Lending Club - Case Study

Anand TV
Ankur Gupta

Requirement

- Lending Club, a consumer finance marketplace specializing in offering a variety 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
- The data contains information about past loan applicants and whether they 'defaulted' or not.
- The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Preprocessing

Data Cleaning

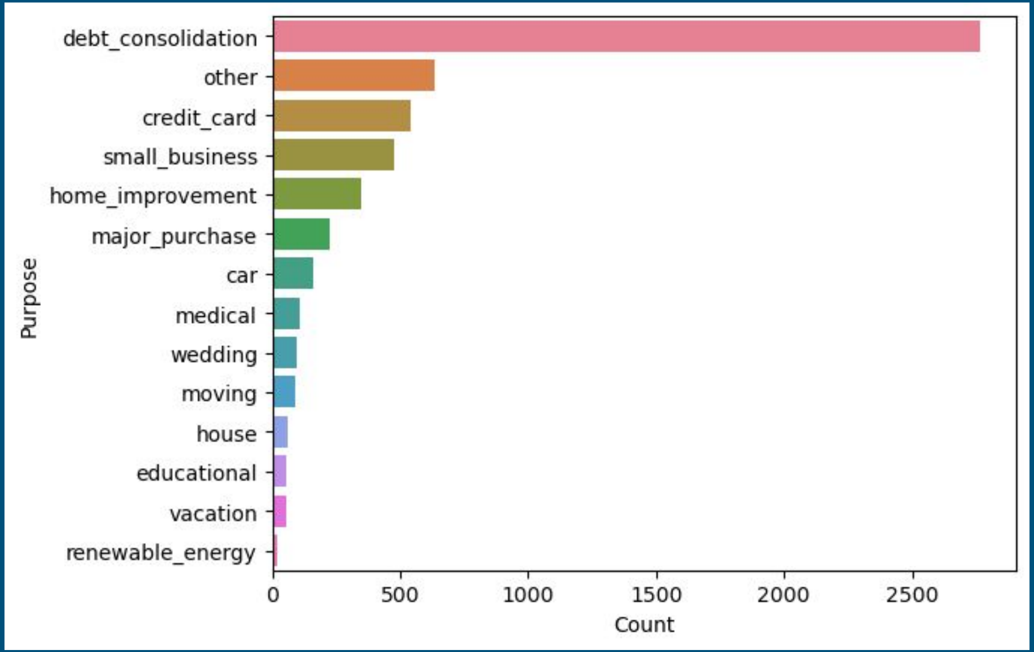
- Drop the columns if,
  - The missing values for that column is > 50%
  - Remove the column based on unique counts percentage.
- Drop the rows
  - Remove the duplicate rows.
  - Consider only *Closed* and *Charged off* loans for majority of the scenarios.
- Outlier removal
  - annual_inc

Data Preparation

- Convert the data type & Values
  - Int_rate
  - Term
  - issue_date
- Derive category bucket: interest rate, loan amount, Annual income and dti.
- Derive new fields: Year and Month
- Convert the values of Employee Length

