Fintech Financial Risk Assessment

CREDIT RISK ANALYSIS

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.
- In this Case Study we use **EDA** to understand how consumer attributes and loan attributes influence the tendency to default.

Business Objective

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

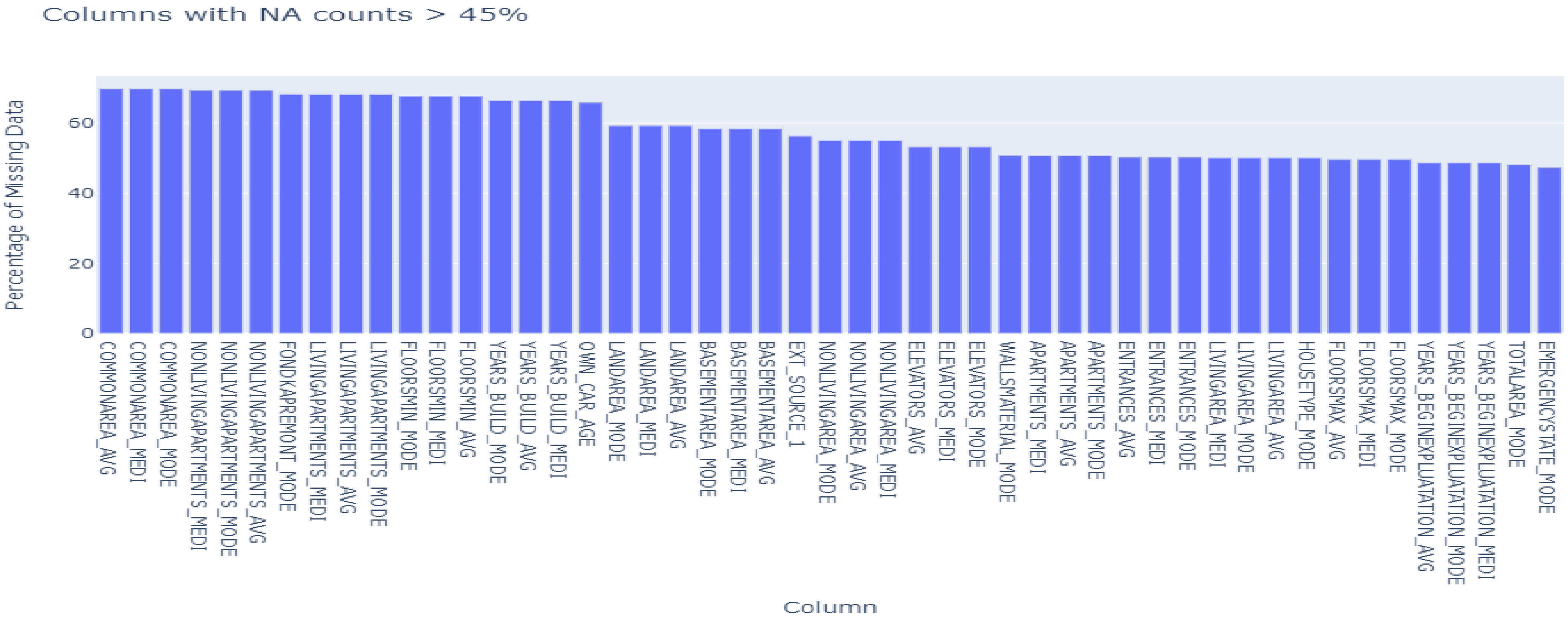
Please note that 'XNA' has been considered as Null Value across all datasets used

- Three data sets have been used:
- *1. 'application_data.csv'* contains all the information of the client at the time of application. The data is about whether a **client has payment difficulties.**
- *2. 'previous_application.csv'* contains information about the client's previous loan data. It contains the data on whether the previous application had been **Approved, Cancelled, Refused or Unused offer.**
- *3. 'columns_description.csv'* is data dictionary which describes the meaning of the variables.

What have we covered under this case Study?

- 1) Dimensions of the data
- 2) Cleaning the data / Correcting negative values
- 3) Missing Values Analysis (Including XNA Values)
- 4) Univariate Analysis
- 5) Bivariate Analysis

