GRAMENER CASE STUDY

SUBMISSION

Group Name:

1. Aditya Varma
2. Mohammed Mohiuddin
3. Bimal Balan
4. Saransh Sahu

Abstract

Introduction: This case study is done to perform risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers using EDA.

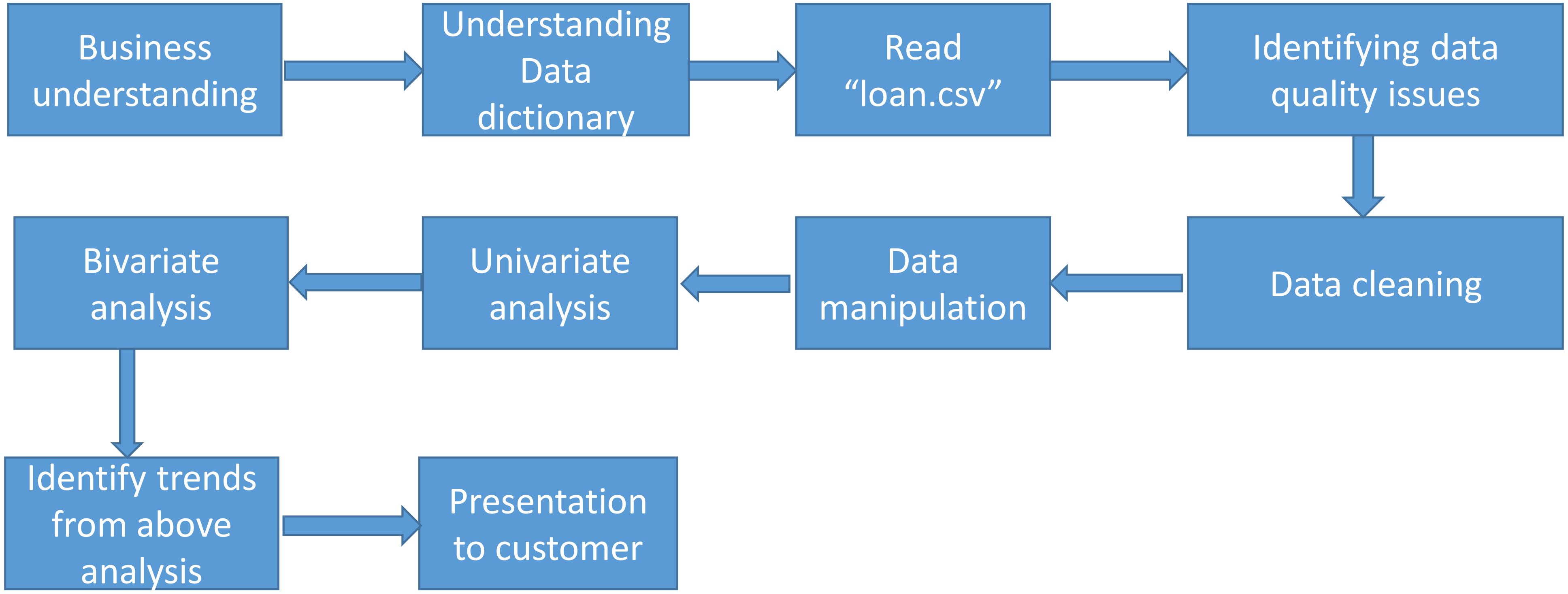
Methods: The study was conducted on data about past loan applicants and whether they ‘defaulted’ or not. Collected data sets were analyzed by python3.6. Collected data consists one “loan” dataset & Data dictionary which explains the variables.

Assumptions:

- ‘Current’ value in ‘loan_status’ column is still a running loan this is excluded from analysis.
- Interest rates depends on Loan grade and subgrade. Hence, interest column is excluded from analysis.
- 'purpose' and 'title' which is essentially giving same information. Hence, title is excluded from analysis.

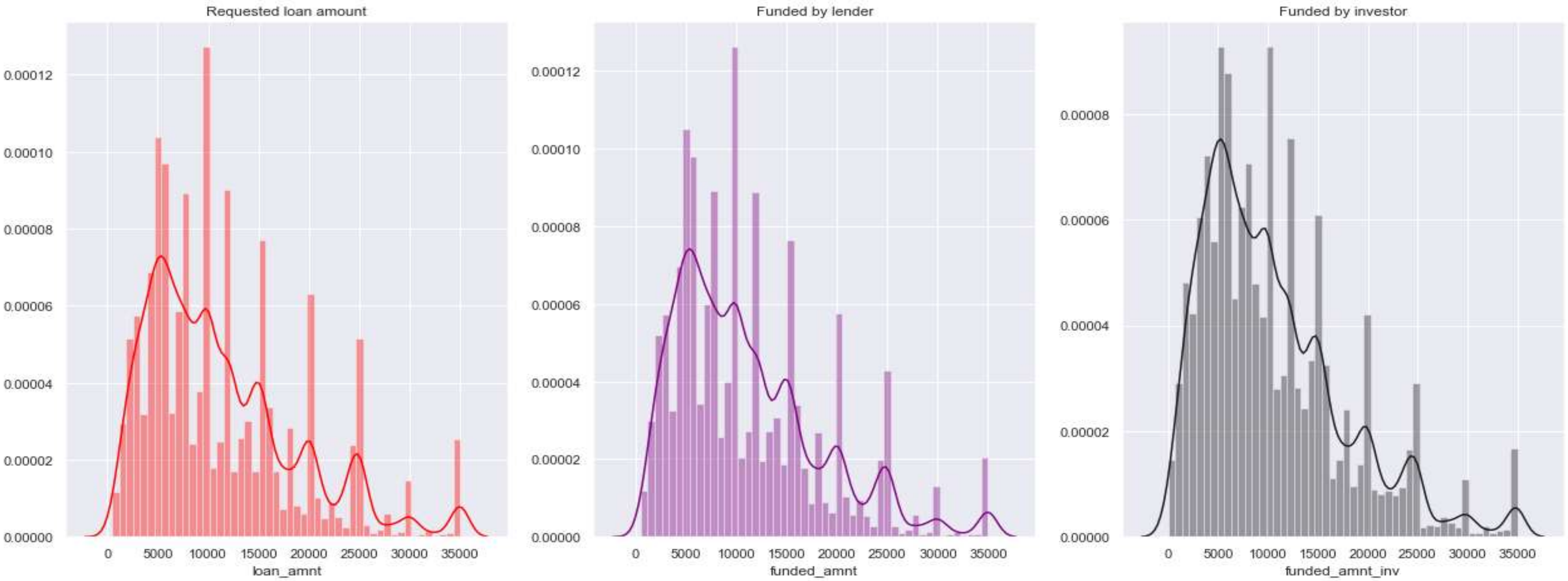
Results: The findings of this analysis indicates most charged off loans are dependent on variables such as employment term, purpose, Verification status, etc. which will be detailed in subsequent slides

