zenius | Kampus Merdeka INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 14
Nama Mentor: Erwin Fernanda

Accelerated Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka

Final Project

Link Tugas :
https://colab.research.google.com/drive/10E3SXkI6lwSf-X24cgZhuc-Oh0HhIFkD?usp=sharing

Kelompok 14

Ratu Sondang Elishabet S

Universitas Diponegoro.

Reynaldi Mulyawan

UPN Yogyakarta.

Ade Amalia

Universitas Alma Ata

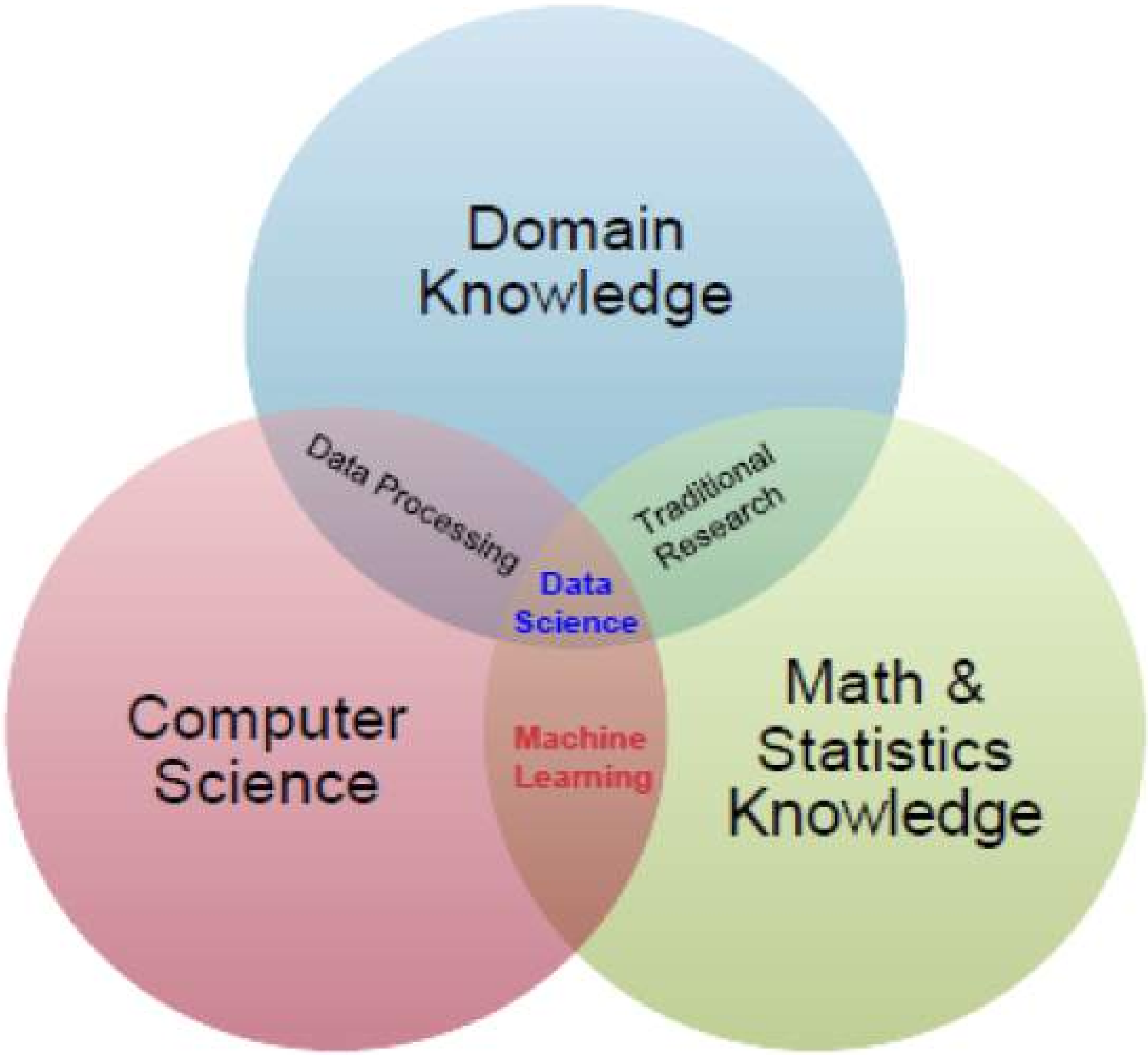
Pendahuluan

“**Data Science** is an **interdisciplinary** field about processes and systems to **extract knowledge** or insights **from data** in various forms”

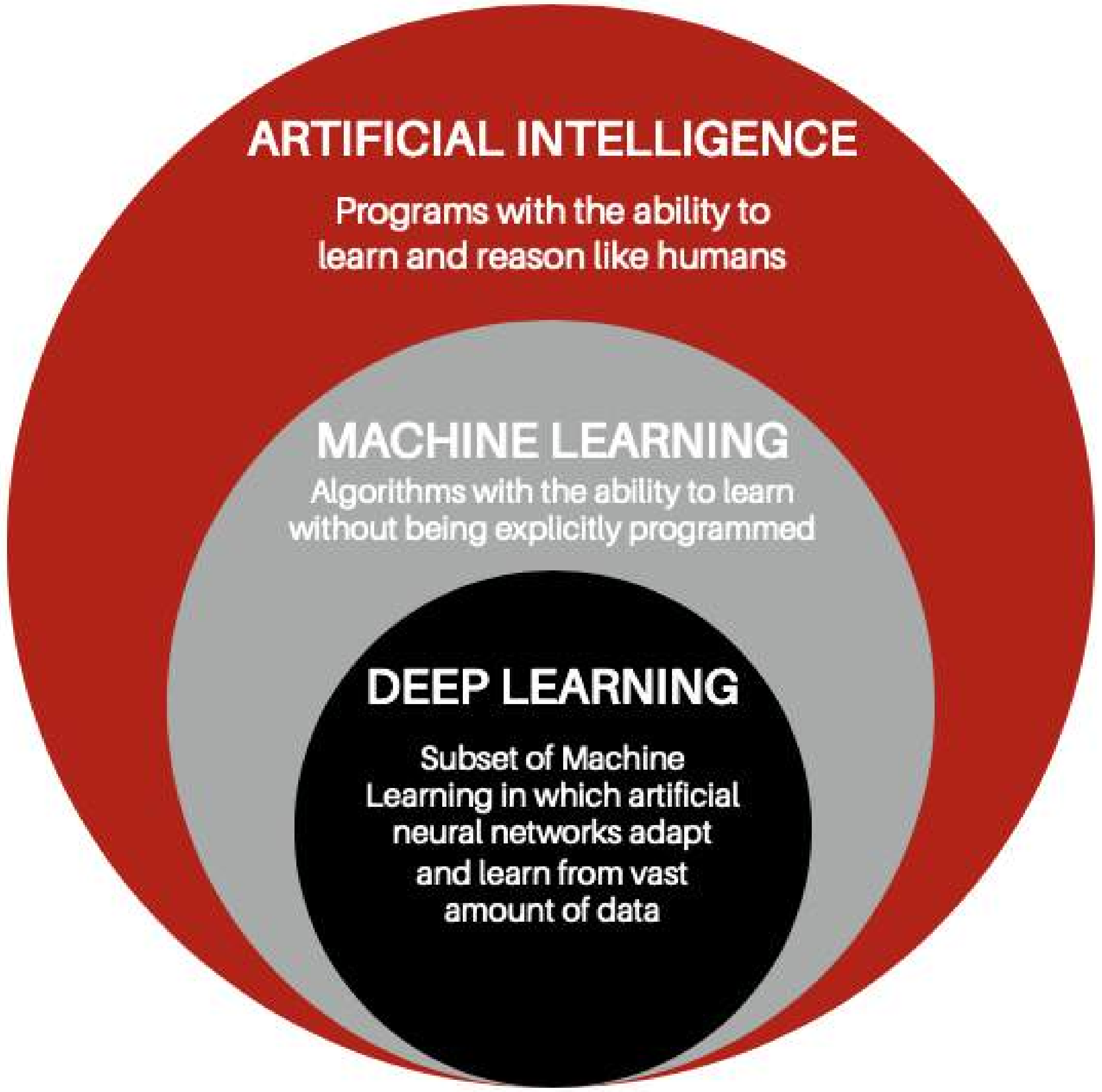
– *Wikipedia*

Therefore, executing data science projects require three key skills:

1. Programming skills,
2. Math & Statistics,
3. Business or subject matter expertise for a given area of scope



Source: Drew Conway, IA Ventures


ARTIFICIAL INTELLIGENCE
Programs with the ability to learn and reason like humans
MACHINE LEARNING
Algorithms with the ability to learn without being explicitly programmed
DEEP LEARNING
Subset of Machine Learning in which artificial neural networks adapt and learn from vast amount of data

zenius × Kampus Merdeka INDONESIA JAYA

Business Understanding

- Pada Final Project ini, kita akan melakukan analisa dan juga prediksi dari dataset Homecredit, dari situs Kaggle, untuk melakukan analisaapakah nasabah yang akan diberi pinjaman mempunya kemungkinan besar kreditnya akan lancar atau tidak.
- **Homecredit** merupakan perusahaan yang menyediakan layanan peminjaman untuk keperluan kredit perlengkapan rumah, peralatan elektronik dll.
- **Problem statementnya** adalah **membuat model yang memprediksi** seberapa tinggi kemampuan konsumen untuk membayar angsuran.

Data Understanding

Ada 7 sumber data yang berbeda:

- **"Application_train/application_test"**: Dataset ini terdiri dari data pelatihan dan pengujian utama dengan informasi tentang setiap aplikasi pinjaman di Home Credit. Setiap pinjaman ditandai dengan fitur SK_ID_CURR. Data aplikasi pelatihan dilengkapi dengan TARGET.

- **bureau**: Dataset yang terdiri dari data mengenai kredit klien sebelumnya dari lembaga keuangan lain. Setiap kredit sebelumnya memiliki barisnya sendiri di biro, tetapi satu pinjaman dalam data aplikasi dapat memiliki beberapa kredit sebelumnya.

- **bureau_balance**: Dataset yang terdiri dari data mengenai data bulanan tentang kredit sebelumnya di biro. Setiap baris adalah satu bulan dari kredit sebelumnya, dan satu kredit sebelumnya dapat memiliki beberapa baris, satu untuk setiap bulan dari panjang kredit.

Data Understanding

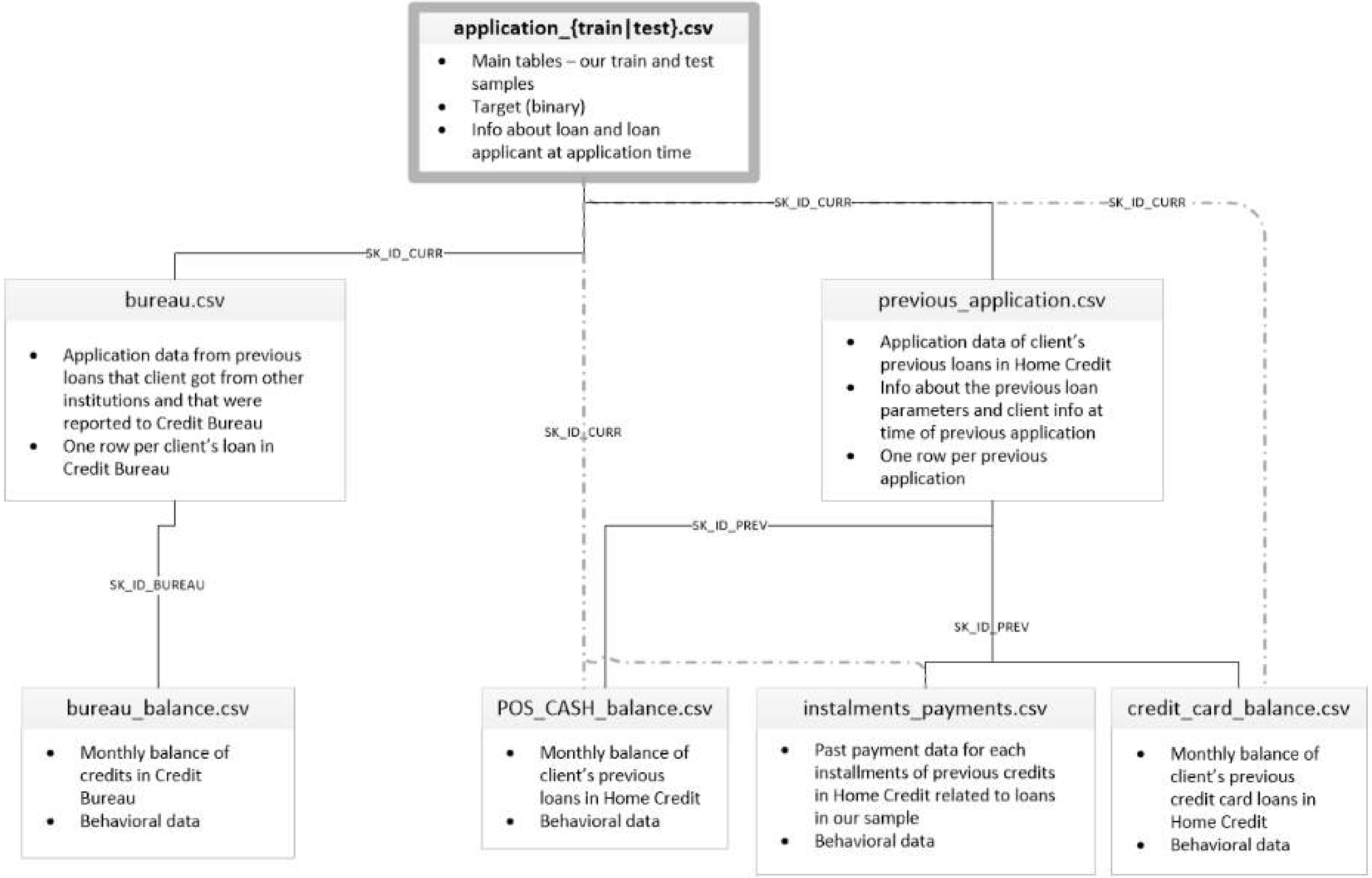
- **previous_application**: Dataset yang terdiri dari data mengenai aplikasi sebelumnya untuk pinjaman di Home Credit klien yang memiliki pinjaman dalam data aplikasi. Setiap aplikasi sebelumnya memiliki satu baris dan ditandai dengan fitur SK_ID_PREV.

- **POS_CASH_BALANCE**: Dataset yang terdiri dari data mengenai data bulanan tentang titik penjualan sebelumnya atau pinjaman tunai yang dimiliki klien dengan Home Credit. Setiap baris adalah satu bulan dari titik penjualan atau pinjaman tunai sebelumnya, dan satu pinjaman sebelumnya dapat memiliki banyak baris.

- **credit_card_balance**: data bulanan tentang kartu kredit sebelumnya yang dimiliki klien dengan Home Credit. Setiap baris adalah satu bulan dari saldo kartu kredit, dan satu kartu kredit dapat memiliki banyak baris.

- **installments_payment**: riwayat pembayaran untuk pinjaman sebelumnya di Home Credit. Ada satu baris untuk setiap pembayaran yang dilakukan dan satu baris untuk setiap pembayaran yang terlewatkan.

Data Preparation



application_{train|test}.csv

- Main tables – our train and test samples
- Target (binary)
- Info about loan and loan applicant at application time

bureau.csv

- Application data from previous loans that client got from other institutions and that were reported to Credit Bureau
- One row per client's loan in Credit Bureau

previous_application.csv

- Application data of client's previous loans in Home Credit
- Info about the previous loan parameters and client info at time of previous application
- One row per previous application

bureau_balance.csv

- Monthly balance of credits in Credit Bureau
- Behavioral data

POS_CASH_balance.csv

- Monthly balance of client's previous loans in Home Credit
- Behavioral data

instalments_payments.csv

- Past payment data for each installments of previous credits in Home Credit related to loans in our sample
- Behavioral data

credit_card_balance.csv

- Monthly balance of client's previous credit card loans in Home Credit
- Behavioral data

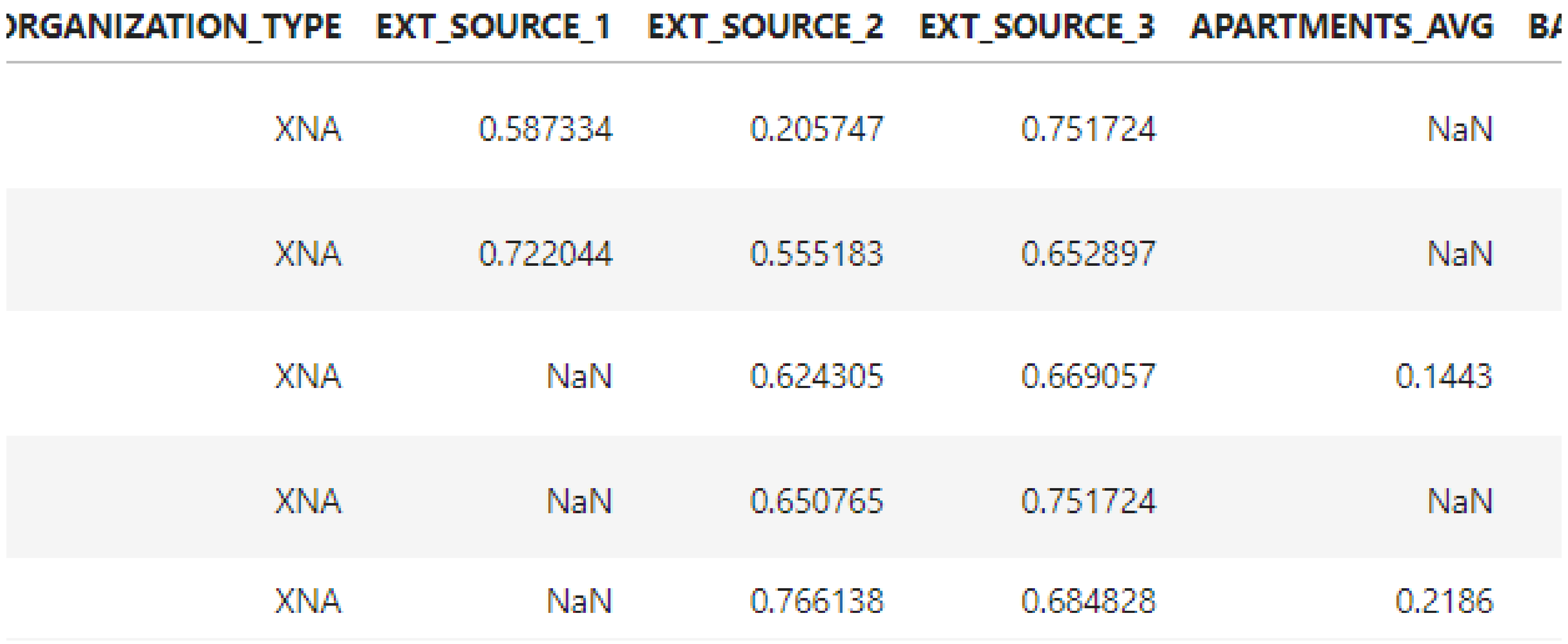
Data Preparation (Cleaning Data)

- Check for Duplicates
- Handling Data for Some Column
- Handling Missing Value
- Handling Outliers for Numerical Data

Data Preparation (Cleaning Data)

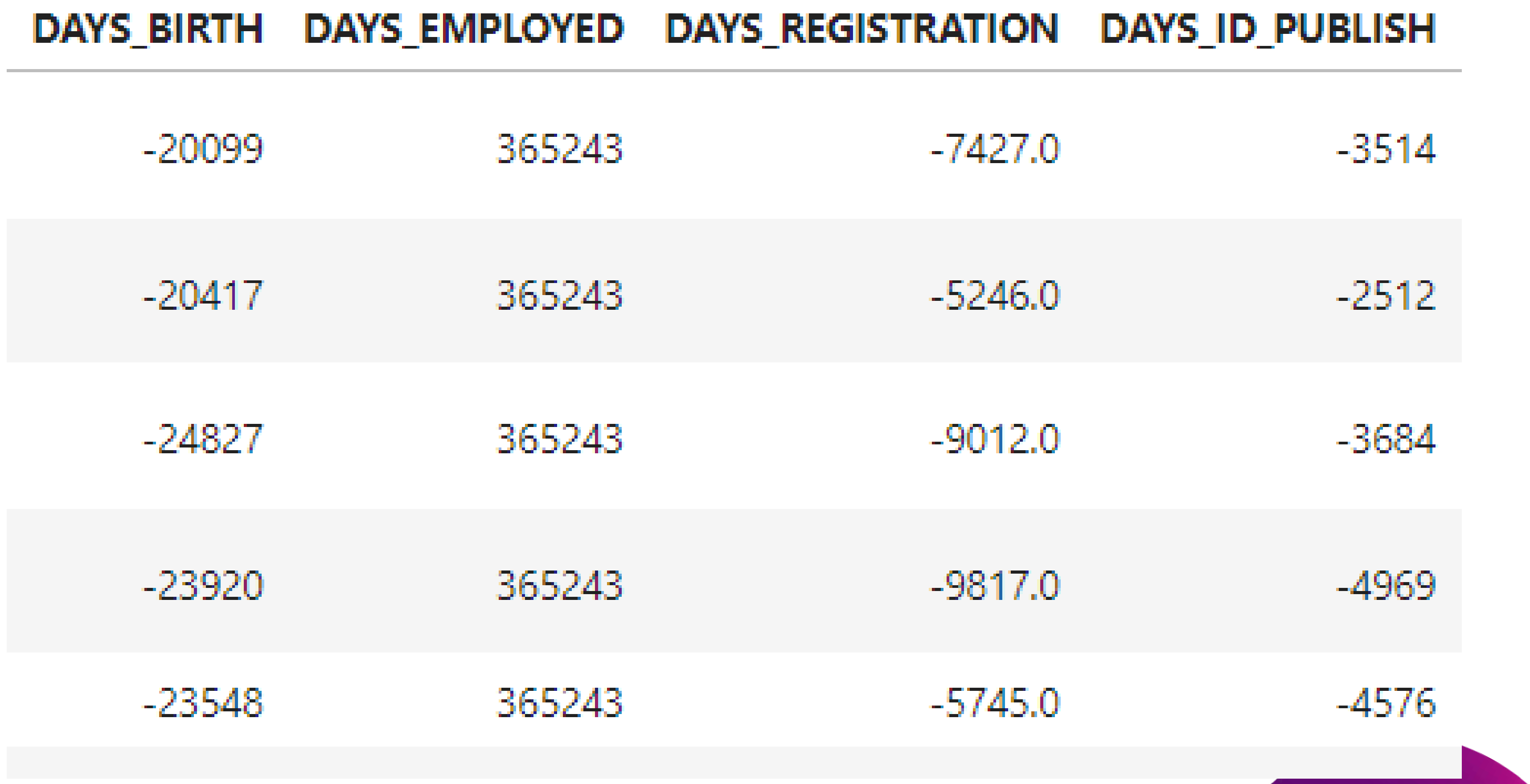
Handling data for some column

Replace XNA with NaN



RGANIZATION_TYPE	EXT_SOURCE_1	EXT_SOURCE_2	EXT_SOURCE_3	APARTMENTS_AVG	BA
XNA	0.587334	0.205747	0.751724	NaN	
XNA	0.722044	0.555183	0.652897	NaN	
XNA	NaN	0.624305	0.669057	0.1443	
XNA	NaN	0.650765	0.751724	NaN	
XNA	NaN	0.766138	0.684828	0.2186	

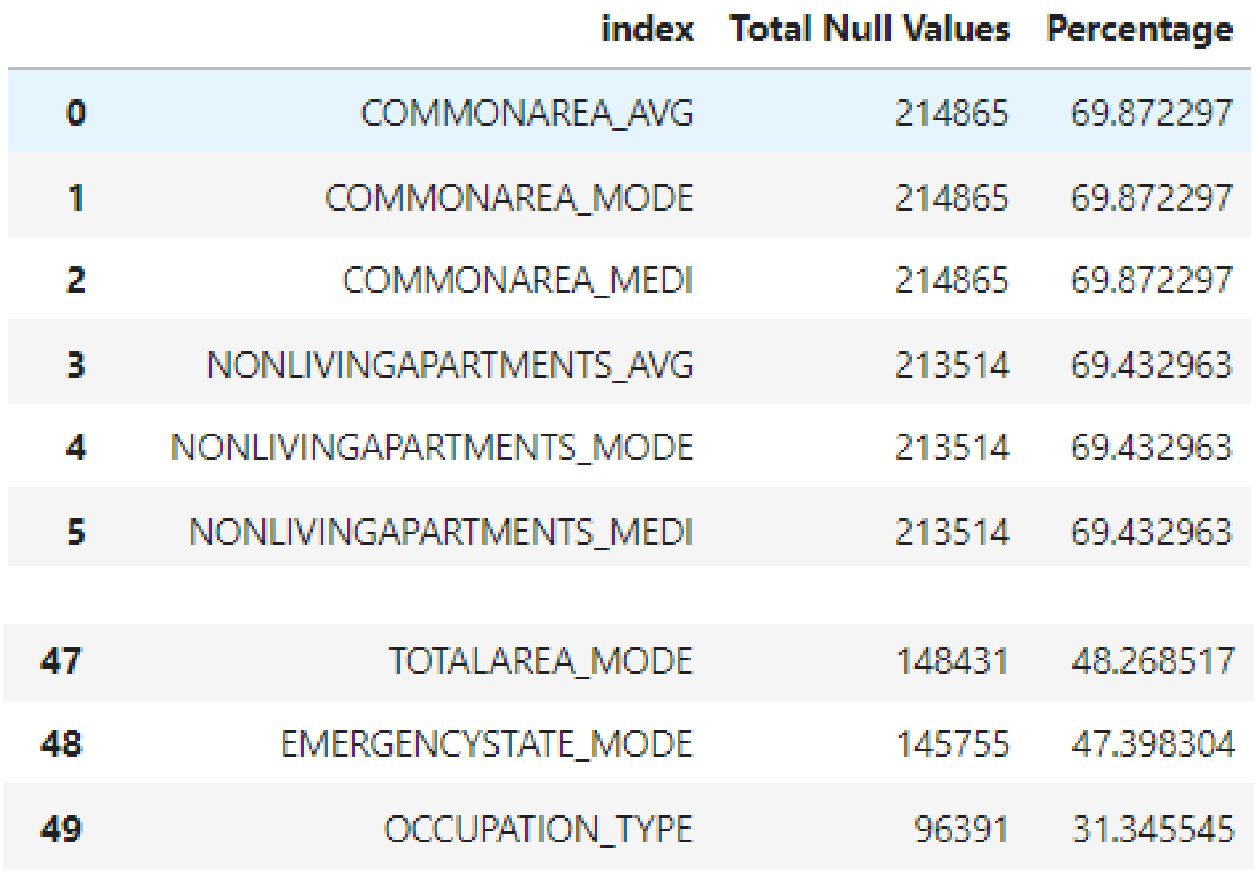
Change days to years



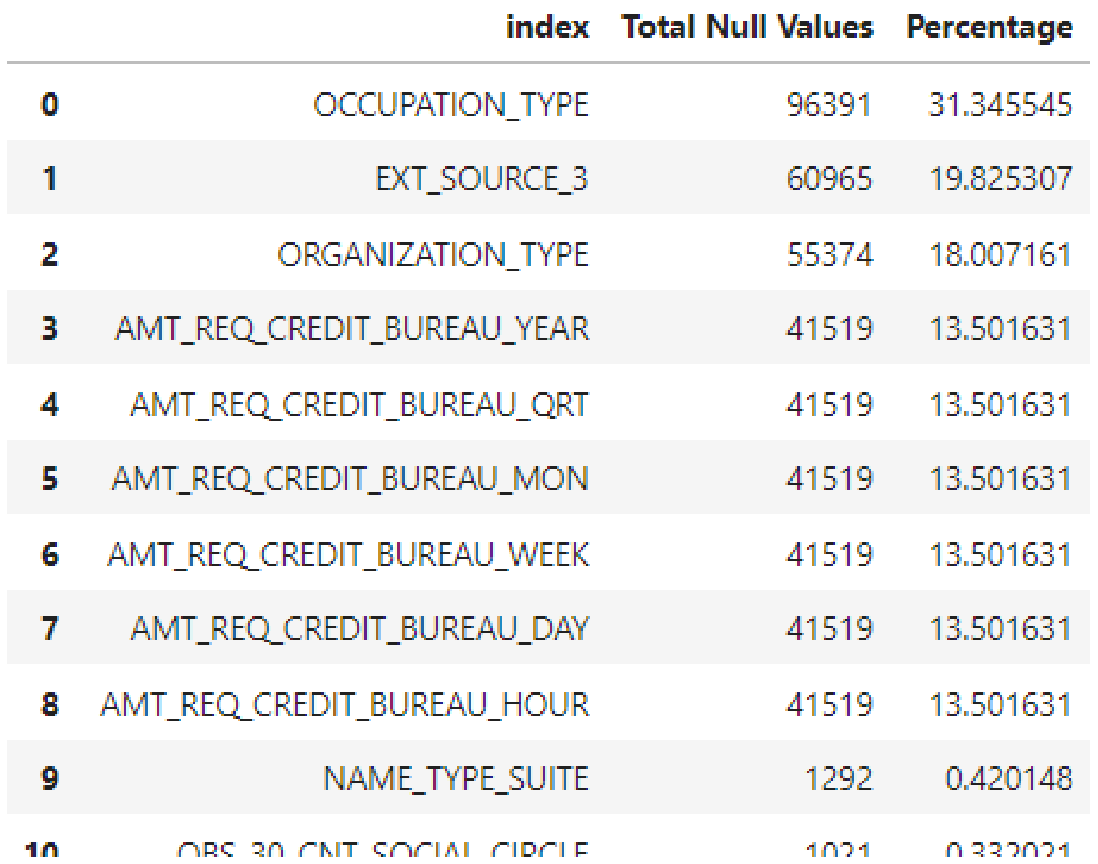
DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH
-20099	365243	-7427.0	-3514
-20417	365243	-5246.0	-2512
-24827	365243	-9012.0	-3684
-23920	365243	-9817.0	-4969
-23548	365243	-5745.0	-4576

DataPreparation (CleaningData)

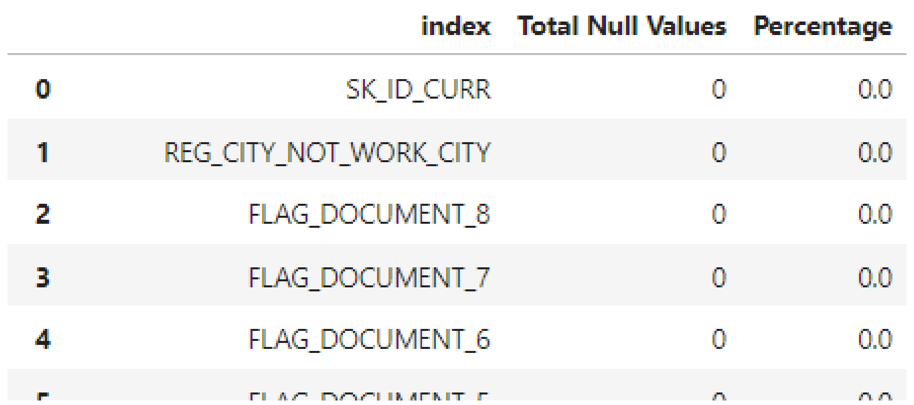
Handling Missing Values



	index	Total Null Values	Percentage
0	COMMONAREA_AVG	214865	69.872297
1	COMMONAREA_MODE	214865	69.872297
2	COMMONAREA_MEDI	214865	69.872297
3	NONLIVINGAPARTMENTS_AVG	213514	69.432963
4	NONLIVINGAPARTMENTS_MODE	213514	69.432963
5	NONLIVINGAPARTMENTS_MEDI	213514	69.432963
47	TOTALAREA_MODE	148431	48.268517
48	EMERGENCYSTATE_MODE	145755	47.398304
49	OCCUPATION_TYPE	96391	31.345545



	index	Total Null Values	Percentage
0	OCCUPATION_TYPE	96391	31.345545
1	EXT_SOURCE_3	60965	19.825307
2	ORGANIZATION_TYPE	55374	18.007161
3	AMT_REQ_CREDIT_BUREAU_YEAR	41519	13.501631
4	AMT_REQ_CREDIT_BUREAU_QRT	41519	13.501631
5	AMT_REQ_CREDIT_BUREAU_MON	41519	13.501631
6	AMT_REQ_CREDIT_BUREAU_WEEK	41519	13.501631
7	AMT_REQ_CREDIT_BUREAU_DAY	41519	13.501631
8	AMT_REQ_CREDIT_BUREAU_HOUR	41519	13.501631
9	NAME_TYPE_SUITE	1292	0.420148
10	OBS_30_CNT_SOCIAL_CIRCLE	1021	0.332021

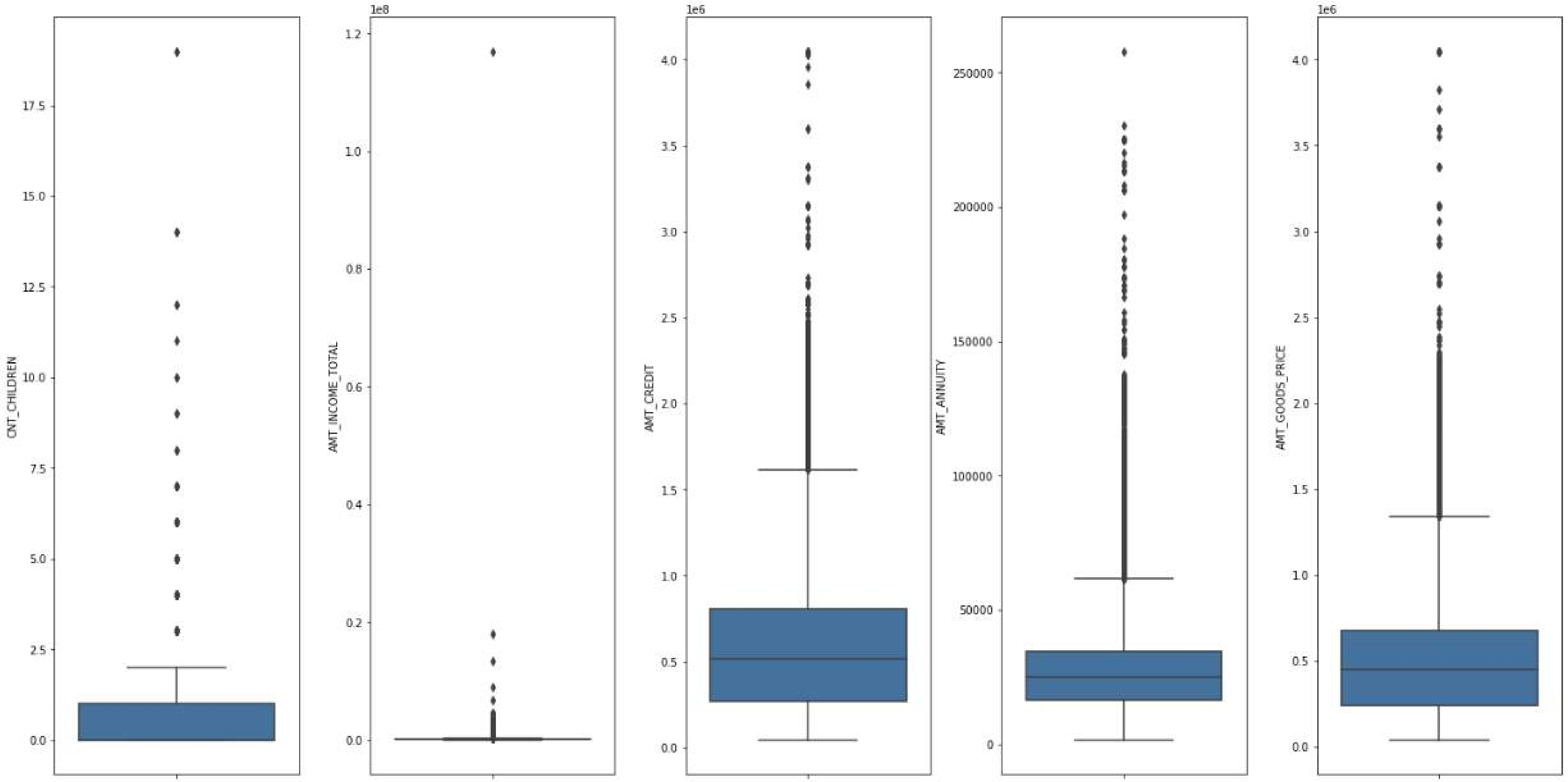


	index	Total Null Values	Percentage
0	SK_ID_CURR	0	0.0
1	REG_CITY_NOT_WORK_CITY	0	0.0
2	FLAG_DOCUMENT_8	0	0.0
3	FLAG_DOCUMENT_7	0	0.0
4	FLAG_DOCUMENT_6	0	0.0
5	FLAG_DOCUMENT_5	0	0.0

Ada 48 data dengan persentase missing value yang tinggi. Oleh karena itu kita melakukan drop column dan input data dengan modus dan median

Data Preparation (Cleaning Data)

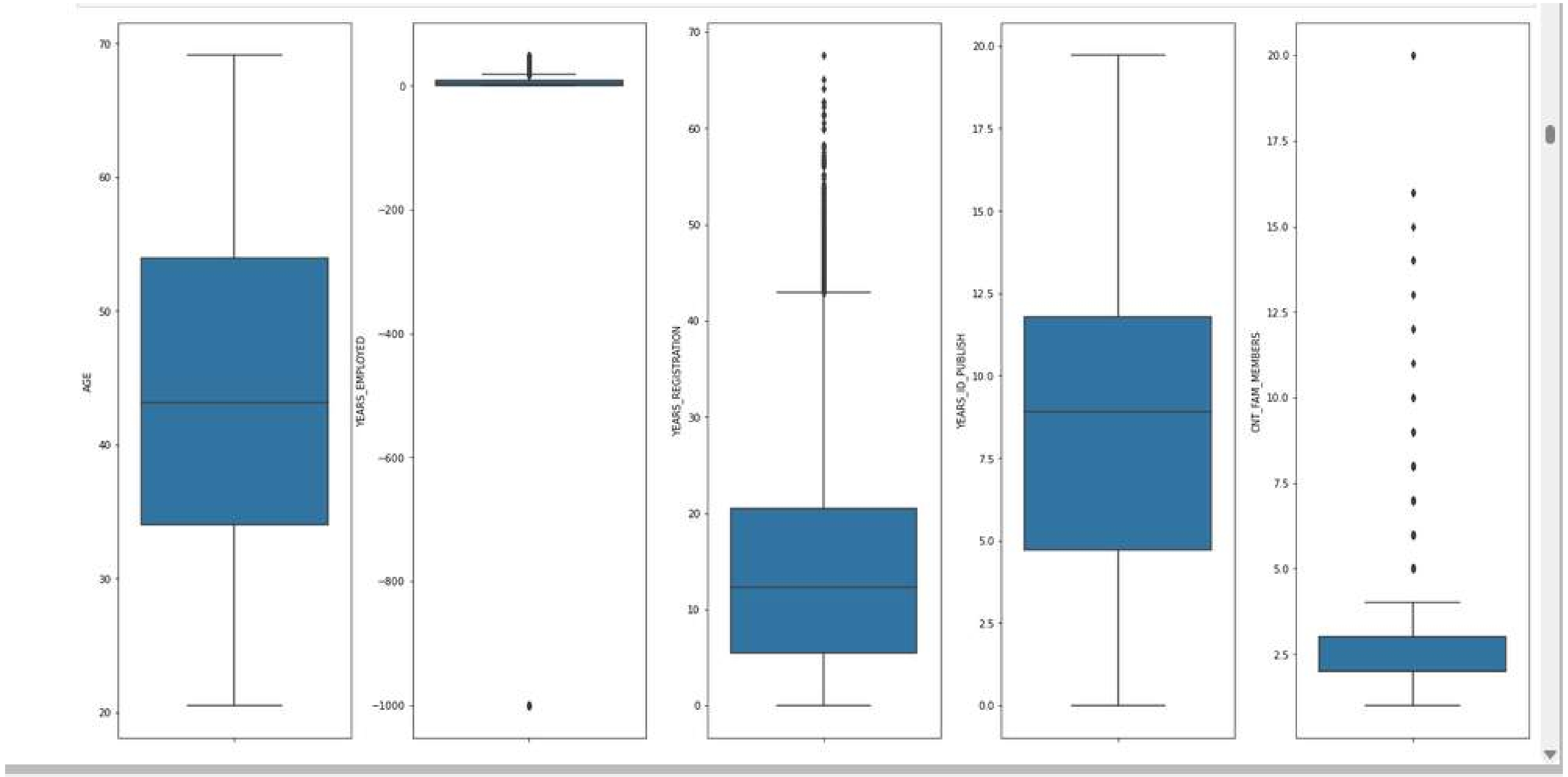
Handling Outliers for Numerical Data



Boxplot dari kolom "Count Children, Amount Income, Amount Credit, Amount Atenuity, Good Price"

Data Preparation (Cleaning Data)

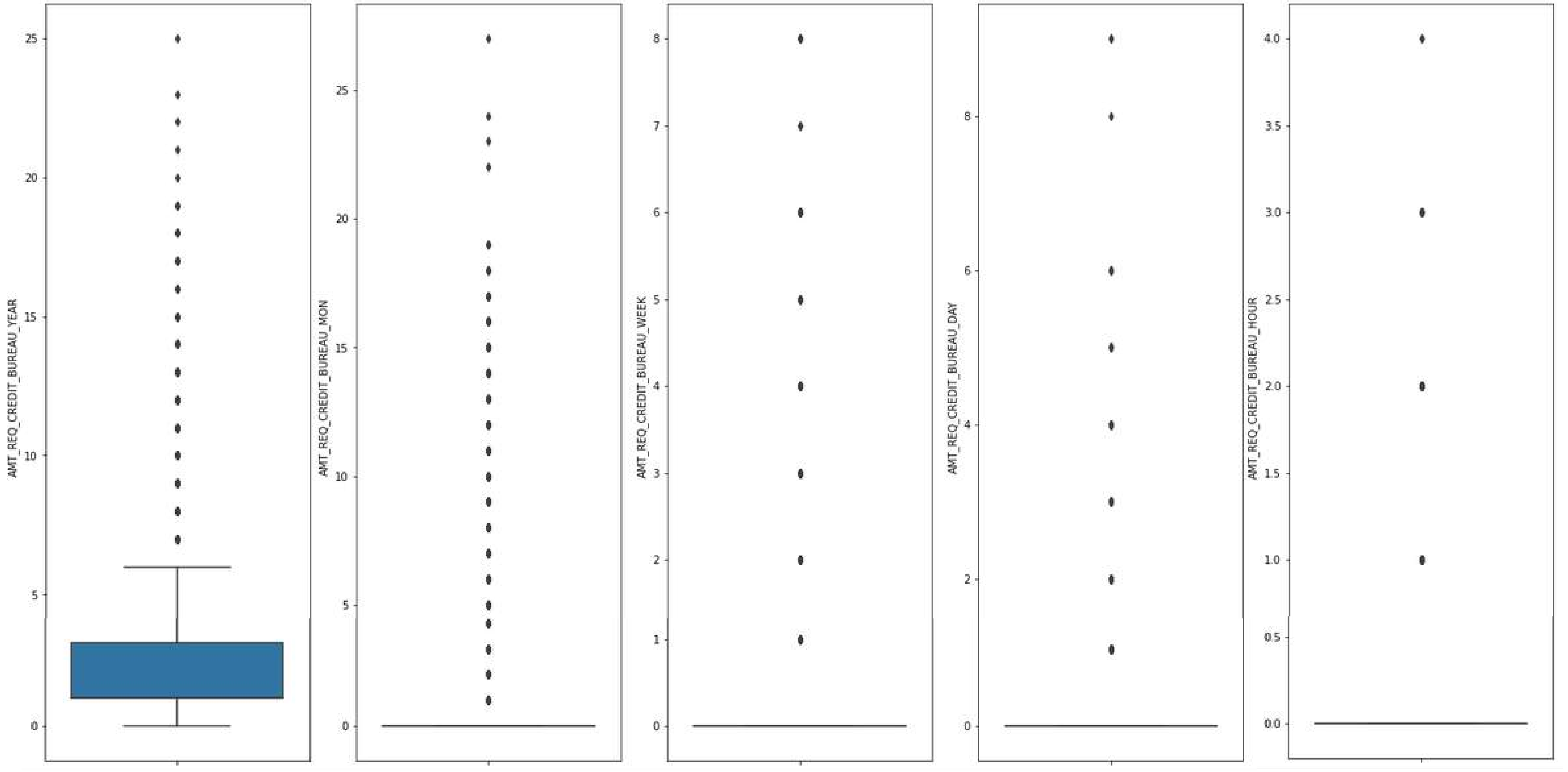
- **Handling Outliers for Numerical Data**



Boxplot dari kolom "AGE", "YEARS_EMPLOYED", "YEARS_REGISTRATION", "YEARS_ID_PUBLISH", "CNT_FAM_MEMBERS"

Data Preparation (Cleaning Data)

