Lending Club Case Study

- Idhaya R and Yash Yennam

08-12-22

Outline

- Problem statement
- Decision making risks
- Analysis approach
- Data Handling
- Data Analysis
- Conclusions

Problem statement

- A consumer finance company specializes in lending various types of loans to urban customers . When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

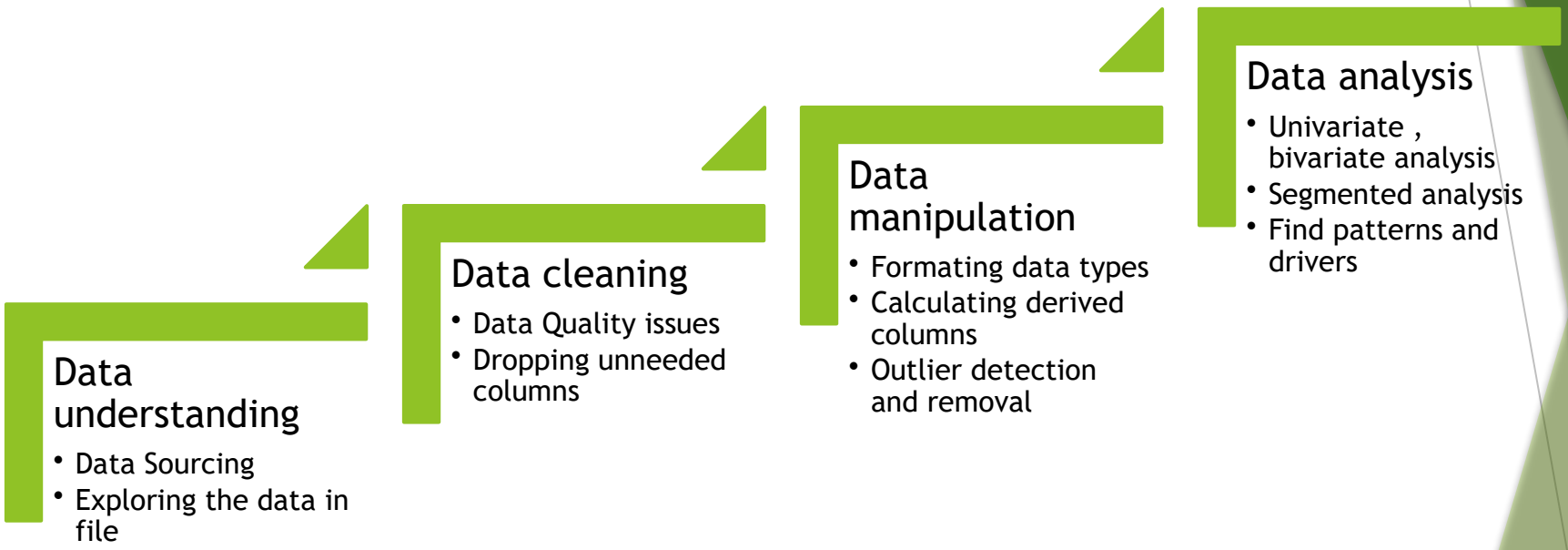
- The data given is given for past loan applicants whose loan was approved and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Decision making Risks

Two types of risks are associated with bank 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

Analysis approach



Data cleaning

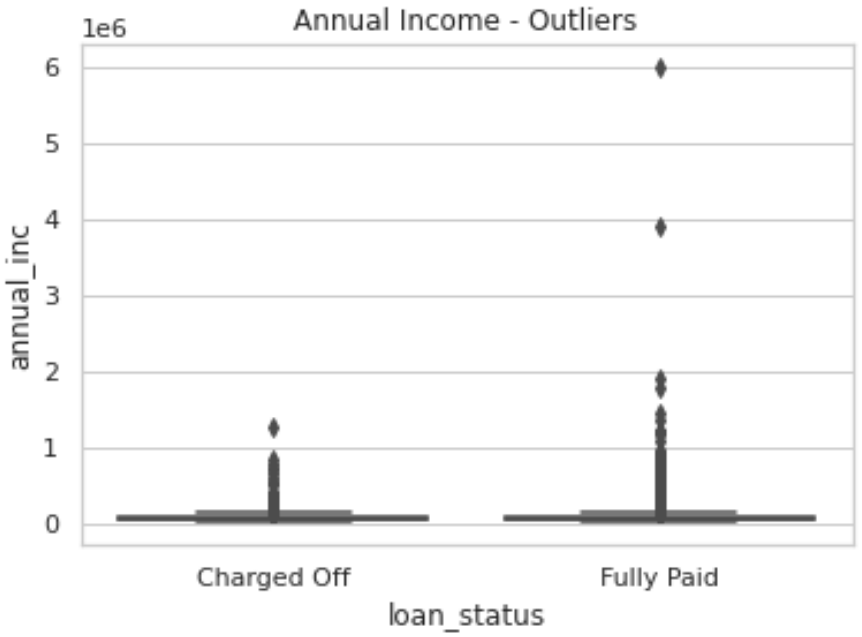
Addressing quality issues and outliers

How did we approach data cleaning and formatting?