Problem solving methodology



Loan attribute variable Analysis

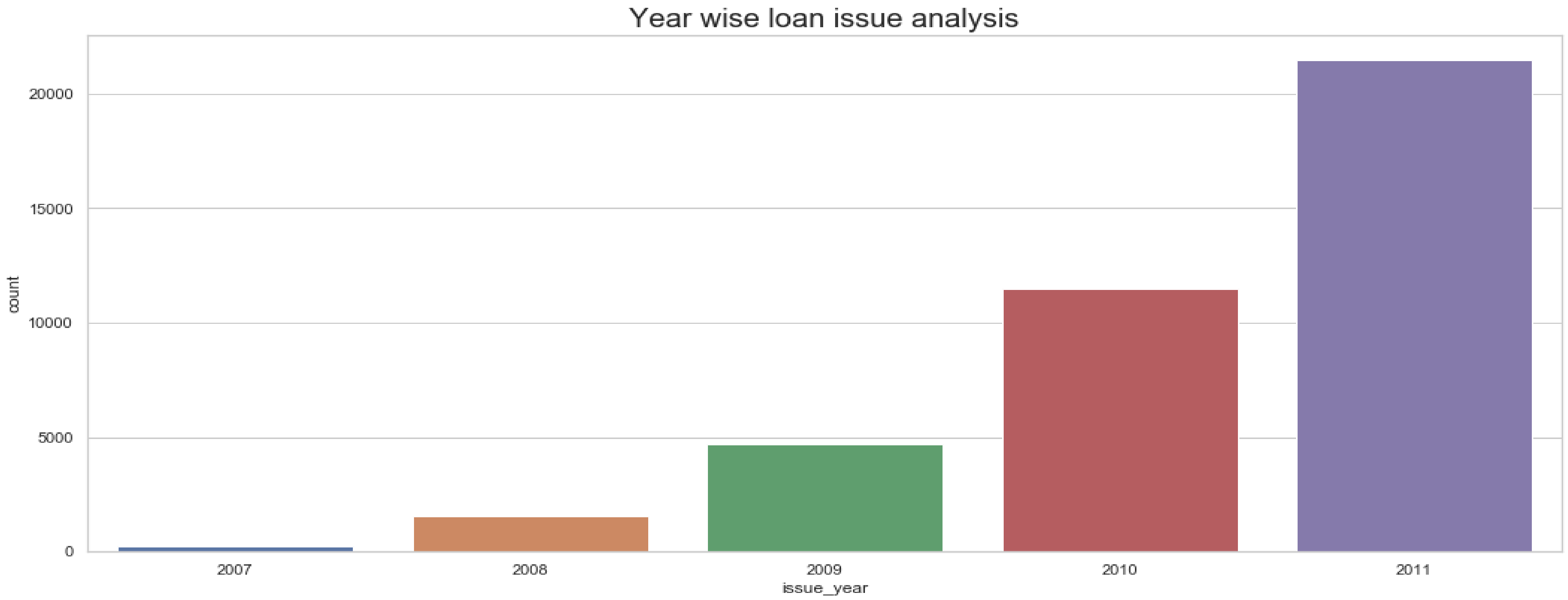
This analysis is done to identify how distributed are loan attribute variables. loan amount, funded amount by lender and amount committed by investors are the 3 variables fall in this category.



Result: From the above visualization it is clear that loan amount, funded amount by lender and amount committed by investors are almost equally distributed and it shows there is sufficient balance between credit and funding

Year wise loan issue analysis

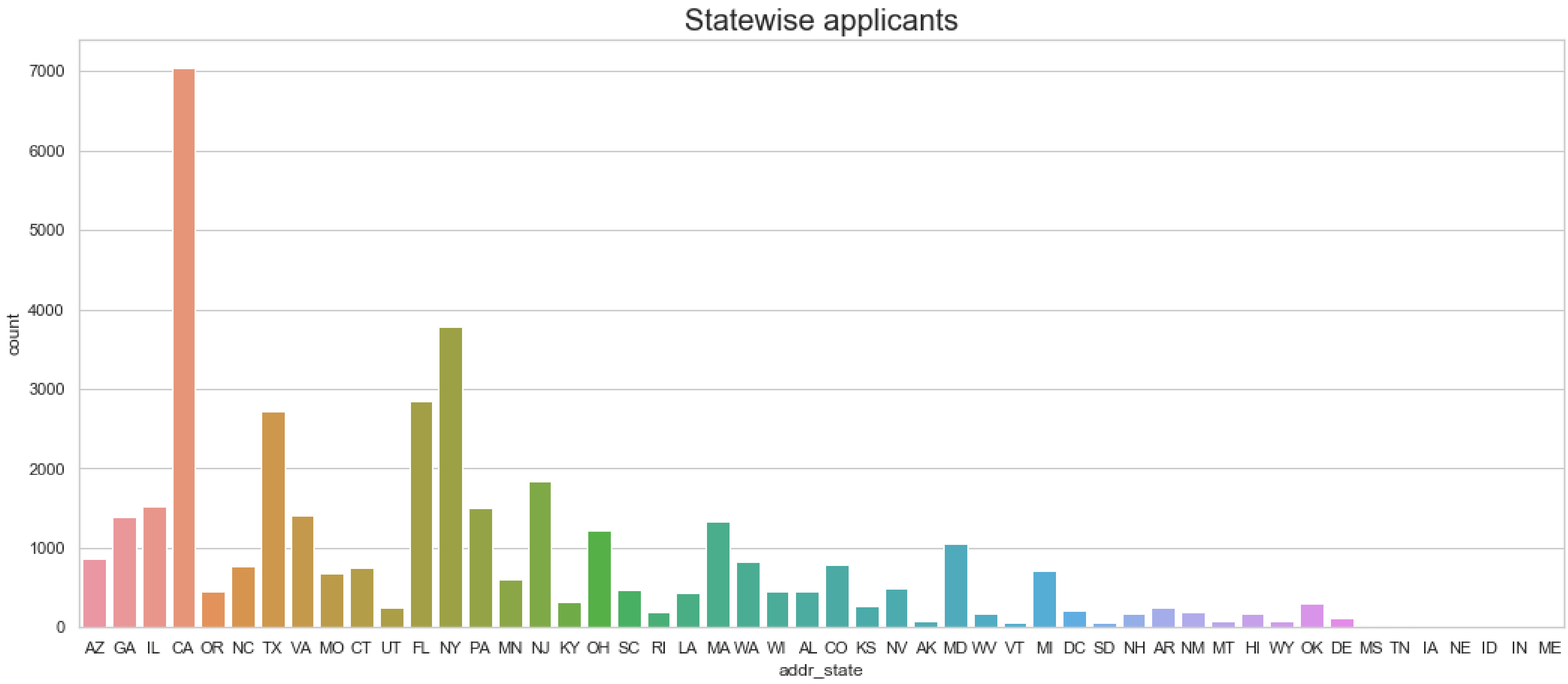
This analysis is done to identify how customer is issuing loans from the year 2007 to 2011.



Result: From the above visualization it is clear that customer is considerably increasing the loan issue from 2007 and there is huge increase in 2011.

