# Lending Club Case Study

Devendran Mani
Mandar Ambuskar

09.08.2023

# Background – Lending Club Case Study

**Background**

1. A bank has given their customer's loan performance statistics data to analyse and find the probable future loan defaulters.
2. We are expected to do EDA on the given data to get the insight future expected problem.
3. All the key loan performance ex: prepose of loan, Interest rate, Funded amount, Invested amount etc. Shall be used for this analysis

**Business Objective**

1. Objective of this assignment is to perform EDA on the given loan performance data and give recommendations to bank to identify the probable future loan defaulter.

# Problem solving approach

The analysis is divided into four main parts

1. Data understanding
2. Data cleaning (cleaning missing values, removing redundant columns etc.)
3. Data Analysis
4. Recommendations

# Data understanding

1. Basic study of each column entry to know how the different customer profile looks. Some of the important columns in the dataset are loan amount, term, interest rate, grade, sub grade, annual income, purpose of the loan etc.

2. Identified the target variable, which we want to compare across the independent variables, is loan status. The strategy is to figure out compare the average default rates across various independent variables and identify the ones that affect default rate the most.

# Data Analysis

Univariate analysis

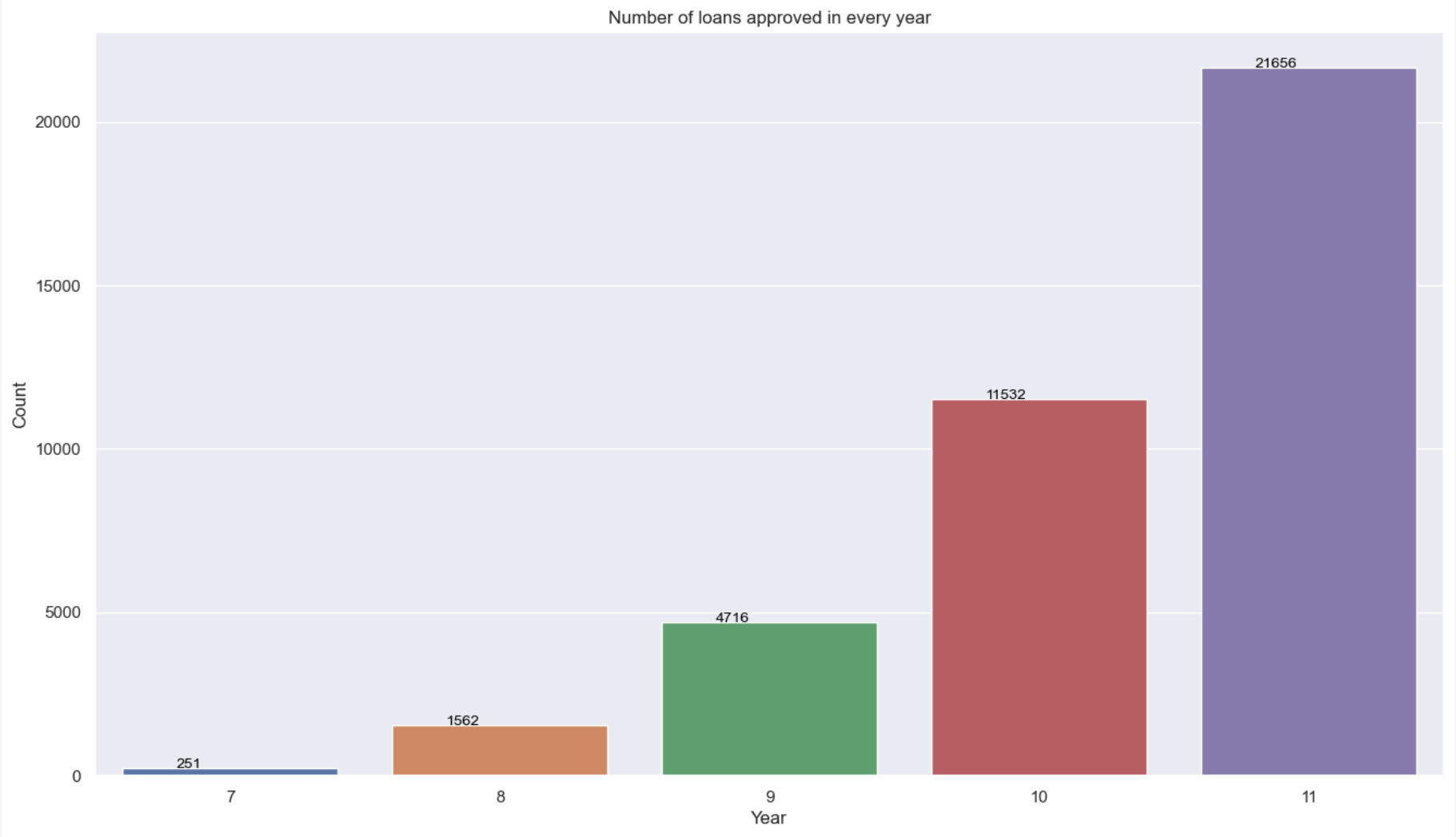
1. Studied the no of loan approvals happened year wise
2. Studied loan amount approved year wise
3. Studied count of loan approved loan tenure wise
4. Studied loan amount, funded amount and investment of funded amount
5. Studies customer nature based their grade category, also studied their population.
6. Studied customers employment duration to know their repaying capacity based on their job stability.
7. 8. Studied customer verification status to filter out the non-verified customer.

Bivariate analysis

1. Studied customer's Annual income v/s Funded amount to check their re-paying capacity.
2. Studied customer's Funded amount v/s home ownership to check their re-paying capacity.
3. Studied the last payment histories to check the recent financial status of customers.
4. Binned/Segmented the customer based on interest rate.
5. Binned/Segmented the customer based on loan amount wise.

