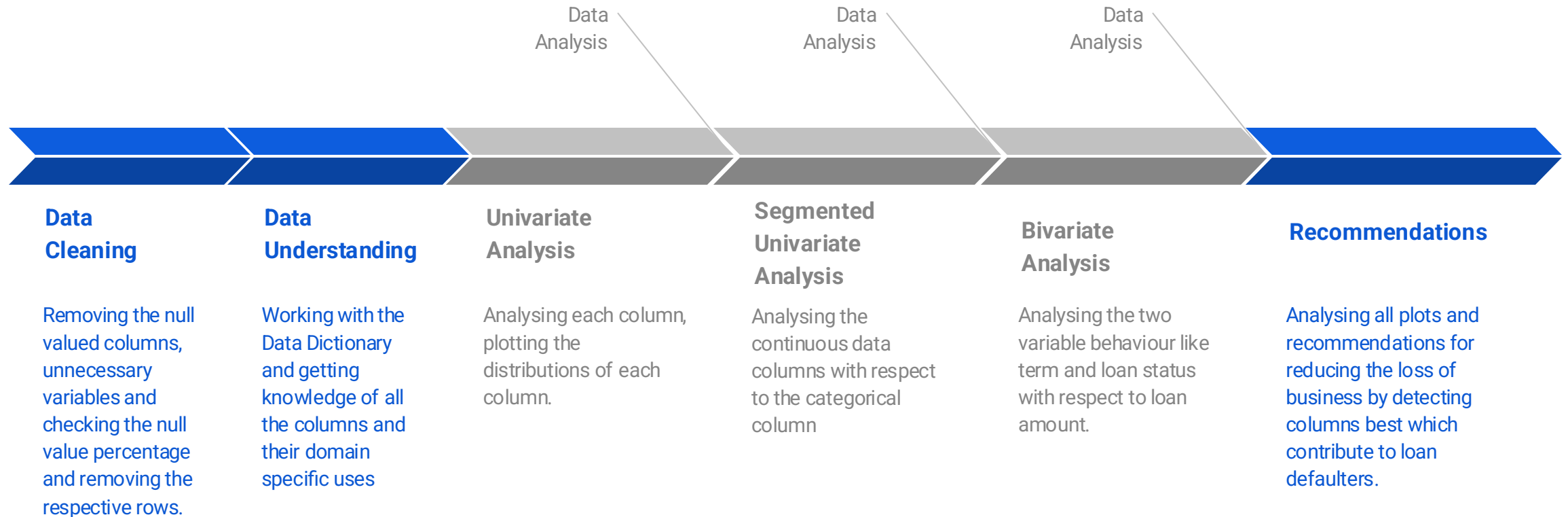# Lending Club Case Study

Group Members:
Ashutosh Kumar Chandel
Ganesh Kumbhar

Abstract

- Lending club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

- Borrowers can easily access lower interest rate loans through a fast online interface.

- The objective of analysis is to use the information about past loan applicants and find whether they 'defaulted' or not.

Problem solving methodology

Data Analysis

Data Analysis

Data Analysis

Data Cleaning

Removing the null valued columns, unnecessary variables and checking the null value percentage and removing the respective rows.

Data Understanding

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Univariate Analysis

Analysing each column, plotting the distributions of each column.

Segmented Univariate Analysis

Analysing the continuous data columns with respect to the categorical column

Bivariate Analysis

Analysing the two variable behaviour like term and loan status with respect to loan amount.

Recommendations

Analysing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.

Data Cleaning

Removing Rows

- Removing Rows which is having loan status as current. Because this will not contribute for the analysis.

Removing Columns

- Dropping columns which has unique identity in nature. They don't contribute anything to loan analysis.
- Dropping text/description columns which won't contribute to overall analysis.
- Behaviour data of the customers are captured post the loan approval, so it does not make any sense to keep.
- Dropping all columns whose all the values are NA.
- Dropping all columns with all zero values
-

Data Understanding

- Univariate Analysis of Quantitative variables
- Unordered Categorical Variable analysis.

- Ordered Categorical Variable Analysis

- Derived Variable Analysis

Univariate Analysis of Quantitative variables

Loan Amount
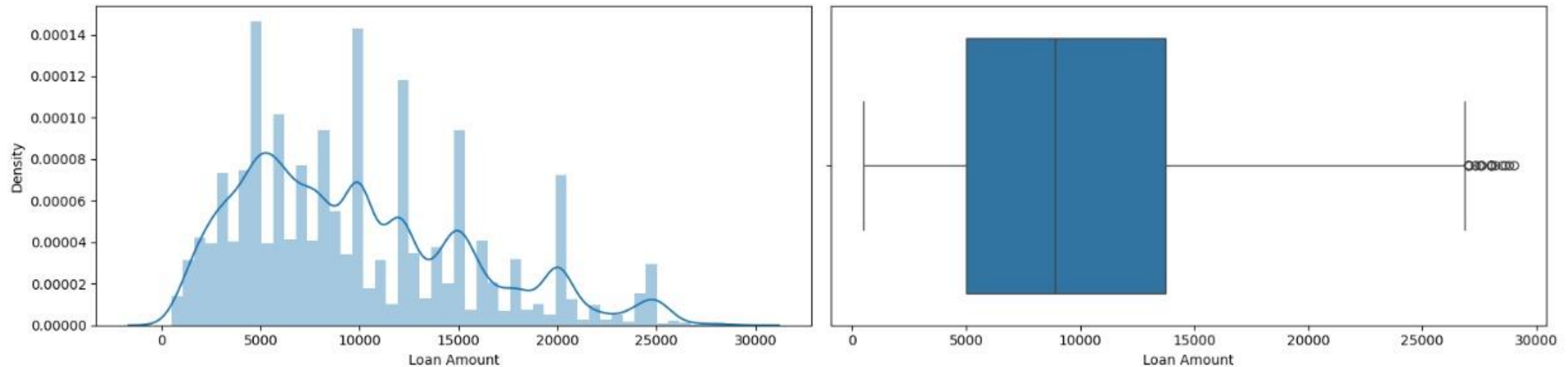
- Majority of the loan amount is in the range of 5k to 14k
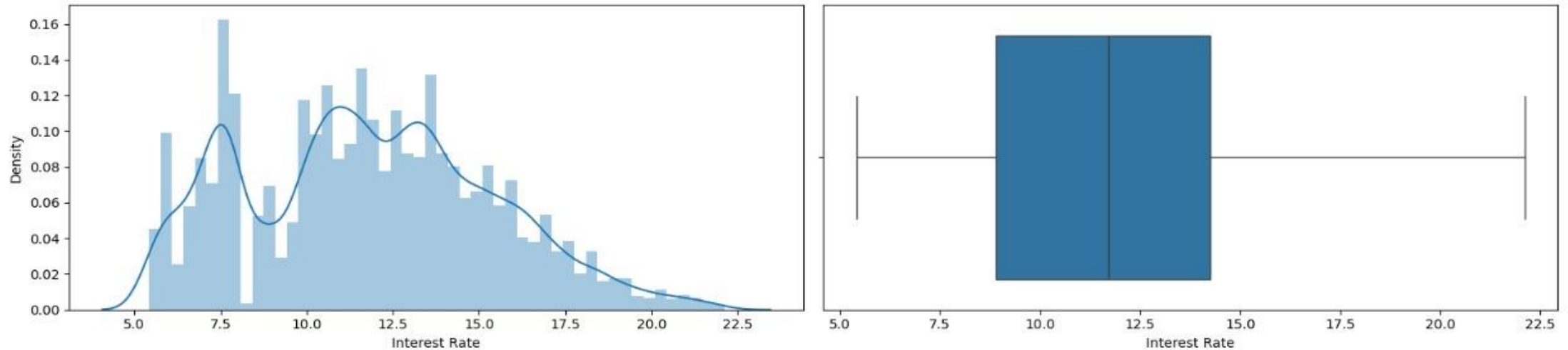
Univariate Analysis of Quantitative variables

