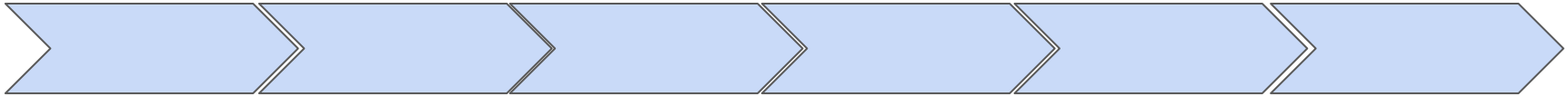
Lending club case study

Group Members:
Mohit Kumar Dubey
Raghuveer Kona

Agenda

- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Like most other lending companies,not lending loans and lending loans to ‘risky’ applicants are the largest source of financial loss (called credit loss).
- The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.
- The company can utilise this knowledge for its portfolio and risk assessment

Problem solving approach



Data Cleaning

Remove the more than 40% null values. Remove the columns that will not contribute in the analysis

Data Standardisation

- Standardise the data by removing the extras from value and standardising the data type
- Remove the outliers

Derive columns

- Derive new columns to create buckets with huge variation
- Derive new columns from existing values (like get month from date)

Univariate analysis

Analyse each column by plotting the distribution of values of that column and create

Bivariate analysis

Analyse the behaviour and relation between two columns and the correlation between them

Observations

Observe and understand all the relations between columns and give recommendations to reduce the loss

Data Cleaning

Dropped columns based on below

- Initial data was consisting of 111 columns
- Around 57 columns were have more than 40% of null values
- 9 columns were having only 1 value
- Around 15 columns were not contributing much to the loan analysis as the data had several data values that could not be categorized, consisted month or dates, were identifiers for individual entries and more.,

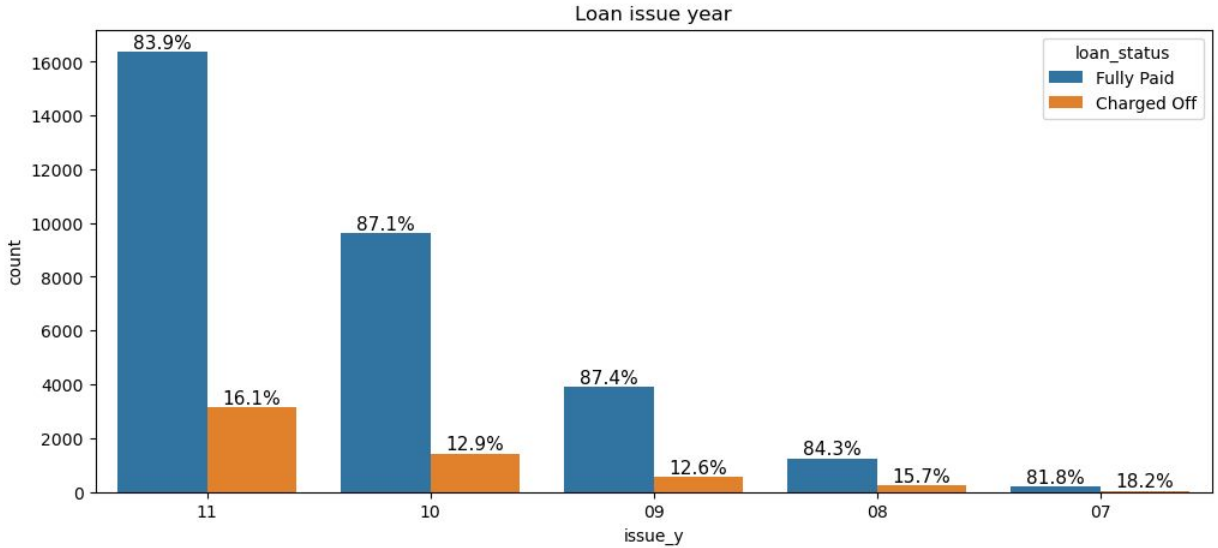
Data Standardisation

- Few columns were having extra delimiter, on removing those, these columns became useful for analysis(such as percentage).
- Removed the outliers from annual_inc.

