LOAN APPLICATION ANALYSIS

SOLON KUMAR DAS

BUSINESS UNDERSTANDING

LOAN APPLICATION ANALYSIS

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse the patterns present in the data. This will ensure that the applicants are capable of repaying the loan are not rejected.

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.

The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:

- **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.

The dataset has 2 files as explained below:

- **'application_data.csv'** :- This file contains all the information of the client at the time of application and the data is about whether a client has payment difficulties.
- **'previous_application.csv'** :- This file contains information about the client's previous loan data and tt contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

When a client applies for a loan, there are four types of decisions that could be taken by the client/company):

1. **Approved:** The Company has approved loan Application
2. **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.
3. **Refused:** The company had rejected the loan (because the client does not meet their requirements etc.).
4. **Unused offer:** Loan has been cancelled by the client but on different stages of the process.

In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

DATASET UNDERSTANDING

'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.**

Columns:

```
['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER', 'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN', 'AMT
_INCOME_TOTAL', 'AMT_CREDIT', 'AMT_ANNUITY', 'AMT_GOODS_PRICE', 'NAME_TYPE_SUITE', 'NAME_INCOME_TYPE', 'NAME_EDUCATIO
N_TYPE', 'NAME_FAMILY_STATUS', 'NAME_HOUSING_TYPE', 'REGION_POPULATION_RELATIVE', 'DAYS_BIRTH', 'DAYS_EMPLOYED', 'DAY
S_REGISTRATION', 'DAYS_ID_PUBLISH', 'OWN_CAR_AGE', 'FLAG_MOBIL', 'FLAG_EMP_PHONE', 'FLAG_WORK_PHONE', 'FLAG_CONT_MOBI
LE', 'FLAG_PHONE', 'FLAG_EMAIL', 'OCCUPATION_TYPE', 'CNT_FAM_MEMBERS', 'REGION_RATING_CLIENT', 'REGION_RATING_CLIENT_
W_CITY', 'WEEKDAY_APPR_PROCESS_START', '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'
, 'ORGANIZATION_TYPE', 'EXT_SOURCE_1', 'EXT_SOURCE_2', 'EXT_SOURCE_3', 'APARTMENTS_AVG', 'BASEMENTAREA_AVG', 'YEARS_B
EGINEXPLUATATION_AVG', 'YEARS_BUILD_AVG', 'COMMONAREA_AVG', 'ELEVATORS_AVG', 'ENTRANCES_AVG', 'FLOORSMAX_AVG', 'FLOOR
SMIN_AVG', 'LANDAREA_AVG', 'LIVINGAPARTMENTS_AVG', 'LIVINGAREA_AVG', 'NONLIVINGAPARTMENTS_AVG', 'NONLIVINGAREA_AVG',
'APARTMENTS_MODE', 'BASEMENTAREA_MODE', 'YEARS_BEGINEXPLUATATION_MODE', 'YEARS_BUILD_MODE', 'COMMONAREA_MODE', 'ELEVA
TORS_MODE', 'ENTRANCES_MODE', 'FLOORSMAX_MODE', 'FLOORSMIN_MODE', 'LANDAREA_MODE', 'LIVINGAPARTMENTS_MODE', 'LIVINGAR
EA_MODE', 'NONLIVINGAPARTMENTS_MODE', 'NONLIVINGAREA_MODE', 'APARTMENTS_MEDI', 'BASEMENTAREA_MEDI', 'YEARS_BEGINEXPLU
ATATION_MEDI', 'YEARS_BUILD_MEDI', 'COMMONAREA_MEDI', 'ELEVATORS_MEDI', 'ENTRANCES_MEDI', 'FLOORSMAX_MEDI', 'FLOORSMI
N_MEDI', 'LANDAREA_MEDI', 'LIVINGAPARTMENTS_MEDI', 'LIVINGAREA_MEDI', 'NONLIVINGAPARTMENTS_MEDI', 'NONLIVINGAREA_MEDI
', 'FONDKAPREMONT_MODE', 'HOUSETYPE_MODE', 'TOTALAREA_MODE', 'WALLSMATERIAL_MODE', 'EMERGENCYSTATE_MODE', 'OBS_30_CNT
_SOCIAL_CIRCLE', 'DEF_30_CNT_SOCIAL_CIRCLE', 'OBS_60_CNT_SOCIAL_CIRCLE', 'DEF_60_CNT_SOCIAL_CIRCLE', 'DAYS_LAST_PHONE
_CHANGE', 'FLAG_DOCUMENT_2', 'FLAG_DOCUMENT_3', 'FLAG_DOCUMENT_4', 'FLAG_DOCUMENT_5', 'FLAG_DOCUMENT_6', 'FLAG_DOCUME
NT_7', 'FLAG_DOCUMENT_8', 'FLAG_DOCUMENT_9', 'FLAG_DOCUMENT_10', 'FLAG_DOCUMENT_11', 'FLAG_DOCUMENT_12', 'FLAG_DOCUME
NT_13', 'FLAG_DOCUMENT_14', 'FLAG_DOCUMENT_15', 'FLAG_DOCUMENT_16', 'FLAG_DOCUMENT_17', 'FLAG_DOCUMENT_18', 'FLAG_DOC
UMENT_19', 'FLAG_DOCUMENT_20', 'FLAG_DOCUMENT_21', 'AMT_REQ_CREDIT_BUREAU_HOUR', 'AMT_REQ_CREDIT_BUREAU_DAY', 'AMT_RE
Q_CREDIT_BUREAU_WEEK', 'AMT_REQ_CREDIT_BUREAU_MON', 'AMT_REQ_CREDIT_BUREAU_QRT', 'AMT_REQ_CREDIT_BUREAU_YEAR']
```

DATASET UNDERSTANDING

'previous_application.csv : This file contains information about the client's previous loan data and tt contains the data whether the previous application had been Approved, Cancelled, Refused or Unused offer.

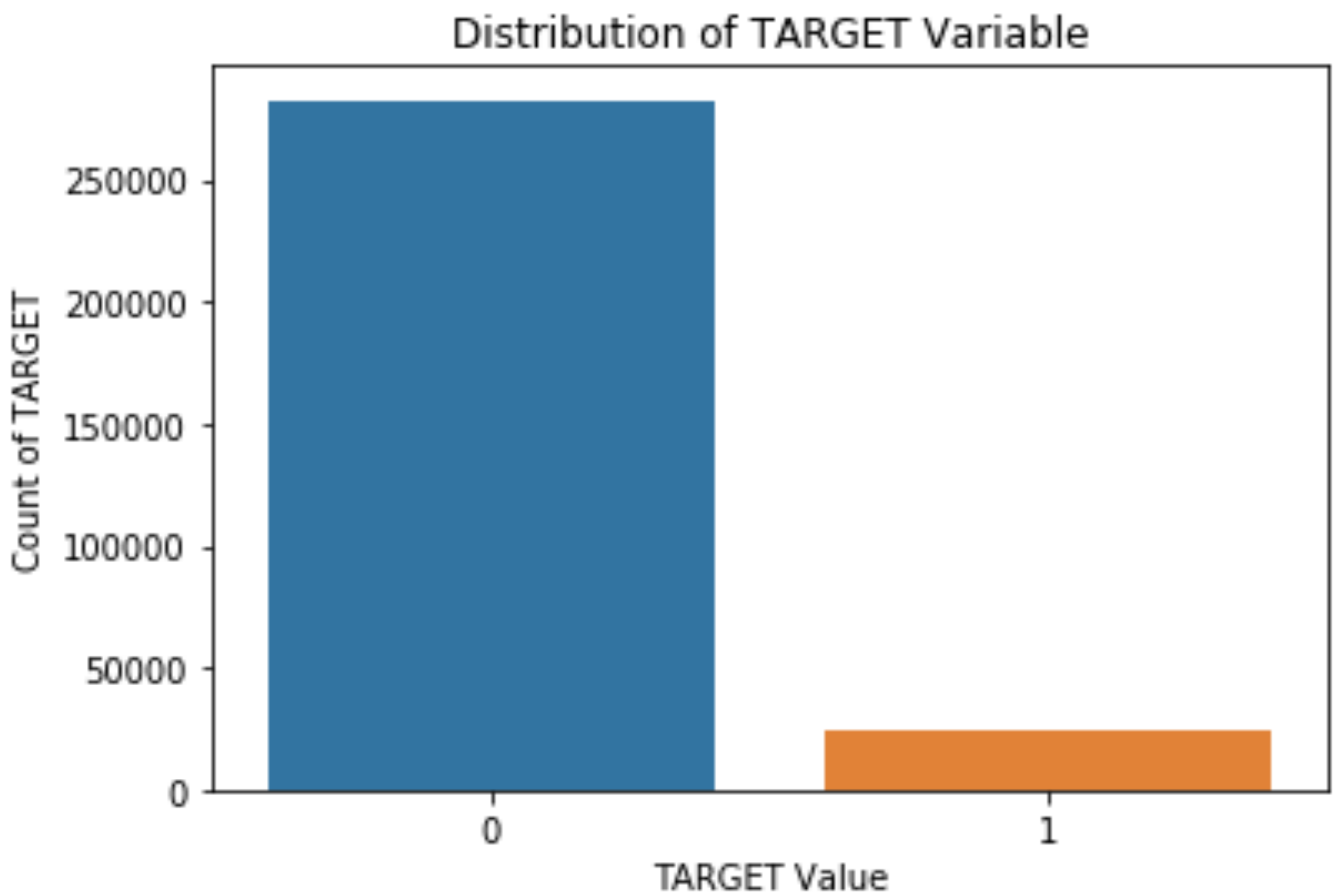
Columns:

```
['SK_ID_PREV', 'SK_ID_CURR', 'NAME_CONTRACT_TYPE', 'AMT_ANNUITY',
'AMT_APPLICATION', 'AMT_CREDIT', 'AMT_DOWN_PAYMENT', 'AMT_GOODS_PRICE',
'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START',
'FLAG_LAST_APPL_PER_CONTRACT', 'NFLAG_LAST_APPL_IN_DAY', 'RATE_DOWN_PAYMENT',
'RATE_INTEREST_PRIMARY', 'RATE_INTEREST_PRIVILEGED',
'NAME_CASH_LOAN_PURPOSE', 'NAME_CONTRACT_STATUS', 'DAYS_DECISION',
'NAME_PAYMENT_TYPE', 'CODE_REJECT_REASON', 'NAME_TYPE_SUITE',
'NAME_CLIENT_TYPE', 'NAME_GOODS_CATEGORY', 'NAME_PORTFOLIO',
'NAME_PRODUCT_TYPE', 'CHANNEL_TYPE', 'SELLERPLACE_AREA',
'NAME_SELLER_INDUSTRY', 'CNT_PAYMENT', 'NAME_YIELD_GROUP',
'PRODUCT_COMBINATION', 'DAYS_FIRST_DRAWING', 'DAYS_FIRST_DUE',
'DAYS_LAST_DUE_1ST_VERSION', 'DAYS_LAST_DUE', 'DAYS_TERMINATION',
'NFLAG_INSURED_ON_APPROVAL']
```

Analysis & Insights of Applications Data

Data Imbalance Check

Analysis



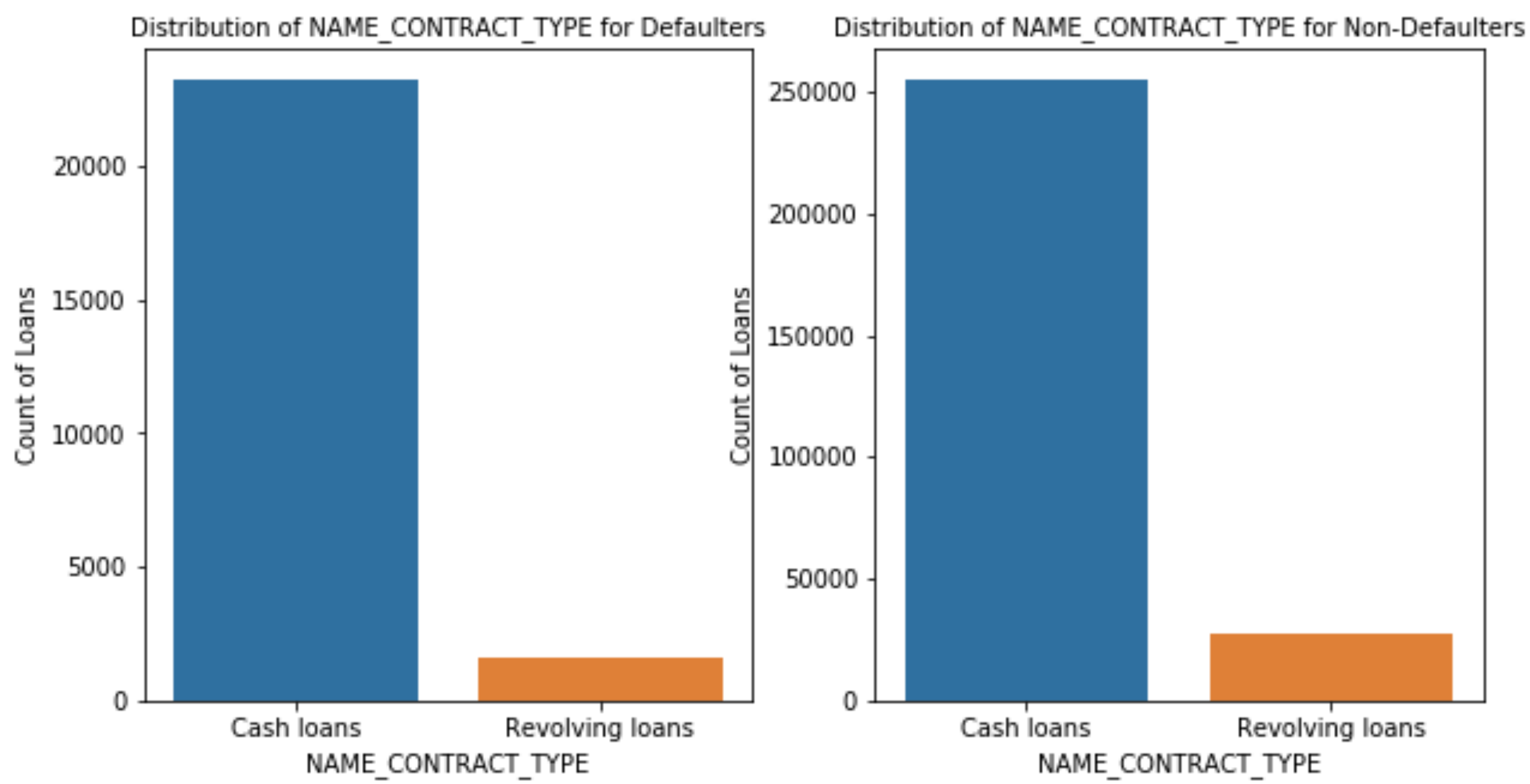
We see that this is a highly imbalanced dataset because there are far more loans that were repaid on time than the loans that were not repaid.
More than 250000 loans were repaid whereas less than 50000 loans were not repaid.

