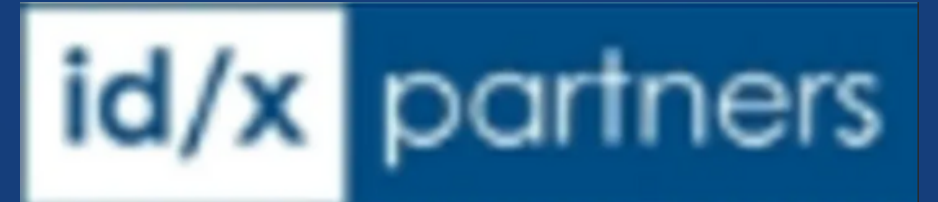




FINAL PROJECT ID/X PARTNERS – VIRTUAL INTERNSHIP



Credit risk Predicition

A Q S H A L M A U L A N A H A F I Z A R D I A N T H A





LATAR BELAKANG

Credit risk merupakan **risiko** yang harus ditanggung oleh sebuah lembaga **pemberi pinjaman** ke seseorang **individu atau instansi** yang **risikonya** sendiri merupakan **tidak bisa dibayarkannya pokok dan bunga pinjaman** sehingga mengakibatkan berbagai macam **kerugian**, seperti **gangguan arus kas** bagi **pemberi pinjaman** dan **meningkatkan biaya penagihan**.

Salah satu **penyebab** terjadinya hal tersebut ialah **penetapan** kepada **calon kreditur** yang **tidak tepat** sehingga menyebabkan **calon kreditur** yang **berpotensi diterima digugurkan** sedangkan **memberikan pinjaman** kepada **calon kreditur** yang **bermasalah**.

Untuk **memperkecil risiko kredit** yang ada, dapat dilakukan **proses penilaian resiko** sebelum diberikan pinjaman yang bisa disebut dengan **credit scoring** dan **credit rating** pada **pihak peminjam**.





TUJUAN

Dengan adanya **Credit scoring** dan **credit rating**, hal ini dapat **memperkecil risiko** yang dapat terjadi pada **lembaga peminjam**, serta dapat menciptakan **kecepatan** dan **efisiensi** untuk **memprediksi** dan **menilai pengajuan kredit** yang **optimal**, dimana prediksi ini **menjadi penentu** apakah pengajuan pinjaman **diterima** atau **ditolak** oleh **lembaga pemberi pinjaman**.

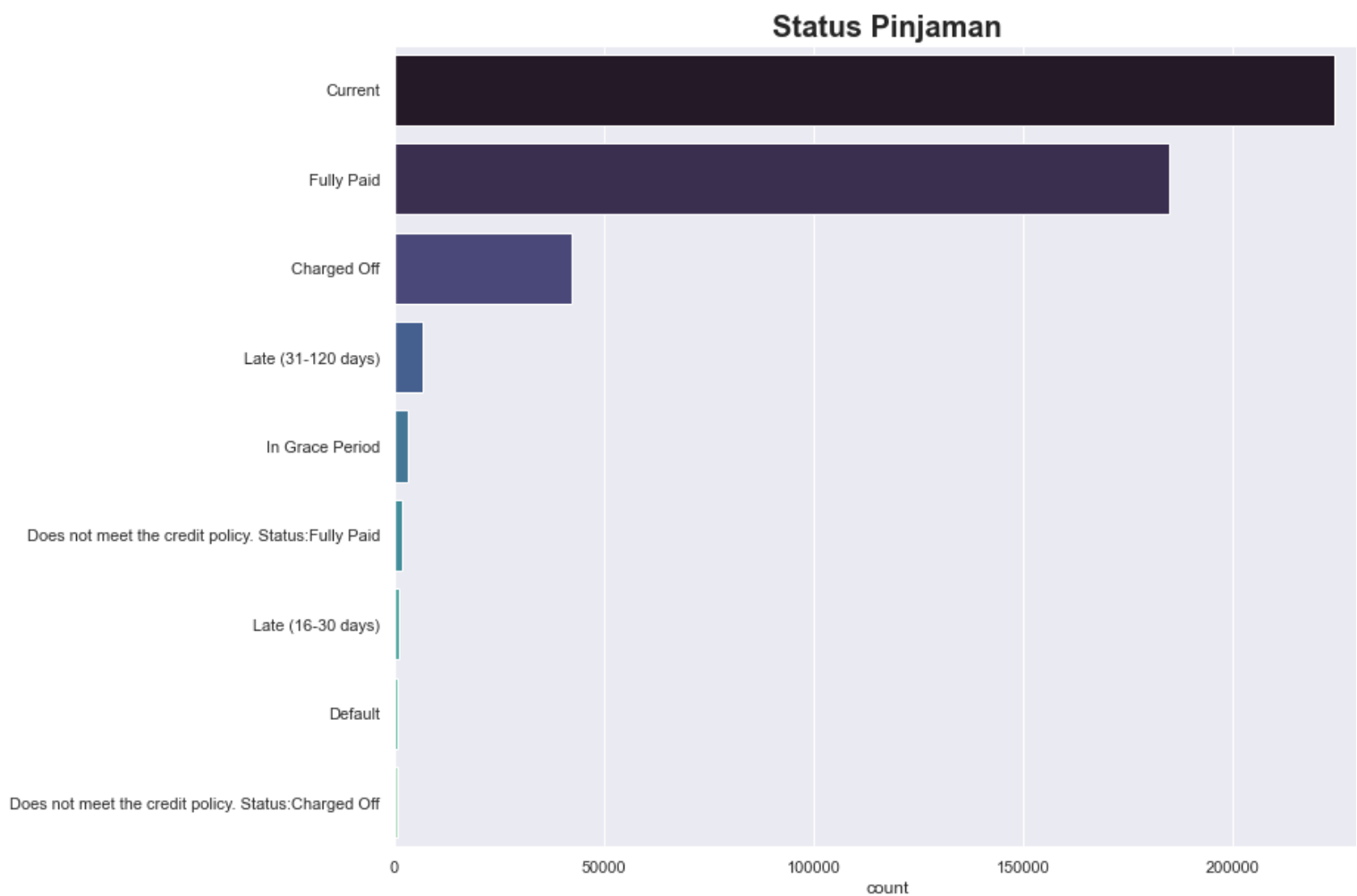
Selain itu, hal tersebut dapat **memperlancar arus kas** dari **perusahaan peminjam** serta membuat **aset** dari perusahaan secara pasti dapat **meningkat**.





