

Paylater loan data

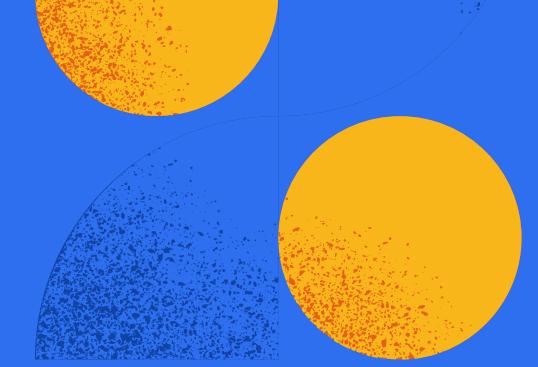
Kaggle is the world's largest data science community with powerful tools and resources to help you achieve your data science goals.

kaggle.com

Project BCC 2023 Data Science

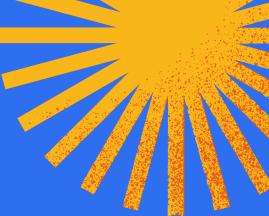
Final Pitching

Kelompok 5



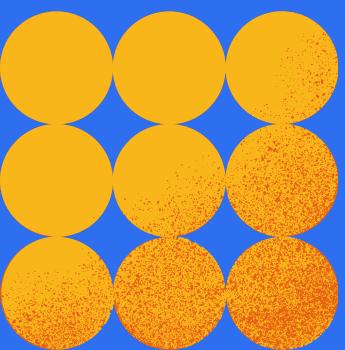
1. Memahami Dataset





Data Set

Analisis Pinjaman Paylater



Masalah

Banyak Klien gagal membayar tepat waktu

Solusi

Analisis Faktor yang berkontribusi terhadap
Kemampuan Membayar Pinjaman



Manfaat Pengolahan Dataset



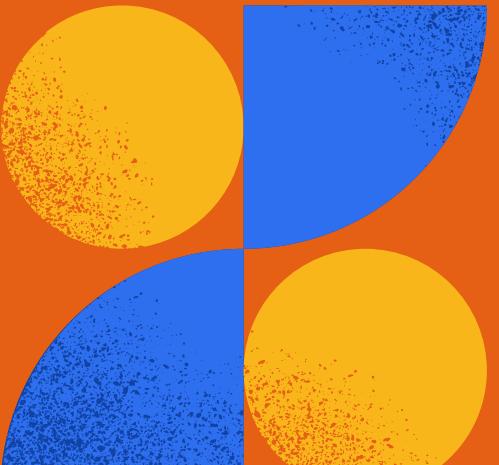
Penentuan Risiko Kredit yang lebih Akurat