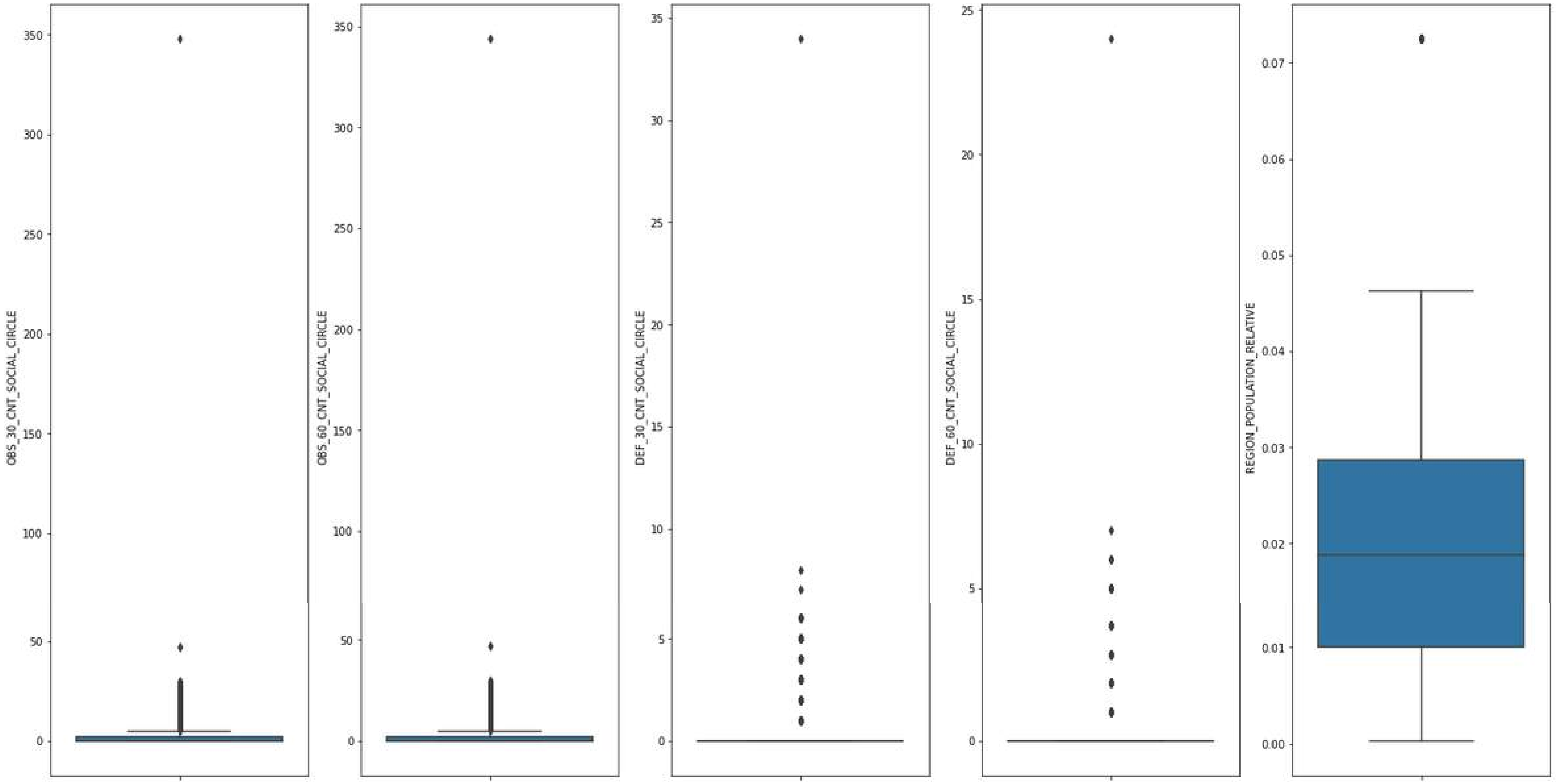
- Handling Outliers for Numerical Data



Boxplot dari kolom "AMT_REQ_CREDIT_BUREAU_YEAR", "AMT_REQ_CREDIT_BUREAU_MON", "AMT_REQ_CREDIT_BUREAU_WEEK", "AMT_REQ_CREDIT_BUREAU_DAY", "AMT_REQ_CREDIT_BUREAU_HOUR"

Data Preparation (Cleaning Data)

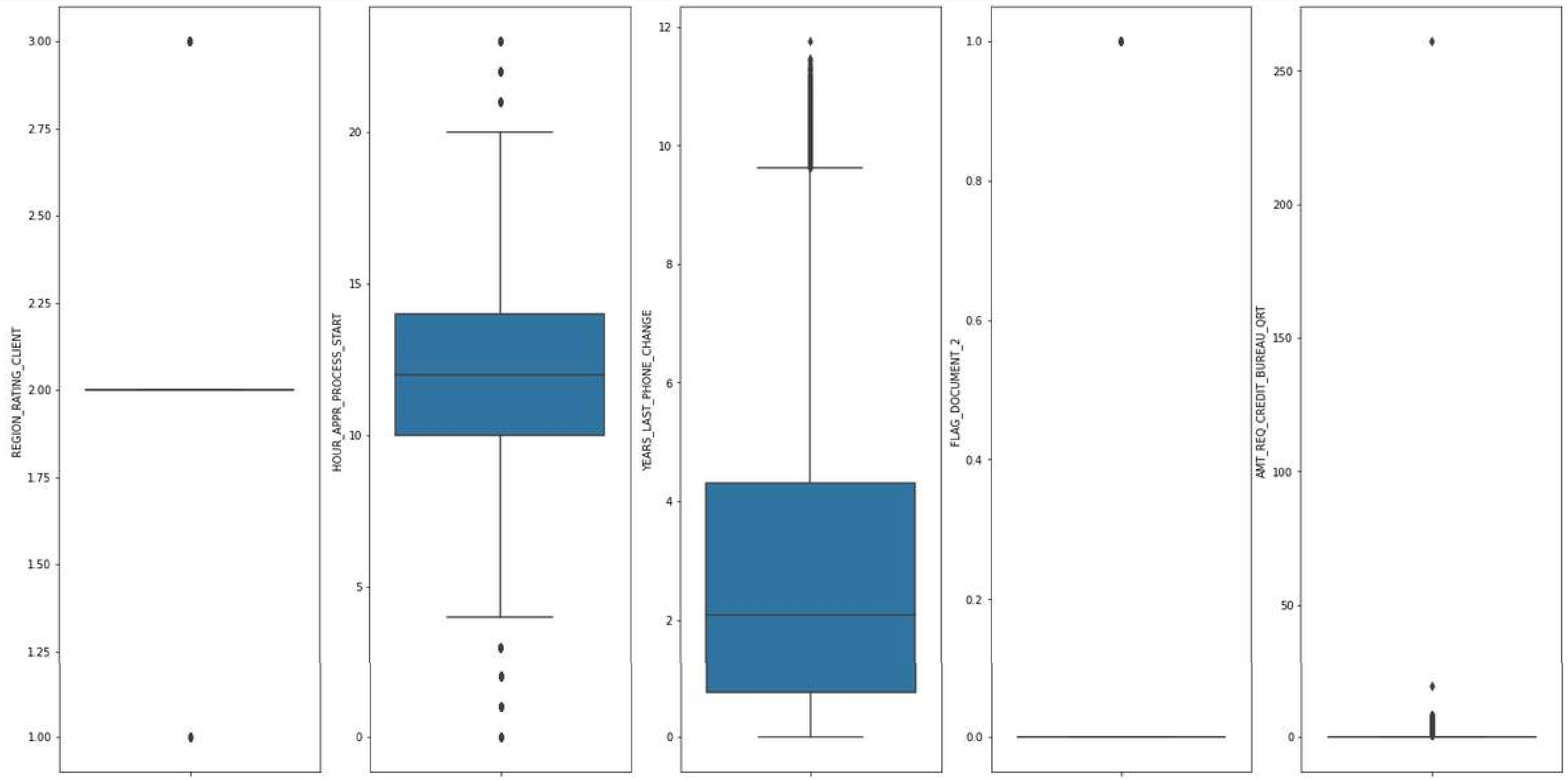
- **Handling Outliers for Numerical Data**



Boxplot dari kolom "OBS_30_CNT_SOCIAL_CIRCLE", "OBS_60_CNT_SOCIAL_CIRCLE", "DEF_30_CNT_SOCIAL_CIRCLE", "DEF_60_CNT_SOCIAL_CIRCLE", "REGION_POPULATION_RELATIVE"

Data Preparation (Cleaning Data)

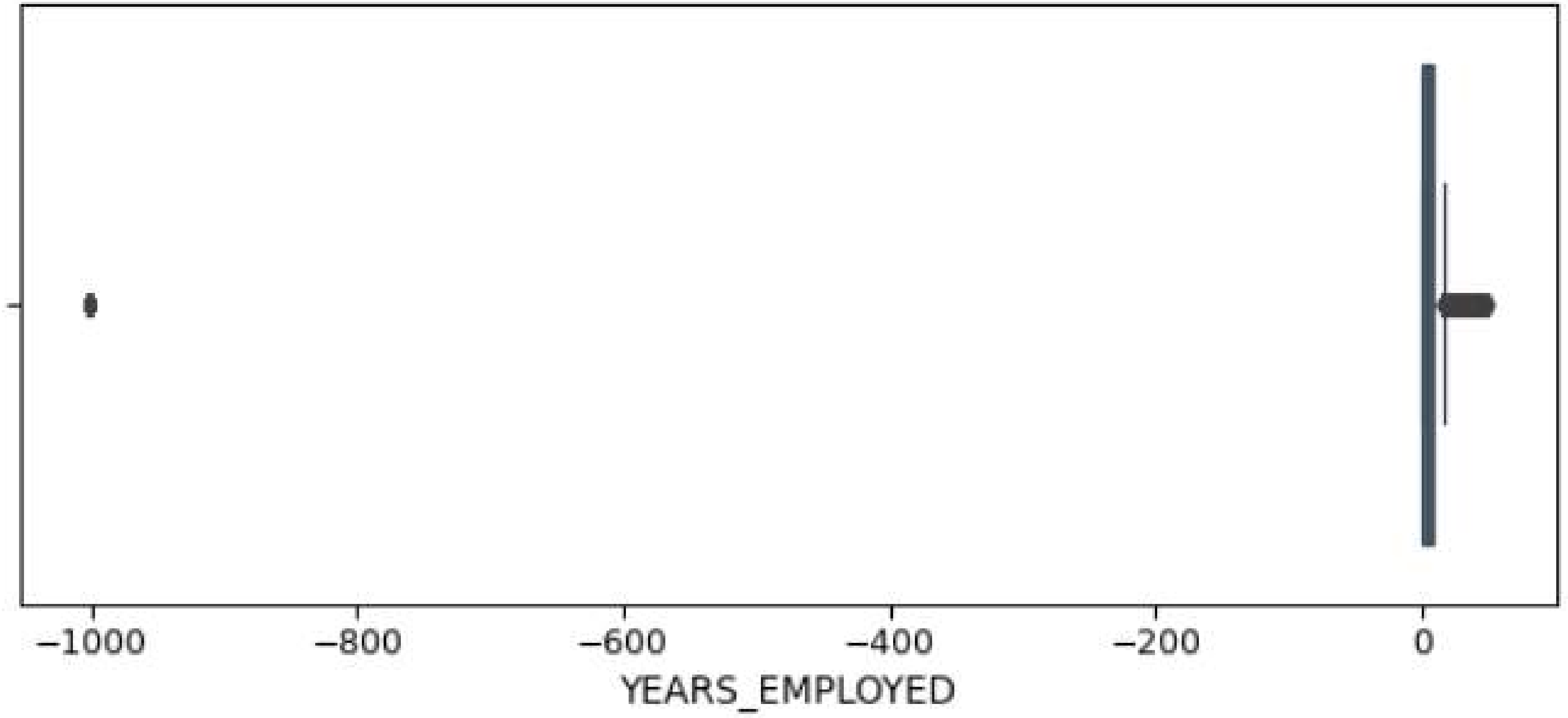
- **Handling Outliers for Numerical Data**

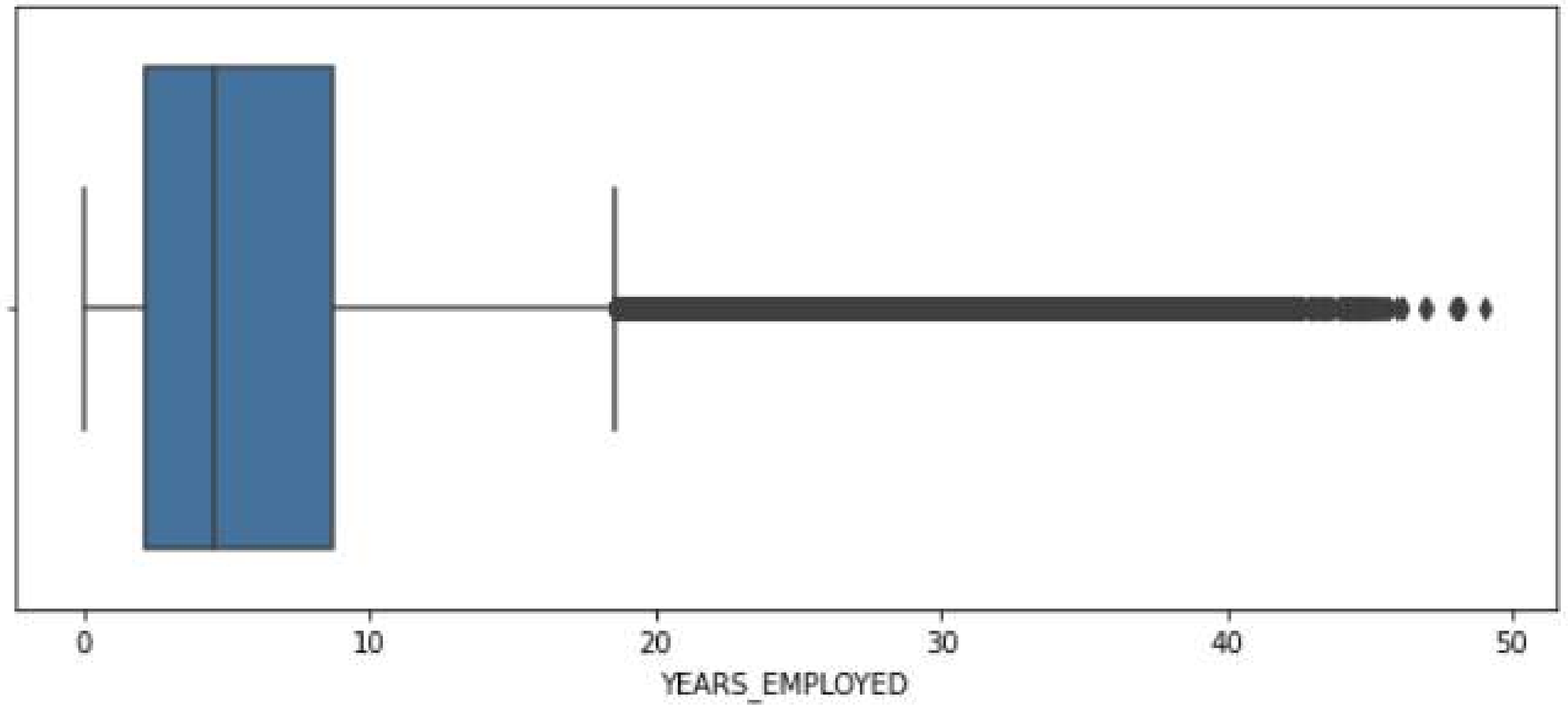


Boxplot dari kolom "REGION_RATING_CLIENT", "HOUR_APPR_PROCESS_START", "YEARS_LAST_PHONE_CHANGE", "FLAG_DOCUMENT_2", "AMT_REQ_CREDIT_BUREAU_QRT"

Data Preparation (Cleaning Data)

Handling the Outliers (Years Employed)

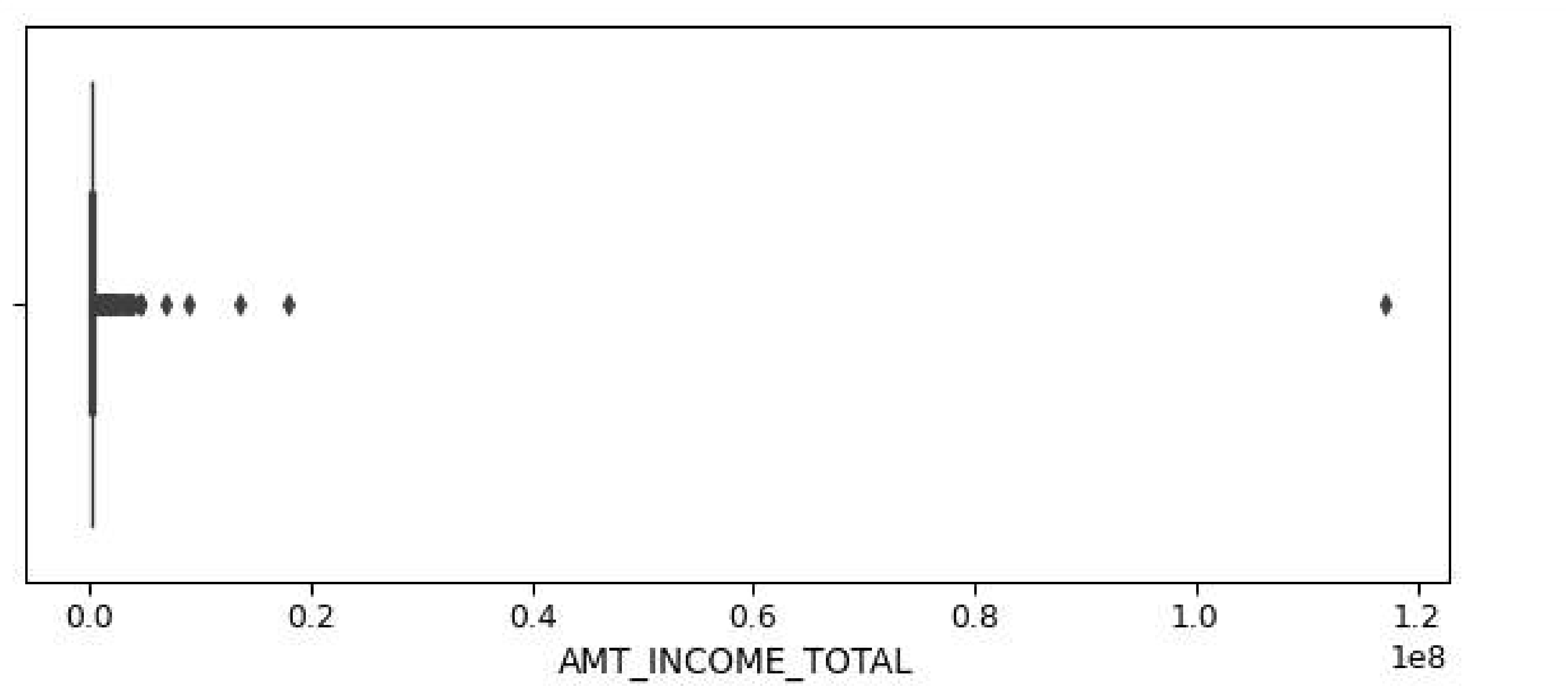


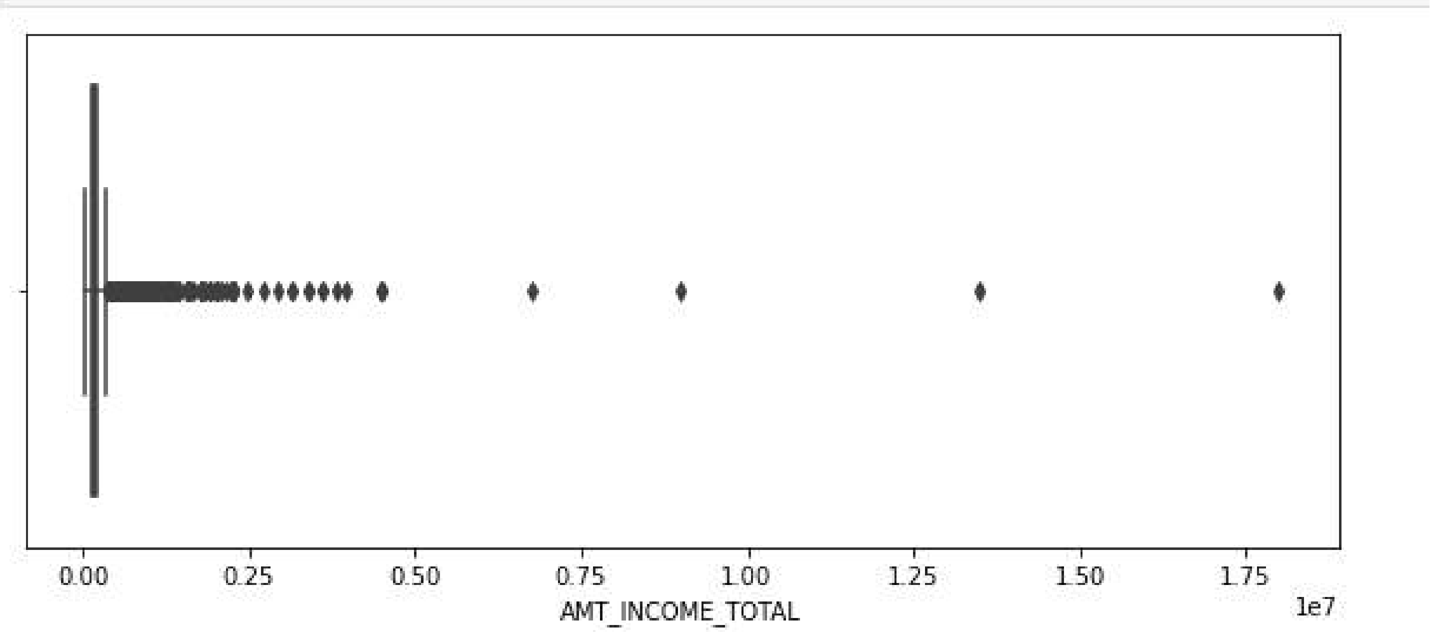


Boxplot YEARS_EMPLOYED

Data Preparation (Cleaning Data)

Handling the Outliers (Amount Income Total)

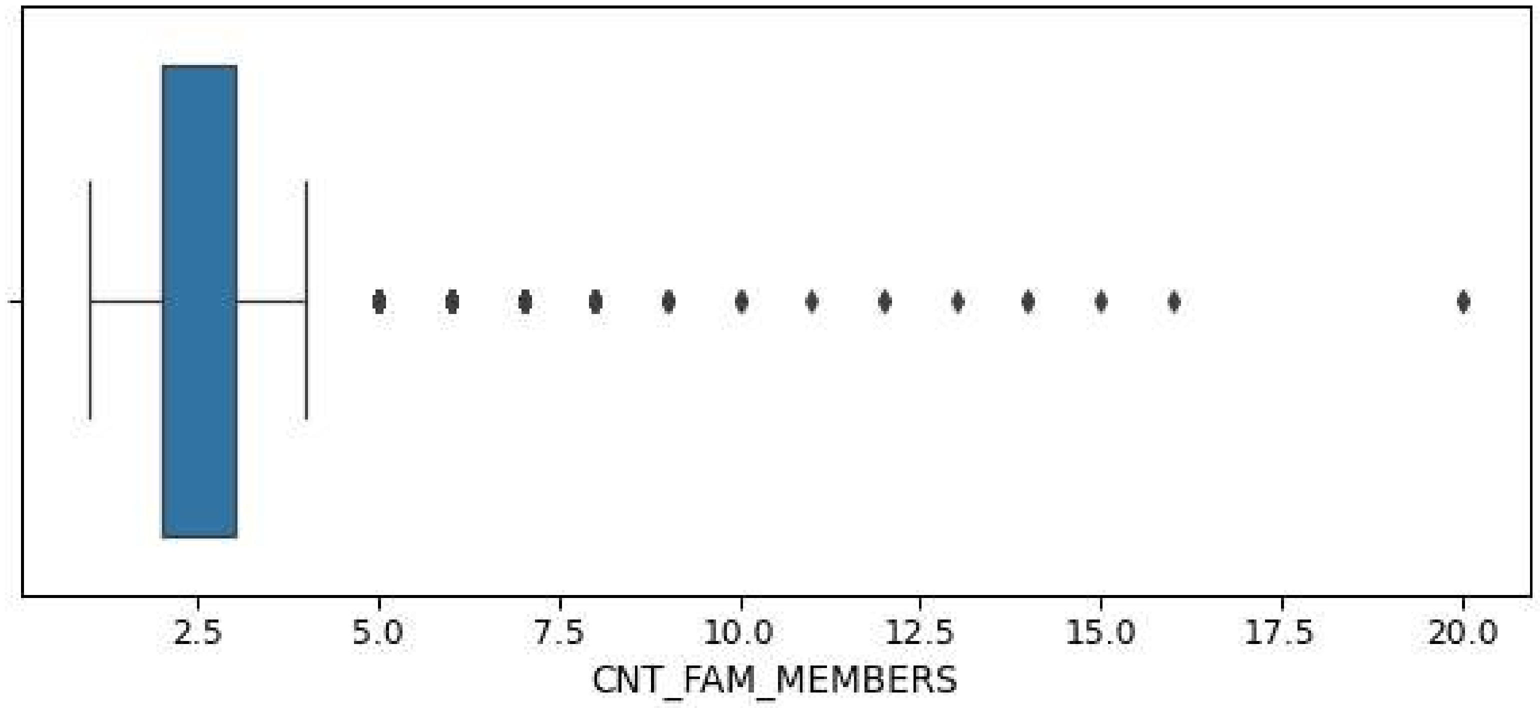


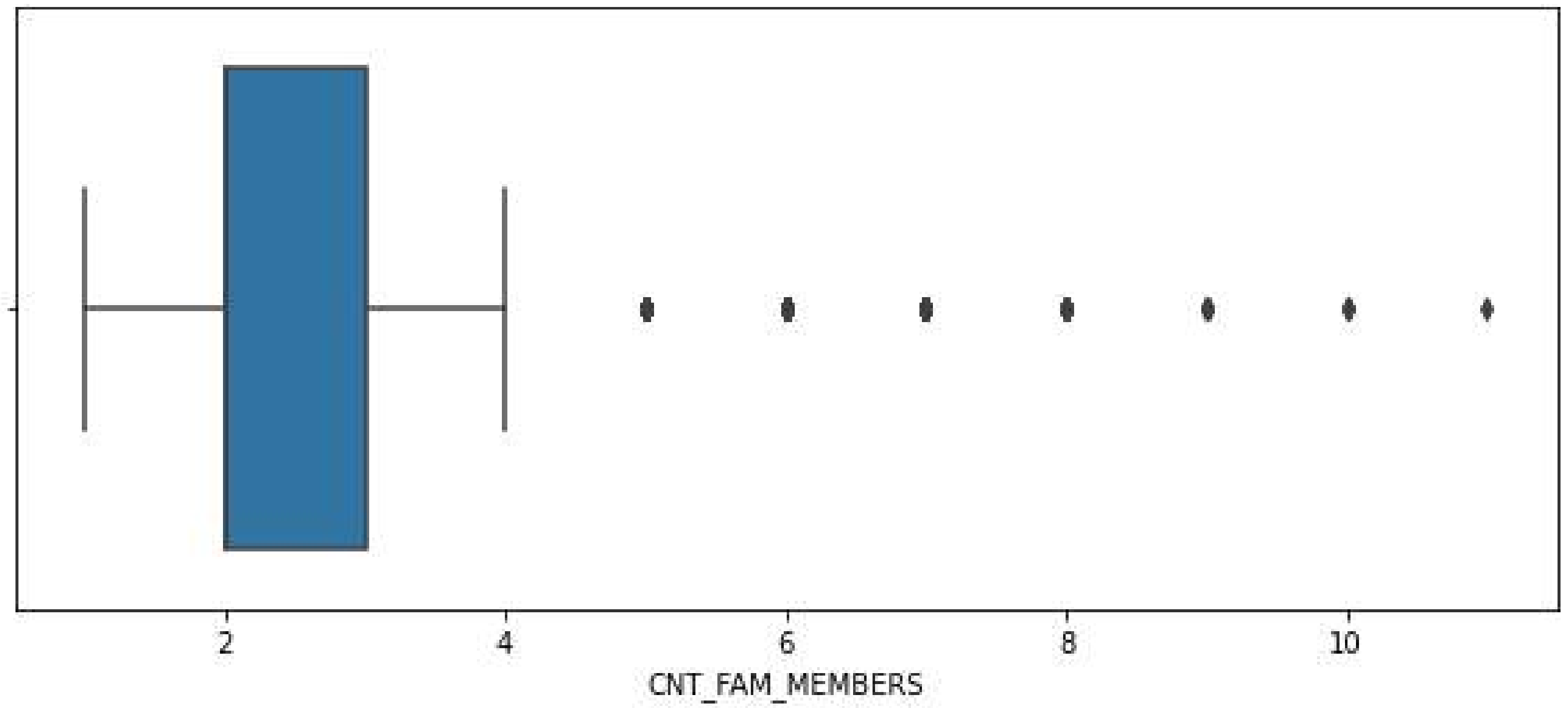


Boxplot Amount Income Total

Data Preparation (Cleaning Data)

Handling the Odd (Count Children)

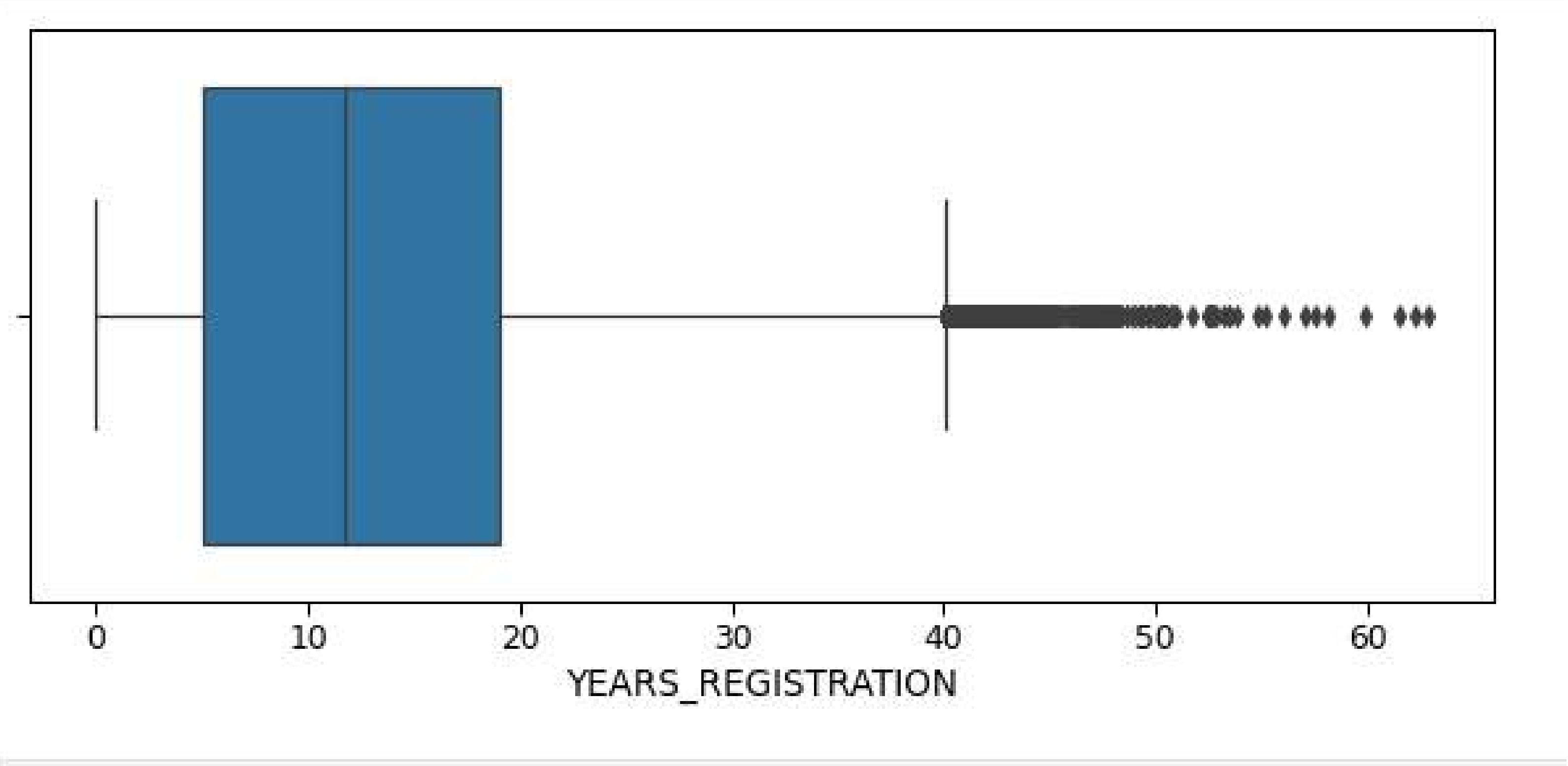


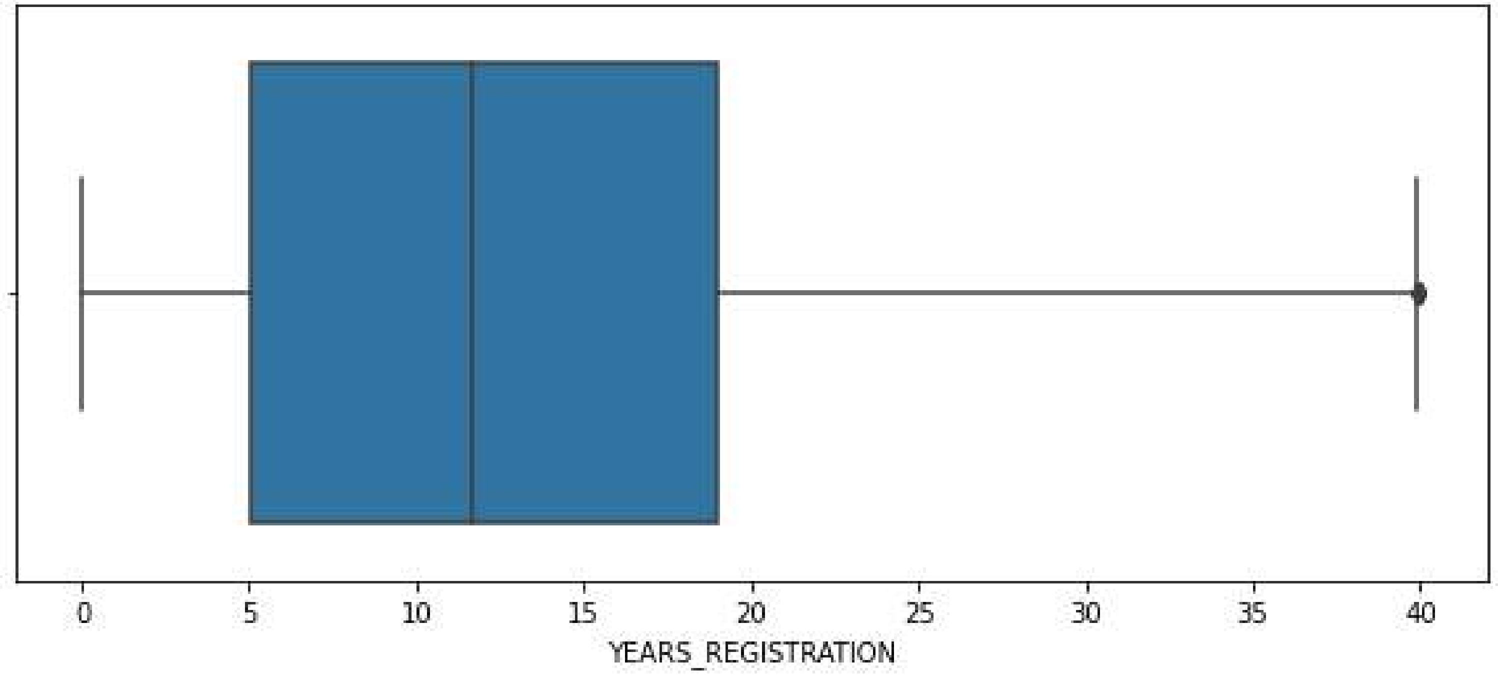


Boxplot Fam Members

Data Preparation (Cleaning Data)

Handling the Outliers (Years Registration)

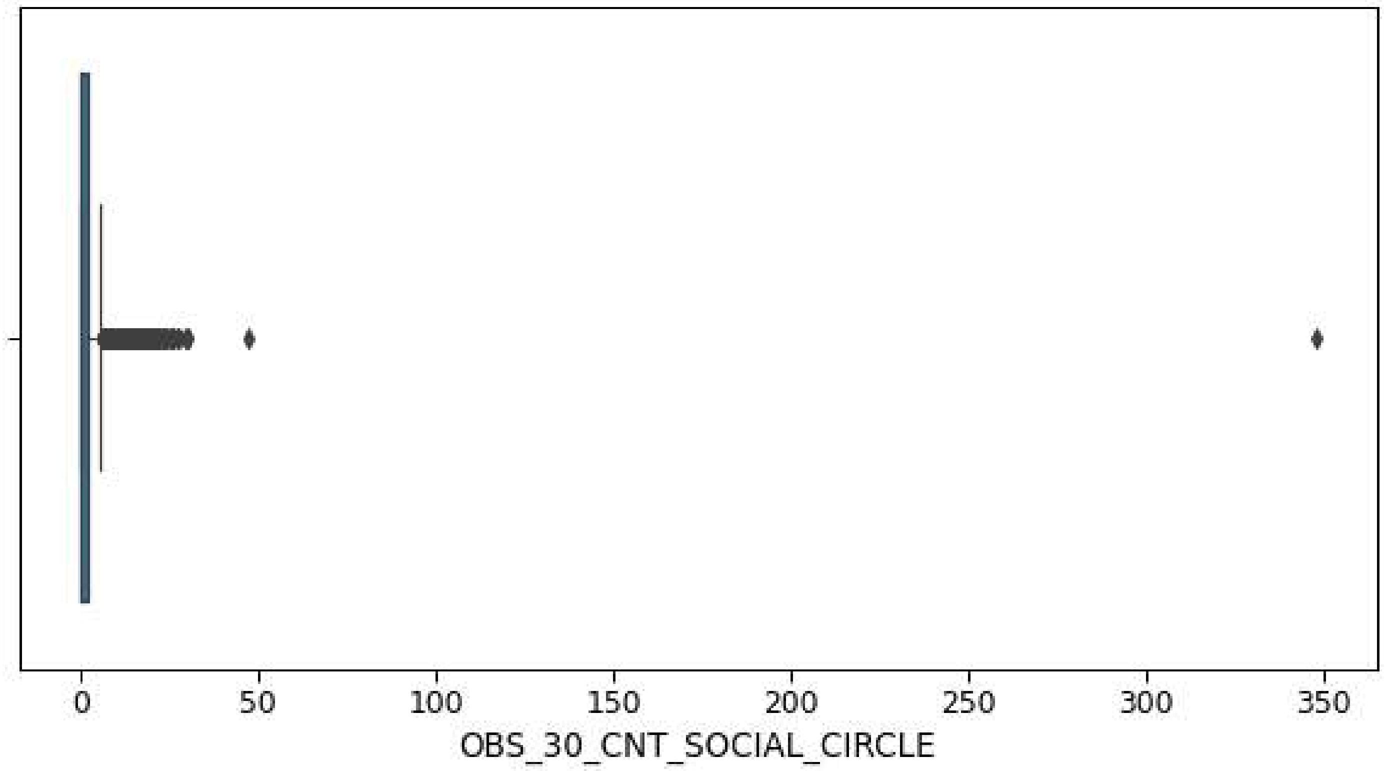


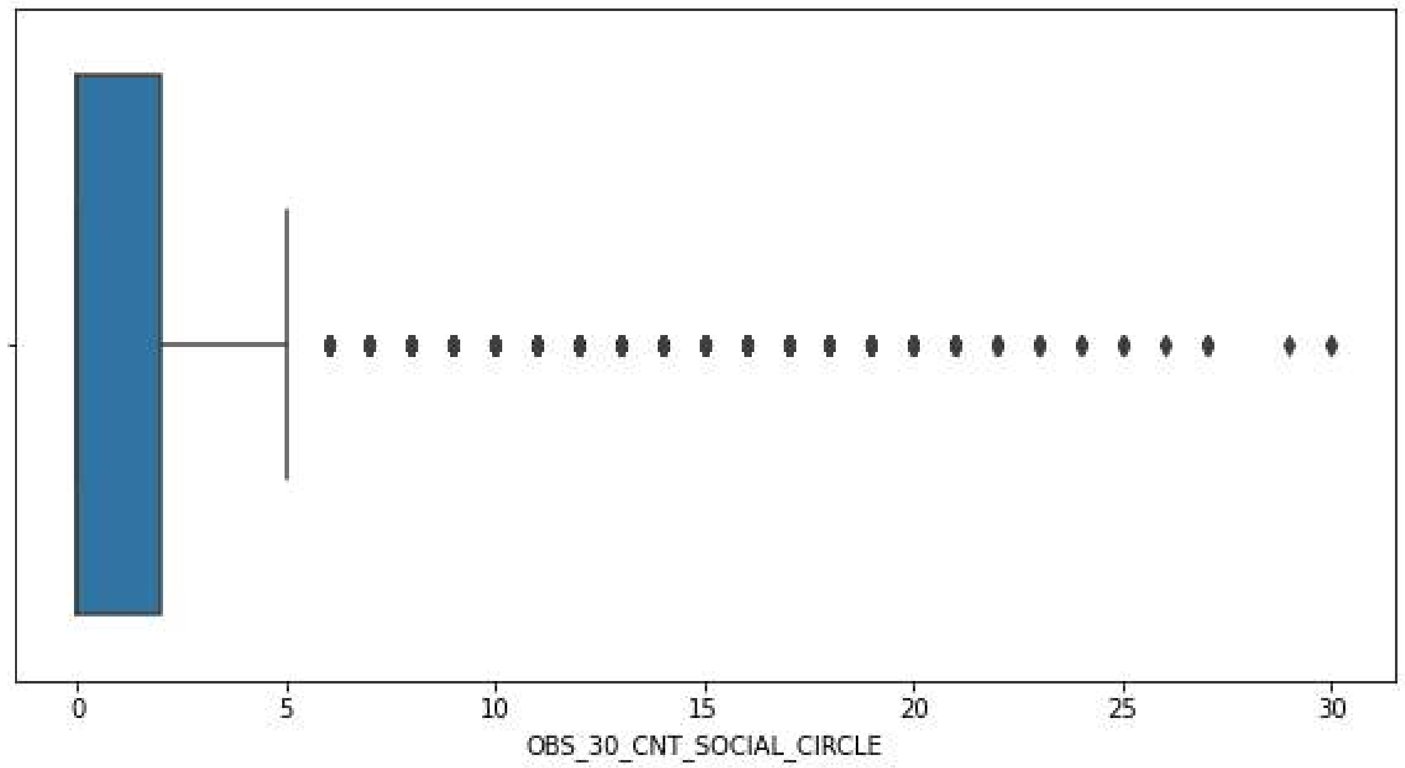


Boxplot Years Registration

Data Preparation (Cleaning Data)

Handling the Outliers (Social Circle)

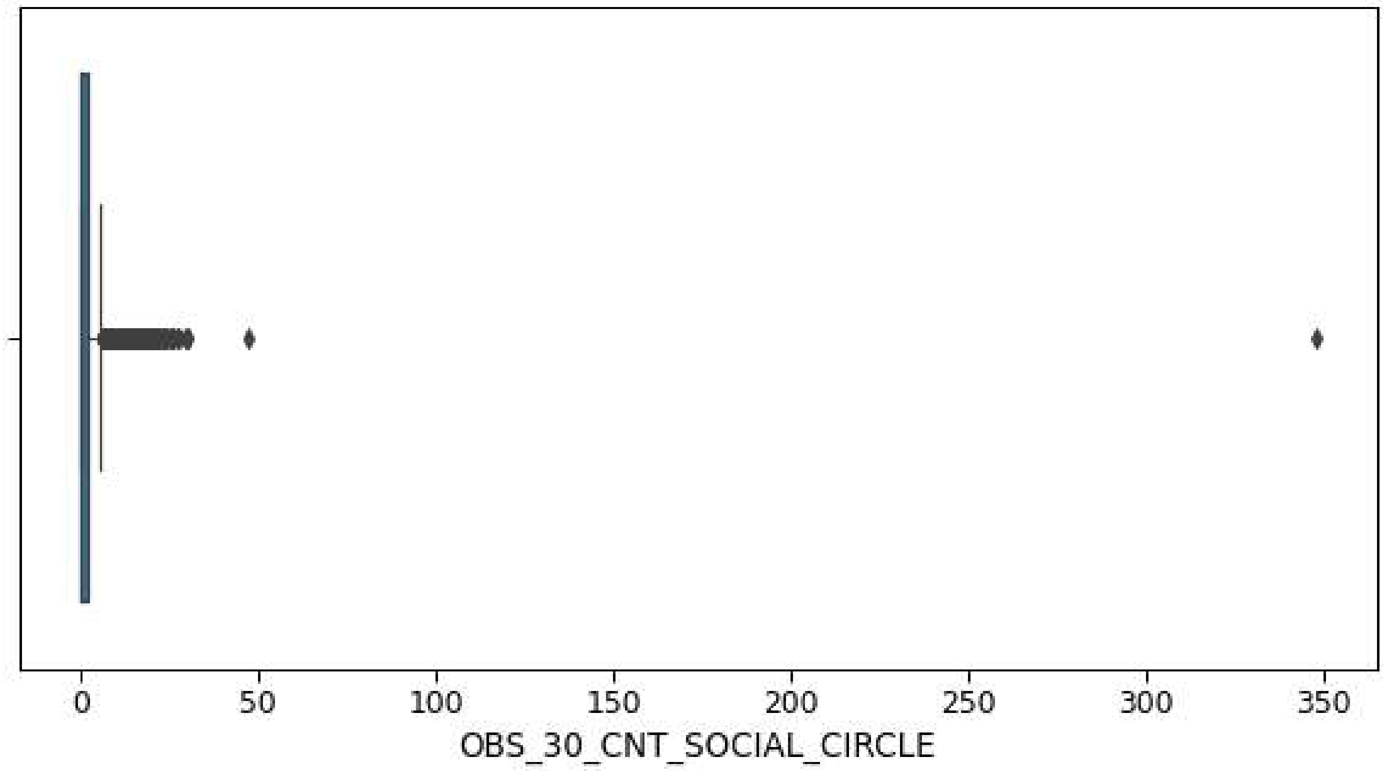


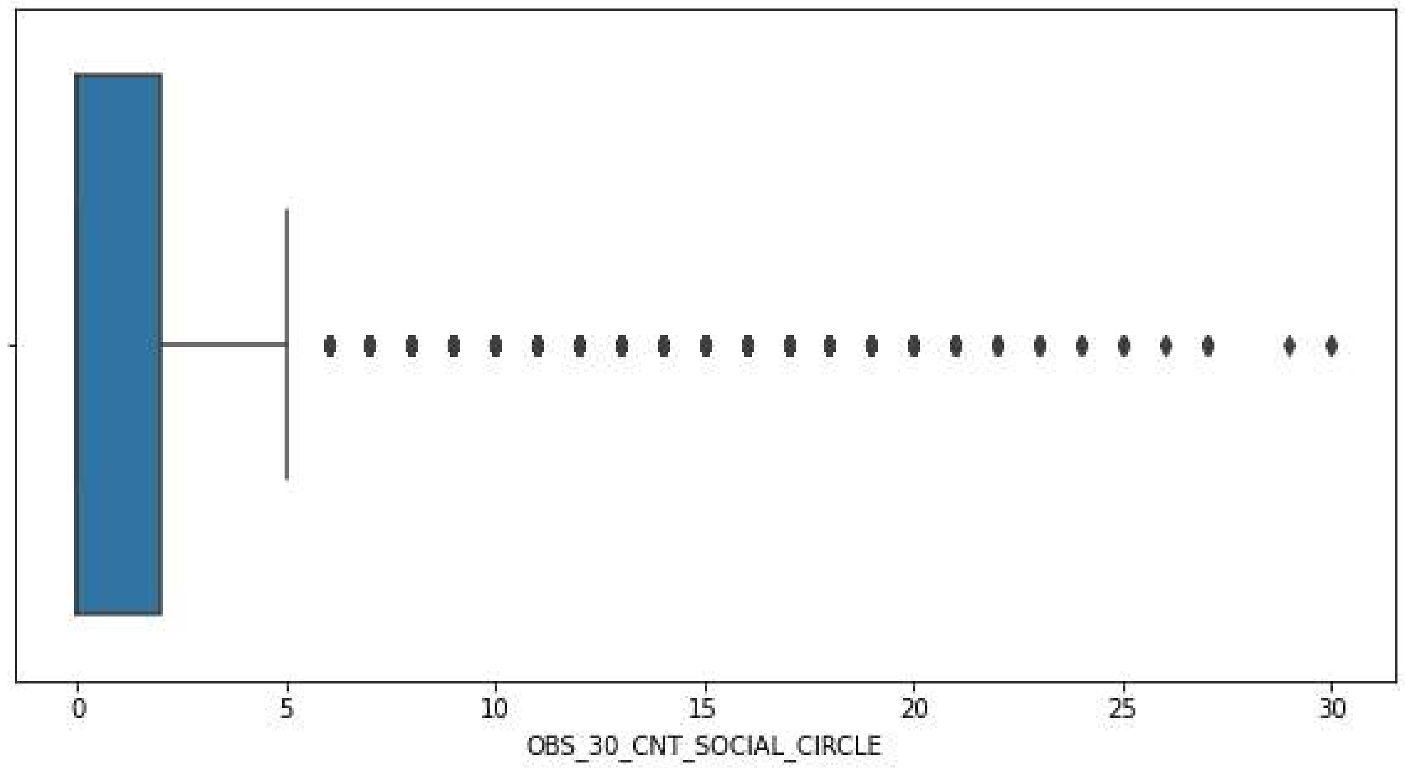


Boxplot Social Circle

Data Preparation (Cleaning Data)

