Lending Club Case Study

-Amogh Gadagi

Lending Club Case Study

- **Problem Statement:**

1. Lending Club 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.
2. Lending loan to risky applicant is the largest source of financial 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.

- **Objectives:**

1. Identify the driving factors behind loan getting charged off and identify risky loan applicants, and taking appropriate steps such as denying loan, reducing loan amount at higher interest rate, etc.
2. Be able to identify the trend and give business insights to the loan lending organizations.

Problem Solving Methodology

- Data Understanding:

1. Loan Data contains information about loan issued during the period 2007-2011.
2. Data Dictionary explains the terms used in the data set.

- Data Cleaning:

1. We have dropped all the null values along columns (axis=1). More than 60% columns are eliminated by this step.
2. We have removed the columns which are not significant for our analysis.
3. We have removed date corresponding to ‘Current’ Value in loan status feature.
4. We have imputed the mode value for null values in emp_length feature.
5. We have filled the na values in pub_rec_bankruptcies column to ‘Not Known”
6. We have derived multiple columns which are useful for analysis.
7. We have defined a function to calculate Default Ratio.

Exploratory Data Analysis

- Univariate Analysis.

1. Histograms and Bar charts to check distribution.
2. Box plot to detect outliers

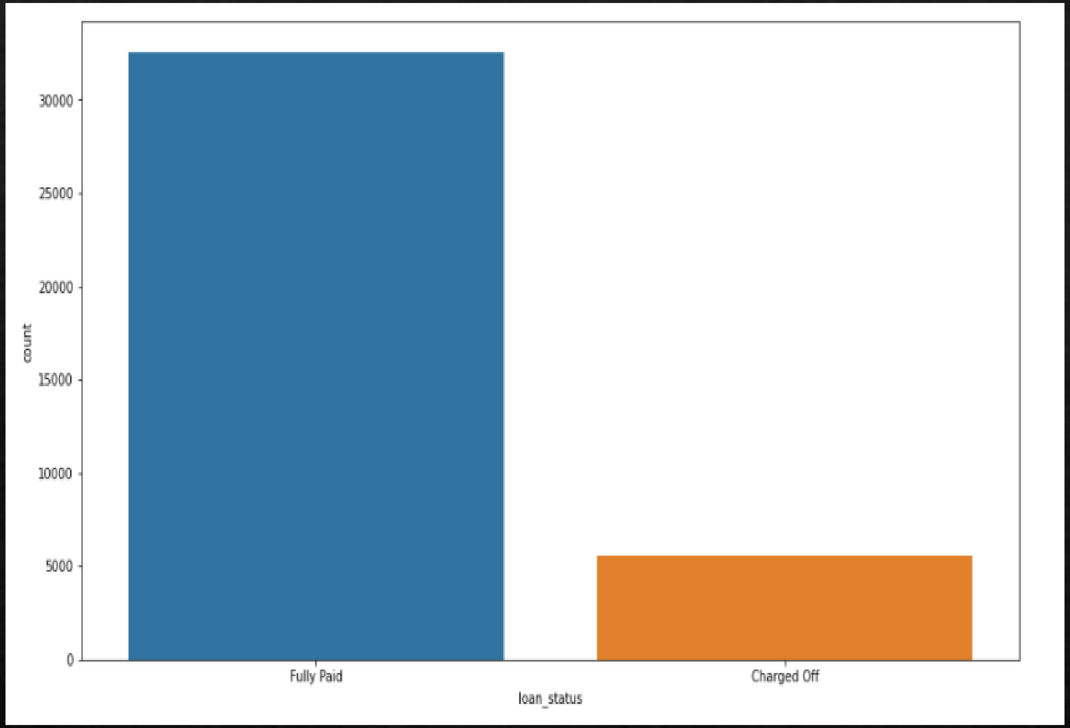
- Bivariate Analysis

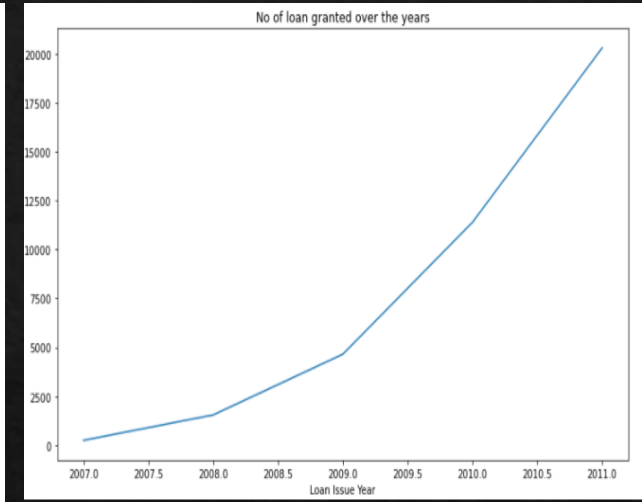
1. Count plots
2. Histograms to check distribution.

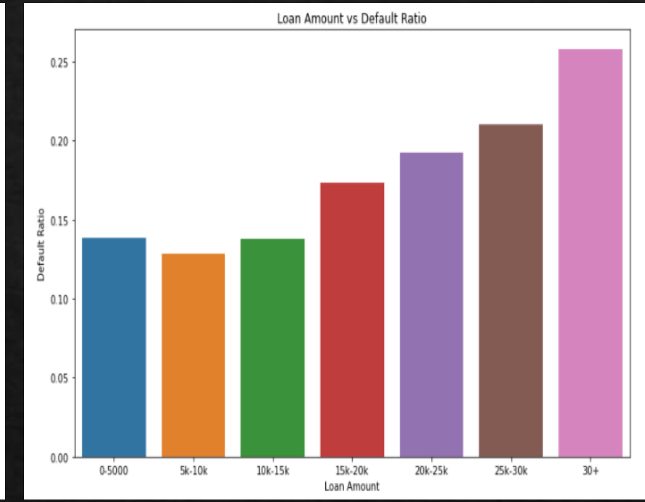
- Multivariate Analysis

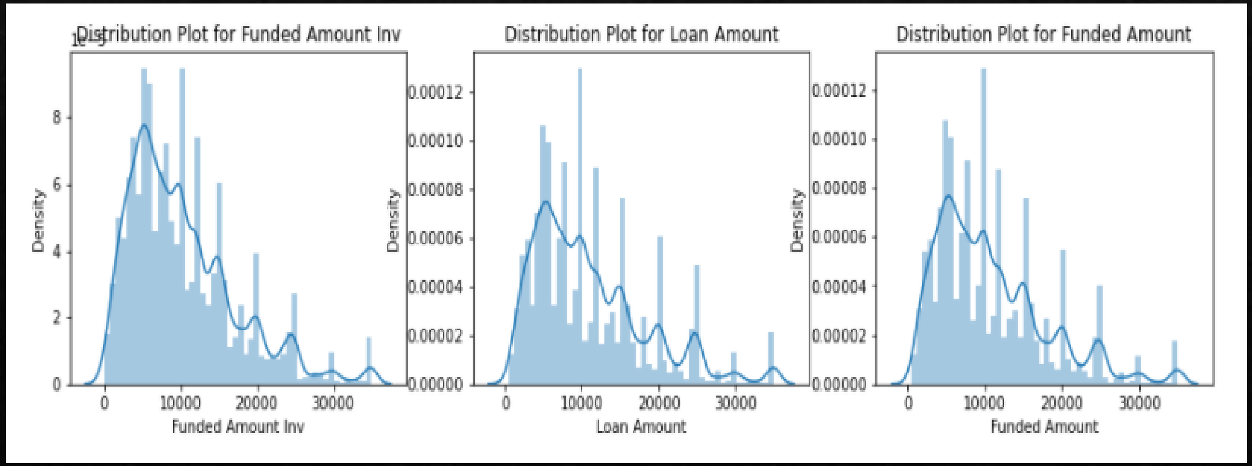
1. Cluster map- to check correlation between different variables.
2. Pair plots- Allows us to see distribution between two features.

