CREDIT EDA ASSIGNMENT

SUSIL PATRO (DATA SCIENCE JULY, 2022 BATCH – DS46)

DATE OF SUBMISSION: 25^{TH} SEP 2022

BUSINESS UNDERSTANDING

- There are two datasets provided (application data and previous application data) ; this can provide insights to loan process of consumer finance company.
- With help of exploratory data analysis on these two datasets, we can make inferences on the main cause of loan defaulters and accordingly we can provide suggestions to business decisions.
- These iinferences could help to increase business by tapping to customer segments where customers less likely to default. Also this will minimize risk by not doing business with customers who are more likely to default.

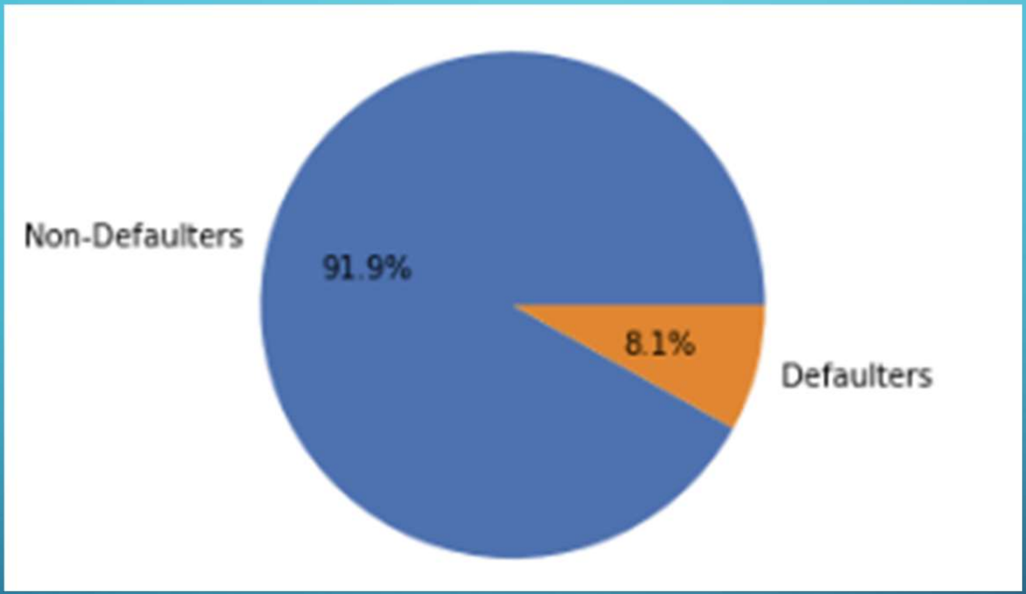
OBJECTIVE

- Assignment objective is to review applicants who are defaulters and also to know who less likely to default
- There could be two main approaches to make business decision with help of exploratory data analysis
  - To understand the parameters properly and target those potential customers who are less likely default. If client has capacity to repay the loan then not approving their loan will be loss of business
  - If such applicants who do not capacity to repay the loan and approving such loans will result to increase in risk as a result loss of revenue Some time to cover the risk, risk officers can make decision to amend interest rate.
  - To understand impact of loan history of applicants on current application.

STEPS FOR EXPLORATORY DATA ANALYSIS

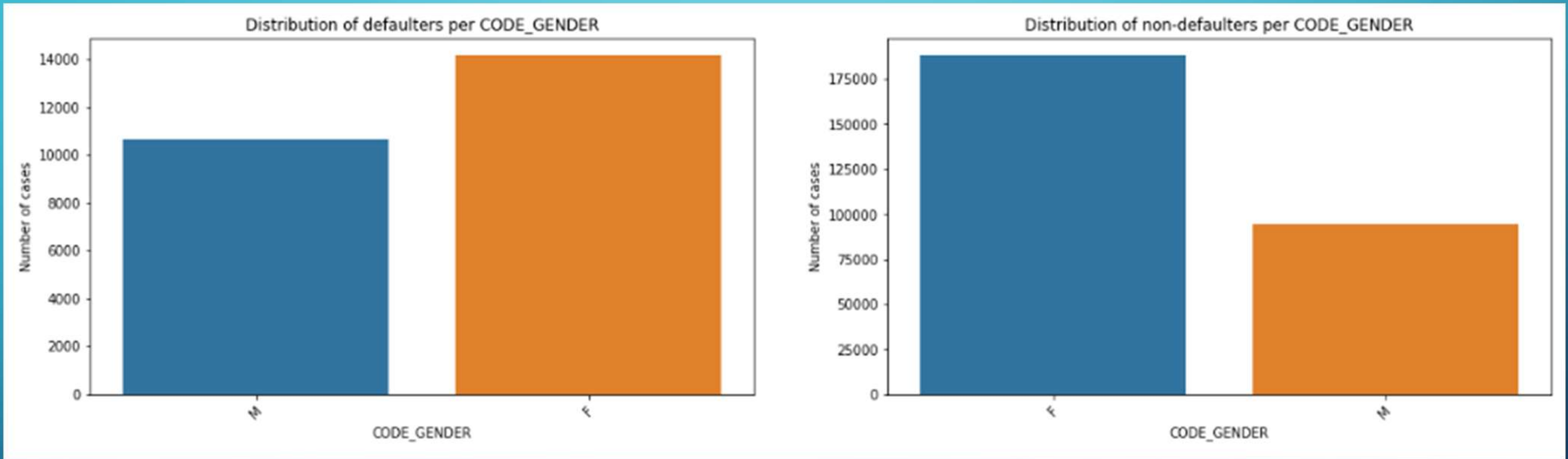
- Business understanding on problem statement
- Understand current application details and previous credit history (details of previous application)
- Import python libraries for data analysis, computation and plotting graph to perform analysis
- Read csv files for application data and previous application data into respective data frames
- Data Inspection (routine structured check)
  - Data Quality and missing values / null inspections
  - Remove columns which have more then 40 percent null values
  - Treatment for outliers (suggestions to impute the values)
  - Binning some numerical values to categorize them
- Merge the previous history dataset and repeat the same treatments and analysis with application data along with history information.
- Analysis
  - Univariate
  - bivariate and multivariate plots for analysis
- Make inferences which can be used for business decisions
- Final recommendation

DISTRIBUTION OF TARGET VARIABLES



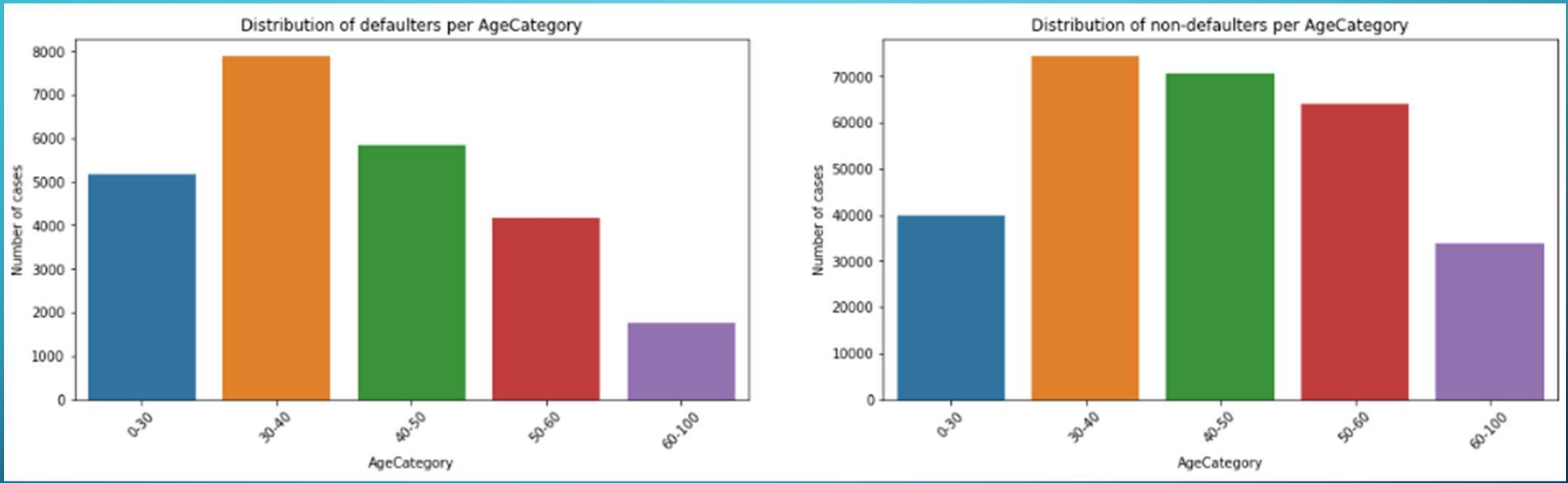
- **In current application 8.1% defaulters observed. We need to study this demography details of these population**

DEMOGRAPHY - GENDER



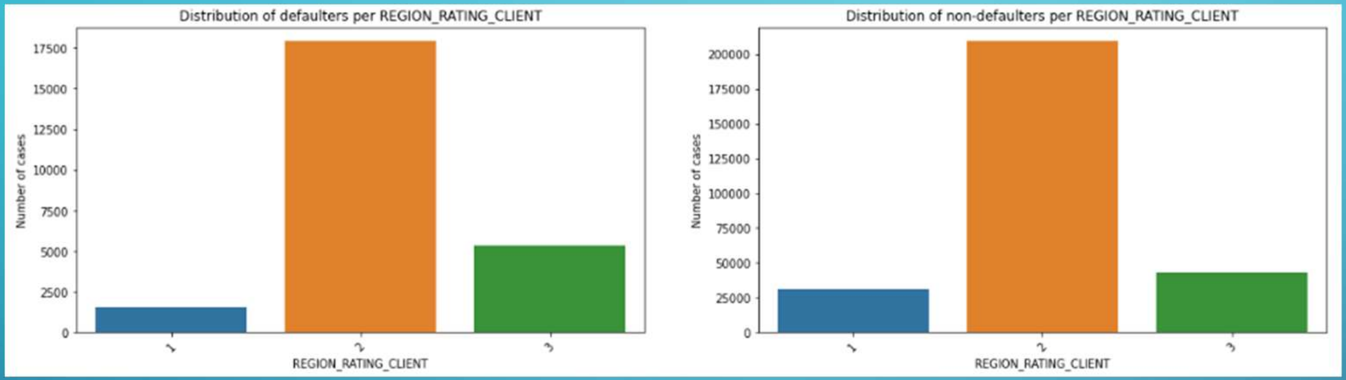
- **Inference: There are more female applicants and defaulters are also more in female then male.**

