CREDIT EDA CASE STUDY

Bijay Khanal

- There are two datasets:
- One is application dataset which is the current and the other is the previous dataset which is of the past.
- In our analysis we have used target variable
- Target variable (1 - client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in our sample, 0 - all other cases)

Business Objectives

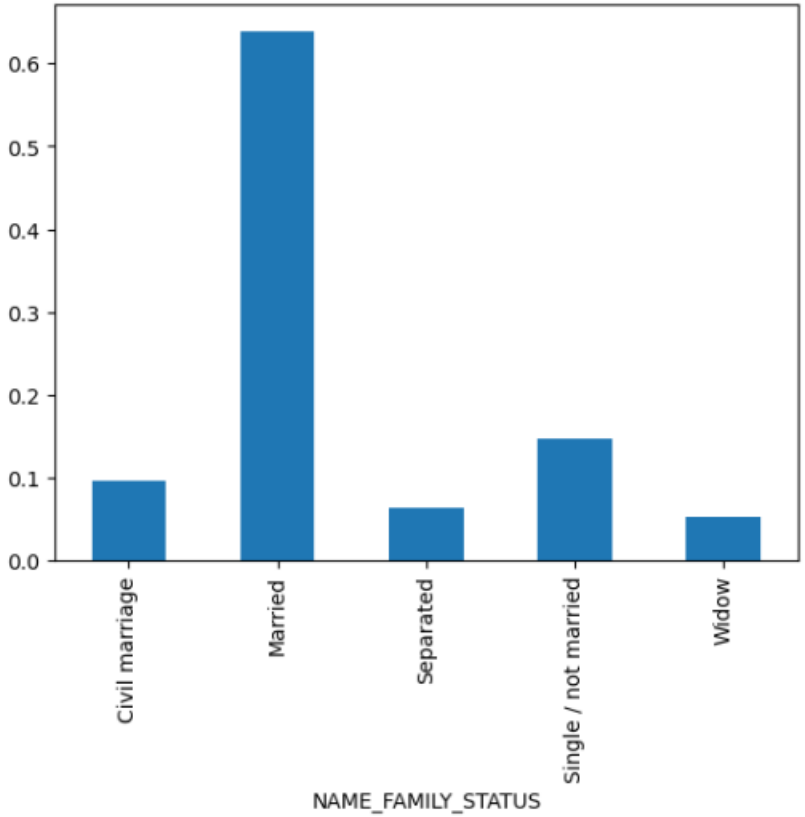
- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

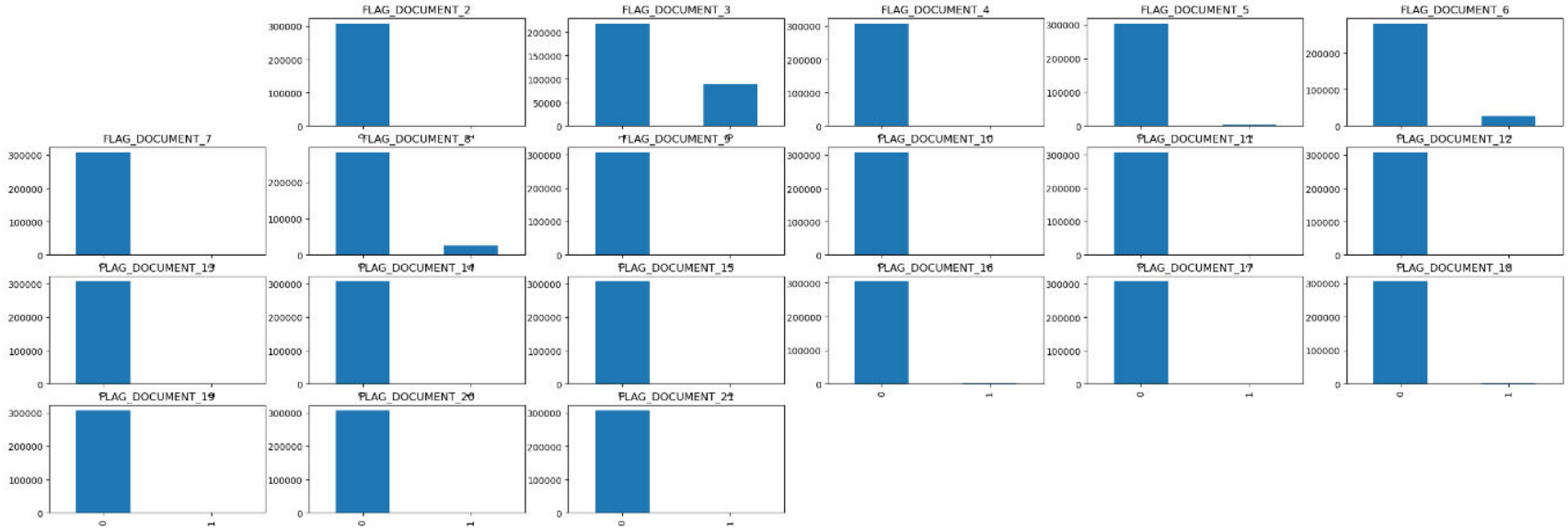
Steps for analysis

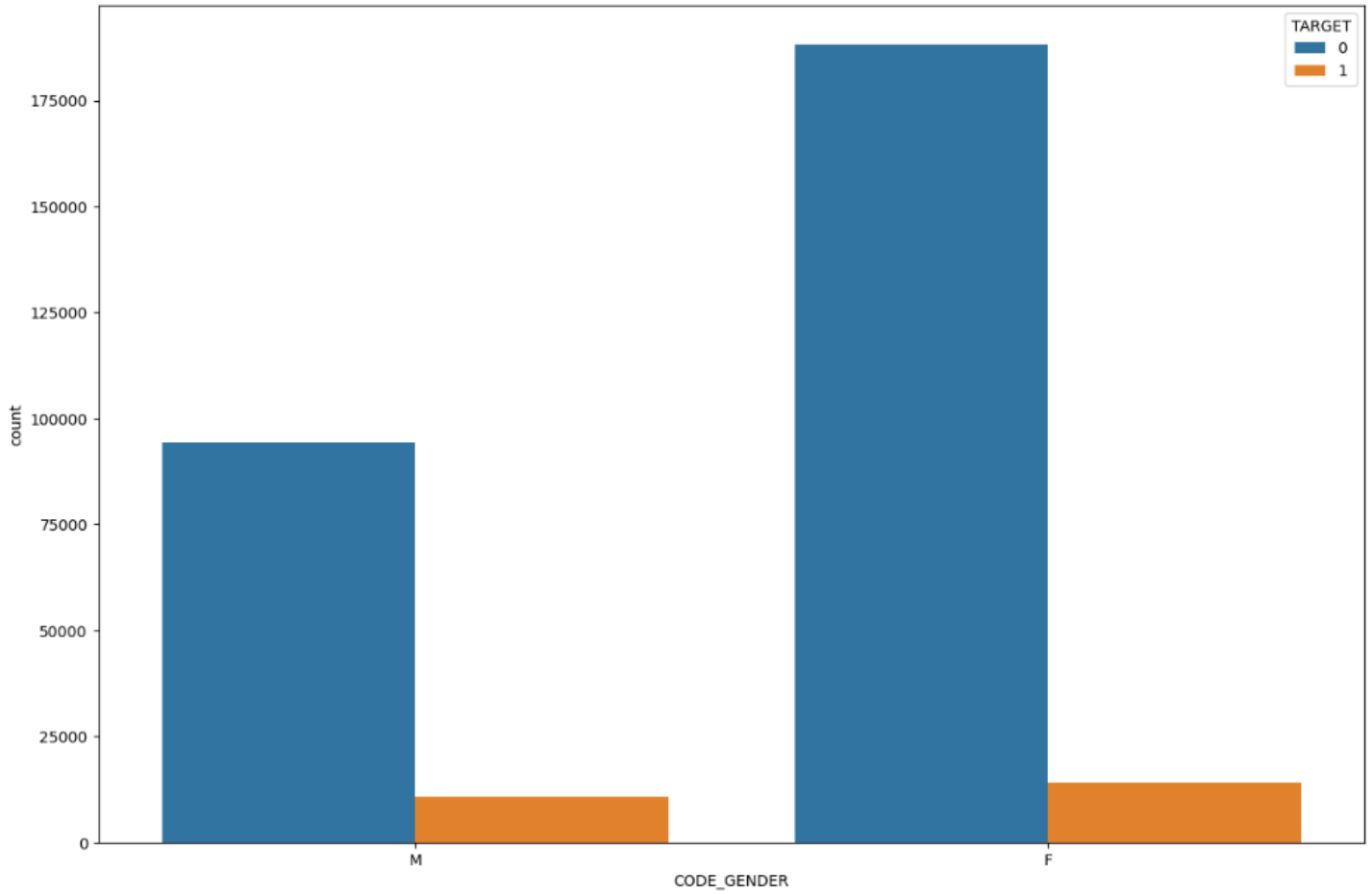
- There are two datasets
- Steps that are taken:
- Data loading and cleaning
- Unnecessary rows columns removal
- Handling outliers
- Handling null values and imputing