Derive Columns

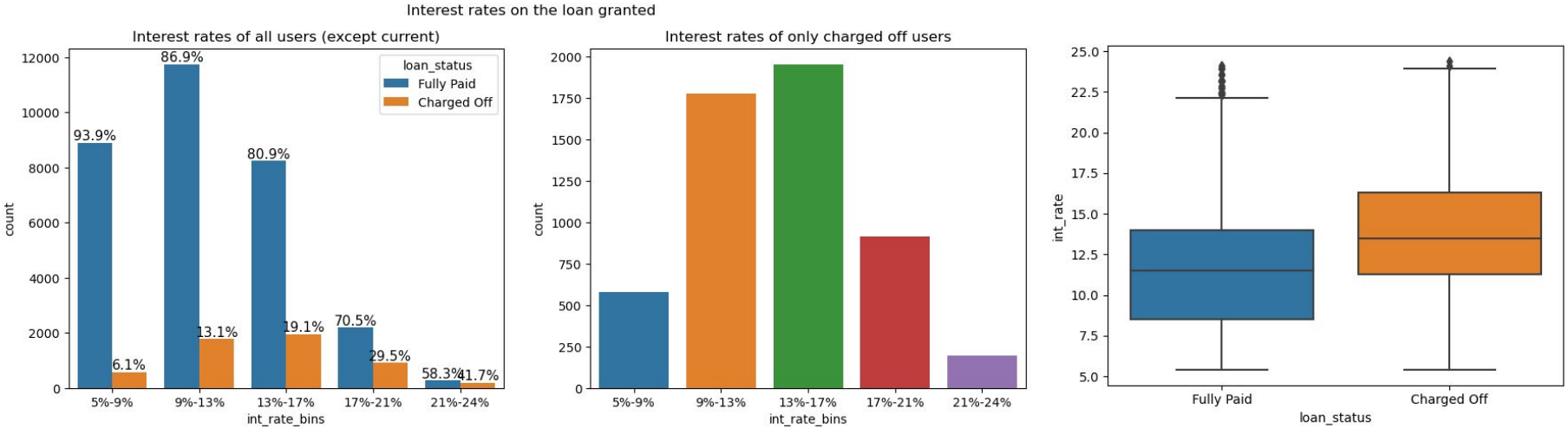
- Created bins for few columns that has very high variation in value like annual_inc, loan_amnt, int_rate etc
- Created new columns with month and year from issue_d column

Observation on loan



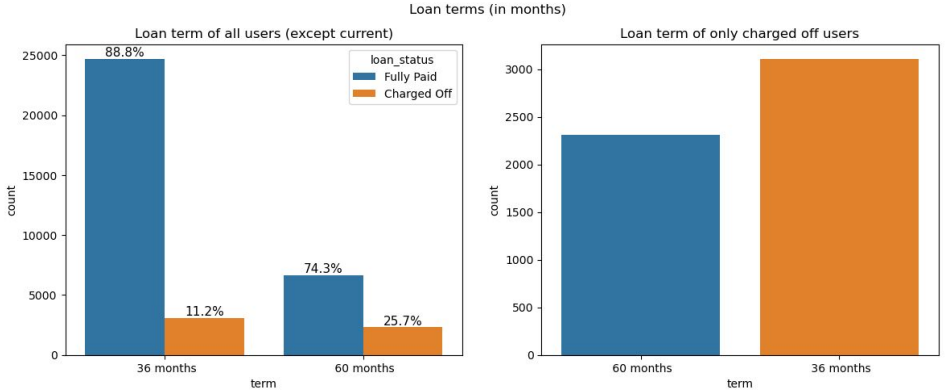
- Number of loans each year has increased from ‘07 to ‘11
- Highest defaulters % w.r.t all loans granted can be observed for loans issued in ‘07, ‘11 and ‘08
- Though we cannot draw any conclusion from the data alone, Considering the market knowledge we estimated that the reason could be the real state crisis that happened in year 2008

