

RISK ANALYSIS & LOAN DEFAULT PREDICTION

FOR SMALL BUSINESSES

- Arya Wiratmaji -

DATA UNDERSTANDING

Dataset yang digunakan merupakan informasi tentang pinjaman SBA dari tahun 1987 hingga 2014

Small Business Administration (SBA) adalah badan pemerintah Amerika Serikat yang menyediakan dukungan finansial dan konsultasi kepada usaha kecil untuk membantu mereka tumbuh dan berkembang

BACKGROUND

SBA ingin mengetahui aspek apa saja yang berpengaruh positif ataupun negatif terhadap tingkat default, dan memprediksi kemungkinan default peminjam

PROBLEMS

SBA ingin menurunkan tingkat default pinjaman, namun belum mengetahui aspek apa saja yang penting dan berpengaruh terhadap loan default

OBJECTIVE

- Menganalisa aspek yang berpengaruh terhadap kelancaran pinjaman, serta menggali informasi lebih dalam
- Model prediksi kemungkinan default pinjaman

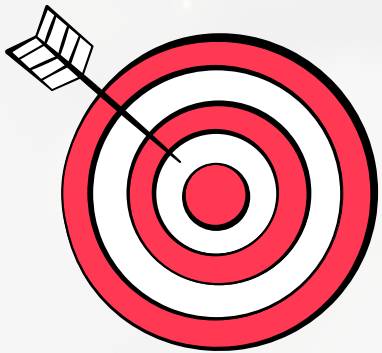
DATASET INFORMATION



899,163 Rows



27 Features



1 Target (MIS_Status)

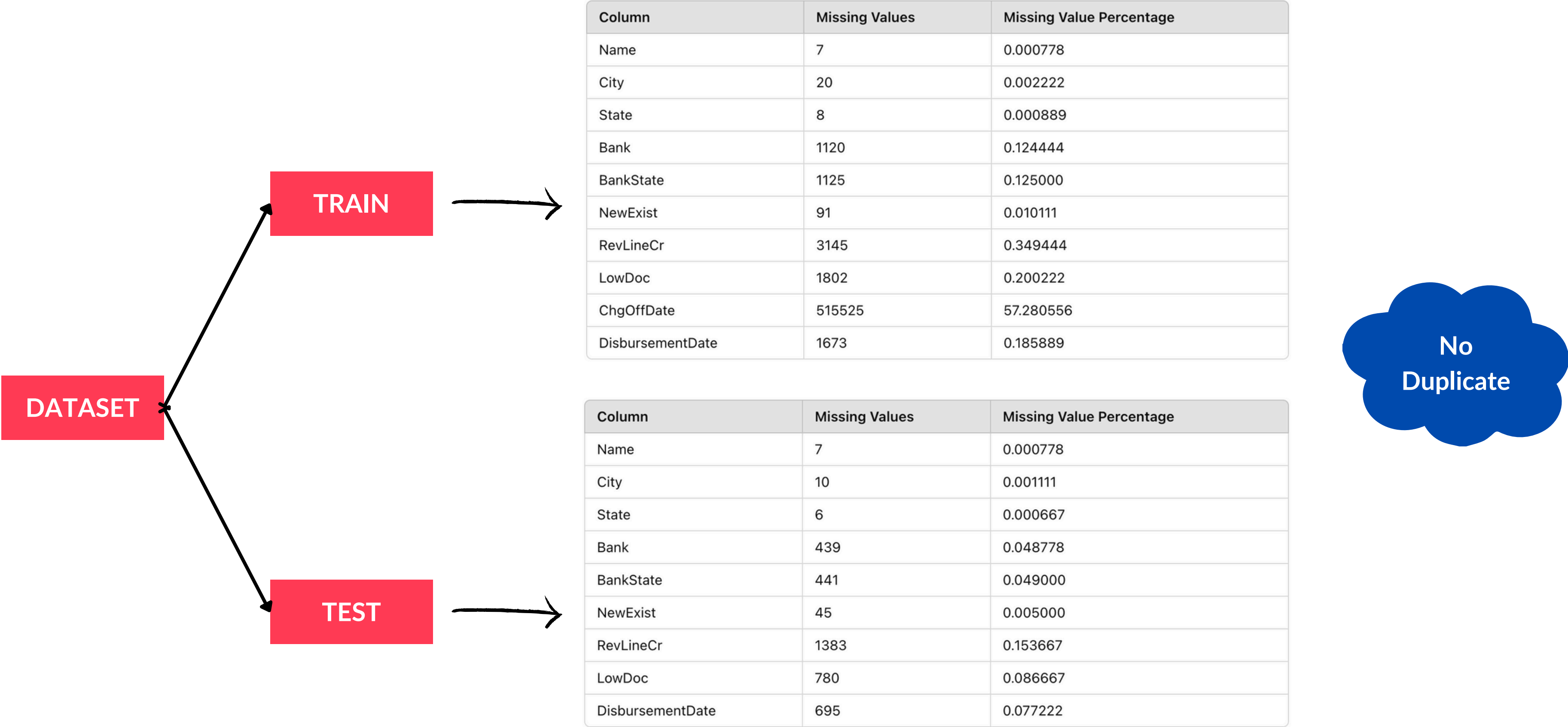
- P I F (Paid in Full)
- CHGOFF (Charge Off)



- LoanNr_ChkDgt
- Name
- City
- State
- Zip
- Bank
- BankState
- NAICS
- ApprovalDate
- ApprovalFY
- Term
- NoEmp
- NewExist
- CreateJob
- RetainedJob
- FranchiseCode
- UrbanRural
- RevLineCr
- LowDoc
- ChgOffDate
- DisbursementDate
- DisbursementGross

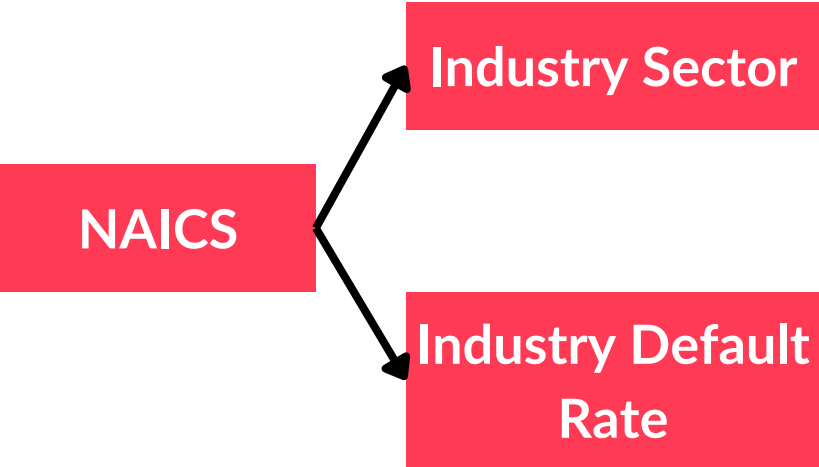
- BalanceGross
- MIS_Status
- ChgOffPrinGr
- GrAppv
- SBA_Appv

MISSING & DUPLICATE



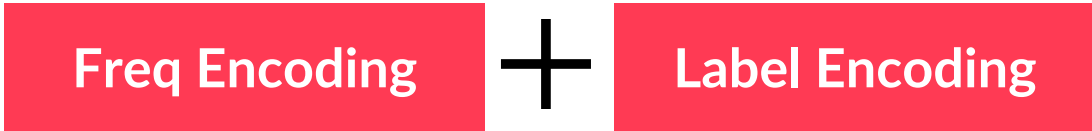
PREPROCESSING

FEATURE ENGINEERING



- +
- Franchise Code -> Is Franchise
 - NewExist -> New Business
 - CreateJob -> Is Created Job

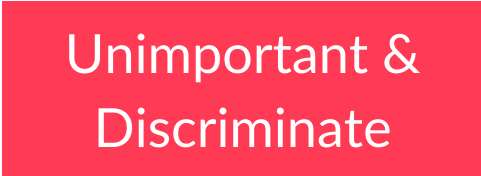
FEATURE ENCODING



SCALING (STANDARD SCALER)



