Credit EDA Case Study

Business Objectives

- This case study aims to identify patterns which indicate if a client has difficulty paying their instalments 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.

Goals and Objectives

Data

This dataset has 3 files as explained below:

1. *'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.**

2. *'previous_application.csv'* contains information about the client's previous loan data. It contains the data whether the previous application had been **Approved, Cancelled, Refused or Unused offer.**

3. *'columns_description.csv'* is data dictionary which describes the meaning of the variables.

Main Column

- 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)

EDA

An **exploratory data analysis** is the process of using statistical tools and ideas to examine data in order to describe their main features.
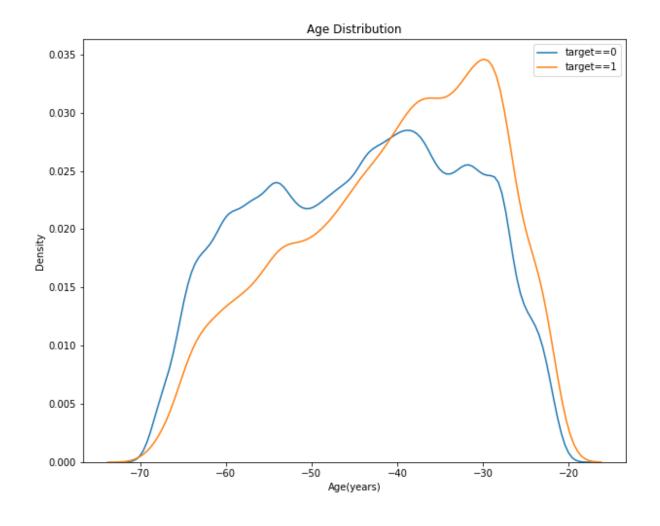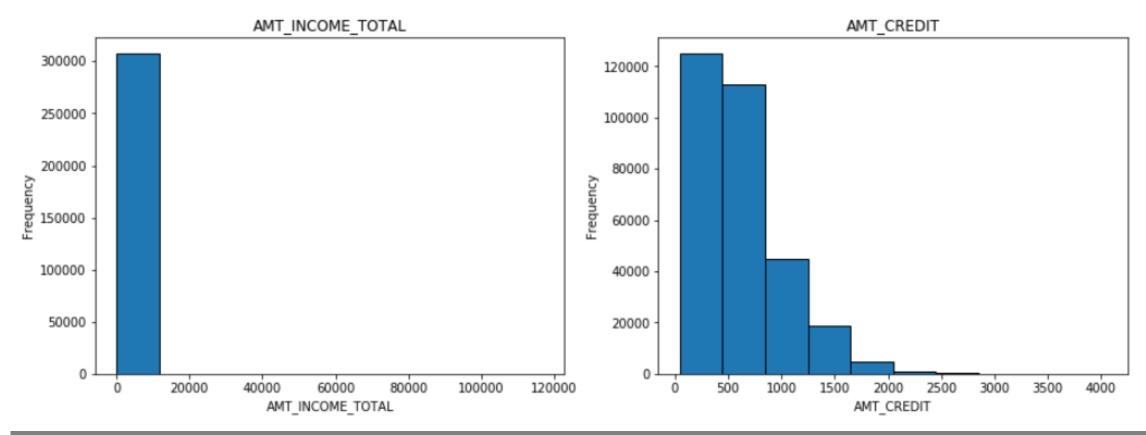
EXPLORING DATA

- Begin by examining each variable by itself. Then move on to study the relationships among the variables.

- Begin with a graph or graphs. Then add numerical summaries of specific aspects of the data.
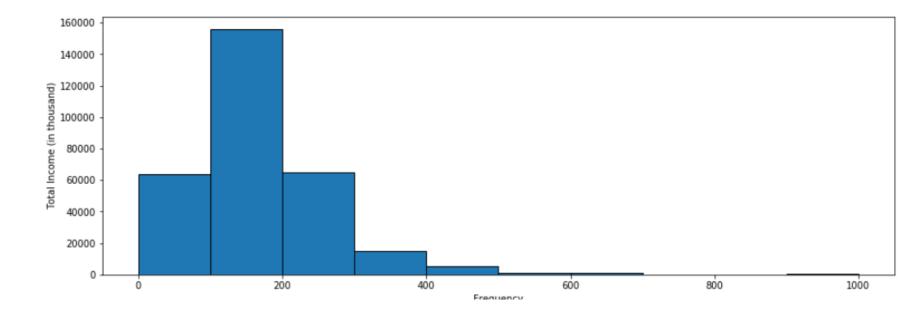
Age distribution for Target

- As can see in graph that 35 to 28 age we have more density for target 1 and lower for target0