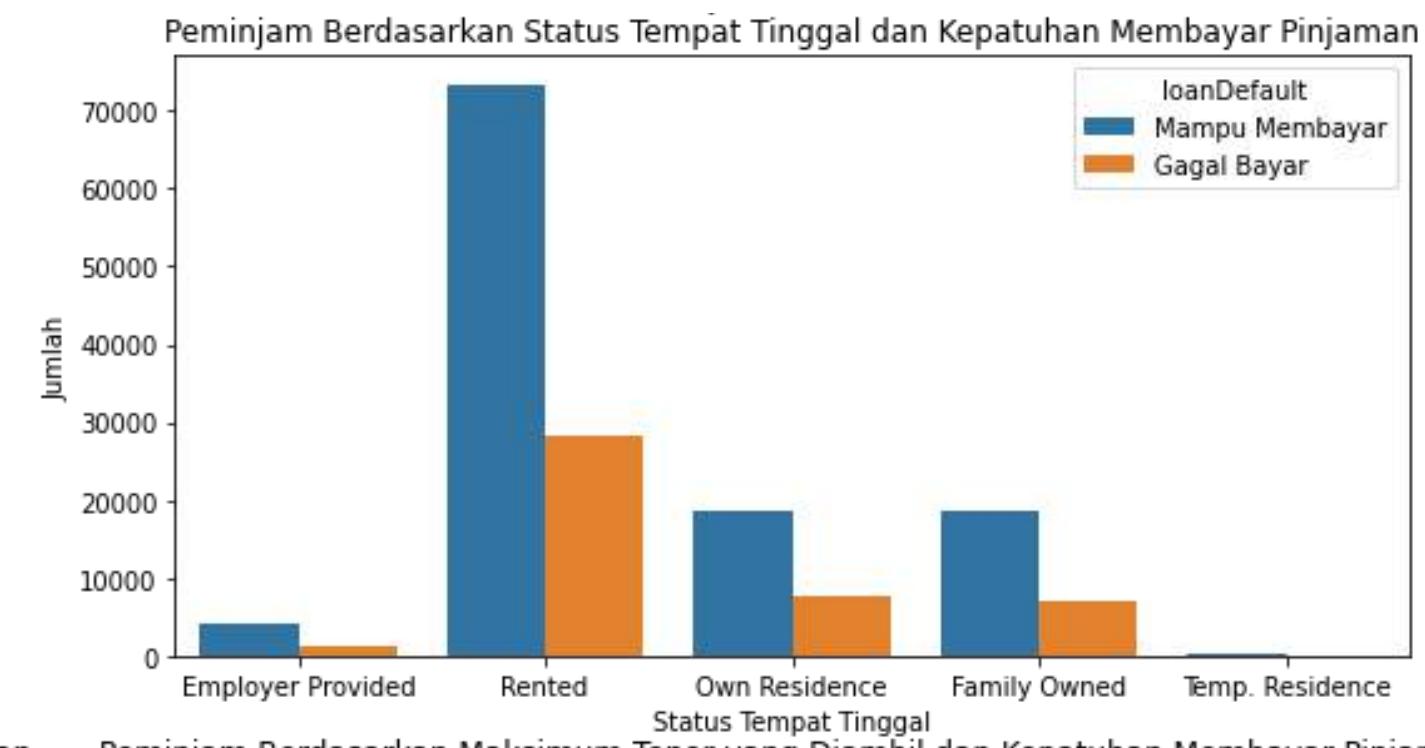
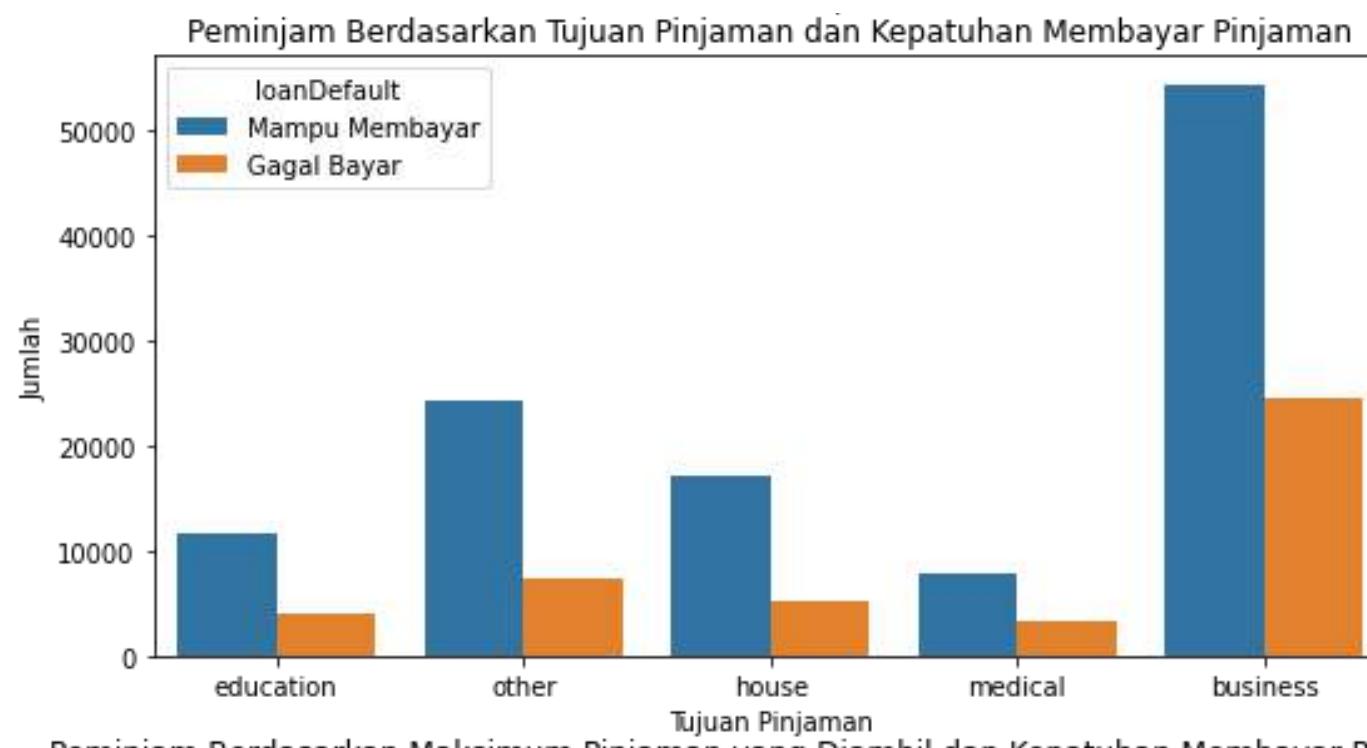
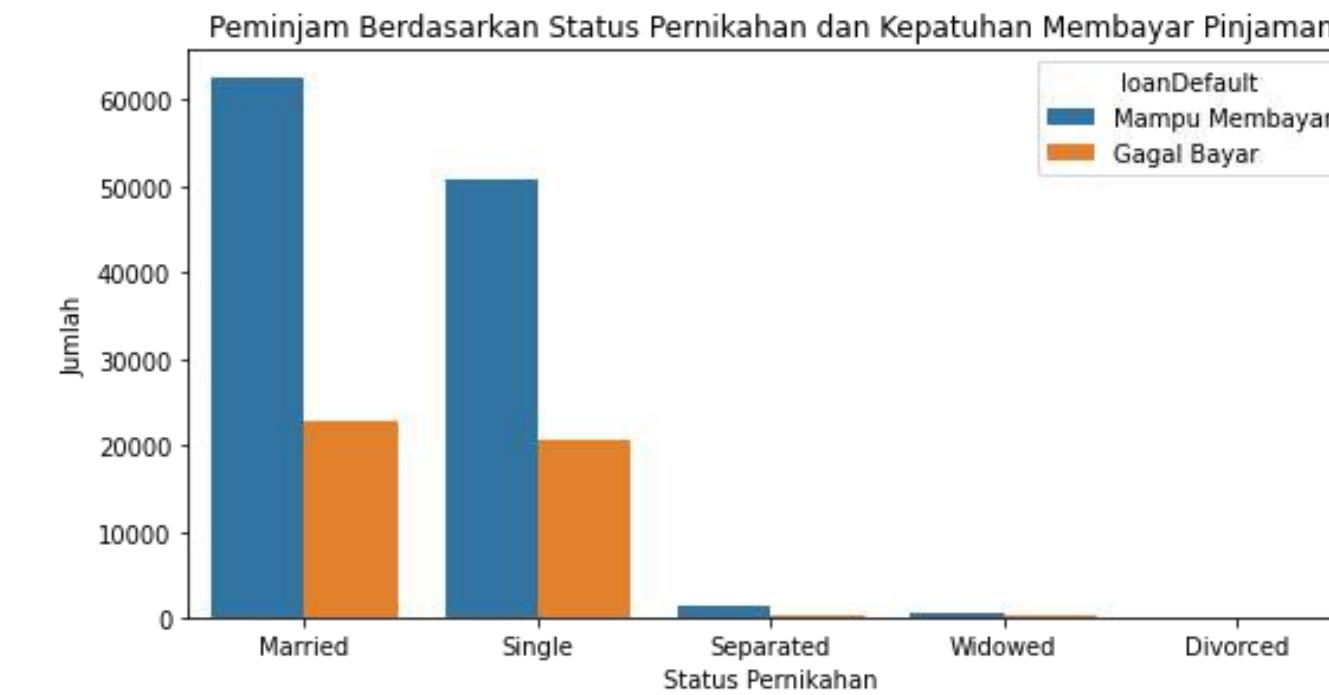
