

# EDA Credit Assignment Case Study

---

NISHANT KUMAR MUKHI

DS C40

# Problem Statement

---

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. EDA is required to analyse the patterns present in data to ensure that applicants capable of repaying the loan are not rejected.

# Analysis Approach

---

I have taken below mentioned approach while doing this analysis

1. Imported given data set and started understanding the data using data dictionary
2. Understood the structure of data
3. Performed data quality checks like Missing data
4. Dropped the columns with more than 45% of missing values
5. Dropped some extra columns which may not have helped in analysis
6. Imputed missing values appropriately in the columns where missing values were less than 13%
7. Checked data type of all columns
8. Updated data types of columns which were incorrect
9. Updated negative data into positive data
10. Removed data rows which were logically incorrect like Employment tenure more than age of the client

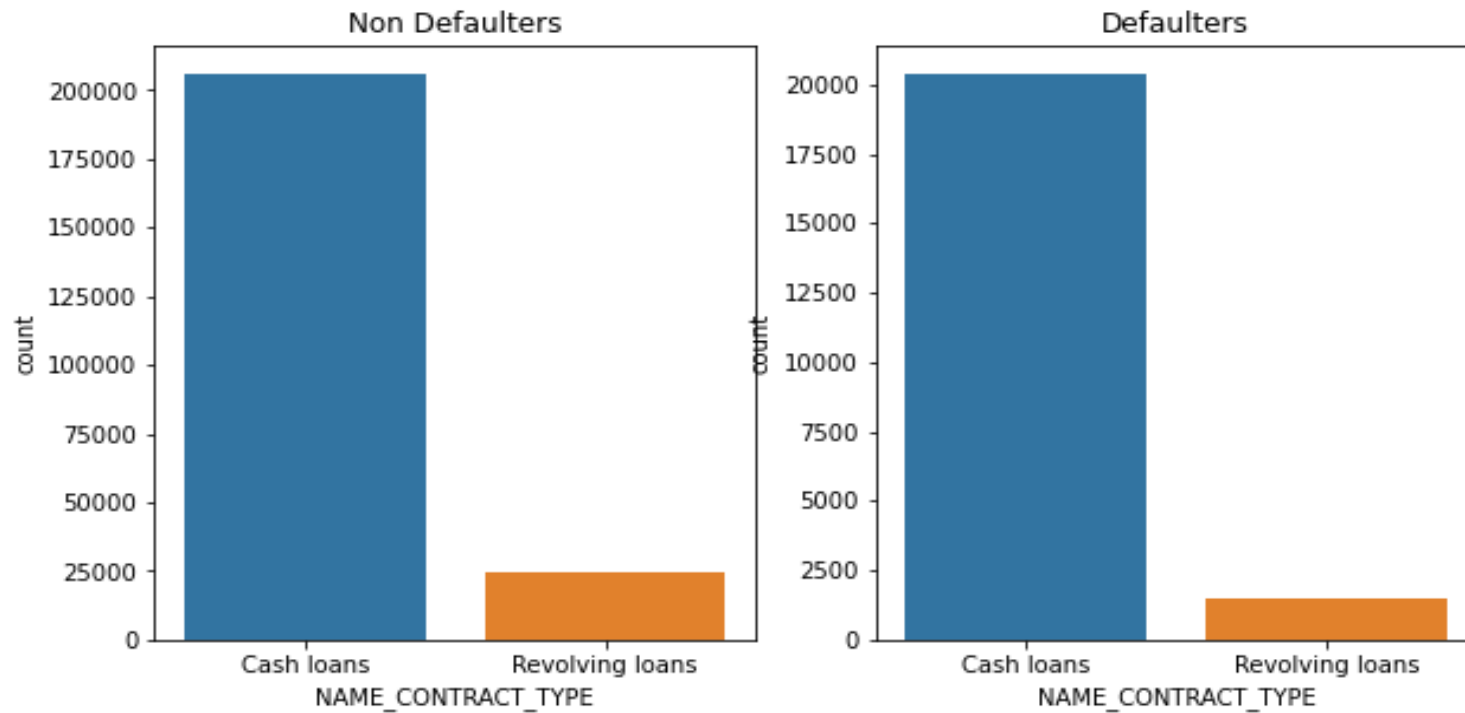
# Analysis Approach Contd..

---

- 11. Checked the outliers and reported them with reasoning
- 12. Performed Binning of continuous variables
- 13. Checked imbalance percentage of target variable
- 14. Divided data sets basis on target variable
- 15. Performed Univariate & Bivariate analysis on both data sets
- 16. Concluded basis on above mentioned analysis

# Univariate Analysis - NAME\_CONTRACT\_TYPE

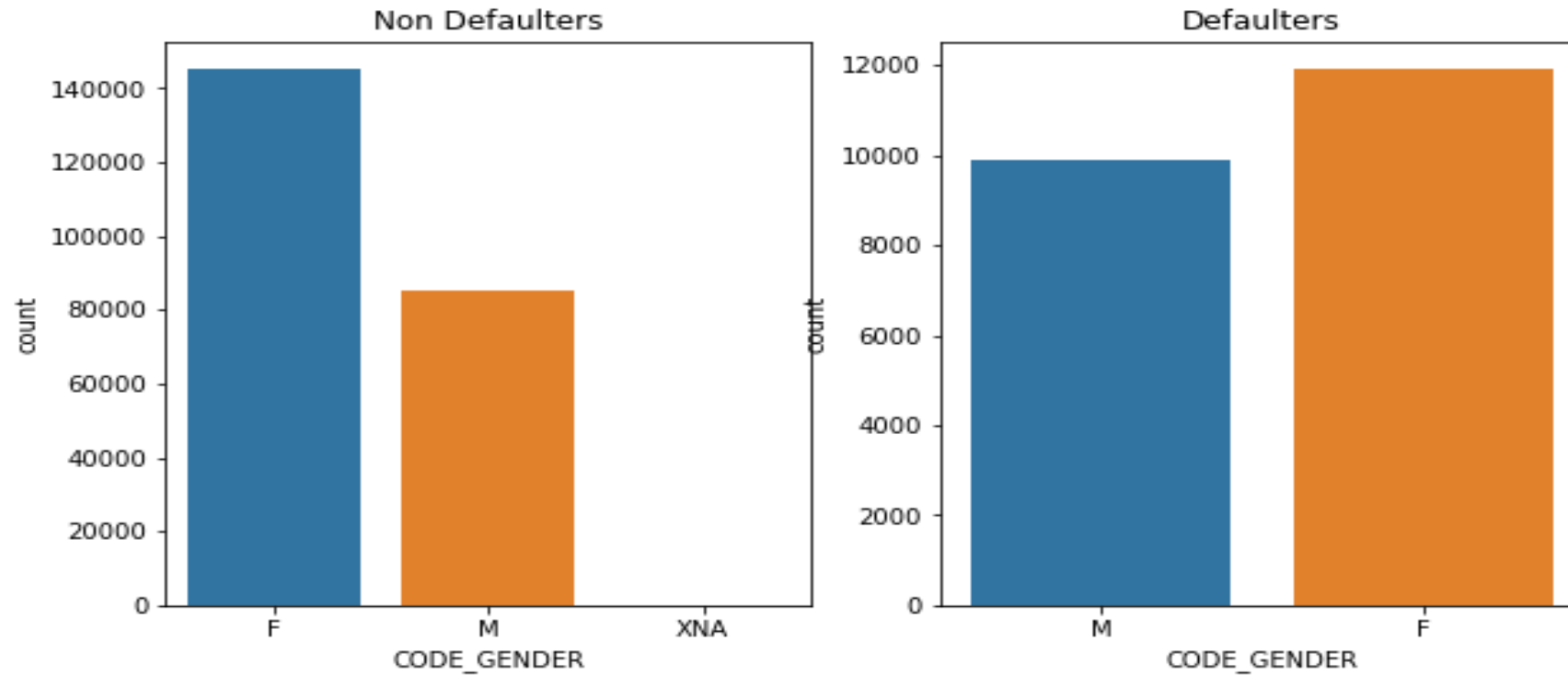
---



---

Observation - Ratio of Revolving loans over Cash loans looks almost same in both defaulters and non defaulters case so nothing conclusive from this variable

