HOME CREDIT SCORECARD MODEL

ACHMAD TULUS HERMAWAN AJI

CONTENT REVIEW

1. Problem Research
2. Data Preprocessing
3. Business Insights
4. Machine Learning Model
5. Business Recommendation

1. PROBLEM RESEARCH

PROJECT BACKGROUND

Many people struggle to get loans due to insufficient or non-existent credit histories. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. In order to make sure this underserved population has a positive loan experience. Home Credit makes use of a variety of alternative data to predict their clients' repayment abilities. Doing so will ensure that clients capable of repayment are not rejected and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.

DATA SOURCE

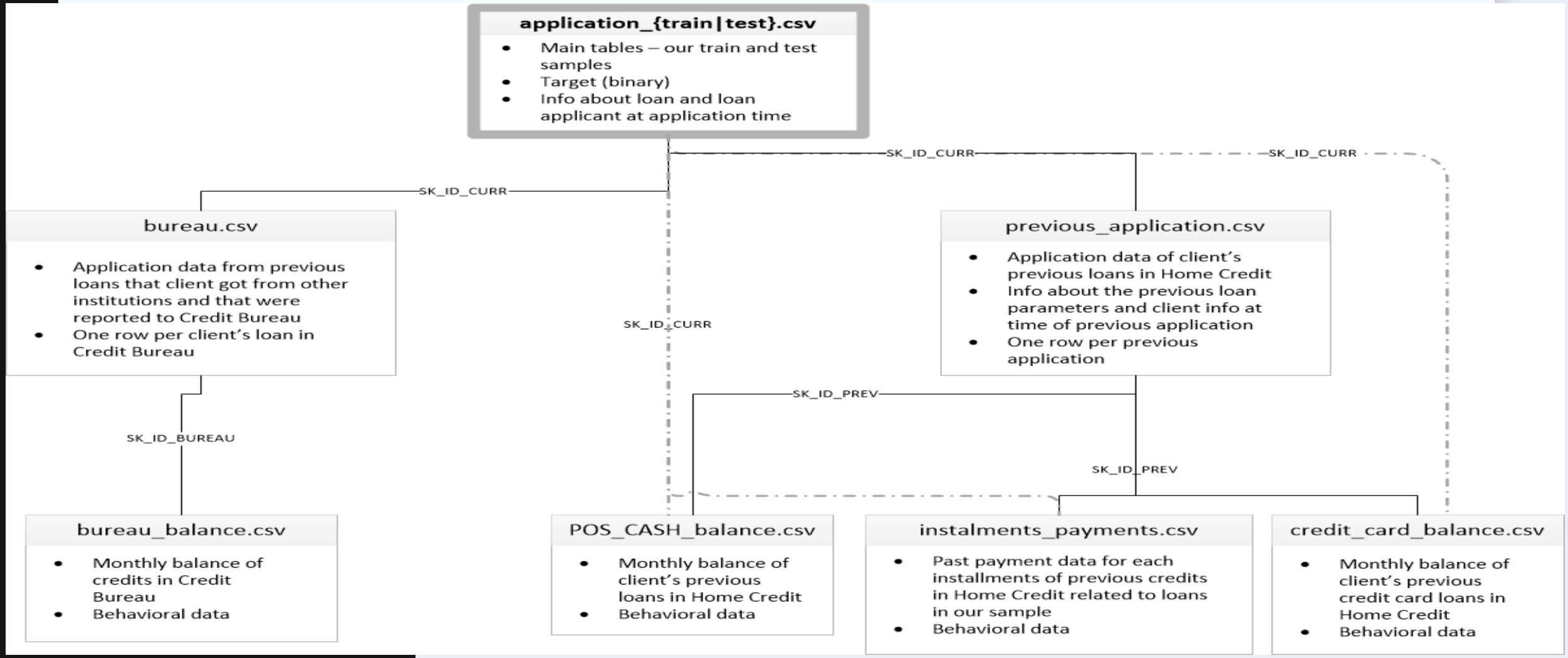
The data used are application train and application test. There are our main table, broken into two files for train (with TARGET) and test (without TARGET)

