# CREDIT EDA CASE STUDY PRESENTATION

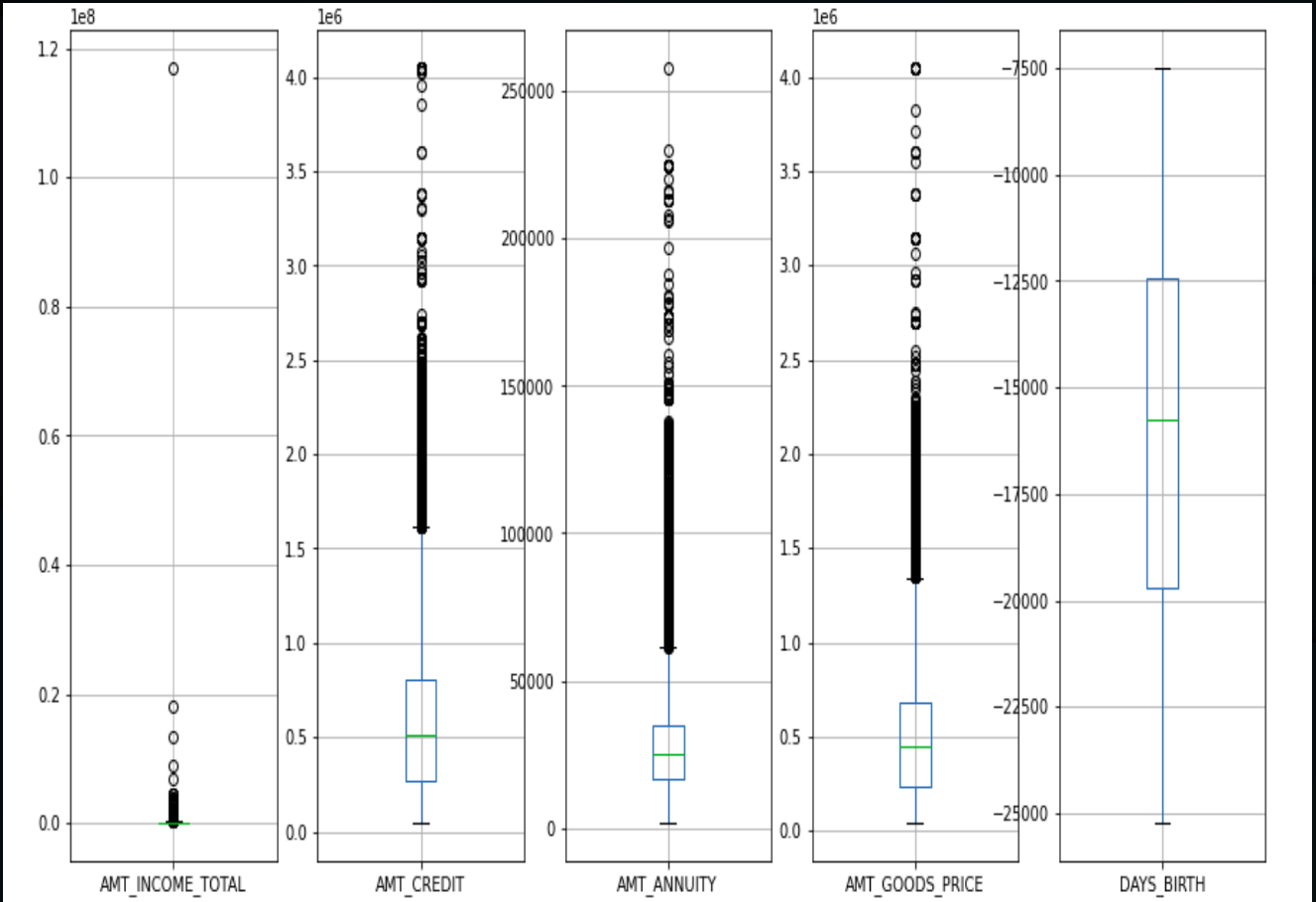
# BUSINESS OBJECTIVE

This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

# STEPS TO FOLLOW

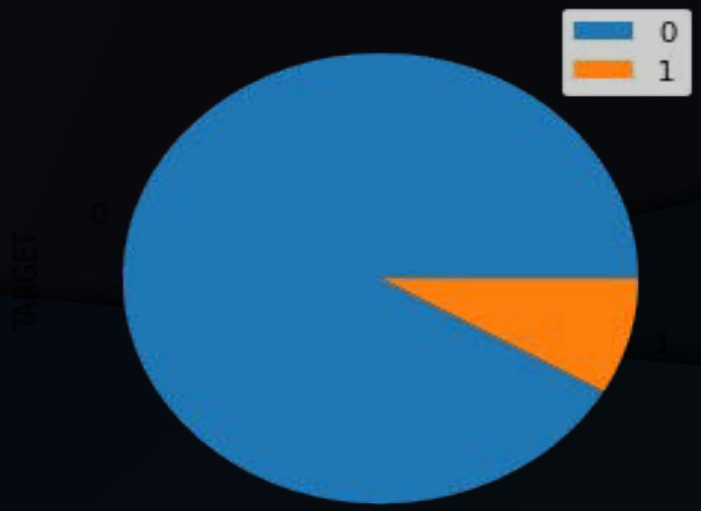
- Data Understanding
- Present the overall approach of the analysis in a presentation. Mention the problem statement and the analysis approach briefly.
- Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value)
- Identify if there are outliers in the dataset. Also, mention why do you think it is an outlier.
- Explain the results of univariate, segmented univariate, bivariate analysis, etc. in business terms.
- dentify if there is data imbalance in the data. Find the ratio of data imbalance.
- Find the top 10 correlation for the Client with payment difficulties and all other cases
- Include visualisations and summarise the most important results in the presentation.

