

Lending Club Group Case Study

Executive PG Programme in ML & AI – May 2022

MADHUSMITA PAUL - GROUP FACILITATOR
ANISH DESHPANDE



Problem Statement & Approach to solution

To develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

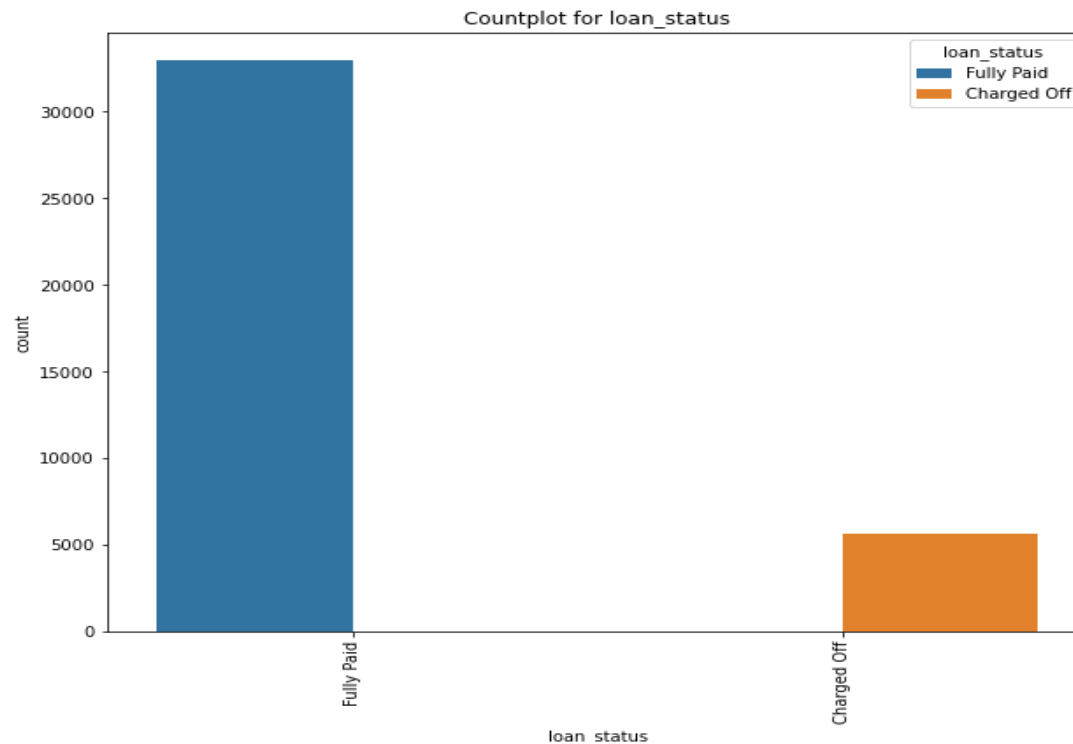
To use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

To understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default

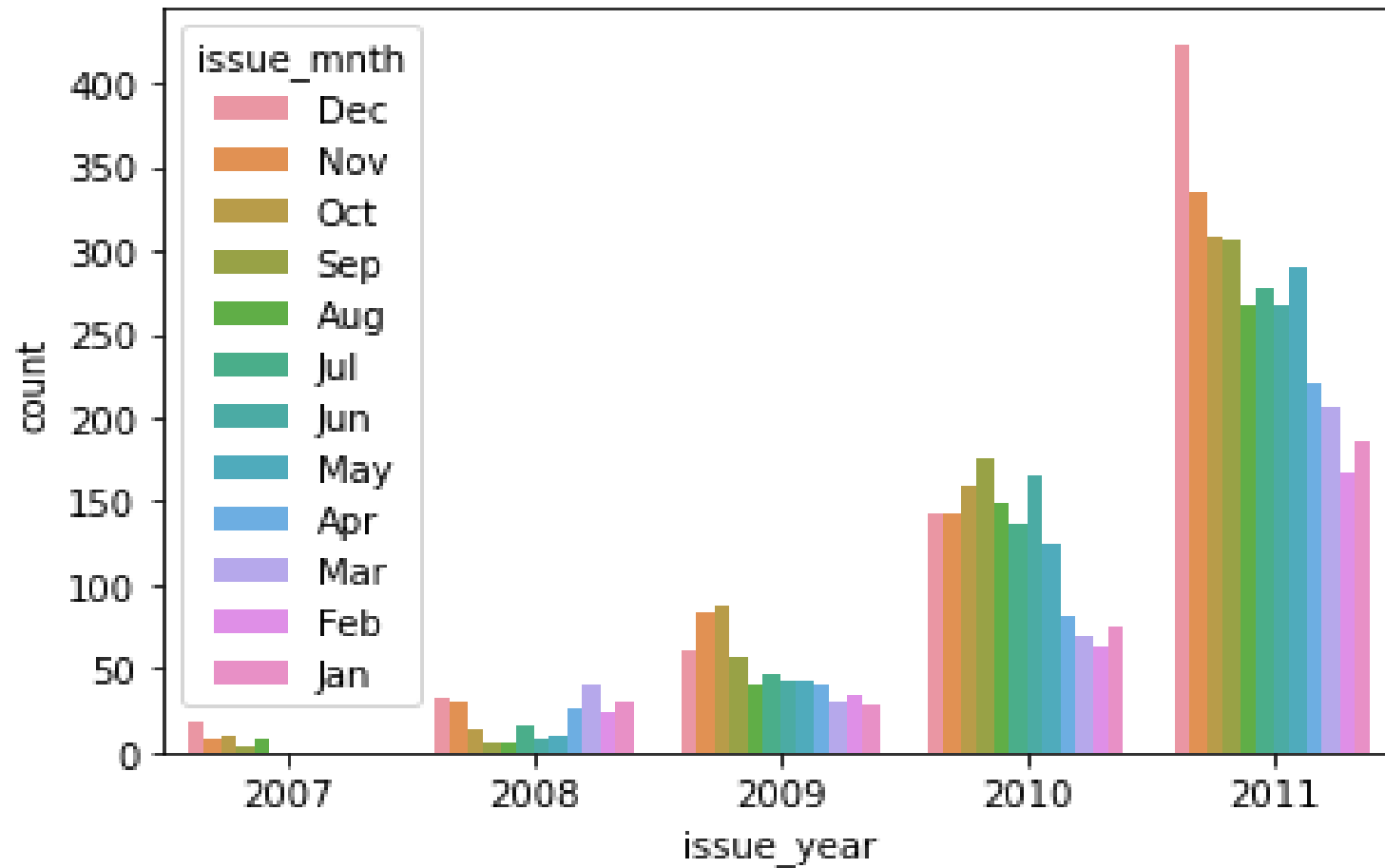
In the following slides we will show how we have used EDA to show the relation between different variables and how they affect the decisions taken by the company

Type of data considered

- The complete **loan data** for all loans issued through the time period **2007 to 2011**.
- We also have a data dictionary which describes the meaning of these variables
- Since there were a lot of redundancies in the data, many of the columns had to be deleted and missing values replaced. After this the final data was considered for analysis
- We have focused on only Fully paid and Charged off as displayed in the graph below



