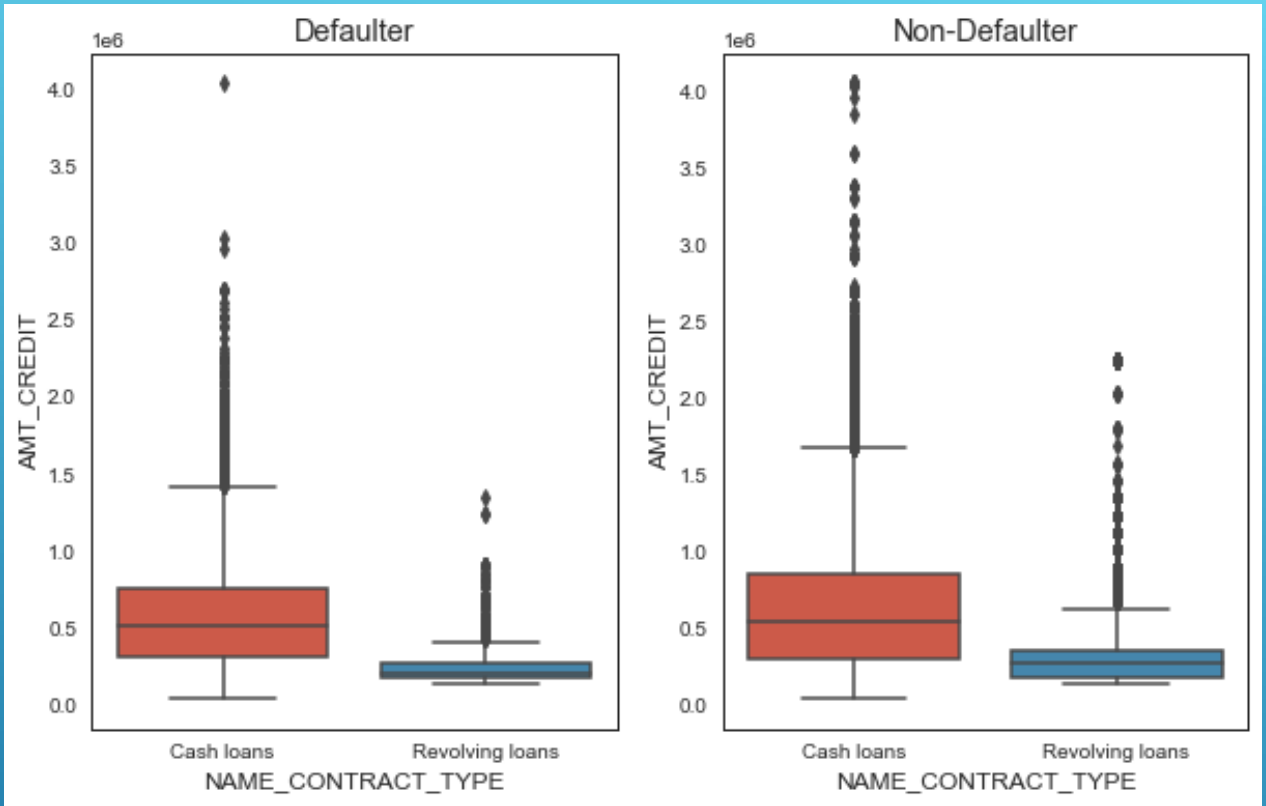


❖ GENDER Vs CREDIT AMOUNT –

By comparing above two plots on the basis of Gender and Credit Amount, The % of Non-Defaulter is increase for both Male and Female as compare to % of Defaulter.



- CREDIT AMOUNT RANGE Vs CREDIT AMOUNT –
  By comparing above two plots on the basis of Credit Amount Range and Credit Amount, The % of Non-Defaulter is increasing for all the Credit Amount Range as compare to % of Defaulter.



- CONTRACT TYPE Vs CREDIT AMOUNT –
  By comparing above two plots on the basis of Contract Type and Credit Amount, The % of Non-Defaulter increases for Cash Loan and Revolving Loan as compare to % of Defaulter.

MULTIVARIATE ANALYSIS

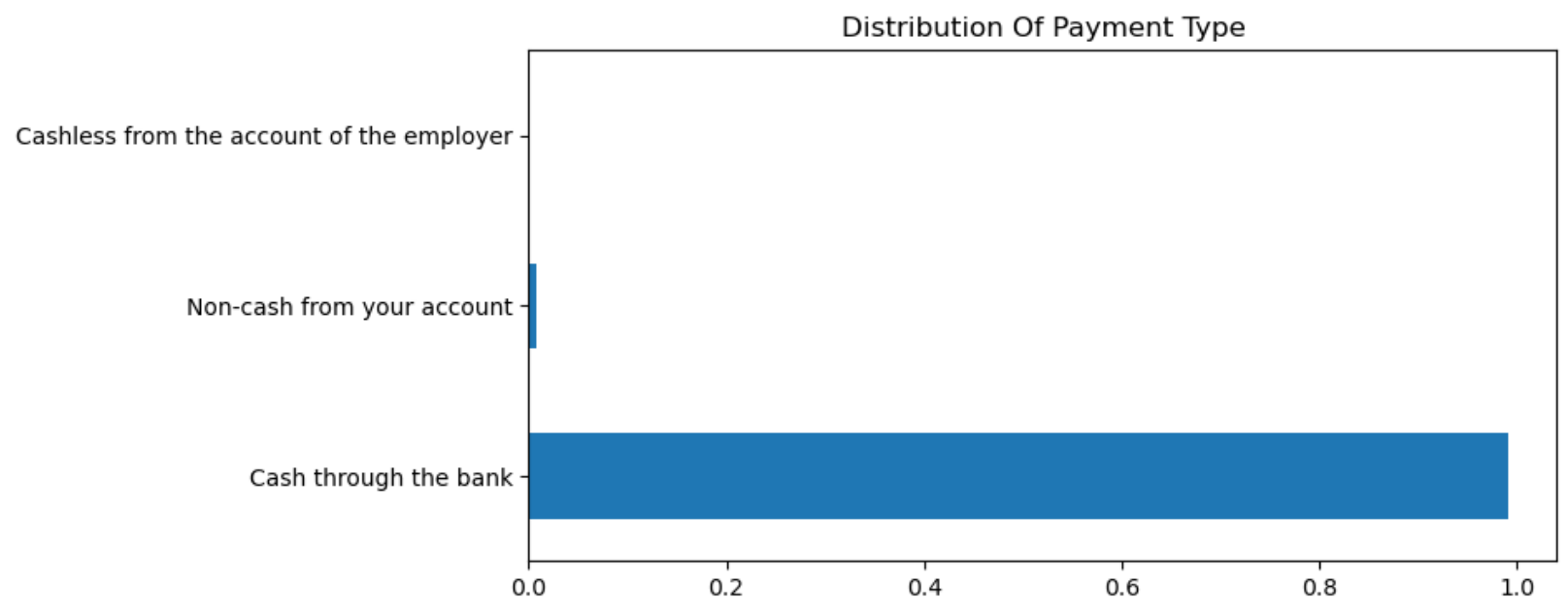
- By observing above heatmap we can say that there is a high correlation between Maternity leave income type and Poor income range.

