

Output :

1. <https://colab.research.google.com/drive/1chM96EfpXtRBLPgsns-gcKtQ0DjJQvCD>
2. https://colab.research.google.com/drive/1ki0cHl6qm01JWFK0_sUftZ_e372jSsmc
3. <https://colab.research.google.com/drive/1phfcEJYvH6R4uV-dCEvK9sI0sm1QkiQw>
4. <https://lookerstudio.google.com/reporting/b9b058f8-df07-4f8f-ac71-63d985fb2455/page/XDzTD>

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok : 30
Nama Mentor : Erwin Fernanda

Data Analytics Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



KELOMPOK 30



Dinny Meilinda Sari
Universitas Komputer
Indonesia



Christ Jordan Baeha
Universitas Katolik Santo
Thomas



Zulfa Nabilah Nurvitasari
Universitas Brawijaya



Fiqih Imanul Haq
Universitas Muhammadiyah
Purwokerto



Sania Salsabila Agustin
Universitas Negeri Semarang

Data Analysis

Home Credit Default Risk

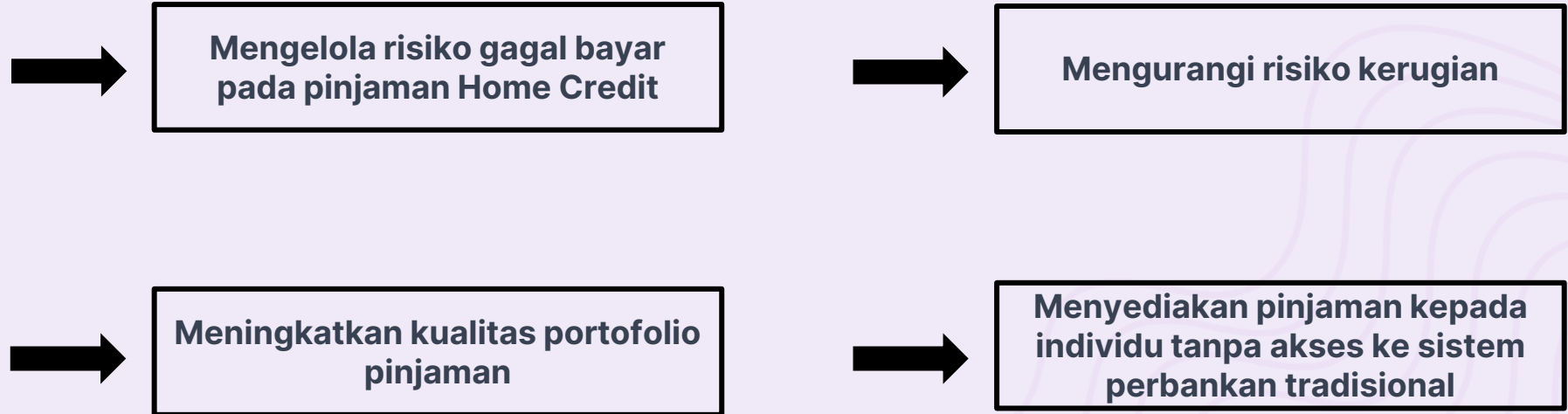
- 1. Latar Belakang**
- 2. Explorasi Data dan Visualisasi**
- 3. Modelling**
- 4. Kesimpulan**

Latar Belakang

Karena adanya pemohon yang tidak mengembalikan peminjaman sebanyak **8,1%** maka dibutuhkan analisis bagi penyedia Pinjaman untuk mengetahui pemohon Pinjaman yang kemungkinan besar akan membayar kembali pinjaman tersebut. Dengan cara ini perusahaan dapat menghindari kerugian dan mendatangkan keuntungan besar.

Exploratory Data Analysis

Bussines Understanding



Data Collection

Sumber Data : Home Credit Default Risk

<https://www.kaggle.com/competitions/home-credit-default-risk/data>

Dataset	Jumlah Record	Jumlah Kolom	Missing Values
application_train.csv	307.511	122	67 kolom, > 50%
bureau.csv	1.716.428	17	7 kolom, > 50%
previous_application.csv	1.670.214	37	16 kolom, >55%

Data Preparation

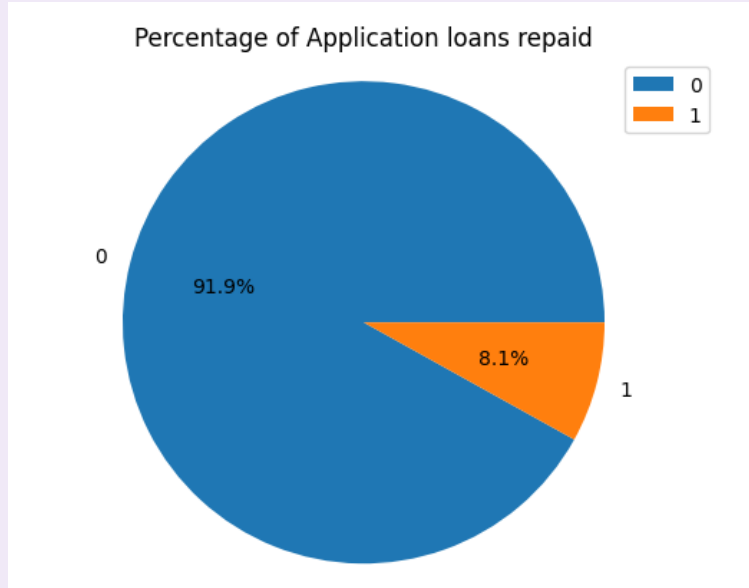
- Melakukan encoding dan mendapat 4 kolom yang berhasil di encoding.
- Menambahkan 4 kolom baru sebagai Domain Knowledge Features yaitu :

