



**id/x** partners

# DATA SCIENCE: LOAN DATASET

DIMAS ARI LUMINTANG

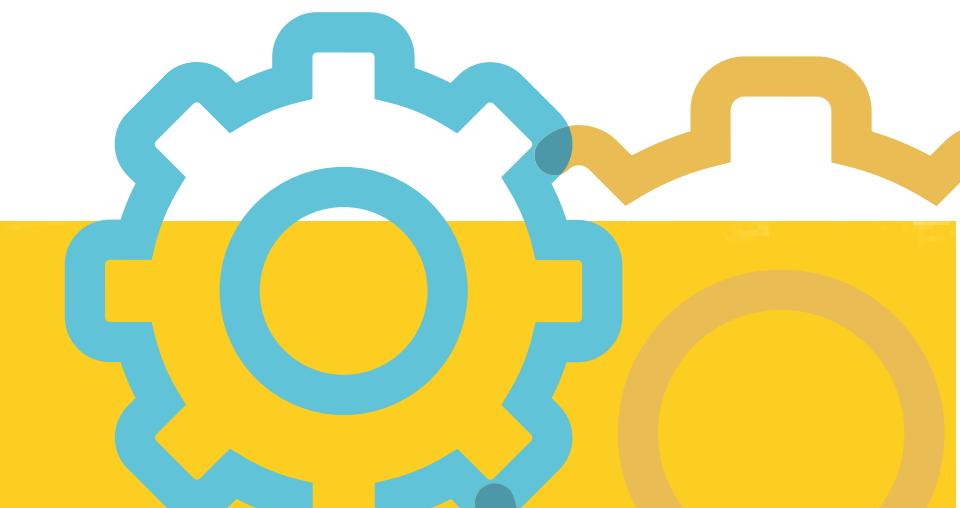
Project-based Intern: Data Scientist Virtual Internship  
Experience by id/x partner dan rakamin academy

## TABLE OF CONTENTS

- Business Plan Summary
- Analyze Point
- Data collection methodology
- Data Preprocessing
- Insight

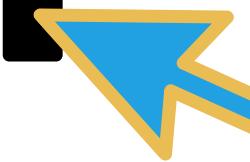
# Business Plan Summary

Lending merupakan kegiatan penyaluran atau meminjamkan dana yang telah dikumpulkan kepada peminjam individu maupun instansi. Jadi perusahaan pemberi pinjaman adalah objek yang menjadi jembatan antara pemberi pinjaman atau pemilik dana (borrower) dan peminjam dana (lender). Permasalahan dalam proyek ini adalah bagaimana cara untuk membangun model yang dapat memprediksi credit risk menggunakan dataset yang disediakan oleh company yang terdiri dari data pinjaman yang diterima dan yang ditolak. Selain itu kamu juga perlu mempersiapkan media visual untuk mempresentasikan solusi ke klien.





# ANALYZE POINT



1. Bagaimana status pinjaman yang ada di perusahaan sekarang?
  2. Variabel apa yang memiliki korelasi paling tinggi dengan status pinjaman?
  3. Berdasarkan dataset perusahaan, apa saja tujuan peminjaman yang dilakukan oleh nasabah?
  4. Seperti apa status kepemilikan rumah dari nasabah, karena kebanyakan nasabah menggunakannya sebagai jaminan?
  5. Bagaimana korelasi antara tanggal dan status pinjaman?
- 

# **DATA COLLECTION METHODOLOGY**

---

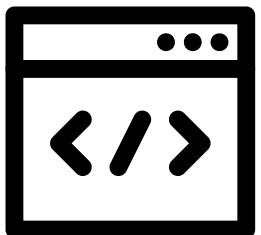
Data yang digunakan disini adalah dataset yang telah diberikan dalam kegiatan ID/X Partners Data Scientist Virtual Internship untuk tugas akhir, yaitu: Loan Dataset ([loan\\_data\\_2007\\_2014.csv](#))





