

Lending Club Case Study

Performing exploratory data analysis (EDA) on the loan approval system.

The problem

Company

The company consumer finance company which specialises in lending various types of loans to urban customers.

Context

Two risks associated with the bank's decision:

- Not approving the loan may result in a loss of business if the applicant is likely to repay.
- Approving the loan may lead to a financial loss if the applicant is likely to default.

Problem statement

The company aims to identify key factors driving loan default, which are strong indicators of default. This knowledge can be used for portfolio and risk assessment purposes.

Challenges deep-dive

Challenge 1

Variable Understanding

Understanding the variables in the provided dataset to identify the most valuable ones for performing operations.

Challenge 2

Cleaning the data set

Conducting a range of operations

- including removing unnecessary rows and columns
- cleaning invalid data
- standardizing the data
- and filling in missing values

Challenge 3

Performing analysis

Performing univariate and bivariate exploratory data analysis (EDA) and generating visual analysis.

Solution

Analyzing data to identify high-risk loan applicants.

Engaging in a thorough data analysis process to gain a comprehensive understanding of the factors and patterns that can help identify applicants with a higher risk of loan default.

Implementation

Univariate analysis on unordered data

1. Analysis on Loan status.

- a. A total of 21,212 individuals have successfully repaid their loans, while 3,188 individuals have experienced loan charge-offs.

2. Analysis on verification status.

- a. 11,739 individuals are unverified, 6,733 are source verified, and 5,928 are fully verified, indicating a relatively lenient loan approval process with minimal verification.

3. Analysis on home ownership status

- a. The majority of loan applicants either have an existing mortgage on their house or reside in rental properties.



Univariate analysis on ordered data

1. Analysis on loan term in months

- a. 19,346 individuals choose 36-month loans, while 5,054 individuals opt for 60-month loans.

2. Analysis on applicant Grade

- a. In the grade data analysis, grade A is the most common with 7,399 individuals, followed by grade B with 7,511 individuals. Grade C ranks third with 4,852 individuals.

3. Analysis on Sub-grades

- a. A5 and A4 have the highest counts, while B5, B3, and B4 exhibit closely comparable values.

4. Analysis on purpose of taking loan

- a. The majority of individuals have chosen to take out loans for debt consolidation purposes.



Univariate analysis on Quantitative data

1. Analysis Loan amount variable

- a. The analysis of requested loan amounts indicates that the highest amount requested by borrowers is \$10,000, followed closely by the second-highest requested loan amount of \$5,000.

2. Analysis Funded amount

- a. Upon analysis, it is evident that the highest funded loan amount is \$10,000, closely followed by the second-highest funded loan amount of \$5,000.

3. Analysis Funded amount by investor

- a. Max funded amount by investor: \$10,000
- b. 2nd highest funded amount by investor: \$6,000 (approx.)

4. Analysis on interest rate

- a. Most of the loan applicants fall between the 10 -16 % interest rate.

5. Analysis on Employment length

- a. A significant proportion of applicants possess 10 or more years of professional experience.

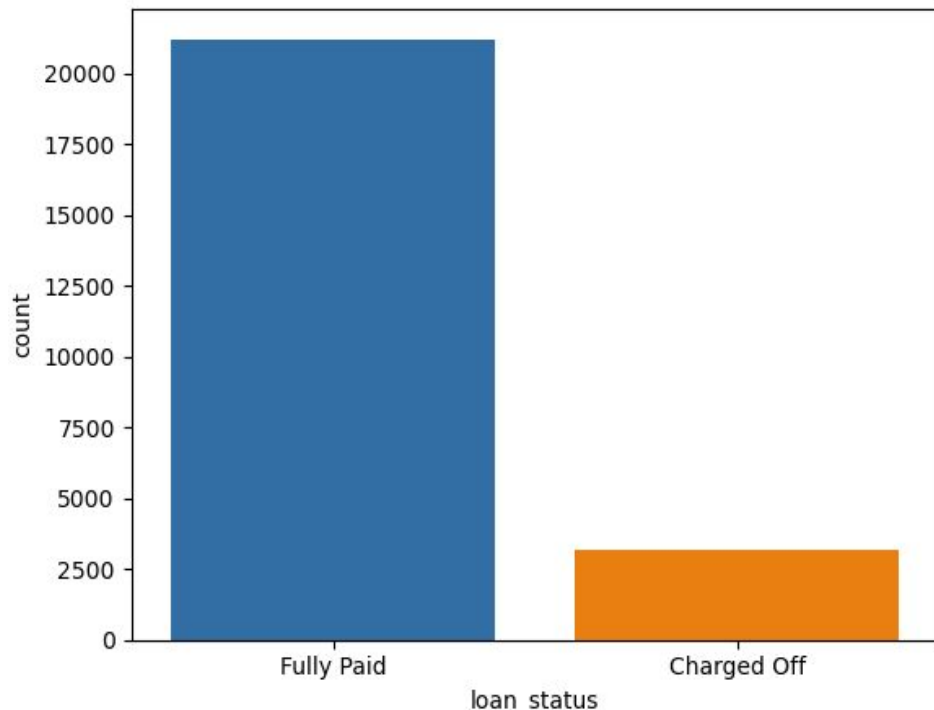
6. Analysis on annual income

- a. The majority of applicants have an annual income ranging between \$25,000 and \$60,000.

