Lending Club Case Study

GROUP MEMBERS:

ISHITA BHARGAVA

JATIN SINGH

Business Objective

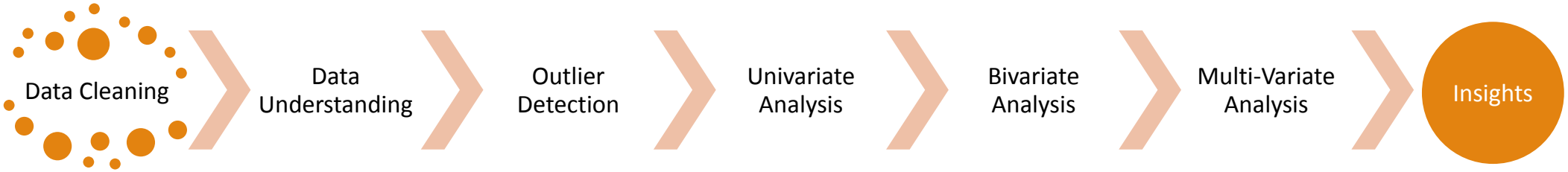
1. Lending Club is Consumer Finance Company which provide loans to different segment of urban society according to their needs and wants.

2. When company receive the loan application, it goes through screening to check if loan can be given to customer or not based on applicant's profile.

3. Two types of risk are associated with it to make decision:
   a. If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company.
   b. 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.

4. The company aims to identify the key factors influencing customer behavior to determine which customers are at risk of defaulting and which are likely to remain loyal.

Exploratory Data Analysis



Data Cleaning

1. We have Removed 57 Columns as they have more 90% missing values.
2. We have Dropped ‘Desc’, ‘Title’ column because title is a subcategory of ‘purpose’ and ‘Desc’ is the description of ‘Title’ in detail. We decided to keep ‘Purpose’ for our analysis and drop ‘Desc’ and ‘Title’.
3. We also removed columns which contains only single values or more than 80% row values are same.
4. Imputing missing values according to column category i.e. Either Categorical or Numerical.
5. Standardizing quantitative values in the dataset and correcting the data types for date related columns.
6. Extracted Year columns as derived metrics to understand the dataset.

Data Understanding

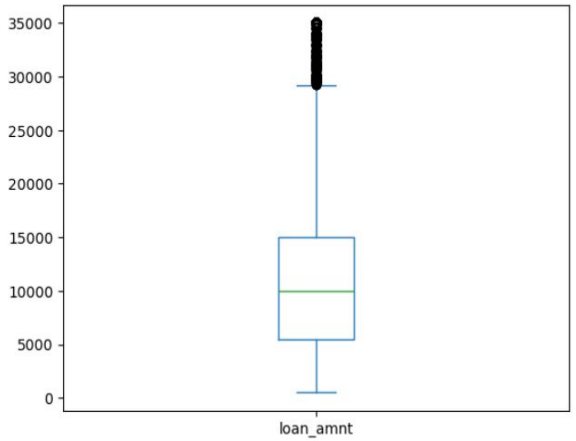
Categorical
• term
• grade
• sub_grade
• emp_*length*
• home_ownership
• verification_status
• issue_d
• loan_status
• purpose
• zip_code
• addr_state
• earliest_cr_line
• inq_last_6mths
• last_pymnt_d
• last_credit_pull_d

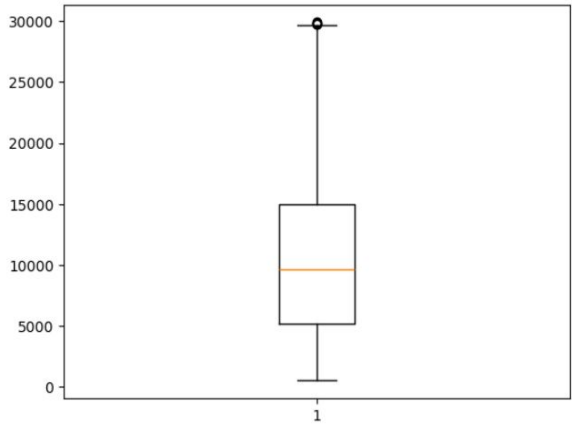
Numerical
• Loan_amnt
• funded_amnt
• int_rate
• installment
• annual_inc
• dti
• revol_bal
• revol_util
• total_acc
• total_pymnt
• total_rec_prncp
• total_rec_int
• total_rec_late_fee
• last_pymnt_amnt

