



RAKAMIN ACADEMY X HOME CREDIT INDONESIA



# HOME CREDIT: CREDIT SCORE MODELLING

Presented By Radhimas Januar Rachman

# TABLE OF CONTENTS

- 1 Problem Research
- 2 Data Cleansing
- 3 Business Insights
- 4 Machine Learning Model
- 5 Business Recommendation

# BACKGROUND PROBLEM

Numerous individuals encounter difficulties in securing loans because of limited or absent credit records. Home Credit aims to expand financial access for those without banking services by offering a secure and favorable borrowing environment. To ensure a positive loan journey for this underserved demographic, Home Credit utilizes diverse alternative data sources to assess their clients' repayment capacities. 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.



# OBJECTIVES & ACTIONS



## Objective

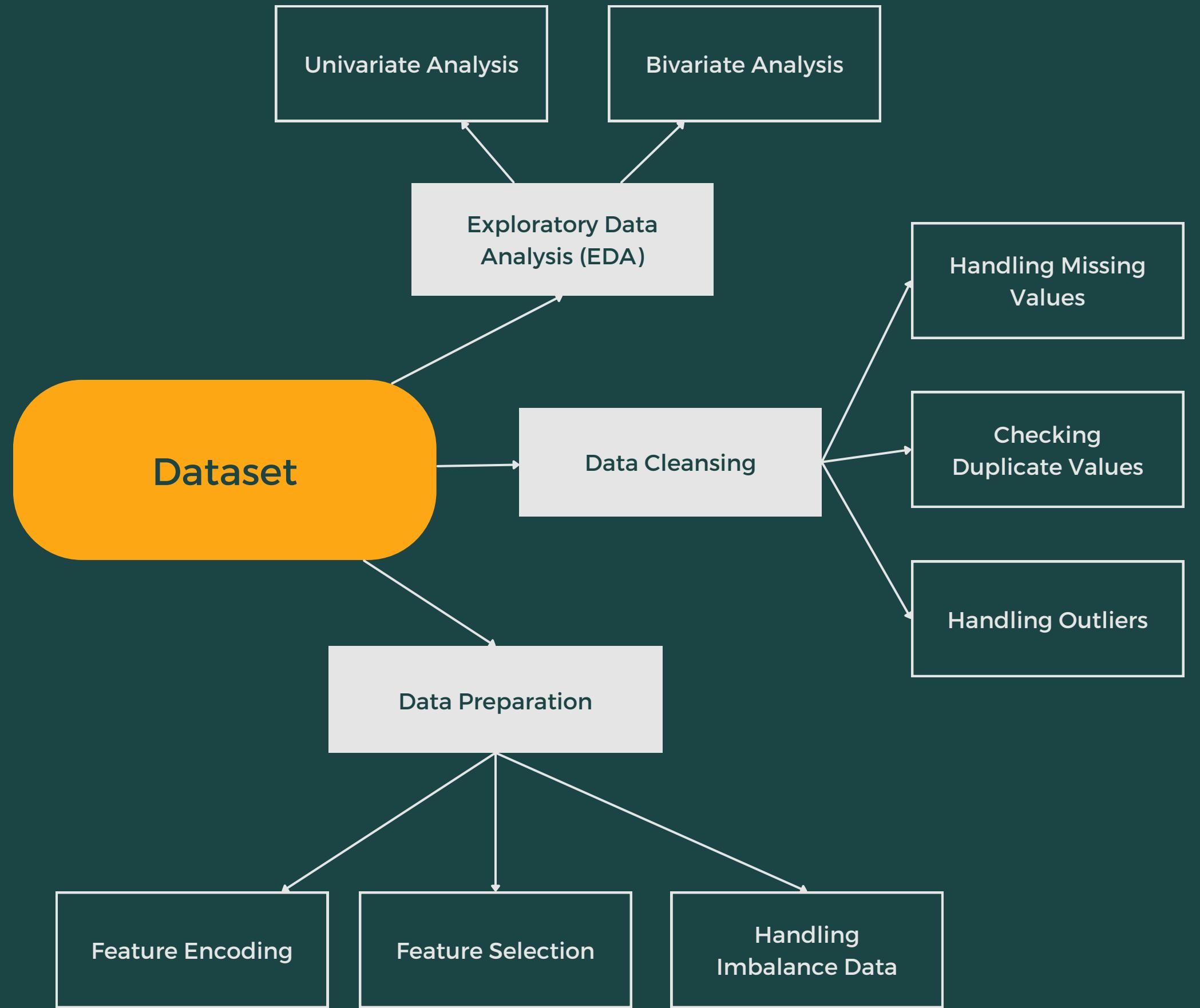
## Actions

1. Identify characteristics of potential clients who will have difficulty repaying loans.
2. Predict client's repayment abilities.

