Tanzeel Iqbal

Credit EDA Assignment

Problem Statement - I

Business Understanding

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 to 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 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.

Problem Statement - II

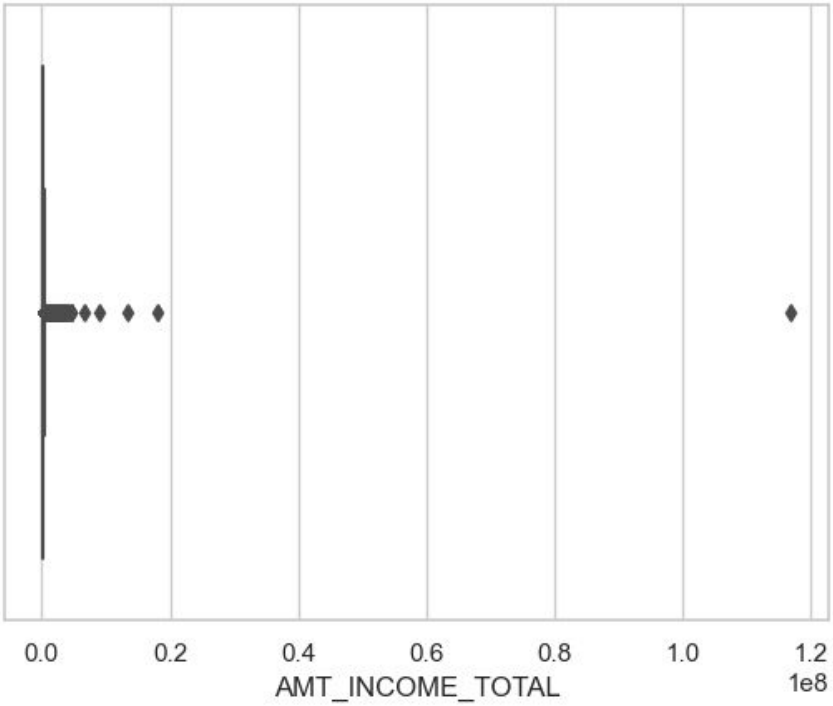
Present the overall approach of the analysis in a presentation. Mention the problem statement and the analysis approach briefly.

Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value)

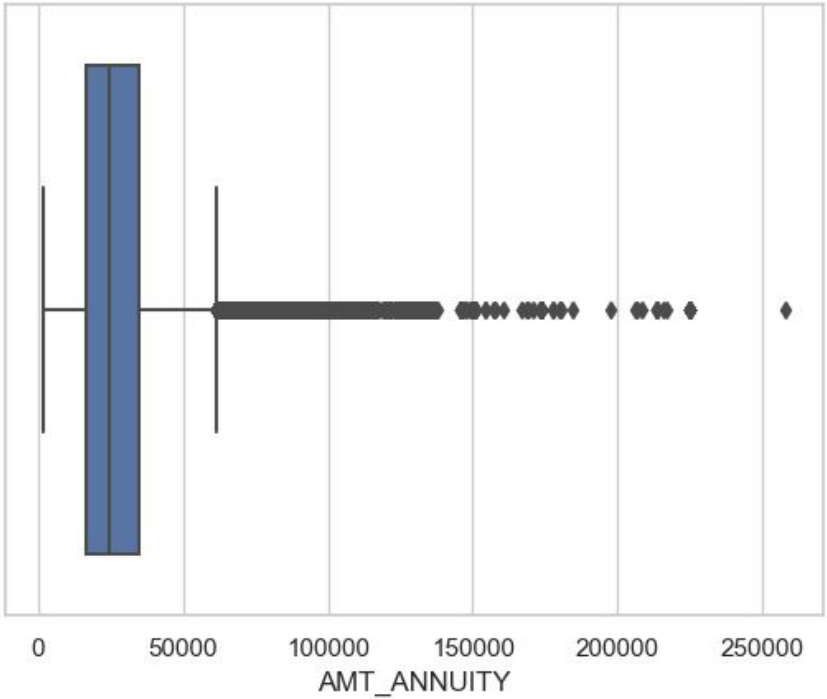
Outlier Analysis

These are highly deviated values from our dense region of data. These need to be analysed and dropped if needed.

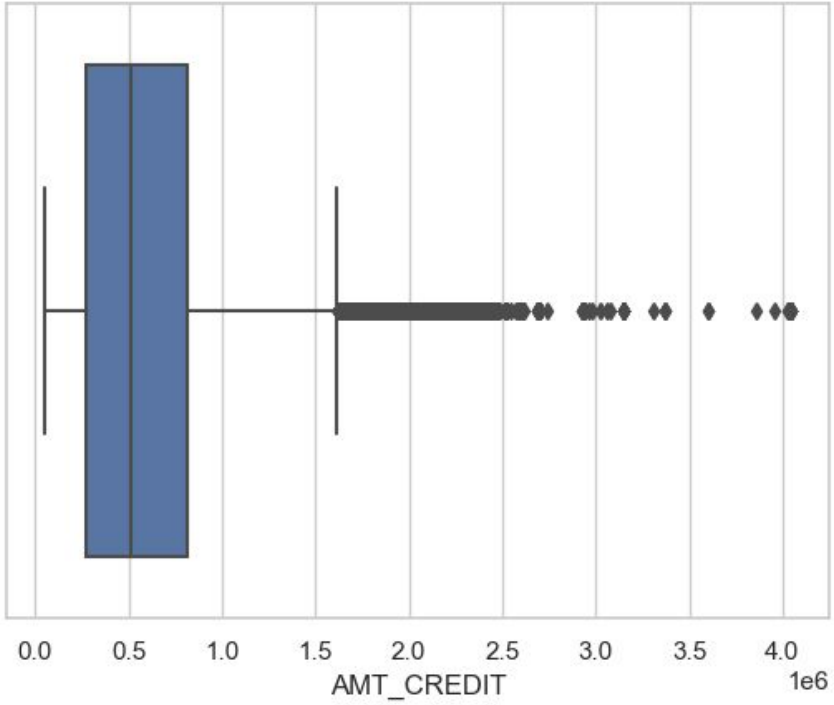
```
#PS Everything above 1M is an outlier
```



```
#PS values above 69898 seems to be outliers
```



```
#PS values above 1843771 seems to be outliers
```