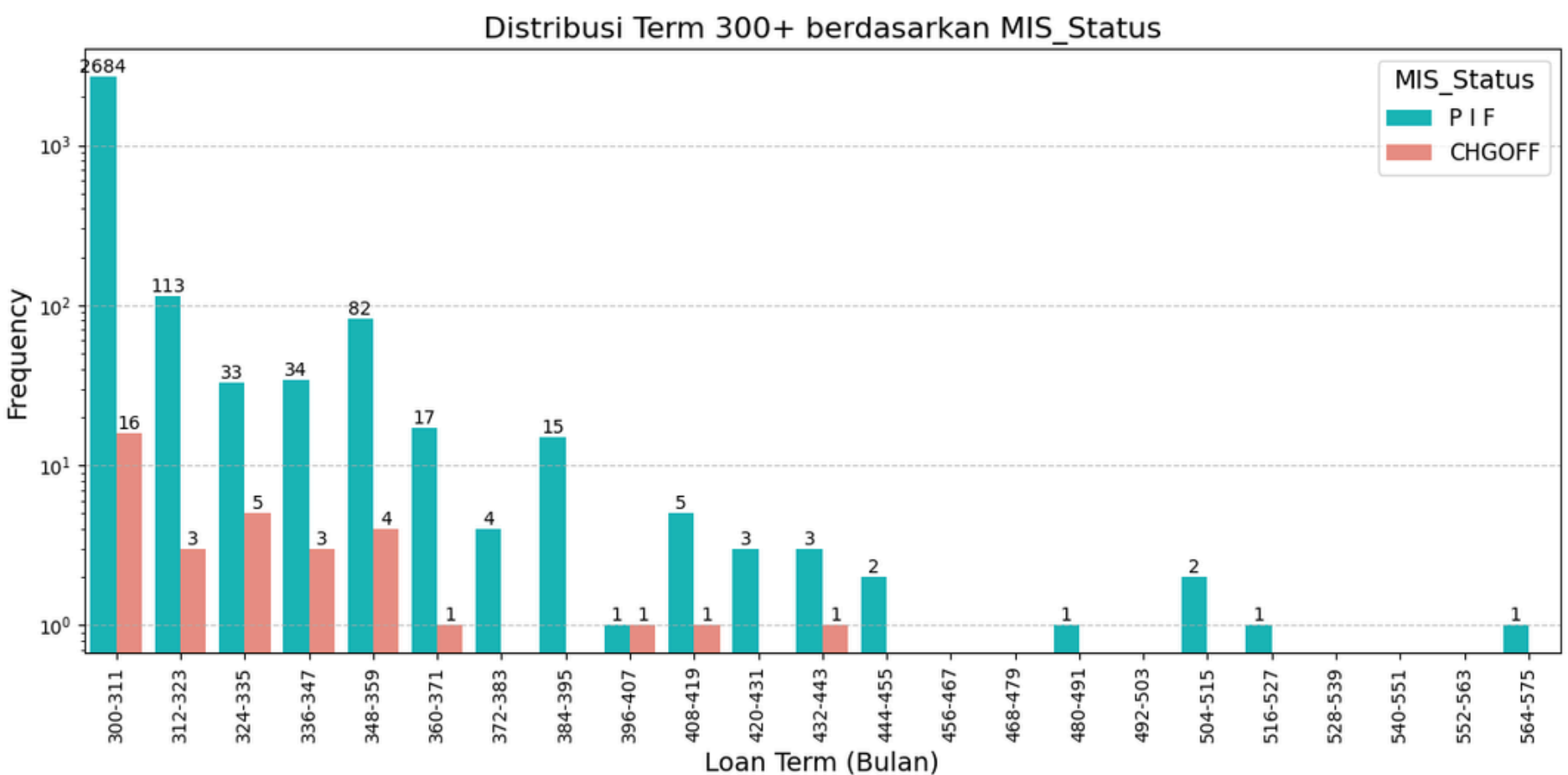
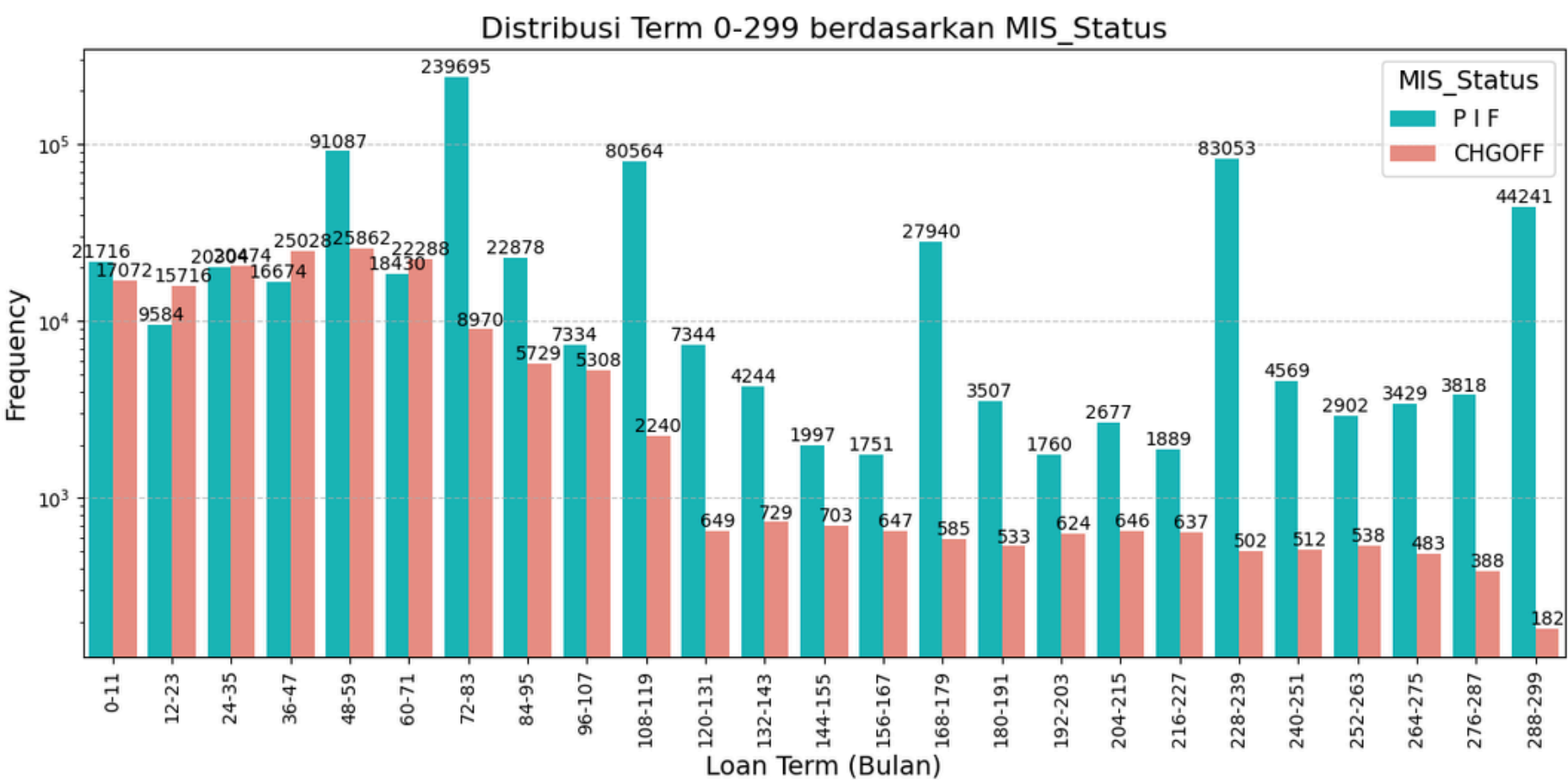
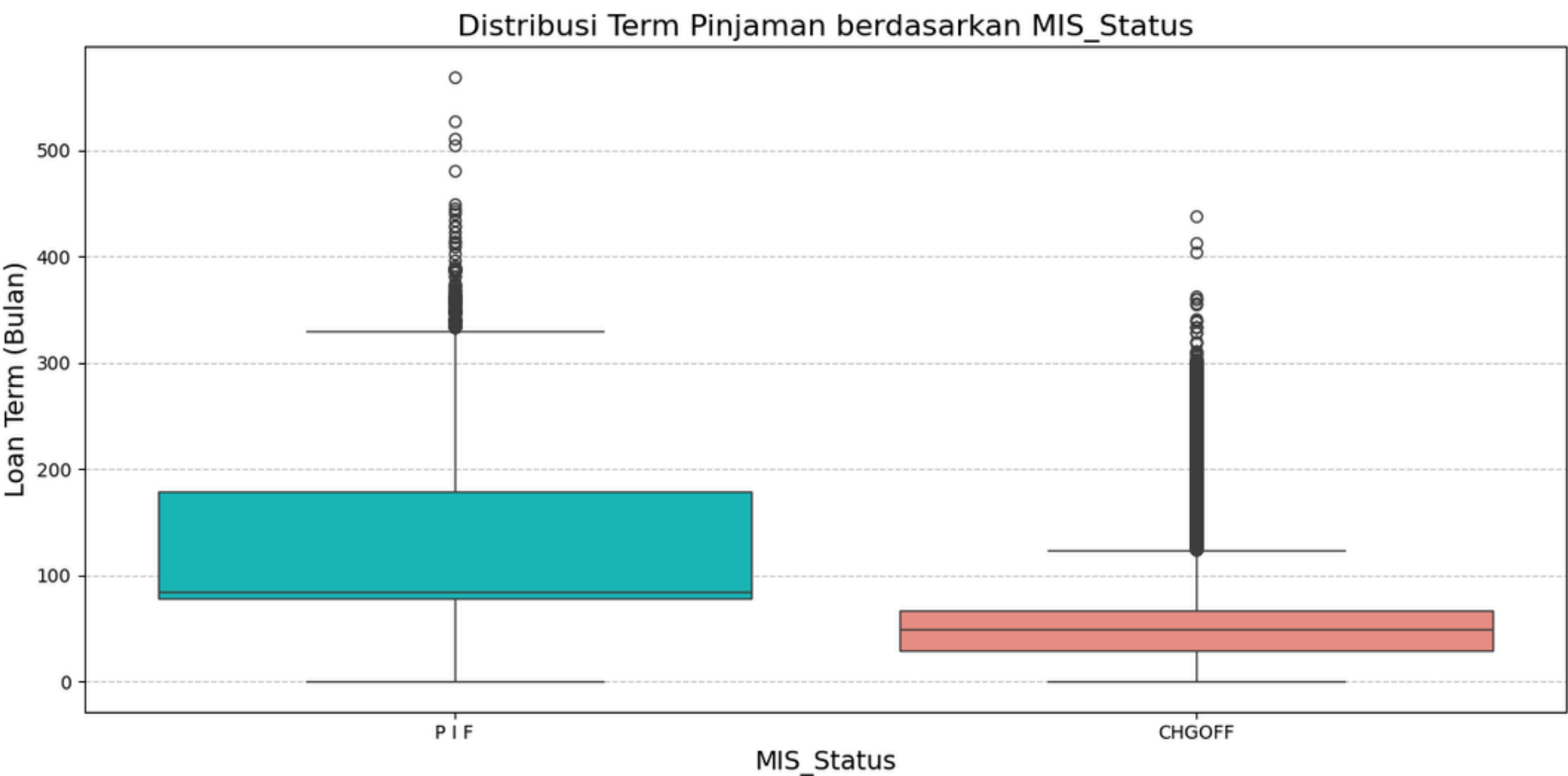
DROP FEATURES