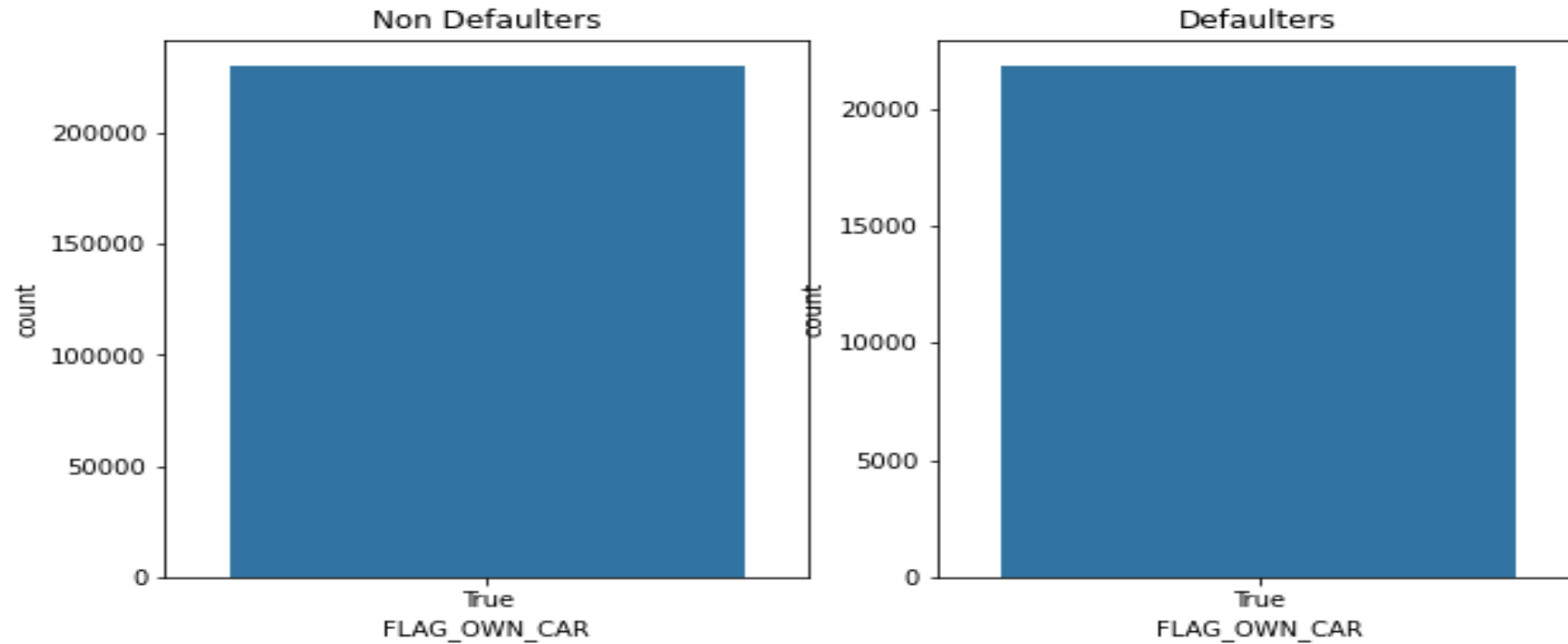
# Univariate Analysis – CODE\_GENDER



Observation - Gender does not give any clear conclusion as Females are high in both defaulters and non-defaulters, however percentage of female non defaulters is higher than female defaulters though it is reverse in case of men.

# Univariate Analysis - FLAG\_OWN\_CAR

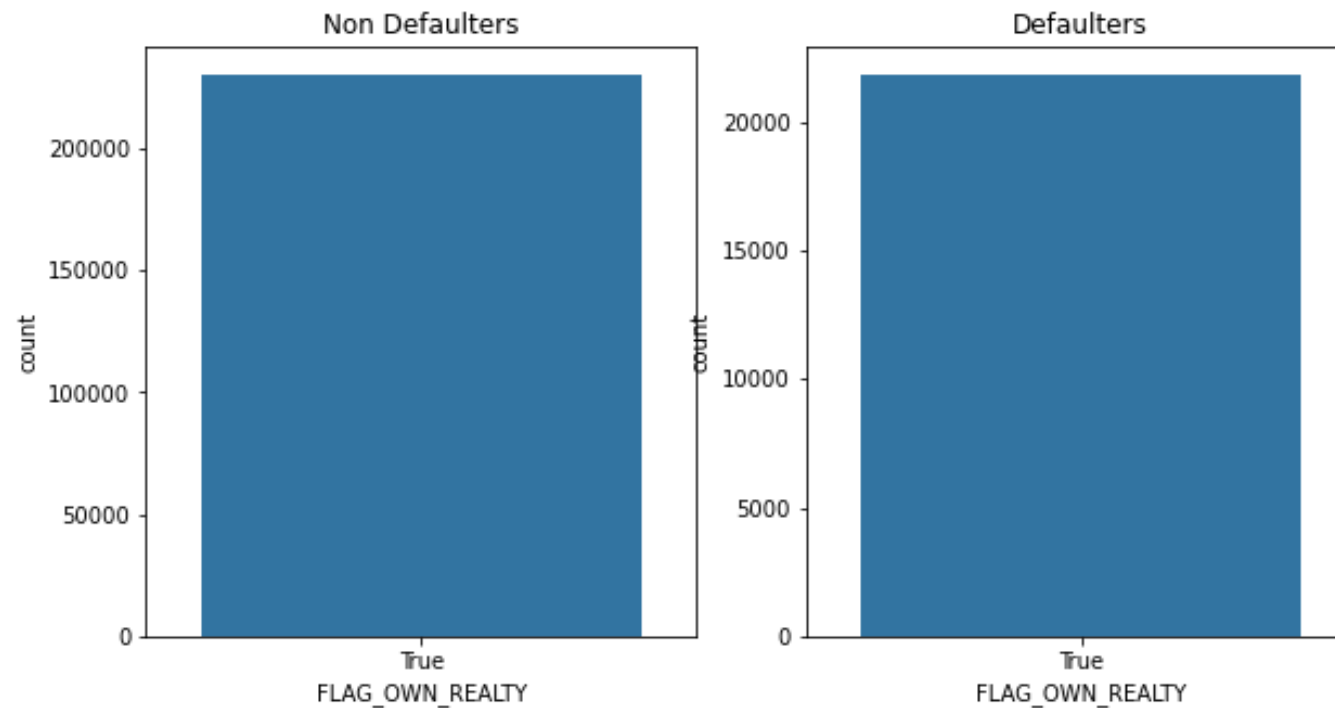
---



Observation - This analysis does not give any conclusion as all defaulters and non-defaulters owns car

