

Lending Club Case Study



Submitted by:
Kifah Muqalled

Problem Statement

This organization is the largest online loan marketplace for personal, business, and medical loans. A rapid internet interface lets borrowers get lower-interest loans.

Lending to 'risky' candidates is the biggest source of credit loss for most lending organizations. The lender loses credit when the borrower defaults or flees. Thus, defaulting borrowers cost lenders the most. Customers branded 'charged-off' are 'defaulters'.

By identifying hazardous loan applicants, credit loss can be reduced. In this case study, EDA is used to identify such applications.

The company needs to know the primary drivers of loan default, or variables that strongly predict default. This might help the organization assess portfolio and risk.

Analysis Approach

1. Understanding the data set
2. Data Cleaning
3. Data preparation
4. Data Analysis
5. The Conclusions

Data Cleaning Process

1. Check duplicate rows
2. Check the missing values and drop columns that have >90% of missing values
3. Filter the data set to include the important columns only
4. Refill missing values with a proper value

Data Preparation

1. Categorize columns to numerical and categorical
2. Change column format if needed

Data Analysis

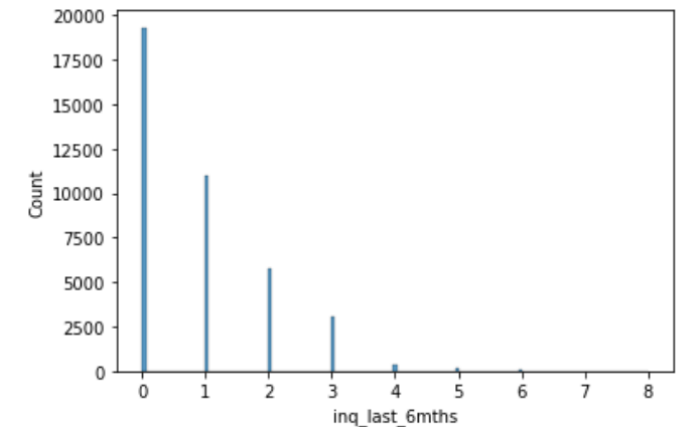
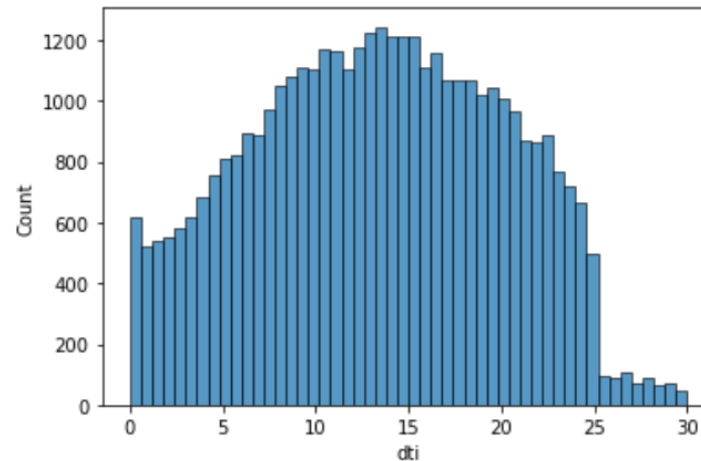
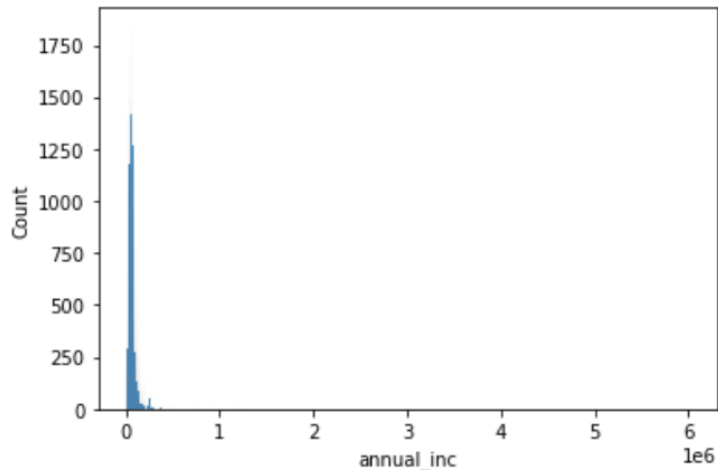
I did below data analysis approach and provide the observations

1. Univariate analysis

a. Numerical analysis

Finding and Observation :