DISTRIBUSI TERM PINJAMAN

INSIGHTS

- Pada seluruh interval *Term*, jumlah pinjaman yang **lunas hampir selalu lebih tinggi dibandingkan default**
- Pinjaman Term > 300 sedikit. Mayoritas -> 300 - 311.
- Pinjaman Term > 300 lebih banyak lunas dibandingkan default



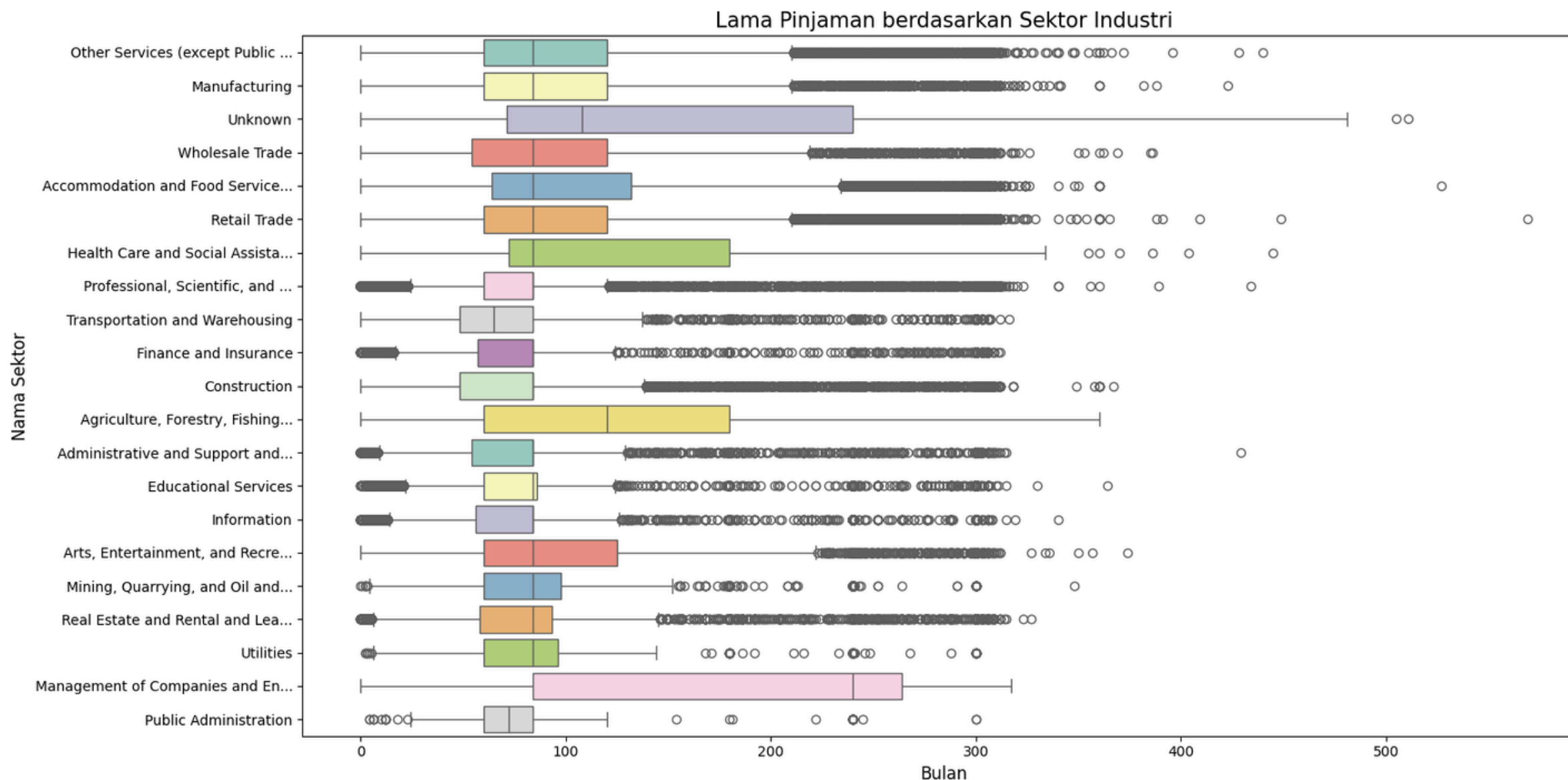
DISTRIBUSI TERM PINJAMAN (CONT.)

INSIGHTS

- Term pinjaman bervariasi secara signifikan antara berbagai sektor industri
- Beberapa sektor cenderung ambil pinjaman jangka pendek, beberapa cenderung panjang

BUSINESS RECOMMENDATIONS

- Pertimbangkan untuk menawarkan opsi Term yang lebih panjang kepada peminjam (Bisa membantu arus kas & waktu yang cukup untuk melunasi pinjaman)
- Pertimbangkan penyesuaian kebijakan suku bunga untuk Term jangka panjang



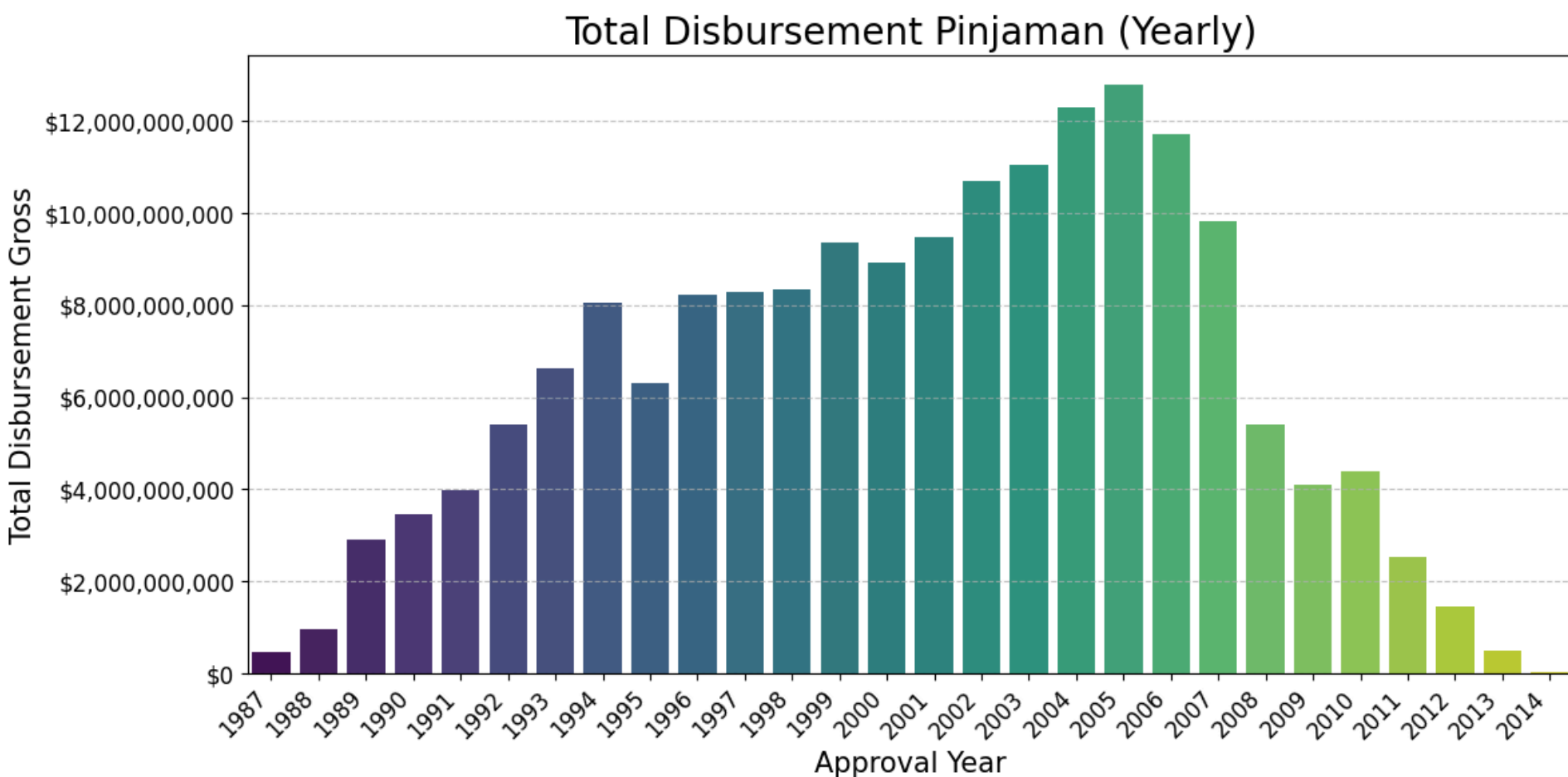
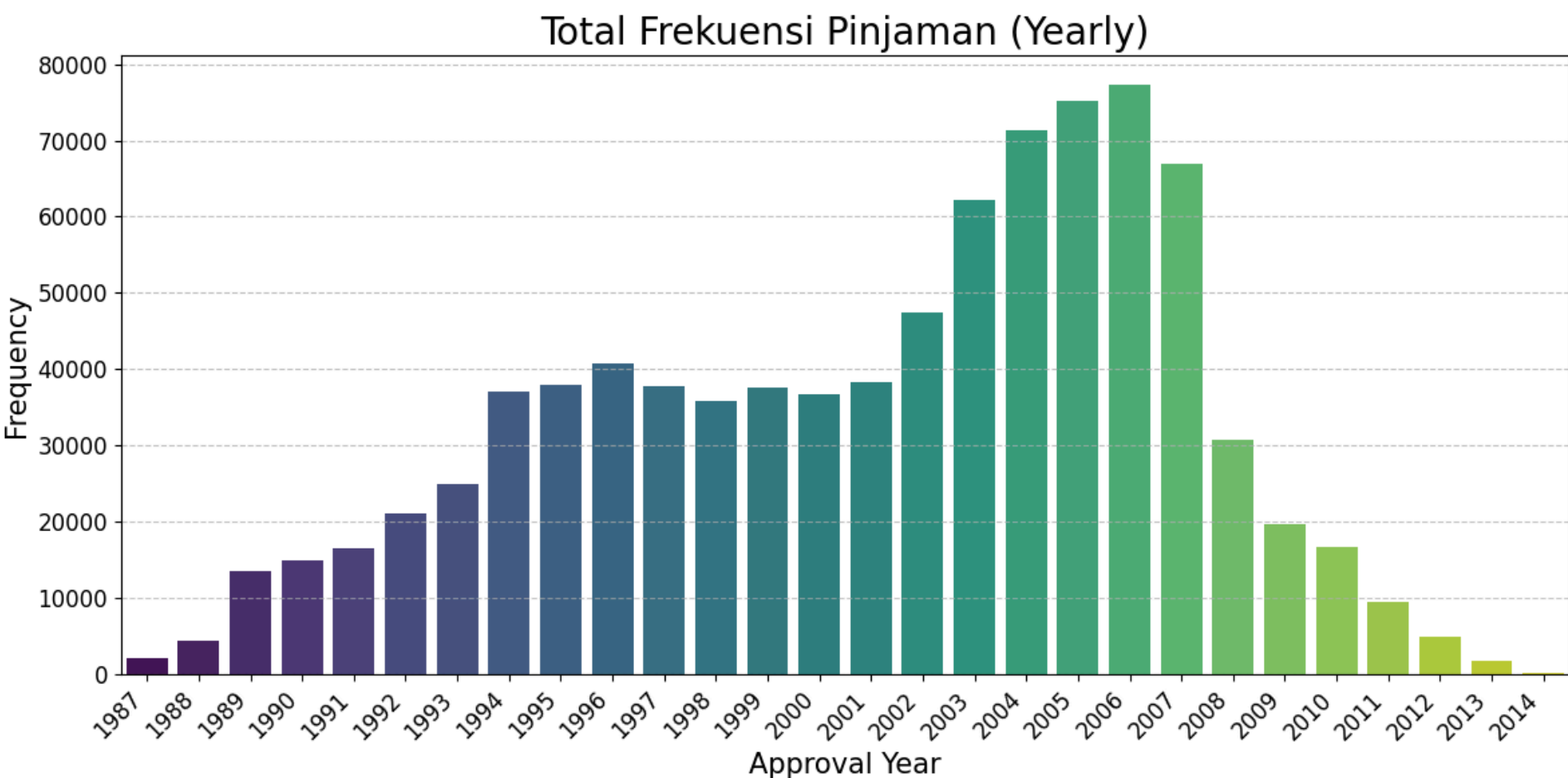
DISTRIBUSI FREQ & DISBURSEMENT (YEARLY)