ANALYSIS DONE ON PREVIOUS APPLICATION DATA

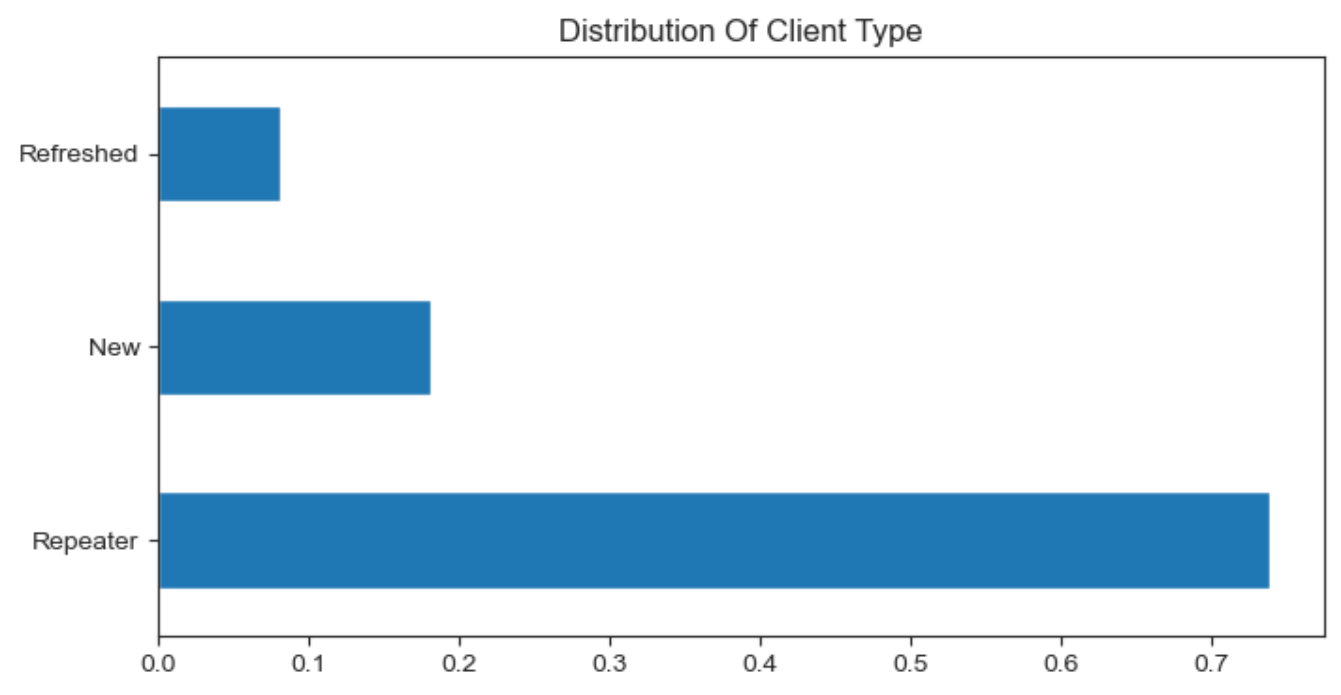
- Data Inspection
- Data Type Errors
- Check Null Values
- Check Outliers
- Check Data Imbalance
- Univariate Analysis
- Byvariate Analysis
- Multivariate Analysis