Loan Status, Issuance by Year and Loan Amount Distribution



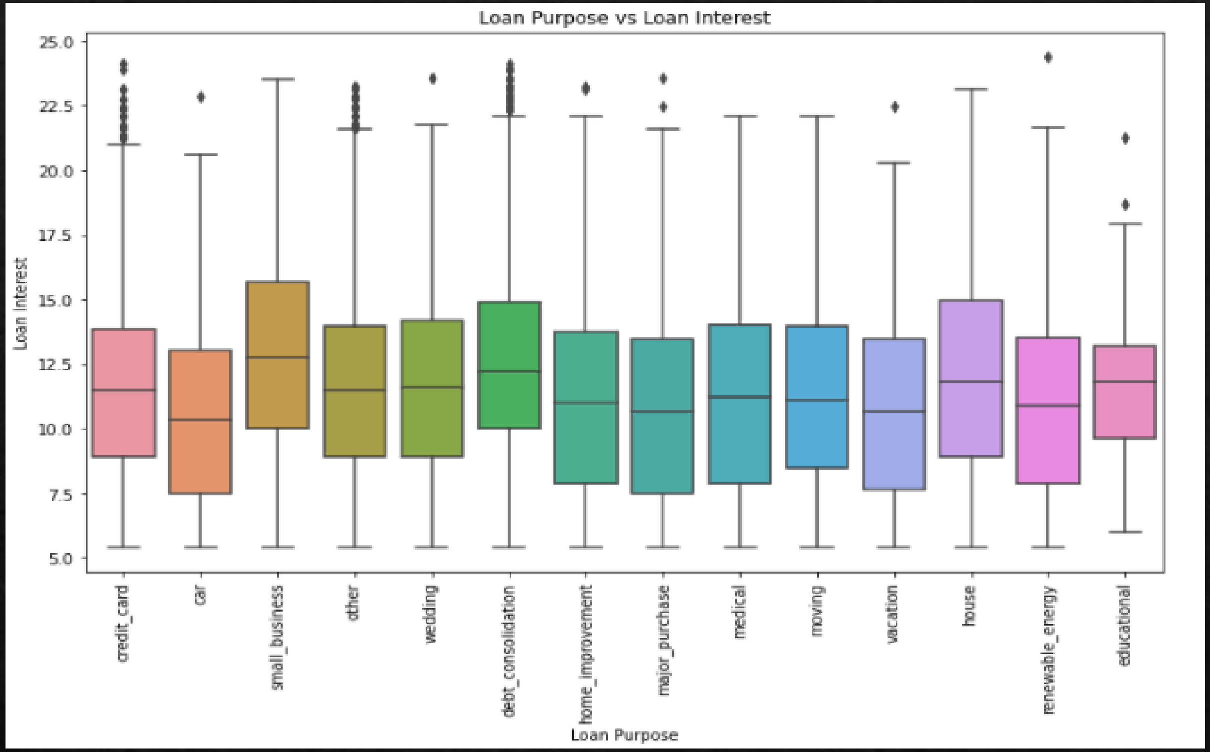


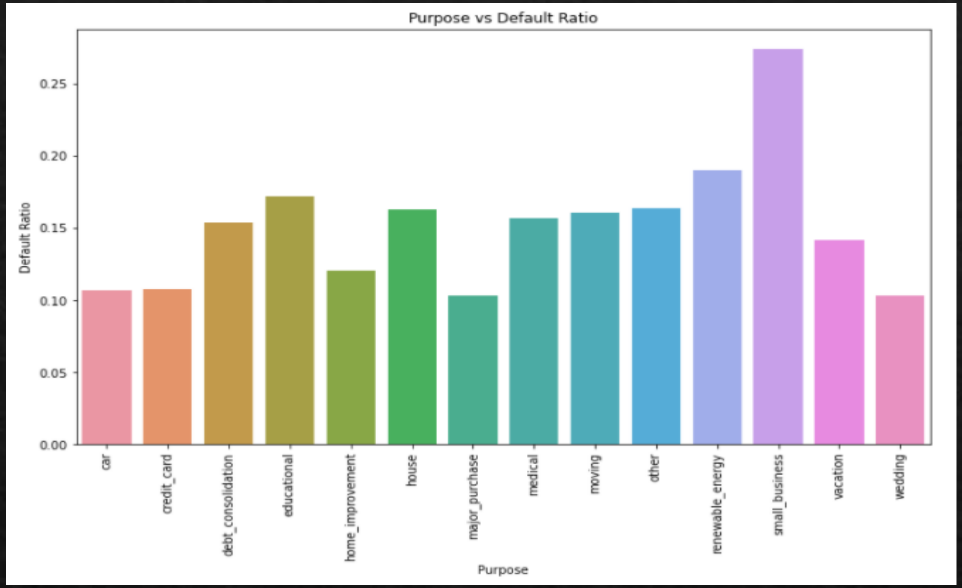


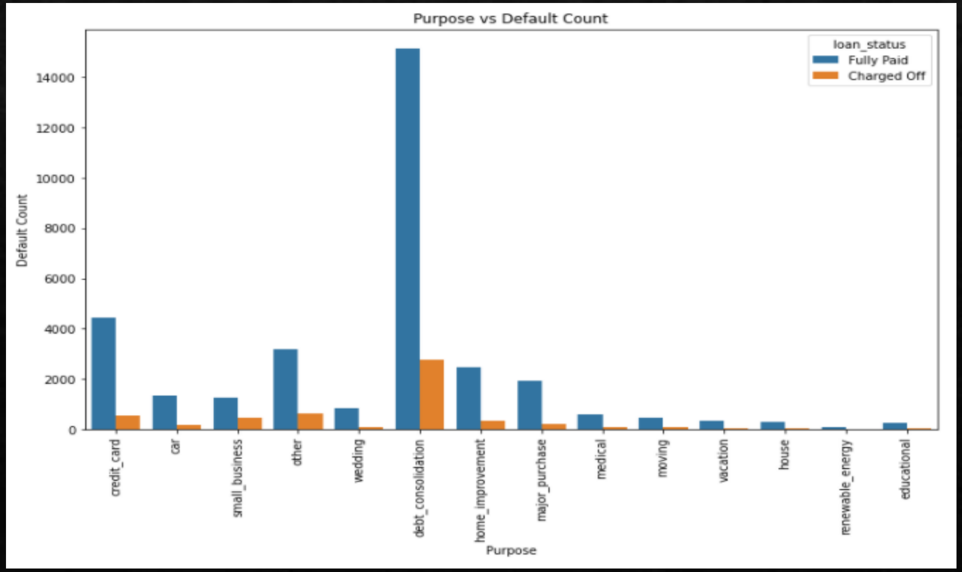


1. Most of the loans are fully paid. Overall charged off ratio is around 14% (left top corner).
2. More loans are being issued every passing year (top middle).
3. Defaults increase with increase in loan amount (top right corner).
4. Distribution of the all three amounts are almost similar and hence we will use only one amount for our analysis (lower right).