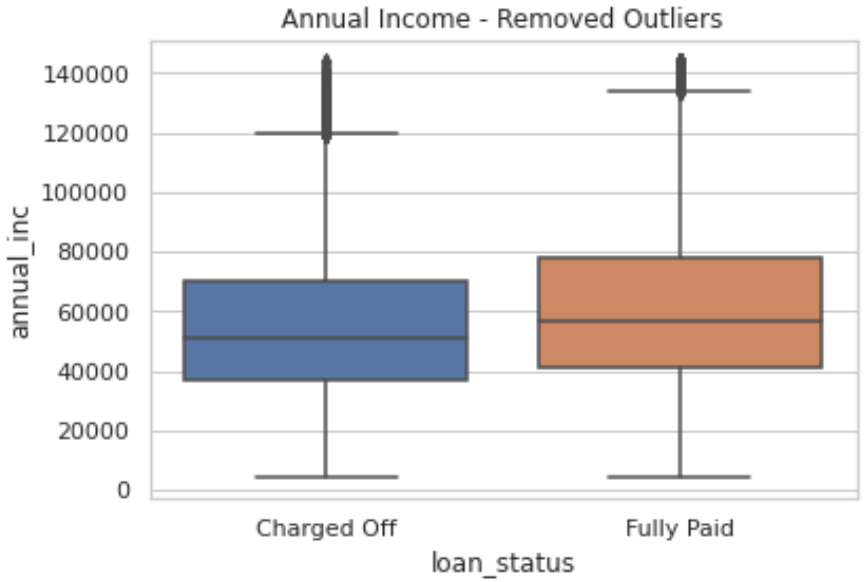
- Calculate missing percentage
- Drop columns where missing percentage > 50 and hold invalid or single values
- Drop duplicate rows
- Convert column data types according to data.
- Box plot to identify possible outliers (graphs in next slide).
  - Logic here?????????
- Create derived columns based on the columns and business understanding.

