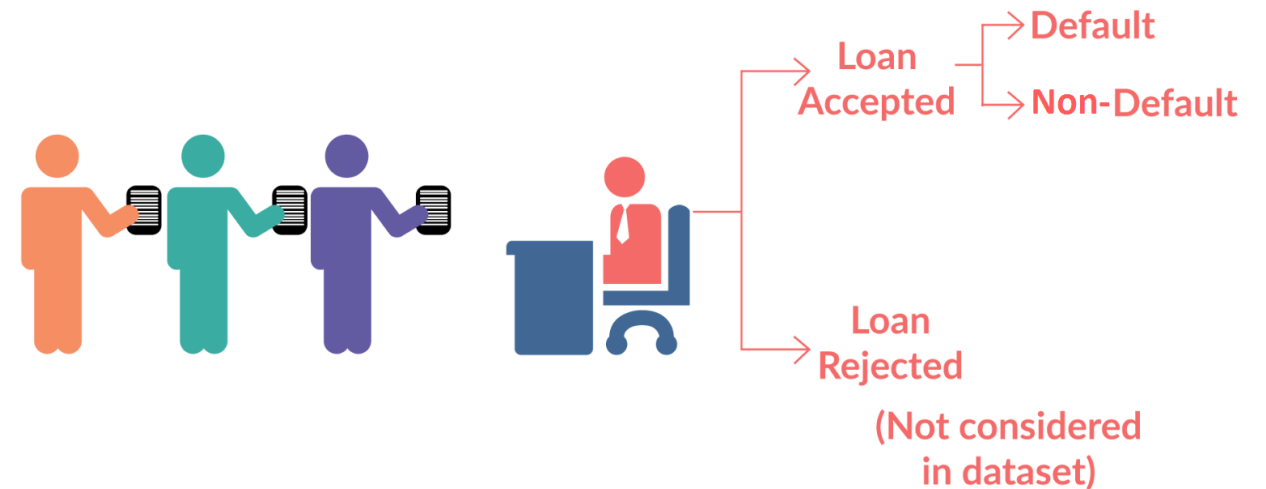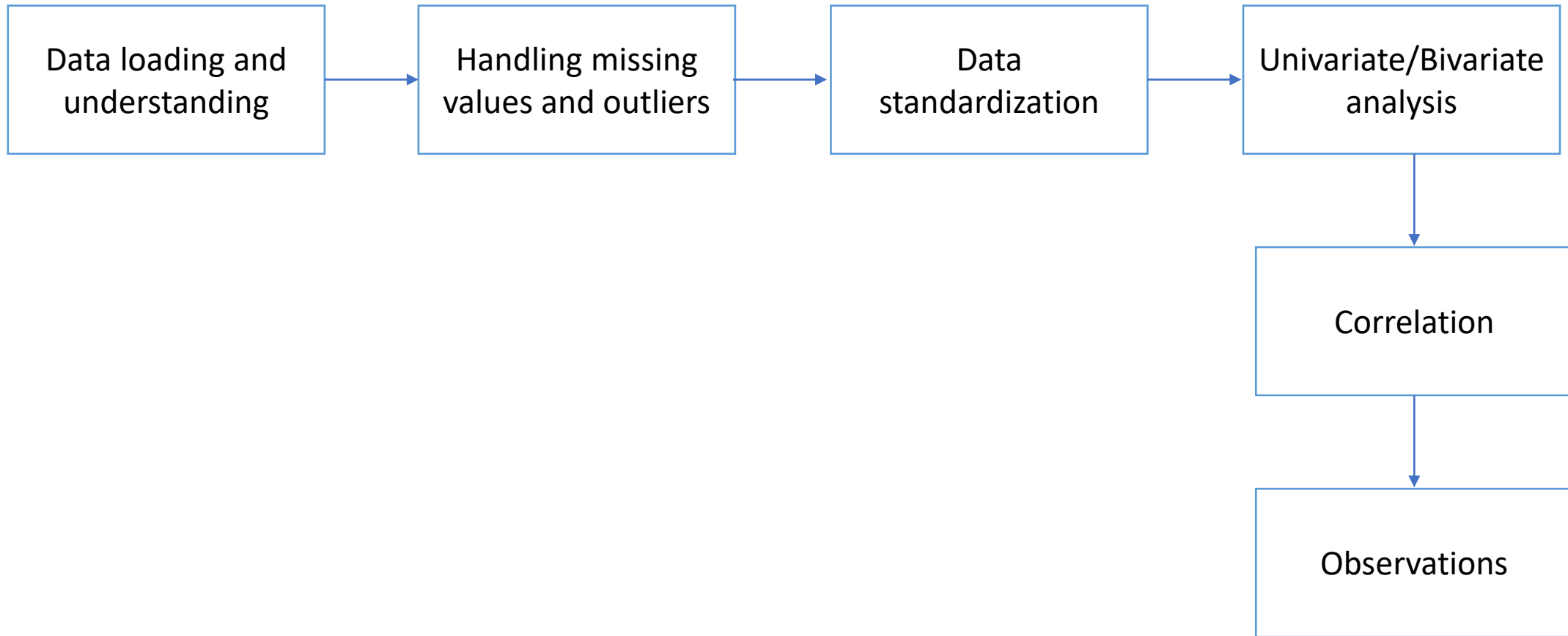# LENDING CLUB CASE STUDY

INDERPREETKAUR ARORA

Business objective

- A lending firm, basis the borrowers' data wants to analyze:
    - Driving factors to accept or reject loan

