BFSI CASE STUDY

PROBLEM STATEMENT

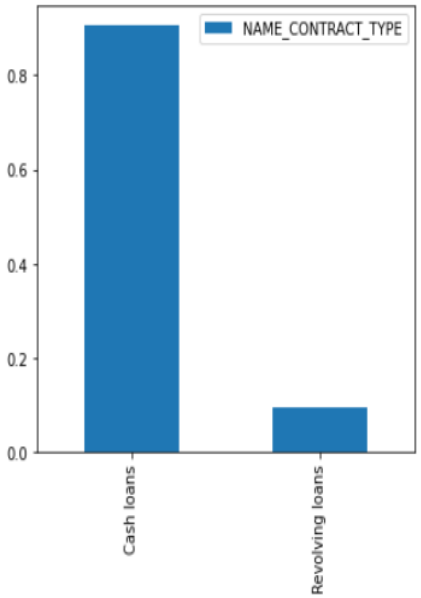
All BFSI institutions are faced with a major default problem: not every individual that takes a loan has the willingness, ability and/or integrity to pay it back. Thus, an average of 2-5% default rate is observed across banks for different loan categories like personal loan, education loan, vehicle loan, business loan etc. Given the fact that banks can never get this number to zero, it has to keep it within limits, and rather keep it at the lowest possible levels to be able to retain the money inflow.

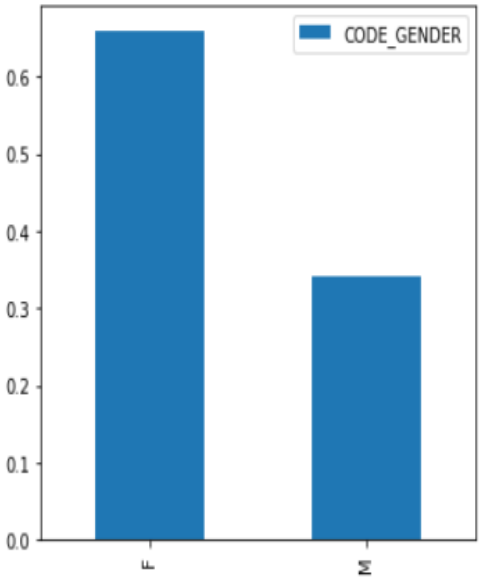
The primary objective of this study is to assist Home Credit in deciding which loan applications should be disbursed, and which should be rejected, based on the applicant's past behavior and application information.

