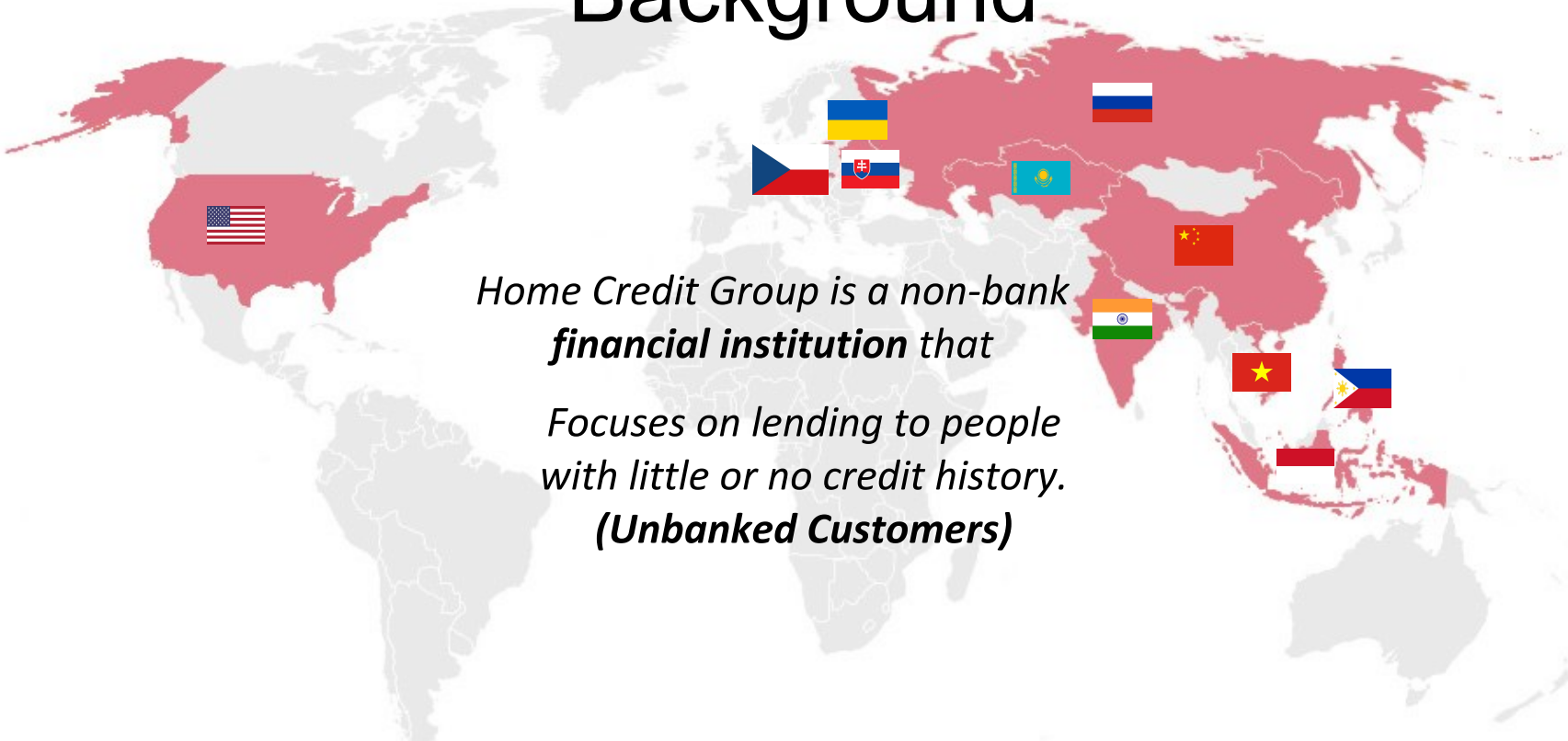
Credit Risk Prediction

Isaac Tan

Background



Home Credit Group is a non-bank ***financial institution*** *that*

Focuses on lending to people with little or no credit history. ***(Unbanked Customers)***

Unbanked Population Example

Just 31% of Vietnamese adults have bank accounts and more than 95% of payments are made with cash and gold, according to the government.