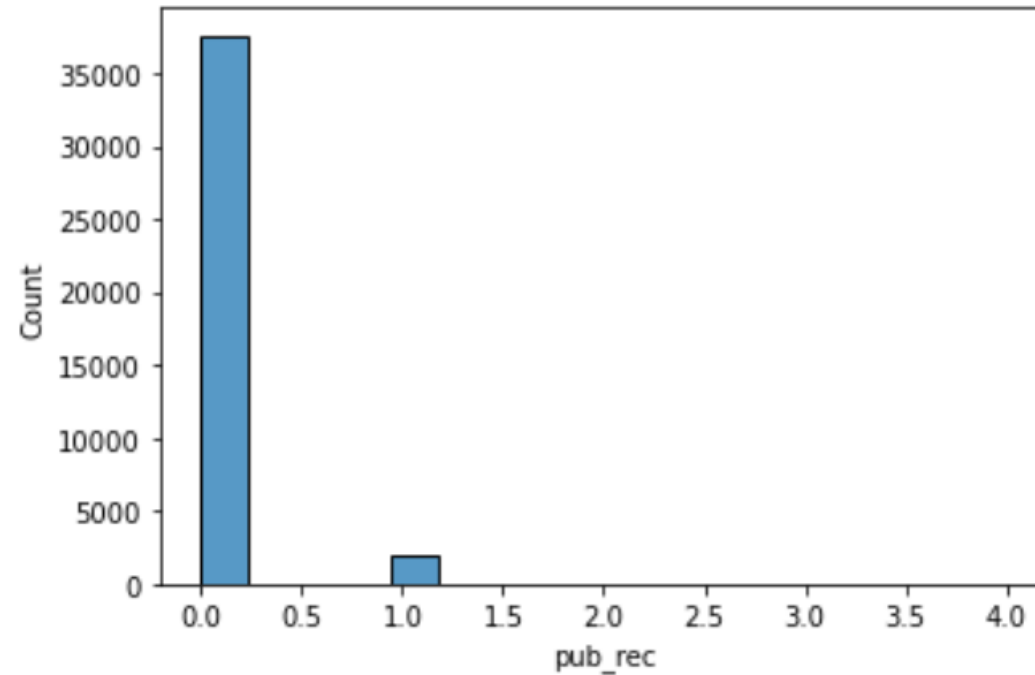
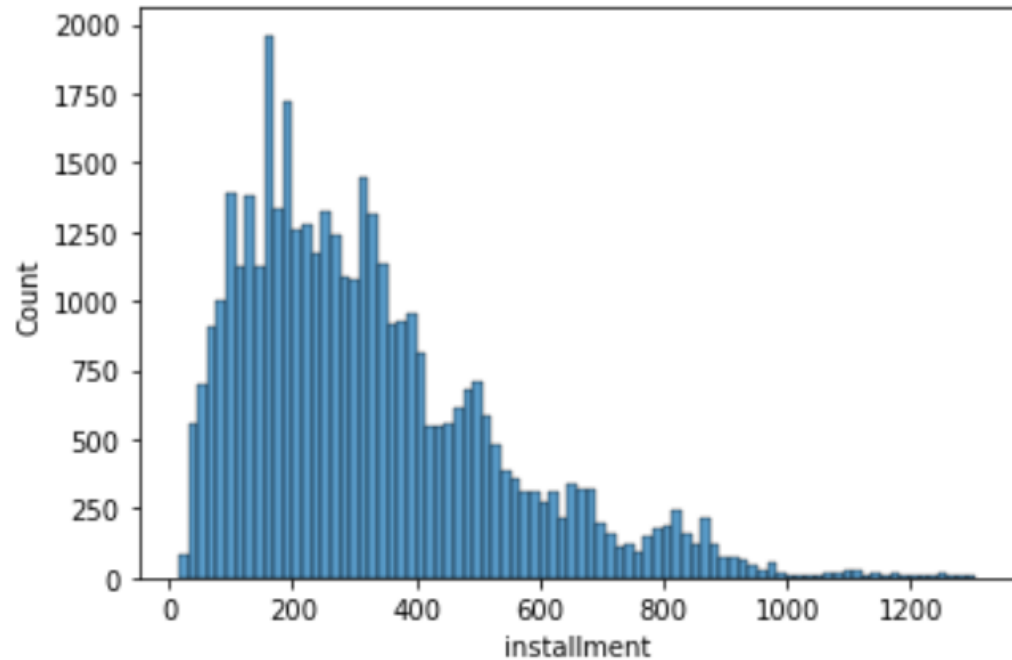
- Annual_inc majority near to 0
- There is a big drop for DTI after 25 .
- inq_last_6mths start with highest value at 0 , then decrease



a. Numerical analysis (Continue)

Finding and Observation :

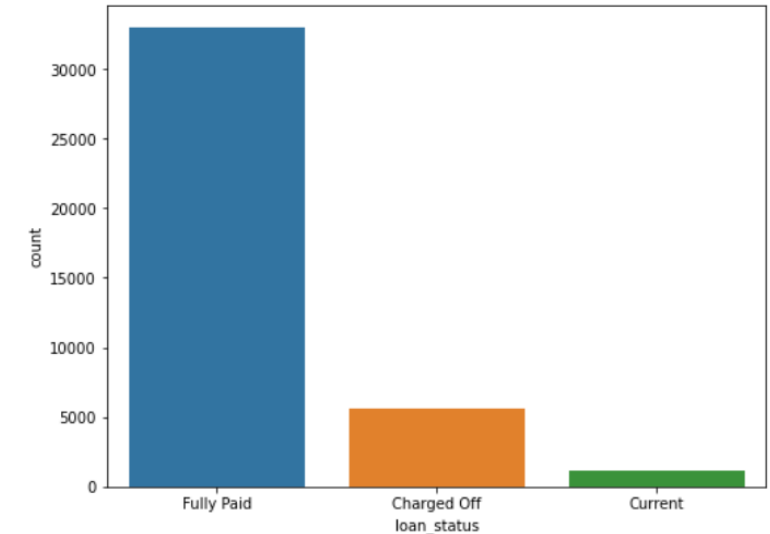
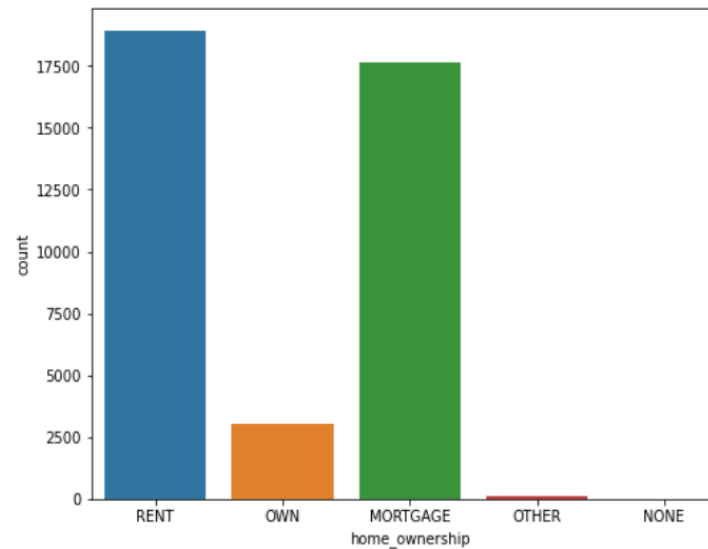
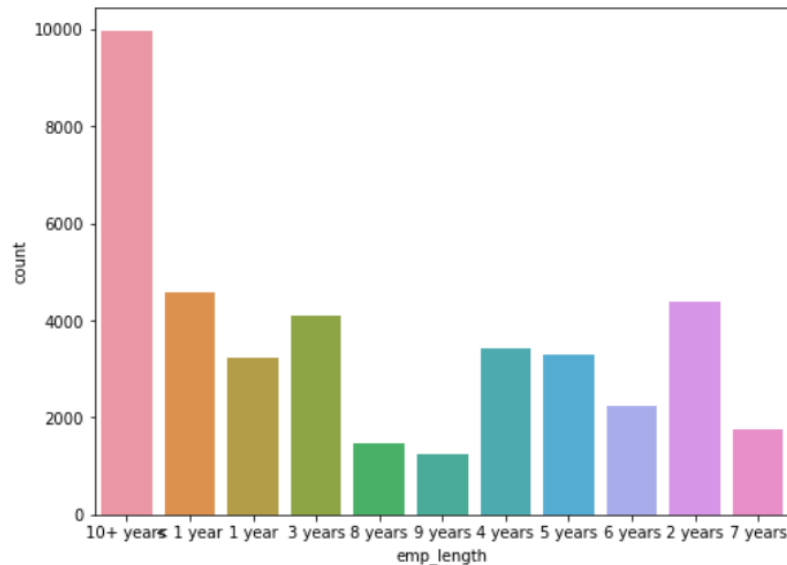
- instalment majority between 50 and 400
- The majority of pub_rec at 0 to 2