FINAL PROJECT ID/X PARTNERS – VIRTUAL INTERNSHIP

TARGET AND MODEL DEFINITION



Data target ditentukan dengan **karakteristik dari para individu** terhadap **pinjaman/kredit** yang mereka bayarkan, apakah individu tersebut termasuk **bad loan (peminjam yang buruk)** atau **good loan (peminjam yang baik)**.

Definisi dari **bad or good loan** dapat berbeda tergantung dengan **kebutuhan bisnis** yang ada. Disini saya menggunakan keterlambatan **diatas 1 bulan/30 hari** dan yang **lebih buruk** dari itu untuk dikategorikan sebagai **bad loan**.

maka kita dapat membagi menjadi **2 kategori** :
good loan = Fully paid
Bad loan = Charged Off, Default, dan Does not meet the credit policy,Late (31-120 days).

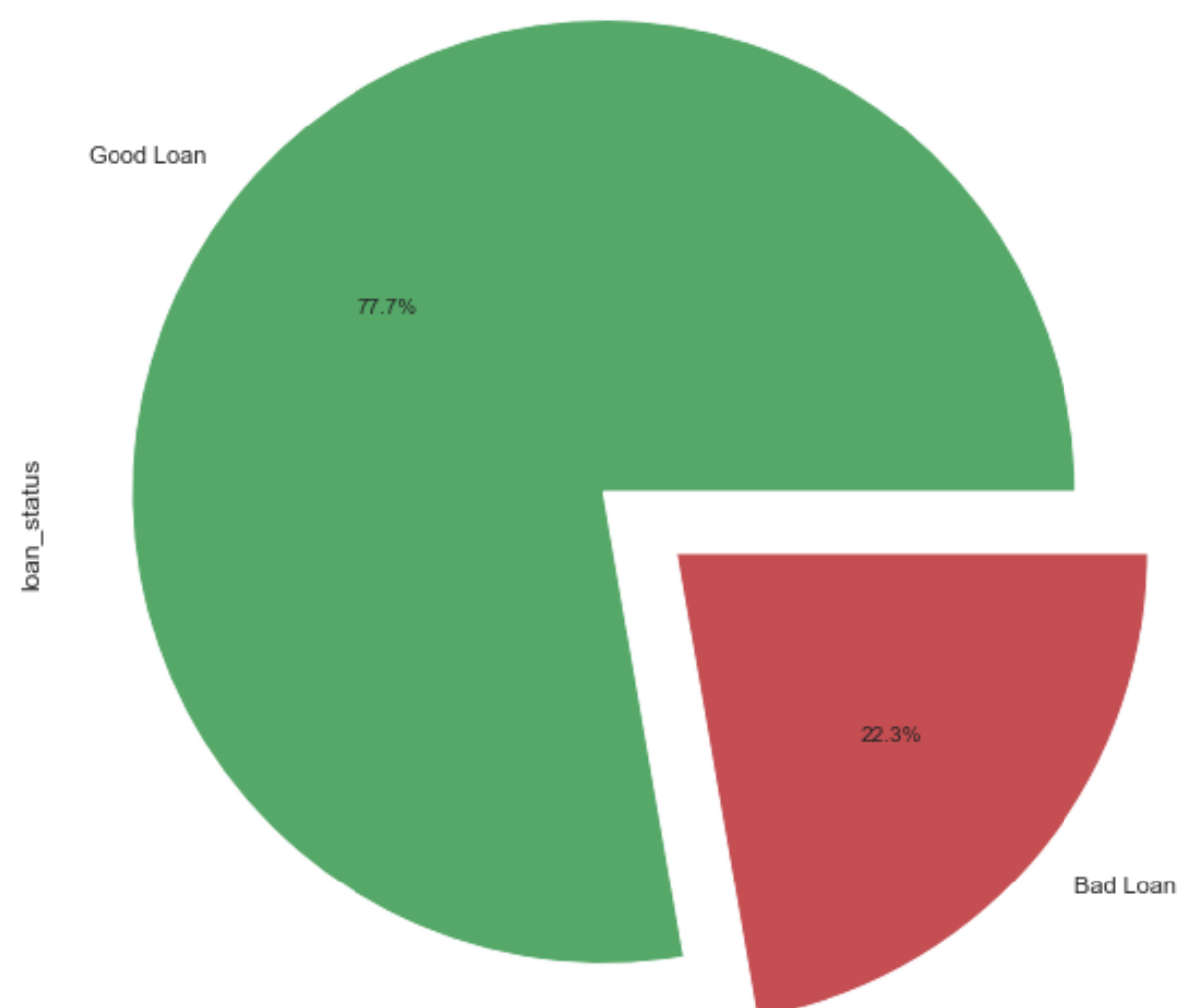
Pada feature **current, in_grace_perid, dan late, tidak dapat** digunakan karena ketiga feature tersebut menandakan bahwa pengajuan peminjaman nasabah **masih berlangsung**

Current	48.09%
Fully Paid	39.62%
Charged Off	9.11%
Late (31-120 days)	1.48%
In Grace Period	0.67%
Does not meet the credit policy. Status:Fully Paid	0.43%
Late (16-30 days)	0.26%
Default	0.18%
Does not meet the credit policy. Status:Charged Off	0.16%
Name: loan_status, dtype: object	





Proposi Status Peminjaman



Dengan **Good Loan** dilabeli dengan **1** yaitu peminjam dengan **karakteristik yang baik** dengan presentase sebesar **77,7%**

Bad Loan dilabel dengan **0** yaitu peminjam dengan **karakteristik buruk** dengan presentase sebesar **22,3%**

Dapat dilihat bahwa proporsi **peminjaman yang baik** jauh **lebih besar** dibandingkan Peminjam yang buruk.