Source: Bloomberg News dated *28 May 2019, 17:36 GMT+8*

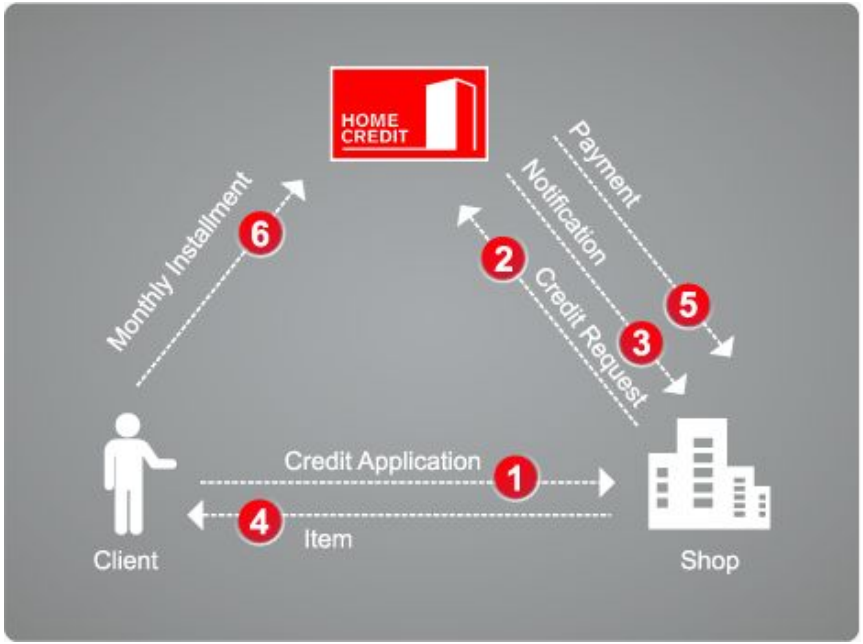
https://www.bloomberg.com/news/articles/2019-05-27/vietnam-s-next-revolution-a-cashless-society-for-the-masses

Business Model

1) Acquire new customers by targeting **unbanked population** who are purchasing durable goods at retail shops.

2) Offer them **credit facilities**:

- Revolving loans
- Cash loans



Problem Statement

- **Credit Scoring**
  - Customers mostly have low exposure to banking services.
  - Credit Bureau records are usually lacking/ non-existent.

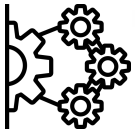
- **Predict clients' repayment abilities**
  - Can prediction be done accurately with alternate data like housing, telco and transactional information ?

Data Preprocessing

Data Cleaning

- Dropped Columns that has more than 60% null values
  (mostly housing features - year built, common area size etc.)
- Impute missing values with median as there are many outliers.

Data Integration



- Merged main dataset with credit bureau and other loans dataset.

Dataset Description

Data Integration

- Merged main dataset with credit bureau and other loans dataset.
- **340,000 unique users**
- **139 columns, 250 features after label and one-hot encoding**
  - ❖ Income, Loan Amount, Type of loan, Occupation, Housing features, Documents submitted, Credit Bureau data, etc.