- These metrics should give the likelihood so that the lending firm:
    - Earns profit by not rejecting the applicants who are most likely to pay the loan
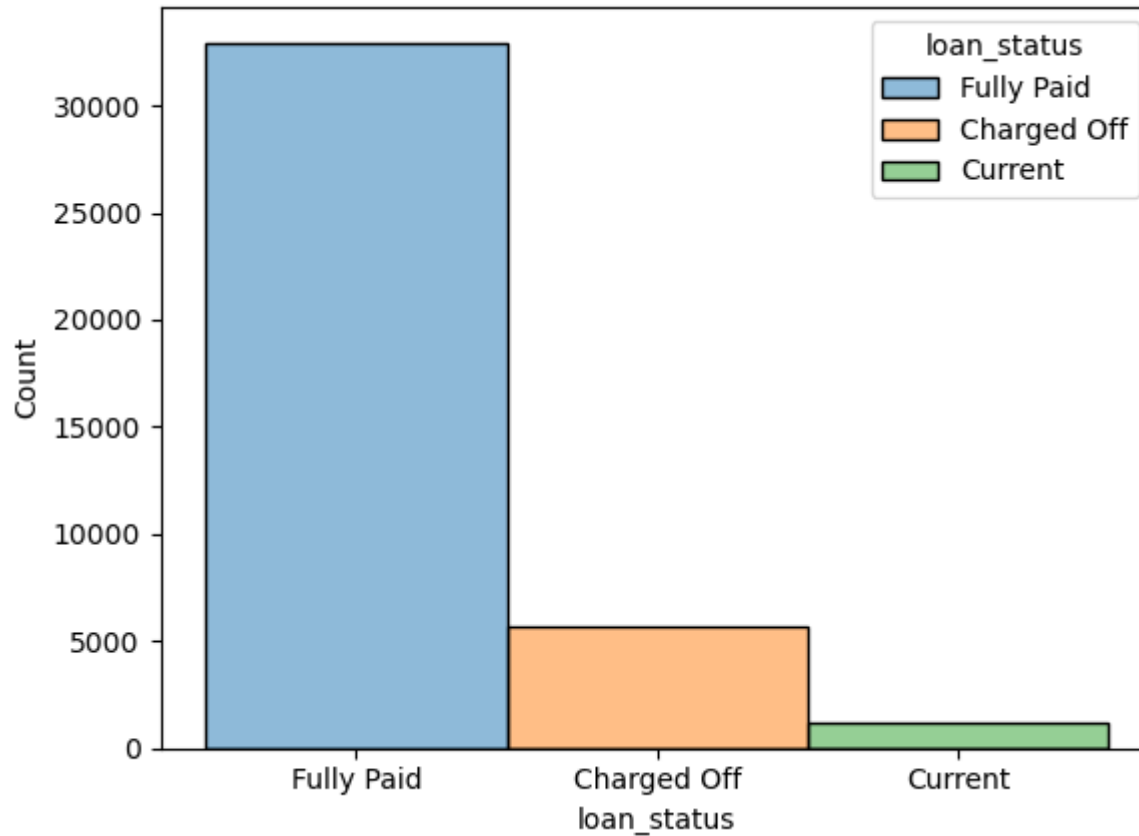   - Reduces losses by rejecting the applicants who are most likely to default

LOAN DATASET

Loan Accepted → Default
Loan Accepted → Non-Default

Loan Rejected
(Not considered in dataset)

Approach

```
┌─────────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
│   Data loading and  │ ───▶ │   Handling missing  │ ───▶ │        Data         │ ───▶ │ Univariate/Bivariate│
│     understanding   │      │  values and outliers│      │   standardization   │      │       analysis      │
└─────────────────────┘      └─────────────────────┘      └─────────────────────┘      └─────────────────────┘
          │                                                                                        │
          ▼                                                                                        ▼
                                                                                        ┌─────────────────────┐
                                                                                        │     Correlation     │
                                                                                        └─────────────────────┘
                                                                                                 │
                                                                                                 ▼
                                                                                        ┌─────────────────────┐
                                                                                        │    Observations     │
                                                                                        └─────────────────────┘
```

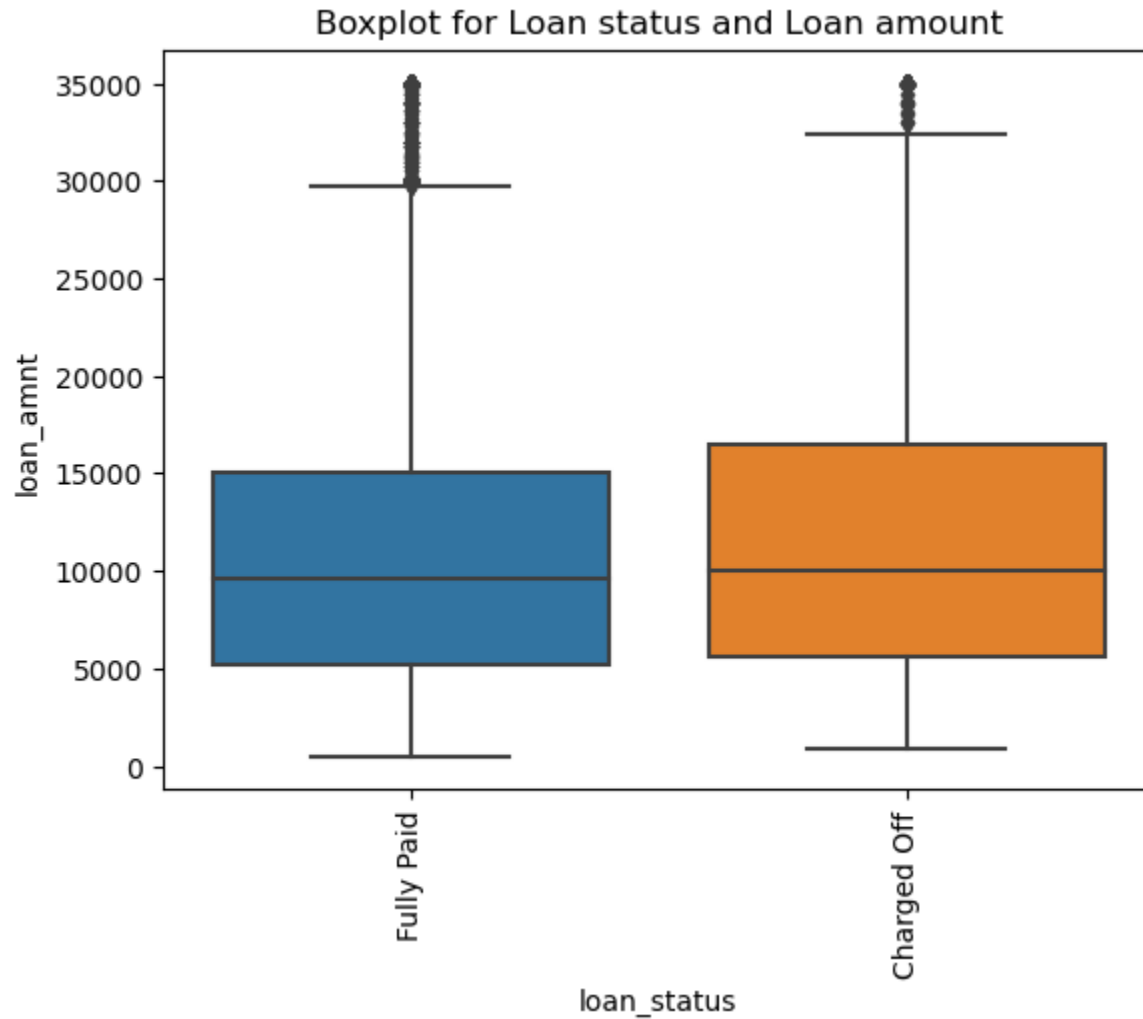# Univariate Analysis

Fully Paid   32950
Charged Off   5627
Current   1140
Name: loan_status
dtype: int64

