Lending Club Case Study Analysis Submission

Nagendra Avadanam

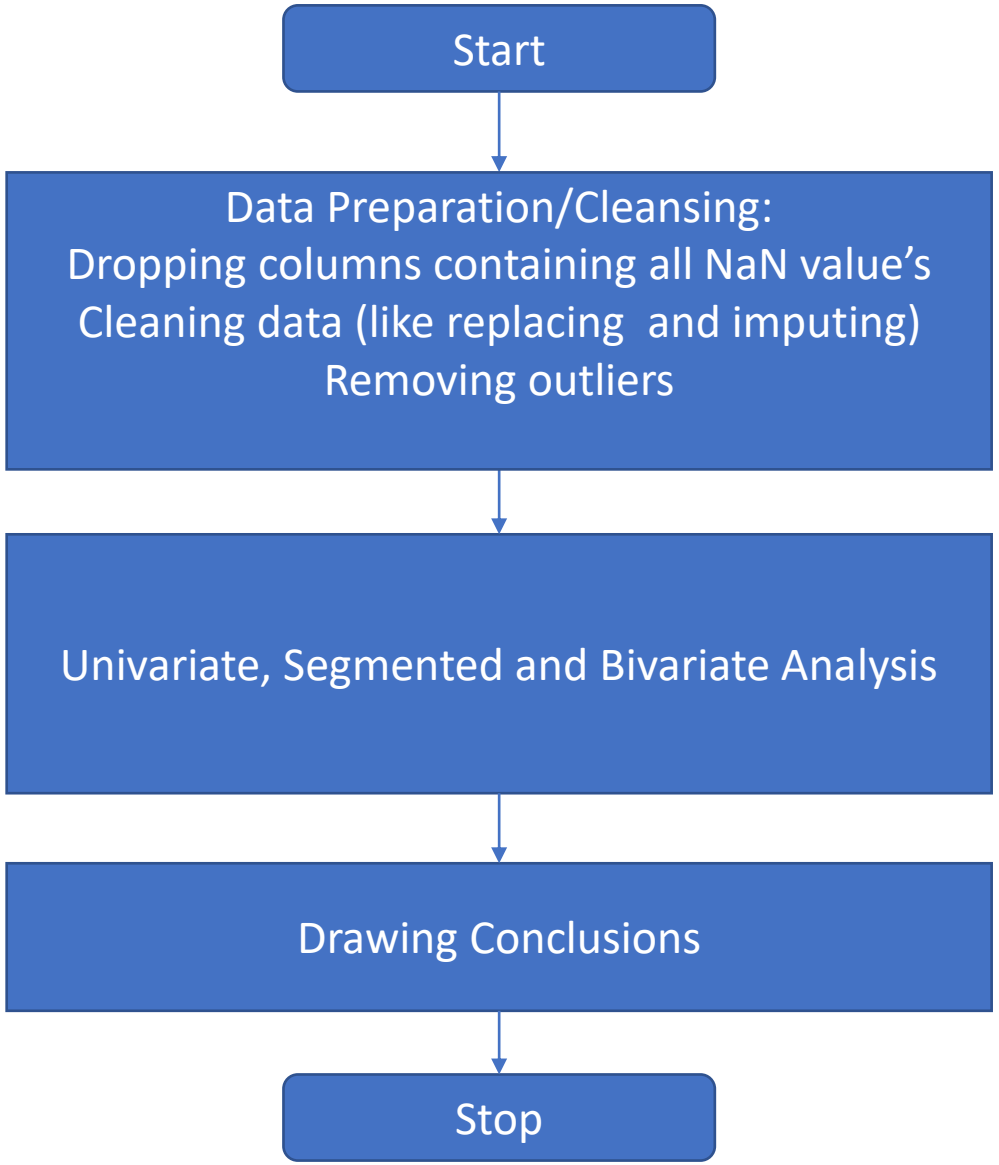
Case Study Objective and Problem Statement

Lending Club: Lending club helps members to pay down-high interest debts, save money and take control of their financial future.

Case Study Objective: Lending club wants to understand the driving factors behind loan defaulters via variables which are strong indicators of default.

Problem Statement: Identifying the loan defaulters from the provided dataset, by using the EDA to understand how attributes are moving to the tendency of default.