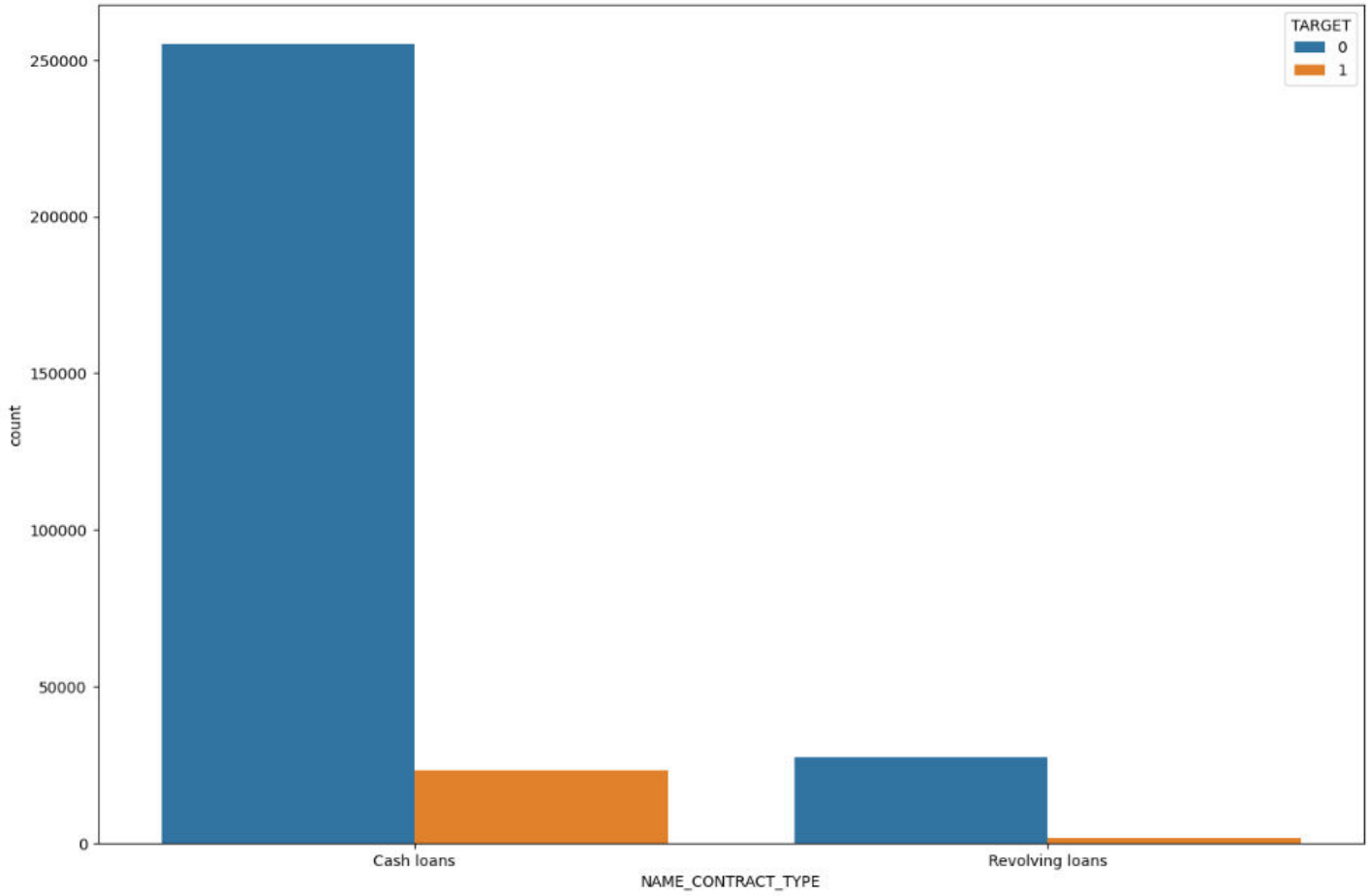
Application dataset

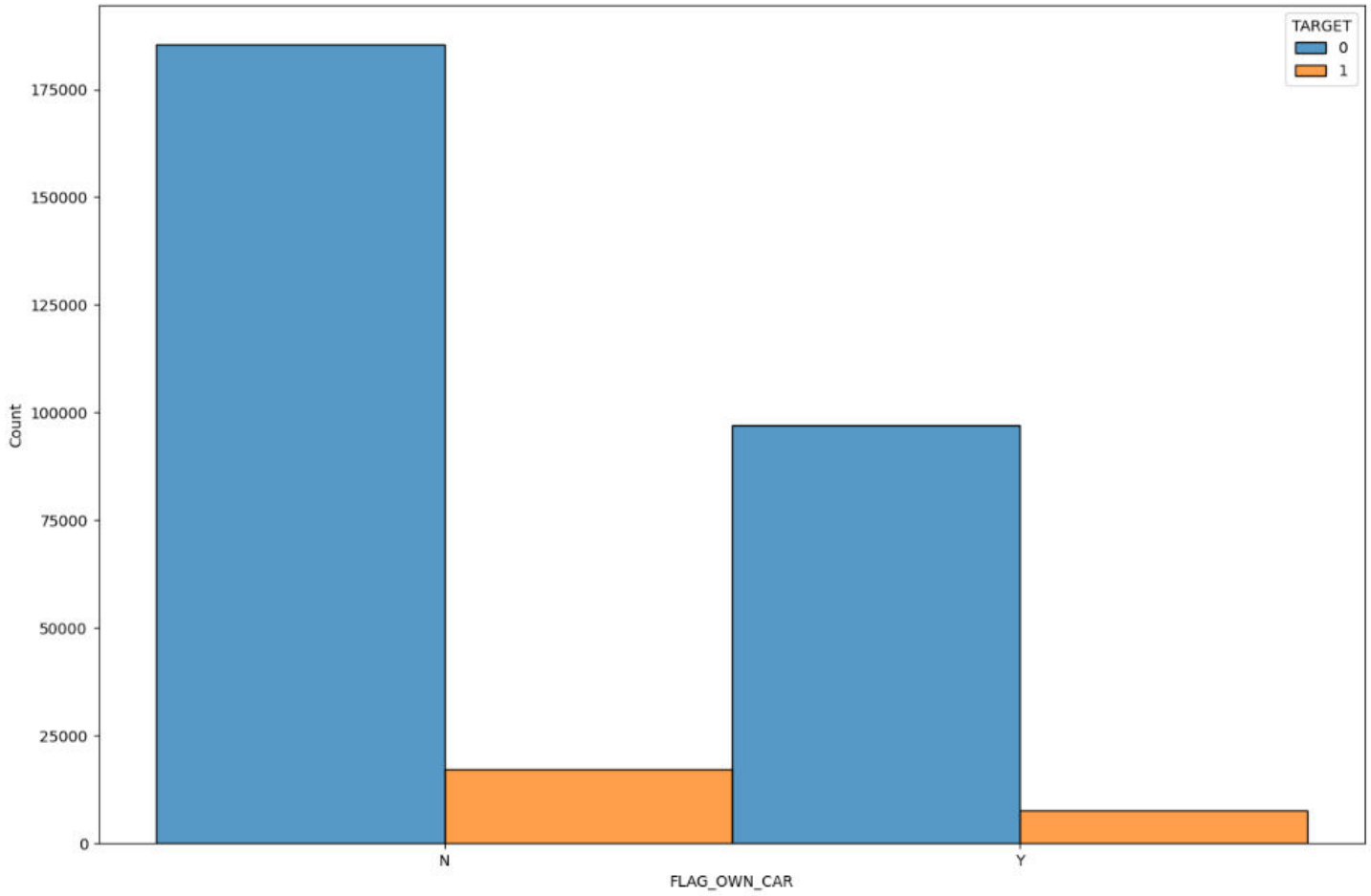
Univariate analysis



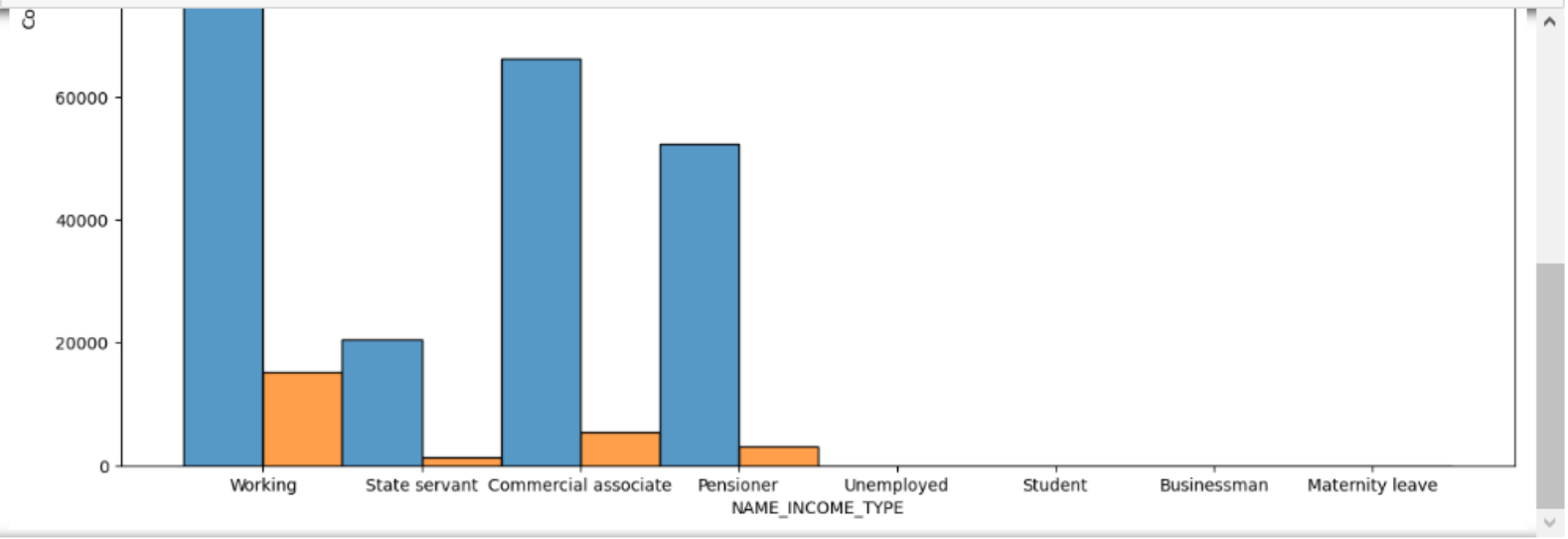

FLAG_DOCUMENT_2
FLAG_DOCUMENT_3
FLAG_DOCUMENT_4
FLAG_DOCUMENT_5
FLAG_DOCUMENT_6
FLAG_DOCUMENT_7
FLAG_DOCUMENT_8
FLAG_DOCUMENT_9
FLAG_DOCUMENT_10
FLAG_DOCUMENT_11
FLAG_DOCUMENT_12
FLAG_DOCUMENT_13
FLAG_DOCUMENT_14
FLAG_DOCUMENT_15
FLAG_DOCUMENT_16
FLAG_DOCUMENT_17
FLAG_DOCUMENT_18
FLAG_DOCUMENT_19
FLAG_DOCUMENT_20
FLAG_DOCUMENT_21


TARGET
0
1
175000
150000
125000
100000
75000
50000
25000
0
count
M
F
CODE_GENDER

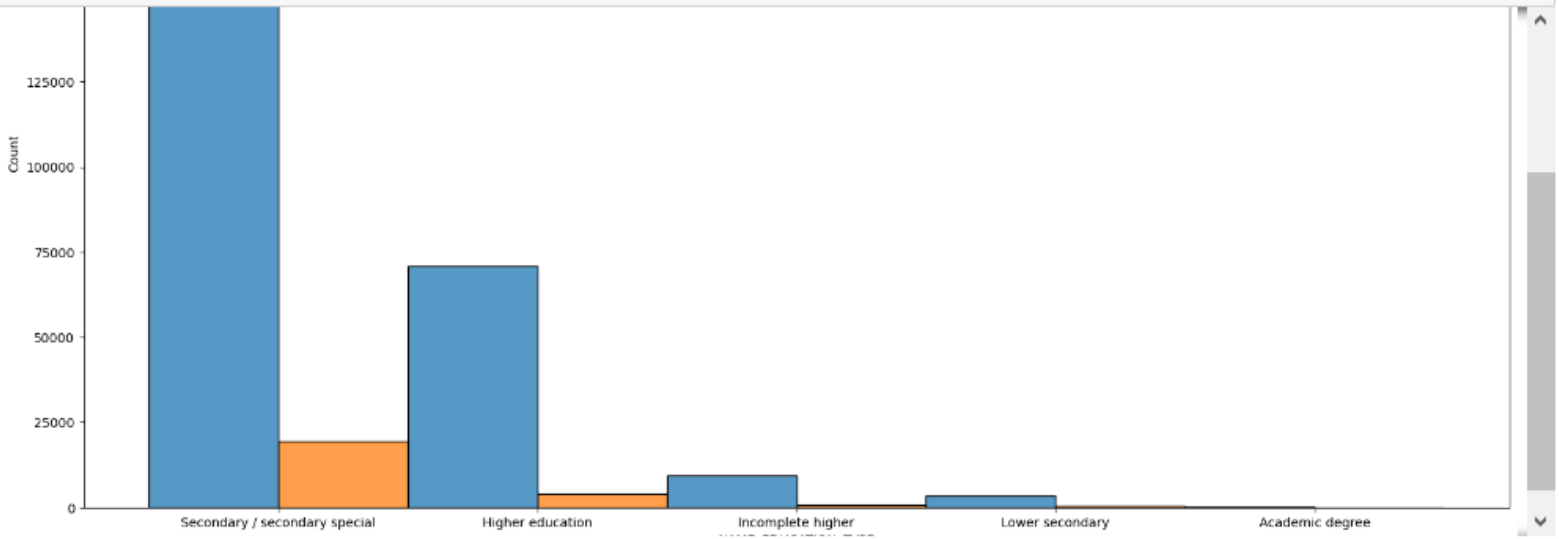

TARGET
0
1
250000
200000
150000
100000
50000
0
count
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

