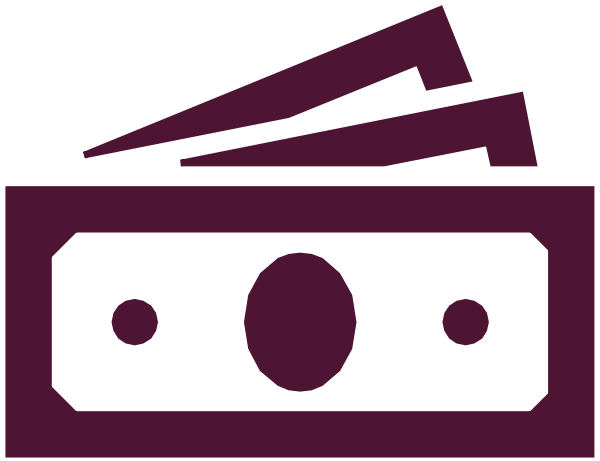


CREDIT EDA CASE STUDY

BY

SOWNDARYA VENKATESWARAN

PALLAVI GODAVARTHY

BUSINESS OBJECTIVE

- The aim of this case study is to identify the pattern of which indicate if the client has difficulty in paying their instalments. - Loss of the Loan amount to the bank.
- Based on which we can take actions like
  - denying the loan,
  - reducing the amount of loan,
  - lending (to risky applicants) at a higher interest rate, etc.
- This will also ensure that the consumers that are capable of repaying the loan are not rejected. - Loss of interest amount to the bank.

BUSINESS UNDERSTANDING

- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company.

BUSINESS UNDERSTANDING

- Two scenarios about the loan application at the time of applying for the loan:
  - **The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
  - **All other cases:** All other cases when the payment is paid on time.

BUSINESS UNDERSTANDING

Decisions that can be taken for a loan application by the company

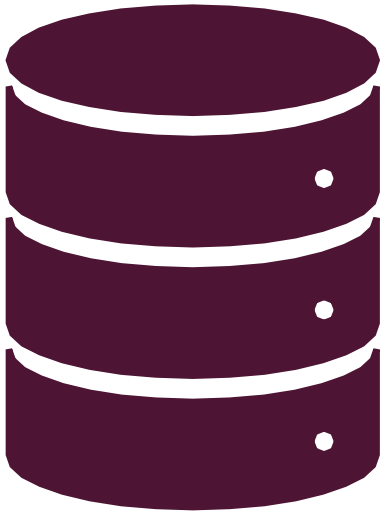
Approved: The Company has approved loan Application

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 on different stages of the process.

DATA UNDERSTANDING



- This dataset has 3 files 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.
  - '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.
  - 'columns_description.csv' is data dictionary which describes the meaning of the variables

ANALYSIS OF CLIENT INFORMATION AT THE TIME OF APPLICATION

MISSING VALUE ANALYSIS

- HANDLING THE MISSING VALUES LESS THAN 40%

Occupation_type is the applicant's occupation type.

- Imputing Null values with value "Unknown" will help in further analysis of the loan application.

EXT_SOURCE_3

- Median value is 0.54,Mean is 0.51 and mode (most repeated value) is 0.75
- The missing Value can be imputed with mean or median as they are closer.
- For further analysis we are not imputing the missing values with mean or median because the missing value % is 19.83% which is higher.

AMT_REQ_CREDIT_BUREAU fields

- We can see that Mode of all the 'AMT_REQ_CREDIT_BUREAU_MONTH', 'AMT_REQ_CREDIT_BUREAU_WEEK', 'AMT_REQ_CREDIT_BUREAU_DAY', 'AMT_REQ_CREDIT_BUREAU_HOUR', 'AMT_REQ_CREDIT_BUREAU_QRT', and 'AMT_REQ_CREDIT_BUREAU_YEAR' is "0".
- Although it makes more sense to impute the more recurring value which is Mode = 0, we are not imputing these columns with "0" to avoid any bias since the missing value % is 13.50% which is quite high.

NAME_TYPE_SUITE

- We observe that maximum clients (248526/306219) are unaccompanied when they are applying for the loan
- NAME_TYPE_SUITE has a total of 1292 null values which is 0.42%
- Imputing the null values as Unaccompanied will help in further analysis.

MISSING VALUE ANALYSIS

- ***HANDLING THE MISSING VALUES LESS THAN 40%***

CNT_SOCIAL_CIRCLE

- The Median and Mode for CNT_SOCIAL_CIRCLE is 0
- The null percentage is 0.33% which is smaller and will help us to produce unbiased results.
- We can impute the null values with mode which is 0.
- We also observe outliers for these columns.

EXT_SOURCE_2

- The Mean (0.51) and Median (0.56) values are closer.
- Null percentage is 0.21% is small.
- Imputing missing values with mean will not bias the analysis.
- There are no Outliers in EXT_SOURCE_2.

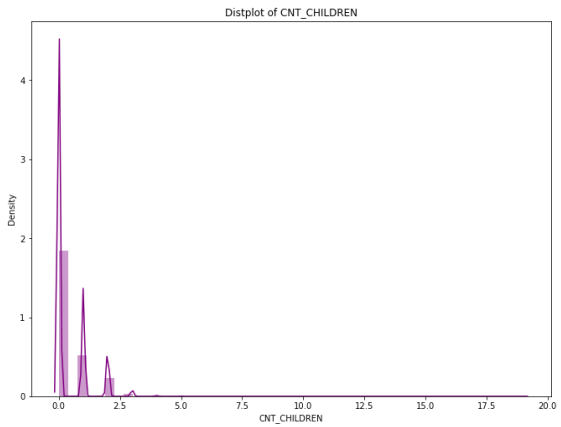
AMT_GOODS_PRICE

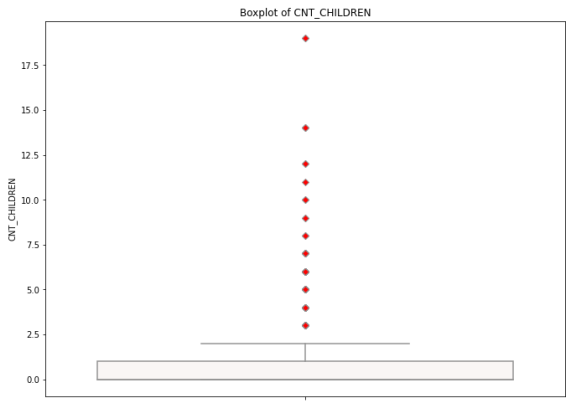
- The Mean and Mode values are exactly the same 450000.0.
- We can impute the null value with 450000.0, since the % of null values is very less.
- Although the box plot shows Outliers, we cannot consider it as outliers as it could be the amount of the good.

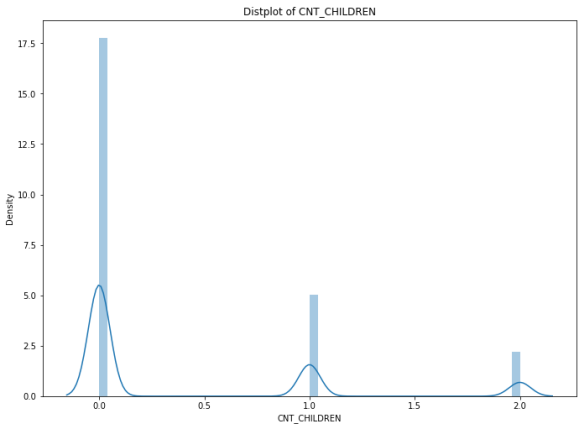
OUTLIER ANALYSIS

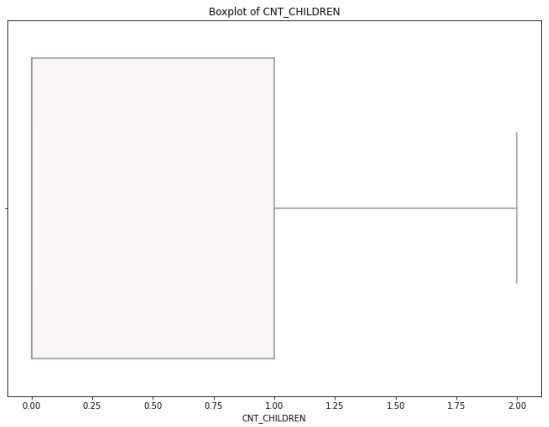
OUTLIERS OF CNT_CHILDREN

- It is oserved that Values above Max value = 2.5 can be treated as outlier.
- It is also observed that for above 7 children, the count of applicants are very minimal (2 or 3 in each category)



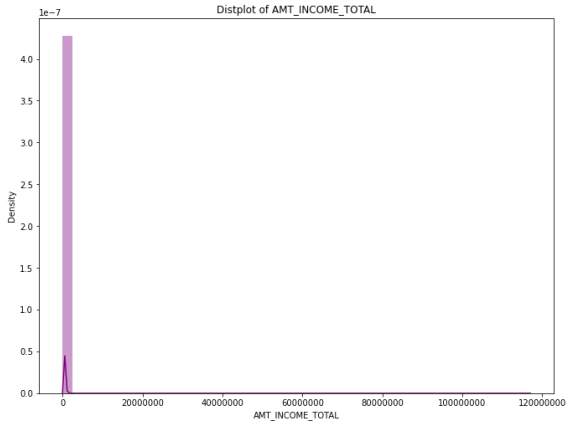


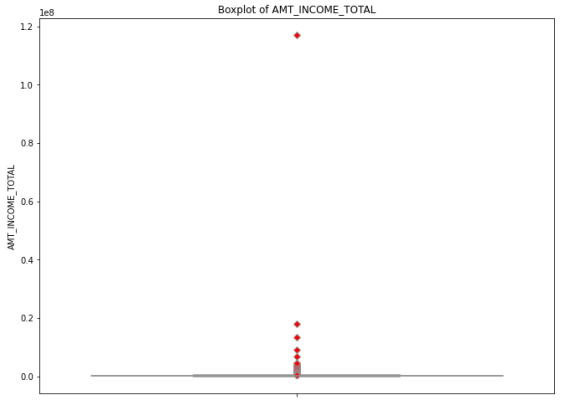


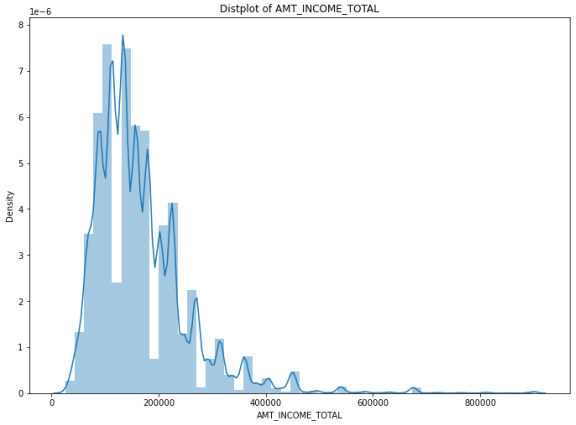


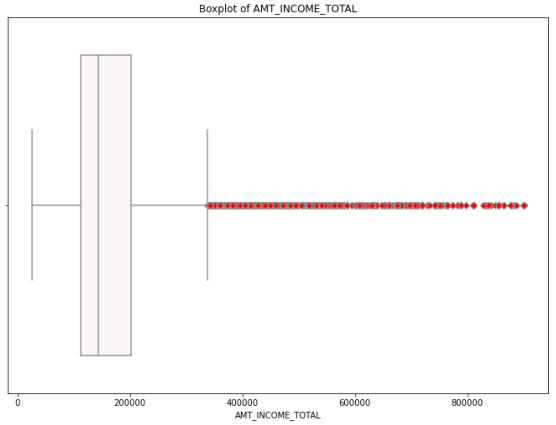
OUTLIERS IN AMT_INCOME_TOTAL

- It is observed that plots are thin and the outlier can be seen at appx 120K.
- Dist and Box plots show that the values above 99th Percentile (900K) are clearly outliers



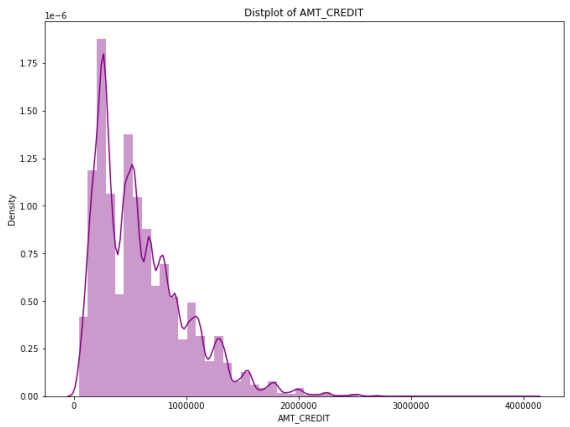


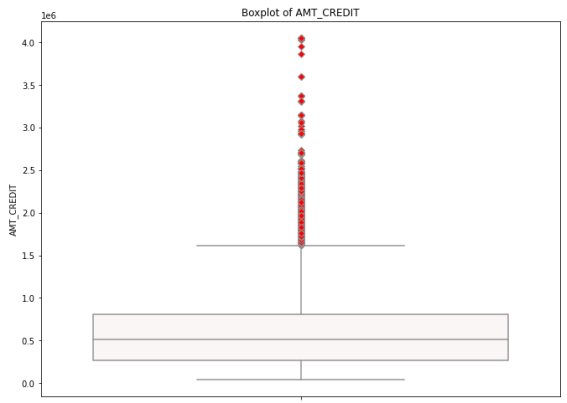


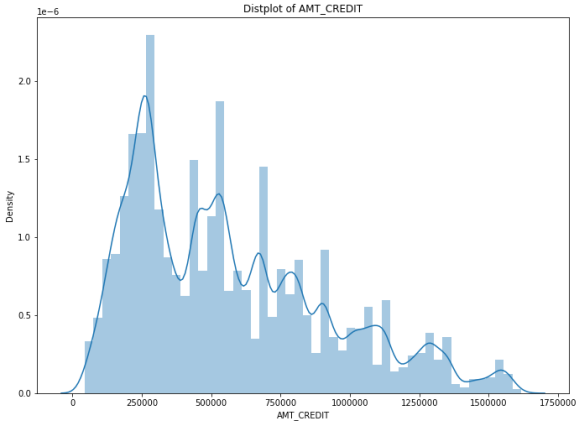


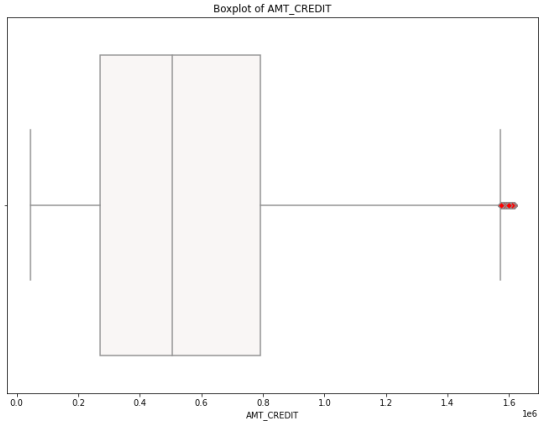
OUTLIERS IN AMT_CREDIT

- AMT_CREDIT value above 1.6 are observed to be outliers
- From the plots we can conclude that values above 1616625 can be considered as Outliers



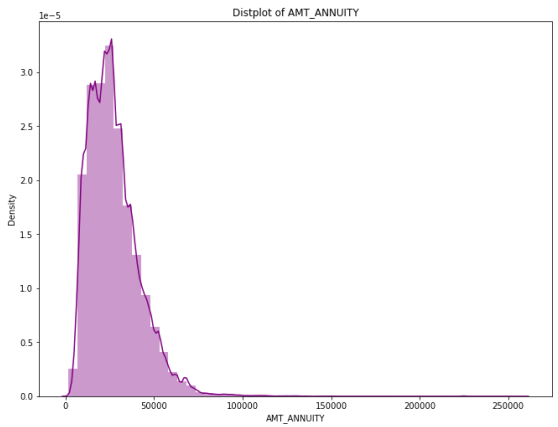


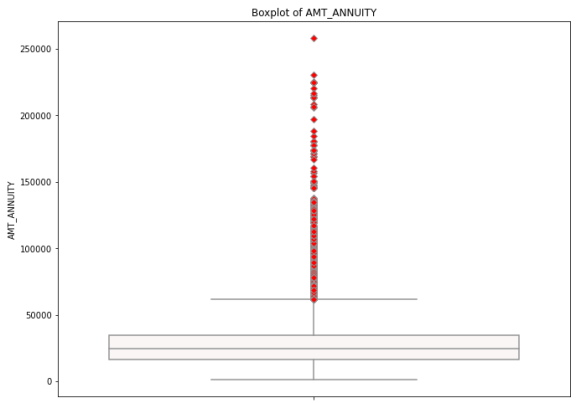


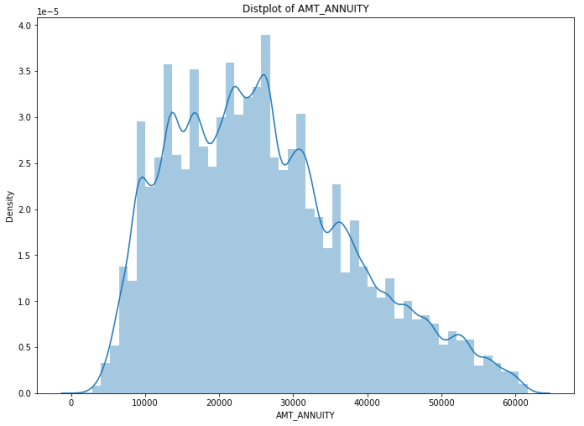


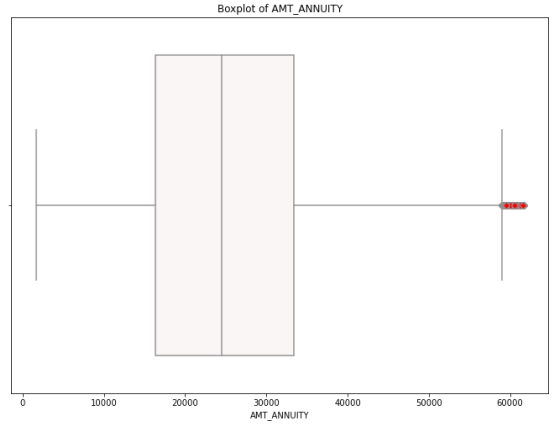
OUTLIERS IN AMT_ANNUITY

- THE VALUES ABOVE 61704.0 AMT_ANNUITY ARE CLEARLY OUTLIERS
- Plot shows 61704.0 AMT_ANNUITY are clearly outliers



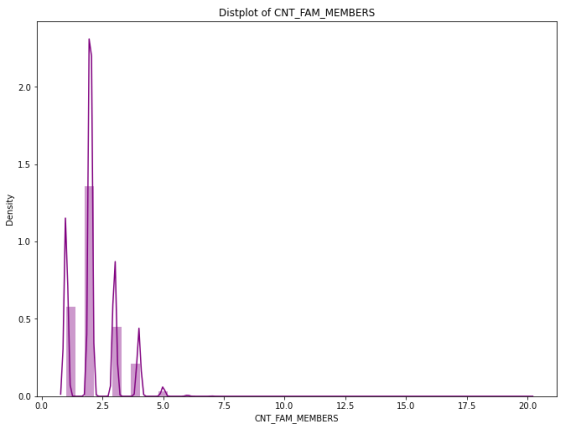


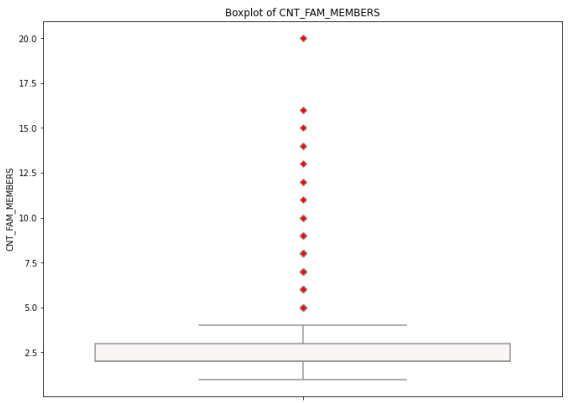


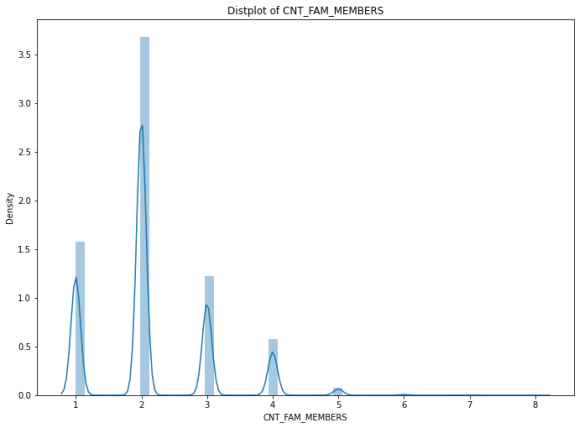


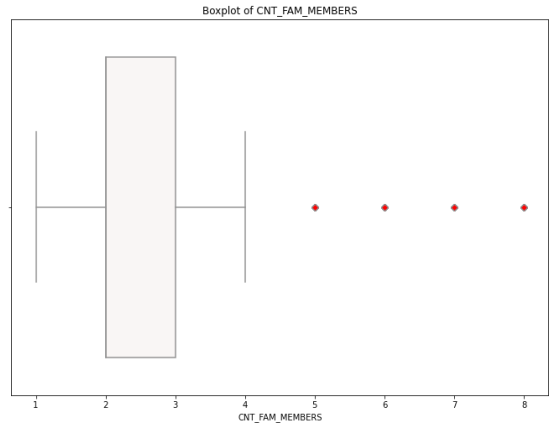
OUTLIERS IN CNT_FAM_MEMBERS

- IQR is 1.0 1.Min value before which outlier exist is 0.5 2.Max value after which outlier exist is 4.5
- From the plots and IQR - Max value calculation, it is clear that values above 4.5 are Outliers





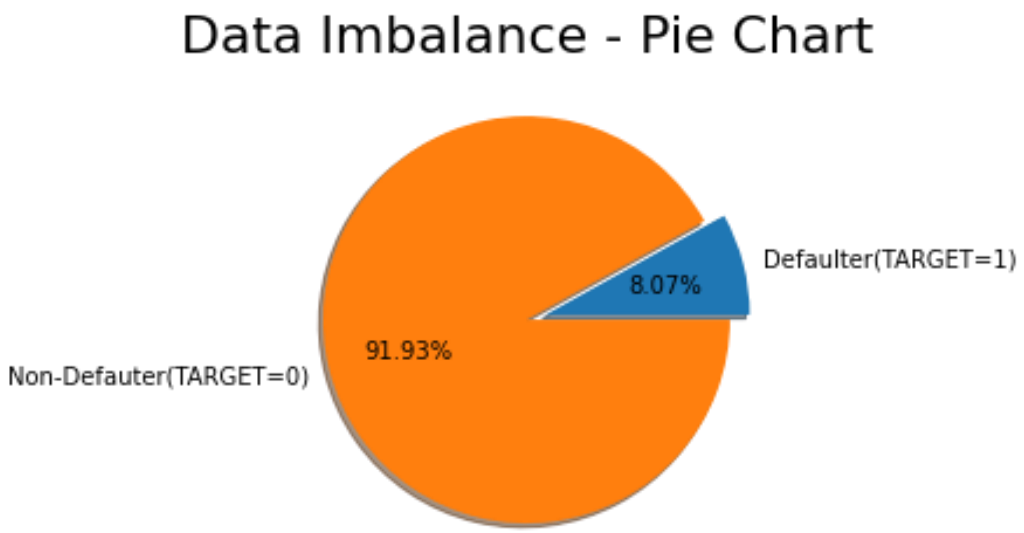




CHECKING IMBALANCE FOR "TARGET"

IMBALANCE FOR TARGET

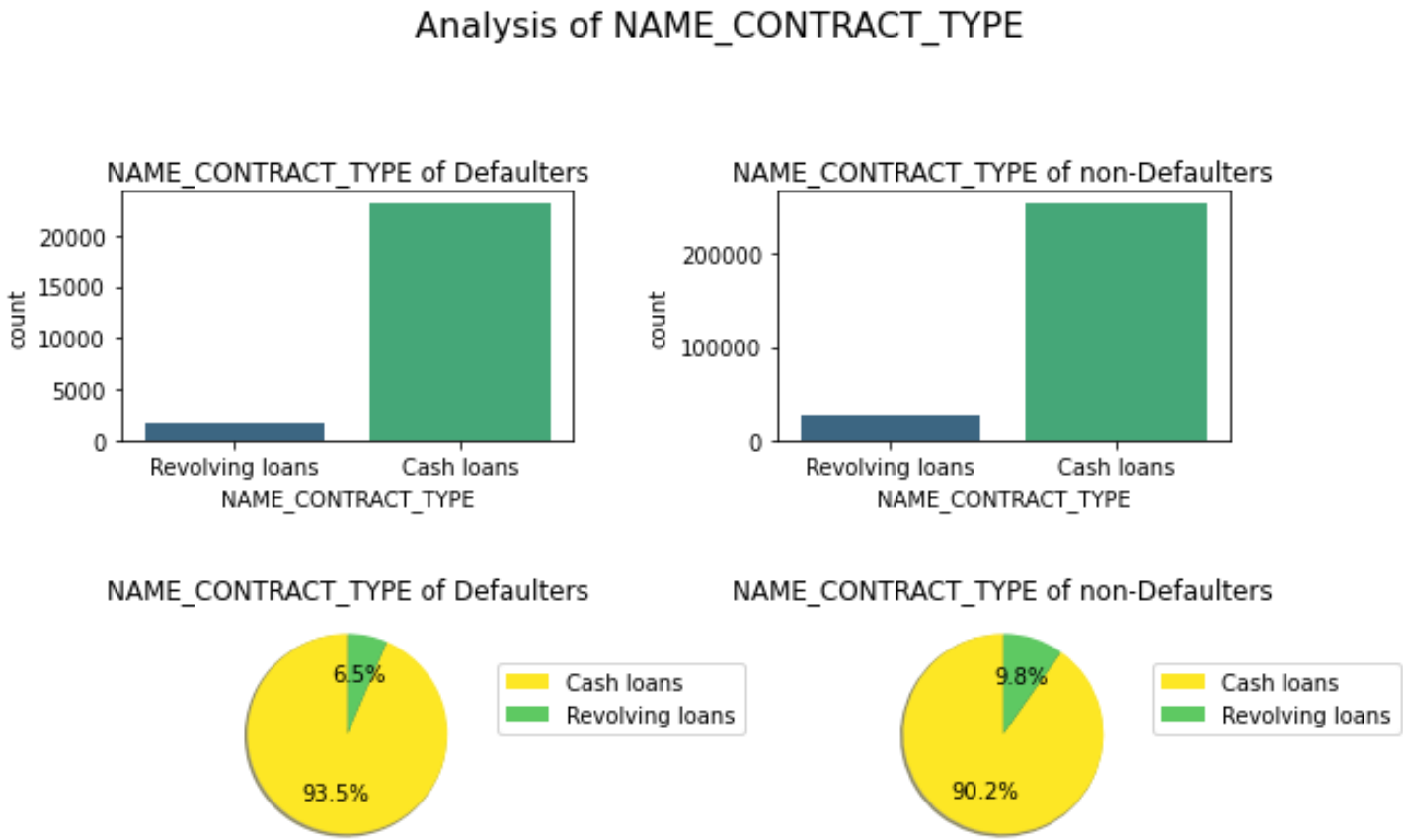
- Observation - From the above plot, we can conclude that-
  - TARGET - 1 which represents clients with payment difficulties are appx 8% [Defaulters]
  - TARGET - 0 which represents clients with no payment difficulties are appx 92% [Non-defaulters]
- The Imbalance Ratio between defaulters and non-defaulters is 11.39%



