


Lending Club Case Business Presentation

by: Yassine Elhallaoui.
Jan_2021

Contents

- Business Problem Overview.
 - Data Overview.
 - Data Cleaning.
 - Exploratory Data Analysis (EDA).
 - Insights and Recommendations.
- 

Business Problem Overview and Solution Approach

- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'. If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study. process, earn more through the interest on loans.
- Challenges will be to Identify a potential customers that will get the loan without putting the company at risks.

Data Overview

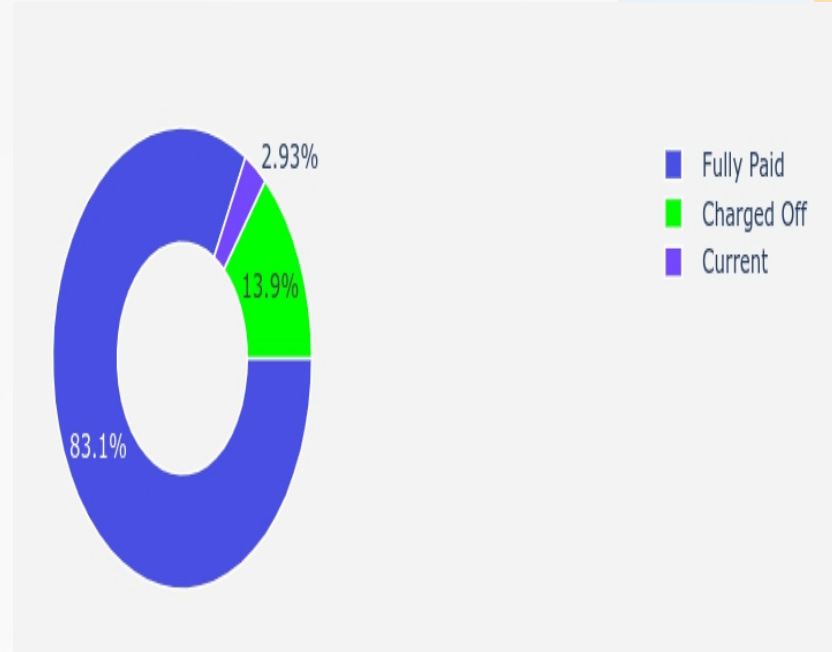
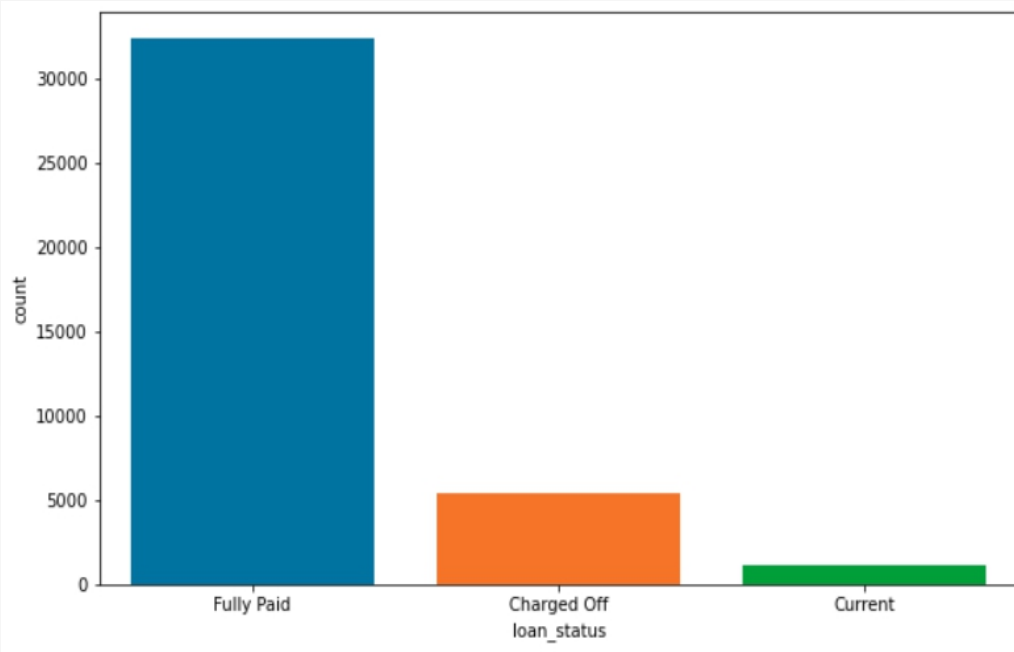
- The data contains information about 39717 customers and their information with 111 columns.
- The details for each customer includes ID, loan_amnt, installment, purpose of the loan request, Annual income, the status of the loan in two types of decisions:
 - **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
 - Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
 - Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
 - Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan.
 - **Loan rejected:** The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset). All the provided variables are numerical and no conversion is needed.
- Some unnecessary columns need to be eliminated to make the analysis clearer.
- Data Contain many data-types : [Categorical, Numerical and Dates] that need to be formatted.

Data Cleaning

- No duplicated observation were located on the dataset.
- Some Features need to be eliminated because all null values shows same amount as the total observation [No data 100%].
- All the values in 'id' column are unique, and it should be dropped.
- Features ["term", "int_rate", "revol_util", "zip_code"] formatted by removing extra character to keep only the numerical values.
- Features ["issue_d", "earliest_cr_line", "last_pymnt_d", "last_credit_pull_d"] are date values, and we are interesting about the year for our analysis.
- Features ['collections_12_mths_ex_med', 'acc_now_delinq', 'chargeoff_within_12_mths', 'delinq_amnt', 'tax_liens'] are dropped because all statistical values are with 0 value.
- Outliers imputed by the limits value of the Quantiles.
- Data now after cleaning is with 38889 observations and 45 columns.

