Lending Club Case Study

Aman
Swapnil GATHIBANDHE

Agenda -

1. Problem Statement
2. Possible actions
3. Life cycle [Process]
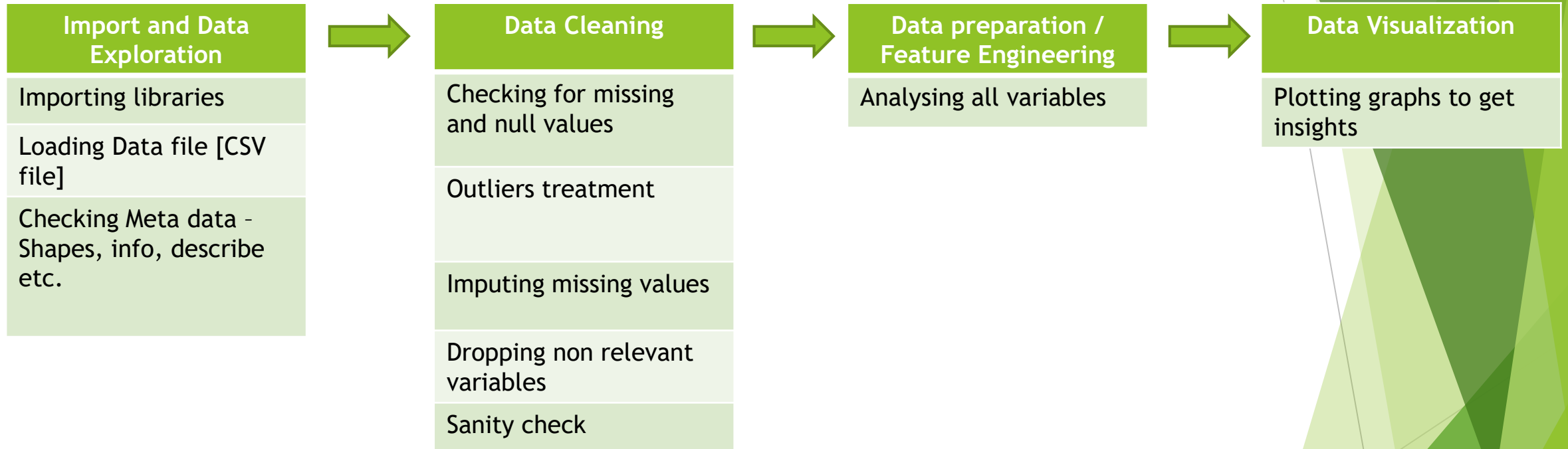4. Analysis of collected data
5. Conclusion/Recommendations

Problem Statement

▶ You work for a **consumer finance company** that specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

▶ If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company

▶ If the applicant **is not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company
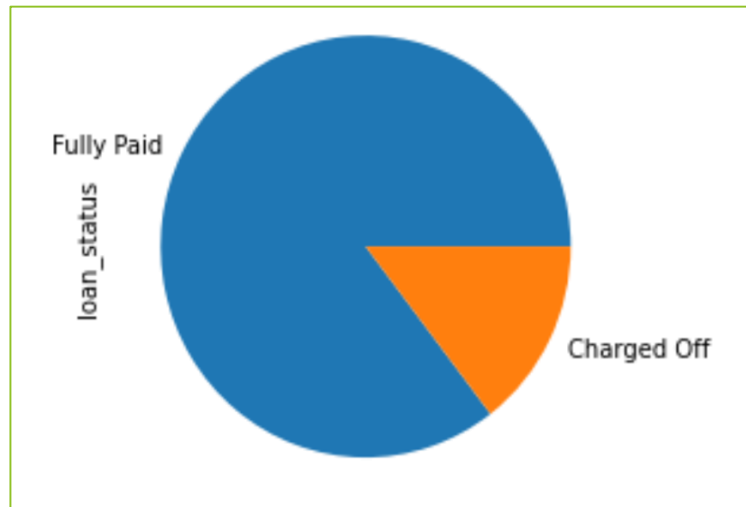
Possible action -

▶ When a person applies for a loan, there are two types of decisions that could be taken by the company:

▶ **Loan accepted**: If the company approves the loan, there are 3 possible scenarios described below:

▶ **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)

▶ **Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

▶ **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

▶ **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Life Cycle [Process] -

Import and Data Exploration

Importing libraries

Loading Data file [CSV file]

Checking Meta data – Shapes, info, describe etc.

