Lending Club Case Study

Group Member: Priyesh Kailash Shah

Business Understanding

Lending club is a consumer finance company which specialises 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. Two **types of risks** are associated with the bank's 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

Problem Statement

This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Analysis steps

Data Loading and Cleaning

- Load data
- Sanity checks
- Data cleanup
- Fill Missing data
- Derive columns

Univariate Analysis

- Analyse data for each column

Segmented Univariate Analysis

- Analyse charged off data set for each column

Bivariate Analysis

- Analyse data columns with target column (loan_status)
- Check correlation between different variables

Summary

- Provide suggestions

Univariate and Segmented Univariate Analysis

Loan status :

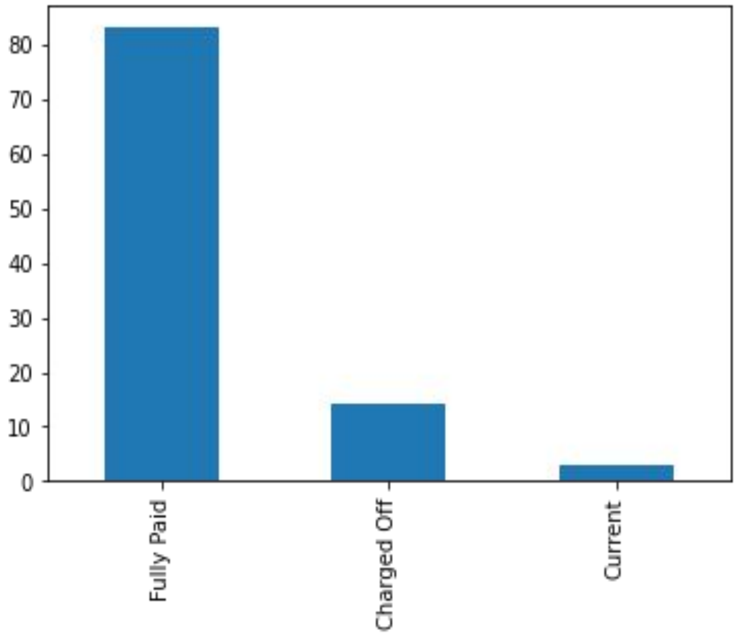
82.96% loan is Fully Paid

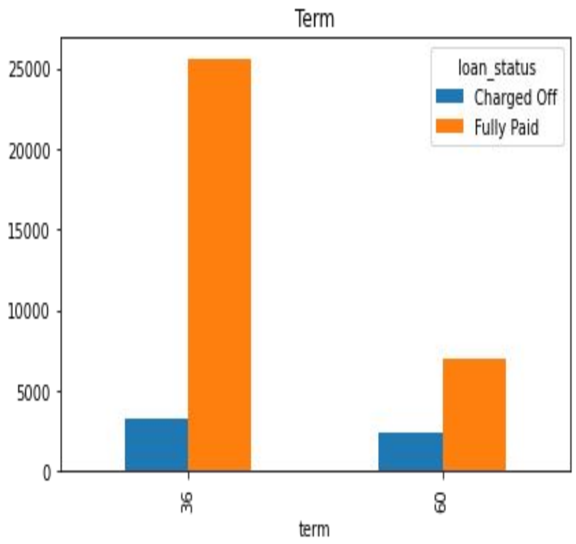
14.17% loan is Charged Off(Defaulted)

2.87% loan is Current status (In Progress)

Observation :

- Out of data provided from year 2007 to 2011, 14.17% borrowers are defaulted.



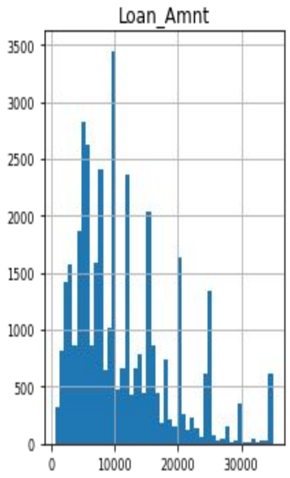


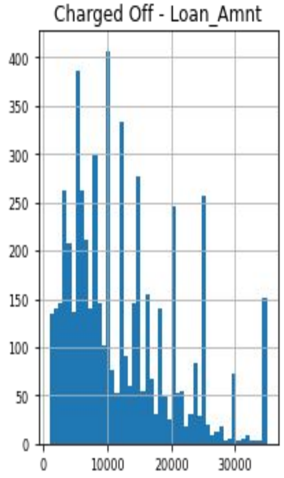
Loan Amount: Loan amount varies from 500 to 35K .Based on trends from loan amount defaulted in all categories

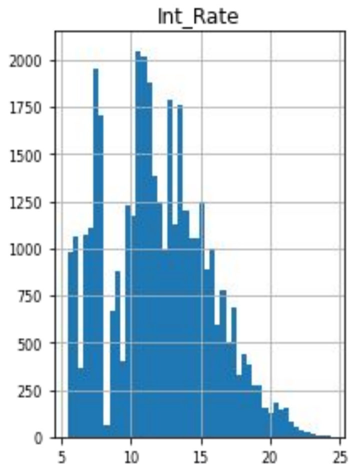
Term : Loan amount payment is for 36 and 60 month term. Loans are getting defaulted more for 36 months.

