BANK LOAN CASE STUDY

Ankita Yadav

AGENDA

Project Description

Approach

Tech-stack Used

Insights

Results

PROJECT DESCRIPTION

In this project, we have performed Exploratory Data Analysis (EDA) on Bank Loan Dataset.

The aims of this project is to give you an idea of applying EDA techniques in a real business scenario. It develops the basic understanding of the RISK ANALYSIS in banking and financial sectors.

The loan providing companies find it hard to give loans to people due to their insufficient or non-existent credit history. Because of that, some consumers use it to their advantage by becoming defaulters.

PRIMARY GOALS

Identification of applicants who wont be able to pay loans using EDA techniques

APPROACH

1. Using MS Excel to study and understand dataset and especially Columns dataset.
2. Using Jupyter Notebook, studying datasets and performing EDA on the dataset.
3. Cleaning dataset, handling missing values and errors, finding outliers, data imbalance, performing variance analysis and finding correlations using various statistical functions.
4. Presenting various plots and graphs and withdrawing insights.

TECH-STACK USED

1. MS EXCEL 2019
2. JUPYTER NOTEBOOK
3. MS POWERPOINT 2019

UNDERSTANDING DATASET

application_data.csv

- It contains all the information of the client at the time of application. The data is about whether a client has payment difficulties.

column_description.csv

- It is data dictionary which describes the meaning of the variables.

previous_application.csv

- It 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.

ANALYSIS OF application_data.csv

LOADING DATASET

```
In [3]: df_app = pd.read_csv('application_data.csv')
        df_app.head()
```

Out[3]:

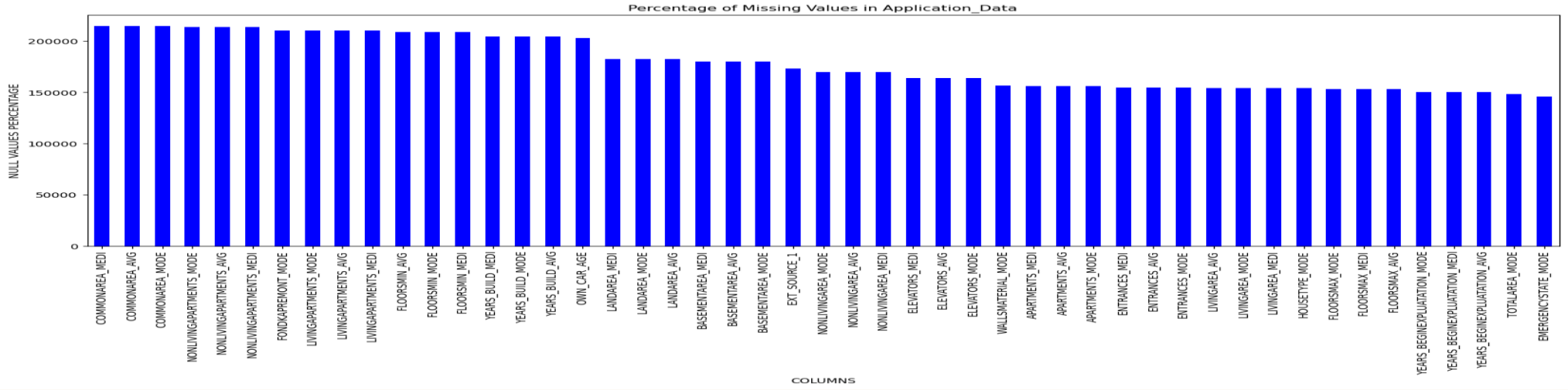
	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREI
0	100002	1	Cash loans	M	N	Y	0	202500.0	40659
1	100003	0	Cash loans	F	N	N	0	270000.0	129350
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	13500
3	100006	0	Cash loans	F	N	Y	0	135000.0	31268
4	100007	0	Cash loans	M	N	Y	0	121500.0	51300

5 rows × 122 columns

Dataset was loaded and all the columns were read for further data inspections (122 columns were detected).

DATA INSPECTION & CLEANING

- Reading the total no. of rows and columns- shape().
- Checking datatypes- info().
- Summarizing numeric columns- describe().
- Calculating the Null Value Percentage(NVP).
- The columns given in picture below having NVP > 35% would be removed and others would be imputed- mode() and median().



ERROR HANDLING

- Converting Negative Days Column to Positive days columns- abs() and unique().
- Replacing Flag Columns from 'Y' and 'N' to 1 and 0.
- Replacing XNA values to 'F' in Gender Column.

```
[ 9461 16765 19046 ...  7951  7857 25061]
[  637  1188   225 ... 12971 11084  8694]
[ 3648.  1186.  4260. ... 16396. 14558. 14798.]
[2120  291 2531 ... 6194 5854 6211]
[1134.  828.  815. ... 3988. 3899. 3538.]
```

	FLAG_OWN_CAR	FLAG_OWN_REALTY
0	0	1
1	0	0
2	1	1
3	0	1
4	0	1

```
Out[19]: 0    M
         1    F
         2    M
         3    F
         4    M
         Name: CODE_GENDER, dtype: object
```

BINNING DATA

- The following list of columns were binned into groups for better readability and analysis:-

1. AMT_INCOME_TOTAL to Income Group
2. AMT_CREDIT to Credit Group
3. DAYS_BIRTH to Age Group

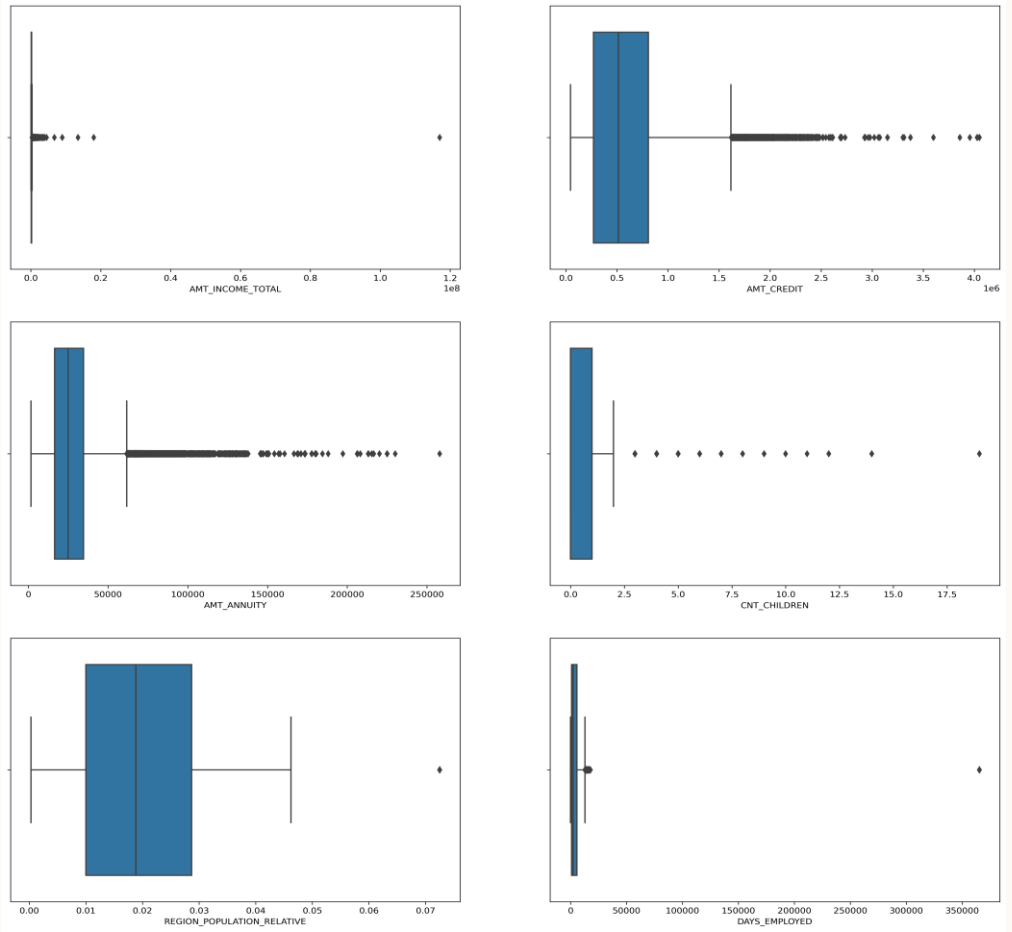
- Unwanted Columns were removed.

```
#Listing the Unwanted Columns

unwanted=['FLAG_MOBIL', 'FLAG_EMP_PHONE', 'FLAG_WORK_PHONE', 'FLAG_CONT_MOBILE',
       'FLAG_PHONE', 'FLAG_EMAIL','REGION_RATING_CLIENT','REGION_RATING_CLIENT_W_CITY','FLAG_EMAIL', 'REGION_RATING_CLIENT',
       'REGION_RATING_CLIENT_W_CITY', 'FLAG_DOCUMENT_2', 'FLAG_DOCUMENT_3','FLAG_DOCUMENT_4', 'FLAG_DOCUMENT_5', 'FLAG_DOCUMENT_6
       'FLAG_DOCUMENT_7', 'FLAG_DOCUMENT_8', 'FLAG_DOCUMENT_9','FLAG_DOCUMENT_10', 'FLAG_DOCUMENT_11', 'FLAG_DOCUMENT_12',
       'FLAG_DOCUMENT_13', 'FLAG_DOCUMENT_14', 'FLAG_DOCUMENT_15','FLAG_DOCUMENT_16', 'FLAG_DOCUMENT_17', 'FLAG_DOCUMENT_18',
       'FLAG_DOCUMENT_19', 'FLAG_DOCUMENT_20', 'FLAG_DOCUMENT_21']
```

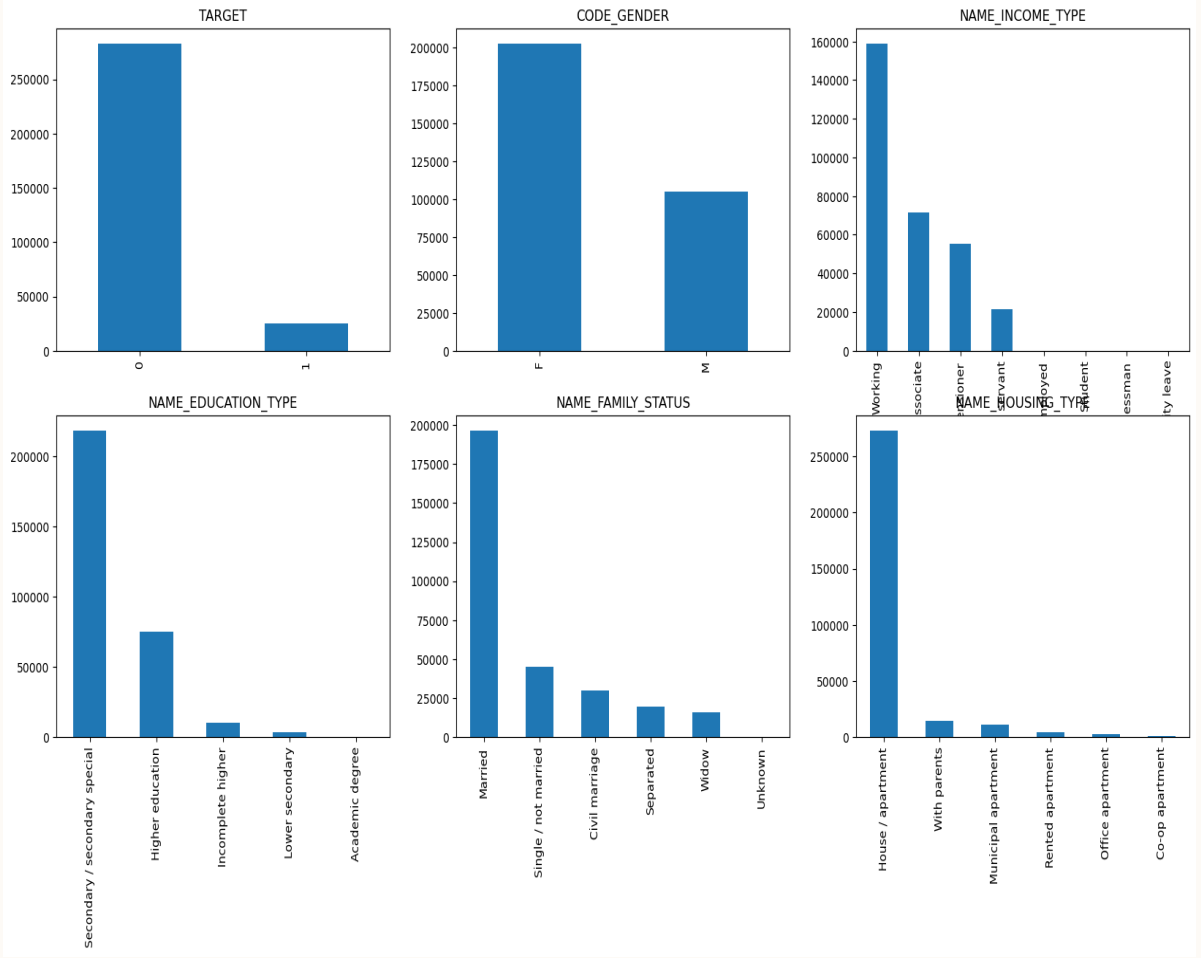
OUTLIERS ANALYSIS

1. IQR for AMT_INCOME_TOTAL is very slim and it have a large number of outliers.
2. Third quartile of AMT_CREDIT is larger as compared to First quartile which means that most of the Credit amount of the loan of customers are present in the third quartile. And there are large number of outliers present in AMT_CREDIT.
3. Third quartile of AMT_ANNUITY is slightly larger than First quartile and there are large number of outliers.
4. CNT_CHILDREN have outlier values having children more than 5.
5. IQR for DAYS EMPLOYED is very slim. Most of the outliers are present below 25000.And a outlier is present 375000.



DATA IMBALANCE

- Data imbalance was calculated for various columns.
- From the graphs we found that TARGET column have the highest imbalance.
- On calculating we found the Data Imbalance Ratio to be **11.39 : 1**
- TARGET 0 – Re-payers (Non- Defaulters).
- TARHET 1 – Defaulters.