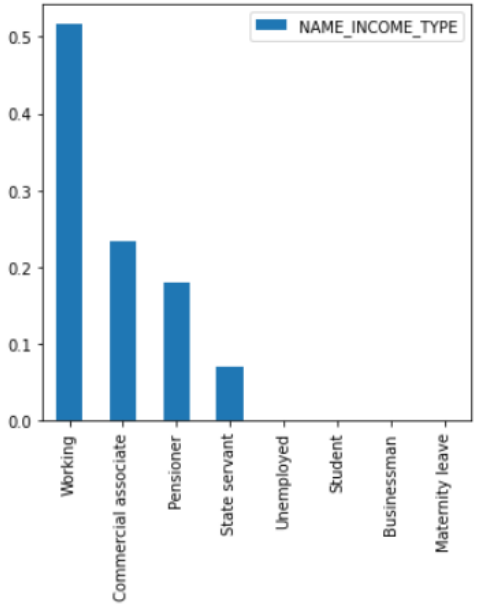
APPROACH

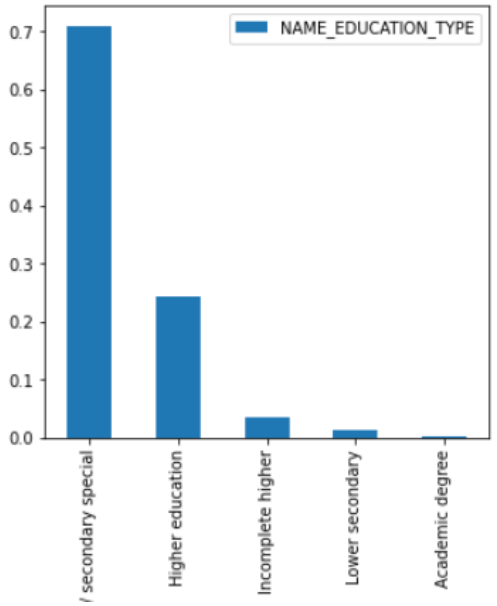
- Data Exploration
- Data Cleaning
  - Missing values treatment
- Exploratory Data Analysis
  - Checking of Outliers
  - Checking of unwanted variables
- Data Preparation
  - Standardizing the data
  - Creating Dummy variables
  - Dropping repeated variables
- Feature Engineering
  - Feature Scaling
  - Feature selection using RFE
- Model Building using Logistic Regression
- Model Evaluation using CV

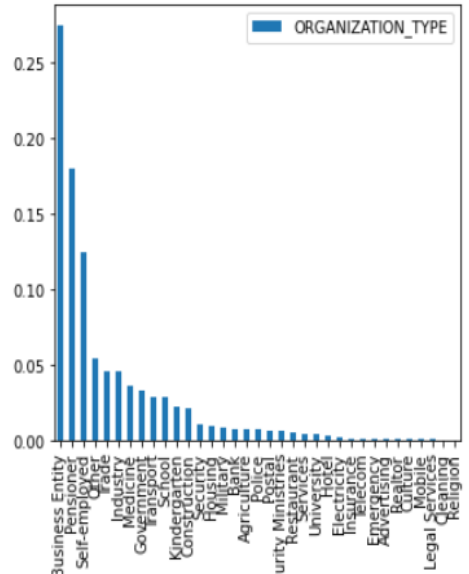
EXPLORATORY DATA ANALYSIS
UNIVARIATE ANALYSIS