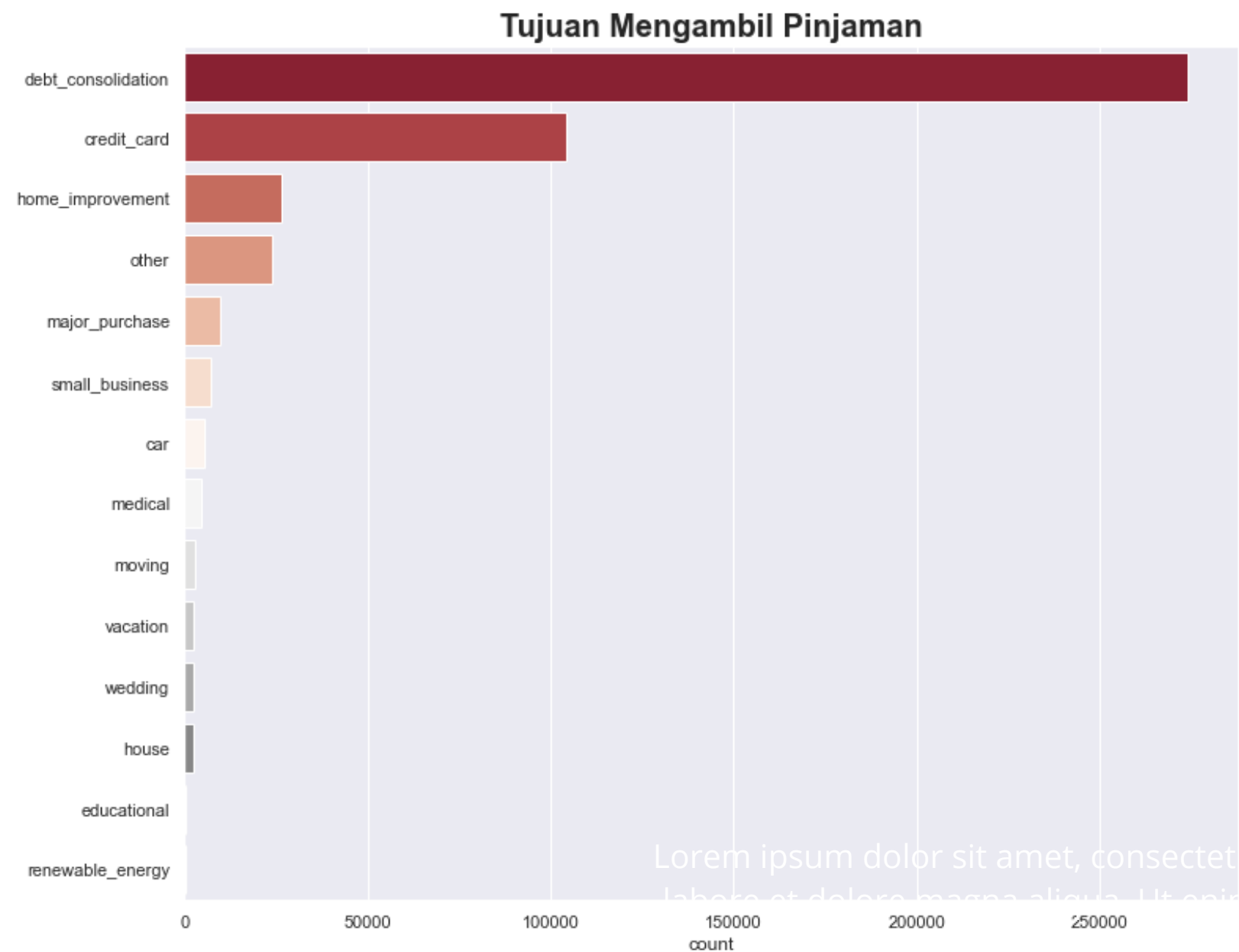




FINAL PROJECT ID/X PARTNERS – VIRTUAL INTERNSHIP

EXPLORATORY DATA ANALYSIS (EDA)

