

Lending Club Case Study

SUBMISSION

Group Members:

Neelam Mahapatro
Chinju K Raveendran

Dec 7 – 2022

Batch – C46

Lending Club Case Study

Introduction:

Lending Club (LC) is the world's largest peer-to-peer lending company, headquartered in San Francisco, California ▪ It was the first peer-to-peer lender to register its offerings as securities with the Securities and Exchange Commission (SEC), and to offer loan trading on a secondary market.

How it Work's:

- Customers interested in a loan complete a simple application at LendingClub.com.
- Lending Club leverages data and technology to screen borrowers, facilitate the transaction with appropriate interest rates, and service the loan.
- Qualified applicants receive loan offers and can evaluate their options with no impact to their credit score.
- Investors select the loans they want to invest in based on their own risk tolerance, investment portfolio goals, and time horizon.

Objective :

- Identify variables that provide strong indicators of potential loan default thus helping Lending club to decide approval/rejection of loan.

Data cleaning steps

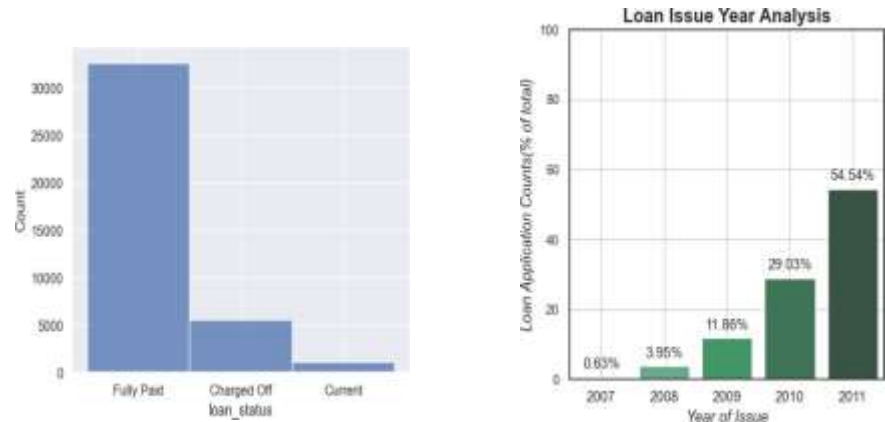
- Delete columns: Delete unnecessary columns.
- Removed column with null values: which are having more than 90 percent of values as null.
- Remove outliers: Remove high and low values that would disproportionately affect the results of your analysis.
- Missing values: Treat missing values with appropriate approach.
- Cleaning of Duplicate data: Remove identical rows, remove rows where some columns are identical.

Analysis

- The essence of the whole project is to analyze and understand how consumer attributes and loan attributes are influencing the tendency of defaulting.
- We performed data cleaning and preparation on the Loan dataset:
 - Imputed the NA values for all the variables
 - Created two new columns:
 - Profit and Loss column
 - Ratio of funded amount and annual income
- During univariate analysis we have created:
 - 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.

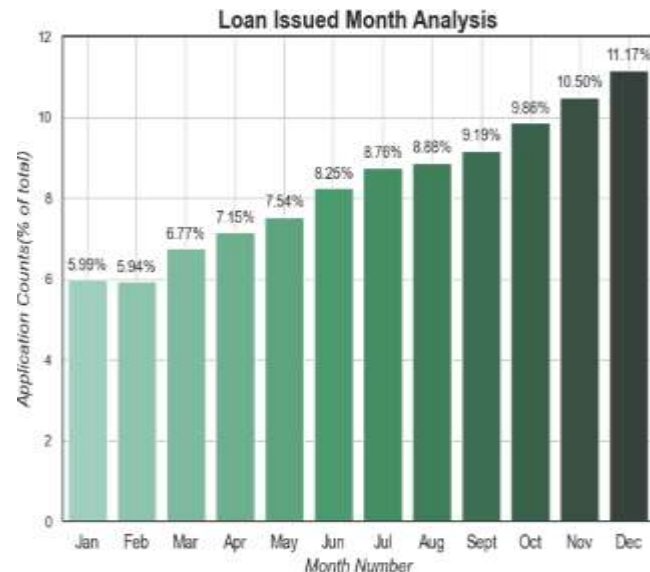
Univariate Analysis: Insights

- Overall Default Rate stands at 14.59%.