UNIVARIATE ANALYSIS

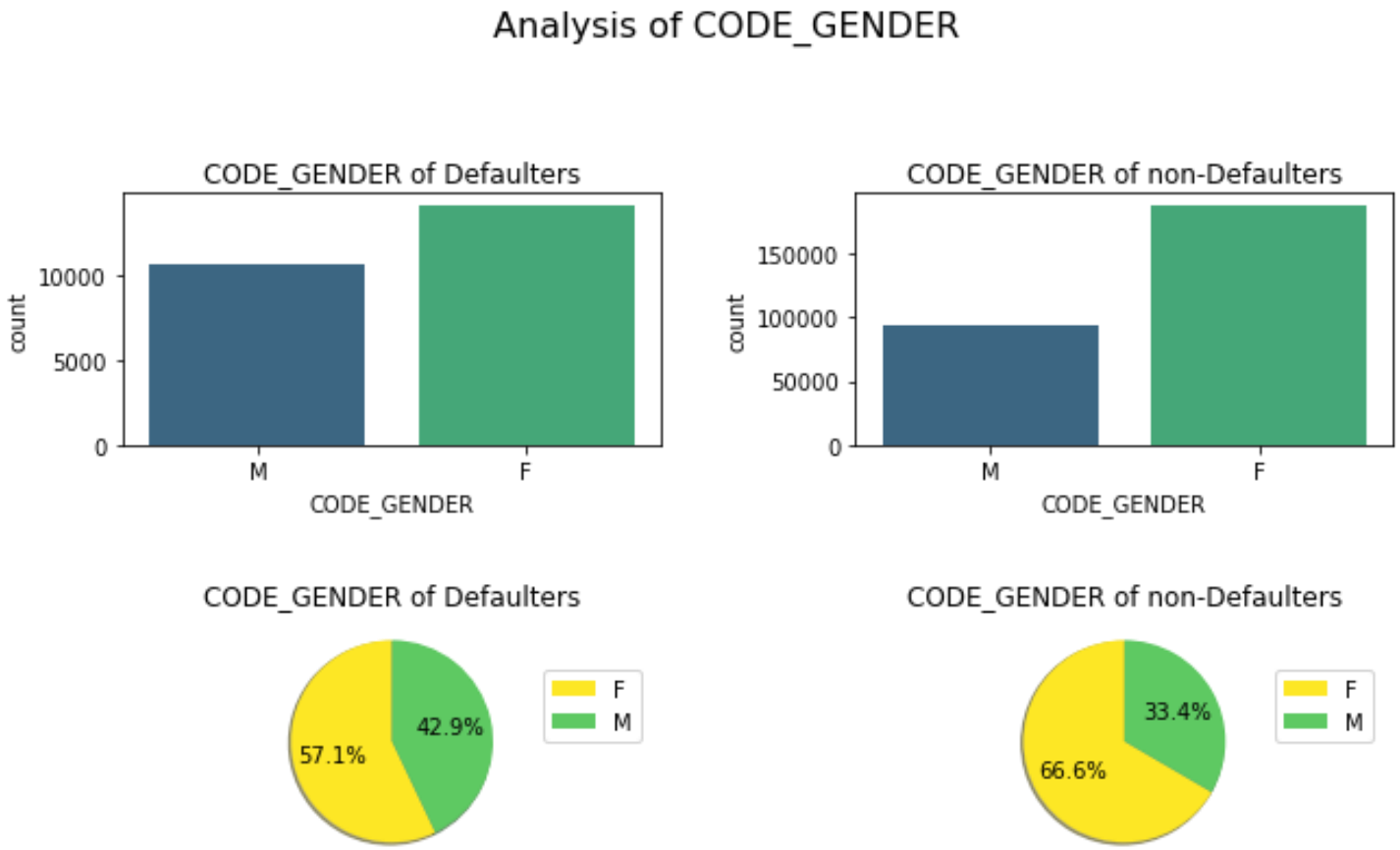
NAME_CONTRACT_TYPE

- With NAME_CONTRACT_TYPE, unable to see any significant difference between Defaulters and non defaulters.
- NAME_CONTRACT_TYPE - does not provide any conclusive evidence in favor of defaulting or non defaulting clients.



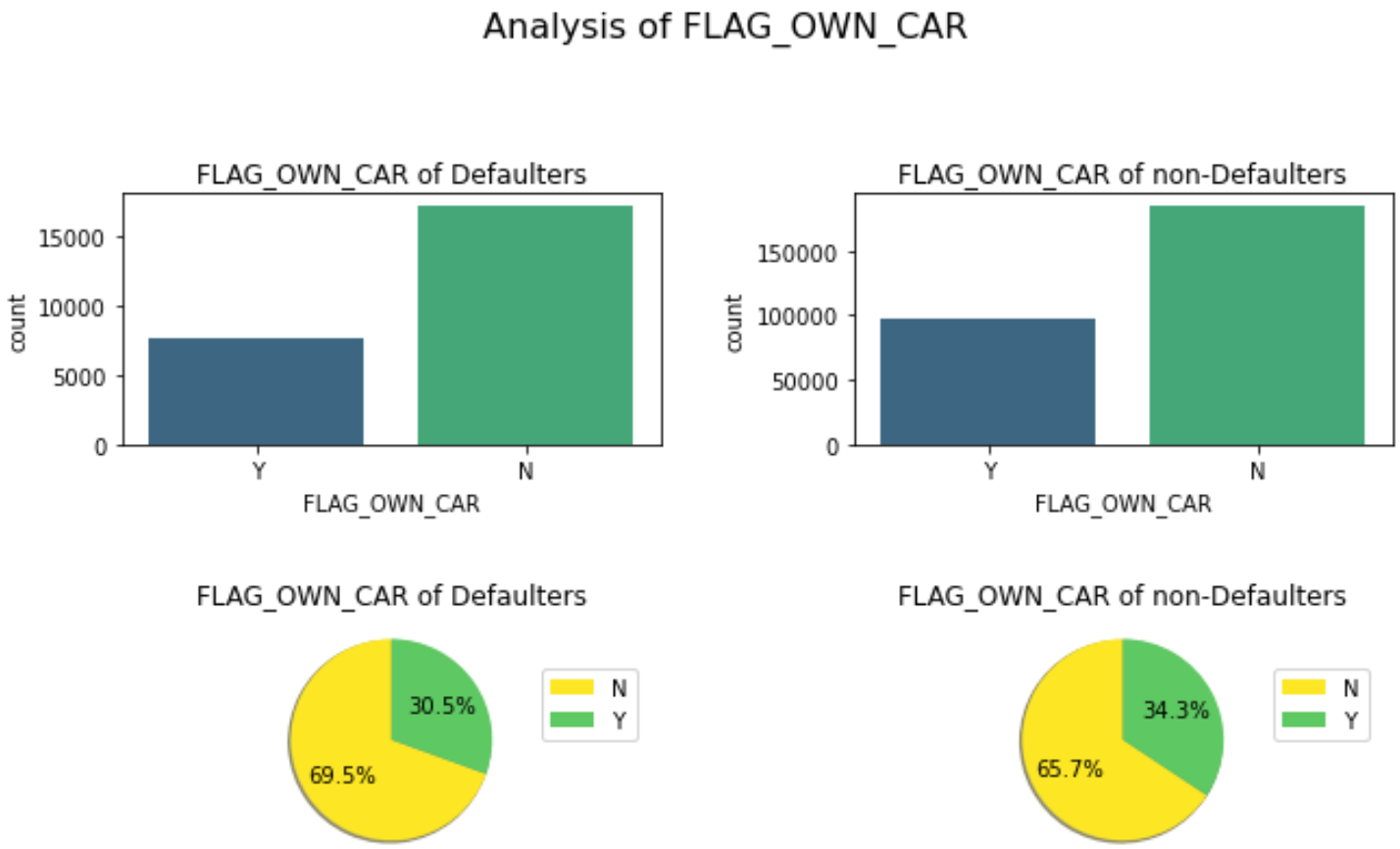
CODE_GENDER

- It is observed that Female applicants are more in numbers than Male applicant.
- This trend is oserved in both Defaulters and Non defaulters.
- The repayment trend shows that
  - Both men and women find it difficult to repay.
  - Men find it more difficult to repay than Women.



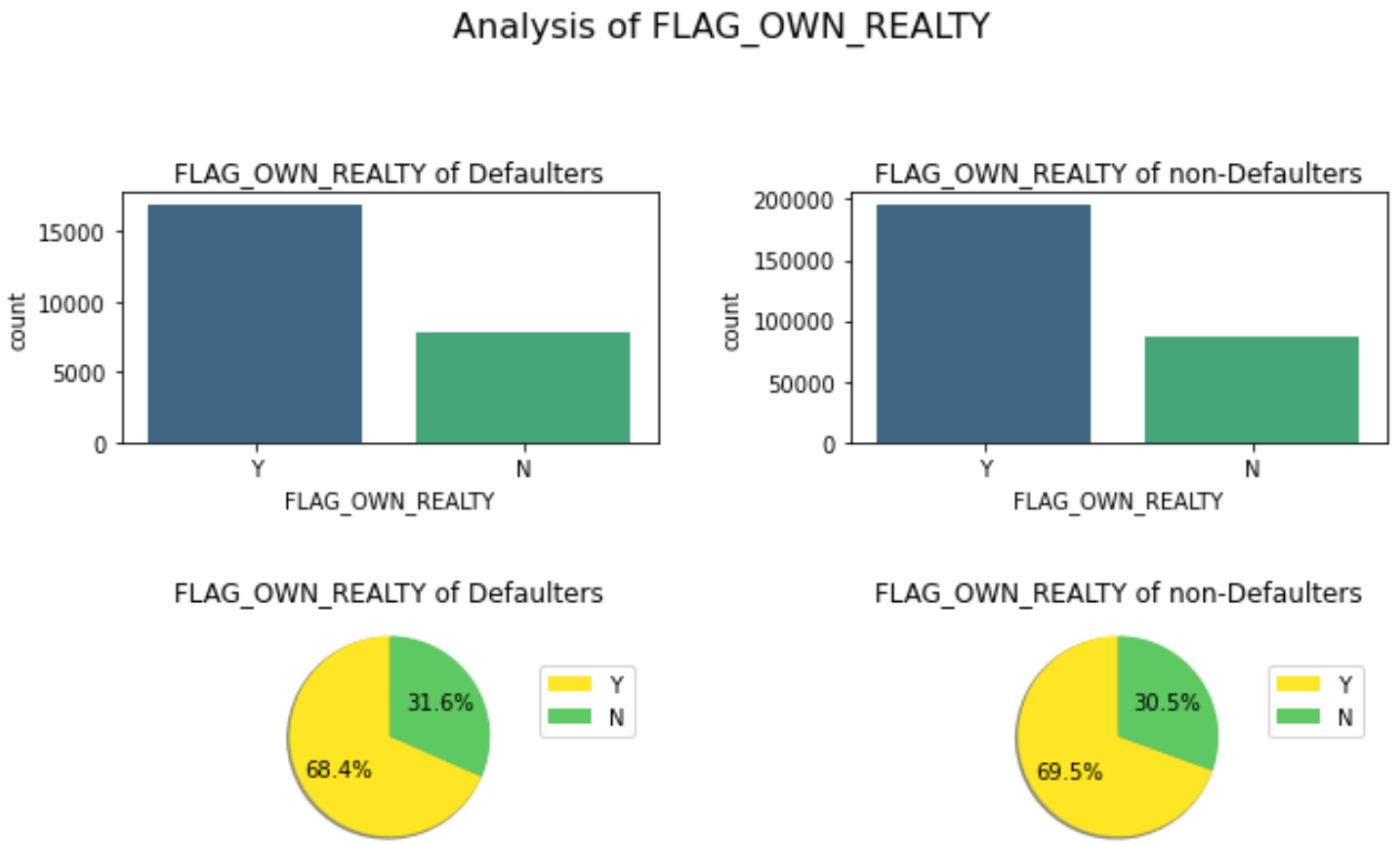
FLAG_OWN_CAR

- There is no significant difference in repayment of loan between car owner and non car owner.
- FLAG_OWN_CAR - variable does not help us in providing any conclusive evidence between Defaulters and non-defaulters.



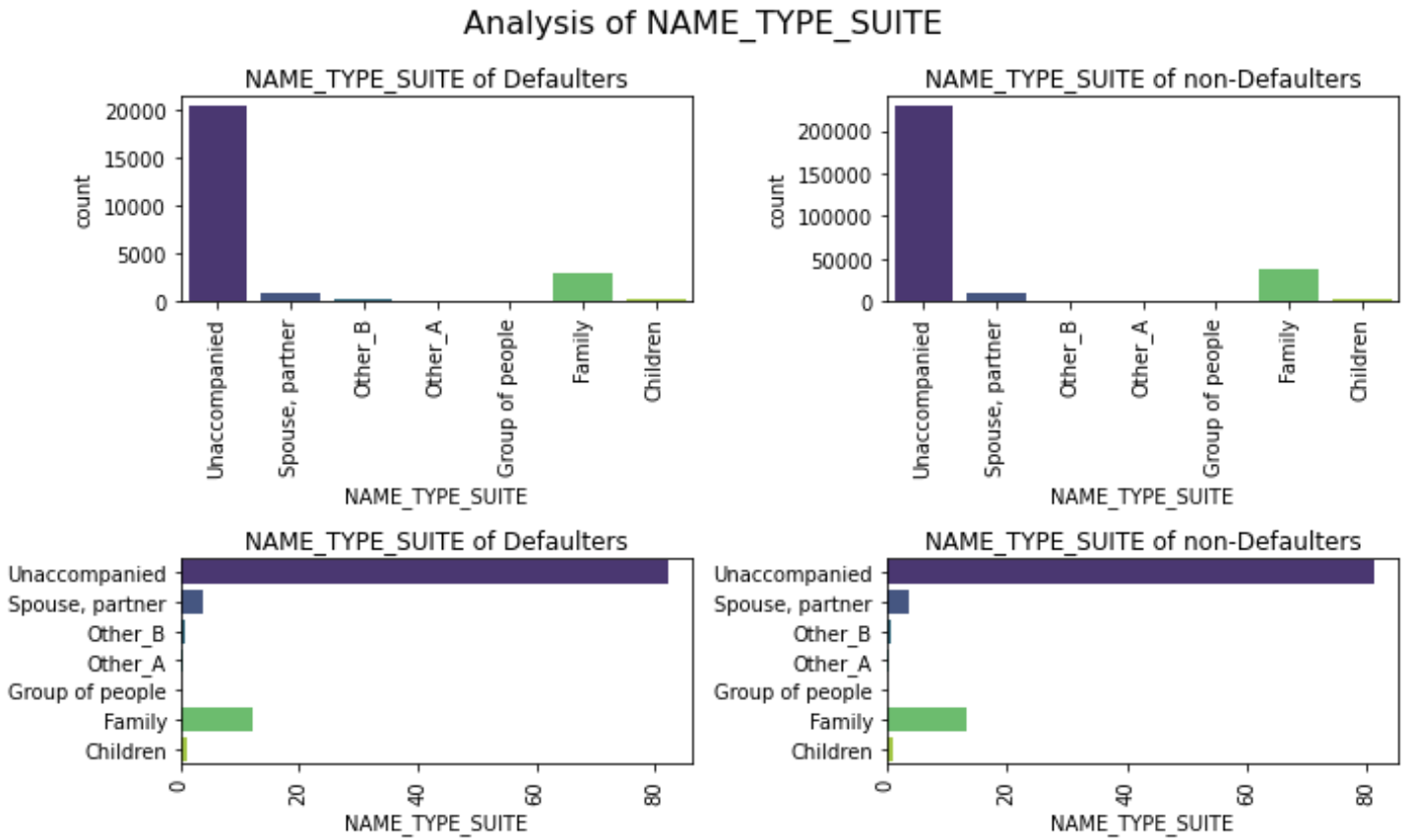
FLAG_OWN_REALTY

- There is no significant difference in repayment of loan between Realty owner and non Realty owner.
- FLAG_OWN_REALTY - variable does not help us in providing any conclusive evidence between Defaulters and non-defaulters.



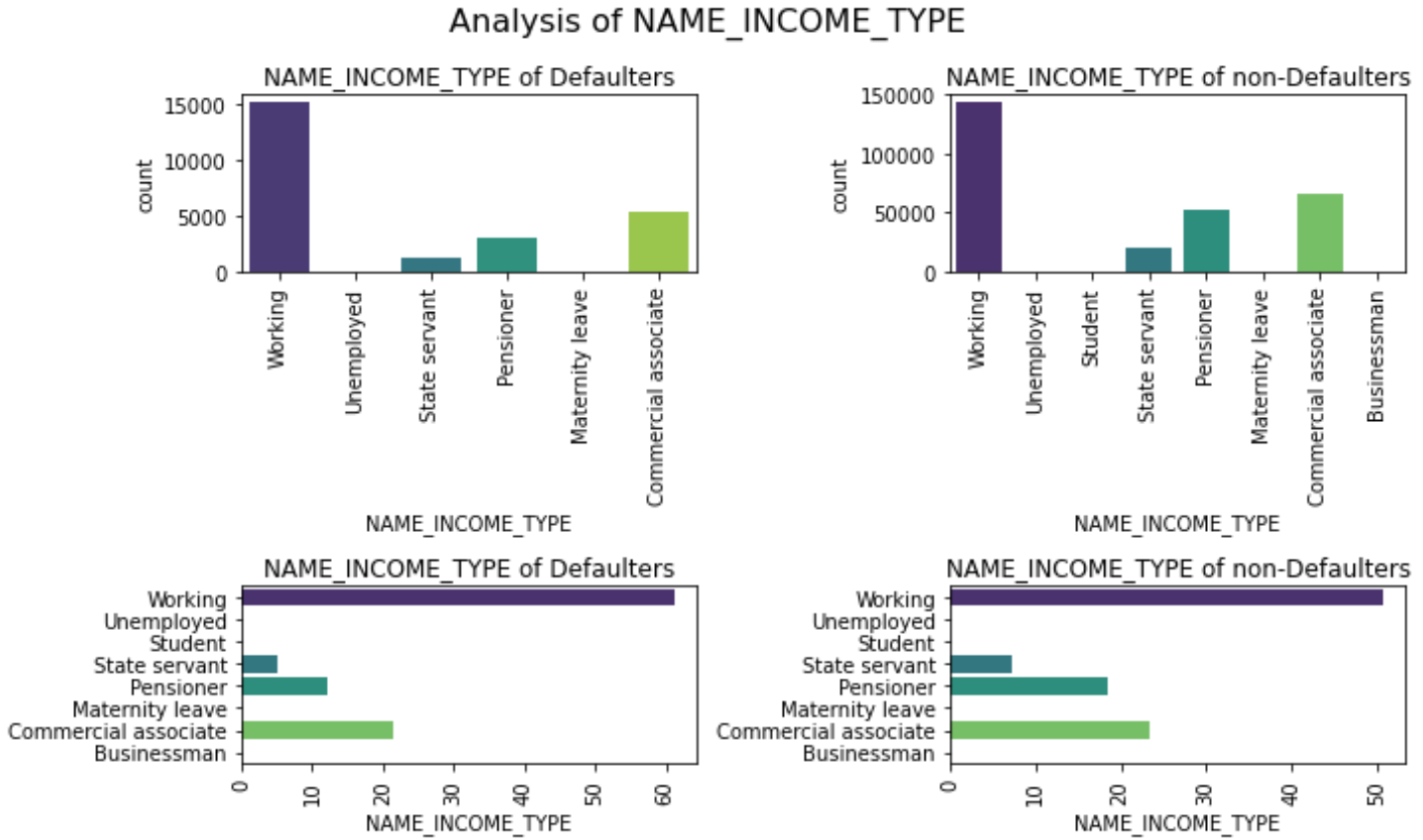
NAME_TYPE_SUITE

- NAME_TYPE_SUITE data does not give us conculsive evidence between Defaulter and non Defaulters



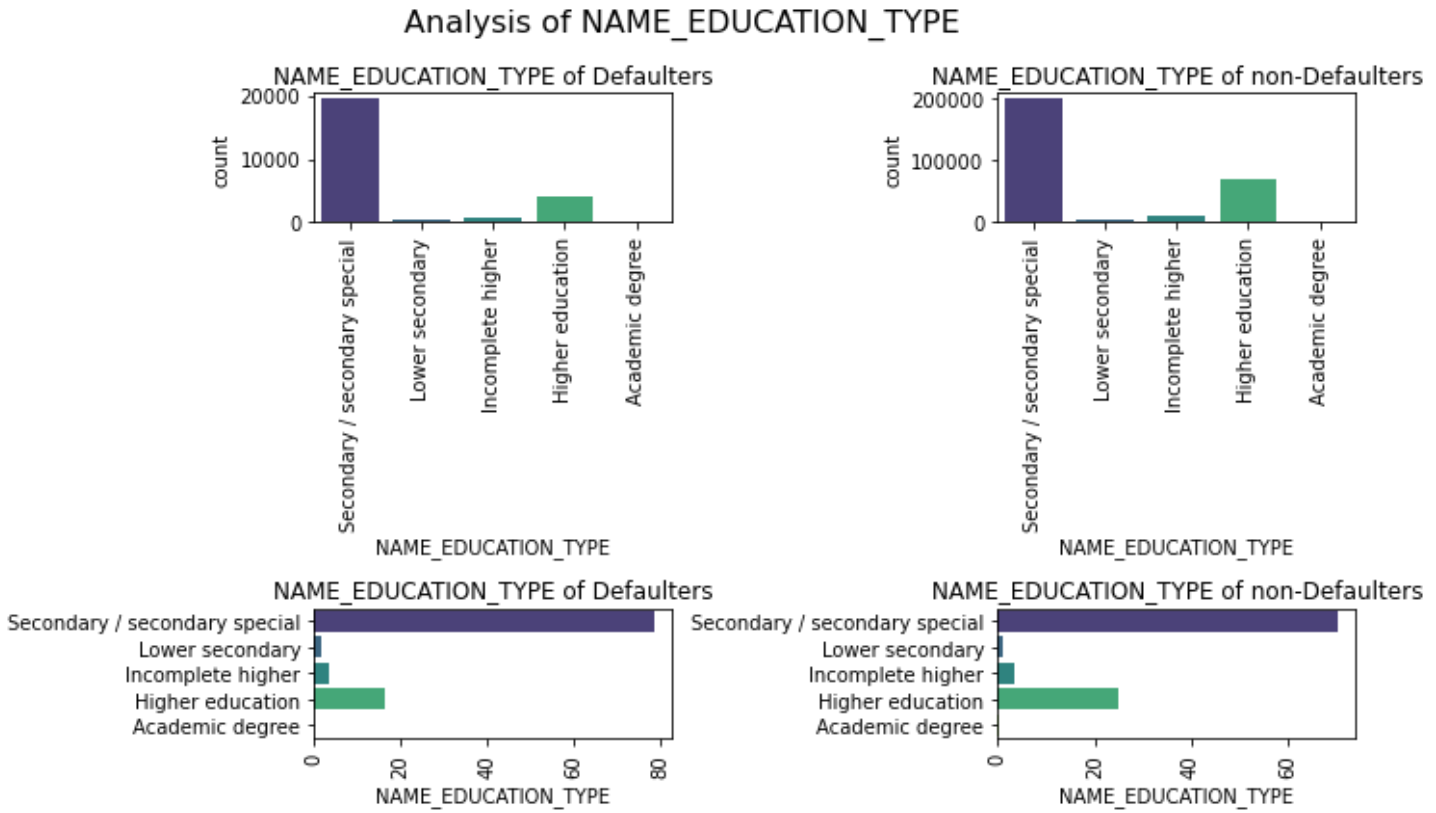
NAME_INCOME_TYPE

- Pensioner have a better on repayment trend.
- Students and Businessmen are able to make on-time payments.



