CREDIT EDA ASSIGNMENT

RISK ANALYTICS IN BANKING FINANCIAL SERVICES

PROBLEM STATEMENT

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

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

ASSUMPTIONS

- Entire analysis and visualizations is done in python
- Features unwanted for analysis are dropped
- XNA values are considered as NA
- Dataset are loaded into Jupiter notebook
- Understanding the datatype present in the dataset (continuous, categorical, ordinal)

APPROACH

And then data cleaning

- Deleting unwanted columns
- Renaming columns (if required)
- Creating new columns (if required)

Fixing rows and columns

- First determining percentage(%) of null values in the data set
- Setting a threshold percentage(%)
- Dropping any columns whose null percentage is beyond Threshold percentage
- identify the continuous variables with missing values and replace them with the median value.
- Identify the categorical variables with missing values and replace them with mode vales and in some cases create a new category.

Outliers

- Identify outliers using boxplot

APPROACH

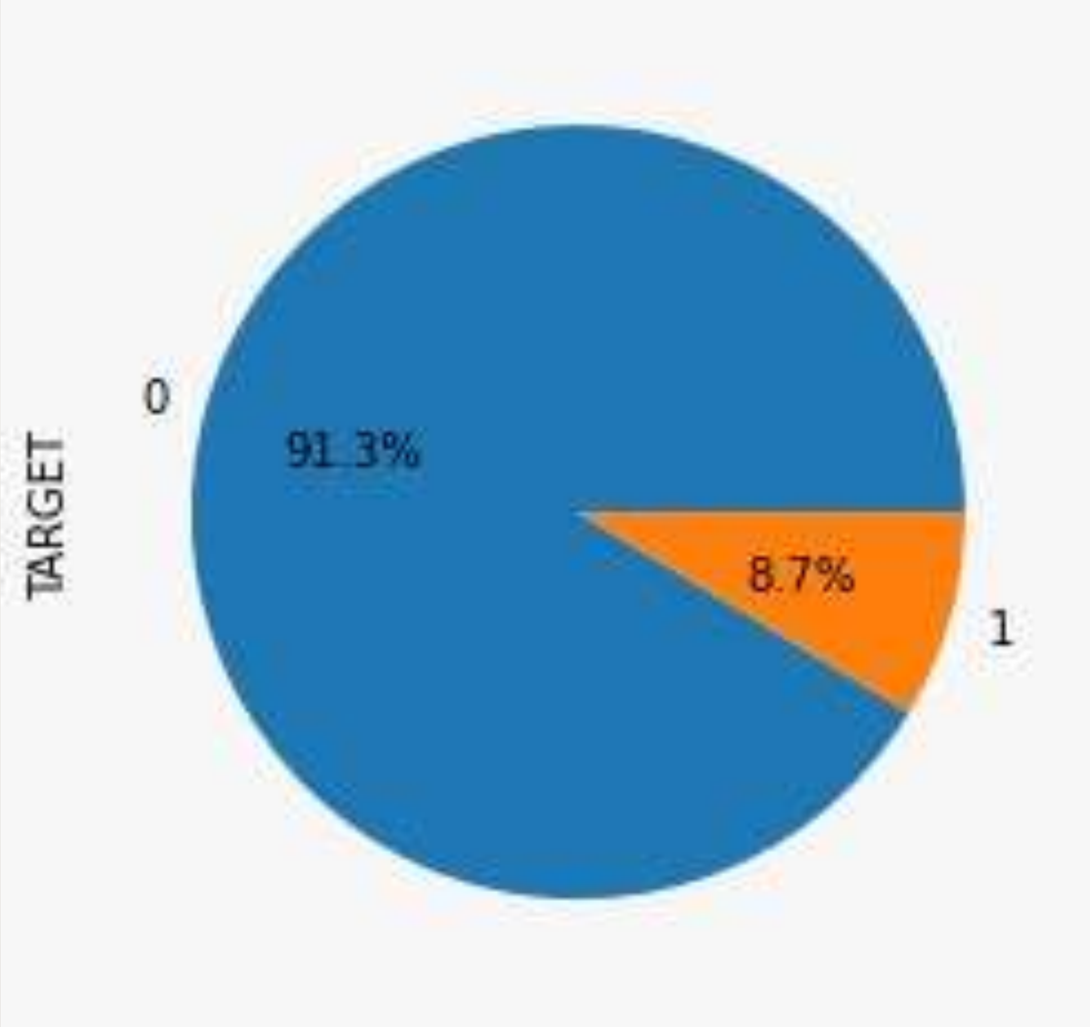
- For continuous columns with outliers ,treat the outliers capping with upper bound values and flooring with lowerbound values.
- Binning the variables wherever required
- Use imputation methods for getting rid of outliers.

Standardizing values

- Identify the columns that need a change in datatype and change accordingly.
- Univariate analys is Perform univariate analysis to understand the distribution of variables across the datasets
- Categorical unordered univariate analysis
- Categorical ordered univariate analysis
- Numeric variable univariate analysis
- Segmented univariate analysis splitting the dataset based on Target variable.
    - Getting the top10 correlation between variables on segmented datasets
    - Bivariate Analysis§ Multivariate Analysis§ Merging the datasets new application and previous application datasets
    - Understanding the distribution based on previous application status
    - Deriving the insight