Handling the Outliers (Social Circle)

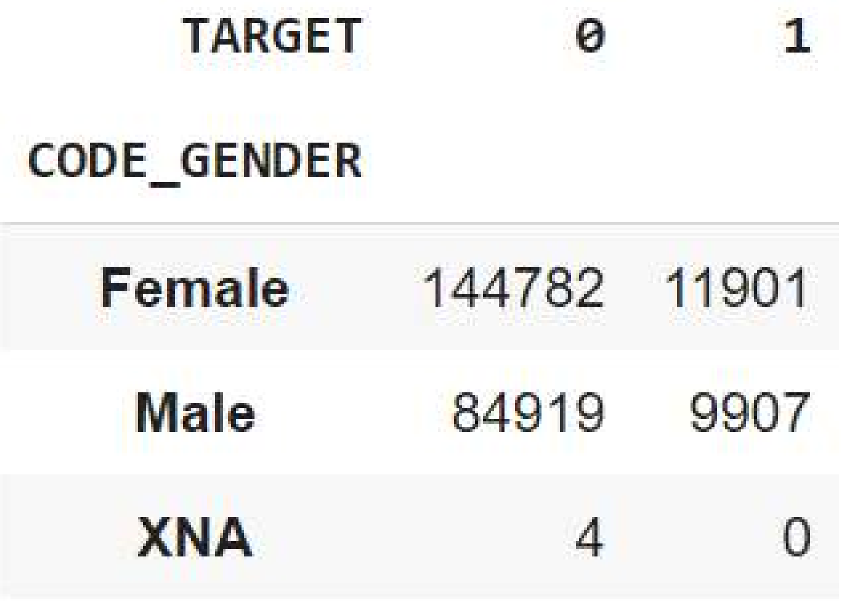




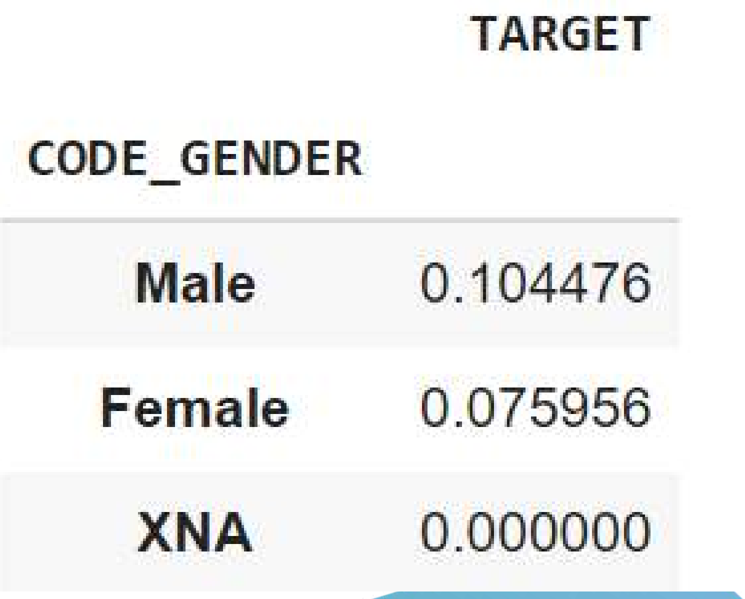
Boxplot Social Circle

Data Preparation (EDA)

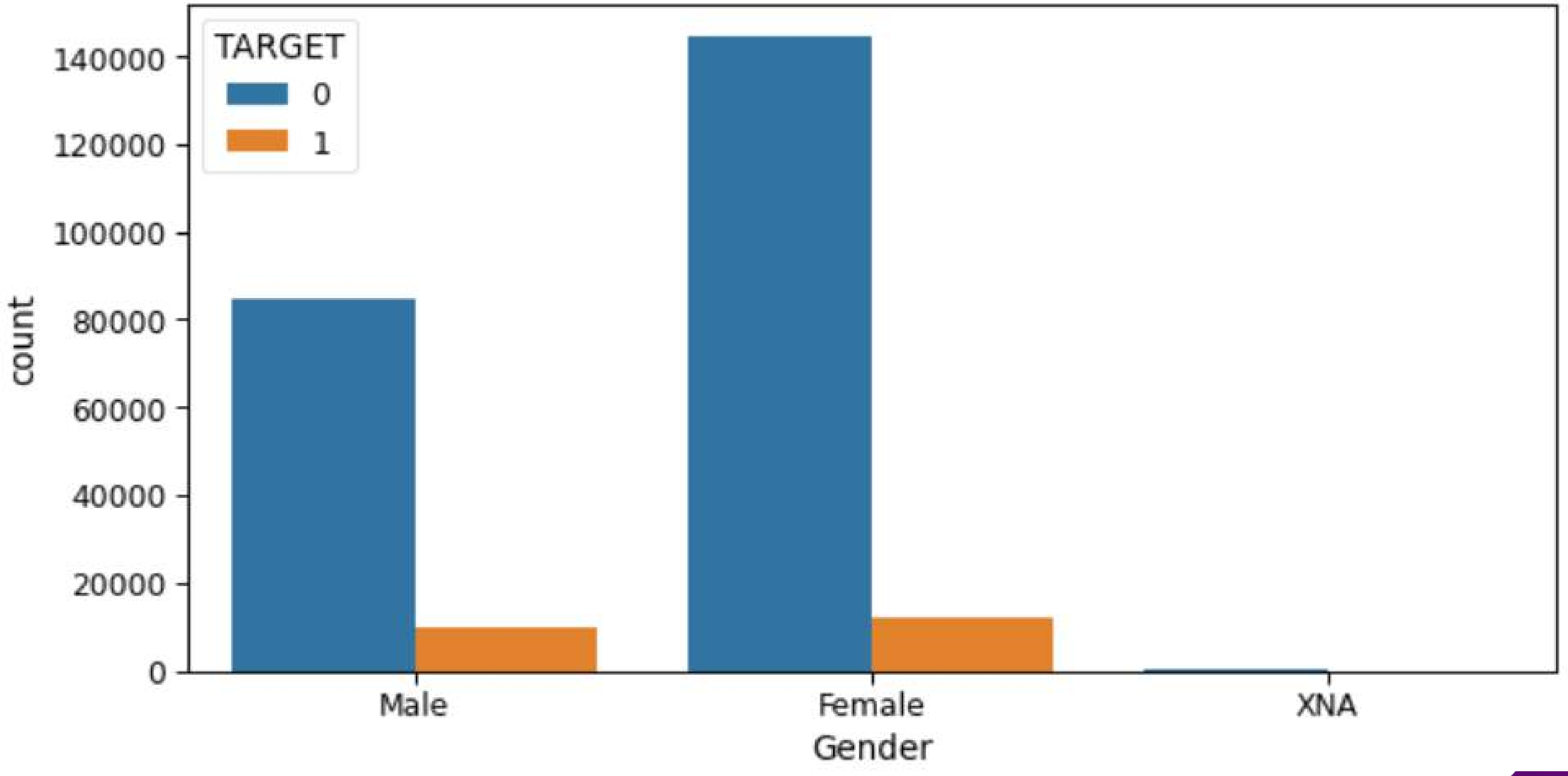
Gender vs Target



TARGET	0	1
CODE_GENDER		
Female	144782	11901
Male	84919	9907
XNA	4	0

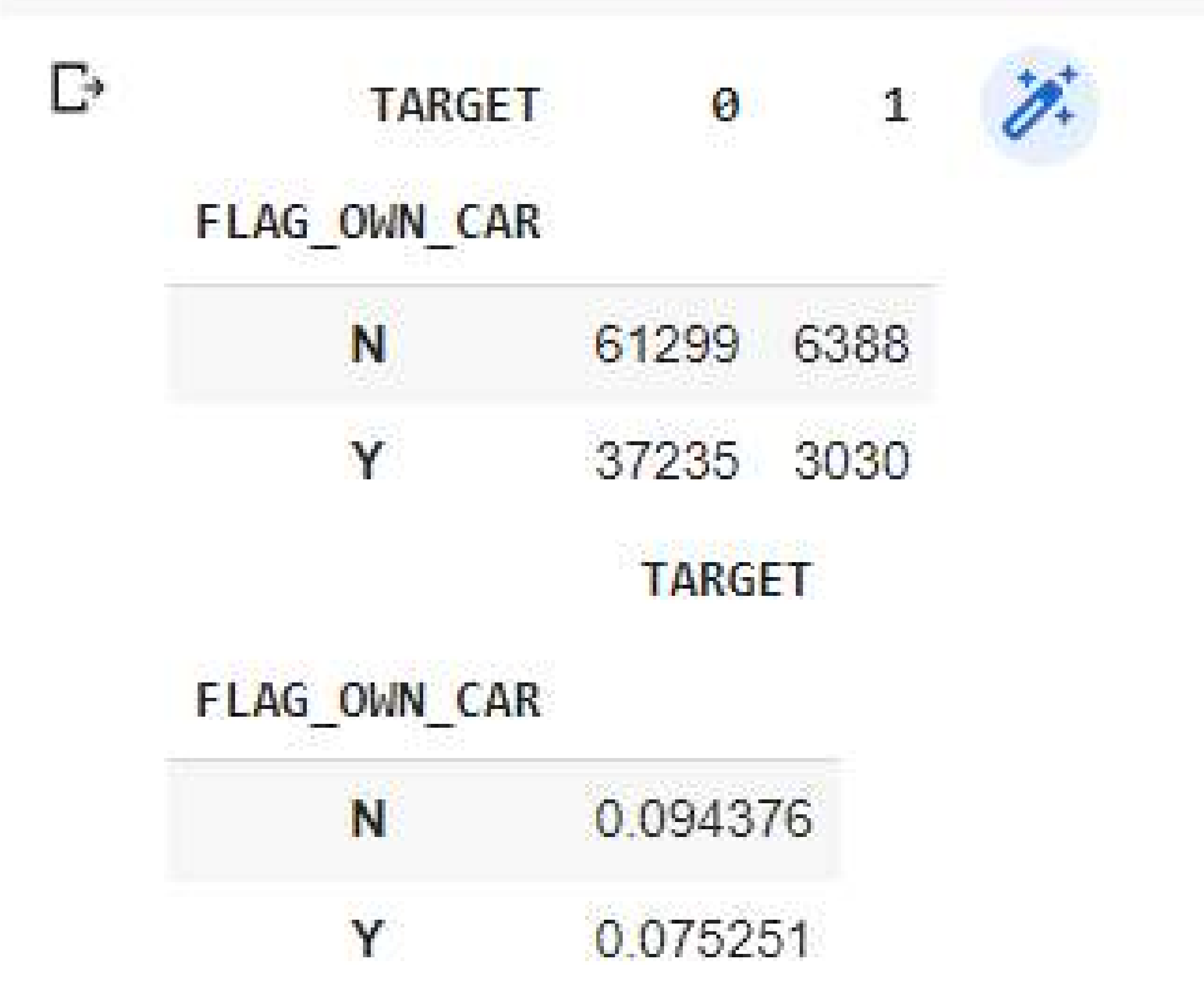


	TARGET
CODE_GENDER	
Male	0.104476
Female	0.075956
XNA	0.000000

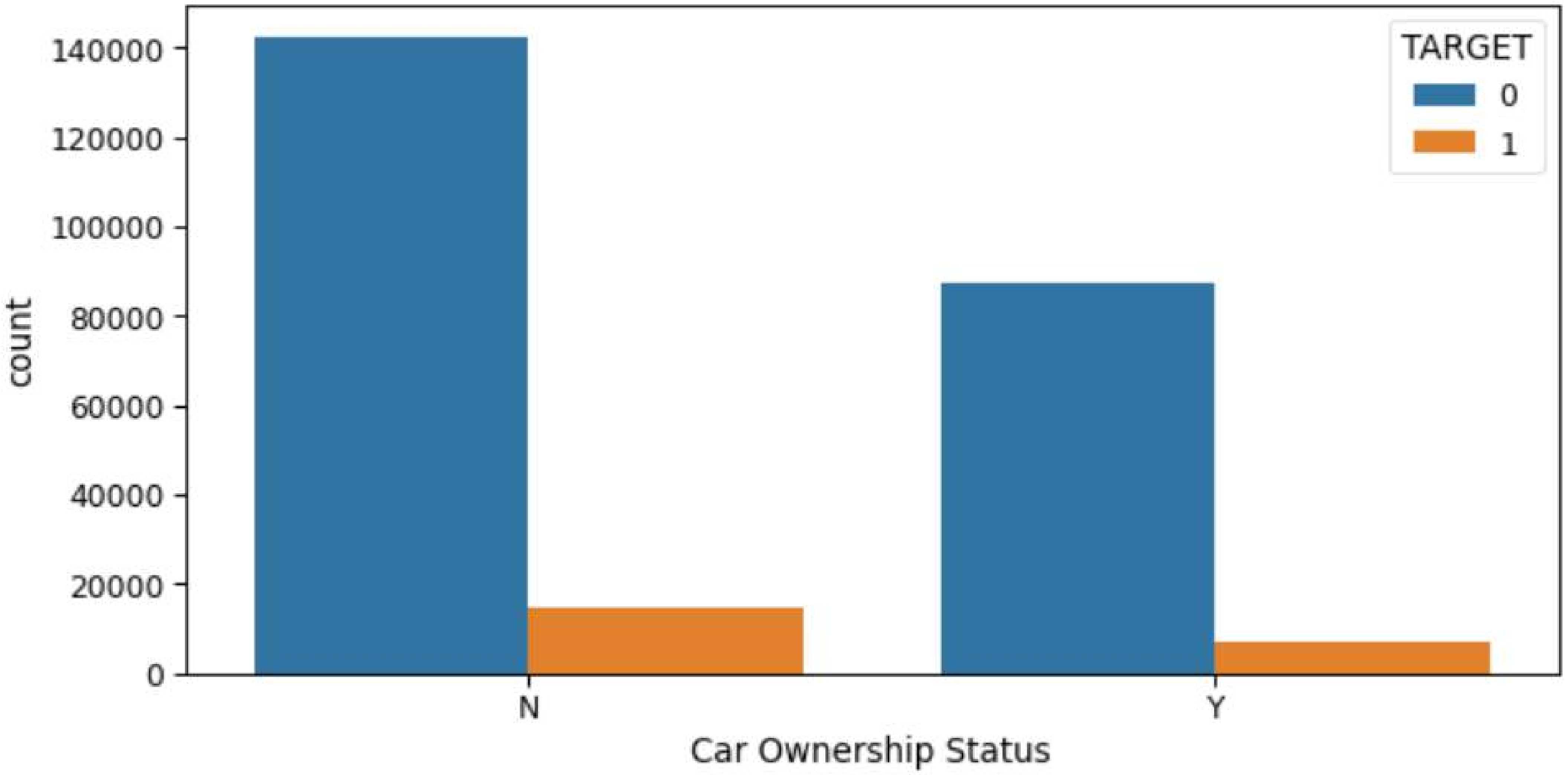


Data Preparation (EDA)

Car Ownership vs Target

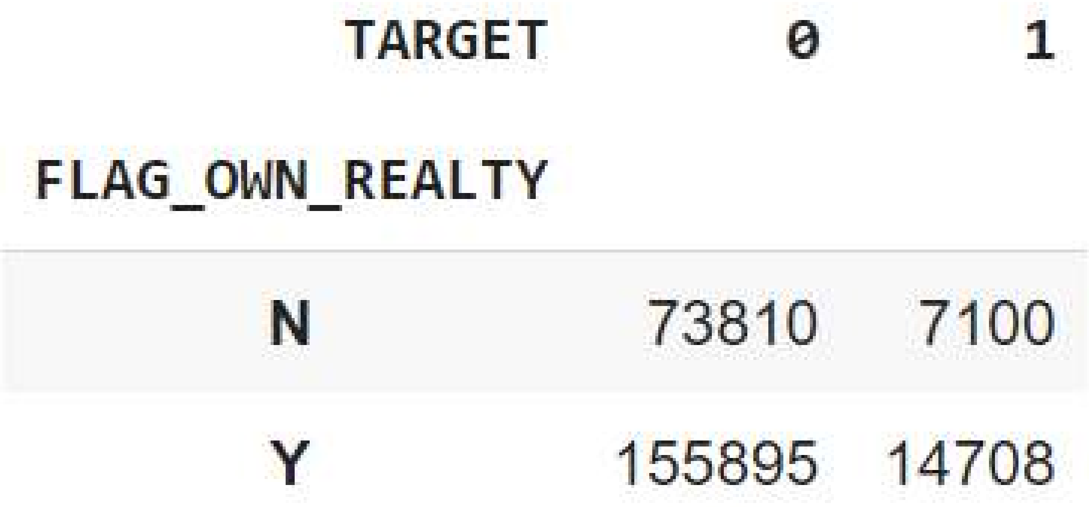


```
        TARGET      0      1
FLAG_OWN_CAR
             N  61299   6388
             Y  37235   3030
                    TARGET
FLAG_OWN_CAR
             N  0.094376
             Y  0.075251
```

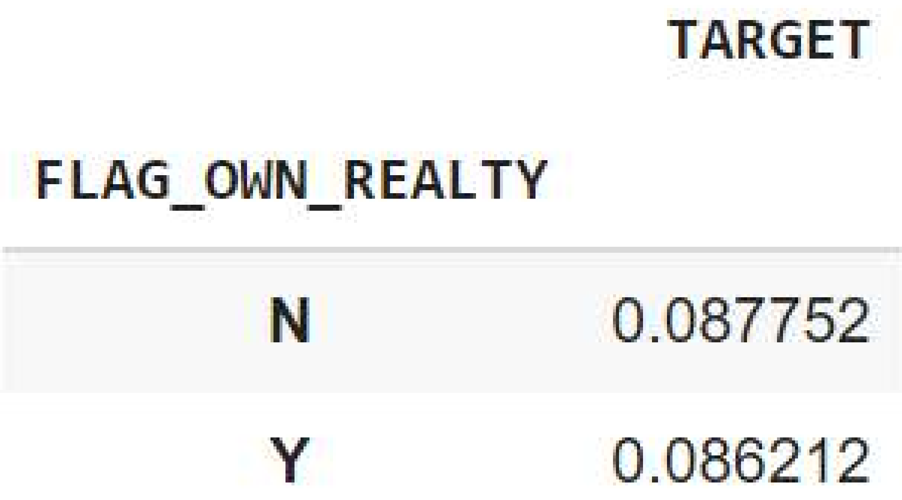


Data Preparation (EDA)

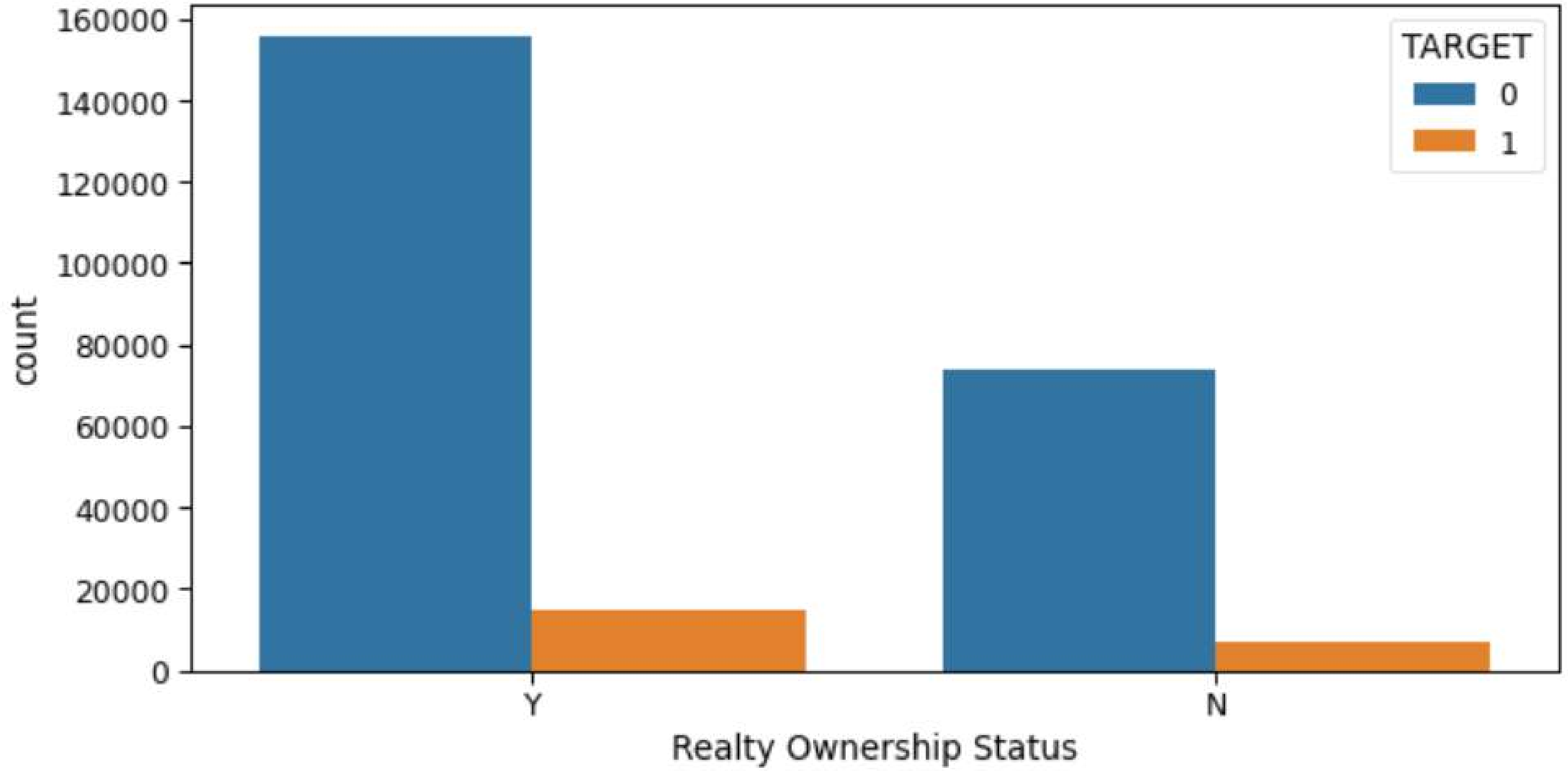
Realty Ownership vs Target



TARGET	0	1
FLAG_OWN_REALTY		
N	73810	7100
Y	155895	14708

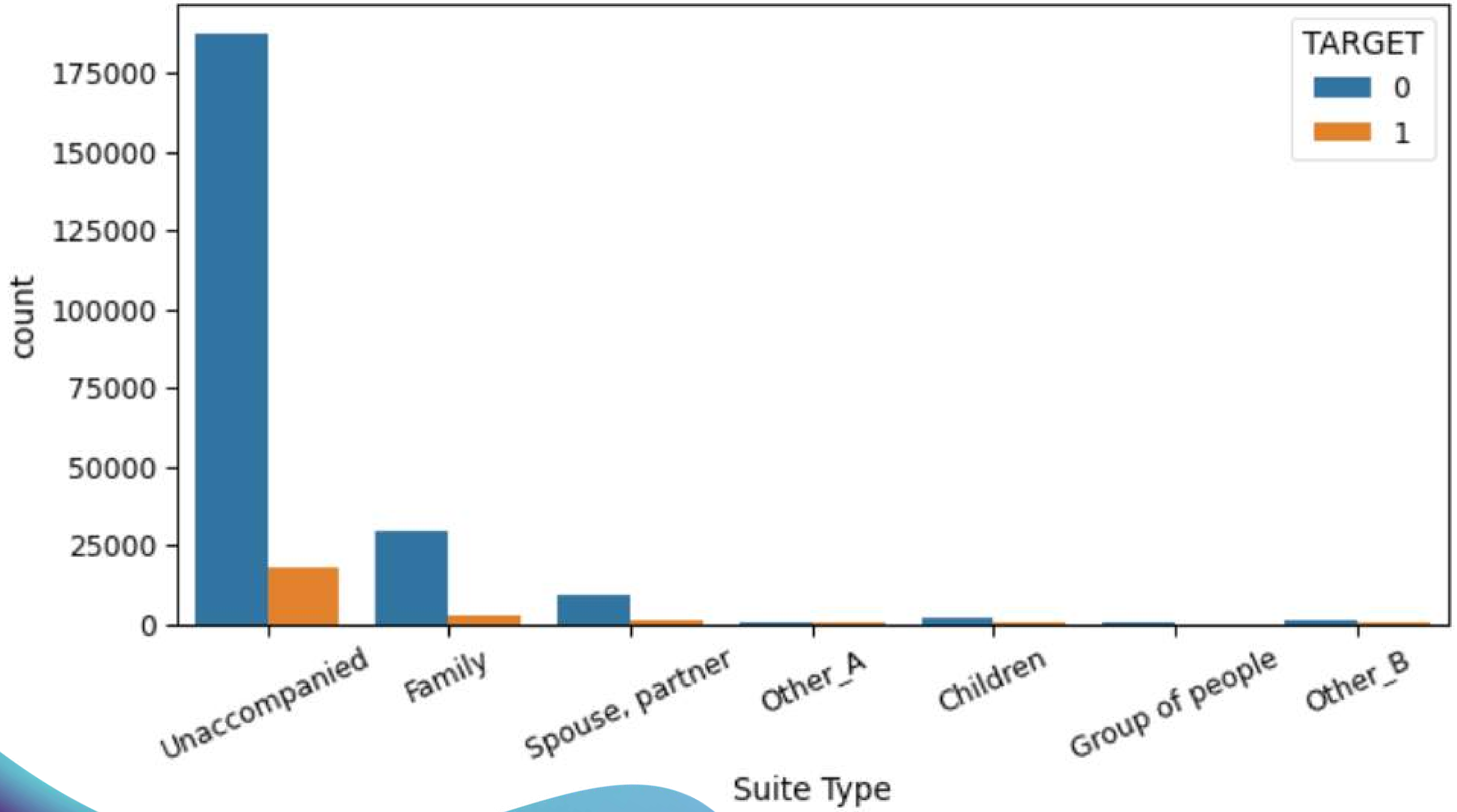


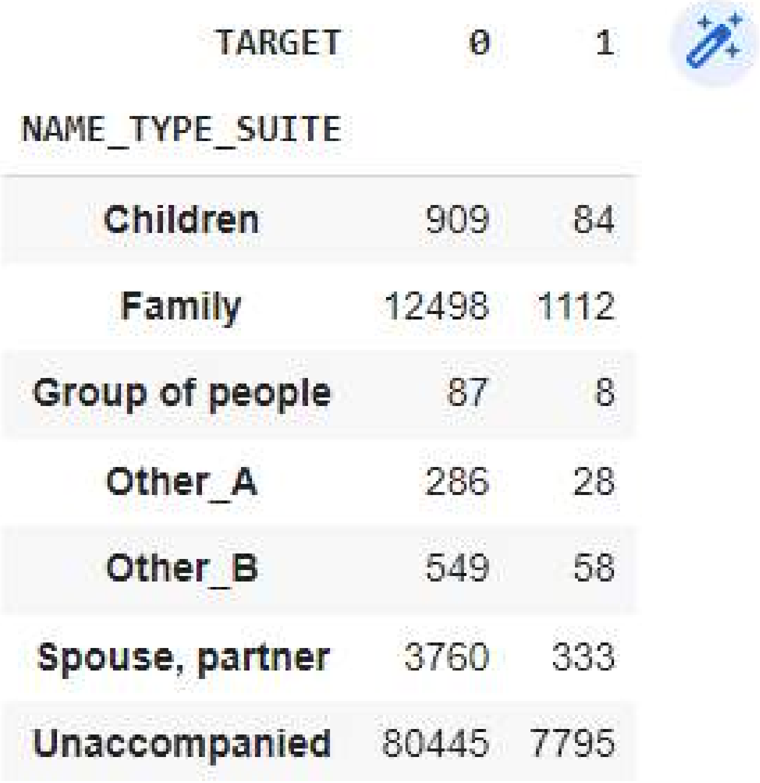
	TARGET
FLAG_OWN_REALTY	
N	0.087752
Y	0.086212



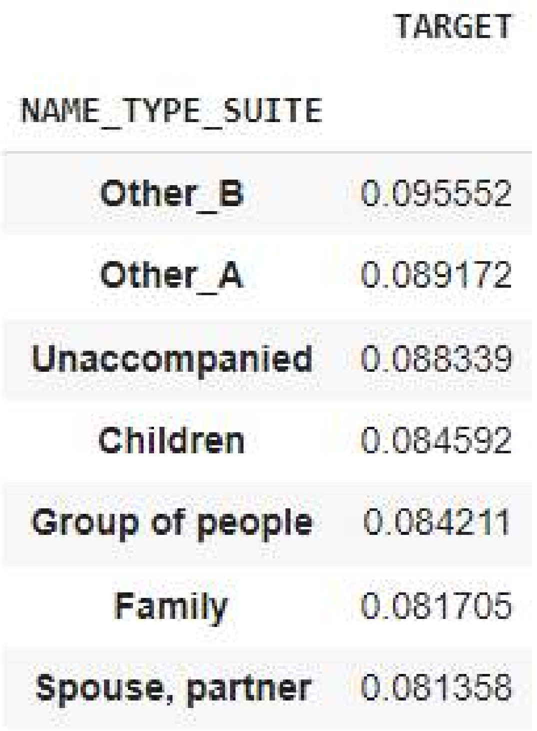
Data Preparation (EDA)

Suite Type vs Target





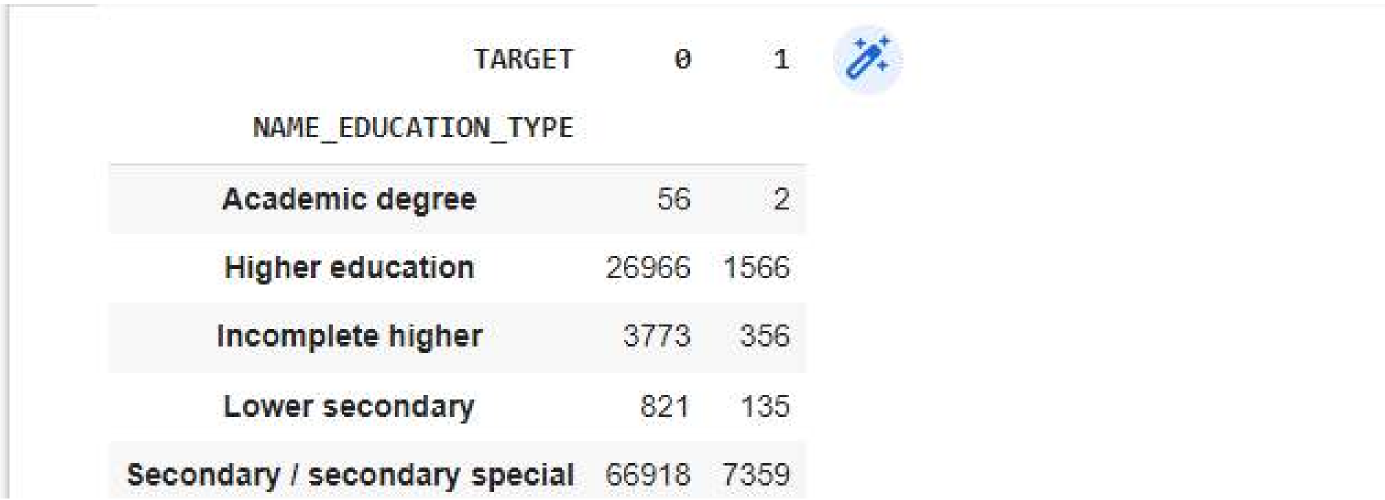
TARGET	0	1
NAME_TYPE_SUITE		
Children	909	84
Family	12498	1112
Group of people	87	8
Other_A	286	28
Other_B	549	58
Spouse, partner	3760	333
Unaccompanied	80445	7795



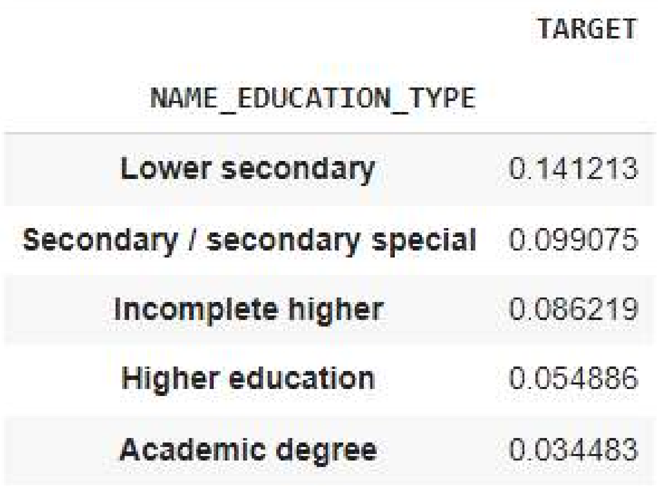
	TARGET
NAME_TYPE_SUITE	
Other_B	0.095552
Other_A	0.089172
Unaccompanied	0.088339
Children	0.084592
Group of people	0.084211
Family	0.081705
Spouse, partner	0.081358

Data Preparation (EDA)

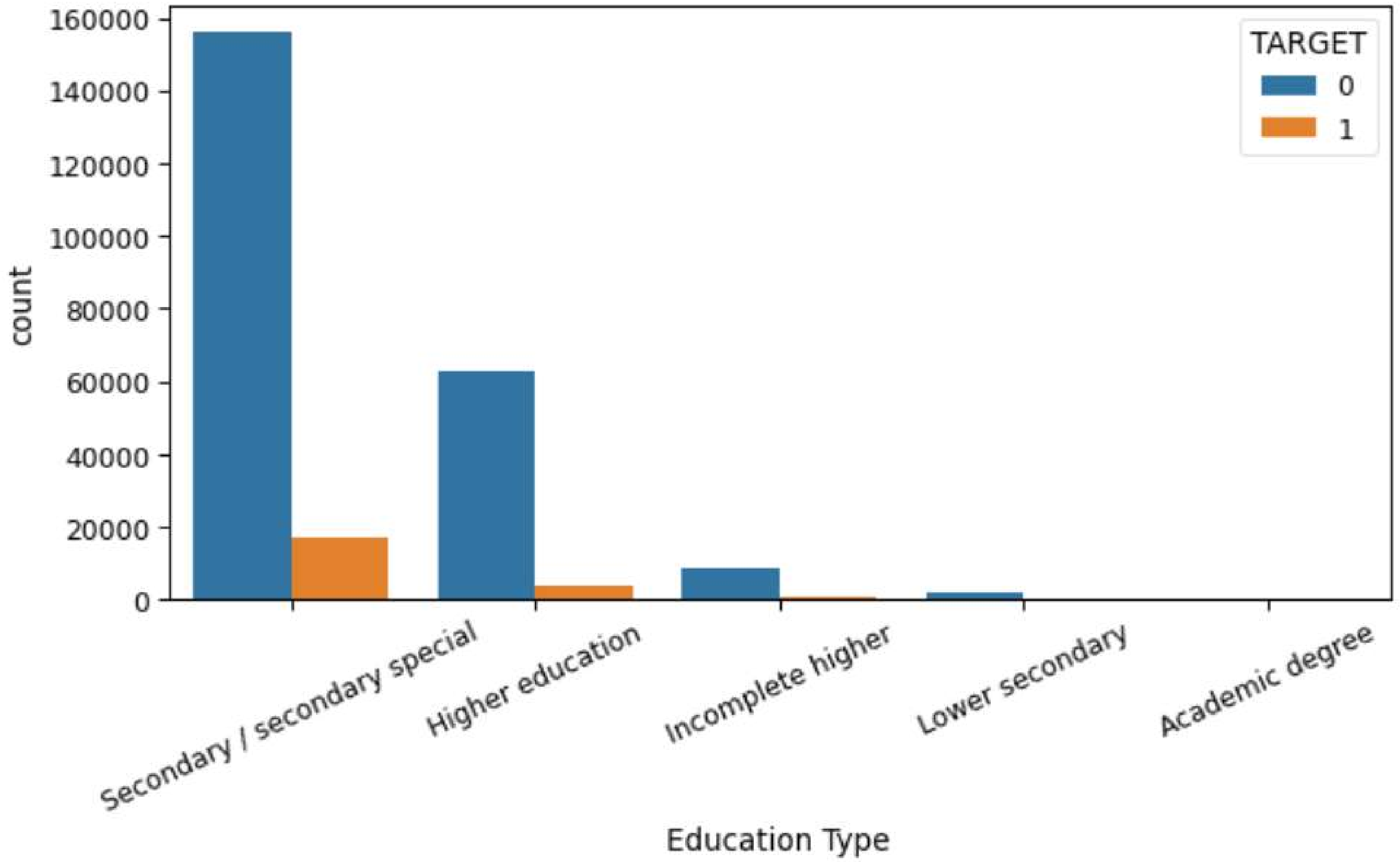
Education Type vs Target



TARGET	0	1
NAME_EDUCATION_TYPE		
Academic degree	56	2
Higher education	26966	1566
Incomplete higher	3773	356
Lower secondary	821	135
Secondary / secondary special	66918	7359



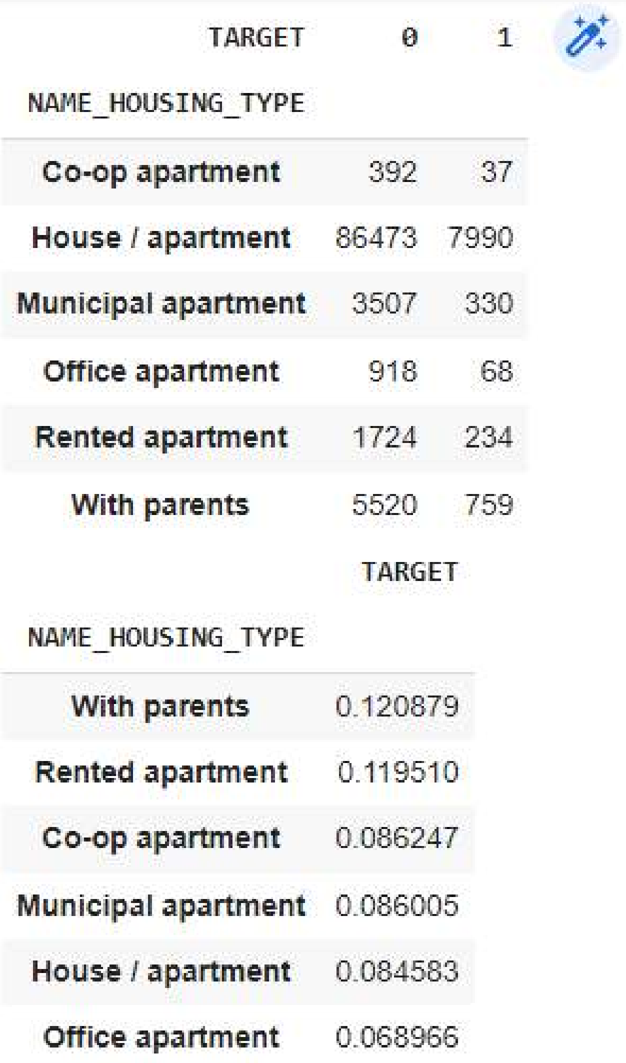
	TARGET
NAME_EDUCATION_TYPE	
Lower secondary	0.141213
Secondary / secondary special	0.099075
Incomplete higher	0.086219
Higher education	0.054886
Academic degree	0.034483



Pada hasil menunjukkan persentase dari nasabah yang lancar membayar dengan variabel pendidikan. Tabel dibagian bawah adalah persentase hasil tabel dibagian atas. Dimana 0 berarti dapat membayar, sedangkan 1 berarti tidak dapat membayar.

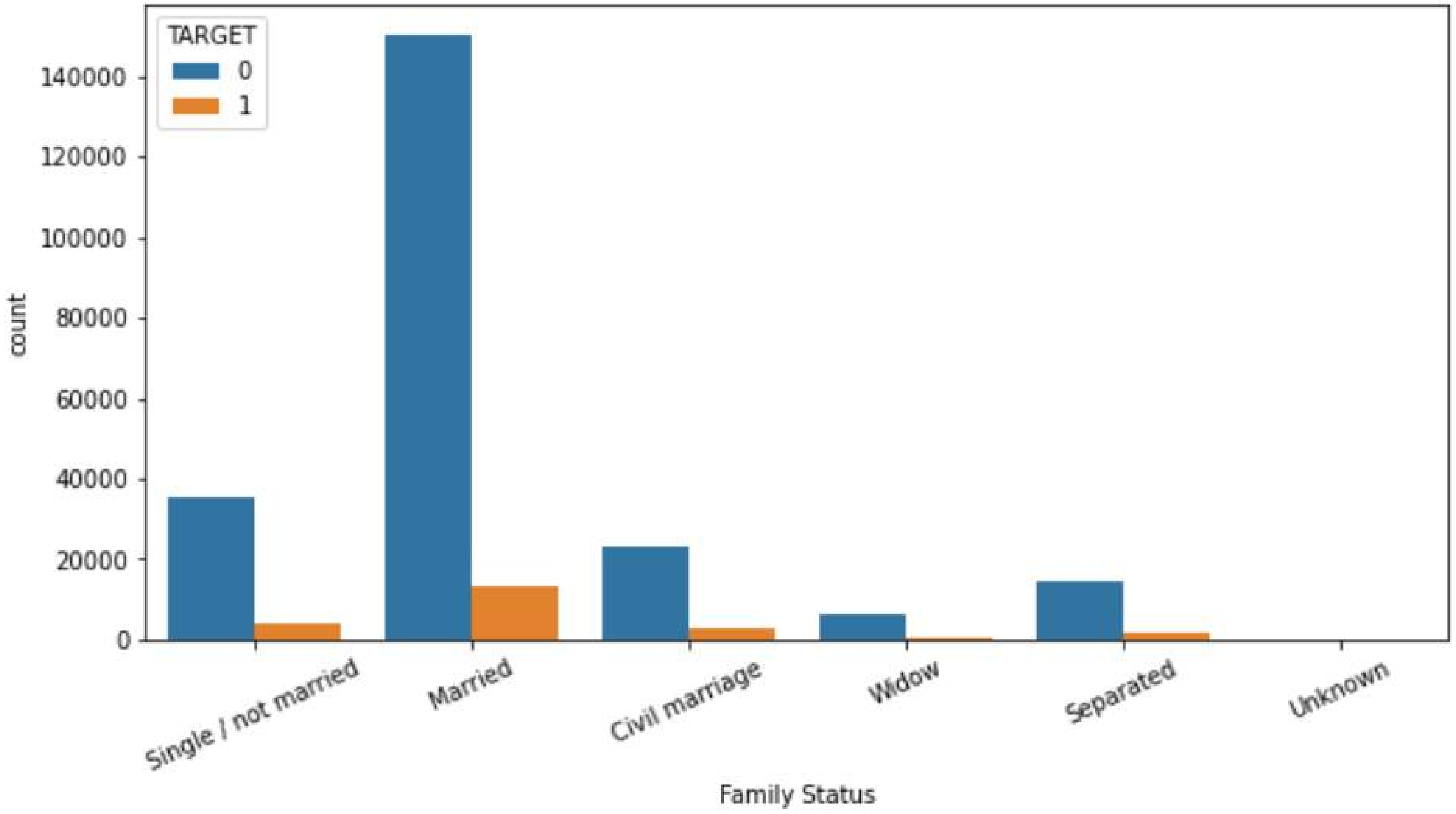
Data Preparation (EDA)

Family Status vs Target



TARGET	0	1
NAME_HOUSING_TYPE		
Co-op apartment	392	37
House / apartment	86473	7990
Municipal apartment	3507	330
Office apartment	918	68
Rented apartment	1724	234
With parents	5520	759

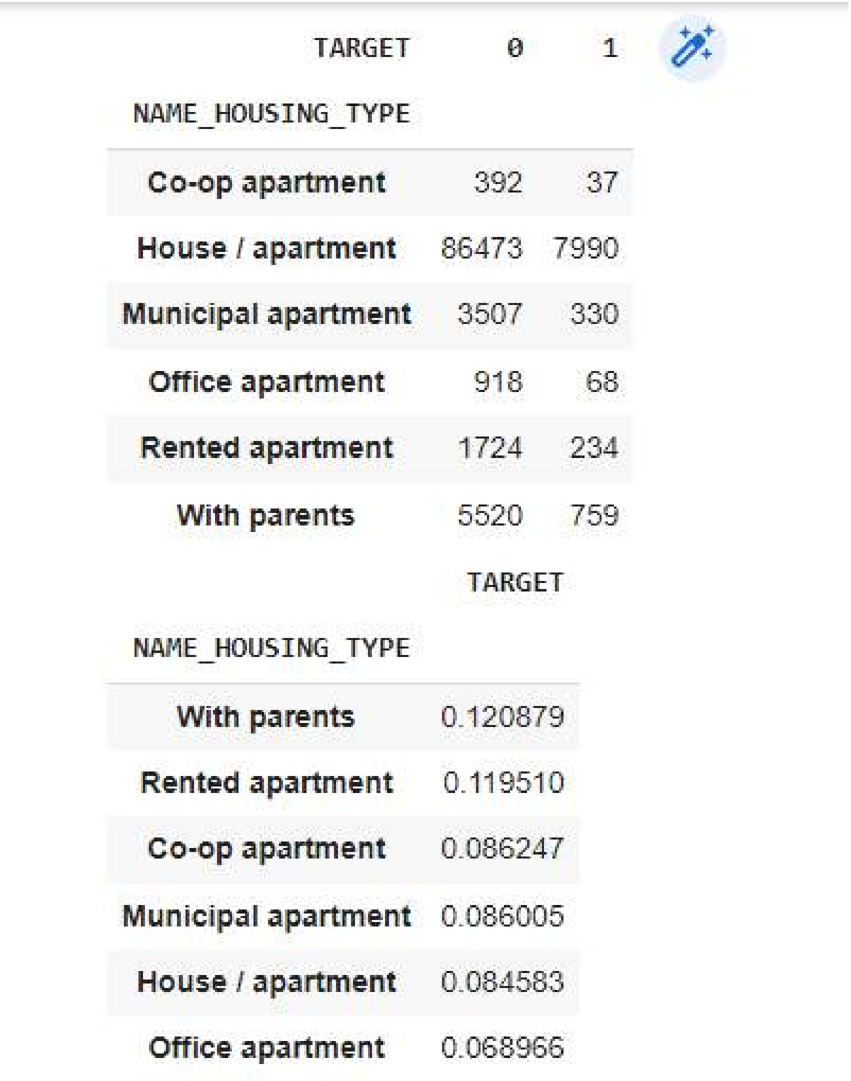
	TARGET
NAME_HOUSING_TYPE	
With parents	0.120879
Rented apartment	0.119510
Co-op apartment	0.086247
Municipal apartment	0.086005
House / apartment	0.084583
Office apartment	0.068966



Pada data menunjukkan persentase dari nasabah yang lancar membayar dengan variabel status keluarga. Tabel dibagian bawah adalah persentase hasil tabel dibagian atas. Dimana 0 berarti dapat membayar, sedangkan 1 berarti tidak dapat membayar.

