Lending Club Case Study (EDA)

- Suyash Kumar Singh
- Srinivasan Radhakrishnan

Agenda

- Problem statement
- Analysis approach
- Observations
- Conclusions

Problem statement

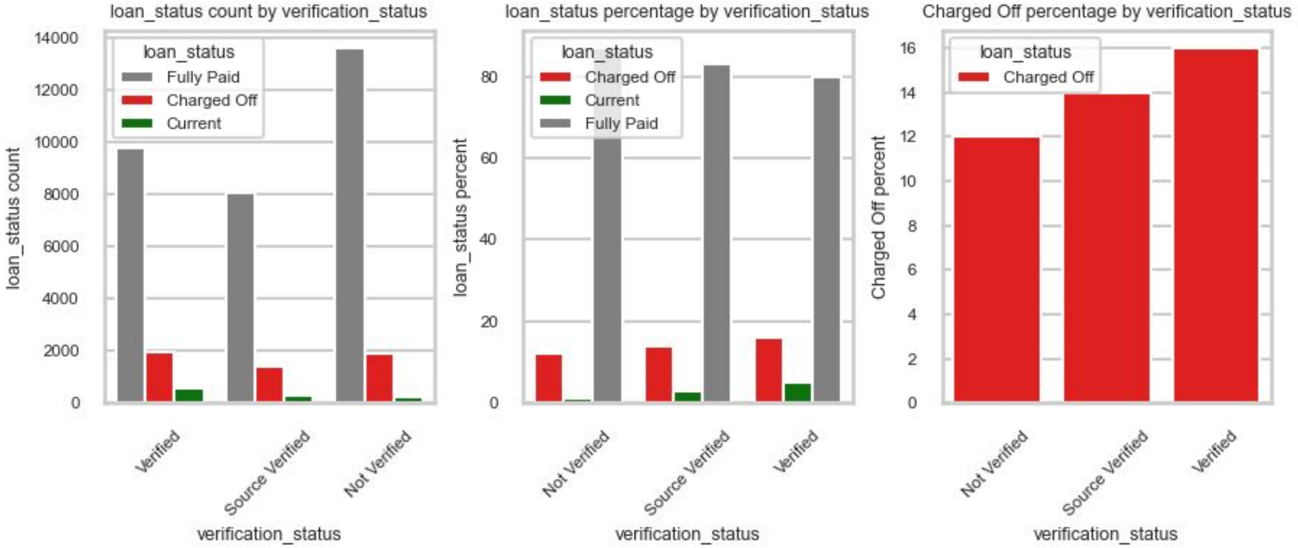
This project is an exploratory data analysis (EDA) on the lending club case study dataset. Lending club is a US based lending platform. Unlike banks, where the loans are provided partly by the investors and partly by the lending platform.

The objective of this case study is to identify possible reasons for loan defaults (a.k.a "Charge Off"). There are various categorical and numerical variables present in the dataset relevant to the entire journey of lending and payments(or recoveries in case of "Charge Off"s). More information describing the dataset can be found in the below sections. Subsequent sections will also show data manipulations, analysis and conclusions.

Analysis Approach

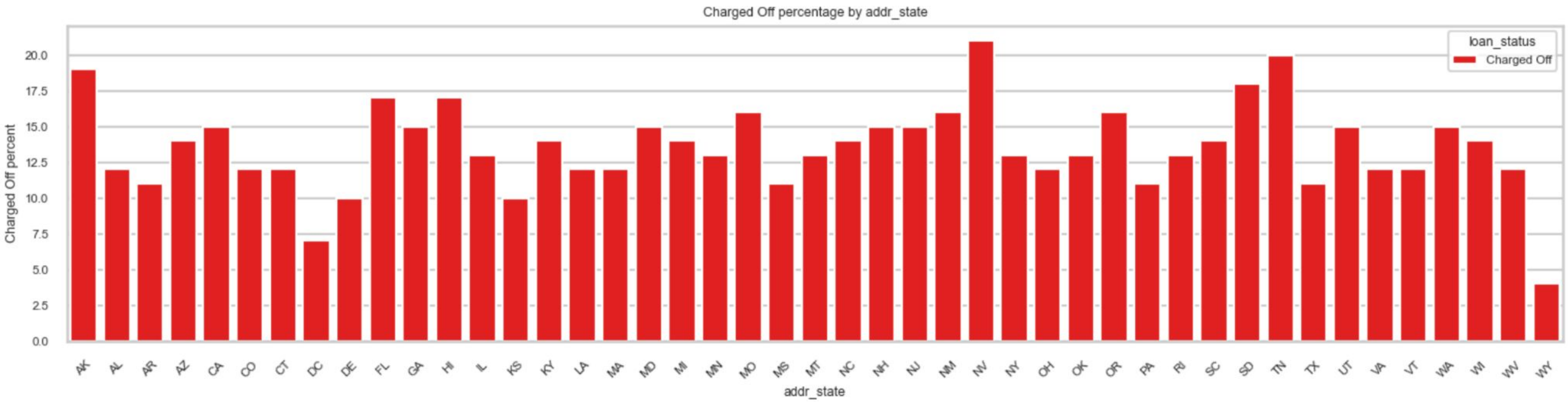
- Domain understanding with the help of tools like chat-gpt and bucketing the relevant data columns to respective entities
- Data cleaning and manipulation
    - Removal of empty rows and columns
    - Removal of highly correlated columns and keeping one out of the set
    - Removal of single value columns
    - Removal of columns based on business needs
    - Expanding date columns, fixing string columns
- Data analysis as per business needs (loan status as the central variable)
    - Segmented Univariate analysis for both categorical and binned numerical columns
    - Bivariate and Segmented Bivariate analysis for numerical columns
    - Multivariate analysis
- Observations and Conclusions