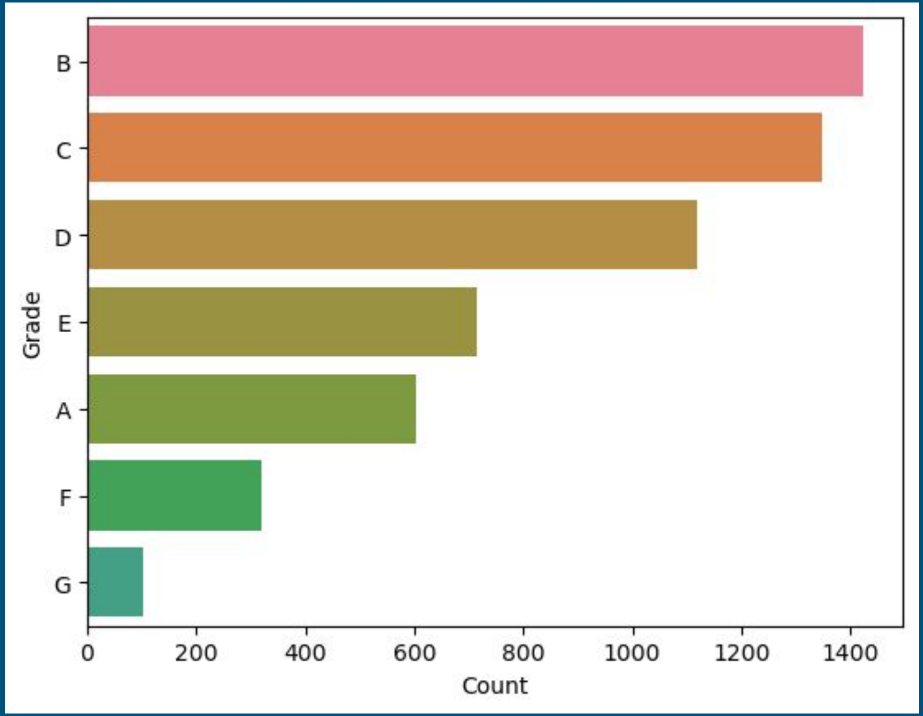
Analysis Summary

Loan Purpose



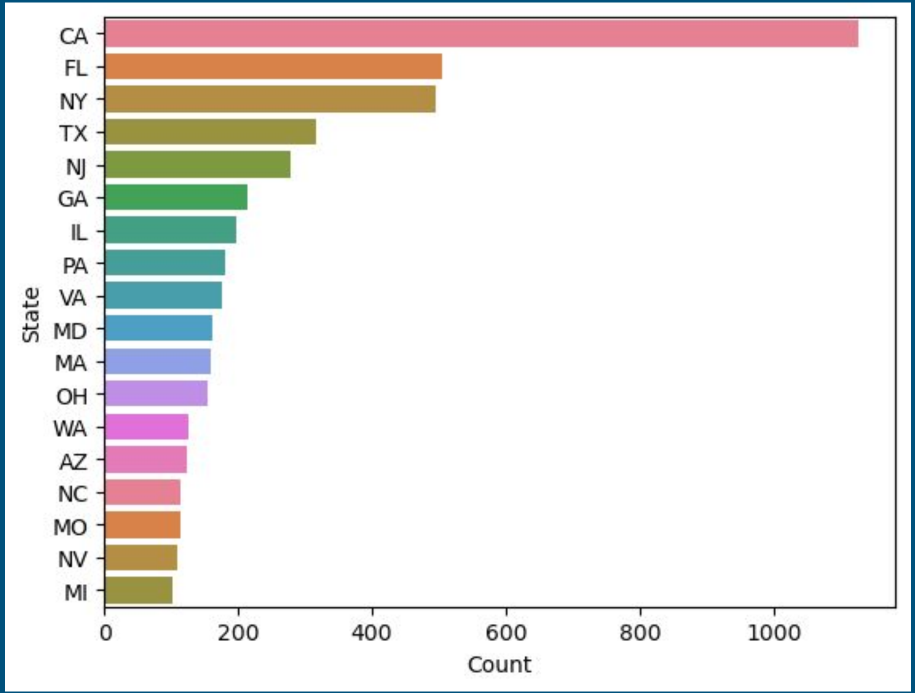
- According to the chart, it is clear that if a consumer's purpose for a loan is to consolidate existing debt, there is a high likelihood that the loan will be charged off.

Loan Grade



- There is a high chance that a loan will be charged off if the loan grade is B, C, or D.

