Credit EDA Case study

Anvi Radhakrishna Naik
G.M.Anitha Priyadarshini

Batch: DS EPG C31
Aug 2021

Credit Risk Analysis

Credit risk analysis in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

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.

Data set provided :

Two data sets were provided :

1.Application data : contains all information of the applicants at the time of application.

2.Previous application data : contains informations about the applicant's previous loan details.

Data Understanding

Need to perform data quality checks and handle missing values,we have considered application data set and performed necessary actions.

- Missing values where imputed on certain necessary columns
- Also missing values > 40% where been dropped from both the data set.
- Identified outliers and taken required steps on the same in analysis.

Data manipulations

- No changes on the data type is done for the given data.
- However we have augmented few columns necessary that would ease for analysis.
- Application data set after the manipulation is :

  [307511 , 73]

- Pervious application data set after the manipulation is :

  [1670214 , 26]

Problem Statement

It contains two types of scenarios:

- The applicants with payment difficulties: he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- All other cases: All other cases when the payment is paid on time.

To perform this ; we have split the data set :

- TARGET value 1 represents applicants with payment difficulties i.e. 8.07% of the data.

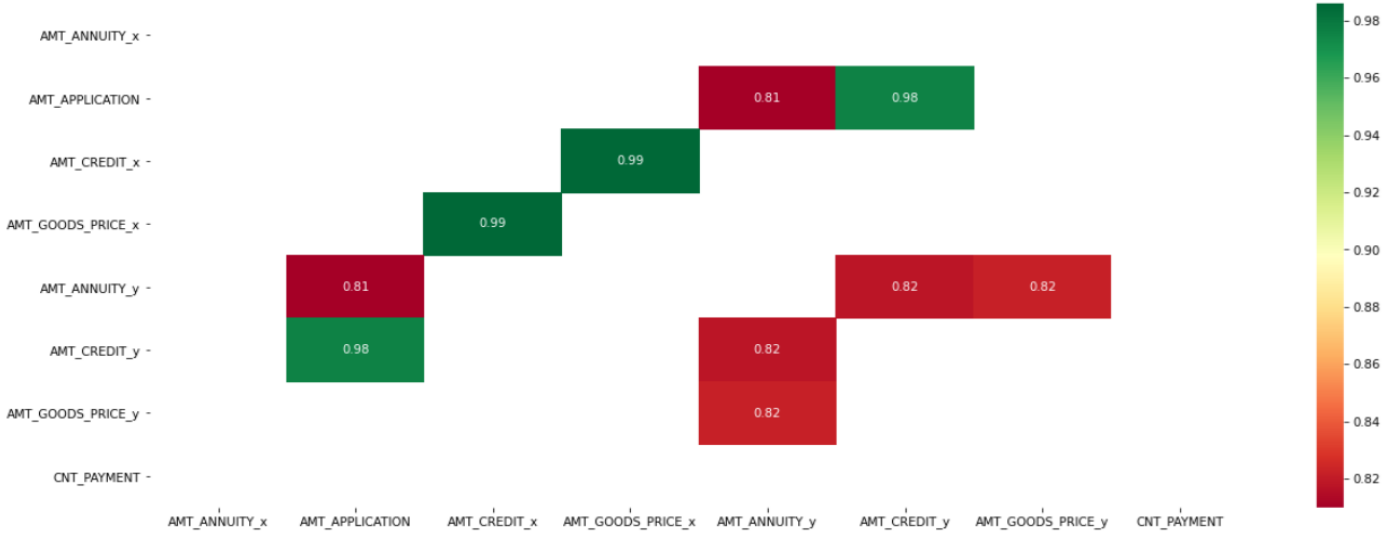
  (contains [24825, 81] data points)

- TARGET value 0 represents all other cases than 1 i.e. 91.93% of the data.

  (contains [282686, 81] data points)

Data Analysis

- Univariate Analysis on categorical and numerical variables
- Bivariate / multivariate analysis been performed
- Correlation Analysis was done and variables have been identified.