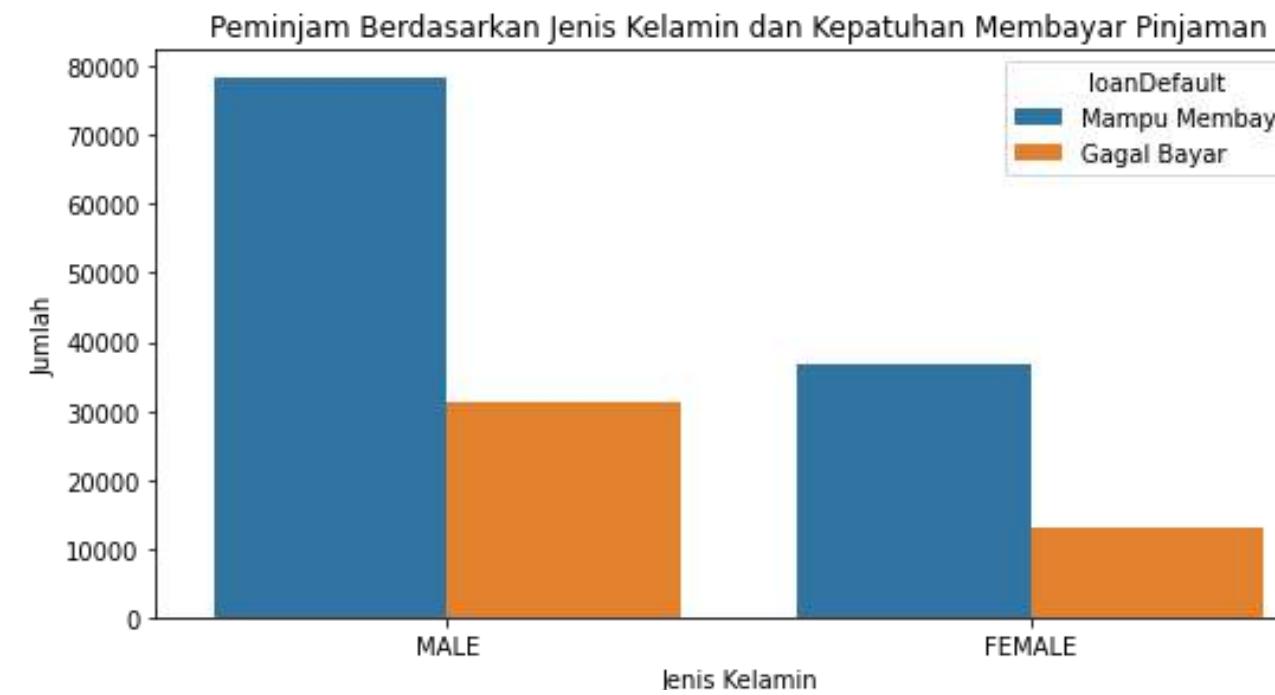
Peningkatan Profit Perusahaan dengan
Analisis Profil Pelanggan



