



# Credit EDA Assignment

---

Abhishek Raikwar

# Exploratory Data Analysis



- The case study provided to us is to use EDA in a real life business scenario.
- We are asked to perform a risk analysis on the data shared to understand if the client applying for a loan in the organisation is capable to pay the installments on time by using previous loan details.

# Credit Risk Analysis



- Credit Risk Analysis is evaluating a borrower's ability to pay back a loan and determine the likelihood of default.
- **Types of Credit Risk:**
  - 1) Credit default risk- unable to pay the loan obligation in full or when the borrower is already 90 days past the due date of the loan repayment
  - 2) Concentration risk- exposure to a single counterparty or sector, and it offers the potential to produce large amounts of losses that may threaten the lender's core operations.
  - 3) Country risk- when a country freezes foreign currency payments obligations, resulting in a default on its obligations

# Steps for Credit Risk Analysis

## Exploratory Data Analysis for Credit Risk Modeling



Steps followed for this EDA are:

1. Data review
2. Research Question formulation
3. Data cleaning and pre-processing
4. Data analysis -Uni, Bi, Multivariate Analysis
5. Data visualization
6. Data presentation

# Research Questions



- To identify patterns which indicate if a client has difficulty paying their instalments; thereby help company decide taking actions such as denying the loan, reducing the amount of loan, lending loans to risky clients at a higher interest rate, and other decisions.
- To ensure consumers capable of repaying the loan are not rejected.

# Data Cleaning



Data Cleaning

- Data cleaning involved the following variables:
  - AMT\_Annuity; missing were replaced to median
  - CNT\_FAM\_MEMBERS; missing were replaced to mode
  - Gender; XNA were changed to females
  - Missing salaries were removed
  - Organization types with XNA were removed
  - Income were binned
  - Age was changed from days to years and then binned
  - Credit amount was binned
  - TARGET variable was separated for individual analysis

# Data Imbalance Ratio



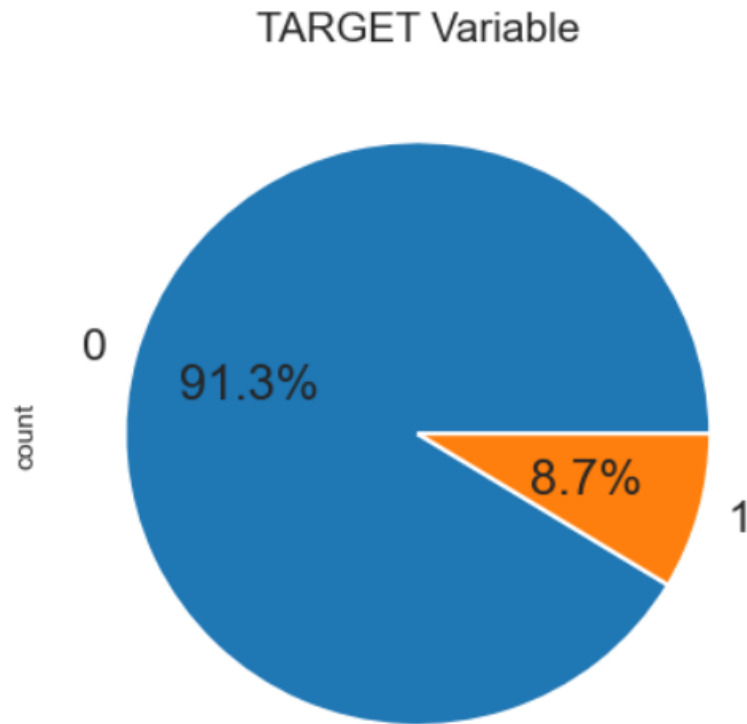
- For the ease of understanding, Target group 1 will be called defaulters and Target group 0 will be non-defaulters.
- The data imbalance ratio was found to be 10.55.
- This indicates that Target group 0 is 10.55 times more than Target group 1. This data is fairly good as this indicates lesser clients are defaulters.

```
[69]: # checking imbalance percentage in the data set  
      round(len(app1)/len(app2),2)
```

```
[69]: 10.55
```

The Imbalance ratio is 10.55.

# Distribution of Defaulters and Non-Defaulters by pie chart



- 1 in "Target" describes Defaulters or clients who are unable to pay their loans and 0 are Non defaulters who are able to pay their loans on time.
- In this pie chart describing "Target", we can see that the clients with financial difficulties to pay are at 8.1% and all other cases are at 91.3%.



# Distribution of Previous Application Status

- In the pie chart, the percentage of previous approved loans were are 62.1% which was significantly higher.
- Very less percentage of clients have unused offer in the past.
- Loans for clients in the 17.4% of refused loans need to be reviewed with care.

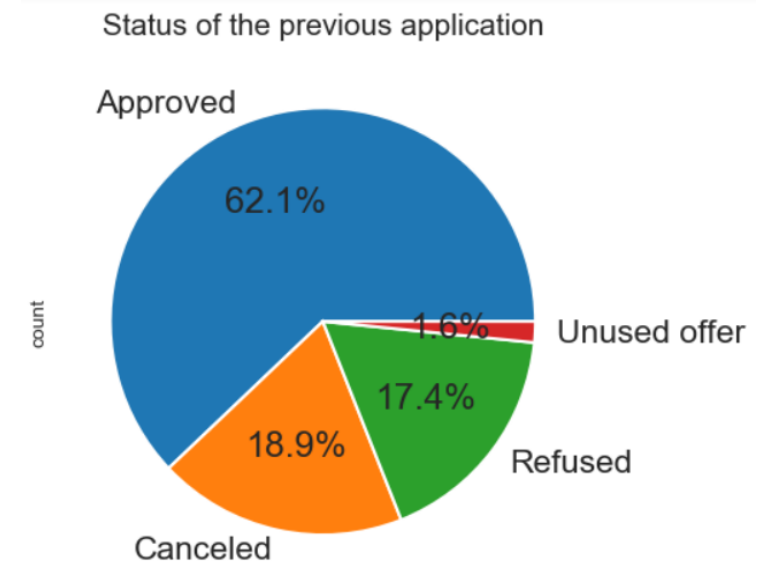
When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

**Approved:** The Company has approved loan Application

**Cancelled:** The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client, he received worse pricing which he did not want.

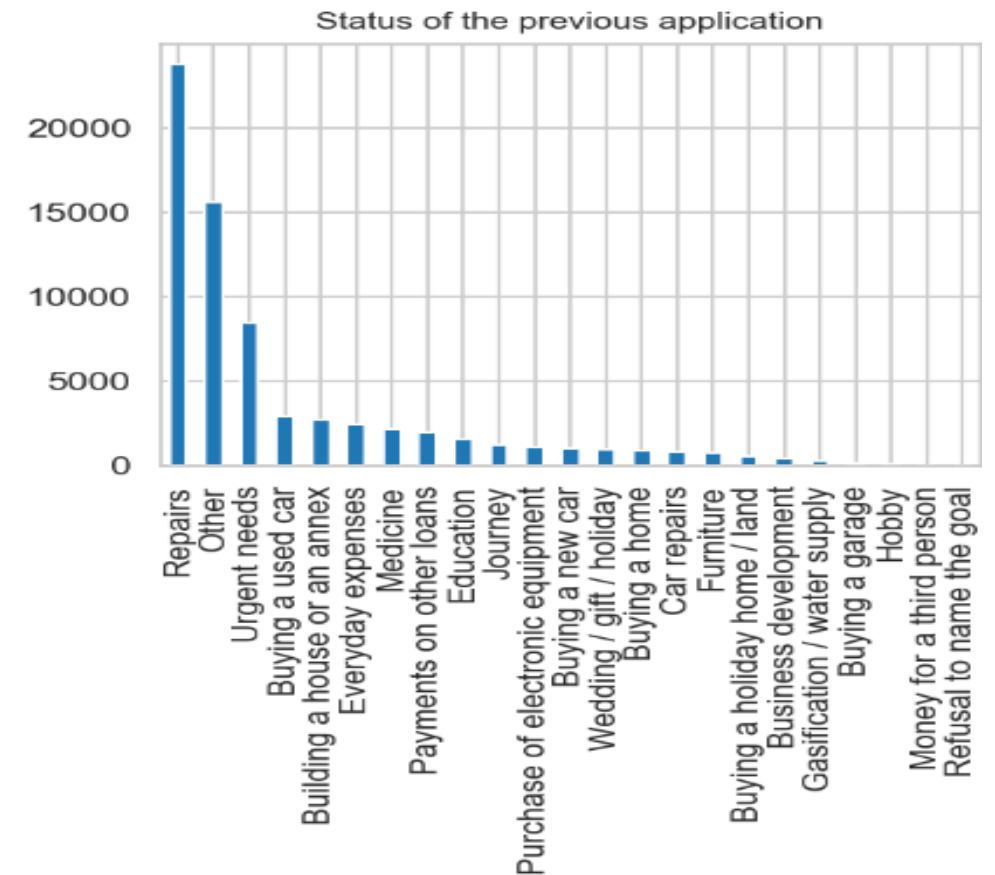
**Refused:** The company had rejected the loan (because the client does not meet their requirements etc.).