b. Categorical analysis

Finding and Observation :

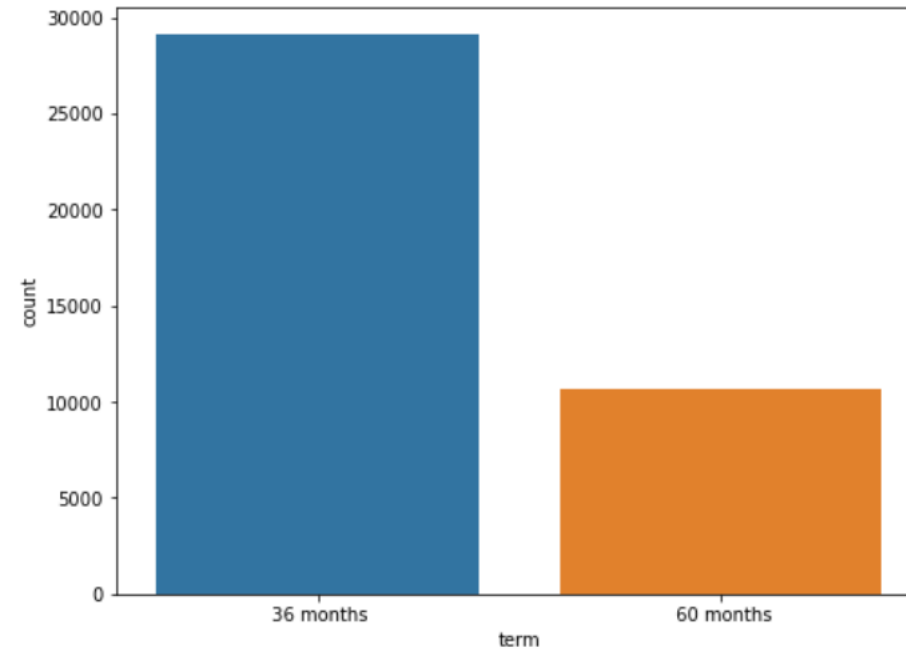
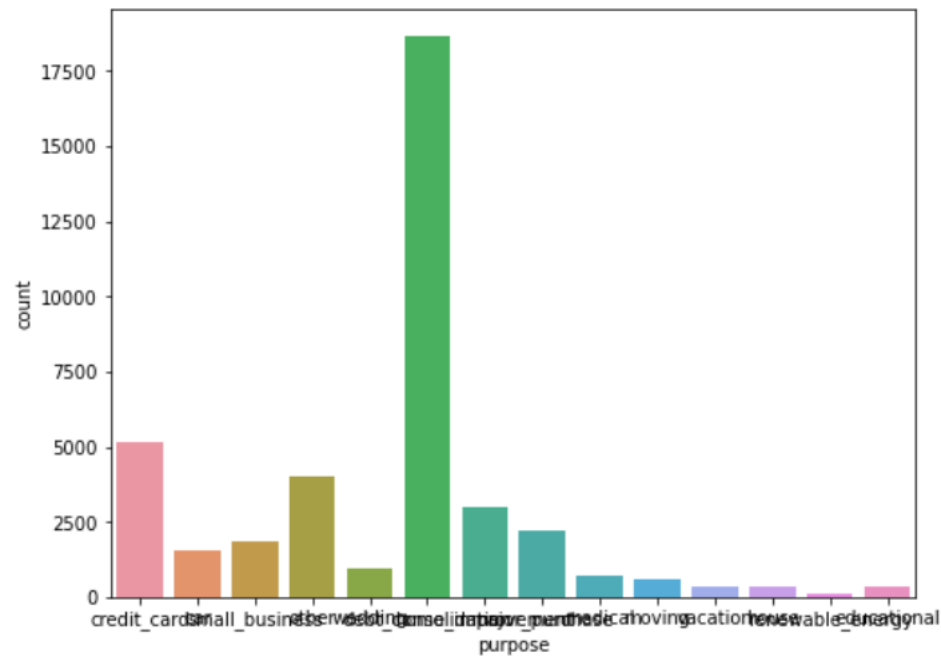
- The majority of emp_length is 10+ .
- The majority of home_ownership is rent , then mortgage .
- The majority of loan_status is fully paid and charge off less than 75 % of fully paid
- The majority of its purpose is debt consolidation .
- The majority of the term is 36 months (60 months , almost half 26 months) .



b. Categorical analysis (Continue)

Finding and Observation :

- The majority of its purpose is debt consolidation .
- The majority of the term is 36 months (60 months , almost half 26 months) .