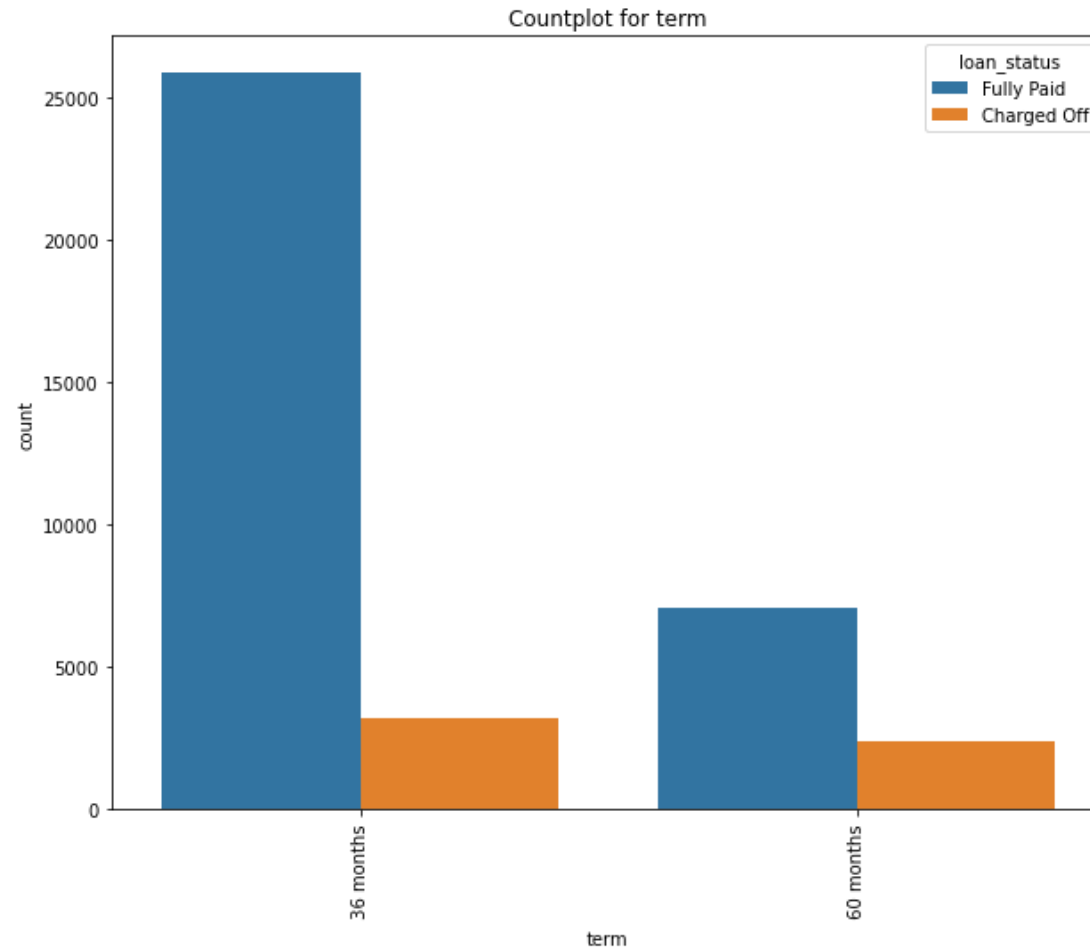
Overview of Charged Off loans issued across all months through 2007 to 2011



Insight :

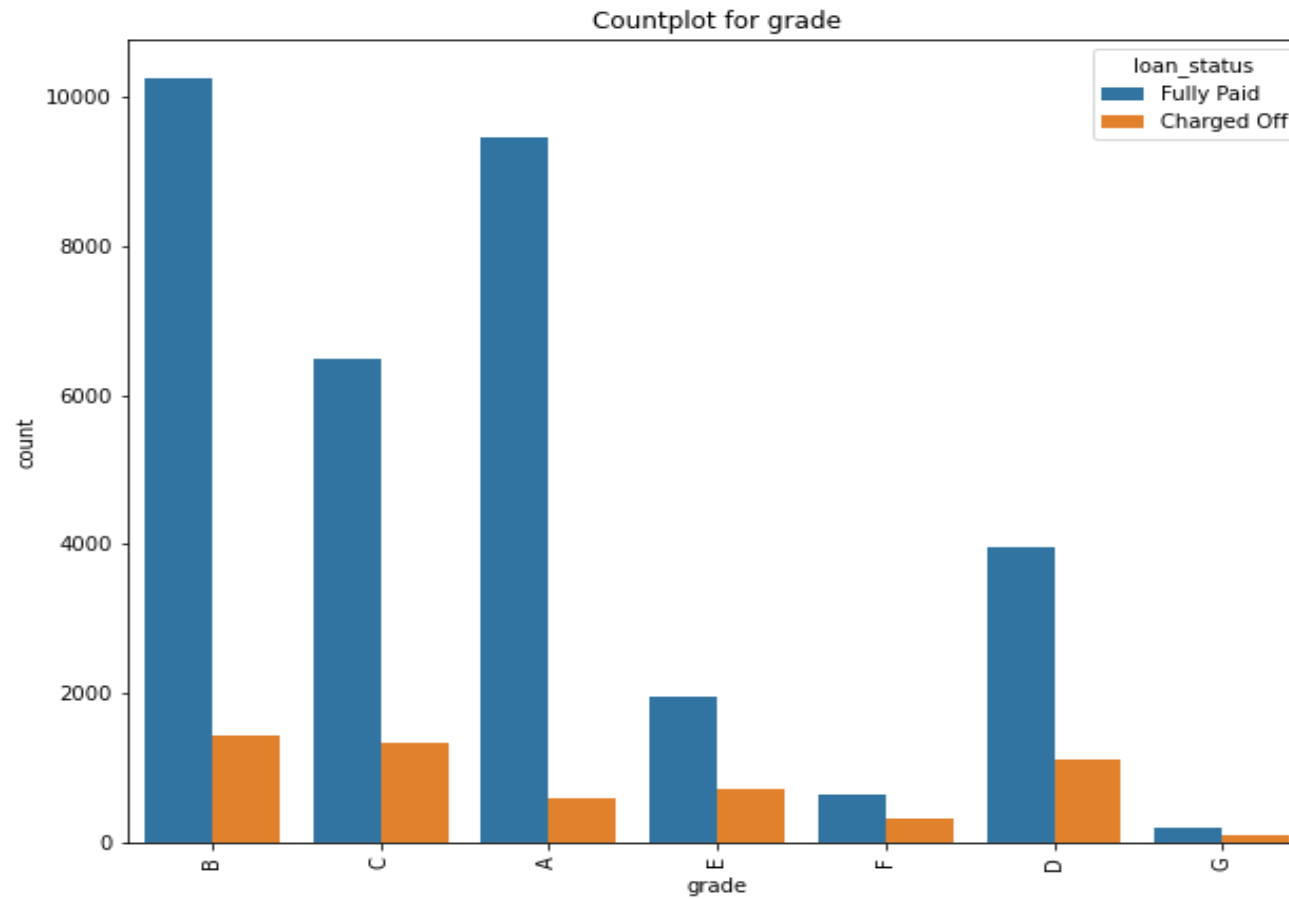
Maximum charged off are the loans that were issued in Dec2011. This might be as seen earlier most loans were issued in 2011 Dec