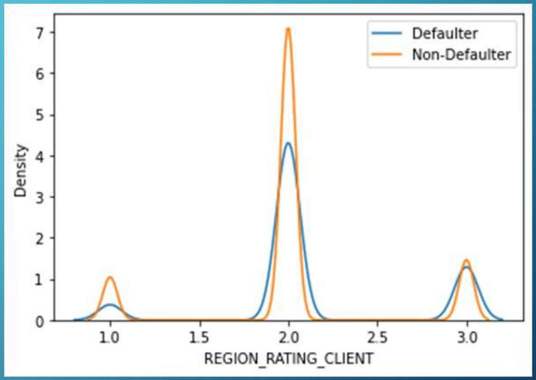
DEMOGRAPHY – AGE CATEGORY



- **Inference: The more senior people more then 40 years of age, they well manage their finances and less defaulters in comparision to applicants who are in 30-40 age bracket.**

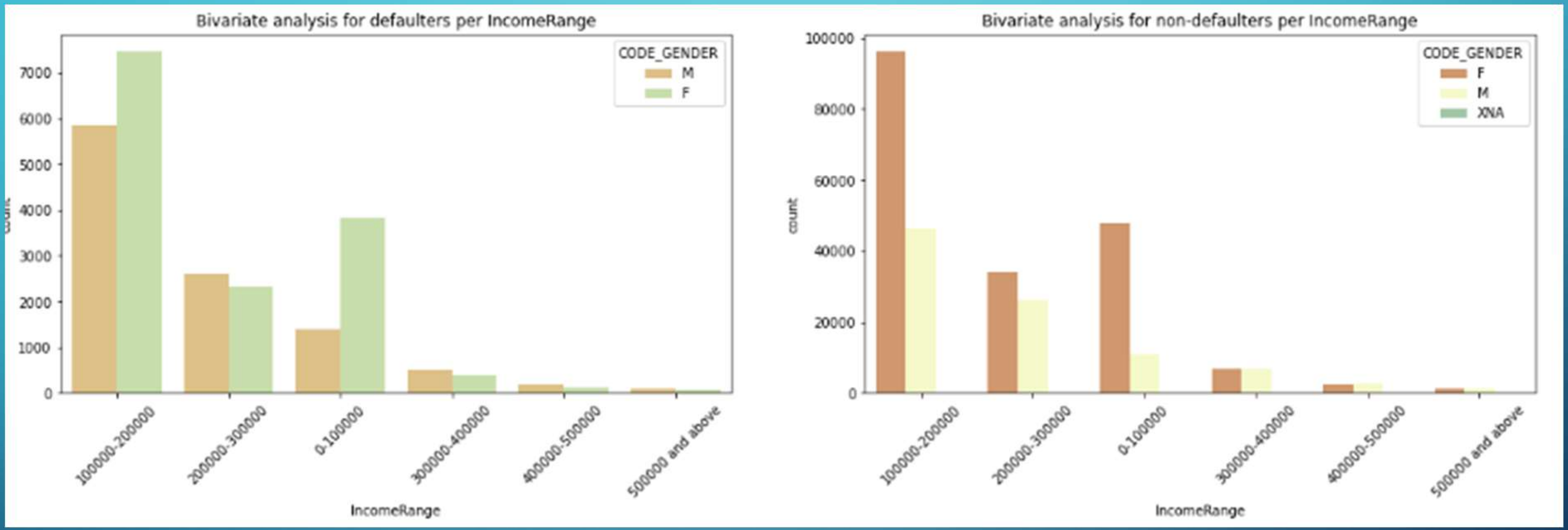
REGIONAL IMPACT – RATING OF REGION OF CLIENT





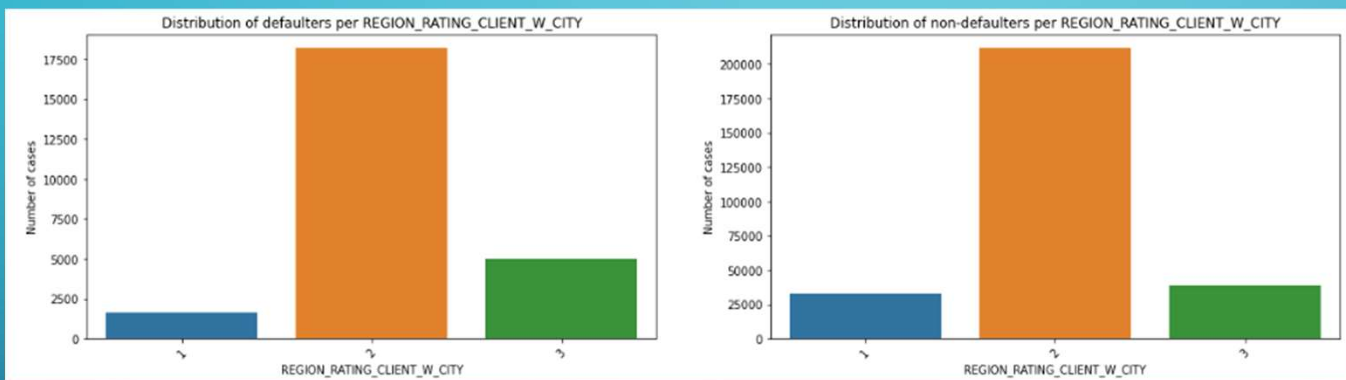
- **Inference:**
- **More applicants are taking loan where region rating is 2 and defaulters are also found maximum in these regions.**
- **More percentage of defaulters with applicants where region rating is 3**

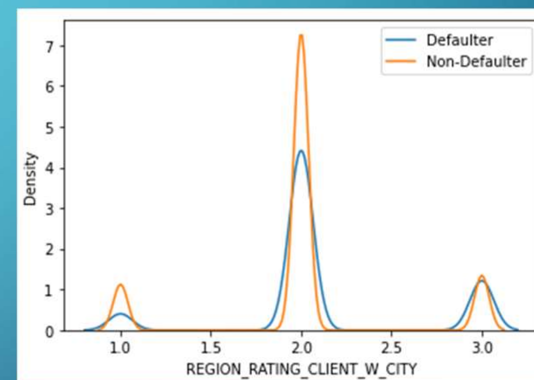
ECONOMY STATUS – INCOME RANGE & GENDER



- **Inference:**
- **Female population shows higher percentage of applications**
- **Female income range as well as well as number of defaulters are higher especially income range 1 lakh to 2 lakhs**

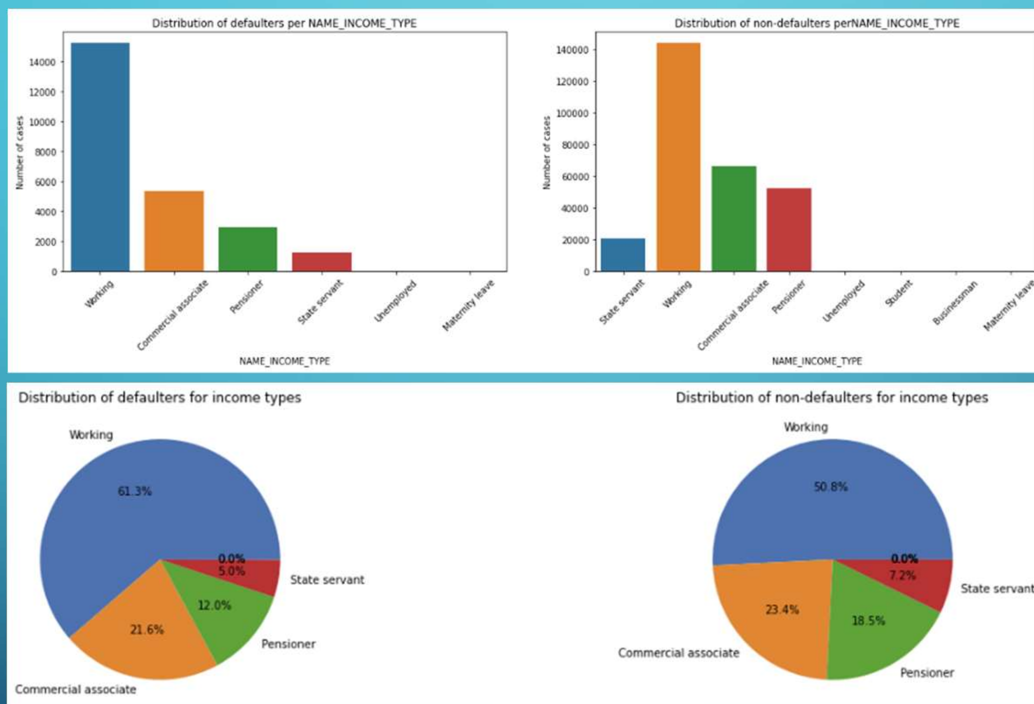
REGIONAL IMPACT – RATING OF REGION OF CLIENT (CITY)