Outlier detection – Annual income

Before



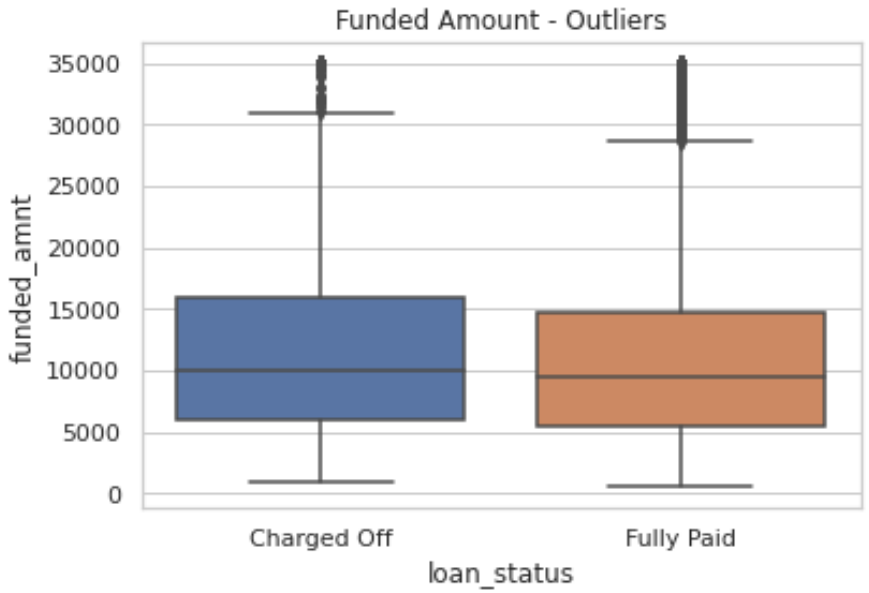
After



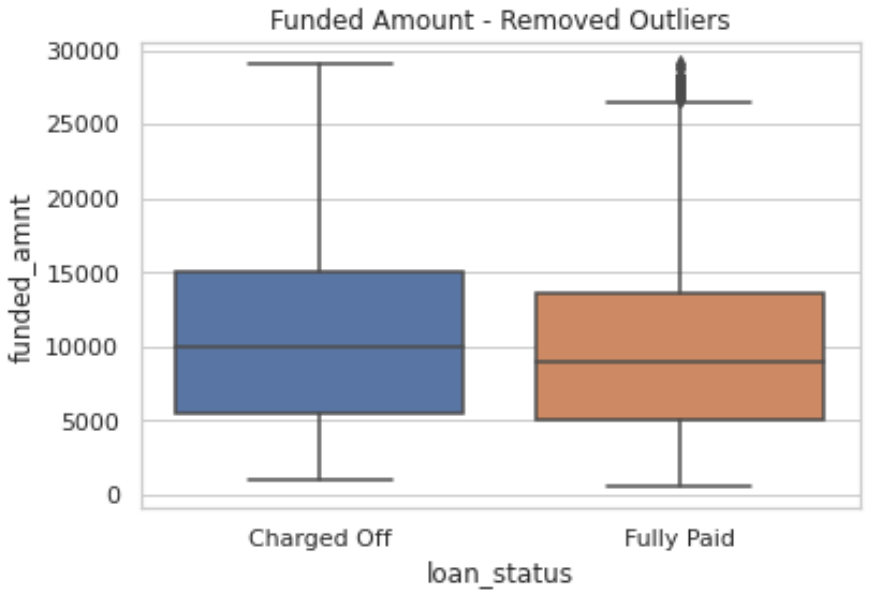
Outlier Detection – Funded amount

(Similar approach used for loan amount and funded amount inv to remove outliers)

