

A black and white photograph showing the lower portion of several modern skyscrapers. The building on the left has a distinctive diagonal, stepped facade with many windows. The building next to it is more rectangular with a grid of windows. The sky is bright and overexposed.

IDX PARTNERS

MARCH 2025



DATA OVERVIEW

Dataset Summary

- Total Data: 19,158 kandidat
- Periode Data: 2018 - 2023
- Label 0 (tidak berisiko): 86.28%
- Label 1 (berisiko tinggi): 13.72%

Business Context

- Tujuan: memprediksi risiko gagal bayar pinjaman
- Kategori risiko ditentukan berdasarkan variabel finansial dan demografis
- Model ini untuk mengoptimalkan keputusan pemberian pinjaman

01

- Tidak terdapat duplikasi data, sehingga dataset sudah bersih dan tidak memerlukan proses deduplikasi.
- Semua kolom memiliki data type yang sesuai dengan nilainya (misalnya, variabel numerik dalam format float/int, variabel kategori dalam format string).
- Encoding telah dilakukan untuk variabel kategorikal agar dapat digunakan dalam machine learning.

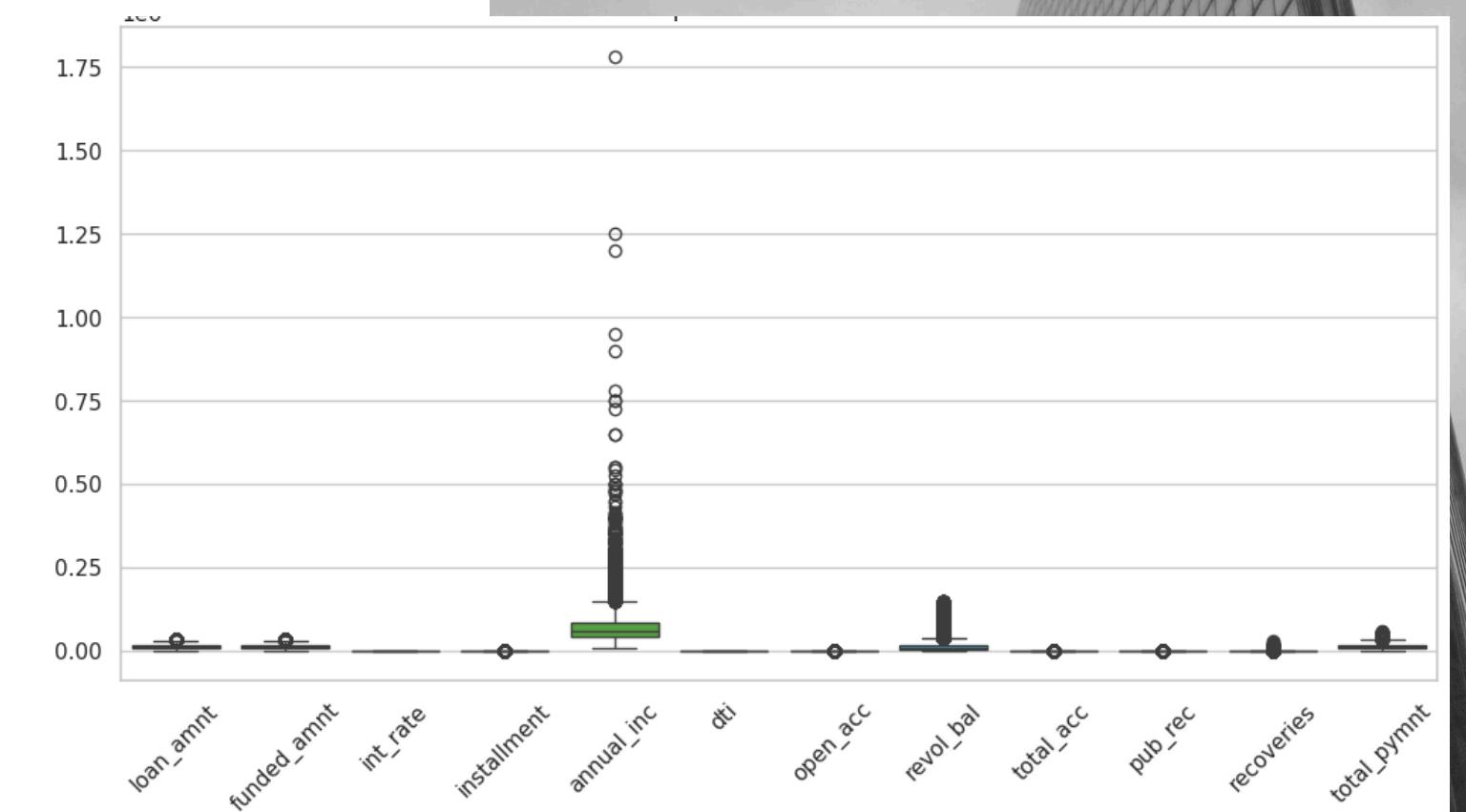
DATA OVERVIEW

```
df.info()
df.duplicated().sum()
df.drop_duplicates(inplace=True)
df.duplicated().sum()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22617 entries, 0 to 22616
Data columns (total 75 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   Unnamed: 0      22617 non-null   int64  
 1   id              22617 non-null   int64  
 2   member_id       22617 non-null   int64  
 3   loan_amnt       22617 non-null   int64  
 4   funded_amnt     22617 non-null   int64  
 5   funded_amnt_inv 22617 non-null   float64 
 6   term             22617 non-null   object  
 7   int_rate         22617 non-null   float64 
 8   installment      22617 non-null   float64 
 9   grade            22617 non-null   object  
 10  sub_grade        22617 non-null   object  
 11  emp_title        21120 non-null   object  
 12  emp_length       21825 non-null   object  
 13  home_ownership   22617 non-null   object  
 14  annual_inc       22617 non-null   float64 
 15  verification_status 22617 non-null   object  
 16  issue_d          22617 non-null   object  
 17  loan_status       22617 non-null   object  
 18  pymnt_plan       22617 non-null   object  
 19  url              22617 non-null   object  
 20  desc              13288 non-null   object  
 21  purpose           22616 non-null   object  
 22  title             22616 non-null   object  
 23  zip_code          22616 non-null   object  
 24  addr_state        22616 non-null   object  
 25  dti               22616 non-null   float64 
 26  delinq_2yrs       22616 non-null   float64 
 27  earliest_cr_line 22616 non-null   object  
 28  inq_last_6mths    22616 non-null   float64 
 29  mths_since_last_delinq 7653 non-null   float64 
 30  mths_since_last_record 1187 non-null   float64 
 31  open_acc          22616 non-null   float64 
 32  pub_rec           22616 non-null   float64 
 33  revol_bal         22616 non-null   float64 
 34  revol_util        22606 non-null   float64 
 35  total_acc          22616 non-null   float64 
 36  initial_list_status 22616 non-null   object  
 37  out_prncp          22616 non-null   float64 
 38  out_prncp_inv     22616 non-null   float64 
 39  total_pymnt        22616 non-null   float64 
 40  total_pymnt_inv    22616 non-null   float64 
 41  total_rec_prncp    22616 non-null   float64 
 42  total_rec_int      22616 non-null   float64 
 43  total_rec_late_fee 22616 non-null   float64 
 44  recoveries         22616 non-null   float64 
 45  collection_recovery_fee 22616 non-null   float64 
 46  last_pymnt_d       22578 non-null   object  
 47  last_pymnt_amnt    22616 non-null   float64 
 48  next_pymnt_d       2056 non-null    object  
 49  last_credit_pull_d 22616 non-null   object  
 50  collections_12_mths_ex_med 22616 non-null   float64 
 51  mths_since_last_major_derog 0 non-null    float64 
 52  policy_code        22616 non-null   float64
```

OUTLIER

- Dari boxplot, terlihat bahwa annual_inc memiliki banyak outlier dengan nilai yang sangat tinggi, menunjukkan beberapa peminjam berpenghasilan jauh di atas mayoritas.
- Recoveries menunjukkan adanya outlier yang cukup tinggi, yang mungkin terkait dengan kredit macet dan pemulihan dana yang tidak merata.
- Total Payment juga memiliki distribusi dengan ekor panjang, menunjukkan beberapa peminjam yang membayar dalam jumlah jauh lebih besar dibandingkan yang lain.

