



Lending Club Case Study

Author:
SHIVAKEERTHI H



Abstract

- The aim of this study is to analyze lending data from Lending Club using Python to identify prospective borrowers who may default on their loans.
- In this study, we employ Python programming language and various data analysis techniques to explore the Lending Club dataset. We analyze features such as borrower characteristics, loan attributes, credit scores, and historical payment behavior to identify patterns and correlations associated with loan defaults.
- Our analysis includes exploratory data analysis (EDA) to gain insights into the dataset, feature engineering to preprocess and transform the data
- The findings of this study provide valuable insights for lending institutions to assess credit risk, make informed lending decisions, and mitigate potential losses due to loan defaults.

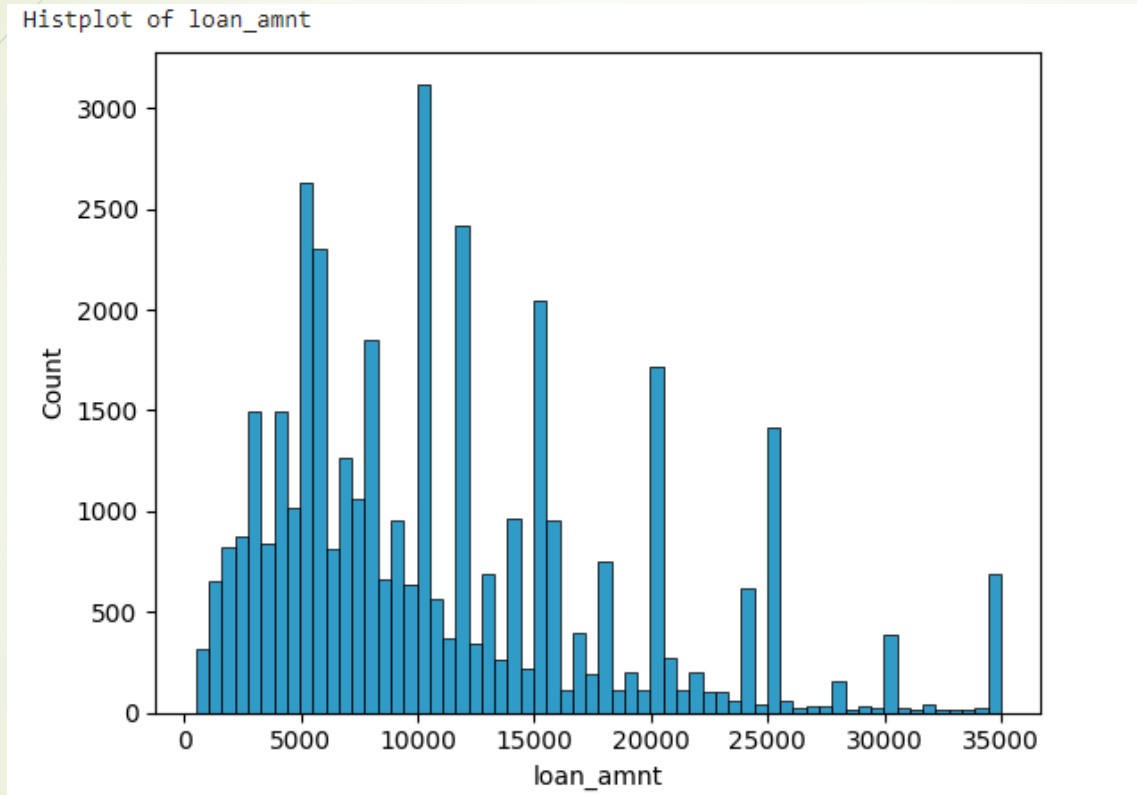


Problem-solving methods



- Data analysis often involve a structured approach that includes various steps, such as data cleaning, univariate analysis, bivariate analysis, and multivariate analysis.
- Data cleaning involves identifying and correcting errors, inconsistencies, and missing values in the dataset. This step ensures that the data is accurate, complete, and suitable for analysis.
- Univariate analysis focuses on examining one variable at a time to understand its distribution, central tendency, variability, and other properties.
- Bivariate analysis explores the relationship between two variables in the dataset. It examines how changes in one variable are associated with changes in another variable.
- Multivariate analysis extends the analysis to multiple variables simultaneously. It explores the complex relationships and interactions among multiple variables in the dataset.