Top 6 highly correlated variables

- AMT_GOODS_PRICE_x , AMT_CREDIT_x having the correlation value has 0.99
- AMT_CREDIT_y , AMT_APPLICATION having the correlation value has 0.98
- AMT_ANNUITY_y , AMT_GOODS_PRICE_y having the correlation value has 0.82
- AMT_CREDIT_y , AMT_ANNUITY_y having the correlation value has 0.82
- AMT_APPLICATION , AMT_ANNUITY_y having the correlation value has 0.81
- AMT_GOODS_PRICE_x , AMT_ANNUITY_x having the correlation value has 0.73
- AMT_CREDIT_x , AMT_ANNUITY_x having the correlation value has 0.73

Actionable Insight :

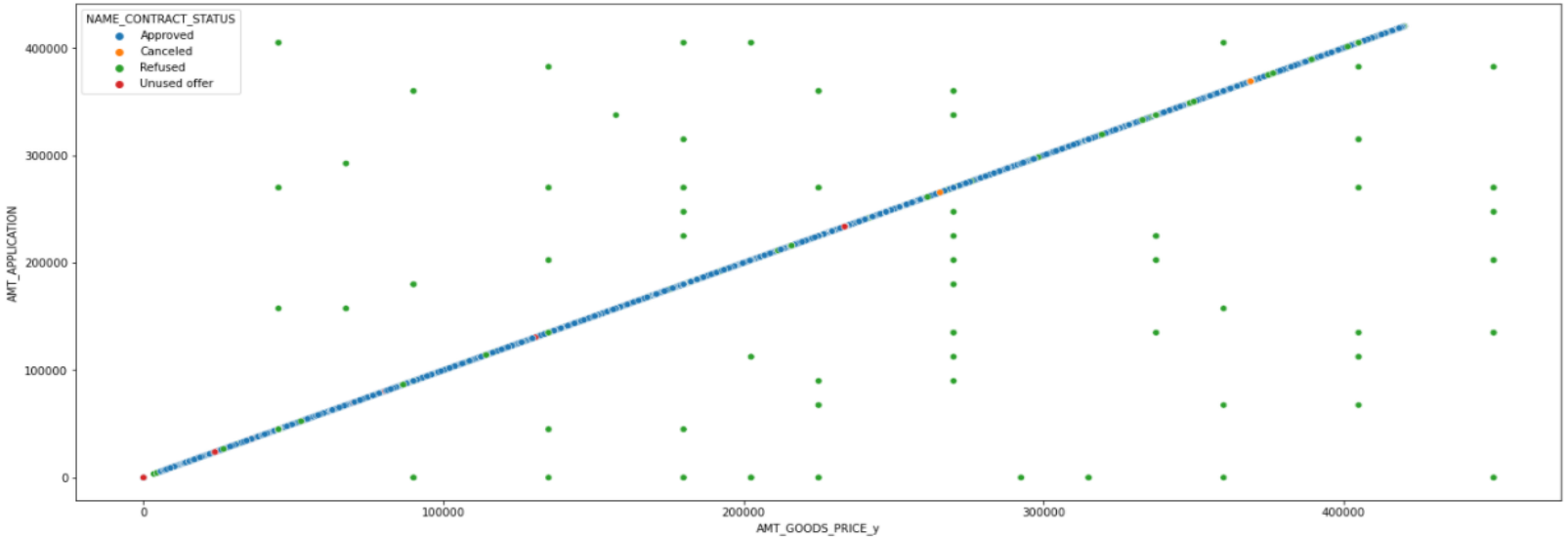
Applicant to be targeted while providing loans are:

- Applicants who are employed for more than 7 years are on time payers ; where has more approval rate of applicants who are employed less than 7 years.
- Applicants in the age range 30-40 and 40-50
- Applicants who are Married
- Applicants who are Repeaters
- Applicants who have done Secondary/secondary special education type
- Applicants who are from Business Entity Type 3 organisations
- Male applicants with Academic Degree always pay the loan on-time
- Students and Businessman seems to be an attractive category to give loans to , since they pay on time

