

The background features a dark blue gradient with several overlapping, semi-transparent financial charts. On the left, a line chart with white circular markers and orange centers is visible. In the center, a bar chart with blue bars is partially shown, with a data label '289.33' next to one of the bars. On the right, another bar chart with blue bars is visible, with a data label '894' above one of the bars. The overall aesthetic is modern and professional, typical of a business presentation.

LENDING CLUB CASE STUDY

Group:

Sourav Kumar Nag and Debojyoti Mallick

Lending Club Case Study - Business Understanding

You work for a consumer finance company which 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. 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.

When a person applies for a loan, there are two types of decisions that could be taken by the company

- Loan accepted: If the company approves the loan, there are 3 possible scenarios described below
 - Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
 - Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
 - Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan
- Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Lending Club Case Study - Business Objectives

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.

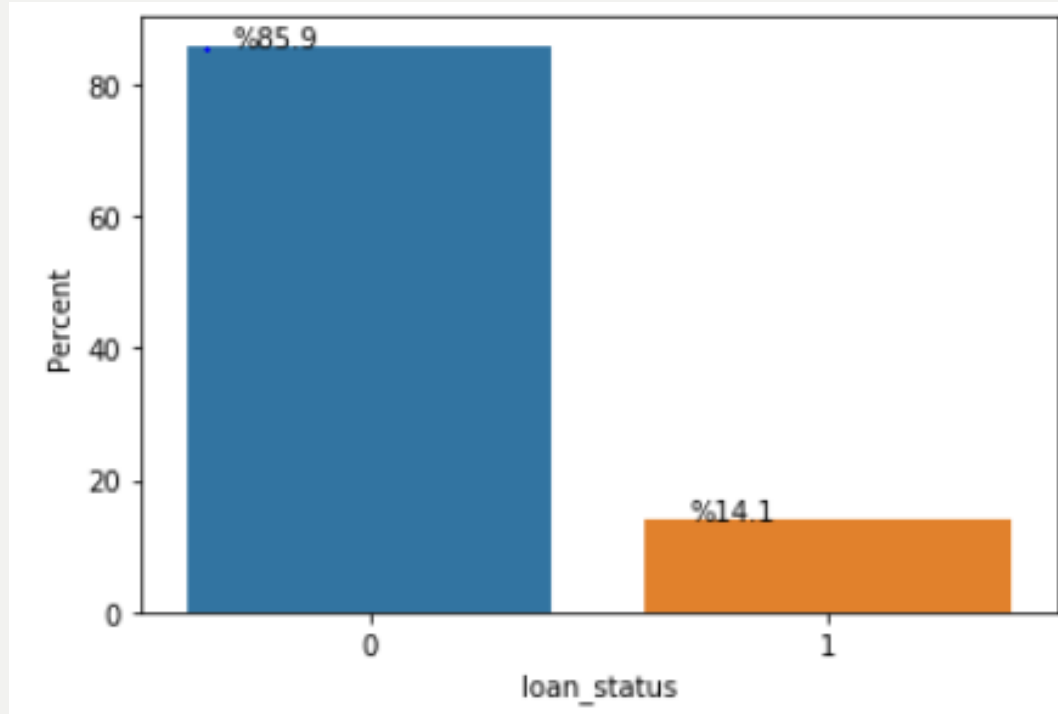
Data Provided:

1. Loan Data Set - Contains Loan data for for all loans issued through the time period 2007 to 2011.
2. Data Dictionary - Description of the fields in the loan data sheet.

Data Cleaning and finalization

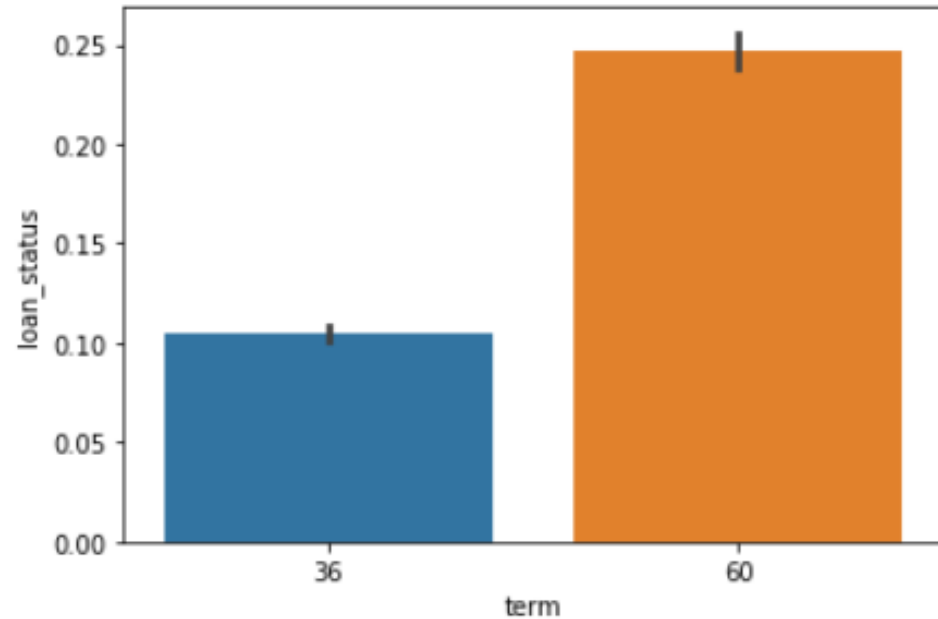
- The first step was to analyze and understand the data
- Purpose of this step was to identify the most important data from the dataset and remove irrelevant data, impute or drop fields that has NA values, change datatypes where required, derive fields if needed, remove unwanted redundant expressions from certain fields
- The variables were broadly classified under 3 categories:
 - *Loan characteristics variables*
 - *Applicant characteristics variables and*
 - *Applicant behavioral variables*
- Out of the above 3 types of variables we only will concentrate on the first 3 types since applicant behavioral variables will impact the loan once the loan is approved and on-going, hence are irrelevant for our case study analysis
- 3 types of analysis is also done on the variables to arrive at conclusions
 - *Univariate analysis*
 - *Segmented univariate analysis*
 - *Bivariate analysis*

Loan Status



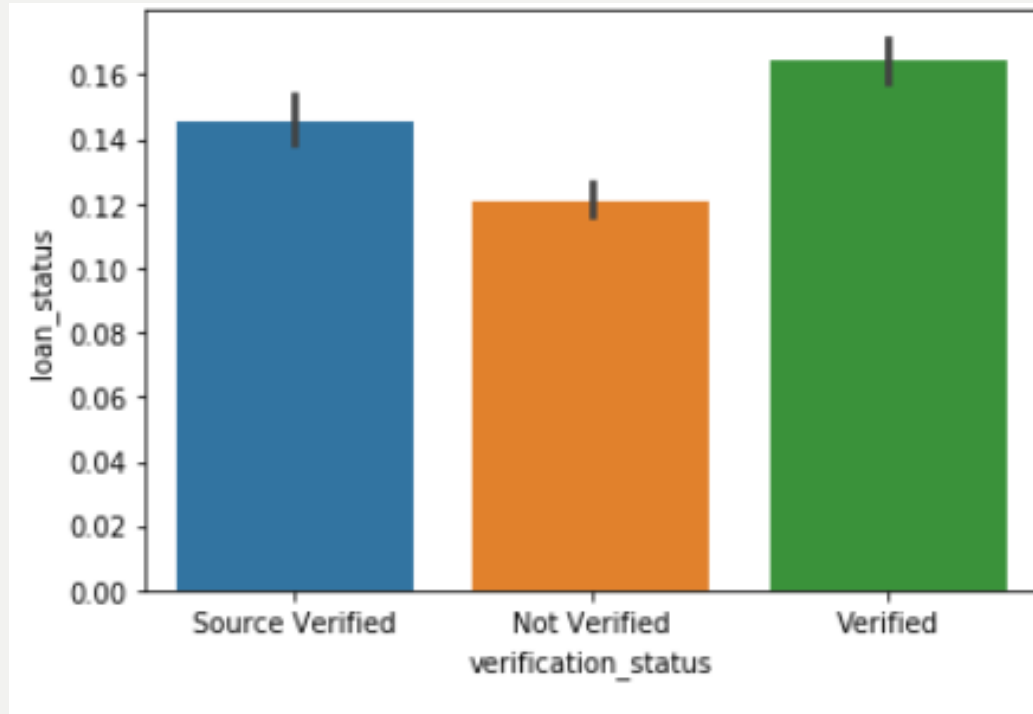
- From the given dataset we see that the overall percentage of loans that are fully paid is 85.9%
- We further see that the overall percentage of loans that has defaulted is 14.1%