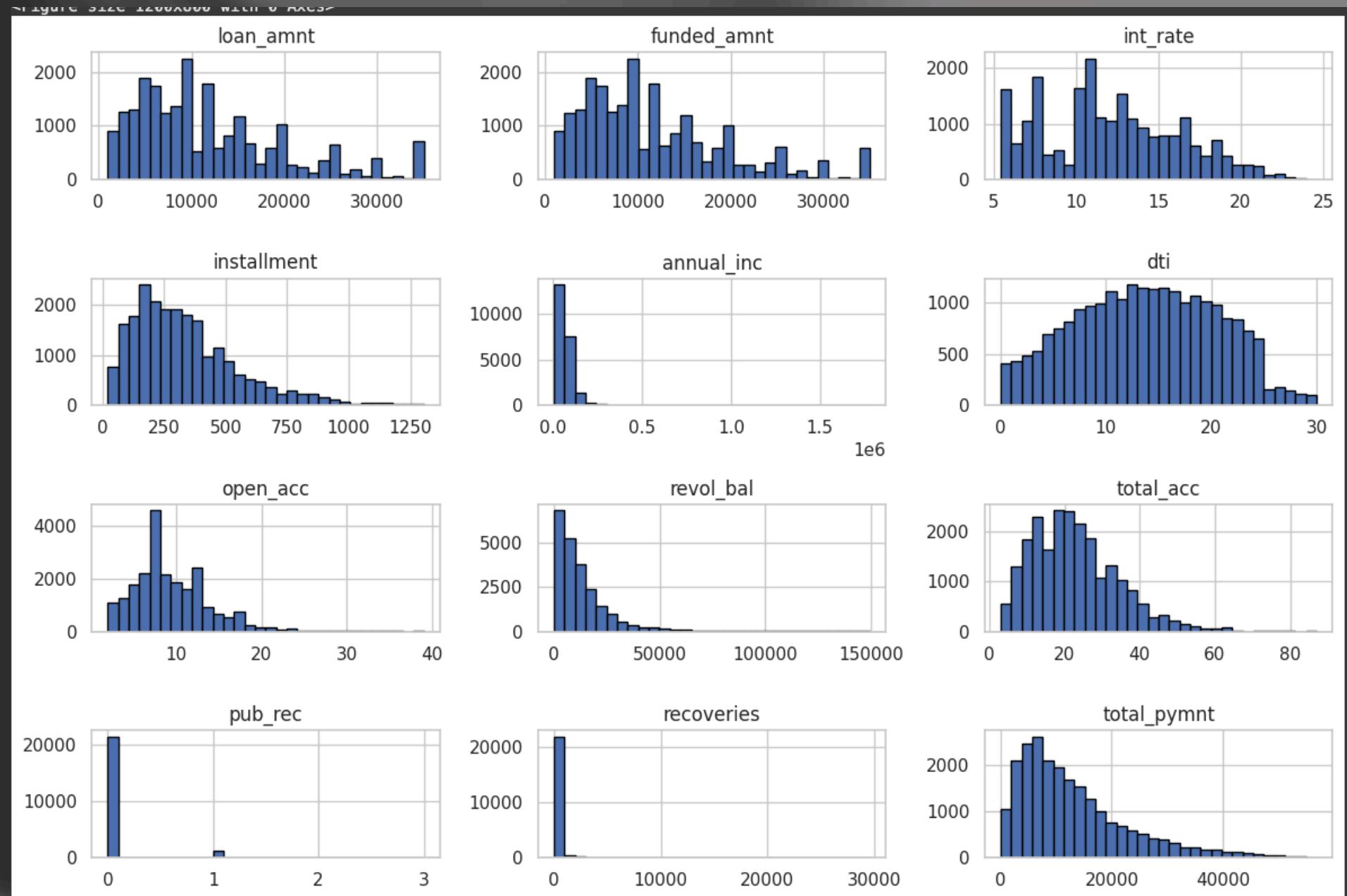


EXPLORATORY DATA ANALYSIS (EDA)

Kolom Numerik yang Digunakan

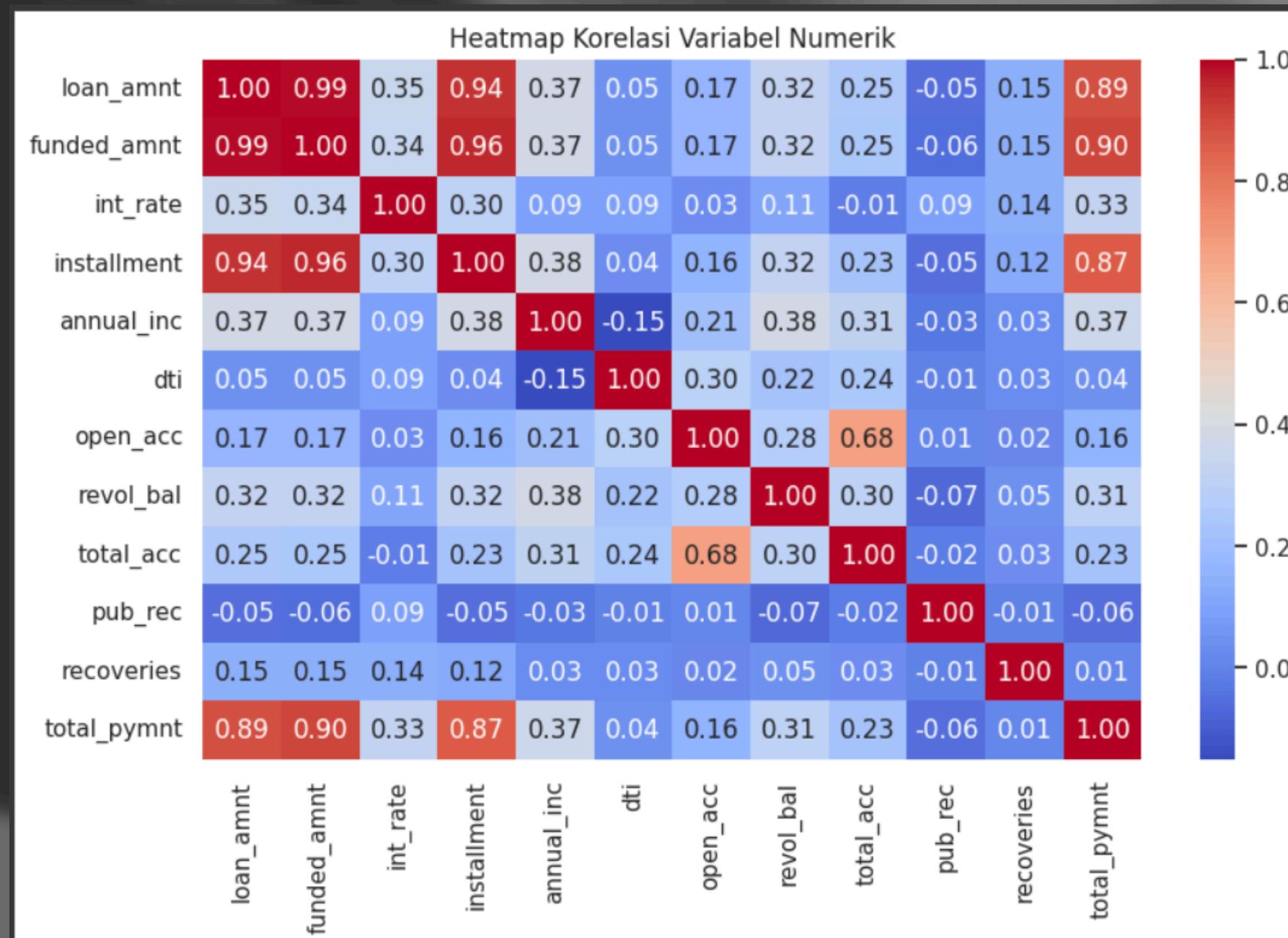
- Dataset ini berisi variabel numerik yang berpengaruh terhadap risiko gagal bayar, seperti loan amount, interest rate, debt-to-income ratio (DTI), dan total payment. Variabel ini memberikan gambaran tentang kondisi finansial peminjam dan kemampuannya dalam membayar pinjaman.
- Faktor yang Mempengaruhi Risiko
- Loan Amount & Funded Amount → Semakin besar pinjaman, semakin tinggi risiko gagal bayar.
- Interest Rate → Bunga tinggi meningkatkan beban finansial peminjam.
- DTI (Debt-to-Income Ratio) → DTI tinggi menunjukkan risiko pembayaran yang lebih besar.
- Public Records (Pub Rec) → Adanya riwayat kredit macet meningkatkan risiko gagal bayar.
- Recoveries → Jika banyak dana yang dipulihkan dari kredit macet, maka risiko gagal bayar di dataset ini tinggi.

