LENDING CLUB CASE STUDY

ASSIGNMENT SUBMISSION

Name: Nabakishore Neog and Sidharth Kriplani

Case Description: Objectives

Lending Club is the largest online market place where borrowers can easily access lower interest rate loans through an online interface.

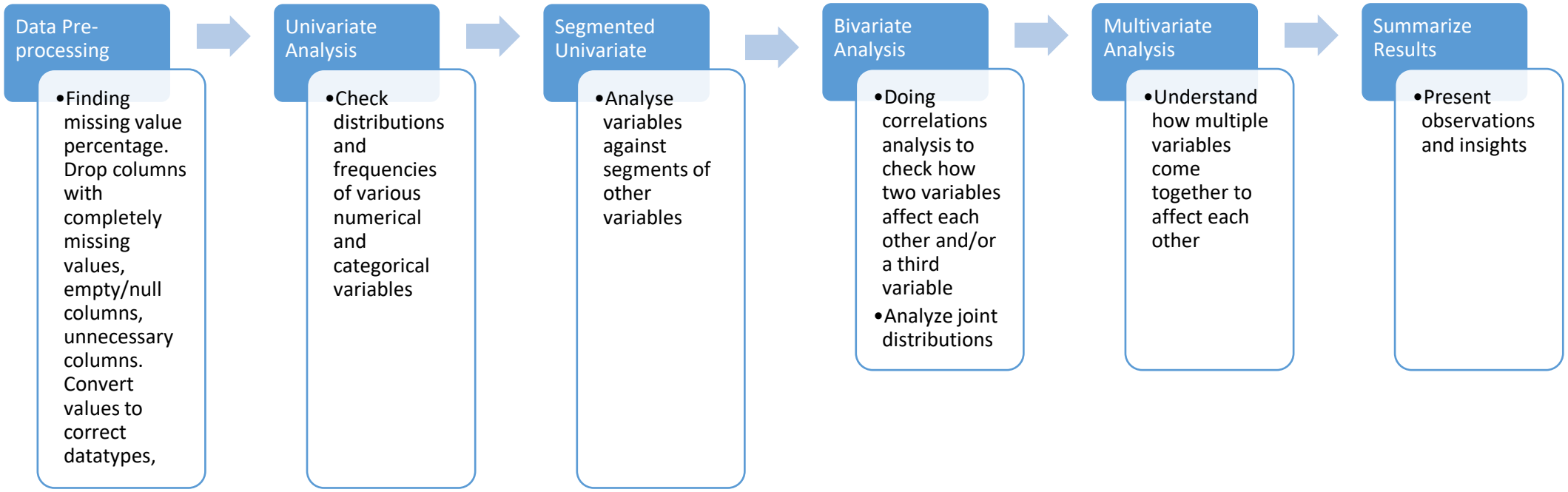
BUSINESS UNDERSTANDING:

- Lending Club wants to understand the driving factors behind the loan default, i.e the driver variables which are strong indicators of defaults
- Lending Club wants to understand the above data for risk assessment while processing loan applications

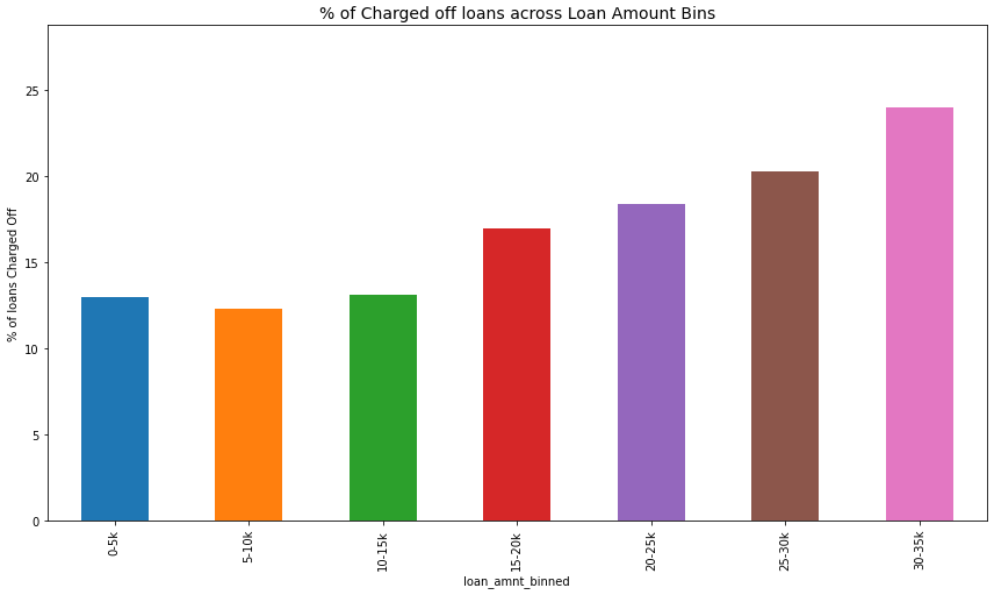
ROLE TO PLAY HERE AS A DATA SCIENTIST:

- Lending Club has a dataset containing information about the past loan applicants
- We have to use EDA to understand how consumer attributes and loan attributes influence the tendency of the default

Problem solving methodology



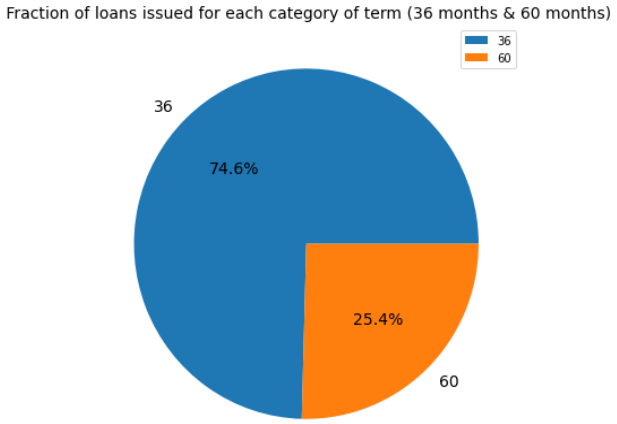
Analysis

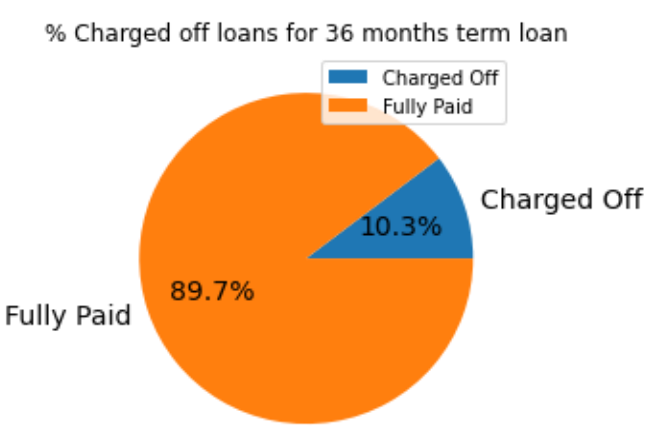


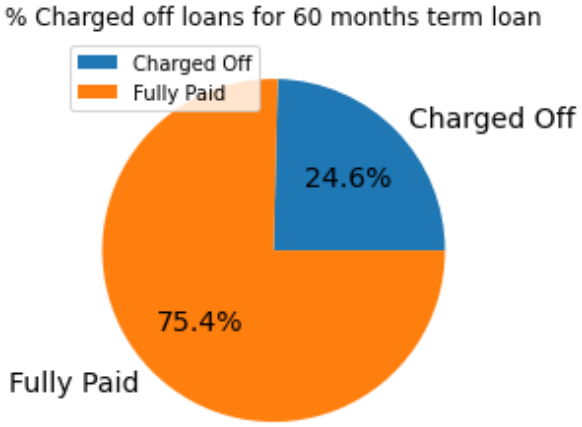
Here, we can see that as the loan amount increases the percentage of charged off loans also increases which means that loan amount is a strong indicator of likelihood of a default occurring.

Lending Club should keep the loan amounts low in this case to reduce the tendency of a default to occur.

Analysis







Most Loans are given out for a term of three years. These are 74.5% of the total loans given out.

