LENDING CLUB CASE STUDY – EXPLORATORY DATA ANALYSIS

Name : - Sumit Suryanarayan Patro
Email ID : - patrosumit33@gmail.com
Contact Number : - +91 8779487231

OBJECTIVES : –

- Through Exploratory Data Analysis, we aim to display patterns and insights within the dataset to distinguish characteristics associated with loan defaulters. By analysing the data completely, we seek to disclose key factors influencing borrower behaviour providing the lending company with actionable information to make informed decisions in risk assessment.

- This investigation involves through into various features such as applicant information, financial indicators, and loan terms to identify correlations and trends that can be indicative of potential defaults. The ultimate goals is to empower the lending company with predictive knowledge, enabling them to implement targeted strategies and mitigate financial losses by avoiding loans to high-risk individuals.

- By leveraging EDA techniques, we intend to uncover nice patterns, relationships, and outliers within the dataset, shedding light on the factors that significantly contribute to loan default. This proactive approach allows the lending company to refine its credit approval process, strengthen risk management protocols, and enhance overall decision-making when extending loans.

PROVING SOLVING PLAN : –

- Importing Required Libraries & Loading Data
- Data Understanding
- Data Cleaning
- Data Formatting
- Univariate Analysis
- Unordered Categorical Variable Analysis
- Ordered Categorical Variable Analysis
- Derived Variable Analysis
- Bivariate Analysis
- Correlation Analysis

IMPORTING REQUIRED LIBRARIES & LOADING DATA: –

- numpy as np
- pandas as pd
- matplotlib.pyplot as plt
- seaborn as sns

DATA UNDERSTANDING: –

- Total Number of Rows 39717
- Total Number of Columns 111
- Types of Data are float, int and object

DATA CLEANING: -

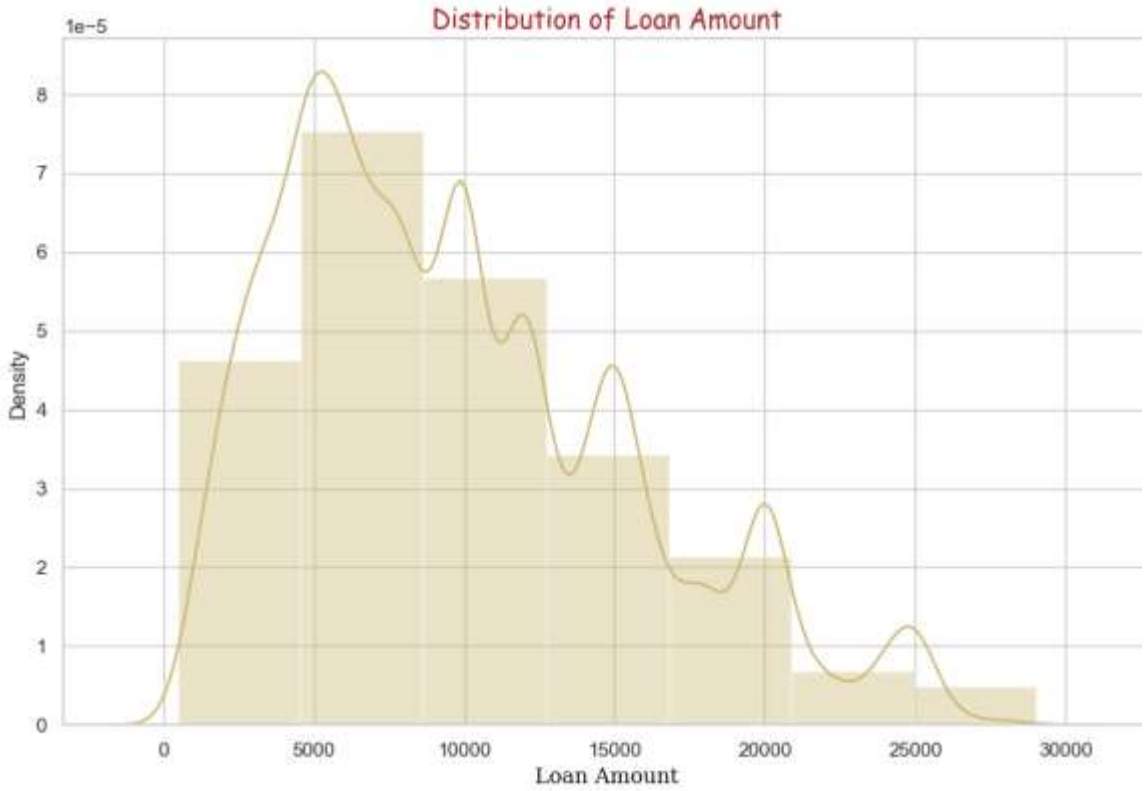
- Removed all the null values columns.
- Removed all the single value columns.
- Removed unique values columns [‘id’ and ‘url’].
- Removed the text and description values columns.
- Removed ‘subgrade’ column as we are limiting our analysis till grade level.
- Removed Behavioural Columns also.
- Removed the columns which is having more than 60% of null values which are not in use in our analysis.
- We have removed the records whose loan status is current ['loan_status' with 'Current' is in progress and cannot rectify if the loan is going bad or not.].
- We have gone through all the column and formatted our data as per our need.
- We have renamed the columns for better understanding while plotting.

DATA FORMATTING: -

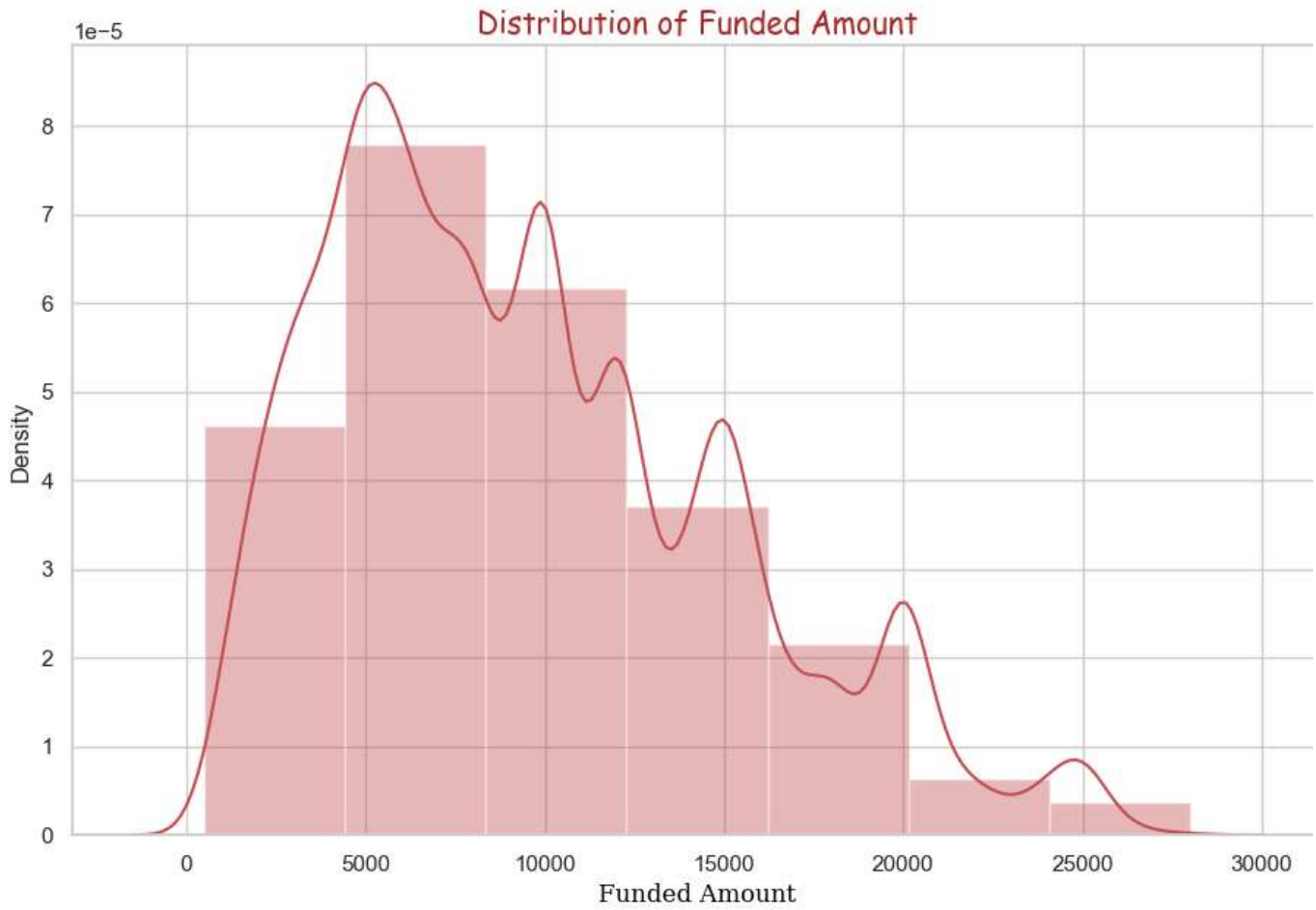
- loan_amnt: - We have converted the datatypes from int to float for better EDA.
- funded_amnt: - We have converted the datatypes from int to float for better EDA.
- term: - While checking in dataset we have observed string part [months]. Hence, we have removed.
- int_rate: - While analysing the dataset we have observed string part [%]. Hence, we have removed.

UNIVARIATE ANALYSIS

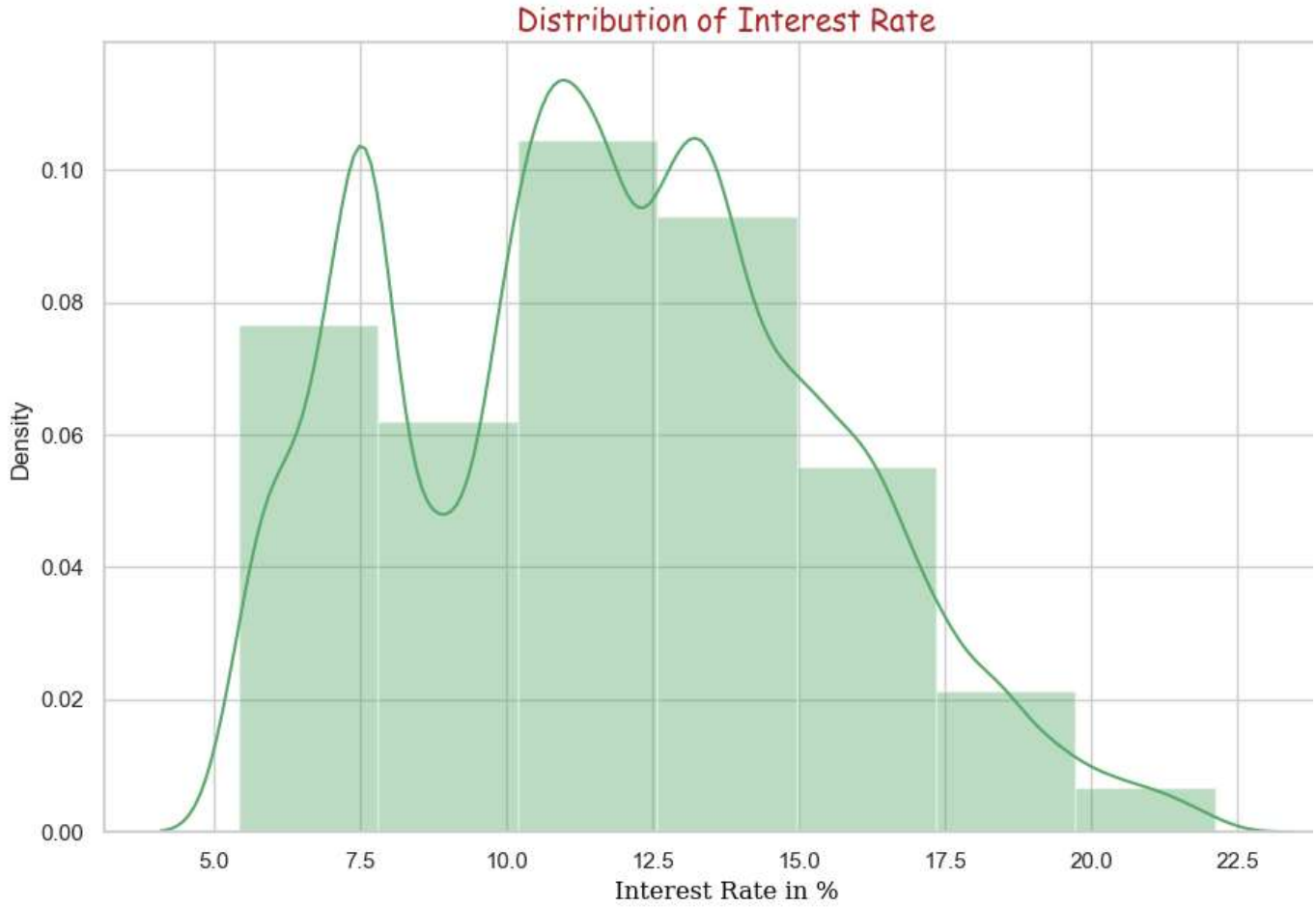
- It has been observed that maximum number of Loan Amount is in between 5000 to 13000.