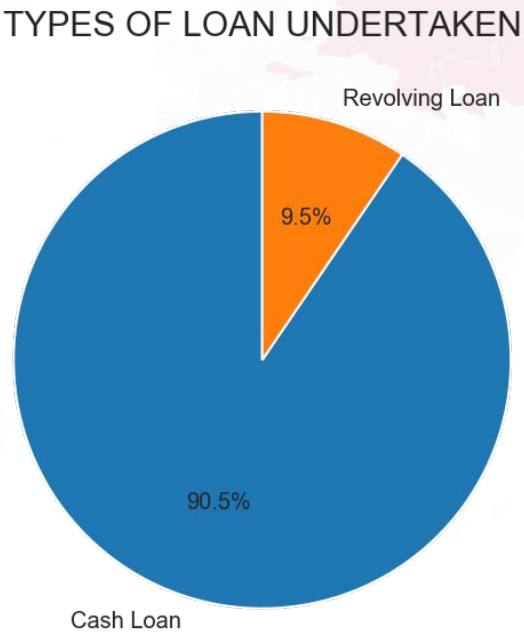
EDA - Class Imbalance

Binary Classification

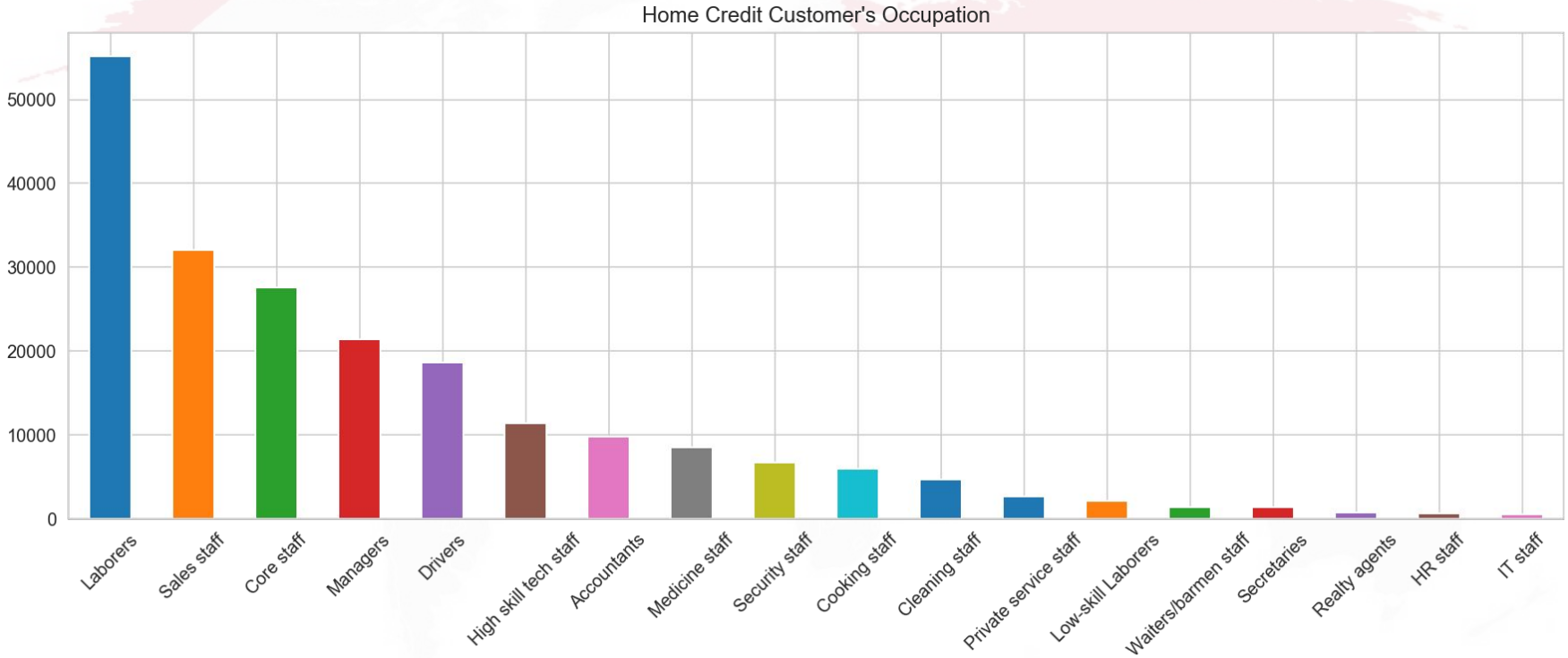
0: Successful Repayment (*282,686*)

1: Defaulted on loan (*24,825*)