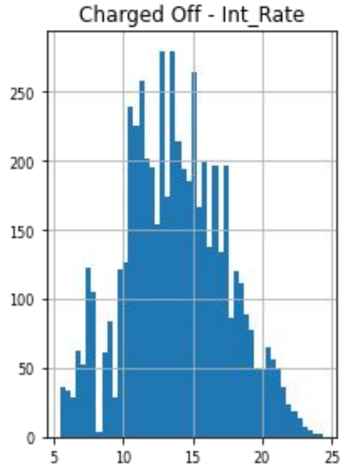
Int Rate : Int rate varies from 5% to 24%. Loans are getting defaulted when Int rates are higher i.e ~10% to ~18%

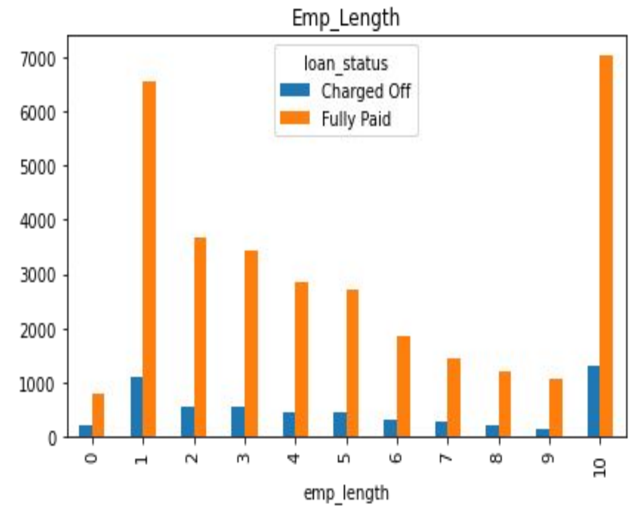
Employee Length : Borrowers from 0 to 10+ experience are taking loans. Also the loans are getting defaulted mainly for 1 and 10 year experience borrowers.





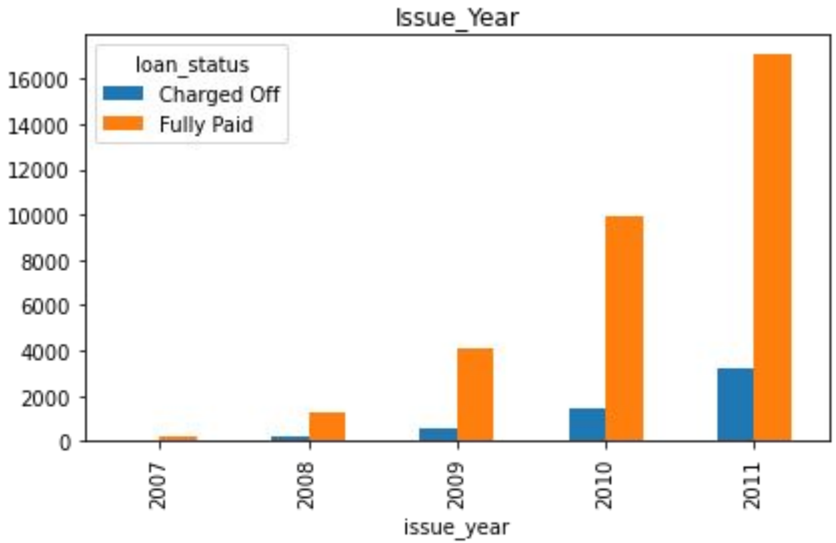


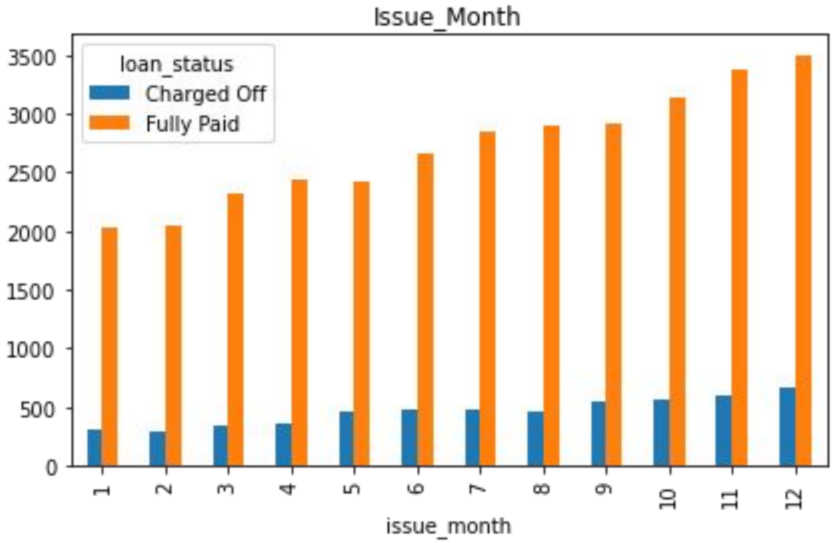




Issue Year and Issue Month : Loans applications are increased every year and also loan defaults are increased every year from 2007 to 2011.

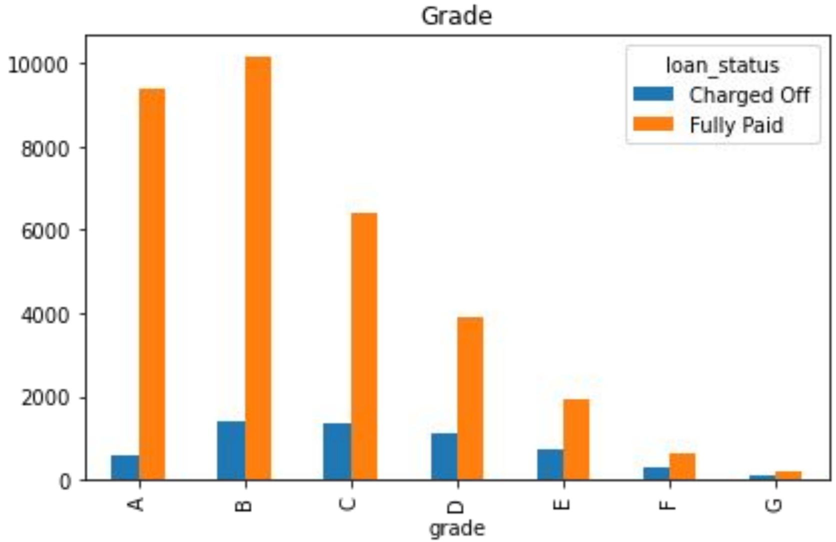
Loans are taken more at 12th month (Dec) and that shows the trend in defaults too.

