Lending Club Case Study

- Idhaya R and Yash Yennam

08-12-22

Outline

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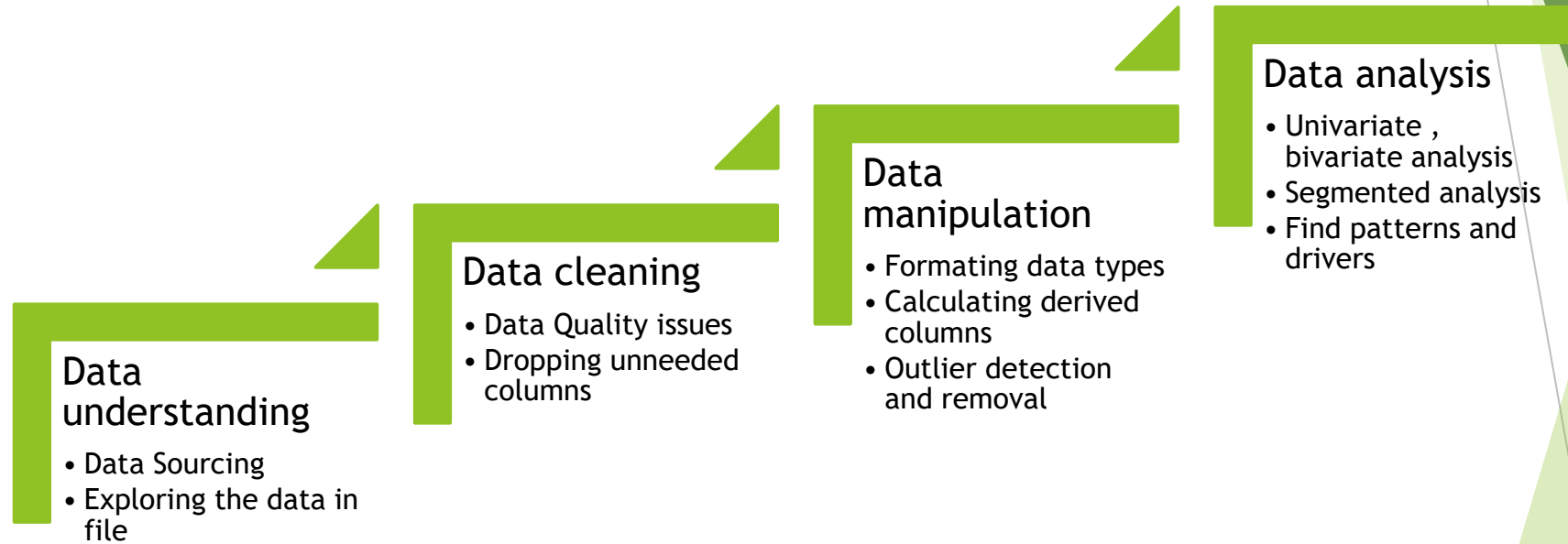- ❑ Conclusions

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
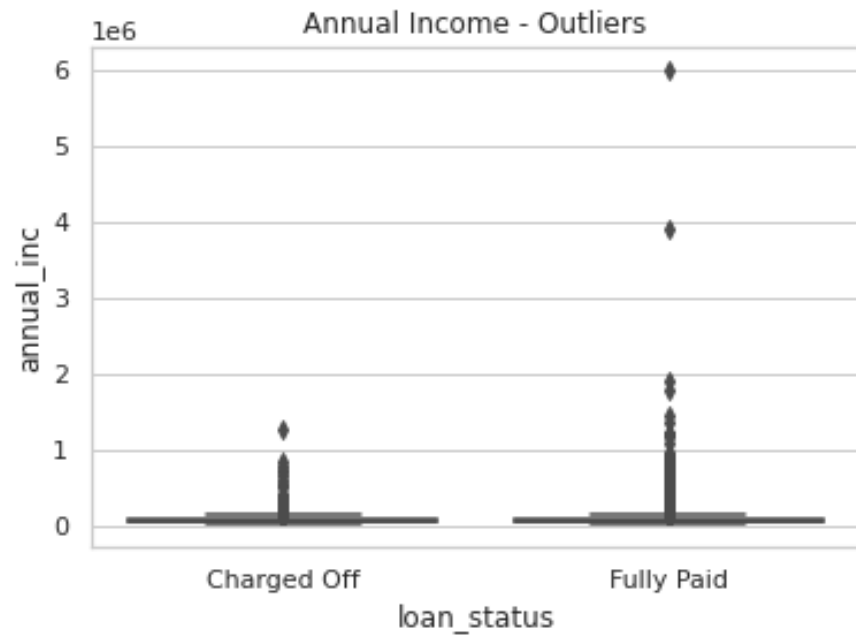
Analysis approach