# DATA PREPROCESSING





# • Read Data

## LOAD DATASET

```
import pandas as pd

[7] #read data
data = pd.read_csv('./data/loan_data_2007_2014.csv')
df = pd.DataFrame(data)

#read data description
data_desc = pd.read_excel("./data/LCDataDictionary.xlsx", sheet_name="browseNotes")
```

```
#show data description
data_desc[["Variable","Description"]]
```

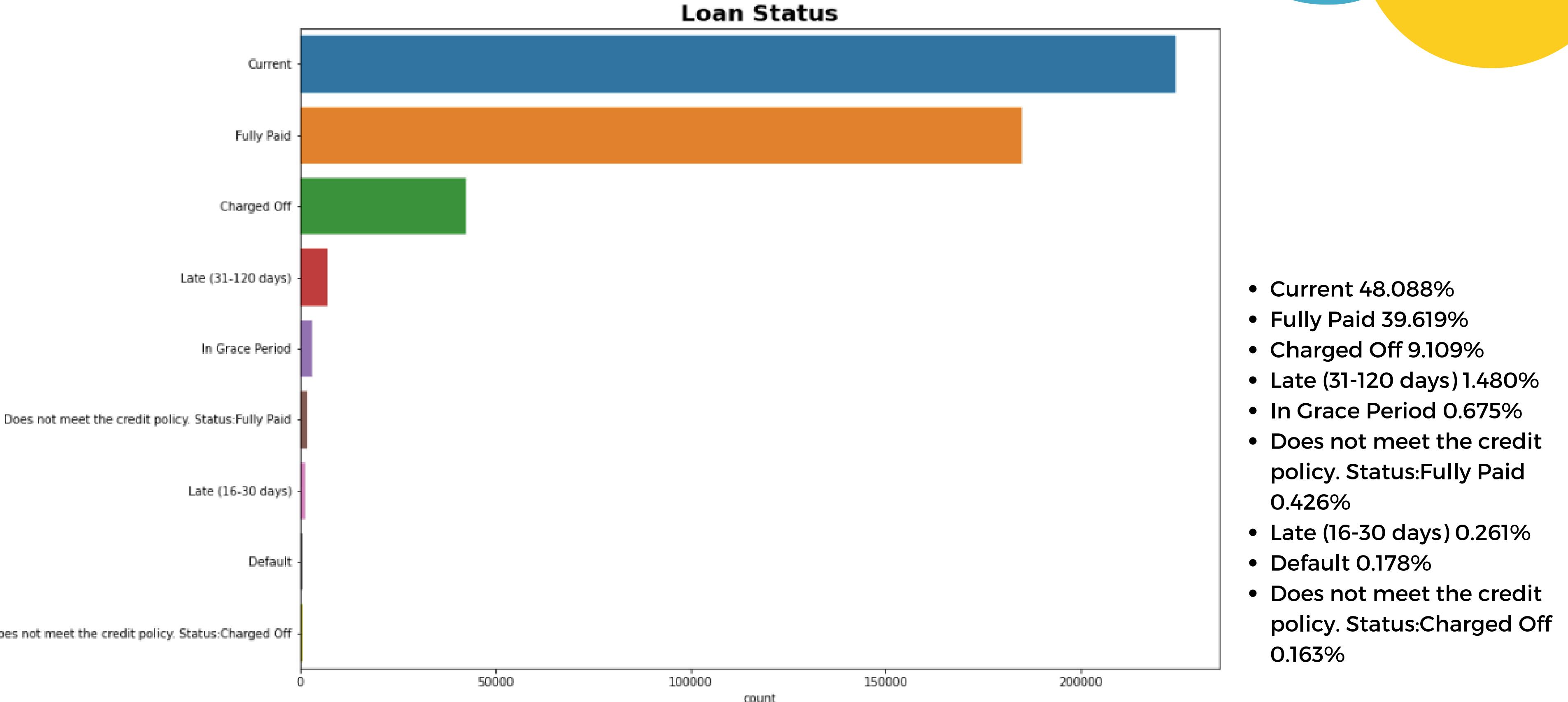
	Variable	Description
0	acceptD	The date which the borrower accepted the offer
1	accNowDelinq	The number of accounts on which the borrower i...
2	accOpenPast24Mths	Number of trades opened in past 24 months.
3	addrState	The state provided by the borrower in the loan...
4	annual_inc_joint	The combined self-reported annual income provi...
...	...	...
103	inq_hi	Number of personal finance inquiries
104	total_cu_t	Number of finance trades
105	inq_last_12m	Number of credit inquiries in past 12 months
106	Nan	Nan
107	Nan	* Employer Title replaces Employer Name for al...

