Bank Loan Default Risk Analysis
(EDA CASE STUDY)

By:-

Ranjana Gupta

Kamal Kumar

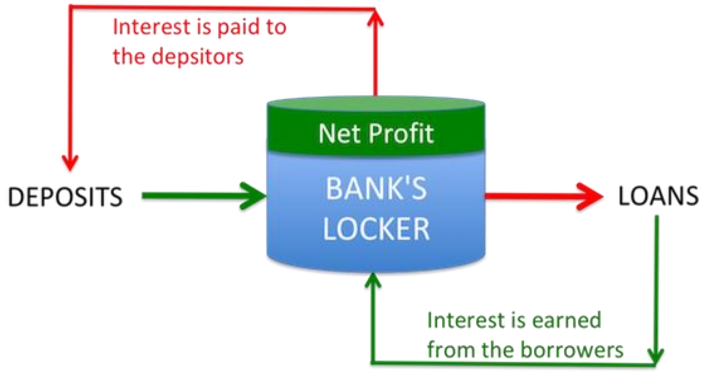
Bank Loan Default Risk Analysis

Table of Contents

1. Introduction

Business Understanding:

- The primary business of any bank revolves around managing the spread between the deposits.
- In other words, when the interest that a bank earns from loans is greater than the interest it pays on deposits, it generates income from the interest rate spread.
- Clearly, the major part of revenues of any bank is attained through the loans they give to the people. But there are chances that the loans may not be paid back by few of the customers, making it a bad loan.



"The only good loan is one that gets paid back."

Business Scenario:

- When a customer applies for a loan, there are four types of decisions that could be taken by the lender /applicant :

- **Approved:**
  The Company approved the loan Application.
- **Cancelled:**
  The client cancelled the application sometime during approval.
- **Refused:**
  The company rejected the loan.
- **Unused offer:**
  Loan has been cancelled by the client but on different stages of the process.

Business Profitability:

- Insufficient or non-existent credit history of an Urban Customer puts the bank lending company in position of dilemma about approving the loan.
- This dilemma revolves around the likelihood that a customer would pay back the loan or not, and can potentially result in 2 types of loss:
  - **Credit Loss**
    If an applicant is not likely to repay the loan, then approving the loan may lead to a financial loss for the company.
  - **Interest Loss**
    If an applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.

Chances that a customer will be a

	Defaulter	Non-Defaulter
Loan Approved	Bad Loan (Credit Loss)	Good Loan (Loan will be repaid with Interest as profit)
Loan Denied	Good Decision (saved from Credit Loss)	Wrong Decision (Interest Loss)

2. Business Objective:

- With this case study, we aim to understand the **strong driving factors behind loan default**.
- This will ensure that the consumers capable of repaying the loan are not rejected, and certain adaptable actions can be taken on client basis, if they are likely to face difficulty paying their installments in future.
- The result of this Risk analysis would help the bank to **identify the patterns**, which indicate if a client has difficulty paying their installments.
- This can further **influence the decisions** such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, collateral etc.

The company may utilize this knowledge for its portfolio and risk assessment.

3.Dataset Understanding

- This Risk analysis is done based on 2 datasets as explained below:

  - ✓ **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. Data related to applicant's socio-economic status is also available
  - ✓ **previous_application.csv** contains information about the client's previous loan data. It contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

Overview structure of the datasets

	Application Data	**Previous Application Data**
Number of Rows	307511	1670214
Number of Columns	122	37
Number of Columns with Null Values	67	15
Number of Columns with more than 50% Null Values	41	4

For further analysis, we have dropped the columns with more than 50% missing values as imputing would bias the analysis & ignoring missing values will not help us with efficient insights.

Remaining all data are considered for next level analysis.

Data Cleansing

Null Values - Strategy

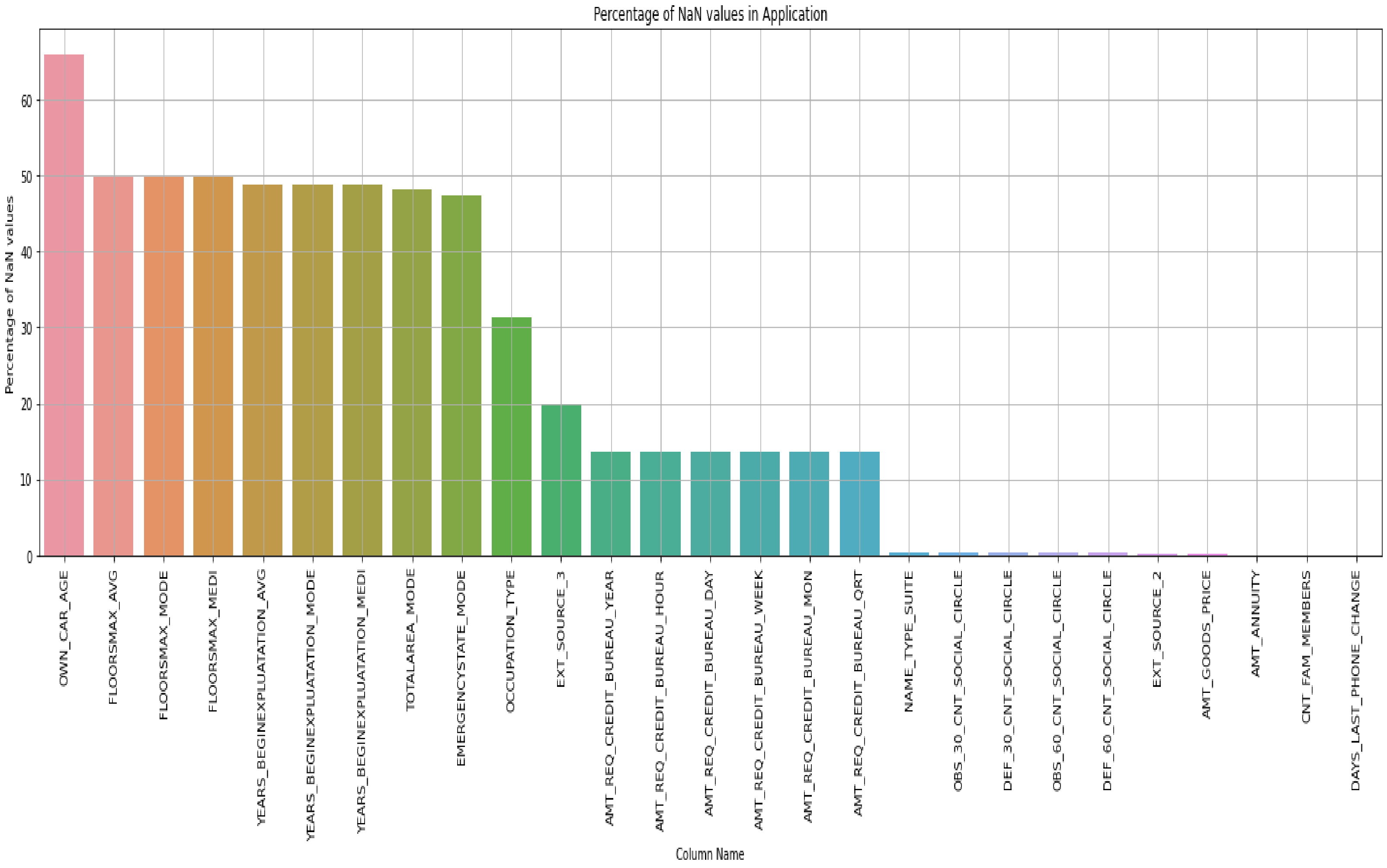
- 20 columns related to FLAG_DOCUMENT_% are dropped, as univariate analysis did not show any meaningful value due to lack of clarity.

- Few columns with around 47% null values are also dropped. This includes YEARS_BEGINEXPLUATATATION_% and FLOORSMAX_%.

- Impute Mode of column value for < 13 % nulls

- Age of car has intentionally been retained for possible insight

Application_data : Columns with Nulls after >50 % deletion



Data Type Conversion

For further analysis we changed the data type to correct format after logical check of the data :
i.e. object to numeric

- AMT_REQ_CREDIT_BUREAU_WEEK
- AMT_REQ_CREDIT_BUREAU_MON
- AMT_REQ_CREDIT_BUREAU_HOUR
- AMT_REQ_CREDIT_BUREAU_DAY
- AMT_REQ_CREDIT_BUREAU_QRT
- DAYS_REGISTRATION
- DEF_30_CNT_SOCIAL_CIRCLE
- DEF_60_CNT_SOCIAL_CIRCLE
- OBS_30_CNT_SOCIAL_CIRCLE
- OBS_60_CNT_SOCIAL_CIRCLE
- Applied abs() to column CNT_FAM_MEMBERS

Columns representing No. of Days prior to an event were changed to obsolete values

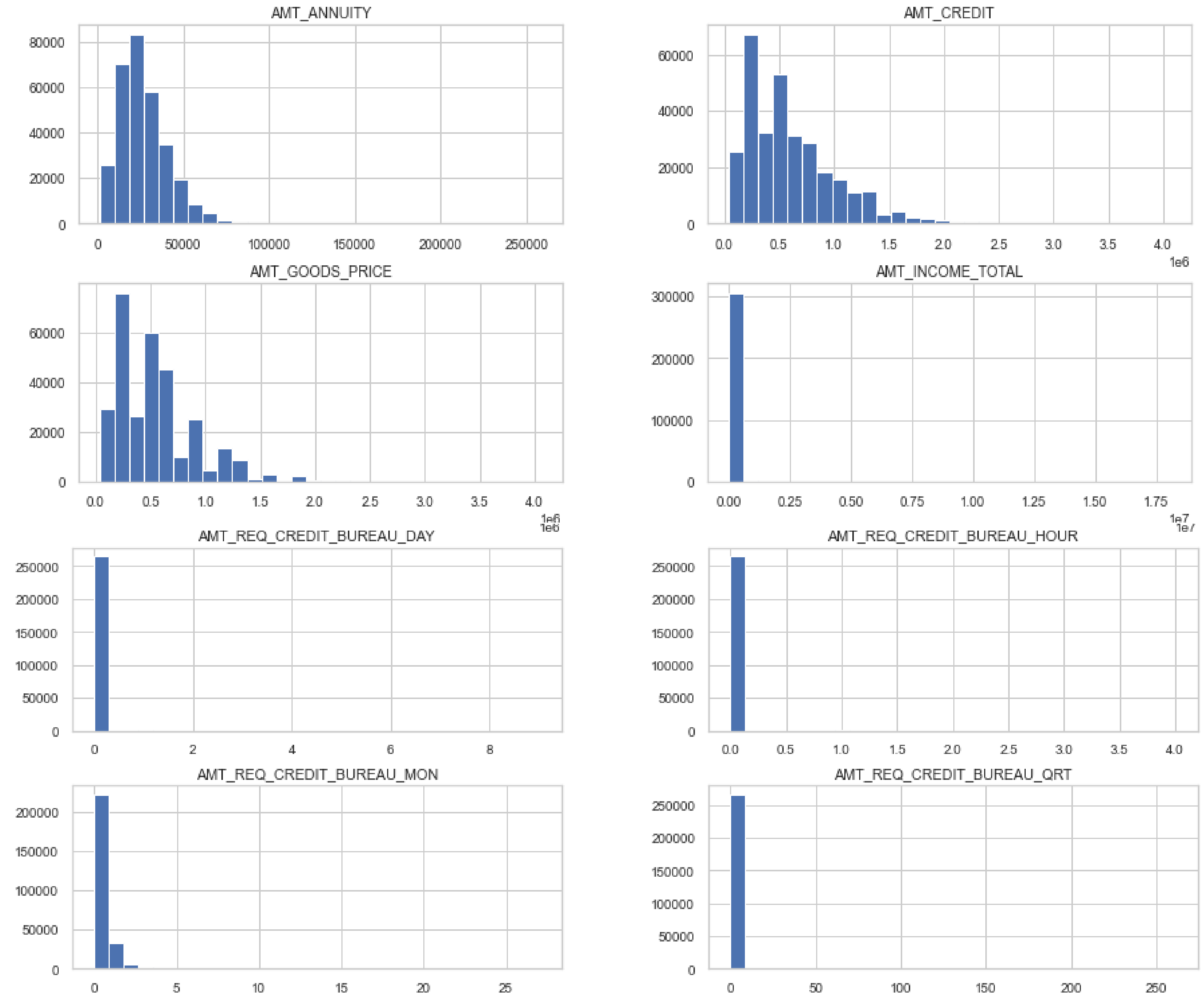
- DAYS_BIRTH
- DAYS_EMPLOYED
- DAYS_EMPLOYED
- DAYS_REGISTRATION
- DAYS_ID_PUBLISH
- DAYS_LAST_PHONE_CHANGE

Outliers Inspection

- Continuous Variables

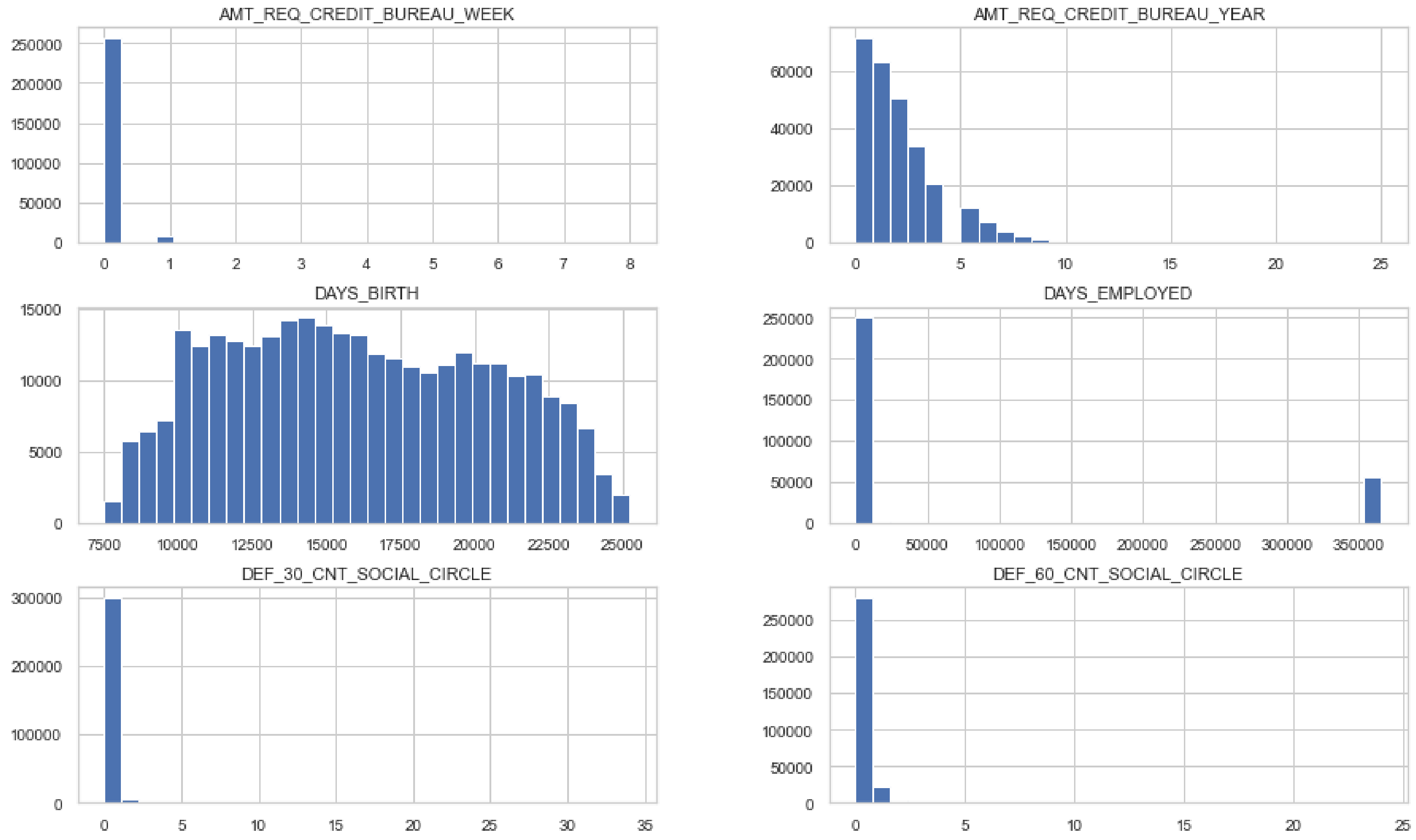
To understand the distribution of the data for each variable, we did plot a series of count plots and distribution plots and the histogram : This was run over following variables:

- AMT_CREDIT
- AMT_GOODS_PRICE
- DAYS_EMPLOYED
- OBS_30_CNT_SOCIAL_CIRCLE
- OBS_60_CNT_SOCIAL_CIRCLE
- AMT_REQ_CREDIT_BUREAU_HOUR
- AMT_REQ_CREDIT_BUREAU_WEEK
- AMT_REQ_CREDIT_BUREAU_QRT
- AMT_INCOME_TOTAL
- REGION_POPULATION_RELATIVE
- DEF_30_CNT_SOCIAL_CIRCLE
- DEF_60_CNT_SOCIAL_CIRCLE
- AMT_REQ_CREDIT_BUREAU_DAY
- AMT_REQ_CREDIT_BUREAU_MON
- AMT_REQ_CREDIT_BUREAU_YEAR
- AMT_ANNUITY
- DAYS_BIRTH
- EXT_SOURCE_2
- EXT_SOURCE_3