INSIGHTS

- 1987 – 2006 Freq & Disb mengalami peningkatan
- 2007 – 2014 (last data) Freq & Disb mengalami penurunan
- Penurunan tajam karena The Great Recession. Bank & lembaga keuangan (incl. SBA) memperketat kredit & hati2 memberi

BUSINESS RECOMMENDATIONS

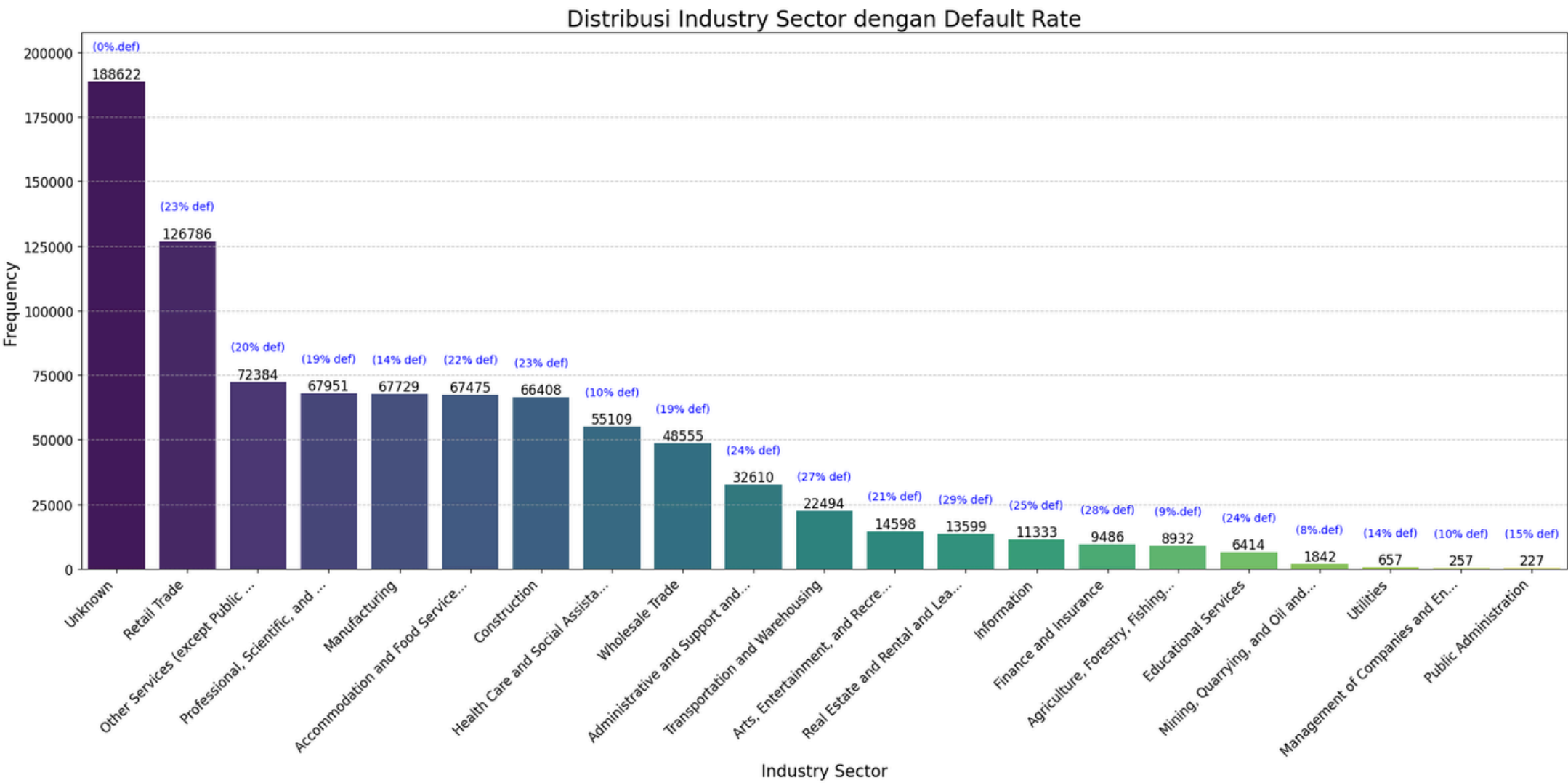
- Harus waspada perubahan ekonomi makro
- Mengembangkan kebijakan kredit fleksibel & adaptif terhadap perubahan ekonomi makro
- Model prediksi yang dibuat berdasarkan data lampau. Tidak bijak digunakan ketika ekonomi bergejolak.



DISTRIBUSI LOAN BERDASARKAN INDUSTRY

INSIGHTS

- Unknown tertinggi. Kategori 'NAICS' dataset ini tidak lengkap
- Retail Trade tertinggi (setelah Unknown) dengan 23% industry default rate.
- Mostly freq pinjaman tertinggi default ratenya moderate
- Mostly default rate rendah ada di freq pinjaman rendah juga



BUSINESS RECOMMENDATIONS

- Unknown (NAICS = 0) perlu ditinjau lebih lanjut. Data yang lebih akurat akan membantu dalam analisis risiko yang lebih tepat.
- Freq pinjaman tinggi + industry default rate tinggi berisiko. Untuk sekarang ini top 3 freq teratas default ratenya moderate (23%, 20%, 19%). Jika nextnya terjadi perubahan, SBA harus waspada.
- Berikan bantuan tambahan sector default rate tinggi

HIGHEST

Industry Sector	Default Rate
Real Estate and Rental and Leasing	29
Finance and Insurance	28
Transportation and Warehousing	27