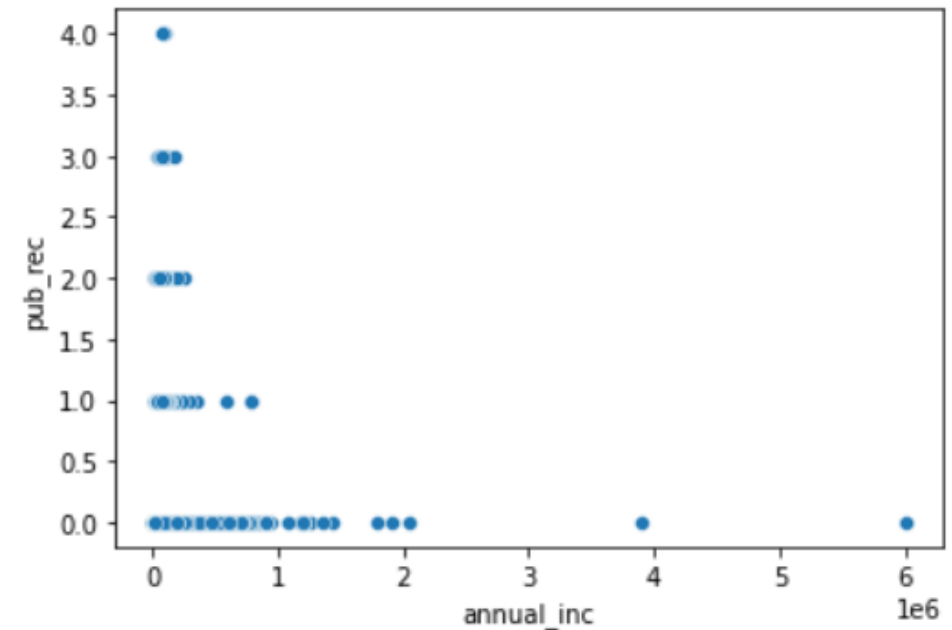
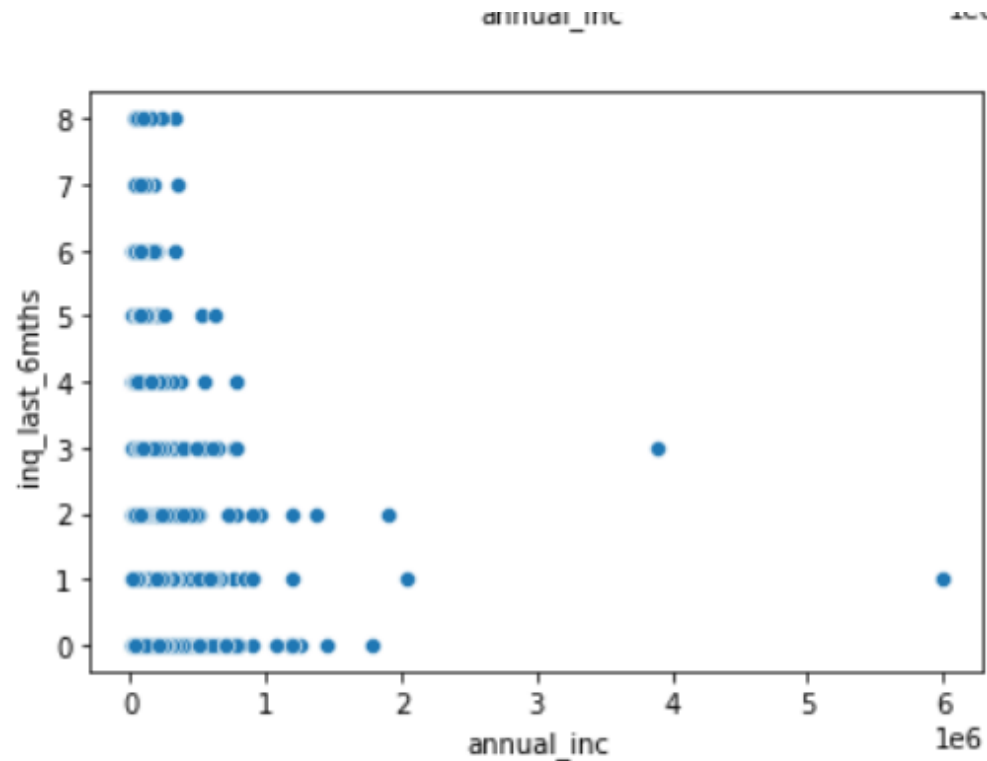


2. Bivariate Analysis

a) Numerical Vs Numerical:

Finding and Observation :