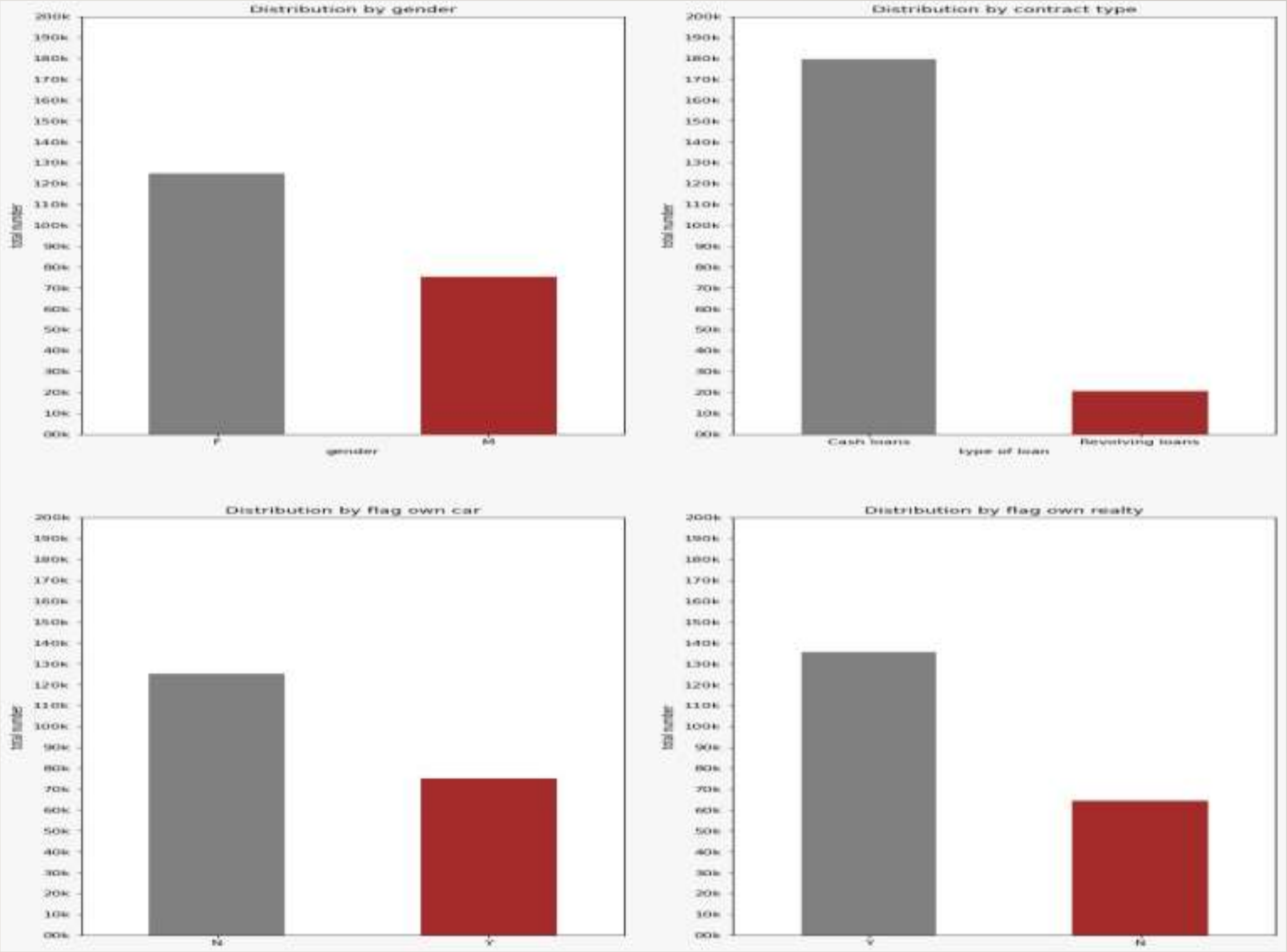
VARIBLE-IMBALANCE

- As can you see the majority of the people are defaulters and minor part is Non- defaulters
- The you there is a major difference almost 10 times



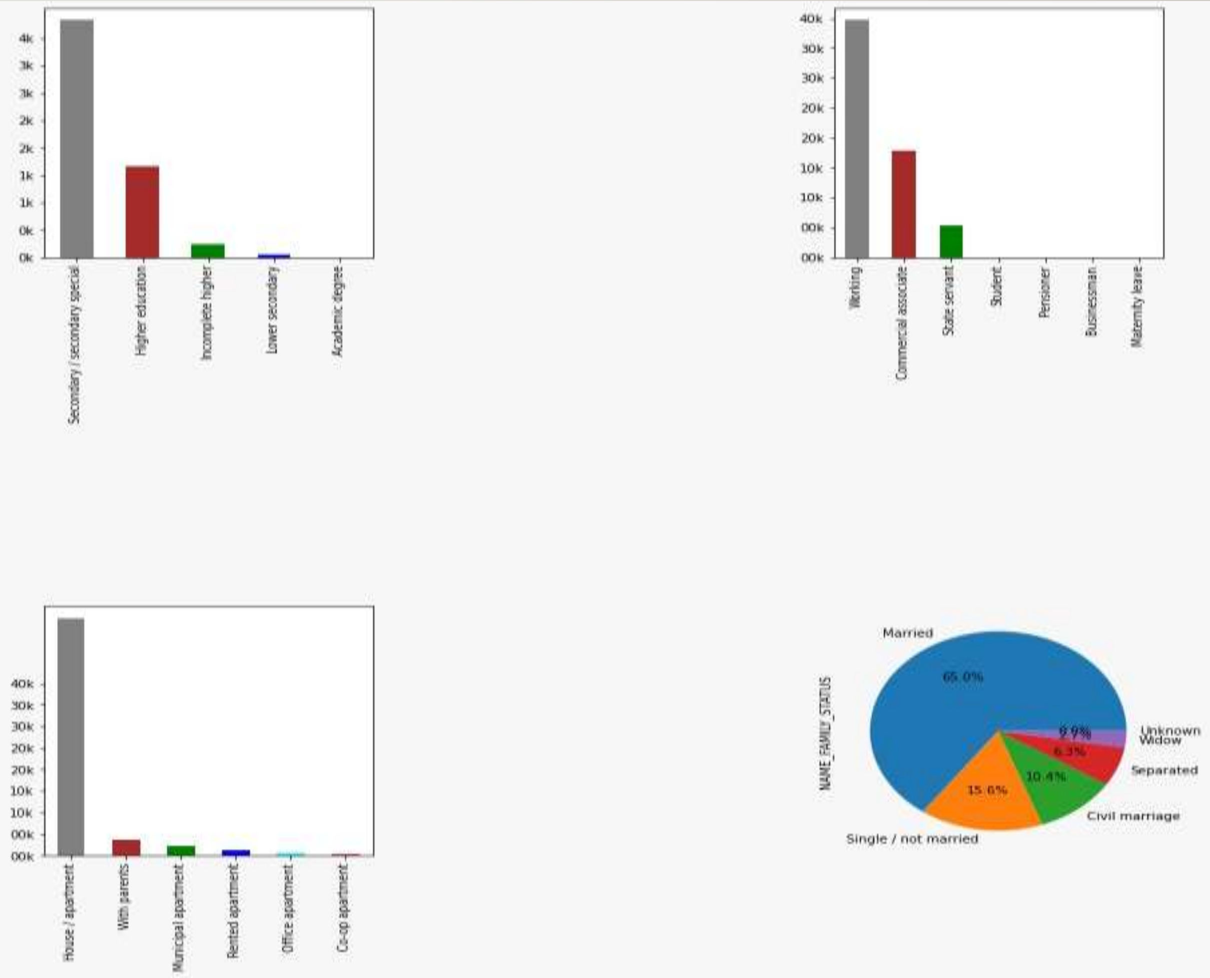
VARIBLE-IMBALANCE

- Here there is 4 graphs each represents gender, type of loan, if person owns a car or not, if person owns a flat or not
- Insight: females are mor than males
- Cash loans are more in percentage
- Majority of people does not own a car same goes for flat