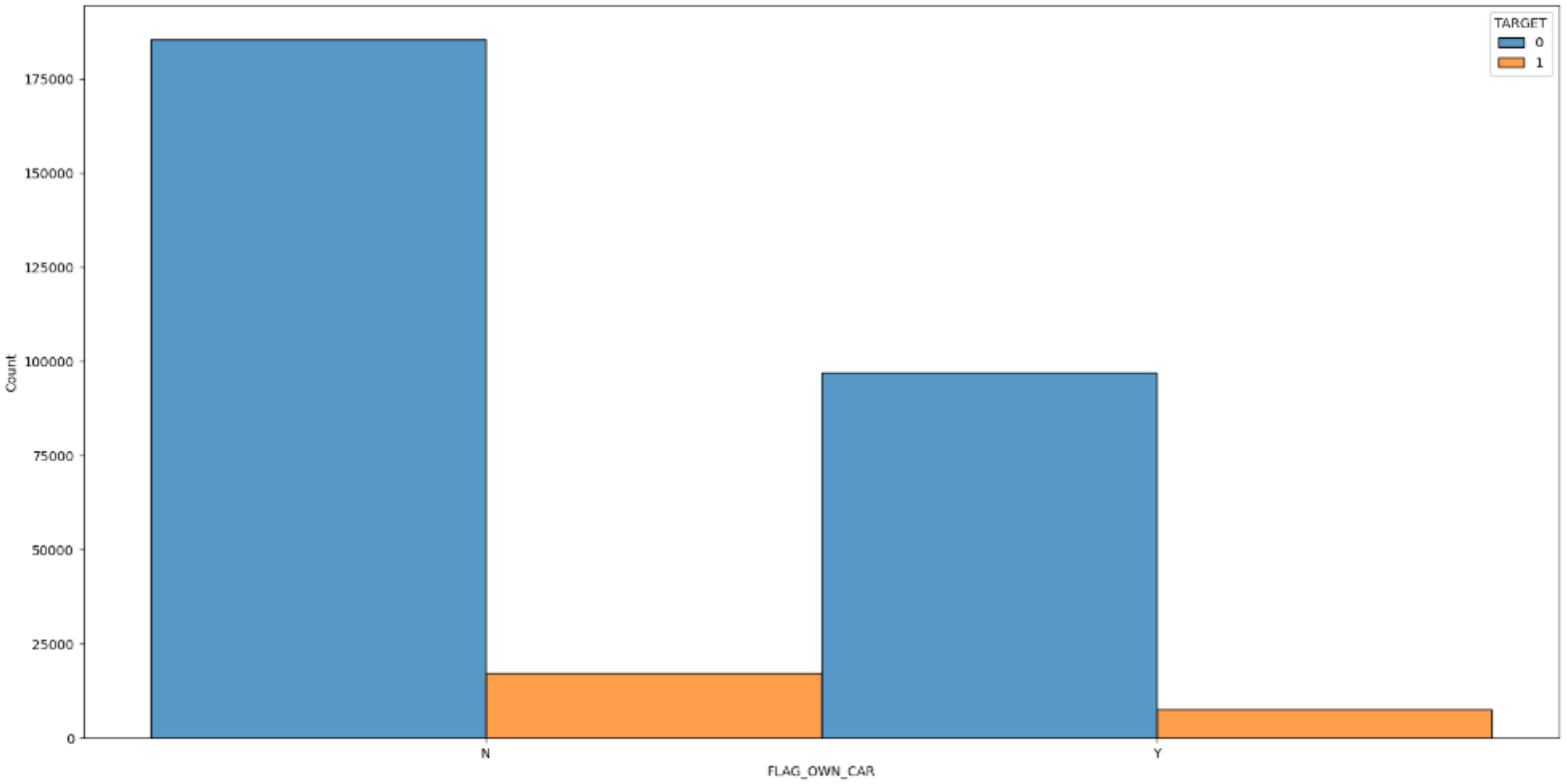

TARGET
0
1
175000
150000
125000
100000
75000
50000
25000
0
Count
N
Y
FLAG_OWN_CAR

In [618]:
```
plt.figure(figsize=(15,10))
plt.subplot(1,1,1)
sns.histplot(data=df1, x='NAME_INCOME_TYPE', hue='TARGET', multiple='dodge', discrete=True);
```

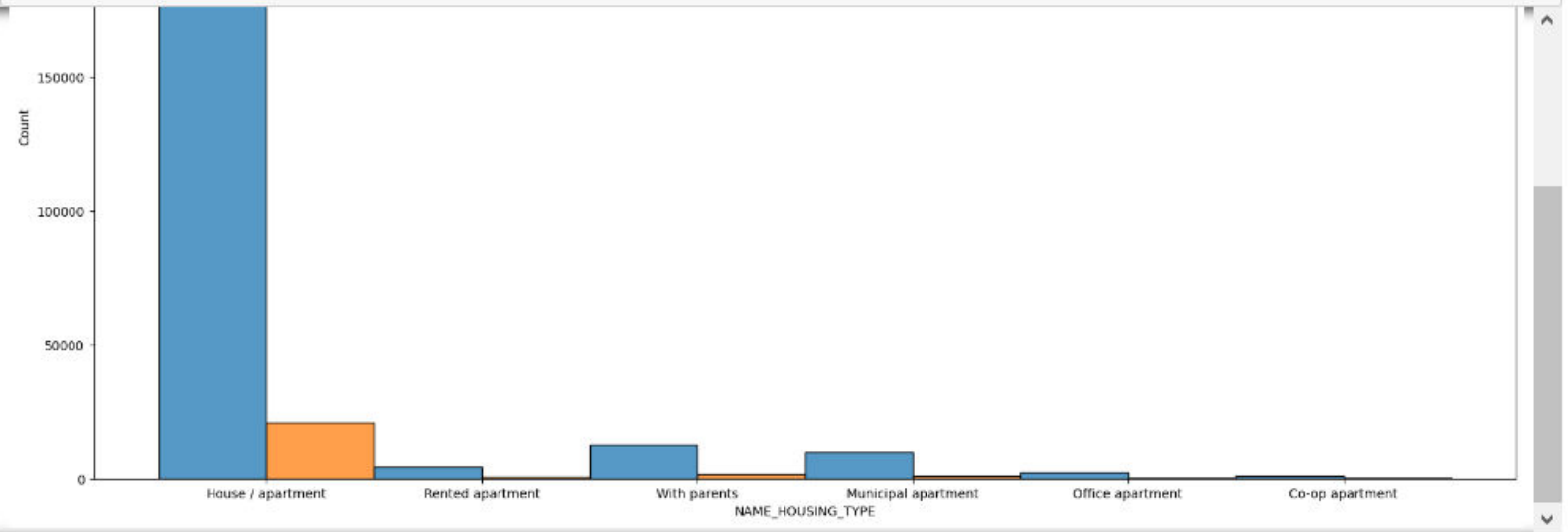


In [624]:
```
plt.figure(figsize=(20,10))
plt.subplot(1,1,1)
sns.histplot(data=df1, x='NAME_EDUCATION_TYPE', hue='TARGET', multiple='dodge', discrete=True);
```