Loan Purpose Analysis

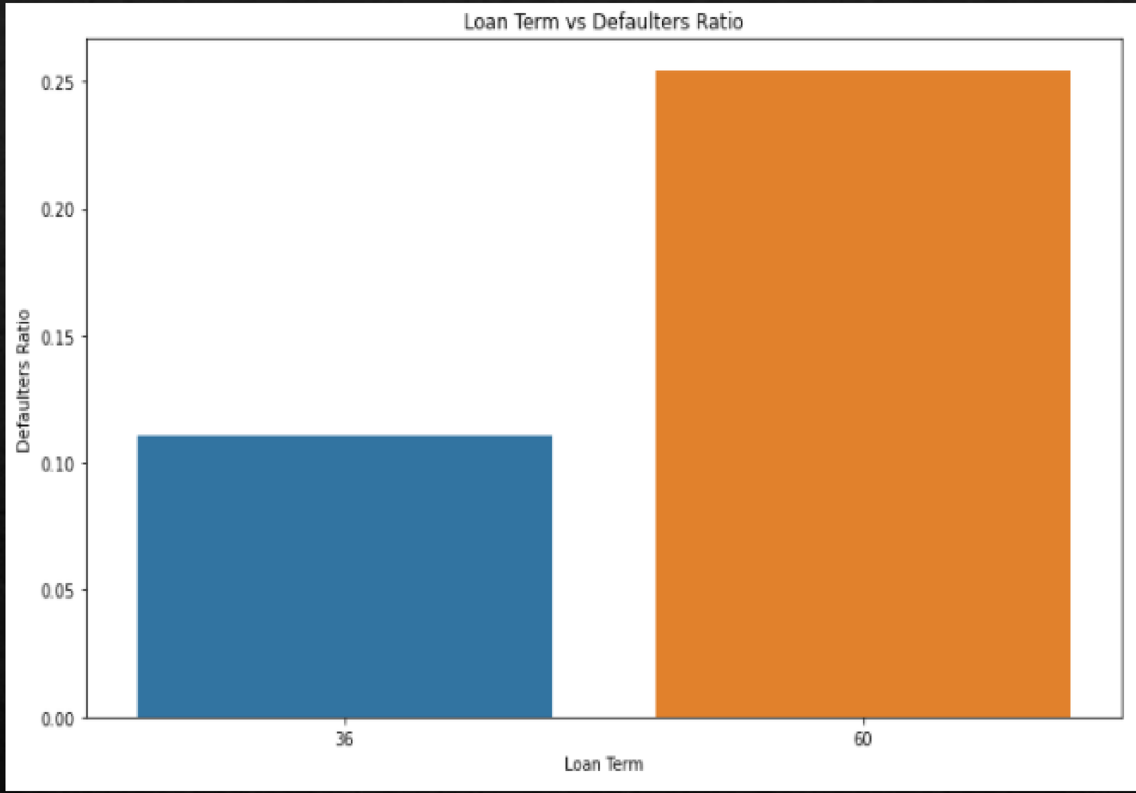


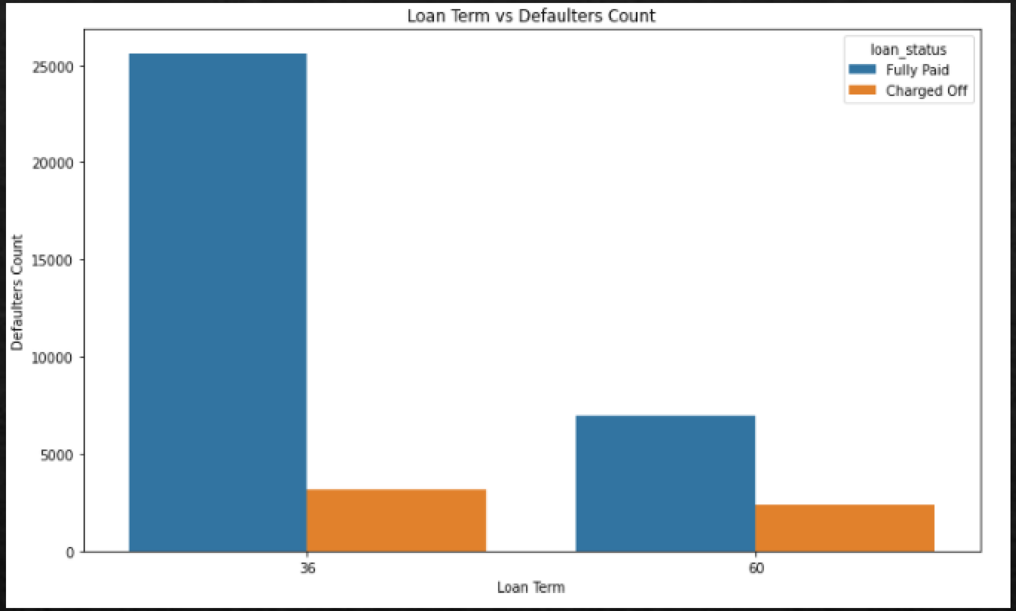


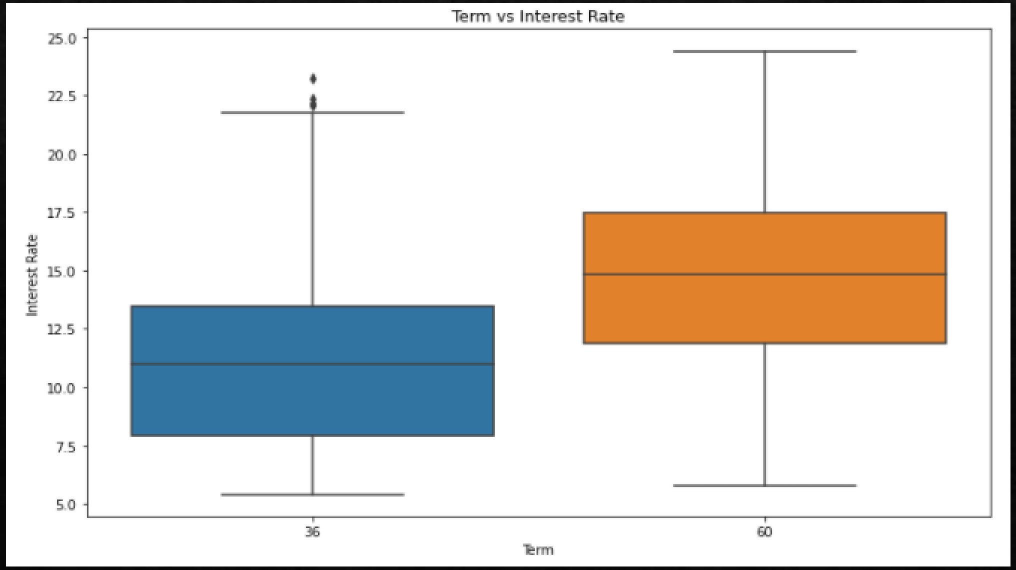


1. Interest rate is high for loans taken for small business. This might be the reason for higher default ratio (top left corner).
2. Default ratio i.e., charged off/total is comparatively high for small business than others. Hence lending org. should be careful before lending out money to people starting/owning small business (top right corner).
3. Most of the loans taken are for debt consolidation and credit card, hence more the default count (lower right corner), however the default ratio is less compared.

Loan Term Analysis

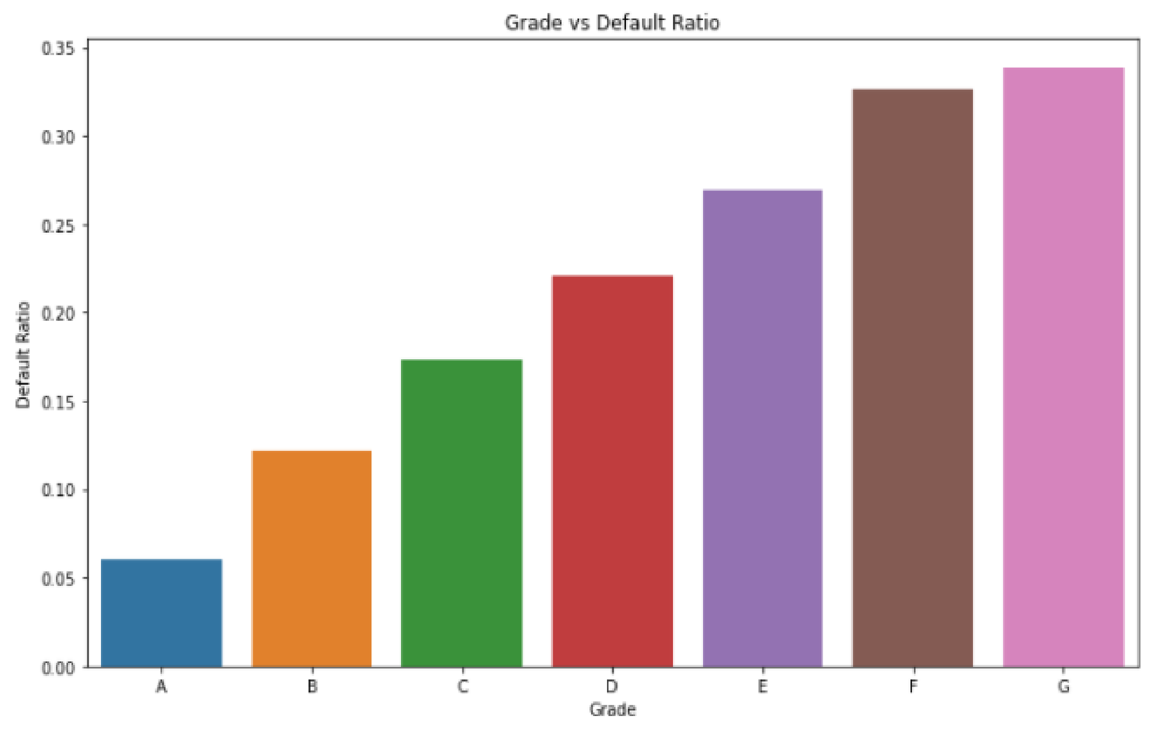


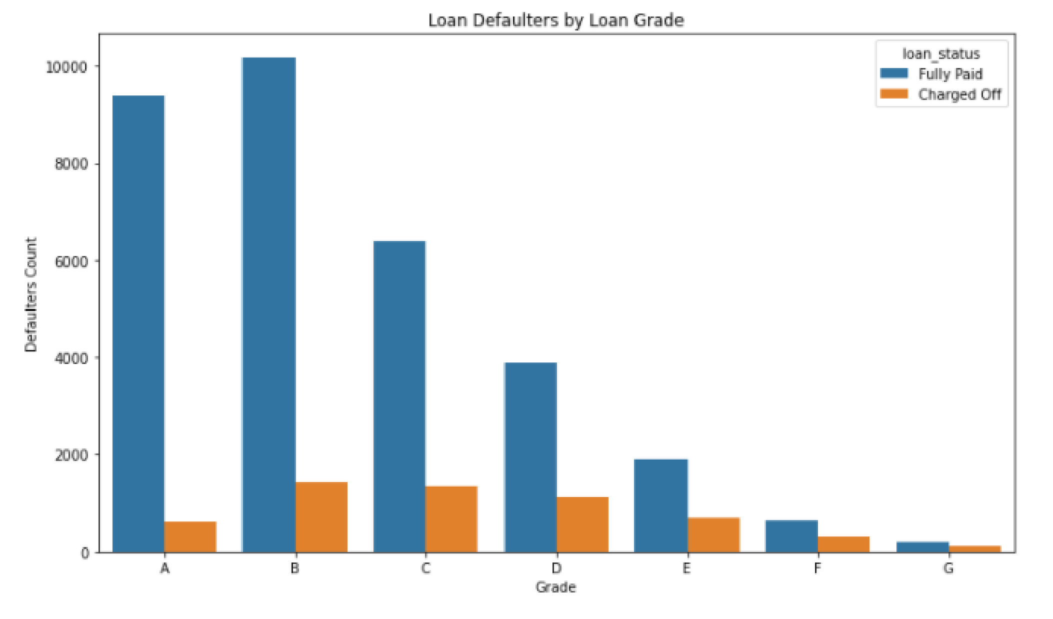


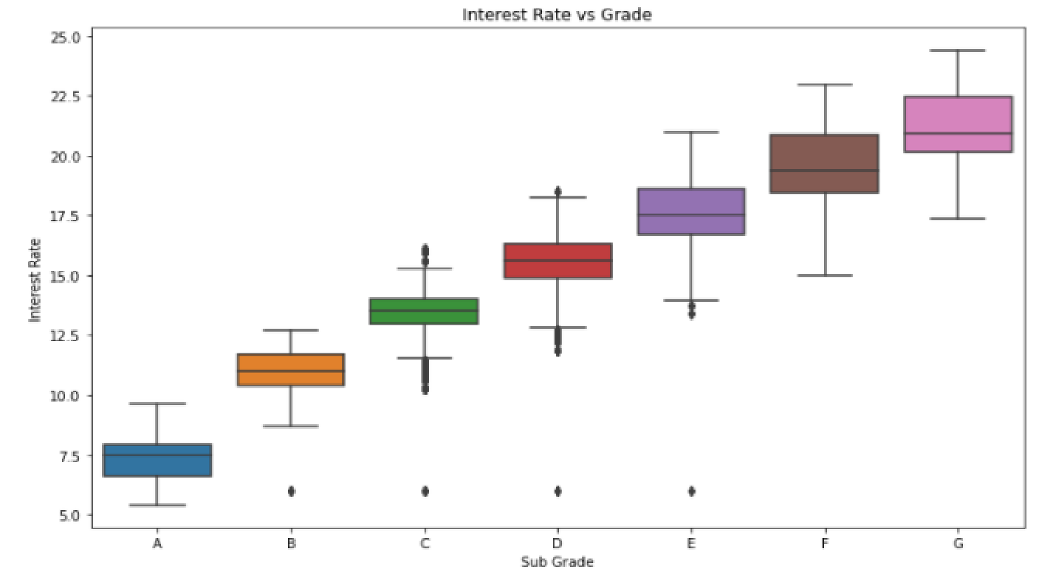


1. Default ratio is higher in longer term loans i.e. 60 months. Hence lending club should evaluate the underlying reasons and consider giving out loans.
2. Most of the time shorter term loans are paid back and there is no reason to not approve such loans.
3. This might be because of the higher interest rate that lending club charges for longer term loans. Reconsidering the interest rate might help is lesser defaults.

