

# HOME CREDIT SCORE CARD MODEL

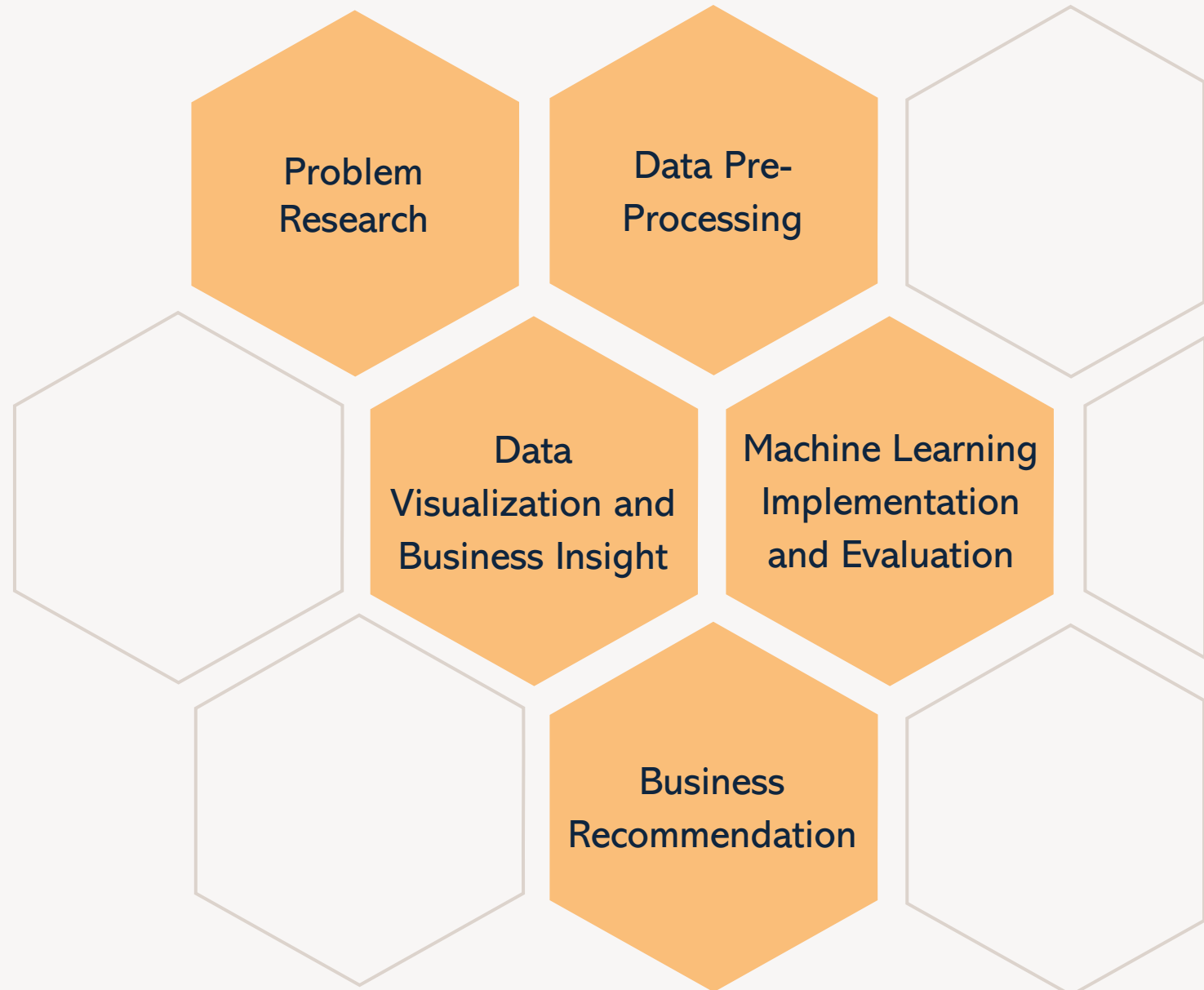
*Machine Learning Modelling and Impact Driven Business Insight for  
Score Card Prediction*