State



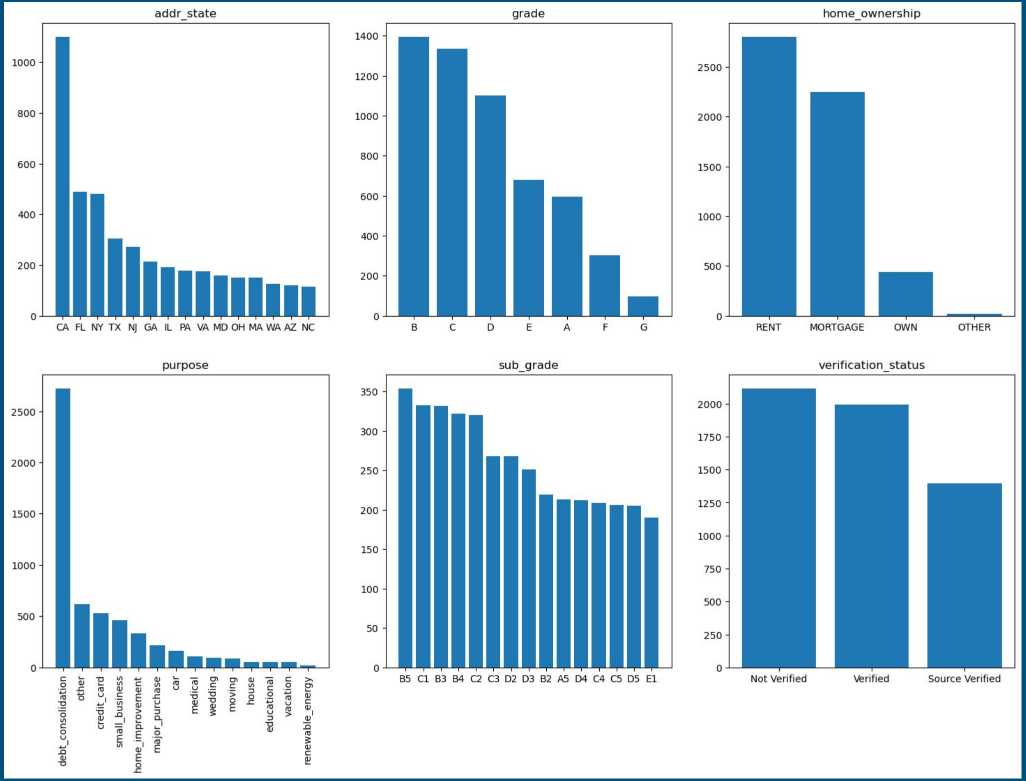
- Borrowers residing in the states of CA and FL are not likely to repay the amount

Univariate: Observations

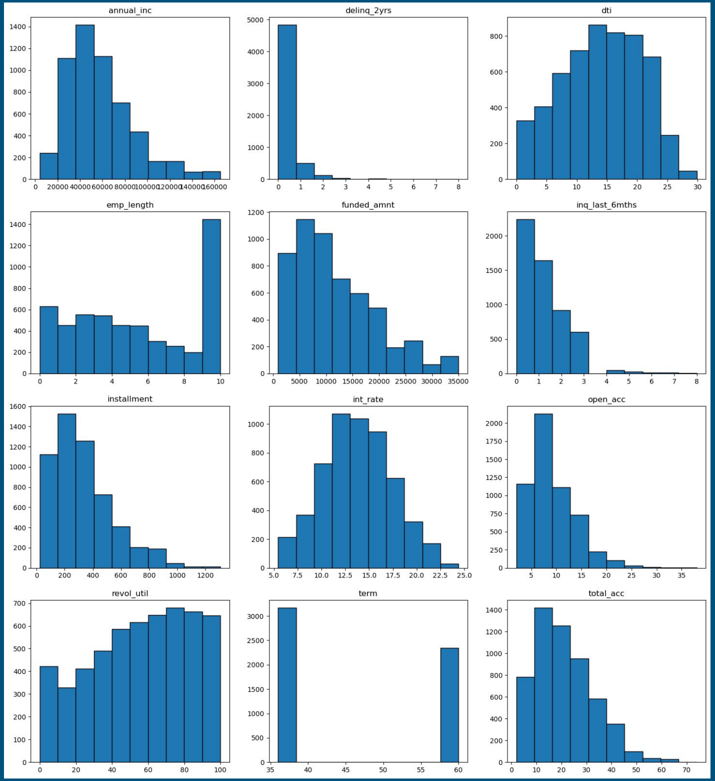
The likelihood of a loan being charged off is based on the following conditions.