Before



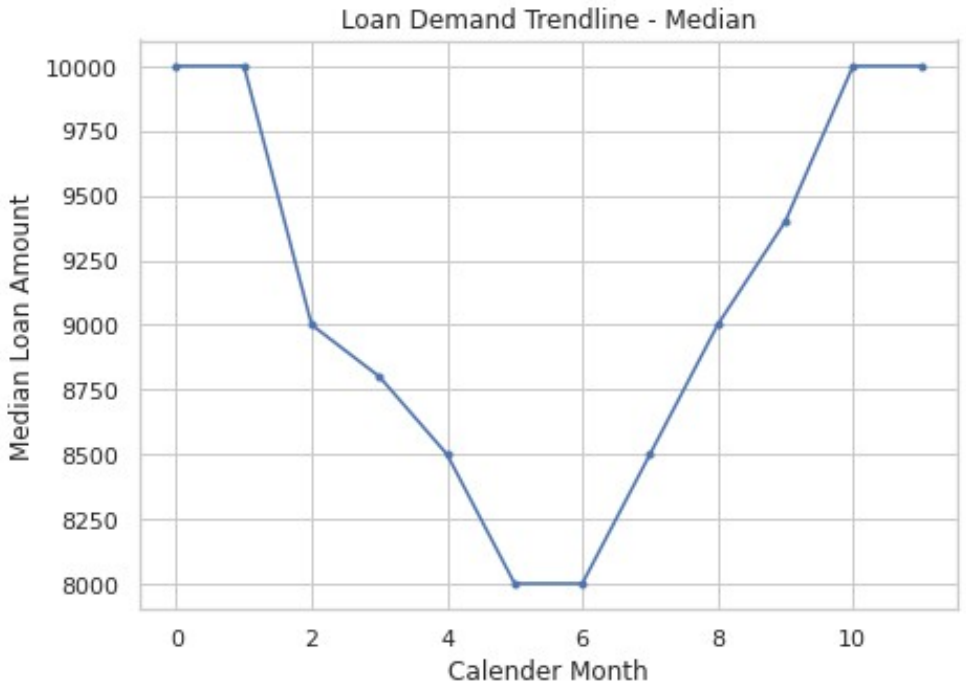
After



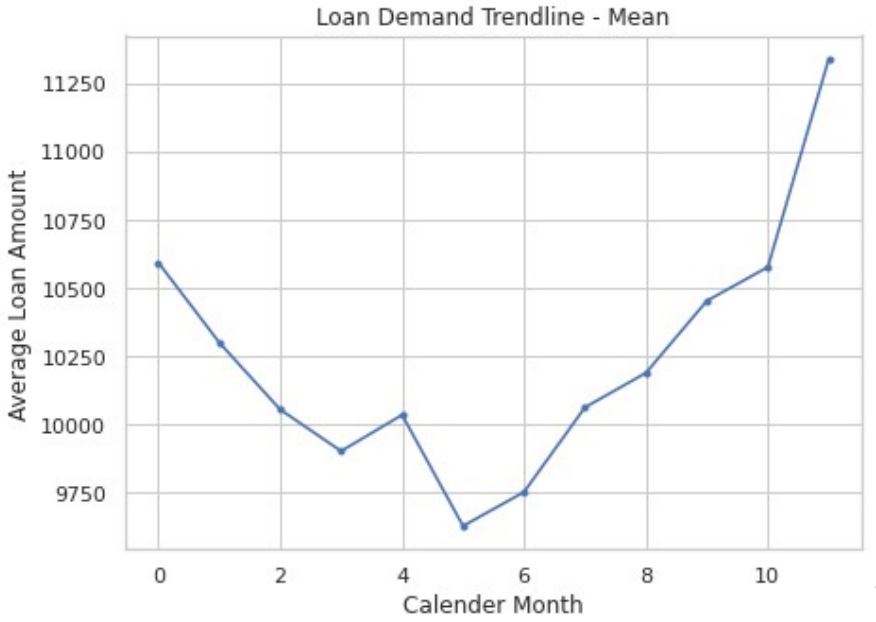
Univariate analysis

Conclusion : The demand is high during festive seasons during the month of December and around.

Loan Demand Trends



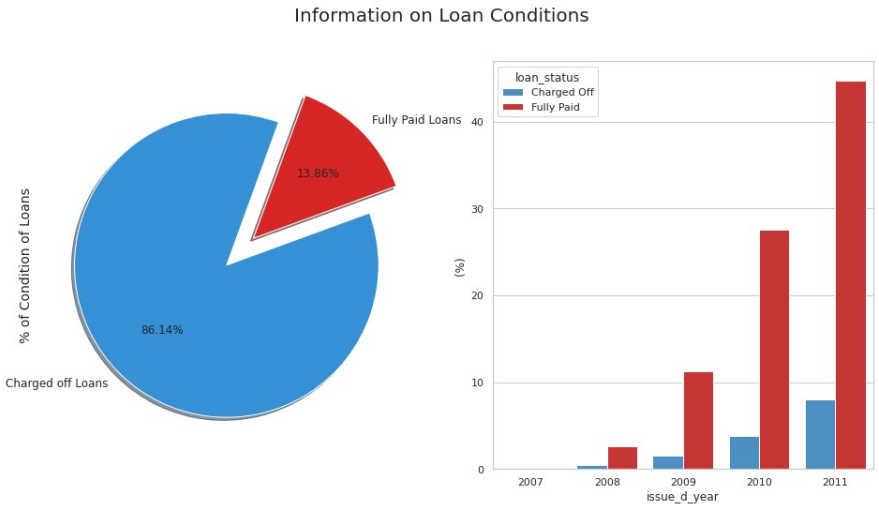
Loan Demand trends



Uni/Bivariate Analysis

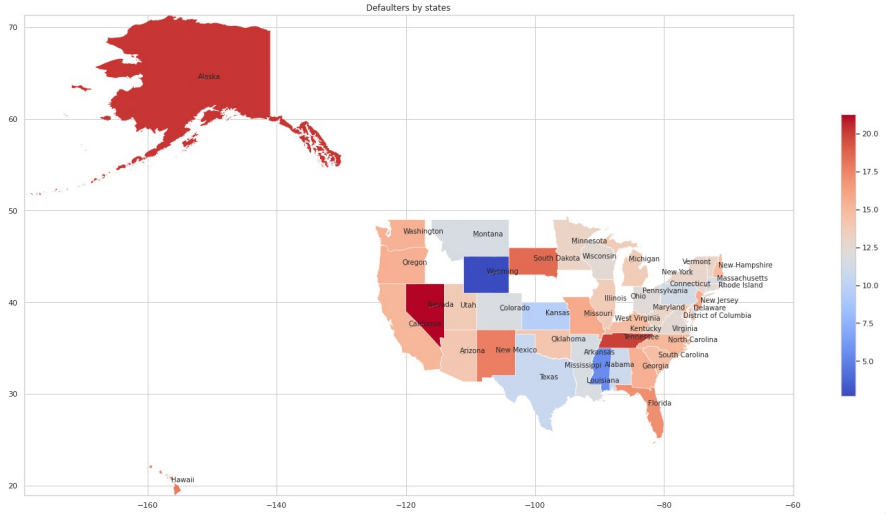
Percentage of defaulters by year

Conclusion: The percentage is below 16% , so the company is in safe zone.