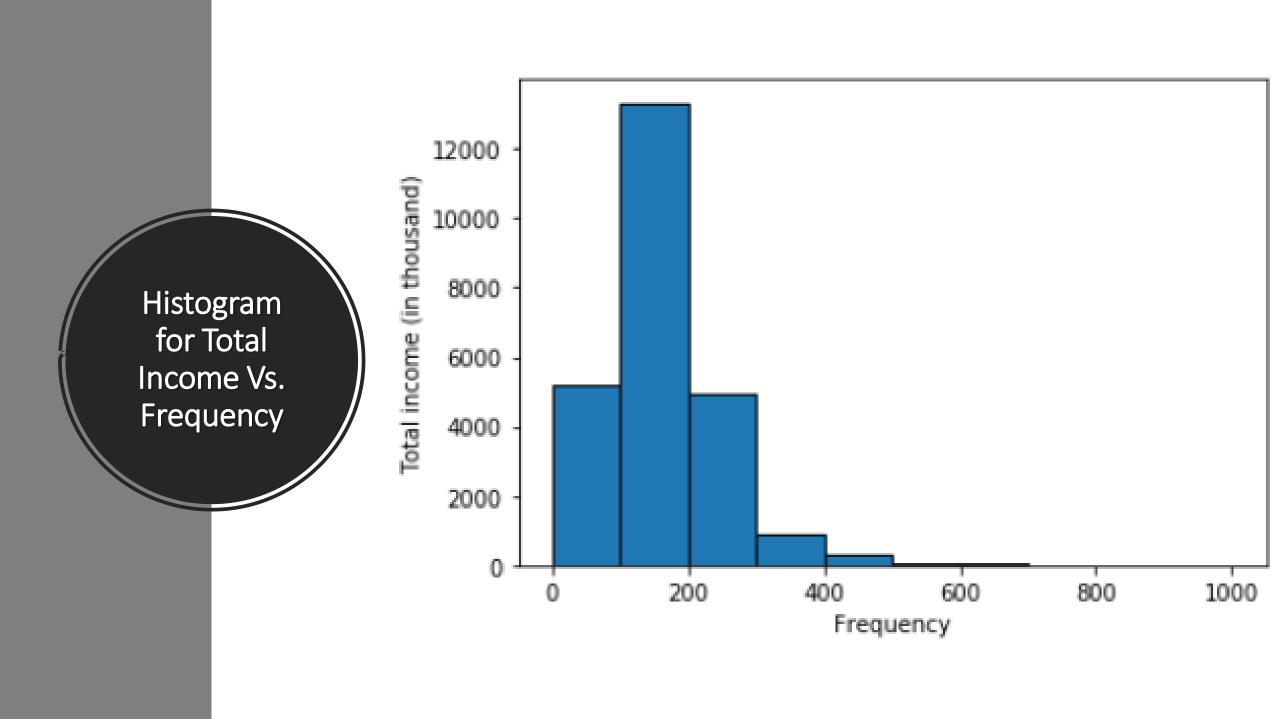- For 65 to 50 age density is higher as compare to target 1
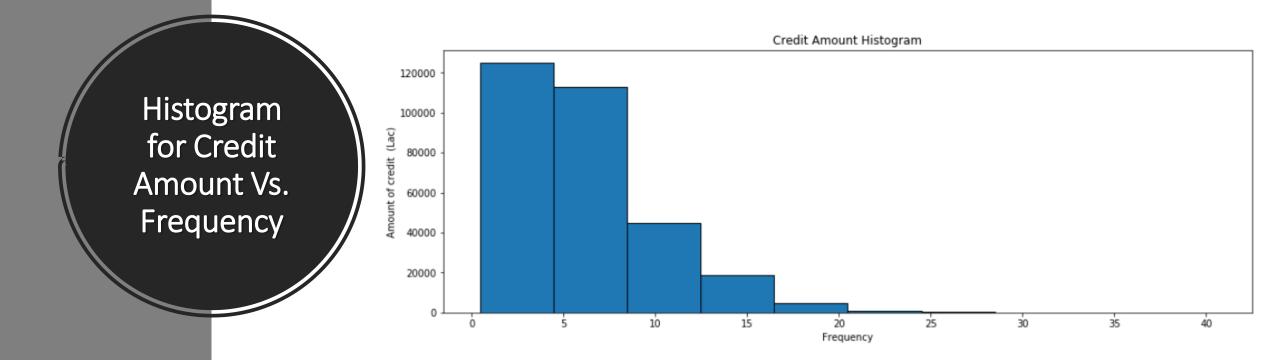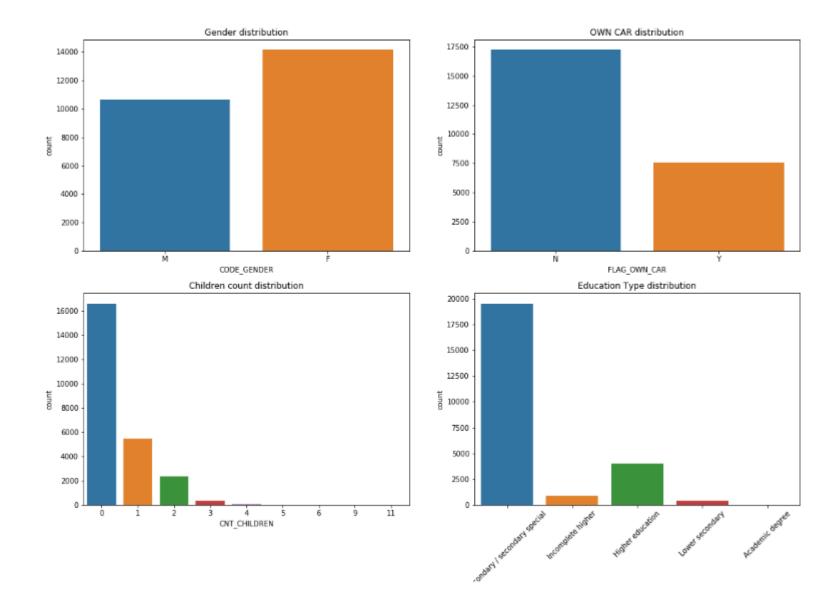