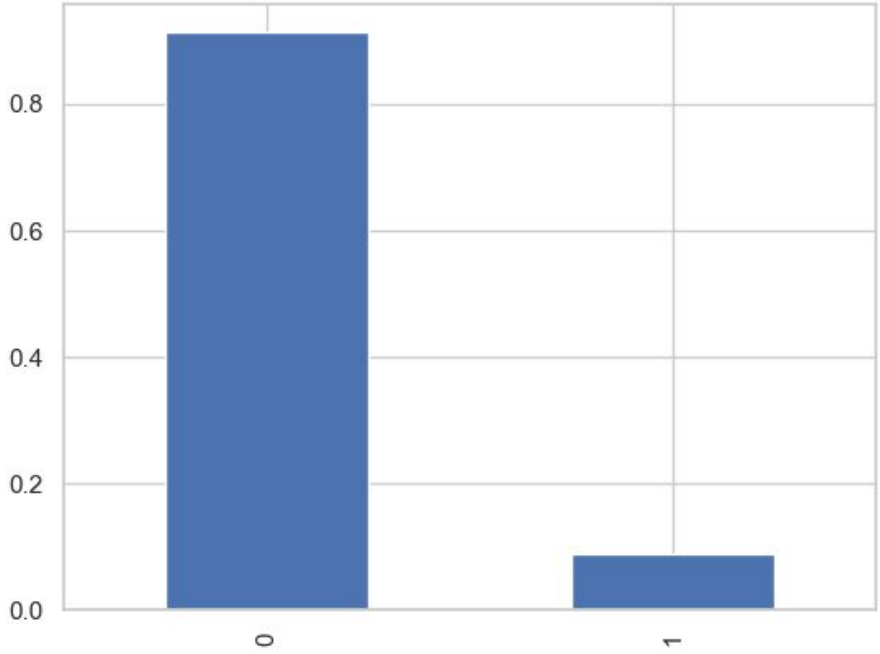
Target Analysis

Target = 1 are clients who are having difficulty repaying

Target = 0 are the clients who have been paying on time

```
#PS less than 10% people have difficulty paying back
```

]: <AxesSubplot:>

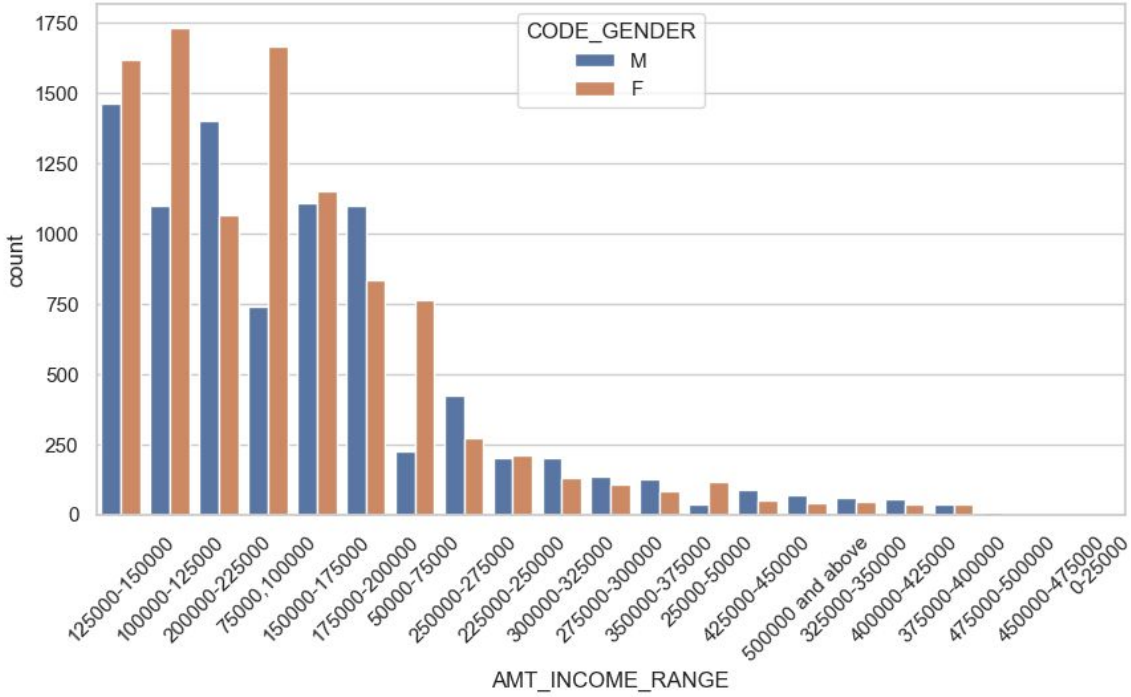


Deep dive into Target = 1

Analyse data and figure out patterns for our target clients

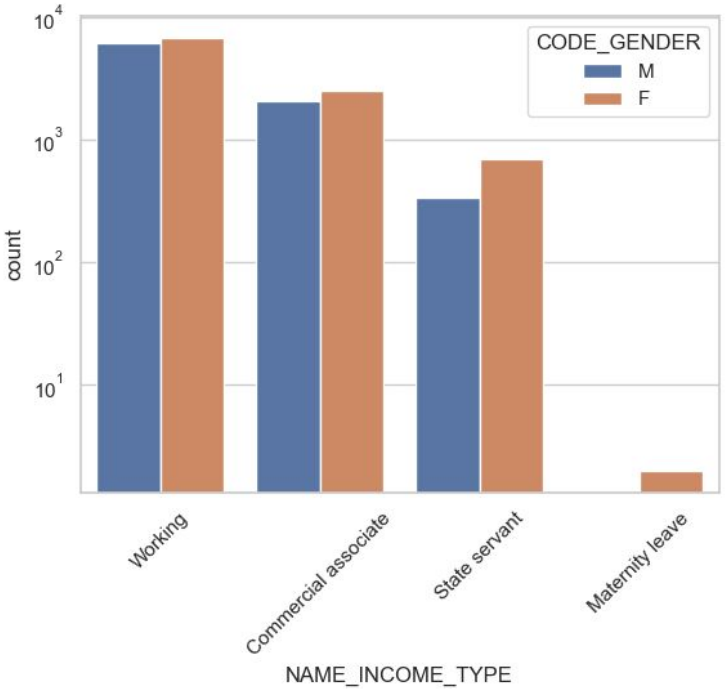
#Plotting AMT_INCOME_RANGE with CODE_GENDER

```
#The following plot shows that female count of default is higher than male
```



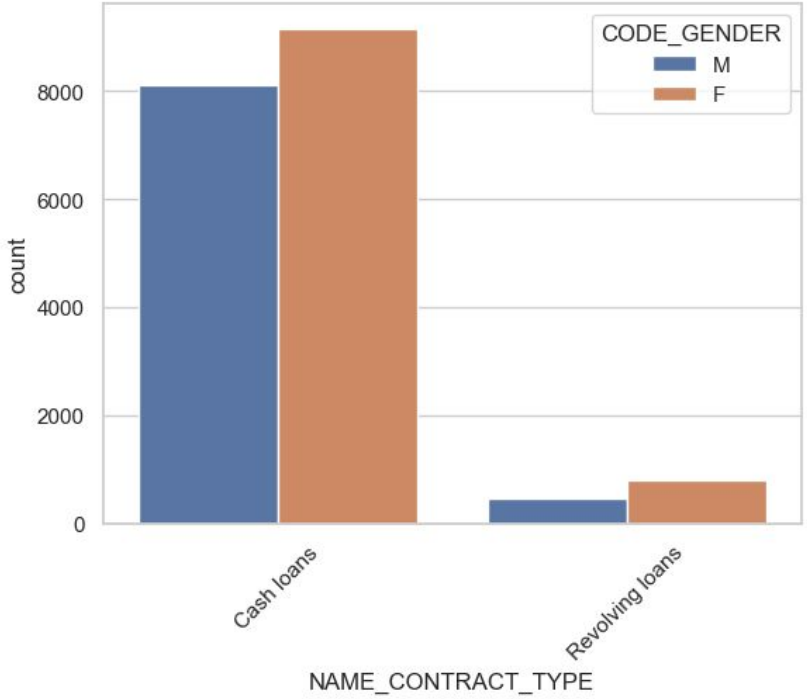
#Plotting NAME_INCOME_TYPE with CODE_GENDER

```
#The following plot shows that working people are most of the clients
#There are more female clients in the top 3 categories of income type
```