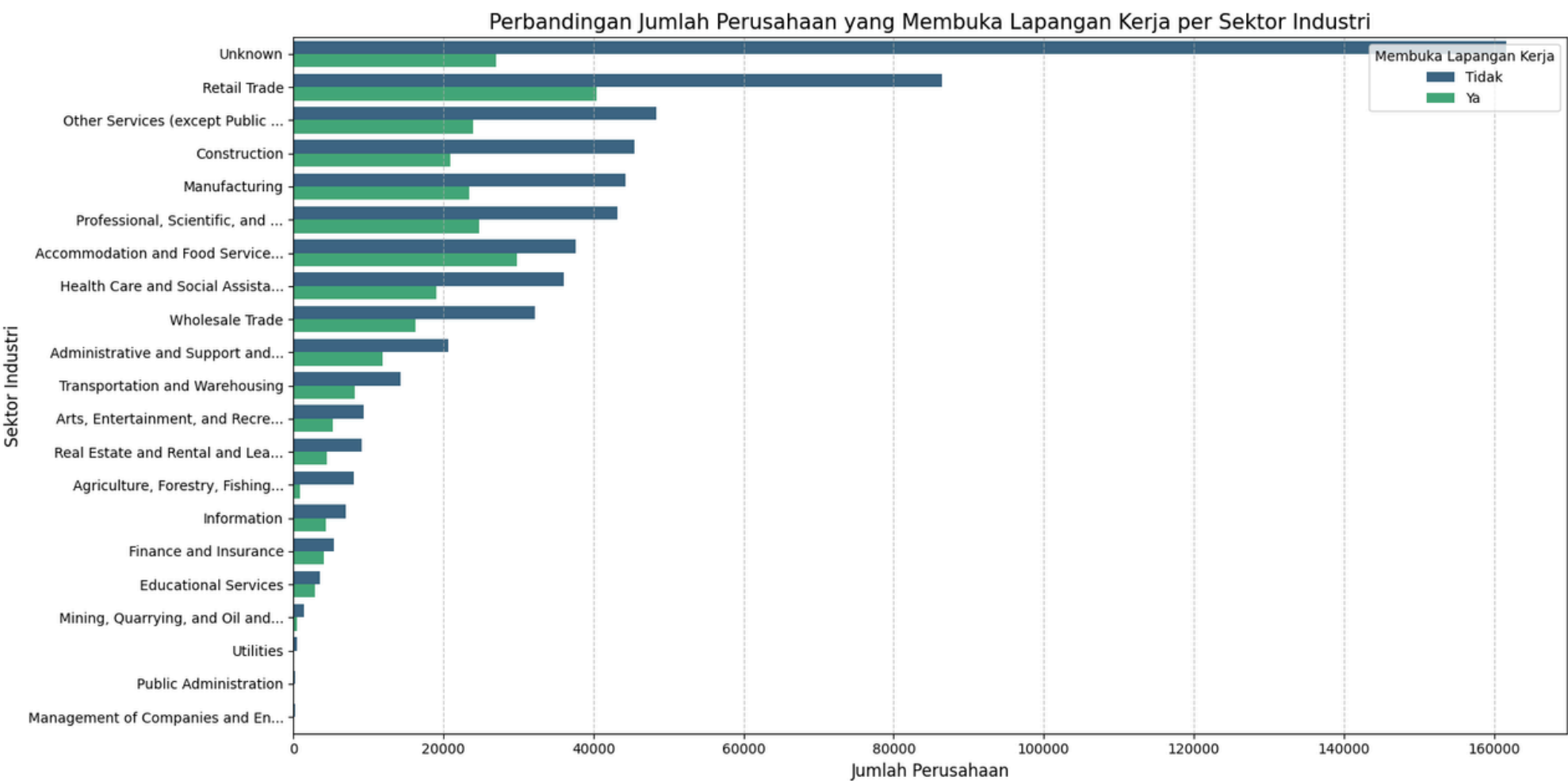
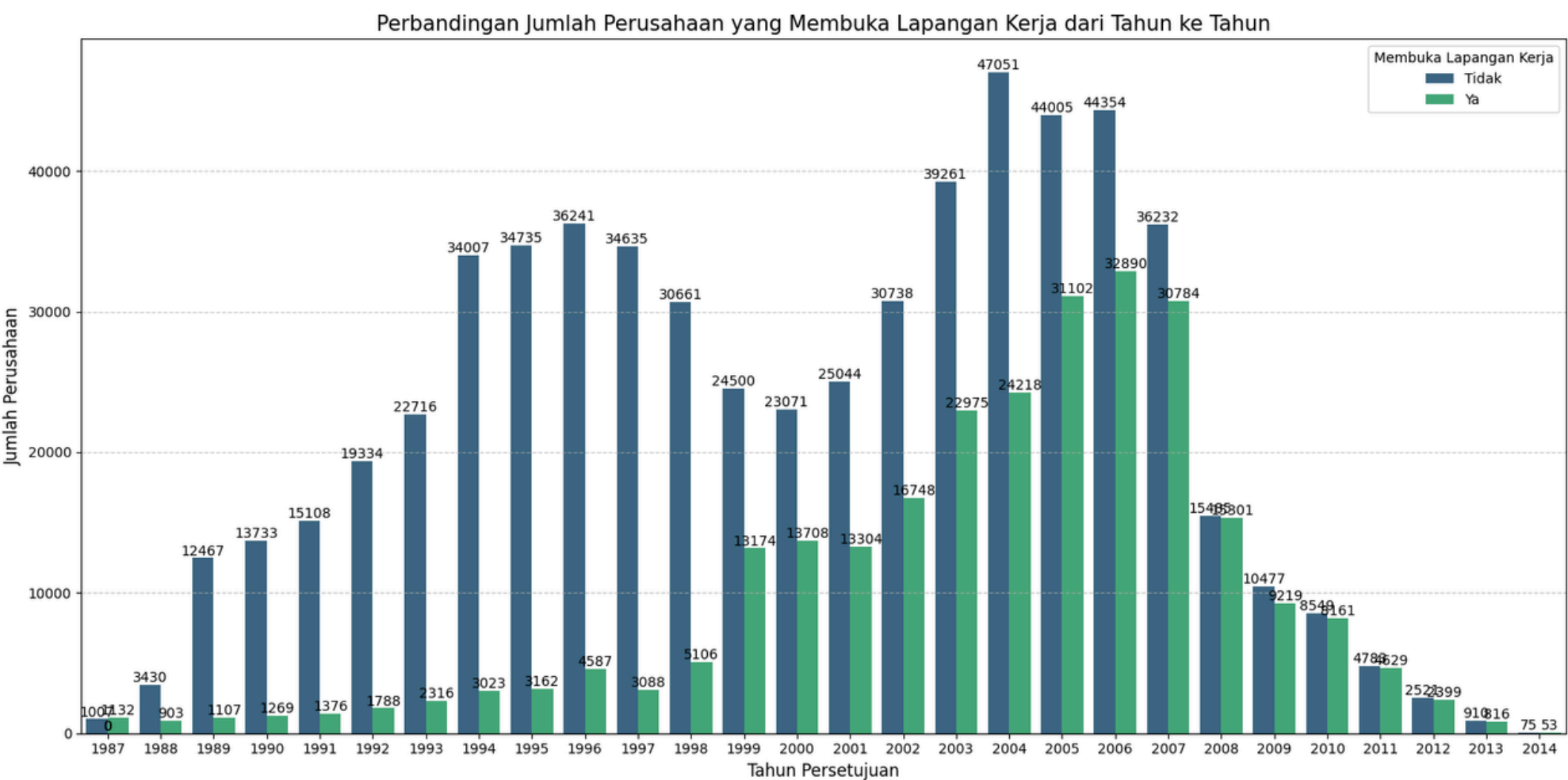
LOWEST

Industry Sector	Default Rate
Mining, Quarrying, and Oil and Gas Extraction	8
Agriculture, Forestry, Fishing and Hunting	9
Health Care and Social Assistance	10

PEMBUKAAN LAPANGAN KERJA

INSIGHTS

- Jumlah perusahaan buka lapangan kerja meningkat 1988 - 2006
- 2007 - 2014 (last data) jumlah perusahaan yang disetujui pinjamannya menurun, tapi ratio buka lapangan kerja vs tidak meningkat
- "Retail Trade" & "Accommodation and Food Services" paling banyak buka lapangan kerja.
- "Public Administration" dan "Management of Companies" paling sedikit buka lapangan kerja



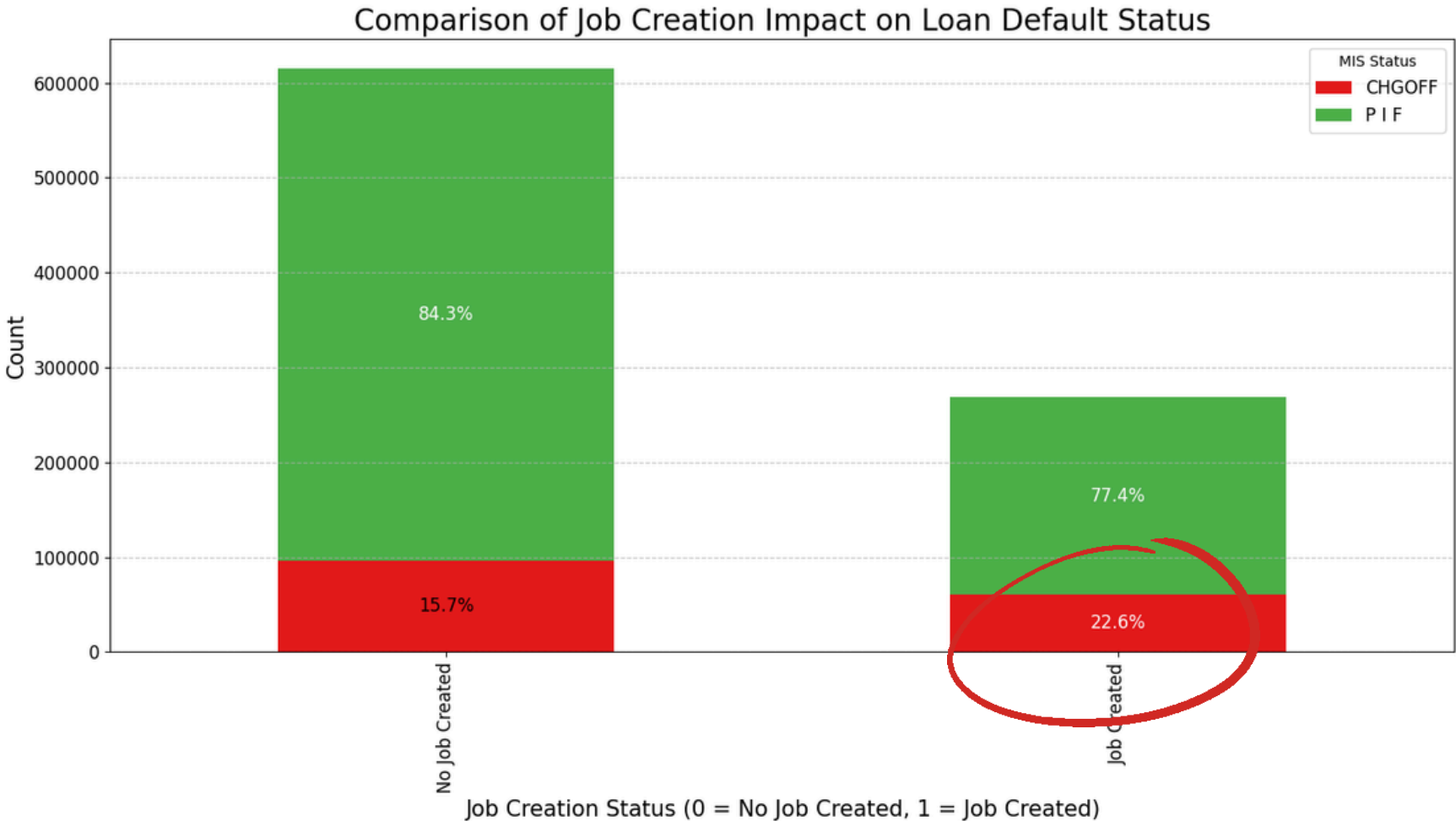
PEMBUKAAN LAPANGAN KERJA (CONT.)