Observations



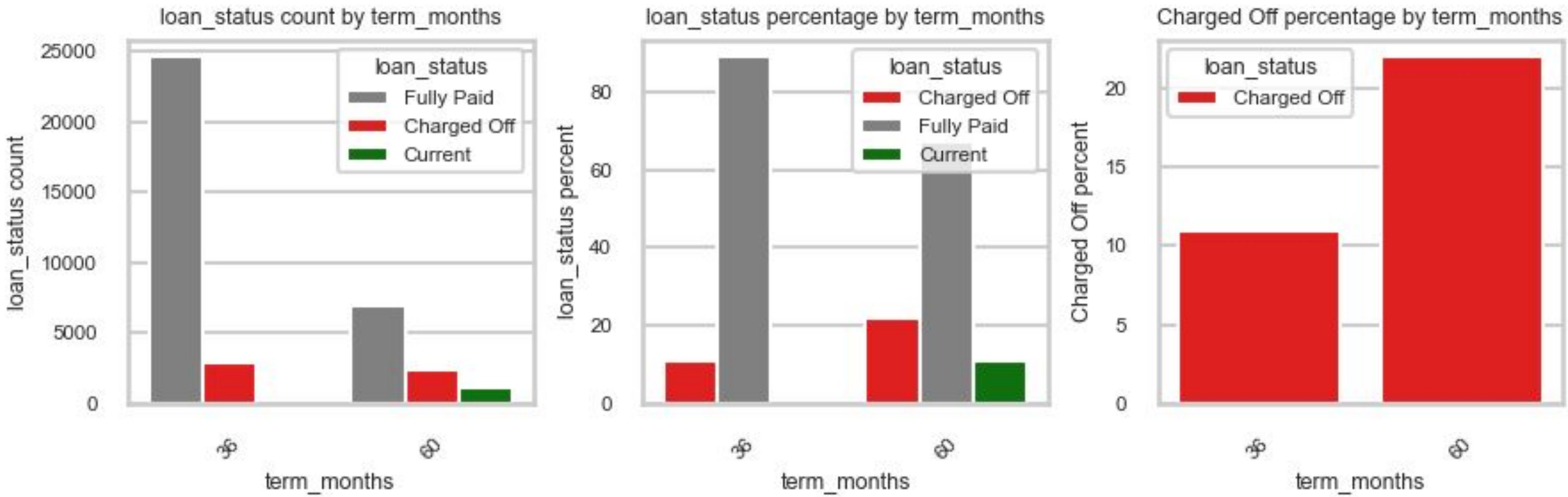
The borrowers whose annual income was verified default more

Observations



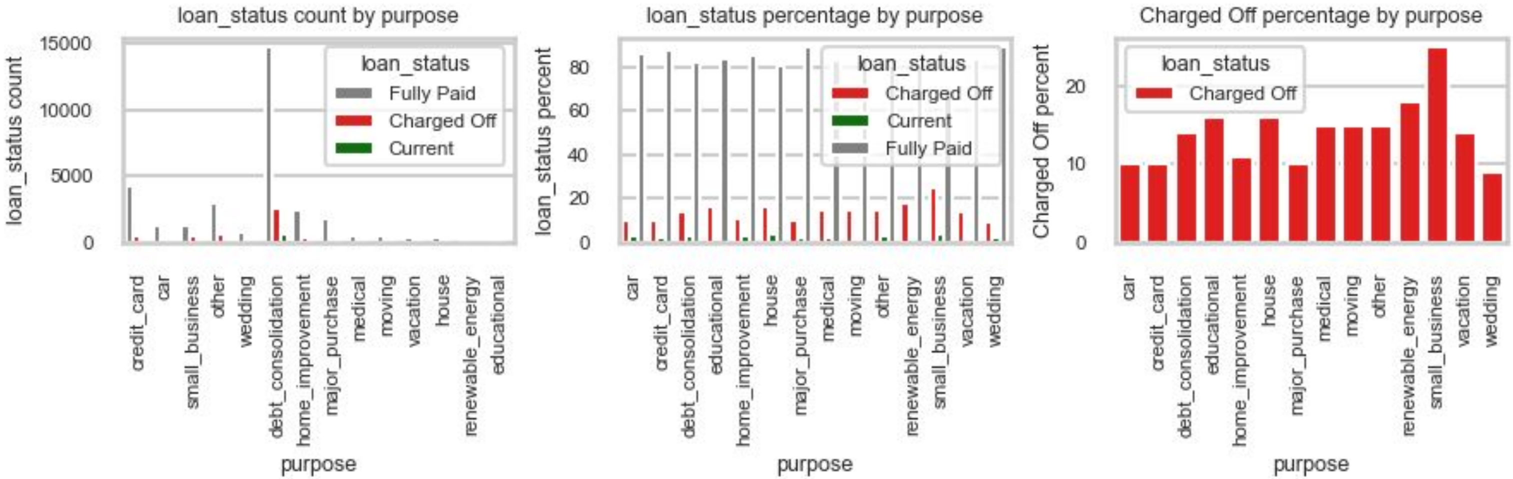
The state of Nevada has proportionally more defaulters