Variable : Term



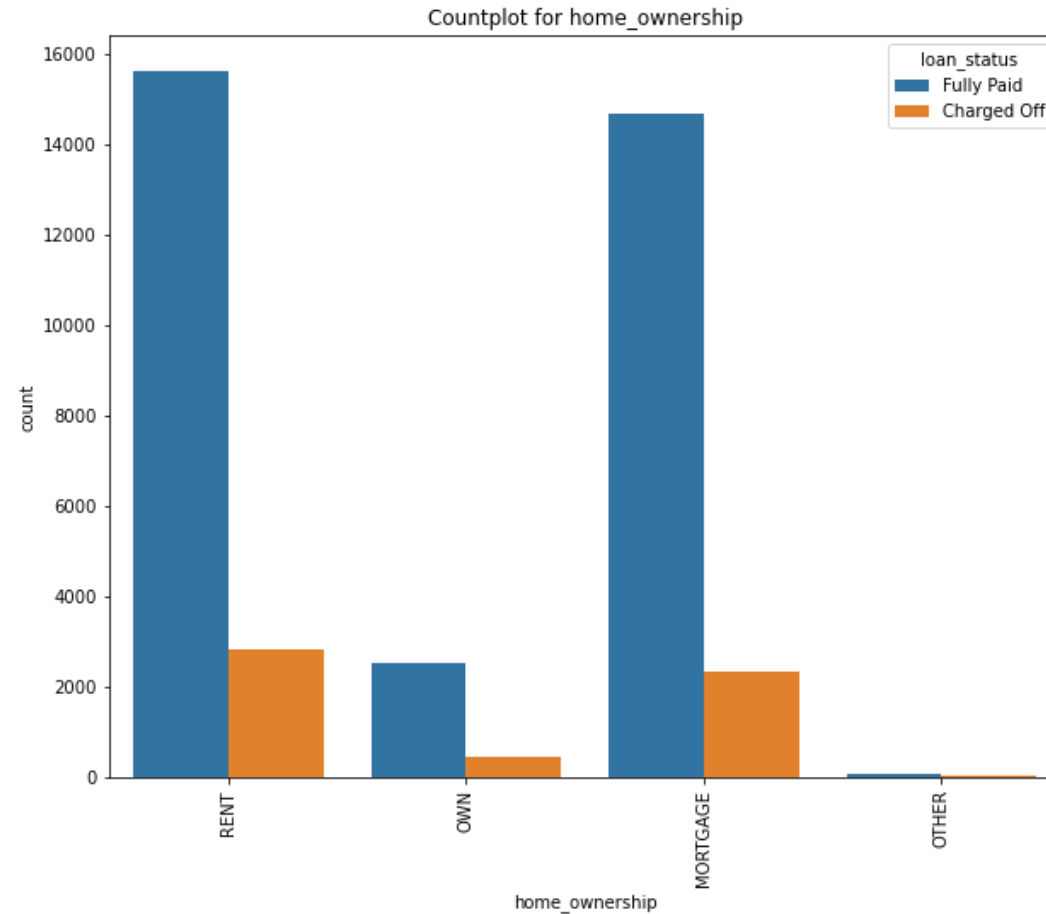
The data corresponding to loan tenure of 36 months is almost 80% of that for 60 months.

Variable : Grade



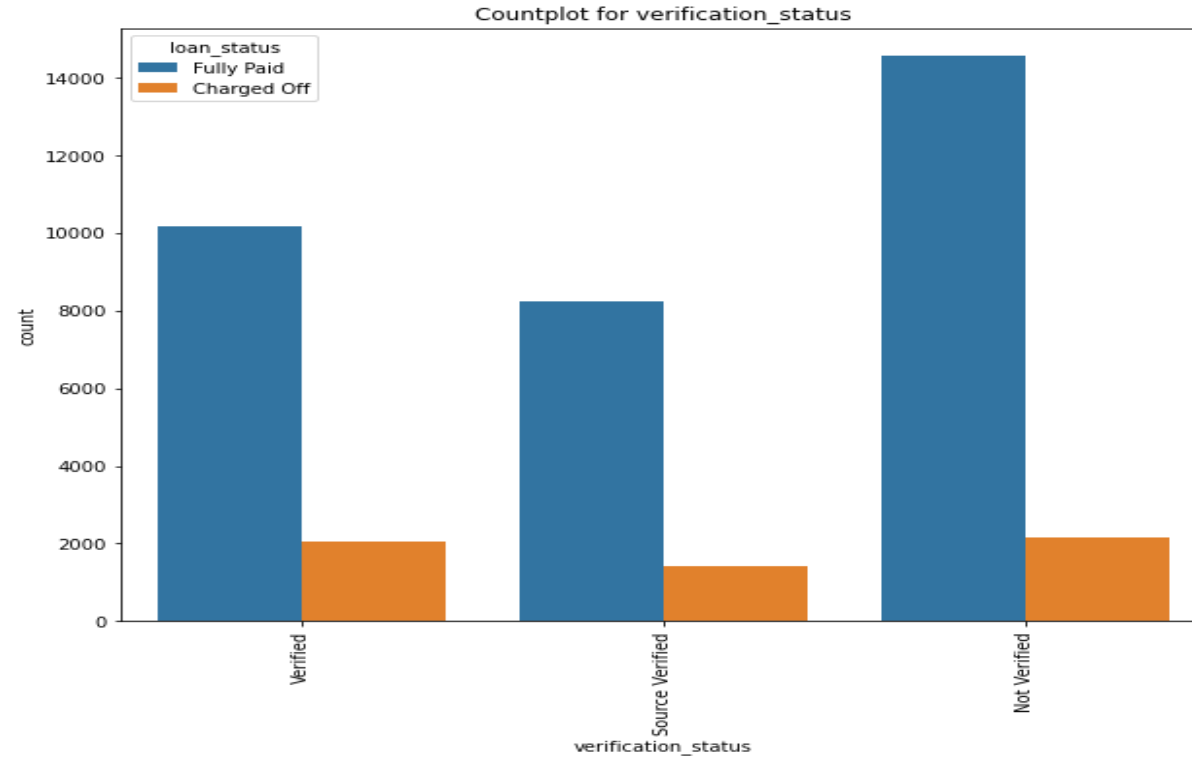
Grades 'B','A' & 'C' are highest respectively followed by 'D', whereas 'F' and 'G' account for very minor portion of the data.