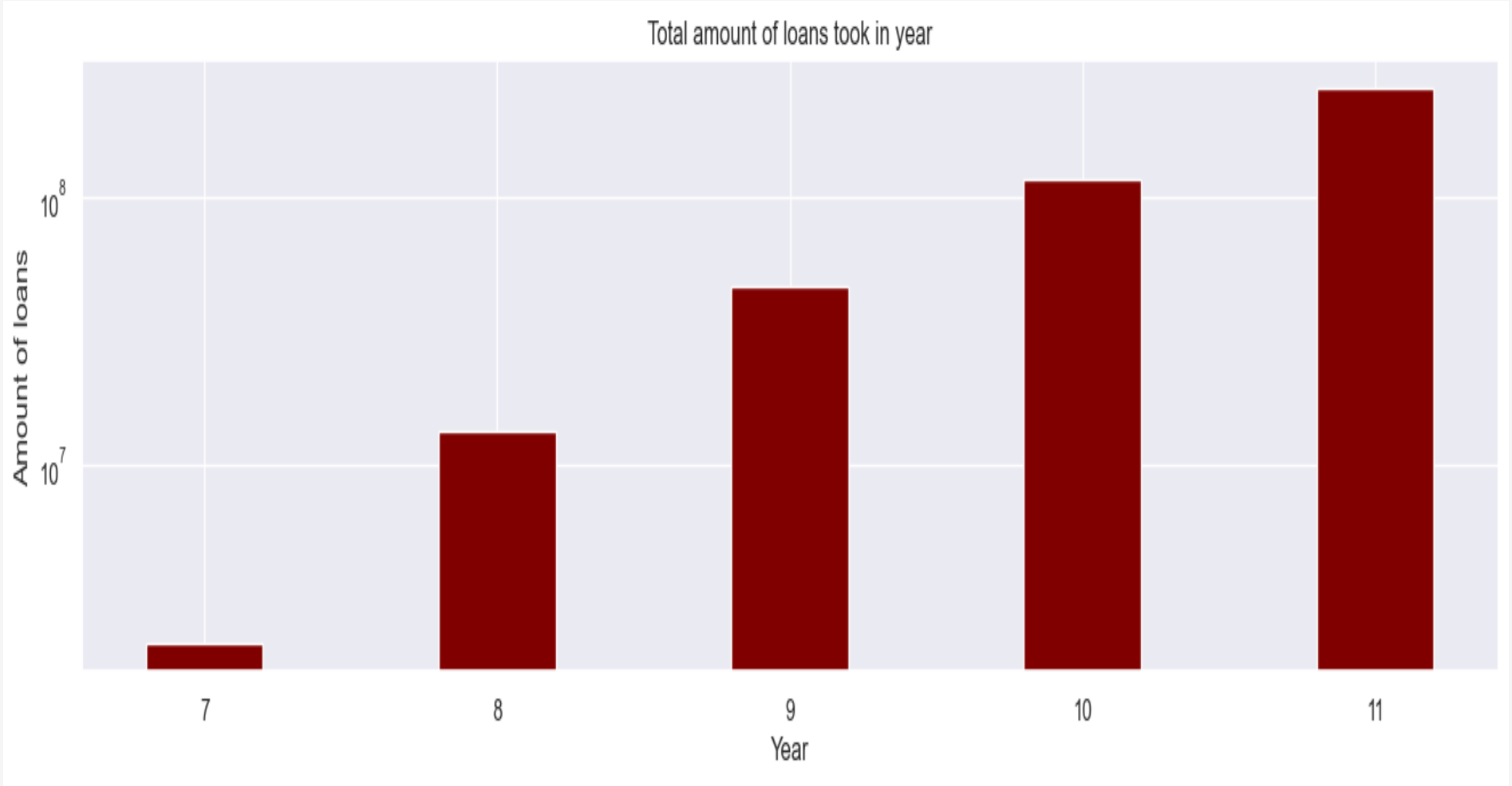
# Number of loans approved in every year