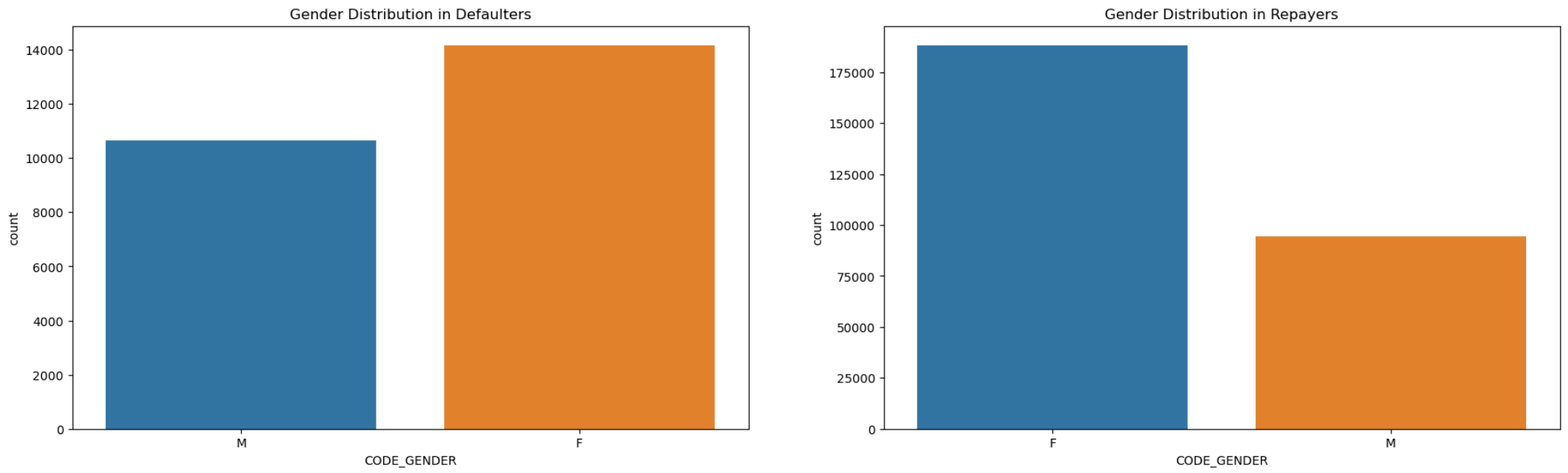
UNIVARIATE ANALYSIS

- Univariate analysis explores each variable in a data set, separately. It looks at the range of values, as well as the central tendency of the values. It describes the pattern of response to the variable. It describes each variable on its own. Descriptive statistics describe and summarize data.
- Analysis was performed based on Categorical and Continuous Data
- Count Plots were generated for Categorical Data.
- Distribution Plots were generated for Continuous Data.

INSIGHTS (CATEGORICAL DATA)

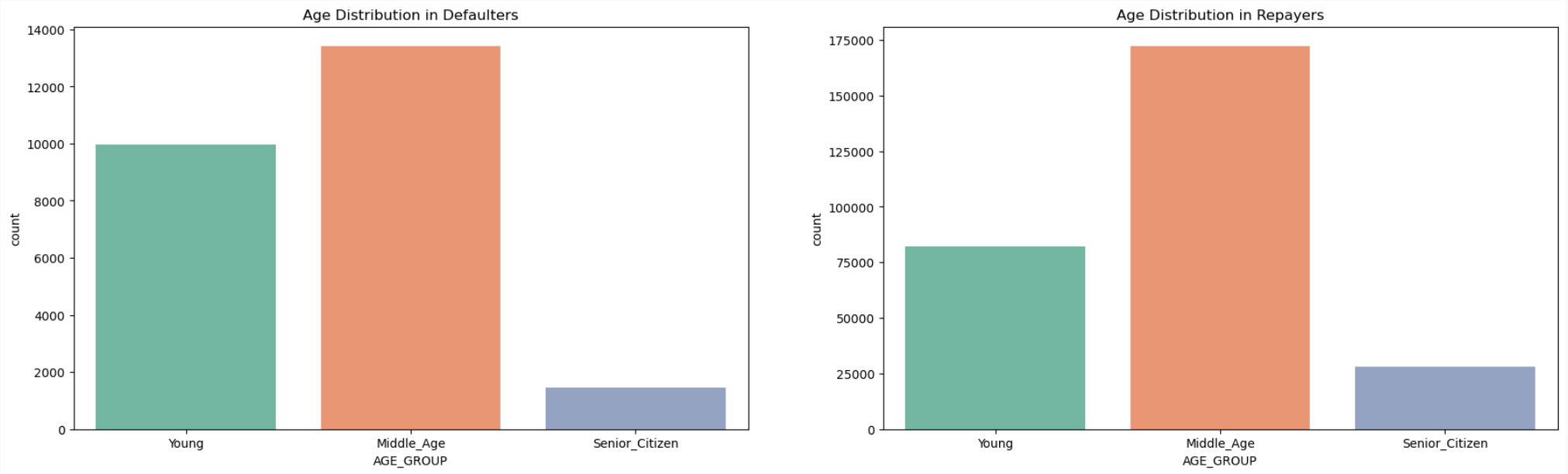
- Based on Gender

1. Defaulters - We can see that females are slightly more in number of defaulters than male.
2. Non-defaulters - The same pattern continues for non-defaulters as well. The females are more in number here than male.



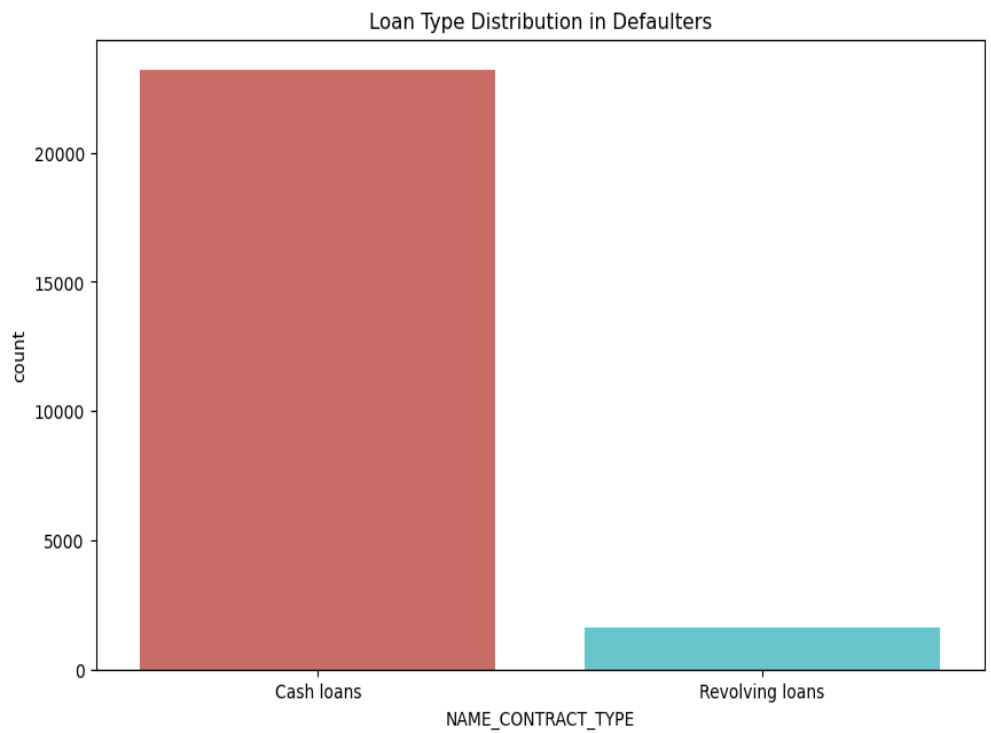
- Based on Age

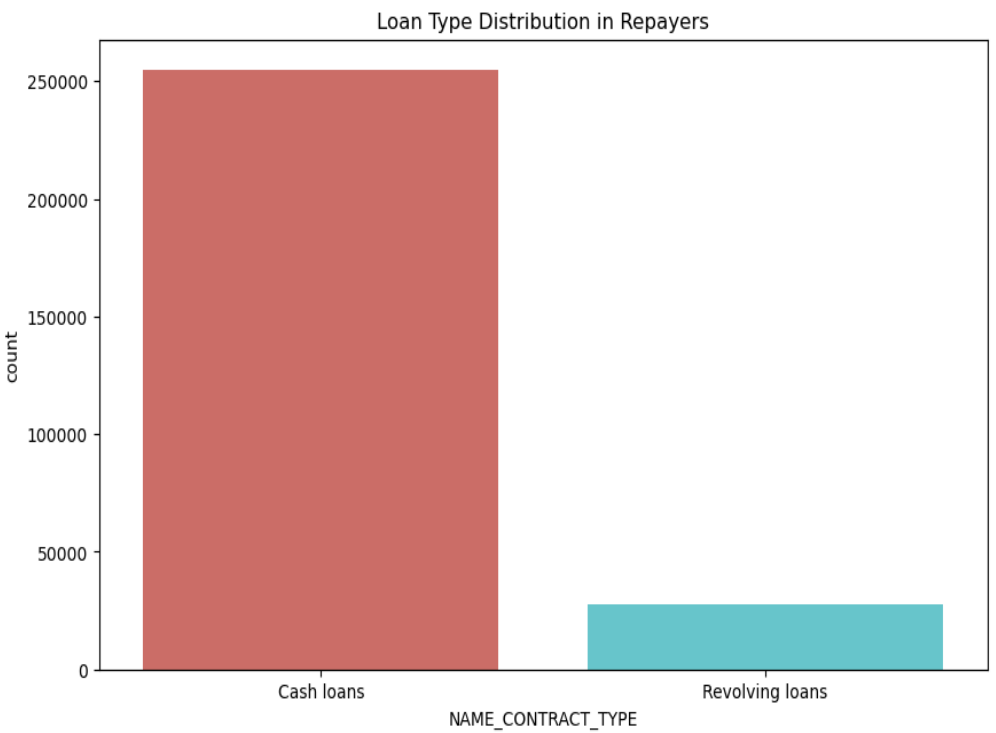
1. Defaulters - We can see that Middle_Age group are most in number of defaulters while Senior_citizens are least.
2. Re-payers - The same pattern continues for non-defaulters as well.



- Based on Loan Type

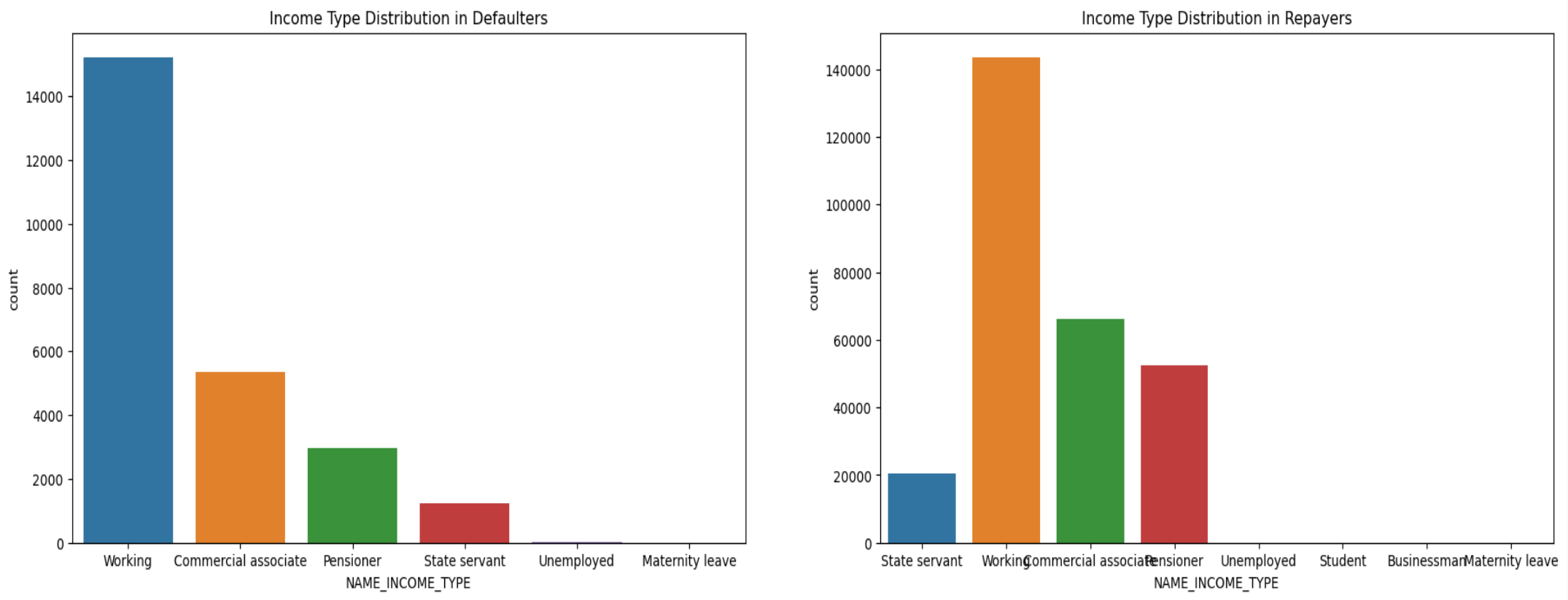
We see in both the cases that Revolving loans are very less in number compared to Cash loans.





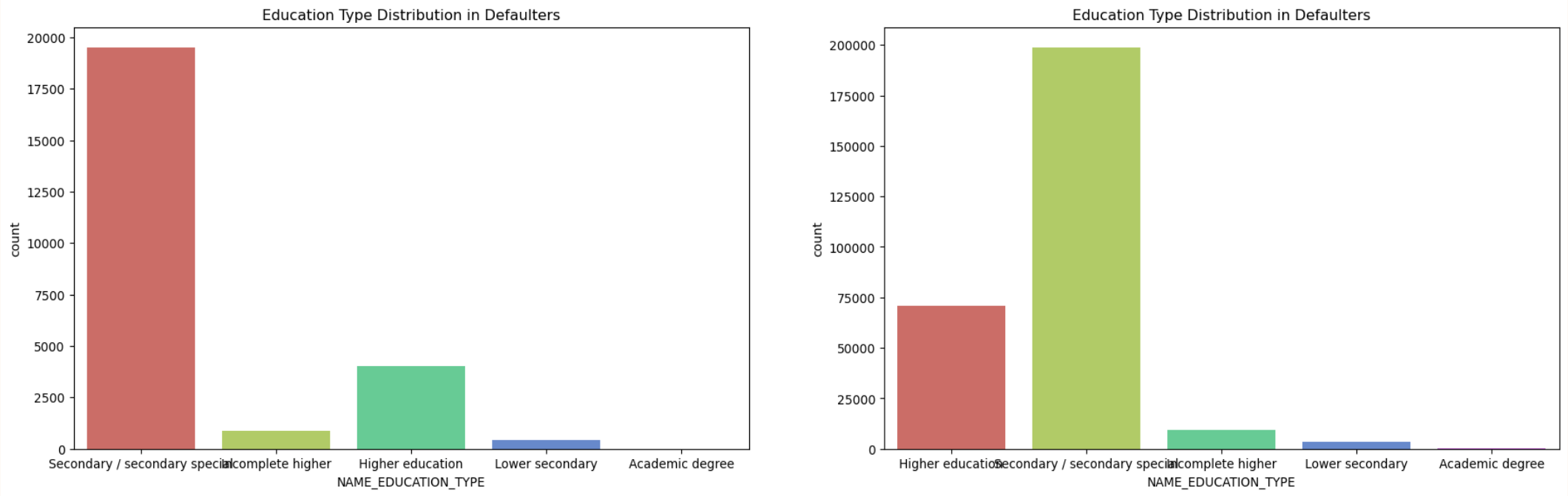
- Based on Income Type

1. Defaulters - Working people are mostly defaulted as their numbers are high with compare to other pfrofessions.
2. Repayers - Similarly here also working people are more in number who are not defaulted.



