LENDING CLUB CASE STUDY

Aatir Kirmani

CASE STUDY OVERVIEW

- The consumer finance company “Lending Club” specializes in lending various types of loans to its urban customers for a range of purposes. The purpose of the case-study is to establish key driver variables that lead to charge-off (default).
- Lending is a very risky business. If loans are accepted for risky applicants, it has a higher probability to result in a financial loss to the company/investors. Therefore, it is essential to find these risky applications and approve loans with stricter conditions (or reject in some cases) to minimize the losses.
- This can also help the company to model risk and create robust risk mitigating system/rules.

DATASET OVERVIEW

- The dataset contains records of 39717 unique applications, where each application belongs to one applicant only.
- The dataset spans multiple years 2007 – 2016 and contains applications from various states of the US.
- The dataset contains attributes/features about the loan and the applicant.
- The applicant features include annual income, employment length, home ownership etc. while the loan features include purpose, interest rate, loan amount.
- The dataset also includes features from the credit history of the applicant such as delinquencies in last 2 years, public records, number of credit inquiries etc.

DATA PREPROCESSING

- Converting the columns to the correct format such as converting date columns to the appropriate format, removing % symbols for the numeric columns etc.
- Checking for the missing values and treating the features accordingly.
- Some columns in the dataset such as policy code, application type etc. contained same value for all the records and therefore would have not provided any meaningful information towards the Loan charge-off. Such columns were removed from the dataset.
- Columns with more than 30% Null values were also dropped, since imputing would have resulted in significant changes in the distributions of the data values.
- Columns such as public record bankruptcies, employment length etc. had only a handful of Null values and therefore median was used for imputation.

UNIVARIATE ANALYSIS

- The univariate analysis revealed certain interesting details about the dataset.
- The dataset is heavily imbalanced, the percentage of Charged Off applications are only 14.16%, whereas the Fully Paid applications are around 85%, rest are Current (active applications).
- Since, analyzing the Current applications would not have made much sense, therefore the data records corresponding to the Current applications were also removed.
- Most applicants are from California (18%), the next highest being New York (9.6%).
- Most applicants live in a RENTED (48%) or MORTGAGED (44%) properties.
- Most applications are for the debt consolidation (46.8%), next one being credit card at a distant 13%.

BIVARIATE AND MULTIVARIATE ANALYSIS

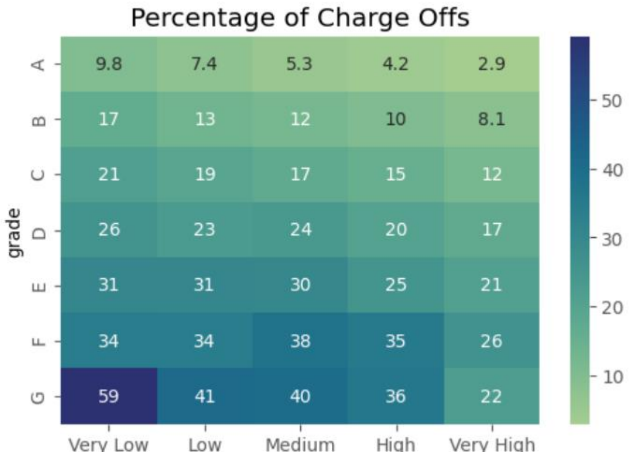
- From the Bivariate analysis of categorical features, following key results were obtained:
  - Percentage of charge-offs from grade **G** are highest, while for **A** it is the lowest.
  - Percentage of charge-offs from sub-grade **F5** are highest, while for **A1** it is the lowest.
  - No significant differences in the percentage of charge-offs for difference categories of home ownership.
  - No significant differences in the percentage of charge-offs for difference categories of verification status.
  - Nebraska has the highest percentage of charge-off as compared to any other states, while Wyoming has the highest percentage of Fully Paid loans.
  - Loans for small business has the highest percentage of charge-offs.

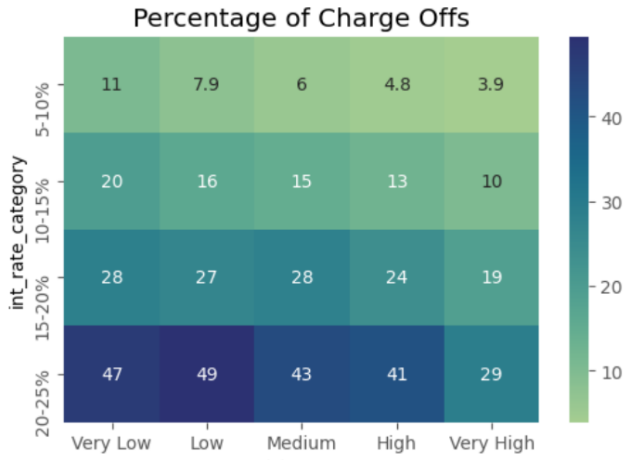
BIVARIATE AND MULTIVARIATE ANALYSIS

- From the Bivariate Analysis of Categorical – Numerical features, following key results were obtained:
  - Higher interest rates are associated with higher charge-offs.
  - Loan amount is significantly high for G grade and sub-grade categories while it is the least for A grade/sub-grades.
  - Loan amount is significantly higher for small business and debt consolidation purposes while it is lowest for vacation, moving and educational purposes.
  - Applicants with higher annual income have lesser charge-offs, tend to take loan for longer term and primarily for home-improvement, renewable energy, small business and housing purposes.
  - Higher dti (debt to income ratio) is associated with higher charge-offs on average.
  - Higher dti is also associated with loans for debt consolidation and credit card as expected.

