Lending Club Case Study

Submitted by:
1. Debaratna Nath
2. Pragya Shuchi

Problem Statement

- We “at Lending Club" are analyzing our past customer data to develop efficient risk mitigation strategies for our lending businesses.
- As a part of this process , we are leveraging data analytics to generate insights regarding the current customers’ base, customers’ defaulting behaviors & the attributes which might form the baseline for segregating good vs bad customers
- As we traverse through this deck , we are trying to generate meaningful recommendations and findings from the exploratory data analysis performed on the available data.

Objectives

- Examine the past customer data to conclude on the possible driving factors of defaulters
- Results can be leveraged in the application screening process for the incoming customers to filter out the potential defaulters in the initial phase itself.
  - This saves costs of other services like sending agents for house-income verification, bank statement parsers , pulling credit reports like CIBIL.
  - Reduces the dependency on underwriters
  - Boosts the TAT and real time processing time significantly
- Decision making & strategizing through analytics can be a key instrument in minimizing default rate & computation cost

Approach

Understanding the data

- Data Attributes & Size
- Data Variance & Quality

Data Pre-processing

- Outlier & Missing Values treatment
- Data Type Formatting
- Feature Reduction based on Business driven & Data driven approach

Univariate Analysis

- Box Plots & Visualisations to understand customer attributes

Segmented Univariate & Bivariate Analysis

- Analysing attributes wrt. Loan status

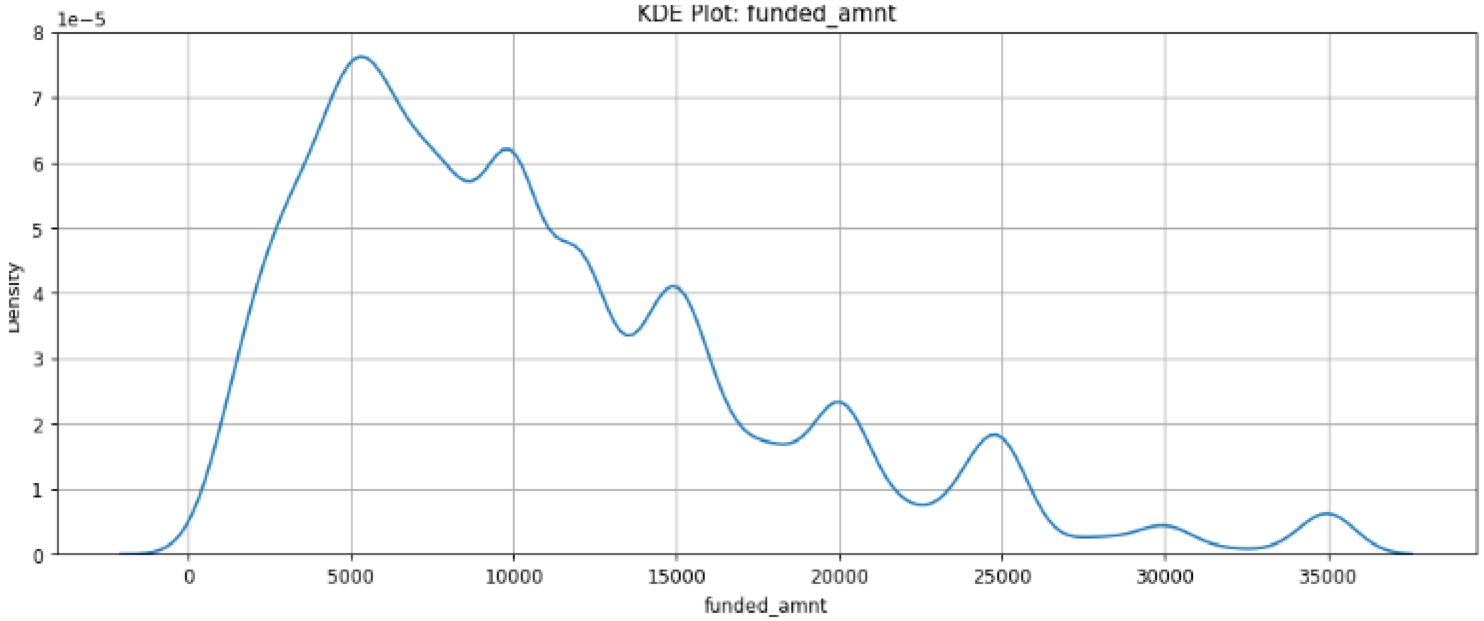
Multivariate Analysis

- Analysing the relationship of two or more variables in alignment with loan status