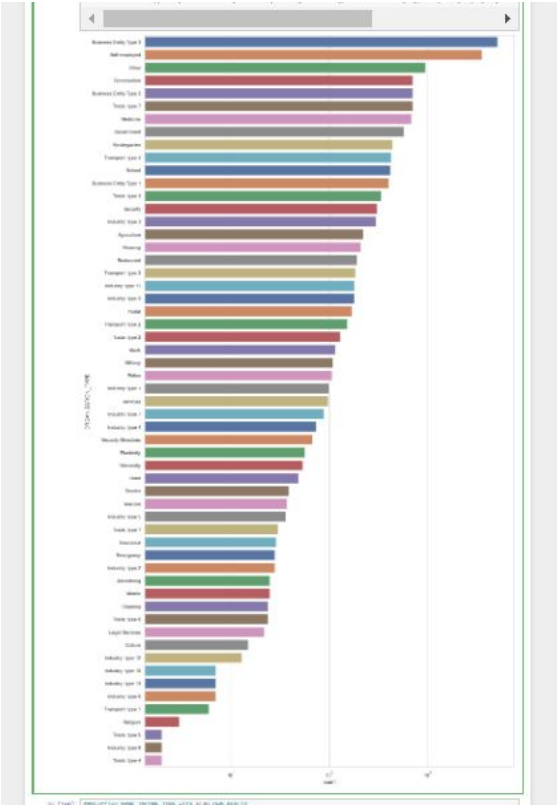


#Plotting NAME_CONTRACT_TYPE with CODE_GENDER

```
#Cash loans clearly has more clients per credit compared to revolving loans. In both cases,
```

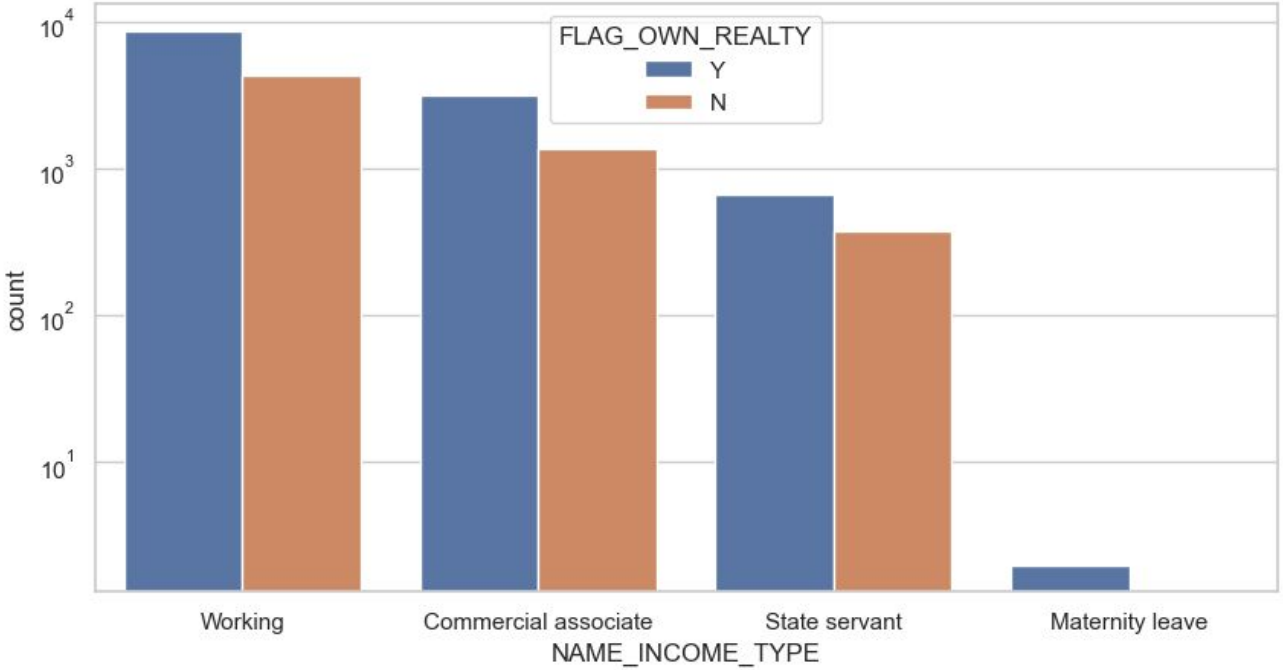


Plotting for ORGANIZATION_TYPE



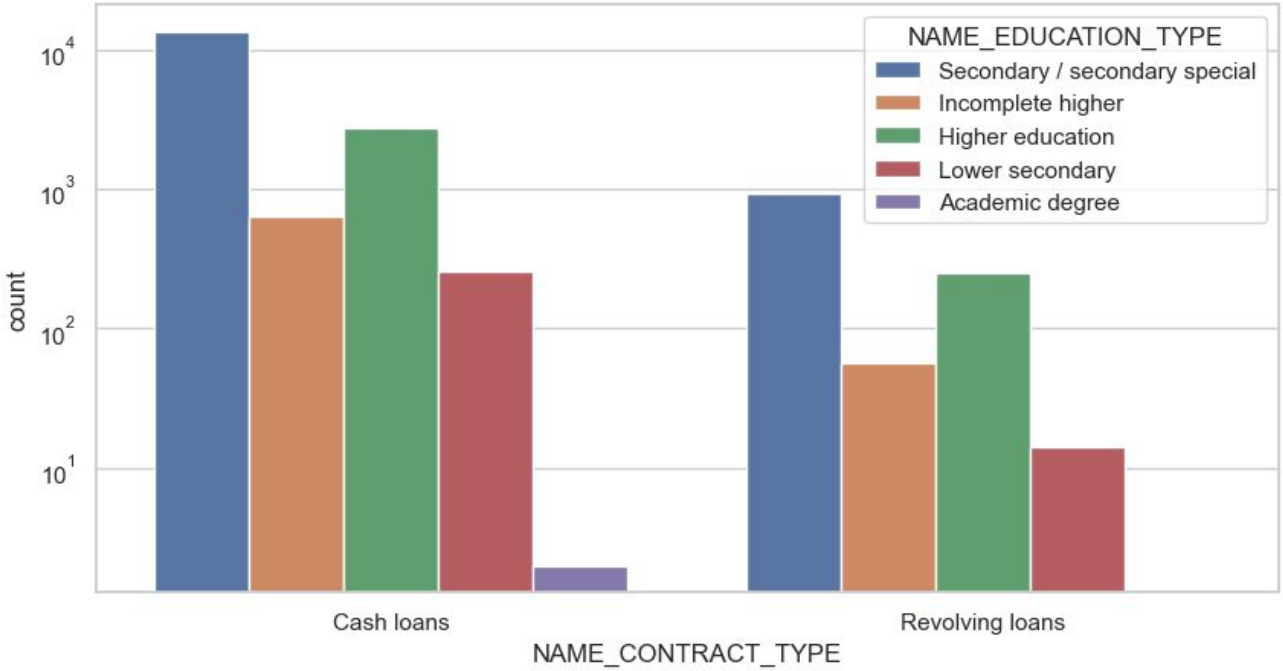
#Plotting NAME_INCOME_TYPE with FLAG_OWN_REALTY

```
#Working customers, obviously, have a higher count.
#As we can see, most customers do have their own property (house or a flat) but a large number of customers can be stat
```



