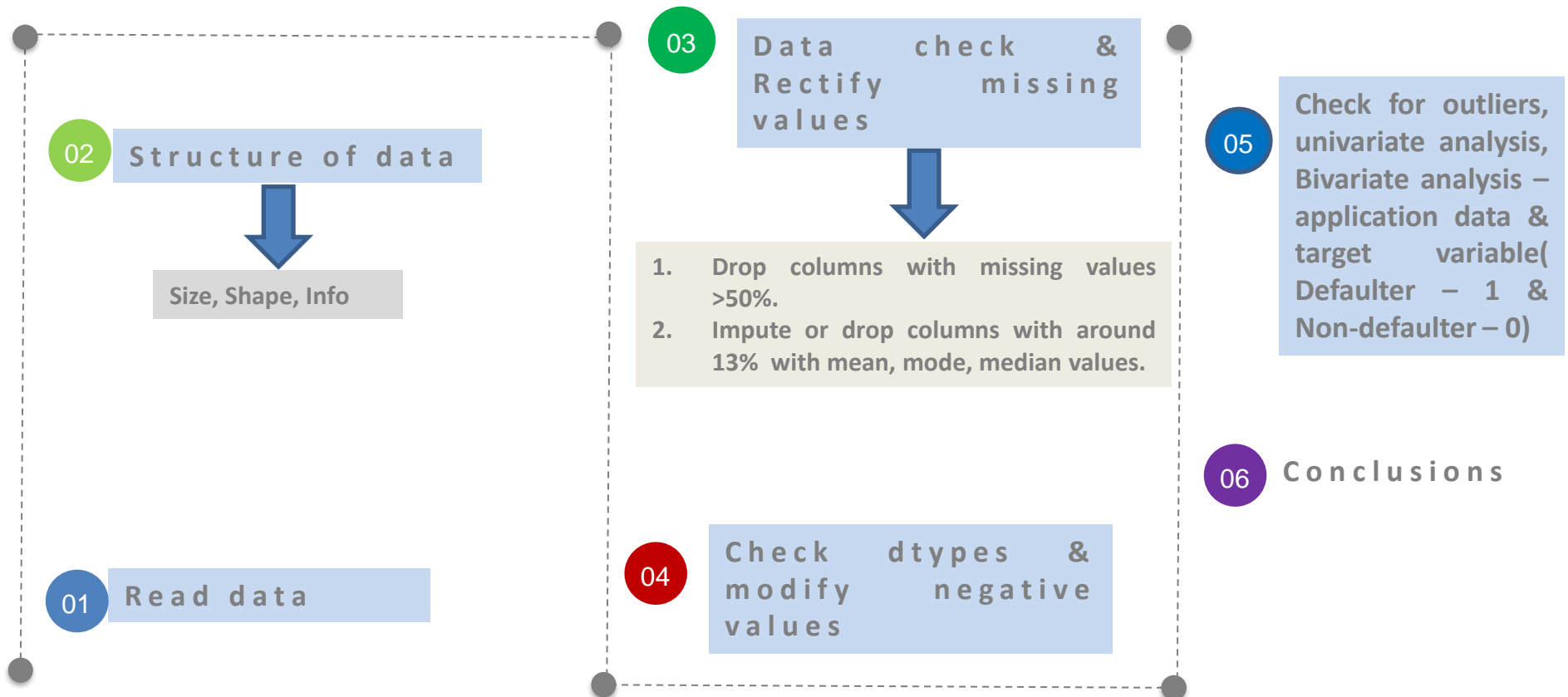


EDA CASE STUDY

Report

METHODOLOGY

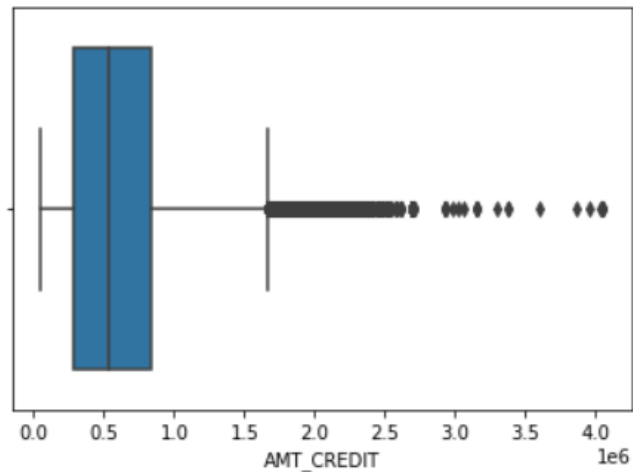
EDA ASSIGNMENT



NUMERICAL COLUMNS

Outliers

```
sns.boxplot(inp0['AMT_CREDIT'])  
plt.show()
```

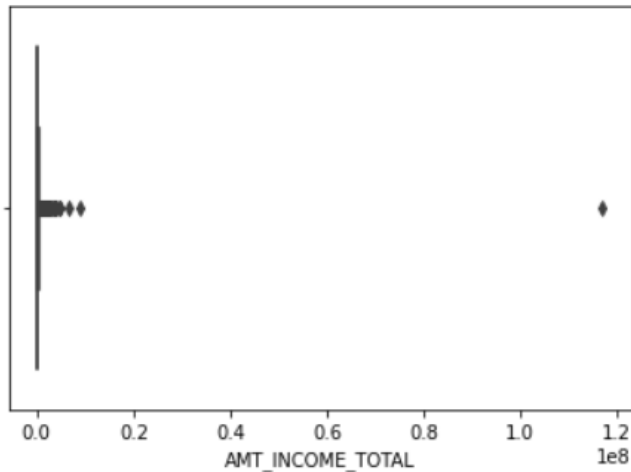


The interquartile range of data for amount credited has many outliers. The credit amount of above 16 million and above are the outliers. The data is positively skewed.

NUMERICAL COLUMNS

Outliers

```
sns.boxplot(inp0['AMT_INCOME_TOTAL'])  
plt.show()
```