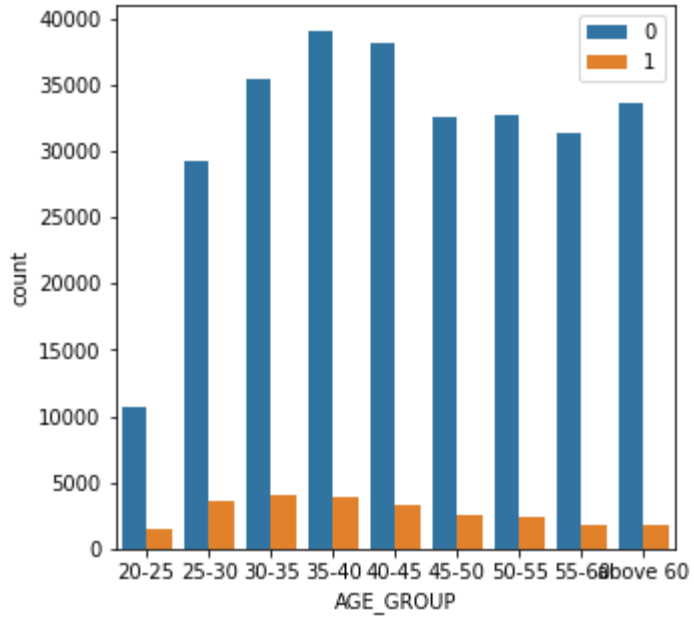


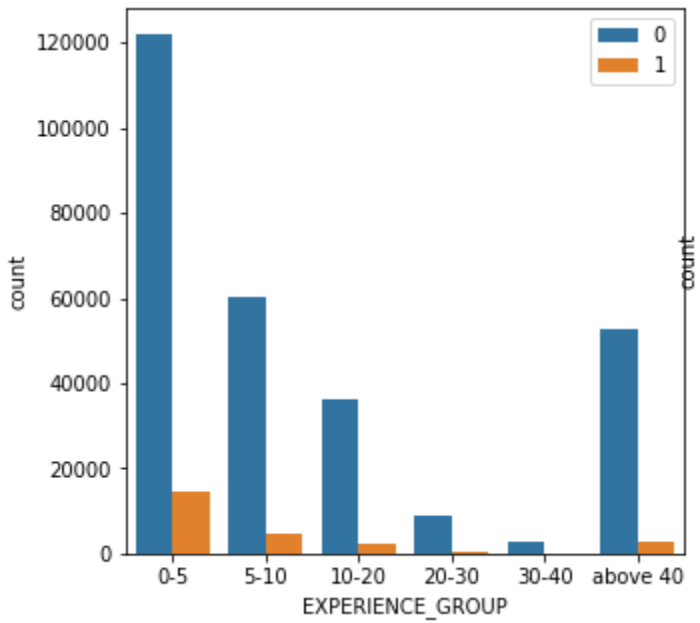


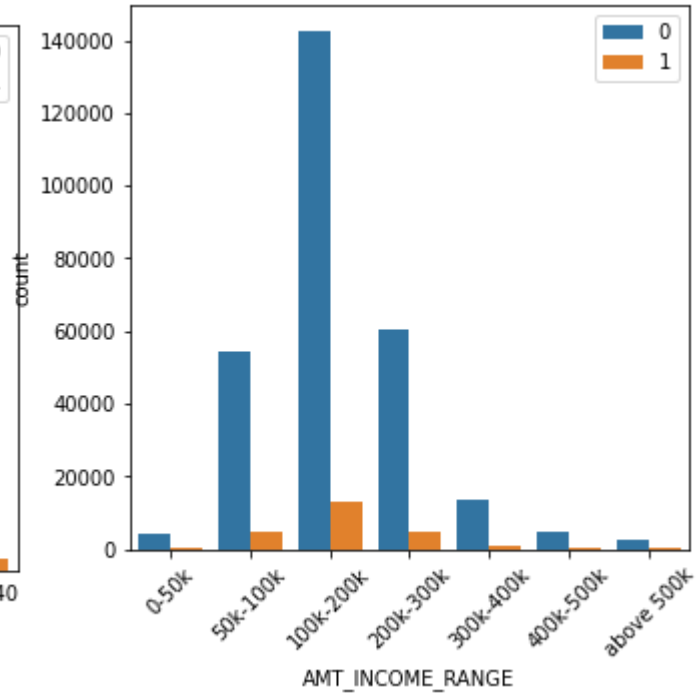
Insights:

- 90% of applicant applied for cash loans,10% of applicant applied for Revolving loans
- 65% of applicants are Female, 35% are male
- 66% of applicants dont owns car, 69% of applicants owns realty
- Most of the applicants (81%) are unaccompanied during application of loan
- Most of the applicants (52%) are working class , 71% of applicants are completed Secondary education
- 64% of applicants are married
- Majority of applicants are from Business background