Observation:
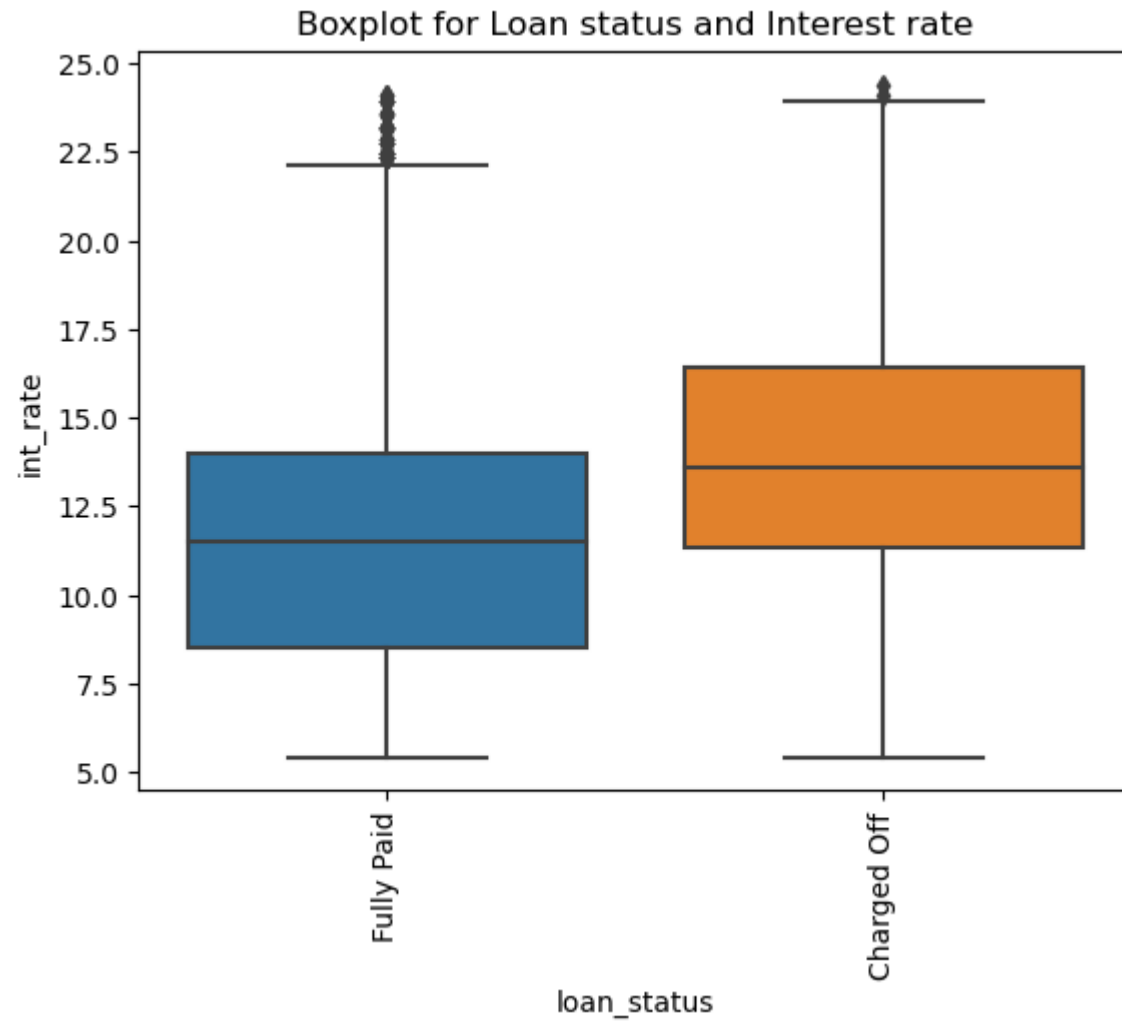- 83% have fully paid
- 14% have defaulted
- 3% are current

Bivariate Analysis

Boxplot for Loan status and Loan amount

Loan amount:
- The defaulted ones have:
  - Higher median
  - Higher min
  - Higher max

Bivariate Analysis

Boxplot for Loan status and Interest rate

Interest rate:
- The charged off ones have:
  - Higher median
  - Higher min
  - Higher max