Data Cleaning

Checking for missing and null values

Outliers treatment

Imputing missing values

Dropping non relevant variables

Sanity check

Data preparation / Feature Engineering

Analysing all variables

Data Visualization

Plotting graphs to get insights

Analysis of collected data - Univariate analysis

▶ Loan Status - More than 80% of the loans are fully paid and only around 20% are charged off.
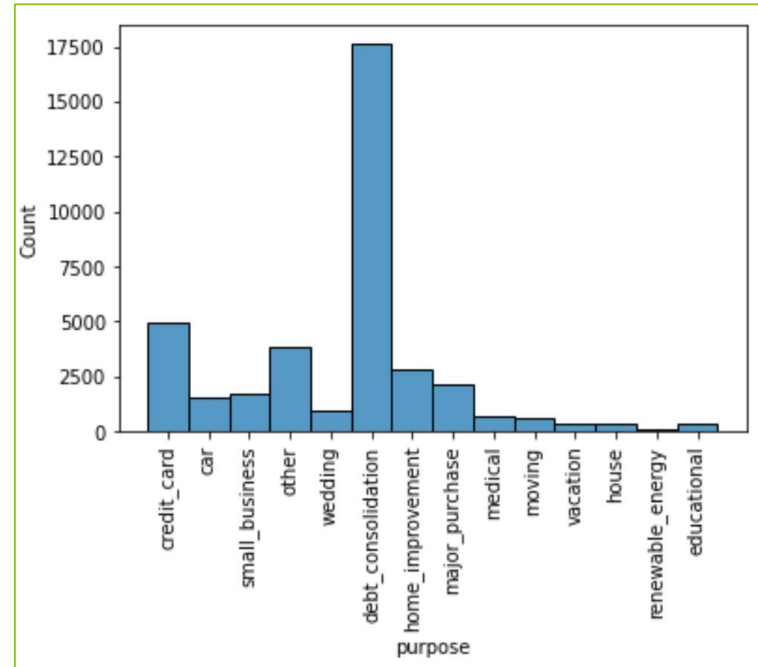
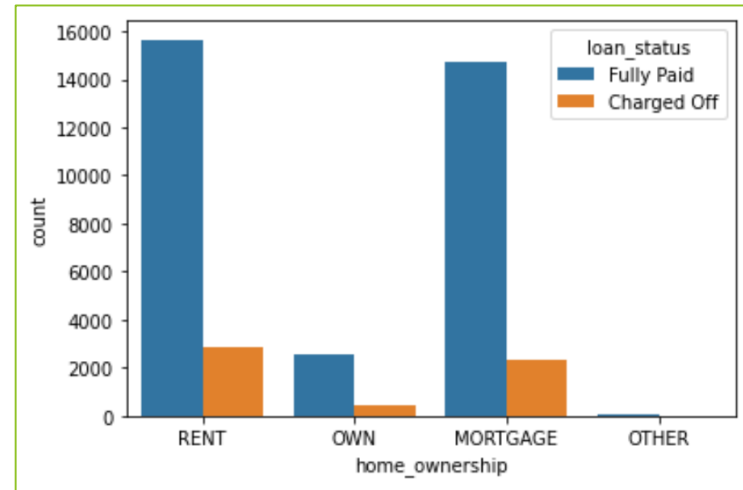▶ Most of the loans are taken for debt consolidation and least have the purpose of renewable energy.
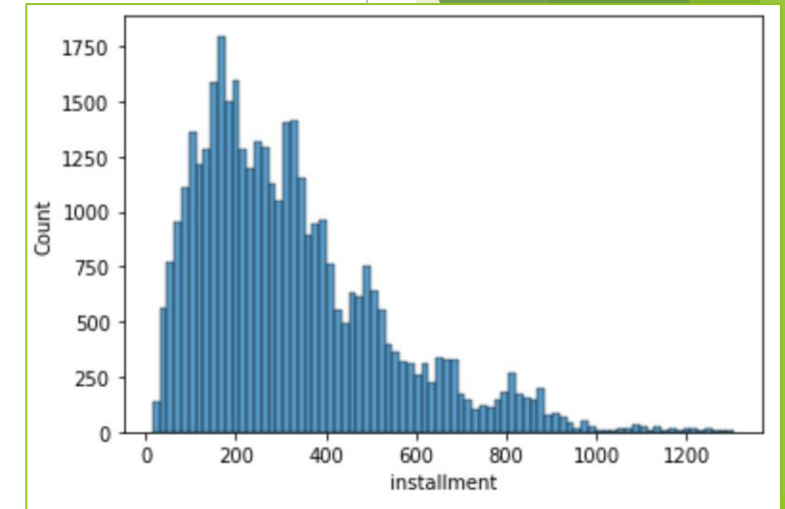
▶ The highest count of installments is found at 180.