**Unused offer:** Loan has been cancelled by the client but at different stages of the process.

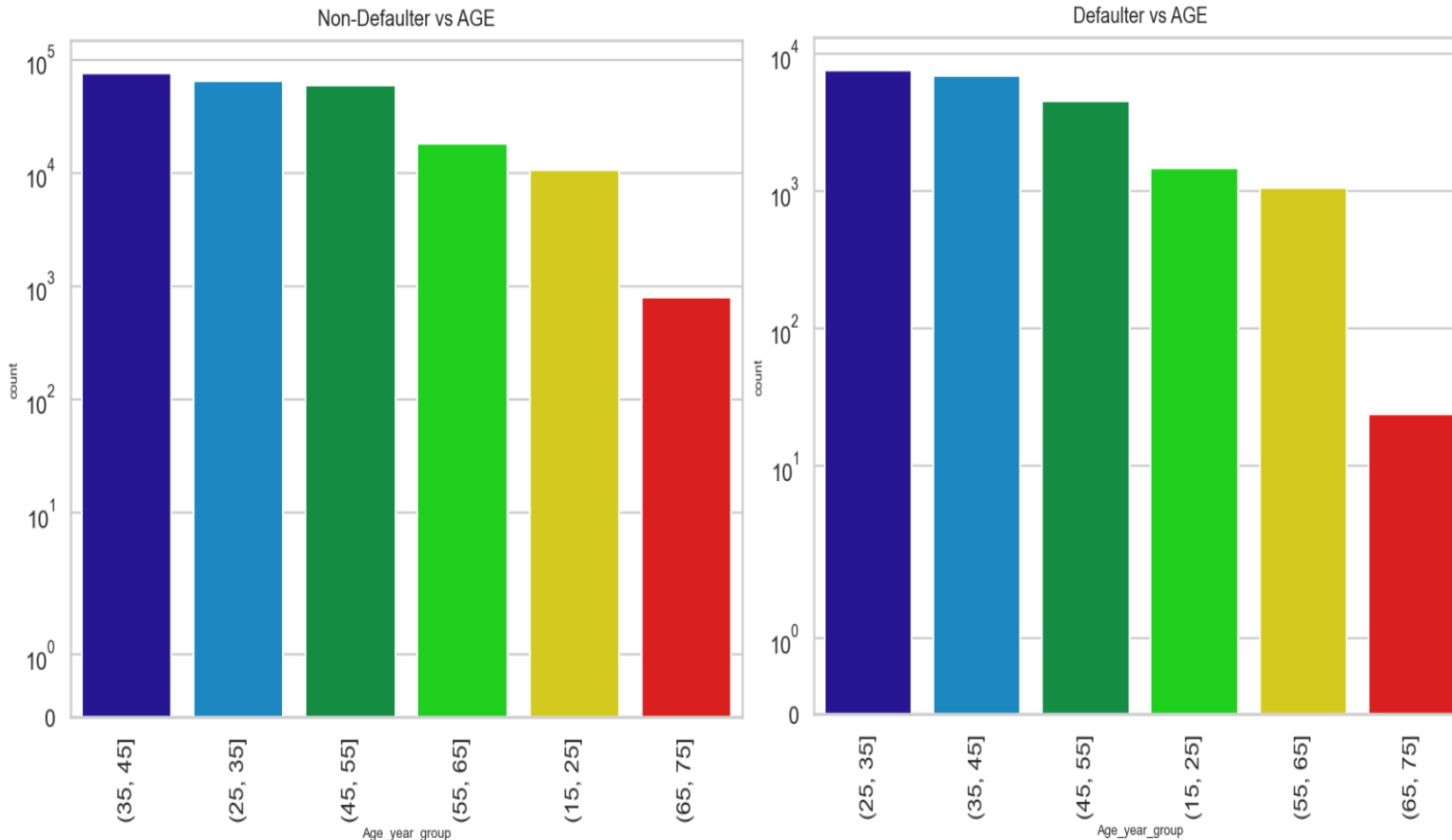


# For what purposes did the clients apply for a loan previously?

- The bar graph shows various reasons why the clients have applied for loans previously.
- From this data, we can infer that majority of the clients took loans for repairs.

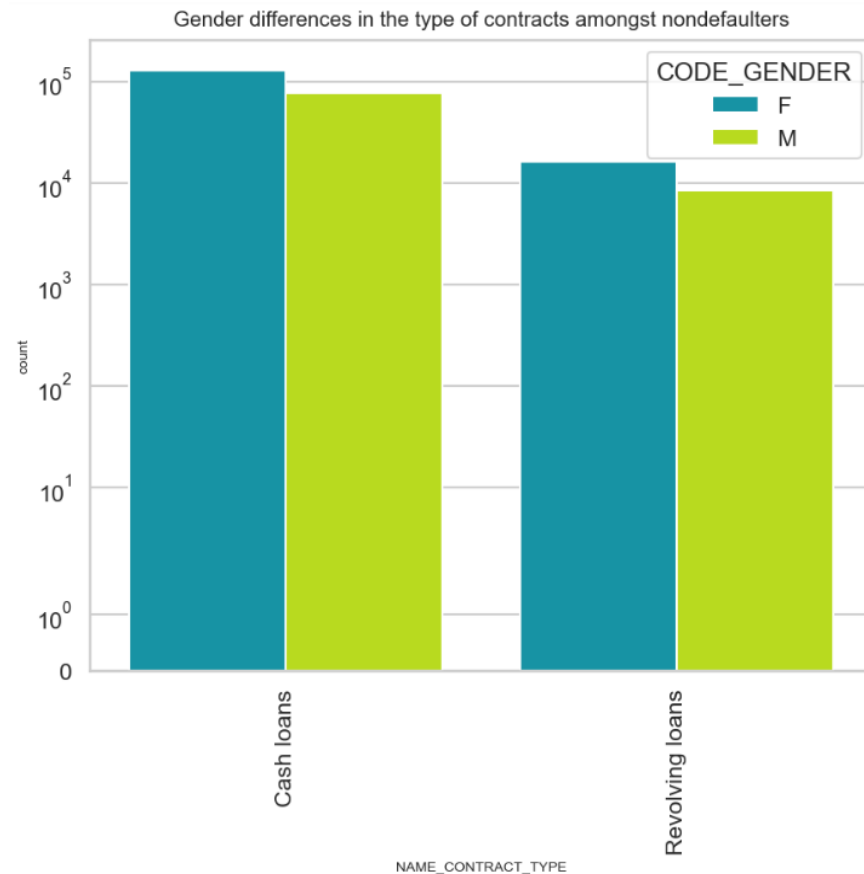


# Relation between TARGET groups and Age categories



- On comparison, defaulters are mainly aged between 25 to 45 years.
- On the other hand, non-defaulters are mainly aged between 25 to 55 years of age.
- We can also conclude that there are lesser defaulters in the age group of 65 to 75 years of age.