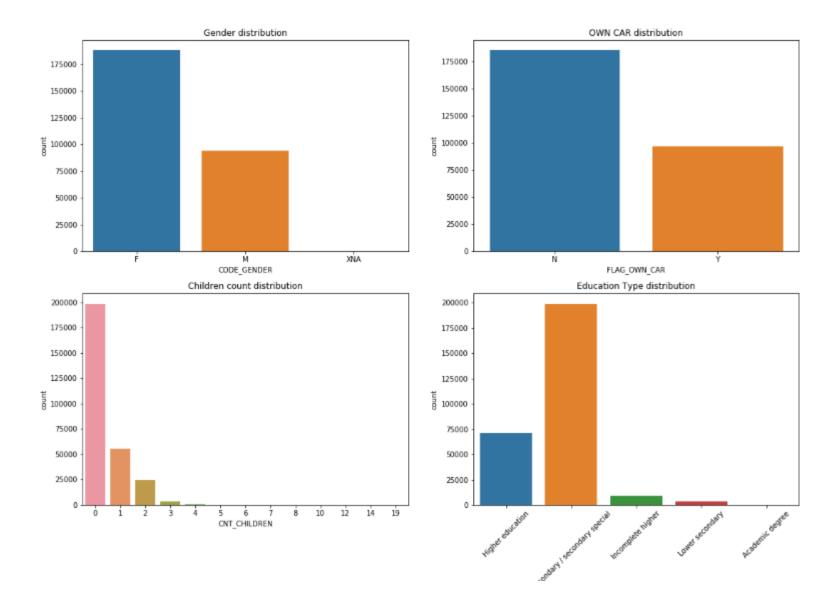Hist plot for AMT INCOME TOTAL and AMT CREDIT
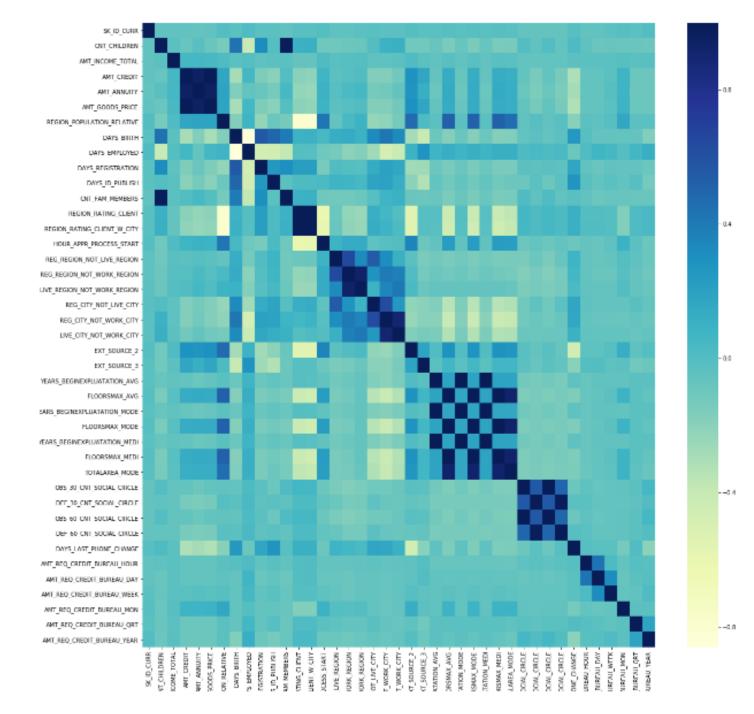
Histogram AMT_INCOME_TOTAL
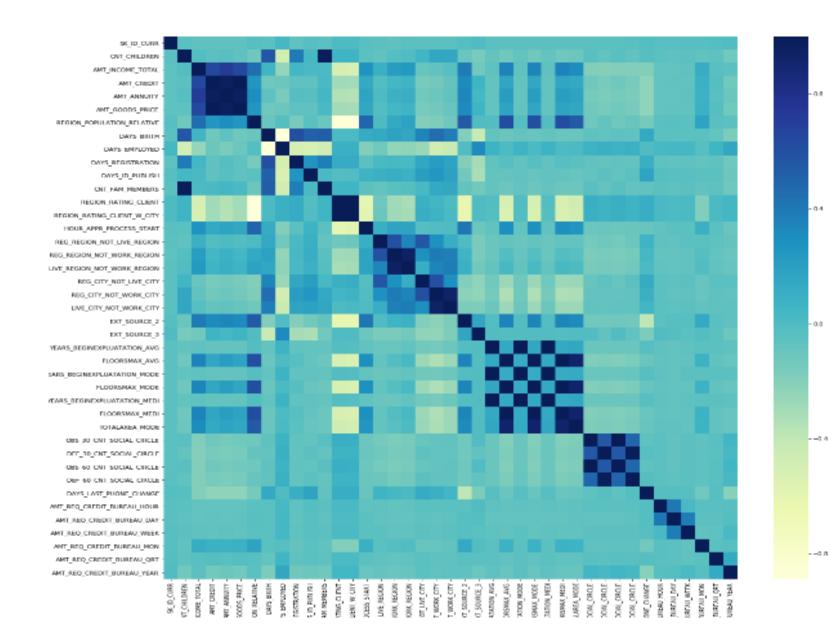
Histogram for Total Income Vs. Frequency