Funded amount
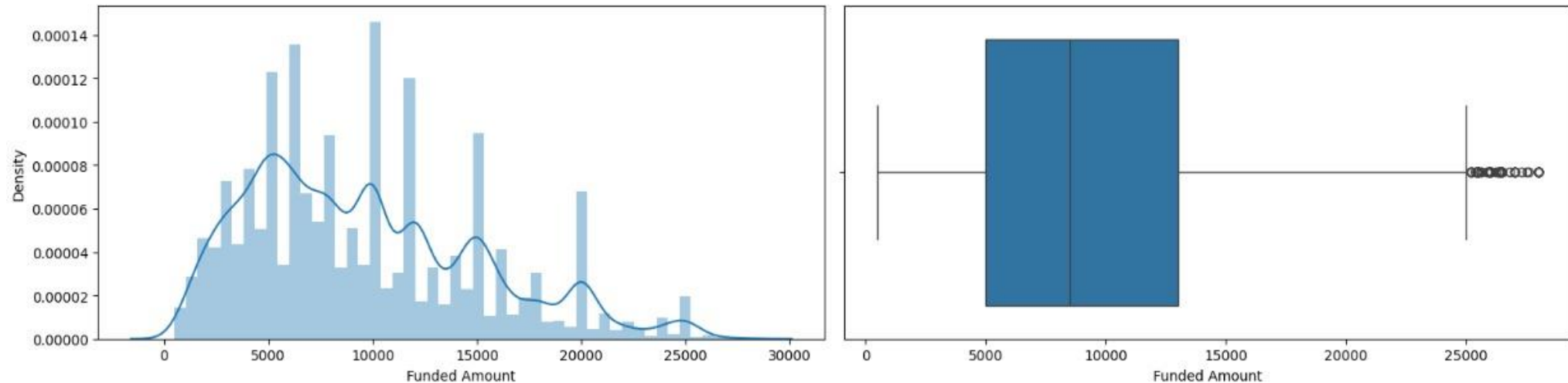
- Majority of the funded amount (Funded_amnt) is in the range of 5k to 12k

Univariate Analysis of Quantitative variables

Interest Rate

- Majority of the interest rate is in the range of 5% to 16% going max at 22%
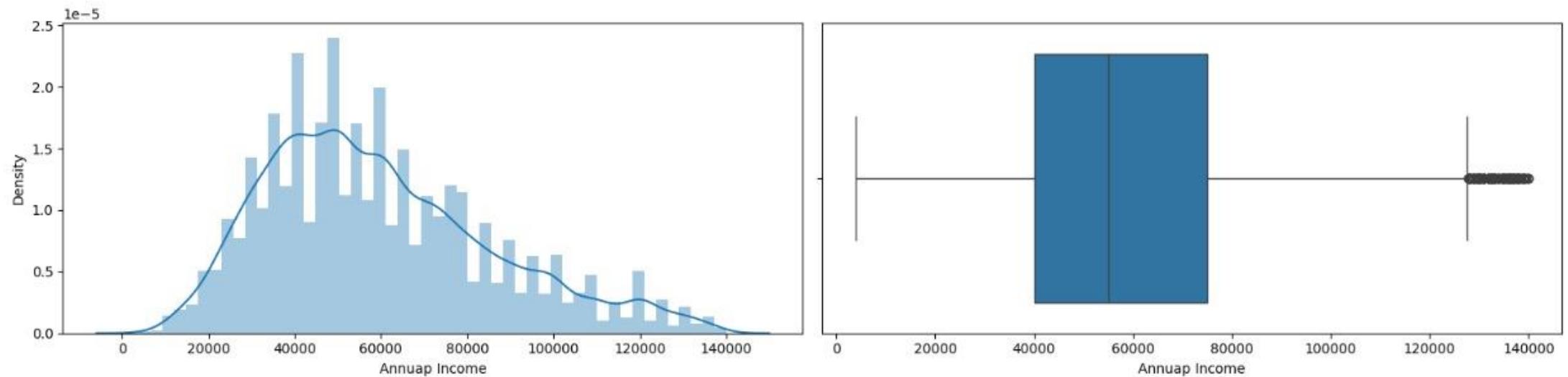
Univariate Analysis of Quantitative variables

Installments
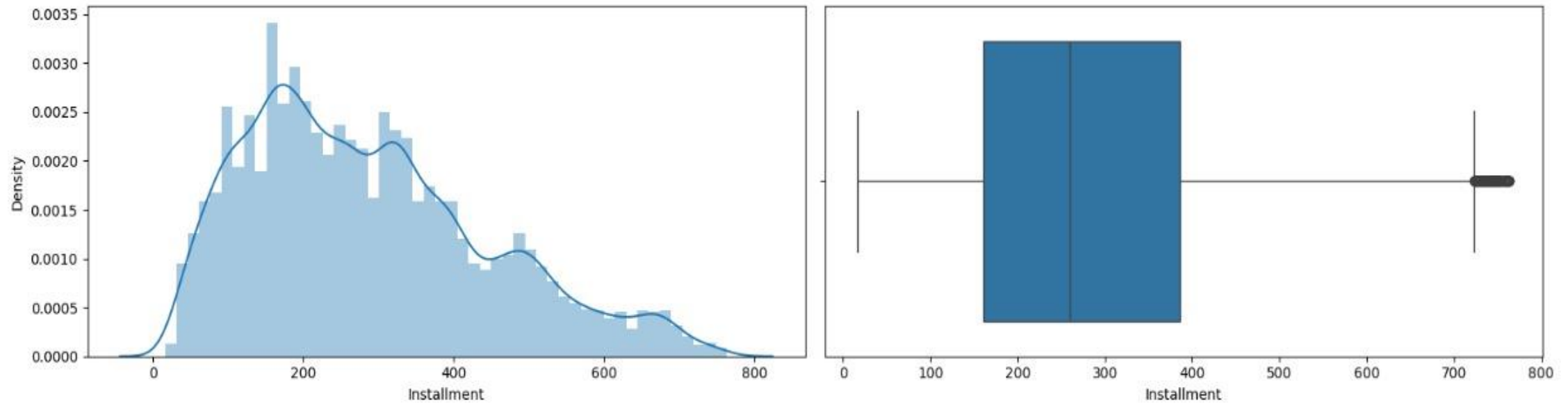
- Majority of the installment is in the range of 20 to 400 going at the max to 700.
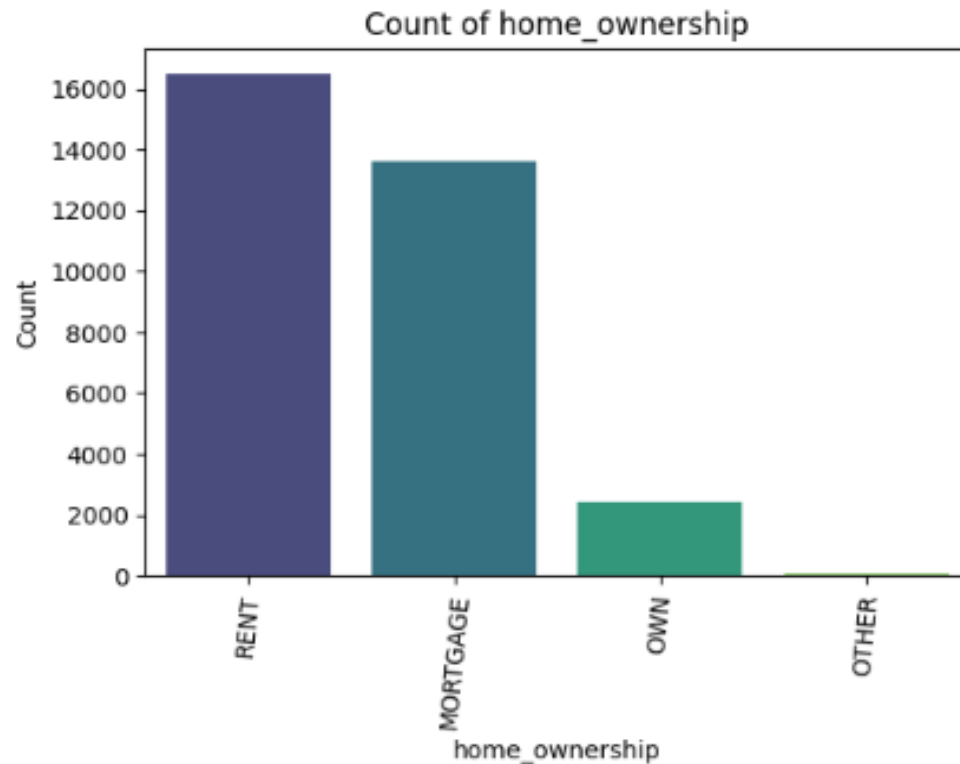
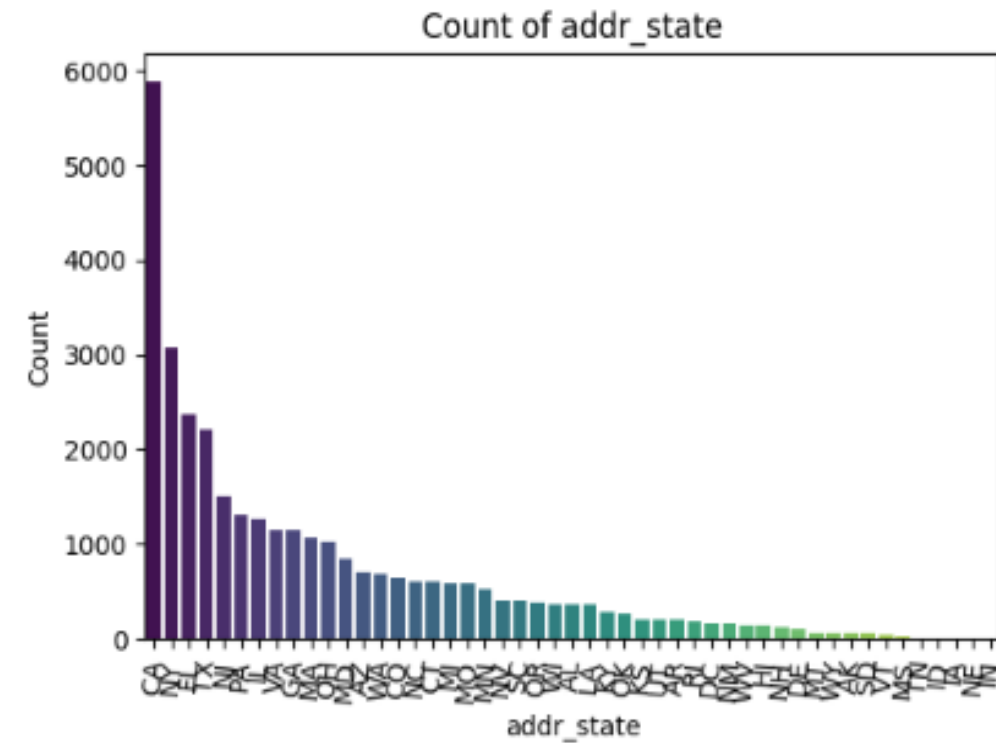# Univariate Analysis of Quantitative variables