Nama Kolom	Keterangan
CREDIT_INCOME_PERCENT	persentase jumlah kredit relatif terhadap pendapatan klien. (<code>'AMT_CREDIT' / 'AMT_INCOME_TOTAL'</code>)
ANNUITY_INCOME_PERCENT	persentase anuitas pinjaman relatif terhadap pendapatan klien. (<code>'AMT_ANNUITY' / 'AMT_INCOME_TOTAL'</code>)
CREDIT_TERM	lamanya pembayaran dalam beberapa bulan (karena anuitas adalah jumlah bulanan yang jatuh tempo). (<code>'AMT_ANNUITY' / 'AMT_CREDIT'</code>)
DAYS_EMPLOYED_PERCENT	persentase hari bekerja relatif terhadap usia klien. (<code>'DAYS_EMPLOYED' / 'DAYS_BIRTH'</code>)

Data Preparation

- **Mengubah seluruh isi Missing Value/Nilai yang hilang menjadi Median kolom tersebut.**
- **Membuat dataset yang telah diencoding, dan ditambahkan Domain Knowledge Features dan diisi nilai Missing Value nya menjadi cleandata.csv**

Persentase pinjaman yang dilunasi

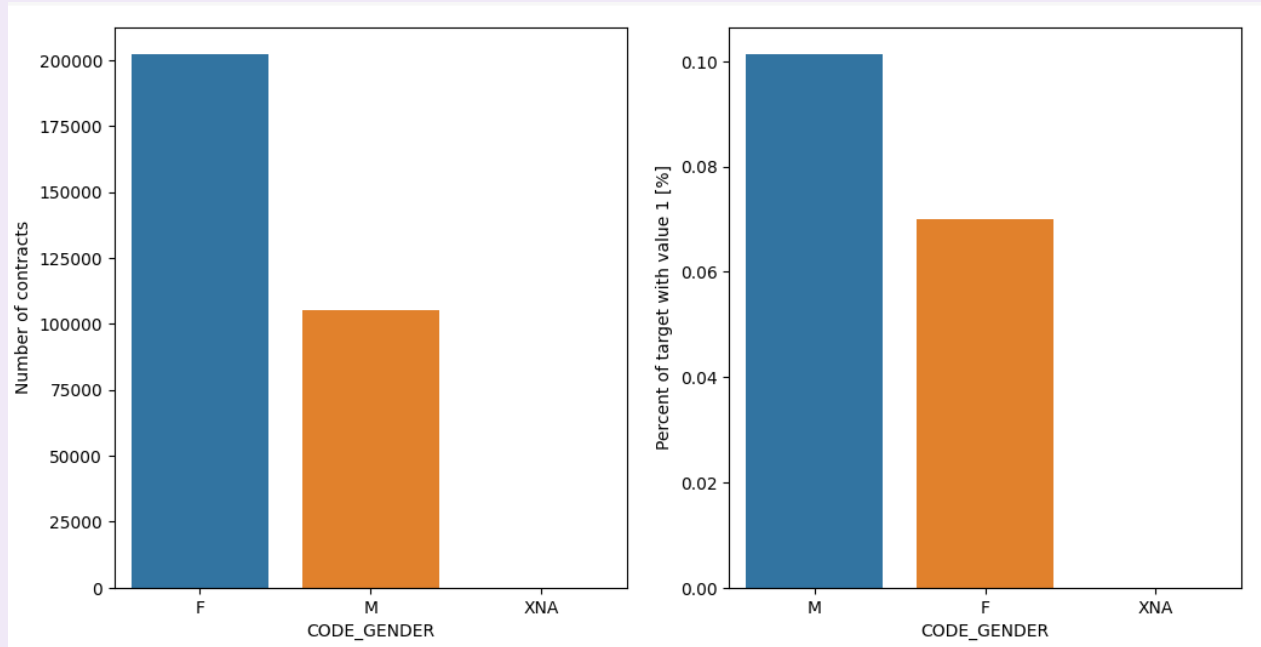


Ket. :

0 = pinjaman dilunasi

1 = pinjaman tidak dilunasi

Persentase pinjaman (berdasarkan jenis kelamin klien)



Tabel Kontingensi

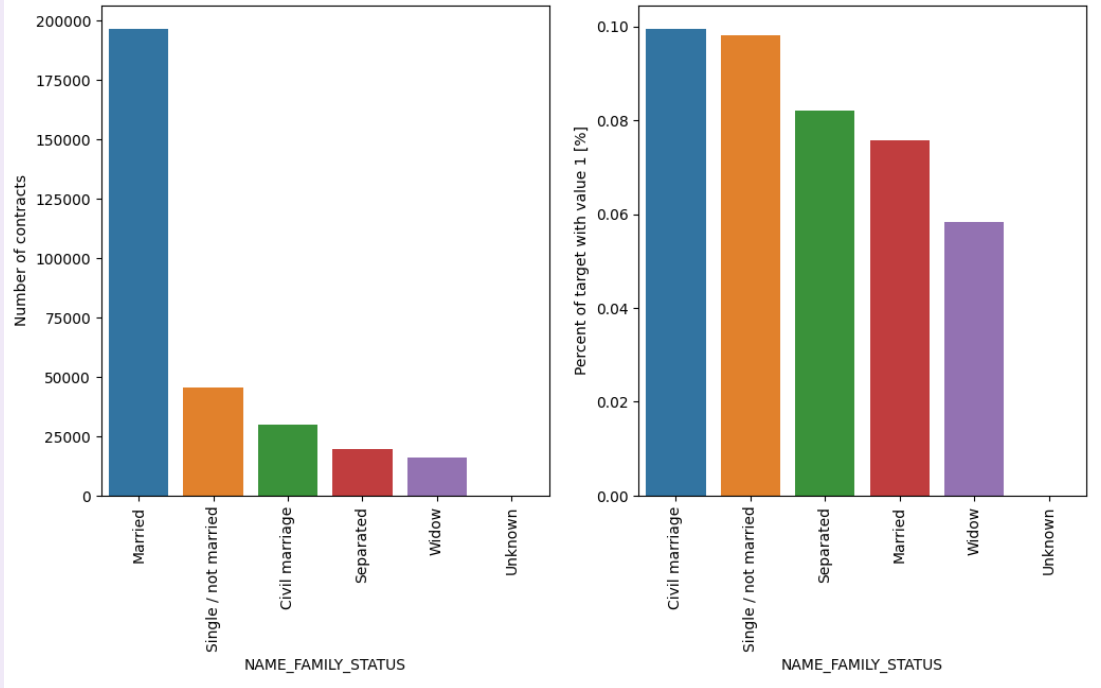
TARGET	0	1
CODE_GENDER		
F	188278	14170
M	94404	10655
XNA	4	0

