Credit EDA Case Study

BY: Shikha Kakkar & Pieyush C Joy

Problem Statement

- Two types of risks are associated with the bank's decision:
  - If the applicant i 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.
  - **The company wants to understand the driving facotes(or driver variables) behind loan default, i.e. the variables which are strong indicators.**

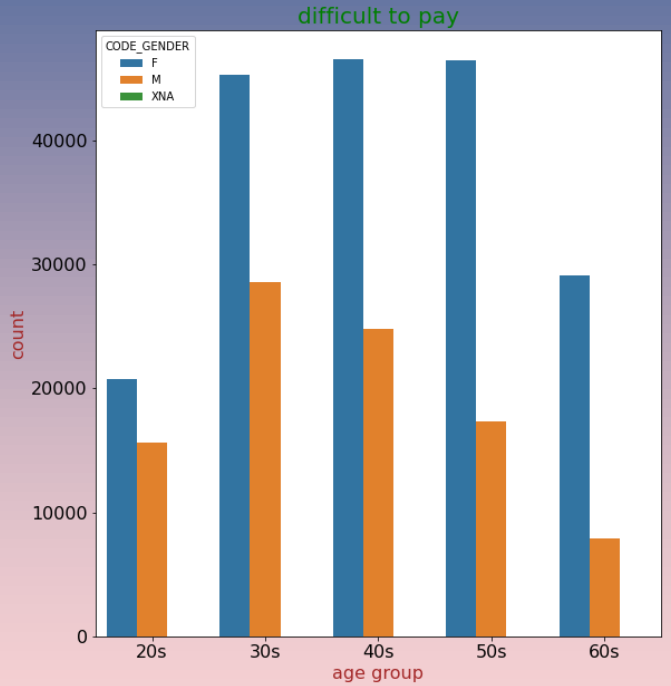
EDA Techniques

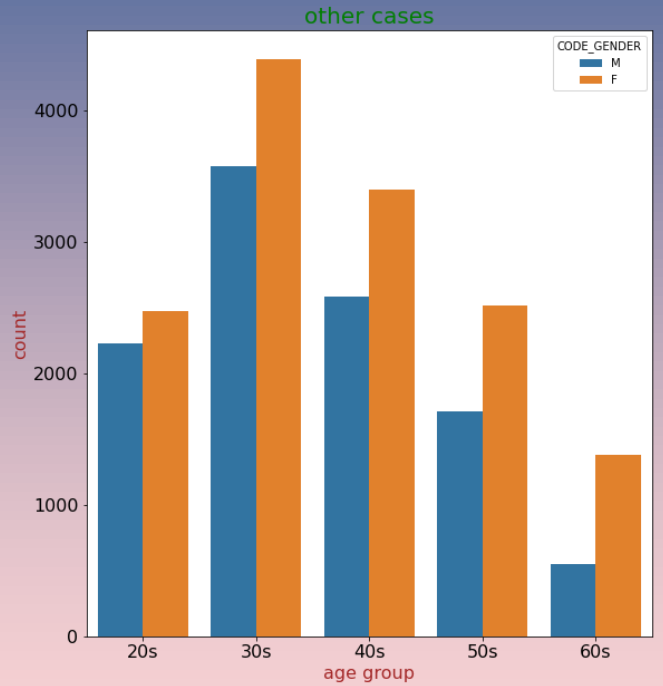
- Start by importing the ‘application_data.csv’
- Check the structure of the data (Normal Routine Check)
- Data Quality Check and Missing Values
  - Find the percentage of missing values for all the columns.
  - Remove Columns with high missing percentage.
  - For columns which has less percentage (around 13% or so), we need to check what will be the best metric to impute the missing values.. If the column are checking is a categorical column, check which category we can use to fill the nulls. For other check does it have a mean or median can be imputed r not. Other cases may be imputing with 0. We need to do this task for some variables and not fall, say 5.

Approach towards problem statement.

- We will be using techniques of Exploratory Data Analysis to meet to the conclusion of our statement.
- We will study the data and use Univariate and Bivariate Analysis.

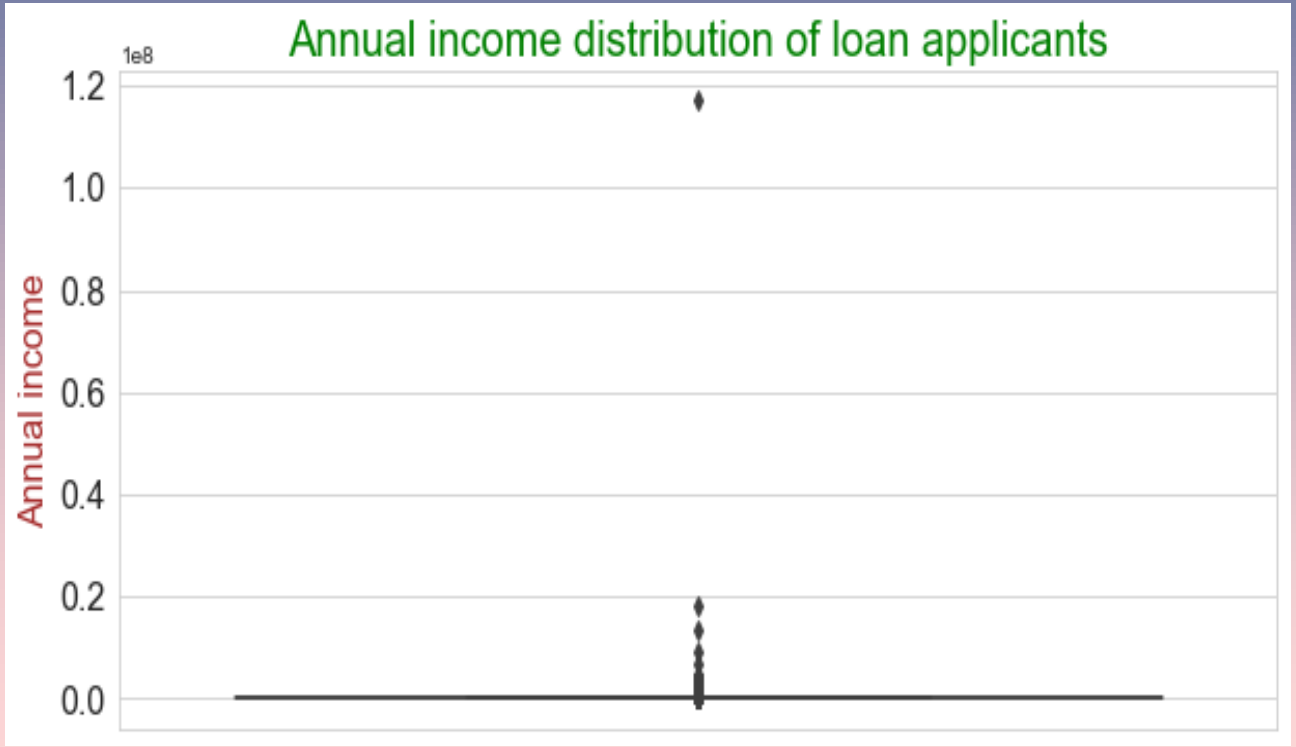
Loan Annuities- Both Male and Female





- Annuities are loans that are paid back over a set period of time at a set interest rate with consistent payments each period. A mortgage or car loan are simple examples of an annuity.
- Here what we observe is the nature of repayment of loan by two categories of people. Who can be differentiated as people who are facing an issue/ difficult to repay the loans in time and then the other people who face other issues like bank issue, payee default, etc.
- Both male and female fall under the two categories.
- In Graph 1 we can infer that female are facing an issue in repaying loans of loans above 40,000Rs and more. Male on the other hand do not face much of an issue when it comes to paying the loans.
- In Graph 2 we can infer that female are facing an issue in repaying loans of loans of all loan rates between ages 30-60. Male on the other hand do face an issue when it comes to paying the loans.

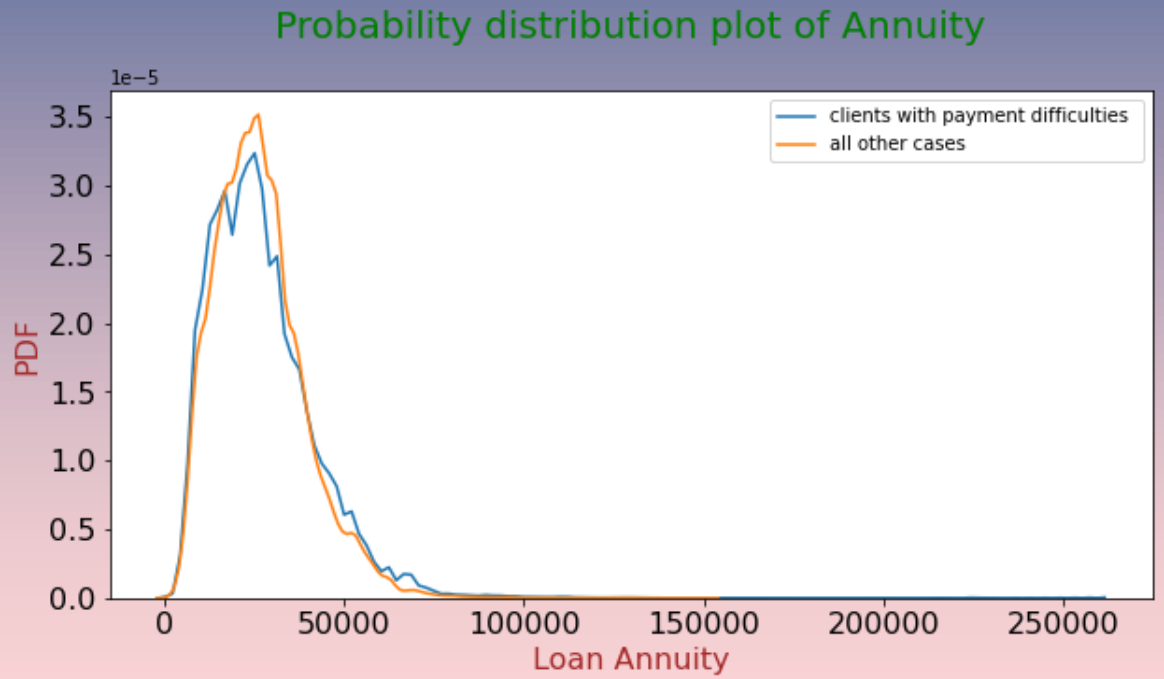
Annual income distribution of loan applicants