1. Conduct data cleansing and visualization to get business insights.
2. Construct models using machine learning algorithms and determine the most optimal model.
3. Provide recommendations for the company to enhancing the success rate of clients' loan applications.

# DATA PREPROCESSING





## EDA

### Univariate Analysis

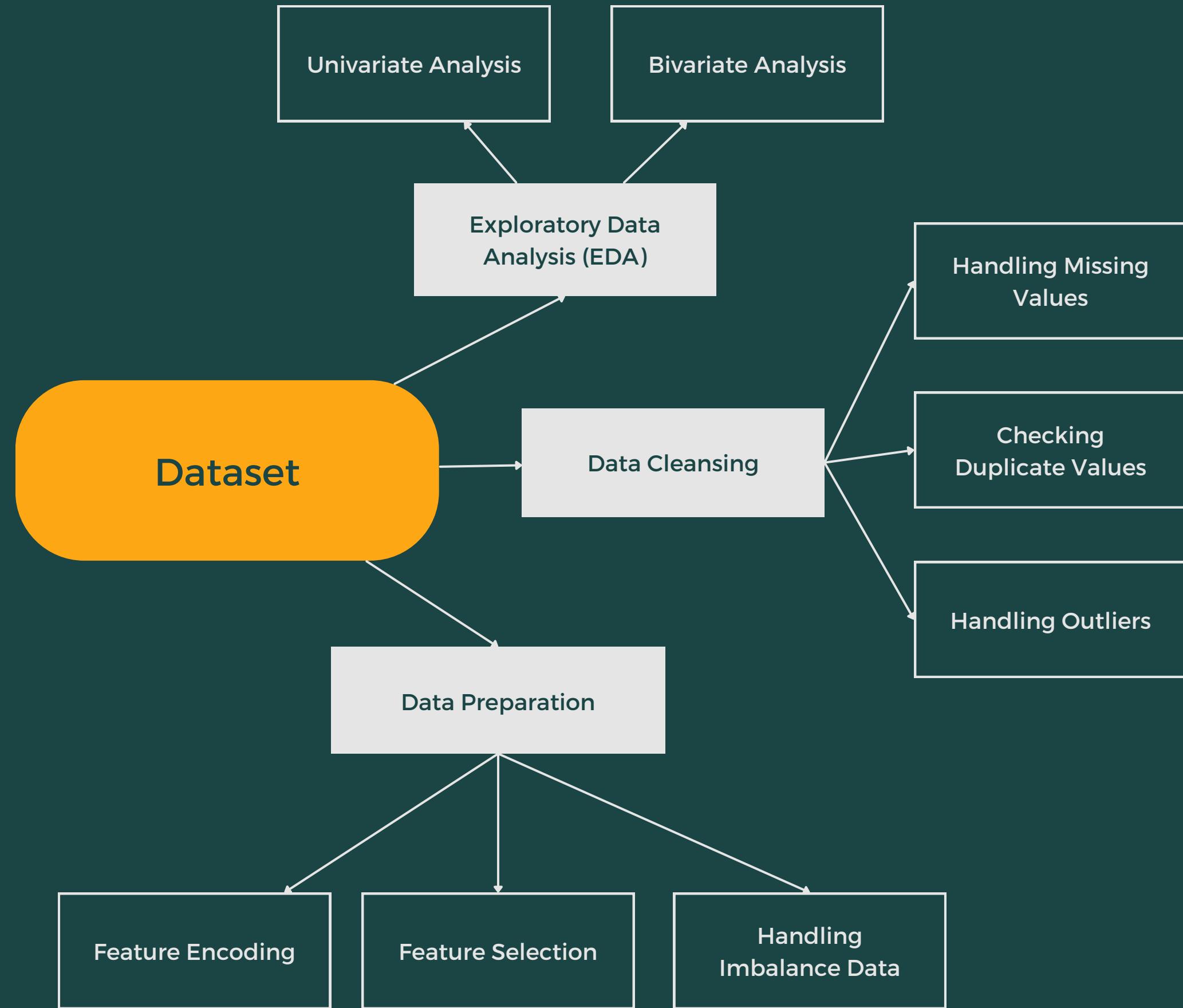
- Visualization of categorical and numerical features