Univariate Analysis – Term vs Loan Status



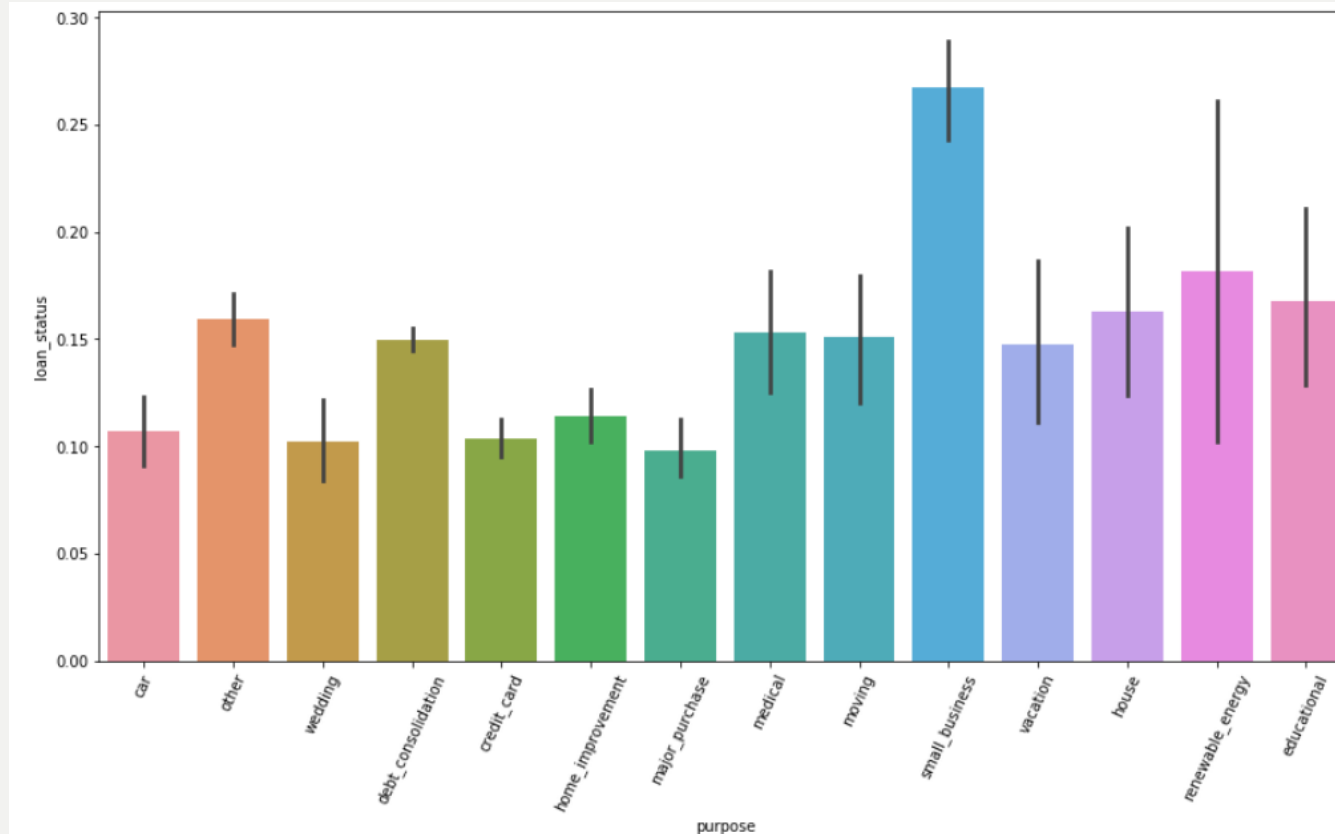
- Loan of term 60 months tend to default more than those of 36 months
- Default rate for loans of term 36 months is less than the mean default rate whereas the default rate of loan of 60 months is much higher than the mean default rate

Univariate Analysis – Verification Status vs Loan Status



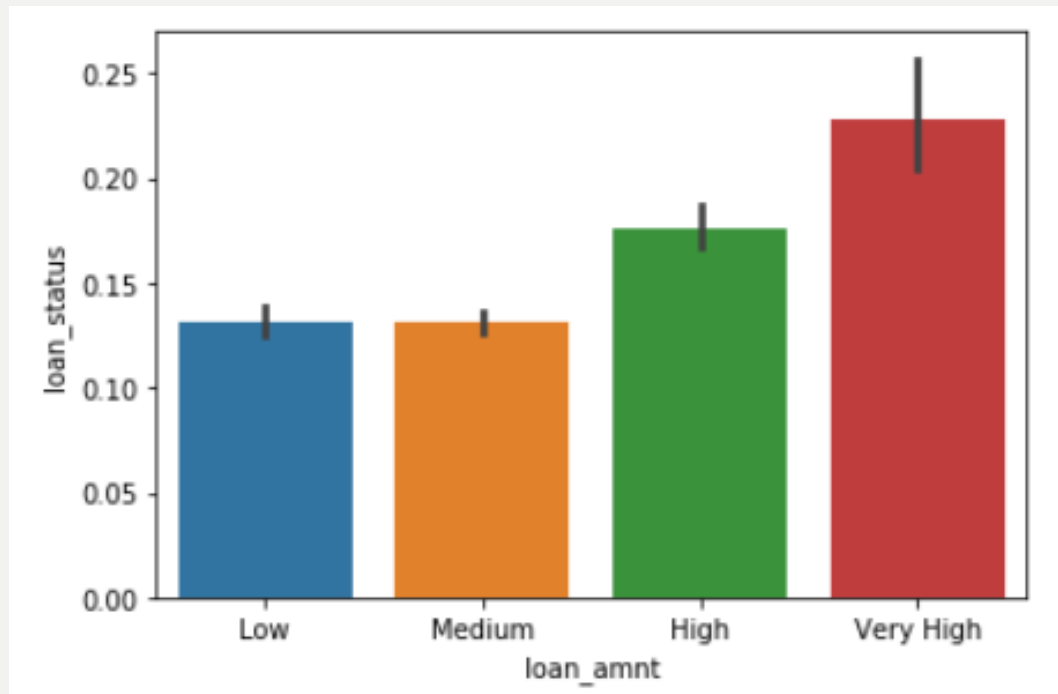
- Loans taken by applicants whose income was verified by LC has a higher default rate than the ones whose income was not verified.
- Default rate of income not verified applicants is lower than mean default rate whereas the same for verified income applicants is higher than the mean rate

Univariate Analysis – Purpose vs Loan Status



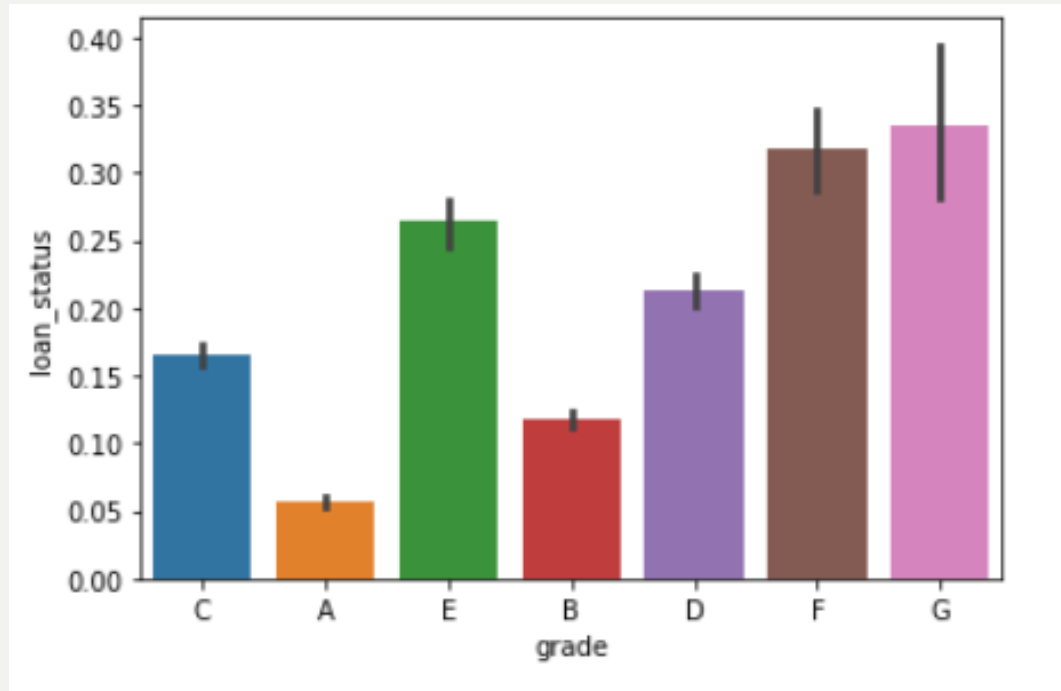
- Loan taken for small_business will likely to default while loan taken for major_purchase has least chance of getting defaulted
- The top 5 purpose that usually is seen to default:
 1. Small Business
 2. Renewable Energy
 3. Educational
 4. House
 5. Other