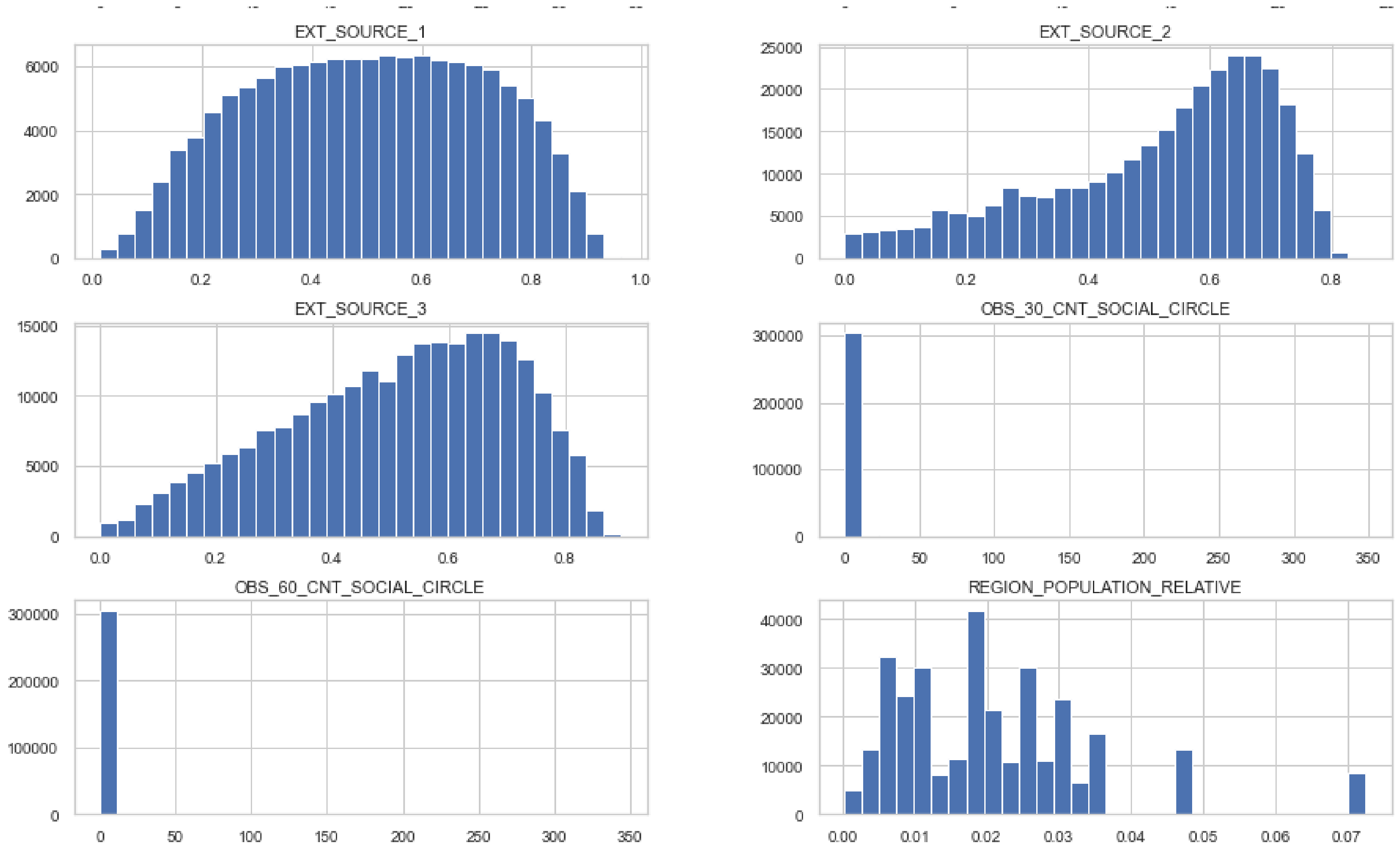
Outliers Inspection

- Continuous Variables



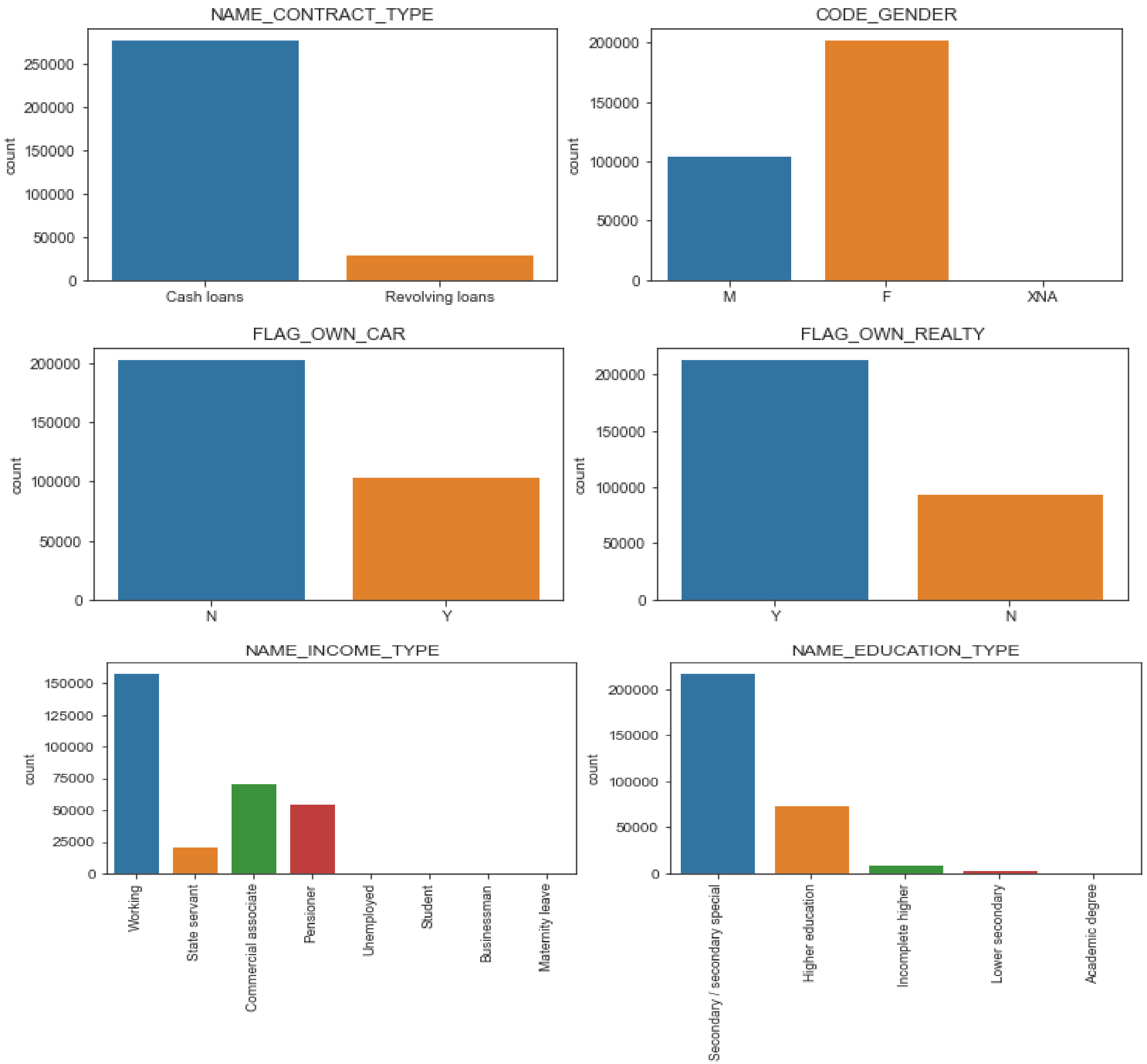
Outliers Inspection

- Continuous Variables



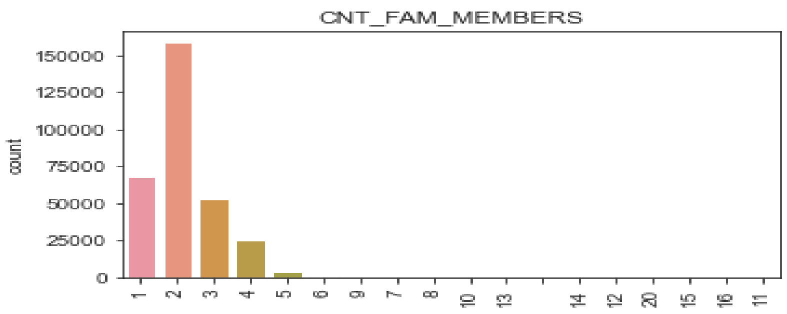
Data Inspection

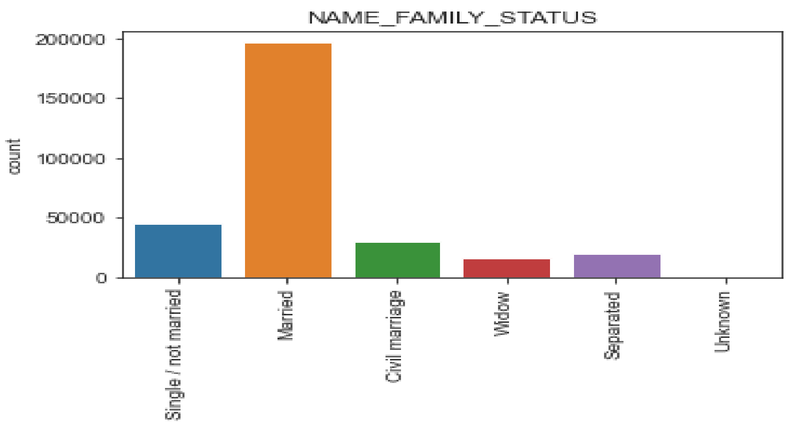
- Categorical Variables

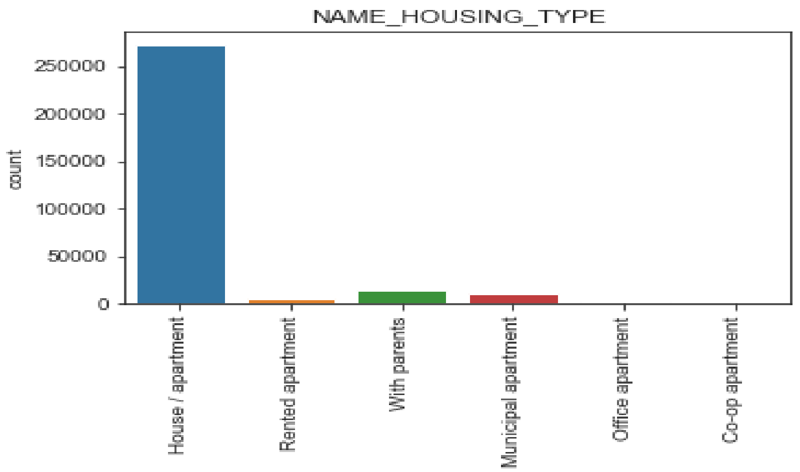


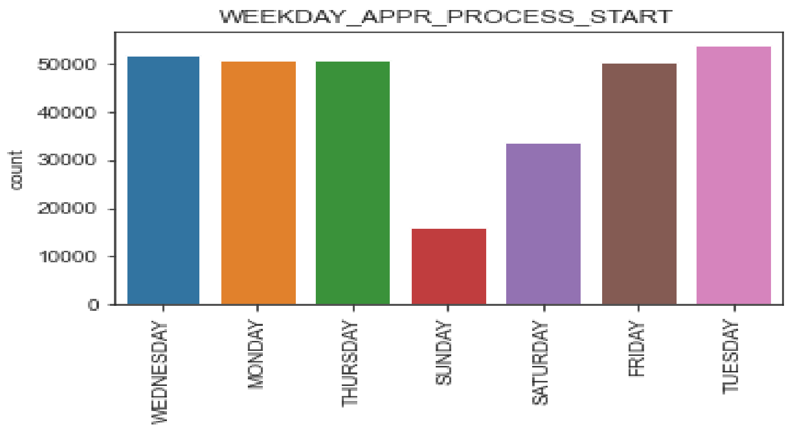
Data Inspection

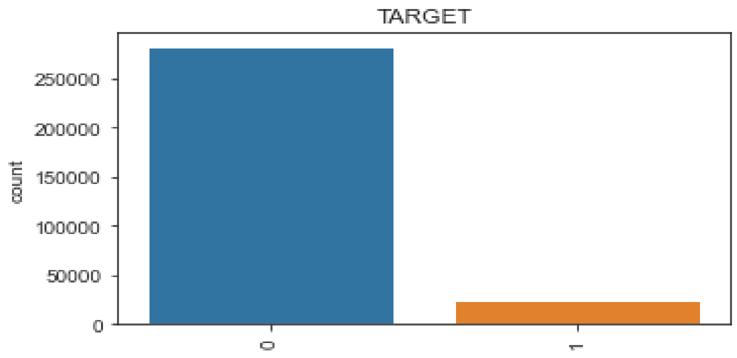
- Categorical Variables

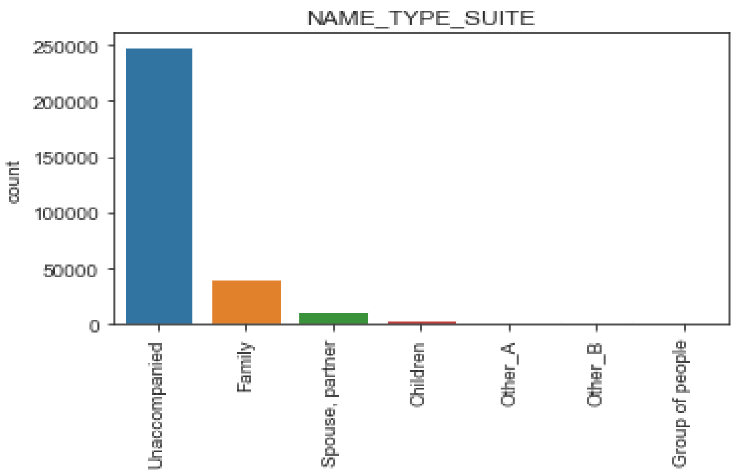






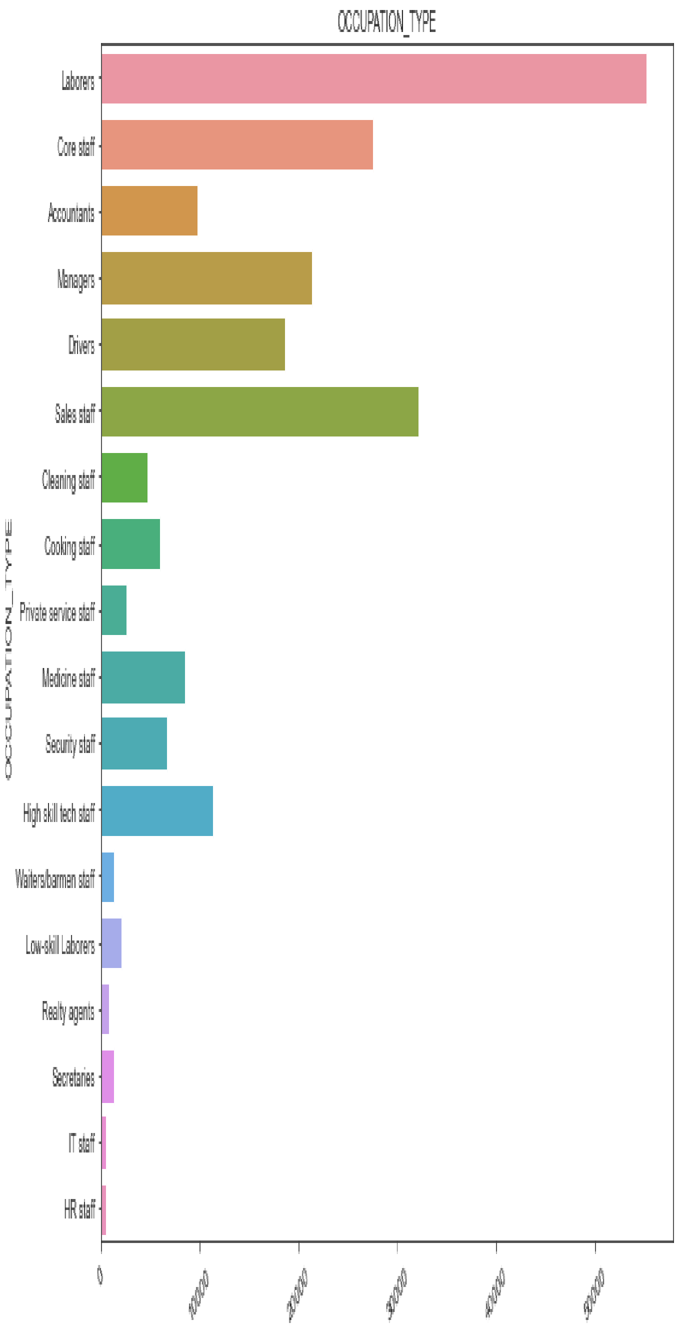


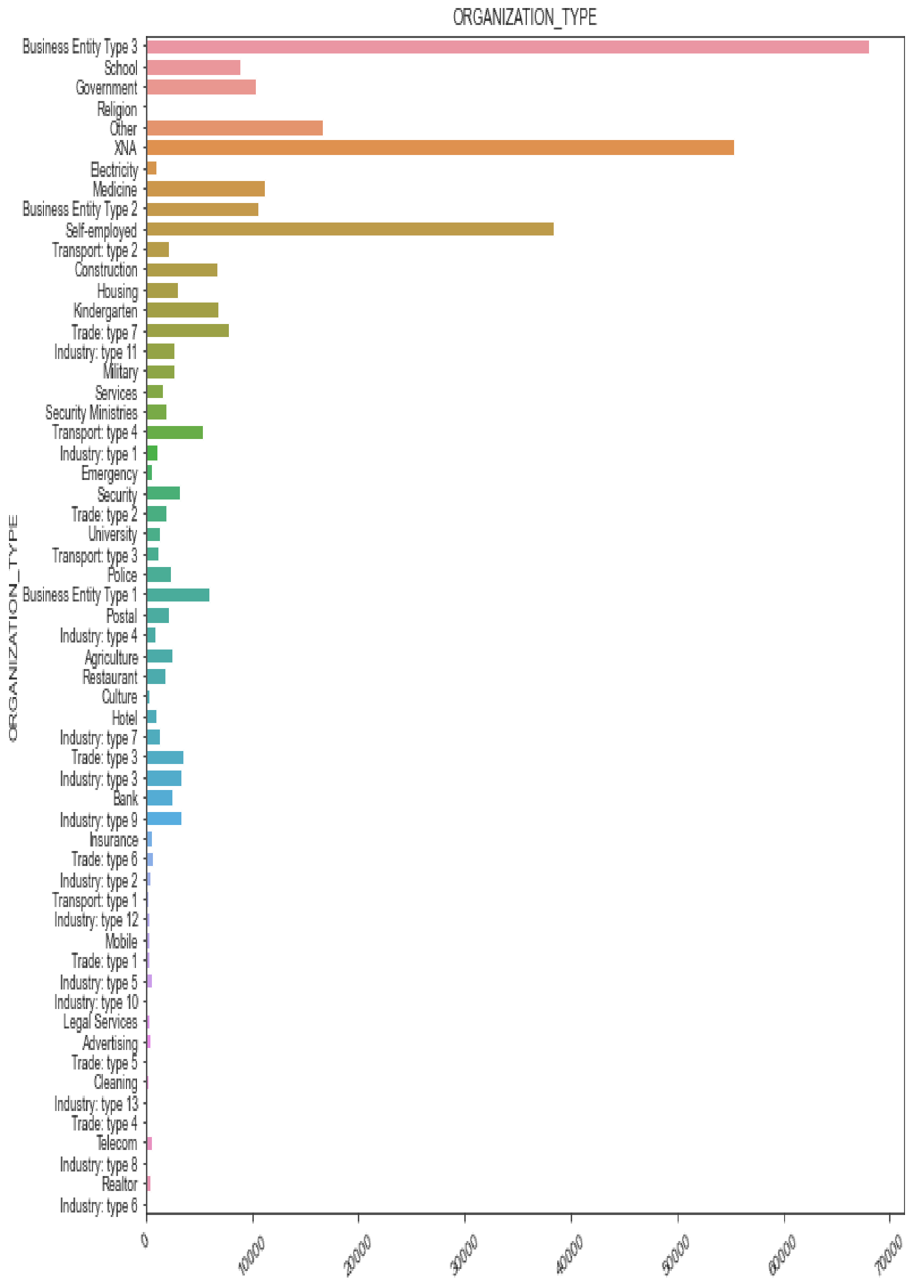




Data Inspection

- Categorical Variables





General Observations on Application data:

- Two types of Contracts - Cash Loans & Revolving Loans, where number of Cash Loans are much higher as compared to Revolving Loans.
- Women have taken more loans as compared to Men.
- Small population of applicants own a car.
- A large proportion of applicants own a house or apartment.
- Borrowers were mostly not accompanied by anyone when applying for the loan.
- Higher number of applicants are Laborers.
- Majority of applicants are Employed or Pensioner.
- Persons owning business or student constitute a Low volume of Loan Applicants.
- Majority of the Loan Applicants were Educated, Married and with 0 or max 2 children, while very Less applicants are having more than 2 children
- Majority of applicants are in Region rating of 2, assume a medium rating by credit bodies i.e. defaults are not very high.

Application _data Final Set for Analysis

The Application_data .csv, dataframe = dfn is now reduced to 56 columns and ready for EDA .

Data Type	No of Columns
Float	23
Integer	19
Object	13

Recommendations: Handling of Outliers

- AMT_INCOME_TOTAL 23 % > 202500 which leads to a +ve skewness of the data, we will retain this as there may be super rich applicants. Abnormally high values to be dropped as such cases are approx. .05 % of the population . On listing the percentile values, 98% data was in the acceptable range. 2 % of the data may be deleted.

- AMT_ANNUITY,AMT_CREDIT,AMT, AMT_GOOD_PRICE seem to have a linear relationship , hence large values will not be treated as outliers individually for now. Percentile value trend for all 3 are similar.

- DAYS_EMPLOYED 18 % = 365243 which appeared to be an error, but INCOME_TYPE shows 99 percent are Pensioners for this subset of data.
  We can impute the DAYS_EMPLOYED to 0 as pensioners are not employed.

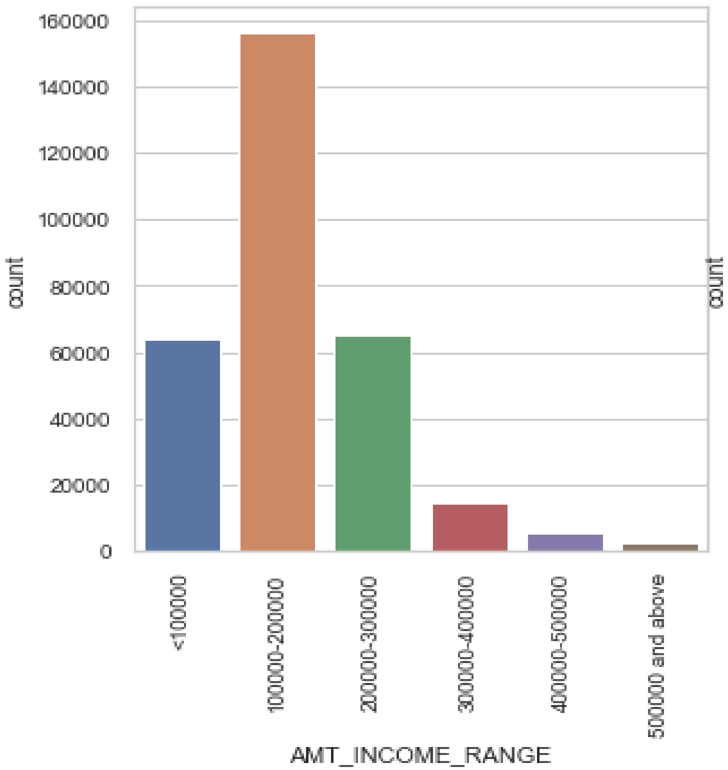
- ORGANIZATION_TYPE = XNA, since pensioners are not employed, XNA may be retained and will not an outlier. It can be interpreted as NA for this segment of applicants.
  Missing values can be imputed based in income bracket bins, imputing the mode of occupation type in the respective bins.

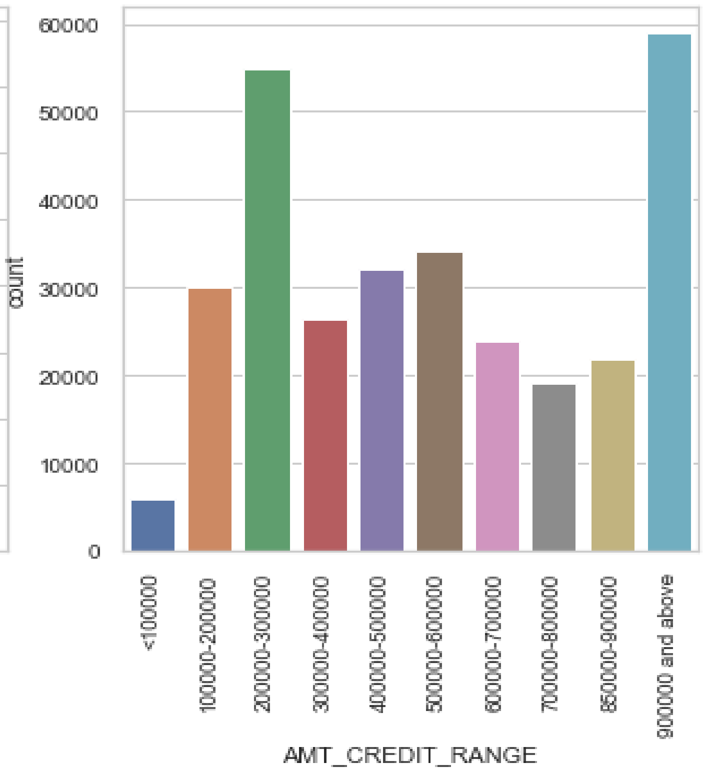
- EXT_SOURCE _2, 3: These are important criteria (Ratings from External agencies) , the mean value of the scores may be used as the final rating of a customer.

Binning –

AMT_INCOME_TOTAL
bins = '<100000', '100000-200000','200000-300000','300000-400000','400000-500000', '500000 and above'

2.AMT_CREDIT