Columns with missing data > 45% have been dropped

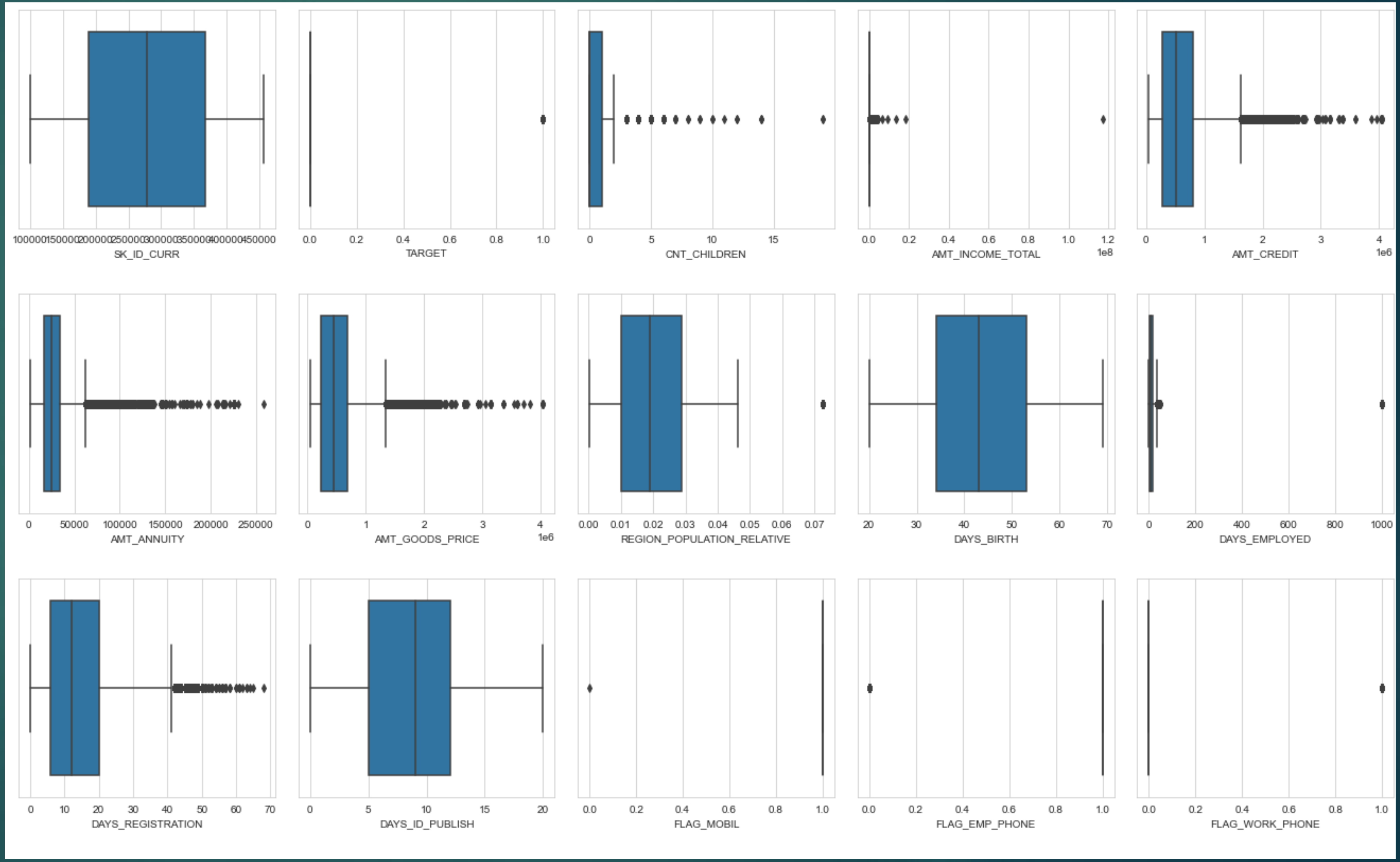


Outlier Analysis

There are many outliers in the dataset.

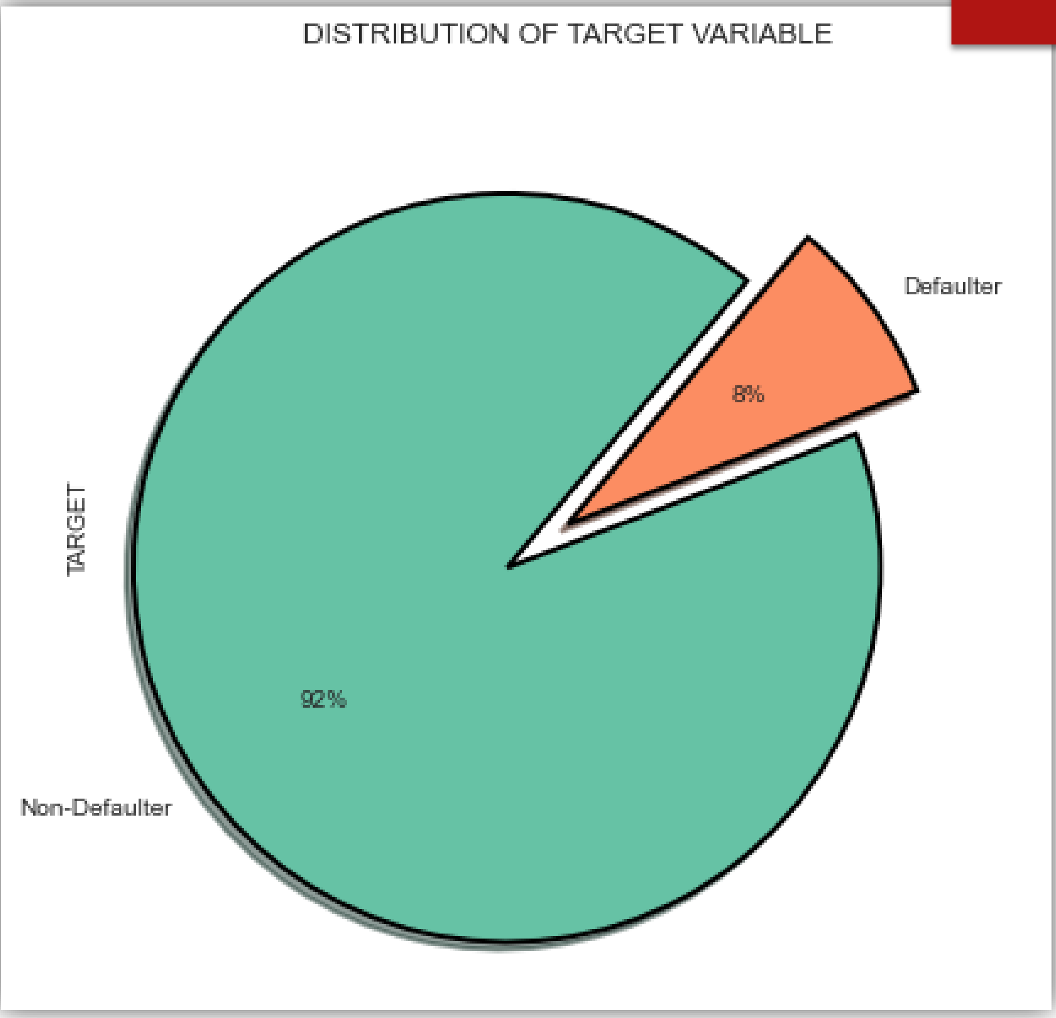
Except DAYS_BIRTH ,DAYS_ID_PUBLISH and EXT_SOURCE_2,EXT_SOURCE_3,FLAG_DOCUMENT_3 , all the other columns have outliers.