# OUTLIERS



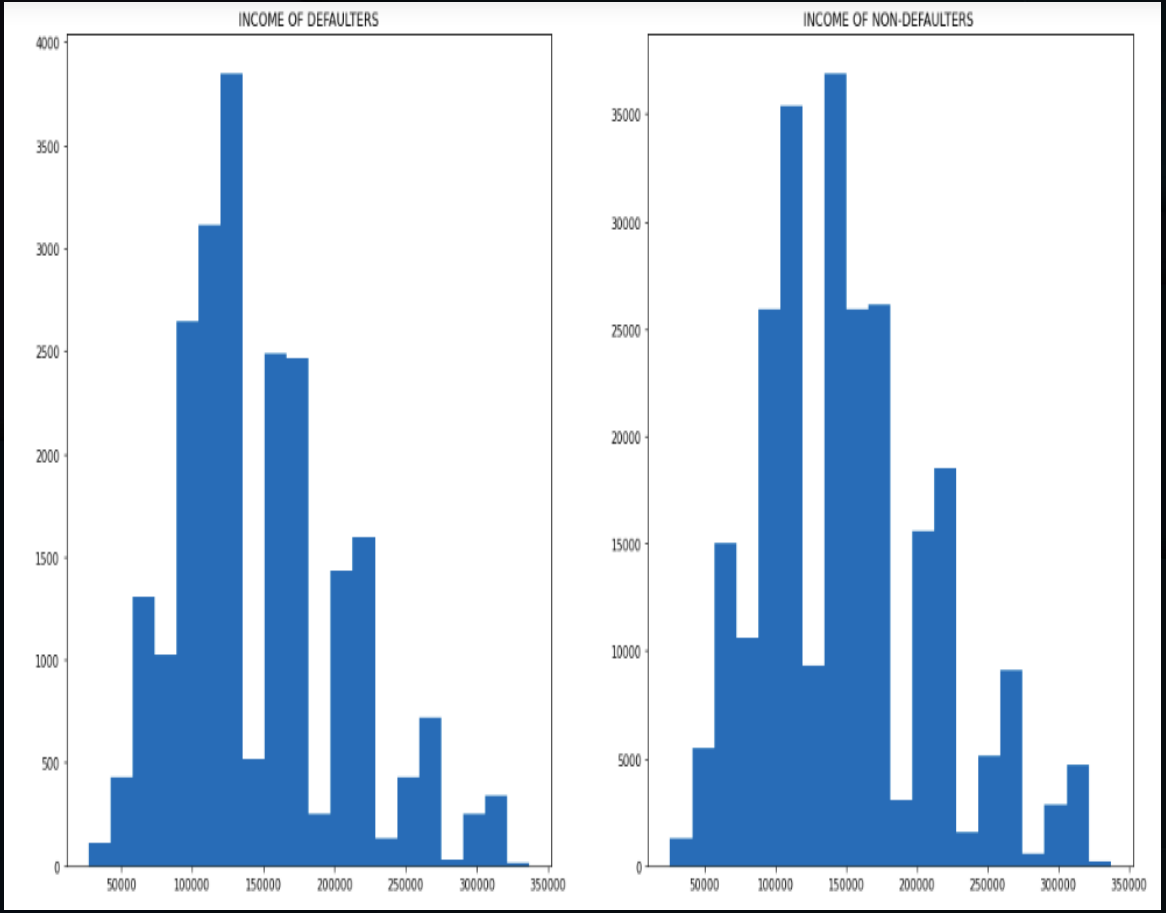

We can see outliers in all the column except Days_Birth

# IMBALANCE BETWEEN TARGET VALUES



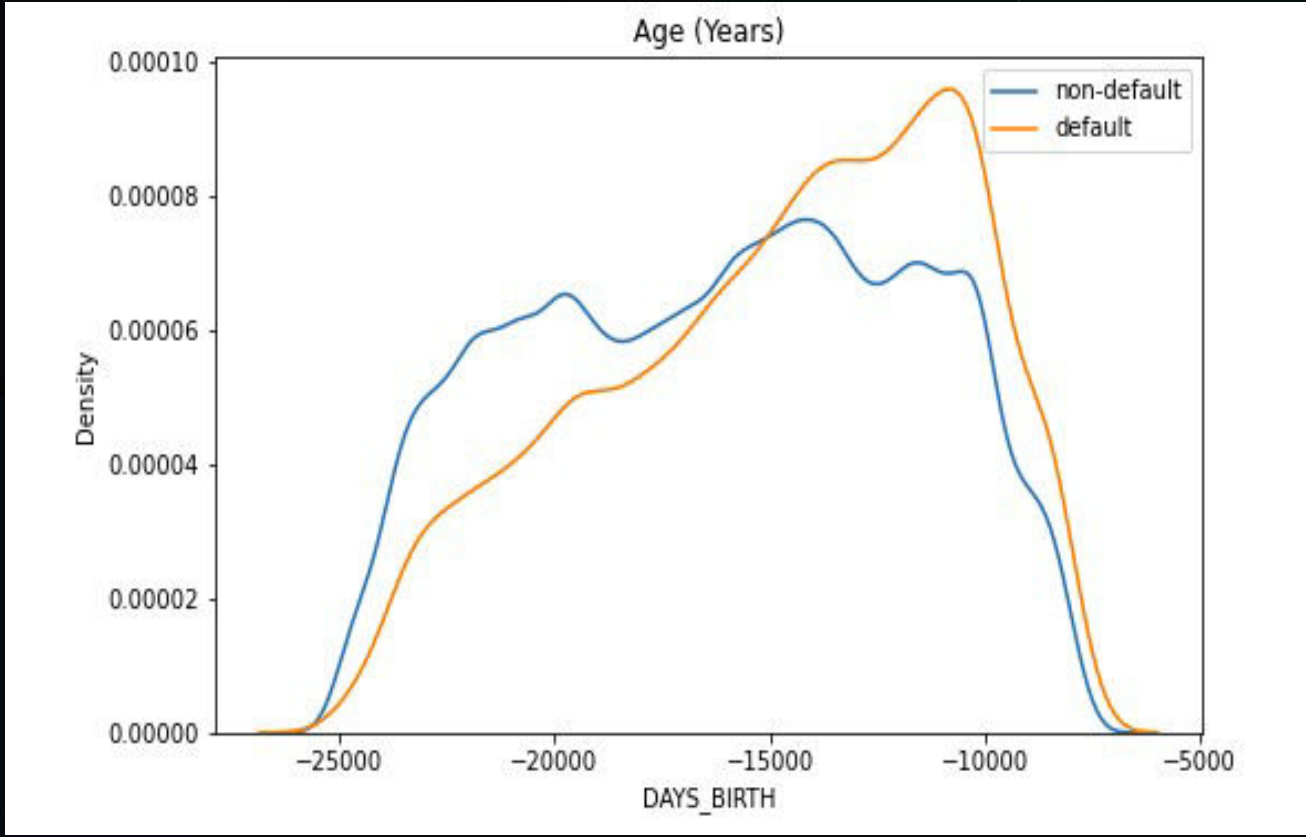

**There is a big imbalance between Target values which can differentiate between default (Target = 1) and non-default (Target = 0). Ratio of target values [non-default/default] is 10.9:1 which is less but it does not mean that we just ignore them there is probability of those defaulters who has taken big loans. So we need to carry on the Analysis.**

# UNIVARIATE ANALYSIS OF CONTINUOUS DATA



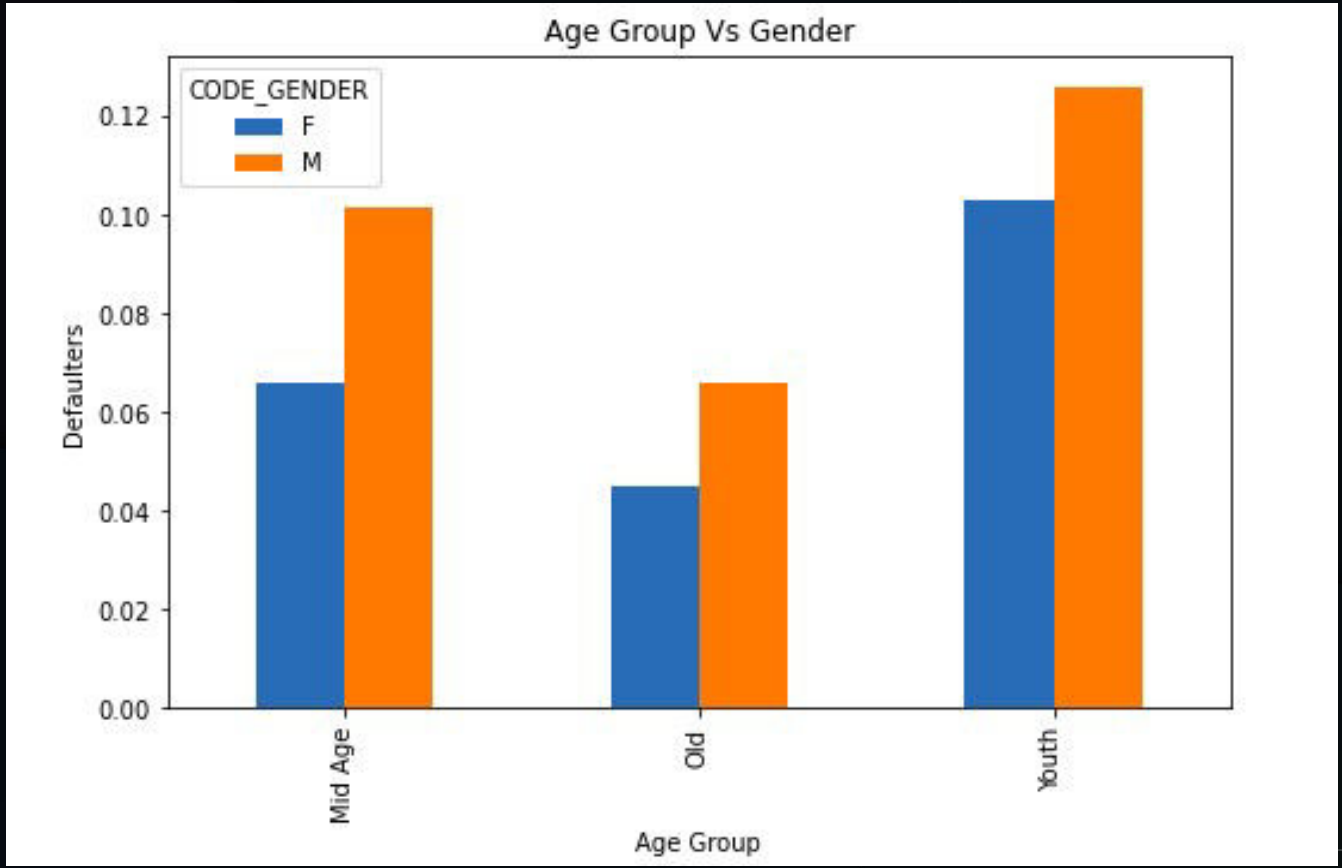

In both cases, maximum people are from income group 1-1.75 lakhs.