The percentage of people who have paid their loans is 91.93%
The percentage of people who have not paid their loans is 8.07%
The Ratio of Data Imbalance is 11.39

Univariate Analysis on Categorical Variables

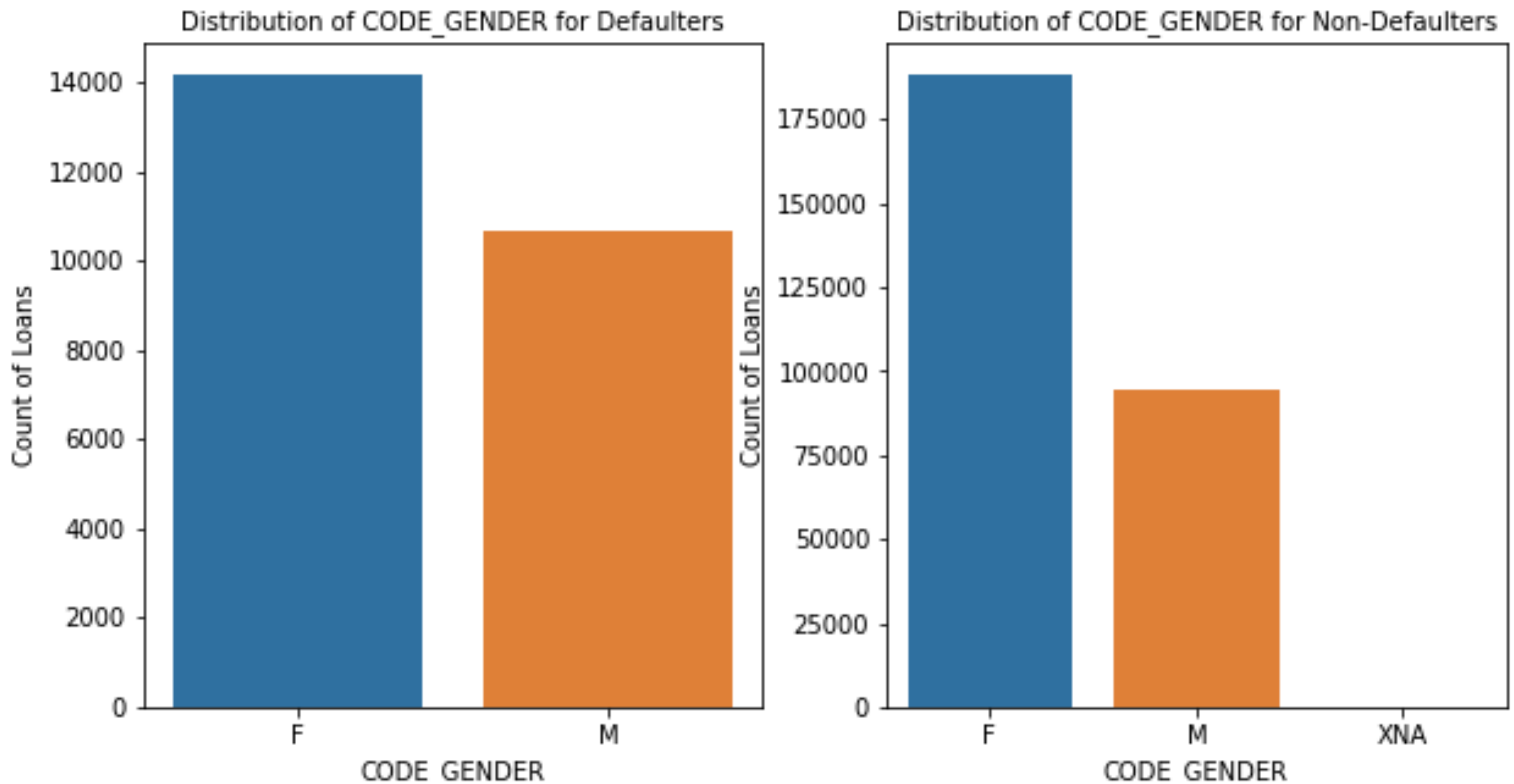
Analysis

Variable: **NAME_CONTRACT_TYPE**



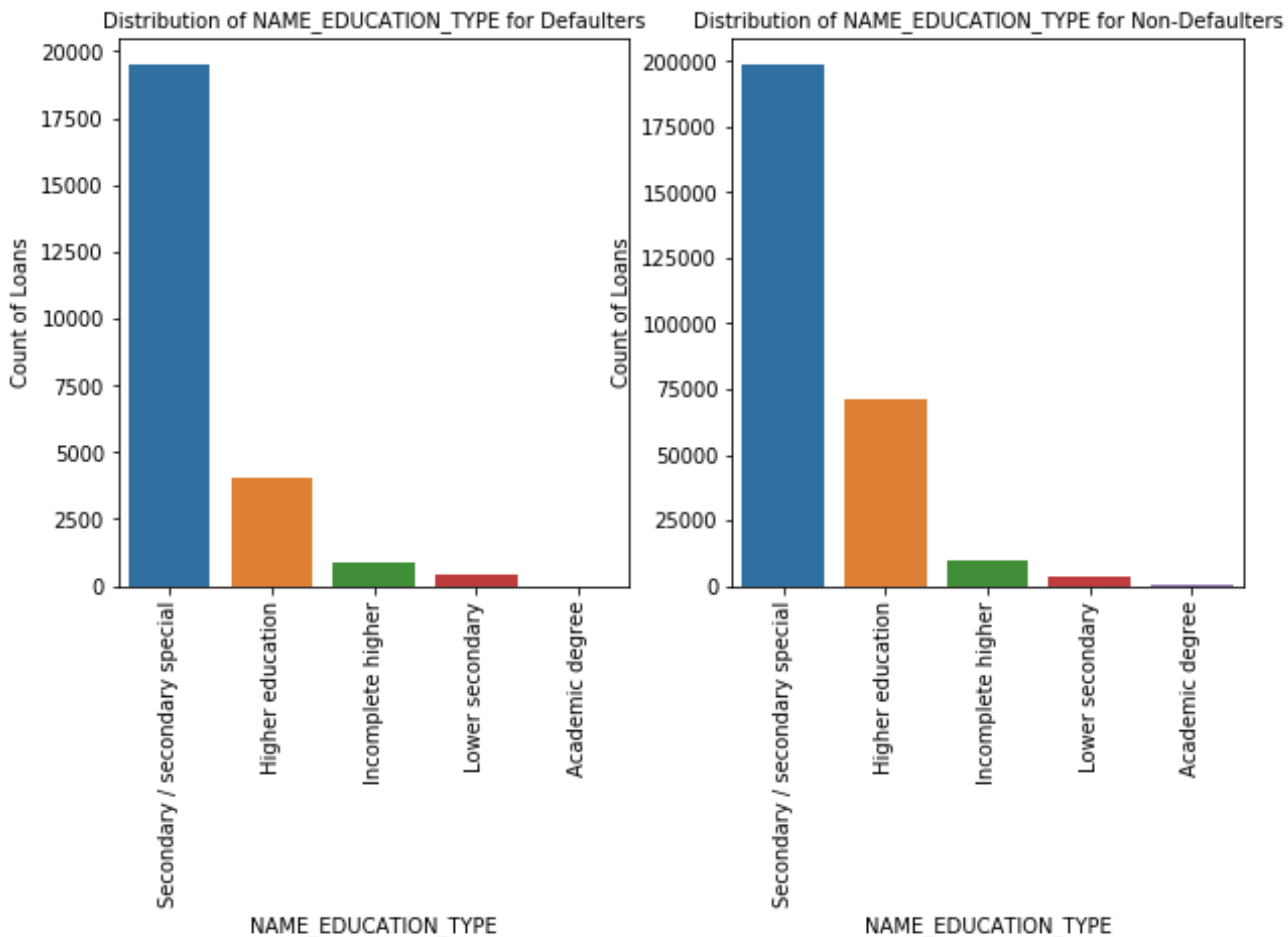
Hence we see for both the targets the cash loans is much higher than Revolving Loans

Variable: **CODE_GENDER**



We see that for both Targets Females are taking more loans than Males

Variable: **NAME_EDUCATION_TYPE**



We see that people with Academic degree rarely take loans and if they take it is all repaid. Hence potential customers.

The ratio of loan count for higher education is also more for non defaulters. So even the higher education has no difficulty

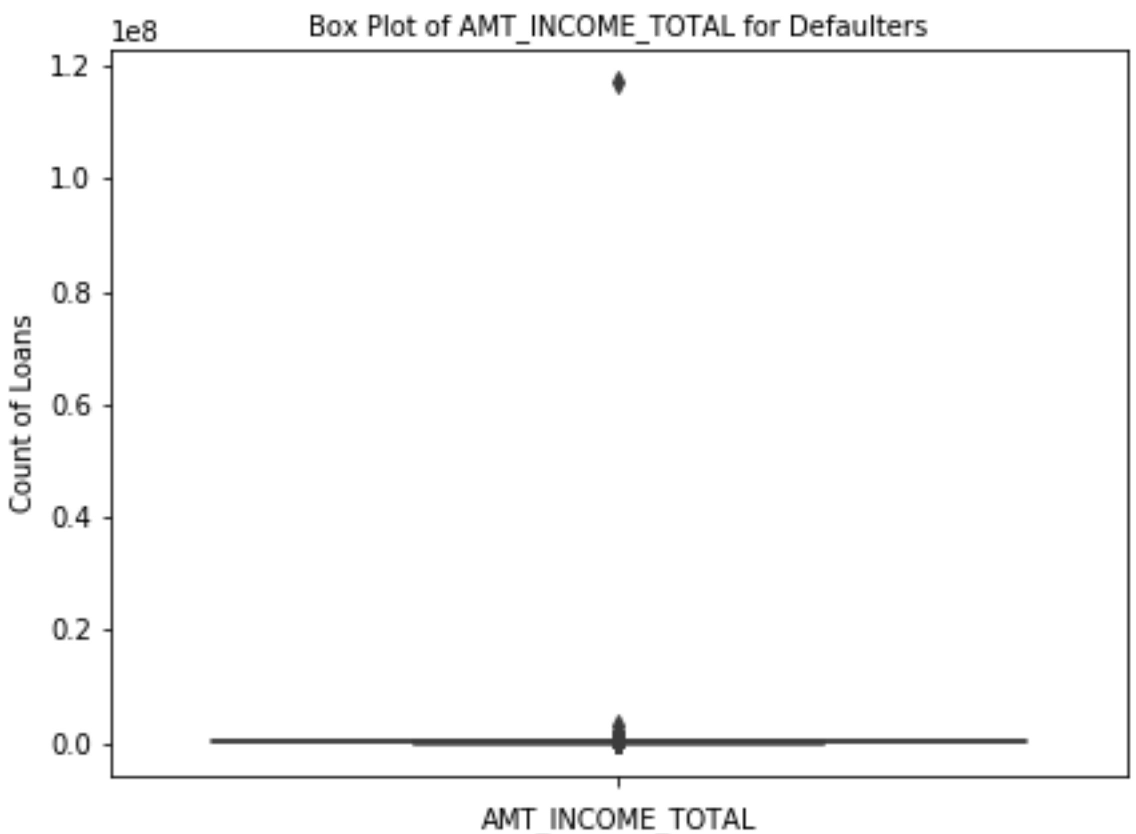
Univariate Analysis on Numeric Variables

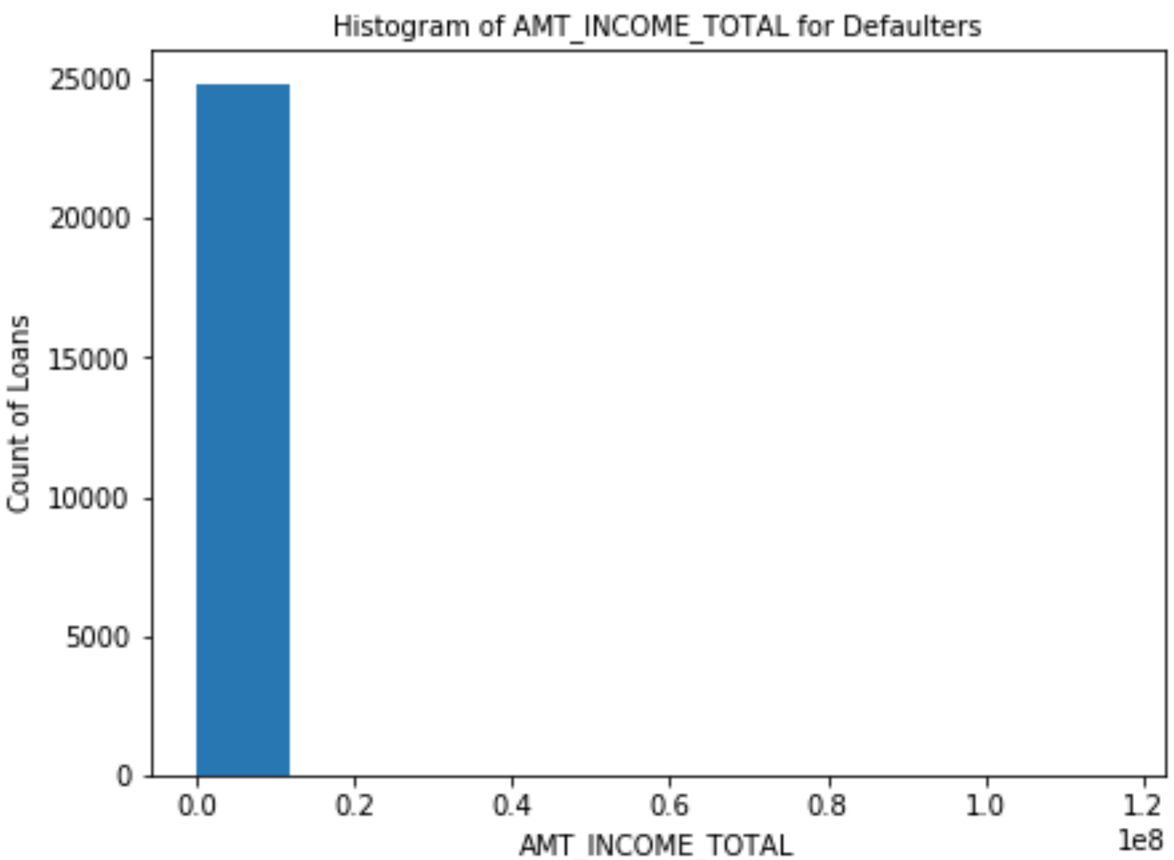
Analysis

Variable: **AMT_INCOME_TOTAL**

There were outliers in this data column because income of few people can be considerably high or considerably low so we need to remove those outliers to see the actual spread of data and where the majority income bracket lies.