Annual Income

- Majority of the annual income is in the range of 4k to 40k going max at 120k.
  This column need outliers treatment.
  probable cause of outliers is many customers fake their annual income to get more loan.

Unordered Categorical Variable Analysis
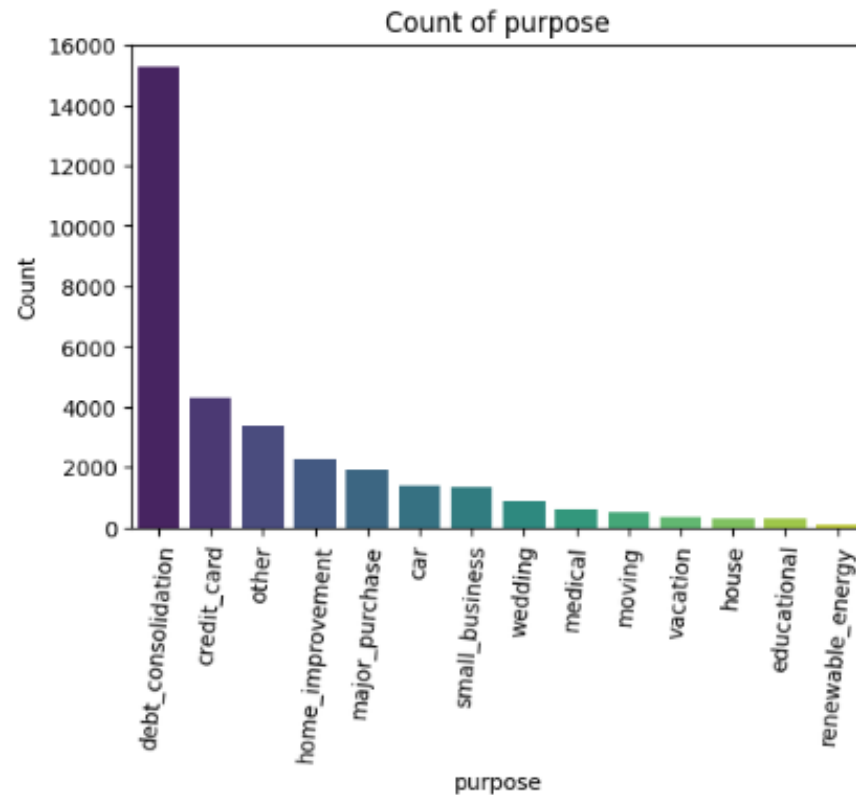
Home Ownership status

Count of Address state

- Majority of the loan takers are stay in rentals.

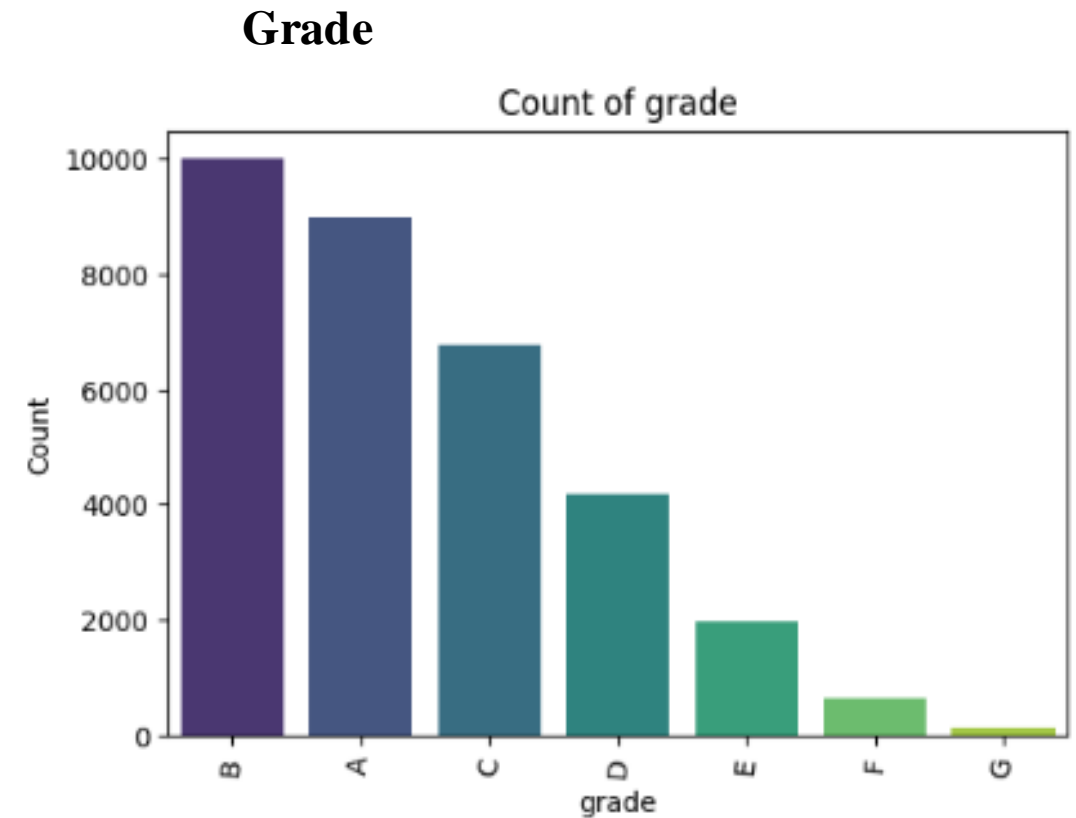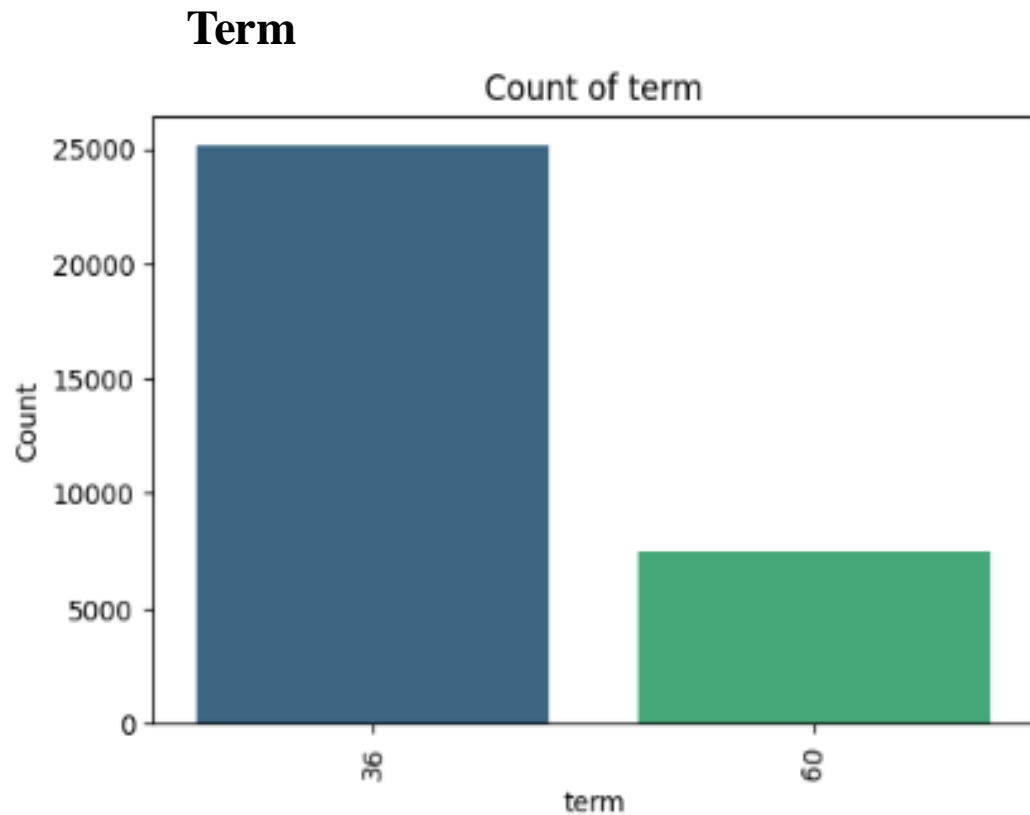- Majority of the loan takers are from CA state.