Annexure :

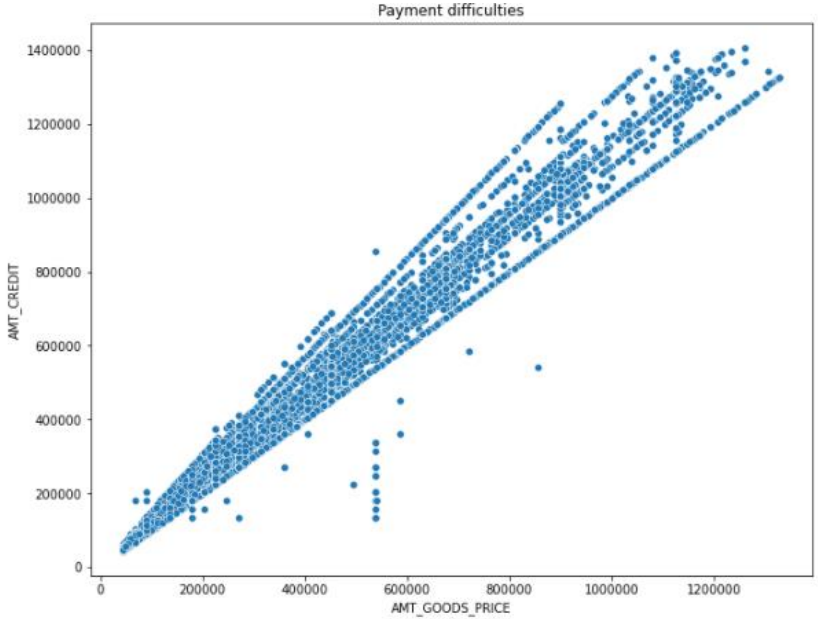
Analysis of AMT_APPLICATION V/S AMT_GOODS_PRICE_y V/S NAME_CONTRACT_STATUS

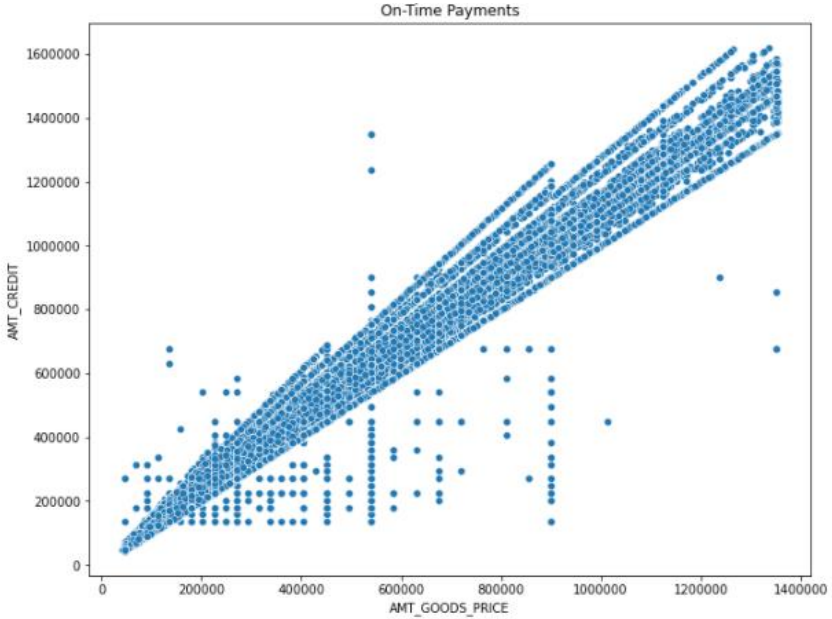
- AMT_APPLICATION has strong positive correlation with AMT_GOOD_PRICE_Y