Distribution Graph for Target 1

Correlation Heat map for Target 1

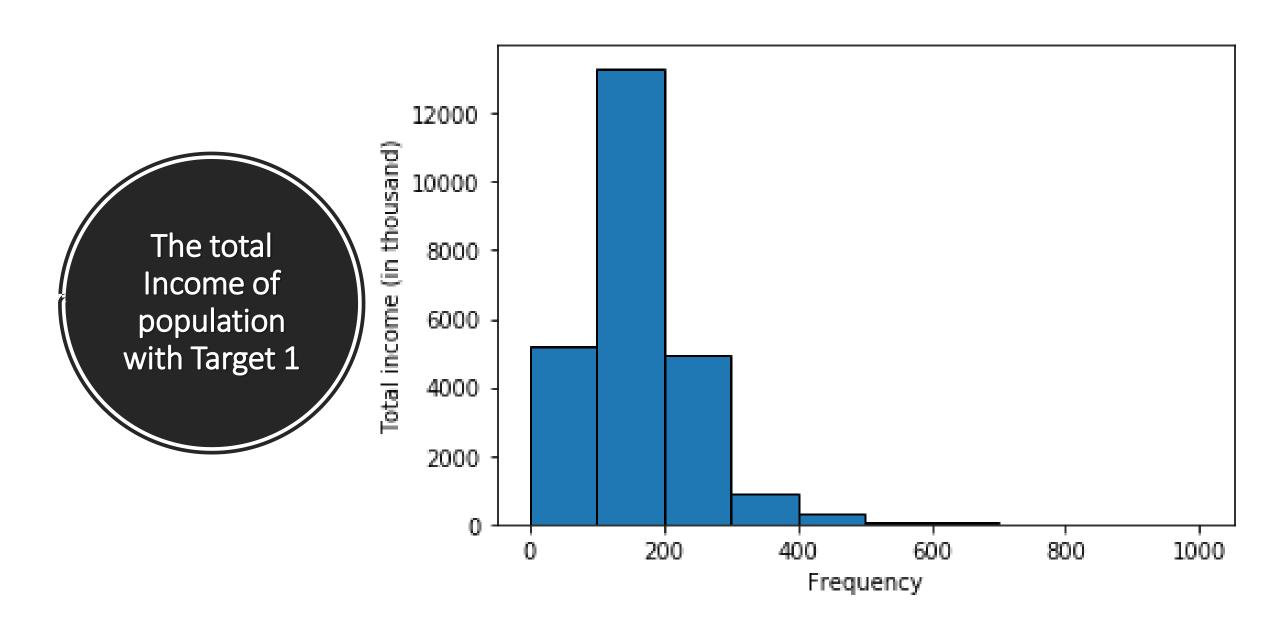Correlation Heat map for Target 0

TOP 10 correlation for the TARGET 0

	Var1	Var2	Corr
41	CNT_CHILDREN	SK_ID_CURR	0.000716
82	AMT_INCOME_TOTAL	SK_ID_CURR	0.001739
83	AMT_INCOME_TOTAL	CNT_CHILDREN	0.027397
123	AMT_CREDIT	SK_ID_CURR	0.000342
124	AMT_CREDIT	CNT_CHILDREN	0.003081
125	AMT_CREDIT	AMT_INCOME_TOTAL	0.342799
164	AMT_ANNUITY	SK_ID_CURR	0.000068
165	AMT_ANNUITY	CNT_CHILDREN	0.020905
166	AMT_ANNUITY	AMT_INCOME_TOTAL	0.418953
167	AMT_ANNUITY	AMT_CREDIT	0.771309