Observations



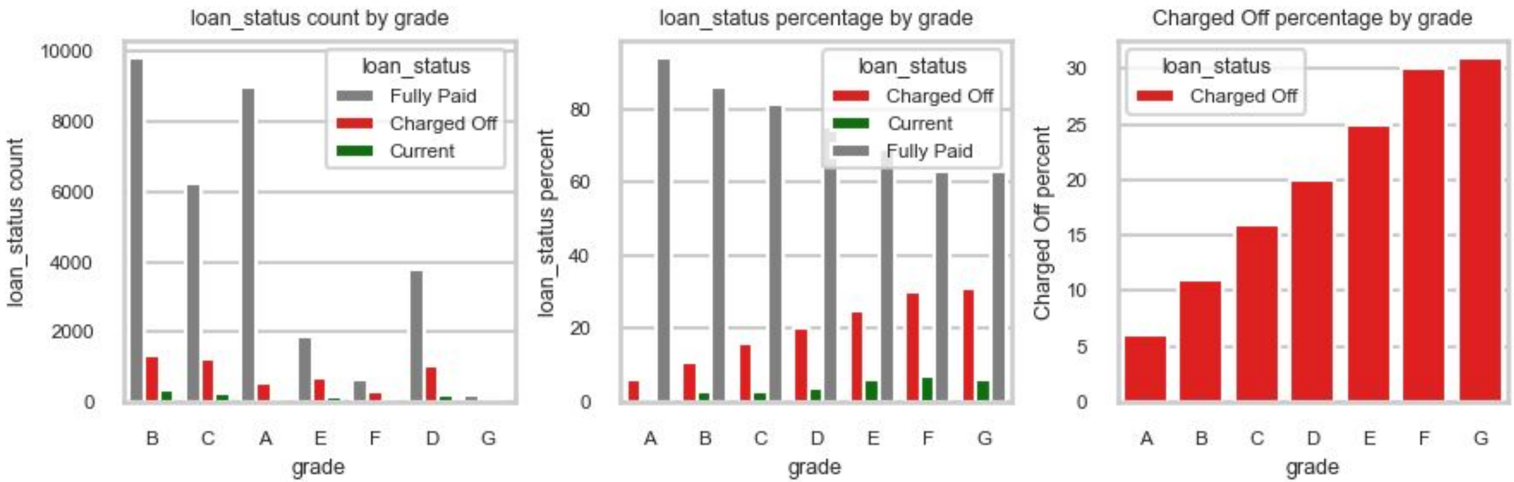
Borrowers who take long term loans default more

Observations



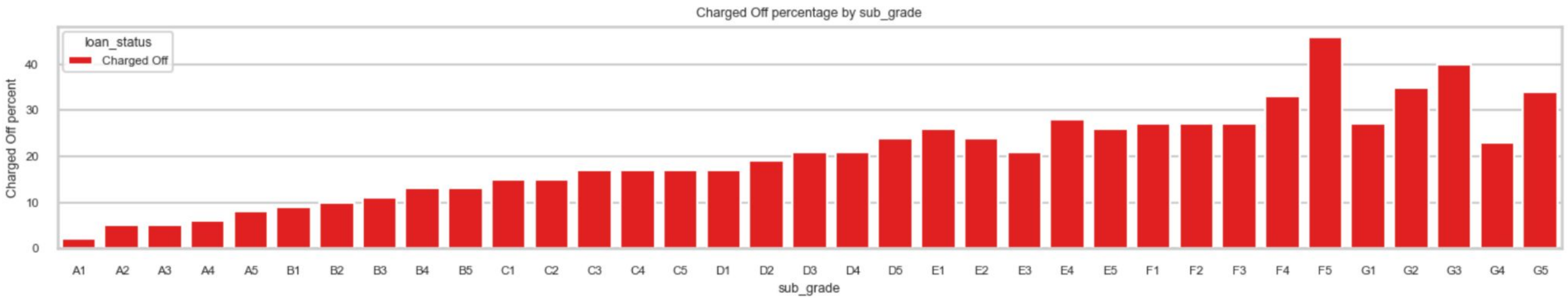
loan for small businesses tend towards more defaults as compared to other purposes

Observations



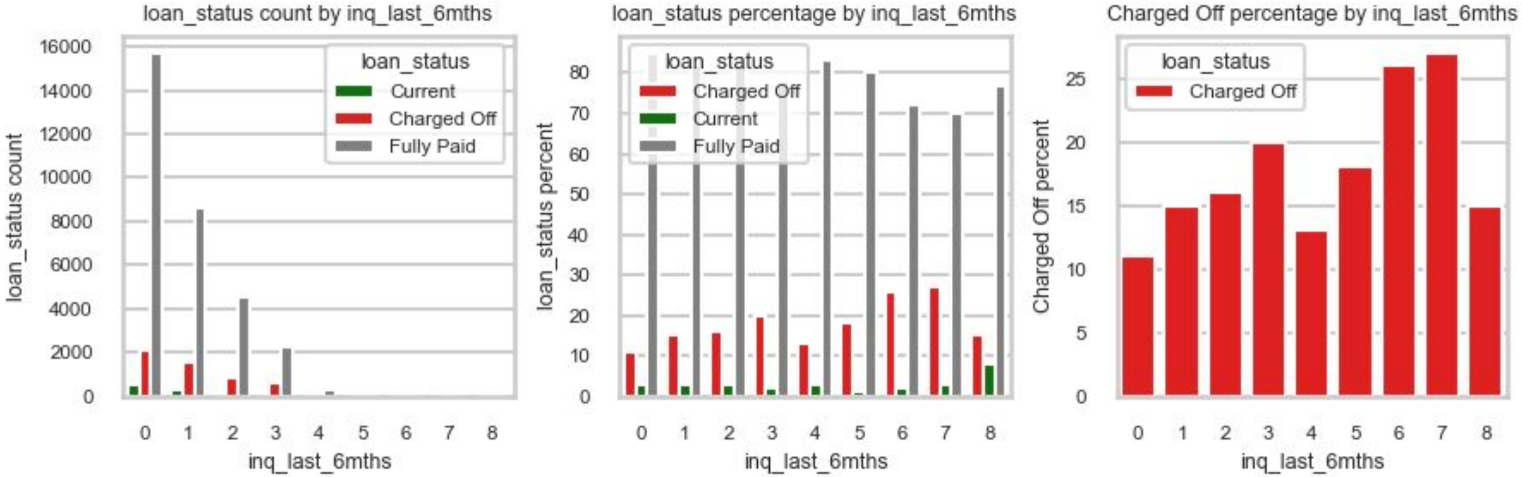
The percentage wise charge off as per the grades is as expected