We see a similar trend across both the datasets



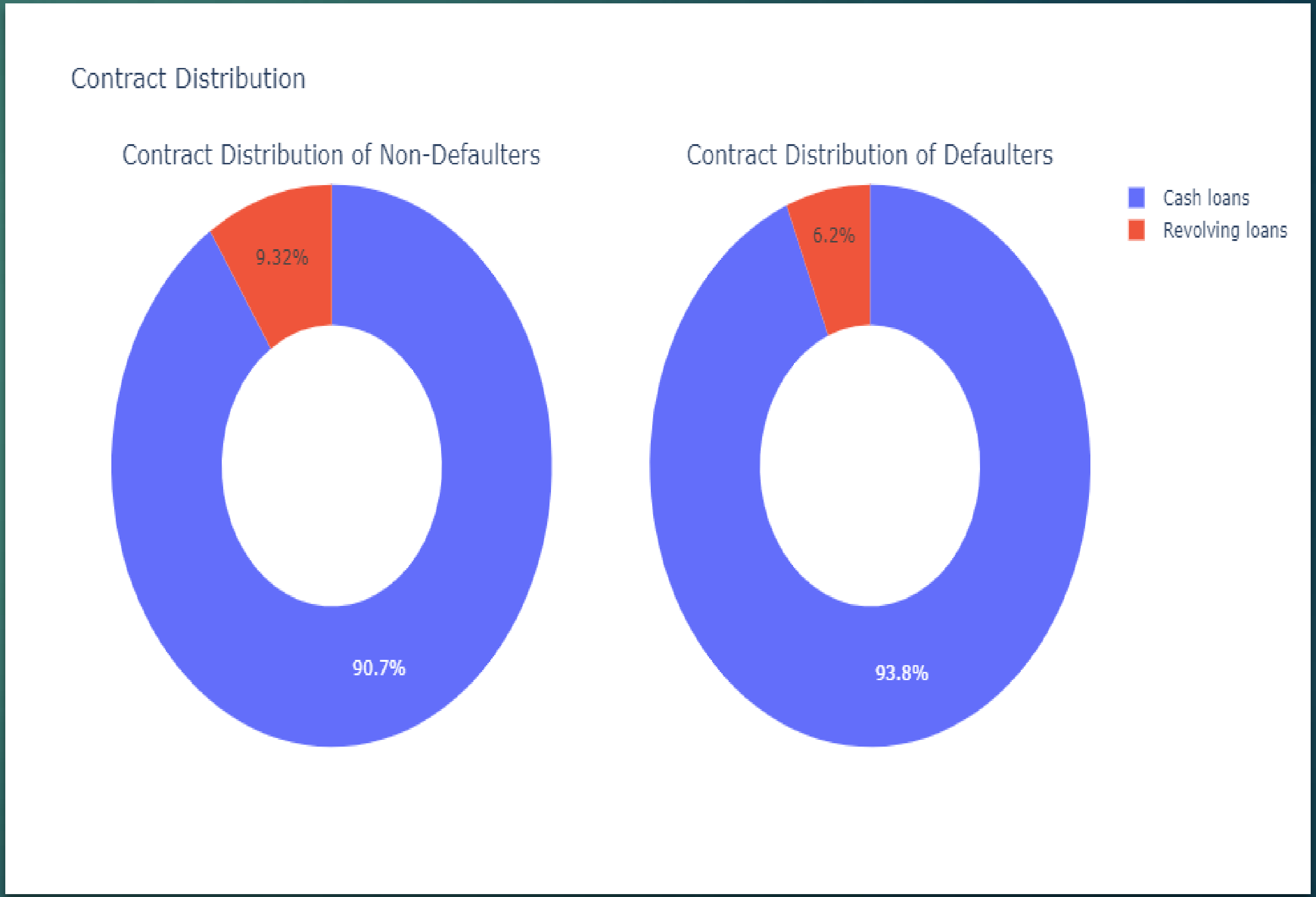
Distribution of Target Variable

We observe that there is a heavy imbalance in the Defaulter and Non-Defaulter data