INSIGHTS

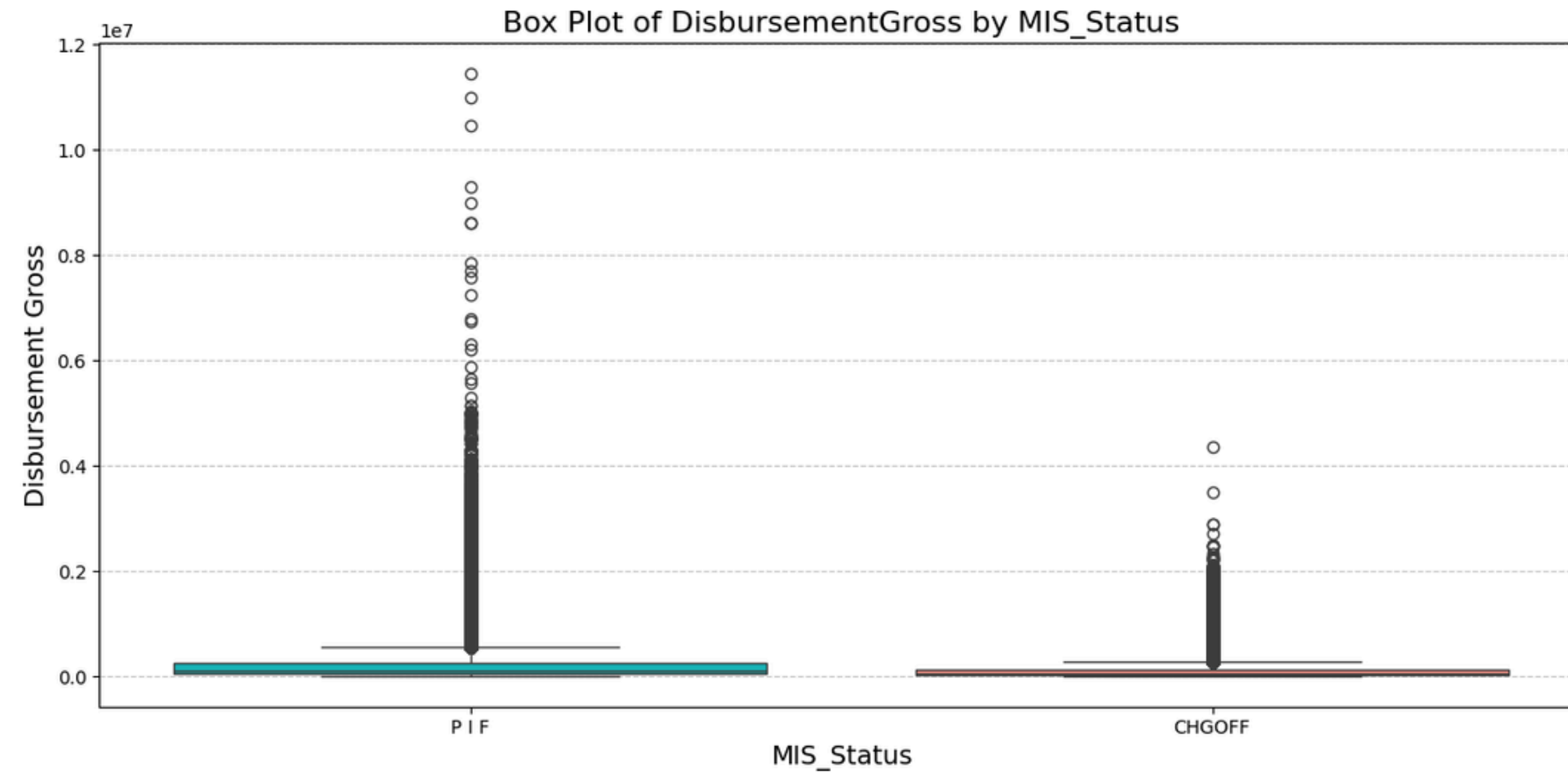
- Ternyata persentase rasio default perusahaan yang membuka lapangan kerja > perusahaan yang tidak membuka
- Bisa jadi dikarenakan faktor seperti biaya operasional lebih tinggi (karena buka kerja), manajemen kurang efektif, dll

BUSINESS RECOMMENDATIONS

- Tetap memberikan pinjaman pada sektor track record baik dalam pembukaan kerja untuk kesejahteraan masyarakat, namun tetap waspada
- Perketat evaluasi kelayakan bisnis sebelum peminjaman, terutama bisnis yang rencana buka lapangan kerja.
- Untuk perusahaan yang rencana buka lapangan kerja, SBA bisa beri bantuan pelatihan (seperti manajemen & keuangan)



HUBUNGAN JUMLAH (\$) PINJAMAN TERHADAP DEFAULT



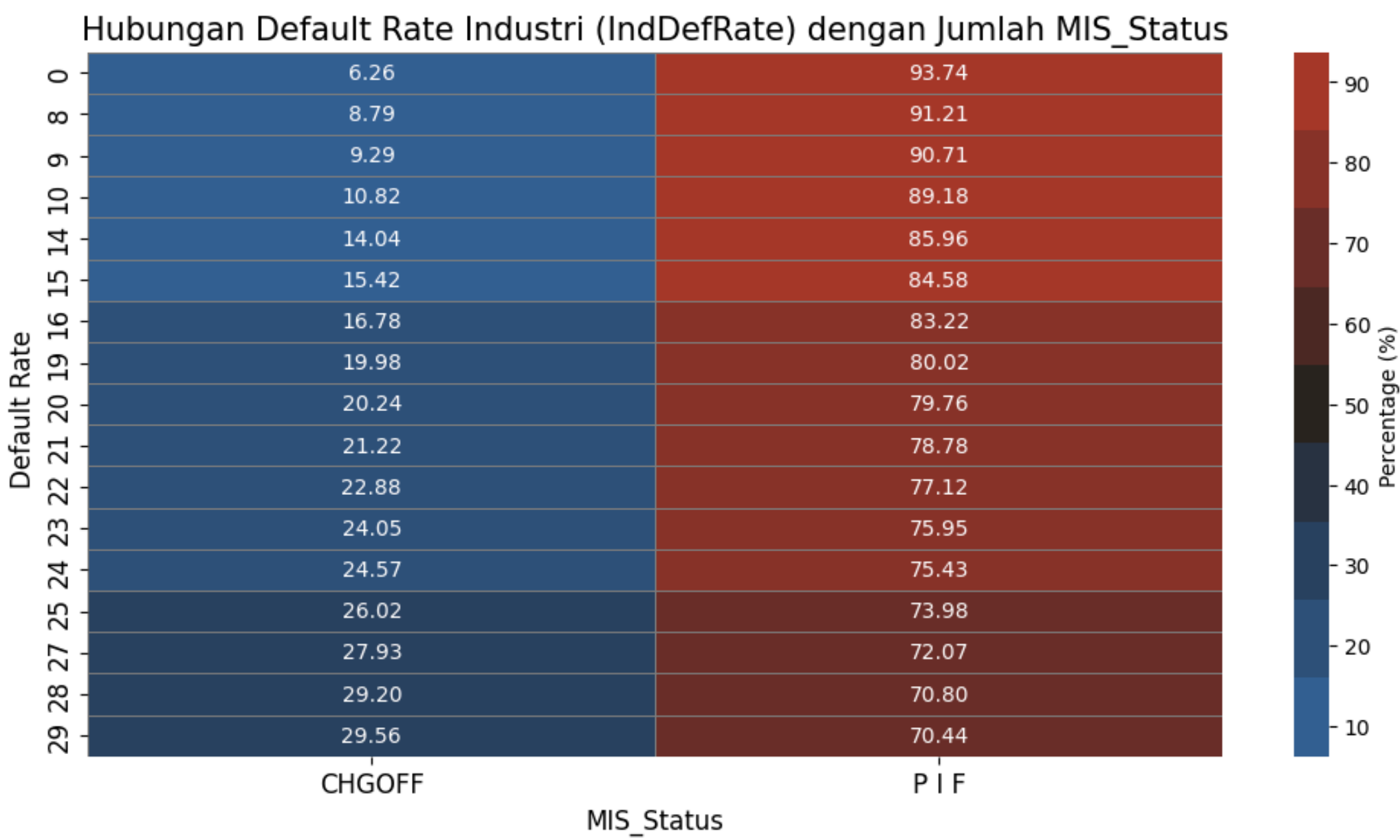
INSIGHT

- Pinjaman yang lunas memiliki distribusi disbursement yang lebih luas dibanding default
- Pinjaman yang lunas beberapa disbursementnya lebih tinggi dibanding default
- SBA sudah cukup baik dalam evaluasi penyetujuan pinjaman dengan jumlah besar

BUSINESS RECOMMENDATIONS

- Terus meningkatkan evaluasi penyetujuan pinjaman jumlah besar, serta pertimbangkan aspek eksternal & ekonomi makro
- Berikan program pendampingan & monitoring, terutama untuk peminjam besar

HUBUNGAN INDUSTRY DEFAULT RATE TERHADAP JUMLAH DEFAULT



INSIGHT

- Hubungan yang selaras antara data Industry Default Rate & Default Status pada dataset ini (SBA).
- Default rate rendah = Rasio default ratenya rendah
- Default rate tinggi = Rasio default rate tinggi

BUSINESS RECOMMENDATIONS

- SBA bisa terapkan requirement yang lebih ketat untuk industri dengan default rate yang tinggi (cth: persyaratan jaminan tambahan, dll)
- Analisis lebih lanjut faktor yang memengaruhi industri default rate tinggi (cth: kond pasar, regulasi).