Bivariate Analysis

Boxplot for Loan status and DTI
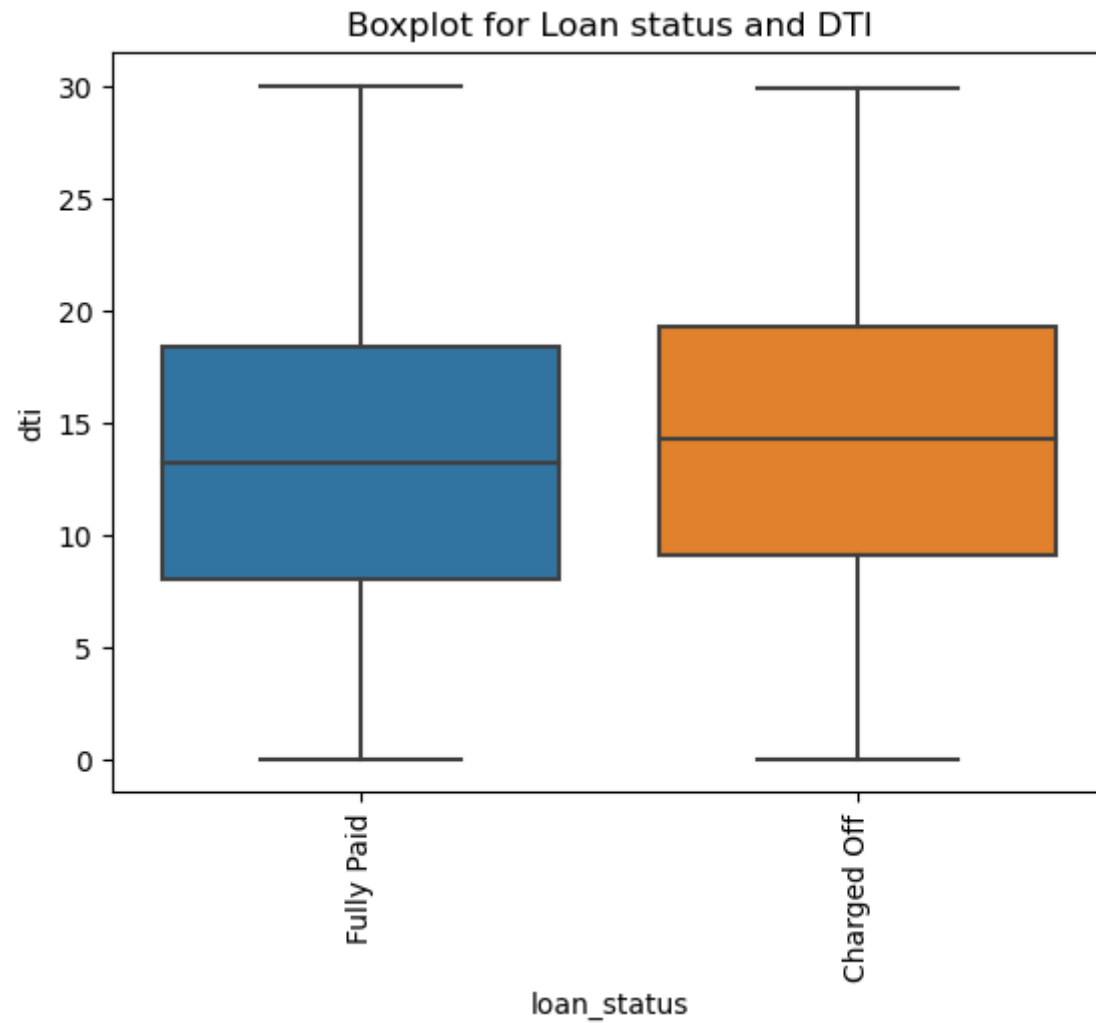
Debt to Income ratio:
- The charged off ones have:
    - Higher median
    - Higher min
    - Higher max

Bivariate Analysis
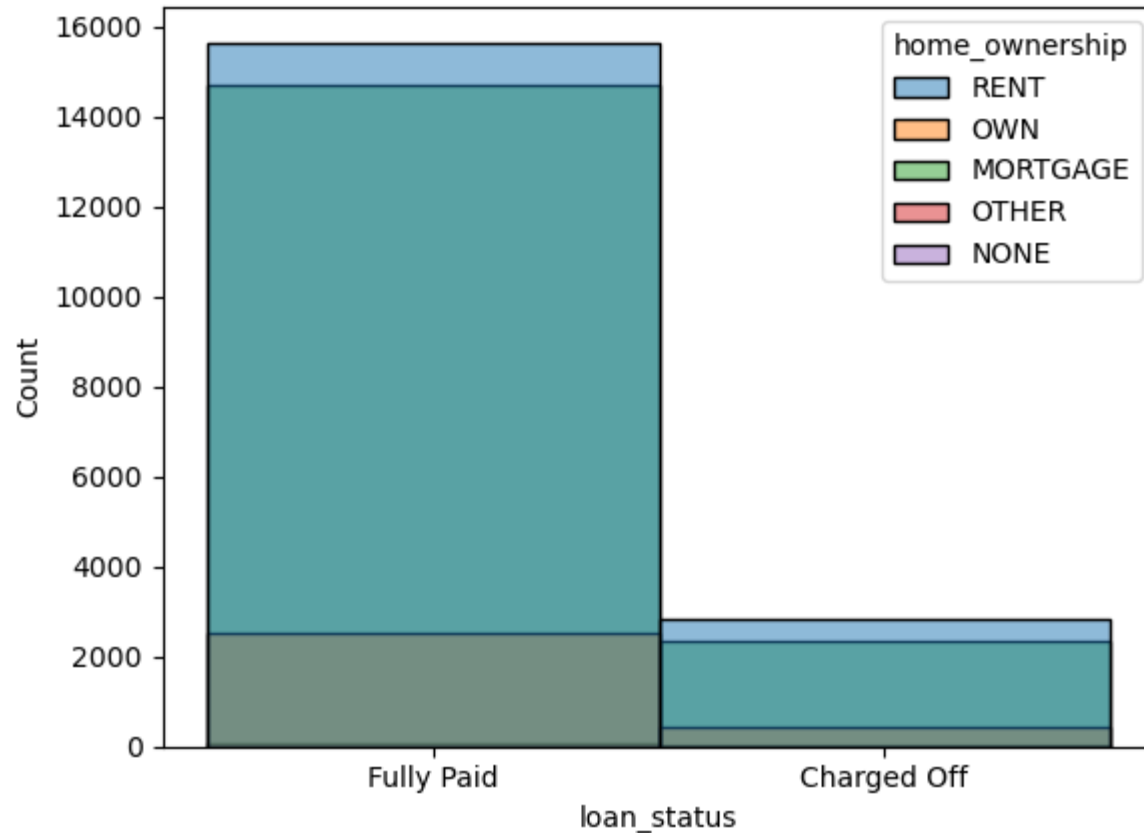
Fully paid
- RENT   15641
- MORTGAGE   14694
- OWN   2532
- OTHER   80
- NONE   3
- Name: home_ownership
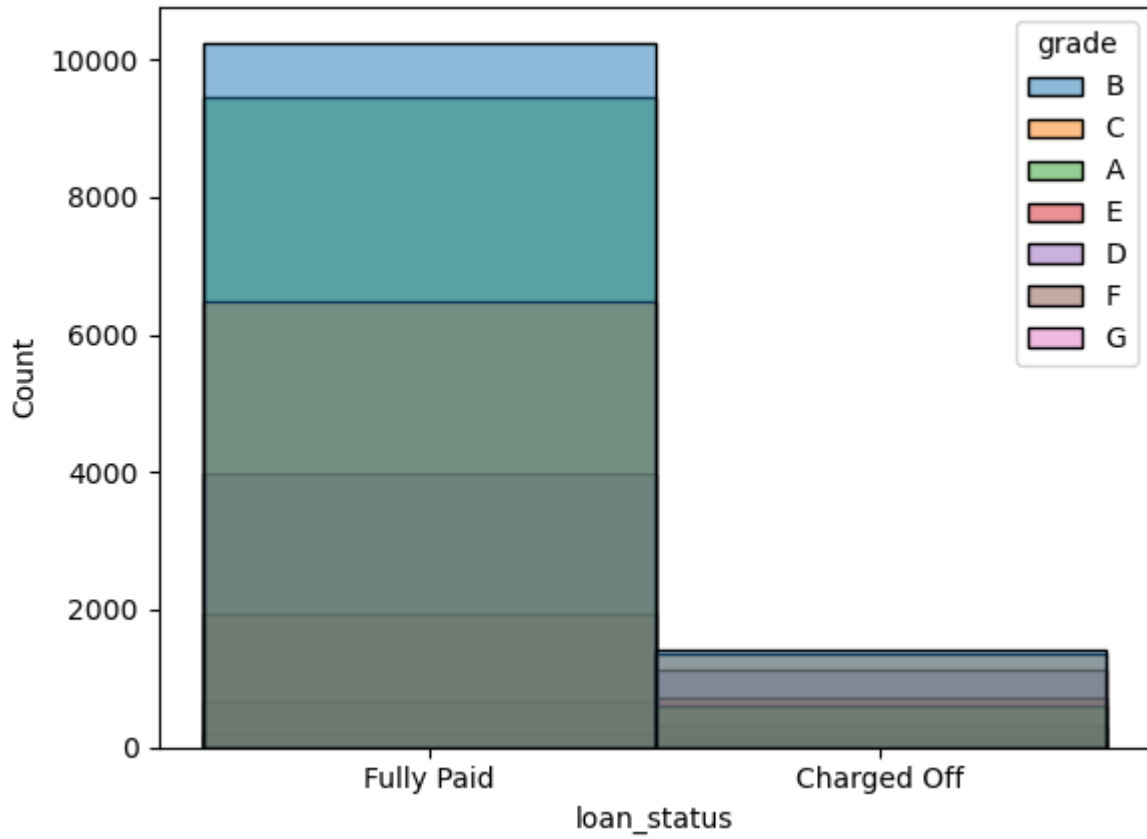- dtype: int64