Loan Grade Analysis

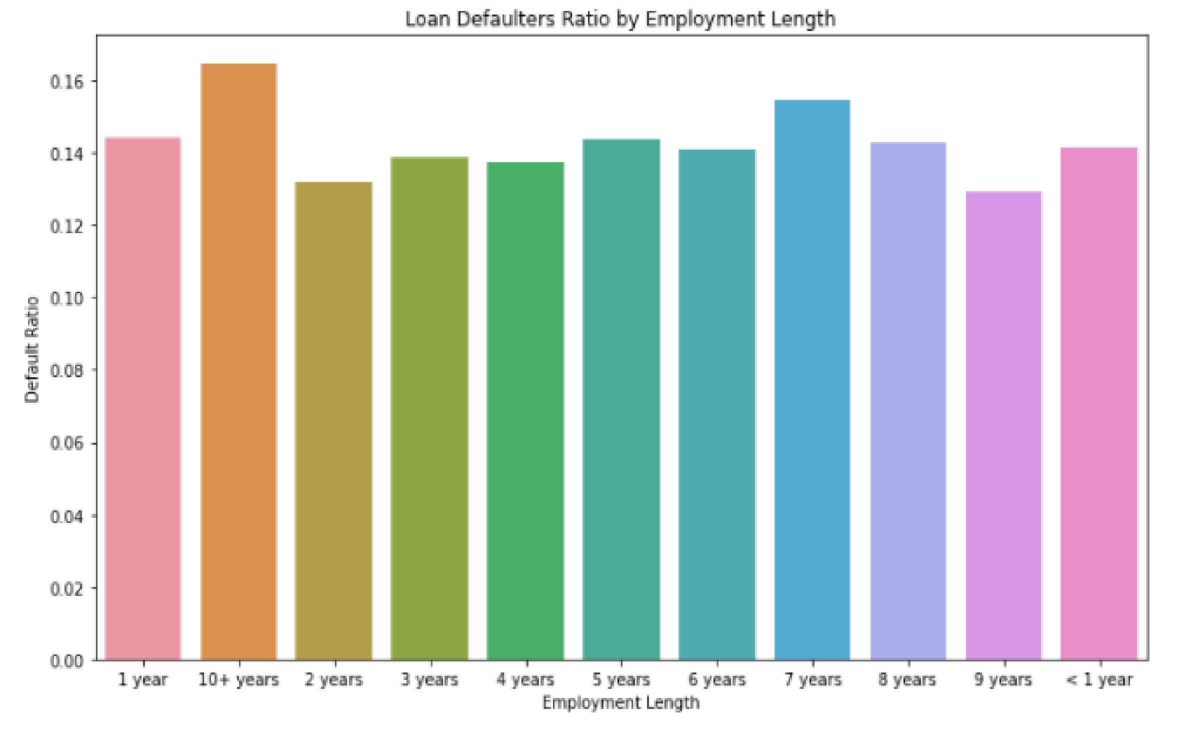


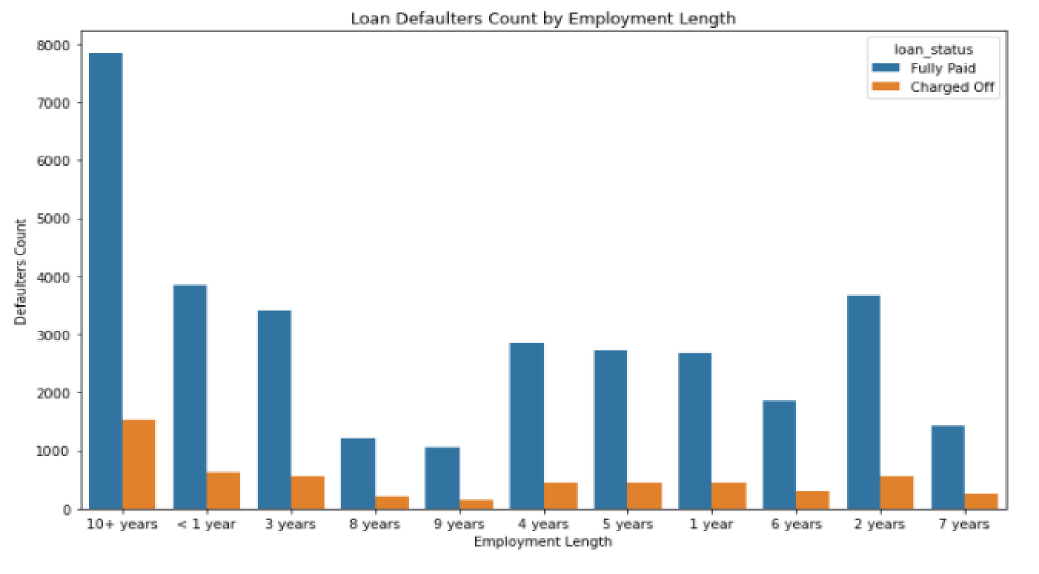


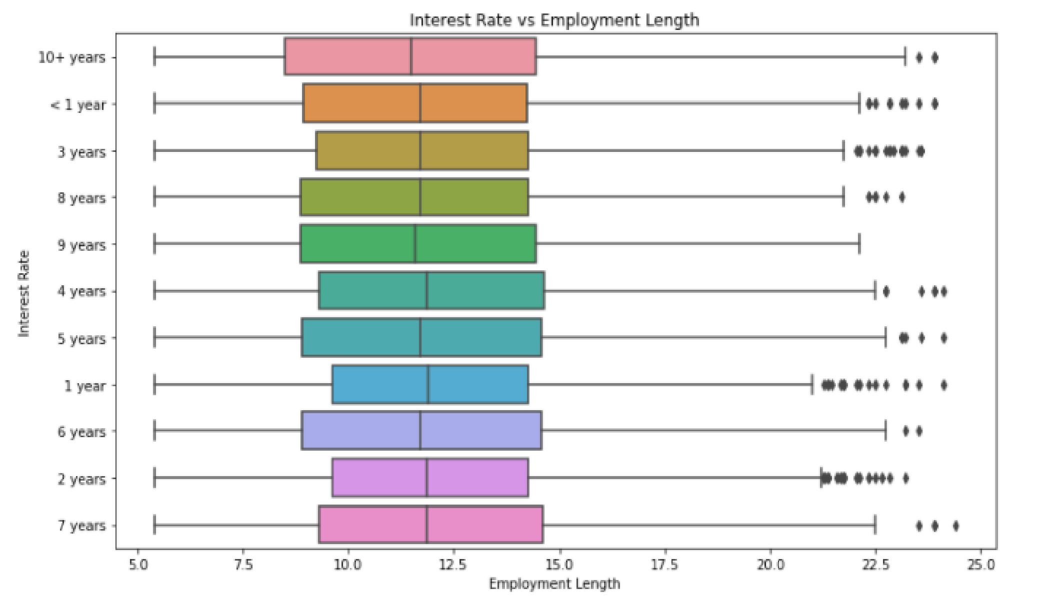


1. Default ratio is higher with higher loan grades. This ratio increase with increase in loan grade (top left corner)..
2. Most loans issued are of grade B and hence the default count is highest, however pay off is also more (top right corner).
3. This reason in higher default ratio as the loan grade increases might be because of the interest the lending club charges for it (lower right corner).