UNIVARIATE ANALYSIS

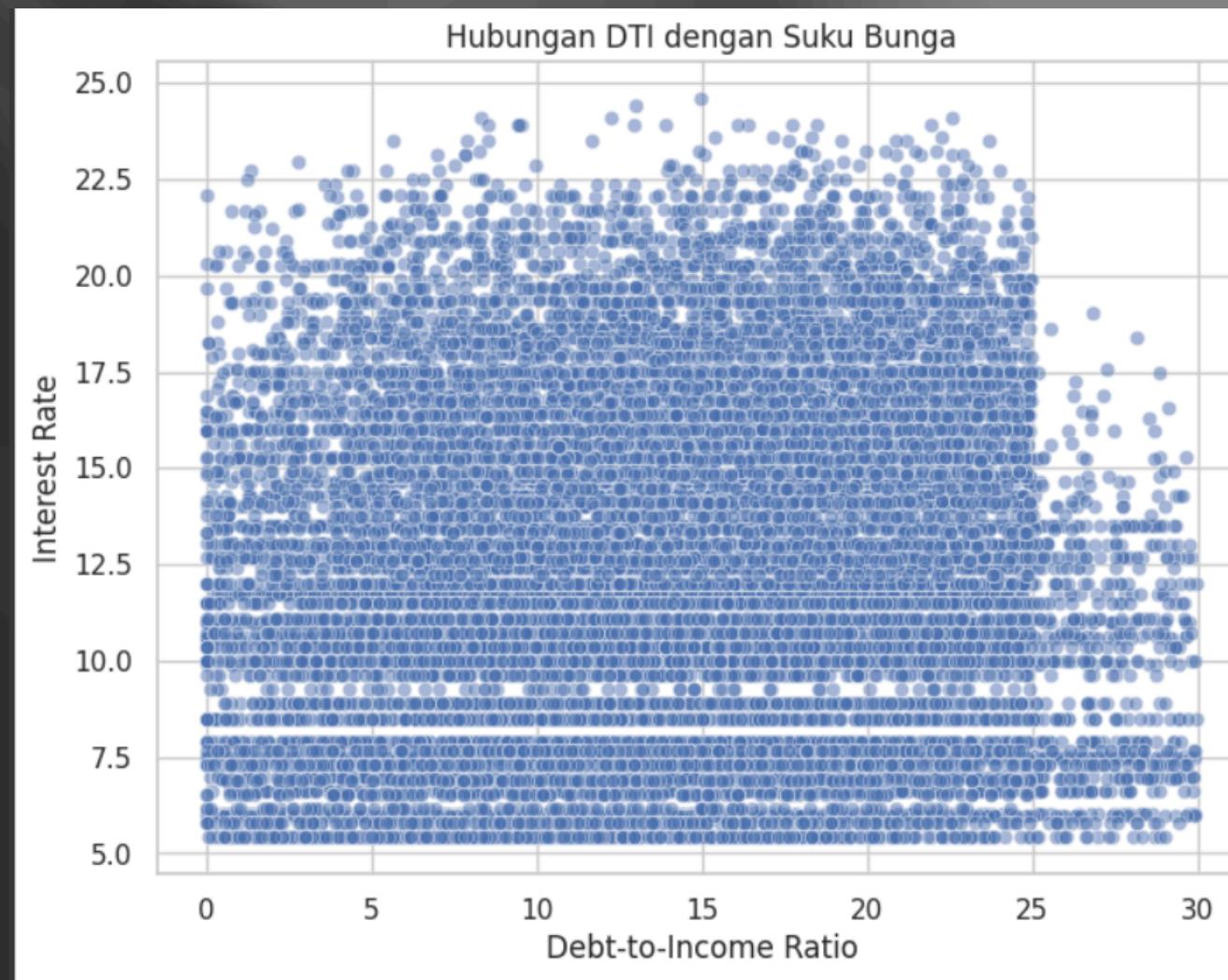


- Loan Amount dan Funded Amount menunjukkan distribusi yang bervariasi, dengan beberapa peminjam mengambil pinjaman dalam jumlah besar.
- Annual Income sangat bervariasi, dengan mayoritas peminjam berpenghasilan rendah hingga menengah.
- Interest Rate tersebar dalam kisaran 5% - 25%, menunjukkan variasi risiko kredit yang signifikan
- Mayoritas peminjam memiliki DTI antara 10-30%, yang menunjukkan tingkat utang yang cukup tinggi dibandingkan pendapatan mereka.
- Open Accounts dan Total Accounts memiliki distribusi yang cenderung miring ke kanan, menunjukkan sebagian besar peminjam memiliki sedikit akun kredit.
- Public Records (catatan gagal bayar) dan Recoveries (jumlah dana yang dipulihkan dari kredit macet) memiliki distribusi yang menunjukkan beberapa peminjam dengan riwayat kredit bermasalah.