Number of loan counts are increasing yearly

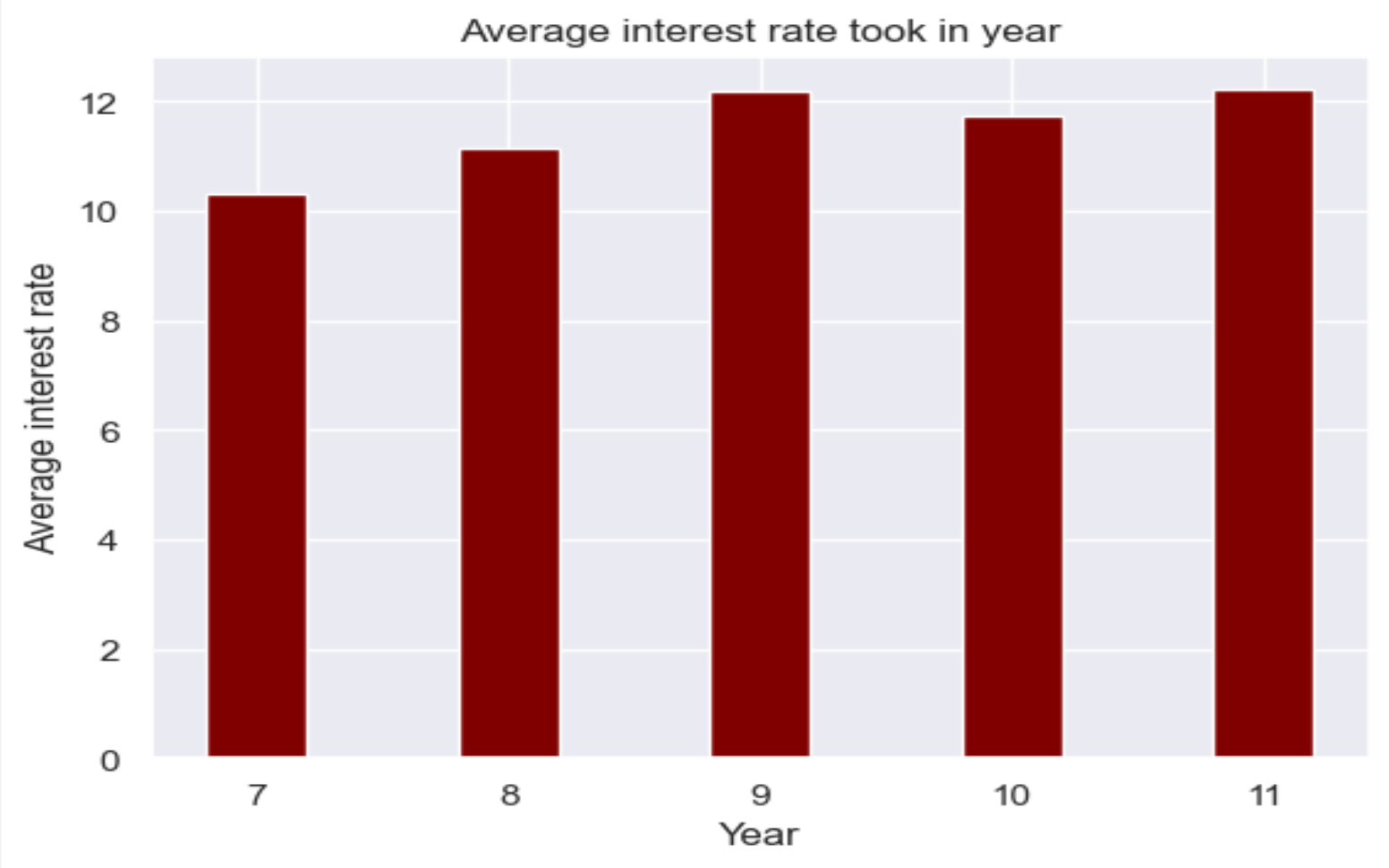


# Total amount approved for loan per year

Total number of loan amount are increasing yearly