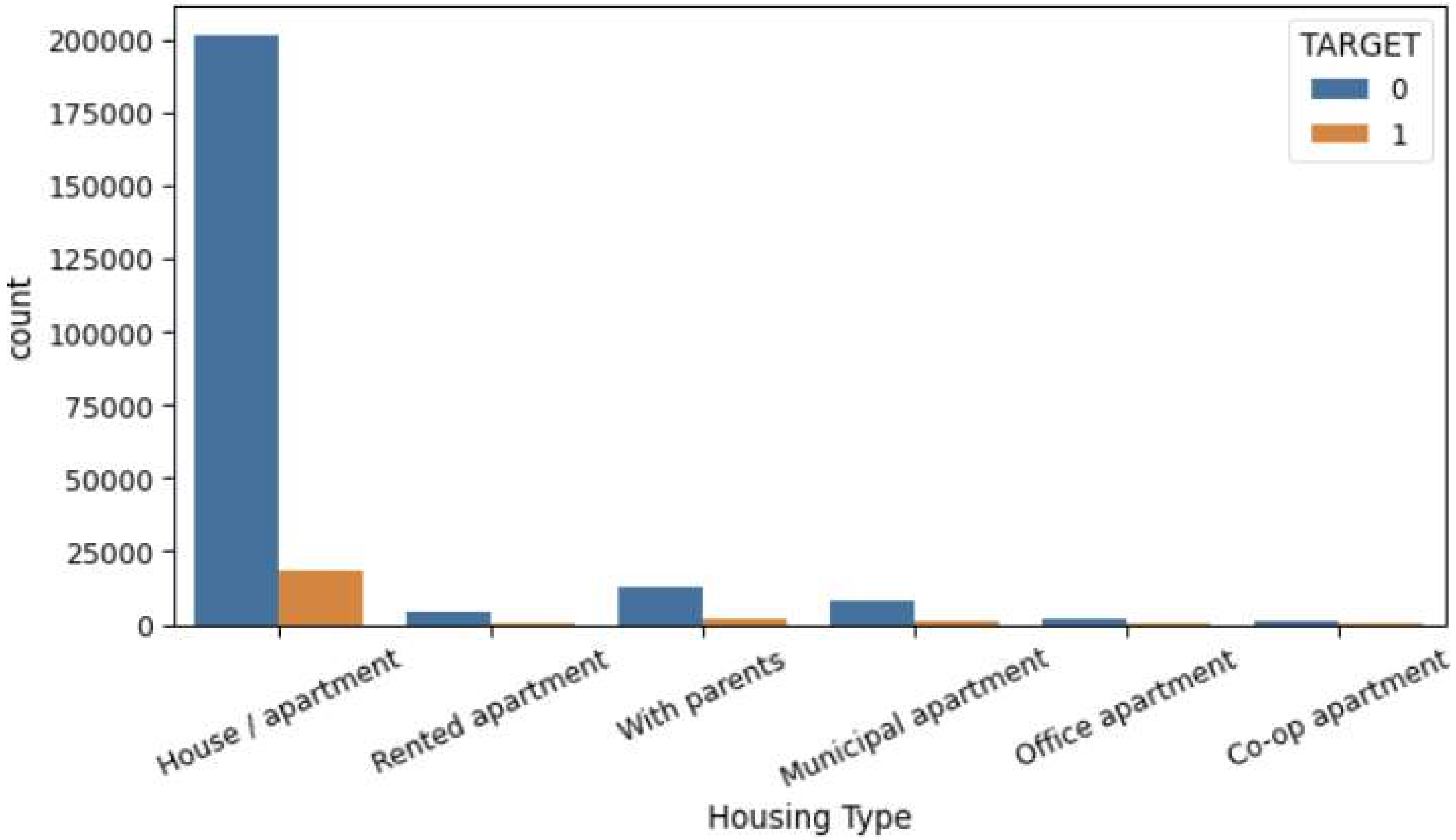
Data Preparation (EDA)

Housing Type vs Target



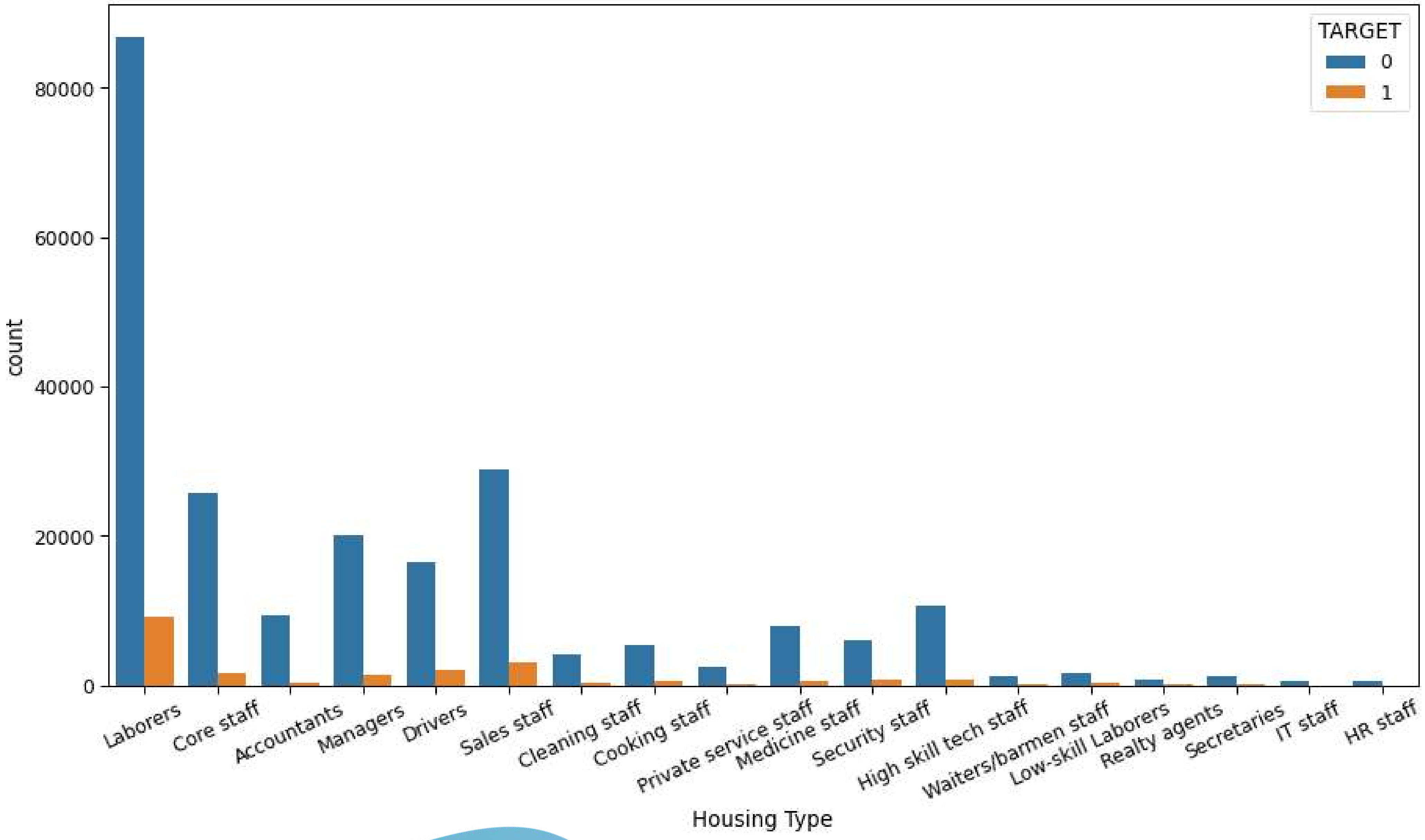
TARGET	0	1
NAME_HOUSING_TYPE		
Co-op apartment	392	37
House / apartment	86473	7990
Municipal apartment	3507	330
Office apartment	918	68
Rented apartment	1724	234
With parents	5520	759

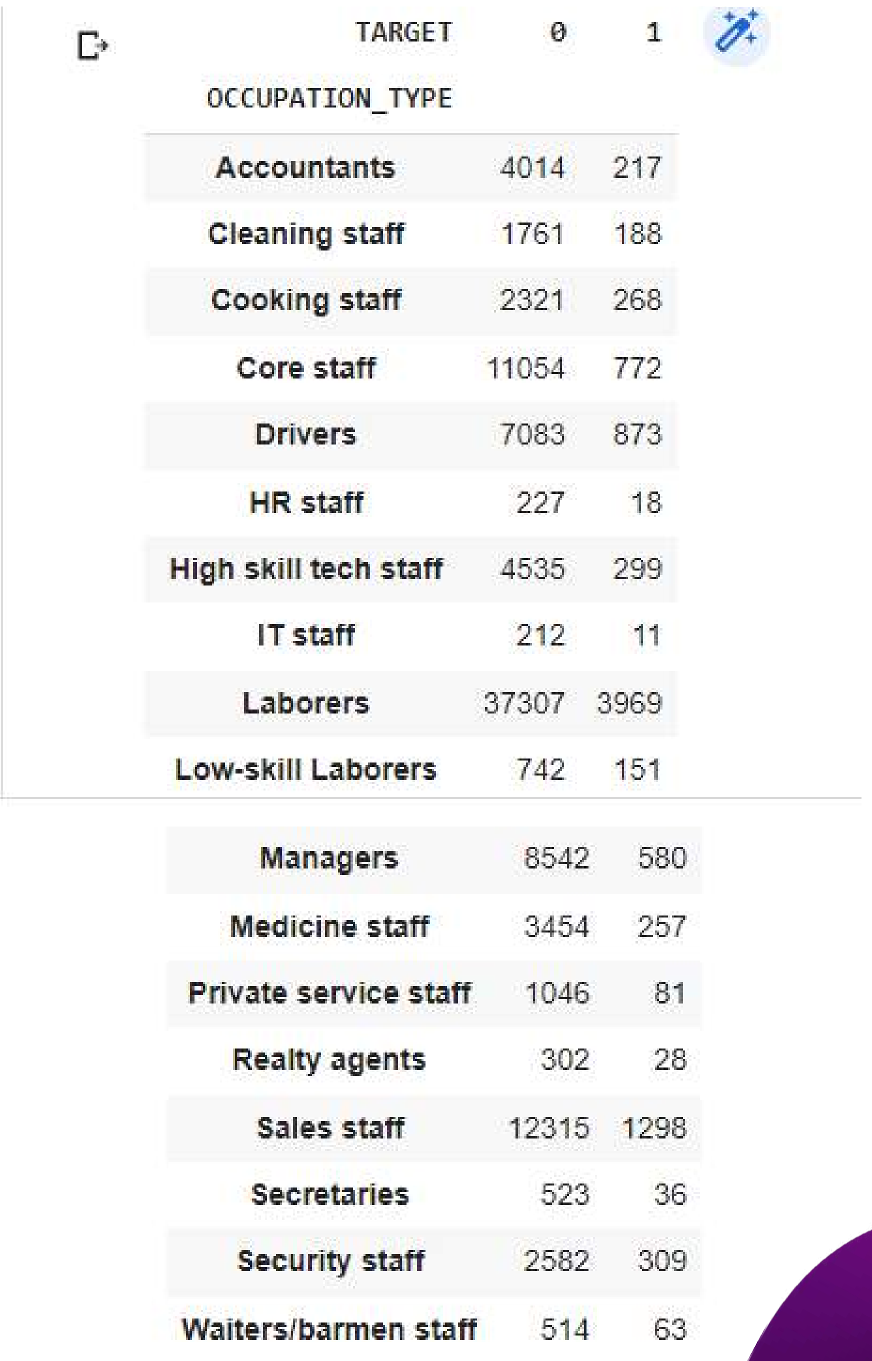
	TARGET
NAME_HOUSING_TYPE	
With parents	0.120879
Rented apartment	0.119510
Co-op apartment	0.086247
Municipal apartment	0.086005
House / apartment	0.084583
Office apartment	0.068966



Data Preparation (EDA)

Ocupation Type vs Target

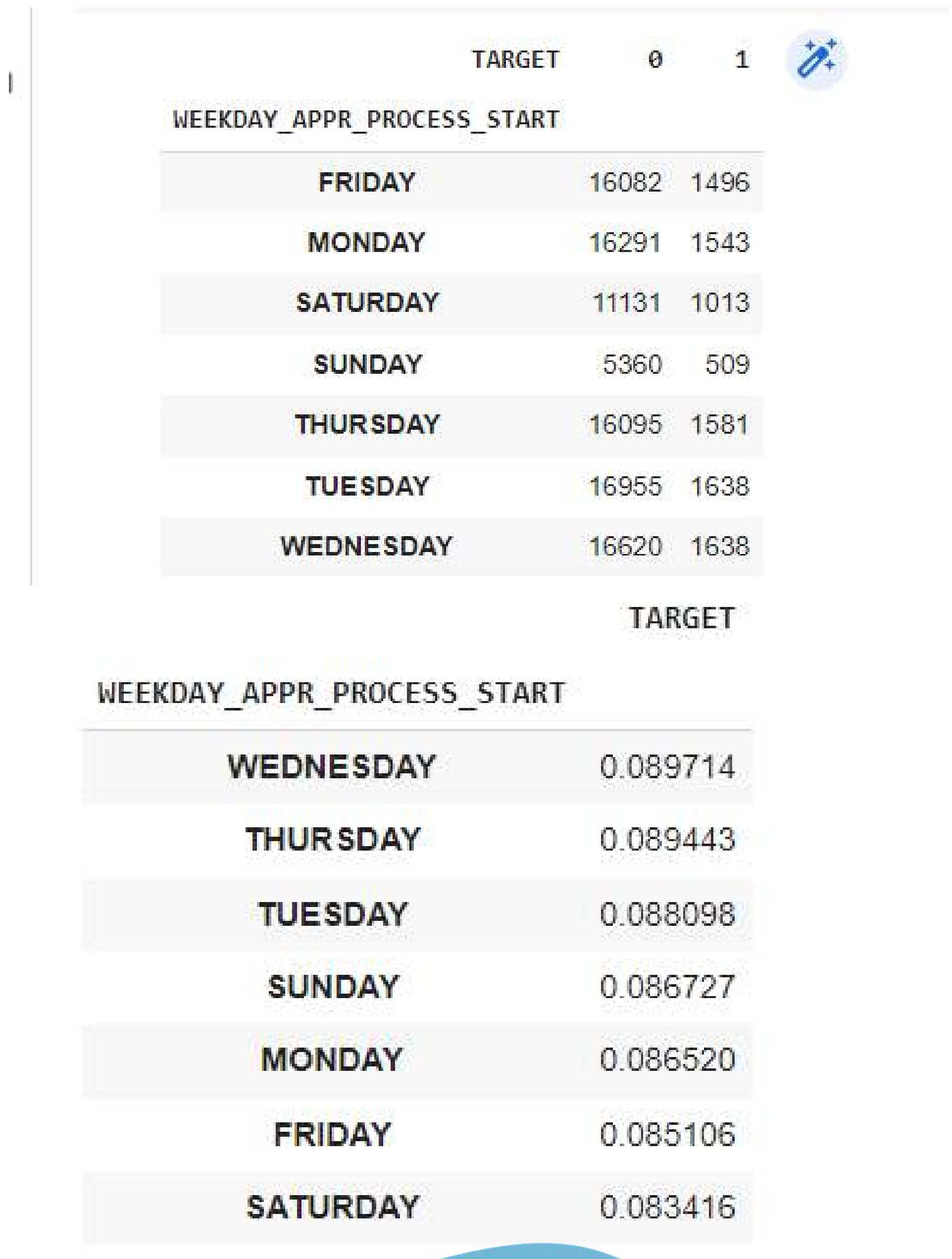




TARGET	0	1
OCCUPATION_TYPE		
Accountants	4014	217
Cleaning staff	1761	188
Cooking staff	2321	268
Core staff	11054	772
Drivers	7083	873
HR staff	227	18
High skill tech staff	4535	299
IT staff	212	11
Laborers	37307	3969
Low-skill Laborers	742	151
Managers	8542	580
Medicine staff	3454	257
Private service staff	1046	81
Realty agents	302	28
Sales staff	12315	1298
Secretaries	523	36
Security staff	2582	309
Waiters/barmen staff	514	63

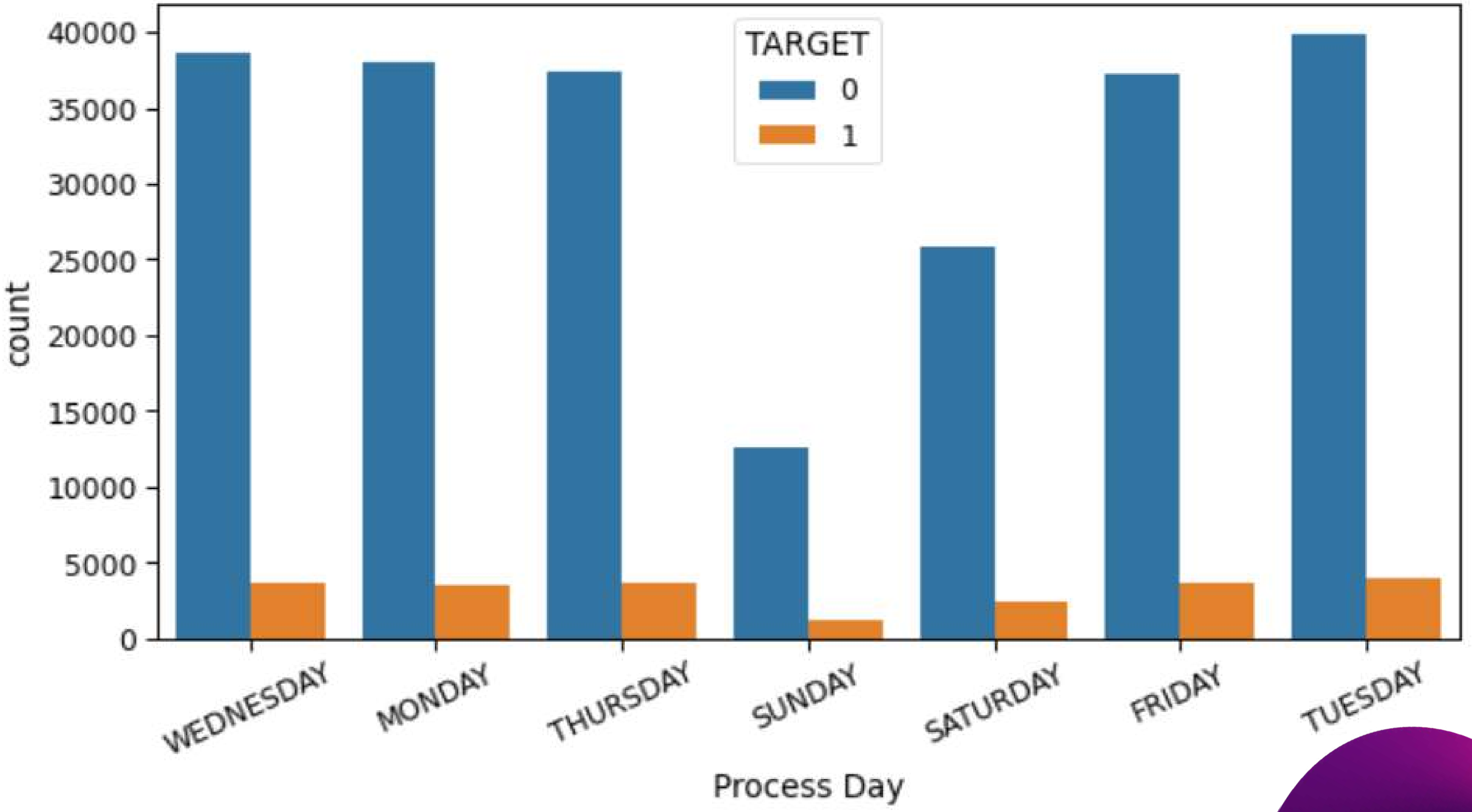
Data Preparation (EDA)

Process Day vs Target



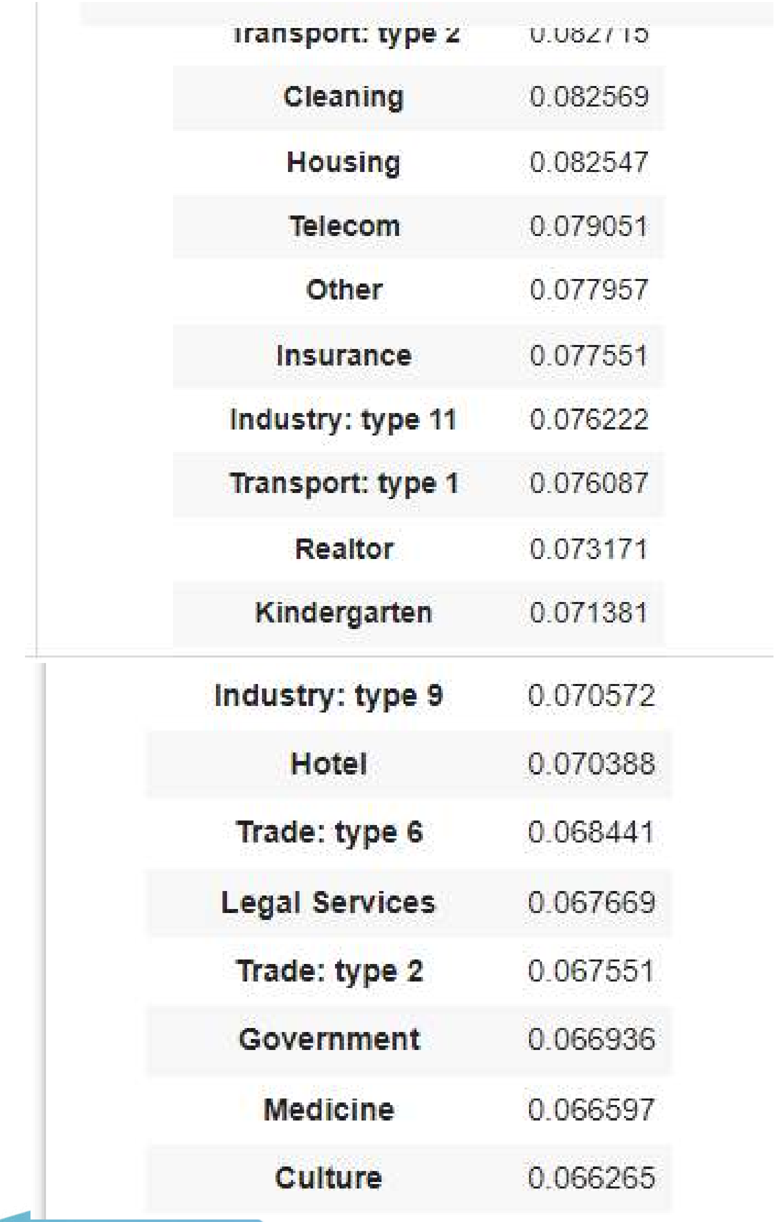
TARGET	0	1
WEEKDAY_APPR_PROCESS_START		
FRIDAY	16082	1496
MONDAY	16291	1543
SATURDAY	11131	1013
SUNDAY	5360	509
THURSDAY	16095	1581
TUESDAY	16955	1638
WEDNESDAY	16620	1638

	TARGET
WEEKDAY_APPR_PROCESS_START	
WEDNESDAY	0.089714
THURSDAY	0.089443
TUESDAY	0.088098
SUNDAY	0.086727
MONDAY	0.086520
FRIDAY	0.085106
SATURDAY	0.083416

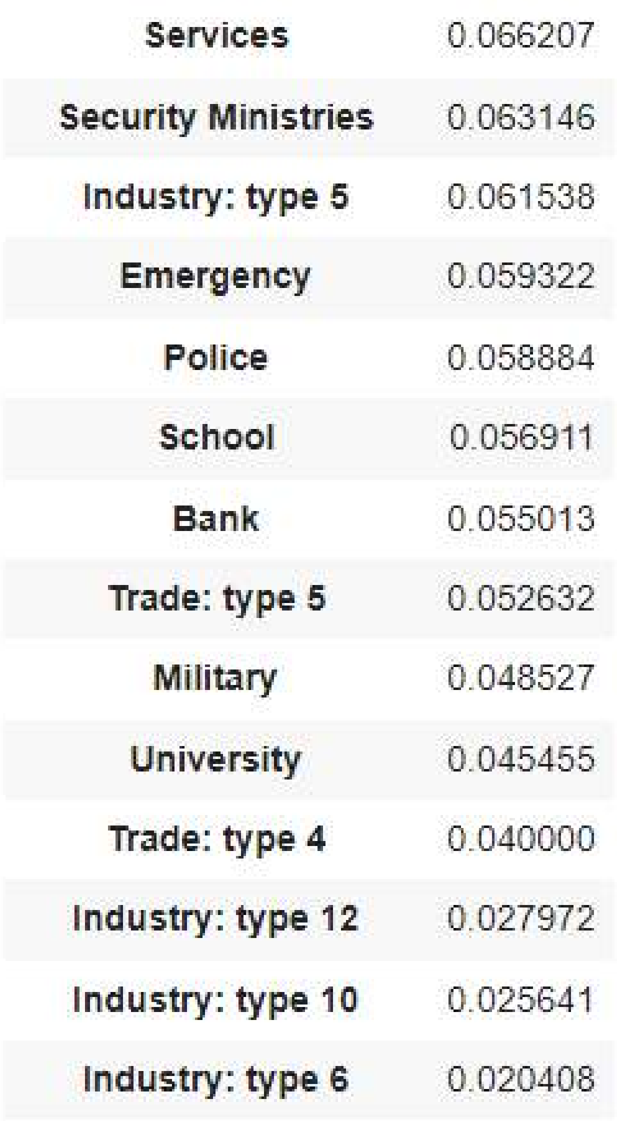


Data Preparation (EDA)

Organization Type vs Target



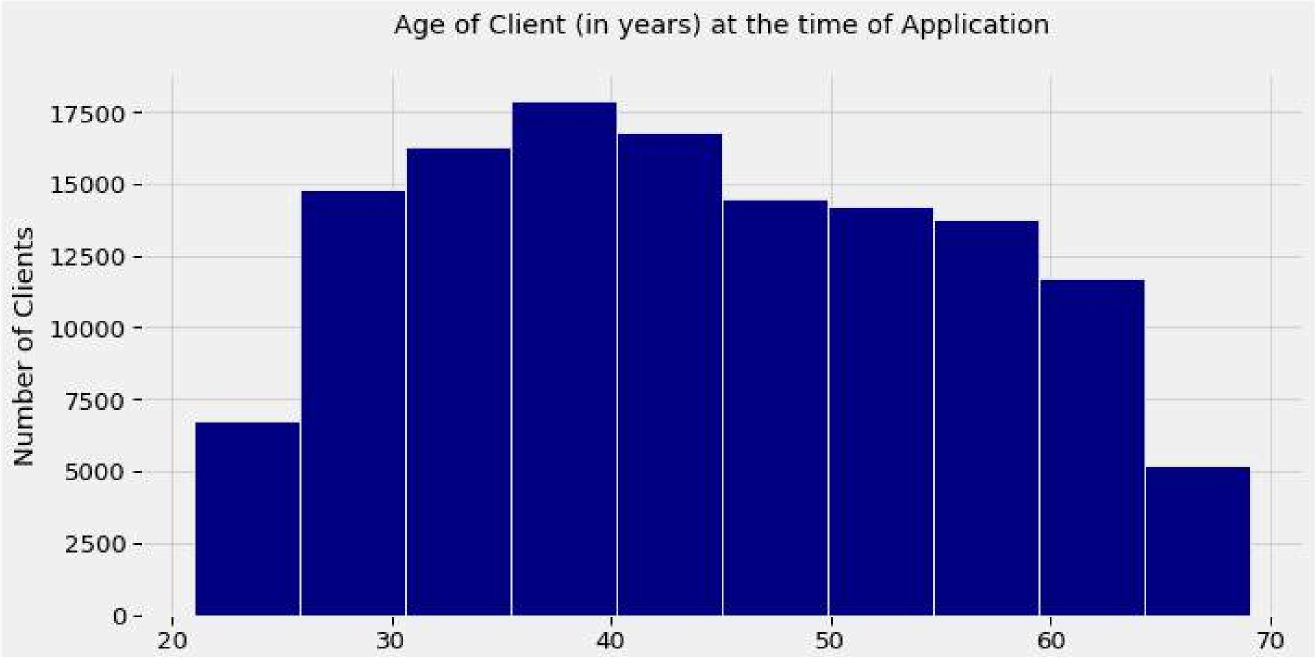
Transport: type 2	0.082715
Cleaning	0.082569
Housing	0.082547
Telecom	0.079051
Other	0.077957
Insurance	0.077551
Industry: type 11	0.076222
Transport: type 1	0.076087
Realtor	0.073171
Kindergarten	0.071381
Industry: type 9	0.070572
Hotel	0.070388
Trade: type 6	0.068441
Legal Services	0.067669
Trade: type 2	0.067551
Government	0.066936
Medicine	0.066597
Culture	0.066265



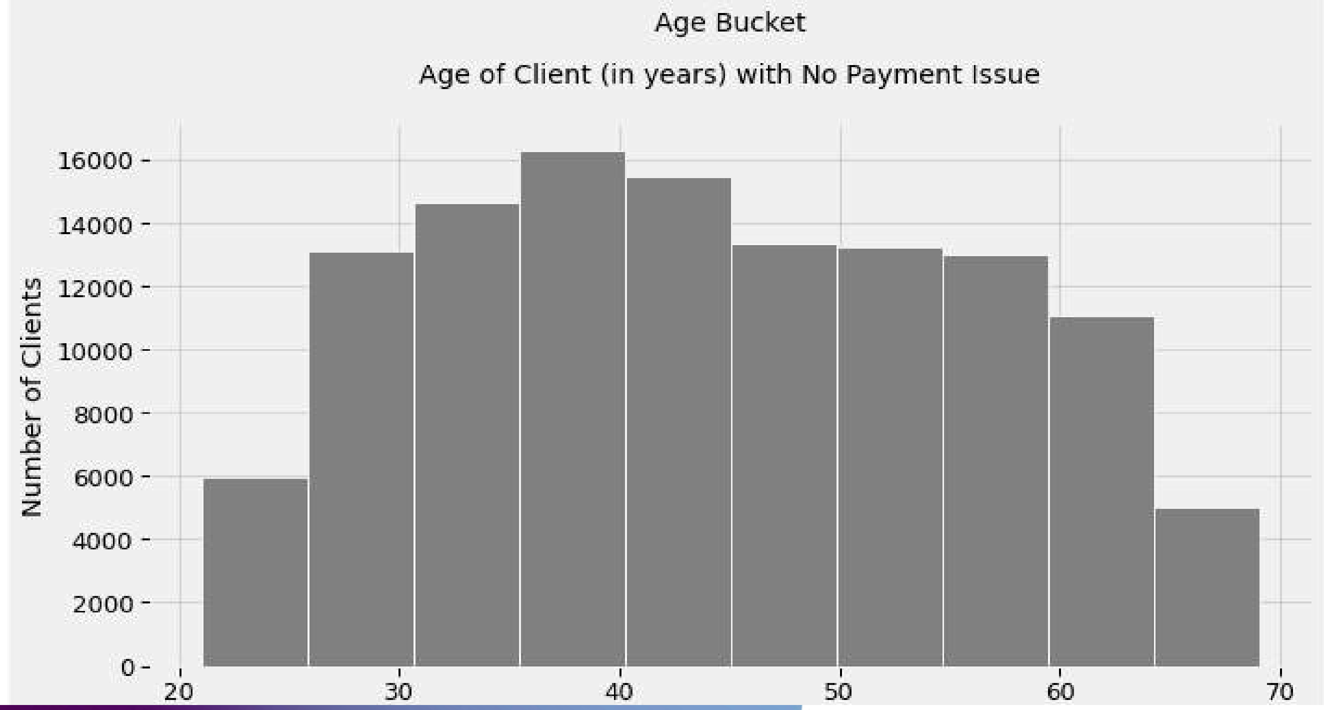
Services	0.066207
Security Ministries	0.063146
Industry: type 5	0.061538
Emergency	0.059322
Police	0.058884
School	0.056911
Bank	0.055013
Trade: type 5	0.052632
Military	0.048527
University	0.045455
Trade: type 4	0.040000
Industry: type 12	0.027972
Industry: type 10	0.025641
Industry: type 6	0.020408

Data Preparation (EDA)

Age vs Target

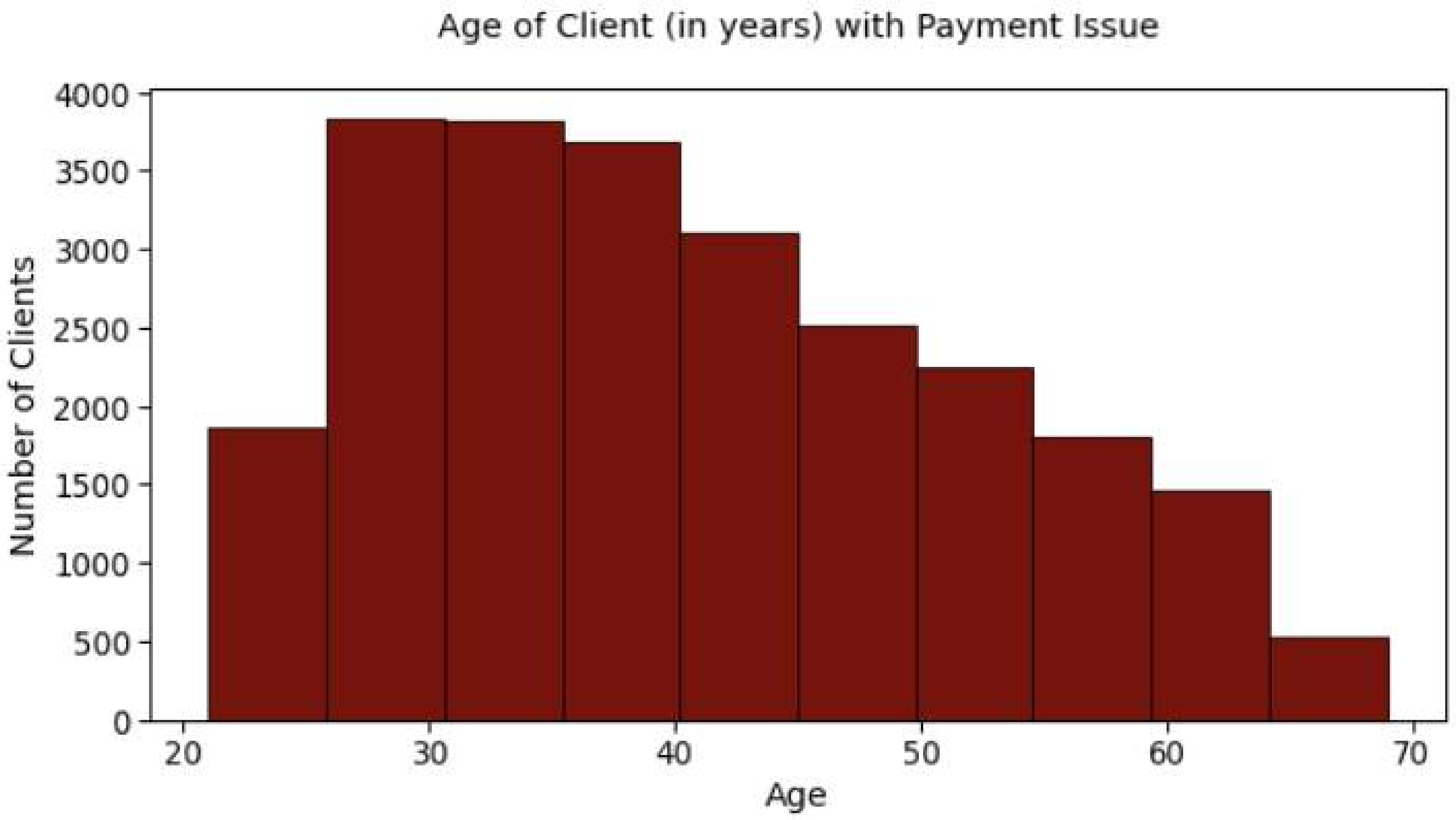


Age of Client (in years) at the time of Application



Age Bucket

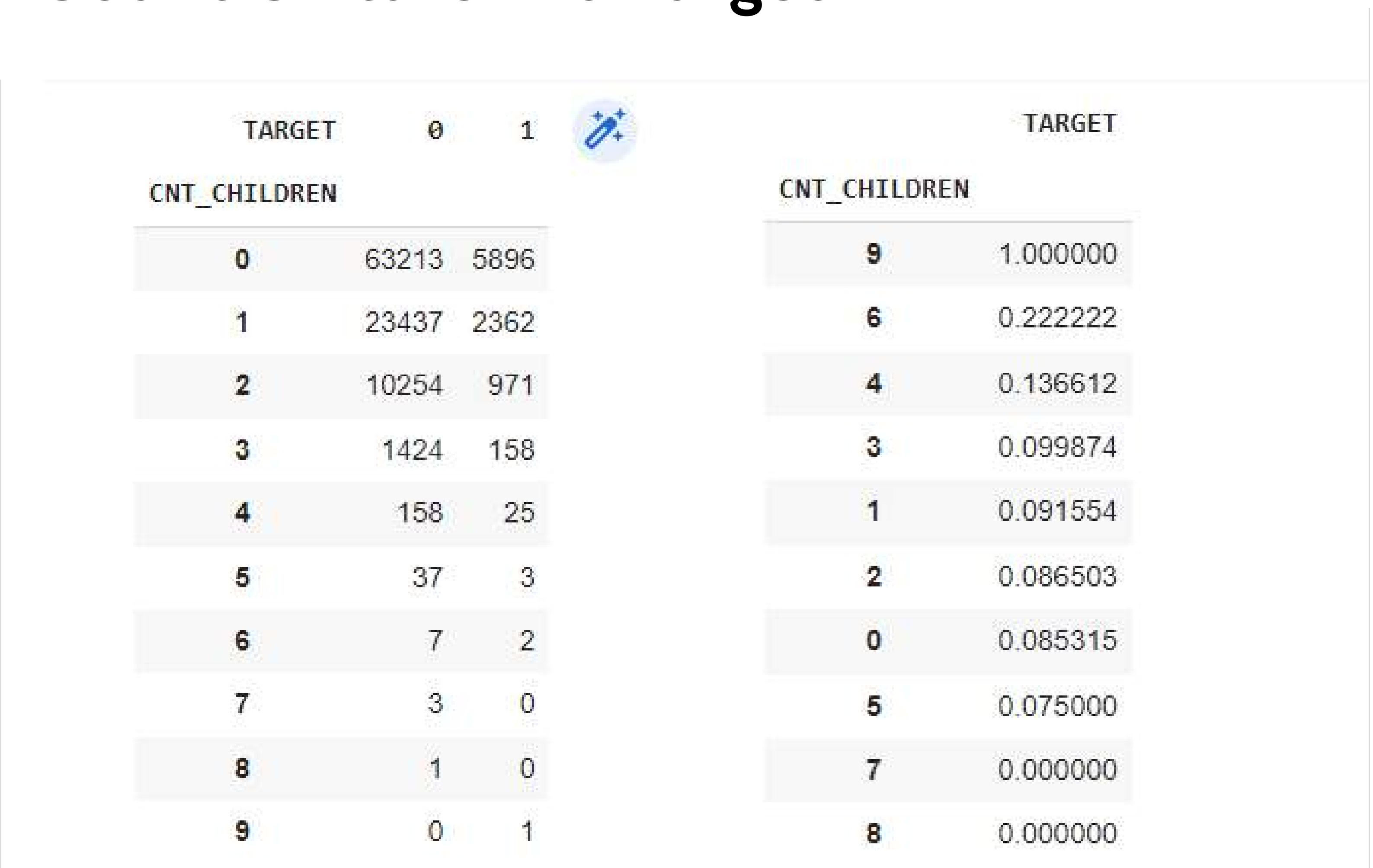
Age of Client (in years) with No Payment Issue



Age of Client (in years) with Payment Issue

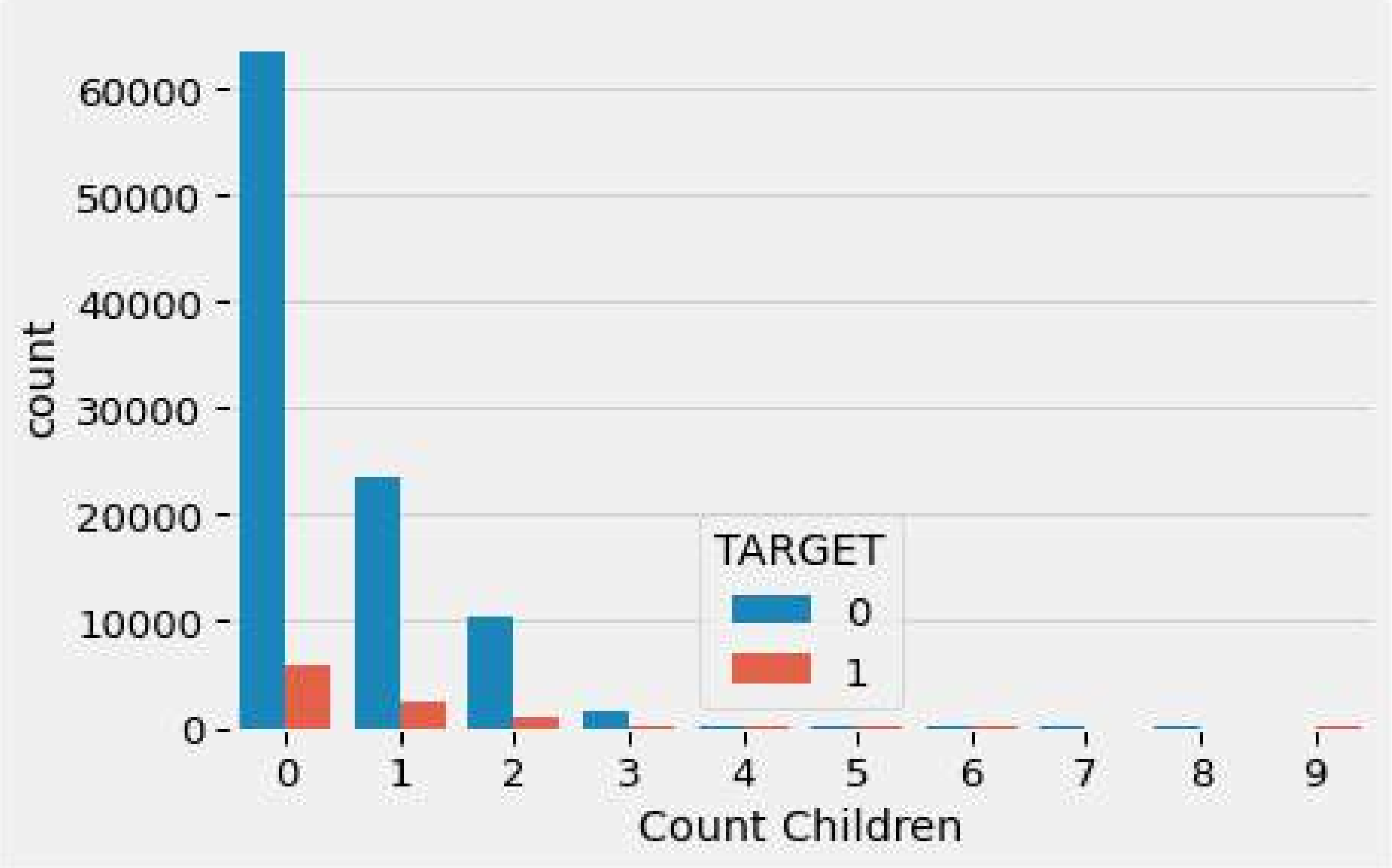
Data Preparation (EDA)

Count Children vs Target



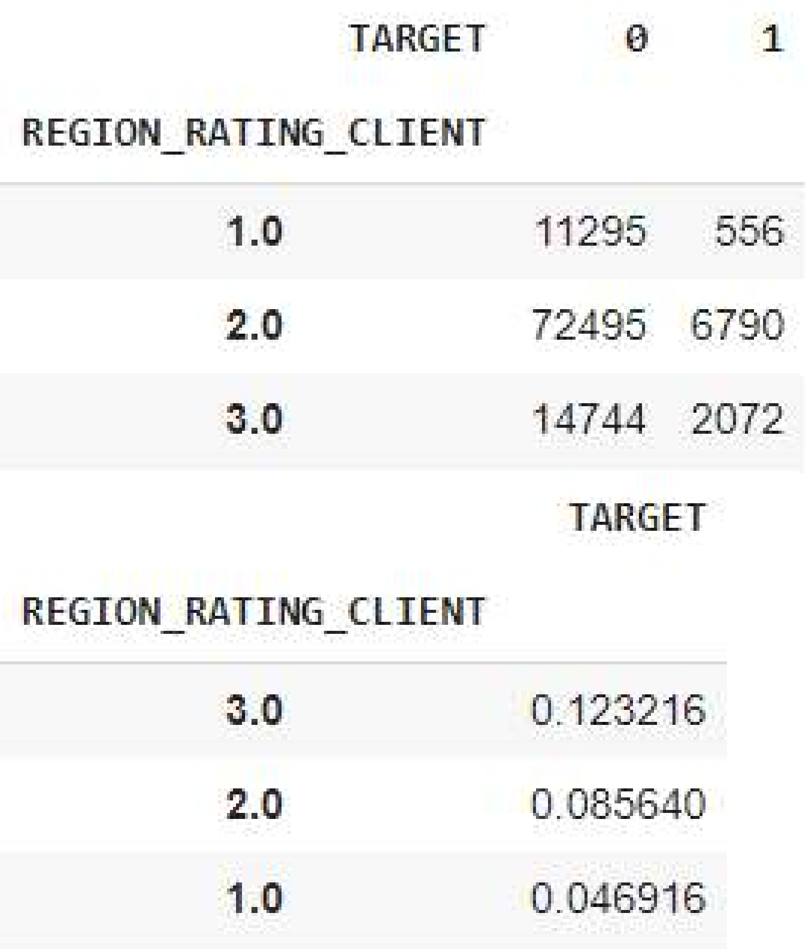
TARGET	0	1
CNT_CHILDREN		
0	63213	5896
1	23437	2362
2	10254	971
3	1424	158
4	158	25
5	37	3
6	7	2
7	3	0
8	1	0
9	0	1