### Bivariate Analysis

- Visualization between 2 features

### Data Cleansing

- Some missing values are dropped, some are imputed
- No duplicate values detected
- Removed outliers using IQR



## Data Preparation

### Feature Encoding

- Categorical features are encoded using OHE

### Feature Selection

- Identify top 25 features using correlation score

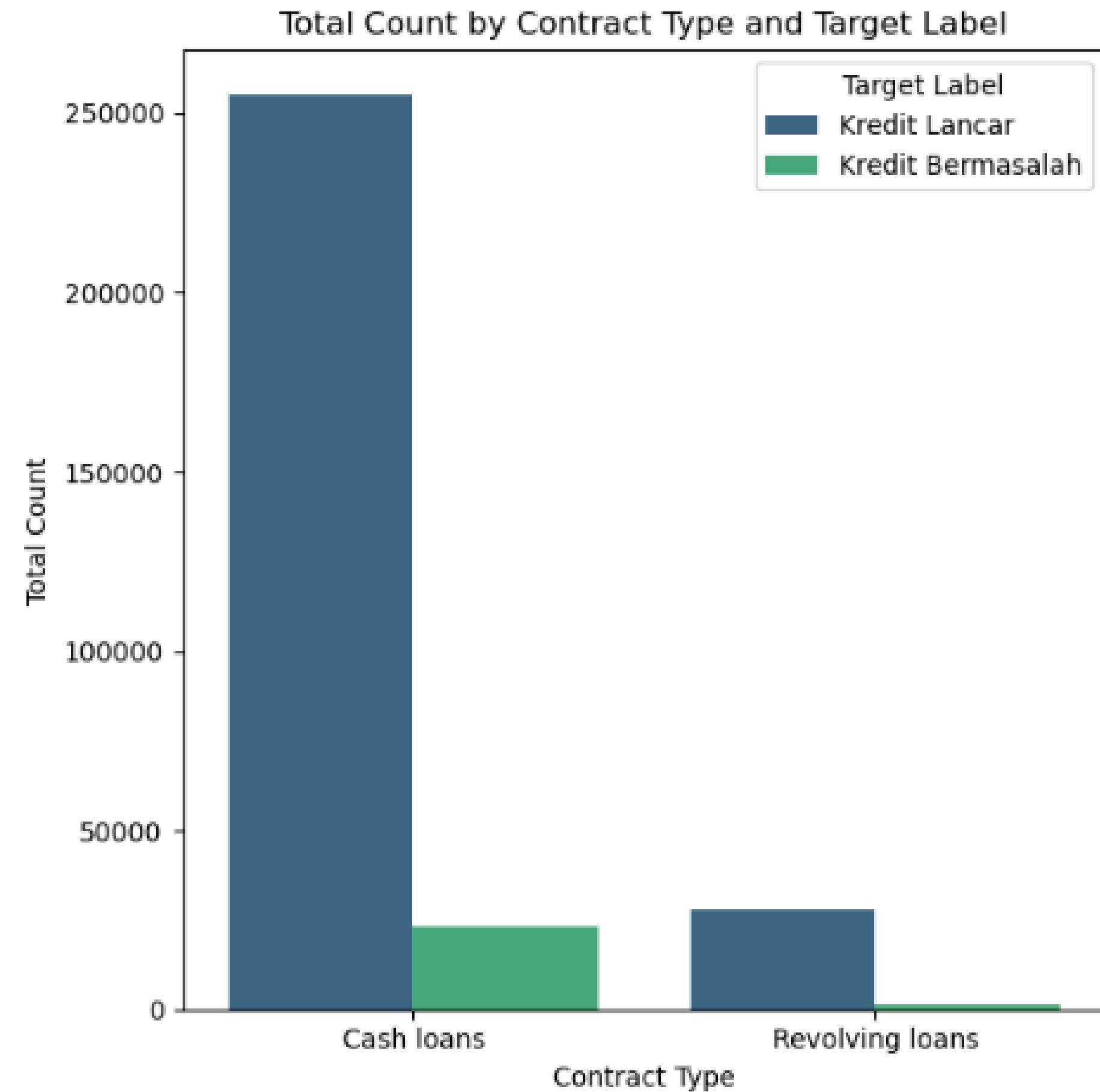
### Handling Imbalanced Data

- Resampling data so that target label is balanced.