Others
• id
• member_id
• open_acc

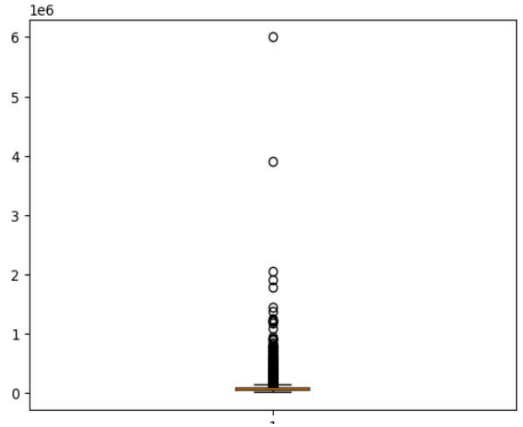
Outlier Detection

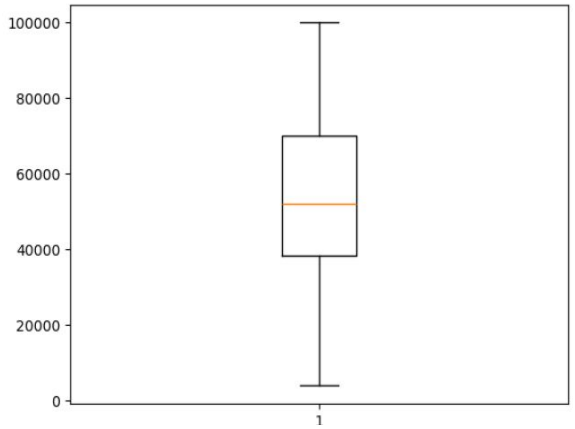
Checking Outliers in ‘loan amount’ using boxplot and taking loan amount upto 30000 only.





Removing Outlier from ‘annual_inc’ columns and keeping only till 100,000.

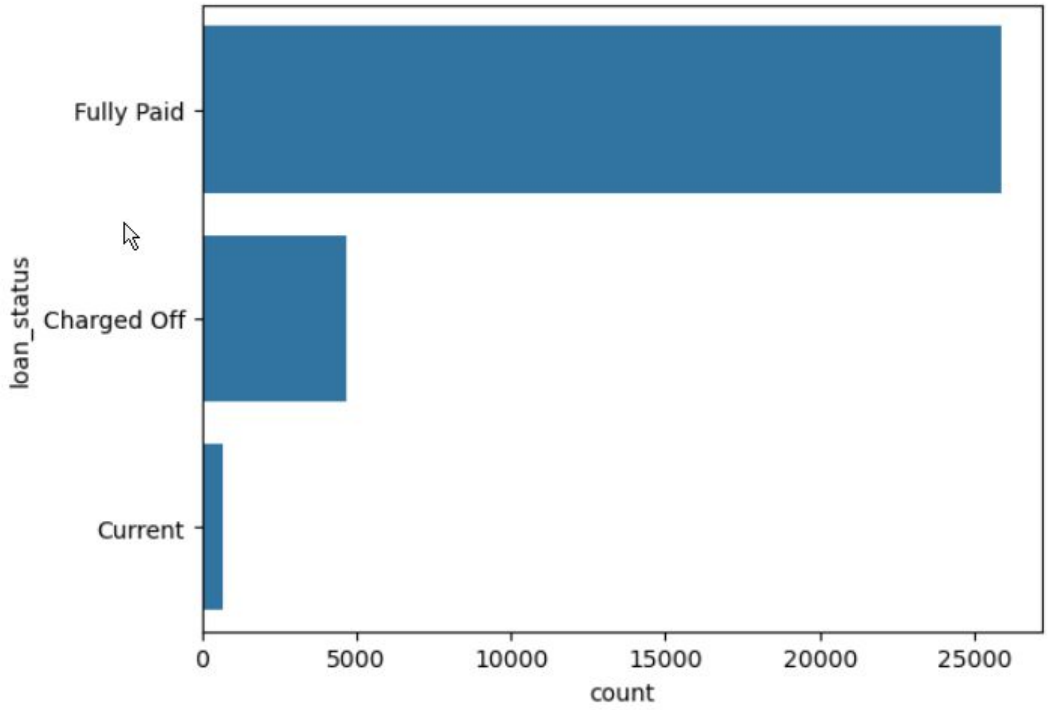




1. Using Box plot and countplot we have removed outliers from other major column as well like ‘funded_amnt’, ‘int_rate’ etc.
2. Removing unnecessary columns also helped to remove outliers from dataset.

Univariate Analysis

Using countplot on ‘loan_status’ shows that majorly customers are divided into 3 subcategories in dataset. Using ‘loan_status’ as anchor columns we have further drill downed our analysis for other columns.

