Lending Club Case Study

Submitted by:

- Khyati Desai
- Utkarsha Shripanavar

Problem Statement

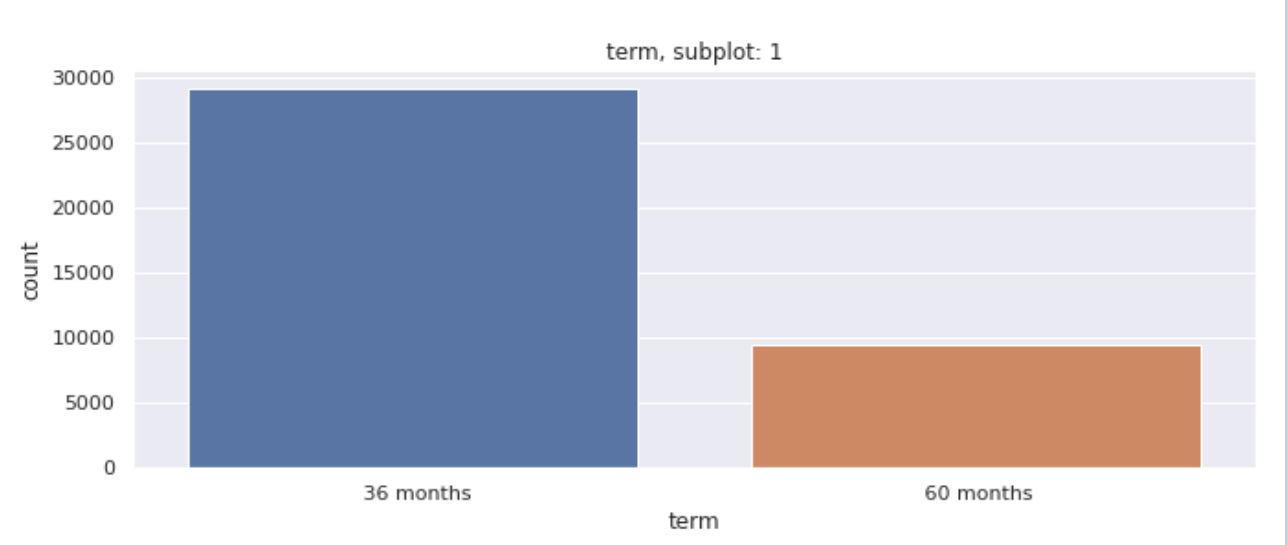
- **What is Lending Club?**
  - Lending Club is like a mediator who matches a borrower seeking a loan with investors who are willing to invest money and get return.
- **What is this analysis being done for?**
  - Here, we are analysing the data from Lending Club's perspective. There are 2 types of risk associated with company's decision of whether or not to approve an applicant's loan request:
    - Approving the loan of a person who defaults (thus making a financial loss)
    - Not approving the loan of a person who will fully pay the loan (thus making a business loss)
  - We are analysing the given data to identify the driving factors that lead to defaults. Identifying these factors can save the company from future losses, at least to some extent.

Analysis Approach

- The analysis on the data is carried out in multiple steps including the following:
  1. Data reading and understanding
  2. Data cleaning
  3. Univariate analysis
     1. Categorical
     2. Quantitative
  4. Segmented univariate analysis – against loan status
     1. Categorical
     2. Quantitative
  5. Bivariate analysis

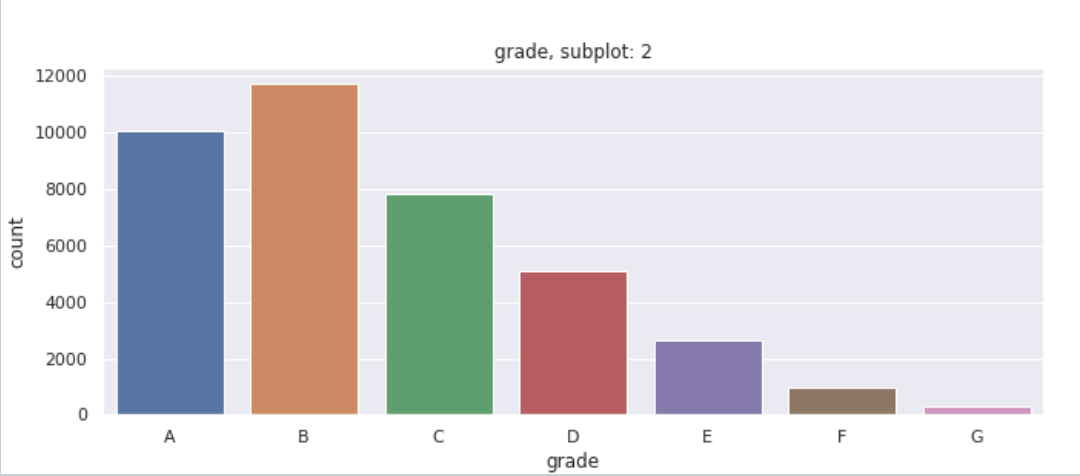
Results of Univariate Analysis - Categorical

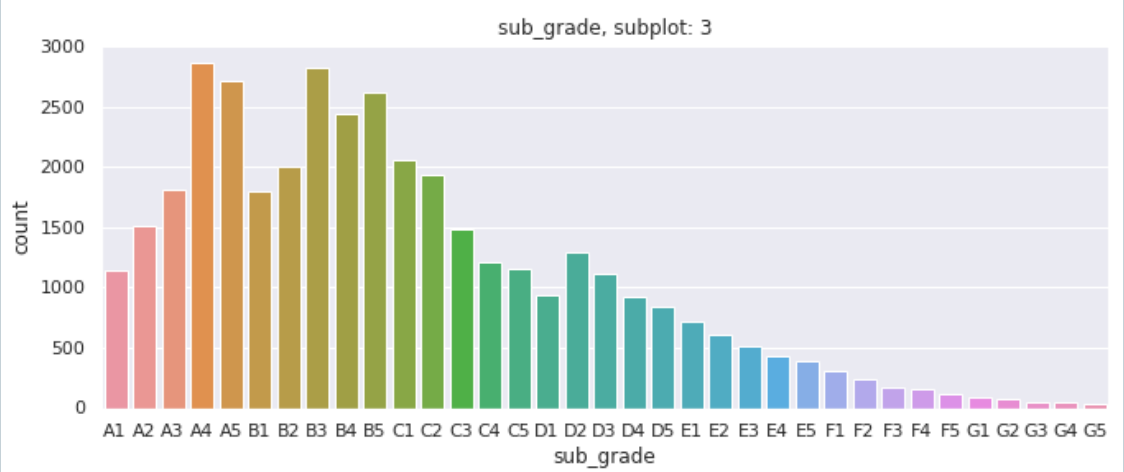
- We analysed frequencies of all the categorical columns.
- There are not much important insights we got from this analysis, except for there being some major variations of frequencies among different categories in some variables.
- The following graphs shows variations in frequencies of different categories of these variables:



- The graph shows that most of the people take a loan of 36 months term.