# Density plot using Seaborn for DAYS_BIRTH



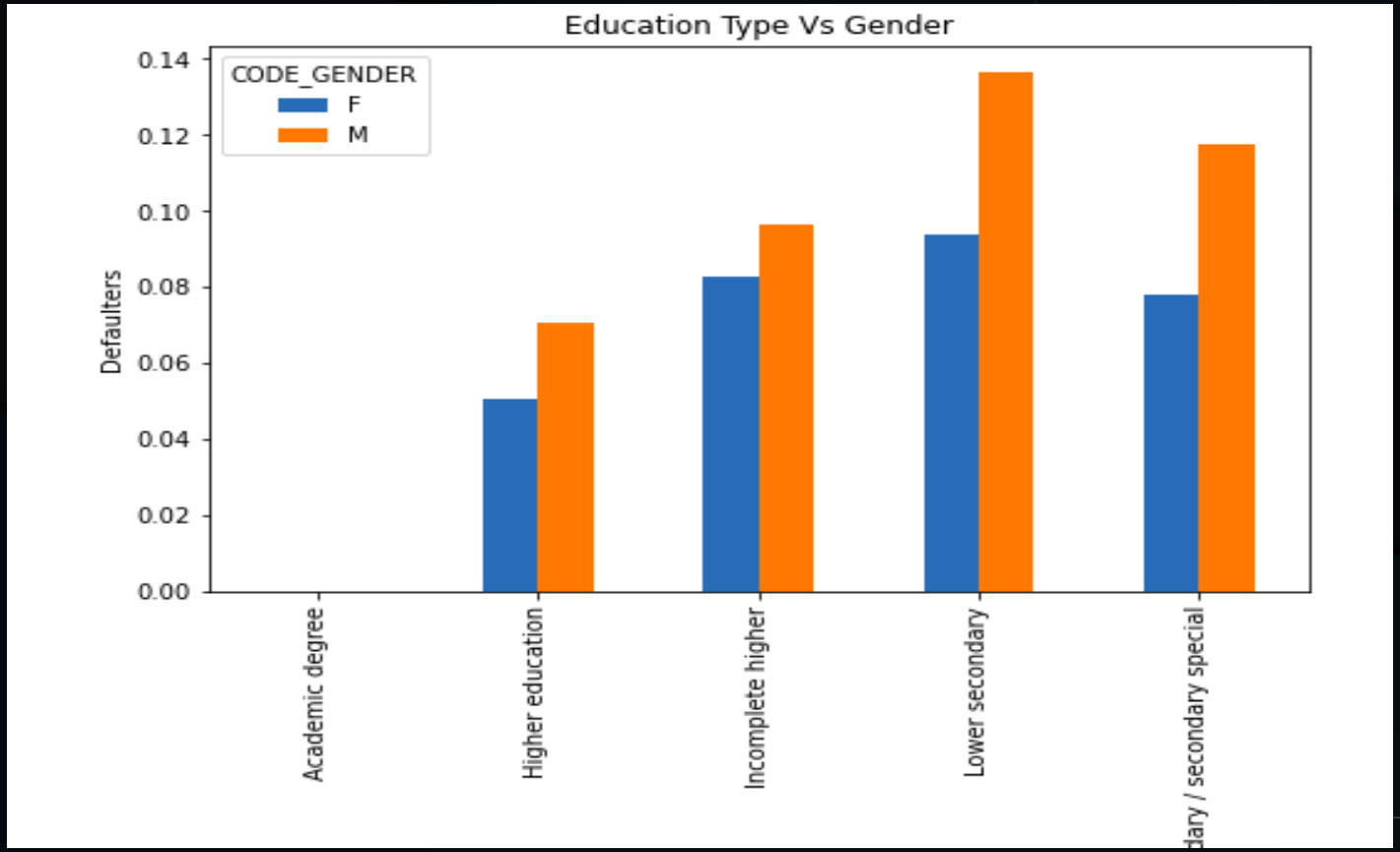

According to Graph, age group of (20-30) have high density of defaulters. So we should take priority to take actions on defaulters of age group(20-30).

# SEGMENTED ANALYSIS



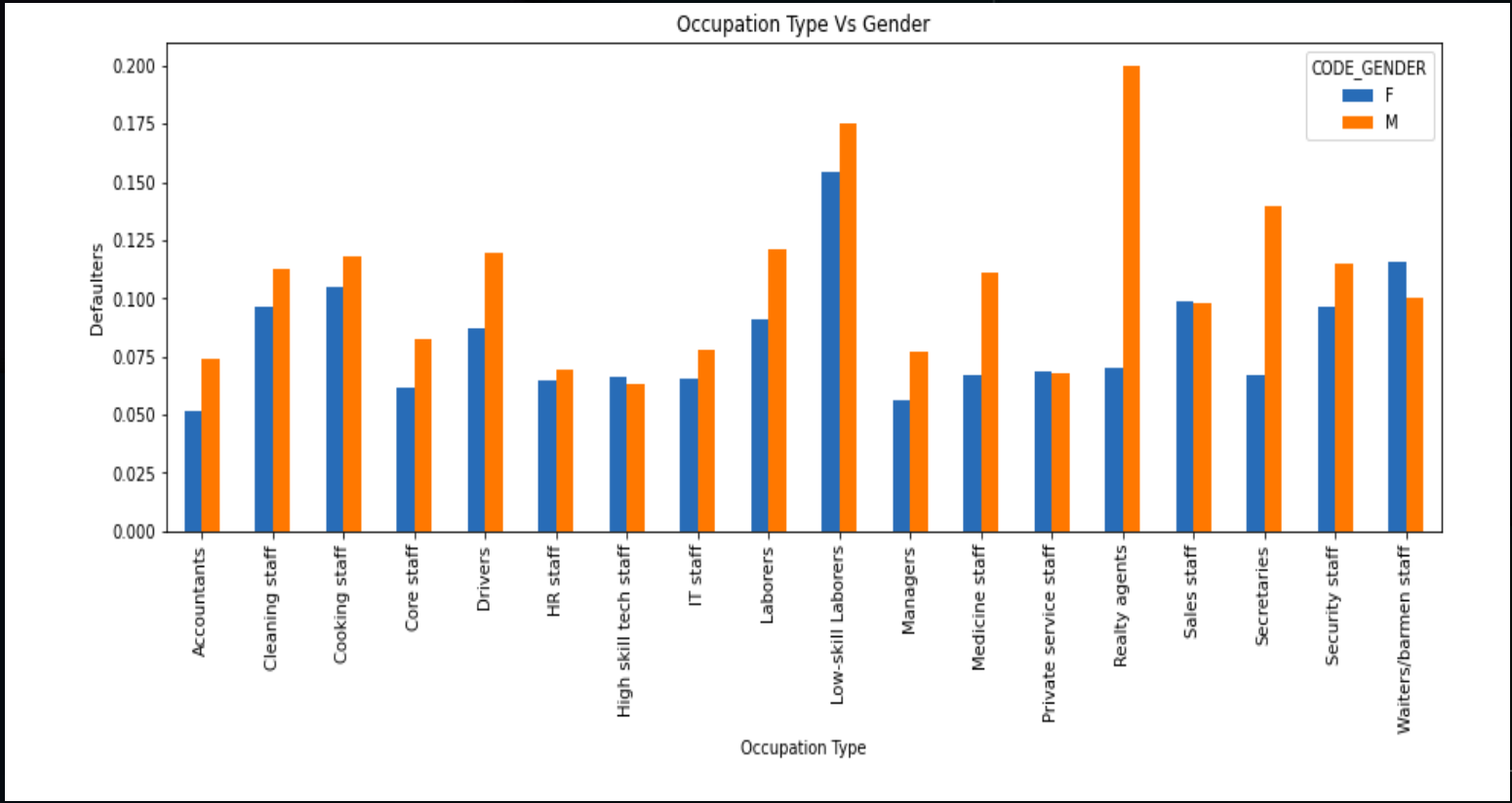

From we able to know In every Age group, Males defaulters are more In number than females. The highest number of defaulters are seen amongst the youth .

