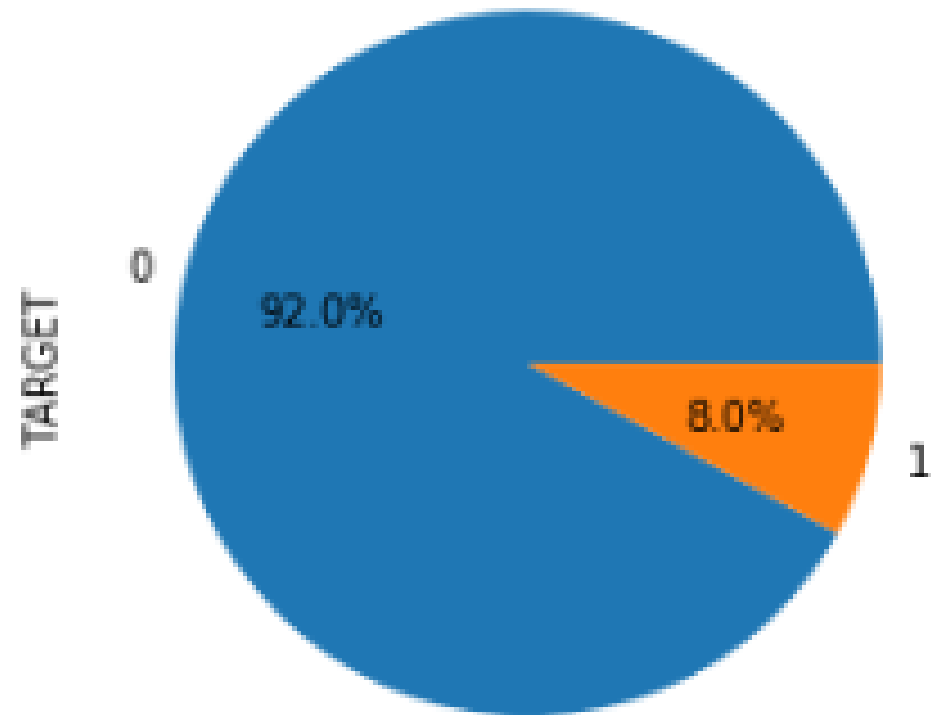




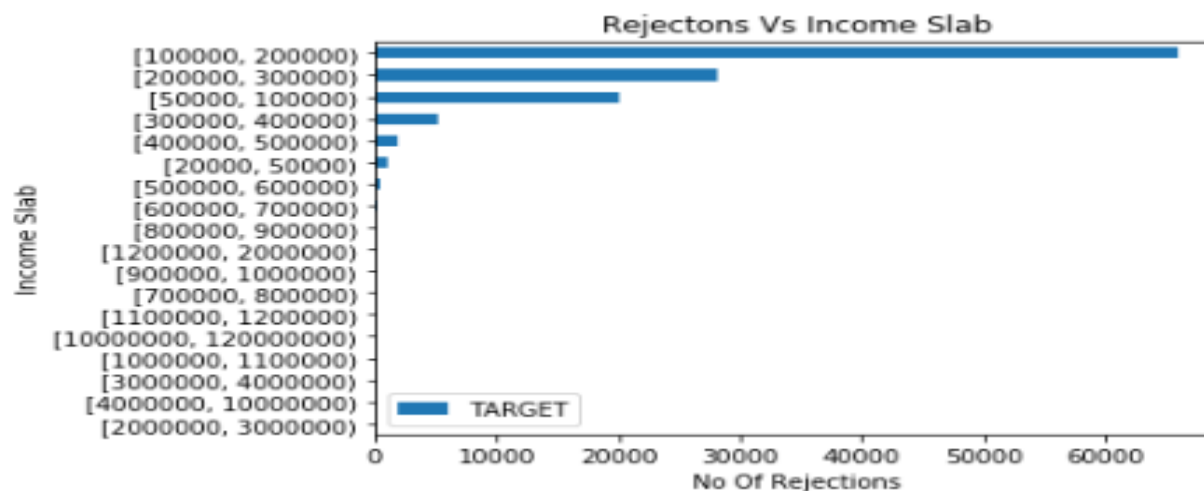
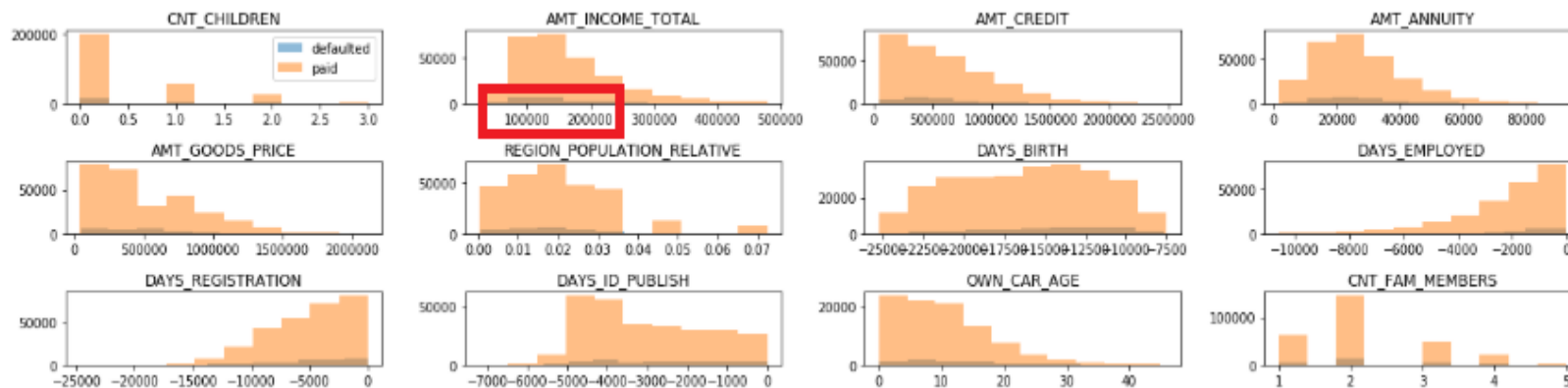
EDA CASE STUDY