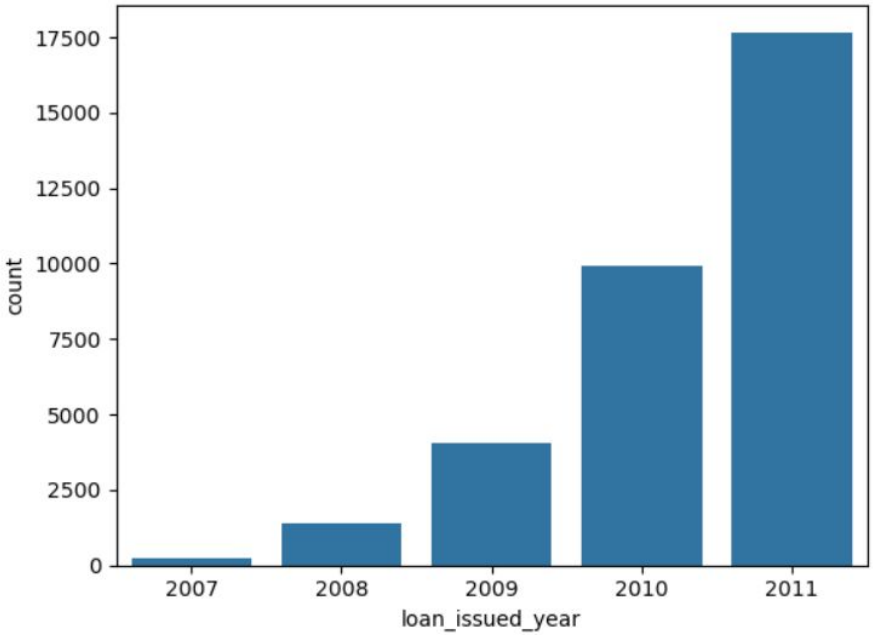


Segregating data based on loan_status into fully paid, charged off, current

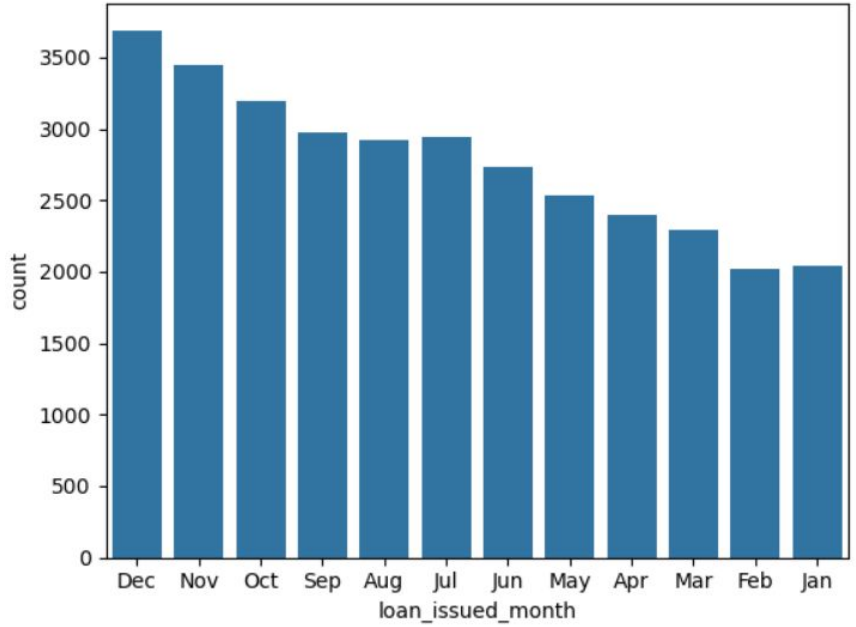
```
data_fullypaid=df[df['loan_status']=='Fully Paid']
data_chargedoff=df[df['loan_status']=='Charged Off']
data_current=df[df['loan_status']=='Current']
```

Total count of ‘Loan status’ in the dataset

Contd…



count of loans vs ‘loan_issued_year’