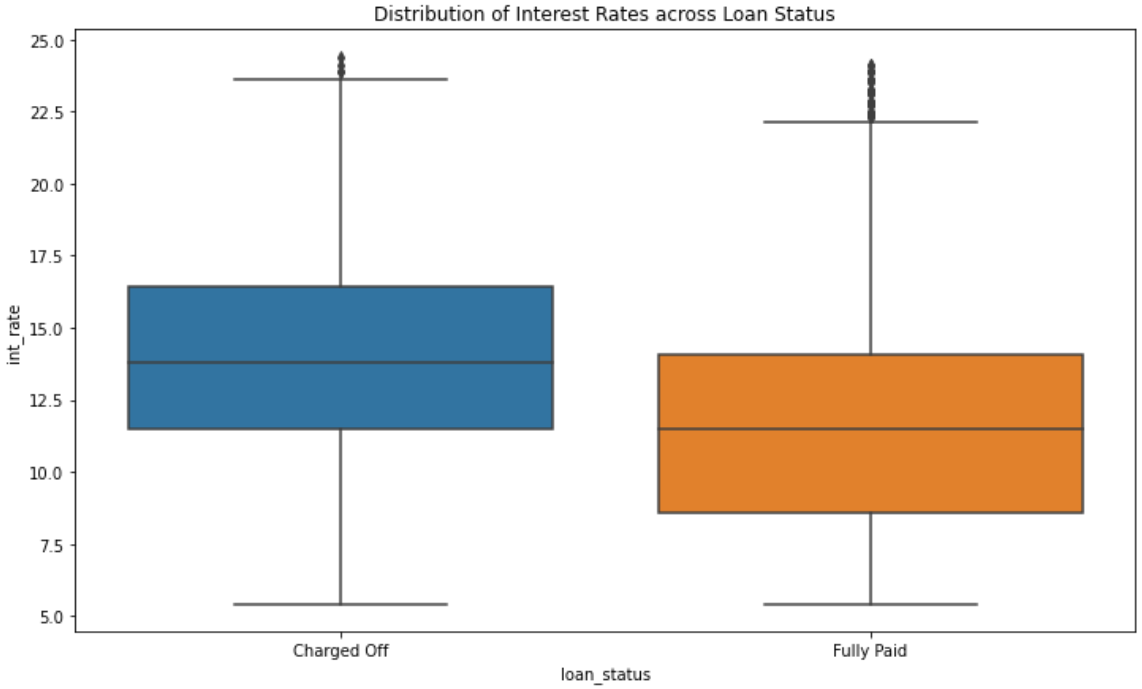
Of the loans given out for a term of three years, only 10.3% were charged off.

Of the loans given out for a term of 5 years, only 24.6% were charged off.

It implies that term has a linear relation with the tendency of default occurring on a loan.

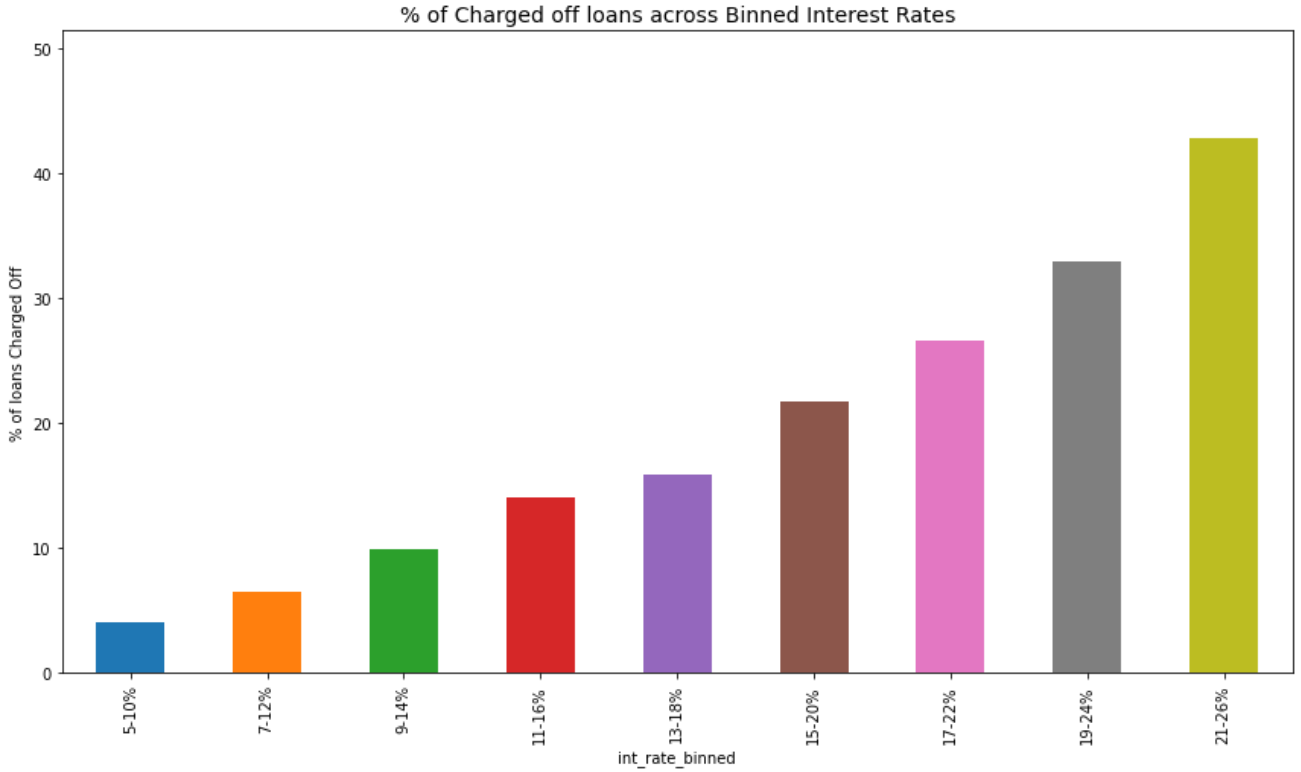
Thus, based on the provided data, loan terms should be for three years and no more to minimize the risk of default.

Analysis



Here we can see that the interest rate in the case of fully paid loans is less than that charged in the case of loans that were charged off which makes sense. So to further minimize the tendency of defaults occurring, interest rates can be only increased.

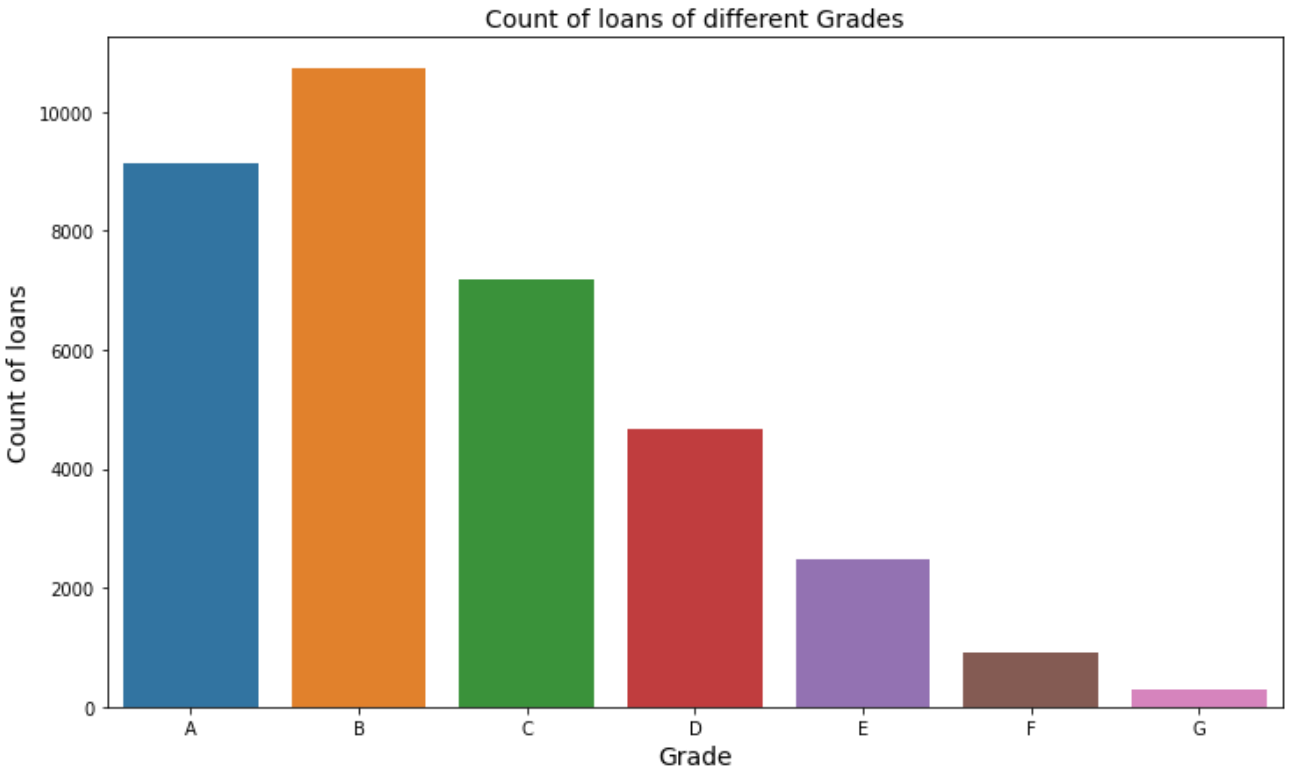
Analysis



Here we can see that higher the interest rate, more are the charged off loans.

