Lending Club Case Study

Understanding Loan Risk Analytics in for a consumer finance company using Exploratory Data Analysis (EDA).

-- Suddhasish Kar

Agenda

- Background of case study
- Problem Statement
- Process
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Conclusion and Key Driving Factors for Loan Defaults
- Recommendations for Risk Mitigation

Case Study Background

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

- If one can identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using **Exploratory Data Analysis** (EDA) is the aim of this case study.

- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Problem Statement

For a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company must decide 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 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.

Process

Technology Used: Python (Numpy,Pandas,Matplotlib, Seaborn) → Data Source: Loan.csv dataset being used to analyze the problem statement. → Data Cleanup and Manipulation → Univariate Analysis → Segmented Univariate Analysis → Bivariate Analysis → Insights / Recommendation

Used Variable

Numerical:

- loan_amnt: The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
- annual_inc: The self-reported annual income provided by the borrower during registration.
- inq_last_6mths: The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
- total_pymnt: Payments received to date for total amount funded
- revol_util: Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
- funded_amnt: The total amount committed to that loan at that point in time.
- last_pymnt_amnt: Last total payment amount received.
- int_rate: Interest Rate on the loan
- emp_length: Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
- revol_bal: Total credit revolving balance.
- earliest_cr_line: The month the borrower's earliest reported credit line was opened
- emp_length: Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
- recoveries: post charge off gross recovery.
- revol_util: Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.

Used Variable

Categorical:

- loan_status: Current status of the loan
- total_rec_late_fee: Late fees received to date
- purpose: A category provided by the borrower for the loan request.
- home_ownership: The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
- verification_status: Indicates if income was verified by LC, not verified, or if the income source was verified.
- verification_status: Indicates if income was verified by LC, not verified, or if the income source was verified.
- Grade: LC assigned loan grade.

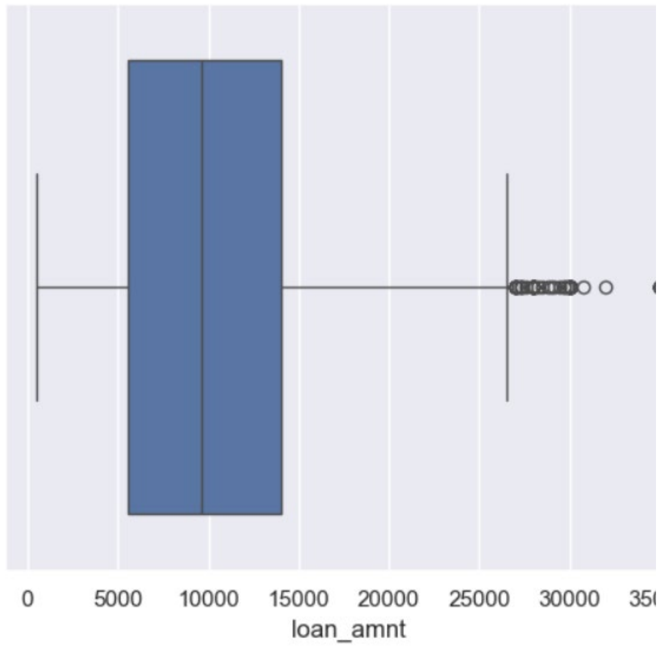
Data Cleanup

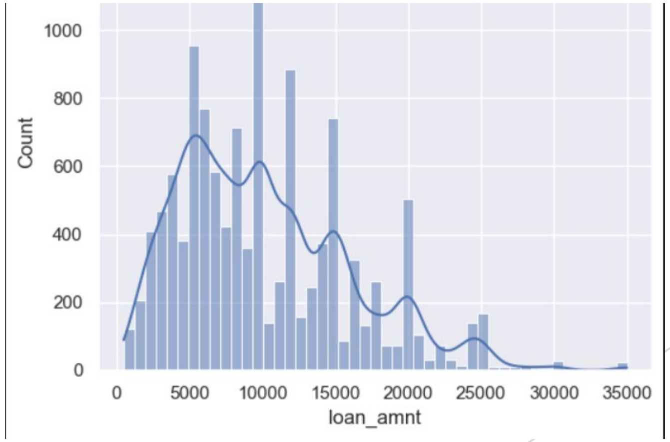
- Dropped Columns where we have maximum number of null value
- Dropped columns where we have Repetitive values and not Useful for analysis.
- Removed Rows which have null values.
- Derived umber of days from Date time value and created 2 new columns