Exploratory Data Analysis

Univariate analysis

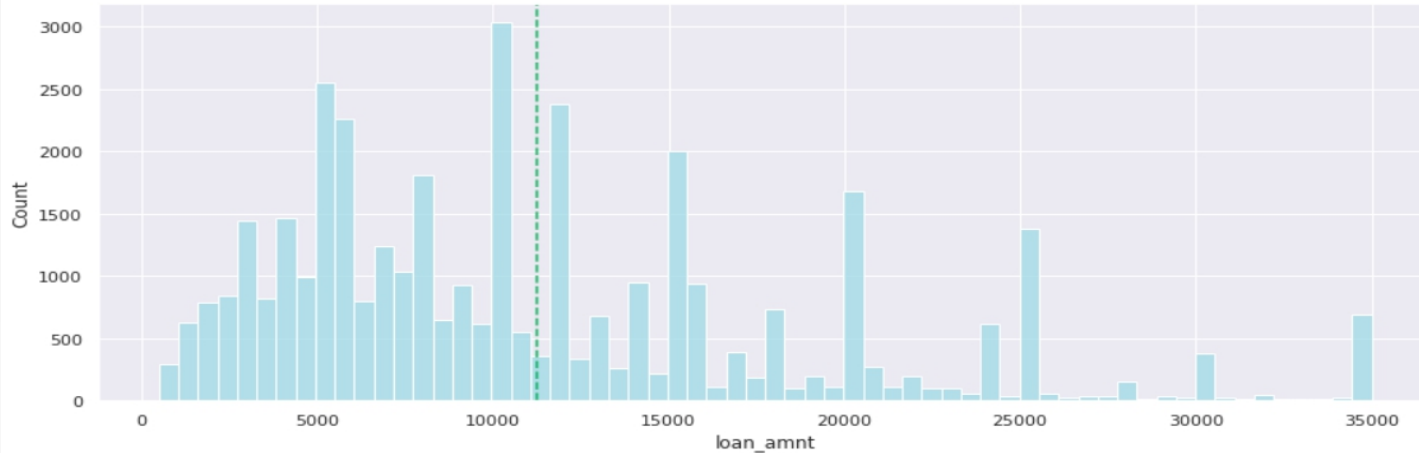


- Majority of customers are fully-paid {83%} followed by the customers that are charged Off {14%} and didn't pay the due installment on time.
- Few customers are with loan status 'current' {3%} which mean the loan is not yet completed.

Exploratory Data Analysis

Univariate analysis

```
loan_amnt  
Skew : 1.06
```

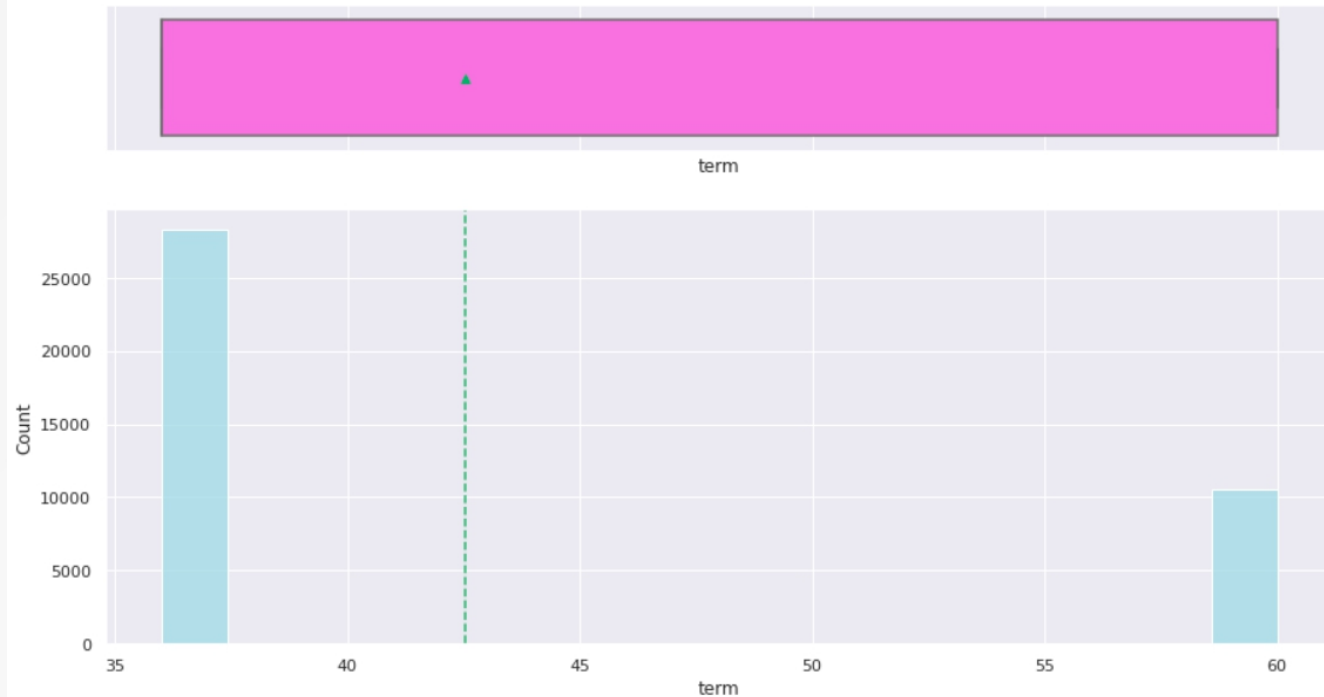


- The Loan amounts are between 5000 and 15000 with a presence of outliers on the variable.
[cleaned]

