Lending Club Case Study

Analyzed & Presented By:
Reetesh Nigam
Prakher Singhal

Abstract

- *Lending Club* company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures
- Lending loans to *risky* applicants is the largest source of financial loss (called credit loss) when the borrower refuses to pay 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

Business Objective

- Lending Club wants to understand the *driving factors* (or driver variables) which are strong *indicators of default*
- The company can utilize this knowledge for its portfolio and risk assessment

Loan Data Set

- Loan data is provided for all loans issued through the time period 2007 to 2011
- To understanding the types of variables and their significance, a Data Dictionary is provided that describes the meaning of these variables

Analysis Approach

Data Understanding	Data Cleaning	Univariate Analysis	Segmented Univariate Analysis	Bivariate Analysis	Recommendations
Analyze the Raw data and its variables. Use Data Dictionary to understand the past loan data	Remove Nulls and unwanted strings. Standardize data to focus on data set that matters in the analysis.	Single variable analysis with the help of data distribution charts.	Analyze continuous and categorical variables.	Analyze two variables' trends and its relation-ship to reveal insights.	Conclude the analysis by suggesting variables that have resulted in loan defaults.

Data Understanding

- Historical loan data of *Fully Paid*, *Current*, and *Charged Off* category customers from year 2007 to 2011
- Total rows = 39717, Total columns = 111
- Example of Categorical Variables:
  - Grade, Sub-Grade, Term, Employment Length, Address State, etc.
- Example of Numerical Variables:
  - Annual Income, Loan amount, DTI, Monthly installment, etc.
- Data Dictionary are used to understand each column semantic
- Mixed data type is present in the data (Object, int, and Float64)

Data Cleaning and Imputation

- Dropped variables with ALL NULL values
- Dropped variables with >75% NULL because imputing them can lead to data inconsistencies
- Dropped variables with only single value because they are contacts and aren't helping in our analysis
- Dropped variables which are of type Personal details, URL, Post loan data, Payment dates, Last payment date, etc.
- Filled emp_length with Median value to represent the 50^{th} Quantile
- Dropped ~900 Rows with NULL values out of total 39717 rows to level all variable rows with data
- ***Final Data Frame: Rows: 38970 x Columns: 29***

Data Standardization

- Removed Leading and Training spaces from *term* variable
- Replaced character *(%)* from *Interest Rate* and *revol_until* variable with empty character and converted to float Data Type.
- No changes are made in variables which are found fit for analysis.
  - Example: Grade, sub-grade, employment length, loan status, address state, etc.
- Converted *Loan Issued Date* and *Earliest Credit Line* variable from Object → DATE TIME format

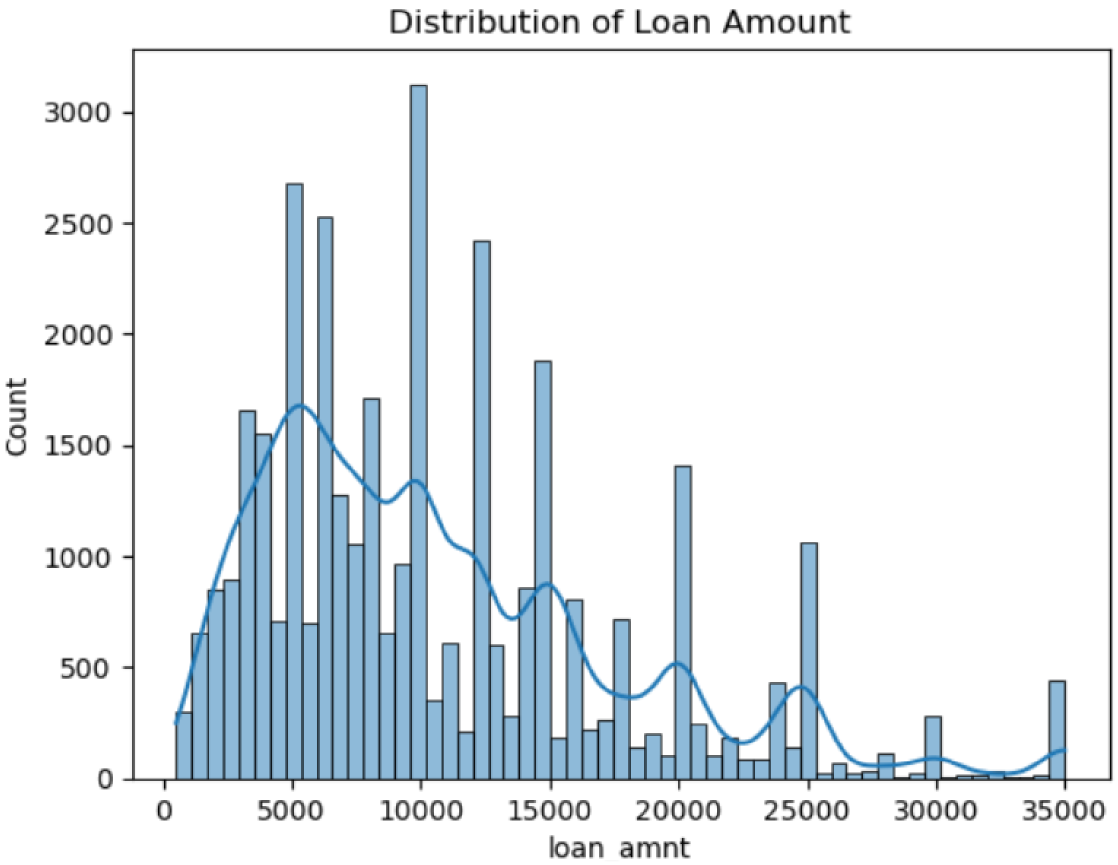
Derived Metrics

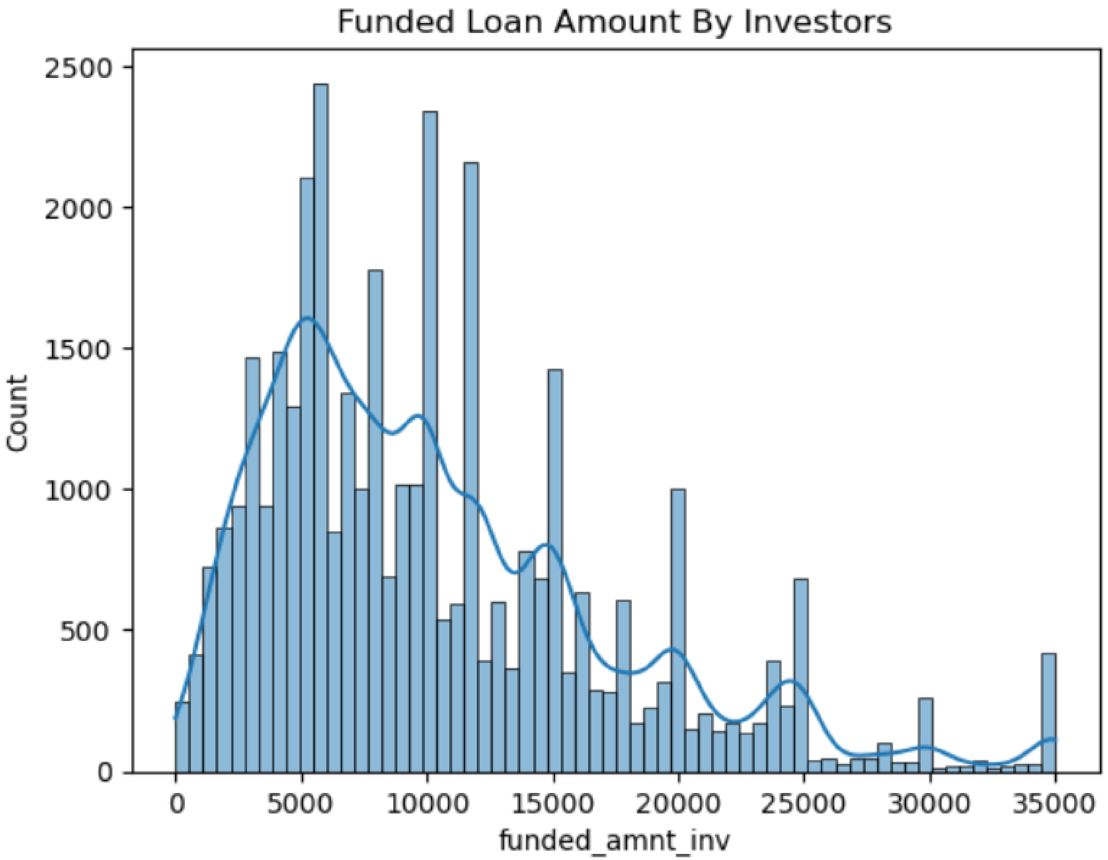
- Created following new columns that can help in our analysis:
  - Loan Issued Date *Month*
  - Loan Issued Date *Year*
  - Earliest Credit Line *Month*
  - Earliest Credit Line *Year*
  - Loan Amount Group (*7 bins*)
  - Funded Amount Intestors Group (*7 bins*)

Data Filtering

- Keeping "*Charged Off*" and "*Fully Paid*" loan Status data and Filtering off "Current" type data
- Reason: We need to analyze variables which can indicate a possible defaulters before lending a loan