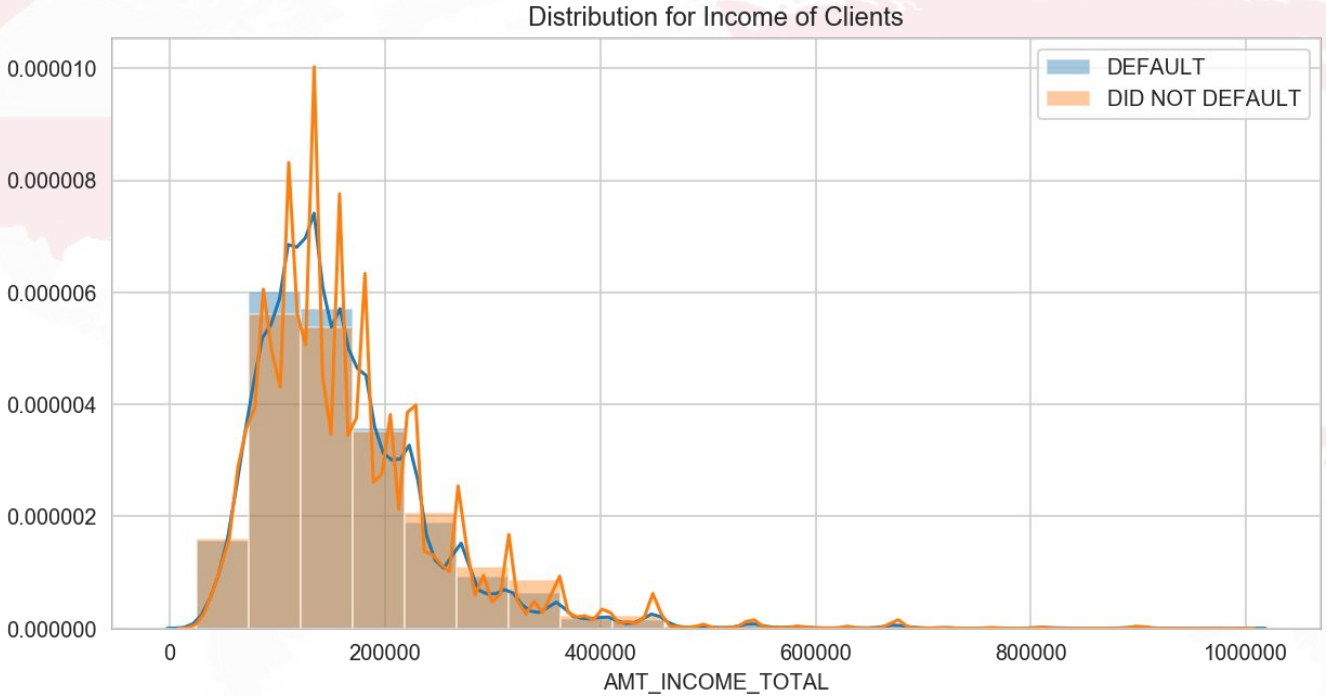
EDA - Loan Types



EDA - Occupation

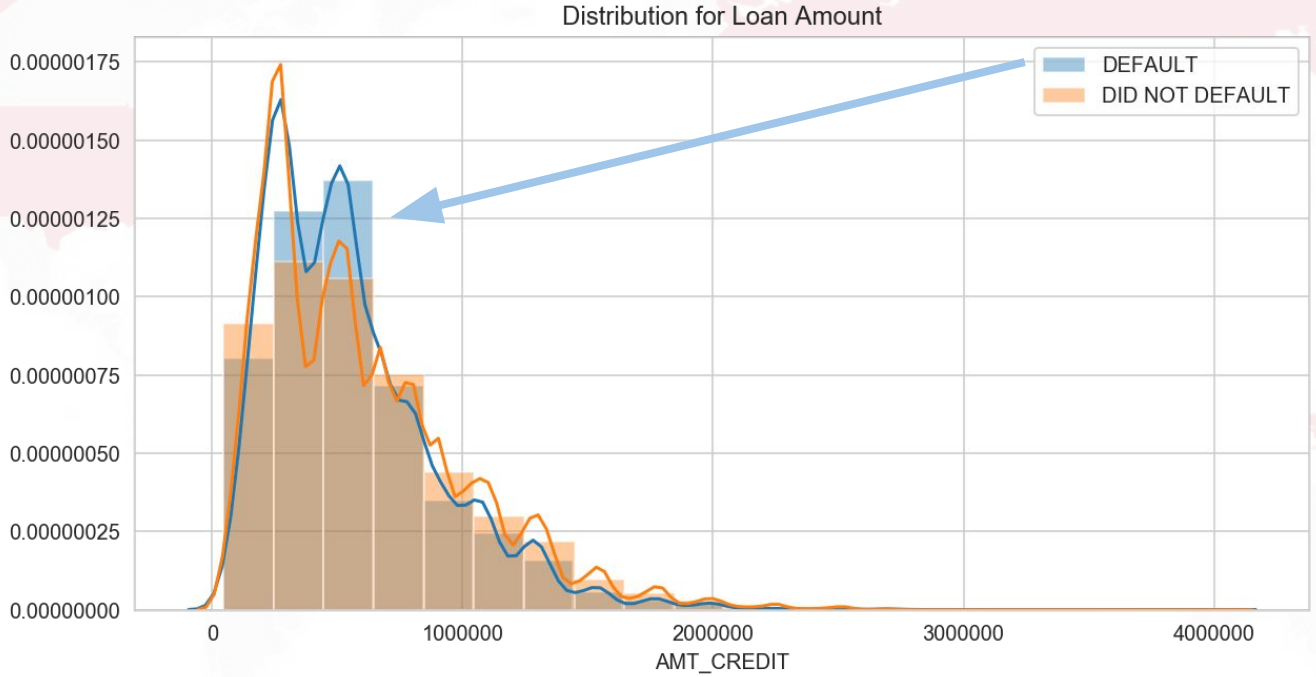


EDA - Income Distribution

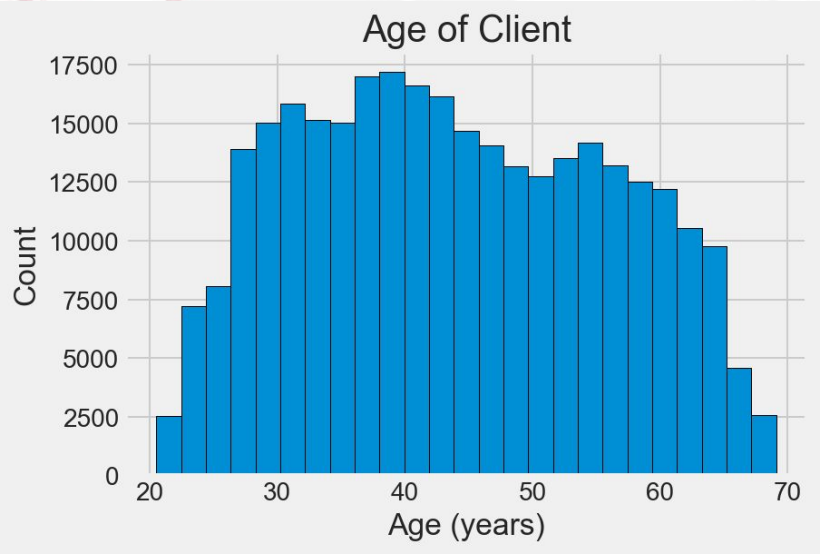


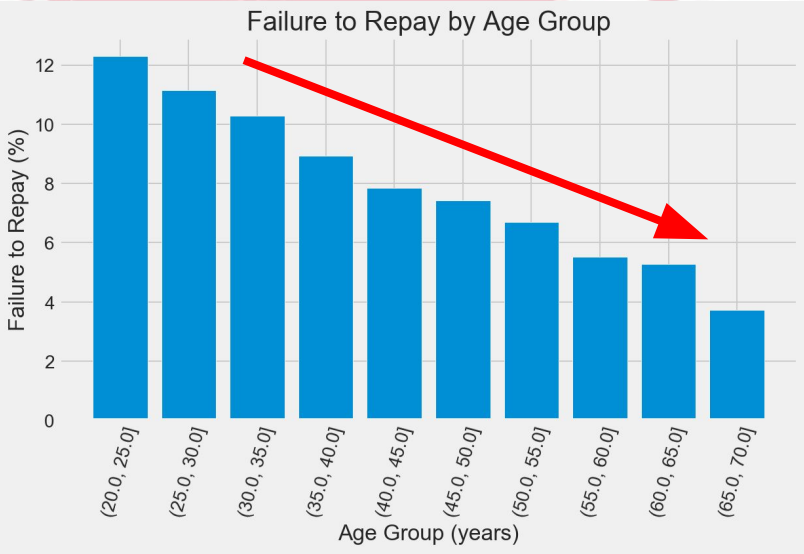
HOME CREDIT

EDA - Loan Amounts

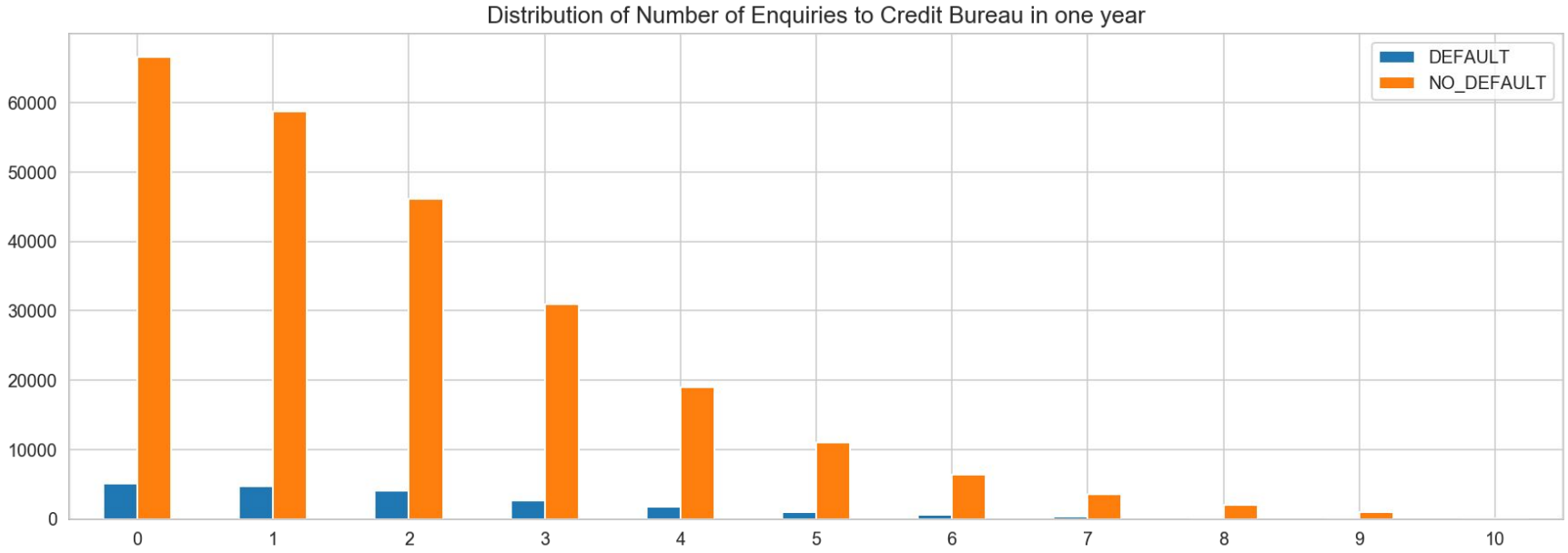


EDA - Client's Age and Failure to Repay





EDA - Credit Bureau Enquiries made in a year



Features Engineered

Feature	**Description**
credit_term	Loan Period
previous_loan_counts	Number of loans with Home Credit
other_loan_counts	Number of loans with other credit providers

Feature Selection

Attempted

RFECV	
SelectKBest	

Final Features : Top-down Manual Feature Elimination

Removed Collinear Variables	Pearson Correlation Coefficient > 0.9
Feature Importance Elimination	Removing features with lowest importance

HOME CREDIT

Project Goal:

Reduce the number of loan **rejections** for clients who were actually capable of repaying.

(Reducing Type I Error, False Positive)