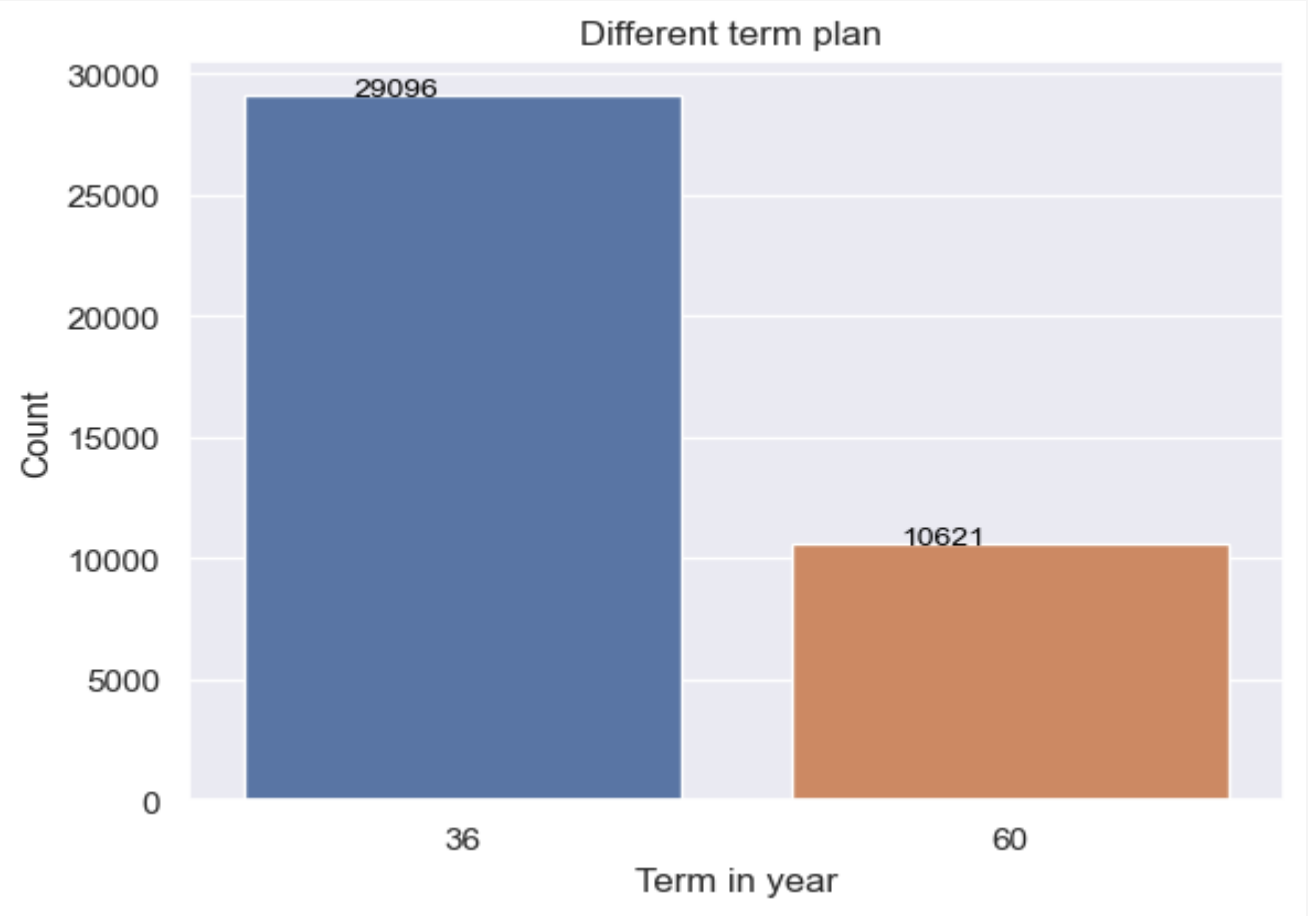


# Average interest rate

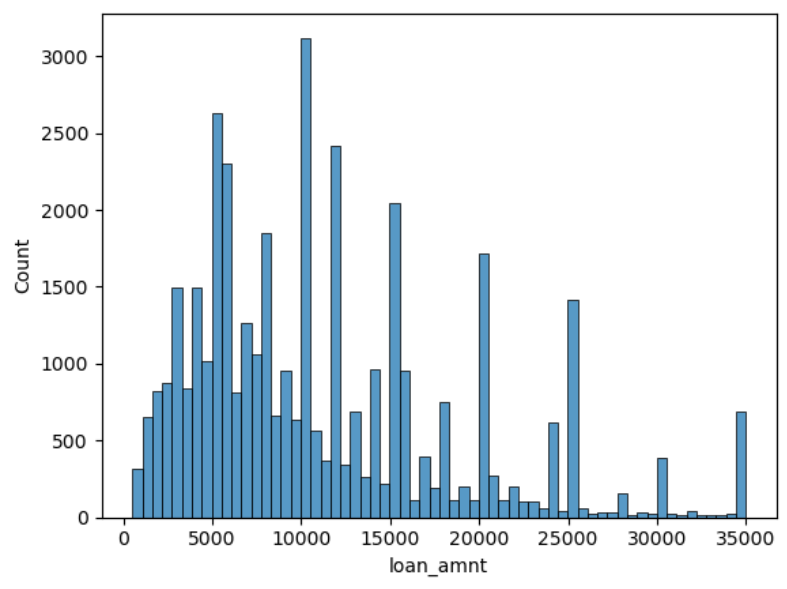


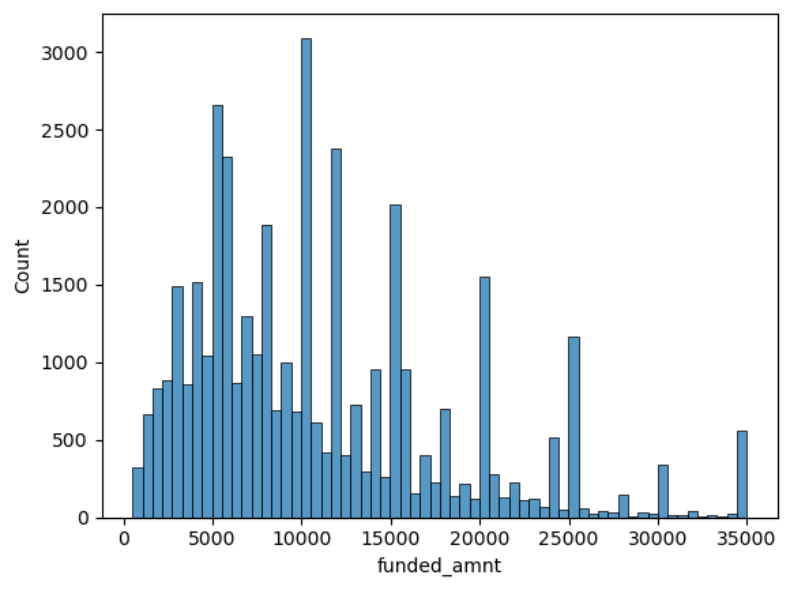
# Different term plan

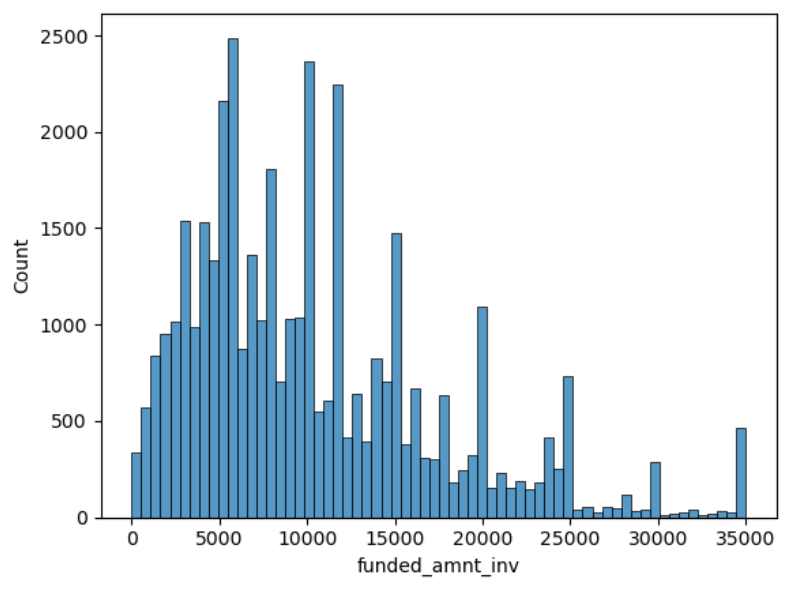