Key Insights from Univariate Field Data Analysis

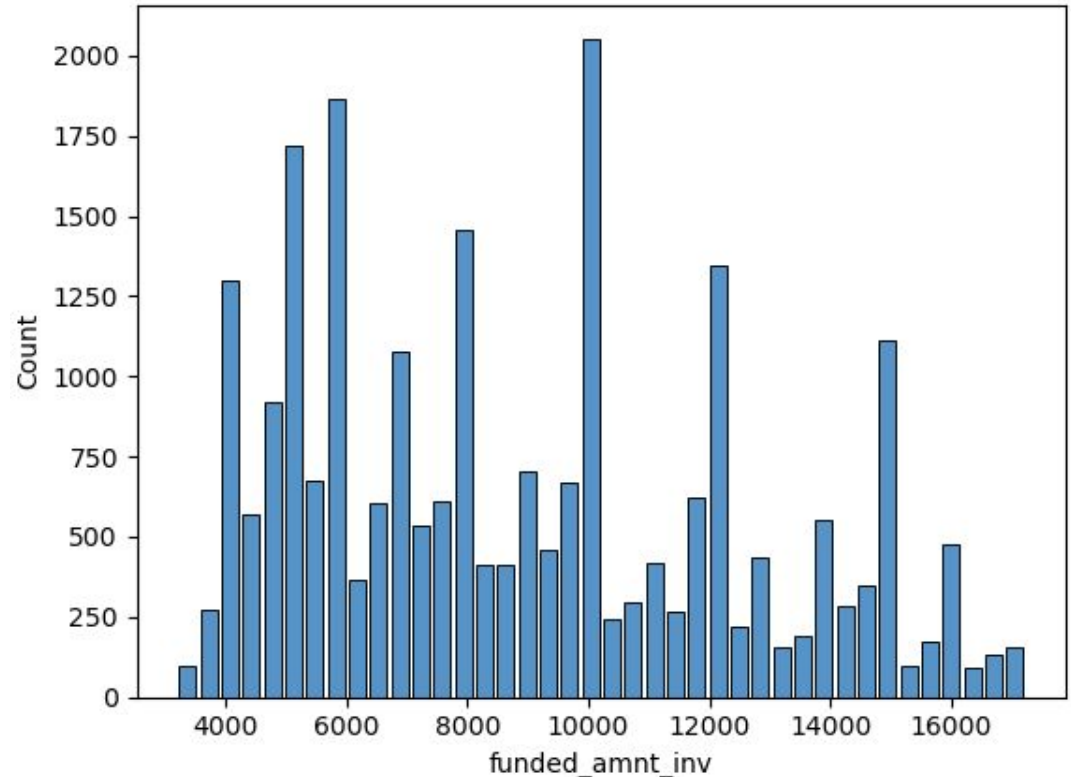
Loan Status Variable

- Within the loan portfolio, a total of 21,212 individuals have successfully fulfilled their payment obligations, while 3,188 individuals have experienced loan charge-offs.



