

HOME Virtual Internship
CREDIT Program

CREDIT SCORING

CREATED BY ERLANDO REGITA



“BACKGROUND

As a data science intern at Home Credit, I was assigned to make a credit risk analysis and credit scoring. My objective here is to make a good prediction model that can classify whether the clients are having payment difficulties or not.

I used logistic regression as a model to predict credit risk and implemented weight evidence & information values to perform feature selection. Metrics used in this project are ROC-AUC score and KS-Statistic. My goals here are:

- build a prediction model with good ROC-AUC score (>0.7) and good KS-Statistic score (>0.3)
- build credit score (score card) each borrower and threshold recommendation list



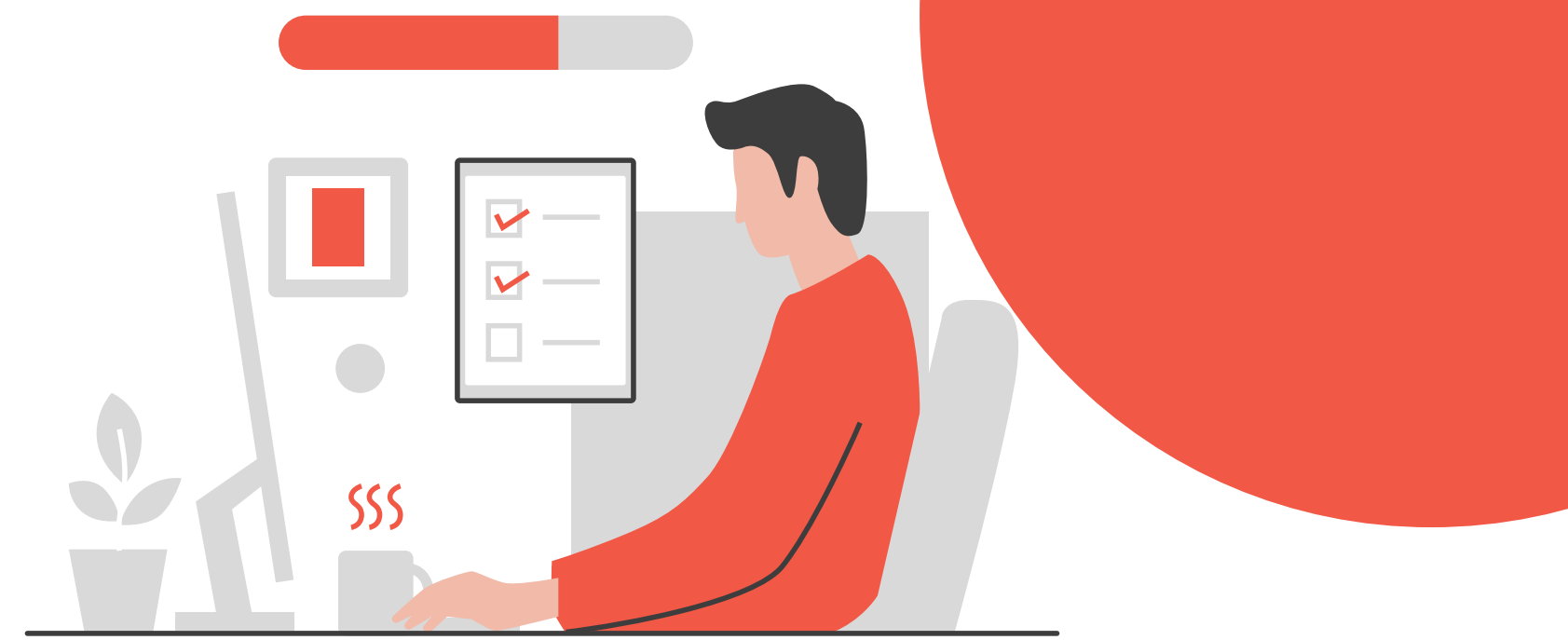
DATASET OVERVIEW

- This datasets is about client application record for credit loans
- The dataset consists of 121 features and 0.37 million records training dataset and 120 features + 38k record testing dataset
- There are 50 features containing >20% missing values, we dropped it and did missing values imputation for the rest features
- The target feature is labelled as client's payment difficulties, 1 stands for the client with payment difficulties.
- Target feature is highly imbalanced (91:9), we handled with SMOTE technique
- We only used training application dataset for modeling , we will only use additional dataset for gaining insights