Univariate Analysis

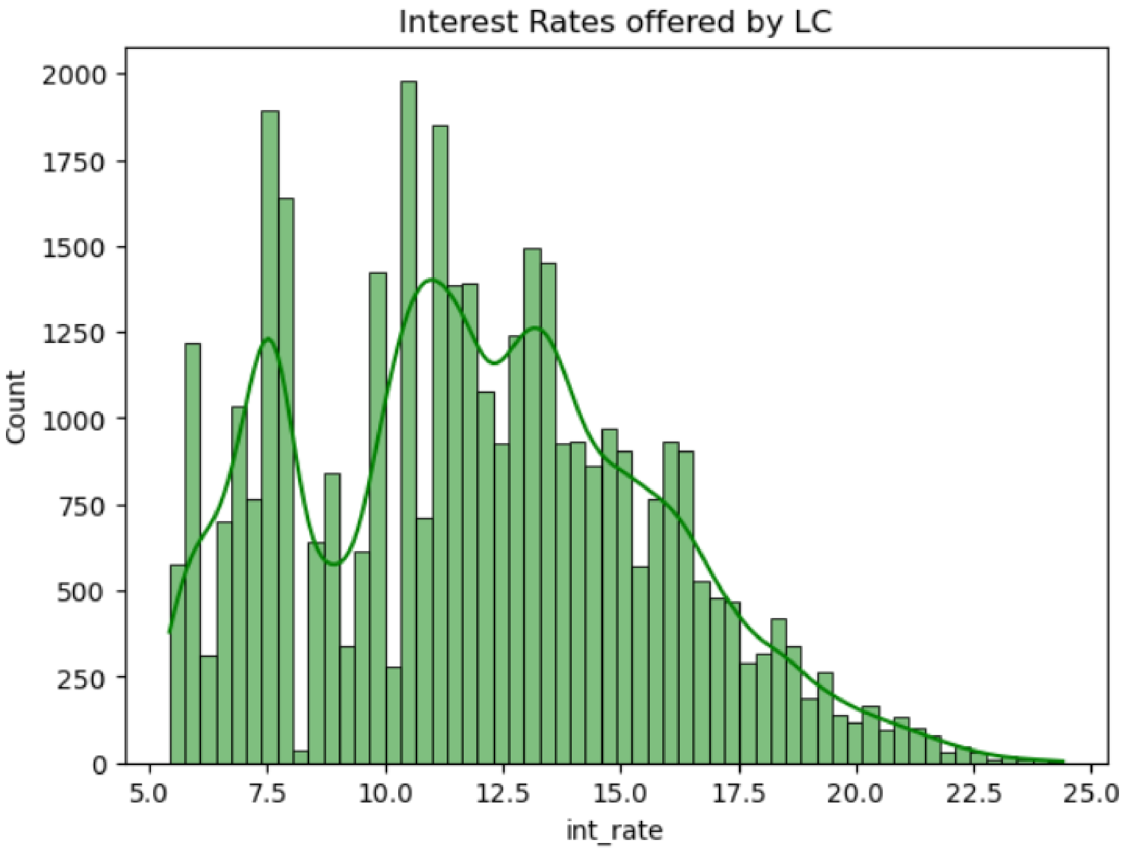


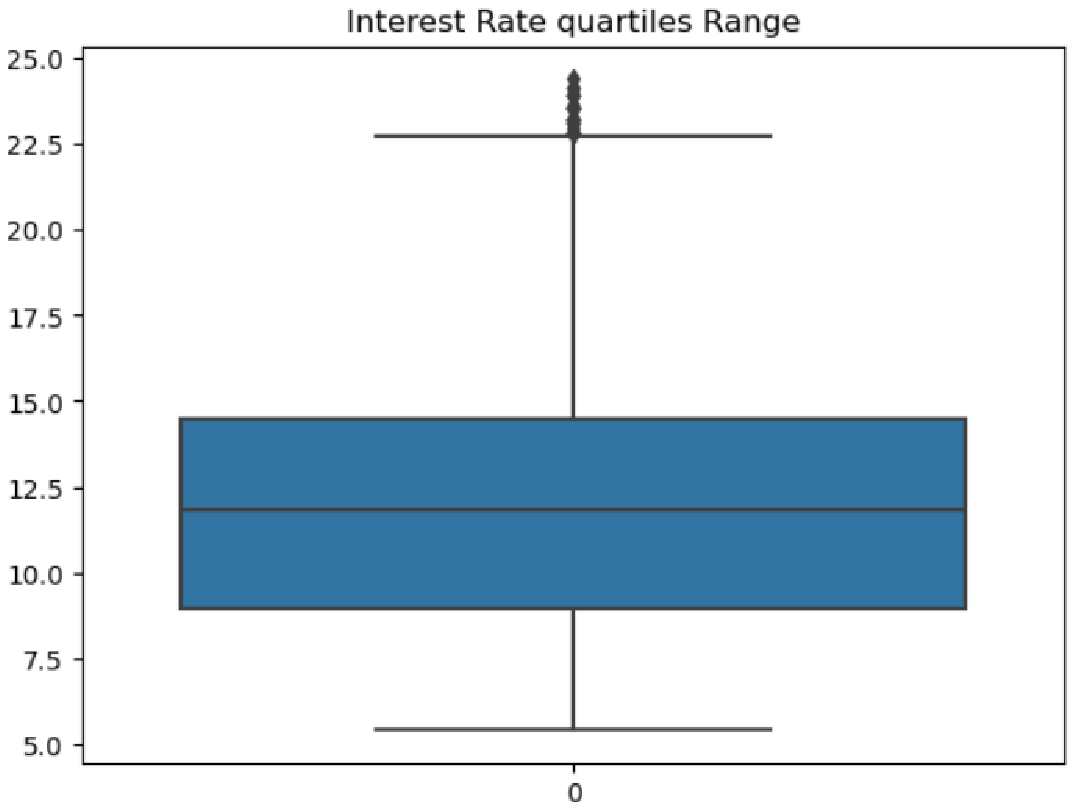


Observations:

- Funded amount distribution is very close the applied loan distribution. Both data shows median around 9600.
- This is indicating that applied loan amount are mostly approved by lender

Univariate Analysis (contd.)