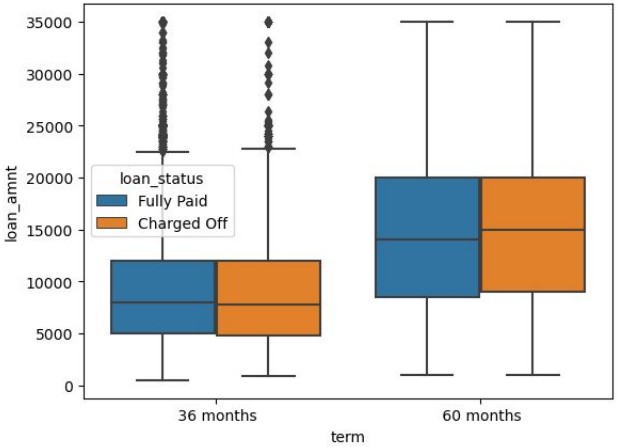
Observation on interest rates

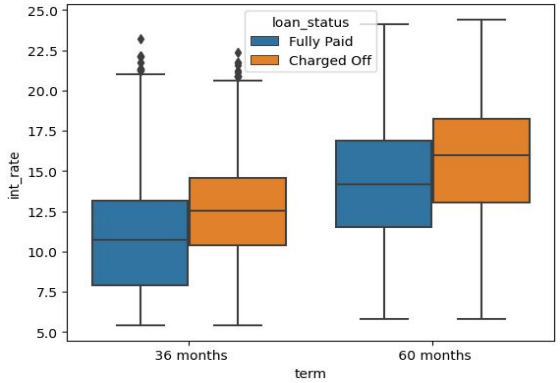


- The count of loans granted is high for interest bucket 9%-17% and so is the count of defaulters as well.
- If we observe the highest count of defaulters is for 13%-17% bucket
- But if we compare the ratio of fully paid vs defaulters for various interest buckets. **We can observe that the ratio of defaulter % is increasing as the interest is increasing**
- **If the interest rate is high, the probability of a person defaulting is also high**