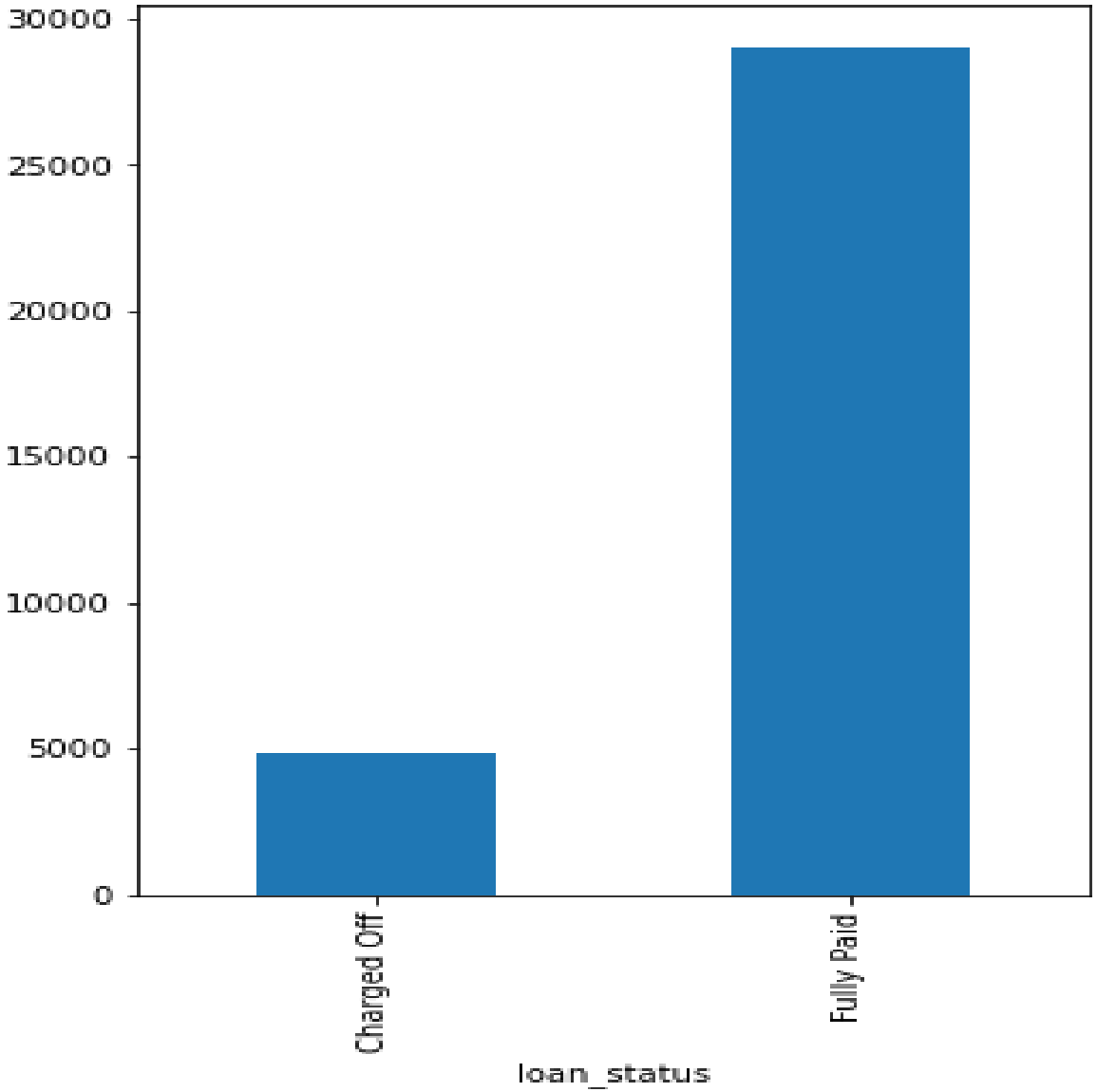
Analysis Approach (Flowchart)



Data Preparation/Cleansing:

- Delete attributes, where there are maximum number of Null/NaN values.
- Remove outliers: Remove high and low values that would disproportionately affect the results of your analysis.
- Missing values: Treat missing values with appropriate approach.
- Duplicate data: Remove identical rows, remove rows where some columns are identical.
- Filter rows: Filter by segment, filter by date period to get only the rows relevant to the analysis.
- Performed data cleaning and preparation on the Loan dataset:
- Created 16 new attributes: 'earliest_cr_line_month', 'earliest_cr_line_year', 'issue_d_month', 'issue_d_year', 'loan_amnt_bin', 'income_bin', 'dti_bin', 'revol_bal_log', 'revol_util_bin', 'last_pymnt_d_month', 'last_pymnt_d_year', 'last_pymnt_amnt_log', 'last_credit_pull_d_month', 'last_credit_pull_d_year', 'ratio', 'ratio_bin'
  Profit and Loss column
- During univariate analysis created distribution plots and box plots for all attributes
- Histograms and Bar charts to check out the distribution of all the driver variables
- Box plots to detect the Outliers
- Performed the Multivariate analysis to understand how different variables interact with each other.

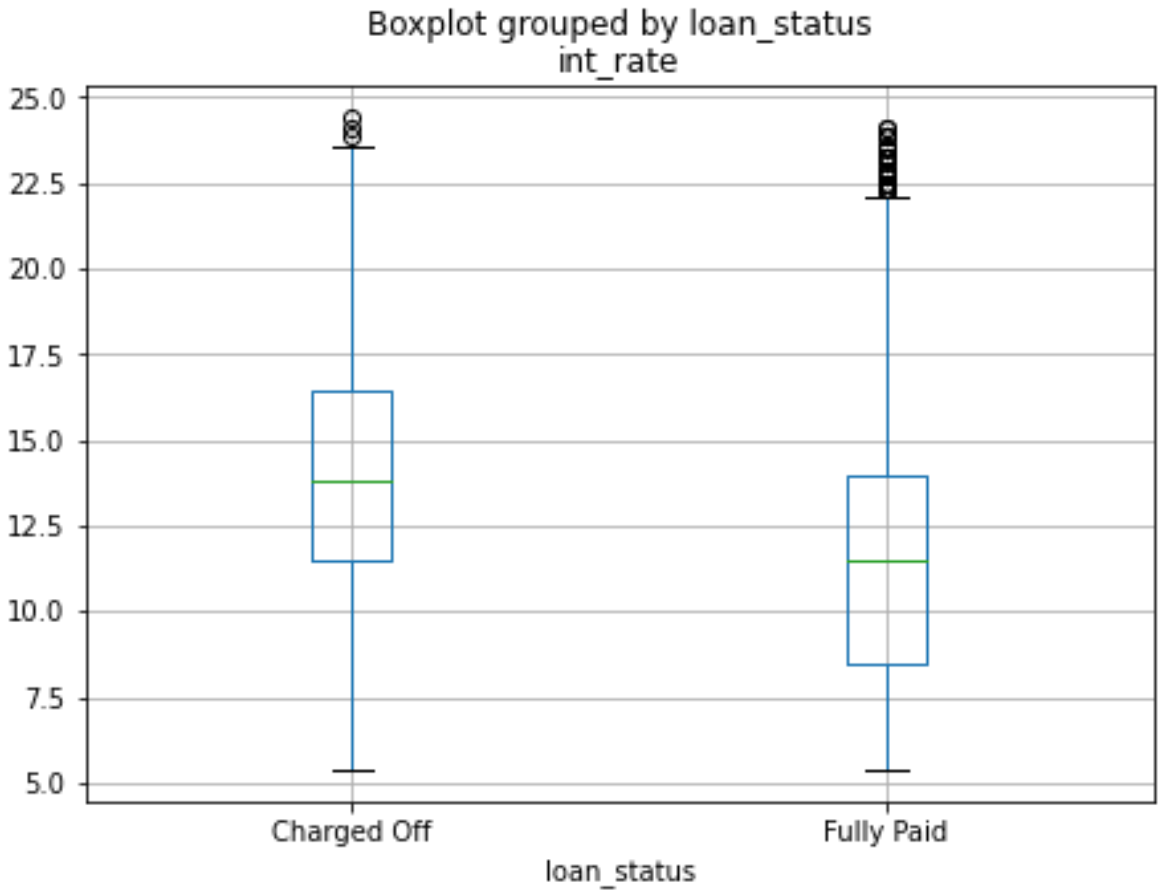
Loan Status Analysis



Observation:

1. Most of the loans are Fully Paid.
2. About 14% of loan are having status as defaulters.

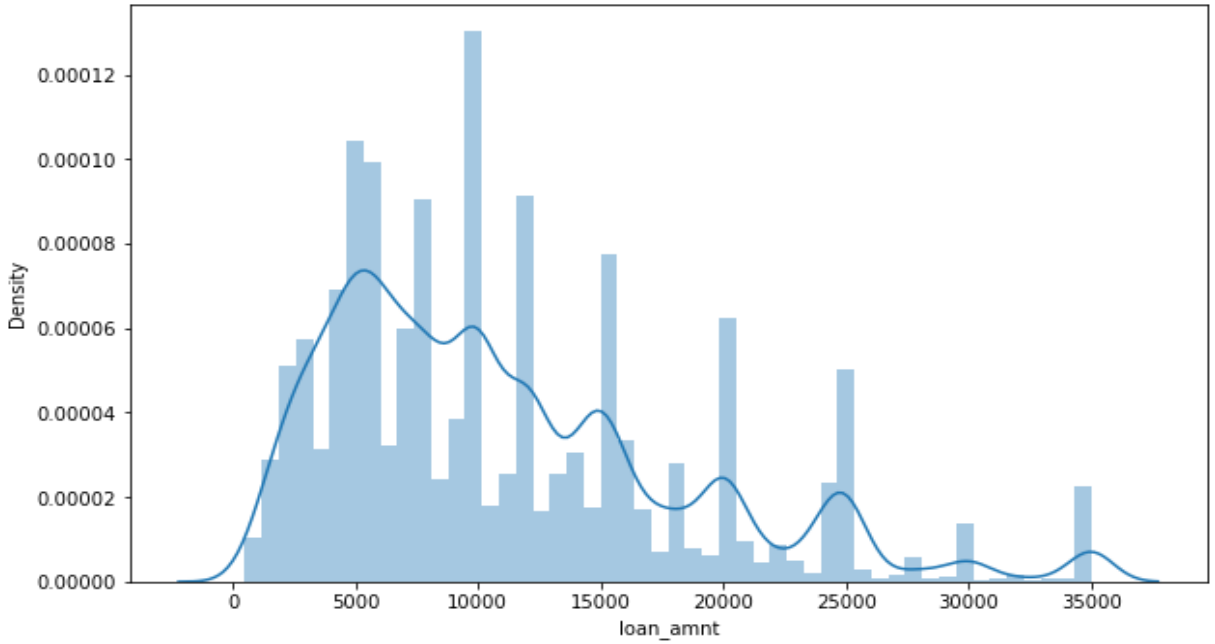
Interest_rates vs Loan Status Analysis

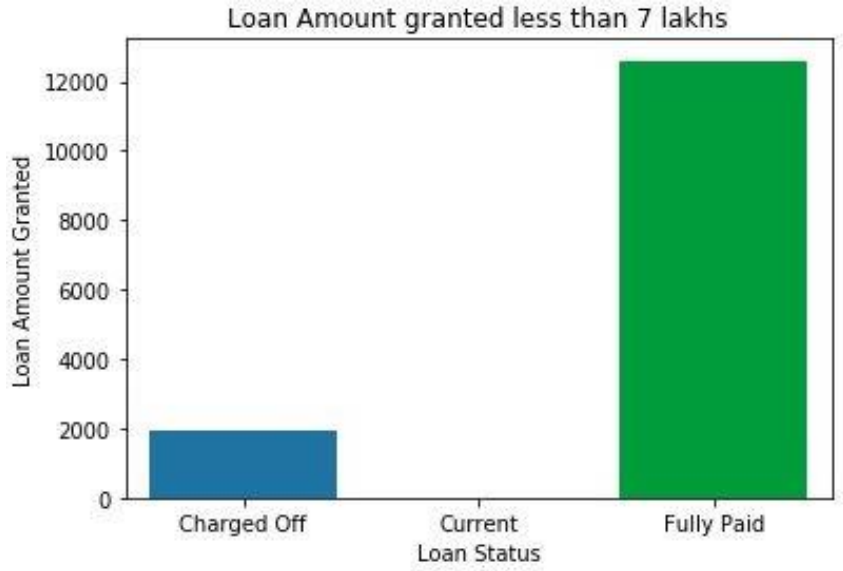


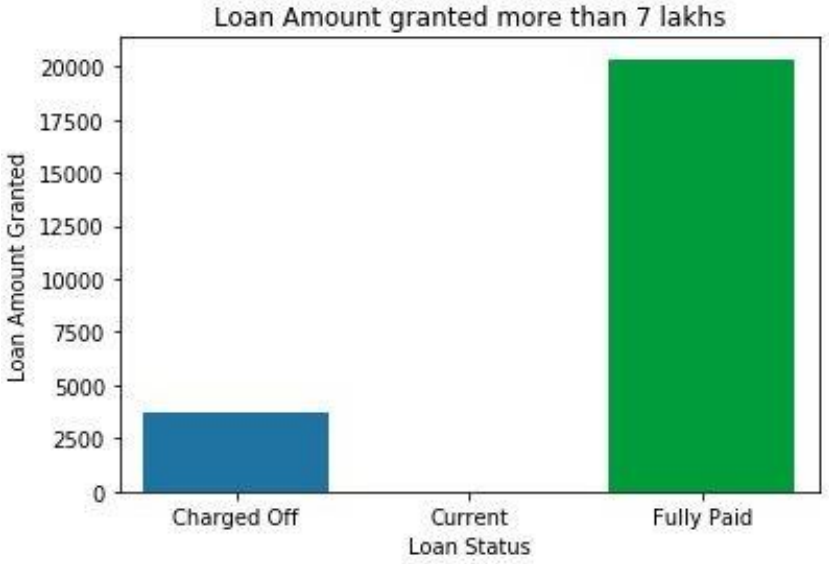
Observations**:**

Loans at a higher interest rate are more likely to be Charged Off.