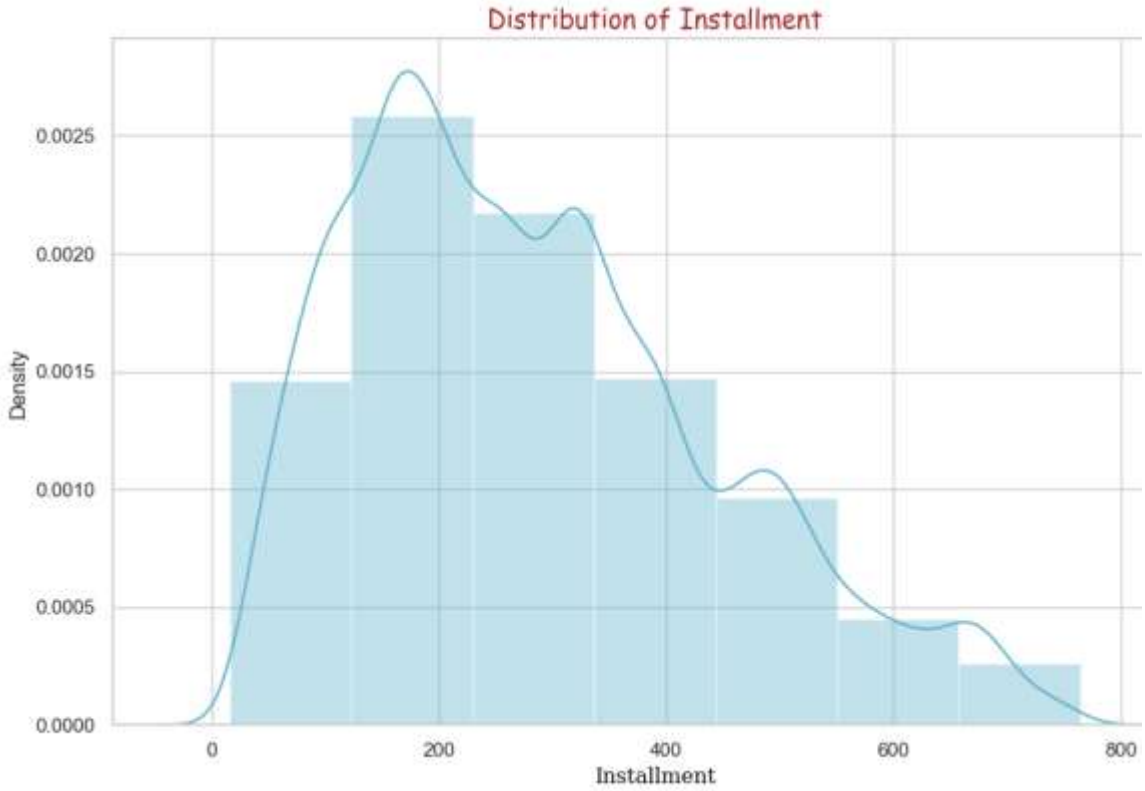
- It has been observed that maximum number of Funded Amount is in between 5000 to 12000.



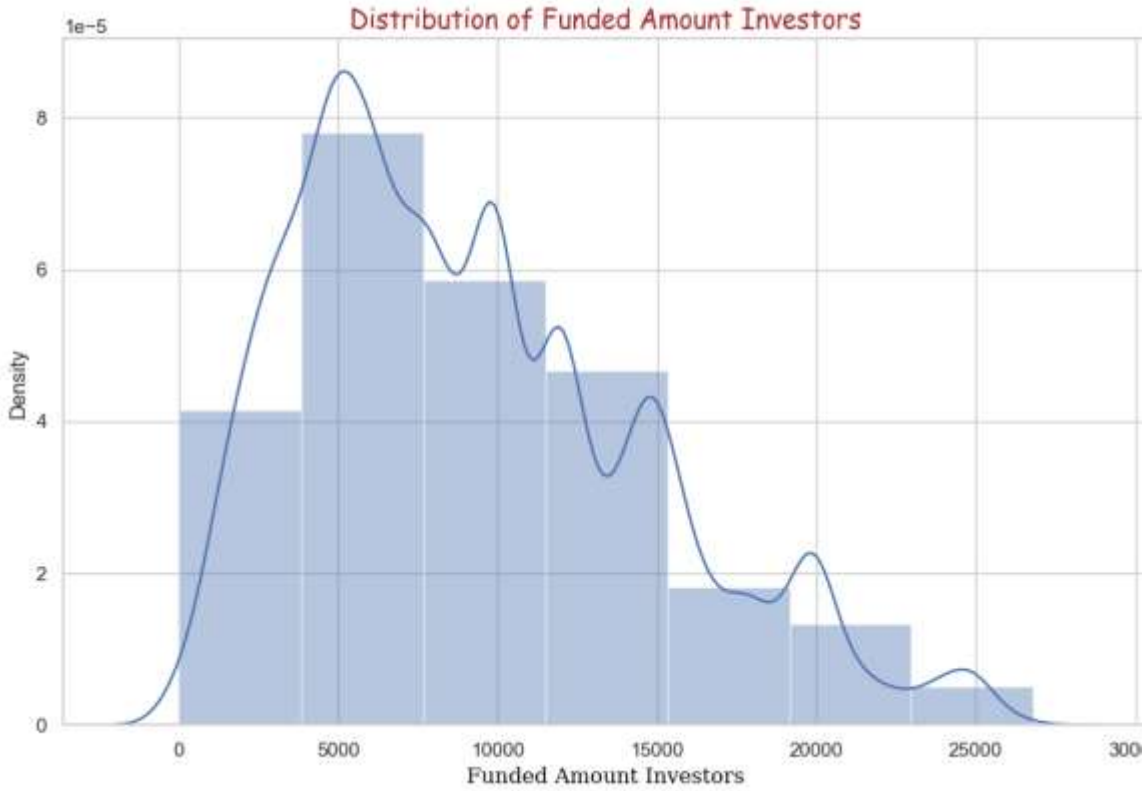
- It has been observed that maximum number of Interest Rate is in between 5% to 15%.



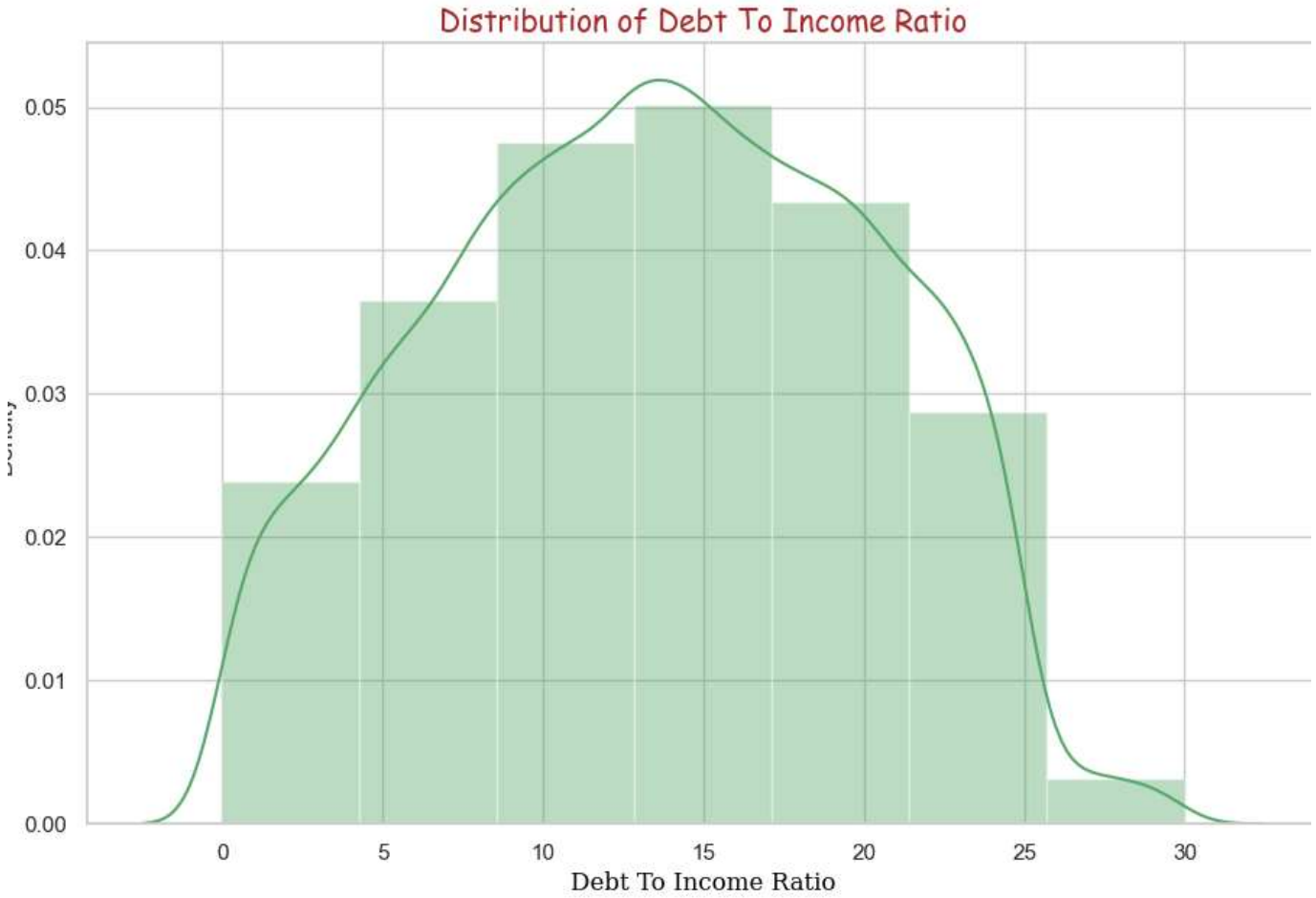
- It has been observed that maximum number of Installment amount is in between 20 to 400.



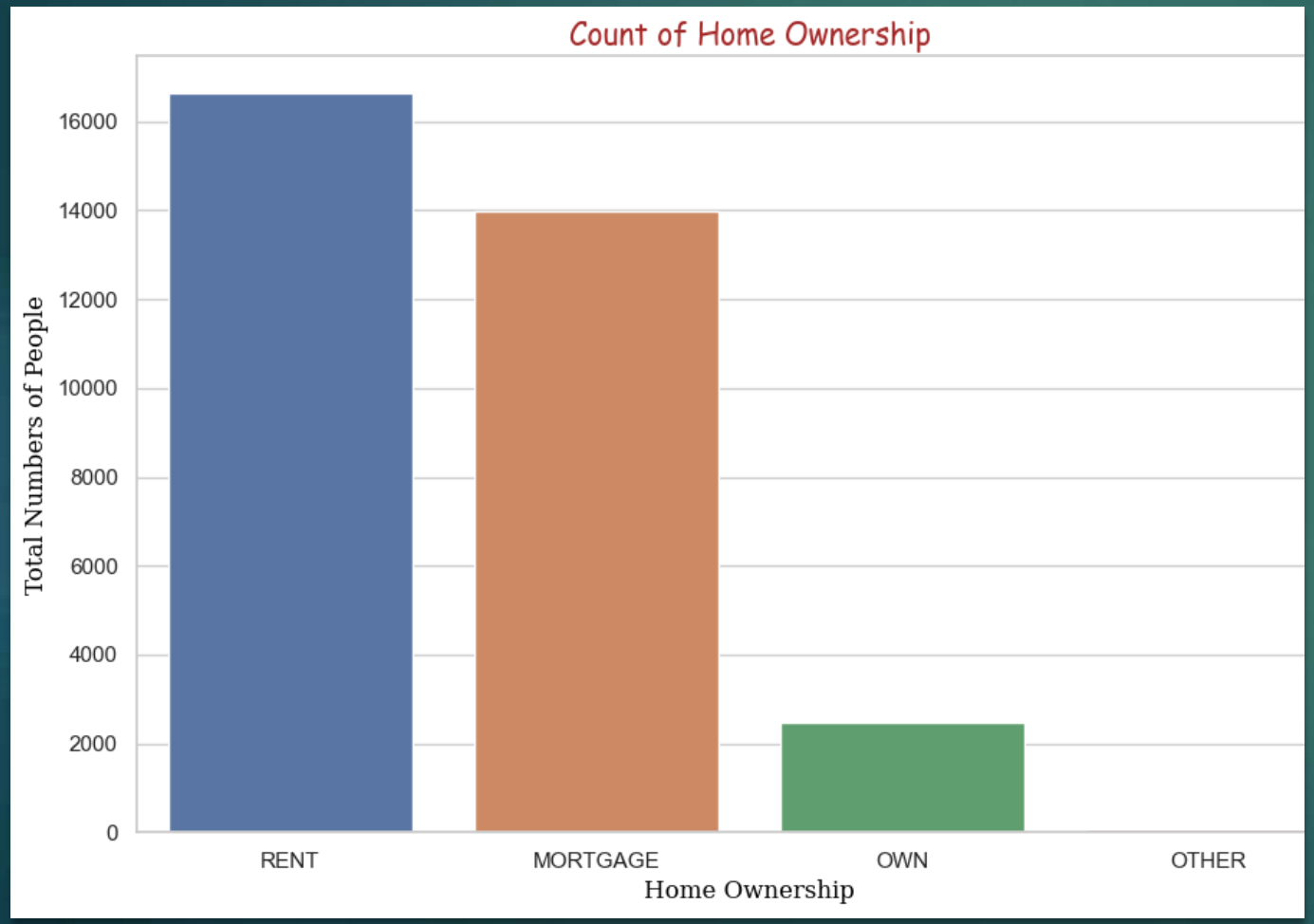
- It has been observed that maximum number of Funded Amount Investors is in between 5000 to 11000.



- It has been observed that maximum number of Debt To Income Ratio amount is in between 5 to 22.