State-wise applicant analysis

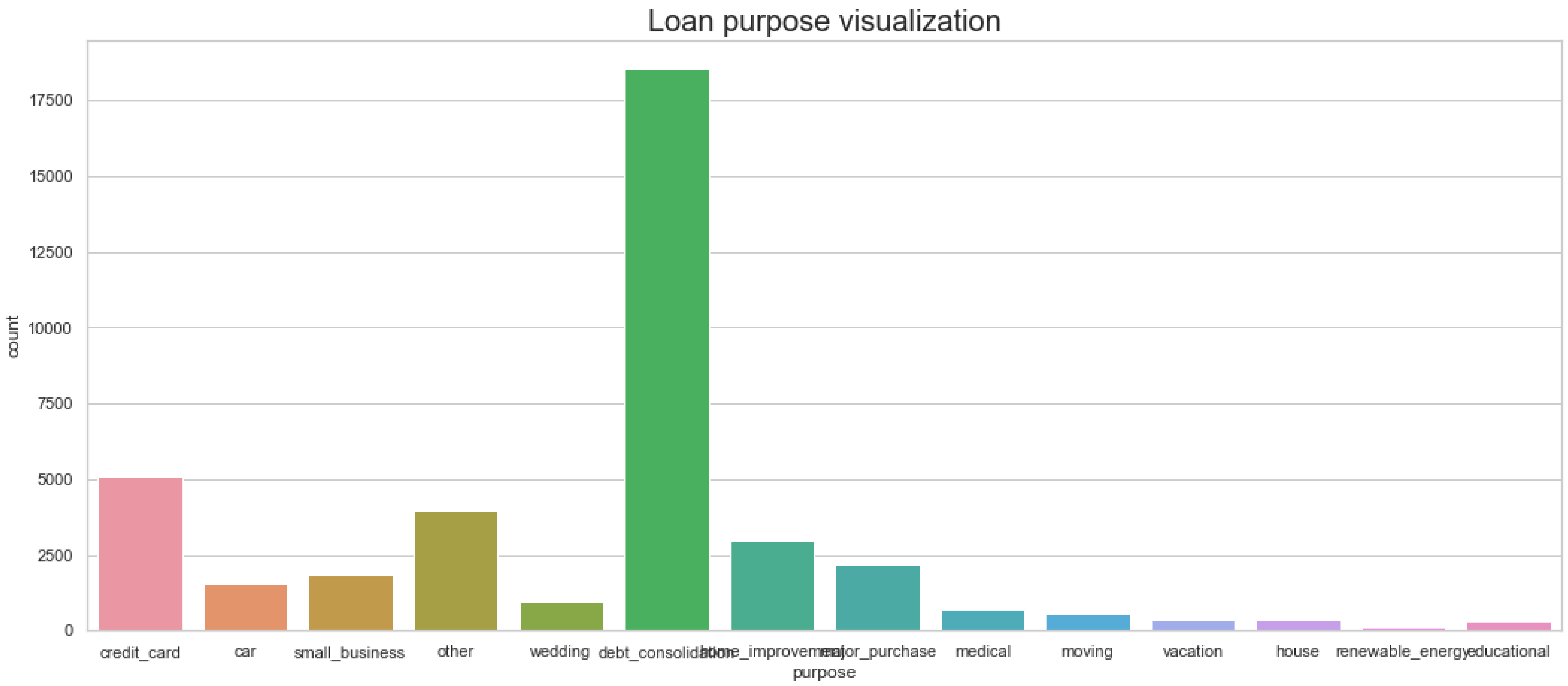
This analysis is done to identify from which states there are more number of applicants



Result: From the above visualization it is evident that CA, NY & FL are top 3 states which has more number of applcants

Loan purpose analysis

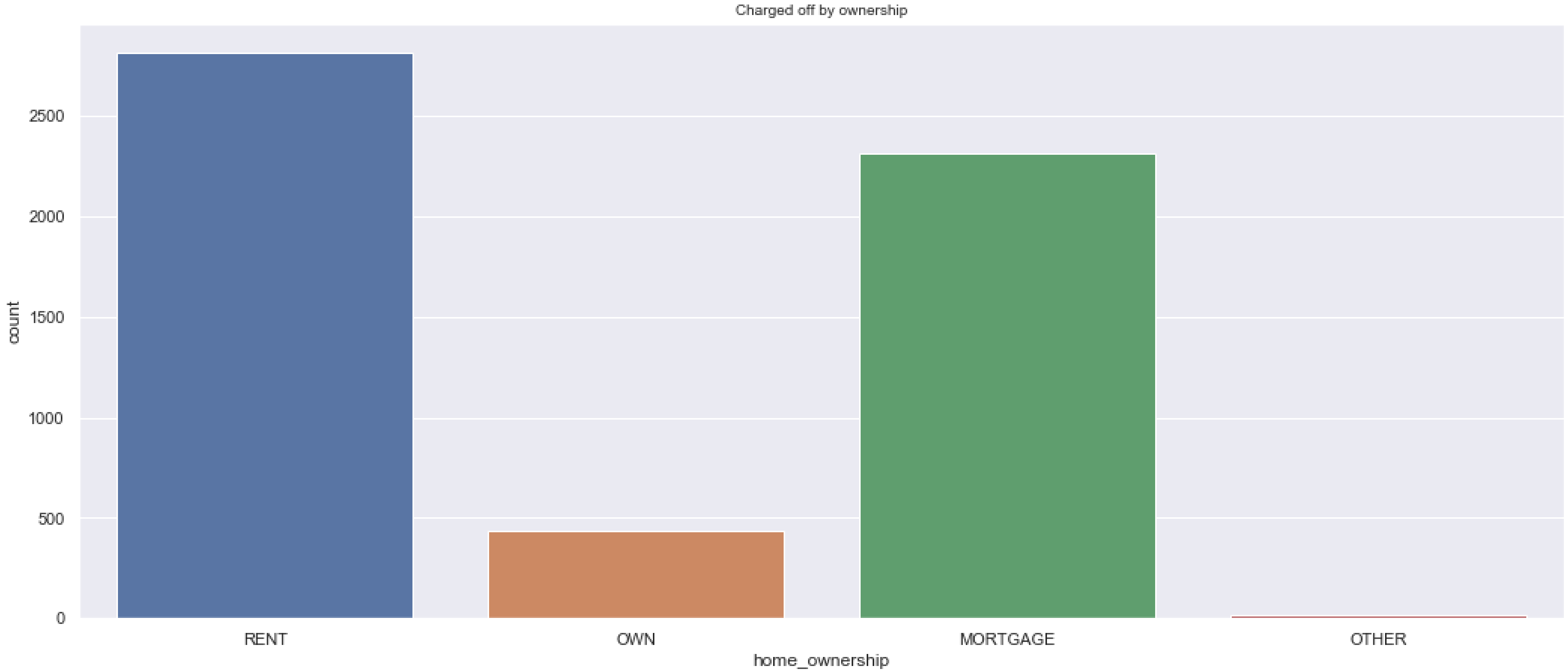
This analysis is done to identify the most frequent loan purpose



Result: Debt Consolidation stands as top loan purpose, with more than 18K loans — or **47%** from the total followed by credit card at **13%** and other by **10%**. Combinedly these 3 purpose comprising of **70% from the total**

Home ownership analysis on charged off loans

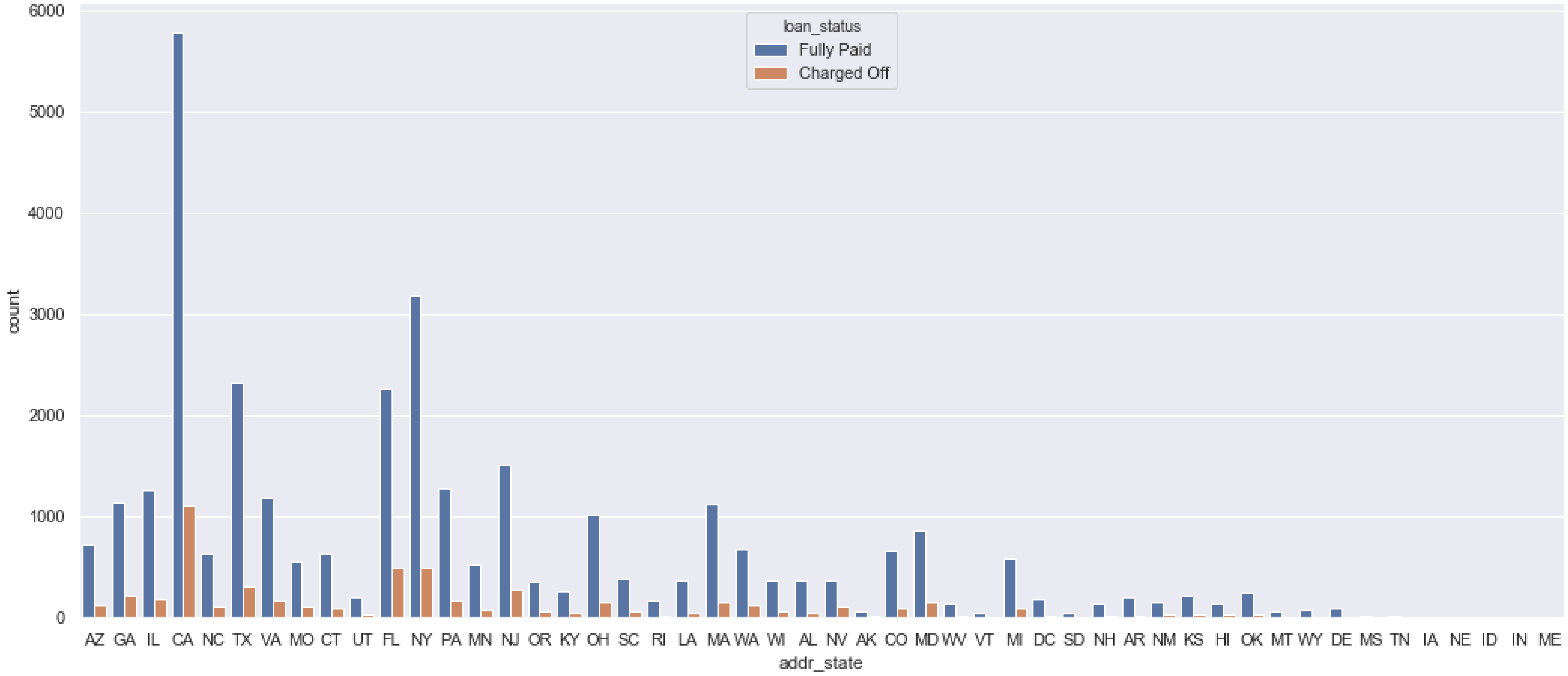
This analysis is done to identify the defaulters of which home ownership