Analysis on Loan Amount



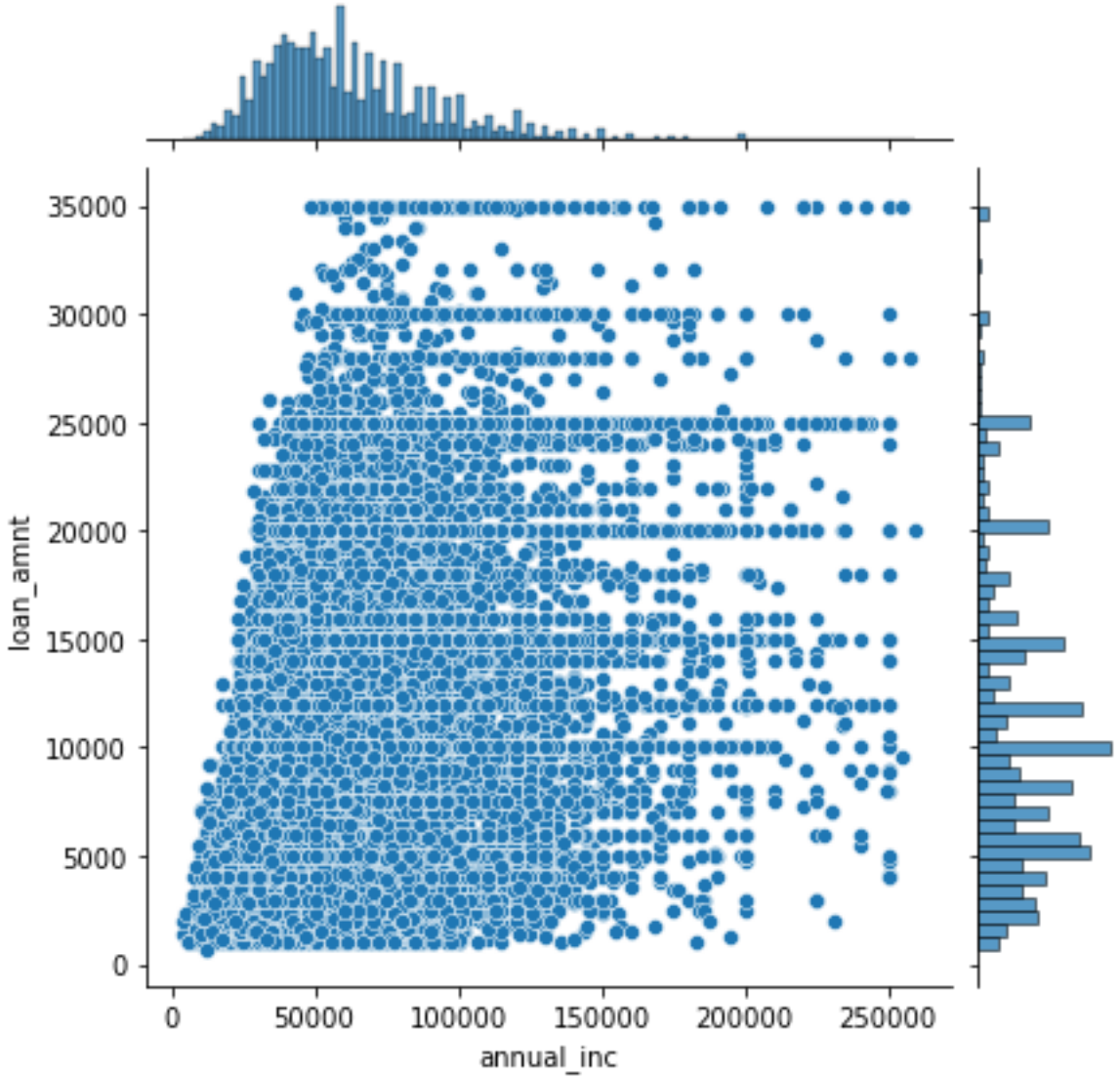




Observation:

1. Funded amount is left skewed. Most of the loan amount given is below 7 lakhs.
2. Probability of people with loan amount greater than 7 lakhs tends to default is more than the people with less than 7 lakh

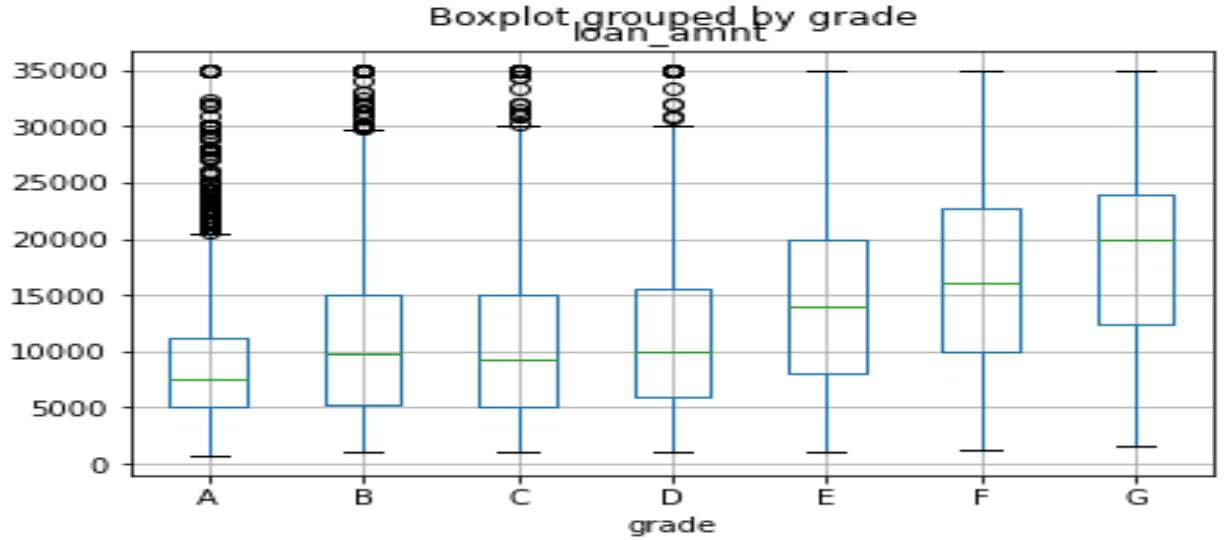
Analysis on Loan Amount Vs Annual Income