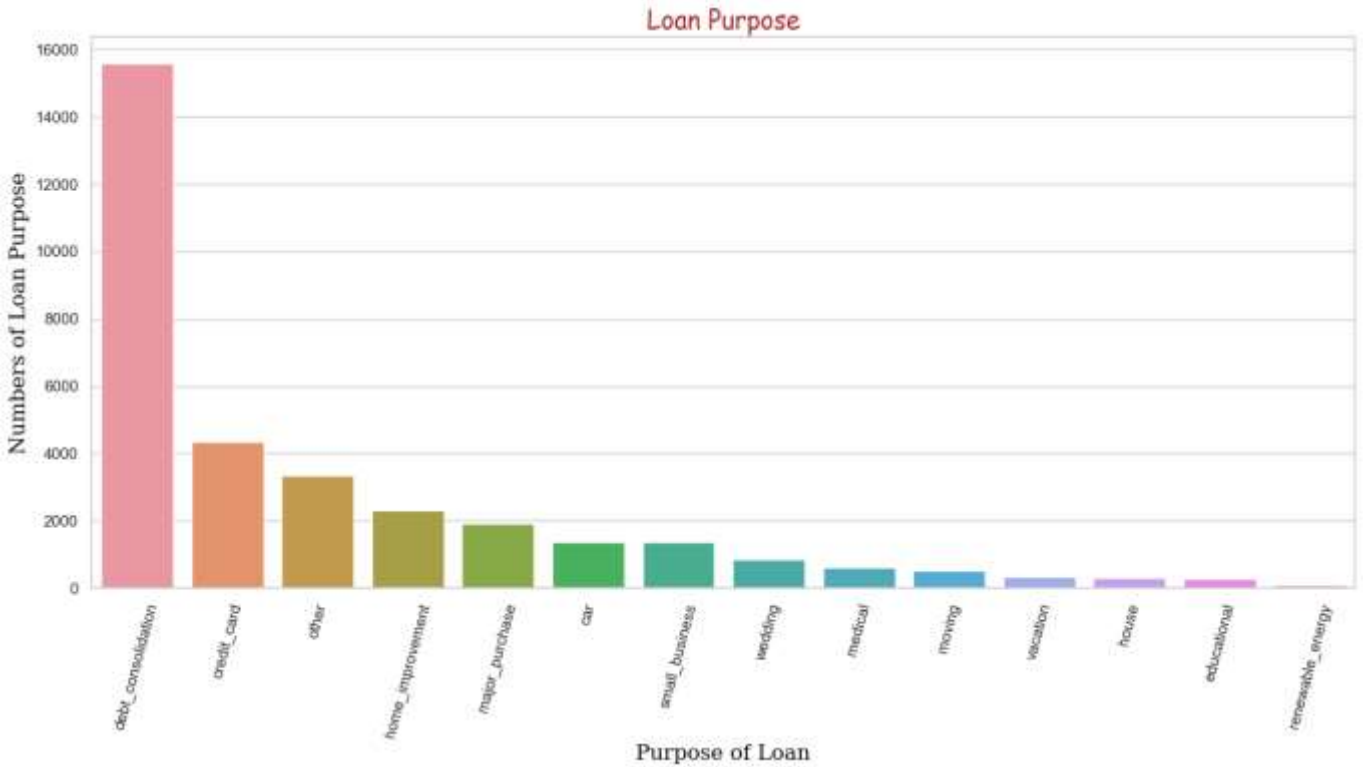


❑ UNORDERED CATEGORICAL VARIABLE ANALYSIS

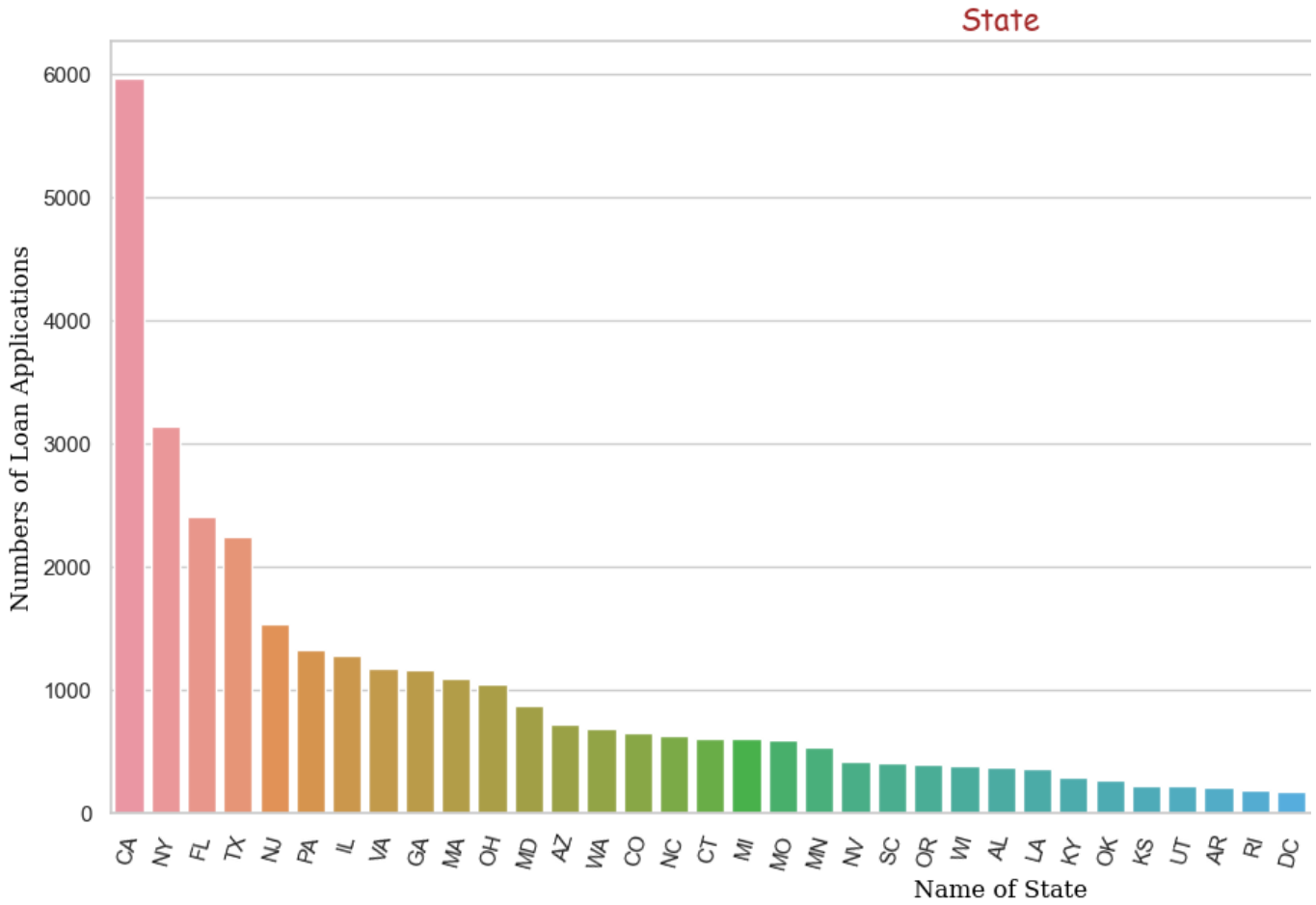


- Rent and Mortgage exhibit the highest among Home Owner Status Categories

- Majority of loan application are in the category of debt_consolidation.

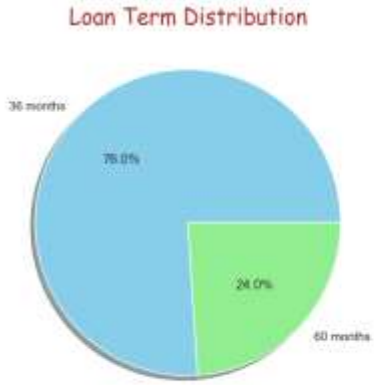


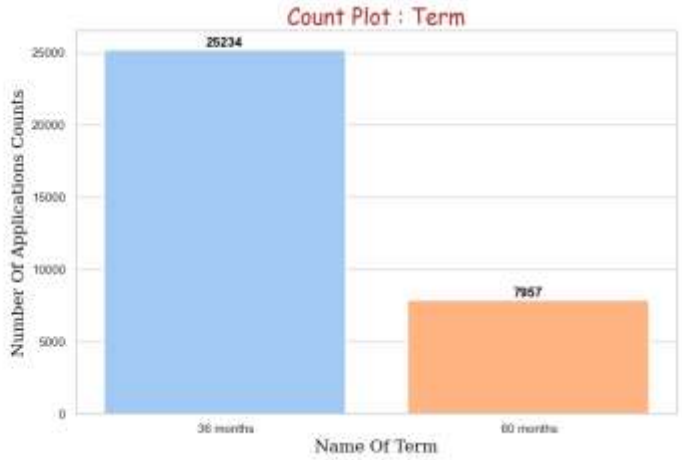
- CA state has the highest number count of loan applications



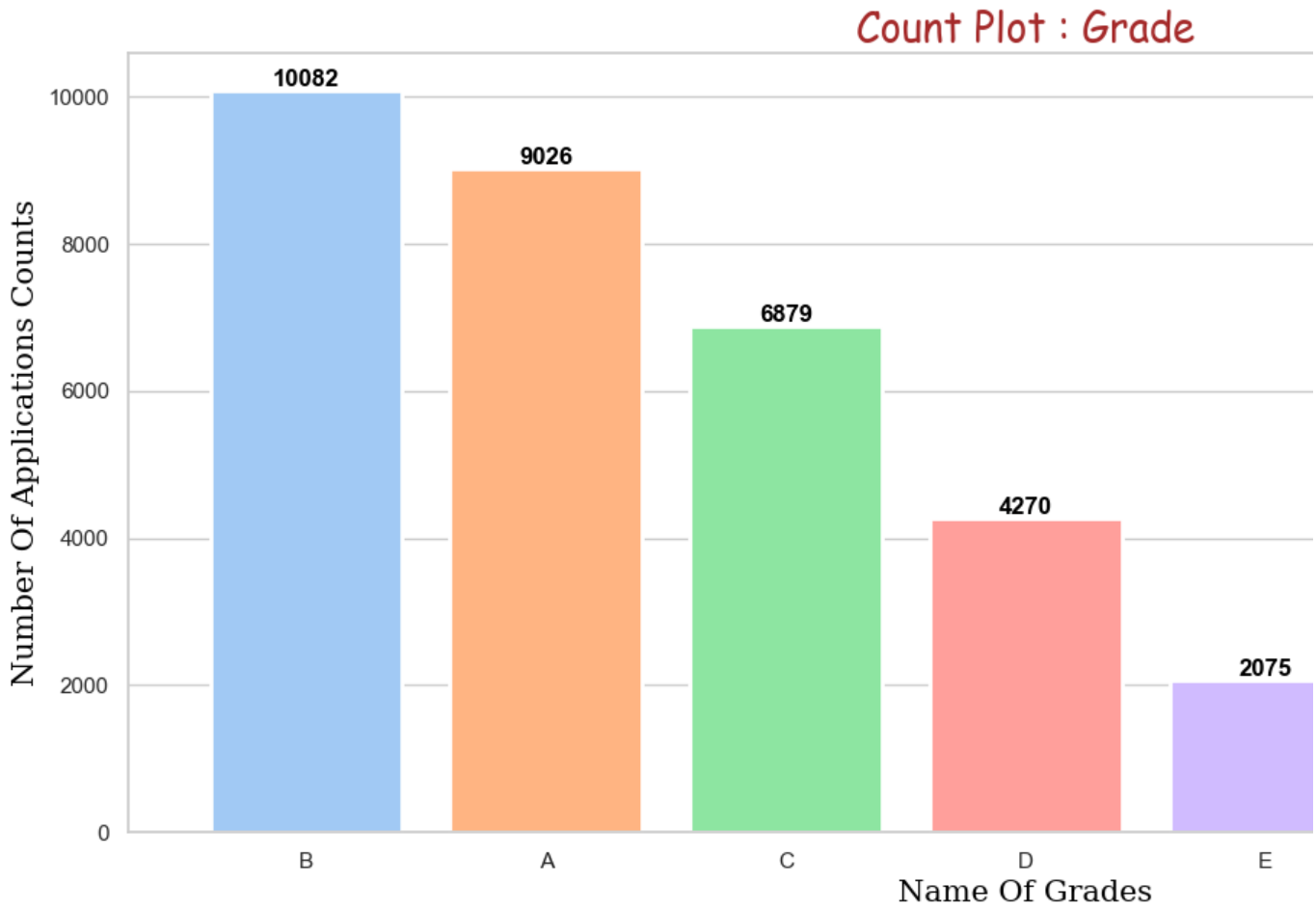
ORDERED CATEGORICAL VARIABLE ANALYSIS

- 76.0 % of loan applications counts are in term 36 months and 25234 applications count.
- 24.0 % of loan applications counts are in term 60 months and 7957 applications count.