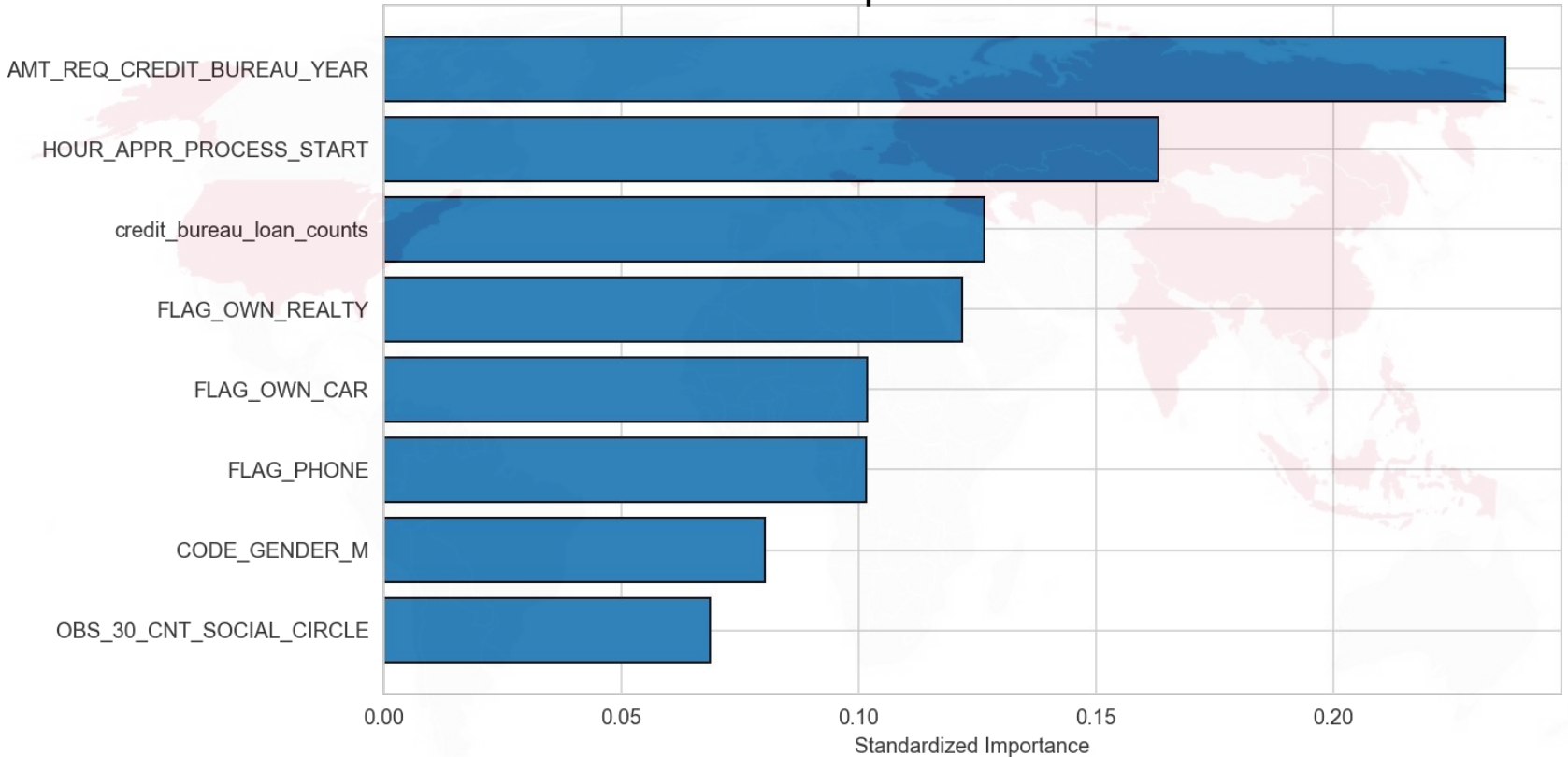
Model Selection

Models\ Metrics	AUC-ROC (Kaggle Submission)	AUC-ROC	Precision
Logistic Regression	0.74	0.74	0.16
Random Forest	0.72	0.71	0.32
dmlc XGBoost	0.71	0.71	0.33
Yandex CatBoost	0.77	0.77	0.52
Ensemble with VotingClassifier	0.75	0.76	0.46

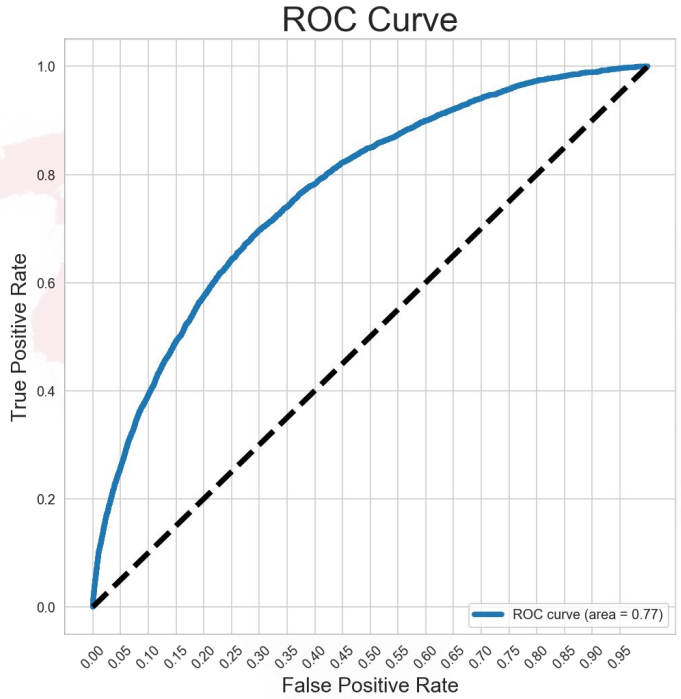
Yandex CatBoost

HOME CREDIT

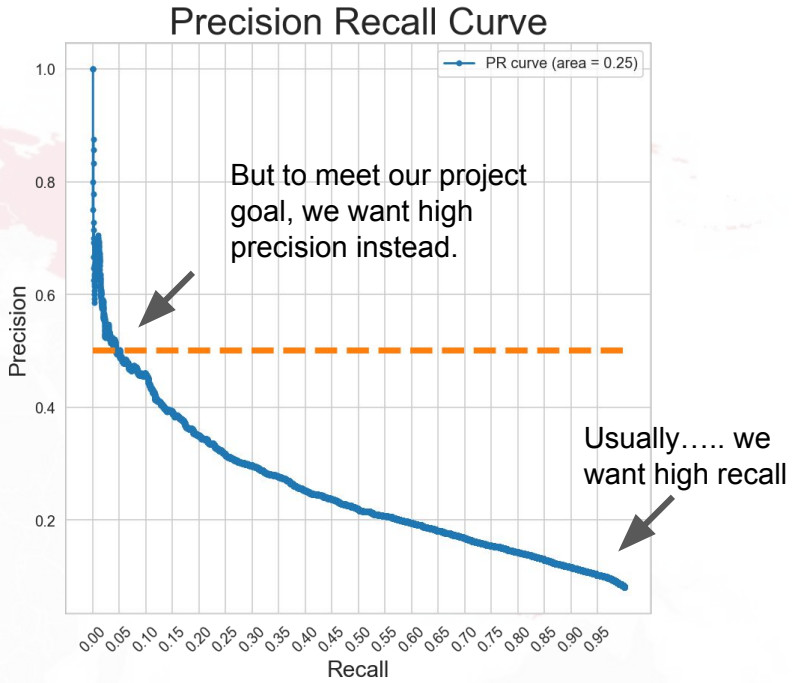
Feature Importance



HOME CREDIT



77% chance that model will be able to distinguish between positive class and negative class.
(Kaggle Competition Winner: 80%)



Trade-off: High Precision with Low Recall.