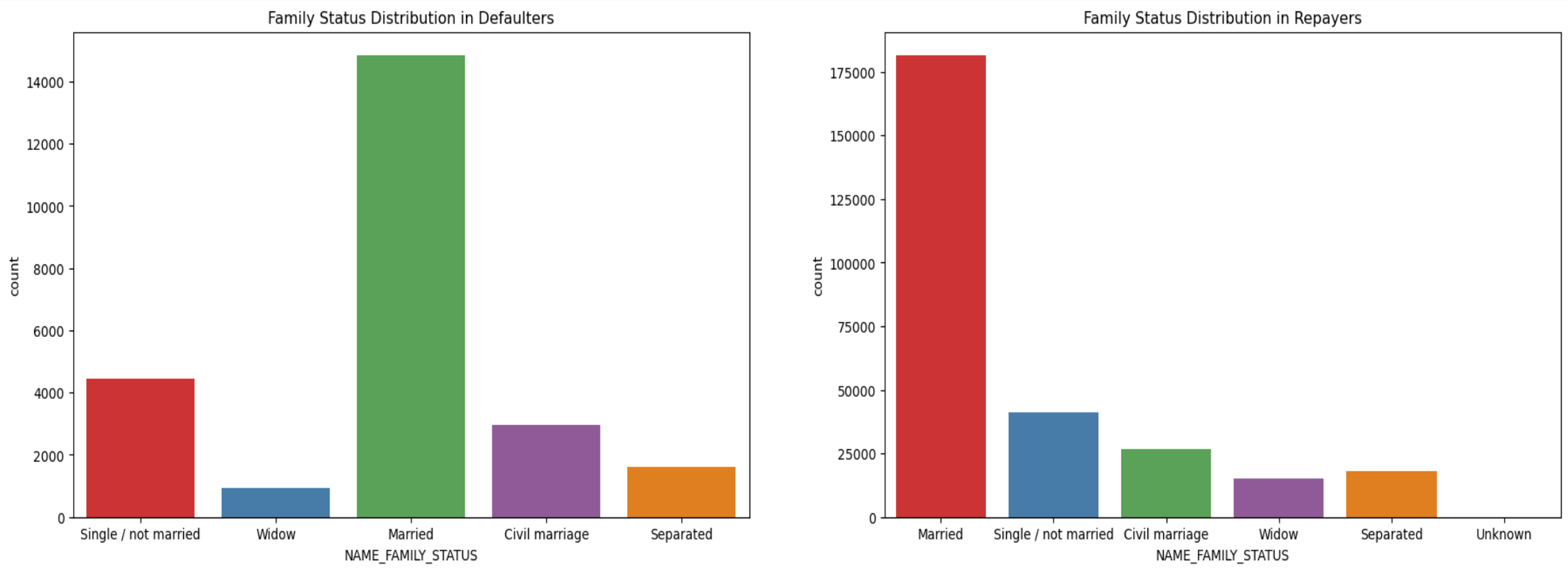
- Based on Education Type

1. Defaulters - Education with Secondary/Secondary special customers are more number in defaulters compare with other level of educated people.
2. Re-payers - Same pattern follows.



- Based on Family Status

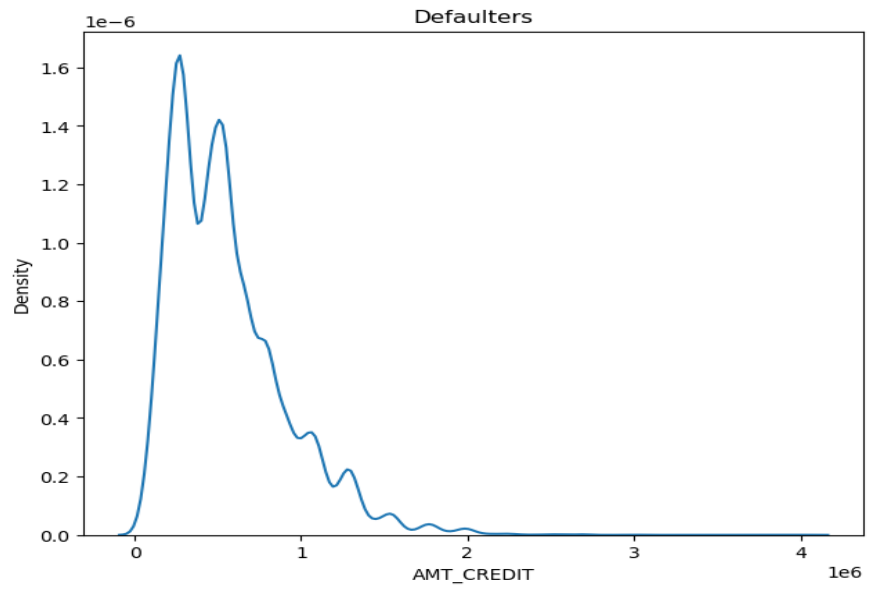
1. Defaulter- Married people are mostly defaulters.
2. Re-payers- Same pattern follows.

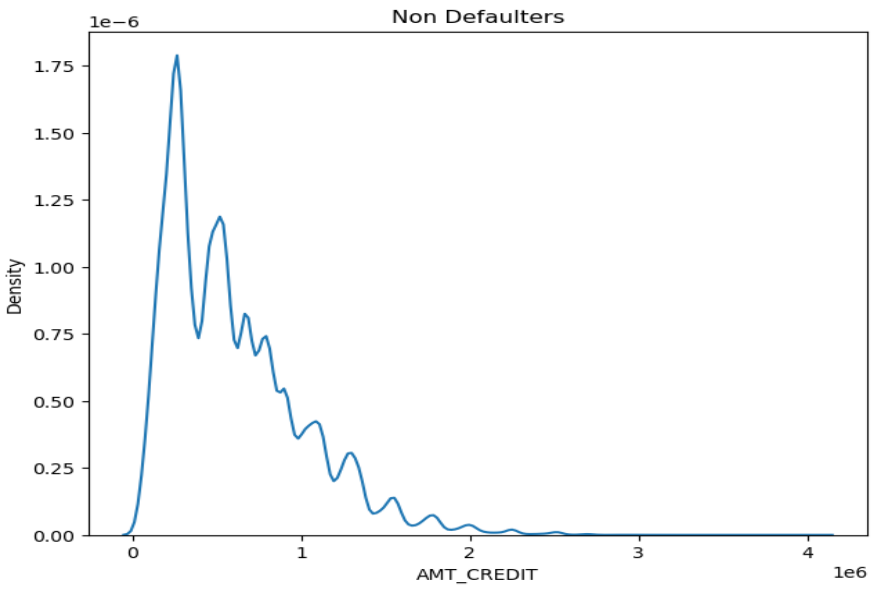


INSIGHTS(CONTINUOUS DATA)

- Based on Credit Amount

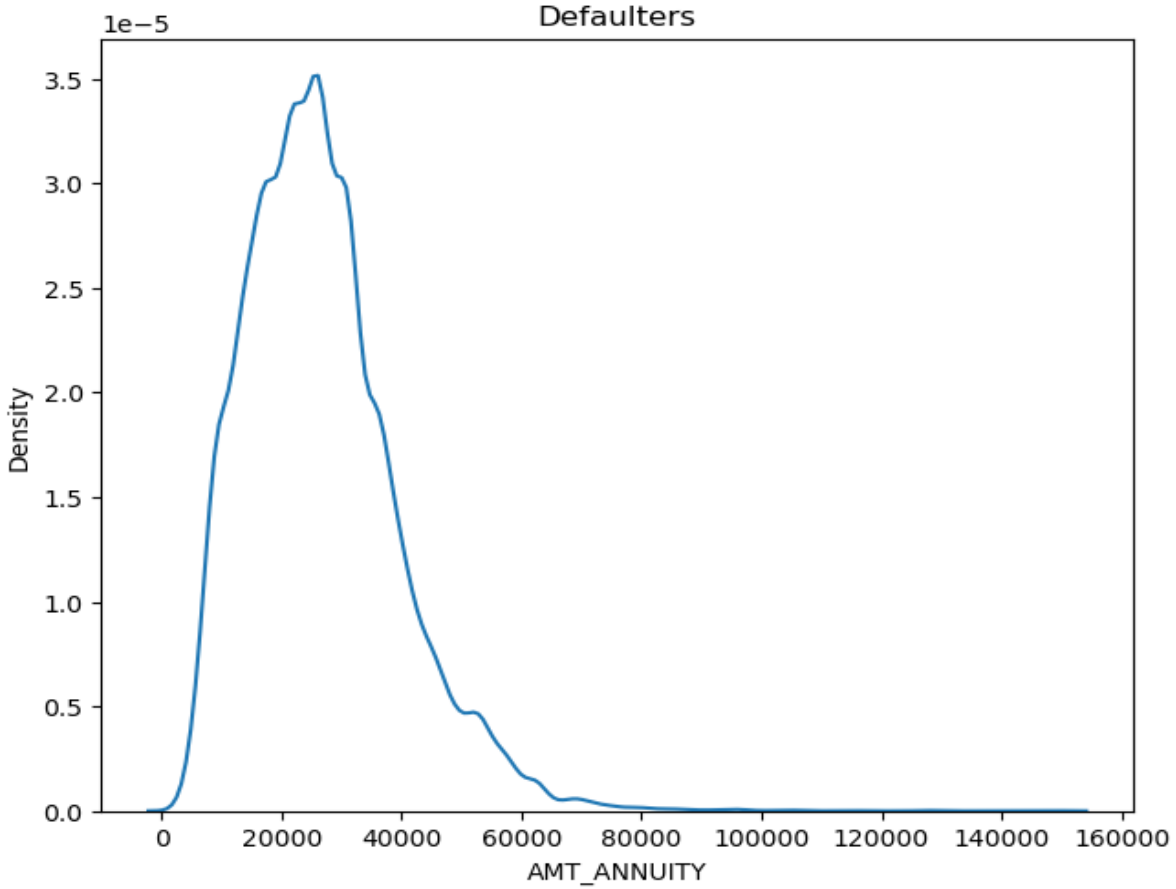
1. Defaulters - We can notice that the lesser the credit amount of the loan, the more chances of being defaulter. The spike is till 500000.
2. Non defaulters - If the credit amount is less, there is lesser chance of being defaulted. And gradually the chance is being decreased with the loan credit amount.