Many customers taking 36 term plan



# Histogram for Loan amount, Funded amount, Funded investment amount

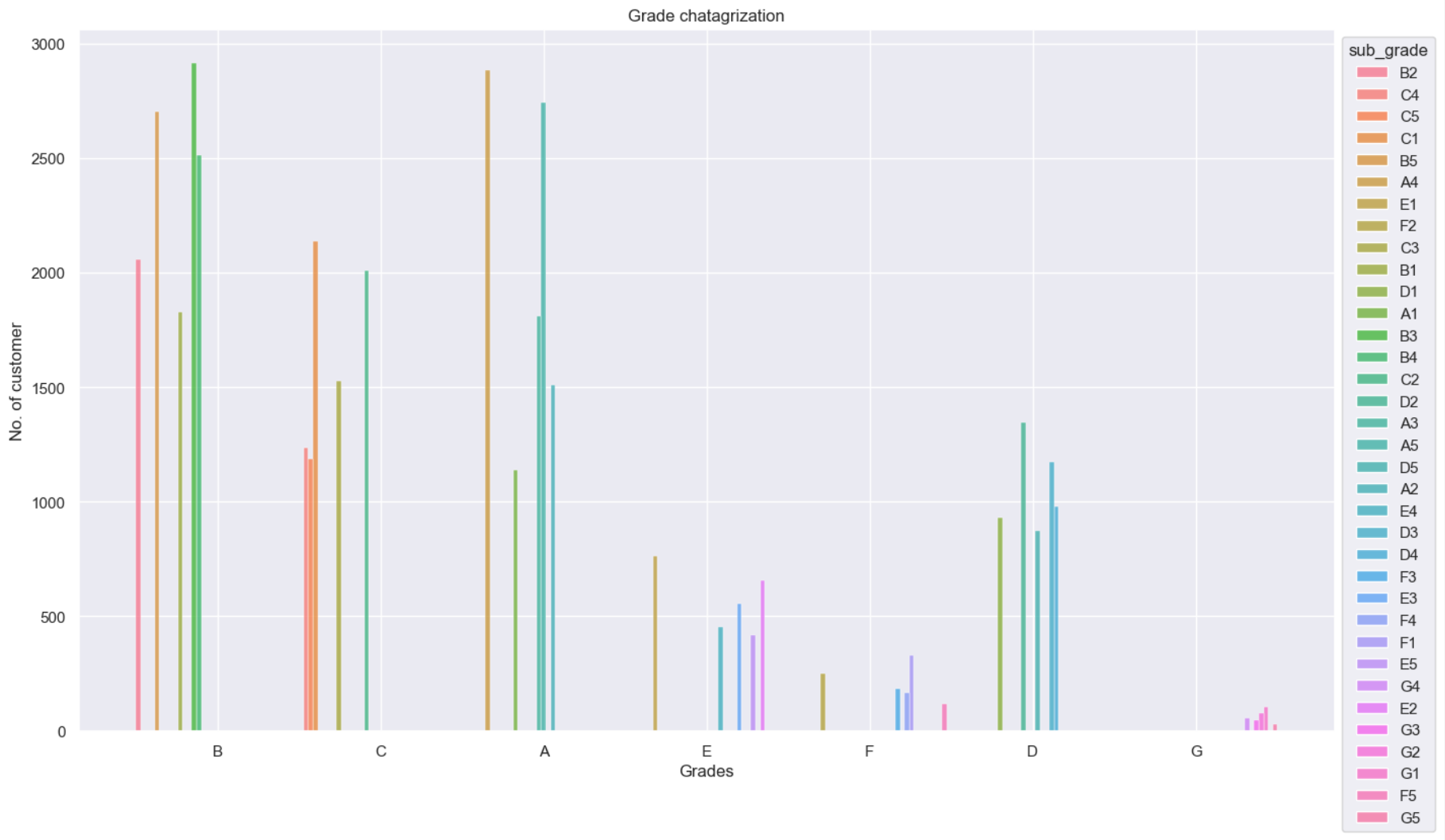


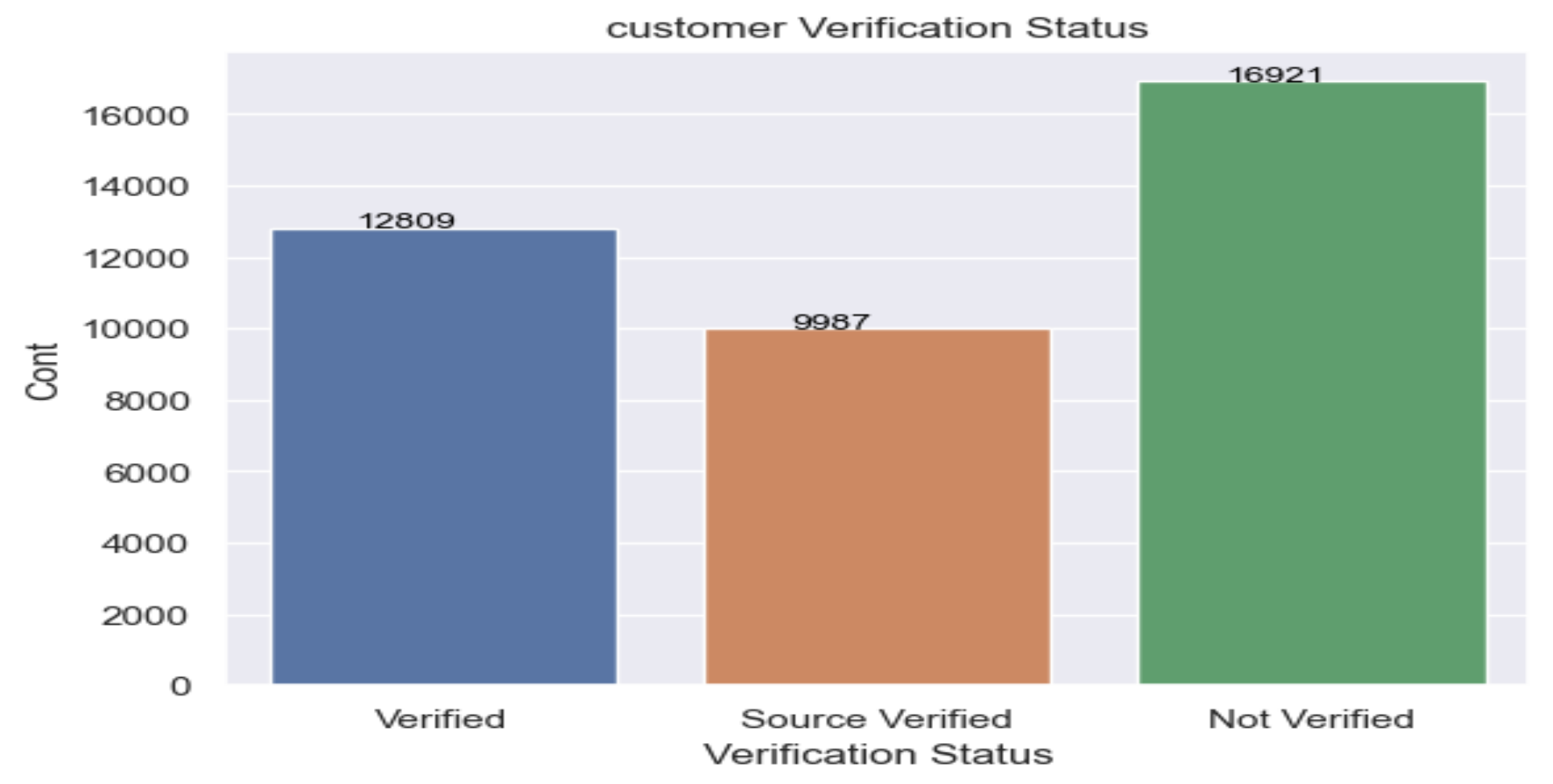