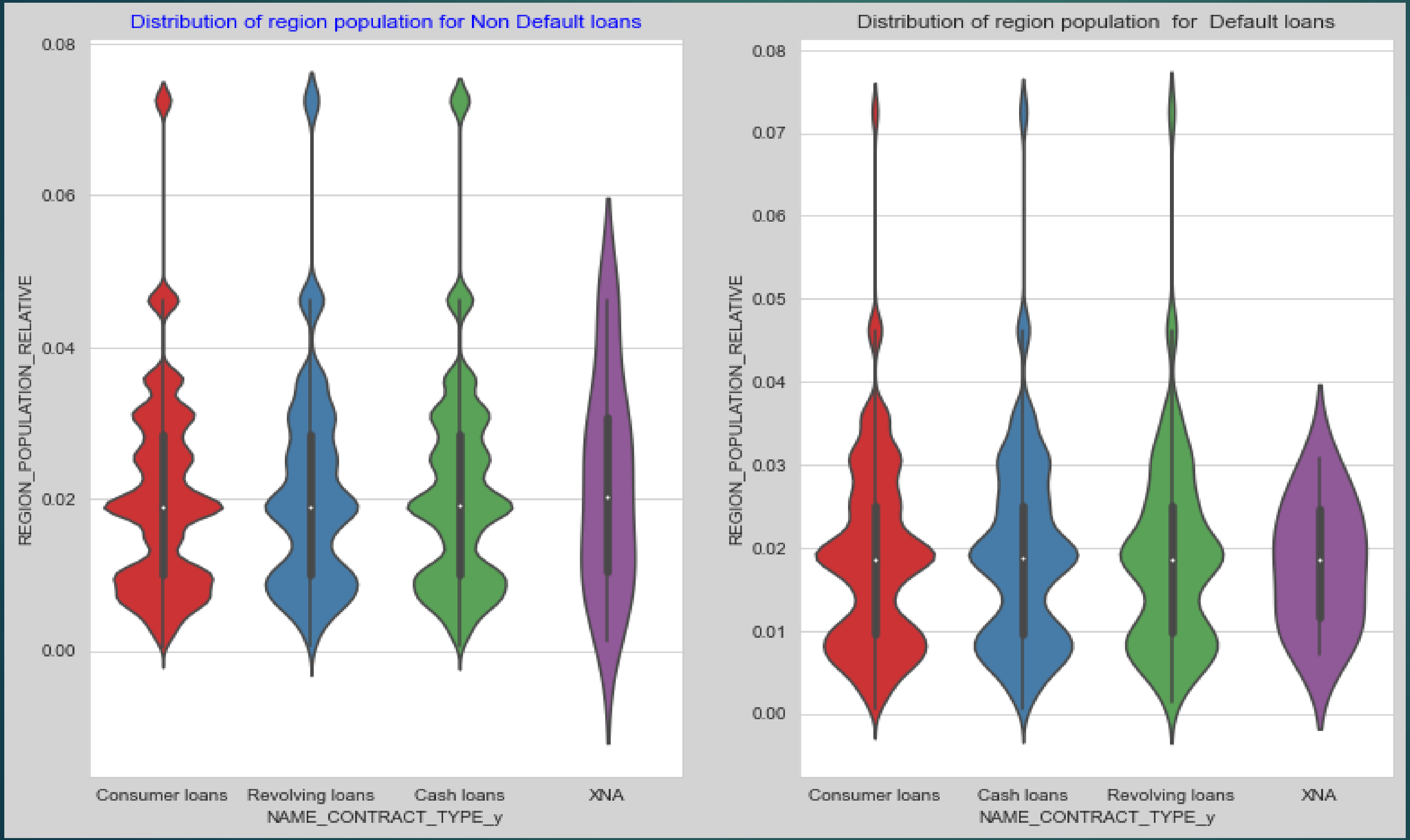


Does Contract Type effect loan payment?

Yes, we observe that people with Revolving loans are less likely to default as compared to Cash loans.