Result: From the above plot it is evident that applicants with 'Rent' & 'Mortgage' as a value are more likely to default when compared to Own house applicants. In total default percentage rental is **50.5%** and mortgage contributes **41.43%**

State-wise loan status analysis

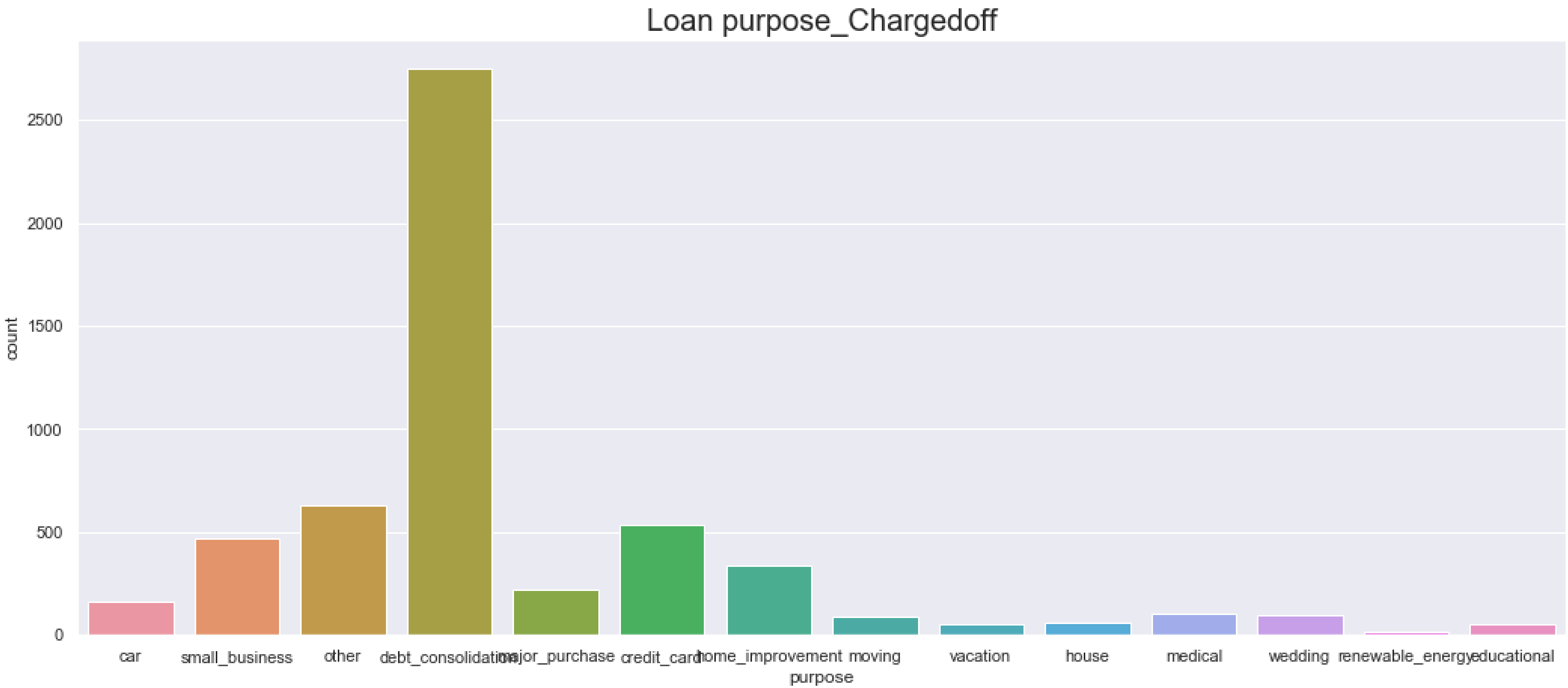
This analysis is done to identify the fully paid and charged off loan status state-wise



Result: For fully paid CA, NY & TX are the top 3 states. For charged off CA, FL, NY are top 3 states with the contribution of 19.92%, 8.93% & 8.84%. The 3 states combined contributes 37% of charged off loans

Loan purpose analysis for charged off

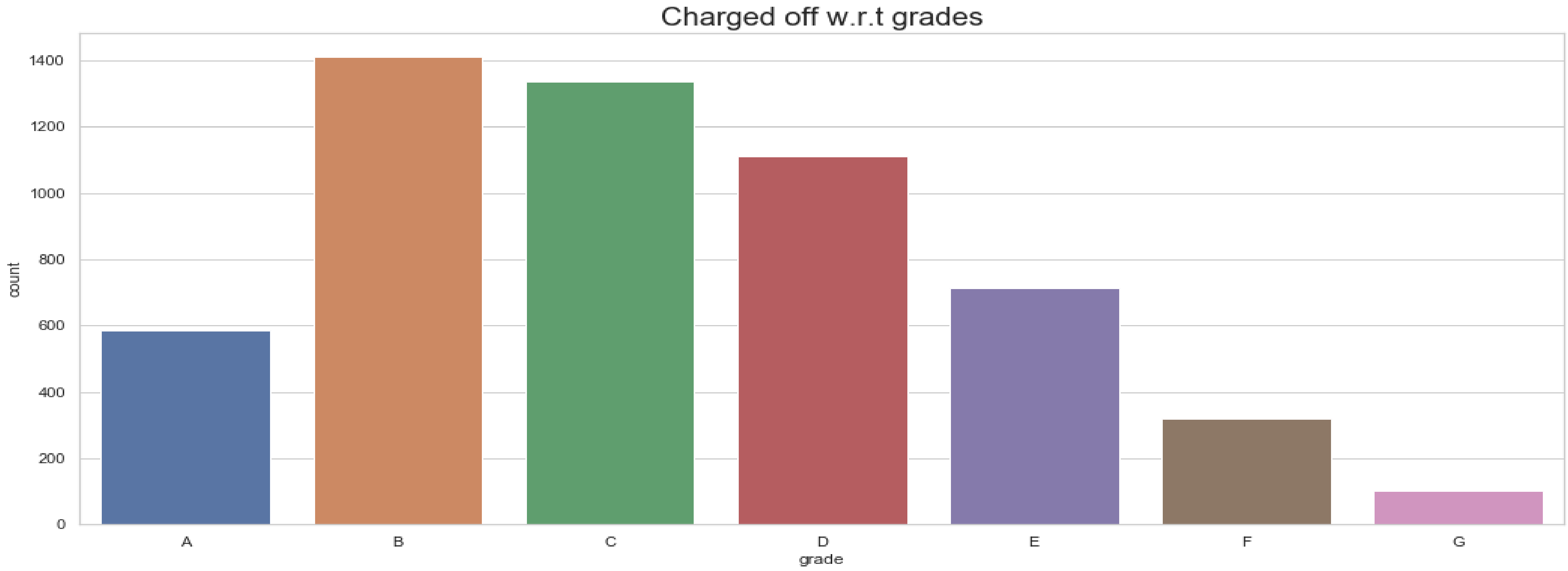
This analysis is done to identify the most frequent loan purpose which is being charged off.