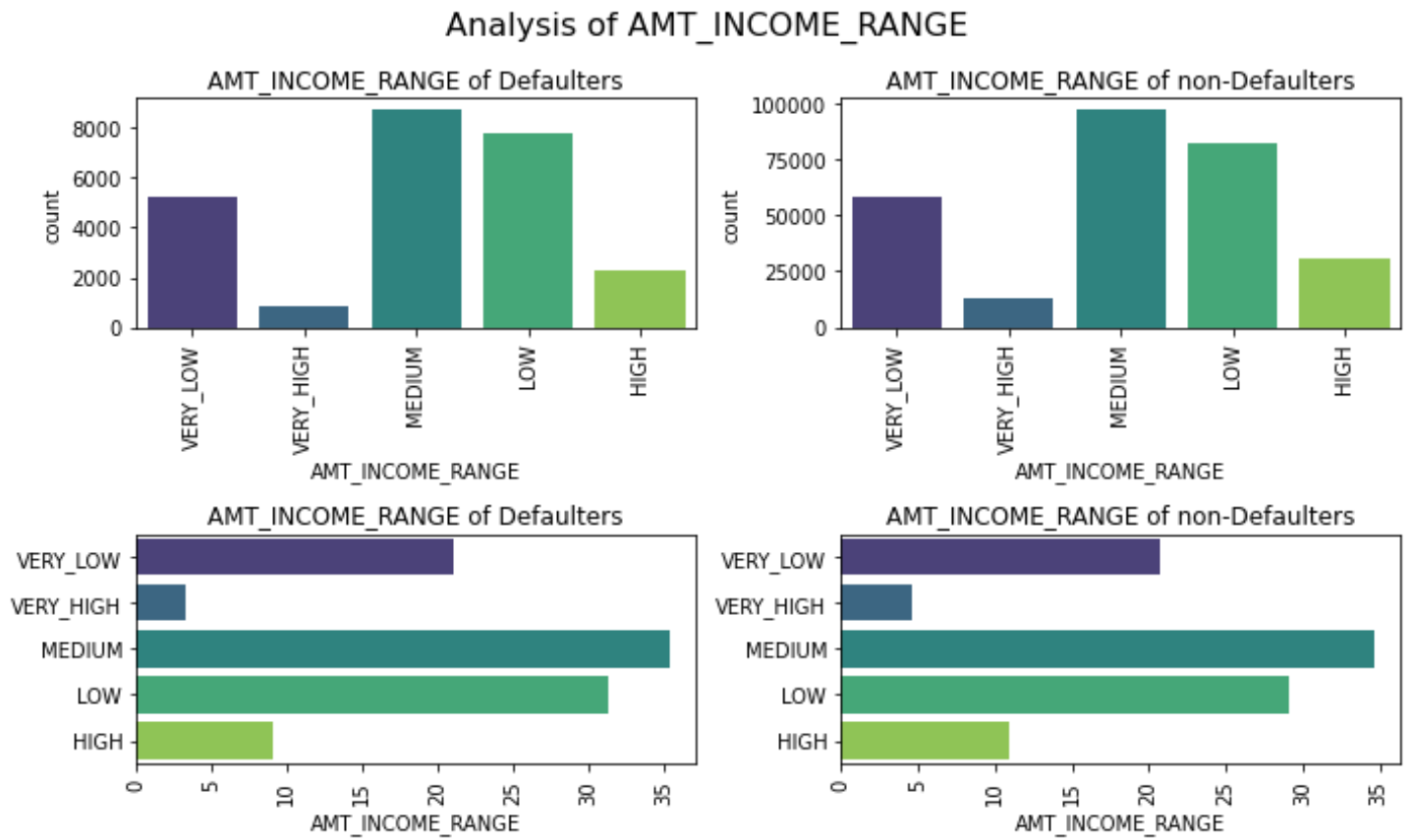
NAME_EDUCATION_TYPE

APPLICANTS WITH HIGHER EDUCATION MAKE ON TIME PAYMENTS BETTER.



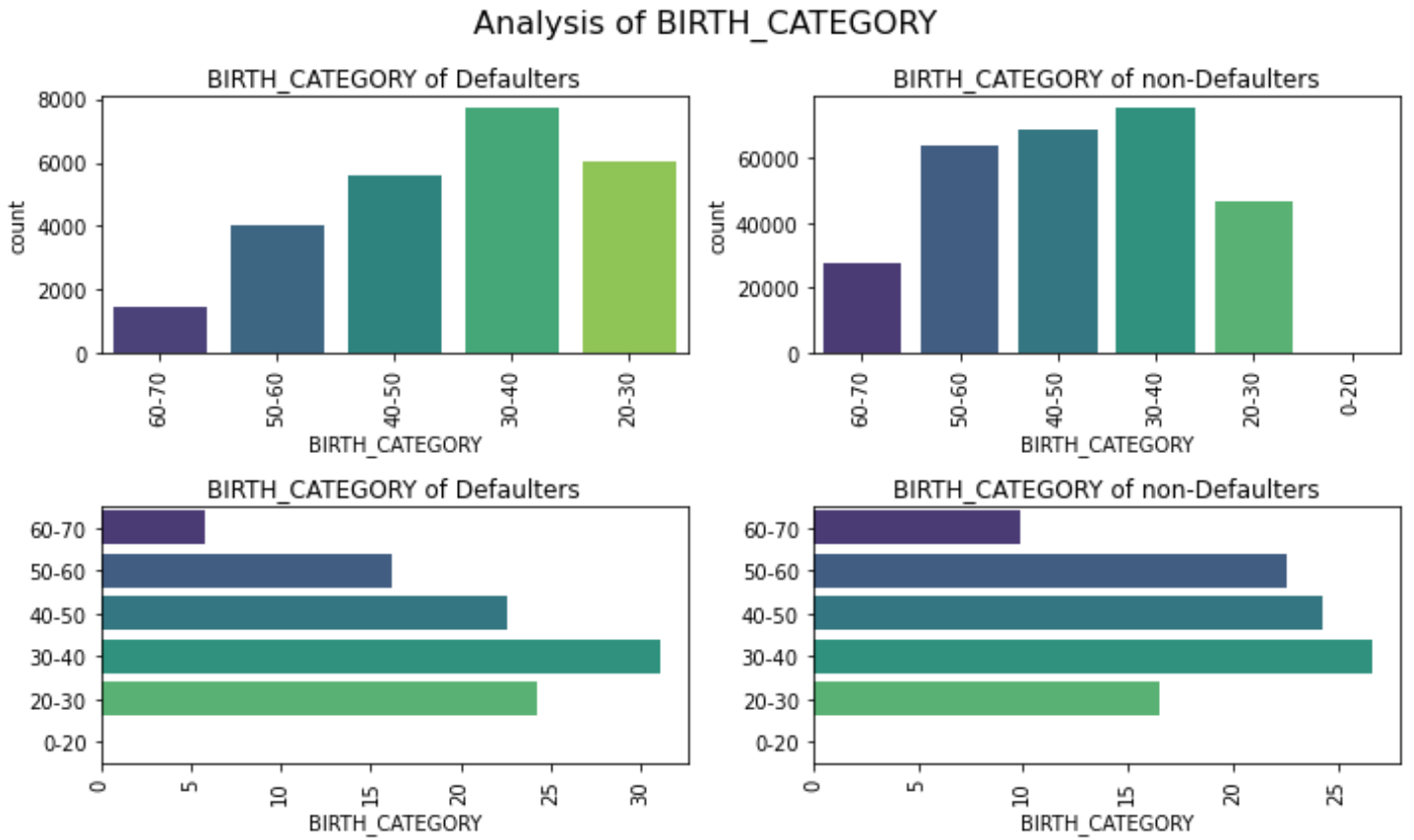
AMT_INCOME_RANGE

- Applicants with Very High and High Income range have moke better on time payments.



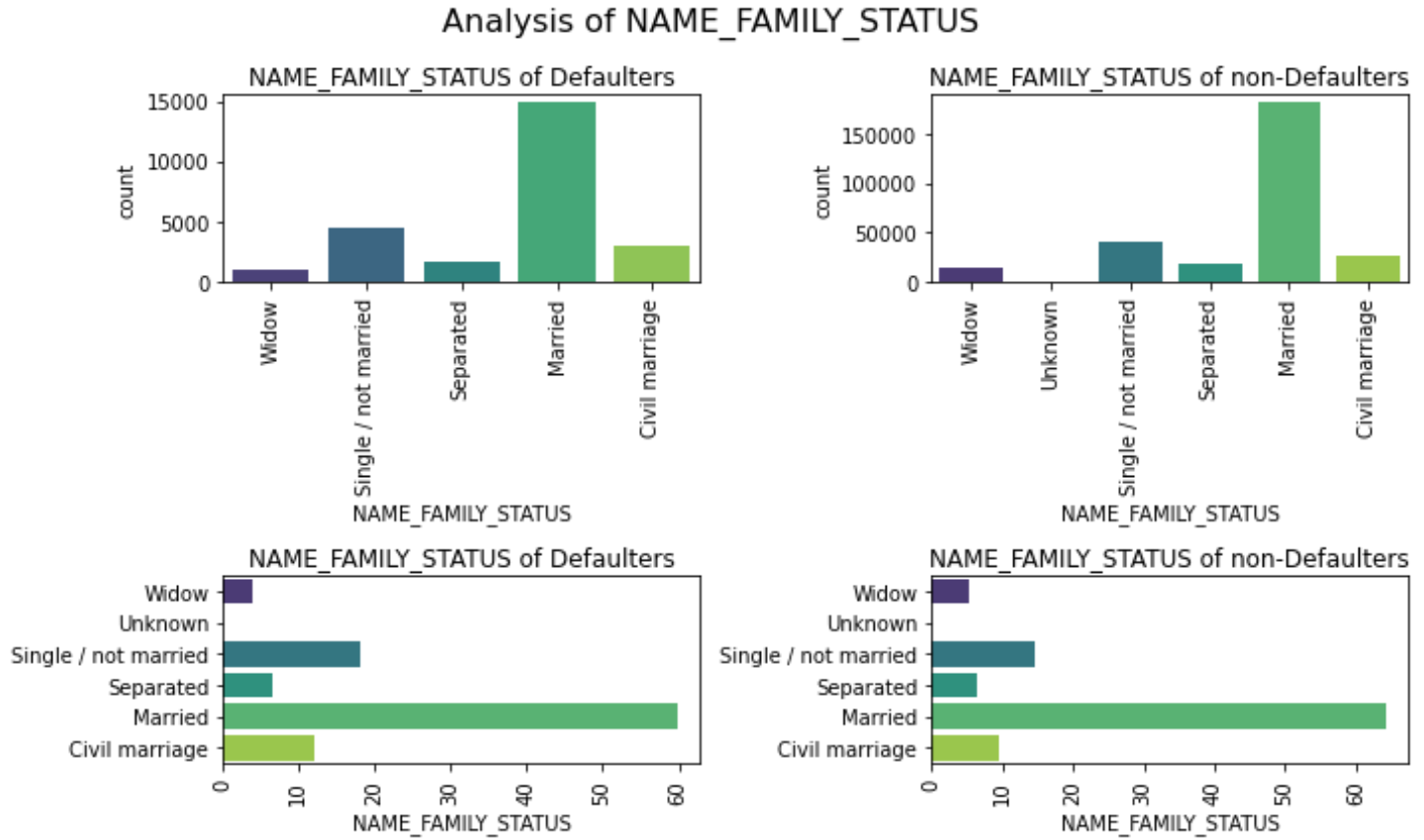
BIRTH_CATEGORY

- Applicants aged more than 30-40 make better ontime payments.
- Applicants in the age range 20-30 have a less impressive repayment trend.



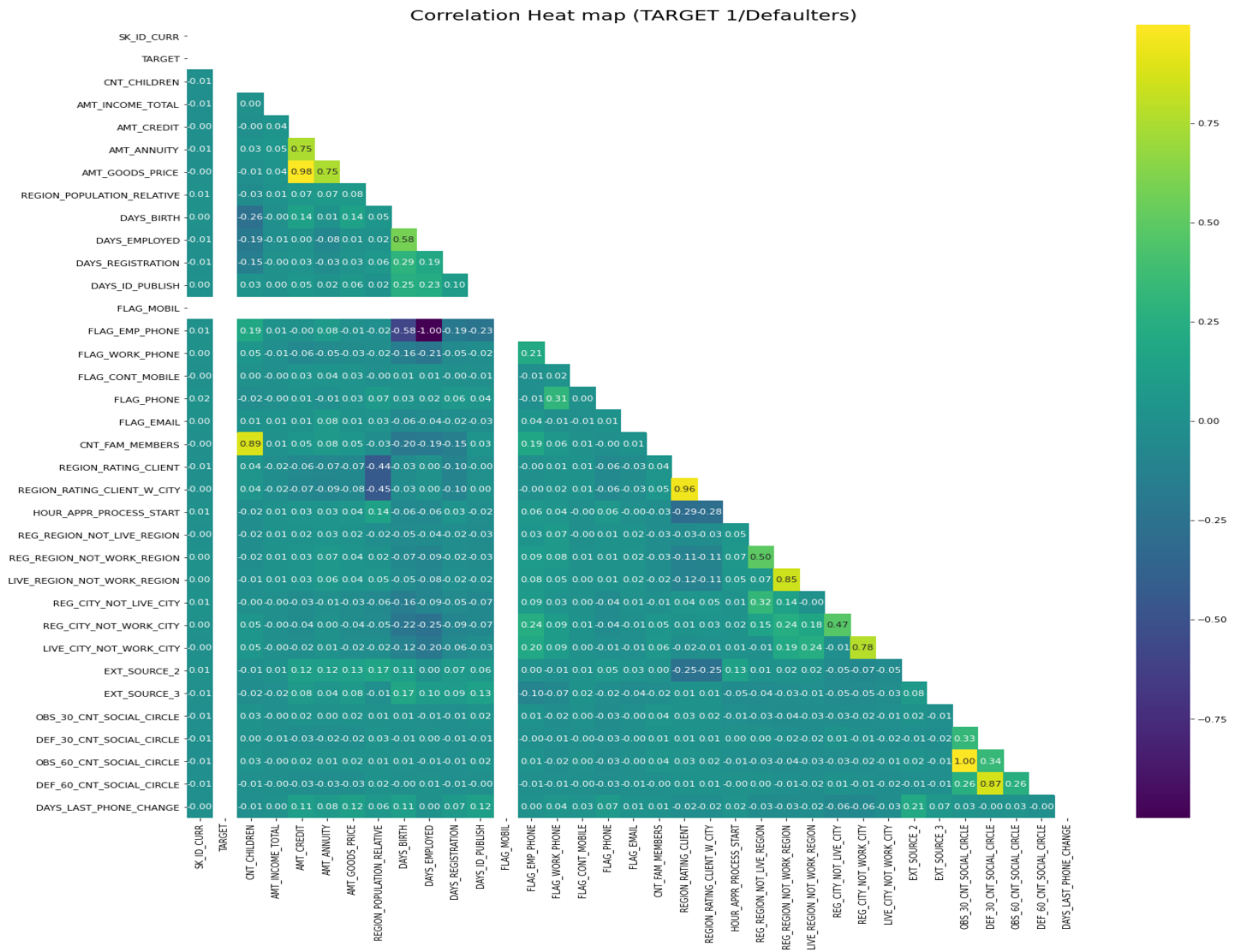
NAME_FAMILY_STATUS

- Applicants who are Widow,Single/Not Married and in a Civil marriage are better on time payers.
- Applicants in Married Category find it difficult to repay.



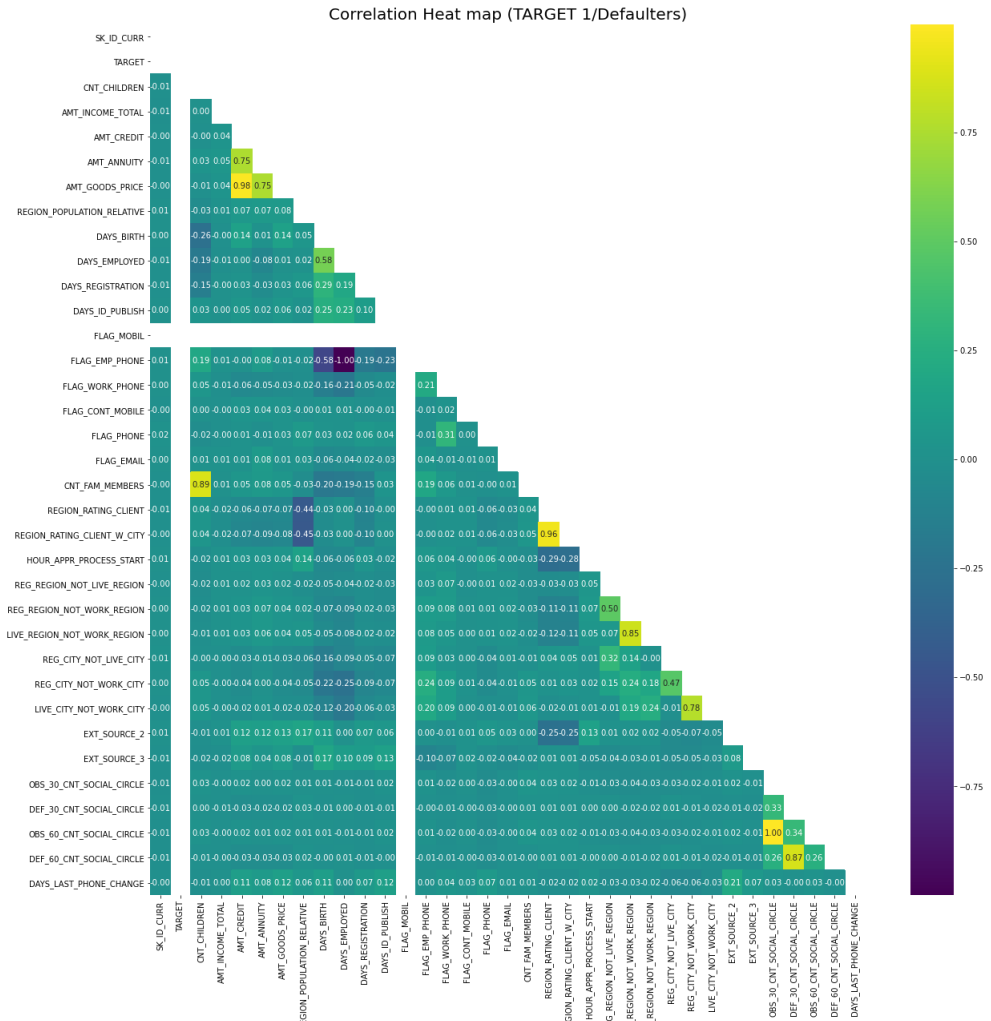
ANALYSIS OF NUMERICAL VARIABLES - *CORRELATION ANALYSIS OF NUMERICAL VARIABLES*

- Top 10 Correlation of Defaulters



ANALYSIS OF NUMERICAL VARIABLES - *CORRELATION ANALYSIS OF NUMERICAL VARIABLES*

- Top 10 Correlation non Defaulters



TOP 10 CORRELATION OF DEFAULTERS VS NON DEFAULTERS

- Correlation between AMT_CREDIT and AMT_GOODS_PRICE in the the highest between Defaulters and non Defaulters.
- The correlation between Defaulters and non Defaulters are almost similar.

Defaulters

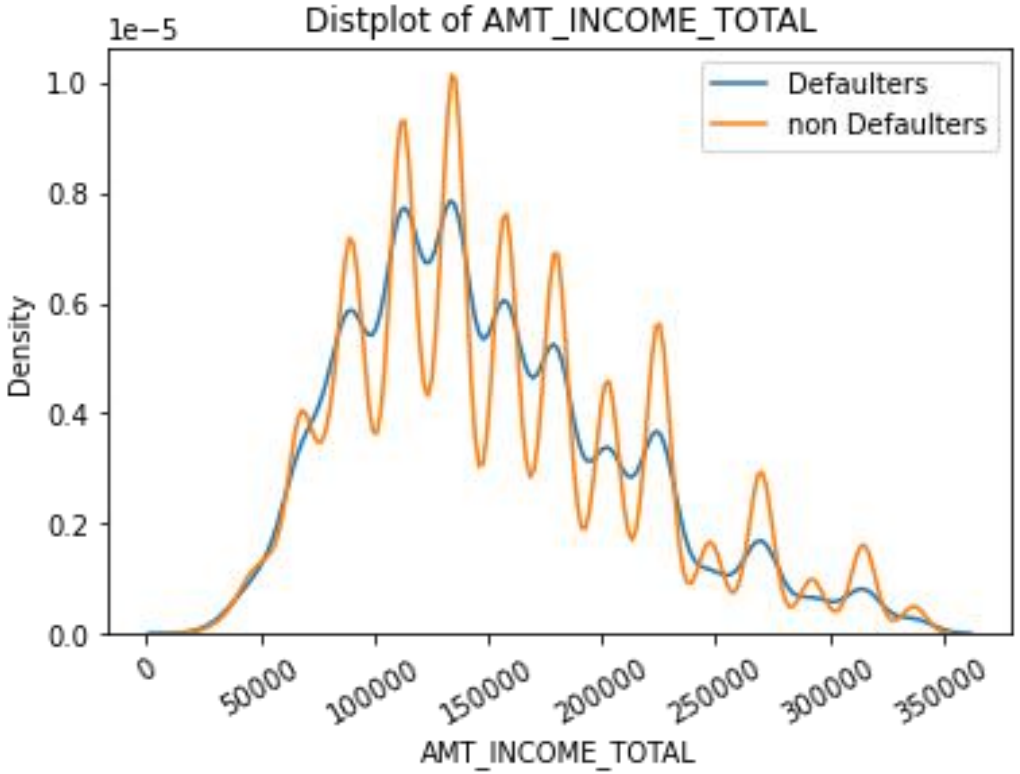
AMT_CREDIT	AMT_GOODS_PRICE	98.31
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	95.66
CNT_FAM_MEMBERS	CNT_CHILDREN	88.55
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	86.90
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	84.79
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	77.85
AMT_GOODS_PRICE	AMT_ANNUITY	75.27
AMT_CREDIT	AMT_ANNUITY	75.22
DAYS_EMPLOYED	DAYS_BIRTH	58.24
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	49.79
REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	47.21

Non - Defaulters

AMT_GOODS_PRICE	AMT_CREDIT	98.73
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	95.01
CNT_CHILDREN	CNT_FAM_MEMBERS	87.86
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	86.19
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	85.93
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	83.04
AMT_ANNUITY	AMT_GOODS_PRICE	77.67
AMT_CREDIT	AMT_ANNUITY	77.13
DAYS_BIRTH	DAYS_EMPLOYED	62.60
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	44.61
REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	43.55

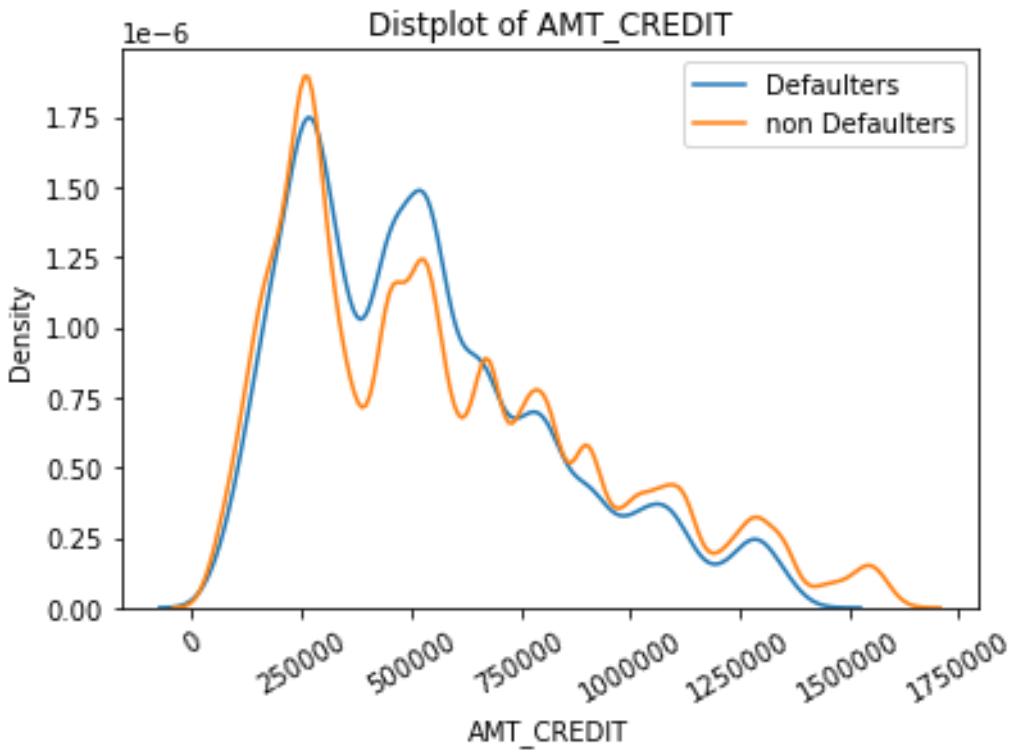
UNIVARIATE ANALYSIS OF NUMERICAL VARIABLES
- AMT_INCOME_TOTAL

- AMT_INCOME_TOTAL
  AMT_INCOME_TOTAL of Defaulters --> IQR: 90000.0 Min_value: -22500.0 Max_value: 337500.0 AMT_INCOME_TOTAL of non Defaulters --> IQR: 90000.0 Min_value: -22500.0 Max_value: 337500.0
- Both Defaulters and non defaulters is similar to a normal distribution.
- Non Defaulters trend of repayment shows spikes, however, is inconclusive of anything.