Bins = <100000', '100000-200000','200000-300000','300000-400000','400000-500000', '500000-600000', '600000-700000','700000-800000','850000-900000','900000 and above'





Observations after Binning

- AMT_INCOME_RANGE :
  - Max loans by Applicants with lower & middle income (<10000 -300000)
  - Affluent applicants have a very low loan frequency.

- CREDIT_RANGE :
  - High frequency for CREDIT ranging 20000 -300000 , > 900000
  - Possible defaulters in HIGH Credit range which may become apparent during comparative analysis of defaulter & non defaulters.

Recommendation : Analysis on the above findings for default behavior.

4. EDA

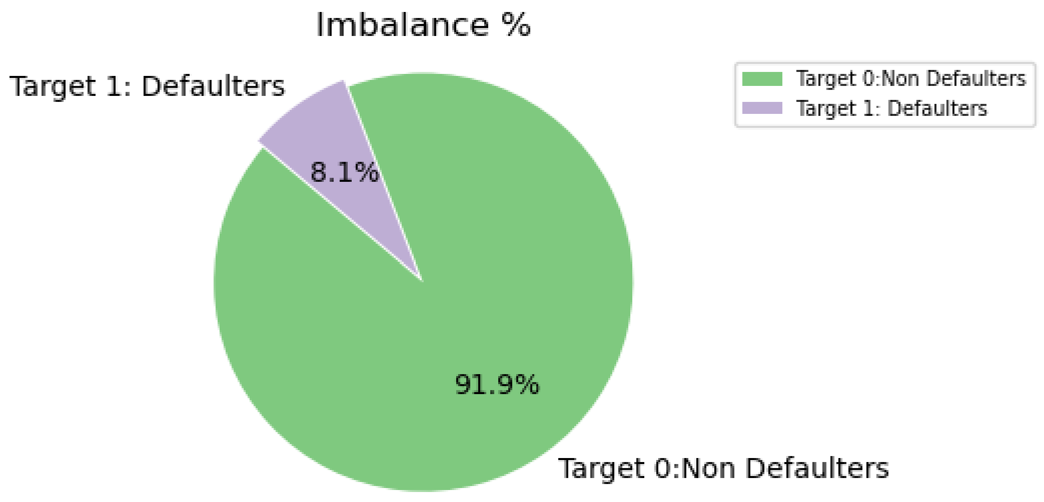
Imbalance %

TARGET Variable

Compute the imbalance % in the variable ‘TARGET’, which has data categorizing the applicants as follows :

Target 0 = Non Defaulters

Target 1 = Defaulters



Data is Imbalanced

TARGET 0 = 91.9 % - Non-Defaulters - Majority Class
TARGET 1 = 8.07% - Defaulters - Minority class

Imbalance ratio = Size of Minority Class / Size of Majority Class
= .0878

Due to imbalanced dataset, we will use log scale for plotting the graphs
Reason : to handle the skewness in the data (which is towards Target = 1)

Since there are very few people who default, the data related to their background & behavior is important to predict the likelihood of a default

A need to focus on outliers too, as they could be the differentiating factor between Defaulter and Non-Defaulter.

Univariate Analysis for Categorical Variables

To understand the behavior of Defaulters & Non-Defaulters , we did the Comparative Univariate Analysis on following Categorical Variables:

Target = 0 Non-Defaulters , Target = 1 Defaulters

- NAME_CONTRACT_TYPE
- NAME_TYPE_SUITE
- OCCUPATION_TYPE
- CNT_CHILDREN
- FLAG_OWN_CAR
- NAME_FAMILY_STATUS
- ORGANIZATION_TYPE
- CODE_GENDER
- NAME_EDUCATION_TYPE
- NAME_INCOME_TYPE
- REGION_RATING_CLIENT
- FLAG_OWN_REALTY
- NAME_HOUSING_TYPE

Few of them are shown in the upcoming slides….>>>>

Univariate Analysis on Categorical Variables

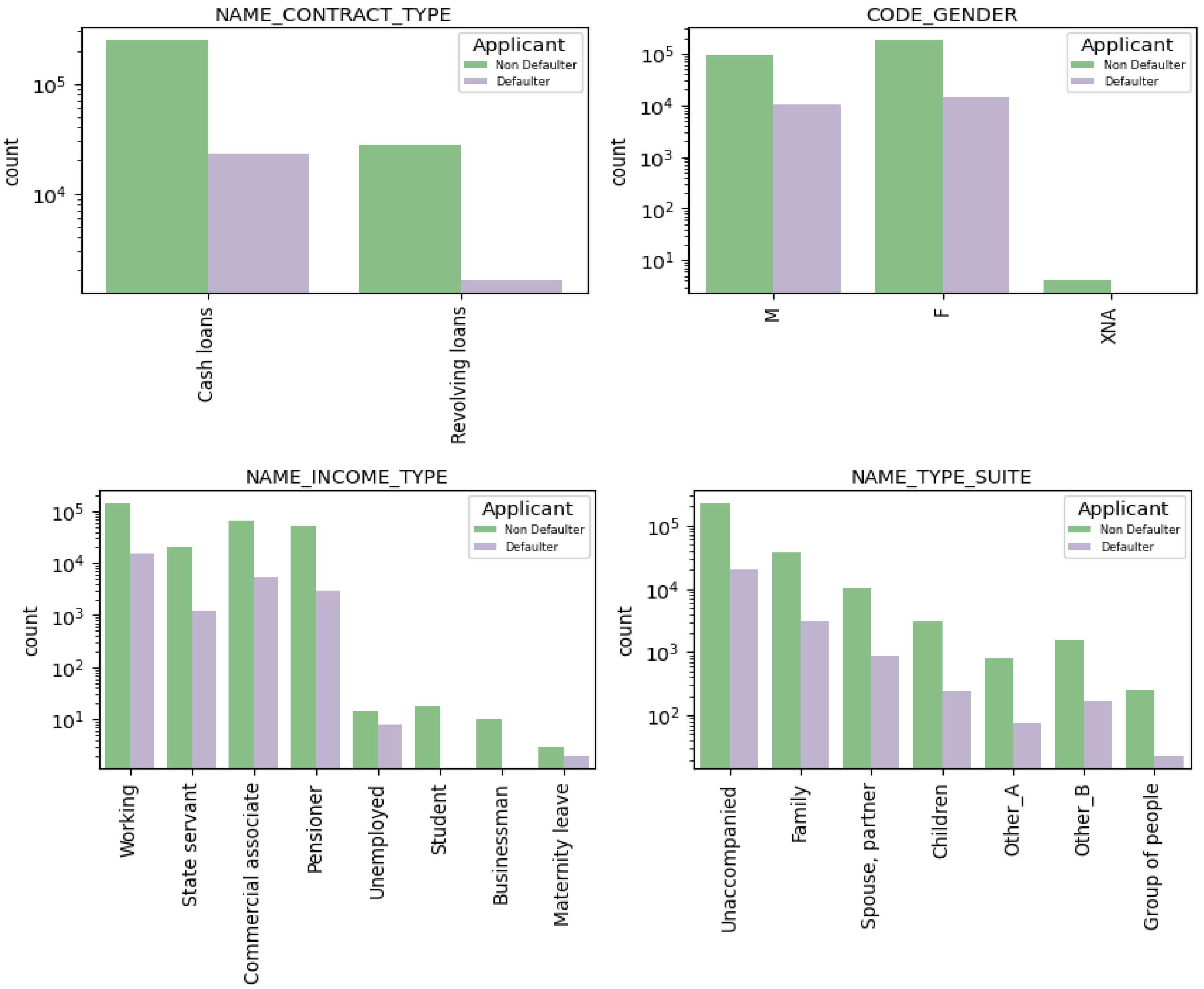
Contract Type:
•Cash Loans is preferred choice across all.
•Cash Loans : 1/3rd are Defaulters
•Low Default rate on Revolving loans

Gender
•Similar distribution of Male /Female applicants for both Defaulter and Non-Defaulters. Approx. .40 % Default rate
•Females have a slightly higher tendency to repay loans on time

INCOME TYPE :
•Major applicants are from : Working , Commercial associate, Pension, State Government.
•This indicates that applicant with a steady source of income tend to take loans frequently and have a high rate of default
•Students, Unemployed , Businessmen have a very low frequency of taking loan - No Defaults
• Unemployed , maternity leave - high defaulters

TYPE_SUITE
•Unaccompanied applicants display high rate of default

•Persons having family are also loan seekers and prone to default



Univariate Analysis

FAMILY
Hight Loan frequency & default rate amongst :

- Applicants with < 3 children,
- Applicants with Family size < 5 members , highest in size 2

Possible reasons :

- Families with two members - maybe newly weds setting up their house, hence avail loans more frequently