This just supports the inference drawn by the previous slide and the deduction here is that the interest rate should be increased according to the common financial wisdom, to deter the loan applicants from taking the loan when it seems that there is a likelihood for those applicants to turn defaulters.

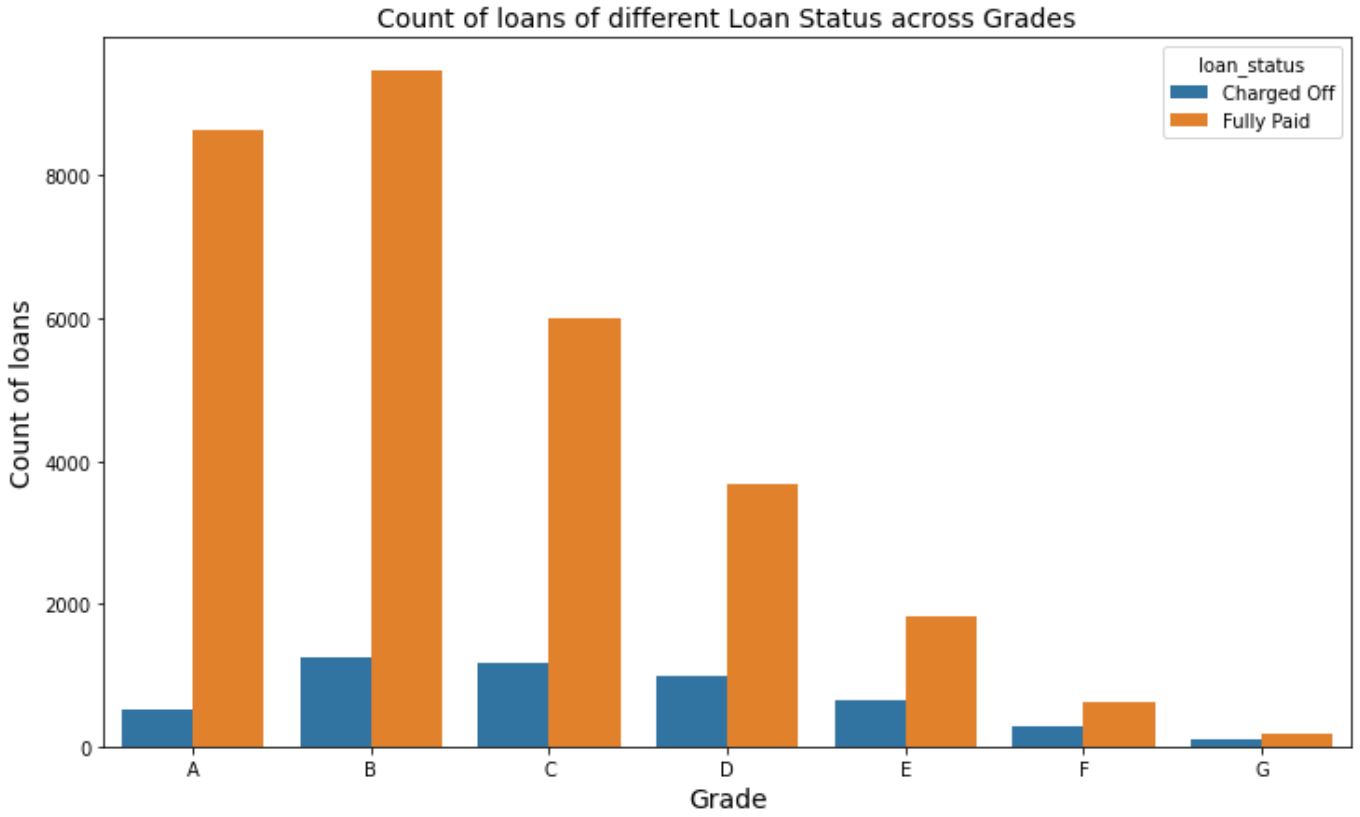
Analysis



Grade B has the most loans followed by Grade A.

Grade G has the lowest number of loans.

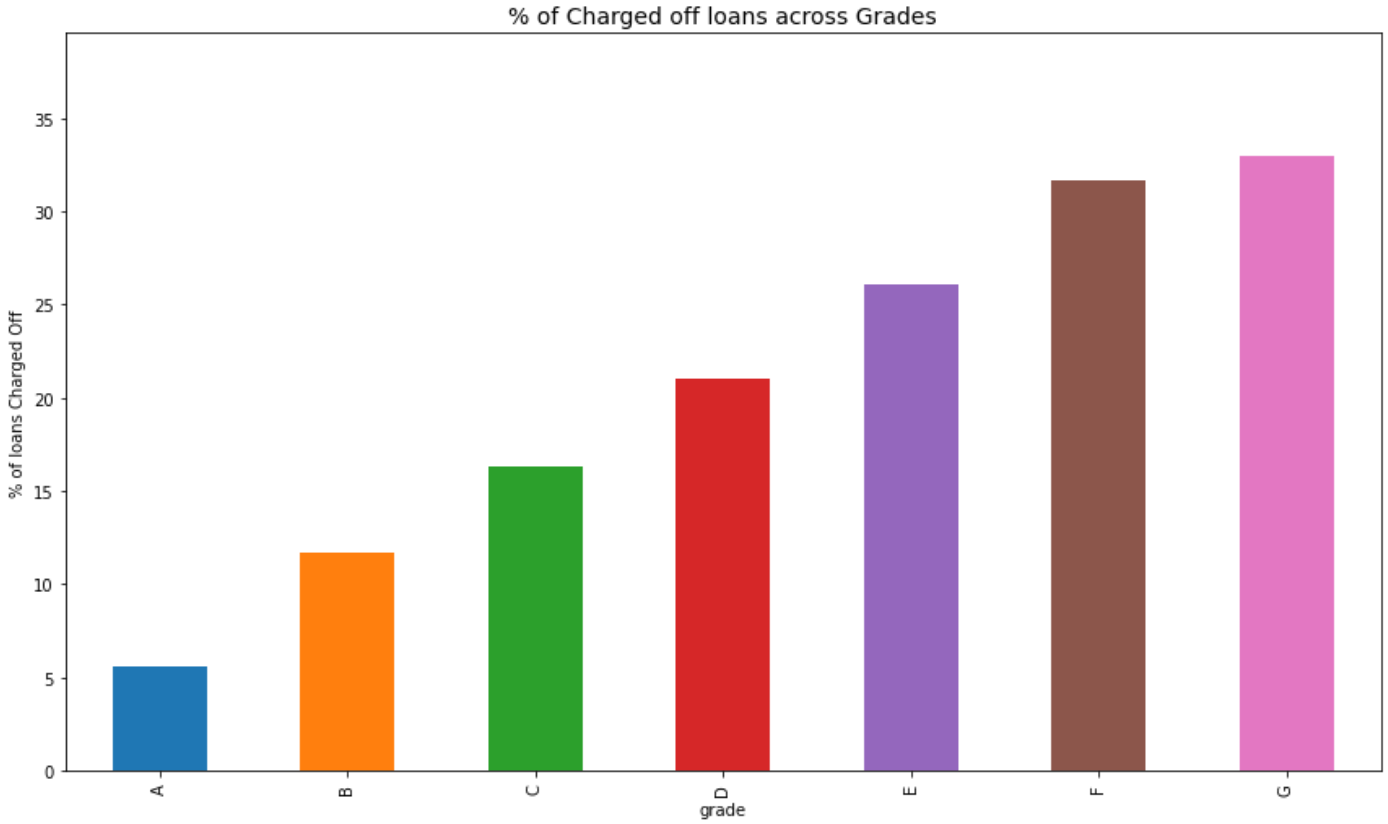
Analysis



Following the previous slide, segmented univariate analysis in the grade of loans against loan count suggests that the behavior of charged off and fully paid loans is the same for all loan grades.

But what we can conclude is that more Grade B and Grade A loans should be given to maximize the likelihood of their being no defaulters on any loans.