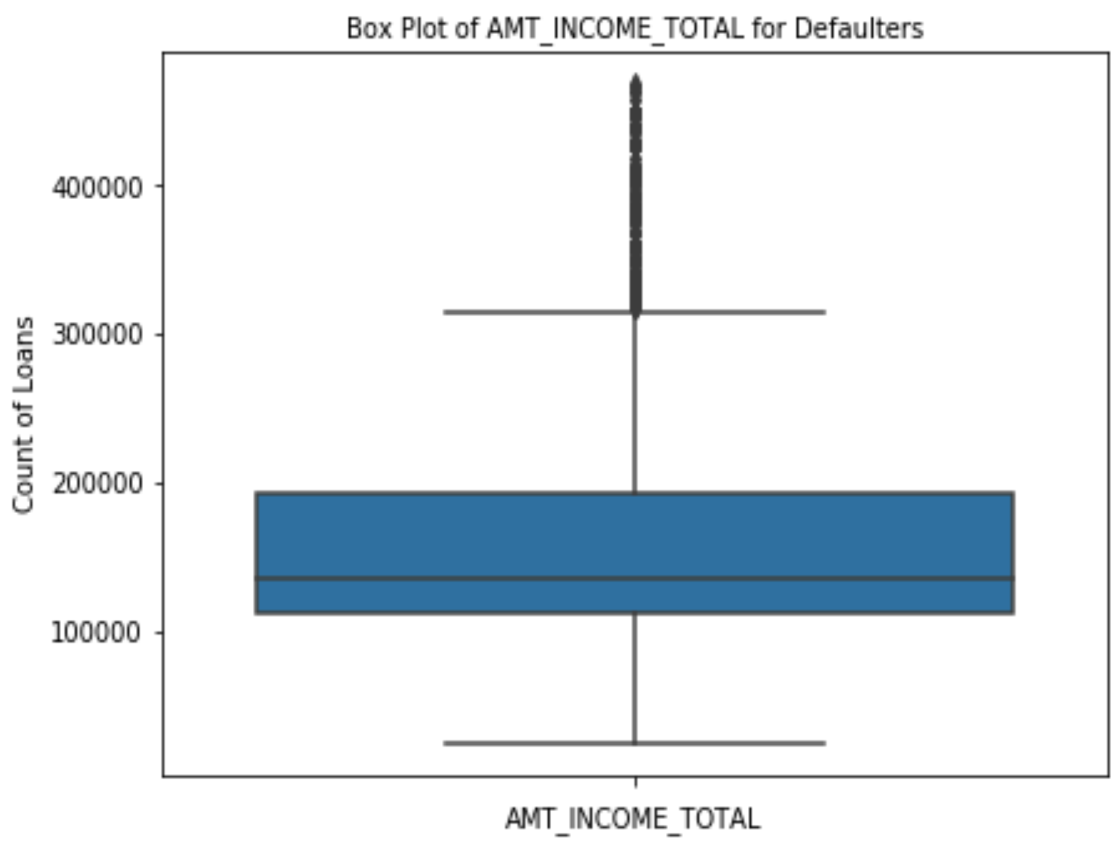
Before Outlier Removal

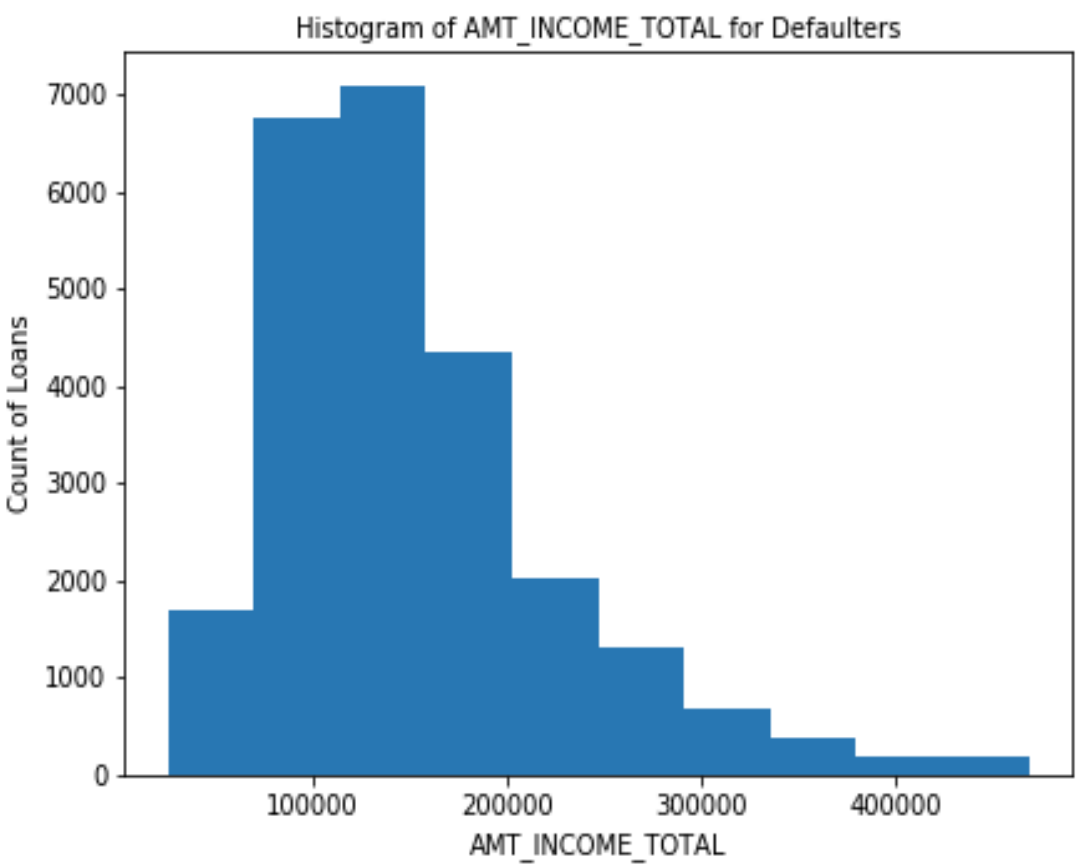




We can’t see the spread of data clearly so. We need to remove the outliers

After Outlier Removal

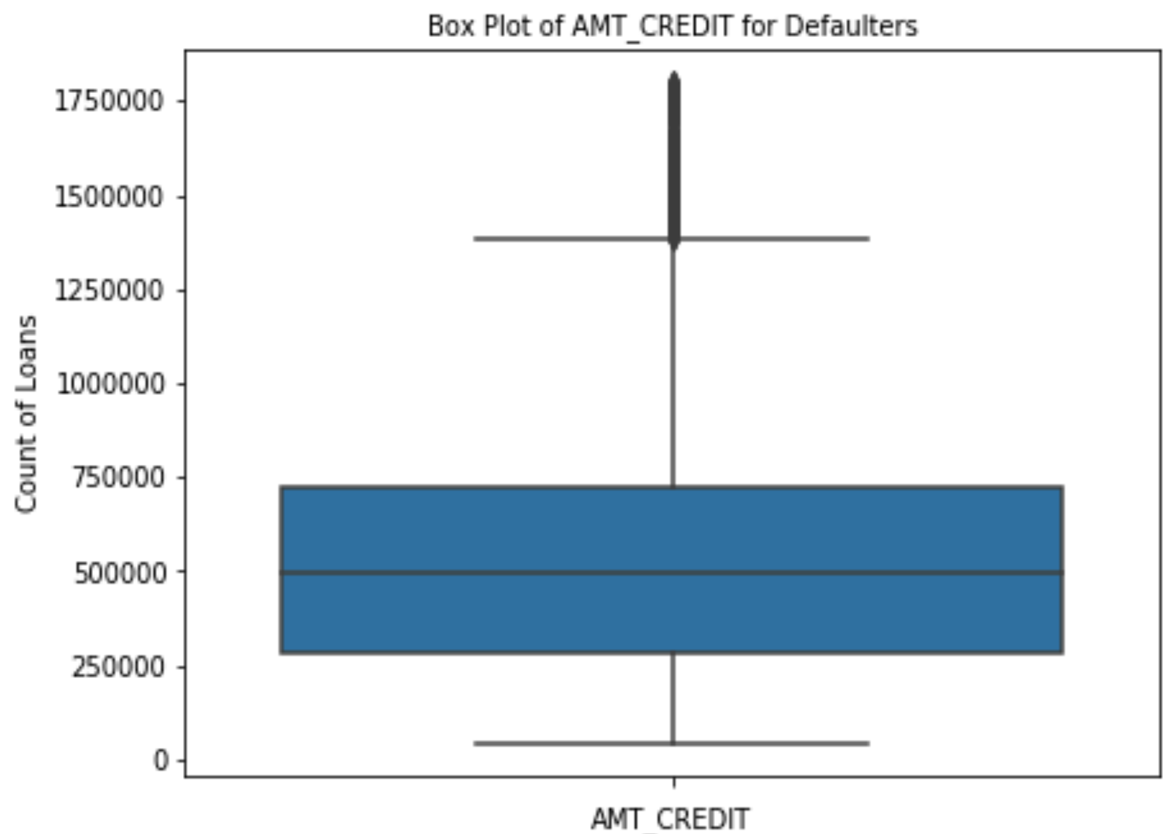


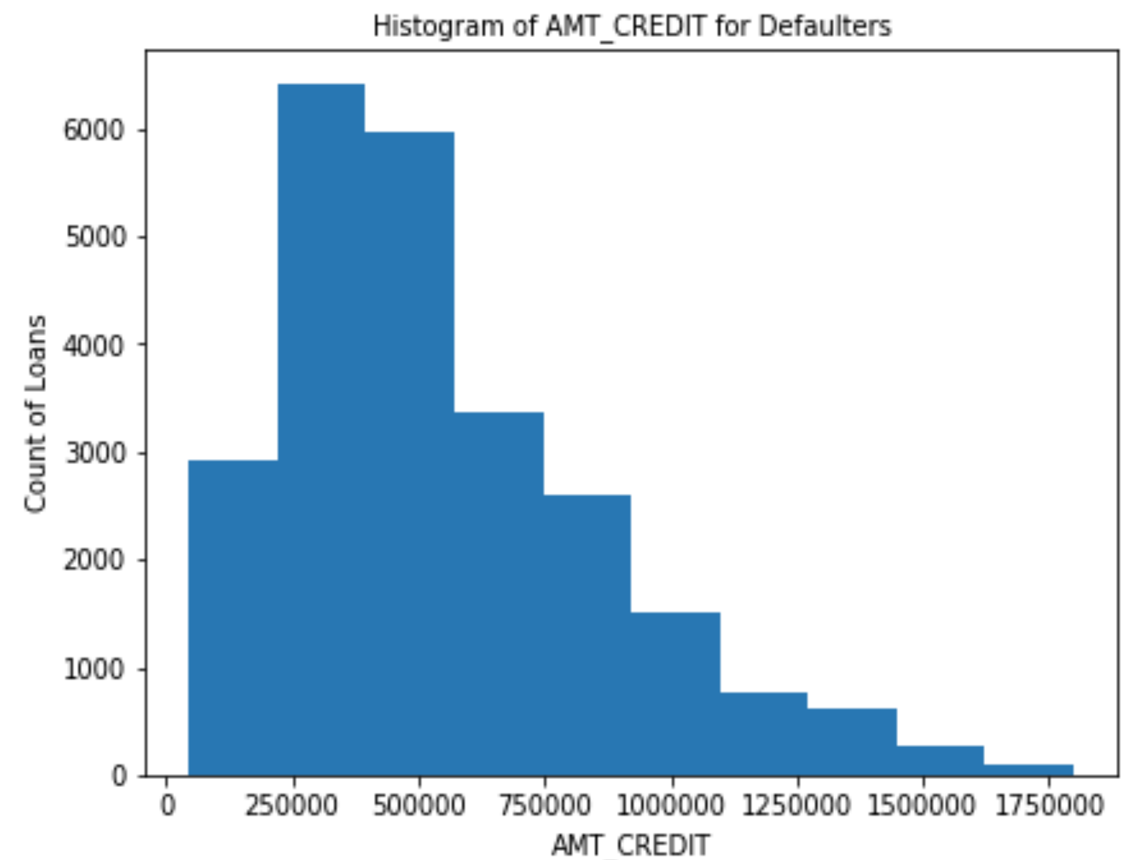


Now that we can see the spread of data clearly
This tells us that most people who take loans are in the range of 100000 to 200000
Number of loans taken are very less outside this spread of value

Variable: **AMT_CREDIT**

After Outlier Removal



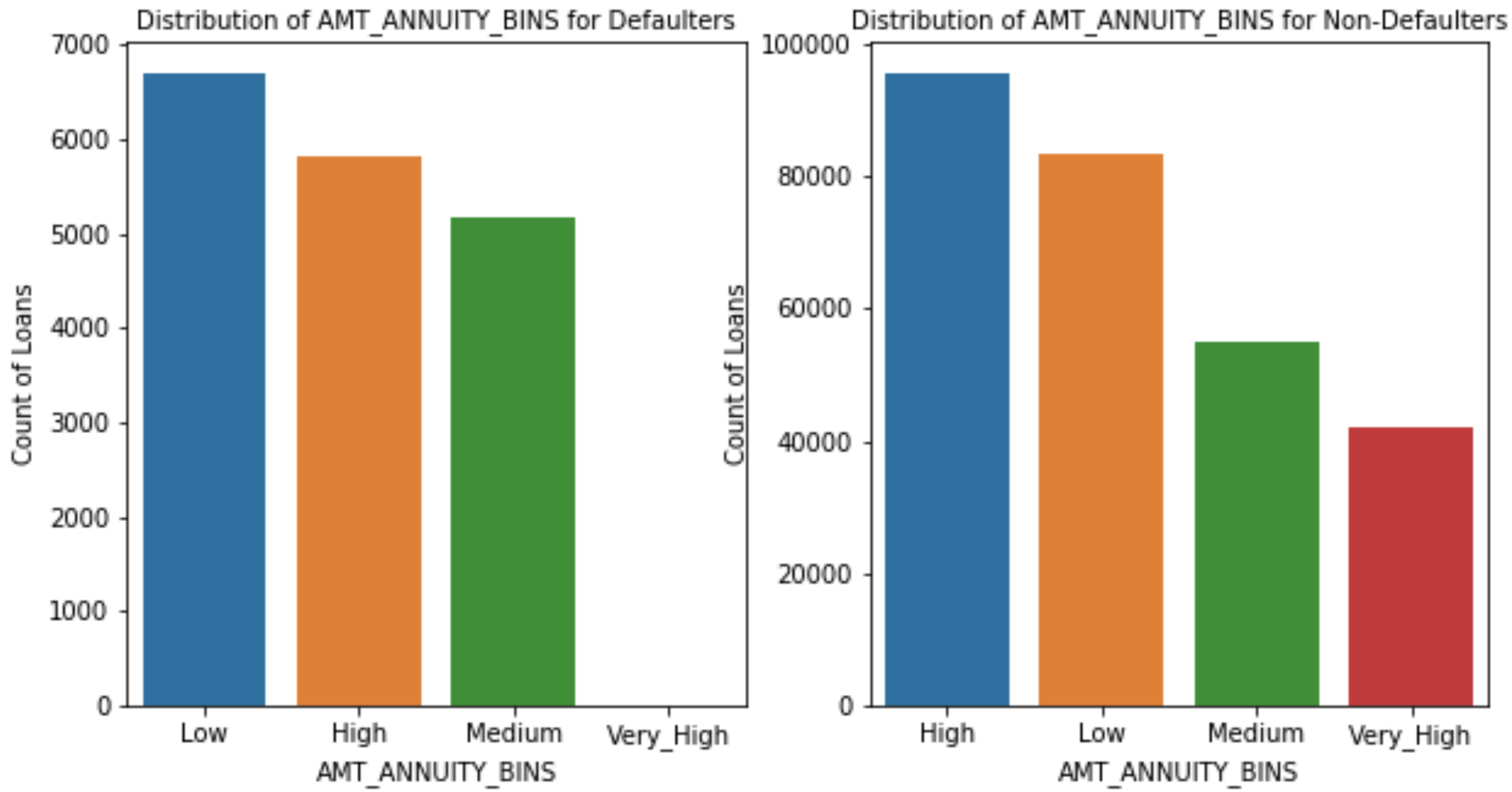


We see that the Amount credited lies in the range of 250000 and 750000

Binning on continuos variables

Analysis

Variable: **AMT_ANNUITY**

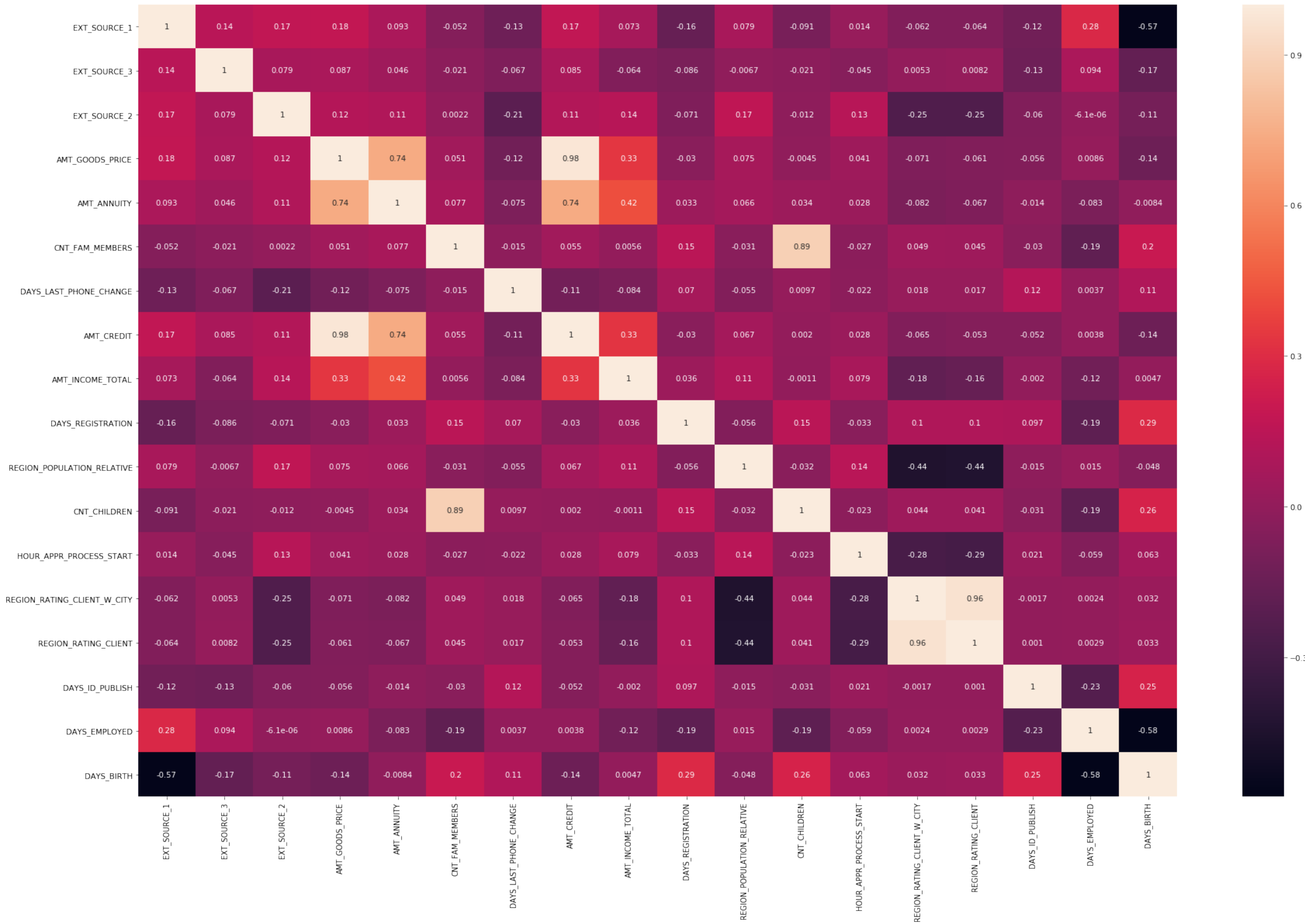


Maximum number of defaulters have Low annuity values, while maximum number of non-defaulters have high annuity