UNIVARIATE ANALYSIS

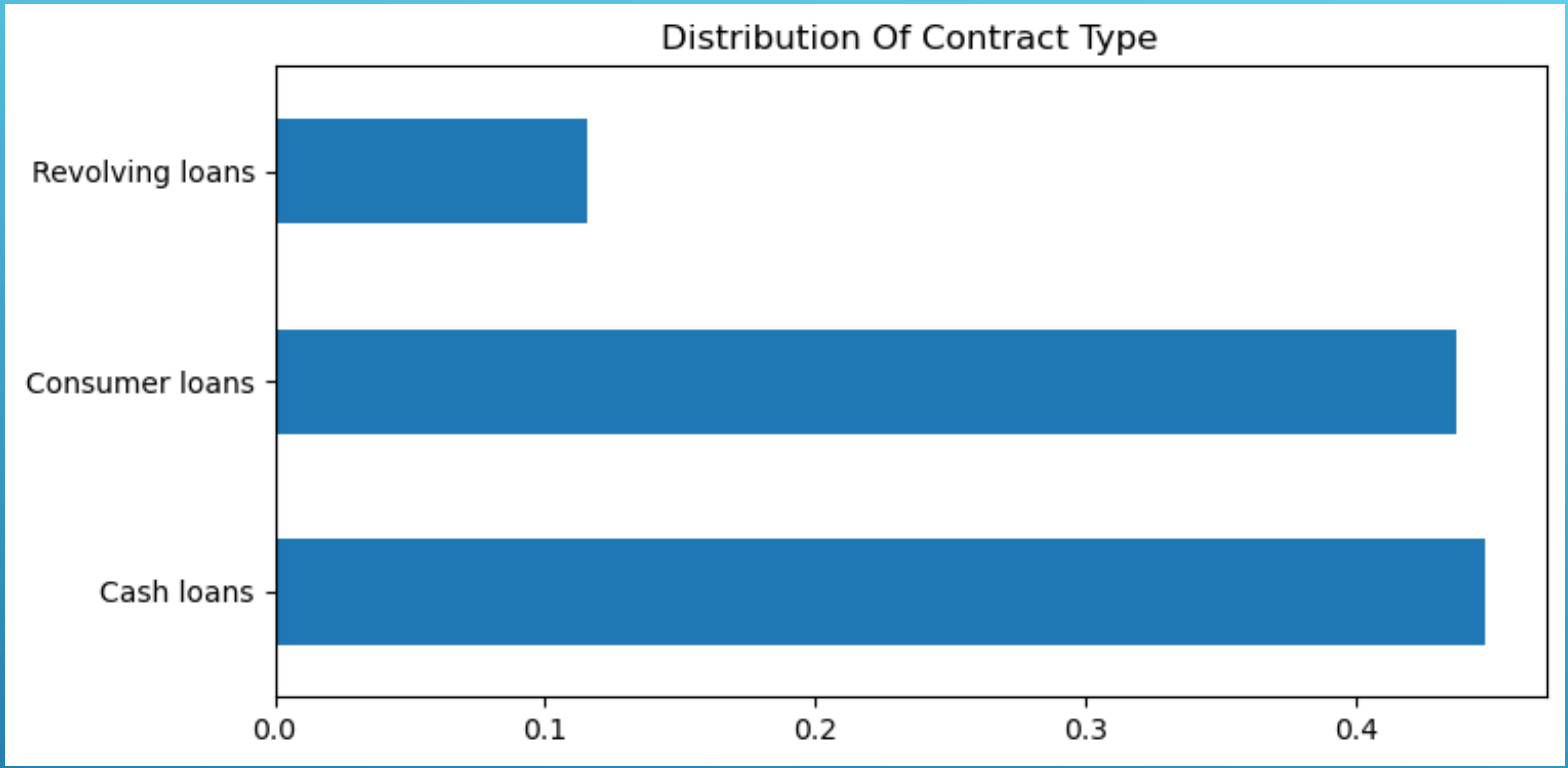
- Categorical Variable



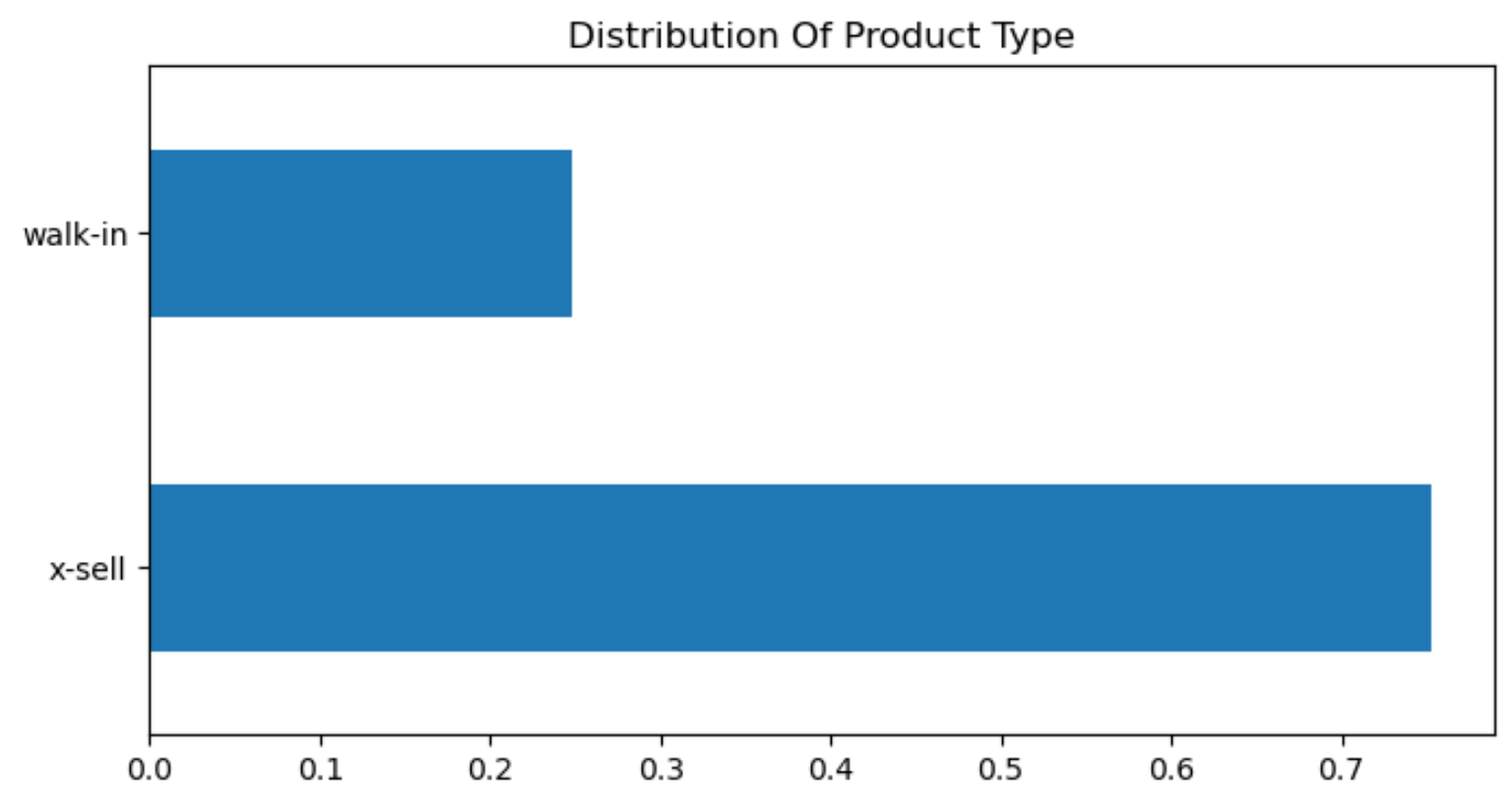
- PAYMENT TYPE –
  By observing above horizontal bar plot, we can say that most of the people chose to pay cash through Bank.



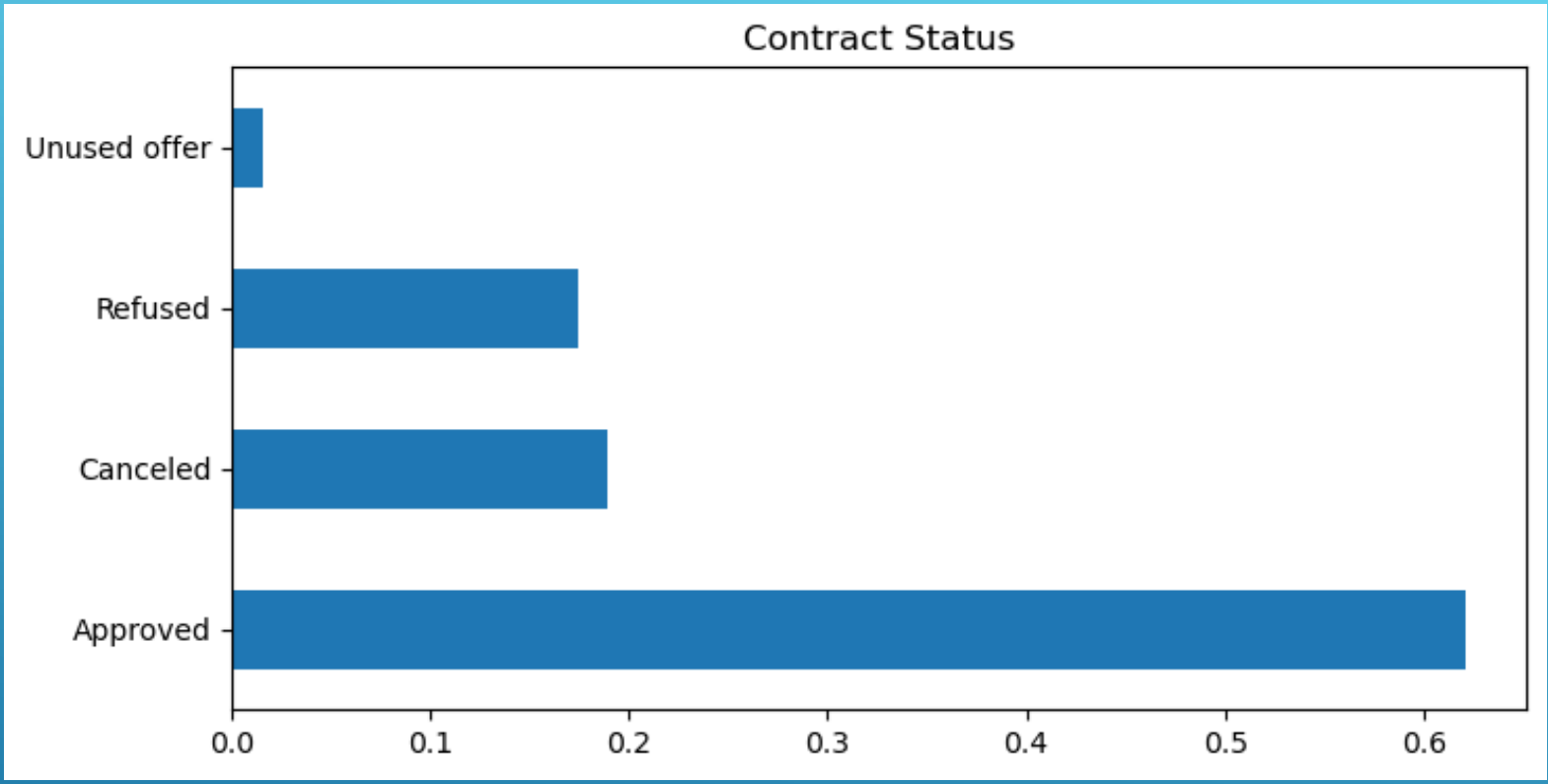
- CLIENT TYPE –
  By observing above horizontal bar plot, we can say that most of the client are Repeater type.