Data understanding
- Data Sourcing
- Exploring the data in file

Data cleaning
- Data Quality issues
- Dropping unneeded columns

Data manipulation
- Formating data types
- Calculating derived columns
- Outlier detection and removal

Data analysis
- Univariate , bivariate analysis
- Segmented analysis
- Find patterns and drivers

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 columns that 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).

  - If outliers are present , then the rows between the range (25 percentile –1.5*IQR and 75 percentile + 1.5* IQR) is kept.

- Create derived columns based on the columns and business understanding.

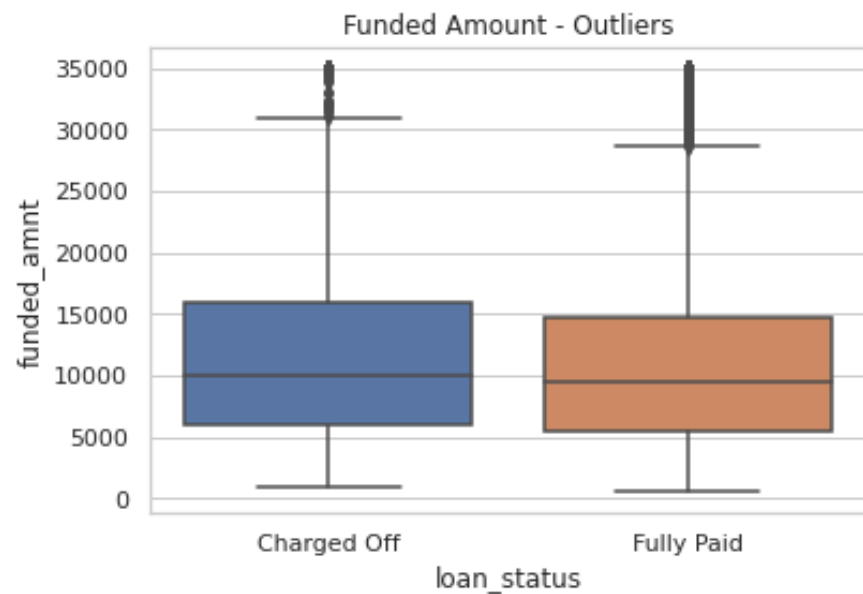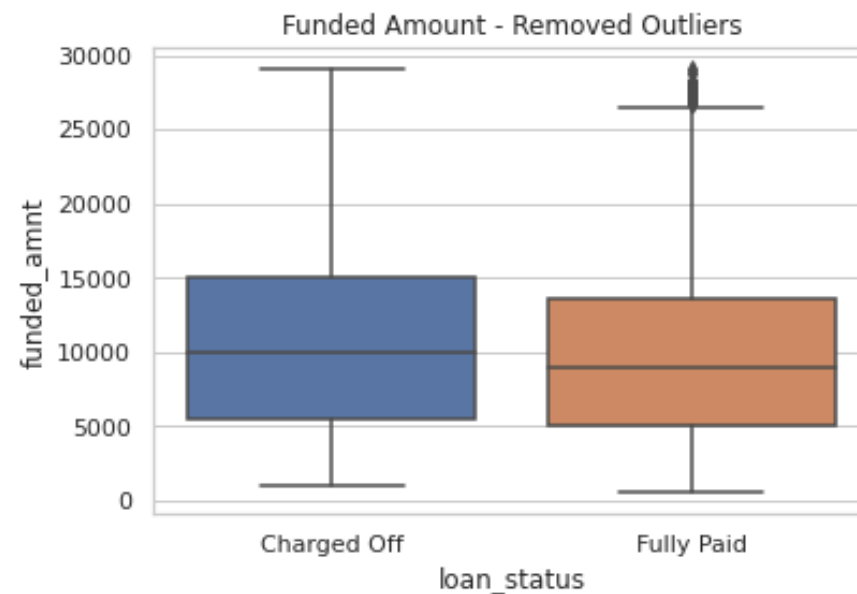# Outlier detection – Annual income