Loan Status on Employment Length

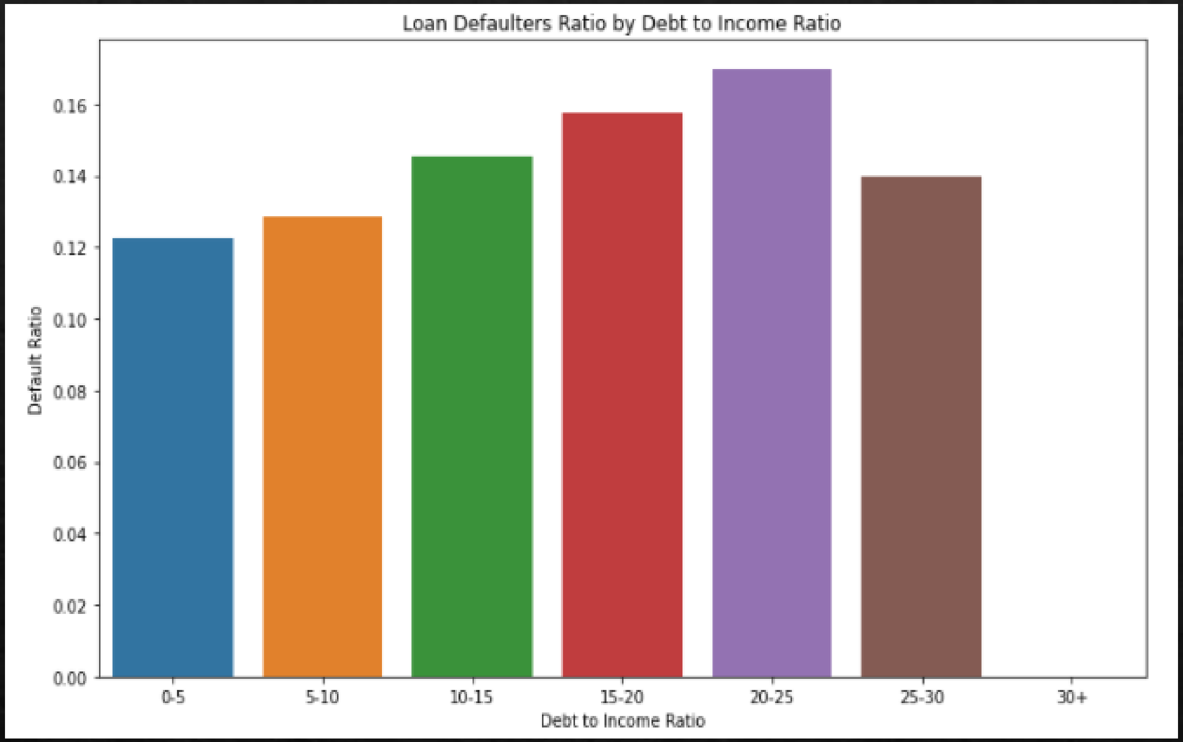


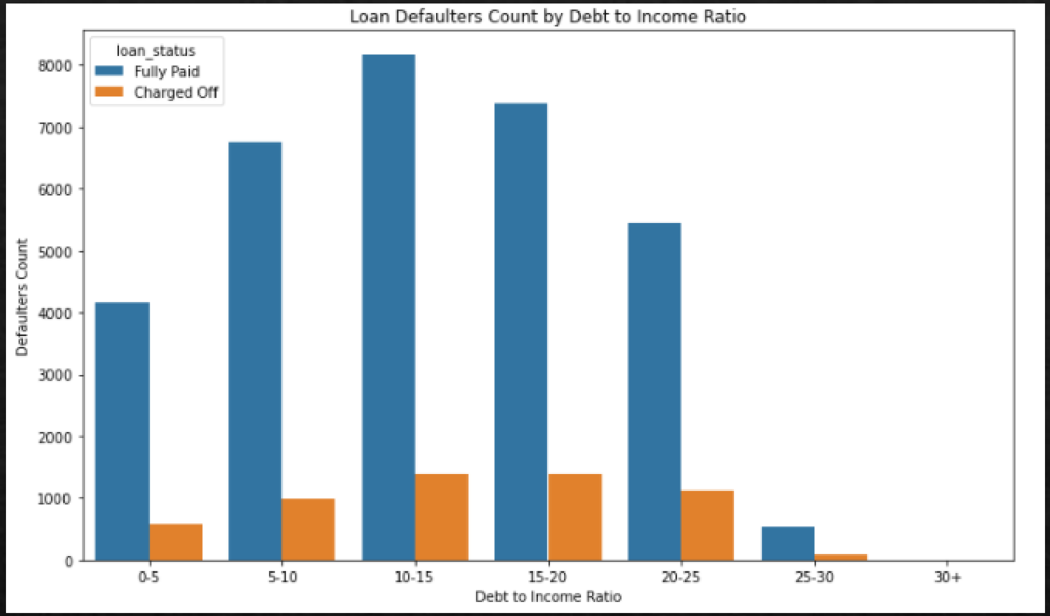


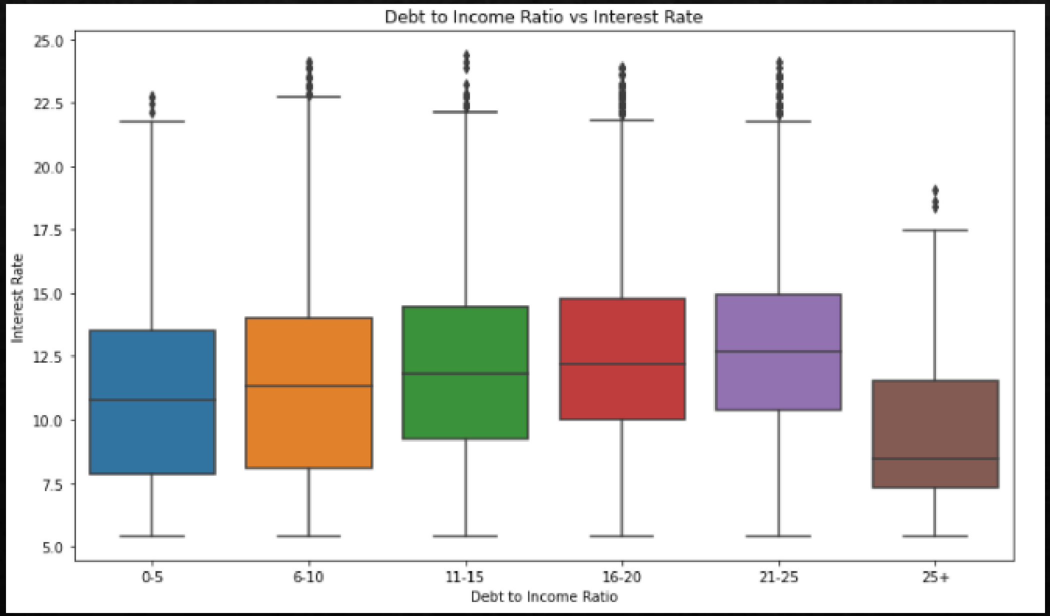


1. It seems like employment length doesn't matter when giving out the loan.
2. This might be because similar interest rates for everyone irrespective of employment length.
3. The default ratio is also similar for all categories; hence employment length doesn't play a major role wile giving out loans.

Debt to Income Ratio on Loan Status

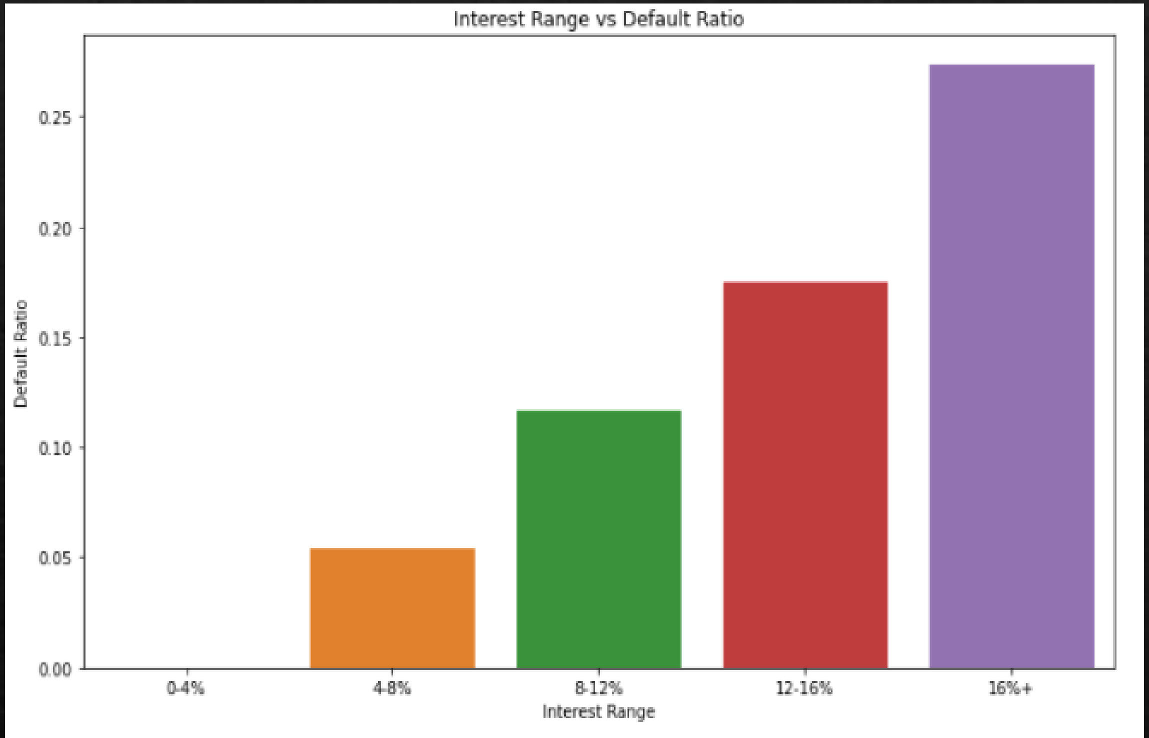


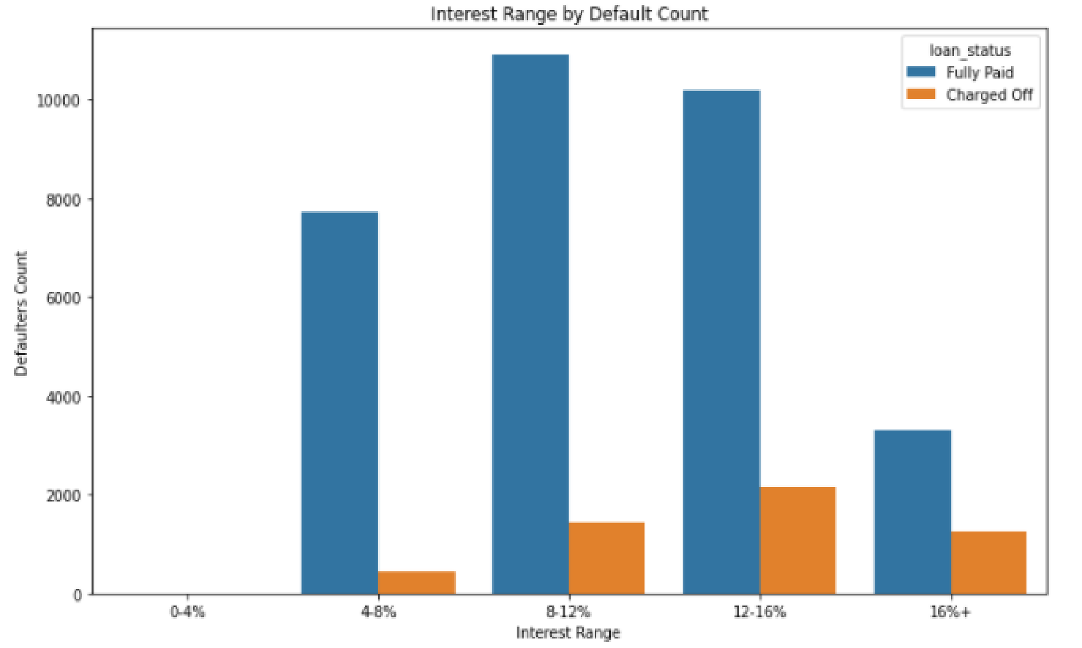




1. There is increase in default ratio with increase in debt-to-income ratio.
2. The interest rate is high for higher debt-to-income ratio . The last boxplot indicates presence of outliers.
3. There is no significant evidence, hence this might not be a serious variable to be considered while giving out loans.