Univariate Analysis – Loan Amount vs Loan Status



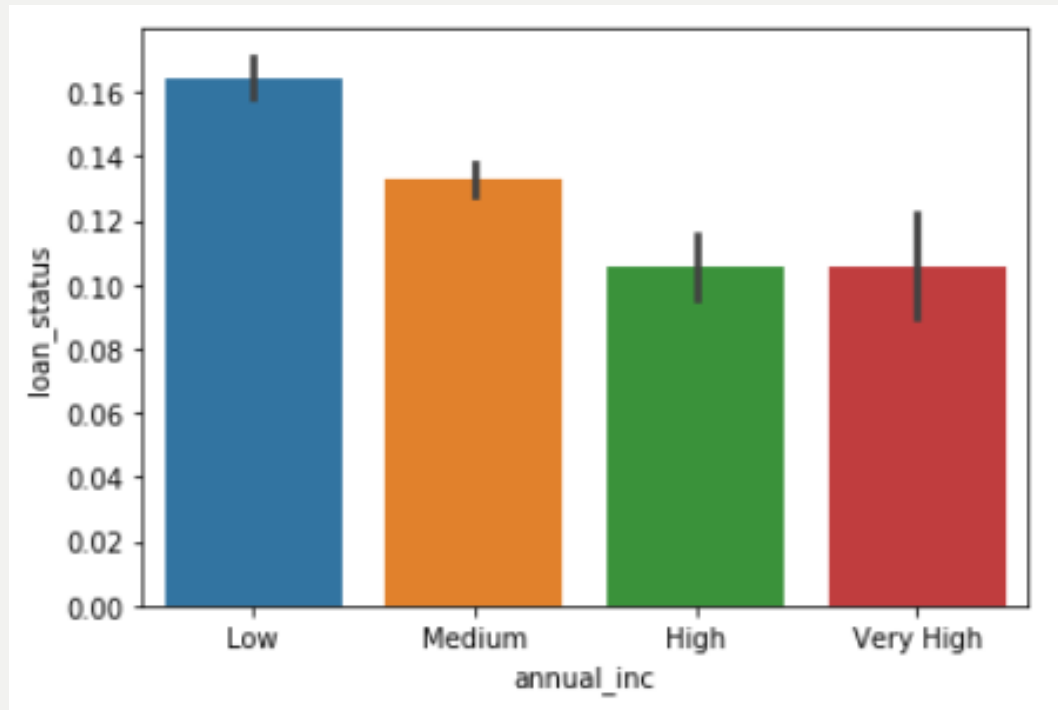
- We have binned the loan amount as follows:
 1. Low < 6000
 2. Medium 6000 to 18000
 3. High 18000 to 30000
 4. Very High > 30000
- It is clear that default rate increases with the increase in loan amount

Univariate Analysis – Loan Grade vs Loan Status



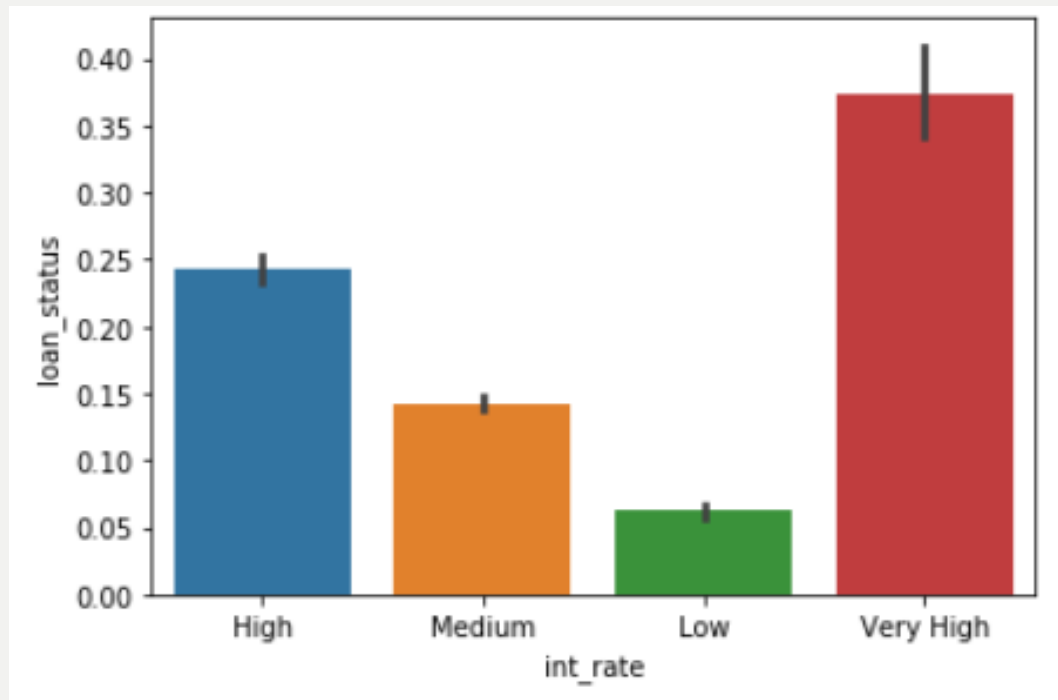
- From the graph it is clear that default ratio is directly proportional to the grade of the loan, the higher the grade more is the default rate
- Highest default rate is at loan grade: G
- Lowest default rate is at loan grade: A