Results of Univariate Analysis - Categorical

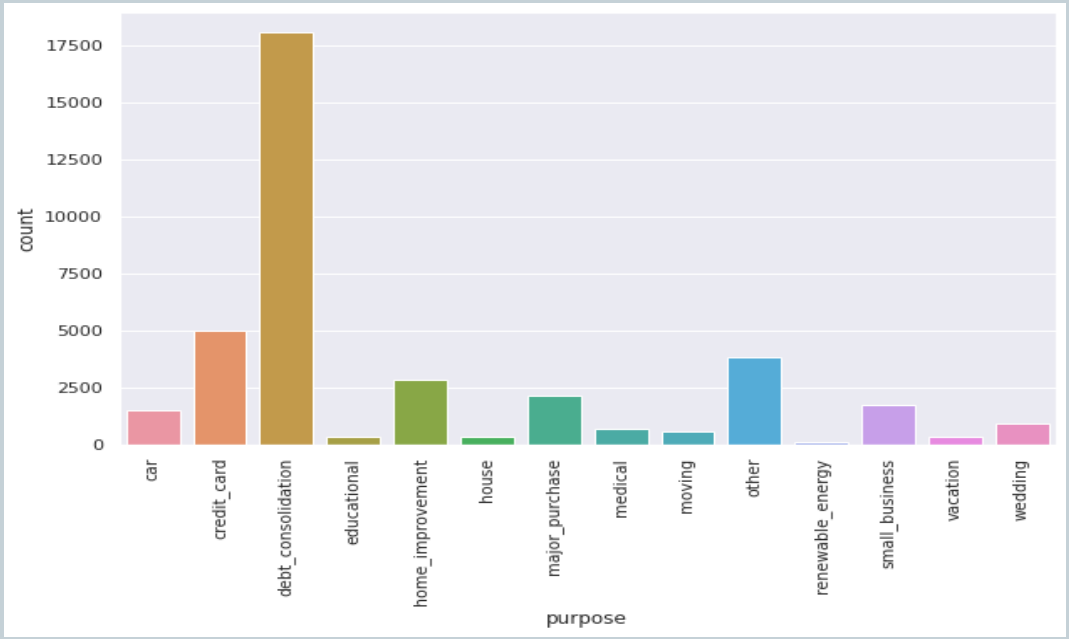




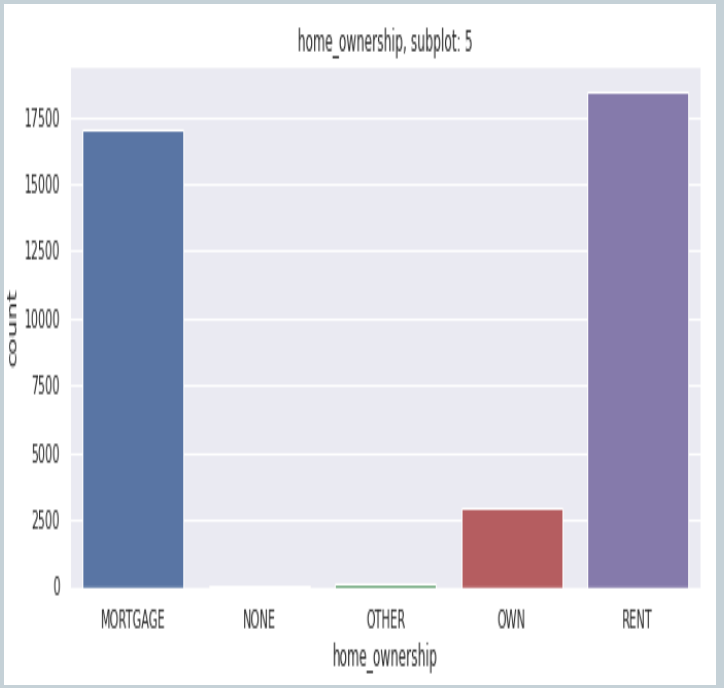
- The graph indicates that most of the people take a loan of B grade category.
- According to data, higher the grade, lower the interest. A grade loans has the lowest interest rate, and G has the highest.

- Although most of the people are taking a B grade loan, if we look in more detail in subgrade, most of the people are taking A4 grade loan. It is slightly more than B3.

Results of Univariate Analysis - Categorical



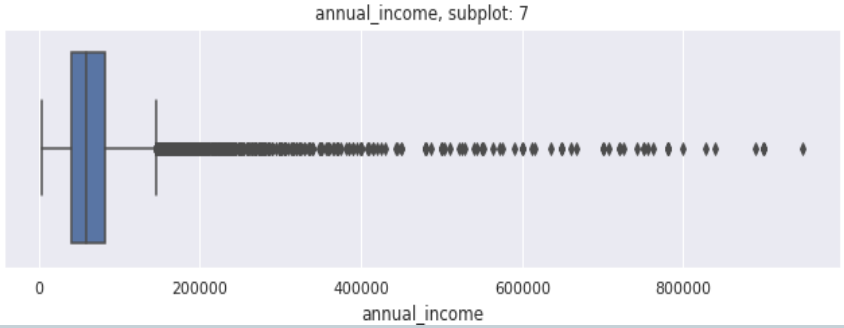
- **Most of the people are taking loan for the purpose of debt consolidation.**

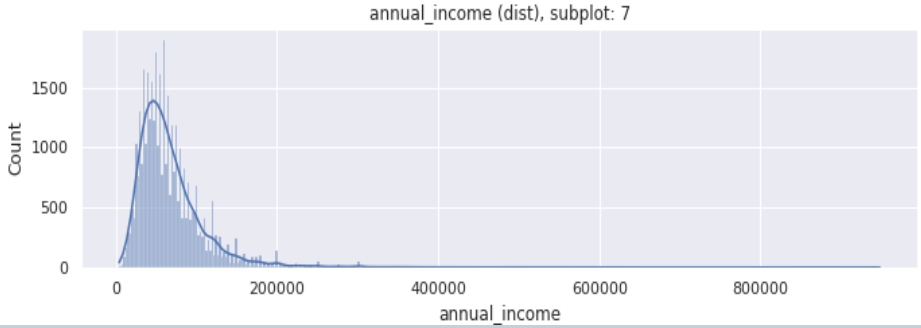


- **The graph shows that most of the people taking a loan are having a rental home, or mortgage. Very few people have their own home.**