Exploratory Data Analysis

Univariate analysis

```
term  
Skew : 1.02
```

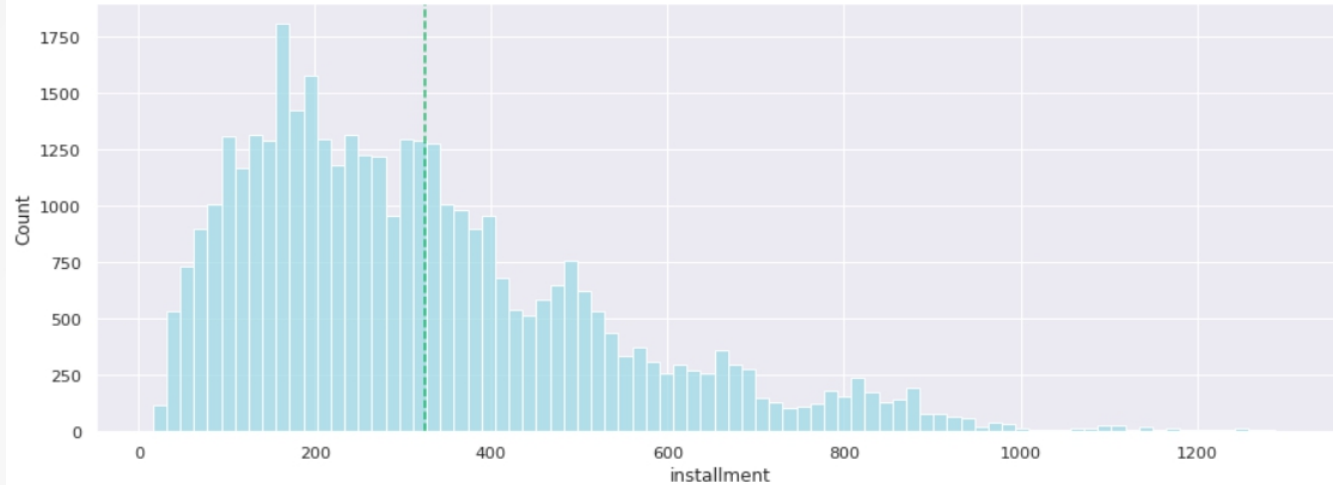


- Term represents the number of payments on the loan, and the majority are choosing 36 months.

Exploratory Data Analysis

Univariate analysis

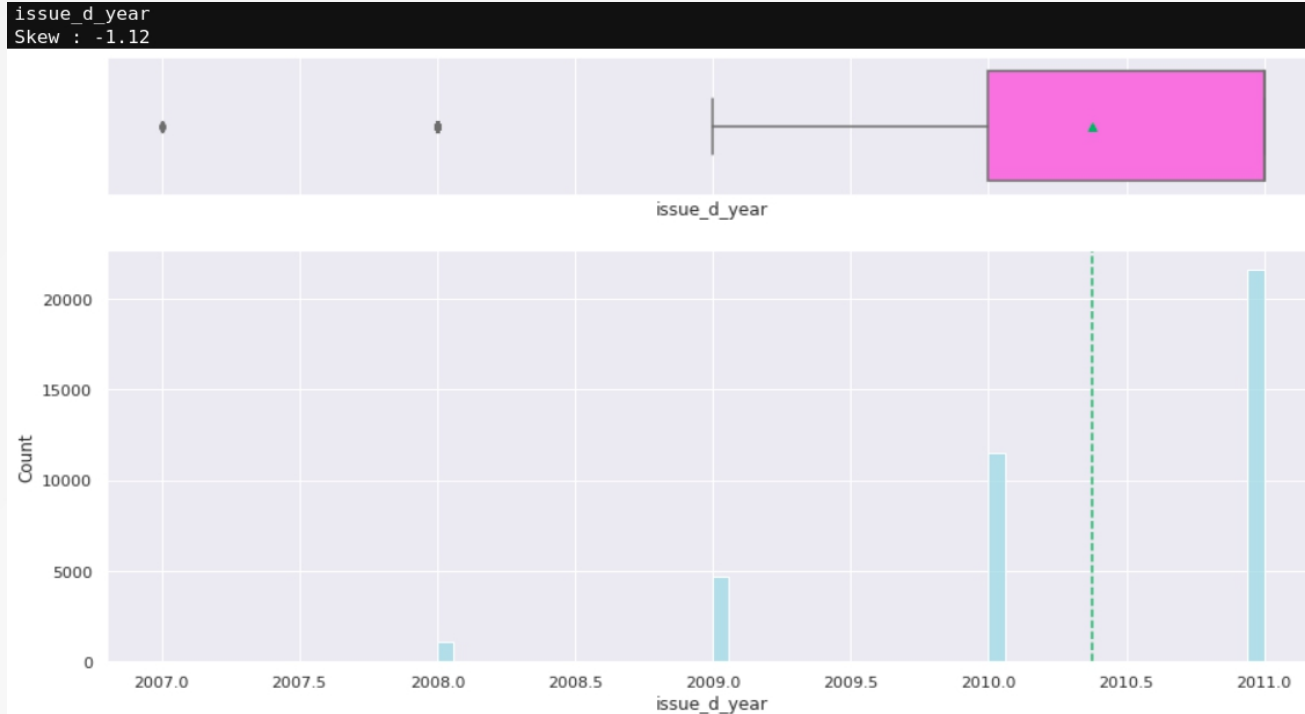
```
installment  
Skew : 1.13
```



- Installment values are concentrated on the range of 150 and 450 payment per month, with a maximum of 850 per month, outliers also present on this variable. [cleaned]