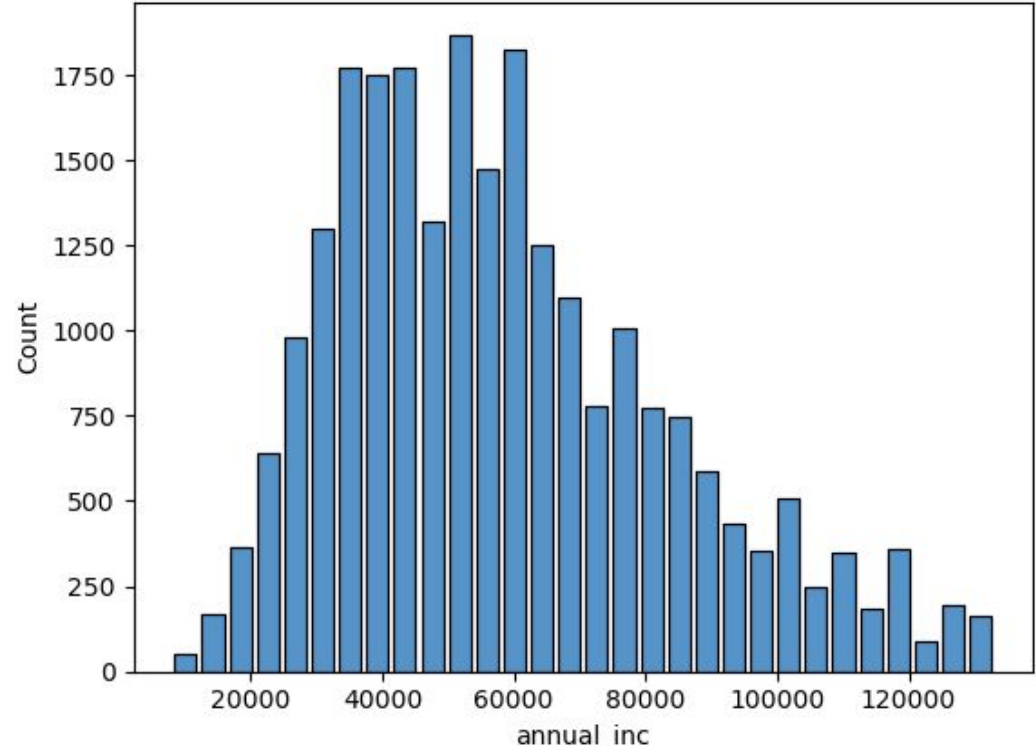
Funded loan amount

- The highest funded loan amount by an investor is \$10,000, while the second highest funded loan amount is approximately \$6,000.



Annual income Variable

- During the analysis, it was observed that a significant portion of the applicants has an annual income falling within the range of \$25,000 to \$60,000. This income bracket represents the majority of individuals seeking loans.



Bivariate analysis on unordered data

1. Analysis between purpose of loan and loan status.
 - a. The majority of charged off loan applicants had taken the loan for debt consolidation purposes.
2. Analysis between home ownership and loan status
 - a. The likelihood of loan charge-offs is higher among individuals who reside in mortgaged houses, rental properties, or other types of housing arrangements.
3. Analysis between verification of application and loan status
 - a. The percentage of loan defaulters appears to be higher among applicants who have undergone verification, including both source verified and fully verified individuals.



Bivariate analysis on ordered variables

1. Analysis between grade and loan status

- a. The count of charged off loan applicants is highest among individuals with grade B, C, and D. These grades represent the categories with the highest number of loan defaulters.

2. Analysis between term of loan and loan status

- a. While it is true that individuals with a 36-month loan term have a higher count of charge-offs, when considering the percentage, it is actually more likely for individuals with a 60-month loan term to experience charge-offs.

3. Analysis between sub-grade of applicant and loan status

- a. Loan applicants with grades C1, B5, B3, C2, and B4 exhibit a significant count of loan defaults, indicating a higher propensity for defaulting among individuals in these specific grade categories.



Bivariate analysis on Quantitative Variable and loan status

1. Analysis annual income and loan status

- a. Individuals with incomes close to \$40,000 and around \$60,000 have a higher count of charged-off loans, indicating a greater likelihood of loan defaults within this income range.

2. analysis loan amount and loan status

- a. Applicants with loan amounts around rounded values such as \$10,000, \$12,000, \$4,000, and \$5,000 exhibit a higher count of charged-off loans.

3. Analysis employment length and loan status

- a. Individuals with employment length exceeding 10 years demonstrate the highest count of charged-off loans.

4. Analysis rate of interest and loan status

- a. Higher interest rates are associated with an increased likelihood of loan charge-offs.

5. Analysis funded amount by investor and loan status

- a. Loan applicants with funded investor amounts close to rounded values like \$10,000, \$12,000, \$4,000, and \$5,000 tend to have a higher count of charged-off loans.

6. Analysis funded amount after approval and loan status

- a. Applicants with funded loan amounts closer to rounded values, such as \$10,000, \$12,000, \$4,000, and \$5,000, tend to exhibit a higher count of charged-off loans.

Bivariate analysis on Quantitative Variables and verification status


1. Analysis on funded amount and verification status

- a. Loan funds below \$12,000 are granted more easily with either no verification or only source verification requirements.

2. Analysis on interest rate and verification status

- a. Verified users typically encounter higher interest rates compared to non-verified users.