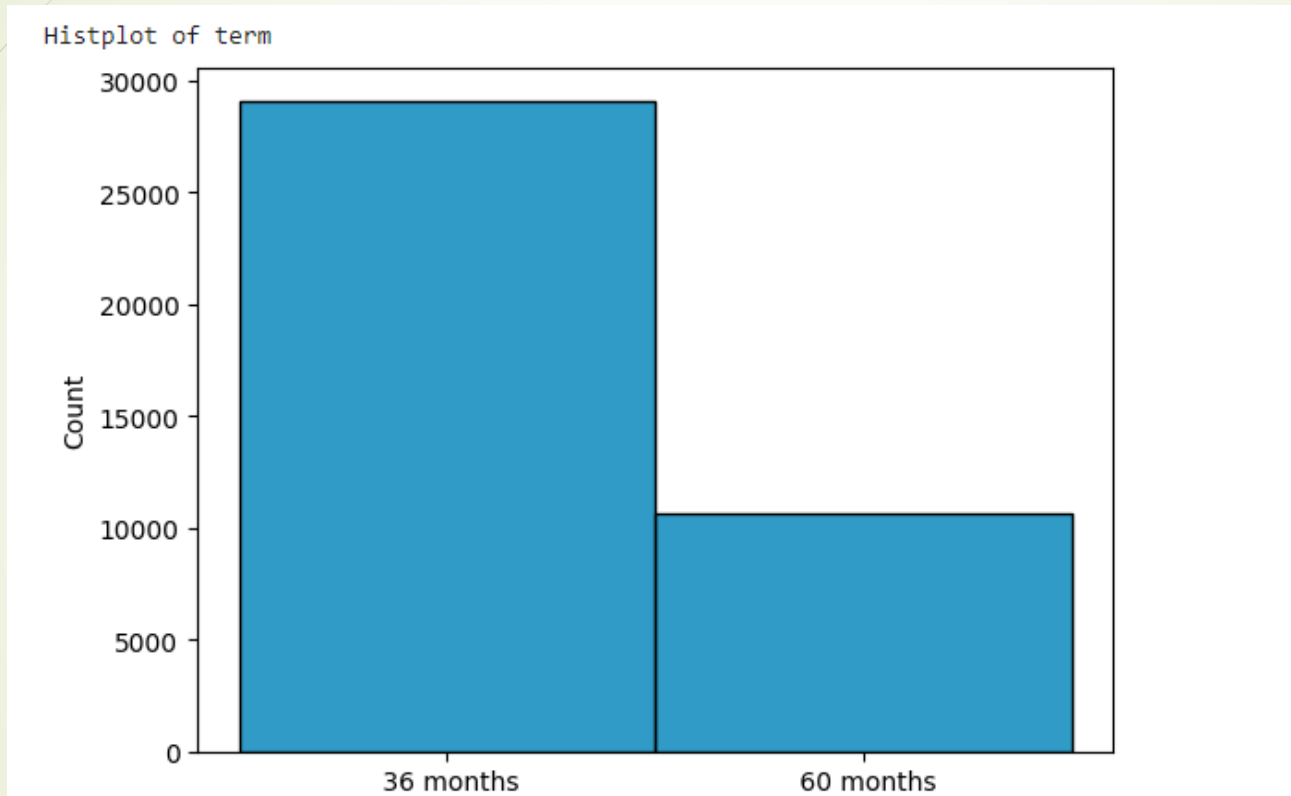
Univariate Analysis – Loan amount



Observation:

1. Most number of people have applied for loan for around 10k amount.

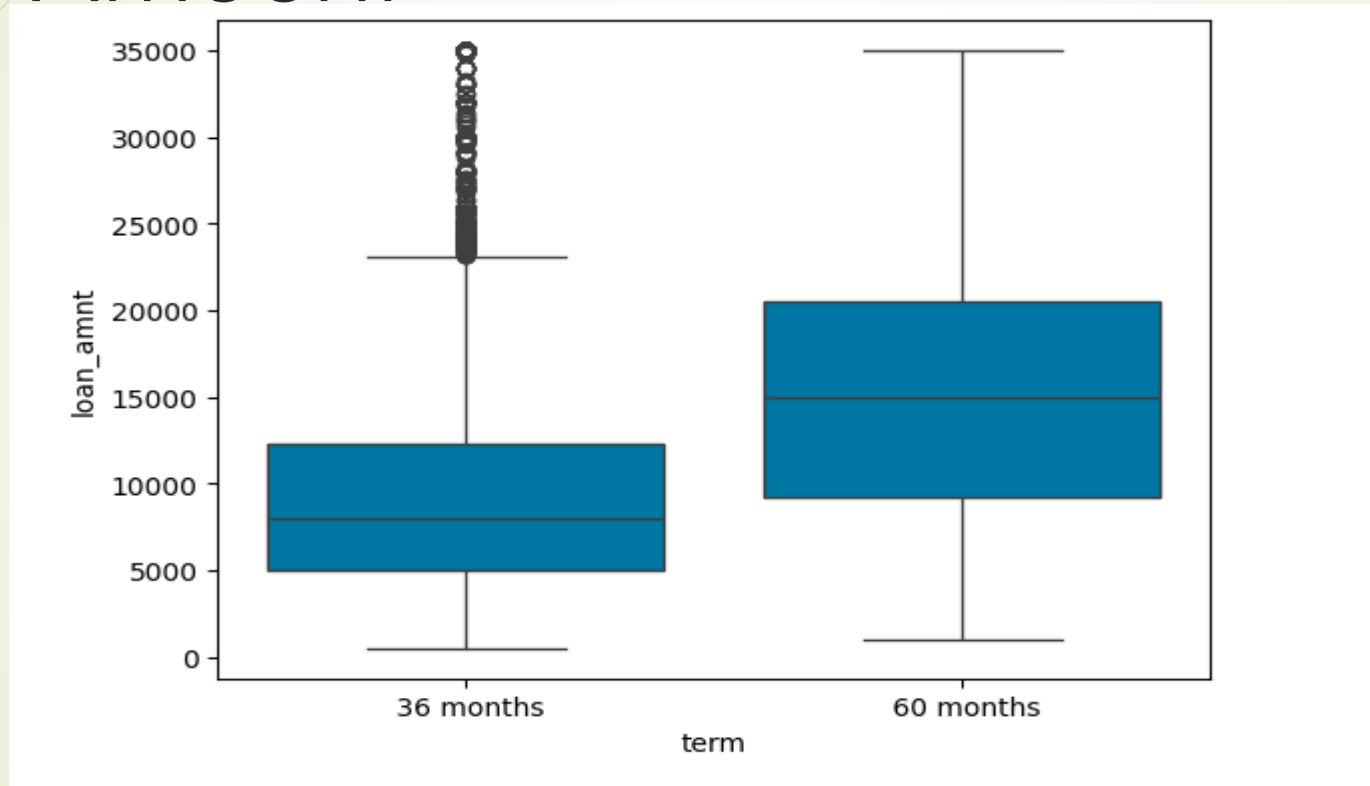
Univariate Analysis – Term



Observation:

1. Maximum number of people have opted for 36 month term.

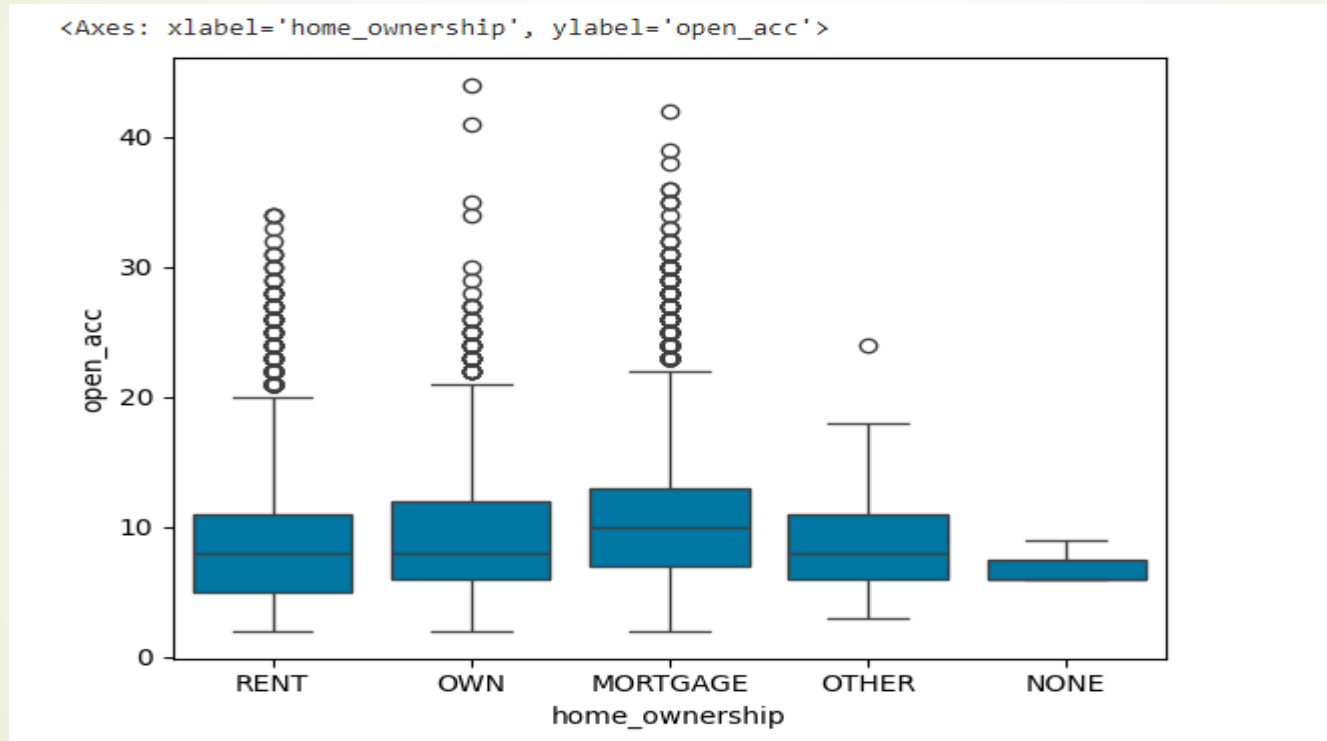
Bivariate Analysis – Term vs Loan Amount