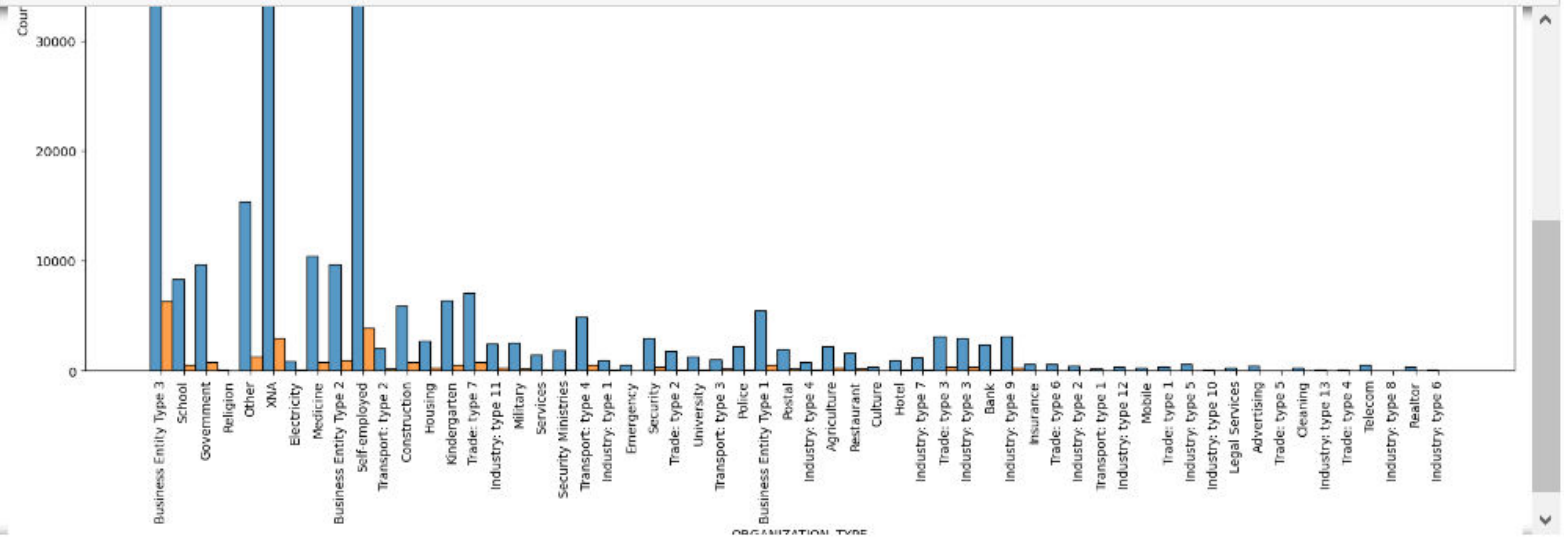


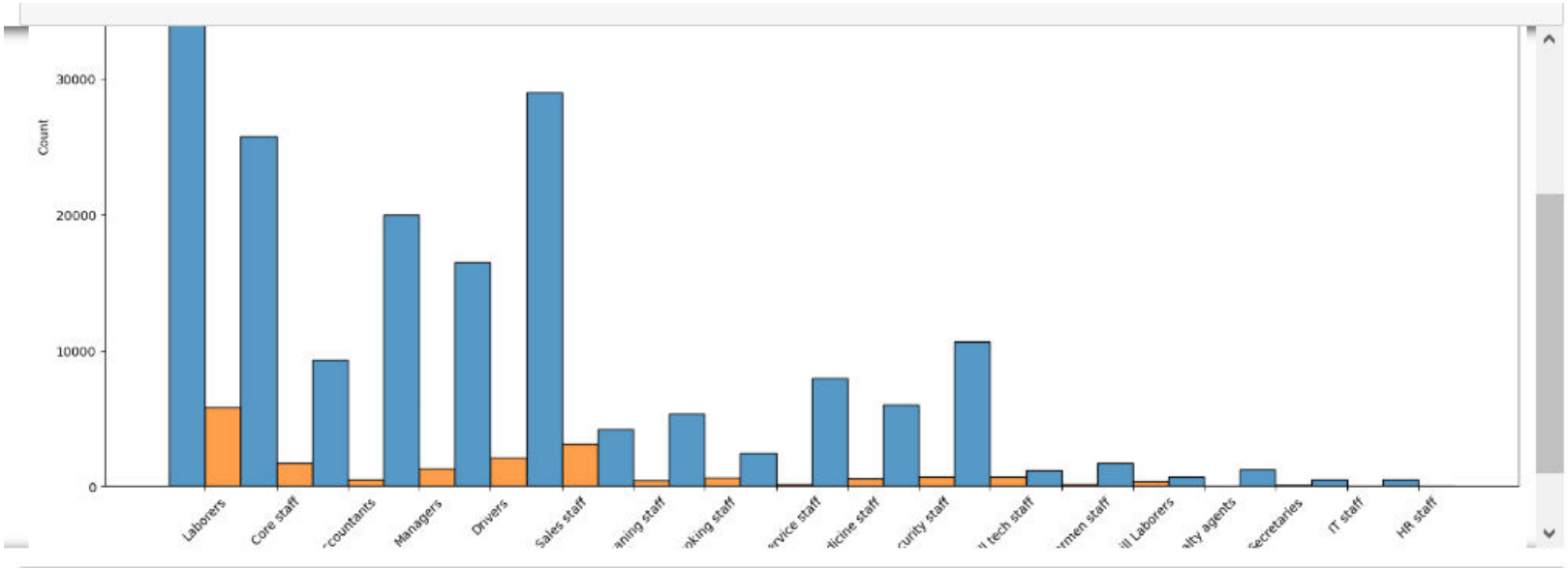

TARGET
0
1
175000
150000
125000
100000
75000
50000
25000
0
Count
N
Y
FLAG_OWN_CAR

In [639]:
```
plt.figure(figsize=(20,10))
plt.subplot(1,1,1)
sns.histplot(data=df1, x='NAME_HOUSING_TYPE', hue='TARGET', multiple='dodge', discrete=True);
```



In [645]:
```
plt.figure(figsize=(20,10))
plt.subplot(1,1,1)
sns.histplot(data=df1, x='ORGANIZATION_TYPE', hue='TARGET', multiple='dodge', discrete=True);
plt.xticks(rotation = 90);
```

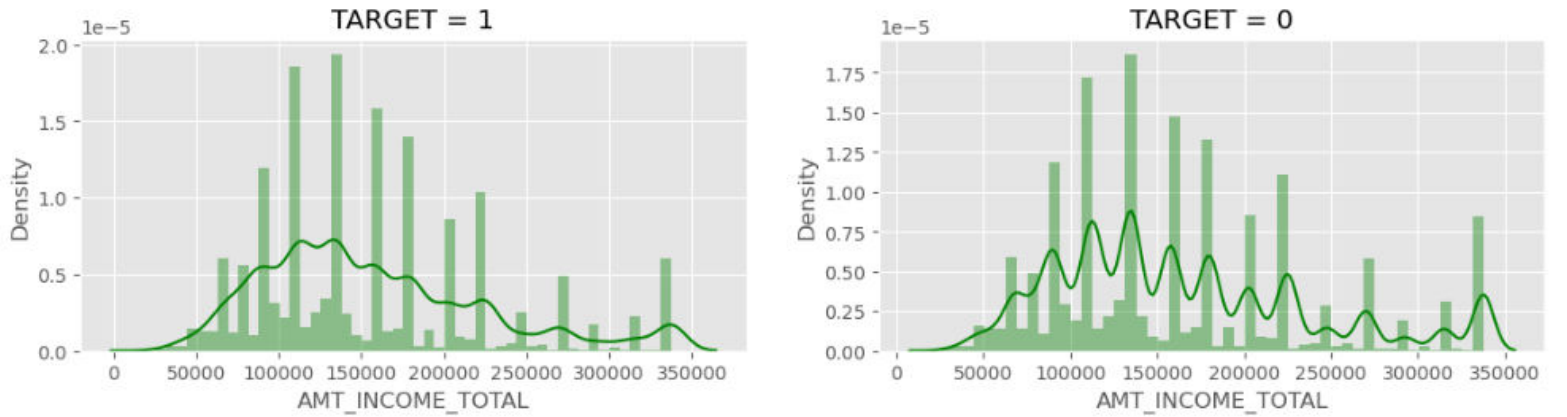



Count
30000
20000
10000
0
Laborers
Core staff
Managers
Drivers
Sales staff
HR staff

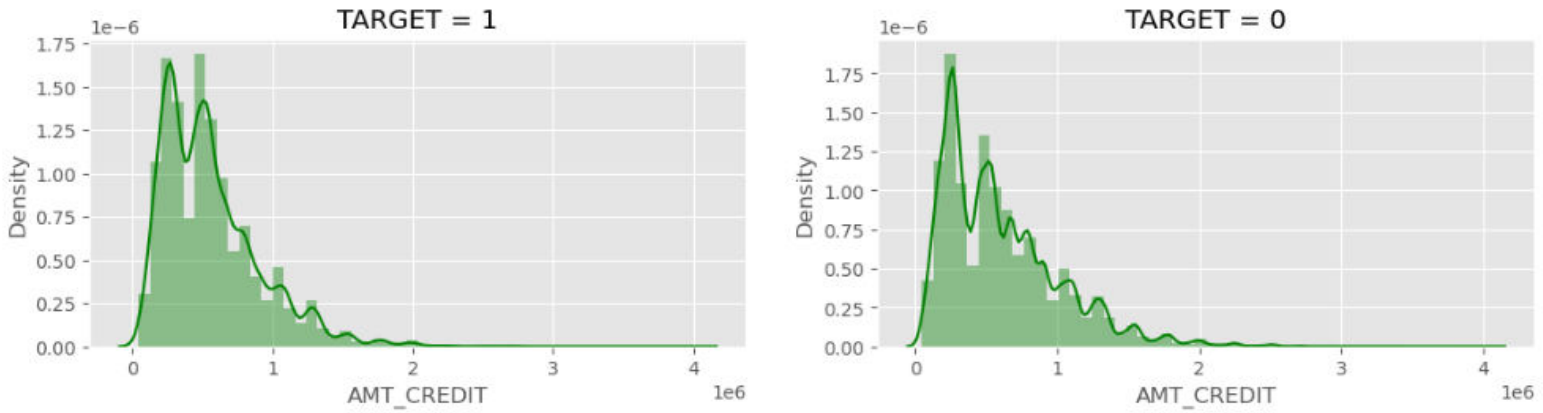
Bivariate analysis

BIVARIATE ANALYSIS

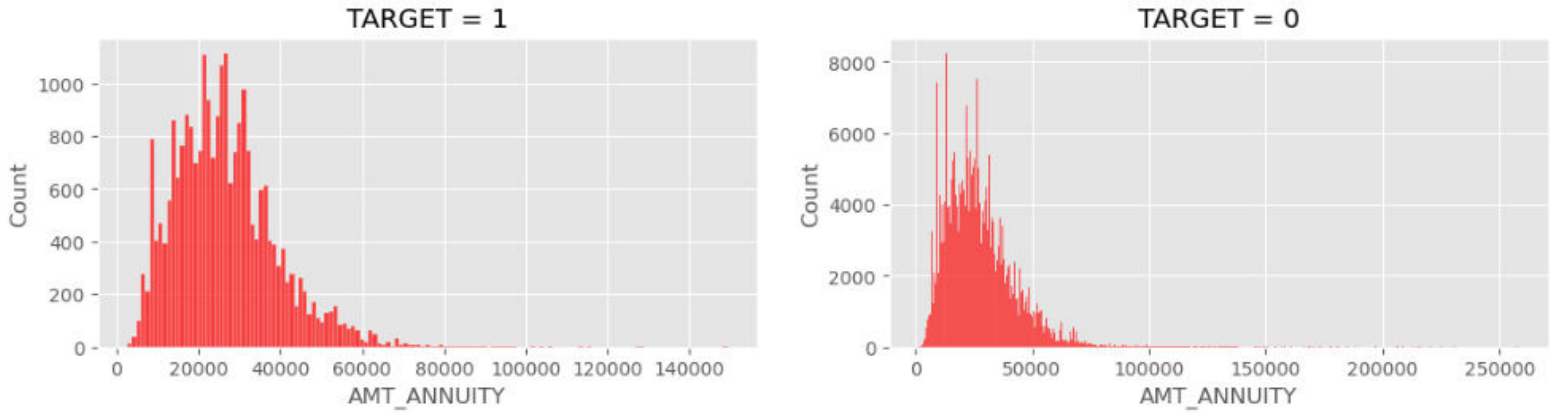
In [654]:
```
plt.style.use('ggplot')

plt.figure(figsize = (14,3))

plt.subplot(1,2,1)
sns.distplot(df1[df1.TARGET == 1]['AMT_INCOME_TOTAL'], color = 'Green');
plt.title('TARGET = 1')

plt.subplot(1,2,2)
sns.distplot(df1[df1.TARGET == 0]['AMT_INCOME_TOTAL'], color = 'Green')
plt.title('TARGET = 0');
```