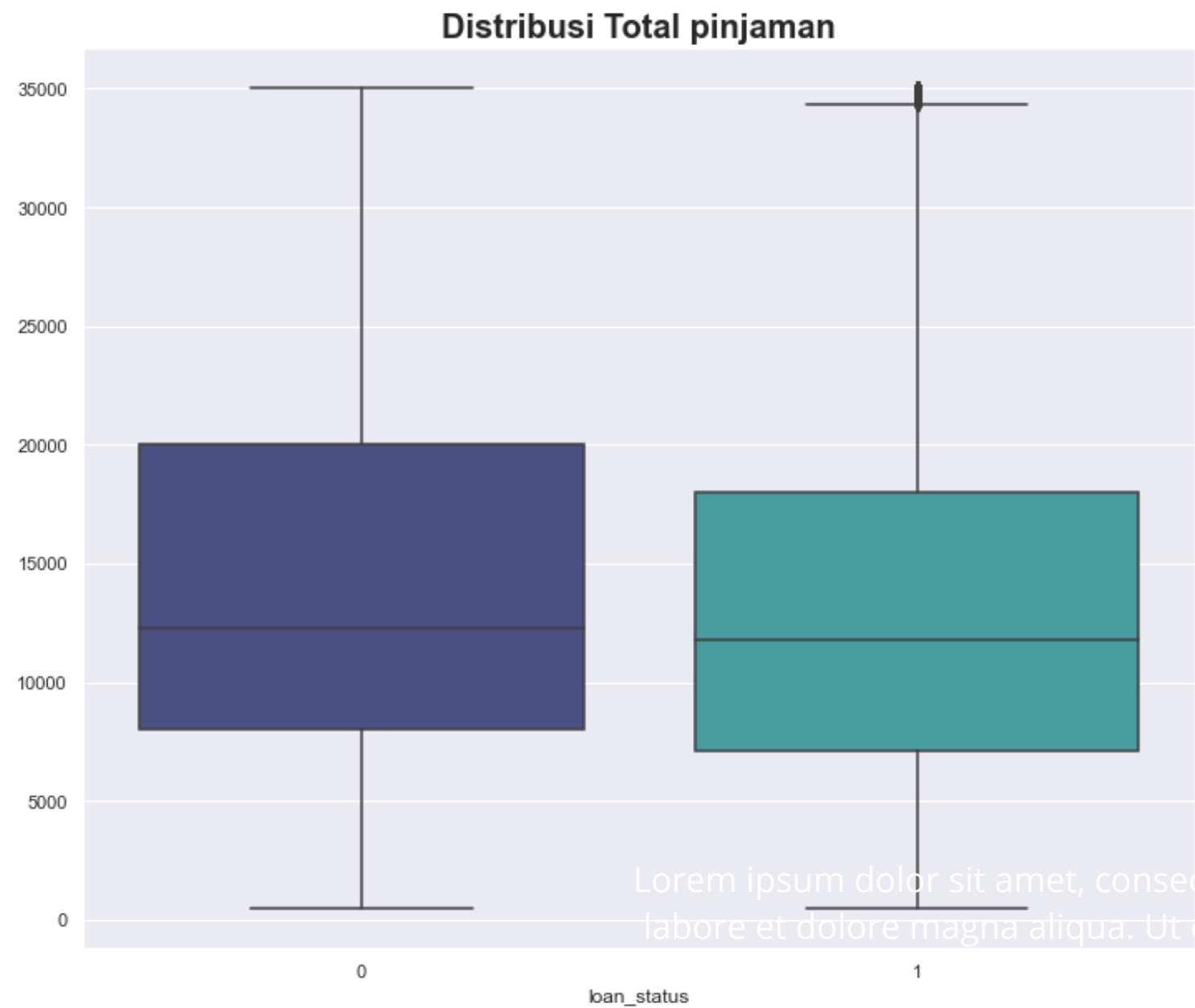




Tujuan pinjaman didominasi untuk **menutup pinjaman sebelumnya**, maka dapat dikatakan kebanyakan peminjam melakukan sistem "**gali lubang tutup lubang**".

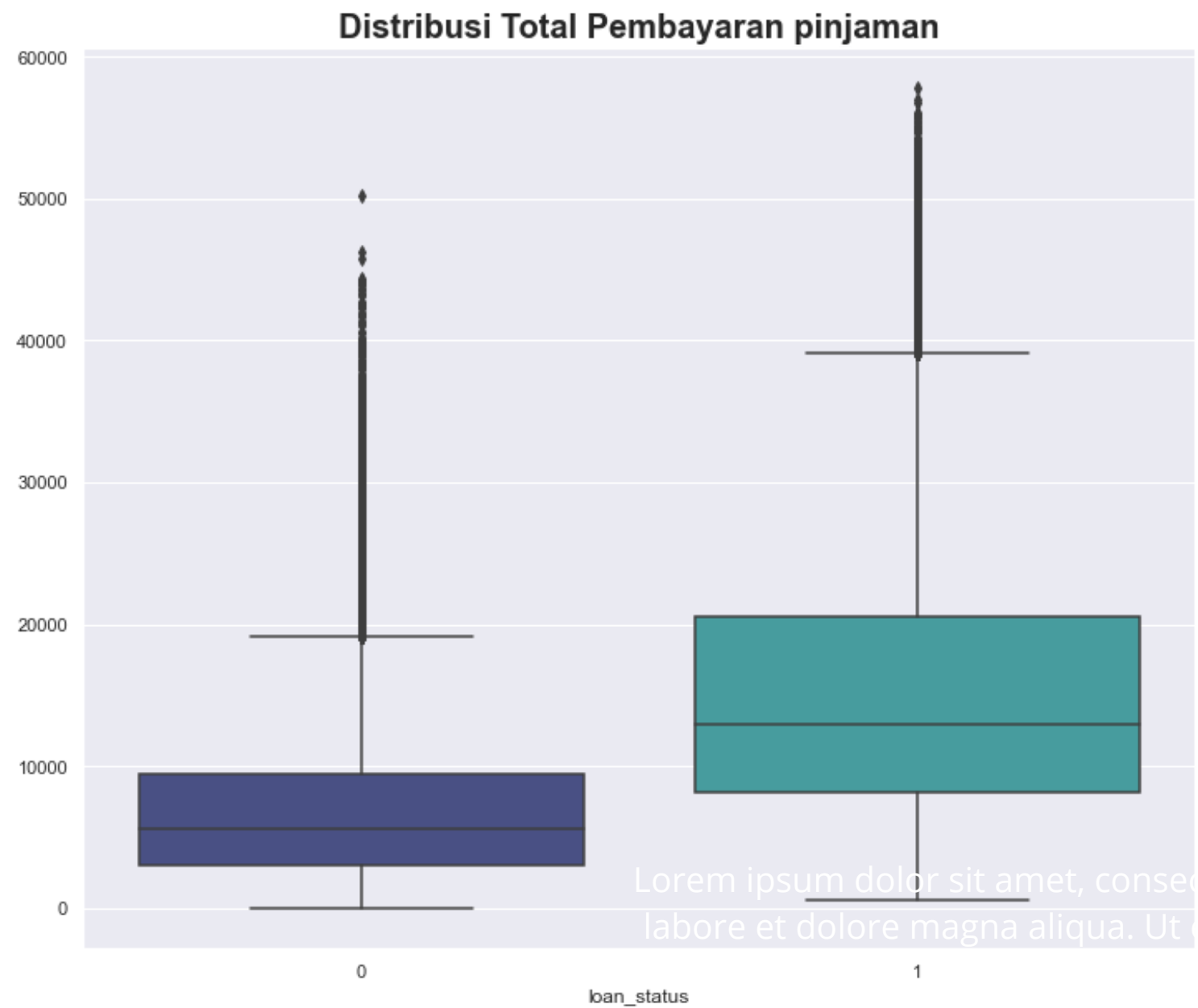
Dapat dilihat pula **tujuan pinjaman** lebih banyak untuk **kegiatan konsumtif** dari setiap individu, daripada untuk tujuan seperti berbisnis, dan biaya pendidikan.