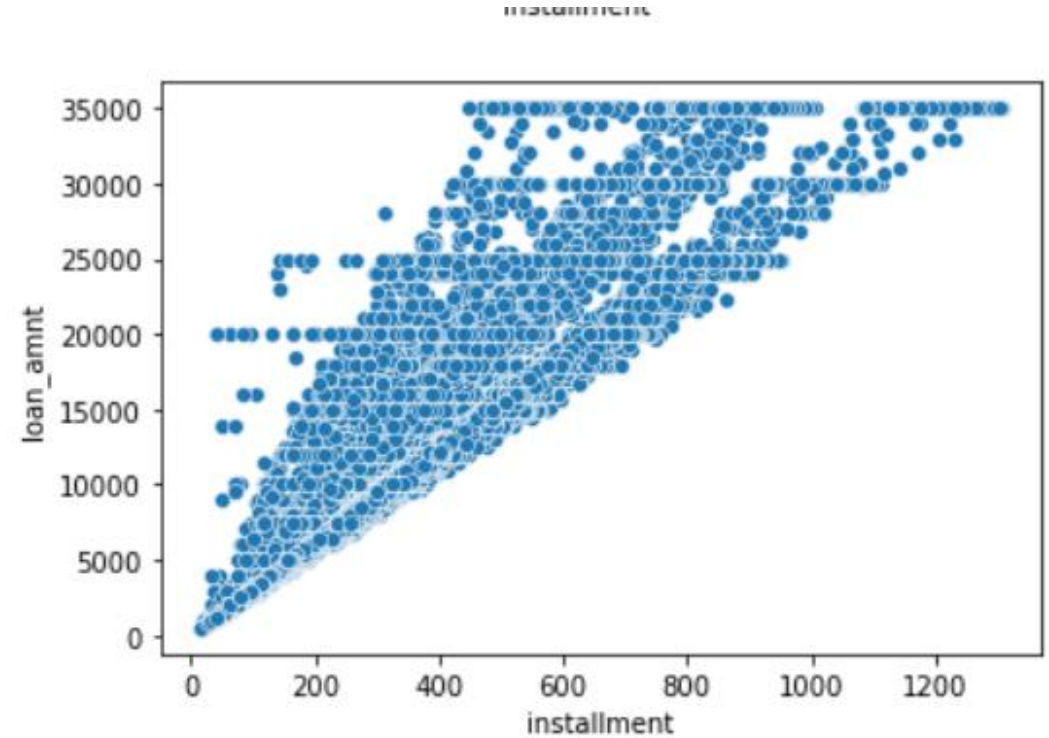
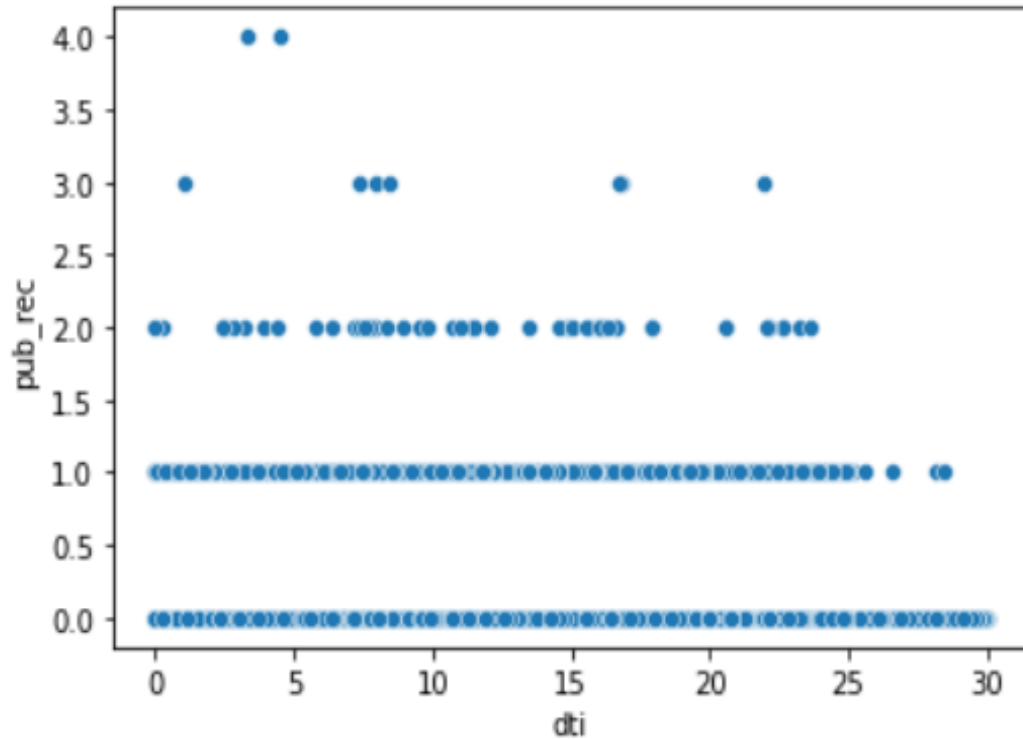
- There are outliers between annual_inc and inq_last_6mths
- The majority of annual_inc values between 0 and 2 , while 0 for pub_rec



a) Numerical Vs Numerical: (Continue)

Finding and Observation :