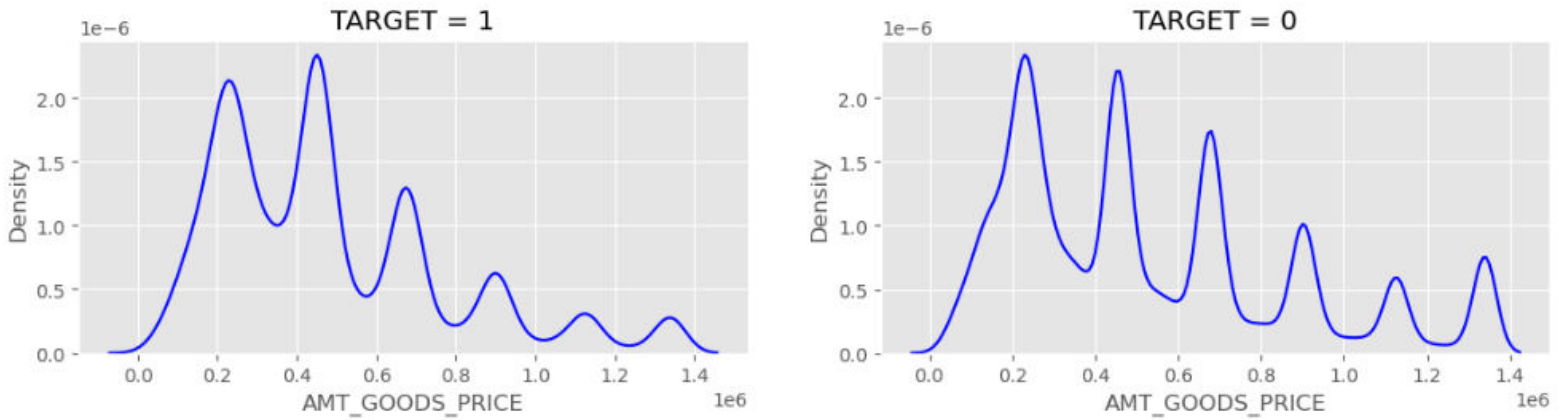
In [656]:
```
plt.style.use('ggplot')

plt.figure(figsize = (14,3))

plt.subplot(1,2,1)
sns.distplot(df1[df1.TARGET == 1]['AMT_CREDIT'], color = 'Green');
plt.title('TARGET = 1')

plt.subplot(1,2,2)
sns.distplot(df1[df1.TARGET == 0]['AMT_CREDIT'], color = 'Green')
plt.title('TARGET = 0');
```



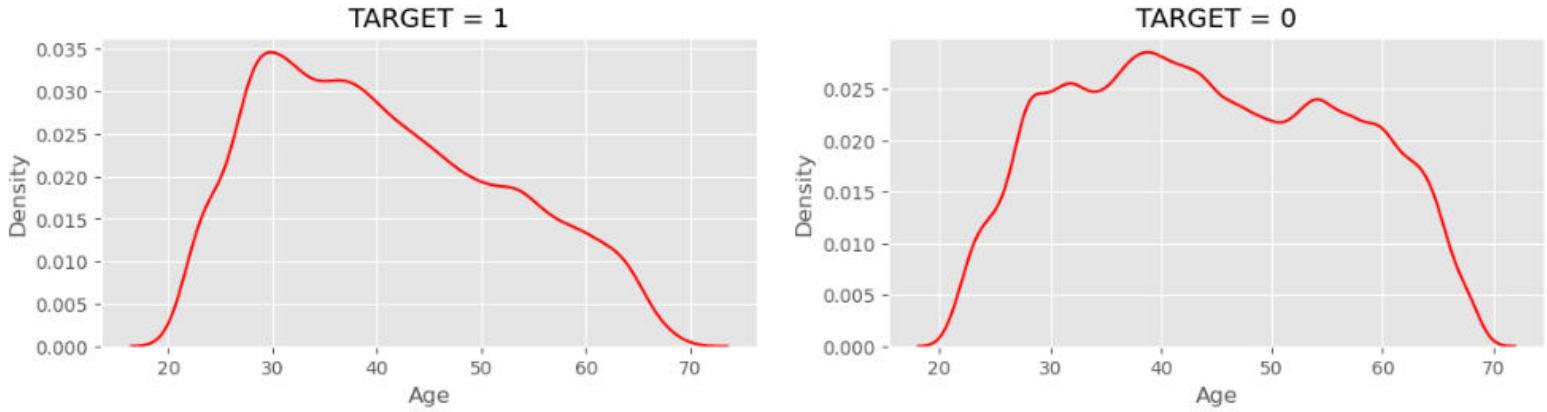
In [658]:
```
plt.style.use('ggplot')

plt.figure(figsize = (14,3))

plt.subplot(1,2,1)
sns.histplot(df1[df1.TARGET == 1]['AMT_ANNUITY'], color = 'Red');
plt.title('TARGET = 1')

plt.subplot(1,2,2)
sns.histplot(df1[df1.TARGET == 0]['AMT_ANNUITY'], color = 'Red');
plt.title('TARGET = 0');
```



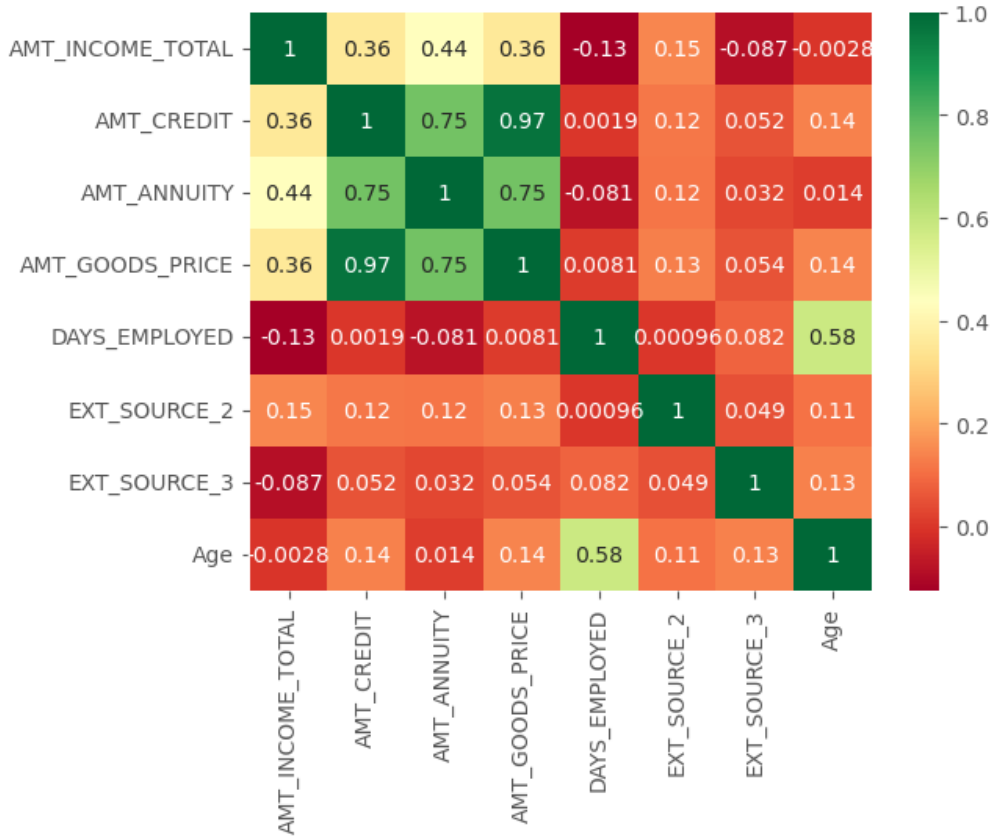
In [660]:
```
plt.style.use('ggplot')

plt.figure(figsize = (14,3))

plt.subplot(1,2,1)
sns.kdeplot(df1[df1.TARGET == 1]['AMT_GOODS_PRICE'], color = 'Blue');
plt.title('TARGET = 1')

plt.subplot(1,2,2)
sns.kdeplot(df1[df1.TARGET == 0]['AMT_GOODS_PRICE'], color = 'Blue');
plt.title('TARGET = 0');
```



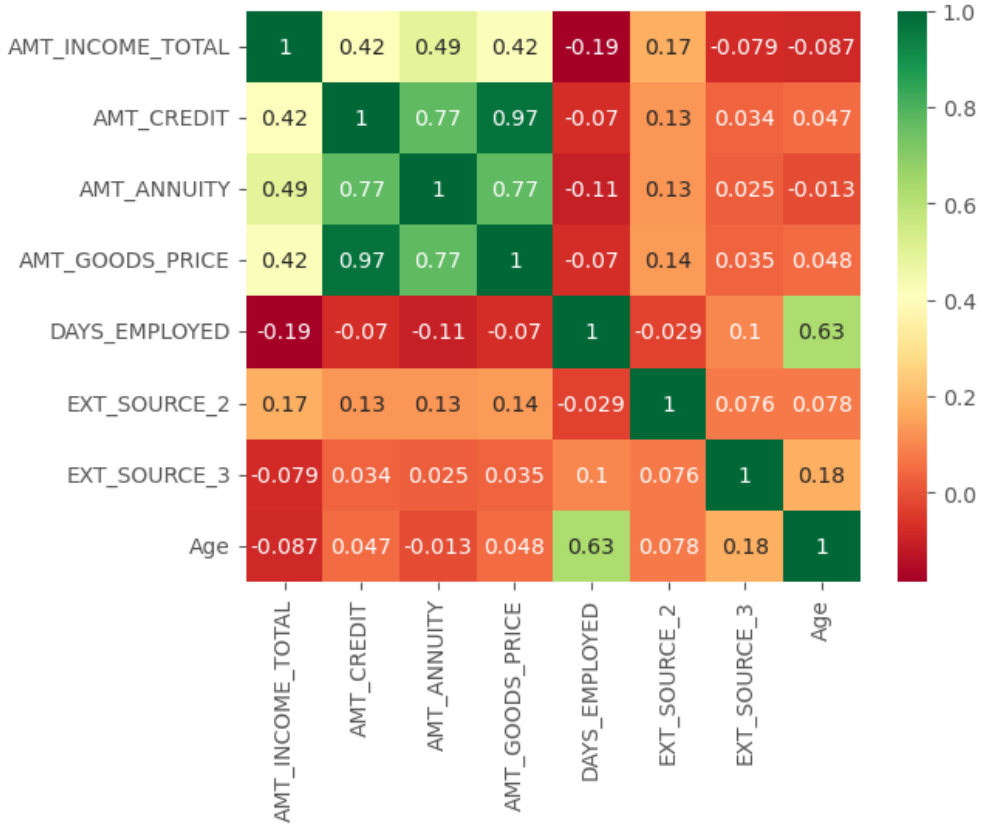
In [662]:
```
plt.style.use('ggplot')

plt.figure(figsize = (14,3))

plt.subplot(1,2,1)
sns.kdeplot(df1[df1.TARGET == 1]['Age'], color = 'Red');
plt.title('TARGET = 1')

plt.subplot(1,2,2)
sns.kdeplot(df1[df1.TARGET == 0]['Age'], color = 'Red');
plt.title('TARGET = 0');
```



In [669]:
```
sns.heatmap(df1[df1.TARGET == 1][num_columns].corr(), annot = True, cmap = 'RdYlGn');
```