Annual Income – Loan Applicants

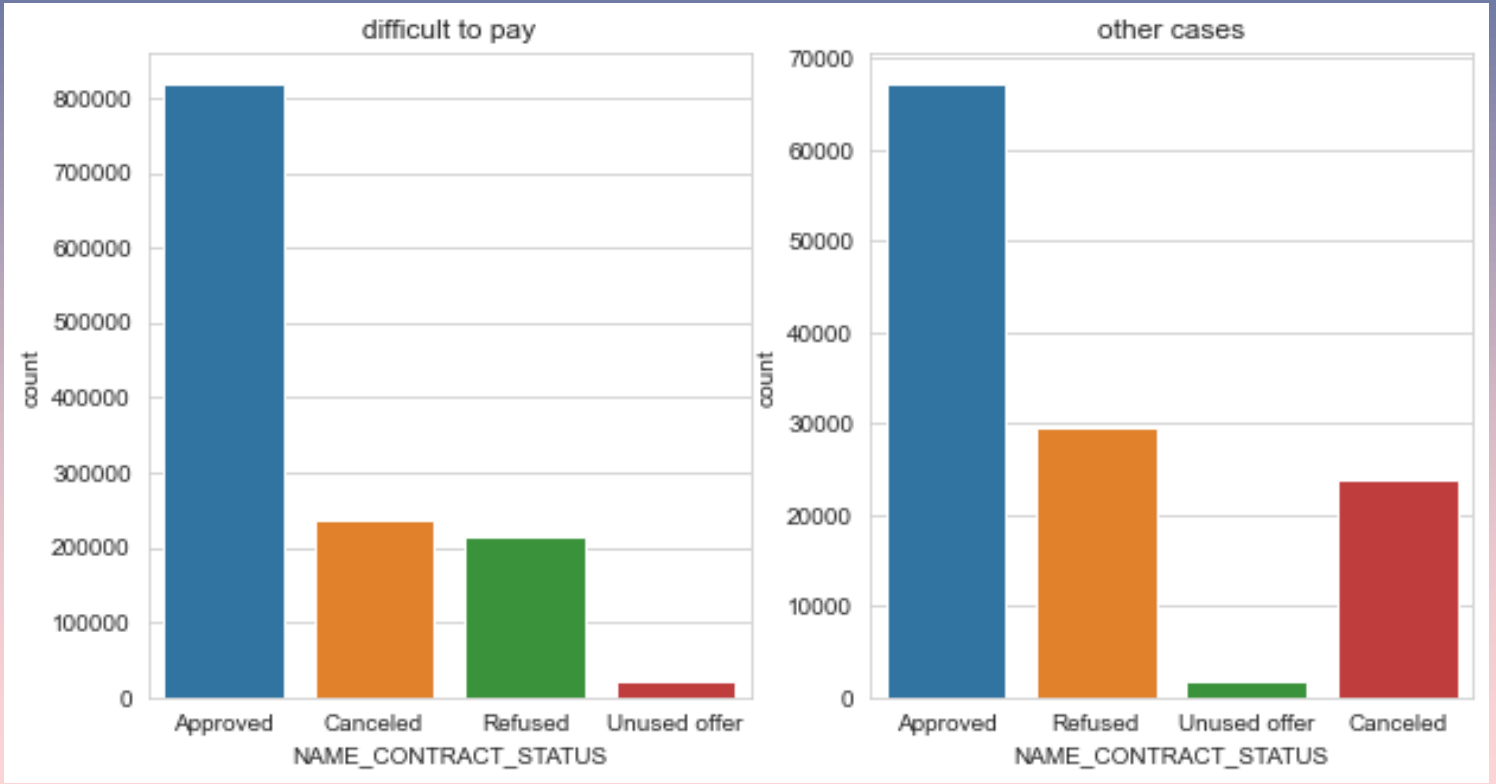
- Outliers are important to show, where the primary sample lies at.
- We can infer from this chart, that the most number of loan applicants fall under the annual income of 2 lakhs or less.
- The other loan applicants can be considered as loan borrowers for various business/ investment purposes as they fall under the annual income of 10 lakhs or more.

Annual income distribution of loan applicants



- This graph clearly represents both categories and the probability of repayment of the loans.
- We see clearly see that the payment probability is high within the annuity rates from 1000-50000/- but will drastically reduce given two circumstances as people who are facing an issue/ difficult to repay the loans in time and then the other people who face other issues like bank issue, payee default, etc.
- We can infer that payments above the rate of 50,000/- must be avoided in loan lending.

Difficult to pay

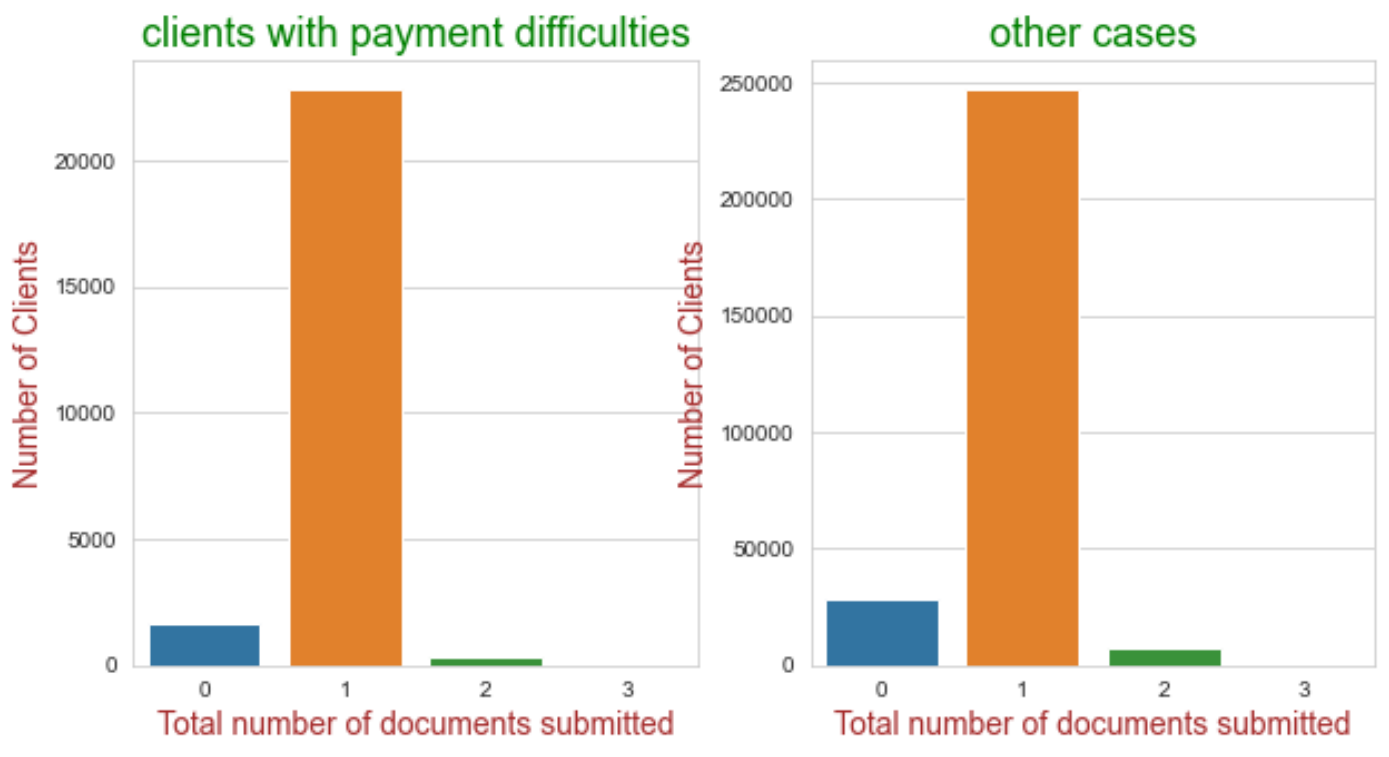


the number of clients who were refused earlier are more in number difficult to pay cases than other cases

the number of clients who got approved applications previously has less number of difficult to pay cases than other cases

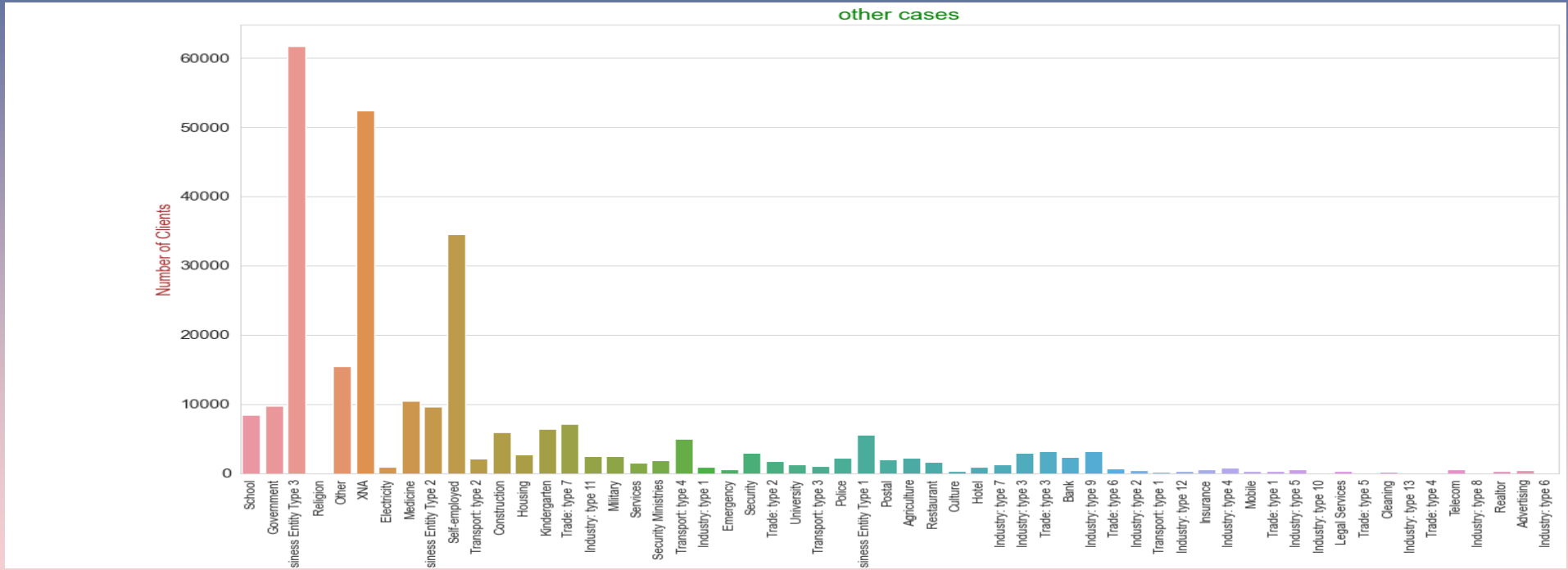
in 'cancelled' and 'unused offer' categories all the clients have both type of behaviour