Charged off
- RENT   2839
- MORTGAGE   2327
- OWN   443
- OTHER   18
- Name: home_ownership
- dtype: int64

Home ownership proportion and trend:
- Do not give any insights basis the loan status

Bivariate Analysis

Fully paid
- B   10250
- A   9443
- C   6487
- D   3967
- E   1948
- F   657
- G   198
- Name: grade
- dtype: int64

Charged off
- B   1425
- C   1347
- D   1118
- E   715
- A   602
- F   319
- G   101
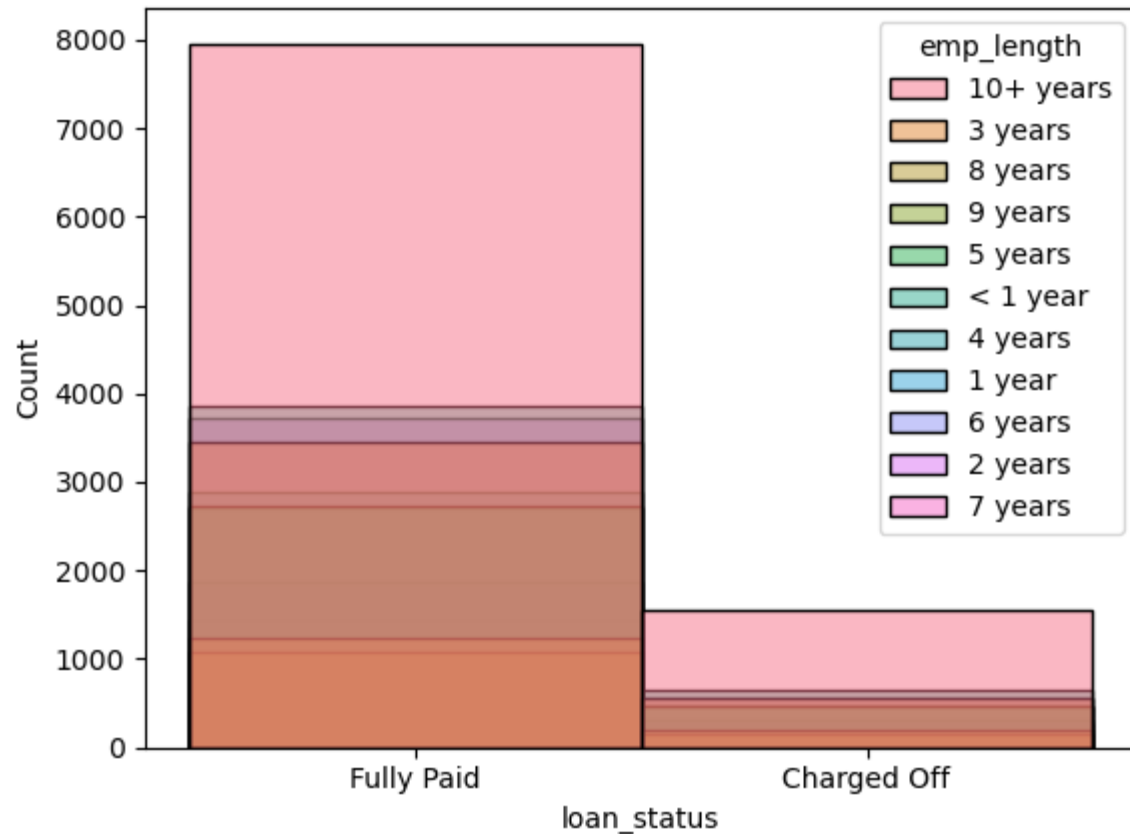- Name: grade
- dtype: int64

Fully paid & Charged off
- B   11675
- A   10045
- C   7834
- D   5085
- E   2663
- F   976
- G   299
- Name: grade
- dtype: int64