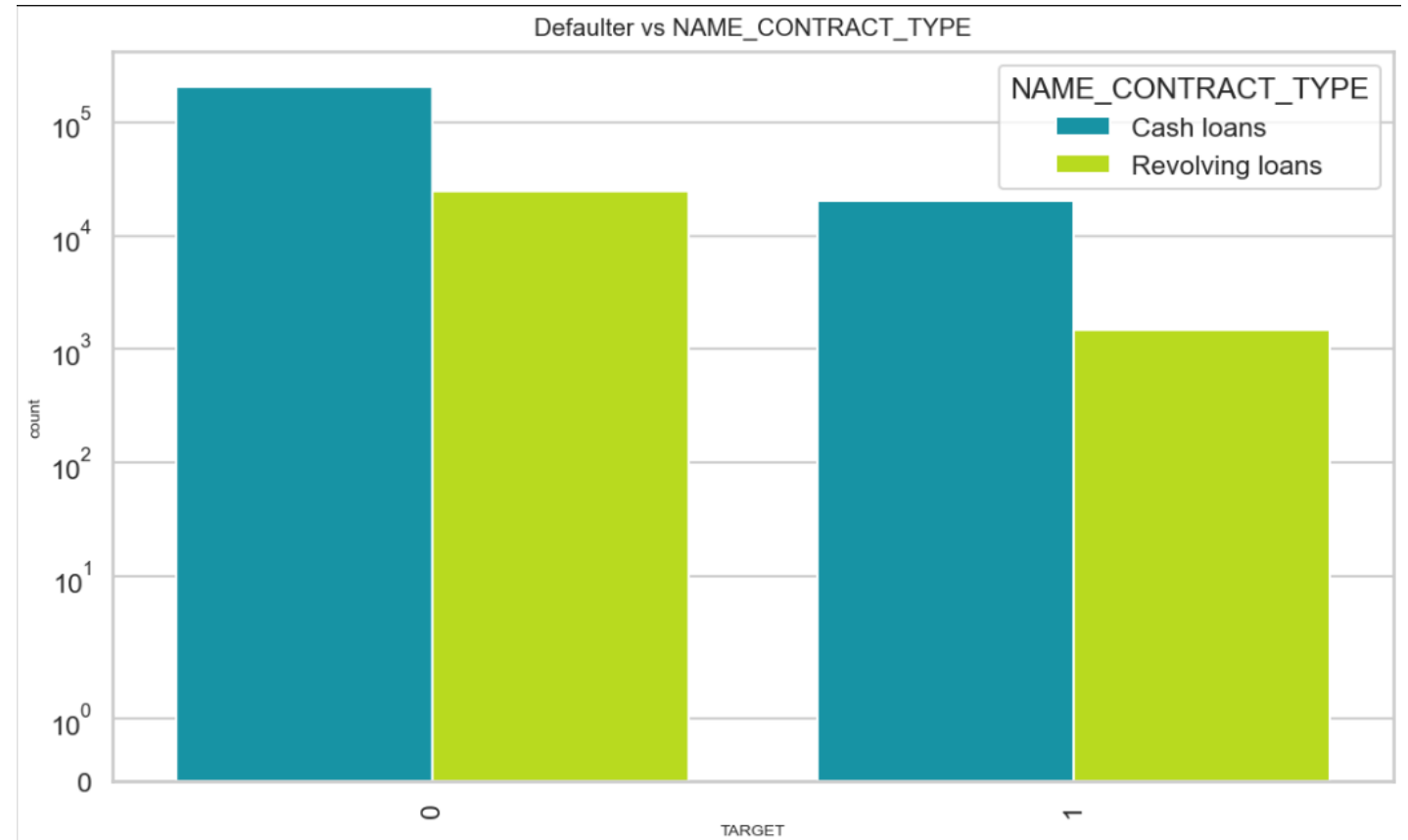
# Relation between the Loan contracts and Genders within Non Defaulter clients



- Non defaulters are able to pay their loans better in case of cash loans than revolving loans.
- Within the genders, female tend to pay their loans sooner than men.

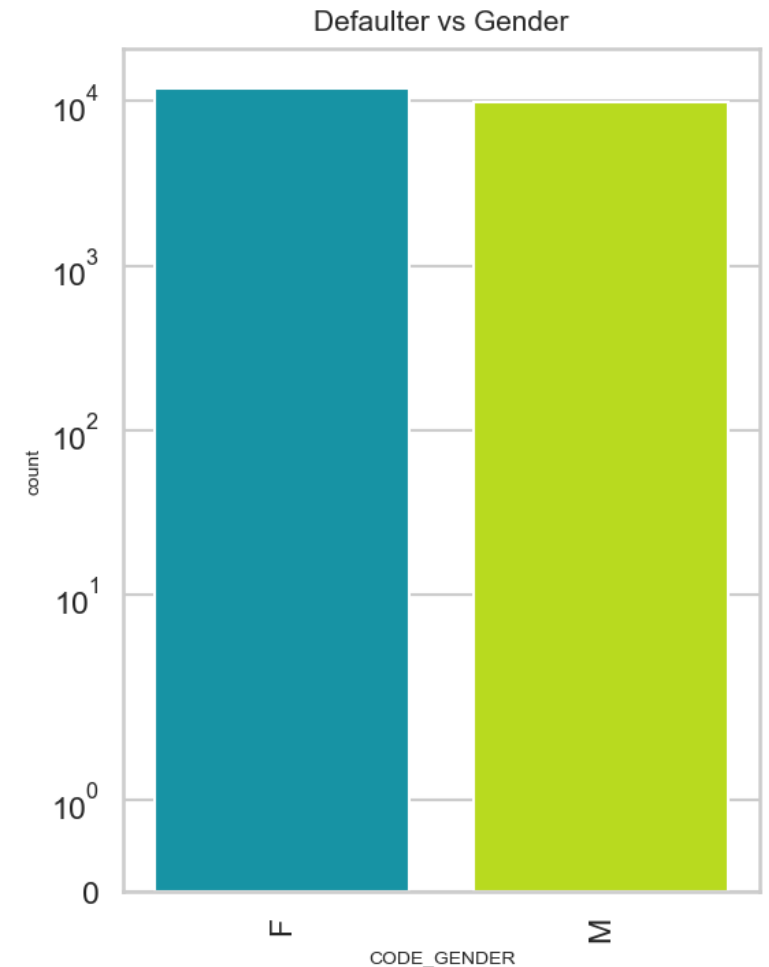
## Relation between the Loan contracts and Gender differences in Defaulter clients

Defaulters are able to pay their loans better in case of revolving loans than cash loans.