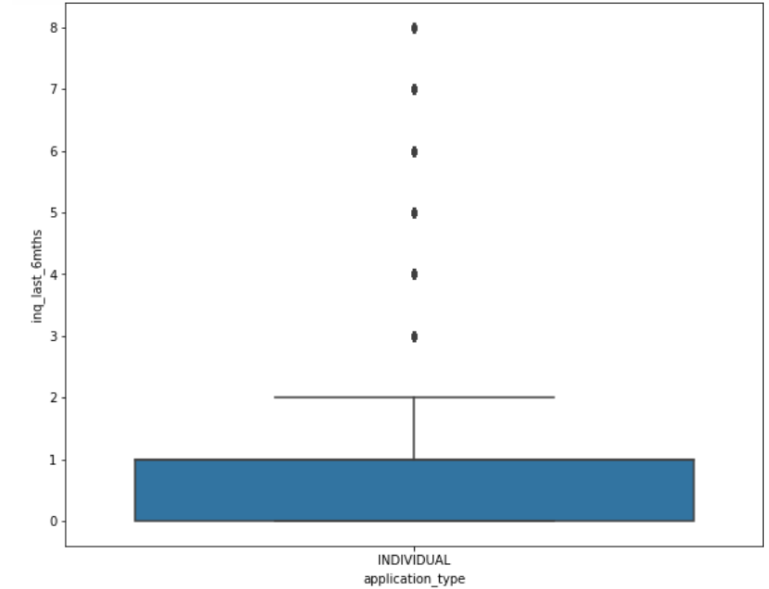
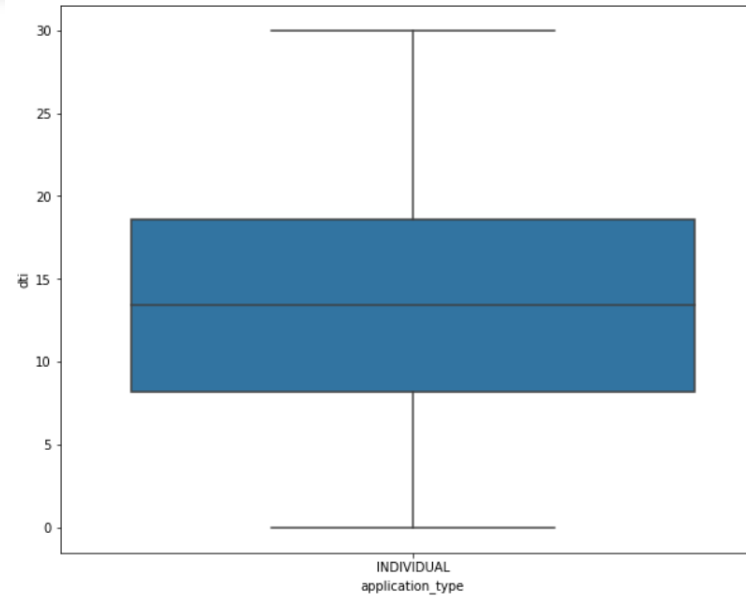
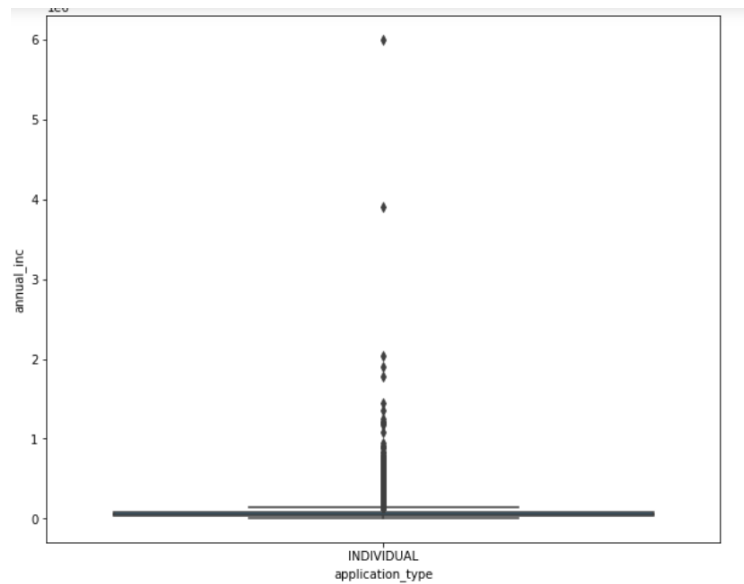
- The majority of pub_rec between 0 and 1 while distributed in DTI
- The a loan_amnt and instalment have a positive relationship as they increase together .



b) Numerical Vs Categorical:

Finding and Observation :

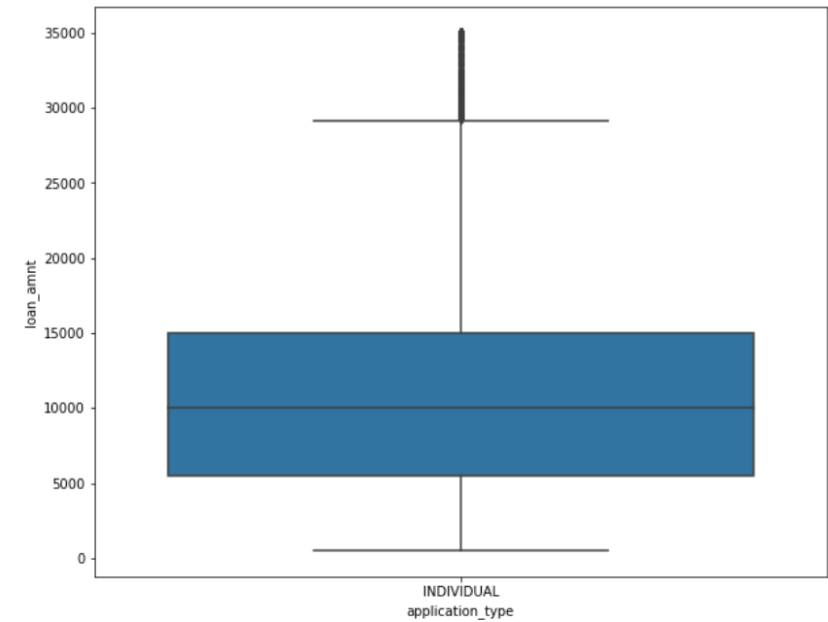
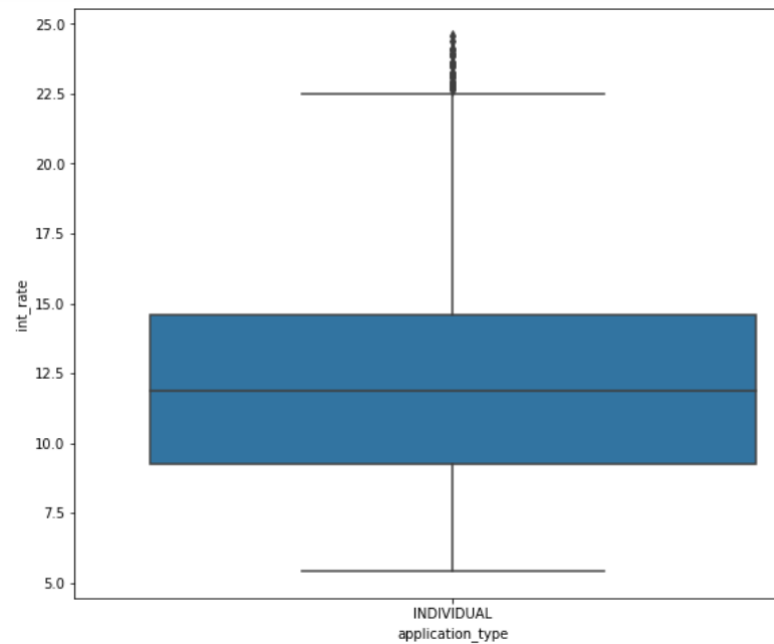
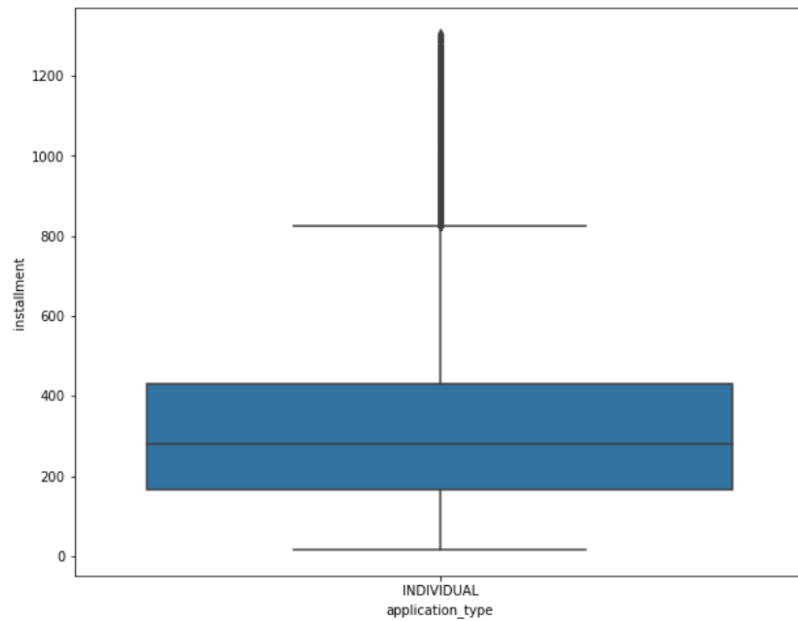
- The majority of application_type has 0 annual_inc
- The majority of application_type between 8 and 19 for dti
- The majority of application_type between 0 and 1 for inq_last_6mths.