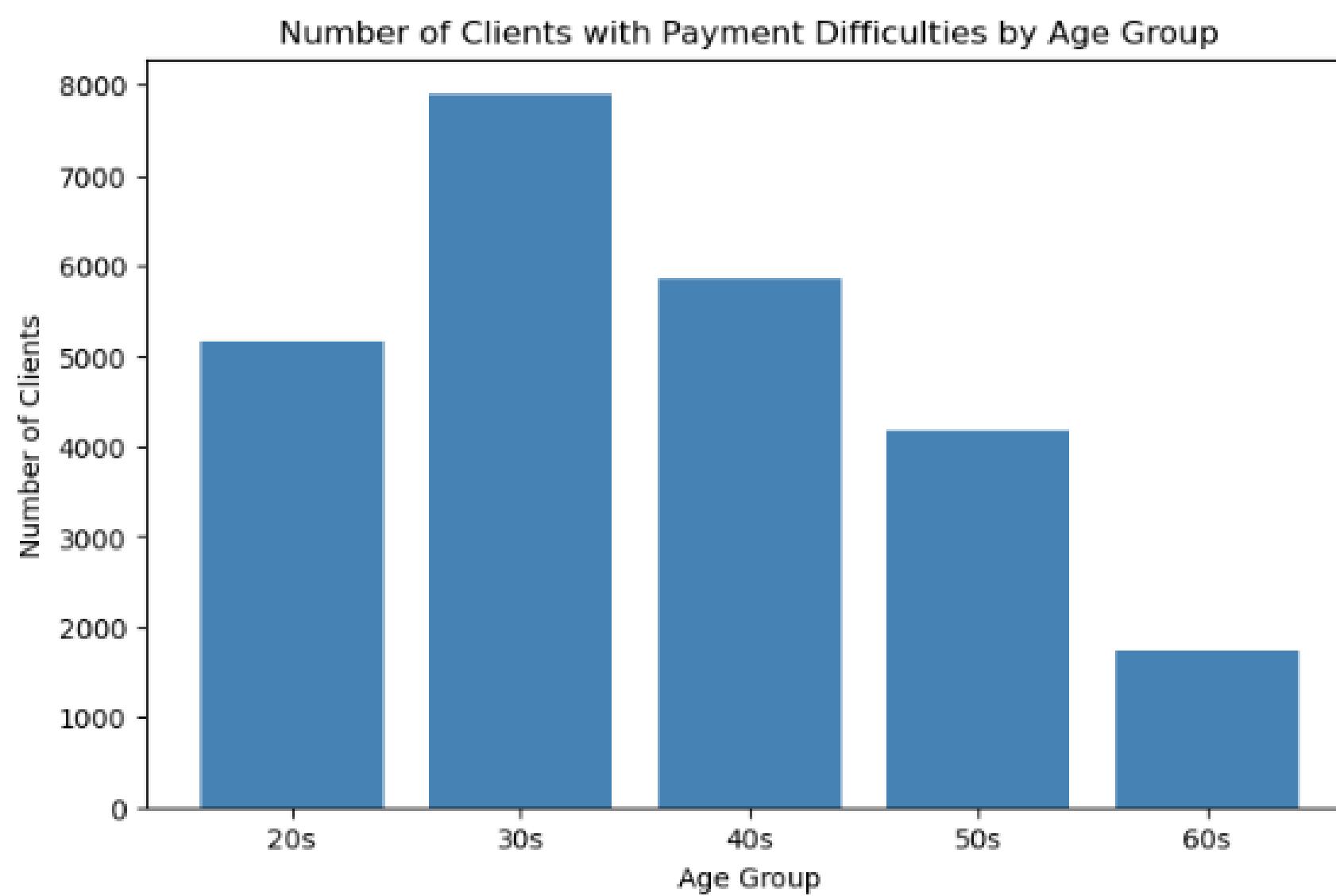