Observation on loan term

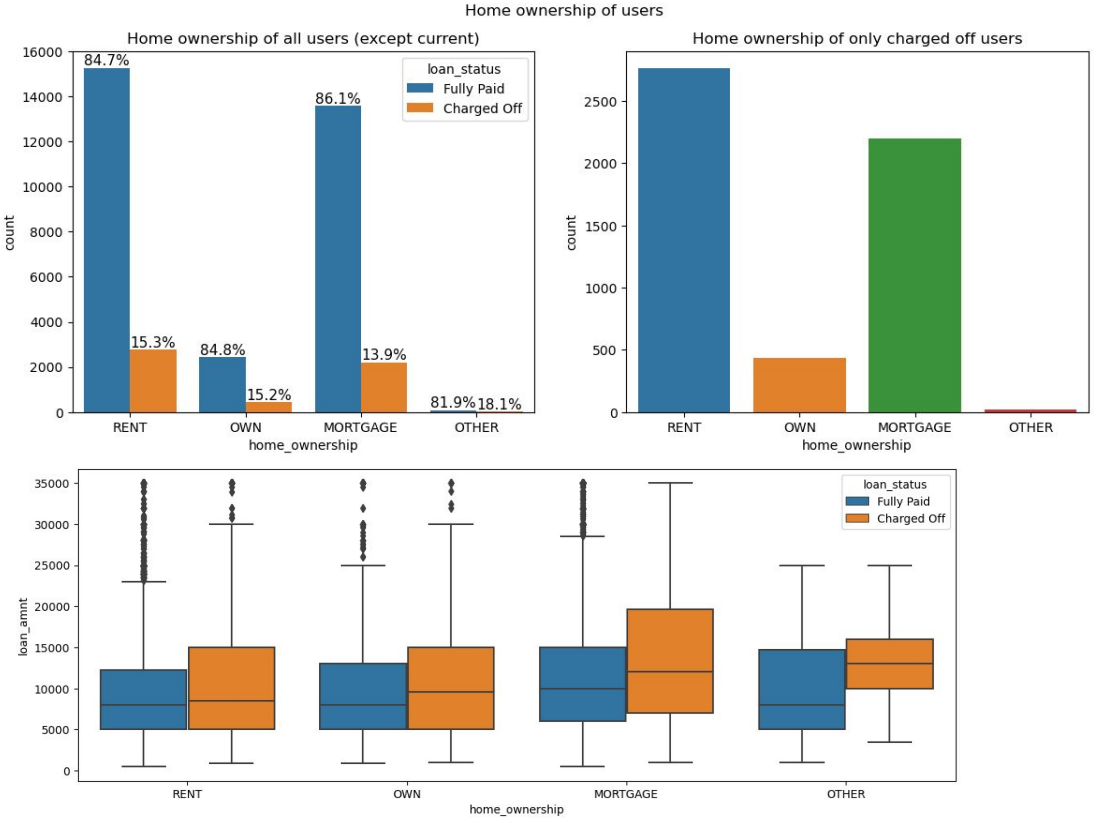






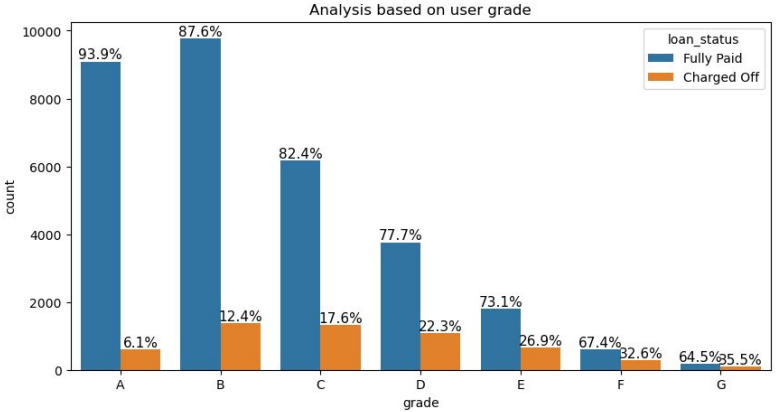
- The count of short term loan (36 months) is high and so the defaulter if we compare the count of defaulters to term then it is high for 36 months because approved loan is high
- But if we compare the ratio of fully paid vs charged off then we can easily conclude that the **defaulter % is very high for 60 months loan term**
- Users took high loan at high interest rate for longer term and then failed to pay back

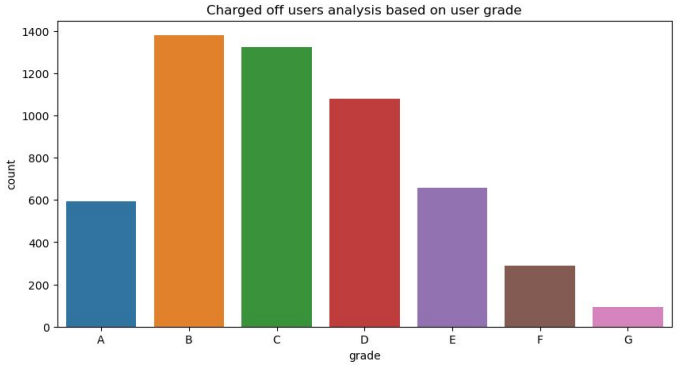
Observation on home ownership

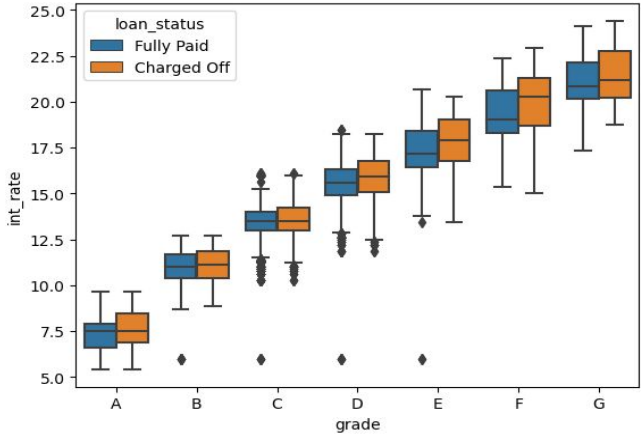


- The defaulter percentage is high for Rent and Mortgage
- The data of Other is very small to draw any conclusion