	TARGET
CNT_CHILDREN	
9	1.000000
6	0.222222
4	0.136612
3	0.099874
1	0.091554
2	0.086503
0	0.085315
5	0.075000
7	0.000000
8	0.000000



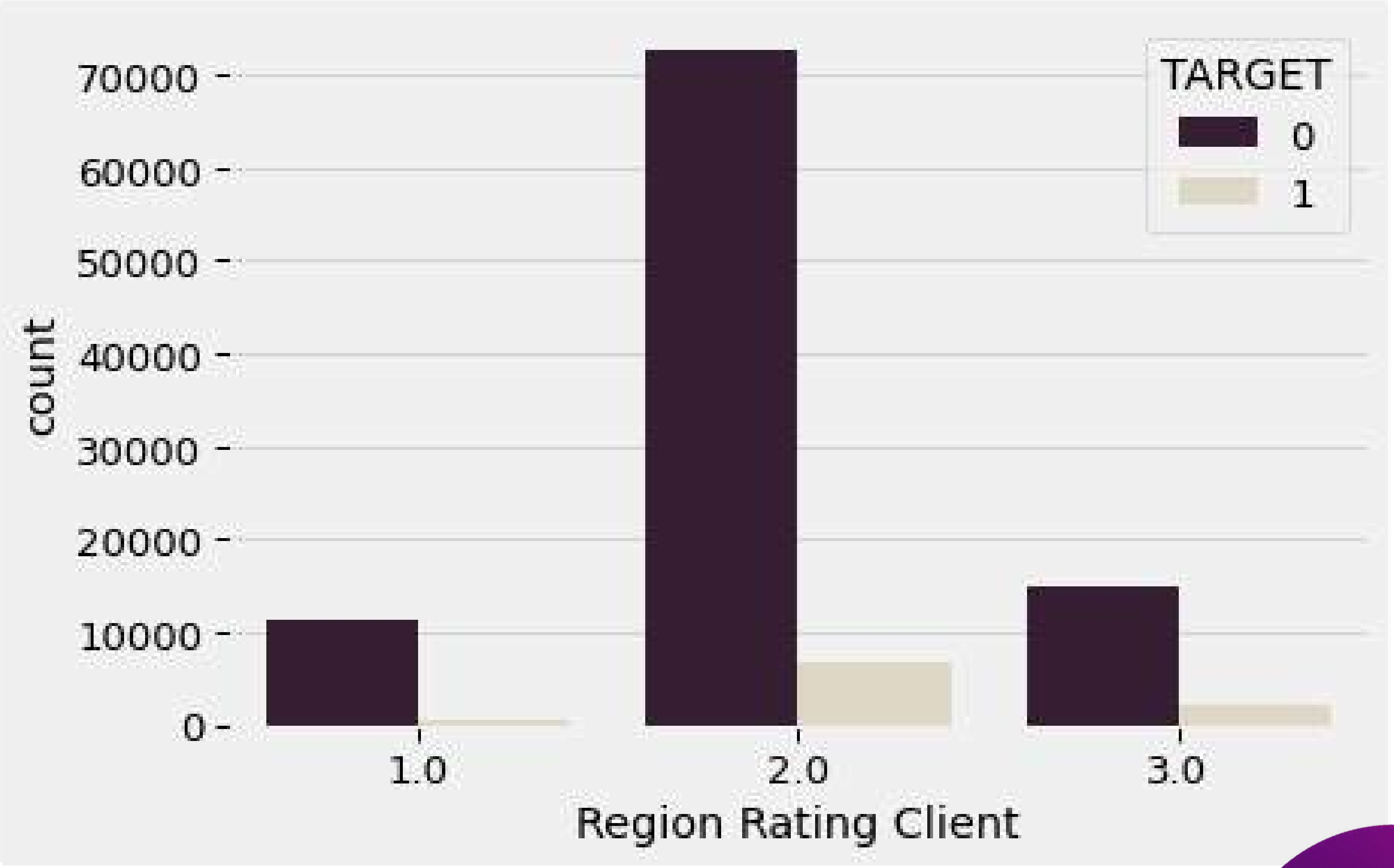
Data Preparation (EDA)

Region Rating Client vs Target



TARGET	0	1
REGION_RATING_CLIENT		
1.0	11295	556
2.0	72495	6790
3.0	14744	2072

	TARGET
REGION_RATING_CLIENT	
3.0	0.123216
2.0	0.085640
1.0	0.046916

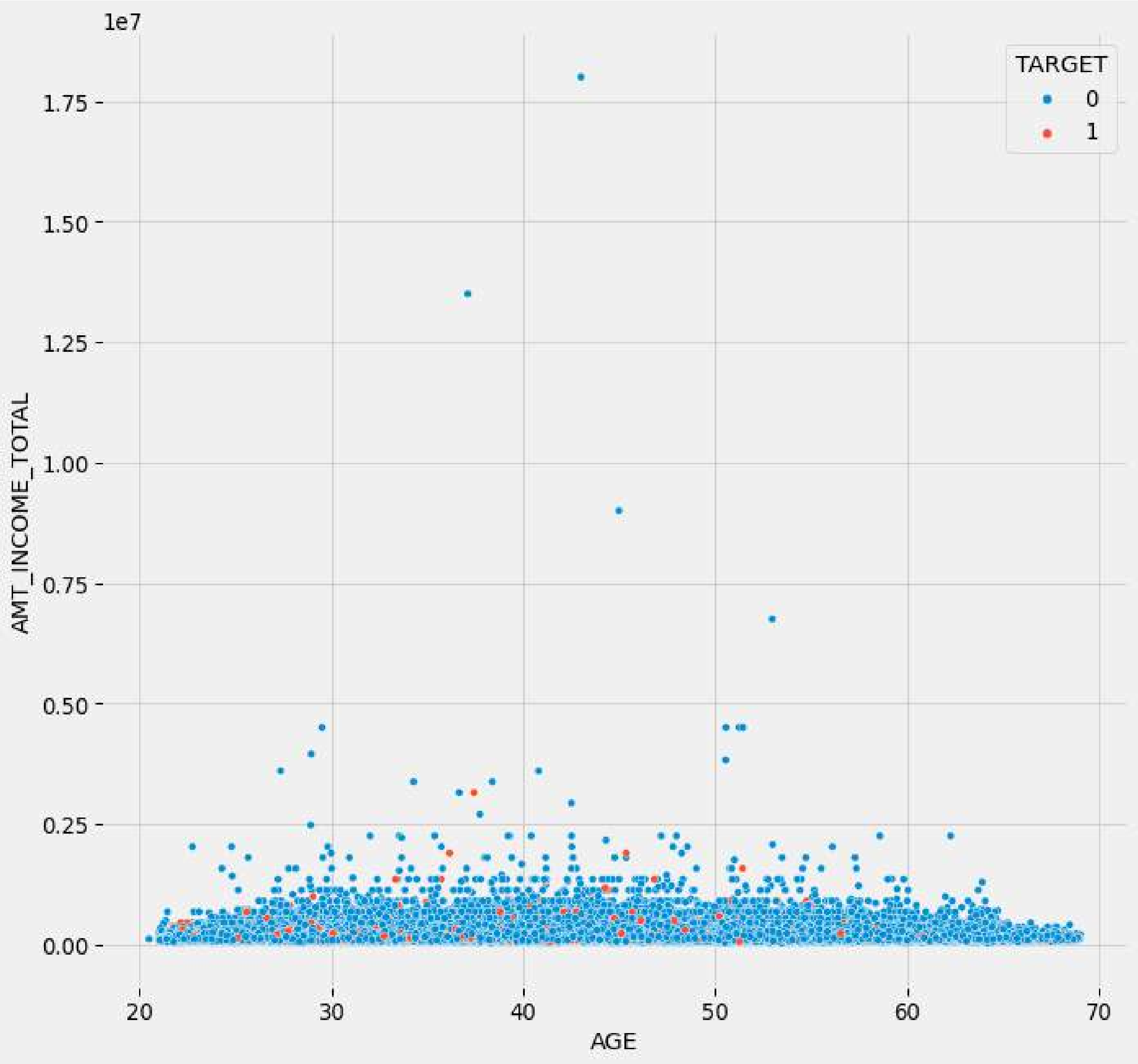


Data Preparation (EDA)

Multivariate Analysis

Analisis multivariat (MVA) didasarkan pada prinsip-prinsip statistik multivariat. Biasanya, MVA digunakan untuk mengatasi situasi di mana beberapa pengukuran dilakukan pada setiap unit eksperimental dan hubungan antara pengukuran ini dan strukturnya penting.

Data Preparation (EDA)

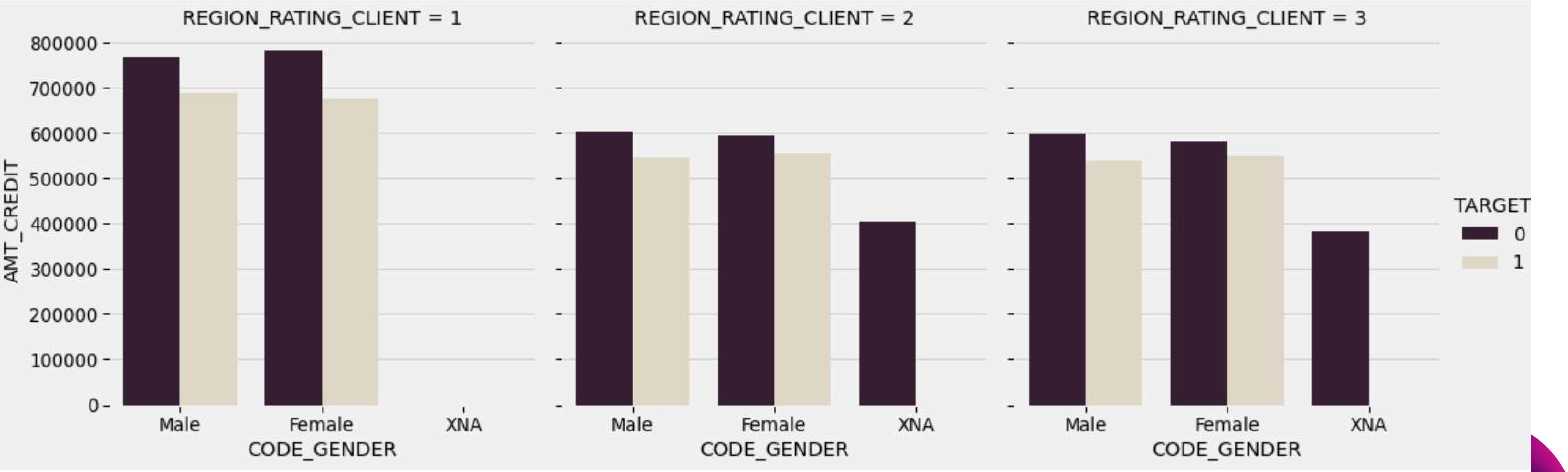


Amount Credit vs Target

Hasil disamping adalah grafik yang menunujukkan analisis dari Amount Credit dari Target. Dimana berdasarkan grafik dapat dilihat bahwa penyebaran income/pemasukan target yang dilihat dari umur.

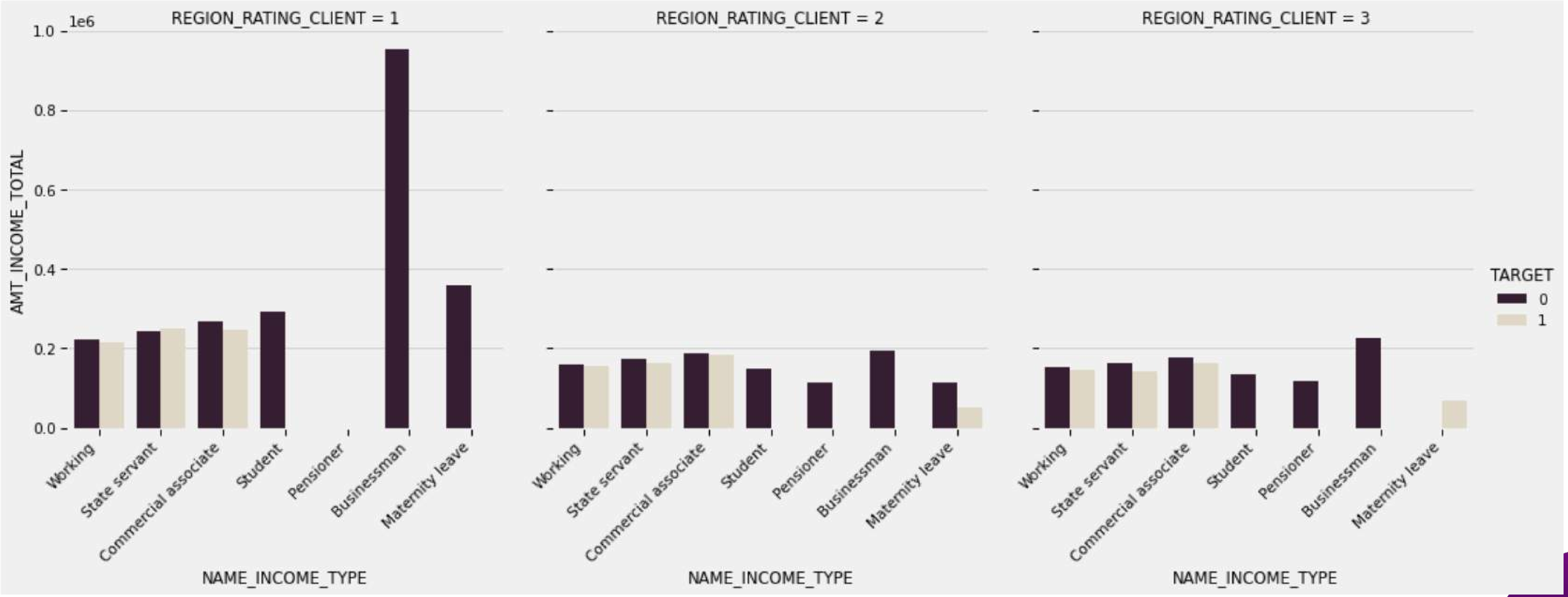
Data Preparation (EDA)

Gender, Amount Credit, Target, and Region Rating



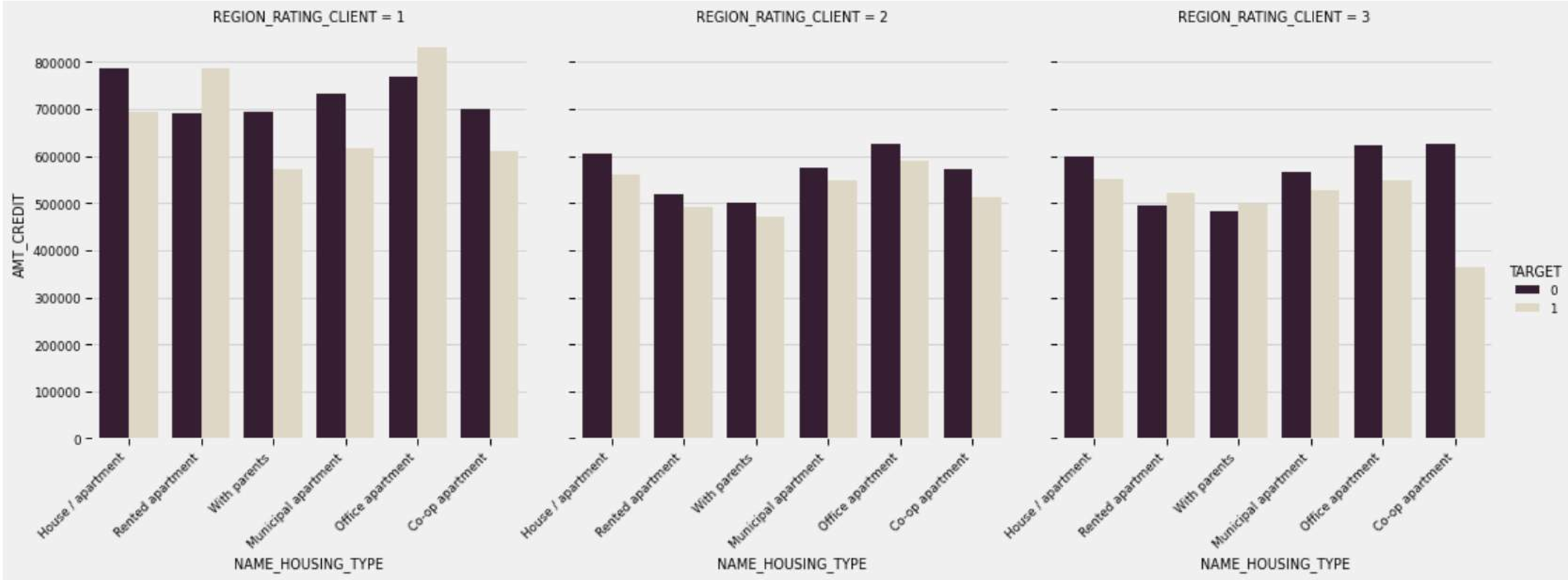
Data Preparation (EDA)

Income Type, Amount of Income, Target, and Region Rating



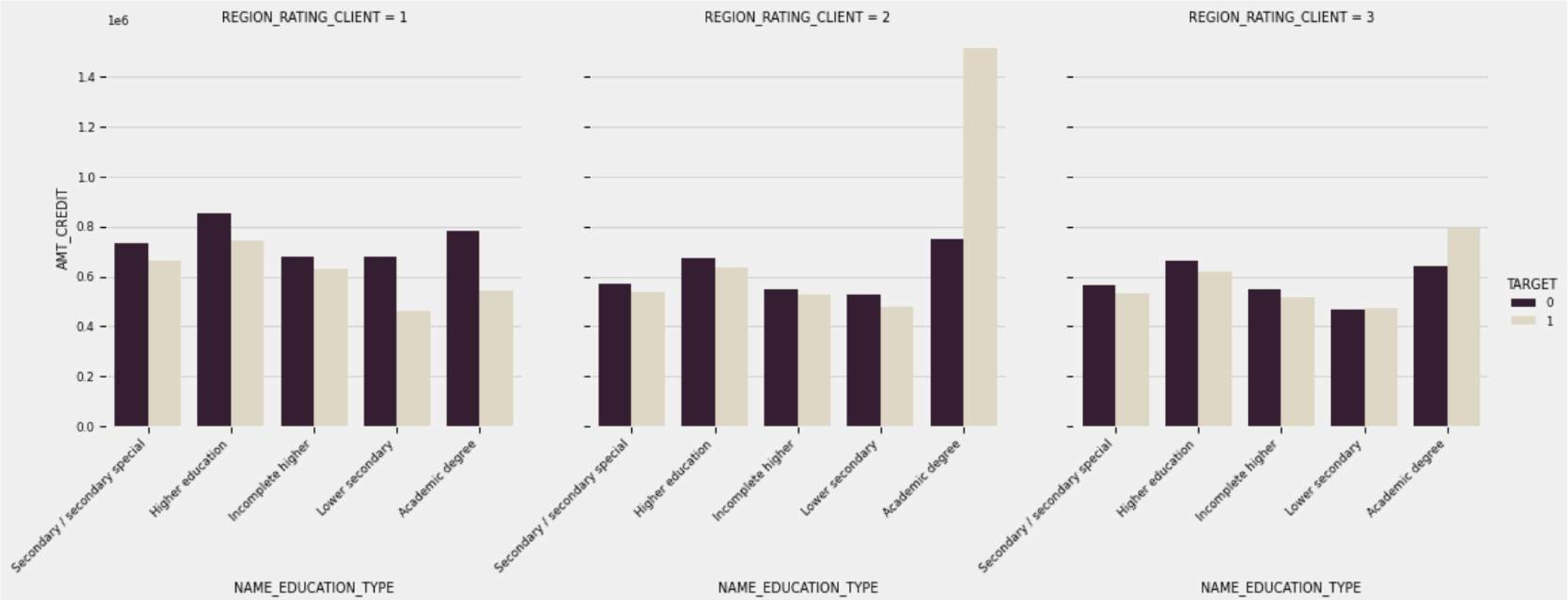
Data Preparation (EDA)

Housing Type, Amount of Credit, Target, and Region Rating



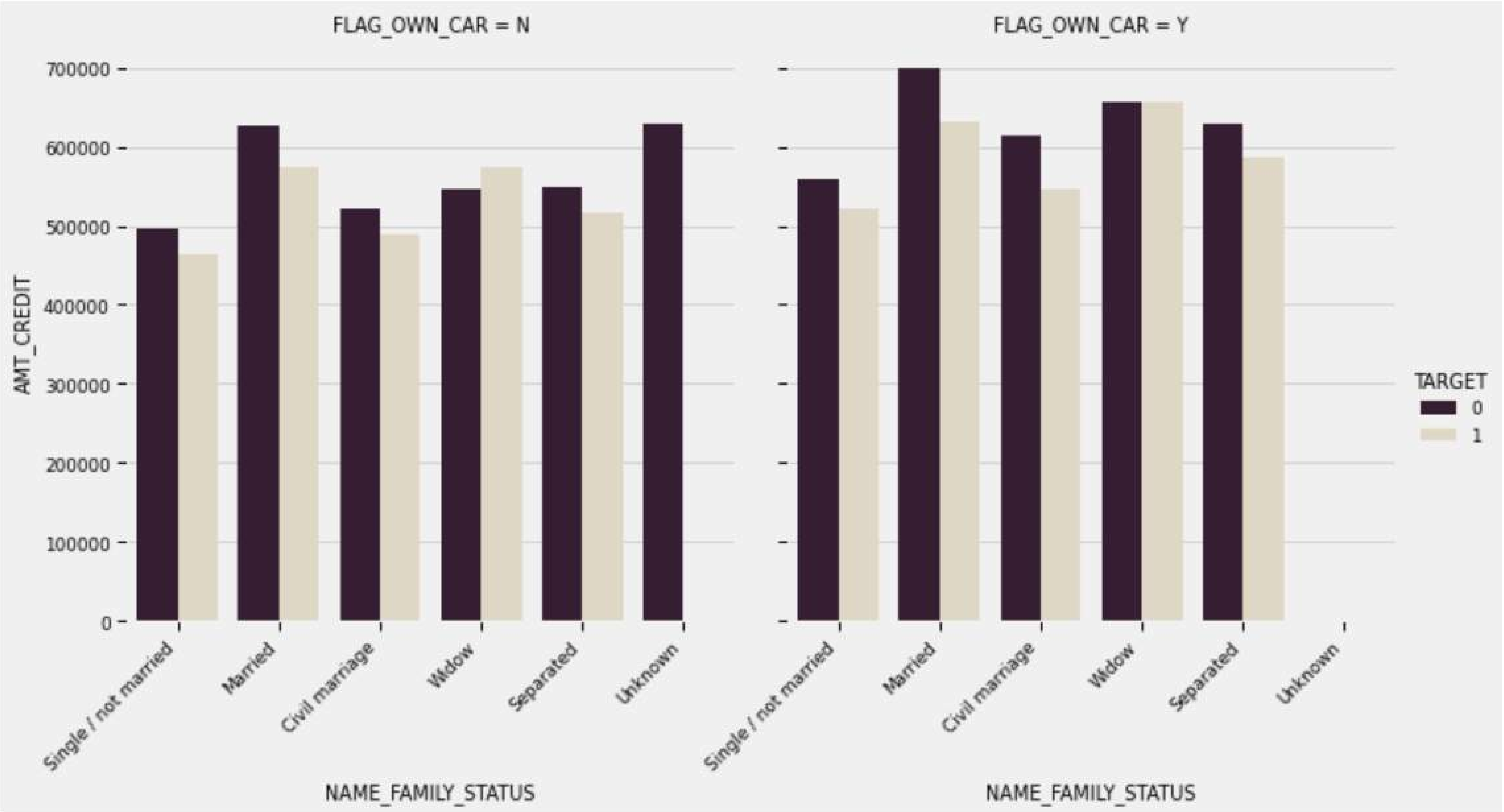
Data Preparation (EDA)

Education Type, Amount of Credit, Target, and Region Rating



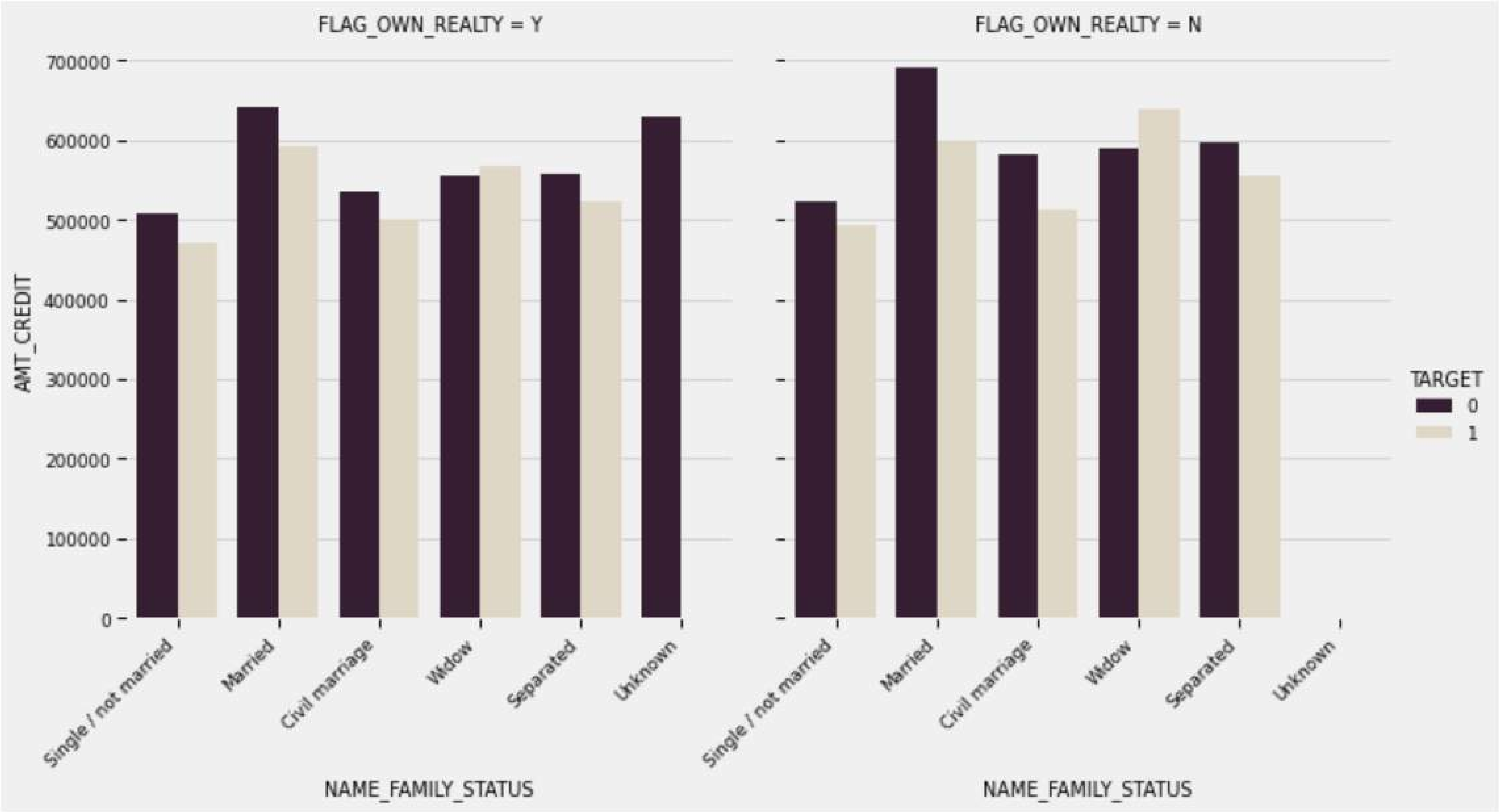
Data Preparation (EDA)

Family Status, Amount of Credit, Target, and Car Ownership



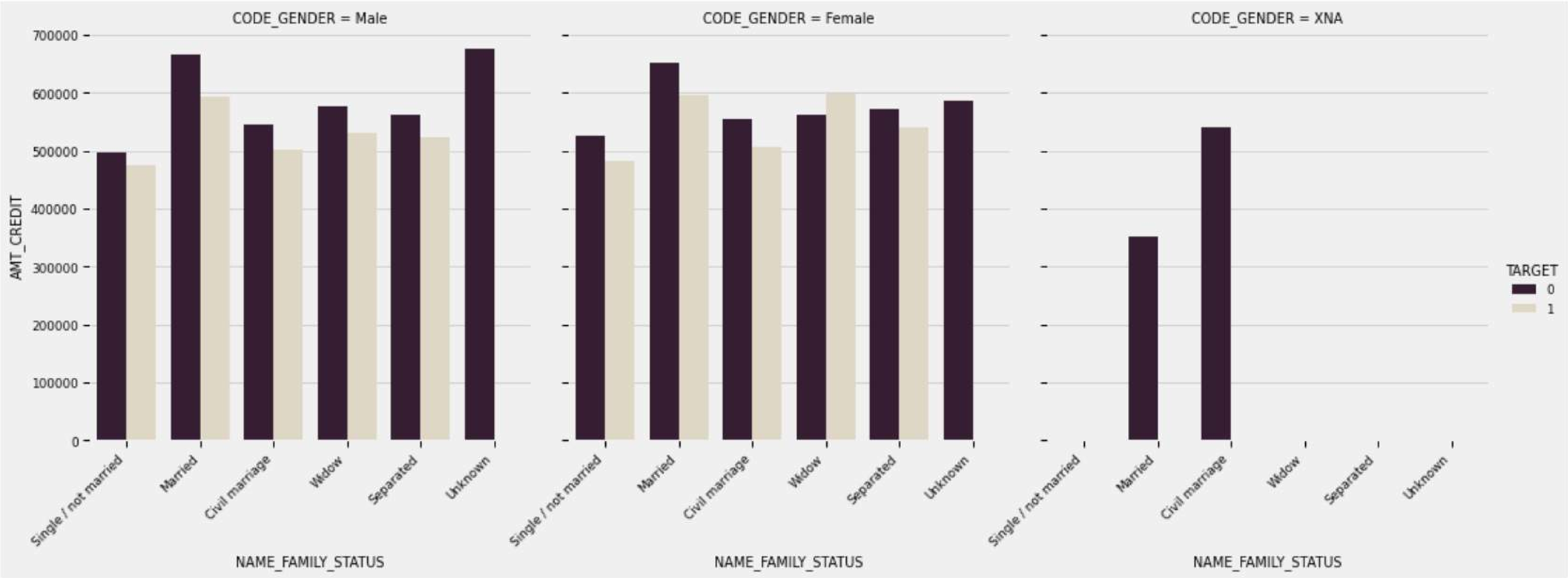
Data Preparation (EDA)

Family Status, Amount of Credit, Target, and Car Ownership



Data Preparation (EDA)

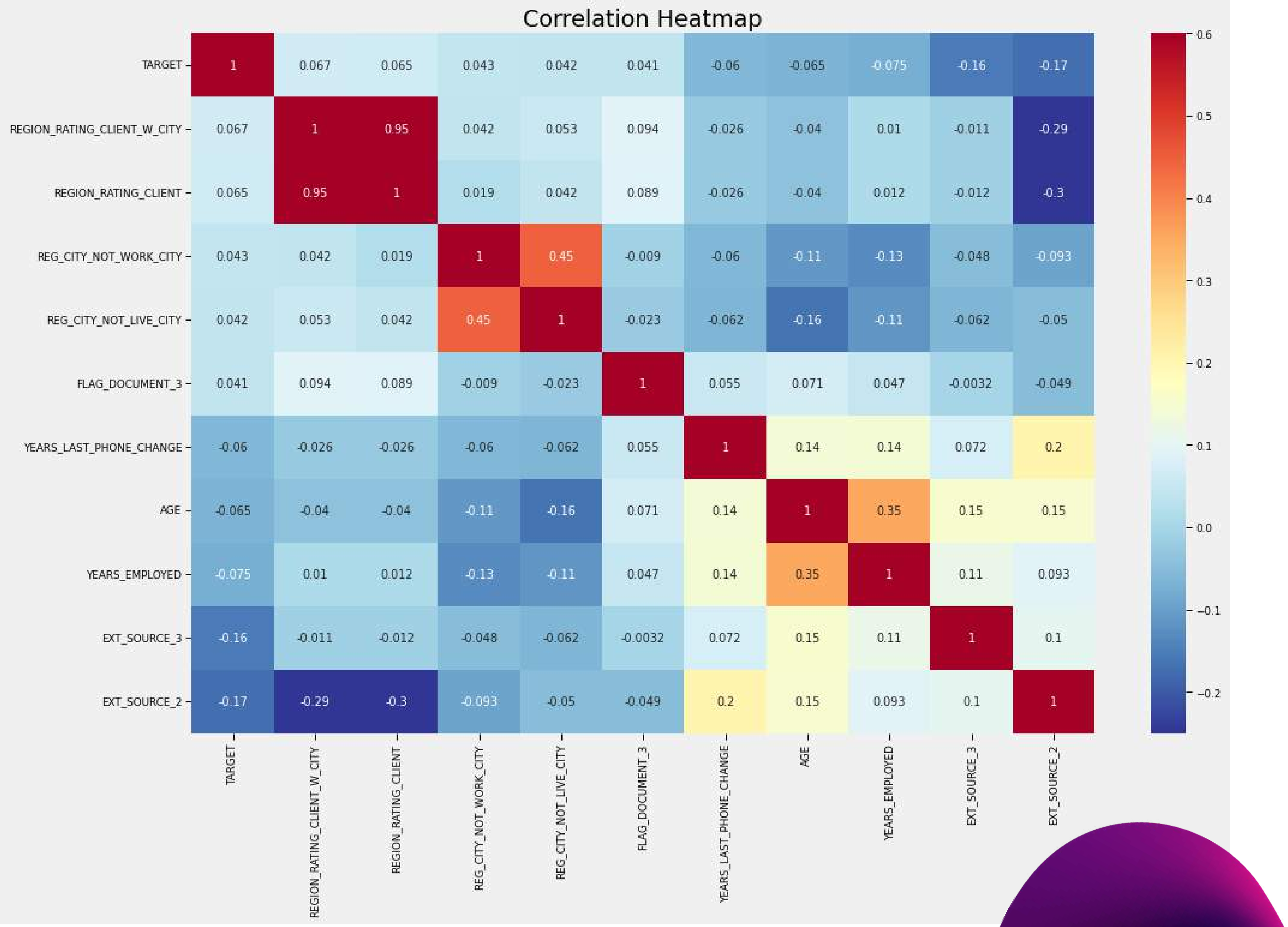
Family Status, Amount of Credit, Target, and Gender



Data Preparation (EDA)

Corelation

- Peta panas korelasi adalah grafik visual yang menunjukkan bagaimana setiap variabel dalam himpunan data berkorelasi satu sama lain. -1 menandakan korelasi nol, sedangkan 1 menandakan korelasi sempurna.
- Dalam hal ini kita dapat melihat visual dari dataset yang kita gunakan.

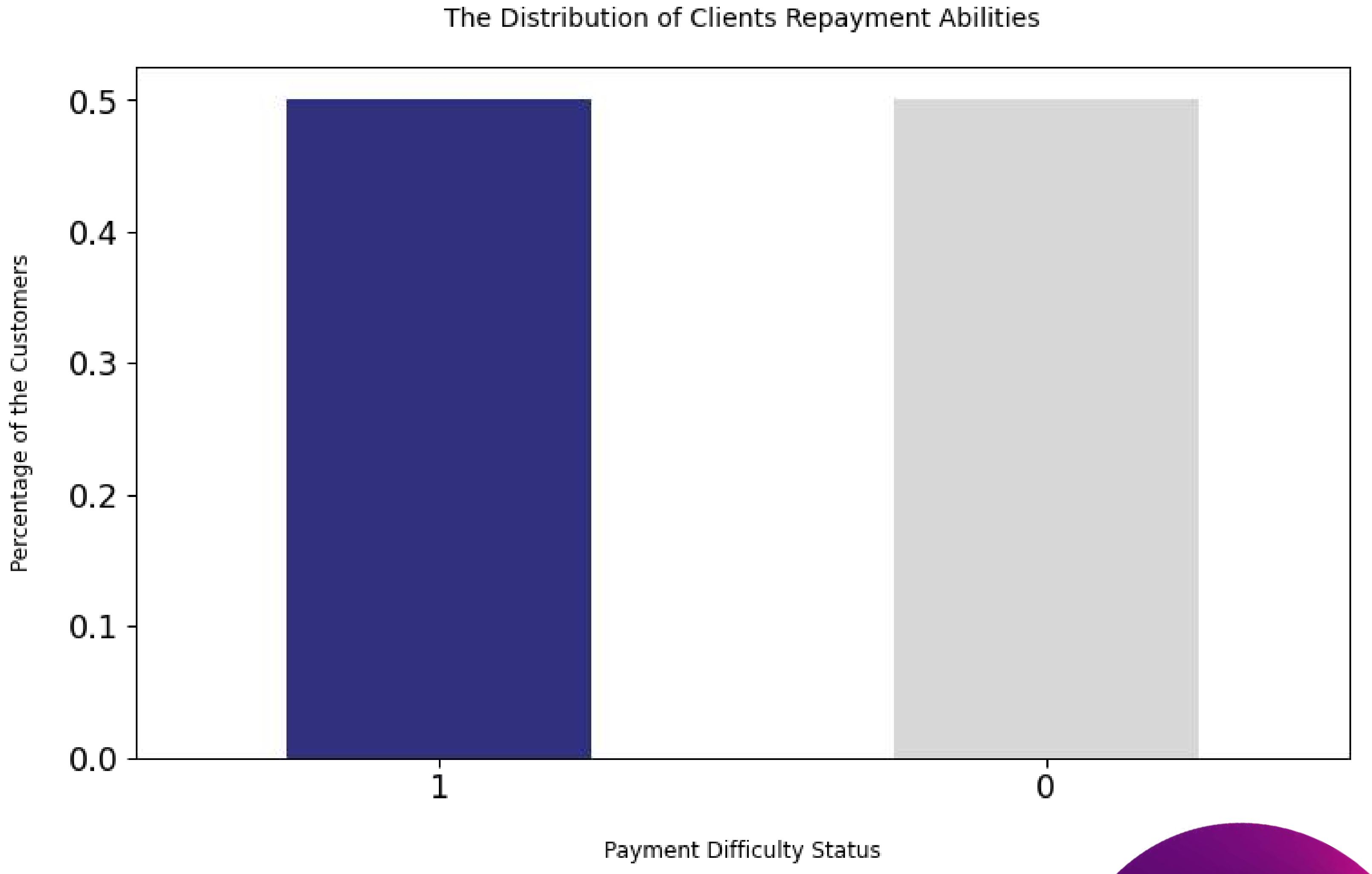


Machine Learning Modelling

Handling Imbalance Data

Handling imbalance data adalah cara untuk menangani ketidakseimbangan data. Imbalanced Dataset biasanya diolah secara klasifikasi dengan salah satu kelas/label pada datanya mempunyai nilai yang sangat jauh berbeda jumlahnya dari kelas lainnya. Pada imbalanced dataset, biasanya kita memiliki data dengan kelas yang sedikit (rare class) dan data dengan kelas yang banyak (abundant class).

```
0    229705
1     21808
Name: TARGET, dtype: int64
```



Machine Learning Modelling

Categorical Encoding

- One-Hot Encoding adalah teknik populer lainnya untuk memperlakukan variabel kategoris. Ini hanya membuat fitur tambahan berdasarkan jumlah nilai unik dalam fitur kategoris. Setiap nilai unik dalam kategori akan ditambahkan sebagai fitur.
- Pada langkah selanjutnya, kita akan membandingkan pemisahan data dengan dan tanpa pemilihan fitur, jadi kita akan menggunakan Label Encoding sebagai gantinya, tetapi juga kita melampirkan kode untuk One Hot Encoding

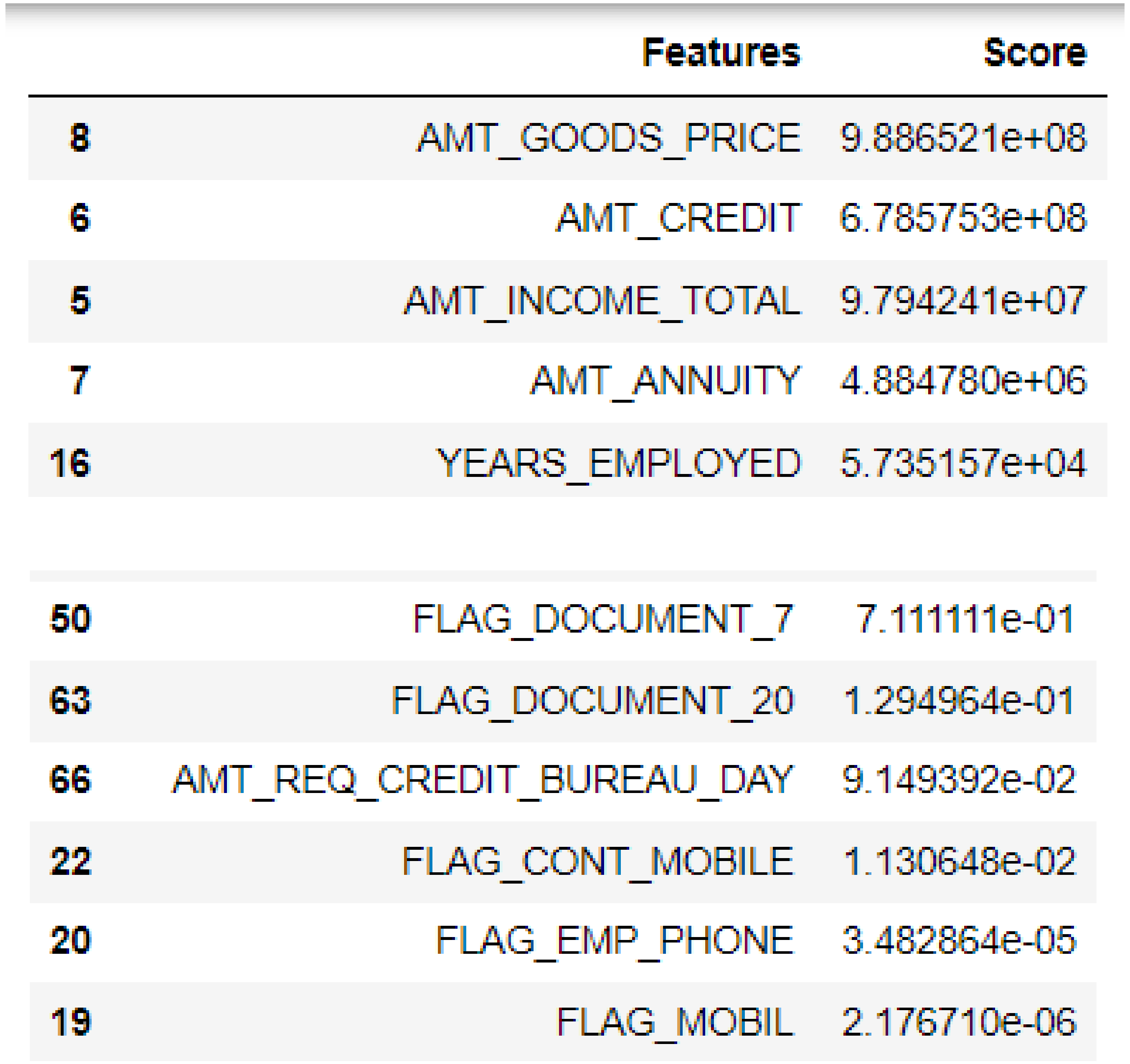
Machine Learning Modelling

Train and Test Split

Untuk membandingkan model dengan dan tanpa Pemilihan Fitur. Kita akan membedakan datanya.
"train, test" = Data tanpa Pemilihan Fitur
"train1, test1" = Data dengan Pemilihan Fitur

Machine Learning Modelling

Feature Selection



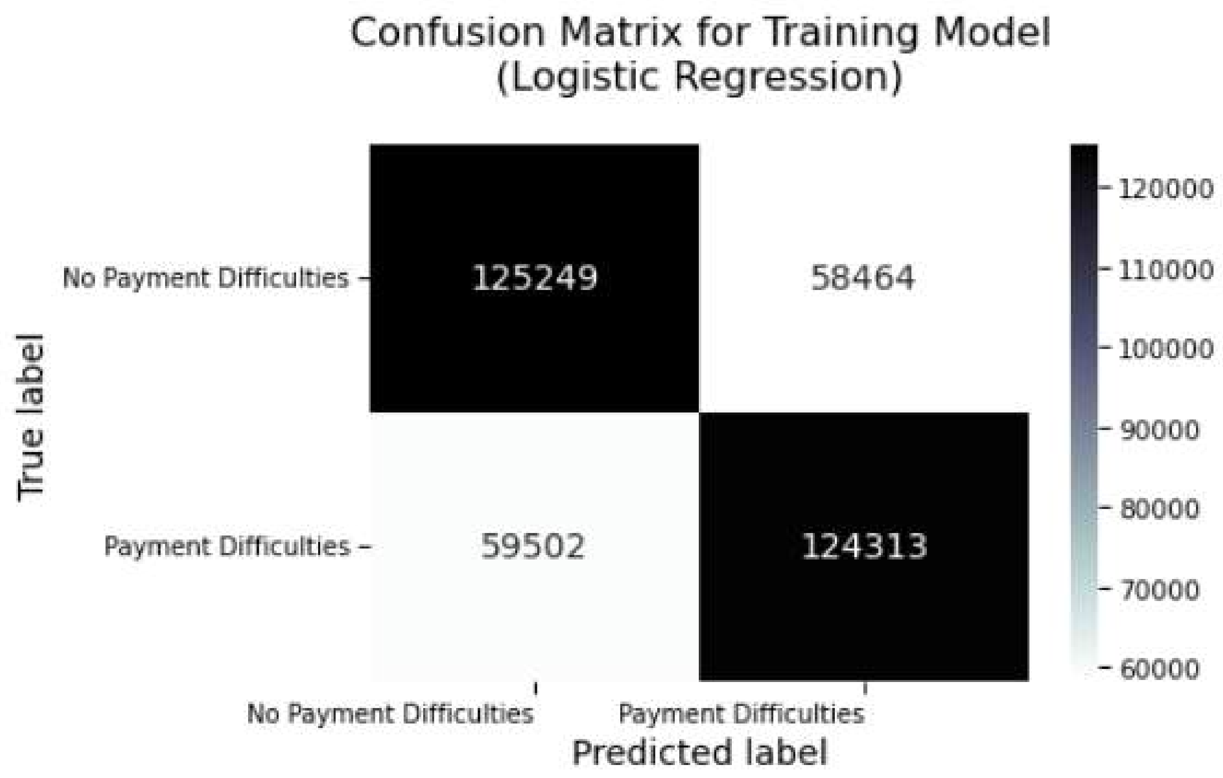
	Features	Score
8	AMT_GOODS_PRICE	9.886521e+08
6	AMT_CREDIT	6.785753e+08
5	AMT_INCOME_TOTAL	9.794241e+07
7	AMT_ANNUITY	4.884780e+06
16	YEARS_EMPLOYED	5.735157e+04
50	FLAG_DOCUMENT_7	7.111111e-01
63	FLAG_DOCUMENT_20	1.294964e-01
66	AMT_REQ_CREDIT_BUREAU_DAY	9.149392e-02
22	FLAG_CONT_MOBILE	1.130648e-02
20	FLAG_EMP_PHONE	3.482864e-05
19	FLAG_MOBIL	2.176710e-06