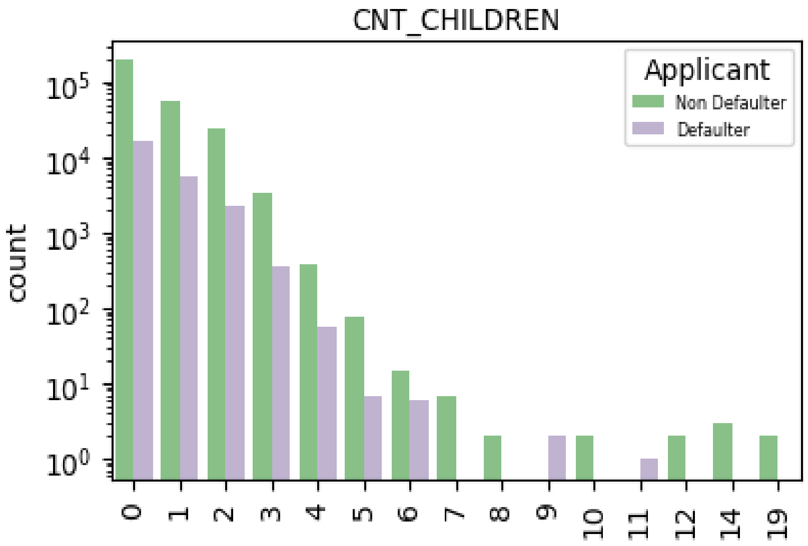
- Applicants with children may require education loans

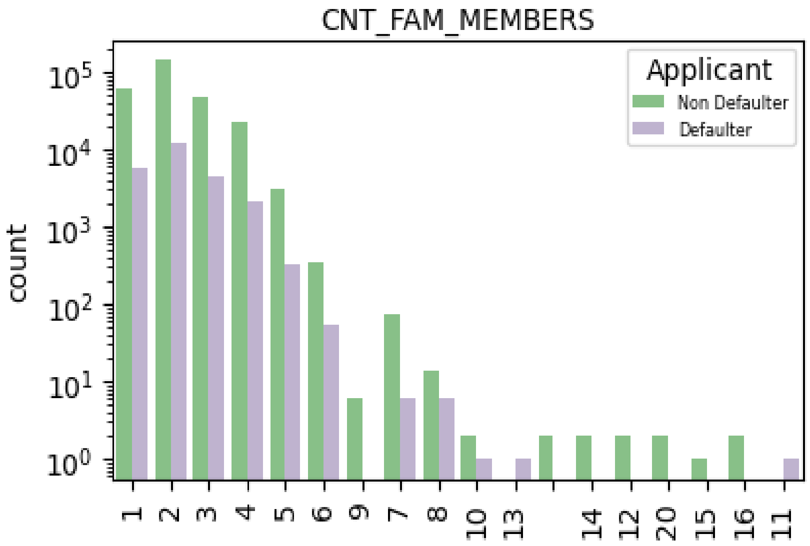
OCCUPATION:

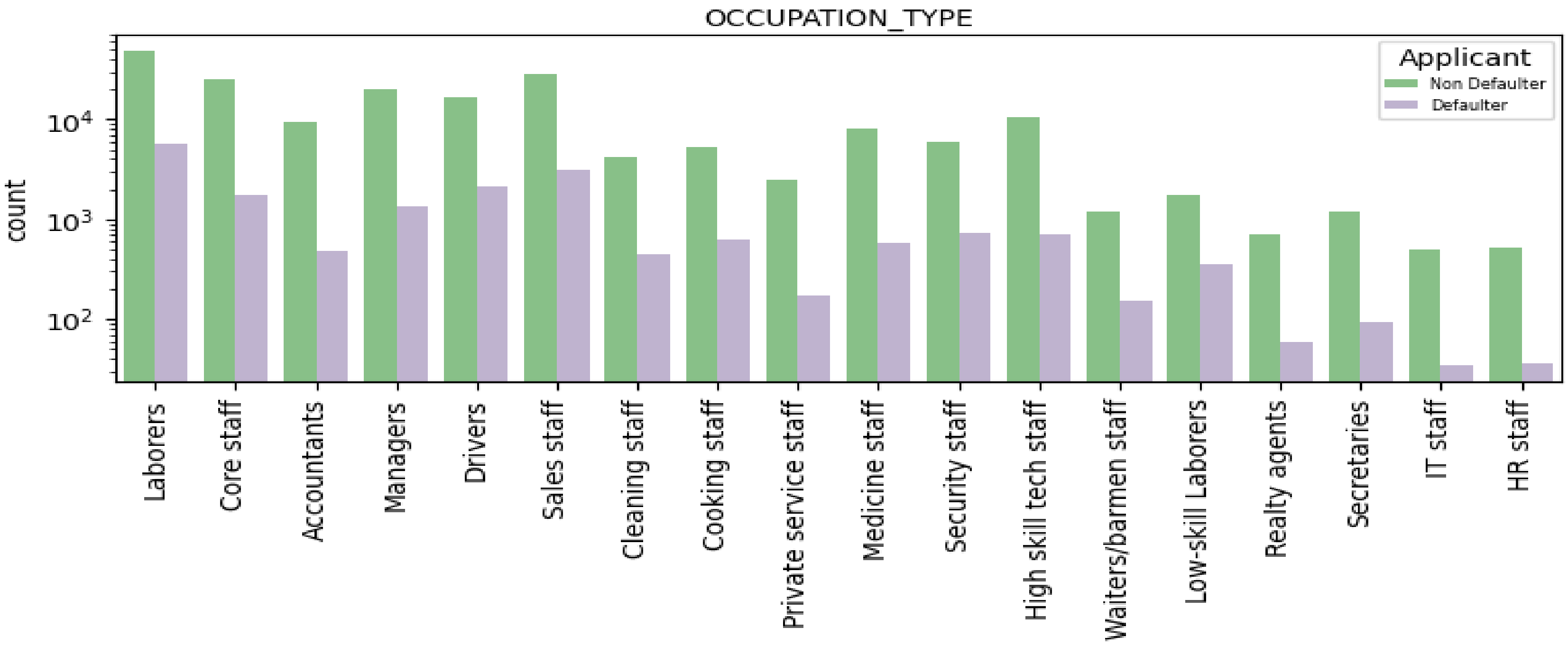
- Loan Applicants mainly : Labourer, Sales staff, Core staff, Drivers ,Managers

- High Default rate in Labourers, Sales Staff, Drivers , Managers, Core Staff - may be the low-income category

- Low Default amongst IT Staff, HR Staff Realty Agents, High Skill tech staff - they may be having a reasonably good income





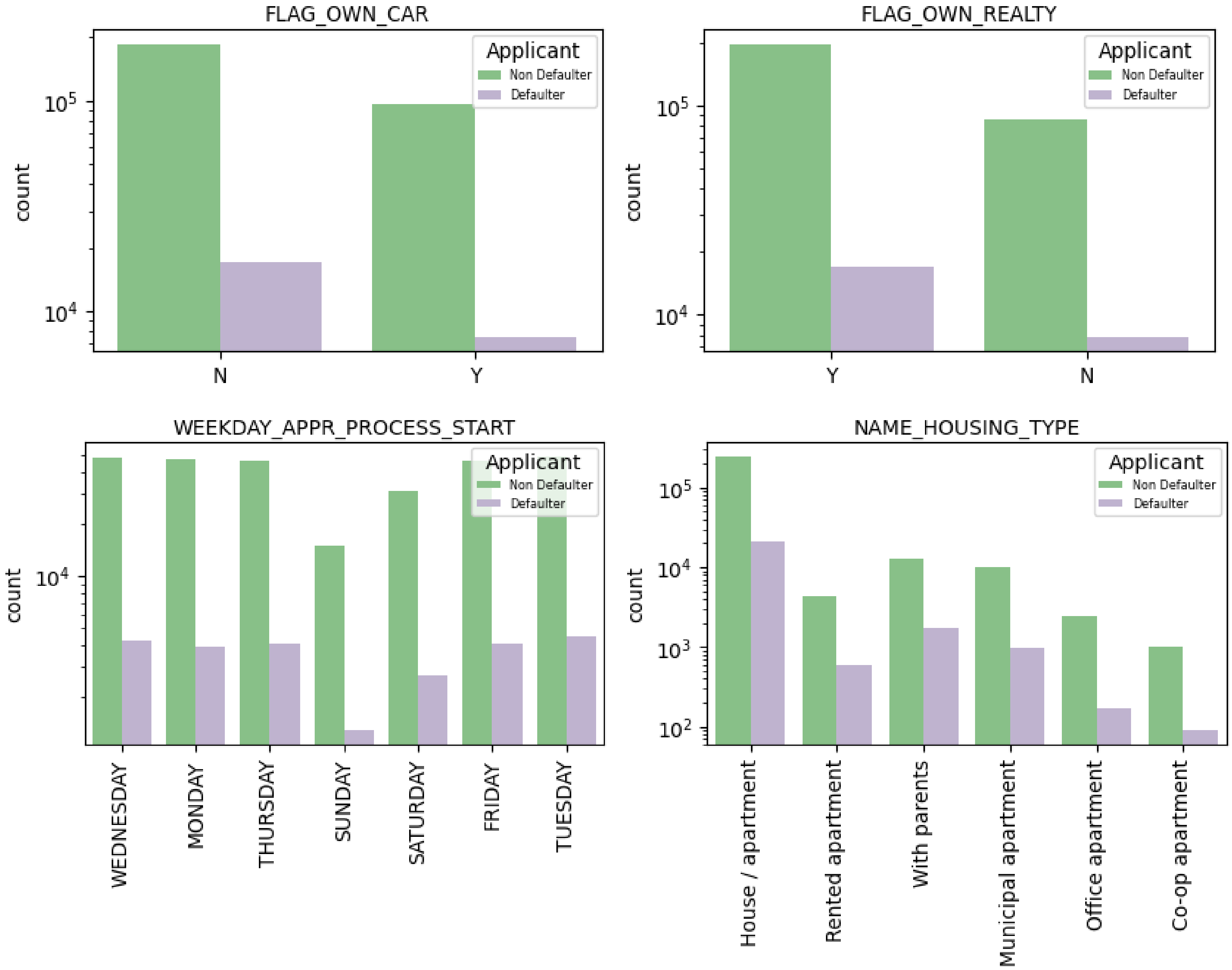


Univariate Analysis

ASSETS:

- Applicants not owning cars default more, this may indicate the applicant may be in the low-income category.
- Applicants owning a House tend to default more maybe the applicant is already having a house loan.
- Housing type: Applicants staying in own house / apartments indicate good income , but default rate is approx. 40 %
- Applicants living with parents are second in the default rank - possibly aged parents have health issues and expenses increase
- Low rate of default amongst people sharing apartments: expenses are possibly shared leading reduce cost of living.

✓ Check loan history of the applicant, and socio-economic information



Univariate Analysis

AMT_INCOME_RANGE

- Highest Default/Highest Loan applicants : Income range 100000-200000 which could represent Lower Middle Class who have limited resources

- Next category of defaulters - Below 100000 - students/widows , 200000-300000 - Middle Income group

✓ Their loan history, occupation, Income source to be checked during loan application

AMT_CREDIT_RANGE:

- Approx. 40 % defaulters in various credit amt ranges, credit amount < 100000 has least defaulters - could be revolving loans

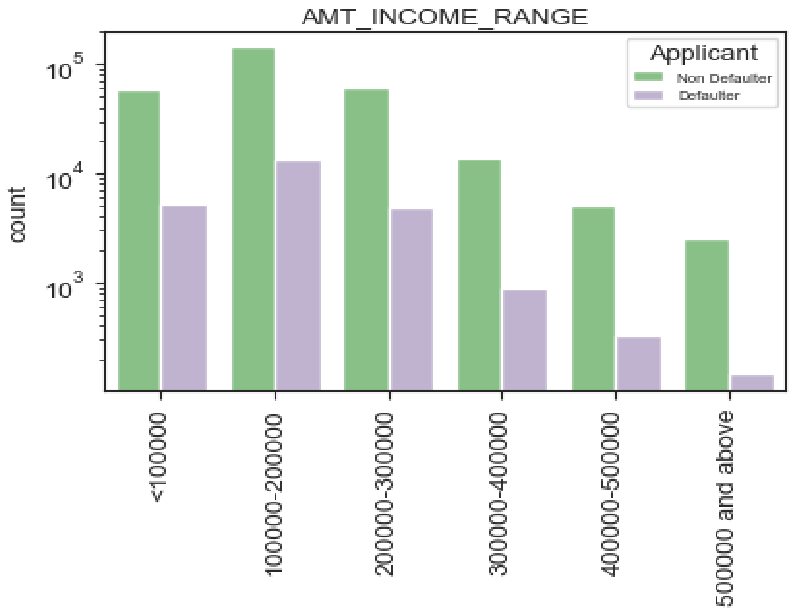
REGION RATING/W CITY of Client:
Assumption : Scale = 1.High 2.Medium 3 Low

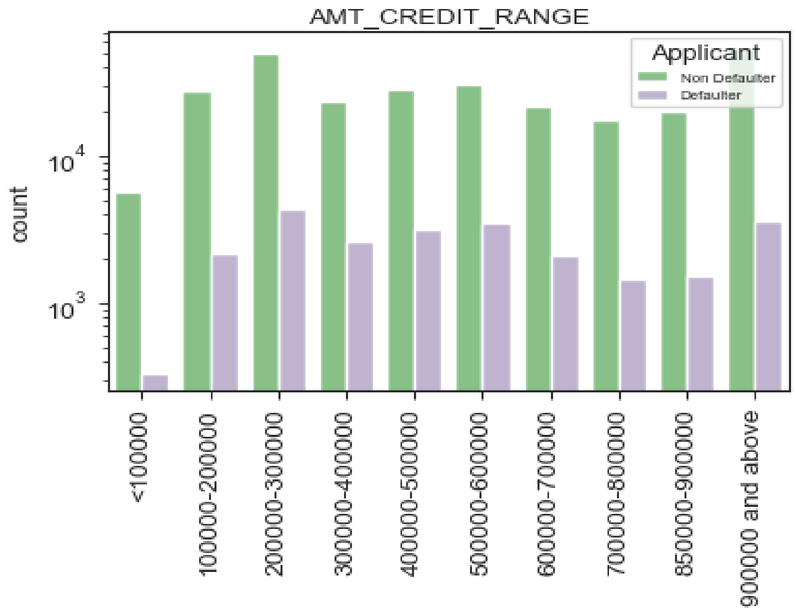
- Low defaults in High Rated regions/cities

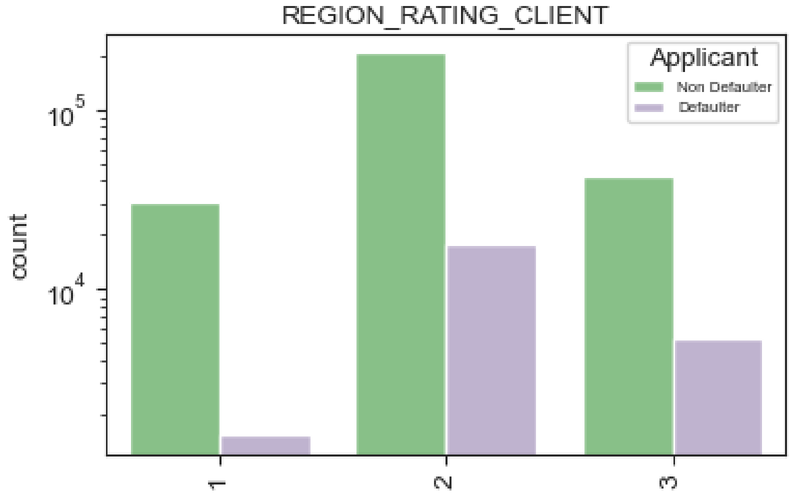
✓ Highest defaults in Medium Rated region/cities

- Moderate defaults in Low Rated Region/cities

- Applicants residing in Medium rated regions/cities apply for more loans as compared to 1,3 regions







Univariate Analysis

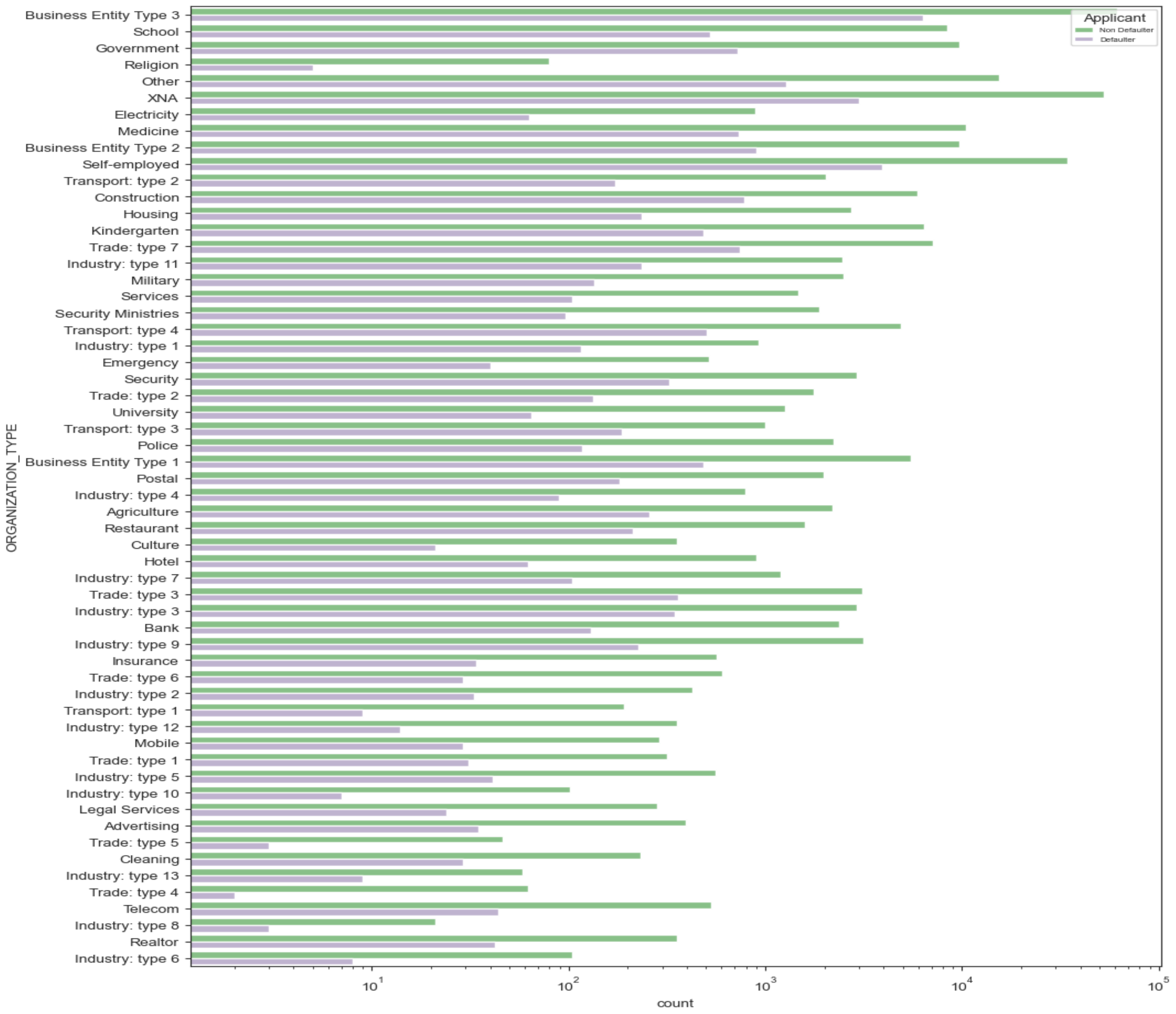
ORGANIZATION_TYPE

XNA, Others will have to handled using some logic or more information – high %

- Highest Loan applicants : Business Entity Type 3 may be sole proprietor company like a startup which need heavy investment, Self Employed, Medicine, XNA, OTHERS
- High Defaulters : Business Entity Type 3, Self Employed, Trade ,Medicine
- Moderate Defaulters : Government, Schools

✓ Watchout for Small Businesses , Self employed, Small Traders, Persons in Medicine

✓ Government , School applicants – to be careful



Univariate Analysis Continuous Variables

Trend of Default /Non-Default same for the foll.