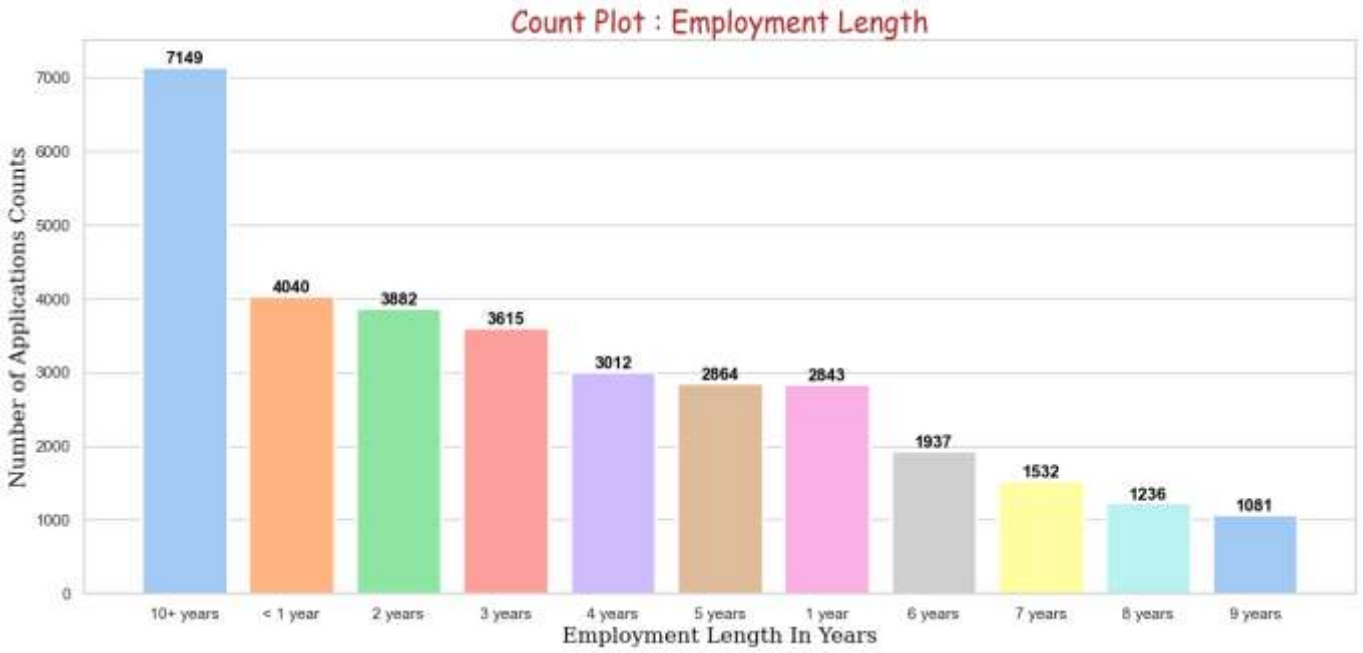




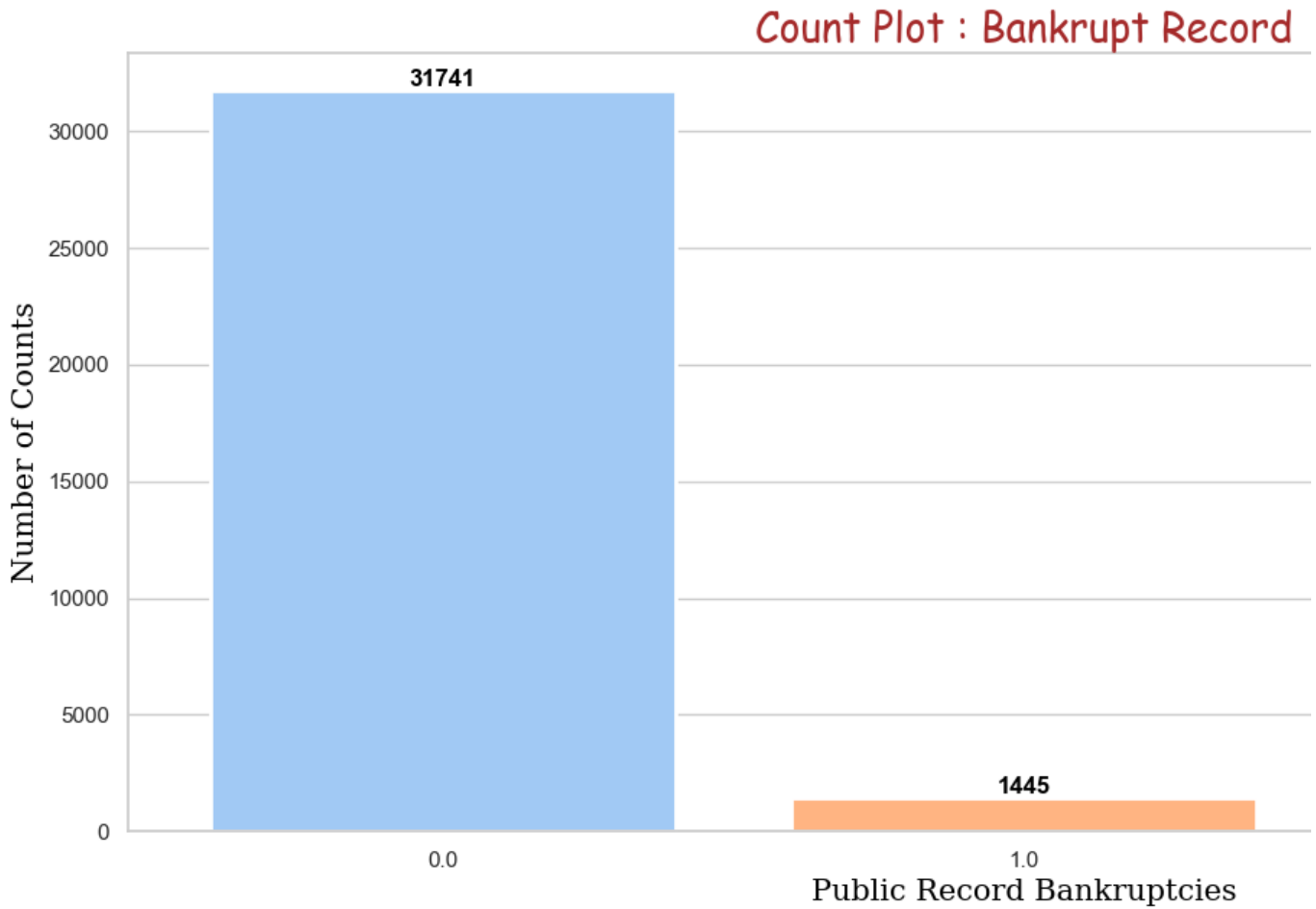
- Grade B has the highest number of loan application count - 10082.
- Grade G has the lowest number of loan application count - 159.



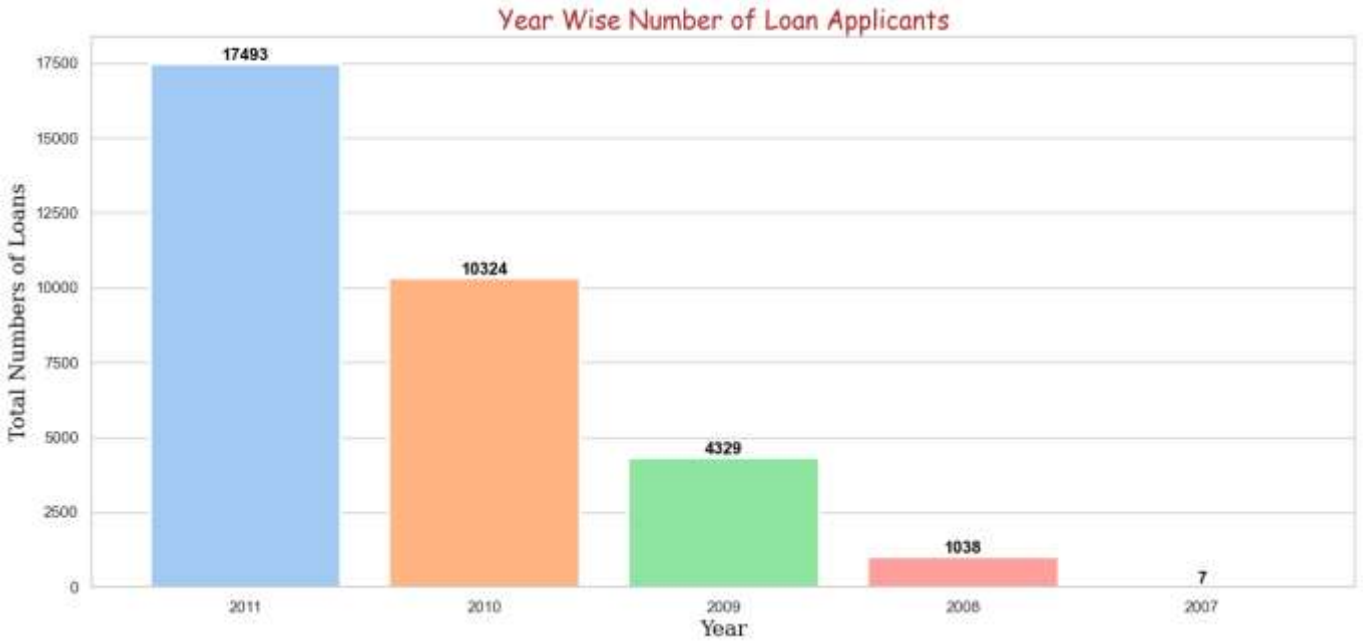
- 10+ years of employment length has the highest number of applications count with 7149 numbers.
- 9 years of employment length has the lowest number of applications count with 1081 numbers.



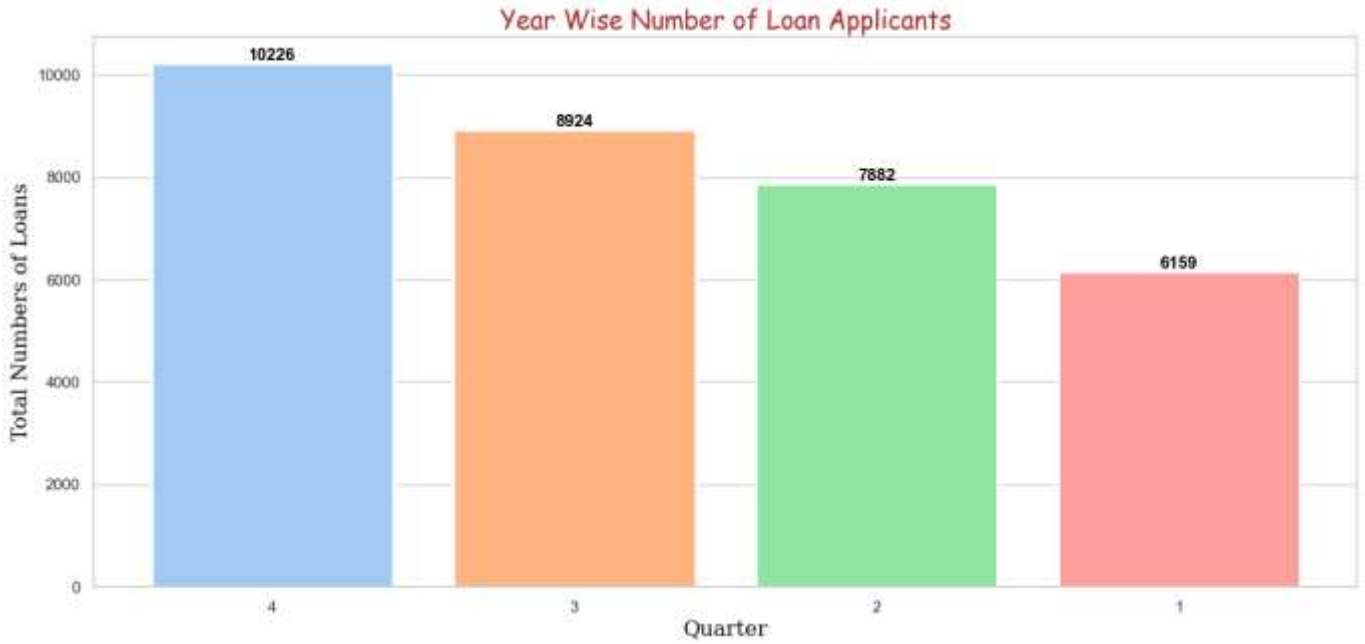
- 31741 customer out of 33191 does not have any bankrupt records.



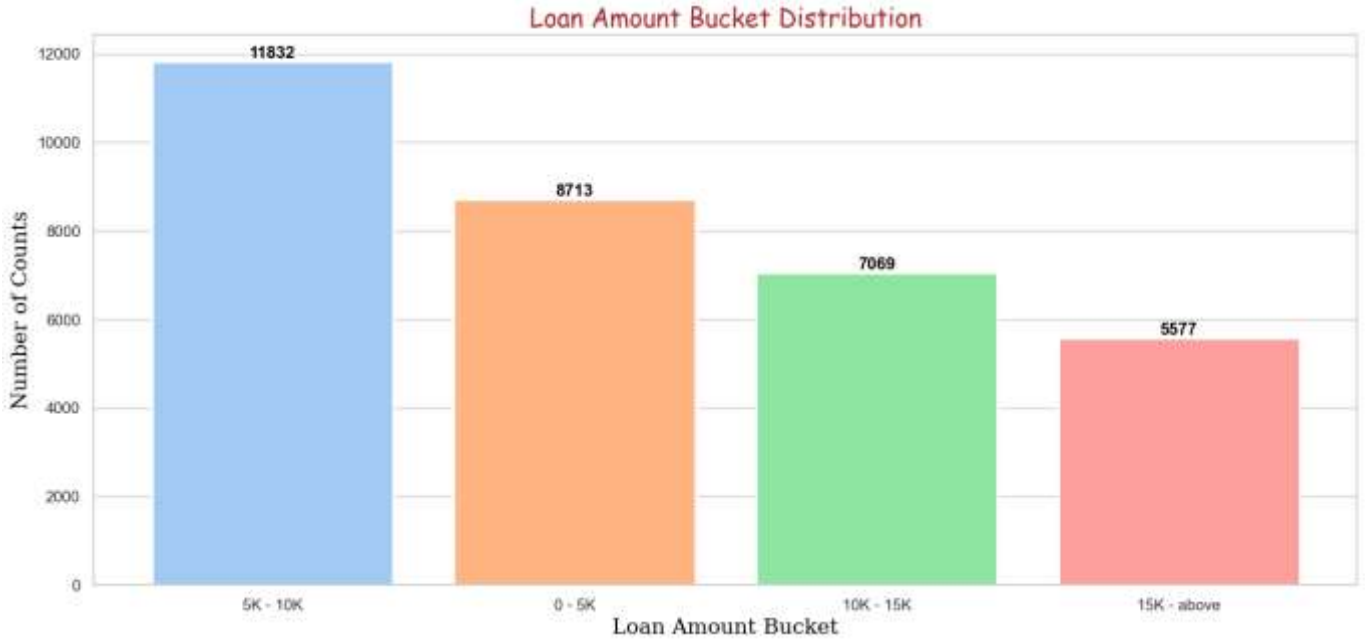
- Loan application counts are increasing year over year. Maybe the risk exposure is increasing over the year



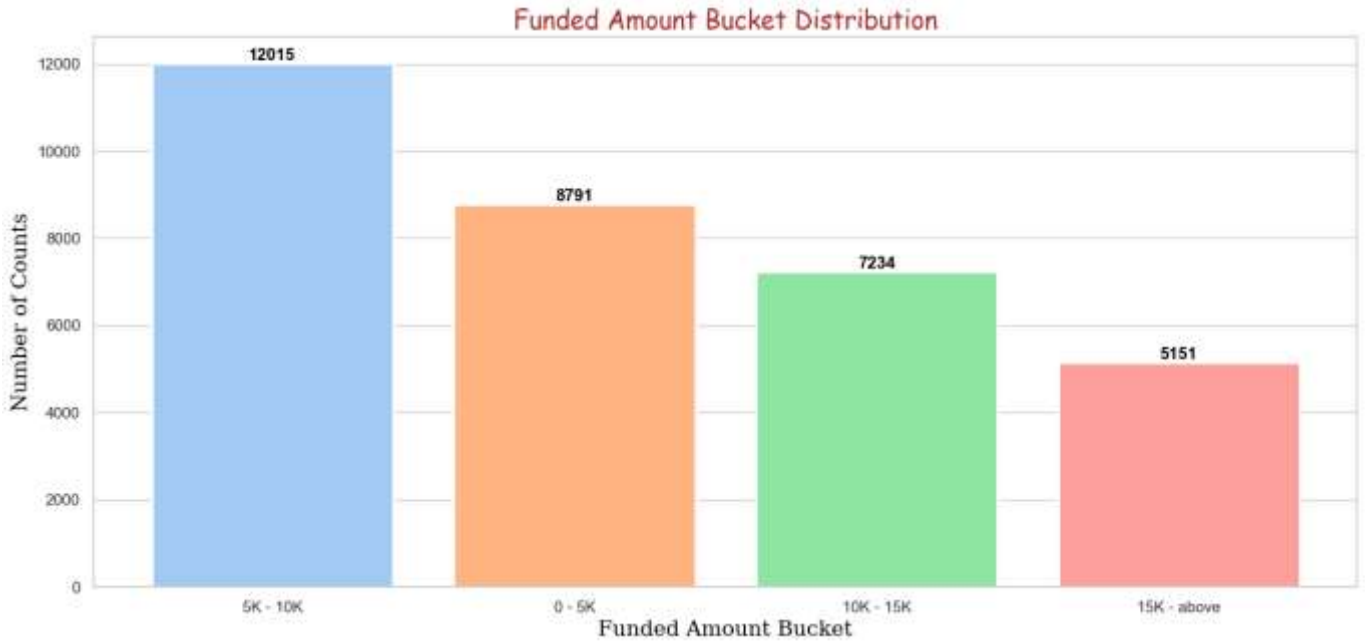
- Highest Counts are in 4th Quarter which is 17493.