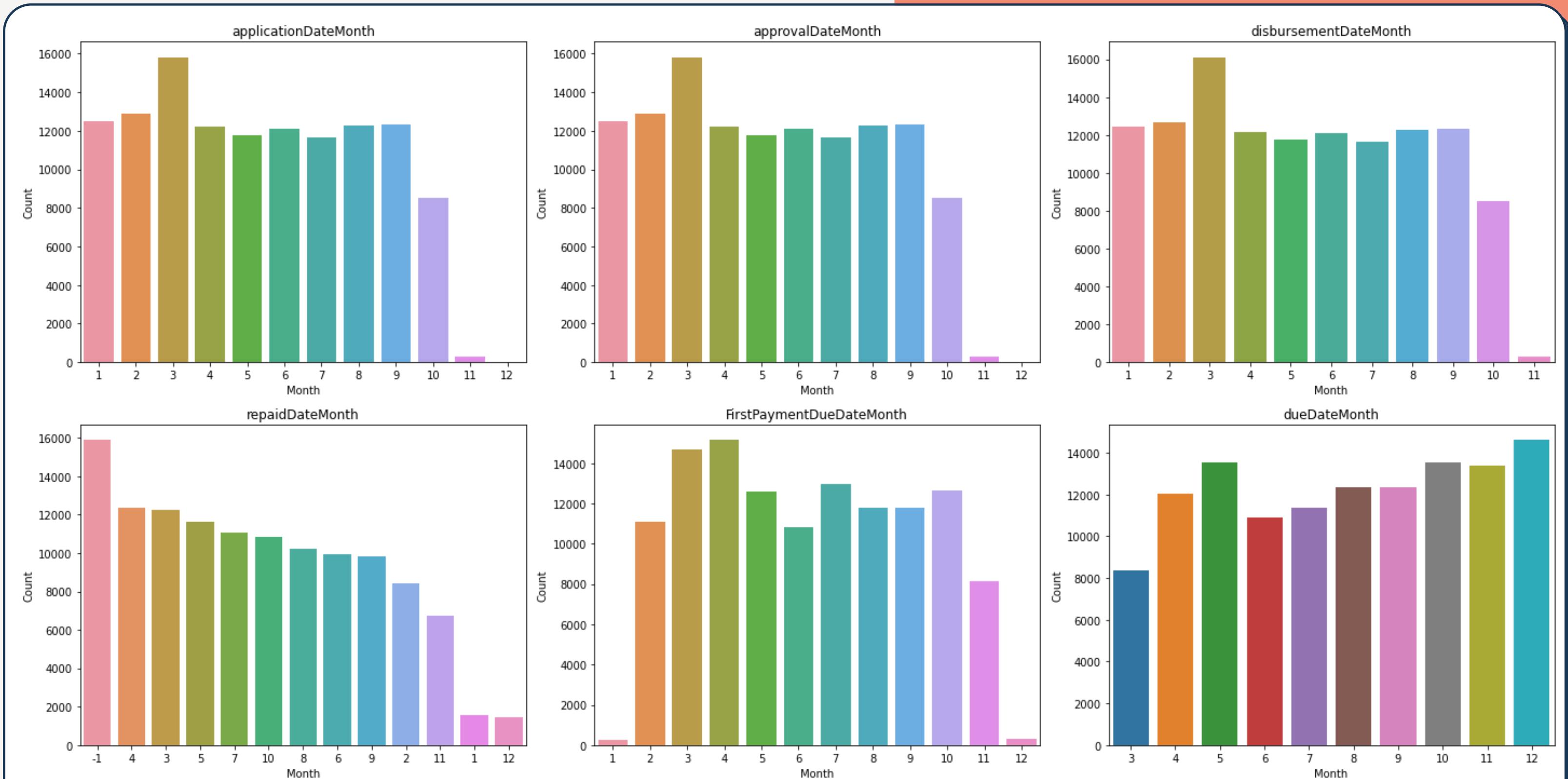
2. Eksplorasi Data



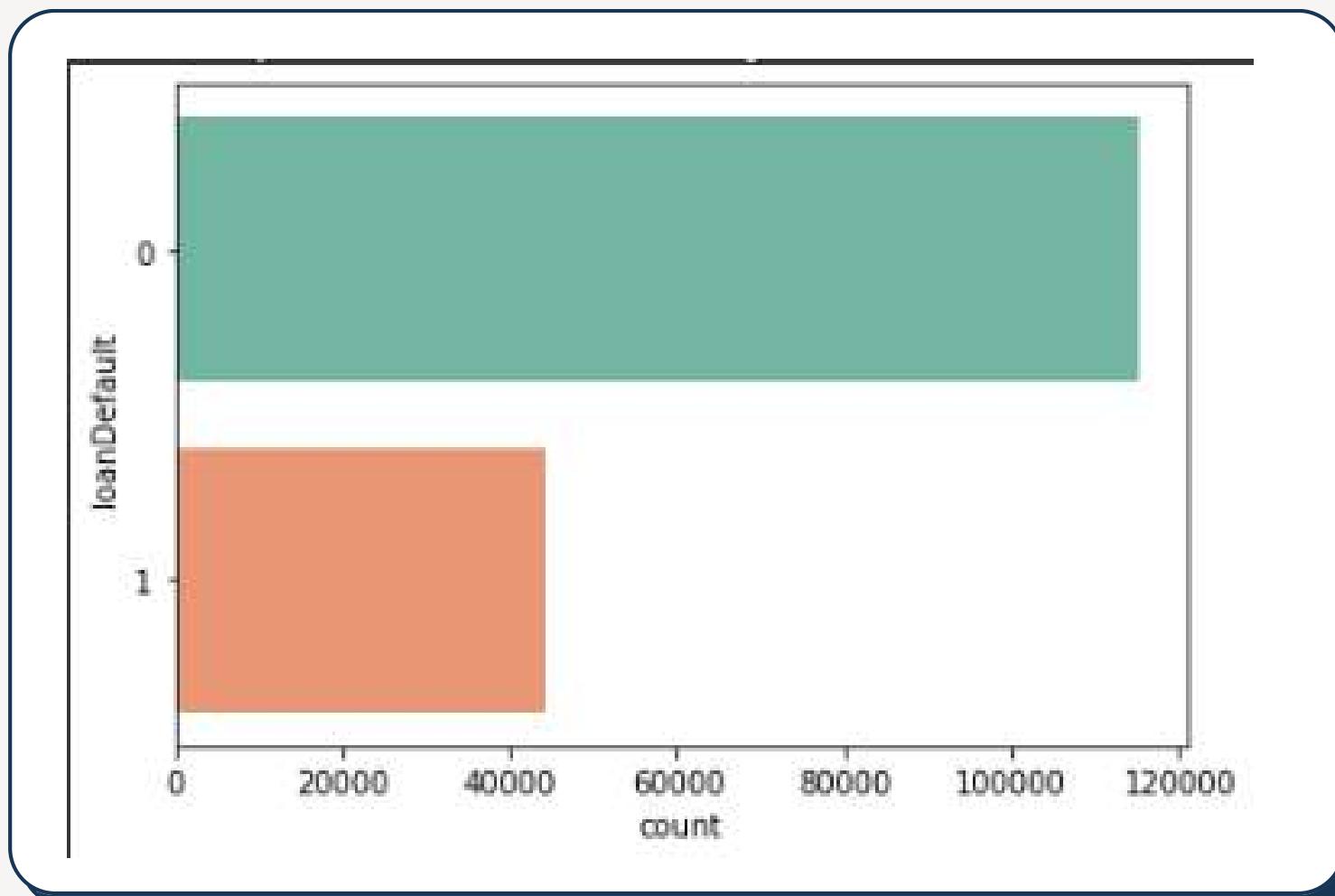
Visualisasi Atribut Profil Pelanggan



Visualisasi Atribut Waktu Pinjaman



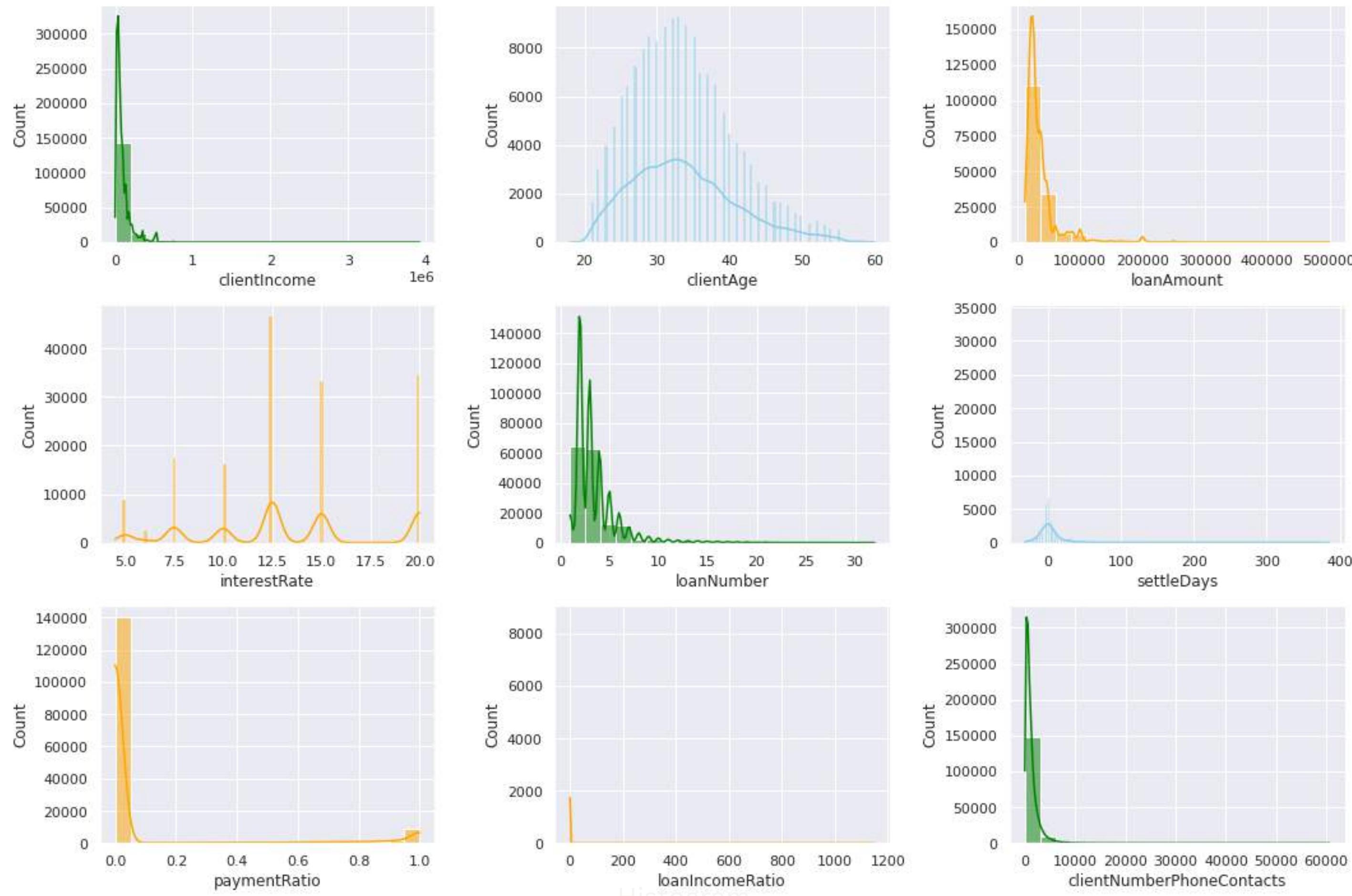
Visualisasi Target Variabel



Rasio Perbandingan Nilai Default (0 : 1)
115321 : 44275

Terdapat Imbalance Data

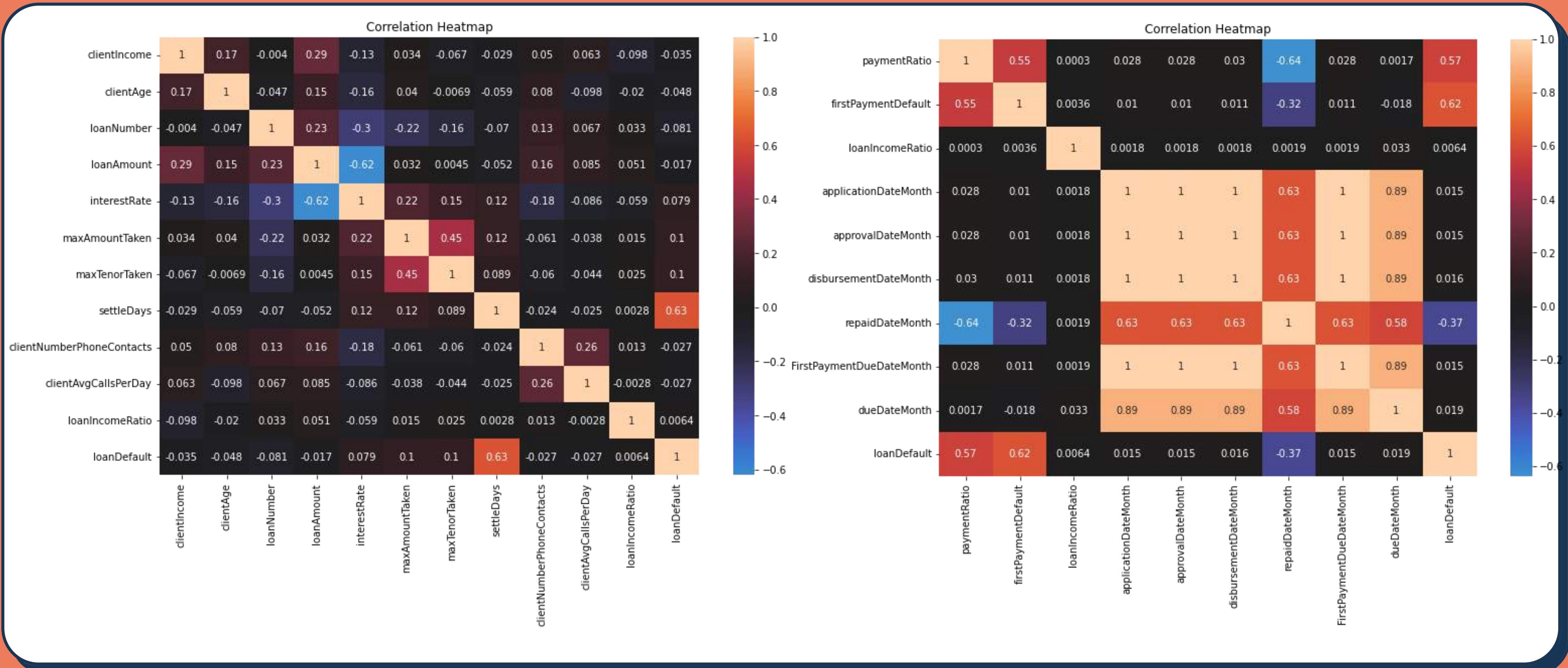
Visualisation



Histogram

Numerical

Visualisasi Heatmap



Heatmap

Insight dari Visualisasi

BCC 2023