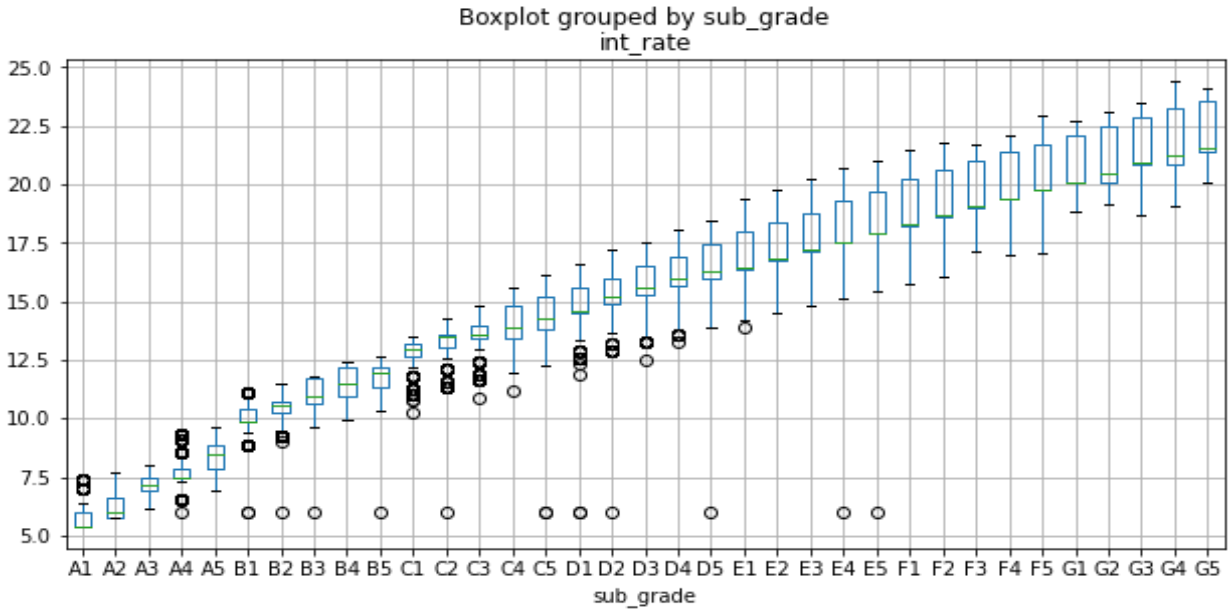


Observation:

There are people with average income lower than 50000 taking loans of 25000 or higher. These would be risky loans.

Analysis on grade vs loan_amnt and subgrade vs int_rate



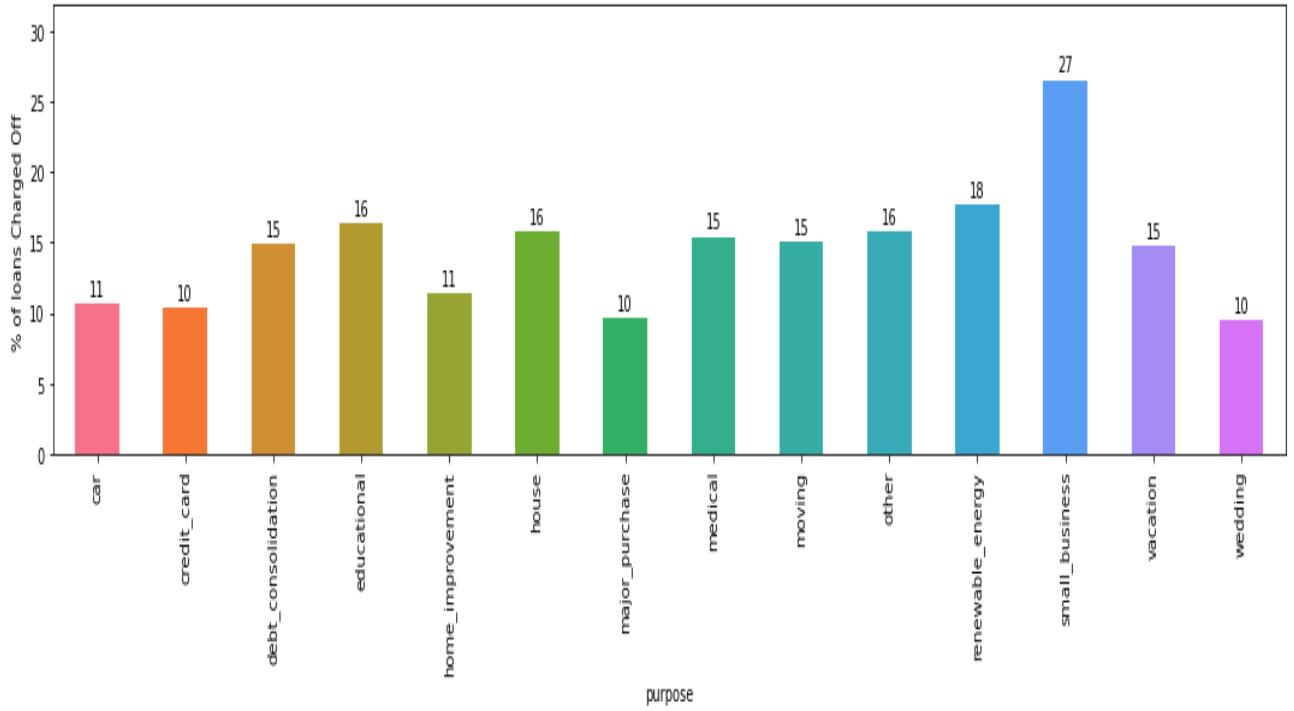


Observations:

Larger loans generally appear to be given a lower grade, with the median loan amount for a grade G loan being almost 10000 higher than that of a grade A, B, or C loan.