# Univariate Analysis - FLAG\_OWN\_REALTY

---

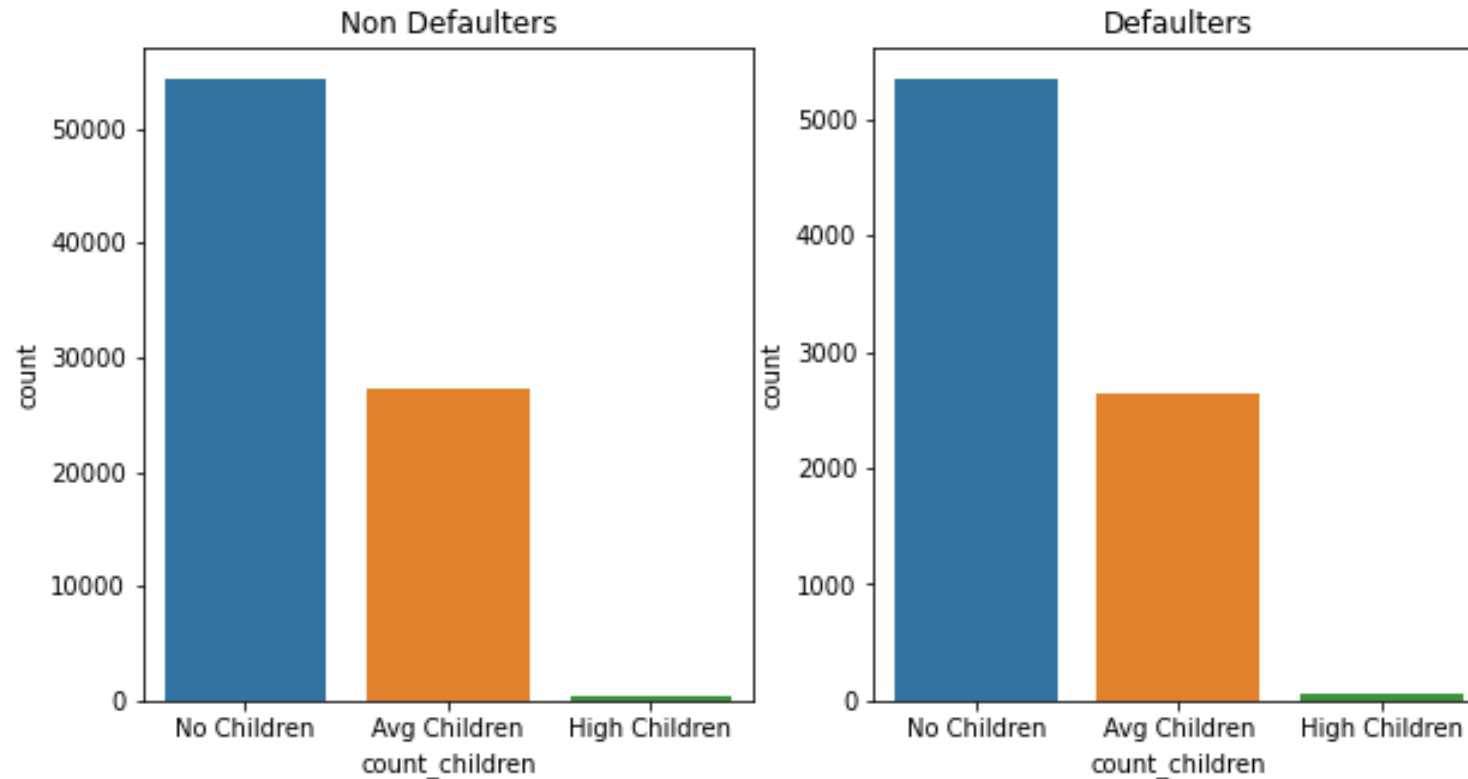


Observation - This analysis does not give any conclusion as all defaulters and non-defaulters own house/flat



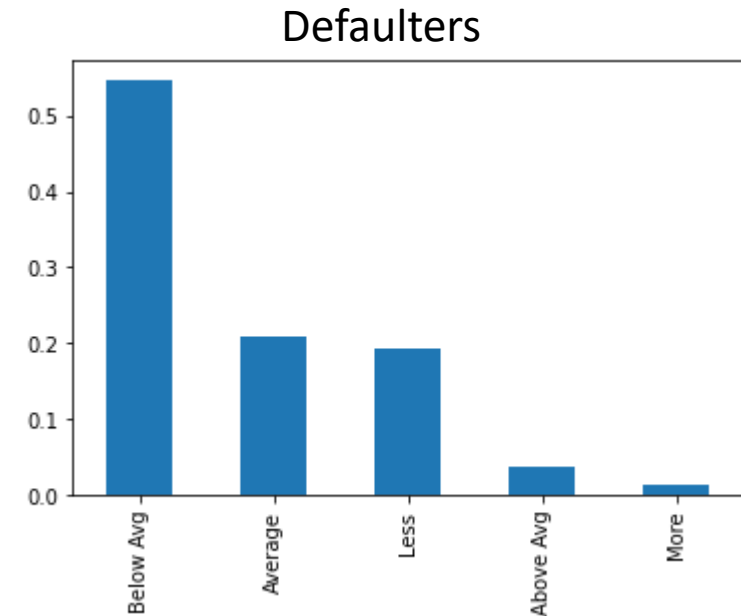
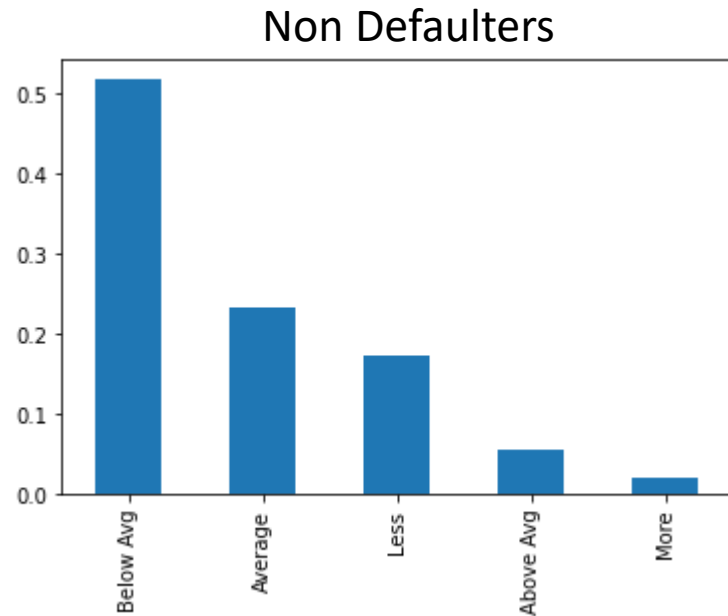
# Univariate Analysis - CNT\_CHILDREN

---



Observation - Number of children is not making any difference in person defaulting the loan

# Univariate Analysis - AMT\_INCOME\_TOTAL

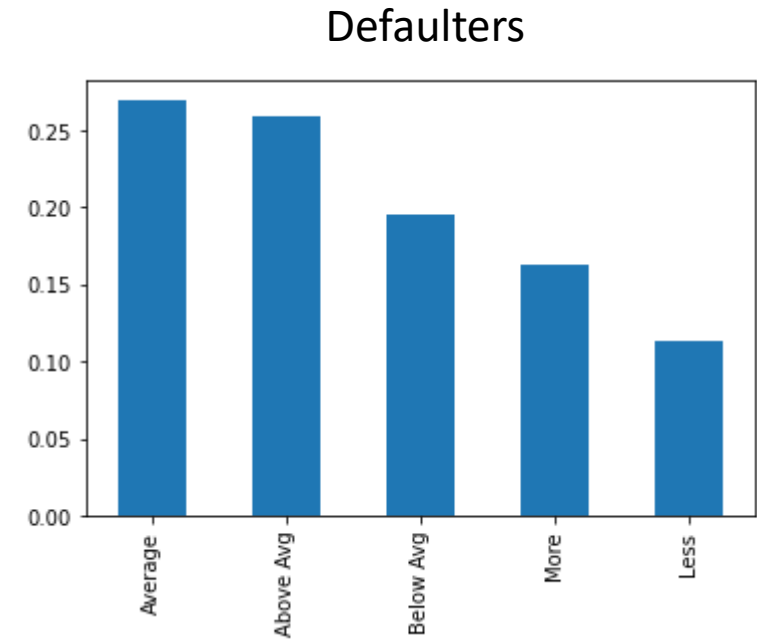
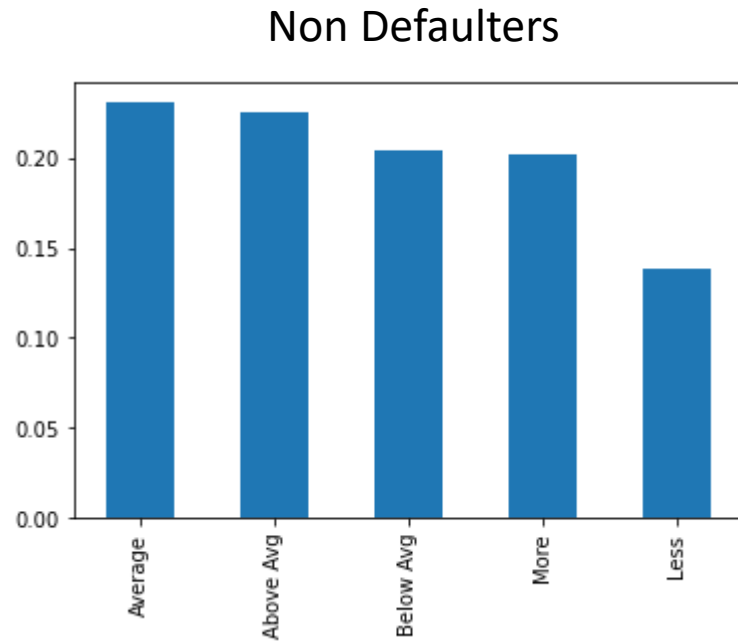


Observation - Approx. 74% people in defaulters have less/Below Average salaries whereas Approx. 69% people in Non-defaulters have less/Below Average salaries.