Univariate Analysis – Annual Income vs Loan Status

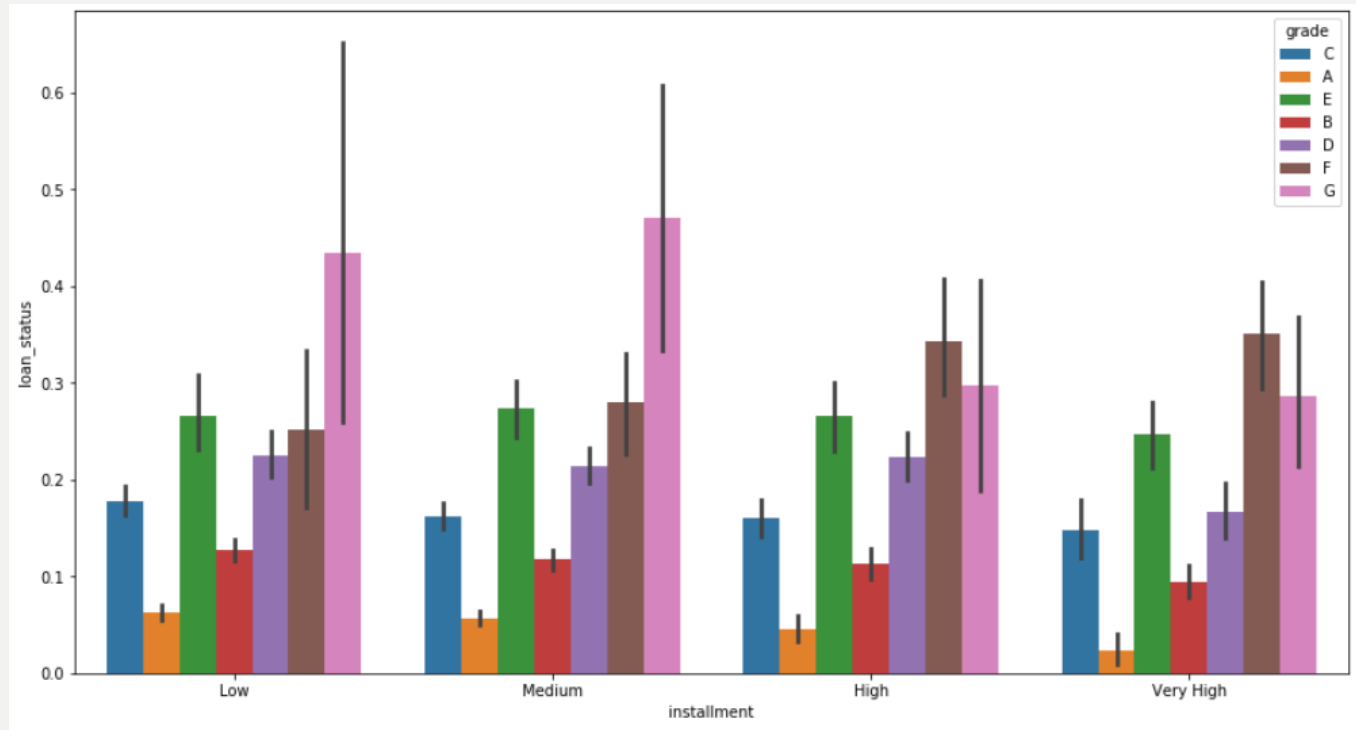


- We have binned the applicant's annual income amount as follows:
 1. Low < 50000
 2. Medium 50000 to 100000
 3. High 100000 to 150000
 4. Very High > 150000
- It is clear that default rate increases with the decrease in applicant's annual income

Univariate Analysis – Interest Rate vs Loan Status

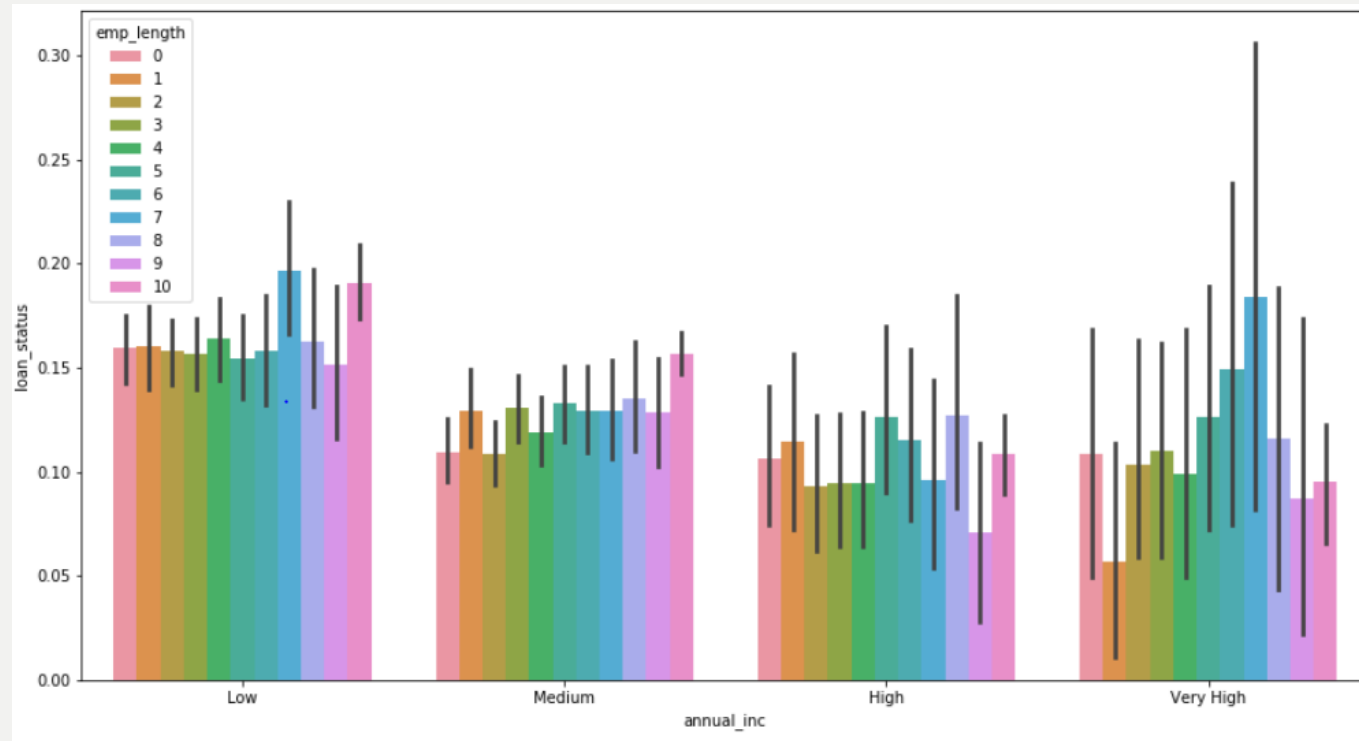


- We have binned loan's interest rate as follows:
 1. Low < 10
 2. Medium 10 to 15
 3. High 15 to 20
 4. Very High > 20
- It is clear that default rate increases with the increase of loan interest



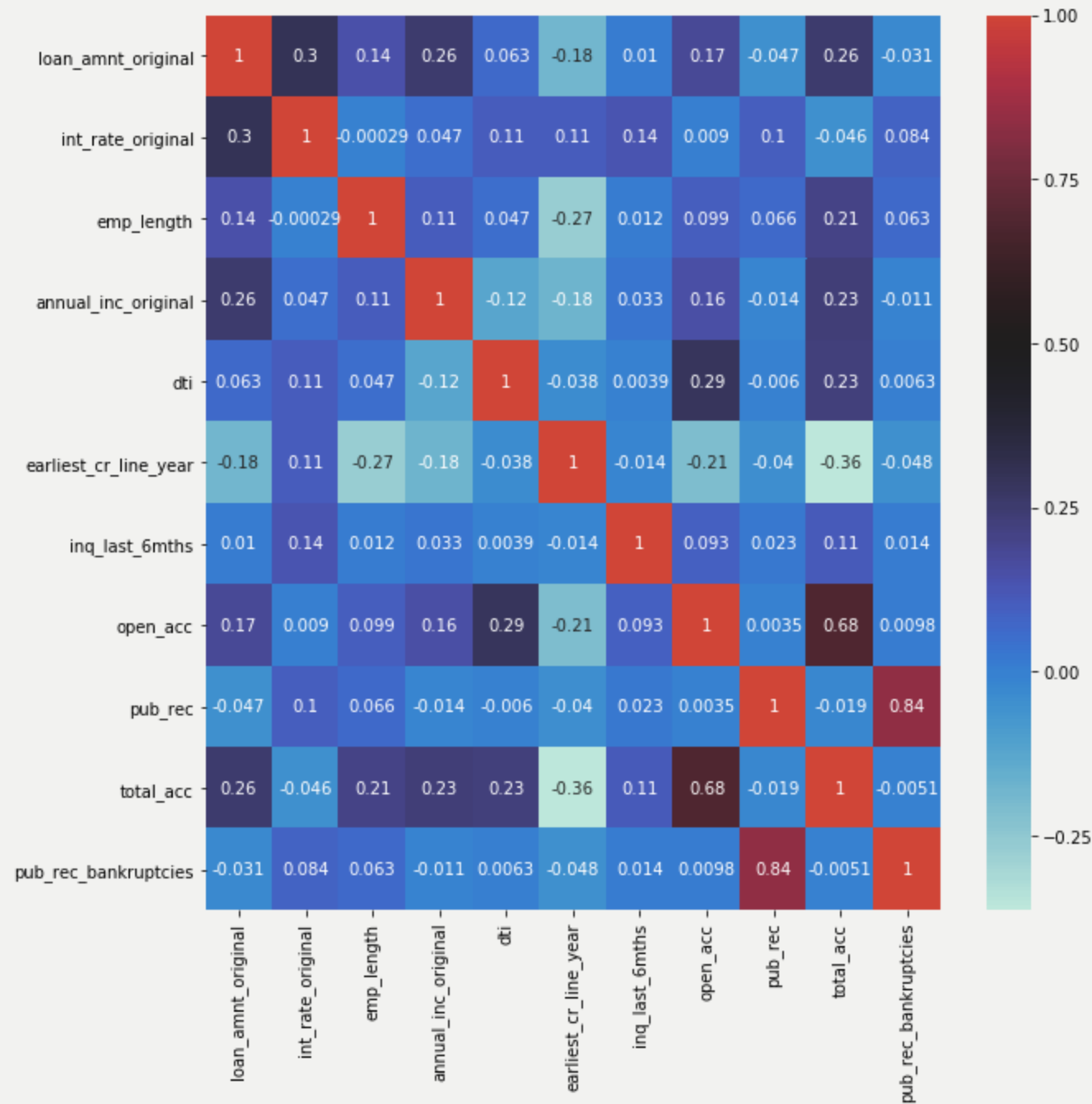
Segmented Univariate Analysis – Installment vs Loan Status segmented by grade