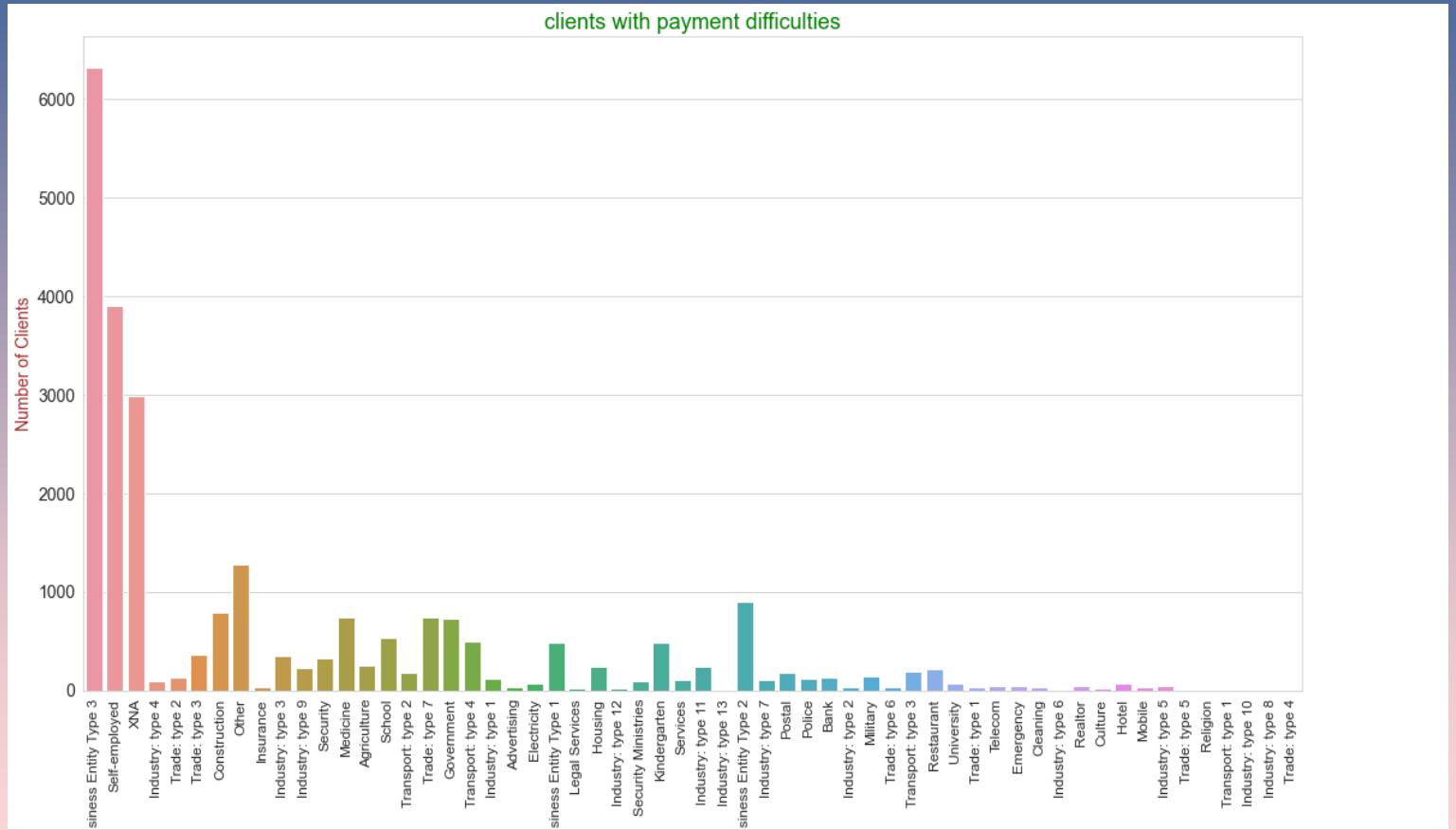
Document submission



- Above chart depicts the number of documents submitted by the number of clients.
- If we closely observe that in both the cases ie. Clients with payment difficulties as well as other cases do not have any impact over the loan repayment.
- We can hence conclude that on both the cases document submission is not a factor of repaying loans.

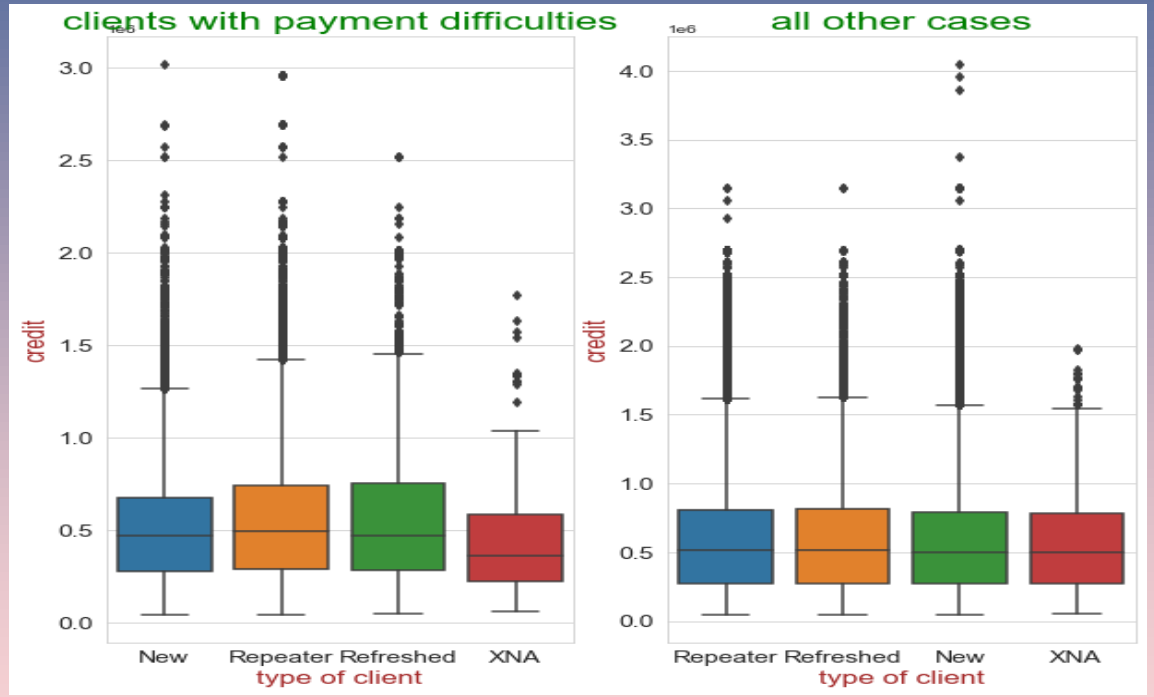
Organization Type




clients with payment difficulties
Number of Clients
0
1000
2000
3000
4000
5000
6000
siness Entity Type 3
Self-employed
XNA
Industry: type 4
Trade: type 2
Trade: type 3
Construction
Other
Insurance
Industry: type 3
Industry: type 9
Security
Medicine
Agriculture
School
Transport: type 2
Trade: type 7
Government
Transport: type 4
Industry: type 1
Advertising
Electricity
siness Entity Type 1
Legal Services
Housing
Industry: type 12
Security Ministries
Kindergarten
Services
Industry: type 11
Industry: type 13
siness Entity Type 2
Industry: type 7
Postal
Police
Bank
Industry: type 2
Military
Trade: type 6
Transport: type 3
Restaurant
University
Trade: type 1
Telecom
Emergency
Cleaning
Industry: type 6
Realtor
Culture
Hotel
Mobile
Industry: type 5
Trade: type 5
Religion
Transport: type 1
Industry: type 10
Industry: type 8
Trade: type 4

ORGANIZATION TYPE

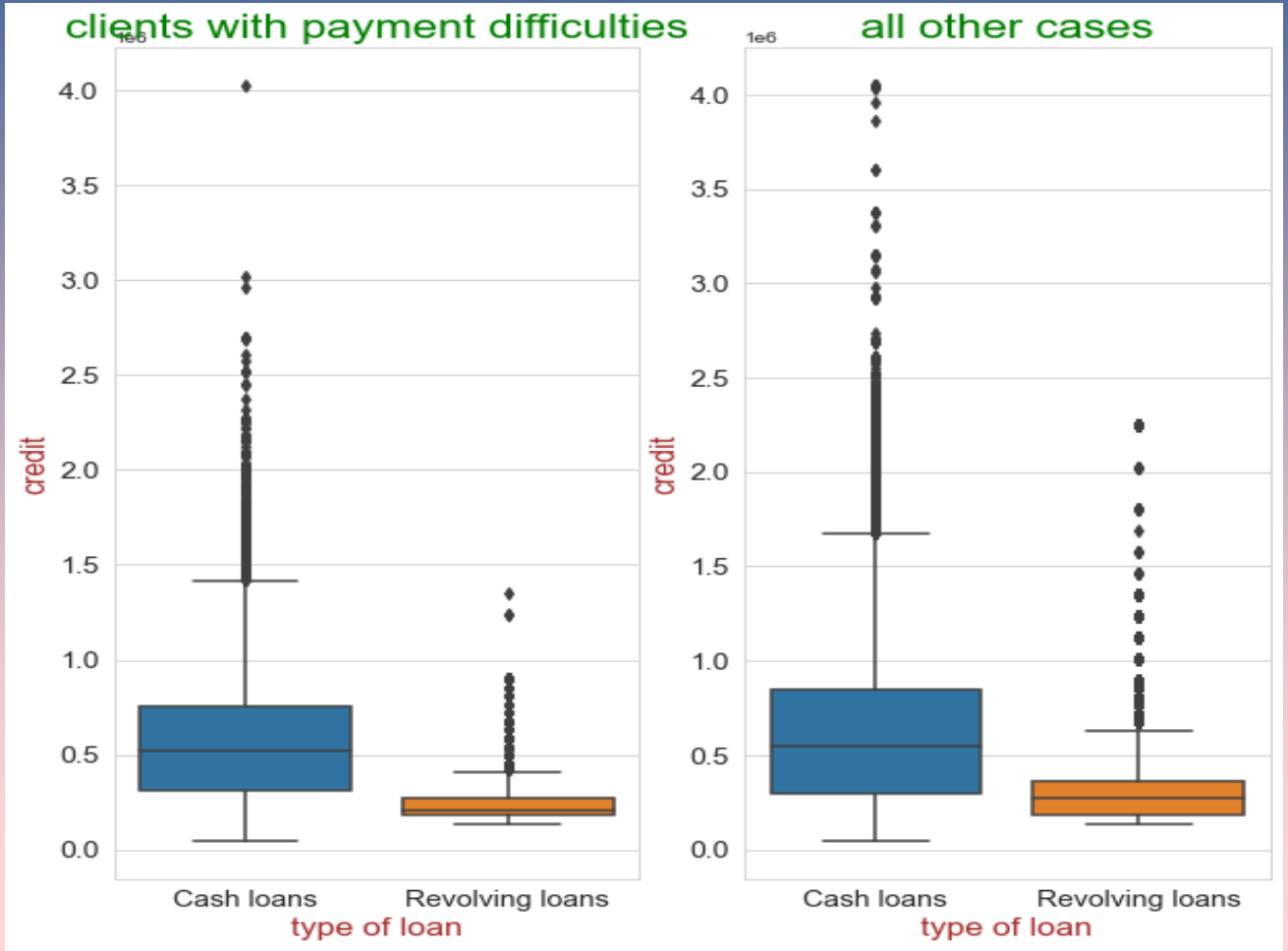
- This chart depicts the number of clients and their respective occupation/organization through which we can analyse what is the purpose of them applying for the loans.
- We see that the main people who avail for loans whilst having payment difficulties are the clients who come under the Business Entity, Self Employed respectively.
- We see that the main people who avail for loans who fall under the category of ‘other cases’ are the clients who come under the Business Entity, Self Employed respectively.
- Hence we can conclude that the main purpose of applying for loans are mainly by the business enterprises and self employed clients.


