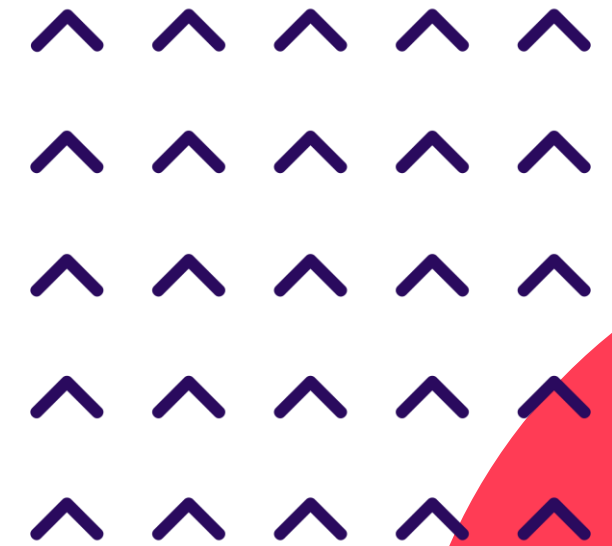


Decreasing The Number of Customer with Difficulties



**Analysis by Pinandito Wahyu
using Company Datasets**

**HOME
CREDIT**

Latar Belakang



Home Credit Indonesia sebagai perusahaan *fin-tech* tidak pernah berhenti mengoptimalisasi produk dan system-nya agar dapat terus berkembang.

Dalam eksplorasi dataset yang telah dilakukan, ditemukan salah satu permasalahan yang berkaitan dengan *customers*, yaitu: jumlah **Customer with Difficulties** yang saat ini berada di angka 8,1% dari keseluruhan *customers* Home Credit Indonesia.

Customers with Difficulties ini diartikan sebagai *customers* yang memiliki kesulitan dalam pembayaran cicilannya.

Dalam analisis dan eksplorasi data yang dilakukan menggunakan pemodelan dan visualisasi, dapat ditemukan cara untuk mengatasi permasalahan ini.

Problem Research

Business Metrics:
Customers Acquisition
Unique Buyers



Problem:

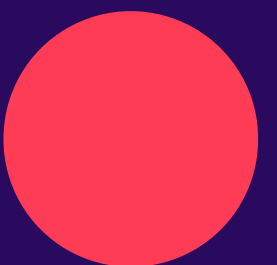
8,1% customers mengalami kesulitan/bermasalah dalam pembayaran cicilan kredit

Goals:

Menurunkan angka persentase **Customers with Difficulties** menjadi kurang dari 5%

Objective:

Mempelajari latar belakang customers yang bermasalah dan menciptakan *predictive model* yang tepat untuk diterapkan pada sistem deteksiajuan kredit.



Exploratory Data Analysis ●●●

```
df_feature.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  -
0   SK_ID_CURR          307511 non-null int64
1   TARGET              307511 non-null int64
2   NAME_CONTRACT_TYPE  307511 non-null object
3   NAME_FAMILY_STATUS  307511 non-null object
4   CNT_CHILDREN        307511 non-null int64
5   NAME_HOUSING_TYPE   307511 non-null object
6   NAME_EDUCATION_TYPE 307511 non-null object
7   NAME_INCOME_TYPE    307511 non-null object
8   AMT_INCOME_TOTAL    307511 non-null float64
9   AMT_ANNUITY         307499 non-null float64
10  AMT_CREDIT           307511 non-null float64
dtypes: float64(3), int64(3), object(5)
memory usage: 25.8+ MB
```

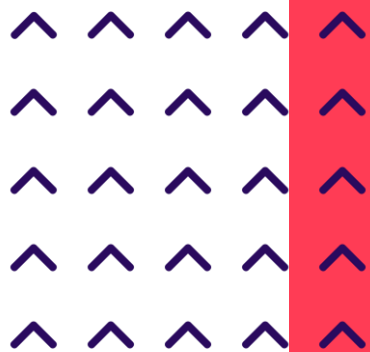
```
cats_dtypes=['object']
cats_df=df_feature.select_dtypes(include=cats_dtypes).columns
cats_df

Index(['NAME_CONTRACT_TYPE', 'NAME_FAMILY_STATUS', 'NAME_HOUSING_TYPE',
      'NAME_EDUCATION_TYPE', 'NAME_INCOME_TYPE'],
      dtype='object')

df_feature[cats_df].describe()
```