Result: Debt Consolidation stands as top loan purpose, with more than 2K loans — or **49%** from the total followed by other at **11%** and credit card by **10%**. Combined these 3 purpose comprising of **70% from the total**

Grade analysis for charged off

This analysis is done to identify the grades which is being charged off.



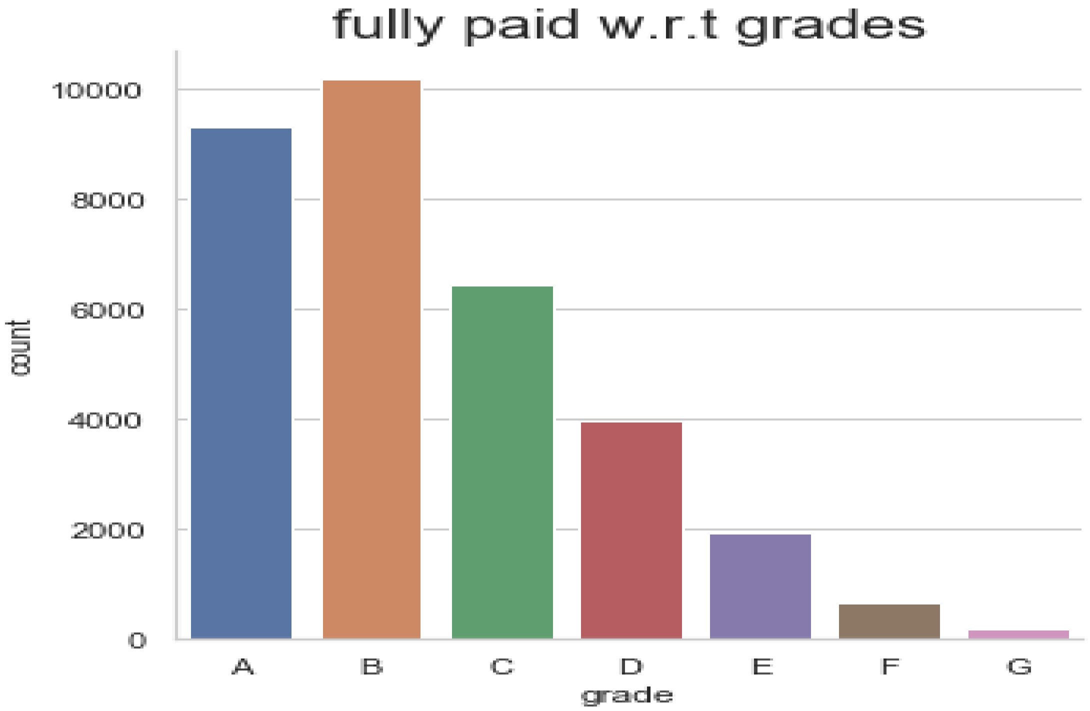
Result: As per industry expectation grades from A to C are considered as low risk and D to G are considered as High risk. But from the above plot it can be observed that Grades B, C & D are having maximum defaulters

Out of 5576 total charged off loans below is the percentage of defaulters in each grade

A->10.51%, B->**25.30%**, C->**23.96%**, D->19.92%, E->12.79%, F-> 5.70%, G->1.81%

Grade analysis for fully paid

This analysis is done to identify the grades which is being fully paid.

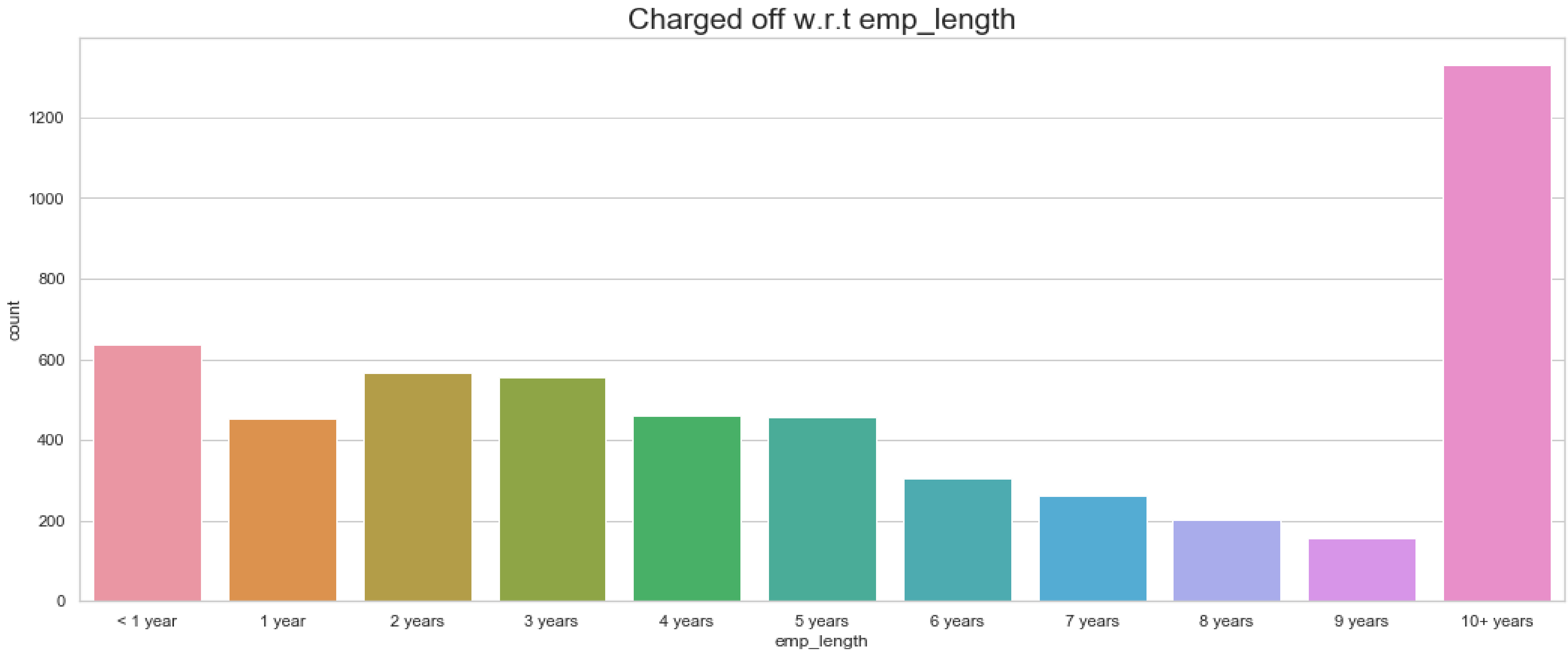


Result: As per industry expectation grades from A to C are considered as low risk and D to G are considered as High risk. It is as expected from above plot that grades A,B,C has maximum number of fully paid

So from grade analysis we have indications that loan grades itself is not an ideal option for decision making

Employee term analysis for charged off

This analysis is done to identify the employee terms which is majorly contributing for charged off loans.