Hangga Bayu





# Agenda





# Problem Research

Home Credit Indonesia strives to broaden financial inclusion for the unbanked population by providing a safe and positive borrowing experience. The credit score prediction analysis unlocks the full potential of the data. Doing this, will ensure that clients who are able to make repayments are not rejected when applying for loans and that loans are given with a principal, maturity, and repayment calendar that will empower clients to be successful



# Objective and Action

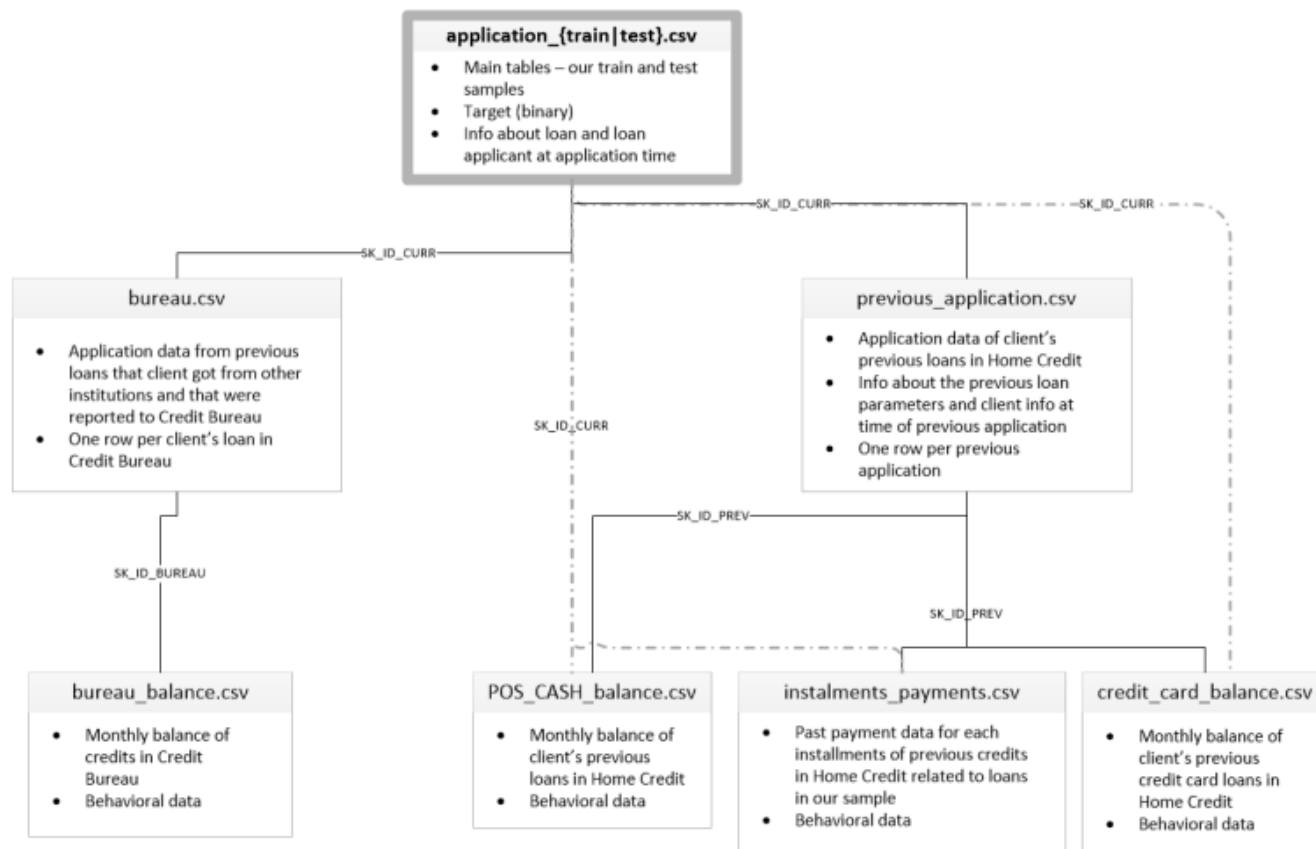
## Objective

- Predict the client's loan payment abilities
- Determine the key features that are correlated with the category of having payment difficulties
- Provide solutions to prevent and minimize clients who are approved but have loan payment difficulties