Observations



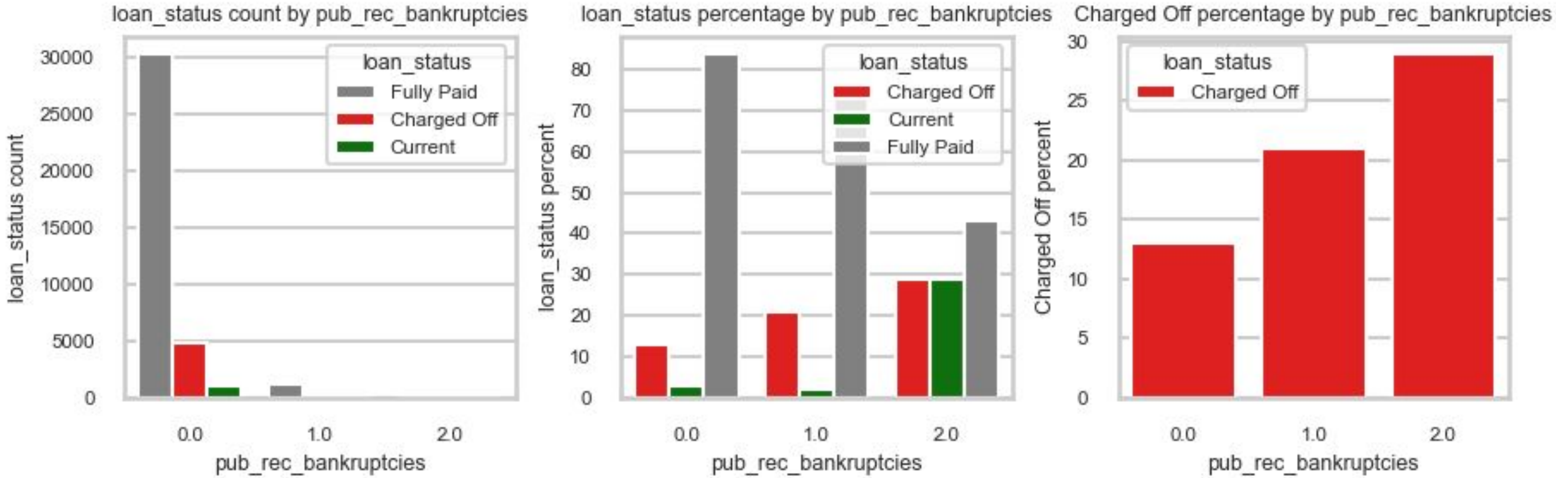
the assigned sub-grades are mostly matching to their expectations except for "G" sub-grades

Observations



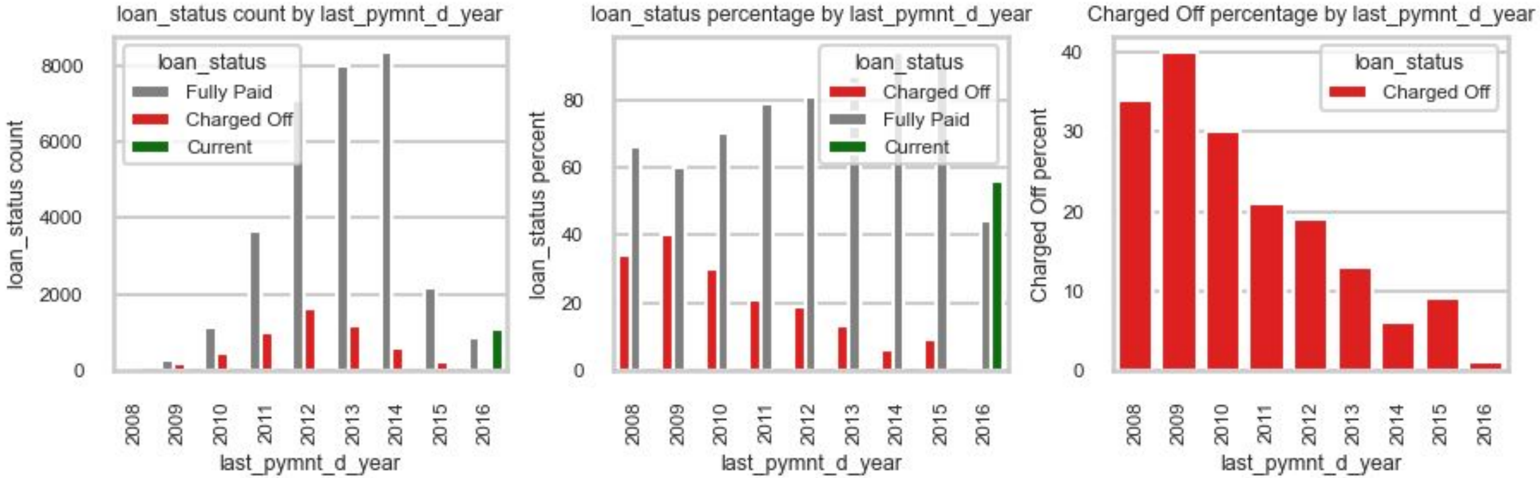
proportionally, the more enquiries that happen on the credit profile in the last 6 months, tend to charge off more