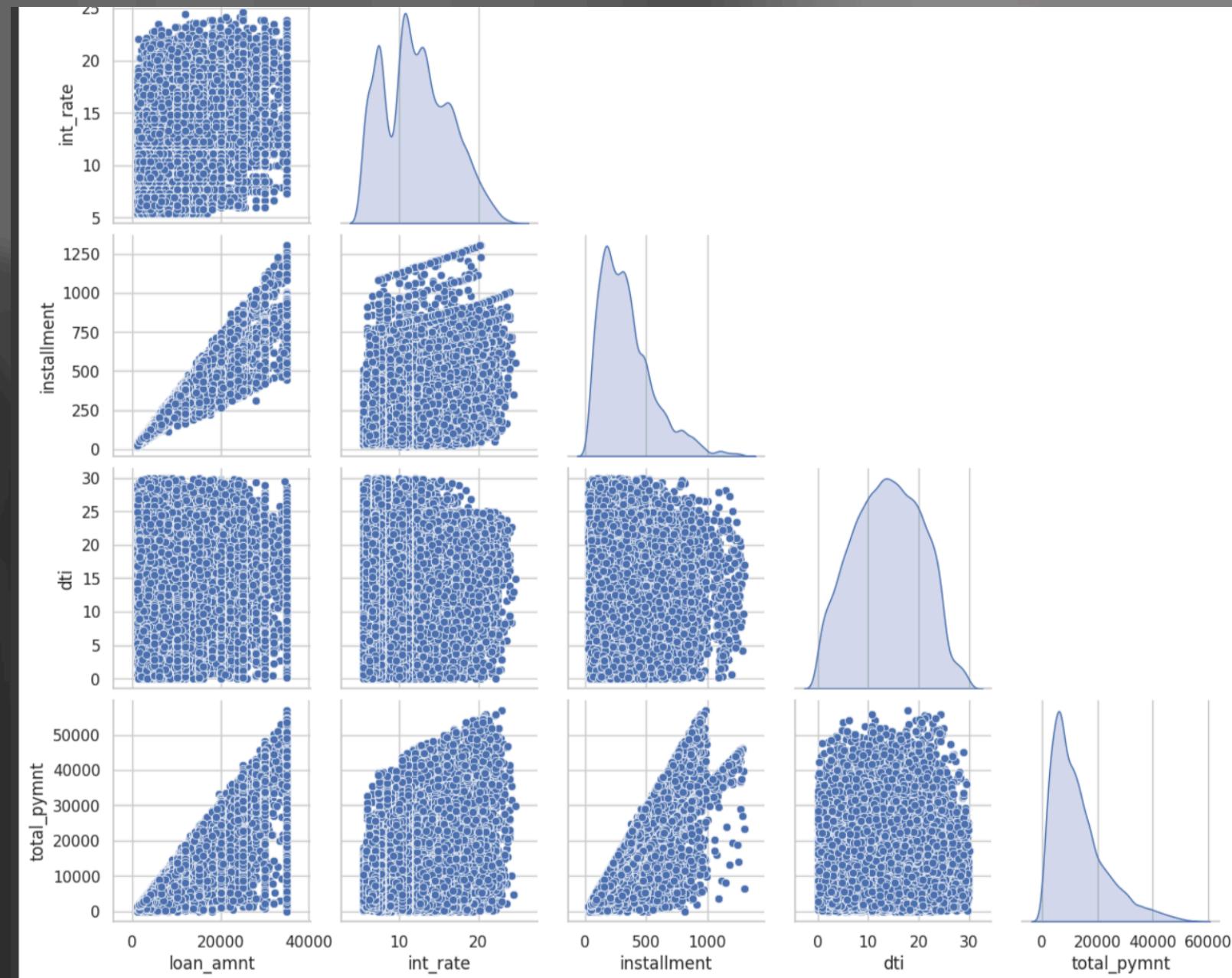
MULTIVARIATE ANALYSIS



- Loan Amount memiliki korelasi sangat tinggi dengan Funded Amount (0.99), yang menunjukkan bahwa jumlah pinjaman yang diajukan hampir selalu sama dengan jumlah yang disetujui.
- Installment memiliki korelasi kuat dengan Loan Amount (0.94), artinya semakin besar jumlah pinjaman, semakin besar pula cicilan yang harus dibayarkan.
- Annual Income memiliki korelasi yang relatif rendah dengan jumlah pinjaman (0.37), yang berarti pendapatan tidak selalu menjadi faktor utama dalam menentukan jumlah pinjaman.
- Open Account dan Total Account memiliki korelasi sebesar 0.68, menandakan bahwa semakin banyak akun kredit yang dibuka, semakin tinggi total akun yang dimiliki seseorang.



- Scatterplot menunjukkan bahwa suku bunga (interest rate) tidak memiliki pola hubungan yang jelas dengan DTI.
- Sebaran data menunjukkan bahwa ada kelompok tertentu yang mendapatkan suku bunga lebih tinggi meskipun memiliki rasio utang terhadap pendapatan (DTI) yang rendah.



MULTIVARIATE ANALYSIS

- **loan_amnt** dan **installment** menunjukkan korelasi positif yang kuat dan cenderung linear. Ini sangat masuk akal, karena semakin besar jumlah pinjaman, biasanya semakin besar pula cicilan bulanannya. Titik-titik membentuk pola garis lurus yang naik dari kiri bawah ke kanan atas.
- **loan_amnt** dan **int_rate** menunjukkan tidak ada korelasi yang kuat atau jelas. Titik-titik tersebar secara acak tanpa membentuk pola yang signifikan. Ini mengindikasikan bahwa jumlah pinjaman tidak secara langsung menentukan tingkat bunga, atau mungkin ada faktor lain yang lebih dominan mempengaruhi tingkat bunga.
- **loan_amnt** dan **total_pymnt** menunjukkan korelasi positif yang kuat dan cenderung linear. Semakin besar jumlah pinjaman awal, semakin besar pula total pembayaran yang telah dilakukan (hingga saat ini). Polanya mirip dengan hubungan antara **loan_amnt** dan **installment**.
- **int_rate** dan **installment** menunjukkan korelasi positif yang lemah hingga sedang. Terdapat kecenderungan bahwa tingkat bunga yang lebih tinggi berhubungan dengan cicilan yang lebih tinggi, tetapi sebarannya lebih luas dibandingkan dengan hubungan antara **loan_amnt** dan **installment**.
- **int_rate** dan **total_pymnt** menunjukkan korelasi positif yang lemah hingga sedang. Pinjaman dengan tingkat bunga yang lebih tinggi cenderung menghasilkan total pembayaran yang lebih tinggi seiring waktu, tetapi faktor lain seperti durasi pinjaman juga memainkan peran penting.
- **installment** dan **total_pymnt** menunjukkan korelasi positif yang kuat dan cenderung linear. Cicilan bulanan yang lebih tinggi akan menghasilkan total pembayaran yang lebih besar dalam periode waktu tertentu.

DATA QUALITY ASSESSMENT & DATA CLEANSING

- Data menunjukkan beberapa kolom memiliki jumlah missing values yang sangat tinggi, yaitu 22617.
- Penghapusan Kolom: Menghapus kolom (axis=1) yang memiliki lebih dari 50% nilai yang hilang
- Pengisian Nilai Numerik: Mengisi nilai-nilai yang hilang pada kolom numerik dengan nilai median dari kolom tersebut.
- Pengisian Nilai Kategorik: Mengisi nilai yang hilang dengan nilai modus (nilai yang paling sering muncul) dari masing-masing kolom.

```
▶ #Data Cleaning and Handling Missing Values
# Mengecek jumlah missing values
print(df.isnull().sum())

# Menghapus kolom dengan terlalu banyak missing values
df.dropna(axis=1, thresh=len(df)*0.5, inplace=True) # Hapus kolom dengan >50% missing

# Mengisi missing values numerik dengan median
df.fillna(df.median(numeric_only=True), inplace=True)

# Mengisi missing values kategorik dengan mode
for col in df.select_dtypes(include=['object']):
    df[col] = df[col].fillna(df[col].mode()[0])


```

	Unnamed: 0	0
id	0	0
member_id	0	0
loan_amnt	0	0
funded_amnt	0	0
	...	
all_util	22617	22617
total_rev_hi_lim	22617	22617
inq_fi	22617	22617
total_cu_tl	22617	22617
inq_last_12m	22617	22617
Length:	75,	dtype: int64