EXPLORATORY DATA ANALYSIS
BIVARIATE ANALYSIS



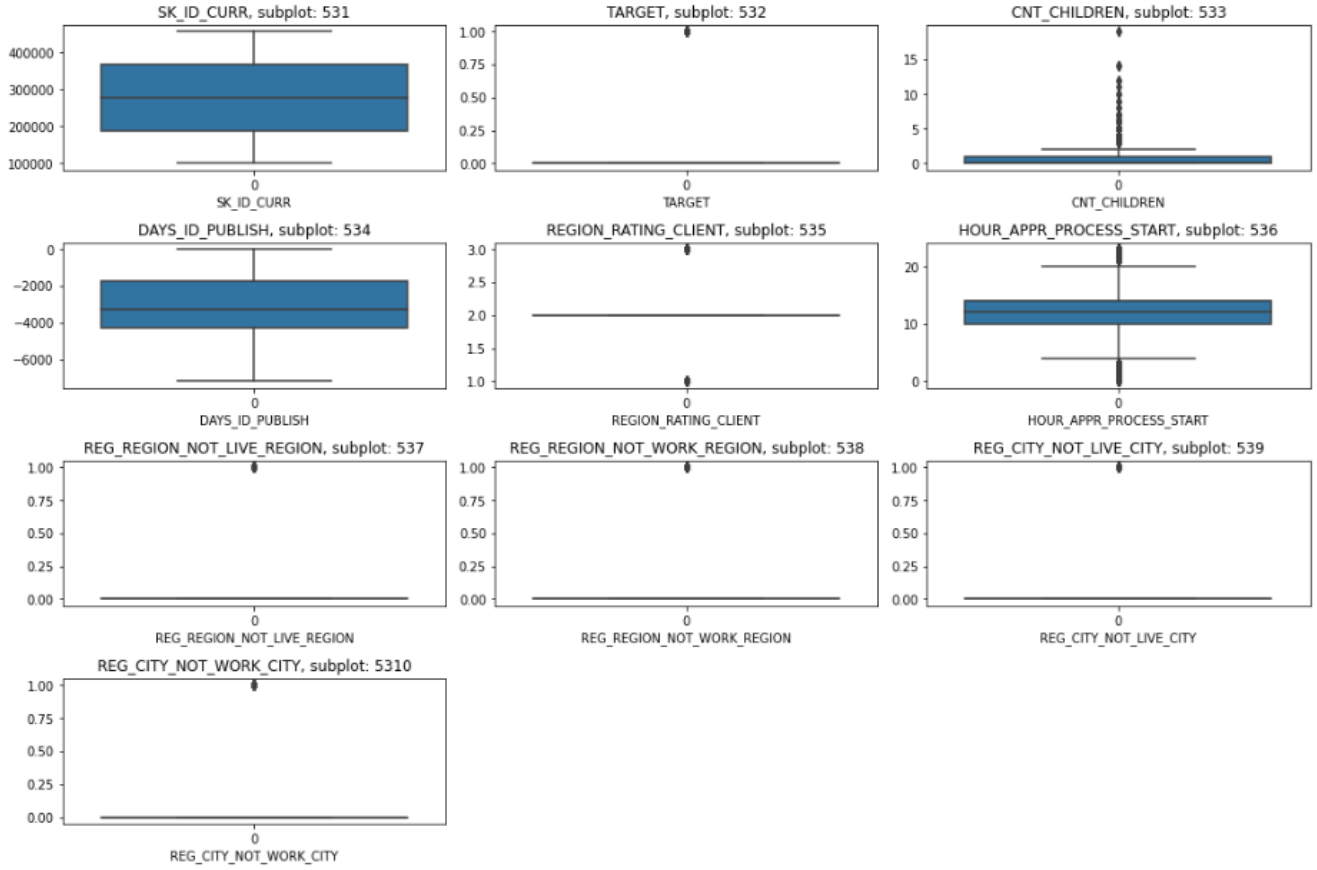




Insights:

- Age group 35-40 having high loans and age from 25 to 45 seems to have high default
- Experience of less than 5 years seems to have more loans and with payment difficulties
- Salary with 100-200k seems to have more loans

