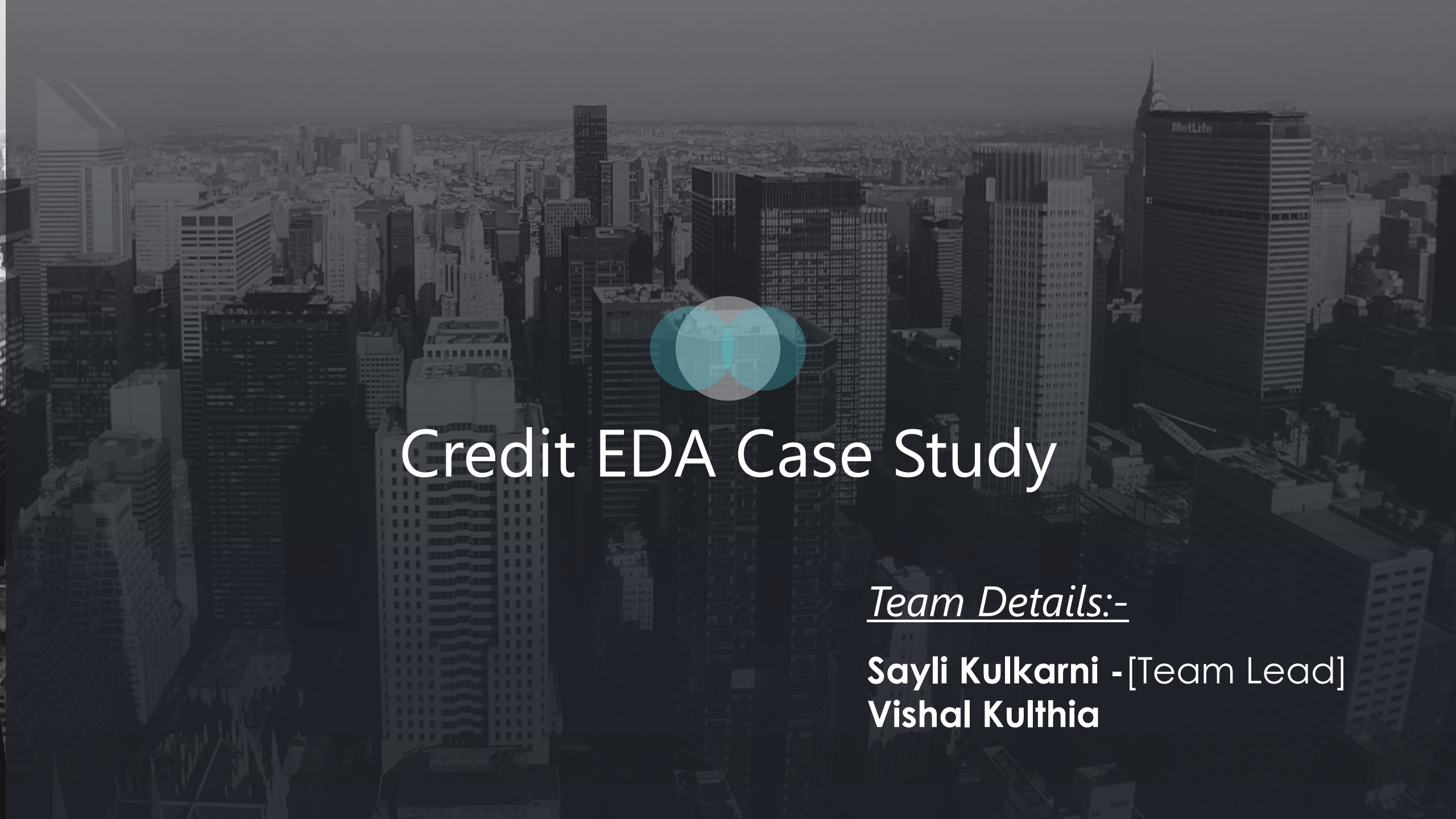


Credit EDA Case Study

Team Details:-

Sayli Kulkarni -[Team Lead]
Vishal Kulthia

Approach Followed – Application data

- Business Problem Description
- Approach Flow diagram
- Imbalance of Target Variable
- Univariant Analysis
- Bivariant Analysis
- Correlation Analysis

Business Problem Description

- The loan providing companies finds it challenging to issue loan to their customers due to insufficient financial history & details about their customers. This results in customer being unable to repay the loan on time and gets tagged as defaulter which directly causes loss to the business.

- Our objective is to analyse the customer data provided and apply data manipulation and Exploratory data analysis concepts (EDA) to come up with a approach/framework which identifies the key driving factors behind loan defaults and analyse the trend in the data to identify customers that will be able to repay the loan within the defined duration.

- Going forward the Business will then use this analysis and key driving factors to decide whether or not a customer is eligible for the loan as part of risk assessment activity.

Approach Followed

- Business Problem Description
- Approach Flow diagram
- Imbalance of Target Variable
- Univariant Analysis
- Bivariant Analysis
- Correlation Analysis

Approach Flow Diagram

Develop Business Understanding

- Analyse the problem statement and final objective to be achieved
- Go through Data dictionary to develop pellucid understanding about significance of each attribute and business importance
- Examine the data and type of values etc.

Data Analysis

- Load the data
- Examine the metadata/structure of the data
- Data Cleansing, Missing Value Check & fix
- Outliers handling

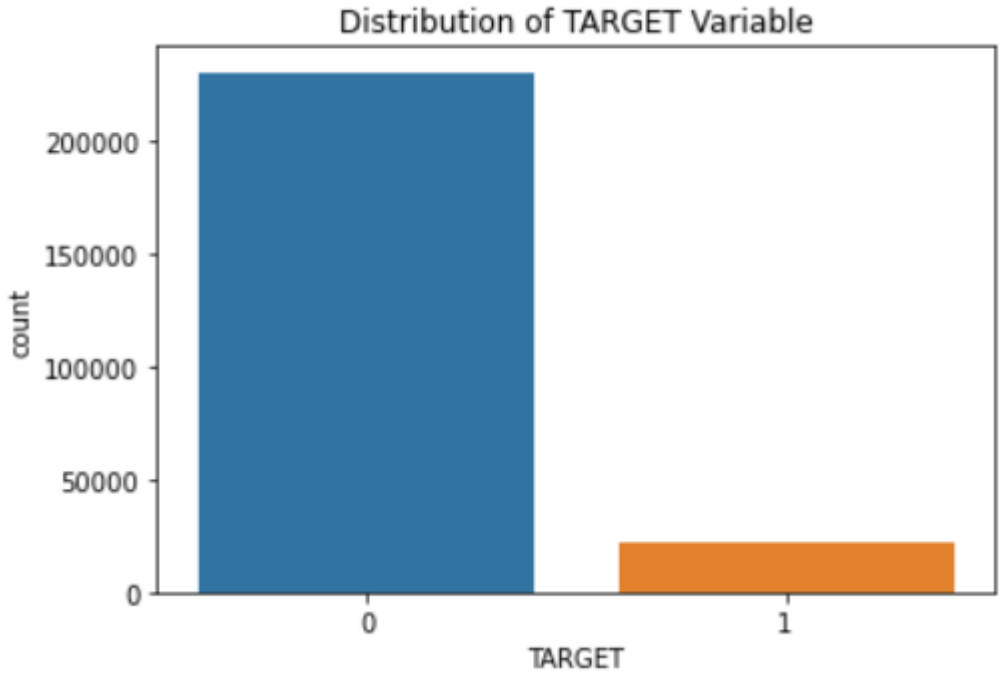
Exploratory Data Analysis

- Determine imbalance percentage in key driving attributes
- Top 10 Attribute Correlation analysis using heat maps
- Bivariant and univariant analysis

Approach Followed

- Business Problem Description
- Approach Flow diagram
- Imbalance of Target Variable
- Univariant Analysis
- Bivariant Analysis
- Correlation Analysis

Imbalance of target variable



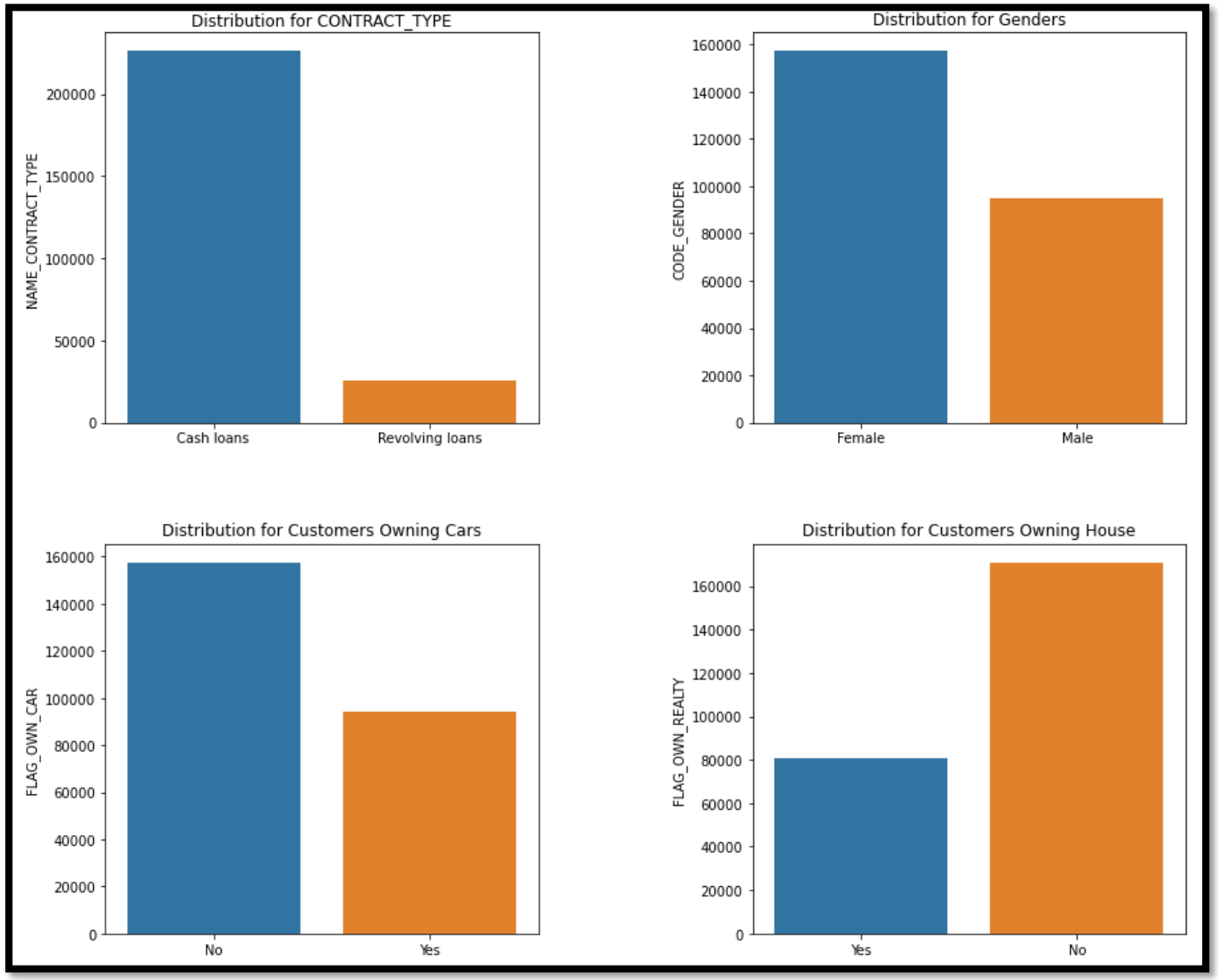
Conclusion Drawn: -

1. Percentage of customer being defaulter(unable to repay loan) is greatly imbalanced as compare to not being defaulter
2.The gap between the percentage of target values are significant