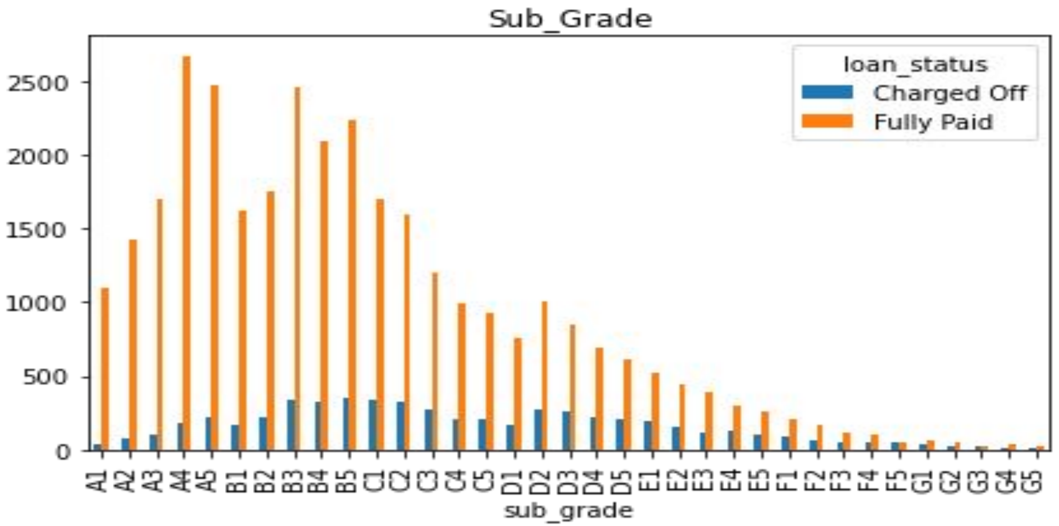




Grade and SubGrade : Many loans are marked as Grade B. (Need further investigation on Grading system)

Loans are getting defaulted for B,C,D more.Same trend can be seen in sub grade

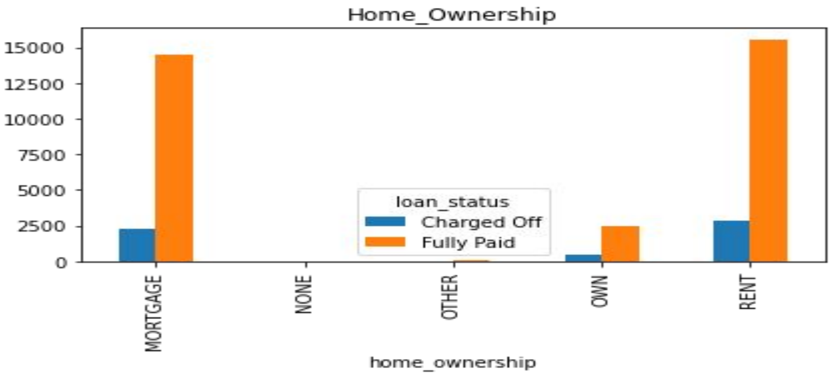


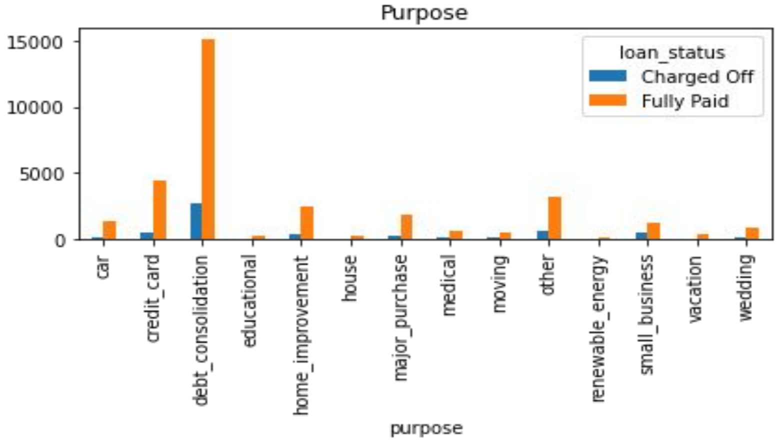


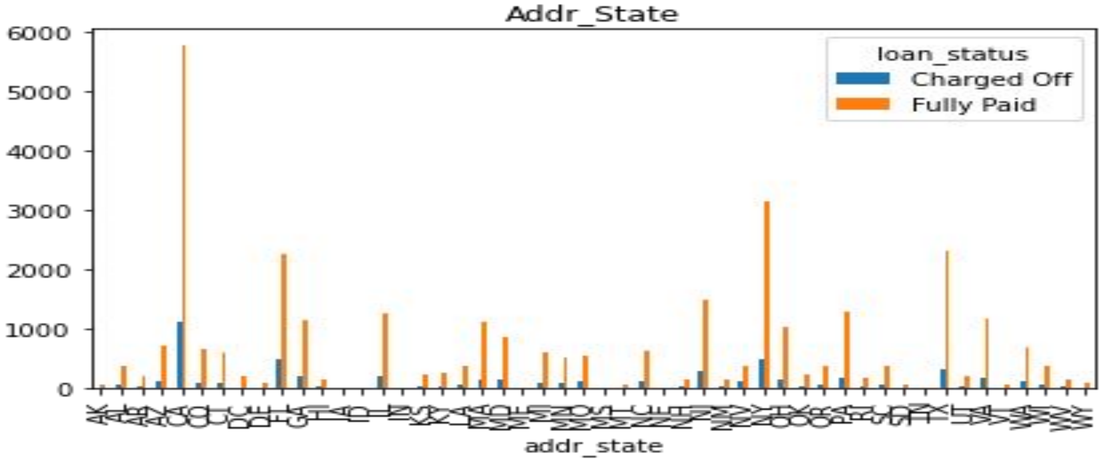
Home Ownership : Borrowers who are having RENT and MORTGAGE homes are taking loans,also they are the one getting defaulted often.