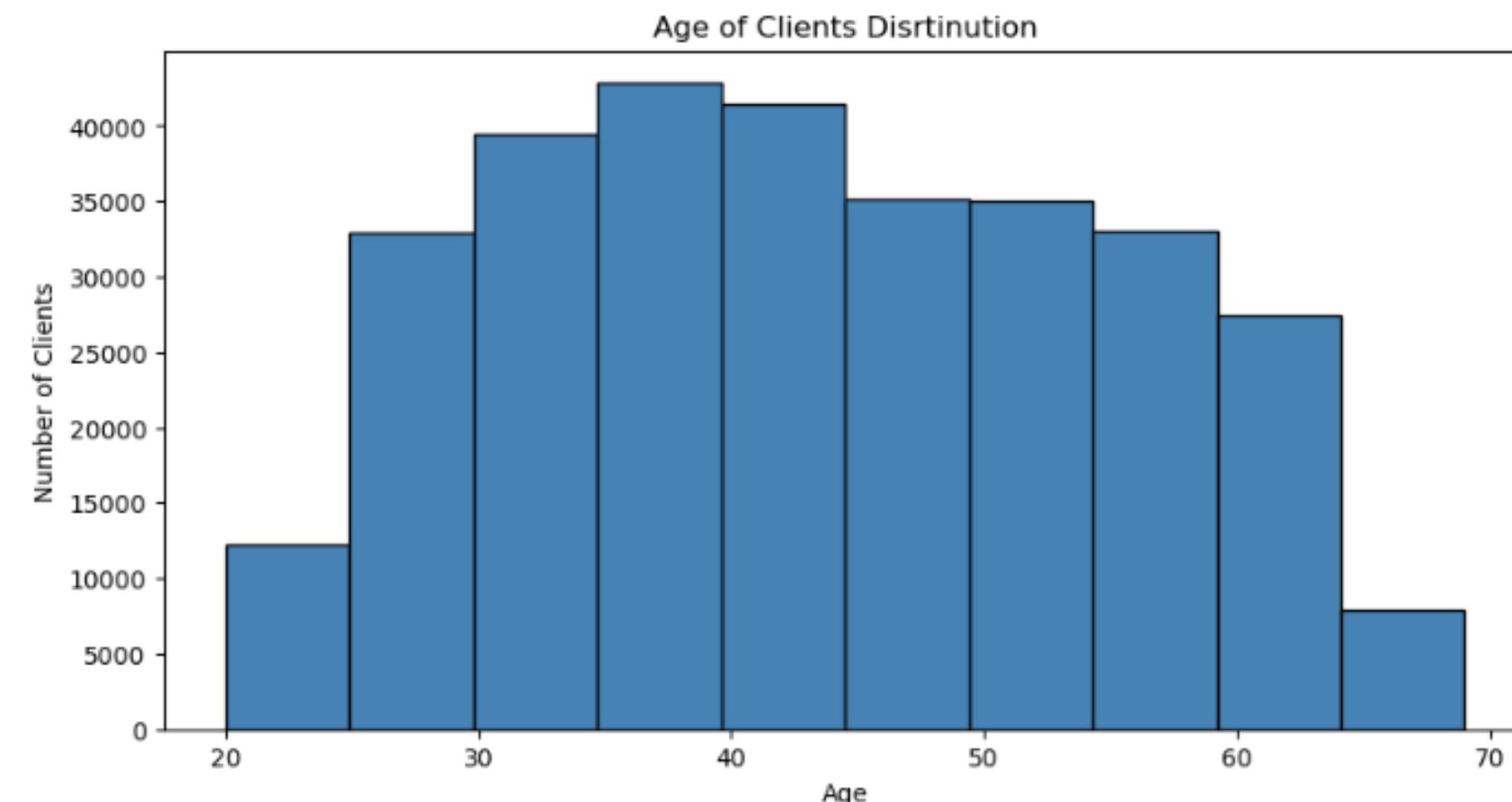
# BUSINESS INSIGHTS