Bivariate Analysis on variables

LOAN APPLICATION ANALYSIS

Analysis on Defaulter Column



5 most Positive correlations:

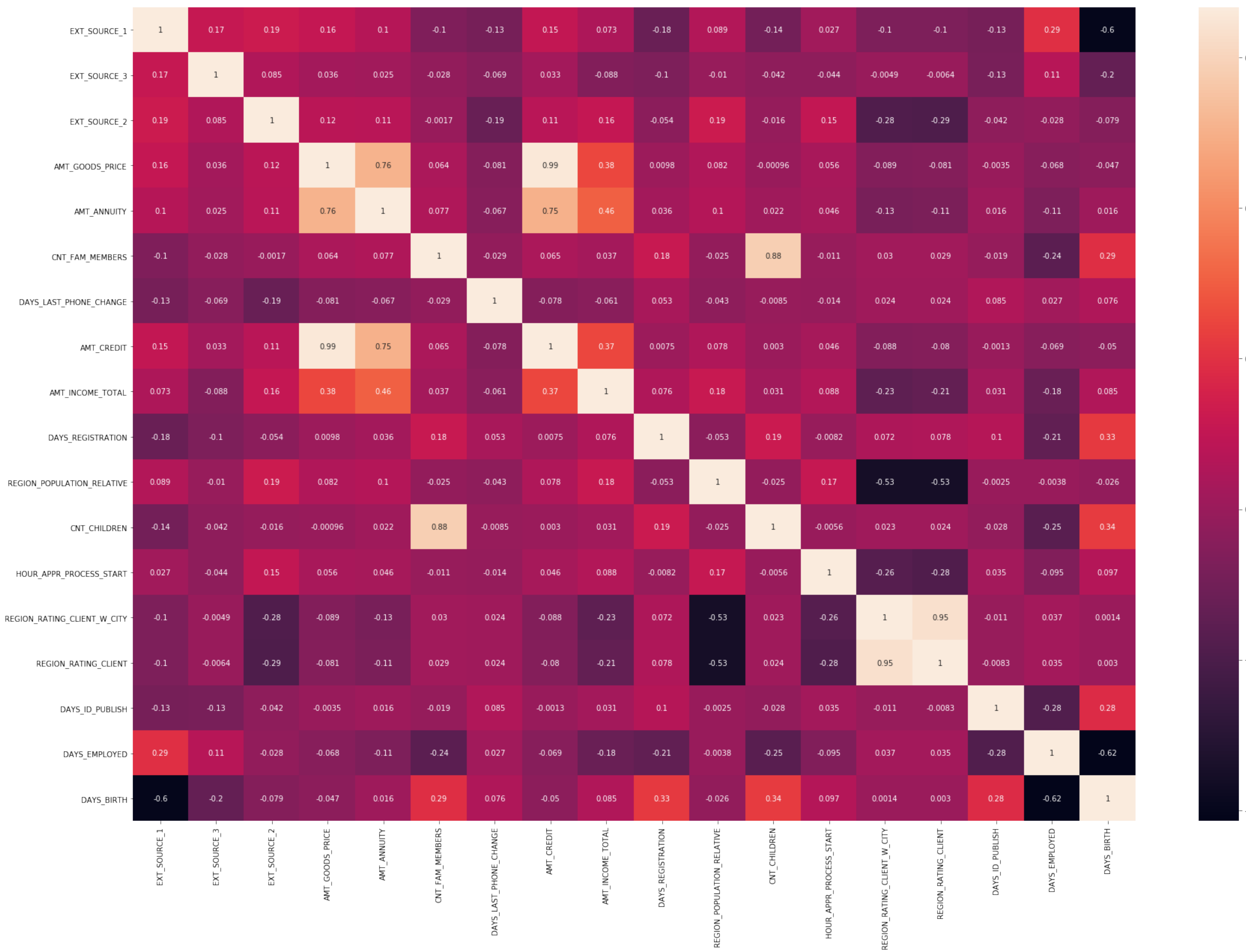
- AMT_CREDIT - AMT_GOODS_PRICE
- REGION_RATING_CLIENT_W_CITY - REGION_RATING_CLIENT
- CNT_CHILDREN - CNT_FAM_MEMBERS
- AMT_CREDIT - AMT_ANNUITY
- AMT_GOODS_PRICE - AMT_ANNUITY

5 most Negative correlations:

- HOUR_APPR_PROCESS_START - REGION_RATING_CLIENT_W_CITY
- REGION_RATING_CLIENT - HOUR_APPR_PROCESS_START
- REGION_POPULATION_RELATIVE - REGION_RATING_CLIENT
- REGION_RATING_CLIENT_W_CITY - REGION_POPULATION_RELATIVE
- EXT_SOURCE_1 - DAYS_BIRTH

LOAN APPLICATION ANALYSIS

Analysis on Non-Defaulter Column



5 most Positive correlations:

- AMT_CREDIT - AMT_GOODS_PRICE
- REGION_RATING_CLIENT_W_CITY - REGION_RATING_CLIENT
- CNT_CHILDREN - CNT_FAM_MEMBERS
- AMT_GOODS_PRICE - AMT_ANNUITY
- AMT_ANNUITY - AMT_CREDIT

Now that we have seen some analysis and insights from Applications Data data set Let’s now look at the previous application data set in the next slides

Analysis & Insights of Previous Applications Data

Data Analysis & Checks

Data Analysis

Let's check is there any null present in this dataset.

In [10]:
```
df_prevappln.isnull().values.any()
```

Out[10]:
```
True
```

So data set has **null** values. Let start the cleaning

Determine *count* and *percentage (%)* of **null** values in columns where **null** values are present

In [11]:
```
for column in df_prevappln.columns:
    if df_prevappln[column].isnull().any():
        print('{0} has {1} , {2} % null values'.format(column,
                                                       df_prevappln[column].isnull().sum(),
                                                       round(100*(df_prevappln[column].isnull().sum()/len(df_prevappln)))))
```

```
AMT_ANNUITY has 372235 , 22.0 % null values
AMT_CREDIT has 1 , 0.0 % null values
AMT_DOWN_PAYMENT has 895844 , 54.0 % null values
AMT_GOODS_PRICE has 385515 , 23.0 % null values
RATE_DOWN_PAYMENT has 895844 , 54.0 % null values
RATE_INTEREST_PRIMARY has 1664263 , 100.0 % null values
RATE_INTEREST_PRIVILEGED has 1664263 , 100.0 % null values
NAME_TYPE_SUITE has 820405 , 49.0 % null values
CNT_PAYMENT has 372230 , 22.0 % null values
PRODUCT_COMBINATION has 346 , 0.0 % null values
DAYS_FIRST_DRAWING has 673065 , 40.0 % null values
DAYS_FIRST_DUE has 673065 , 40.0 % null values
DAYS_LAST_DUE_1ST_VERSION has 673065 , 40.0 % null values
DAYS_LAST_DUE has 673065 , 40.0 % null values
DAYS_TERMINATION has 673065 , 40.0 % null values
NFLAG_INSURED_ON_APPROVAL has 673065 , 40.0 % null values
```

As you can see from above, there are columns which are almost or have 100% null values. Lets drop those columns.

Data Analysis

Data Cleaning

In [12]:
```
cols_to_drop = ['RATE_INTEREST_PRIMARY','RATE_INTEREST_PRIVILEGED']
df_prevappln.drop(cols_to_drop,axis=1,inplace=True)
```

Remove columns where number of unique value is only 1

In [13]:
```
unique = df_prevappln.nunique()
unique = unique[unique.values == 1]
```

In [14]:
```
df_prevappln.drop(labels = list(unique.index), axis =1, inplace=True)
```

Let us examine type of loans

In [15]:
```
df_contractType = pd.pivot_table(df_prevappln, index =['NAME_CONTRACT_TYPE'])
df_contractType
```

Out[15]:

	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	CNT_PAYMENT	DAYS_DECISION	DA
NAME_CONTRACT_TYPE								
Cash loans	26119.82	274760.43	304061.49	0.00	443072.95	28.35	-571.82	
Consumer loans	10128.05	93787.83	91524.59	7152.16	93962.70	11.21	-1249.57	
Revolving loans	11384.79	97816.24	173505.60	3260.82	202849.07	0.00	-684.36	
XNA	nan	0.00	0.00	nan	nan	nan	-415.49	

As you can see from above that there is contract type *XNA* for which there are columns which have **most nan values**.
Remove all those rows where contract type is *XNA*

In [16]:
```
df_prevappln.drop(df_prevappln.loc[df_prevappln['NAME_CONTRACT_TYPE']=='XNA'].index, inplace=True)
```

LOAN APPLICATION ANALYSIS

In [17]:
```
# Recheck
df_contractType = pd.pivot_table(df_prevappln, index =['NAME_CONTRACT_TYPE'])
df_contractType
```

Out[17]:

NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	CNT_PAYMENT	DAYS_DECISION	DAY
Cash loans	26119.82	274760.43	304061.49	0.00	443072.95	28.35	-571.82	
Consumer loans	10128.05	93787.83	91524.59	7152.16	93962.70	11.21	-1249.57	
Revolving loans	11384.79	97816.24	173505.60	3260.82	202849.07	0.00	-684.36	

Drop columns which have more than 30% null values

In [18]:
```
threshold = 0.3 # 30%
col_names = list(df_prevappln.columns[(df_prevappln.isnull().sum()/len(df_prevappln)) > threshold])
df_prevappln.drop(labels = col_names,axis =1,inplace=True)
print("Number of Columns dropped\t: ",len(col_names))
print("List of Columns dropped\t: ",col_names)
```

```
Number of Columns dropped      :  9
List of Columns dropped :  ['AMT_DOWN_PAYMENT', 'RATE_DOWN_PAYMENT', 'NAME_TYPE_SUITE', 'DAYS_FIRST_DRAWING', 'DAYS_FIRST_DUE',
'DAYS_LAST_DUE_1ST_VERSION', 'DAYS_LAST_DUE', 'DAYS_TERMINATION', 'NFLAG_INSURED_ON_APPROVAL']
```

Check which columns have nulls/nan values

In [19]:
```
print(round(100*(df_prevappln.isnull().sum()/len(df_prevappln))))
```

```
SK_ID_PREV                      0.00
SK_ID_CURR                      0.00
NAME_CONTRACT_TYPE              0.00
AMT_ANNUITY                    22.00
AMT_APPLICATION                 0.00
AMT_CREDIT                      0.00
AMT_GOODS_PRICE                23.00
WEEKDAY_APPR_PROCESS_START      0.00
HOUR_APPR_PROCESS_START         0.00
FLAG_LAST_APPL_PER_CONTRACT     0.00
NFLAG_LAST_APPL_IN_DAY          0.00
NAME_CASH_LOAN_PURPOSE          0.00
NAME_CONTRACT_STATUS            0.00
DAYS_DECISION                   0.00
NAME_PAYMENT_TYPE               0.00
CODE_REJECT_REASON              0.00
NAME_CLIENT_TYPE                0.00
NAME_GOODS_CATEGORY             0.00
NAME_PORTFOLIO                  0.00
NAME_PRODUCT_TYPE               0.00
CHANNEL_TYPE                    0.00
SELLERPLACE_AREA                0.00
NAME_SELLER_INDUSTRY            0.00
CNT_PAYMENT                    22.00
NAME_YIELD_GROUP                0.00
PRODUCT_COMBINATION             0.00
dtype: float64
```

That leaves us with 3 columns **(AMT_ANNUITY,AMT_GOODS_PRICE,CNT_PAYMENT)** where we still have null or nan values. Referring to the data dictionary we know that -
AMT_ANNUITY is Annuity of previous application
AMT_GOODS_PRICE is Goods price of good that client asked for (if applicable) on the previous application
CNT_PAYMENT is the Term of previous credit of the previous application
Let's analyse these columns and check if we can impute values

In [20]:
```
df_prevappln.loc[:, ['AMT_ANNUITY', 'AMT_GOODS_PRICE','CNT_PAYMENT']].describe()
```

Out[20]:

	AMT_ANNUITY	AMT_GOODS_PRICE	CNT_PAYMENT
count	1297979.00	1284699.00	1297984.00
mean	15955.12	227847.28	16.05
std	14782.14	315396.56	14.57
min	0.00	0.00	0.00
25%	6321.78	50841.00	6.00
50%	11250.00	112320.00	12.00
75%	20658.42	234000.00	24.00
max	418058.15	6905160.00	84.00

Imputation and Reasons/Explanation

AMT_ANNUITY : Impute the value of this column to the median as that is the lowest(compared between mean and median)

In [21]:
```
df_prevappln.loc[np.isnan(df_prevappln['AMT_ANNUITY']), ['AMT_ANNUITY']] = df_prevappln['AMT_ANNUITY'].median()
```

AMT_GOODS_PRICE : Impute the value of this column to 0 as we cannot say from the given the data if the client asked for goods price on the prev application

In [22]:
```
df_prevappln.loc[np.isnan(df_prevappln['AMT_GOODS_PRICE']), ['AMT_GOODS_PRICE']] = 0
```

Check unique values for each value in remaining column

CNT_PAYMENT : Impute the value of this column to median(50%) value as from above we can see most number (approx 20%) of term applications are closer or equal to median(50%) value

In [24]:
```
df_prevappln.loc[np.isnan(df_prevappln['CNT_PAYMENT']), ['CNT_PAYMENT']] = df_prevappln['CNT_PAYMENT'].median()
```

Re-check nulls / nan values

In [25]:
```
print(round(100*(df_prevappln.isnull().sum()/len(df_prevappln))))
```

```
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               0.00
WEEKDAY_APPR_PROCESS_START    0.00
HOUR_APPR_PROCESS_START       0.00
FLAG_LAST_APPL_PER_CONTRACT   0.00
NFLAG_LAST_APPL_IN_DAY        0.00
NAME_CASH_LOAN_PURPOSE        0.00
NAME_CONTRACT_STATUS          0.00
DAYS_DECISION                 0.00
NAME_PAYMENT_TYPE             0.00
CODE_REJECT_REASON            0.00
NAME_CLIENT_TYPE              0.00
NAME_GOODS_CATEGORY           0.00
NAME_PORTFOLIO                0.00
NAME_PRODUCT_TYPE             0.00
CHANNEL_TYPE                  0.00
SELLERPLACE_AREA              0.00
NAME_SELLER_INDUSTRY          0.00
CNT_PAYMENT                   0.00
NAME_YIELD_GROUP              0.00
PRODUCT_COMBINATION           0.00
dtype: float64
```

In [26]:
```
new_rowcount=df_prevappln.shape[0]
new_colcount=df_prevappln.shape[1]
print("We are now left with -> ",(new_rowcount,new_colcount) ,
      "rows & columns as compared to original ->",(org_rowcount,org_colcount))
```

```
We are now left with ->  (1669868, 26) rows & columns as compared to original -> (1670214, 37)
```

Now that we have some relevant & cleaned data for further analysis, let us go through the attributes and start making observations.