Observation on user grade

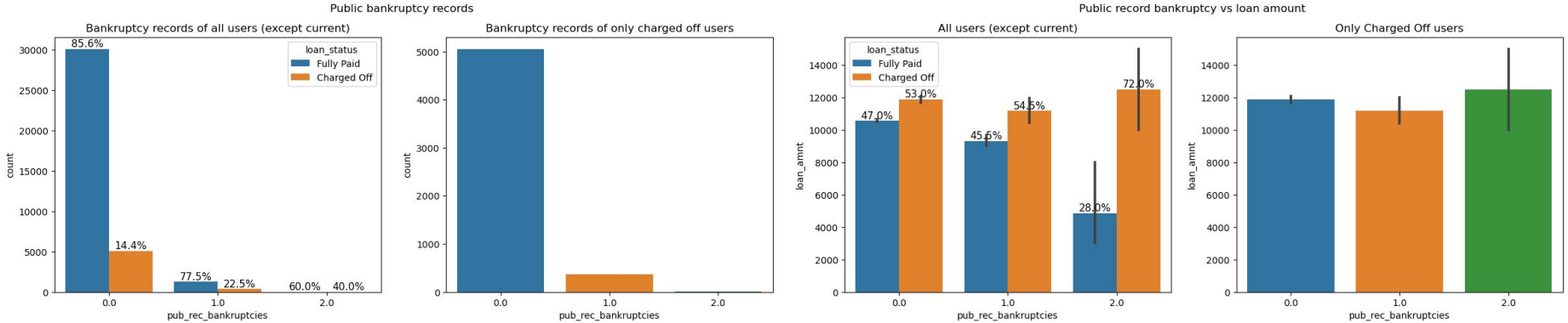






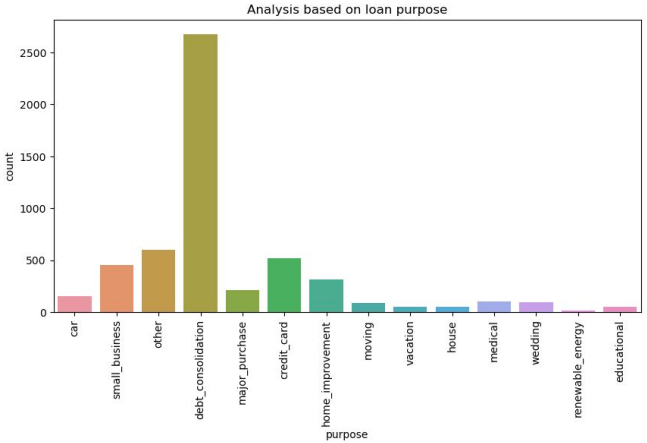
- As the grade decreases the ratio % of defaulters increases
- The count of defaulters is highest for B as the loans granted is also highest for B
- Granting high loan to a low grade user is risky

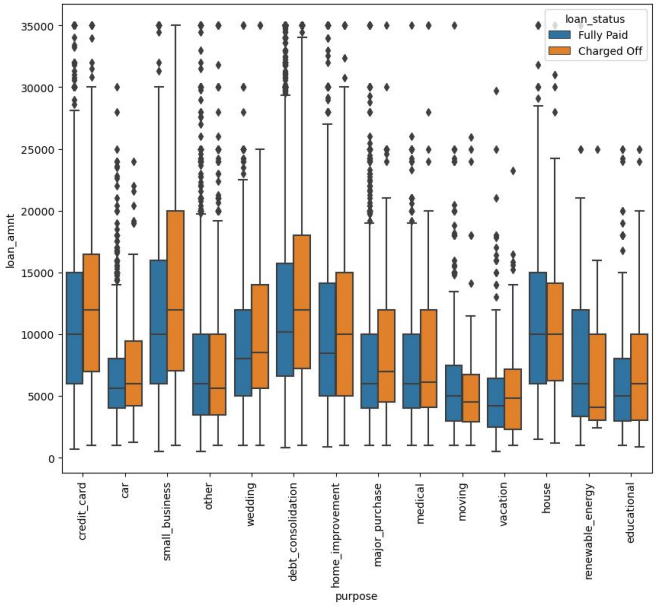
Observation on public record bankruptcy



- As the bankruptcy record increases the chances of a loan being defaulted also increases (The data for bankruptcy 2 is very low to conclude but the percentage distribution is very high)
- The count of defaulters is highest for 0 bankruptcy records because the loan granted is also highest to users with 0 bankruptcy records
- More loan is requested from users with bankruptcy records and hence resulted in defaulters