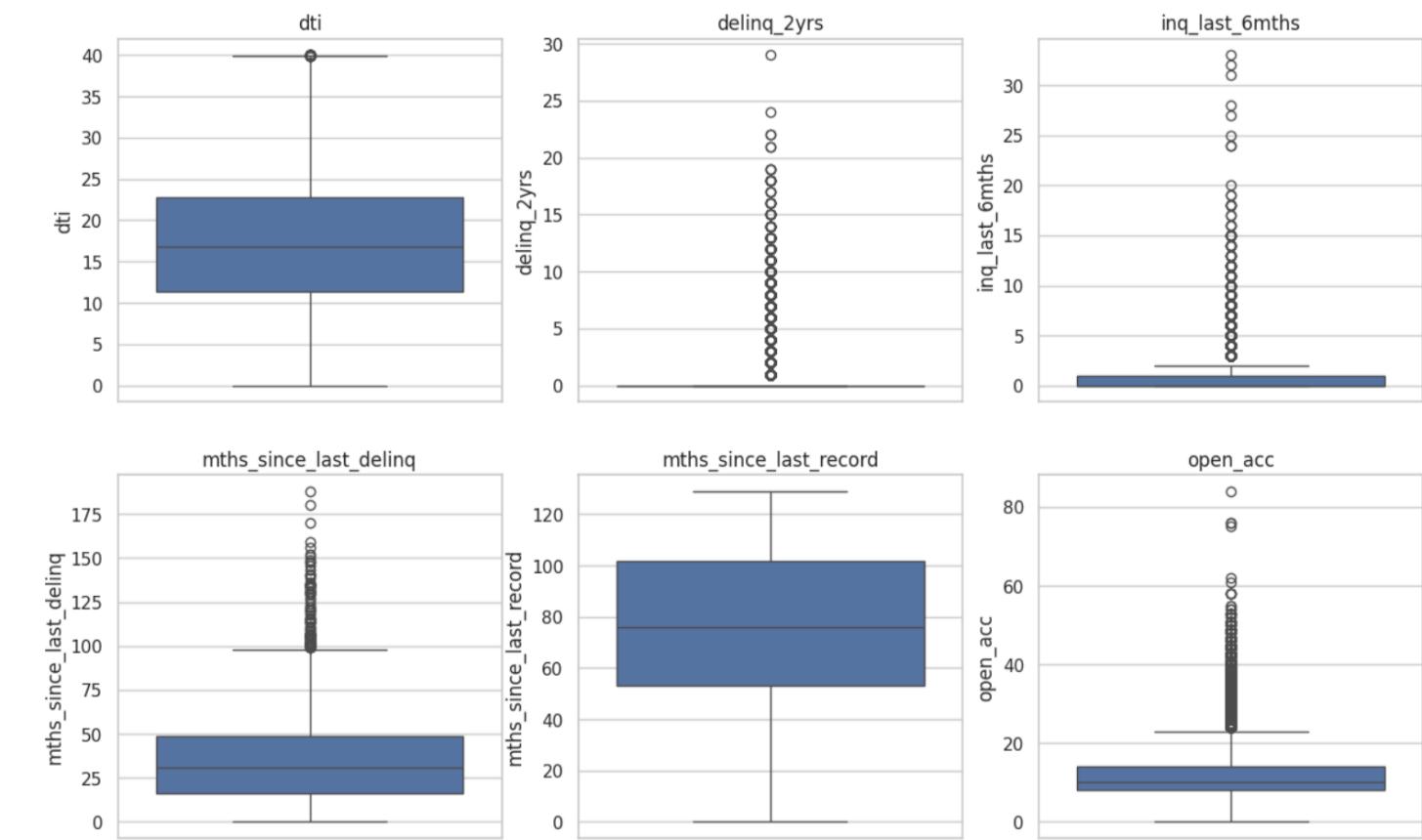
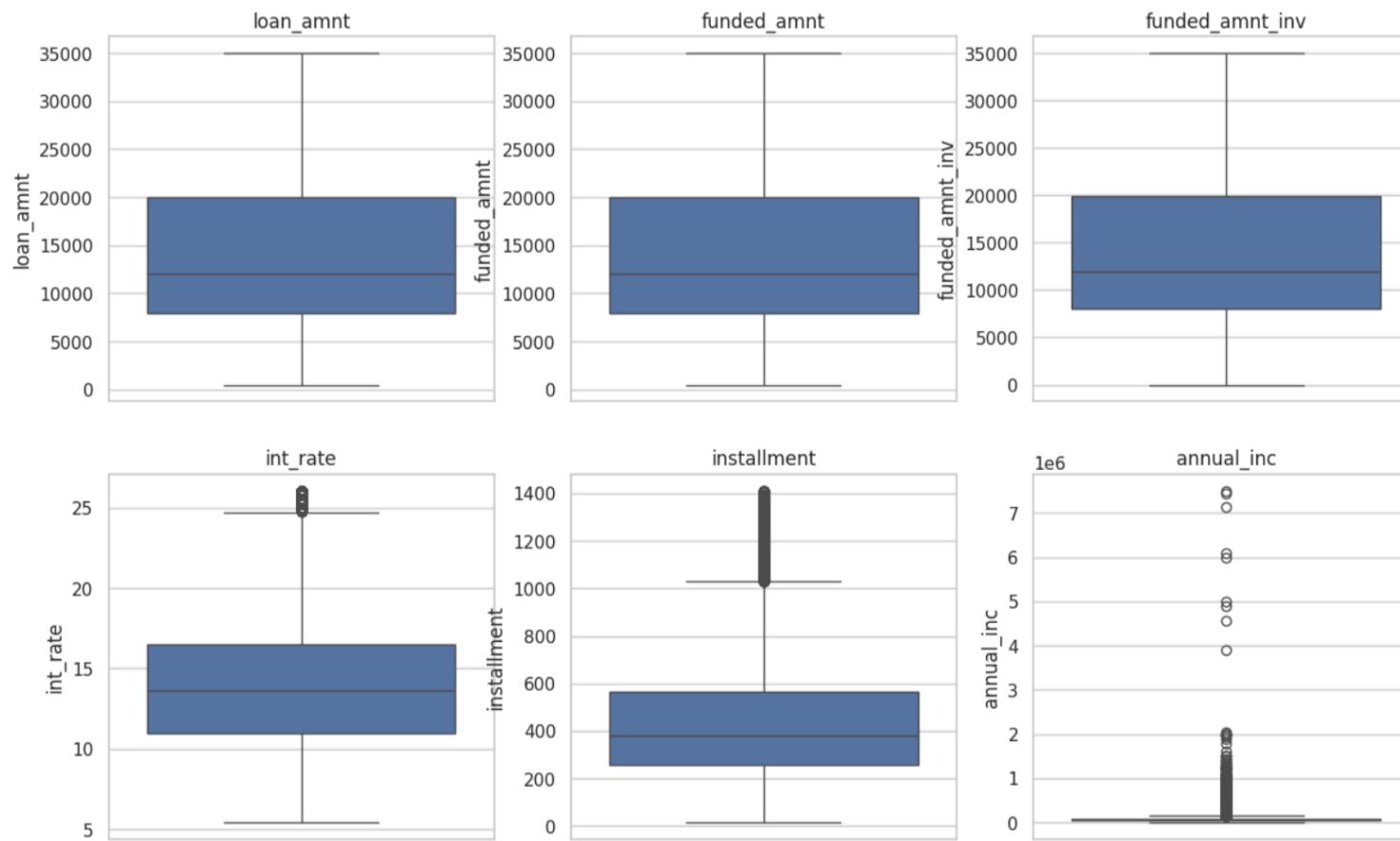
DATA QUALITY ASSESSMENT & DATA CLEANSING

```
➡ Unique values in categorical columns:  
term: ['36 months' '60 months']  
grade: ['B' 'C' 'A' 'E' 'F' 'D' 'G']  
sub_grade: ['B2' 'C4' 'C5' 'C1' 'B5' 'A4' 'E1' 'F2' 'C3' 'B1']  
emp_title: [nan 'Ryder' 'AIR RESOURCES BOARD' 'University Medical Group'  
'Veolia Transportaton' 'Southern Star Photography' 'MKC Accounting '  
'Starbucks' 'Southwest Rural metro' 'UCLA']  
emp_length: ['10+ years' '< 1 year' '1 year' '3 years' '8 years' '9 years' '4 years'  
'5 years' '6 years' '2 years']  
home_ownership: ['RENT' 'OWN' 'MORTGAGE' 'OTHER']  
verification_status: ['Verified' 'Source Verified' 'Not Verified']  
issue_d: ['Dec-11' 'Nov-11' 'Oct-11' 'Sep-11' 'Aug-11' 'Jul-11' 'Jun-11' 'May-11'  
'Apr-11' 'Mar-11']  
loan_status: ['Fully Paid' 'Charged Off' 'Current' 'Default' 'Late (31-120 days)'  
'In Grace Period' 'Late (16-30 days)']  
 pymnt_plan: ['n']  
url: ['https://www.lendingclub.com/browse/loanDetail.action?loan_id=1077501'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1077430'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1077175'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1075863'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1075358'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1075269'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1069639'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1072053'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1071795'  
'https://www.lendingclub.com/browse/loanDetail.action?loan_id=1071570']  
desc: [' Borrower added on 12/22/11 > I need to upgrade my business technologies.<br>'  
' Borrower added on 12/22/11 > I plan to use this money to finance the motorcycle i am looking at. I plan to have it paid off as soon as possible/when i sell my old bike. I only need this money because the deal im looking at is to good to pass up.  
nan  
" Borrower added on 12/21/11 > to pay for property tax (borrow from friend, need to pay back) & central A/C need to be replace. I'm very sorry to let my loan expired last time.<br>"  
" Borrower added on 12/21/11 > I plan on combining three large interest bills together and freeing up some extra each month to pay toward other bills. I've always been a good payor but have found myself needing to make adjustments to my budget.  
' Borrower added on 12/18/11 > I am planning on using the funds to pay off two retail credit cards with 24.99% interest rates, as well as a major bank credit card with a 18.99% rate. I pay all my bills on time, looking for a lower combined payment.  
' Borrower added on 12/16/11 > Downpayment for a car.<br>'  
' Borrower added on 12/21/11 > I own a small home-based judgment collection business. I have 5 years experience collecting debts. I am now going from a home office to a small office. I also plan to buy a small debt portfolio (eg. $10K for $1M)  
" Borrower added on 12/16/11 > I'm trying to build up my credit history. I live with my brother and have no car payment or credit cards. I am in community college and work full time. Im going to use the money to make some repairs around the house.  
' Borrower added on 12/15/11 > I had received a loan from Citi Financial about a year ago, I was paying 29.99 intrest, so the refinance is to cut that rate since cleaning up my credit I have been paying everything on time as shown on my credit report.  
purpose: ['credit_card' 'car' 'small_business' 'other' 'wedding'  
'debt_consolidation' 'home_improvement' 'major_purchase' 'medical'  
'moving']  
title: ['Computer' 'bike' 'real estate business' 'personel' 'Personal'  
'My wedding loan I promise to pay back' 'Loan' 'Car Downpayment'  
'Expand Business & Buy Debt Portfolio' 'Building my credit history.'][br/>zip_code: ['860xx' '309xx' '606xx' '917xx' '972xx' '852xx' '280xx' '900xx' '958xx'  
'774xx']  
addr_state: ['AZ' 'GA' 'IL' 'CA' 'OR' 'NC' 'TX' 'VA' 'MO' 'CT']  
earliest_cr_line: ['Jan-85' 'Apr-99' 'Nov-01' 'Feb-96' 'Jan-96' 'Nov-04' 'Jul-05' 'Jan-07'  
'Apr-04' 'Sep-04']  
initial_list_status: ['f' nan]  
last_pymnt_d: ['Jan-15' 'Apr-13' 'Jun-14' 'Jan-16' 'Apr-12' 'Nov-12' 'Jun-13' 'Sep-13'  
'Jul-12' 'Oct-13']  
next_pymnt_d: [nan 'Feb-16' 'Jan-16']  
last_credit_pull_d: ['Jan-16' 'Sep-13' 'Jan-15' 'Sep-15' 'Dec-14' 'Aug-12' 'Mar-13' 'Dec-15'  
'Aug-13' 'Nov-12']  
application_type: ['INDIVIDUAL' nan]
```