Approach Followed

- Business Problem Description
- Approach Flow diagram
- Imbalance of Target Variable
- Univariant Analysis
- Bivariant Analysis
- Correlation Analysis

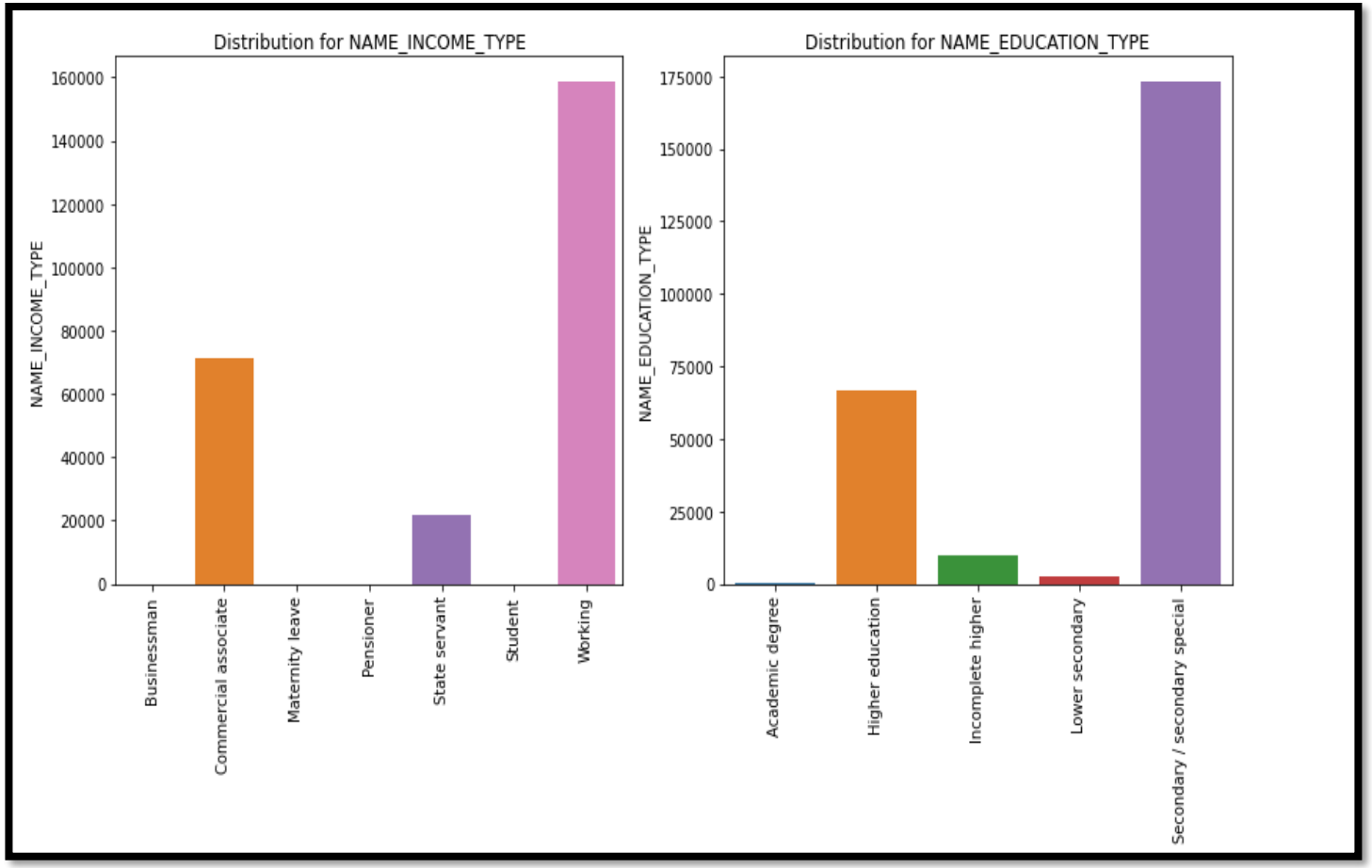
Univariant Analysis for categorical variables



Conclusion Drawn: -

1. Huge difference exist between categories of contract type i.e. Cash & Revolving Loans
2. Females customers are almost double in numbers as compared to male customers
3. Customers who doesn't own a car are more in number as compared to those owning a case
4. Customers who doesn't own a house are more in number as compared to those owning a house

Distribution of clients income type and highest education

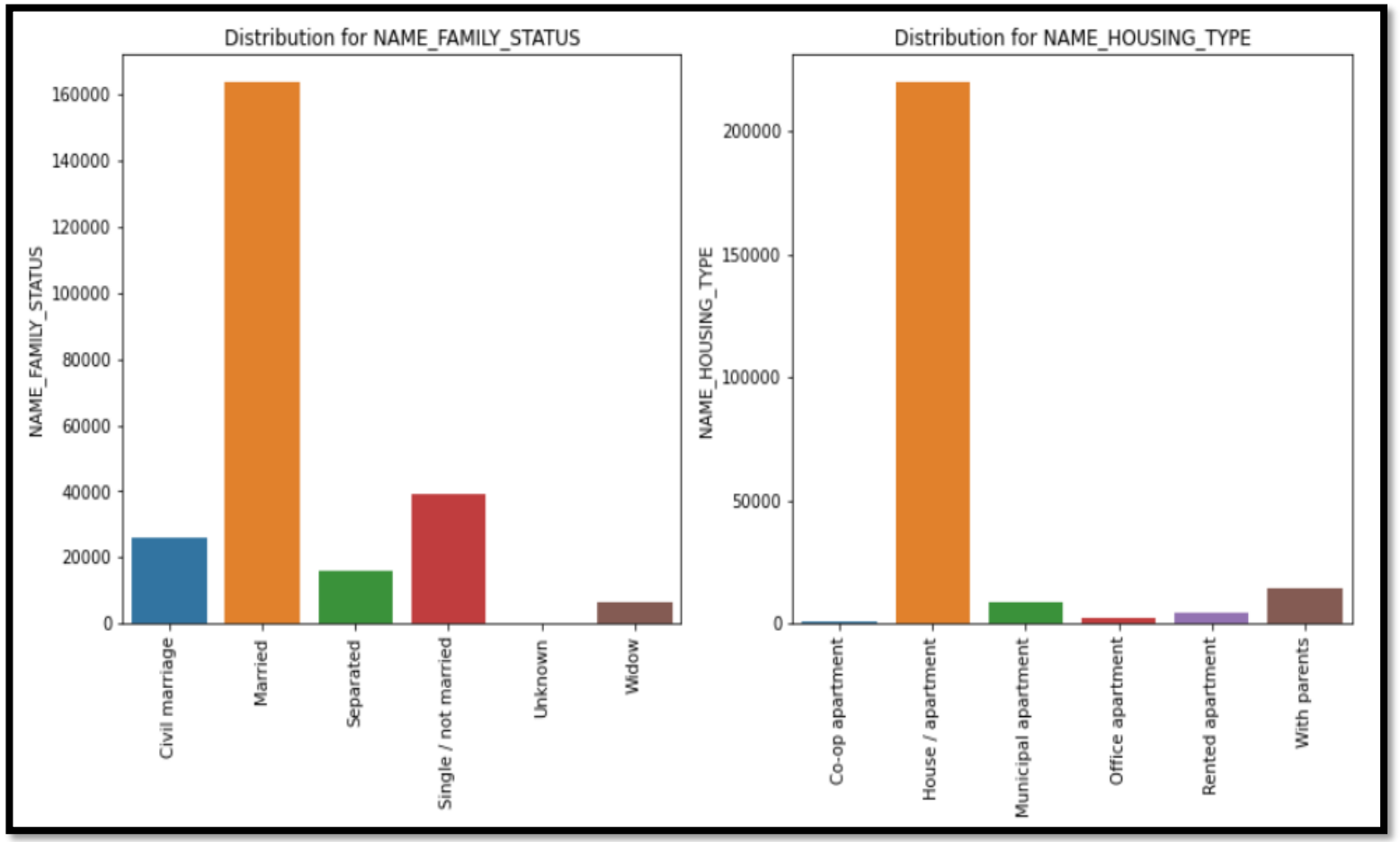


Conclusion Drawn: -

1. It has been observed that people having income type as student, pensioner etc. are extremely low as compared to other categories like working, state servant etc.