Dapat dilihat dari **distribusi Total pinjaman** yang ada, rata-rata **peminjam** dengan **karakteristik yang buruk** memiliki **rata-rata pinjaman** yang **lebih tinggi** dibandingkan dengan **peminjam** dengan **karakteristik baik**.

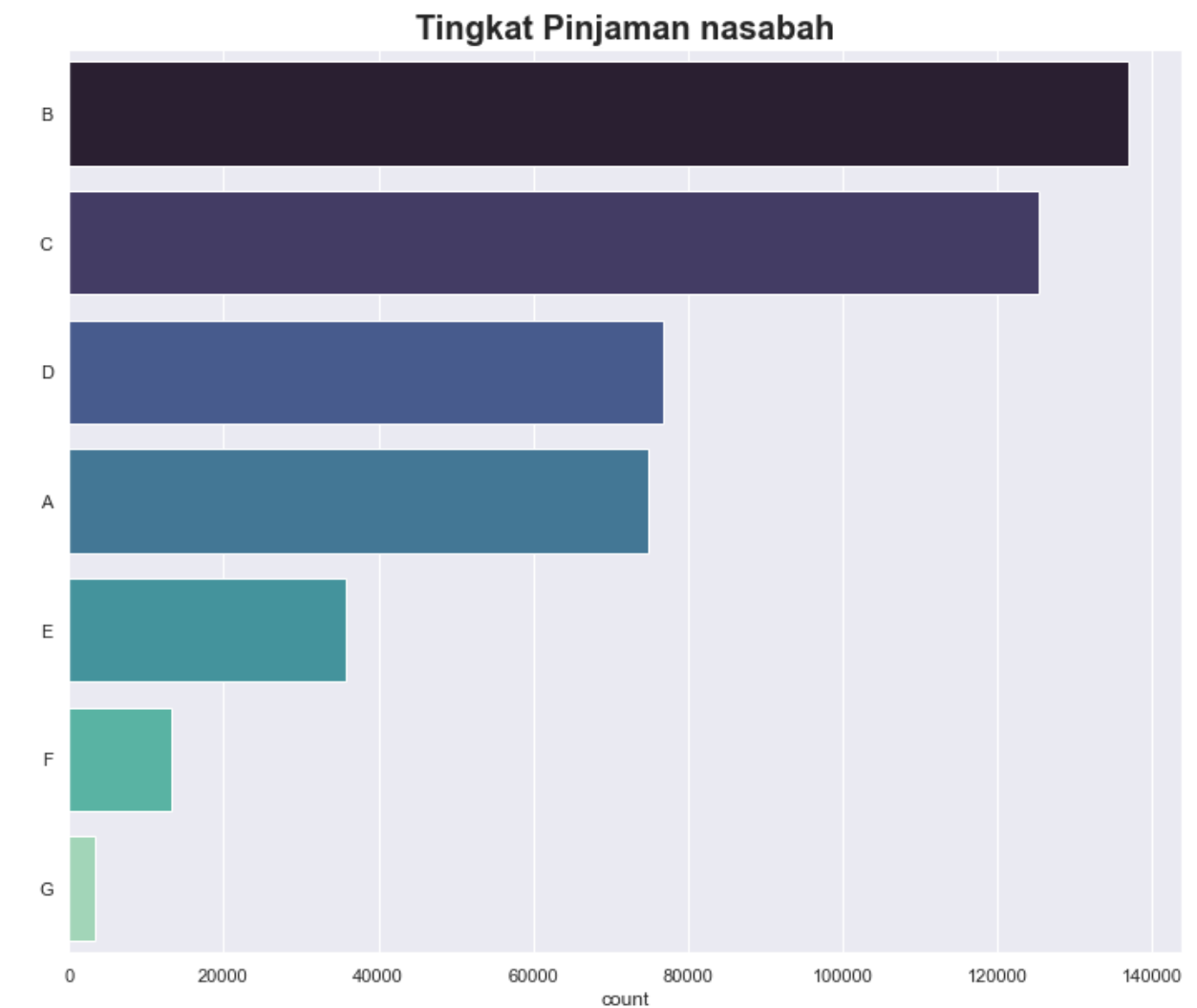
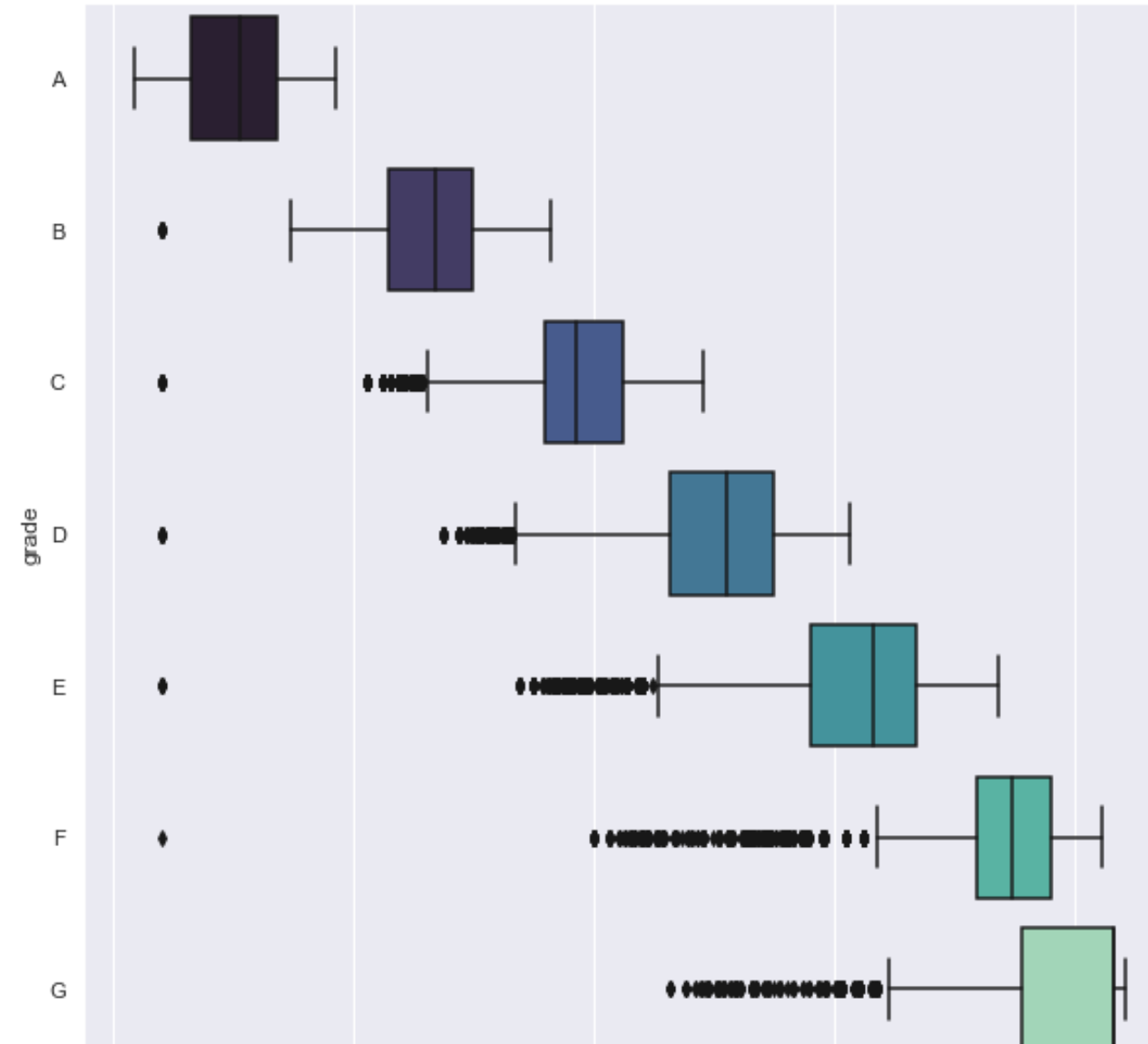