OBJECTIVE

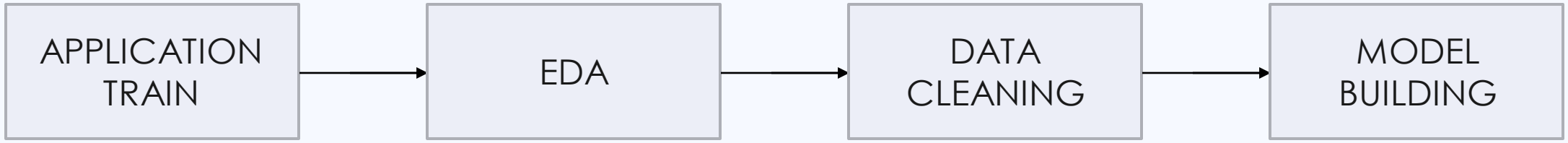
Build machine learning models to predict whether a loan applicant is capable of repaying the intended borrowed amount.

2. DATA PREPOCESSING

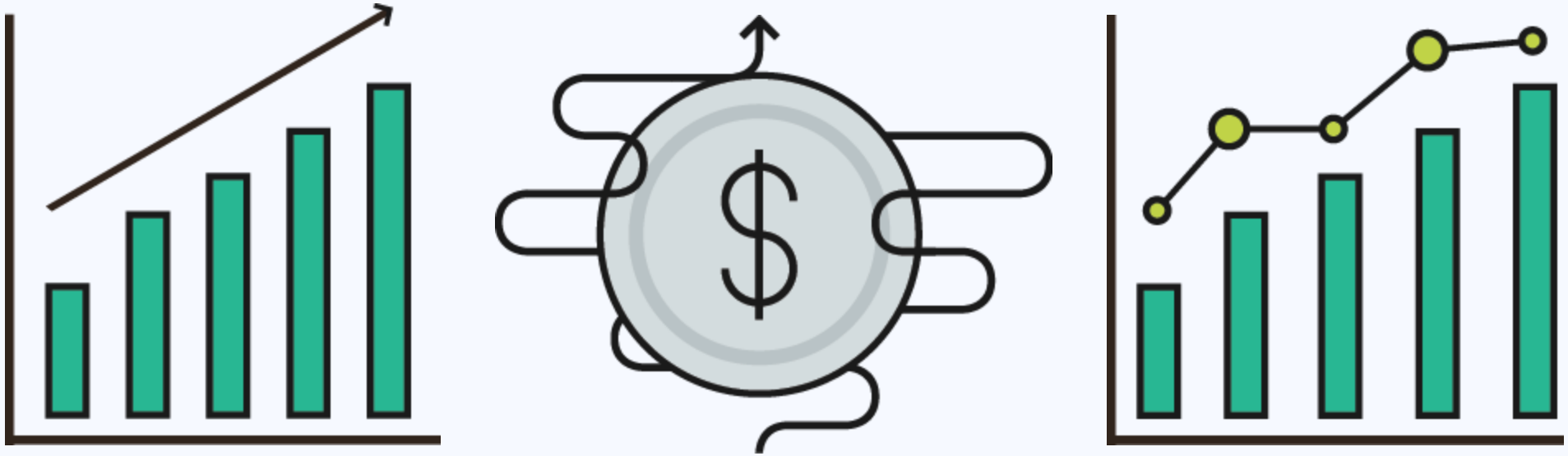
DATASET

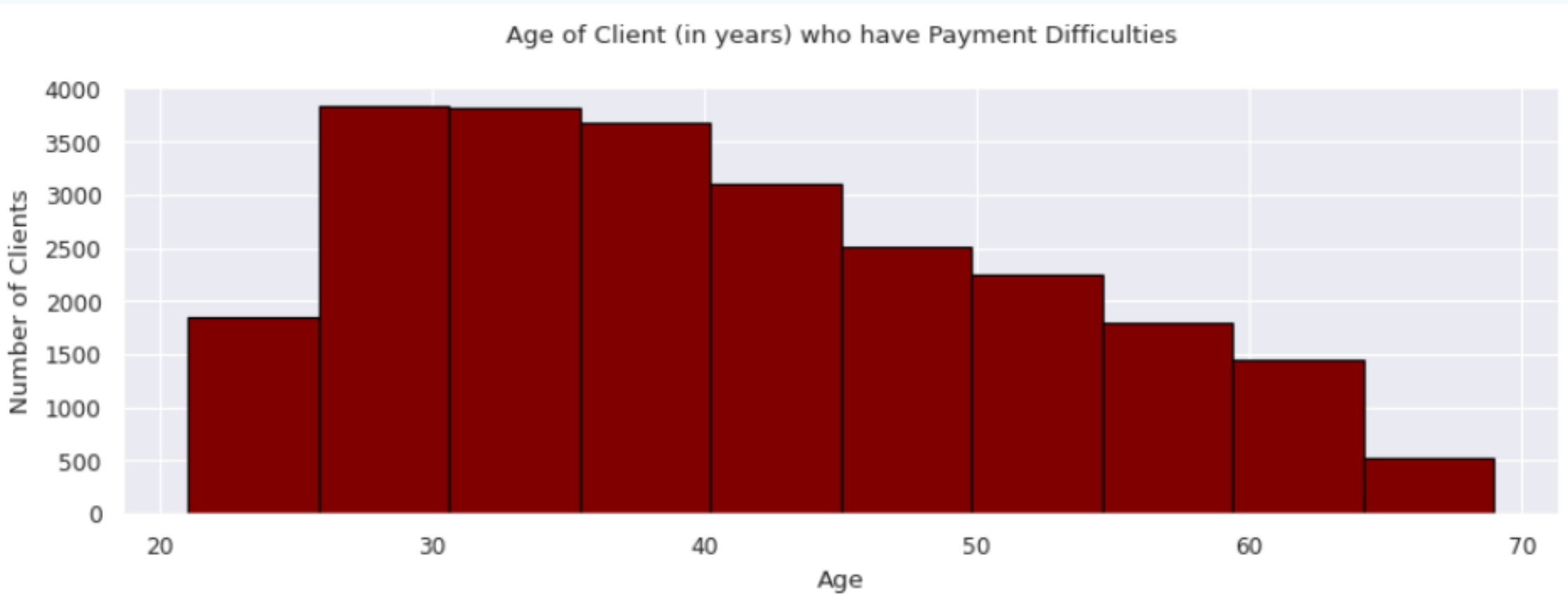

application_{train|test}.csv
Main tables – our train and test samples
Target (binary)
Info about loan and loan applicant at application time
SK_ID_CURR
SK_ID_CURR
SK_ID_CURR
bureau.csv
Application data from previous loans that client got from other institutions and that were reported to Credit Bureau
One row per client's loan in Credit Bureau
previous_application.csv
Application data of client's previous loans in Home Credit
Info about the previous loan parameters and client info at time of previous application
One row per previous application
SK_ID_BUREAU
SK_ID_PREV
SK_ID_PREV
bureau_balance.csv
Monthly balance of credits in Credit Bureau
Behavioral data
POS_CASH_balance.csv
Monthly balance of client's previous loans in Home Credit
Behavioral data
instalments_payments.csv
Past payment data for each installments of previous credits in Home Credit related to loans in our sample
Behavioral data
credit_card_balance.csv
Monthly balance of client's previous credit card loans in Home Credit
Behavioral data