Derived Metrics

Derived Metrics for Visualisation

```
In [4]: # weekdays as a dictionary (considering Monday as first working day of the week)
weekDays = {'MONDAY':1,'TUESDAY':2,'WEDNESDAY':3,'THURSDAY':4,'FRIDAY':5,'SATURDAY':6,'SUNDAY':7}
```

Derived Metrics for Visualisation

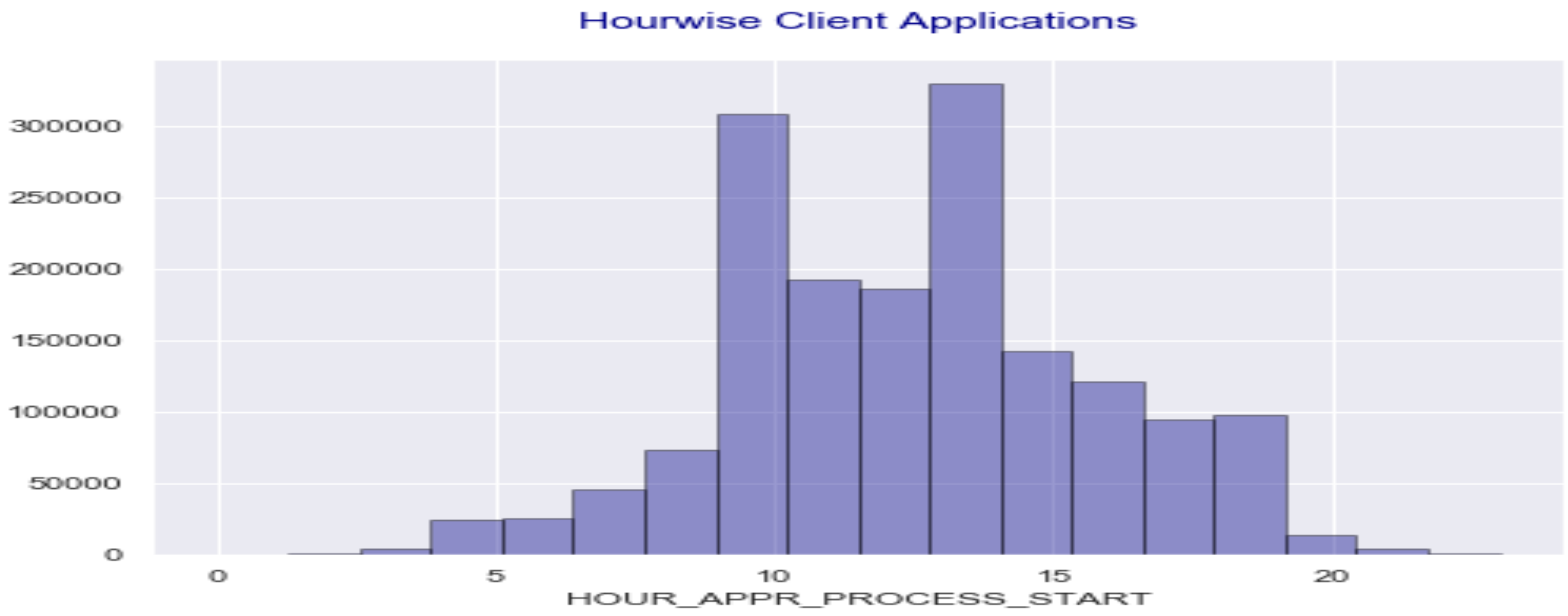
```
In [29]: df_prevappln['DAY_OF_WEEK'] = df_prevappln['WEEKDAY_APPR_PROCESS_START'].apply(lambda day : weekDays[day])
```

Univariate Analysis

LOAN APPLICATION ANALYSIS

Analysis

Variable: HOUR_APPR_PROCESS_START



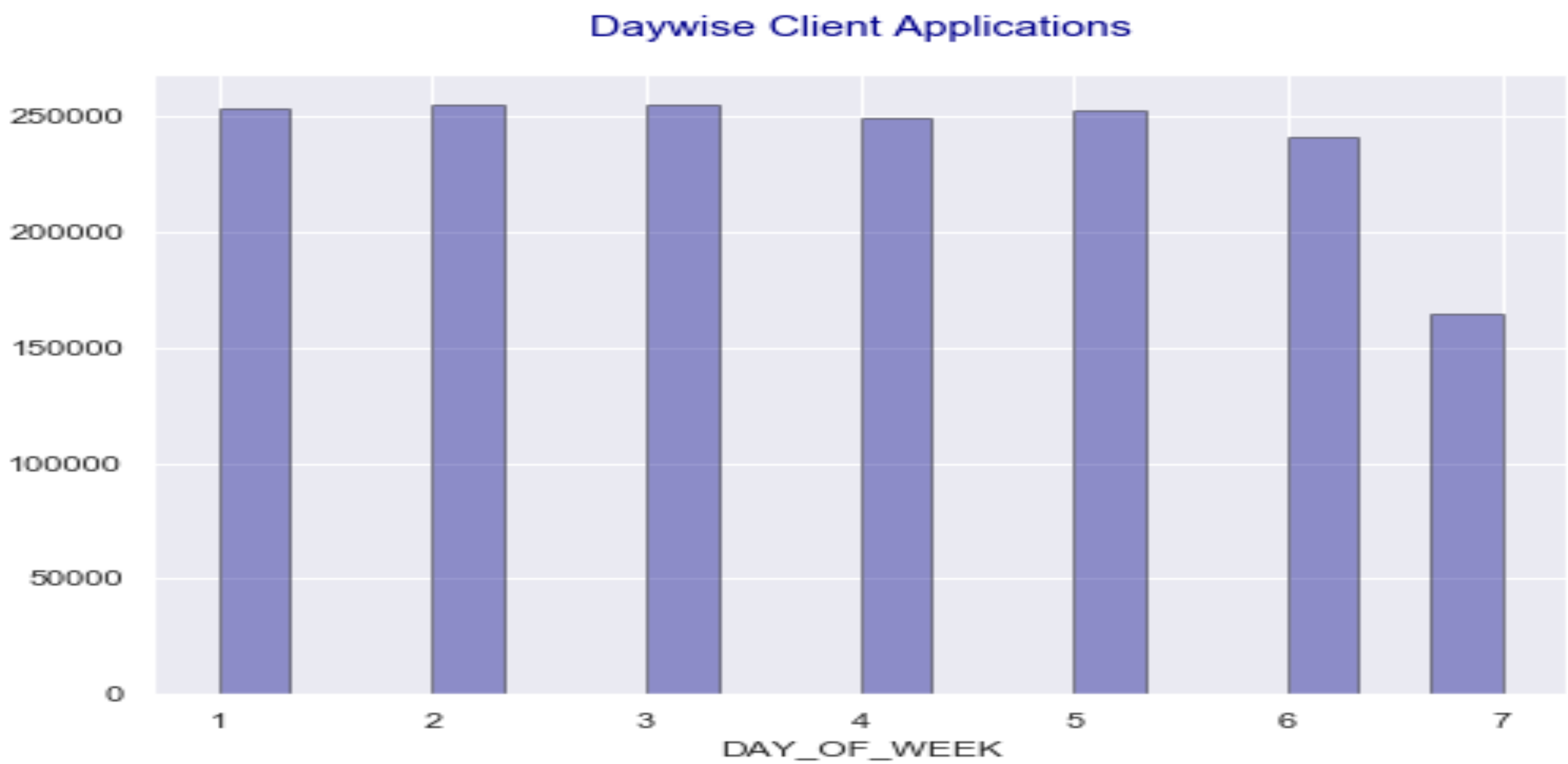
Insights :
Most of the applications are processed during noon.
Most of the applications are processed during work hours.

LOAN APPLICATION ANALYSIS

Analysis

Variable: DAY_OF_WEEK

DERIVED METRIC from variable WEEKDAY_APPR_PROCESS_START



Insights :

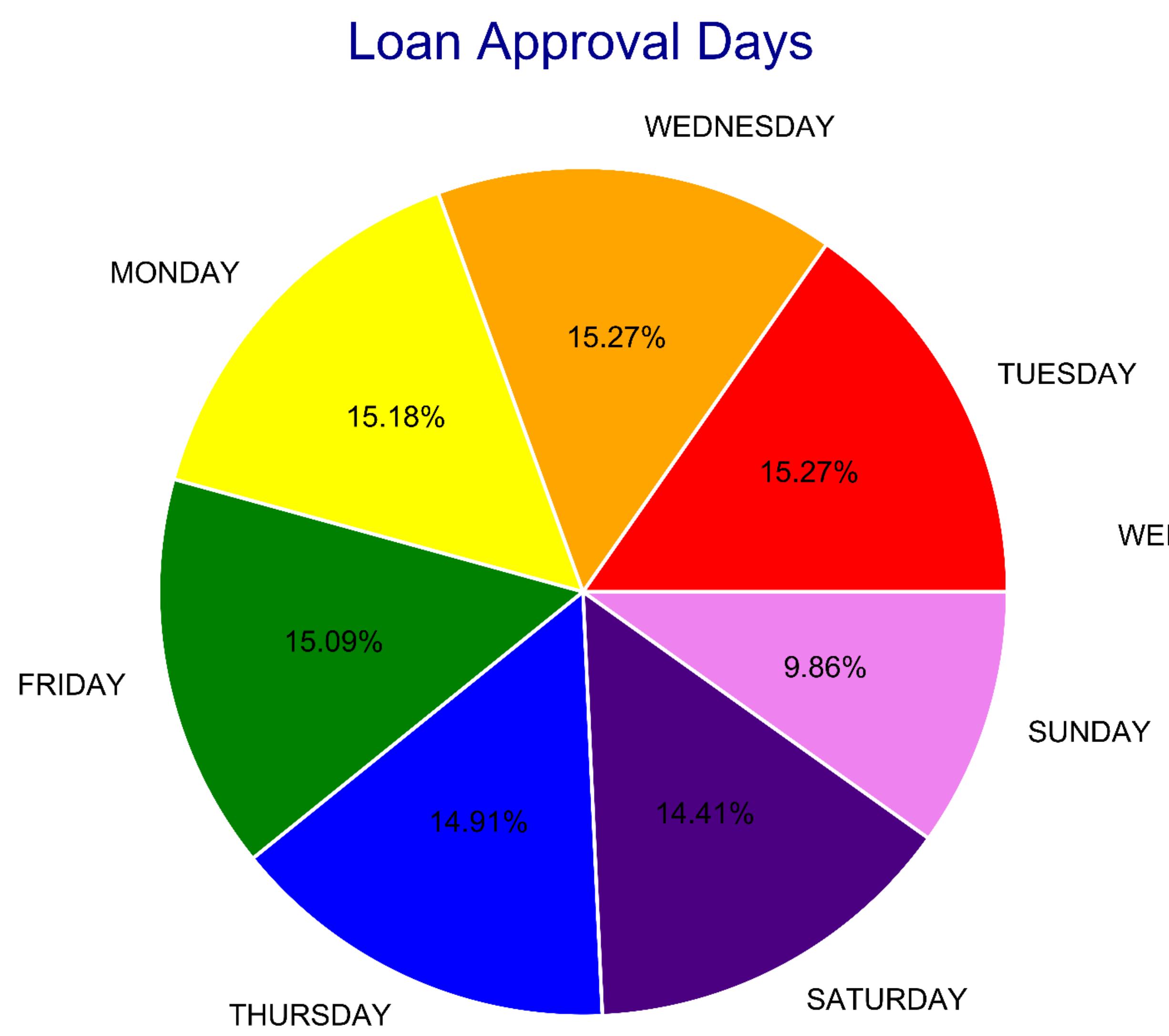
Client has applied for loan amount applications almost on all days except 7th day (Sunday).

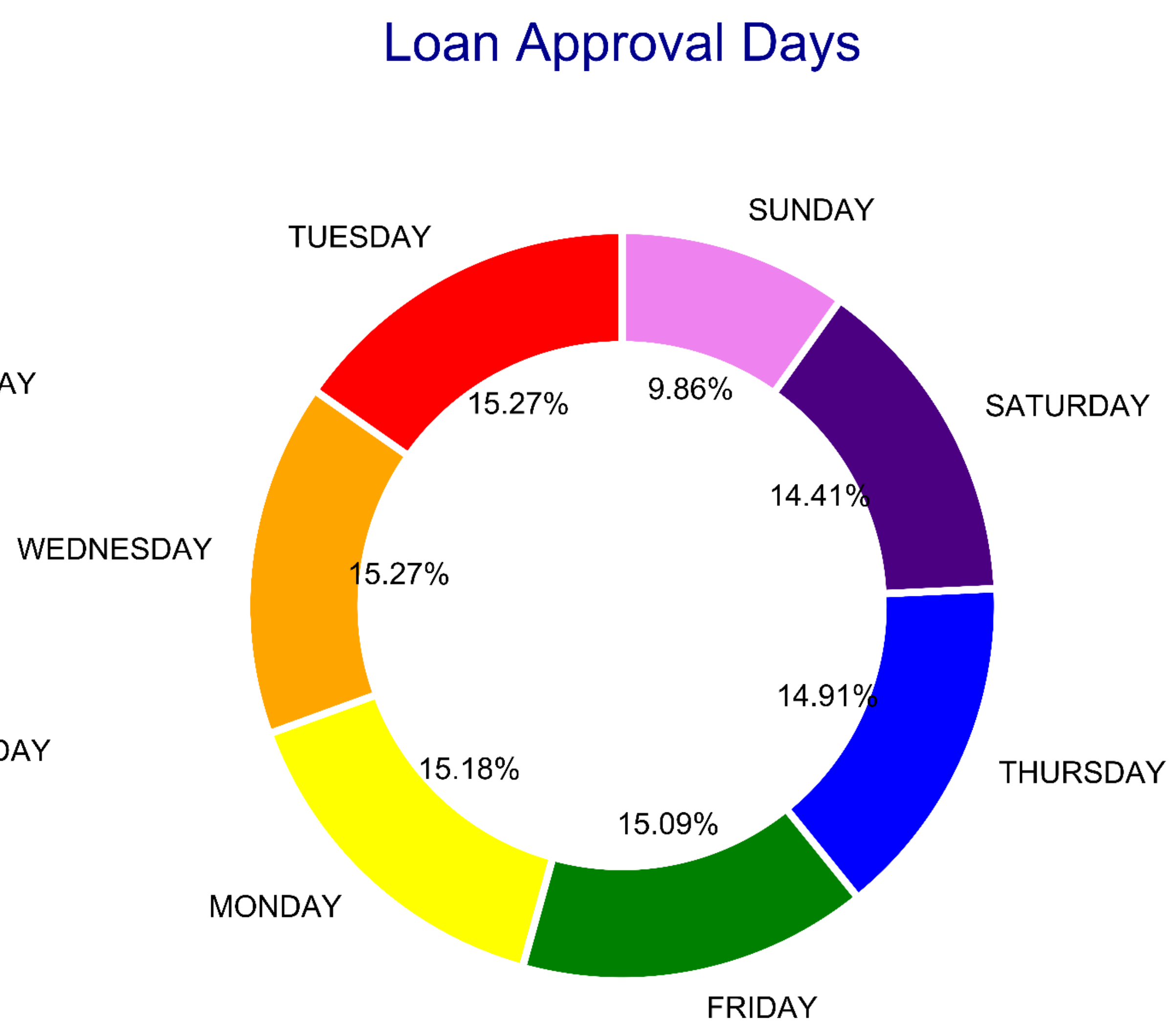
Loan providing companies always have high number of requests. Almost heavy rush of requests is seen on almost every weekday