Purpose : Major purpose of taking loan is debt consolidation followed by credit card Payments. Defaulters are seen more in debt consolidation,others and credit card

Address State : State CA,NY,FL are getting defaulted more compared to other states.







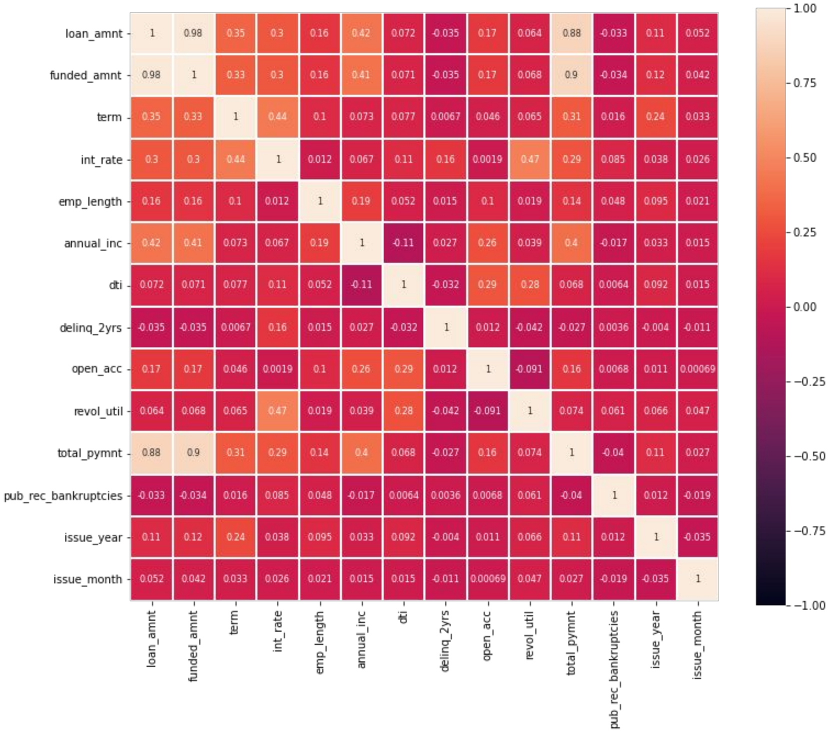
Bivariate Analysis

Correlation Matrix:

- dti is negatively impacting annual income -0.11

  i.e if annual income is low,dti (debt to income ratio) is high and vice versa

- loan_amnt,funded_amnt,total_pymnt are correlated with 0.9

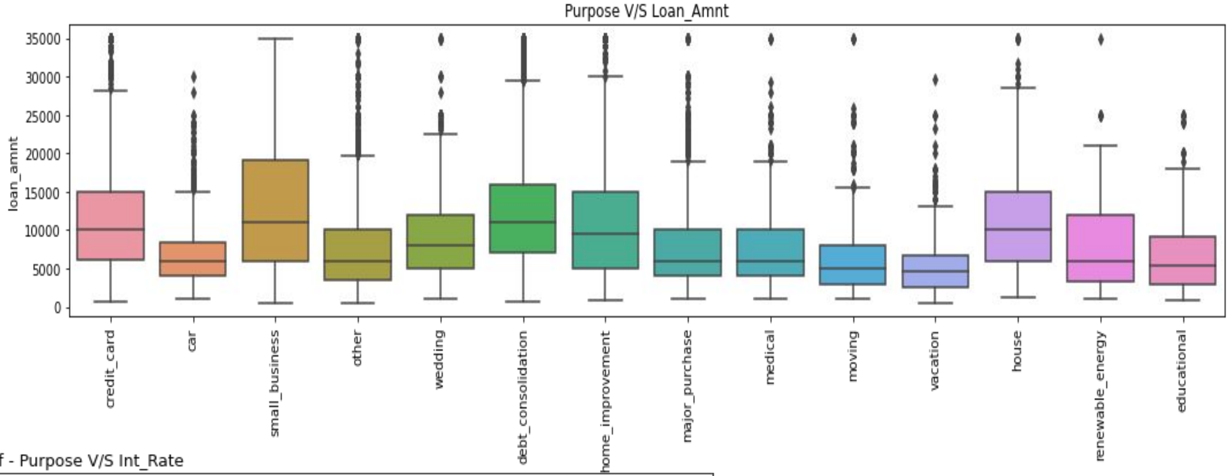


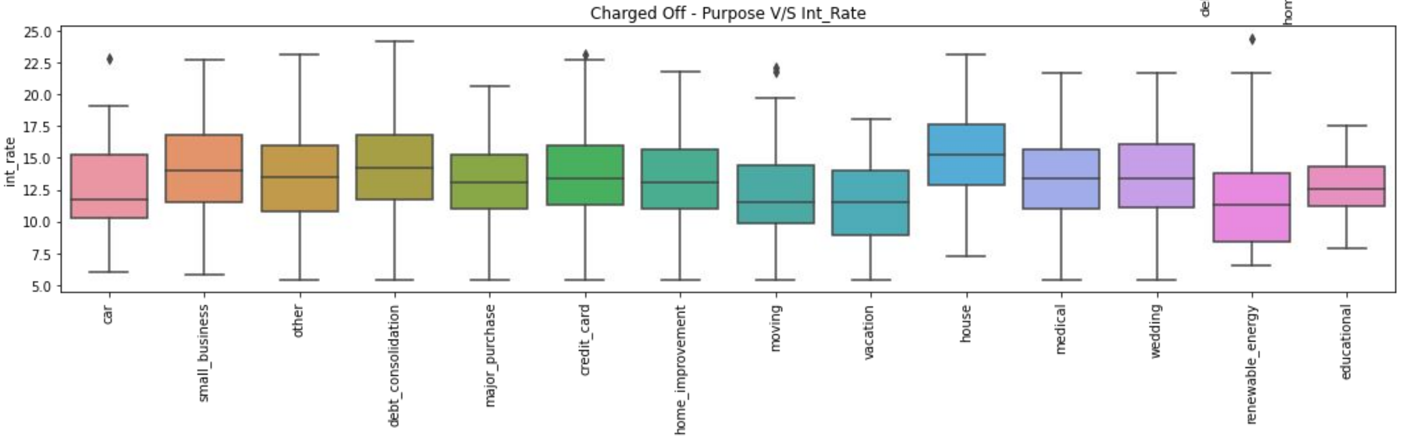
Purpose and Loan Amount

Major purpose of taking loan is for small business.