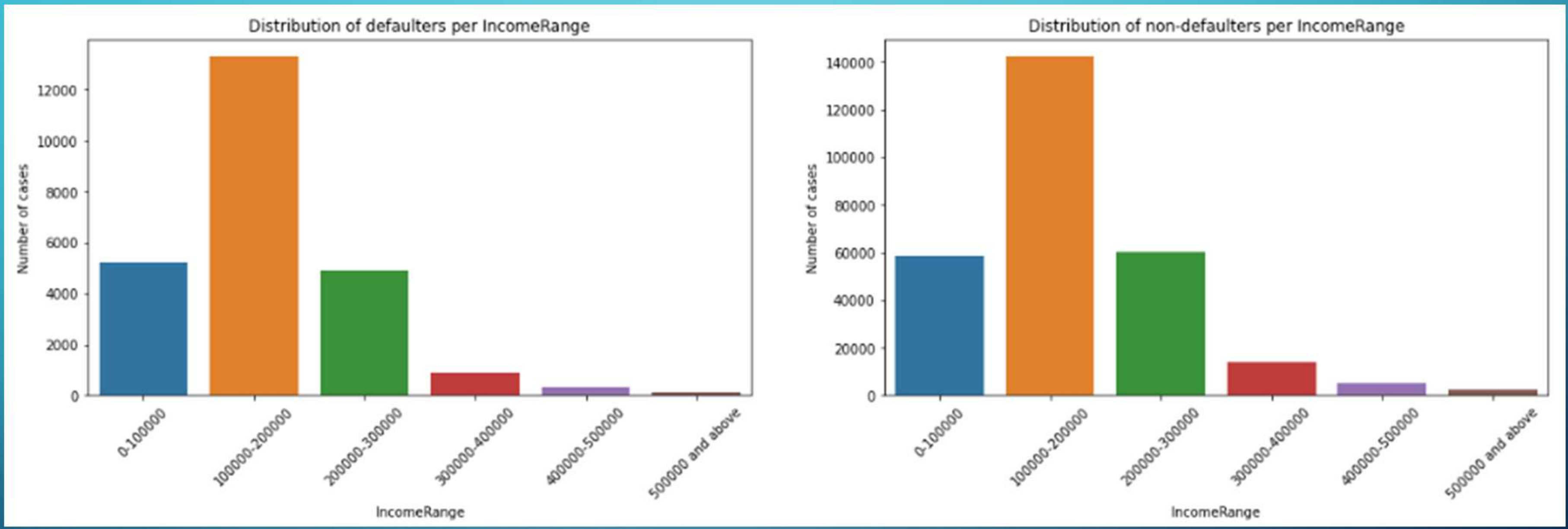
- **Inference: More applicants are taking loan where region rating (with city in account) is 2 and defaulters are also found maximum in these regions. Also defaulters seen that the applicants with rating 1.**

ECONOMY STATUS – INCOME TYPE



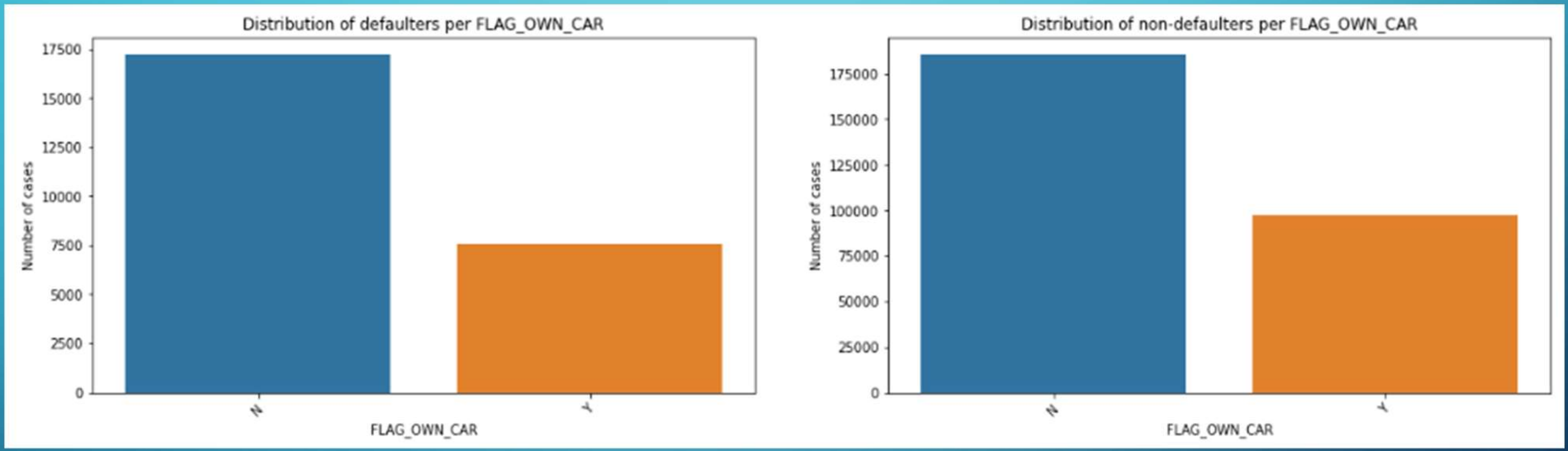
- **Inference: Working class has high defaulters , next commercial associate then pensioner and then state servant.**

ECONOMY STATUS – INCOME RANGE



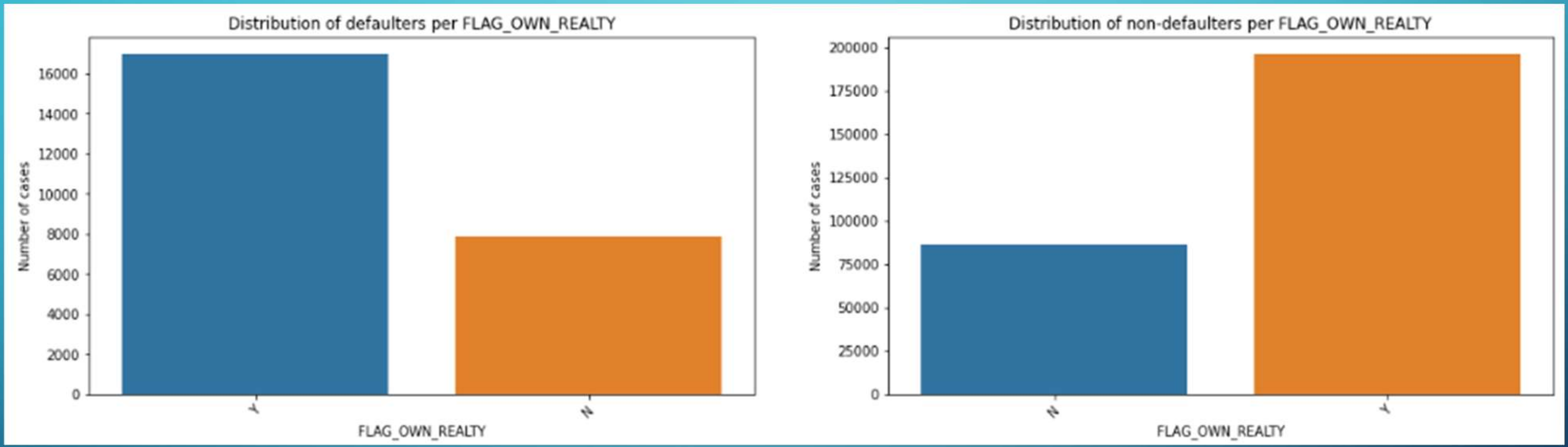
- **Inference: There are more applicants in the range of 1 lakh to 2 lakh and they have maximum defaulters.**

ECONOMY STATUS – CAR OWNERS



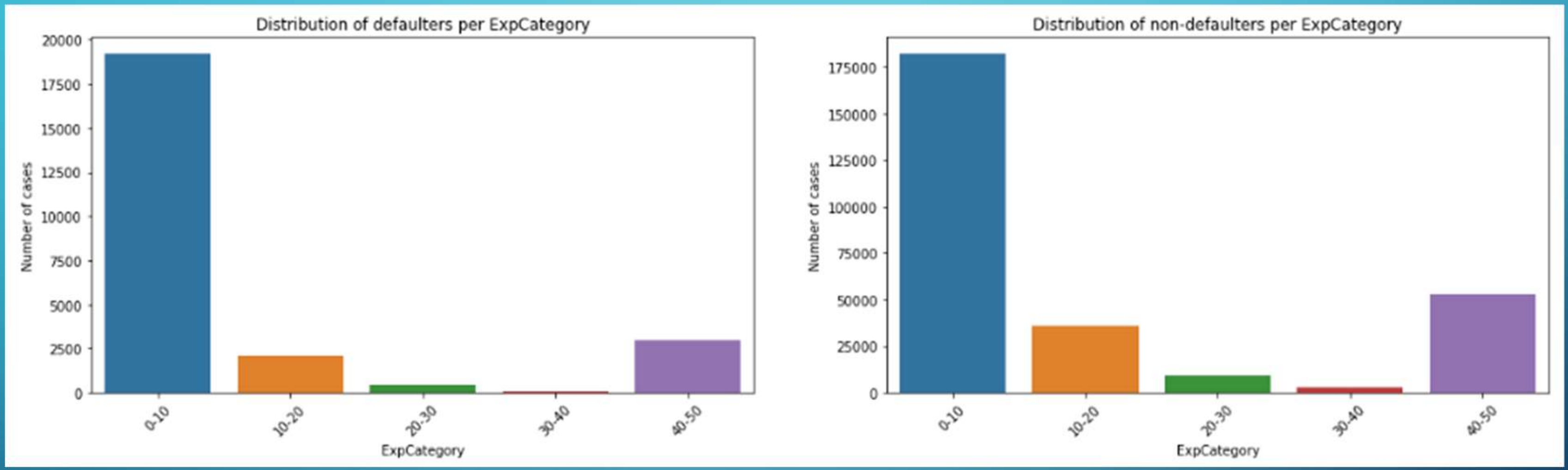
- **Inference: owning a car does not lead to defaulter.**
- some assumptions which need further research: Normally people who does not have a car try to go for a loan and normally the default cases are also more. This could be because people who are financial stable (can afford car) does not go for loan and hence less defaults might observed.