[GITHUB REPO](#)

66 DATA PREPROCESSING



- Data preprocessing: dropped unused features, missing values imputation, datetime feature engineering, and feature encoding
- Feature Selection: using Weight of Evidence and Information Values
- Feature encoding and feature binning
- Split for training and testing
- Handling imbalanced target feature
- We got 79 features ready to be training in machine learning model



[GITHUB REPO](#)

“CREDIT SCORE MODELING

We found that logistic regression is the most proper algorithm in this case, and the result are



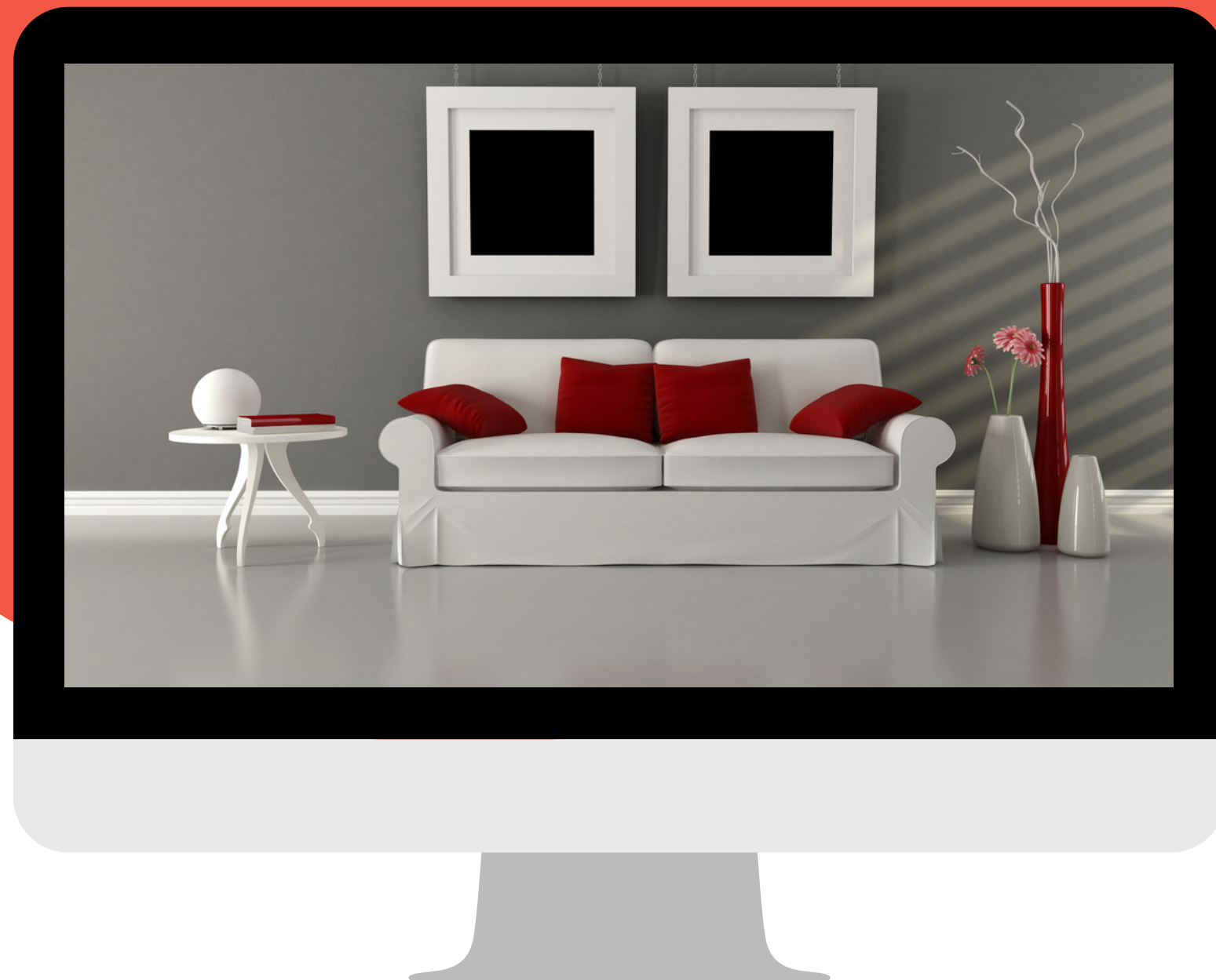