Total Pembayaran Pinjaman serta status pinjaman

Terlihat jelas bahwa individu yang berstatus **good loan** memiliki **jumlah pembayaran terbanyak** pada pinjaman.

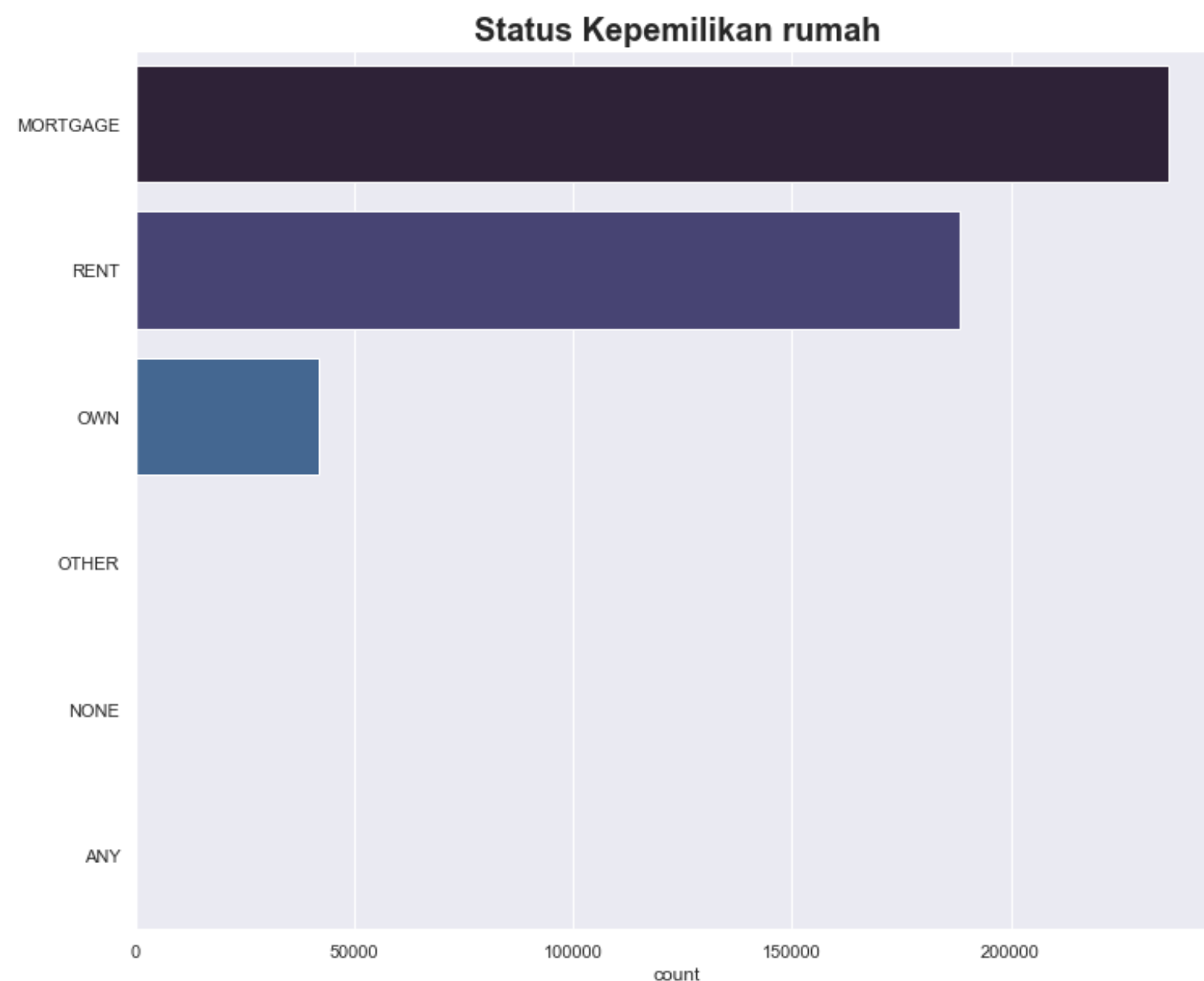




Grade tingkat A,B,C adalah grade dengan **resiko rendah** sedangkan grade D,E,F,G memiliki **resiko tinggi**.