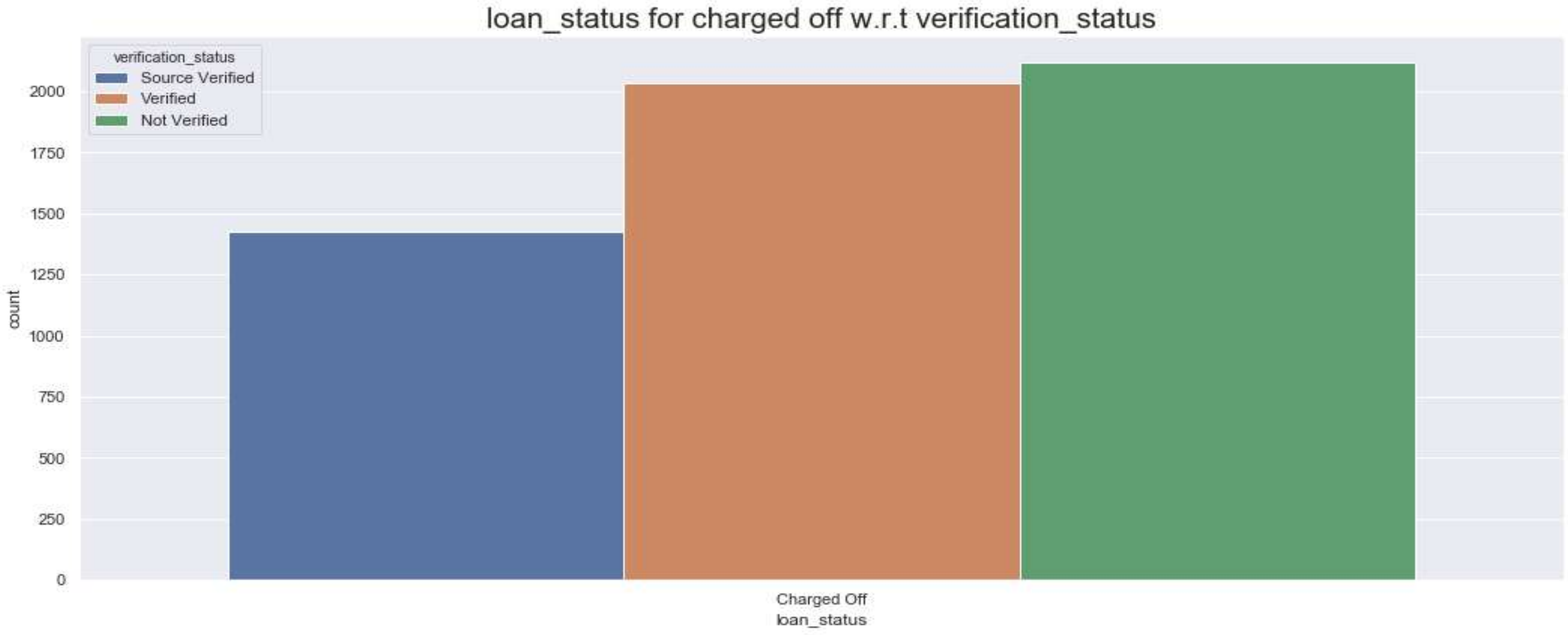


Result: From the above plot it is evident that employees with terms 10+years(more risky), <1 year & 2 years are major defaulters with below %. So loans for lesser employee term should be preferred rather that more than 10 years

10+ years -> **24.67%**, < 1 year -> 11.84%, 2 years -> 10.51%

Verification status analysis for charged off

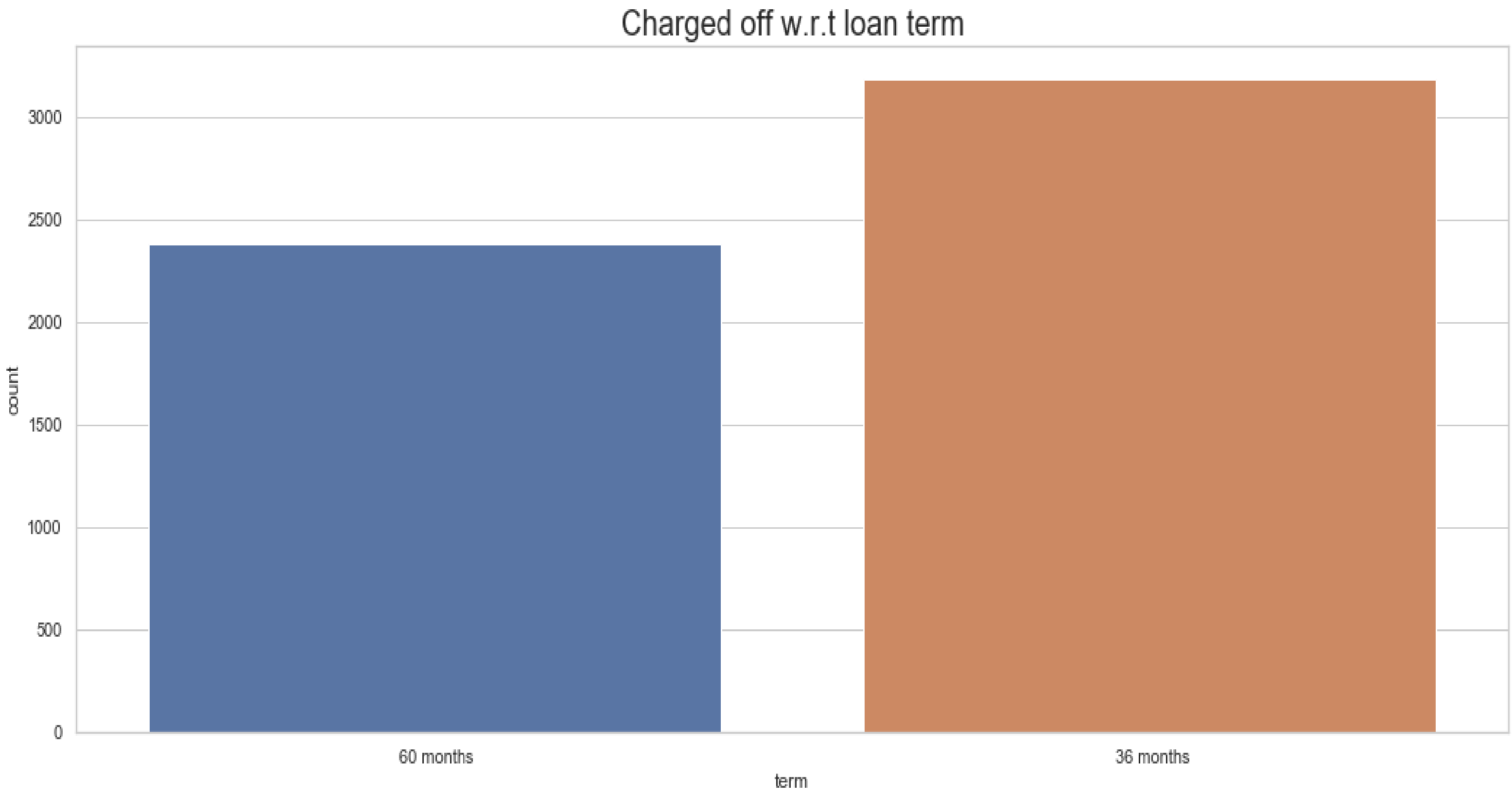
This analysis is done to identify the verification status which is majorly contributing for charged off loans.



Result: Charged off loans high for both Not Verified and Verified. **38%** of not verified and **36.5%** of verified status are charged off which is alarming. Verification process should be reviewed and make necessary improvements.