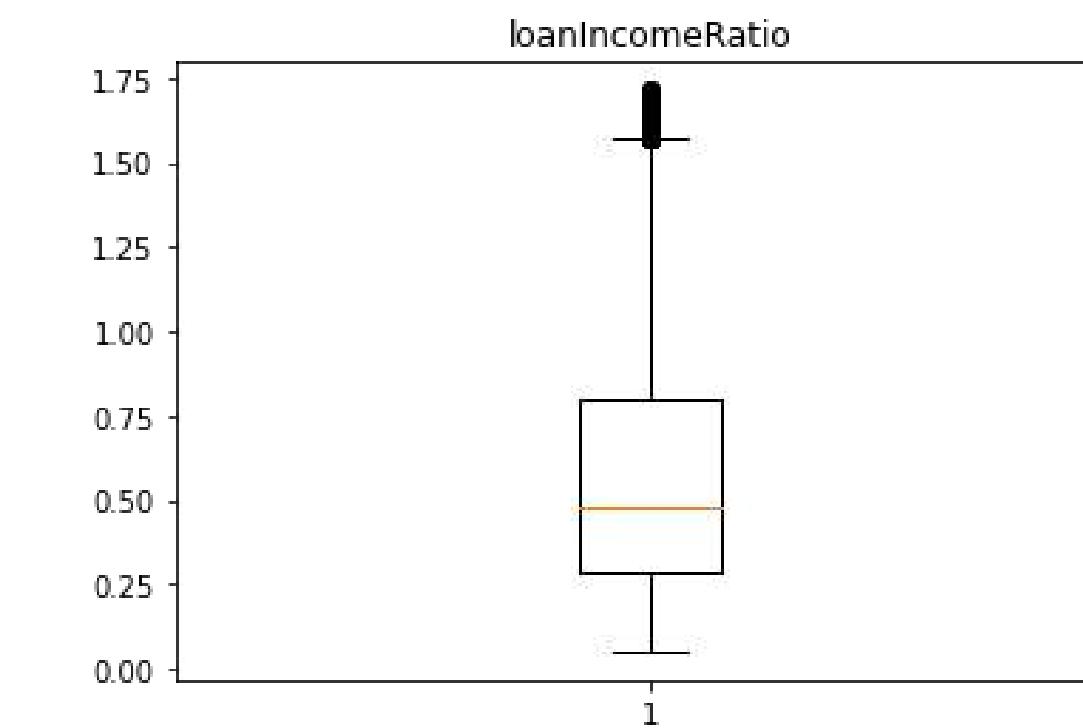
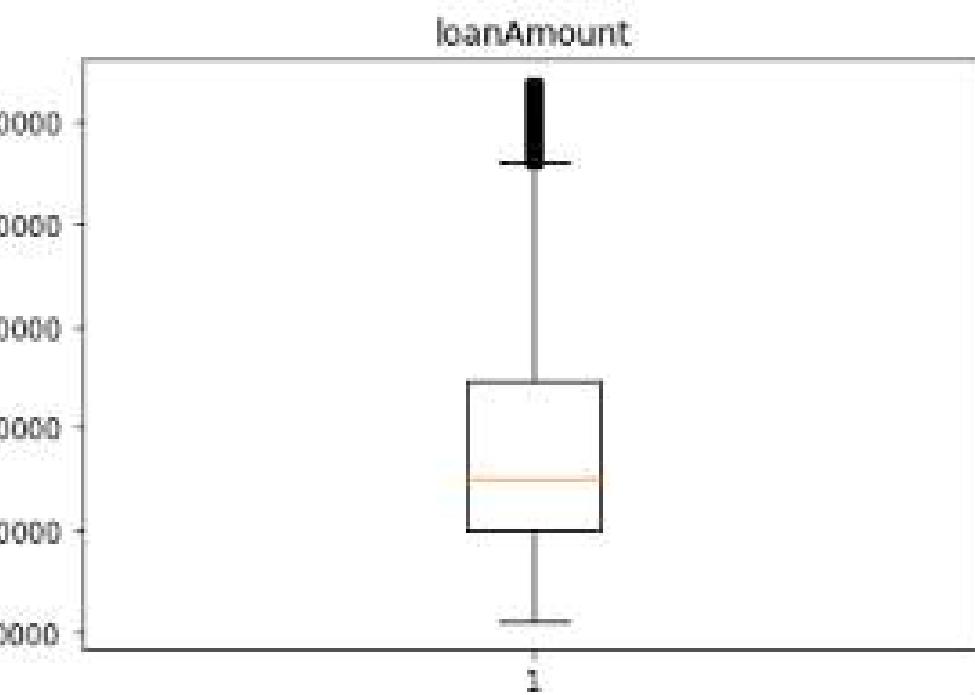
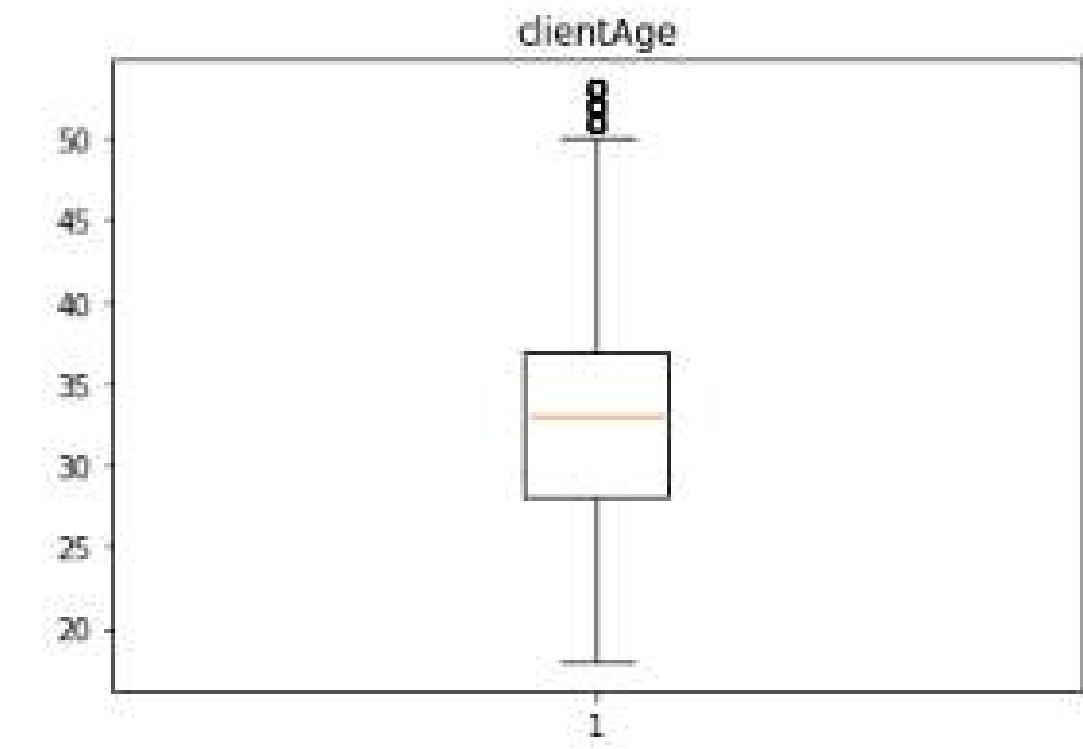
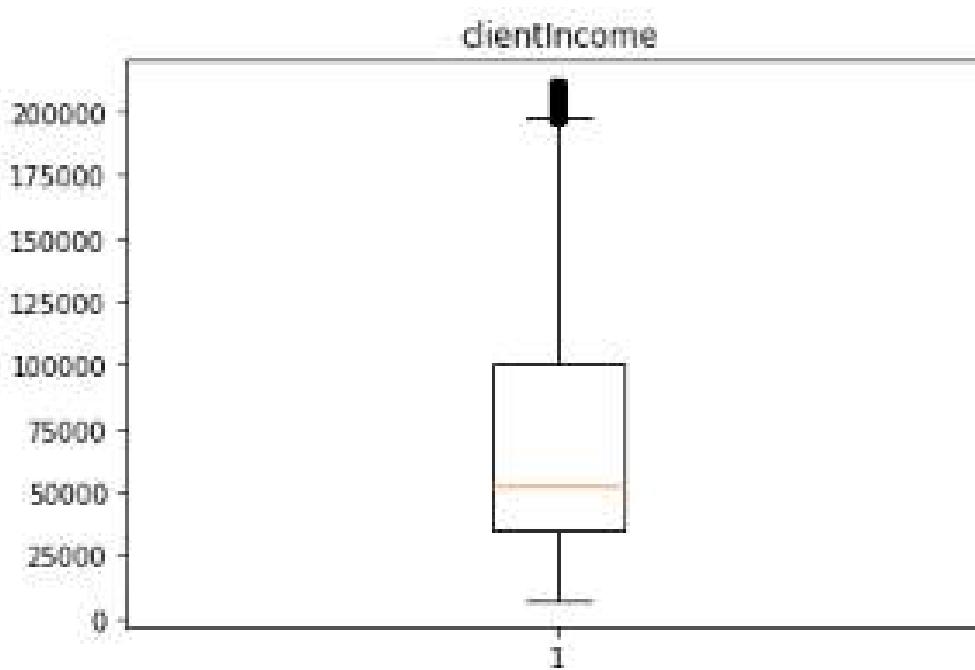
- 1.** Atribut Profil Peminjam Didominasi oleh Laki-Laki, Single & Telah Menikah, Tujuan Pinjaman untuk Bisnis dan Rumah Sewa
- 2.** Atribut terkait Waktu Peminjaman terdistribusi Normal, sehingga tidak ada waktu khusus yang berpengaruh terhadap kegiatan pinjaman
- 3.** Atribut Numerikal tidak terdistribusi Normal
- 4.** Visualisasi Heatmap menghasilkan Atribut SettleDays, paymentRatio dan firstPaymentDefault yang berkorelasi kuat dengan loanDefault

3. Preprocessing Data



Handling Outlier

Penerapan Teknik IQR
dan Visualisasi
Boxplot untuk Cek
Outlier



Feature Selection



Fitur Selection dengan Metode Correlation Based

'interestRate', 'loanTerm', 'paymentRatio',
'firstPaymentDefault', 'loanIncomeRatio', 'approvalDateMonth',
'disbursementDateMonth', 'FirstPaymentDueDateMonth',
'dueDateMonth'.