	NAME_CONTRACT_TYPE	NAME_FAMILY_STATUS	NAME_HOUSING_TYPE	NAME_EDUCATION_TYPE	NAME_INCOME_TYPE
count	307511	307511	307511	307511	307511
unique	2	6	6	5	8
top	Cash loans	Married	House / apartment	Secondary / secondary special	Working
freq	278232	196432	272868	218391	158774

- Insight yang dapat diambil:
- **Total data customers** adalah sebanyak **307511 baris**,
 - **Data terdiri dari 10 kolom**, jumlah feature data customer yang dimiliki
 - Tipe Data: **Numeric** (int64/integer), **Categorical** (object/string), dan **Floating** (float64)
 - **Tidak ada Null Values** (data yang kosong)
 - Tidak ada data yang anomaly
 - Tidak ada perbedaan signifikan antara mean dan median sehingga dapat disimpulkan bahwa data **terdistribusi normal**.



Feature Engineering ●●●



2

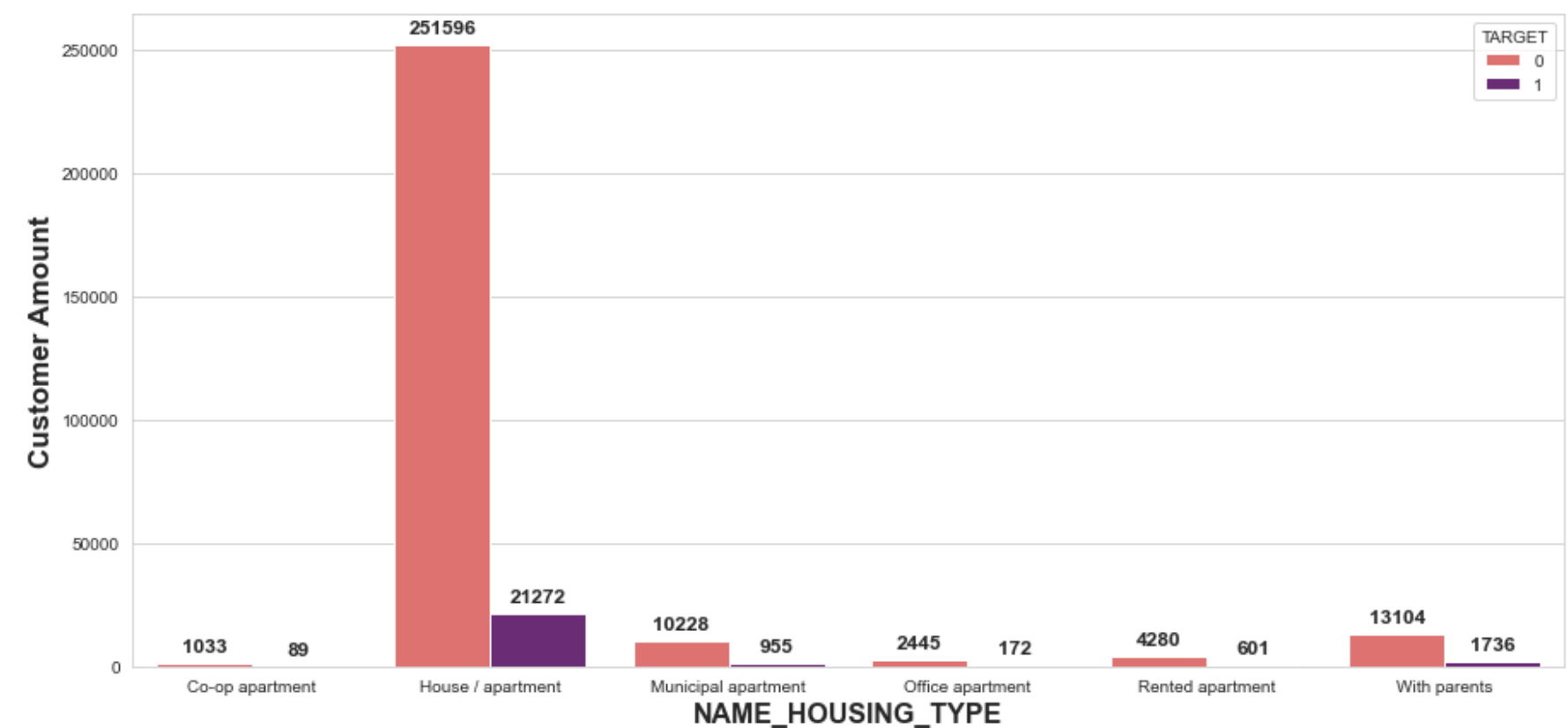
```
df_drop2['INCOME'] = np.where(df_drop2['AMT_INCOME_TOTAL'] > 500000, '500000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 450000, '450000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 400000, '400000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 350000, '350000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 300000, '300000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 250000, '250000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 200000, '200000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 150000, '150000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 100000, '100000',  
                             np.where(df_drop2['AMT_INCOME_TOTAL'] > 50000, '50000', '10000'))))))))
```