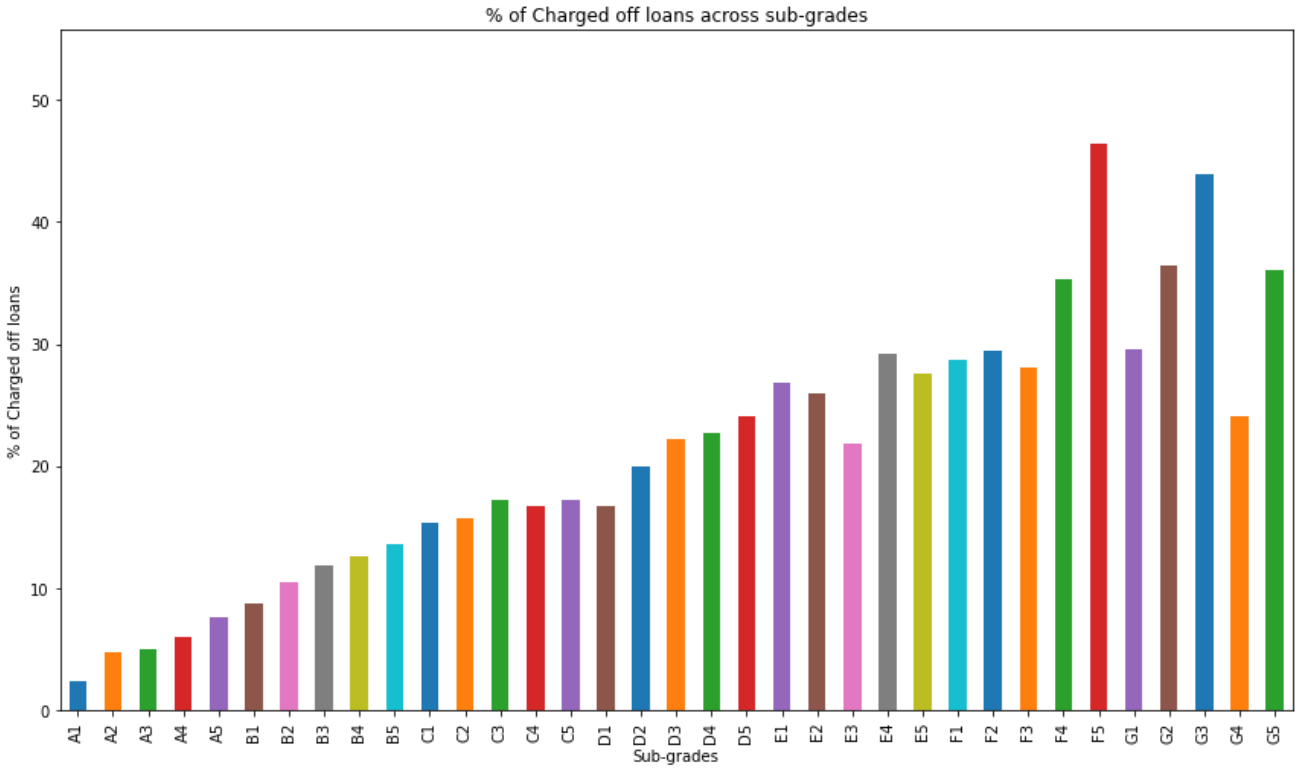
Analysis



Further analysis suggests that from Grade G to Grade F and alphabetically going down, the number of charged off loans keeps decreasing.

This is the reverse of the conclusion from the previous slide, and this simply suggests that Grade G and Grade F loans should be reduced to decrease the likelihood of there being defaulters.

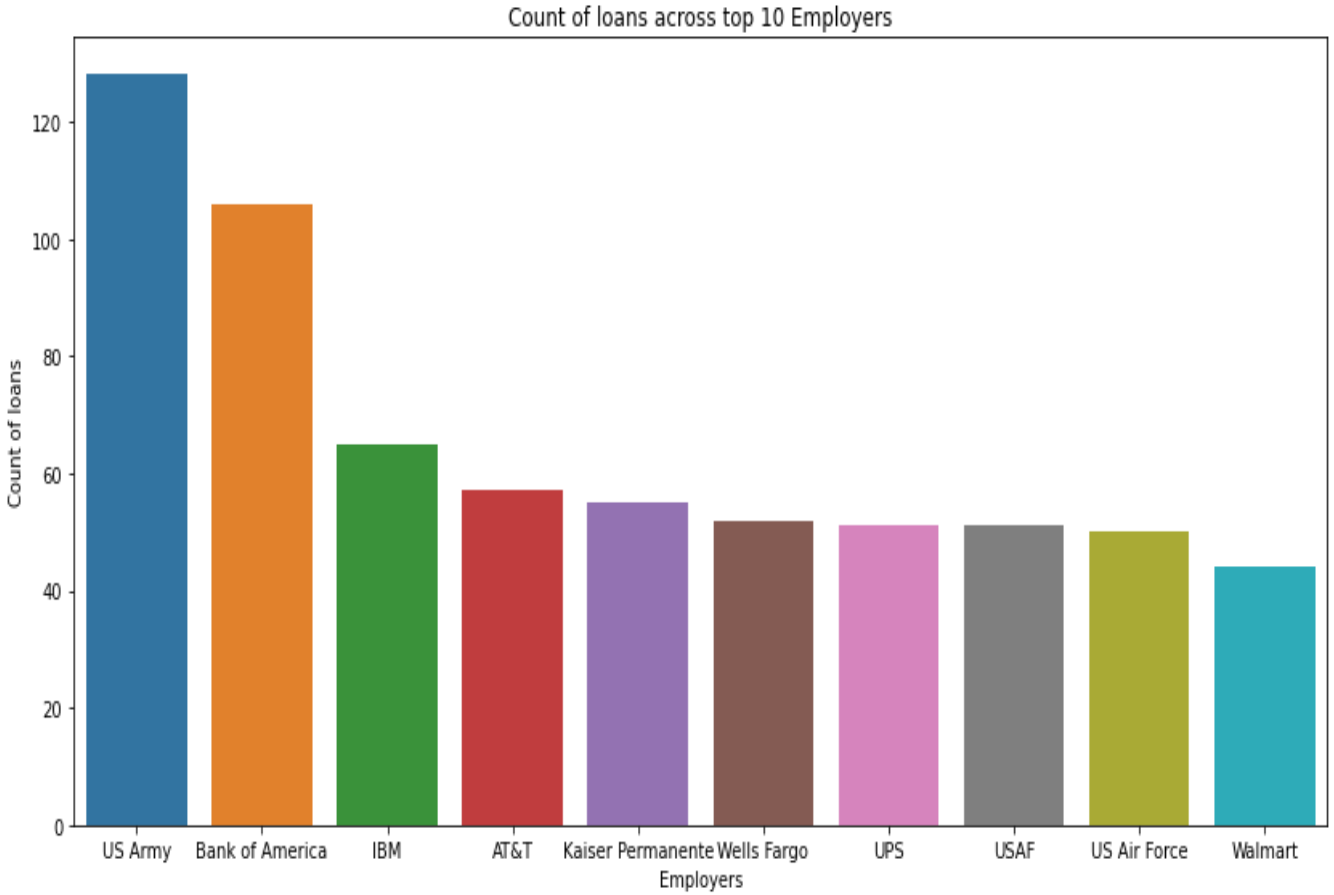
Analysis



Here we can see that subgrades F5 and G3 are where the highest number of charged off loans were given out. A1 subgrade being the lowest with the number of charged off loans, is thus, the safest loan subgrade as well.

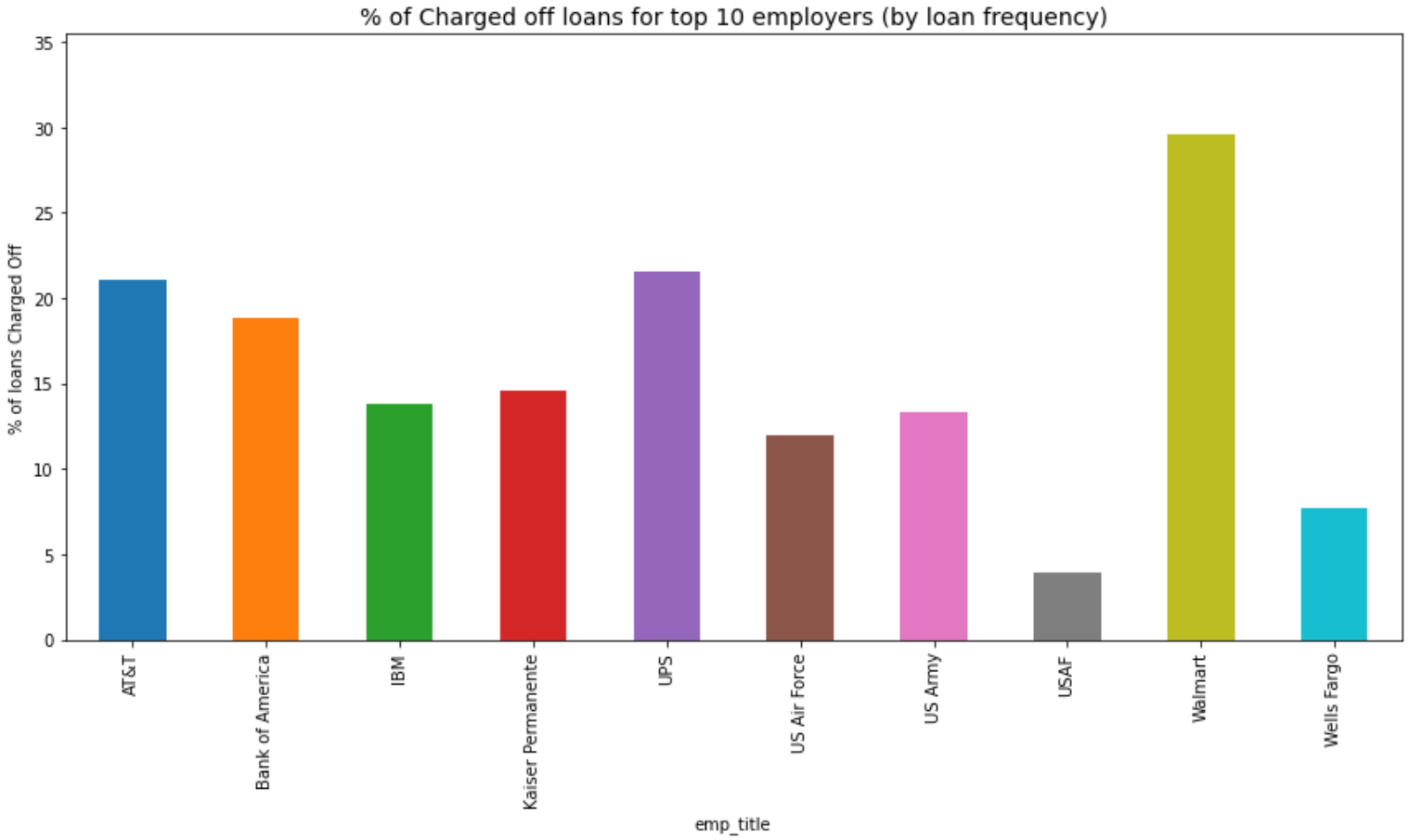
Loan subgrades show a linear variation as the alphabetical and numerical increment keeps happening.