ROC-AUC SCORE

After tuning hyperparameter, we got 0.72 AUC Score



KS STATISTIC

And we got a 0.33 KS Statistic score, which is good because generally with >0.7 AUC score and > 0.3 KS Statistic in risk modelling are categorized as a good model



[GITHUB REPO](#)

CREDIT SCORE

*Glimpse of credit score using FICO scale

ID	CREDIT SCORE
429673	533
261440	623
389176	663

[Full Credit Score FICO Scale Records](#)

*Glimpse of credit score using 0-100 scale

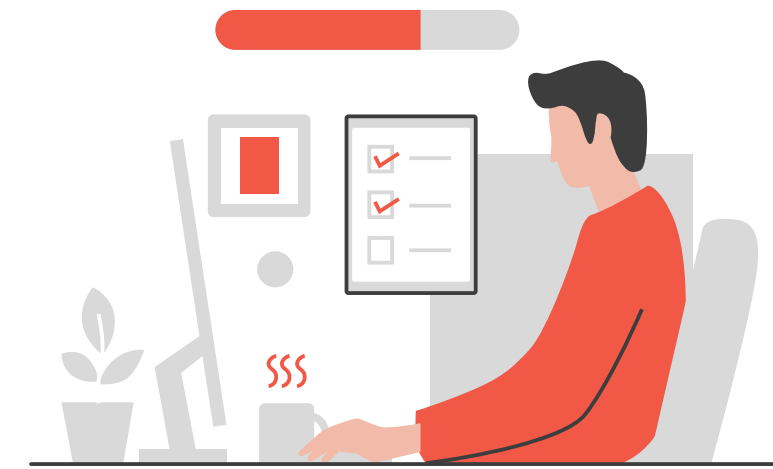
ID	CREDIT SCORE
227506	0.56
138480	0.53
389176	0.52

[Full Credit Score 100 Scale Records](#)