## Actions

- Determine the goal, objectives, and metrics of the problem.
- Conduct information mining related to the initial data condition.
- Perform Data Cleaning and Data Processing.
- Conduct insight discovery with reference to the set objectives.
- Perform modeling with various methods (including Logistic Regression) and their hyperparameters.
- Evaluate the modeling results.
- Measure the impact of modeling and business recommendations on the predetermined metrics.





Out[2]:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT
0	100002	1	Cash loans	M	N	Y	0	202500.0	400000.0
1	100003	0	Cash loans	F	N	N	0	270000.0	120000.0
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	100000.0
3	100006	0	Cash loans	F	N	Y	0	135000.0	50000.0
4	100007	0	Cash loans	M	N	Y	0	121500.0	50000.0

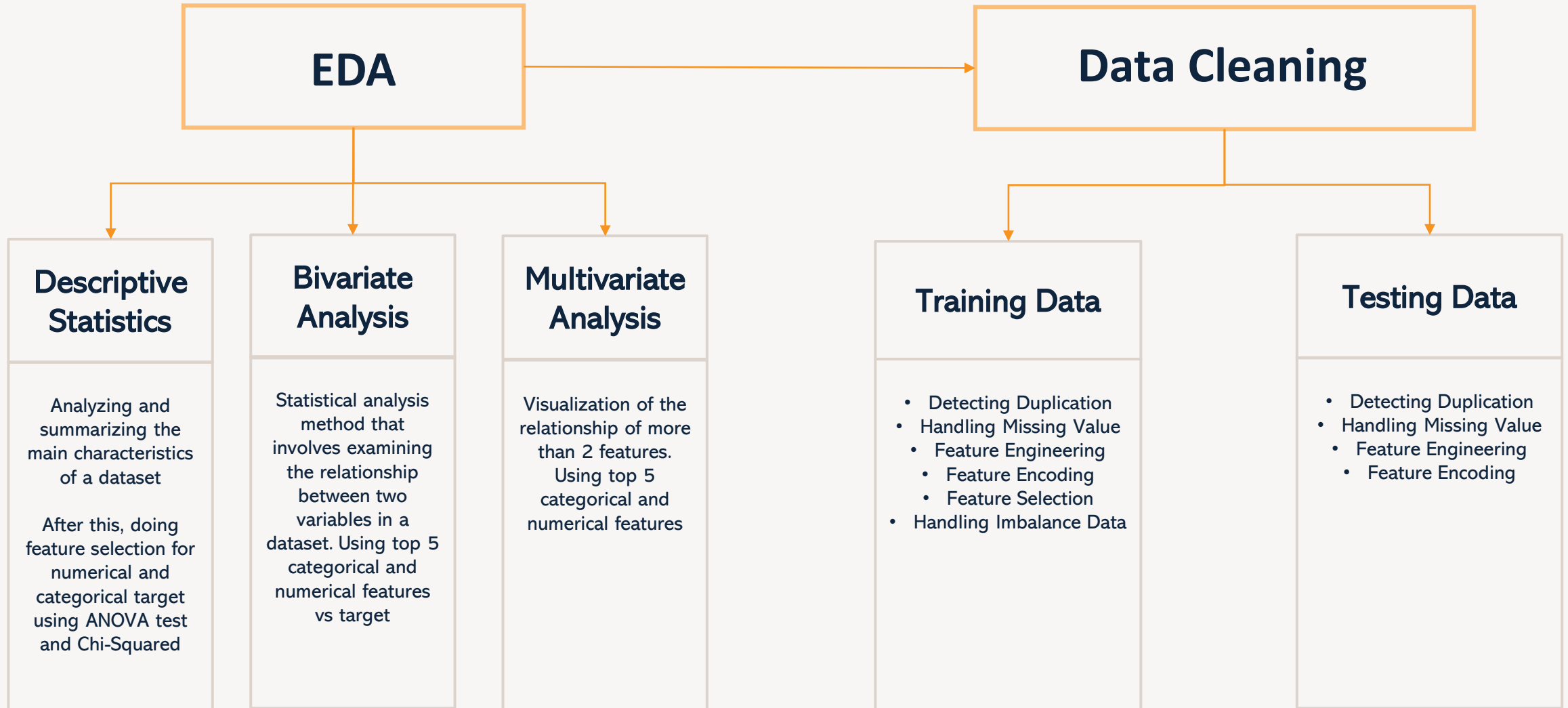
The train data has 307511 rows and 122 columns