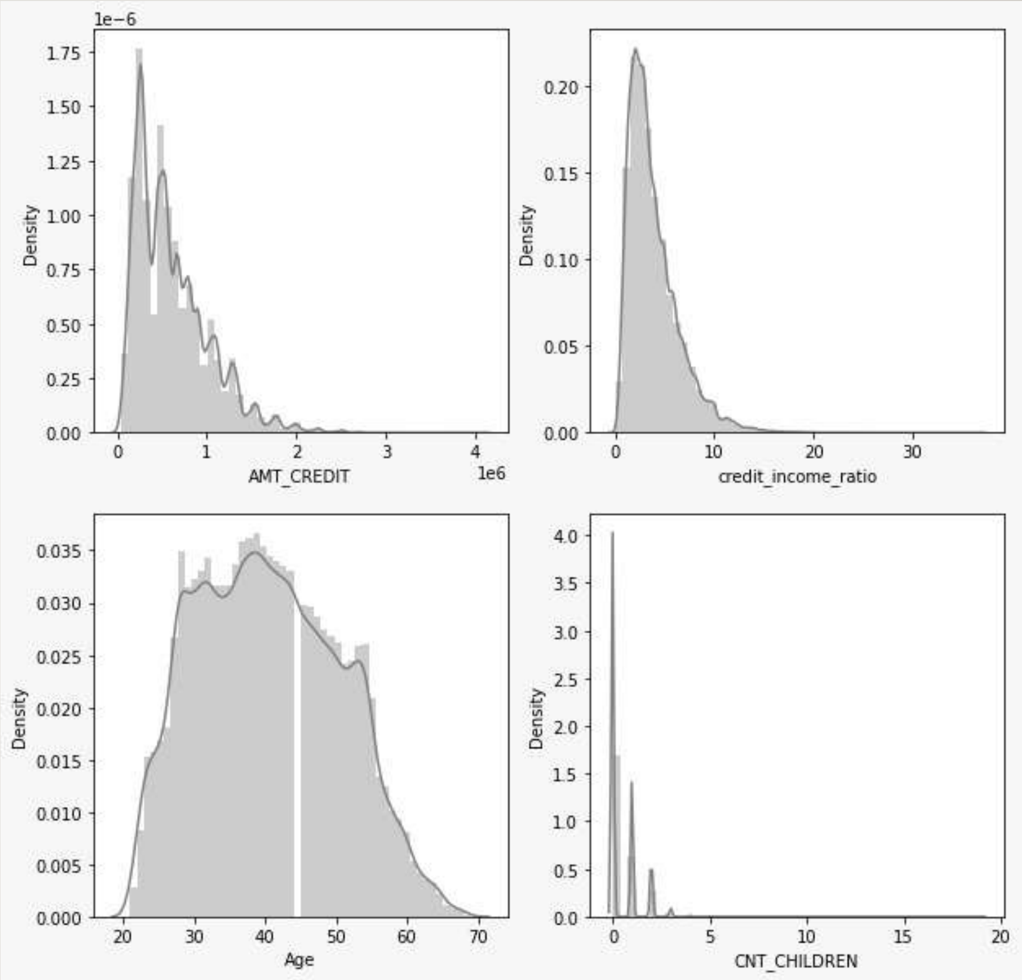
VARIBLE-IMBALANCE

- Insight: the majority of people are secondary education
- Where as in jobs like working and commercial are more in number
- Majority of people are staying at house or apartment
- And majority of them are married followed by non marred/single