Ket. :

0 = pinjaman dilunasi

1 = pinjaman tidak dilunasi

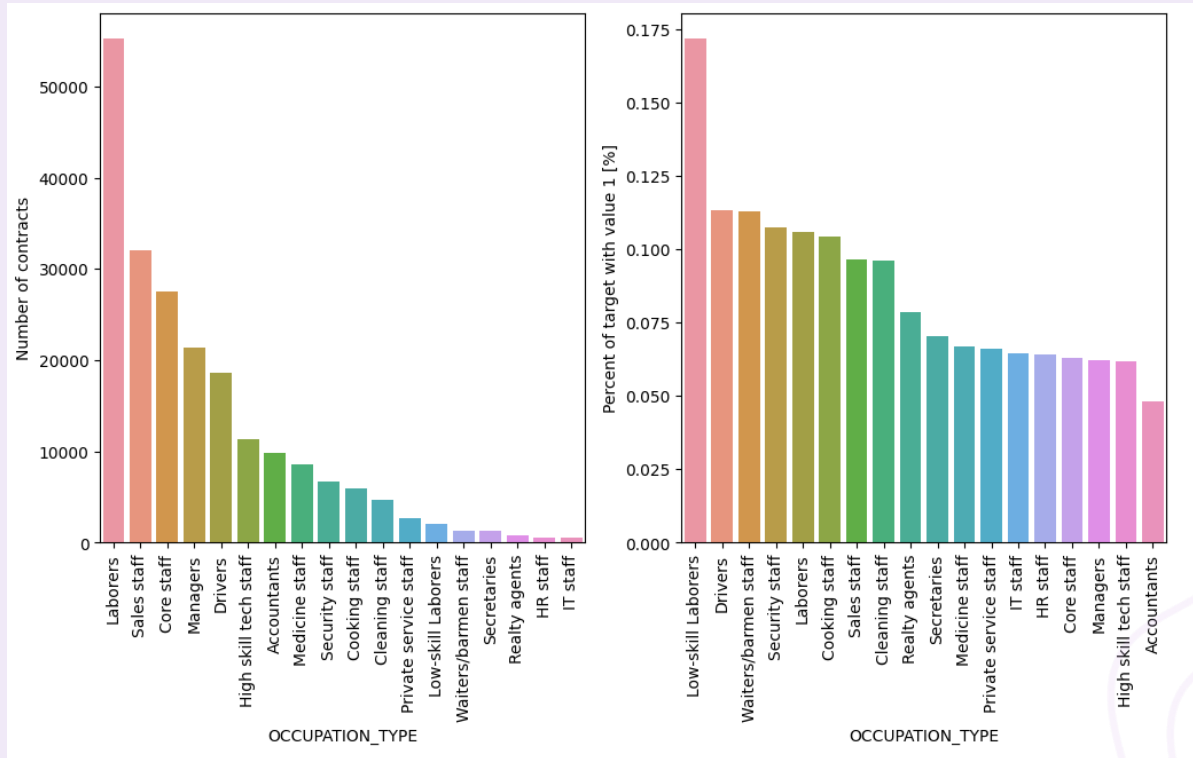
Persentase pinjaman (berdasarkan status keluarga)



Kategori klien tertinggi yang melakukan peminjaman adalah sudah menikah, diikuti “Single/belum menikah” dan “civil marriage”.

Dalam hal persentase tidak dilunasi pinjaman, “civil marriage” memiliki persentase tidak dilunasi tertinggi (10%), dengan “Janda” yang terendah (pengecualian “unknown”).

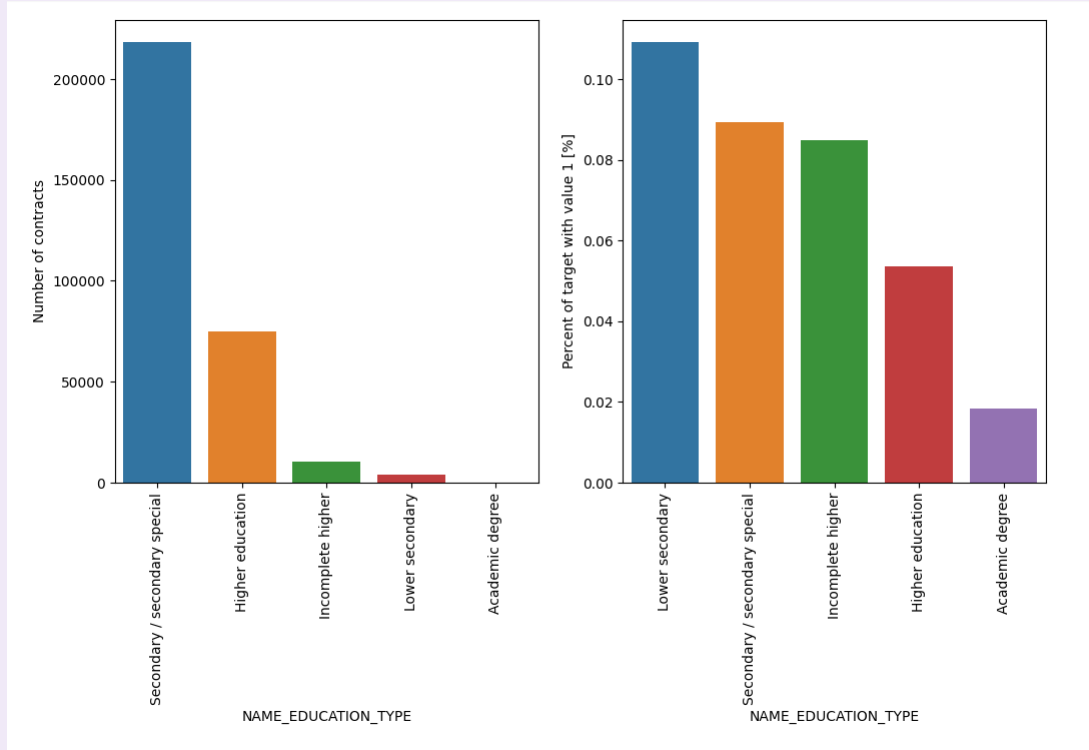
Persentase pinjaman (berdasarkan pekerjaan)



Kategori klien tertinggi pinjaman diambil oleh **"Buruh"**, diikuti oleh **"Staf Penjualan"**. **"Staf IT"** mengambil jumlah pinjaman terendah.