Proof of Imbalance or class 0 and 1



Data Imbalace

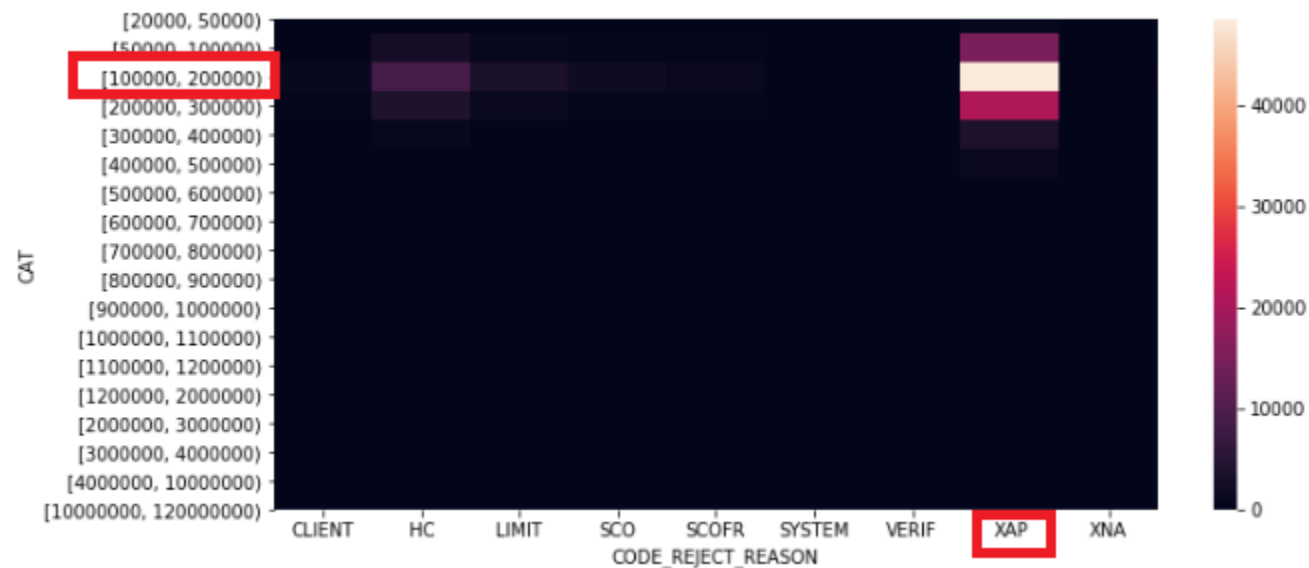
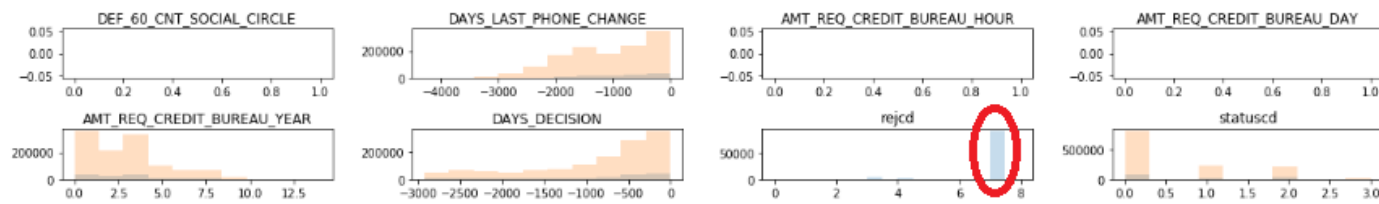
- Majority Class 0
- Minority Class 1