# SEGMENTED EDUCATION AND GENDER



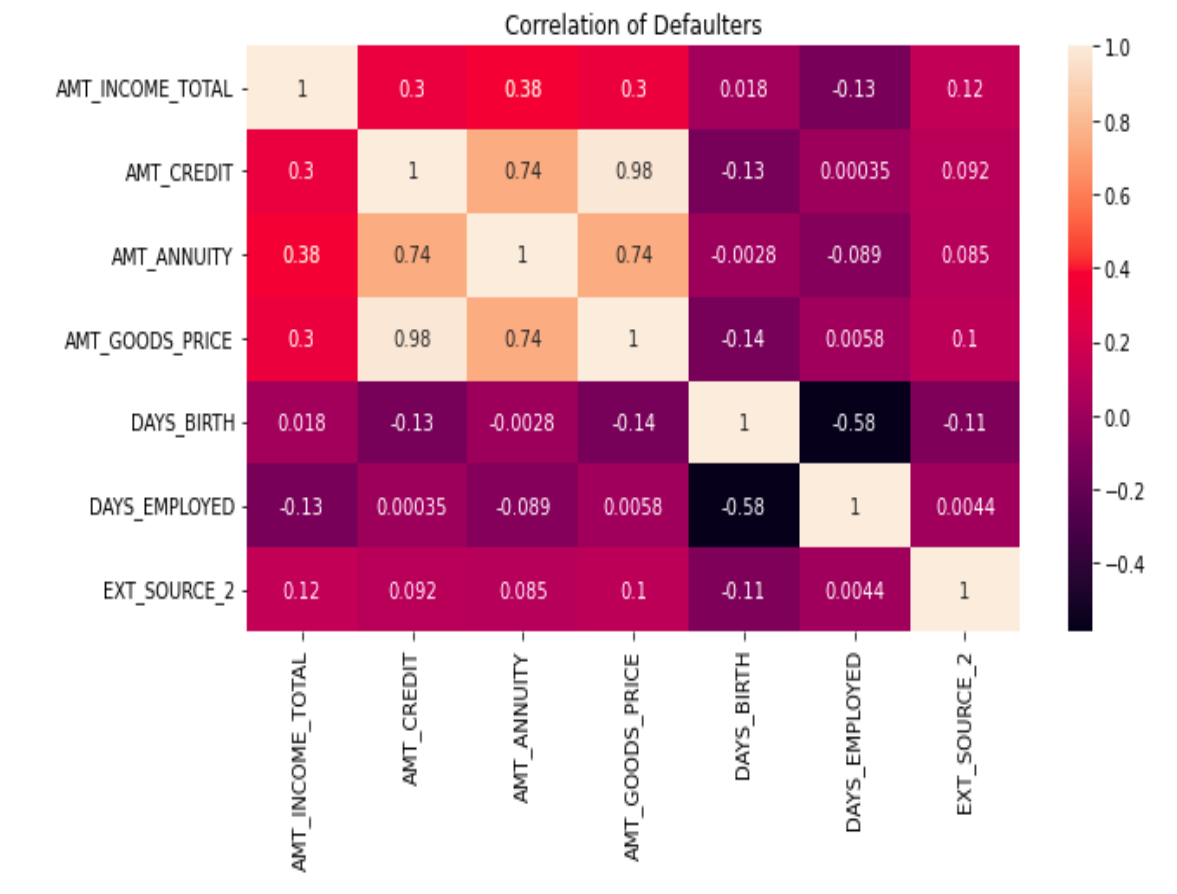

From the grabh we able to know Males defaulters are higher In number As compared to female defaulters.Also the defaluters are highest In Lower Secondary group.

# SEGMENTED OCCUPATION TYPE vs GENDER



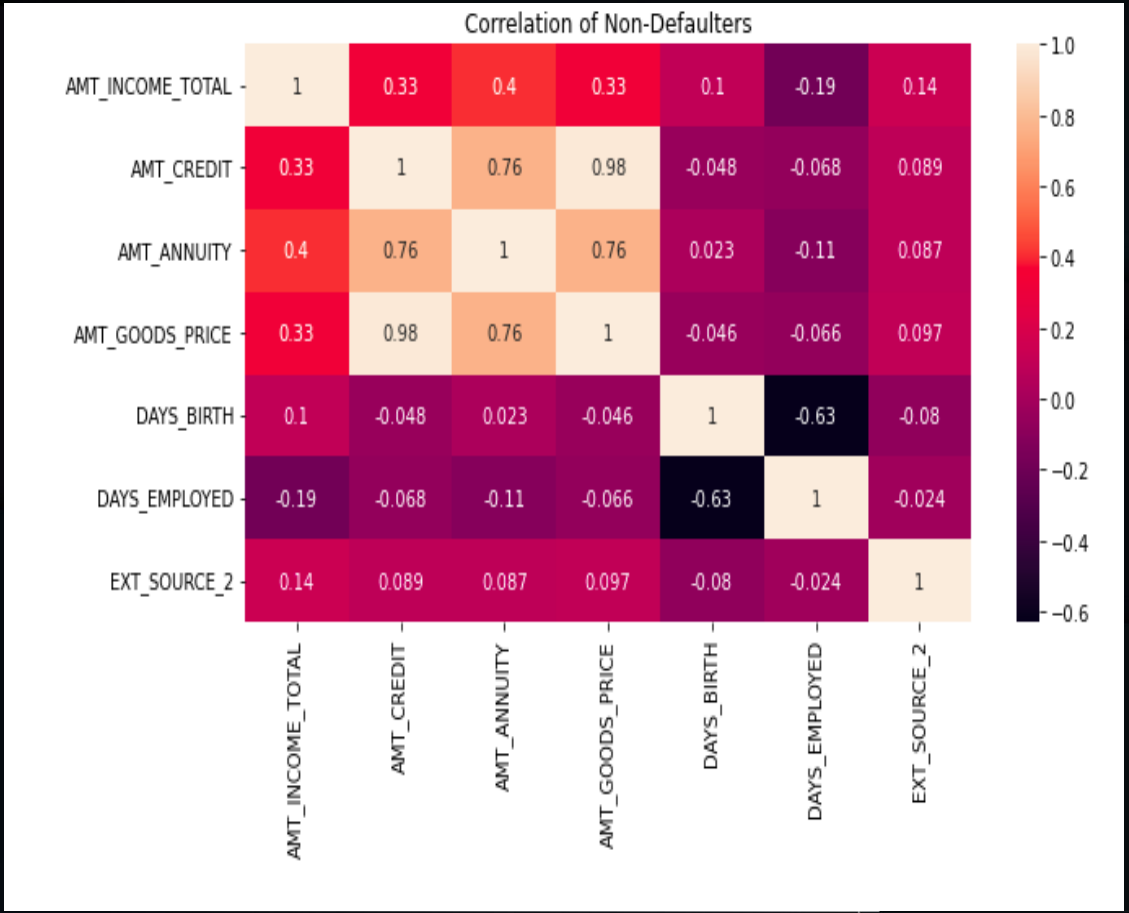

After looking into the graph Realty Agent Male defaulters are highest In the category.

# The Co-relation of continuous variables of Defaulter and Non-Defaulter dataframe will concluded through this analysis



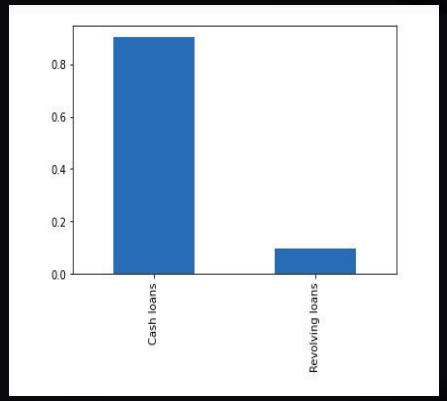

From above plot we can see that there Is a high correlation value between GOODS_PRICE and AMT_CREDIT, AMT_ANNUTY and AMT_AMT_CREDIT .