Results of Univariate Analysis - Quantitative

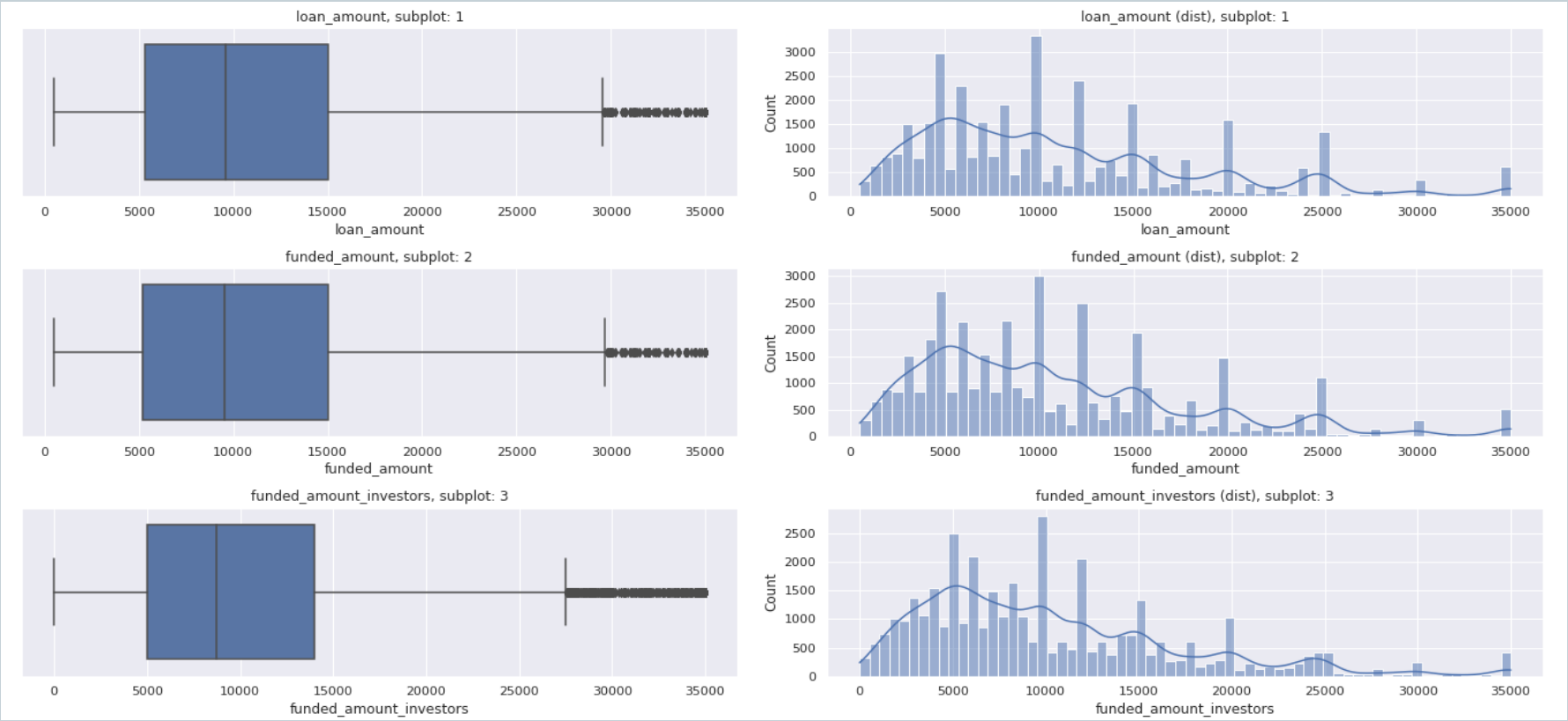
- We analysed the distribution of data for different quantitative variables.
- We found some outliers in the data, but since they were useful for the analysis, we have not removed them all.
- Following graphs shows distribution of data for various quantitative variables:





- The graph clearly indicates that most of the people have annual income between 0 to 100000. Then the distribution just decreases.
- Also, there are some outliers, but the outliers area is also pretty dense. So we have not removed those. And the outliers on the extreme right are important for analysis.

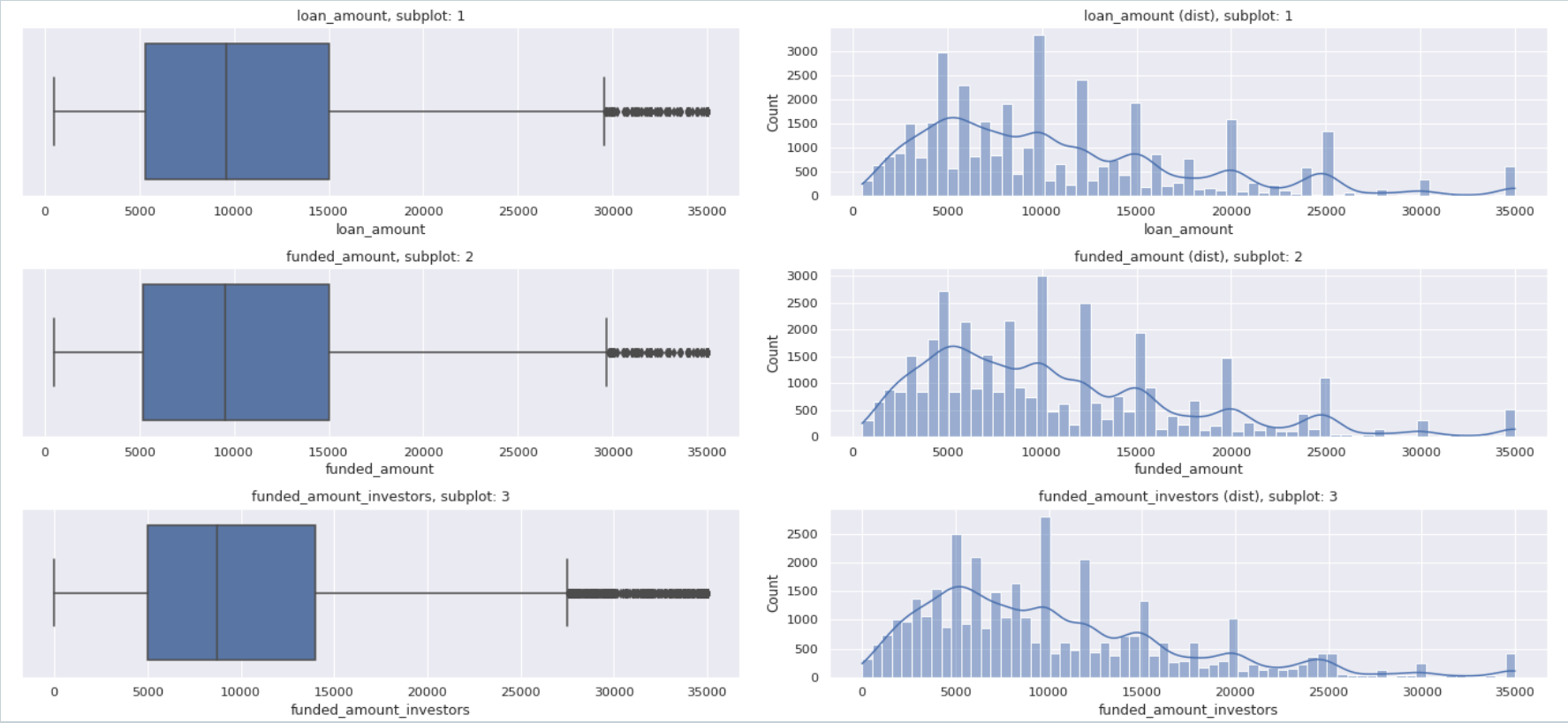
Results of Univariate Analysis - Quantitative



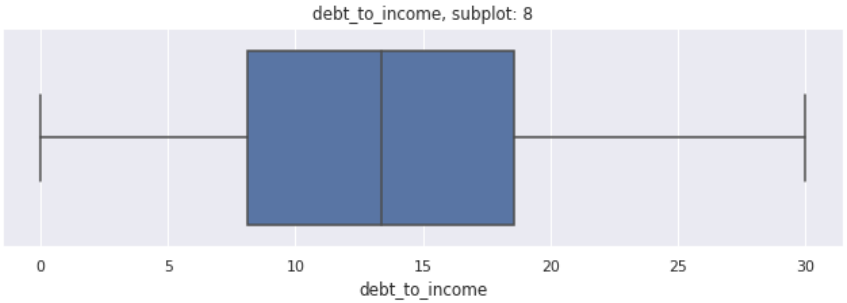
- All these graphs are pretty much the same.
- Amounts is mostly spread on the lower-to-middle side of the range.

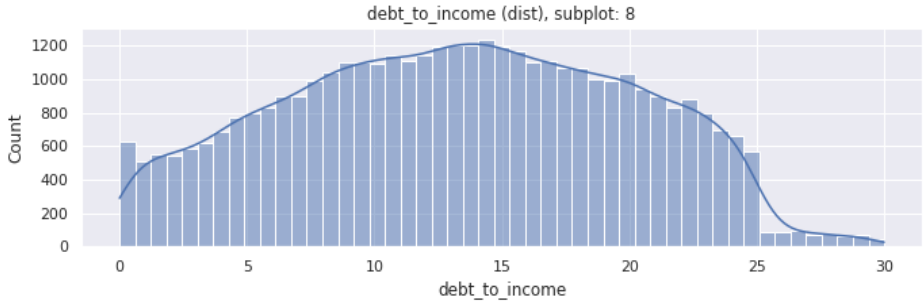
Results of Univariate Analysis - Quantitative

- We can also see sudden spikes at round numbers like 5000, 10000, 20000 etc. This maybe because person mostly applies for loan in the round amount. And investors also invest the same way.