Tingkat pinjaman nasabah masih didominasi di **tingkat B** yang menandakan **bunga pinjaman** yang dibayarkan **kebanyakan nasabah tidak begitu besar** jumlahnya.

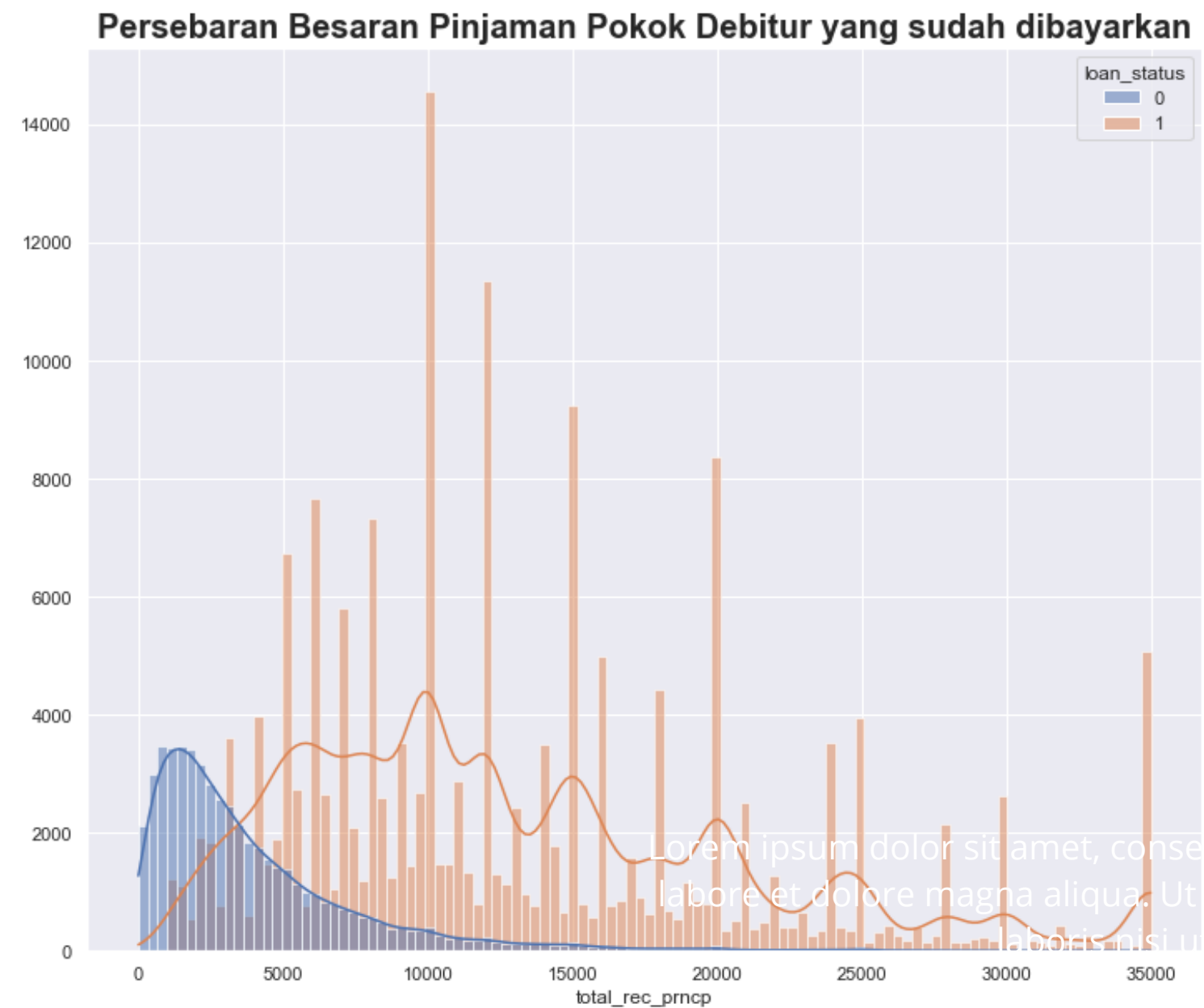




Kebanyakan peminjaam **menjamin rumahnya** sebagai **jaminan pinjaman** jika peminjam **tidak sanggup** untuk **membayarkan seluruh pinjaman** dan bunga pinjaman yang ada.

Padahal **status kepemilikan** rumah **sendiri** dari setiap individu masi tergolong **sedikit**.





Pinjaman Pokok yang Sudah Dibayarkan

Dapat dilihat jika nasabah yang berstatus **bad loan** banyak yang **belum bisa** untuk **membayarkan uang pokok** yang telah **dipinjamkan** sehingga pembayaran mereka sampai **jatuh tempo**.