Variable : Home Ownership



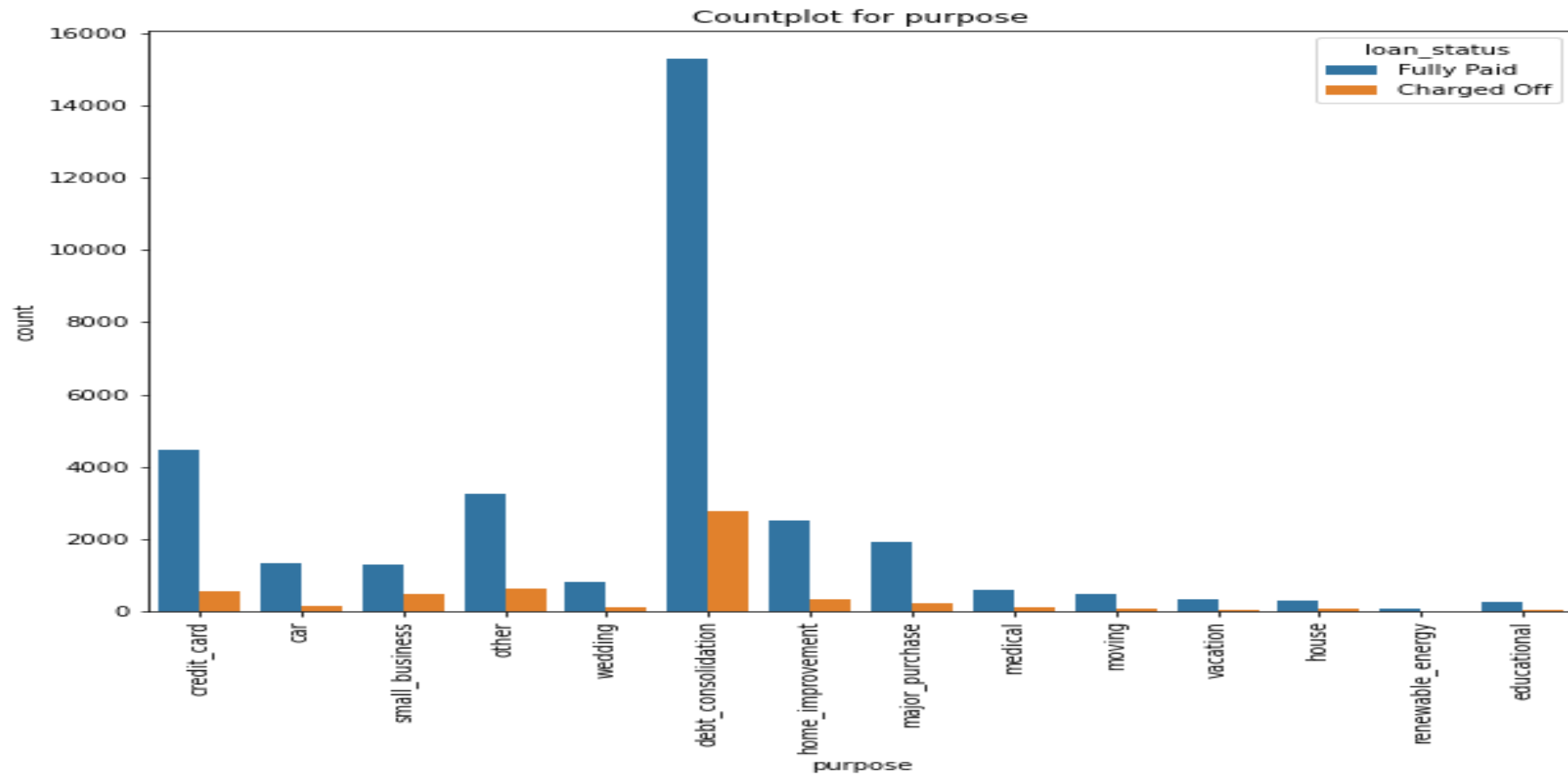
Maximum Charged Off correspond to MORTGAGE & RENT(though looking at the graph MORTGAGE charged off percentile seems higher

Variable : Verification Status



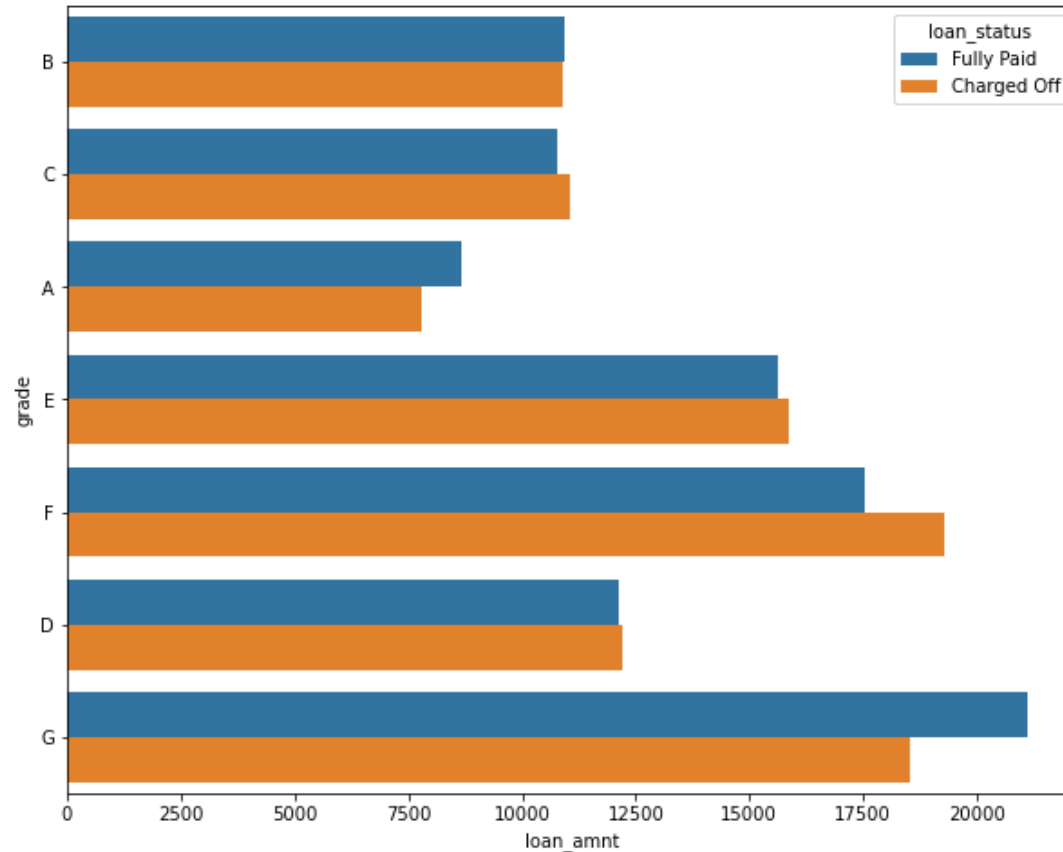
Most of the loan approved falls under Not Verified category, whereas verified and source verified are quite close in numbers.