Unordered Categorical Variable Analysis

Purpose of the loan

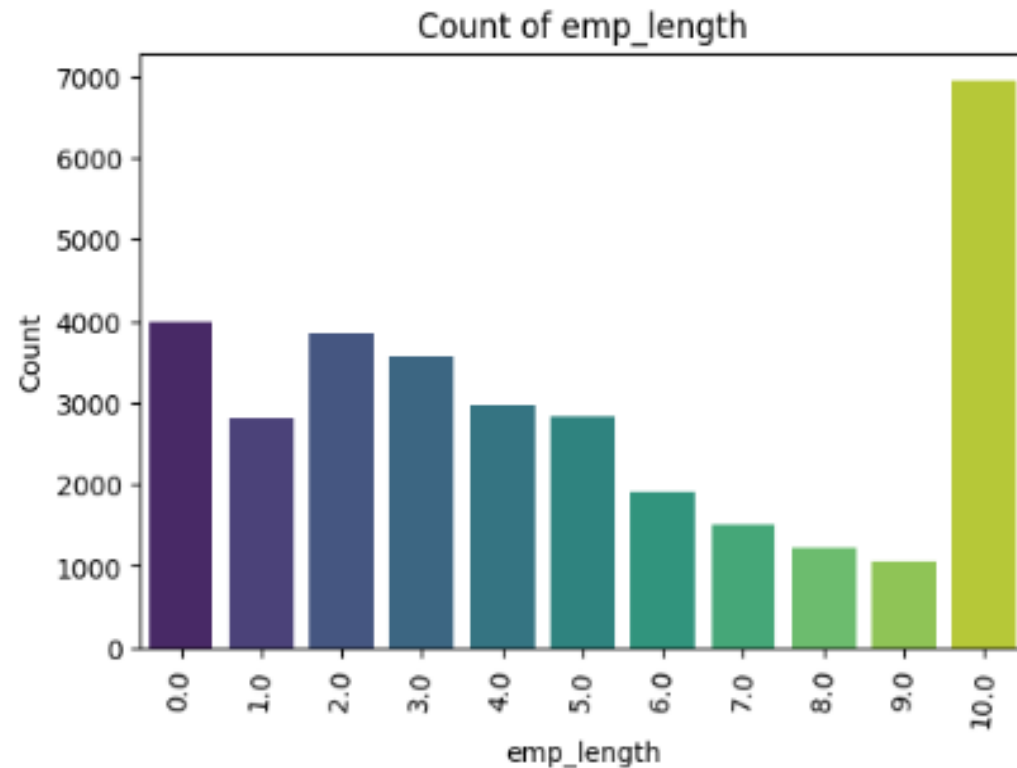- Majority of the loan amount taken for debt consolidation.

Ordered Categorical Variable Analysis
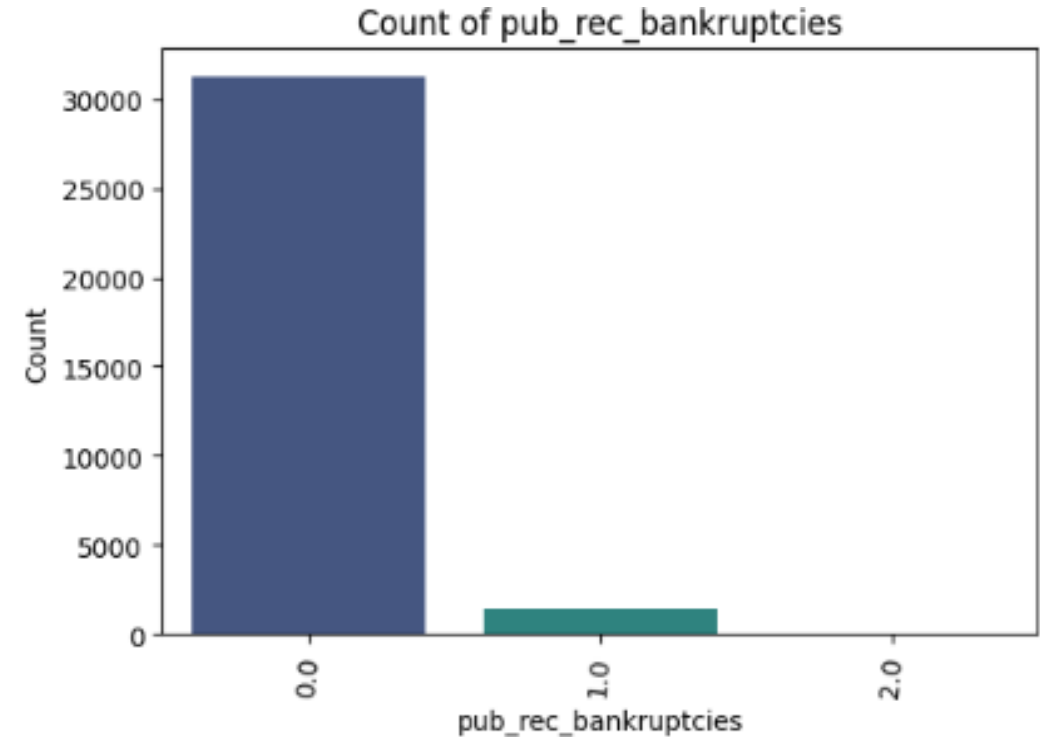
Term

Grade

- Majority of the loan application counts are in the term of 36 months.
- Majority of the loan application counts fall under the category of Grade B