Recommendations & Conclusion

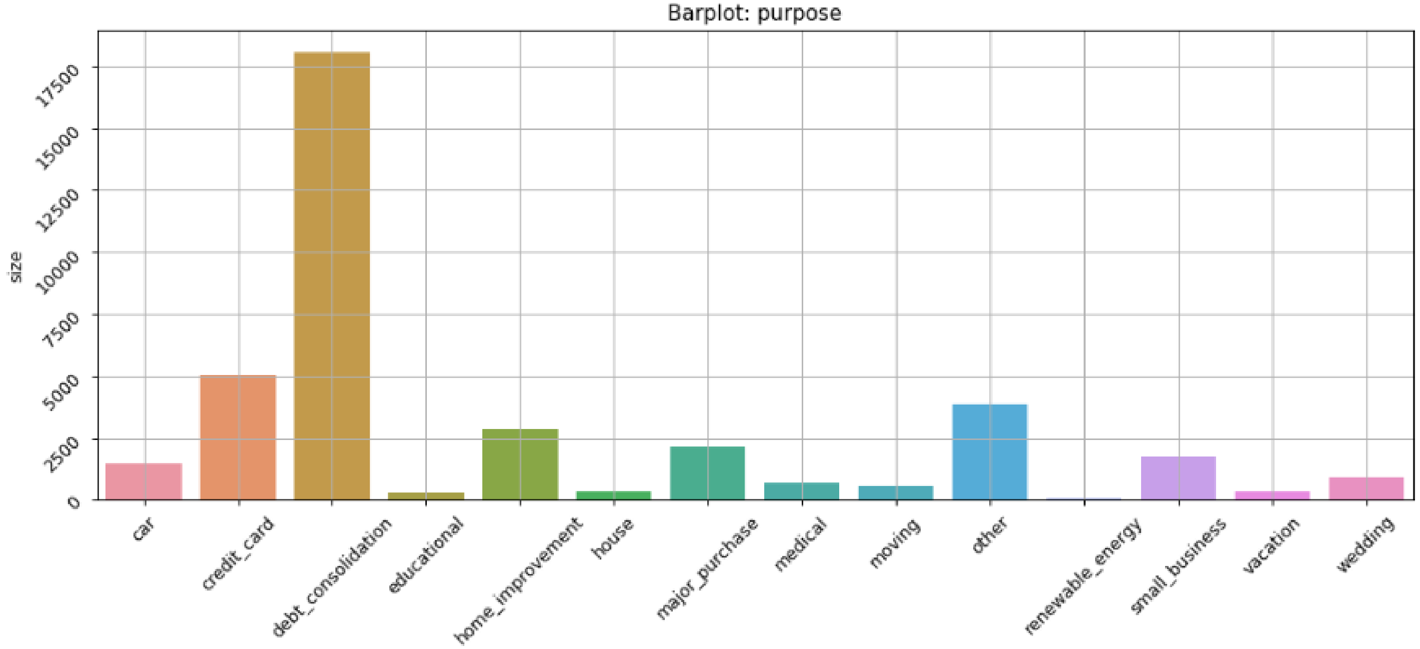
Analysis

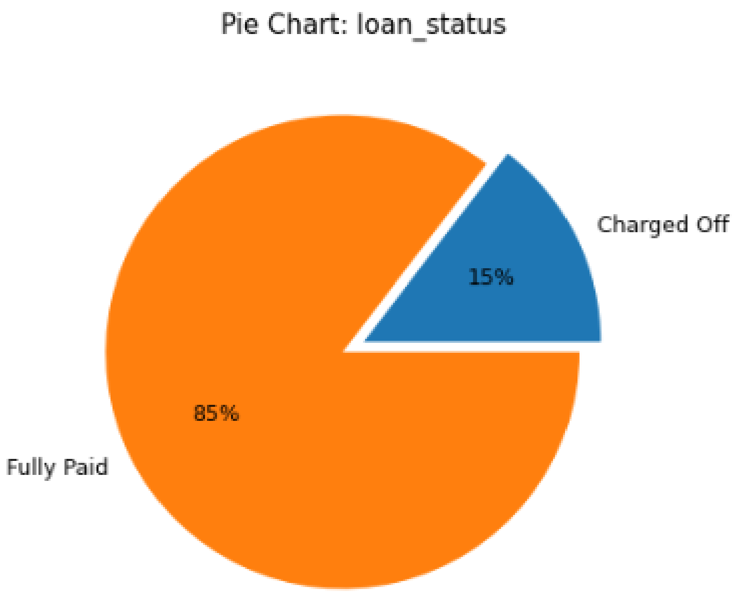
- Understanding the Data
    - ~40 K records at customer level with 111 attributes
    - Data Quality Checks:
        - More than 50 columns with 100 % null values
        - Limited Cardinality in data
    - Data Categories
        - Customer Attribute/Demographic Data
        - Loan Attribute Data
        - Customer Credit Behavior
- Univariate Analysis
    - Analyzing the distribution of few of the attributes provided important inferences.



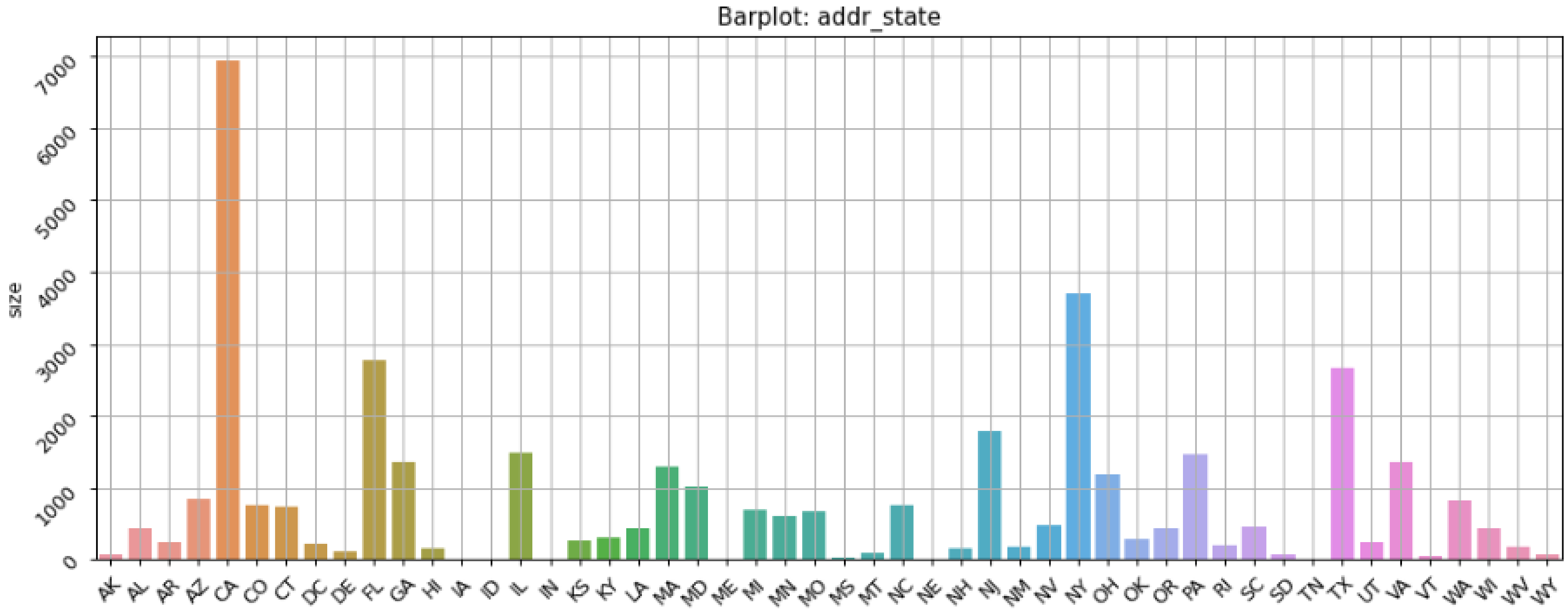
Majority of customers apply for loan amount ranging between 5k – 15 k.