  `last_pymnt_d_tilldate, earliest_cr_line_tilldate`

Univariate Analysis

- Loan Amount: Based on Univariate analysis we get to know loan is most frequently distributed around $10000 amount loan issues normally between $5000-$14000



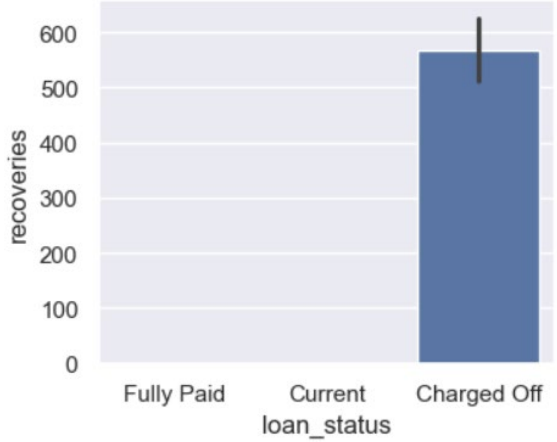


Univariate Analysis (Segmented)

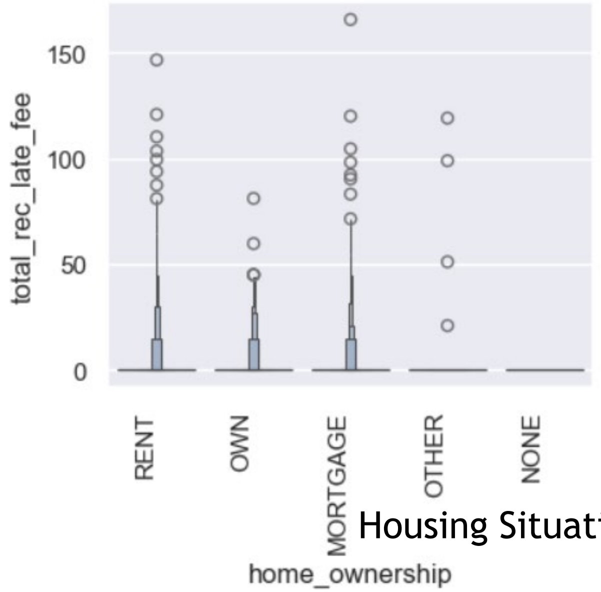
Univariate Analysis Insights:

- **Recovery Indicator:** Recovery is a strong indicator of a loan being charged off, as current or fully paid loans do not require recovery efforts.
- **Recovery Rate by Housing Situation:** Higher recovery rates are observed for customers who rent or have mortgages.
- **Recovery Rate by Loan Purpose:** Elevated recovery rates are associated with loans for small businesses, credit cards, and debt consolidation.
- **Recovery Rate by Loan Grade:** Recovery rates are lower for higher-grade loans (A, B, C) and higher for lower-grade loans (G, F, E, D), suggesting that lower-grade loans are more likely to be charged off.
- **Loan Distribution:** The company has issued more loans to categories deemed riskier.
- **Interest rate by Loan Grade:** Loan provided at higher interest rate for in this sequence A,B,C,D,E,F,G