Ordered Categorical Variable Analysis
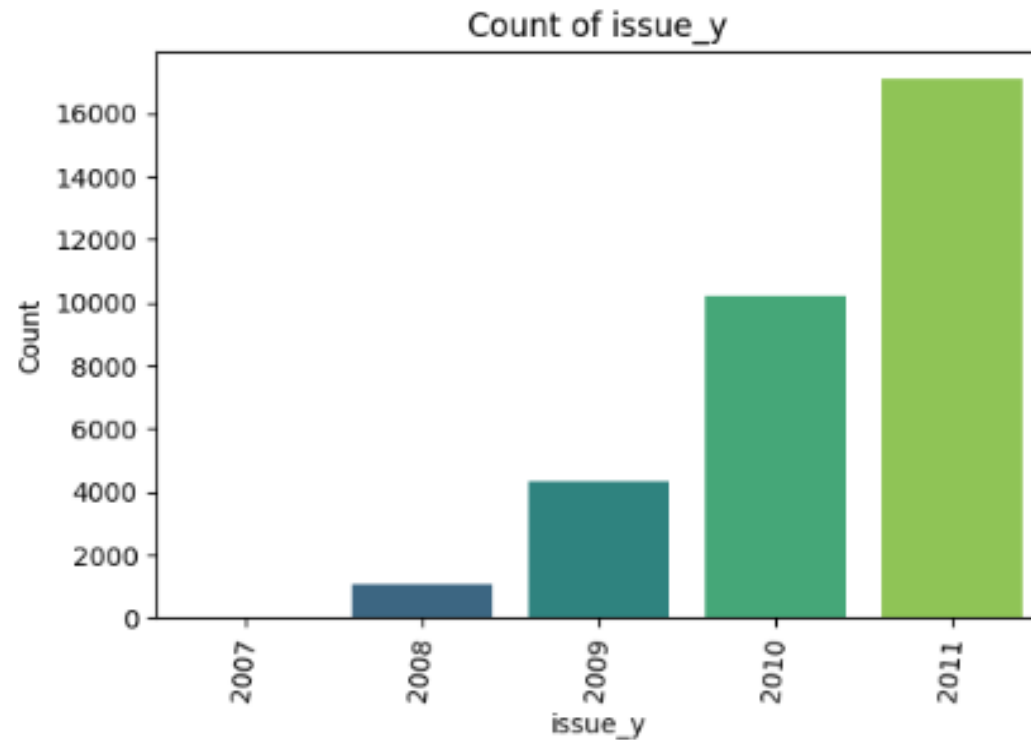
Employment length
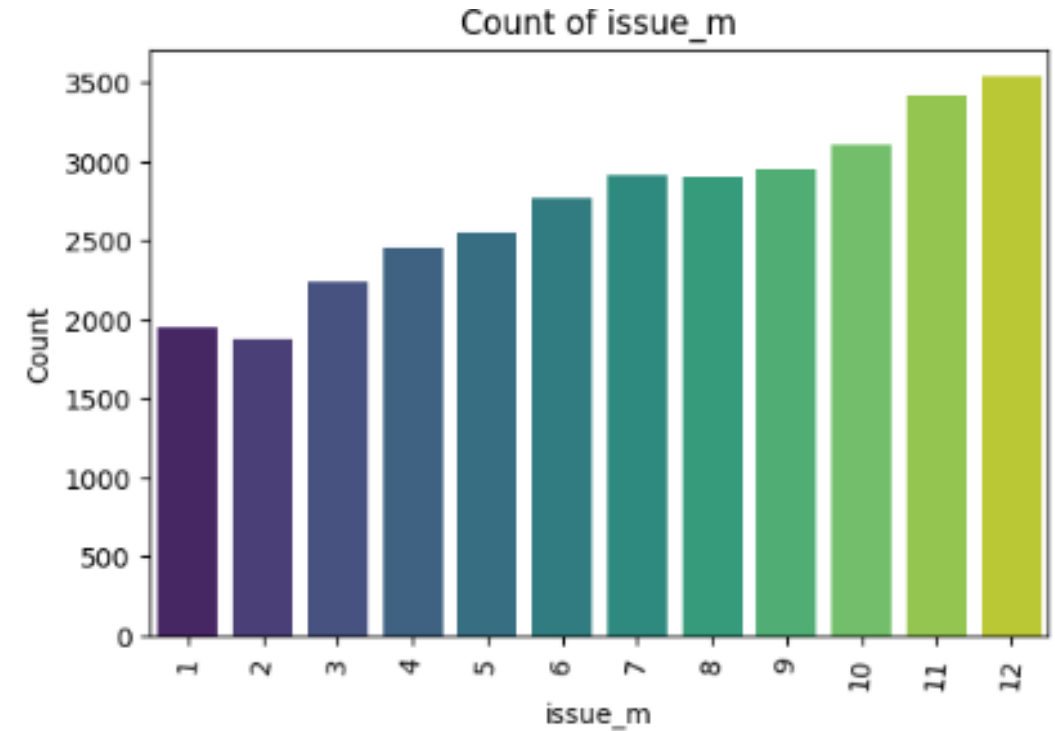

Public record of bankruptcies

- Majority of the employment length of the customers are 10+ years.
- Majority of the loan applicants don't have public record of bankruptcies.
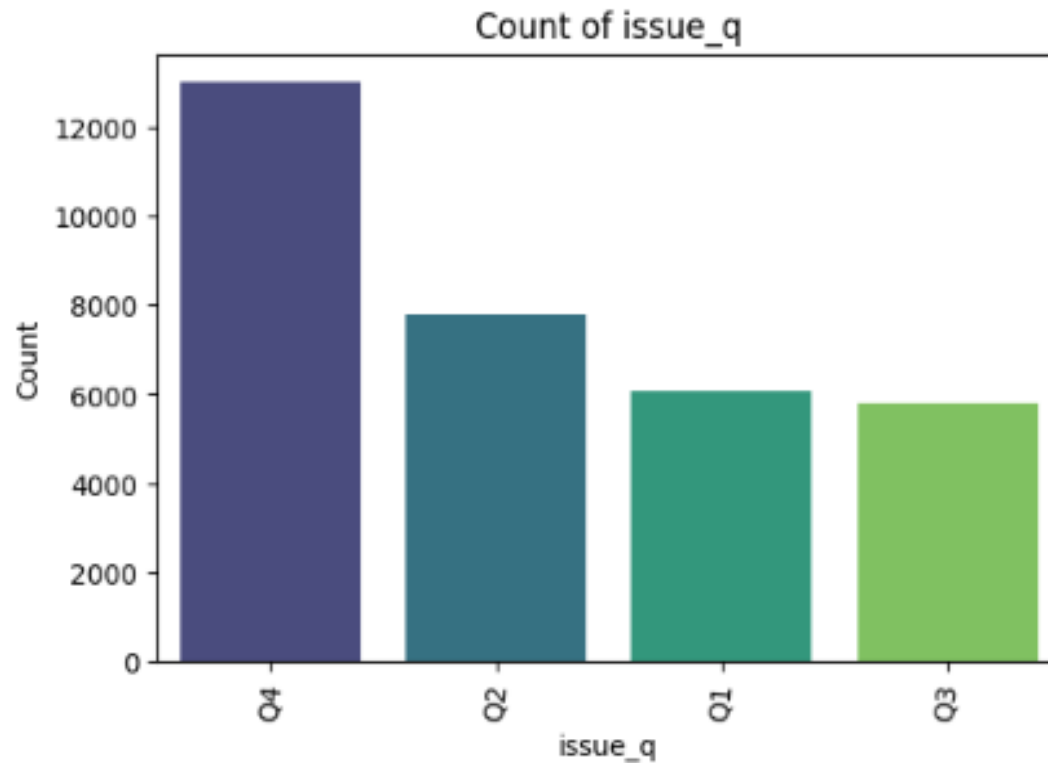
Derived Variable Analysis

Issue Year

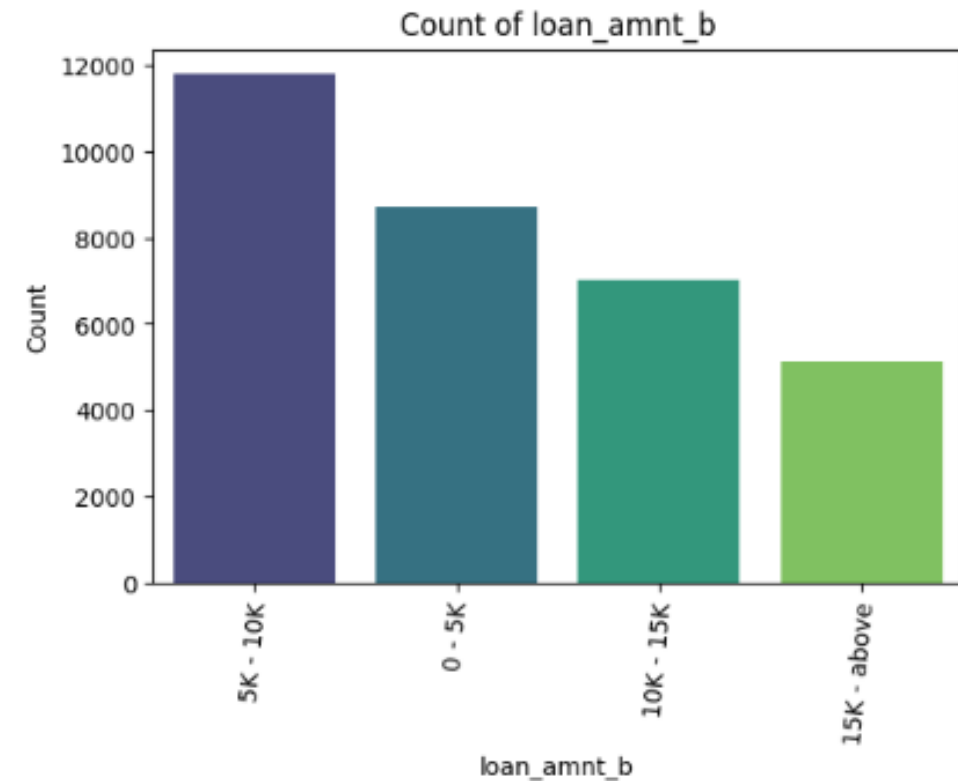

Count of issue_y

Issue Month

Count of issue_m

- Year wise loan applicants count increasing.