2. Majority of the customers existing have either secondary education followed by higher education

Distribution for plot for Family Status and Housing Type

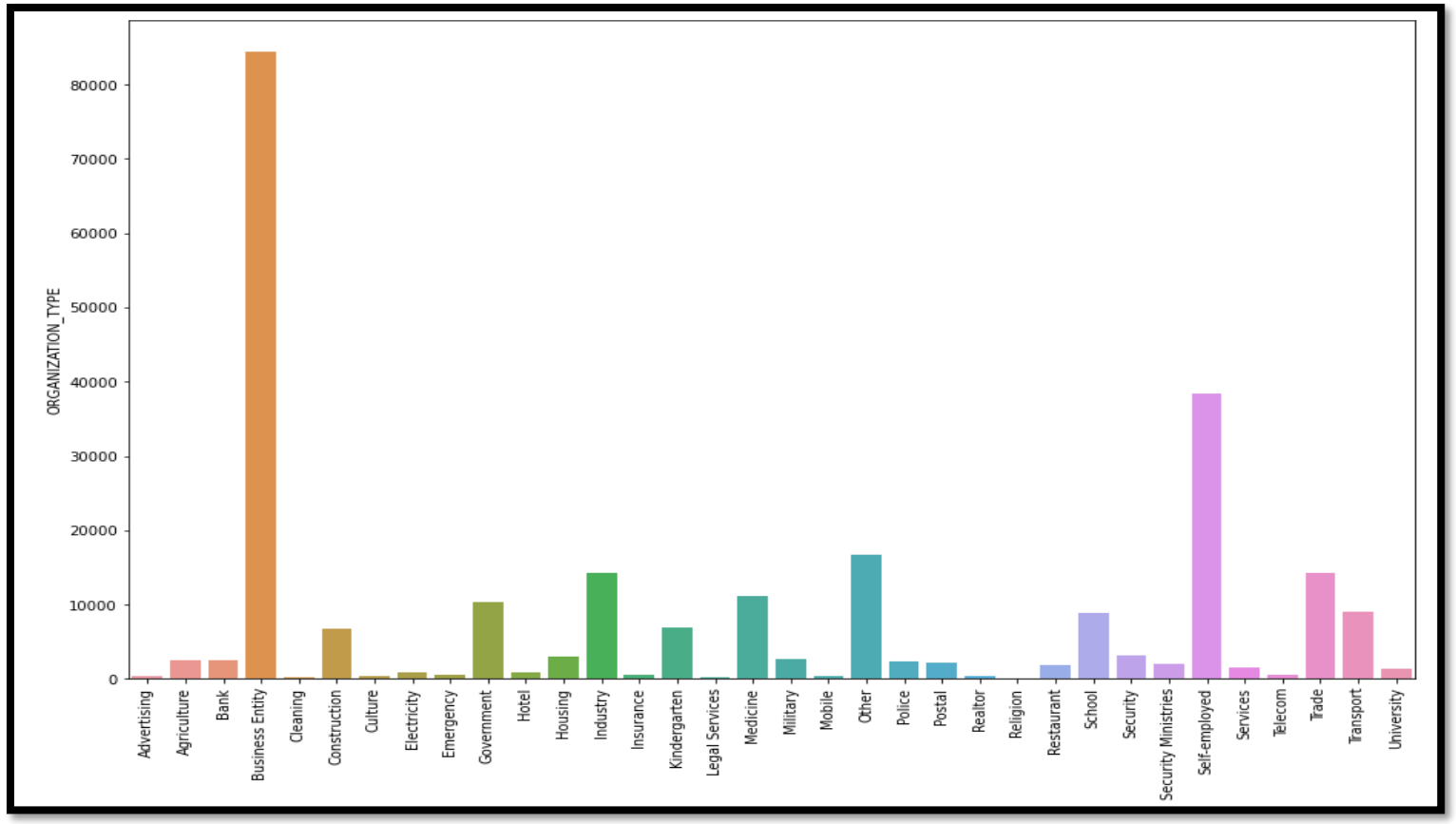


Conclusion Drawn: -

1. Majority of the customers existing are either married and or single in family status category/group

2. Analysis shows that majority proportionate of customers actually live in houses/apartments which in a way indicates the financial condition/stability of customers

Distribution plot of Customers Organization Type



Conclusion Drawn: -

1. It is observed from the distribution plot that most of the customers have business entities as their organization type followed by self employed

2. The lowest strength of the customer organization type are namely advertising, cleaning, culture, religion etc.

Approach Followed

- Business Problem Description
- Approach Flow diagram
- Imbalance of Target Variable
- Univariant Analysis
- Bivariant Analysis
- Correlation Analysis

Bivariant Analysis

We will perform three types of Bivariate analysis

- Bivariate Analysis of Numerical - Numerical Variables
- Bivariate Analysis of Numerical- Categorical Variables
- Bivariate Analysis of Categorical- Categorical Variables

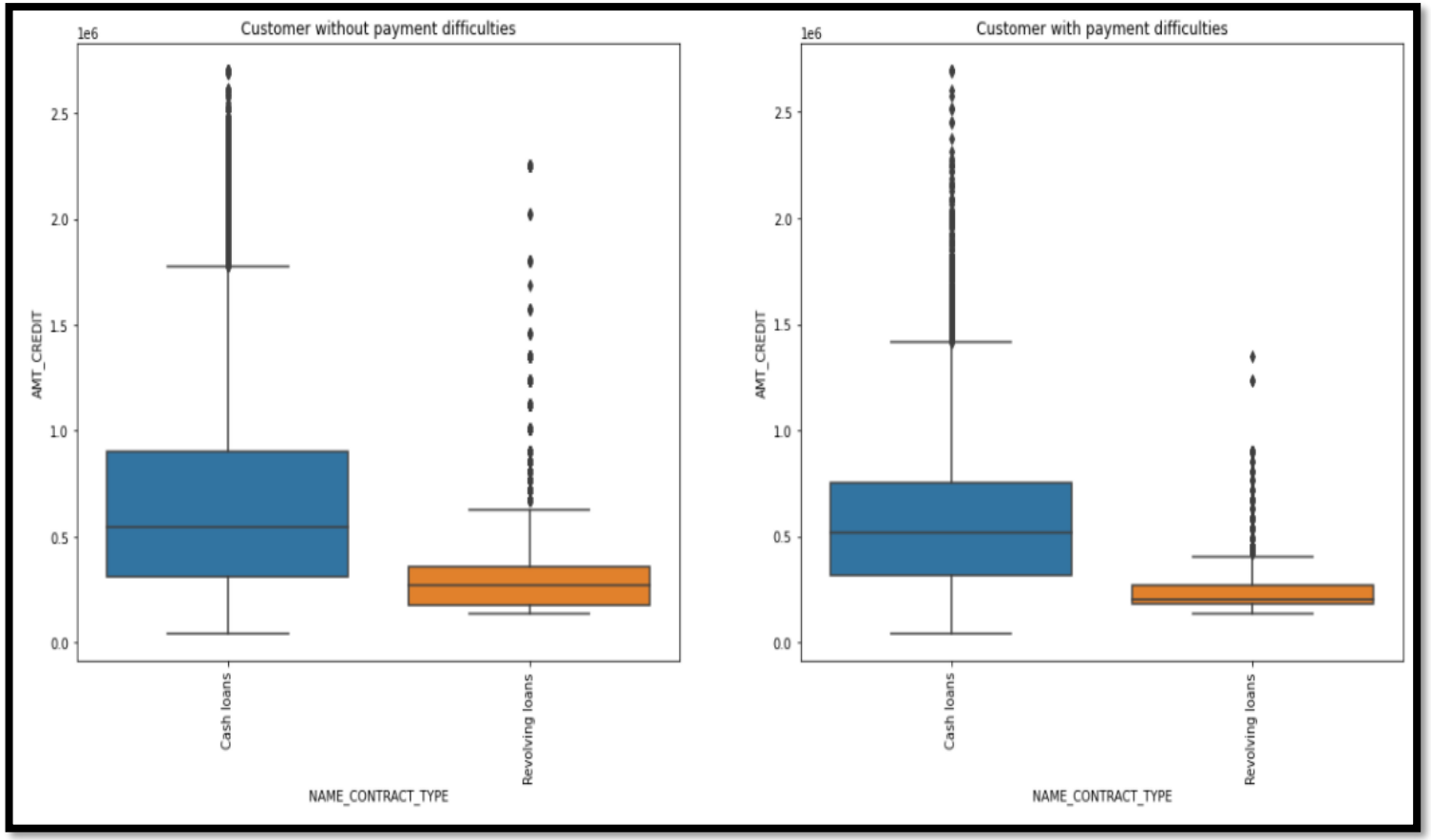
To further simplify the analysis, two data frames with the following names have been created –

1. *Client without payment difficulties : Target value =0*

2. *Client with payment difficulties : Target value =1*

Bivariate Analysis on above two mentioned data frames have been carried out separately.

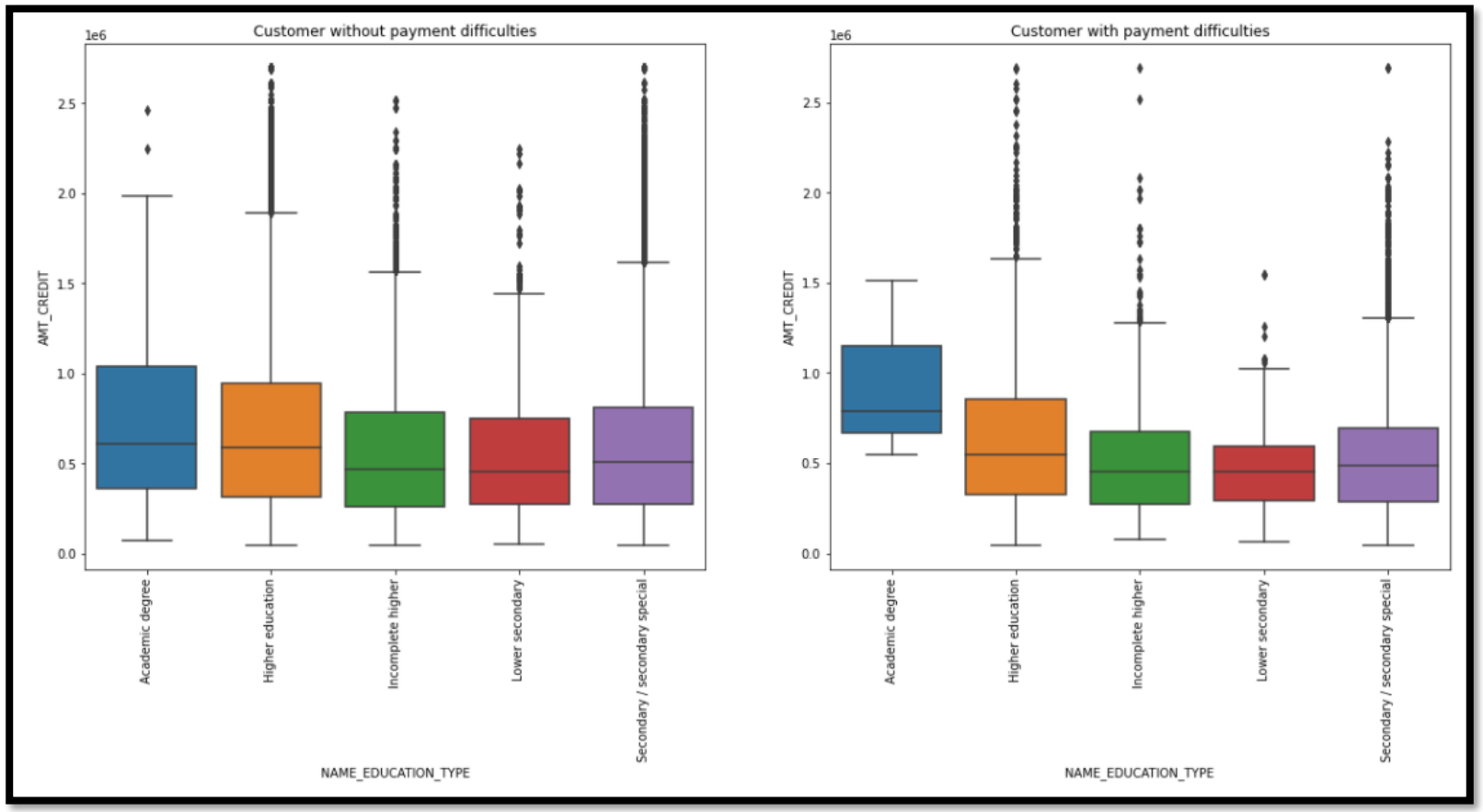
Distributions for Credit amount in last application vs Identification if loan is cash or revolving



Conclusion Drawn: -

1. Median for Customer without payment difficulties and Customer with payment difficulties is almost same
2. Huge difference exist between 75th quartile of Customer without payment difficulties and those with payment difficulties
3. There are some outliers present in Customer with payment difficulties