FINAL PROJECT ID/X PARTNERS – VIRTUAL INTERNSHIP

FEATURE SELECTION AND DATA PREPROCESSING



Langkah **data preprocessing dan feature selection** yang saya lakukan yaitu :

1. **Cleansing data** yang meliputi :

- **Handle Duplicated Data** dengan menghapus data yang berduplikat.
- **Handle Missing Value** dengan mean, modus, dan median sesuai dengan distribusi data setiap variabel
- **Handle Missing Value** menghapus feature yang memiliki missing value kosong $> 30\%$, dan feature yang terlalu banyak memiliki unique value
- **Handle Outlier** menggunakan Z-Score atau IQR

2. **Data Normalization** menggunakan StandardScaler

3. **OneHot Encoding** pada data kategorikal



FINAL PROJECT ID/X PARTNERS – VIRTUAL INTERNSHIP

MACHINE LEARNING (LOGISTIC REGRESSION)





Logistic Regression

Dilakukan **pembagian** menjadi **train Data** dan **test data** dengan **train data** digunakan untuk **melatih model** dan **test data** untuk **menguji model**.

Saya menggunakan **permodelan Logistic Regression** untuk memprediksi **2 label kategori** pada data target.

Dengan bobot **80 : 20** dengan pembagian kategori **good loan** diberikan **bobot sedikit** dan kategori **bad loan** diberikan **bobot yang banyak** karena mengingat perbandingan dari kedua **rasio** tersebut **tidak seimbang**.



Accuracy

96,6%

Train Data

97%

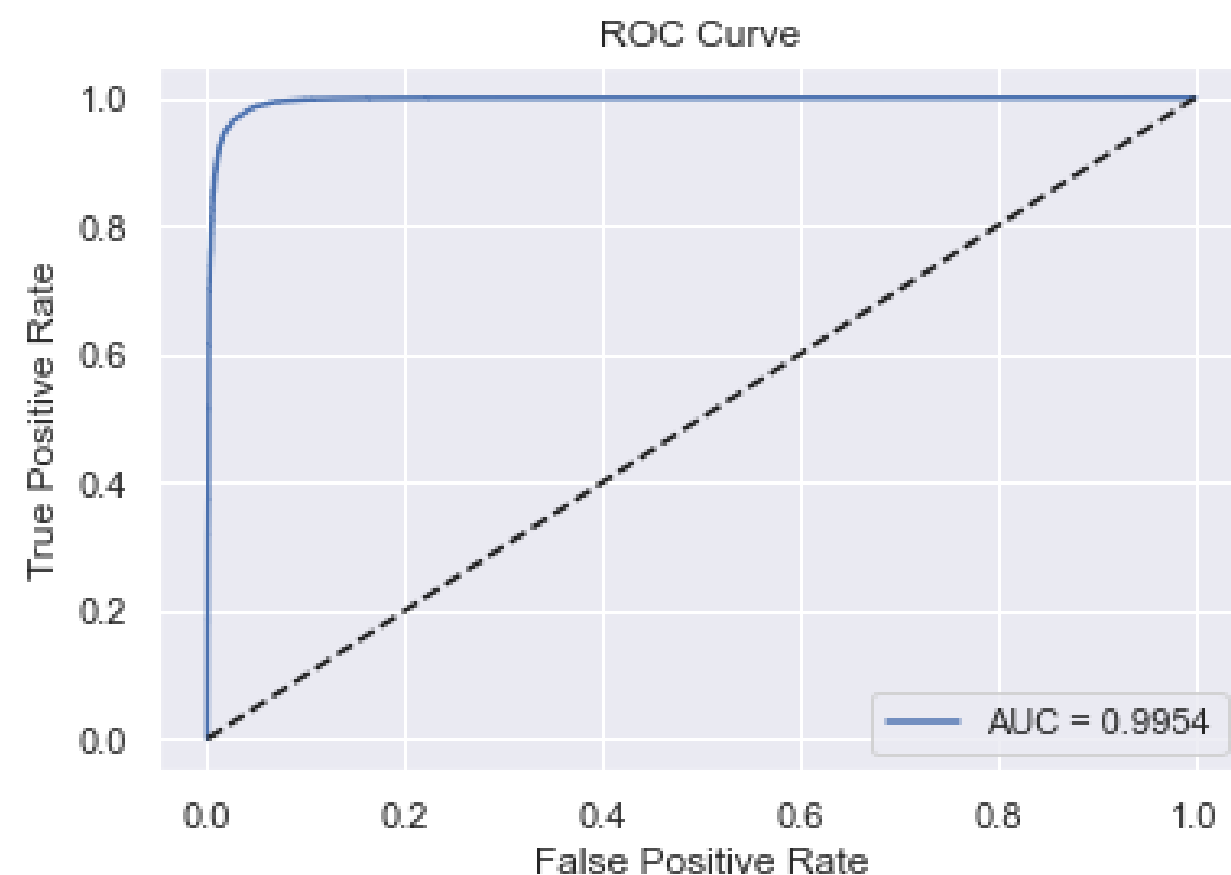
Test Data

Model dapat **memprediksi karakteristik peminjam** dengan tepat dan memiliki **akurasi diatas 90%**



Model Evaluation

	precision	recall	f1-score	support
0	0.89	0.97	0.93	10591
1	0.99	0.97	0.98	36948
accuracy			0.97	47539
macro avg	0.94	0.97	0.95	47539
weighted avg	0.97	0.97	0.97	47539



Terlihat presentase nilai **precision, Recall dan f1-Score** yang cukup **baik** untuk prediksi model yang ada.

Model dapat menghasilkan performa **AUC = 0.99**, umumnya AUC **di atas 0.7** dapat digolongkan kedalam performa yang baik.



KESIMPULAN

Dengan model **Logistic Regression** yang dibangun, kita dapat **memprediksi** karakteristik dari peminjam yang ada, apakah termasuk **peminjam baik (Good loan)** atau **peminjam yang buruk (Bad loan)** dengan tingkat akurasi sebesar 97%.

Kita pun juga dapat menciptakan **kecepatan** dan **efisiensi** dalam mengklasifikasikan karakteristik dari peminjam yang ada, serta dapat mengurangi **risiko** yang harus ditanggung oleh sebuah **lembaga pemberi pinjaman**.



FINAL PROJECT ID/X PARTNERS – VIRTUAL INTERNSHIP

Thank You