Ada beberapa inconsistent entries

Kolom pymnt_plan: Terlihat nilai yang kosong ("") dan kemungkinan nilai lain. Kolom title: Terlihat beberapa contoh judul pinjaman yang diberikan oleh peminjam. Di sini, potensi inkonsistensi sangat tinggi karena peminjam bisa memasukkan teks yang bebas.

HANDLE OUTLIERS



```
# Outlier
numeric_cols = df.select_dtypes(include=[np.number]).columns
num_cols = len(numeric_cols)
fig, axes = plt.subplots(nrows=(num_cols // 3) + 1, ncols=3, figsize=(15, 5 * ((num_cols // 3) + 1)))
axes = axes.flatten()

for i, col in enumerate(numeric_cols):
    sns.boxplot(y=df[col], ax=axes[i])
    axes[i].set_title(col)

# Hapus subplot kosong jika ada
for j in range(i + 1, len(axes)):
    fig.delaxes(axes[j])

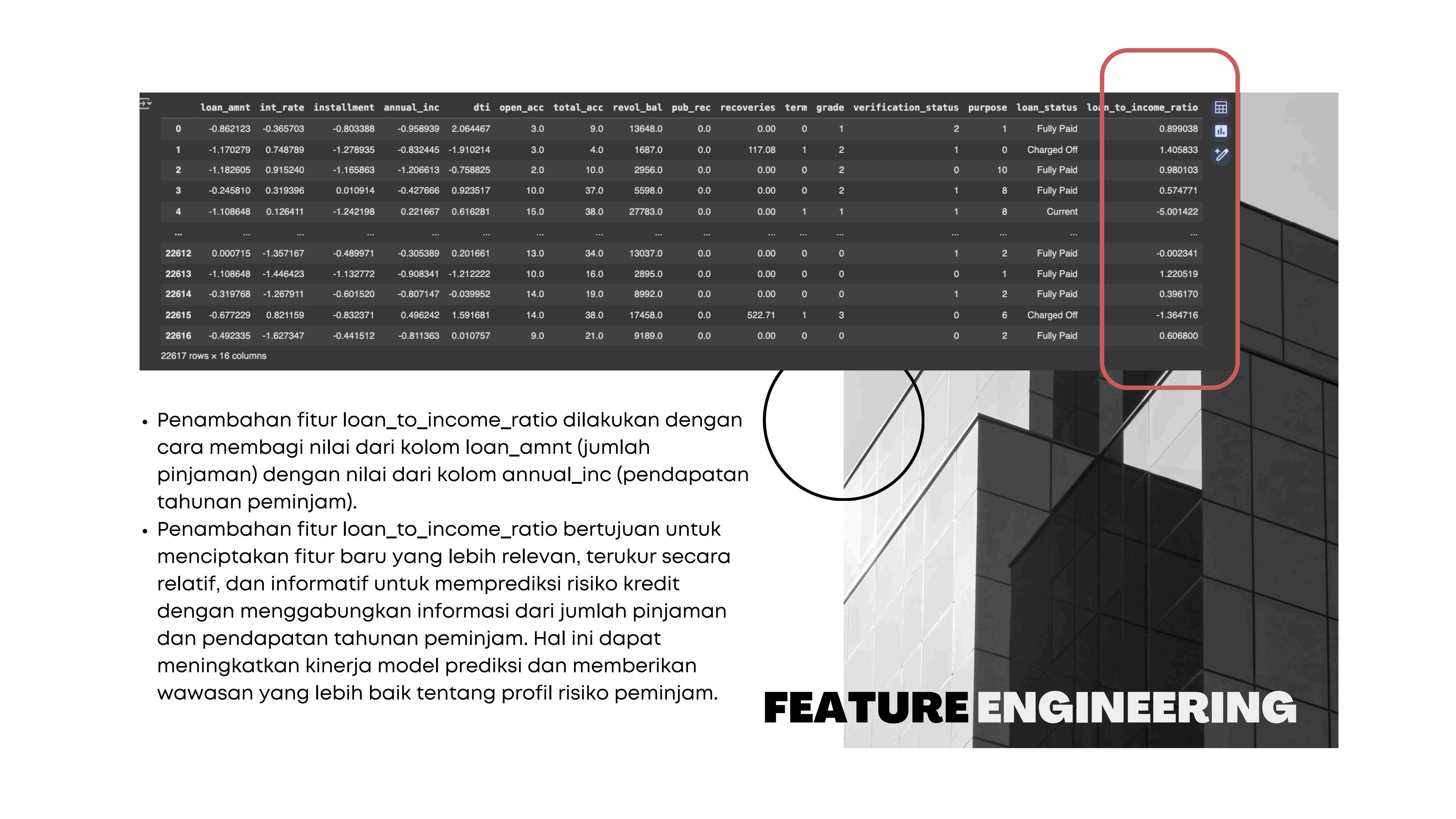
plt.tight_layout()
plt.show()

# Mendeksi Outlier Menggunakan IQR
def detect_outliers_iqr(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = data[(data[column] < lower_bound) | (data[column] > upper_bound)]
    return outliers

outlier_counts = {}
for col in numeric_cols:
    outliers = detect_outliers_iqr(df, col)
    outlier_counts[col] = len(outliers)

print("\nJumlah outliers per kolom (IQR Method):")
print(outlier_counts)
```