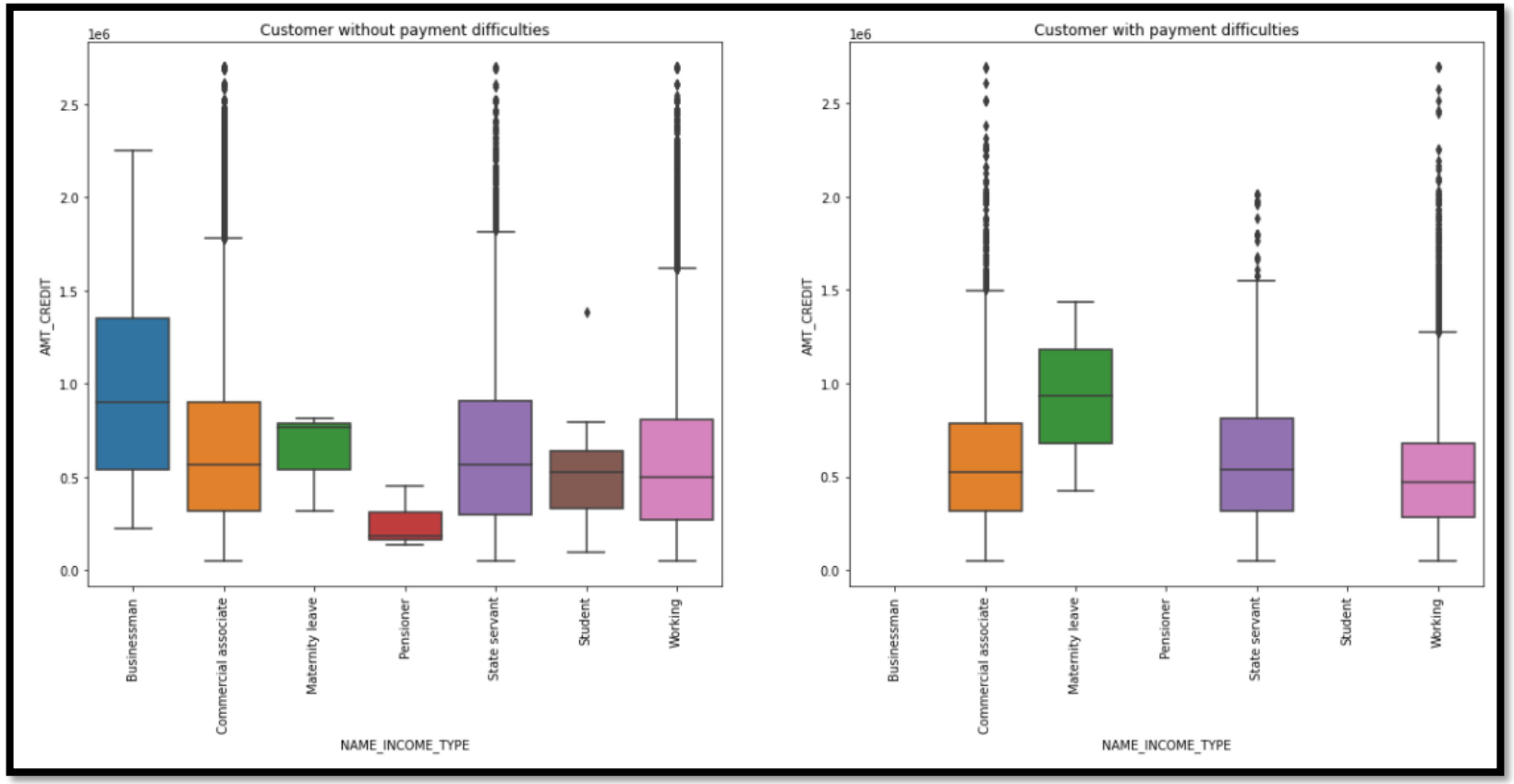
Distributions for Credit amount of the loan vs Level of highest education of client



Conclusion Drawn: -

1. Median for Academic degree is highest and Incomplete higher is lowest for both Customer with and without payment difficulties
2. For Customer without payment difficulties , median is approximately same for each education type
3. Minimal outliers present in Education type Academic degree for both Customer with & without payment difficulties

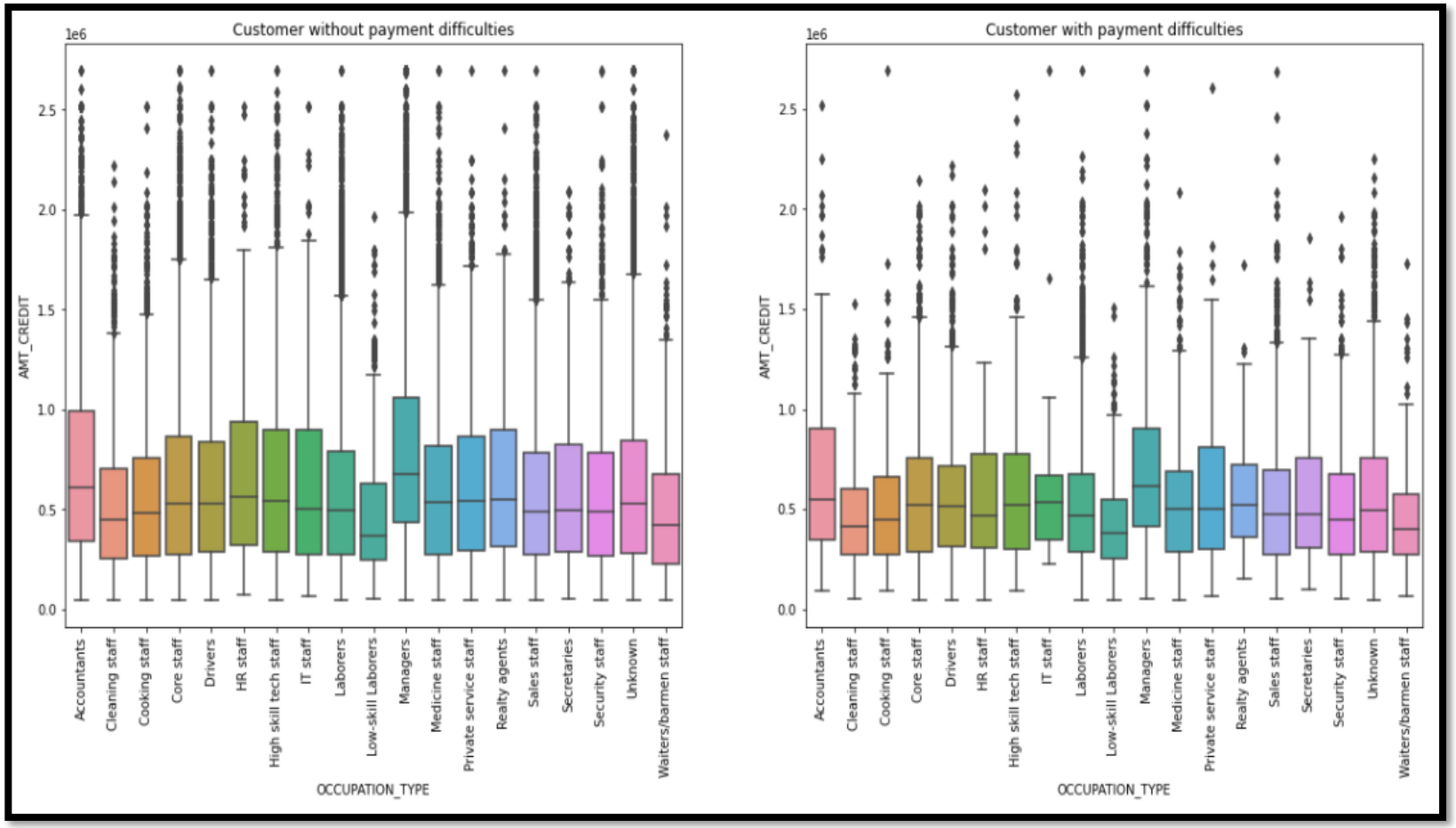
Distributions for Final credit amount on the previous application vs Clients income type



Conclusion Drawn: -

1. Businessman, Pensioner and Students are not facing any payment difficulties
2. Number of outliers are more for Customer without payment difficulties as compared to Customer with payment difficulties
3. Median is highest for Businessman & Maternity leave for Customer without & with payment difficulties respectively

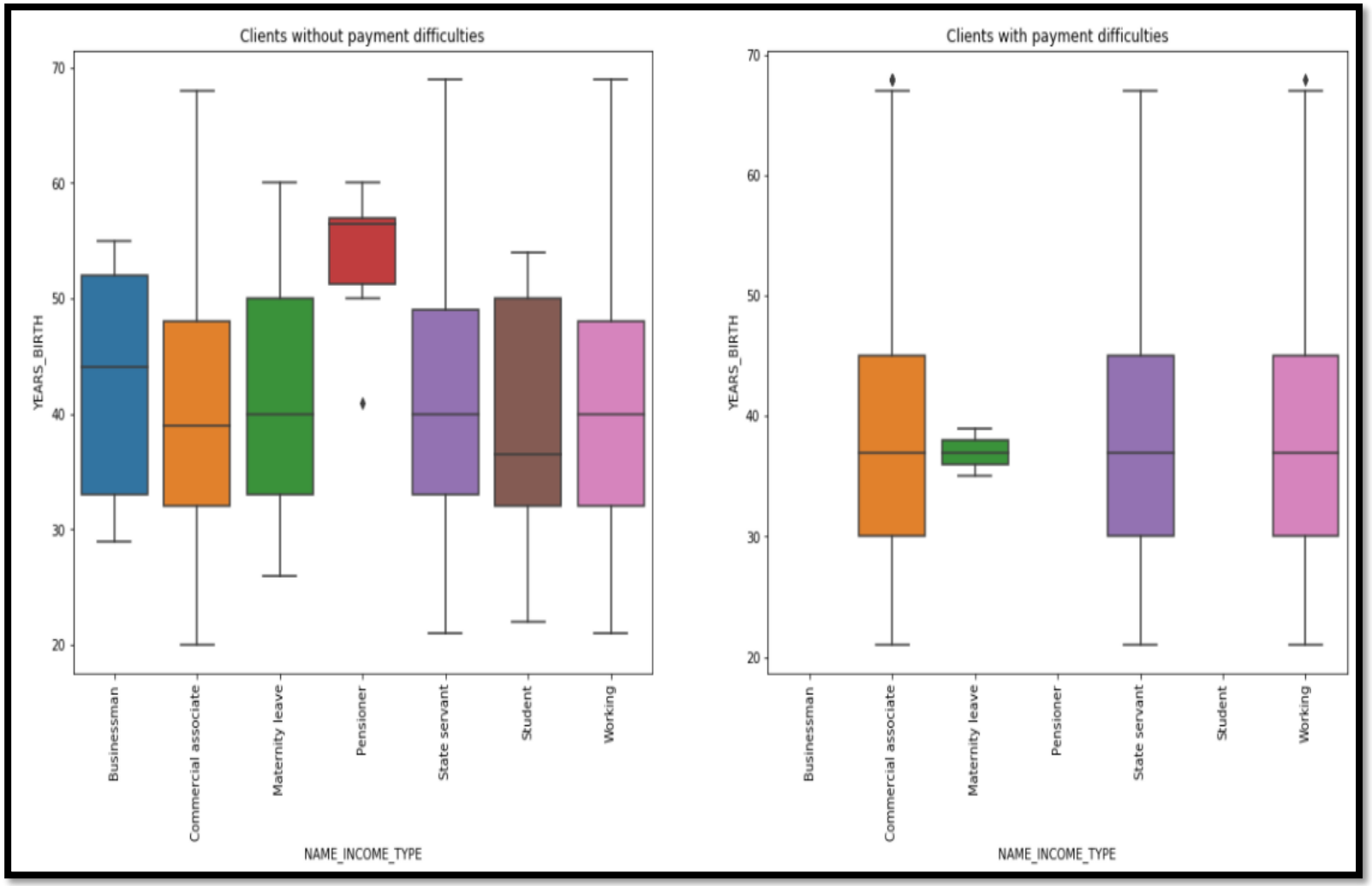
Distributions of final credit amount on the previous application vs Occupation type