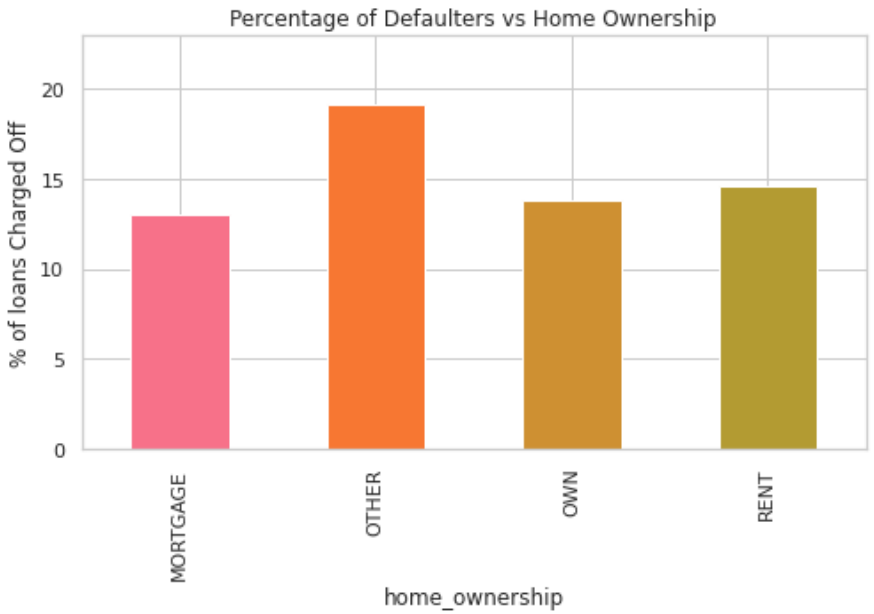
Percentage of defaulters by geography

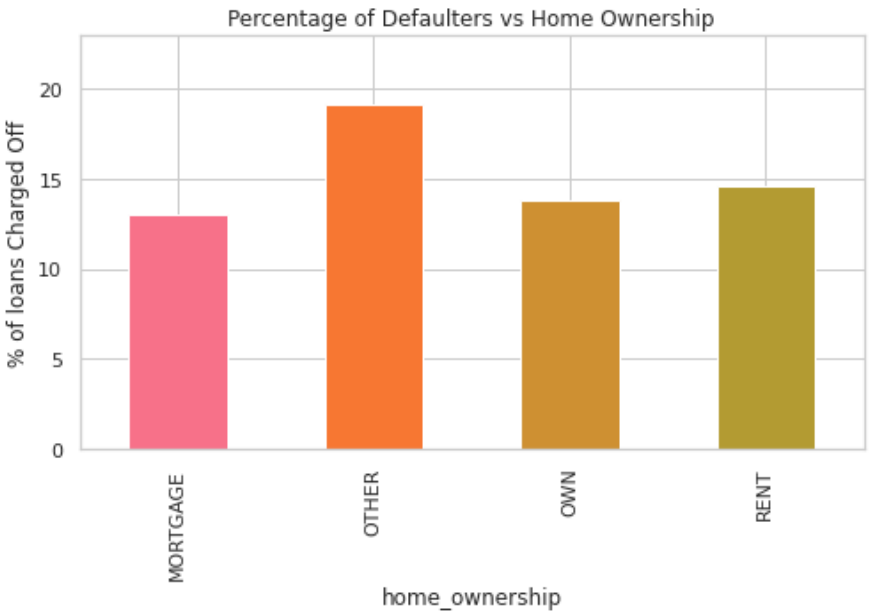
Conclusion: The states () have high defaulters.