Loan term analysis for charged off

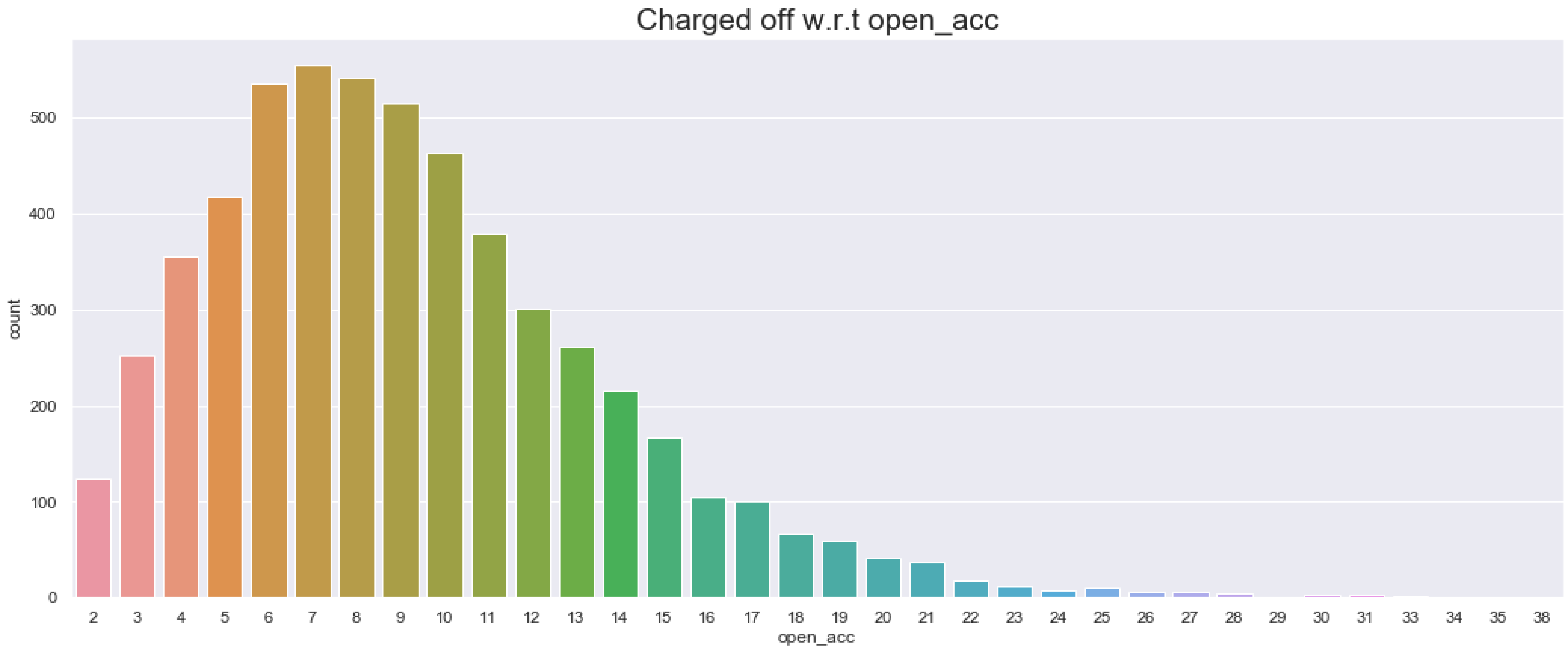
This analysis is done to identify the loan term which is majorly contributing for charged off loans.



Result: Defaulters are higher for 36 months when compared to 60 months. 36 months tenure is contributing **57.19%** of charged off loans.

Open account analysis for charged off

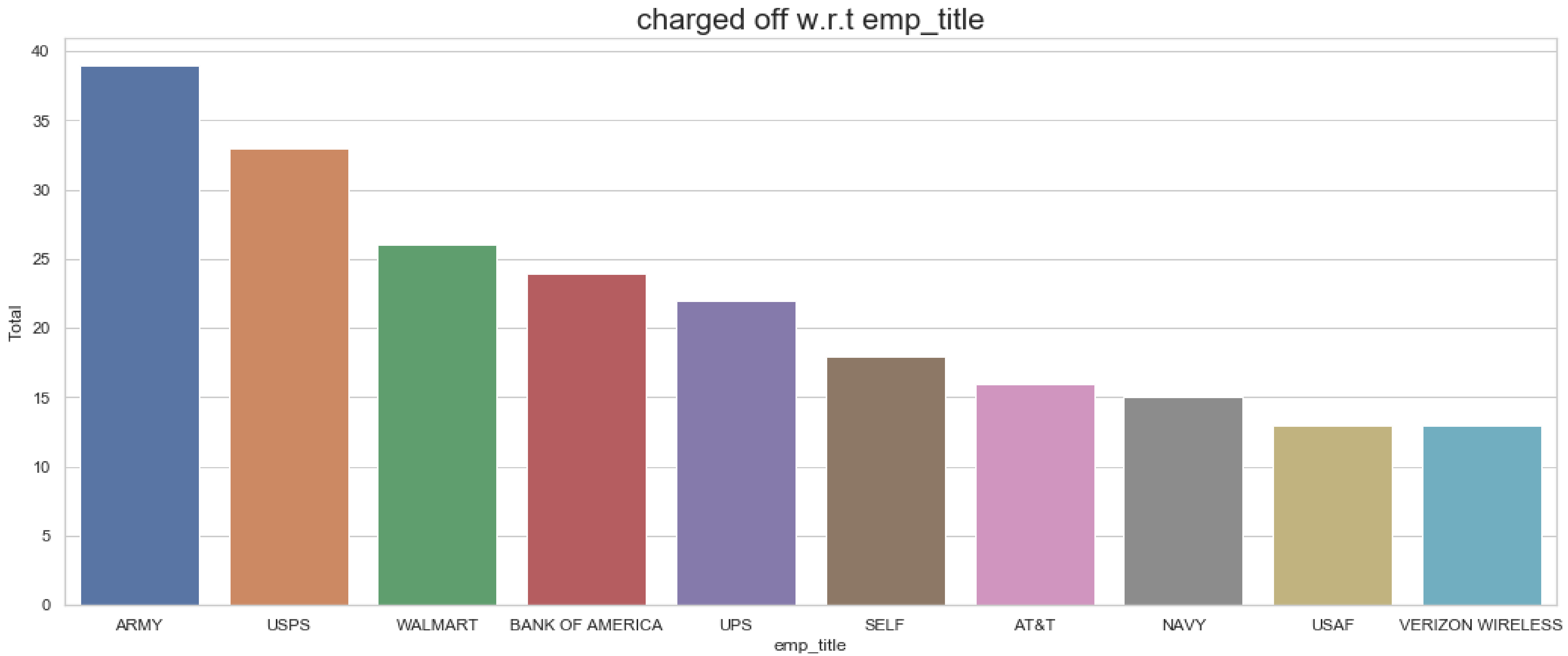
This analysis is done to identify the number of open credit lines in the borrower's credit file which is majorly contributing for charged off loans.



Result: Charged off loans high in the range of 5 to 10 open accounts. All these open accounts contributing **54.30%** for charged off loans.