108 rows × 2 columns

```
[13] df.info()
```

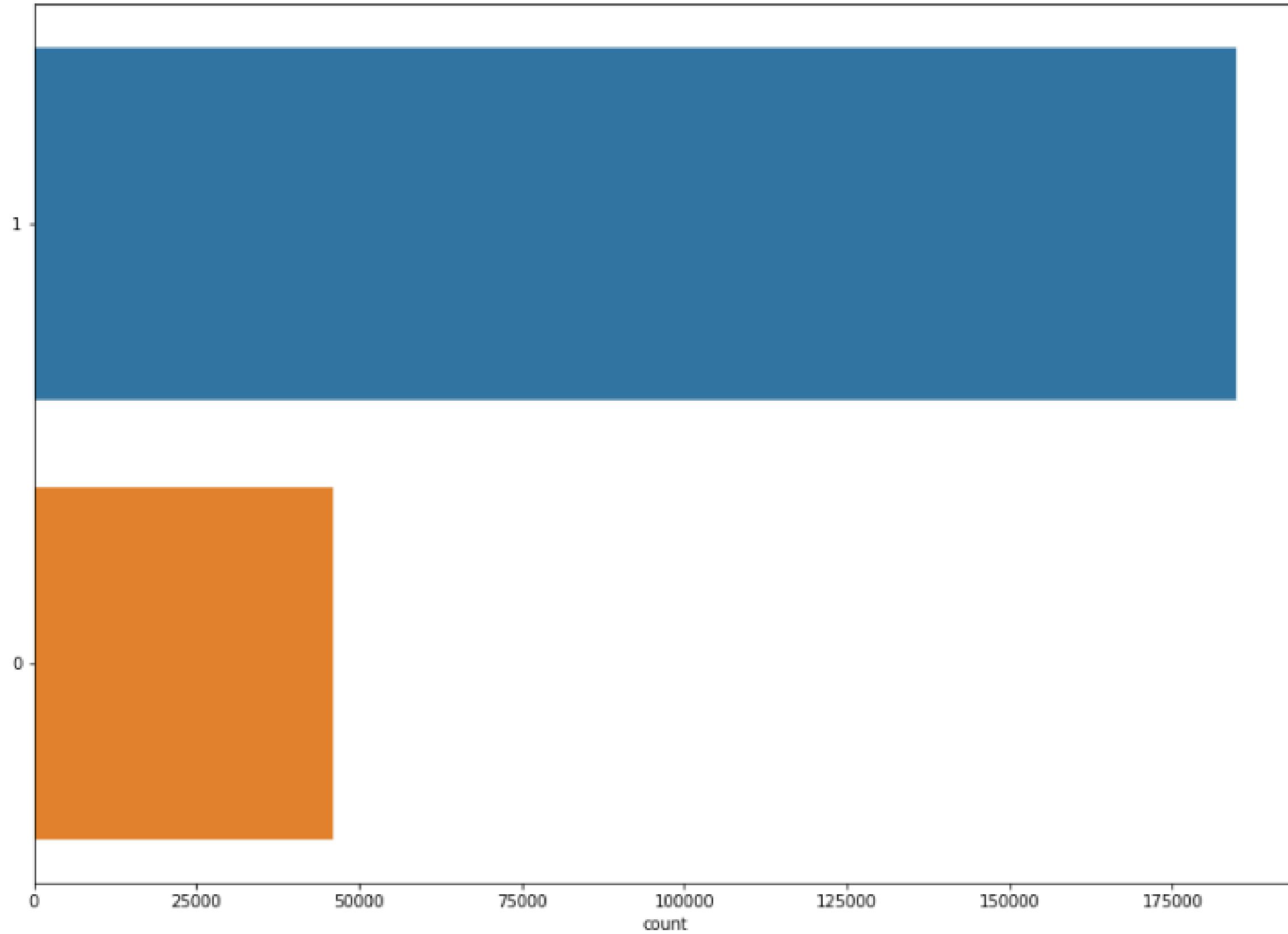
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 466285 entries, 0 to 466284
Data columns (total 75 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Unnamed: 0        466285 non-null  int64  
 1   id               466285 non-null  int64  
 2   member_id        466285 non-null  int64  
 3   ...
```

- Bagaimana status pinjaman yang ada di perusahaan sekarang?





**Loan Status**



- 1 -> 80.045%
- 0 -> 19.955%
- 1 adalah pinjaman yang disetujui
- 0 adalah pinjaman yang ditolak