Observations



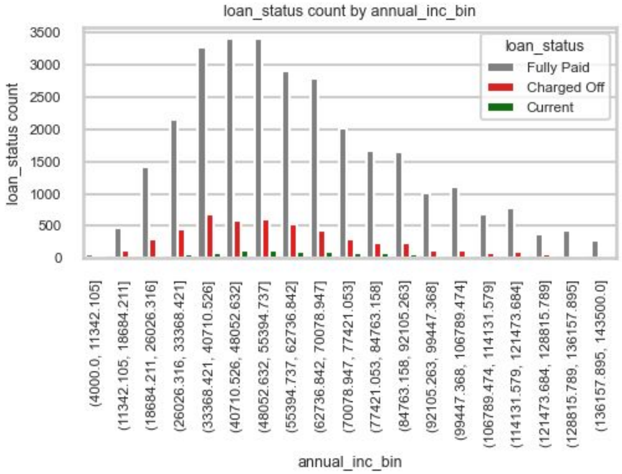
not a lot of borrowers have publicly declared bankruptcies, but those who have are more likely to charge off

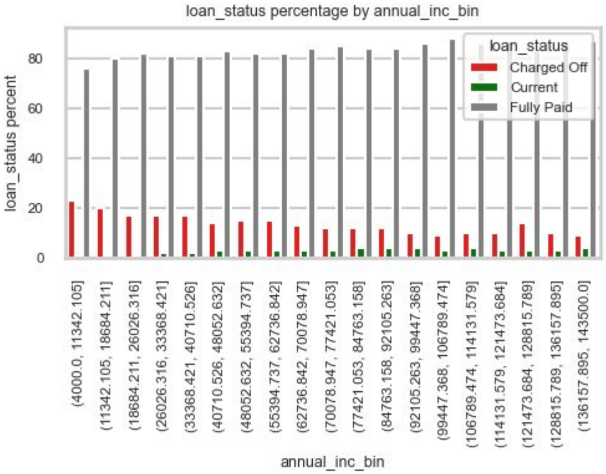
Observations

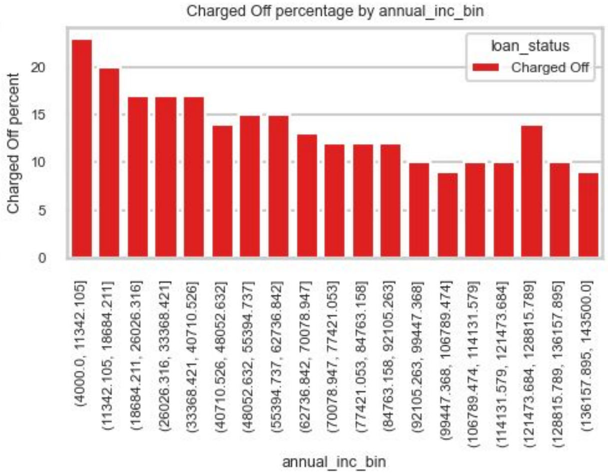


a lot of last payments happened in the year of 2008 and 2009 due to recession and were declared charged off since they couldn't pay after 2009