Almost similar percentage so it is inconclusive that salary has any role to play

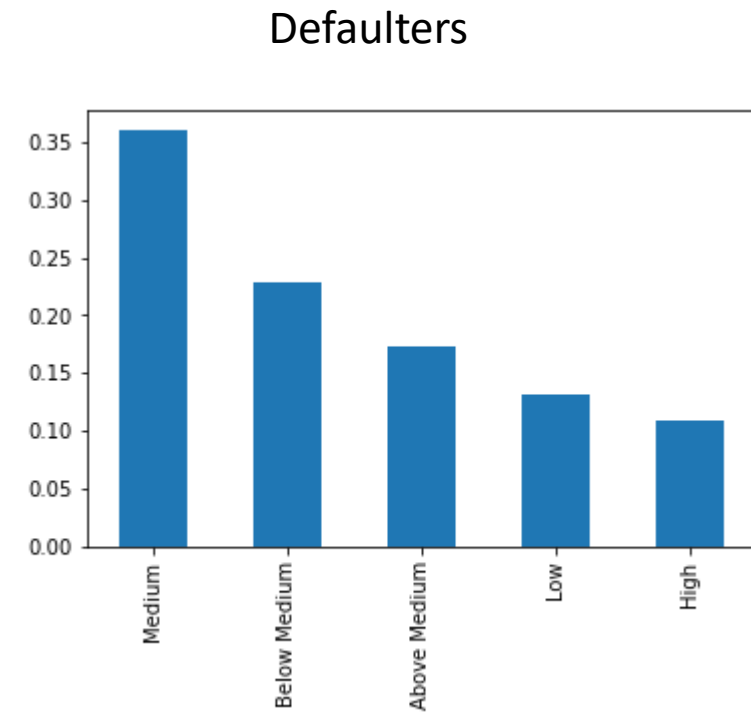
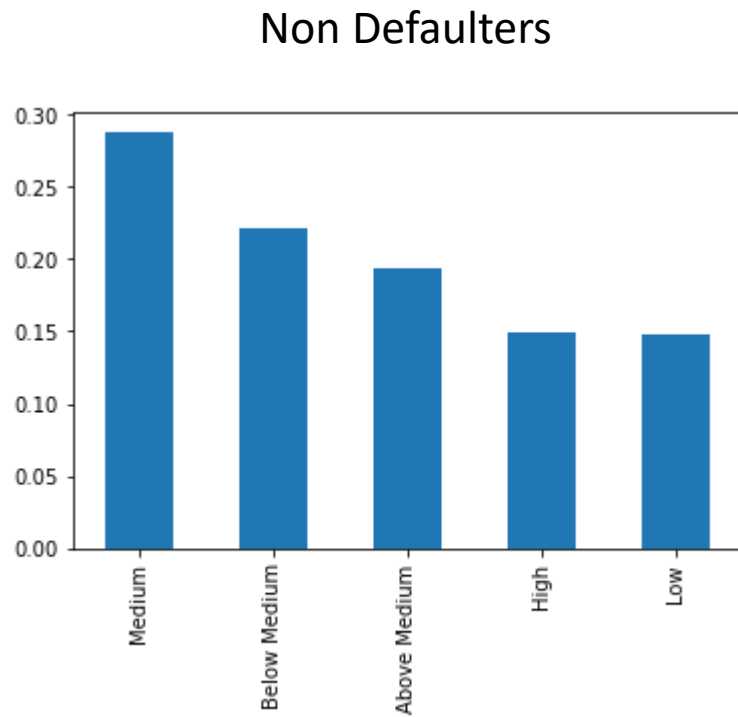
# Univariate Analysis - AMT\_CREDIT

---



Observation - There is not much difference w.r.t Loan amount between two data sets so nothing conclusive from this variable

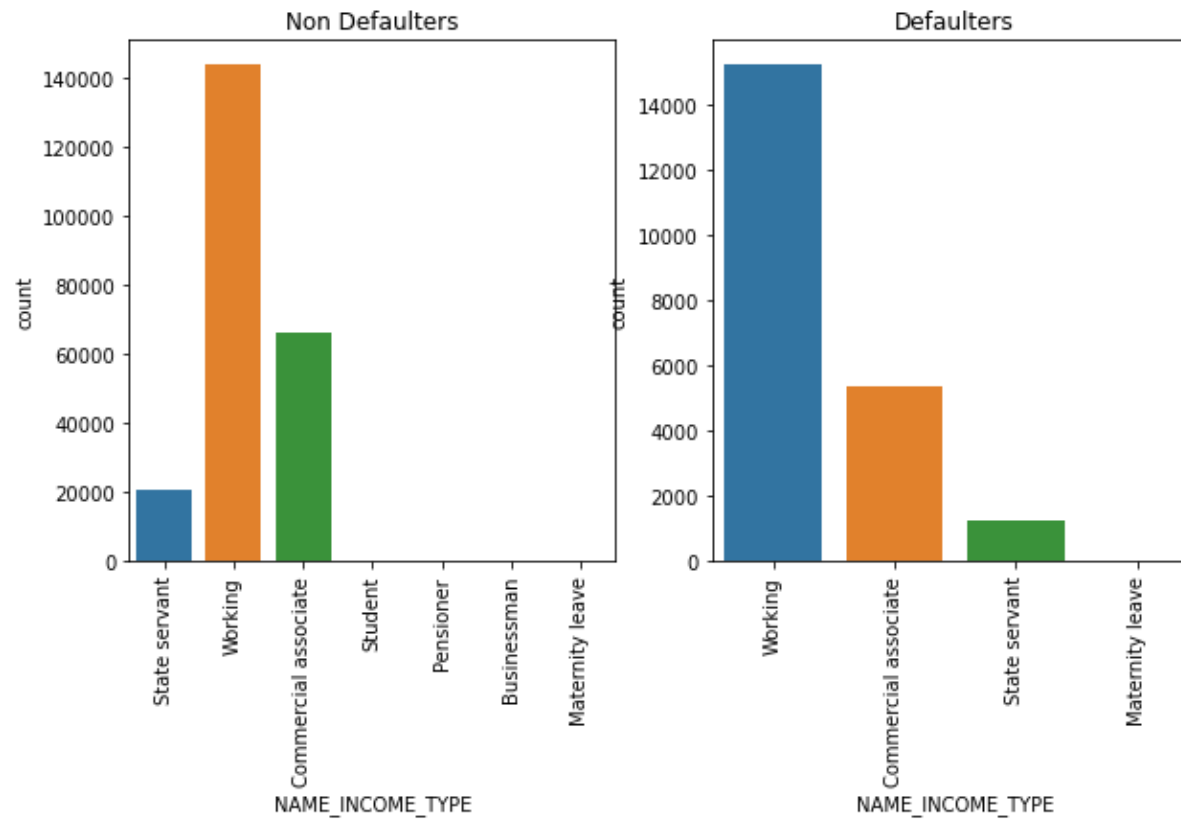
# Univariate Analysis - AMT\_GOODS\_PRICE



Observation - Most of the people (in both data sets) have bought goods of 3 to 5 Lacs for which they have taken loan