Analysis



Most number of loans were given to the soldiers of the US army, followed by the employees of Bank of America and so on.

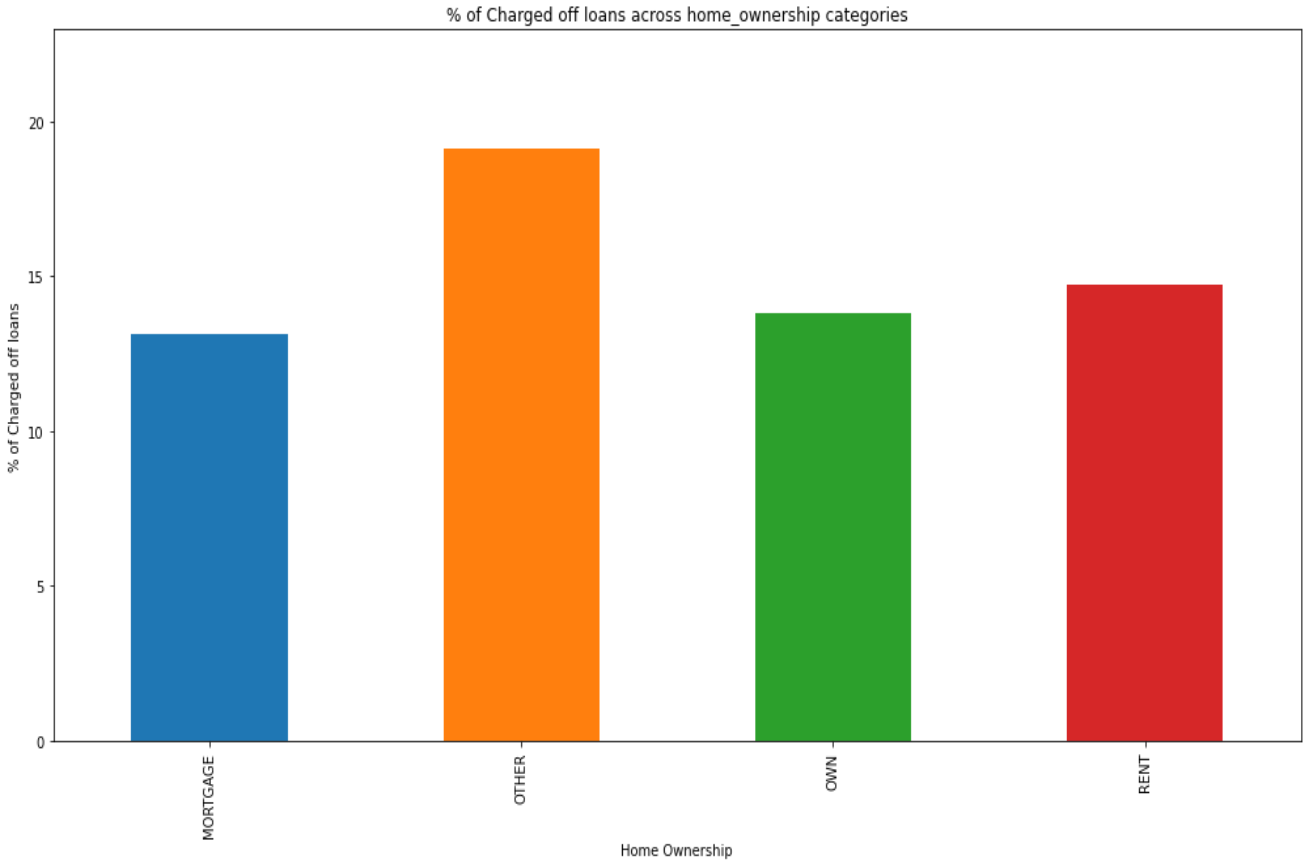
Analysis



The bar chart on the left side suggests that the most number of charged off loans were given out to employees of Walmart, followed by those of UPS and AT&T and then Bank of America.

Lending Club should be careful before proceeding with loan disbursal to the employees of the above mentioned companies.

Analysis

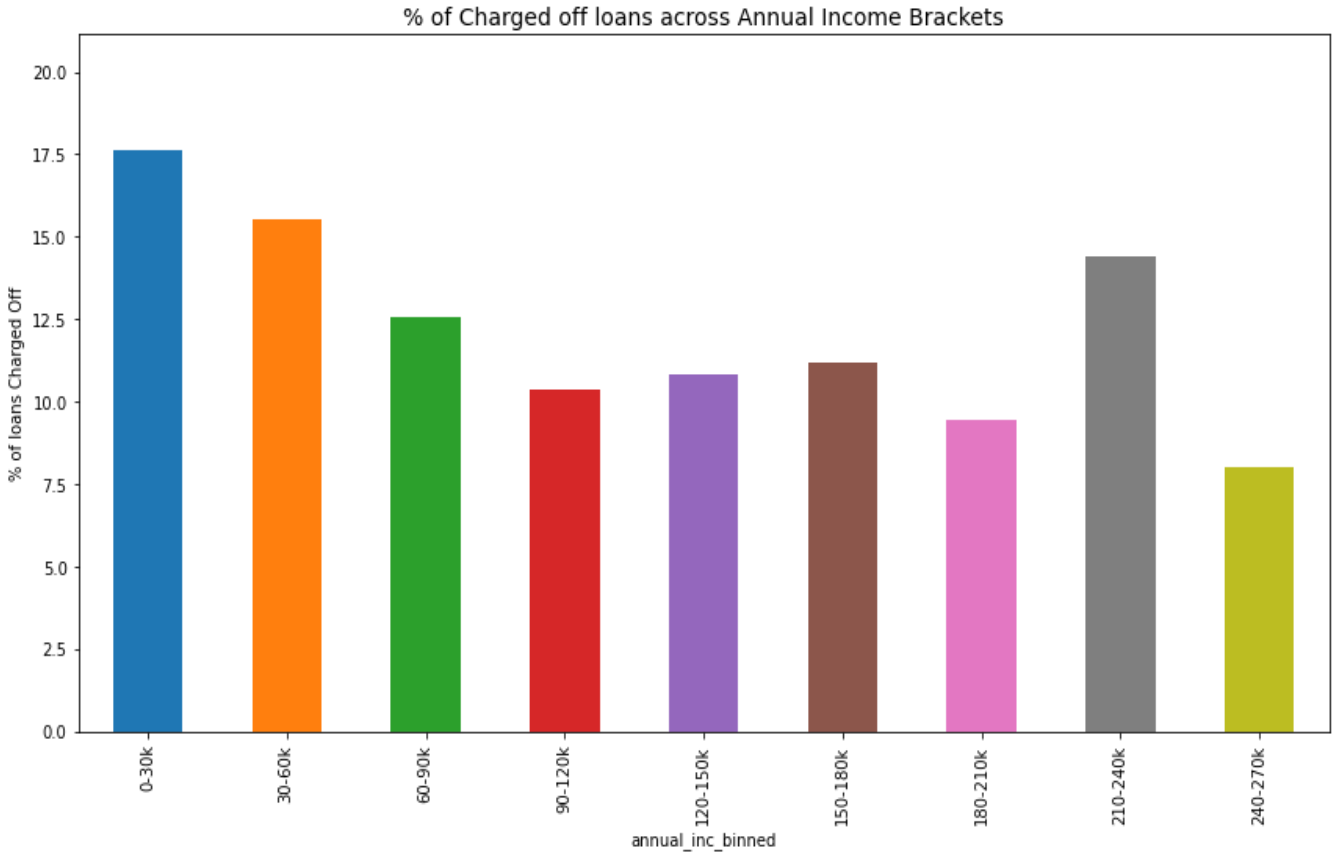


The most loans that were charged off belong to the Other category.