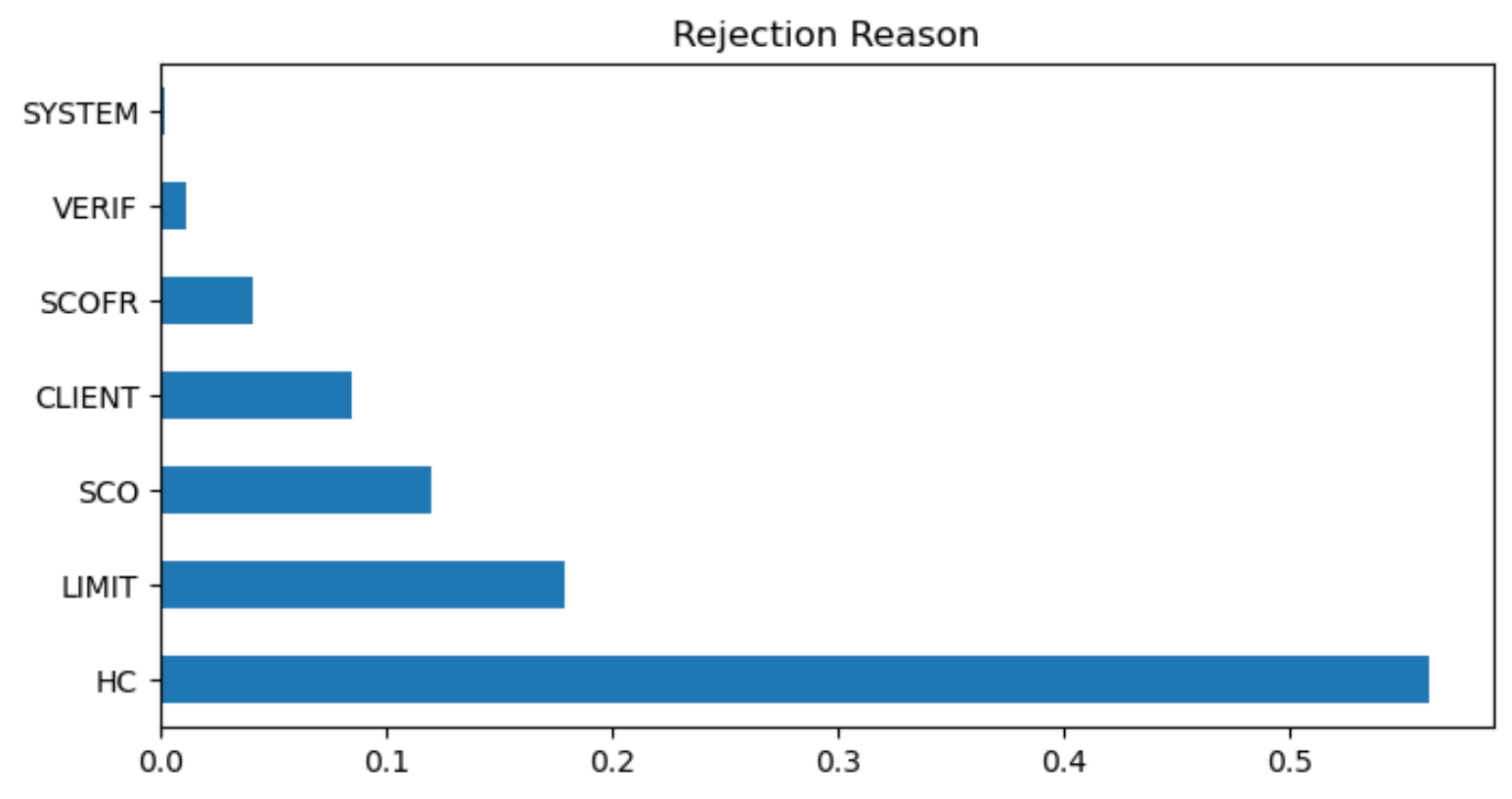
- CONTRACT TYPE –
  By observing above horizontal bar plot, we can say that most of the people prefer Cash loans and Consumer loans.



- PRODUCT TYPE –
  By observing above horizontal bar plot, we can say that most of the people chose x-sell product type.

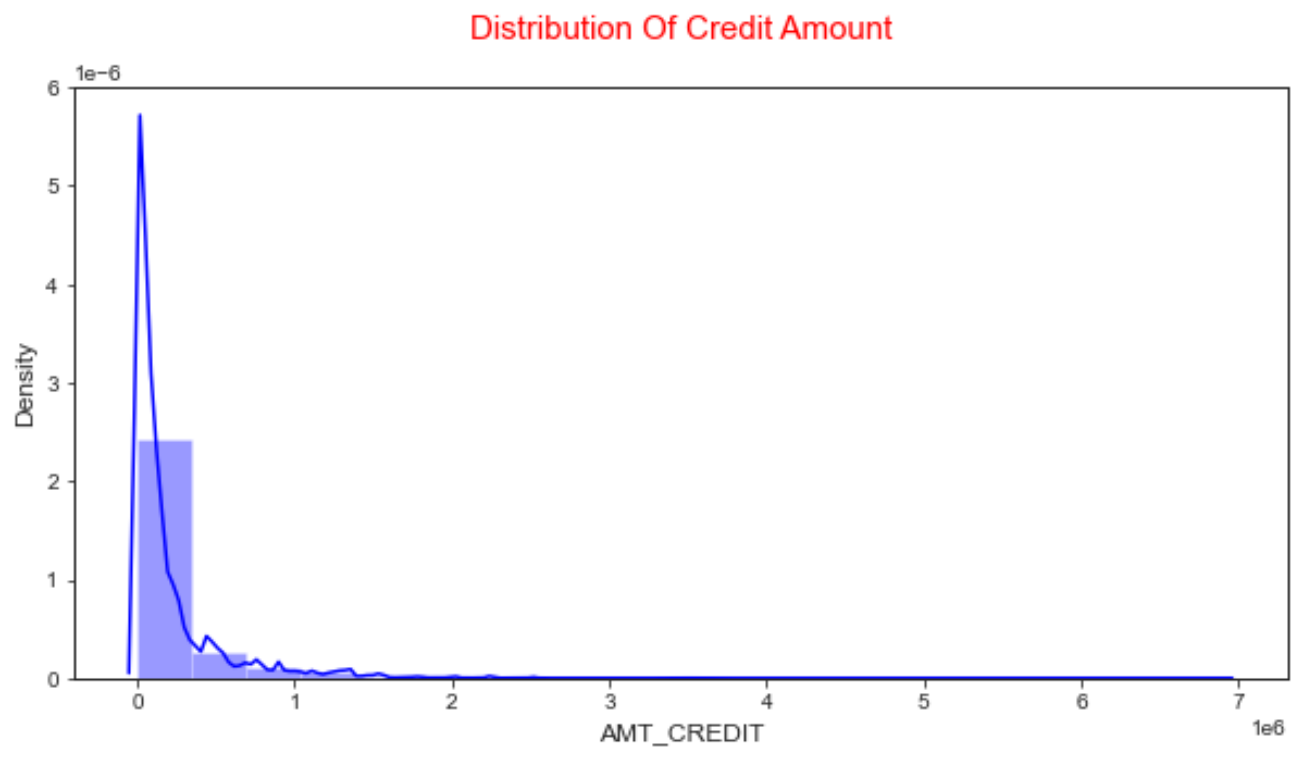


- CONTRACT STATUS –
  By observing above horizontal bar plot, we can say that majority of the loans are approved.