4. Modelling & Hyperparameter Tuning



Tahap Pipeline

01



Import Library yang dibutuhkan

02



Pemisahan Target Variabel loanDefault

03



Data Splitting menjadi Data Training dan Data Valid 75 : 25

04



Inisiasi Variabel untuk Fitur Terseleksi menjadi Array

05



Membangun Pipeline dan SMOTE untuk handling Imbalance Data

06



Definisi Parameter untuk Hyperparameter Tuning

Tahap Pipeline

07



Definisi RandomizedSearchCV
untuk Algoritma

08



Fit RandomizedSearchCV
untuk Data Training

09



Uji Prediksi dengan Data Valid

10



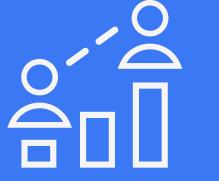
Evaluasi Model Dengan
Metriks F1-Score

11



Cross Validation

12



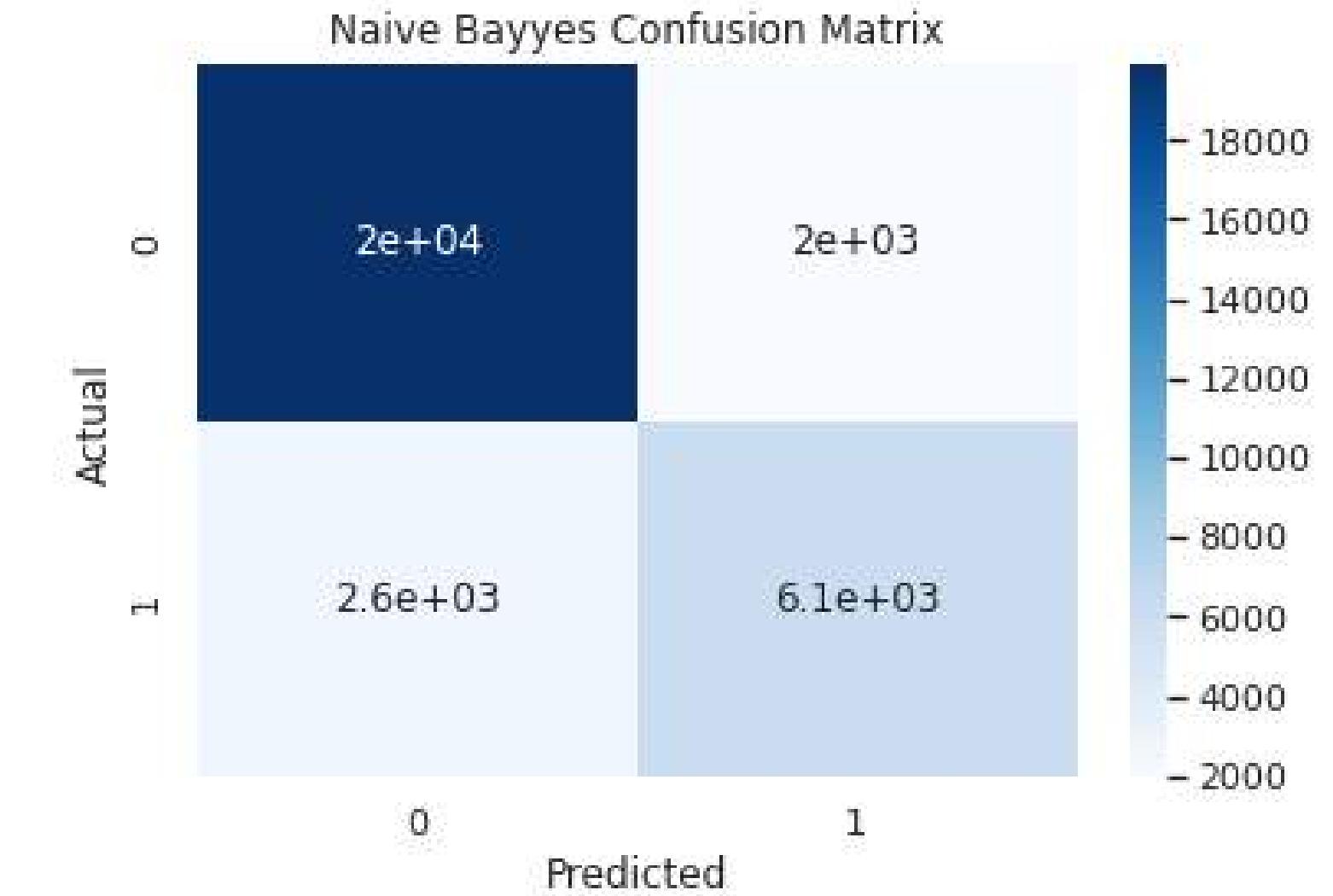
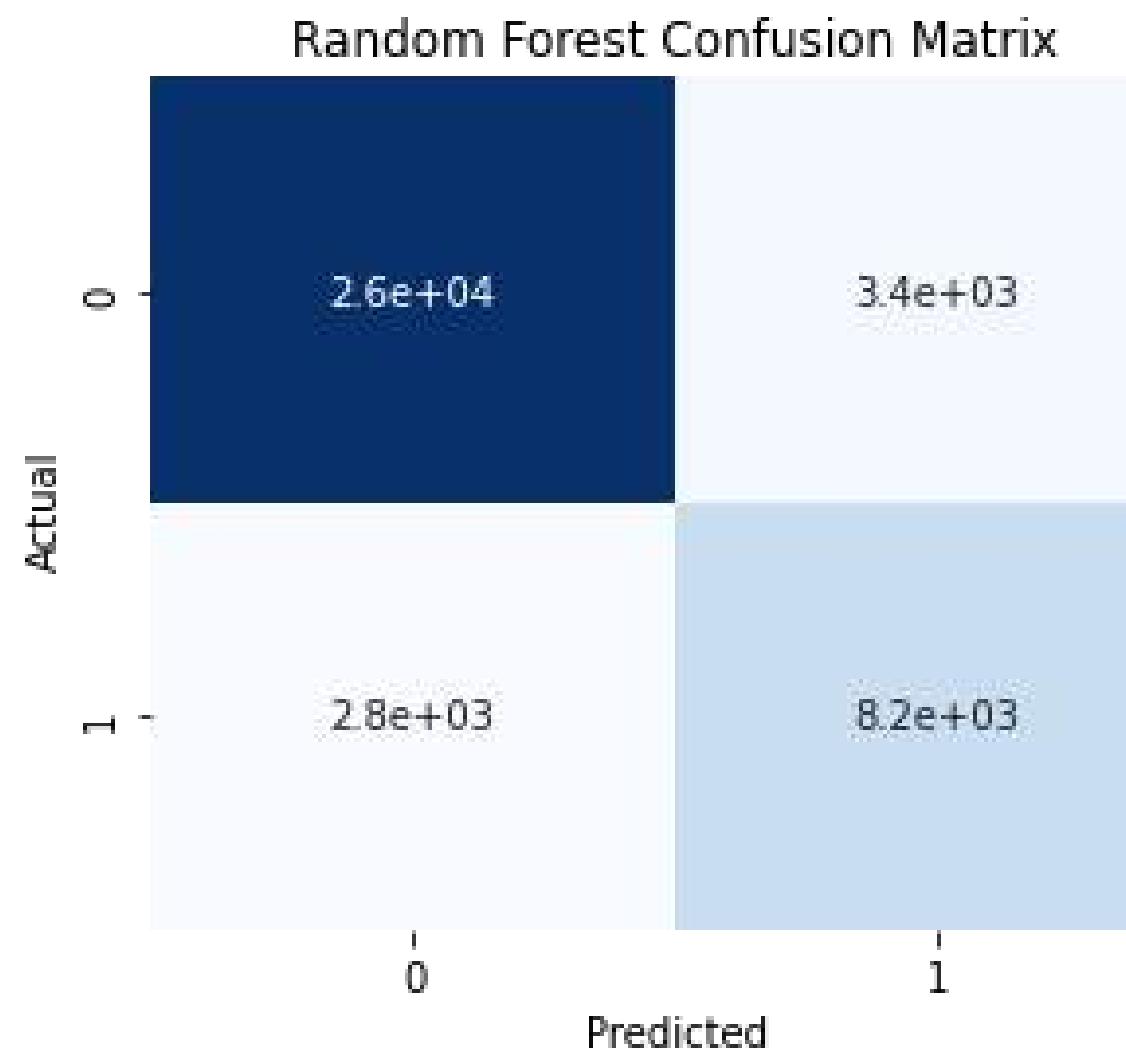
Visualisasi Confusion Matrix