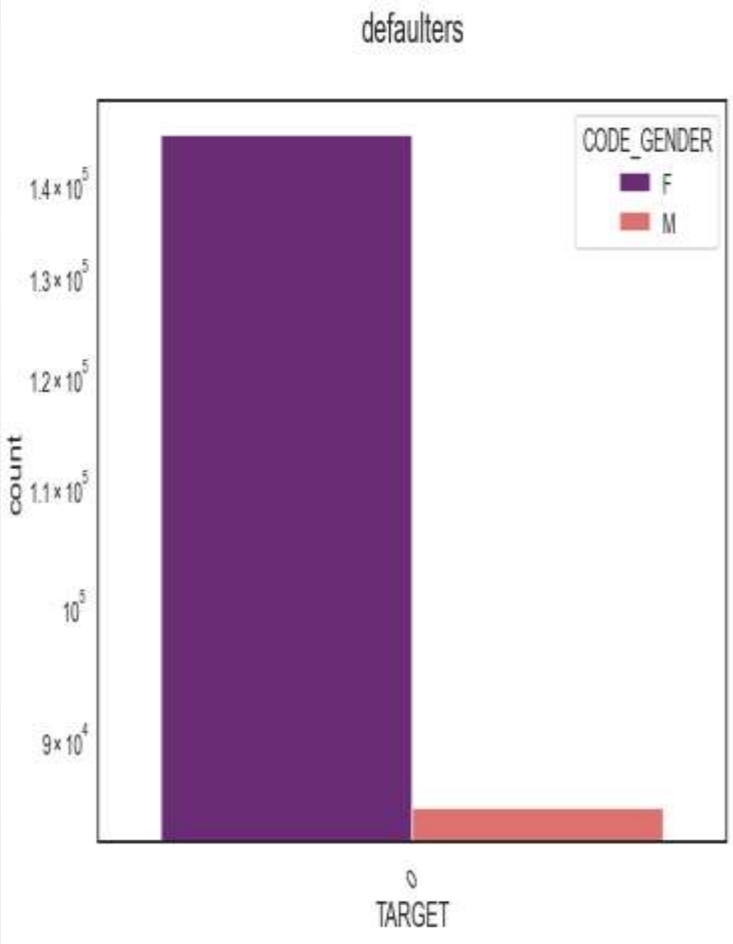
UNIVARIATE ANALYSIS

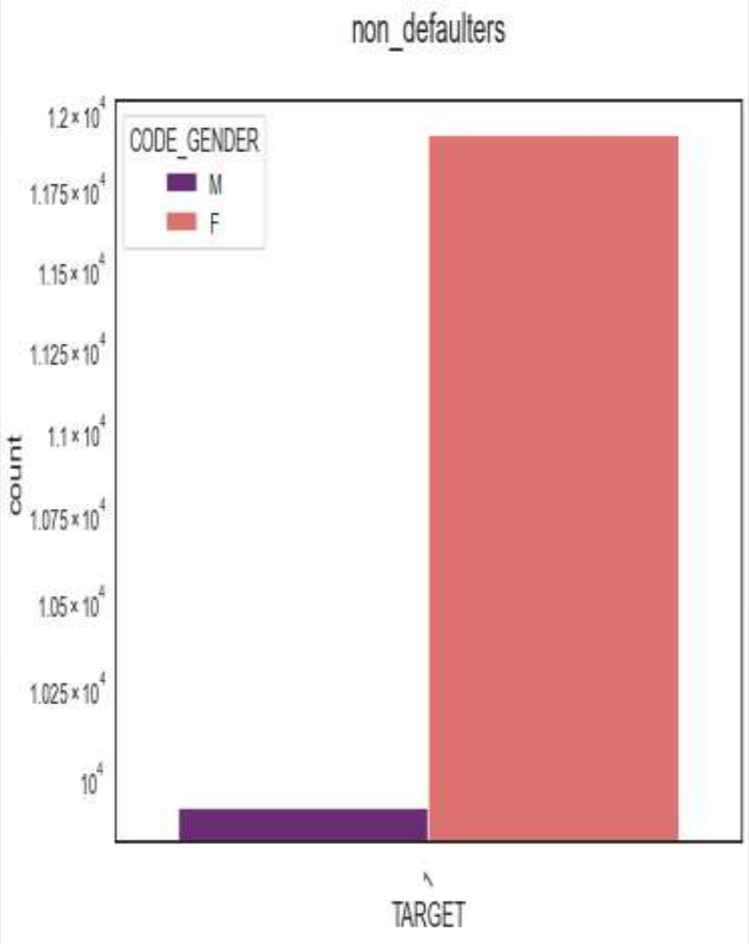
- The four pictures shows different graphs
- Insight: the density of applicants with low credits are high
- Majority of applicants have low credit income ratio
- Most Applicants are in between 35-40 years old
- Majority of family's have <=3 children



SEGMENTED UNIVARIATE ANALYSIS

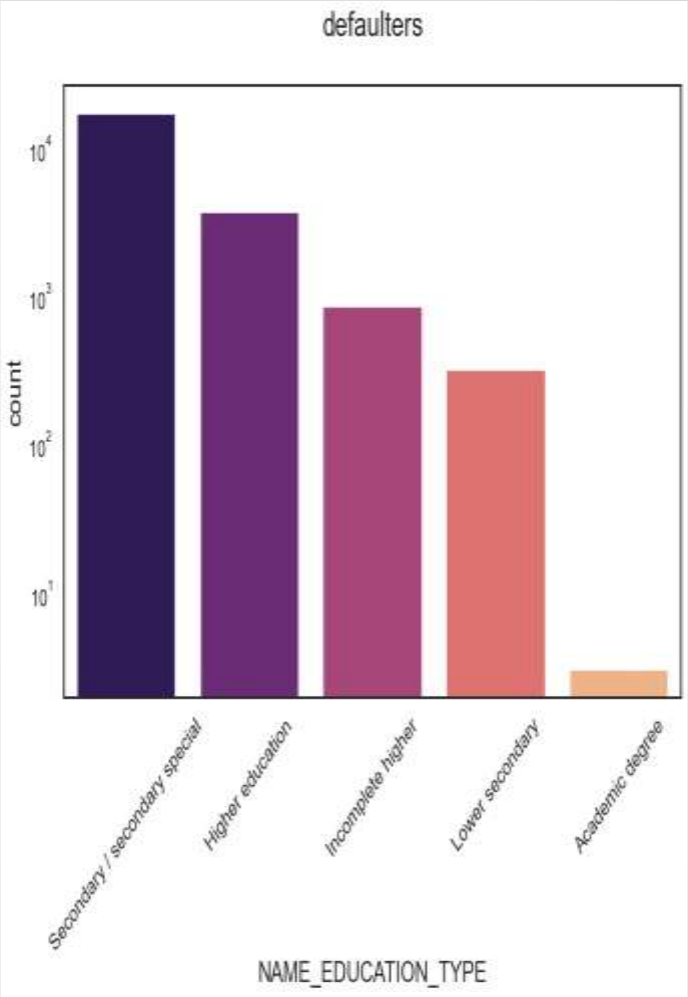
- It is obvious that female have high percentage in both defaulters and Non-defaulters
- Since they are high percentage than male
- The rate of defaulters in less in female when compared to males