- Loan applications are increasing at the yearend.

Derived Variable Analysis

Issue quarter

Count of issue_q

Loan amount
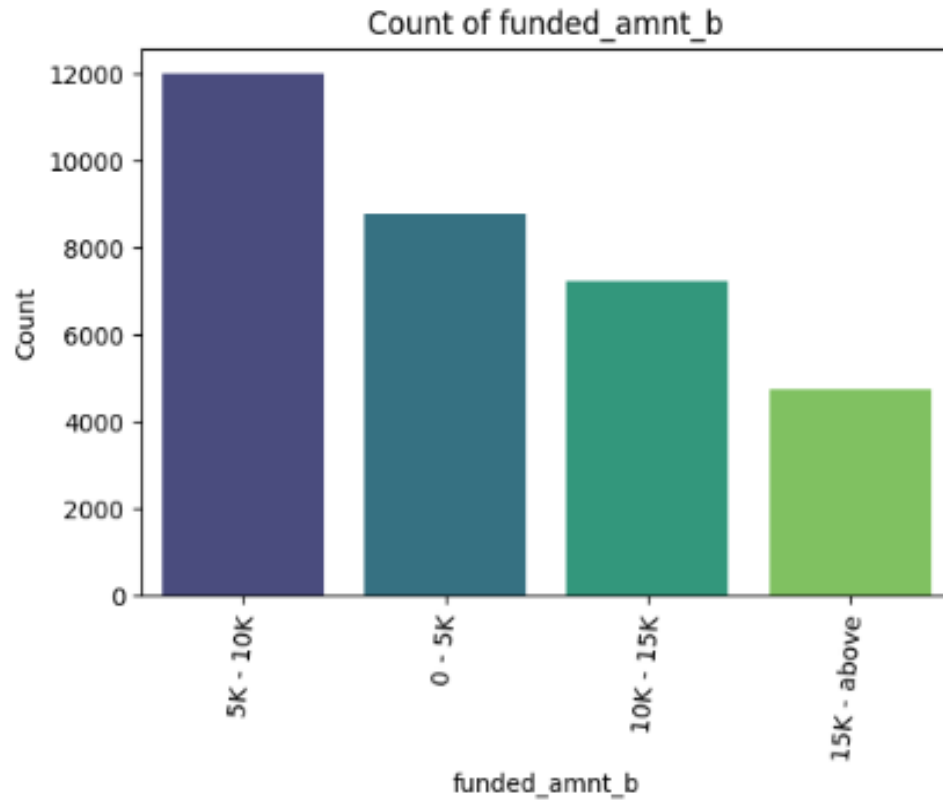
Count of loan_amnt_b

- Highest loan issued in the quarter- 4

- Highest loan amount applications fall in the range of 5k to 10k

Derived Variable Analysis

Funded amount

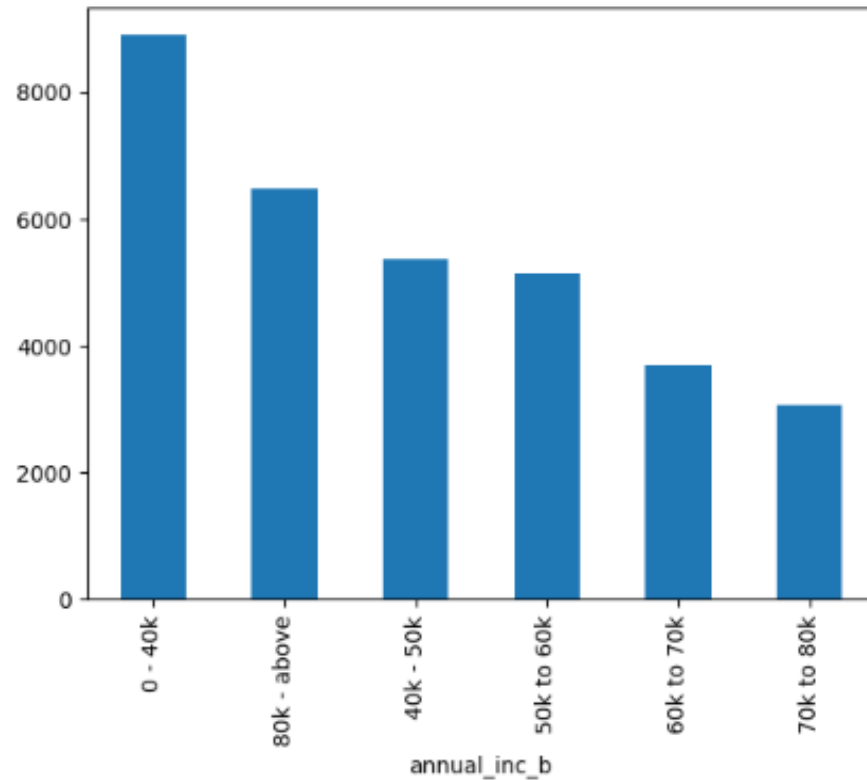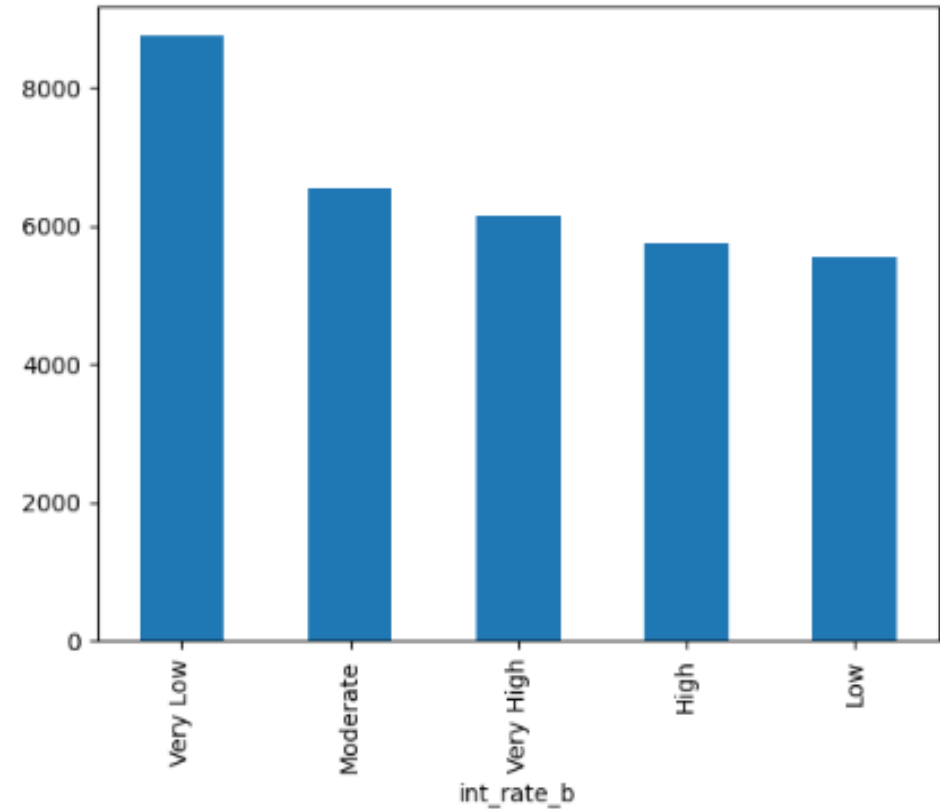**Loan amount**
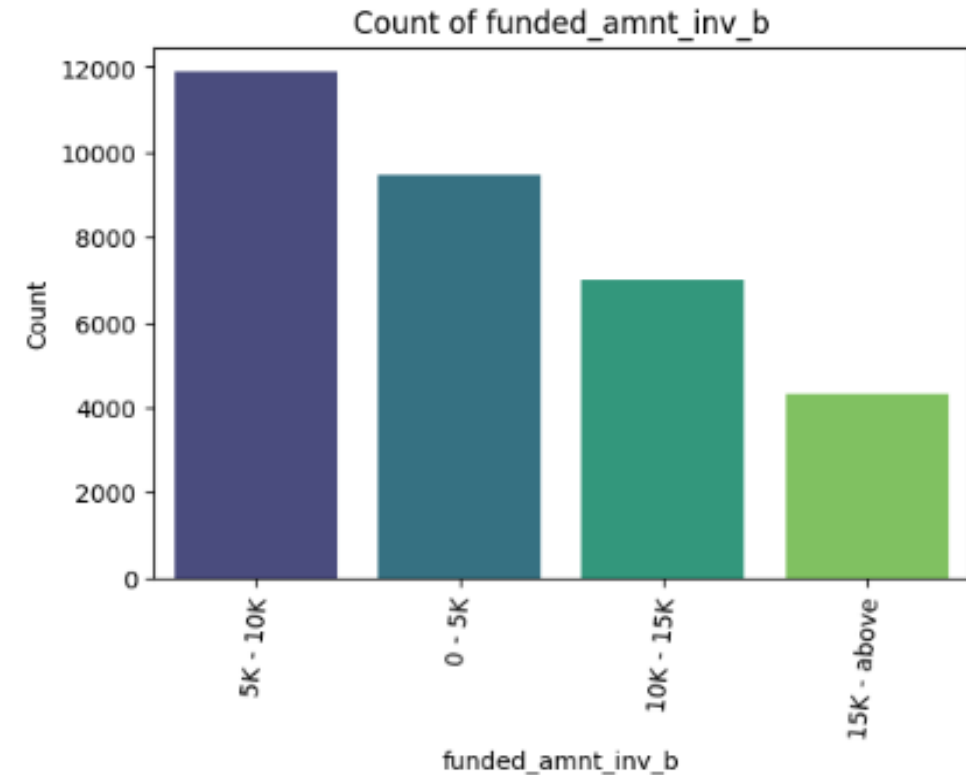
- Highest funded amount applications fall in the range of 5k to 10k
- Highest loan amount applications fall in the range of 5k to 10k