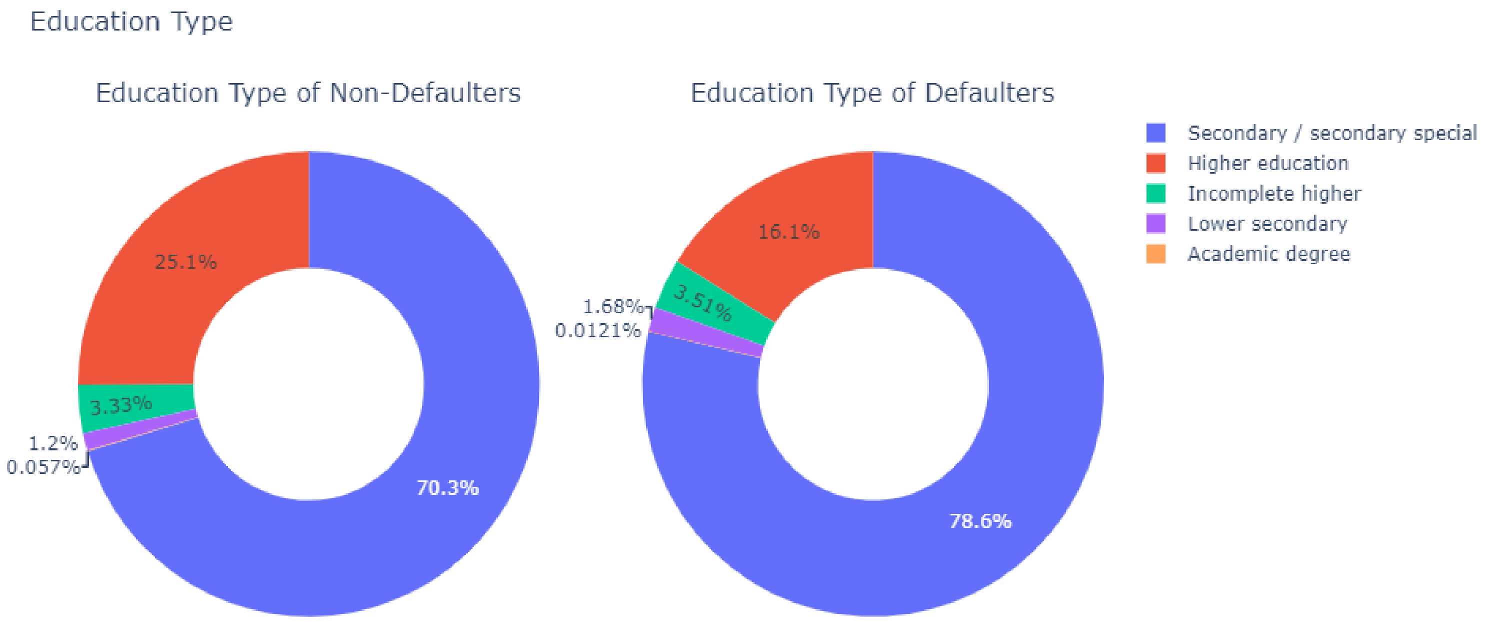
Does distribution of region population affect loan payment?



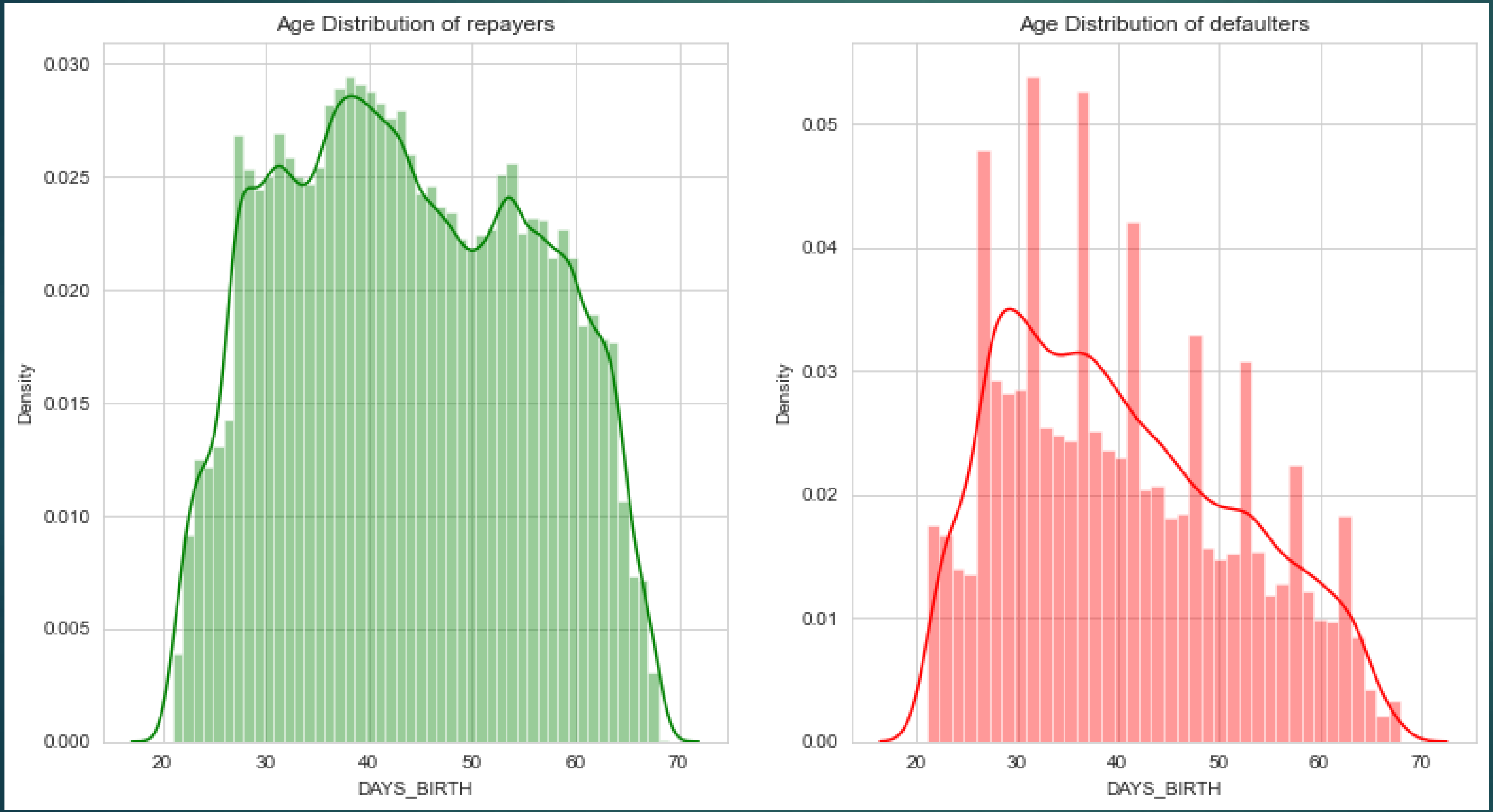
We see that clients in regions with High population density are less likely to default on loans.