	SK_ID_CURR	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT
0	100001	Cash loans	F	N	Y	0	135000.0	568800.0
1	100005	Cash loans	M	N	Y	0	99000.0	222768.0
2	100013	Cash loans	M	Y	Y	0	202500.0	663264.0
3	100028	Cash loans	F	N	Y	2	315000.0	1575000.0
4	100038	Cash loans	M	Y	N	1	180000.0	625500.0

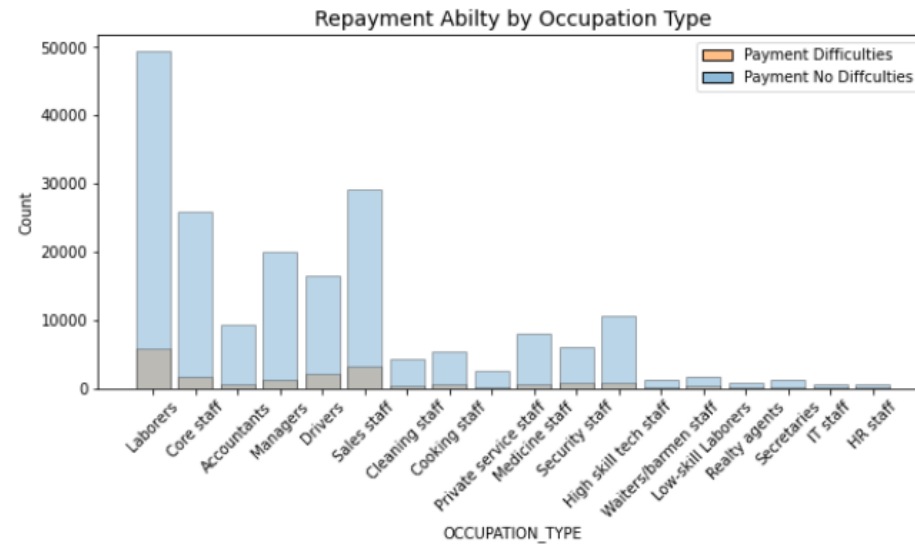
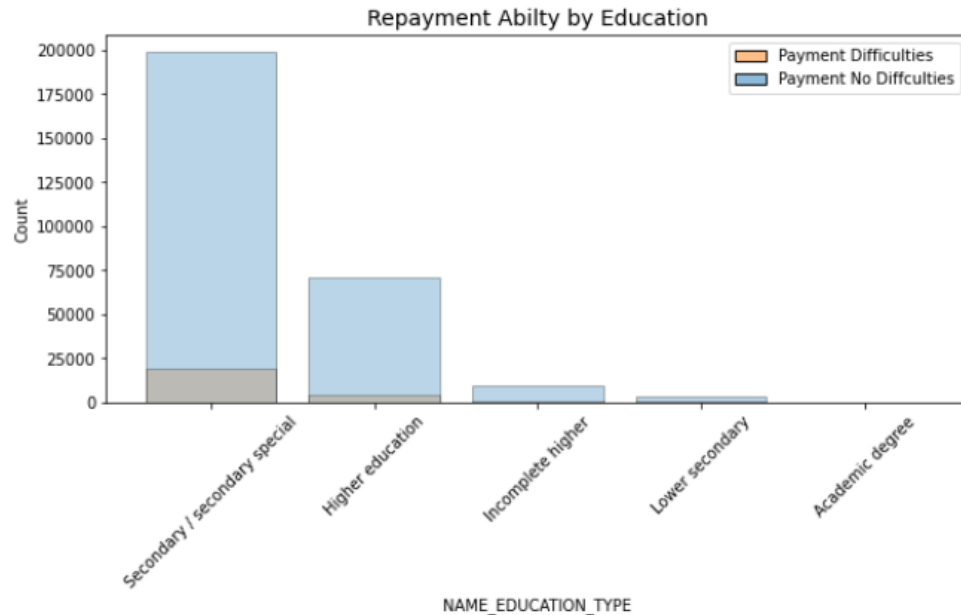
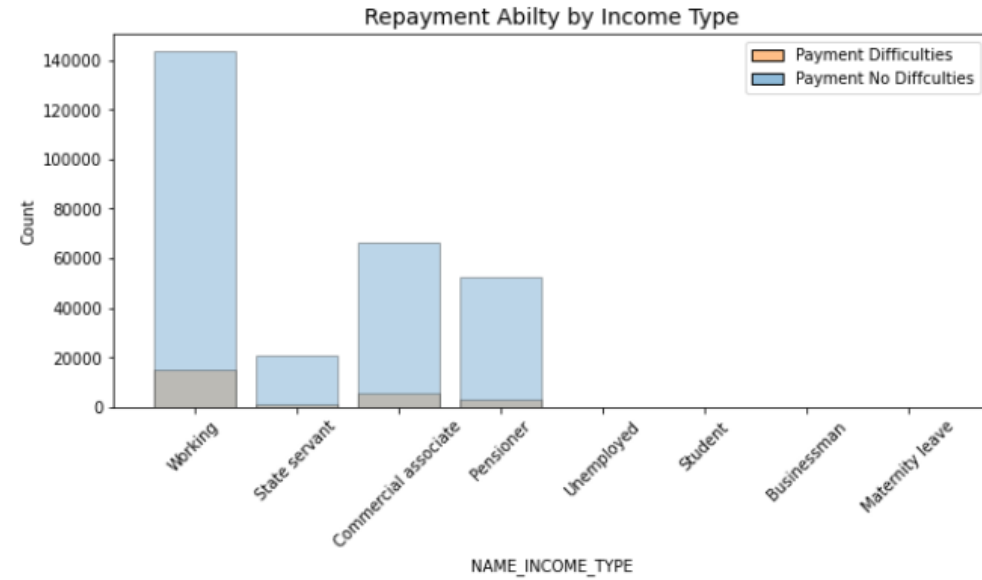
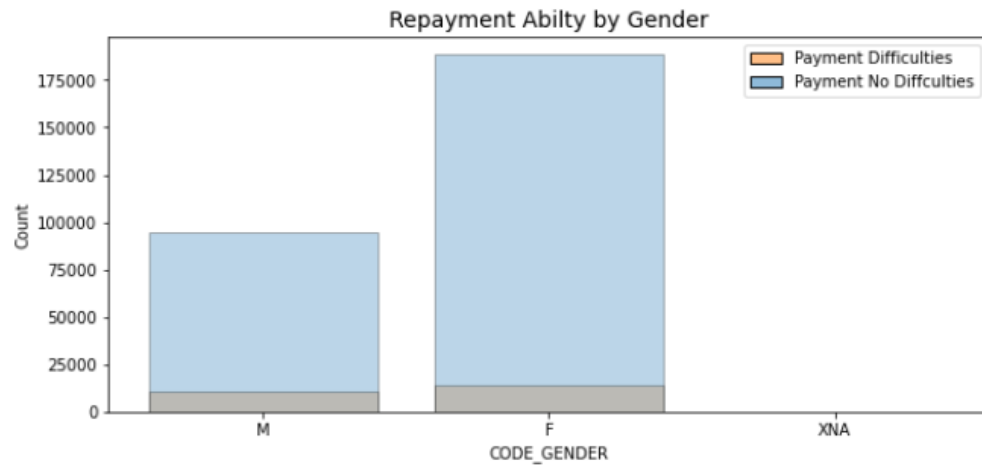
The test data has 48744 rows and 121 columns