clients with payment difficulties
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
credit
New
Repeater
Refreshed
XNA
type of client
all other cases
1e6
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
credit
Repeater
Refreshed
New
XNA
type of client

EDA Technqiues

- We see that there are two categories of loan applicants, the ones who are fresh applicants, repeaters and refreshed applicants.
- From the above chart we can note that most of the new applicants do not have access to avail high amount of loans.. whereas the repeaters do avail high amount of loans.

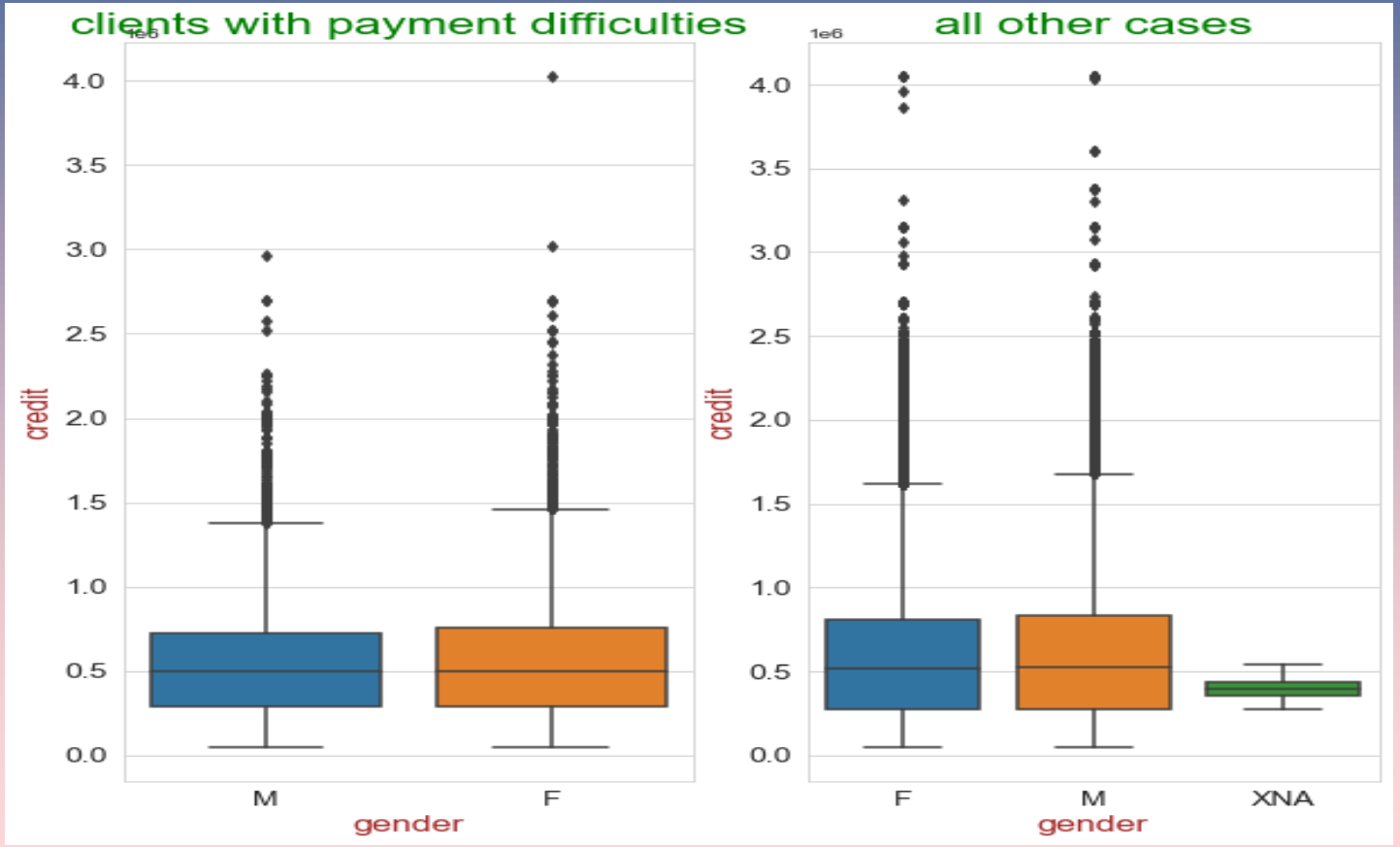
Amount credited across different type of loans



EDA Technqiues

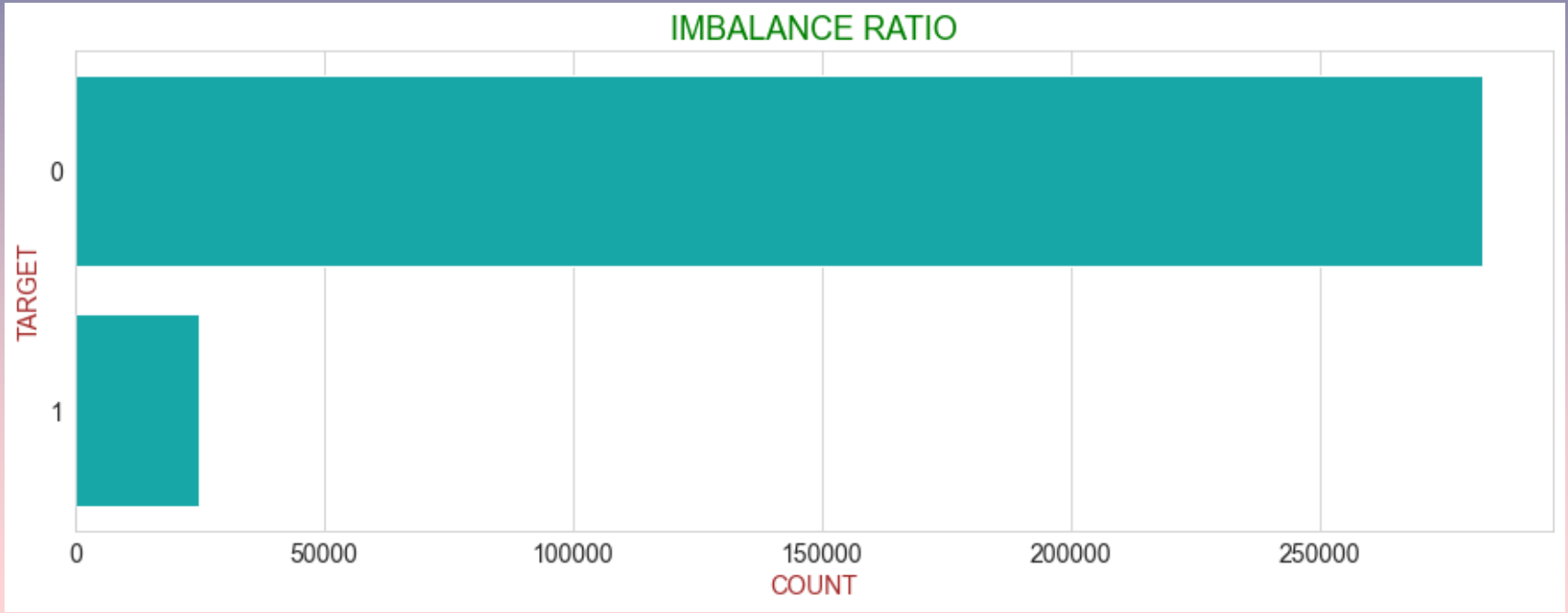
- We see that cash loans are the most popular type of loans among both the cases ie. With payment difficulties as well as other cases.
- The cash loans range from a value ranging up to a credit of 1.5 Lakhs to 3.0 Lakhs and the most availed loans fall below 1 Lakh.
- Revolving loans on the other hand do not have a high credit value compared to cash loans, we can hence conclude that cash loans are the most preferred type of loan in both cases.
- There are two segments shown here revolving and cash loans.
- Box size is larger for cash loans for both the difficult to pay cases and all other cases
- Median is for cash loans is almost equal for both the cases this implies 50 percent of clients has taken same amt of loan.Even maximum credit in this segment is same in both the cases.
- But in revolving loan segment it is different median is lower for clients difficult to pay case.