Kategori dengan persentase pinjaman tidak dilunasi tertinggi adalah **"Buruh Keterampilan Rendah"** (di atas 17%), diikuti oleh **"Pengemudi"** dan **"Staf Pelayan/Barmen"**, **"Staf Keamanan"**, **"Buruh"** dan **"Staf Memasak"**.

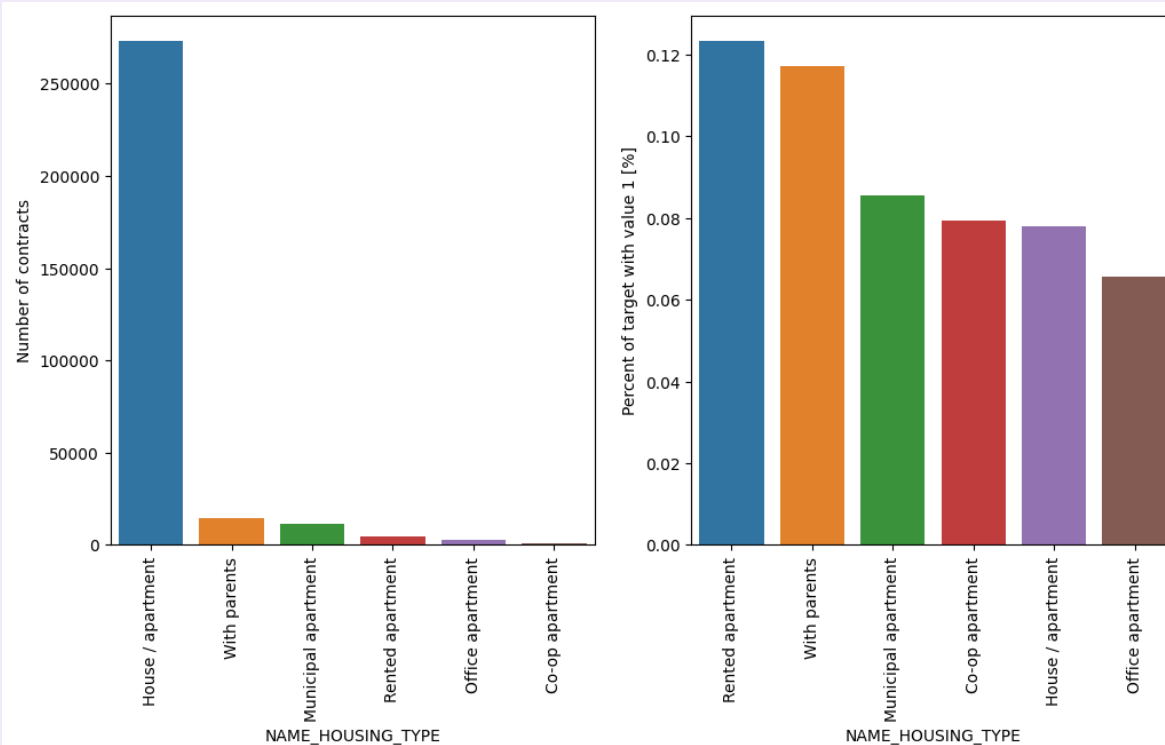
Persentase pinjaman (berdasarkan pendidikan)



Mayoritas klien berpendidikan menengah/menengah khusus, diikuti oleh klien berpendidikan tinggi. Hanya sedikit sekali yang bergelar sarjana.

Kategori menengah bawah, meskipun jarang, memiliki tingkat tidak mengembalikan pinjaman terbesar (11%). Orang-orang dengan gelar Akademik memiliki tingkat tidak membayar kurang dari 2%.

Persentase pinjaman (berdasarkan jenis tempat tinggal)

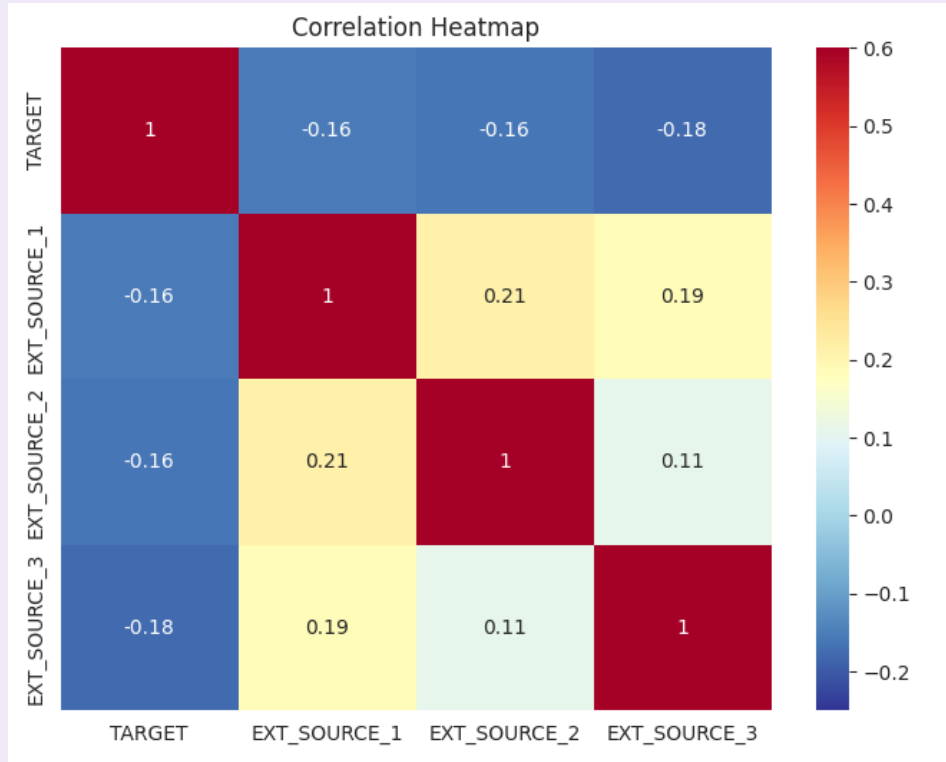


Lebih dari 250.000 pemohon kredit mendaftarkan rumah mereka sebagai Rumah/apartemen.