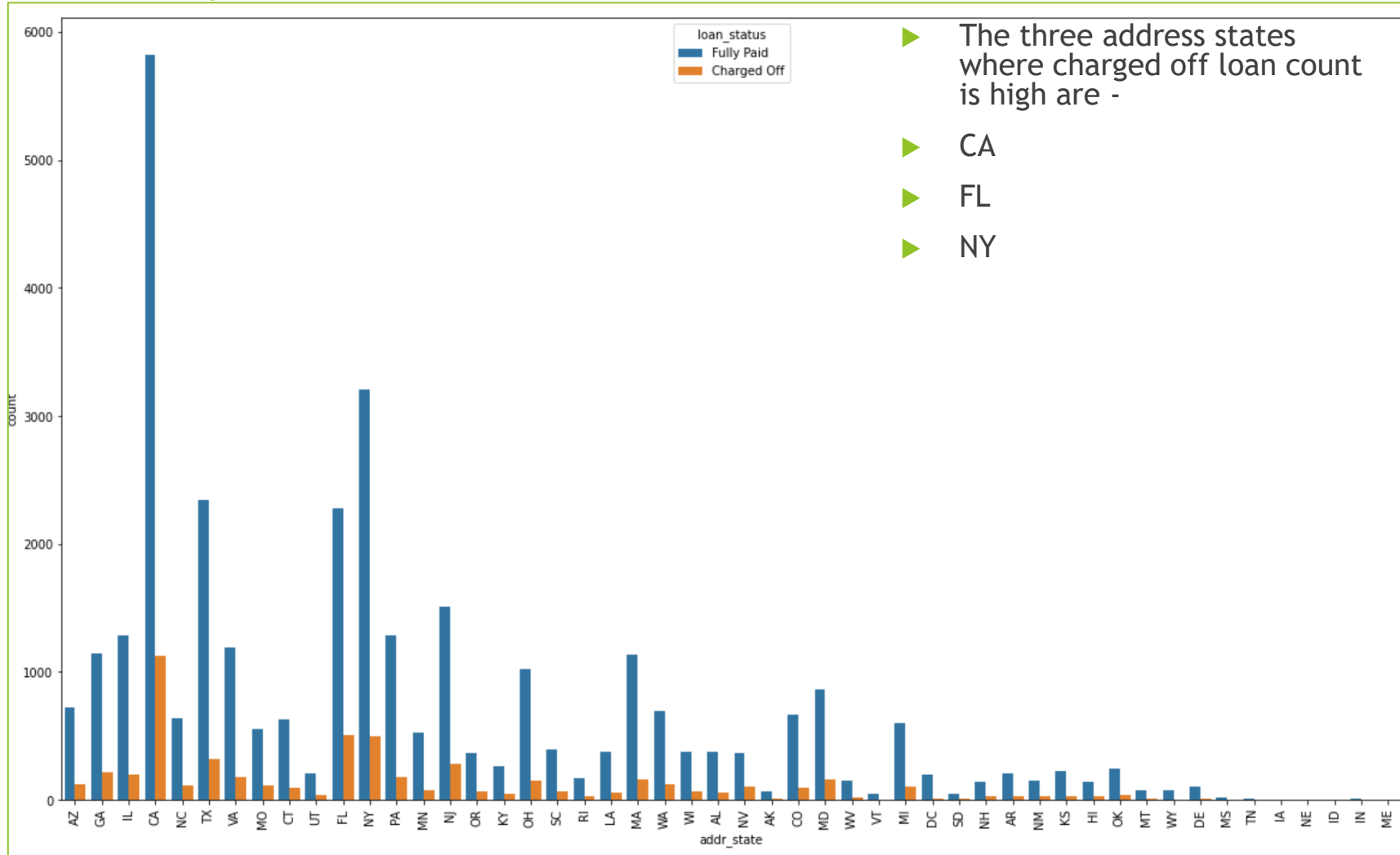# Analysis of collected data – Segmented Univariate analysis
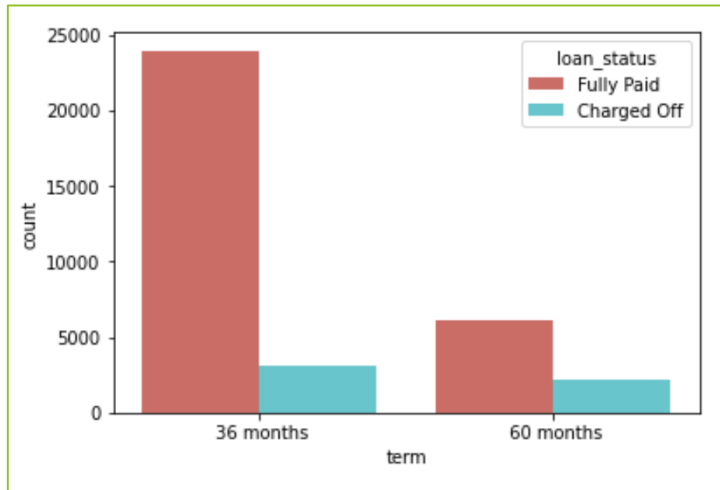
► Loans with payment term of 60 months are more in % than loans having payment term of 36 months.
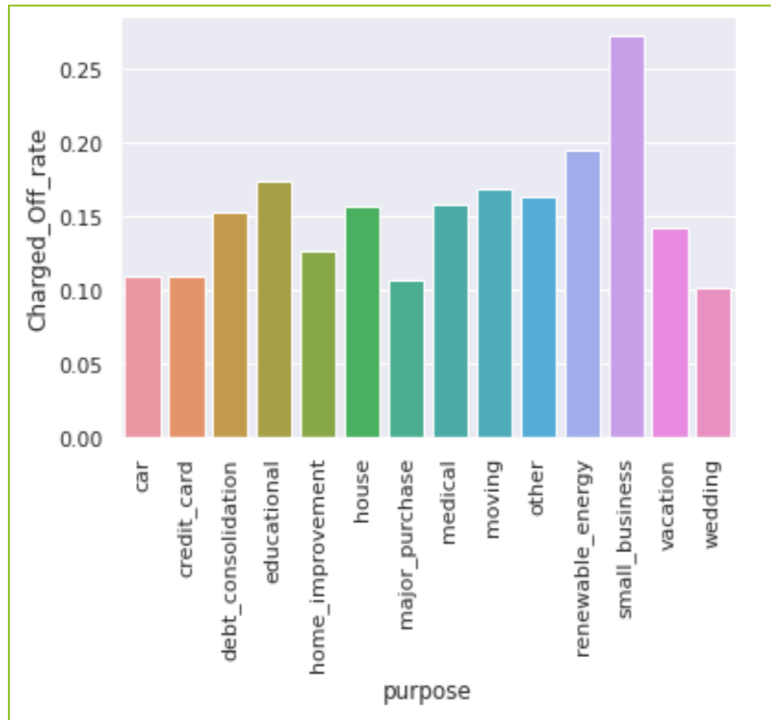
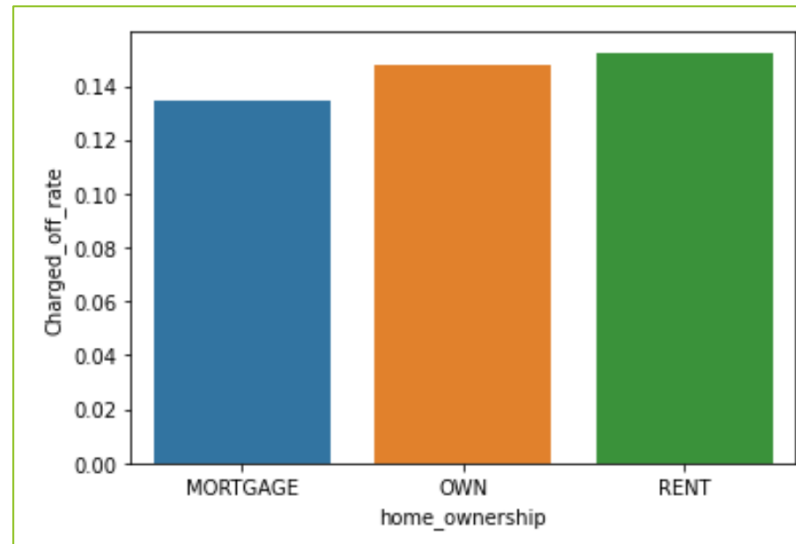► Charged Off rate is for people with either rent or have home ownership on mortgage.