Amount credited among both the genders



IMBALANCE RATIO

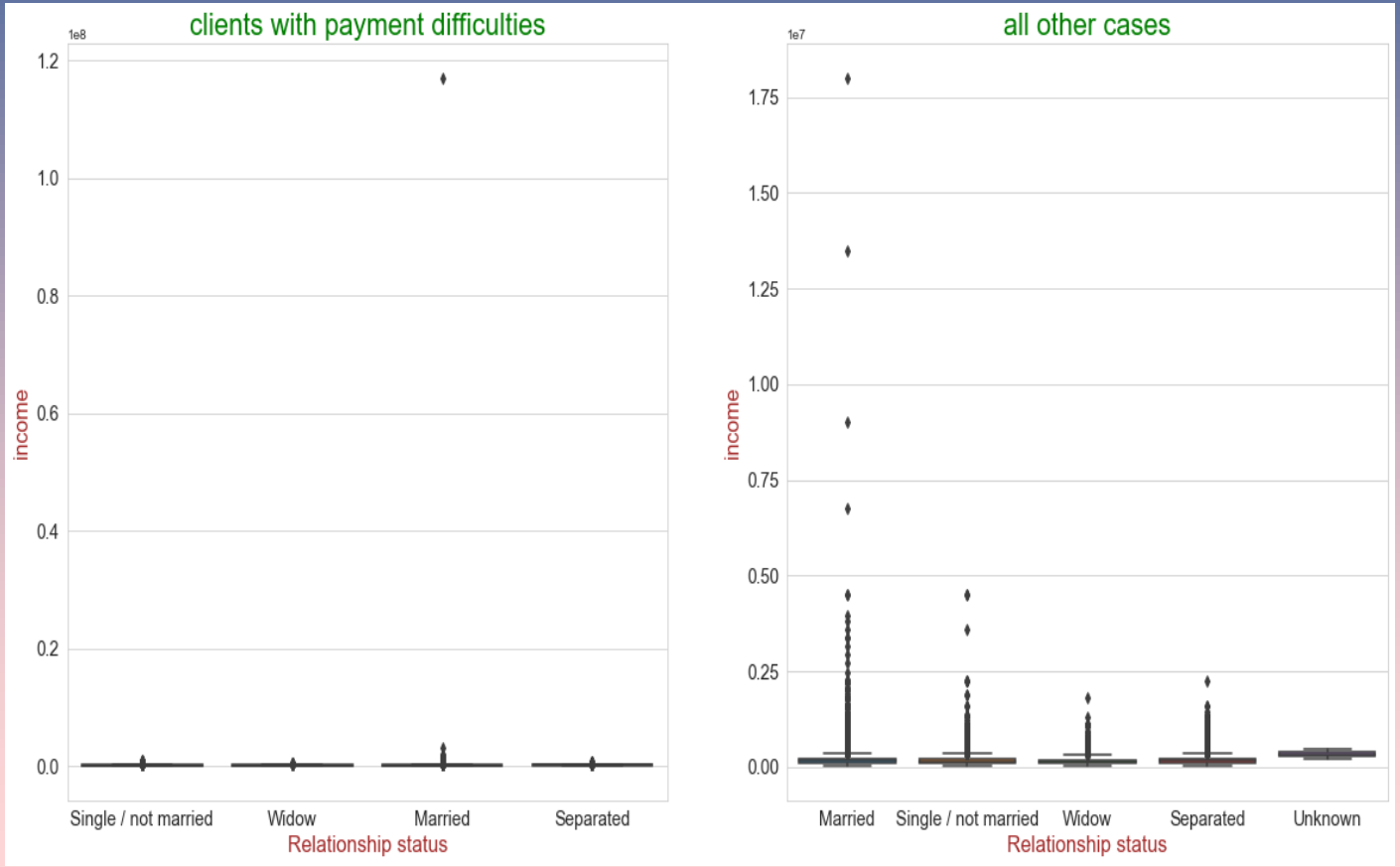
This graph shows imbalance values of target variable with 1s as 8% indicating clients with payment difficults and 92% as 0s indicating all other cases.



EDA Technqiues

- This chart clearly depicts the amount of loan availed by both male and female gender.
- If we carefully observe female gender is availing more loans in both case scenarios ie. Clients with payment difficulties as well as other case.
- Even it is observed that female gender do access more higher amount of loans as well.

Relationship status of Clients



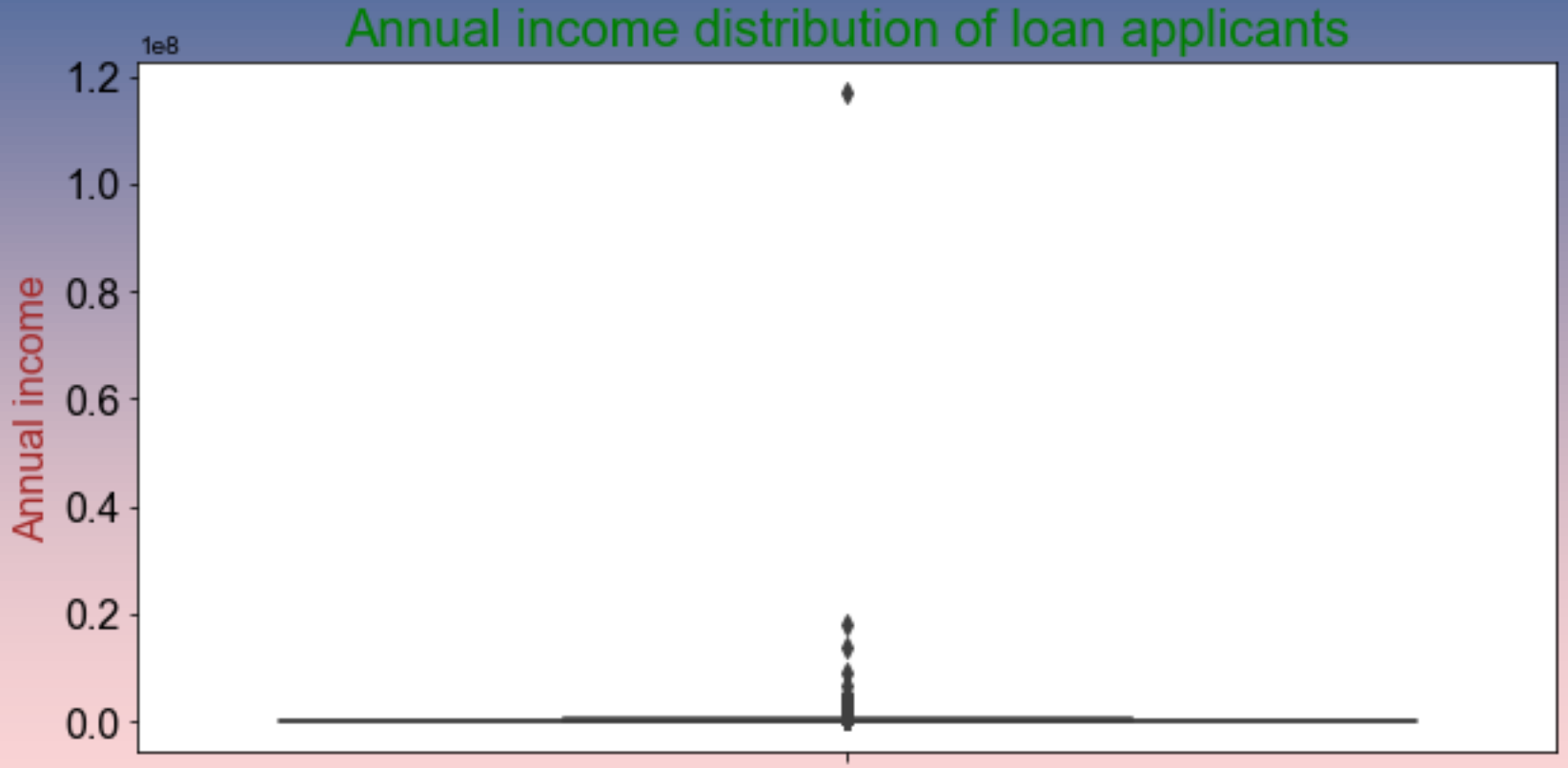
EDA Techniques

- Start by importing the ‘application_data.csv’
- Check the structure of the data (Normal Routine Check)

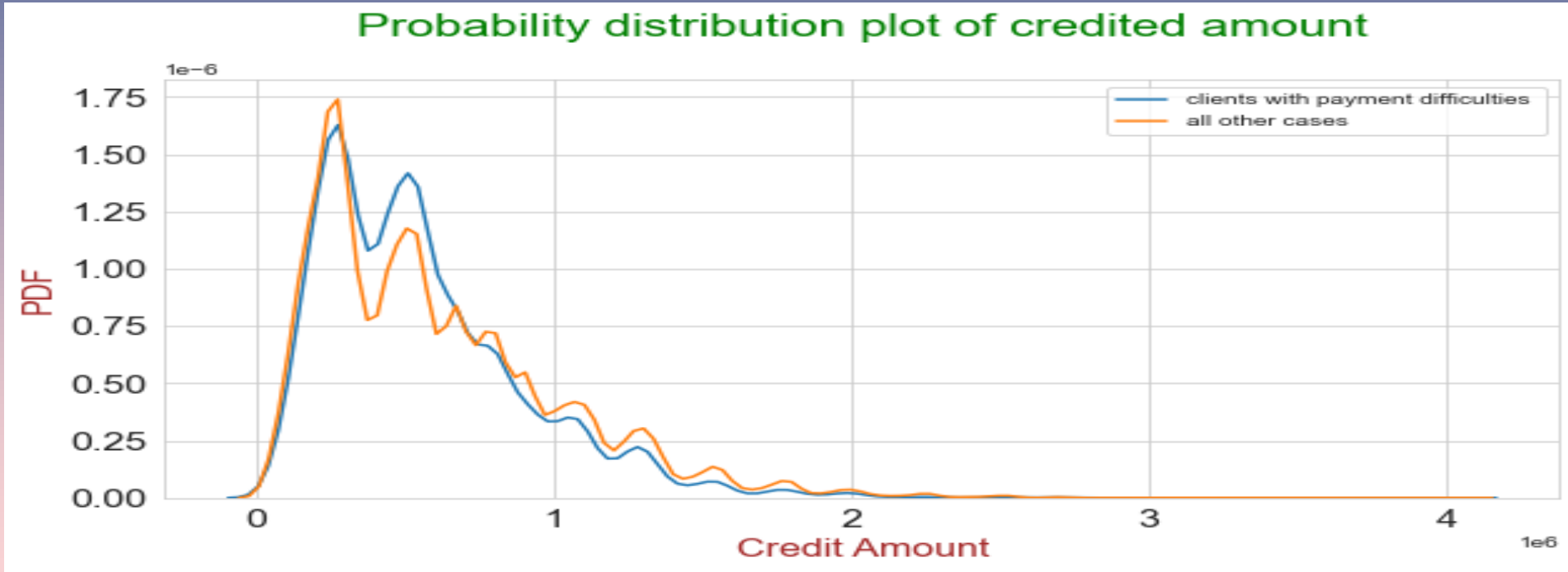
EDA Techniques

- This chart depicts two scenarios where clients who are facing ‘payment difficulty’ and clients who fall under ‘other cases’.
- Here we have categorised the clients by their relationship status and loan amount. The relationship status is divided by single/divorced/married/widowed & unknown.
- We clearly see that in the above mentioned categories, under other cases we see a spike in loan availing under the other cases section.