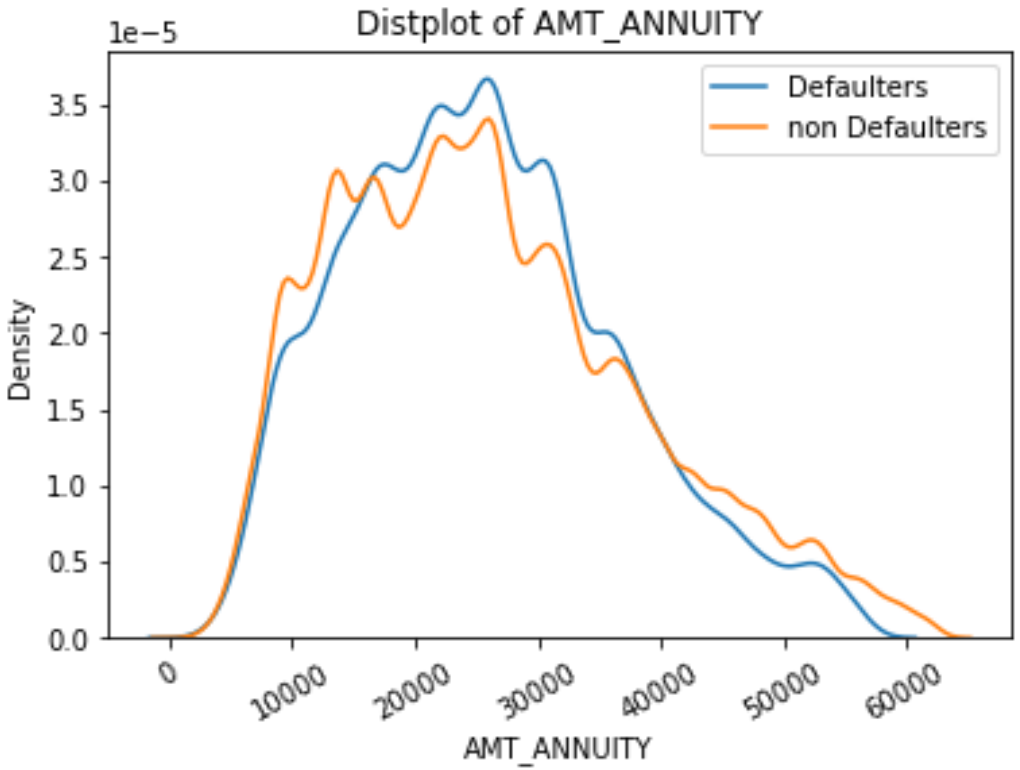
AMT_CREDIT

- AMT_CREDIT of Defaulters --> IQR: 448915.5 Min_value: -388973.25 Max_value: 1406688.75 AMT_CREDIT of non Defaulters --> IQR: 540000.0 Min_value: -540000.0 Max_value: 1620000.0
- It is observed that from appx 250000 until 600000, the Defaulters are high.
- For AMT_CREDIT appx 250000 and above 750000, the applicants are able to make payments on time.



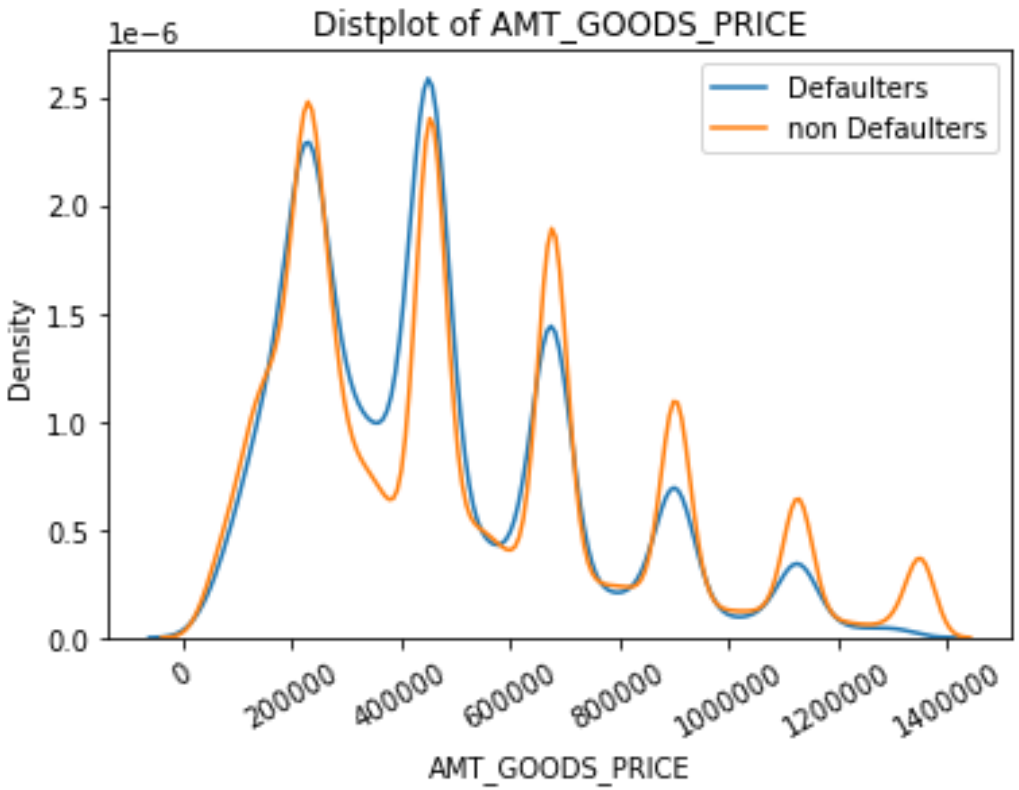
AMT_ANNUITY

- AMT_ANNUITY of Defaulters --> IQR: 15615.0 Min_value: -6061.5 Max_value: 56398.5 AMT_ANNUITY of non Defaulters --> IQR: 18292.5 Min_value: -10982.25 Max_value: 62187.75
- For AMT_ANNUITY, less tha 15000 we can see there are lessno of defaulters.
- AMT_ANNUITY in the range 15000 till appx 40000 we are able to see there are more deaulters.
- Repayment trend is higher as the AMT_ANNUITY increases above 40000



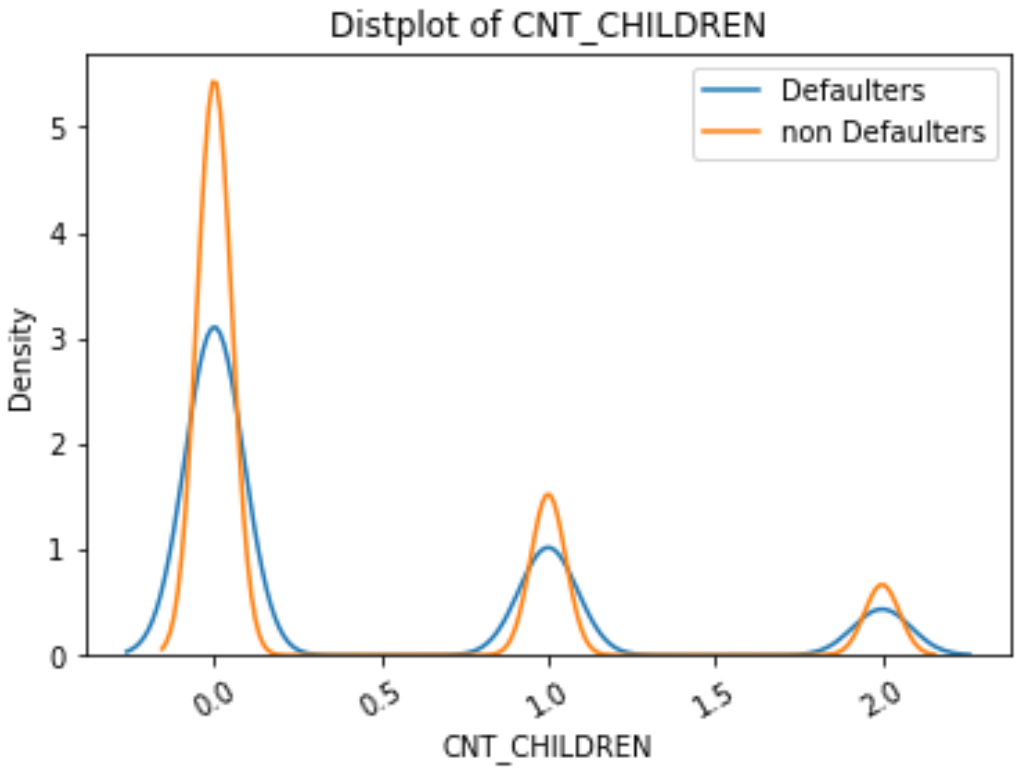
AMT_GOODS_PRICE

- AMT_GOODS_PRICE of Defaulters --> IQR: 436500.0 Min_value: -416250.0 Max_value: 1329750.0 AMT_GOODS_PRICE of non Defaulters --> IQR: 450000.0 Min_value: -436500.0 Max_value: 1363500.0
- For AMT_GOODS_PRICE between ~250000 and ~550000, there are more clients with Payment difficulties.



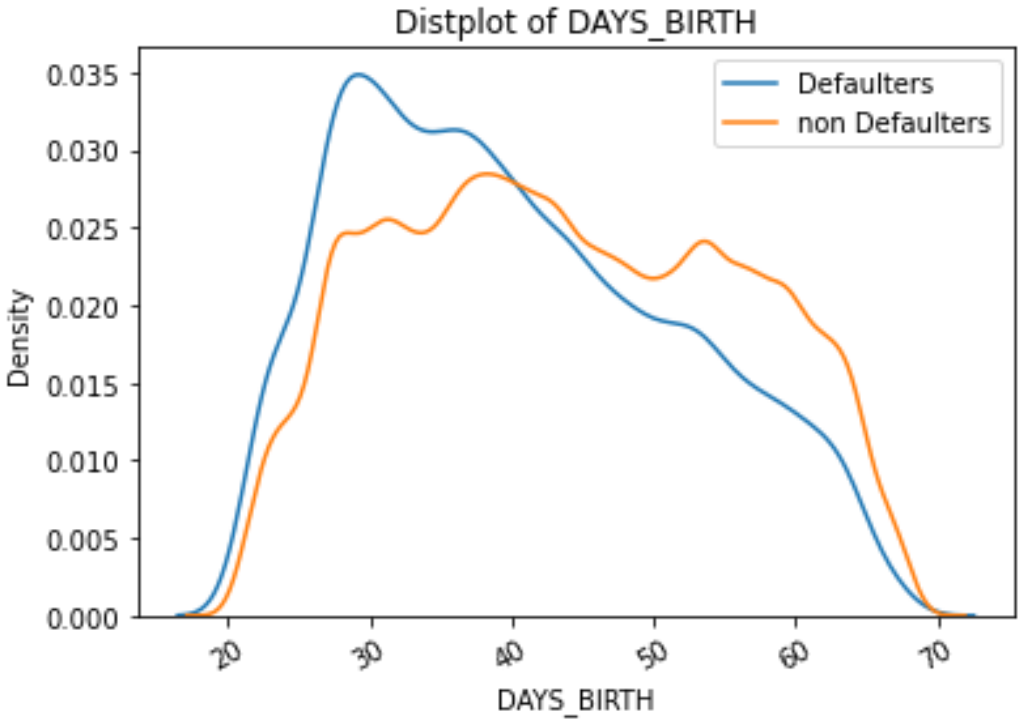
CNT_CHILDREN

- CNT_CHILDREN of Defaulters --> IQR: 1.0 Min_value: -1.5 Max_value: 2.5
  CNT_CHILDREN of non Defaulters --> IQR: 1.0 Min_value: -1.5 Max_value: 2.5
- For CNT_CHILDREN 0 , there are lots of clients with On-Time Payments.
- For CNT_CHILDREN with 1 OR 2, there are few more clients with On-Time Payments.
- This does not provide us with enough evidence.



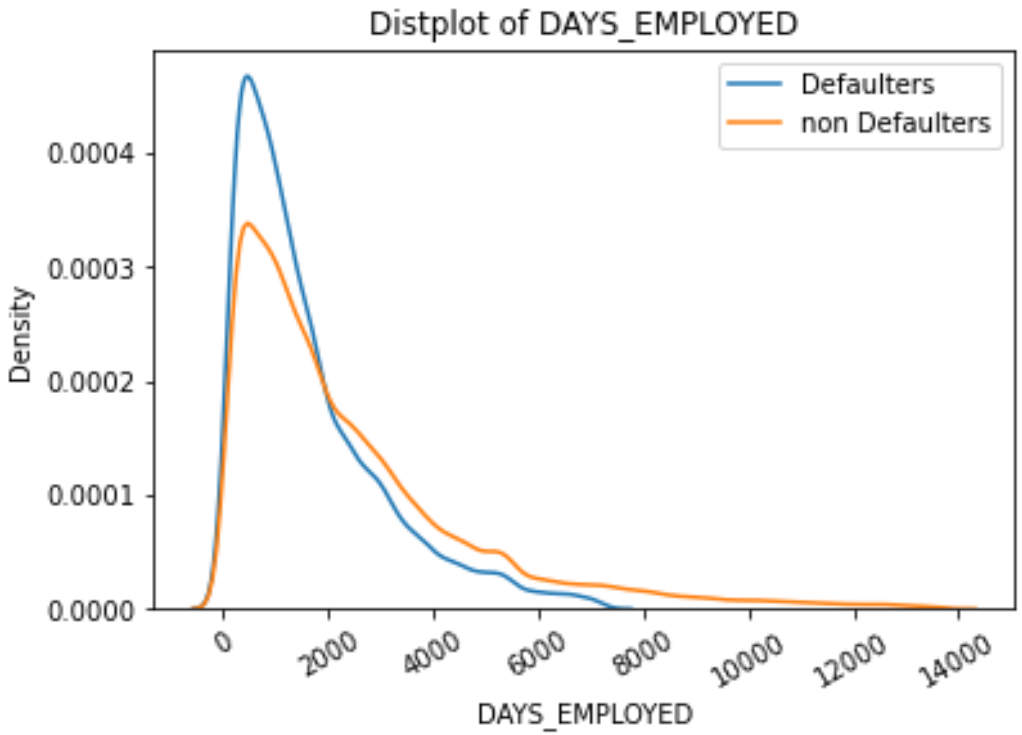
DAYS_BIRTH

- DAYS_BIRTH of Defaulters --> IQR: 18.0 Min_value: 4.0 Max_value: 76.0 DAYS_BIRTH of non Defaulters --> IQR: 20.0 Min_value: 4.0 Max_value: 84.0
- For DAYS_BIRTH less than 40 there is a bad trend in repayment.
- For DAYS_BIRTH ablove 40 does have a better on time repayment.



DAYS_EMPLOYED

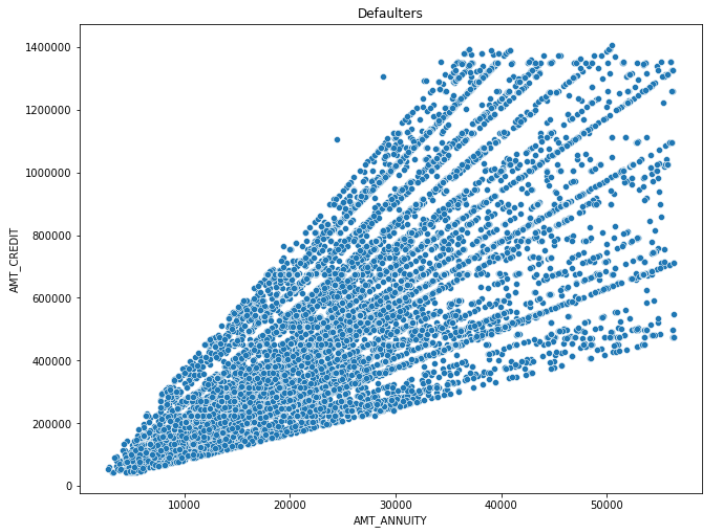
- DAYS_EMPLOYED of Defaulters --> IQR: 2603.0 Min_value: -3227.5 Max_value: 7184.5 DAYS_EMPLOYED of non Defaulters --> IQR: 5107.0 Min_value: -6693.5 Max_value: 13734.5
- For DAYS_EMPLOYED less than 2000, there are more clients with Payment difficulties.
- But, for DAYS_EMPLOYED > 2000 , there are more clients with On-Time Payments.
- This means that those who are employed longer have better chances of repaying the loa

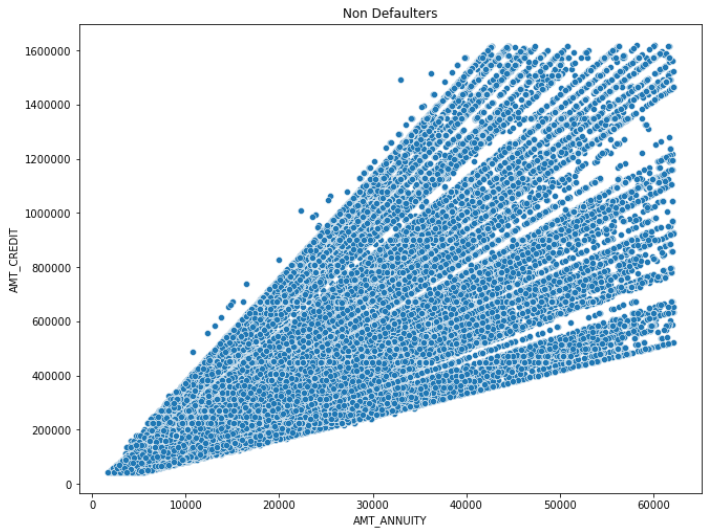


BIVARIATE / MULTIVARIATE ANALYSIS

CONTINUOUS V/S CONTINUOUS VARIABLES
ANALYSIS OF AMT_ANNUITY V/S AMT_CREDIT

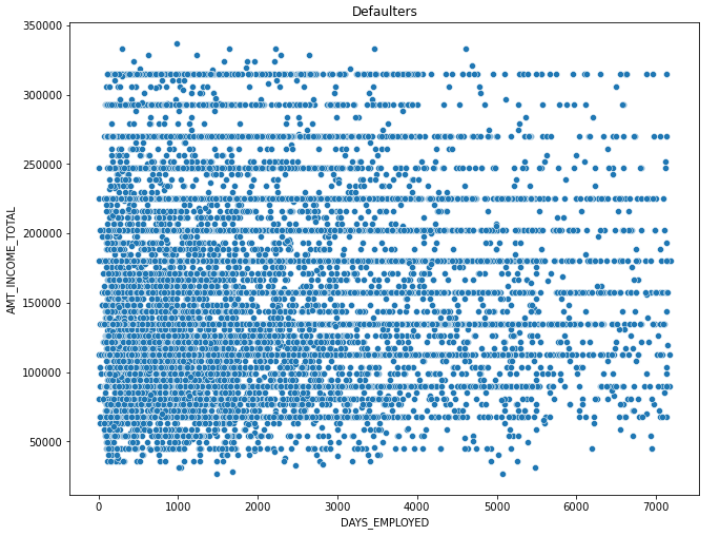
- AMT_ANNUITY and AMT_CREDIT are positively correlated.

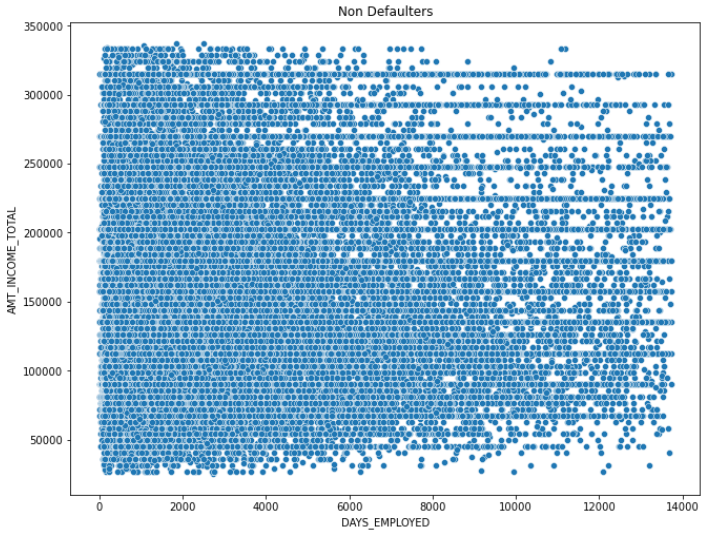




ANALYSIS OF DAYS_EMPLOYED V/S AMT_INCOME_TOTAL

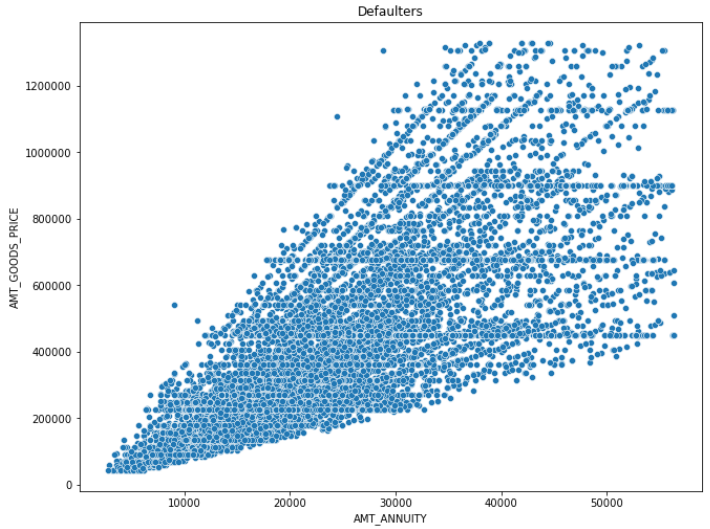
- There is no Correlation between DAYS_EMPLOYED and AMT_INCOME_TOTAL

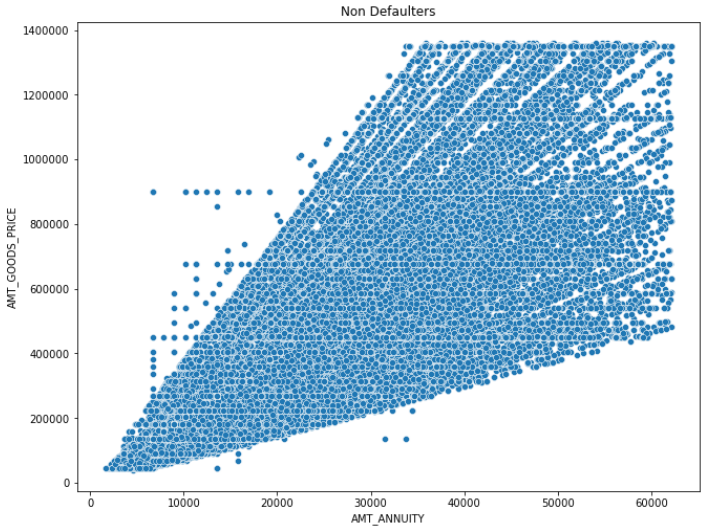




ANALYSIS OF AMT_ANNUITY V/S AMT_GOODS_PRICE

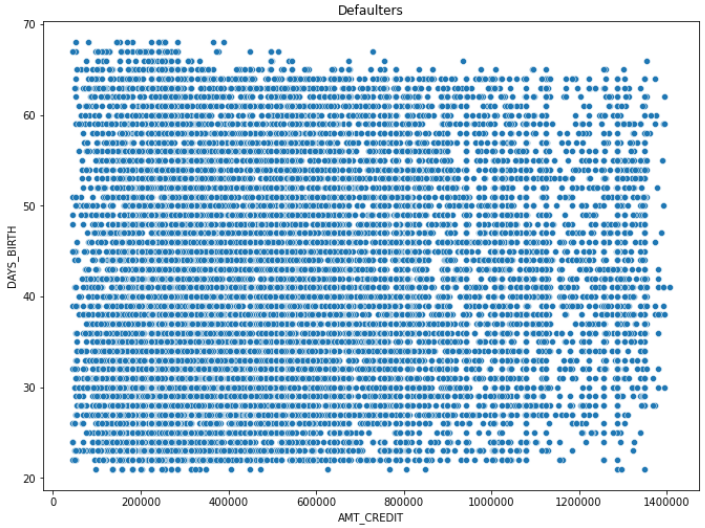
- AMT_ANNUITY and AMT_GOODS_PRICE are positively correlated