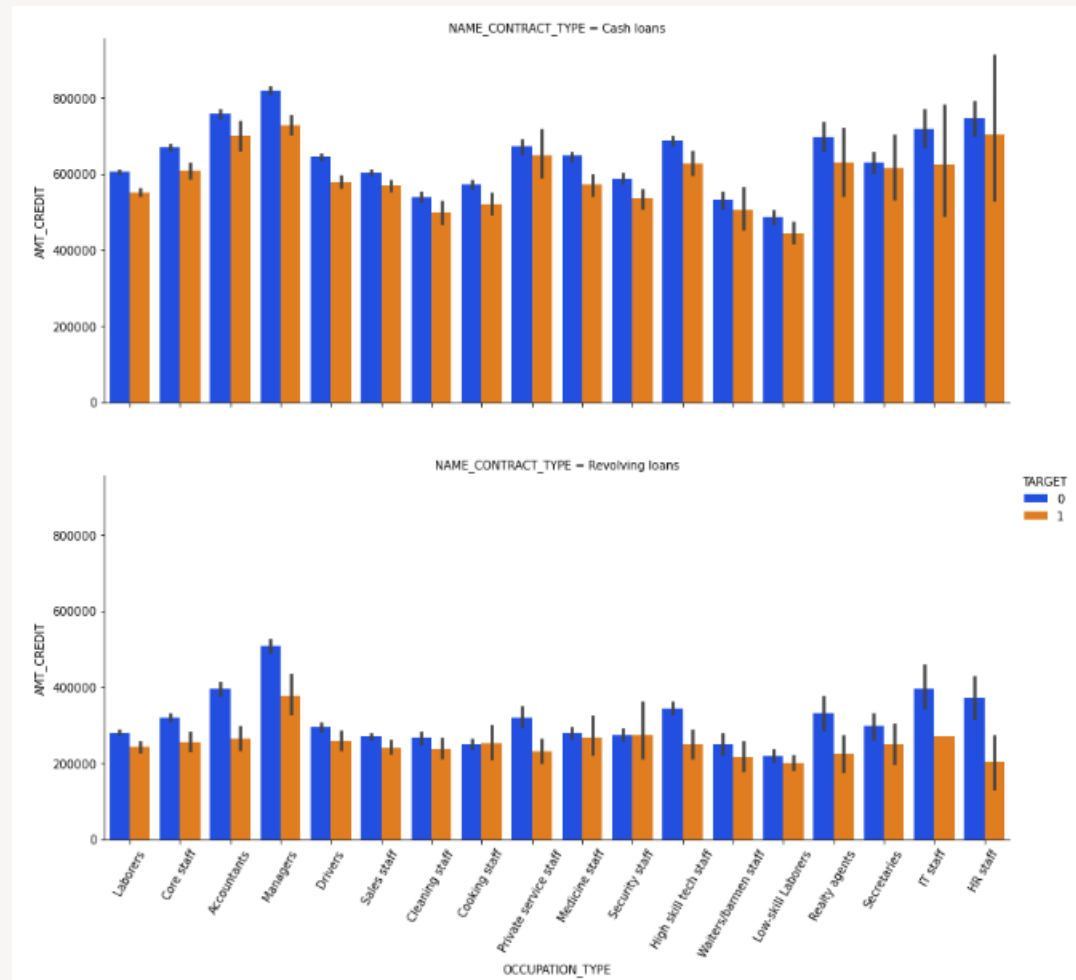