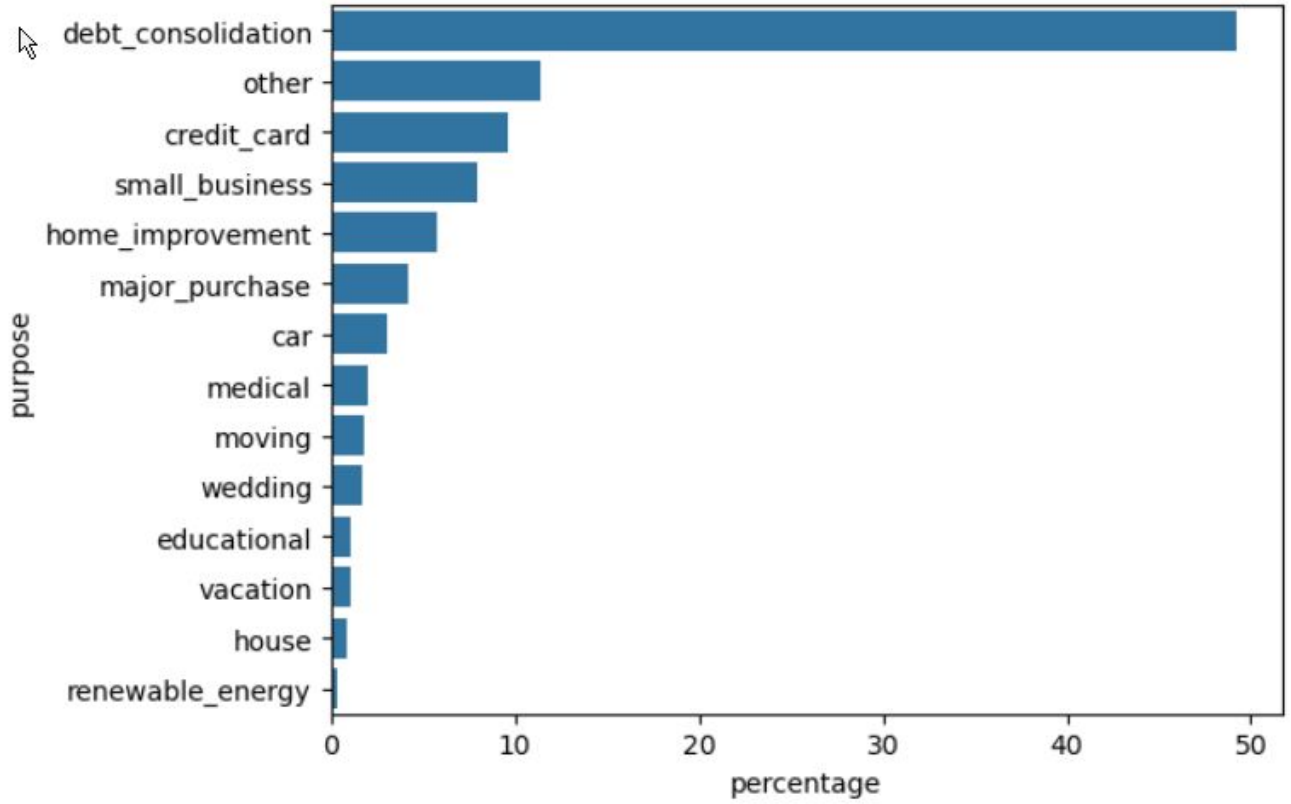


count of loans vs ‘loan_issued_month’

Number of customer taking loans are increasing significantly every year.

Customers are taking more loans during Nov,Dec due to festive season.

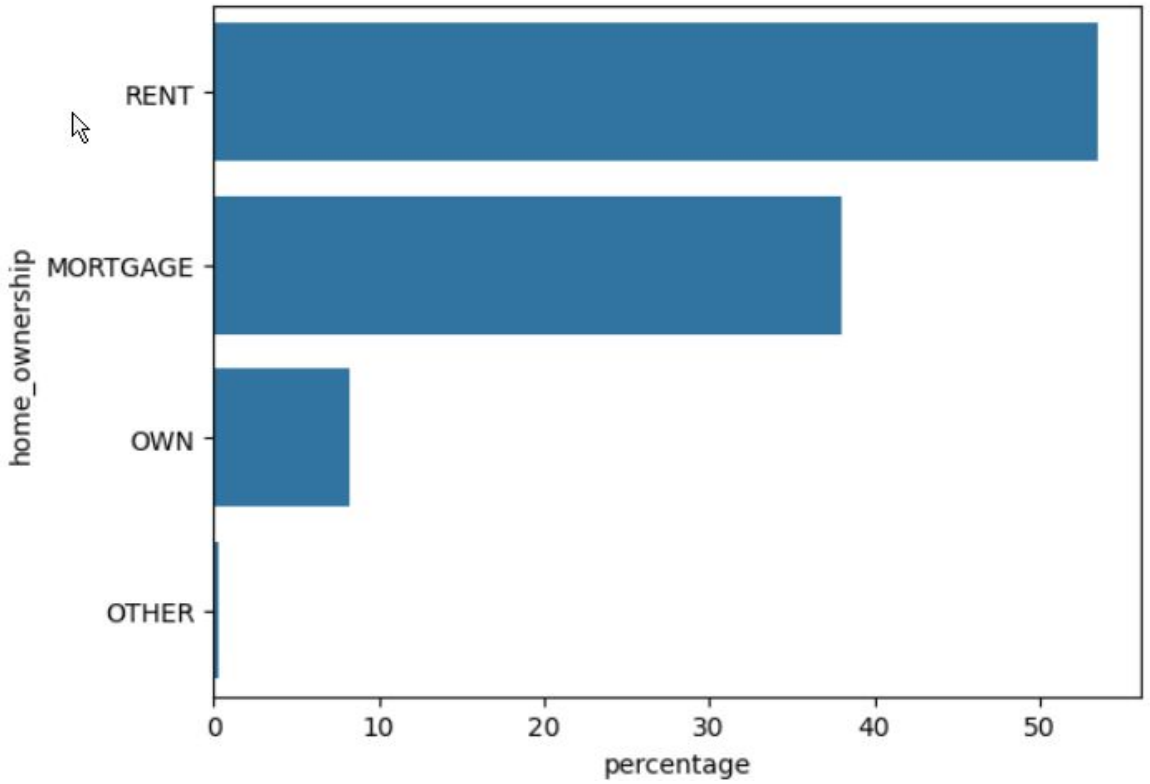
Contd…



‘Purpose’ of Charged-off borrowers

Approximately 48% customer who have taken loan for debt consolidation are mostly defaulted.