The reason for applying in 47% of the cases is to settle their debt consolidation followed by credit card payment.





85 % of past loans have been fully paid. It indicates the need of an effective application screening process for the remaining 15% of loans that went bad.



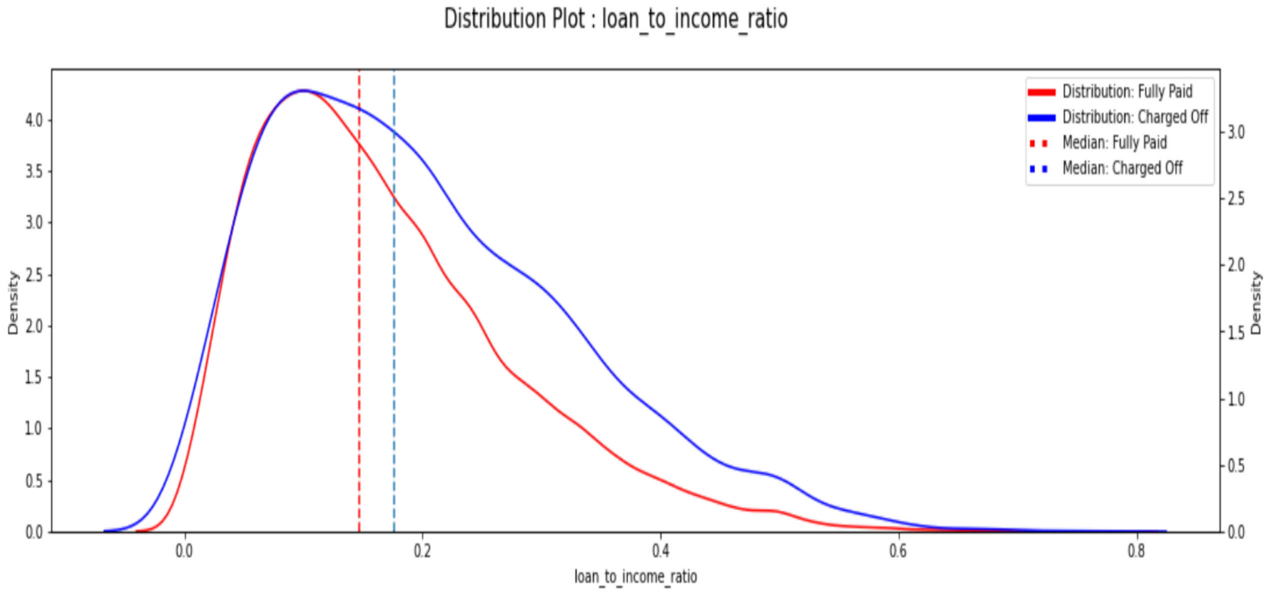
CA followed by NY dominates the customer market demographically.