- INCOME Applicant Income is right skewed
- CREDIT : Credit amt is in the in the lower range and also a very high value range - pls refer to CREDIT_RANGE plot
- ANNUITY - Maximum concentration 300000-500000
- GOODS_PRICE - Shows similar distribution as Credit& Annuity Defaulters show a similar trend

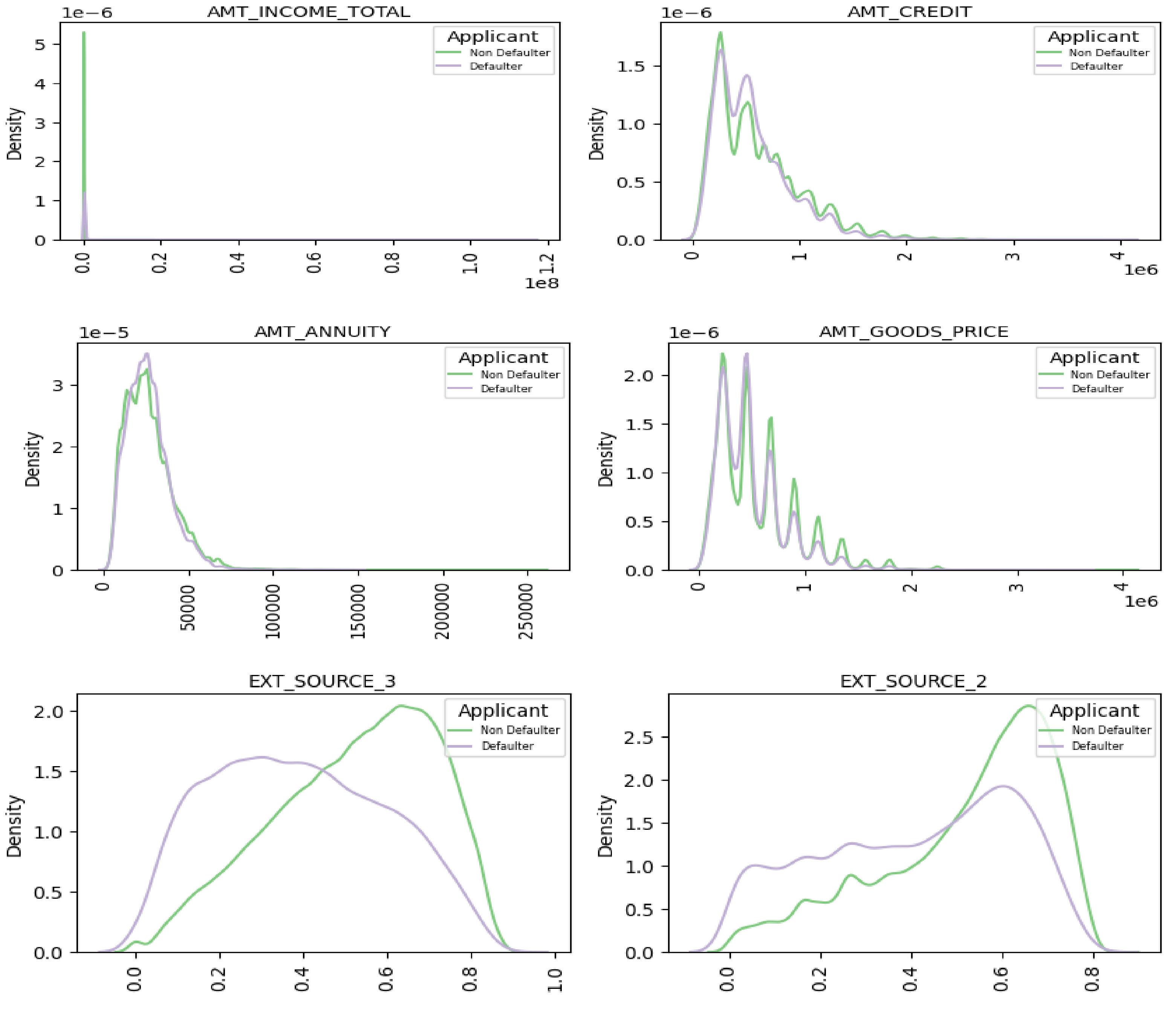
Ratings3 : EX_DATA_SOURCE3

- Defaulters have a low scores 0.2 to 0.4
- Non-Defaulters have a higher score 0.6 to 0.8

Ratings2 :EXT_DATA_SORUCE2

- Defaulter have low scores 0.4 to 0.6
- Non-Defaulter have higher score of 0.6 to 0.8

✓ Ratings from External Resource is an important differentiator for loan approval process



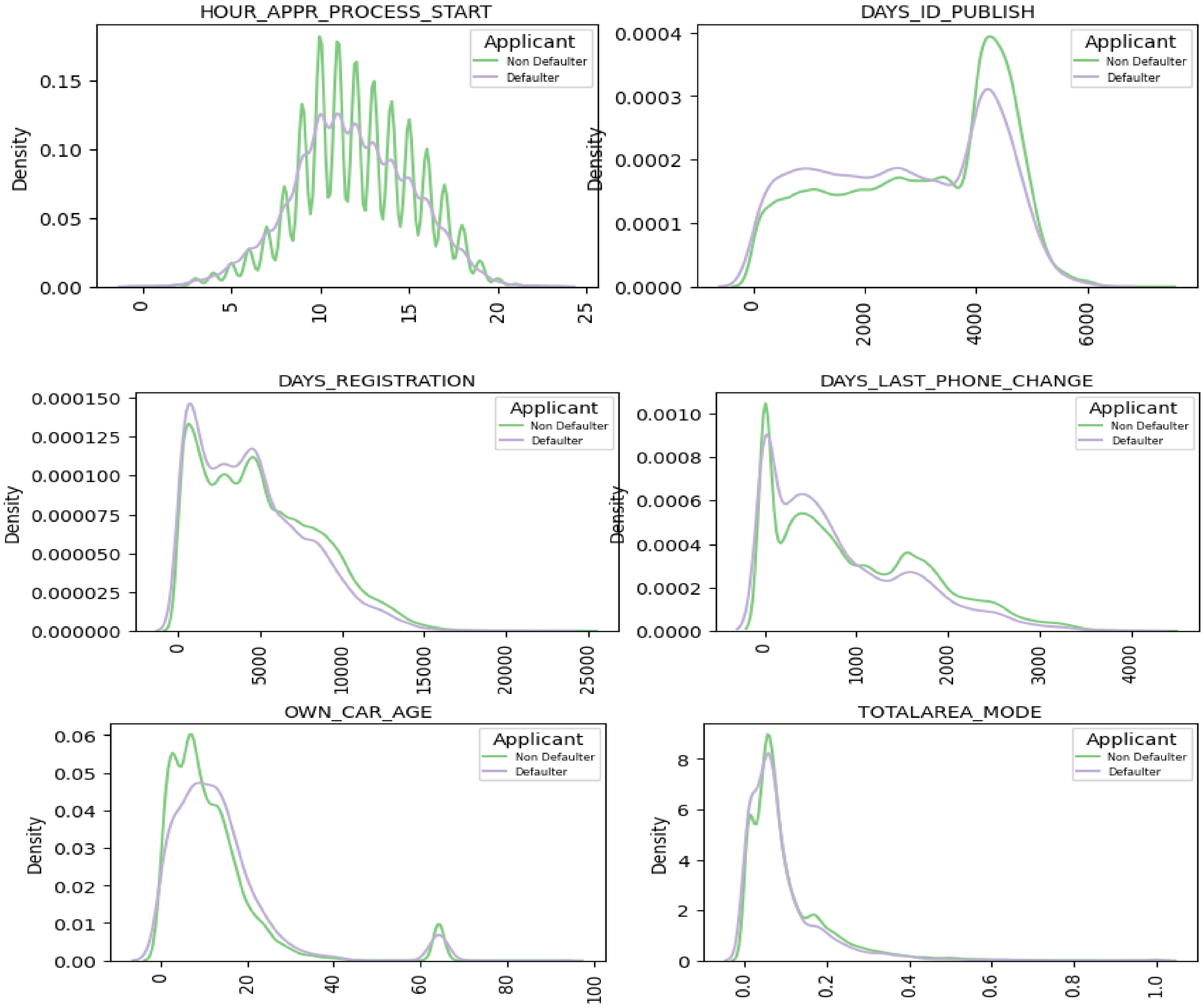
Univariate Analysis Continuous Variables

DAYS ID PUBLISHED :
- Not much to infer

DAYS REGISTRATION
- Defaulters have a higher tendency to change their registration during loan filing
- ✓ This behavior may be important

DAYS LAST PHONE CHANGE
- Not much to infer

CAR AGE
- Defaulters have slightly older cars
- ✓ Banks may use this information in case of default



Bivariate Analysis

Credit Amount - Education Status

Defaulters :

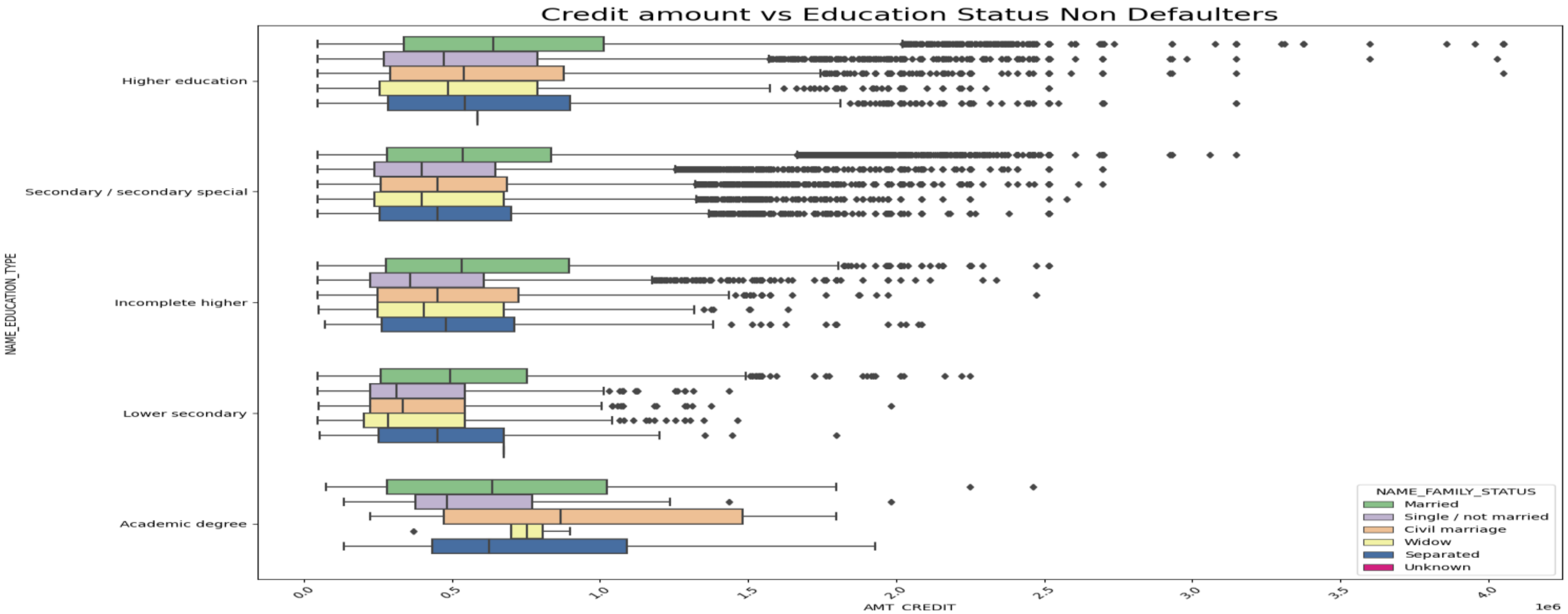
- Highest tendency in Married , Highly educated with degrees & high credit amount
- ✓ They may be having past loans for education

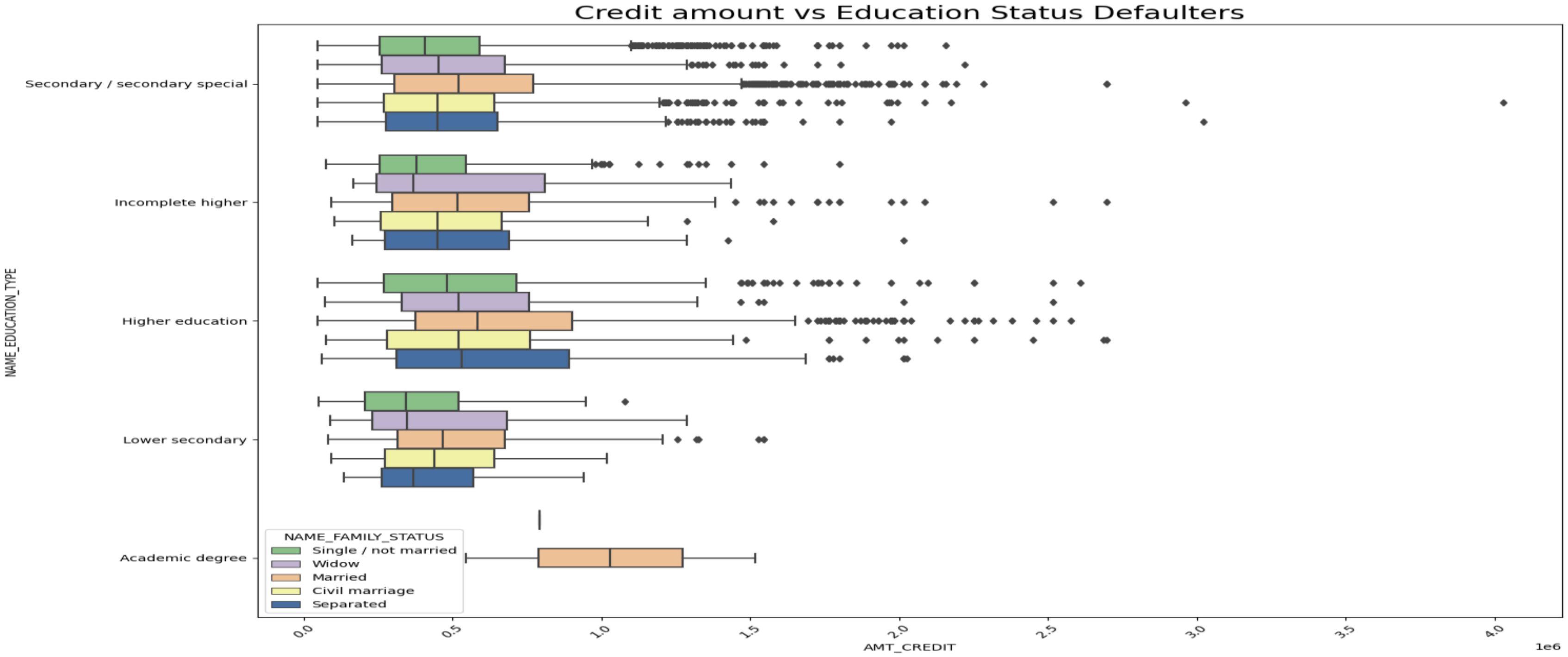
- Widows - tendency to default in all lesser education category