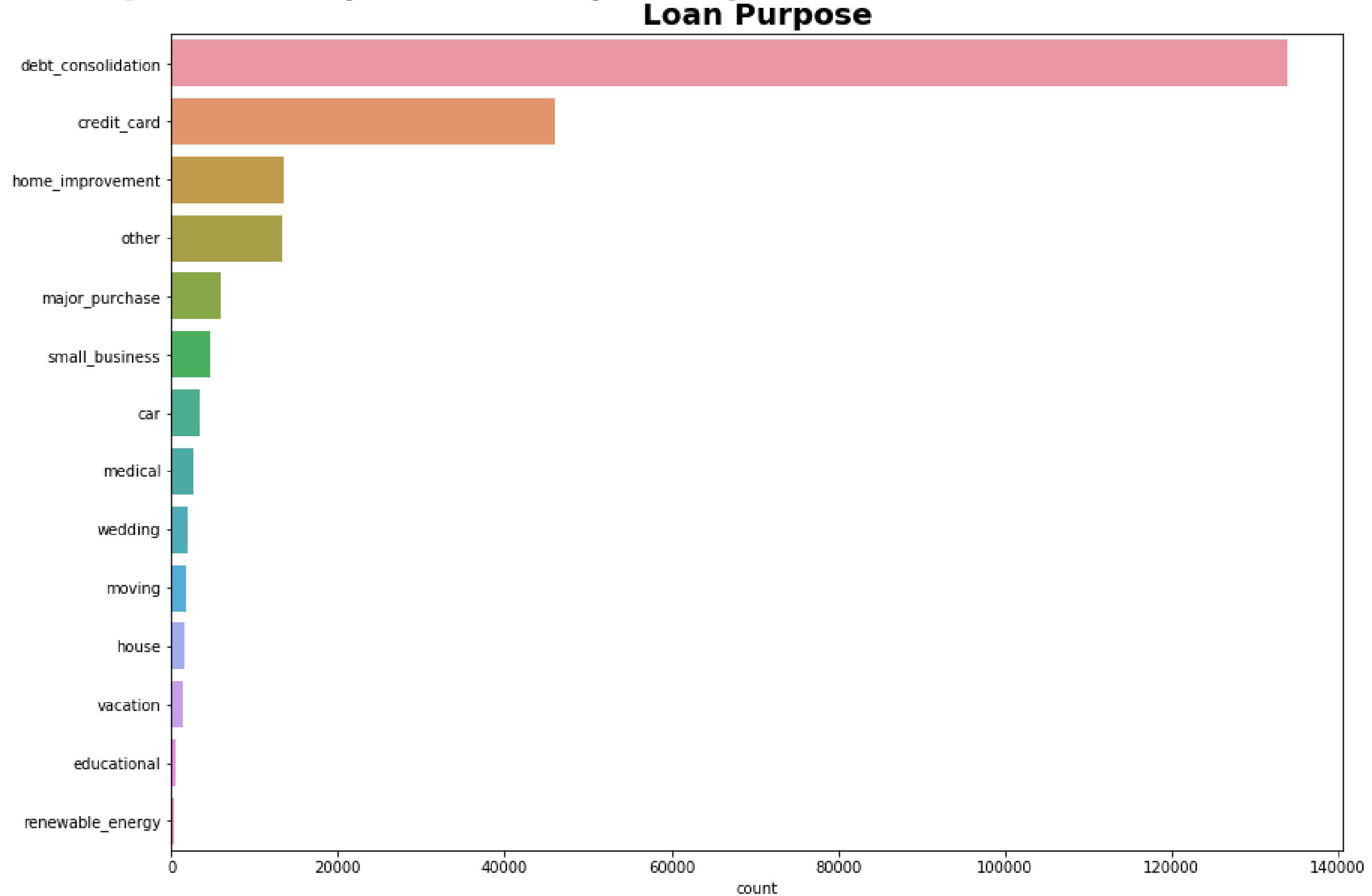
- Variabel apa yang memiliki korelasi paling tinggi dengan status pinjaman?



Dari heatmap diatas terdapat beberapa variabel yang memiliki pengaruh terhadap status pinjaman, diantaranya:

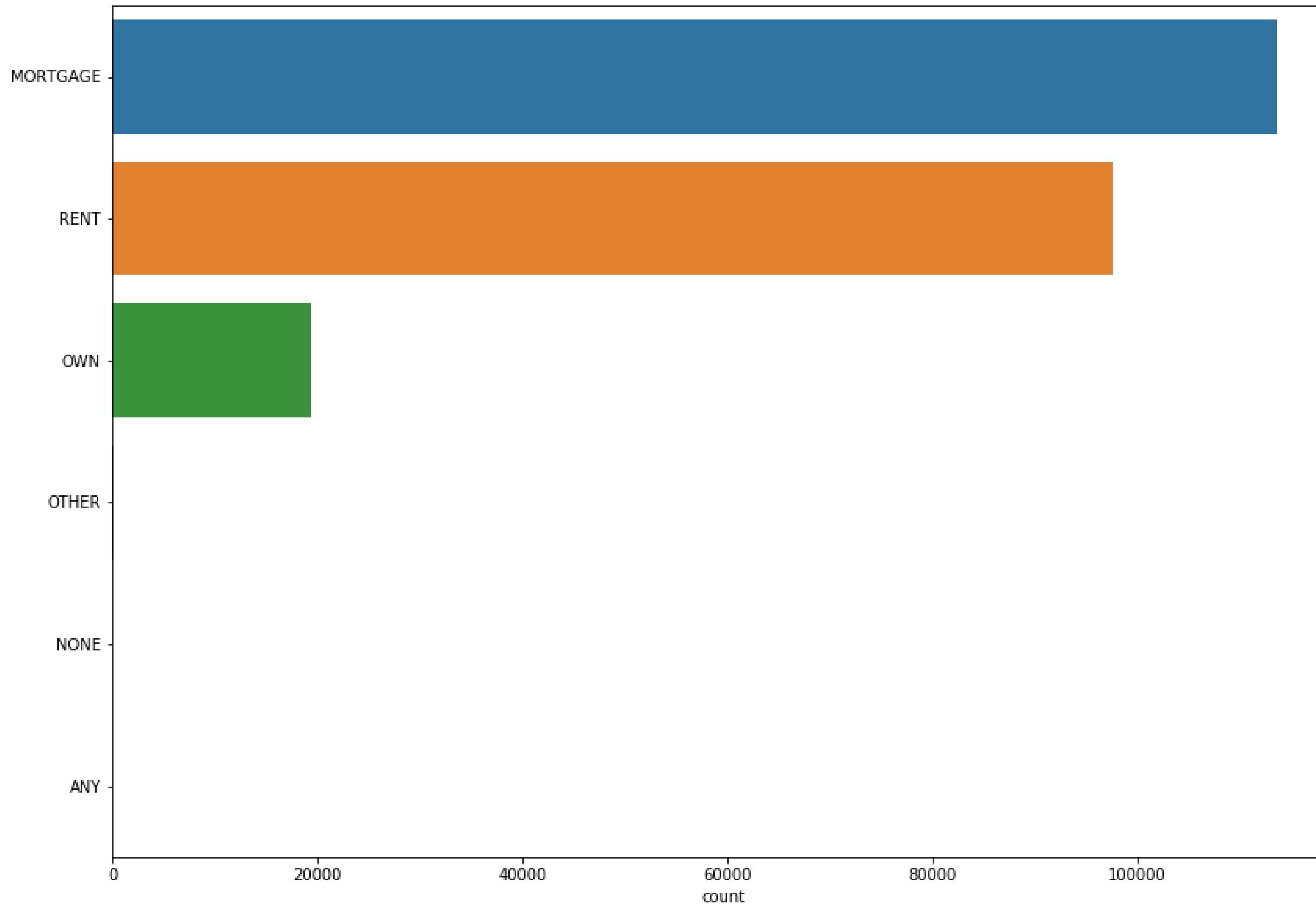
- **dti 0.107257**
- **inq\_last\_6mths 0.116393**
- **total\_rec\_late\_fee 0.152231**
- **int\_rate 0.241059**
- **collection\_recovery\_fee 0.300094**
- **total\_pymnt 0.355349**
- **total\_pymnt\_inv 0.358588**
- **last\_pymnt\_amnt 0.384781**
- **recoveries 0.445422**
- **total\_rec\_prncp 0.460338**