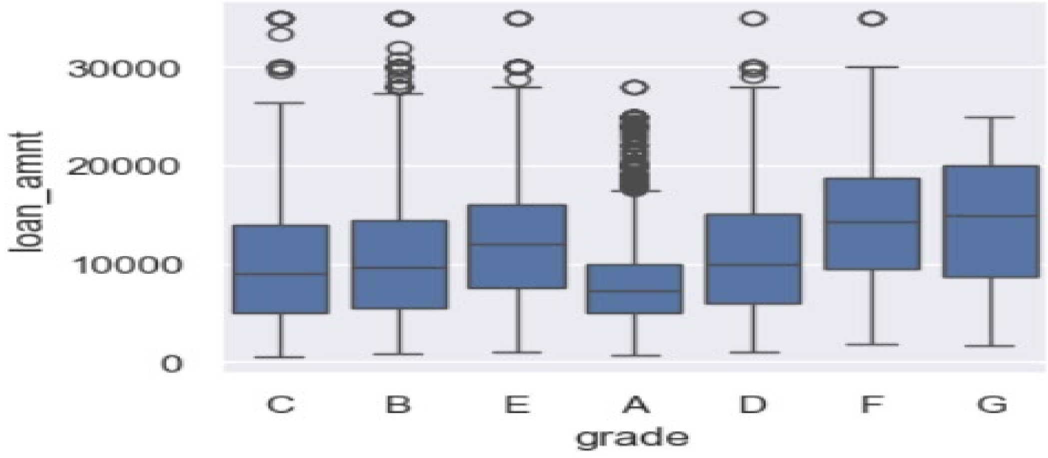
Segmented Univariate Plots



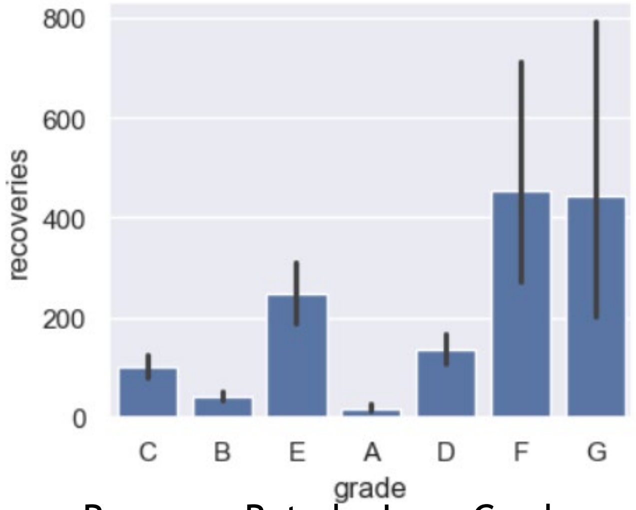
Recovery Indicator



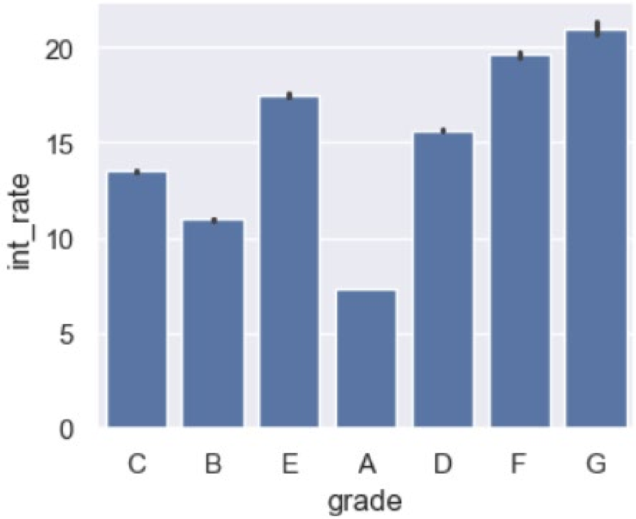
Housing Situation



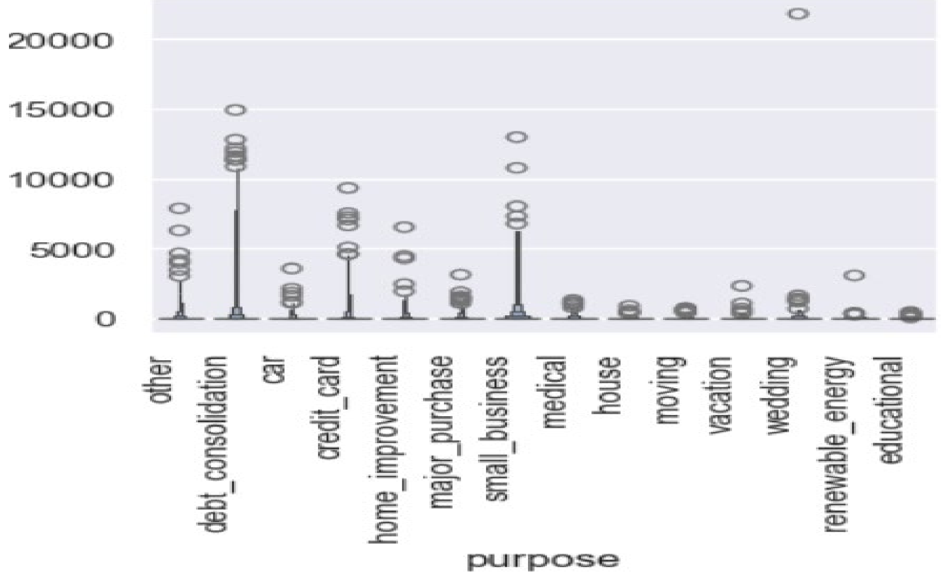
Loan Distribution



Recovery Rate by Loan Grade



Interest rate by Loan Grade



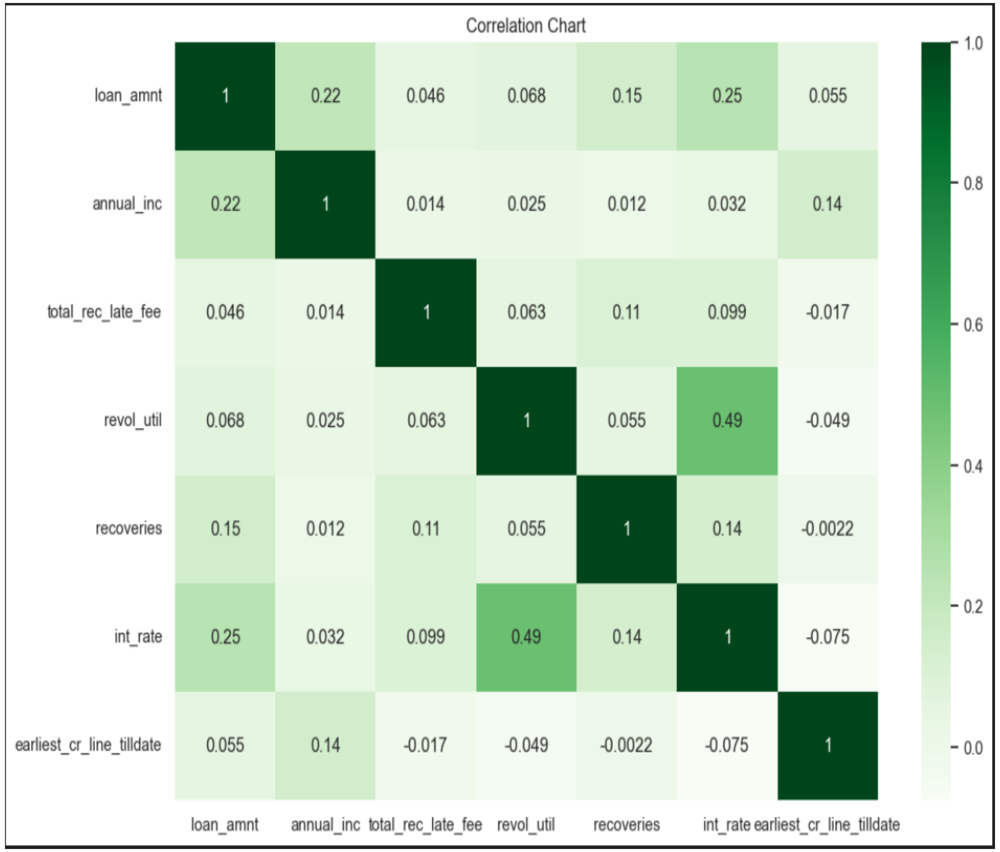
Loan Purpose

Bivariate Analysis

Bivariate Analysis Insights:

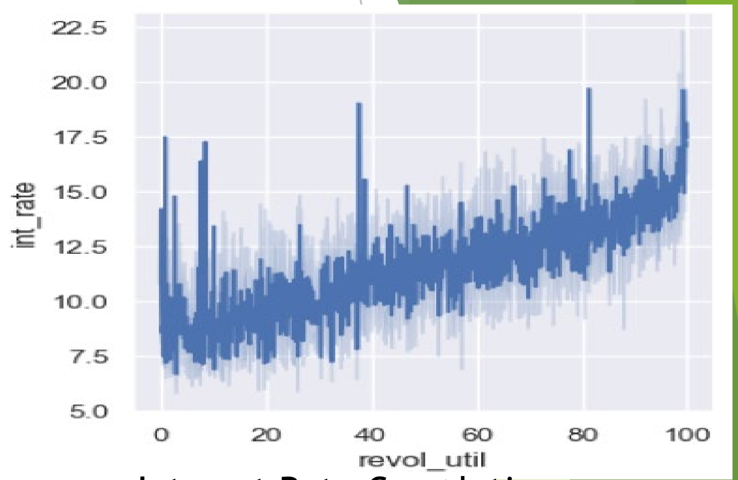
- **Loan Amount Correlations:** There is a positive correlation between loan amount and factors such as the earliest credit line, revolving utilization, and annual income.
- **Revolving Utilization:** High revolving utilization (over 60%) is linked to a higher likelihood of loan charge-offs.
- **Interest Rate Correlations:** Interest rates are positively correlated with revolving utilization but negatively with the age of the earliest credit line.
- Loan Grade and Credit Line Age: Higher loan grades are associated with older credit lines (over 10,000 days).
- Credit Line Age and Loan Status: Charged-off loans tend to have newer credit lines compared to loans that are fully paid or current.