- It is clear that when installment amount is in the low and medium range (less than 400/month), maximum default happens in grade 'G' type of loan.
- For installment in the higher ranges (greater than 400/month), maximum default happens in grade 'F'



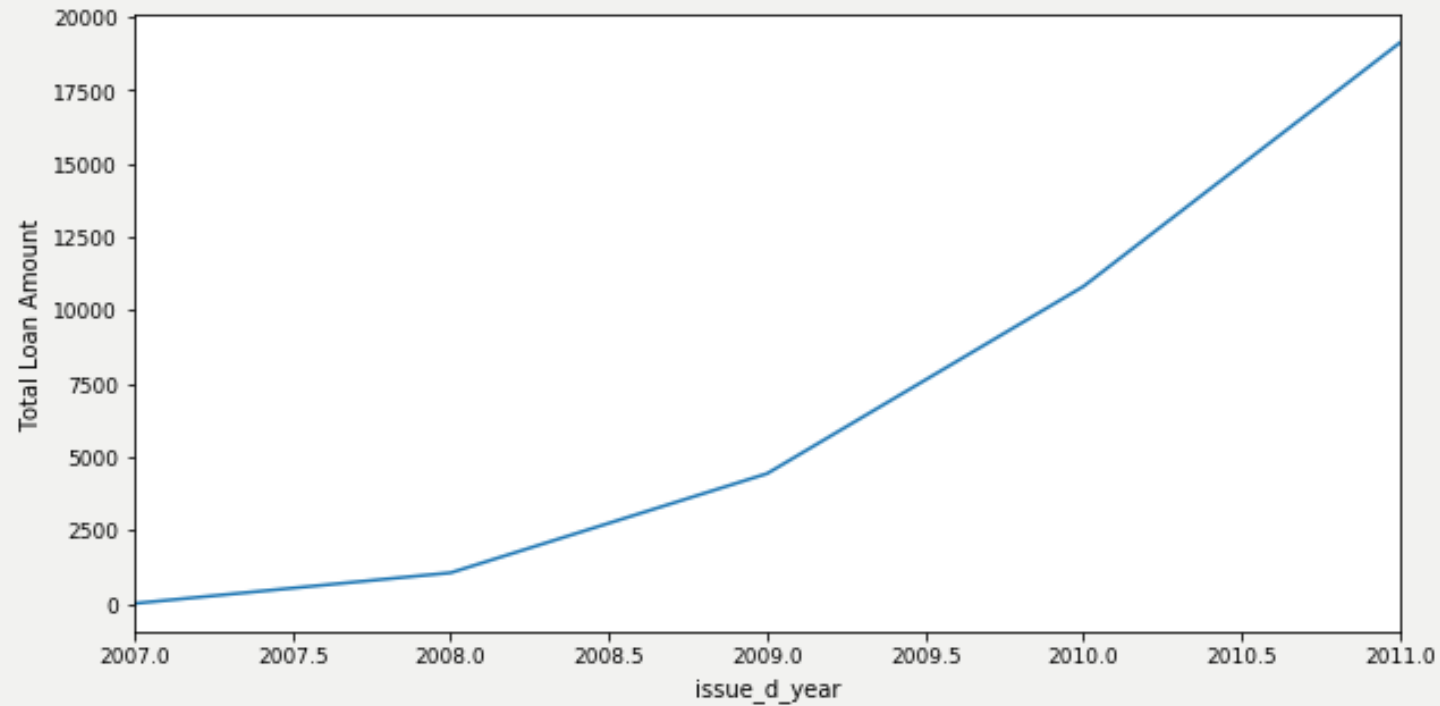
Segmented Univariate Analysis – Annual Income vs Loan Status segmented by length of employment

- Applicants having 7 to 10 years of experience in the low income group tend to default more while applicants having the same number of years of experience in the high or very high income group default much less.
- We cannot conclude about applicants with other number of years of experience as the graphs show a steady pattern.



BI - Variate Analysis

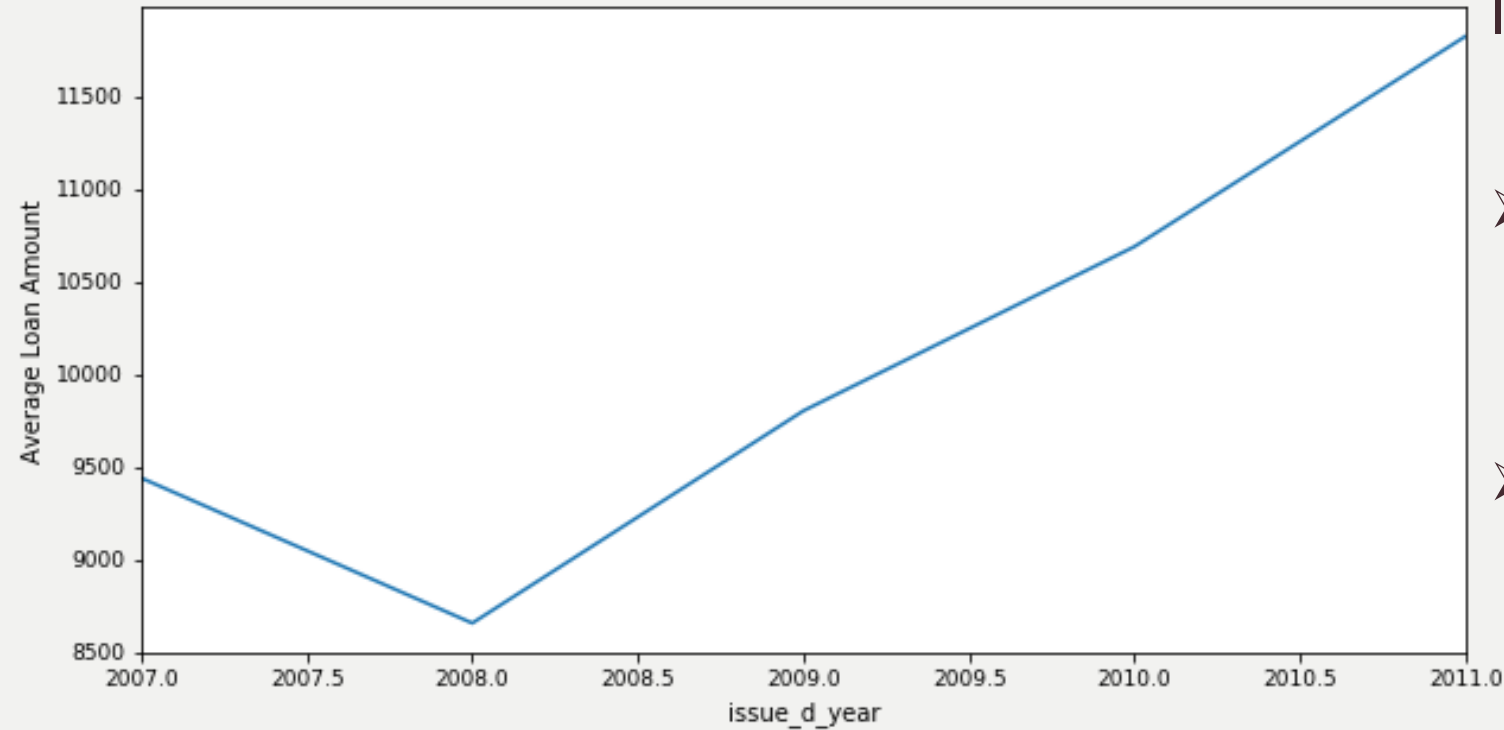
- Public Record Banck and Public records are very highly correlated.
- Total Account and Open accounts are highly correlated.