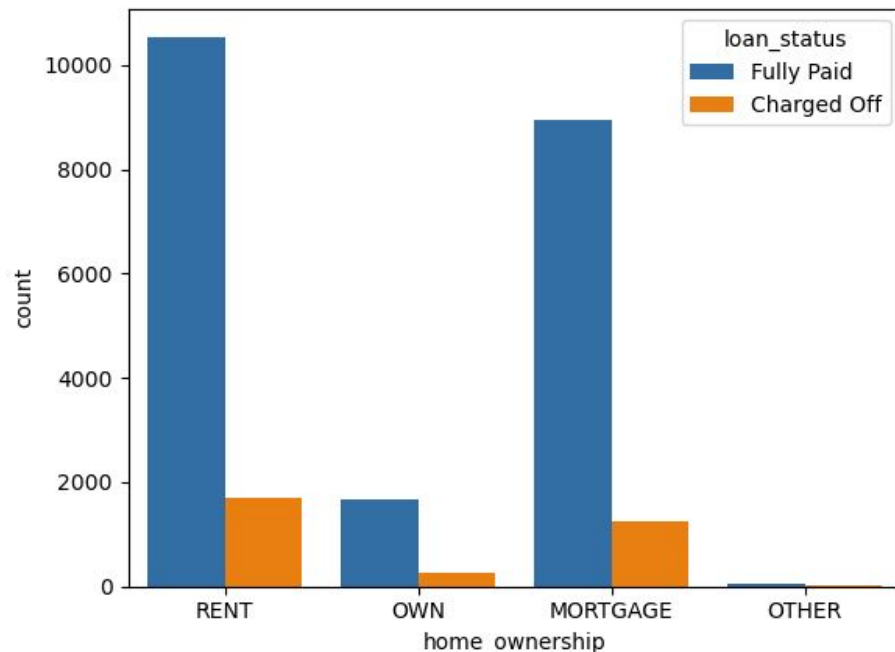
3. Analysis on funded amount and home ownership status

- a. The count of loan applicants requesting loan amounts below \$12,000 is generally lower among homeowners, while for mortgage homes, the count is higher for loan amounts exceeding \$12,000.
 - b. Applicants with rental properties predominantly have interest rates ranging from 10% to 19.29%.
 - c. Furthermore, individuals residing in mortgaged homes exhibit a higher count for interest rates between 5% to 8%.
- 

Key Insights from Bivariate Field Data Analysis

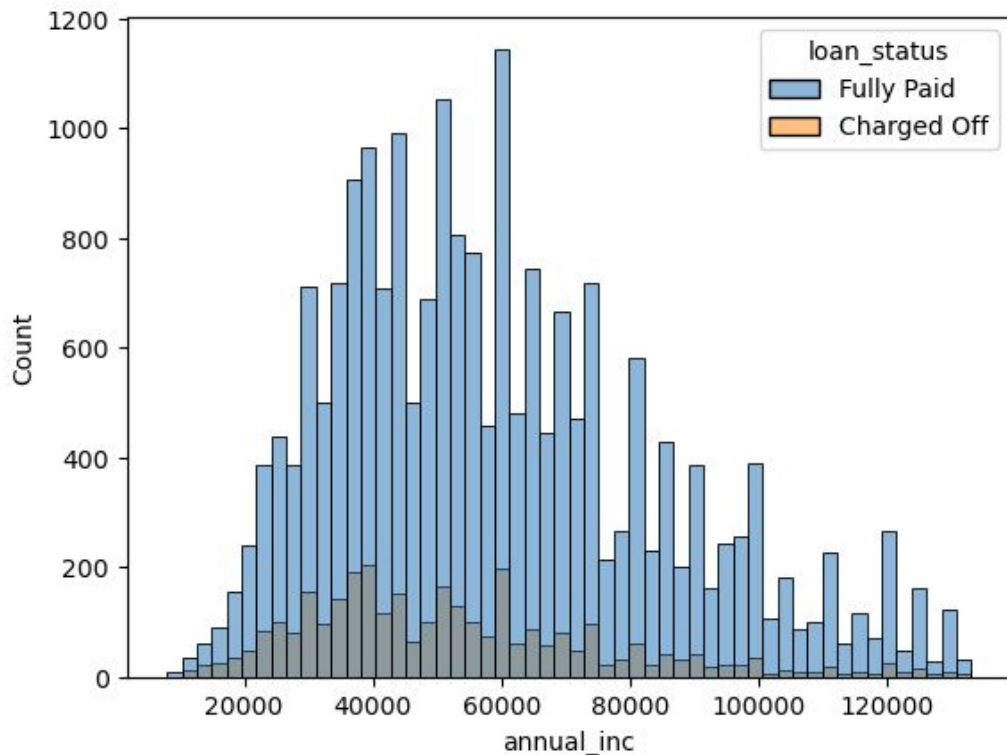
Home owners and loan status

- The graph clearly indicates that individuals residing in mortgaged houses, rental properties, or other types of housing are more likely to experience loan charge-offs.



Annual income and loan status

- Indeed, individuals with incomes close to \$40,000 and around \$60,000 have a higher count of charged-off loans. However, it is important to note that among those with an income around \$60,000, there is also a high count of individuals who have fully paid off their loans, making it the highest category in terms of successful loan repayment.



The team