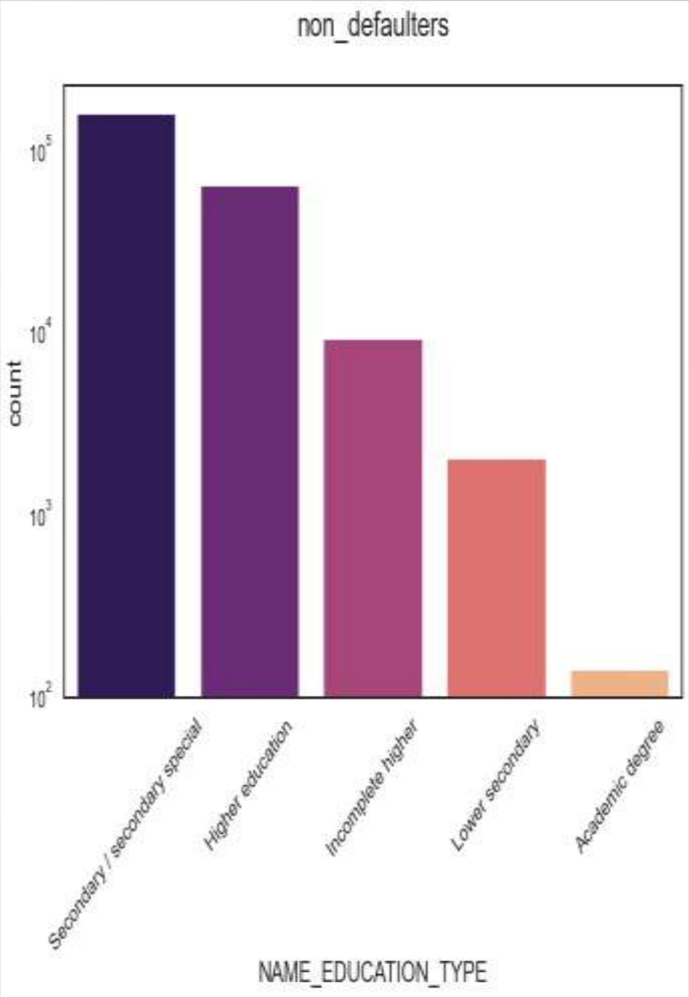




SEGMENTED UNIVARIATE ANALYSIS

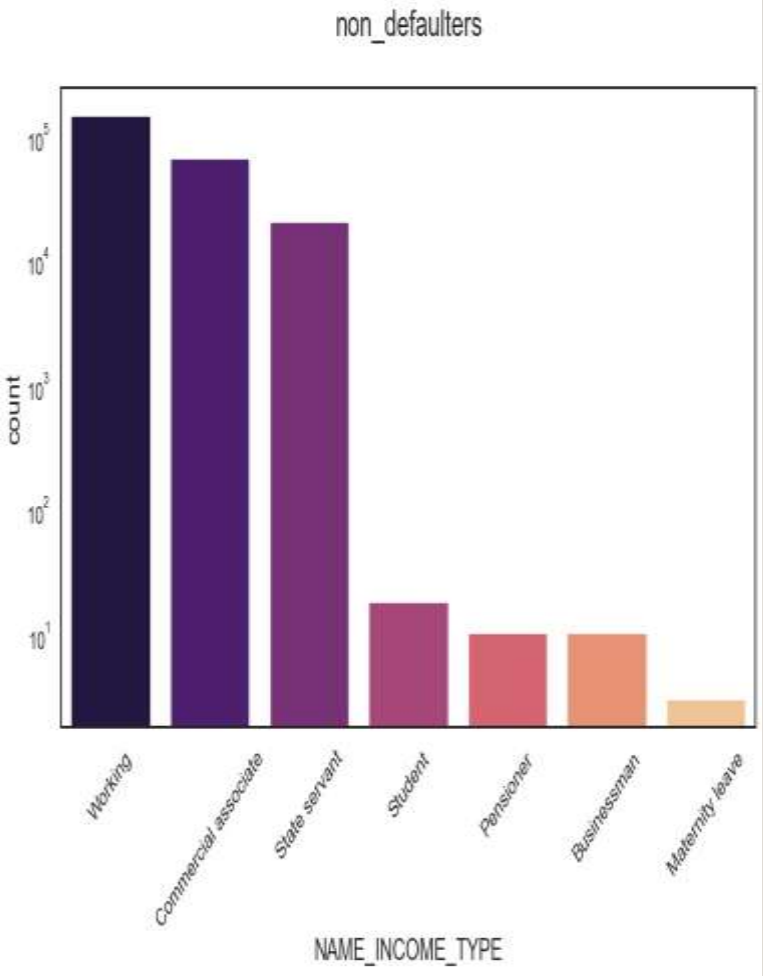
- The majority of applicants are secondary education followed by higher
- Secondary education and lower secondary are high in defaulters