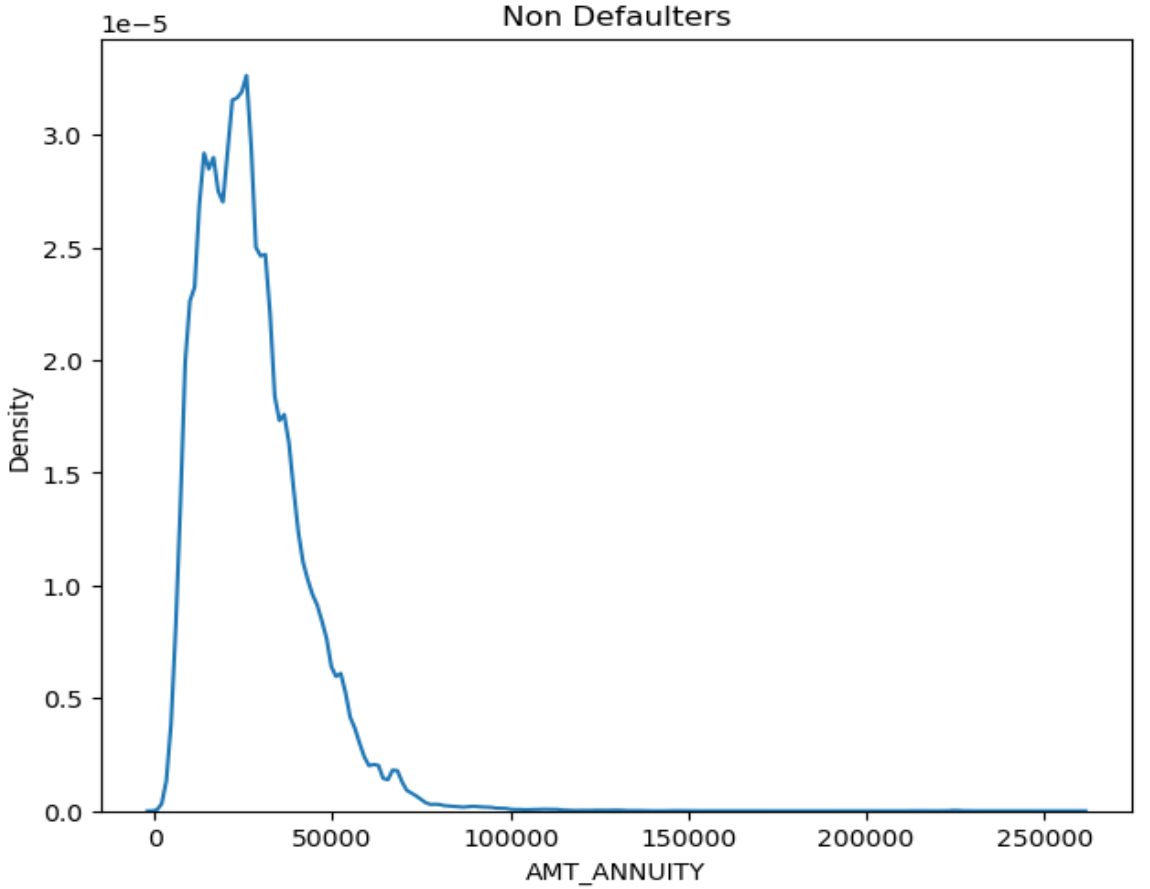




- Based on Amount Annuity

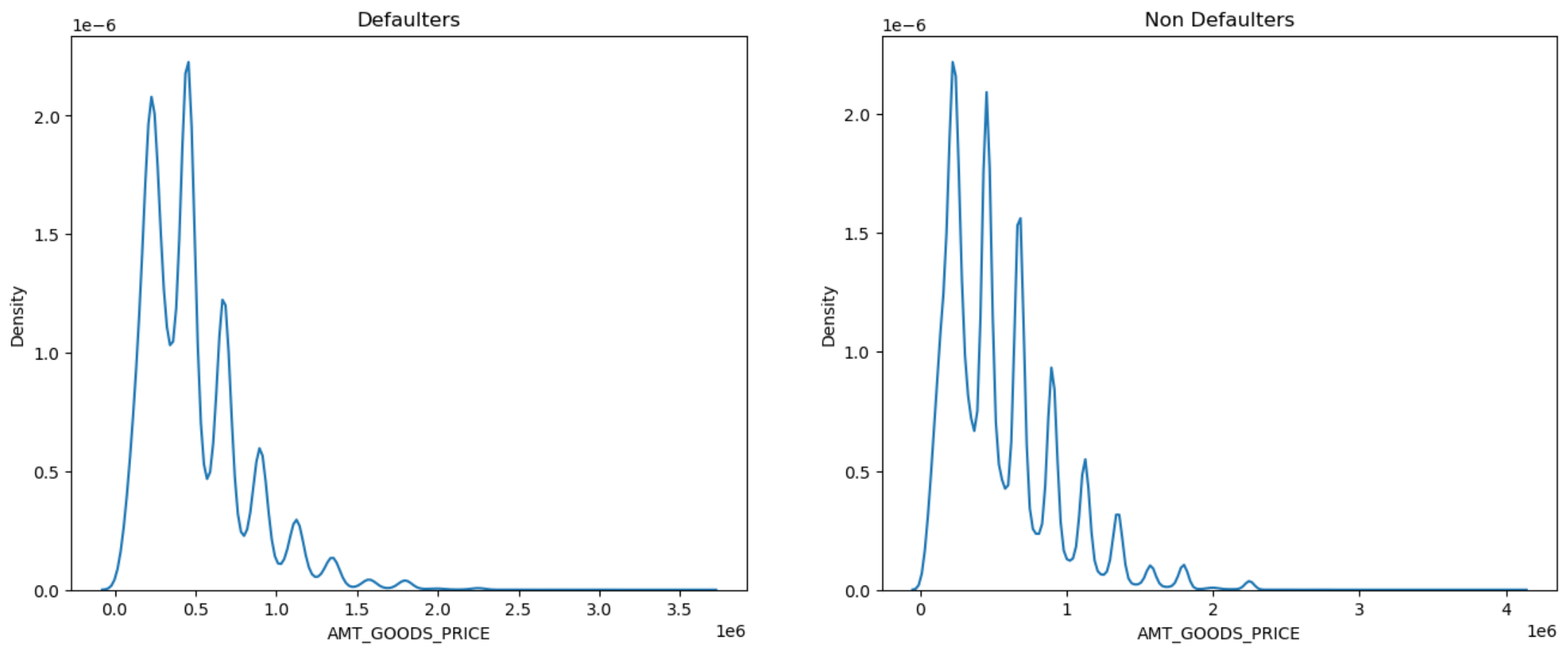
In both the cases the loan annuity is concentrated more from 10000 to 40000.





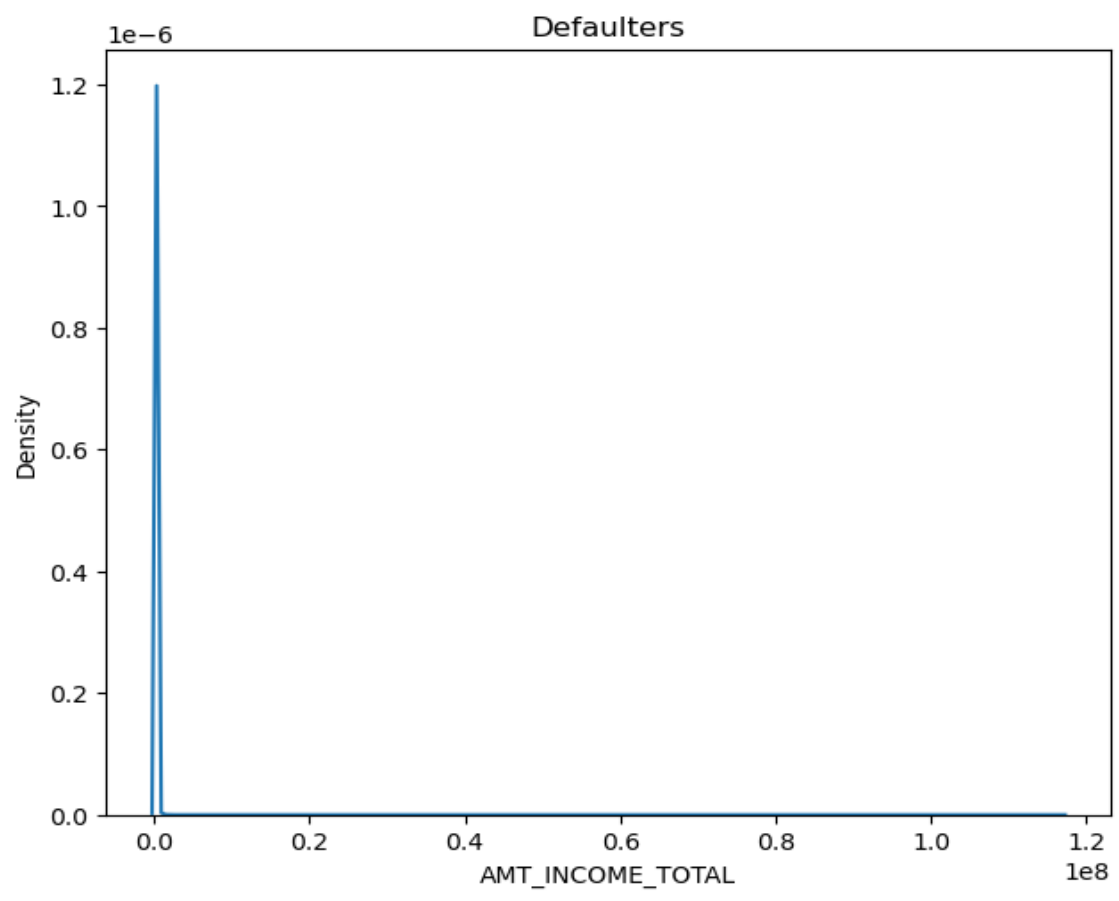
- Based on Goods Price Amount

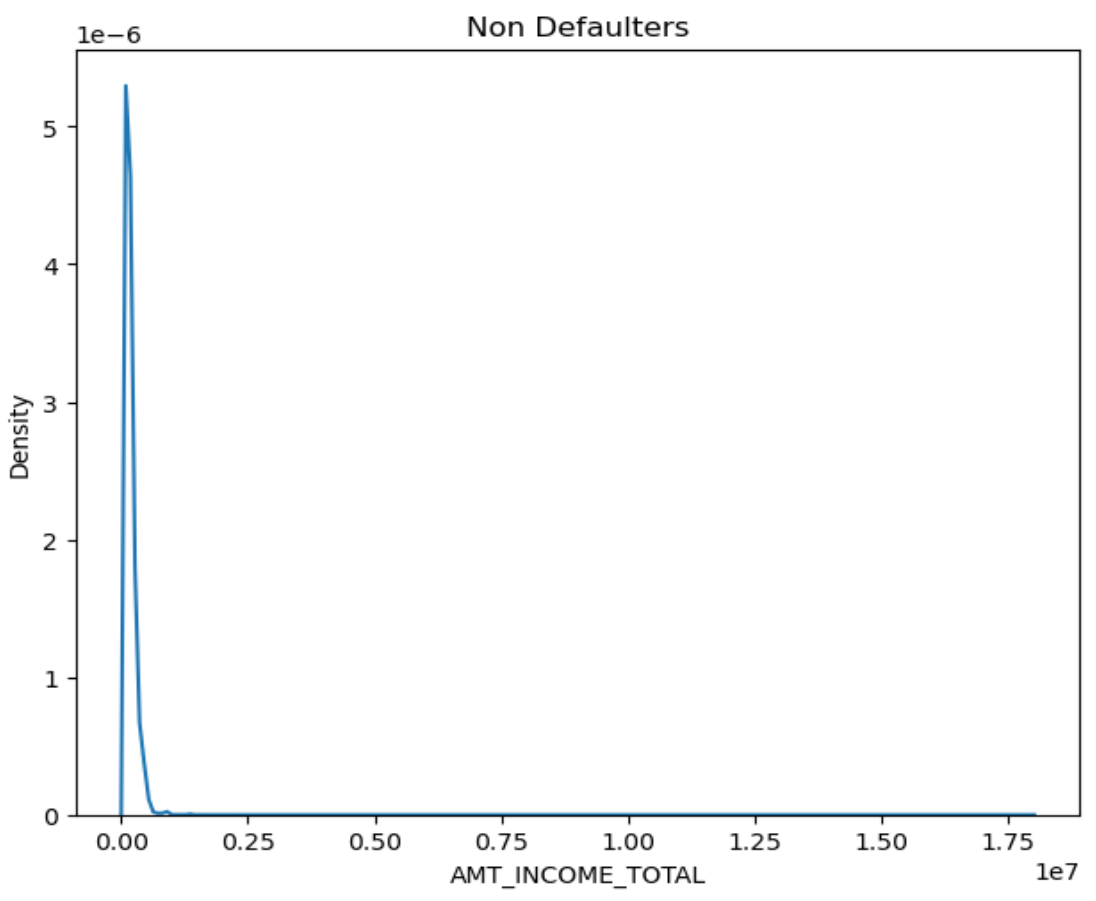
Both Defaulters and Non Defaulters shows same pattern.



- Based on Total Income

1. Defaulters -The pattern that for being a defaulter are almost equal in all income levels.
2. Non defaulters - Same Pattern is observed





BIVARIATE ANALYSIS

- Bivariate analysis is stated to be an analysis of any concurrent relation between two variables or attributes. This study explores the relationship of two variables as well as the depth of this relationship to figure out if there are any discrepancies between two variables and any causes of this difference.
- Analysis was performed based on Numerical, Categorical and Continuous data.
- Correlation matrix and plots were used for the analysis of Numerical Data.
- Scatter Plots were used for analysis of Categorical Data.
- Box plots were used for Continuous Data.

INSIGHTS (NUMERICAL DATA)

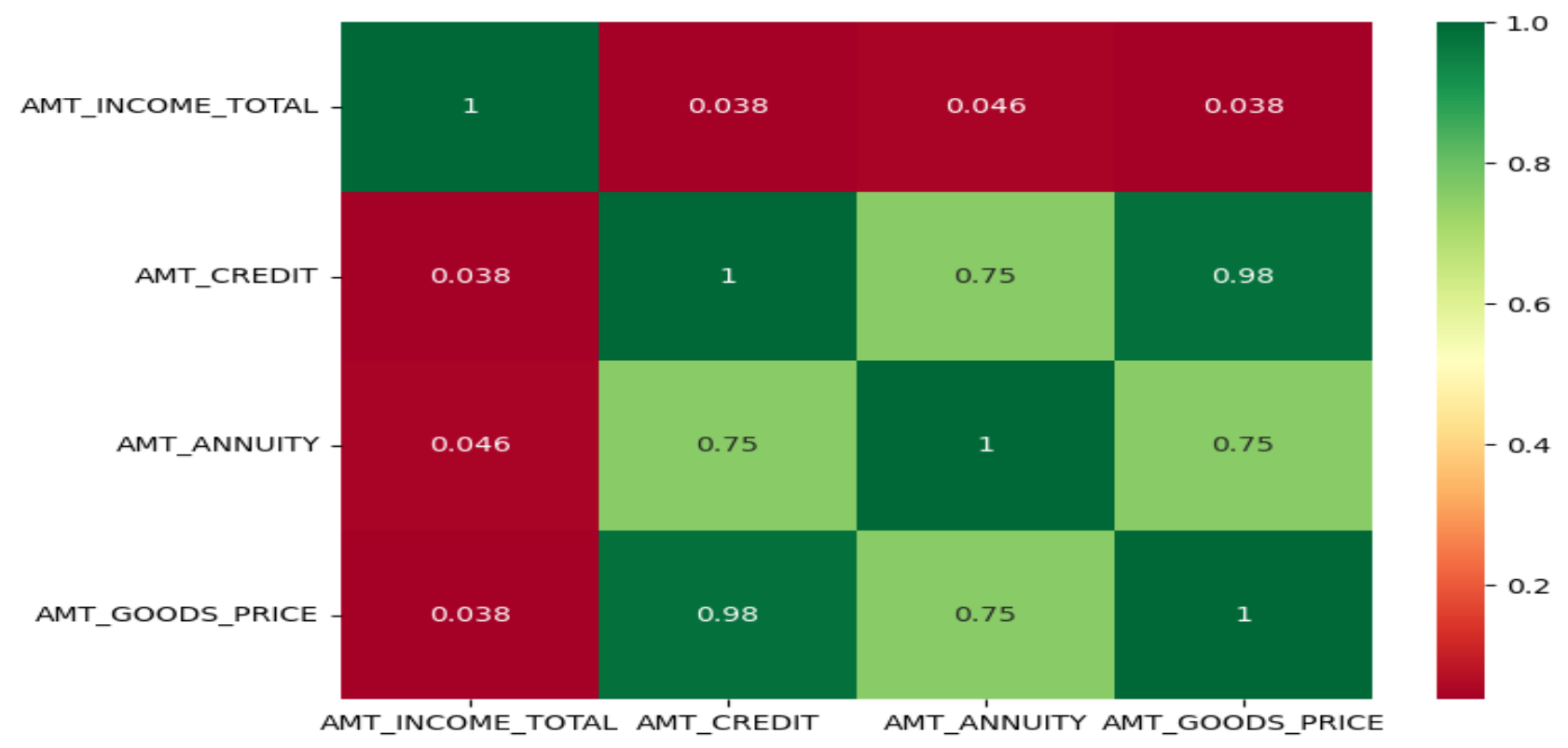
- Correlation matrix was first calculated For Defaulters and Re-payers separately and then graphs were plotted.