# Grade categorization

For observing how many customer in which categories belong

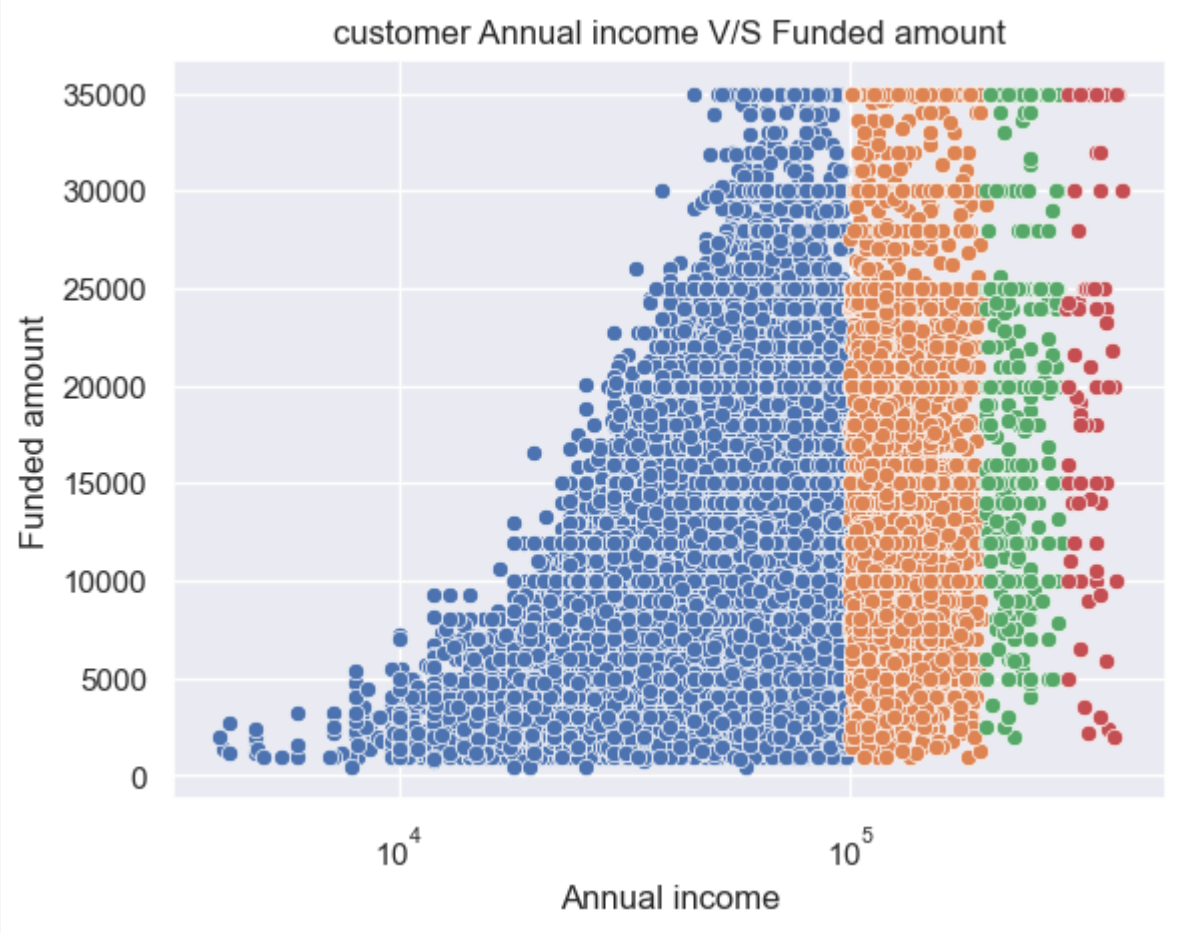


# Customer working year

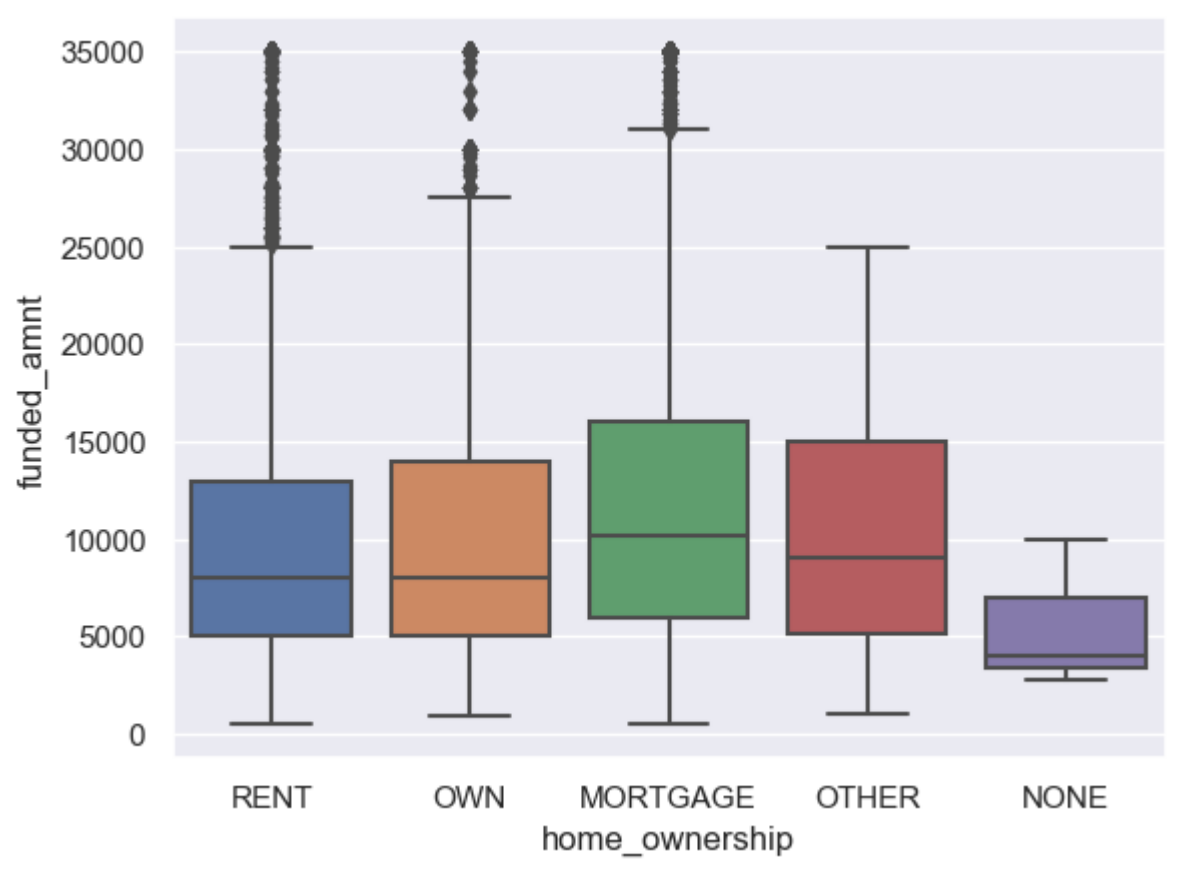