- Married , Separated Non degree holders tend to default more
- ✓ Separated persons may be burdened with alumni or loss of income

Non-Defaulters :

- High Credit Amt – Degree holders



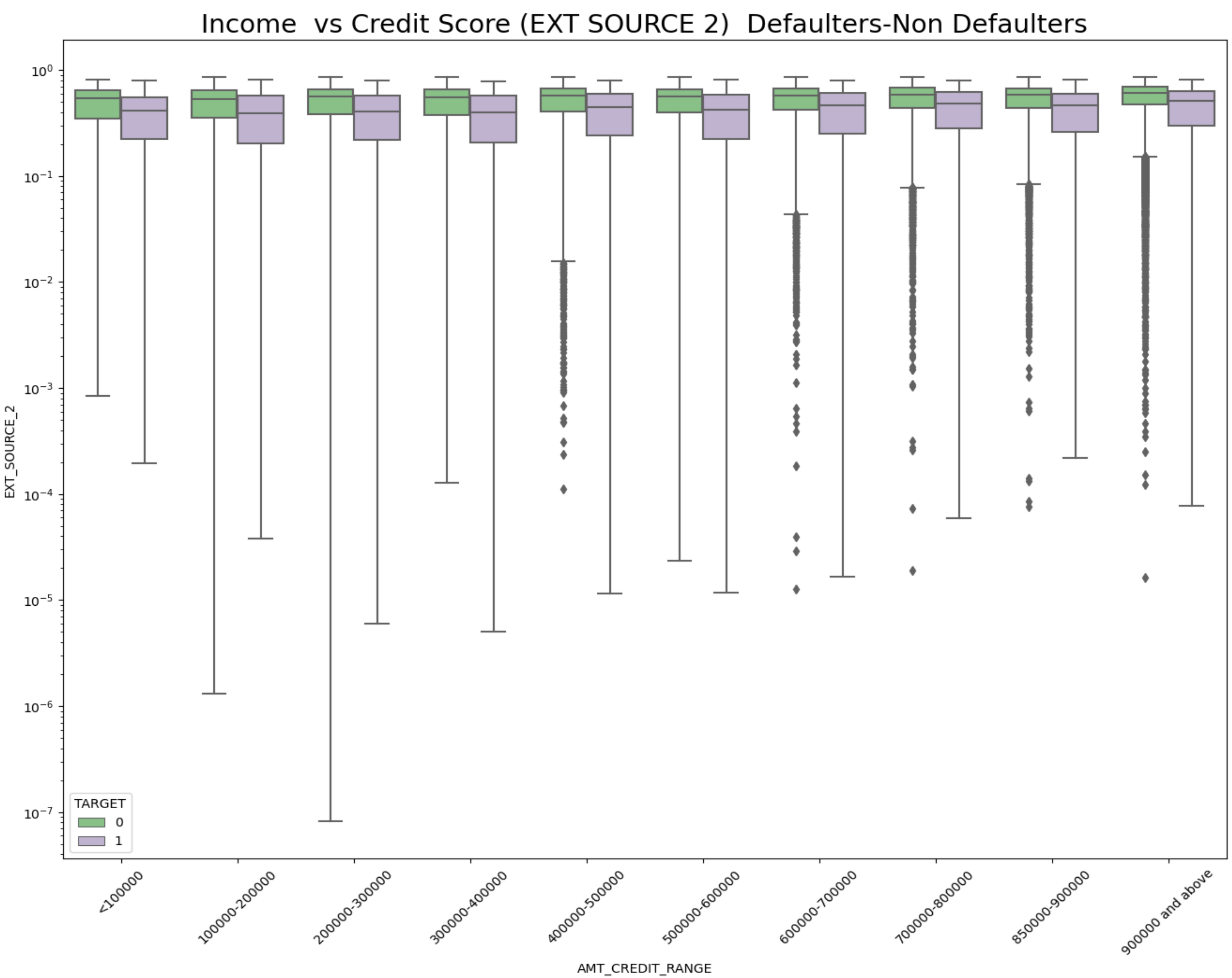


Bivariate Analysis

Credit Amount - Credit Score (EXT_SOURCE_3)
SOURCE_3 taken as it had least null values 1 %

- Defaulters have a low credit rating in all Credit ranges
- Defaulter in the Income range of 100000 to 300000 have extremely low scores & default the most - Need to watch as per earlier

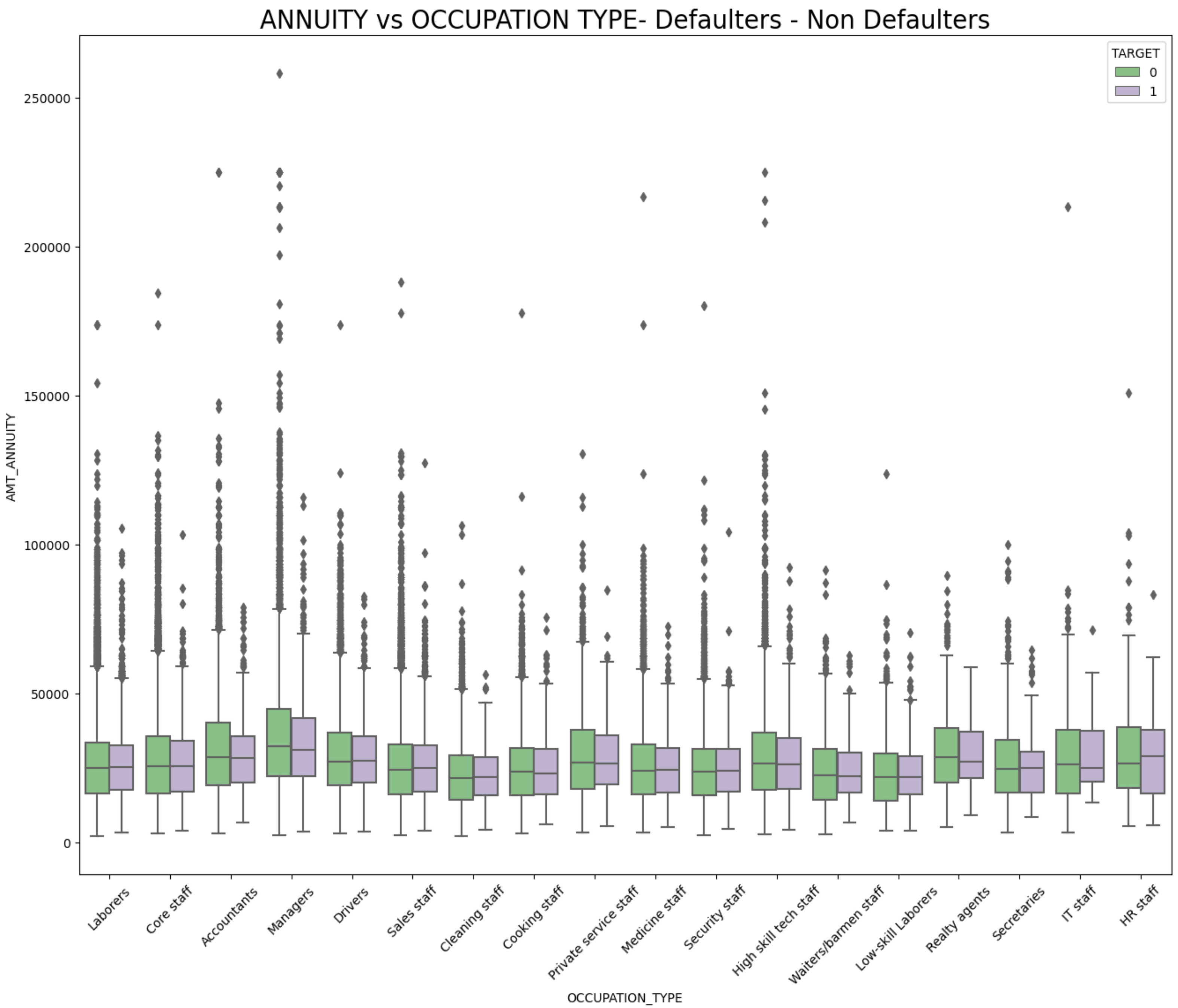
✓ Credit Score : Important criteria for loan approval



Bivariate Analysis

Annuity – Occupation Type

- Highest Annuity is being Accountants & Managers default avg is a bit less as compared to other occupations
- ✓ Have in-depth knowledge of finance , hence would have planned accordingly
- High Skill Tech Staff are the next highest Annuity payers and default rate also lower
- HR Staff are the highest defaulters
- Low rate of default – cleaning/Low skilled labourers and they do not have high Annuity pay out
- ✓ Education, Income play an important role in default tendency, quite obvious



Bivariate Analysis

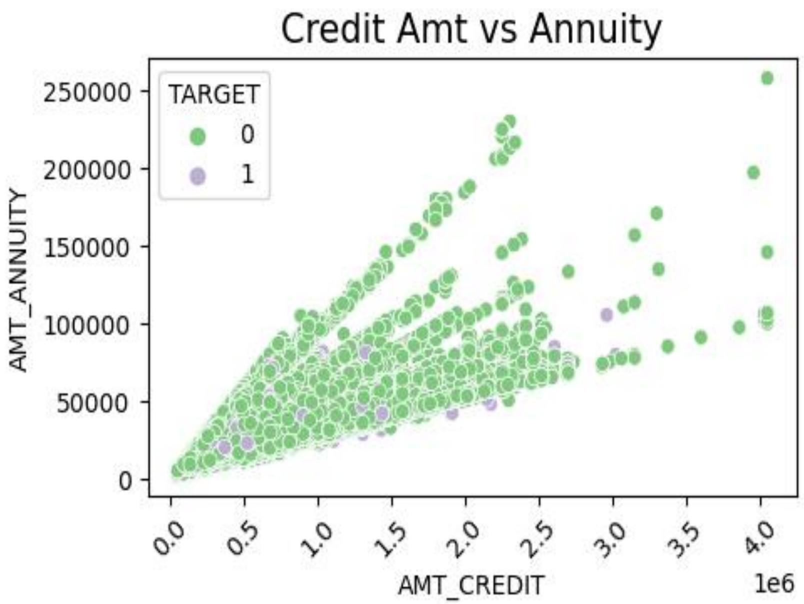
Credit Amt - Annuity, Credit Amt vs Goods Price

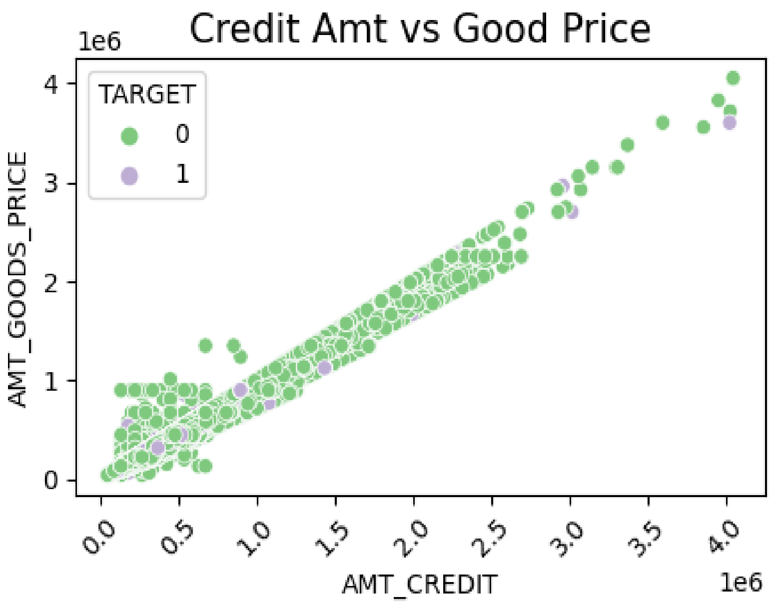
- Linear relationship evident which supports our earlier observation on outliers , hence outliers are genuine date

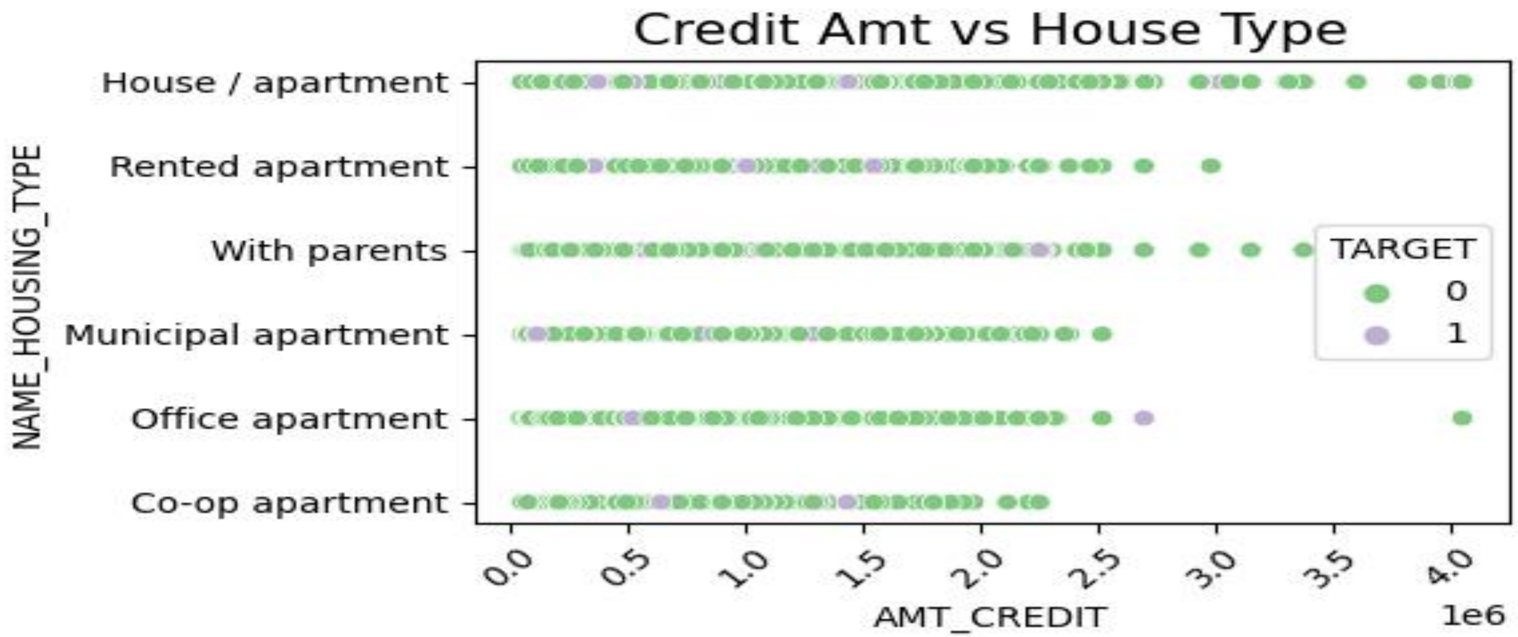
Credit Amt – Housing Type

- Applicants staying with parents have high credit amts and default more – supports the univariate analysis observation

- Applicants owning apartments have high credit amt but low rate of default ie Affluent