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE
AMT_INCOME_TOTAL	1.000000	0.038131	0.046421	0.037591
AMT_CREDIT	0.038131	1.000000	0.752195	0.982783
AMT_ANNUITY	0.046421	0.752195	1.000000	0.752295
AMT_GOODS_PRICE	0.037591	0.982783	0.752295	1.000000

	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE
AMT_INCOME_TOTAL	1.000000	0.342799	0.418953	0.349426
AMT_CREDIT	0.342799	1.000000	0.771309	0.987022
AMT_ANNUITY	0.418953	0.771309	1.000000	0.776433
AMT_GOODS_PRICE	0.349426	0.987022	0.776433	1.000000

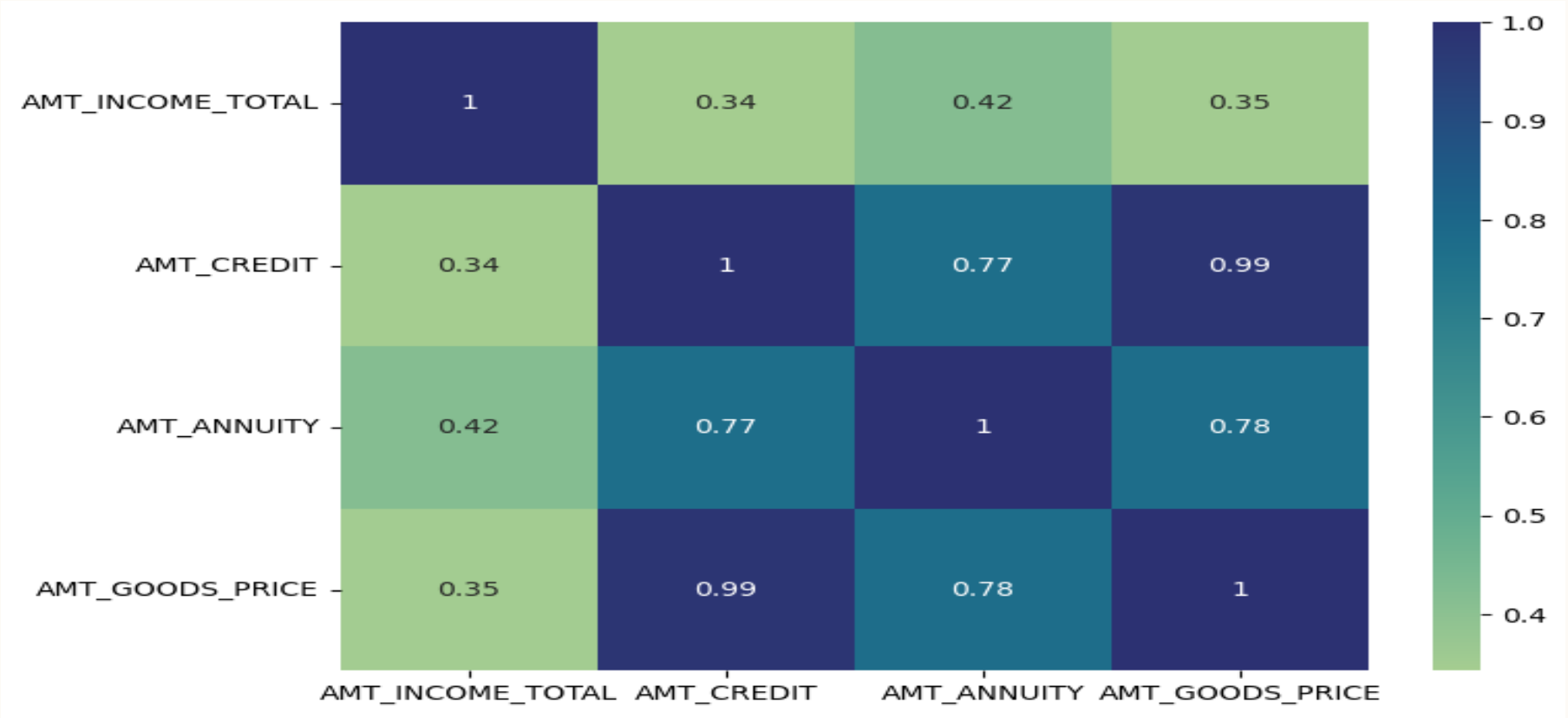
- For Defaulters-

1. AMT_CREDIT and AMT_ANNUITY (0.74)
2. AMT_CREDIT and AMT_GOODS_PRICE (0.98)
3. AMT_ANNUITY and AMT_GOODS_PRICE (0.74)



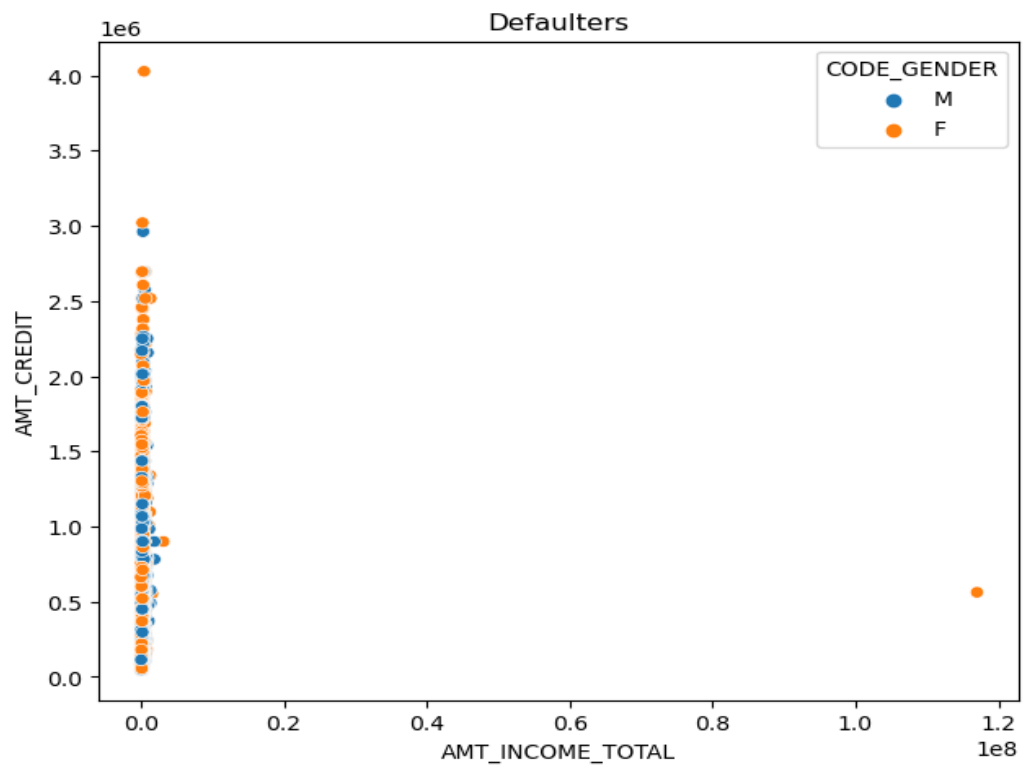
- For Re-payers –

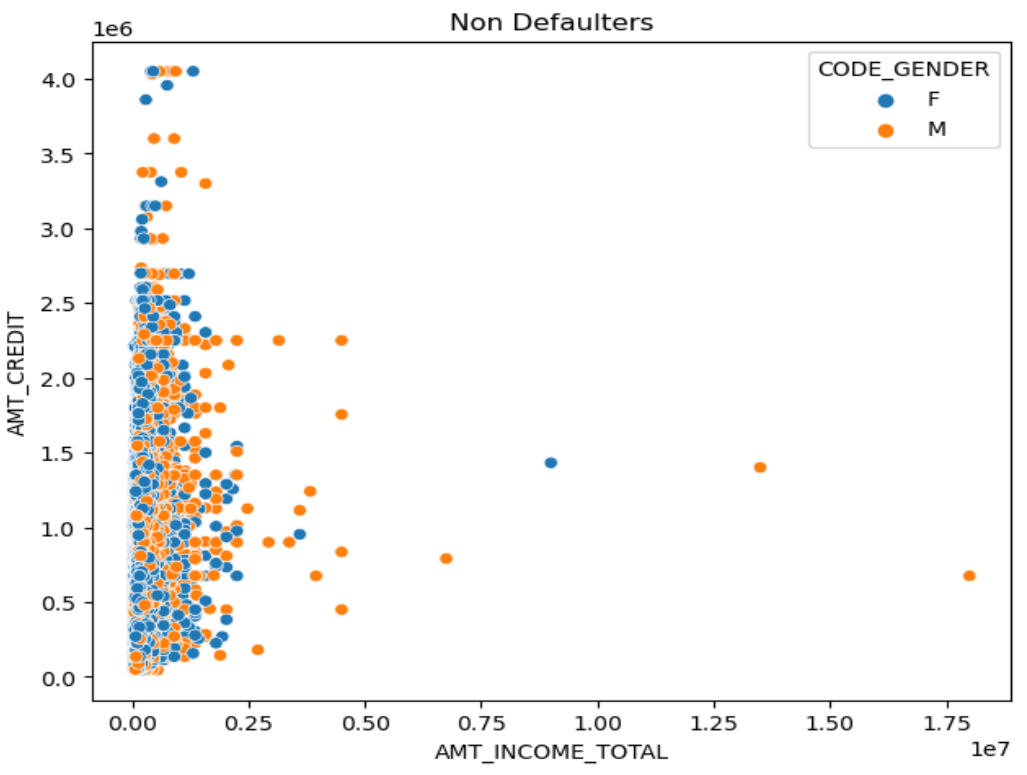
1. AMT_CREDIT and AMT_ANNUITY (0.76)
2. AMT_CREDIT and AMT_GOODS_PRICE (0.98)
3. AMT_ANNUITY and AMT_GOODS_PRICE (0.76)



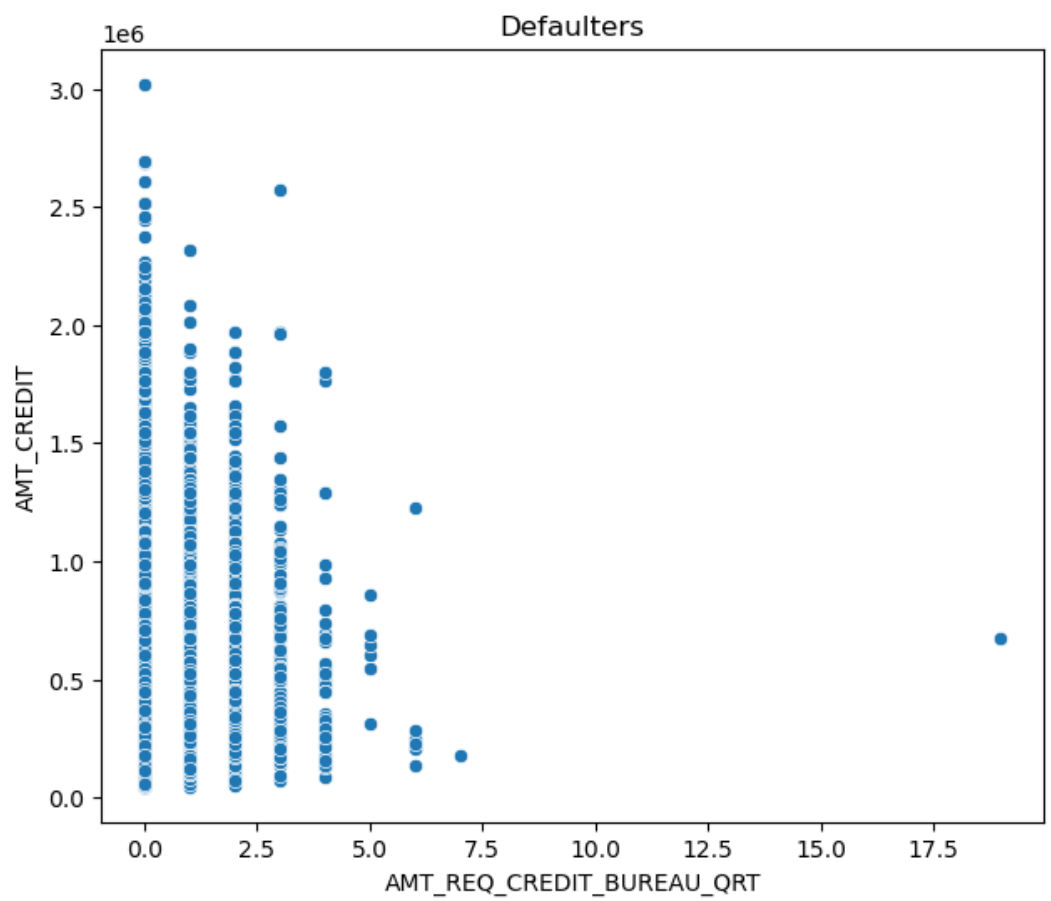
INSIGHTS(CONTINUOUS DATA)