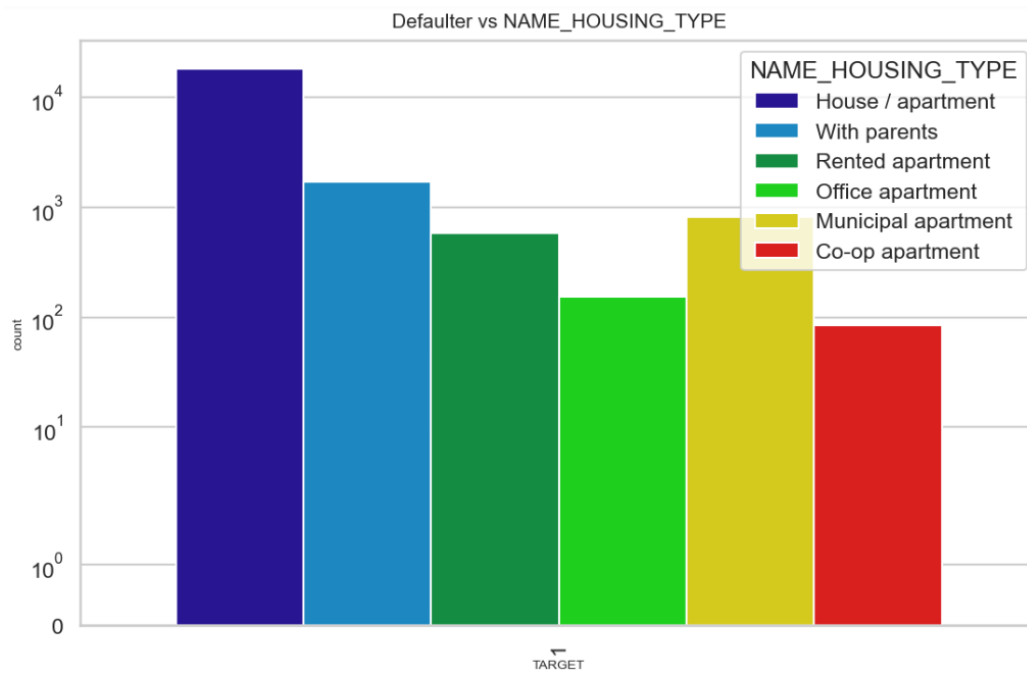


# Relation between Defaulters and Genders

- Amongst the defaulters or clients who are not able to pay their loans on time, there are slightly more females than males.
- As the difference is not high enough, we cannot conclude that women tend to be defaulters.



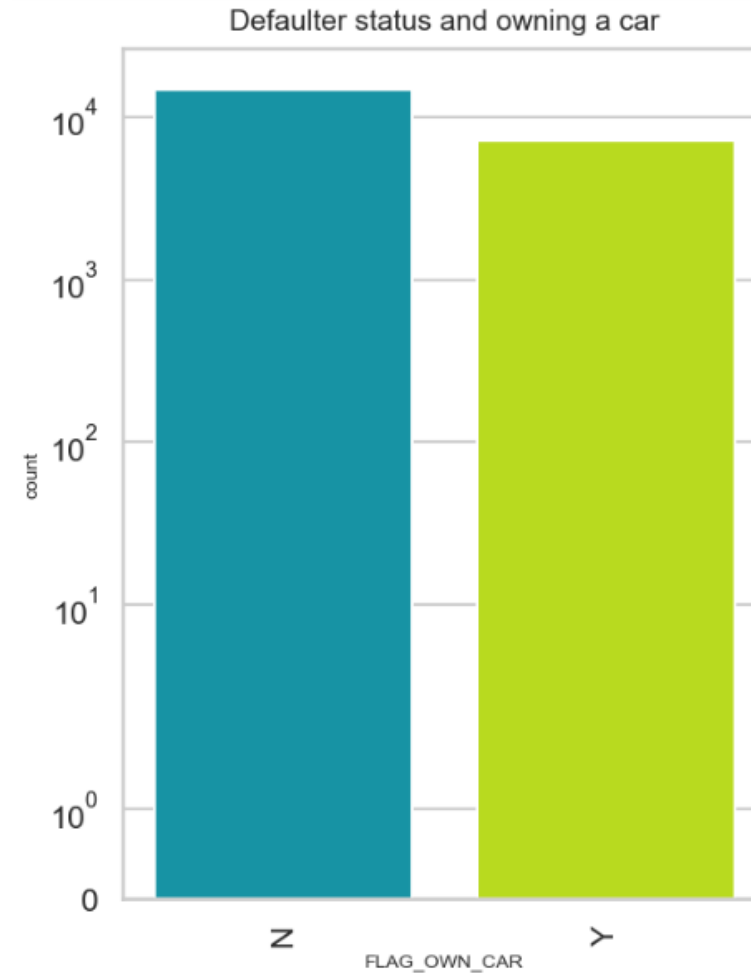
# Relation between the Housing types and Defaulter status



- It can be observed that clients living in house/apartments tend to have difficulty paying their loans.
- Clients living in the office or co op apartments do not face as much difficulties as others in paying their loans.

# Relation between the Defaulter status and Car ownership

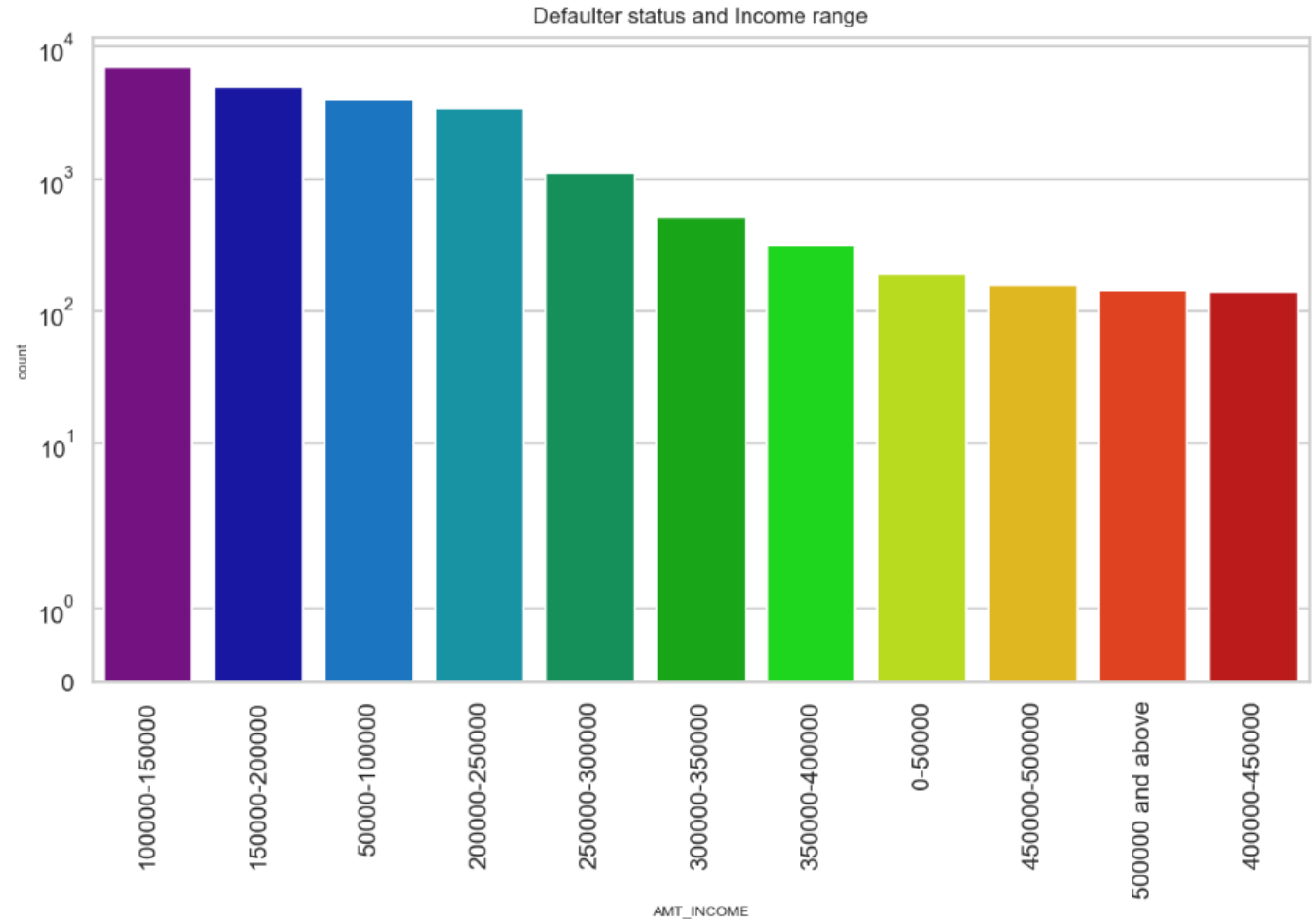
It can be observed that clients who do not own car face slightly more difficulty in paying their loans.