Before

After
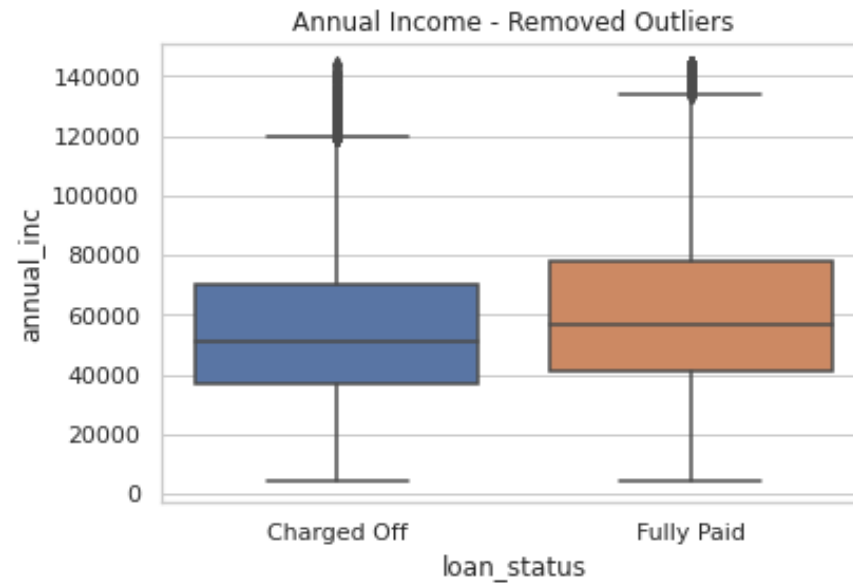
Outlier Detection – Funded amount
(Similar approach used for loan amount and funded amount inv to remove outliers)
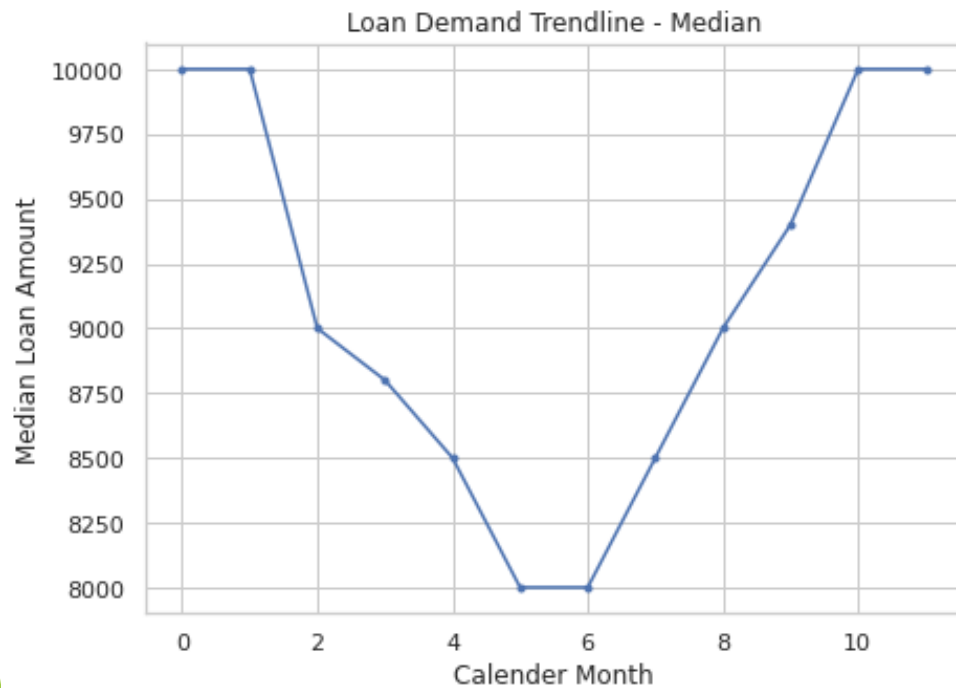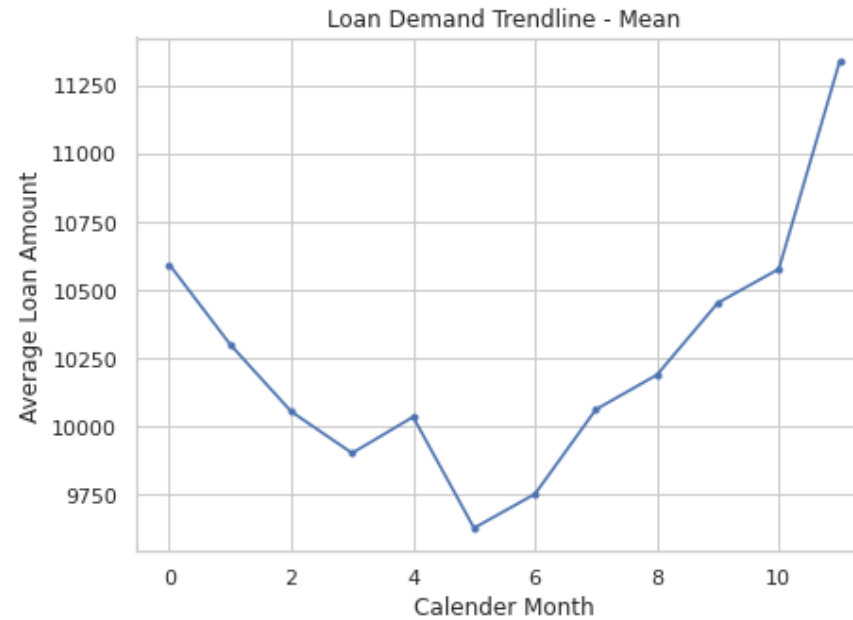
Before