Conclusion Drawn: -

1. Range of clients is more for Customer without payment difficulties as compared to Customer with payment difficulties
2. Number of outliers is more for Customer without payment difficulties as compared to Customer with payment difficulties
3. Median of occupation type Managers is highest for both Customer without & with payment difficulties

Distributions of Year Birth vs Income Type

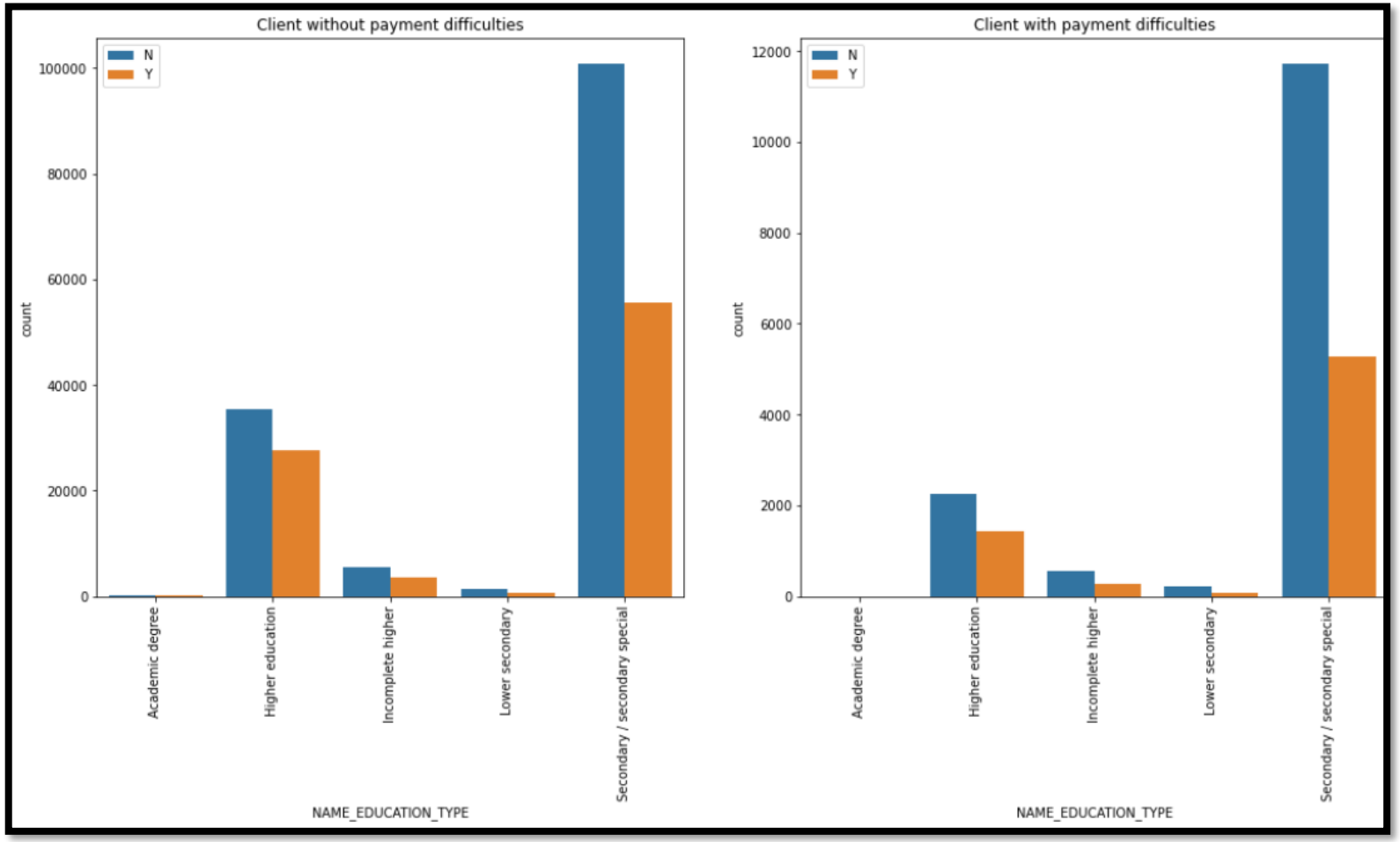


Conclusion Drawn: -

1. Clients with Income type category as Businessman, Pensioner and student doesn't have difficulties in repaying the loan to firm

2. Median of all available category for 'Clients with payment difficulties' is same

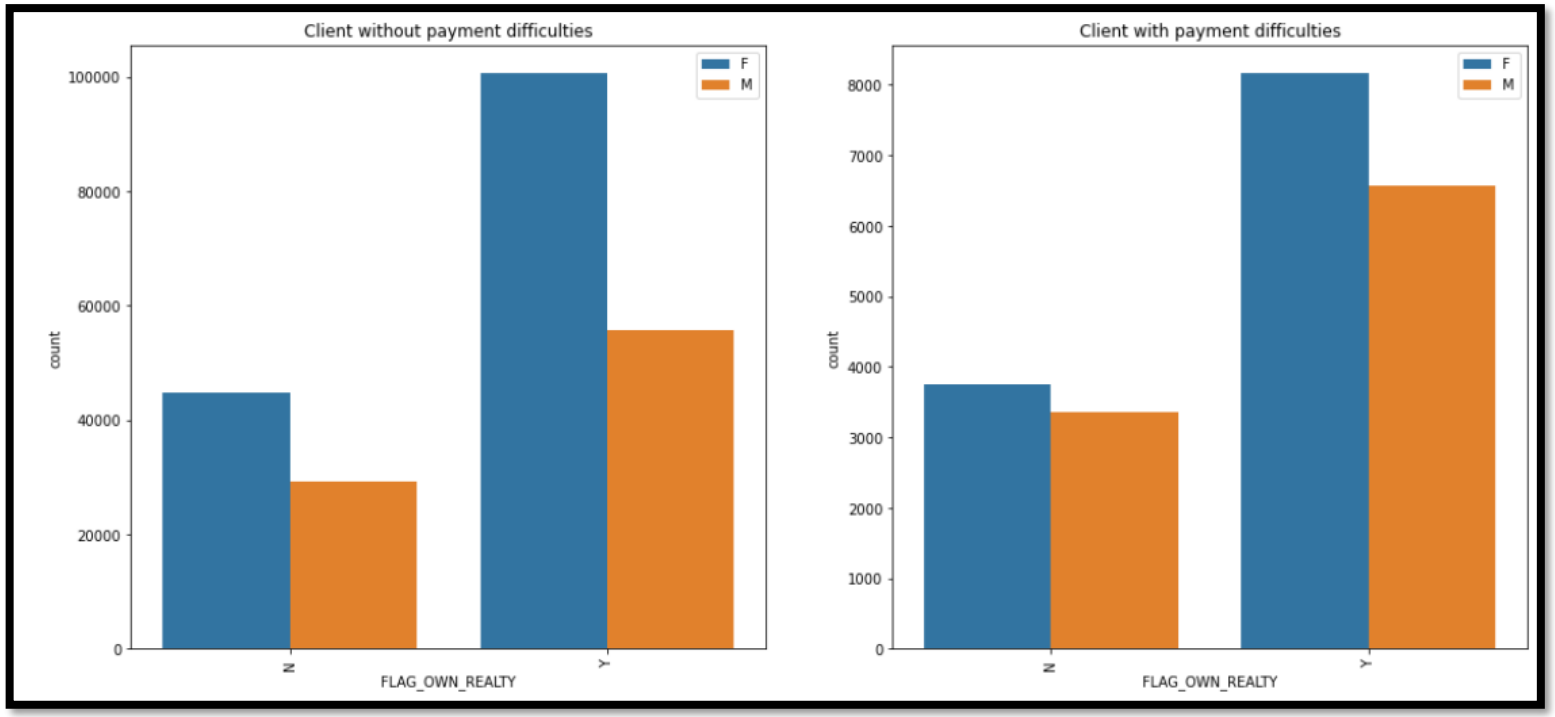
Distributions for Income of the client vs Education Type



Conclusion Drawn: -

1. Clients with Education type academic degree have not mentioned anything related to own a car
2. Clients having education type Secondary/secondary special have heighest number of cars
3. Range of clients is more for Customer without payment difficulties as compared to Customer with payment difficulties