Kategori berikut memiliki jumlah klien yang sangat kecil (bersama orang tua, apartemen Kota).

Dari kategori ini, *Apartemen sewaan* dan *bersama orang tua* memiliki tingkat tidak dapat dilunasi lebih dari 10%.

Correlation Heatmap



Ketiga kolom EXT_SOURCE memiliki korelasi negatif dengan target, yang menunjukkan bahwa seiring dengan peningkatan nilai EXT_SOURCE, klien cenderung membayar kembali pinjamannya.

Most Positive Values

No.	Nama Kolom	Correlation Value
1.	TARGET	1.000000
2.	DAYS_EMPLOYED	0.074958
3.	REGION_RATING_CLIENT_W_CITY	0.060893
4.	REGION_RATING_CLIENT	0.058899
5.	DAYS_LAST_PHONE_CHANGE	0.055218
6.	DAYS_ID_PUBLISH	0.051457
7.	REG_CITY_NOT_WORK_CITY	0.050994
8.	FLAG_EMP_PHONE	0.045982
9.	REG_CITY_NOT_LIVE_CITY	0.044395
10.	FLAG_DOCUMENT_3	0.044346

Most Negative Values

No.	Nama Kolom	Correlation Value
1.	FLOORSMAX_MEDI	-0.043768
2.	FLOORSMAX_MODE	-0.043226
3.	AMT_GOODS_PRICE	-0.039645
4.	REGION_POPULATION_RELATIVE	-0.037227
5.	ELEVATORS_AVG	-0.034199
6.	ELEVATORS_MEDI	-0.033863
7.	FLOORSMIN_AVG	-0.033614
8.	FLOORSMIN_MEDI	-0.033394
9.	LIVINGAREA_MEDI	-0.032739
10.	LIVINGAREA_AVG	-0.032997

Modelling

Supervised Learning

BAGAN

Logistic Regression

Random Forest

Balancing Class

SMOTETomek

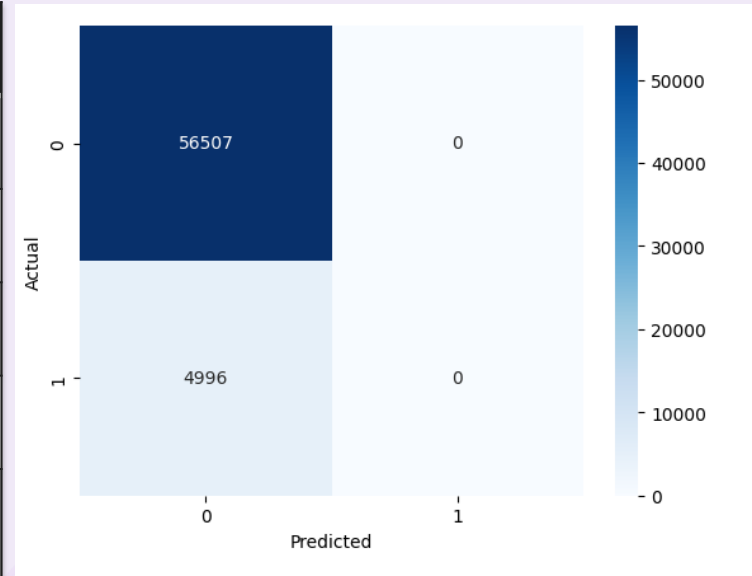
Random Over
Sampler

Logistic Regression

Classification Report

	precision	recall	f1-score	support
0	0.92	1.00	0.96	56507
1	0.00	0.00	0.00	4996
accuracy			0.92	61503
macro avg	0.46	0.50	0.48	61503
weighted avg	0.84	0.92	0.88	61503

Confusion Matrix



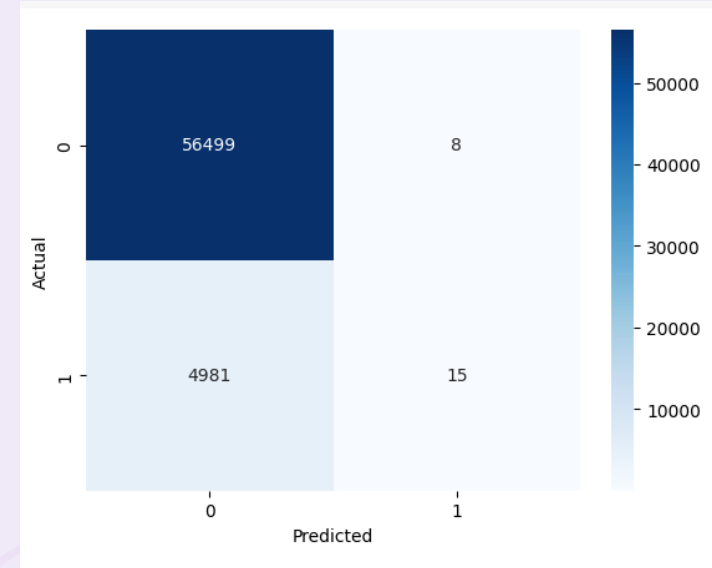
Nilai ROC
AUC
sebesar
0.5.