Observation:

1. As the loan amount increases we see that people have opted for higher loan term i.e 60 months.

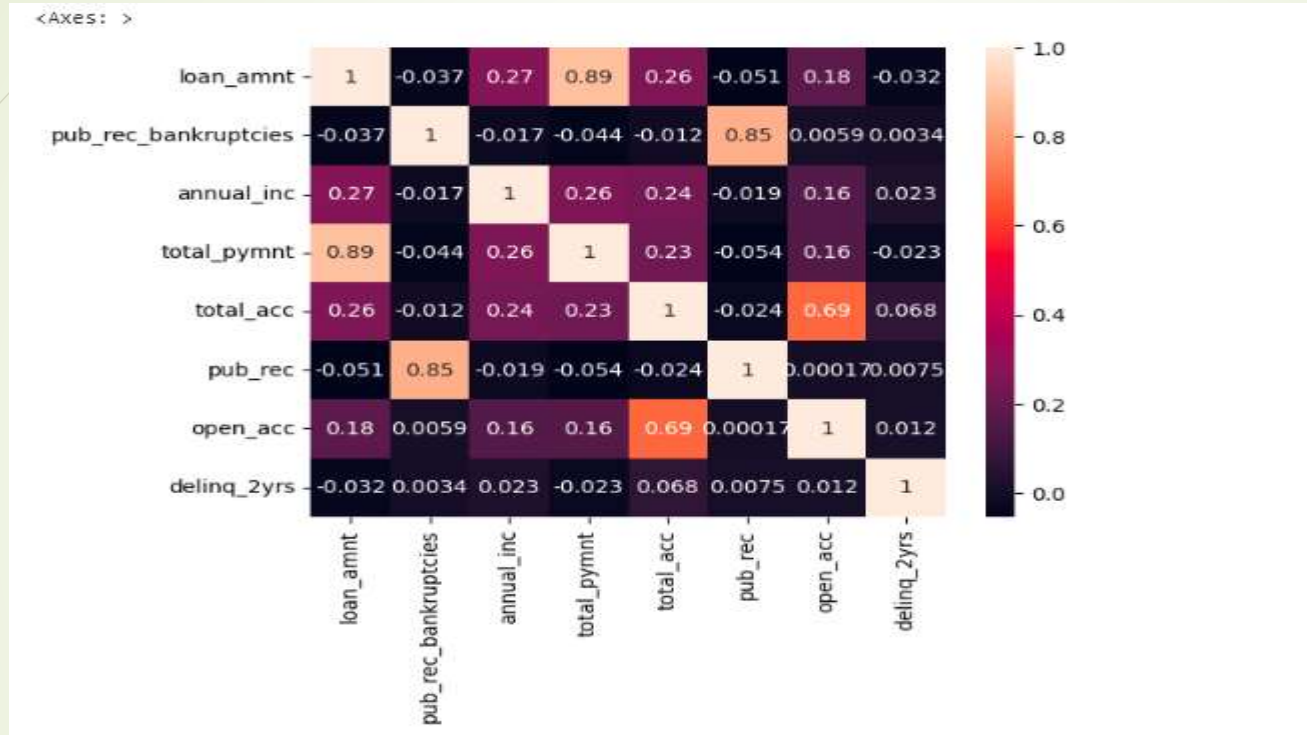
Bivariate Analysis – Home Ownership vs Open Credit Accounts



Observation:

1. Maximum number of open credit accounts are held by the people who have mortgages.

Multivariate Analysis



Observation:

1. We see that loan amount and total payment have high correlation, which is good for lending business.
2. We observe -ve correlation between annual income and public record so we can infer that people with high income have chances of degrading their public record.



Conclusion



- Lending club should reduce high interest rate 60 month term loans as they are more prone to default.
- Grades are good indicators of study the defaulters. Lending club should study low grades G-A before providing loan.
- Borrowers who already have mortgage are opting for higher loan there by increasing the risk of defaulting. Lending club must consider their existing mortgage before issuing loan.
- People with more public derogatory record have higher chance to default.
- Also we observe that the number of open credit line accounts are more with the people with mortgages and such people have higher derogatory public records. Lending club must consider these as well before disbursing the loan.