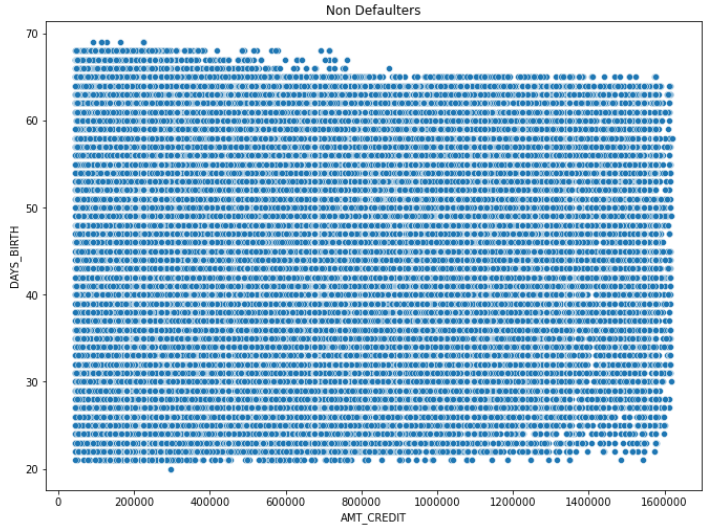




ANALYSIS OF AMT_CREDIT V/S DAYS_BIRTH

- There is no correlation between AMT_CREDIT and DAYS_BIRTH

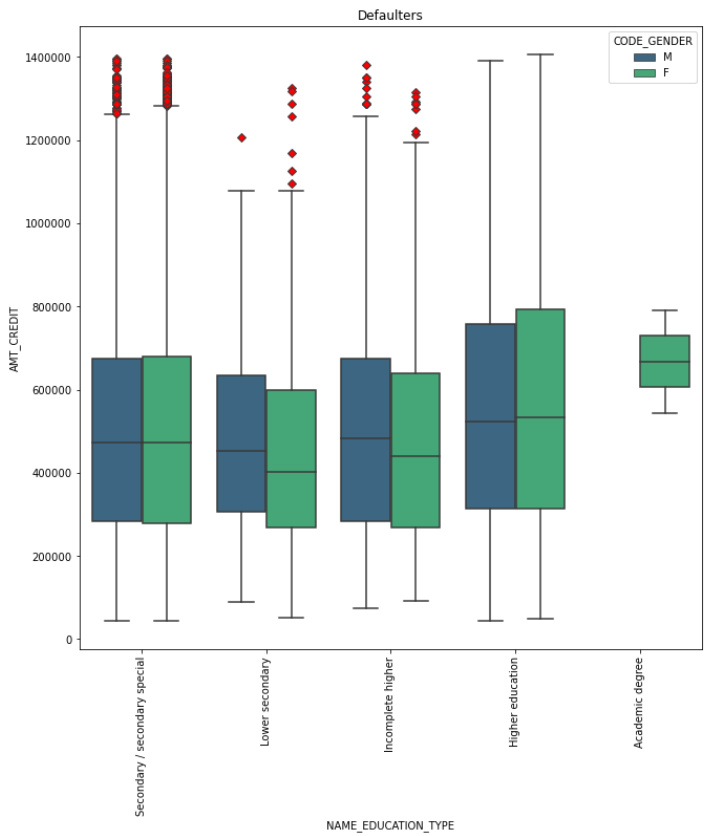


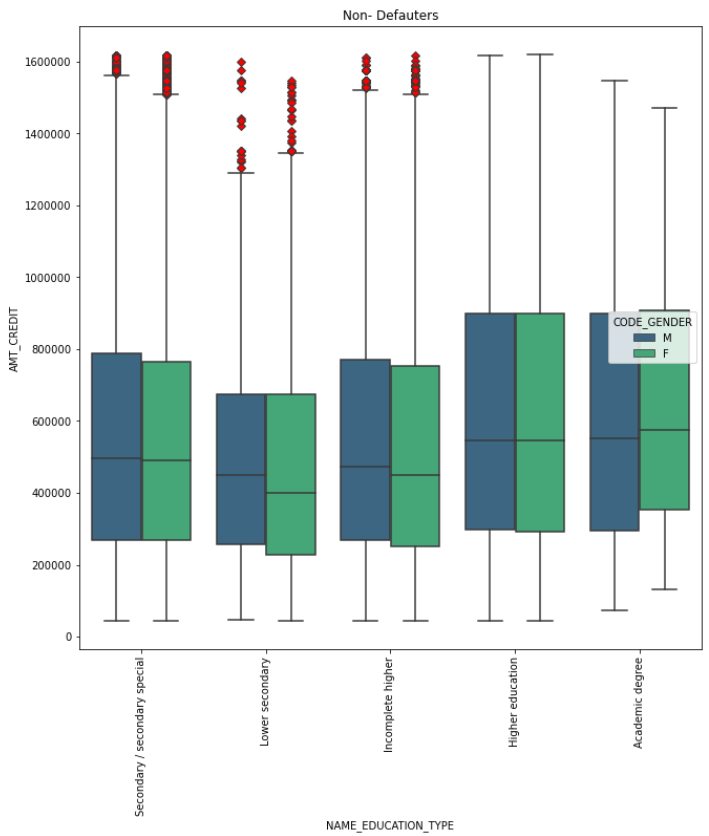


CONTINUOUS V/S CATEGORICAL VARIABLES

ANALYSIS OF NAME_EDUCATION_TYPE V/S AMT_CREDIT V/S CODE_GENDER

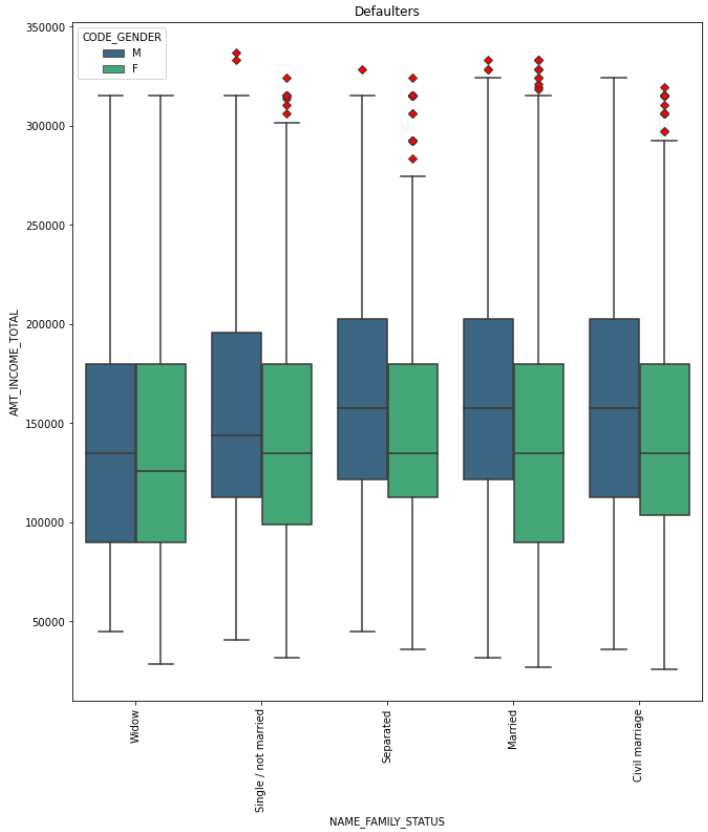
- Male with Academic degree have make all the payments on time.
- Clients with Academic degree are wide range of Amount credit in non defaulters and very less in defaulters.

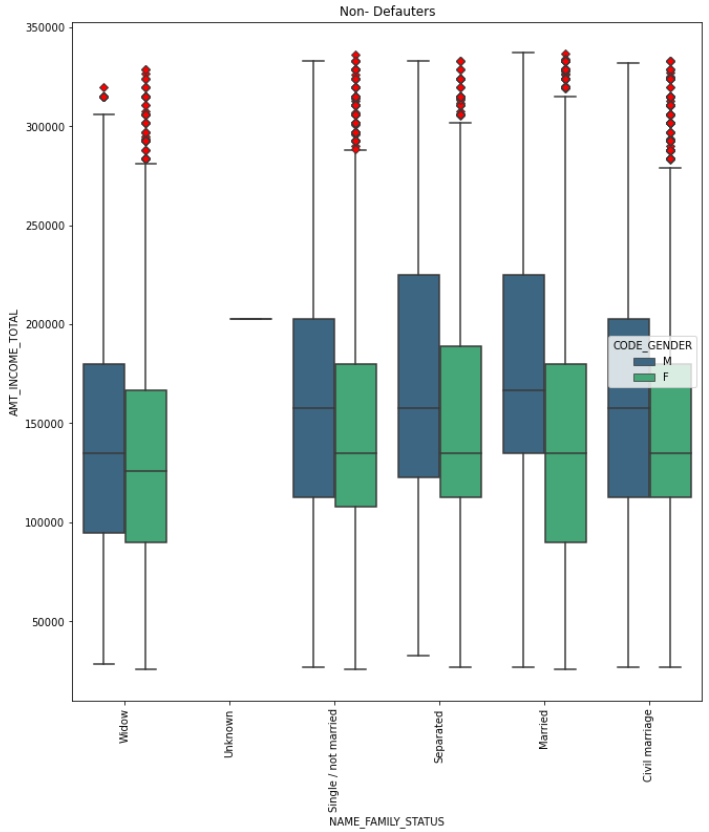




ANALYSIS OF NAME_FAMILY_STATUS V/S AMT_INCOME_TOTAL V/S CODE_GENDER

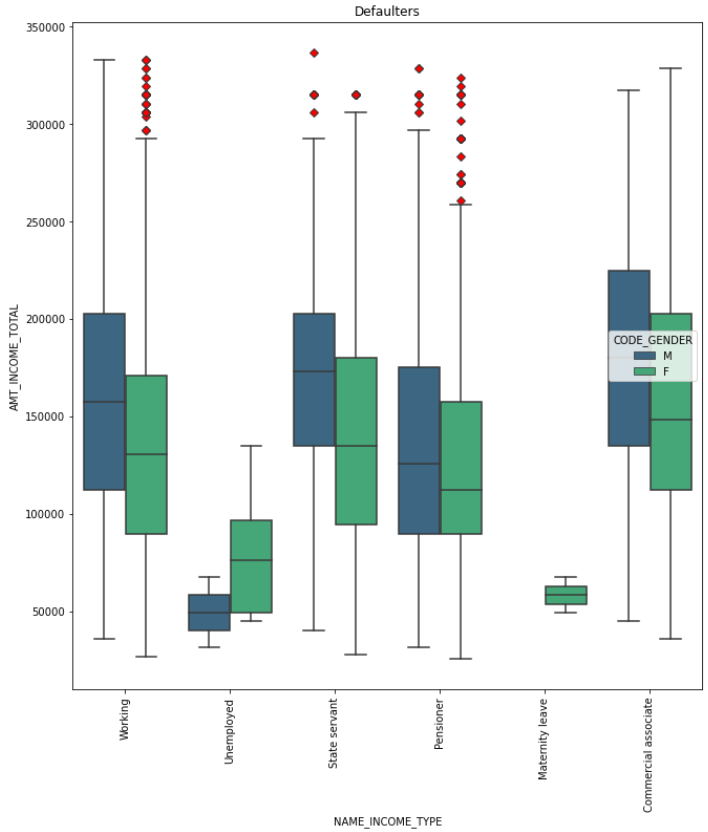
- Married men have high mean and better repayment trend.

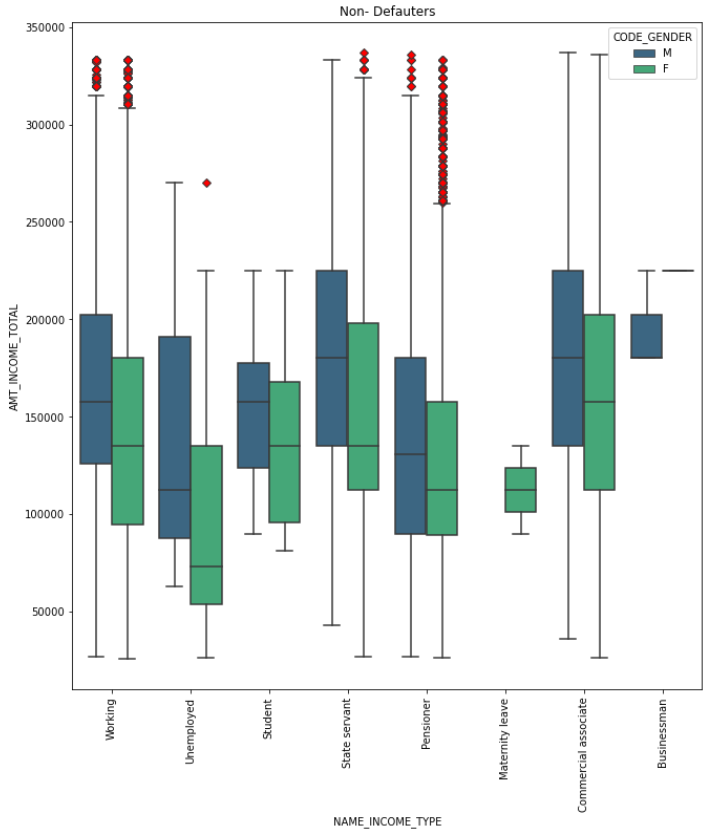




ANALYSIS OF NAME_INCOME_TYPE V/S AMT_INCOME_TOTAL V/S CODE_GENDER

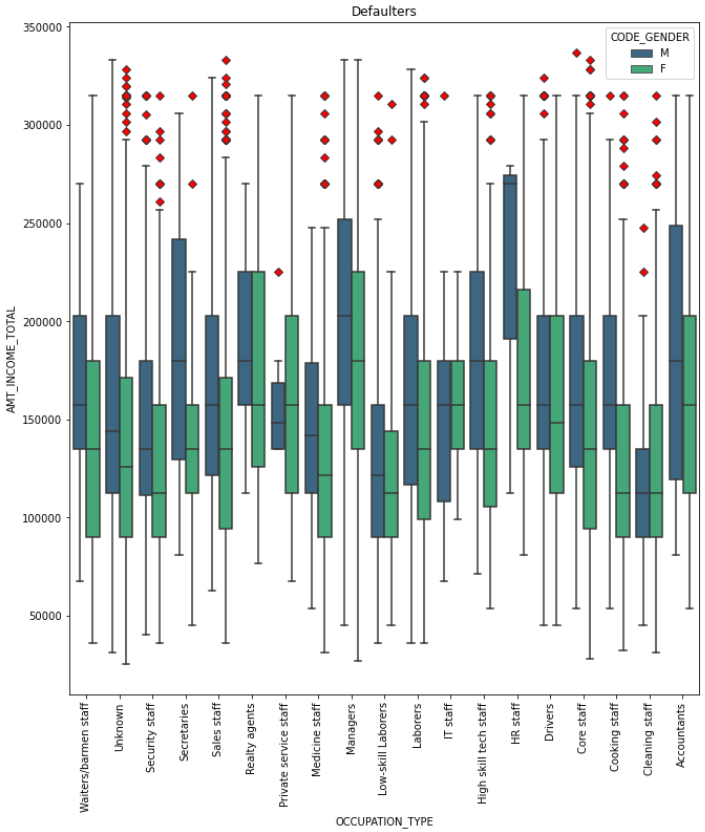
- Applicants who are Unemployed and very high income range have netter repayment trend.
- Student both Male and Female make on time payments.
- Poor repayment trend observed in Male state servants.
- Businessmen (Male and Female) who take loan make are not observed as defaulters.
- Female applicants in maternity and high income have better repayment trend.

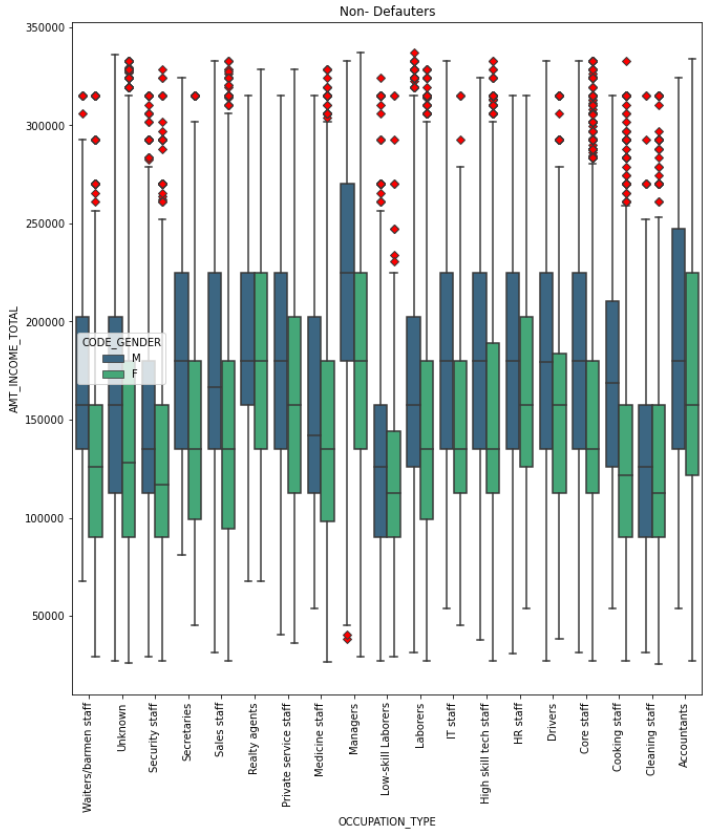




ANALYSIS OF OCCUPATION_TYPE V/S AMT_INCOME_TOTAL V/S CODE_GENDER

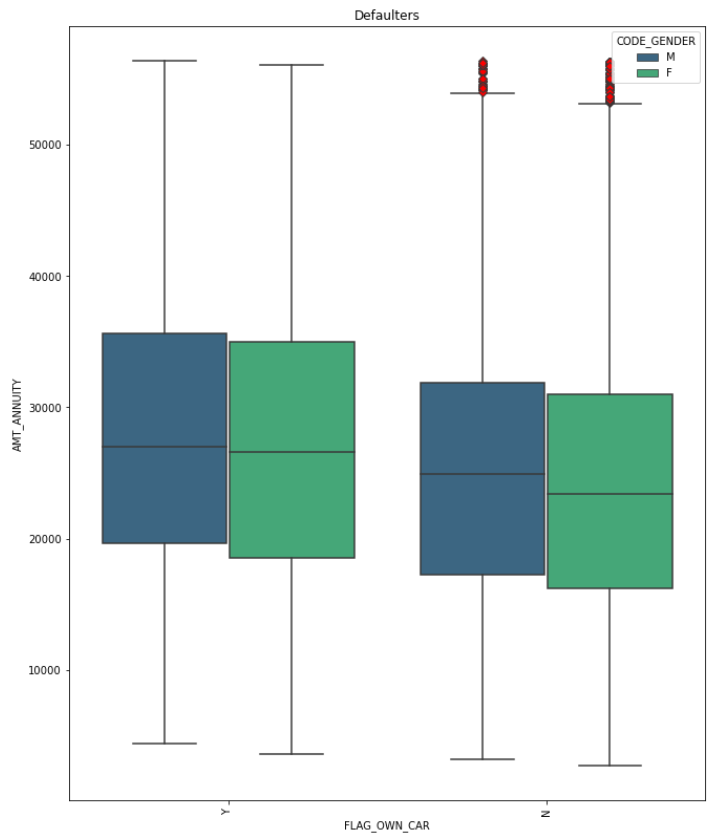
- Cleaning staff Male have poor Income in defaulter than non defaaulters.
- HR staff (Men) with high total income have poor trend in ontime payment.
- IT staffs and Private Services (Men) have a better repayment trend

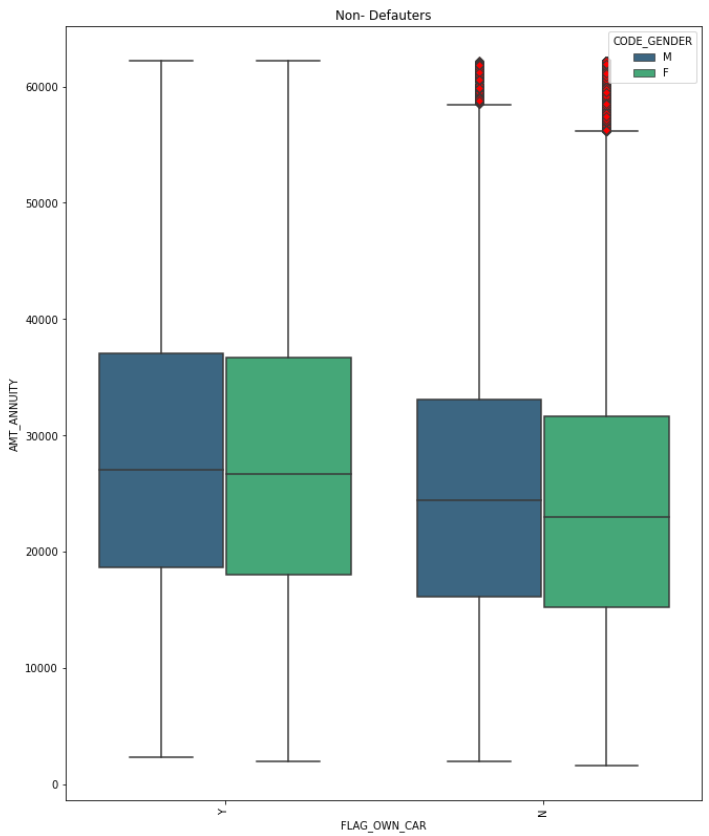




ANALYSIS OF FLAG_OWN_CAR V/S AMT_ANNUITY V/S CODE_GENDER

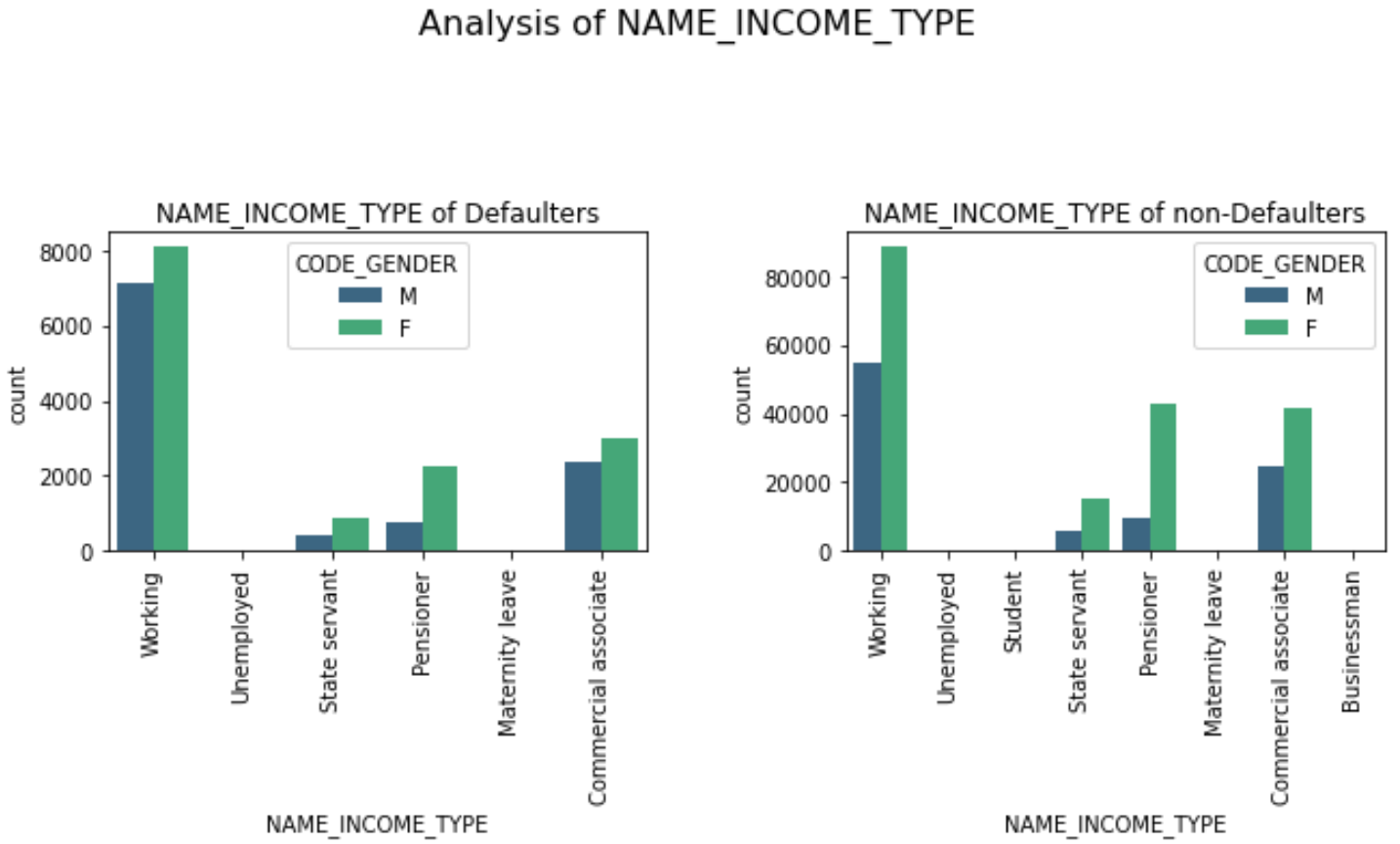
- No significant observations here.





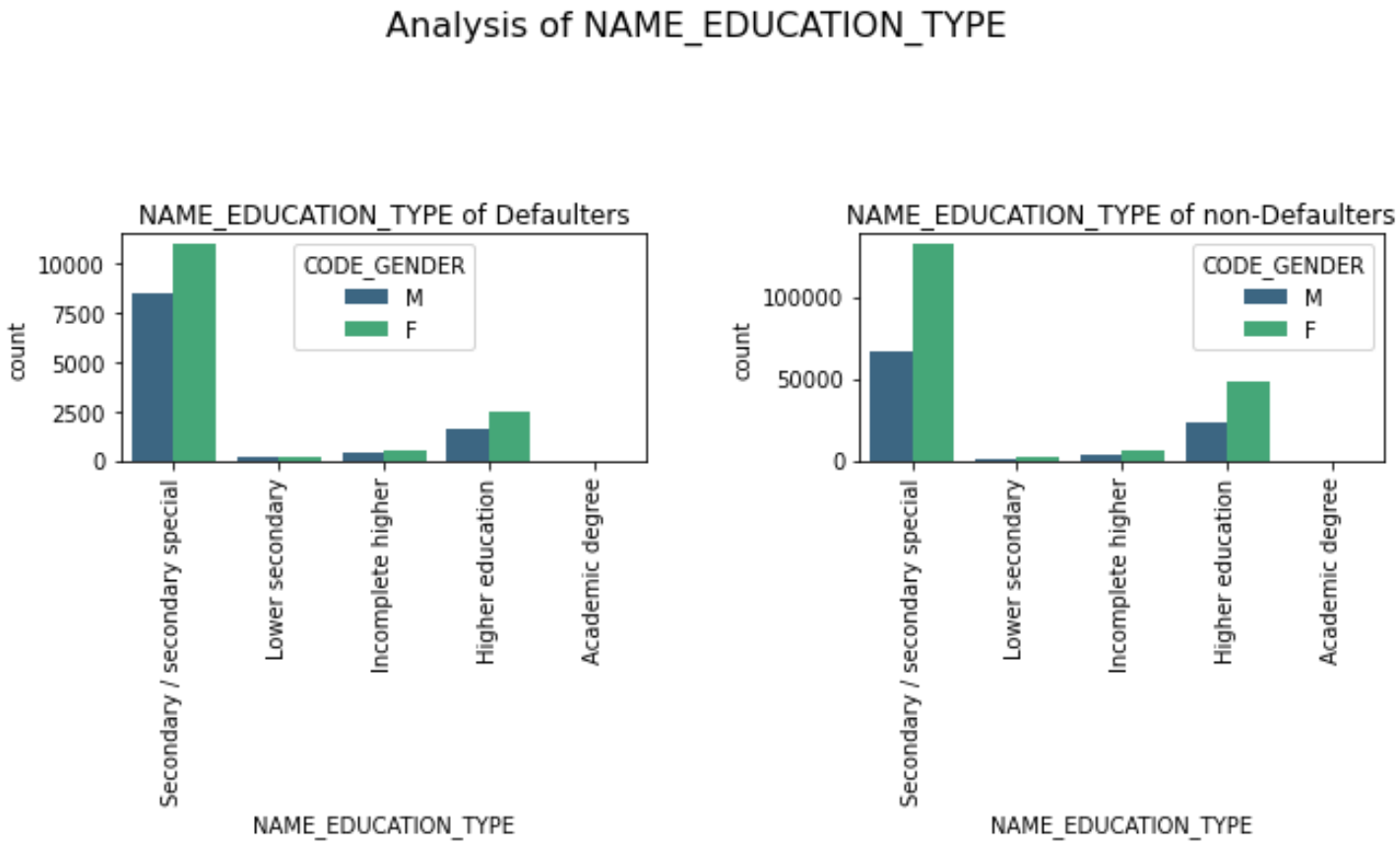
CATEGORICAL V/S CATEGORICAL VARIABLES
ANALYSIS OF NAME_INCOME_TYPE V/S CODE_GENDER

- Businesspeople and Students make on time payments.
- Working Men make lesser on time payments than women.
- Pension especially women have payment difficulties.