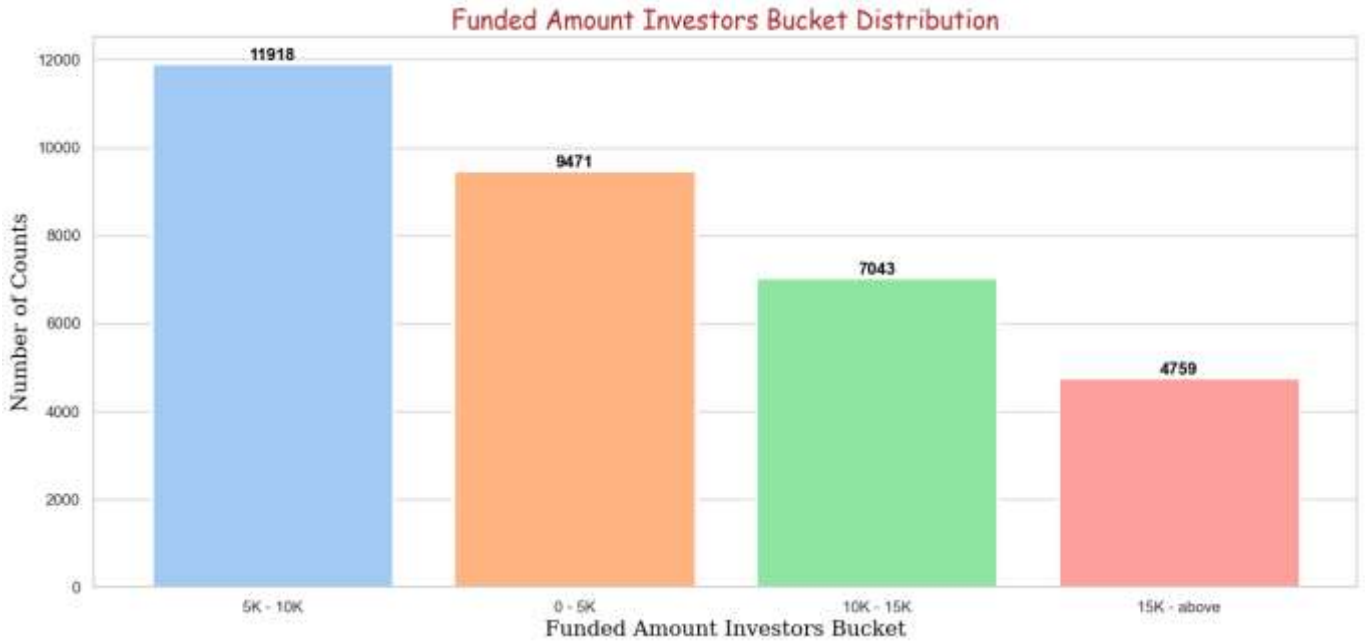
- Loan Amount applications falls in the range of 5K to 10K.
- 11832 which most as compared to other.



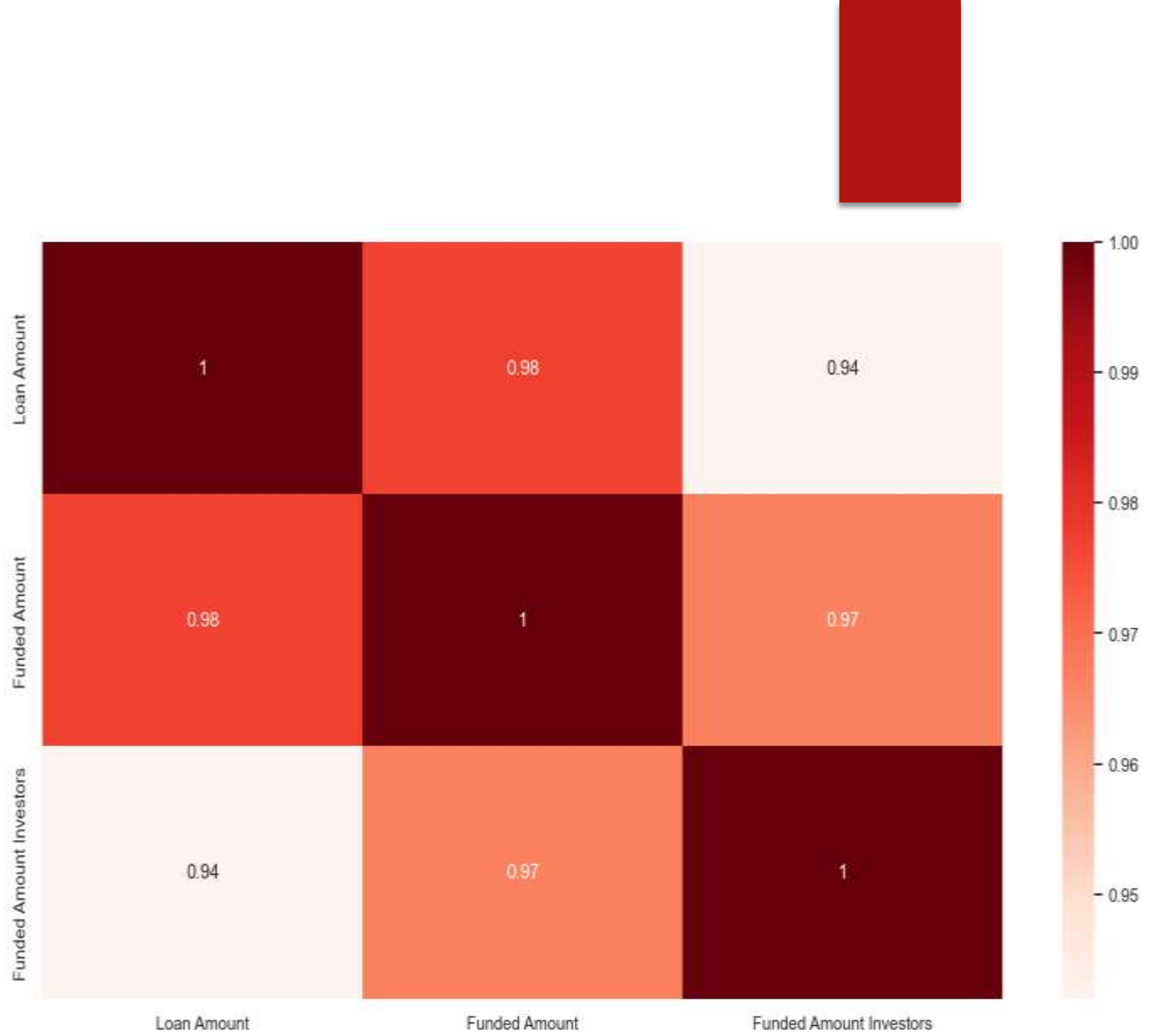
- Funded Amount Bucket applications falls in the range of 5K to 10K.
- 12015 which most as compared to other.



- Funded Amount Investors Bucket applications falls in the range of 5K to 10K.
- 11918 which most as compared to other.

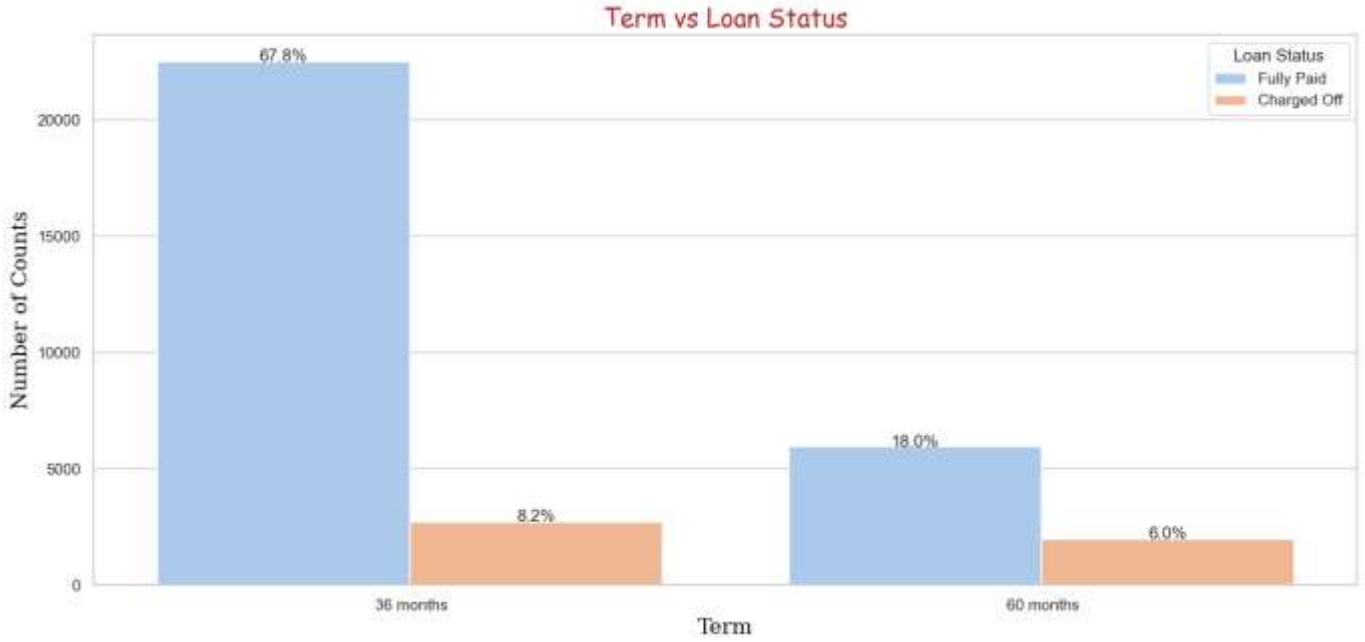


- The Correlation shows that the Loan Amount, Funded Amount, Funded Investors Amount are highly correlated.

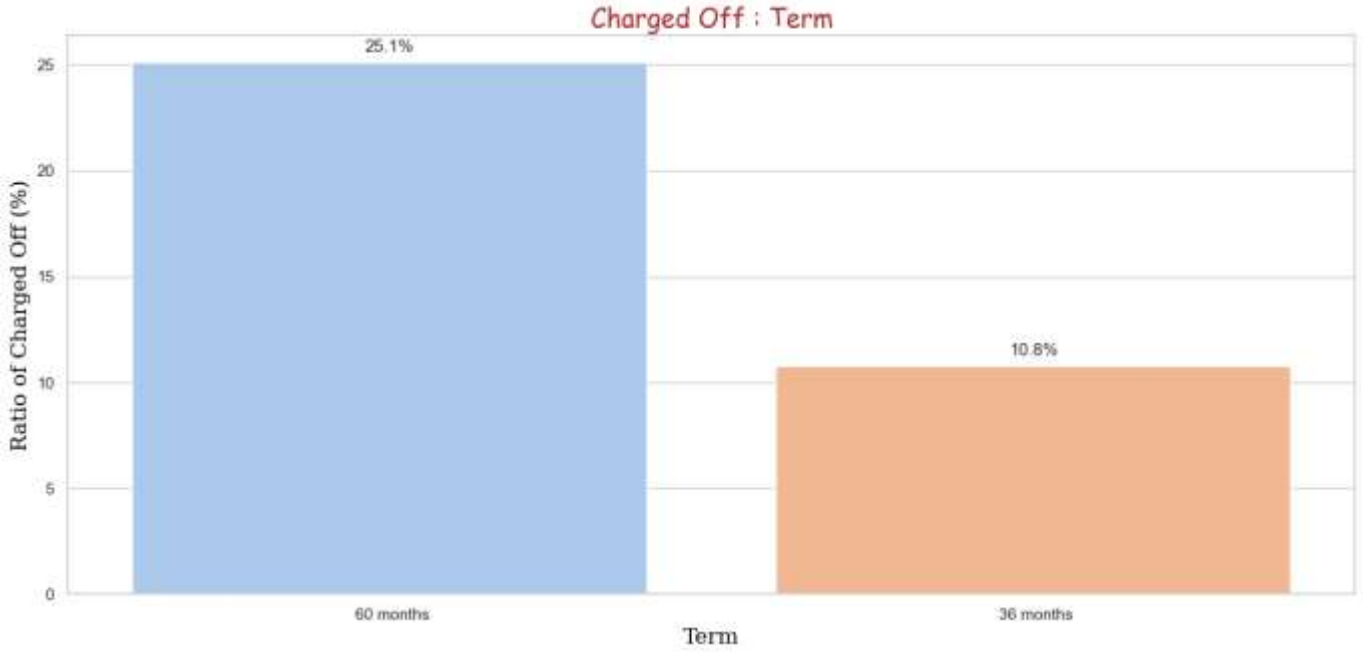


BIVARIATE ANALYSIS

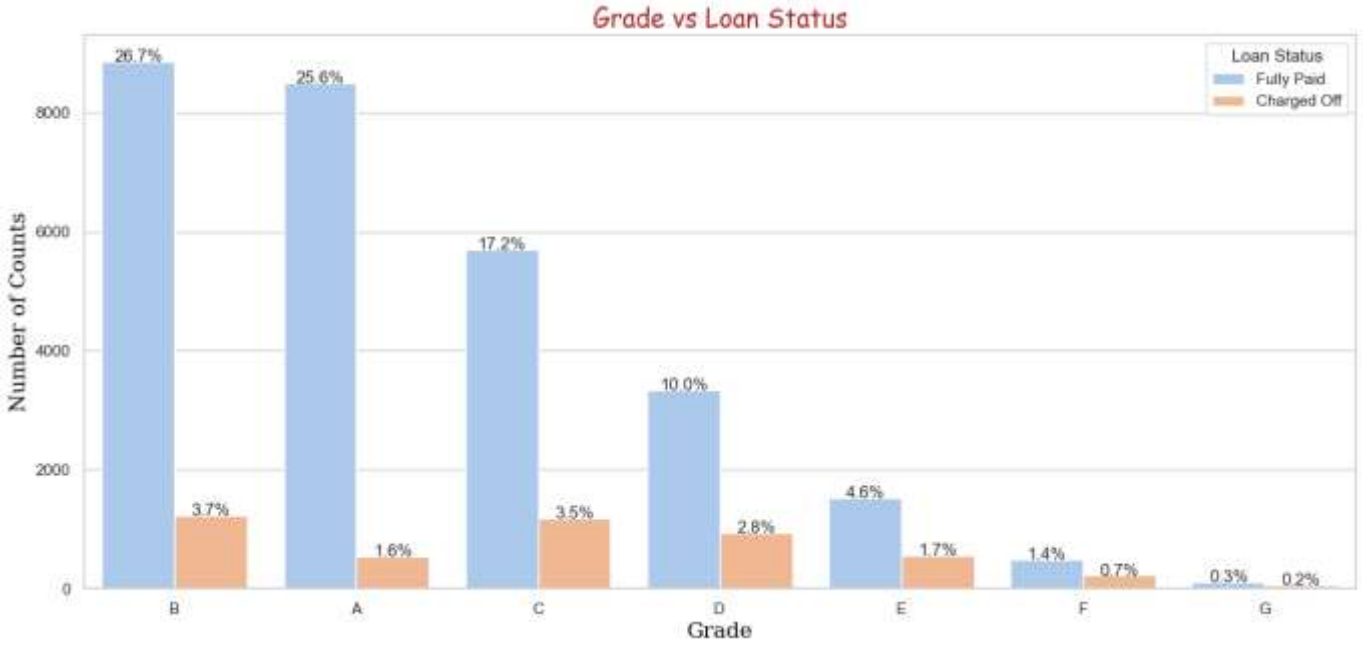
- Overall percentage of36 months Charged Off is bit higher then 60 months
- 36 months which charged off is 8.2% as compared to 60 months 6.0.