ECONOMY STATUS – CLIENT OWNS HOUSE OR FLAT



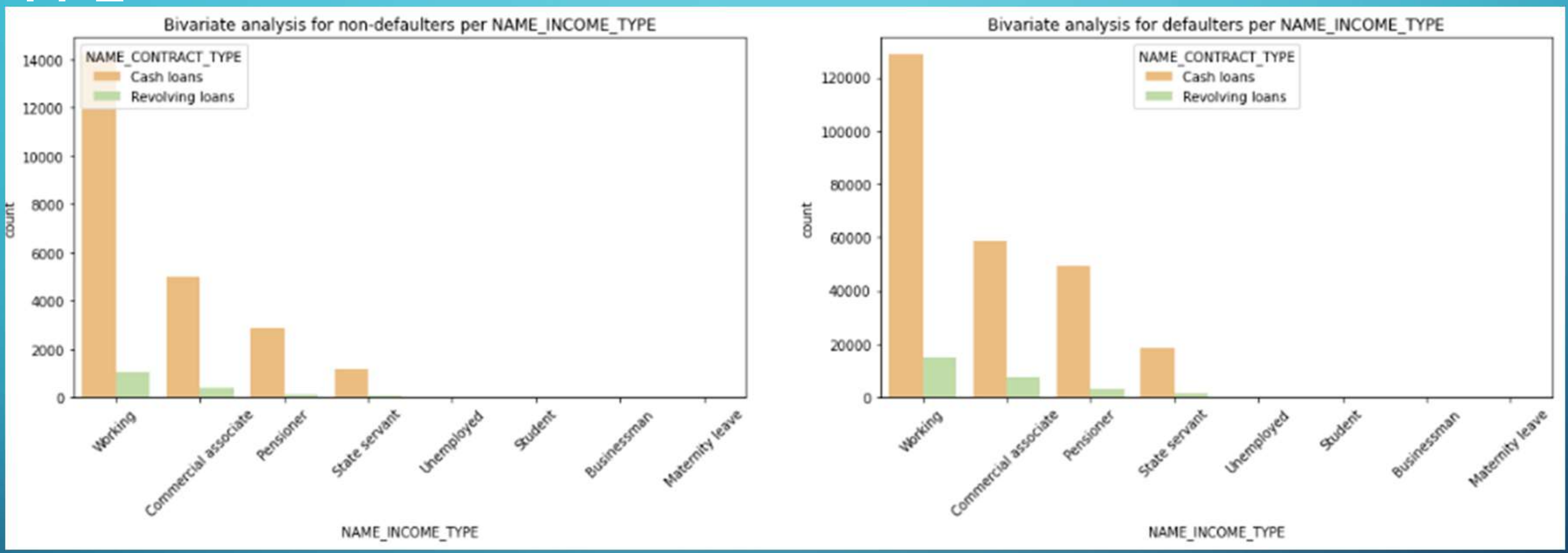
- **Inference: owning a house might be indirect reason where financially lead to defaulter. (Further investigation needed)**

ECONOMY STATUS – EXPERIENCE CATEGORY



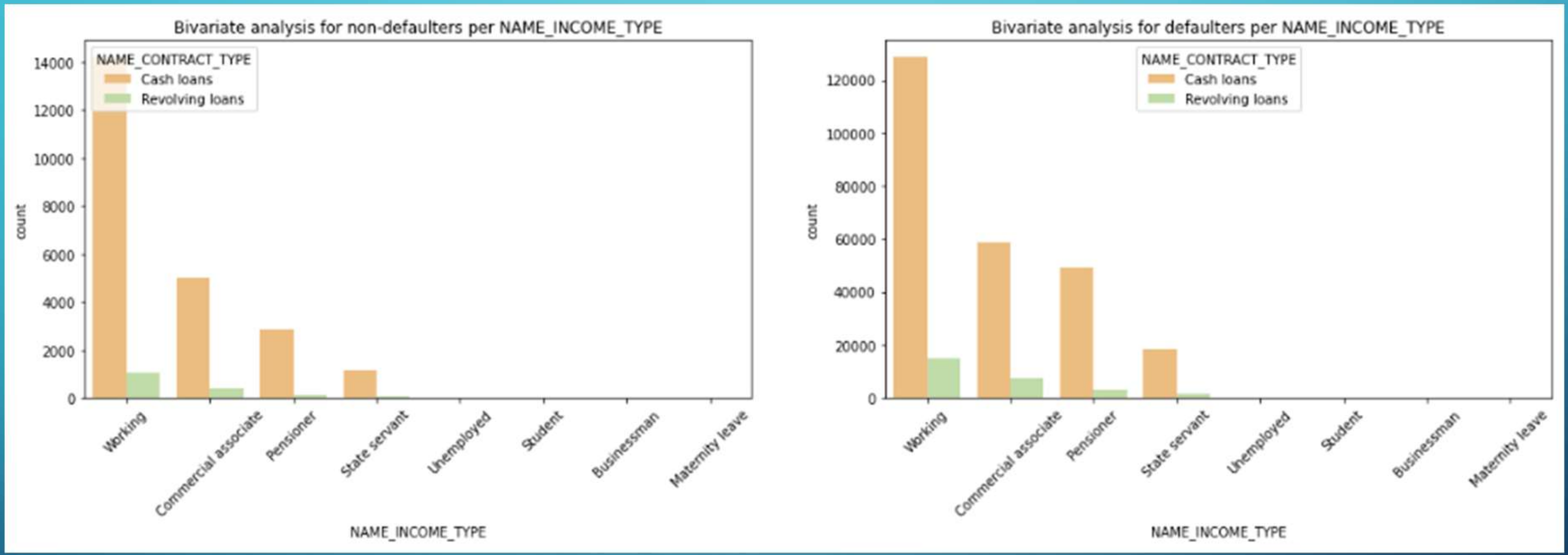
- **Inference: applicants who are having more experience are less succeptable to default.**
- **This could be much correlated to younger age, less children and etc. We can see in bi-variate analysis.**

ECONOMY STATUS – INCOME TYPE & CONTRACT TYPE



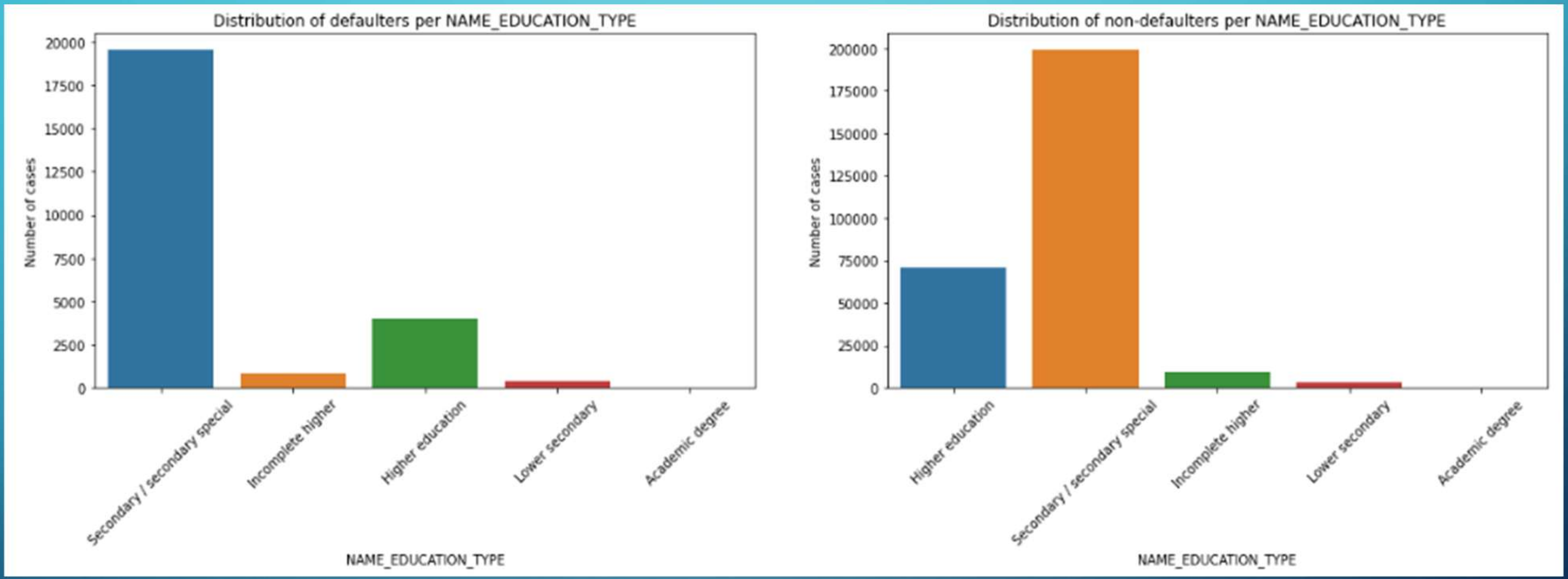
- **Inference: More working applicants defaulted who has taken cash loans**

ECONOMY STATUS – INCOME TYPE AND CONTRACT TYPE



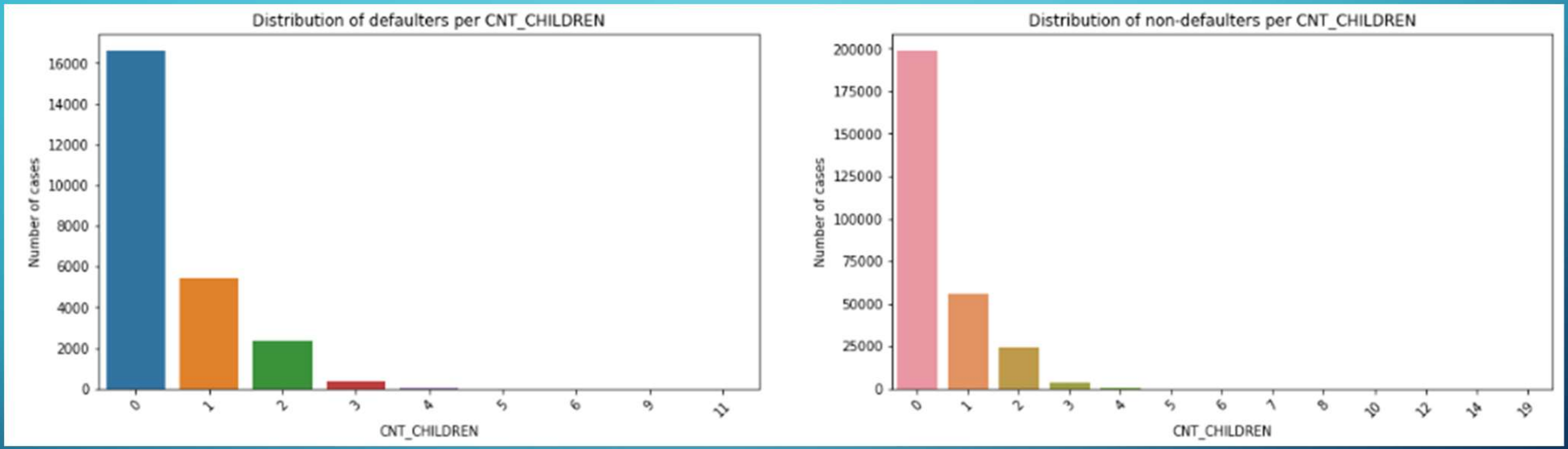
- **Inference: In comparison to revolving loans many applicants who are working class goes for loan and maximum defaulters observed there who has opted for cash loans.**