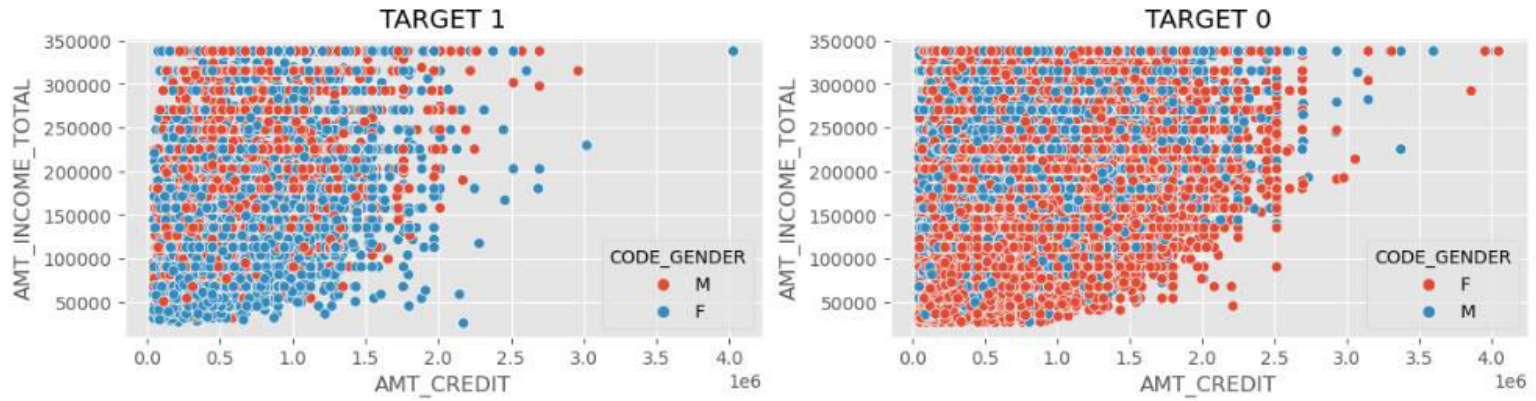


In [672]:
```
sns.heatmap(df1[df1.TARGET == 0][num_columns].corr(), annot = True, cmap = 'RdYlGn');
```



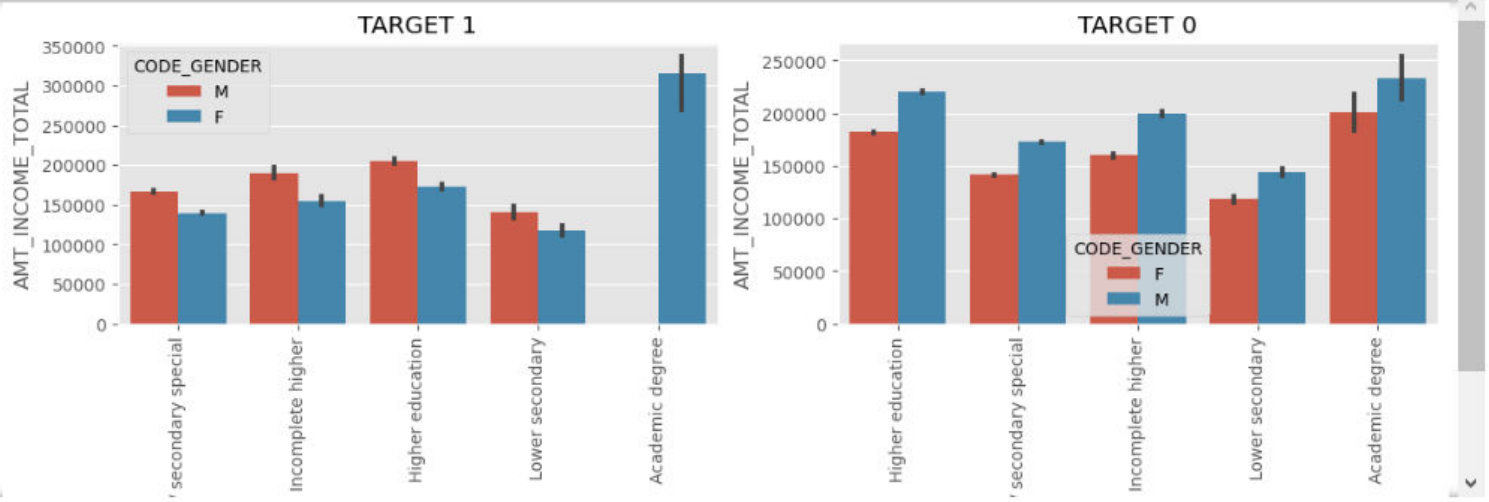
In [674]:
```
plt.figure(figsize = [14, 3])

plt.subplot(1,2,1)
sns.scatterplot(data = df1[df1.TARGET == 1], x = df1[df1.TARGET == 1]['AMT_CREDIT'], y = df1[df1.TARGET == 1]['AMT_INCOME_TOTAL']
                hue= df1[df1.TARGET == 1]['CODE_GENDER']);
plt.title('TARGET 1')

plt.subplot(1,2,2)
sns.scatterplot(data = df1[df1.TARGET == 0], x = df1[df1.TARGET == 0]['AMT_CREDIT'], y = df1[df1.TARGET == 0]['AMT_INCOME_TOTAL']
                hue= df1[df1.TARGET == 0]['CODE_GENDER']);
plt.title('TARGET 0')
```

Out[674]: Text(0.5, 1.0, 'TARGET 0')



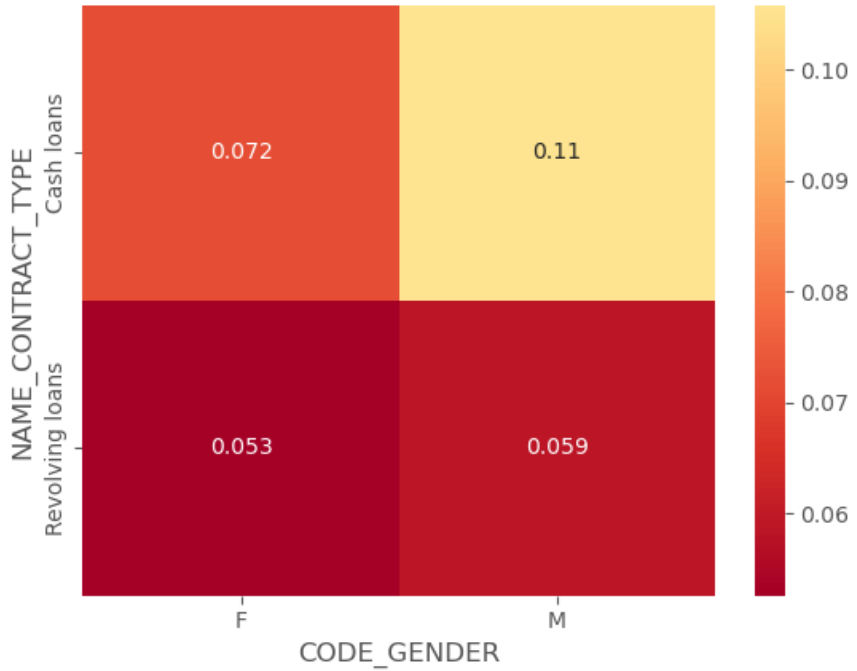
In [676]:

```
plt.figure(figsize=[14,3])

plt.subplot(1,2,1)
sns.barplot(data = df1[df1.TARGET == 1], x = df1[df1.TARGET == 1]['NAME_EDUCATION_TYPE'], y = df1[df1.TARGET == 1]['AMT_INCOME_TO
                hue= df1[df1.TARGET == 1]['CODE_GENDER'])
plt.xticks(rotation = 90)
plt.title('TARGET 1')

plt.subplot(1,2,2)
sns.barplot(data = df1[df1.TARGET == 0], x = df1[df1.TARGET == 0]['NAME_EDUCATION_TYPE'], y = df1[df1.TARGET == 0]['AMT_INCOME_TO
                hue= df1[df1.TARGET == 0]['CODE_GENDER'])
plt.xticks(rotation = 90)
plt.title('TARGET 0');
```

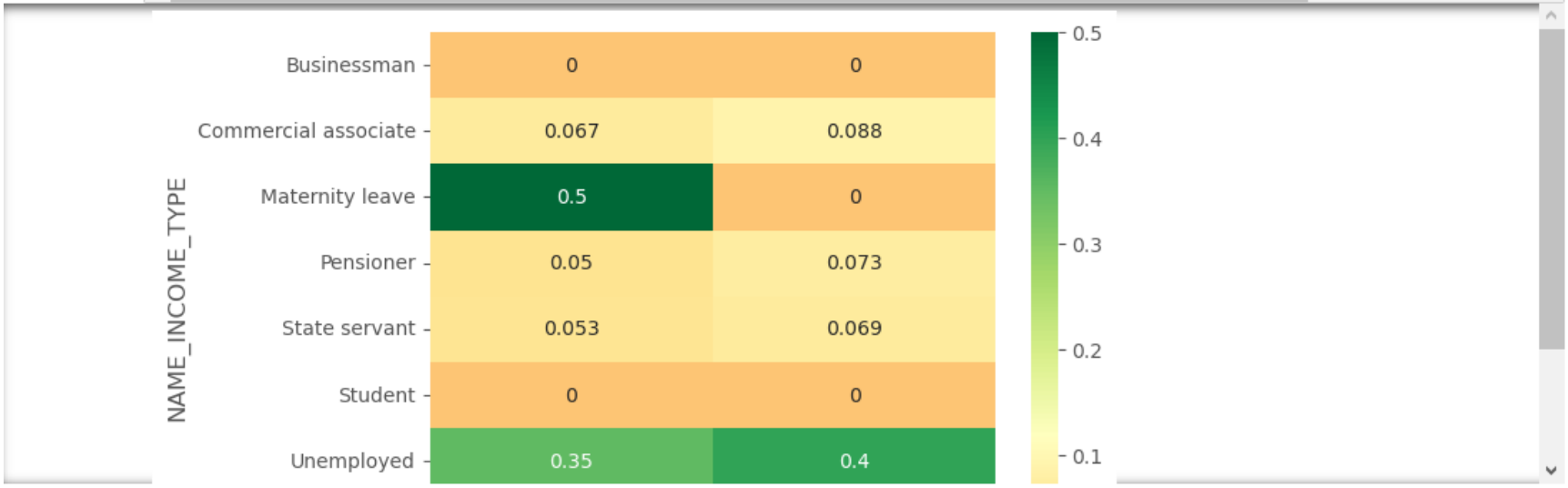


In [677]:

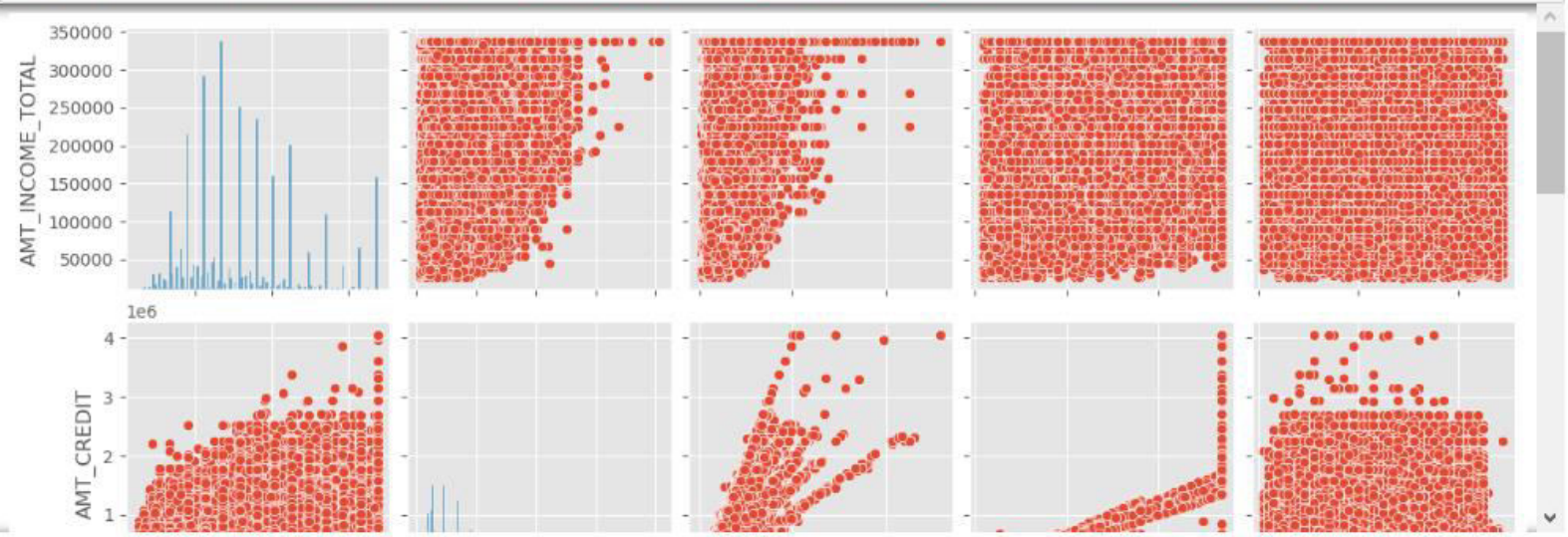
In [687]:
```
sns.heatmap(pd.pivot_table(df1 , index = 'NAME_CONTRACT_TYPE', columns = 'CODE_GENDER' , values = 'TARGET'), cmap = 'RdYlGn', anr
```



```
In [689]: sns.heatmap(pd.pivot_table(df1 , index = 'NAME_INCOME_TYPE', columns = 'CODE_GENDER' , values = 'TARGET'), cmap = 'RdYlGn', annot
```



```
In [690]: sns.pairplot(df1[df1['TARGET'] == 0][['AMT_INCOME_TOTAL', 'AMT_CREDIT', 'AMT_ANNUITY', 'AMT_GOODS_PRICE','Age']] );
```

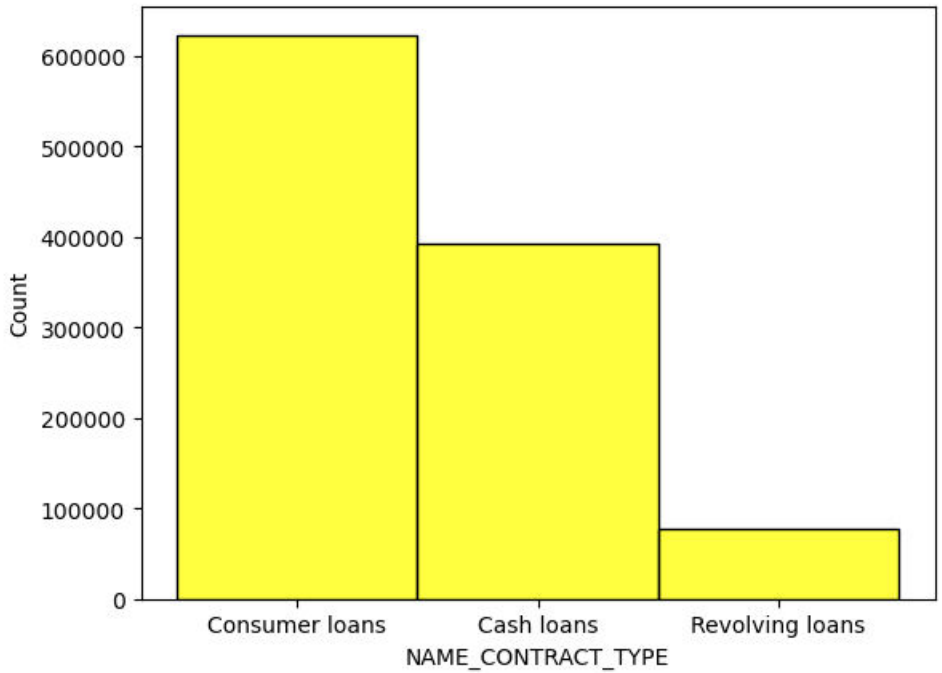



AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
Age
1e6
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
Age

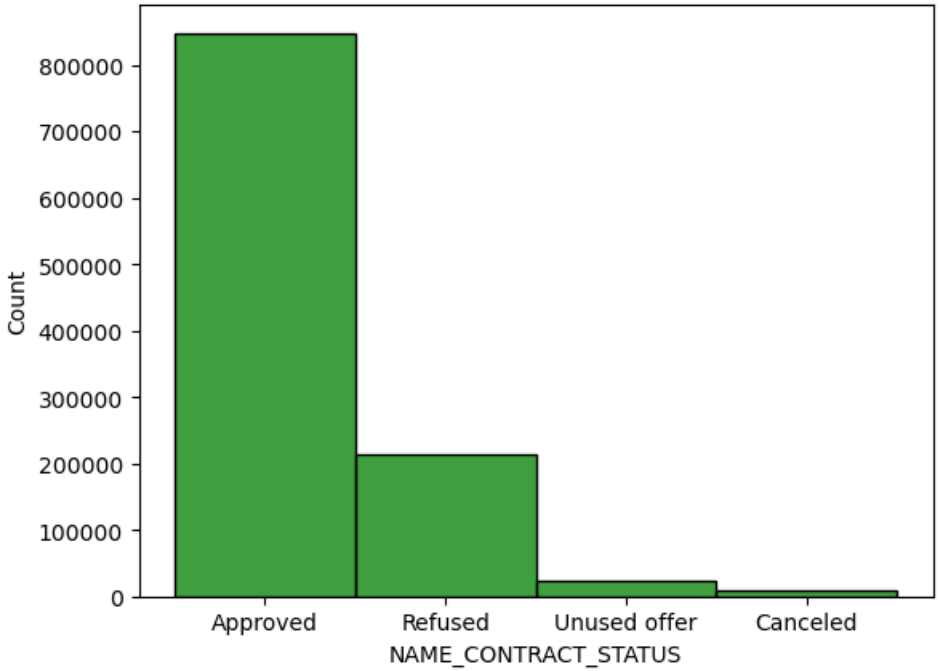
PREVIOUS DATASET

- Univariate analysis

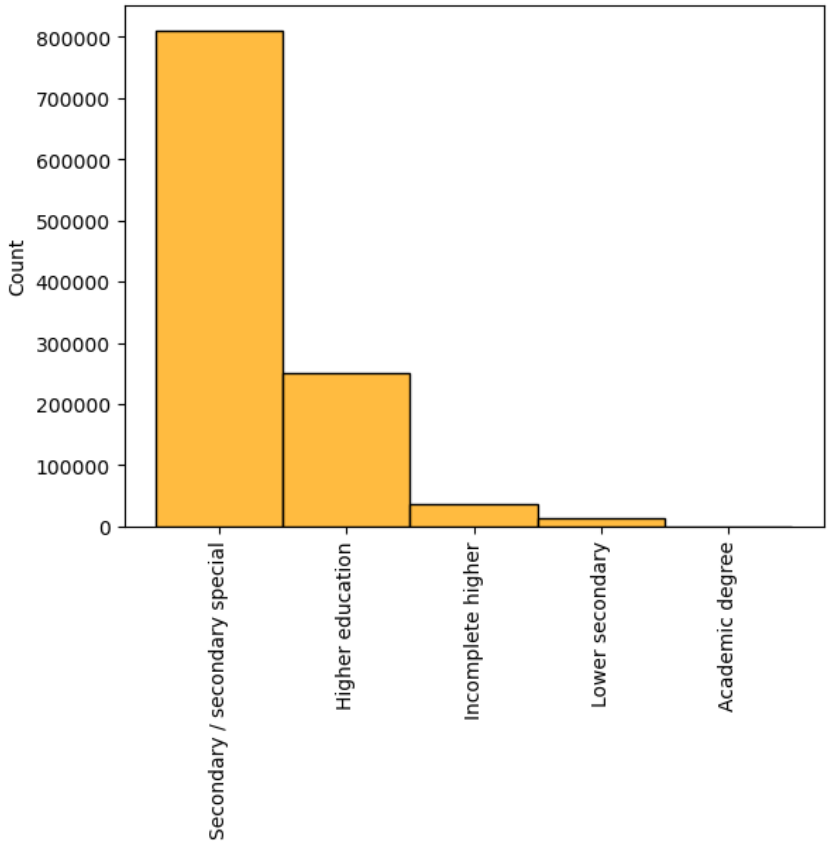
In [764]:
```
# df3['NAME_CONTRACT_TYPE'].dtype

plt.style.use('default')

sns.histplot(data=df3['NAME_CONTRACT_TYPE'], multiple='dodge', discrete=True, color = 'yellow');
```



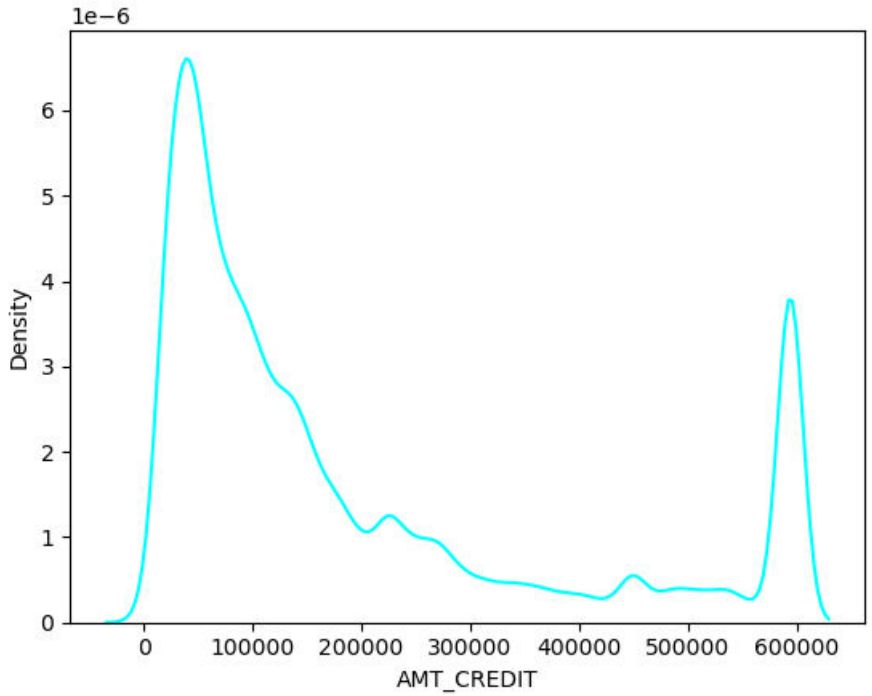
In [765]:
```
sns.histplot(data=df3['NAME_CONTRACT_STATUS'], multiple='dodge', discrete=True, color = 'green');
```