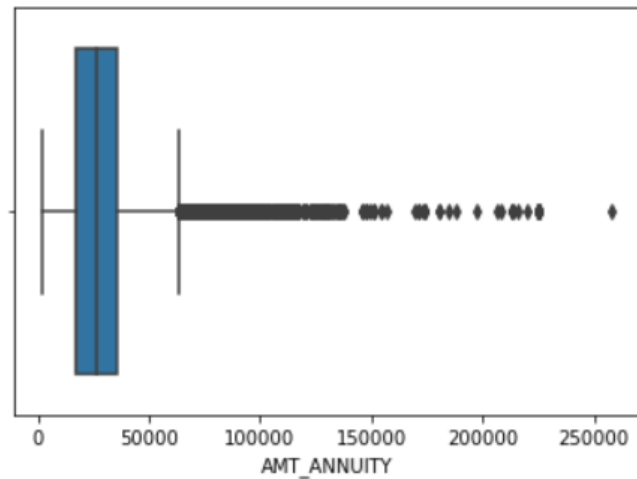
```
inp0[inp0.AMT_INCOME_TOTAL > 100000000]  # Outlier row
```

SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	NAME_INCOME_TYPE
12840	1	Cash loans	F	117000000.0	562491.0	26194.5	454500.0	

NUMERICAL COLUMNS

Outliers

```
sns.boxplot(inp0['AMT_ANNUIITY'])  
plt.show()
```

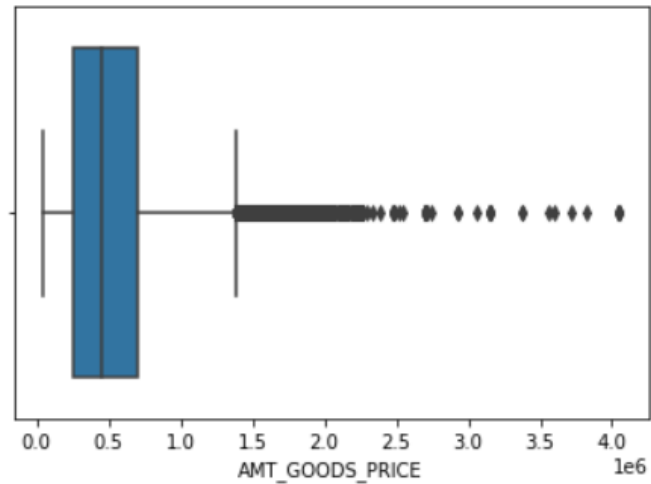


The interquartile range of data for amount credited has many outliers. The annuity amount of above 60000 and above are the outliers. The data is positively skewed.