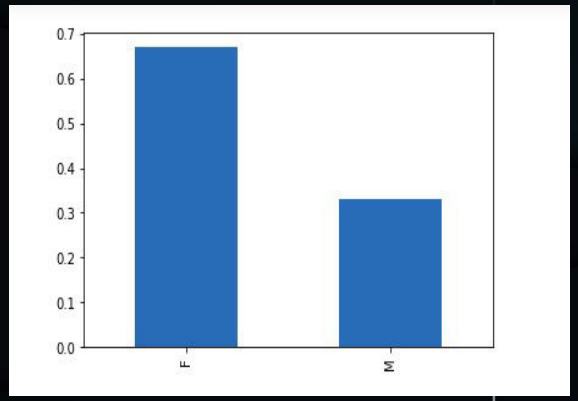
The correlation values of DAYS_EMPLOYED is very low with variables like AMT_INCOME_TOTAL,AMT_CREDIT,AMT_ANNUITY,DAYS_BIRTH () they have negtaive values.

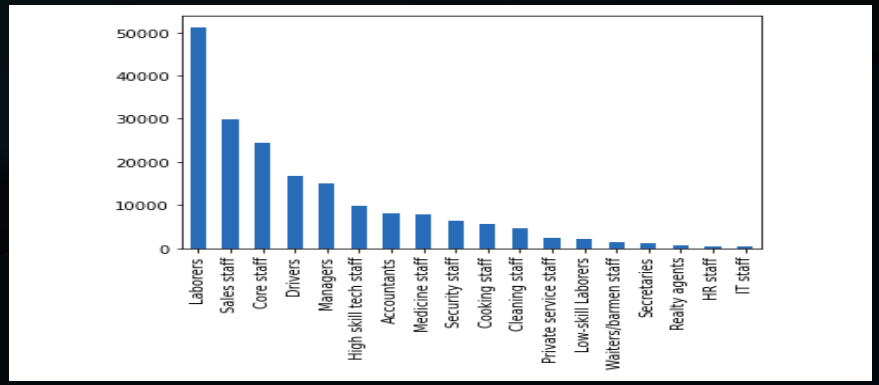


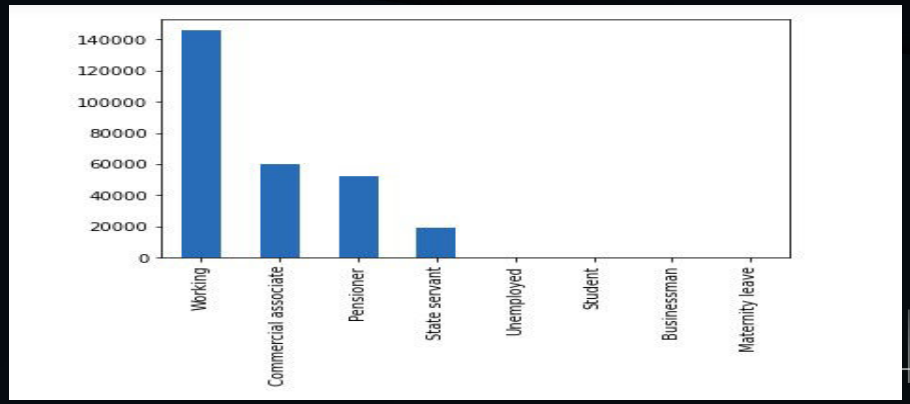

We conclude that there is a high co-relation only between AMT_GOODS_PRICE and AMT_CREDIT. While there is moderate co-relation between DAYS_BIRTH and DAYS_EMPLOYED,AMT_ GOODS_PRCE and AMT_ANNUITY alos AMT_CREDIT and AMT_ANNUITY.

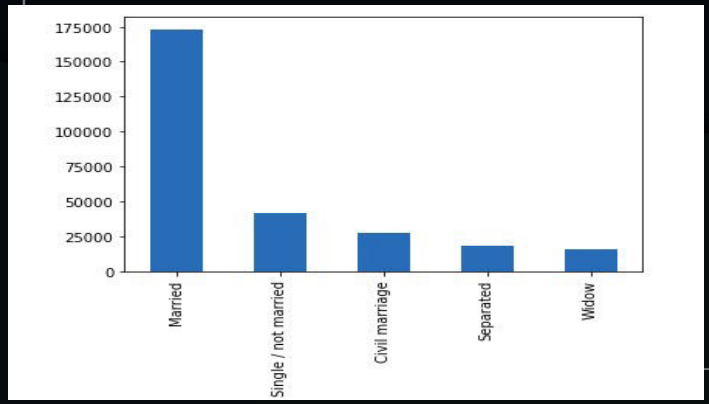
# UNIVARIATE ANALYSIS ON CATEGORICAL DATA