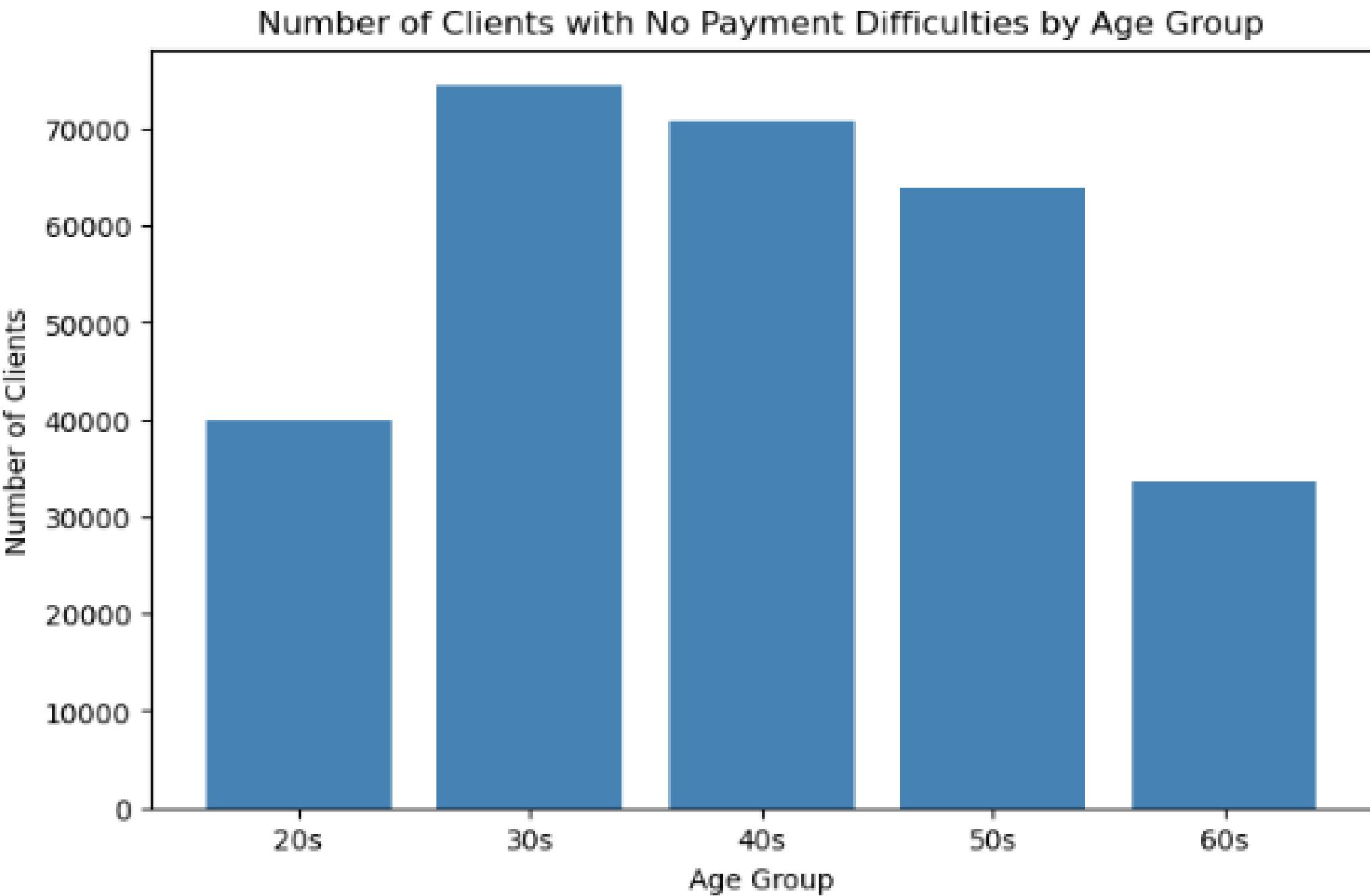
The proportion of "kredit bermasalah" is relatively lower for Cash loans (23.2K) compared to Revolving loans (1.6K). This suggests that clients using Revolving loans tend to have a slightly higher level of non-performing loan risk compared to Cash loans.