#Plotting for NAME_CONTRACT_TYPE with NAME_EDUCATION_TYPE

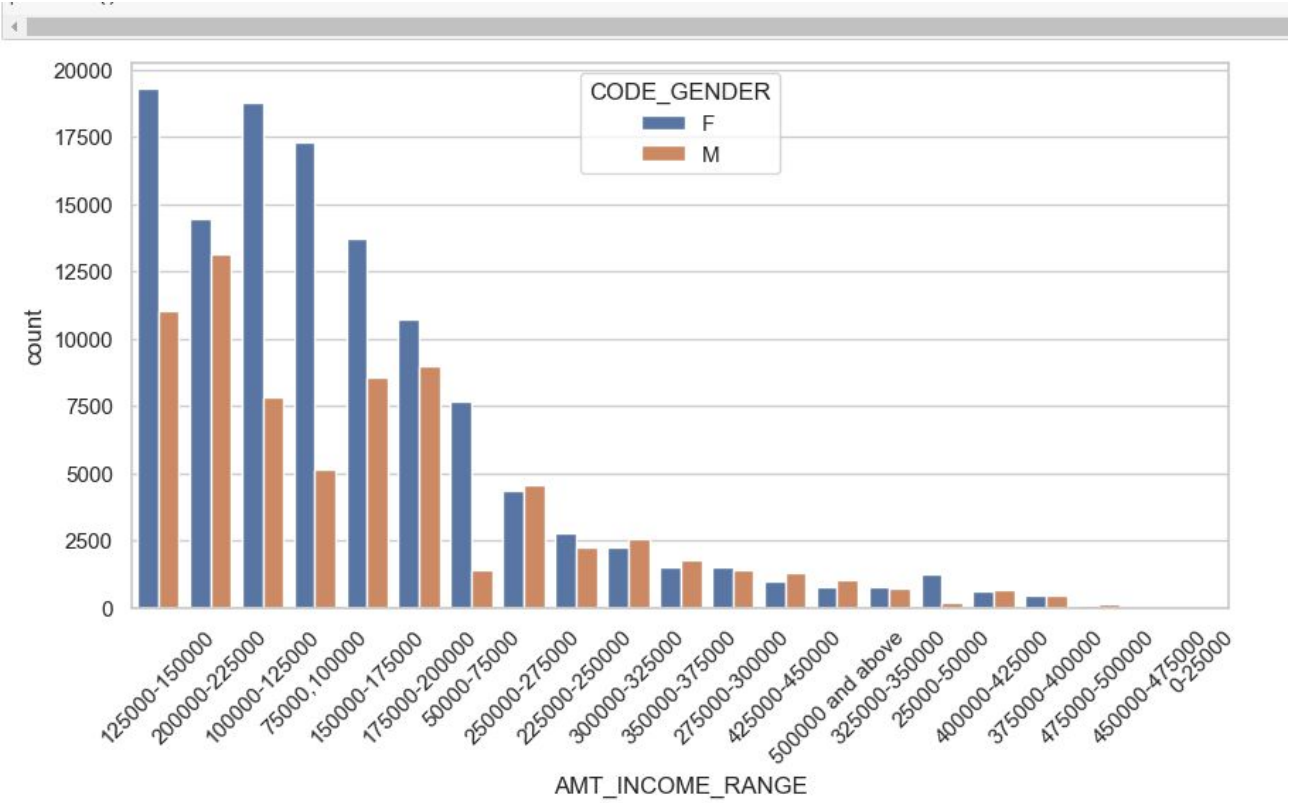
```
#Cash loans, as we can see, are preferred by clients of all education backgrounds with a majority.
#People with only an academic degree do not prefer revolving loans at all.
```



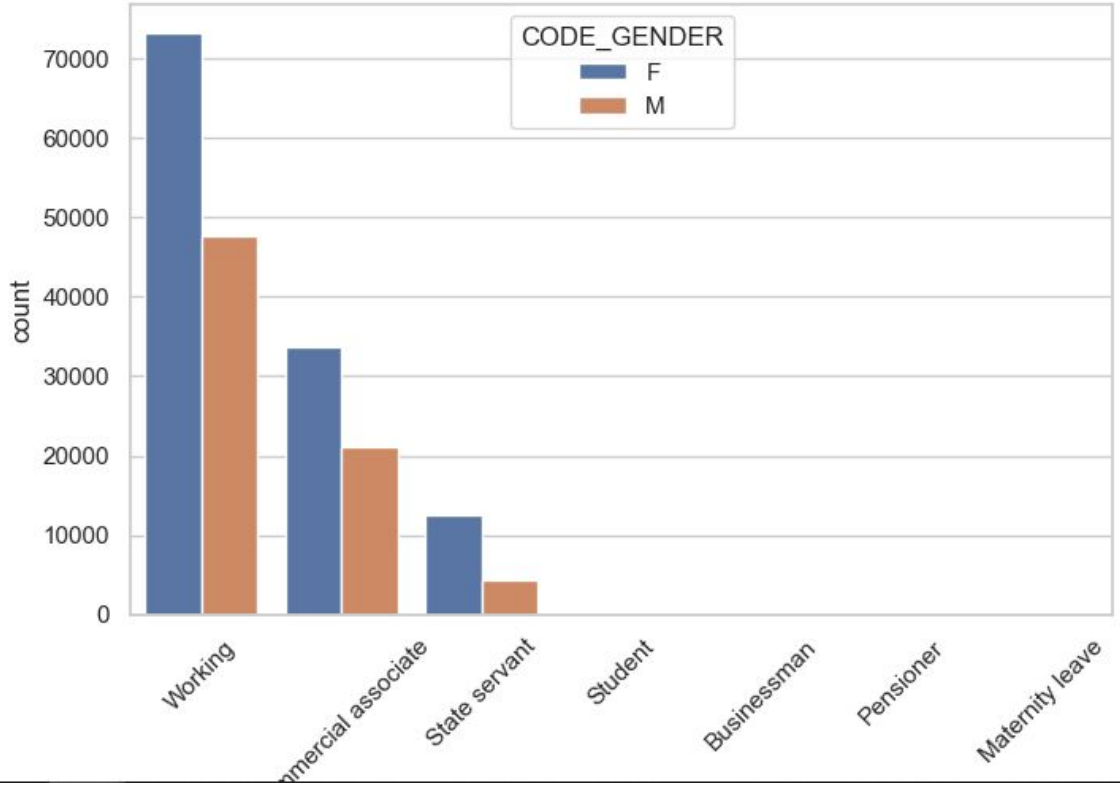
Deep dive into Target = 0

Analyse data and figure out patterns for our clients who have been paying on time