# Customer Verification Status



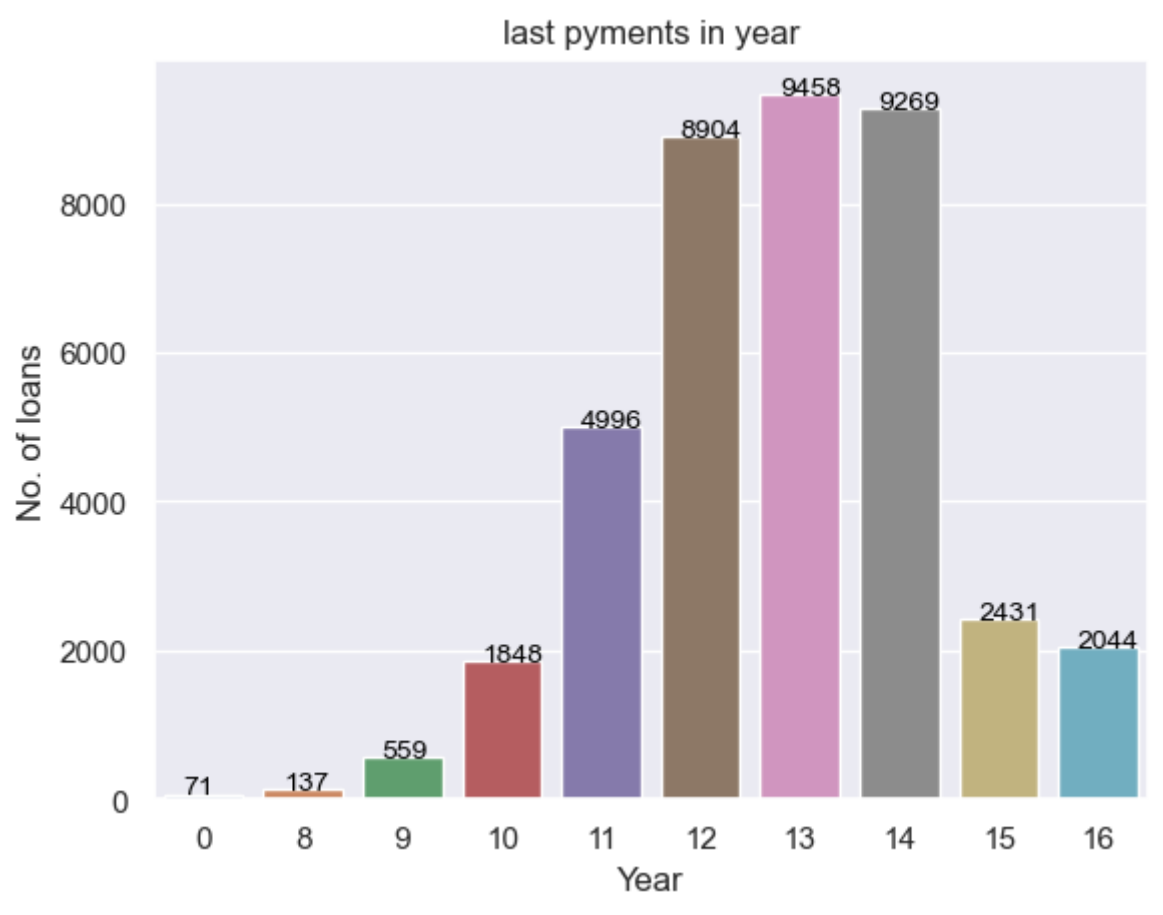
# Customer Annual income V/S funded amount in scatter plot



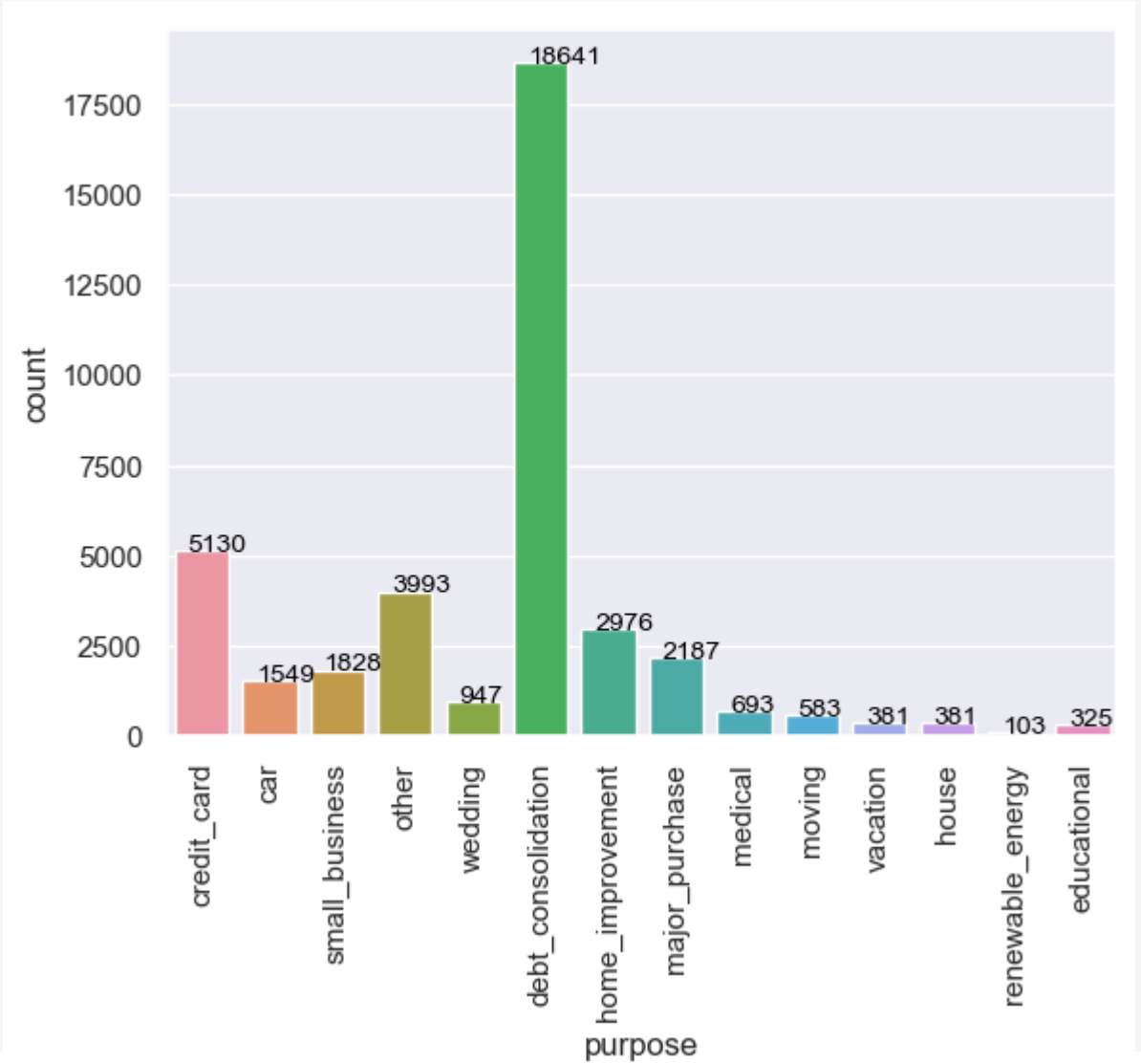
# Funded amount with respect to home ownership