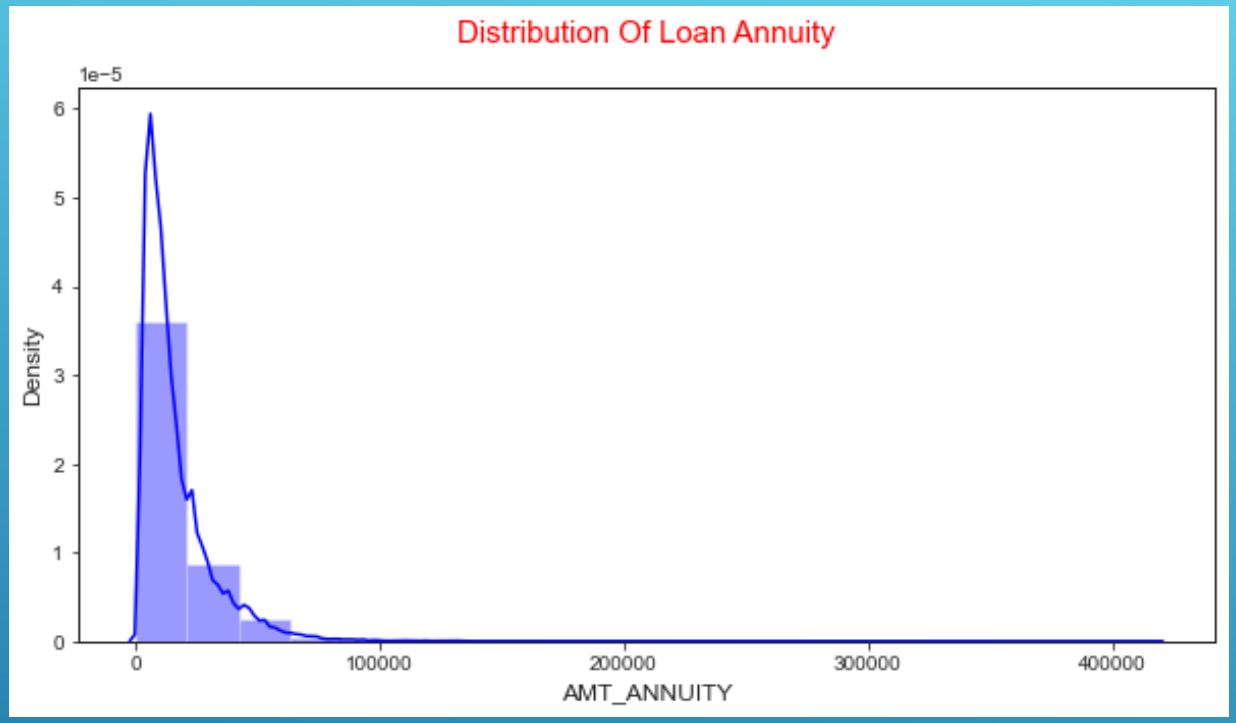
- REJECTION REASON –
  By observing above horizontal bar plot, we can say that major reason for loan rejection is HC.