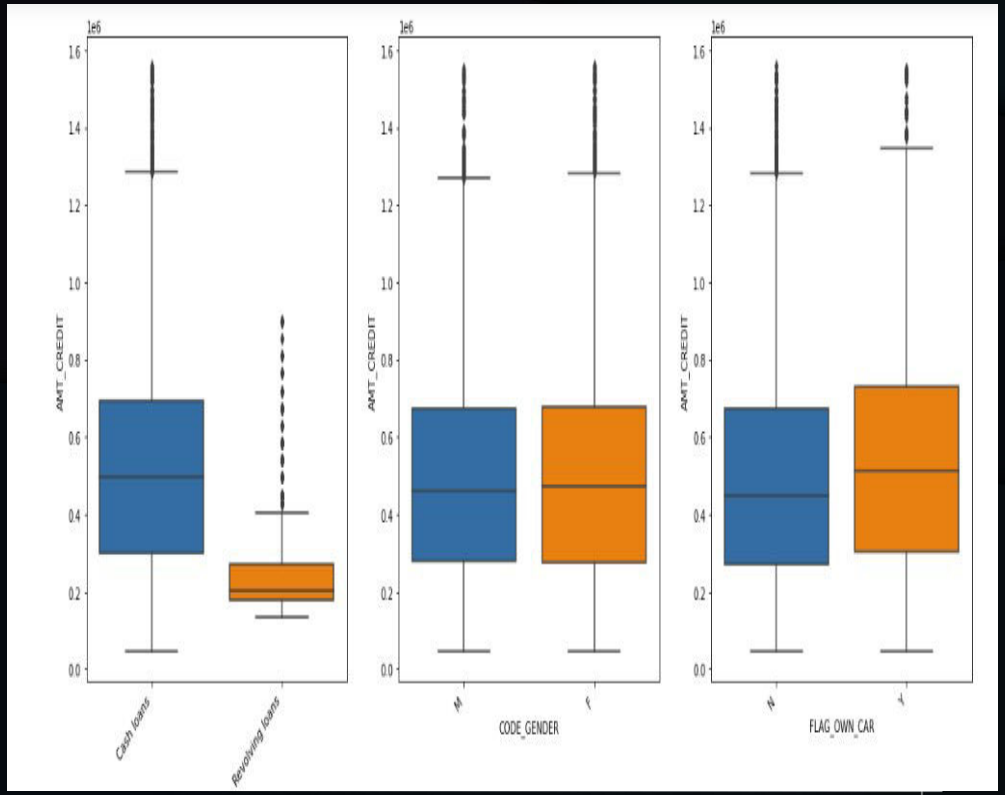



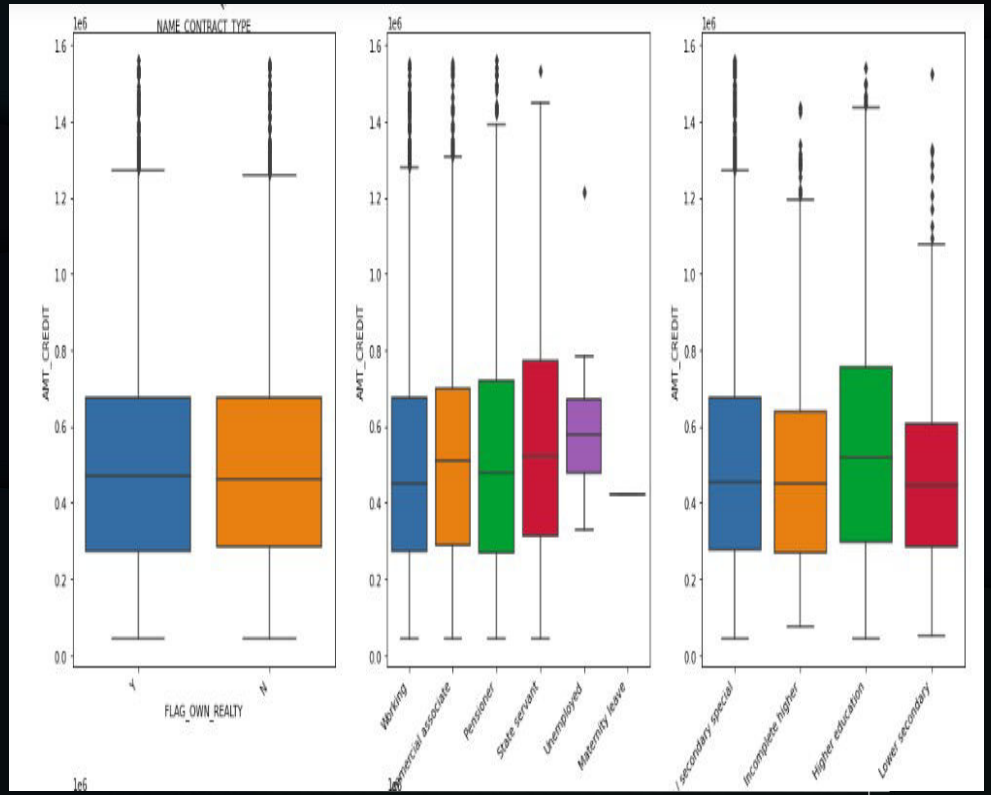



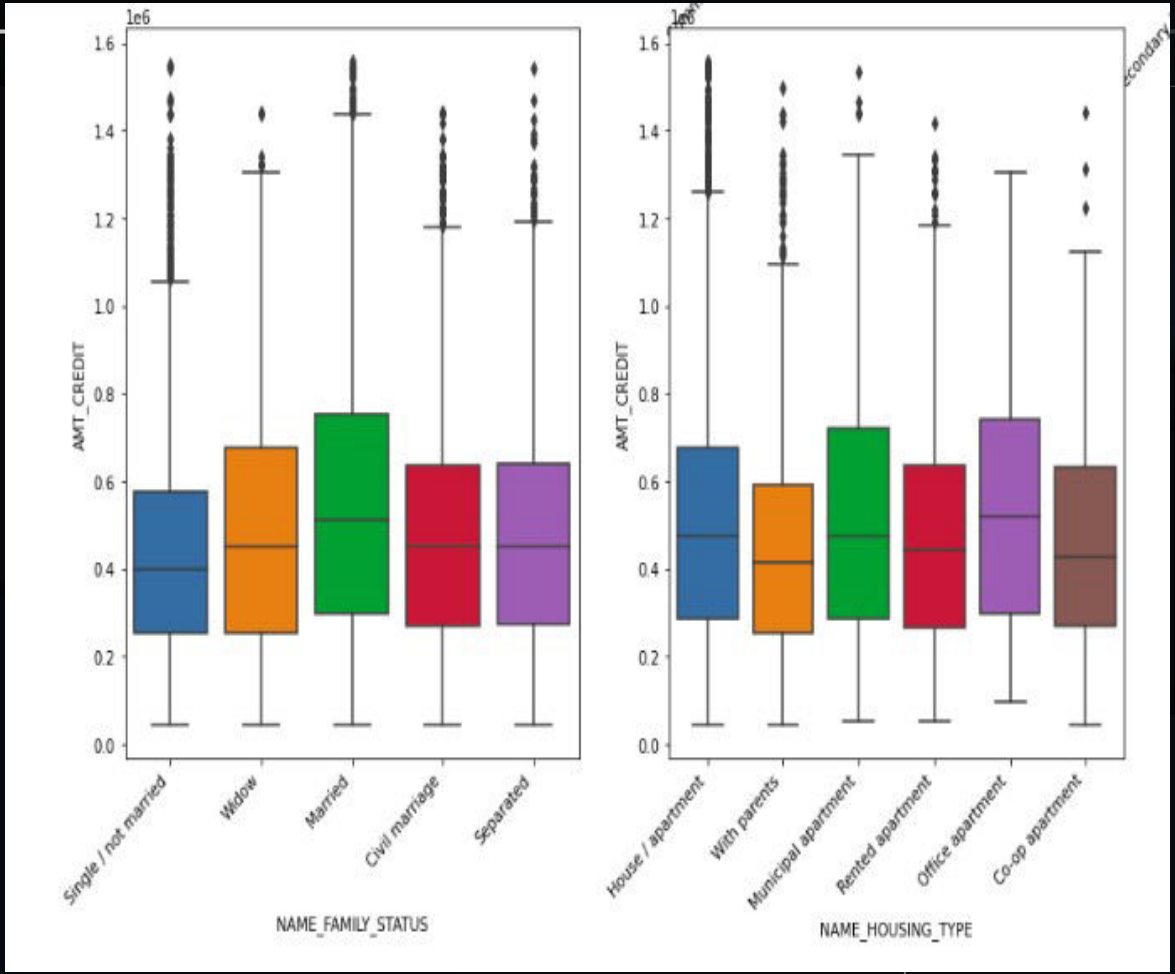



# BIVARIATE ANALYSIS ON CATEGORICAL COLUMN

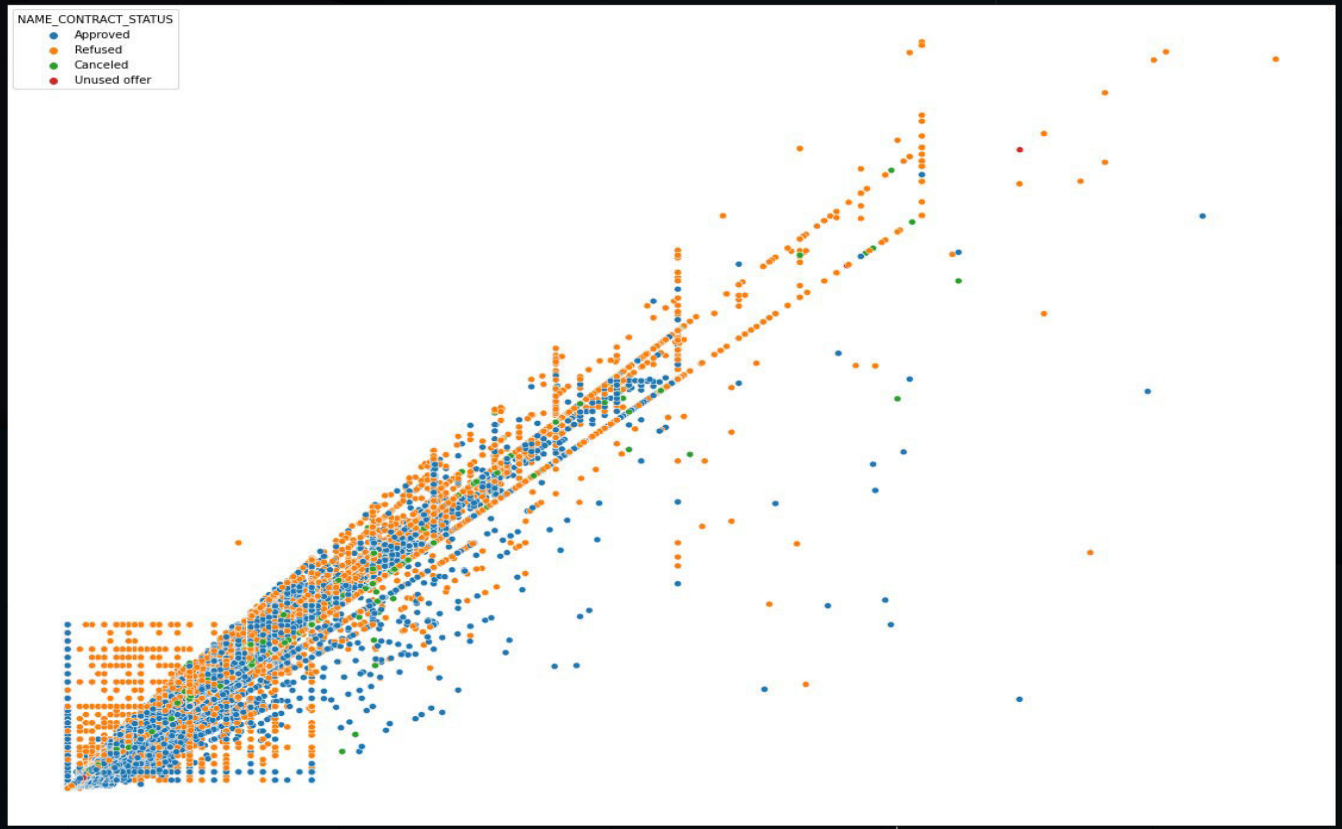




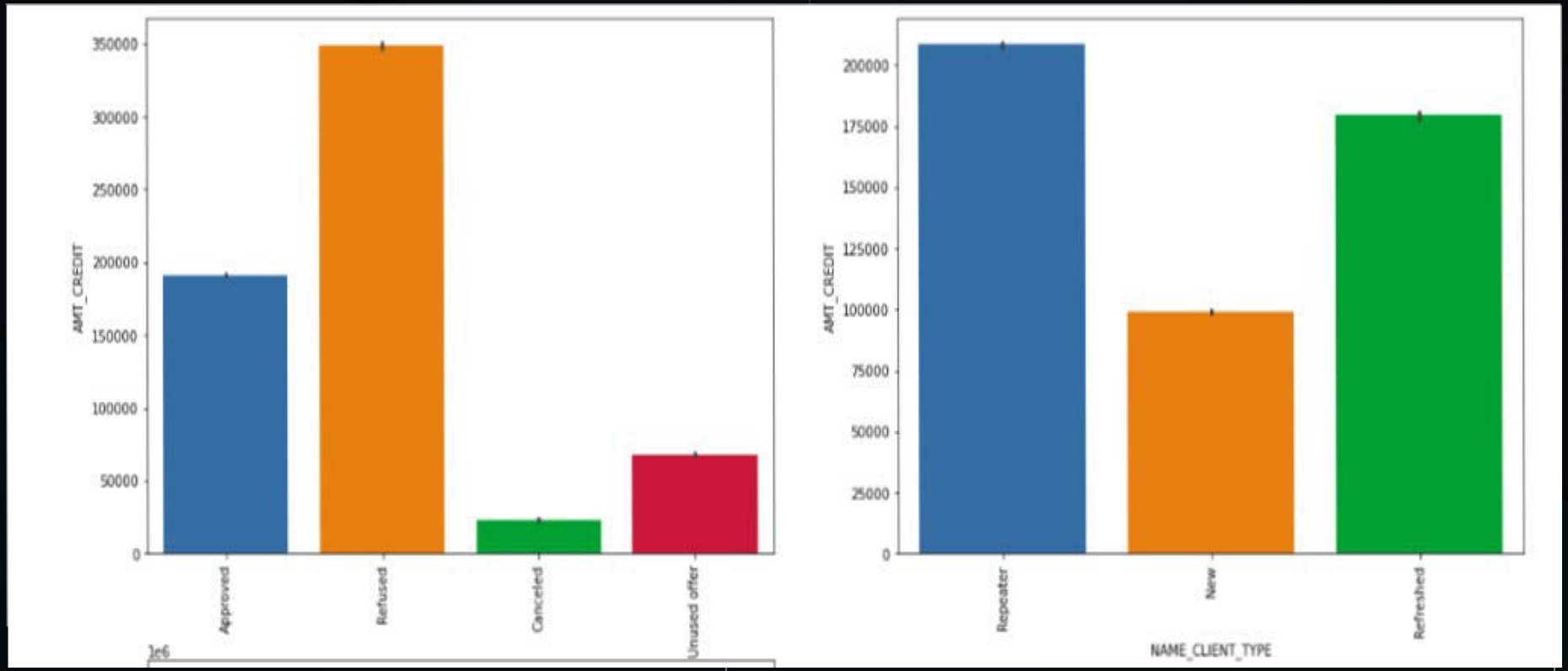

- We can conclude the following things from graps above:-
- Educated pepole get more loans.
- Females got more loans.
- Those who own cars also got more loans.