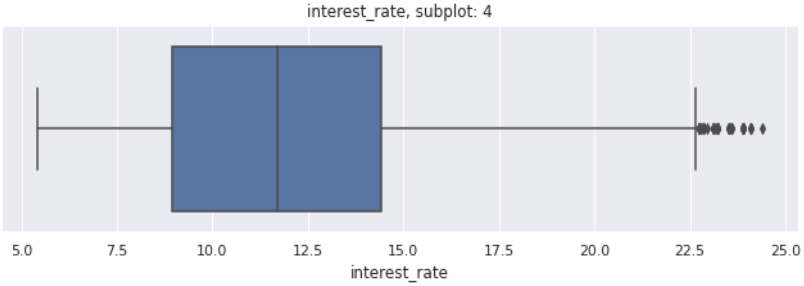


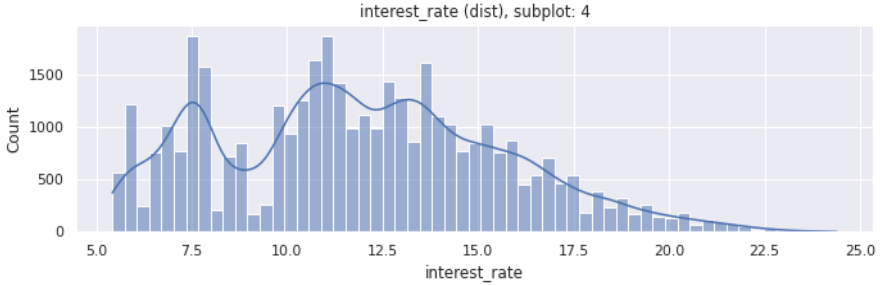
Results of Univariate Analysis - Quantitative





- The debt to income is almost equally distributed and there are no outliers as well.
- But we can see a sudden fall at 25 and then there are very few people having more than 25 debt to income.

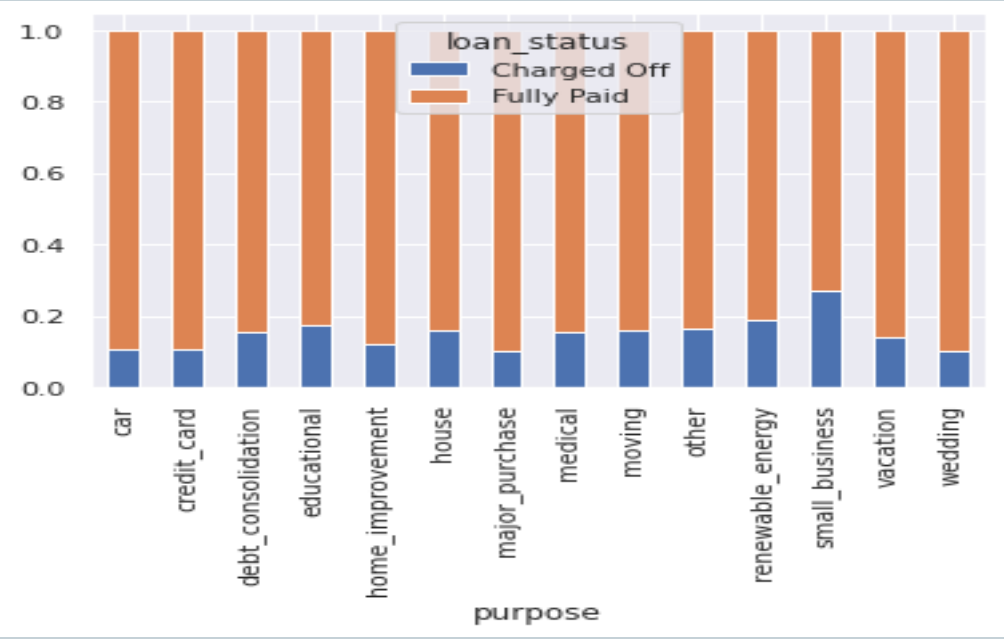




- Interest rate is also has very few outliers. The distribution shows that there is a sudden fall between 7.5 and 10. Less people are taking loan at this rate.

Results of Segmented Univariate Analysis - Categorical

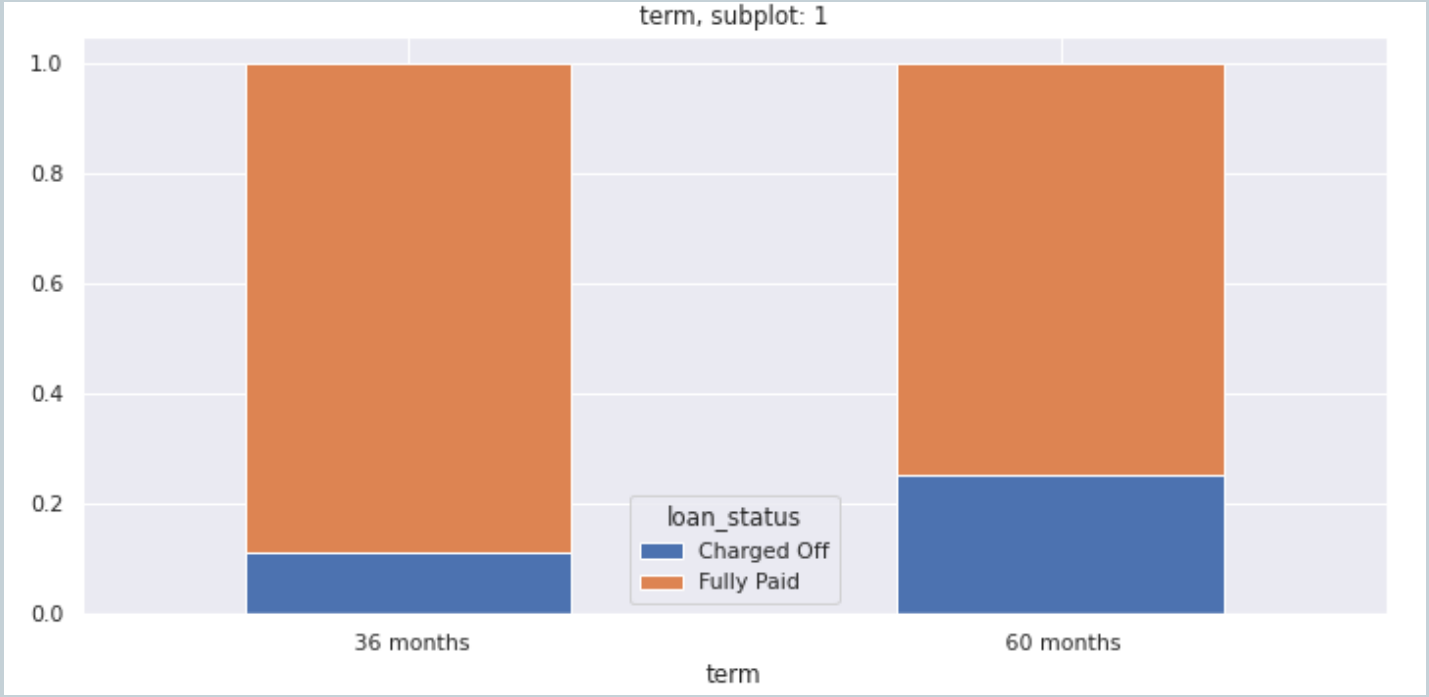
- Since the aim for this analysis is to find the driving factors for loan default, we took the loan status as target variable, and analysed other variables to check which variables are affecting the loan status.
- Following graphs show how other categorical variables are affecting loan status:



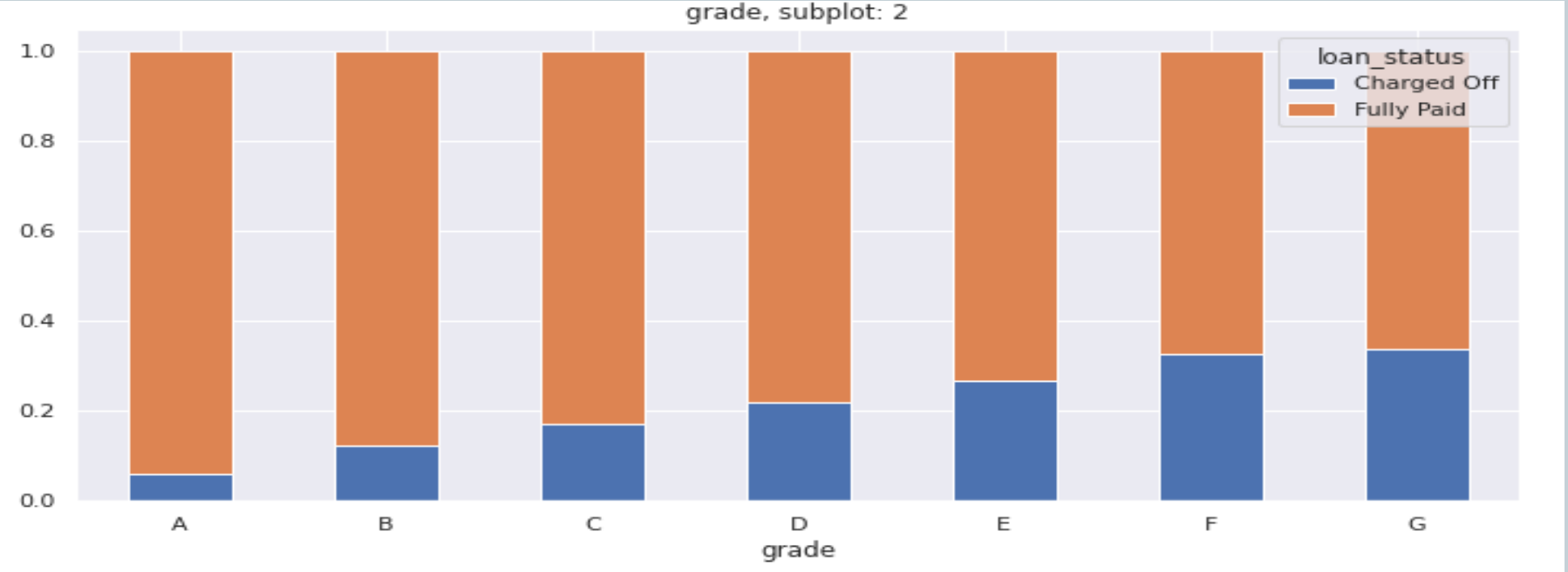
- According to graph, more number of people who take loan for small business, are defaulters.
- Also, as per earlier analysis, not many people take loan for small business, and among them also, there are high number of defaulters.

Results of Segmented Univariate Analysis - Categorical

- As seen in the graph, more number of people who take 60 months of loan are defaulters. And as earlier observed also, there are in fact less number of people who take 60 months loan. And among them also, more are defaulters.

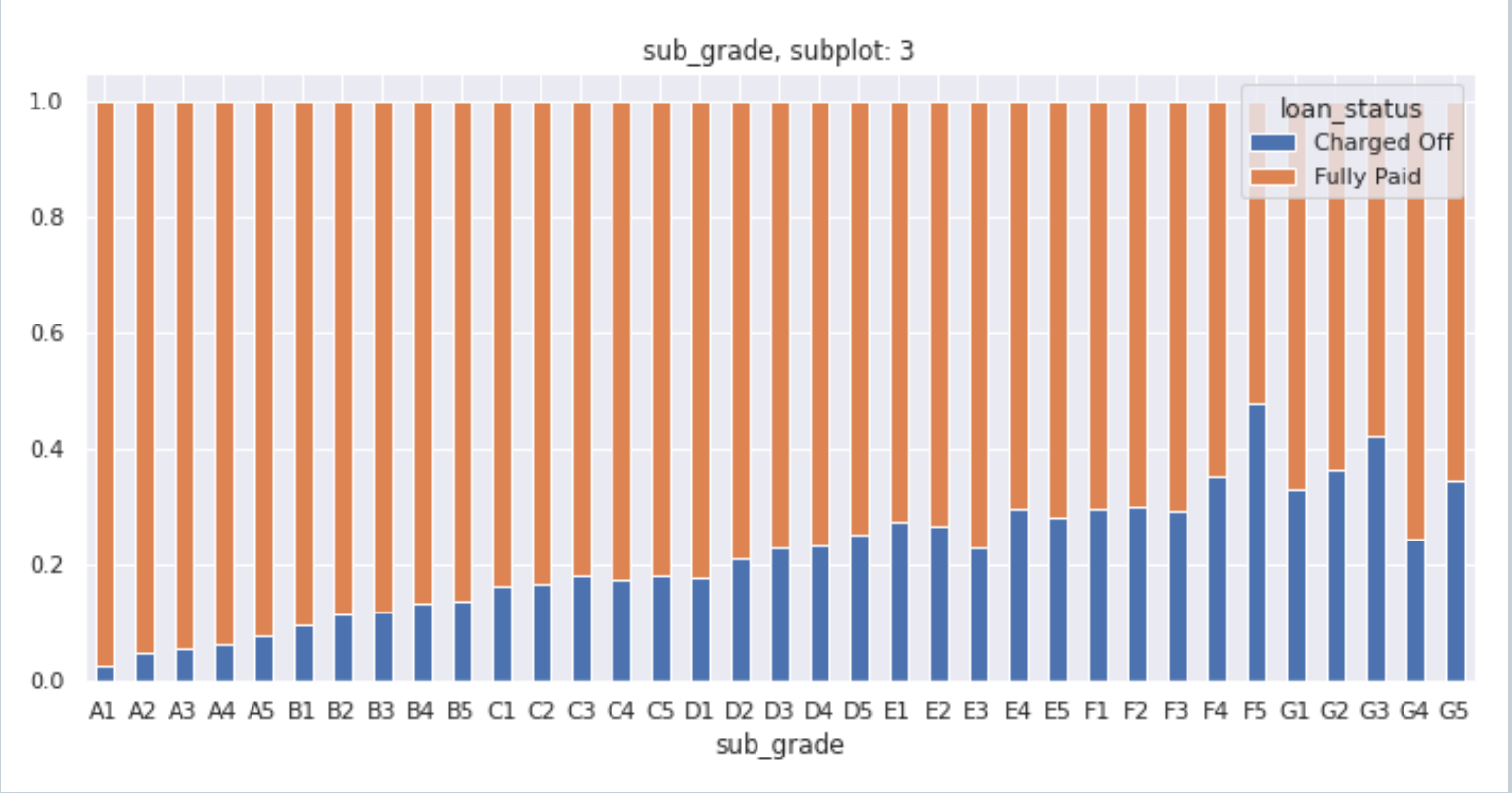


Results of Segmented Univariate Analysis - Categorical



- As per graph, most number of defaulters are the people taking F and G grade loan.
- Already there are very few people taking F and G grade loan, among them also, more than 30% are defaulters.