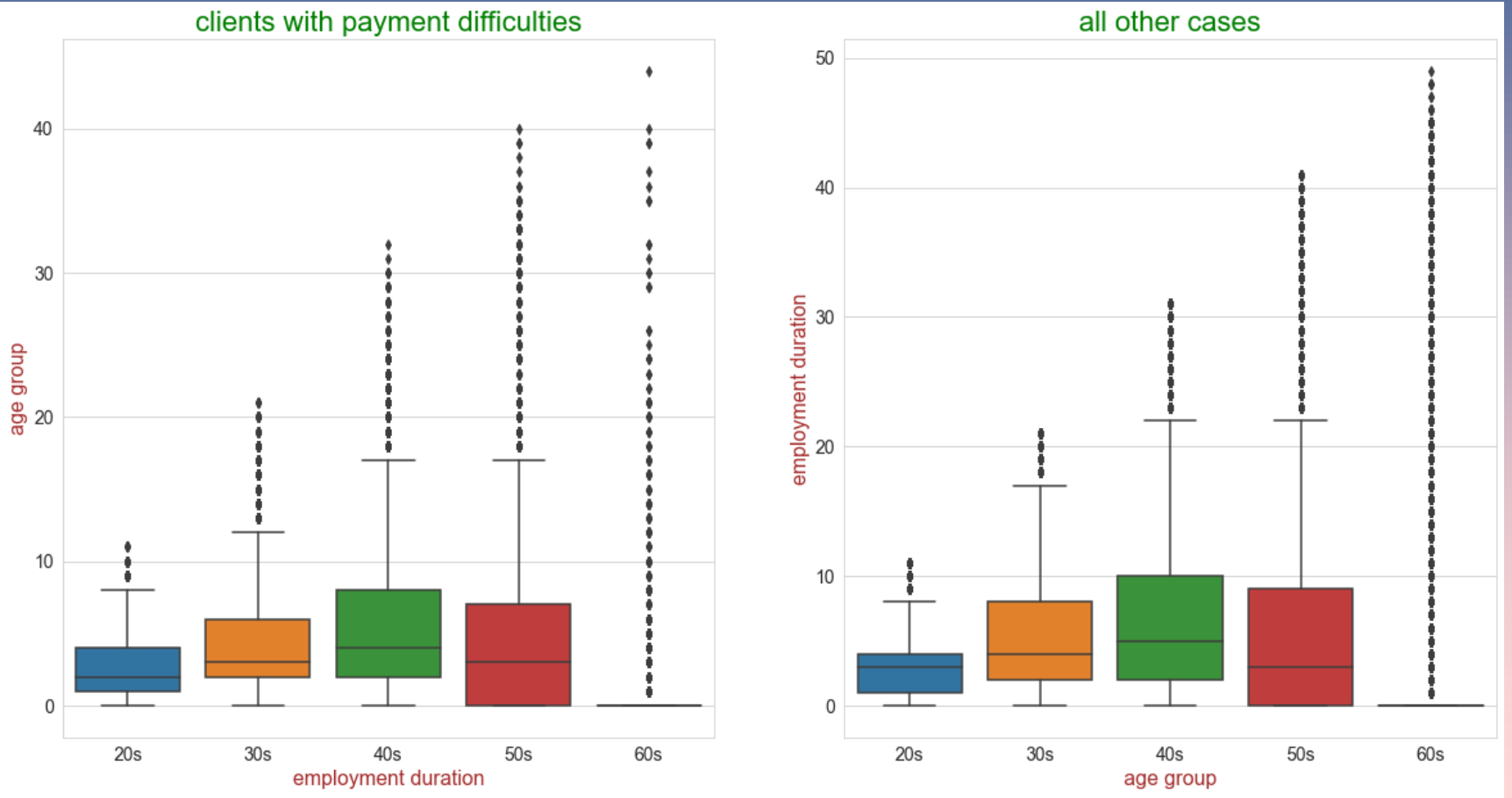
Annual income distribution of loan applicants