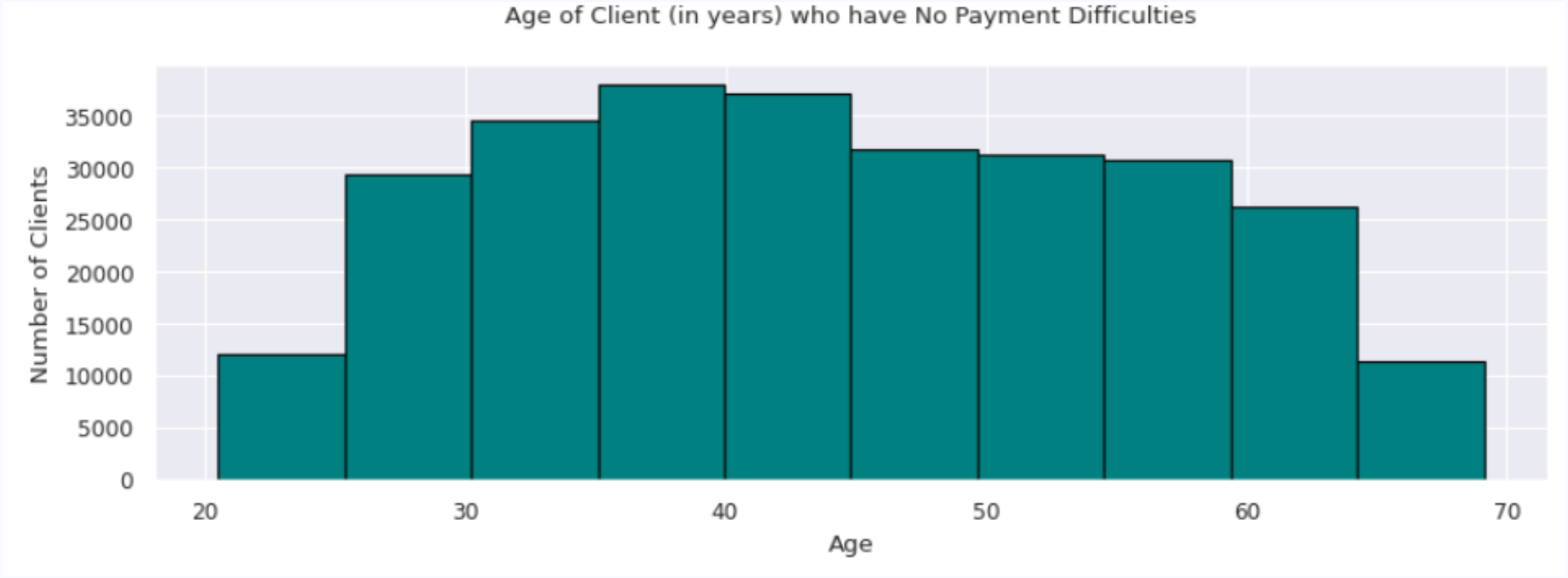
THE MAIN DATASET (APPLICATION TRAIN) CONSISTS OF 122 COLUMNS AND 307511 ROWS. WITH 1 COLUMN TARGET FEATURE ['TARGET']