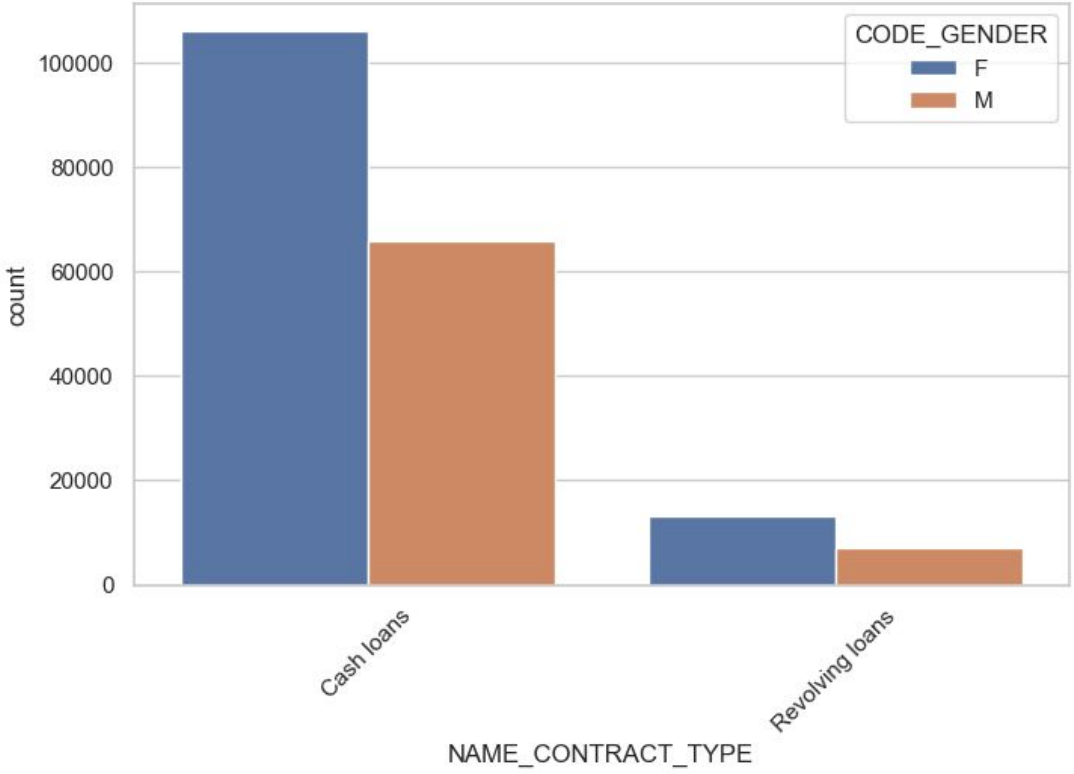
#Plotting AMT_INCOME_RANGE with gender



#Plotting NAME_INCOME_TYPE with gender



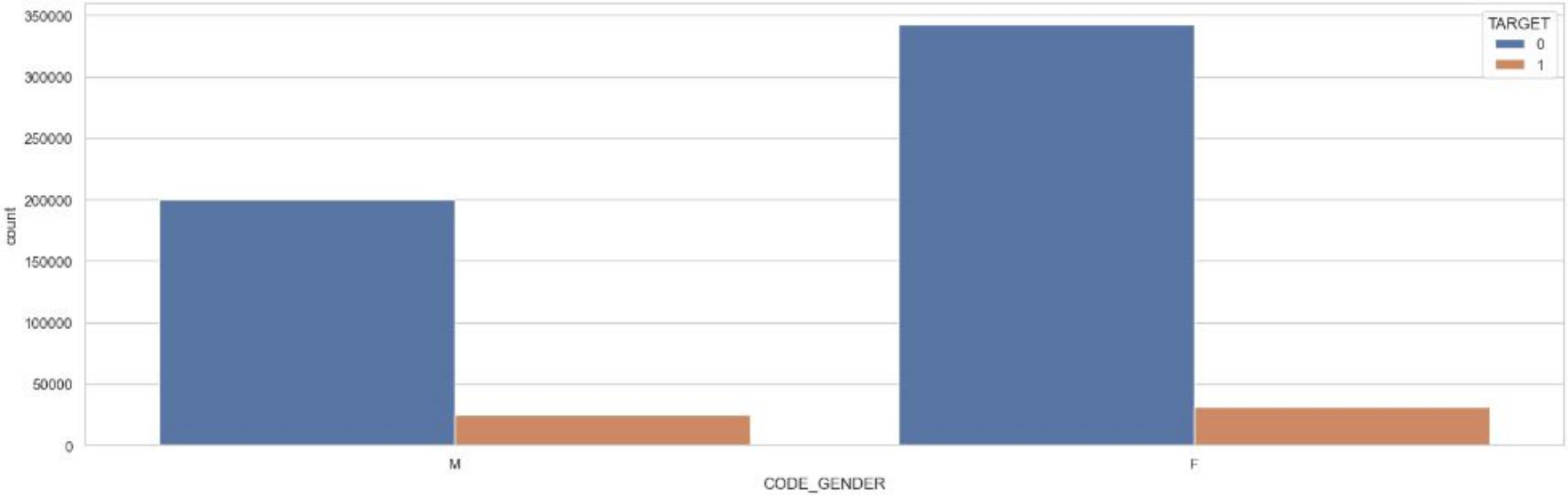
#Plotting NAME_CONTRACT_TYPE with gender