Results of Segmented Univariate Analysis - Categorical

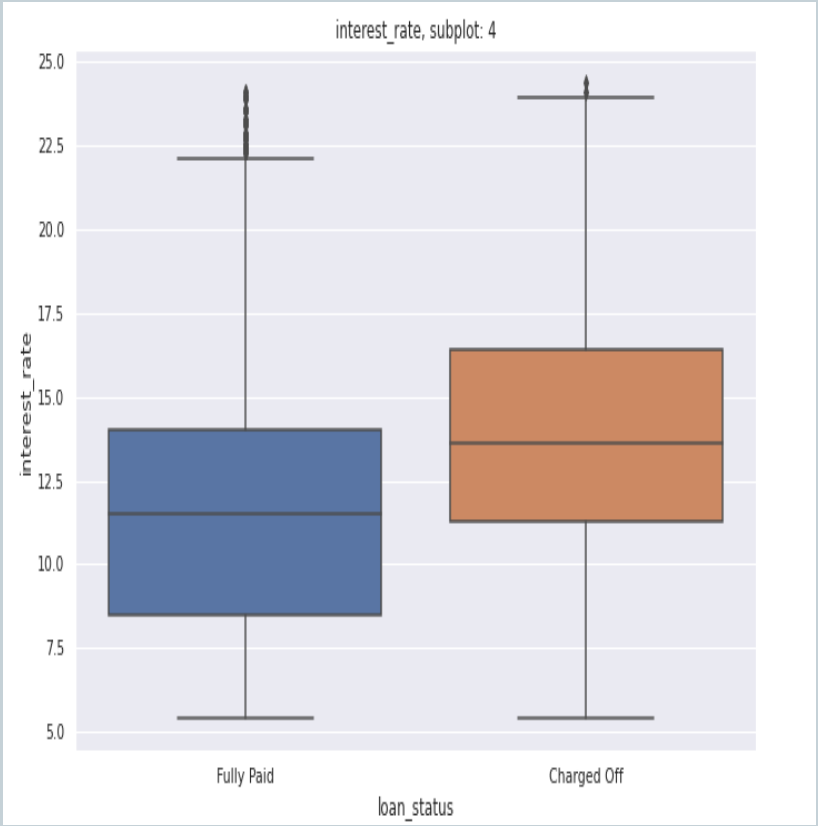


- **Most of the people taking loans of subgrade F4, F5, G2 and G3 have defaulted.**

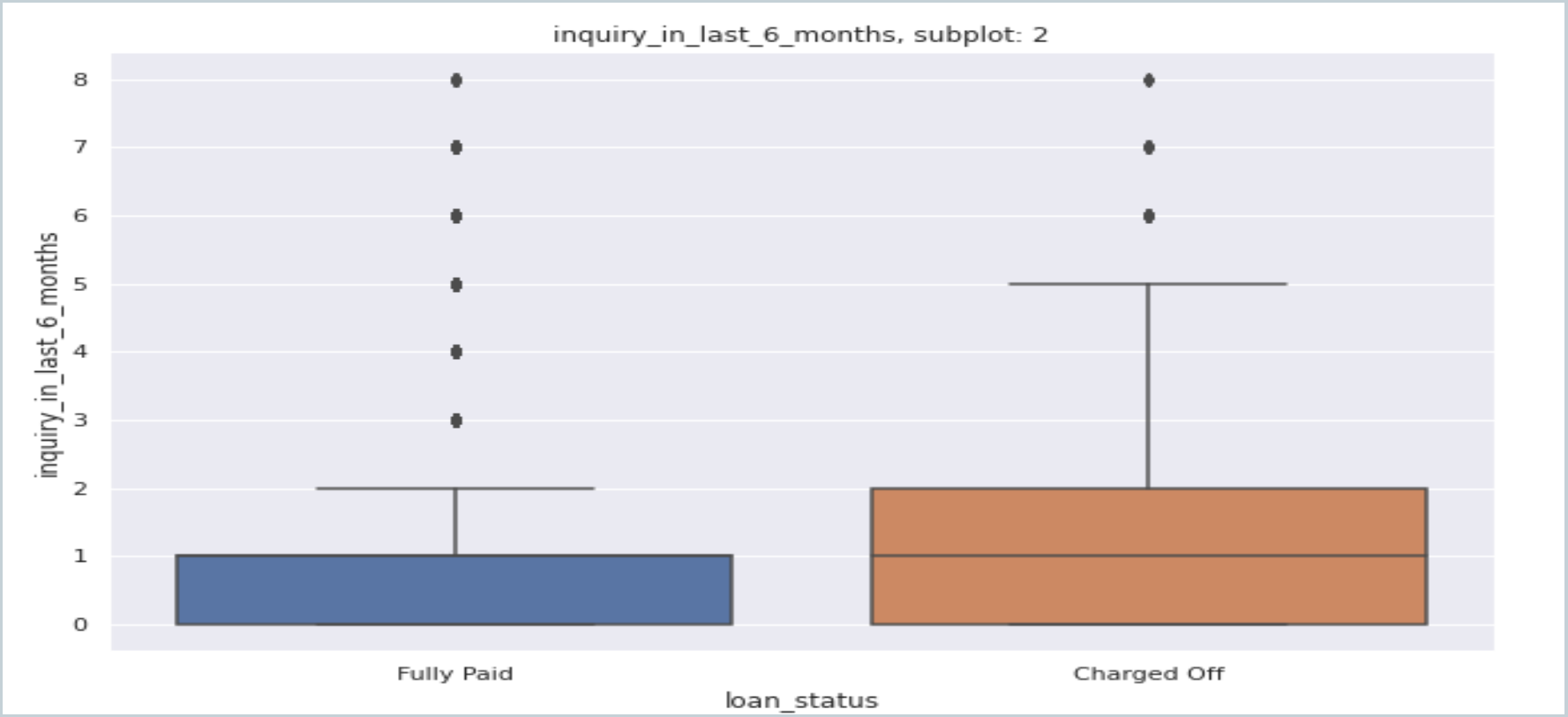
Results of Segmented Univariate Analysis - Quantitative

- **Following graphs show how other quantitative variables are affecting loan status:**

  - **The graph indicates that the people who defaulted took loan with higher interest rate than those who fully paid the loan.**