Bivariate Plots

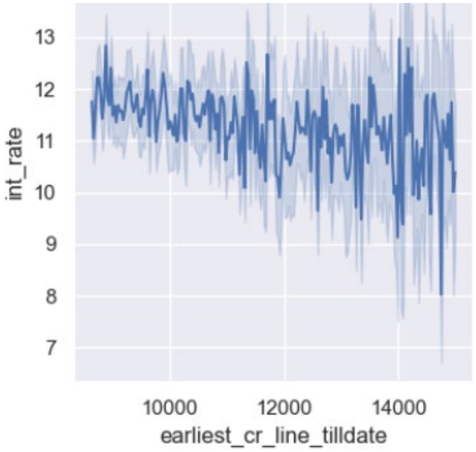


Loan Amount Correlations

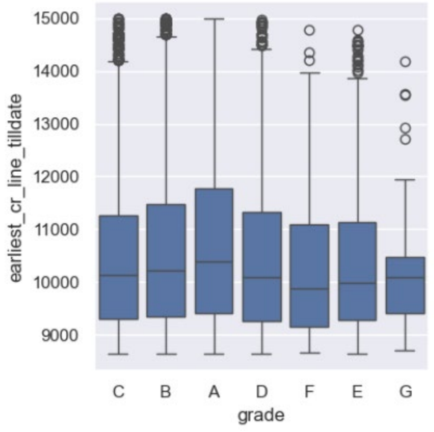
Revolving Utilization



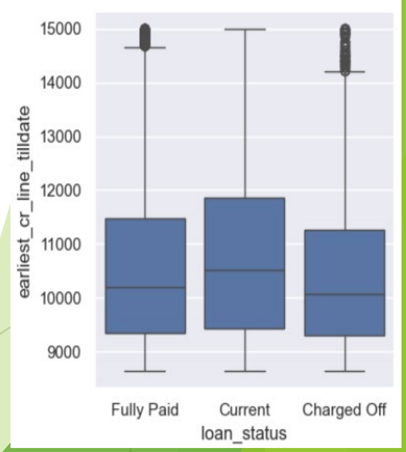
Interest Rate Correlations



Interest Rate Correlations



Loan Grade and Credit Line Age



Credit Line Age and Loan Status

Conclusion and Key Driving Factors for Loan Defaults

The primary factors influencing loan defaults include:

- **Loan Grade:** Lower-grade loans (G, F, E, D) are more susceptible to defaults.
- **Loan Purpose:** Loans for small businesses, credit cards, and debt consolidation show higher default rates.
- **Revolving Utilization:** High revolving utilization rates significantly increase the risk of default.
- **Credit Line Age:** Newer credit lines are more frequently associated with charged-off loans.
- **Late Payments and Last Payment Amount:** These are indicators of financial distress or unwillingness to repay, which correlate with higher default rates.

Recommendations for Risk Mitigation

- **Enhanced Scrutiny for Certain Loans:** Loans falling into lower grades, or for purposes like small business, should be subjected to more rigorous evaluation.
- **Interest Rate Adjustments:** Consider adjusting interest rates based on revolving utilization and credit line age to mitigate risks.
- **Monitoring Late Payments:** Close monitoring of late fees and last payment amounts can provide early warning signs of potential defaults.

By focusing on these driving factors, the company can better identify risky loan applicants and take appropriate actions to reduce credit losses, thereby enhancing profitability and stability in its lending operations.

Thank you