Exploratory Data Analysis

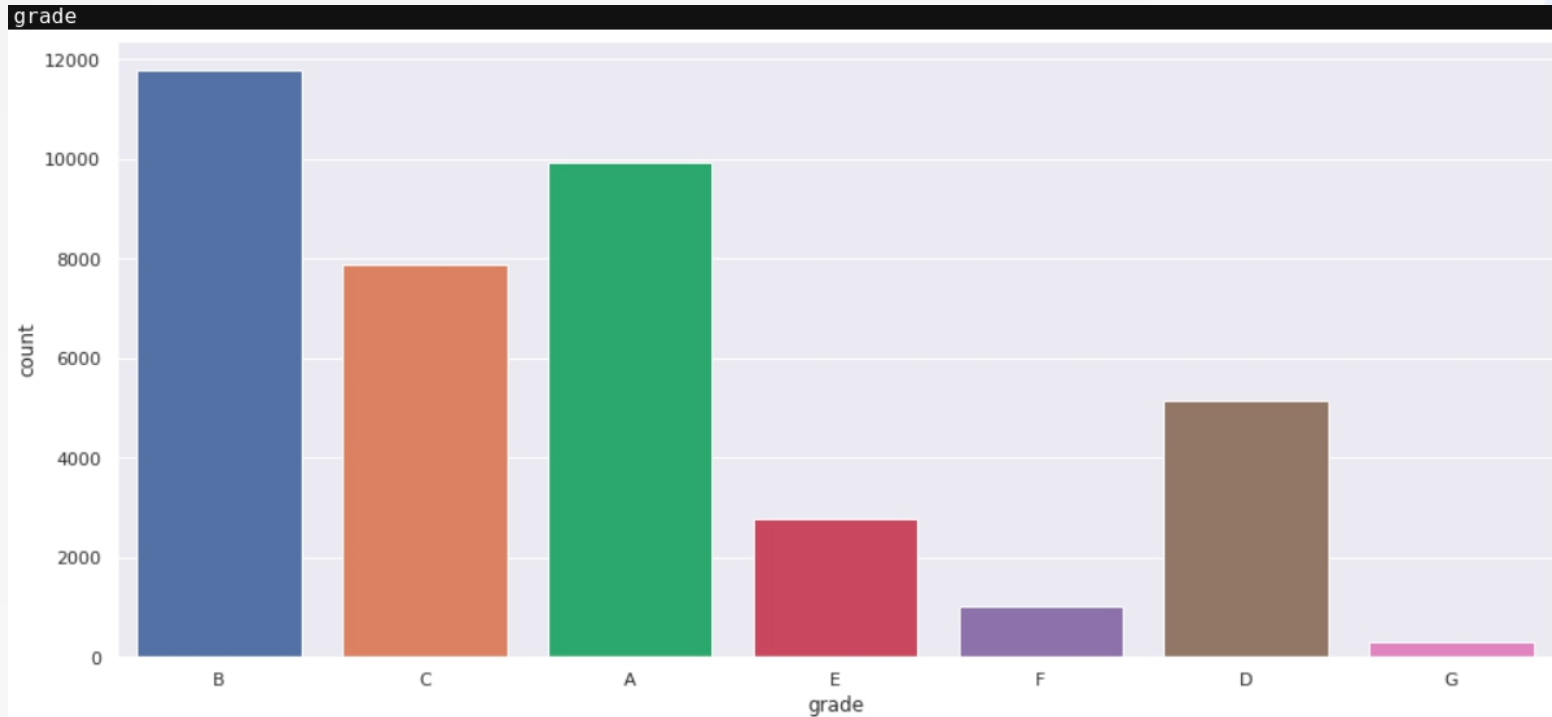
Univariate analysis



- Majority of loans are issued on 2010 and 2011.
- A big uptrend of the issued loans from 2007 till 2011.

Exploratory Data Analysis

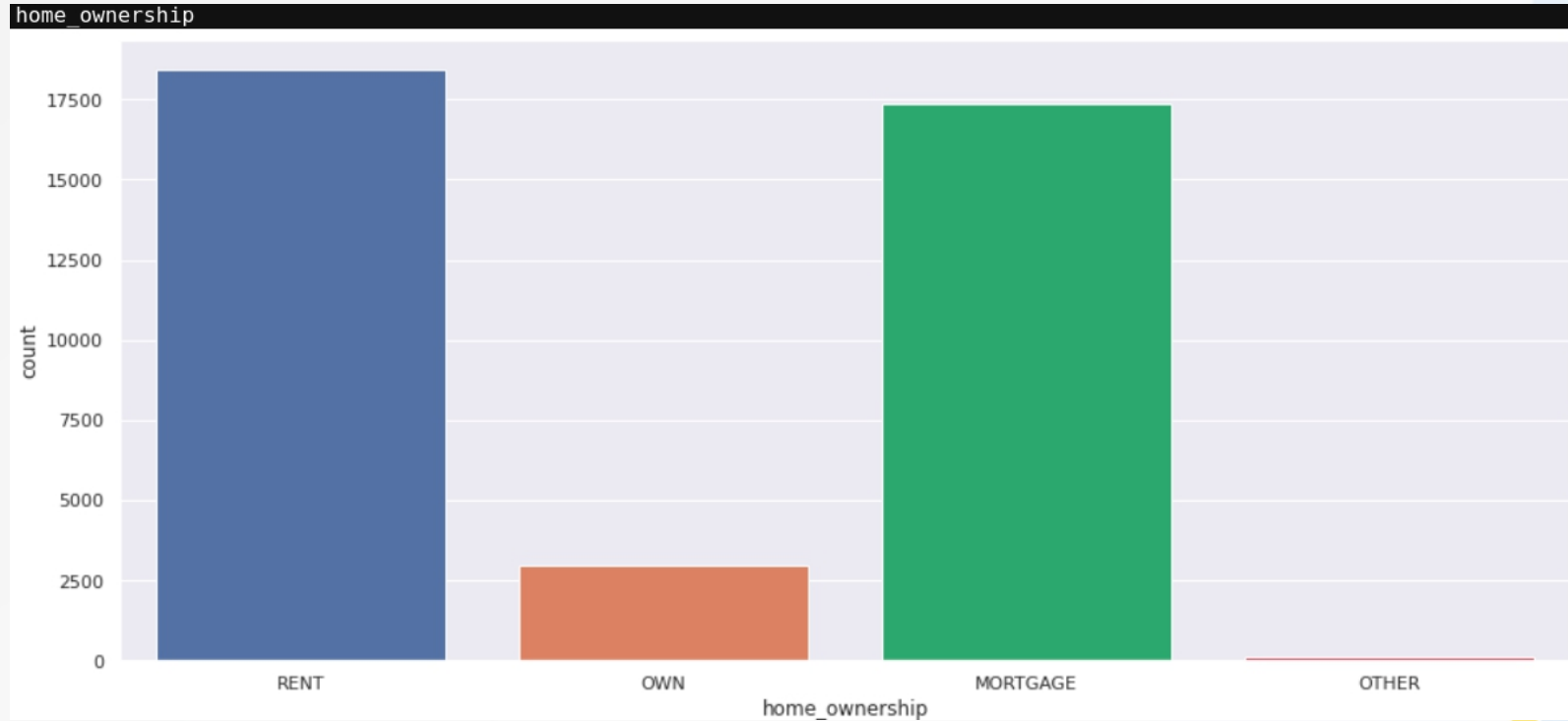
Univariate analysis



- Loan grade B is the most assigned, followed by A, C, D, E, F and G.