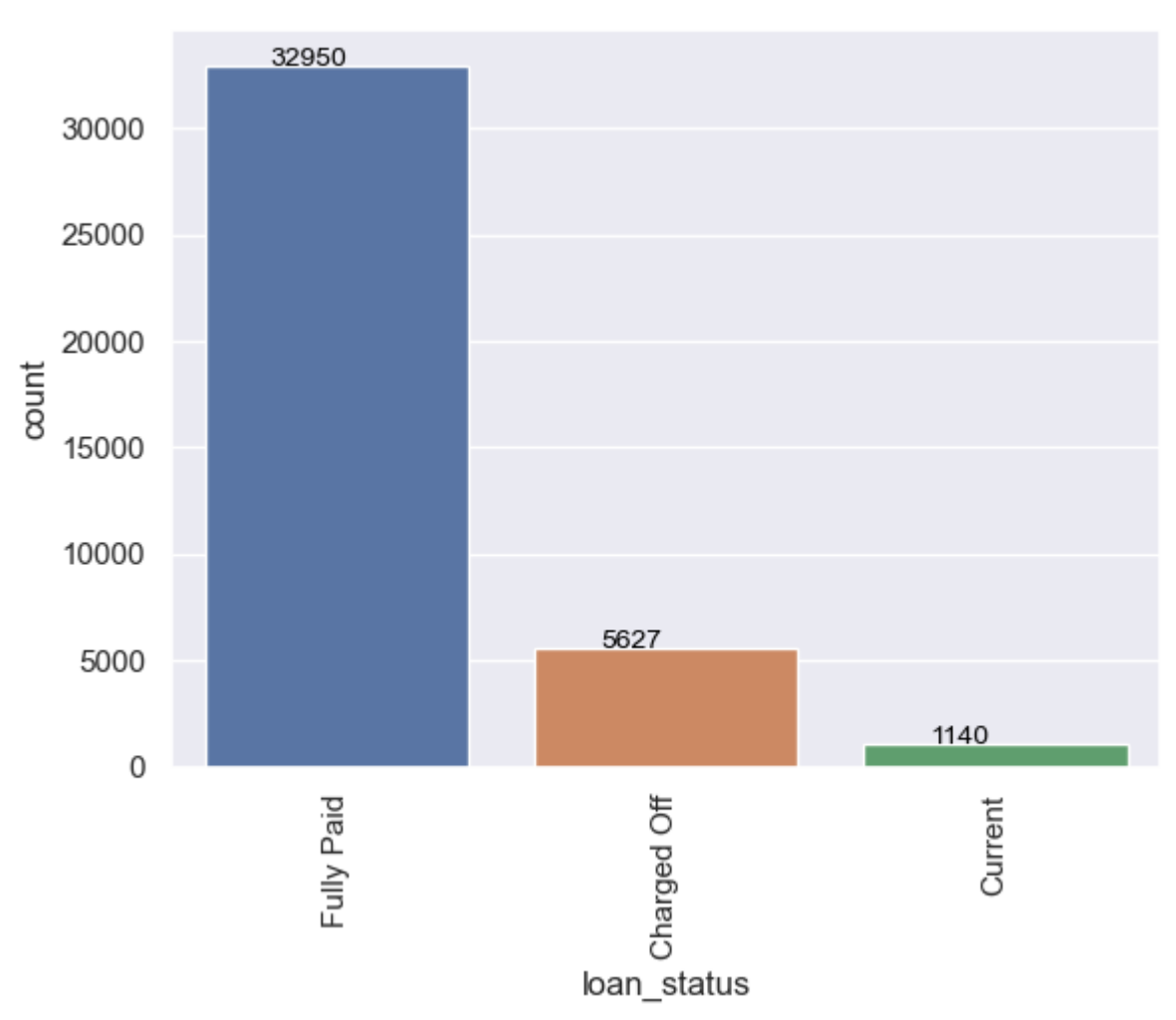
# Last payments year



# Purpose of loan



# Loan status


32950
5627
1140
30000
25000
20000
15000
10000
5000
0
count
Fully Paid
Charged Off
Current
loan_status

# Below the data analysis outcomes

1. Higher A grade customer have good record of re-paying the load, only < 10% customers defaulted the load.
2. Higher non re-payment rate is found in higher interest rate loans.
3. Both Higher and lower income customer loan re-payment commitment is at similar proportion
4. More loan defaulters are seen in the lower loan amount groups.
5. More loan defaulters are seen in verified customer category.

# Recommendations

1. Higher interest rate loans needs to monitored. Close engagement with the customer is needed.
2. More loan defaulters are seen in verified customer category, so verification process needs to be strengthened

Thank you