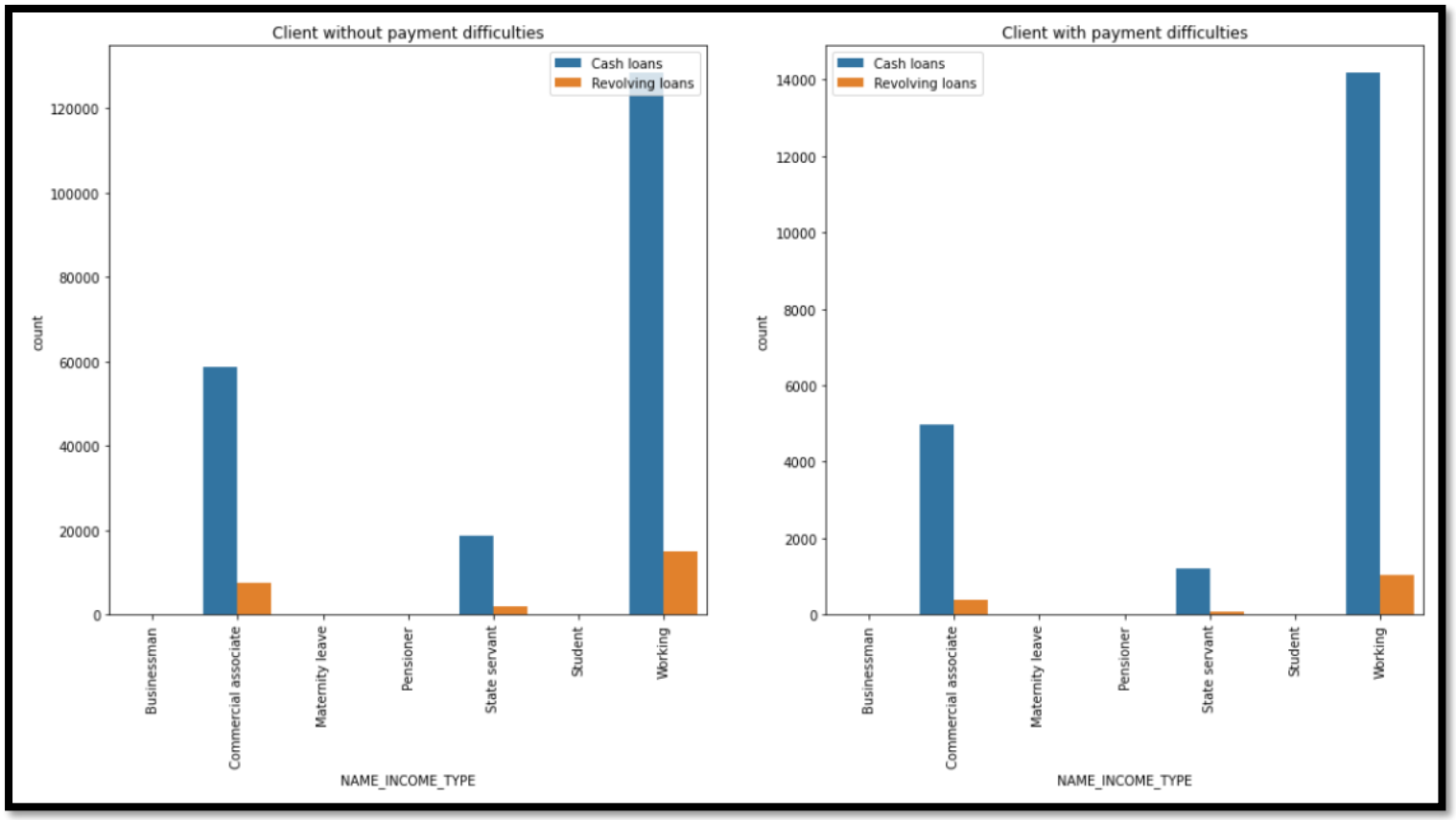
Distributions for Flag if client owns a house/flat vs Gender Code



Conclusion Drawn: -

1. Number of female clients is more who owns reality in both Customer without payment difficulties and Customer with payment difficulties
2. There is not huge difference between female and male clients having reality and not having reality in Customer with payment difficulties as compared to Customer without payment difficulties

Distributions for Clients Income type vs Contract product type of previous application data



Conclusion Drawn: -

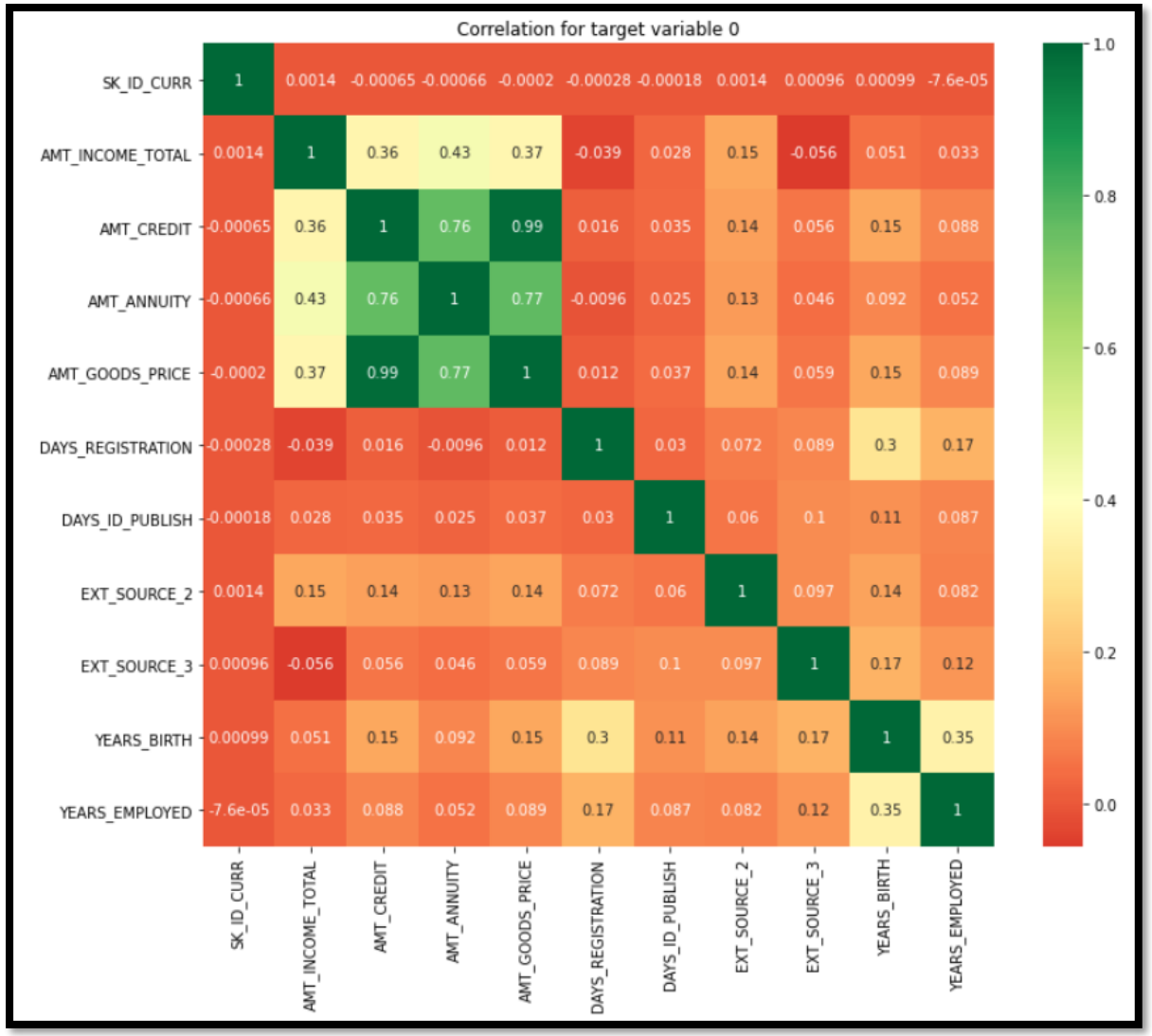
1. There is not any correlation between income type Businessman , Maternity leave, Pensioner , Student and the contract type

2. Number of cash loan client is very high having income type working

Approach Followed

- Business Problem Description
- Approach Flow diagram
- Imbalance of Target Variable
- Univariant Analysis
- Bivariant Analysis
- Correlation Analysis

Correlation Analysis for Customers without payment difficulties

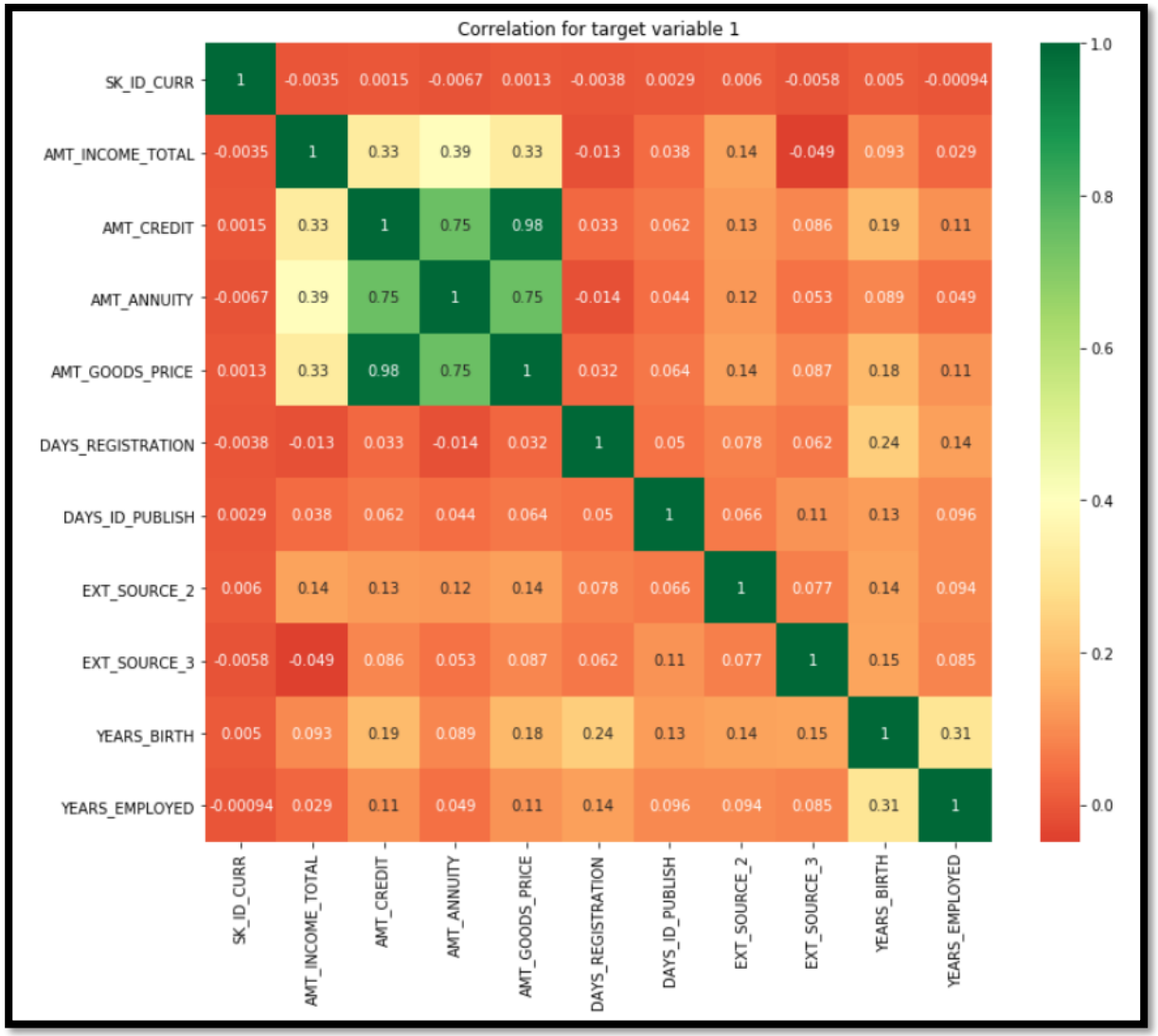


Conclusion Drawn: -

1. The heat map shows that for customers having no difficulty for loan repayment have highest correlation between Goods price of good that client asked for on the previous application and amount credited in last application

2. A strong correlation exist between Loan annuity and goods price of good that client asked for on the previous application