After

Univariate analysis

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

Loan Demand Trends
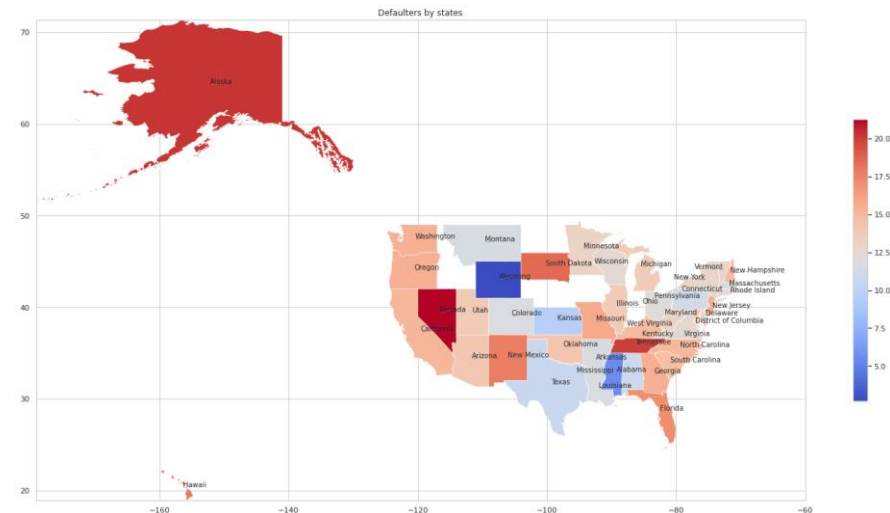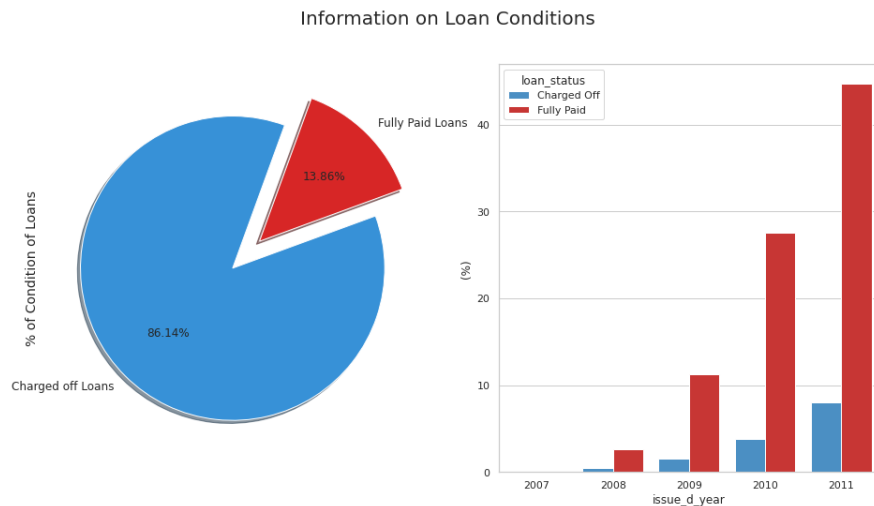

Loan Demand trends