NUMERICAL COLUMNS

Outliers

```
sns.boxplot(inp0['AMT_GOODS_PRICE'])  
plt.show()
```

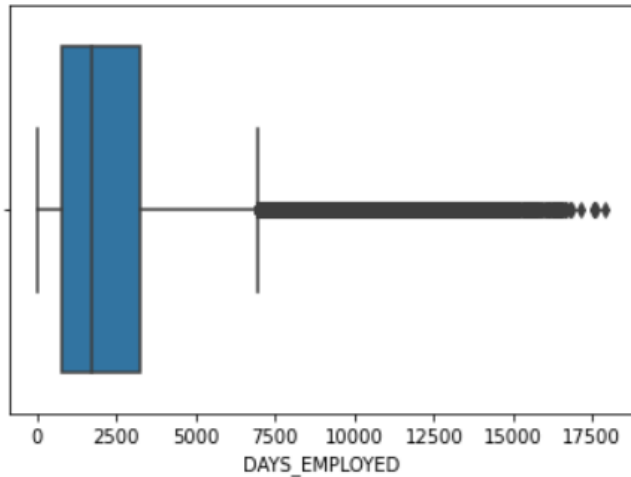


The interquartile range of data for amount credited has many outliers. The goods price amount of above 14.5 million and above are the outliers. The data is positively skewed.

NUMERICAL COLUMNS

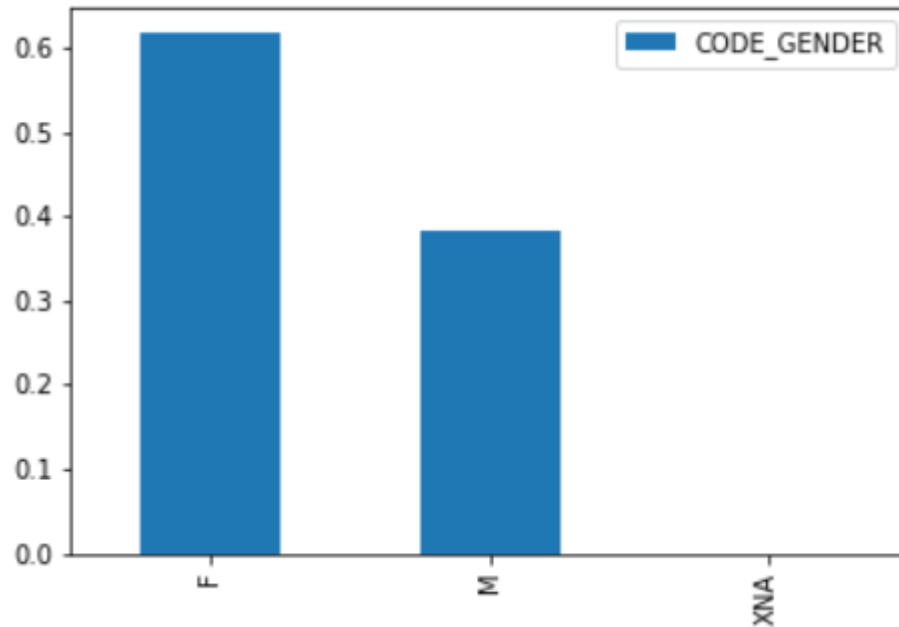
Outliers

```
sns.boxplot(inp0['DAYS_EMPLOYED'])  
plt.show()
```



The interquartile range of data for amount credited has many outliers. The clients who are employed 7300 days and above are the outliers. The data is positively skewed.