- Clients are generally around 30-40 years old
- Clients with credit repayment problems were most likely to be in their 30s, while the least likely to have credit repayment problems were around their 60s



- Clients with no credit repayment problems are generally to be in their 30s to 40s



**Occupations with "Laborers" and "Sales Staff" are the most common occupations in the data, and also tend to have a higher number of "Kredit Lancar" (current loans) than "Kredit Bermasalah"(non-performing loans).**

| OCCUPATION_TYPE       | TARGET_label      | total |
|-----------------------|-------------------|-------|
| Laborers              | Kredit Lancar     | 49348 |
| Sales staff           | Kredit Lancar     | 29010 |
| Core staff            | Kredit Lancar     | 25832 |
| Managers              | Kredit Lancar     | 20043 |
| Drivers               | Kredit Lancar     | 16496 |
| High skill tech staff | Kredit Lancar     | 10679 |
| Accountants           | Kredit Lancar     | 9339  |
| Medicine staff        | Kredit Lancar     | 7965  |
| Security staff        | Kredit Lancar     | 5999  |
| Laborers              | Kredit Bermasalah | 5838  |
| Cooking staff         | Kredit Lancar     | 5325  |
| Cleaning staff        | Kredit Lancar     | 4206  |
| Sales staff           | Kredit Bermasalah | 3092  |

# MACHINE LEARNING MODELLING



# MODEL SELECTION

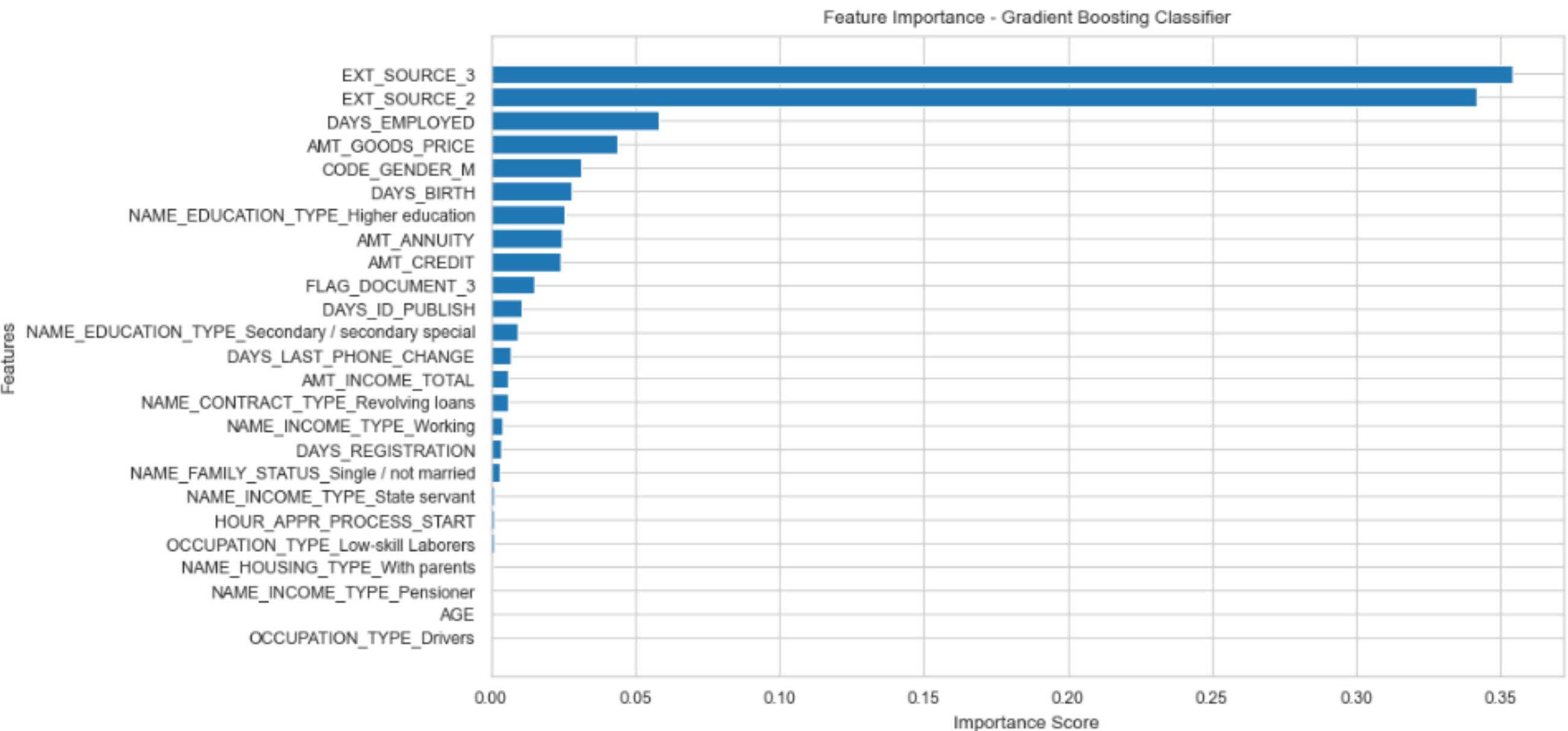
|   | Models              | Training Accuracy Score | Testing Accuracy Score | ROC Train Score | ROC Test Score |
|---|---------------------|-------------------------|------------------------|-----------------|----------------|
| 0 | Random Forest       | 1.000000                | 0.919700               | 1.000000        | 0.712000       |
| 3 | Neural Network      | 0.919700                | 0.919600               | 0.763700        | 0.738000       |
| 2 | Gradient Boosting   | 0.919600                | 0.919700               | 0.750300        | 0.743200       |
| 1 | Logistic Regression | 0.919200                | 0.919600               | 0.734500        | 0.737500       |