LOAN APPLICATION ANALYSIS

Analysis

Variable: WEEKDAY_APPR_PROCESS_START



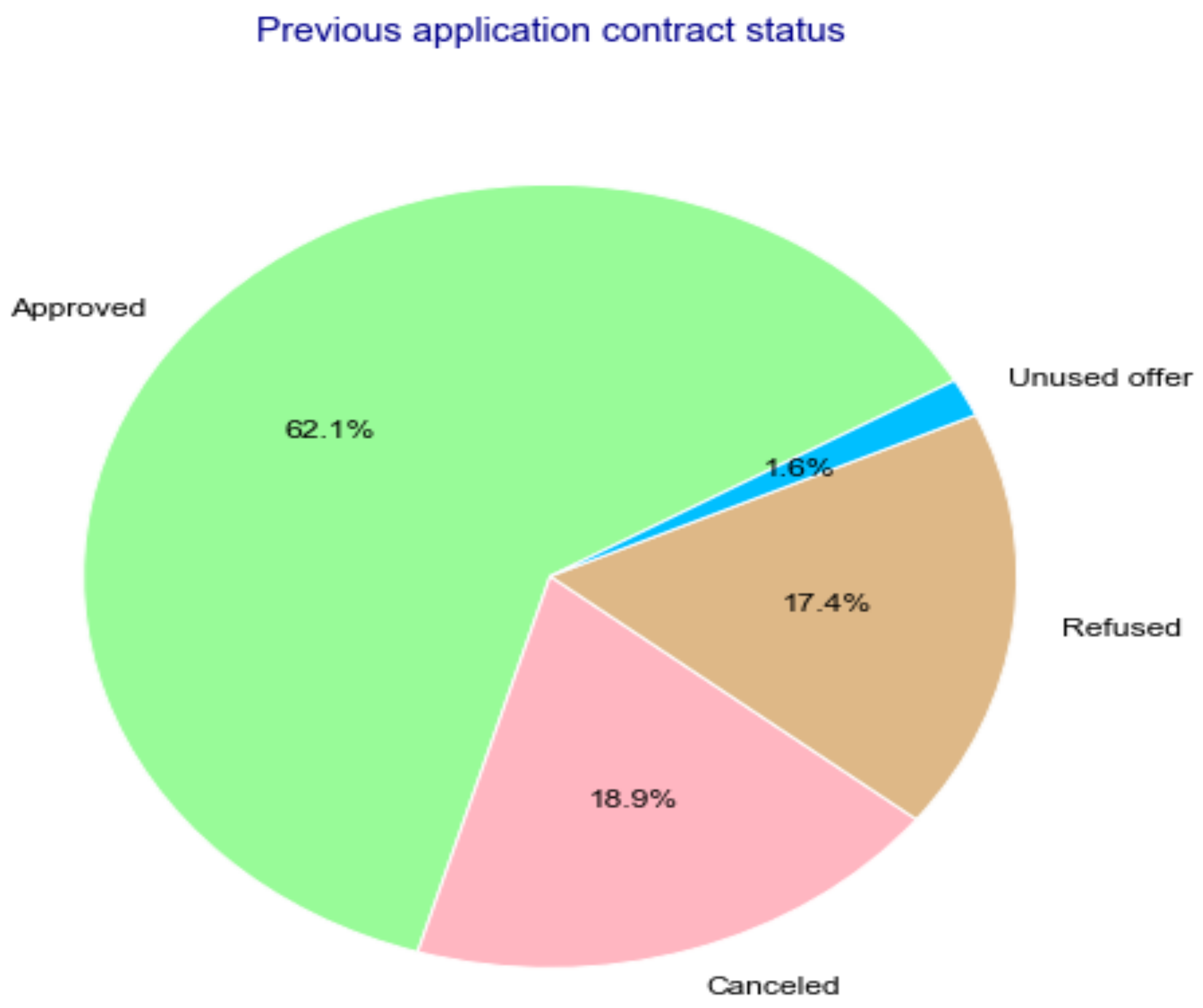


Insights:
Most of the previous loan applications are processed between Monday to Wednesday(1st 3 days of the week)

LOAN APPLICATION ANALYSIS

Analysis

Variable: NAME_CONTRACT_STATUS

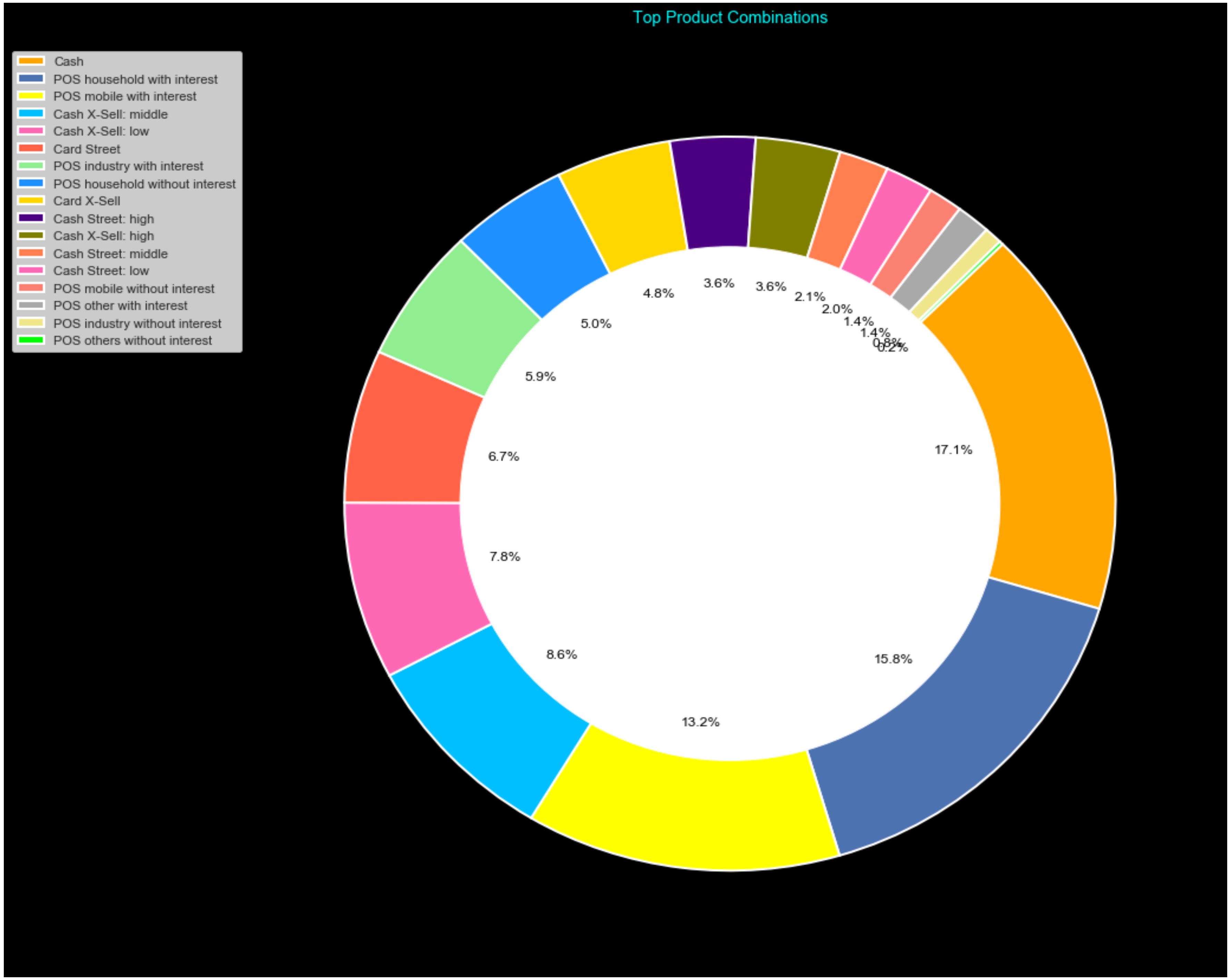


Insights:
High percentage of the previous loan applications were approved

LOAN APPLICATION ANALYSIS

Analysis

Variable: PRODUCT_COMBINATION



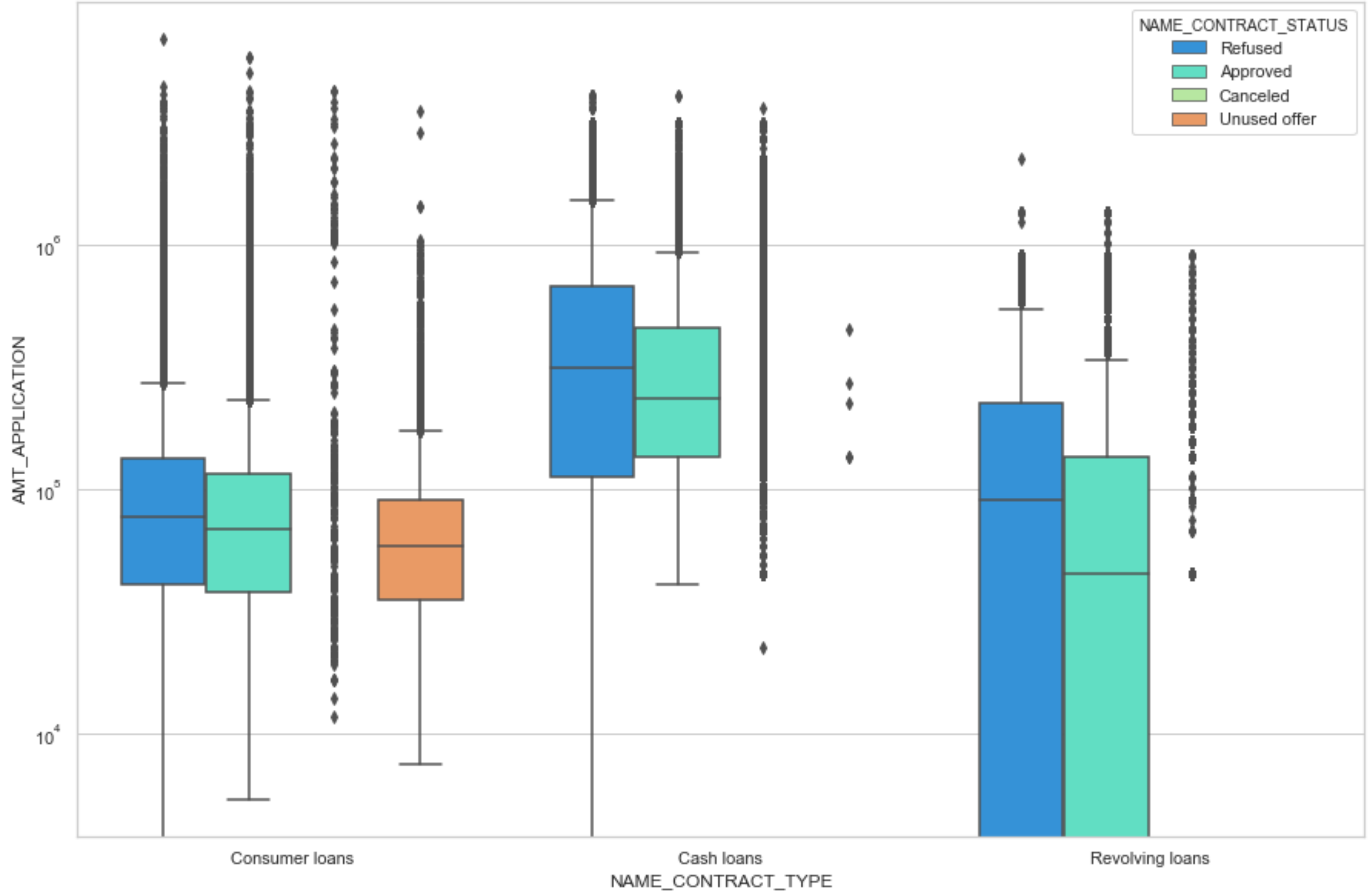
Insights:
Cash is the top product combination for previous loan applications

Bivariate Analysis

LOAN APPLICATION ANALYSIS

Analysis

Variable: NAME_CONTRACT_TYPE and AMT_APPLICATION

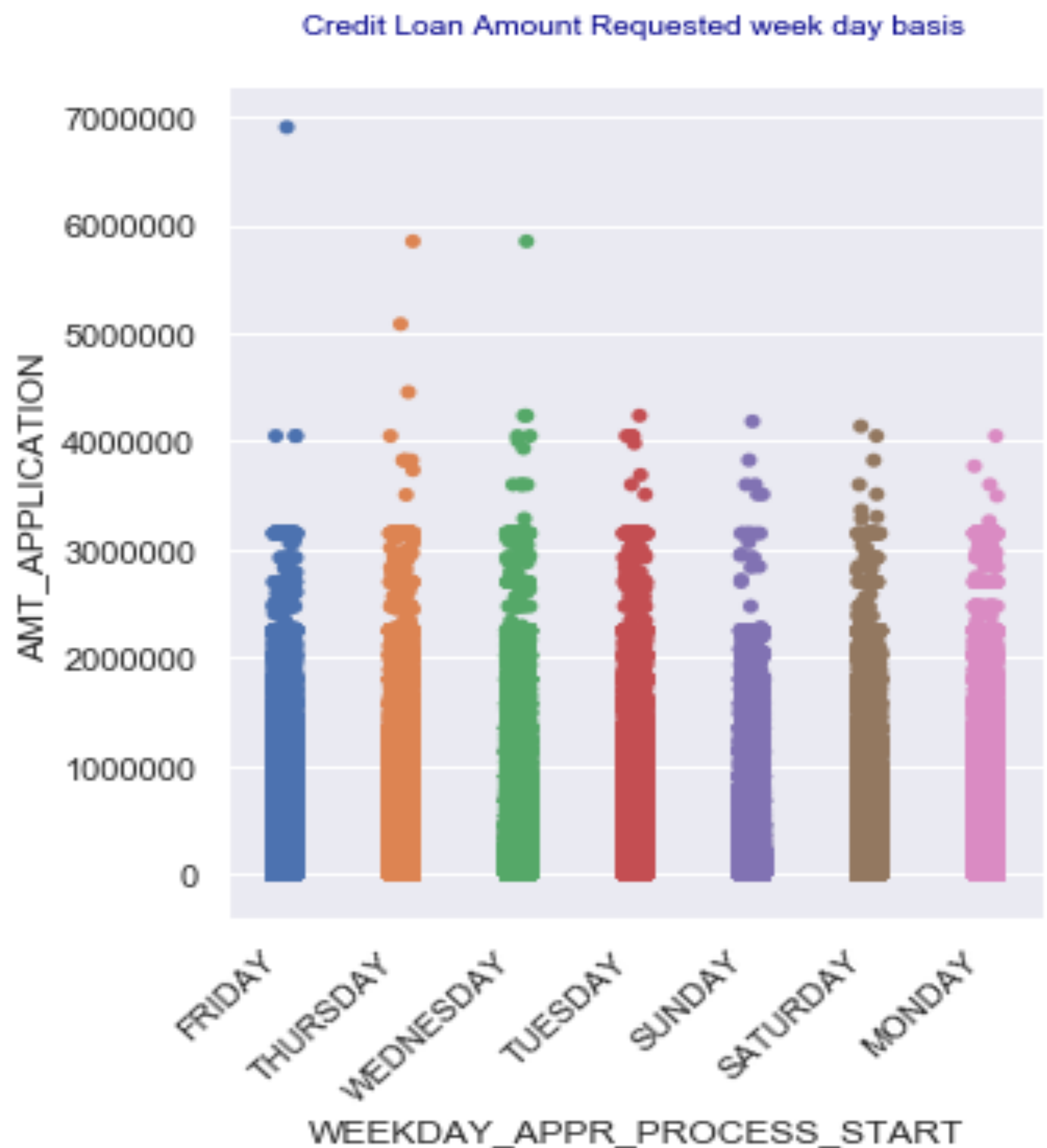


Insights:
Cash Loans are the most refused as compared to other loans.
There are lot of unused offers on the Consumer Loans.
There are no cancellations or unused offers on the Cash and Revolving loans