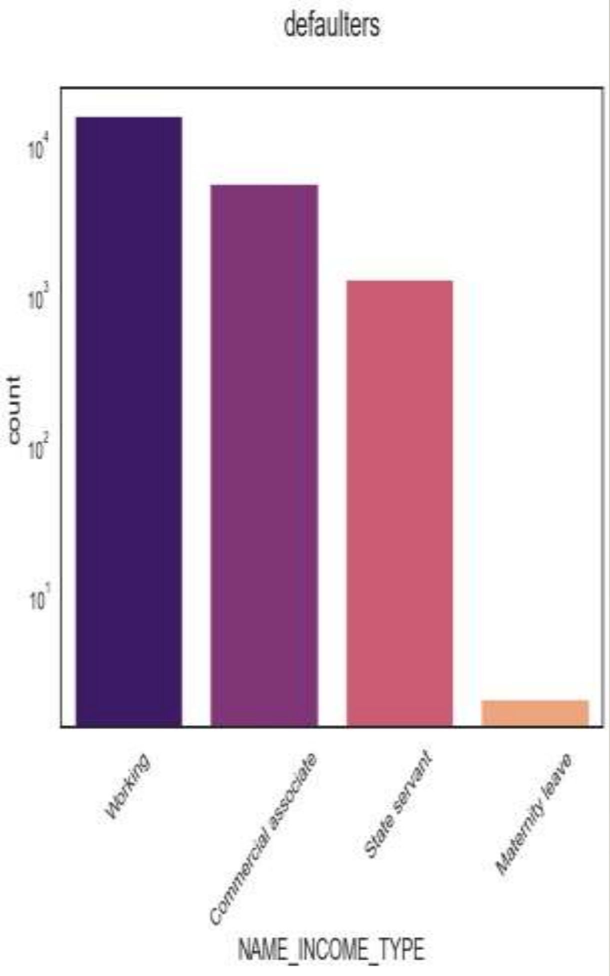




SEGMENTED UNIVARIATE ANALYSIS

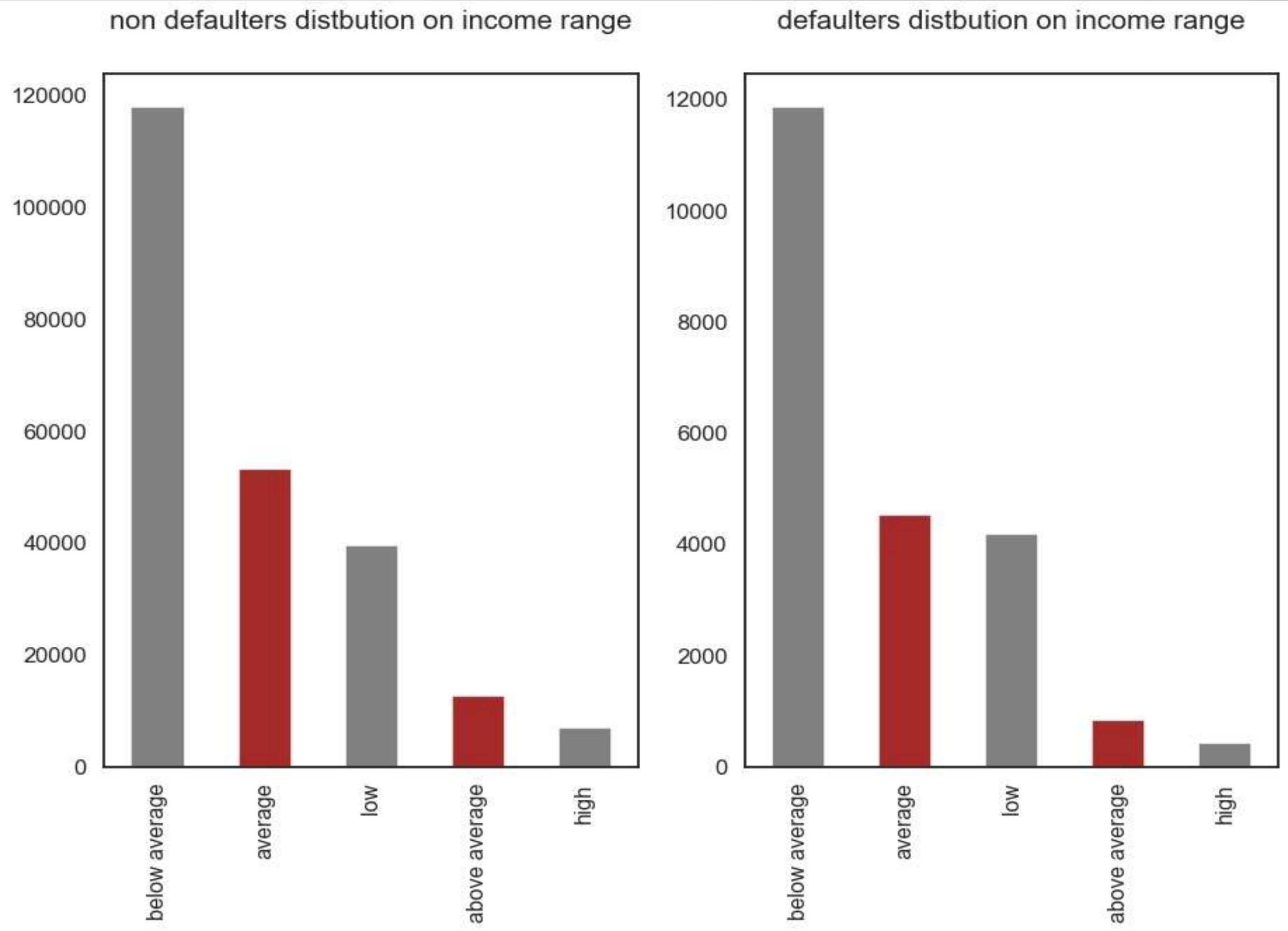
- Students, pensioners, businessman are highly to be non defaulters
- And working are highly to be defaulters followed by commercial associates





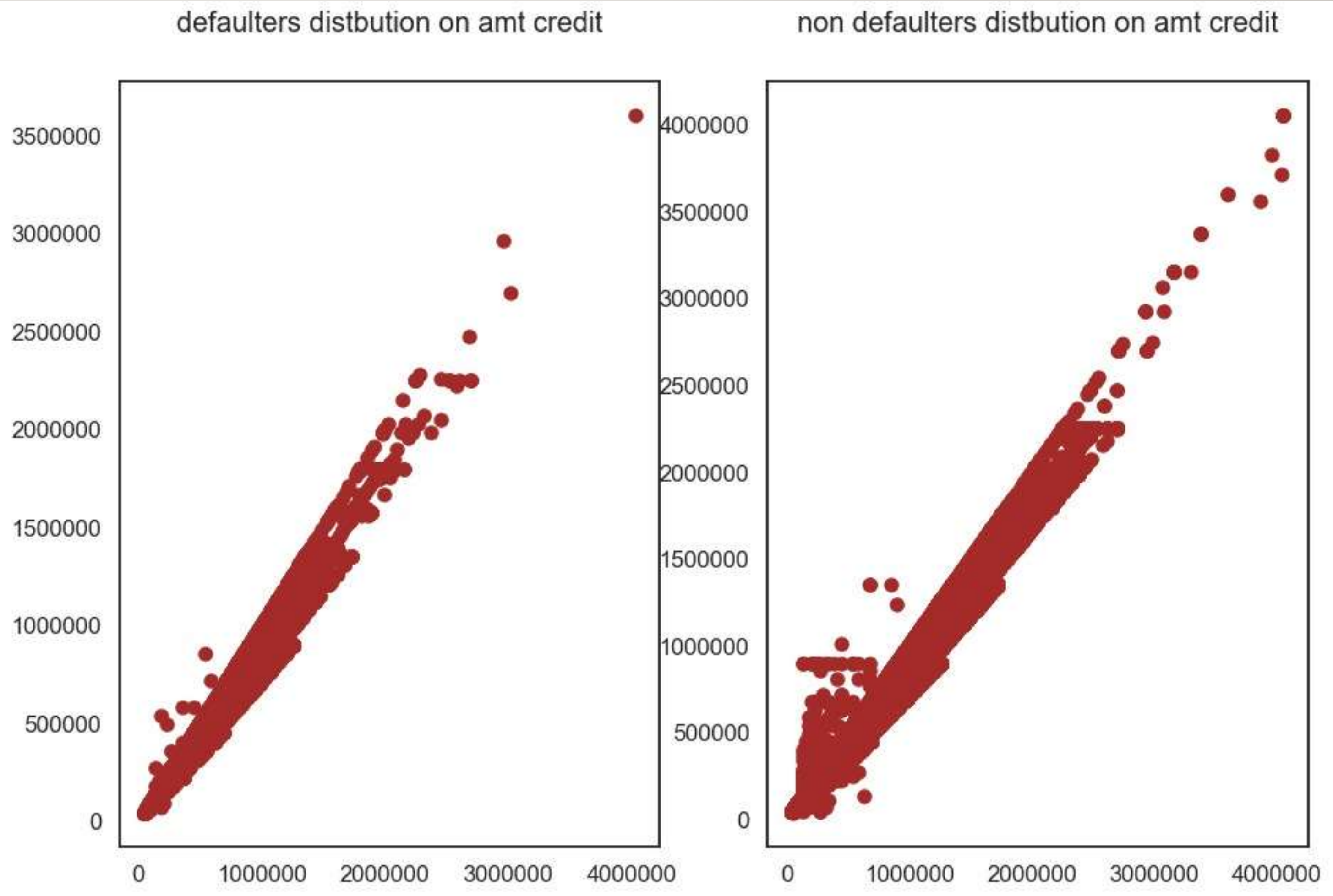
SEGMENTED UNIVARIATE ANALYSIS

- As we can see workers are highly calcified into low so they are mostly high in number
- The low income applicants are mostly likely to be defaulters
- Followed by average