Annexure : Analysis of AMT_GOODS_PRICE V/S AMT_CREDIT

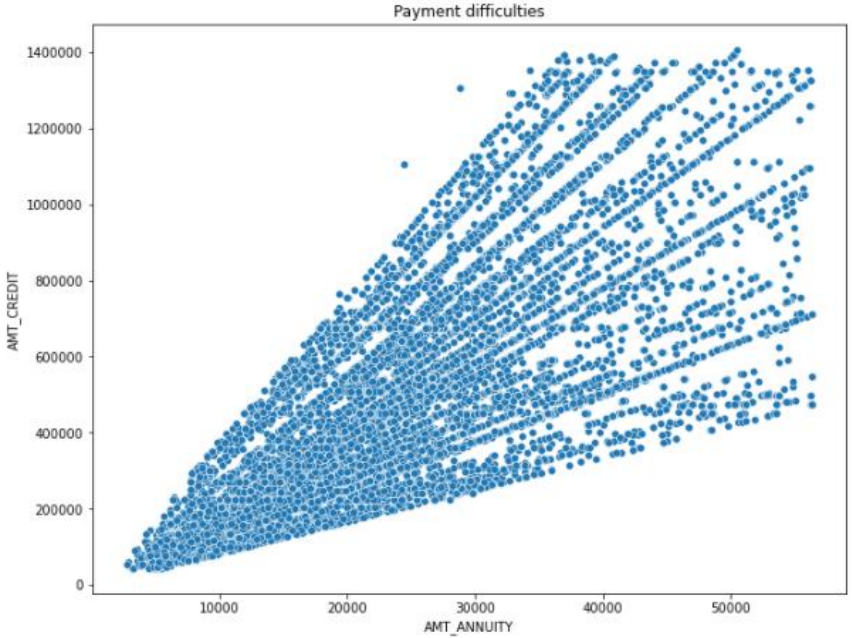
- AMT_GOODS_PRICE and AMT_CREDIT have strong positive correlation.
- This means that as Goods price increases, so does Credit Amount

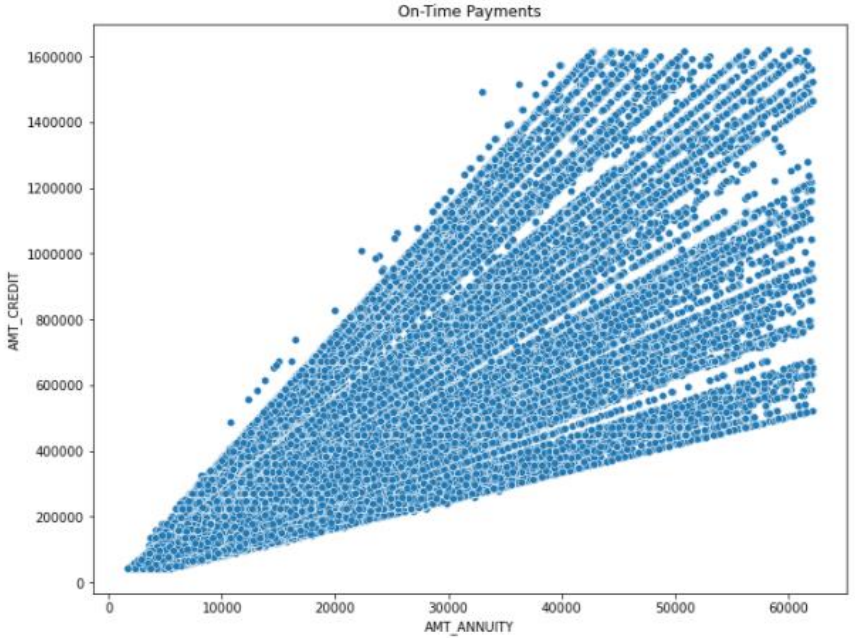




Annexure : Analysis of AMT_ANNUITY V/S AMT_CREDIT

- AMT_ANNUITY and AMT_CREDIT have strong positive correlation.
- This means that as Annuity Amount increases, so does Credit Amount





Annexure : Analysis of AMT_ANNUITY V/S AMT_GOODS_PRICE

- AMT_ANNUITY and AMT_GOODS_PRICE have strong positive correlation.
- This means that as Annuity increases, so does Goods Price