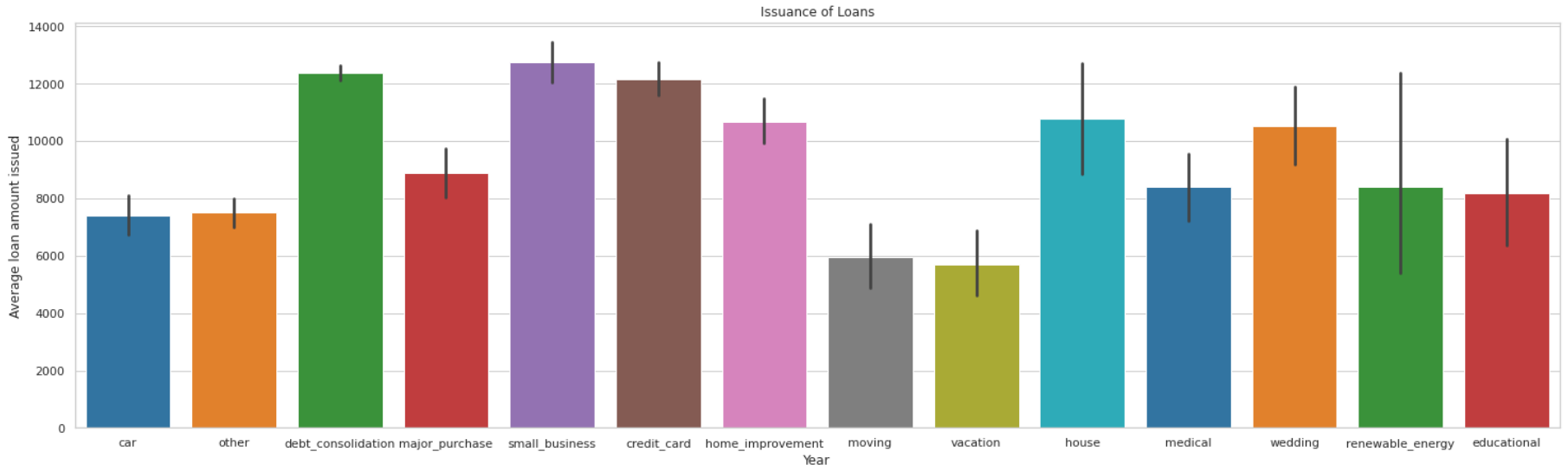
Bi Variate analysis

Conclusion: People who own the home is less likely to default.



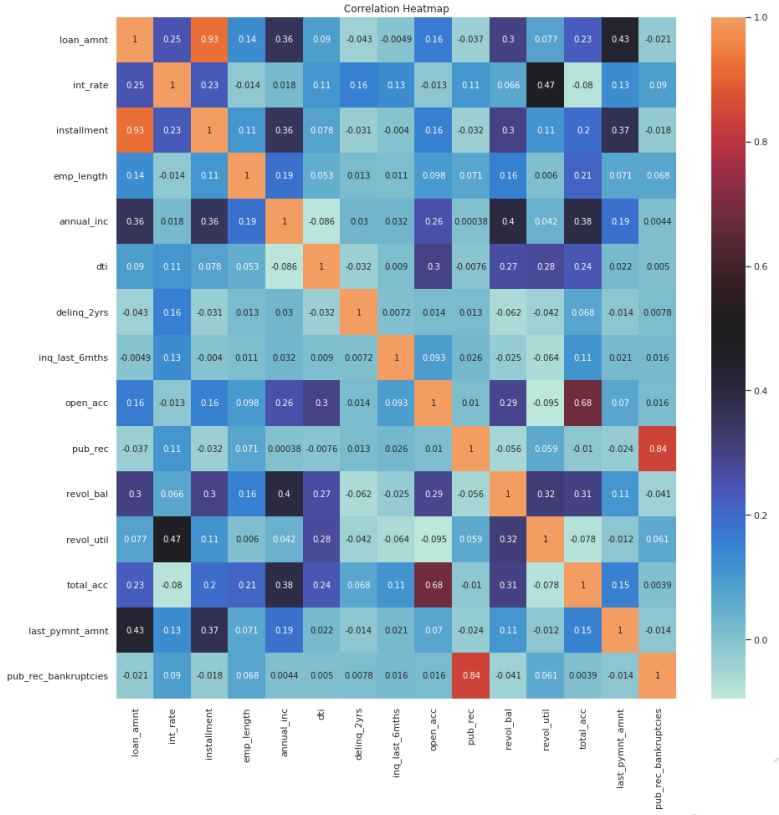


Bivariate analysis
Average loan amount vs issuance of loans



Correlation

- Correlation is used to define the similarity between two variables in the range of -1 to 1.
- Based on correlation and its inference if two or more variables have high correlation then it means we can take one variable and use it for analysis and ignore the remaining.
- After finding the variables, use business understanding to remove the variables.



Drivers of default

- Purpose
- Loan amount
- Home ownership?
- DTI?
- Term of loan?
- Interest rate?
- Last credit pull?