- Berdasarkan dataset perusahaan, apa saja tujuan peminjaman yang dilakukan oleh nasabah?



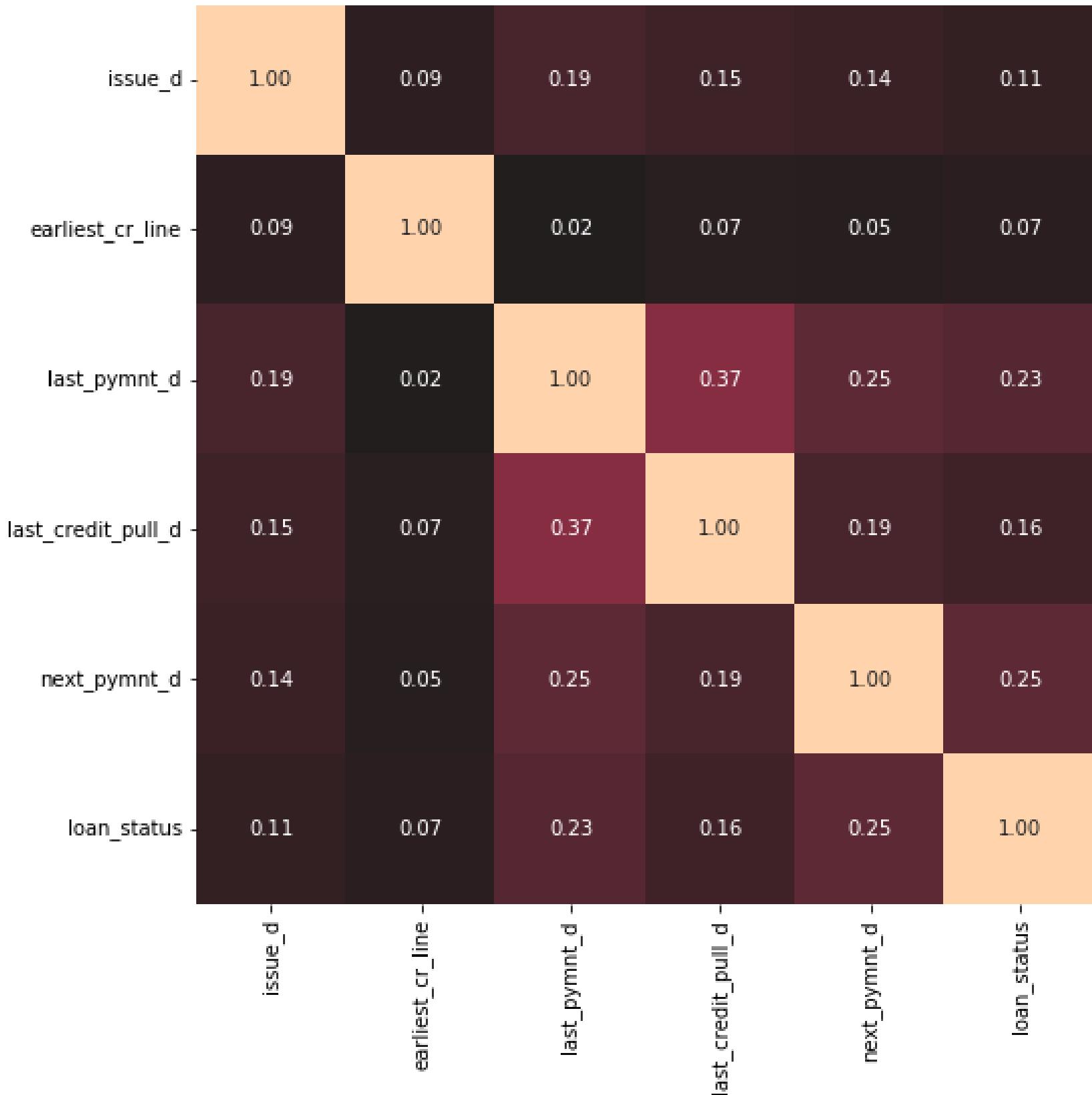
- debt\_consolidation 58.026%
- credit\_card 19.953%
- home\_improvement 5.826%
- other 5.764%
- major\_purchase 2.536%
- small\_business 2.019%
- car 1.499%
- medical 1.145%
- wedding 0.871%
- moving 0.801%
- house 0.649%
- vacation 0.618%
- educational 0.183%
- renewable\_energy 0.110%

- Seperti apa status kepemilikan rumah dari nasabah, karena kebanyakan nasabah menggunakannya sebagai jaminan?



- MORTGAGE 49.233%