Random Forest

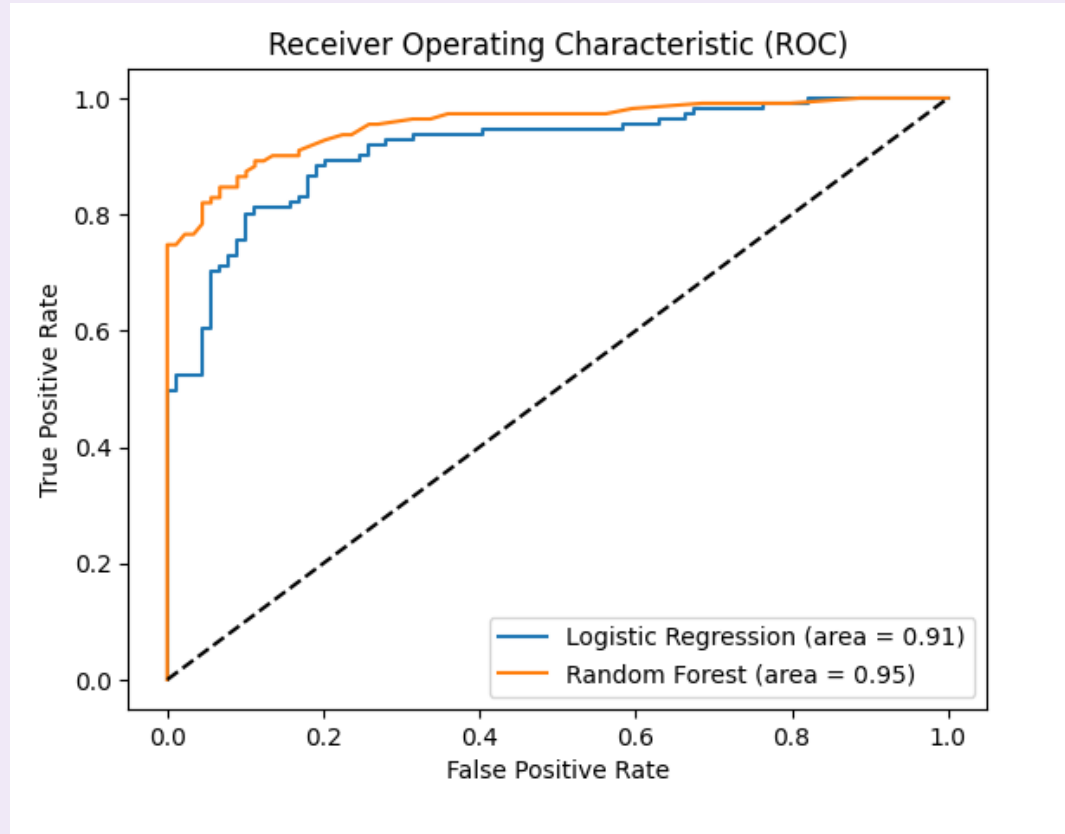
Classification Report

	precision	recall	f1-score	support
0	0.92	1.00	0.96	56507
1	0.65	0.00	0.01	4996
accuracy			0.92	61503
macro avg	0.79	0.50	0.48	61503
weighted avg	0.90	0.92	0.88	61503

Confusion Matrix



Nilai ROC
AUC
sebesar
0.5.

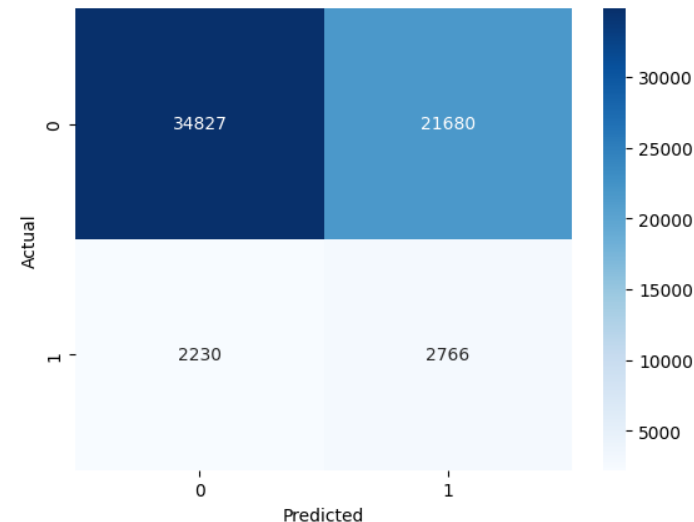


Balancing class - using SMOTETomek (Logistic Regression)

Classification Report

	precision	recall	f1-score	support
0	0.79	0.84	0.82	89
1	0.87	0.82	0.84	111
accuracy			0.83	200
macro avg	0.83	0.83	0.83	200
weighted avg	0.83	0.83	0.83	200

Confusion Matrix



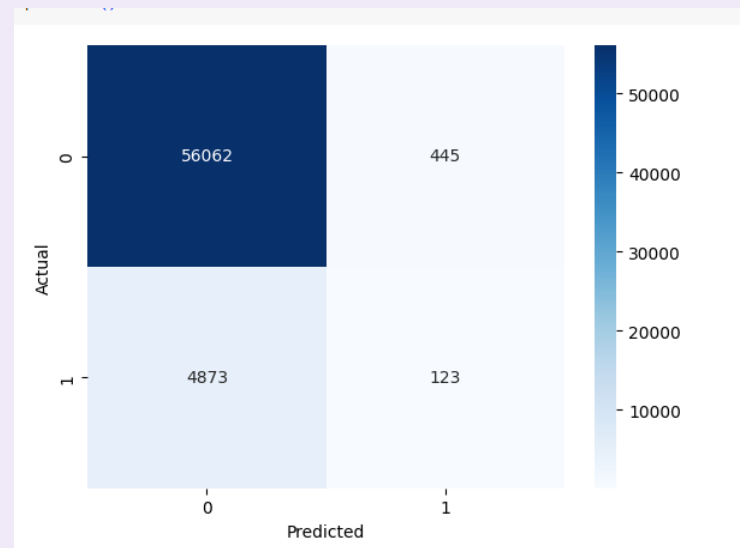
Nilai ROC
AUC
sebesar
0.83.