3

```
df_drop3['TARGET'] = df_drop3['TARGET'].astype('str')  
df_drop3['SK_ID_CURR'] = df_drop3['SK_ID_CURR'].astype('str')
```

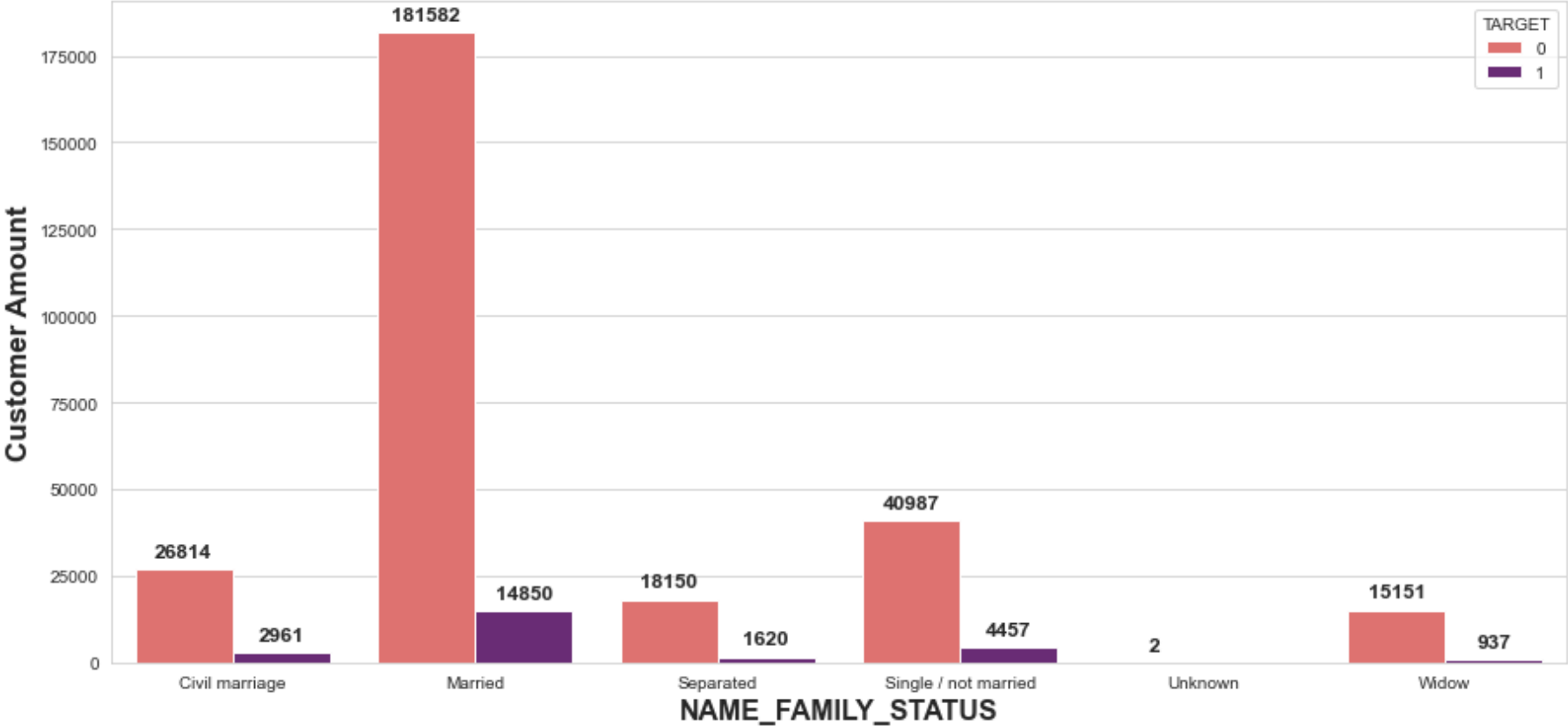
1. Berdasarkan *core heatmap* dari dataset, terlihat bahwa ada *feature-feature* yang sifatnya *redundant* sehingga perlu di *drop*, yaitu: **'AMT_CREDIT'** dan **'AMT_ANNUITY'**
2. Feature **'AMT_INCOME_TOTAL'** memiliki distribusi yang terlalu luas dan outliers yang terlalu banyak sehingga disederhanakan dengan klasifikasi berdasarkan besaran terdekatnya
3. Feature **'TARGET'** dan **'SK_ID_CURR'** diubah menjadi string karena hanya antara 0 dan 1