➢ Numerical Variable

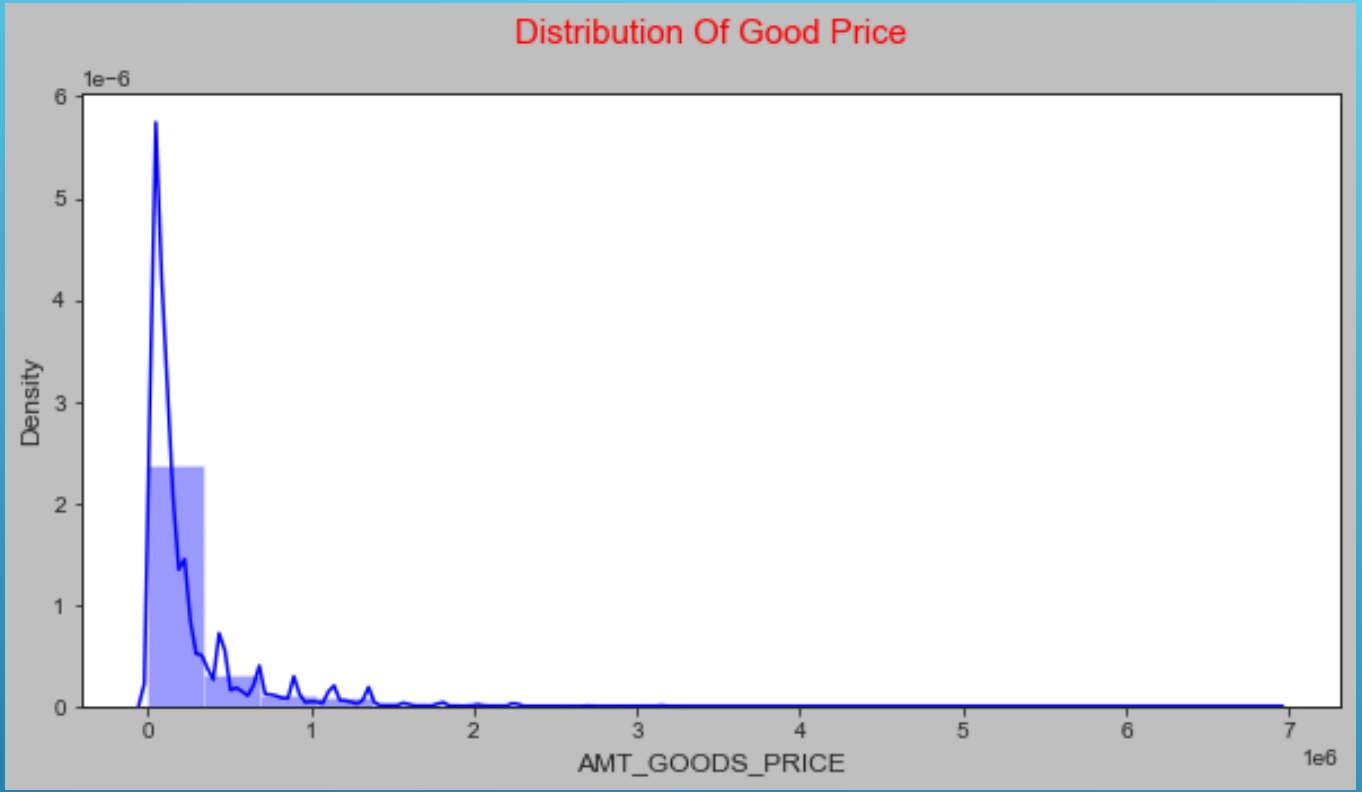


❖ CREDIT AMOUNT –

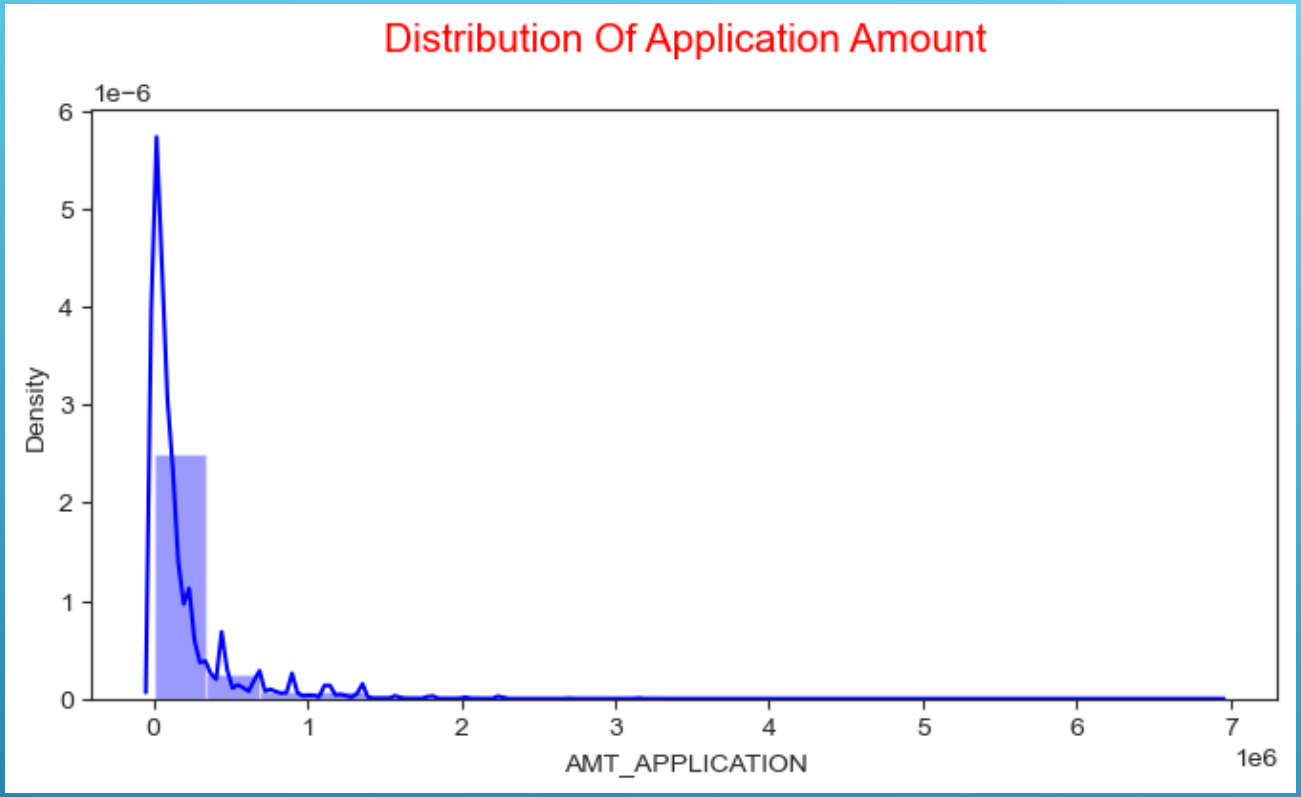
We can observe from the above plot that most of the clients are from first quartile than third quartile. & there are some outliers and the curve is not a bell curve.



- ANNUITY AMOUNT –
  We can observe from the above plot that most of the clients are from first quartile than third quartile. & there are some outliers and the curve is not a bell curve.



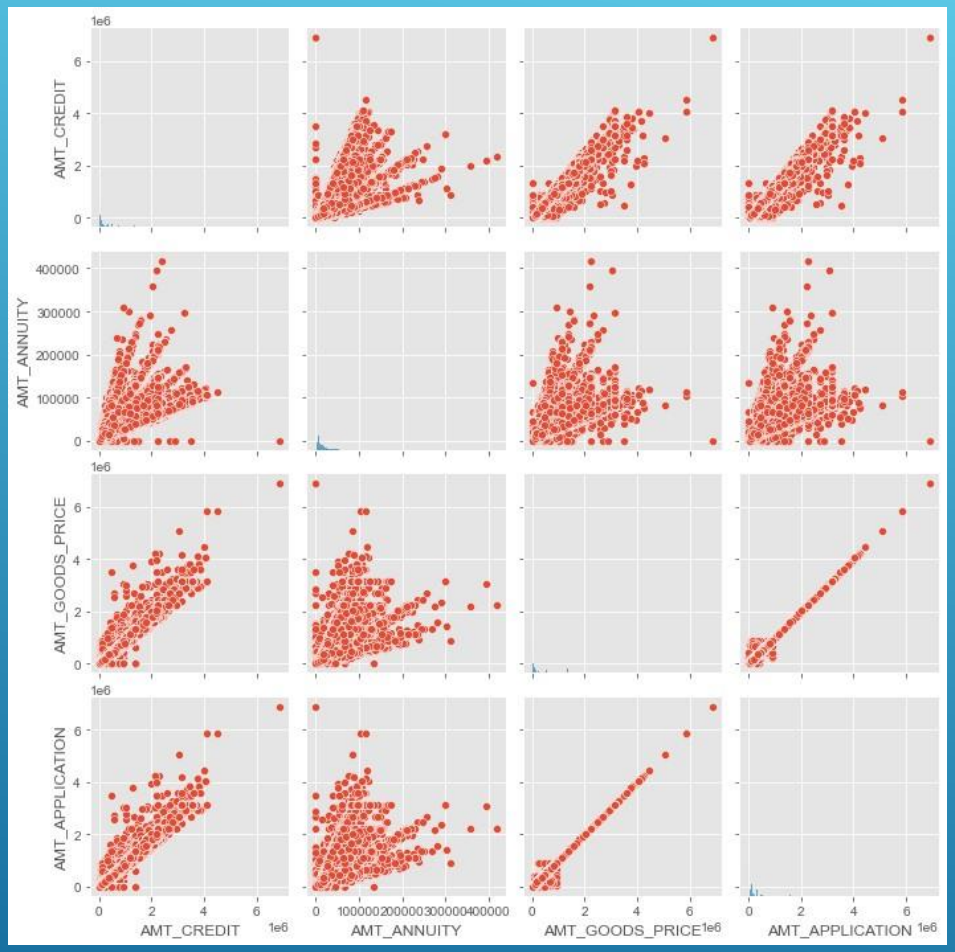
- GOODS PRICE AMOUNT –
  We can observe from the above plot that most of the clients are from first quartile, there are some outliers and the curve is not a bell curve.



- APPLICATION AMOUNT –
  We can observe from the above plot that most of the clients are from first quartile, there are some outliers and the curve is not a normal curve.