What about Education Type?

- We observe that individuals with Secondary higher education are 8% more likely to default
- Clients with Higher education have lower number of defaulters
- Incomplete higher education and Lower secondary are slightly more in the defaulters list



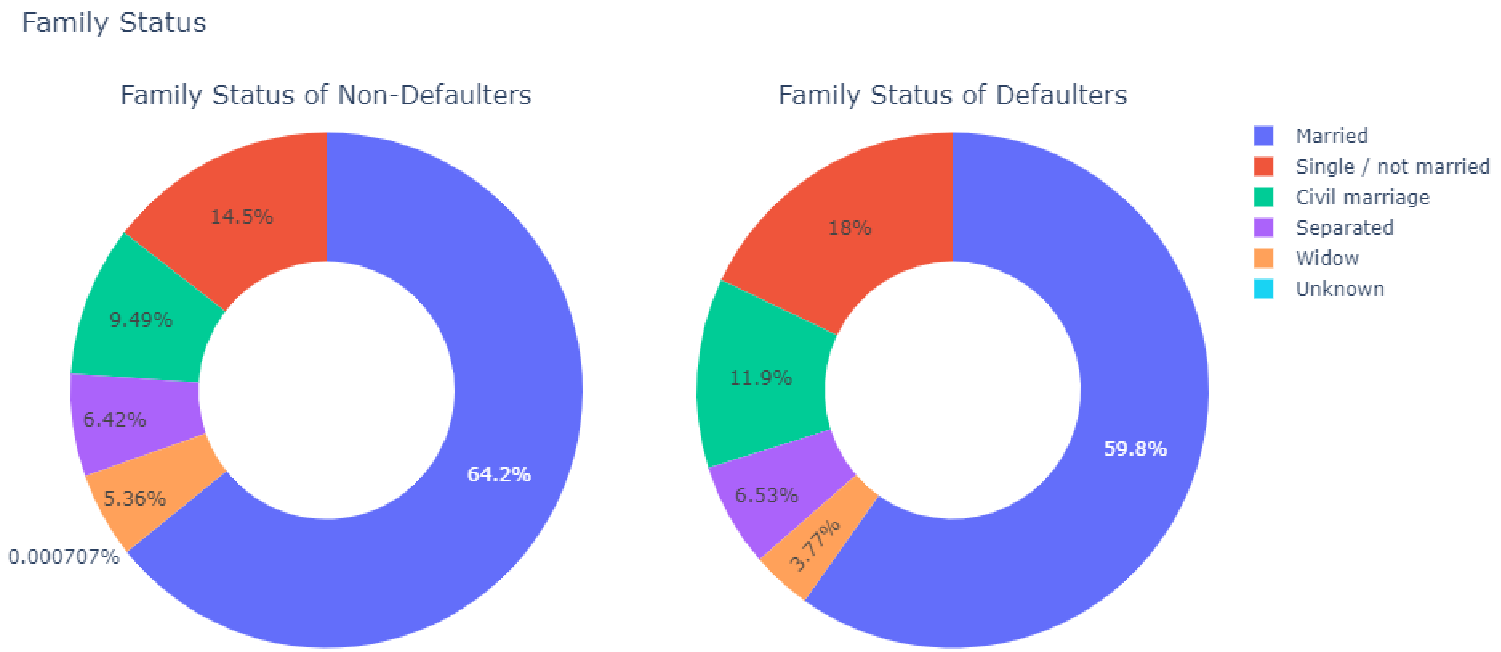
Does Age effect loan payment?