Data Visualization ●●●



BERDASARKAN ASET

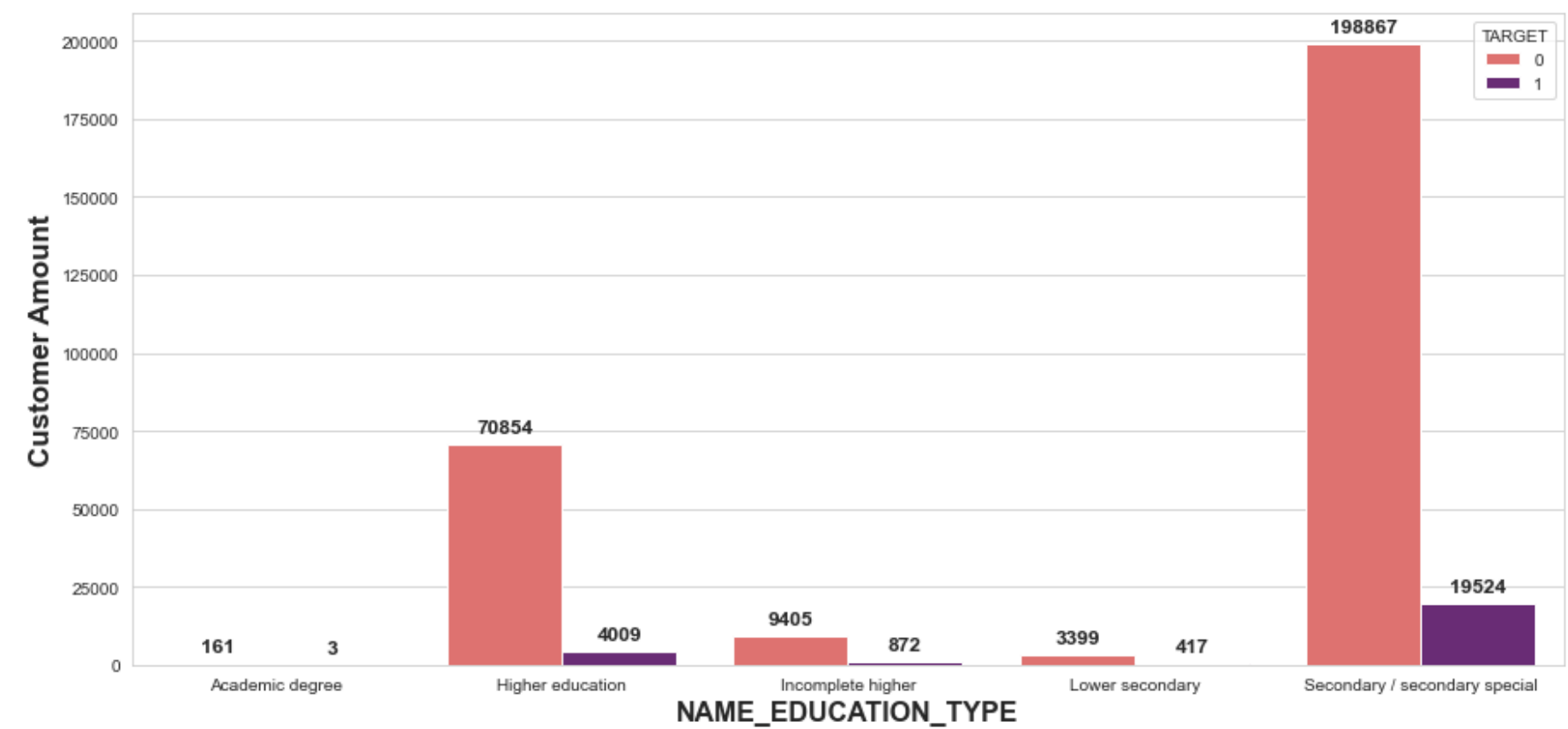
Mayoritas *Customers with Difficulties* memiliki rumah atau apartemen sebagai tempat tinggal. Hal ini diprediksi terjadi karena kecenderungan adanya cicilan lain untuk melunasi pembelian rumah atau apartemen.