- The number of cash loans were more as comapred to revolving loans.
- Married people got more loans as compared to others.
- Also people living in Municipal Apartments got more loan as compared to others.

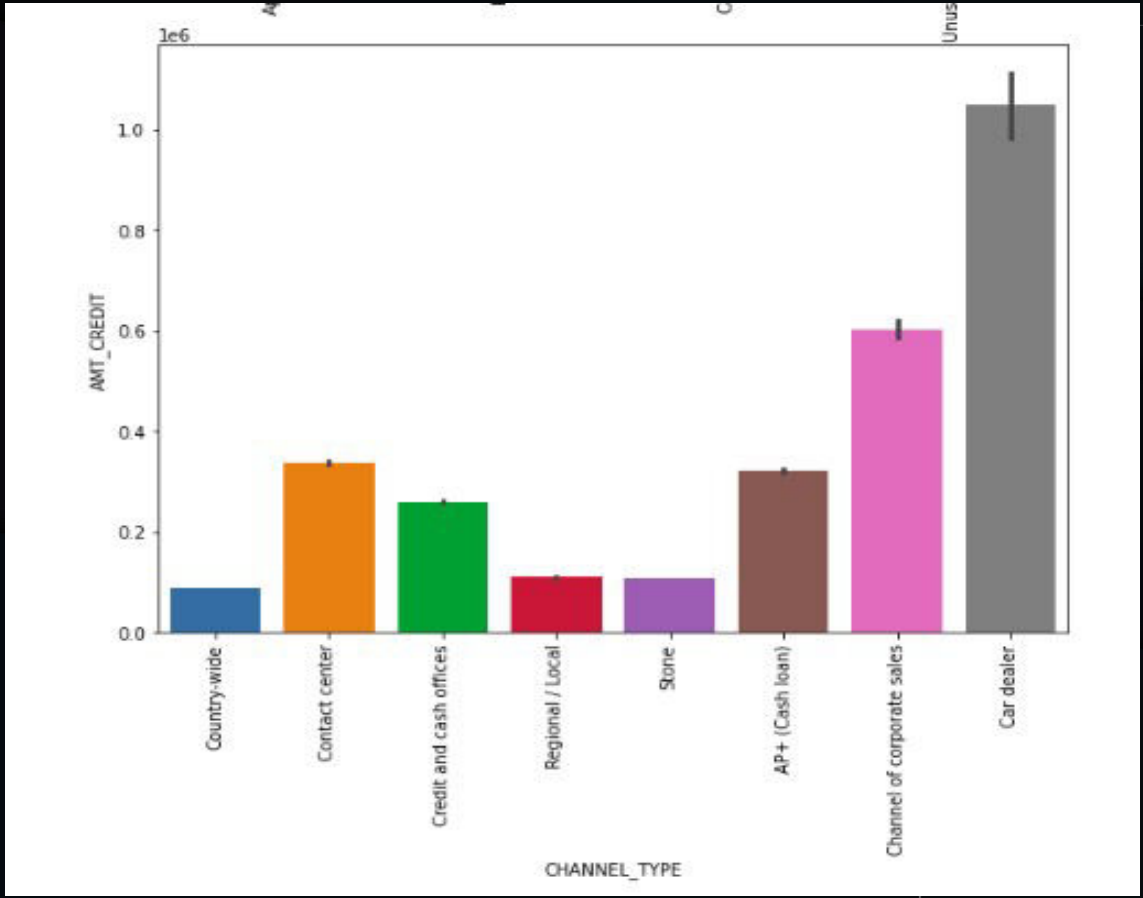
# BIVARIATE IN CONTINUOUS COLUMNS FROM CURRENT DATA AND PREVIOUS DATA




we concluded that they have asked for loan amount in previous application is differ from the amount credit in new application is mostly refused.