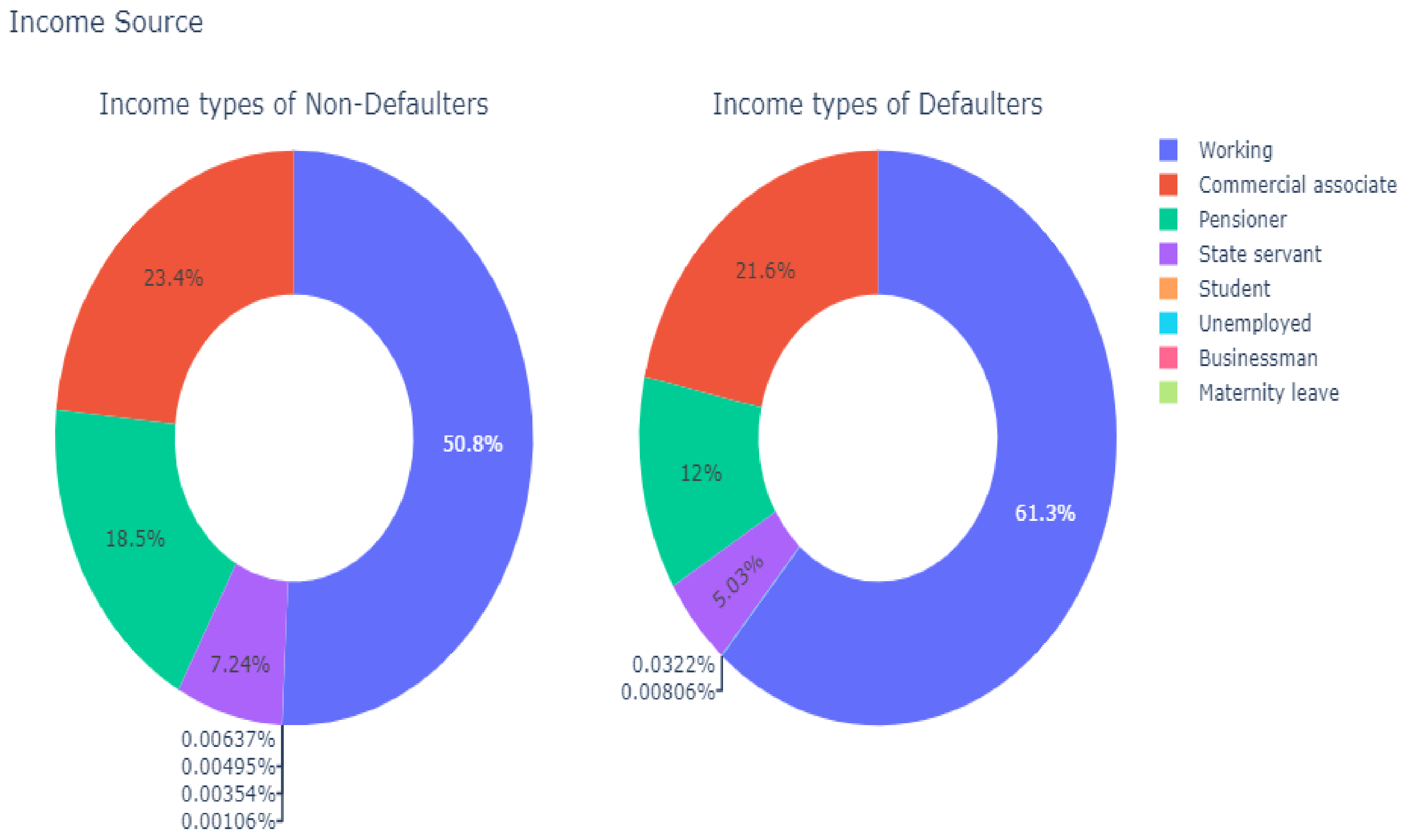


We observed that younger people tend to default more than older people.

Let's compare Family Status

- We note that Married individuals have a lower default %.
- Single, Civil marriage and Seperated clients seem to have more payment difficulties than Widows