Variable : Purpose

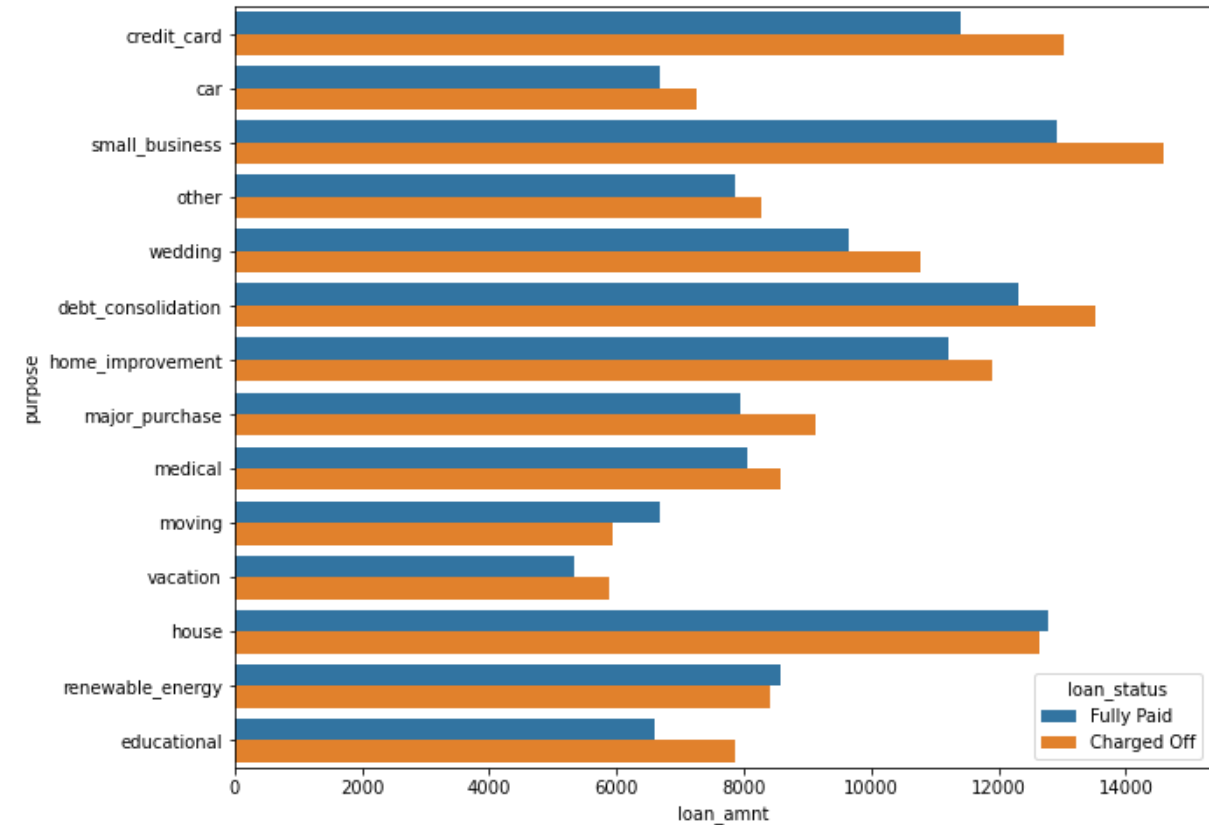


Debt consolidation appears to be the main purpose for the applicants, also charged off is higher

Bi-variate Analysis (Variable : Loan Amount)

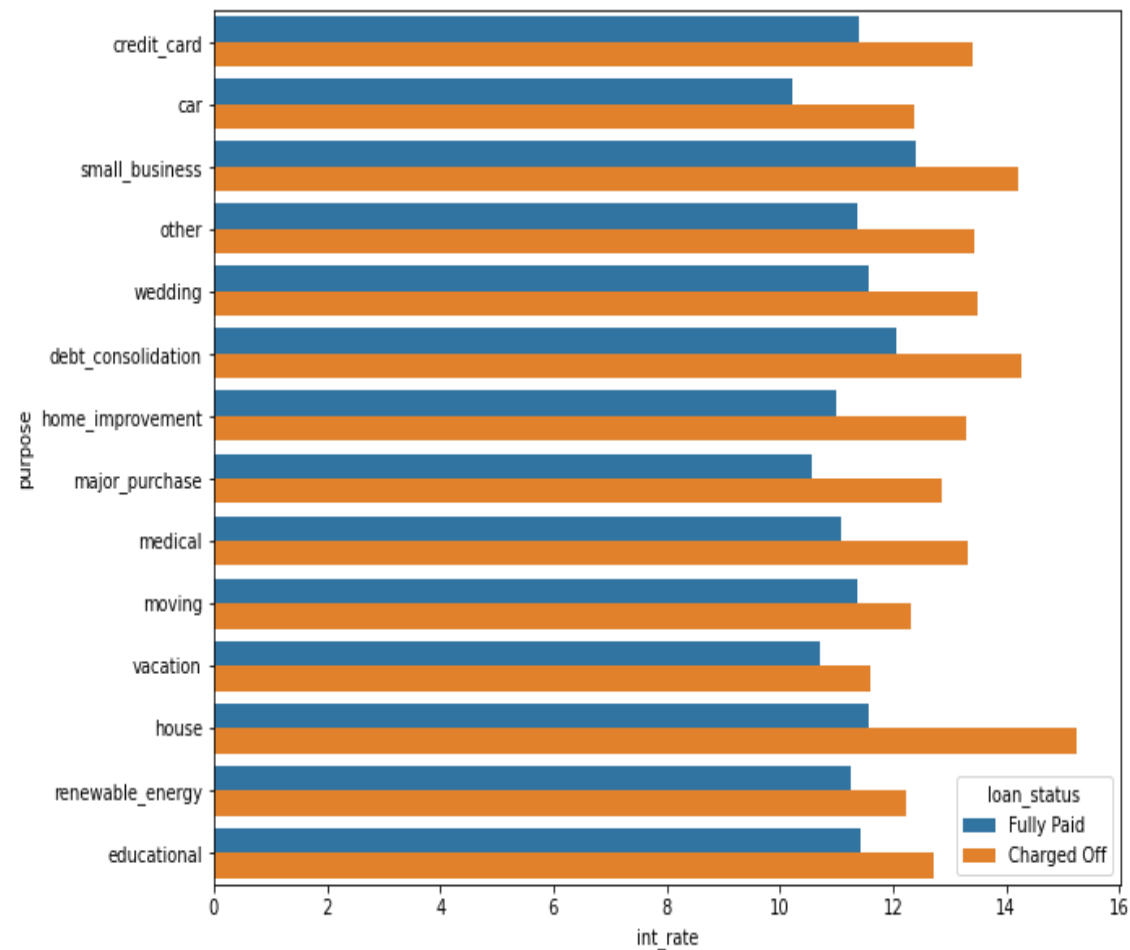
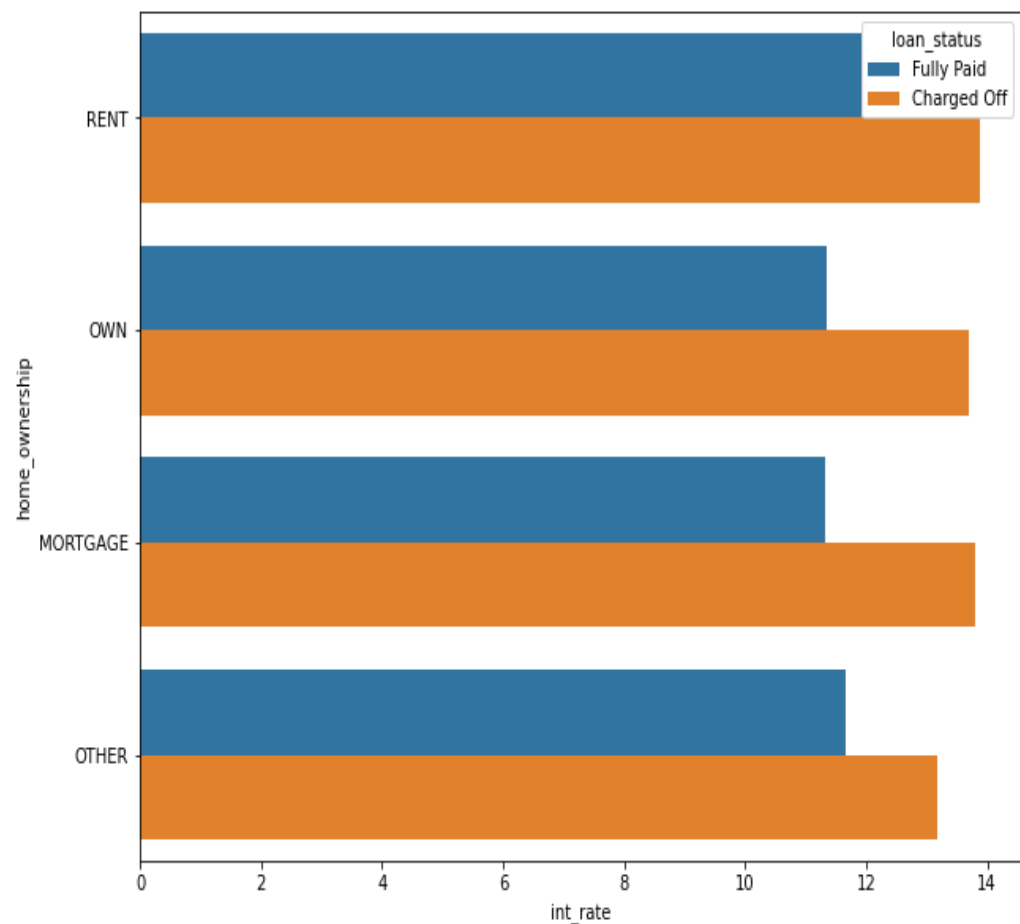