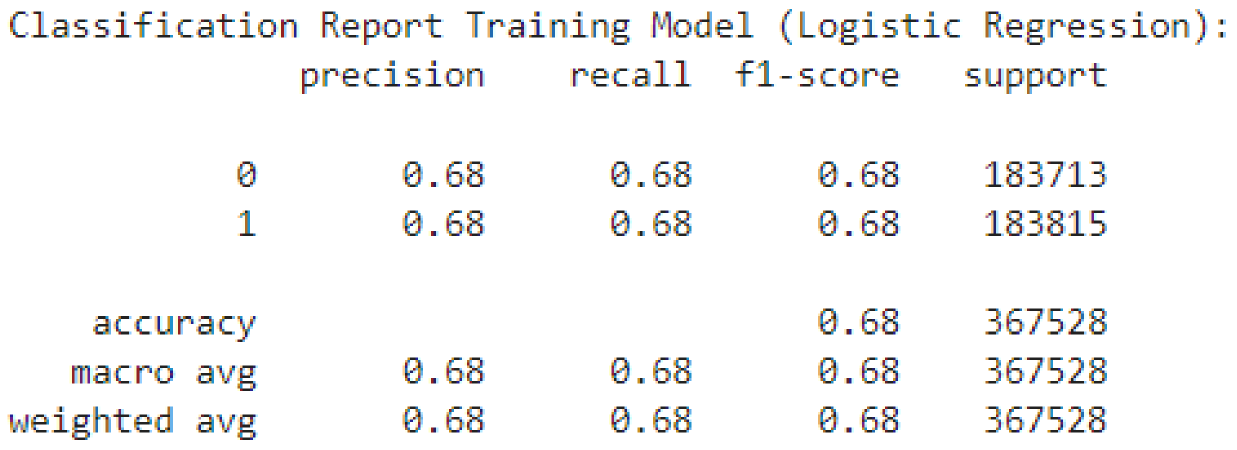
- Best features : YEARS_EMPLOYED, AMT_GOODS_PRICE, and AMT_CREDIT
- Worst features : FLAG_MOBIL, FLAG_CONT_MOBILE, and AMT_REQ_CREDIT_BUREAU_HOUR

Machine Learning Modelling

Logistic Regression

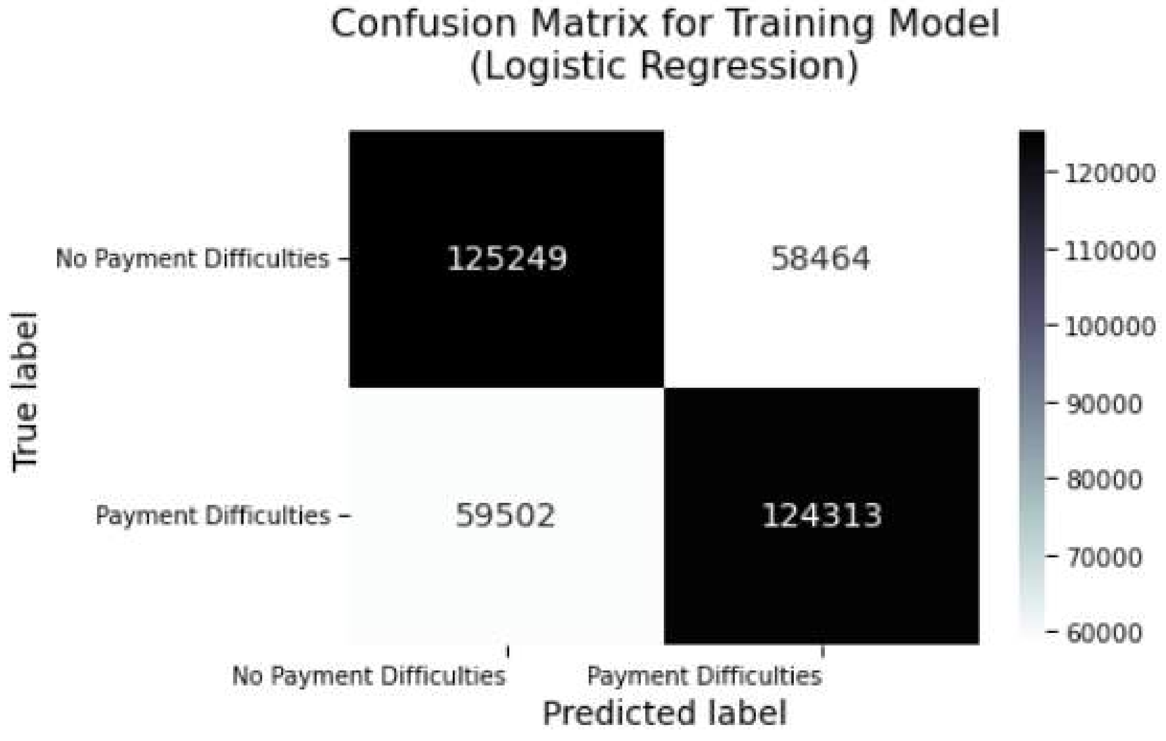
Without Feature Selection

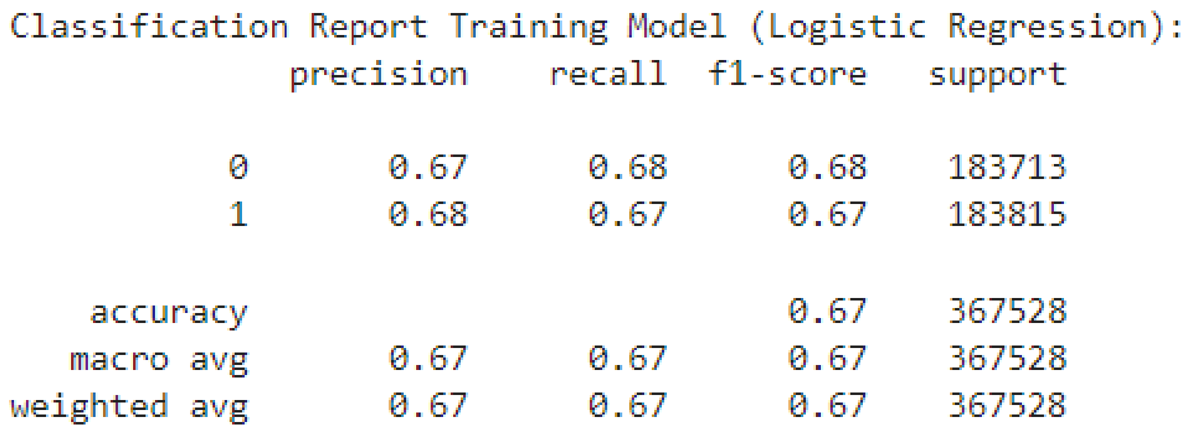




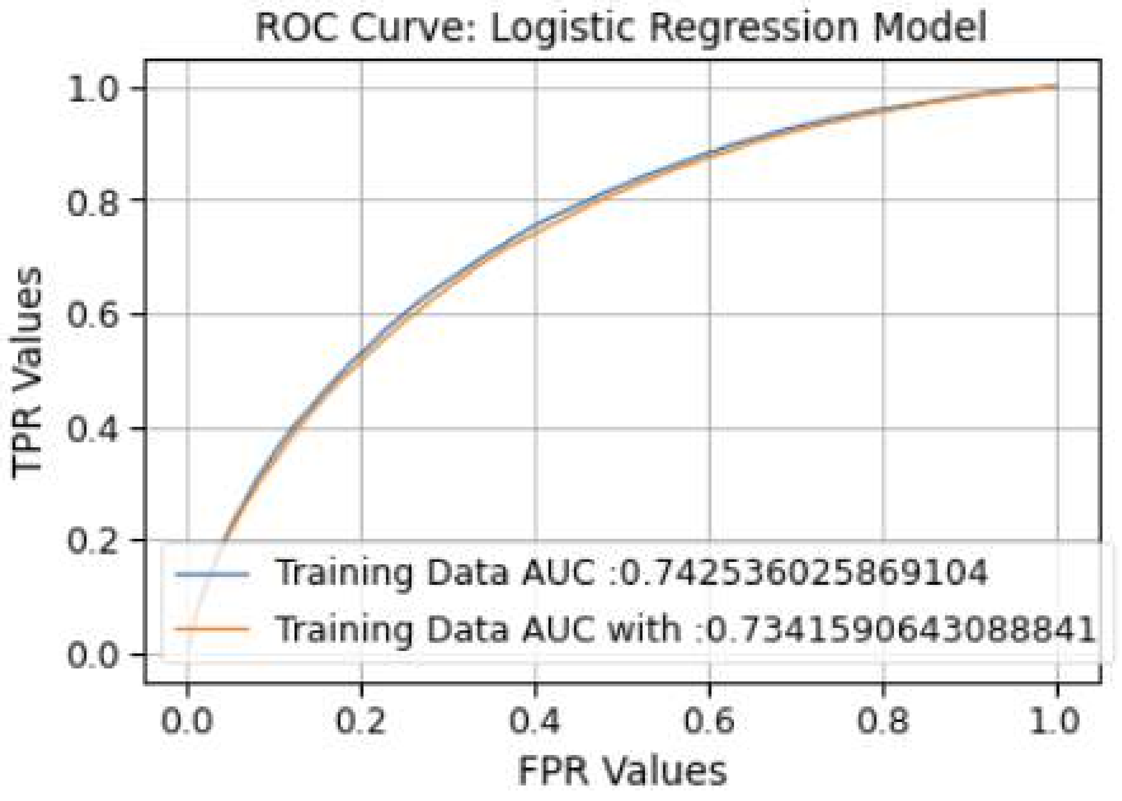
```
Classification Report Training Model (Logistic Regression):
              precision    recall  f1-score   support

           0       0.68      0.68      0.68    183713
           1       0.68      0.68      0.68    183815

    accuracy                           0.68    367528
   macro avg       0.68      0.68      0.68    367528
weighted avg       0.68      0.68      0.68    367528
```

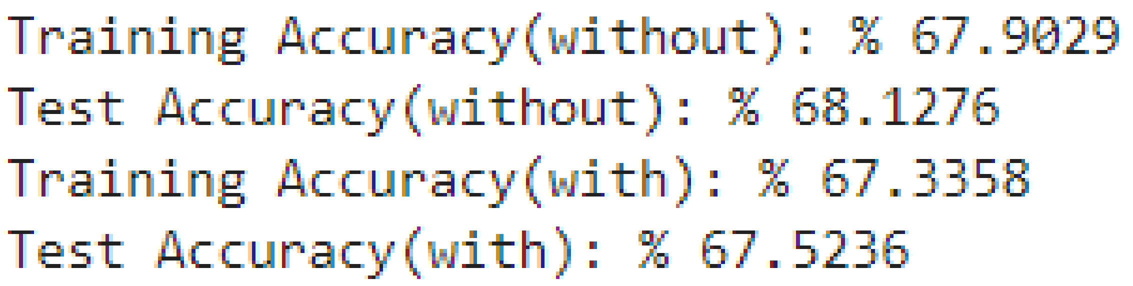
With Feature Selection





```
Classification Report Training Model (Logistic Regression):
              precision    recall  f1-score   support

           0       0.67      0.68      0.68    183713
           1       0.68      0.67      0.67    183815

    accuracy                           0.67    367528
   macro avg       0.67      0.67      0.67    367528
weighted avg       0.67      0.67      0.67    367528
```



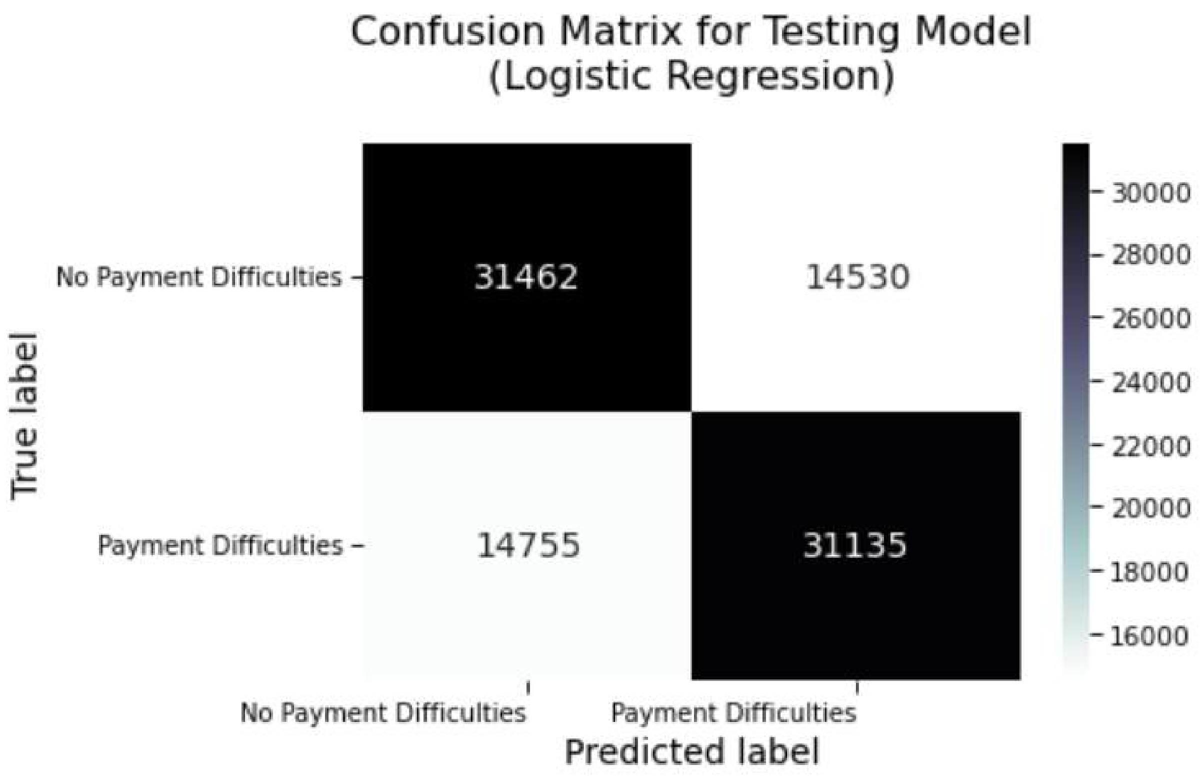


```
Training Accuracy(without): % 67.9029
Test Accuracy(without): % 68.1276
Training Accuracy(with): % 67.3358
Test Accuracy(with): % 67.5236
```

Machine Learning Modelling

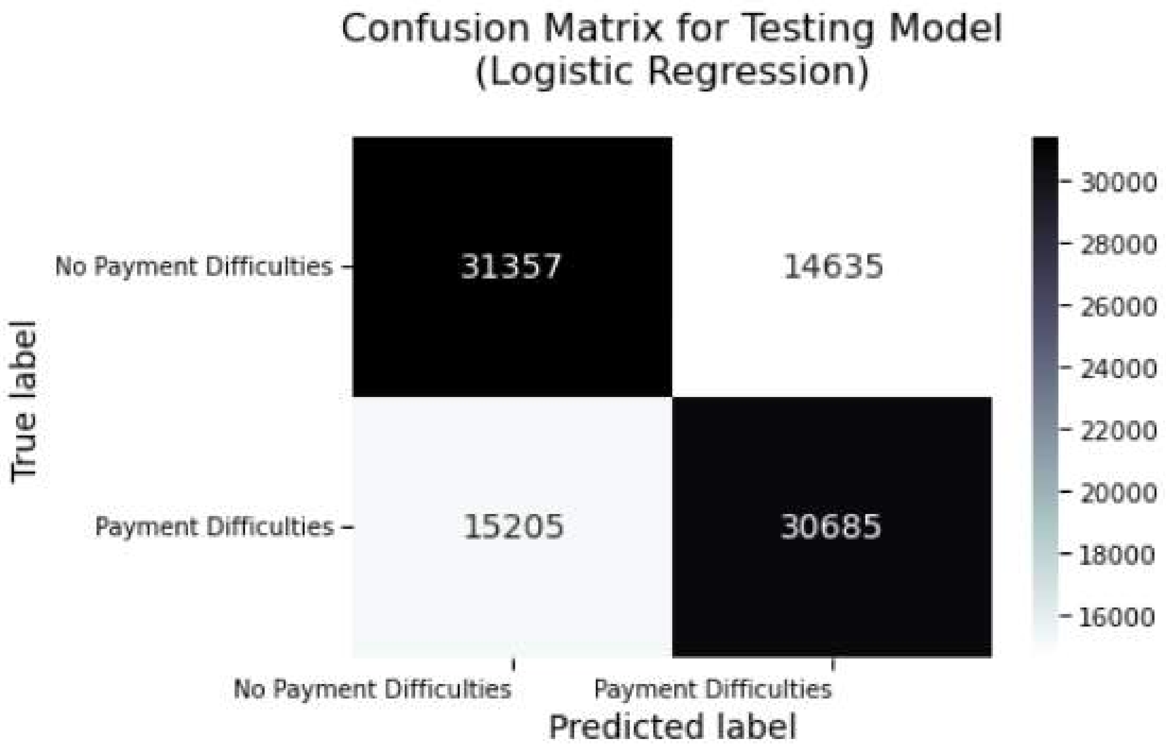
Logistic Regression

Without Feature Selection

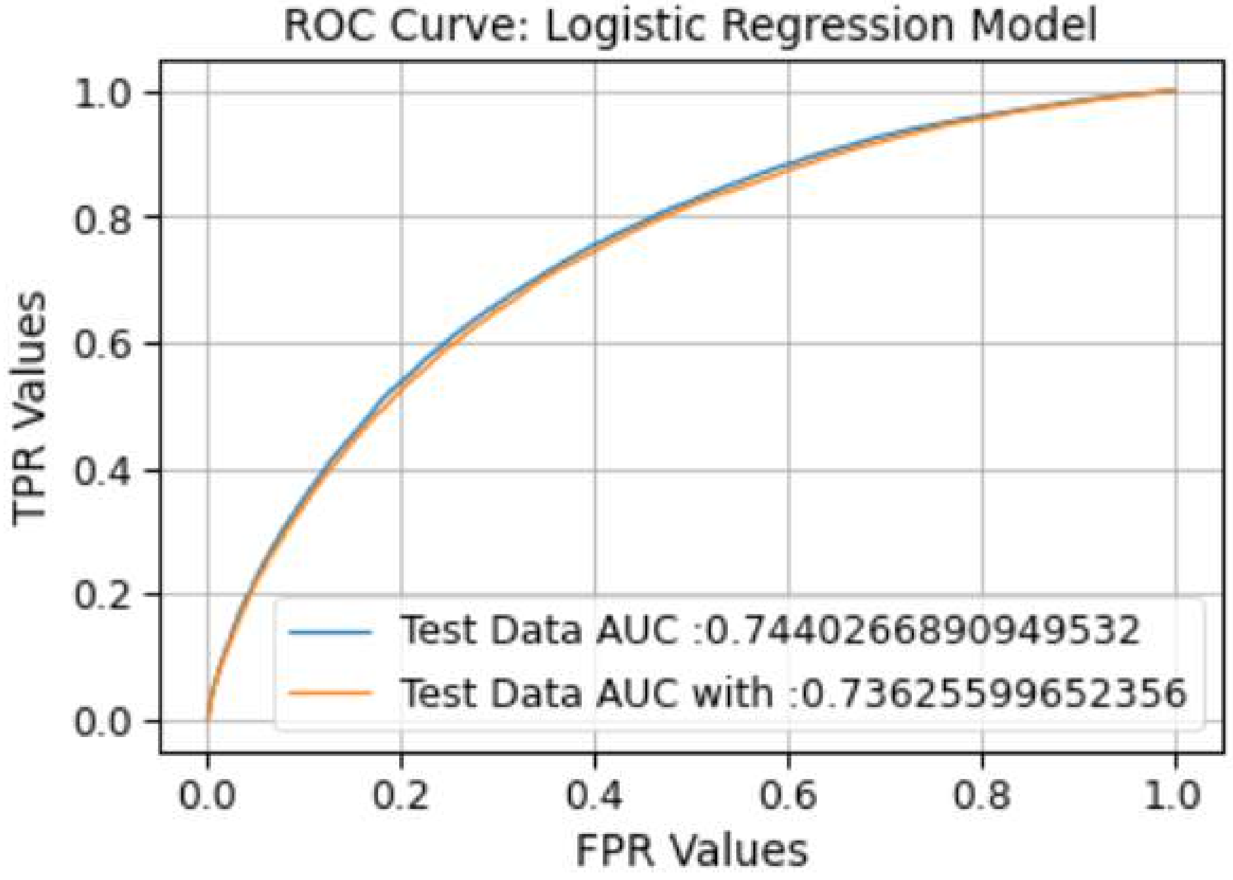


```
Classification Report Testing Model (Logistic Regression):
              precision    recall  f1-score   support

           0       0.68      0.68      0.68     45992
           1       0.68      0.68      0.68     45890

    accuracy                           0.68     91882
   macro avg       0.68      0.68      0.68     91882
weighted avg       0.68      0.68      0.68     91882
```

With Feature Selection



```
Classification Report Testing Model (Logistic Regression):
              precision    recall  f1-score   support

           0       0.67      0.68      0.68     45992
           1       0.68      0.67      0.67     45890

    accuracy                           0.68     91882
   macro avg       0.68      0.68      0.68     91882
weighted avg       0.68      0.68      0.68     91882
```

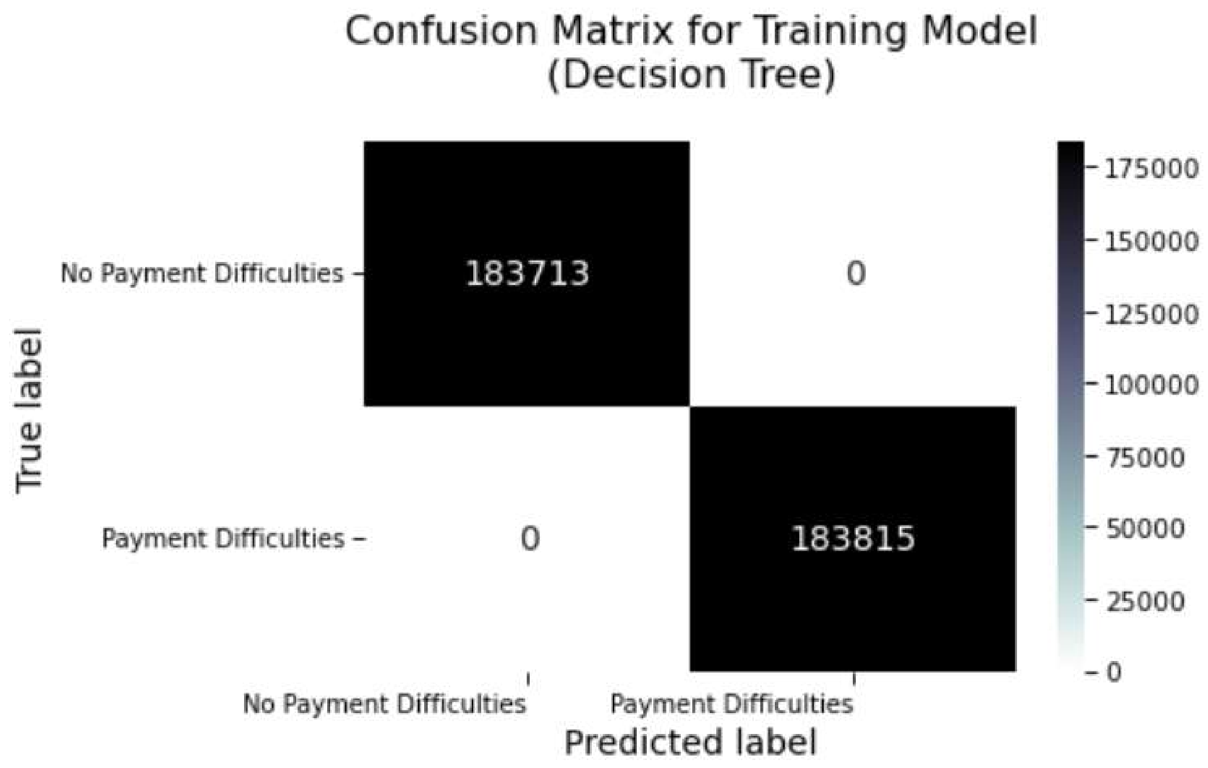


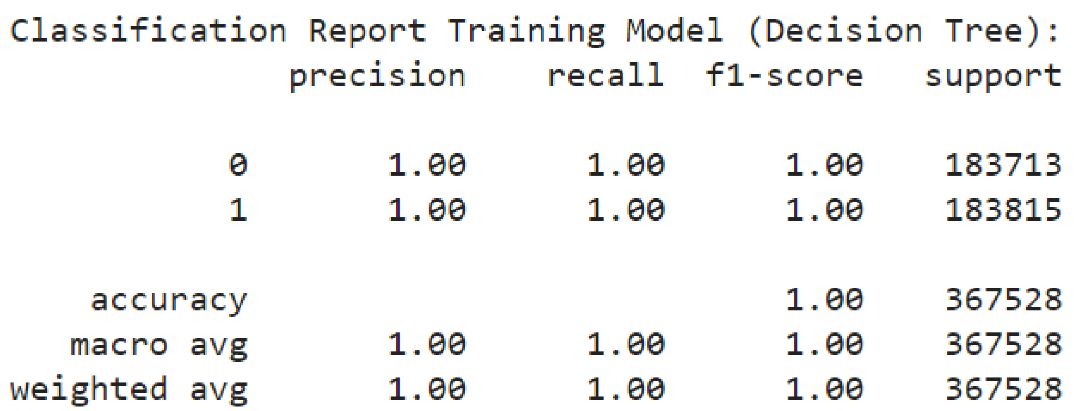
```
Training Accuracy(without): % 67.9029
Test Accuracy(without): % 68.1276
Training Accuracy(with): % 67.3358
Test Accuracy(with): % 67.5236
```

Machine Learning Modelling

Decision Tree

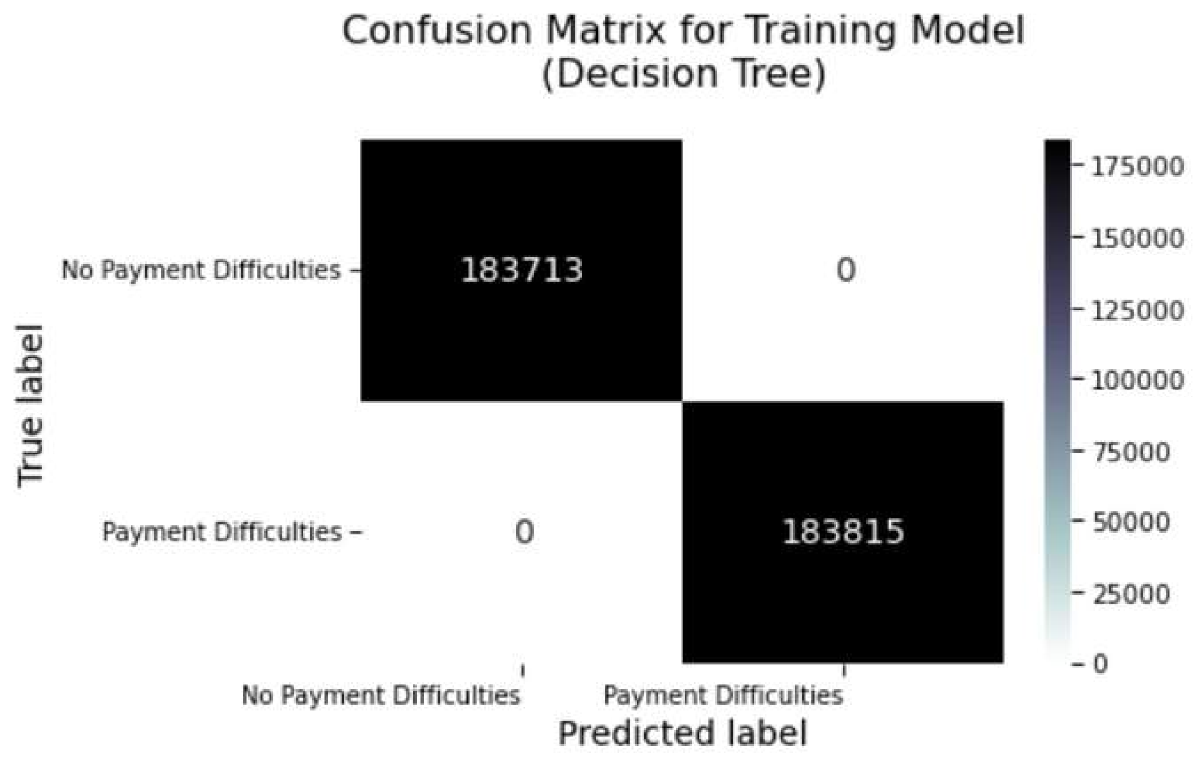
Without Feature Selection

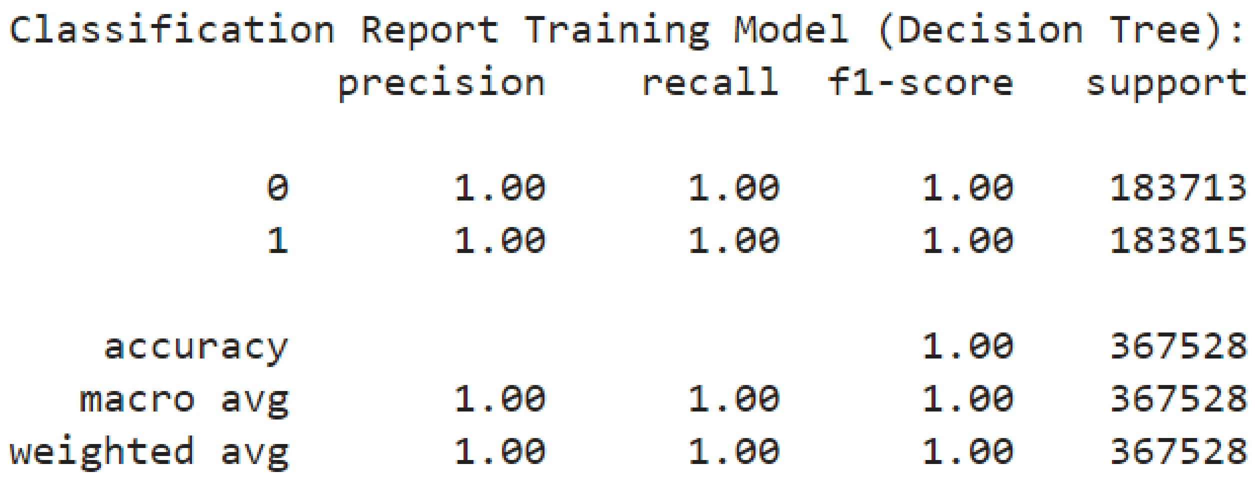




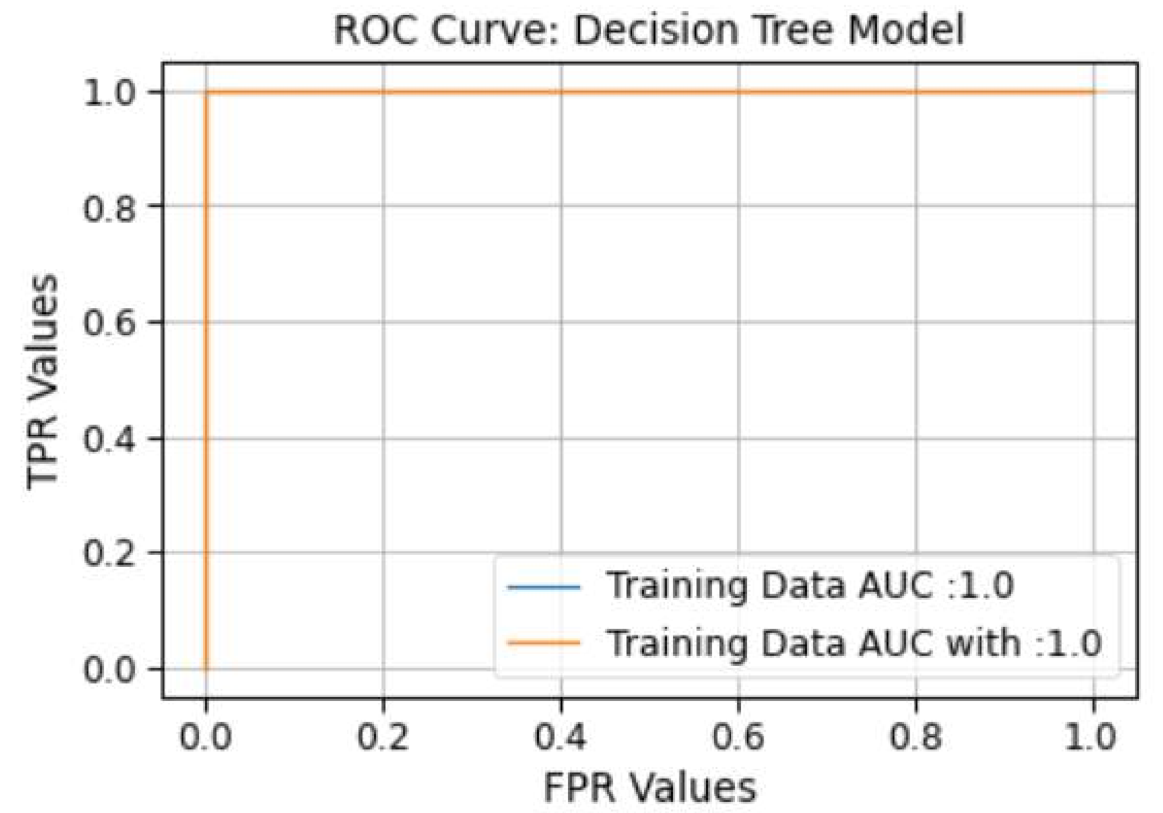
```
Classification Report Training Model (Decision Tree):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    183713
           1       1.00      1.00      1.00    183815

    accuracy                           1.00    367528
   macro avg       1.00      1.00      1.00    367528
weighted avg       1.00      1.00      1.00    367528
```

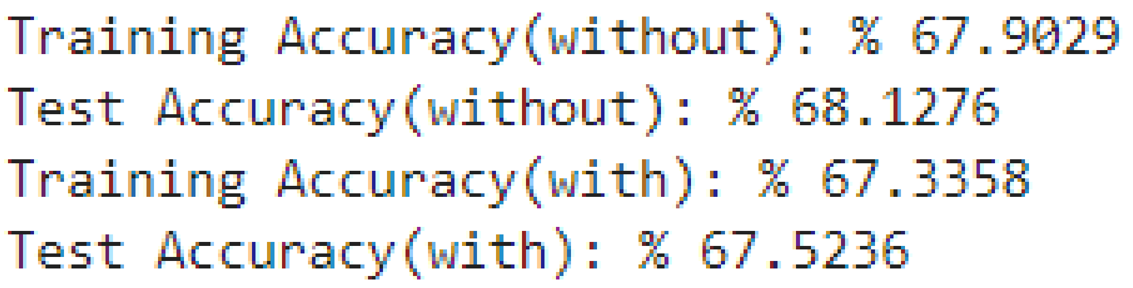
With Feature Selection





```
Classification Report Training Model (Decision Tree):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    183713
           1       1.00      1.00      1.00    183815

    accuracy                           1.00    367528
   macro avg       1.00      1.00      1.00    367528
weighted avg       1.00      1.00      1.00    367528
```





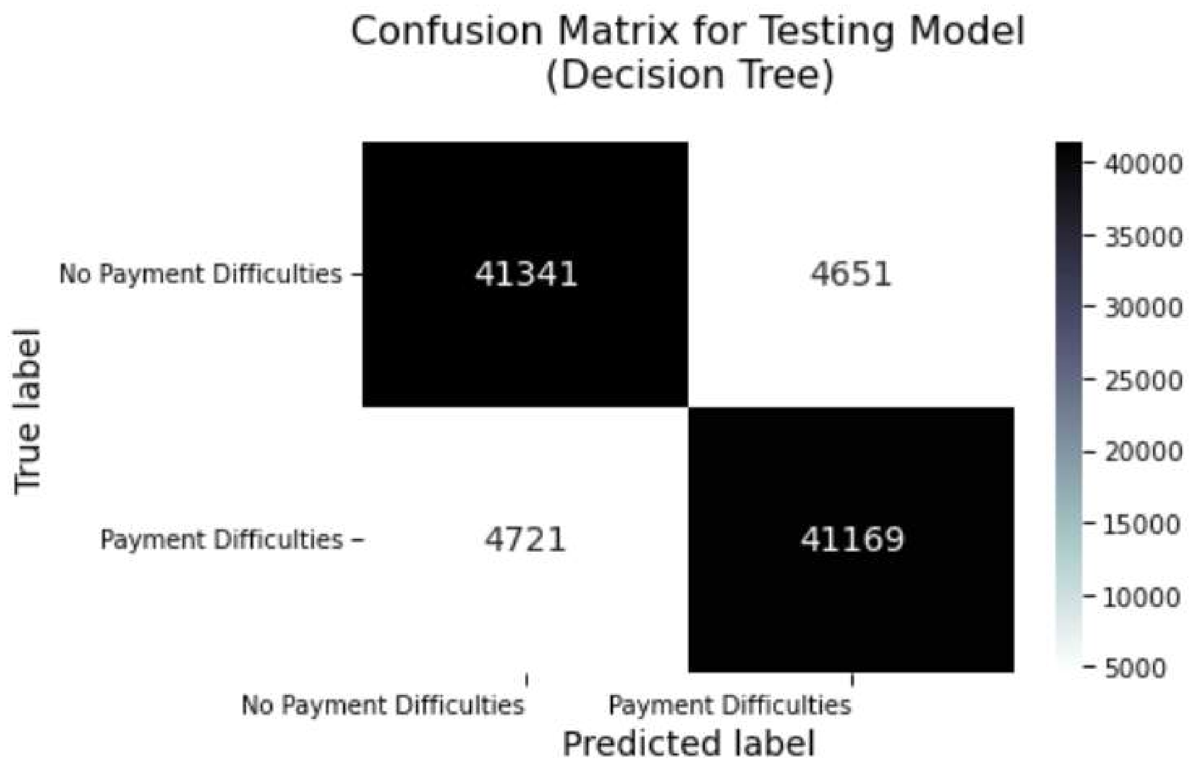
```
Training Accuracy(without): % 67.9029
Test Accuracy(without): % 68.1276
Training Accuracy(with): % 67.3358
Test Accuracy(with): % 67.5236
```

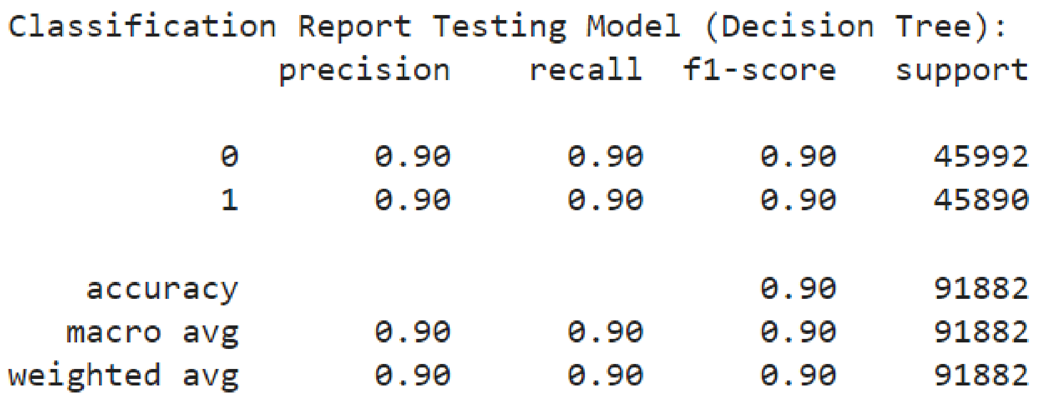
zenius × Kampus Merdeka INDONESIA JAYA

Machine Learning Modelling

Decision Tree

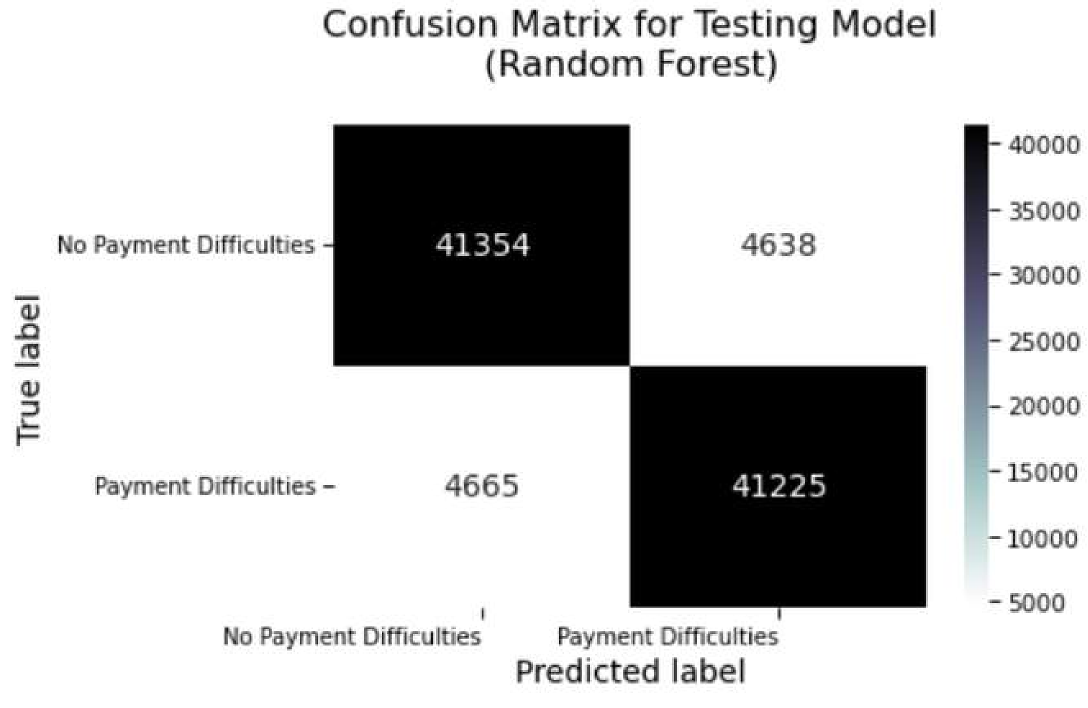
Without Feature Selection

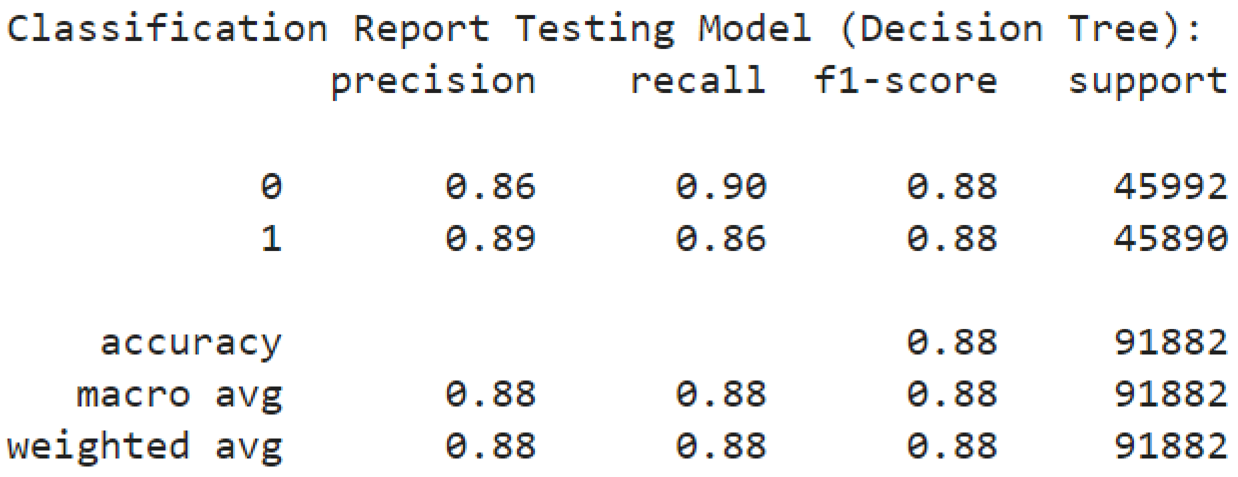




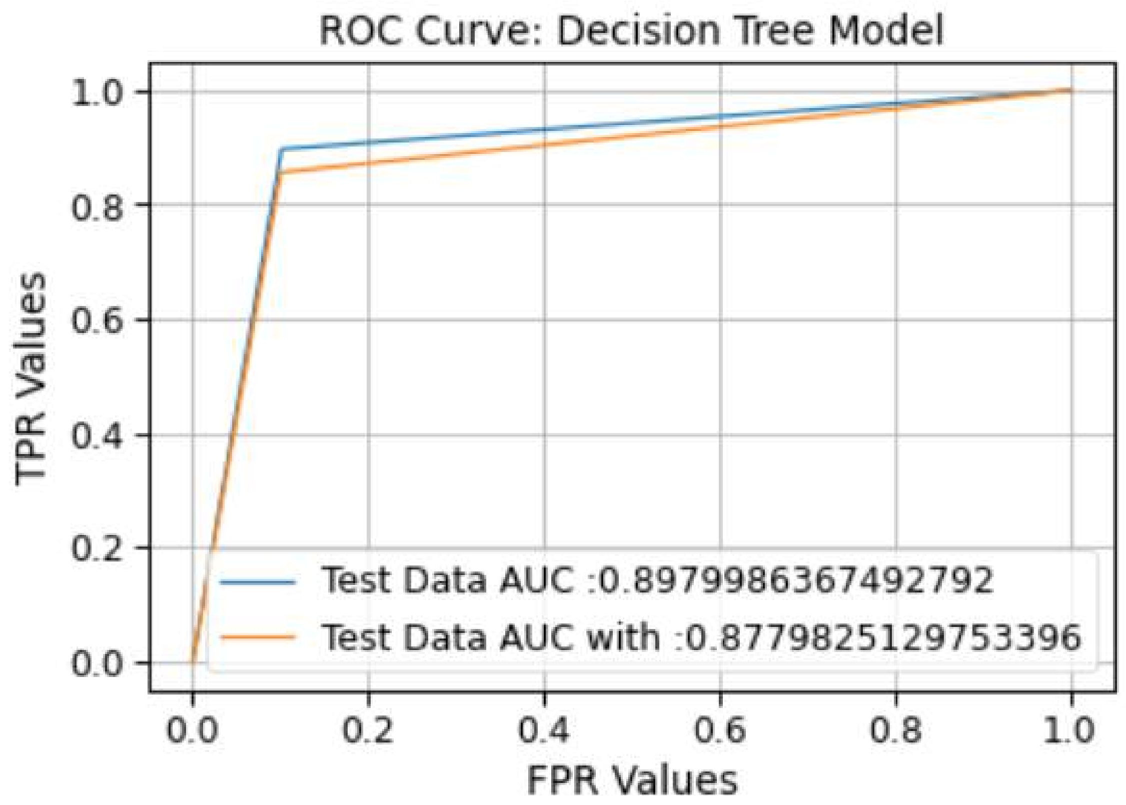
```
Classification Report Testing Model (Decision Tree):
              precision    recall  f1-score   support

           0       0.90      0.90      0.90     45992
           1       0.90      0.90      0.90     45890

    accuracy                           0.90     91882
   macro avg       0.90      0.90      0.90     91882
weighted avg       0.90      0.90      0.90     91882
```