But loans are getting more defaulted for small business,

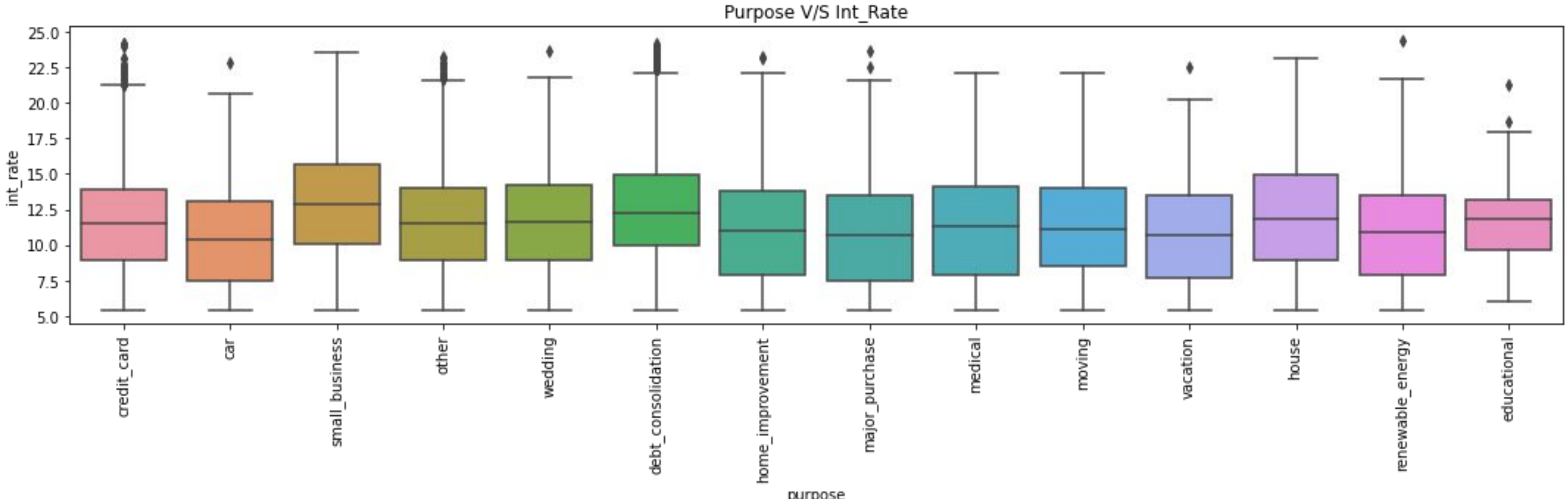
house,debt_consolidation.





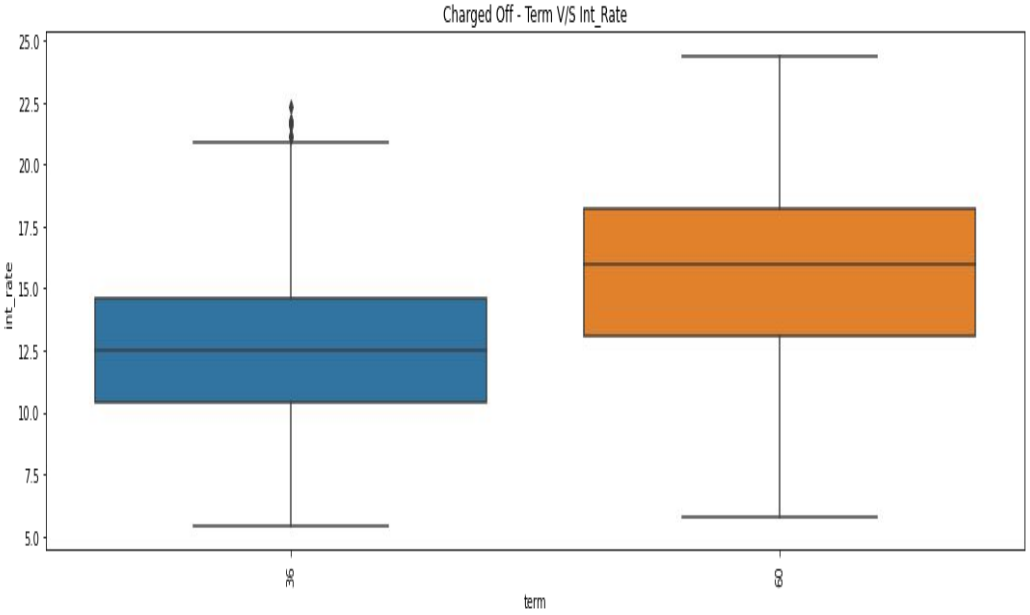
Purpose and Int Rate

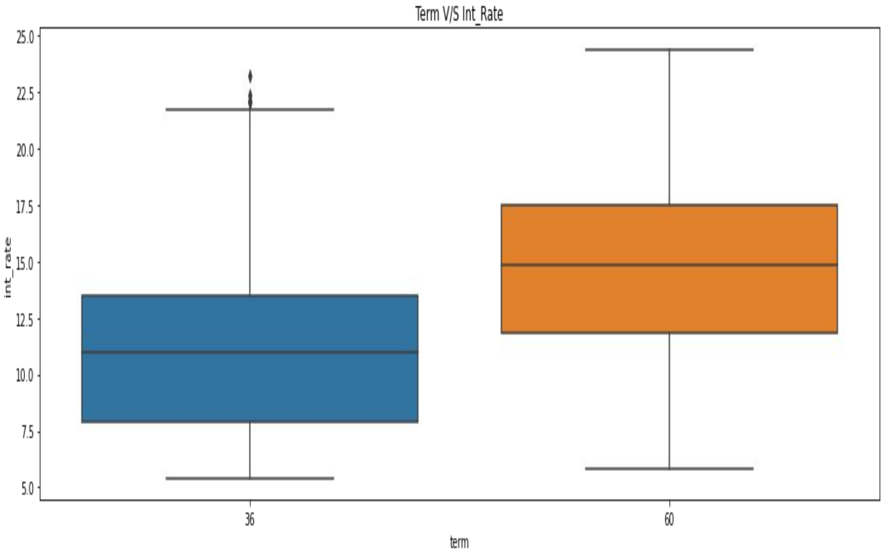
Int rates are higher for small business loans. This shows they are getting defaulted also more in the category.