Top 10
Correlation
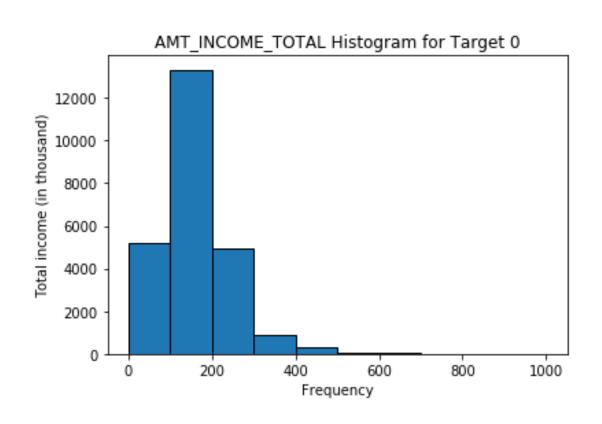
TOP 10 correlation for the TARGET 1
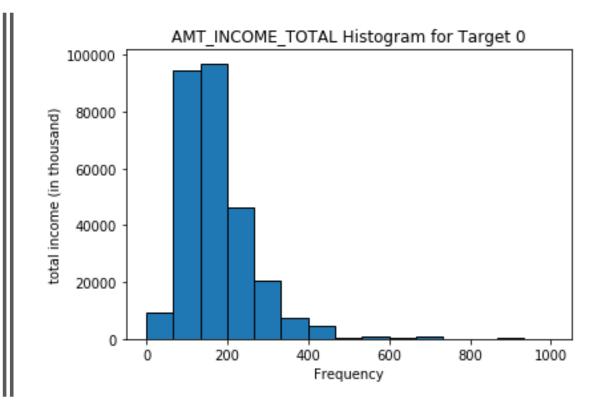
	Var1	Var2	Corr
41	CNT_CHILDREN	SK_ID_CURR	0.005144
82	AMT_INCOME_TOTAL	SK_ID_CURR	0.010165
83	AMT_INCOME_TOTAL	CNT_CHILDREN	0.004796
123	AMT_CREDIT	SK_ID_CURR	0.001290
124	AMT_CREDIT	CNT_CHILDREN	0.001675
125	AMT_CREDIT	AMT_INCOME_TOTAL	0.038131
164	AMT_ANNUITY	SK_ID_CURR	0.007578
165	AMT_ANNUITY	CNT_CHILDREN	0.031257
166	AMT_ANNUITY	AMT_INCOME_TOTAL	0.046421
167	AMT_ANNUITY	AMT_CREDIT	0.752195

Top 10 Correlation

The total Income of population with Target 1
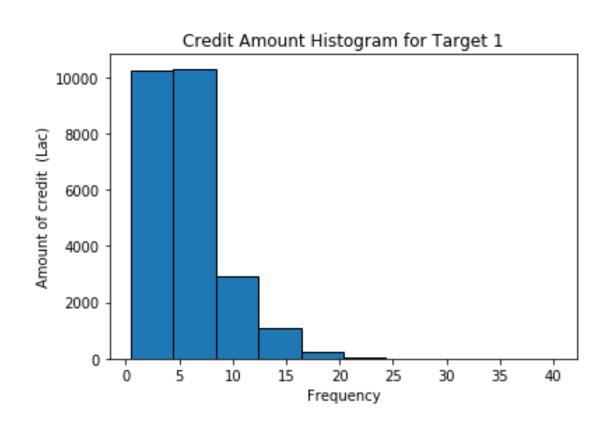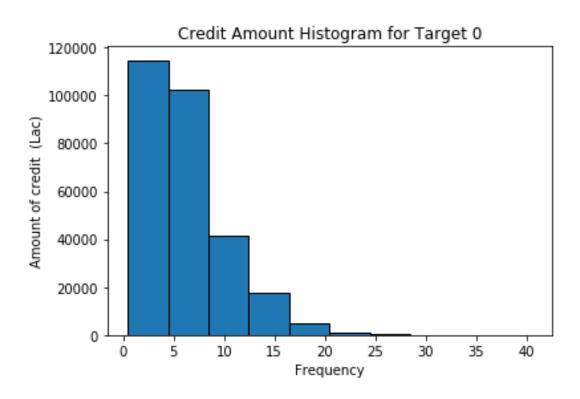
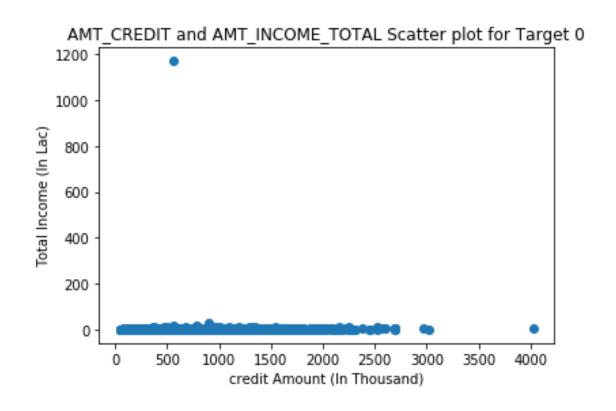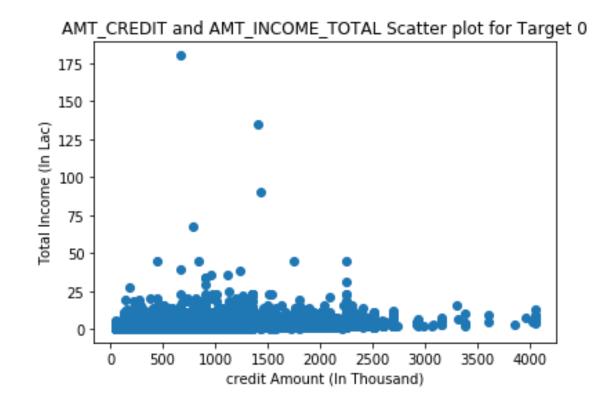# AMT INCOME TOTAL Histogram for Target 0 and 1