IMBALANCE PERCENTAGE

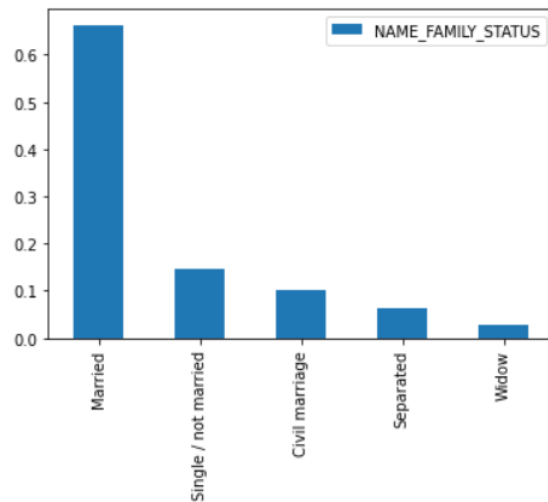


```
inp0['CODE_GENDER'].value_counts(normalize=True)*100 ## Percentage of Female clients(61%) are more than male(38%) clients.
```

```
F      61.670388
M      38.328510
XNA     0.001103
Name: CODE_GENDER, dtype: float64
```


IMBALANCE PERCENTAGE

```
inp0['NAME_FAMILY_STATUS'].value_counts(normalize=True).plot.bar()  
plt.legend()  
plt.show()
```

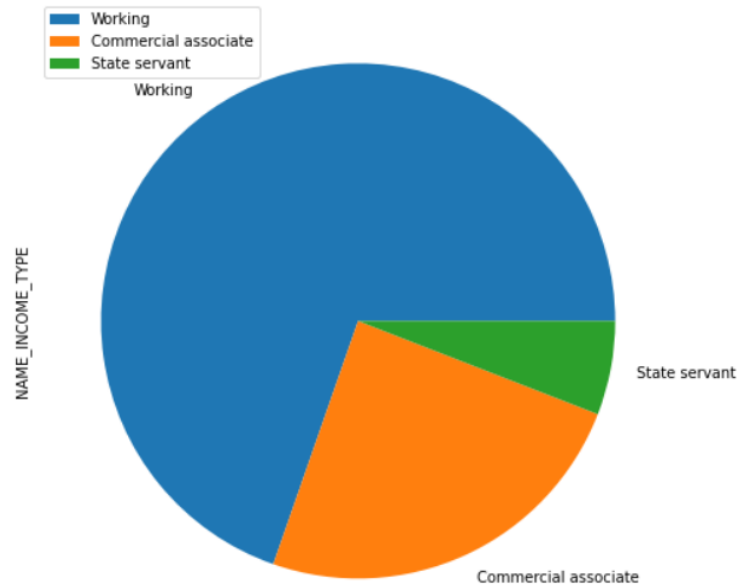


```
inp0['NAME_FAMILY_STATUS'].value_counts(normalize=True)*100
```

```
Married          66.317153  
Single / not married  14.613964  
Civil marriage    10.060366  
Separated         6.385512  
Widow             2.623005  
Name: NAME_FAMILY_STATUS, dtype: float64
```

Clients Family status : Married (66%) are the highest followed by Single/ not married(14%) & Civil marriage(3.61%) respectively.

UNIVARIATE ANALYSIS – TARGET – (1-DEFaulter)