Balancing class - using SMOTETomek (Random Forest)

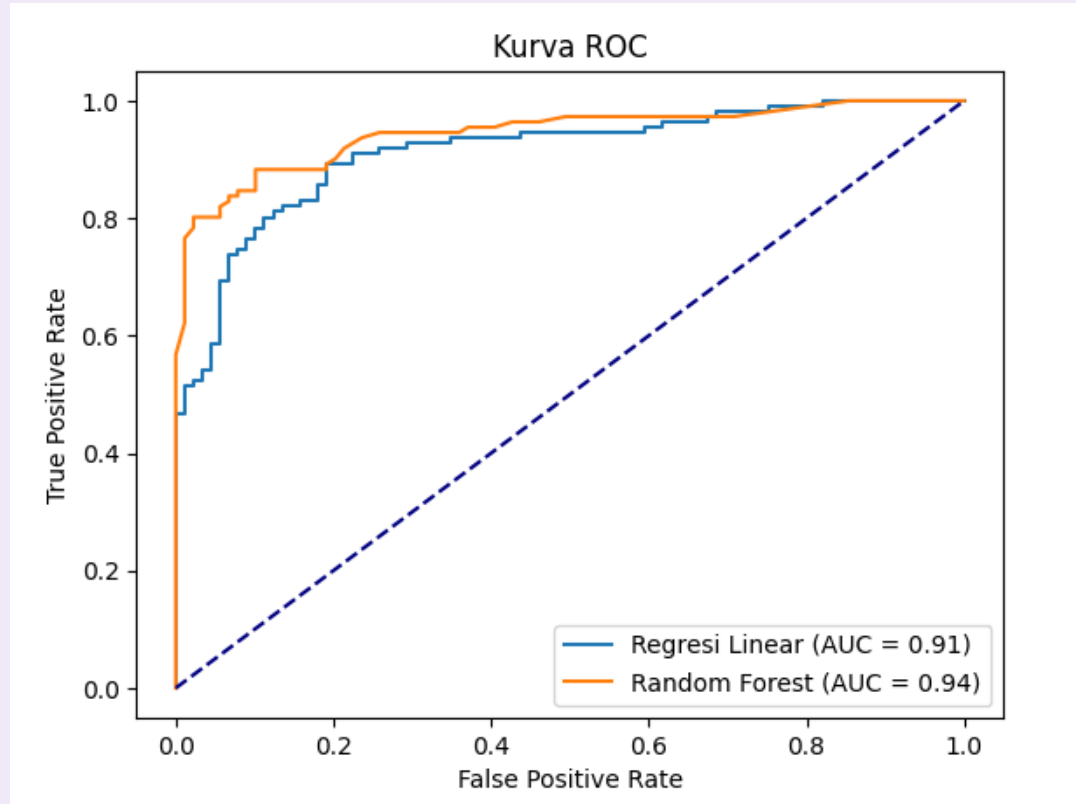
Classification Report

	precision	recall	f1-score	support
0	0.84	0.90	0.87	89
1	0.91	0.86	0.89	111
accuracy			0.88	200
macro avg	0.88	0.88	0.88	200
weighted avg	0.88	0.88	0.88	200

Confusion Matrix



Nilai ROC
AUC
sebesar
0.88.

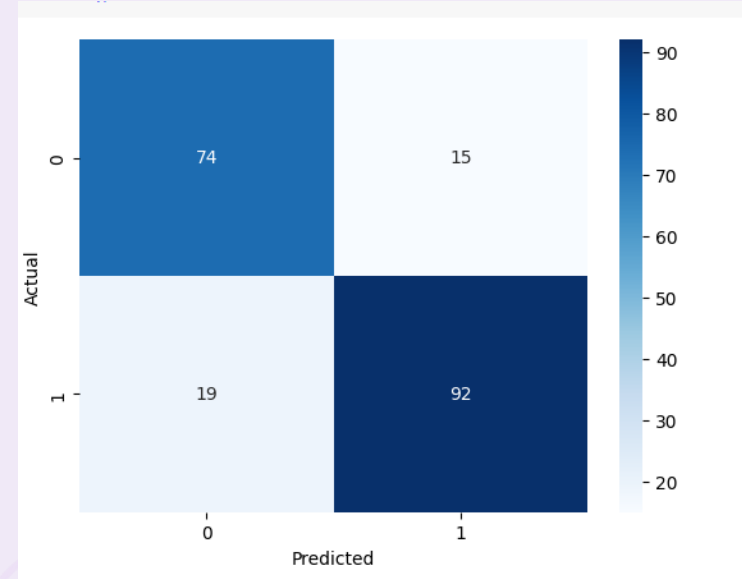


Balancing class - using Random Over Sampler (Logistic Regression)

Classification Report

	precision	recall	f1-score	support
0	0.80	0.83	0.81	89
1	0.86	0.83	0.84	111
accuracy			0.83	200
macro avg	0.83	0.83	0.83	200
weighted avg	0.83	0.83	0.83	200

Confusion Matrix



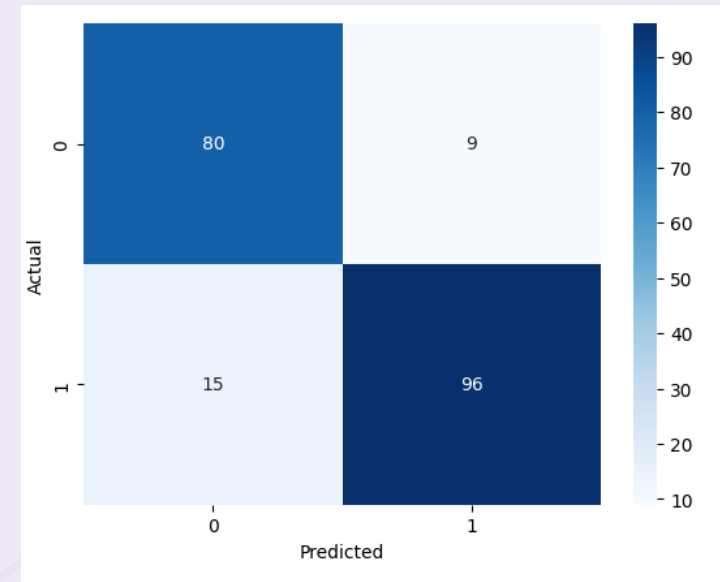
Nilai ROC
AUC
sebesar
0.83.