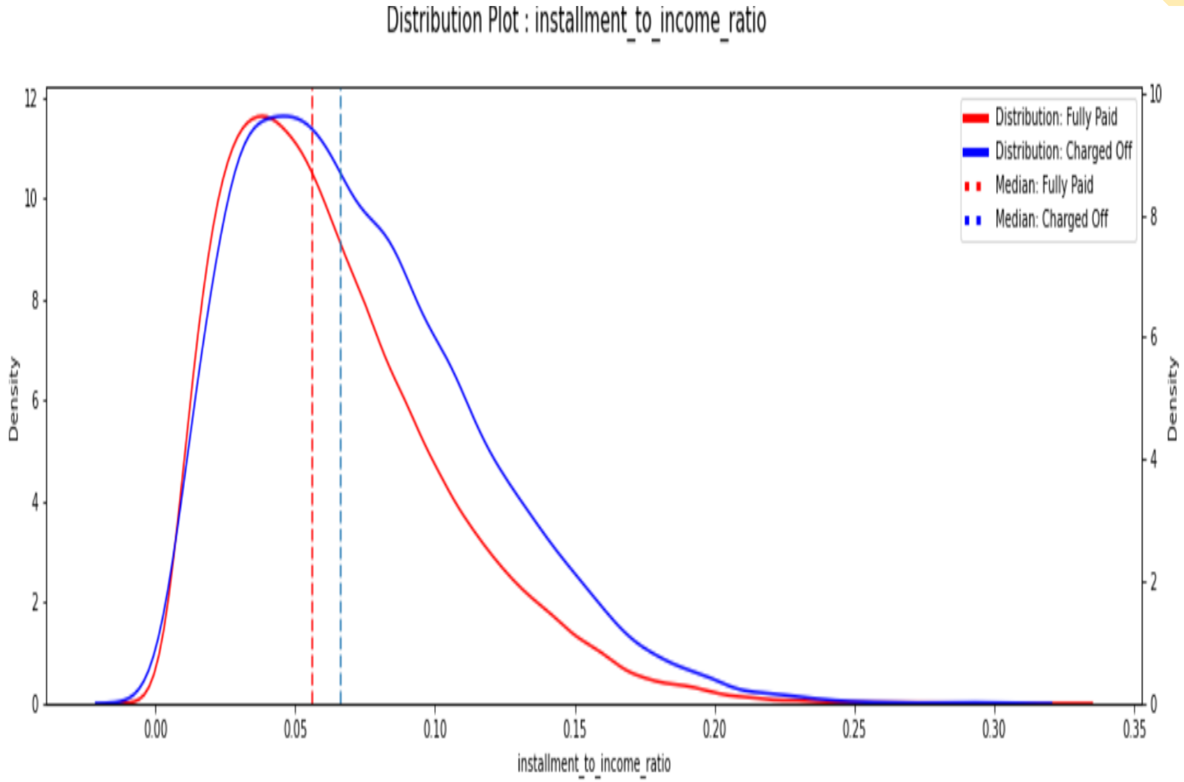
Bivariate Analysis

Derived two variables:

•**loan-to-income-ratio** : The ratio between the amount of loan one received to their annual income.

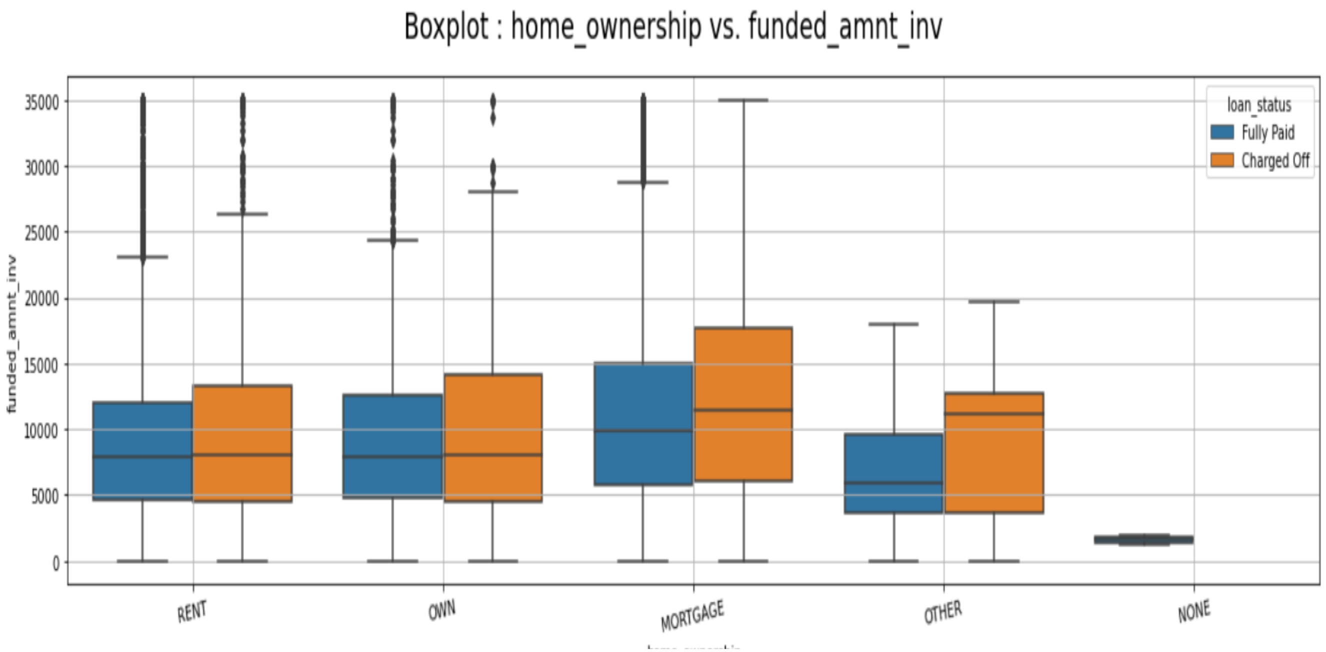
•**Installment-to-income-ratio** : The ratio between the installment and the **monthly** income.