EDA Techniques
Probability distribution plot of credited amount

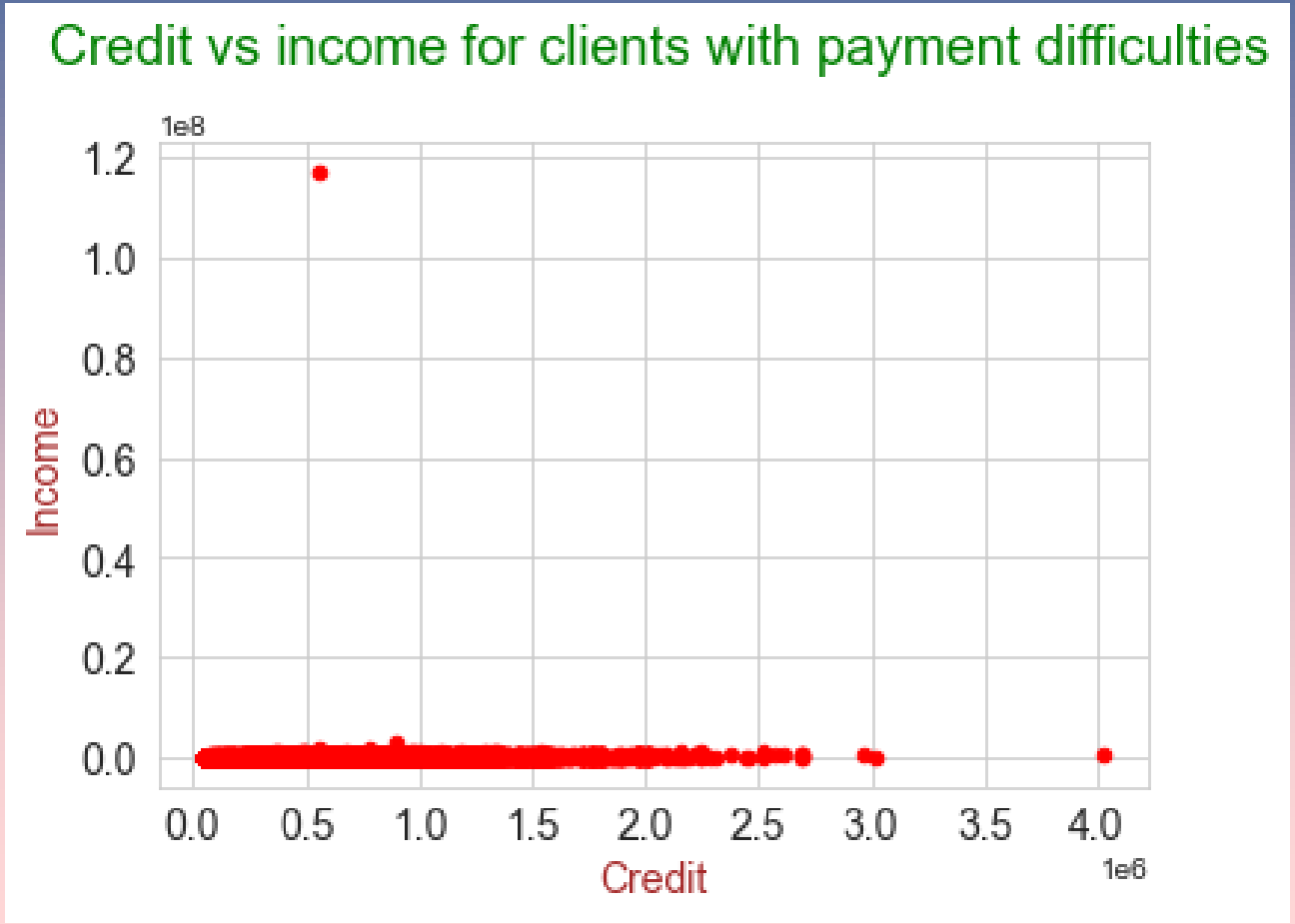


EDA Techniques
clients with payment difficulties based on age

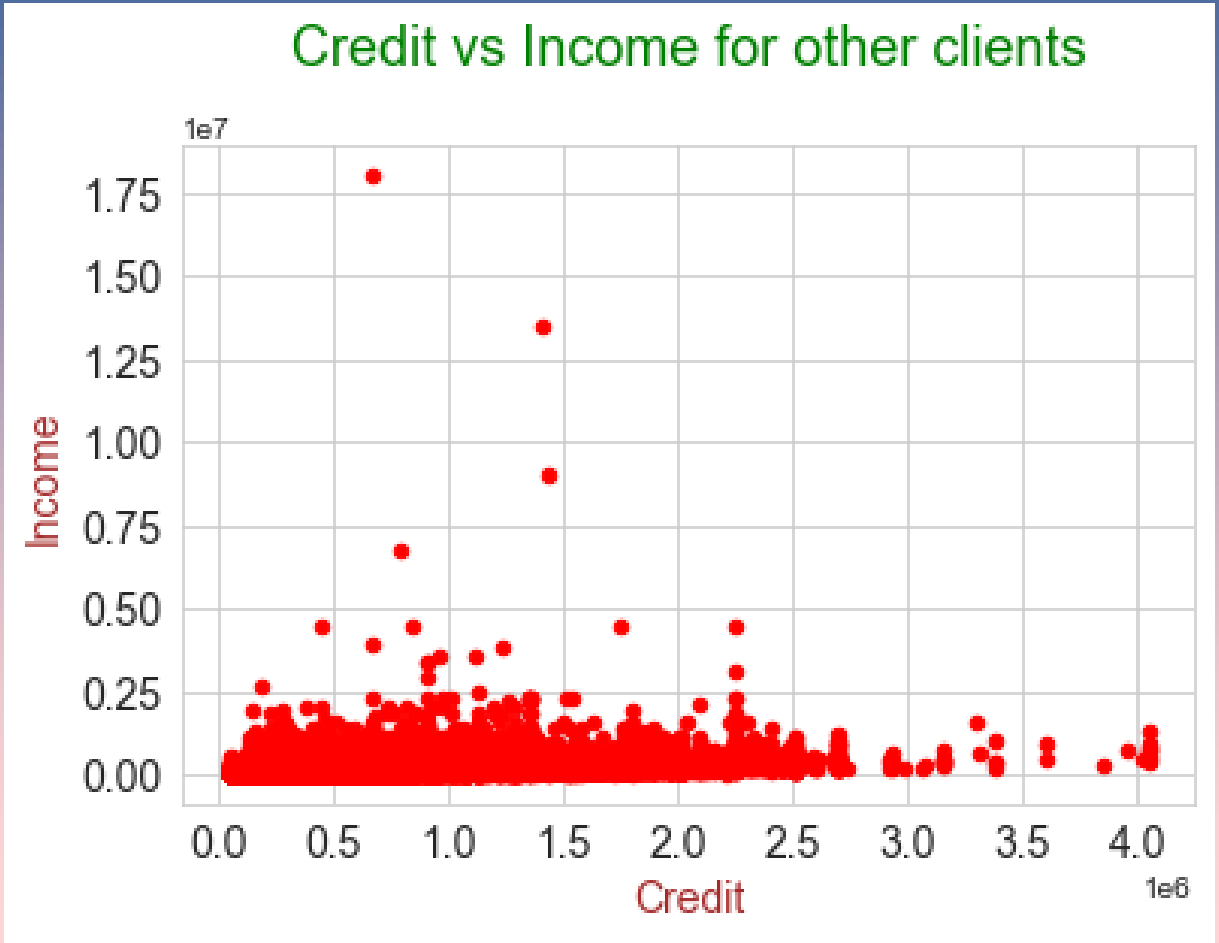


EDA Techniques

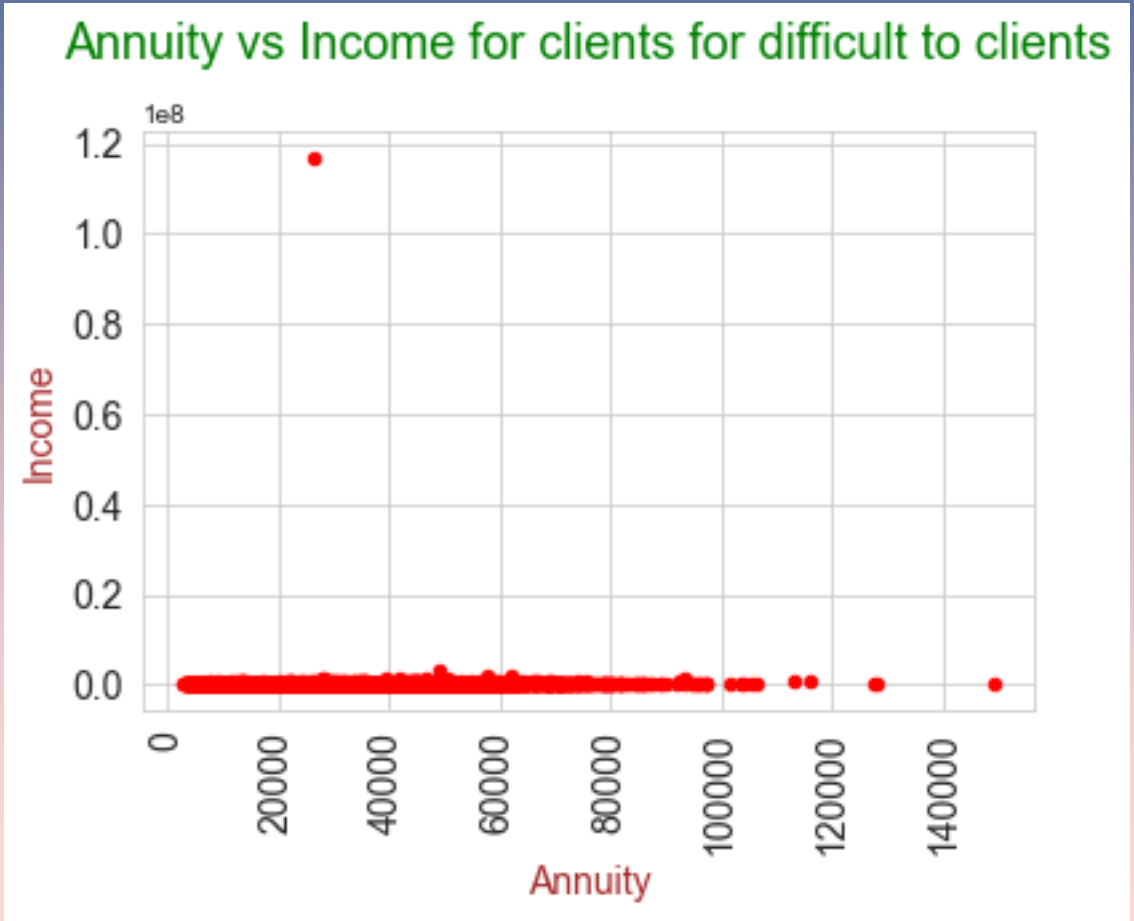
Credit vs income for clients with payment difficulties



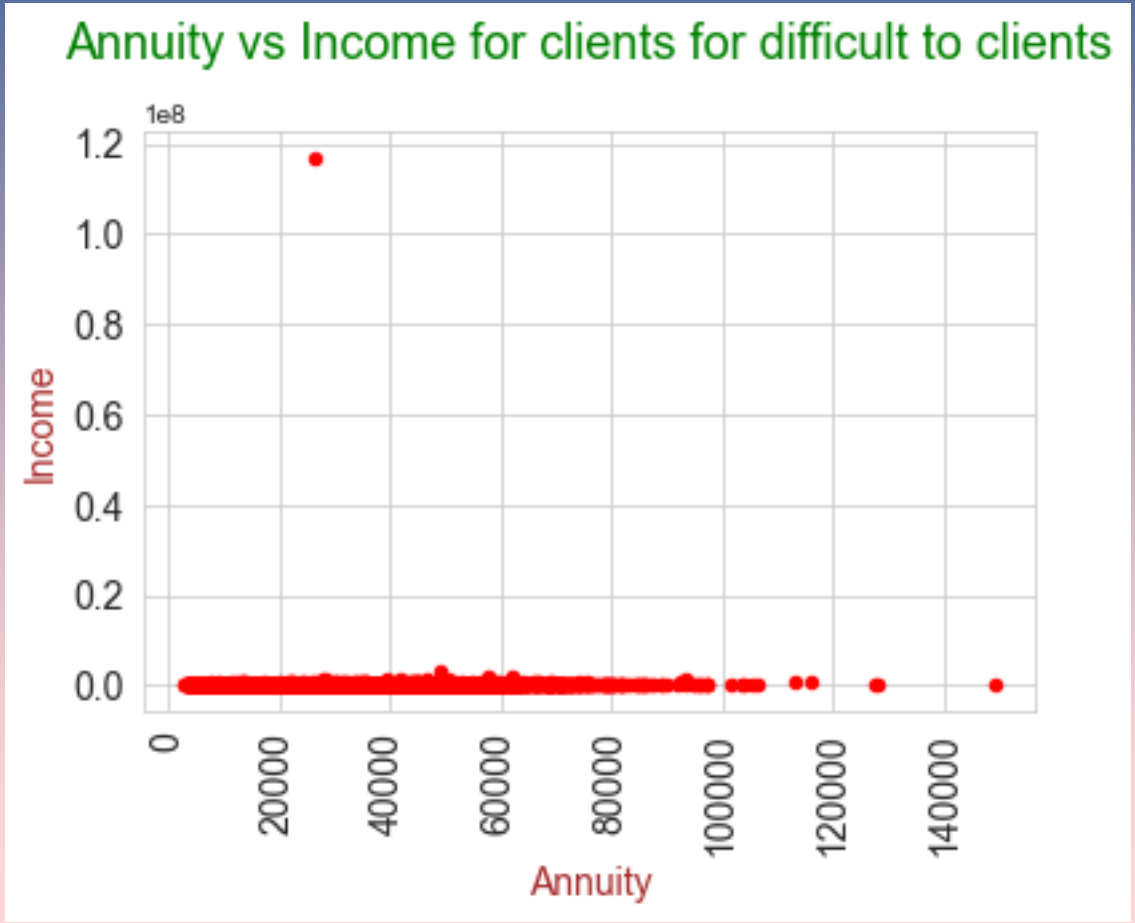
EDQ Techniques
Credit vs Income for other clients



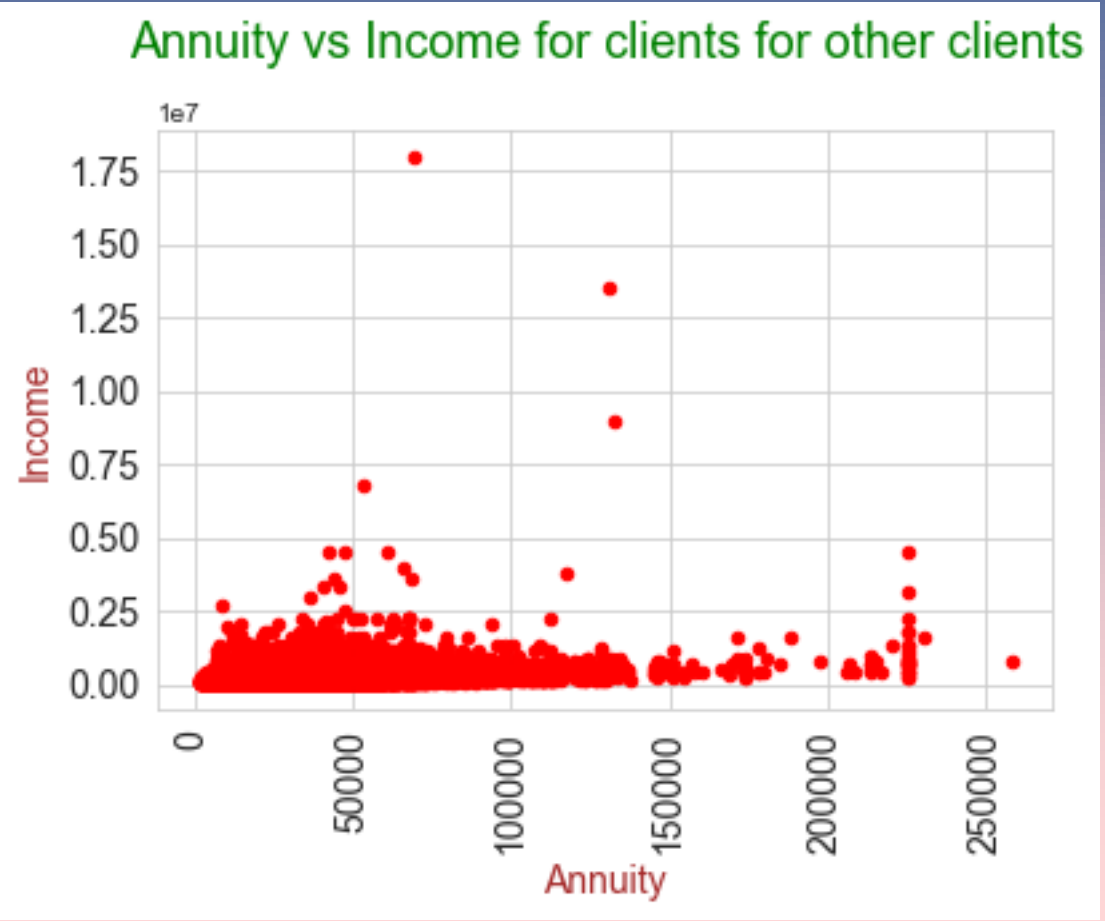
EDA Techniques
Annuity vs Income for clients for difficult to clients