Beberapa variabel, terutama `installment` dan `annual_inc`, menunjukkan sejumlah besar outlier di sisi atas. `loan_amnt`, `funded_amnt`, `funded_amnt_inv`, dan `int_rate` juga memiliki outlier di sisi atas. `annual_inc` memiliki satu outlier di sisi bawah. Outlier ini dapat mempengaruhi statistik deskriptif dan berpotensi memengaruhi model statistik



	loan_amnt	int_rate	installment	annual_inc	dti	open_acc	total_acc	revol_bal	pub_rec	recoveries	term	grade	verification_status	purpose	loan_status	loan_to_income_ratio	
0	-0.862123	-0.365703	-0.803388	-0.958939	2.064467	3.0	9.0	13648.0	0.0	0.00	0	1		2	1	Fully Paid	0.899038
1	-1.170279	0.748789	-1.278935	-0.832445	-1.910214	3.0	4.0	1687.0	0.0	117.08	1	2		1	0	Charged Off	1.405833
2	-1.182605	0.915240	-1.165863	-1.206613	-0.758825	2.0	10.0	2956.0	0.0	0.00	0	2		0	10	Fully Paid	0.980103
3	-0.245810	0.319396	0.010914	-0.427666	0.923517	10.0	37.0	5598.0	0.0	0.00	0	2		1	8	Fully Paid	0.574771
4	-1.108648	0.126411	-1.242198	0.221667	0.616281	15.0	38.0	27783.0	0.0	0.00	1	1		1	8	Current	-5.001422
...	
22612	0.000715	-1.357167	-0.489971	-0.305389	0.201661	13.0	34.0	13037.0	0.0	0.00	0	0		1	2	Fully Paid	-0.002341
22613	-1.108648	-1.446423	-1.132772	-0.908341	-1.212222	10.0	16.0	2895.0	0.0	0.00	0	0		0	1	Fully Paid	1.220519
22614	-0.319768	-1.267911	-0.601520	-0.807147	-0.039952	14.0	19.0	8992.0	0.0	0.00	0	0		1	2	Fully Paid	0.396170
22615	-0.677229	0.821159	-0.832371	0.496242	1.591681	14.0	38.0	17458.0	0.0	522.71	1	3		0	6	Charged Off	-1.364716
22616	-0.492335	-1.627347	-0.441512	-0.811363	0.010757	9.0	21.0	9189.0	0.0	0.00	0	0		0	2	Fully Paid	0.606800

22617 rows × 16 columns

- Penambahan fitur `loan_to_income_ratio` dilakukan dengan cara membagi nilai dari kolom `loan_amnt` (jumlah pinjaman) dengan nilai dari kolom `annual_inc` (pendapatan tahunan peminjam).
- Penambahan fitur `loan_to_income_ratio` bertujuan untuk menciptakan fitur baru yang lebih relevan, terukur secara relatif, dan informatif untuk memprediksi risiko kredit dengan menggabungkan informasi dari jumlah pinjaman dan pendapatan tahunan peminjam. Hal ini dapat meningkatkan kinerja model prediksi dan memberikan wawasan yang lebih baik tentang profil risiko peminjam.

FEATURE ENGINEERING



MODELLING

Pembagian data dengan train-test split memastikan model memiliki data train yang cukup serta data test yang representatif untuk melihat performa model secara obyektif. Sedangkan StandardScaler membantu meningkatkan stabilitas model dengan mengurangi pengaruh outlier, sehingga model dapat melakukan prediksi dengan lebih baik pada distribusi data.

```
X = df.drop(columns=["loan_status"]) # Fitur  
y = df["loan_status"] # Target  
  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  
  
[22] from sklearn.preprocessing import StandardScaler  
  
scaler = StandardScaler()  
X_train_scaled = scaler.fit_transform(X_train)  
X_test_scaled = scaler.transform(X_test)
```

MODELLING



Melakukan perbandingan model dari Logistic Regression, Decision Tree Classifier dan Random Forest Classifier

```
log_model = LogisticRegression(max_iter=1000)
log_model.fit(X_train_scaled, y_train)

# Logistic Regression
LogisticRegression(max_iter=1000)

# Decision Tree
tree_model = DecisionTreeClassifier()
tree_model.fit(X_train, y_train)

# Decision Tree Classifier
DecisionTreeClassifier()

# Random Forest
rf_model = RandomForestClassifier()
rf_model.fit(X_train, y_train)

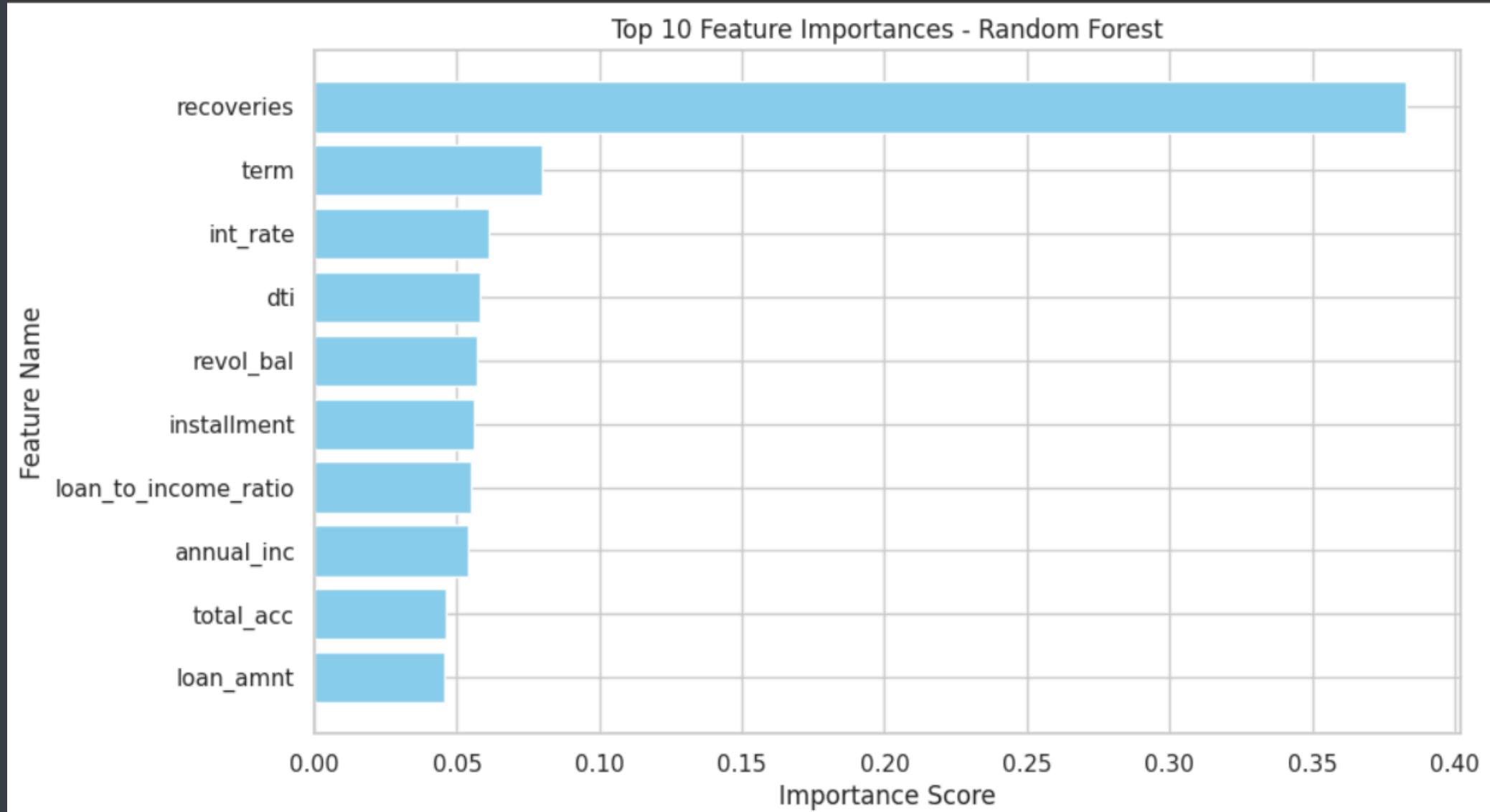
# RandomForestClassifier
RandomForestClassifier()
```