- Loan Attributes
  - The loan term is 36 months.
  - Interest rate is between 10 to 15%.
  - Grade is B or C.
  - Verification status - verified is slightly better in terms of loan closure.
  - Funded amount between 5000 to 12000.
- Consumer Attributes
  - Consumer's annual income between 35000 to 65000.
  - Consumer's DTI between 10 to 20.
  - Borrower's open credit lines is between 5 to 10.
  - The loan purpose is for debt consolidation.
  - Consumer is staying in the state CA.
  - Consumer is staying on Rent or his home is on mortgage.

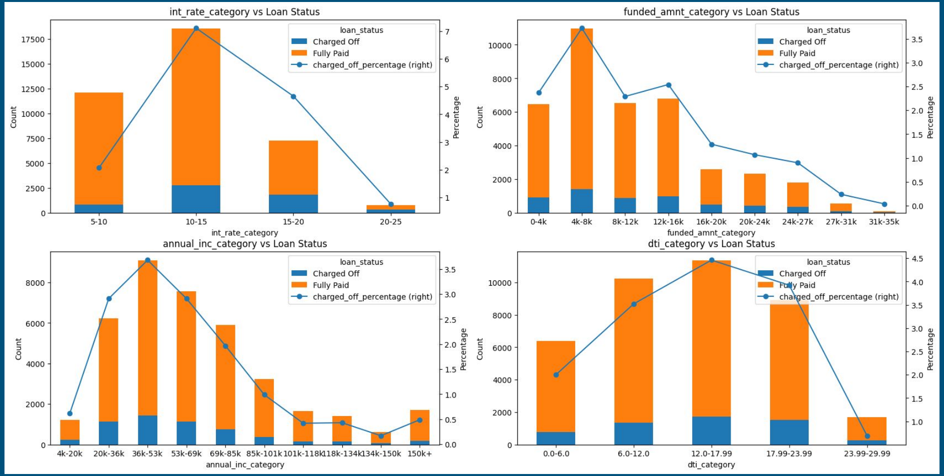
Univariate - Categorical



Univariate - Numerical



Fields vs Loan Status



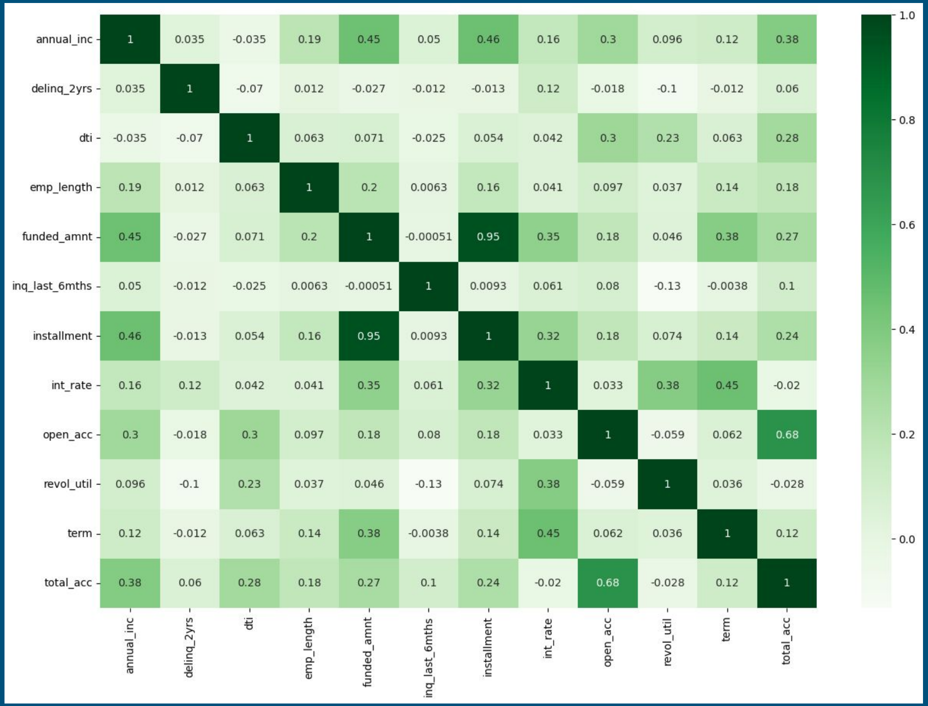
- If the interest rate increases, the likelihood of the loan being charged off also increases.
- The majority of loan amounts fall between 4000 and 8000, and there is a high probability that these loans will be fully repaid.
- As annual income increases, the chances of a loan being charged off decrease.
- The higher the DTI value, the greater the likelihood of the loan being charged off.