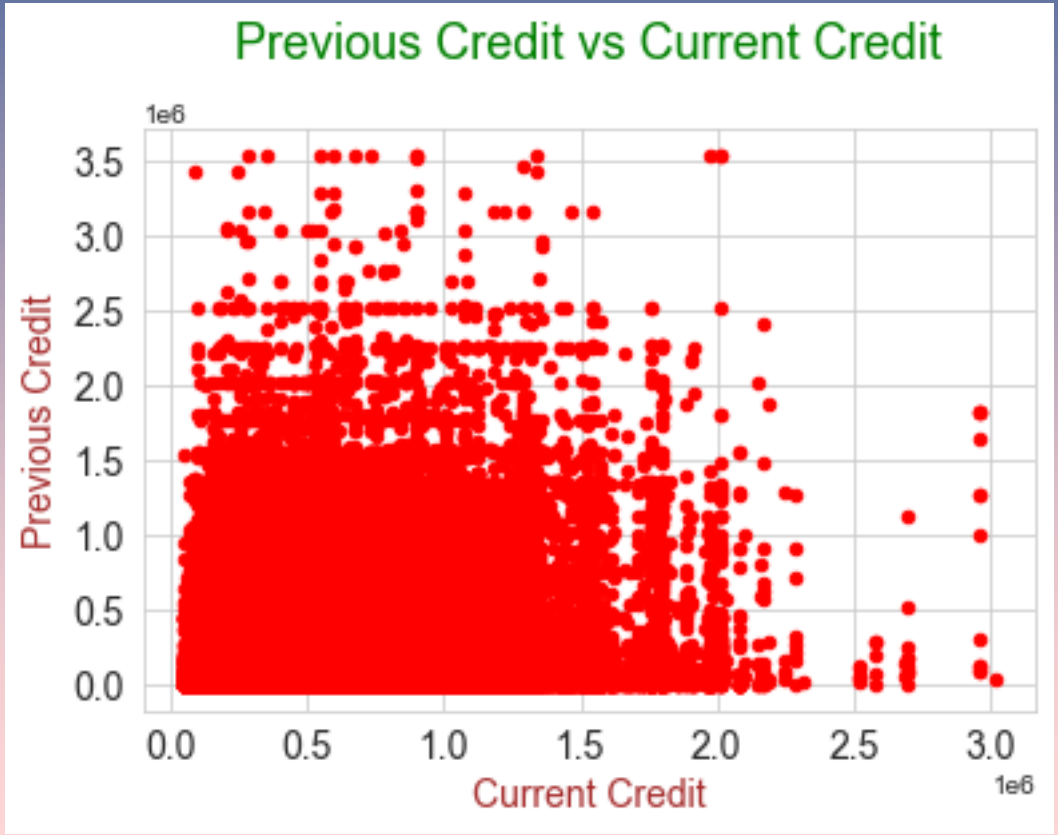
EDA Techniques
Annuity vs Income for clients for difficult to clients



EDA Techniques
Annuity vs Income for clients for other clients

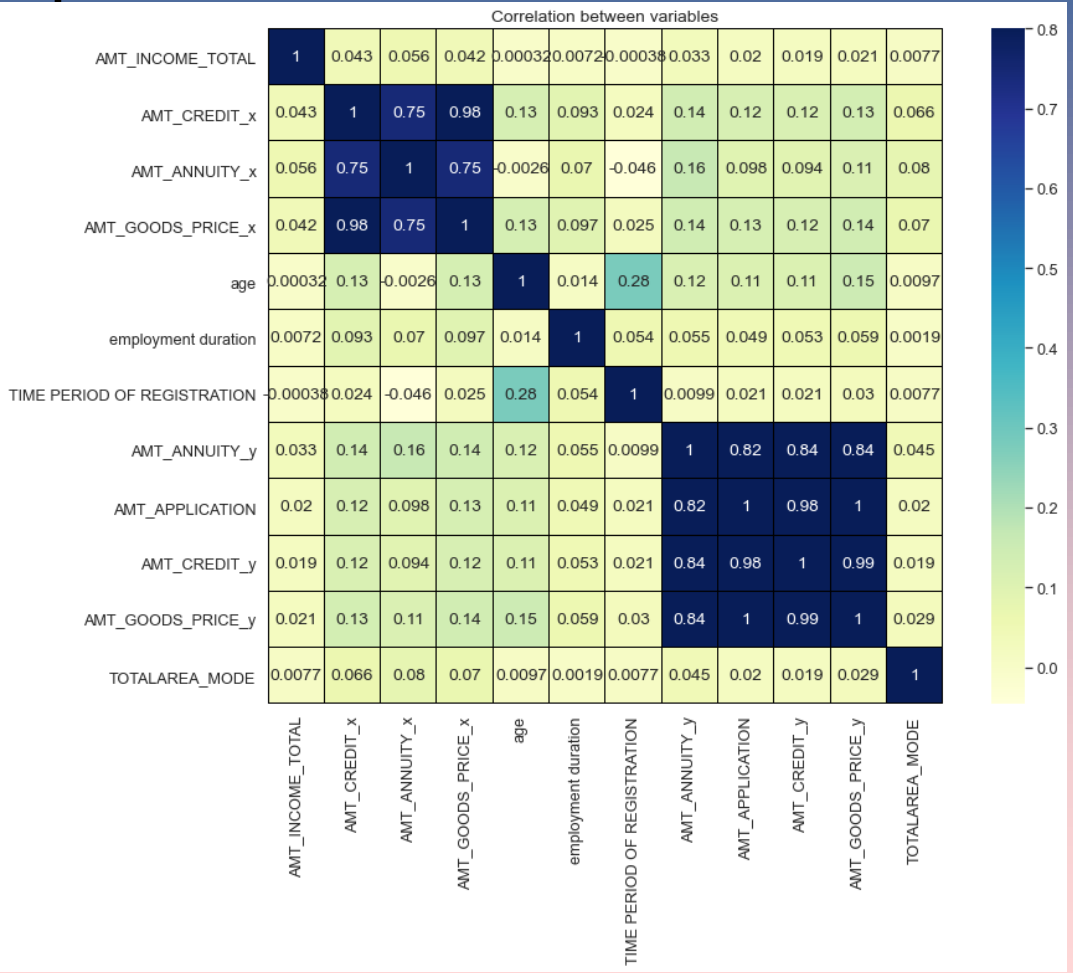


EDA Techniques
Previous Credit vs Current Credit

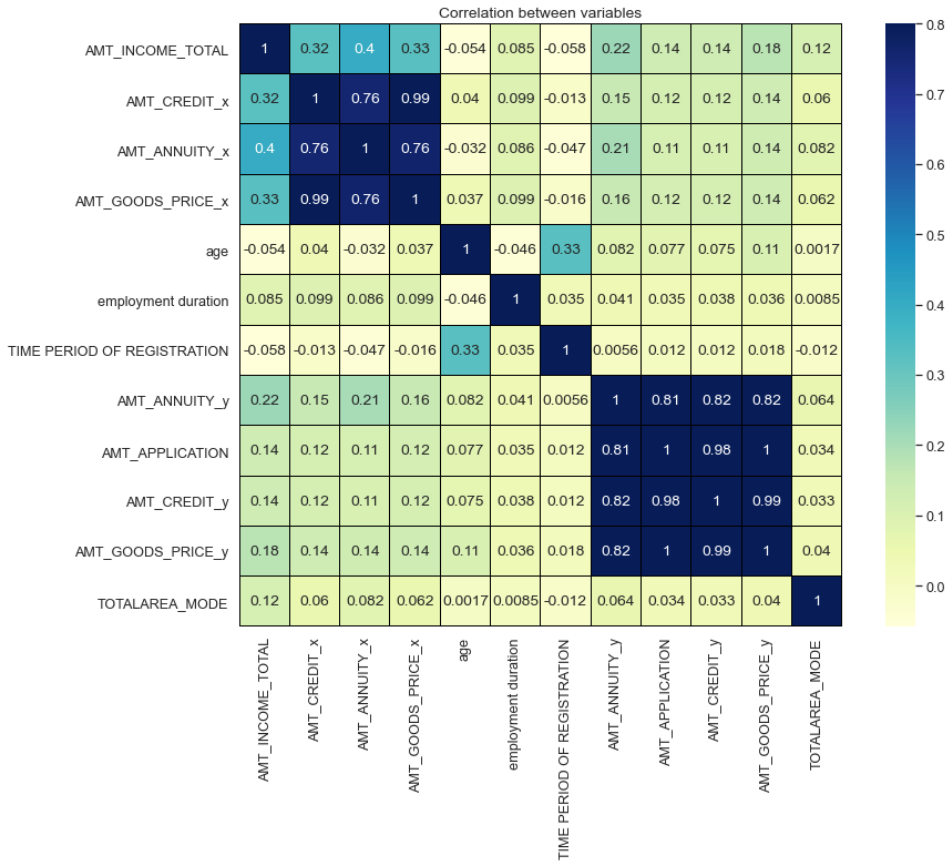


EDA Techniques

plotting heatmap for correlation matrix for clients difficult to pay



EDA Techniques
plotting heatmap for correlation matrix for other cases



Thank You