Observations

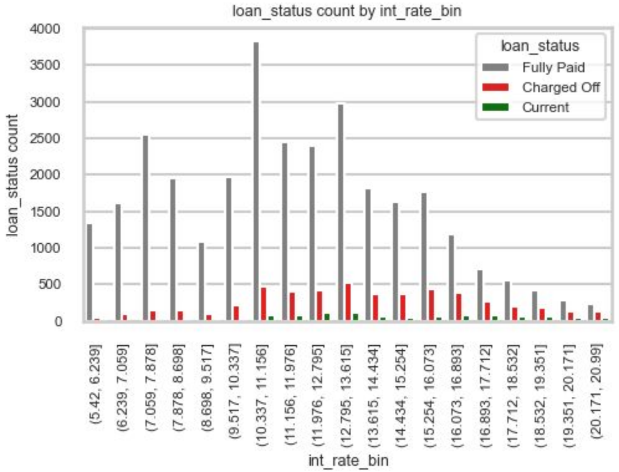


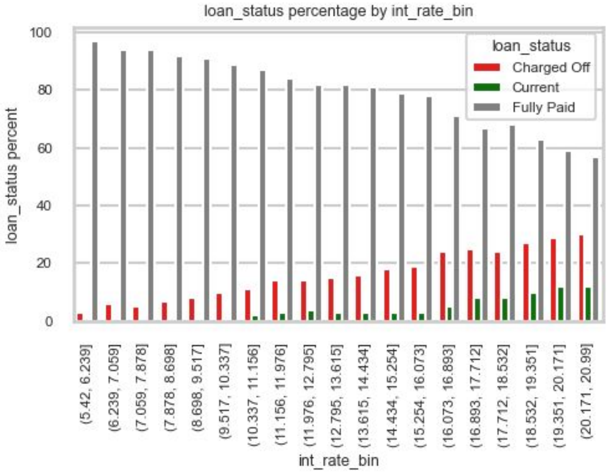


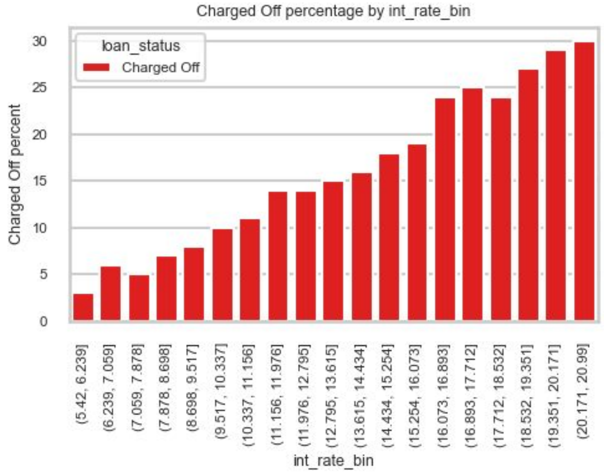


borrowers having high annual income tend to default less

Observations

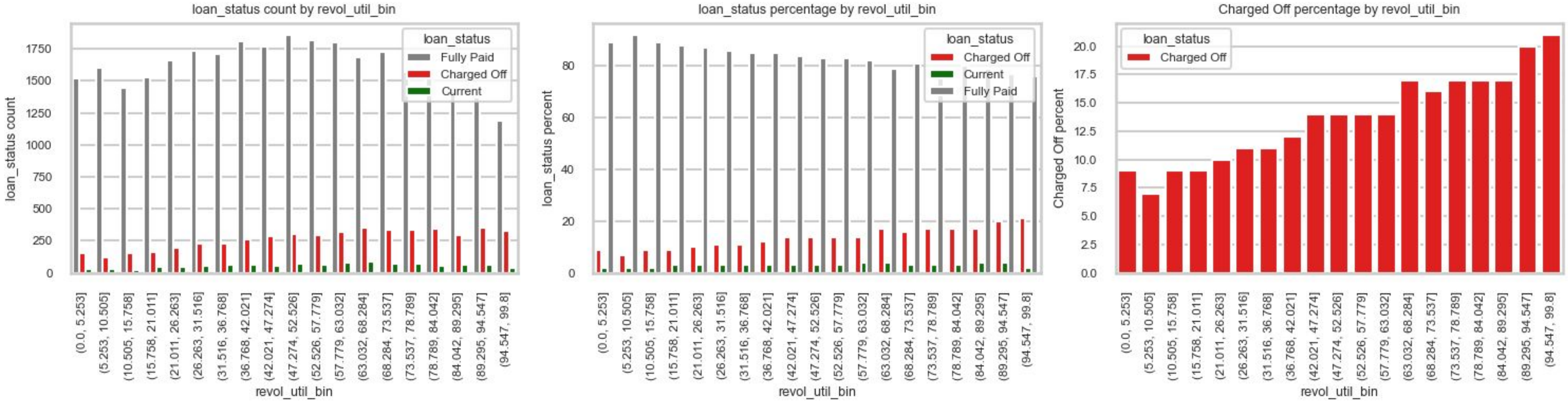






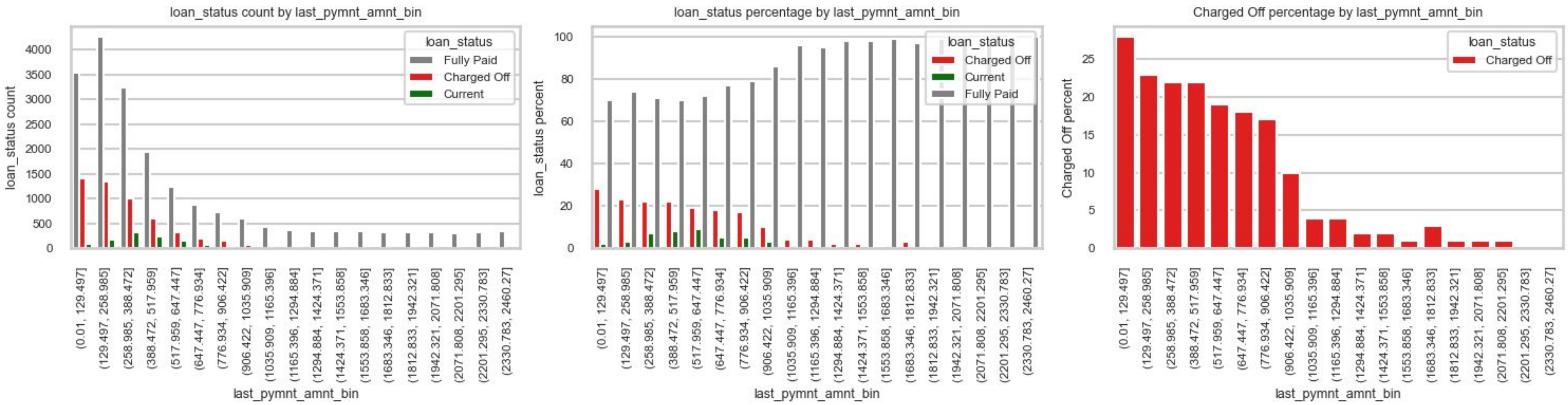
the higher the interest rates, the higher the chances of being charged off

Observations



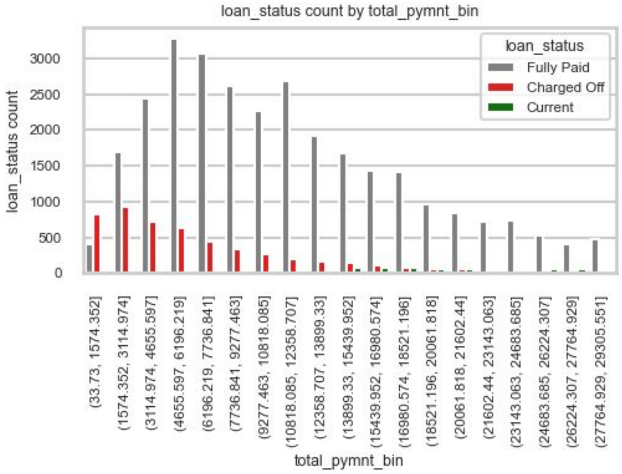
higher utilization of credit limit shows signs of distress and a weak financial stability