Derived Variable Analysis

Annual Income Category

Interest Rate Category

- Majority of the loan applicants are in the range of 0 - 40K annual income
- Majority of the loan applicants are in the category of Very Low interest rates

Univariate Analysis Summary

1. **Customer Demographics**

   - Majority of the loan applicants are in the range of 0 - 40K annual income
   - Majority of the debt to income is in the range of 0 to 20 going at the max to 30
   - Majority of the home owner status are in status of RENT and MORTGAGE
   - Highest loan applications are in the category of debt consolidation
   - CA (California) state has the maximum amount of loan applications
   - Majority of the loan applicants are in the category of not having an public record of bankruptcies
   - Majority of the employment length of the customers are 10+ years and then in the range of 0-2 years

2. **Loan Demographics**

   - Highest loan amount applications fall in the range of 5k to 10k
   - Majority of the interest rate is in the range of 5% to 16% going at the max to 22%
   - Majority of the installment amount is in the range of $20 to $400 to $400
   - Majority of the loan applications counts are in the term of 36 months
   - Majority of loan application counts fall under the category of Grade B

Univariate Analysis Summary

3.Time Based Analysis

- Loan application counts are increasing year over year
- Highest loan application volume in Quarter 4 of every year
- Lowest loan applications are in Q1
  - Possibly because by year ends people face the financial challenges
  - Possibly because of festive seasons
  - Possibly because they are consolidating debt by year end

Inferences

- The customer demographic data shows which segment of customers to target for highest volume of loan
- Indicates more analysis is needed why other categories are not as high as other few
- Indicates the Lending Club to be prepared with volume in Q4
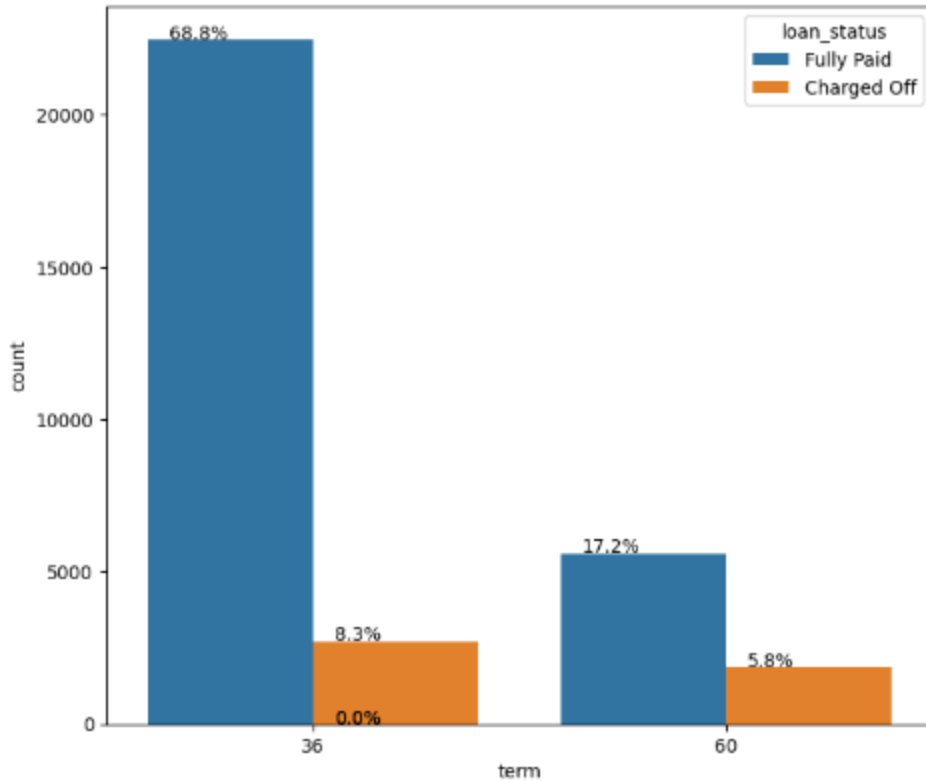- Indicates the Lending Club to target customers in other quarters to increase sales

Bivariate Analysis

Term vs charged off ratio
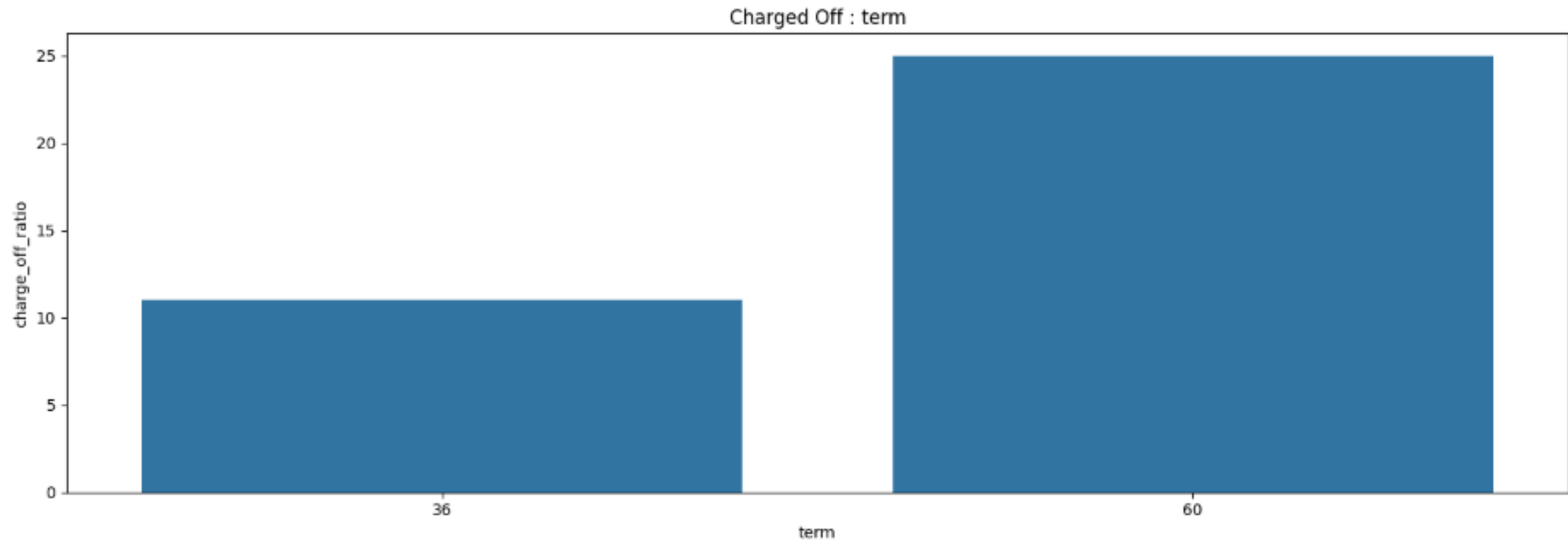
Bivariate Analysis