BIVARIATE AND MULTIVARIATE ANALYSIS

- From the Multivariate Analysis of features, following key results were obtained:
  - Low annual income and high interest rates lead to high percentage of charge-offs.
  - Loans for small business leads to high charge offs irrespective of the income category.
  - Loans for renewable energy leads to high charge offs unless income category is not Very High.
  - Lower Incomes and Higher Grades lead to high charge-offs.

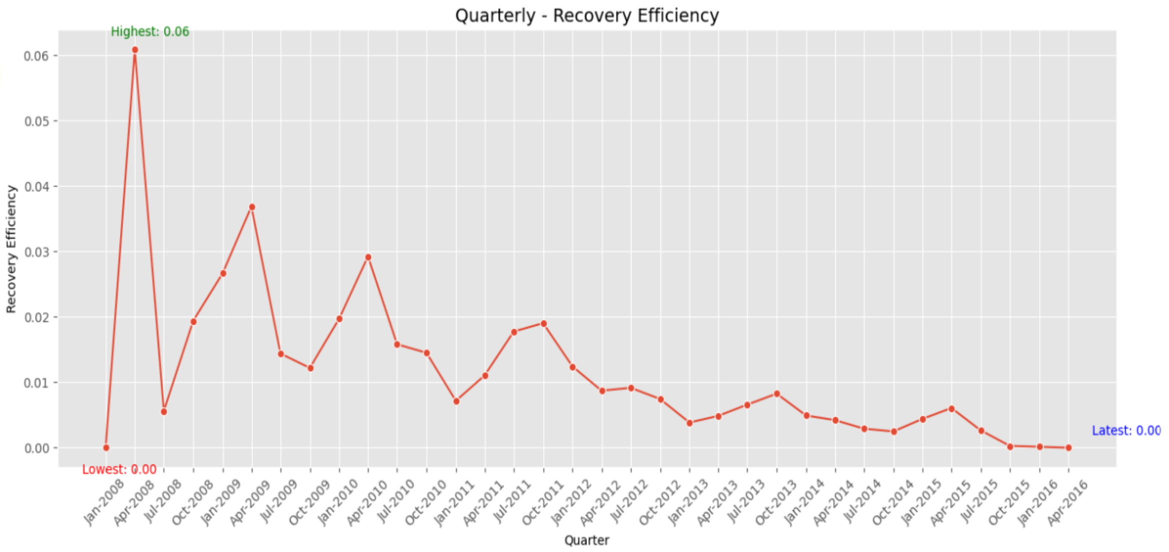




KEY BUSINESS METRICS

Loan Performance Rate: tells us the proportion of the principal amount that borrowers have successfully repaid over the total amount funded by the lender.

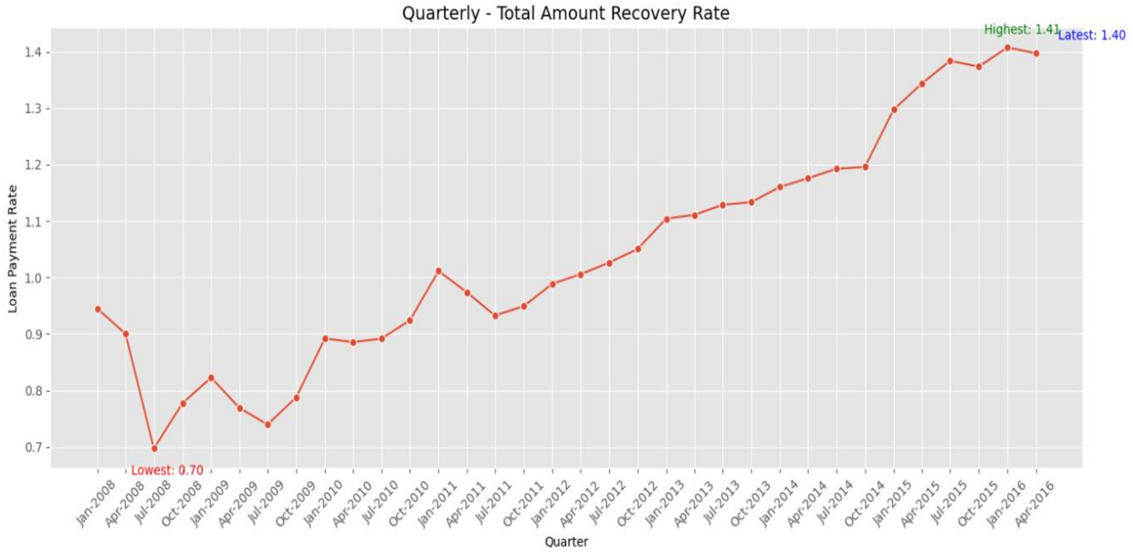
Recovery Efficiency: This metric indicates the effectiveness of the recovery process. It shows the portion of the funded amount that has been recovered after the loan has charged-off.



KEY BUSINESS METRICS

Total Amount Recovery Rate: tells us the proportion of the principal amount that borrowers have successfully repaid over the total amount funded by the lender.

Average Interest Earned: This metric indicates the effectiveness of the recovery process. It shows the portion of the funded amount that has been recovered after the loan has charged-off.



SUMMARY

Key driver variables for the Loan Charge-Offs:

- Annual Income (annual income category)
- Interest Rate (interest rate category)
- Purpose
- Grade
- Sub Grade
- DTI (Debt to Income Ratio)
- State

Recommendations:

- Careful consideration of loan applications where interest rates are high and the annual income of the applicant is low.
- Investigating the charge-offs from Nebraska.
- Higher collateral for loan applications for the purpose of small business.
- Thorough screening of applications where the grade is F or above and sub-grade is F1 or above.
- Careful consideration (often rejecting) of loan applications for grade G where income category is very low (59%) charge-offs.