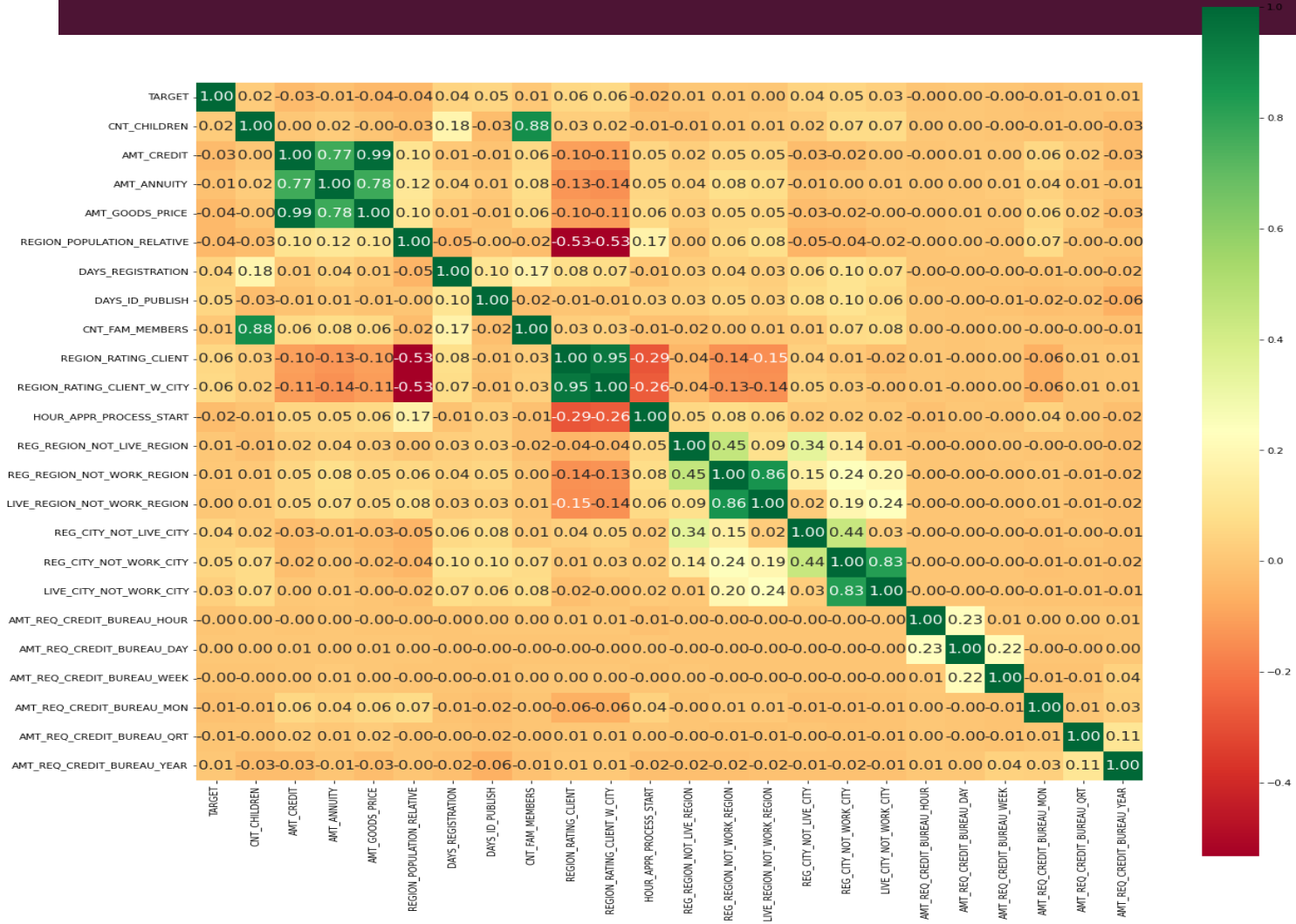
EXPLORATORY DATA ANALYSIS



Insights

- Many columns with int data type are Flag columns. For purpose of calculations we will keep them as int.
- CNT_CHILDREN has outlier which is possible. So it kept as it is

CORRELATION MATRIX



Insights:

Credit amount is highly correlated with Annuity Amount & Goods Price.

Region population relative is inversely correlated with Region Ratings.

Children is highly correlated with family members

DATA PREPARATION

- **Data Preparation**
  - **One-hot encoding for categorical variables**
  - **Aggregating with mean of trade level data to applicant level**
  - **Merging the two data sets**
  - **Dropping repeated & unwanted & highly correlated variables**

PRE MODEL & MODEL BUILDING & EVALUATION

- **Train & Test data split**
- **Feature scaling using standard scaler**
- **Handling class imbalance using SMOTE & TOMEK technique**
- **Feature selection using RFE**
- **Model building using Logistic Regression**
- **Model evaluation**
- **Cross validation**

MODEL EVALUATION

Comparing the values obtained for Train & Test data:

Train Data:

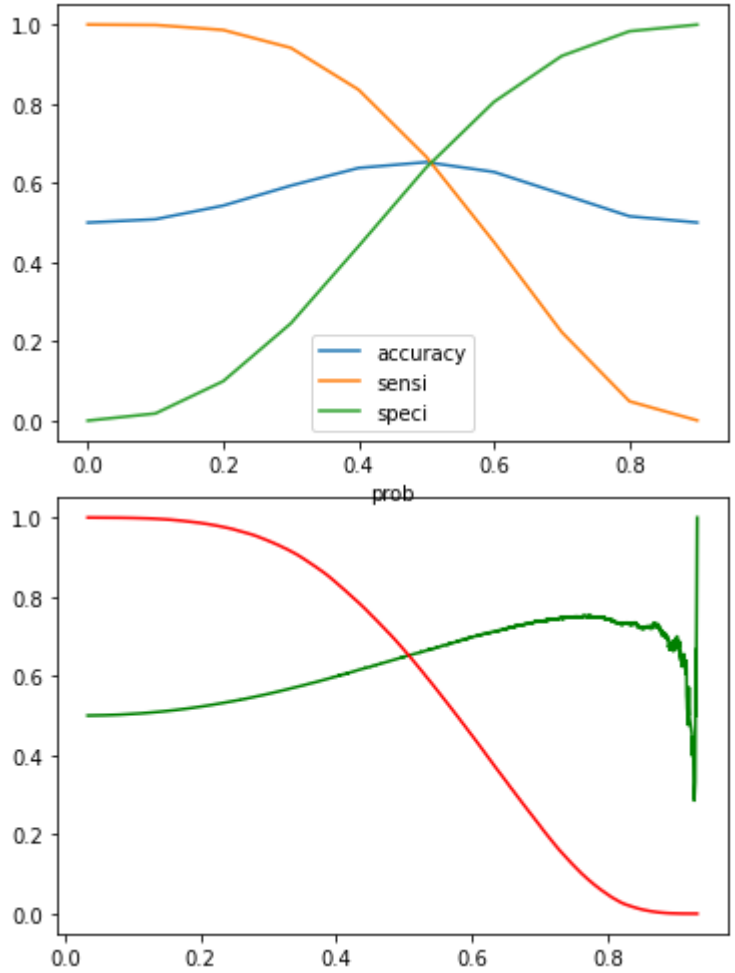
- **Accuracy : 65.26 %**
- **Sensitivity : 66.59 %**
- **Specificity : 63.93 %**

Test Data:

- **Accuracy : 63.53 %**
- **Sensitivity : 60.45 %**
- **Specificity : 63.79 %**

Precision & Recall

- **Precision: 64.86%**
- **Recall: 66.59%**



RECOMMENDATION

- Below 4 variables are the good predictor for separating good & bad borrowers

CODE_GENDER_M
REGION_RATING_CLIENT
DAYS_ID_PUBLISH
DAYS_CREDIT

- Below variables are having inversly proportional to default or having less exposure to default

EXPERIENCE_GROUP_10-20
AGE_GROUP_55-60
AGE_GROUP_above 60
NAME_CONTRACT_TYPE_Revolving loan
NAME_EDUCATION_TYPE_Higher education

ANSWERS:

How to leverage trade level information for Credit Bureaus by aggregating trade level information to applicant level in order to capture their payment behavior?
Answer: Using mean to aggregate trade level information to applicant level in order to capture the payment behaviour.

Which application or payment behavior factors significantly influence borrower's behavior on any new disbursed loan?
Answer: Below four variables highly influence borrowers behavior

CODE_GENDER_M
REGION_RATING_CLIENT
DAYS_ID_PUBLISH
DAYS_CREDIT

After identifying these factors, how to leverage them in the form of a model which can be used for decisioning?
Answer: Using RFE can help to identify the variables. The data should be standardized using standard scaler and target variable should be balanced uding SMOTE TOMEK or anyother technique.