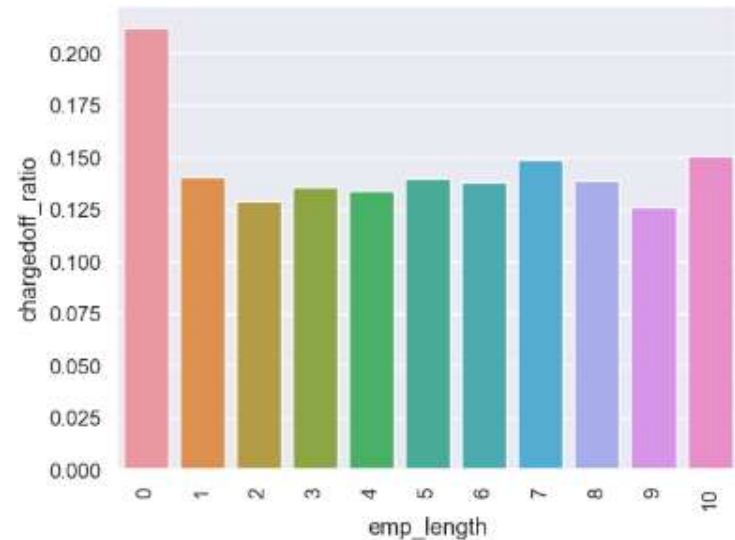
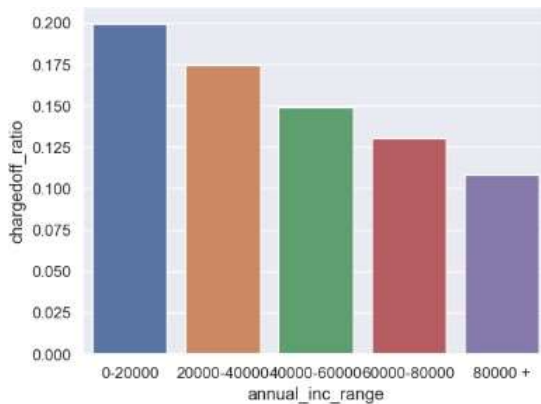
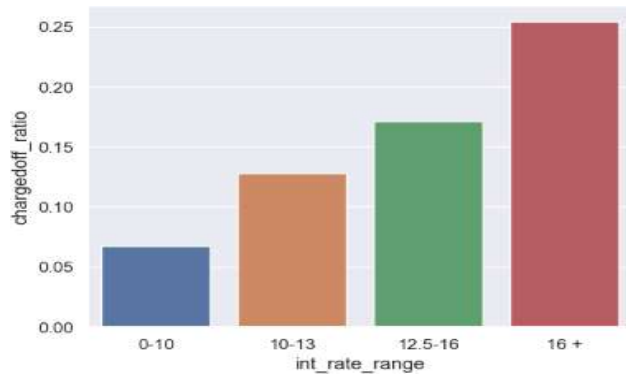


Ordered categorical Variables

- Approval Rate of loans is higher during the Holiday Season (November & December).
- Approved Loan amount almost doubled between years' 2010 and 2011.