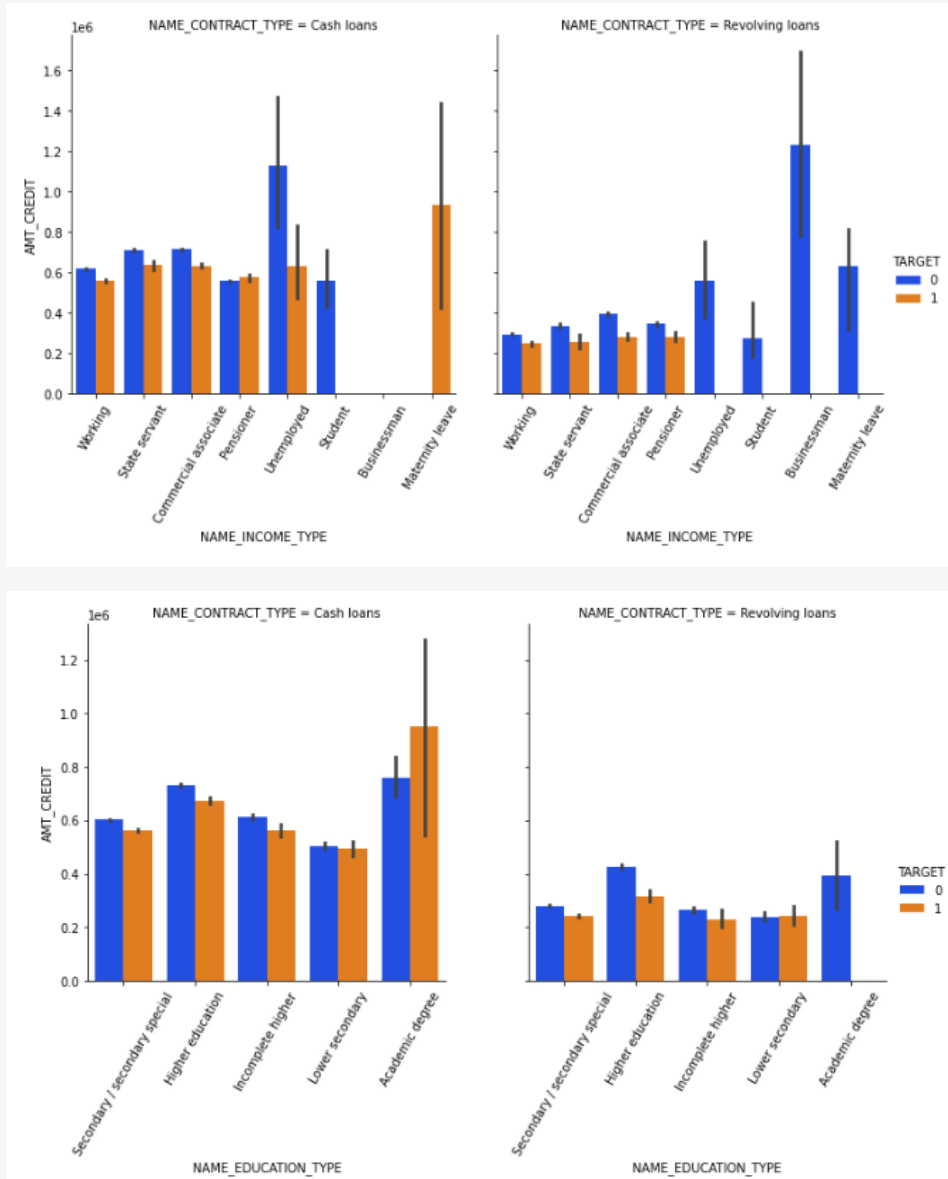
# DATA PRE-PROCESSING