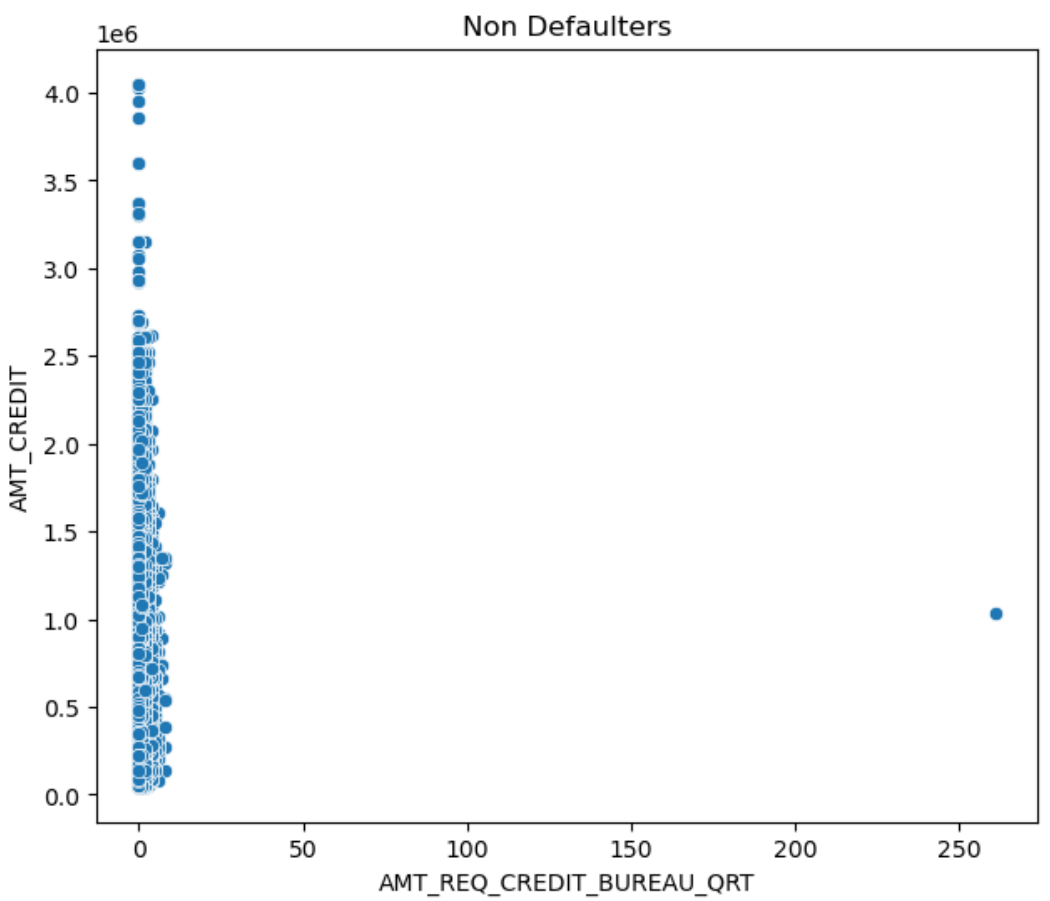
1. Defaulters - We can slightly figure out that the values are more concentrated on the lower income and lower credit of the loan. That means as the income is increased, the amount of loan is also increased. This is true for both the genders.
2. Non defaulters - We can hardly figure out any pattern out of this.





- The more number of enquiries the lesser the amount of loan credited for both defaulters and non defaulters.





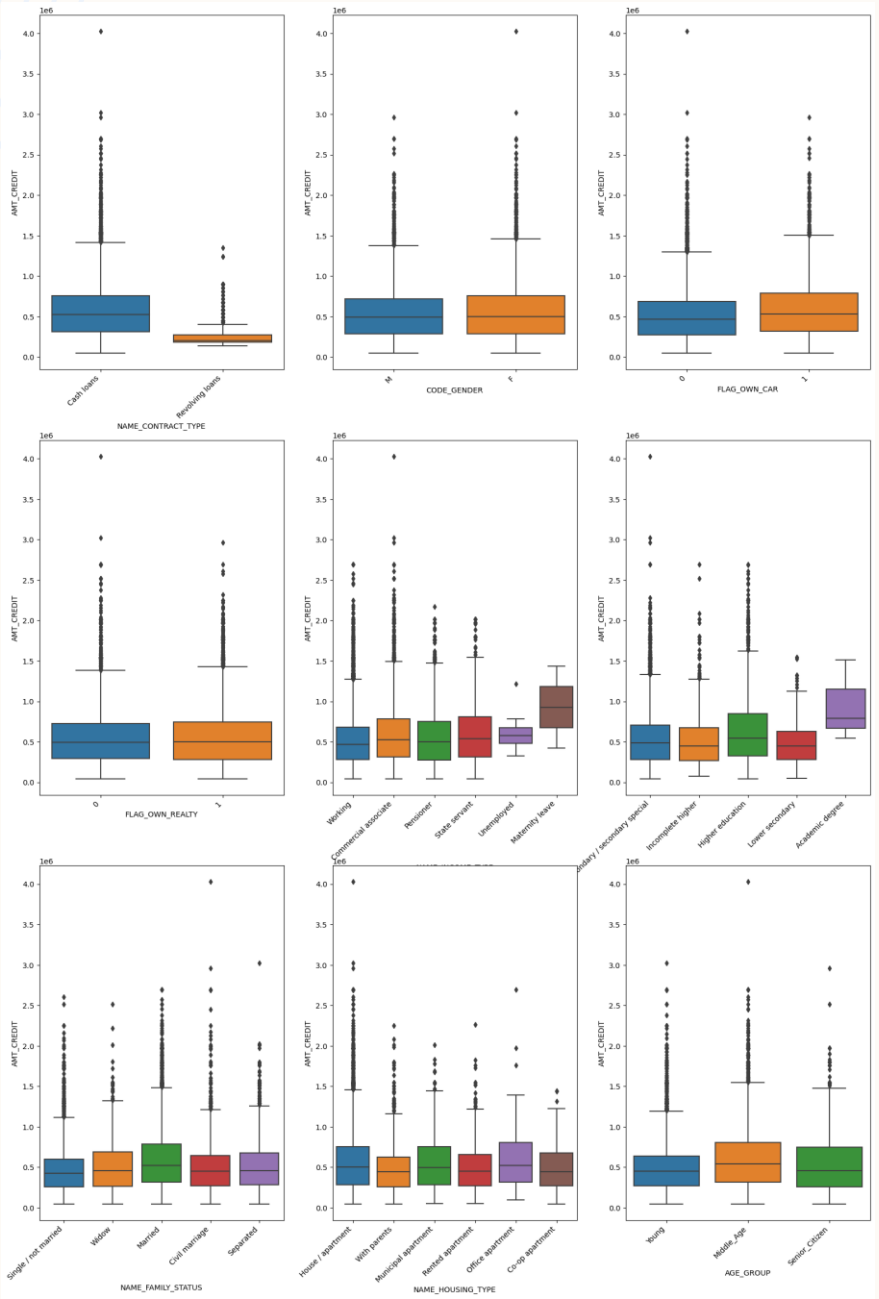
INSIGHTS(CONTINUOUS DATA)

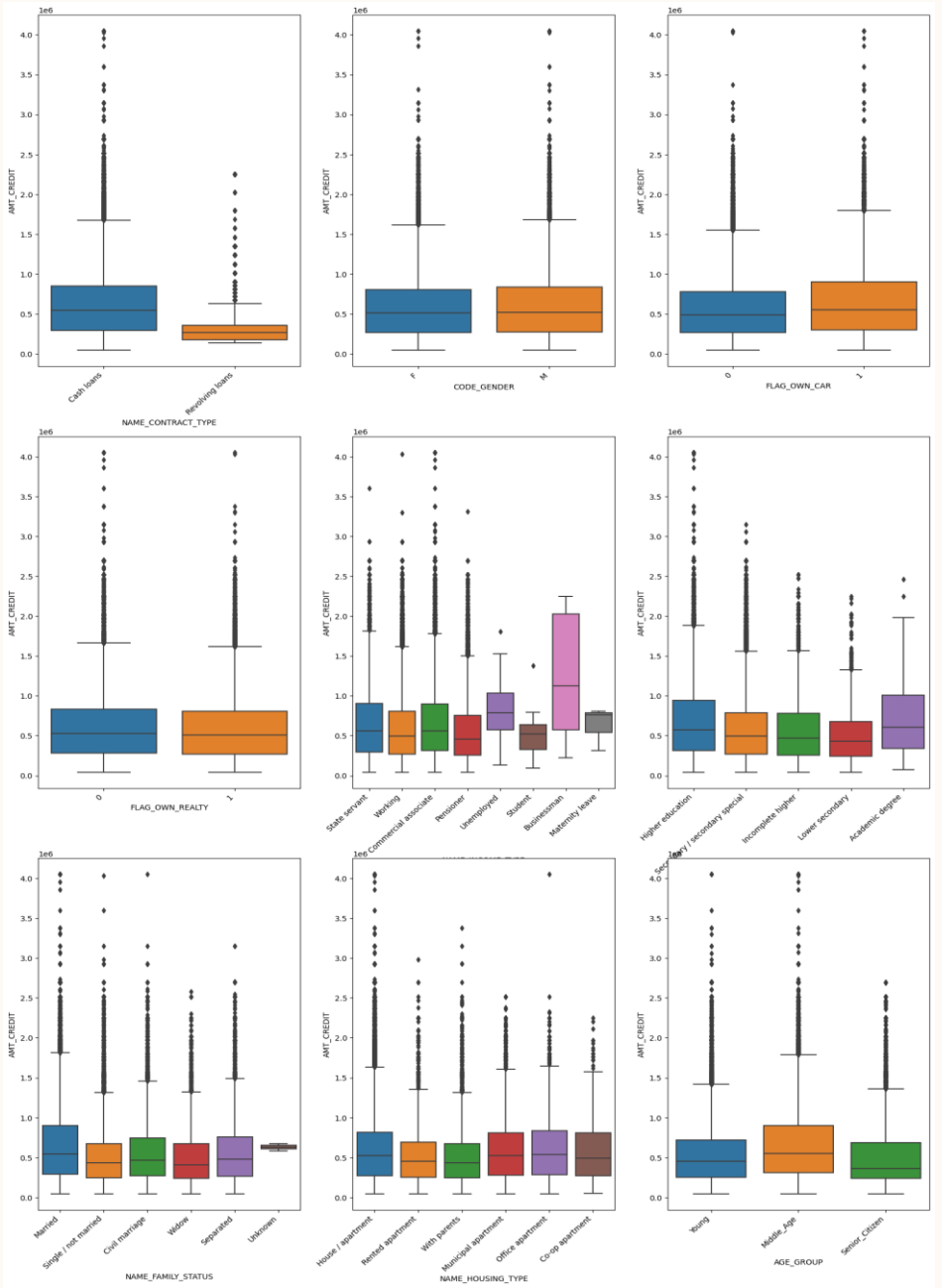
- Case of Defaulters-

1.Credit amount of the loans are very low for Revolving loans
2.There is no credit amount difference between genders, client owning cars or realty.
3.The Young age group got less amount of loan credited compared to mid age and senior citizen.
4.Higher income group have more loan amount credited.
5.Clients having higher external score have more loan amount.

- Case of Non Defaulters-

1. Credit amount of the loans are very low for Revolving loans
2. There is no credit amount difference between genders, client owning cars or realty.
3. The mid age group got more amount of loan credited compared to young and senior citizen.
4. Higher income group have more loan amount credited and lower the lowest.
5. Clients having higher external score have more loan amount.
6. Surprisingly the unemployed people have spike in credit amount of loan
7. The Married people have more loan amount credited.


AMT_CREDIT
1e6
NAME_CONTRACT_TYPE
Cash loans
Revolving loans
CODE_GENDER
M
F
FLAG_OWN_CAR
0
1
FLAG_OWN_REALTY
NAME_INCOME_TYPE
Working
Commercial associate
Pensioner
State servant
Unemployed
Maternity leave
NAME_EDUCATION_TYPE
Secondary / secondary special
Incomplete higher
Higher education
Lower secondary
Academic degree
NAME_FAMILY_STATUS
Single / not married
Widow
Married
Civil marriage
Separated
NAME_HOUSING_TYPE
House / apartment
With parents
Municipal apartment
Rented apartment
Office apartment
Co-op apartment
AGE_GROUP
Young
Middle_Age
Senior_Citizen


AMT_CREDIT
1e6
NAME_CONTRACT_TYPE
Cash loans
Revolving loans
CODE_GENDER
F
M
FLAG_OWN_CAR
0
1
FLAG_OWN_REALTY
NAME_INCOME_TYPE
State servant
Working
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave
NAME_EDUCATION_TYPE
Higher education
Secondary / secondary special
Incomplete higher
Lower secondary
Academic degree
NAME_FAMILY_STATUS
Married
Single / not married
Civil marriage
Widow
Separated
Unknown
NAME_HOUSING_TYPE
House / apartment
Rented apartment
With parents
Municipal apartment
Office apartment
Co-op apartment
AGE_GROUP
Young
Middle_Age
Senior_Citizen