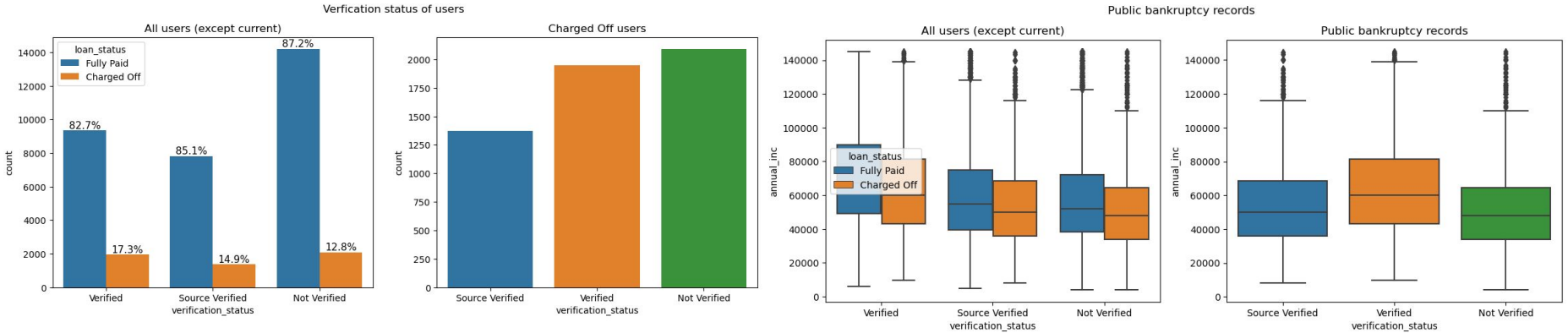
Observation on purpose of loan





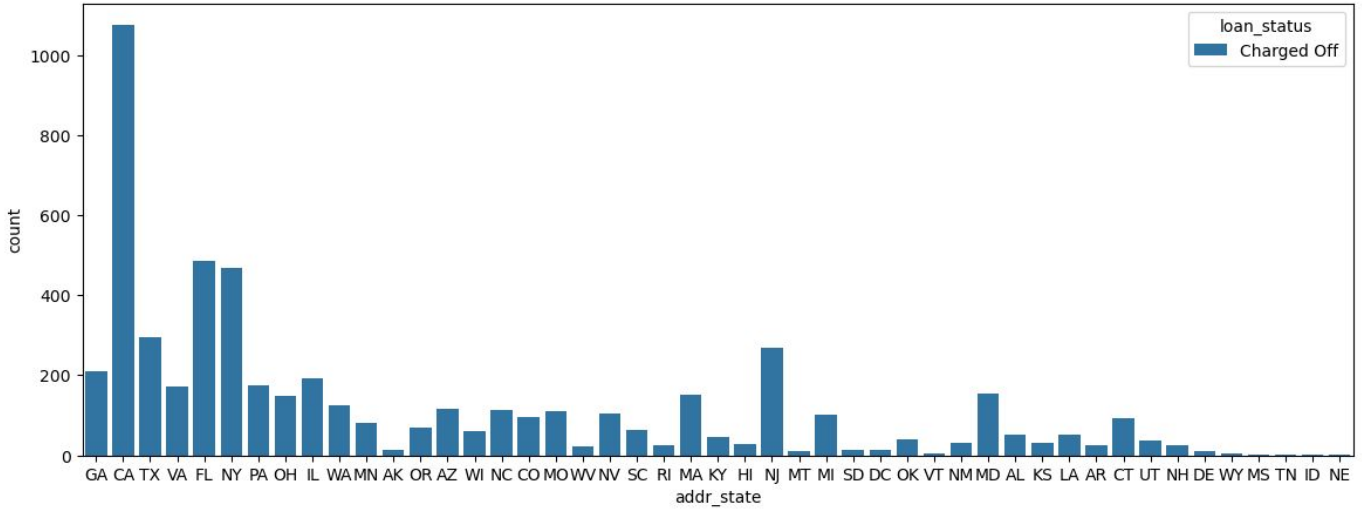
- Higher loan is provided to small businesses followed by debt consolidation, credit card
- Loan amount for defaulters is high for small businesses followed by debt consolidation, credit card

Observation on verification status



- More loans were given to not verified
- The count of defaulters is high for not verified, followed by Verified and then source verified
- The highest ratio defaulter percentage is for verified users
- More loans were provided to not verified users based on annual income

Observation on state loan



- CA, FL and NY has the highest defaulter count

Recommendation

Based on our observations made from the dataset we have observed that the following parameters might be one of the factors to be considered for identifying loan defaults.

- Loans provided on high interest rate are more prone to defaulter.
- Loans provided on longer term (60 months) and high loan value with high interest rates are more prone to defaulter.
- Borrower with rent or mortgage home ownership are more prone to turning out to a defaulter. Lending club should avoid giving high loans to this category.
- Grade could be a good source to validate users. Lending club should be more clinical when issuing loan to a lower grade as lower the grade, higher the probability of turning out to be defaulter
- Public bankruptcy record is also a good metric to analysis. Users with bankruptcy records are more prone to defaulter
- Loans provided to small businesses, debt consolidation and credit card are more prone to defaulter. Lending club should investigate more for loan request for these purposes. They should also issue either less or deny such loan requests
- More loans were provided to not verified users and hence it has more count. However more defaulter % (proportion) is when the user’s verification status is verified.
- Loans provided to users from CA, NL, NY are more prone to defaulter