Exploratory Data Analysis

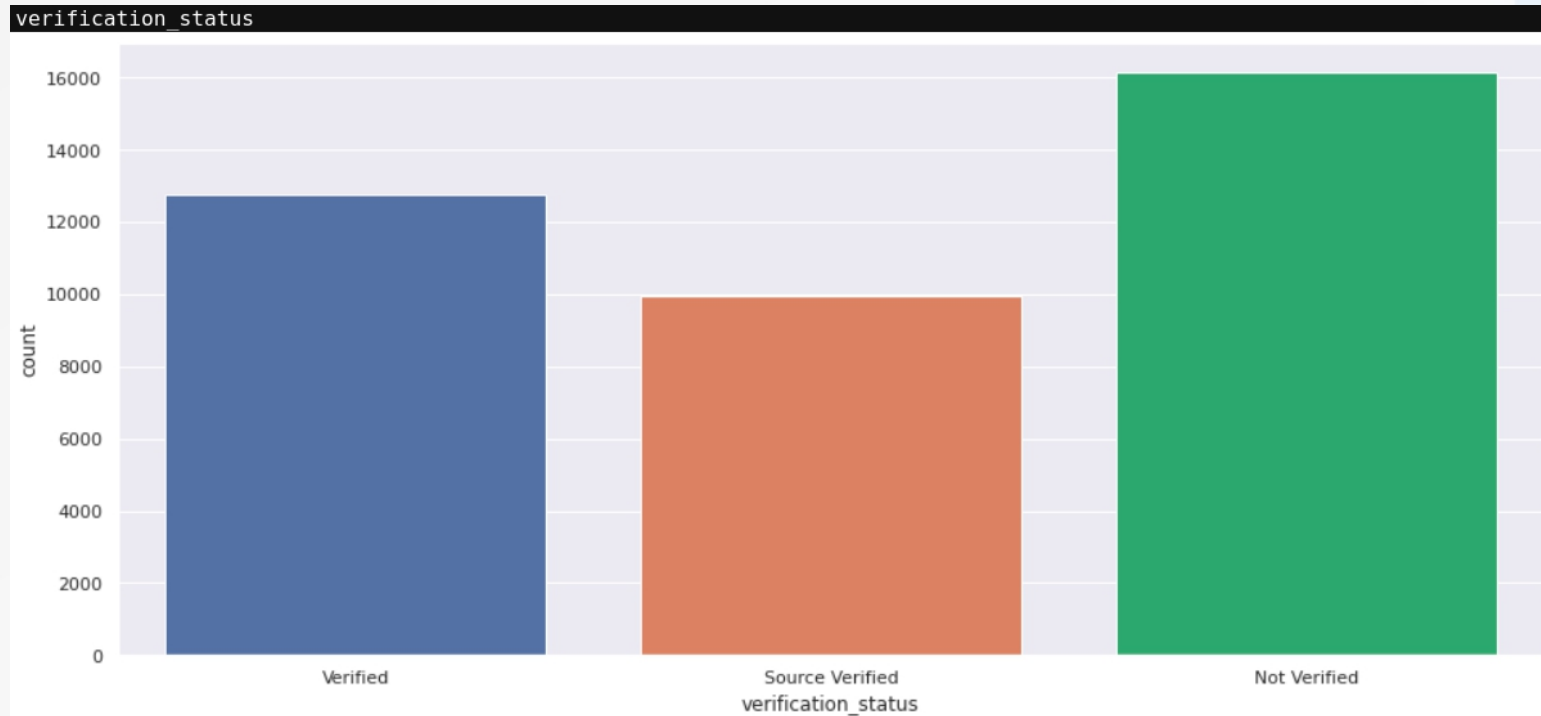
Univariate analysis



- Rent and Mortgage are the most status of home_ownership during the registration for the loan.