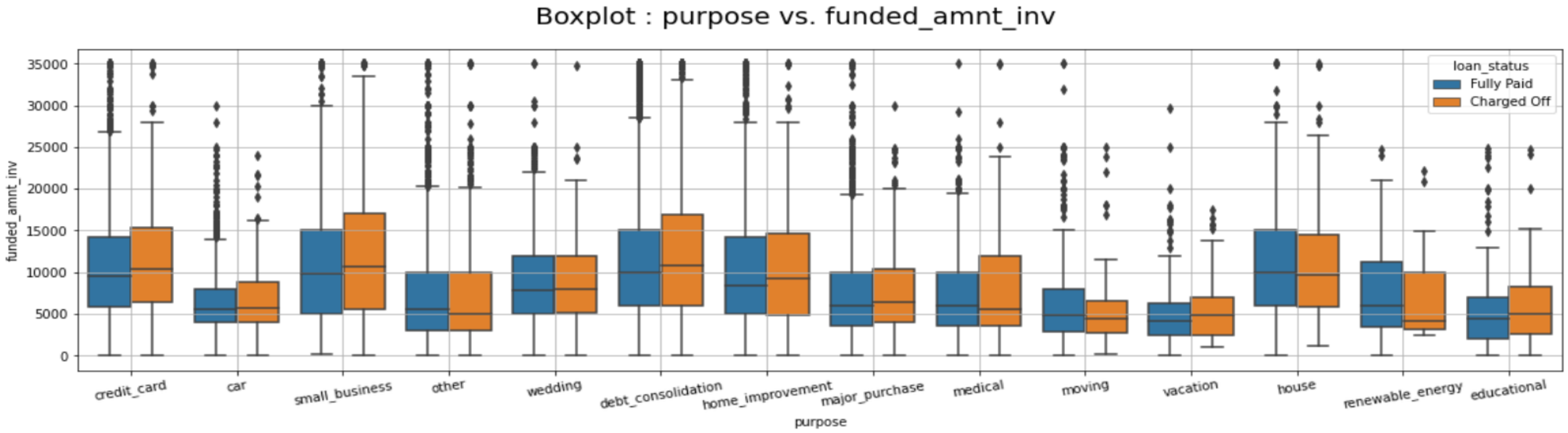


•Applicants who defaulted on their loans had a higher median value of loan_to_income_ratio
•Applicants who defaulted on their loans had a higher median value of installment_to_income_ratio
•Applicants who defaulted on their loans had a significantly higher value of revol_util (which is a measure of the credit utilisation limit of the applicant)

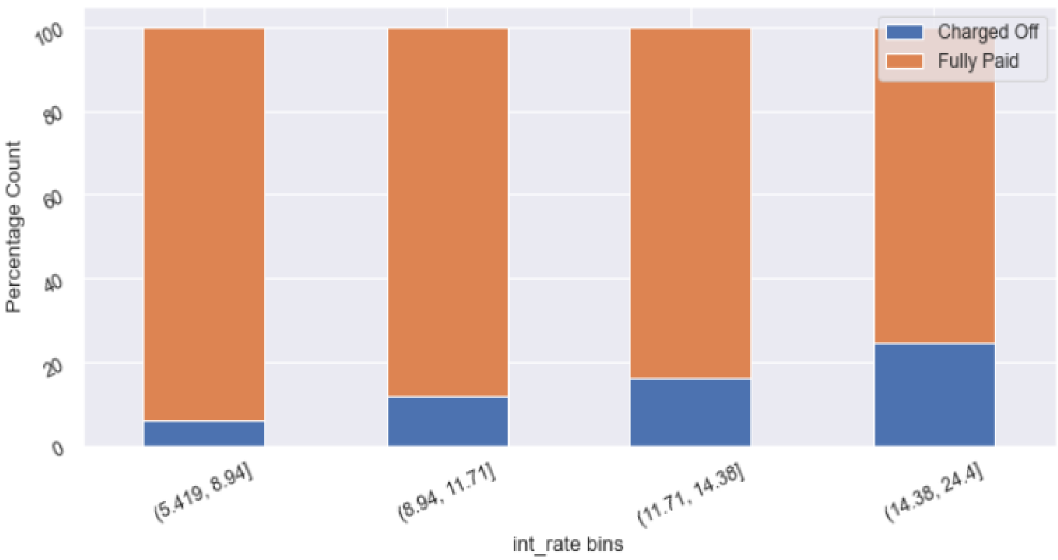
Down: The distribution of 'defaulting' or 'Charge-Off' is more shifted towards higher values in plot for people with 'Mortgaged Homes'. This means that people with an existing Mortgage who applied for higher amount of loan were at a greater risk of defaulting.

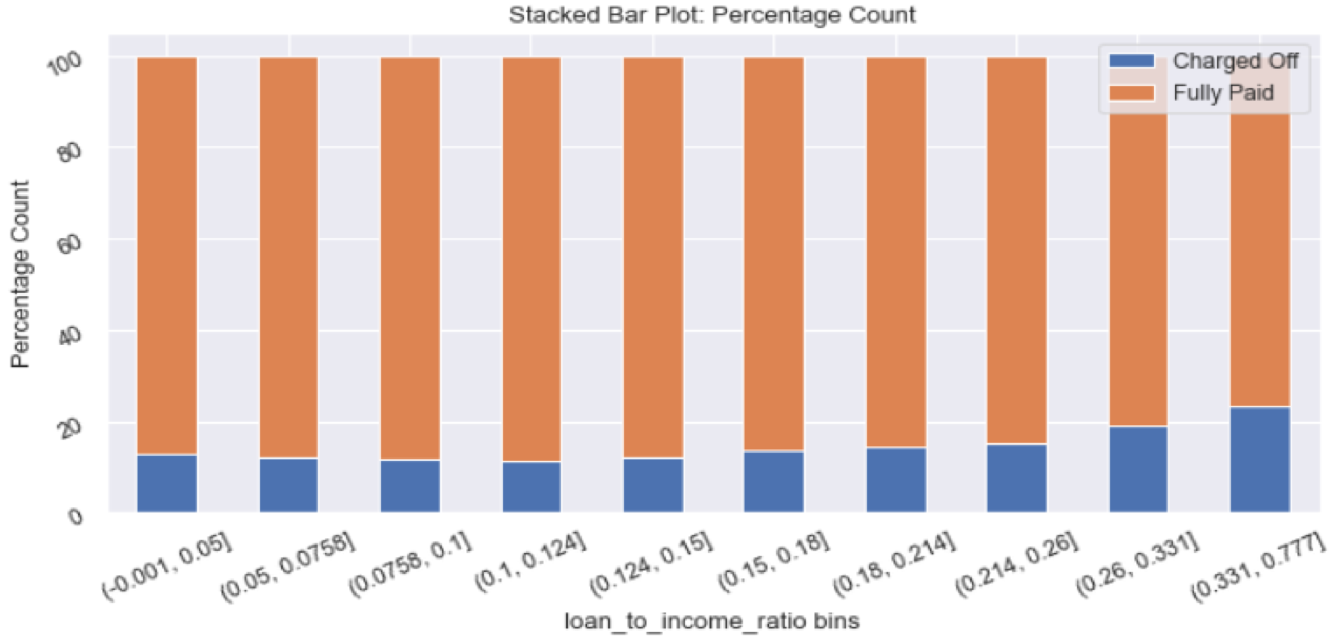


Down: People who applied for high amount of loan, for the purpose of business, debt-consolidation or credit-card were at a higher risk of defaulting.