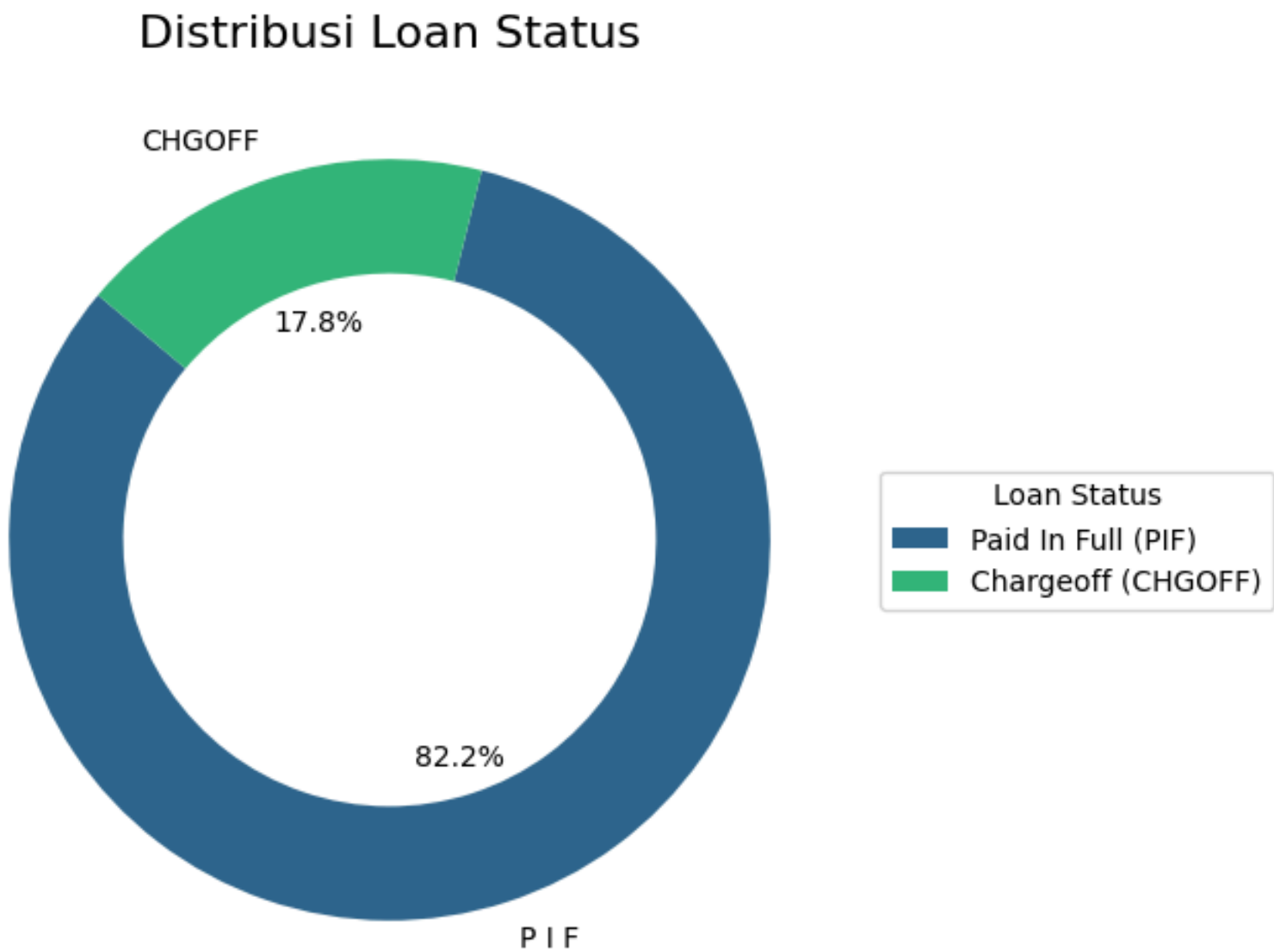
Industry Default Rate:
Persentase pinjaman yang default dalam industri tertentu

Sumber default rate: US Census Bureau



MODELING

VARIABEL TARGET (MIS_STATUS)



Karena target memiliki ketidakseimbangan kelas antara CHGOFF dan P I F, maka metric evaluasi utama yang digunakan di sini adalah F1-Score, dibanding accuracy.

MODELING (TRAIN DATA)

Models	Accuracy	AUC	Recall	Precision	F1
LightGBM	0.94	0.97	0.79	0.87	0.83
XGBoost	0.94	0.97	0.77	0.87	0.81
Random Forest	1.00	1.00	0.99	1.00	0.99
Logistic Regression	0.84	0.84	0.24	0.63	0.35
KNN	0.92	0.97	0.69	0.83	0.76

Accuracy: Persentase prediksi yang benar (keseluruhan)

AUC (Area Under Curve): Model dapat membedakan kelas positif & negatif dengan baik (0-1)

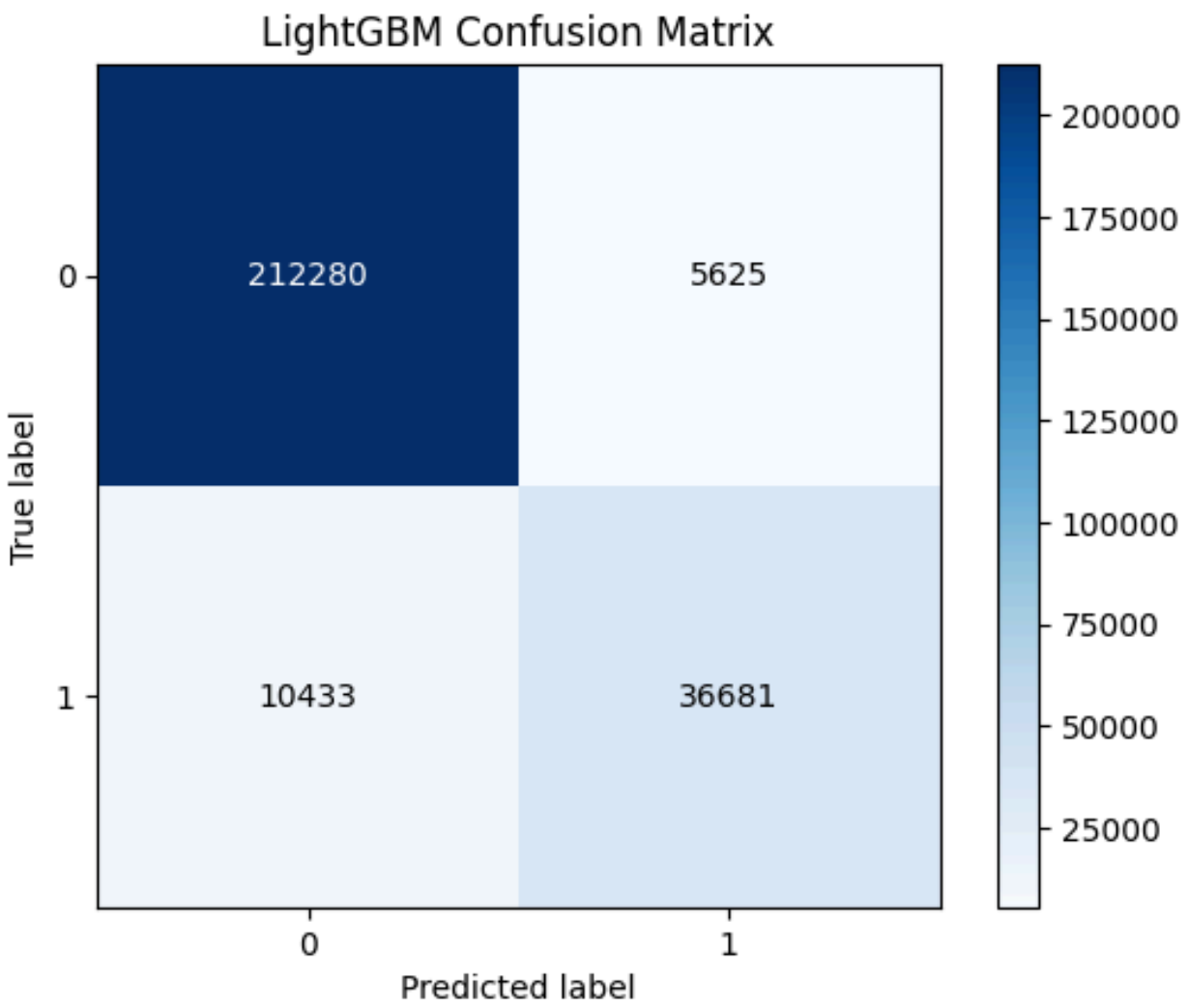
Recall: Proporsi prediksi kelas positif yang berhasil diidentifikasi dengan benar