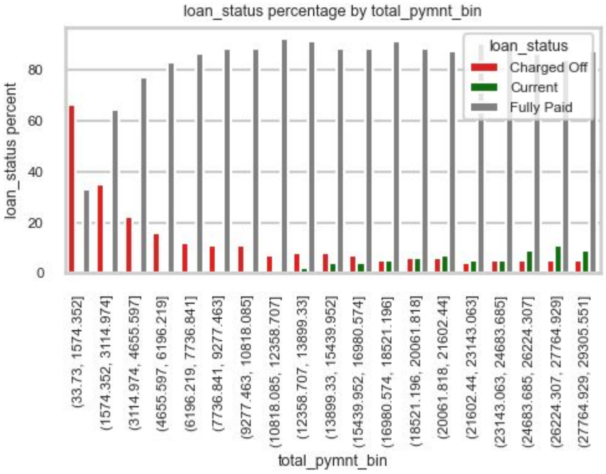
Observations

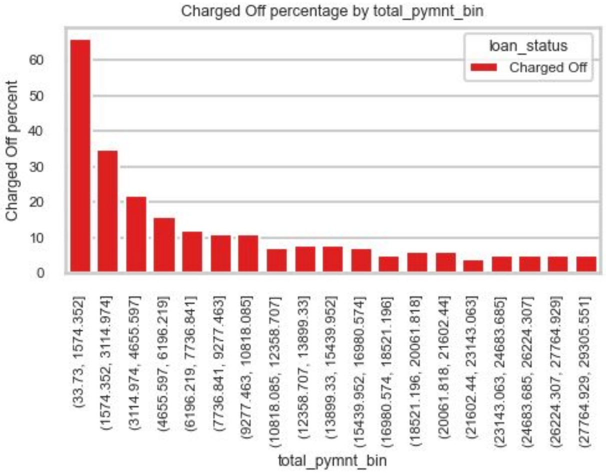


lower payment amounts tend to charge off more

Observations

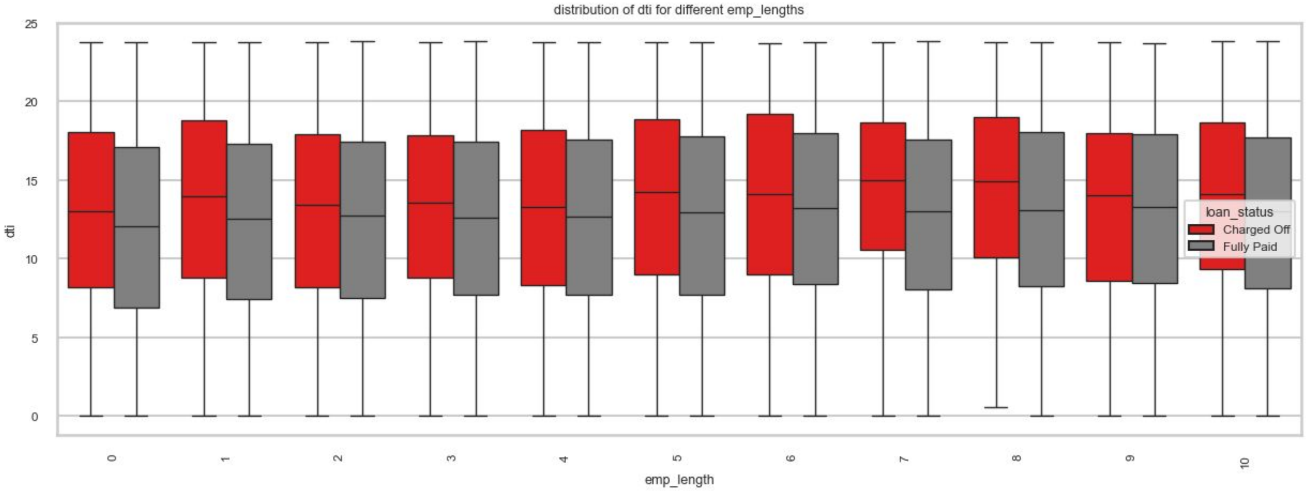






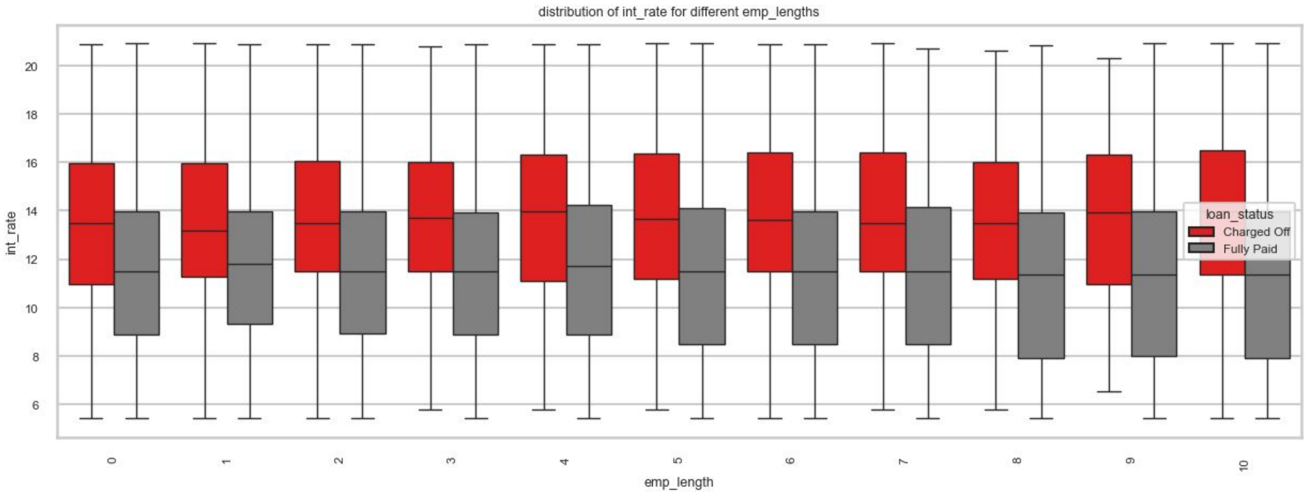
borrowers who have done more payments show intent of fully paying the loan