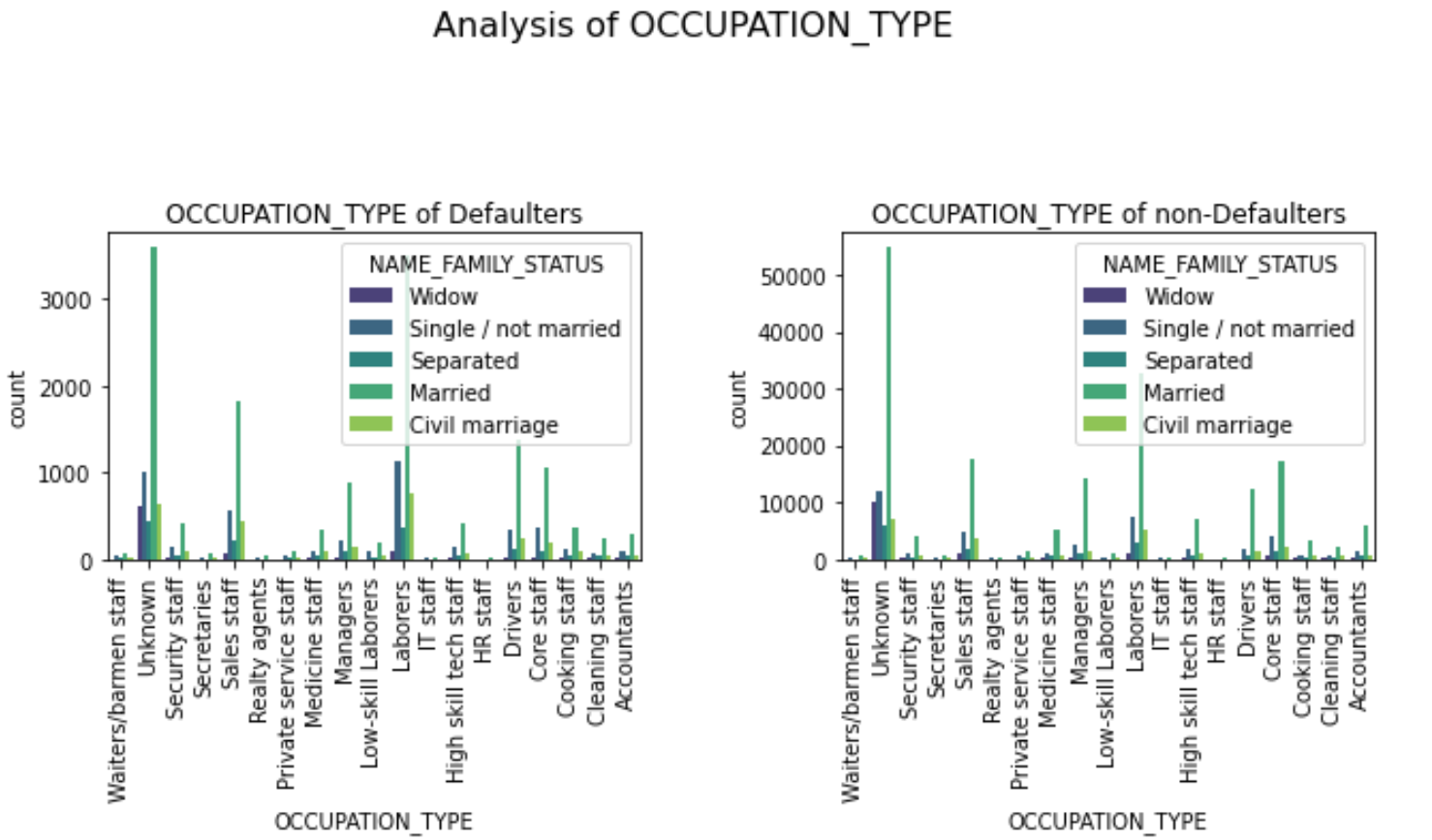
ANALYSIS OF NAME_EDUCATION_TYPE V/S CODE_GENDER

- Applicants having Secondary/Secondary special education and Male have more Payment difficulties compared to On-Time Payments.
- Applicants having Higher education and Female make more On-Time Payments compared to Payment difficulties.



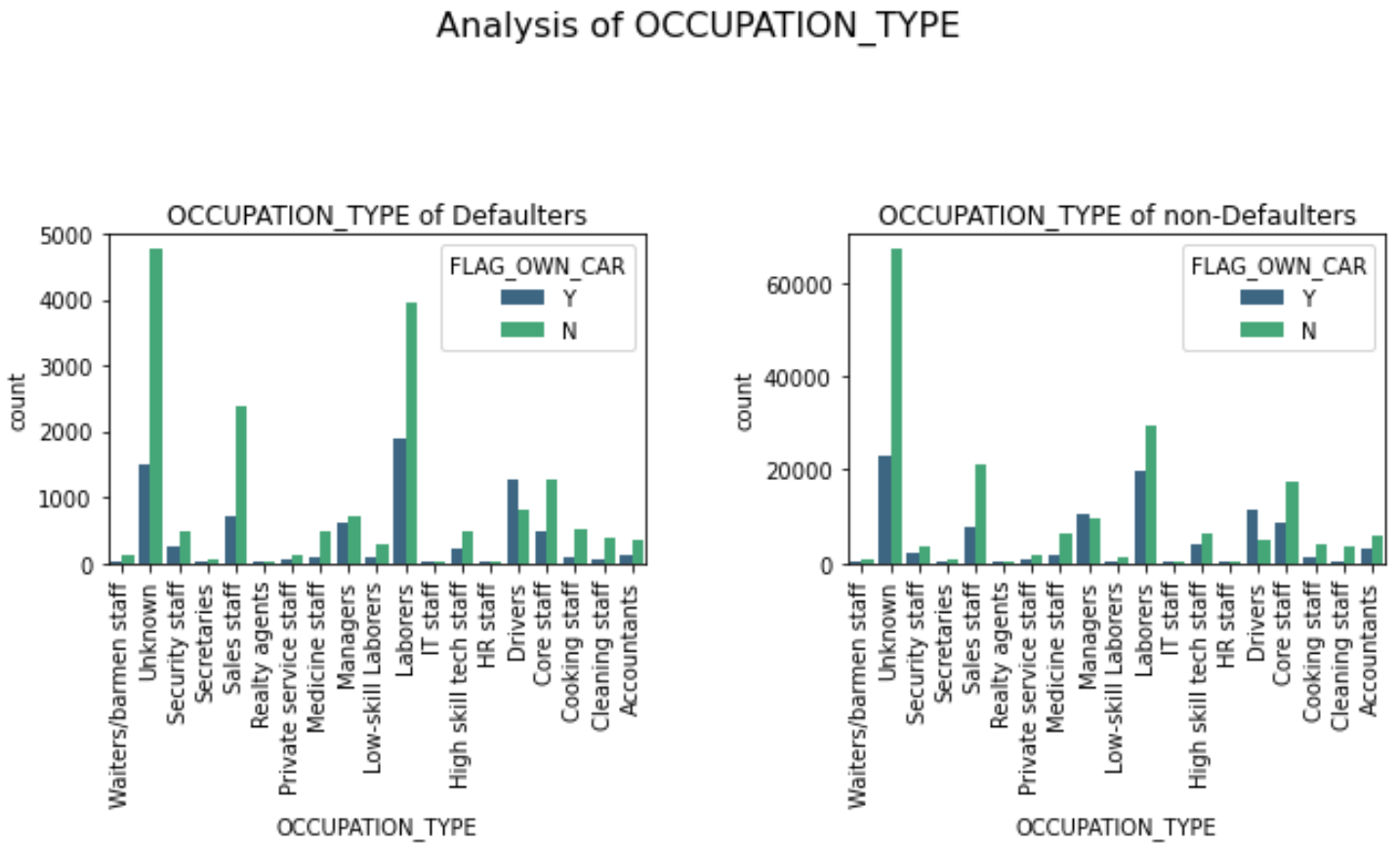
ANALYSIS OF NAME_FAMILY_STATUS V/S OCCUPATION_TYPE

- Applicants Single/not married, Married & Civil marriage and are Waiters/barmen staff have more Payment difficulties compared to On-Time Payments
- Applicants Single/not married & Married and are Laborers are more defaulters.
- Applicants Married and are Drivers default more compared to On-Time Payments Married and Accountants.



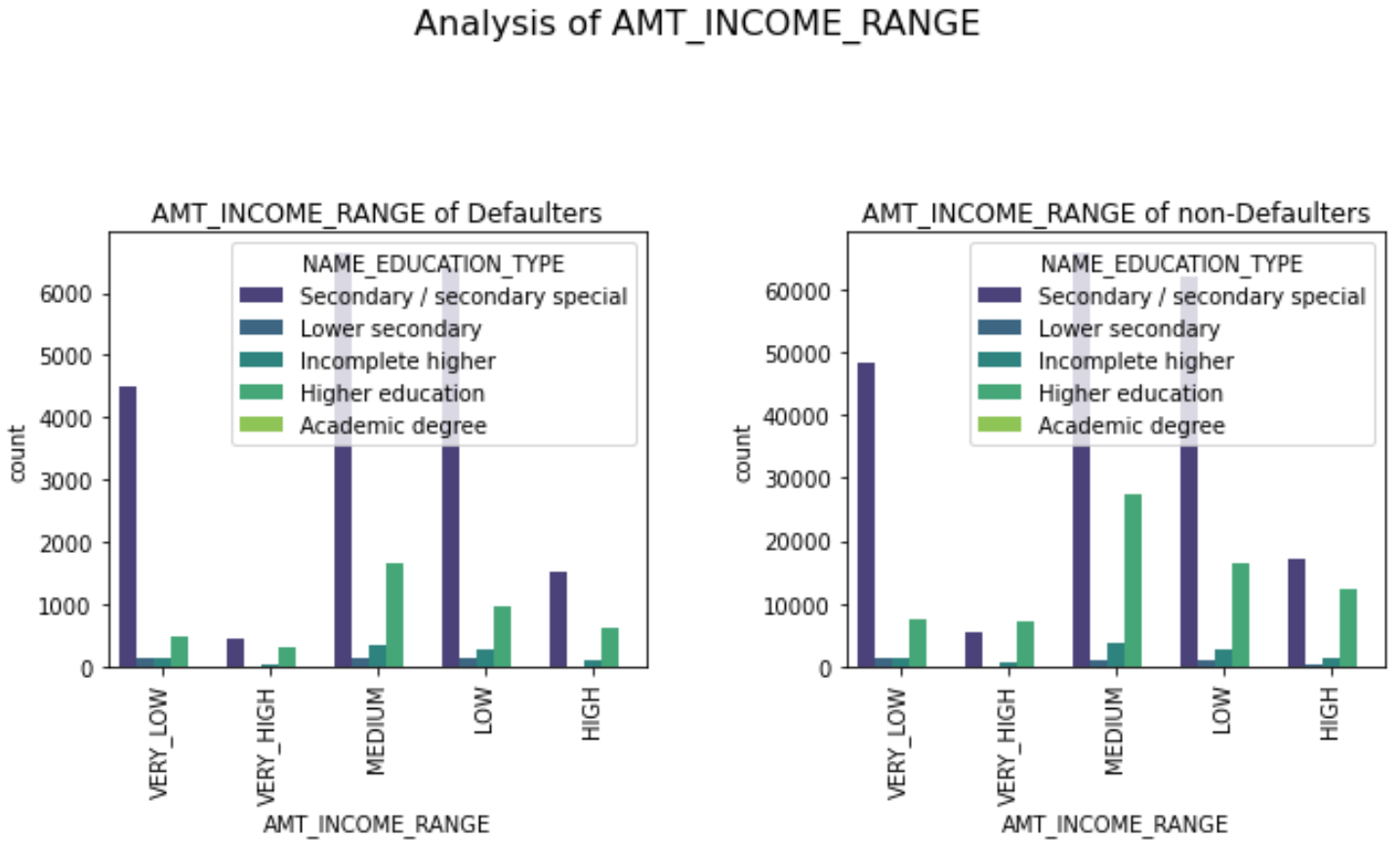
ANALYSIS OF OCCUPATION_TYPE V/S FLAG_OWN_CAR

- Labourers who don't own Car have more Payment difficulties



ANALYSIS OF AMT_INCOME_RANGE V/S FLAG_OWN_CAR

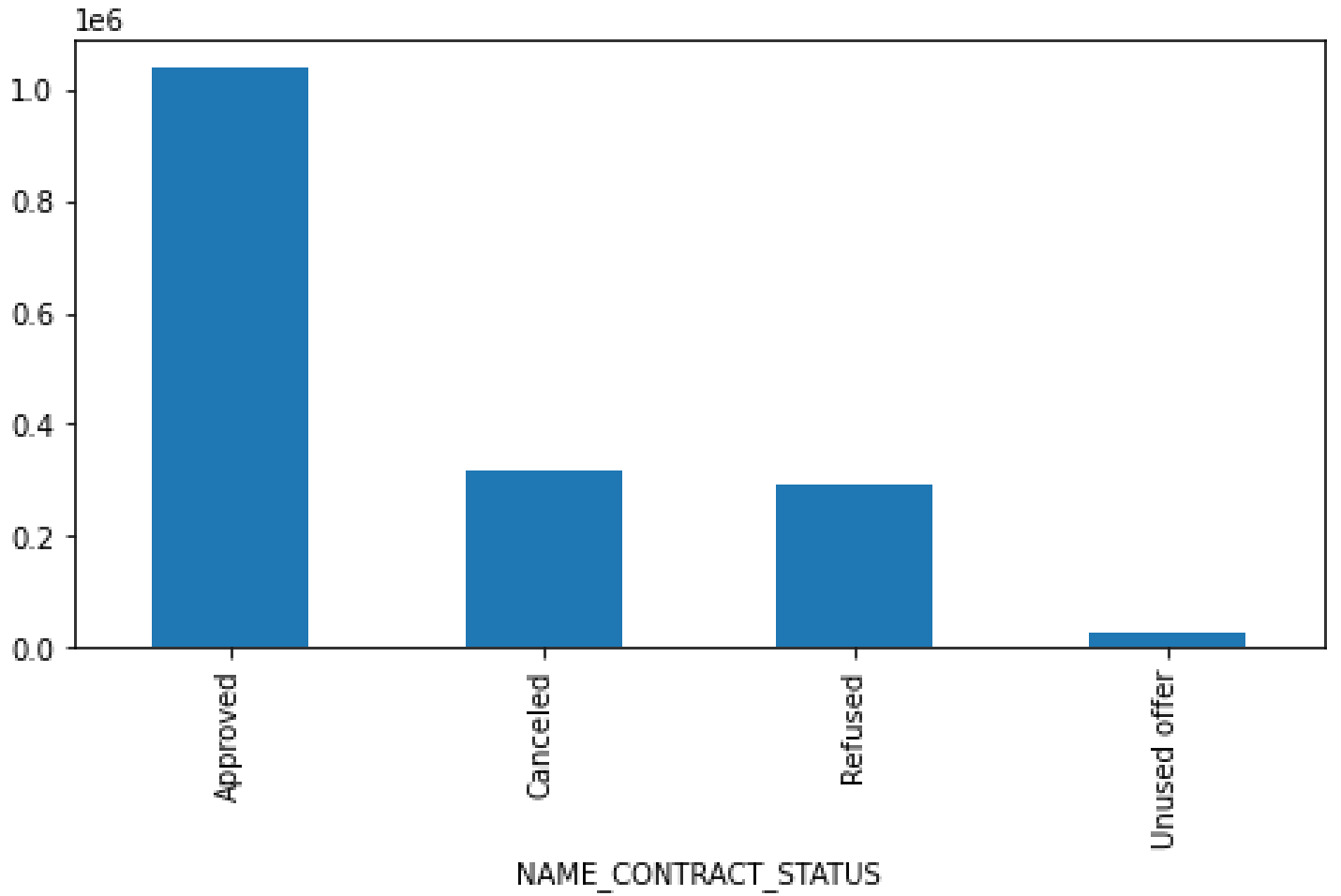
- Applicants with Higher education have a better repayment trend.



PREVIOUS DATA ANALYSIS

UNIVARIATE ANALYSIS - KNOWING THE % OF COLUMNS THAT ARE APPROVED,REFUSED,CANCELLED

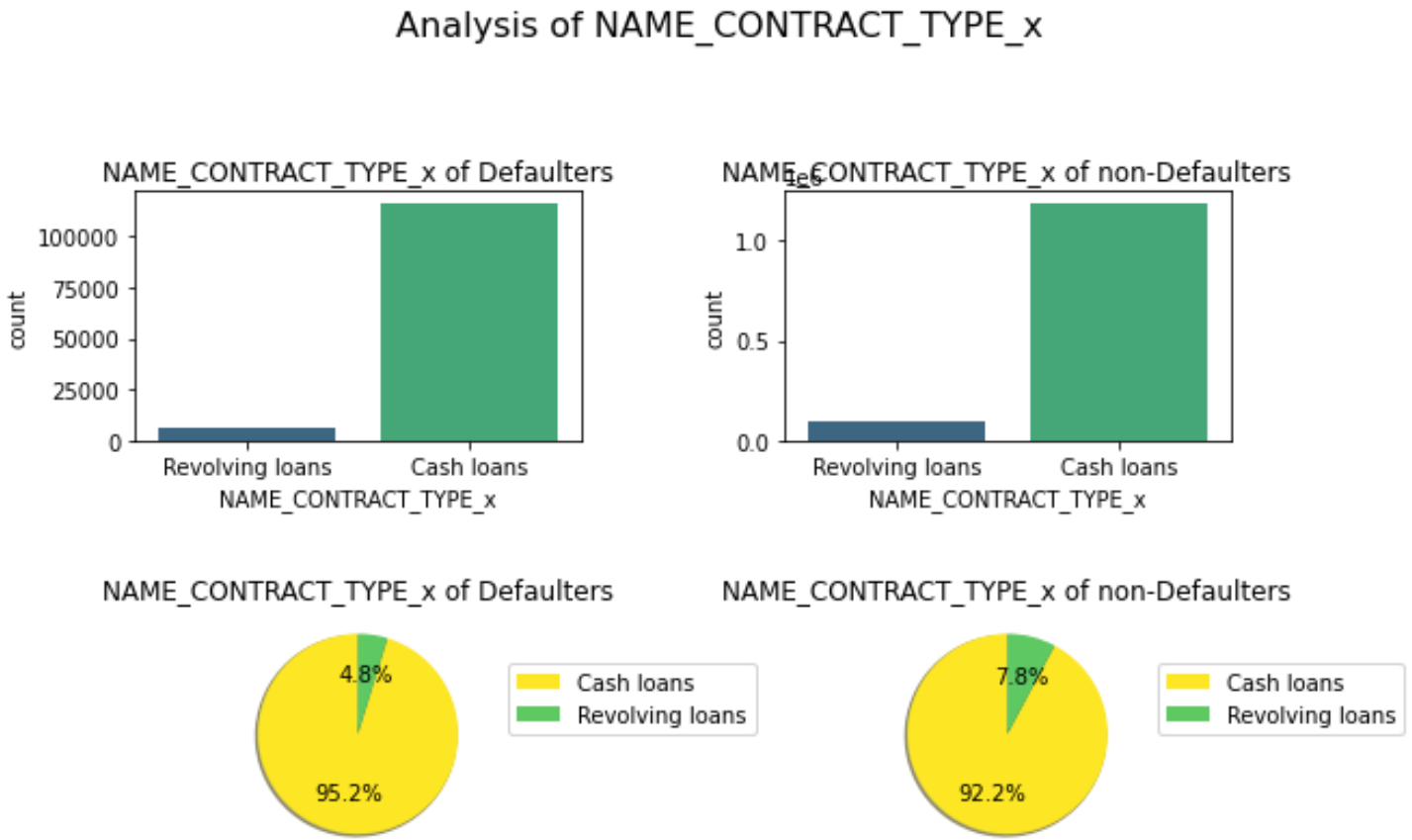
- Previous Application data consist of a greater number of Approved clients

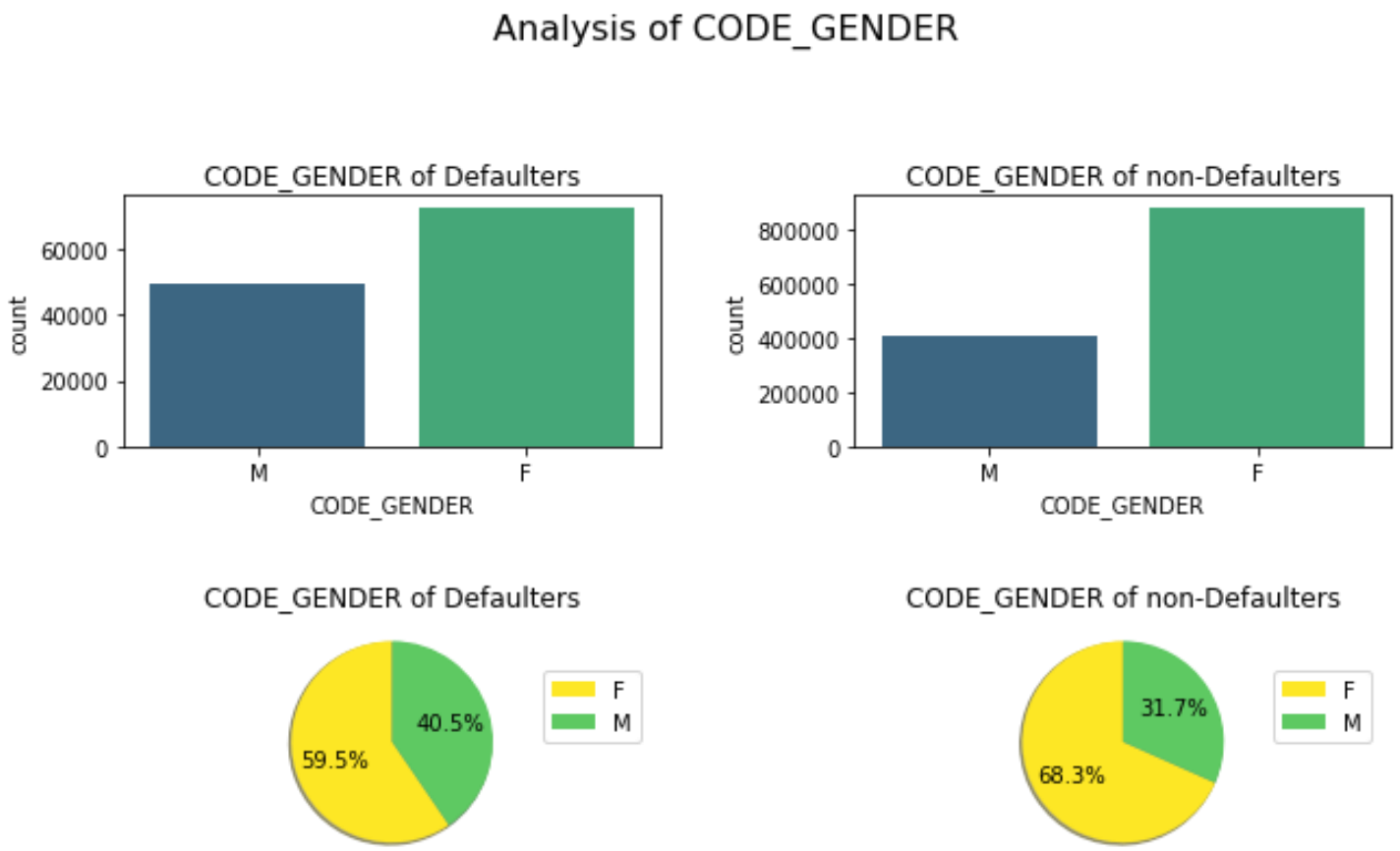


MERGE DATASETS CURRENT APPLICATION AND PREVIOUS APPLICATION INTO DF_MERGE

UNIVARIATE ANALYSIS

- Cash Loans have lesser defaulters when compared to Rev



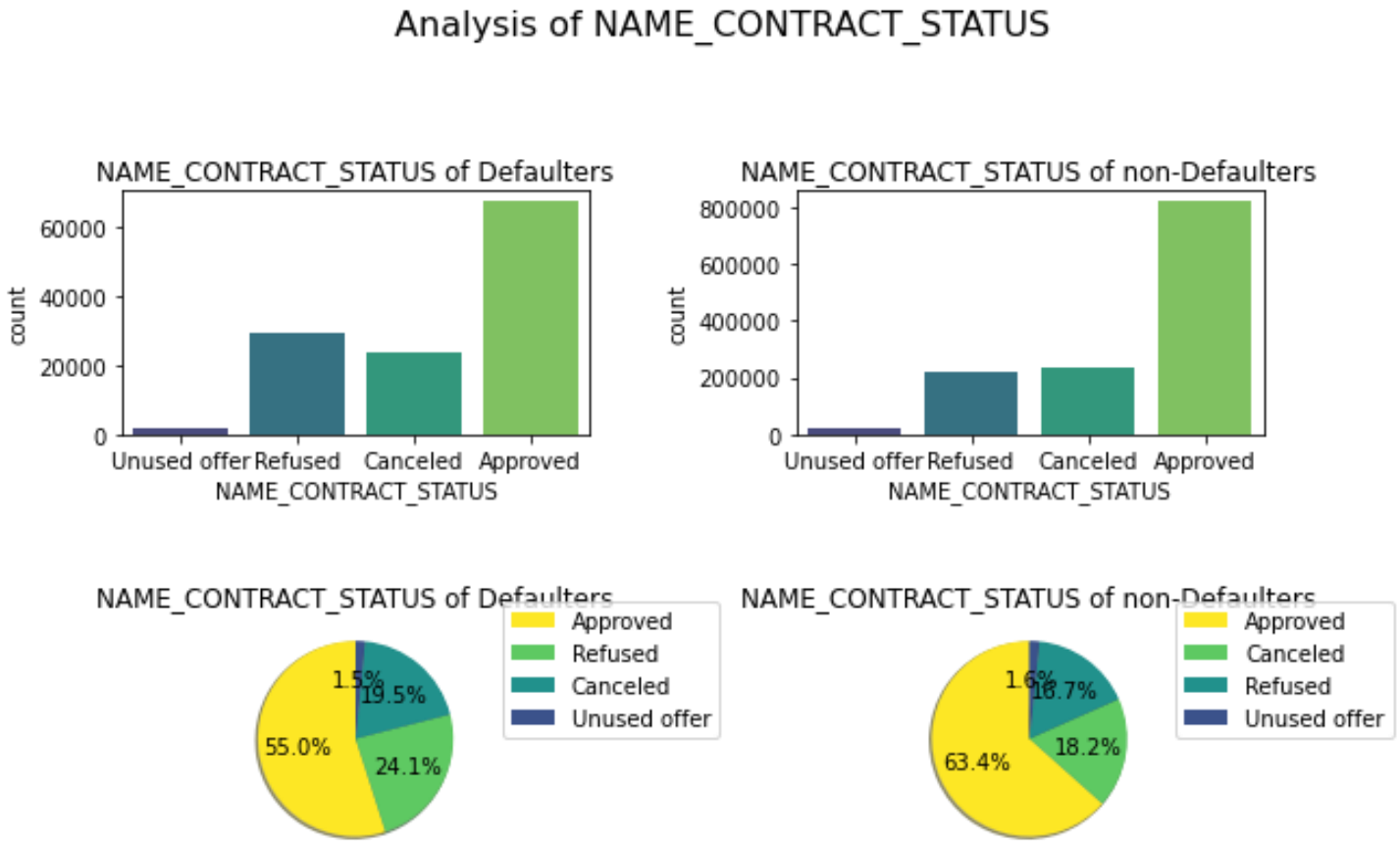


ANALYSIS OF CODE_GENDER

NO MAJOR INFERENCE BASED ON GENDER.