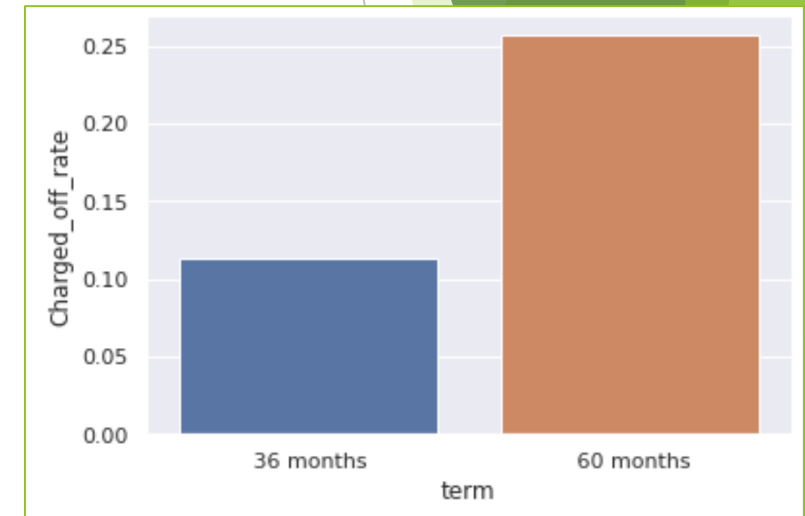# Analysis of collected data – Segmented Univariate analysis