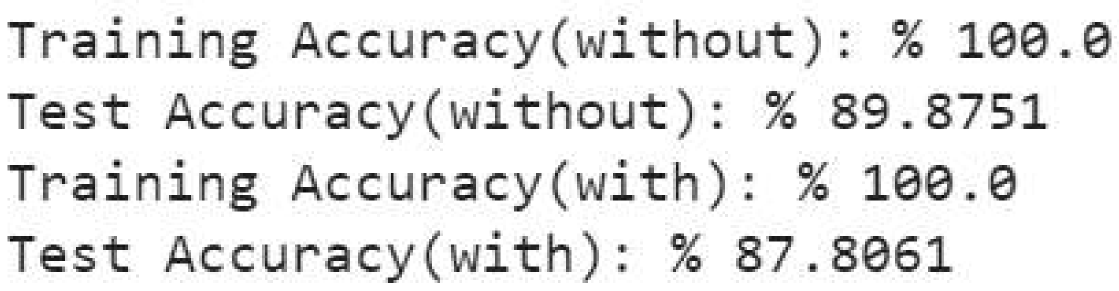
With Feature Selection





```
Classification Report Testing Model (Decision Tree):
              precision    recall  f1-score   support

           0       0.86      0.90      0.88     45992
           1       0.89      0.86      0.88     45890

    accuracy                           0.88     91882
   macro avg       0.88      0.88      0.88     91882
weighted avg       0.88      0.88      0.88     91882
```



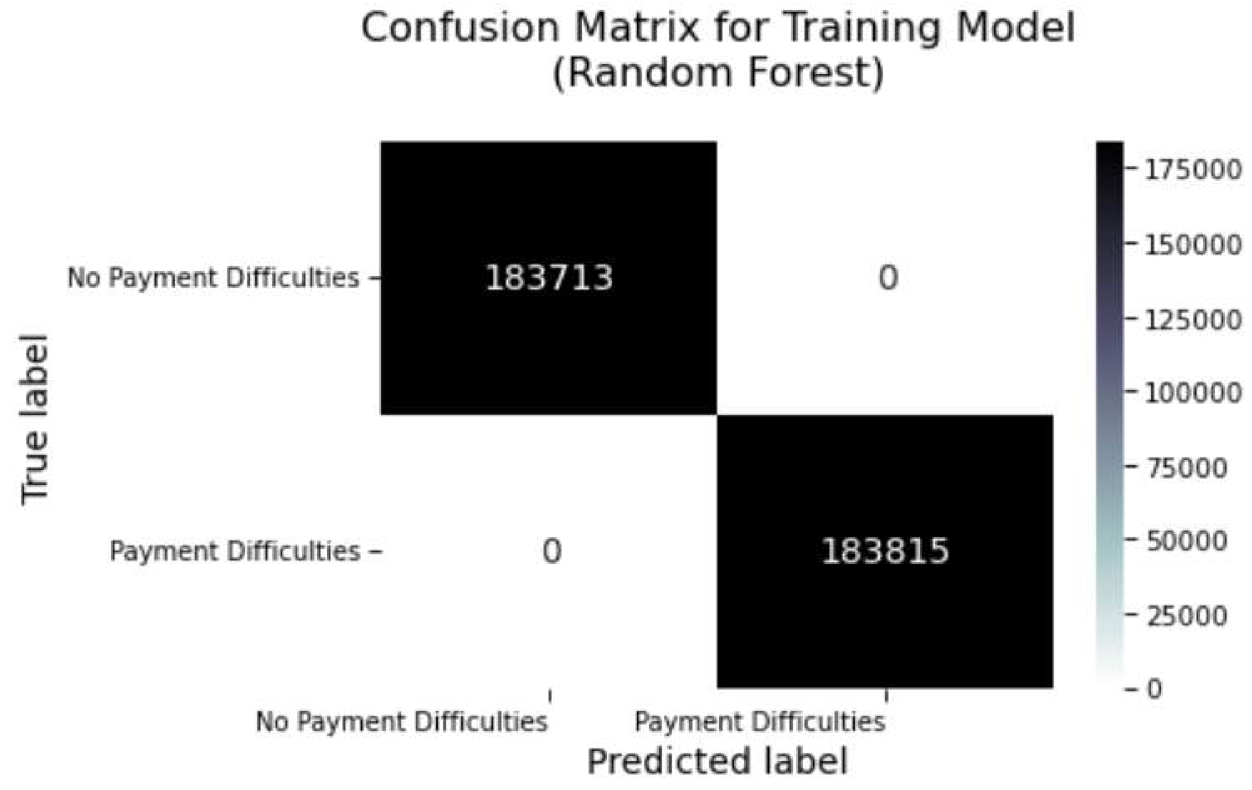


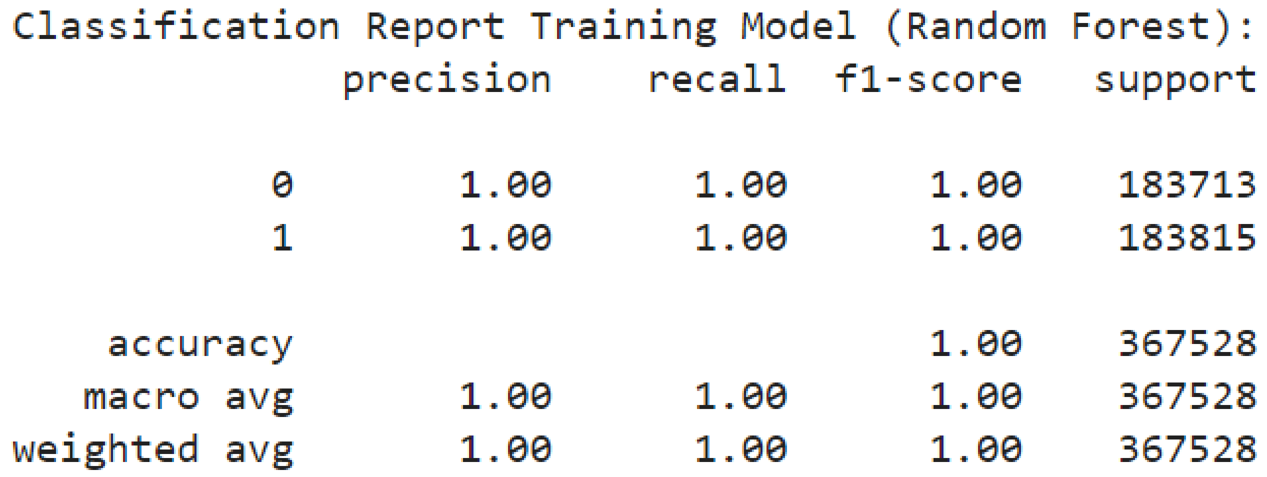
```
Training Accuracy(without): % 100.0
Test Accuracy(without): % 89.8751
Training Accuracy(with): % 100.0
Test Accuracy(with): % 87.8061
```

Machine Learning Modelling

Random Forest

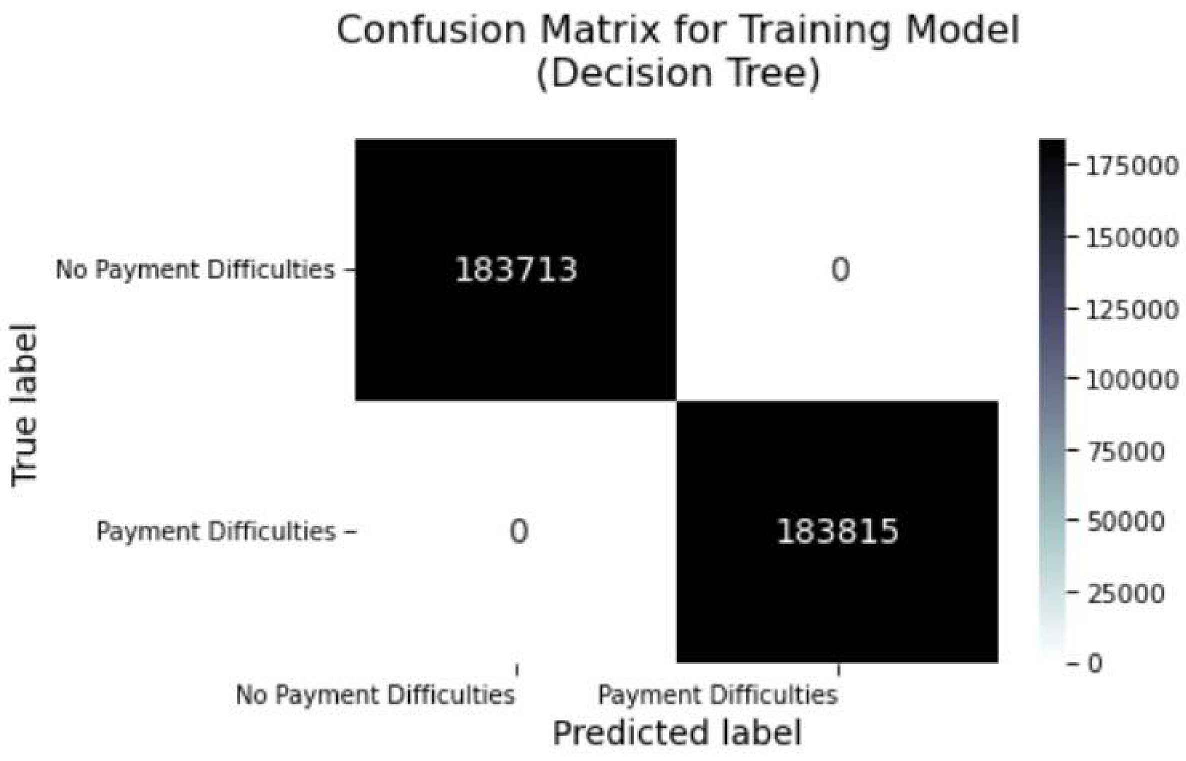
Without Feature Selection

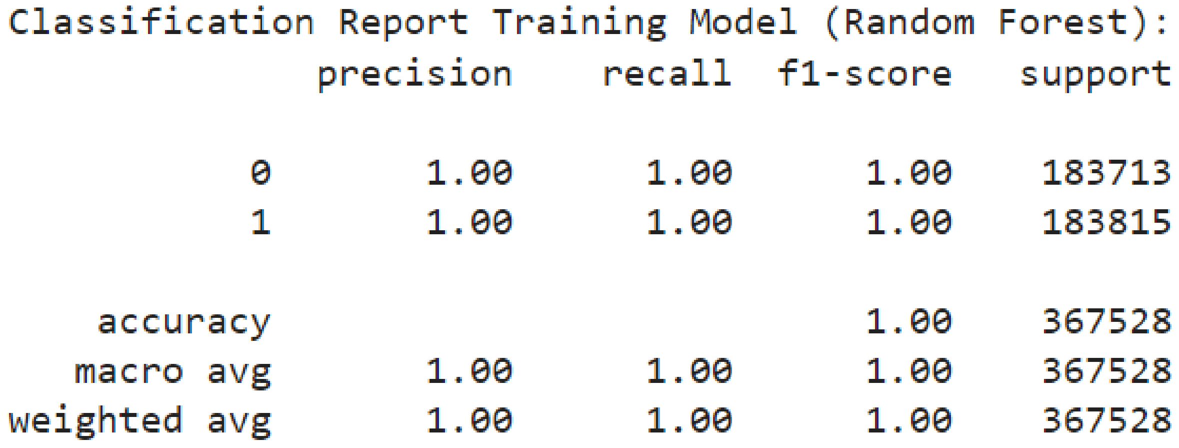




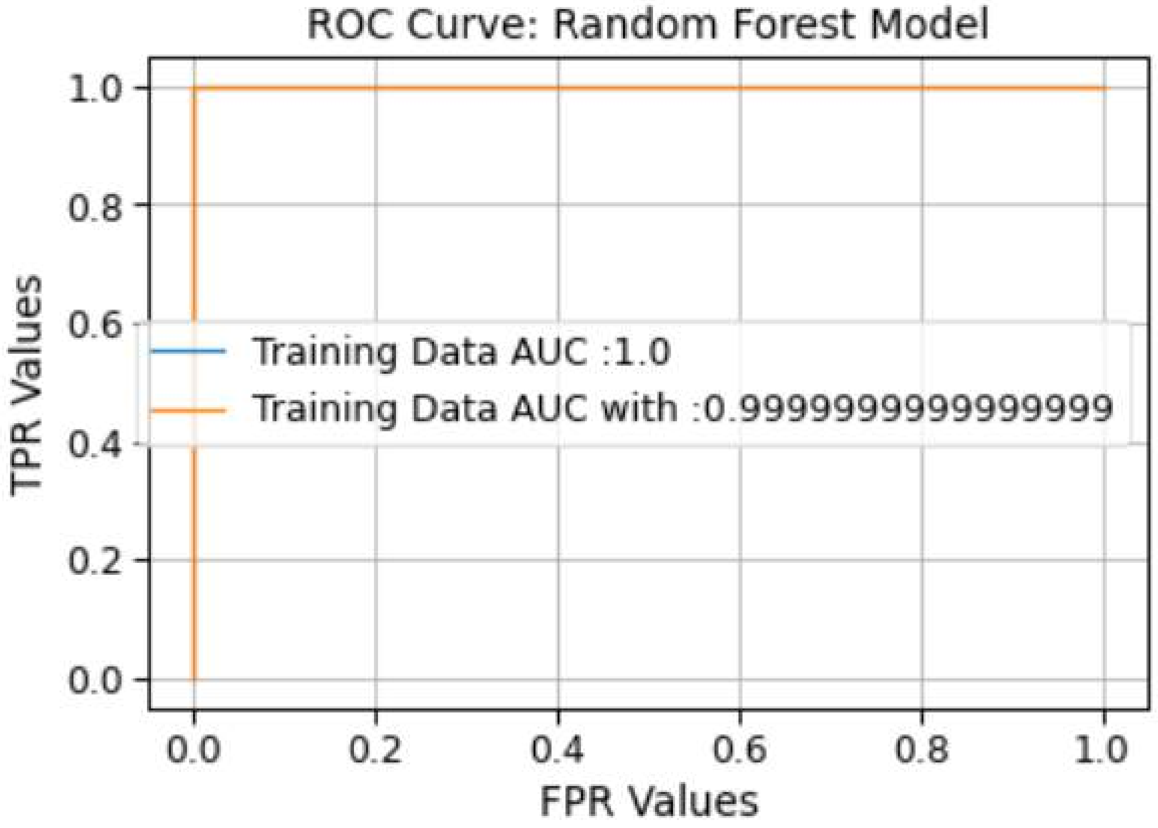
```
Classification Report Training Model (Random Forest):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    183713
           1       1.00      1.00      1.00    183815

    accuracy                           1.00    367528
   macro avg       1.00      1.00      1.00    367528
weighted avg       1.00      1.00      1.00    367528
```

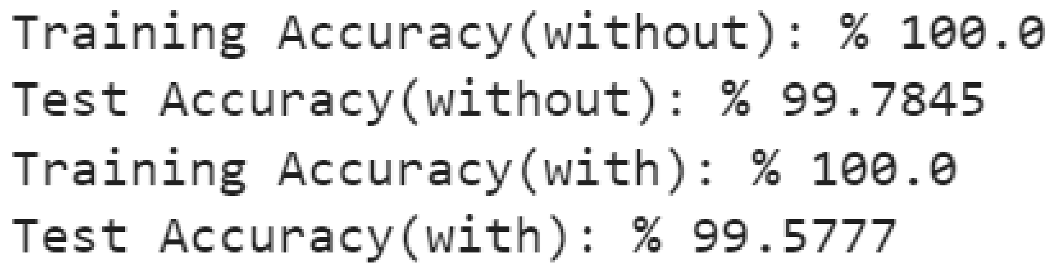
With Feature Selection





```
Classification Report Training Model (Random Forest):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    183713
           1       1.00      1.00      1.00    183815

    accuracy                           1.00    367528
   macro avg       1.00      1.00      1.00    367528
weighted avg       1.00      1.00      1.00    367528
```



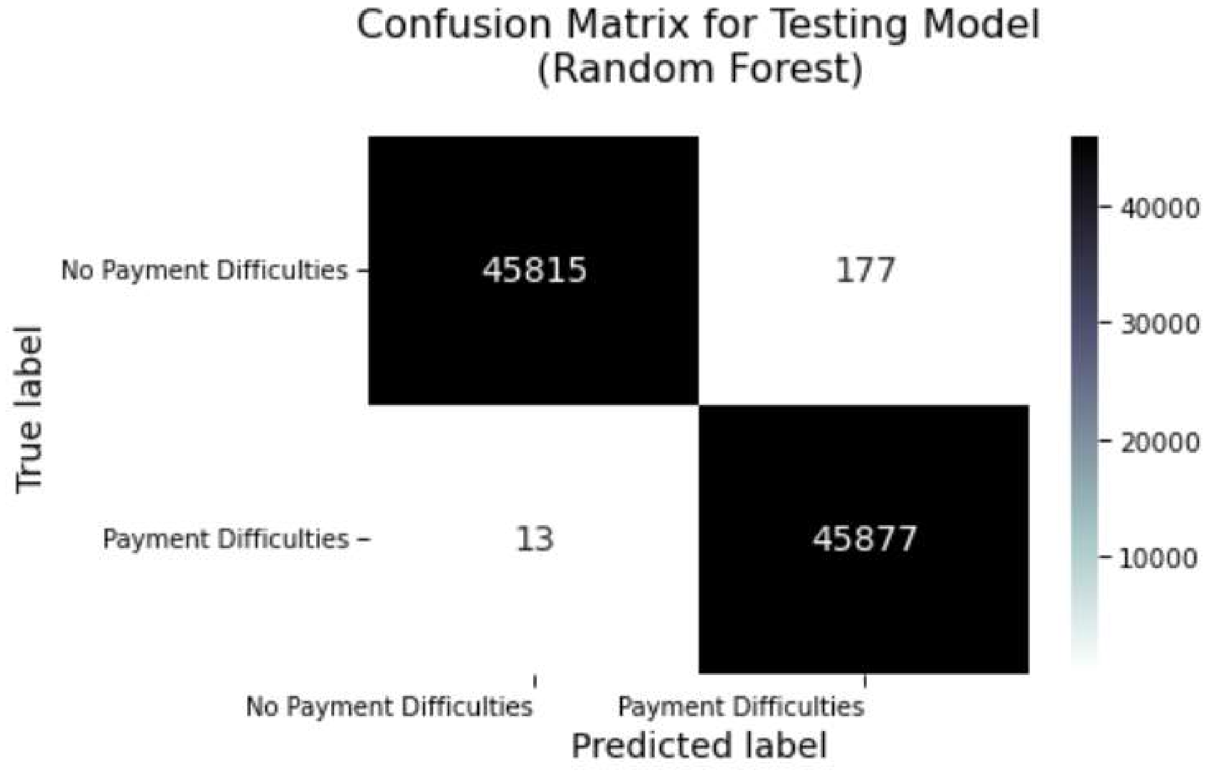


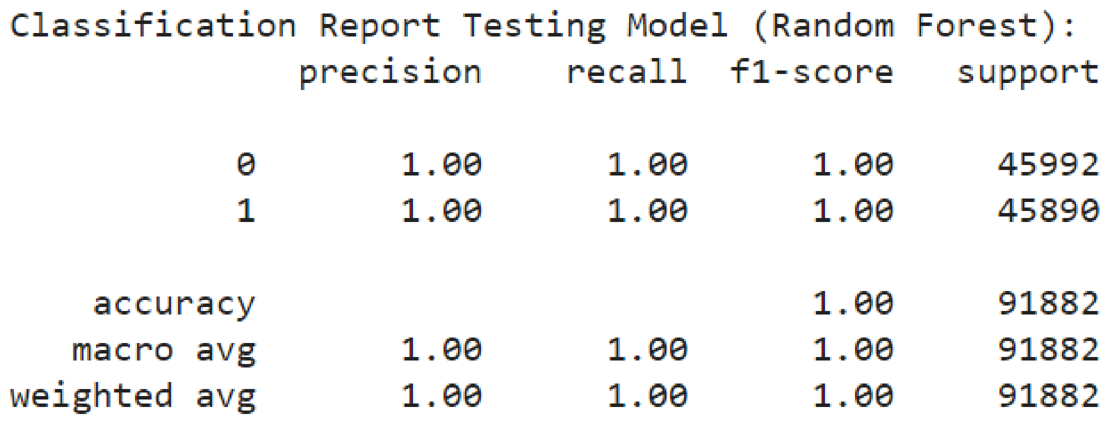
```
Training Accuracy(without): % 100.0
Test Accuracy(without): % 99.7845
Training Accuracy(with): % 100.0
Test Accuracy(with): % 99.5777
```

Machine Learning Modelling

Random Forest

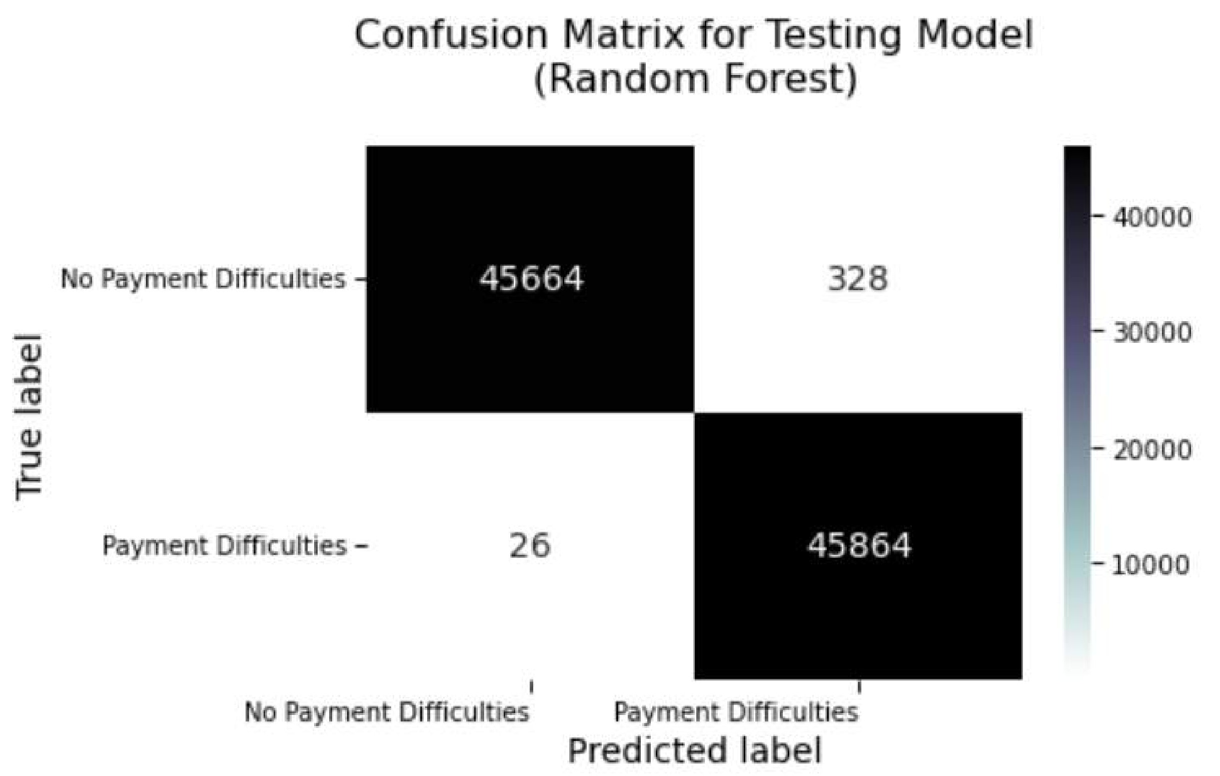
Without Feature Selection

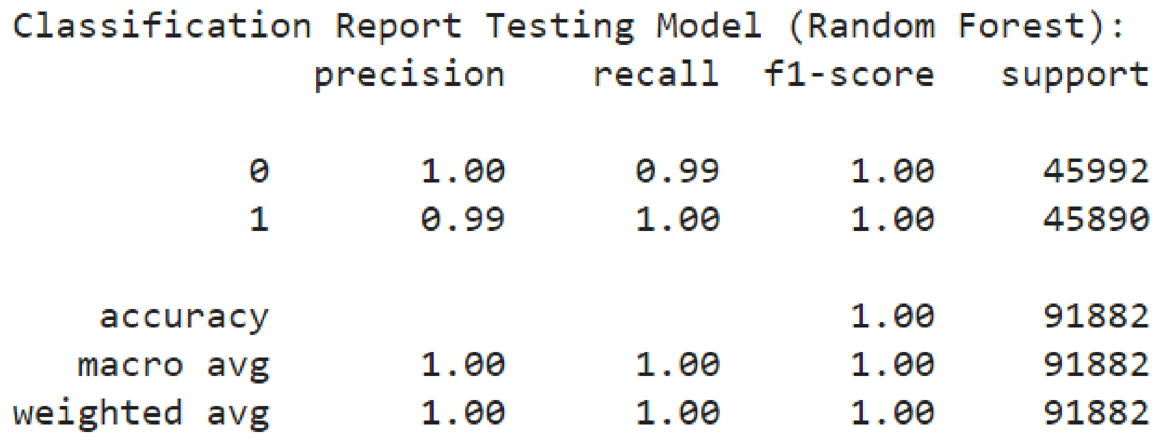




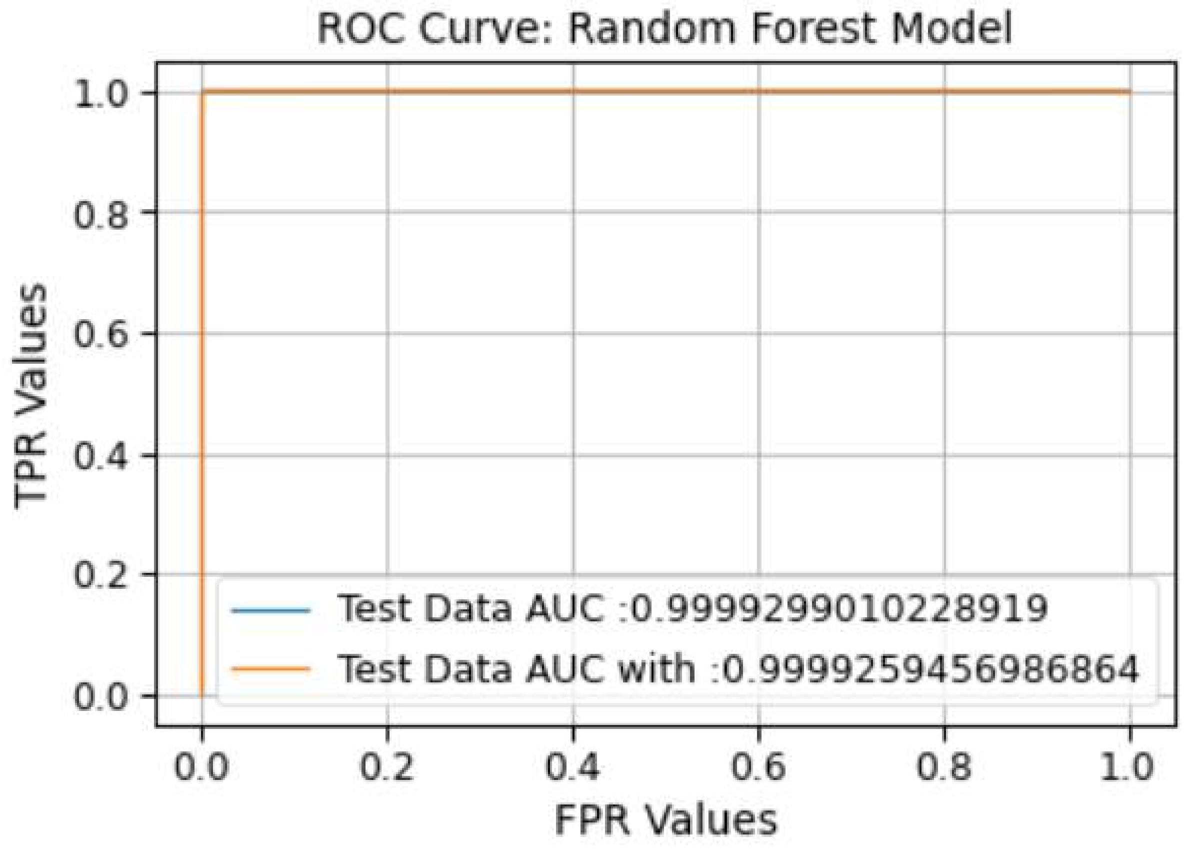
```
Classification Report Testing Model (Random Forest):
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     45992
           1       1.00      1.00      1.00     45890

    accuracy                           1.00     91882
   macro avg       1.00      1.00      1.00     91882
weighted avg       1.00      1.00      1.00     91882
```

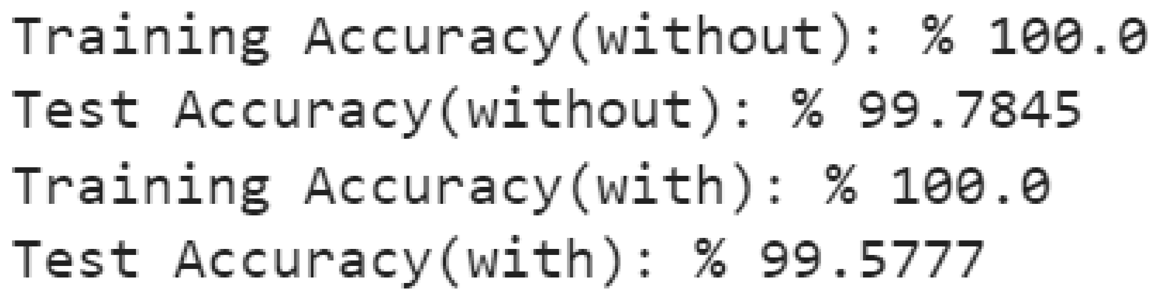
With Feature Selection





```
Classification Report Testing Model (Random Forest):
              precision    recall  f1-score   support

           0       1.00      0.99      1.00     45992
           1       0.99      1.00      1.00     45890

    accuracy                           1.00     91882
   macro avg       1.00      1.00      1.00     91882
weighted avg       1.00      1.00      1.00     91882
```

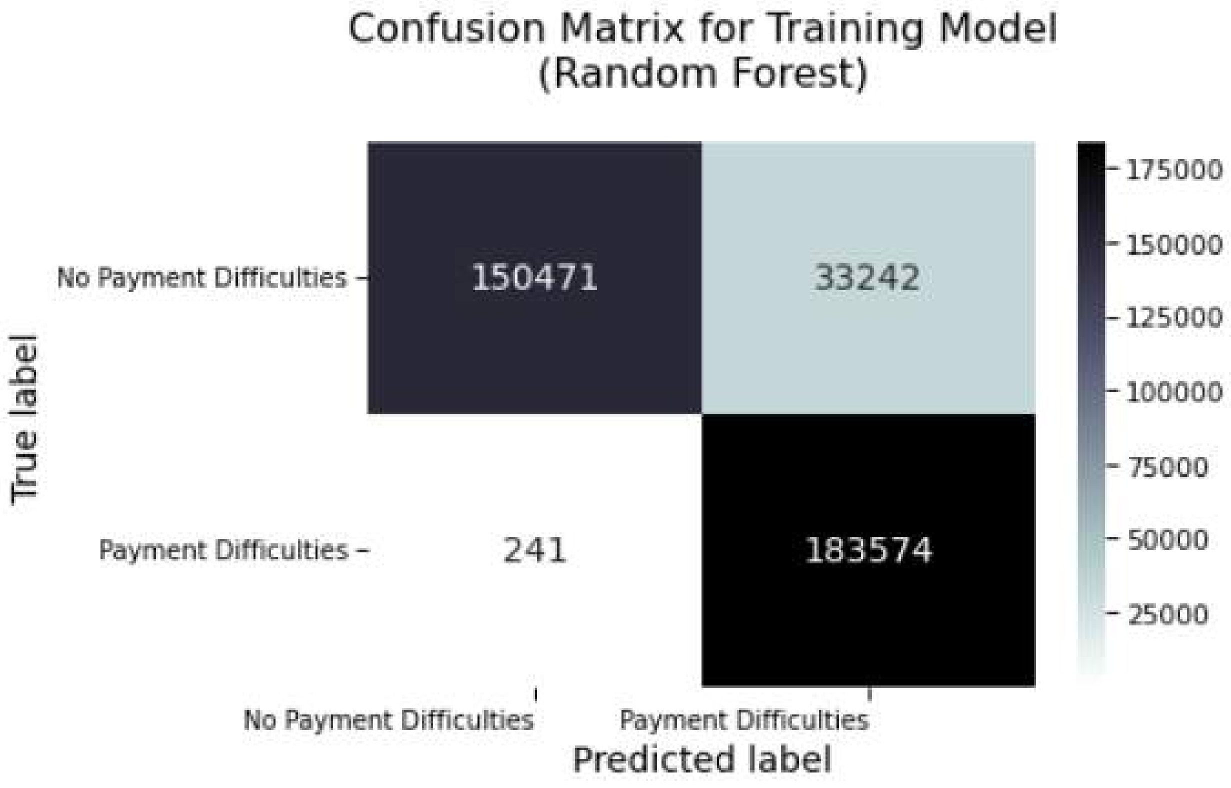




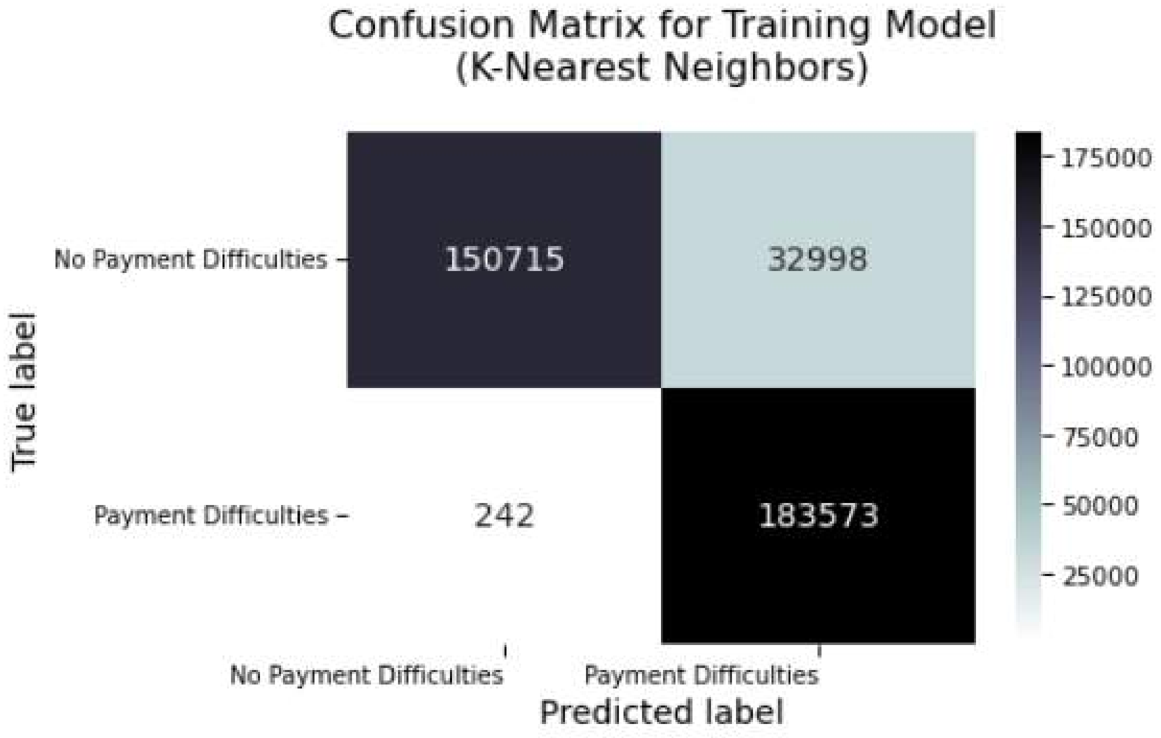
```
Training Accuracy(without): % 100.0
Test Accuracy(without): % 99.7845
Training Accuracy(with): % 100.0
Test Accuracy(with): % 99.5777
```

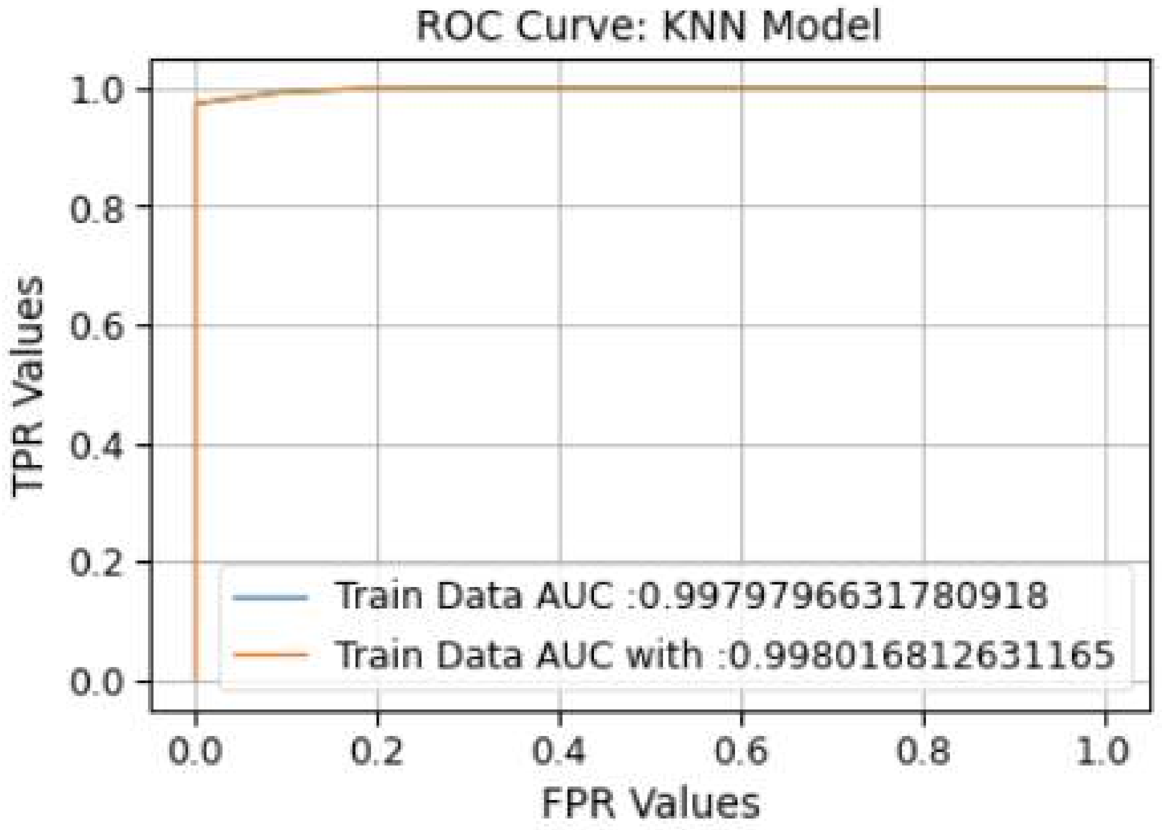
Machine Learning Modelling K-Nearest Neighbors

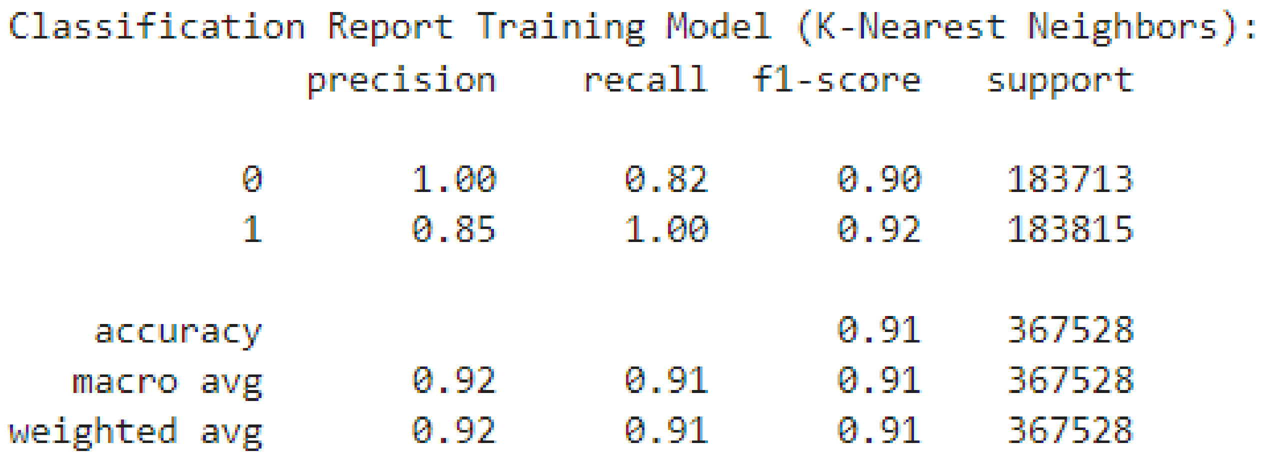
Without Feature Selection



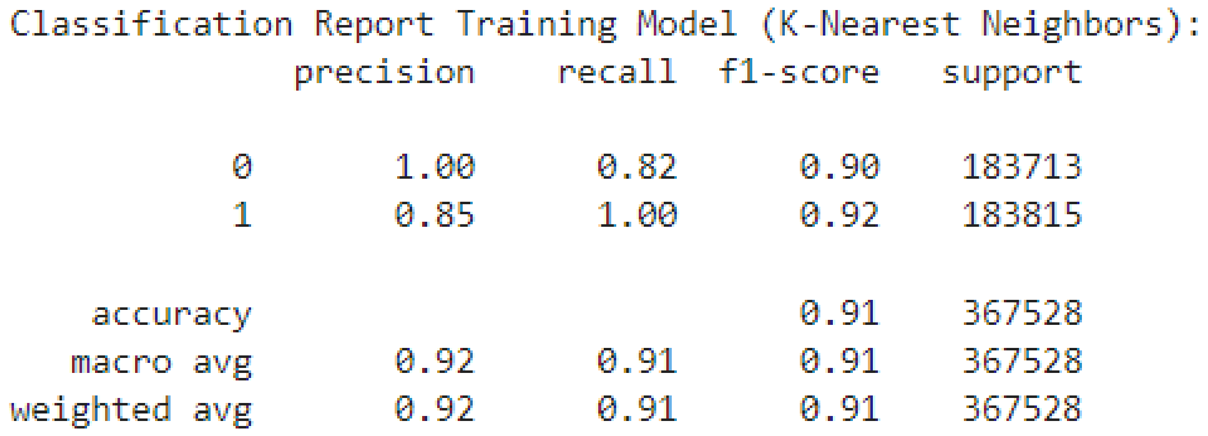
With Feature Selection



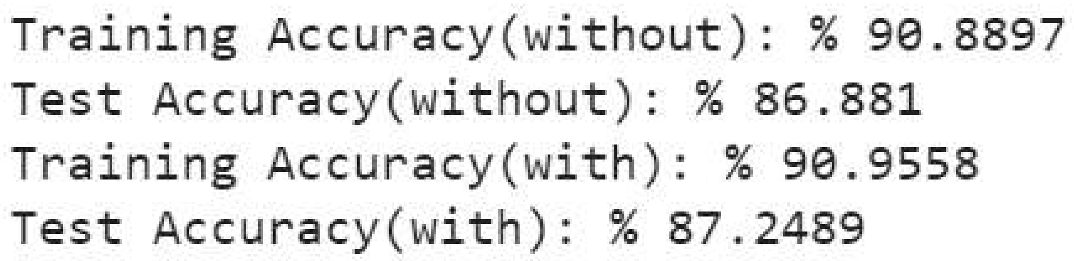




```
Classification Report Training Model (K-Nearest Neighbors):
              precision    recall  f1-score   support

           0       1.00      0.82      0.90    183713
           1       0.85      1.00      0.92    183815

    accuracy                           0.91    367528
   macro avg       0.92      0.91      0.91    367528
weighted avg       0.92      0.91      0.91    367528
```



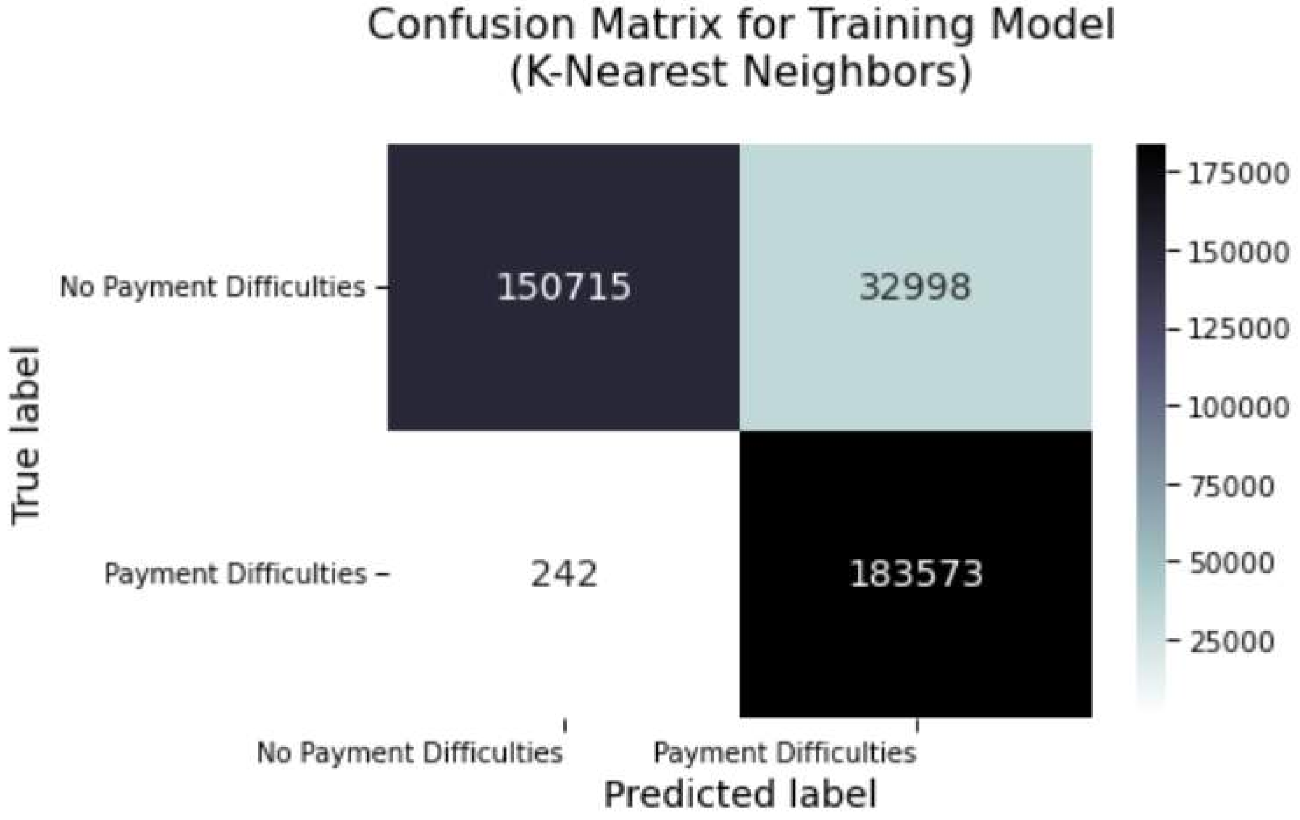
```
Classification Report Training Model (K-Nearest Neighbors):
              precision    recall  f1-score   support

           0       1.00      0.82      0.90    183713
           1       0.85      1.00      0.92    183815

    accuracy                           0.91    367528
   macro avg       0.92      0.91      0.91    367528
weighted avg       0.92      0.91      0.91    367528
```