Intrest rates varies directly with the subgrade. Larger or worst the sub grade, higher are the rate of intrest for the loan.

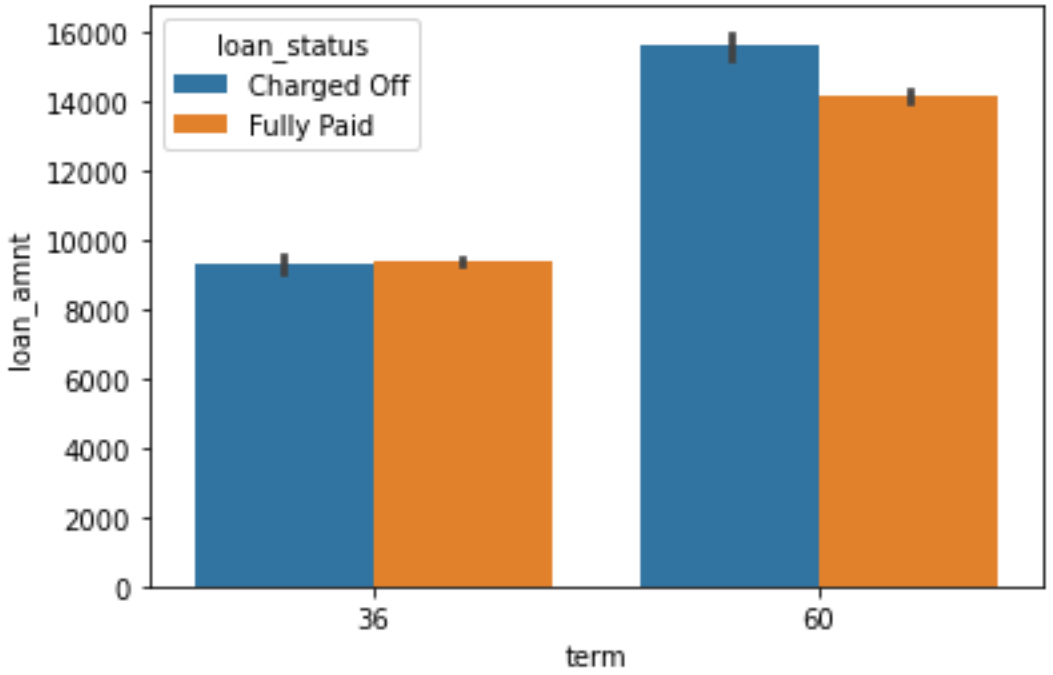
Analysis on Purpose



Observation:

Limiting the number of approvals where purpose is “small business“

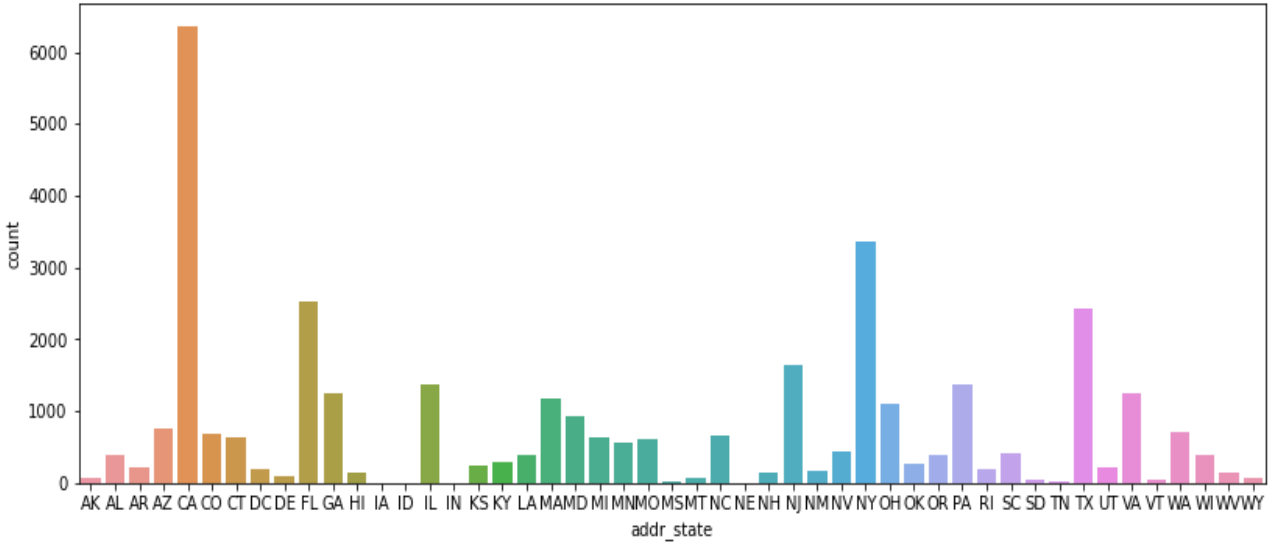
Analysis on loan_amnt vs term vs loan_status