BI - Variate Analysis Total No Loan Amount vs Issue Date(year)

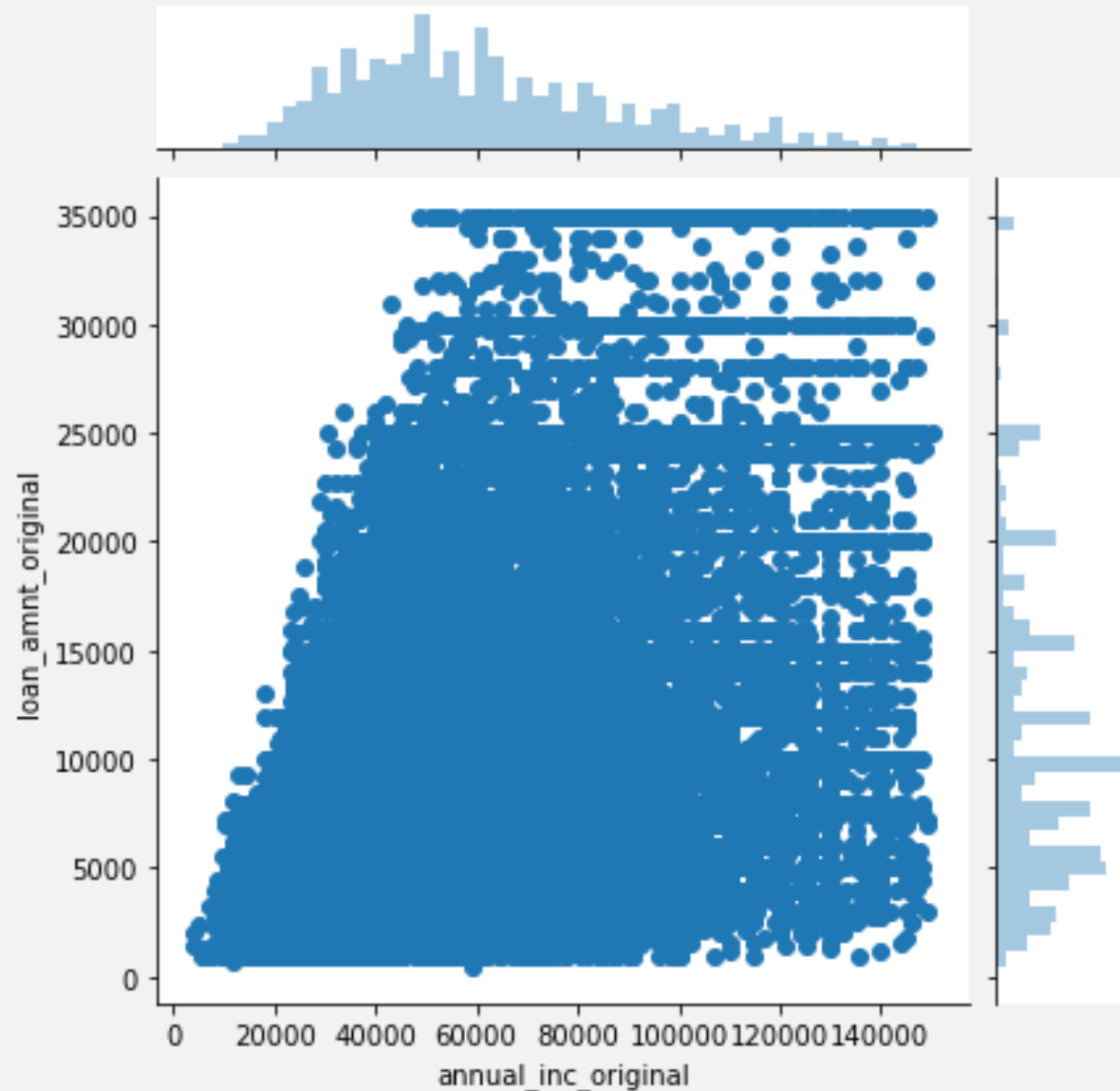
- Form the graph its evident that there is a dip in the loan amount being issued around year 2008.
- There is a steady increase in the loan issued from 2009 onwards.

BI - Variate Analysis Loan Amount Mean vs Issue Date(year)

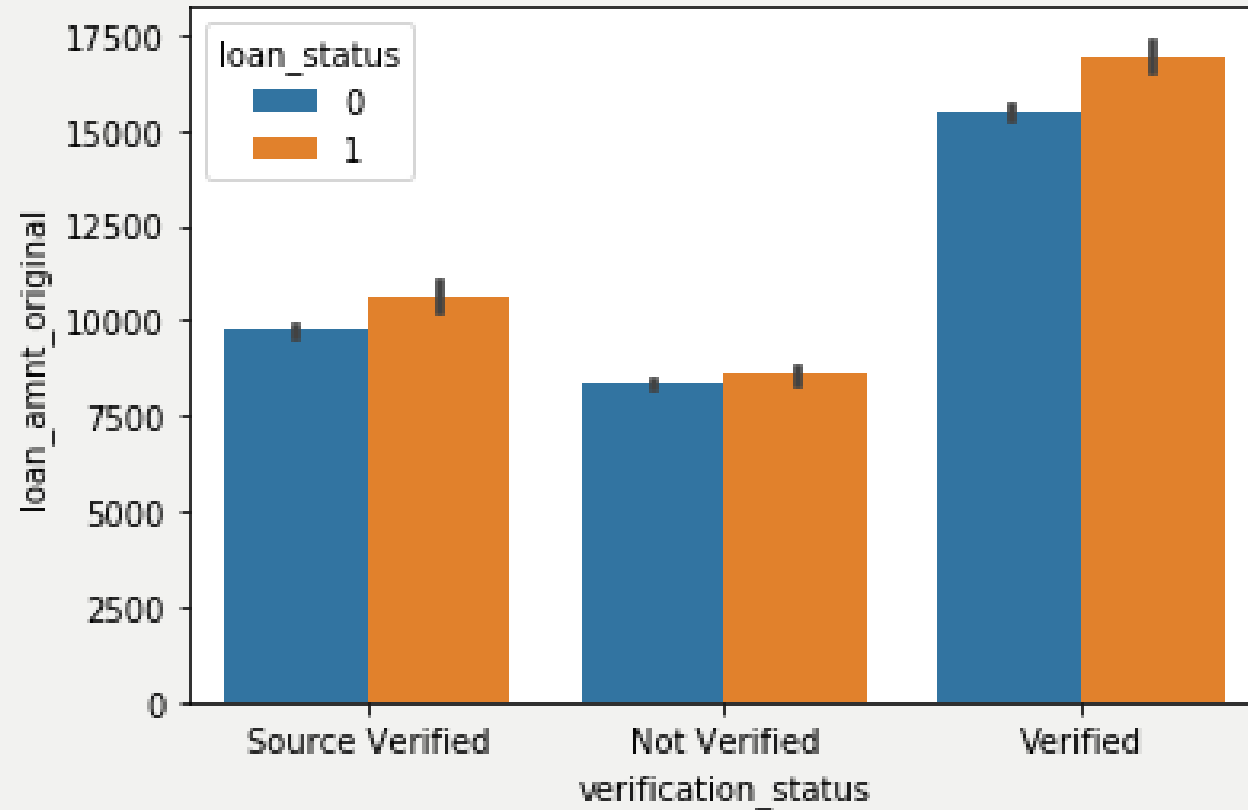


- Form the graph its evident that there is a dip in the loan amount being issued around year 2008.
- There is a steady increase in the loan issued from 2009 onwards.