Term and Int Rate

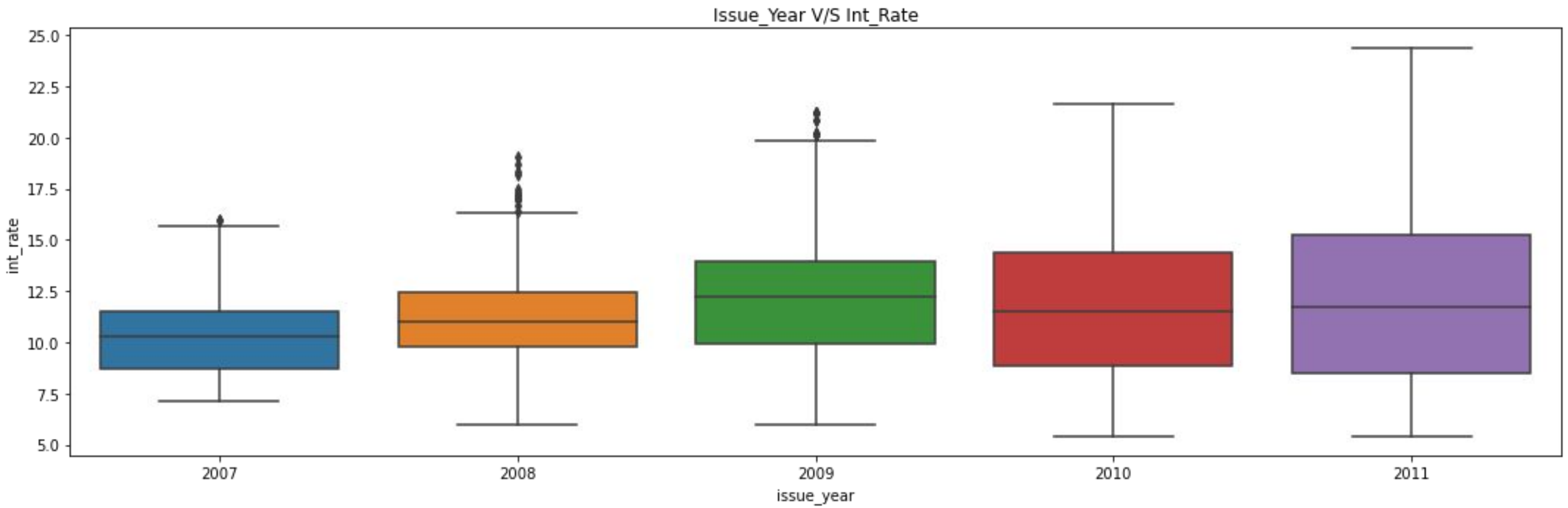
Int rates are higher for 60 months term as compared to 36 months. Loans are getting defaulted if Int rates are higher than 10%





Issue Year and Int Rate

Int rates range is increasing every year based on trend.