Number of loan defaults(bad debt) increases if we want to meet our goal of High Precision.

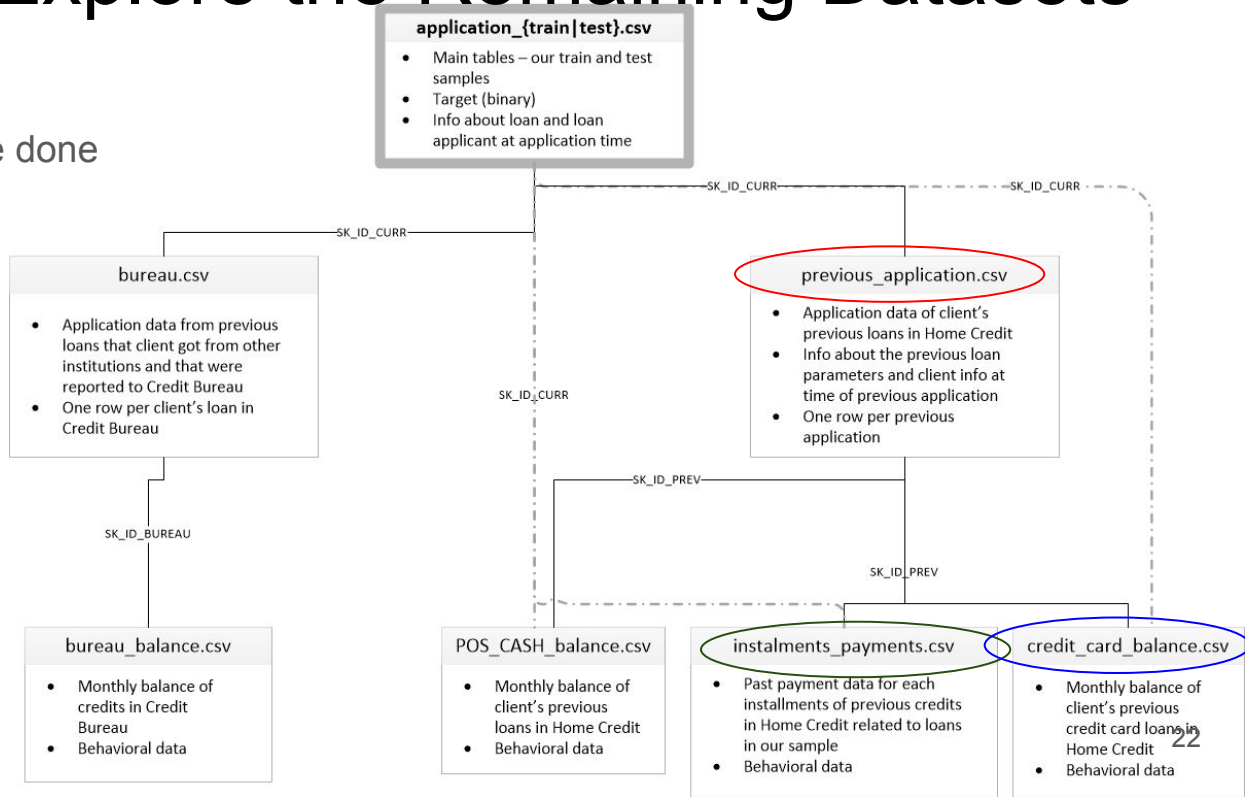
Limitations

- Difficult to predict new customers with no credit bureau records.
- Only managed to utilize 4 out of 7 datasets provided
  - Might be missing out on other important features.

Next Step: Explore the Remaining Datasets

- More features to select
- More Feature Engineering to be done

Application_train	307,511	122
Application_test	48,744	121
Bureau	1,716,428	17
Bureau_balance	27,299,925	3
Previous_application	1,670,214	37
POS_CASH_balance	10,001,358	8
Credit_card_balance	3,840,312	23
Installments_payments	13,605,401	8




Thank
You


Questions
Answers

How to improve your credit score risk grade in Singapore

- Always repay loans on time
- Avoid making multiple loan enquiries in a short time
- Don’t have too many credit facilities open
- Never default on your loans