BI - Variate Analysis Annual Income vs Loan Amount



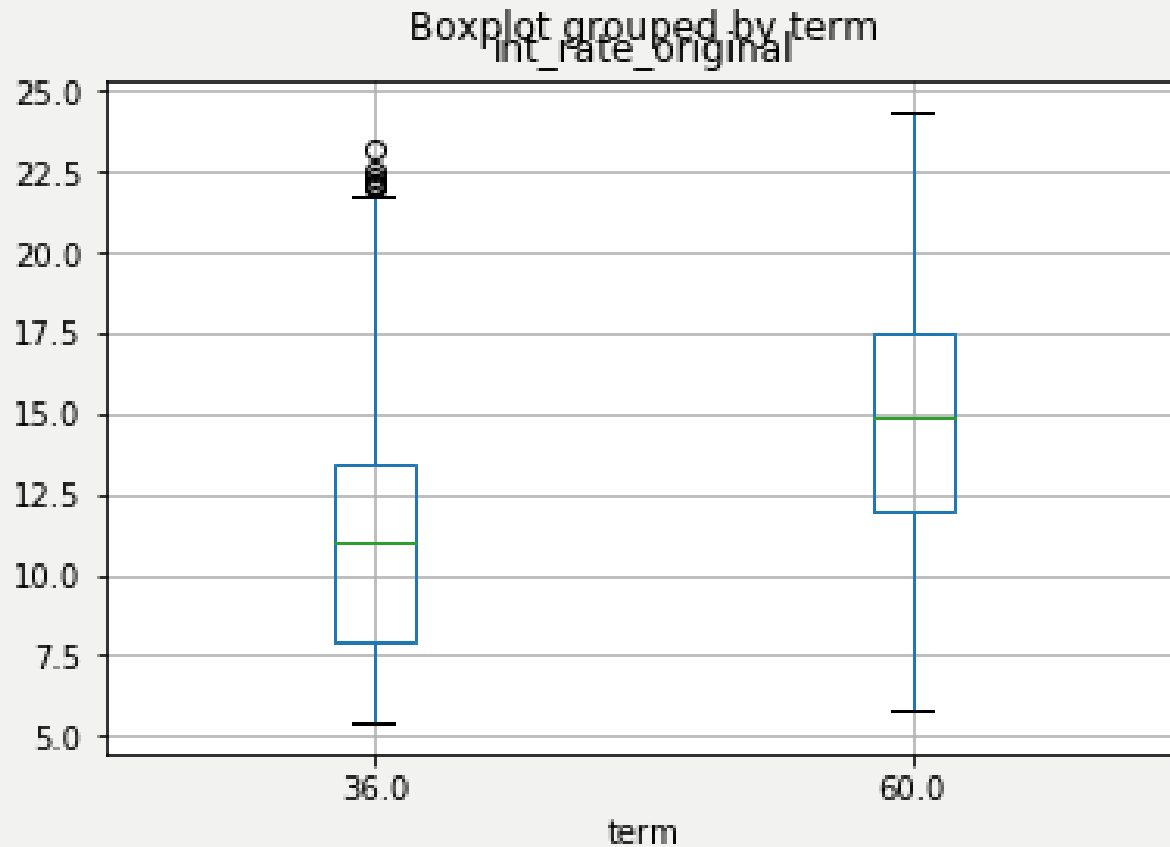
- The riskiest loans are the ones which are offered to customers with annual income between 50k to 80k with loan amount sanctioned of 35k.



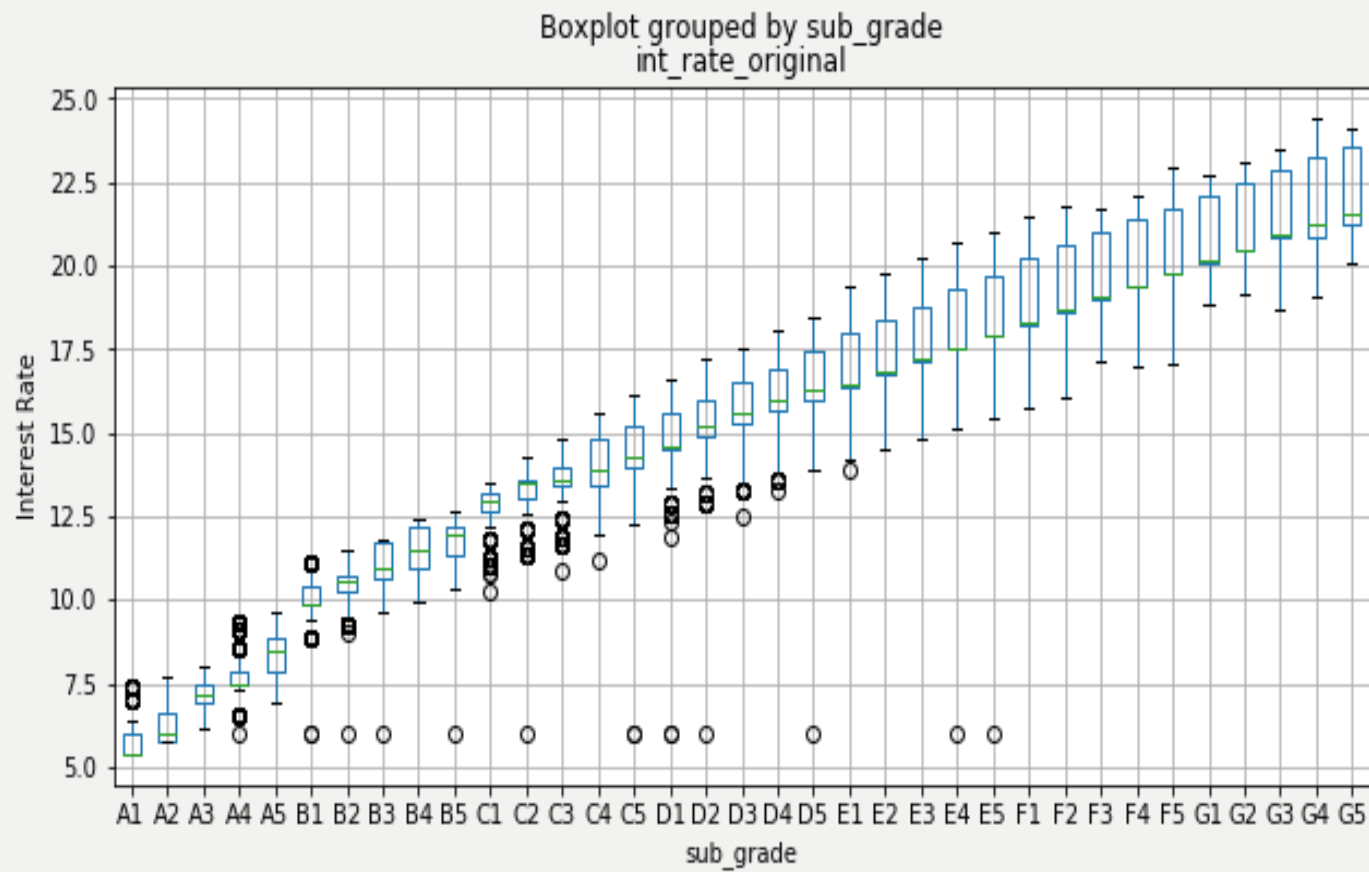
BI - Variate Analysis Loan amount vs Loan status vs Verification Status

- Higher amount loans are also verified which is expected.
- Also Verified loans are the ones which have defaulted more.