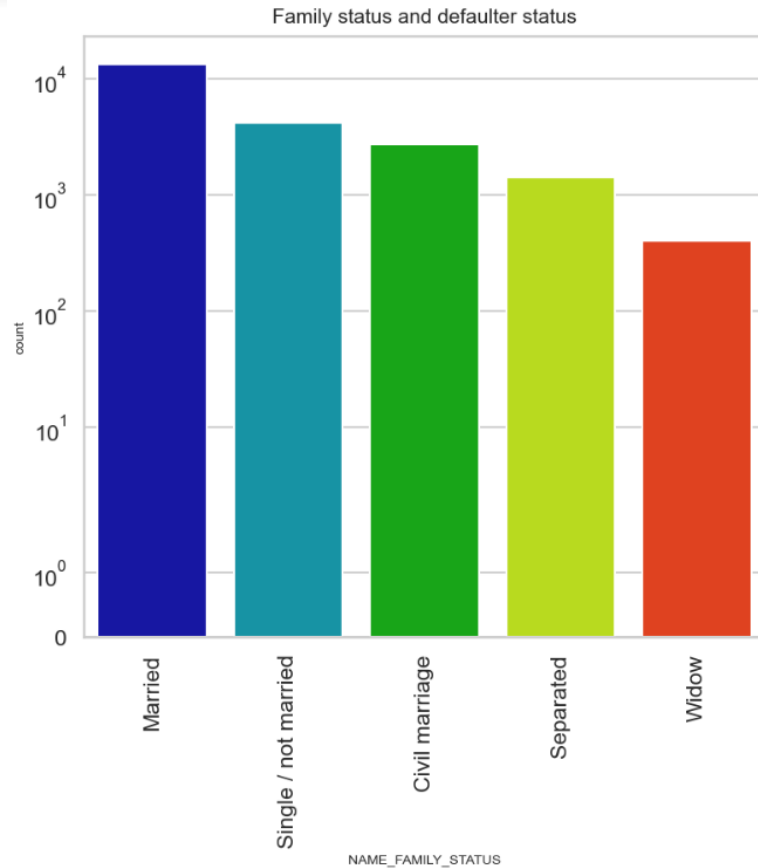


# Relation between the Defaulter status and income range of clients

People with income between 50000 to 300000 face more problems in paying their loan installments; however, this also depends on the amount of loan credited.



# Relation between Defaulter status and client's Family Status

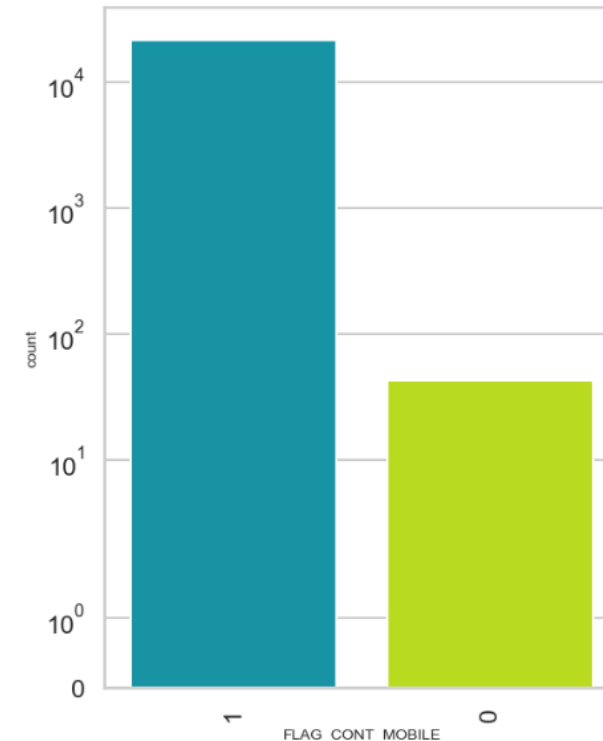


- It can be observed that married clients tend to face difficulty paying their loans.
- Consequently, widows tend to have lesser difficulties paying their loans. This could be due to lower expenses to bear.

## Relation between Defaulter status and reaching their mobile phones

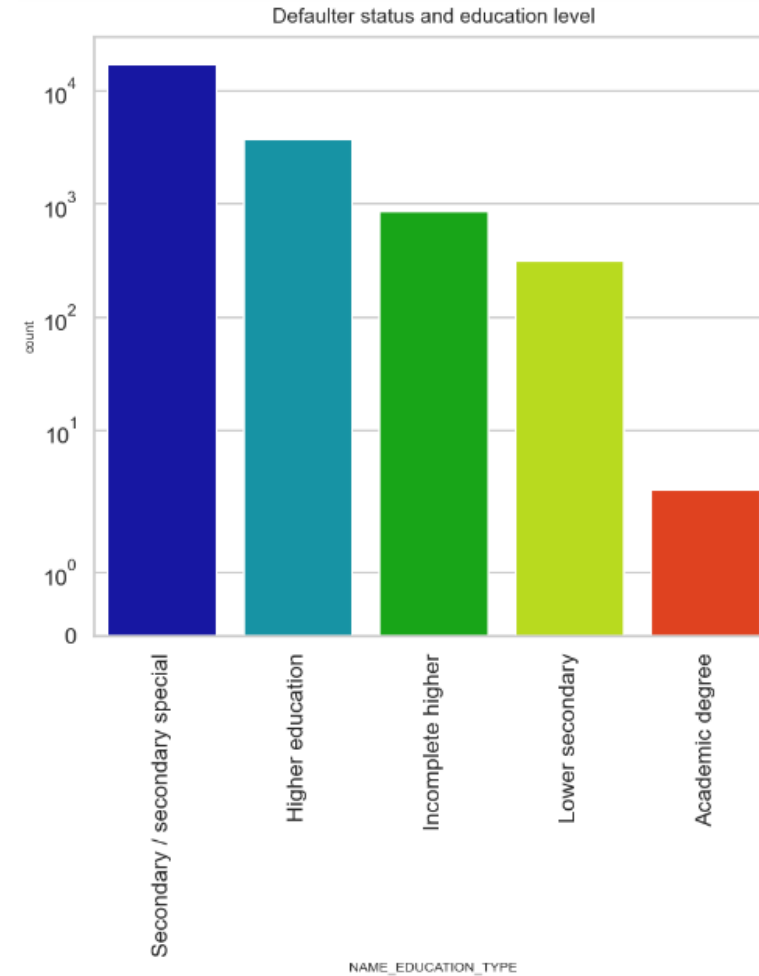
- There was difficulty reaching a significant number of defaulter clients via their cell phones despite providing a phone number.