Correlation Analysis for Customers with payment difficulties



Conclusion Drawn: -

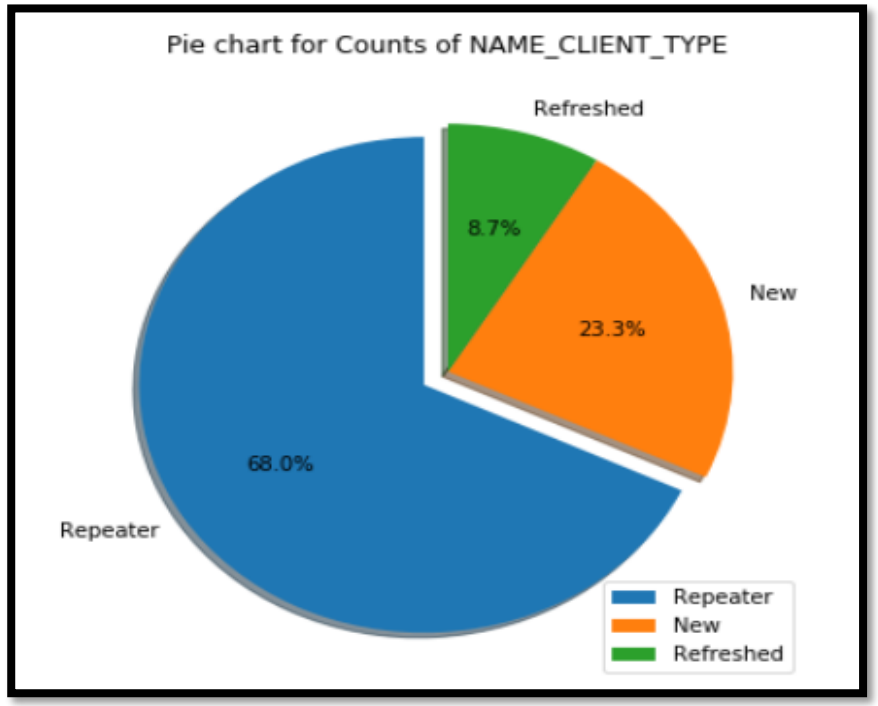
1. The heat map shows that for customers having difficulty for loan repayment have highest correlation between Goods price of good that client asked for on the previous application and amount credited in last application

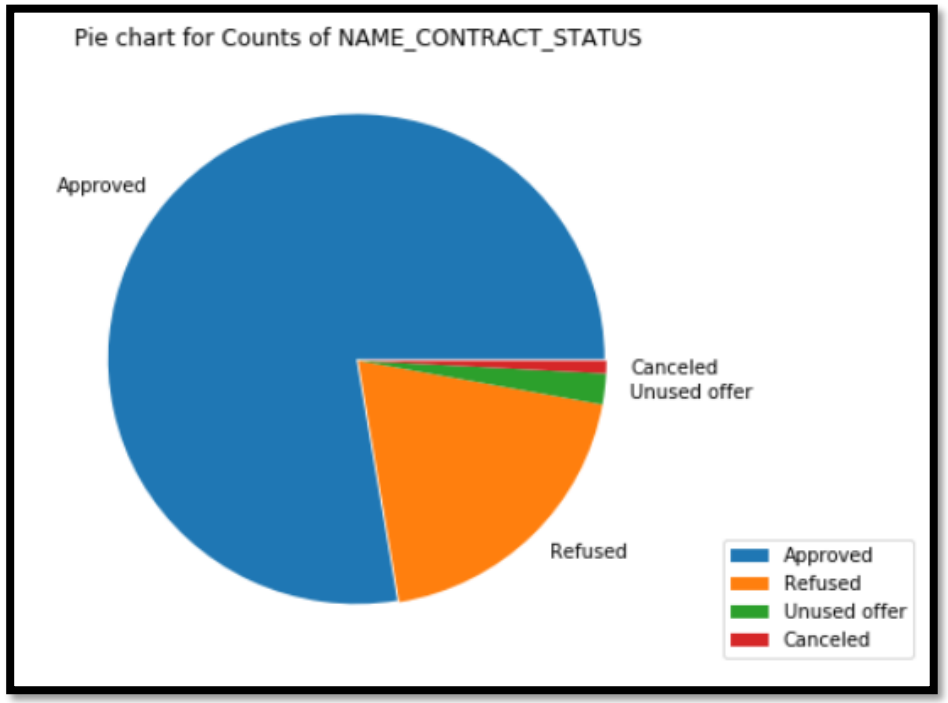
2. A strong correlation exist between Loan annuity and goods price of good that client asked for on the previous application

Approach Followed - Previous Application Data

- Loading and data manipulation
- Univariant Analysis
- Bivariant Analysis
- Merging of Application data with Previous Application Data

Previous Application Data Analysis





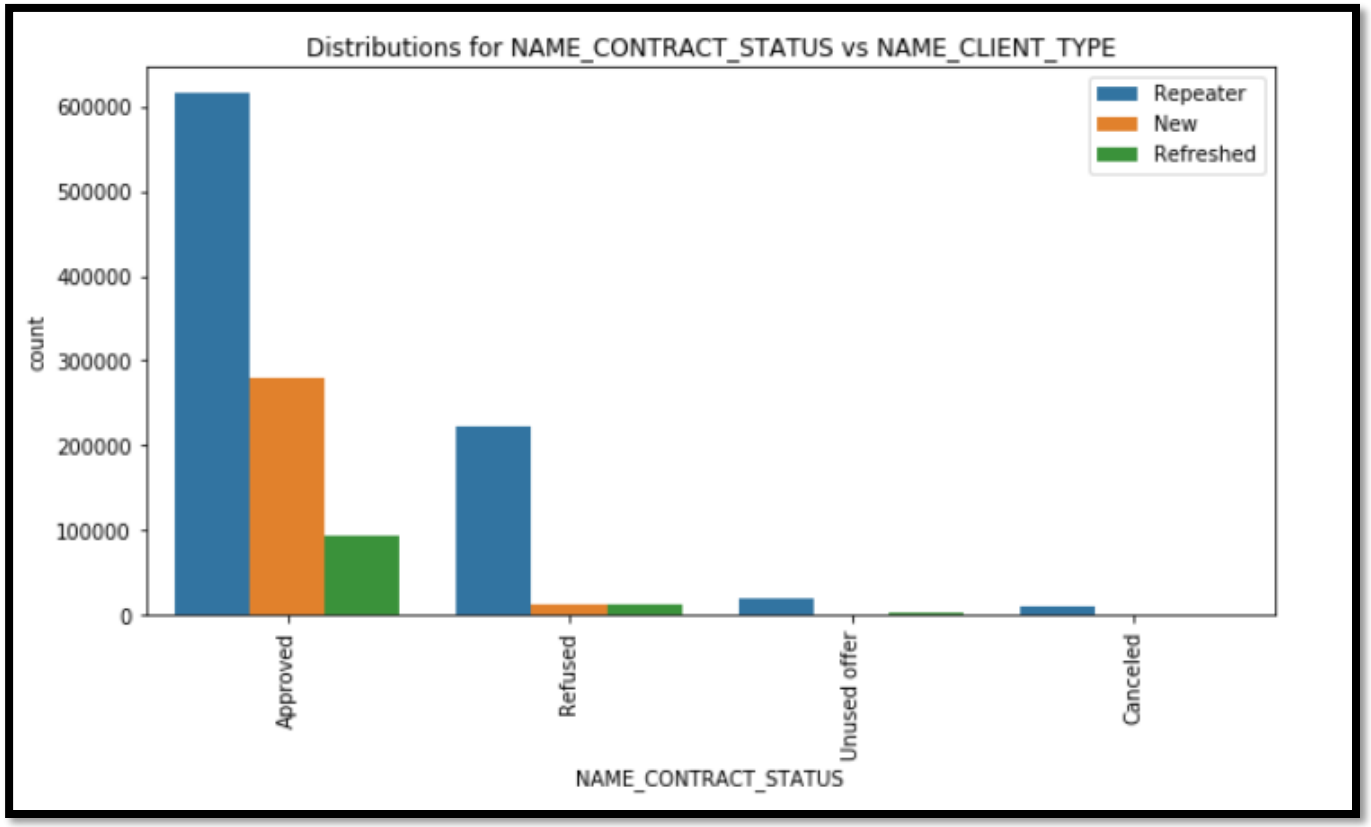
Conclusion Drawn: -

1. Highest percentage of clients are repeater clients which is 73.8%
2. 8.1% of clients are refresher, lowest percentage of the data
3. 18.1% of new clients are available according to data

Conclusion Drawn: -

1. From the available data 62.1 % loans has been approved
2. 1.6 % of clients consists of unused offer
3. There is very less difference between the clients who canceled and refused loans

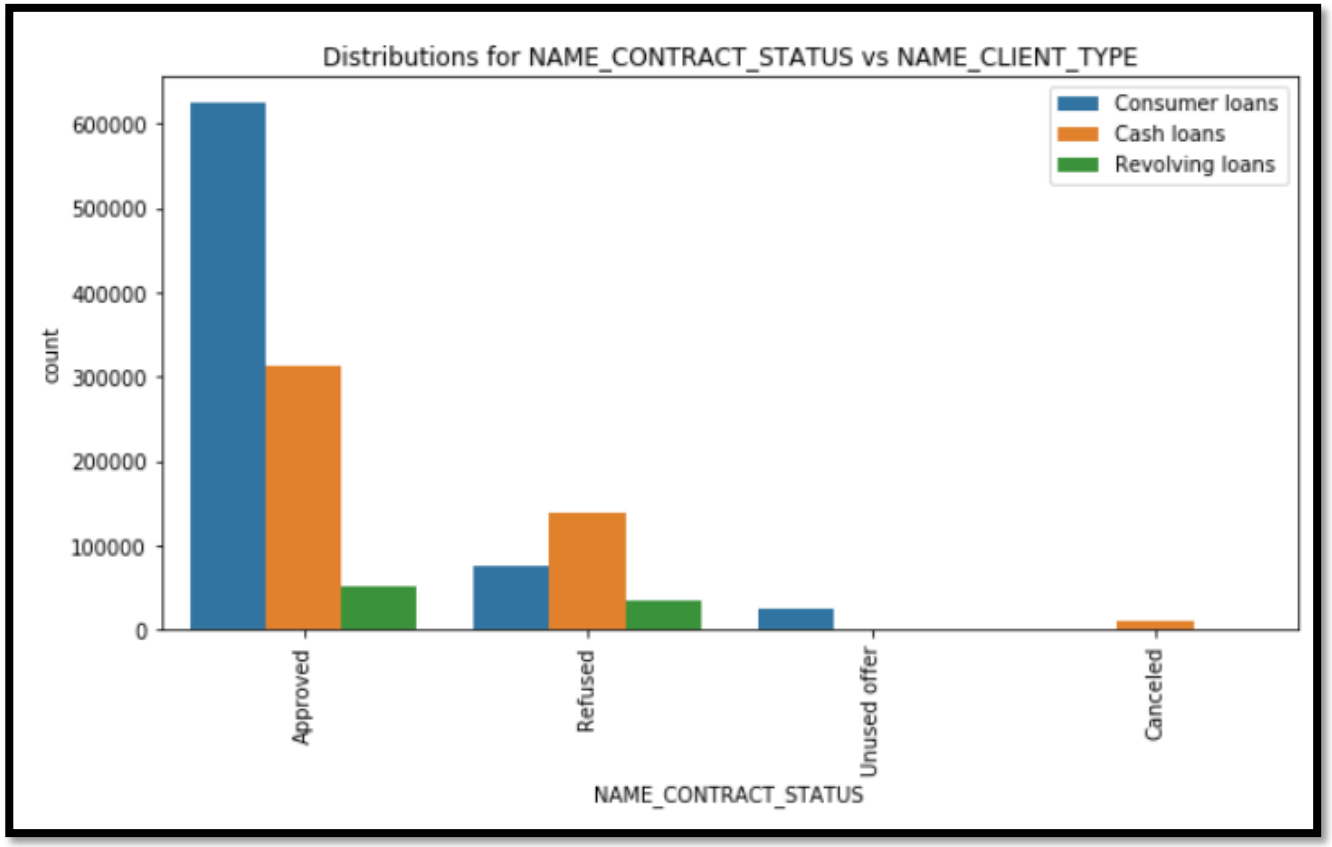
Distributions for Client Type vs Contract Status