Borrower grade proportion and trend:
- Basis the sum of the above, C grade are more likely to default, and A grade are less likely to default

Bivariate Analysis

Charged off
- 10+ years   1559
- < 1 year   639
- 2 years   567
- 3 years   555
- 4 years   462
- 5 years   458
- 1 year   456
- 6 years   307
- 7 years   263
- 8 years   203
- 9 years   158

Fully paid
- 10+ years   7962
- < 1 year   3869
- 2 years   3724
- 3 years   3457
- 4 years   2880
- 5 years   2736
- 1 year   2713
- 6 years   1861
- 7 years   1448
- 8 years   1232
- 9 years   1068

Employee experience proportion and trend:
- Highly experienced and freshers seem to be defaulting more than the other categories
- However, for fully paid dataset, highly experienced are paying off
- So, need to be careful while considering this

Bivariate Analysis
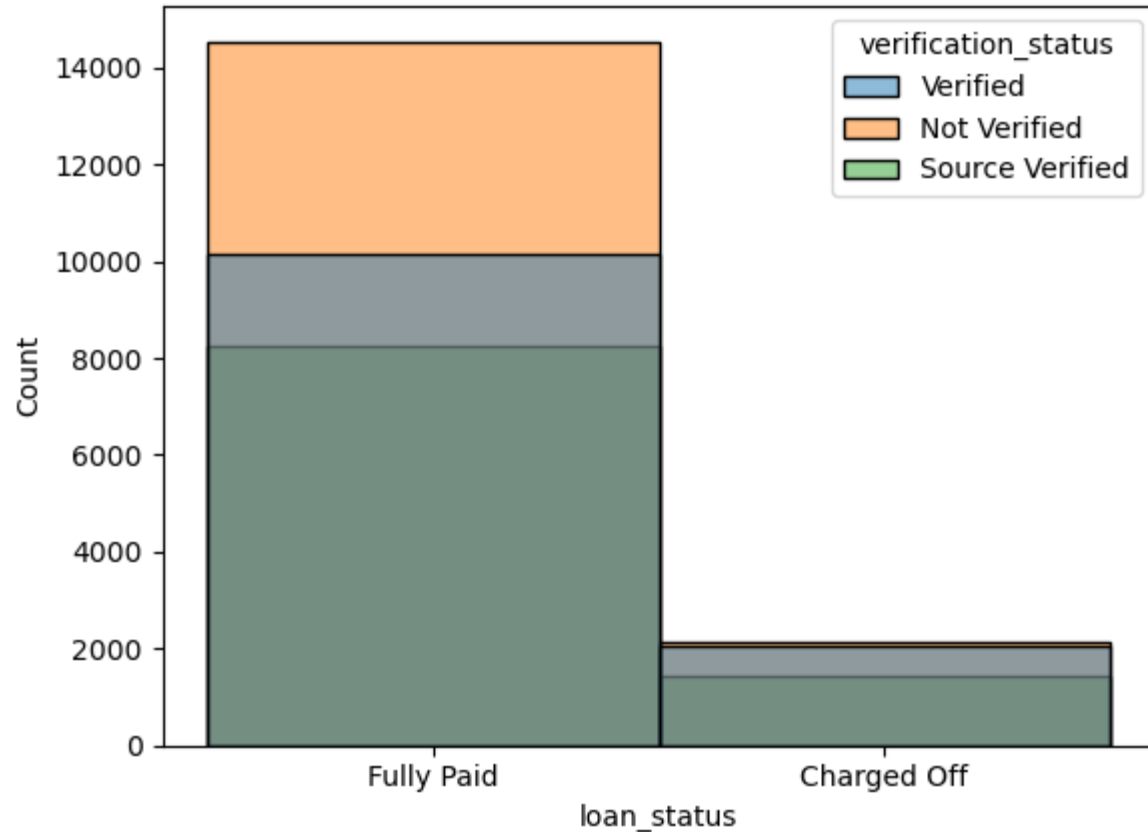
Charged off
- Not Verified   2142
- Verified   2051
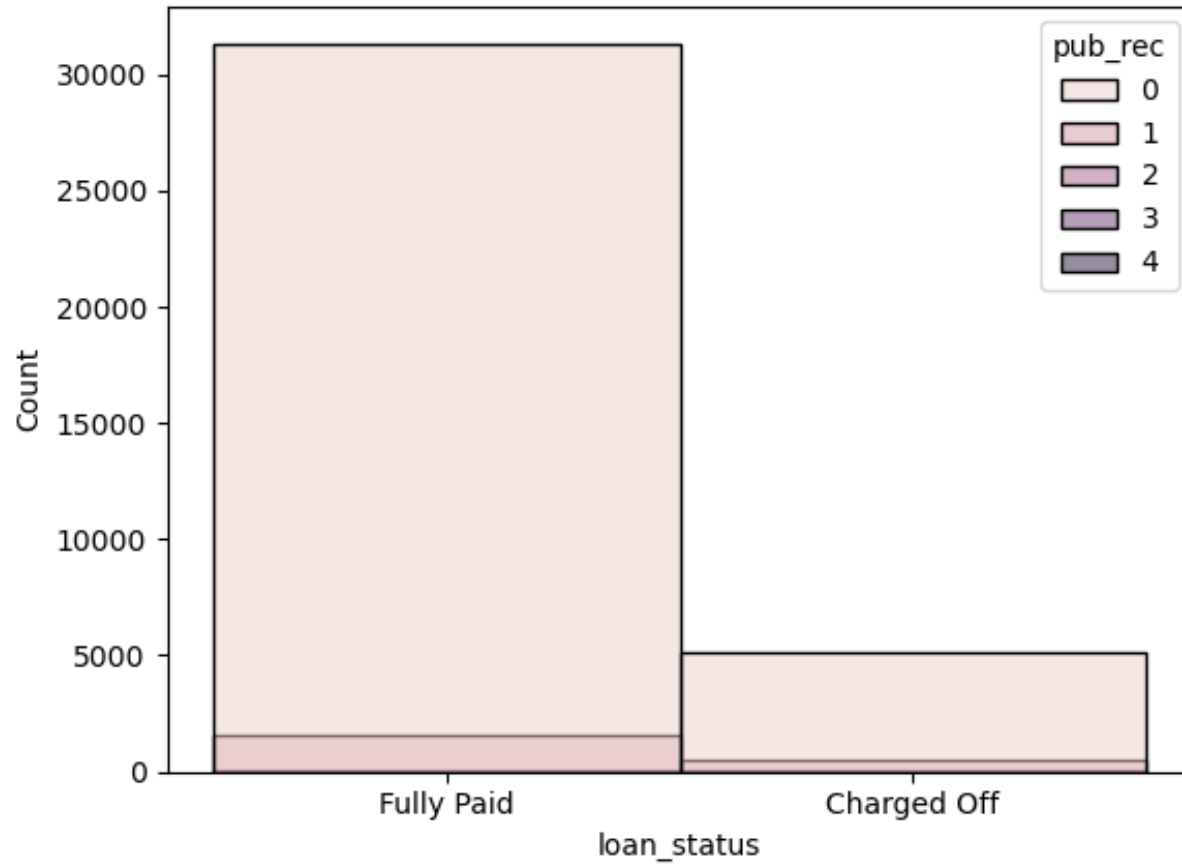- Source Verified   1434