BERDASARKAN STATUS

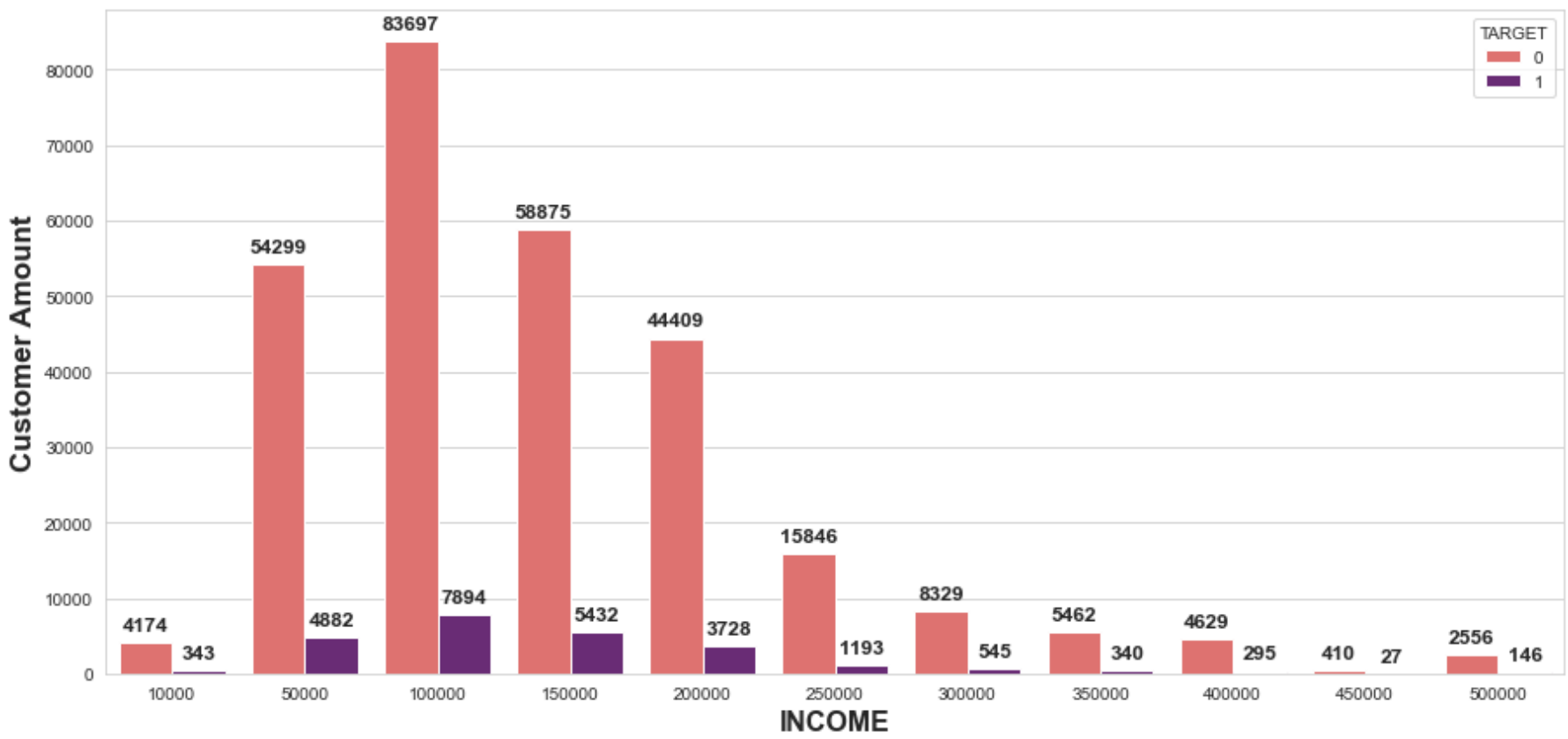
Mayoritas *Customers with Difficulties* berstatus sudah menikah. Hal ini diprediksi terjadi karena kebutuhan biaya hidup yang lebih tinggi untuk individu yang sudah berkeluarga.