Difficulty reaching client through provided mobile number amongst defaulters

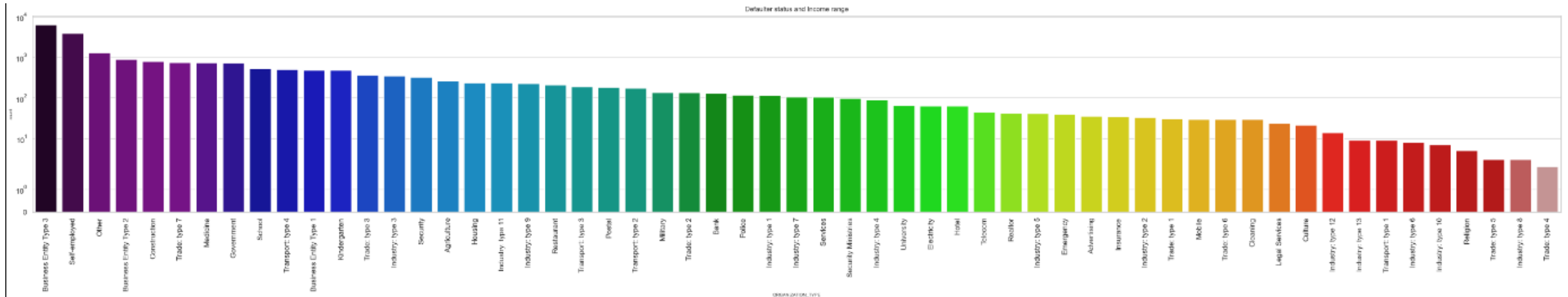


# Relation between Defaulter status and Education level

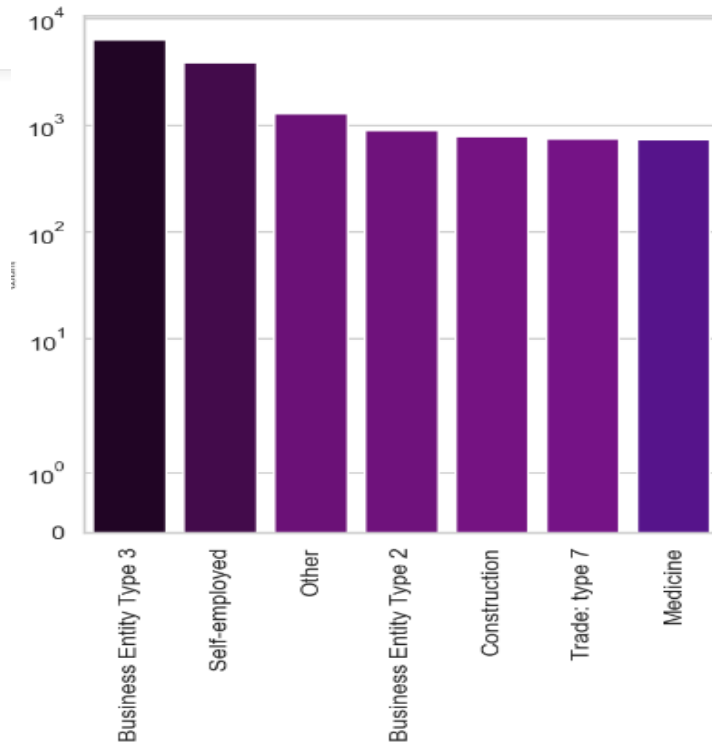
- There are more defaulters with secondary level of education.
- Lesser clients with an academic degree tend to be defaulters.



# Relation between the defaulter status and organization type



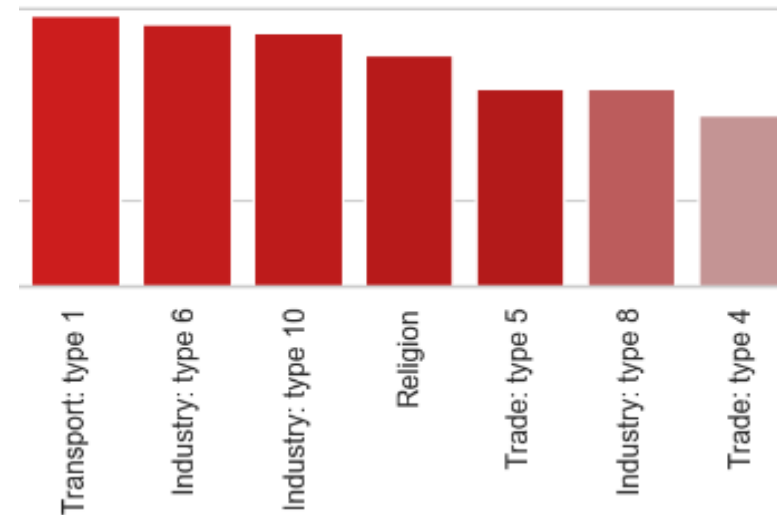
## Defaulter Majority



It can be observed that majority of the defaulters belong to 'Business Entity type 3', 'self employed', 'Other', 'Business Entity type 2' or 'Construction' organization types.