ANALYSIS OF NAME_CONTRACT_STATUS

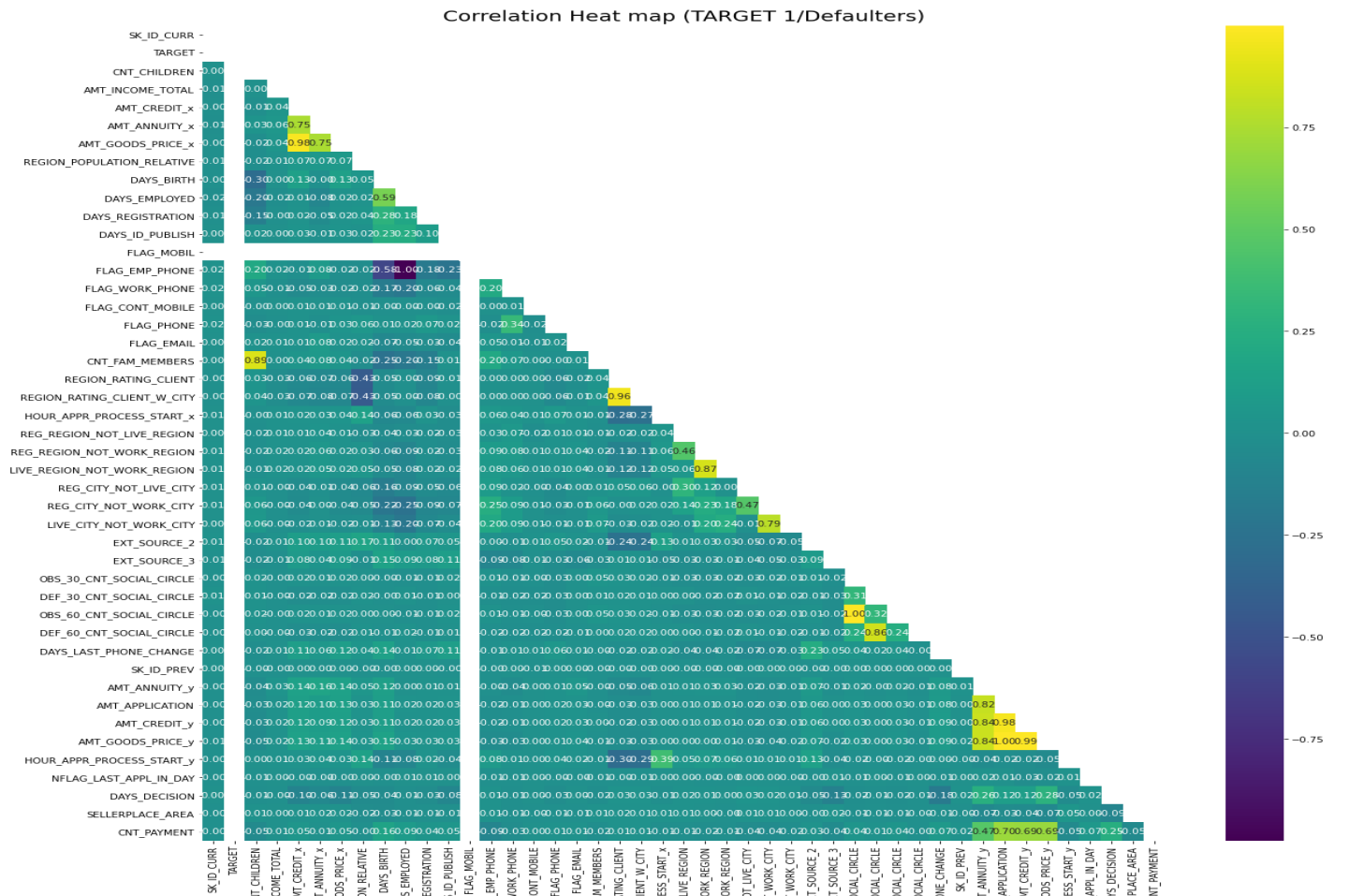
- Refused Applicants have major difficulties in repayment



NUMERICAL CORRELATION OF MERGED DF - DEFAULTERS

REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	95.65
CNT_CHILDREN	CNT_FAM_MEMBERS	88.63
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	87.31
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	85.83
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	79.29
DAYS_EMPLOYED	DAYS_BIRTH	58.71
REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	46.58
REG_REGION_NOT_WORK_REGION	REG_REGION_NOT_LIVE_REGION	45.72
FLAG_WORK_PHONE	FLAG_PHONE	34.14
DEF_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	31.83

dtype: float64



NUMERICAL CORRELATION OF MERGED DF - DEFAULTERS

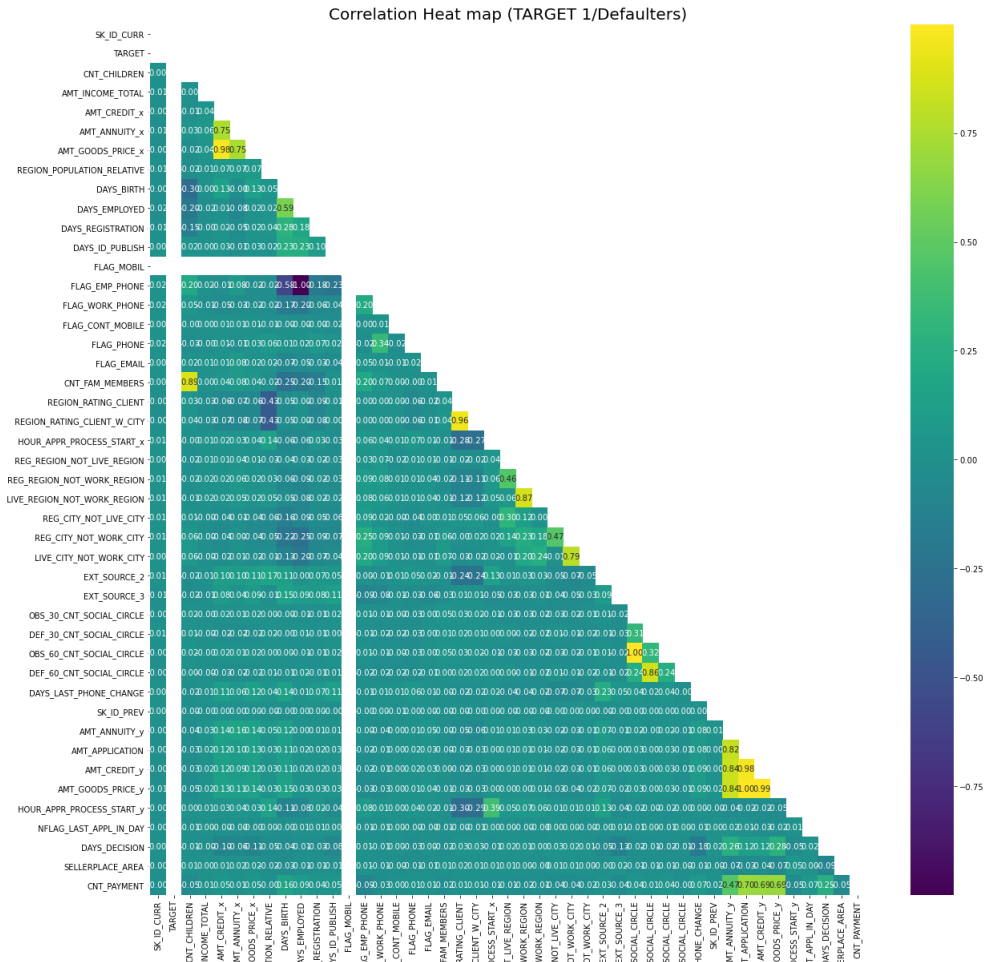
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	94.44
CNT_FAM_MEMBERS	CNT_CHILDREN	87.85
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	87.58
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	86.31
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	83.56
DAYS_EMPLOYED	DAYS_BIRTH	63.50
REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	43.09
REG_REGION_NOT_WORK_REGION	REG_REGION_NOT_LIVE_REGION	42.22
DAYS_REGISTRATION	DAYS_BIRTH	32.80
FLAG_PHONE	FLAG_WORK_PHONE	32.01

dtype: float64

Inference –

- Correlation between REGION_RATING_CLIENT_W_CITY and REGION_RATING_CLIENT in the the highest between Defaulters and non Defaulters.

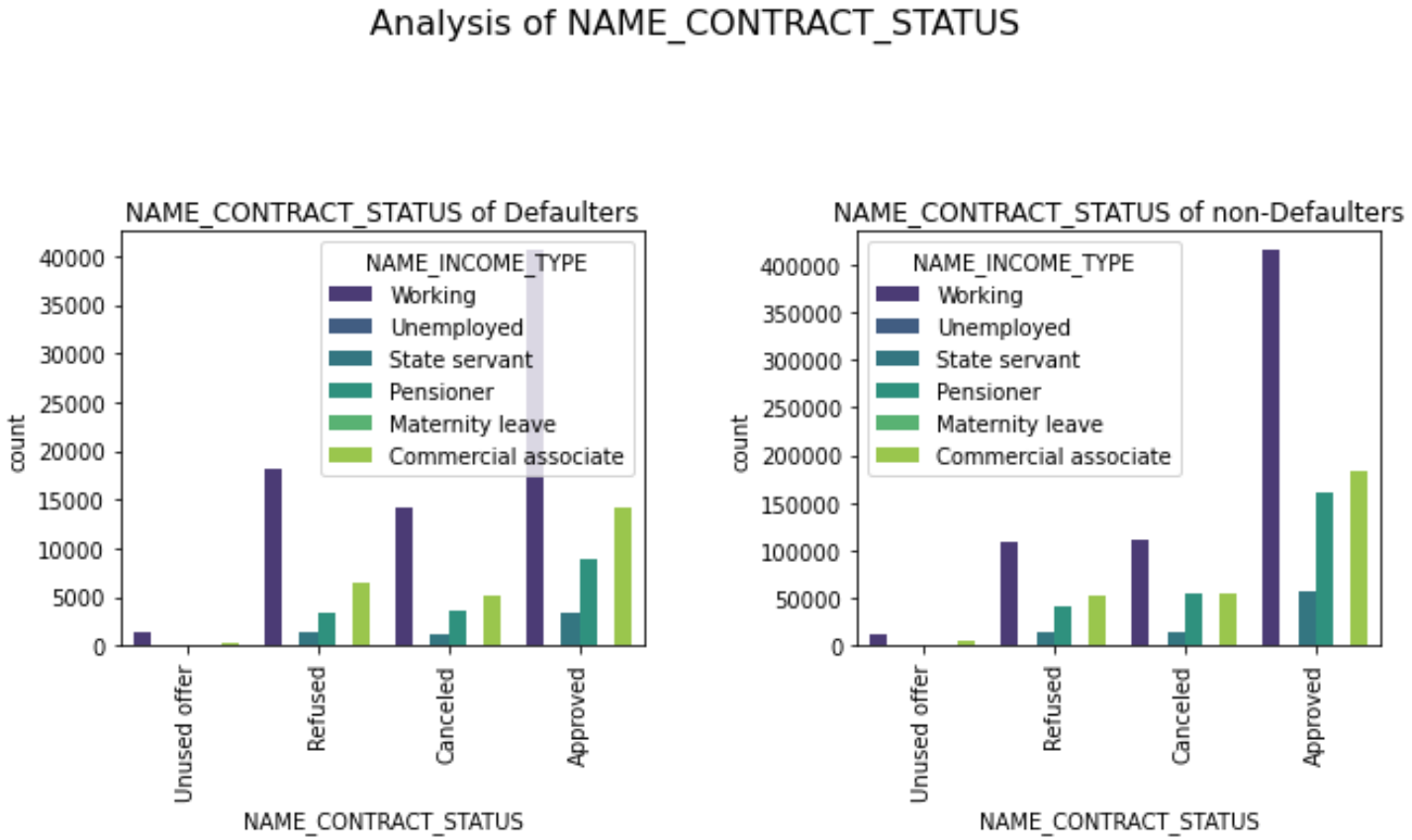
- The correlation between Defaulters and non Defaulters are almost similar.



BIVARIATE ANALYSIS BASED ON NAME_CONTRACT_STATUS

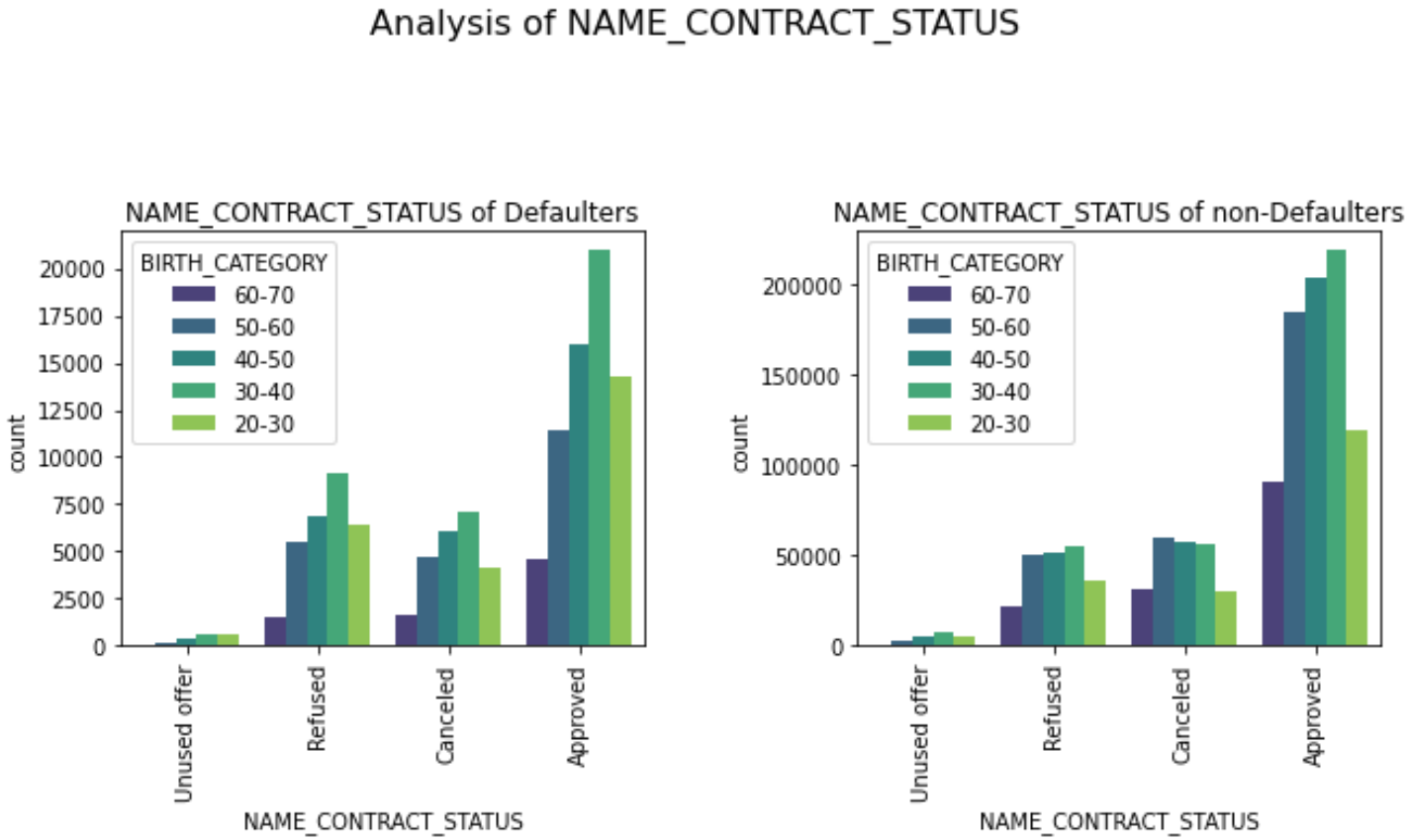
BIVARIATE ANALYSIS BASED ON NAME_CONTRACT_STATUS VS NAME_INCOME_TYPE

- Refused and Cancelled Working Applicants have difficulty in repayment.
- Approved Applicants who are pensioners and Commercial associates have less difficulty in repayment.



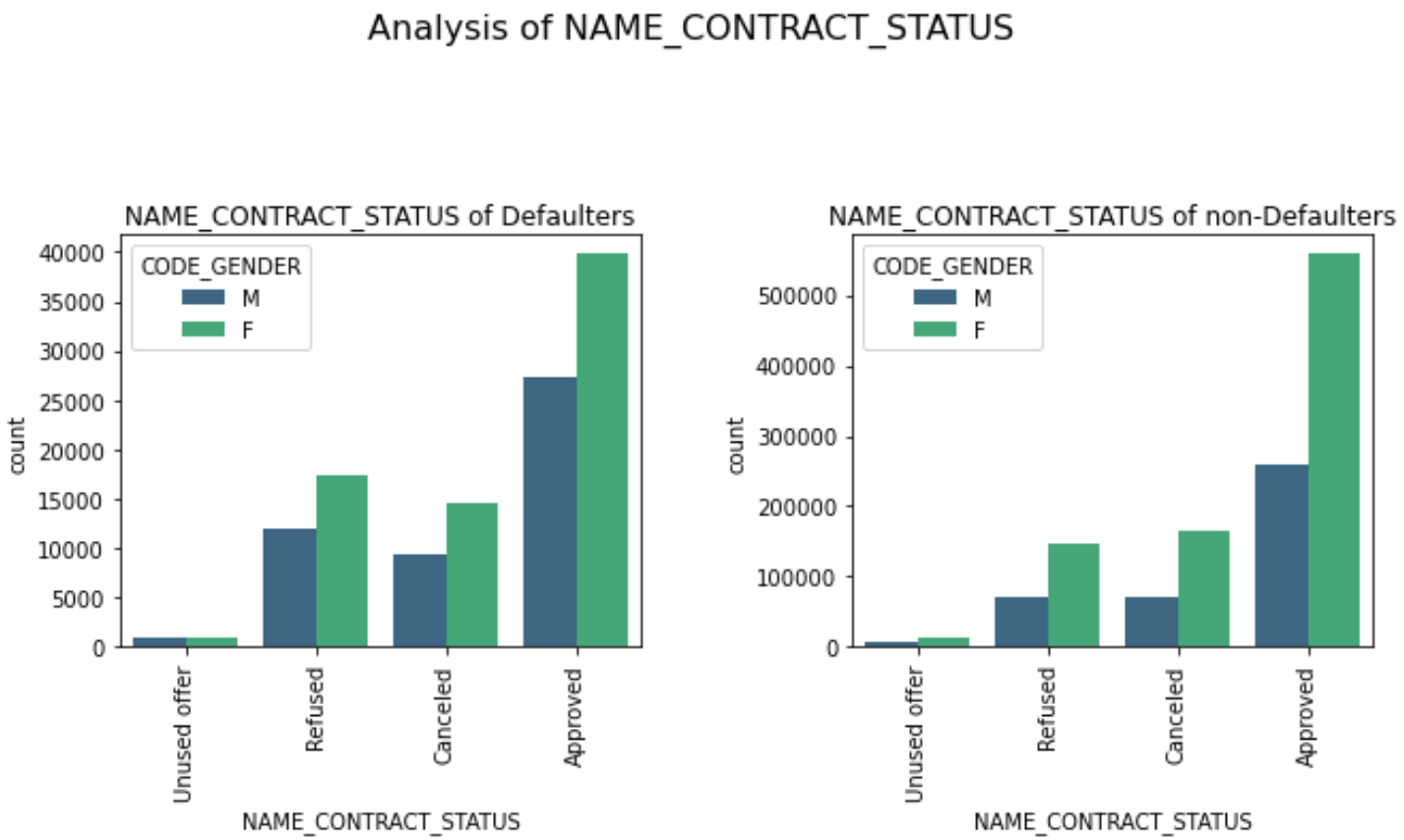
BIVARIATE ANALYSIS BASED ON NAME_CONTRACT_STATUS VS BIRTH_CATEGORY

- Approved Clients above 40 years have less payment difficulties.
- Refused and cancelled clients have more payment difficulties.



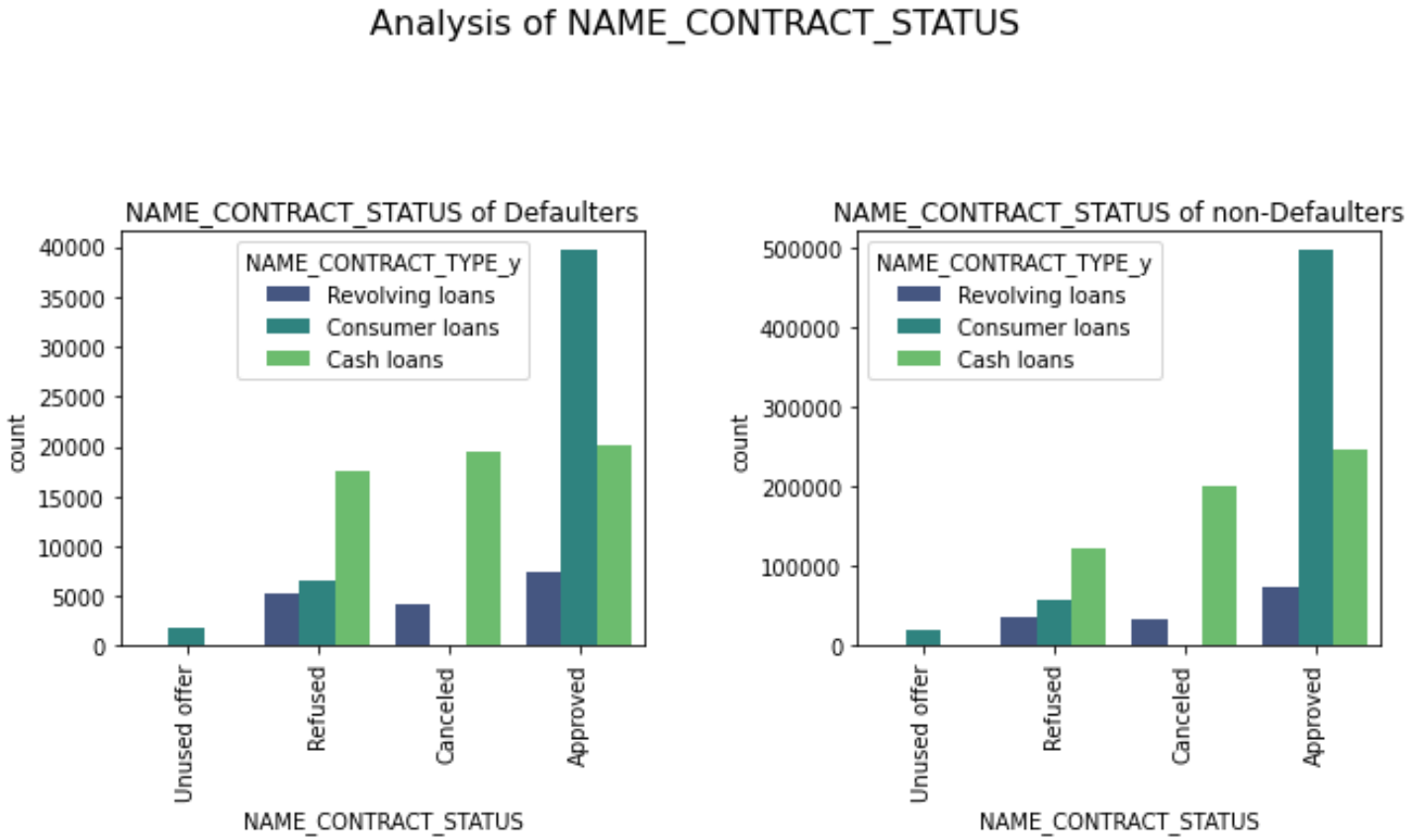
BIVARIATE ANALYSIS BASED ON NAME_CONTRACT_STATUS VS CODE GENDER

- Refused and Cancelled Applicants both male and female have more payment difficulties.
- Approved Male applicants have repayment difficulties.