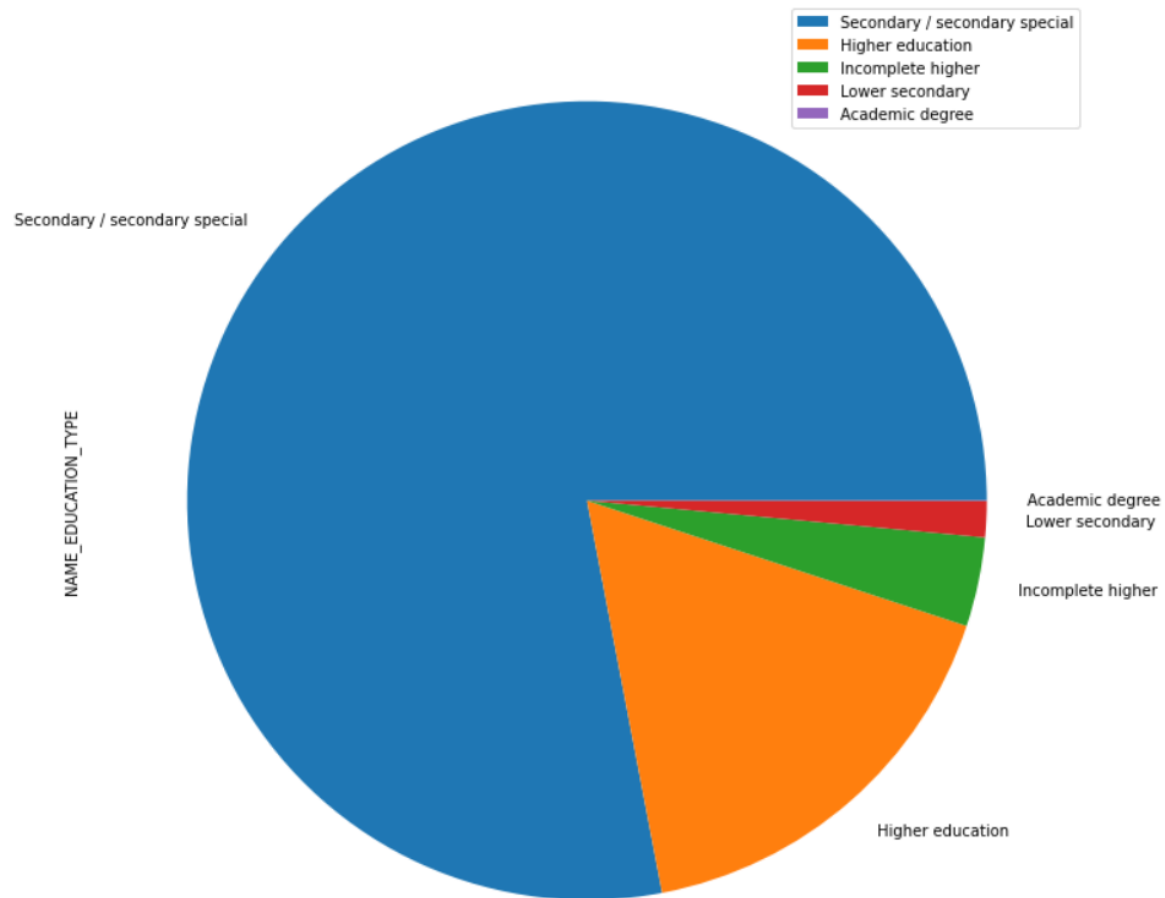


```
3]: df_1['NAME_INCOME_TYPE'].value_counts(normalize=True)*100
```

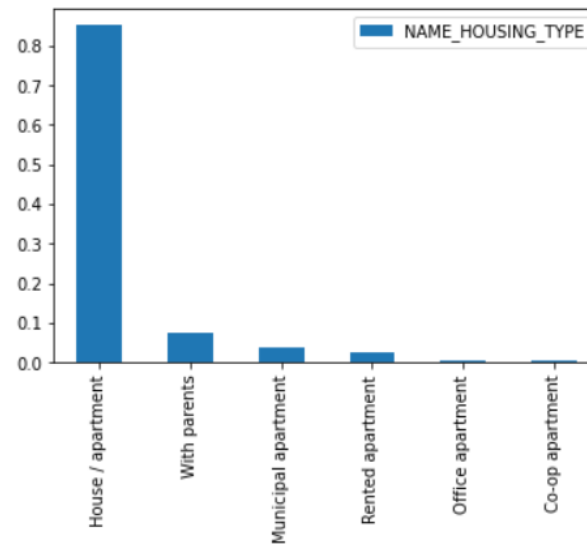
```
3]: Working          69.646243
     Commercial associate  24.462172
     State servant      5.891584
     Name: NAME_INCOME_TYPE, dtype: float64
```

```
3]: ## Percentage of defaulters - Income type variable : 1. Working (69%) 2. Commercial associate (24.4%) & 3. State servant(5.89%)
```

UNIVARIATE ANALYSIS – TARGET – (1-DEFAULTER)



UNIVARIATE ANALYSIS – TARGET – (1-DEFALTER)