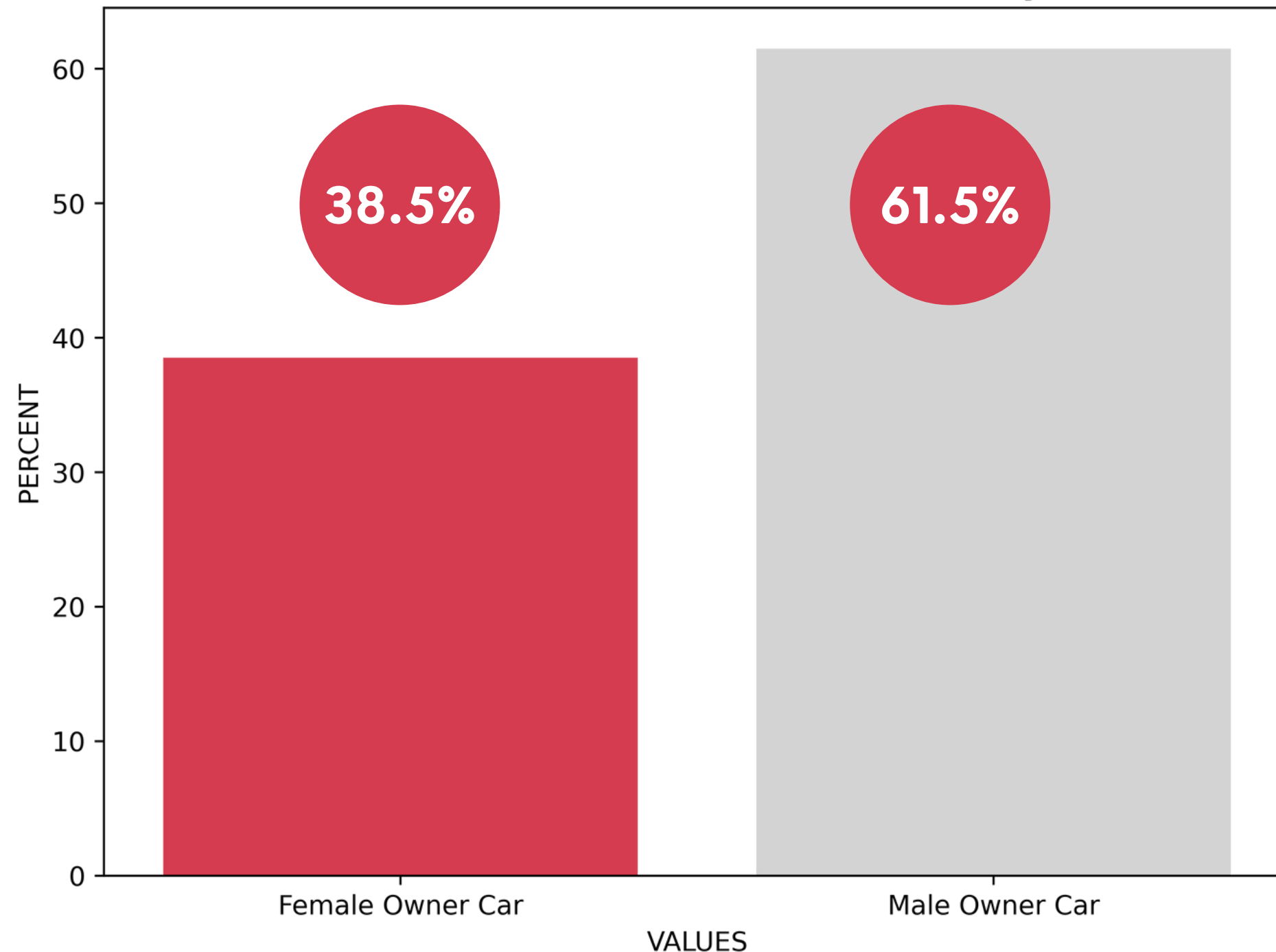


[GITHUB REPO](#)