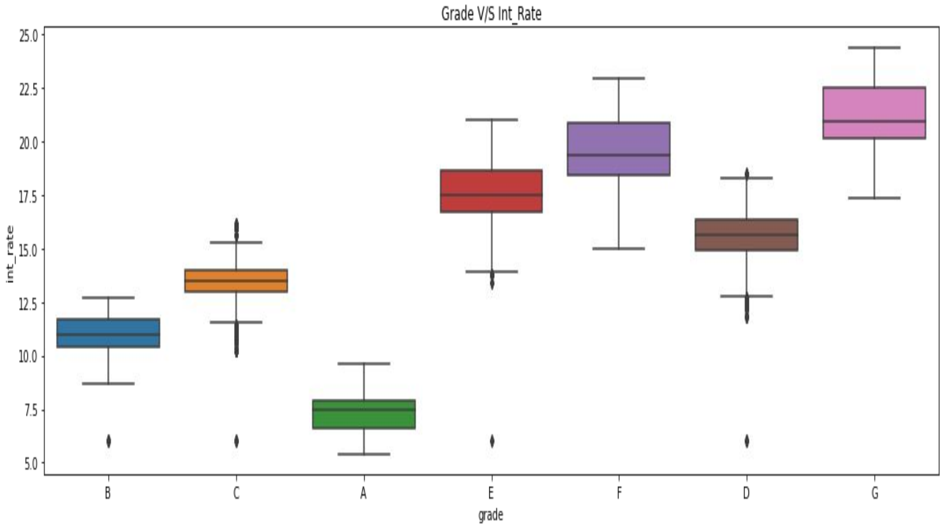


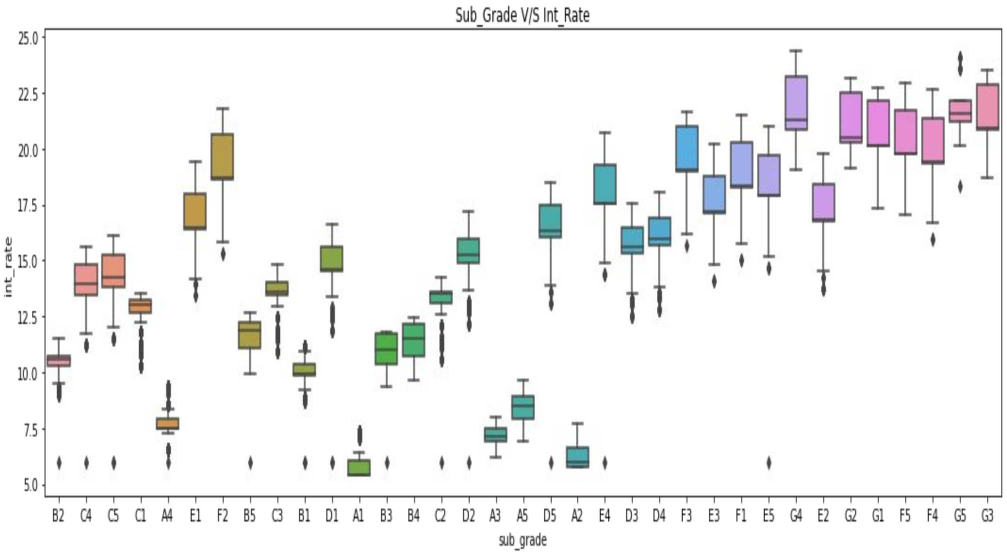
Grade and Int Rate

Int rates are increasing for Grades. This means higher grade having high int rates.

Sub Grades also shows the same trend where Int rates are increasing in

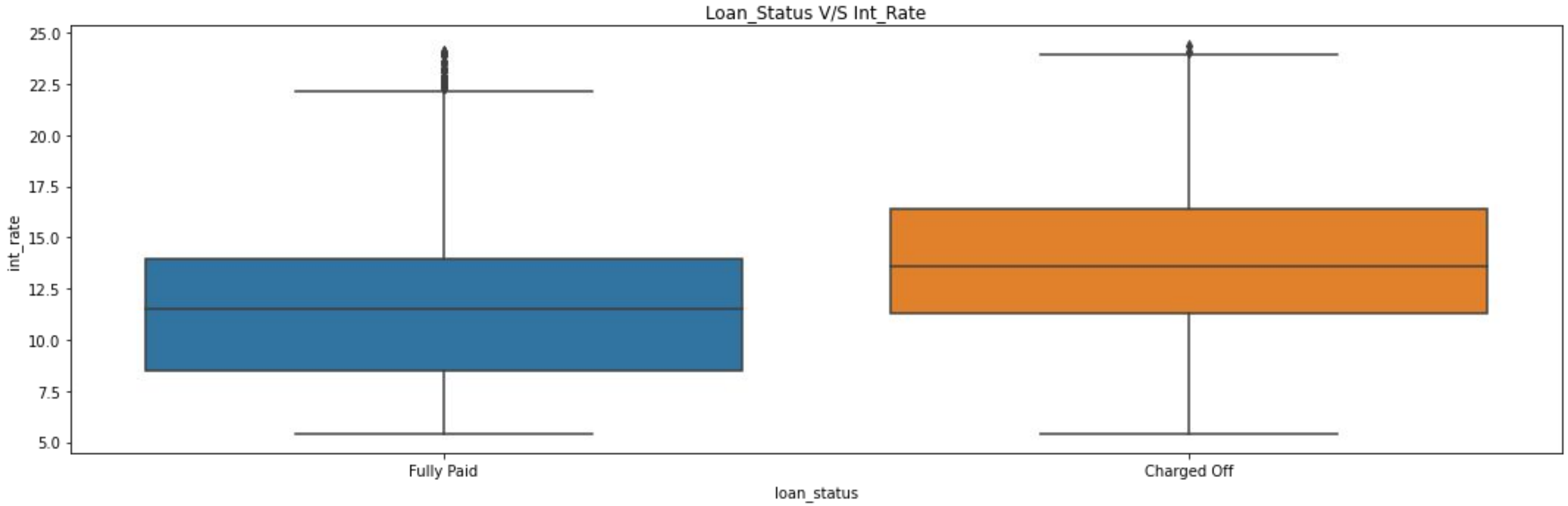
Chronological order from A to G.





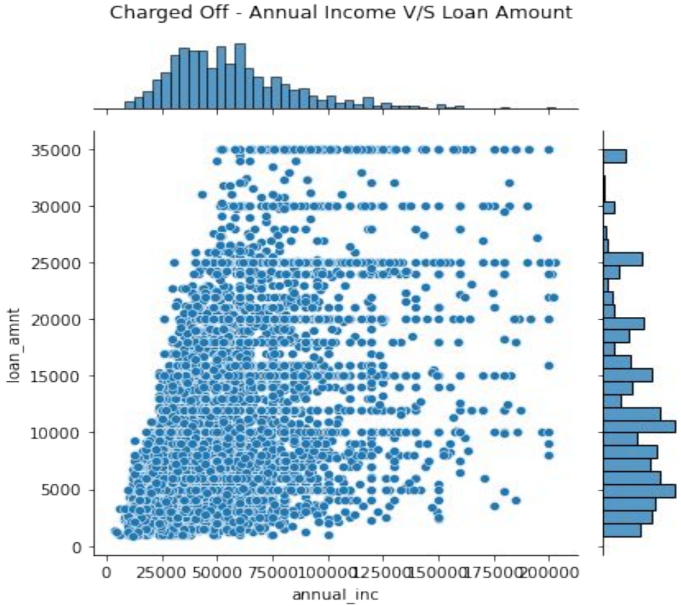
Loan Status and Int Rate

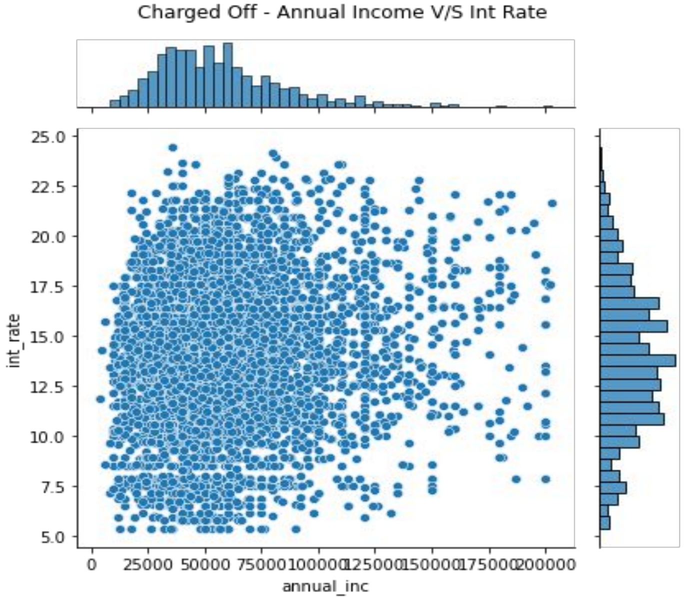
Loans are getting defaulted when Int rates are higher.