- Charged Off ratio of 60 months [25.1%] which is higher than 36 months [10.8%]

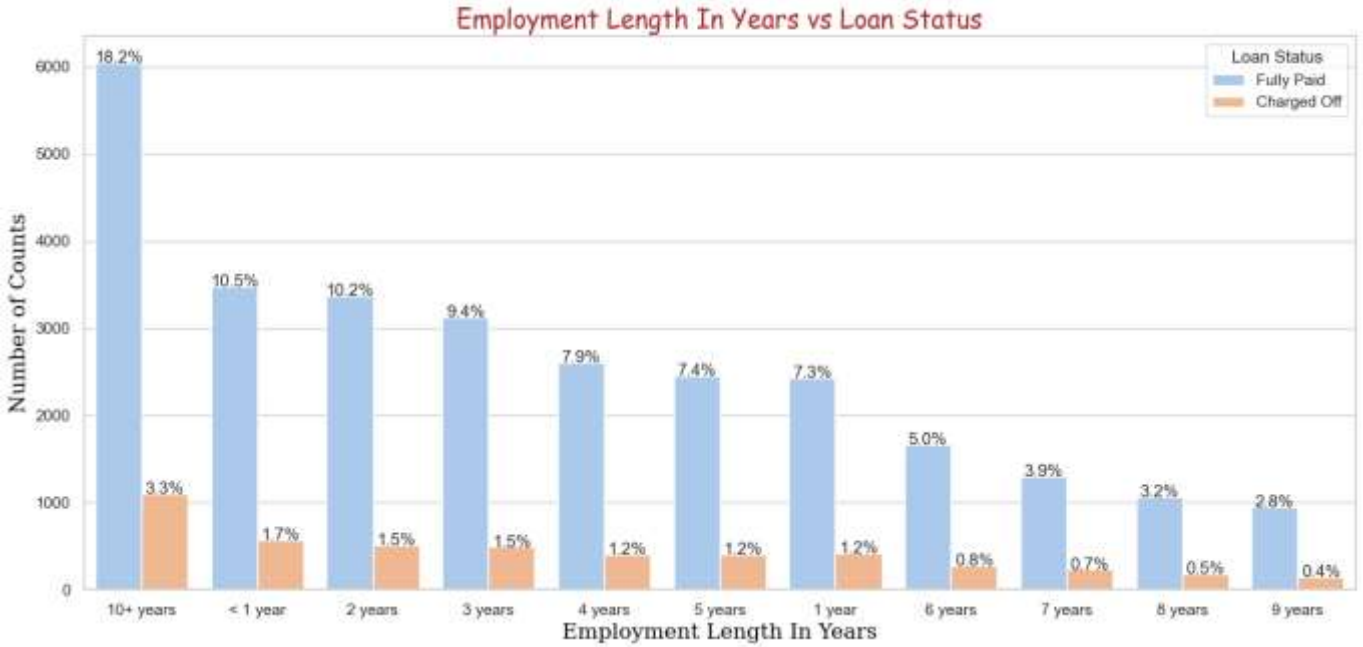


- The maximum number of loan is in Grade B.
- Highest Percentage of overall Charged Off are in Grade B [3.7%] and Grade C [3.6%]

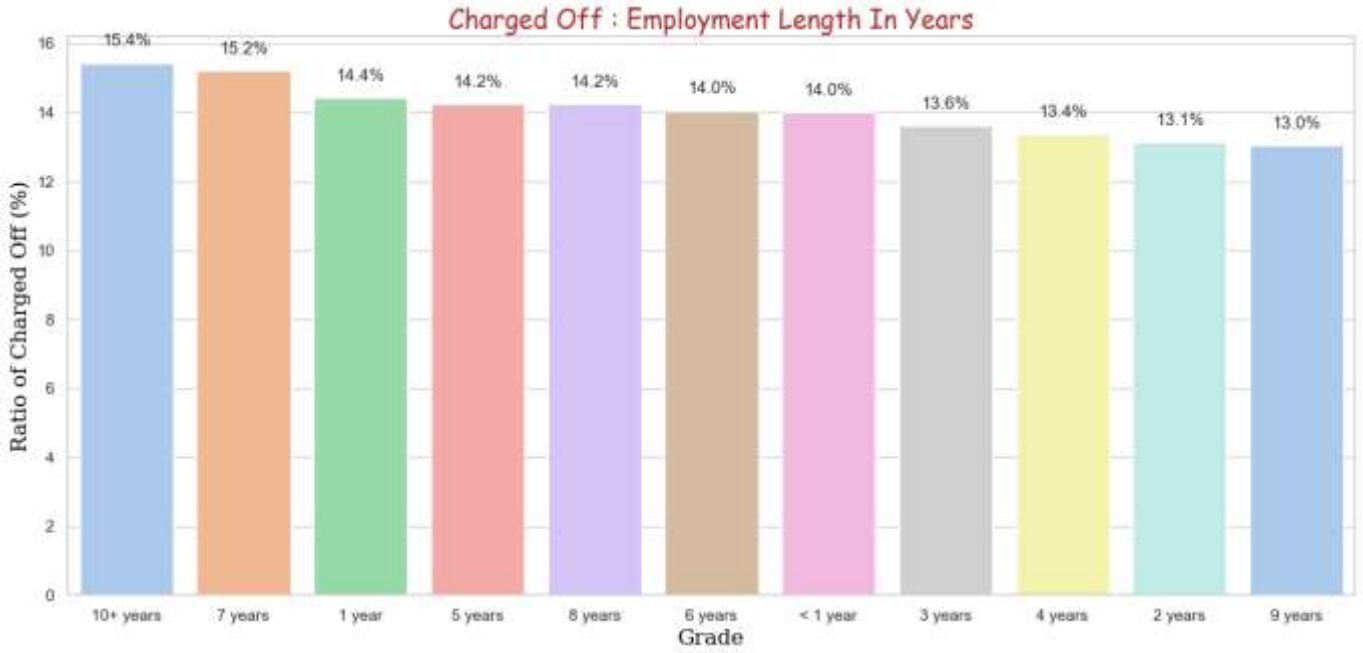


- The highest percent of Charged Off is in the Grade G.
- Probability of Charged Off is consistently increasing from A to G.
- High chances of Charged Off is in Grade F and Grade G.

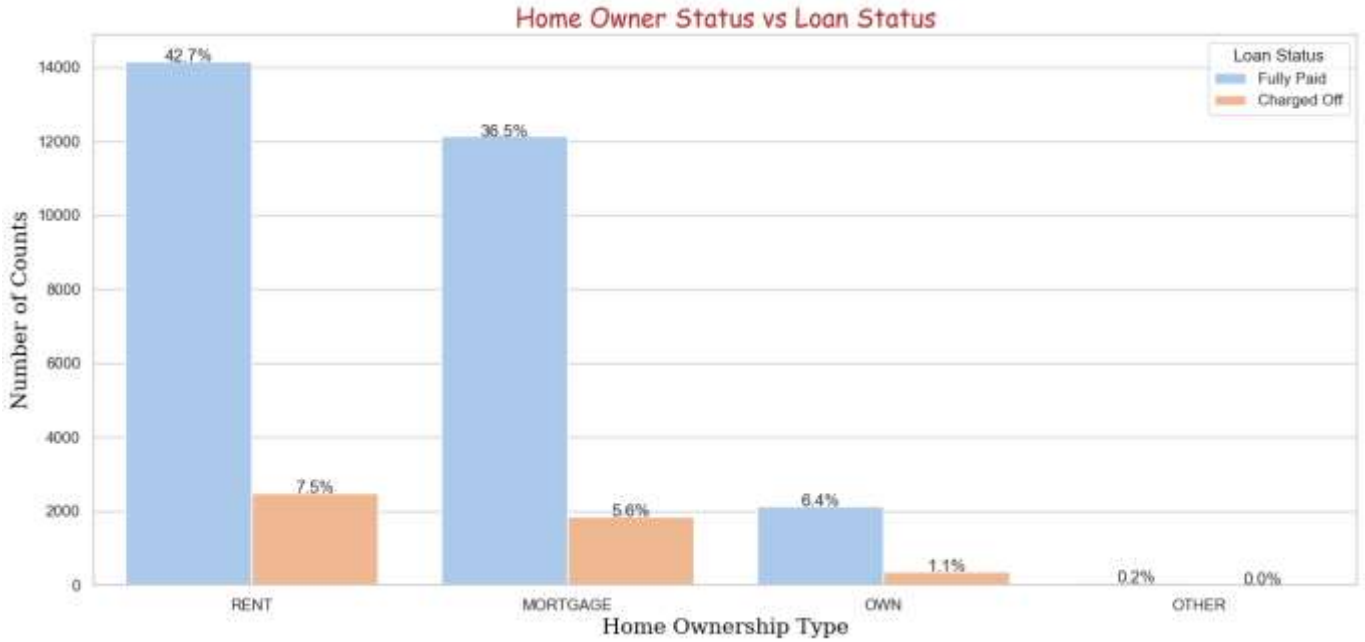
- Maximum number of Charged Off are in 10 Years or above.
- Least number of Charged Off are in 9 Years.



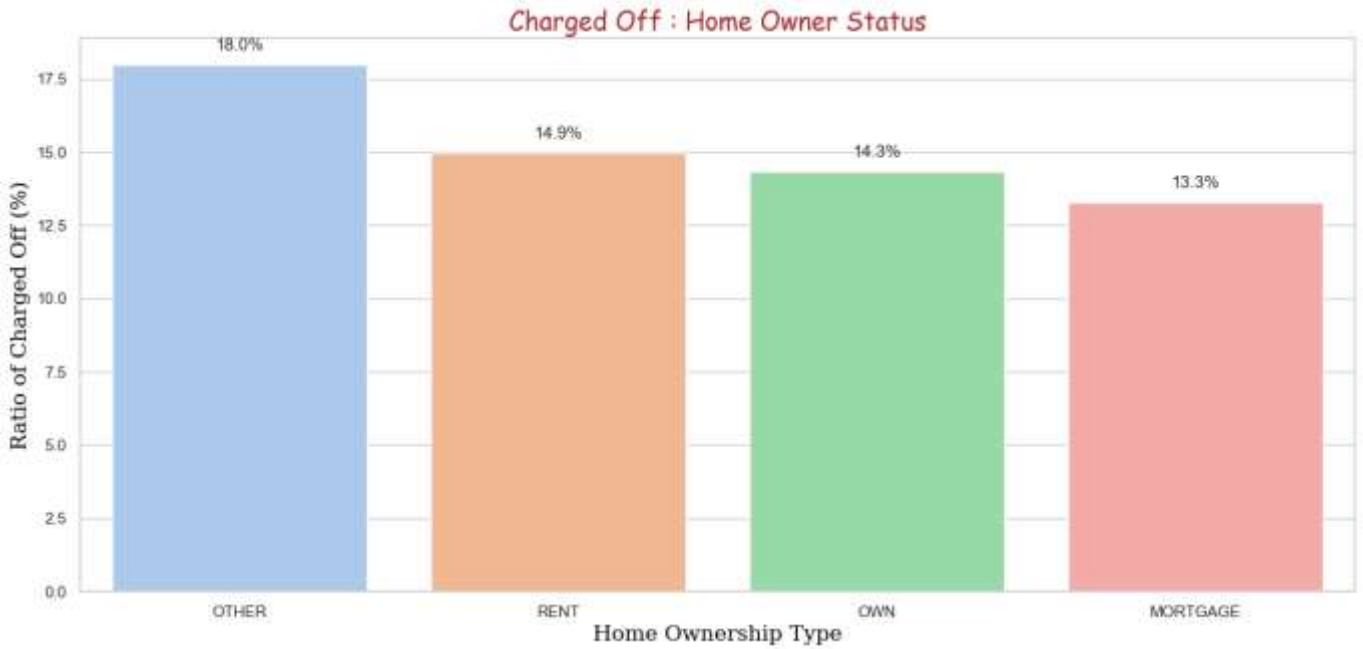
- Maximum number of Charged Off are in 10 Years or above.
- Least number of Charged Off are in 9 Years.