# Univariate Analysis - NAME\_INCOME\_TYPE

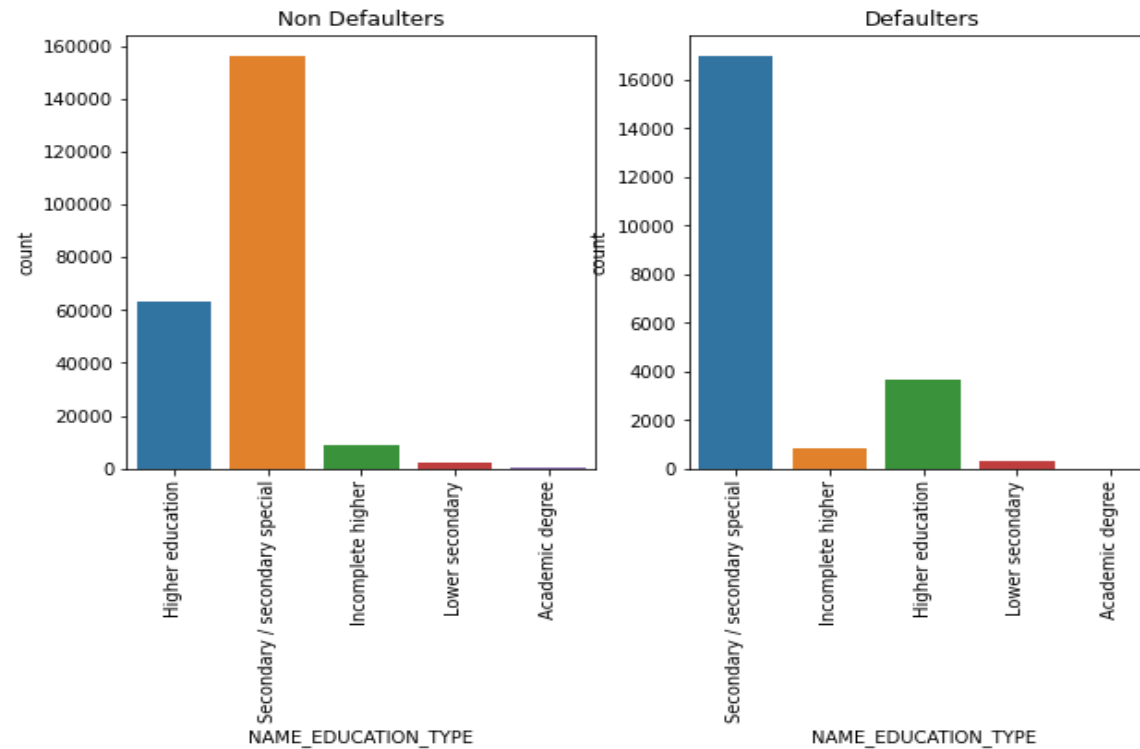
---



Observation - Same pattern is observed in both data sets w.r.t Income type

# Univariate Analysis - NAME\_EDUCATION\_TYPE

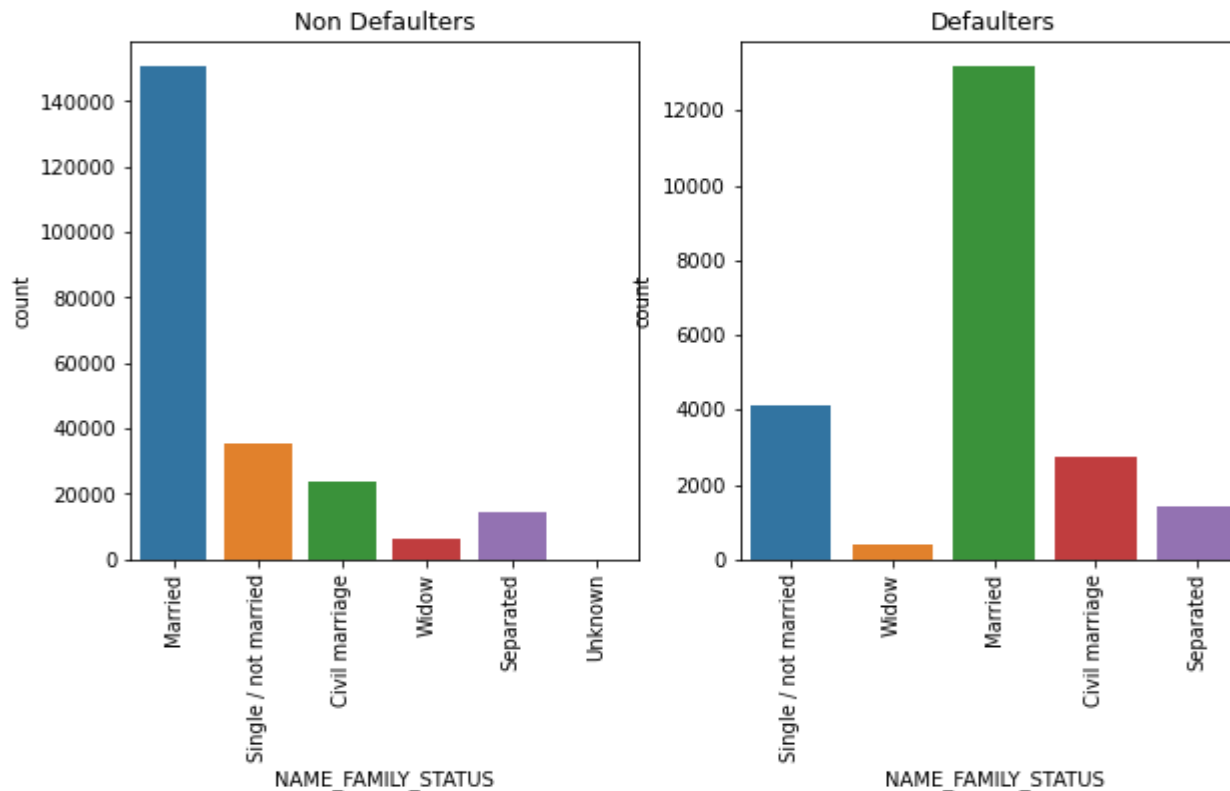
---



Observation - Same pattern is observed in both data sets w.r.t Education type

# Univariate Analysis - NAME\_FAMILY\_STATUS

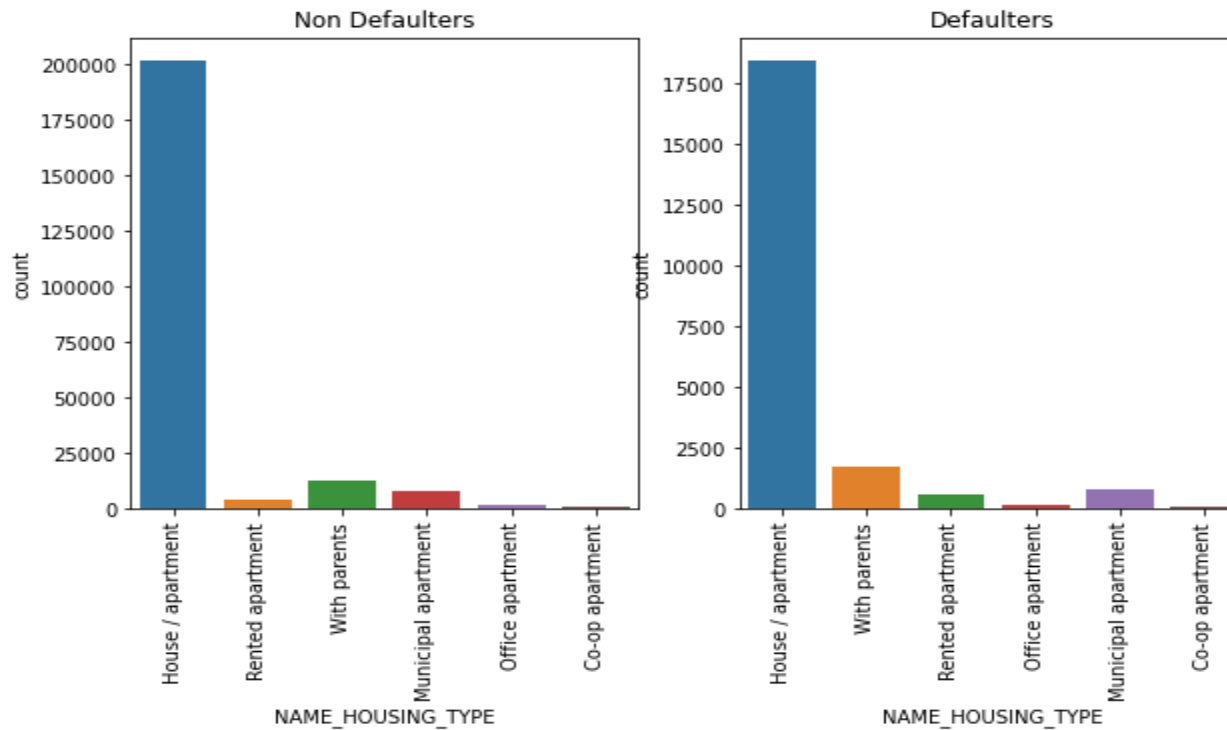
---