Exploratory Data Analysis

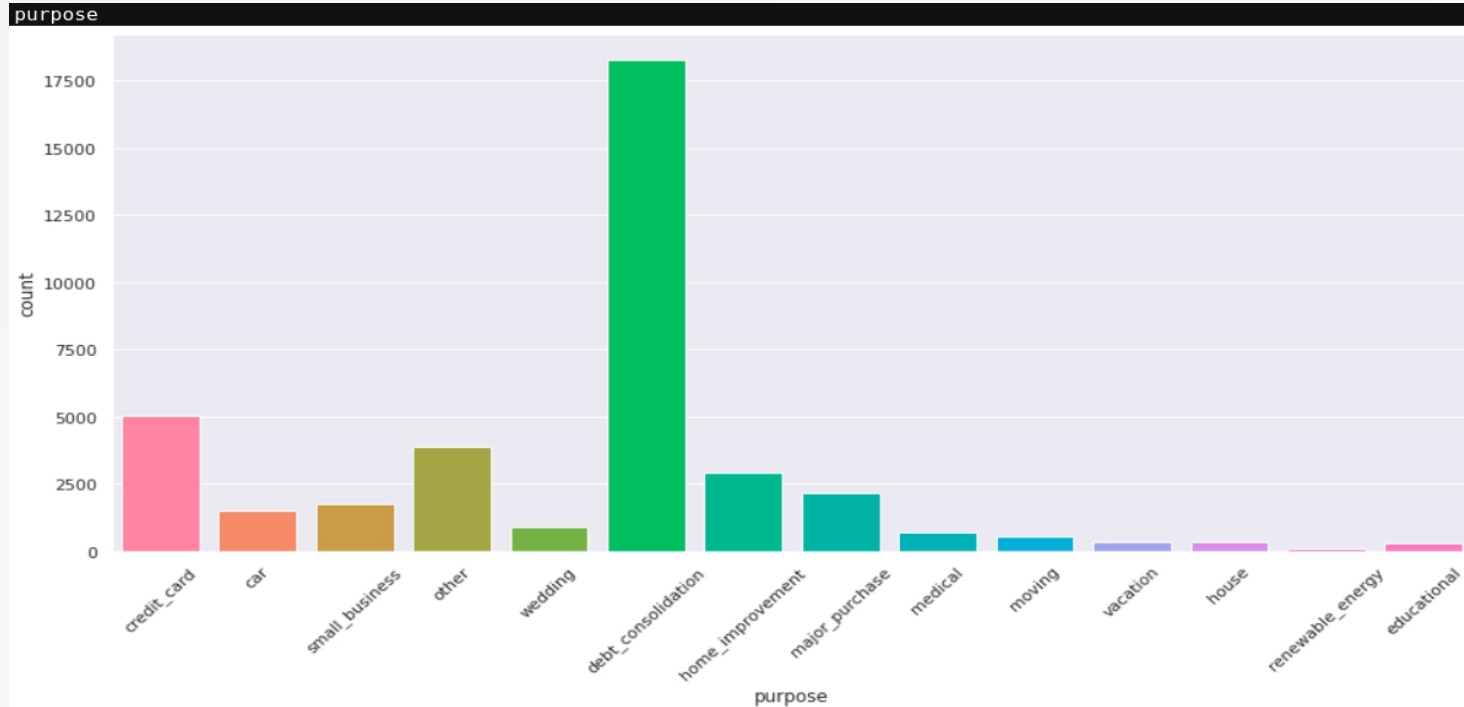
Univariate analysis



- Majority of the customer's income is not verified.

Exploratory Data Analysis

Univariate analysis

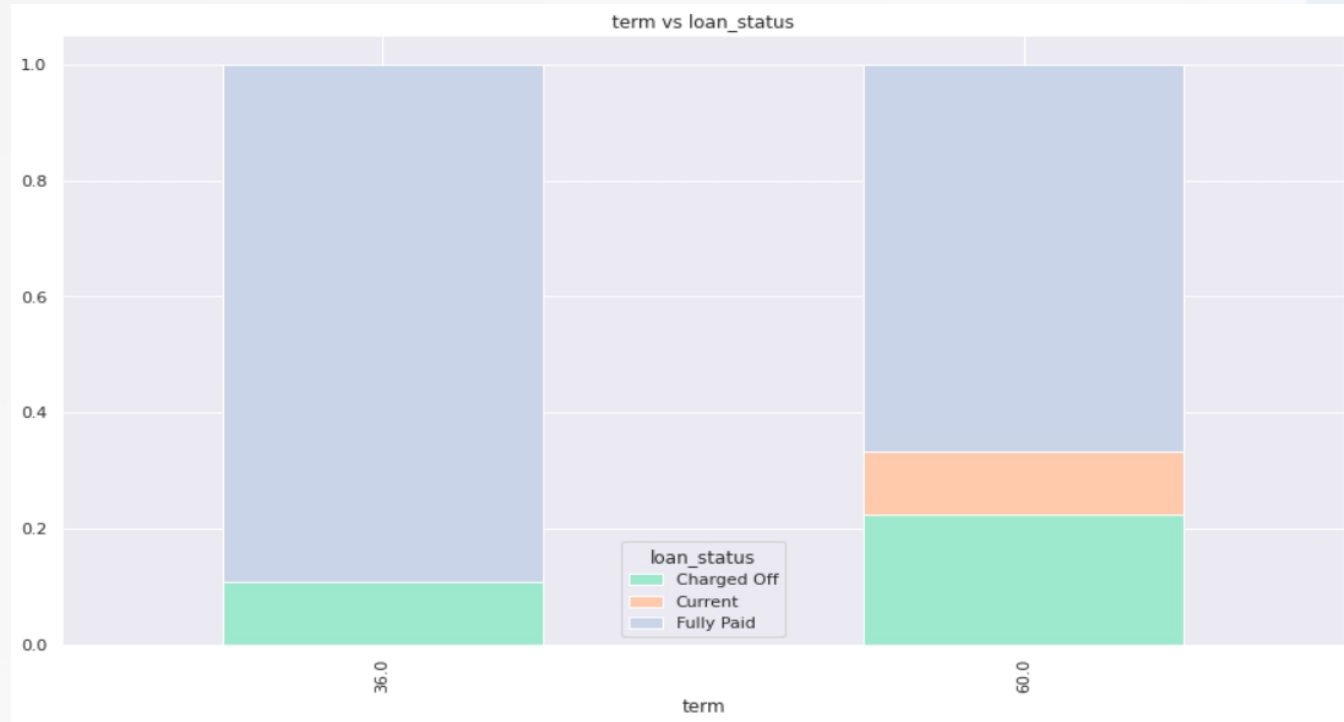


- Surprisingly most of customers purpose for taking the loan is because of weddings mainly.
- credit_card, other, home improvement, major_purchase, small_business, car are other major motive for a customer to request the loan.