BIVARIATE ANALYSIS BASED ON NAME_CONTRACT_STATUS VS NAME_CONTRACT_TYPE_ OF PREVIOUS APPLICATION

- Refused applicants with contract type as ‘Revolving loans’ and 'Cash Loan' in previous application are the driving factors for Loan Defaulters.



FINAL RECOMMENDATION BASED ON THE INFERENCES

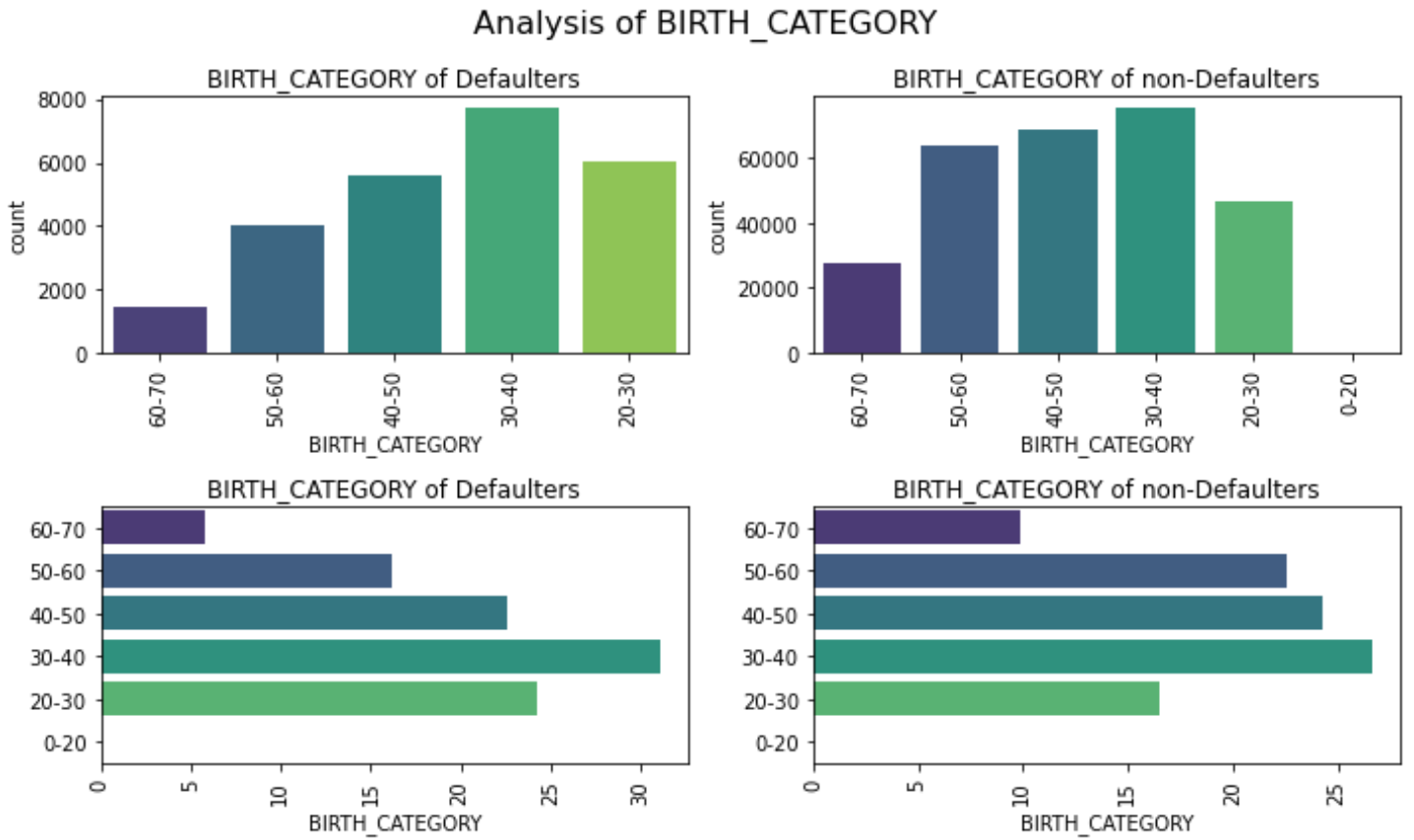
FOR PROVIDING LOAN:

- Applicants in the age range 30-40 and 40-50
- Male Applicants with Academic degree
- Married Applicants
- Applicants who are Students and Businessman
- Applicants who are employed for more than 15 years
- Previously Approved Applicants

REFERENCE FOR THE RESULTS

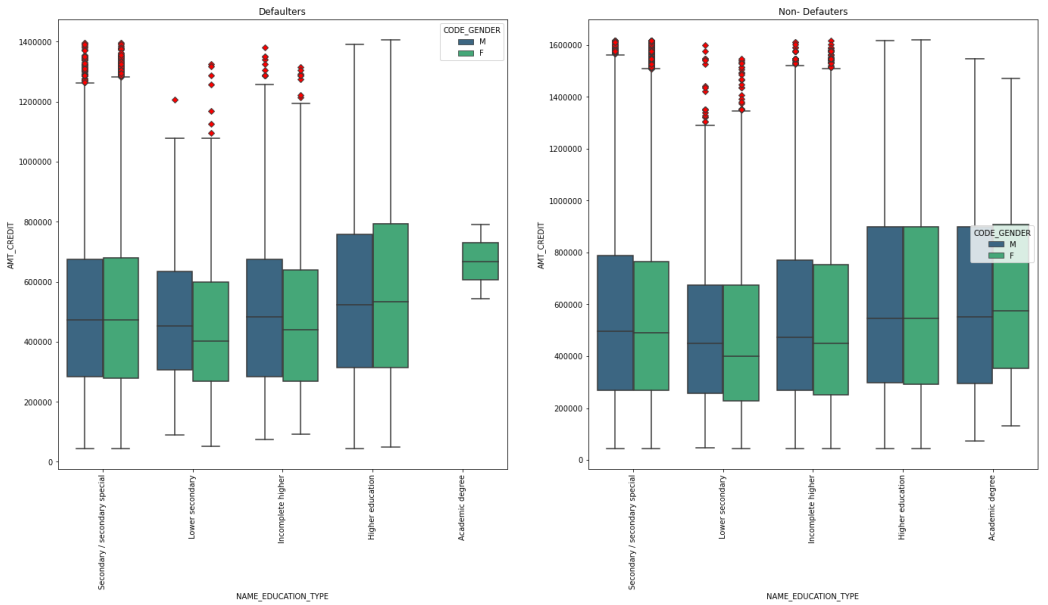
BIRTH_CATEGORY

- Applicants aged more than 30-40 make better ontime payments.
- Applicants in the age range 20-30 have a less impressive repayment trend.



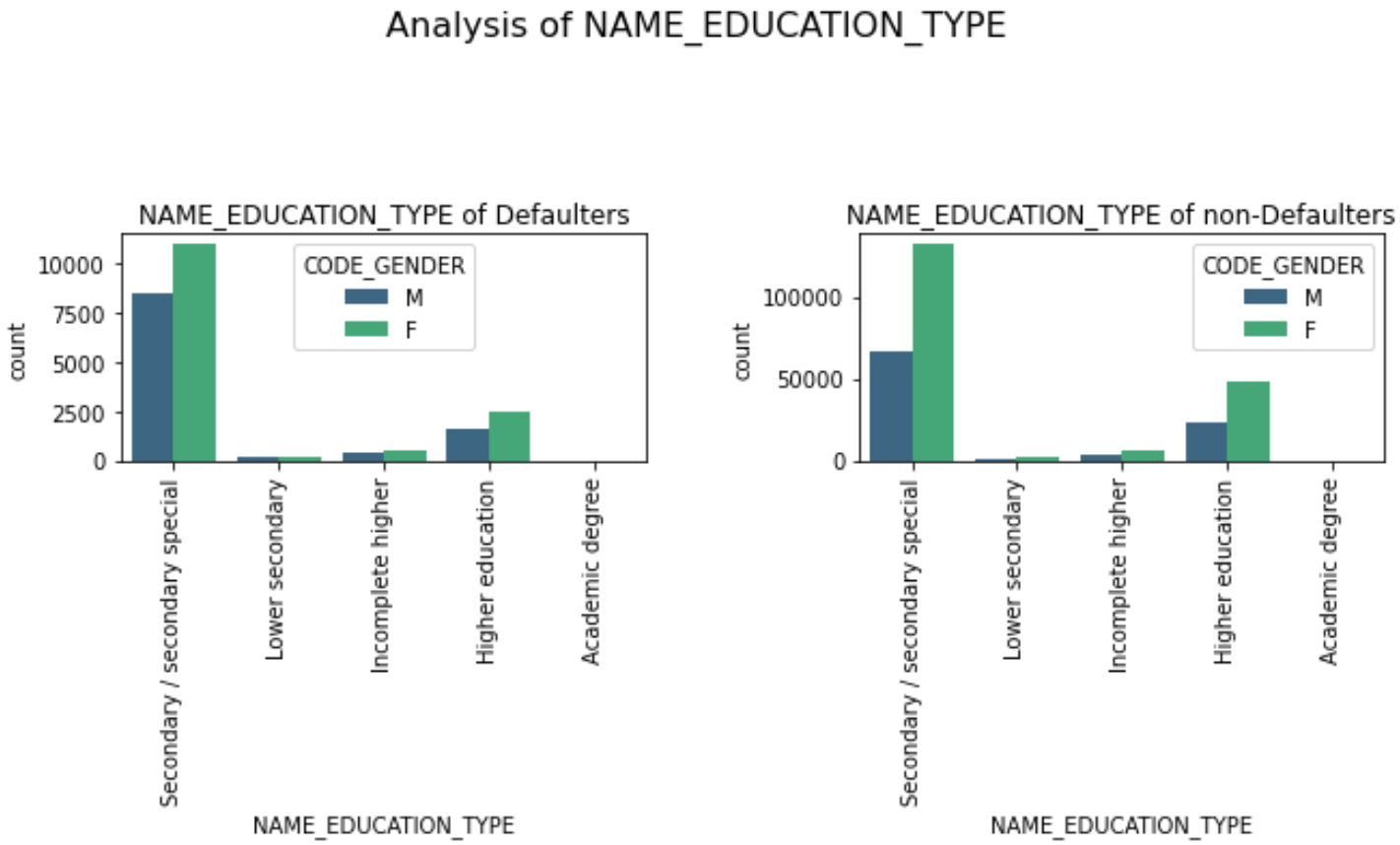
CONTINUOUS V/S CATEGORICAL VARIABLES
ANALYSIS OF NAME_EDUCATION_TYPE V/S AMT_CREDIT V/S CODE_GENDER

- Male with Academic degree make all the payments on time.
- Clients with Academic degree are wide range of Amount credit in non defaulters and very less in defaulters.



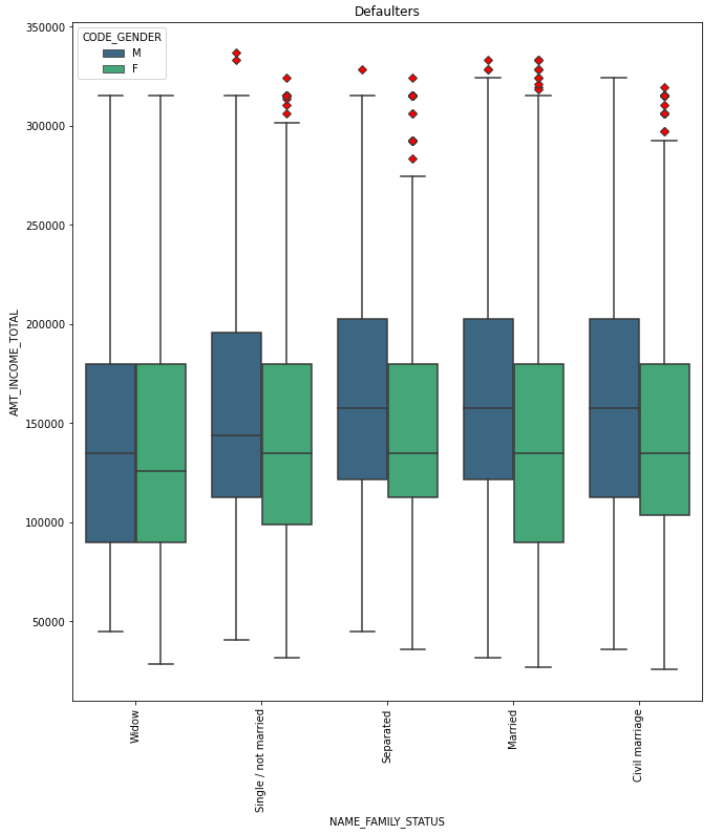
ANALYSIS OF NAME_EDUCATION_TYPE V/S CODE_GENDER

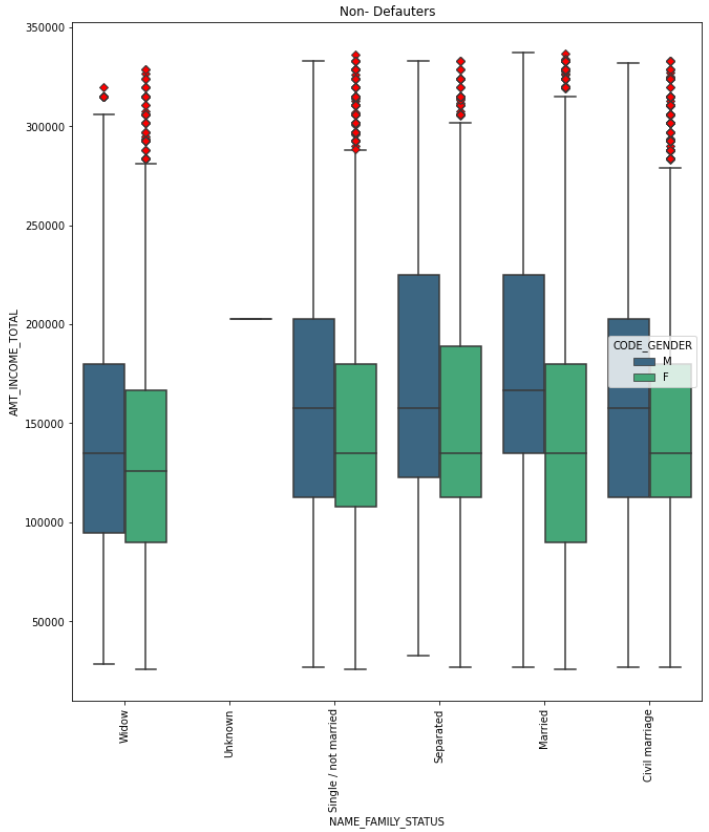
- Applicants having Secondary/Secondary special education and Male have more Payment difficulties compared to On-Time Payments.
- Applicants having Higher education and Female make more On-Time Payments compared to Payment difficulties.



ANALYSIS OF NAME_FAMILY_STATUS V/S AMT_INCOME_TOTAL V/S CODE_GENDER

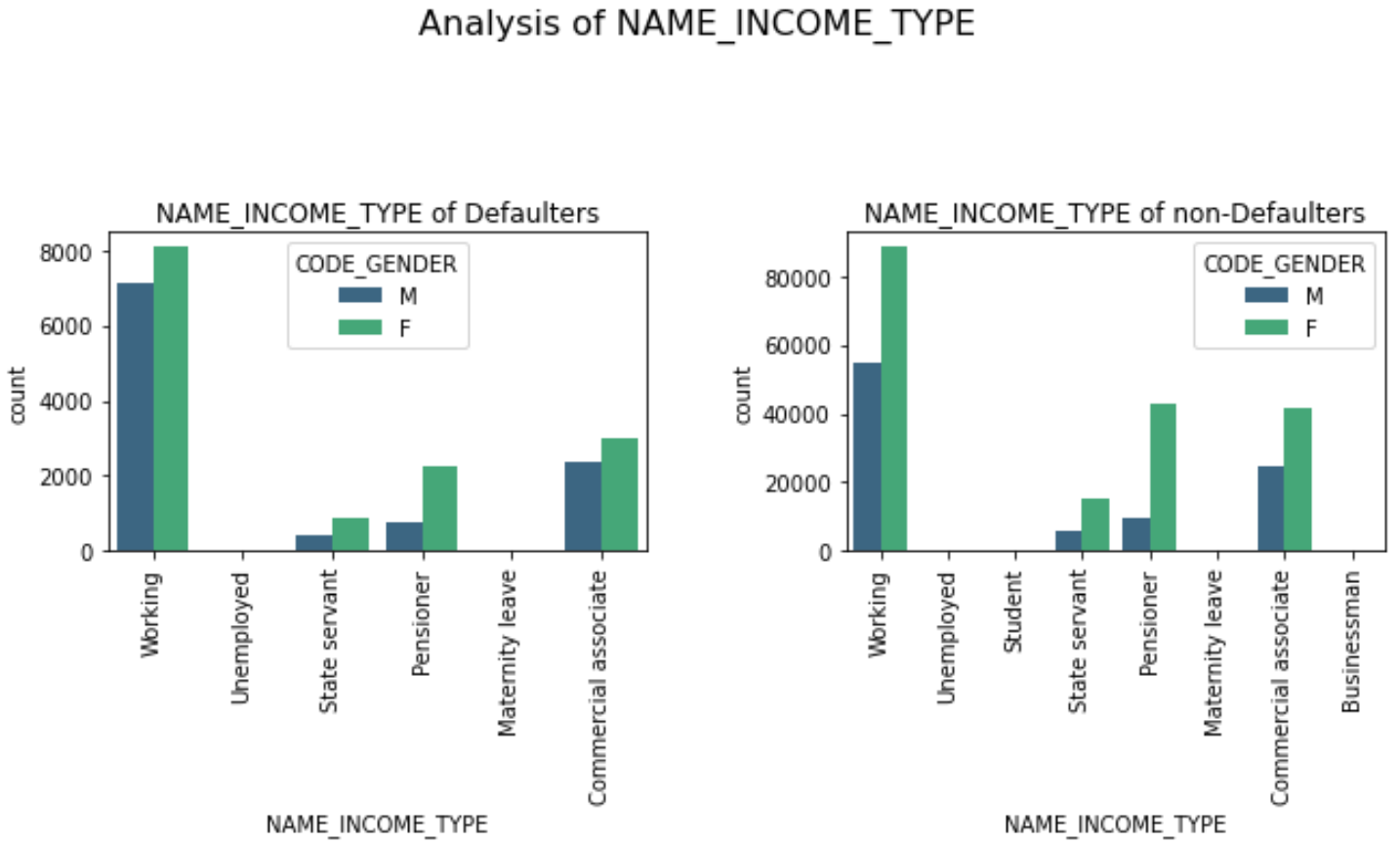
- Married men have high median and better repayment trend.





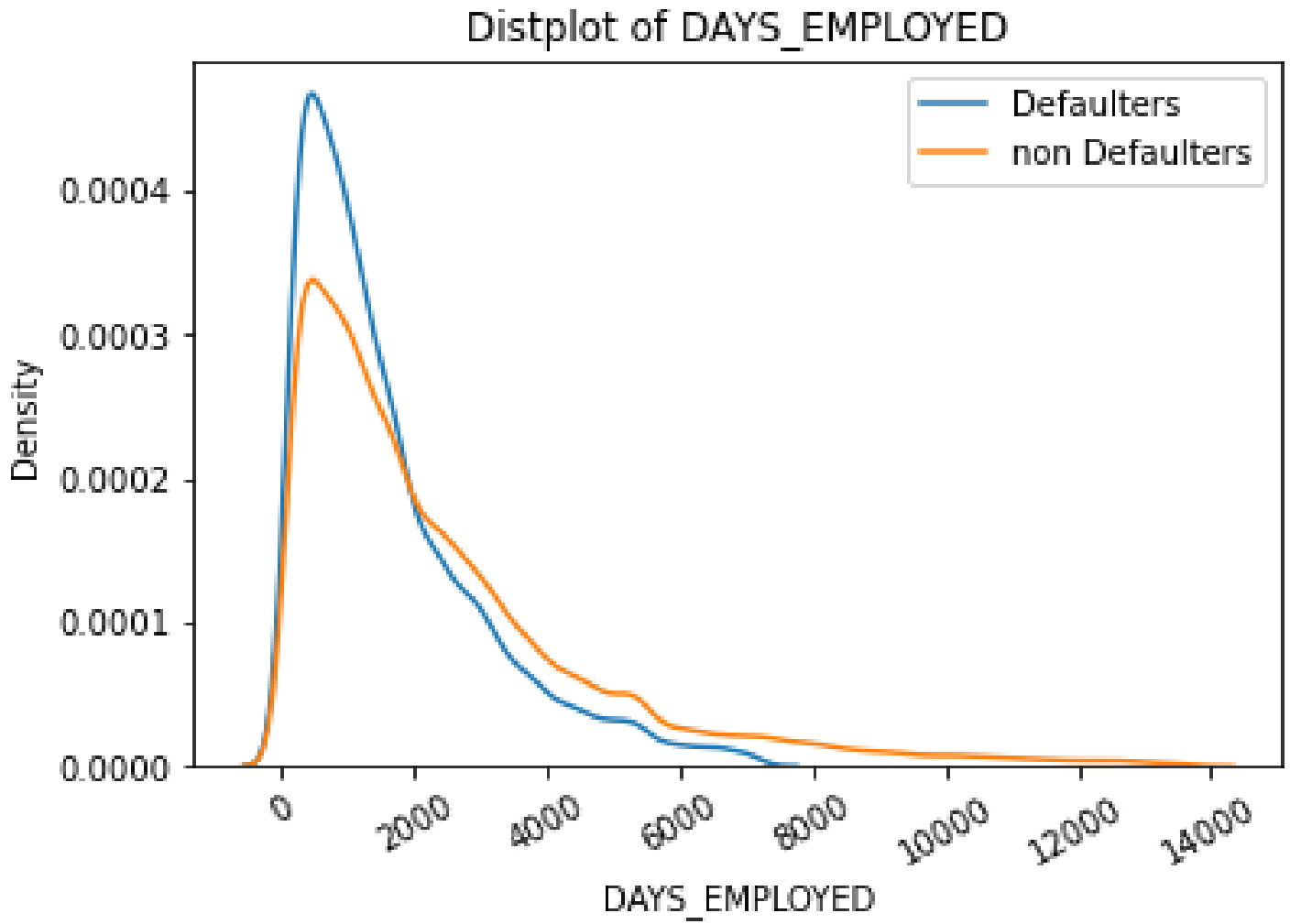
CATEGORICAL V/S CATEGORICAL VARIABLES
ANALYSIS OF NAME_INCOME_TYPE V/S CODE_GENDER

- Businesspeople and Students make on time payments.
- Working Men make lesser on time payments than women.
- Pensioners especially women have payment difficulties.



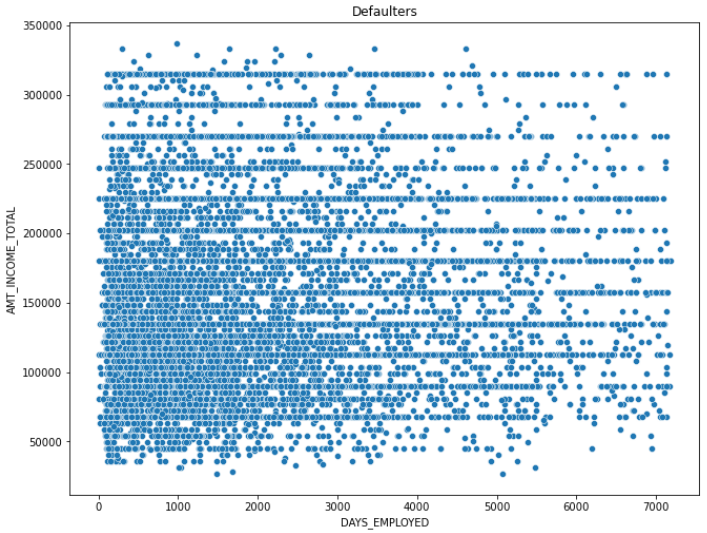
DAYS_EMPLOYED

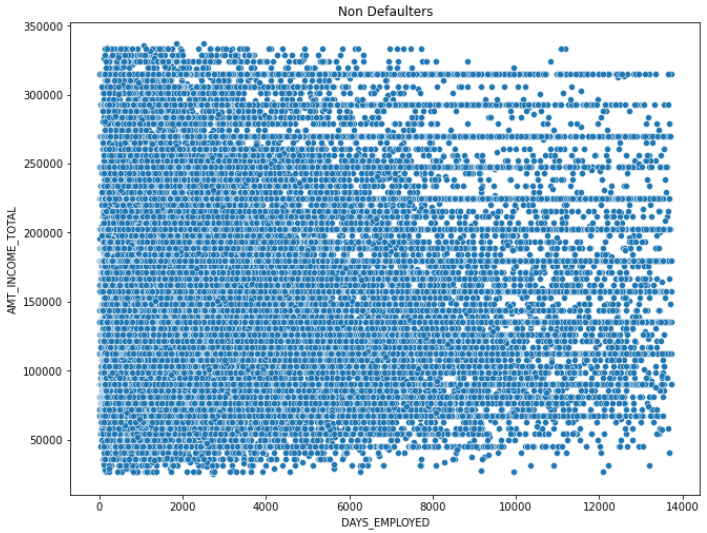
- DAYS_EMPLOYED of Defaulters --> IQR: 2603.0 Min_value: -3227.5 Max_value: 7184.5 DAYS_EMPLOYED of non Defaulters --> IQR: 5107.0 Min_value: -6693.5 Max_value: 13734.5
- For DAYS_EMPLOYED less than 2000, there are more clients with Payment difficulties.
- But, for DAYS_EMPLOYED > 2000 , there are more clients with On-Time Payments.
- This means that those who are employed longer have better chances of repaying the loa



ANALYSIS OF DAYS_EMPLOYED V/S AMT_INCOME_TOTAL

- There is no Correlation between DAYS_EMPLOYED and AMT_INCOME_TOTAL
- However, we can see that as days of Employment increases above 5000 the defaulters reduce.





BIVARIATE ANALYSIS BASED ON NAME_CONTRACT_STATUS VS NAME_CONTRACT_TYPE_ OF PREVIOUS APPLICATION

- Refused applicants with contract type as ‘Revolving loans’ and 'Cash Loan' in previous application are the driving factors for Loan Defaulters.