Grade 'G' & 'F' applicants requested max loan amount. Also, maximum defaulters are Grade 'F' applicants



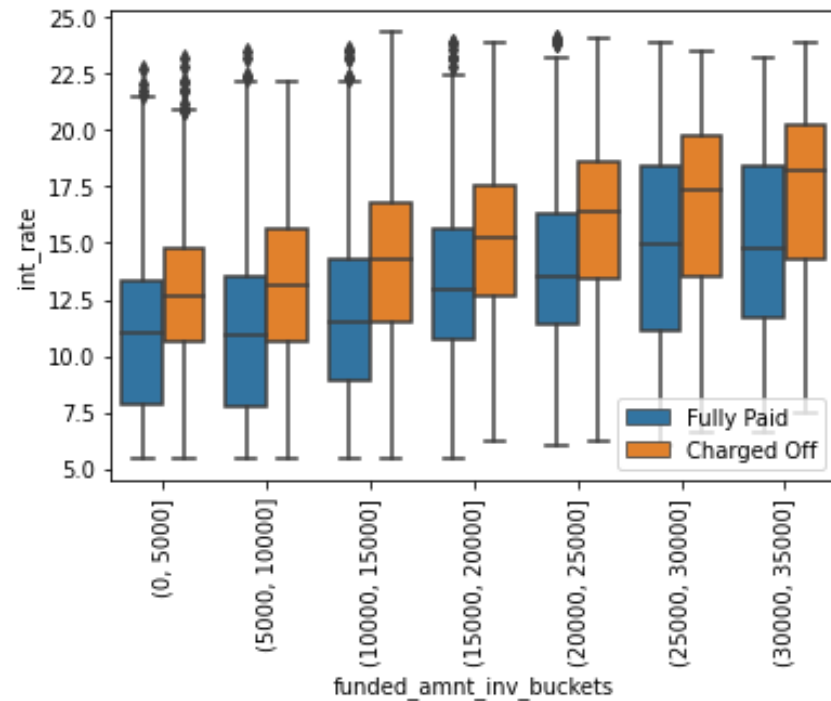
Loan amount requested for Small Business is highest and also have high defaulters followed by Credit Card & Debt Consolidation

Bi-variate Analysis (Variable : Interest Amount)



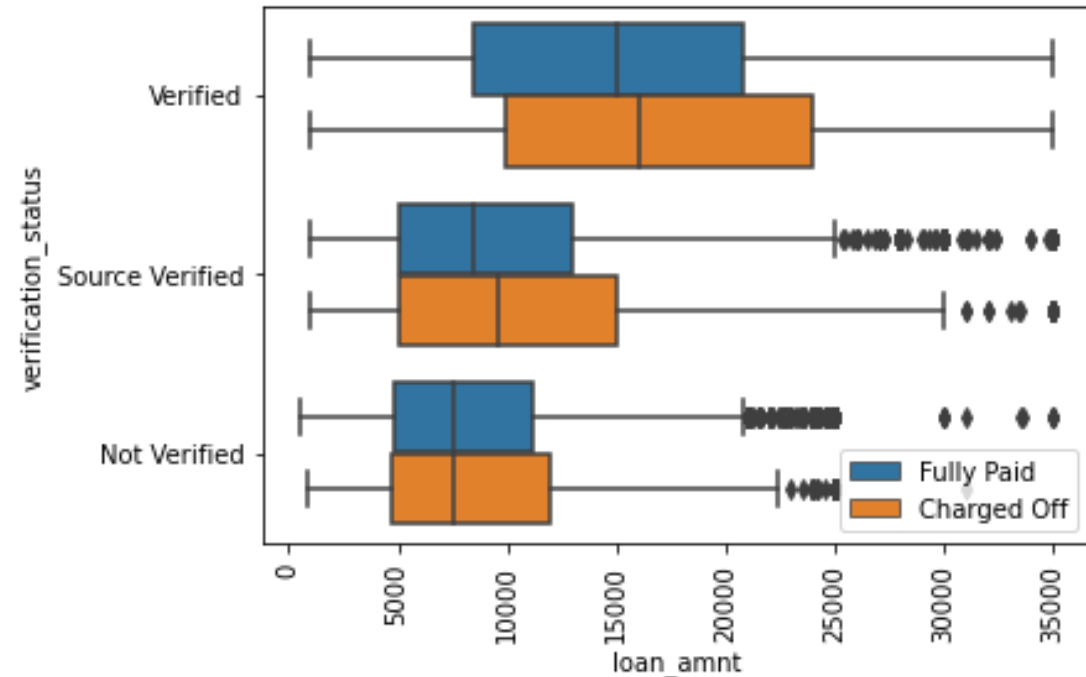
Defaulters are higher for higher interest rate across all categories

Bi-variate Analysis (Variable : Interest Rates vs Funded amt by investors)



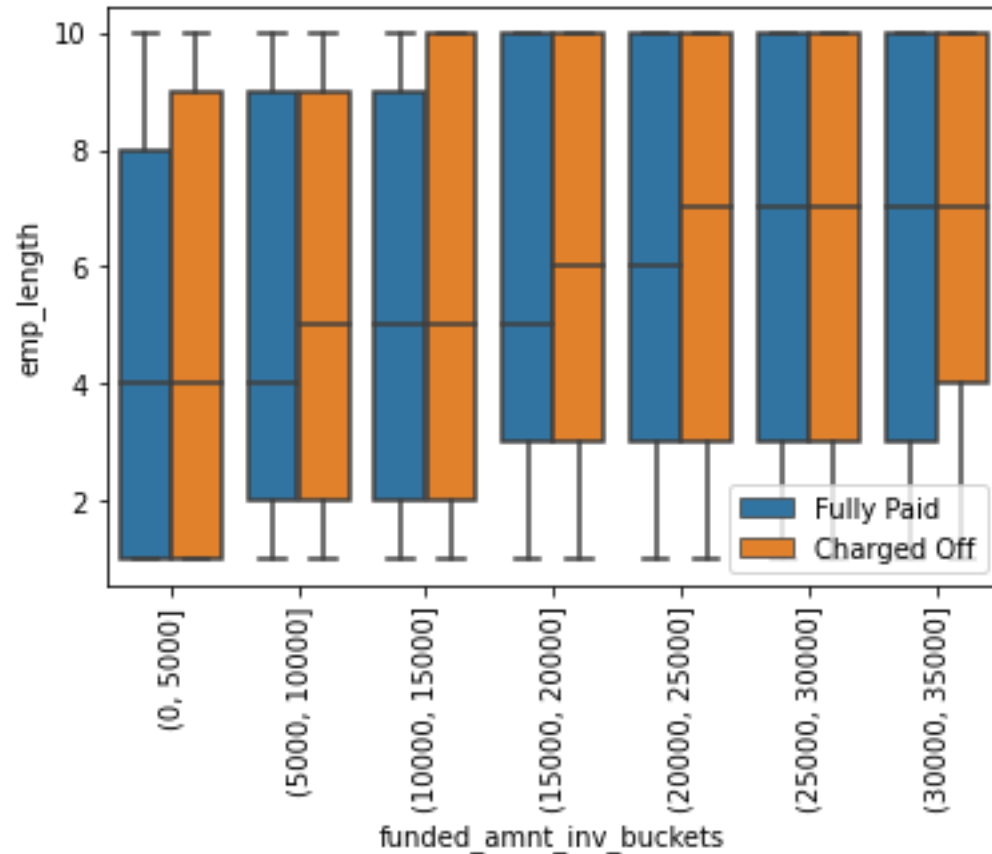
Insight : Interest Rate for Charged Off Loans are quite high compared to Fully Paid loans