Observation - Same pattern is observed in both data sets w.r.t Family Status

# Univariate Analysis - NAME\_HOUSING\_TYPE

---

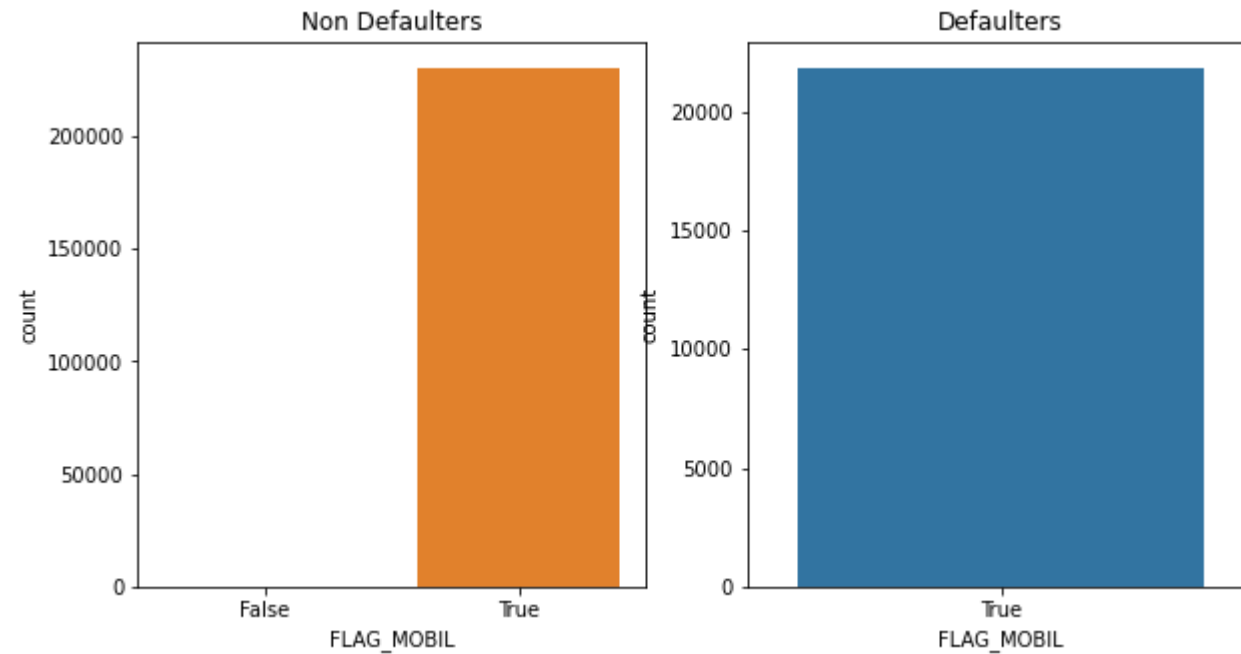


Observation - Same pattern is observed in both data sets w.r.t Housing Type



# Univariate Analysis - FLAG\_MOBIL

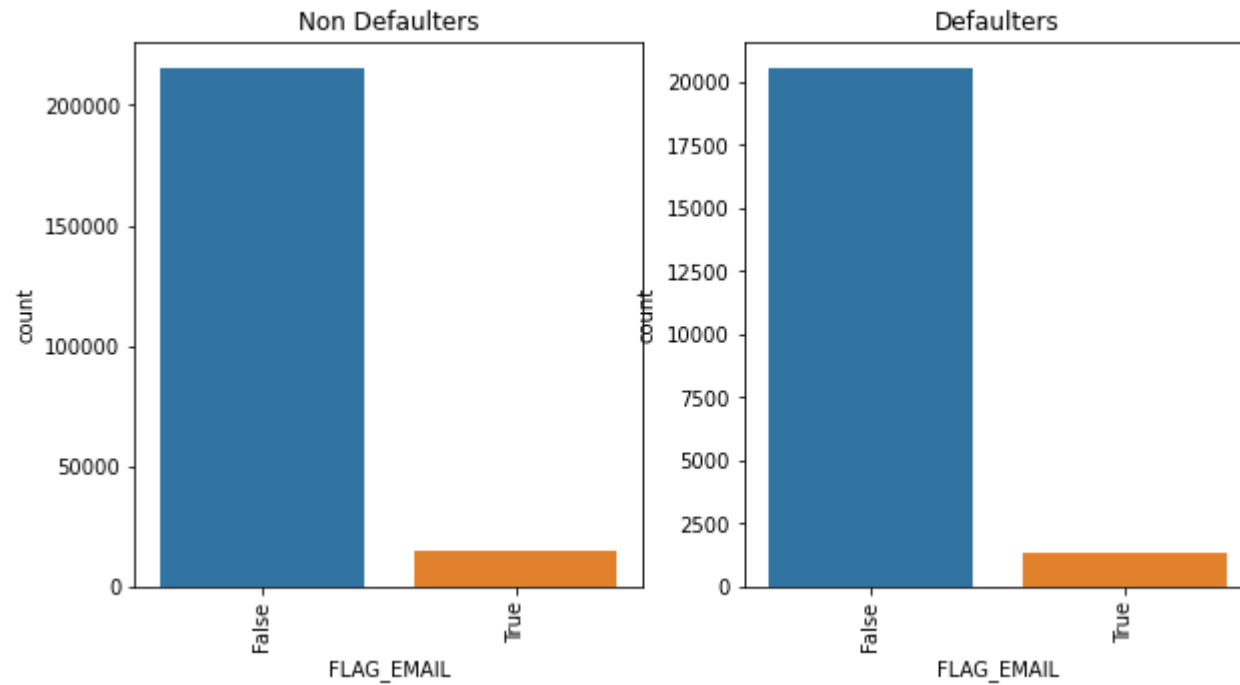
---



Observation - All defaulters provided their mobile phone however there is a minimal set in Non defaulters who didnt provide mobile phones