Conclusion Drawn: -

1. Value counts of New and Refreshed clients is same for Refused category
2. We can say that for New and Refreshed clients has not cancelled loan as value count is not present for unused and cancelled category
3. Approved value count is much higher Repeater clients as compared to New and Refreshed clients

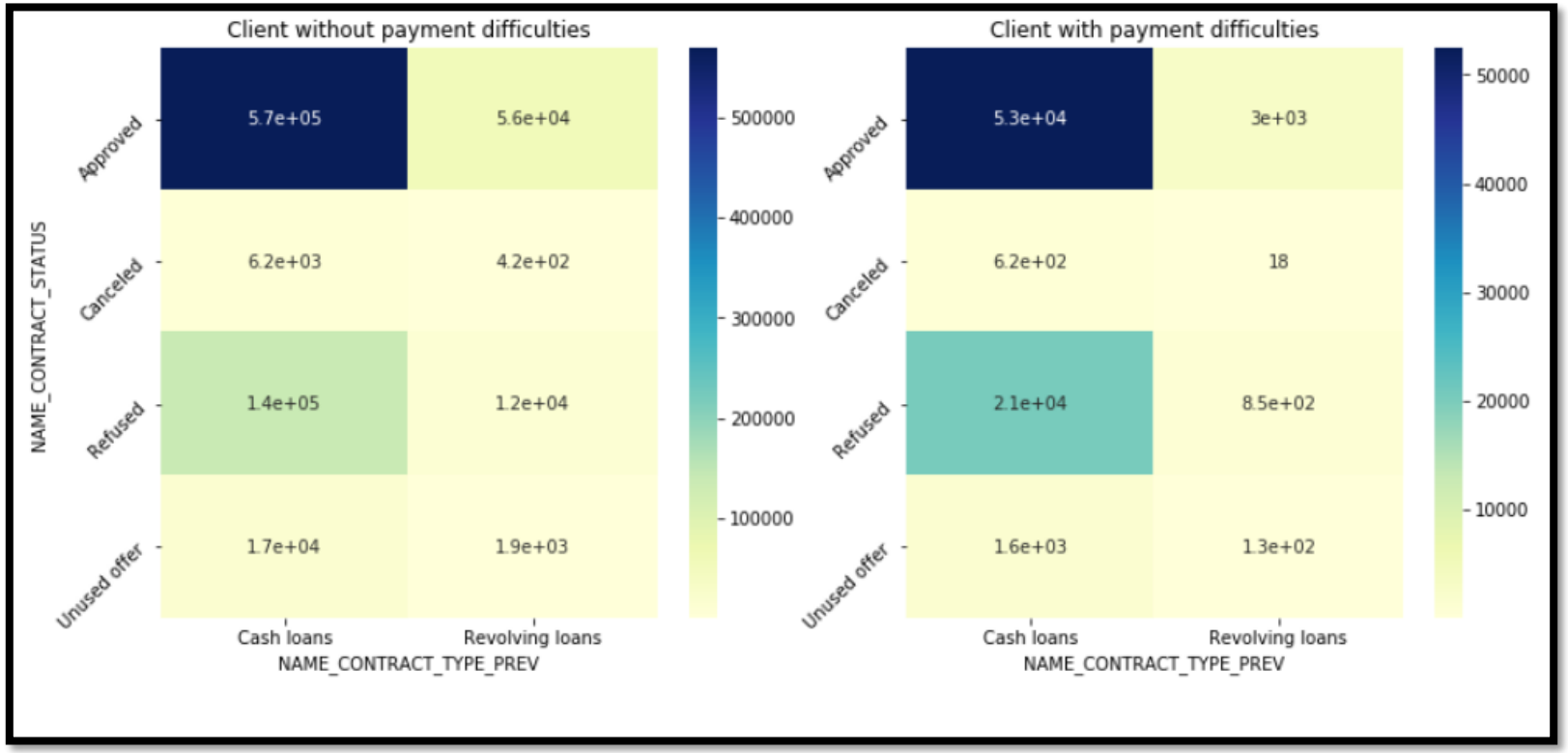
Distributions for Contract Type vs Contract Status



Conclusion Drawn: -

1. There is no correlation between Unused offer vs Cash loan and Unused offer vs Revolving loans
2. There is no correlation between Cancelled category vs Consumer loan and Cancelled category vs Revolving loans
3. More number of clients Refused cash loans as compared with Revolving and Consumer loans
4. Highest number of consumer loans has been approved

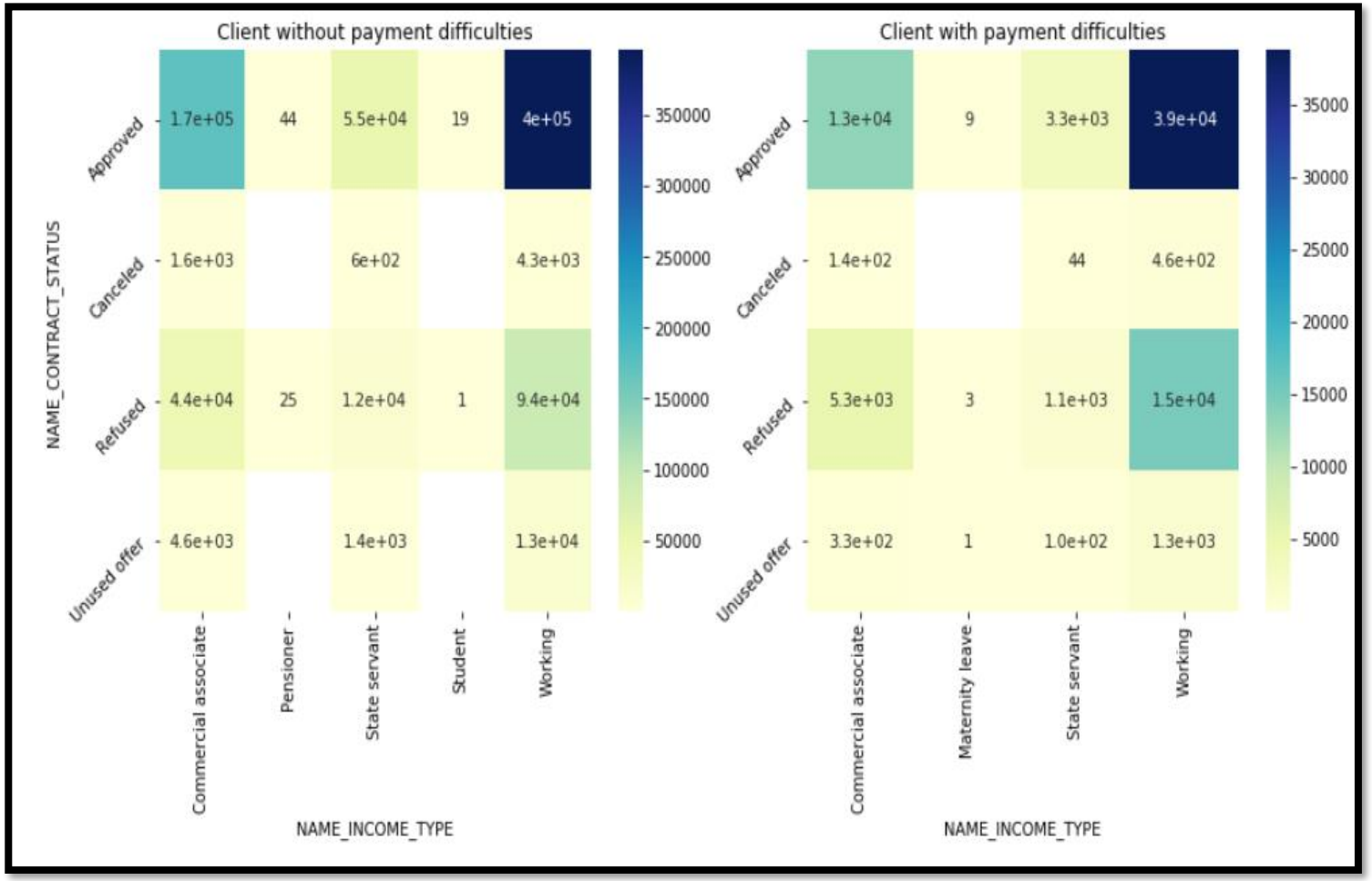
Distributions for Contract Status and Contract Type of Previous Application



Conclusion Drawn: -

1. Very less number of Revolving loans has been cancelled for the Clients with payment difficulties
2. Clients with payment difficulties has been refused to take Cash loans
3. We can infer that many clients applied for cash loans as number of Approved cash loans is very large as compared to revolving loans for both Clients with & without payment difficulties

Distributions for Contract Status and Income Type



Conclusion Drawn: -

1. We can infer that loan procedure of Pensioner clients is not cancelled
2. There is not any correlation between Pensioner vs unused offer , Student vs Cancelled for clients without payment difficulties
3. we can say that , 95% of student loan is approved for clients without payment difficulties
4. There is not any correlation between Maternity leave vs uncancelled for clients with payment difficulties

Case Study Final Conclusion

- Customer Occupation Type greatly effects possibilities of customers being able to repay the company's loan. e.g. Occupation type such as IT staff , HR staff , Managers etc. have lower difficulty in repayment when observed in comparison to Drivers, Cleaning staff ,laborers

- In general 10 to 15 % of the customers falling into categories Married , Single/ Not-Married , Separated etc. and belonging to Repeater , New and Refreshed faces difficulty in repaying the loan

- Analysis revels customers having income type Businessman, Maternity leave, Pensioner and Student are not applying for the loan. And Working and commercial associates are the highest who tends to apply for the loan.

- In general more number of Cash loans are provided by bank in comparison to Revolving loan post chances to get defaulted increases

- To increases the chances of repaying the loan provided by the companies should less focus on the clients having income type as Working


THANK YOU