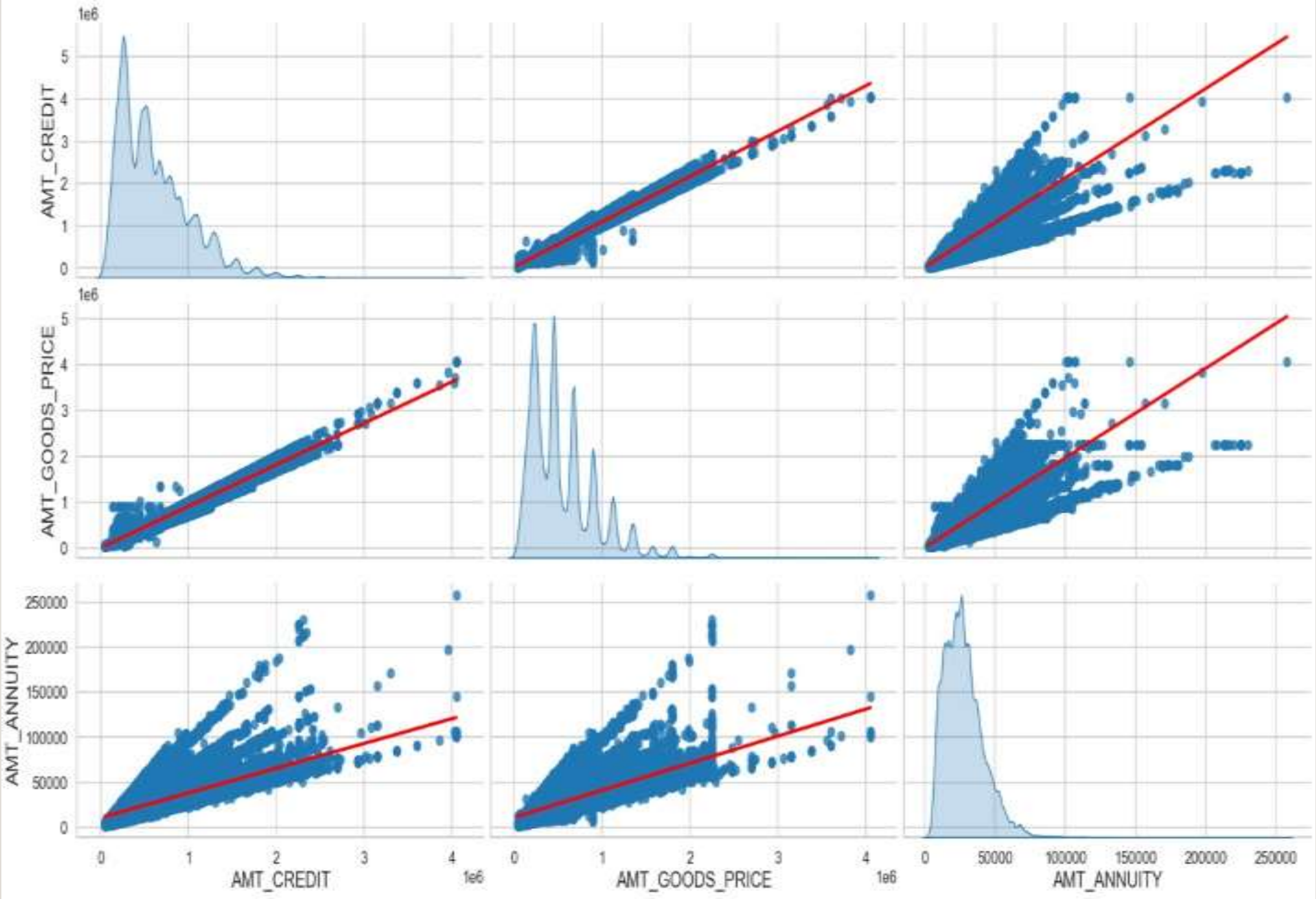
SEGMENTED UNIVARIATE ANALYSIS

- Amount goods price and credit are linear. Higher the goods price higher the credit
- We can say that they are linear proportion
- We can see that as density of defaulters decrease as credit increase
- Insight: applicants with less credits are most likely to be defaulters