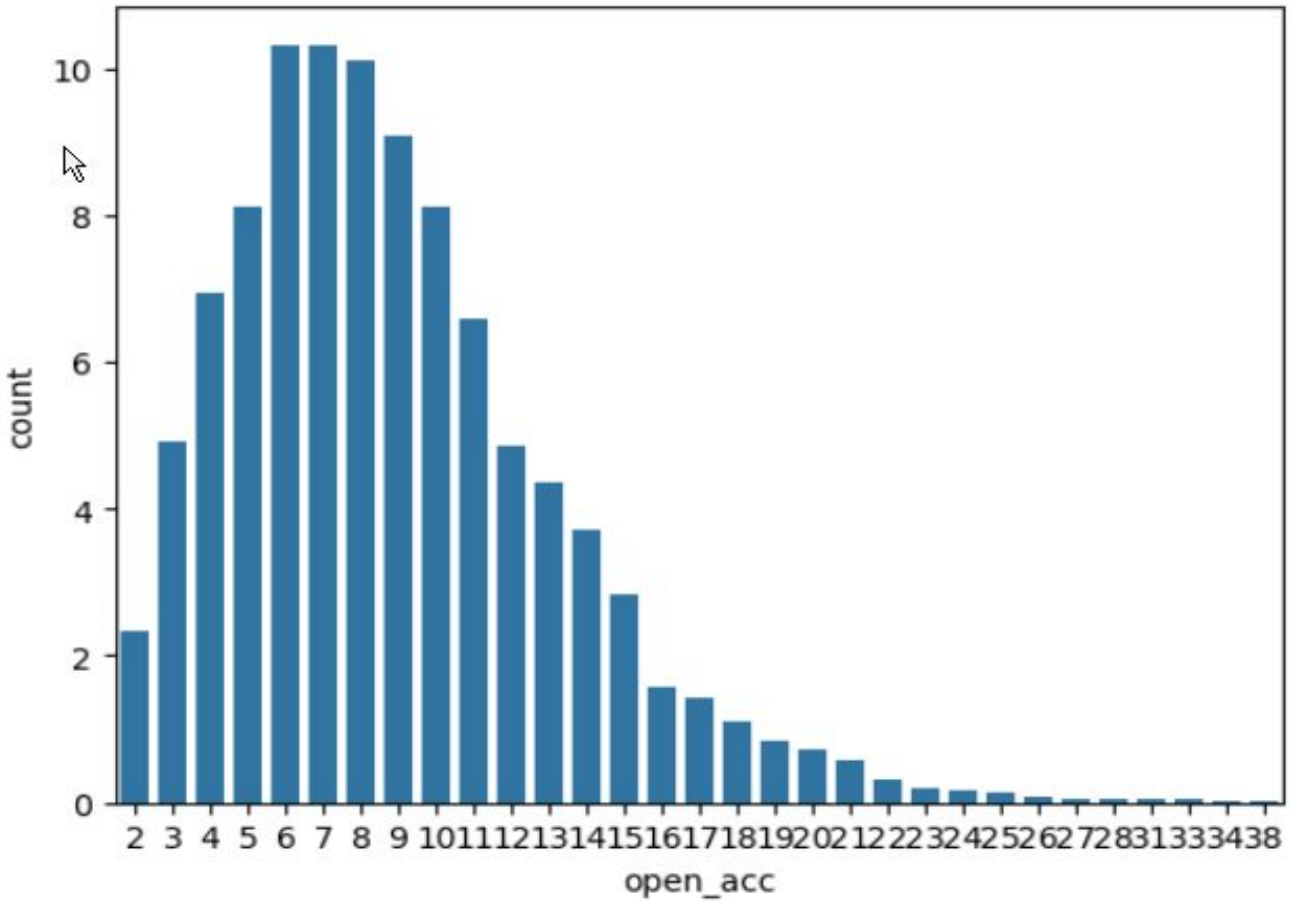
Contd…



'Home_onwership' of Charged-off borrowers

Customers living on rent are more likely to default than the customer having mortgage or own their own house.