Uni/Bivariate Analysis

Percentage of defaulters by year

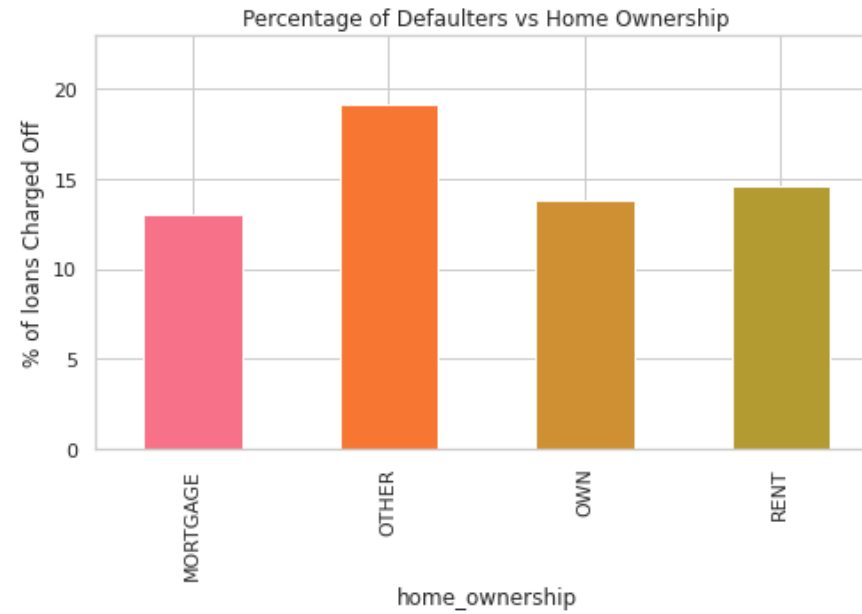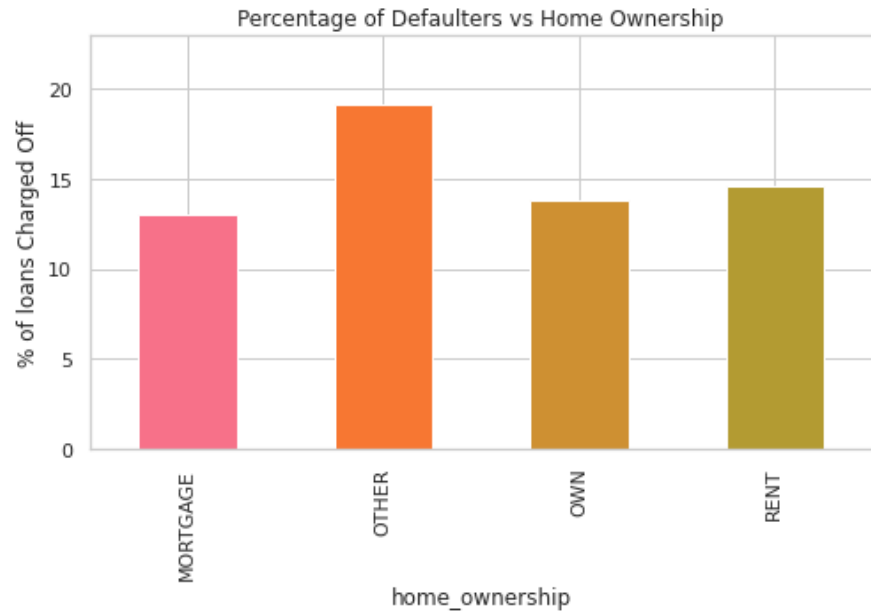*Conclusion: The percentage is below 16% , so the company is in safe zone.*