- RENT 42.291%
- OWN 8.377%
- OTHER 0.078%
- NONE 0.021%
- ANY 0.000%

- Bagaimana korelasi antara tanggal dan status pinjaman?



Korelasi yang tertinggi dari tanggal dan status pinjaman adalah **issue\_d**, **last\_pymnt\_d**, **last\_credit\_pull\_d**, **next\_pymnt\_d** dan yang paling rendah korelasinya adalah **earliest\_cr\_line**

# INSIGHT



Dari data dan diagaram yang telah didapatkan bahwa saat ini masih banyak status pinjaman yang masih berjalan dan ada beberapa orang dengan status pembayaran yang buruk, namun untuk pinjaman yang disetujui masih sebesar 80.045% dan yang ditolak sebesar 19.955%.

Variabel yang paling berhubungan dan berpengaruh dengan status pinjaman adalah dti, inq\_last\_6mths, total\_rec\_late\_fee, int\_rate, collection\_recovery\_fee, total\_pymnt, total\_pymnt\_inv, last\_pymnt\_amnt, recoveries, dan total\_rec\_prncp.

Alasan peminjaman yang dilakukan nasabah kebanyakan digunakan untuk konsolidasi hutang.



# THANK YOU

Open to discuss and Feedback, contact me here:



<https://www.linkedin.com/in/arilumintang/>



[dimasarilumintang10.4@gmail.com](mailto:dimasarilumintang10.4@gmail.com)