LOAN APPLICATION ANALYSIS

Analysis

Variable: WEEKDAY_APPR_PROCESS_START and AMT_APPLICATION



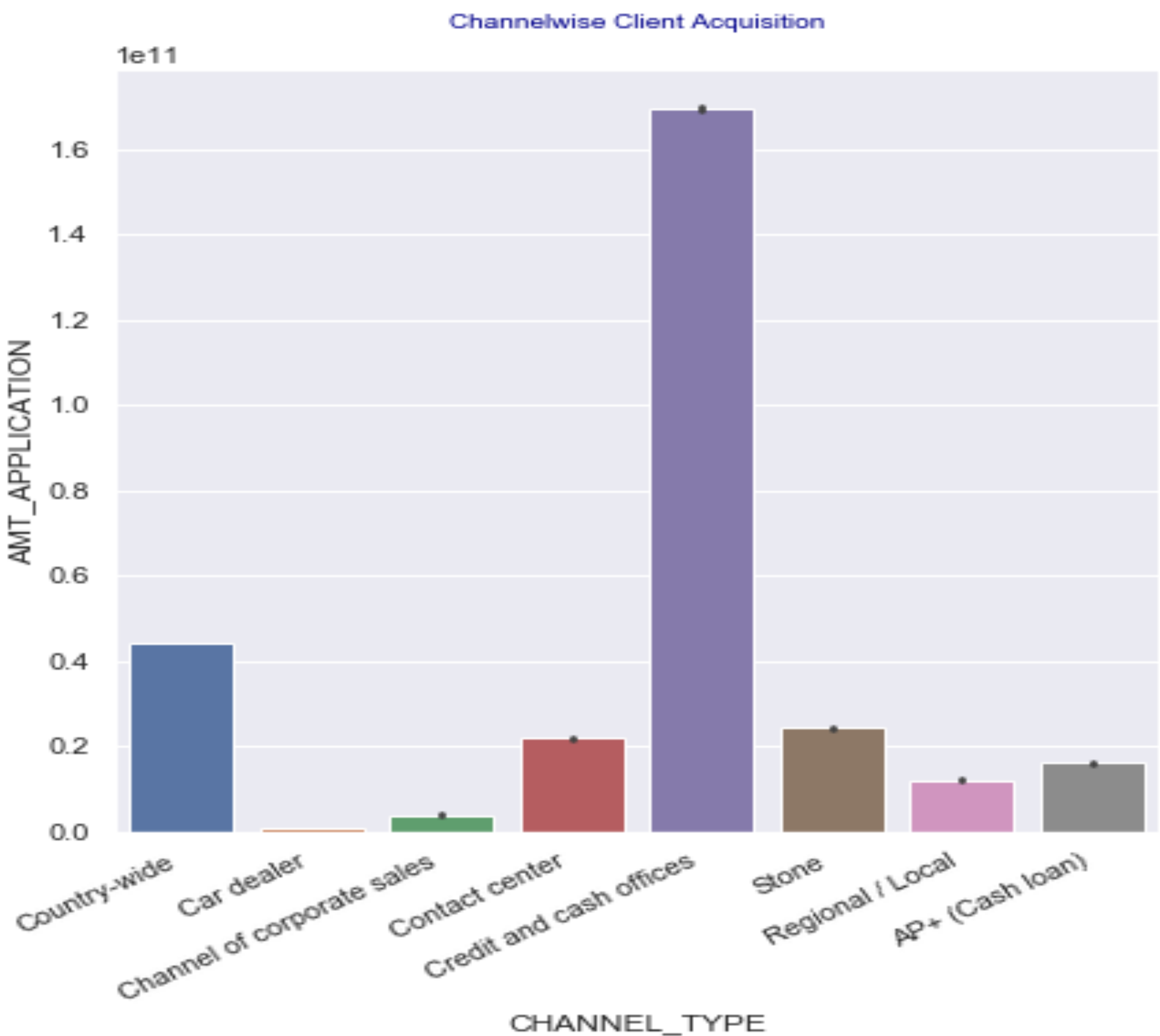
Insights:

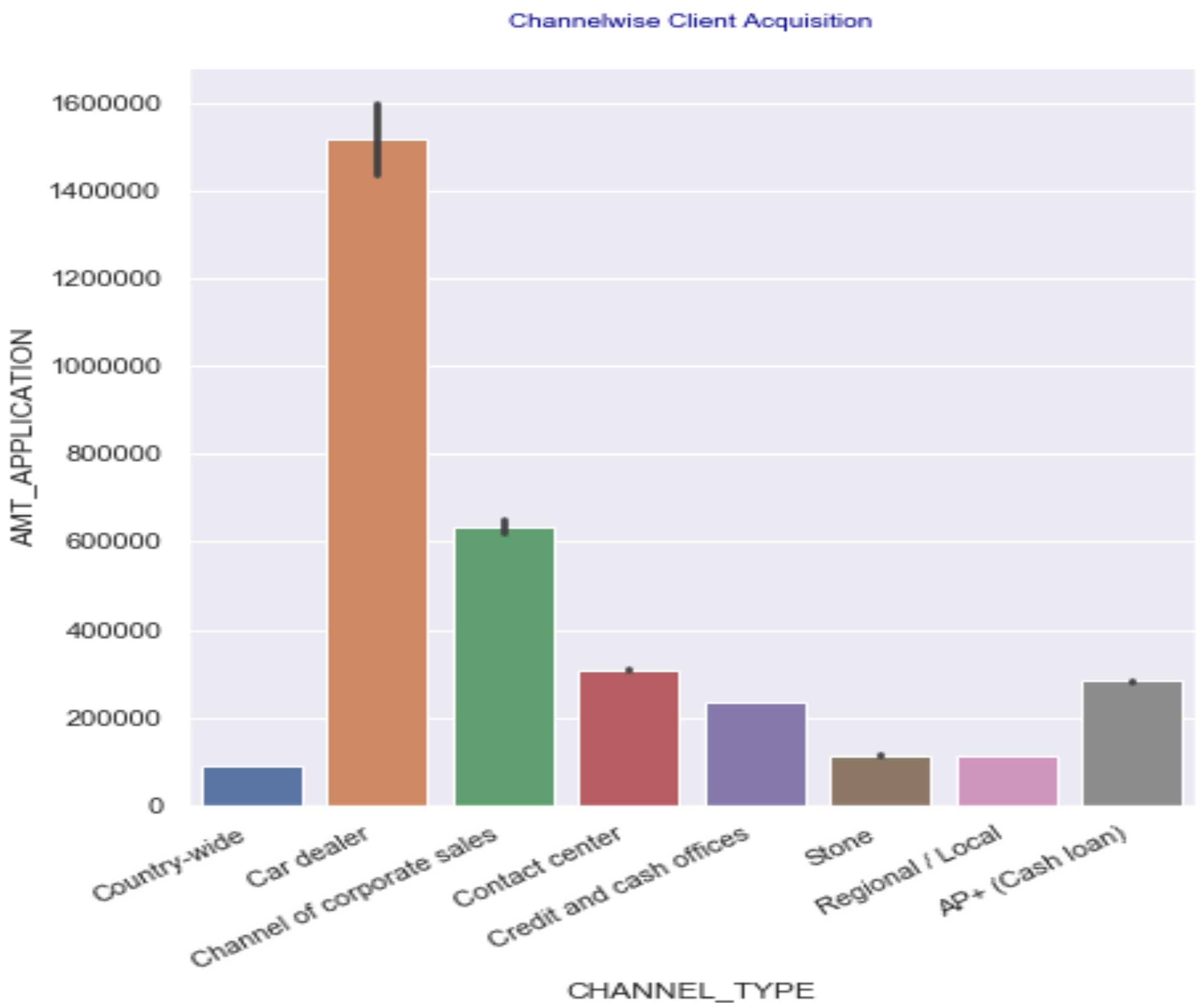
Loan Applications which had high amounts are requested towards the end of the week by clients - between Wed to Fri

LOAN APPLICATION ANALYSIS

Analysis

Variable: CHANNEL_TYPE and AMT_APPLICATION

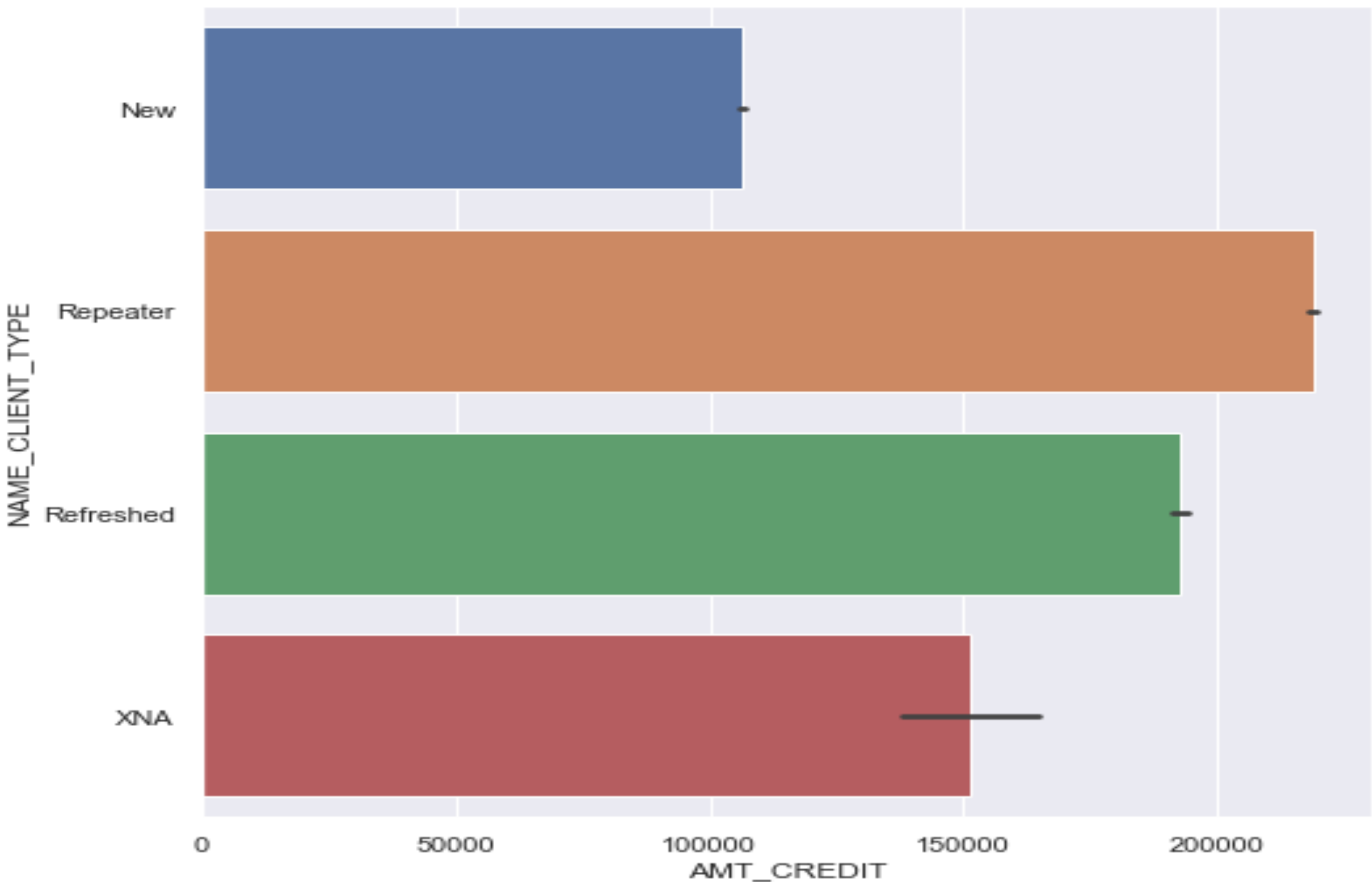




Insights :
Considering the sum of loan amounts asked,clients were acquired the most on through the credit and cash offices channel followed by country wide.
However, considering the average loan amount asked the clients were acquired the most on through the car dealership channel followed by corporate sales

Analysis

Variable: AMT_CREDIT and NAME_CLIENT_TYPE



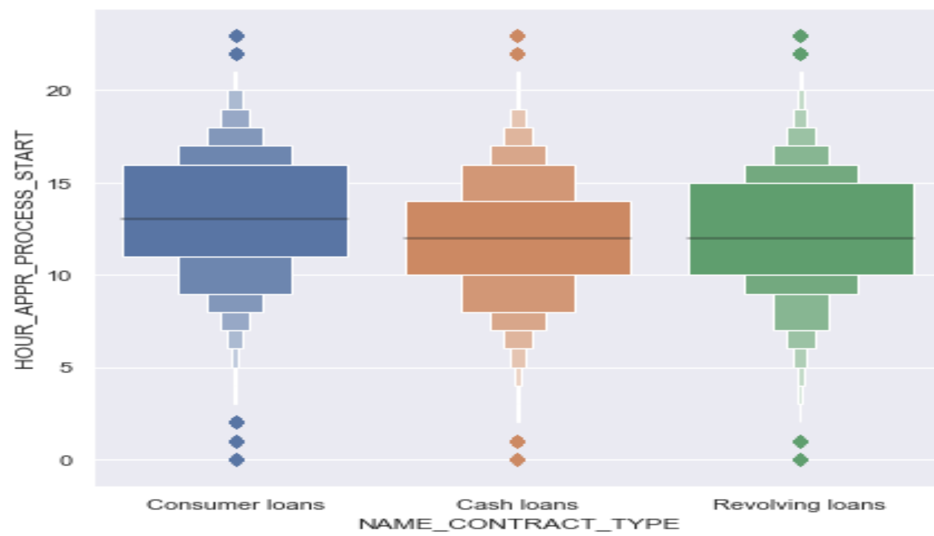
Insights:

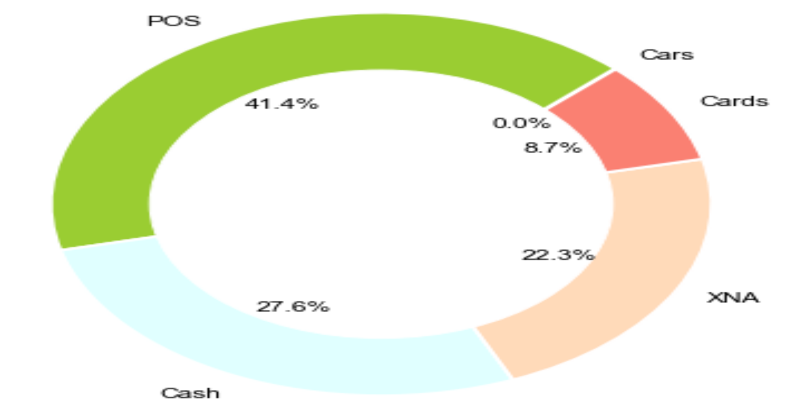
Many people struggle to get loans due to insufficient credit histories. From above it is seen that loan companies have credited loan amounts to repeaters(old) clients as they have credit history. So old clients are the more favored above others to provide loans by the loan comapnies.