- The three address states where charged off loan count is high are -
- CA
- FL
- NY

Analysis of collected data - Bivariate analysis

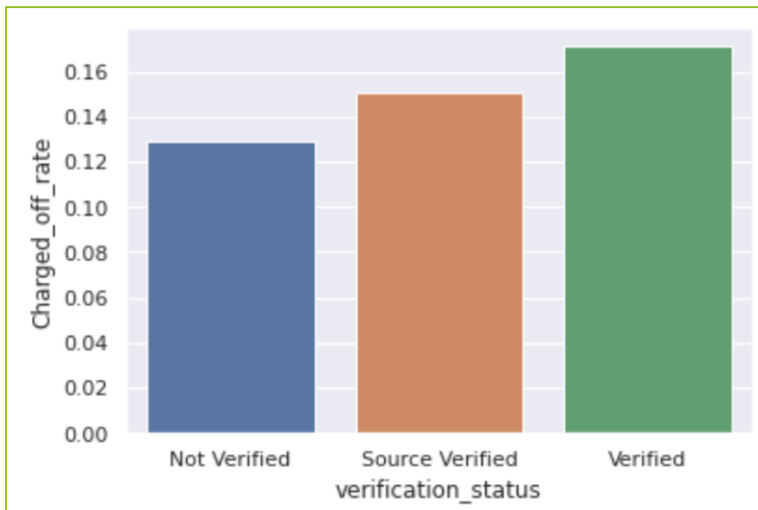► The purpose of small business has a high rate of charge off.

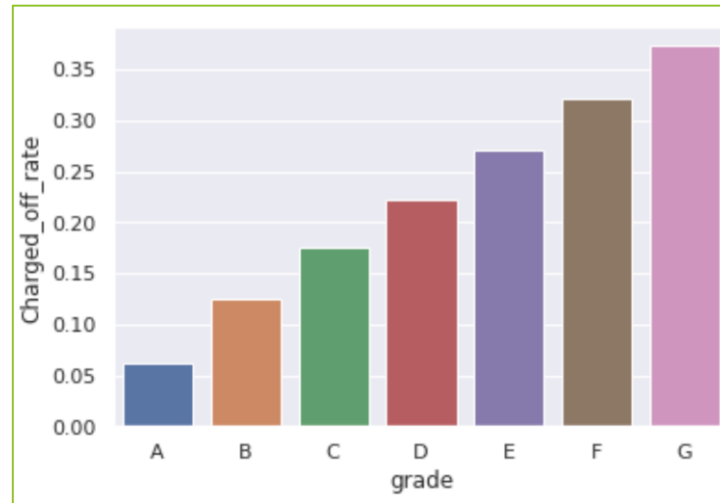► The loan applicants having home ownership as rent are more likely to charge off.