Bi-variate Analysis (Variable : Verification status vs Loan Amt)

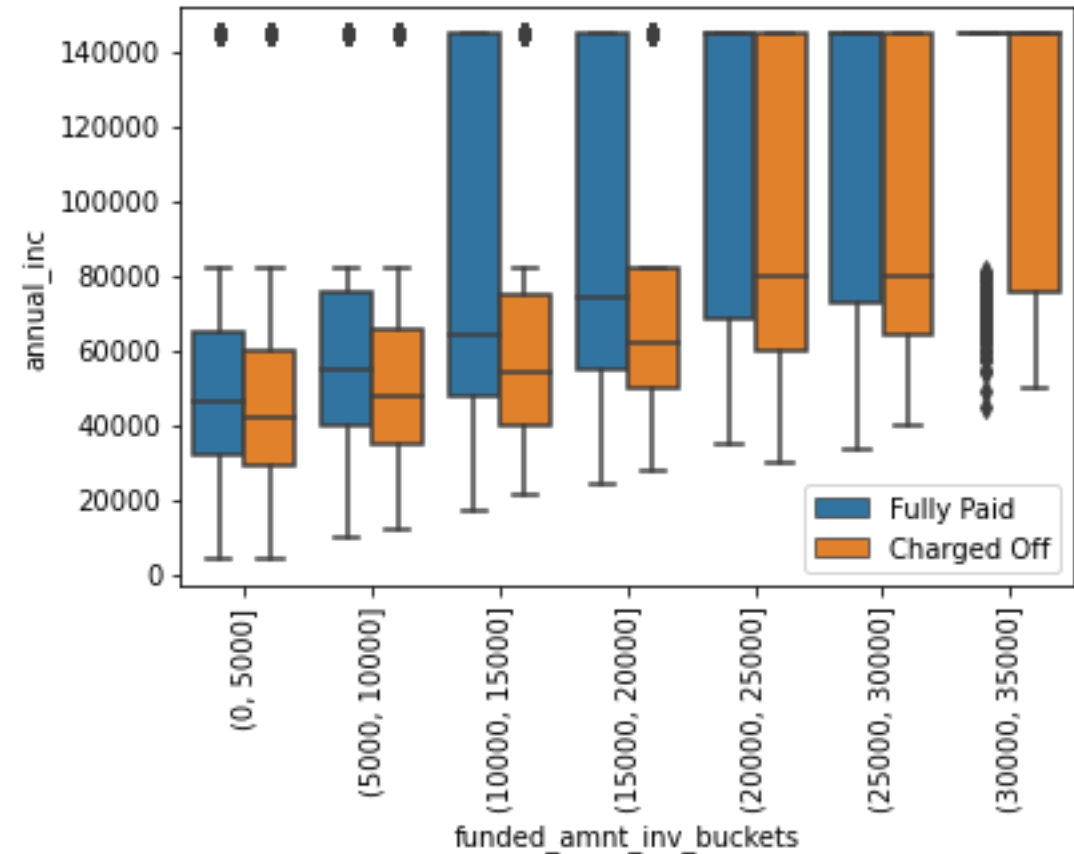


Insight : Loan status is verified and loan amount is 15k & above shows a higher Charged Off

Bi-variate Analysis (Variable : Employment length & Annual income against Funded amount by investors)

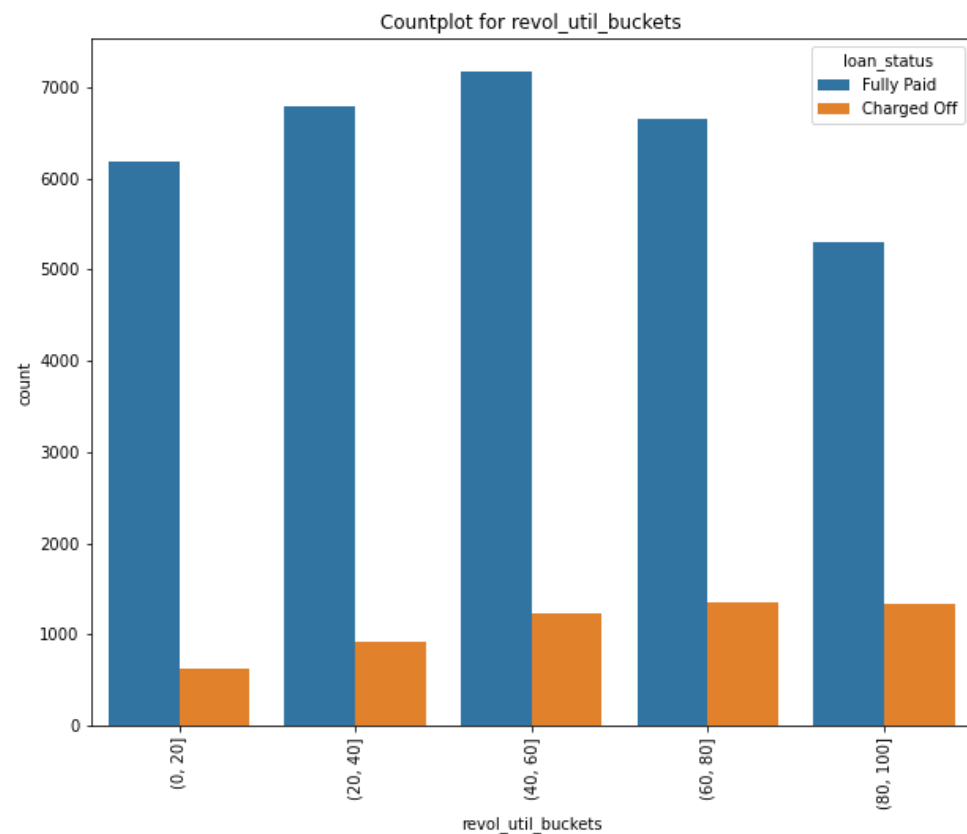


Insight : Defaulters are pretty high for funded amount between 10k to 15k and employee length 10+years