The other categories, Rent, Mortgage and Own rank behind Other in this case, and have almost equal number of Charged off loans.

What we can conclude from this is that it is better for Lending Club to give loans to an applicant belong to any category but ‘Other’.

Analysis

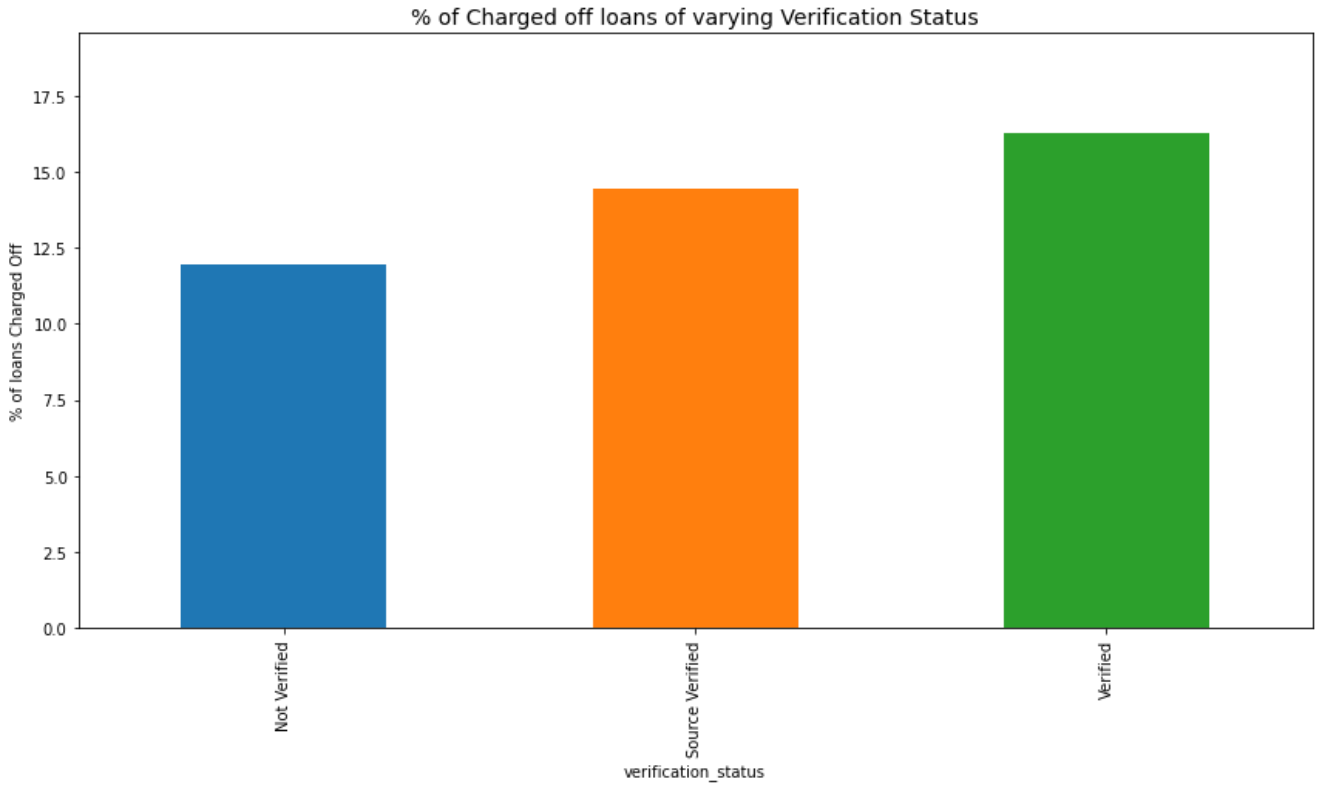


Annual Income of the loan applicant and the tendency to default is inversely related to each other as can be concluded from the adjoining plot.

Most charged off loans belong to the lowest income bin.

Thus, Lending Club should provide loans only to the higher income bins.

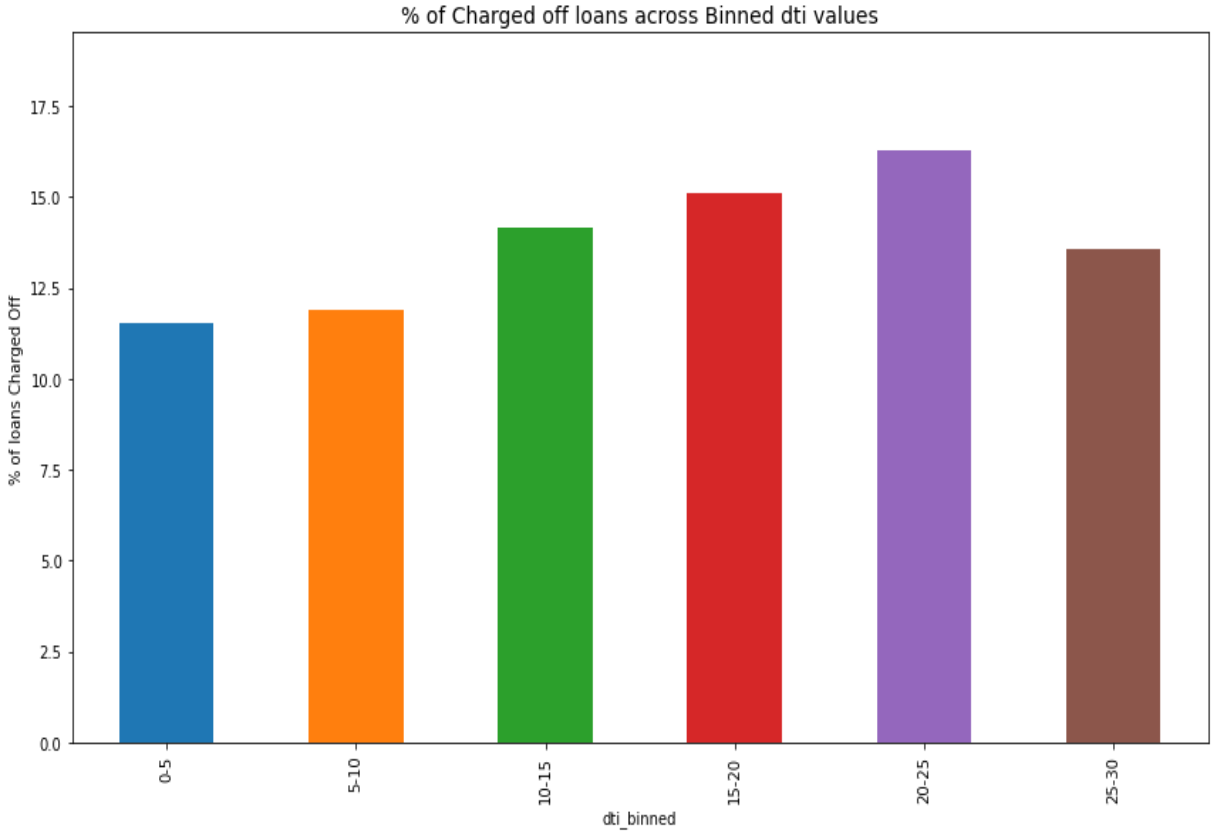
Analysis



This plot compares the variables verification status and number of charged off loans.

Since the most charged off loans happen even when the income source is verified, we cannot draw a precise conclusion here, except the fact that if Lending Club provides loans to only those applicants whose income source is ‘source verified,’ then the likelihood of their being defaults on that loan will be less.