► Loans with long payment term are more likely to charge off.
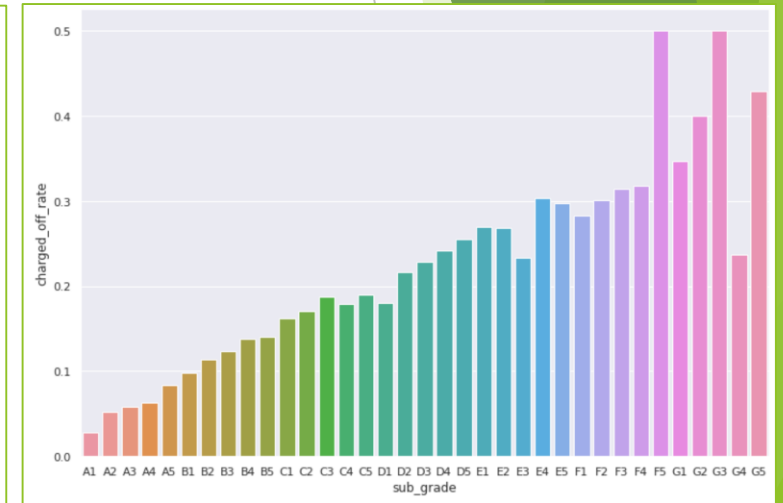
Analysis of collected data – Bivariate analysis

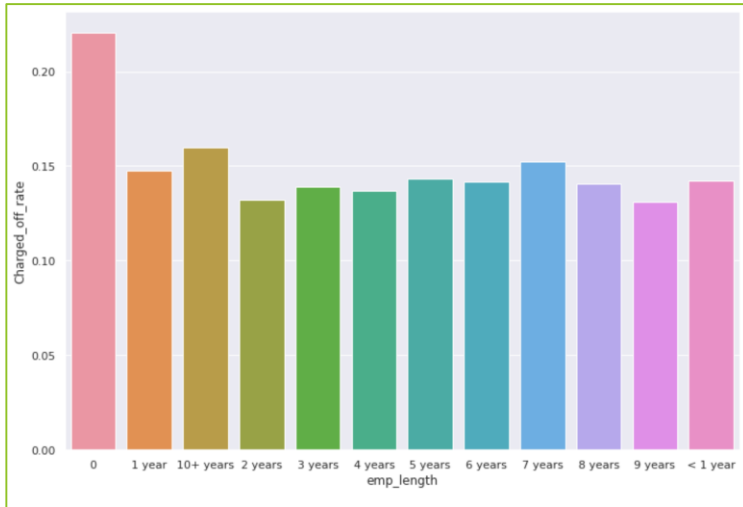▶ Loans having verification status as verified are more likely to charged off.

▶ The loans having grade G are more likely to be charged off.
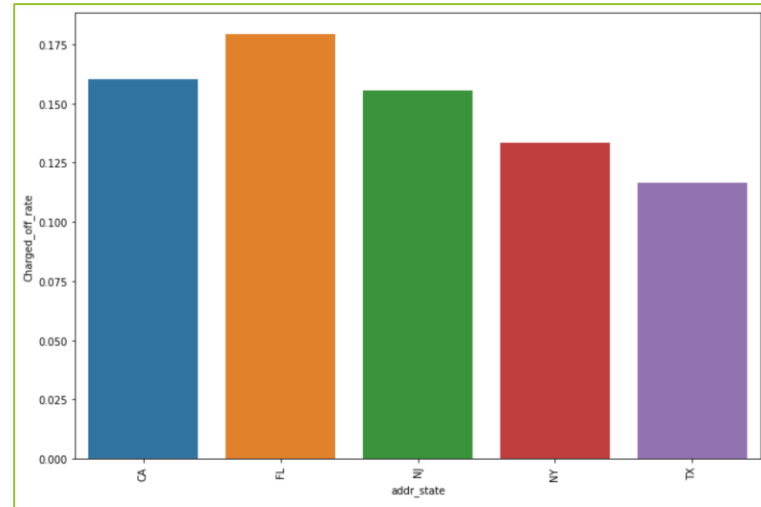

▶ As the subgrade increases, the probability of loan to be charged off increases.

▶ Grade F5 to G5 have most probability to be charged off.

Analysis of collected data – Bivariate analysis

▶ The short the term of employment has most likely to charged off.

▶ The top 5 applicant states in which the FL has high charged off loans.

Conclusion / Recommendation

- Parameters that can help in predicting whether the loan will be fully paid or charged off are –

- Term of repayments

- Purpose

- Homeownership

- Grade and subgrade

- Interest rate

- Annual Income

- Term of employment

Recommendation

- Analyse with respect to the predicting variables to reduce the count of charged off loans.

Thank you