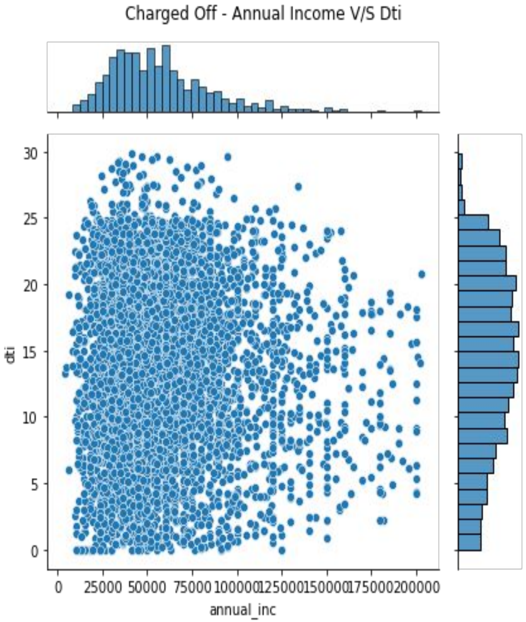


Annual Income and Loan Amount/Int Rate/Dti

Borrowers defaulted if annual income is lower than 50000 and taking loan more than 25000



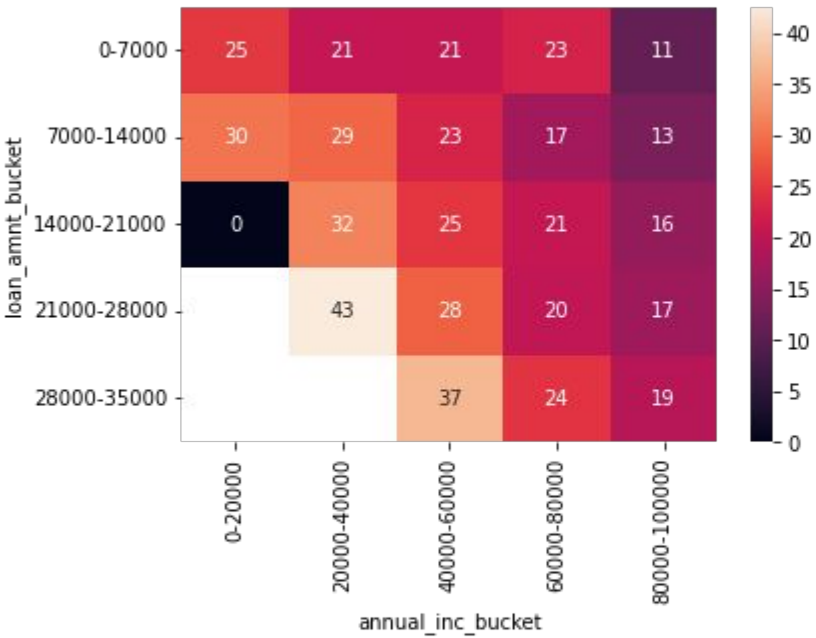


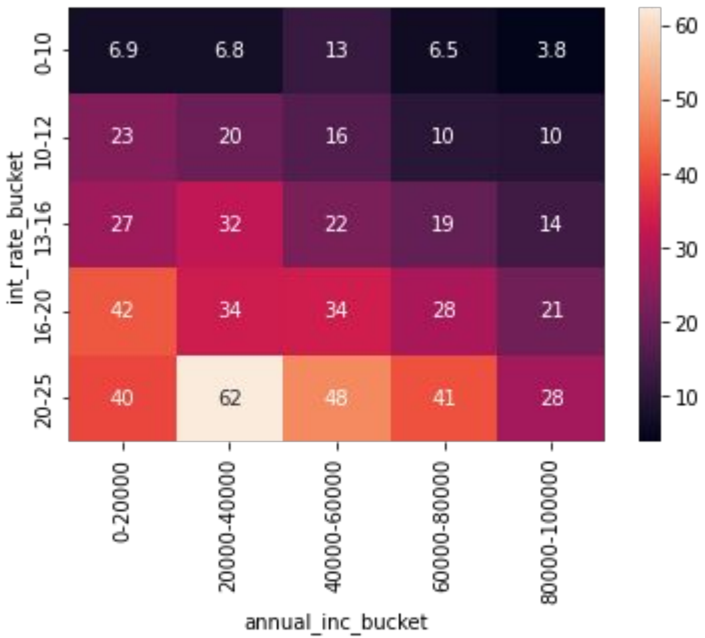


Annual Income and Int Rate/Loan Amount Category

Annual Income range 20000-40000, Loan Amount range 21000-28000,Int Rate 20-25% are

major defaulters.





Summary

- Lending club should stop approving loans if below criterias are majorly fulfilled.
    - Loan applications are from CA,NY,FL state
    - Loan application is for purpose : small business
    - Loan application is in month of December
    - Borrower Annual Income is less than 40000 and loan amount is in range 21000-28000
    - Borrower work experience is 1 or 10+
    - Loan int rates are more than 12%
    - Borrower home ownership is RENT and MORTGAGE
    - Borrower grades are between D to G. The higher the grade i.e G more possibility for default