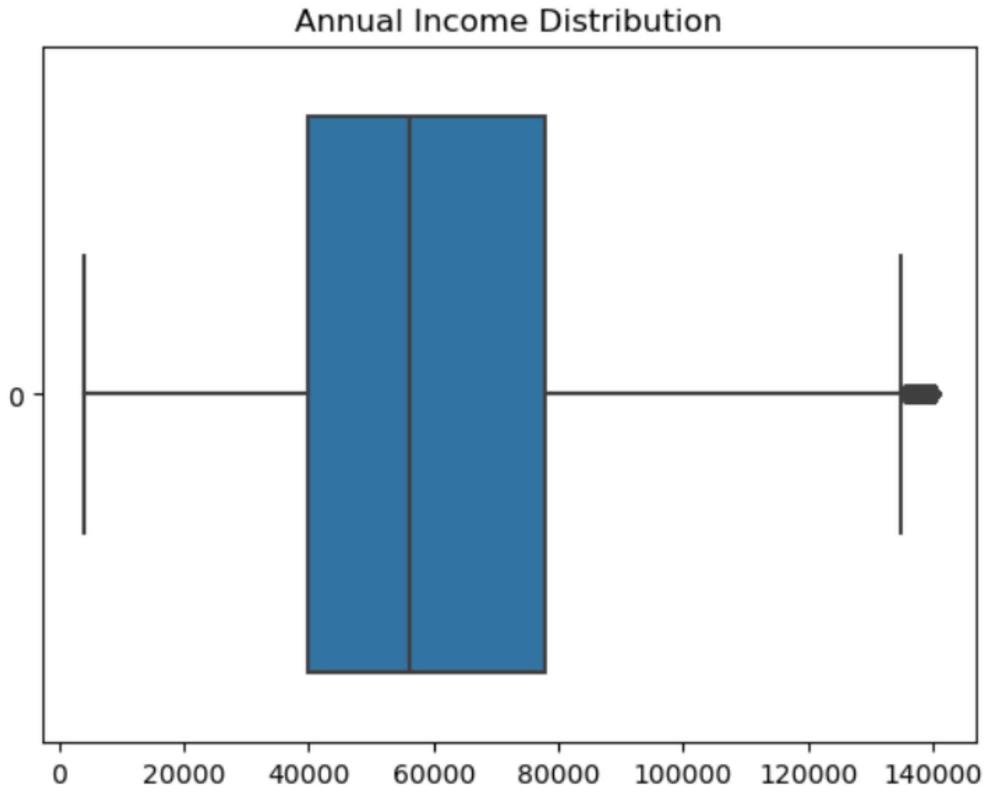


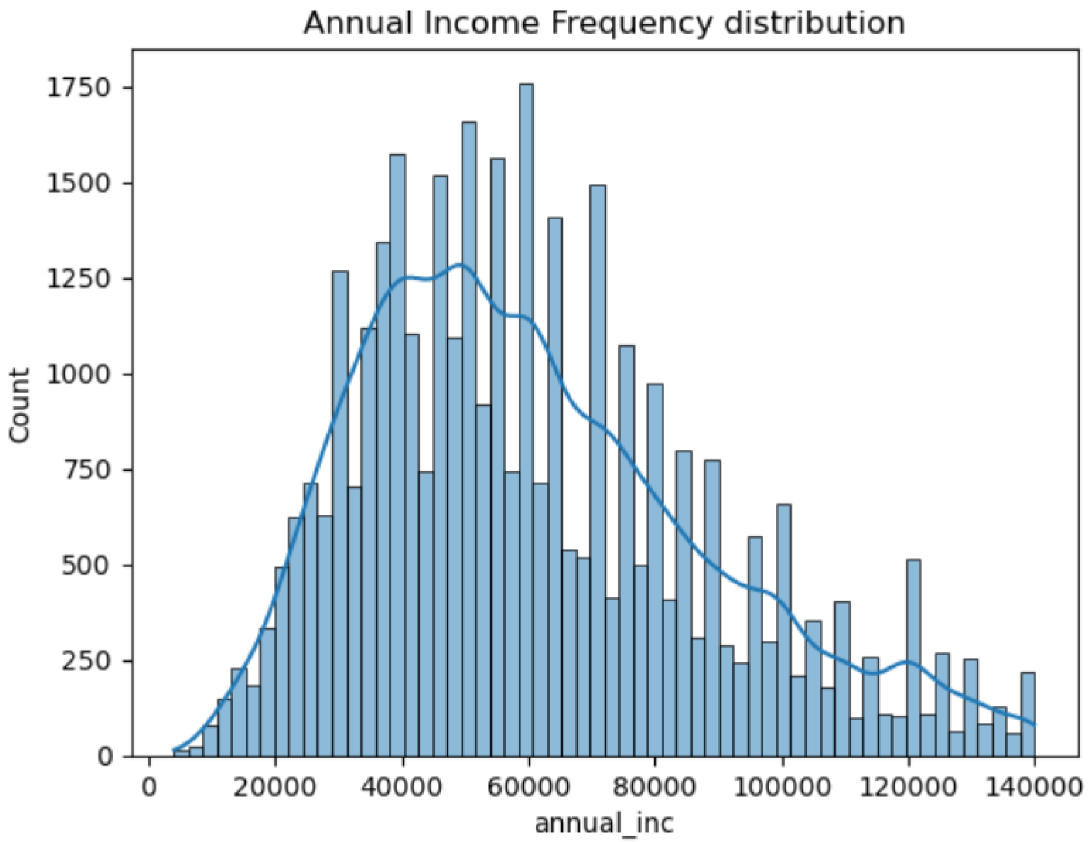


Observations:

- For defaulters, ranges of interest rates offered by LC are:
  - 25th – 75th Percentile loans were offered at an interest rate between: 9% to 14.5%
  - Some loans were offered at lower interest rate: <9%
  - Minimal loans were also offered at higher interest rate: 22.5%

Univariate Analysis (contd.)

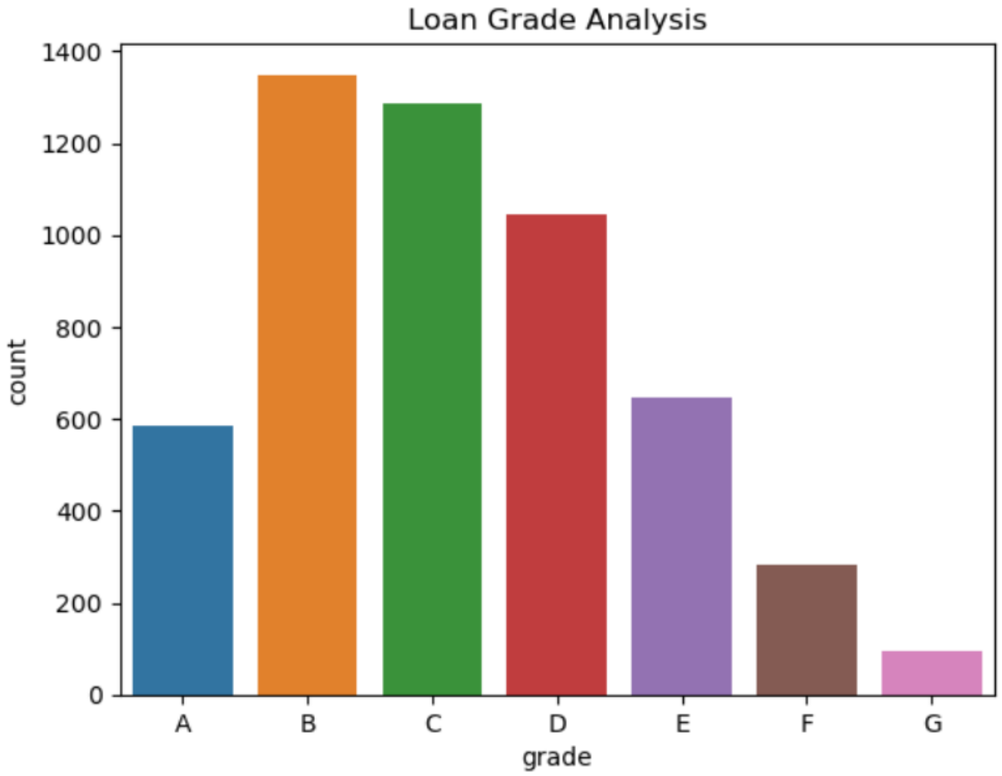


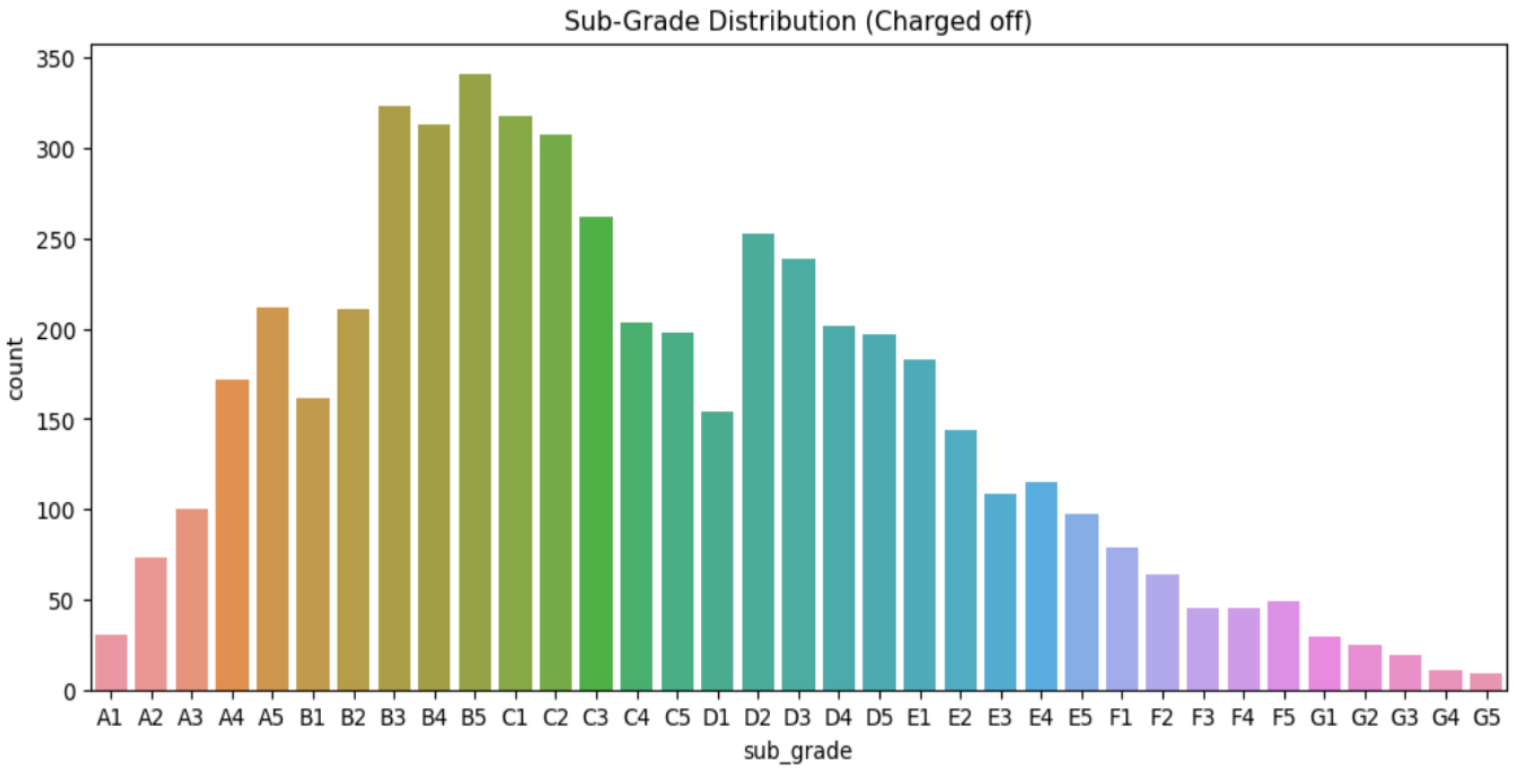


Observations:

- Annual Income boxplot indicates most salary are in ranges from 4000 – 8000 (post outlier removal)
- Maximum 1750 borrowers had salary reported as 60000
- This range is also confirmed by income distribution chart