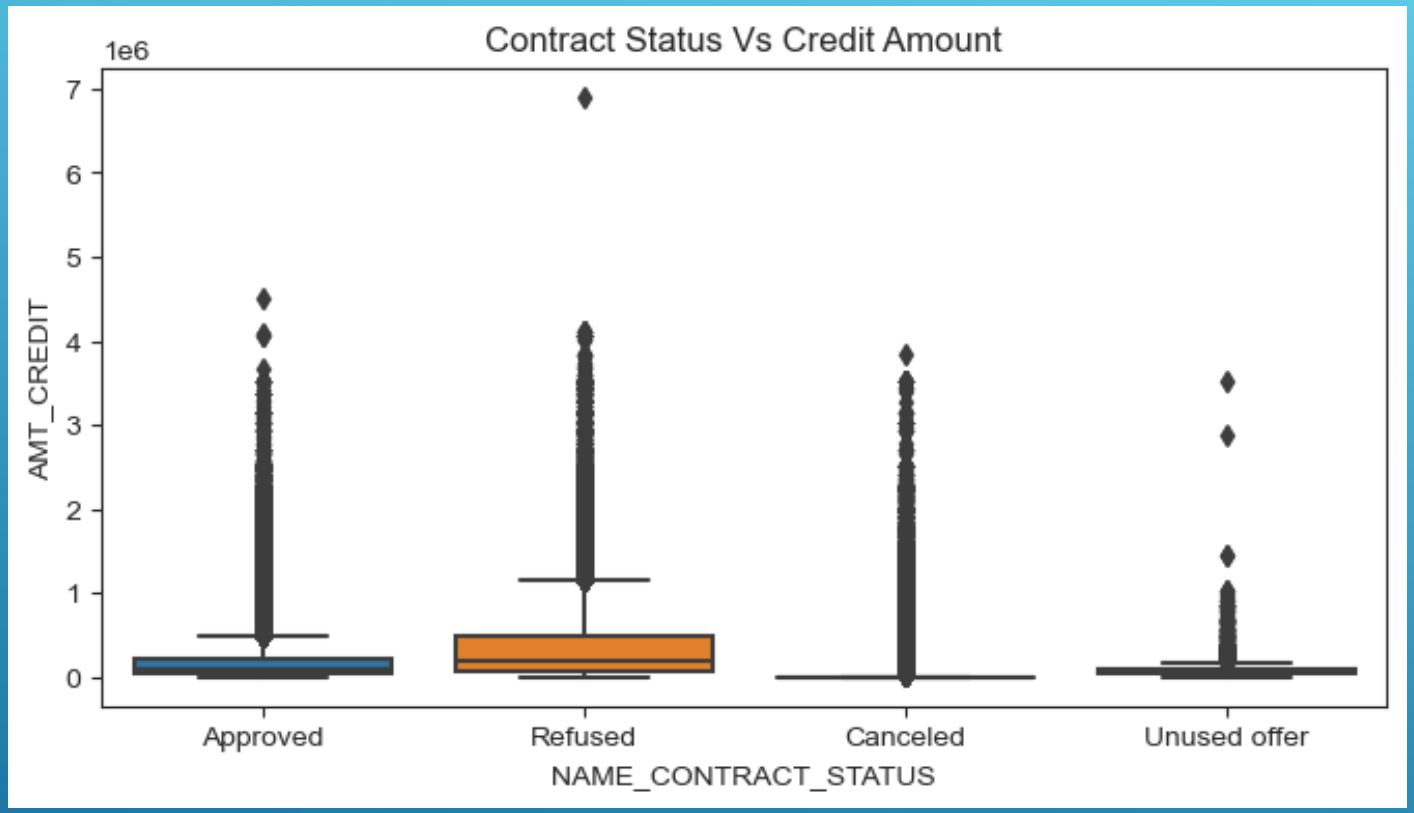
BIVARIATE ANALYSIS

- Numerical Vs Numerical

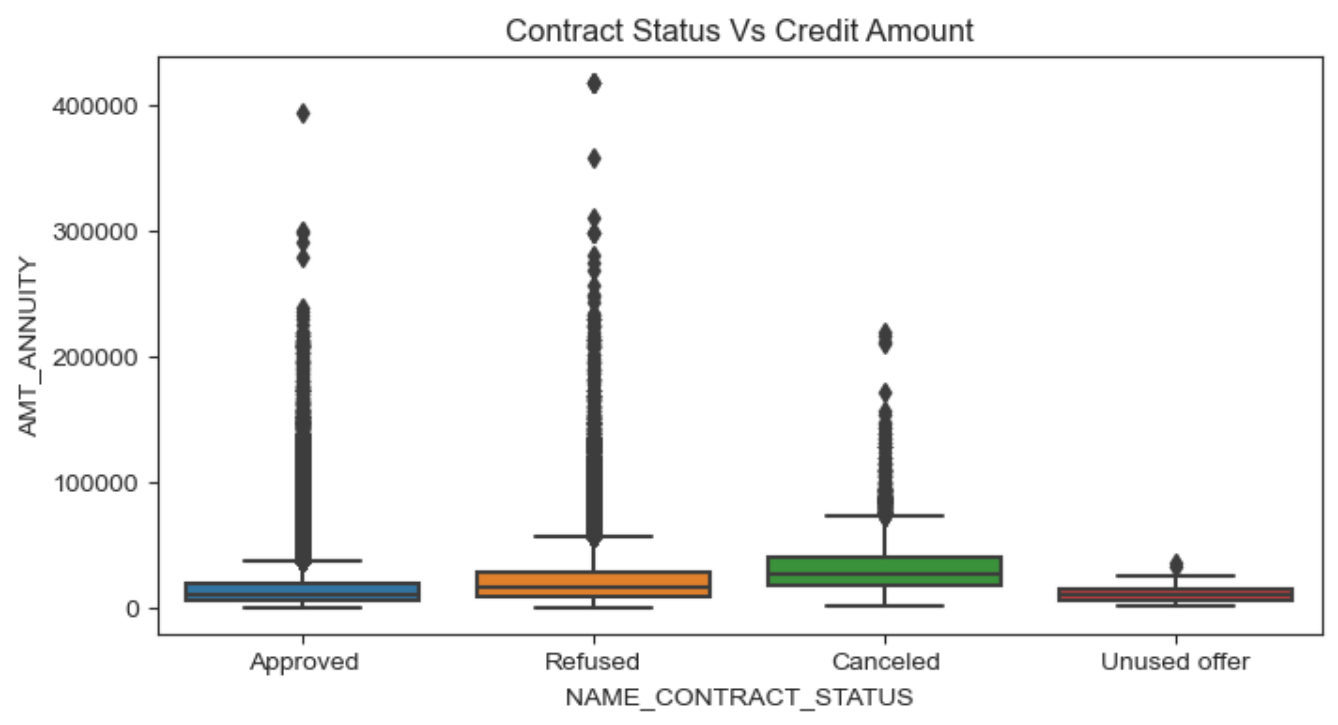


- We can observe from above pair plot that Increase of annuity increases Credit amount and Goods Price for the Previous application Data.