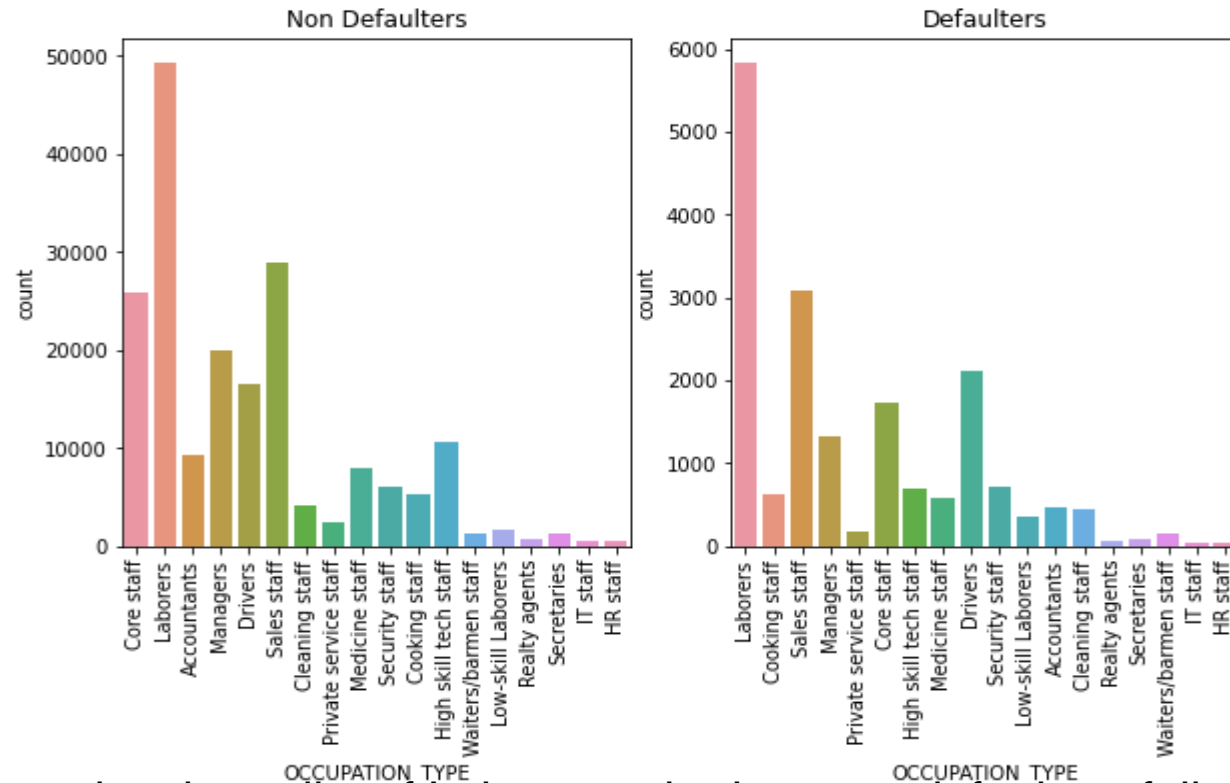
# Univariate Analysis - FLAG\_EMAIL

---



Observation - Same pattern is observed in both data sets w.r.t providing Email details

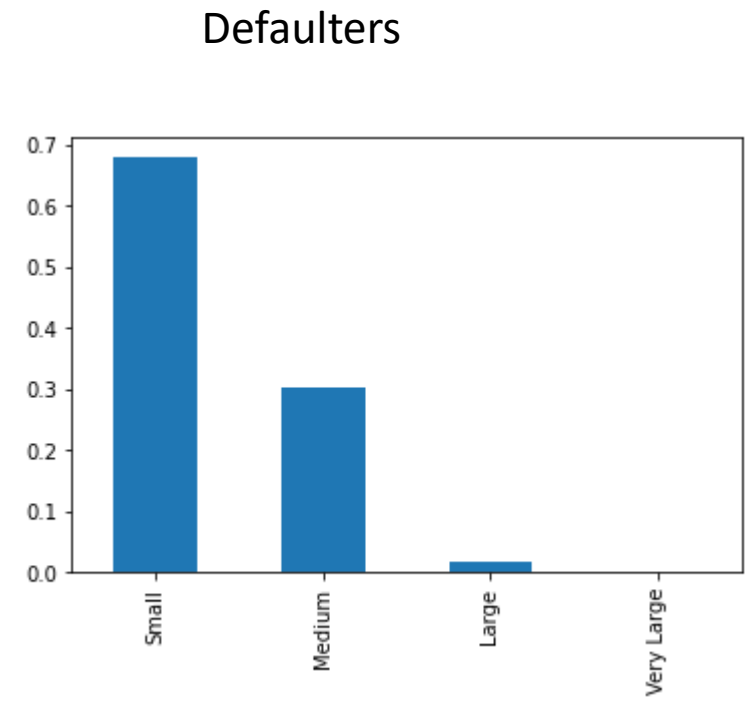
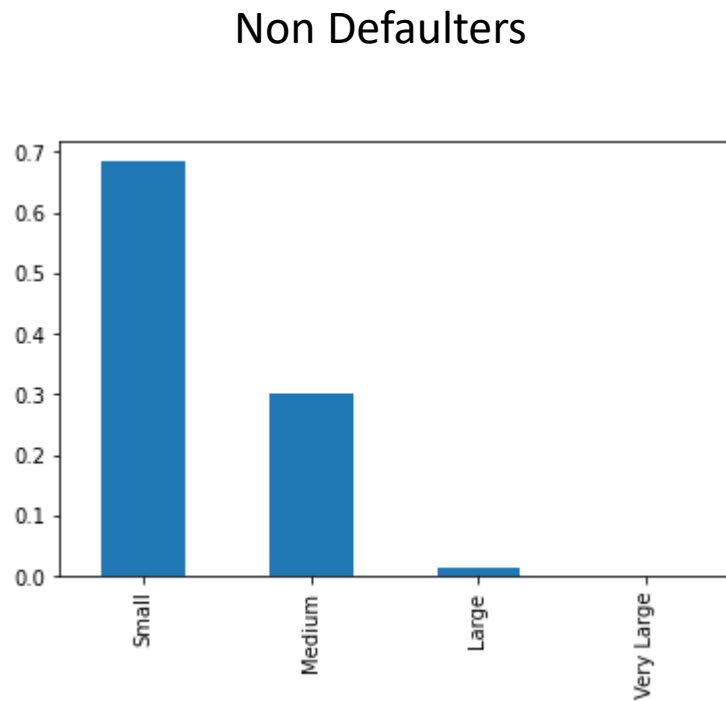
# Univariate Analysis - OCCUPATION\_TYPE



Observation - There is a huge list of Laborers who became defaulters followed by Sales Staff people and then Drivers. However, Laborers are the ones who became less Non defaulters as well followed by Sales staff and then Core staff members

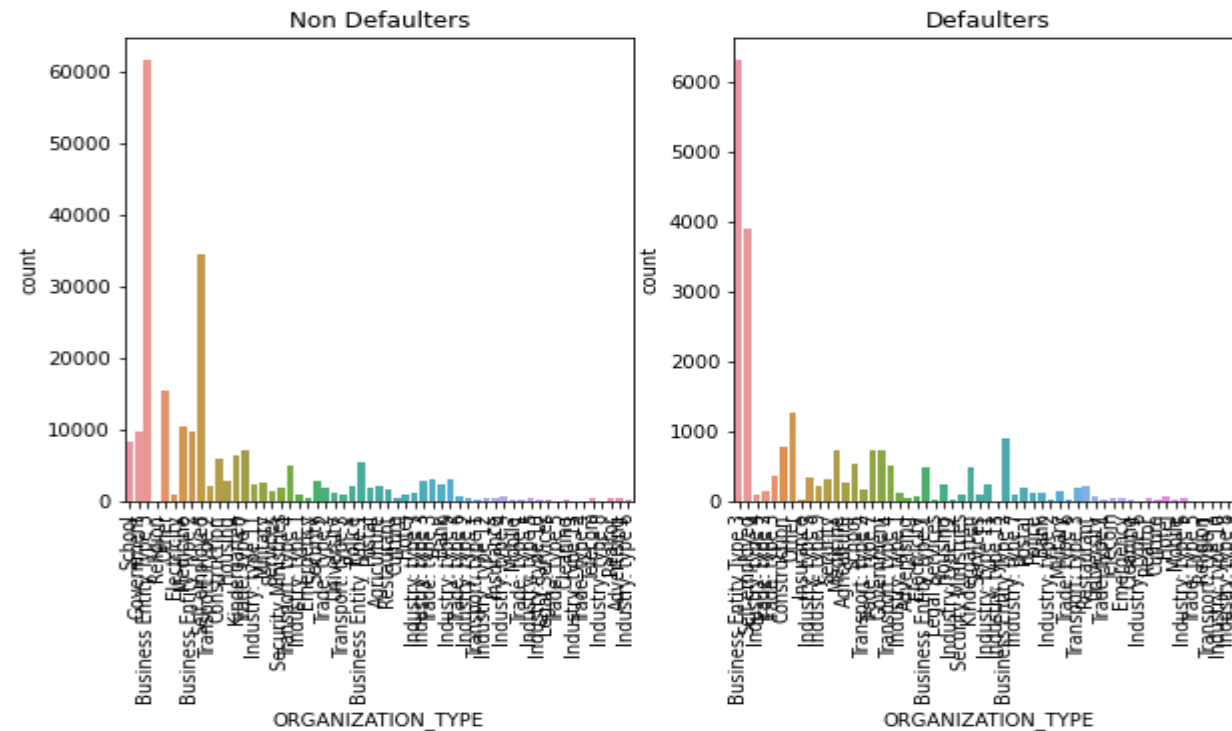
# Univariate Analysis - CNT\_FAM\_MEMBERS

---



Observation - Same pattern is observed in both data sets w.r.t count of family members