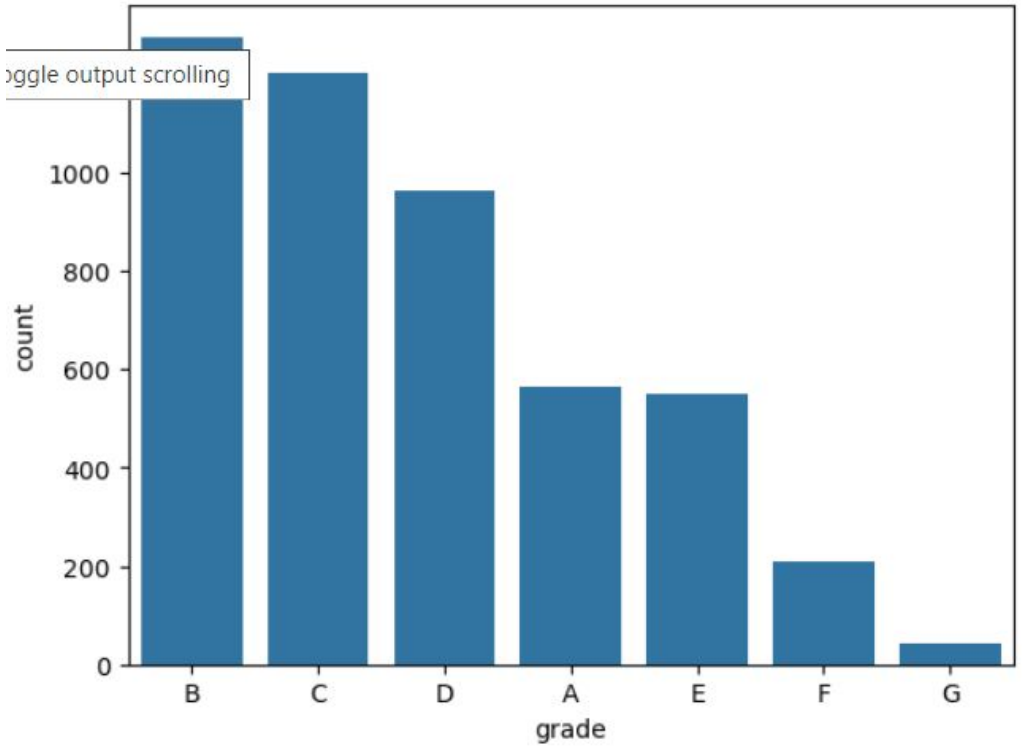
Contd…



‘Open_acc’ on charged-off borrower

We had made an assumption that Defaulters will mostly have more open loans but Data is rightly skewed which shows that defaulter have open loans in range of 5-10.

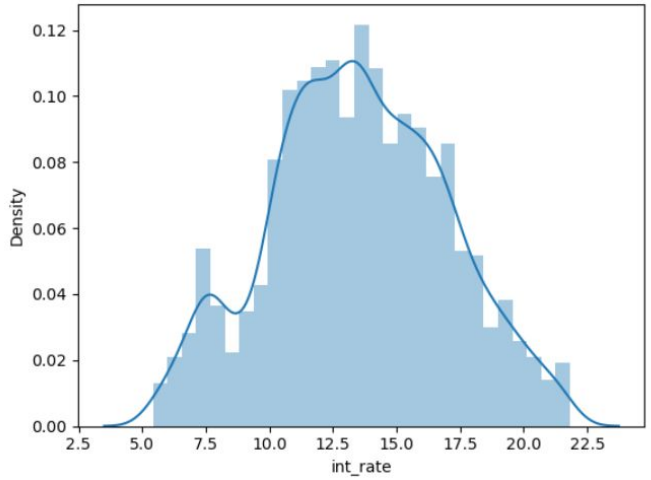
Contd…



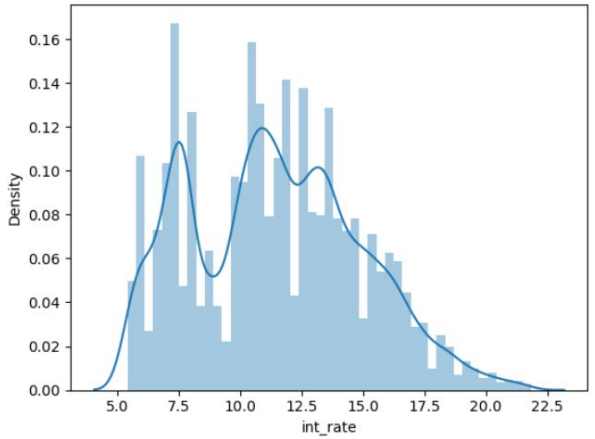
'Grade' count on Charged-off borrowers

Customer who have been given a grade of B, C, D by LC are most likely to default.