Credit Amount Histogram for Target 0 and 1

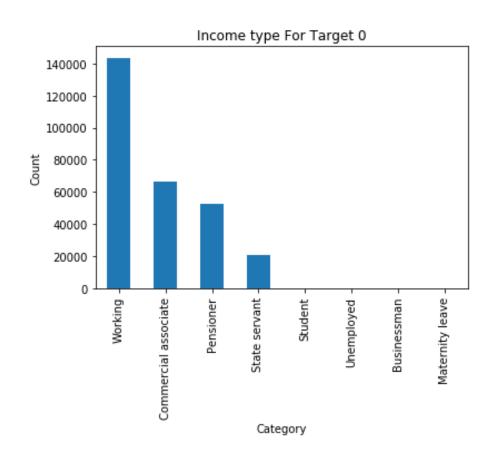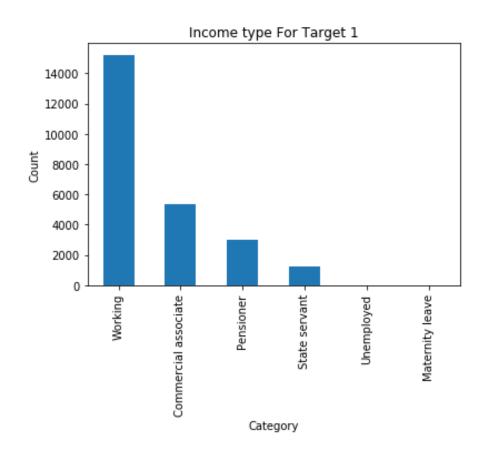AMT_CREDIT and AMT_INCOME_TOTAL Scatter plot for Target 0 and 1