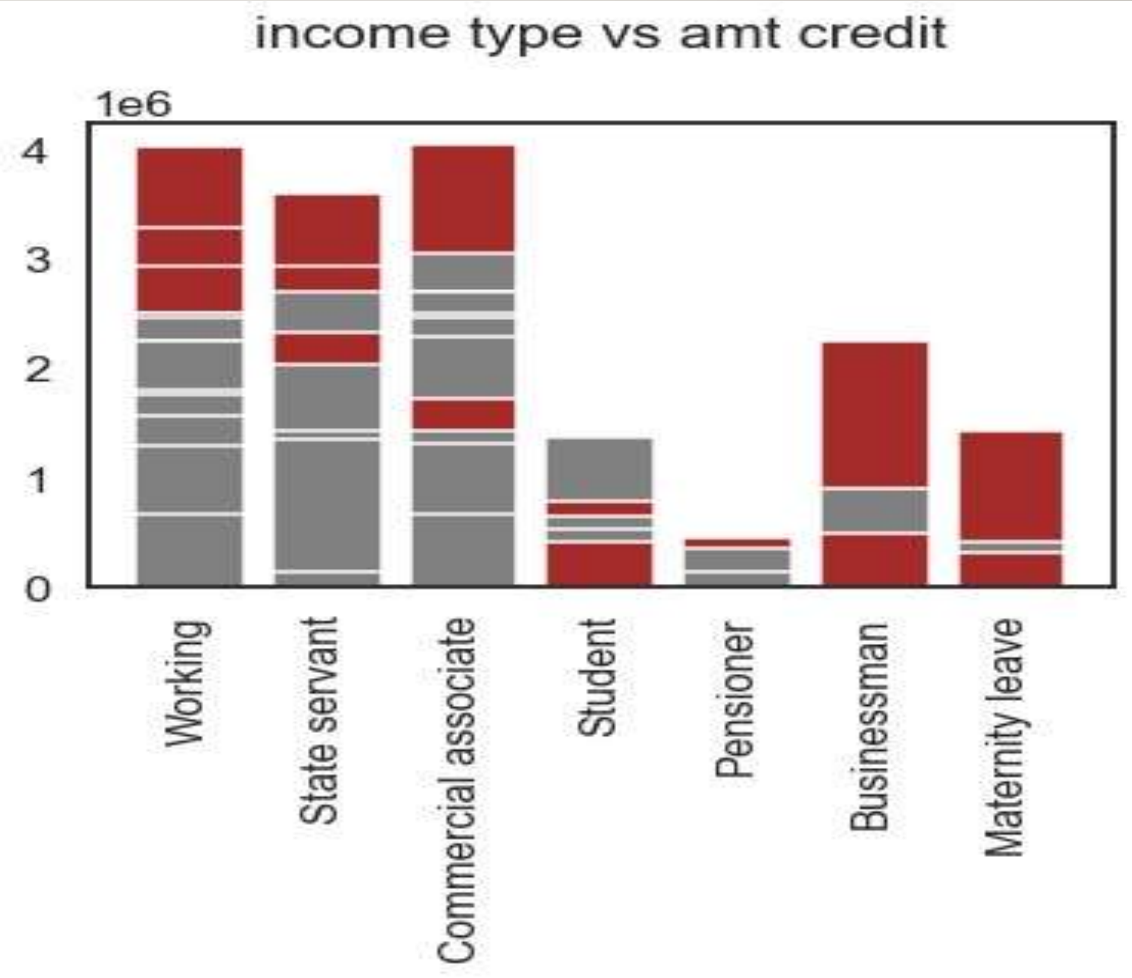
BI VARIATE ANALYSIS

- Amt_credit, Amt_ goods_price, Amt_annuty are in releated to each other if raise raise other raise
- Density of all 3 features to be more to lower level



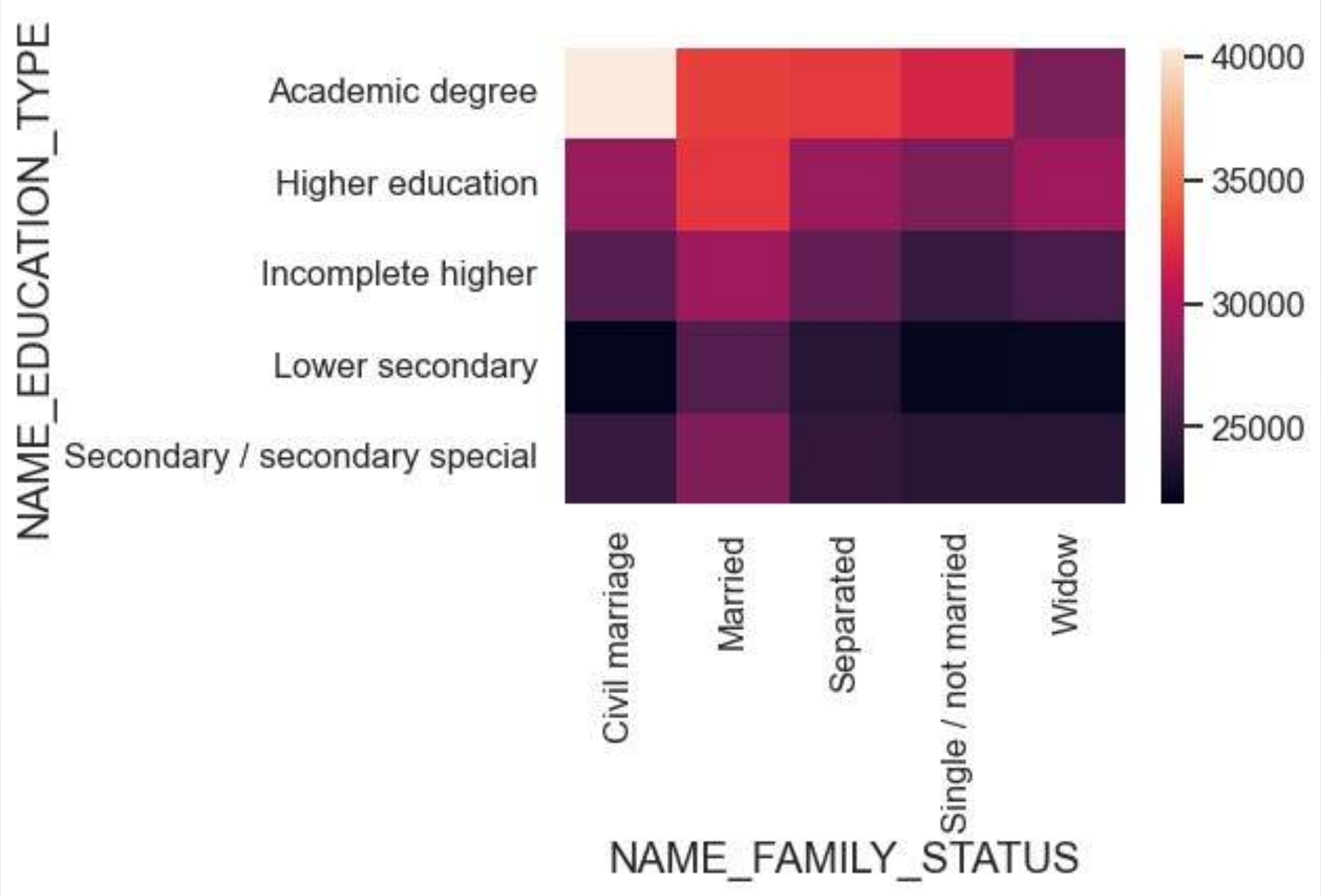
BI VARIATE ANALYSIS

- From the above graphs average credit amount is higher for defaulters than non defaulters
- Working is higher followed by commercial associate and state servant



MULTIVARIATE ANALYSIS

- The amt_anuality is higher for civil marriage and academic degree
- Low for lower secondary education



CONCLUSION:

- From all the graphs and analysis we conclude that
  - Applicants are major in cash type loans
  - The credit of loans increase as goods price increase
  - High risk defaulters
    1. Gender: male
    2. Income type :working
    3. Income group: below average and low
  - It is best to avoid
    1. Age group:60+
    2. Income type :pensioneer
    3. Education: high education
    4. Family status: widow and civil manager

CONCLUSION:

- Best to give
    1. Gender: female
    2. Age: 35-40
    3. High credit
    4. Education: academic degree