Exploratory Data Analysis

Bivariate analysis

term vs loan_status

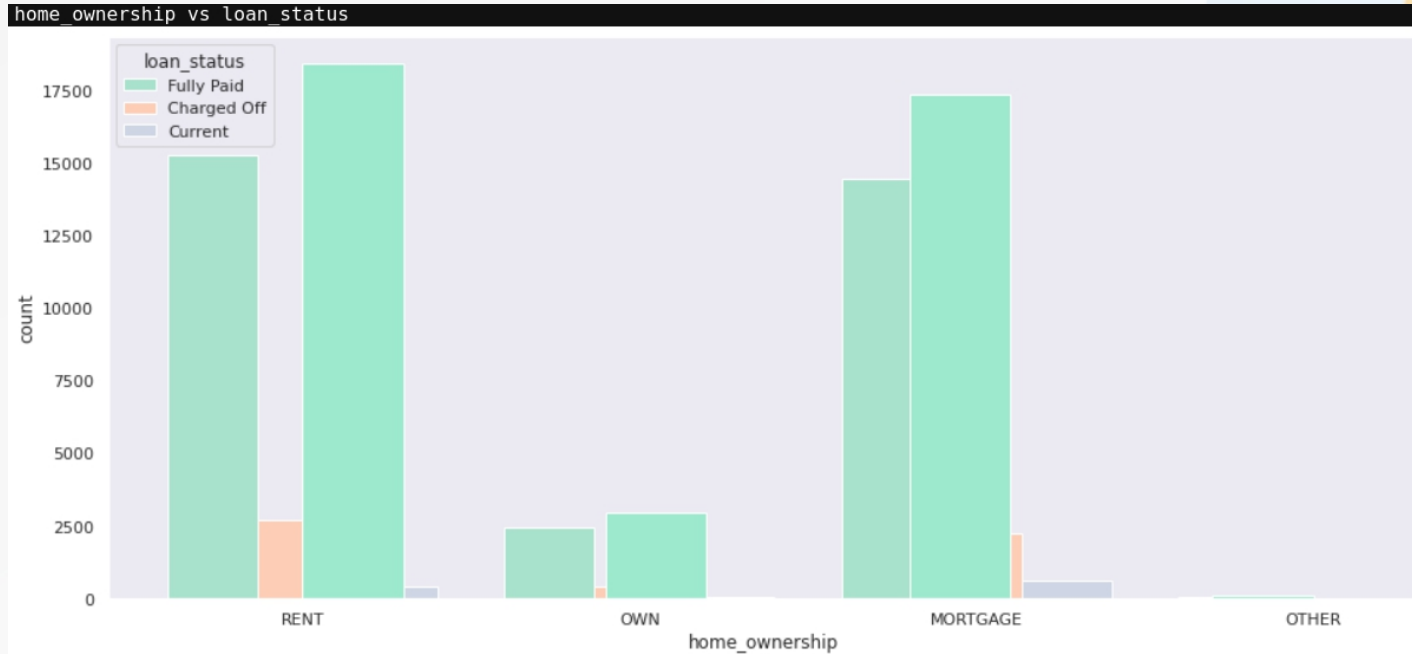


- Loans with “current status” are present only when the term is 60months.

Exploratory Data Analysis

Bivariate analysis

home_ownership vs loan_status

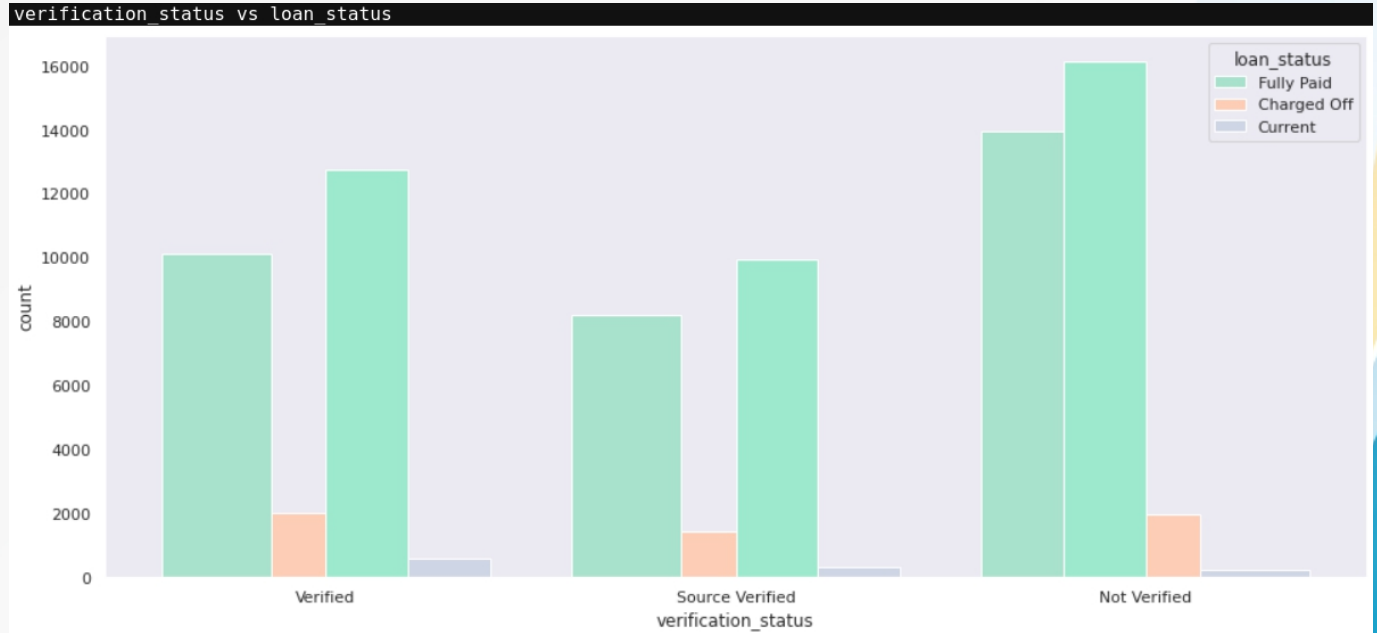


- Fully Paid loan is a common for all the home ownership status.