Percentage of defaulters by geography

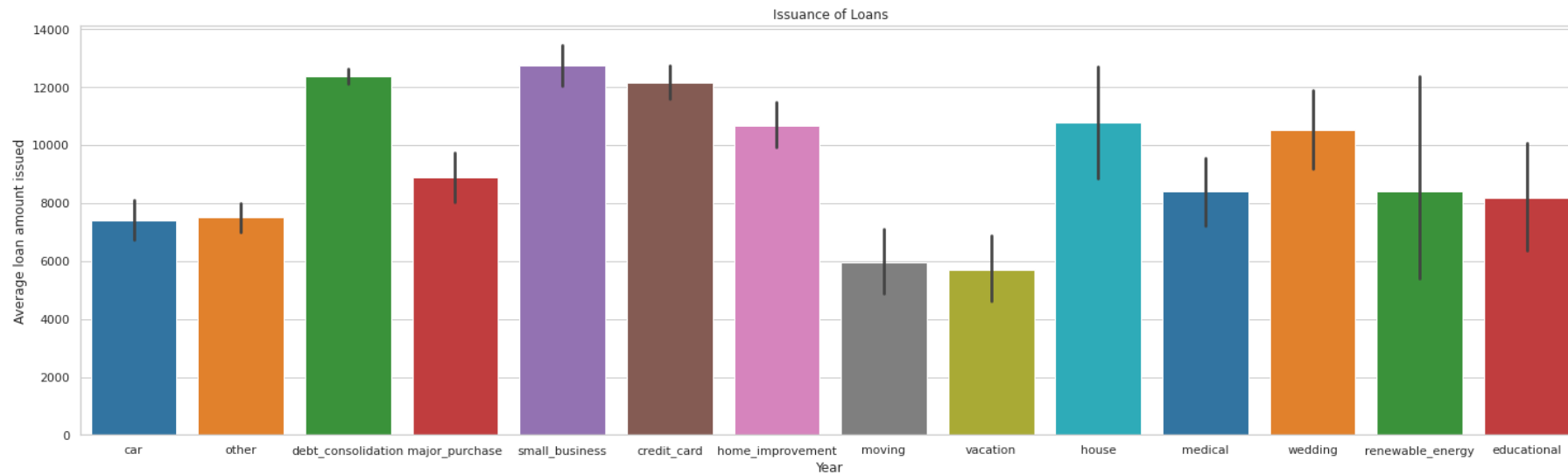*Conclusion: The states (Nevada, Alaska,Tenneessee,South Dakota, Hawaii) have high defaulters.*