SOCIAL STATUS – EDUCATION TYPE



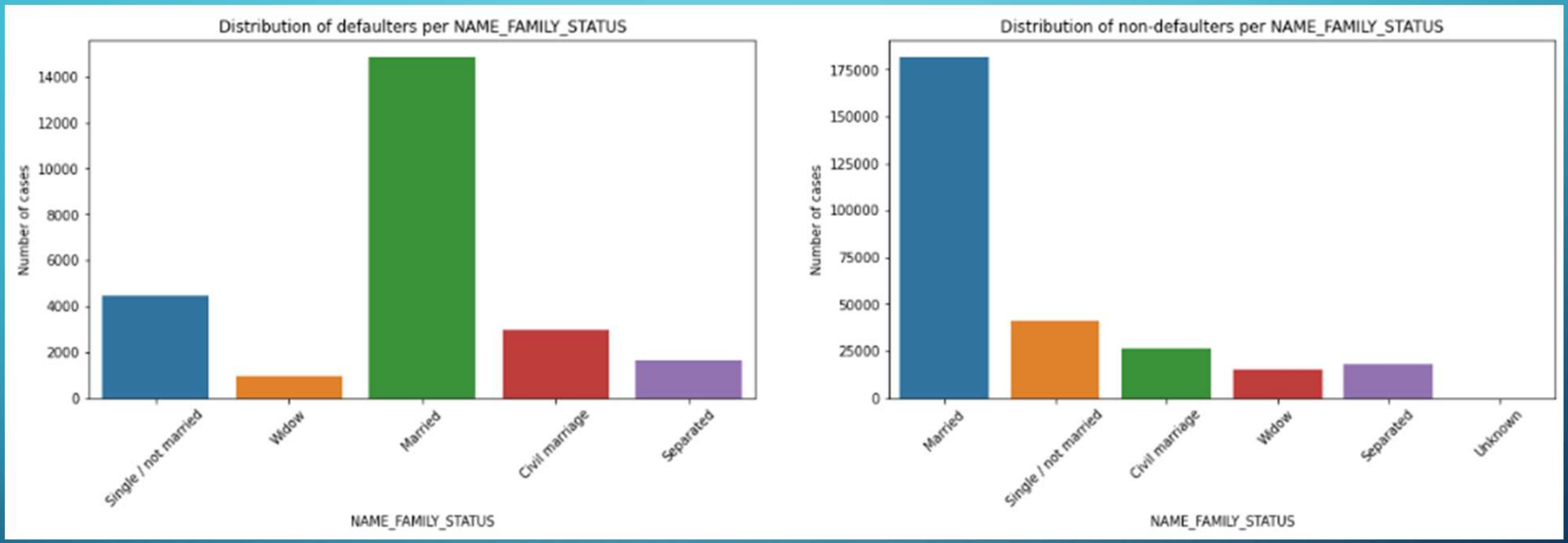
- **Higher educated applicants are well managing their finances and hence there are less defaulters observed.**

SOCIAL STATUS – HAVING CHILDREN



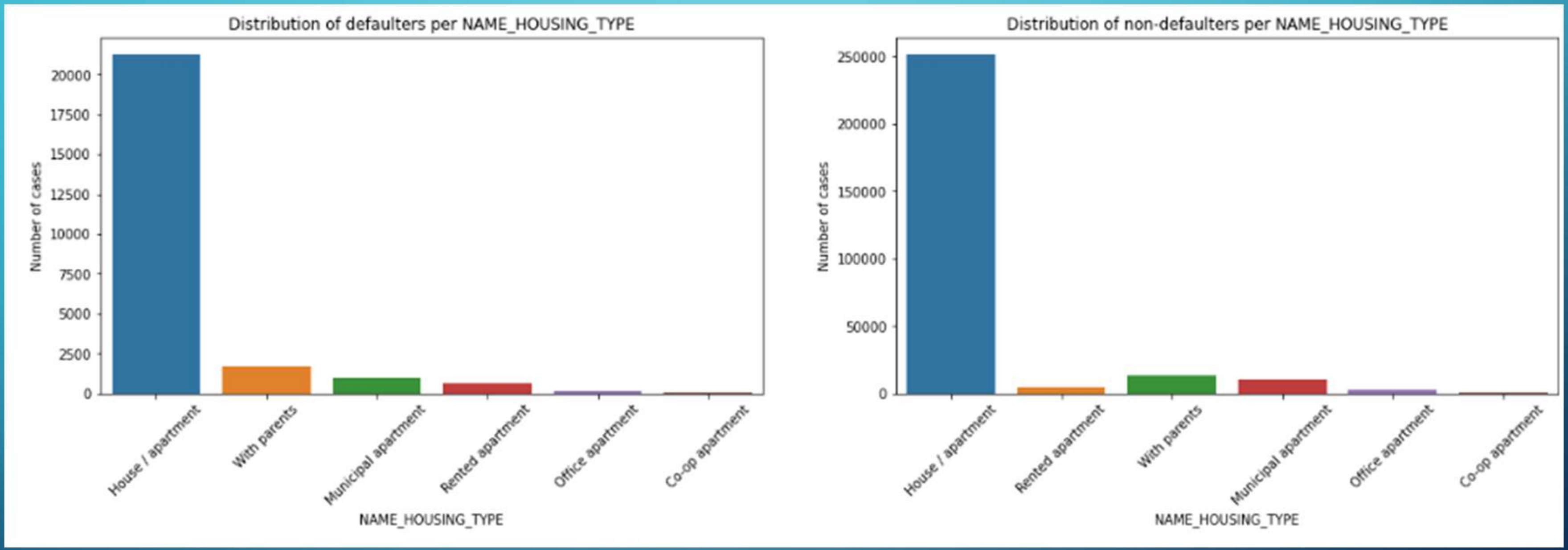
- **Inference: there seems to be not having direct impact in becoming defaulter when applicants have children (age could be a factor though)**
- **Needs further investigation: Looks like applicants who are young who does not have children go for loan and get defaulter as well.**

SOCIAL STATUS – FAMILY STATUS



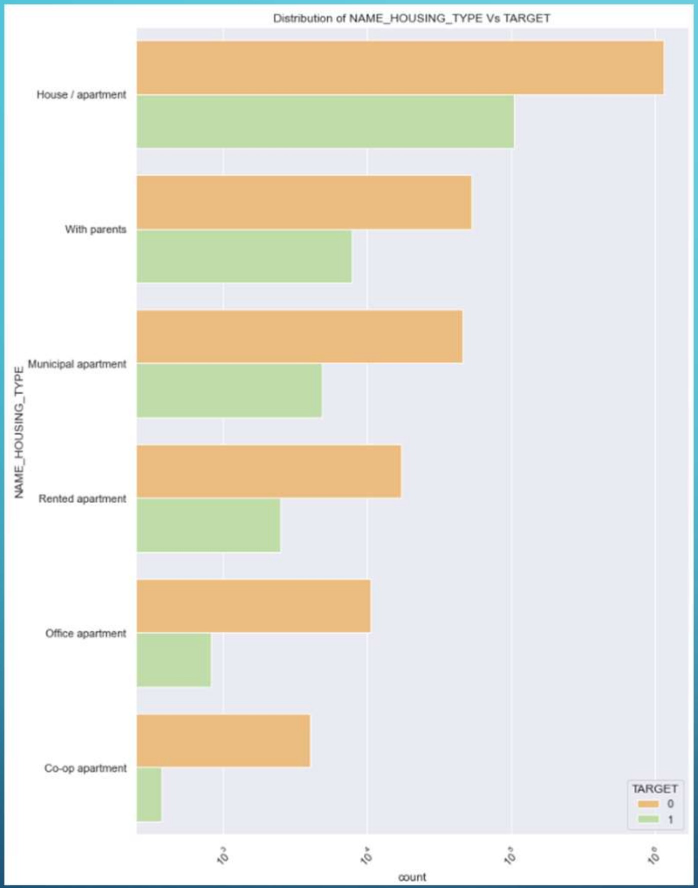
- **Inference: Married man overall goes for loan and default also observed. After marriage responsibilities increases and number of applicants increases. Defaulters also proportionally increases**

SOCIAL STATUS – HOUSING TYPE



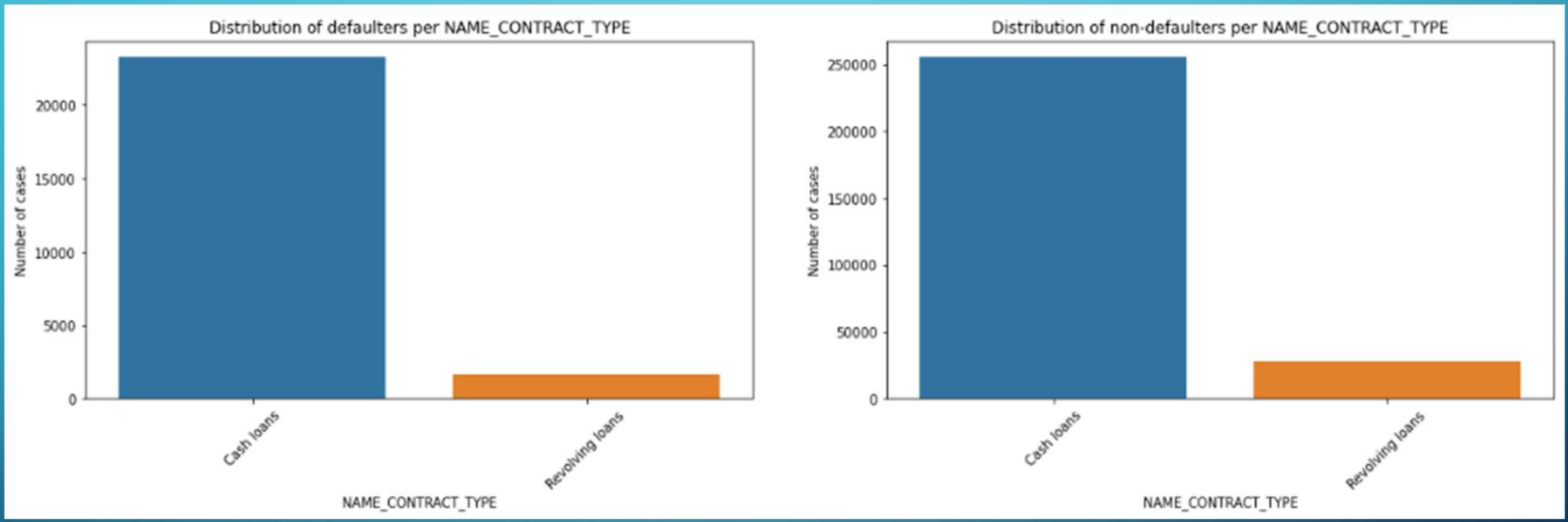
- **Inference: applicants who has house they go for loan and as well as more defaulters in this segment.**