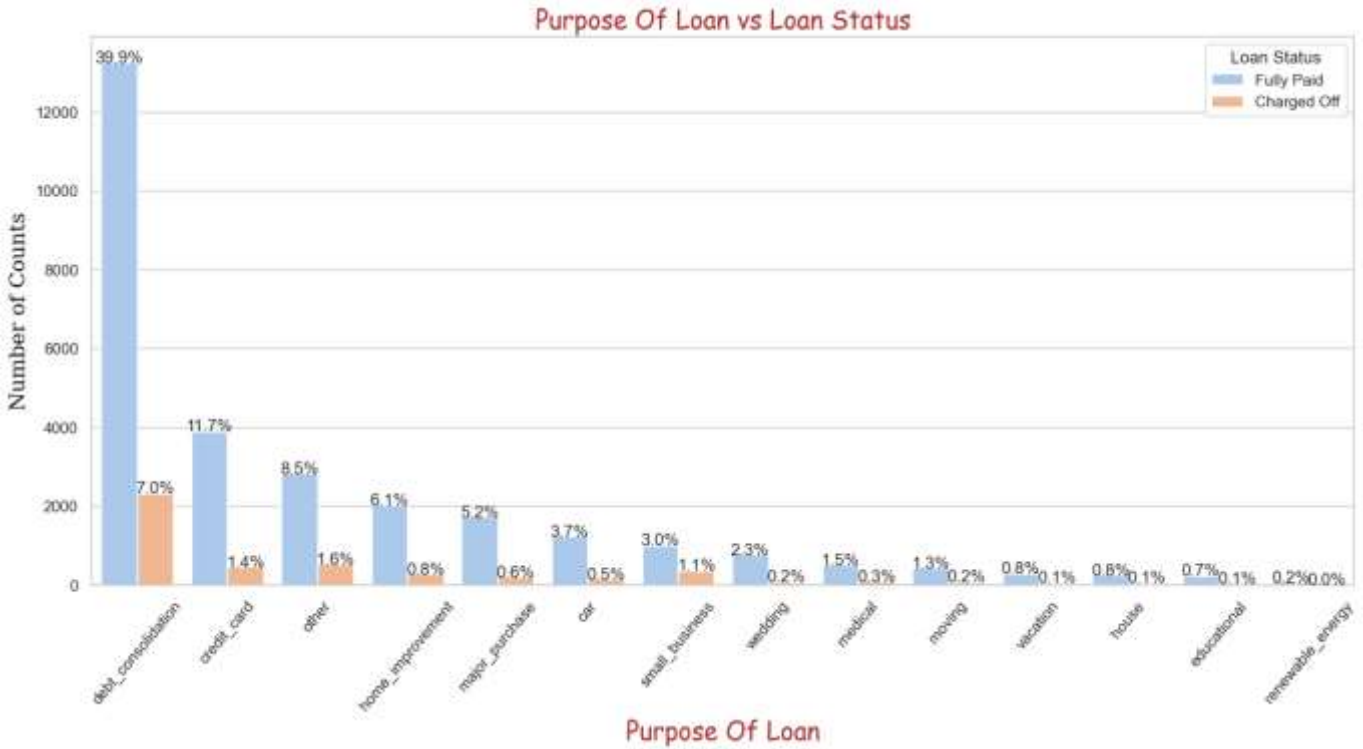
- Overall highest number of Charged Off are in the RENT and MORTGAGE with 42.7% and 36.5% respectively



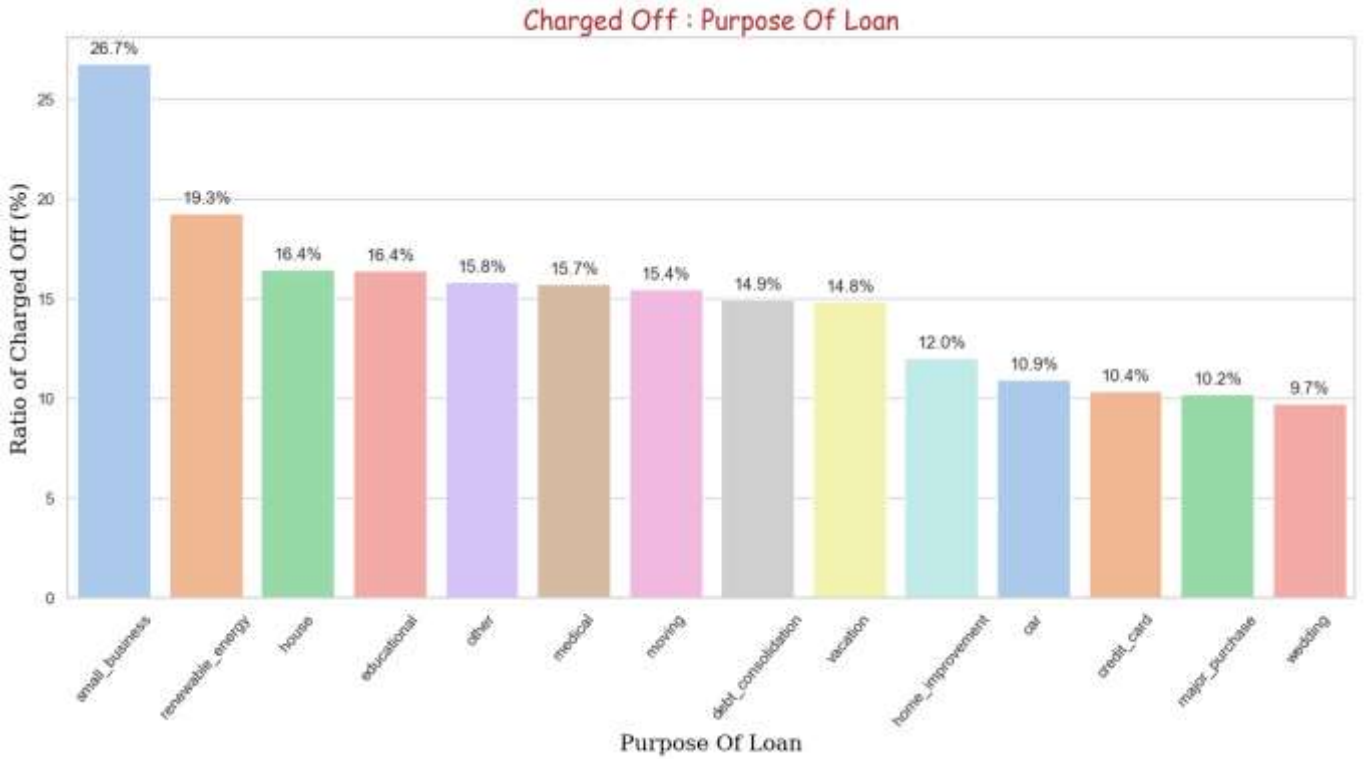
- Within each Home Owner Status the ratio of OTHER is higher.
- Other percent which is 18.0% highest.



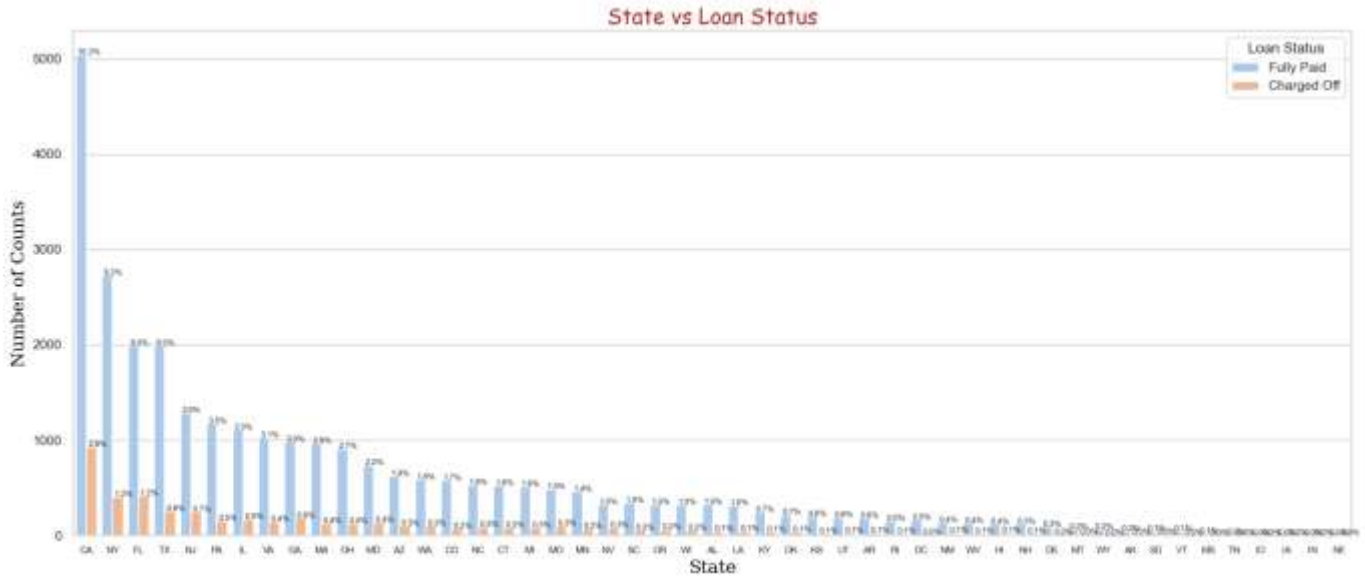
- The category of deb_consolidation has the more risk of Charged Off.



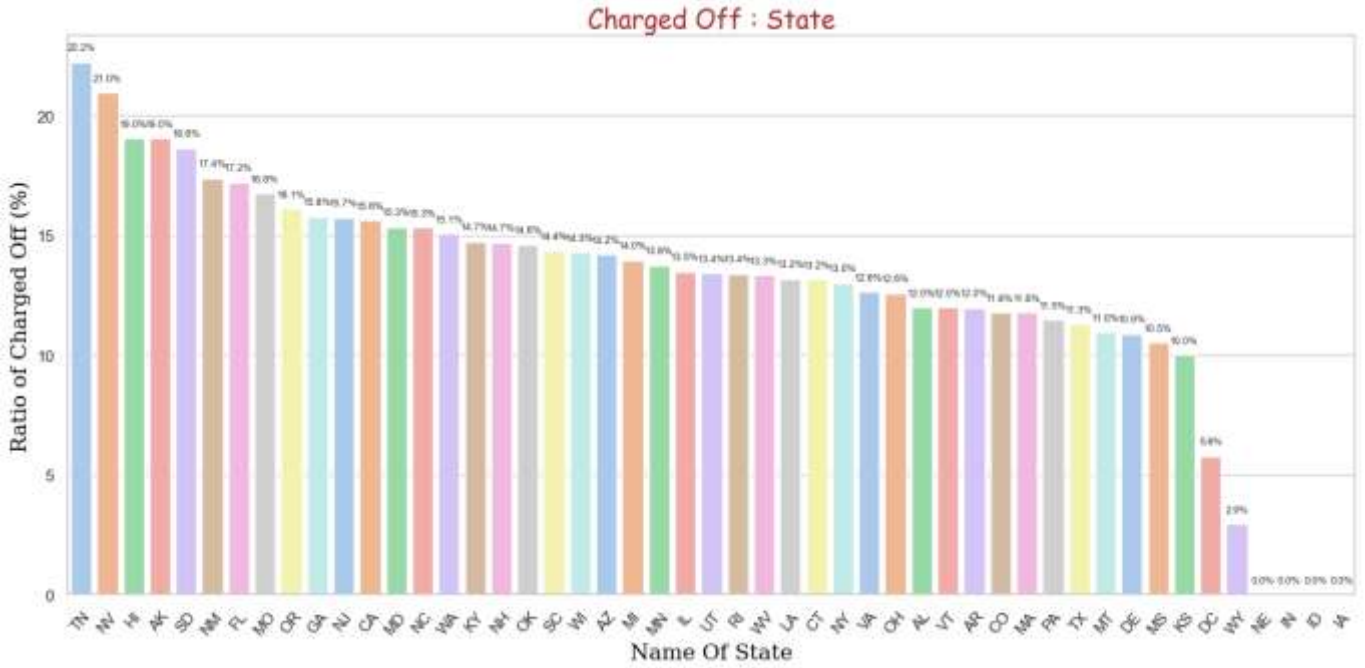
- Highest probability of Charged Off within the category are small_business but the volume is immensely low.



- Highest Volume of Loans are from CA only.

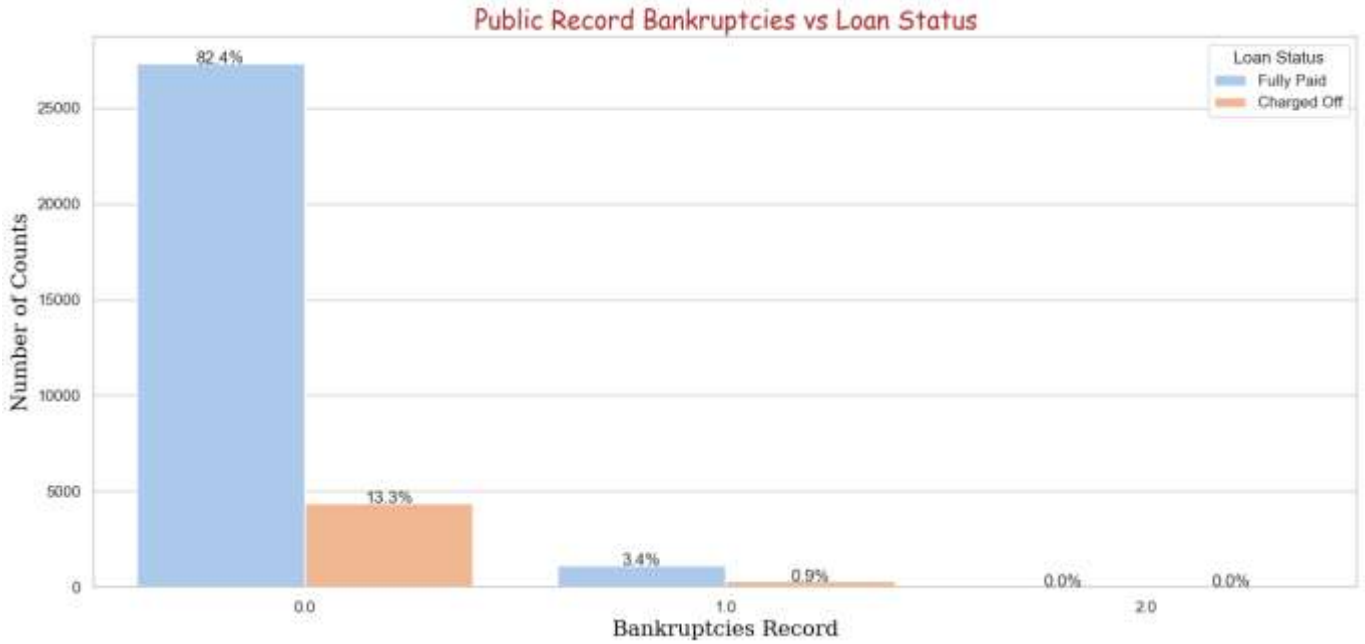


- State TN and NV has the highest Charged Off.
- NE, IN, ID, and IA are the very low volume which cannot be considered.
- Loan applications from TN and NV will have very high risk.



- State TN and NV has the highest Charged Off.
- NE, IN, ID, and IA are the very low volume which cannot be considered.
- Loan applications from TN and NV will have very high risk.