Precision: Proporsi kelas positif yang benar dari semua prediksi positif

F1-Score: Harmonisasi dari Precision & Recall

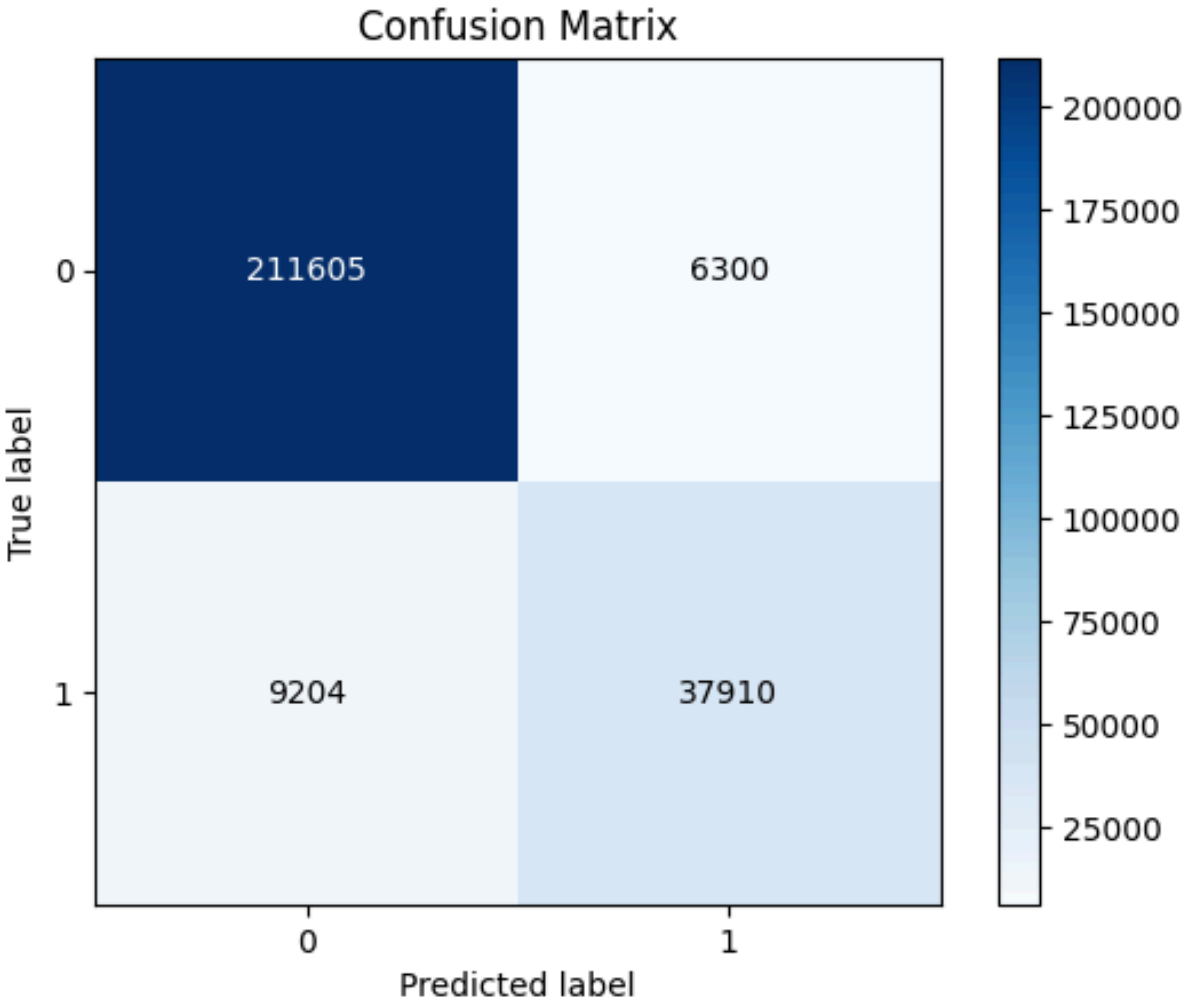
MODELING (TEST DATA)

Models	Accuracy	AUC	Recall	Precision	F1
LightGBM	0.94	0.97	0.78	0.87	0.82
XGBoost	0.92	0.96	0.74	0.81	0.77
Random Forest	0.92	0.96	0.73	0.83	0.77
Logistic Regression	0.84	0.84	0.24	0.62	0.35
KNN	0.89	0.88	0.58	0.74	0.65

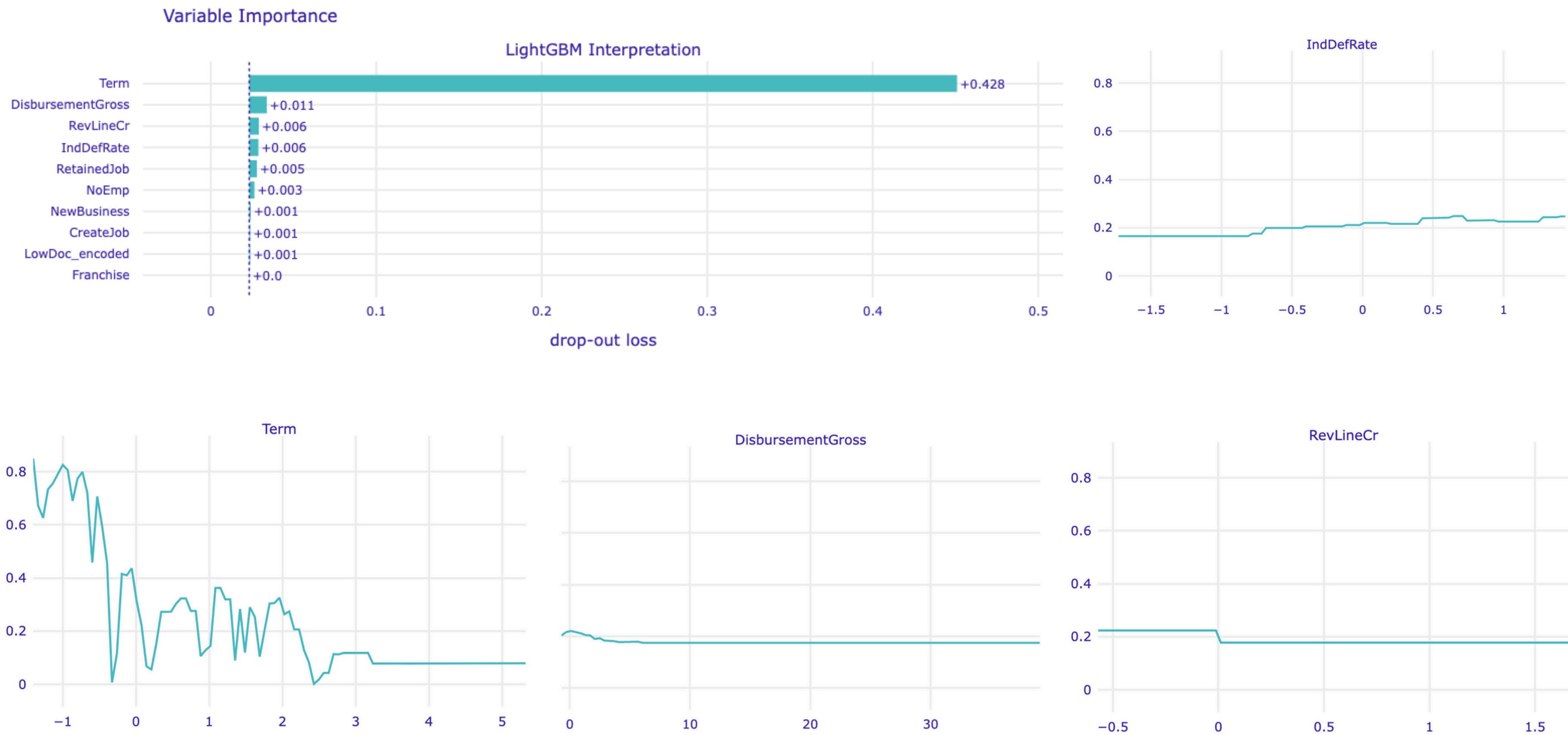


HYPERPARAMETER TUNING (OPTUNA)

Models	Accuracy	AUC	Recall	Precision	F1
Default LightGBM	0.94	0.97	0.78	0.87	0.82
Tuned LightGBM	0.94	0.97	0.81	0.87	0.84
Tuned LightGBM (Test Data)	0.94	0.97	0.80	0.86	0.83



FEATURE IMPORTANCE



KESIMPULAN

- Model prediksi memiliki F1 Score 0.83, menunjukkan bahwa model ini memiliki keseimbangan yang baik antara precision dan recall. Artinya model ini mampu mengidentifikasi akurat proporsi pinjaman yang default sambil meminimalkan kesalahan prediksi.
- Model prediksi memiliki Recall 0.8 yang mana cukup baik. Karena pada kasus ini, False Negative sangat merugikan SBA sebagai pemberi pinjaman
- Dengan menggunakan model prediksi ini, SBA dapat terbantu dalam hal menseleksi pinjaman yang diterima dan yang ditolak
- Menggunakan insight & recommendation EDA, SBA dapat menerapkan kebijakan & bantuan yang ditargetkan untuk industri dan jangka waktu yang berisiko tinggi

RANGKUMAN BUSINESS RECOMMENDATIONS

TERM

- Penawaran opsi term panjang untuk bantu arus kas & memberi waktu cukup
- Penyesuaian suku bunga untuk term panjang

FREQ & DISBURSEMENT

- Kebijakan kredit yang fleksibel dan adaptif terhadap perubahan ekonomi makro.
- Waspada penggunaan model prediksi saat situasi ekonomi buruk

DISTRIBUSI PER INDUSTRI

- Tinjau lebih lanjut 'Unknown' pada NAICS untuk memperoleh data yang lebih akurat
- Beri bantuan tambahan untuk industri default rate tinggi

PEMBUKA LAPANGAN KERJA

- Tetap beri pinjaman pada sektor pembuka lapangan kerja, namun tetap waspada
- Evaluasi kelayakan bisnis, terutama untuk bisnis pembuka lapangan kerja
- Beri bantuan pelatihan manajemen & keuangan untuk bisnis pembuka lapangan kerja

JUMLAH PINJAMAN

- Meningkatkan evaluasi penyetujuan pinjaman besar & mempertimbangkan aspek eksternal & ekonomi makro
- Beri program pendampingan & monitoring pada peminjam besar

INDUSTRY DEFAULT RATE

- Persyaratan lebih ketat untuk industry default rate tinggi (jaminan tambahan, dll)
- Analisis lebih lanjut faktor pengaruh default rate pada bisnis high risk (kondisi pasar, regulasi, dll)



TERIMA KASIH



MISC (MULTICOLLINEARITY HEATMAP)