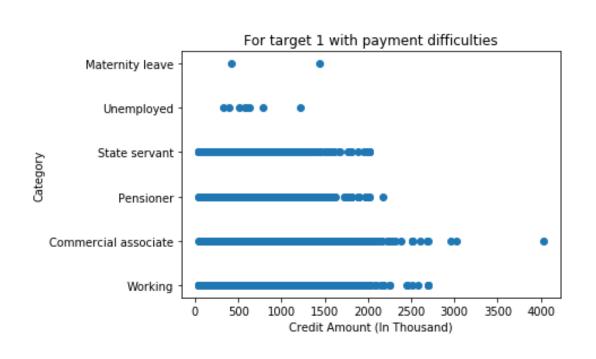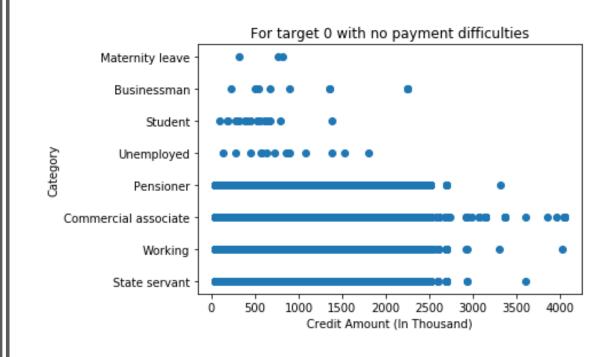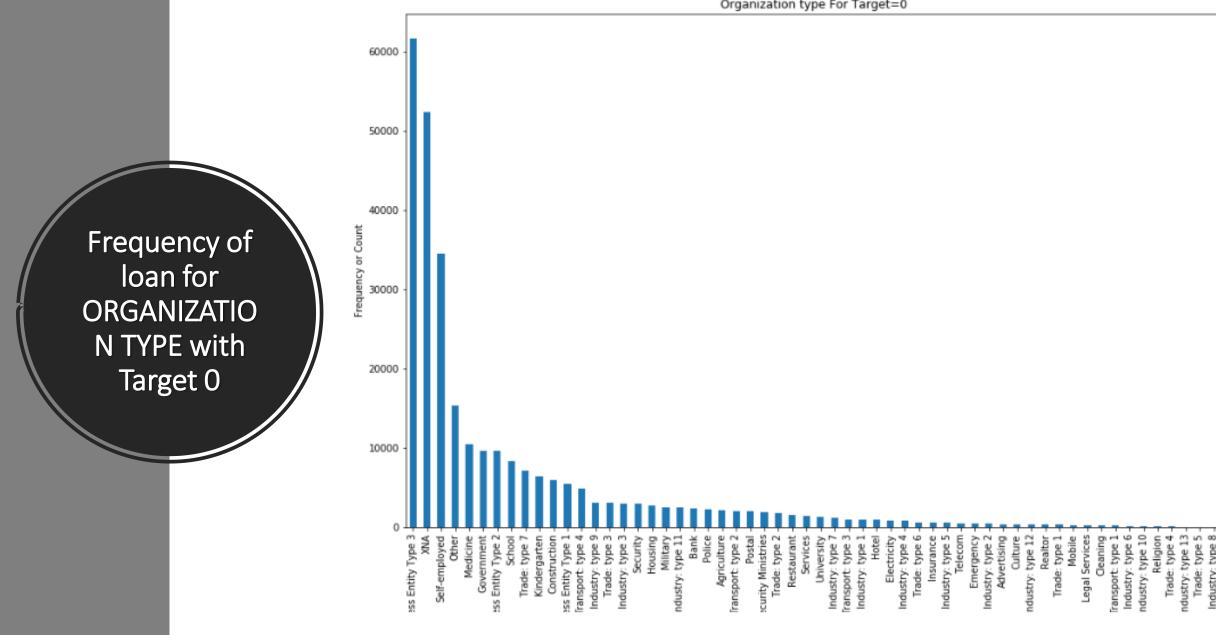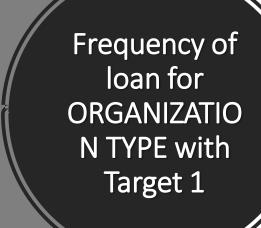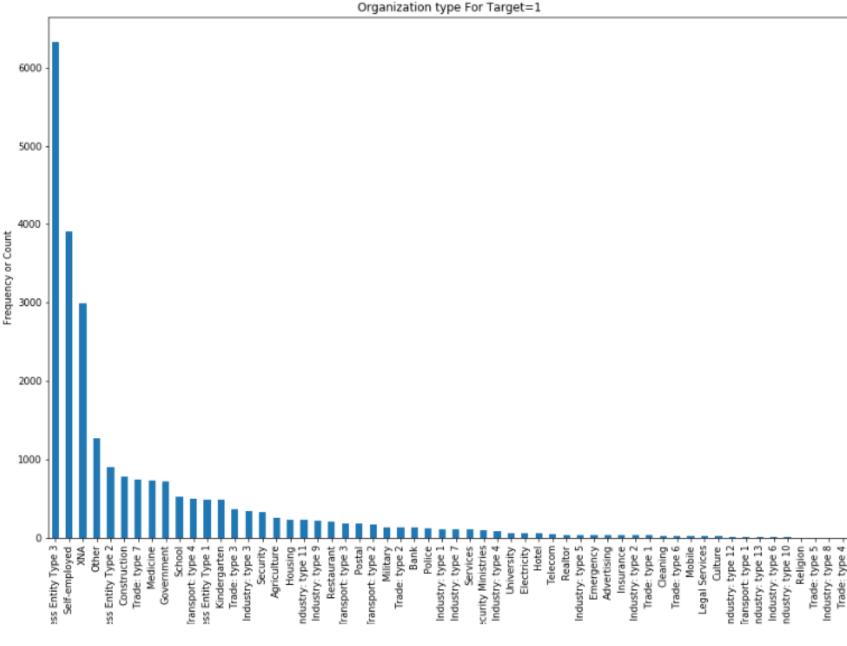Graph for Target 1 having difficulty and for Target 0 no difficulty