Contd…



Interest Rate on
Charged-off borrowers

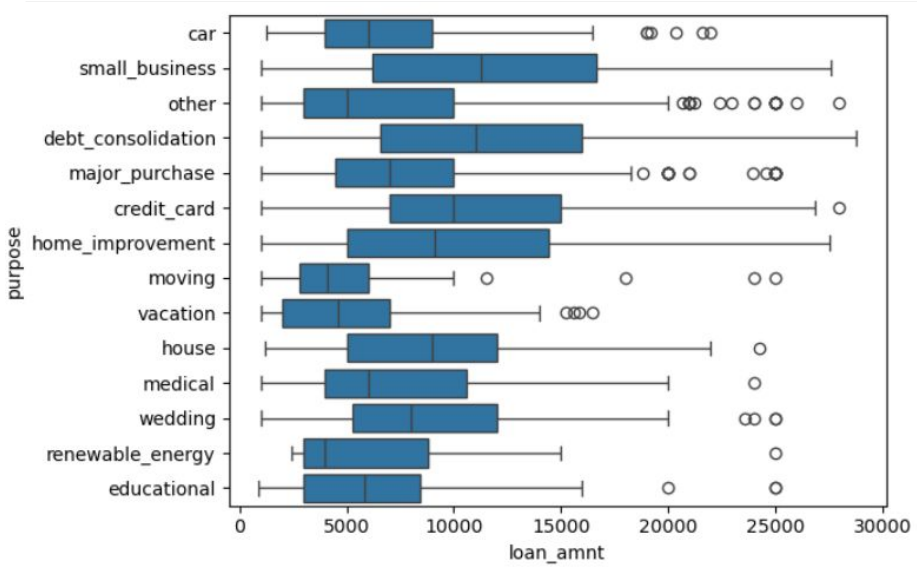


Interest rate on
Fully paid borrowers

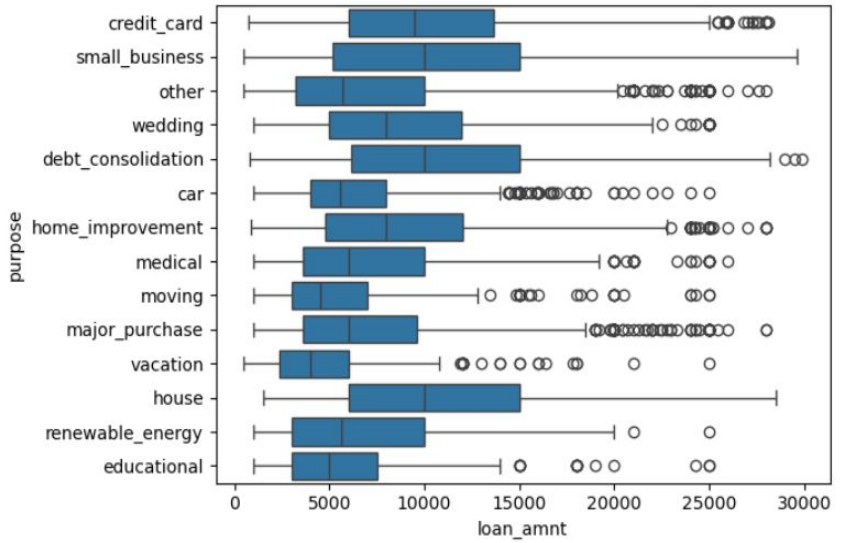
Interest rate for defaulted borrower is ranges around 12%-16%

Interest rate for non-defaulters ranges around 6%-8% and 12.5%-15%. This means borrowers are defaulting more who have high interest rate generally.

Bivariate Analysis



'Purpose' vs 'loan_amnt'
On charged-off borrowers

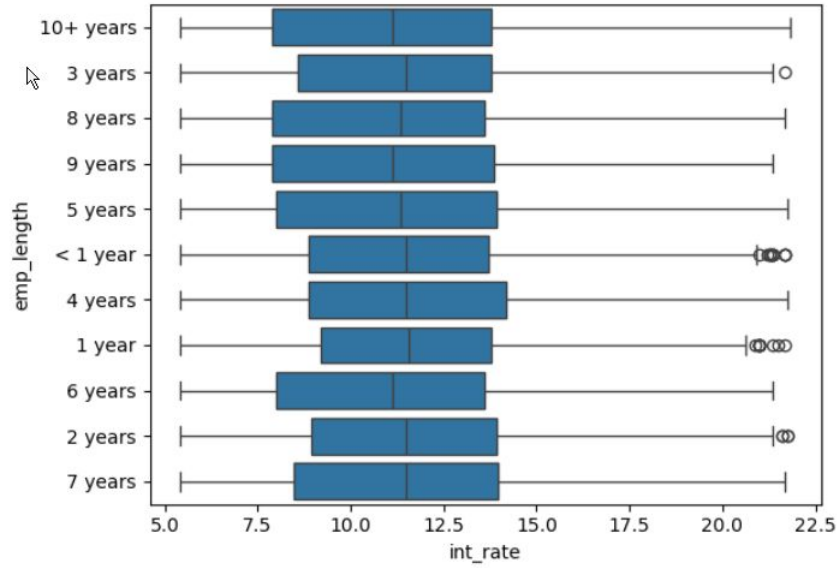


'Purpose' vs 'loan_amnt'
On fully-paid borrowers

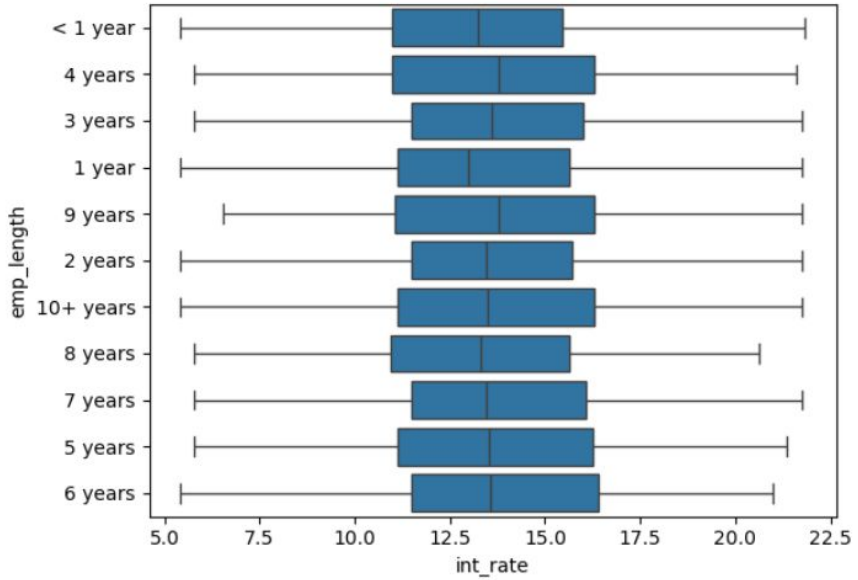
Analysis shows IQR range of **Small business , Debt Consolidation** is more in defaulters as compare to closed loans.

Vacation IQR range is less in fully-paid borrowers, it means borrowers are taking less loans for vacation hence they are not defaulting.