6/25/2024

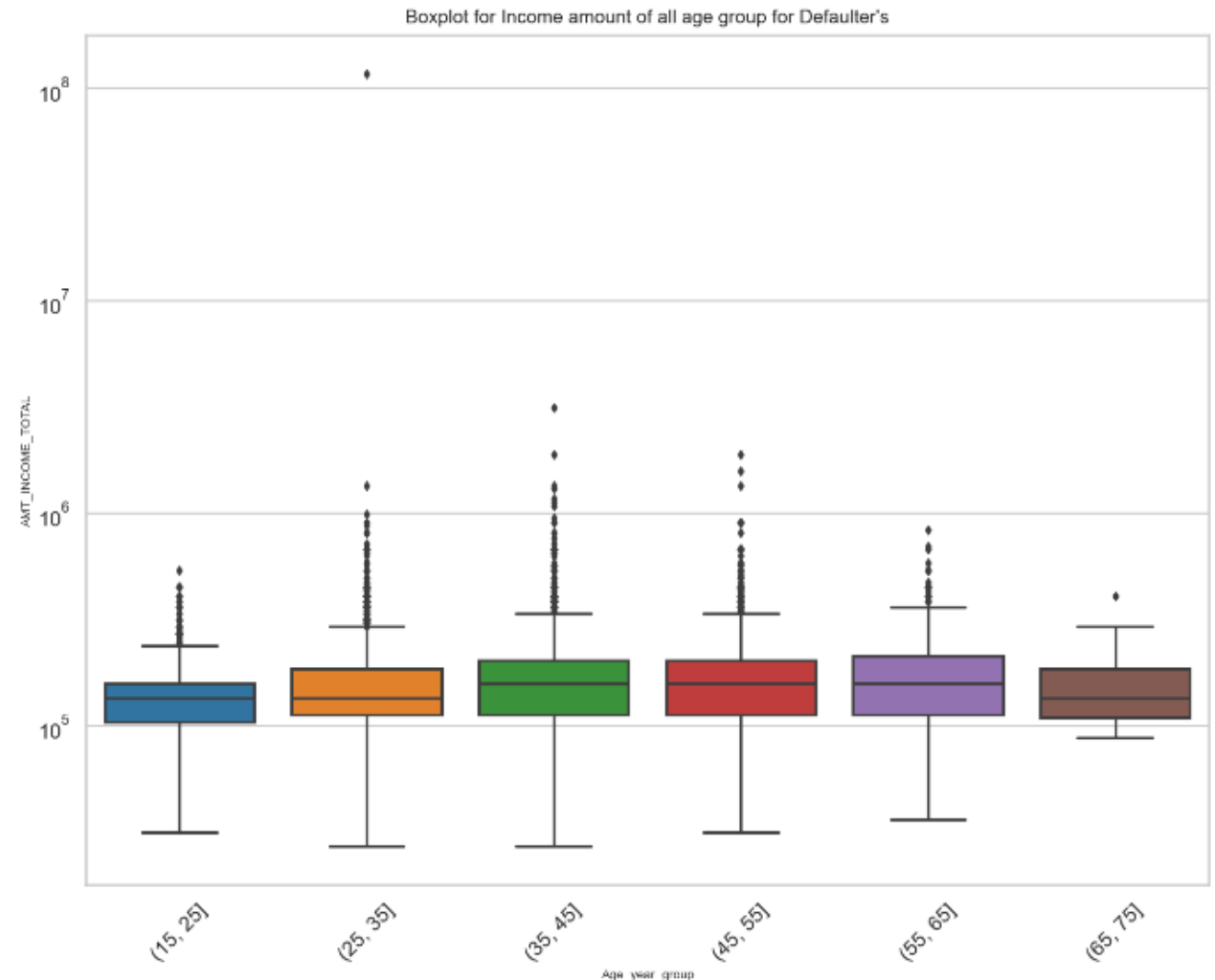
## Defaulter Minority



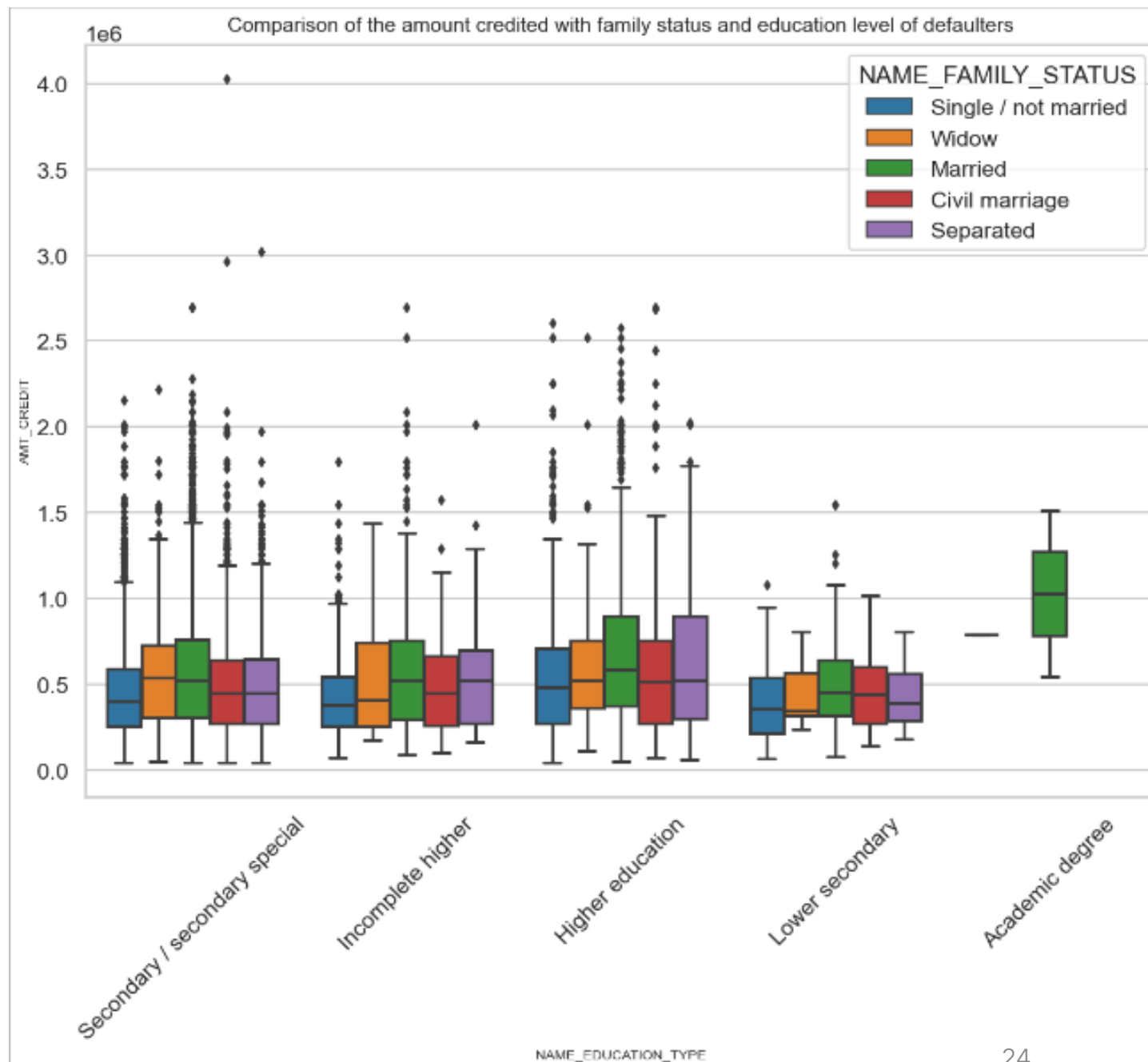
It can be observed that there are lesser defaulters in 'Industry type 10', 'Religion', 'Trade type 5', 'Industry type 8', and 'Trade type 4' organization types.

# Relation between age groups and salary levels amongst the defaulters

- It can be noticed that clients in the age group of 15 to 25 have the lowest range of salary.
- Some clients from 25 to 35 years of age have salaries lesser than clients in other groups.
- The smallest range of salary are for 65 to 75 years of age. Also, they earn more than others and are lesser defaulters.