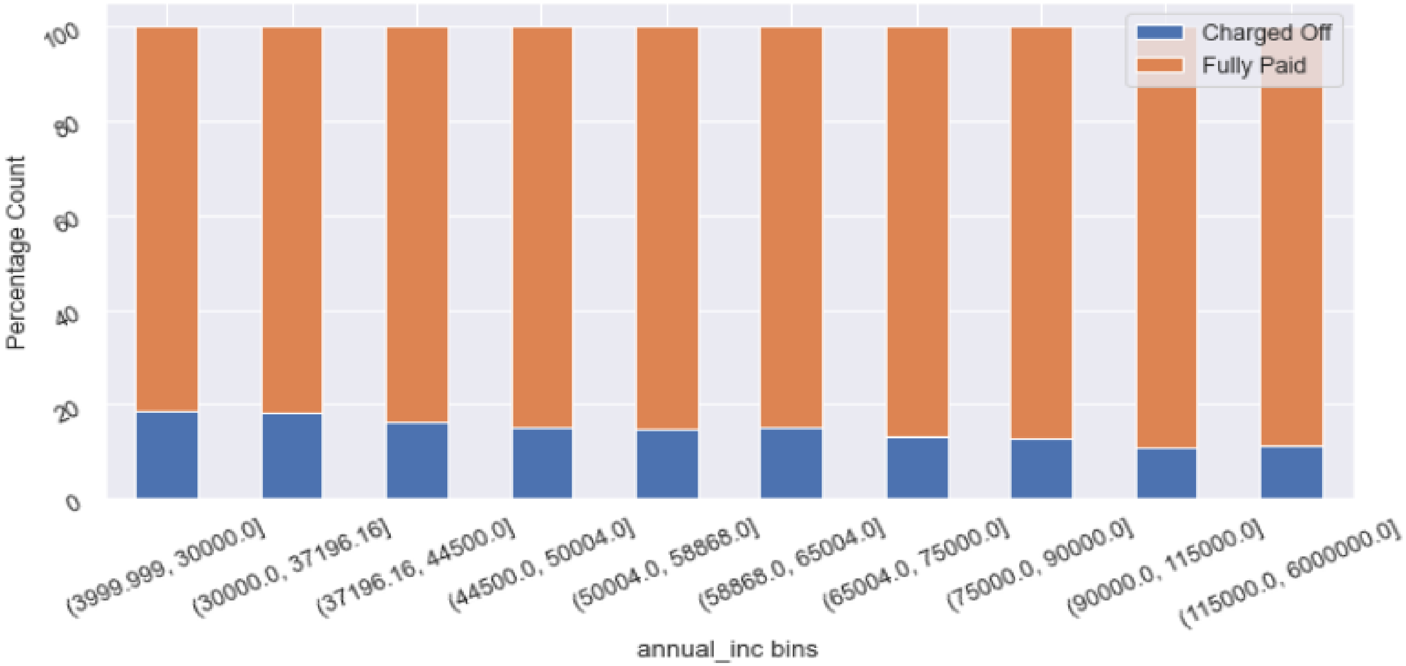
BINNING NUMERICAL VARIABLES & ANALYSING WRT LOAN STATUS



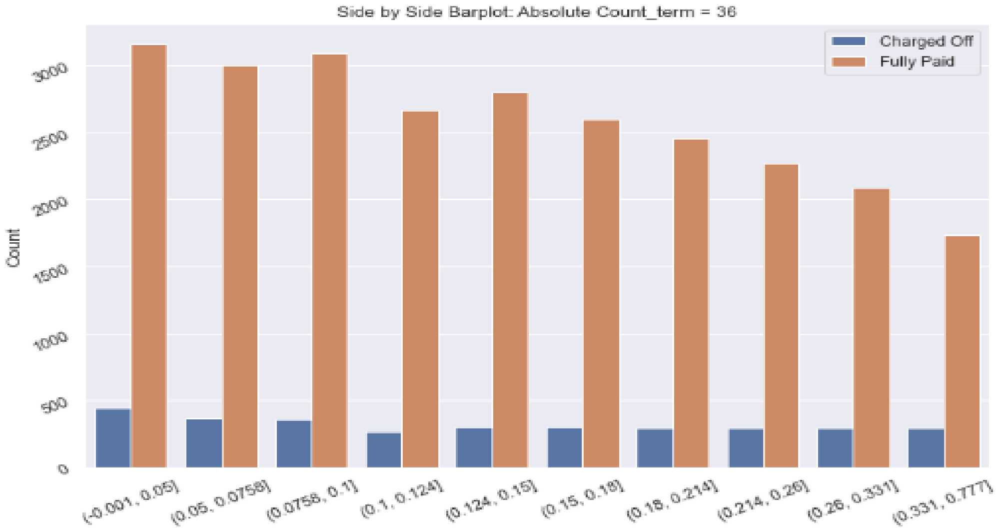


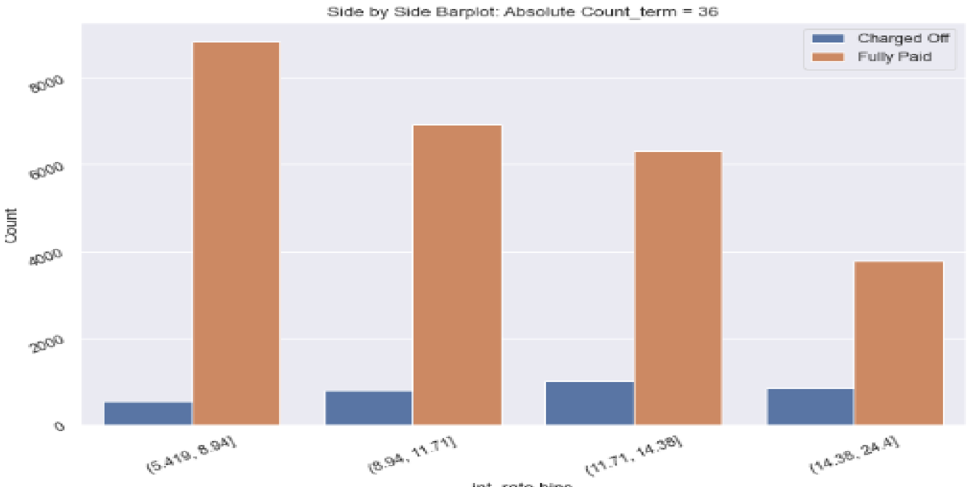
Percentage of defaulting in historical data is more for:

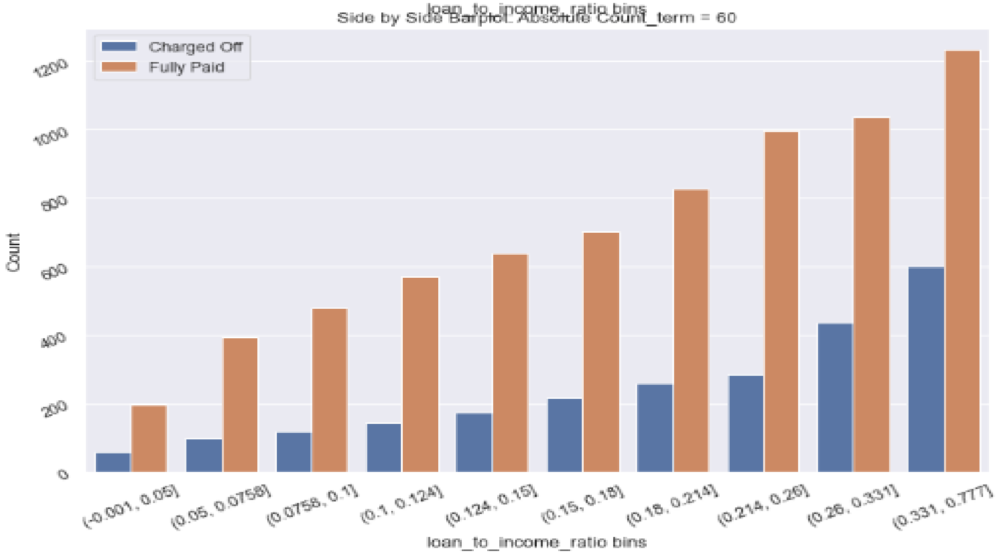
- Applicants with lower annual income. (4k-30k range)
- Applicants with higher loan to income ratio (> 25% of annual income)
- Applicants with higher installment to income ratio (>8% of monthly income)
- Applicants with a higher credit utilisation rate (>77% of total credit)
- When the loan interest is more (> 14% p.a)