Employee title analysis for charged off

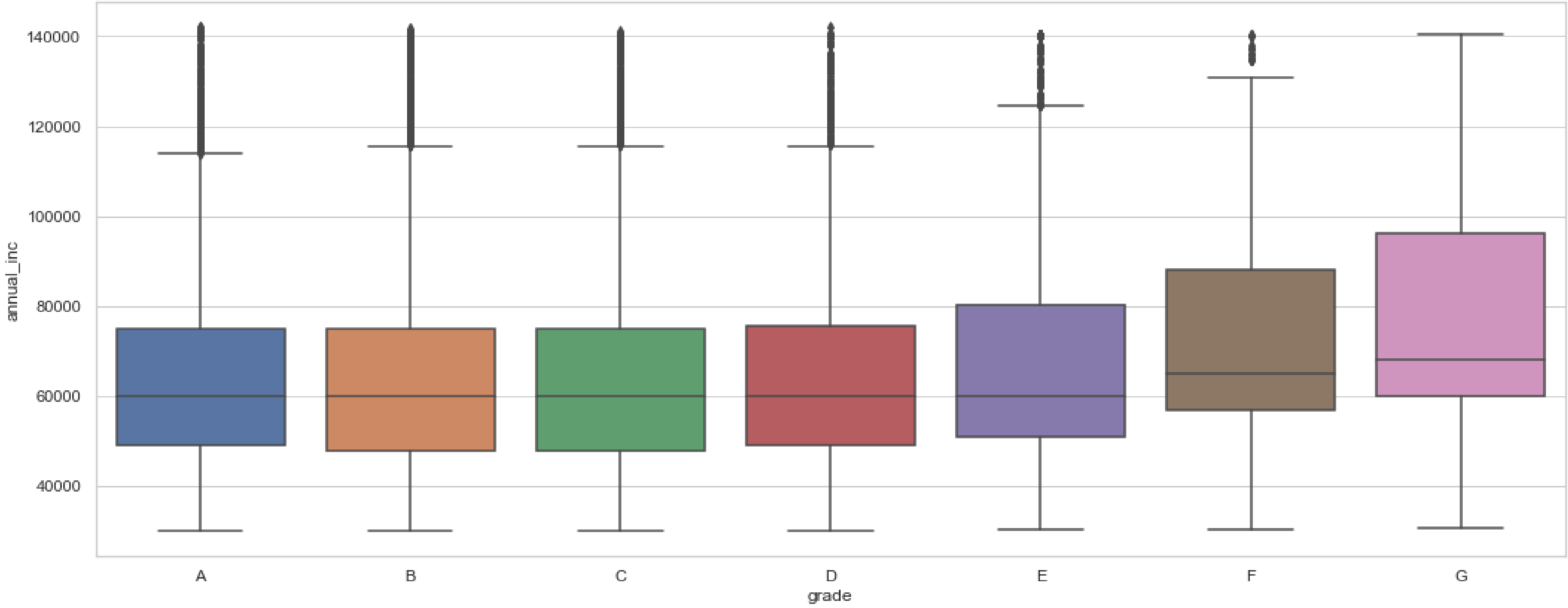
This analysis is done to identify the employee title which is majorly contributing for charged off loans.



Result: Employees from Army, USPS & Walmart are the top 3 defaulters

Bivariate analysis on income & grade

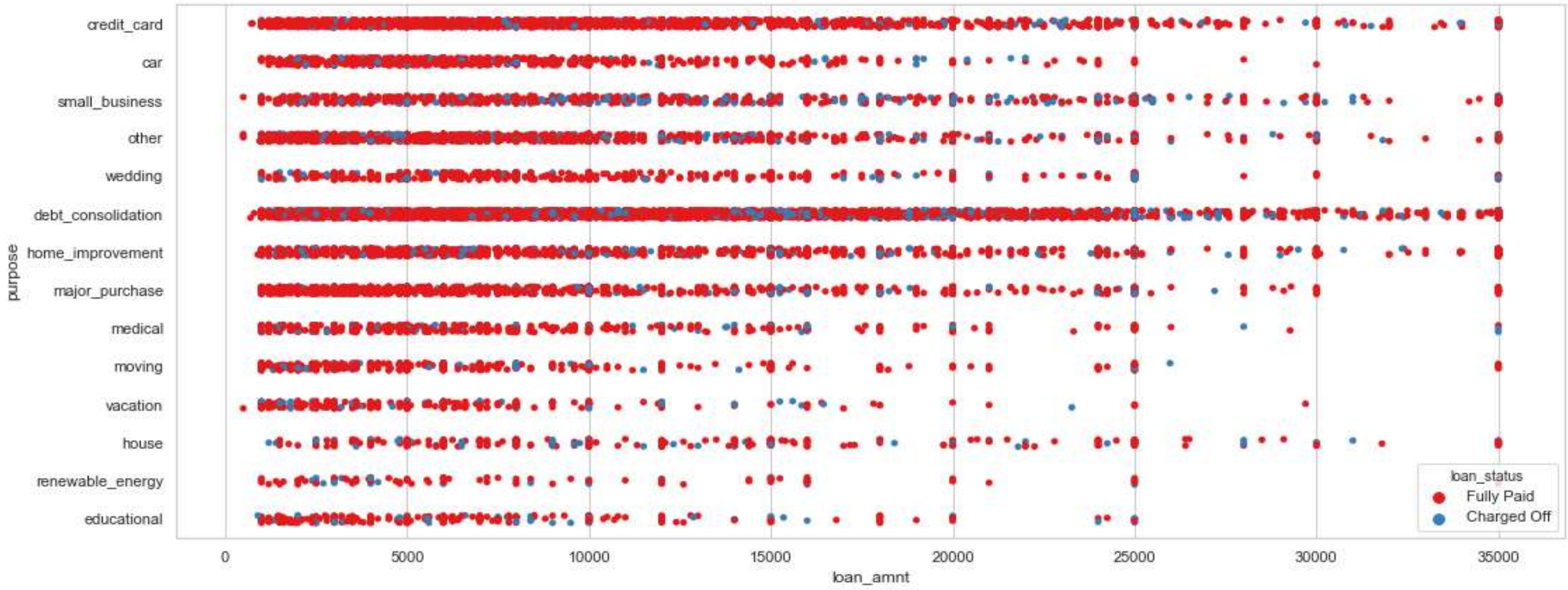
This analysis is done to identify the affect of annual_income on grades.



Result: Mean annual_inc for A,B,C,D remains almost same where as, it gradually increases from E to G

Multivariate strip plot analysis for loan_status

This analysis is done to identify the loan status whether fully paid or charged off based on two more variables loan amount and purpose.



Result: Debt consolidation, small business, other & credit card are the major defaulters. Most number of default loans are observed from 1000 to 15000 range of loan amount. This range contributes **71.12%** of the total charged off loans.

Conclusions- Driving factors behind loan default

The findings indicated that below are the driving factors for loan default.

- **Home ownership**: 'Rent' & 'Mortgage' as a value are more likely to default when compared to Own house applicants. In total default percentage rental is **50.5%** and mortgage contributes **41.43%**
- **Applicant state**: CA, FL, NY are top 3 states with the contribution of 19.92%, 8.93% & 8.84%. The 3 states combined contributes **37%** of charged off loans
- **Loan purpose**: Debt Consolidation stands as top loan purpose, with more than 2K loans — or 49% from the total followed by other at 11% and credit card by 10%. Combined these 3 purpose comprising of 70% from the total
- **Employee term**: Employees with terms 10+years(more risky) contributes **24.67%** of total defaulters. After 1st year there is a decreasing trend in charged off loans till 9 years. But after that there is a considerable amount of increase in defaulters
- **Verification status**: Charged off loans high for both Not Verified and Verified. **38%** of not verified and **36.5%** of verified status are charged off which is alarming. This situation might be occuring due to loan approvals done during month end to reach employee targets. **Verification process should be strictly reviewed and make necessary improvements.**
- **Employee title**: Employees from Army, USPS & Walmart are the top 3 defaulters
- **Open account**: Default rate is high in the range of 5 to 10 open accounts. All these open accounts contributing **54.30%** for charged off loans.
- **Loan term**: Defaulters are higher for 36 months when compared to 60 months. 36 months tenure is contributing **57.19%** of charged off loans.
- **Annual income & Grade**: Mean annual income for A, B, C, D remains almost same where as, it gradually increases from E to G.
- **Loan amount**: Most number of default loans are observed from 1000 to 15000 range of loan amount. This range contributes **71.12%** of the total charged off loans