Balancing class - using Random Over Sampler (Random Forest)

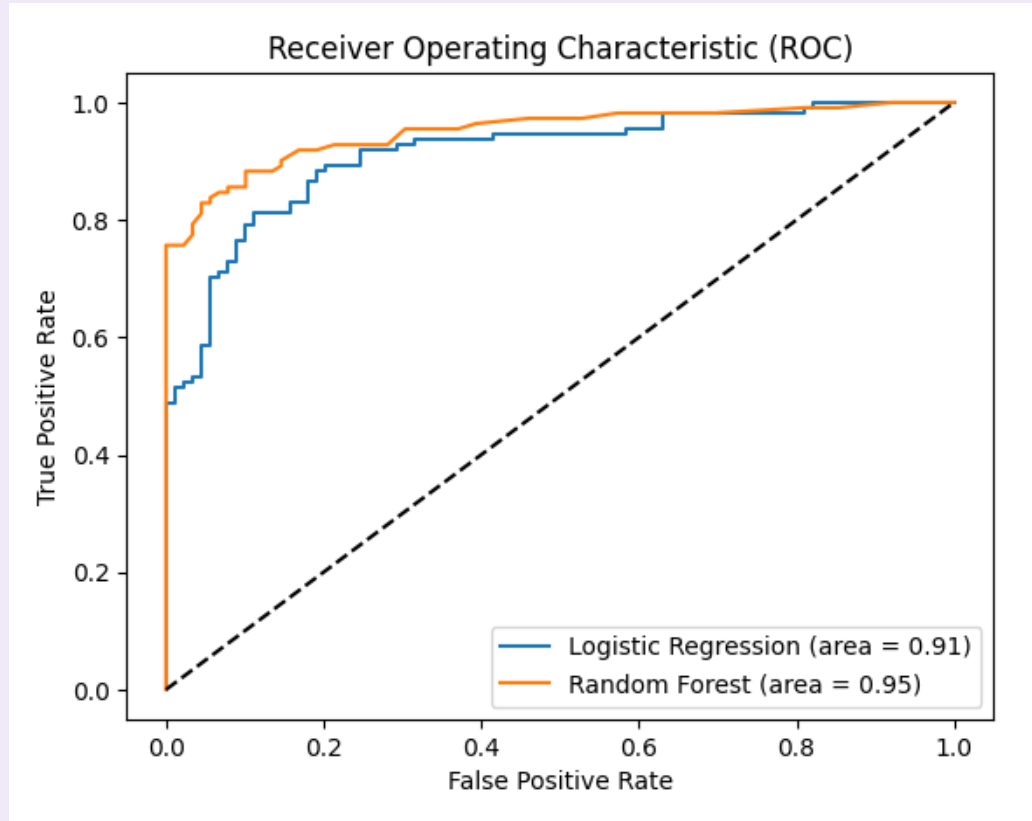
Classification Report

	precision	recall	f1-score	support
0	0.84	0.91	0.87	89
1	0.82	0.86	0.89	111
accuracy			0.88	200
macro avg	0.88	0.88	0.88	200
weighted avg	0.88	0.88	0.88	200

Confusion Matrix



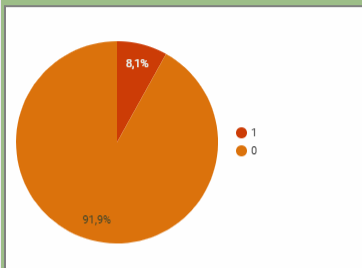
Nilai ROC
AUC
sebesar
0.88.



Dashboard

Home Credit Default Risk Dashboard

Client by Target



Client with Payment Difficult

24.825

Client with Other Cases

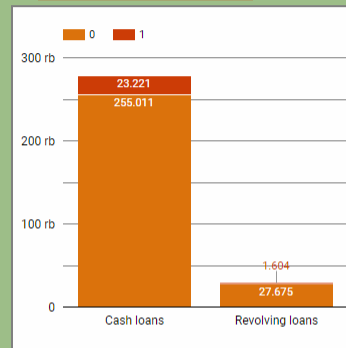
282.686

Total Target by Education Type

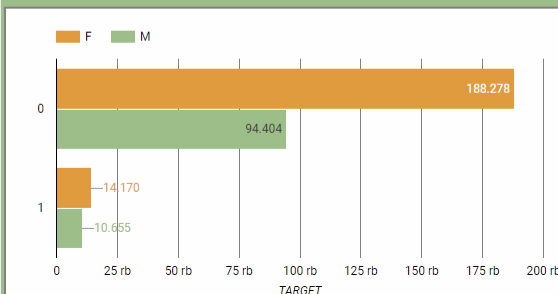
	NAME_EDUCATION_TYPE	TARGET	TARGET ▾
1.	Secondary / secondary special	0	198.867
2.	Higher education	0	70.854
3.	Secondary / secondary special	1	19.524
4.	Incomplete higher	0	9.405
5.	Higher education	1	4.009
6.	Lower secondary	0	3.399
7.	Incomplete higher	1	872
8.	Lower secondary	1	417
9.	Academic degree	0	161
10.	Academic degree	1	3

1 - 10 / 10 < >

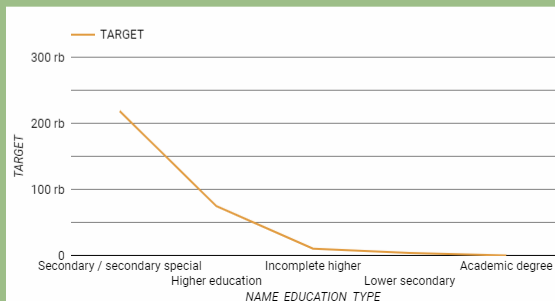
Target by Contract Type



Target by Gender



Target by Education Type



**Terima
kasih!**
Ada pertanyaan?

zenius



**Kampus
Merdeka**
INDONESIA JAYA