# Comparison of the education level, family status, and loan credits amongst the defaulter group

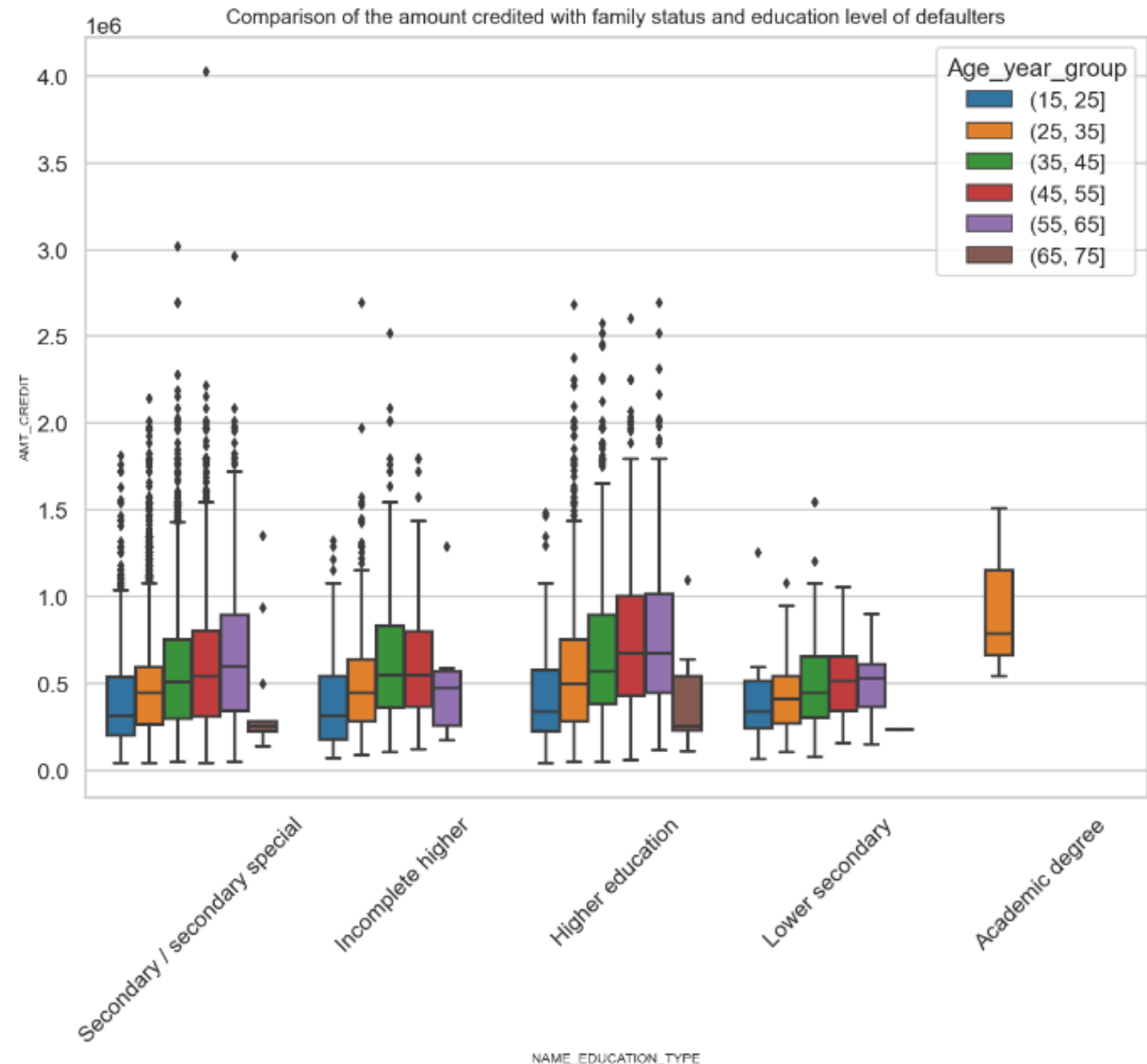




# Comparison of the education level, family status, and loan credits amongst the defaulter group

- This multivariate analysis using boxplots review the relationship between the education level and family groups of defaulters to the amount of loan credited.
- It can be noted that people from windowed clients are better at paying their loans.
- Separated secondary and higher education clients tend to be defaulters.
- We have less data on academic degree holders who are defaulters.

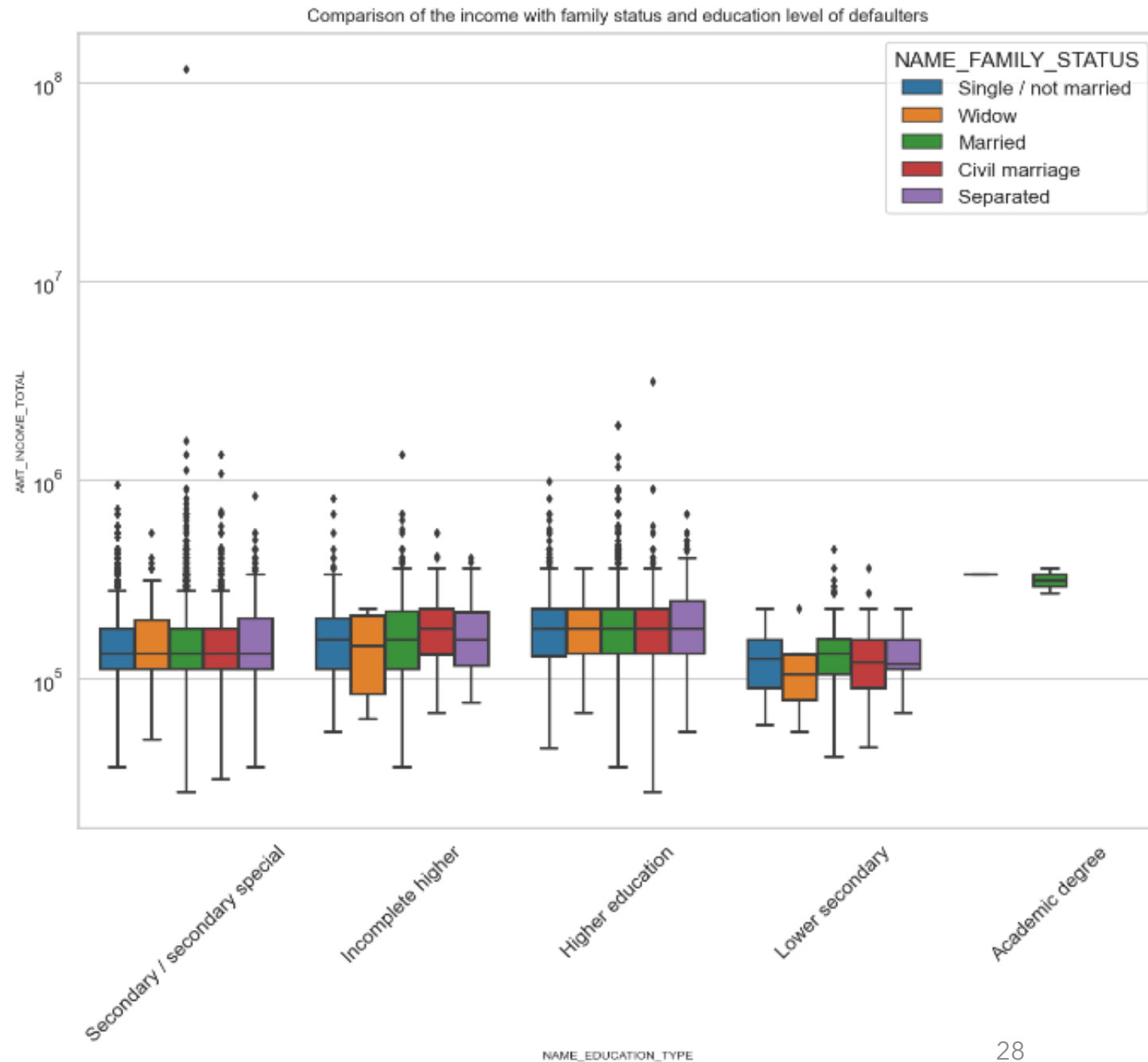
Comparison of the education level, age groups, and loan credits amongst the defaulter group



## Comparison of the education level, age groups, and loan credits amongst the defaulter group

- This multivariate analysis using boxplots review the relationship between the education level and age groups of defaulters to the amount of loan credited.
- It can be noted that people from higher education groups especially in the 55 to 65 years of age are majorly defaulters. They also take up comparatively higher amounts of loan.
- Notably, defaulters with an academic degree are lesser. They take up less loans as well and belong to ages 25 to 35.
- Within the secondary education level, with age there are more defaulters.
- The least risky are clients above 65 years of age and those who have an academic degree.

# Comparison of the education level, family status, and income amongst the defaulter group



# Comparison of the education level, family status, and income amongst the defaulter group

- This multivariate analysis using boxplots review the relationship between the education level and family status of defaulters to the income levels.
- It can be noted that the income levels amongst lower secondary level education is the least.
- Academic degree holders have the highest salaries.
- Amongst the higher education level, the salary levels are almost similar across different family status.



## Final inferences after data analysis

- Majority of the clients (91.3%) are able to pay off their loans.
- 62% of loans are approved, about 17 and 19 percent are refused and cancelled. We need to know the reason for refusal and cancelation.
- The main reason to avail a loan is for repairs.

# Defaulters/Risky Clients



- Age: Usually 25 to 45 years of age
- Gender: Males more than females
- Loan type: Cash loan
- Education level: Secondary and Higher Education
- Income level: 50000 to 300000
- Housing type: House/Apartment and with parents
- Family status: Married and Single
- Car owning status: No
- More previously rejected loans
- Organization type: 'Business Entity type 3', 'self employed', 'Other', 'Business Entity type 2' or 'Construction'
- Companies have difficulty reaching risky clients on phone

# Non Defaulters/ Non Risky Clients

- Age: Usually 65 to 75 years of age
- Gender: Females more than men; however, they take more loans than men.
- Loan type: Revolving loans
- Education level: Academic degree
- Income level: Above 300000
- Housing type: Office and co-op apartments
- Family status: Widowed and separated
- Car owning status: Yes
- More previously approved loans
- Organization type: Industry type 10', 'Religion', 'Trade type 5', 'Industry type 8', and 'Trade type 4'
- Companies do not have difficulty reaching risky clients on phone





# Thank You