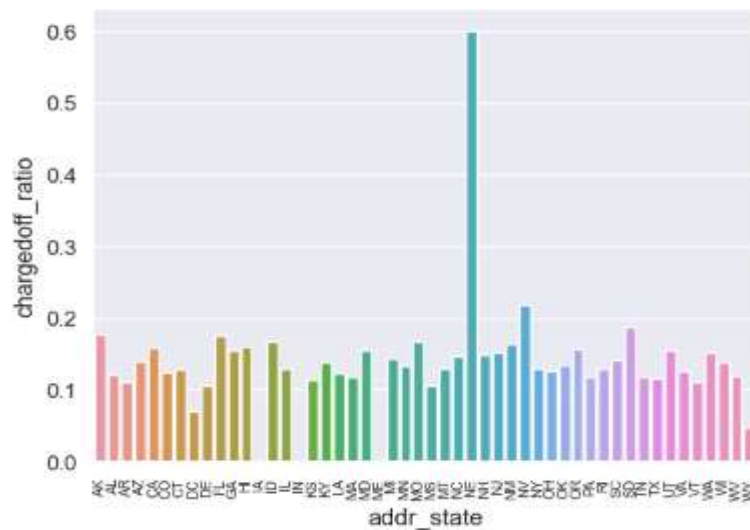
Bivariate Analysis: Insights



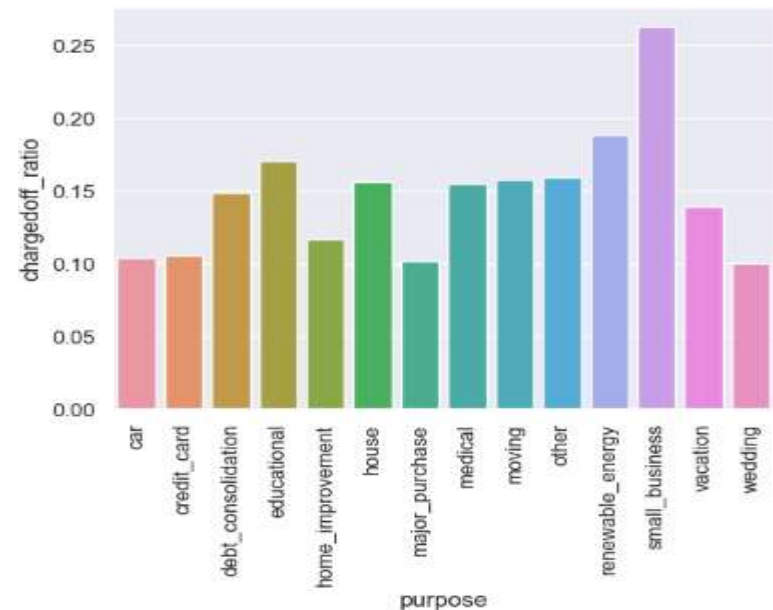
- Charged off ratio increases with interest rate.
- Highest charged off ratio is on zero employee length.
- Annual income of first group between 0-20000 have more charged off ratio. And it reduces with increase in annual income