Contd…



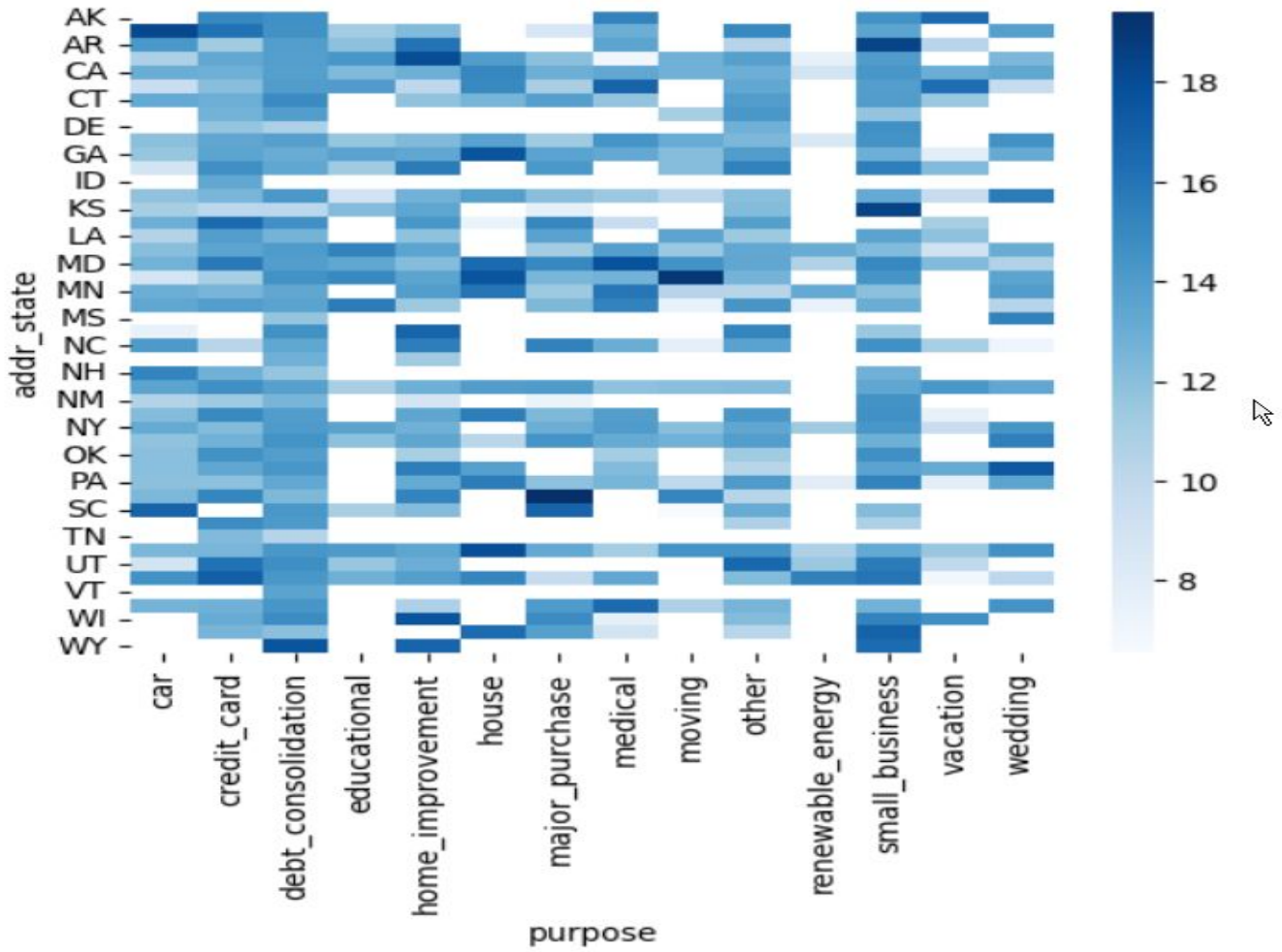
'Int_rate' vs 'emp_length'
For Charged-off borrowers



'Int_rate' vs 'emp_length'
For fully-paid borrowers

All defaulted customers are paying more interest rate as compare to non-defaulter customers.

Multivariate Analysis



'Int_rate' & 'addr_state' & 'purpose'
On Charged-off borrower

1. Renewable Energies have very less defaulters.
2. Debt_consolidation borrowers are generally getting defaulted.
3. CA States have more defaulters as compare to other states

Recommendation

- Lending Club should consider refraining from approving loans for customers seeking loan consolidation, as they are more likely to default. The **purpose** of the loan should be carefully evaluated.
- Borrowers are taking loans for 60 months are 20% more getting defaulted since small amount loans are more their **term** should be less.
- Higher **interest rates** for certain purposes, such as small business loans, tend to attract more defaulters. Lending Club should consider lowering interest rates for such cases to reduce default risk.
- Defaulting customers with lower **annual incomes** are taking out larger loans. Lending Club should implement an upper limit on loan amounts based on the borrower's income.
- Borrowers with lower **debt-to-income (DTI) ratios** are less likely to default.
- LC should review a borrower's **open credit lines** before approving a new loan. Apart from mortgages, borrowers typically maintain 3 to 4 active credit lines.