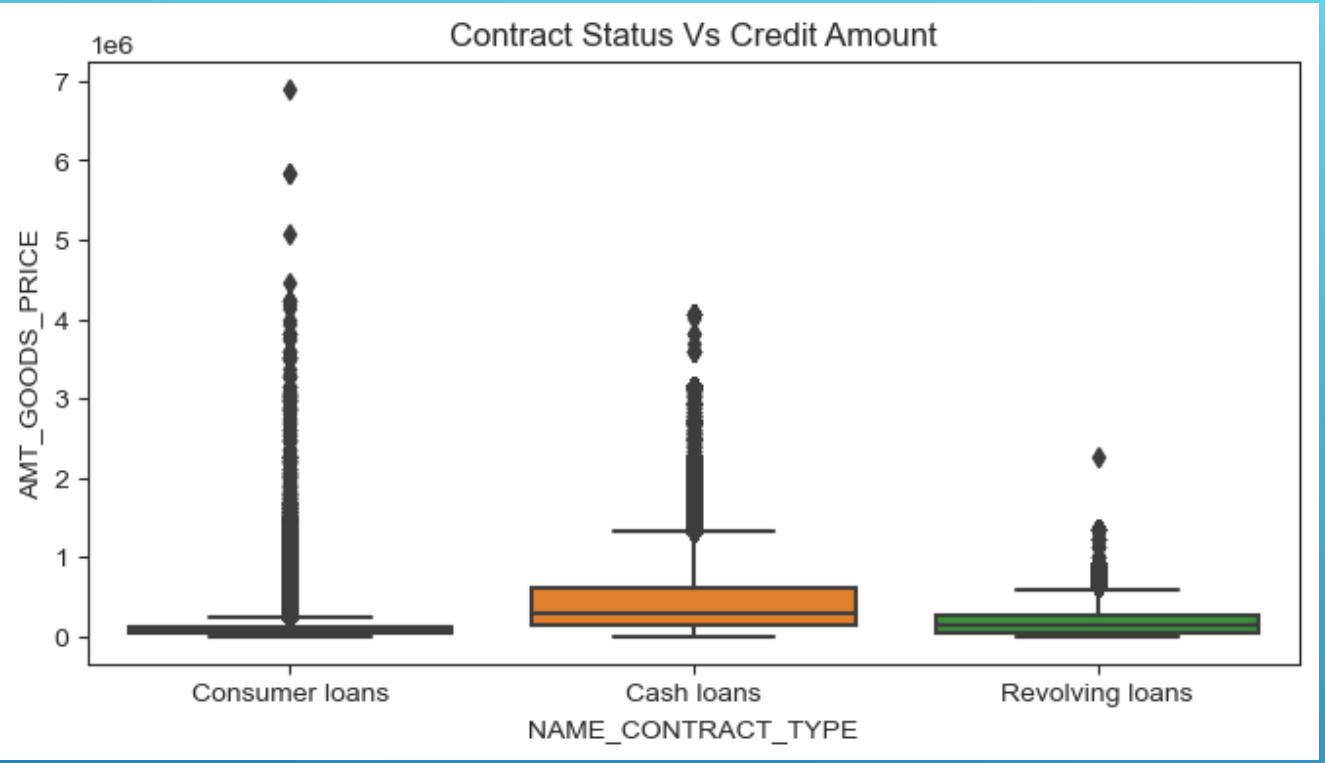
➢ Categorical Vs Numerical



- ❖ CONTRACT STATUS Vs CREDIT AMOUNT –
  We can observe from above boxplot that if Credit amount is low the application may get canceled/Unused.

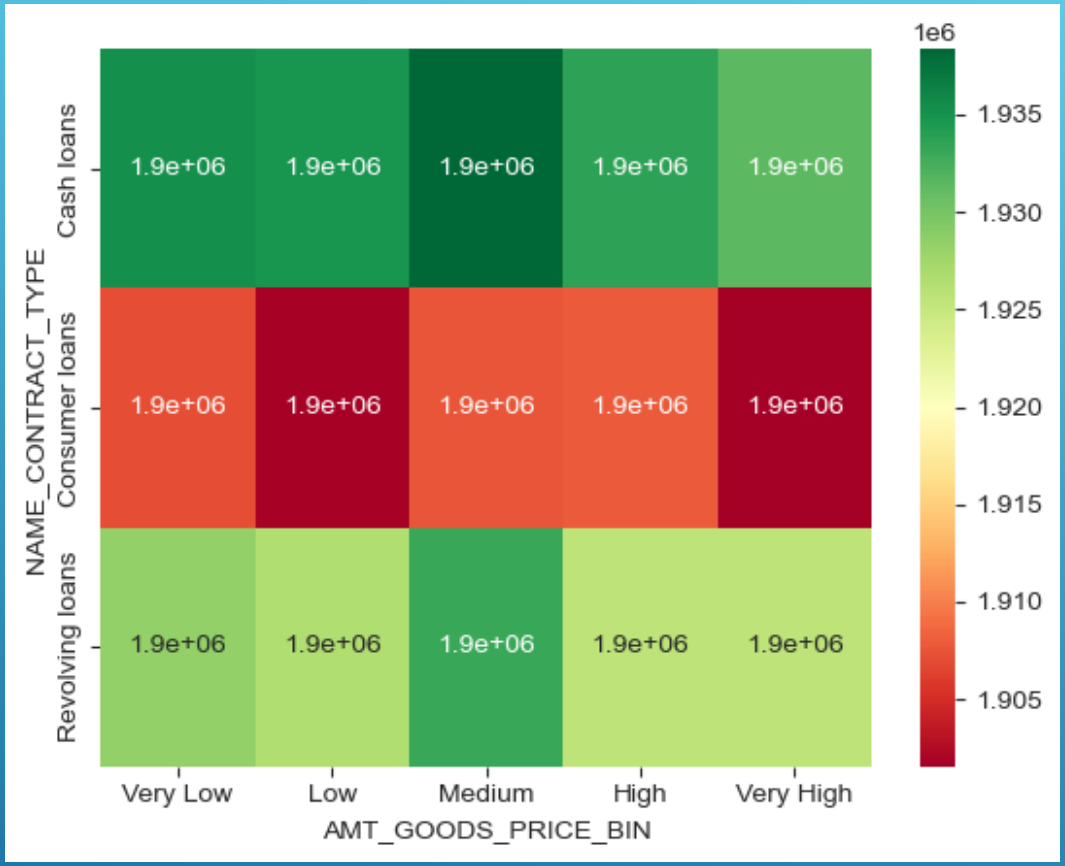


- CONTRACT STATUS Vs ANNUITY AMOUNT –
  We can observe from above boxplot that if Annuity amount is low is low the application may get canceled/Unused and if amount is high application get Approved or sometime Refused.



- CONTRACT TYPE Vs GOODS PRICE –
  We can observe from above boxplot that the loan contact type is Revolving loan if Goods Price is low and if Goods price is high then loan contract type is Consumer loans.

MULTIVARIATE ANALYSIS



- We can observe from above boxplot that the loan contact type is Revolving loan if Goods Price is low and if Goods price is high then loan contract type is Consumer loans.

Conclusions -

EDA on data set revealed that :

- The proportion of Defaulters is 8%.
- Bank majorly provide loan to Females.
- Bank should provide more Revolving Loans.
- Proportions of Defaulters for Working is high.
- Higher educated people default less.
- Proportion of Defaulters for those who lives in Rented apartment is less.
- Married people default more.
- Cash loans have high defaulters, Bank should provide more Revolving Loans.
- Old-age people are safer as they are less default less, middle age people default more.

THANK YOU…