Purpose vs Requested amount

Purpose	Average Loan Amount
car	6714
credit_card	11452
debt_consolidation	12388
educational	6692
home_improvement	11220
house	12712
major_purchase	7970
medical	8103
moving	6506
other	7893
renewable_energy	8508
small_business	13015
vacation	5379
wedding	9752

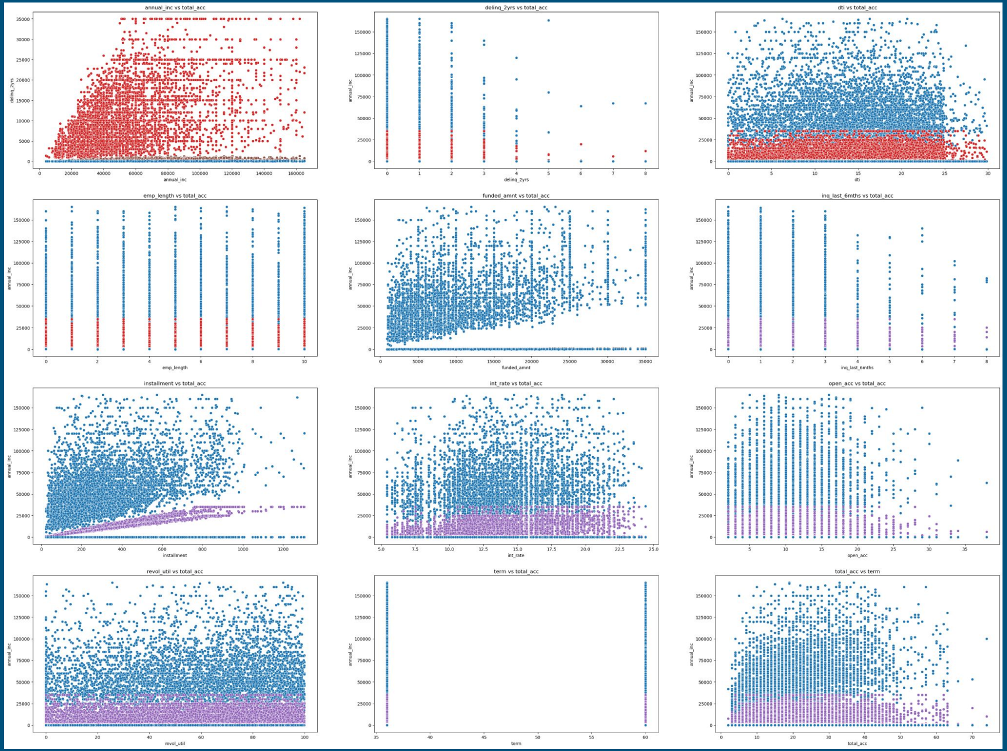
- LC can compare the purpose vs the requested amount, and if the consumer is asking for too much money, then LC can reduce the loan amount.

Correlation Analysis



- installment has a strong correlation with funded_amnt.
- term has a correlation with interest rate and funded_amnt.
- annual_inc has a correlation with funded_amnt.
- dti has weak correlation with most of the fields.
- annual_inc has a negative correlation with dti

Bi-Variate Analysis



Thank You