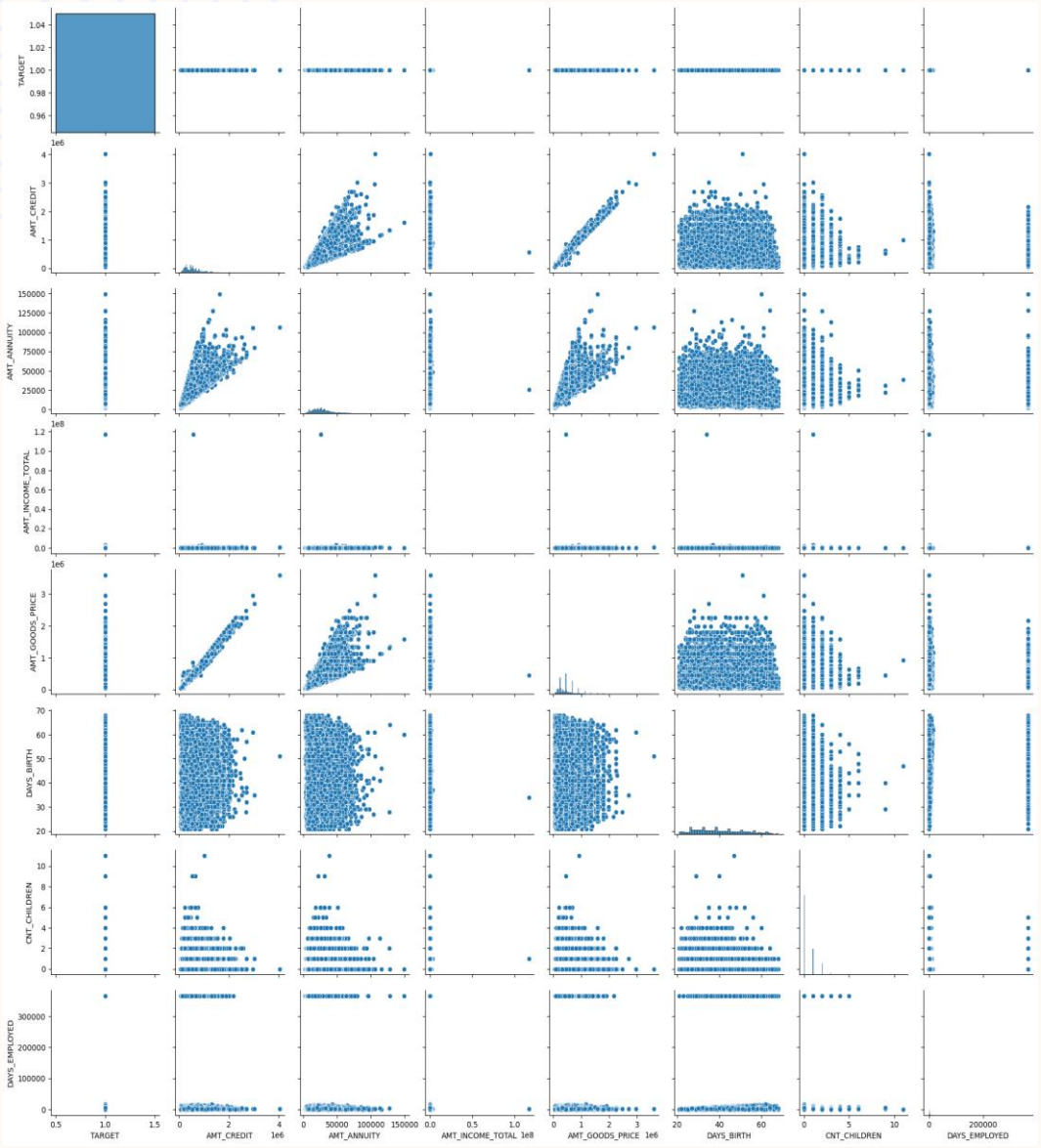
CORRELATION ANALYSIS

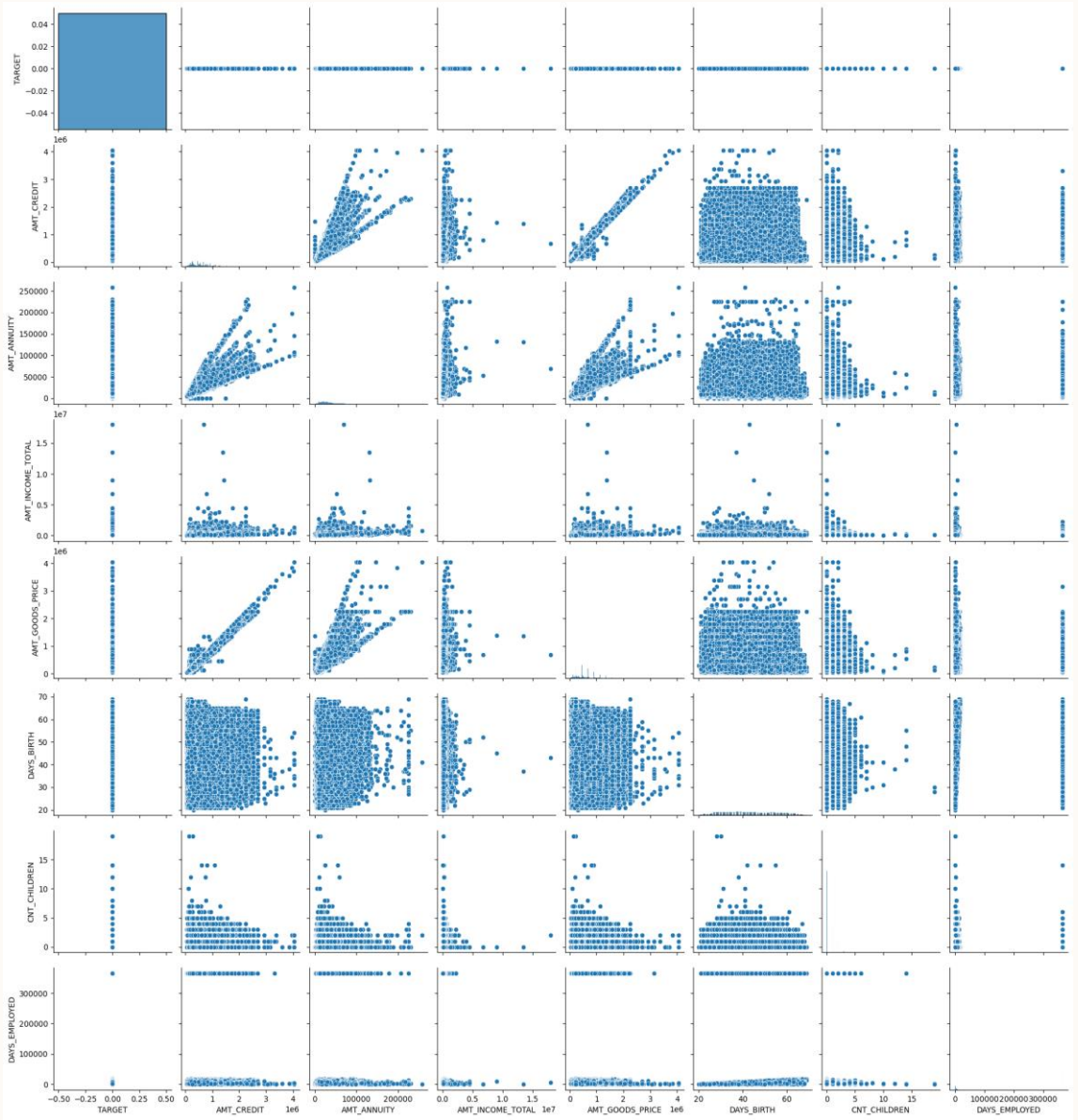
- Correlation Analysis is statistical method that is used to discover if there is a relationship between two variables/datasets, and how strong that relationship may be.
- We have performed this analysis in two ways-

1. Using Pair plots
2. Using Heat maps

INSIGHTS ON PAIR PLOTS

1. AMT_CREDIT and AMT_GOODS_PRICE are highly correlated variables for both defaulters and non-defaulters. So as the home price increases the loan amount also increases.
2. AMT_CREDIT and AMT_ANNUITY (EMI) are highly correlated variables for both defaulters and non-defaulters . So as the home price increases the EMI amount also increases which is logical.
3. All three variables AMT_CREDIT, AMT_GOODS_PRICE and AMT_ANNUITY are highly correlated for both defaulters and non-defaulters, which might not give a good indicator for defaulter detection.


TARGET
AMT_CREDIT
AMT_ANNUITY
AMT_INCOME_TOTAL
AMT_GOODS_PRICE
DAYS_BIRTH
CNT_CHILDREN
DAYS_EMPLOYED


TARGET
AMT_CREDIT
AMT_ANNUITY
AMT_INCOME_TOTAL
AMT_GOODS_PRICE
DAYS_BIRTH
CNT_CHILDREN
DAYS_EMPLOYED

INSIGHTS ON HEATMAPS

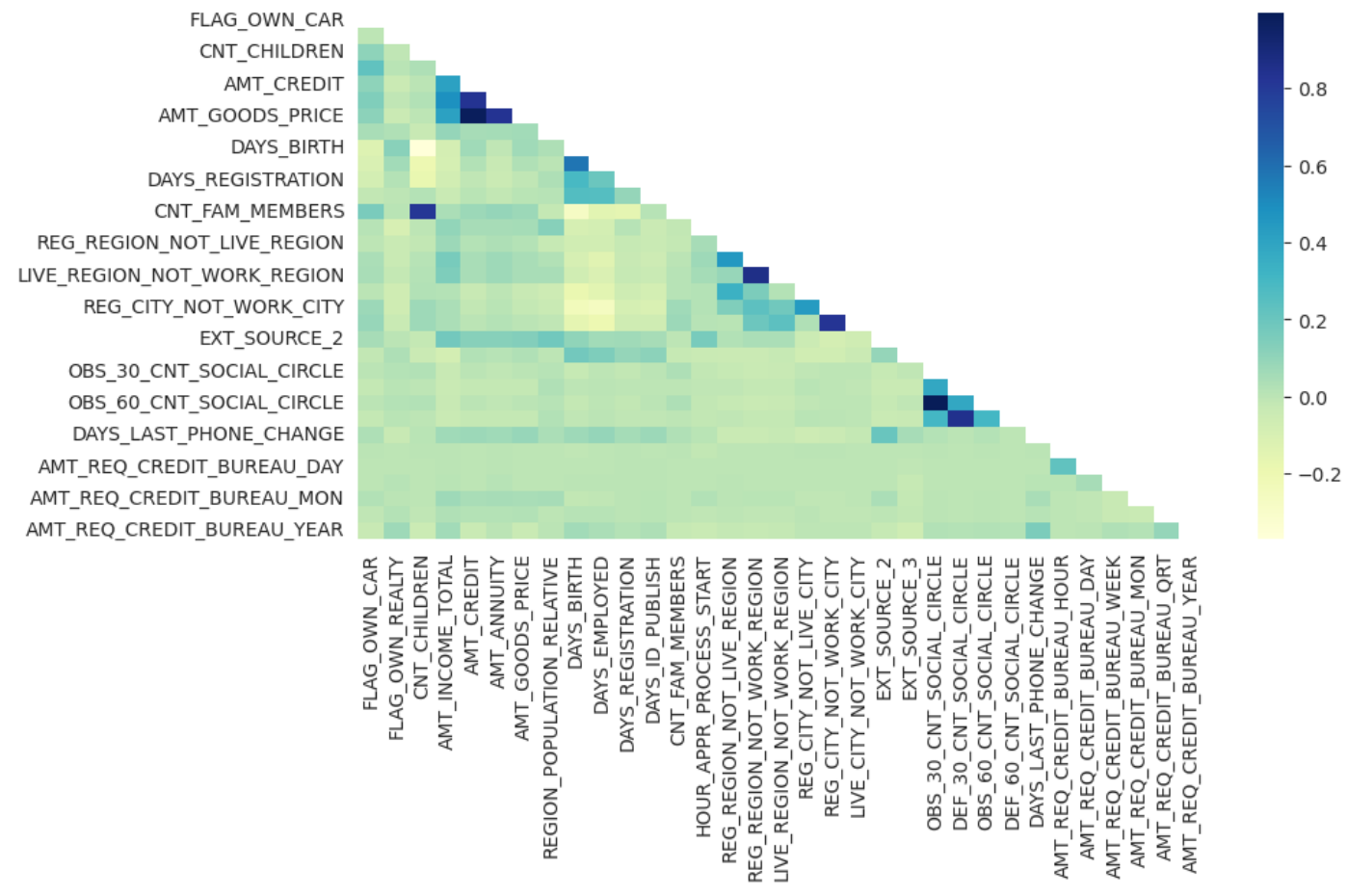
Target 0

1. AMT_CREDIT is inversely proportional to the DAYS_BIRTH , peoples belongs to low-age group taking high Credit amount and vice-versa
2. AMT_CREDIT is inversely proportional to the CNT_CHILDREN, means Credit amount is higher for less children count client have and vice-versa.
3. AMT_INCOME_TOTAL is inversely proportional to the CNT_CHILDREN, means more income for less children client have and vice-versa.
4. less children client have in densely populated area.
5. AMT_CREDIT is higher to densely populated area.
6. AMT_INCOME_TOTAL is also higher in densely populated area.

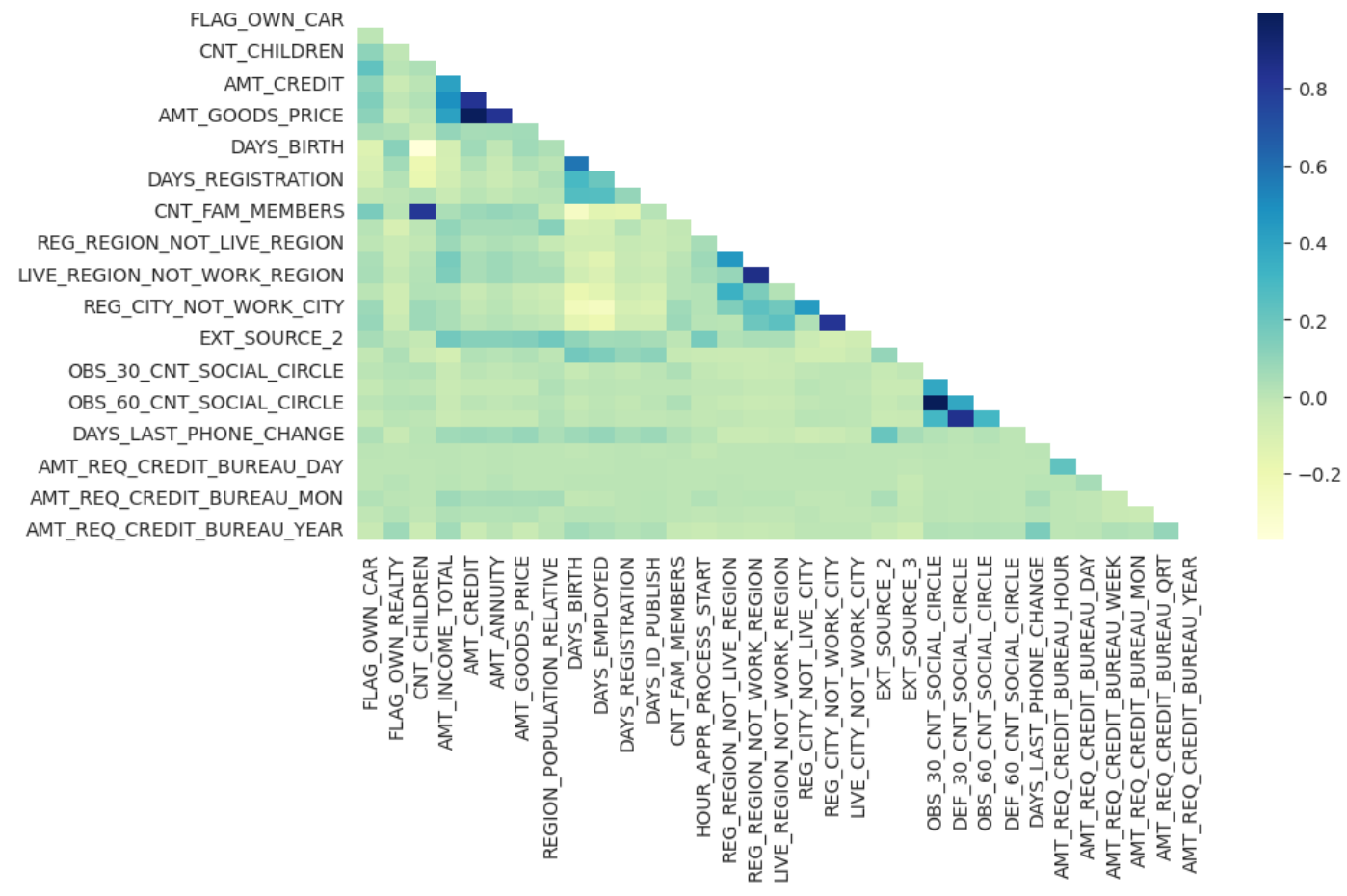
Target 1

1. This heat map for Target 1 is also having quite a same observation just like Target 0. But for few points are different. They are listed below.
2. The client's permanent address does not match contact address are having less children.
3. The client's permanent address does not match work address are having less children.