Observations



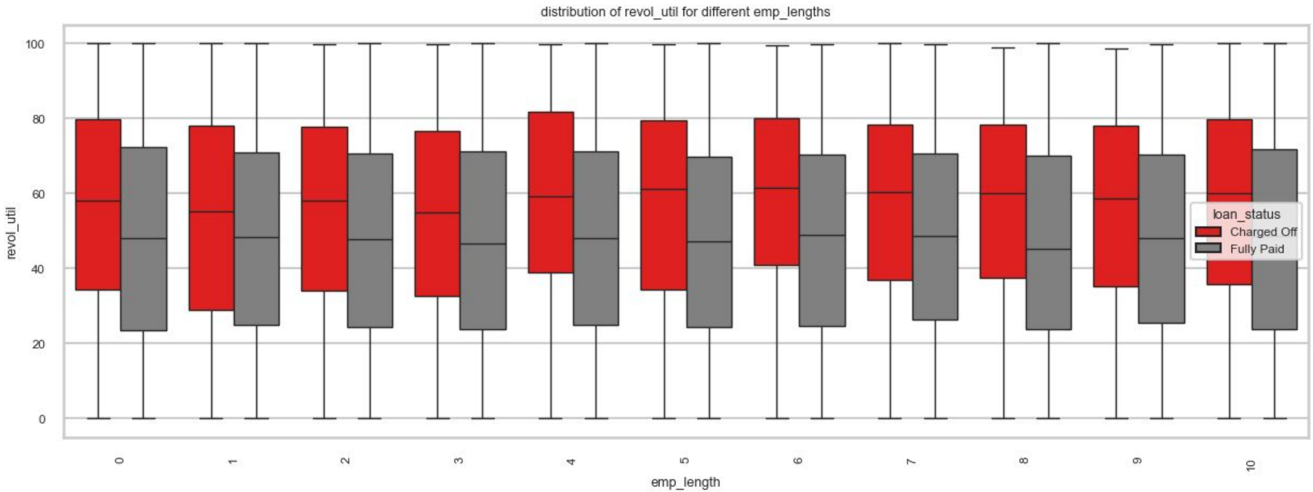
in all the debt to income ratio(dti) boxplots, the median of charge offs is higher than fully paid. Borrowers having high debt to income ratio across all segments have higher chances of charging off

Observations



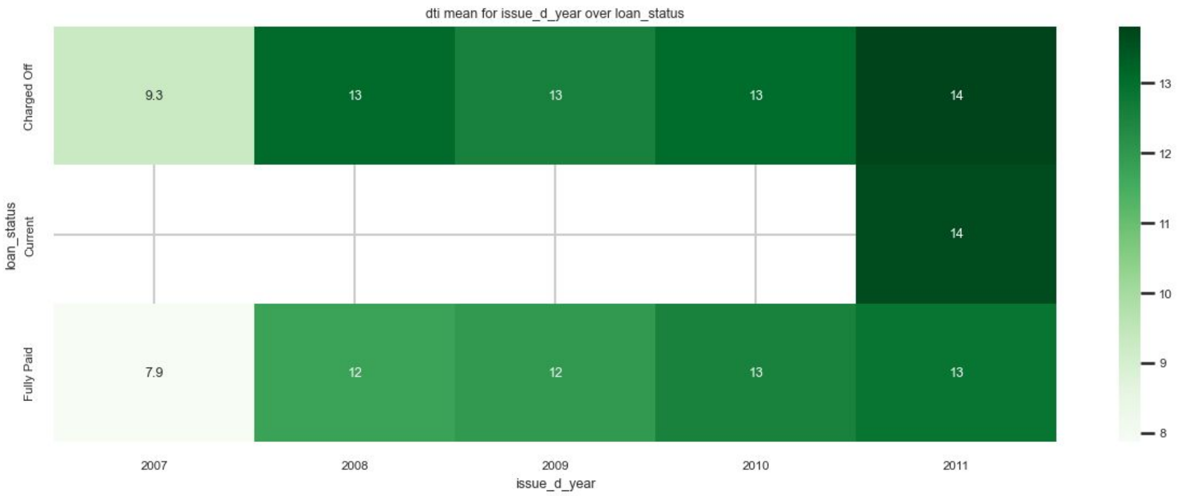
in all the interest rate boxplots, borrowers who charged off have a high median of interest rates as compared to borrowers who fully paid

Observations



in all the revolving utilisation (higher usage of credit limit), borrowers who charged off have a high median of revol_util as compared to borrowers who fully paid

Observations



Debt to income ratio has been increasing over the years, which is also leading to higher charge offs

Conclusions

- Longer terms of loans come with a risk of higher chances of being charged off
- There seems to be a discrepancy with the way borrowers are being verified for their annual incomes because the higher the annual income, the lesser the charge offs.
- a lot of last payments happened in the year of 2008 and 2009 due to recession and were declared charged off since they couldn't pay after 2009
- High values of debt to income ratio (dti), higher interest rates(int_rate) being offered to the borrowers and higher credit limit utilisation (revol_util) tends to show a pattern of the loan being charged off
- Publicly recorded bankruptcies tend to be correlated to charge offs
- The Lending Club Grades & Subgrades looks good enough to describe the risk.