Univariate Analysis

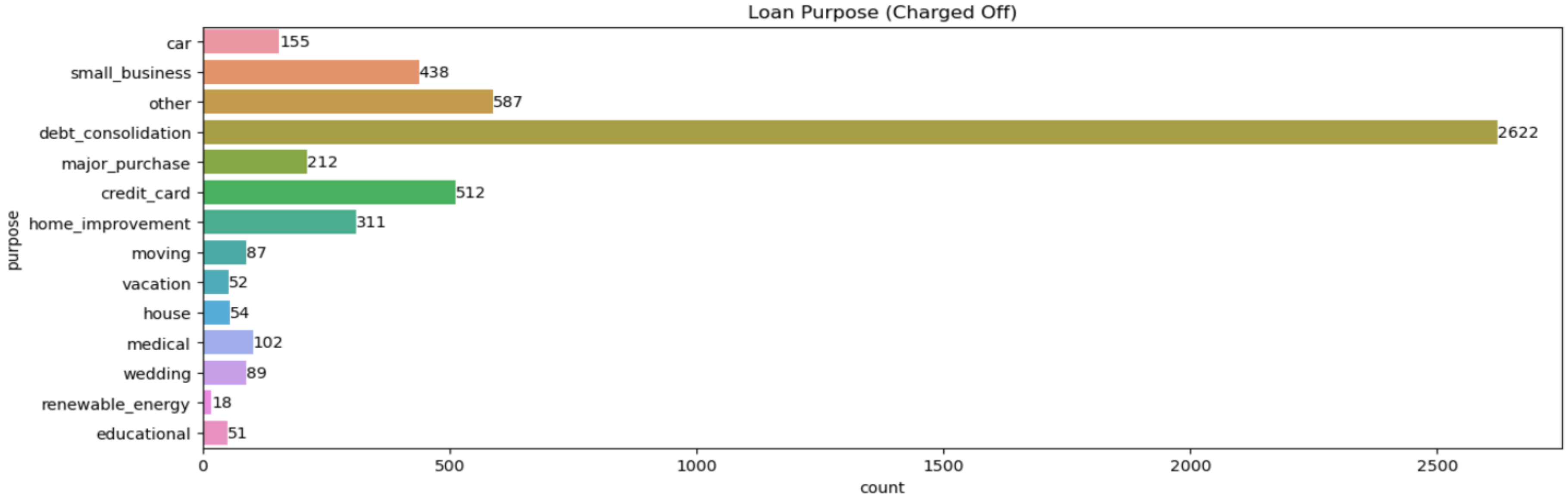




Observations:

- **Grade B, C, and D** have defaulted the most as found in this Univariate analysis
- Sub-categories **B3, B4, B5, C1, C2, C3, D2, D3, D4, and D5** are the most defaulters

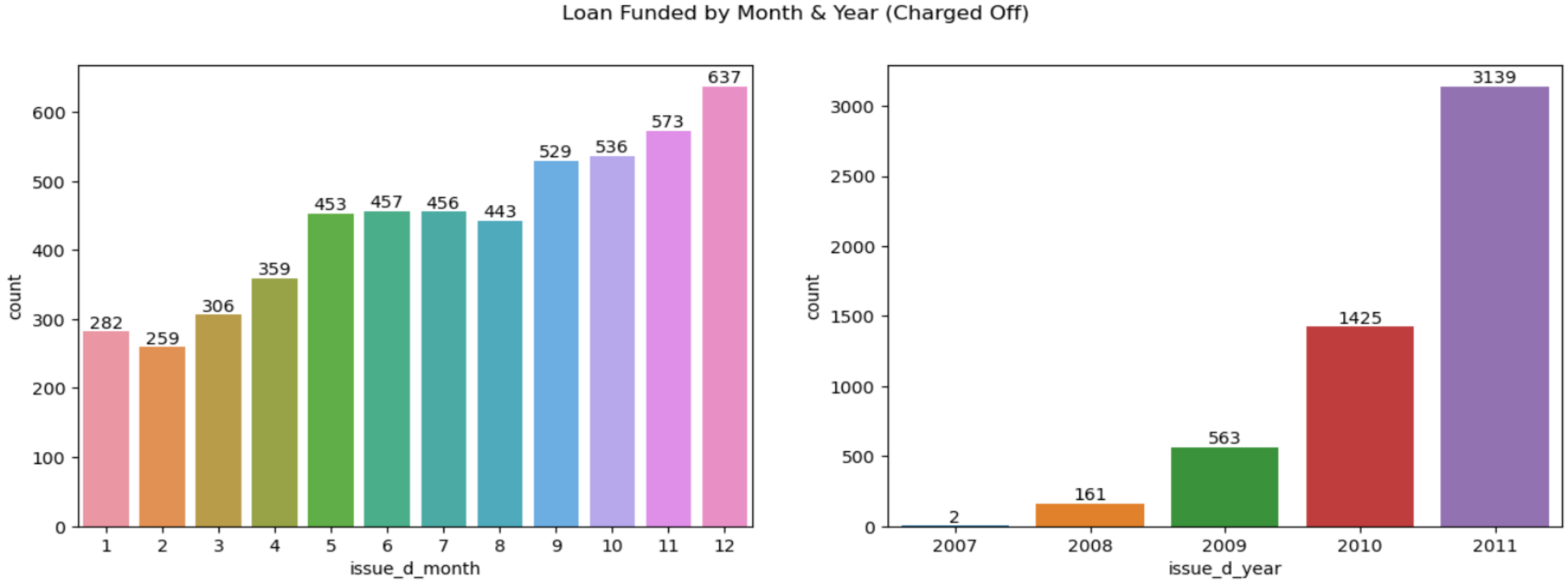
Univariate Analysis

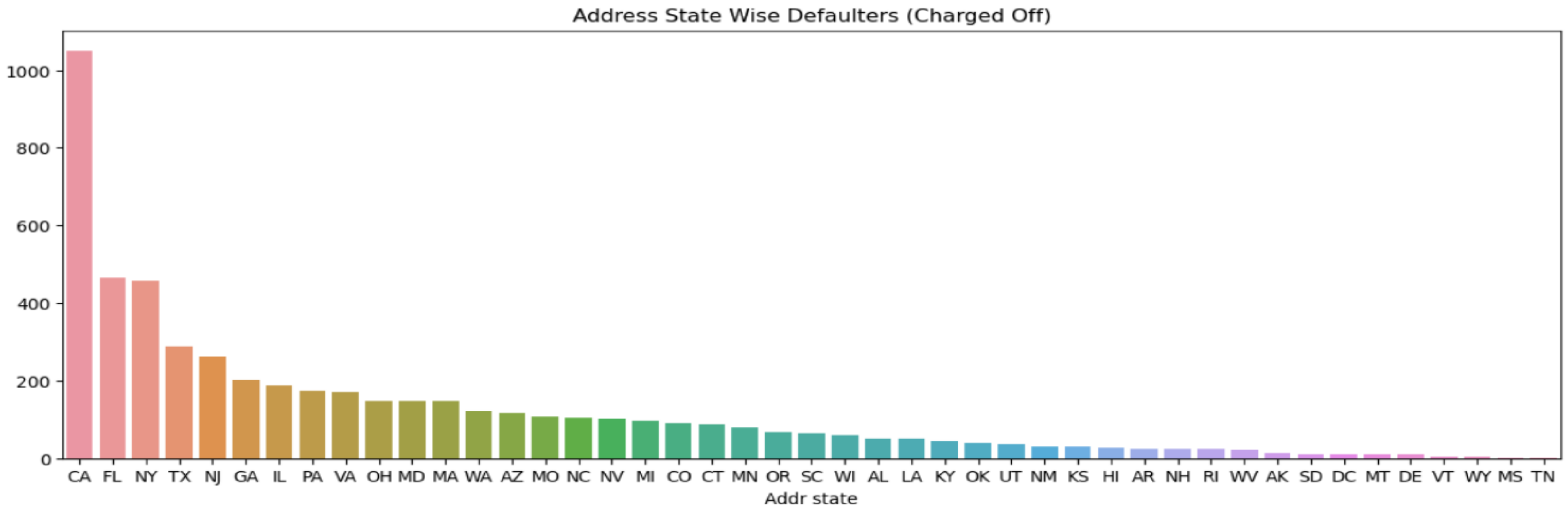


Observations:

- It's strongly indicating that **10+ years** of experience borrowers are the most defaulters
- Debt consolidation is the MAJOR reason provided for taking the loan and borrower defaulting it

Univariate Analysis

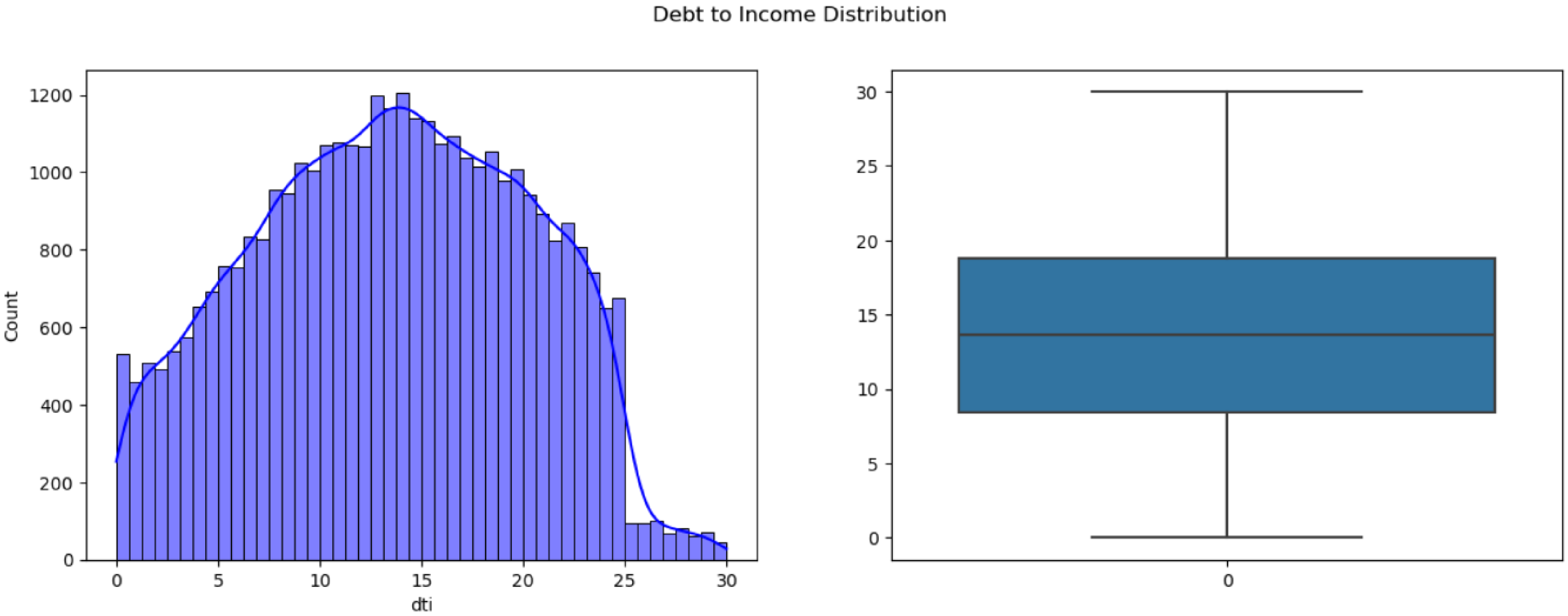




Observations:

- Trend of issuing loan that are defaulted have grown significantly from 2007 - 2011.
- **4th Quarter** of any year is where most of the loans (Charged off) were funded in LC.
- Most Defaulter are from state **CA, NY, and FL**

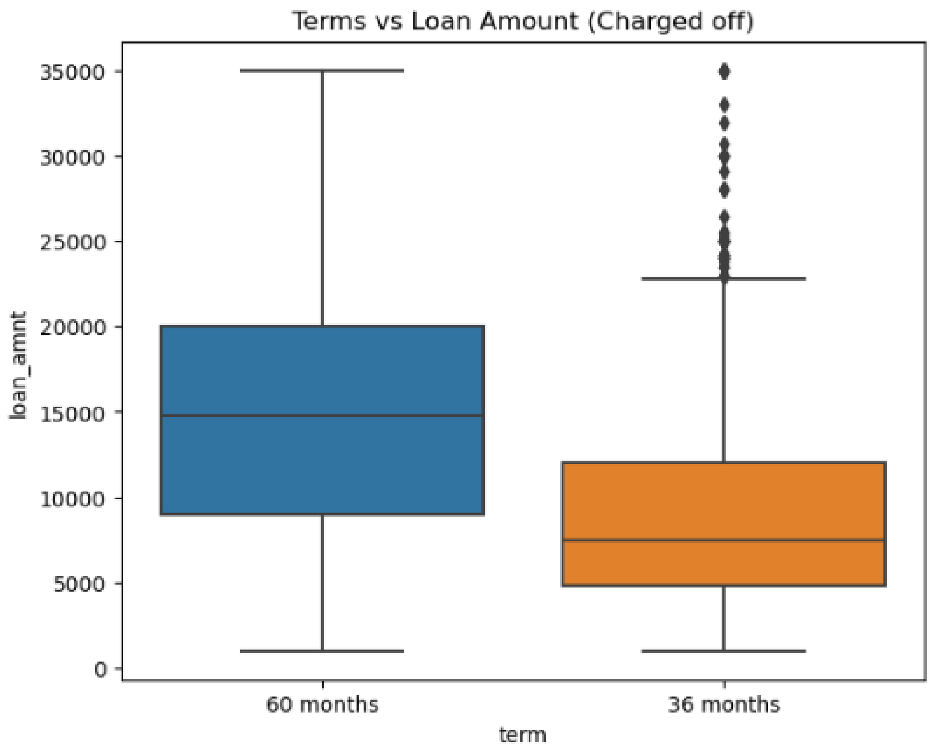
Univariate Analysis

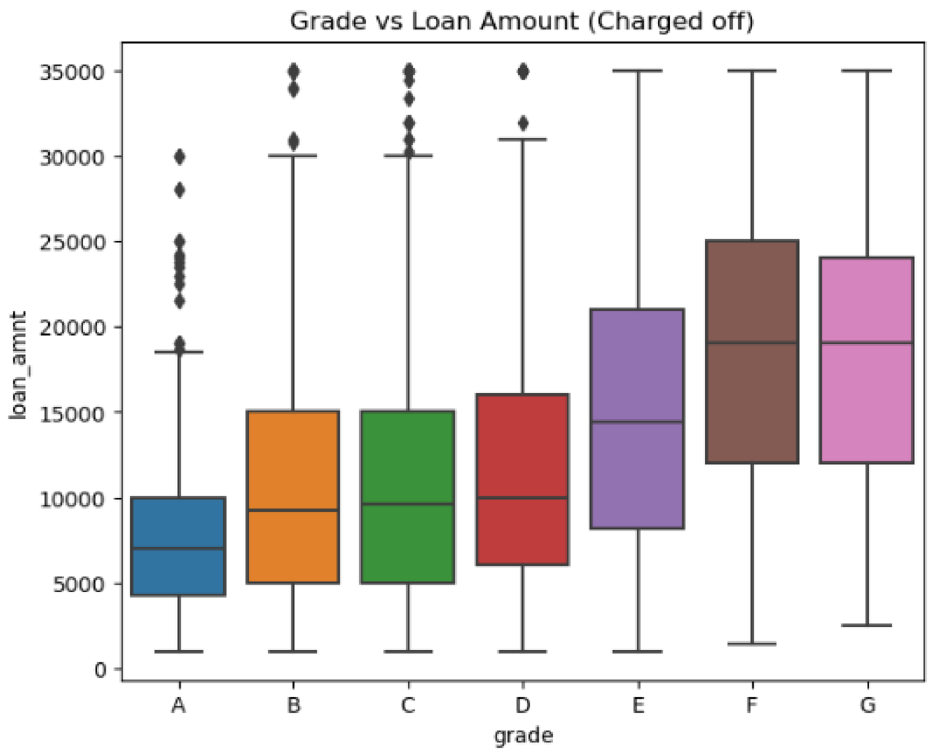


Observations:

- DTI ratio is **<30%** which is **a good indicator** and borrowers are not considered Risky to Lend the money.
- So, it seems there are **other factors/variables** which needs to be analyzed for loan Charged off analysis

Segmented Univariate Analysis

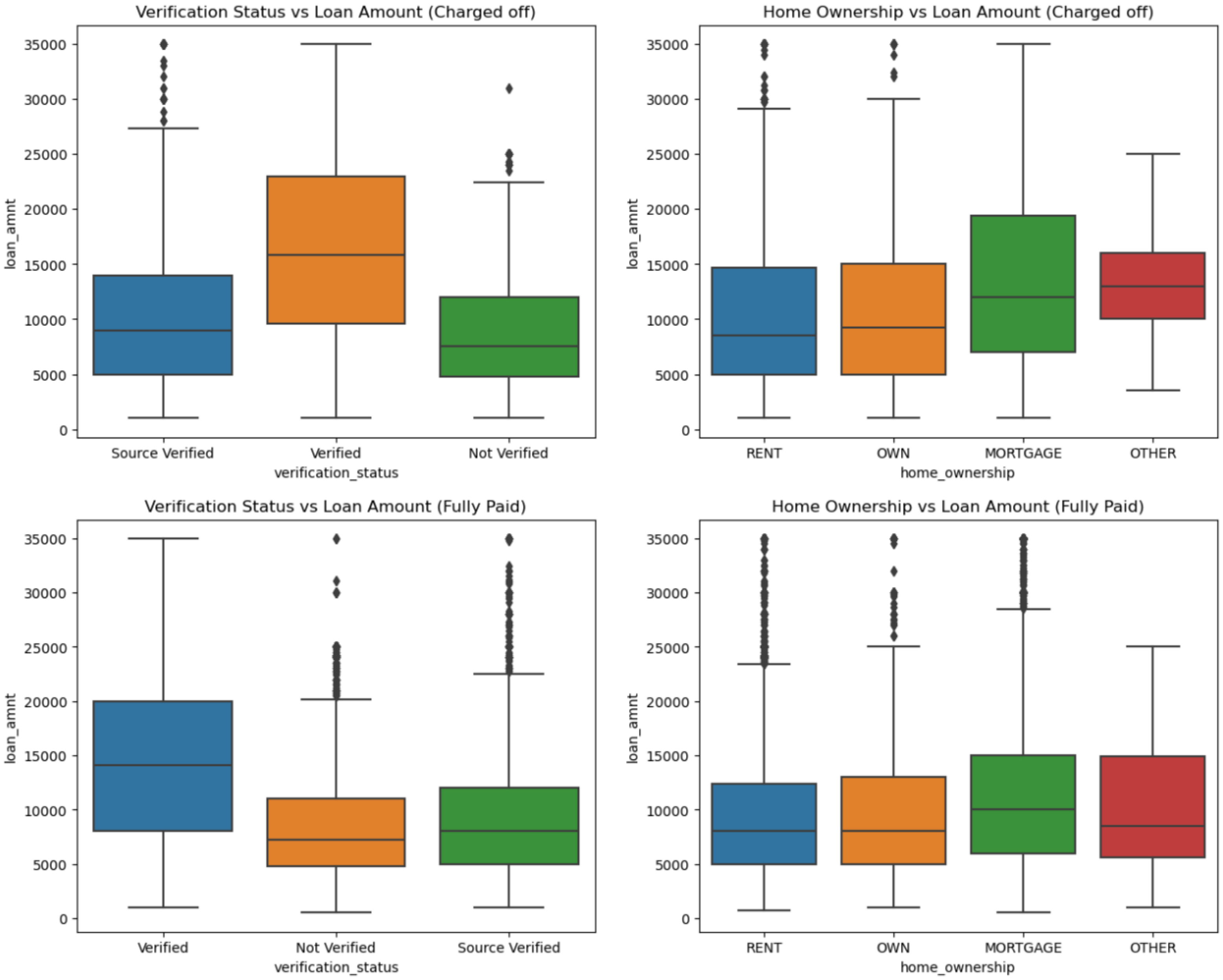




Observations:

- Higher amount loans are taken for longer term which had defaulted
- Grade F and G had taken the large amount loans and are MOST defaulters

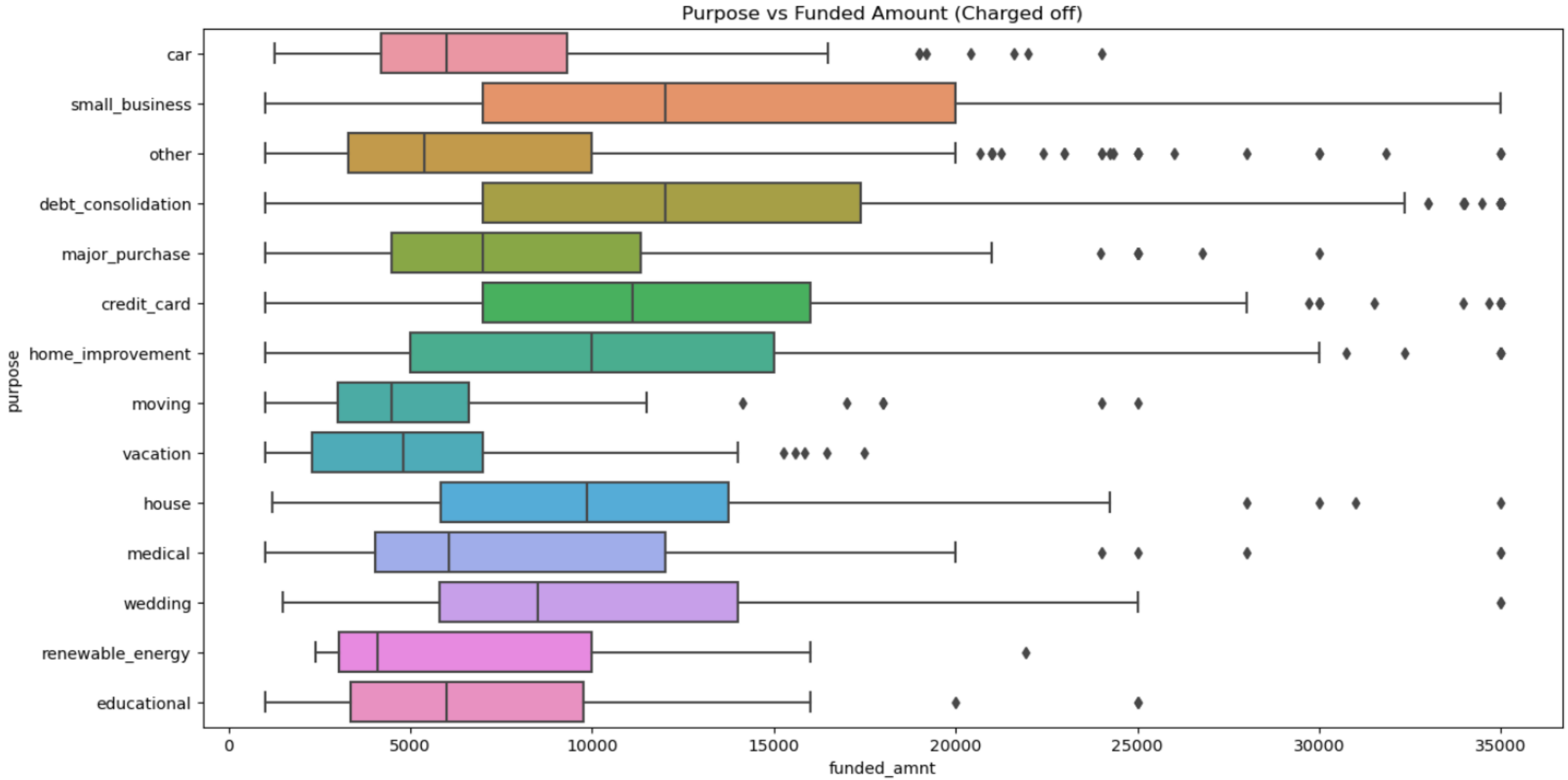
Segmented Univariate Analysis



Observations:

- Charged Off loans belonged to MORTGAGE Home Category
- Source verification Status “verified” were the most defaulters
- But this is not true when compared with Fully Paid one.
- Hence, Verification status is NOT a clear indication of the defaulters.

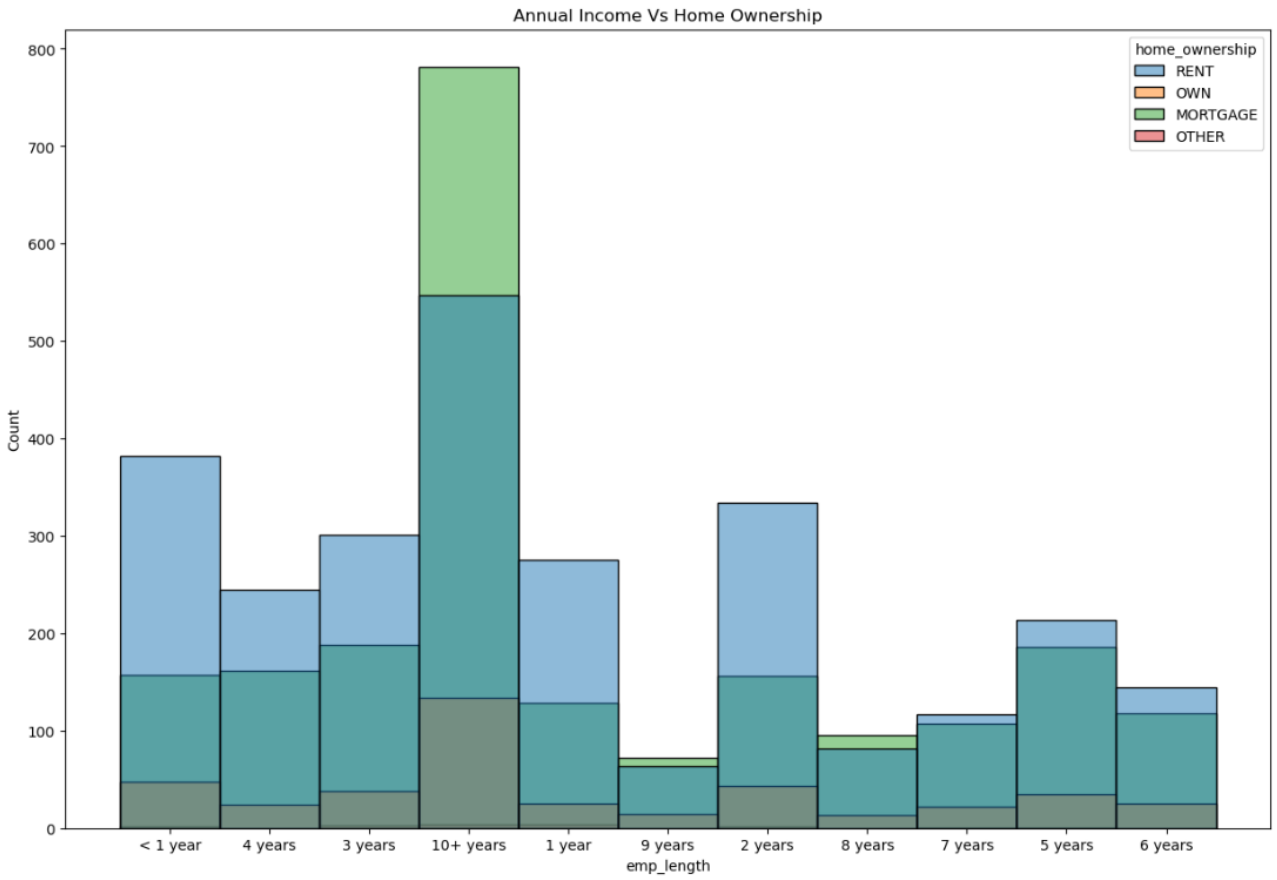
Segmented Univariate Analysis

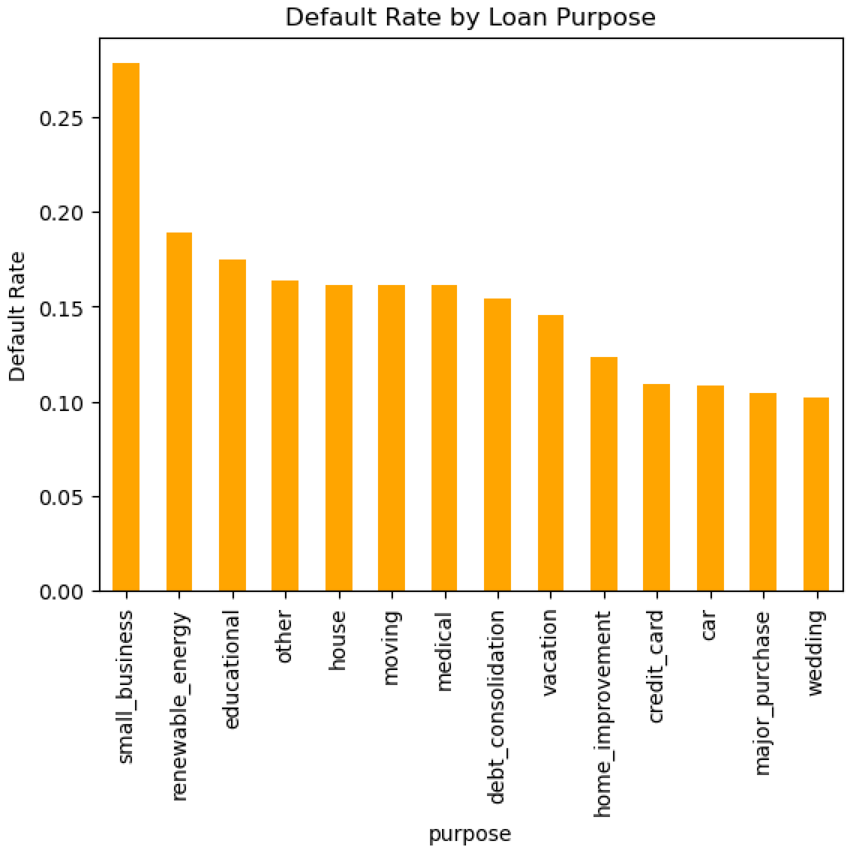


Observations:

- **Small Business** and **Debt Consolidation** Purposes loans have defaulted the most
- LC should limit the loans offered to small businesses purpose or do more verifications

Bivariate Analysis

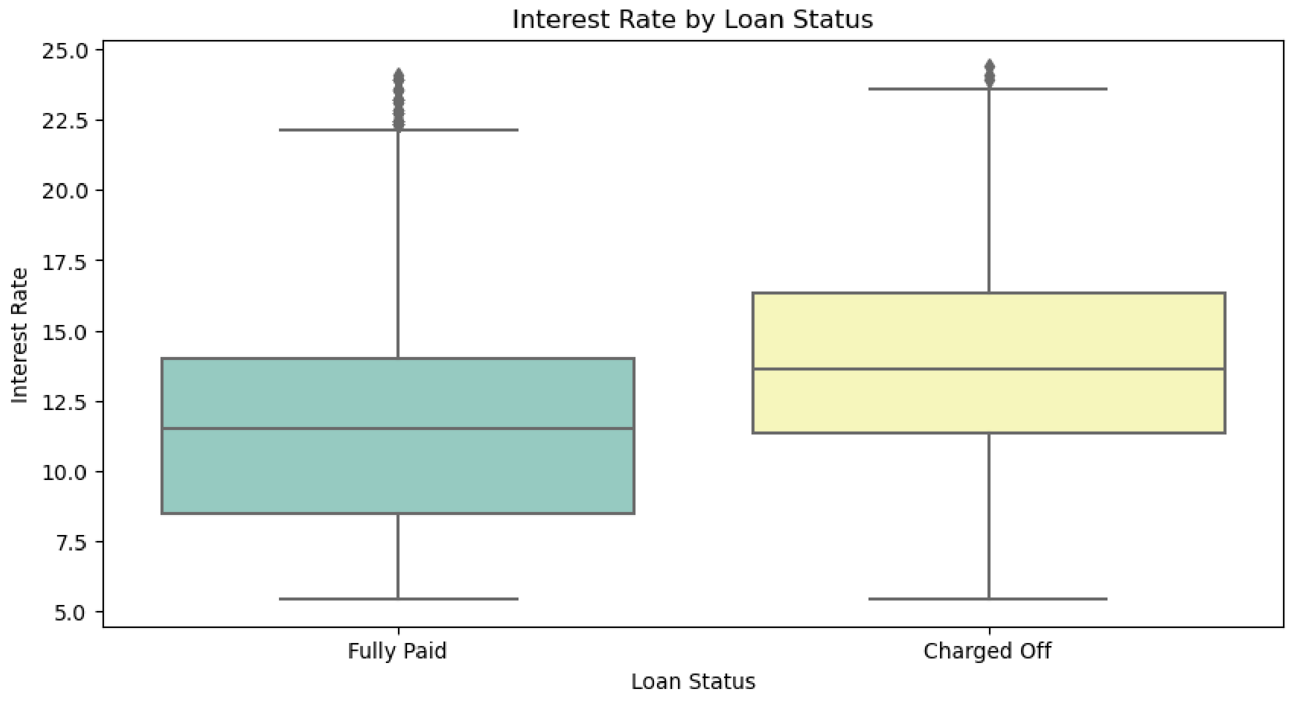


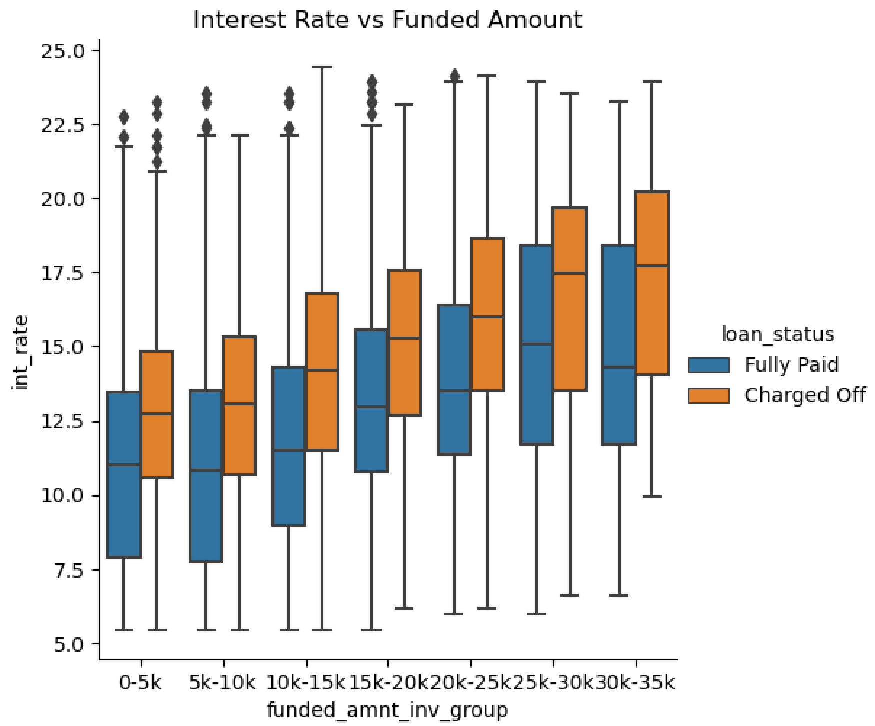


Observations:

- Borrowers, with **all Experience type**, having Home ownership as **RENT** are the also the TOP defaulters.
- Borrowers with **>10 Years of experience** and having Home ownership as **MORTGAGE** are the second most defaulters.
- **Small Businesses** have defaulted the loans a lot more then other Loan purpose categories
- This is a STRONG indicator to consider borrowers who defaulted their loans