TARGET 0


FLAG_OWN_CAR
CNT_CHILDREN
AMT_CREDIT
AMT_GOODS_PRICE
DAYS_BIRTH
DAYS_REGISTRATION
CNT_FAM_MEMBERS
REG_REGION_NOT_LIVE_REGION
LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_WORK_CITY
EXT_SOURCE_2
OBS_30_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
DAYS_LAST_PHONE_CHANGE
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_YEAR
FLAG_OWN_CAR
FLAG_OWN_REALTY
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
CNT_FAM_MEMBERS
HOUR_APPR_PROCESS_START
REG_REGION_NOT_LIVE_REGION
REG_REGION_NOT_WORK_REGION
LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
EXT_SOURCE_2
EXT_SOURCE_3
OBS_30_CNT_SOCIAL_CIRCLE
DEF_30_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
DAYS_LAST_PHONE_CHANGE
AMT_REQ_CREDIT_BUREAU_HOUR
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_WEEK
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_QRT
AMT_REQ_CREDIT_BUREAU_YEAR
0.8
0.6
0.4
0.2
0.0
−0.2

TARGET 1


FLAG_OWN_CAR
CNT_CHILDREN
AMT_CREDIT
AMT_GOODS_PRICE
DAYS_BIRTH
DAYS_REGISTRATION
CNT_FAM_MEMBERS
REG_REGION_NOT_LIVE_REGION
LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_WORK_CITY
EXT_SOURCE_2
OBS_30_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
DAYS_LAST_PHONE_CHANGE
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_YEAR
FLAG_OWN_CAR
FLAG_OWN_REALTY
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
CNT_FAM_MEMBERS
HOUR_APPR_PROCESS_START
REG_REGION_NOT_LIVE_REGION
REG_REGION_NOT_WORK_REGION
LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
EXT_SOURCE_2
EXT_SOURCE_3
OBS_30_CNT_SOCIAL_CIRCLE
DEF_30_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
DAYS_LAST_PHONE_CHANGE
AMT_REQ_CREDIT_BUREAU_HOUR
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_WEEK
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_QRT
AMT_REQ_CREDIT_BUREAU_YEAR
0.8
0.6
0.4
0.2
0.0
−0.2

TOP 10 CORRELATION

- TARGET 0

	VAR1	VAR2	CORRELATION	CORR_ABS
934	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998510	0.998510
286	AMT_GOODS_PRICE	AMT_CREDIT	0.987022	0.987022
494	CNT_FAM_MEMBERS	CNT_CHILDREN	0.878571	0.878571
647	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.861861	0.861861
970	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.859371	0.859371
755	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.830381	0.830381
287	AMT_GOODS_PRICE	AMT_ANNUITY	0.776433	0.776433
251	AMT_ANNUITY	AMT_CREDIT	0.771309	0.771309
395	DAYS_EMPLOYED	DAYS_BIRTH	0.626028	0.626028
611	REG_REGION_NOT_WORK_REGION	REG_REGION_NOT_LIVE_REGION	0.446101	0.446101

TOP 10 CORRELATION

- TARGET 1

	VAR1	VAR2	CORRELATION	CORR_ABS
934	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998270	0.998510
286	AMT_GOODS_PRICE	AMT_CREDIT	0.982783	0.987022
494	CNT_FAM_MEMBERS	CNT_CHILDREN	0.885484	0.878571
647	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.847885	0.861861
970	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.869016	0.859371
755	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.778540	0.830381
287	AMT_GOODS_PRICE	AMT_ANNUITY	0.752295	0.776433
251	AMT_ANNUITY	AMT_CREDIT	0.752195	0.771309
395	DAYS_EMPLOYED	DAYS_BIRTH	0.582441	0.626028
611	REG_REGION_NOT_WORK_REGION	REG_REGION_NOT_LIVE_REGION	0.497937	0.446101

ANALYSIS OF previous_application.csv

LOADING DATASET

```
df_prev = pd.read_csv('previous_application.csv')
df_prev.head()
```

	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	WEEKI
0	2030495	271877	Consumer loans	1730.430	17145.0	17145.0	0.0	17145.0	
1	2802425	108129	Cash loans	25188.615	607500.0	679671.0	NaN	607500.0	
2	2523466	122040	Cash loans	15060.735	112500.0	136444.5	NaN	112500.0	
3	2819243	176158	Cash loans	47041.335	450000.0	470790.0	NaN	450000.0	
4	1784265	202054	Cash loans	31924.395	337500.0	404055.0	NaN	337500.0	

5 rows × 37 columns

Dataset was loaded and all the columns were read for Data Inspection(37 columns were discovered)

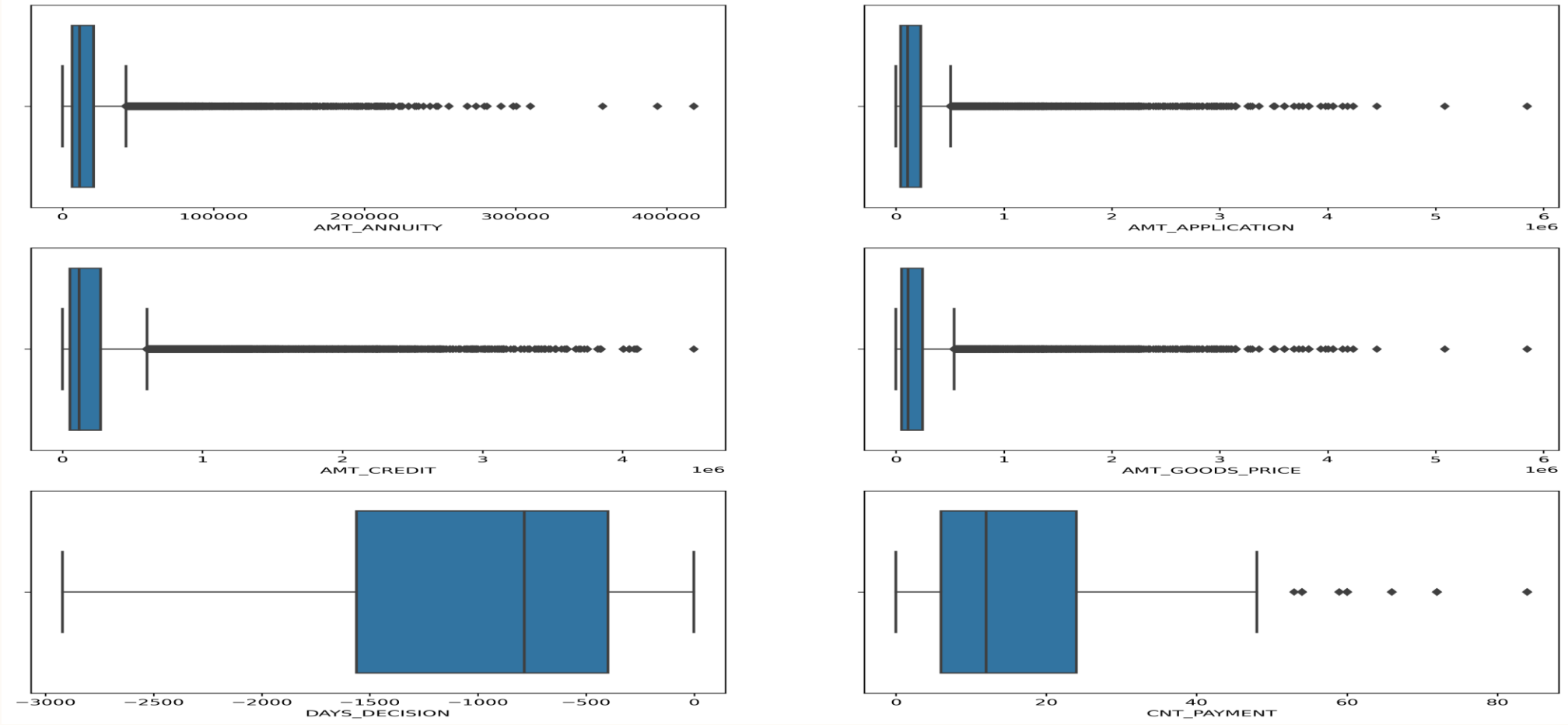
HANDLING MISSING VALUES

- Data inspection was performed similar to previous dataset and missing values were handled using Null Value percentage and imputations were performed.

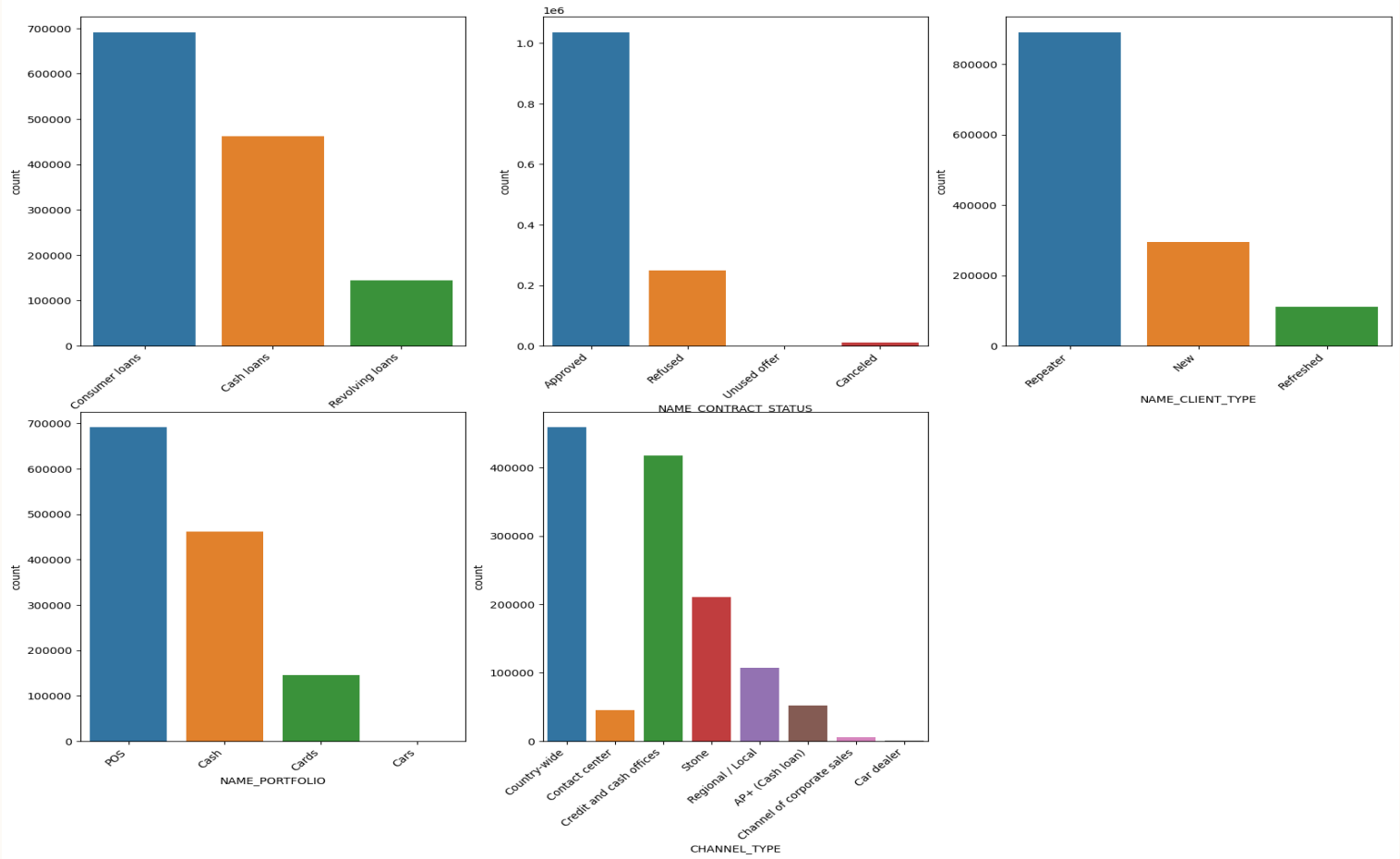
```
]: SK_ID_PREV                0.00
   SK_ID_CURR                0.00
   NAME_CONTRACT_TYPE        0.00
   AMT_ANNUITY               2.87
   AMT_APPLICATION           0.00
   AMT_CREDIT                0.00
   AMT_GOODS_PRICE           3.87
   NAME_CONTRACT_STATUS      0.00
   DAYS_DECISION             0.00
   NAME_CLIENT_TYPE          0.07
   NAME_PORTFOLIO            2.87
   CHANNEL_TYPE              0.00
   SELLERPLACE_AREA          0.00
   CNT_PAYMENT               2.87
   PRODUCT_COMBINATION       0.00
   dtype: float64
```

```
SK_ID_PREV                0.00
SK_ID_CURR                0.00
NAME_CONTRACT_TYPE        0.00
AMT_ANNUITY               0.00
AMT_APPLICATION           0.00
AMT_CREDIT                0.00
AMT_GOODS_PRICE           3.98
NAME_CONTRACT_STATUS      0.00
DAYS_DECISION             0.00
NAME_CLIENT_TYPE          0.07
NAME_PORTFOLIO            0.00
CHANNEL_TYPE              0.00
SELLERPLACE_AREA          0.00
CNT_PAYMENT               0.00
PRODUCT_COMBINATION       0.00
dtype: float64
```

OUTLIERS ANALYSIS



DATA IMBALANCE



RESULTS

1.Learned how a company should manage risk during giving loans to clients.

2. Understood visualizing techniques using python libraries (pandas, matplotlib, seaborn, etc.) and the senseful data from the graphs and charts.

3. Learned to present valuable insights and driving factors from the huge dataset.

4. Helped me to use EDA (Exploratory Data Analysis)and understand Risk Analysis in real business case scenes.

5. Understood about Outliers, Imputations, Handling missing values and errors, Data Imbalances and Variate analysis.

THANK YOU

Ankita Yadav
ankitaydv07@gmail.com