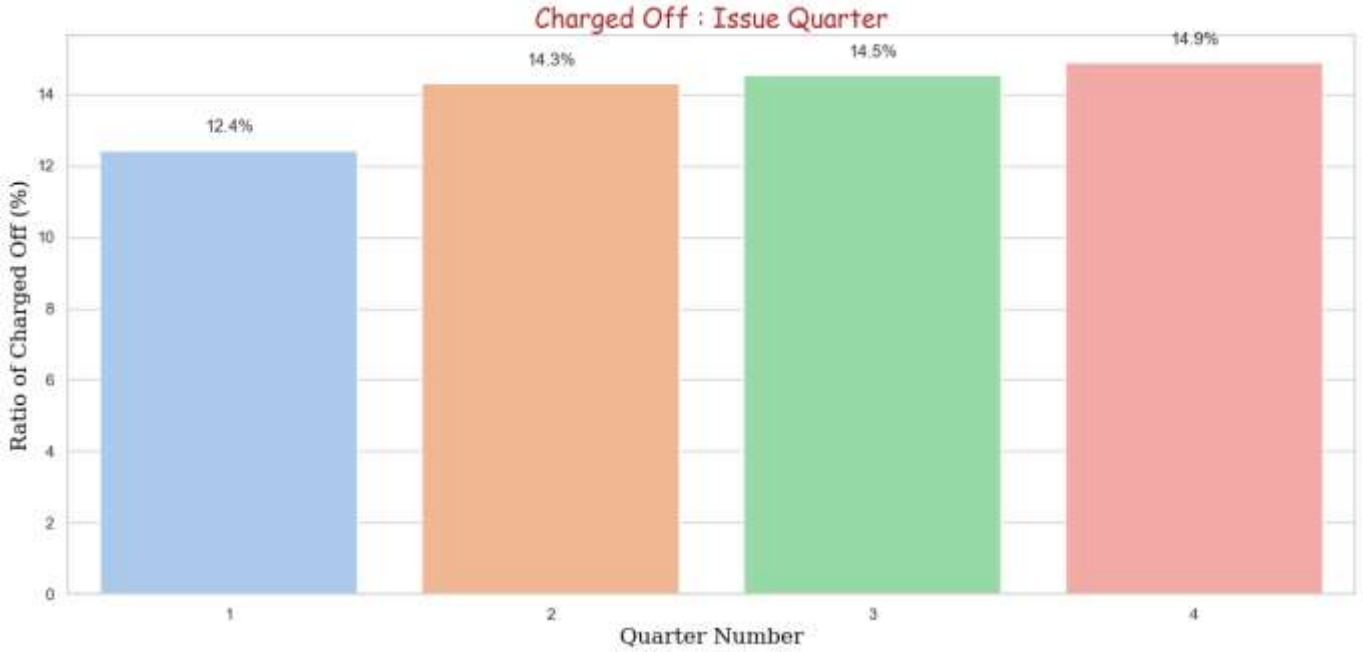
- Higher number of Charged Off are in the category of 0.



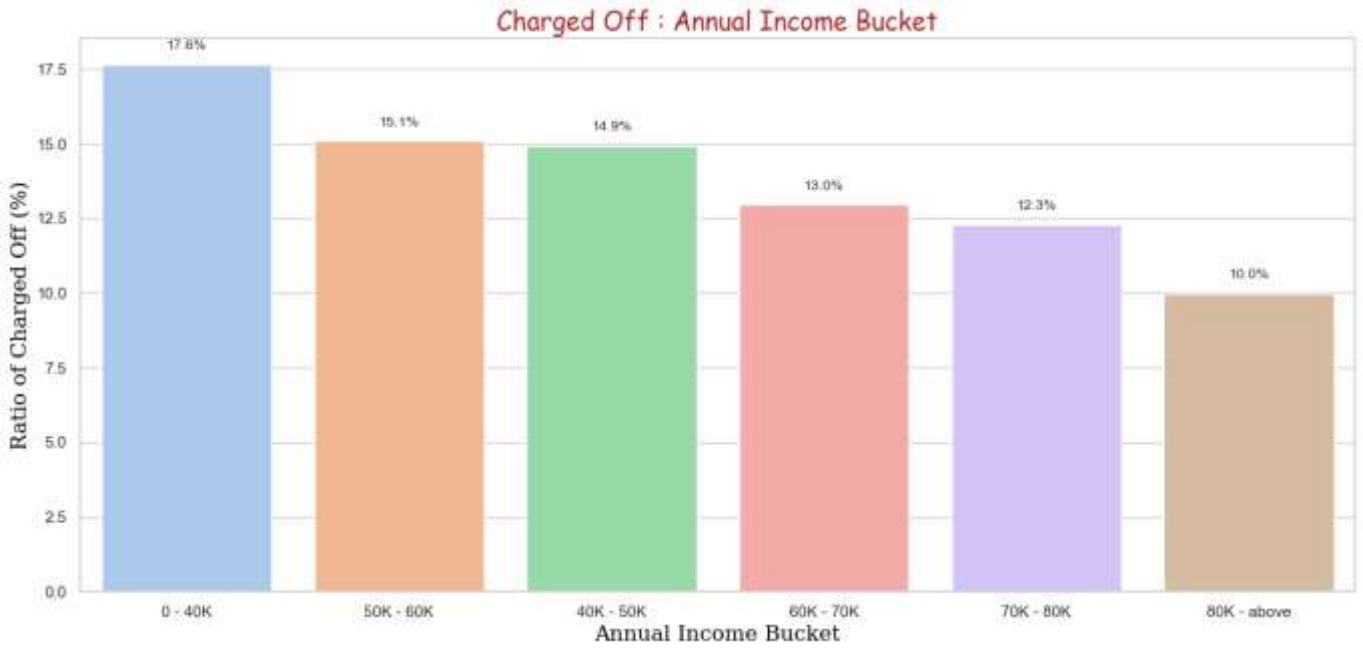
- Customers having Bankruptcy record has high Charged Off ratio.

- Quarter 4 has the maximum number of Loan Application.

- Quarter 4 has the highest ratio of Charged Off as compared to other three quarter.

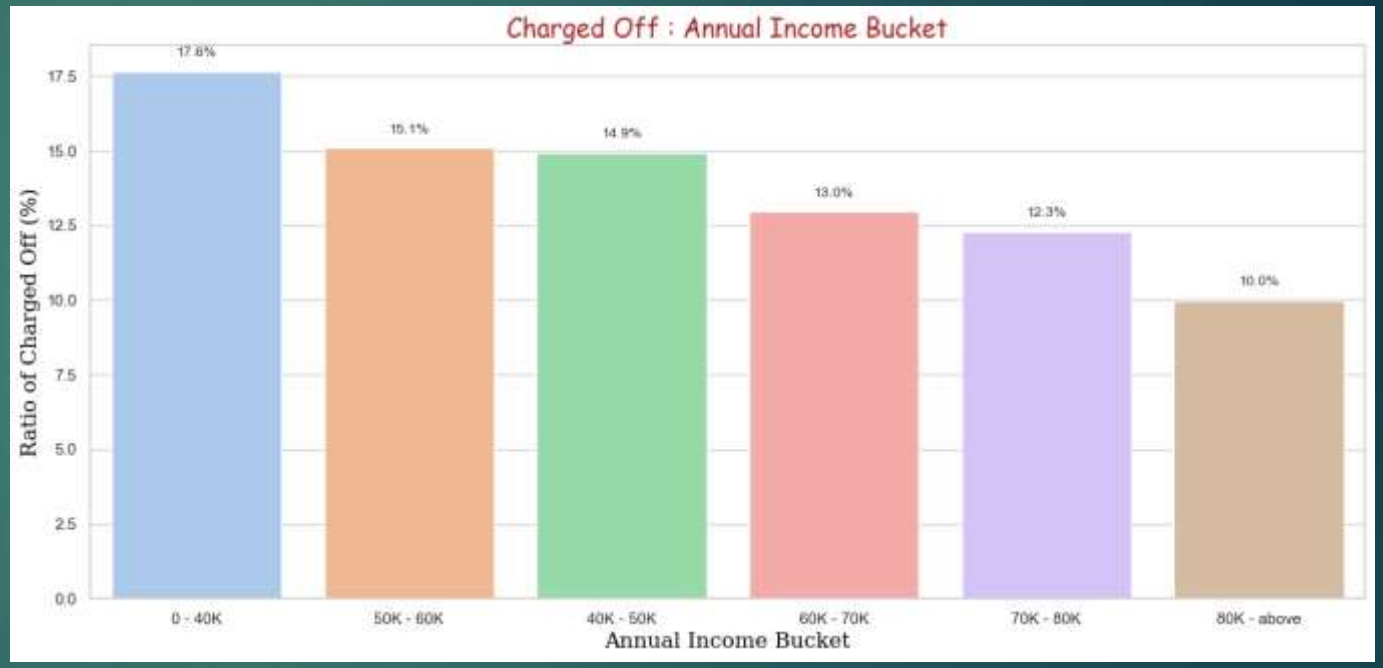


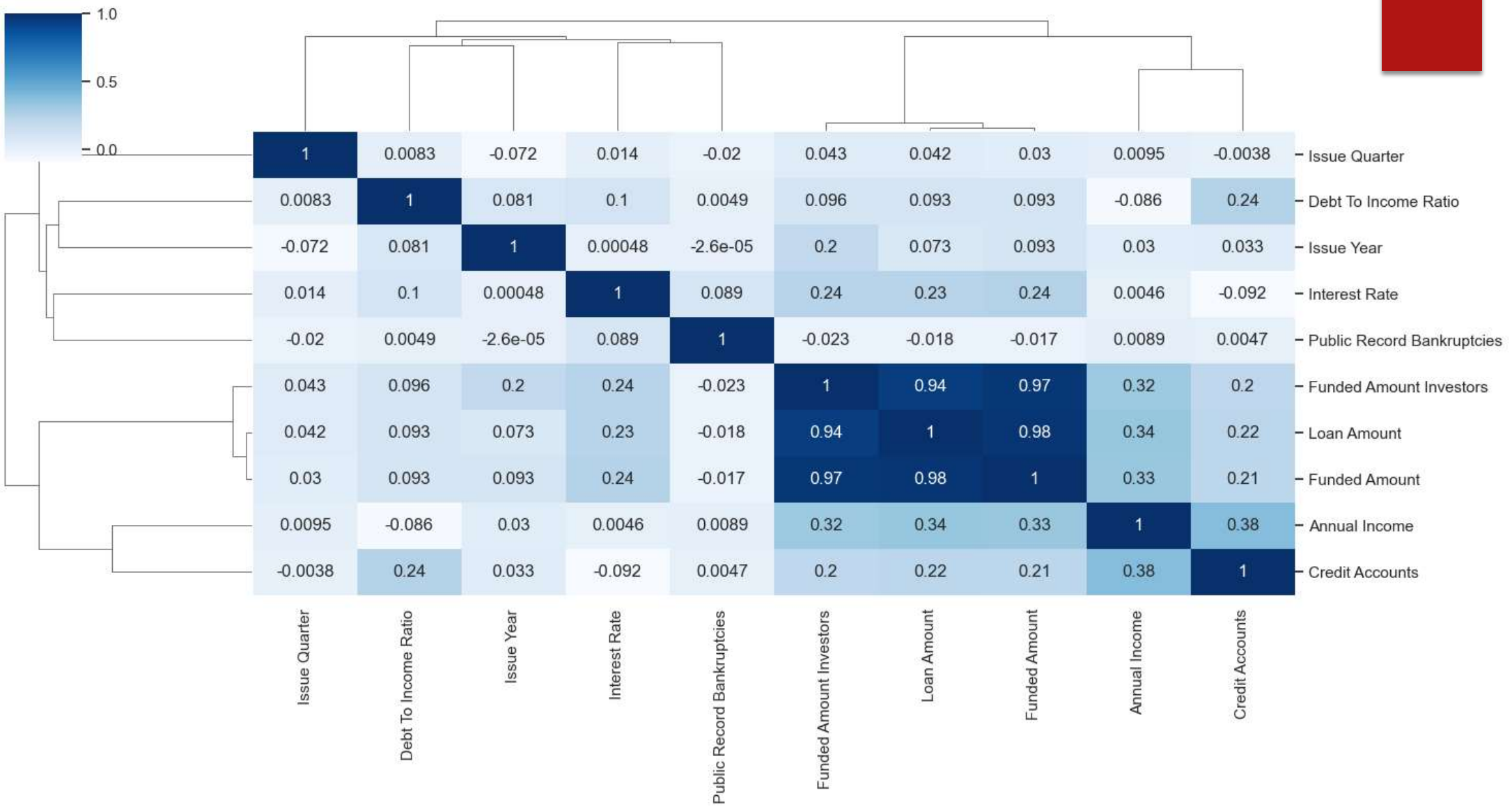
- 0-40K has the highest Charged Off of Annual Income Bucket and Ratio within the bucket as well.



- 0-40K has the highest Charged Off of Annual Income Bucket and Ratio within the bucket as well.
- The Charged Off Ratio of all the custormers within the Loan Amount Bucket of 15K-above is at the highest Charged Off.

- 0-40K has the highest Charged Off of Annual Income Bucket and Ratio within the bucket as well.




1.0
0.5
0.0
Issue Quarter
Debt To Income Ratio
Issue Year
Interest Rate
Public Record Bankruptcies
Funded Amount Investors
Loan Amount
Funded Amount
Annual Income
Credit Accounts
1 0.0083 -0.072 0.014 -0.02 0.043 0.042 0.03 0.0095 -0.0038
0.0083 1 0.081 0.1 0.0049 0.096 0.093 0.093 -0.086 0.24
-0.072 0.081 1 0.00048 -2.6e-05 0.2 0.073 0.093 0.03 0.033
0.014 0.1 0.00048 1 0.089 0.24 0.23 0.24 0.0046 -0.092
-0.02 0.0049 -2.6e-05 0.089 1 -0.023 -0.018 -0.017 0.0089 0.0047
0.043 0.096 0.2 0.24 -0.023 1 0.94 0.97 0.32 0.2
0.042 0.093 0.073 0.23 -0.018 0.94 1 0.98 0.34 0.22
0.03 0.093 0.093 0.24 -0.017 0.97 0.98 1 0.33 0.21
0.0095 -0.086 0.03 0.0046 0.0089 0.32 0.34 0.33 1 0.38
-0.0038 0.24 0.033 -0.092 0.0047 0.2 0.22 0.21 0.38 1

- ## NEGATIVE CORRELATION

Loan Amount has negative correlation with Public Record Bankruptcies.
Annual Income has negative correlation with Debt To Income Ratio.

- ## STRONG CORRELATION

Term has strong correlation with Loan Amount.
Term has strong correlation with Interest Rate.
Annual Income has strong correlation with Loan Amount.

- ## WEAK CORRELATION

Public Record Bankruptcies has weak correlation with maximum number of fields.