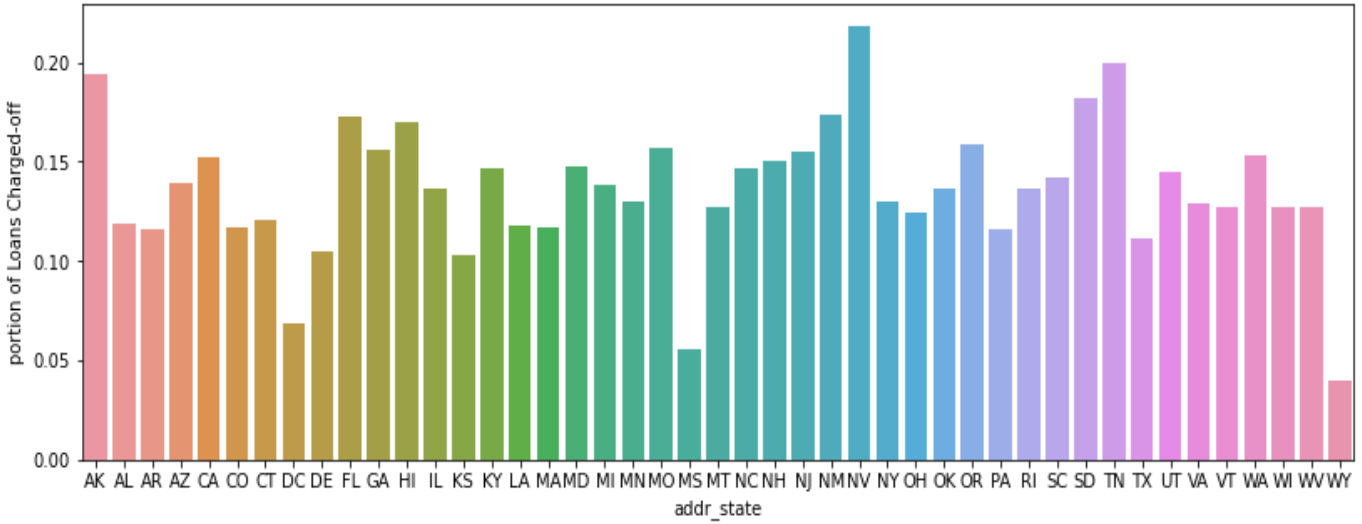


Observations:

Higher loan amount are associated with longer terms and see higher Charge Offs.

Analysis on Addr_state



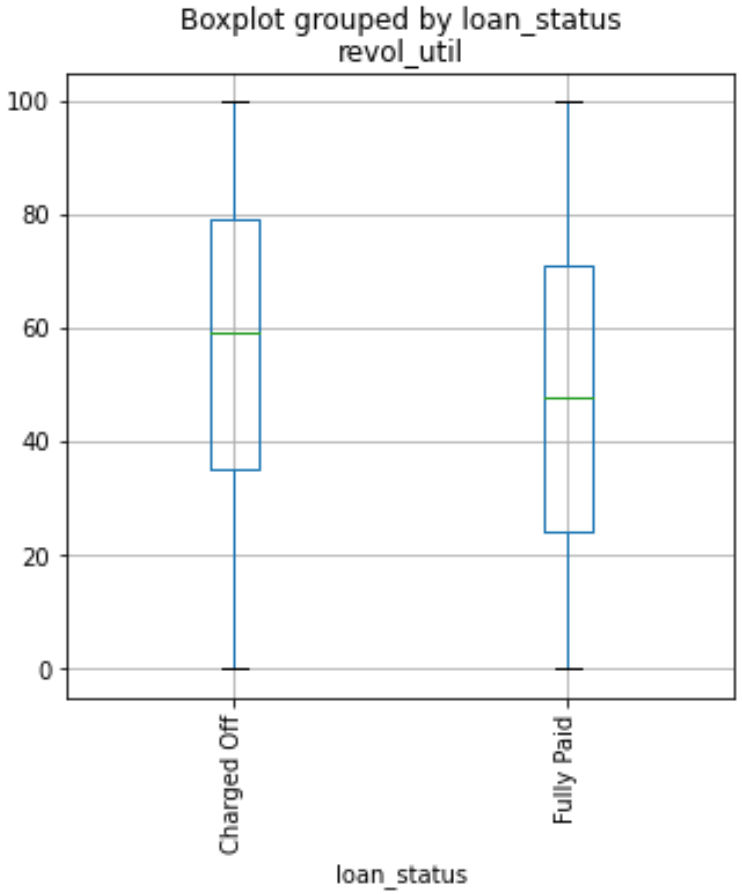


Observation:

Maximum loans ~18% are from California following that is 9.5% from New york state and Florida with 7%. This is to be expected as these are also the three most populous US states

•States with higher Charge Off rates have very low number of loans. The percentage is therefore NOT significant and should be ignored. Overall, this variable does not affect propensity to default.

Analysis on revolv_util

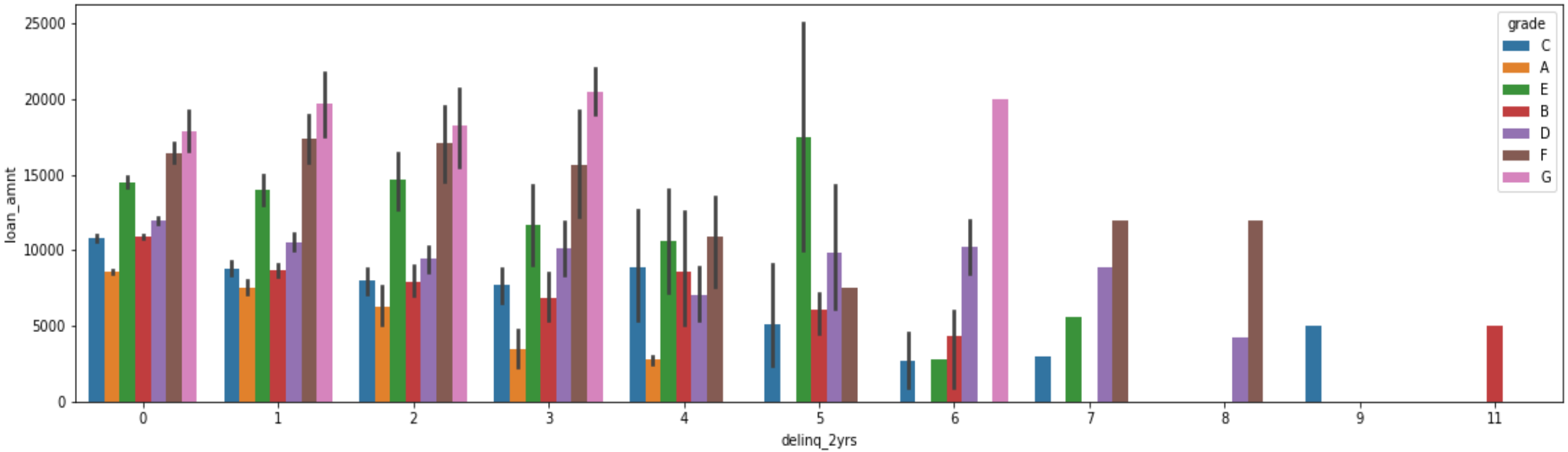


Observation:

The average utilization rate is around 49%

Average revolving Utilisation rate is seen to be higher for Charged Off loans than the Fully Paid loans.