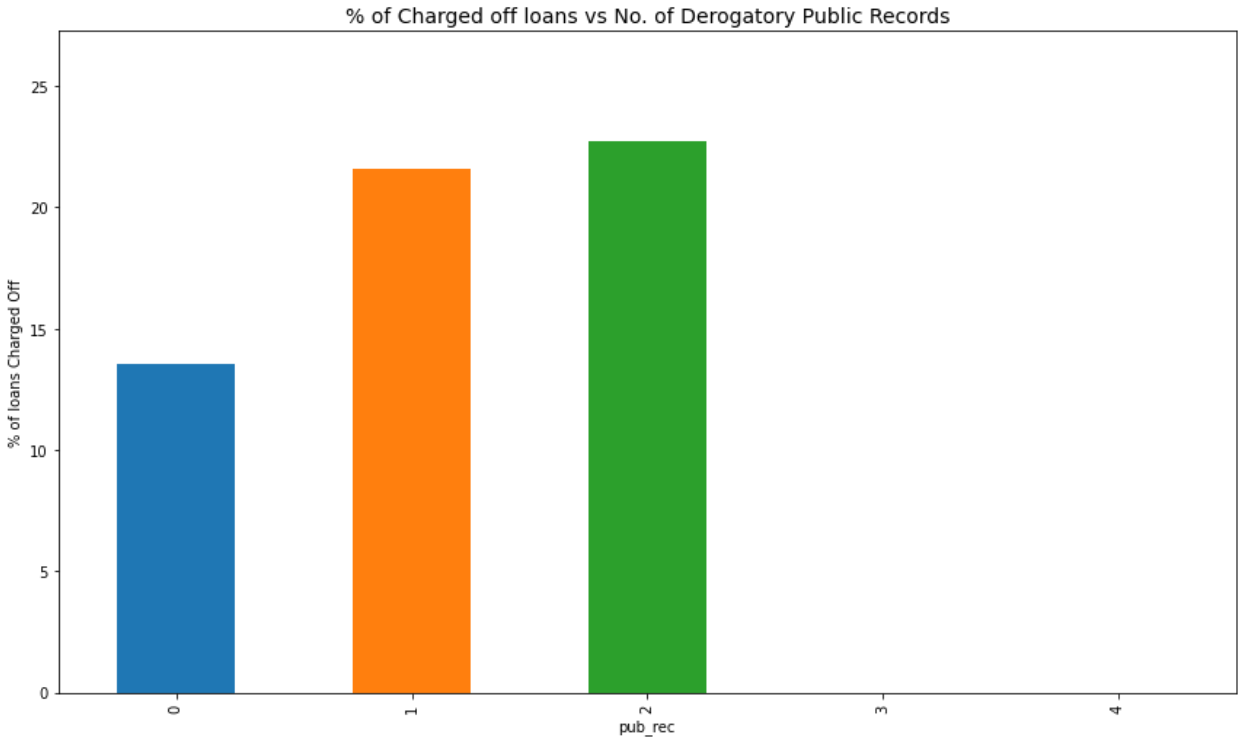
Analysis



Number of charged off loans increases as we keep moving ahead in the dti bins, which just concludes that dti is a strong factor in there being a default or not.

In that case, Lending Club should give loans only where the dti value in the credit score of an applicant is low.

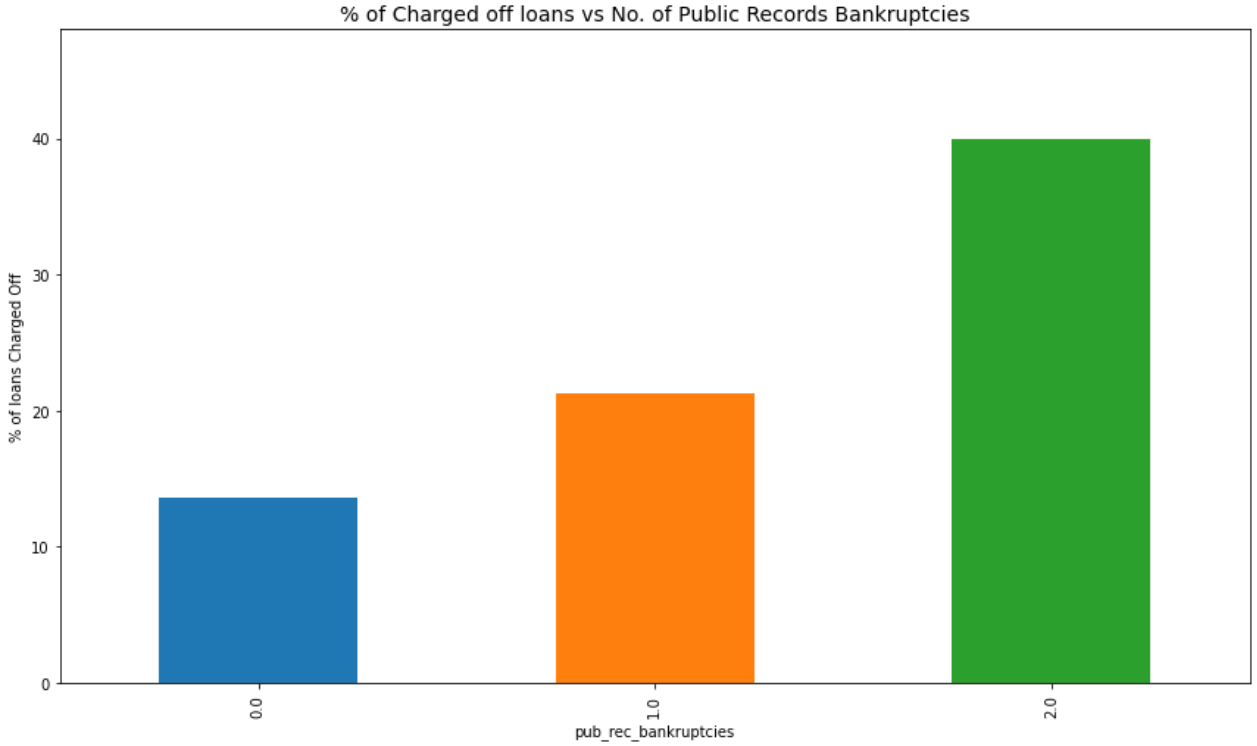
Analysis



This analysis suggests that more the number of public derogatory records, more will be the chances of default occurring.

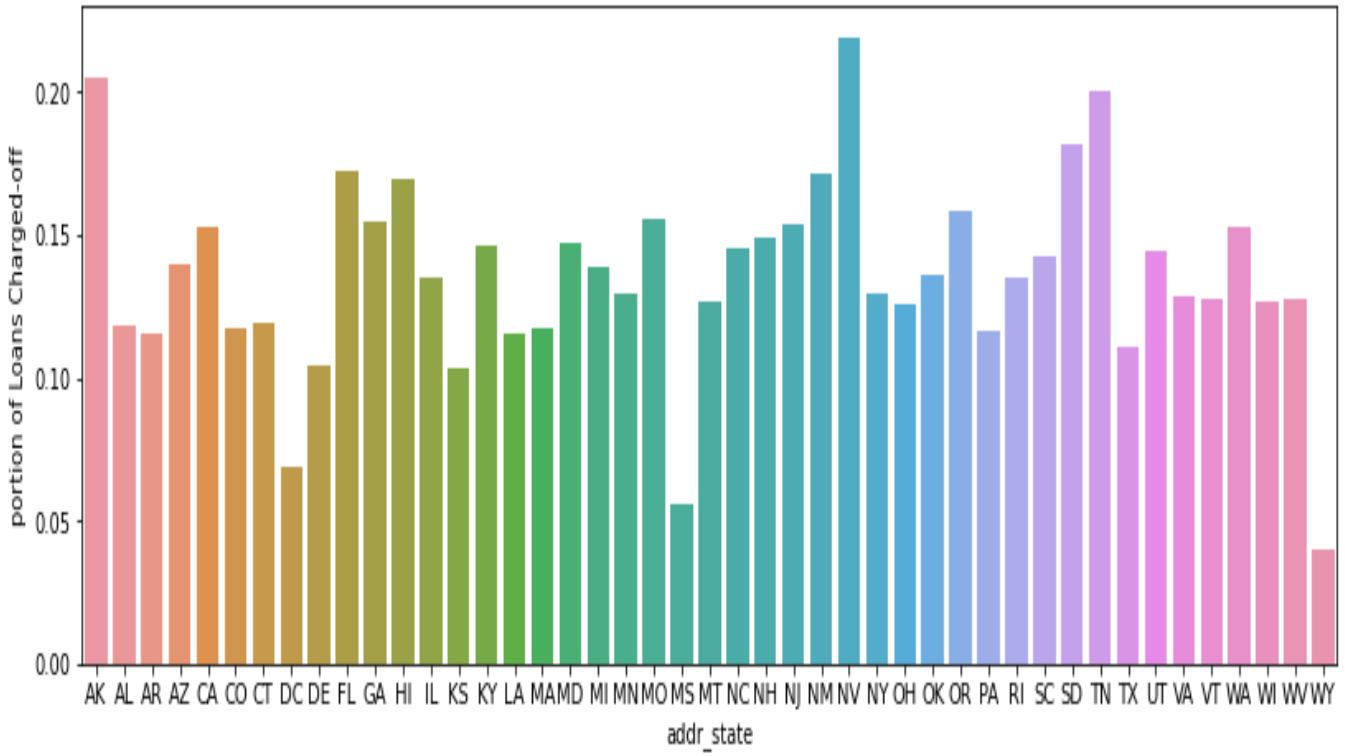
Thus Lending Club should disburse loan to only those who have no public derogatory records.