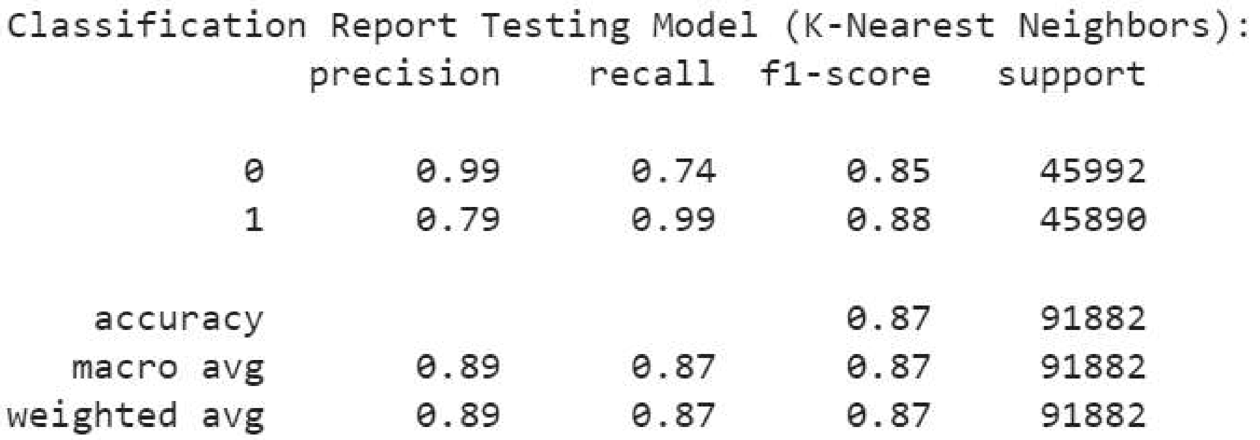


```
Training Accuracy(without): % 90.8897
Test Accuracy(without): % 86.881
Training Accuracy(with): % 90.9558
Test Accuracy(with): % 87.2489
```

Machine Learning Modelling K-Nearest Neighbors

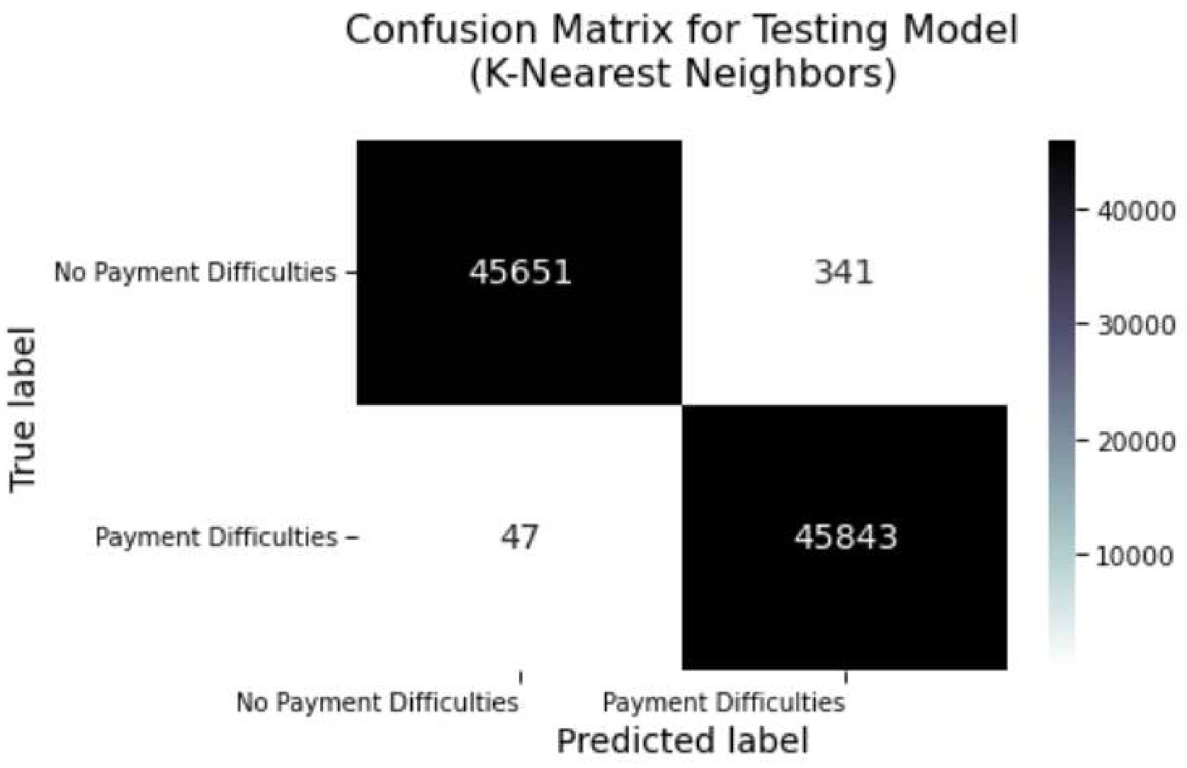
Without Feature Selection

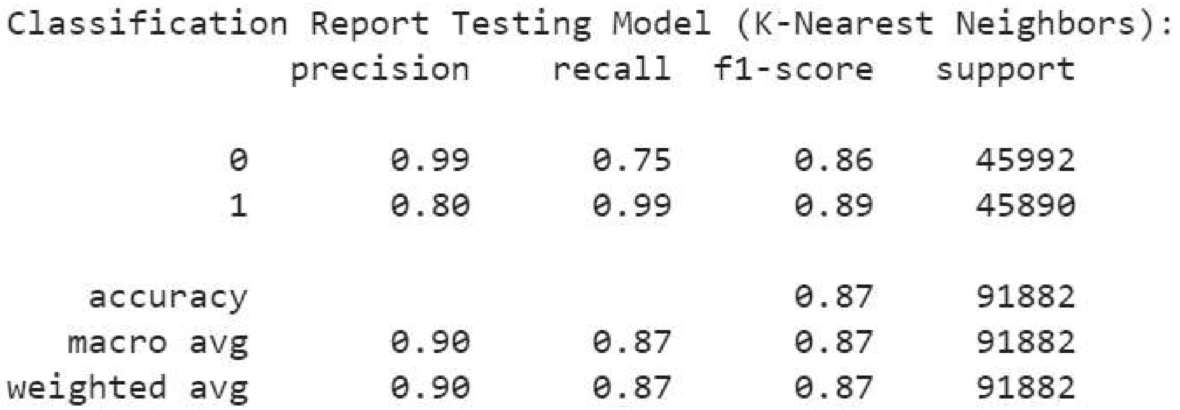




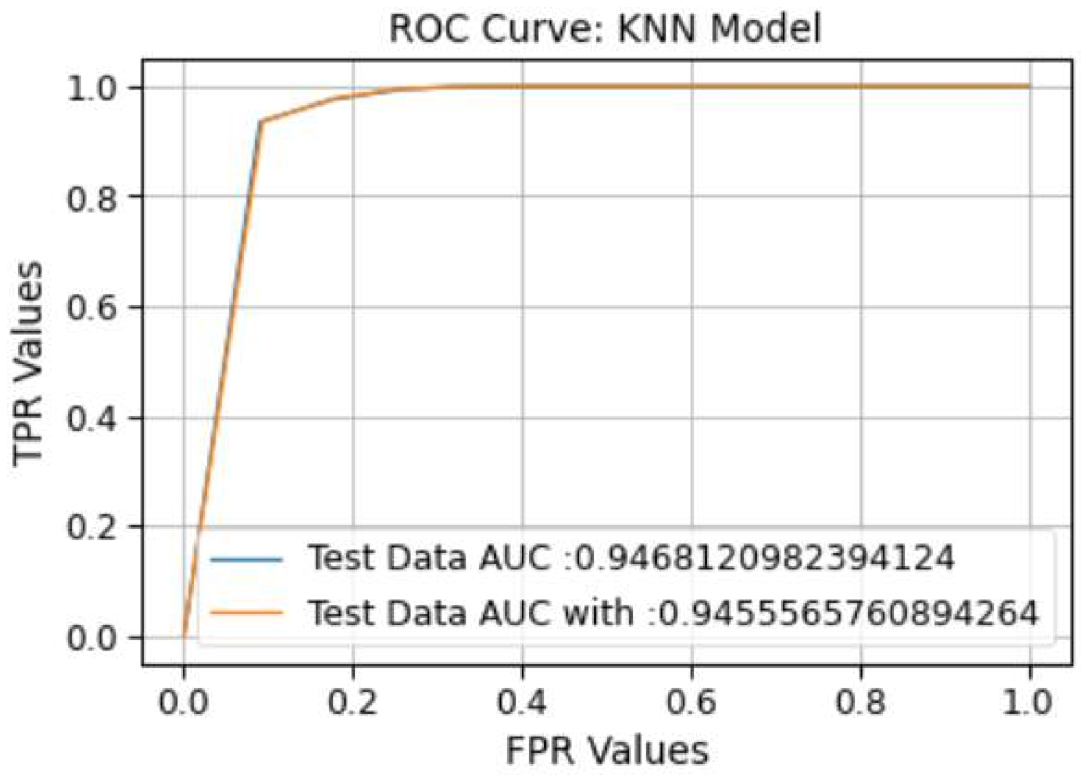
```
Classification Report Testing Model (K-Nearest Neighbors):
              precision    recall  f1-score   support

           0       0.99      0.74      0.85     45992
           1       0.79      0.99      0.88     45890

    accuracy                           0.87     91882
   macro avg       0.89      0.87      0.87     91882
weighted avg       0.89      0.87      0.87     91882
```

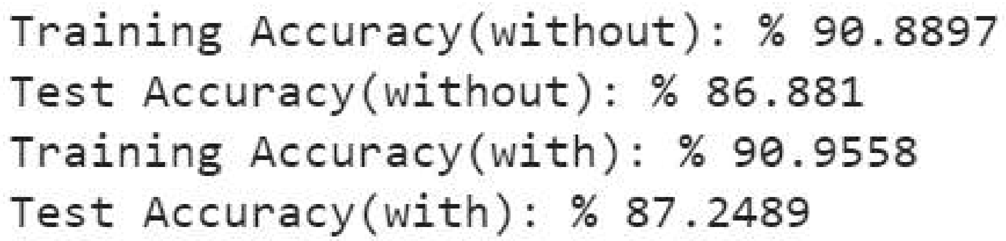
With Feature Selection





```
Classification Report Testing Model (K-Nearest Neighbors):
              precision    recall  f1-score   support

           0       0.99      0.75      0.86     45992
           1       0.80      0.99      0.89     45890

    accuracy                           0.87     91882
   macro avg       0.90      0.87      0.87     91882
weighted avg       0.90      0.87      0.87     91882
```

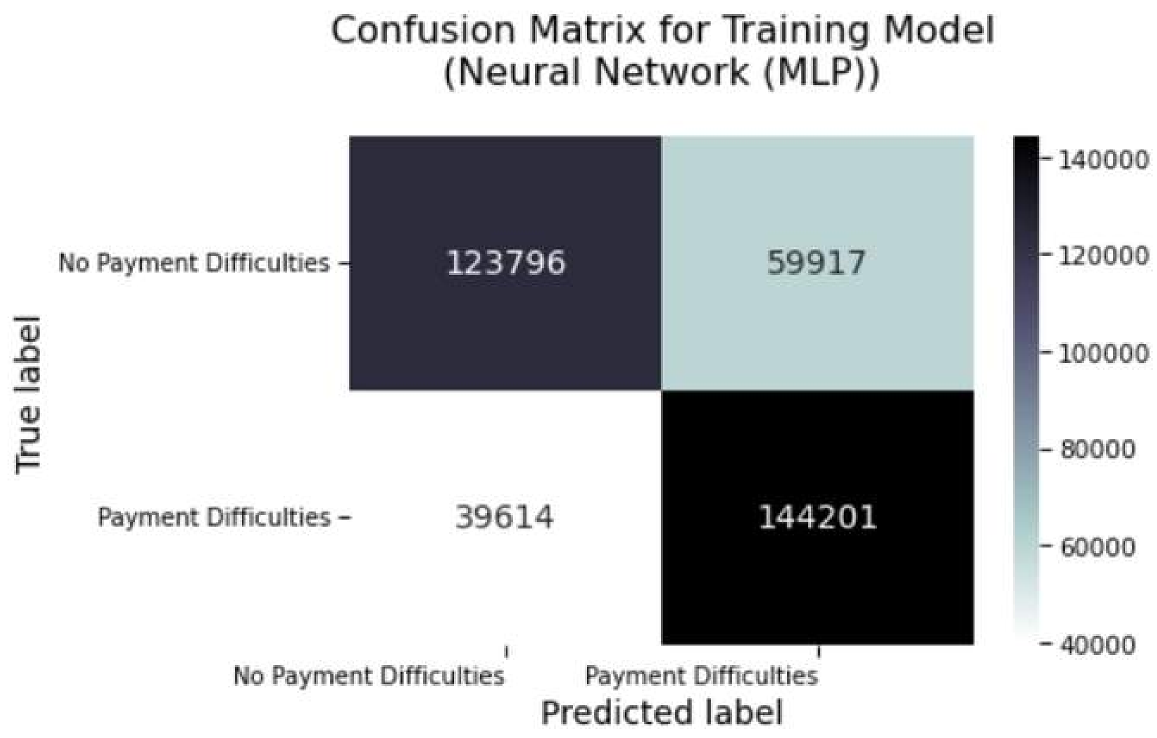


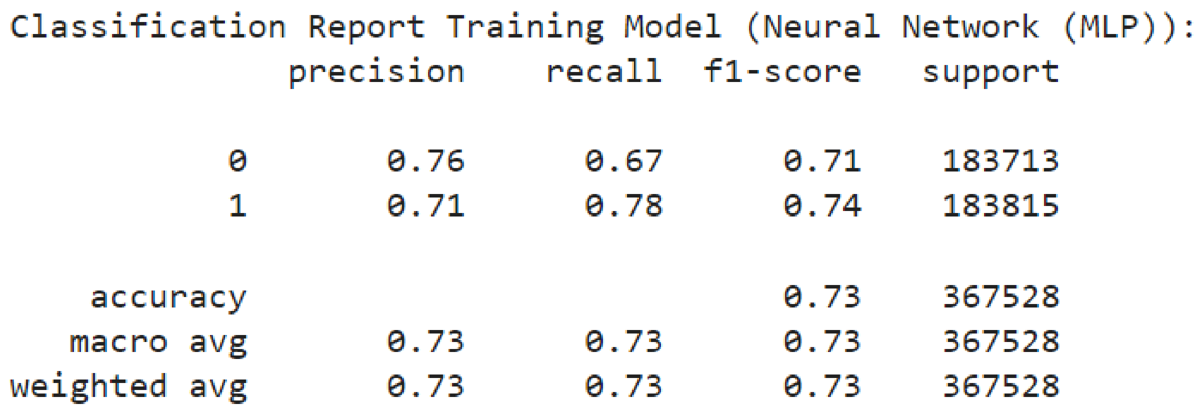


```
Training Accuracy(without): % 90.8897
Test Accuracy(without): % 86.881
Training Accuracy(with): % 90.9558
Test Accuracy(with): % 87.2489
```

Machine Learning Modelling Neural Network (Multi-layer Perceptron)

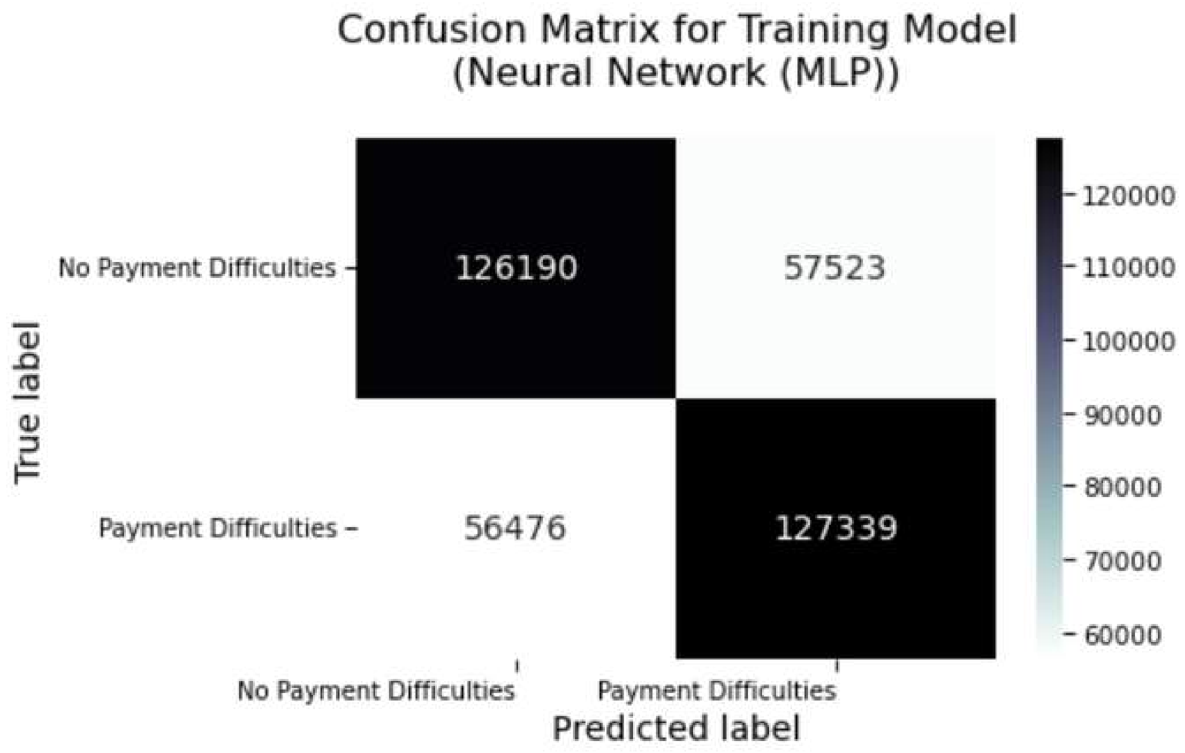
Without Feature Selection

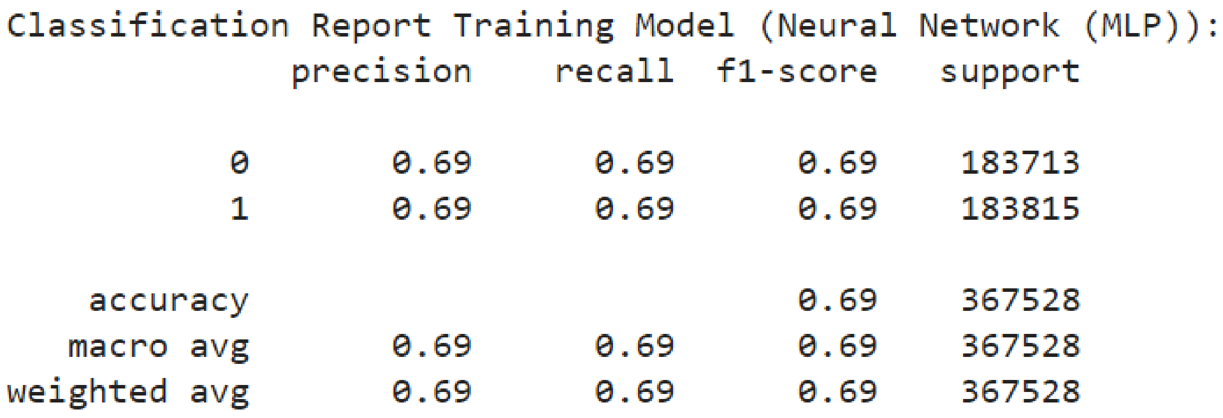




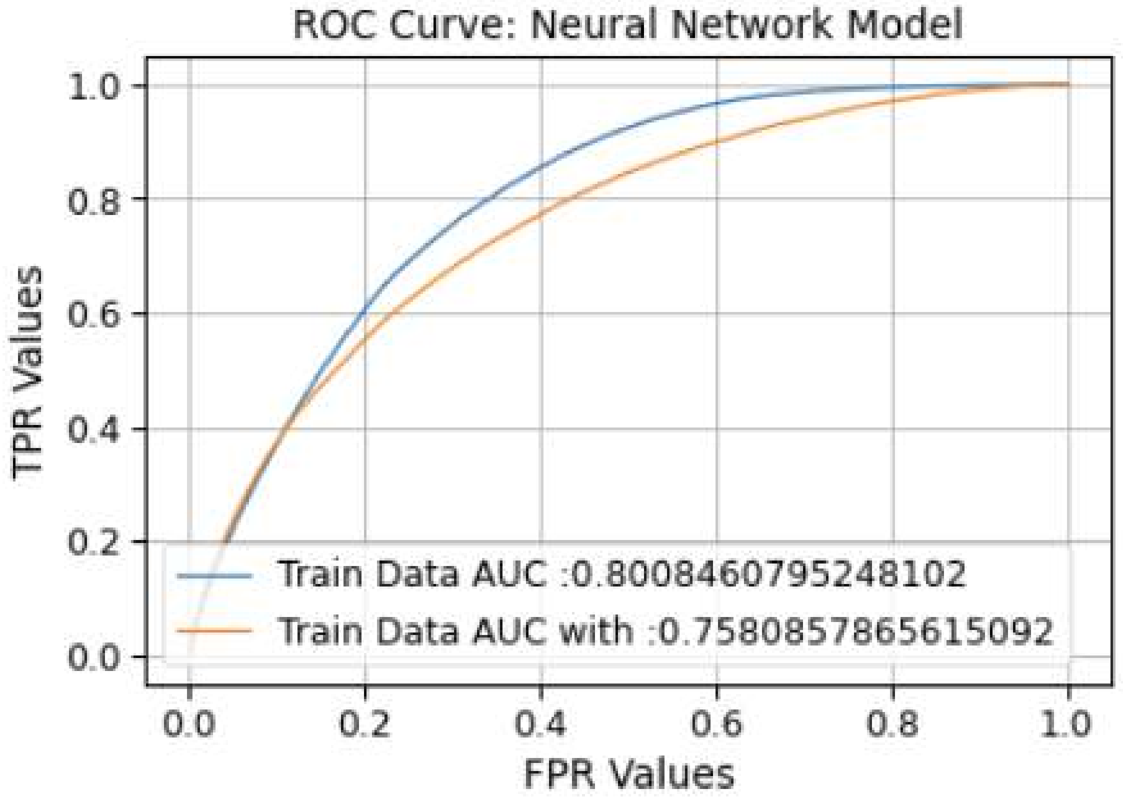
```
Classification Report Training Model (Neural Network (MLP)):
              precision    recall  f1-score   support

           0       0.76      0.67      0.71    183713
           1       0.71      0.78      0.74    183815

    accuracy                           0.73    367528
   macro avg       0.73      0.73      0.73    367528
weighted avg       0.73      0.73      0.73    367528
```

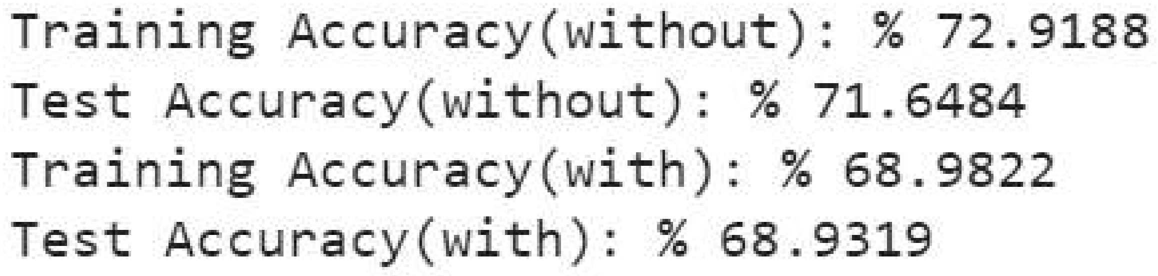
With Feature Selection





```
Classification Report Training Model (Neural Network (MLP)):
              precision    recall  f1-score   support

           0       0.69      0.69      0.69    183713
           1       0.69      0.69      0.69    183815

    accuracy                           0.69    367528
   macro avg       0.69      0.69      0.69    367528
weighted avg       0.69      0.69      0.69    367528
```

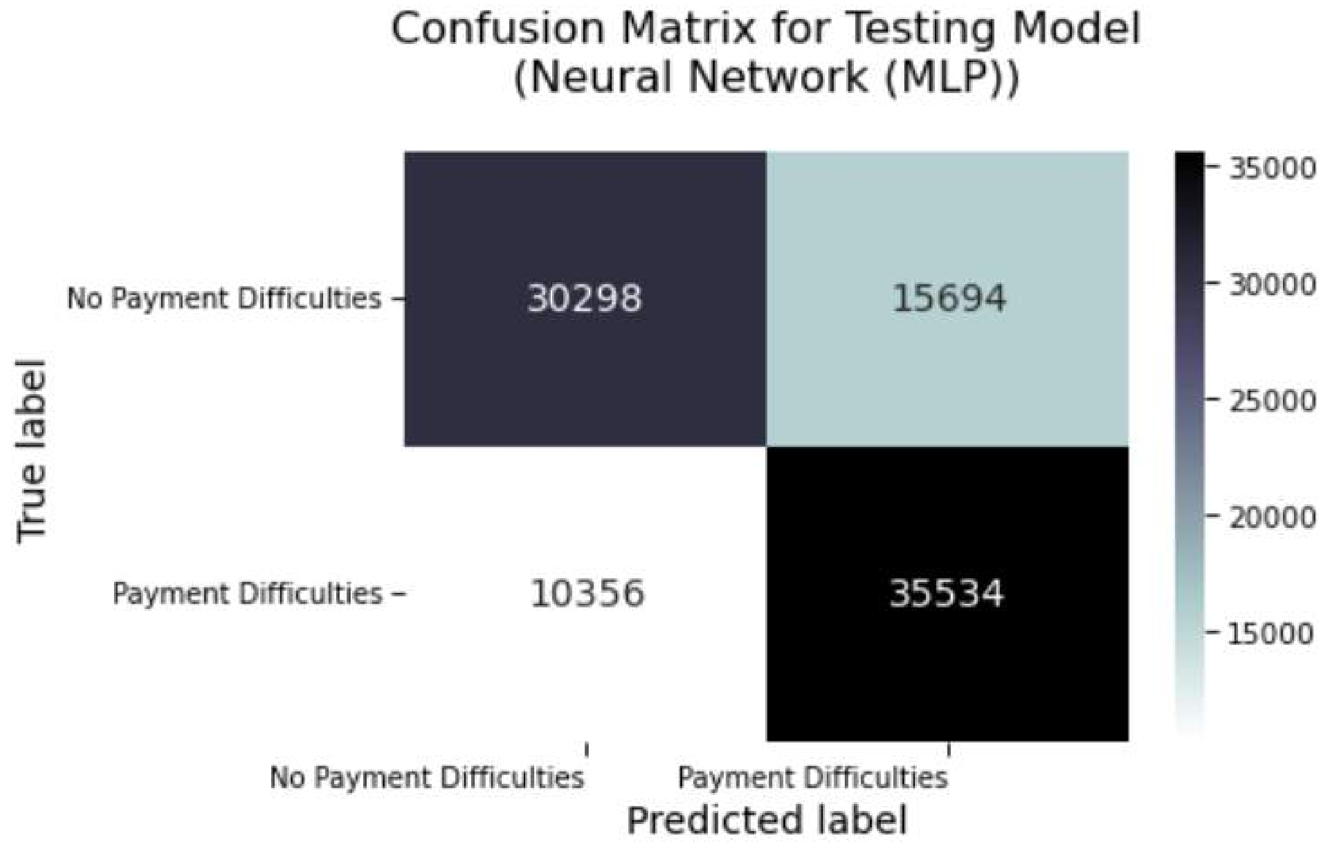


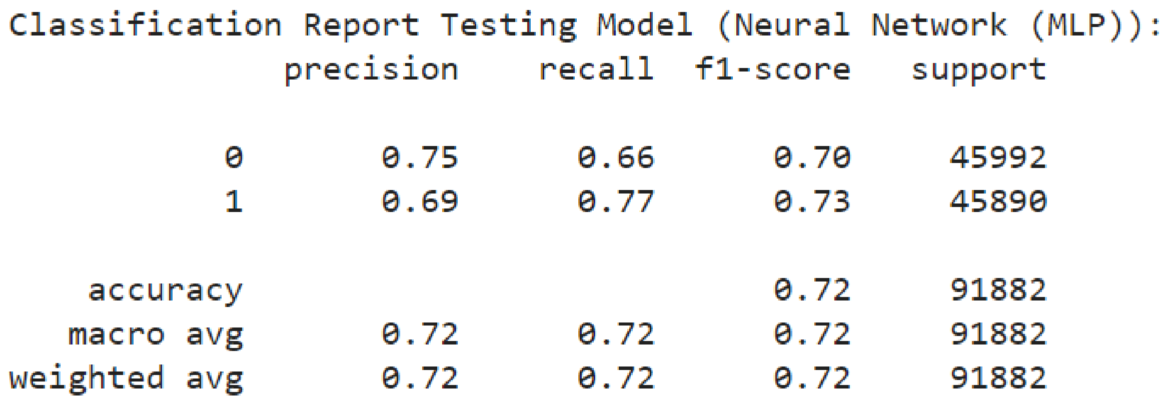


```
Training Accuracy(without): % 72.9188
Test Accuracy(without): % 71.6484
Training Accuracy(with): % 68.9822
Test Accuracy(with): % 68.9319
```

Machine Learning Modelling Neural Network (Multi-layer Perceptron)

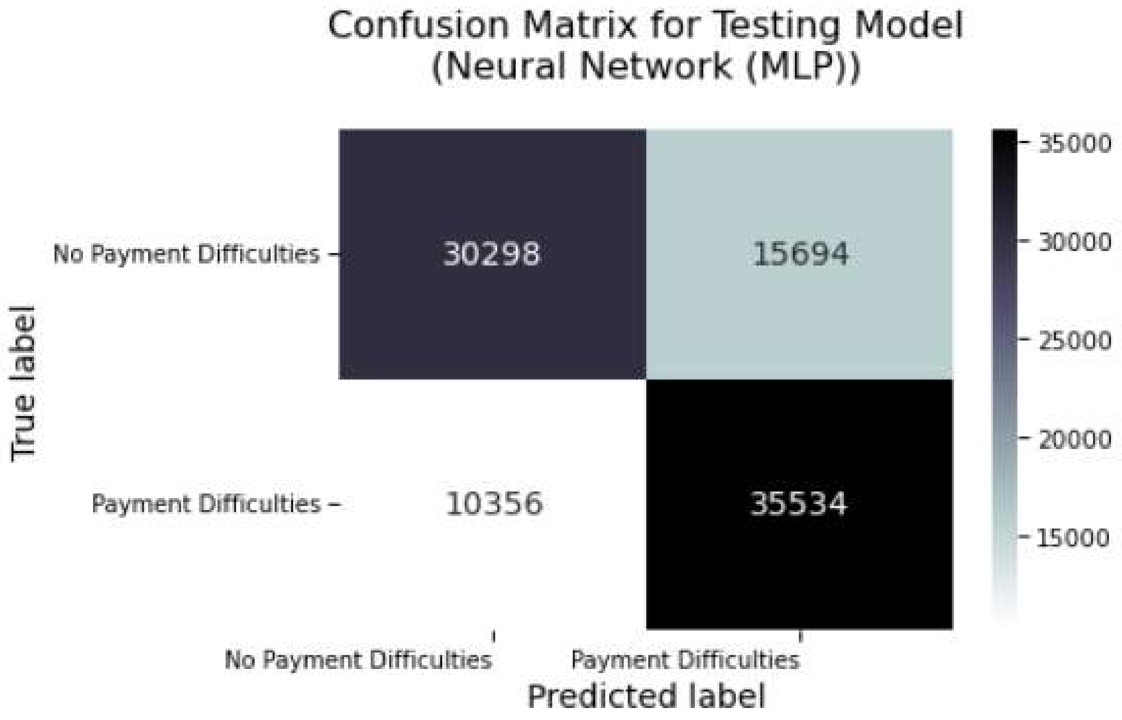
Without Feature Selection

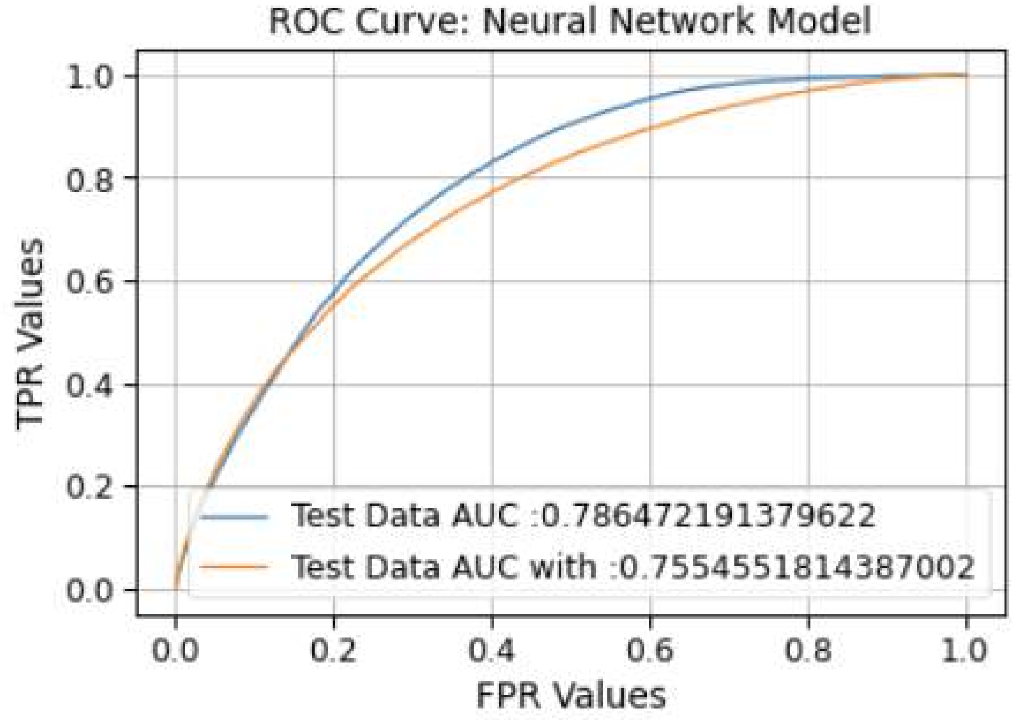


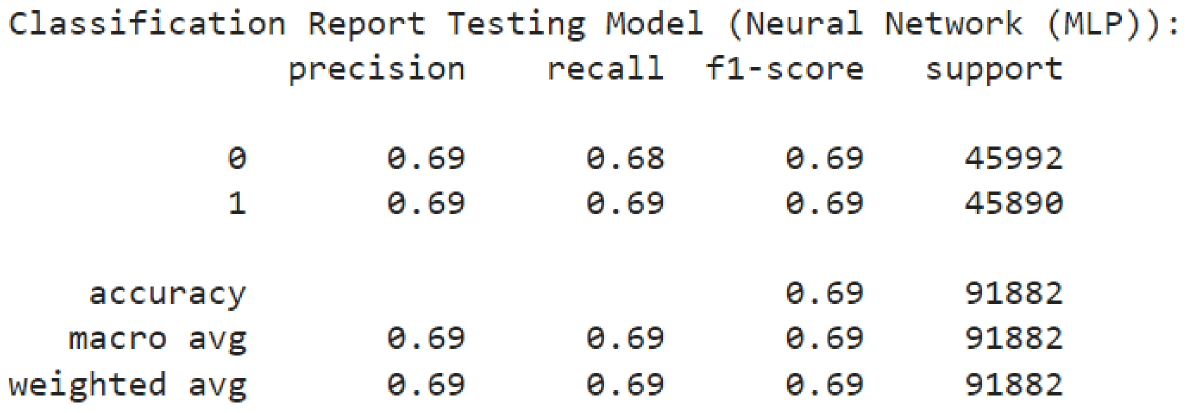


```
Classification Report Testing Model (Neural Network (MLP)):
              precision    recall  f1-score   support

           0       0.75      0.66      0.70     45992
           1       0.69      0.77      0.73     45890

    accuracy                           0.72     91882
   macro avg       0.72      0.72      0.72     91882
weighted avg       0.72      0.72      0.72     91882
```

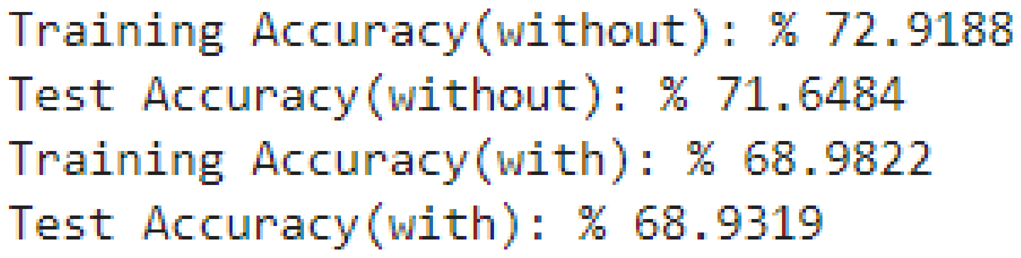
With Feature Selection





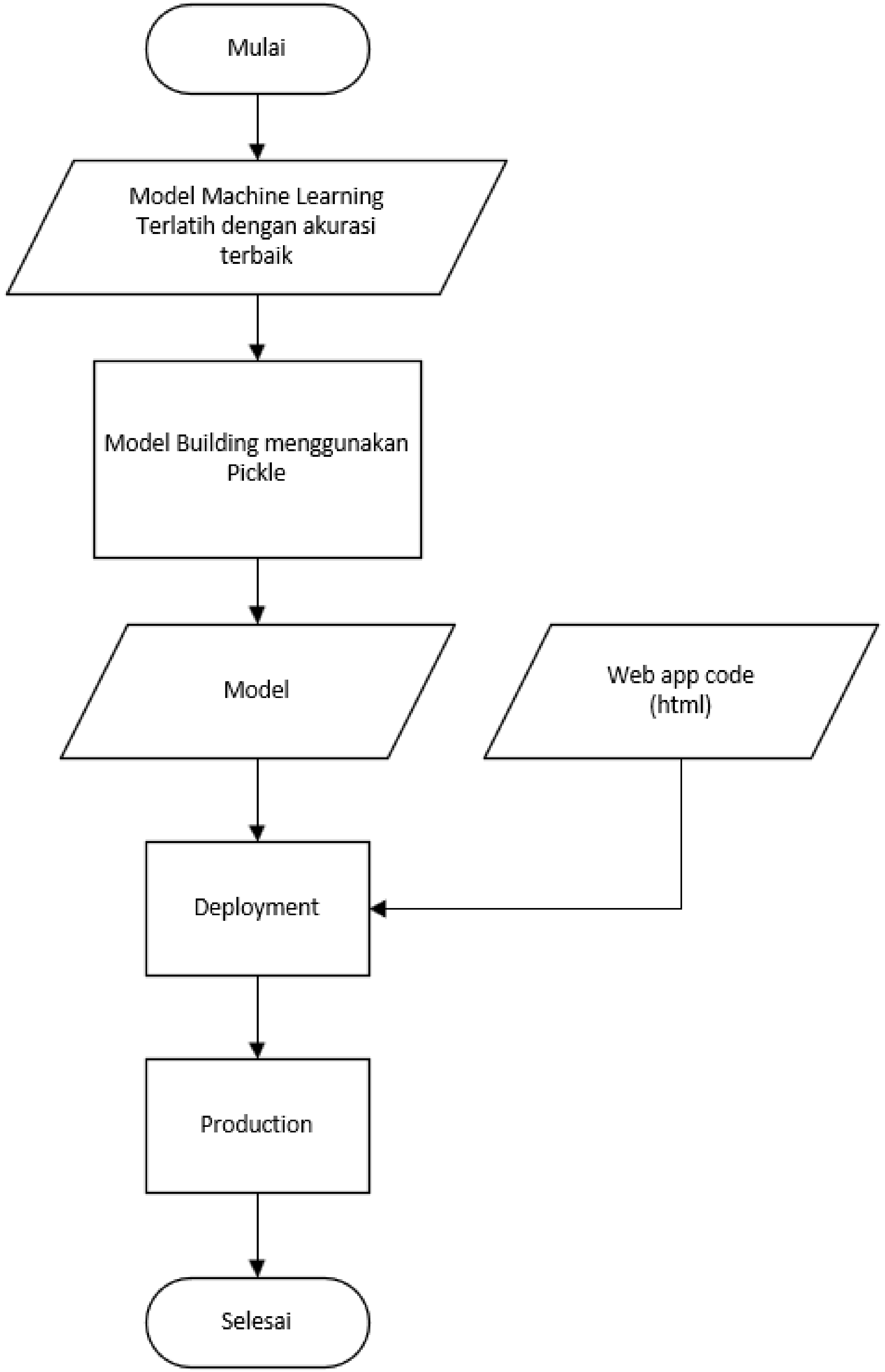


```
Classification Report Testing Model (Neural Network (MLP)):
              precision    recall  f1-score   support

           0       0.69      0.68      0.69     45992
           1       0.69      0.69      0.69     45890

    accuracy                           0.69     91882
   macro avg       0.69      0.69      0.69     91882
weighted avg       0.69      0.69      0.69     91882
```

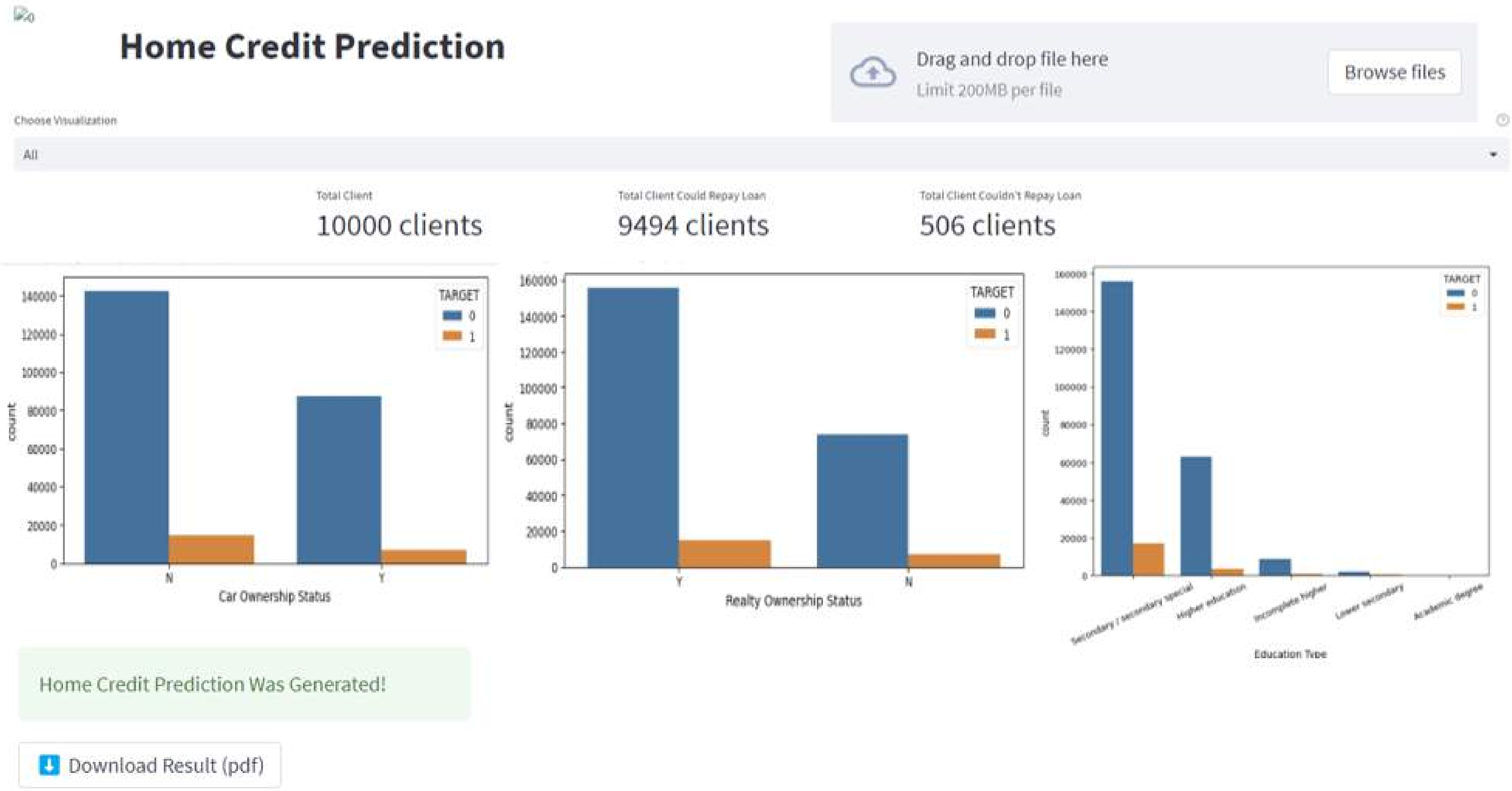


```
Training Accuracy(without): % 72.9188
Test Accuracy(without): % 71.6484
Training Accuracy(with): % 68.9822
Test Accuracy(with): % 68.9319
```

Model Deployment Design



Mulai
Model Machine Learning Terlatih dengan akurasi terbaik
Model Building menggunakan Pickle
Model
Web app code (html)
Deployment
Production
Selesai



Home Credit Prediction

Drag and drop file here

Limit 200MB per file

Choose Visualization

All

Total Client

10000 clients

Total Client Could Repay Loan

9494 clients

Total Client Couldn't Repay Loan

506 clients

Home Credit Prediction Was Generated!