Merging the two data set

Merging previous application data to application data for further analysis

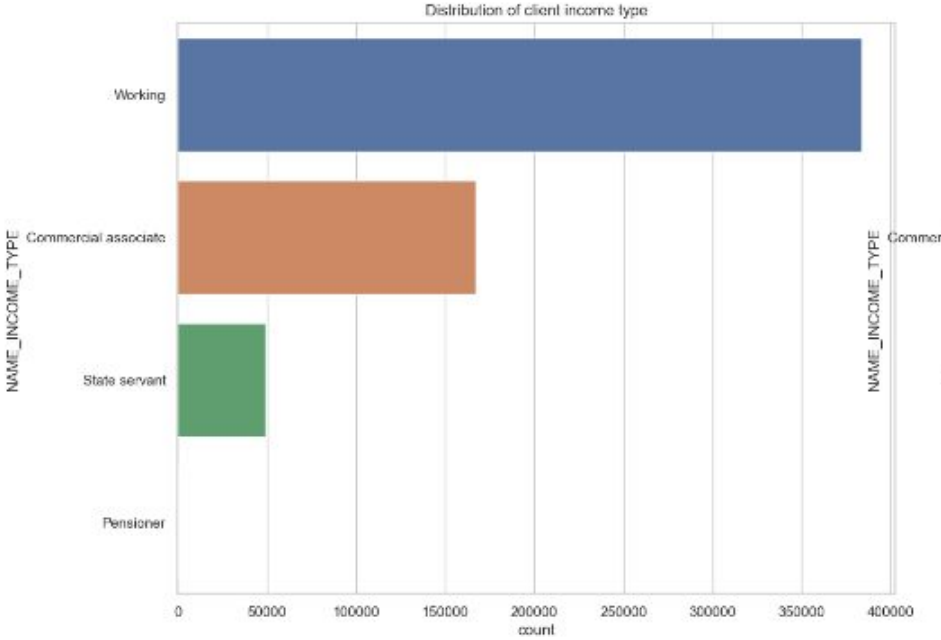
#Distribution of CODE_GENDER by TARGET for repayment status

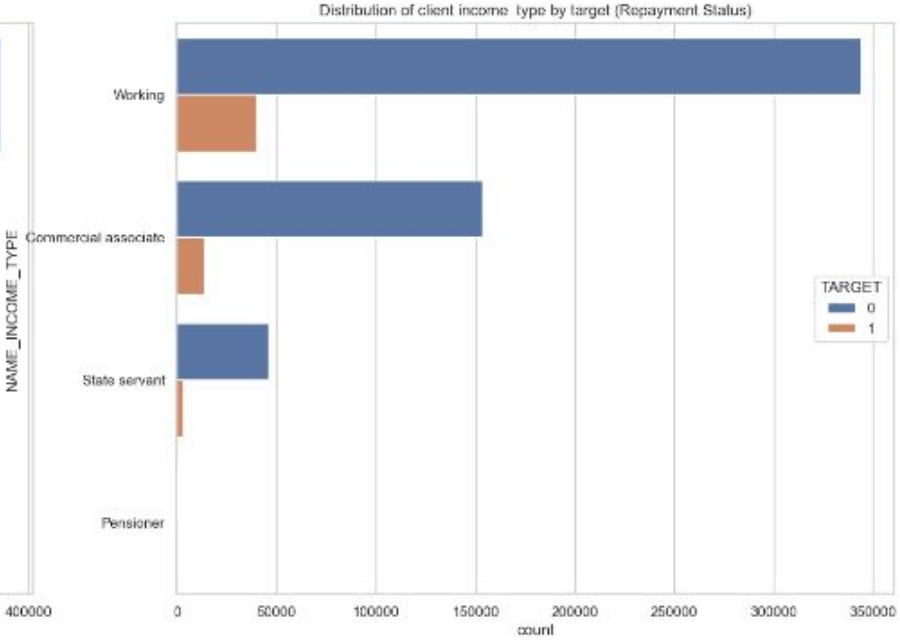


Clearly, female clients are the best repayers of their loan (almost double the amount of males).

Amount of defaulters in both genders are almost equally distributed.

#Distribution of client income



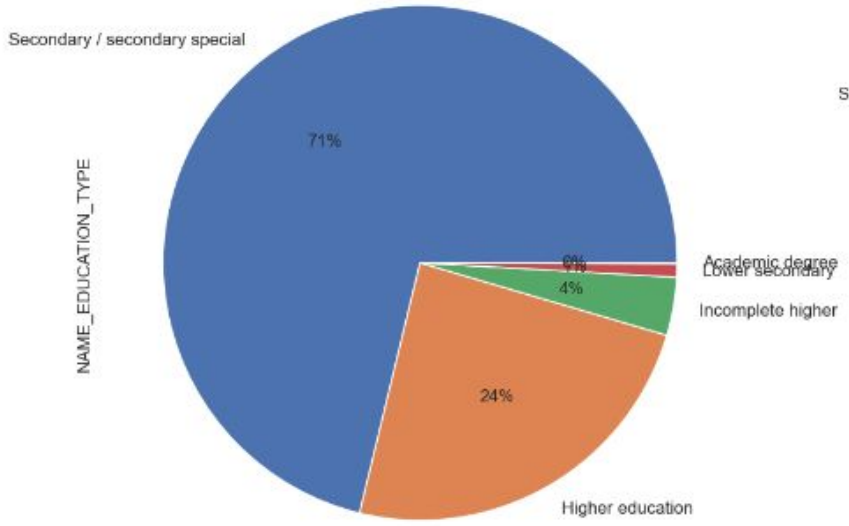


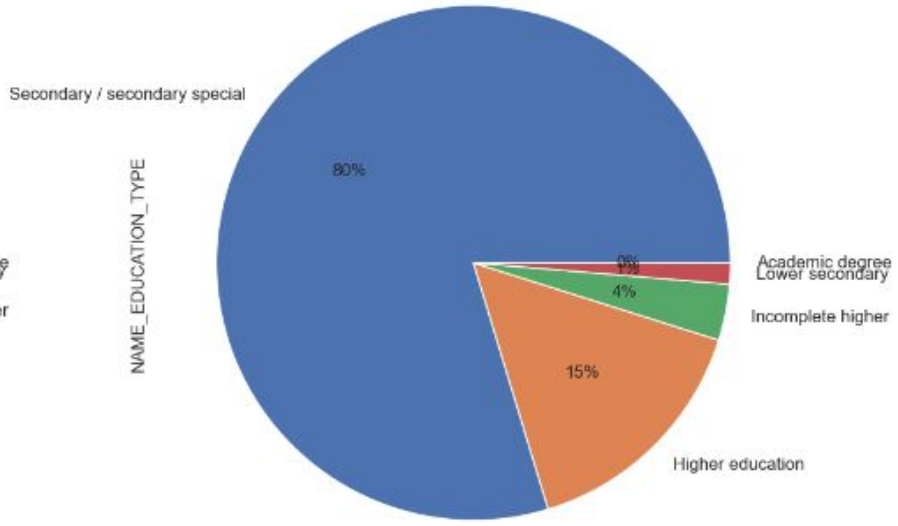
Most clients as per both cases of repayment status, are working

On the other end, the least amount of clients are pensioners

#Distribution of education type by repayment status

```
#For clientd with secondary edu, default is proportionally 9% higher compared to clients who do not default
#In the higher education category, clients who default are 8% fewer.
#In both cases of repayment status, lower secondary and academic degree categories are the minority.
```