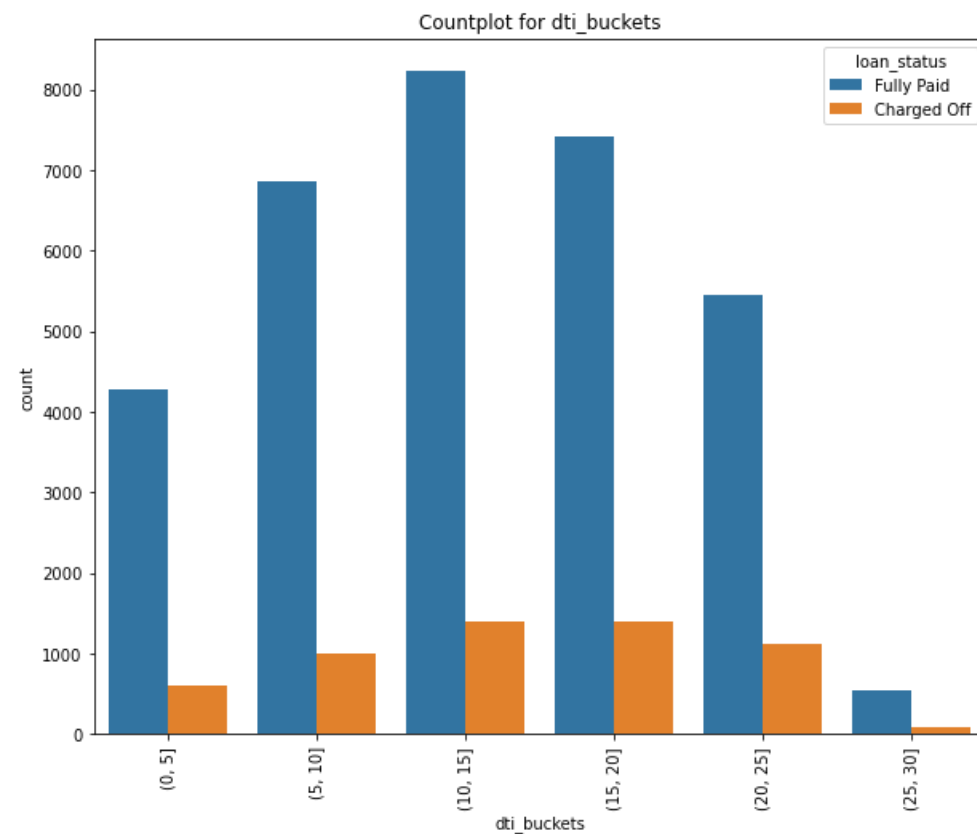


Insights : Chances of defaulting is higher when the funded amount is in range of 20K to 35K and annual income is between 75k to 140k

Bi-variate Analysis (Variable : revol_util & dti)

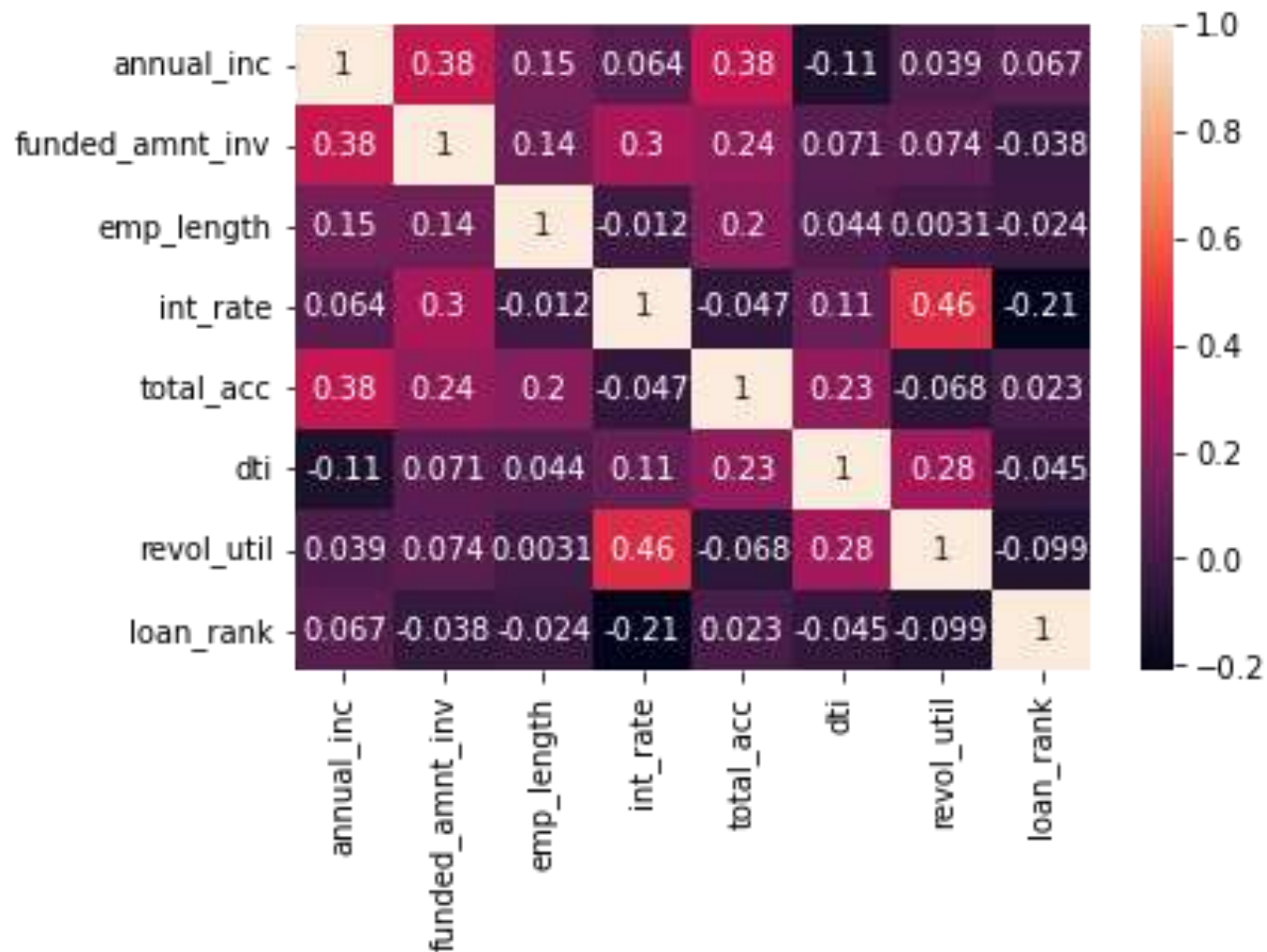


Higher the revol_util charge offs increases



Higher the dti charge offs increases

Overall analysis of the driving factors



Insight :

1. Higher the funded amount, interest rate, dti, revol util and employee length , higher are the chances of loan defaulting
2. Total accounts & Annual income are positively correlated to Loan status

Final Conclusion:

Based on above analysis below are some of the attributes (but not limited to) that can help identify defaulters:

1. Maximum defaulters are higher for applicants with home ownership as **MORTGAGE** or **OTHER**
2. Loan amount requested for **Small Business** is highest and also have high defaulters followed by **Credit Card & Debt Consolidation**
3. Defaulters are higher for **higher interest rate** across all categories
4. Defaulters are pretty high for **funded amount between 10k to 15k** and **employee length 10+years**
5. Chances of defaulting is higher when the funded amount **is in range of 20K to 35K** and **annual income is between 75k to 140k**
6. Long term employees with **purpose as Credit Card** shows a high default rate
7. Higher the dti & revol util, higher are the chances of **defaulting**
8. The dataset shows a very high number of **non-verified loan approvals**.

Recommendations

Based on the study conducted using EDA, we can give the following suggestions to Lending Club:

- ❖ Since the non-verified loans tend to get defaulted, proper verification has to be done vigorously.
- ❖ Issuing loans to home owners seems to be profitable to the company, whereas the most defaults are in mortgage and people living in rented houses. The company can issue them smaller loans after proper verifications & analysis.
- ❖ Issue of loans where the purpose is mentioned as debt consolidation is risky and the company can avoid doing this.
- ❖ Since defaults are pretty high for funded amount between 10k to 15k and employee length 10+years, company can minimize lending out money for such cases.

Thank You