Based on the results provided, the Neural Network model and the Gradient Boosting model have approximately the same performance. However, considering the higher ROC value of the Gradient Boosting model on both train and test data it can be considered slightly better for predicting TARGET labels. So the best model chosen for prediction is Gradient Boosting.

# FEATURE IMPORTANCE

Based on the feature importance visualization, the Top 5 features are:

- EXT\_SOURCE\_3
- EXT\_SOURCE\_2
- DAYS\_EMPLOYED
- AMT\_GOODS\_PRICE
- CODE\_GENDER\_M





# BUSINESS RECOMMENDATION

# SUMMARY

## Target vs Occupation

Occupations with "Laborers" and "Sales Staff" are the most common occupations in the data, and also tend to have a higher number of "Kredit Lancar"(current loans) than "Kredit Bermasalah"(non-performing loans).

Of the two occupations, Sales Staff has a slightly higher ratio of Kredit Lancar than Laborers based on the ratio of total to Kredit Lancar/ Kredit Bermasalah. This means that Sales Staff tend to be slightly more current in repaying loans than Laborers.

---

## Target vs Age

Clients with credit repayment problems were most likely to be in their 30s, while the least likely to have credit repayment problems were in their 60s.

---

## Target vs Contract Type

From the comparison of total and Kredit Lancar, as well as Kredit Bermasalah, it can be seen that Cash loans perform better than Revolving loans:

- 1.The total amount of "Kredit Lancar"(current loans) in Cash loans is higher than Revolving loans.
- 2.The total number of "Kredit Bermasalah"(non-performing loans) in Cash loans is lower than Revolving loans.

Based on this comparison, Cash loans can be considered better performing than Revolving loans.

# RECOMMENDATION

- 1. The occupations Laborers and Sales Staff are among the professions with the largest numbers in the data. Companies can strengthen marketing strategies by focusing more on client segments that tend to have higher current credit, thus targeting marketing more effectively to these segments.**

---
- 2. Clients in their 30s tend to experience credit repayment problems, companies can offer credit products that are more suitable for their needs and financial conditions such as providing special benefits such as more flexible payment terms.**

---
- 3. Cash loans perform better than Revolving loans, the company can increase promotion through campaigns and product development of Cash loans.**

---

# PROJECT DETAILS

GITHUB  
LINK





**THANK YOU**