Dari hasil akurasi, Random Forest lebih tinggi tingkat akurasinya dibandingkan dengan model lain, dengan itu memilih model Random Forest

	Model	Accuracy
0	Logistic Regression	0.102122
1	Decision Tree	0.818744
2	Random Forest	0.869584

FEATURE IMPORTANCE

- **recoveries** (Skor sekitar 0.35 - 0.40): Fitur ini memiliki skor kepentingan yang paling tinggi, secara signifikan lebih tinggi dari fitur lainnya. recoveries kemungkinan merujuk pada jumlah uang yang berhasil dipulihkan setelah pinjaman gagal bayar (Charged Off). Kepentingannya yang tinggi menunjukkan bahwa informasi mengenai riwayat pemulihan pinjaman sebelumnya adalah indikator yang sangat kuat dalam memprediksi hasil pinjaman di masa depan. Pinjaman dengan riwayat pemulihan yang rendah (kerugian besar) mungkin menjadi sinyal risiko yang tinggi.
- **term** (Skor sekitar 0.20 - 0.25): Jangka waktu pinjaman (term) merupakan fitur penting kedua. Ini menunjukkan bahwa durasi pinjaman memiliki pengaruh yang signifikan terhadap prediksi. Pinjaman dengan jangka waktu yang lebih panjang mungkin dianggap lebih berisiko karena adanya periode waktu yang lebih lama bagi peminjam untuk mengalami kesulitan keuangan.
- **int_rate** (Skor sekitar 0.15 - 0.20): Tingkat bunga pinjaman (int_rate) juga merupakan fitur yang cukup penting. Tingkat bunga yang lebih tinggi seringkali dikenakan pada peminjam dengan profil risiko yang lebih tinggi, sehingga fitur ini menjadi prediktor yang baik untuk hasil pinjaman.
- **revol_bal** (Skor sekitar 0.05 - 0.10): Saldo revolving kredit (revol_bal) juga berkontribusi pada prediksi. Saldo revolving yang tinggi mungkin mengindikasikan bahwa peminjam memiliki ketergantungan yang besar pada kredit dan berpotensi memiliki kesulitan keuangan.
- **installment** (Skor sekitar 0.05 - 0.10): Jumlah cicilan bulanan (installment) memiliki tingkat kepentingan yang serupa dengan revol_bal. Cicilan yang tinggi relatif terhadap pendapatan peminjam dapat menjadi indikator tekanan finansial dan meningkatkan risiko gagal bayar.
- **loan_to_income_ratio** (Skor sekitar 0.05 - 0.10): Rasio antara jumlah pinjaman dengan pendapatan tahunan (loan_to_income_ratio) juga merupakan fitur penting. Rasio yang tinggi menunjukkan beban utang yang besar relatif terhadap pendapatan, yang dapat meningkatkan risiko gagal bayar.
- **annual_inc** (Skor sekitar 0.04 - 0.05): Pendapatan tahunan peminjam (annual_inc) memiliki tingkat kepentingan yang lebih rendah dibandingkan rasio loan_to_income_ratio, meskipun tetap berkontribusi pada model. Ini mungkin karena model lebih fokus pada beban utang relatif terhadap pendapatan (yang diwakili oleh rasio) daripada pendapatan absolut.
- **total_acc** (Skor sekitar 0.04 - 0.05): Jumlah total akun kredit (total_acc) yang dimiliki peminjam juga menjadi faktor yang dipertimbangkan oleh model. Jumlah akun kredit yang berbeda mungkin mencerminkan pengalaman kredit dan kompleksitas keuangan peminjam.
- **loan_amnt** (Skor sekitar 0.04 - 0.05): Jumlah pinjaman (loan_amnt) memiliki tingkat kepentingan yang serupa dengan annual_inc dan total_acc.



INSIGHT

- Fitur recoveries menjadi prediktor terkuat. Ini mengindikasikan bahwa riwayat dan potensi pemulihan dana setelah gagal bayar adalah faktor penentu utama dalam memprediksi hasil pinjaman di masa depan. Pinjaman yang memiliki potensi pemulihan rendah adalah sumber risiko terbesar.
- Jangka waktu pinjaman (term) dan tingkat bunga (int_rate) memiliki pengaruh besar. Pinjaman dengan jangka waktu lebih panjang dan tingkat bunga lebih tinggi secara inheren membawa risiko yang lebih tinggi.
- Fitur-fitur yang mencerminkan beban utang relatif terhadap kemampuan finansial peminjam (revol_bal, installment, loan_to_income_ratio) juga merupakan indikator risiko yang kuat. Peminjam dengan saldo revolving tinggi, cicilan besar relatif terhadap pendapatan, dan rasio pinjaman terhadap pendapatan yang tinggi lebih berisiko.
- Meskipun pendapatan tahunan (annual_inc) penting, pengaruhnya lebih kecil dibandingkan rasio loan_to_income_ratio. Ini menunjukkan bahwa kemampuan membayar relatif (dibandingkan dengan besarnya pinjaman) adalah pertimbangan yang lebih kritikal daripada sekadar besarnya pendapatan.
- Jumlah pinjaman (loan_amnt) dan jumlah total akun kredit (total_acc) juga memberikan informasi, meskipun dengan skor yang lebih rendah.

BUSINESS RECOMMENDATION

- Investasikan sumber daya dan strategi yang efektif dalam proses pemulihan pinjaman yang gagal bayar. Analisis faktor-faktor yang berkorelasi dengan tingkat pemulihan yang tinggi untuk mengoptimalkan upaya ini.
- Saat menyetujui pinjaman, pertimbangkan aset atau jaminan yang mungkin dimiliki peminjam yang dapat digunakan untuk pemulihan jika terjadi gagal bayar.
- Terapkan persyaratan yang lebih ketat atau tingkat bunga yang lebih tinggi untuk pinjaman dengan jangka waktu yang lebih panjang. Pertimbangkan batasan maksimum jangka waktu untuk profil risiko tertentu.
- Gunakan tingkat bunga (int_rate) sebagai faktor utama dalam penetapan harga pinjaman, secara akurat mencerminkan risiko yang terkait dengan berbagai profil peminjam.
- Tetapkan batas maksimum untuk rasio utang seperti revol_bal terhadap batas kredit dan loan_to_income_ratio sebagai kriteria kelayakan pinjaman.
- Lakukan analisis yang mendalam terhadap kemampuan membayar peminjam, tidak hanya berdasarkan pendapatan absolut tetapi juga mempertimbangkan kewajiban keuangan lainnya (tercermin dalam installment dan revol_bal).
- Dalam penilaian risiko, berikan bobot yang lebih besar pada loan_to_income_ratio dibandingkan hanya melihat pendapatan tahunan. Peminjam dengan pendapatan tinggi tetapi rasio utang yang tinggi tetap berisiko.
- Integrasikan insight dari feature importance ini ke dalam model penilaian risiko dan proses underwriting. Pastikan faktor-faktor paling penting mendapatkan perhatian yang lebih besar dalam pengambilan keputusan persetujuan pinjaman.

**THANK
YOU!**