SOCIAL STATUS – HOUSING TYPE (DEFAULT)



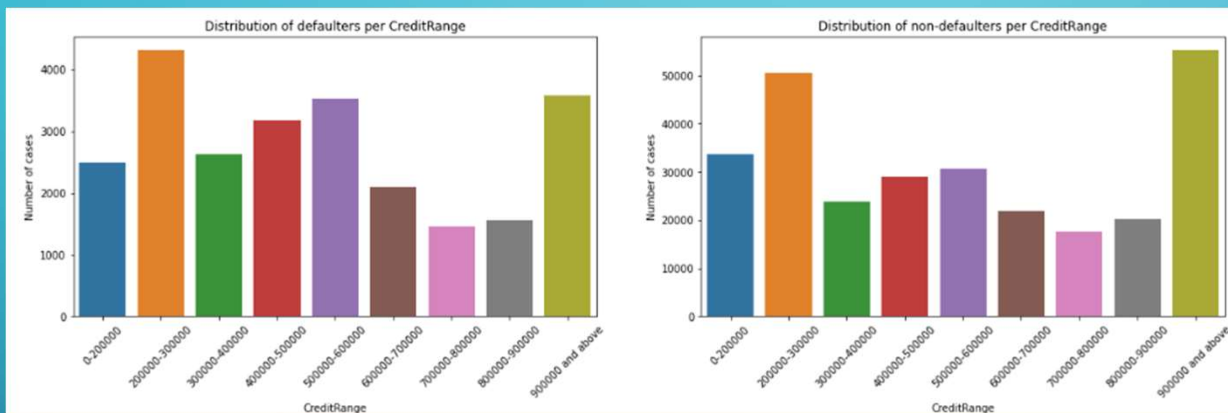
- **Inference: Office apartment and co-op apartment loan applicants has less defaulters; where as house apartment has many applicants and defaulters as well**

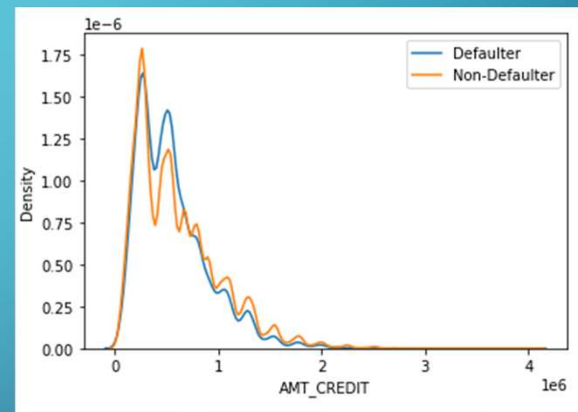
LOAN - TYPE OF CONTRACT



- **More applicants are for cash loans and defaulters also observed more in this segment.**
- **Number of defaulters / non-defaulters for cash loans almost same; where as in case of revolving loans number of defaulters are very less in comparison to the proportion**

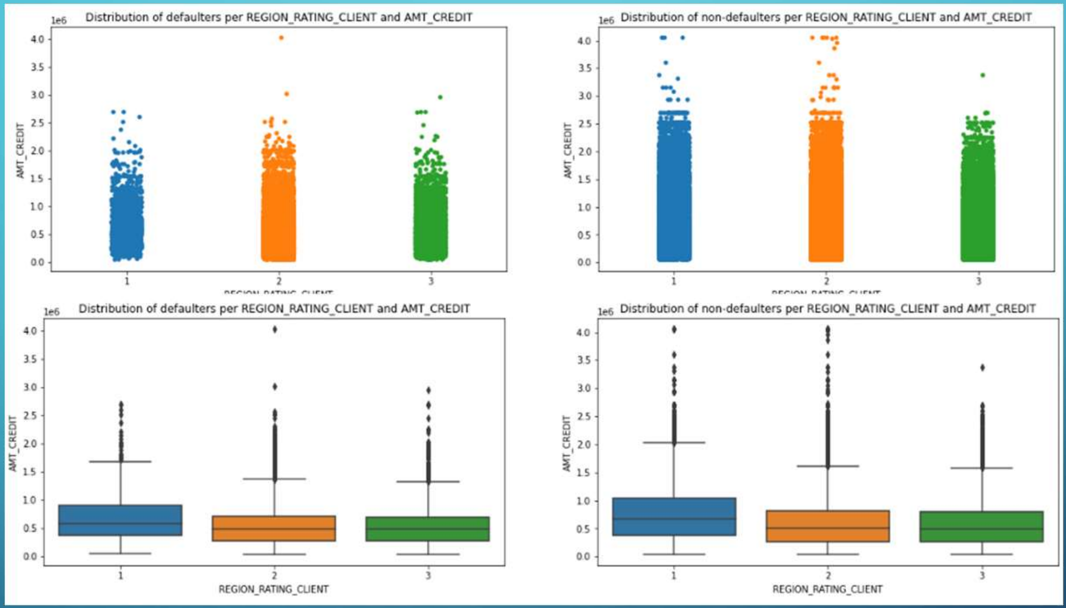
LOAN - CREDIT RANGE





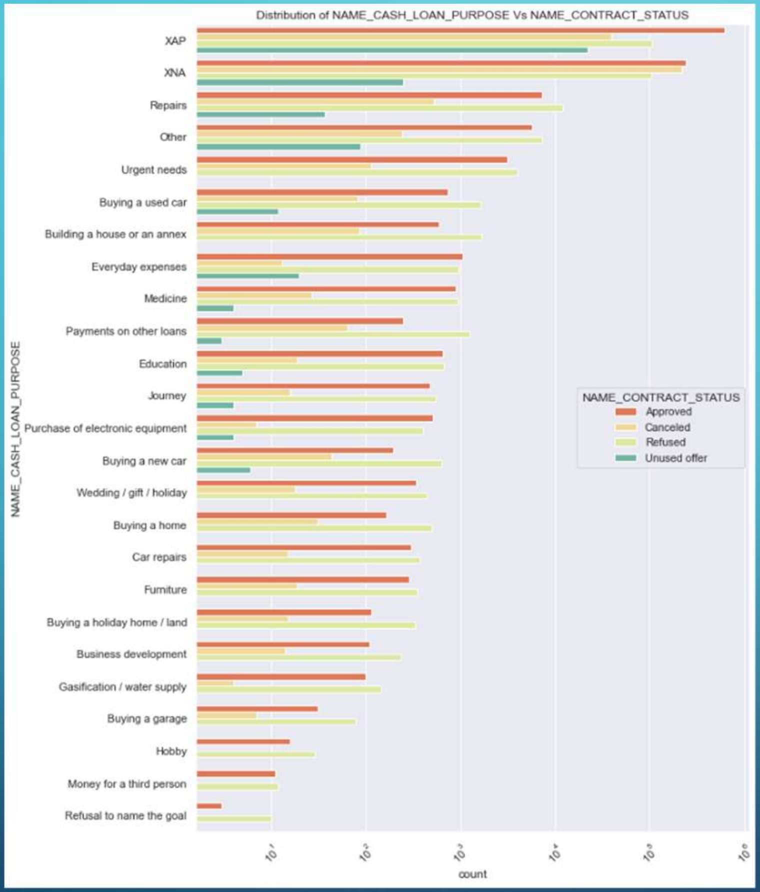
- **Inference: More defaulters are seen for credit range more then 9 lakhs in proportion to number applicants.**

LOAN – RATING OF REGION OF CLIENT & CREDIT AMOUNT



- **Inference: Rating of region where client leaves has some minor impact, more credit given to applicants where region rating is 1**
- **Some outliers are seen.**

CASH LOAN PURPOSE AND CONTRACT STATUS

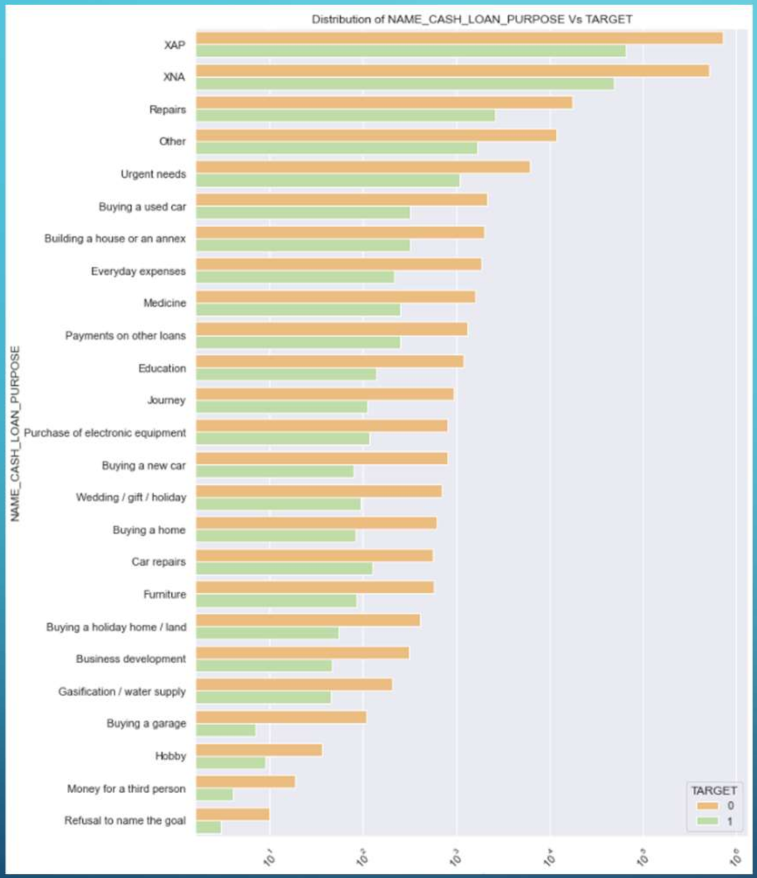


- **Inference:**

In case of following cash loan purposes there are more rejection is seen then approval:

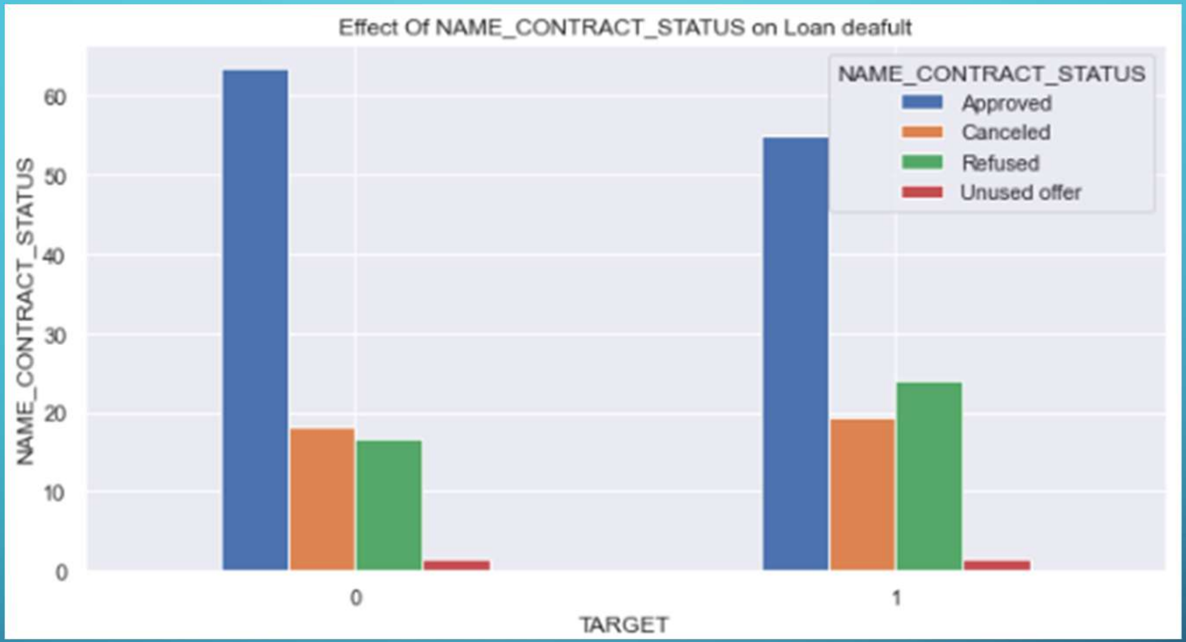
Repairs, buying used car, payment on other loans, buying a new car, buying home, buying a garage

CASH LOAN PURPOSE AND TARGET



- **Inference: Loan history of Repairs are having more difficulty in paying money and they become defaulters.**
- **Where as buying a garage, business development less likely to default.**

PREVIOUS HISTORY ON CONTRACT STATUS

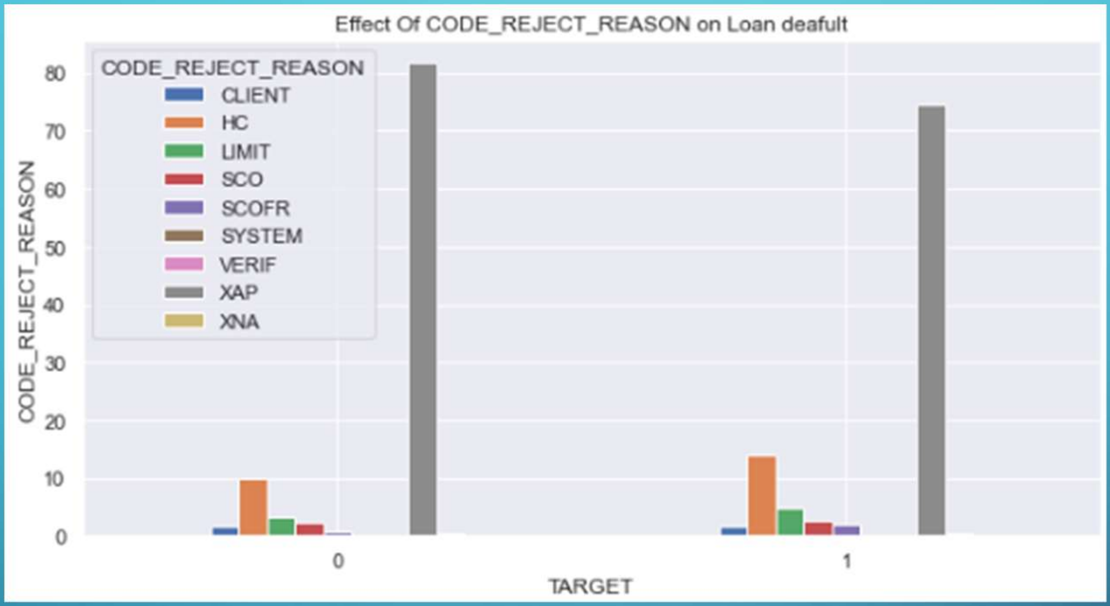


Inference:

Proportion of cancelled and refused loans in previous history; seen more defaulters in current application.

Proportion for approved loans in previous history has slight decline in default

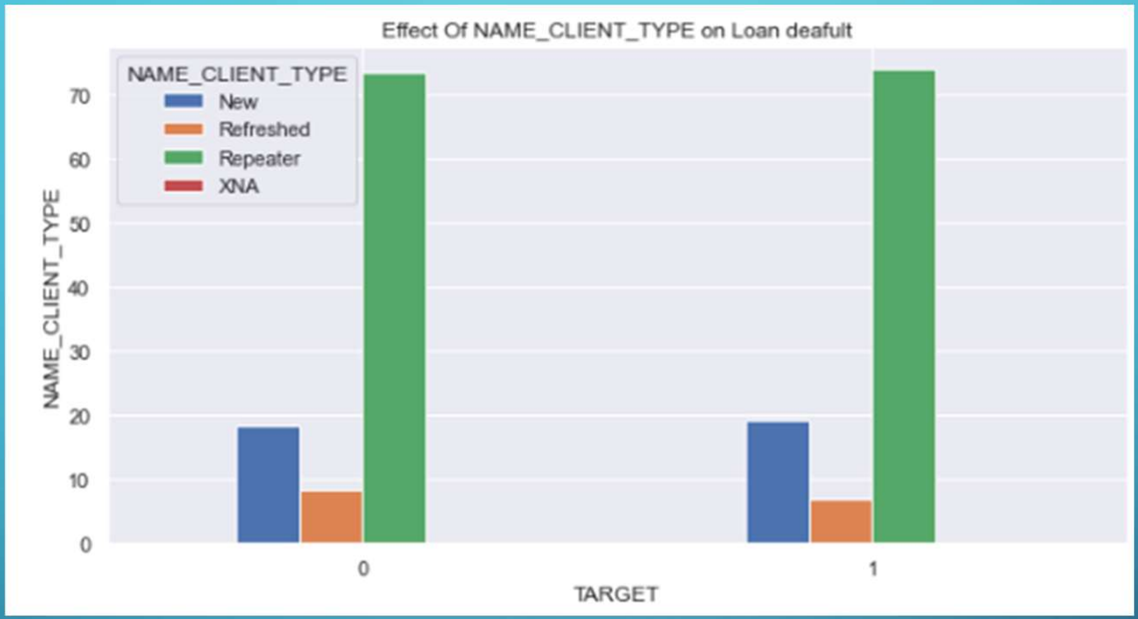
PREVIOUS HISTORY ON CODE REJECT REASON



Inference:

Where previous application reject reason is HC there more default observed. XAP cases are approved cases where default is less observed

PREVIOUS HISTORY ON CLIENT TYPE

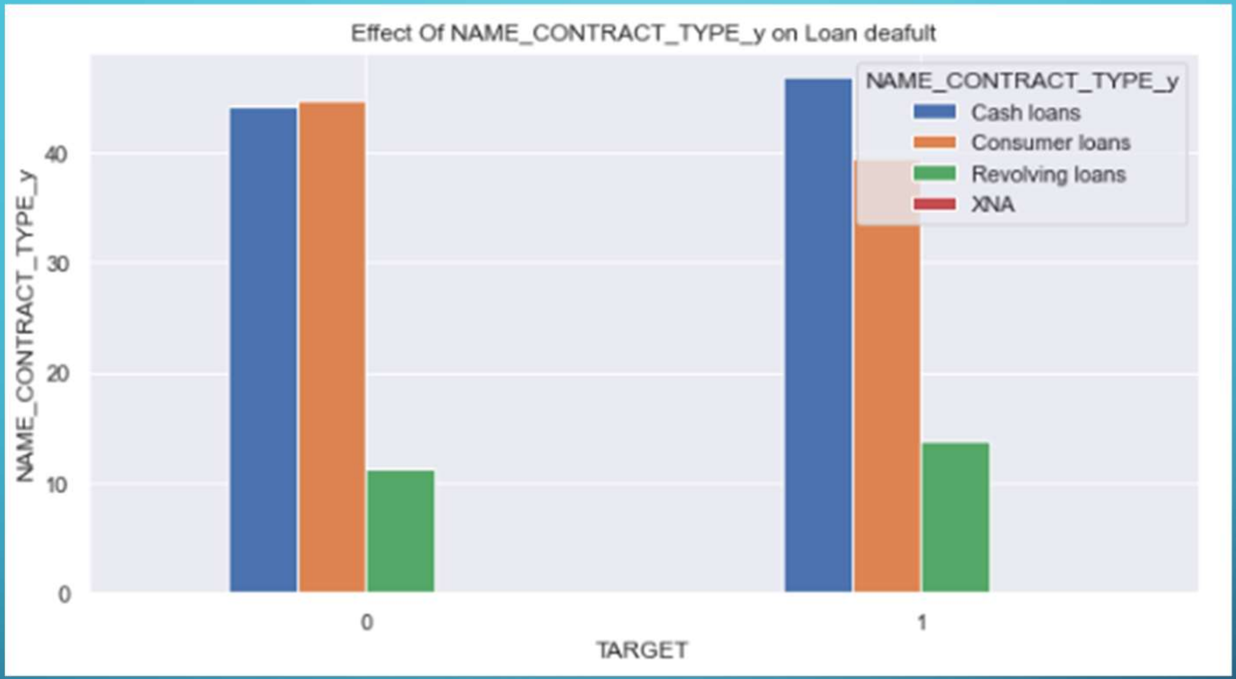


Inference:

Repeater and new customer has same default rate

Refreshed customer has seen to have less default

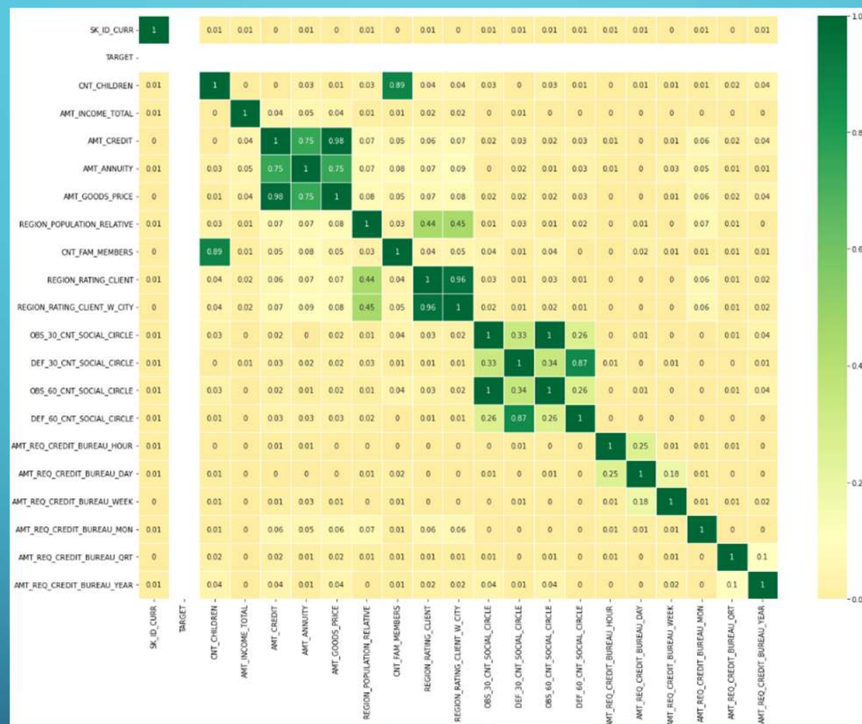
PREVIOUS HISTORY ON NAME CONTRACT TYPE



Inference:

Default in proportion of cash loans has slightly more.
Default in proportion of consumer loans has slightly less.
Default for proportion of revolving loan is slightly more.

CORRELATION

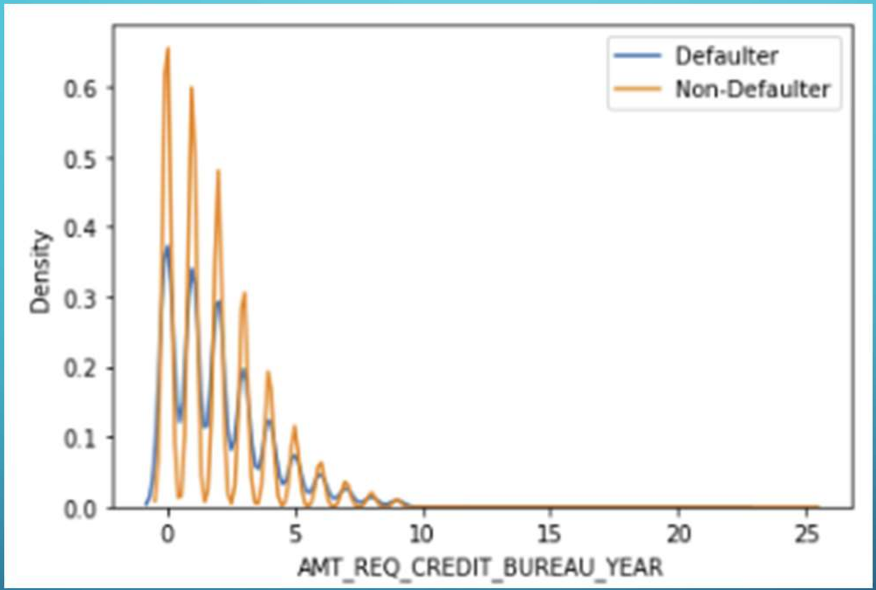


Inference: Most correlated attributes are following:

1) Loan annuity and price of the good for which loan is given are mostly correlated.
2) Rating of region where client leaves with or without city into account are almost same (most corelated).
3) Count of family members and count of children are mostly correlated
4) clients social surroundings observable 30 days past due is almost correlated to defaulted 60 days past due

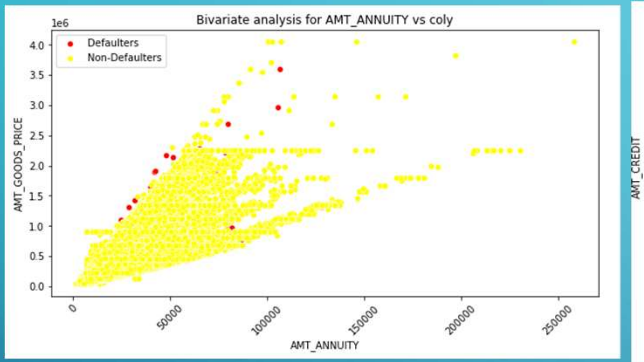
Also plotted pairplot to observe the same correlation found in above heatmap.

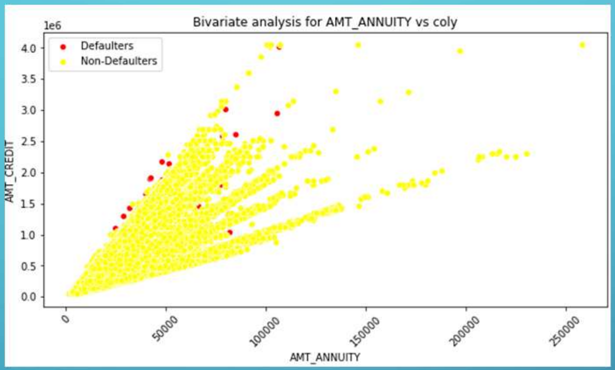
CORRELATION - TO UNDERSTAND CREDIT BUREAU ENQUIRIES TO DEFAULTERS

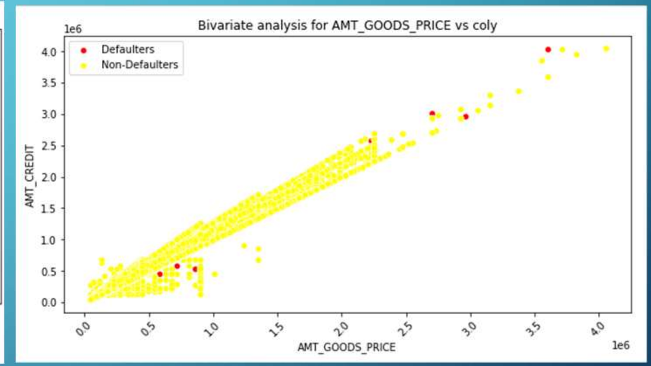


- **Inference: Observed that more enquiries are being done for less applicants.**
- **Need to focus on defaulted applicants where less enquiry is being done where more defaulters seen**

CORRELATION







- **Inference:**
- Amount good price and annuity amount are mostly correlated.
- Amount credit and annuity amount are mostly correlated.
- Amount credit and amount good price are mostly correlated.

RECOMMENDATIONS

Category	Recommendations
Demography	
Gender	Female applicants are promising to be approached for loan – less riskier to default
	Riskier group – 30 to 40 years age bracket. The more senior people more then 40 years of age, they well manage their finances and less defaulters in comparison to applicants who are in 30-40 age bracket.
Region rating (with and without city)	More applicants are taking loan where region rating is 2 and defaulters are also found maximum in these regions. However percentage default is less in these regions in comparison to other regions whose rating is 3. More percentage of defaulters with applicants where region rating is 3.
Economy Status	
Income Type	Working class has high defaulters , next commercial associate then pensioner and then state servant.
Income range	There are more applicants in the range of 1 lakh to 2 lakh and they have maximum defaulters.
Car Owner	Owning a car does not lead to defaulter. This could be because people who are financial stable (can afford car) does not go for loan and hence less defaults might observed. owning a house might be indirect reason where financially lead to defaulter. (Further investigation needed to proof this assumption)
Experience	Applicants who are having more experience are less susceptible to default. (This could be much correlated to younger age, less children and etc. We can see in bi-variate analysis.)
Income Segment	More working applicants defaulted who has taken cash loans In comparison to revolving loans many applicants who are working class goes for loan and maximum defaulters observed there who has opted for cash loans.

RECOMMENDATIONS – CONTD..

Category	Recommendations
	Social Status
Education	Higher educated applicants are well managing their finances and hence there are less defaulters observed. there seems to be not having direct impact in becoming defaulter when applicants have children (age could be a factor though) Needs further investigation: Looks like applicants who are young who does not have children go for loan and get defaulter as well.
Marital Status	Married man overall goes for loan and default also observed. Assumptions - After marriage responsibilities increases and number of applicants increases. Defaulters also proportionally increases applicants who has house they go for loan and as well as more defaulters in this segment.
House Type	Office apartment and co-op apartment loan applicants has less defaulters; where as house apartment has many applicants and defaulters as well
	Type of Loan
Contract Type	More applicants are for cash loans and defaulters also observed more in this segment. Number of defaulters / non-defaulters for cash loans almost same; where as in case of revolving loans number of defaulters are very less in comparison to the proportion
Credit Range	More defaulters are seen for credit range more then 9 lakhs in proportion to number applicants.
Rating region and credit amount	Rating of region where client leaves has some minor impact, more credit given to applicants where region rating is 1. Some outliers are seen.
Cash loan purpose and contract status	In case of following cash loan purposes there are more rejection is seen then approval: Repairs, buying used car, payment on other loans, buying a new car, buying home, buying a garage

RECOMMENDATIONS – CONTD..

Category	Recommendations
Previous history	
Cash loan purposes	In case of following cash loan purposes there are more rejection is seen then approval: Repairs, buying used car, payment on other loans, buying a new car, buying home, buying a garage
Status of previous loans	Proportion of cancelled and refused loans in previous history; seen more defaulters in current application. Proportion for approved loans in previous history has slight decline in default
Reason for rejection	Where previous application reject reason is HC there more default observed. XAP cases are approved cases where default is less observed
Client Type	Repeater and new customer has same default rate Refreshed customer has seen to have less default
Contract Type	Default in proportion of cash loans has slightly more. Default in proportion of consumer loans has slightly less. Default for proportion of revolving loan is slightly more

THANK YOU

- Susil Patro (Data Science Program July, 2022 batch