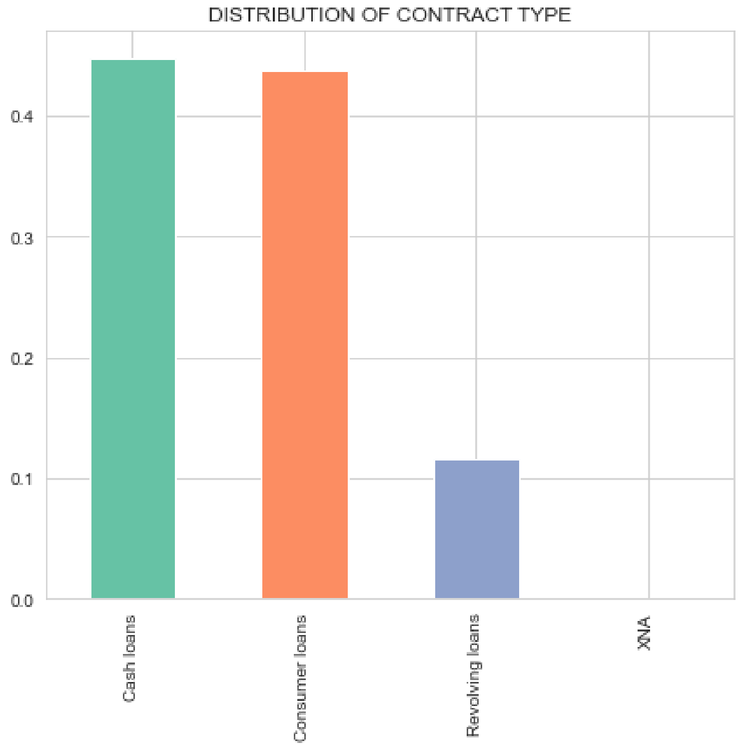


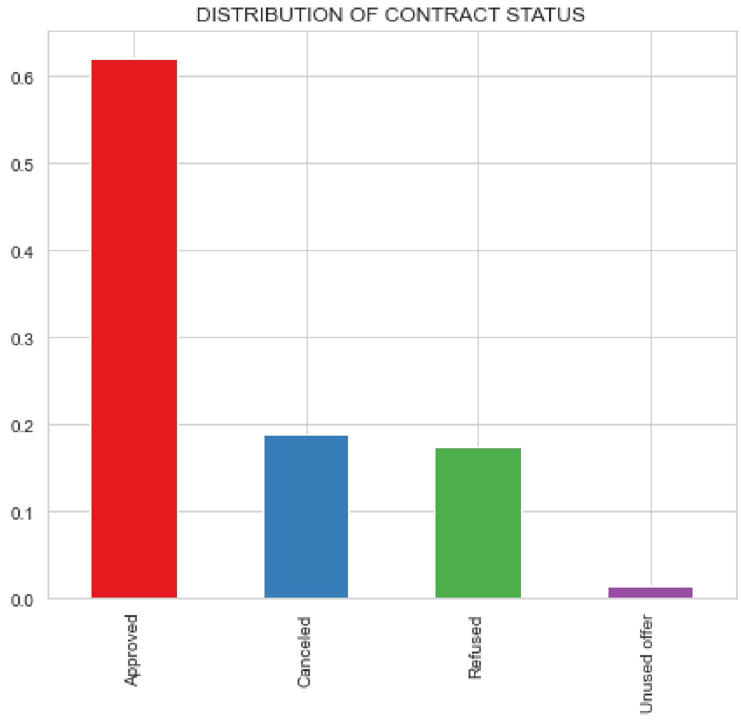
Income Sources

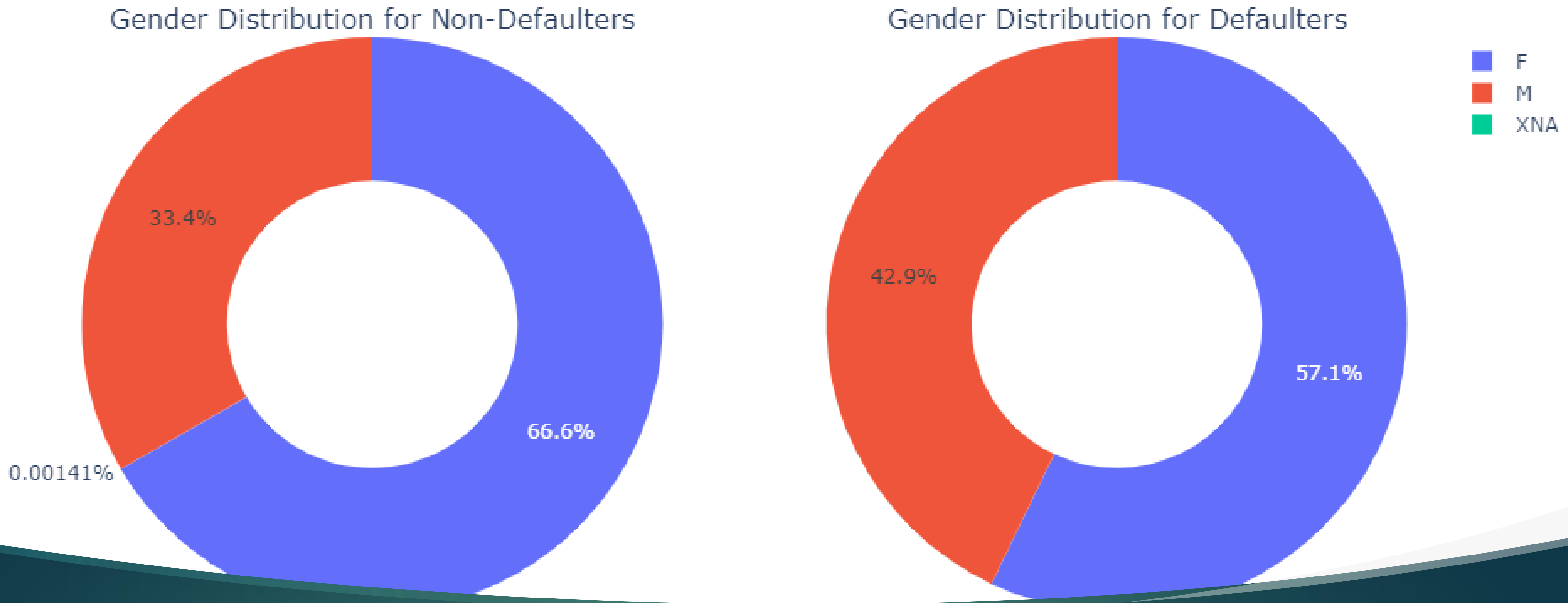
- Here we observe that Commercial Associate, Pensioner, State Servant have less % in the defaulter list. They are less likely to default than Working individuals (classed under 'Labourers' in the data-set).
- Unemployed individuals are the heaviest defaulters.
- Students and Businessmen tend to pay back the amount on time, they have no defaulters.

Contract Type vs Contract Status

- We note that most of the clients opt for Cash loans, followed closely by Consumer loans.
- It also looks like most loans are approved than cancelled or refused.
- 'Unused offer' is comparatively minimal.







Does Gender affect loan repayment?

In the above plot we observe that the number of male defaulters are more than females. Men are more likely to default by 9%

Key take-aways and Conclusion

- All pre-processing steps were performed
- We have covered Null value imputation methods
- Analyzing techniques such as Univariate analysis, Bivariate analysis, Models, were used
- Suggestion: we could look deeper into the Payment plan types arrive at the EMI using a better formula.