“BUSINESS INSIGHT



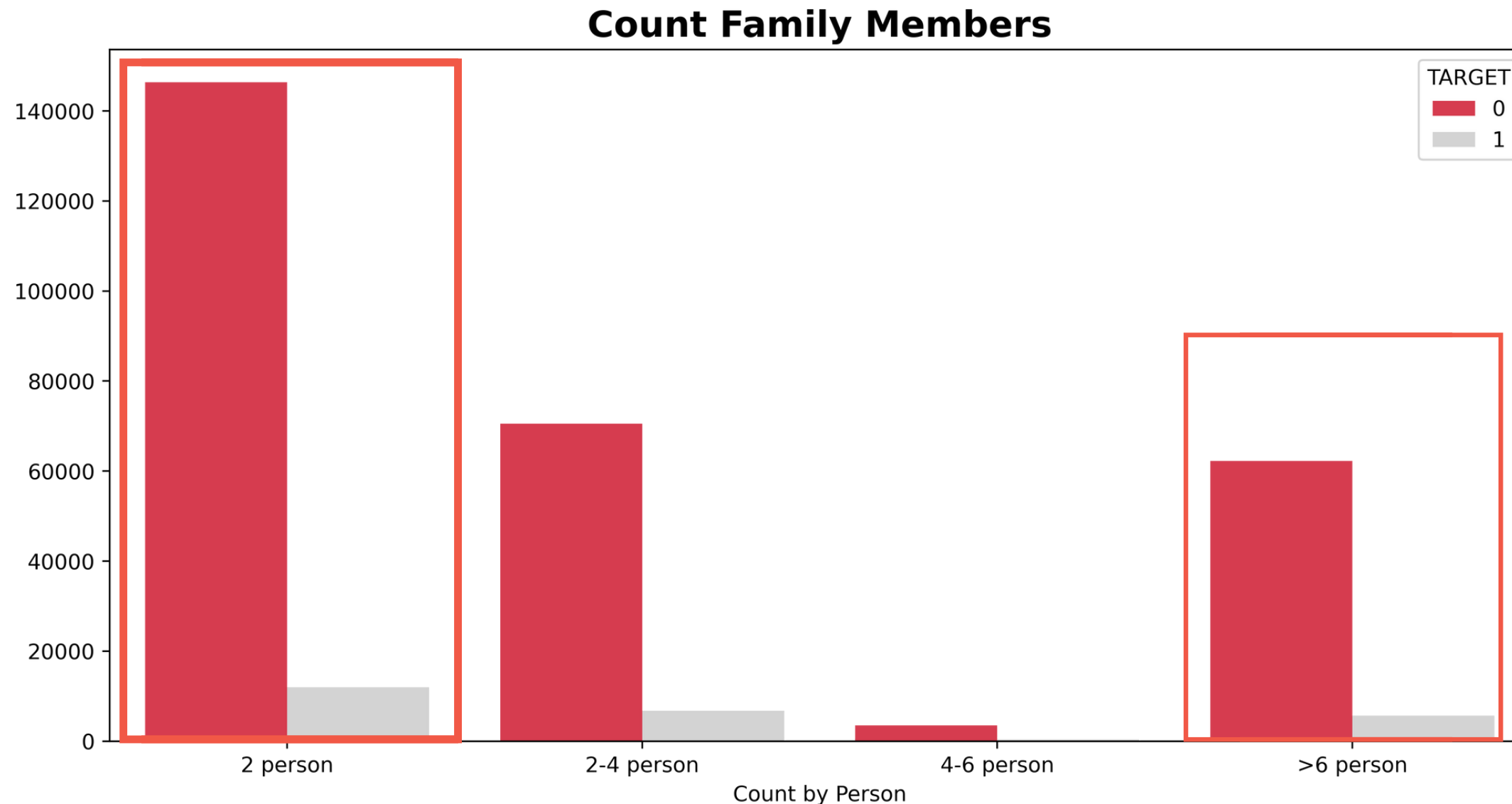
Ratio of Male/Female Car Owners who Have Payment Difficulties



- Gender and Car Ownership are the top 3 important features, demographically male with car ownership have a higher probability to be clients with payment difficulties
- We suggest marketing department to target female clients and/or male prospective clients with no car ownership and personalized campaign with demographic information related to next recommendation



BUSINESS INSIGHT



- Client with no children and have >4 children are the top 6 important features
- Our recommendation in marketing department, focus on client with less children and aware with client that have either no children or too many children
- It's make sense when family with no burdency and with too many burdency to have payment difficulties