Data Visualization ●●●



BERDASARKAN PENDIDIKAN

Mayoritas *Customers with Difficulties* berpendidikan menengah (SMA)



BERDASARKAN EKONOMI

Mayoritas *Customers with Difficulties* memiliki penghasilan tahunan antara 50,000 hingga 200,000

Predictive Modeling ●●●

1

```
df_mod = df.copy()

a = df_mod.drop(columns=['TARGET'])
b = a.drop(columns=['NAME_CONTRACT_TYPE'])
c = b.drop(columns=['NAME_FAMILY_STATUS'])
d = c.drop(columns=['NAME_HOUSING_TYPE'])
e = d.drop(columns=['NAME_EDUCATION_TYPE'])
x = e.drop(columns=['NAME_INCOME_TYPE'])
y = df_mod['TARGET'] # target / Label

#Splitting the data into Train and Test
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3, random_state = 42)
```

```
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import roc_auc_score
from sklearn.metrics import f1_score
from sklearn.metrics import roc_auc_score
```

```
print("Accuracy (Test Set): %.2f" % accuracy_score(y_test, y_pred))
print("Accuracy (Train Set): %.2f" % accuracy_score(y_train, y_pred_train))
```

```
Accuracy (Test Set): 0.92
Accuracy (Train Set): 0.92
```

Jika dilihat dari metrics evaluasi, *Accuracy* yang didapatkan adalah 0.92. Sedangkan *Precision* dan *Recall* tidak dapat dinilai karena feature memiliki lebih dari 2 probabilitas.

2

```
print('auc_score pada y_pred: ', roc_auc_score(y_test, y_pred))
print('auc_score pada y_pred_proba: ',roc_auc_score(y_test, y_pred_proba[:, 1]))
```

```
auc_score pada y_pred: 0.5
auc_score pada y_pred_proba: 0.515121900612166
```

```
print('auc_score pada y_pred_tun: ', roc_auc_score(y_test, y_pred_tun))
print('auc_score pada y_pred_tun_proba: ',roc_auc_score(y_test, y_pred_tun_proba[:, 1]))
```

```
auc_score pada y_pred_tun: 0.5
auc_score pada y_pred_tun_proba: 0.515121900612166
```

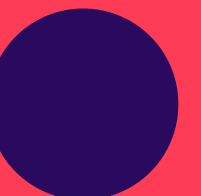
Setelah dilakukan Hyperparameter Tuning terlihat adanya peningkatan pada AUC score yang dihasilkan sebanyak 0.015 menjadi 0.515. Oleh karena itu dapat disimpulkan pemodelan pada dataset kasus ini memiliki performa yang kurang maksimal jika menerapkan algoritma Logistic Regression. Walaupun demikian, data ini masih dapat dipakai untuk membuat rekomendasi bisnis.

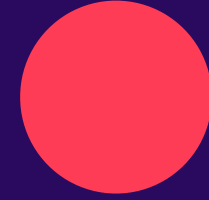


Business Recommendation

Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa untuk dapat mengurangi jumlah Customers with Difficulties, hal yang perlu dilakukan oleh Home Credit Indonesia adalah:

1. Memperketat seleksi persetujuan pengajuan kredit dengan menandai customers yang masih memiliki hutang / cicilan properti maupun barang diluar Home Credit Indonesia
2. Mengutamakan customers yang belum berkeluarga, berpendidikan tinggi (sarjana), dan berpenghasilan dalam range 100,000 – 250,000
3. Menyediakan pinjaman dengan rentang waktu angsuran lebih lama agar angka cicilan bulanan menurun.





**Click on the link below
to access the code**

[Source Code \(Github Repository\)](#)



Pinandito Wahyu

Entry-level Data Scientist

Email: pinanditow@gmail.com

Phone: 0818466226

Linkedin: [Pinandito Wahyu](#)

Github: [@pinanditow](#)