Bi Variate analysis
*Conclusion: People who own the home is less likely to default.*

Bivariate analysis
Average loan amount vs issuance of loans

Correlation

Conclusion: There is correlation between Installments and Loan Amount, which is obvious in this scenario since higher the Loan amount higher will be the installments. **This is a perfect example of *Correlation does not imply causation***

- ▶ 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.

Correlation Heatmap
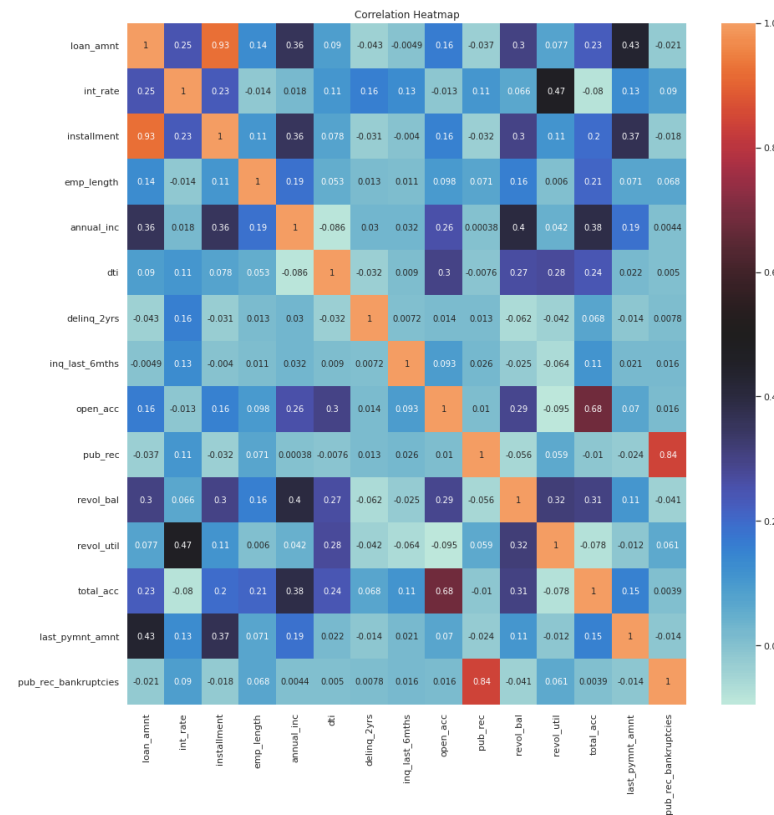
Conclusions from analysis

- Driver 1 : Default percentage

  - Lending club performance is increasing by every year, with higher number of fully paid loans by every passing year.

  - Approx. 14% of lenders have not paid back the loan

- Driver 2: Seasonality

  - The trend varies based on the month, it is higher in winters compared to summers.

  - Recommendation: Can increase offers during summer to attract more volumes in months with less demand.

- Driver 3: States

  - Top 5 States with maximum Charged off Rates : Nevada,Alaska,Tennessee,South Dakota,Hawaii

  - Top 5 States with Minimum Charged off Rates : Wyoming,Mississippi,District of Columbia,Kansas,Delaware

  - Recommendation: Make a stringent verification/Higher Interest rates to States with higher default rates and more offers/Discounts to states with lower charged off rates

Conclusions from analysis

- Driver 4: Purpose / Loan category

    - Small Business tends to make largest category of loan which are charged off

    - Vacation is tends to make least category of loan which are charged off.

- Driver 5: Public record bankruptcies

    - When there is more record of bankruptcies, more likely the loan will be defaulted.

- Driver 6: Percentage of defaulters vs home ownership

    - When the customer owns/rents home he is less likely to default.

Thank you!