Exploratory Data Analysis

Bivariate analysis

verification_status vs loan_status

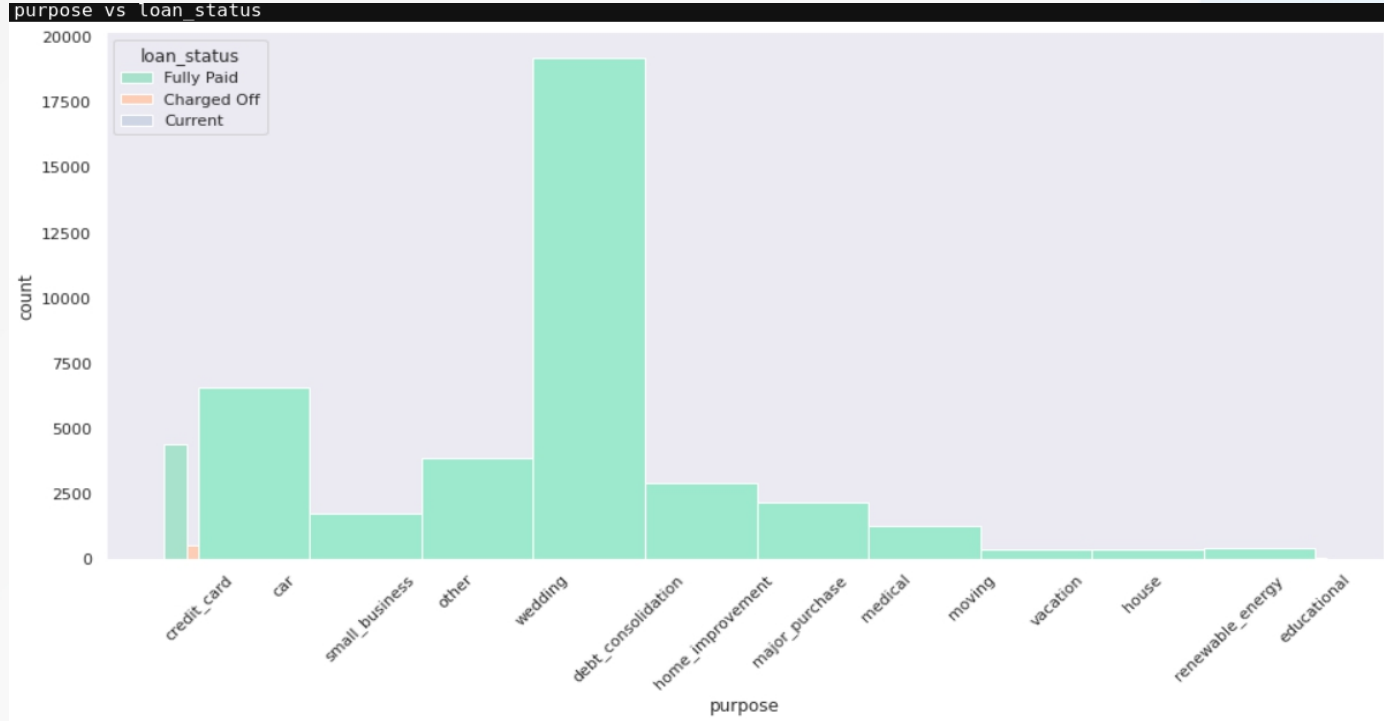


- Verifying the income is not necessary to determine the loan status.

Exploratory Data Analysis

Bivariate analysis

purpose vs loan_status



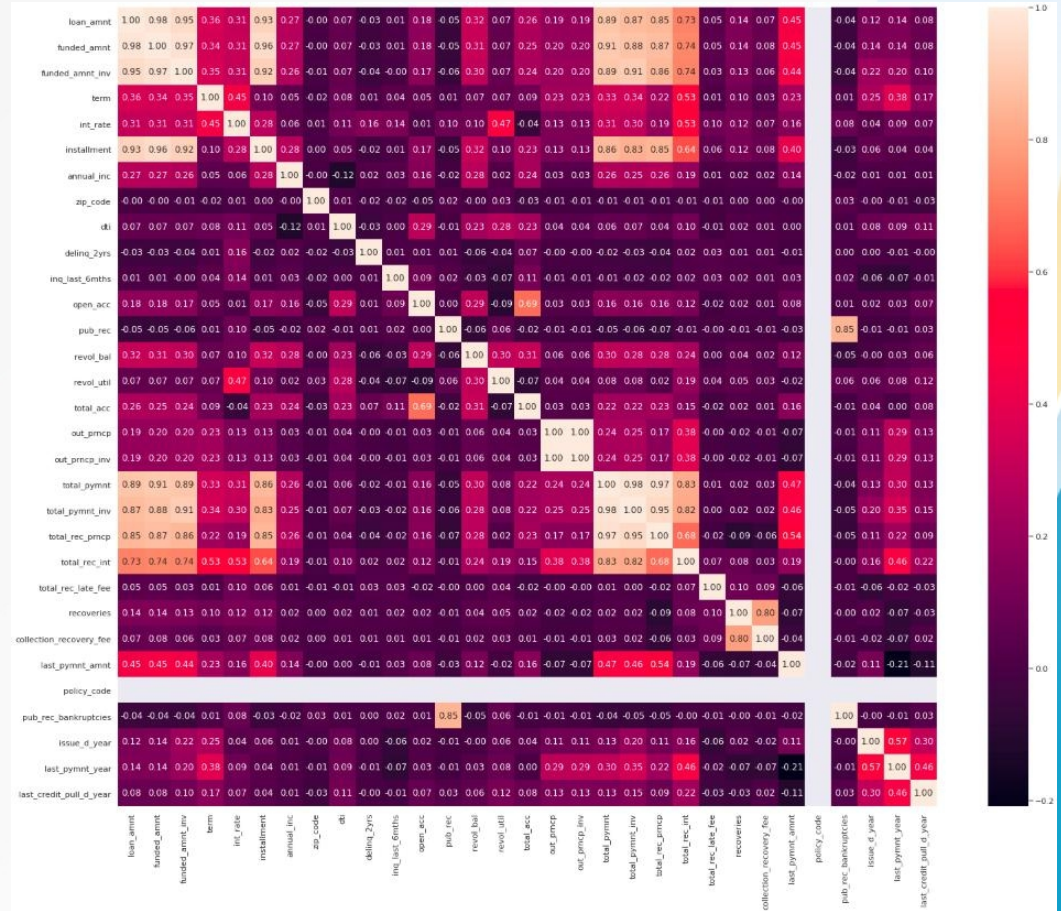
- Majority of purposes of the loan are fully paid, and wedding is the top of the list.

Exploratory Data Analysis

Correlation & Relation between Variables

- Positively high correlated variables between each other :
['loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'installment', 'total_paymnt', 'total_paymnt_inv', 'total_rec_pmcp', 'total_rec_int'].
- pub_rec_bankruptcies is positively high correlated with pub_rec.
- collection_recovery_fee is positively high correlated with recoveries.

Bivariate analysis



Insights and Recommendations

According to our EDA we can summarize that :

- Majority of customers are fully-paid {83%} followed with the customer that are charged Off {14%} and didn't pay the due installment on time.
- Rent and Mortgage is the most status of home_ownership during the registration for the loan.
- Loans with “current status” are present only when the term is 60 months.
- Fully Paid loan is a common for all the home ownership status.
- Majority of purposes of the loans are fully paid, and wedding is the top of the list.
- Verifying the income is not necessary to determine the loan status.
- Installment values are concentrated in the range of 150 and 450 payment per month, with a maximum of 850 per month.

Those above points can be used to have a general idea of any customer requested a loan to estimate his pattern.

Thank You

by: Yassine Elhallaoui.
Jan_2022