b) Numerical Vs Categorical: (Continue)

Finding and Observation :

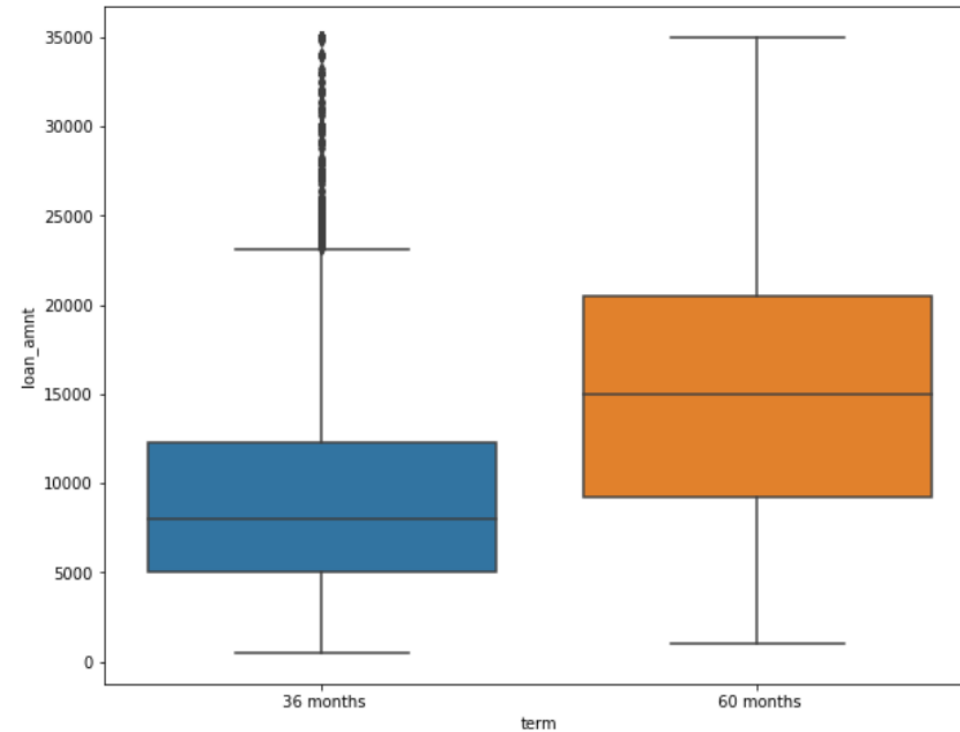
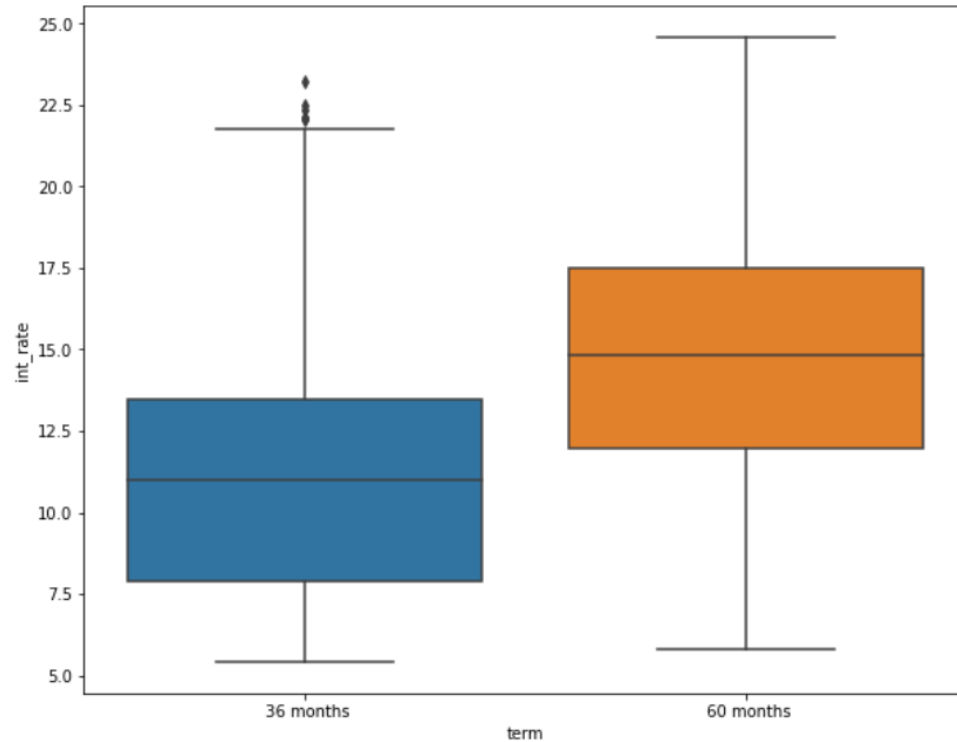
- The majority of application_type between ~200 and ~ 400 for installment
- The majority of application_type between 9 to 15 for int_rate
- The majority of application_type between 5000 to 15000 for loan_amnt