Finding Correlation

Correlations between numerical data can be used to infer

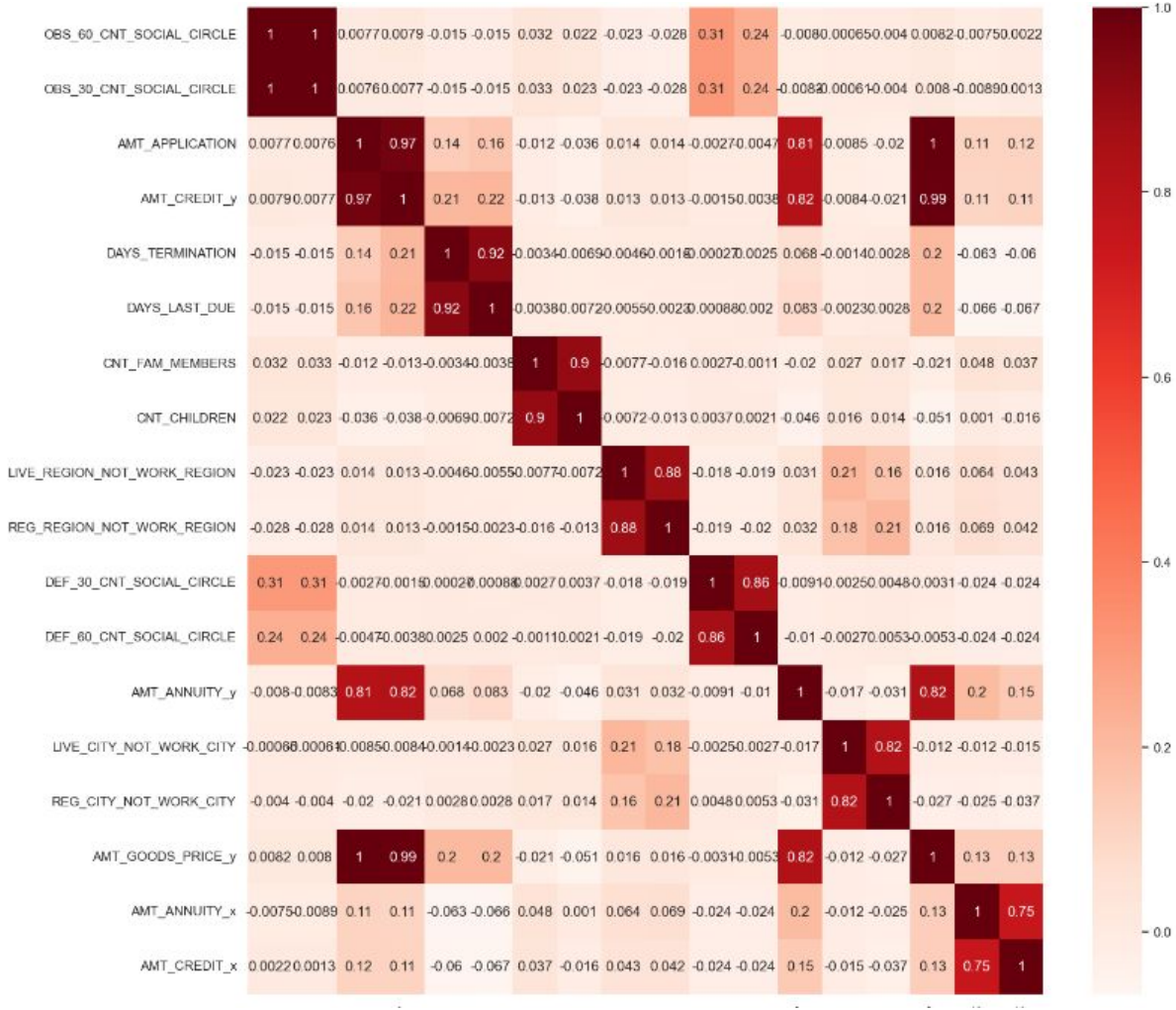
Top 10 correlated columns for repayers

- OBS_30_CNT_SOCIAL_CIRCLE OBS_60_CNT_SOCIAL_CIRCLE 1.00
- AMT_CREDIT_y AMT_APPLICATION 0.97
- DAYS_TERMINATION DAYS_LAST_DUE 0.93
- CNT_FAM_MEMBERS CNT_CHILDREN 0.90
- REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION 0.88
- DEF_30_CNT_SOCIAL_CIRCLE DEF_60_CNT_SOCIAL_CIRCLE 0.87
- AMT_GOODS_PRICE_y AMT_CREDIT_y 0.86
- AMT_APPLICATION AMT_GOODS_PRICE_y 0.85
- REG_CITY_NOT_WORK_CITY LIVE_CITY_NOT_WORK_CITY 0.83
- AMT_CREDIT_y AMT_ANNUITY_y 0.81


OBS_30_CNT_SOCIAL_CIRCLE 1 1 0.0053 -0.014 -0.015 0.031 0.022 -0.027 -0.022 0.31 0.24 0.0053 -0.002 0.00064 0.0051 -0.01
OBS_60_CNT_SOCIAL_CIRCLE 1 1 0.0054 -0.015 -0.015 0.031 0.022 -0.027 -0.023 0.31 0.24 0.0056 -0.0021 0.00057 0.0053 -0.0099
AMT_APPLICATION 0.0053 0.0054 1 0.14 0.16 -0.013 -0.037 0.015 0.015 -0.0024 -0.004 1 -0.02 -0.0082 0.97 0.81
DAYS_TERMINATION -0.014 -0.015 0.14 1 0.92 -0.0029 -0.0068 -0.0011 -0.004 0.0002 0.0028 0.2 0.0011 -0.002 0.21 0.072
DAYS_LAST_DUE -0.015 -0.015 0.16 0.92 1 -0.0031 -0.007 -0.0019 -0.0049 0.00071 0.0021 0.21 0.00098 -0.0028 0.22 0.088
CNT_FAM_MEMBERS 0.031 0.031 -0.013 -0.0029 -0.0031 1 0.9 -0.013 -0.0045 0.00043 -0.0026 -0.022 0.018 0.026 -0.014 -0.021
CNT_CHILDREN 0.022 0.022 -0.037 -0.0068 -0.007 0.9 1 -0.0098 -0.0044 0.0022 0.0018 -0.053 0.014 0.015 -0.039 -0.047
REG_REGION_NOT_WORK_REGION -0.027 -0.027 0.015 -0.0011 -0.0019 -0.013 -0.0098 1 0.88 -0.02 -0.022 0.016 0.21 0.18 0.014 0.033
LIVE_REGION_NOT_WORK_REGION -0.022 -0.023 0.015 -0.004 -0.0049 -0.0045 -0.0044 0.88 1 -0.018 -0.02 0.017 0.17 0.21 0.014 0.031
DEF_30_CNT_SOCIAL_CIRCLE 0.31 0.31 -0.0024 0.0002 -0.00071 0.00043 0.0022 -0.02 -0.018 1 0.86 -0.003 0.0049 -0.00078 -0.0014 -0.009
DEF_60_CNT_SOCIAL_CIRCLE 0.24 0.24 -0.004 0.0028 0.0021 -0.0026 0.0018 -0.022 -0.02 0.86 1 -0.0046 0.0056 -0.00087 -0.0035 -0.01
AMT_GOODS_PRICE_y 0.0053 0.0056 1 0.2 0.21 -0.022 -0.053 0.016 0.017 -0.003 -0.0046 1 -0.026 -0.012 0.99 0.82
REG_CITY_NOT_WORK_CITY -0.002 -0.0021 -0.02 0.0011 0.00098 0.018 0.014 0.21 0.17 0.0049 0.0056 -0.026 1 0.83 -0.021 -0.031
LIVE_CITY_NOT_WORK_CITY 0.00064 0.00057 -0.0082 -0.002 -0.0028 0.026 0.015 0.18 0.21 -0.00078 0.00087 -0.012 0.83 1 -0.0083 -0.017
AMT_CREDIT_y 0.0051 0.0053 0.97 0.21 0.22 -0.014 -0.039 0.014 0.014 -0.0014 -0.0035 0.99 -0.021 -0.0083 1 0.82
AMT_ANNUITY_y -0.01 -0.0099 0.81 0.072 0.088 -0.021 -0.047 0.033 0.031 -0.009 -0.01 0.82 -0.031 -0.017 0.82 1
RCLE RCLE ATION ATION _DUE BERS DREN GION GION RCLE RCLE ICE_y _CITY _CITY EDIT_y UITY_y
1.0
0.8
0.6
0.4
0.2
0.0