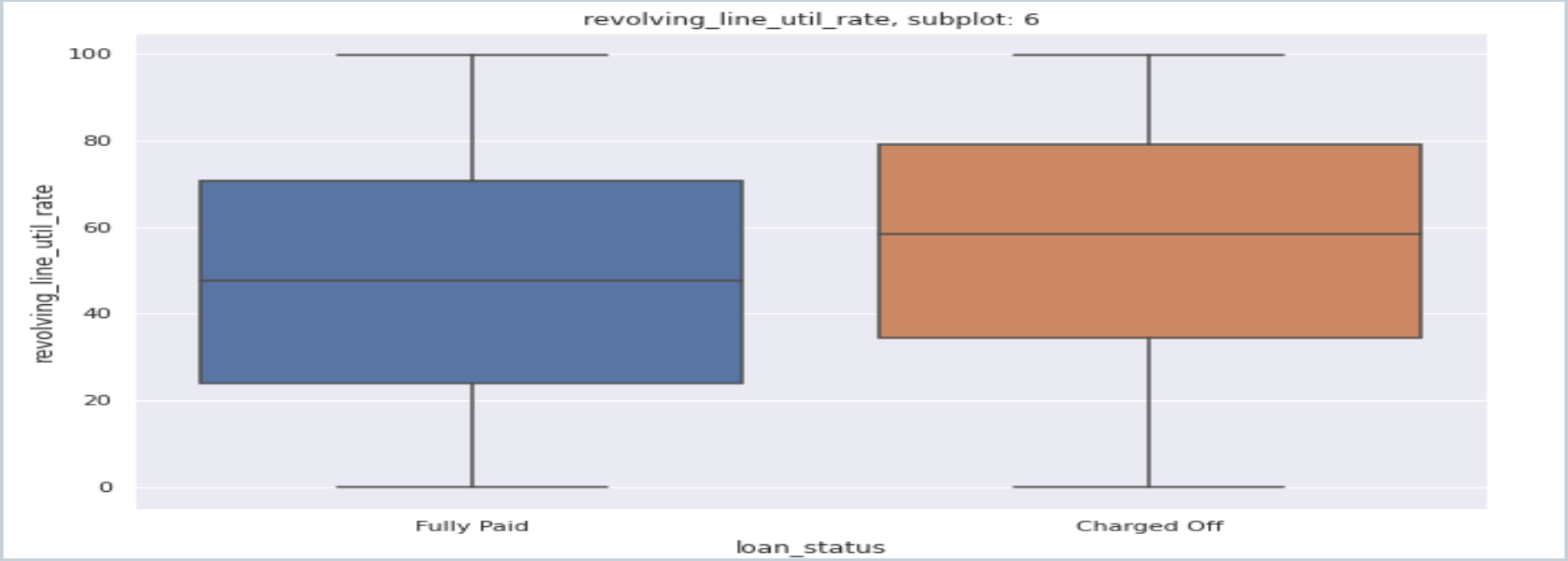


Results of Segmented Univariate Analysis - Quantitative



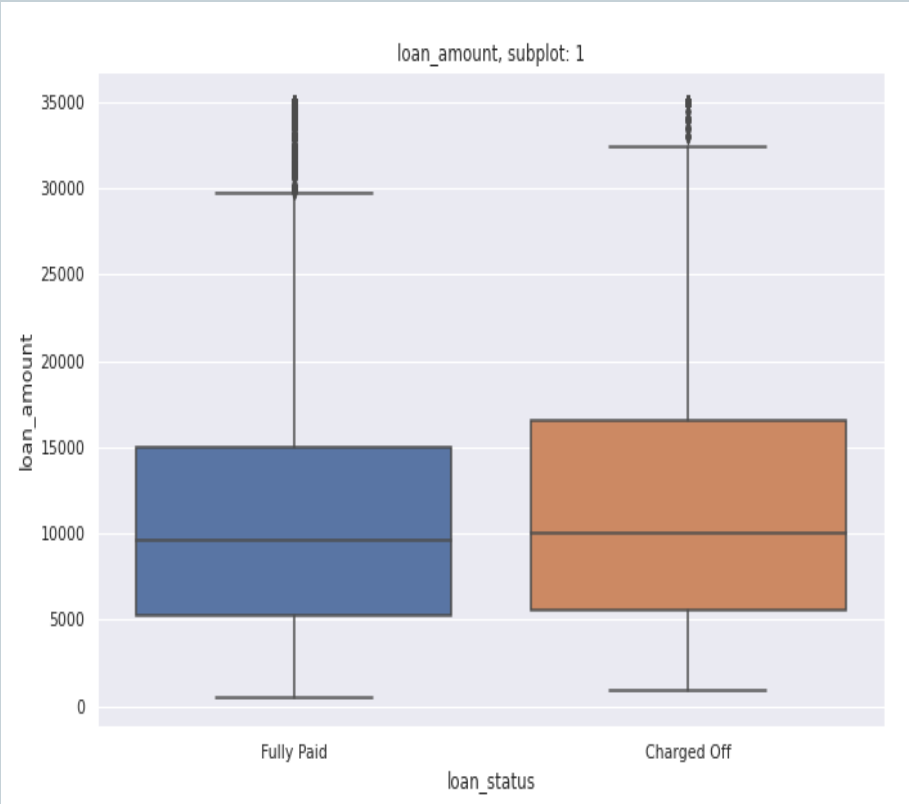
- **The graph clearly shows that the people who inquired more number of times say 3,4,5 are mostly the ones who defaulted.**

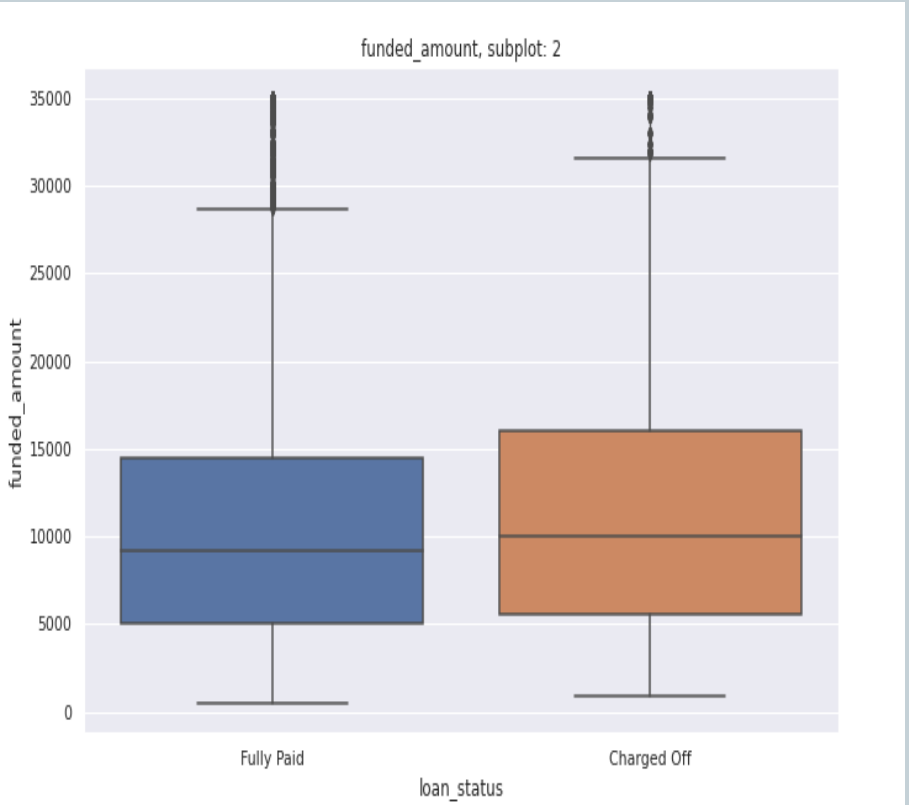
Results of Segmented Univariate Analysis - Quantitative