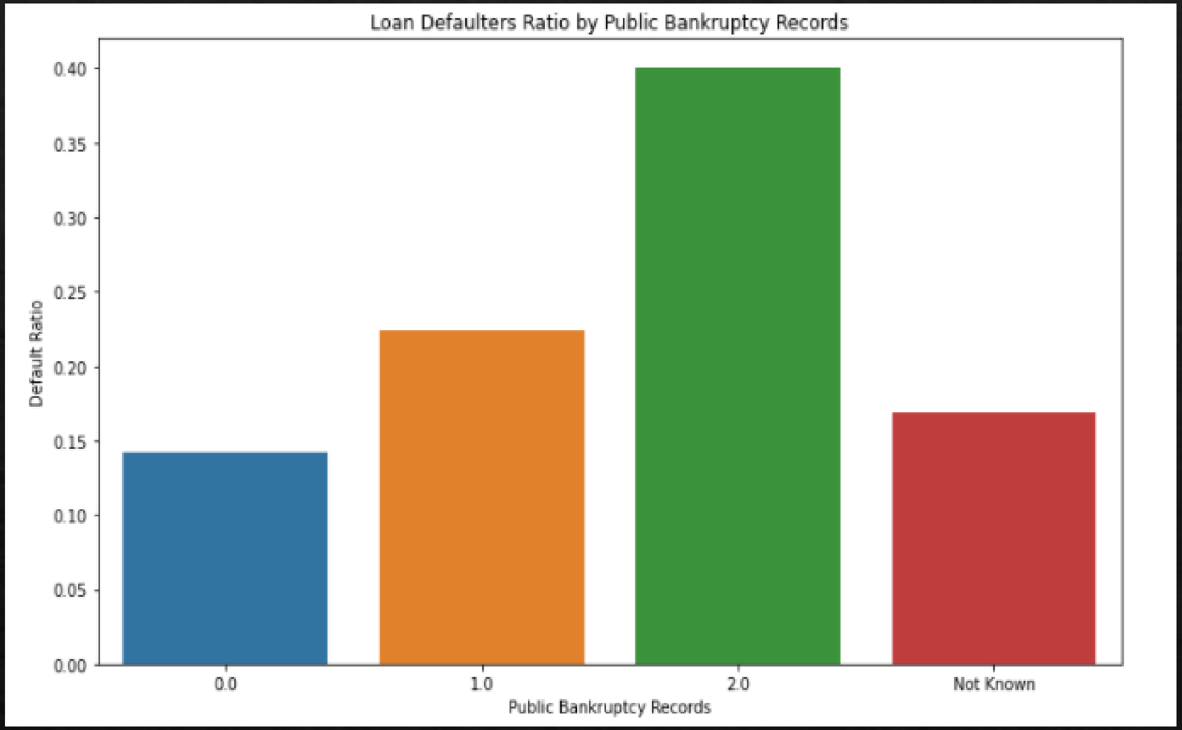
Loan Interest Analysis

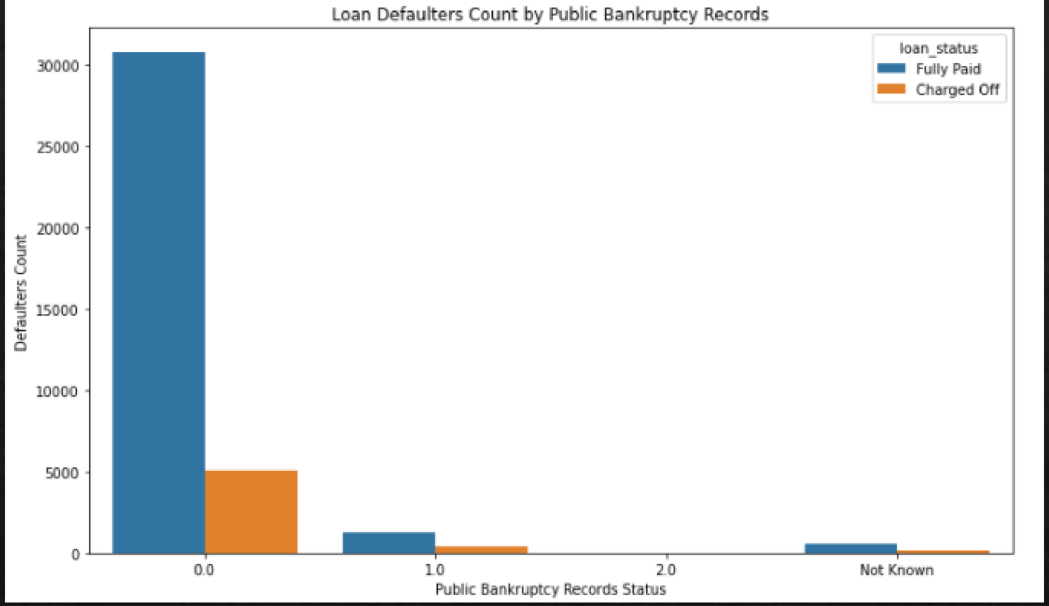


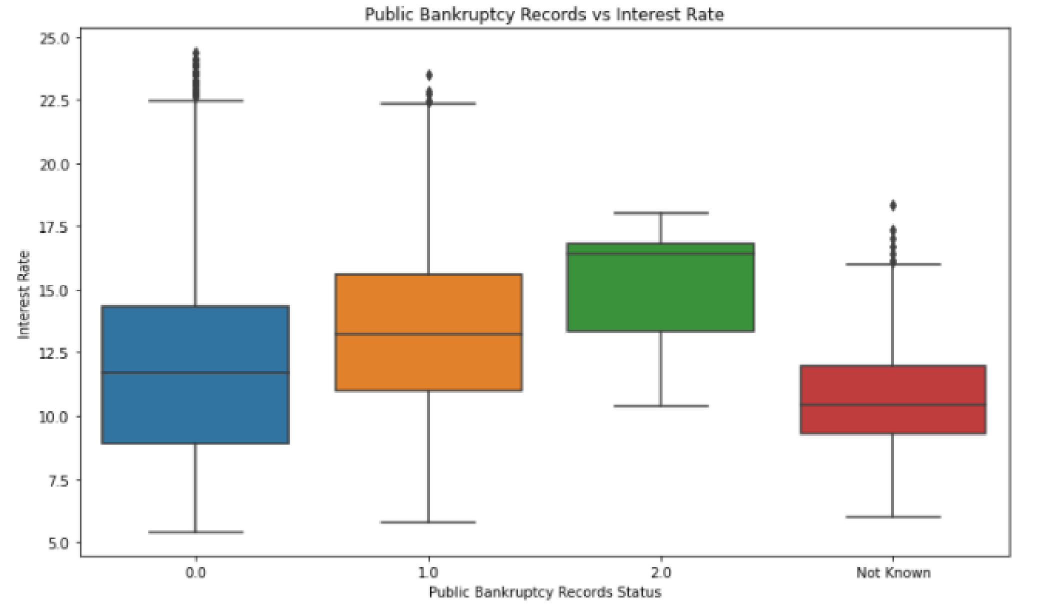


1. There is increase in default ratio with increase in interest rate.
2. The default count is highest for 12-16%.
3. Lending club should examine giving out high interest loans (16%+) as there are more defaults.