Analysis on delinq_2years Vs Loan_amnt vs grade

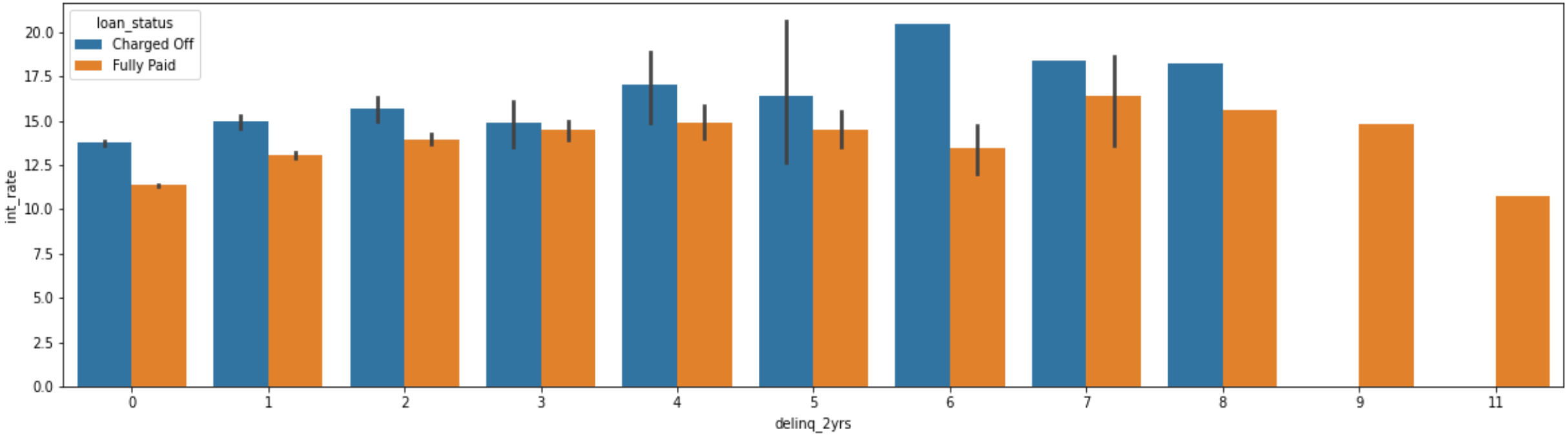


Observation:

1) Not many loans receive invstment with higher number of delinquencies(>3). Despite the low loan amount request, these loans are considered risky and are not invested much in.

2) Lending club should further restrict their investment. We see loan amounts of >15000 on average for people having >=2 delinquencies.

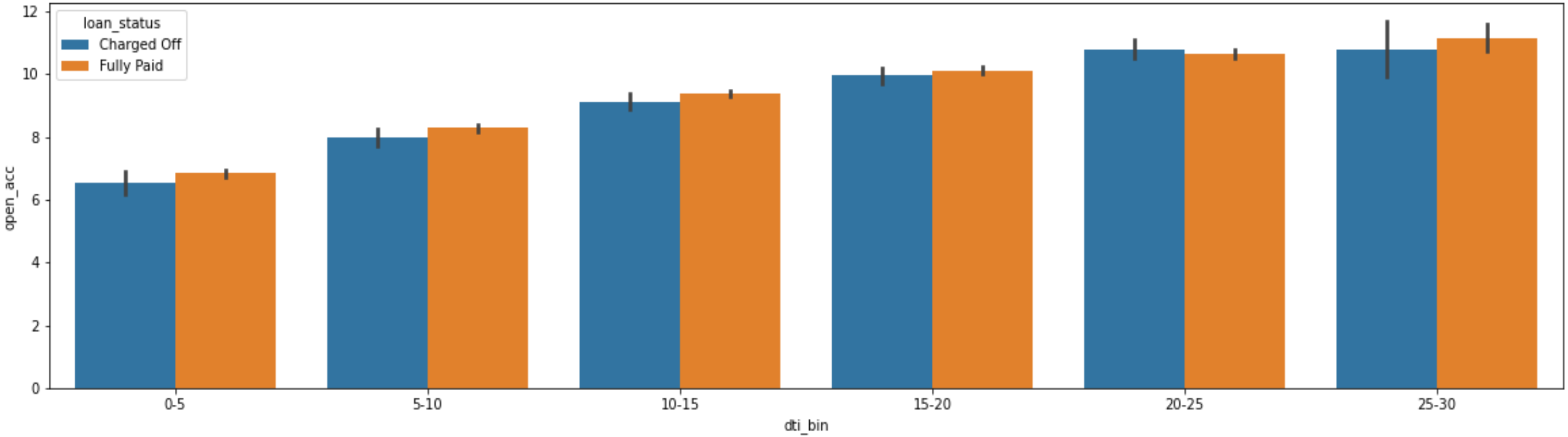
Analysis on delinq_2yrs vs int_rate vs loan_status



Observation

In general, interest rate offered increases with the number of deliquency of the borrower.

Analysis on dti vs open_acc



Observation:

The dti increses with the number of open credit line in borrowers credit file.

Conclusions/Recommendations:

- Check the background of applicant thoroughly if interest rate is high, loans having higher interest rate have more defaulters.
- Restricting approving loans, where amount/income is higher than 30%.
- Doing thorough analysis on applicants belonging to CA state, as tendency to default is high.
- Limiting number of approvals where purpose is small business.
- Restricting approving high-value loans when revolving line utilization rate greater than 75%.
- Low grade loans have high tendency to default. Grading system is working as expected.
- Restricting approving loans to people with prior bad record. Or at least stop approving high-value loans.
- Restricting charging higher interest rates for loans with dti greater than 20 .