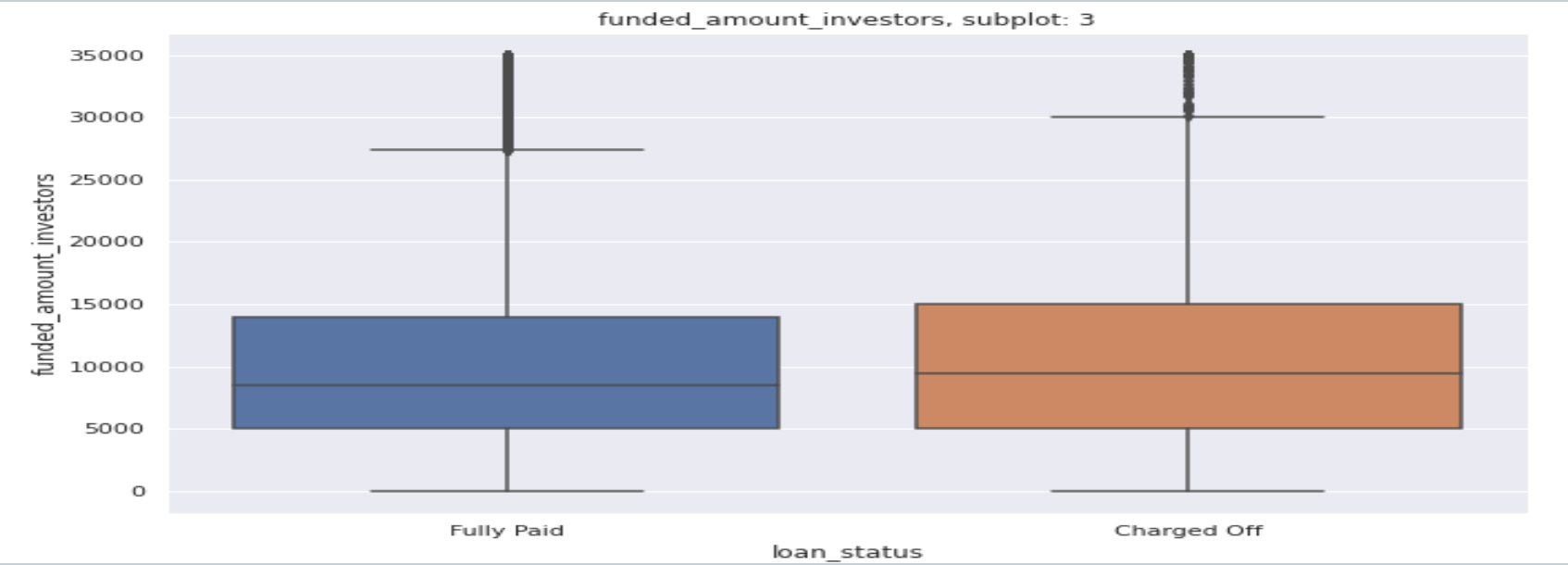
- As per the graph, mostly people having high revolving line utilization rate are more likely to default.

Results of Segmented Univariate Analysis - Quantitative





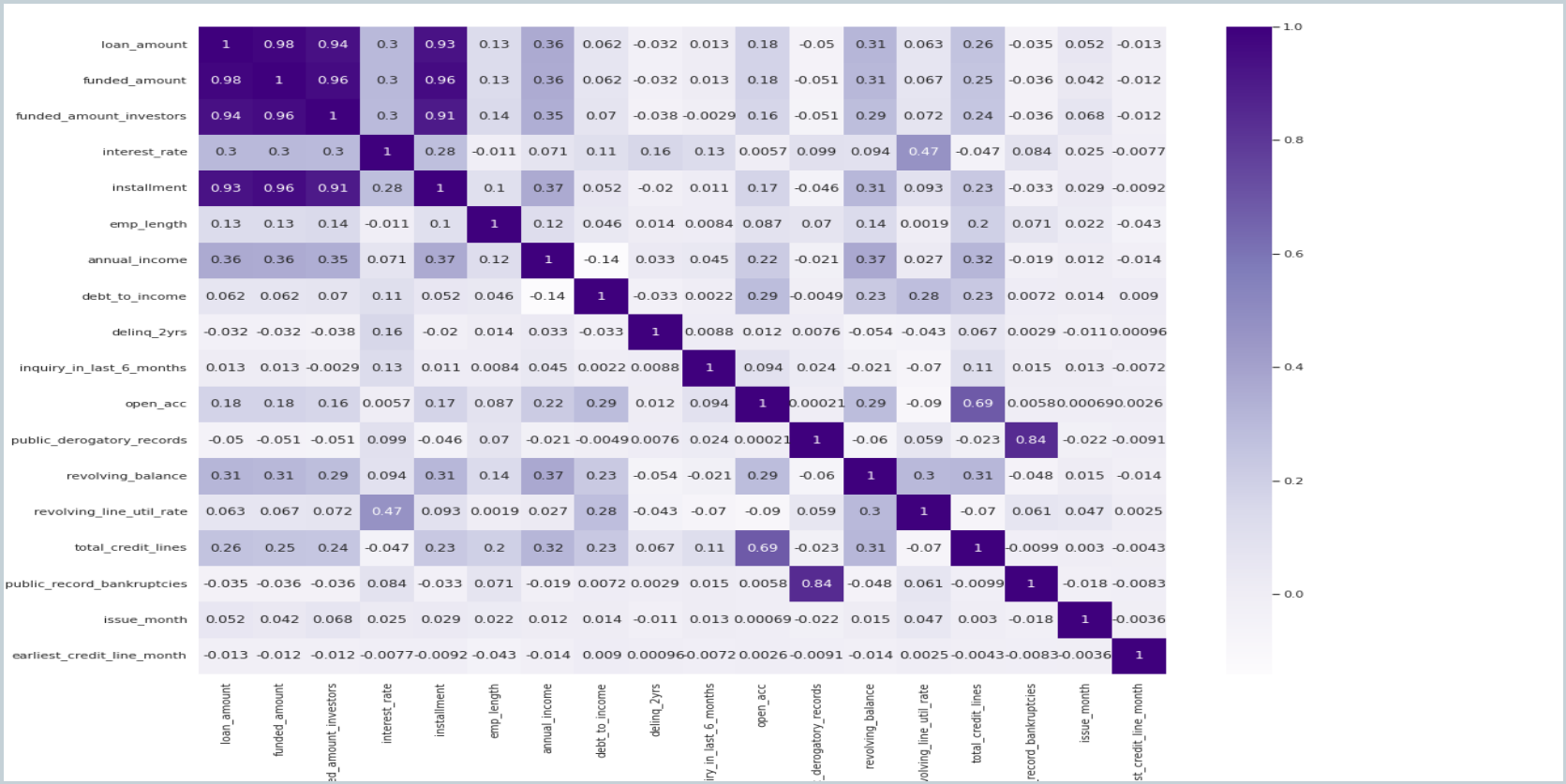
Results of Segmented Univariate Analysis - Quantitative



- The graphs shows that the defaulters asked for relatively higher amount of loan.
- The company also suggested higher amount of money to defaulters more.
- The defaulters also received higher amount of money from investors.

Results of Bivariate Analysis

- **Below is the correlation matrix that describes the correlation of different numeric variables with each other:**



Results of Bivariate Analysis

- As you can see, few variables like: loan_amount, funded_amount and funded_amount_investors are highly correlated.
- These 3 variables are also highly correlated with installment.
- Total_credit_lines and open_acc are highly correlated.
- Public_derogatory_records and public_record_bankruptcies are highly correlated.
- Debt to income and annual income are least correlated, in fact negatively correlated.

Final Results

- As per all the analysis, we can conclude that the driving factors for loan to be likely getting defaulted are: term, purpose, grade, sub grade, interest rate, inquiry in last 6 months and revolving line utilization rate.
- Recommendations:
  - The company should verify the persons details thoroughly if they are applying for 60 months term of loan.
  - The company should do their research before giving loan for small businesses.
  - They should also double check before giving loan of lower grade F and G, subgrade F4, F5, G2 & G3 and by default higher interest rate.
  - They should check the background of the people inquiring multiple times for loan in last 6 months.
  - There is also an insight found from the amount variables. They should also check the backgrounds of people asking for higher amount of loans. And they can also suggest their investors to do the same before investing. They can also suggest a lower amount to the person asking for a very high amount of loan.

Thank You!