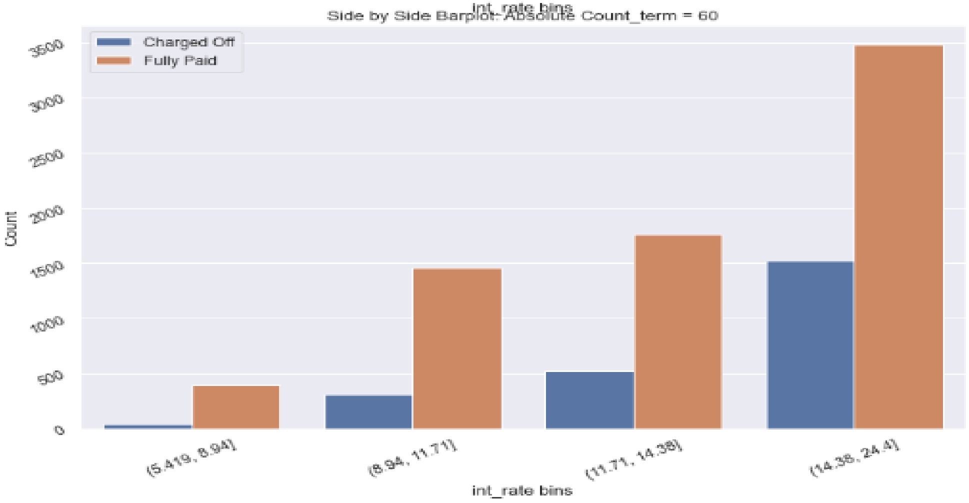


Multivariate Analysis







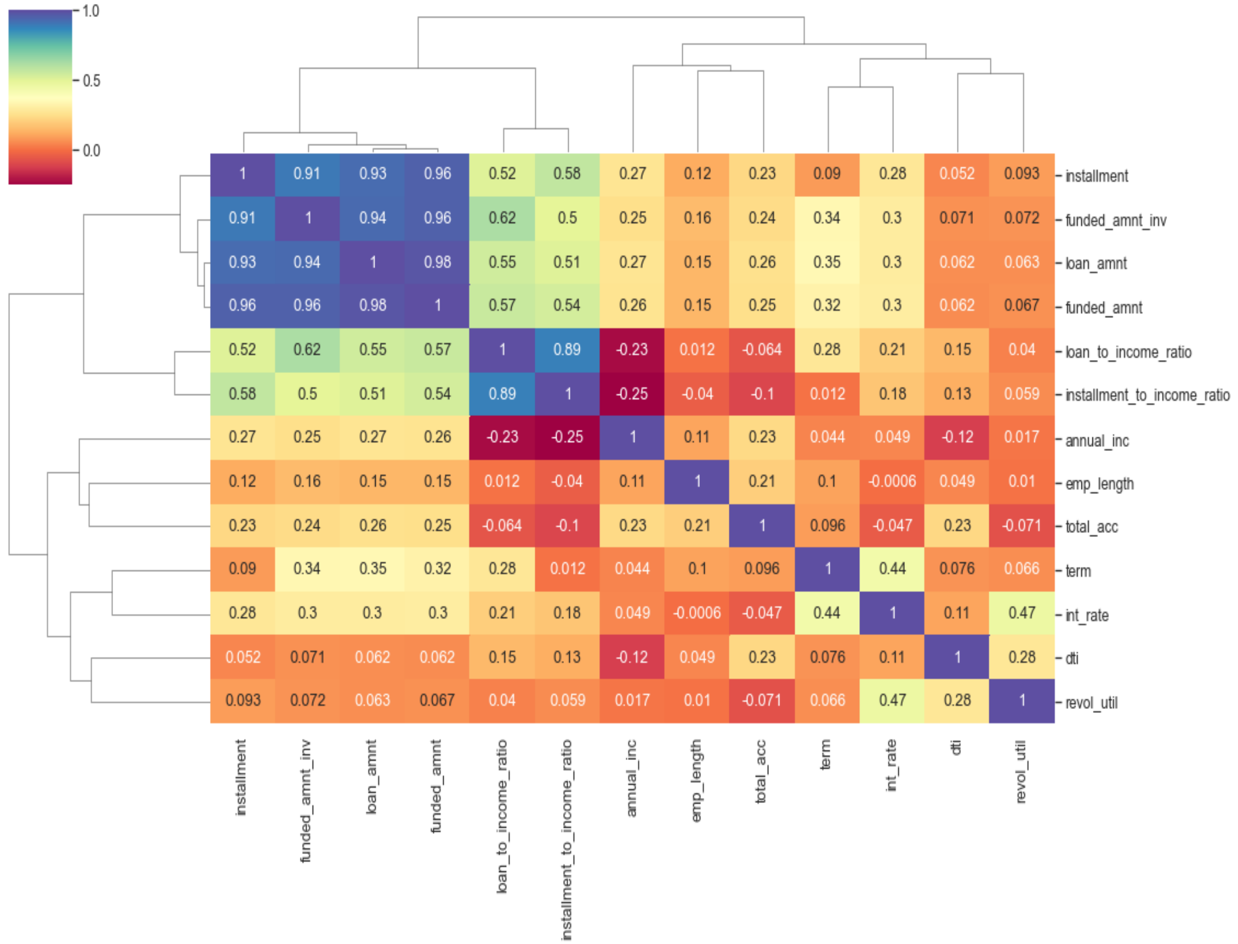


Among all the probable driving factors, viz., loan_to_income_ratio, installment_to_income_ratio, interest_rate, dti, revol_util, the percentage of defaulting/charge off is more when the loans are lent out at a higher term of 60 months

Correlation Matrix

Correlation clustermap suggests some clustered relationships like

- (loan_amnt, funded_amnt, funded_amnt_inv) : high correlation as all these three values are essentially carrying similar information.
- (loan_to_income_ratio, installment_to_income_ratio) : as both are derived from the similar two features.
- moderately high correlation between int_rate and revol_util (0.47): People with high credit utilisation rate are given loans at a higher interest
- term and int_rate (0.44): If we opt to pay off the loan at a higher term, then generally the interest rate offered is also more.



Analysis Report at a Glance

- The Probability of Defaulting increases for applicants with a higher loan to income ratio (> 25% of their yearly income).
- The Probability of Defaulting increases for applicants with a higher installment to income ratio (>9% of their monthly income)
- The Probability of Defaulting Increases for applicants with a Credit Utilisation Rate of greater than 0.77 (or 77%)
- A loan lent out at a higher term of 60 months has more probability of being defaulted
    - For applicants with more than 25% loan_to_income_ratio, the percentage of defaulting is around 30%
    - For applicants with installment amounts greater than 8% of their monthly income, the percentage of defaulting is 30-40%.
    - When the interest rate is more than 14%, the percentage of defaulting is around 30%
- Applicants who live on Rent or Mortgaged property have higher chances of defaulting when the loan amount is high.
- Applicants with purpose of loan as Small Business, Credit-Card loan or Debt Consolidation have higher chances of defaulting when the loan amount is high

Recommendations

- There are few checks which should be a part of application screening(Phase 1) :
  - Filter out applicants with Income level below a certain threshold (e.g. 15 k)
  - Filter out applicants from areas where collections in case of default might be difficult
  - Filter out applicants whose job profiles look non-promising and remote and highly transferable . Collections will be a concern for such folks.
  - Filter out the applications whose DTI crosses 70% (this threshold is as per industry standards, this might vary for lenders)
    - DTI (Debt to Income) can be derived using monthly income , old loans installments & the loan he or she is applying for.
  - Type of accommodation can be a part of this checklist as well
- Screening 2 (Post getting the credit report & bank statements of the customer)
  - Understand his credit behavior and put checks like – no. of times he has gone 60 days bad/late for any loan should not be more than 2.
  - Credit Score should fall between a certain range.
  - Calculate the monthly disposable income using his bank statement – Put strict thresholds on the average balance per month
  - Devise a loan amount strategy based on DTI , Monthly Income & Requested amount
  - Term of loan should be set based on his past credit behavior – e.g. no of times he went 30 days bad / no of total active accounts in Credit report
  - Use the bank statement data to figure out the loans applied in the recent month as these might not be reflecting in Credit report yet.