Public Bankruptcy Records on Loan Status

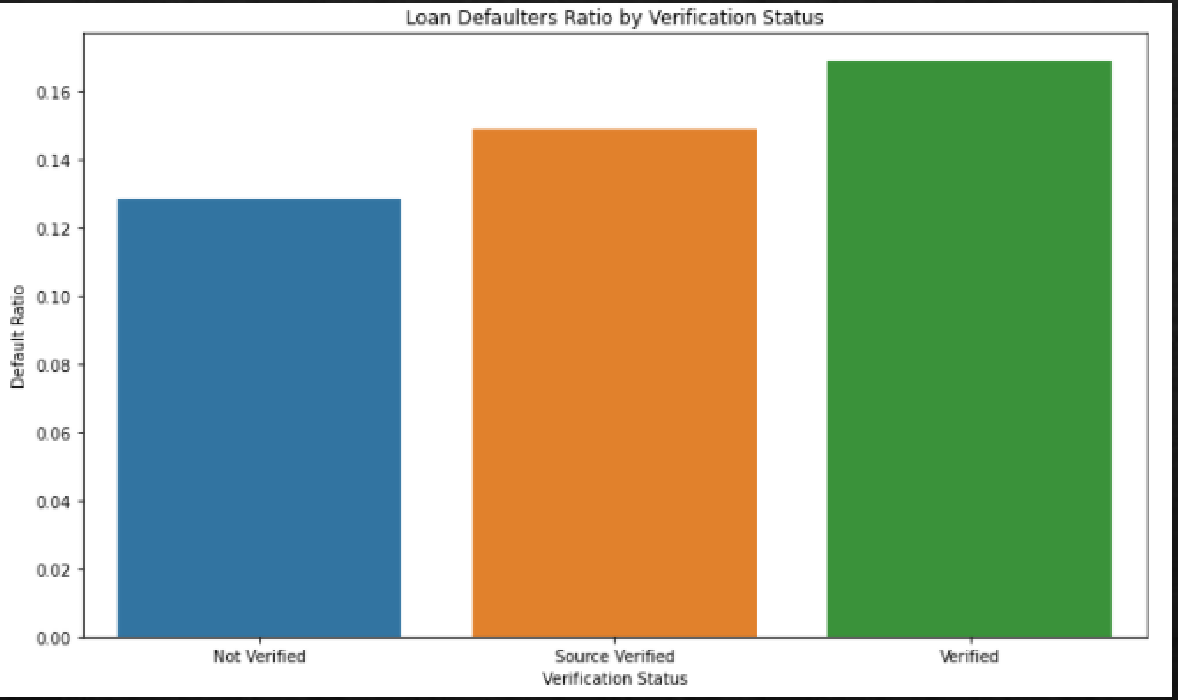


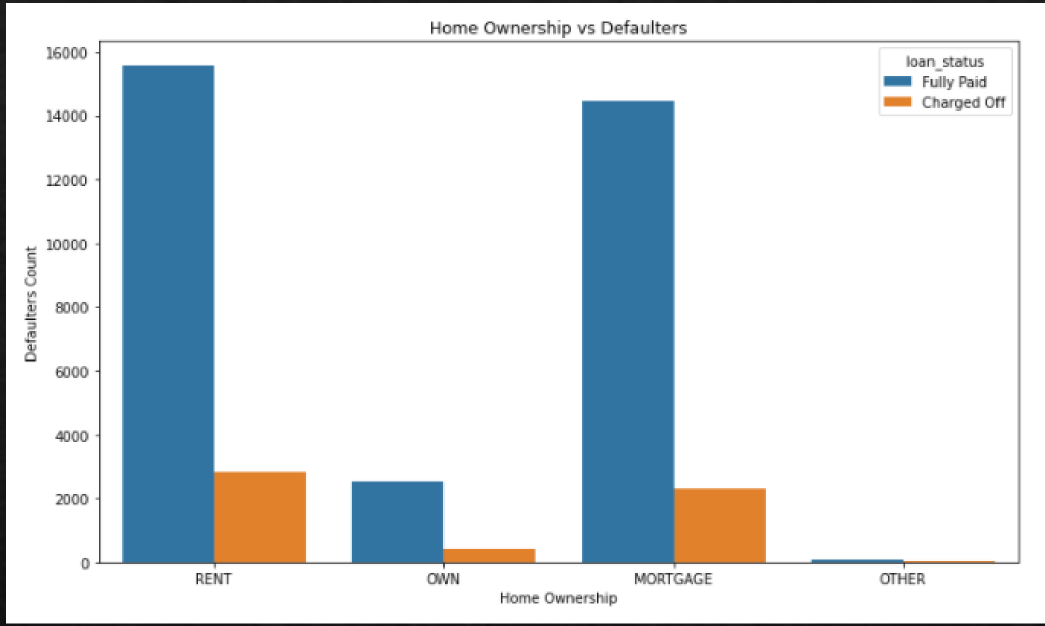




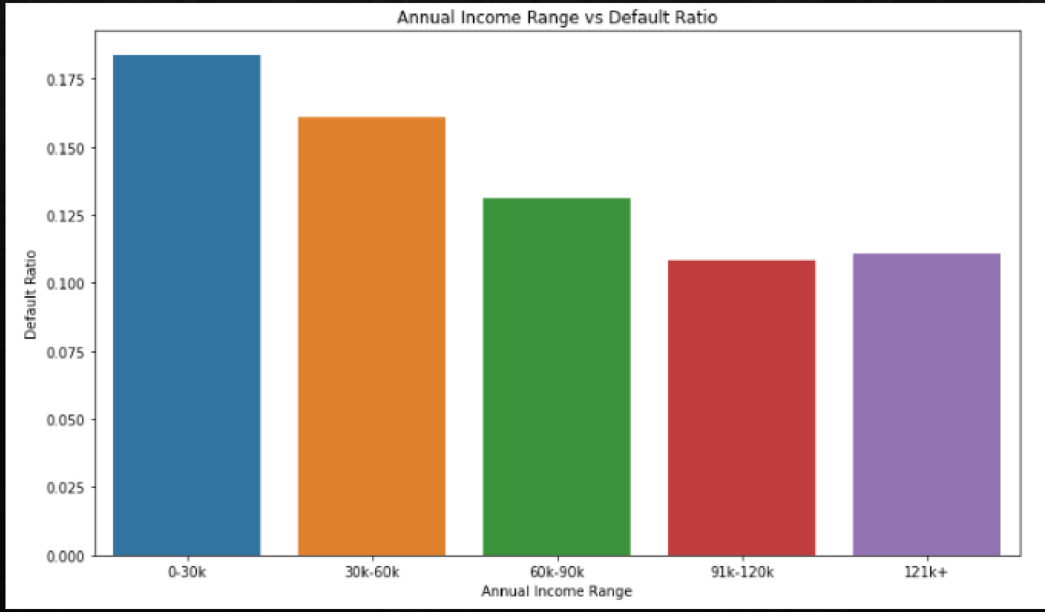
1. Default ratio increases with increase in bankruptcy record.
2. Most of the customers have zero bankruptcy record. The number is not significant enough to come to conclusion in case of people having 1 or 2 bankruptcy records. Its left to organization to decide on loan for such records.
3. Not known is the column where lending org. has not done a good job in collecting the information.
4. Interest rate also increases with increase in bankruptcy records.

Verification Status, Home Ownership and Annual Income on Loan Status

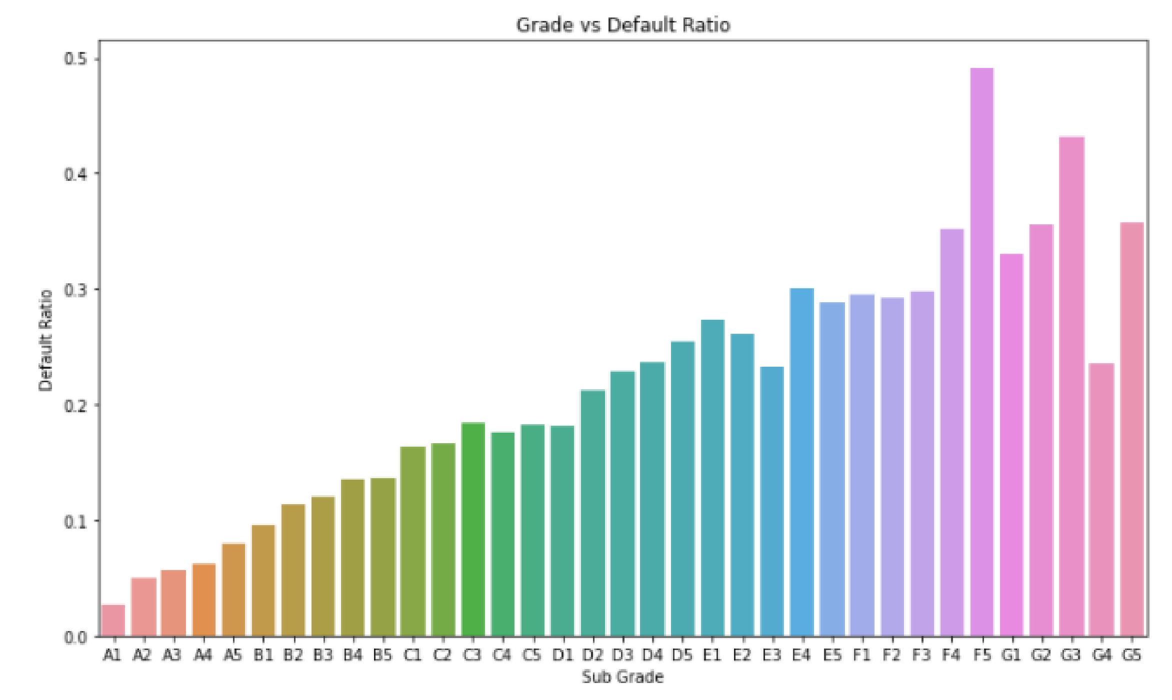


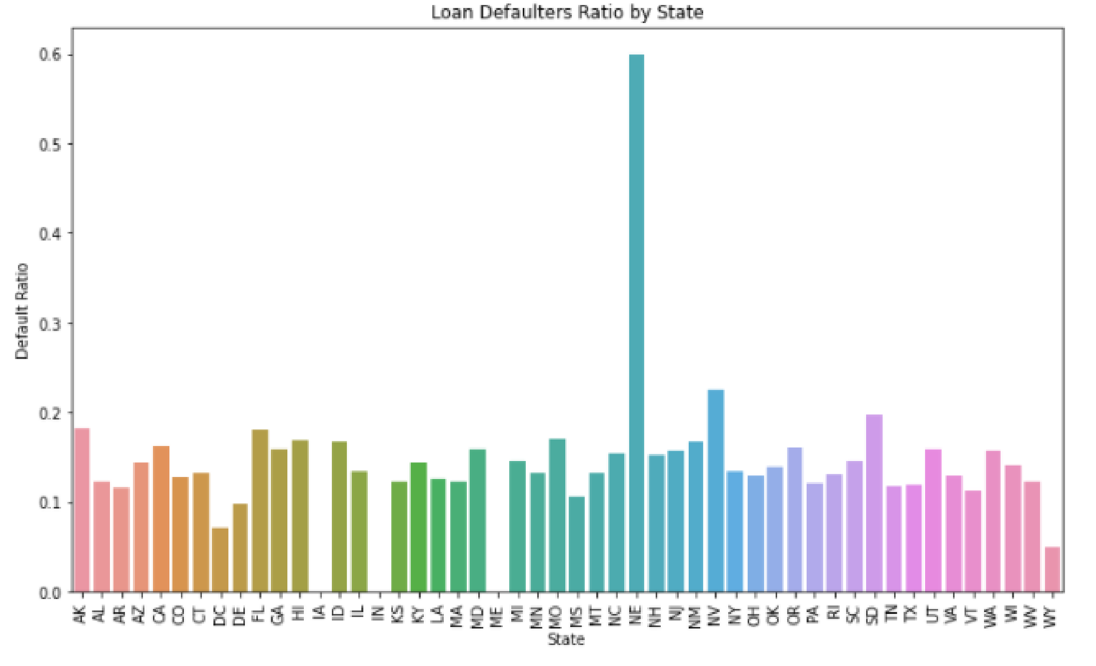


1. Verification Status- Default ratio is almost same for all three categories. This doesn't provide any insight (top left corner)
2. Most of the loans are for applicants who either rent or have a mortgage. Hence the defaults are more (top right corner).
3. Default ratio is decreases with increase in income. Financially well-off applicants are more likely to repay the loan. Bank should consider this as a factor before rolling out a loan (bottom right corner).