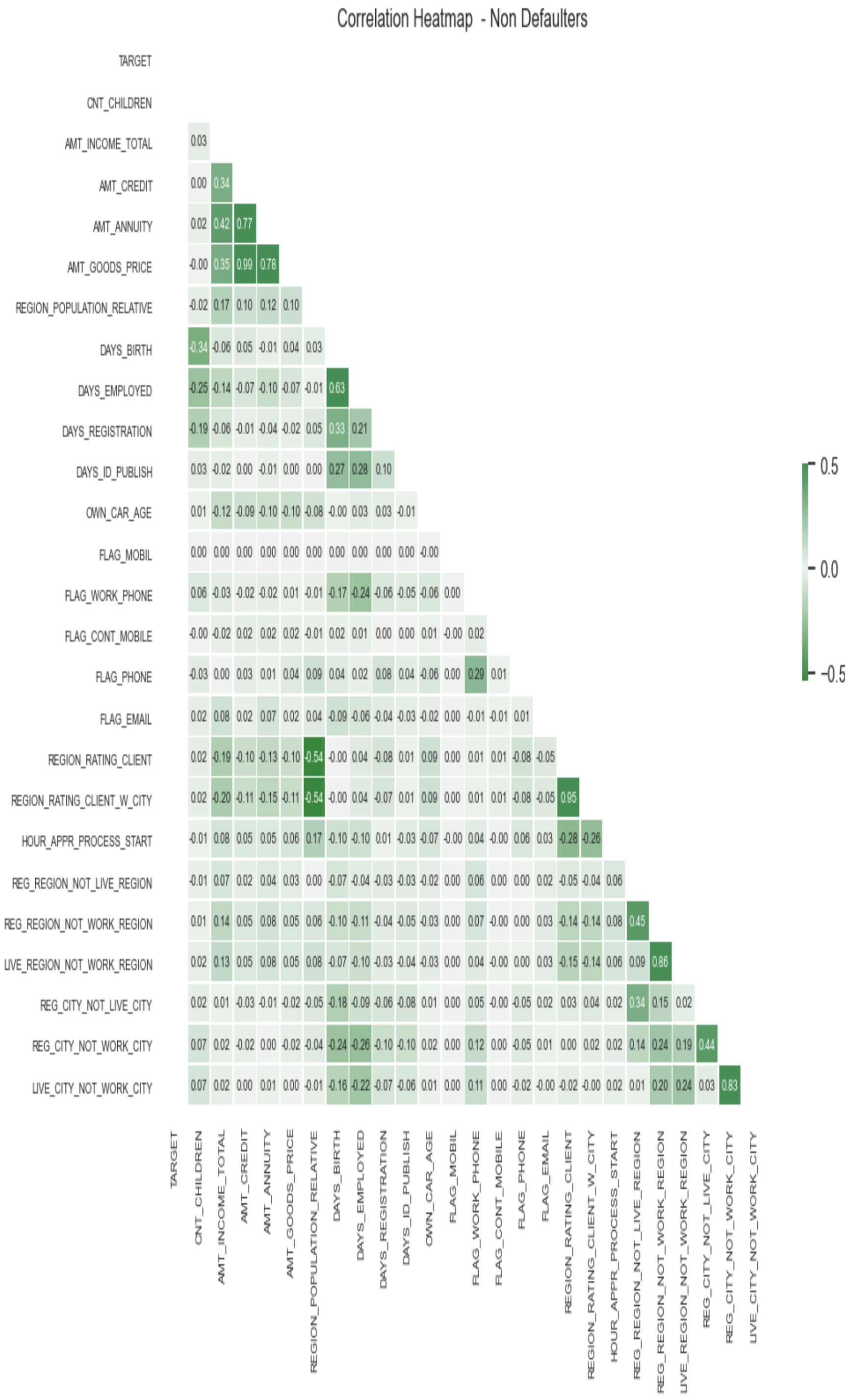
CORRELATION – Heat Map

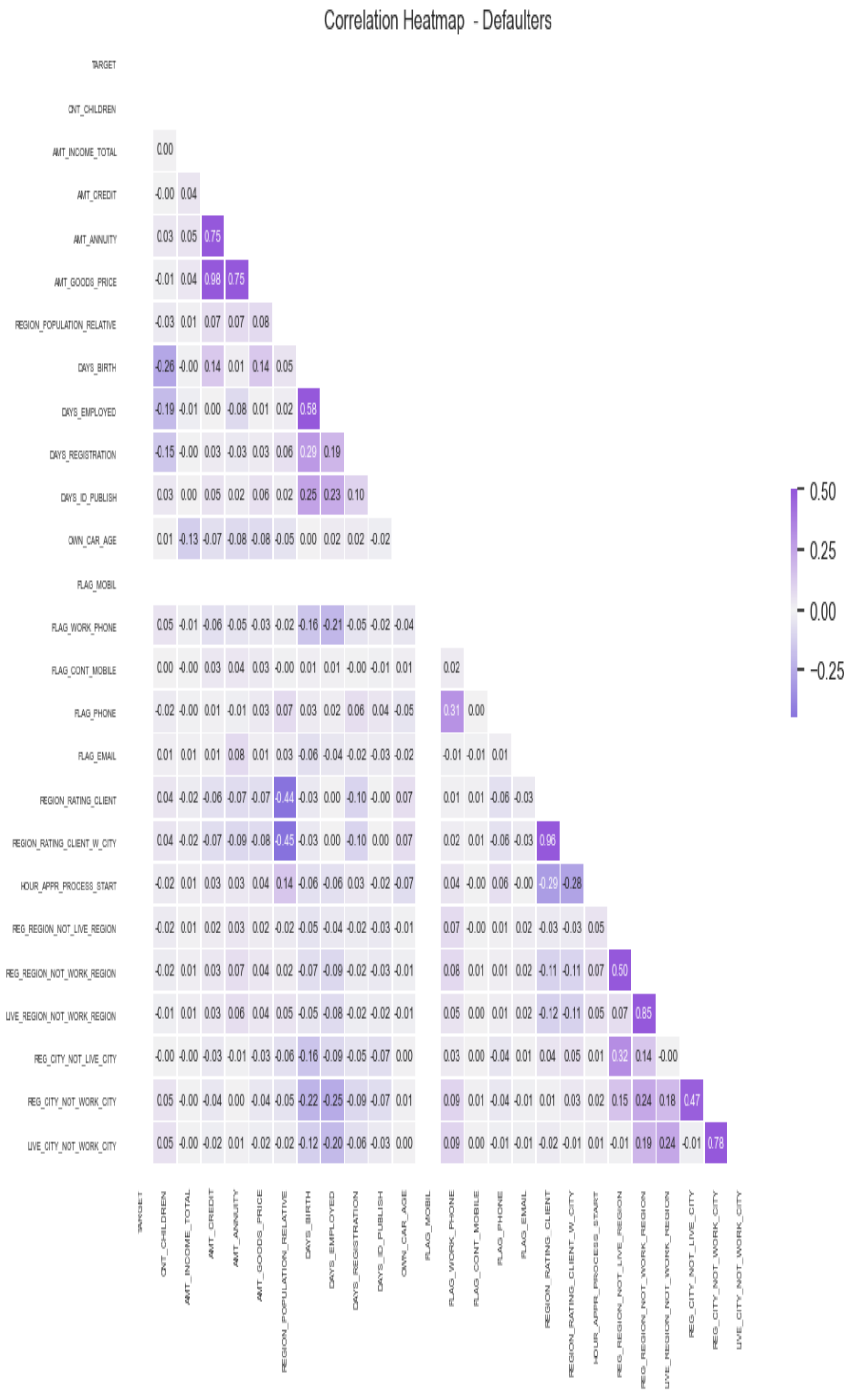
- The Heat Map of Defaulters / Non-Defaulters are similar

- Associations amongst the variables are not very strong

- Strong Correlations observed :
- Credit, Annuity, Goods Price
- Region and City ratings
- Employment Duration & Age

Differentiation between Defaulters and Non-Defaulters cold not be ascertained





CORRELATION

- Top 10 Correlations:

- Goods Price, Credit Amt
- Loan default tendency of neighborhood customers

- Annuity vs credit

- Region Score and Region, City score

- Applicant working in the city of loan application , different city

Top 10 Correlations – Non-Defaulters

	Var-NDF1	Var-NDF2	Correlation
611	FLAG_EMP_PHONE	DAYS_EMPLOYED	1.00
1494	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	1.00
262	AMT_GOODS_PRICE	AMT_CREDIT	0.99
879	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.95
1538	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.86
1055	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.86
1187	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.83
263	AMT_GOODS_PRICE	AMT_ANNUITY	0.78
219	AMT_ANNUITY	AMT_CREDIT	0.77
395	DAYS_EMPLOYED	DAYS_BIRTH	0.63

Top 10 Correlations – Defaulters

	Var1-DF	Var2-DF	Correlation
1494	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	1.00
611	FLAG_EMP_PHONE	DAYS_EMPLOYED	1.00
262	AMT_GOODS_PRICE	AMT_CREDIT	0.98
879	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.96
1538	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.87
1055	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.85
1187	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.78
263	AMT_GOODS_PRICE	AMT_ANNUITY	0.75
219	AMT_ANNUITY	AMT_CREDIT	0.75
1363	TOTALAREA_MODE	FLOORSMAX_MEDI	0.64

CONCLUSIONS

Application Data :

- Analysis shows that Defaulters and Non-Defaulters follow almost similar variable range.
- Data on defaulters very less to really come out with strong differentiating patterns.
- The external source rating 2 and 3 looks like a strong driving factor to identify defaulters.

Key Takeaways to be considered for Loan Processing :

- Credit Score
- Socio economic status
    - Income
    - Education
    - Occupation
    - Assets : House, Car
    - Residential locality

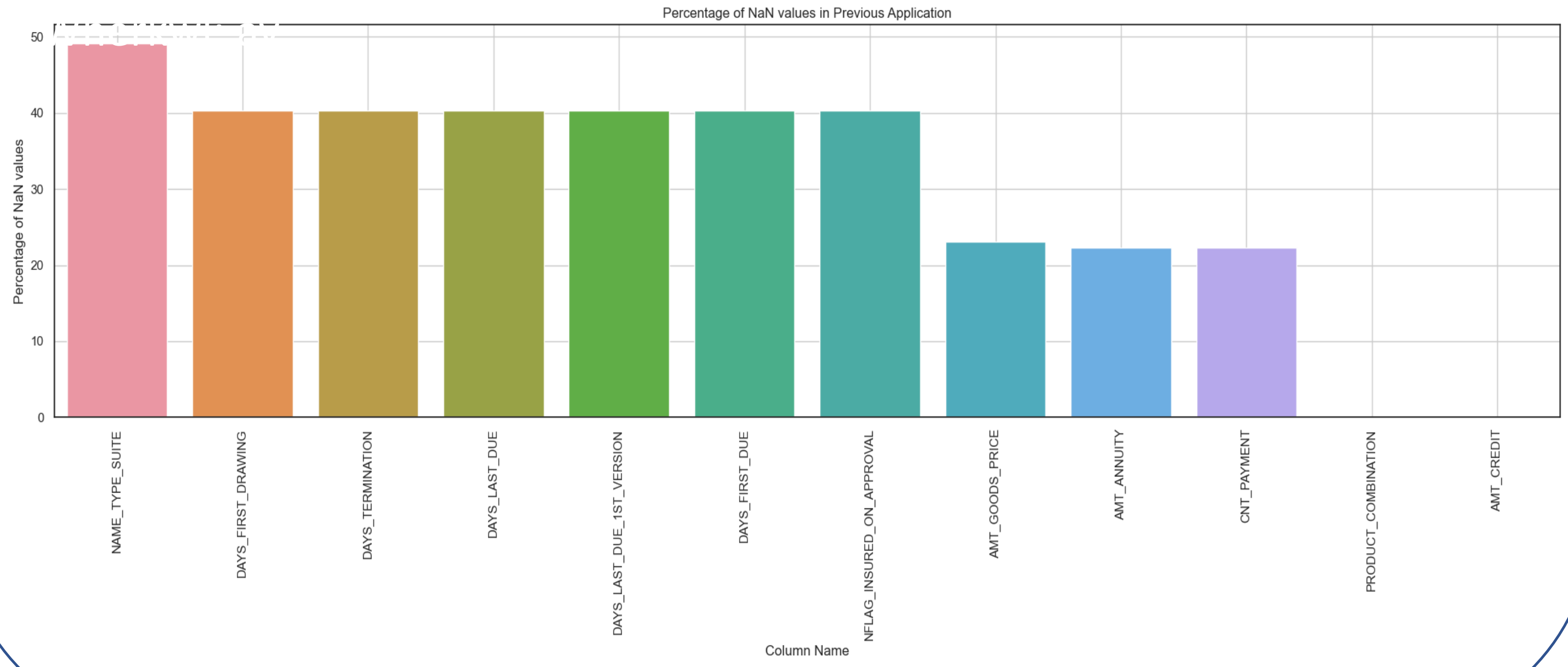
5. Comparative EDA

Previous Loans Data

Data File : Previous_application.csv

No of Rows : 1670214
No of Columns : 37

No of Columns dropped = 4 > 50 % null values



Univariate Analysis

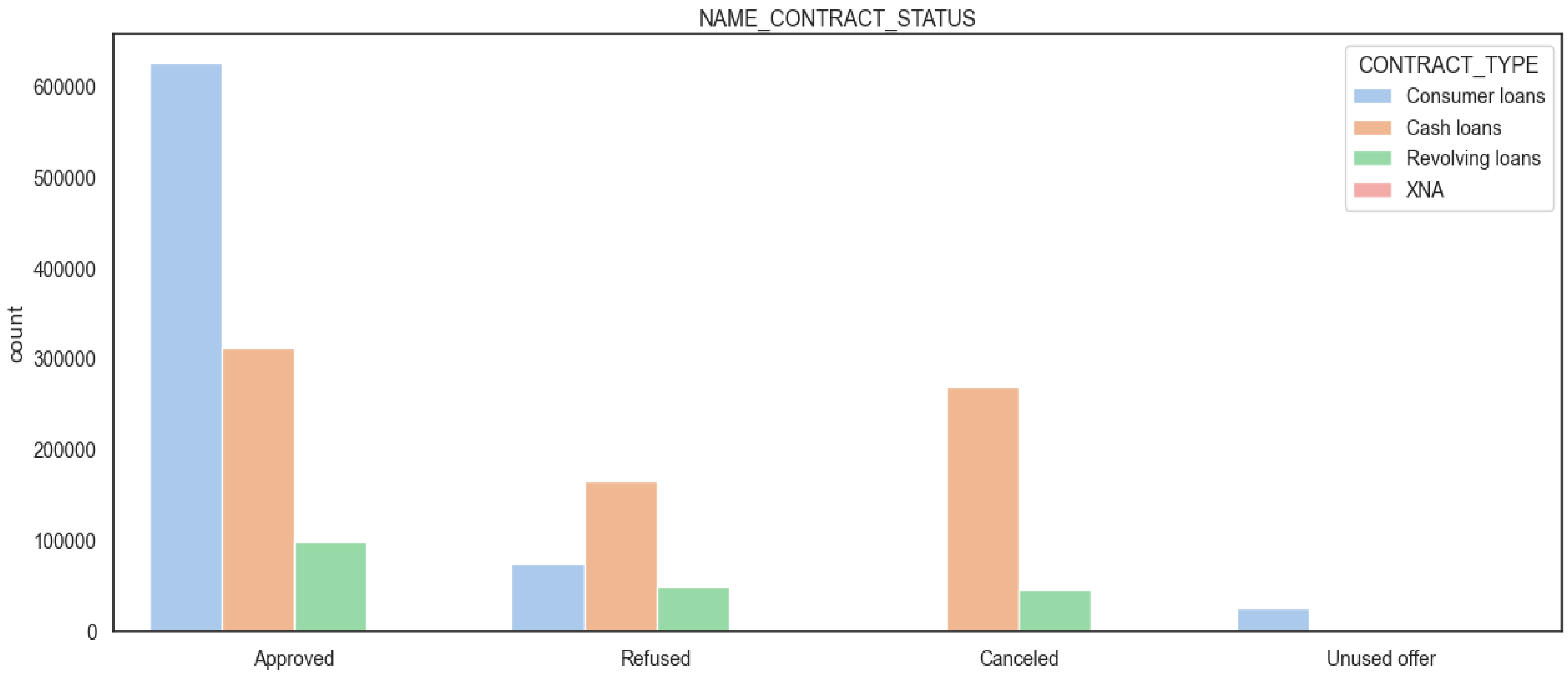
Preliminary Analysis

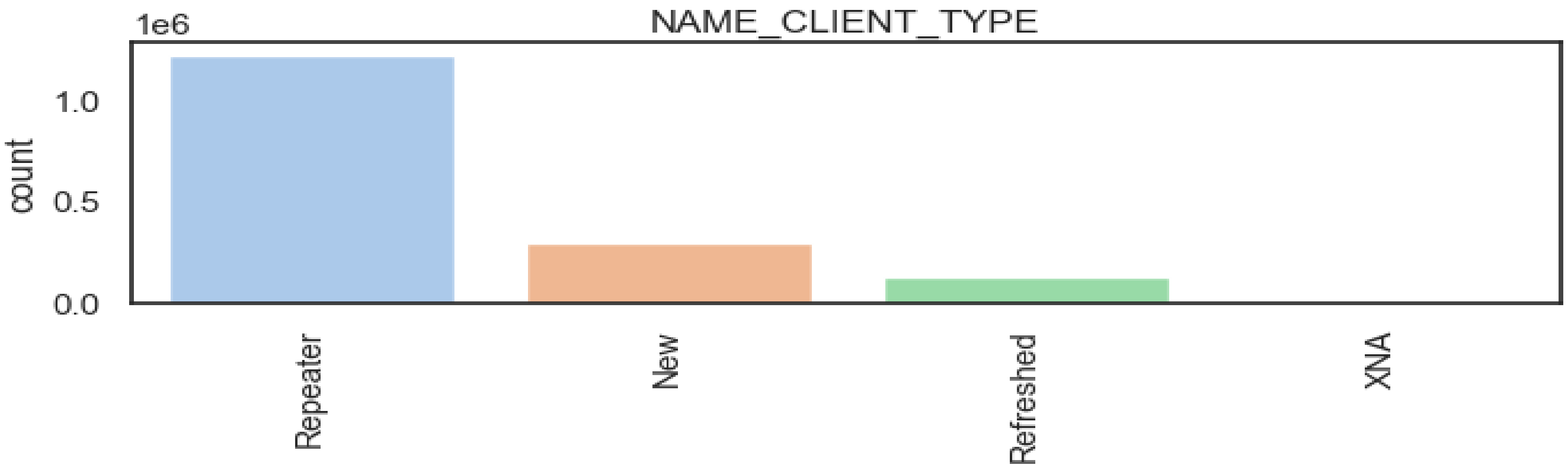
CONTRACT - STATUS

- Consumer Loans : Major Loan type
- Cash Loans : approx. 60 % applications refused or cancelled
- Revolving Loans – Low in frequency , 50 % refusal/ cancellation
- ❖ The application data set did not have information on Consumer loans

CLIENTS :