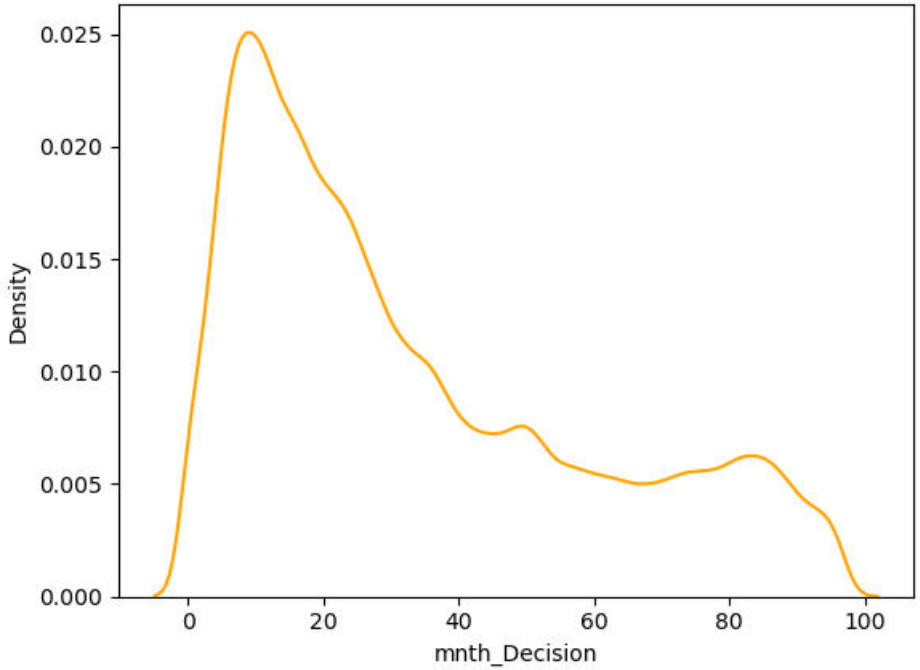
In [766]:
```
sns.histplot(data=df3['NAME_EDUCATION_TYPE'], multiple='dodge', discrete=True, color = 'orange');
plt.xticks(rotation = 90);
```



In [767]:
```
sns.kdeplot(df3['AMT_CREDIT'], color = 'cyan');

# loan amounts till 150,000 has been provided to clients most number of times
# loan amounts from 580,000 to  630,000 has been provided to clients most number of times as well.
```

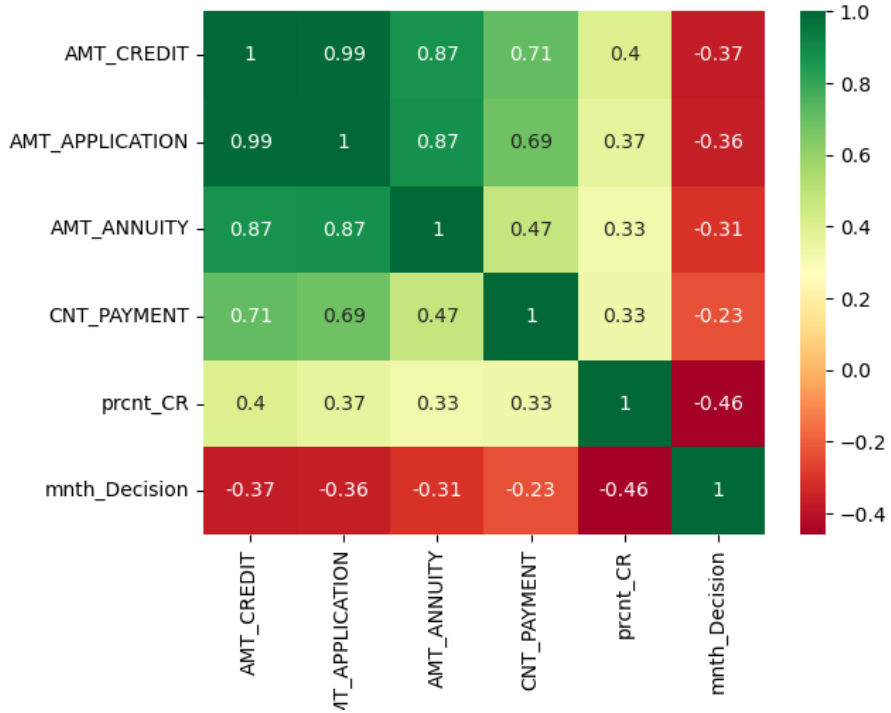


In [768]:
```
sns.kdeplot(df3['mnth_Decision'], color = 'Orange');
```

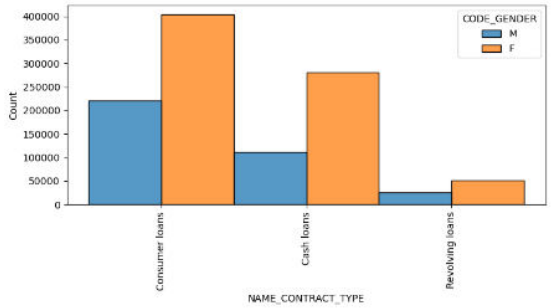


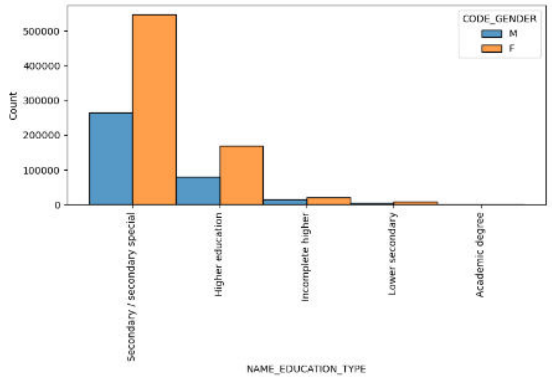
Bivariate analysis

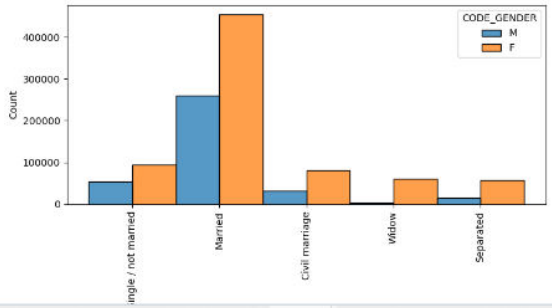
```
In [ ]: # Analysing 'AMT_CREDIT', 'AMT_APPLICATION', 'AMT_ANNUITY', 'CNT_PAYMENT', 'prcnt_CR', 'mnth_Decision'

In [ ]: df3[['AMT_CREDIT', 'AMT_APPLICATION', 'AMT_ANNUITY', 'CNT_PAYMENT', 'prcnt_CR', 'mnth_Decision']].corr()

In [769]: sns.heatmap(df3[['AMT_CREDIT', 'AMT_APPLICATION', 'AMT_ANNUITY', 'CNT_PAYMENT', 'prcnt_CR', 'mnth_Decision']].corr(), cmap = 'RdY
```



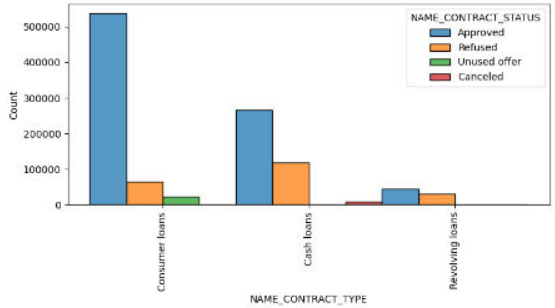
In [770]:
```
j=1
plt.figure(figsize=(21,11))

for i in ['NAME_CONTRACT_TYPE','NAME_EDUCATION_TYPE','NAME_FAMILY_STATUS']:
    plt.subplot(2,2,j)
    sns.histplot(data=df3, x=i, hue = 'CODE_GENDER', multiple='dodge', discrete=True);
    plt.subplots_adjust(wspace=0.4, hspace=1)
    plt.xticks(rotation = 90)
    j=j+1
```

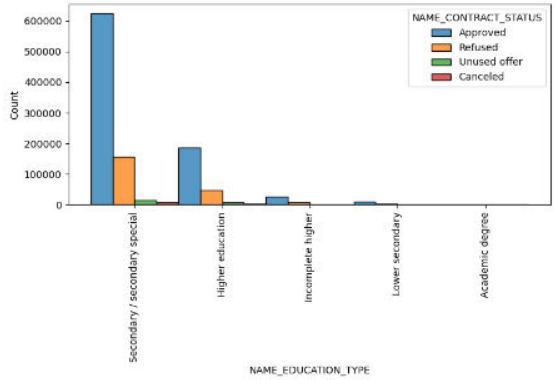


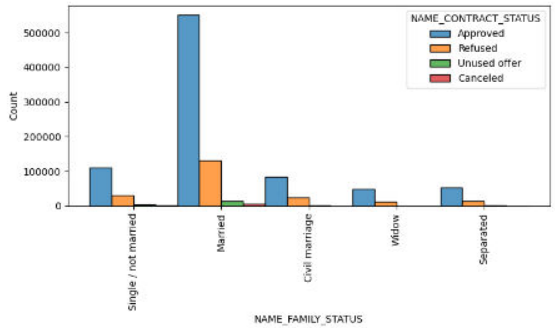




In [771]:
```
j=1
plt.figure(figsize=(21,11))

for i in ['NAME_CONTRACT_TYPE','NAME_EDUCATION_TYPE','NAME_FAMILY_STATUS']:
    plt.subplot(2,2,j)
    sns.histplot(data=df3, x=i, hue = 'NAME_CONTRACT_STATUS', multiple='dodge', discrete=True);
    plt.subplots_adjust(wspace=0.4, hspace=1)
    plt.xticks(rotation = 90)
    j=j+1
```







In [772]:
```
sns.histplot(x = "NAME_CONTRACT_STATUS", hue = "CODE_GENDER", multiple = 'dodge', data = df3 );
```