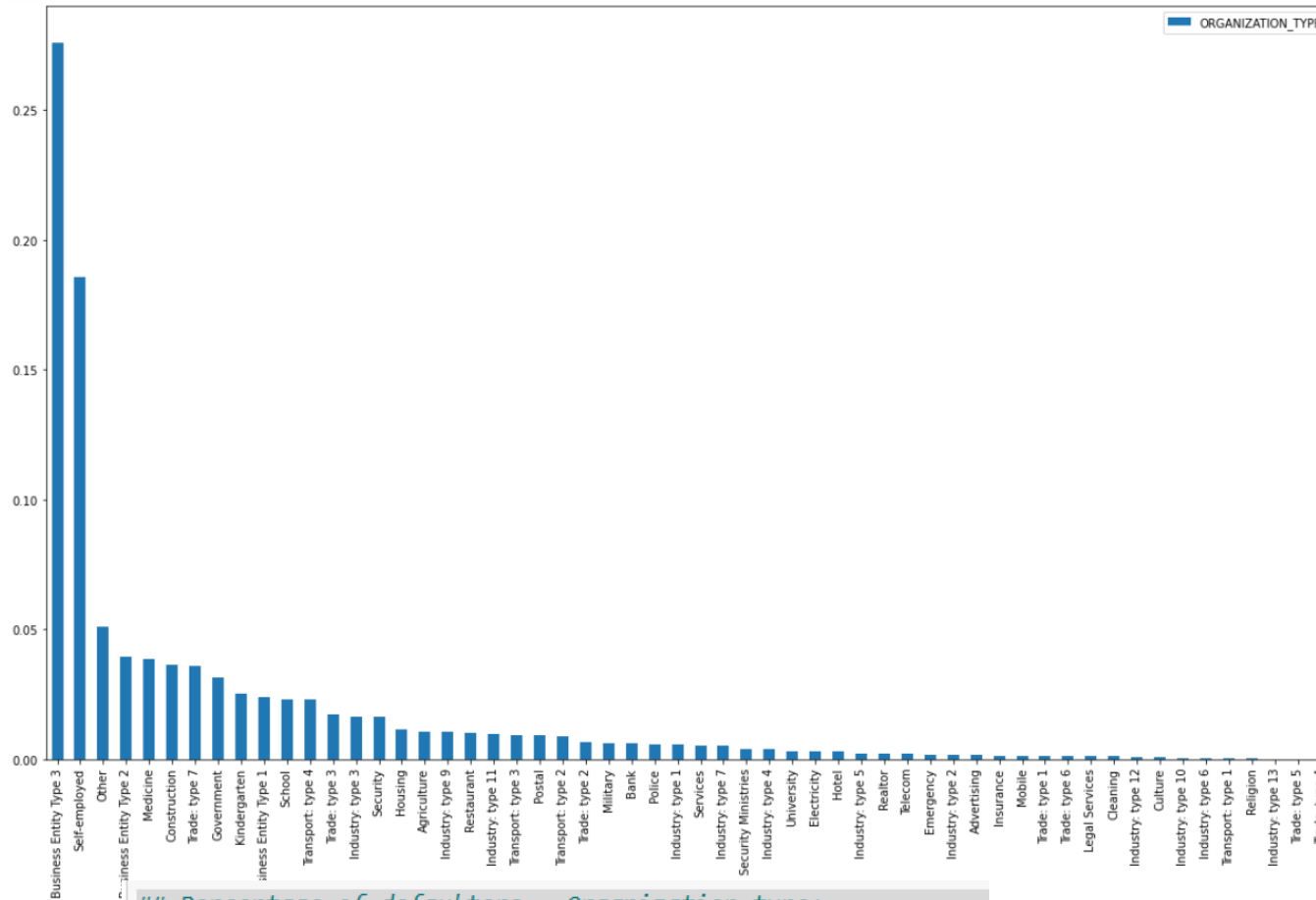


```
df_1['NAME_HOUSING_TYPE'].value_counts(normalize=True)*100
```

```
House / apartment      85.208919
With parents           7.480547
Municipal apartment    3.714118
Rented apartment       2.510953
Office apartment       0.699667
Co-op apartment        0.385797
Name: NAME_HOUSING_TYPE, dtype: float64
```

```
## Percentage of defaulters - Housing type:
#1. House / apartment (85%)
#2. With parents (7%) &
#3. Municipal apartment(3.7%)
```

UNIVARIATE ANALYSIS – TARGET – (1-DEFALTER)



Percentage of defaulters - Organization type:

#1. Business entity - Type 3 (27.61%)

#2. Self employed (18.55%) &

#3. Other (5.08%)

Bottom two - Trade: type 5 & Trade: type 4 - 0.013%