# Univariate Analysis - ORGANIZATION\_TYPE

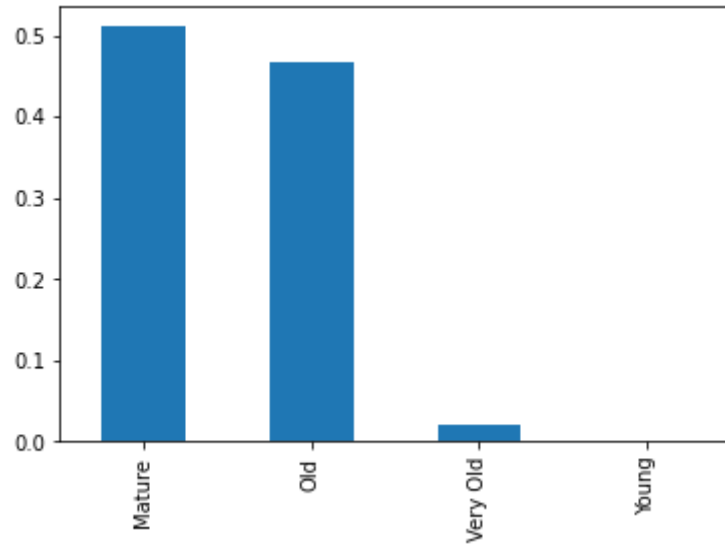


Observation - Not much conclusive due to various org types

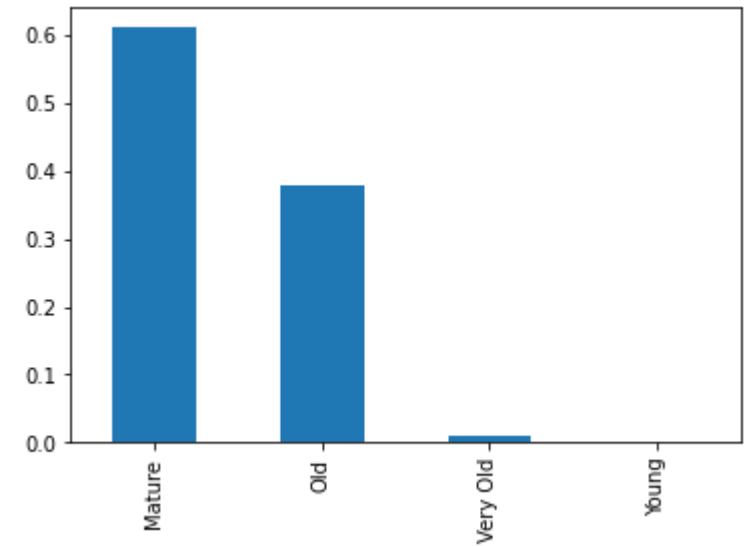
# Univariate Analysis - Client\_Age

---

Non Defaulters



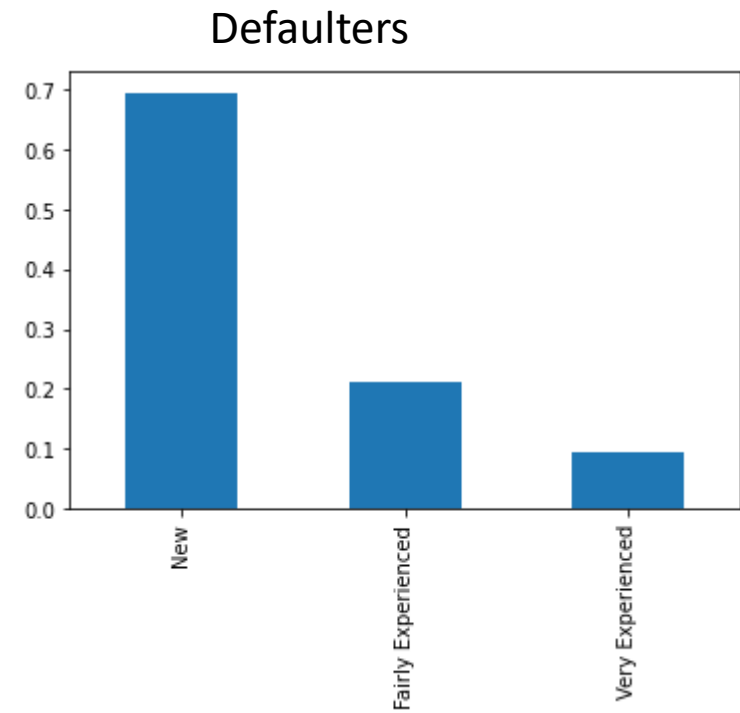
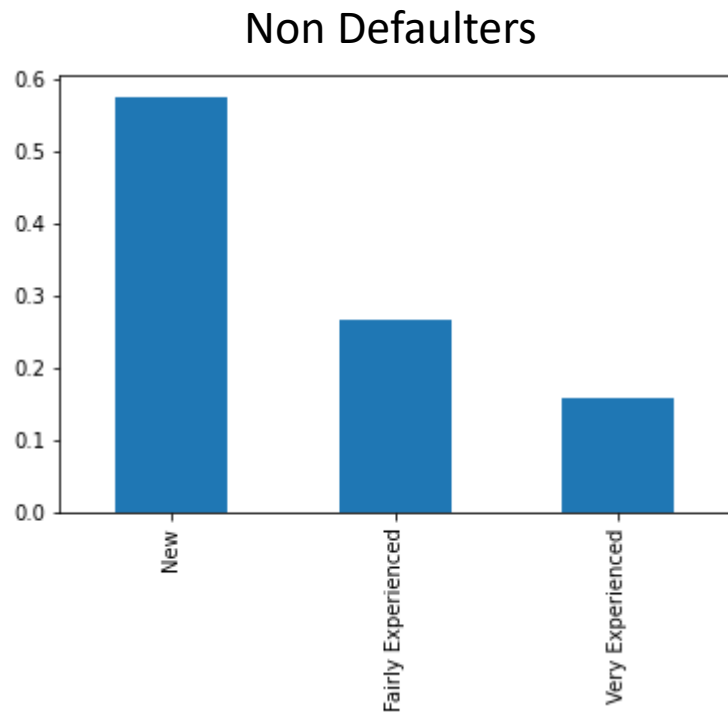
Defaulters



Observation - Same pattern is observed in both data sets w.r.t Client Age

# Univariate Analysis - Employment\_Tenure

---



Observation - Same pattern is observed in both data sets w.r.t Employment Tenure

# Bivariate Analysis - AMT\_INCOME\_TOTAL VS AMT\_CREDIT



Observation - There is huge chunk of people who has taken more loan than their total income. This looks to be one of the primary reasons of them becoming defaulters



# Correlation - AMT\_INCOME\_TOTAL, AMT\_CREDIT , AMT\_GOODS\_PRICE

---

Non Defaulters

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_GOODS_PRICE
AMT_INCOME_TOTAL	1.000000	0.326155	0.333292
AMT_CREDIT	0.326155	1.000000	0.986471
AMT_GOODS_PRICE	0.333292	0.986471	1.000000

Defaulters

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_GOODS_PRICE
AMT_INCOME_TOTAL	1.000000	0.036484	0.036105
AMT_CREDIT	0.036484	1.000000	0.982464
AMT_GOODS_PRICE	0.036105	0.982464	1.000000

Observation - Defaulters have taken goods of more price and hence taken more loan w.r.t their income whereas Amount goods price and Amount credit is lesser in case of Non defaulter

# Previous Application Analysis – Observations

---

Observation 1 - 17% of applications who were previously rejected, given the loan this time and those people defaulted. This data shows that wrong assessment was done this time though previous assessment about them was correct.

Observation 2 - 61% of applications who were previously approved, approved this time as well and those people defaulted. This data shows wrong assessment done both last time and this time

Observation 3 - 17% of applications who were previously rejected, approved this time and those people didn't default. This data shows wrong assessment done last time but correctly assessed this time.

Observation 4 - 62% of applications who were previously approved, approved this time as well and those people didn't default. This data shows correct assessment done last time and this time as well.

# Conclusion

---

Careful assessment is required w.r.t Amount of Loan being taken and Applicant's Income as huge chunk of people defaulted who took more loan than their income