Univariate Analysis

Removing Outliers using sigma clipping method from each numerical column and plotting to see the difference in distribution of features for class 0 and class 1 to identify which feature is able to differentiate between defaulter and who are able to pay better.

Income slab [100000,200000) has the maximum rejections.



Rejection Code XAP

XAP is the major reason for rejection which meaning is not explained but is a clear feature to differentiate between who could pay and having payment difficulty.

```

fig, axes = plt.subplots(18, 4, figsize=(17, 22))
ax = axes.ravel()
for i, col in zip(range(72), studv1):
    quartiles = np.percentile(dfap3[col].dropna(), [25, 50, 75])
    mu = quartiles[1]
    sig = 0.74 * (quartiles[2] - quartiles[0])

    cola = dfap3[(dfap3[col] > (mu - 5 * sig)) & (dfap3[col] < (mu + 5 * sig))][col]
    ax[i].hist(cola, alpha=.25, label="defaulted")
    quartiles = np.percentile(dfap4[col].dropna(), [25, 50, 75])

    mu = quartiles[1]
    sig = 0.74 * (quartiles[2] - quartiles[0])
    colb = dfap4[(dfap4[col] > (mu - 5 * sig)) & (dfap4[col] < (mu + 5 * sig))][col]
    ax[i].hist(colb, alpha=.25, label="paid")
    ax[i].set_title(col)
ax[0].legend()
fig.tight_layout()

```

Sigma Clipping

Method used to remove outliers from numerical column plotting it.

Var1	Var2	Correlation
AMT_GOODS_PRICE	AMT_CREDIT	0.983103
CNT_FAM_MEMBERS	CNT_CHILDREN	0.885484
AMT_GOODS_PRICE	AMT_ANNUITY	0.752699
AMT_ANNUITY	AMT_CREDIT	0.752195
REGION_RATING_CLIENT_W_CITY	REGION_POPULATION_RELATIVE	0.446977
REGION_RATING_CLIENT_W_CITY	AMT_ANNUITY	0.089291
REGION_RATING_CLIENT_W_CITY	AMT_GOODS_PRICE	0.077191
REGION_POPULATION_RELATIVE	AMT_GOODS_PRICE	0.076049
AMT_ANNUITY	CNT_FAM_MEMBERS	0.075711
REGION_POPULATION_RELATIVE	AMT_ANNUITY	0.071690

*Top 10 Correlation
for class 1 /
Minority Class*

Var1	Var2	Correlation
AMT_GOODS_PRICE	AMT_CREDIT	0.987250
CNT_FAM_MEMBERS	CNT_CHILDREN	0.878571
AMT_GOODS_PRICE	AMT_ANNUITY	0.776686
AMT_ANNUITY	AMT_CREDIT	0.771309
REGION_RATING_CLIENT_W_CITY	REGION_POPULATION_RELATIVE	0.537301
AMT_ANNUITY	AMT_INCOME_TOTAL	0.418953
AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.349462
AMT_CREDIT	AMT_INCOME_TOTAL	0.342799
REGION_RATING_CLIENT_W_CITY	AMT_INCOME_TOTAL	0.200466
REGION_POPULATION_RELATIVE	AMT_INCOME_TOTAL	0.167851

*Top 10 Correlation
for Majority Class 0*

Thank
You!