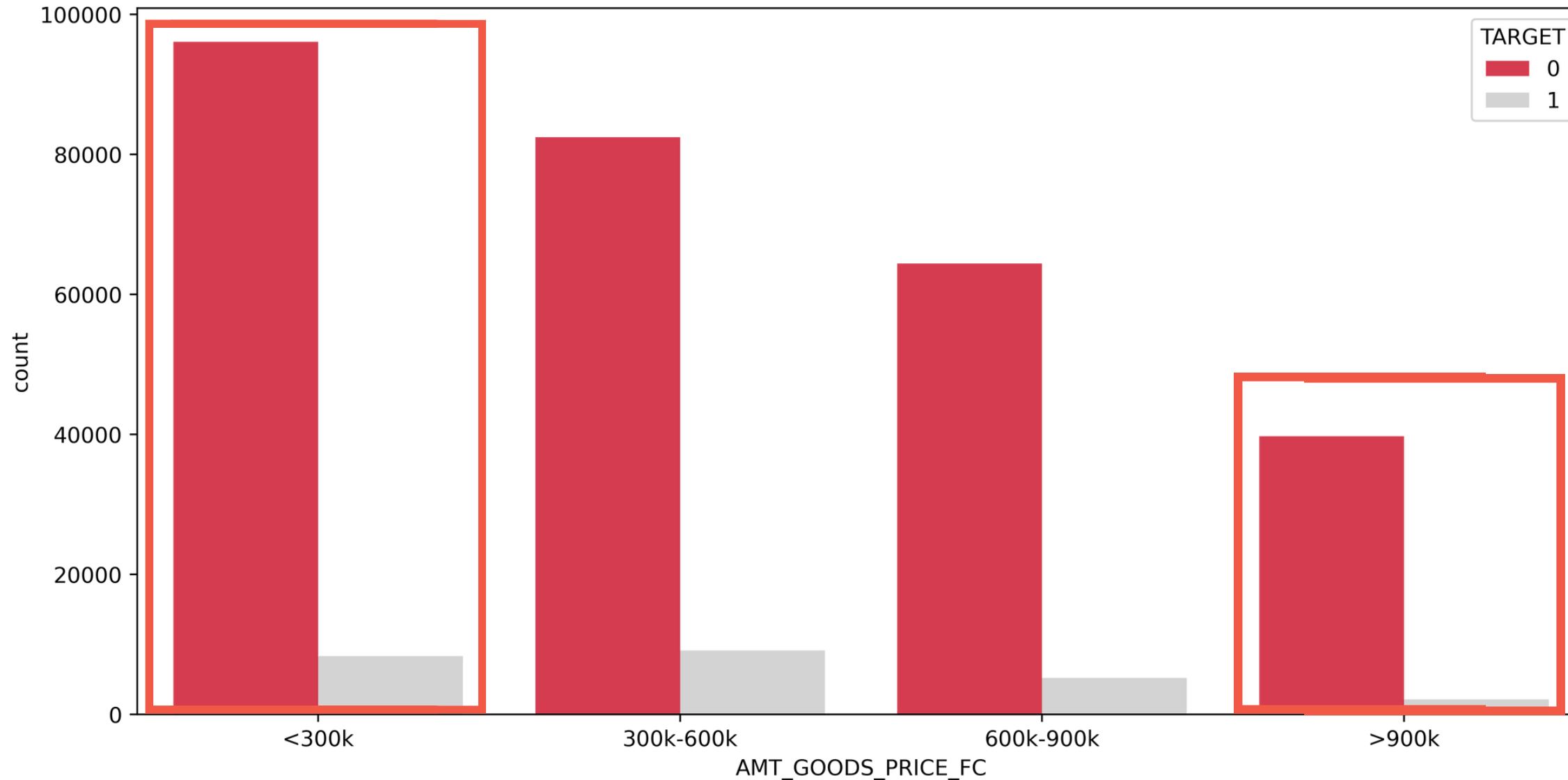


GITHUB REPO



BUSINESS INSIGHT

Price of the Goods for Which the Loan is Given



- Client with amount price for loans < 300k and >900k are top 10 feature importance
- Our recommendation in marketing department, between those range because it has lower probability to have payment difficulties
- Company should aware with too little amount goods price and/or with too high amount goods of price.



[GITHUB REPO](#)

HOME Virtual Internship
CREDIT Program

THANK
YOU