Term v/s charged off percentage

loan_status	term	Charged Off	Fully Paid	total	charge_off_ratio
0	36	2715	22452	25167	11.0
1	60	1880	5608	7488	25.0

- The volume of loans are in the category of term = 36
- The overall percentage of Charge Off's is slightly higher in term = 36 (8%) as compared to term=60 (6%)
- If we calculate the ratio of Charge Off's within a category
- **Charge Offs** ratio is for the term=60 is 25% which is much higher than term=36 (10%)
- **term=60 is the loan applications which require more scrutiny**
- **Inferences**
- Most of the applicants with term=60 potentially will have high Charge Offs

Bivariate Analysis

Grade v/s Charged of counts
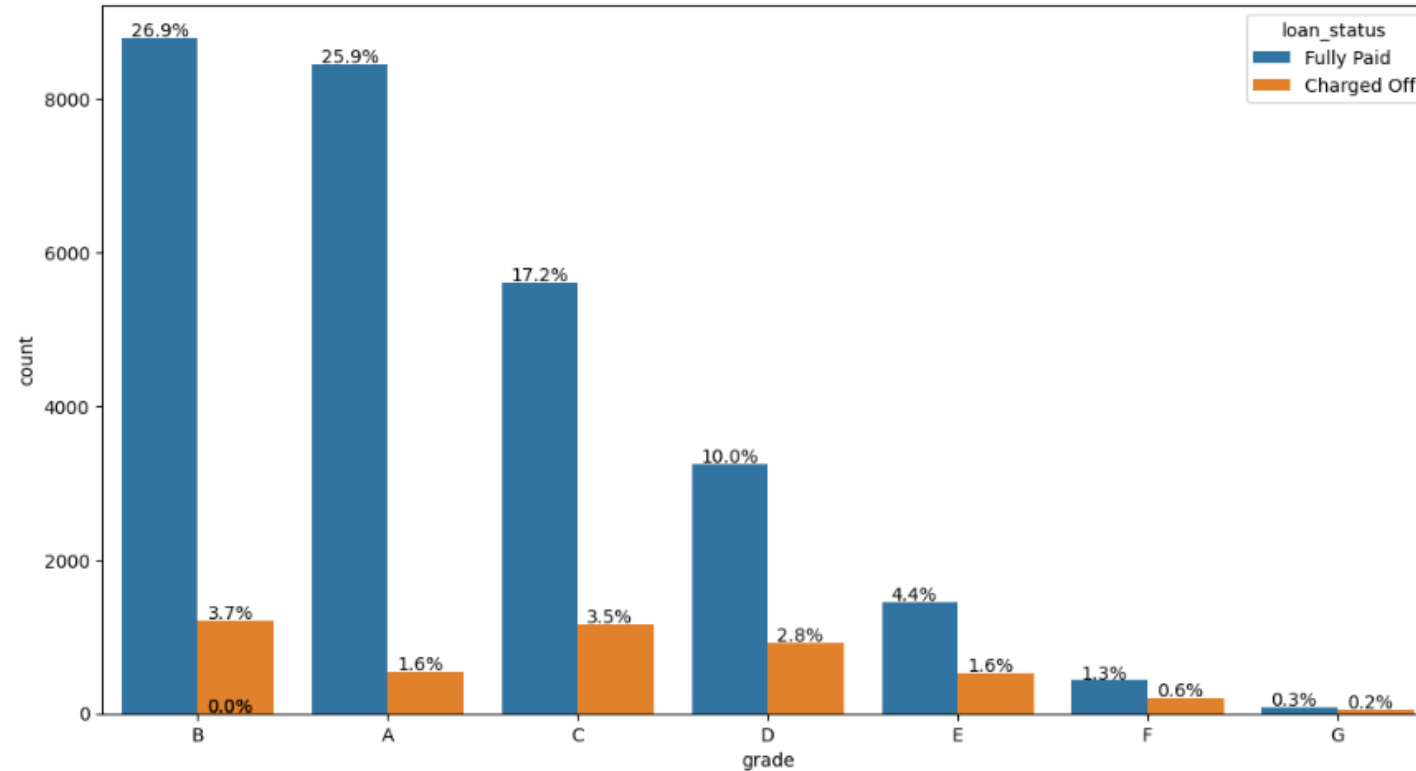
loan_status	grade	Charged Off	Fully Paid	total	charge_off_ratio
0	A	537	8447	8984	6.0
1	B	1213	8781	9994	12.0
2	C	1155	5613	6768	17.0
3	D	918	3253	4171	22.0
4	E	519	1447	1966	26.0
5	F	199	433	632	31.0
6	G	54	86	140	39.0

Bivariate Analysis

Grade v/s Charged of counts
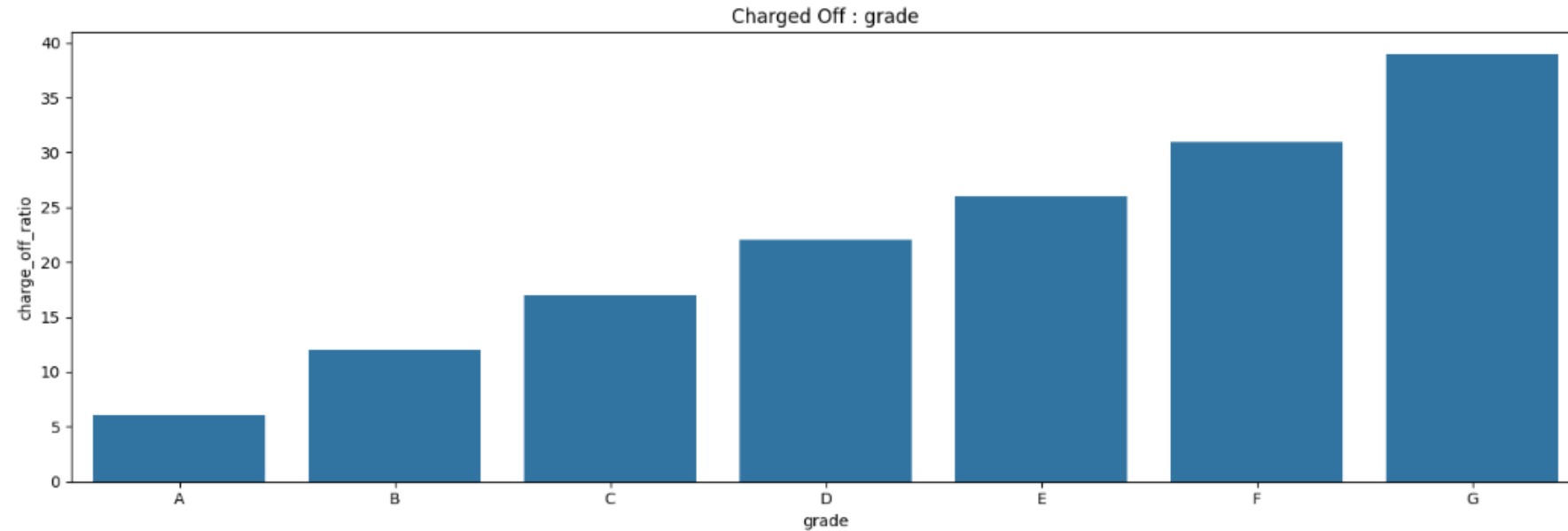
Charged Off : grade

Bivariate Analysis

Grade v/s Charged of counts

- The Majority of *loan volume is in grade=B*
- Highest percentage of overall Charge Offs are in grade B (3.7%) and C(3.6%)
- If we analyse the Charge Off Ratio within a category
  - The highest percentage of **Charge Off**s are in the *grade=G*
  - Highest cluster of **Charge Off**s are in the grades G,F (> 30%)
  - The volume of Grade G is extremely low 158 thus it does not contribute to overall risk significantly
- **Inferences**
  - Highest risk of charge off's are in the grades of B and C
  - Grade "F" and "G" have very high chances of charged off. The columes are low
  - Grade "A" has very less chances of charged off.
  - Probablity of charged off is increasing from "A" to "G"

Recommendations

- Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan default.
- Grades are good metric for detecting defaulters. Lending club should examine more information from borrowers before issuing loans to Low grade (G to A).
- Lending Club should control their number of loan issues to borrowers who are from CA, FL and NY to make profits.
- Small business loans are defaulted more. Lending club should stop/reduce issuing the loans to them.
- Borrowers with mortgage home ownership are taking higher loans and defaulting the approved loans. Lending club should stop giving loans to this category when loan amount requested is more than 12000.
- People with more number of public derogatory records are having more chance of filing a bankruptcy. Lending club should make sure there are no public derogatory records for borrower.