Bivariate Analysis

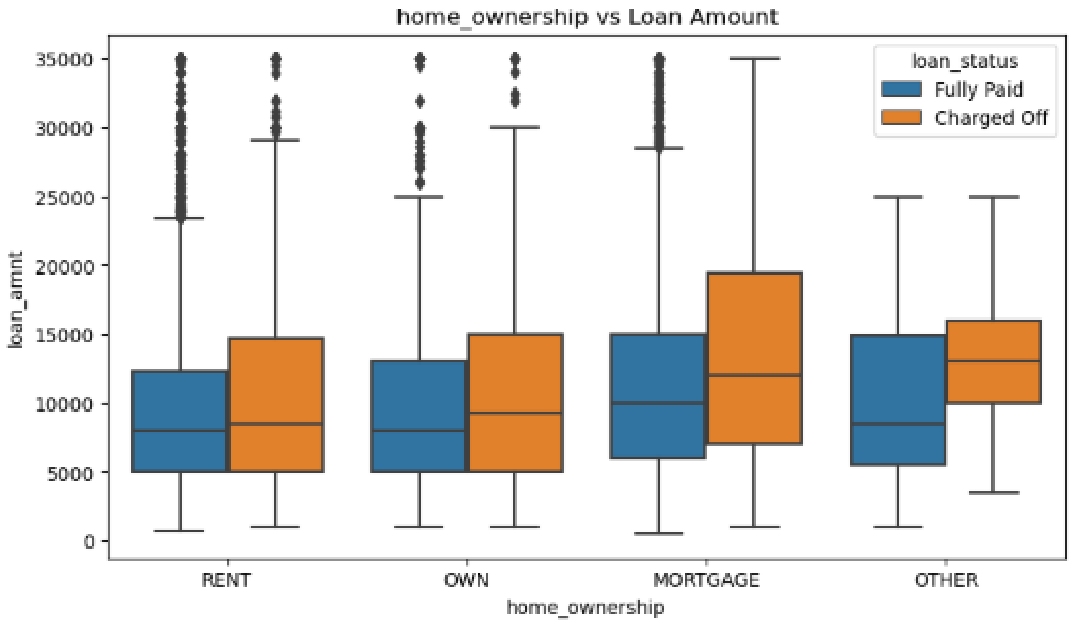


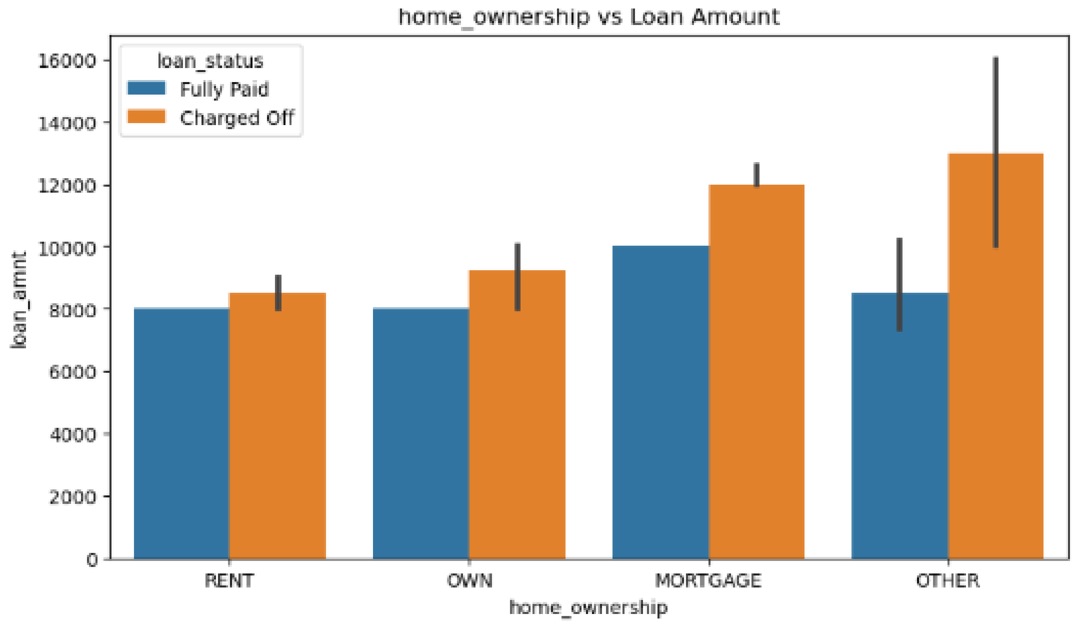


Observations:

- It is observed that when interest rates are higher, defaulter percentile is also increased.
- Higher interest rates are charged for higher loan amount resulting in loan defaulting

Bivariate Analysis

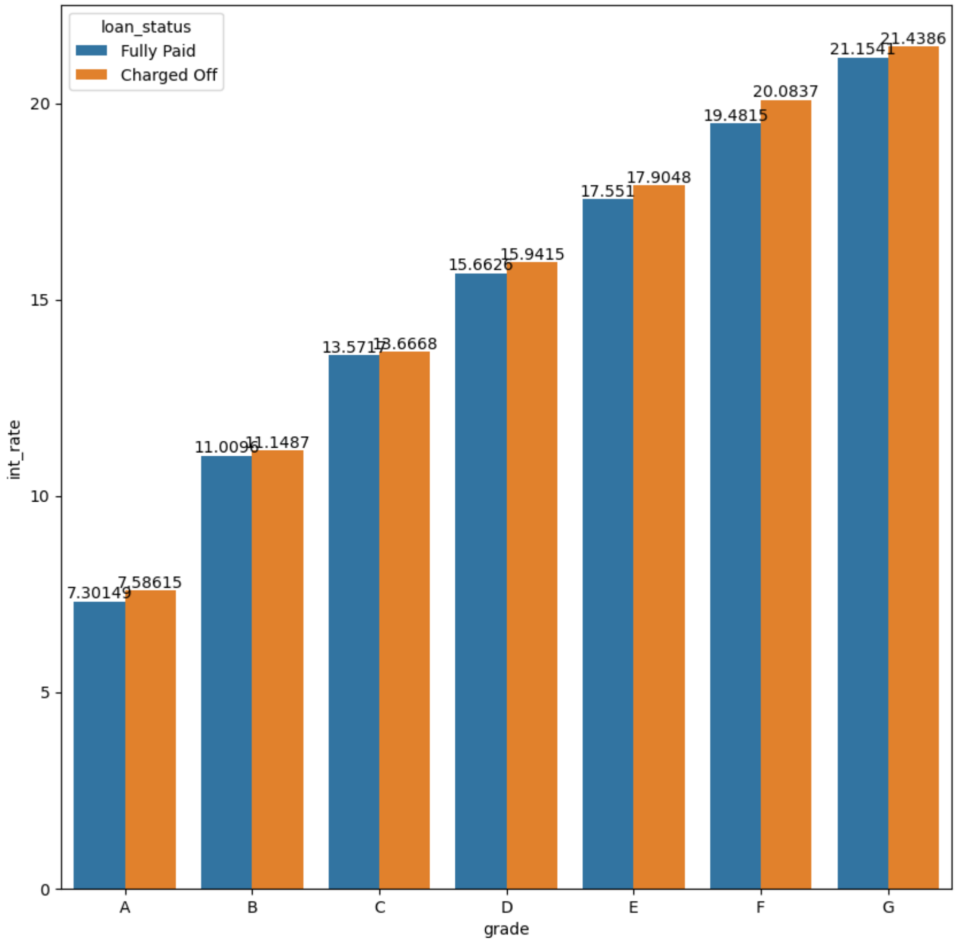


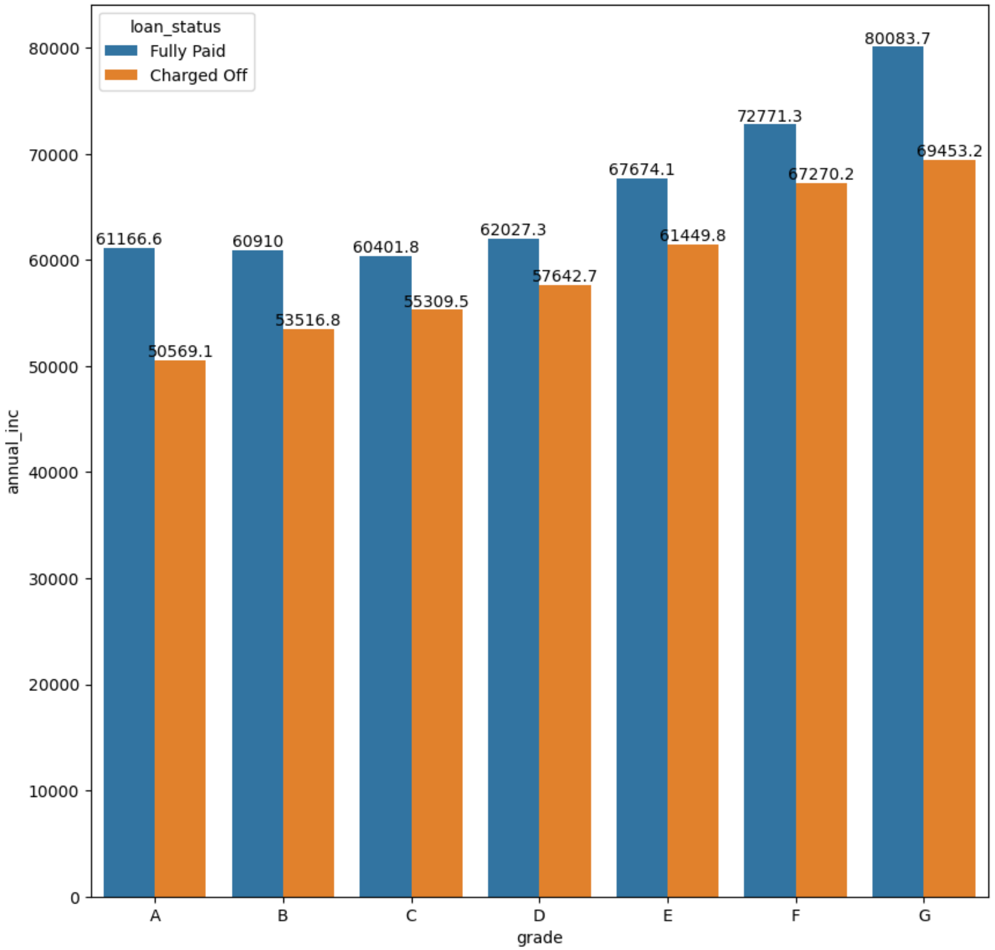


Observations:

- **Higher loan amount** was sanctioned to **MORTGAGE and OTHER** home ownership type
- These categories have resulted in HIGHER loan defaults

Bivariate Analysis





Observations:

- Lower grades loans (E, F, and G) have Very high Interest rates.
- Annual income of Grade (E, F and G) are lower than non-defaulters
- There are high chances that **when High amount loans are funded to Lower income borrowers, they are likely to default**.

Recommendations (Driving Factors behind Loan Defaults)

Our analysis revealed the following Driving factors behind the LC Loan defaults:

Loan Term and Interest Rates:

- 60 months tenure loans have higher interest loans and resulting in higher defaults. Reduce the 60 months tenure or reduce the interest rate for long durations loans.

Loan Grades:

- Significant loan defaulters are found in Lower grade loans (G, F, E). Reduce/Stop issuing lower grade loans for further analysis on defaulters.

Loan Purpose:

- Borrowers with Loan Purpose “Small Businesses” and “Debt Consolidation” have defaulted more then other purpose category. Stop/Reduce funded loan amount to reduce losses.

Home Ownership:

- “Mortgage and Other” home ownership type are taking higher loans and defaulting the funded loans. LC needs to make a decision here whether to offer loans or reduce the loan amount to below 12000.

Address State:

- CA, FL, and NY state are prone to loan defaults. LC needs to check here why these state are most defaulters.

Employment Length:

- Borrowers with >10 Years are taking higher loans with high Interest rates. Try reducing the interest rates to reduce losses in this category.