Fully paid
- Not Verified   14552
- Verified   10155
- Source Verified   8243

Verification status proportion and trend:
- Both verified and not verified have are almost similar in count for Charged off data
- For fully paid, even the not verified, in larger proportion have paid
- So, can not draw a concrete conclusion using this

Bivariate Analysis

Charged off
- 0   5160
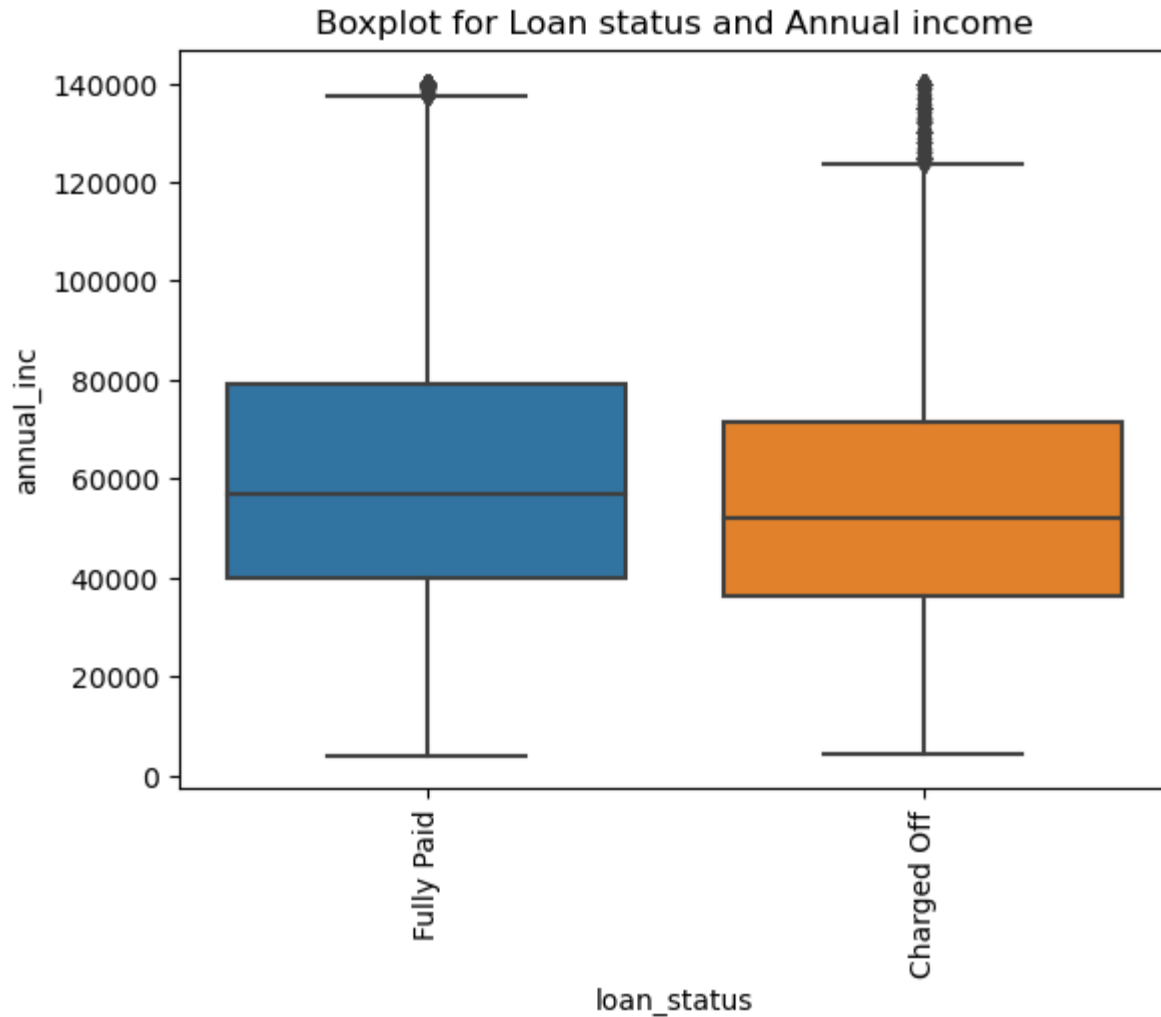- 1   457
- 2   10

Fully paid
- 0   31347
- 1   1556
- 2   38
- 3   7
- 4   2

Derogatory public records proportion and trend:
- Comparing the above 2, no trend observed

Bivariate Analysis

Boxplot for Loan status and Annual income

After the outlier removal beyond 95 percentile,
- Comparing the 2 plots, we observe that the records with less annual income have more probability to default

Charged off dataset Analysis

	loan_amnt	funded_amnt	term	int_rate	installment	annual_inc	dti	inq_last_6mths	open_acc	pub_rec
count	5627.000000	5627.000000	5627.000000	5627.000000	5627.000000	5.627000e+03	5627.000000	5627.000000	5627.000000	5627.000000
mean	12104.385108	11753.398792	46.236360	13.820432	336.175006	6.242730e+04	14.000624	1.058113	9.178070	0.084770
std	8085.732038	7760.516335	11.870746	3.654413	217.051841	4.777601e+04	6.585577	1.154136	4.525937	0.284873
min	900.000000	900.000000	36.000000	5.420000	22.790000	4.080000e+03	0.000000	0.000000	2.000000	0.000000
25%	5600.000000	5575.000000	36.000000	11.310000	168.555000	3.700000e+04	9.050000	0.000000	6.000000	0.000000
50%	10000.000000	10000.000000	36.000000	13.610000	293.870000	5.300000e+04	14.290000	1.000000	8.000000	0.000000
75%	16500.000000	16000.000000	60.000000	16.400000	457.840000	7.500000e+04	19.290000	2.000000	12.000000	0.000000
max	35000.000000	35000.000000	60.000000	24.400000	1305.190000	1.250000e+06	29.850000	8.000000	38.000000	2.000000