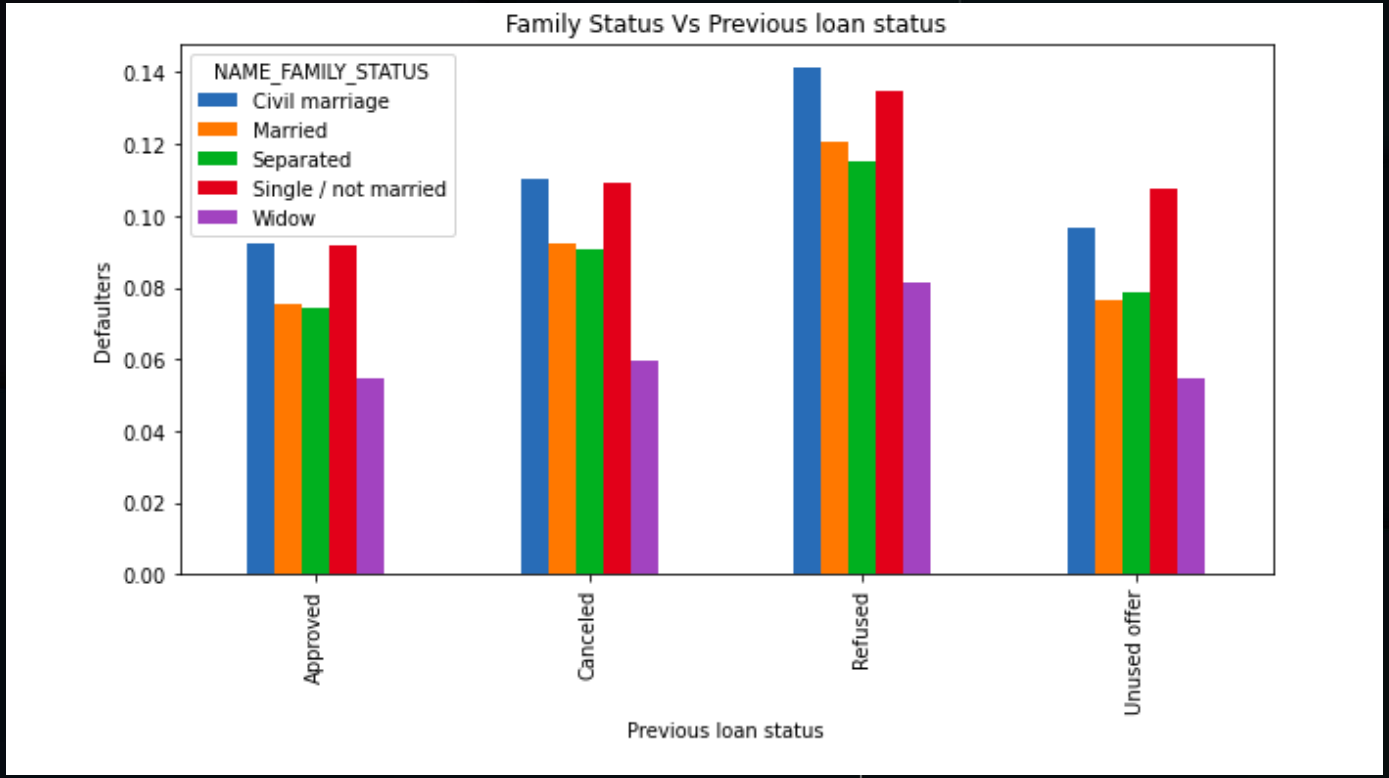
# CATEGORICAL ANALYSIS ON PREVIOUS DATA





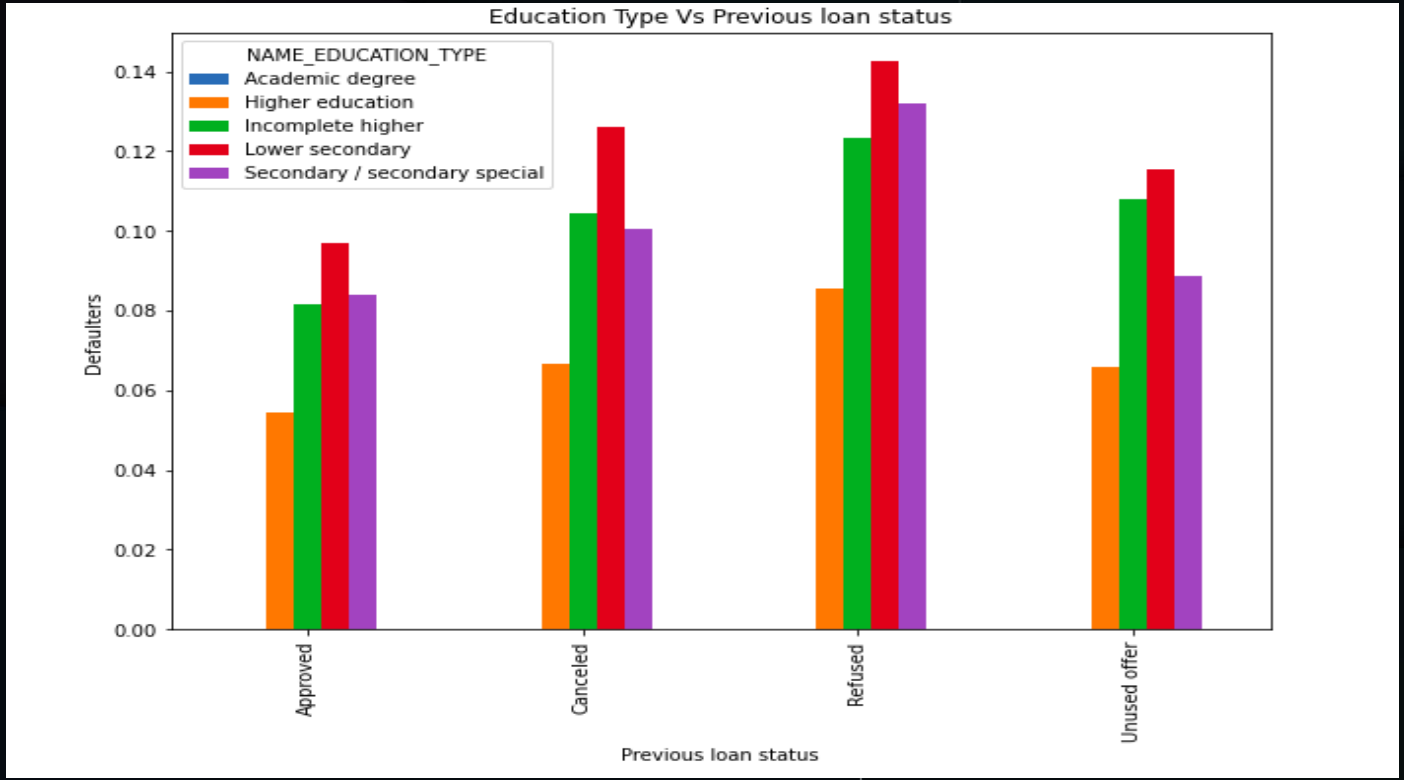

- Credit amount of the loan based on previous applications are mostly refused.
- Repeaters were the client when applying for the previous application
- Through Car Dealer channel we acquired the most client on the previous application.

# FAMILY STATUS VS PREVIOUS DATA LOAN STATUS




From the graph conslusion came Is People who were refused previously are more defaulters.

# EDUCATION TYPE vs PREVIOUS LOAN STATUS




from the graph able to know Analysis Previously refused people with lower secondary education are more defaulted in current application.

# RECOMMENDATIONS

- Clients who are state servants
- Old age people
- Client with high income
- Old female client
- Client with higher education in female category
- Any client whose previos load was approved
- Widow with unused previous loan status
- Refreshed client with unused load status previously

# THANK YOU