b) Numerical Vs Categorical: (Continue)

Finding and Observation :

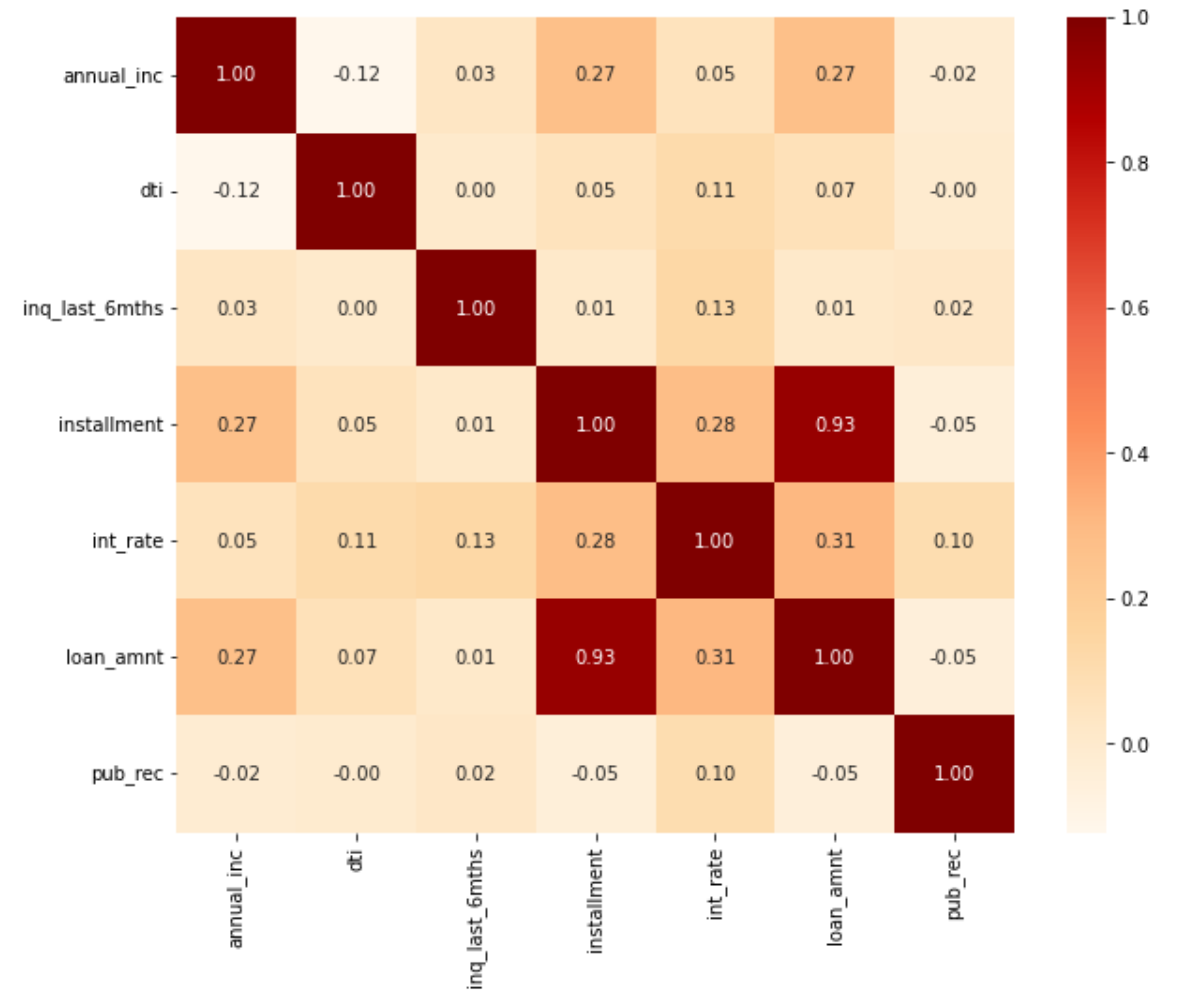
- The 60-month term is higher than the 36-month term for int_rate.
- The 60-month term is higher than the 36-month term for loan_amnt.



3. Multivariate Analysis:

Finding and Observation :

- loan_amt has strong positive correlation with instalment (0.93)
- loan_amt has positive correlation with int_rate (0.31)



The Conclusions

- Applications with a big drop for DTI after 25 should be carefully reviewed.
- Installment majority between 50 and 400 should be given more weight in the decision-making process.
- Applications with emp_length majority at 10+ should be prioritized.
- Loans for debt consolidation purposes should be given preference.
- Preferably accept loan applications with a term of 36 months, as it is the majority.
- Strong positive correlation between loan_amnt and installment should be considered in loan approval decisions.
- Positive correlation between loan_amnt and int_rate should also be considered.