Analysis



Like the previous slide and its judgment, the same applies here: that Lending Club should not provide loans to those loan applicants who have even a single case of bankruptcy in their past.

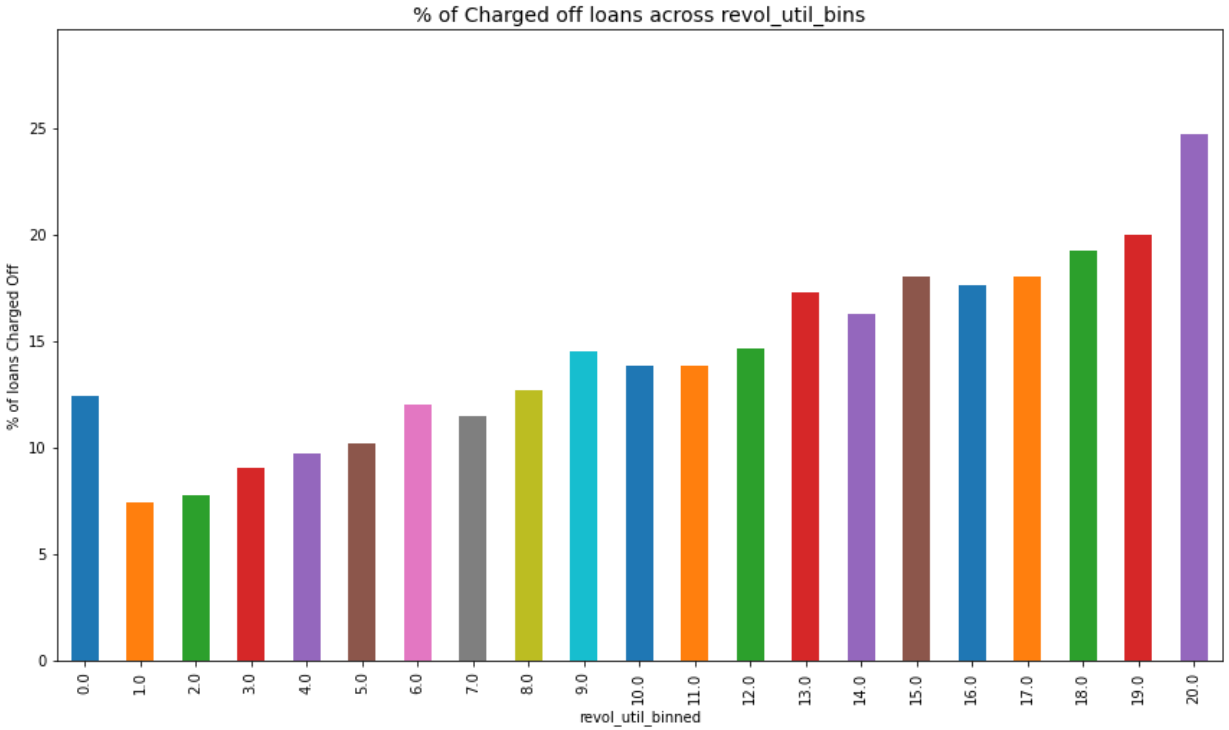
Analysis



From this plot it appears that most of the loans that produced defaulters had their applicants come from the state of Nevada, Arkansas and Tennessee and those that had the least chances of producing defaulters were from the states of Wyoming, Mississippi and Washington DC with all the other states and their likelihood of producing a defaulting loan applicant falling somewhere in the middle of these two extremes.

Thus, Lending Club should be cautious when providing loans to the states of Nevada, Arkansas and Tennessee at the very least.

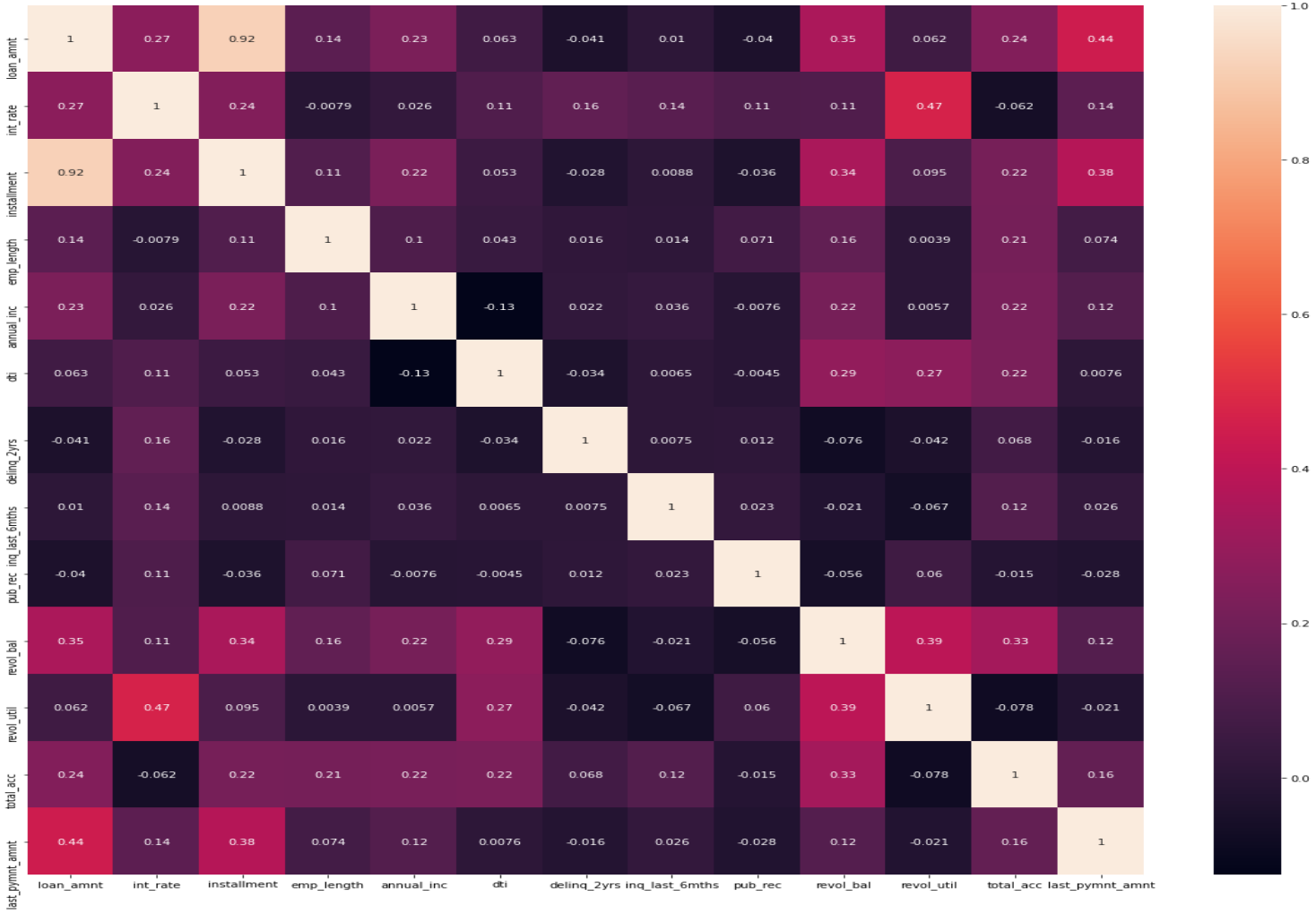
Analysis



From this plot we can deduce that the number of charged off loans increase as the revolving line utilization rate increases.

Thus this is a simple inference here, that more the revolving line utilization rate of a loan applicant, the more is his tendency to default on the loan and thus the Lending Club must not provide loan facility to such applicants.

Analysis



From this we can see that revolving line utilization rate and interest have a high correlation, as they should.

This is a good factor, as we have already inferred that more the revol_util, higher the chances of default, and thus to deter such loan applicants from taking loans, high interest rate should be charged.

Conclusion

After all the analysis we have done, we have arrived at the following conclusion to filter out the most suitable borrowers so Lending Club can have less number of defaults happening and thus save itself from financial/business losses:

1. Loans should be provided only If terms do not exceed three years.
2. Loans should only be provided to those applicants whose income is source verified.
3. Only Grade B and Grade A loans should be provided.
4. If employees of Walmart, UPS and AT&T and residents of Nevada, Arkansas or Tennessee are the loan applicants, then the process should be followed through but cautiously.
5. Loans should be provided only to the high income category applicants.
6. No loan applicant should have past derogatory records or have undergone bankruptcy in the past.
7. The revolving line utilization rate of the loan applicant needs to be low for them to be a safe borrower.
8. When the dti value in the credit profile of a loan applicant is low, they can be considered to be a safe bet.