Frequency of loan for ORGANIZATION TYPE with Target 0

Frequency of loan for ORGANIZATION TYPE with Target 1

Organization type For Target=1

For Target 1 AMT_CREDIT and ORGANIZATION_TYPE Relation
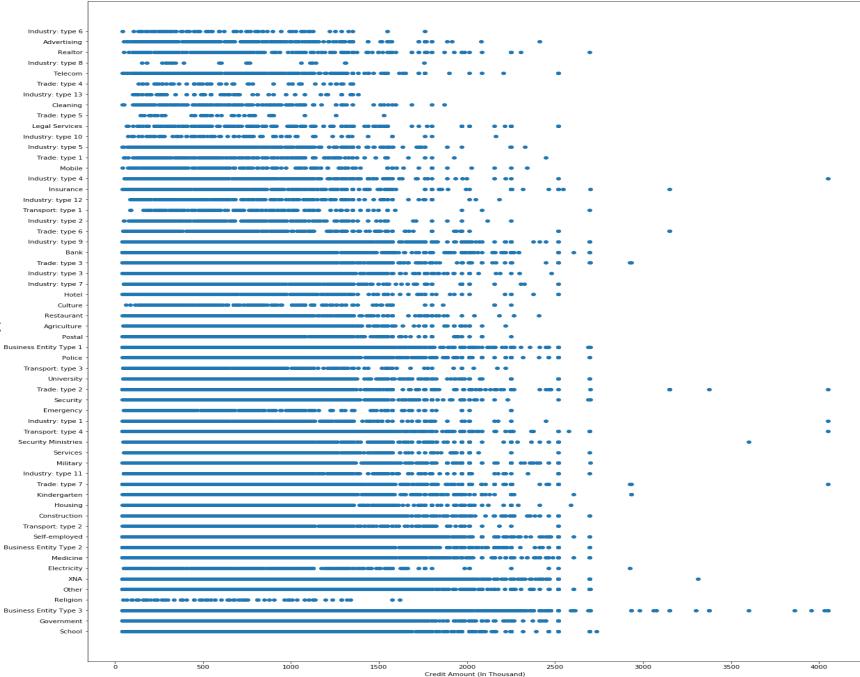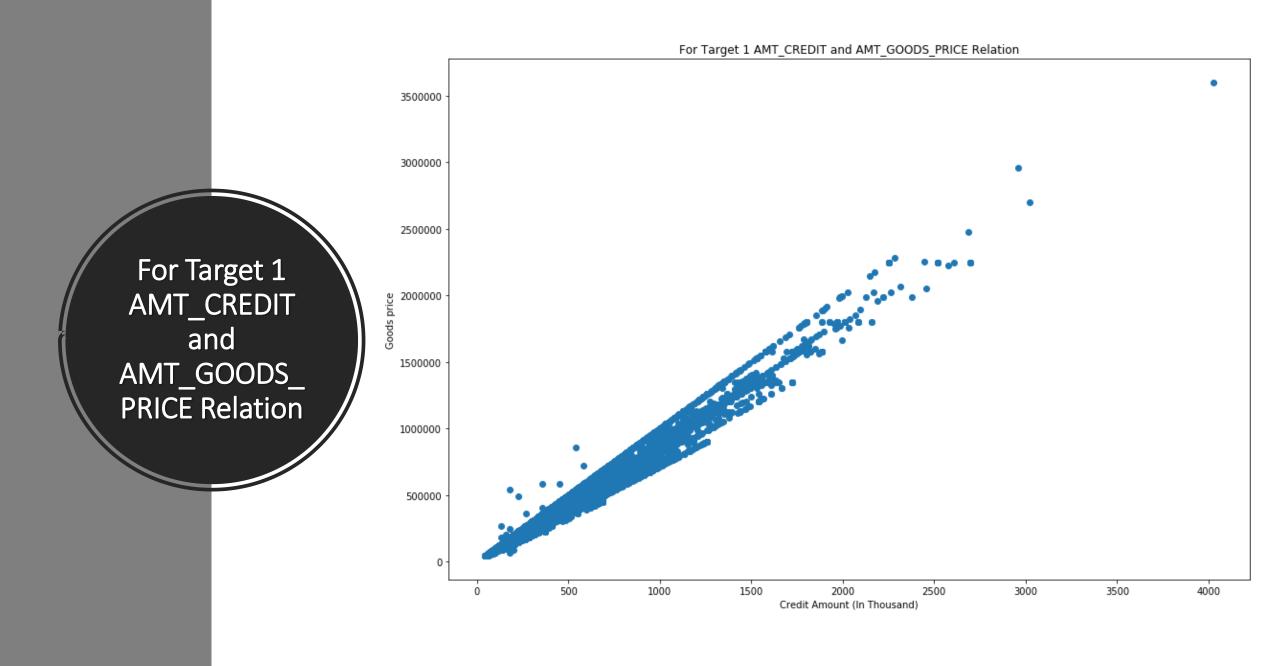
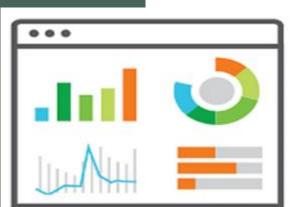For Target 0 AMT_CREDIT and ORGANIZATION_TYPE Relation

For Target 1 AMT_CREDIT and AMT_GOODS_PRICE Relation

Insights

1. For Target 1 customer having Higher Education has difficulties to replaying the loan

2. Can see for Credit amount for Target 0 and 1 - Target 1 between 5 to 20 Lakh have some issue in repaying the amount

3. Can see for Target living with Parent some issue in repaying the amount

4. Maximum application has been applied by working professionals

Conclusion for this case study

Below column are playing the major role to decide for Target 1 and 0

- ✓ **AMT_CREDIT : credit amount**

- ✓ **NAME_INCOME_TYPE : Type of Income type(Working, Pensioner, Commercial associate, state Servant)**

- ✓ **NAME_EDUCATION_TYPE : Secondary, Higher Education , Incomplete Higher, etc.**

Rohit Keshari
&
Rahul Choudhary

upGrad

Thank you