Sub Grade and State on Loan Status

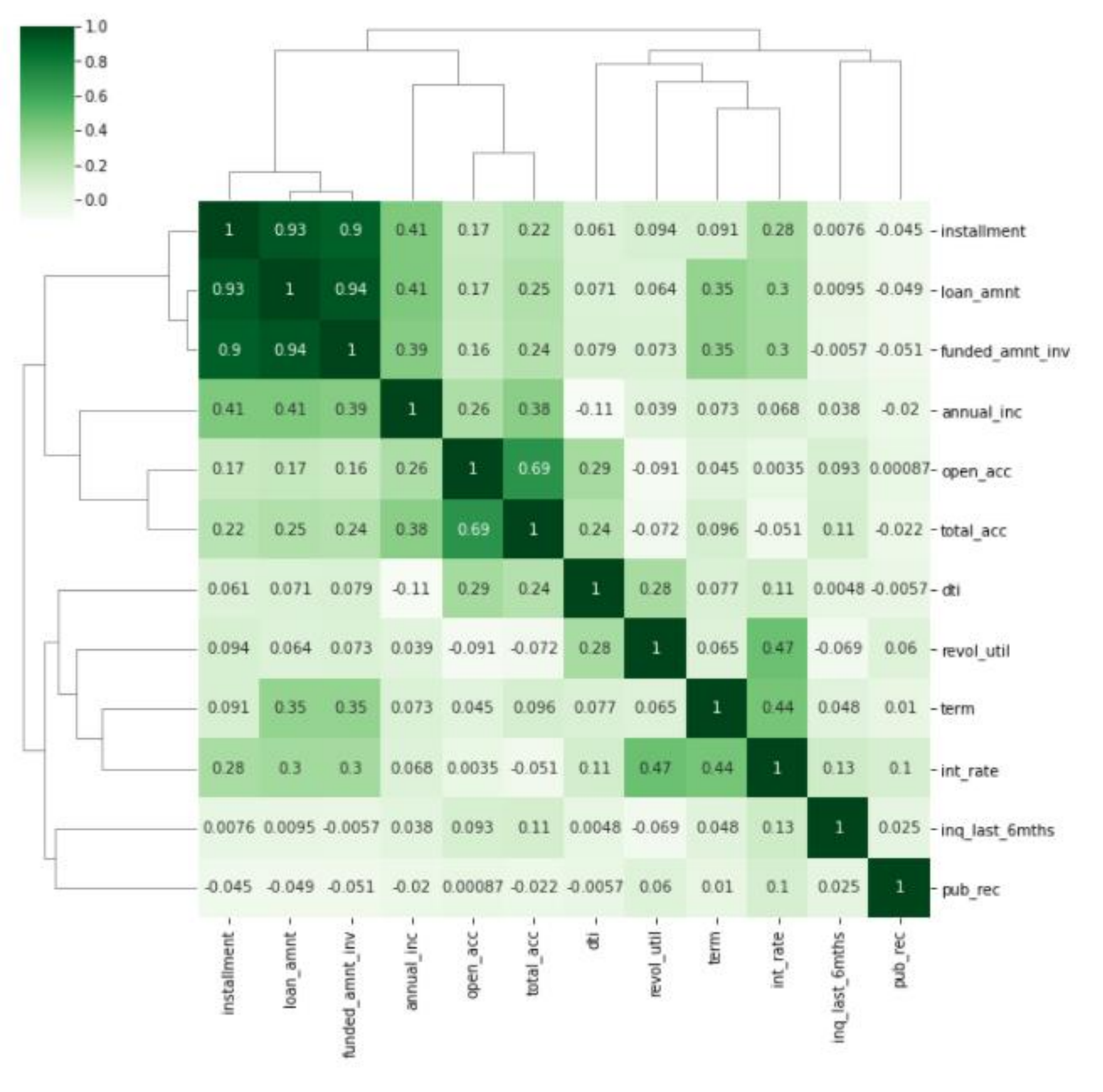




1. Loan of Sub Grade A are less likely to be charged off. Default ratio increase from A1 to F5 and varies from F5 to G5. Hence higher the sub grades higher the chances of default.
2. Default ratio is highest in the state of Nevada.
3. Florida, California and New York are also high in default ratio. Hence lending club should reconsider before rolling out the loans to applicants from these states. It should consider other factors too.

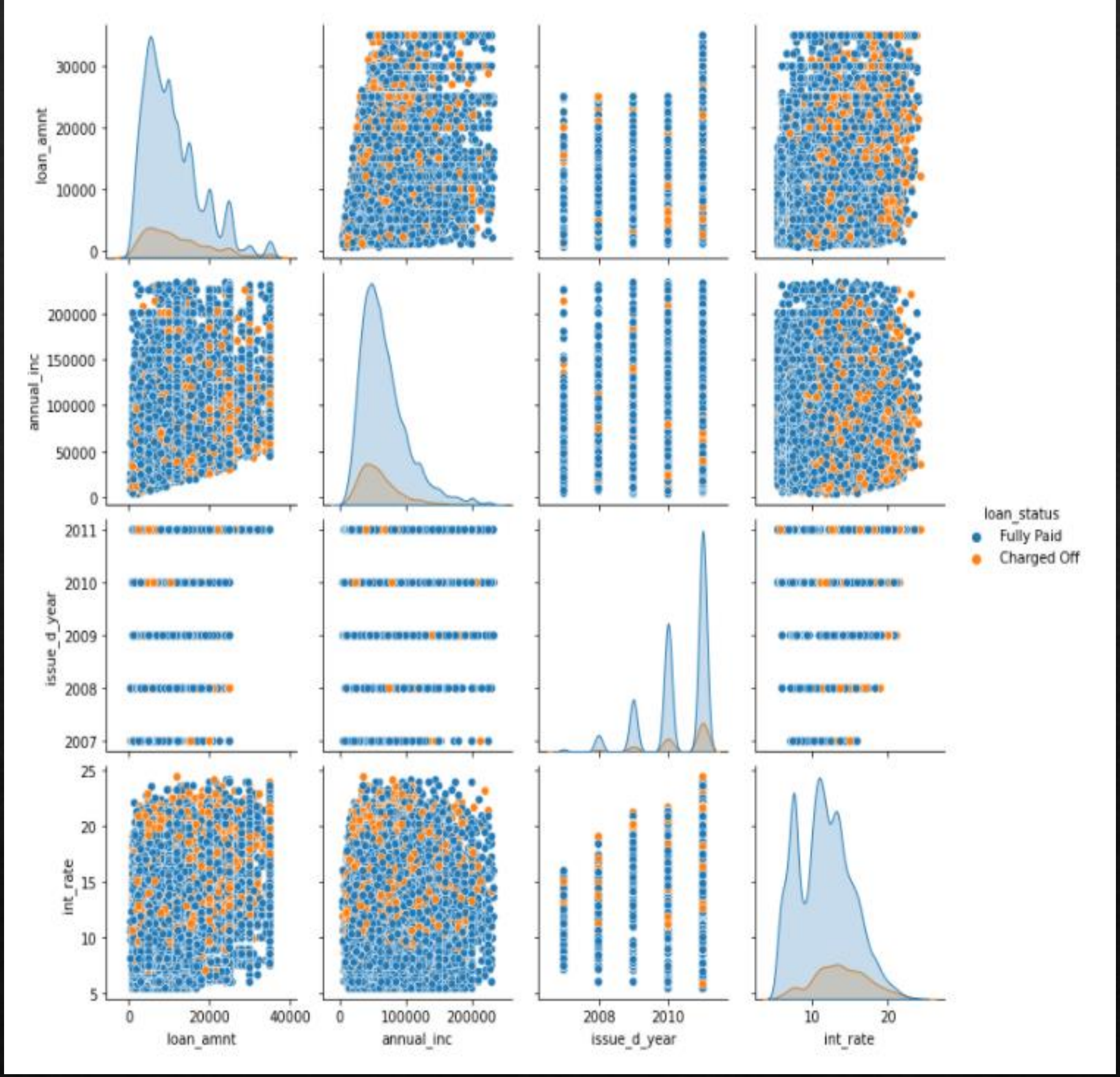
Cluster Map to Determine Closeness and Correlation between Various Loan Features.

1. Studies correlation between different variable important in loan analysis (Numeric Variables).
2. Dendrograms indicate the closeness of variables.
3. Installment and loan amount are positively correlated.
4. Loan amount and Funded amount are very closely related as the dendrogram is small.



Pair plot to check distribution of variables

1. Higher the loan amount, higher the int rate, higher defaults
2. Lower the income, higher the int rate, higher the defaults.
3. Later the issuance of loan, greater the number of defaults (not that significant).



Recommendations

Variable	Recommendations
Loan Amount	Lending up to15k is safe, anything above 15k is a risk.
Purpose	Loans for major purchase, car, credit card, wedding can be considered, whereas loans for small business is a risk.
Loan Term	Shorter term loan (36 months) are safe, whereas longer term loans are not.
Loan Grade	Grades A-C can be considered safe, anything above D should be reconsidered.
Employment Length	Not significant is determining the defaults
Debt-to-Income Ratio	Dti range 0-10 is safe, above 15 is a risk.
Loan Interest	Loan interest ranging from 0-8% can be considered safe. Interest rate above 15 is a risk.
Public Bankruptcy	Zero bankruptcy records is safe, while 2 or above is a risk.
Home Ownership	People who own home are safe whereas people who rent and have a mortgage are risk factors.
Annual Income	Annual Income ranging 90k and above are considered safe, whereas as 0-60k are risk factors.