BI - Variate Analysis Interest Rate vs Term

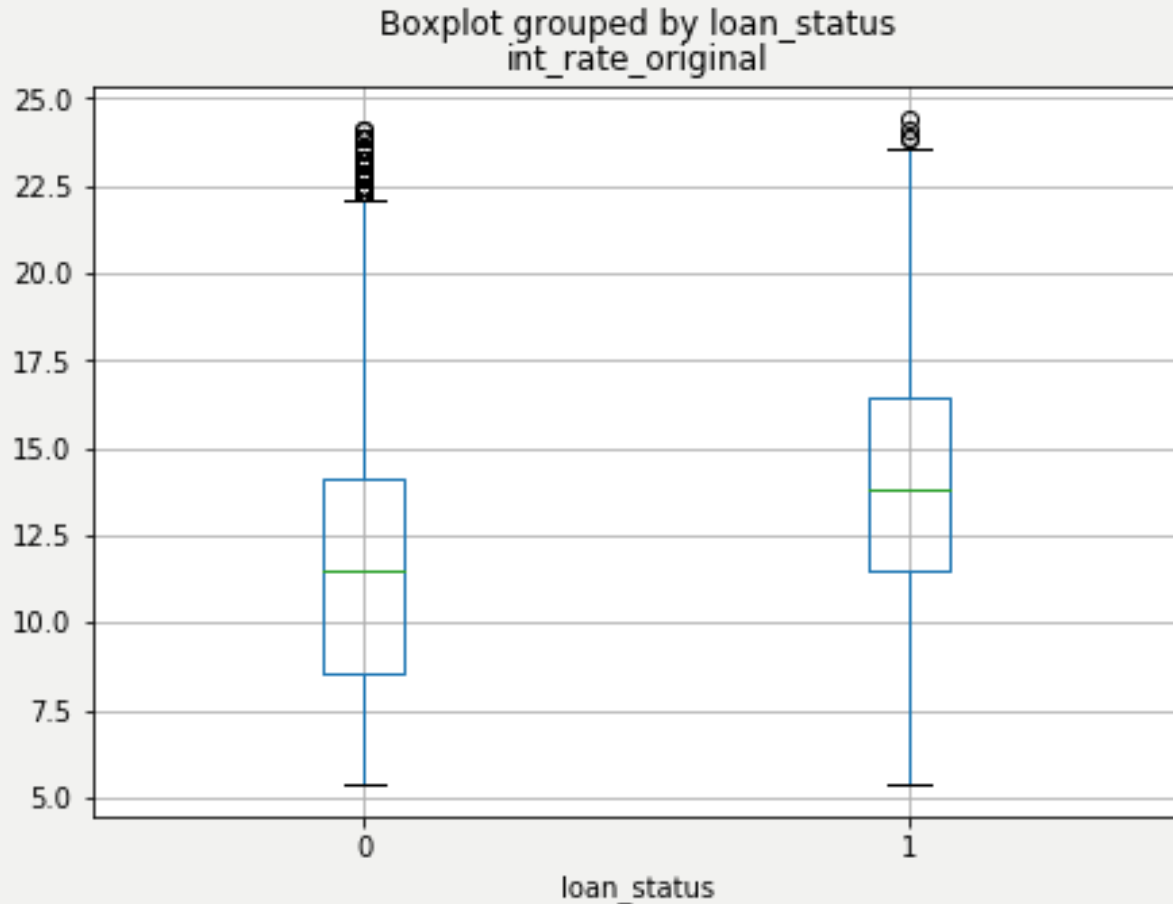


- Loans with longer term have more interest rates.
- 36 month term loans have clearly some outliers who have very high interest rates. Probably the customers with previous default history.
- They should be the most risky customers.



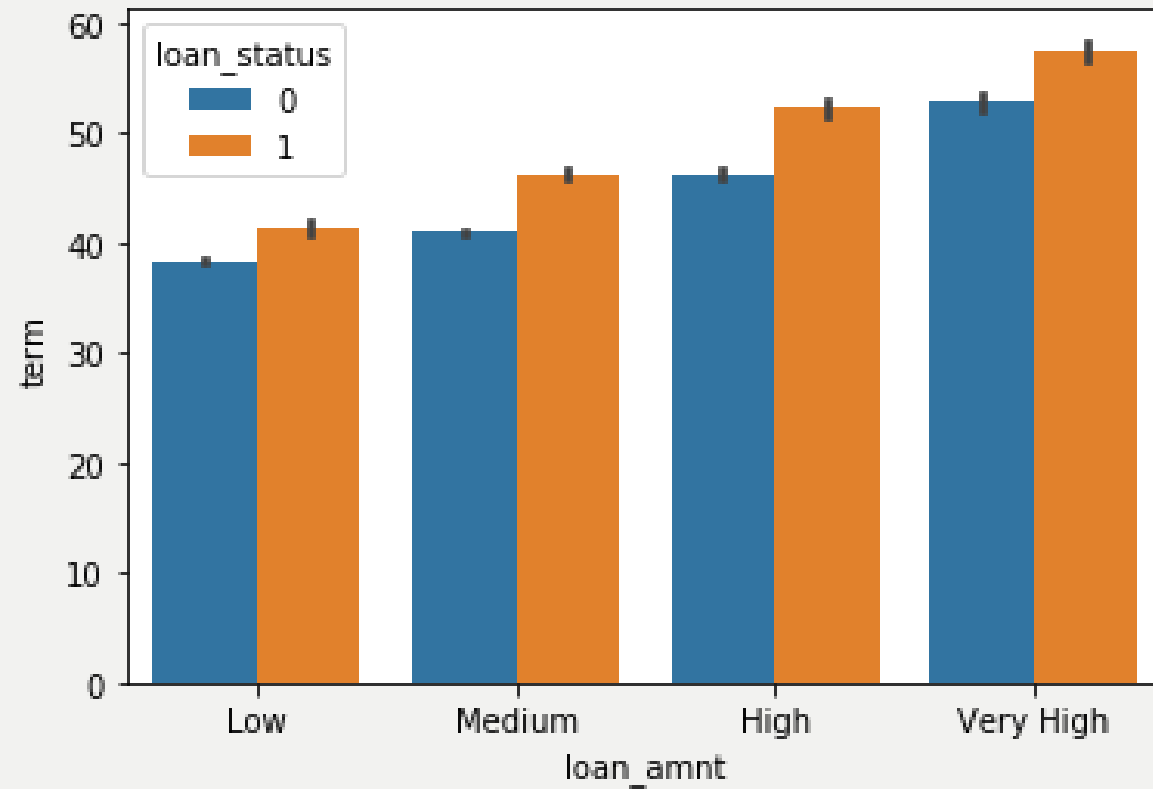
BI - Variate Analysis Sub Grade vs Interest Rate

- Lower the grades the higher the interest rate goes.
- Higher grades have a less spread of the interest rate within the same grade.
- Lower grades have high spread of rate within the same grade.



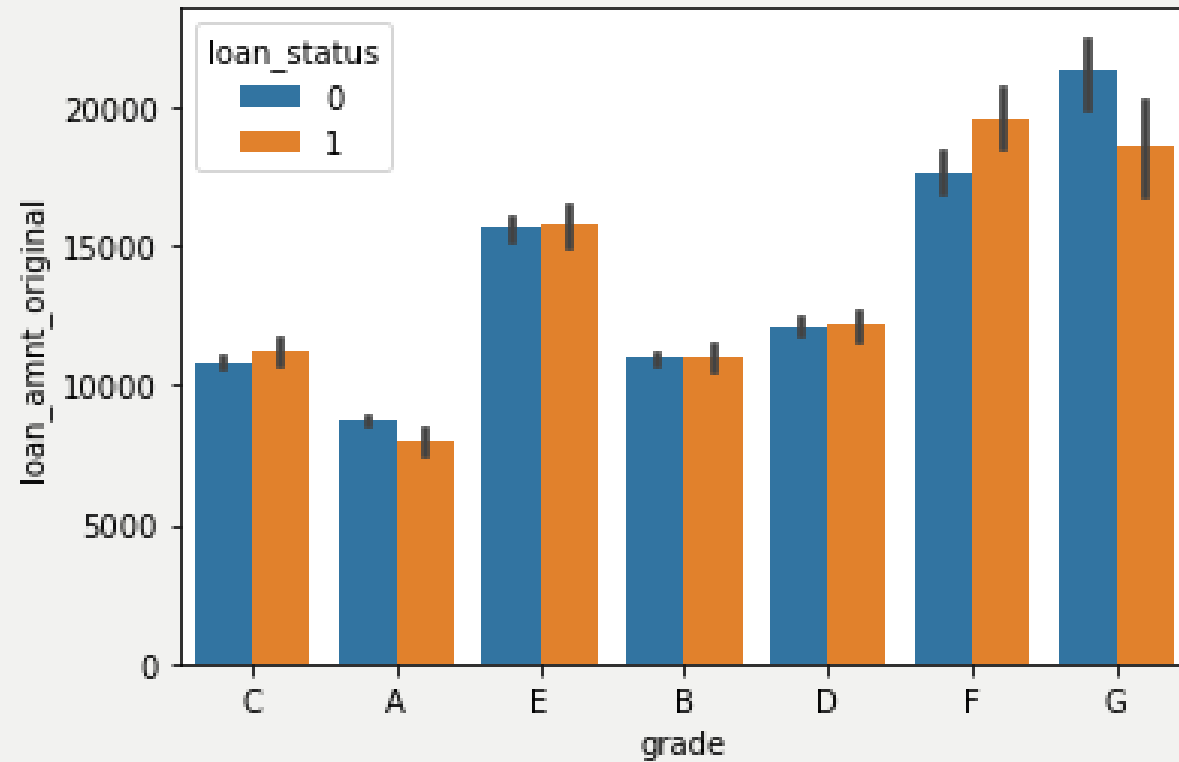
BI - Variate Analysis Interest Rate vs Loan Status

- Paid Off loans have lower interest rates.
- Defaulted loans have higher interest rates.
- For paid off loans some outliers exist which have a very high interest rate. These might be some very risky customers whom loans were given and they still paid off on time.



BI - Variate Analysis Term vs Loan Amount vs Loan Status

- Very high value loans are riskier and gets more defaulted.



BI - Variate Analysis

Loan Amount vs Grade vs Loan Status

- High Value loans offered to lower grade customers(F) got more defaulted.

Recommendations:

- 1. Reduce Number of approvals where purpose in Small Business.
- 2. Approve less loans where Amount to Income ratio goes beyond 28%
- 3. Stop approving high value loans where utilization rate greater than 75%
- Be very cautious when approving loans to people with previous bankruptcy records.
- Start verifying all loans if possible.