# Data Visualization and Business Insight



# Data Visualization and Business Insight



- Student in either cash or revolving has no problem in payment abilities. Maternity leaves have no problem in revolving loans, in contrast in cash loans
- For the education type Academic degree, it has no difficulty in revolving loans, but in cash loans, payment difficulty has a higher number
- Manager, HR staff, IT staff, and realty agents are top 4 in the credit amount of the loan



# Machine Learning Implementation and Evaluation



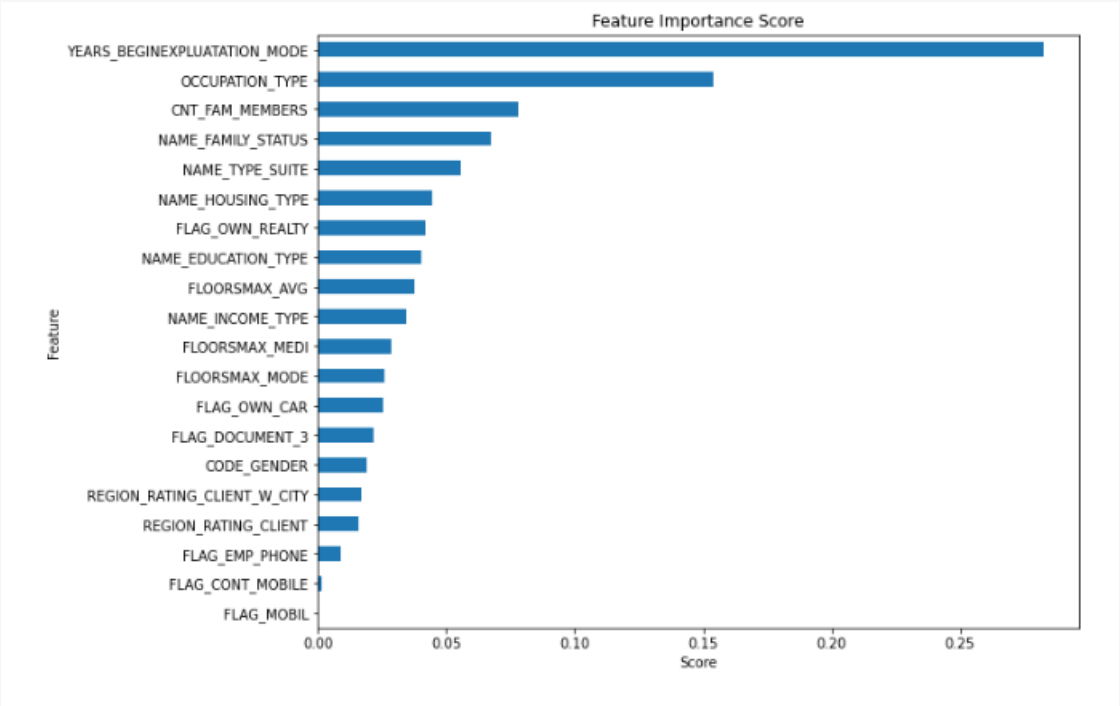
## Modelling Result

Models	Training Accuracy Score	Testing Accuracy Score	ROC AUC Score
Random Forest	0.8641	0.8455	0.8455
Decision Tree	0.8631	0.8277	0.8278
K-Nearest Neighbors	0.7671	0.7381	0.7381
XGBoost	0.6454	0.6393	0.6393
Logistic Regression	0.5922	0.5937	0.5937

- From the table above, the Random Forest model is chosen as the best model to predict client's payment abilities. Compared to other models, random forest has good accuracy and minimum error between training and testing



## Feature Importance from Random Forest Modelling



- The diagram shows the feature importance score from Random Forrest Classifier.

# Business Recommendation

1. High-paid job levels, such as managers, IT staff, and HR staff have a good capability to repay and have a high number credit amount, but it only has a small number of applications compared to other clients
2. Clients with income type of business man and students are potentially profitable clients because both of them have 100 % payment no difficulties. Unfortunately, they only have a small number of applications coming from them

Home Credit need to focus their marketing strategy to increase the number of clients with high-level paid job such as Manager, IT staff, and HR staff.

Create campaigns and special treatments for students and businessmen because the ratio between the number of applicants and payment ability has great potential

It is recommended further analysis (either in survey or modeling method ) choose the feature importance which has influenced to categorize of payment difficulty and payment no difficulty

## Project Documentation:

[https://github.com/HanggaBayu/HomeCredit\\_Score\\_CardModel](https://github.com/HanggaBayu/HomeCredit_Score_CardModel)