- Majority are repeat Clients

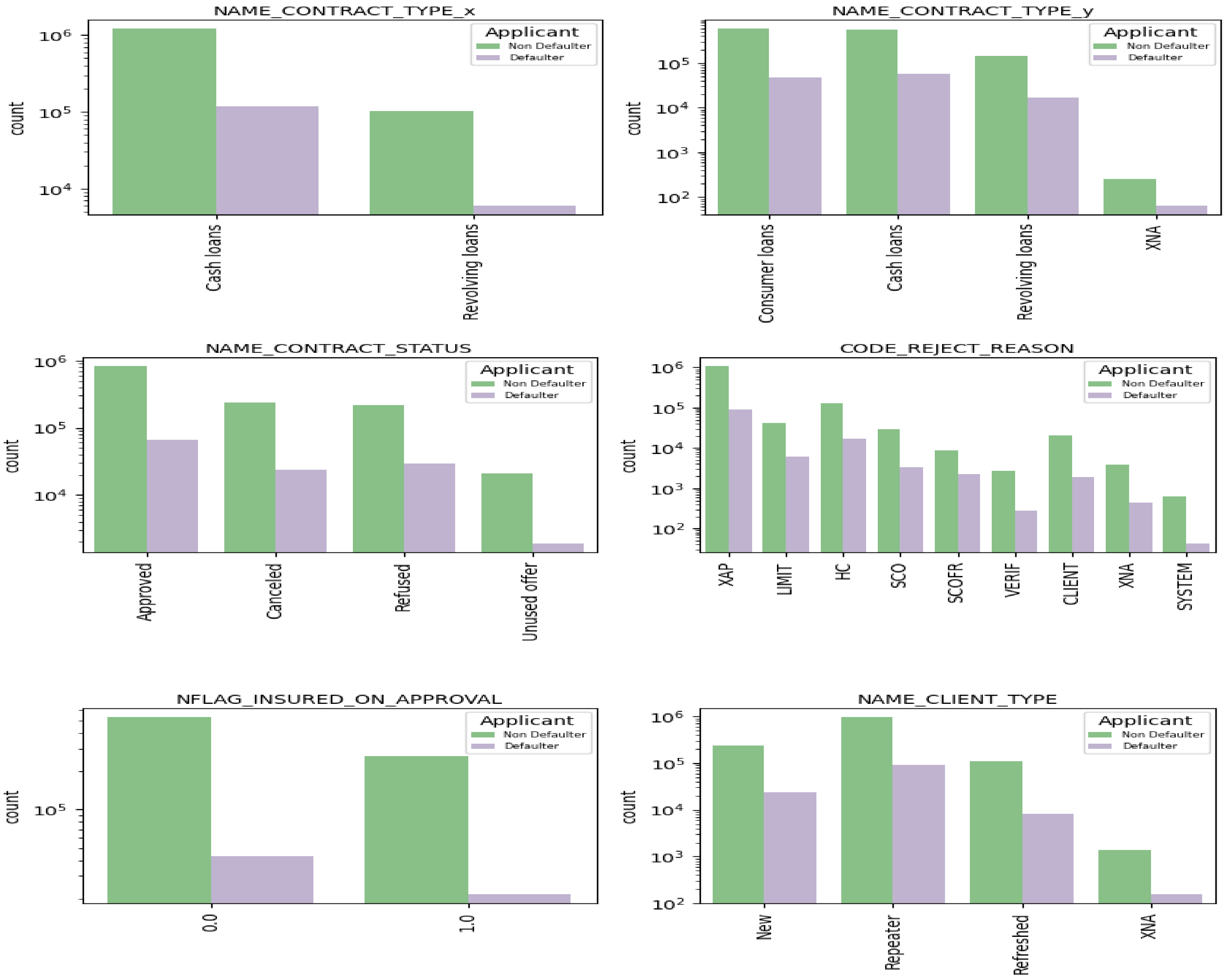




Merged Data :
Application , Previous Loan - Univariate Analysis
- Default vs Nondefault

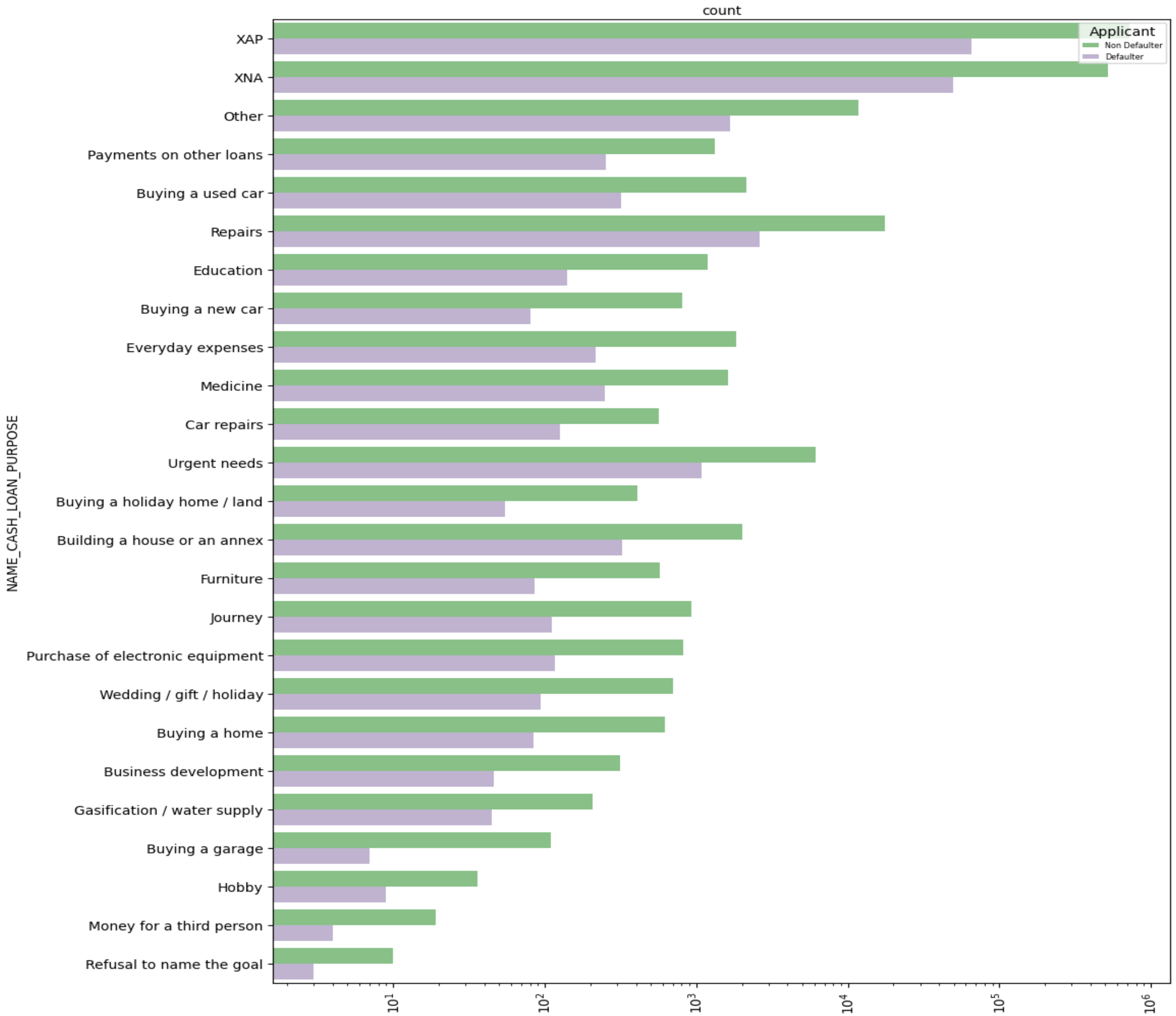
Defaulters:
- Approx. 40 % defaults in all 3 Contract types
- Rejection reason XAP, HC highest (description not known)
- High Defaults in Repeat Clients, followed by New Clients

✓ Repeat Clients – potential defaulters



Merged Data Set

- CASH LOAN PURPOSE :
- High default in XNA,XAP cases – need further information
- Defaults high - Urgent Need, Repairs
- ✓ Higher probability of Cash Loan defaults



Merged Data Set

Top 10 :Correlation Matrix

The merged data is aligned with results of 'application data' , new associations not evident

Possible Reason : The previous application data has loan primarily for CONTRACT_TYPE = Cash , which is not present in Application data.

Top 10 Correlations – Non-Defaulters

	Var1NDF	Var2NDF	Correlation
348	DAYS_FIRST_DRAWING	DAYS_FIRST_DRAWING	1.00
158	AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.99
767	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.94
415	DAYS_TERMINATION	DAYS_LAST_DUE	0.93
863	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.87
959	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.84
159	AMT_GOODS_PRICE_x	AMT_ANNUITY_x	0.76
127	AMT_ANNUITY_x	AMT_CREDIT_x	0.76
443	DAYS_EMPLOYED	DAYS_BIRTH	0.63
721	REGION_RATING_CLIENT	REGION_POPULATION_RELATIVE	0.53

Top 10 Correlations -Defaulters

	Var1DF	Var2DF	Correlation
348	DAYS_FIRST_DRAWING	DAYS_FIRST_DRAWING	1.00
158	AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.98
767	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.96
415	DAYS_TERMINATION	DAYS_LAST_DUE	0.94
863	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.87
959	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.79
127	AMT_ANNUITY_x	AMT_CREDIT_x	0.75
159	AMT_GOODS_PRICE_x	AMT_ANNUITY_x	0.75
443	DAYS_EMPLOYED	DAYS_BIRTH	0.59
410	DAYS_TERMINATION	DAYS_FIRST_DRAWING	0.47

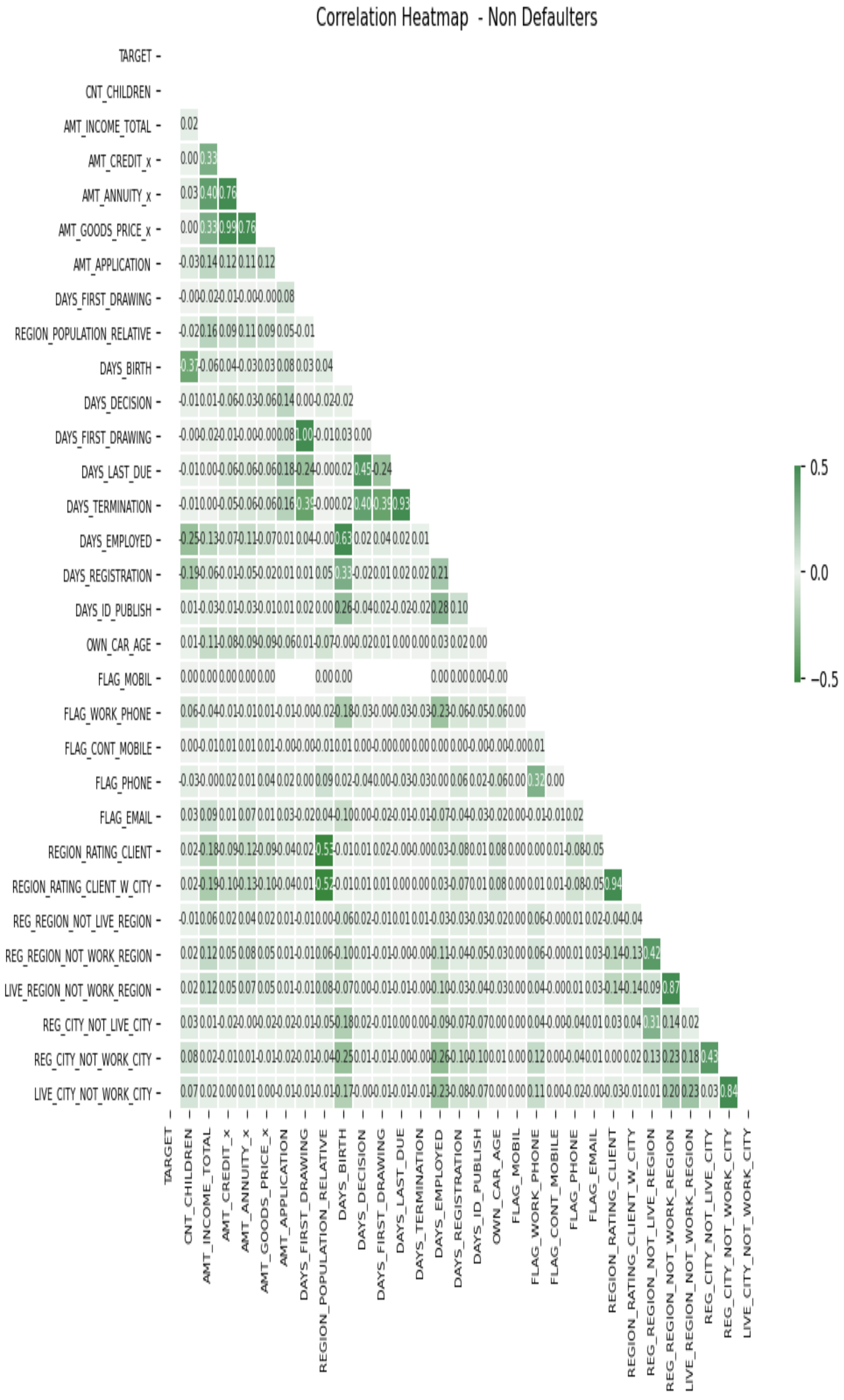
Merged Data Set

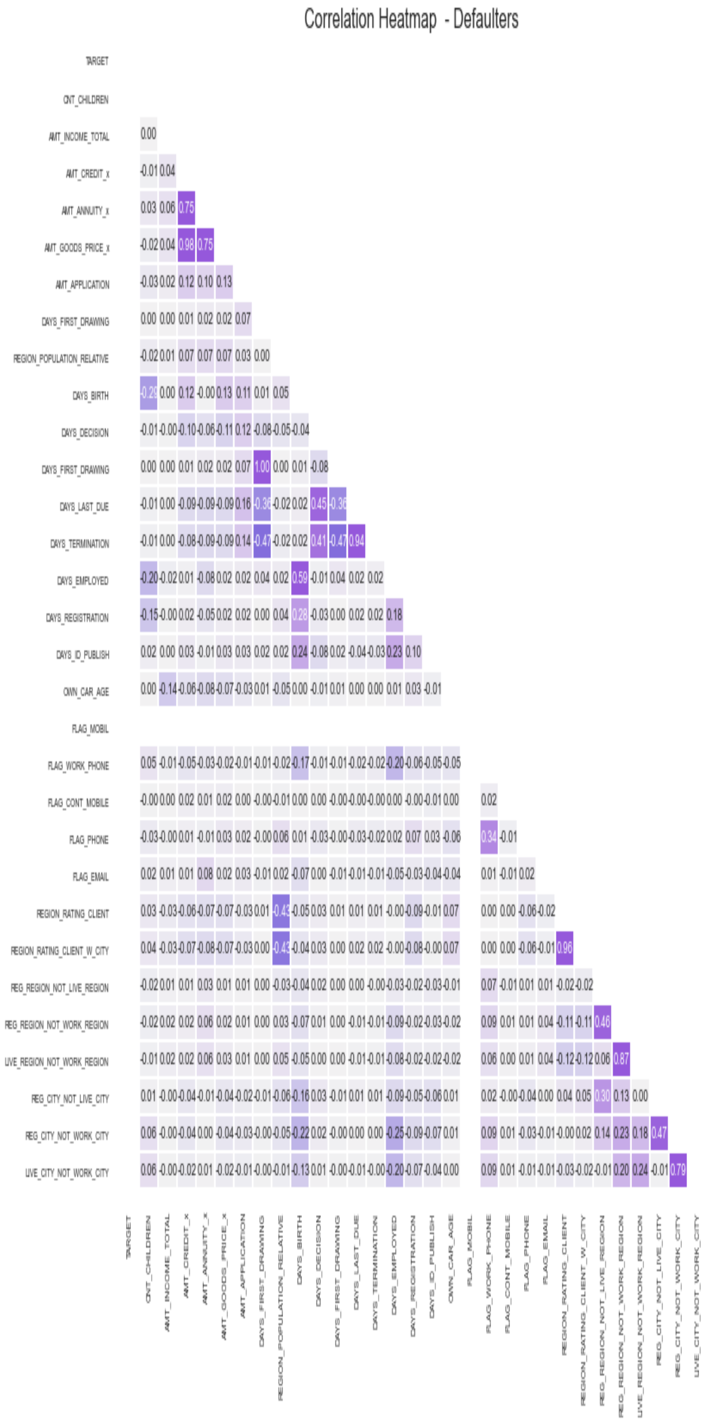
Correlation Heat Map

Very weak correlations

CONCLUSION :

- Defaulters and Non-Defaulters analysis behavior almost similar , No new associations were evident
- Defaults in all 3 Contract Types high approx. 40 %
- Repeat Customers tend to Default repeatedly





6. Final Insight

Final Recommendations to Loan Lending Company

The Below listed Variables are potential identifier for a customer to be a

Likely Defaulter:

- Credit Score - low External Source score
- Income: Range of 100000 to 200000
- Marital Status: Married, Single with high value loan, followed by Widows
- Education : Married , Highly educated with degrees, and high credit amount
- Number of Children 0 to 3 , High number of family members
- Occupation: Labourers, Sales Staff, Drivers , Managers, and Core Staff
- Lifestyle : Living in Own House, Apartments , Staying with parents, or owning a car
- Organization - Small Businesses, Self employed, Small Traders, Persons in Medicine, Government, School applicants

Final Recommendations to Loan Lending Company

The Below listed Variables are potential identifier for a customer to be a

Likely Non-Defaulter:

- People with Higher Ext_Source_Score, especially EXT_SOURCE_3 score above 0.4, and
  EXT_SOURCE_2 score above 0.5
- People With Revolving Loans
- People who started their process over weekends. Those who initiated on Sunday are least like to default, followed by those who initiated on Saturday and then Friday.
- People with an income of 500000 and above, with their credit lying below 10000
- Businessman and Student from Name Income Type
- People who are accompanied by a group of people when came for taking a loan
- People with Academic Degree
- HR Staff, IT Staff, followed by Real state agents and private service staff
- People who own their own car and real estate
- People living in co-operative or office apartment
- People with Unused offer in Name Contract Category

The End

References :

1. https://towardsdatascience.com/how-to-perform-exploratory-data-analysis-with-seaborn-97e3413e841d
2. https://stackoverflow.com
3. https://matplotlib.org
4. https://seaborn.pydata.org
5. https://medium.com/@morganjonesartist/color-guide-to-seaborn-palettes-da849406d44f
6. https://plotly.com/