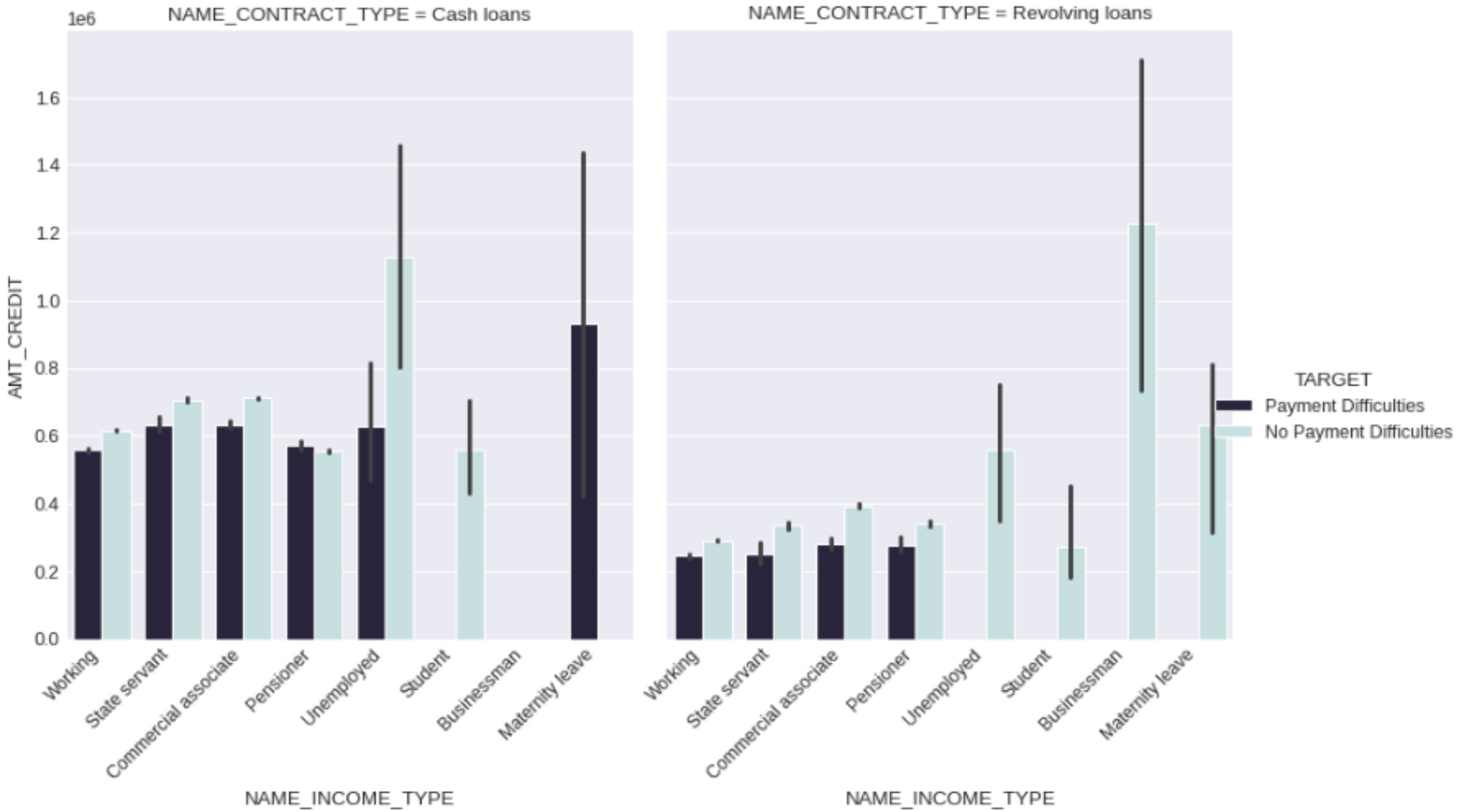


3. BUSINESS INSIGHTS




Age of Client (in years) who have Payment Difficulties
Number of Clients
4000
3500
3000
2500
2000
1500
1000
500
0
20
30
40
50
60
70
Age


Age of Client (in years) who have No Payment Difficulties
Number of Clients
35000
30000
25000
20000
15000
10000
5000
0
20
30
40
50
60
70
Age


NAME_CONTRACT_TYPE = Cash loans
NAME_CONTRACT_TYPE = Revolving loans
1e6
AMT_CREDIT
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Working
State servant
Commercial associate
Pensioner
Unemployed
Student
Businessman
Maternity leave
NAME_INCOME_TYPE
NAME_INCOME_TYPE
TARGET
Payment Difficulties
No Payment Difficulties

INSIGHTS

● Most number of clients who apply for loans are in the range of 35-40 years. ● Meanwhile, the number of applicants for clients aged <25 or age >65 is very low.

● Clients who have no payment difficulties are clients in the range of 35-45 years. You can target these clients as your priority.

● While clients who have payment difficulties are client the range of 25-35 years.

● All student clients have no difficulty repaying the loans whether with cash loan or revolving loan for a low to medium credit amount of the loan.

●For the income type of maternity leave with cash loans, all the clients have problems repaying the loans for a medium credit amount of the loan. While all clients with maternity leaves and revolving loans have no difficulty repaying the loans.

●For unemployed clients with cash loans more than 50% of clients have problems repaying loans with medium credit amounts of the loan. While all unemployed clients with revolving loans have no difficulty repaying the loan.