Bivariate Analysis: Insights

- State vs Charged off

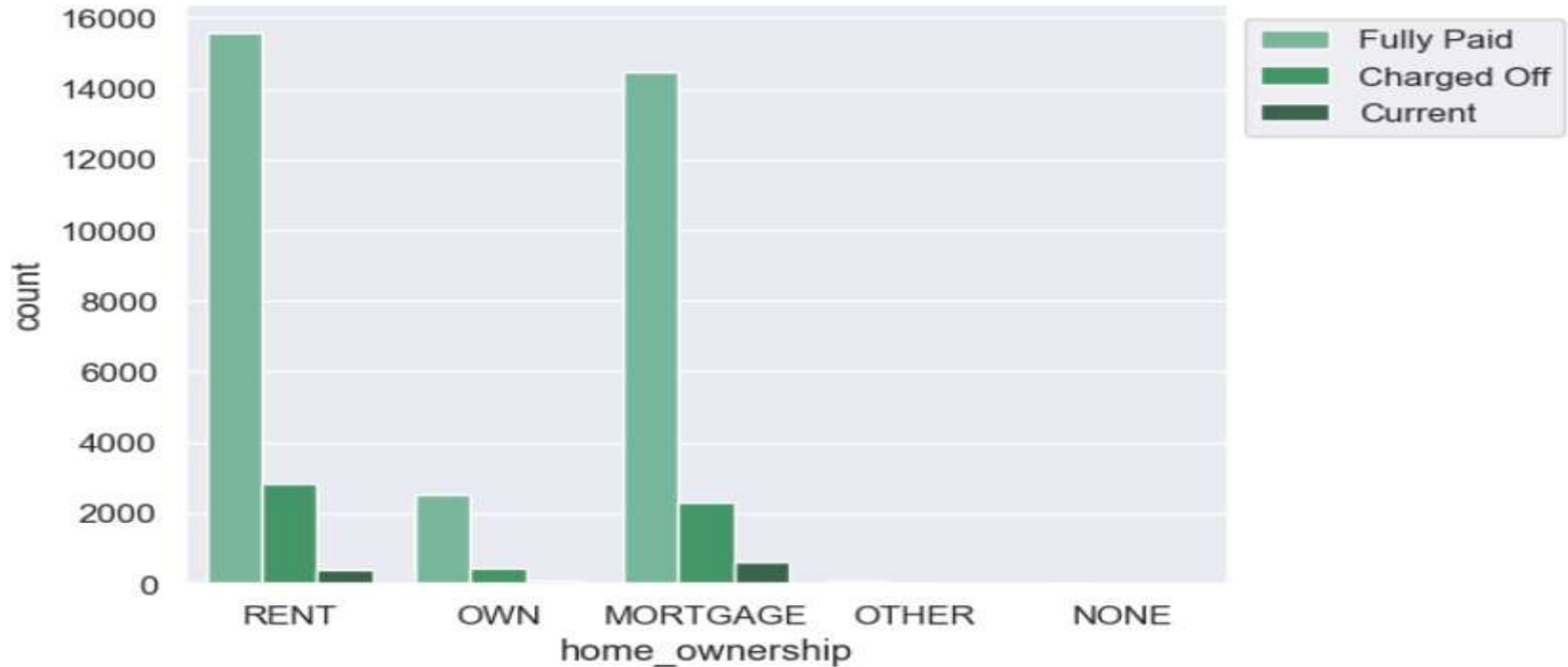


- Loan purpose analysis



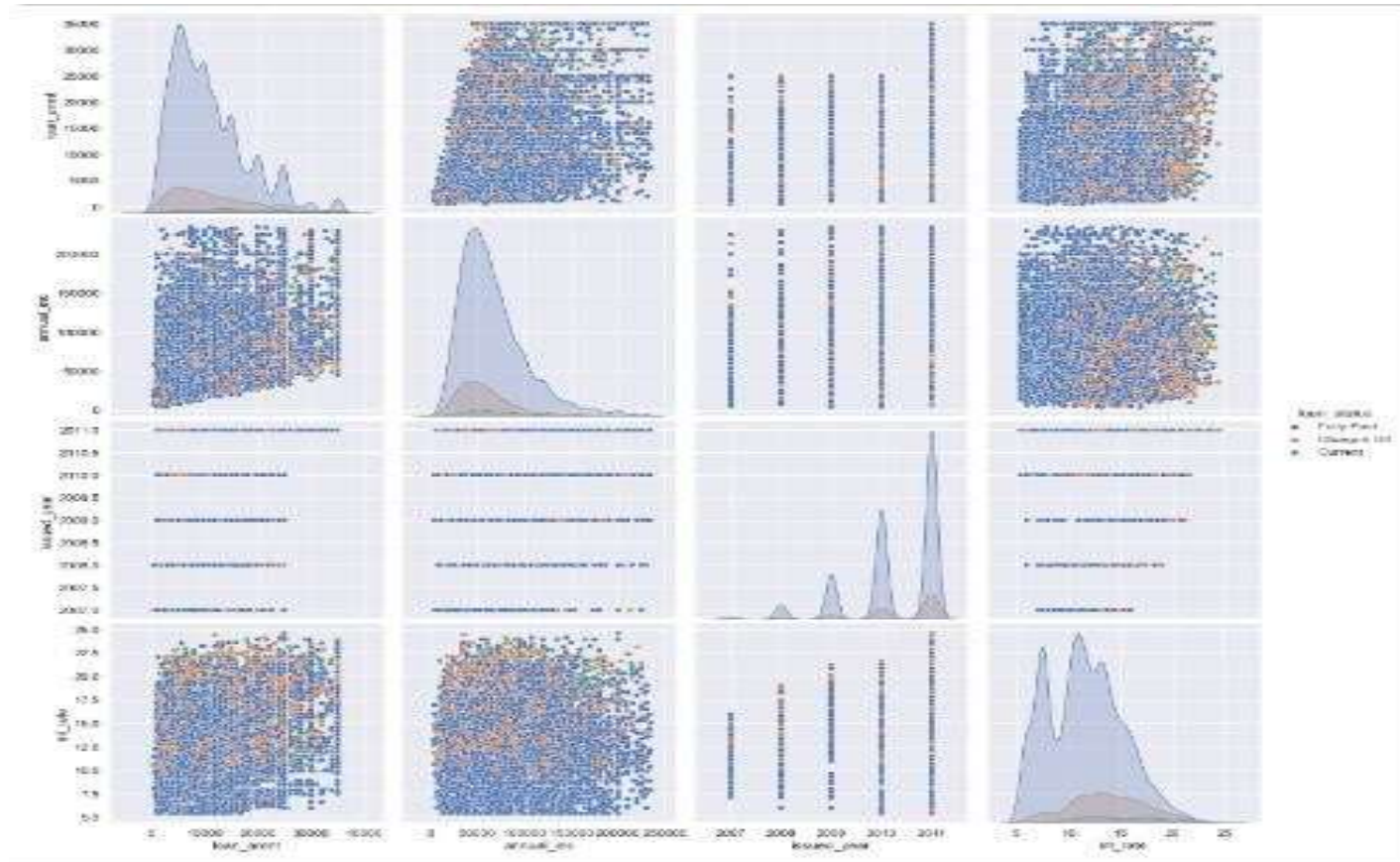
- Small business followed by renewable energy , are the top charged off purposes .
- NE has more than twice as many charged off ratio as compared to any other state.

Bivariate analysis insights



- 91 % of the defaulters already have a mortgage taken out or live on rent.

Multivariate Insights



Pair plot in variables loan amount ,annual income, issued year, Interest rate with loan status (Fully paid, charged off, and current)

Conclusion

Driving Factors for granting loans-

- Annual Income :- Applicants from 'Low'(0-20K) and 'Medium'(40-60K) income group have a greater share of defaulted loans.
- Employment Length : Maximum number of defaulters have experience 0 to 2 years of experience and 10/10+ years of . Hence, LC should be take this aspect into consideration while lending loans.
- Loan Purpose :- The top two reasons for loans are small business and renewable energy. Such applications should be carefully assessed.
- Address state:-Applicants from state NE has more defaulted or charged off Ratio.
- Interest Rate:-Charged off ratio increases with interest rate.