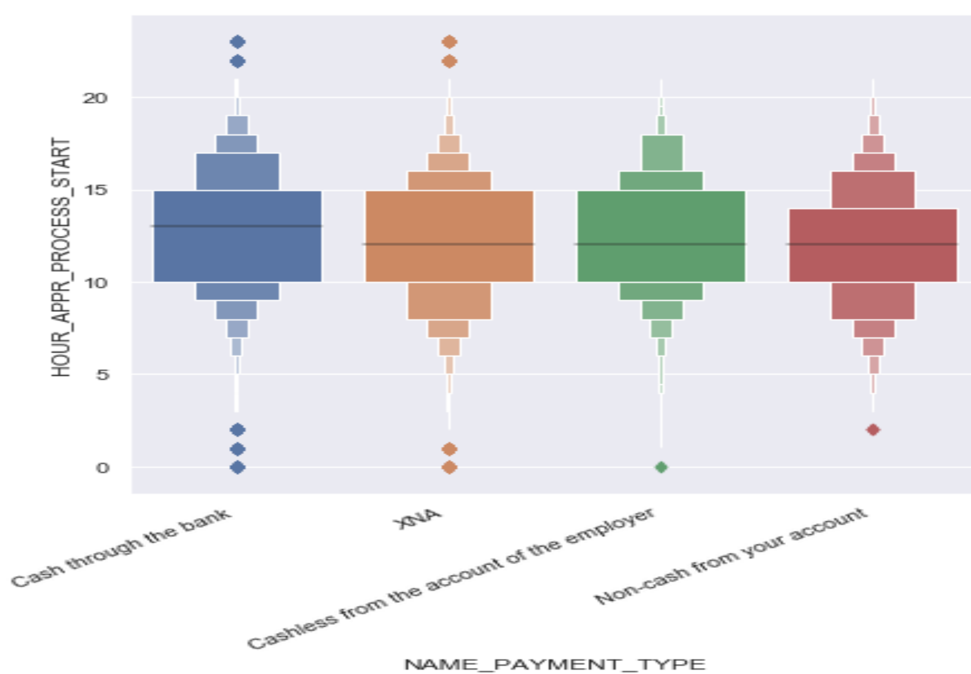
It is also seen that loans are credited to clients for whom the data is not available with the Loan companies(Client Type=XNA). This could be a FRAUD or may some error in th system due to which client data is not captured. It is necessary that the loan companies focus on capturing client data while disbursing loans else there will be huge effort required to recover the loan amount.

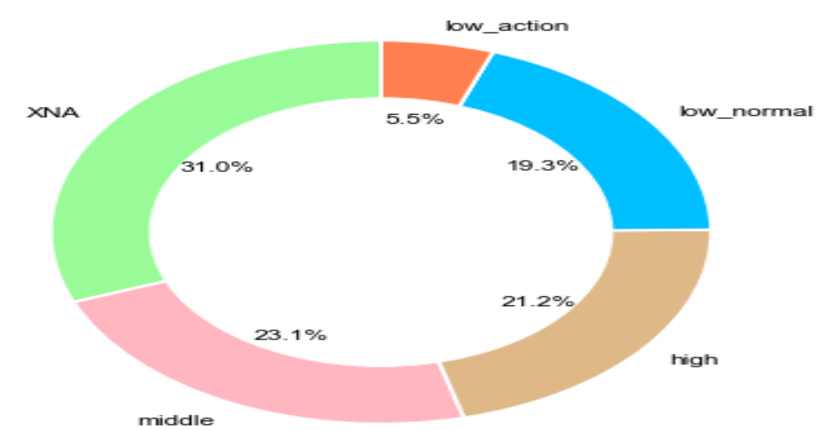
LOAN APPLICATION ANALYSIS

Analysis









Insights :

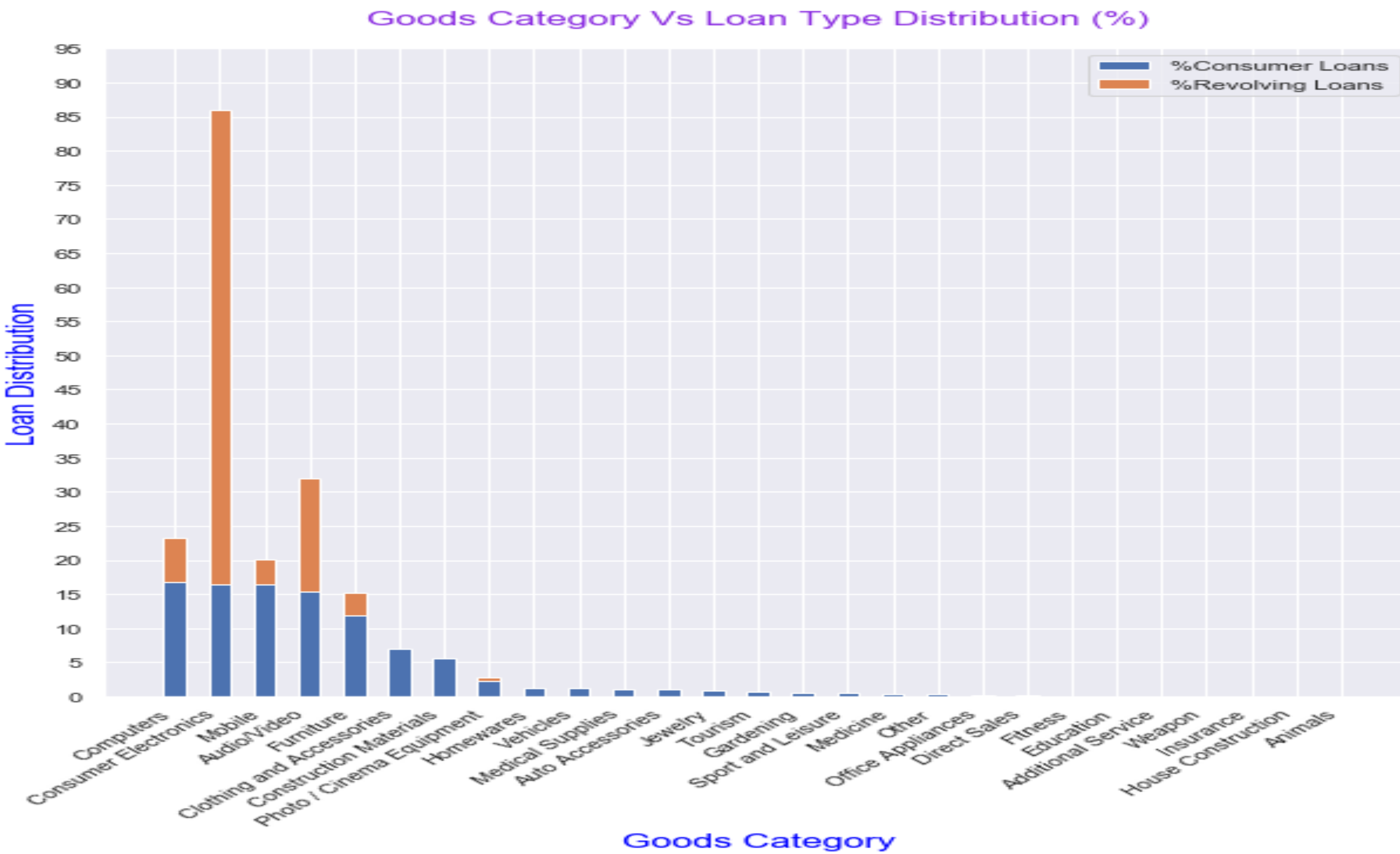
Consumer loans are the highly offered loans and are less likely to default

Mostly the previous applications were POS followed by Cash

On an average 'Cash through the bank' is most used payment method used by the clients for paying loan amounts followed by cashless transaction. There is a need to educate clients and improve security loan payment facilities so that more and more clients use cashless mode.

LOAN APPLICATION ANALYSIS

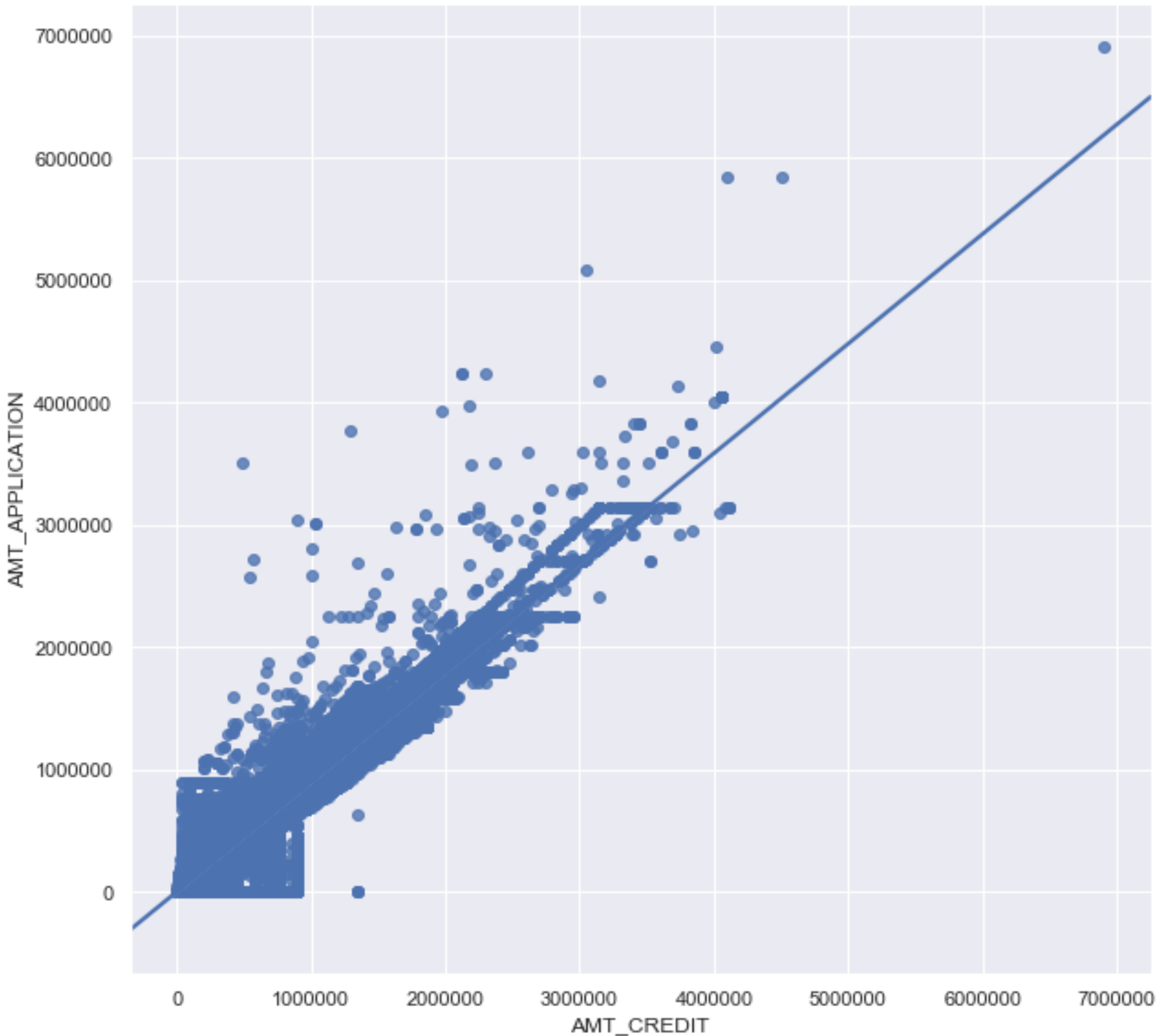
Analysis



Insights : Revolving Loans are <u>more (approx. 70%)</u> for Consumer Electronics than any other goods category

LOAN APPLICATION ANALYSIS

Analysis



Insights :
Following columns are highly correlated
1.AMT_ANNUITY
2.AMT_APPLICATION
3.AMT_CREDIT
4.AMT_DOWN_PAYMENT
5.AMT_GOODS_PRICE

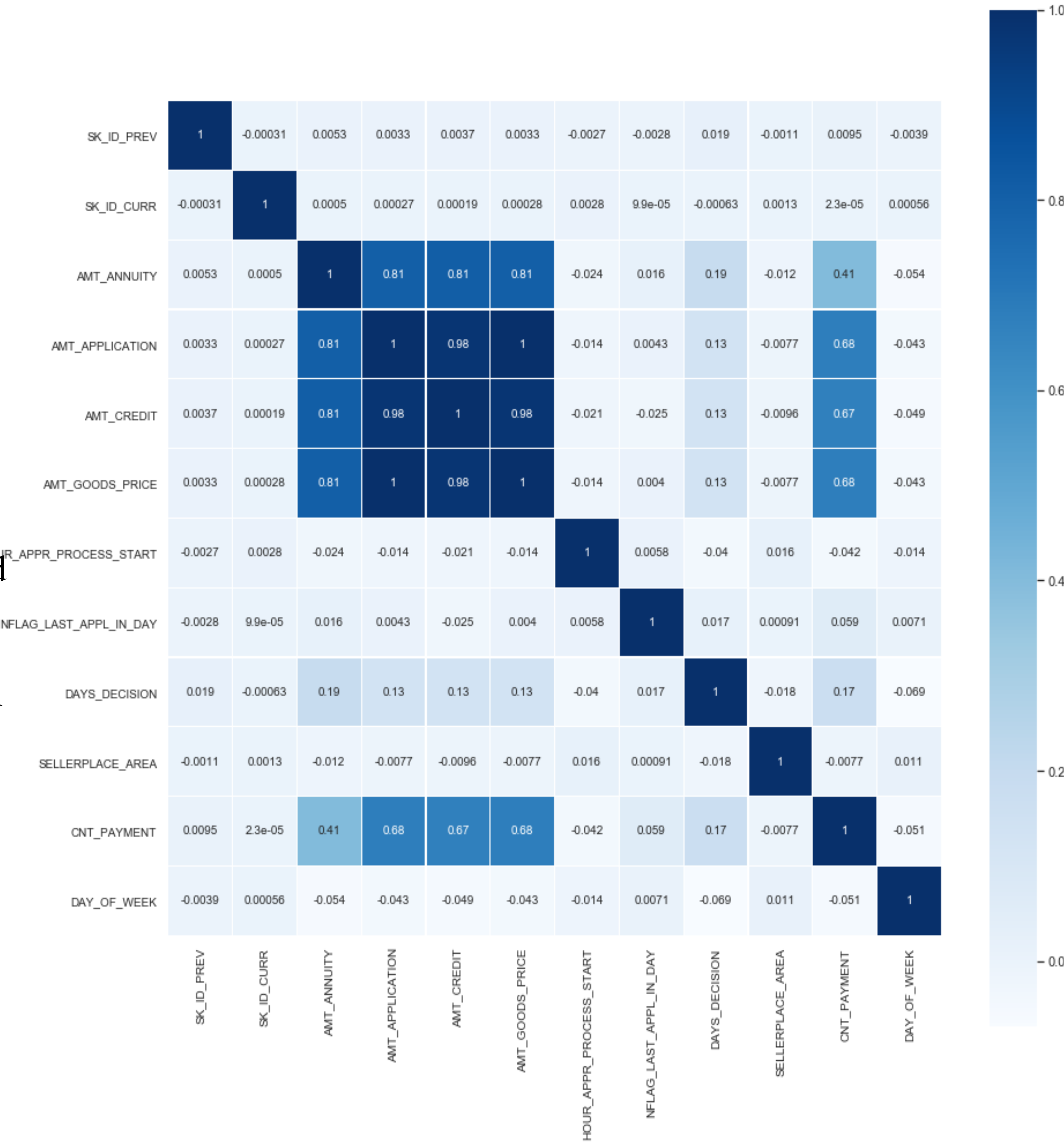
Loan Amount requested by client orignally and final credit loan ammount on the previous application given to the client are highly correlated as seen from above scatter and heat maps. This means the loan companies are providing almost closest amount of for which the client has made an initial request for.
We have already seen earlier that consumer loans are the more offered than other loans

LOAN APPLICATION ANALYSIS

Analysis

Insights :
Following columns are highly correlated
1.AMT_ANNUITY
2.AMT_APPLICATION
3.AMT_CREDIT
4.AMT_DOWN_PAYMENT
5.AMT_GOODS_PRICE

Loan Amount requested by client orignally and final credit loan ammount on the previous application given to the client are highly correlated as seen from above scatter and heat maps. This means the loan companies are providing almost closest amount of for which the client has made an initial request for.
We have already seen earlier that consumer loans are the more offered than other loans



Thank You