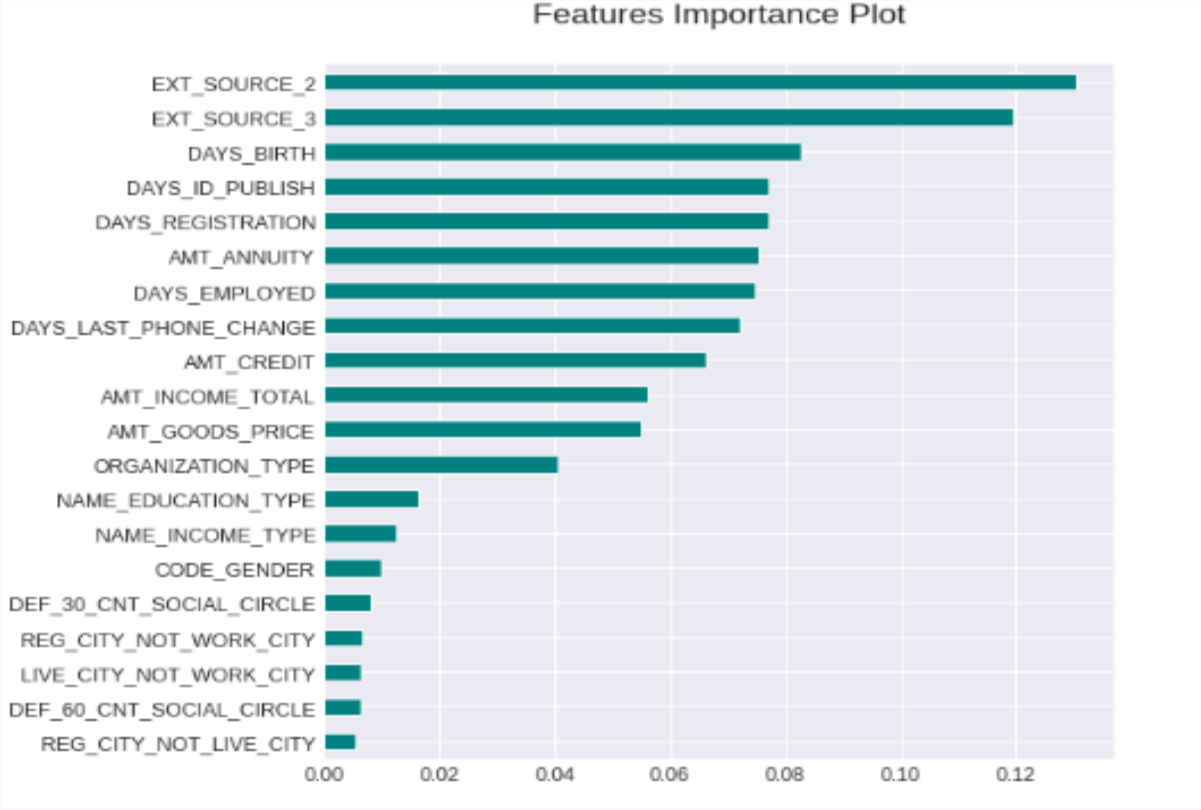
4. MACHINE LEARNING MODEL

MODEL COMPARATION

Random Forest model have the highest score andthe best model to predict client's repayment abilities.

	Models	Training Accuracy Score	Testing Accuracy Score	ROC Score
3	Random Forest	1.000000	0.947800	0.947800
2	Decision Tree	1.000000	0.781300	0.781100
4	K-Nearest Neighbor	0.915900	0.882100	0.882300
5	Neural Network	0.696600	0.688800	0.688700
0	Logistic Regression	0.672300	0.670900	0.670900
1	Gaussian Naive Bayes	0.600600	0.600700	0.604000

FEATURES IMPORTANTFEATURES IMPORTANT



The 5 most important features are EXT_SOURCE_2, EXT_SOURCE_3, DAYS_BIRTH, DAYS_ID_PUBLISH, and DAYS_REGISTRATION

5. BUSINESS RECOMMENDATION

RECOMMENDATION

1. Create a campaign so that more student, accountant, high skill tech staff, manager interested in applying for a loan.

2. Need further analysis, you can survey to find out if there is a problem if a client with maternity leaves or unemployed takes a cash loans contract. So, in the future, if there are clients with that type of income, you can recommend the right contract type so that their applications will be approved.

3.Clients who have no payment difficulties are clients in the range of 35-45 years. You can target these clients as your priority.

GIT HUB :

https://github.com/ajitulus/HOMECREDITVIX