Top 10 correlations for Defaulters

- OBS_60_CNT_SOCIAL_CIRCLE OBS_30_CNT_SOCIAL_CIRCLE 1.00
- AMT_APPLICATION AMT_CREDIT_y 0.97
- DAYS_TERMINATION DAYS_LAST_DUE 0.95
- CNT_FAM_MEMBERS CNT_CHILDREN 0.90
- LIVE_REGION_NOT_WORK_REGION REG_REGION_NOT_WORK_REGION 0.87
- DEF_30_CNT_SOCIAL_CIRCLE DEF_60_CNT_SOCIAL_CIRCLE 0.86
- AMT_CREDIT_y AMT_ANNUITY_y 0.83
- LIVE_CITY_NOT_WORK_CITY REG_CITY_NOT_WORK_CITY 0.78
- AMT_ANNUITY_y AMT_GOODS_PRICE_y 0.76
- AMT_ANNUITY_x AMT_CREDIT_x 0.74


OBS_60_CNT_SOCIAL_CIRCLE 1 1 0.0077 0.0079 -0.015 -0.015 0.032 0.022 -0.023 -0.028 0.31 0.24 -0.0080 0.00065 0.004 0.0082 -0.0075 0.0022
OBS_30_CNT_SOCIAL_CIRCLE 1 1 0.0076 0.0077 -0.015 -0.015 0.033 0.023 -0.023 -0.028 0.31 0.24 -0.0082 0.00061 0.004 0.008 -0.0089 0.0013
AMT_APPLICATION 0.0077 0.0076 1 0.97 0.14 0.16 -0.012 -0.036 0.014 0.014 -0.0027 -0.0047 0.81 -0.0085 -0.02 1 0.11 0.12
AMT_CREDIT_y 0.0079 0.0077 0.97 1 0.21 0.22 -0.013 -0.038 0.013 0.013 -0.0015 -0.0038 0.82 -0.0084 -0.021 0.99 0.11 0.11
DAYS_TERMINATION -0.015 -0.015 0.14 0.21 1 0.92 -0.0034 -0.0069 -0.0046 -0.0016 -0.00027 0.0025 0.068 -0.0014 0.0028 0.2 -0.063 -0.06
DAYS_LAST_DUE -0.015 -0.015 0.16 0.22 0.92 1 0.0038 0.0072 0.0055 0.0023 0.00088 0.002 0.083 -0.0023 0.0028 0.2 -0.066 -0.067
CNT_FAM_MEMBERS 0.032 0.033 -0.012 -0.013 -0.0034 0.0038 1 0.9 -0.0077 -0.016 0.0027 -0.0011 -0.02 0.027 0.017 -0.021 0.048 0.037
CNT_CHILDREN 0.022 0.023 -0.036 -0.038 -0.0069 0.0072 0.9 1 -0.0072 -0.013 0.0037 0.0021 -0.046 0.016 0.014 -0.051 0.001 -0.016
LIVE_REGION_NOT_WORK_REGION -0.023 -0.023 0.014 0.013 -0.0046 -0.0055 0.0077 -0.0072 1 0.88 -0.018 -0.019 0.031 0.21 0.16 0.016 0.064 0.043
REG_REGION_NOT_WORK_REGION -0.028 -0.028 0.014 0.013 -0.0015 -0.0023 -0.016 -0.013 0.88 1 -0.019 -0.02 0.032 0.18 0.21 0.016 0.069 0.042
DEF_30_CNT_SOCIAL_CIRCLE 0.31 0.31 -0.0027 -0.0015 0.00026 0.00088 0.0027 0.0037 -0.018 -0.019 1 0.86 -0.0091 -0.0025 0.0048 -0.0031 -0.024 -0.024
DEF_60_CNT_SOCIAL_CIRCLE 0.24 0.24 -0.0047 -0.0038 0.0025 0.002 -0.0011 0.0021 -0.019 -0.02 0.86 1 -0.01 -0.0027 0.0053 -0.0053 -0.024 -0.024
AMT_ANNUITY_y -0.008 -0.0083 0.81 0.82 0.068 0.083 -0.02 -0.046 0.031 0.032 -0.0091 -0.01 1 -0.017 -0.031 0.82 0.2 0.15
LIVE_CITY_NOT_WORK_CITY -0.00066 -0.00061 0.0085 0.0084 -0.0014 -0.0023 0.027 0.016 0.21 0.18 -0.0025 0.0027 -0.017 1 0.82 -0.012 -0.012 -0.015
REG_CITY_NOT_WORK_CITY -0.004 -0.004 -0.02 -0.021 0.0028 0.0028 0.017 0.014 0.16 0.21 0.0048 0.0053 -0.031 0.82 1 -0.027 -0.025 -0.037
AMT_GOODS_PRICE_y 0.0082 0.008 1 0.99 0.2 0.2 -0.021 -0.051 0.016 0.016 -0.0031 -0.0053 0.82 -0.012 -0.027 1 0.13 0.13
AMT_ANNUITY_x -0.0075 -0.0089 0.11 0.11 -0.063 -0.066 0.048 0.001 0.064 0.069 -0.024 -0.024 0.2 -0.012 -0.025 0.13 1 0.75
AMT_CREDIT_x 0.0022 0.0013 0.12 0.11 -0.06 -0.067 0.037 -0.016 0.043 0.042 -0.024 -0.024 0.15 -0.015 -0.037 0.13 0.75 1
1.0
0.8
0.6
0.4
0.2
0.0

CONCLUSION:

Target

1. Students, Pensioners and Commercial Associates with a housing type such as office/co-op/municipal apartments
2. Clients living with parents
3. Female clients, as they have a high rate of repayment
4. Should target client who own car

Not to target

1. Based on education, as there is no sufficient evidence available
2. Females on maternity leave
3. Clients 'working', as they have highest defaulters
4. 'Repairs' purpose of loans have the most amount of defaulters, loan disbursement should be of low risk

Thank You. It was a pleasure brainstorming and finishing this assignment.