Fully paid dataset Analysis

	loan_amnt	funded_amnt	term	int_rate	installment	annual_inc	dti	inq_last_6mths	open_acc	pub_rec
count	32950.000000	32950.000000	32950.000000	32950.000000	32950.000000	3.295000e+04	32950.000000	32950.000000	32950.000000	32950.000000
mean	10866.455994	10618.520486	41.157633	11.609761	320.125232	6.986250e+04	13.148421	0.839909	9.292049	0.050410
std	7199.629493	6956.047888	9.858243	3.599964	207.081110	6.656225e+04	6.680041	1.053524	4.379854	0.228429
min	500.000000	500.000000	36.000000	5.420000	15.690000	4.000000e+03	0.000000	0.000000	2.000000	0.000000
25%	5200.000000	5050.000000	36.000000	8.490000	165.282500	4.113275e+04	7.980000	0.000000	6.000000	0.000000
50%	9600.000000	9200.000000	36.000000	11.490000	275.650000	6.000000e+04	13.200000	1.000000	9.000000	0.000000
75%	15000.000000	14500.000000	36.000000	13.990000	420.735000	8.400000e+04	18.390000	1.000000	12.000000	0.000000
max	35000.000000	35000.000000	60.000000	24.110000	1295.210000	6.000000e+06	29.990000	8.000000	44.000000	4.000000

Correlation

- For analysis purposes, we use the relevant data to get the correlation
- We use total_pymnt as a factor indicating full payment or part payment
- The probability that the borrower is likely to pay fully is related to:
  - 38% related to annual income
  - 85% related to installment amount
  - 28% to interest rate
  - 31% to term
  - 89% to funded amount
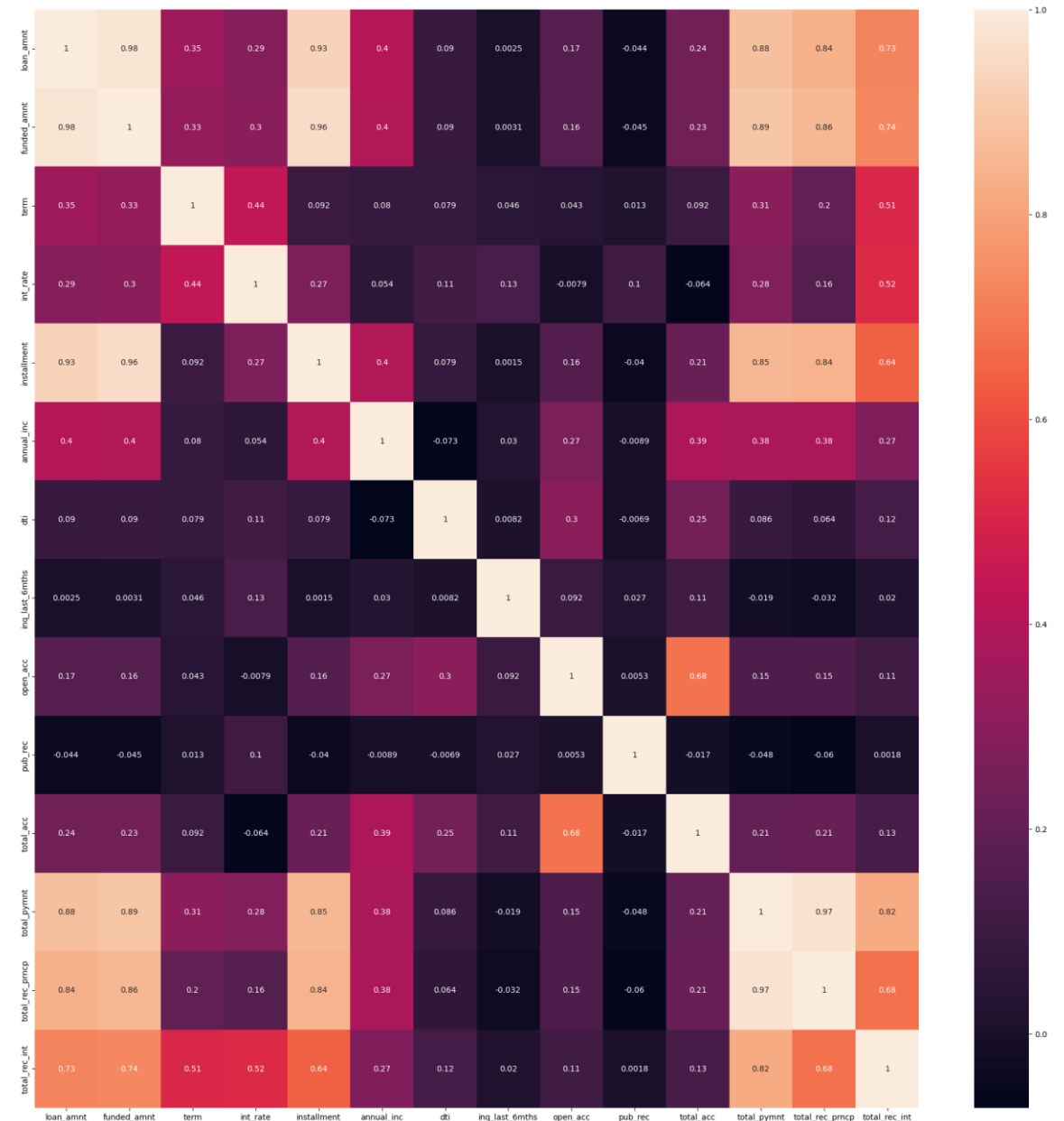
Summary

Basis the dataset, following are the observations on the major driving factors affecting the borrowers' loan payment status:
- Annual income
- Interest rate
- Monthly Installment
- Loan period
- Employment grade
- Employment length
- Debt to income ratio

THANK YOU