Hasil Modelling & Hyperparameter Tuning

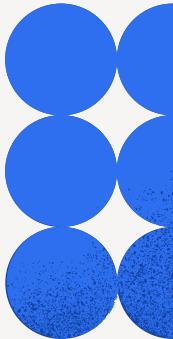
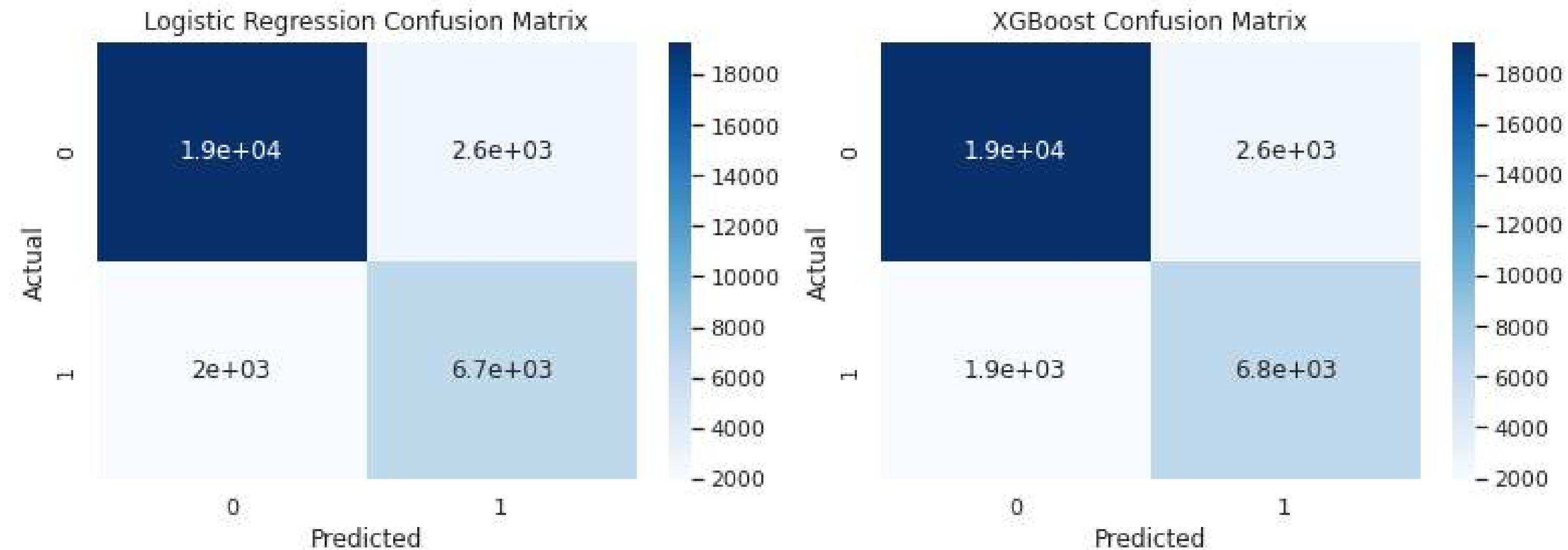
BCC 2023

NO	MODELLING ALGORITMA	SKOR PREDIKSI DEFAULT	SKOR SETELAH HYPERPARAMETER TUNNING
1	Random Forest	67.12%	72.48%
2	XGBoost	74.79%	74.79%
3	Naive Bayes	61.88%	72.92%
4	Logistic Regression	74.82%	74.82%

Visualisasi Confusion Matrix



Visualisasi Confusion Matrix



4. Conclusion

Kesimpulan

1.

Kualitas Data
mempengaruhi skor
prediksi pinjaman

2.

Profil Pelanggan
sebagai Strategi
Pemberian Pinjaman

3.

Modelling Random
Forest menghasilkan
Skor Prediksi yang
cukup tinggi = 72%

REKOMENDASI

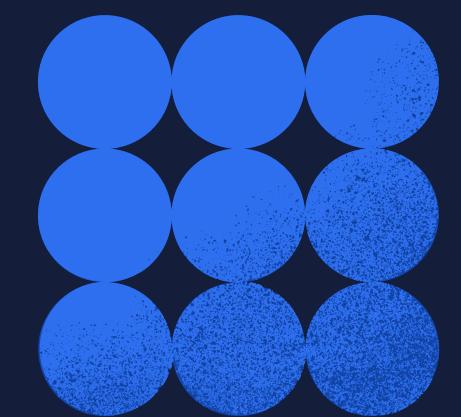
Analisis Data berkala
agar informasi selalu
akurat



Data Science



Thank You....



BCC 2023