

# Credit Risk Analysis: Years 2007 - 2001

Fiorella Tenorio

05/09/2019

## ABOUT THE DATA

I got this dataset from Kaggle (<https://www.kaggle.com/wendykan/lending-club-loan-data> (<https://www.kaggle.com/wendykan/lending-club-loan-data>)). This dataset is very rich containing a lot of variables that the company looks at to determine if they approved the loan or not. I got this information from Kaggle .

The main question I wanted to answer when I chose this dataset is: the US housing financial crisis occurred between 2007-2009 and it had a terrible impact in the economy not only in the US but all over the world; my question is, has the lesson been learned? Have we learned to be more careful deciding to whom do we grant loans? Or are we still accepting the risk of giving a loan to someone that may not be able to pay it back? Or even worse, granting a loan to someone that has a record of not paying their credit responsibilities on time? So, I will focus on trying to understand the client's credit health at the moment of granting the loan.

This is the question I will aim to answer with this dataset. The main challenge I am facing is that since I only have information from one institution, the conclusions I reached can't be easily escalated to the general US market. But it will allow me to get an idea of it. Also, something to consider is that, given that I am using data up until 2011, all the conclusions I draw will be within this window, I won't be able to escalate it to the current situation of the company.

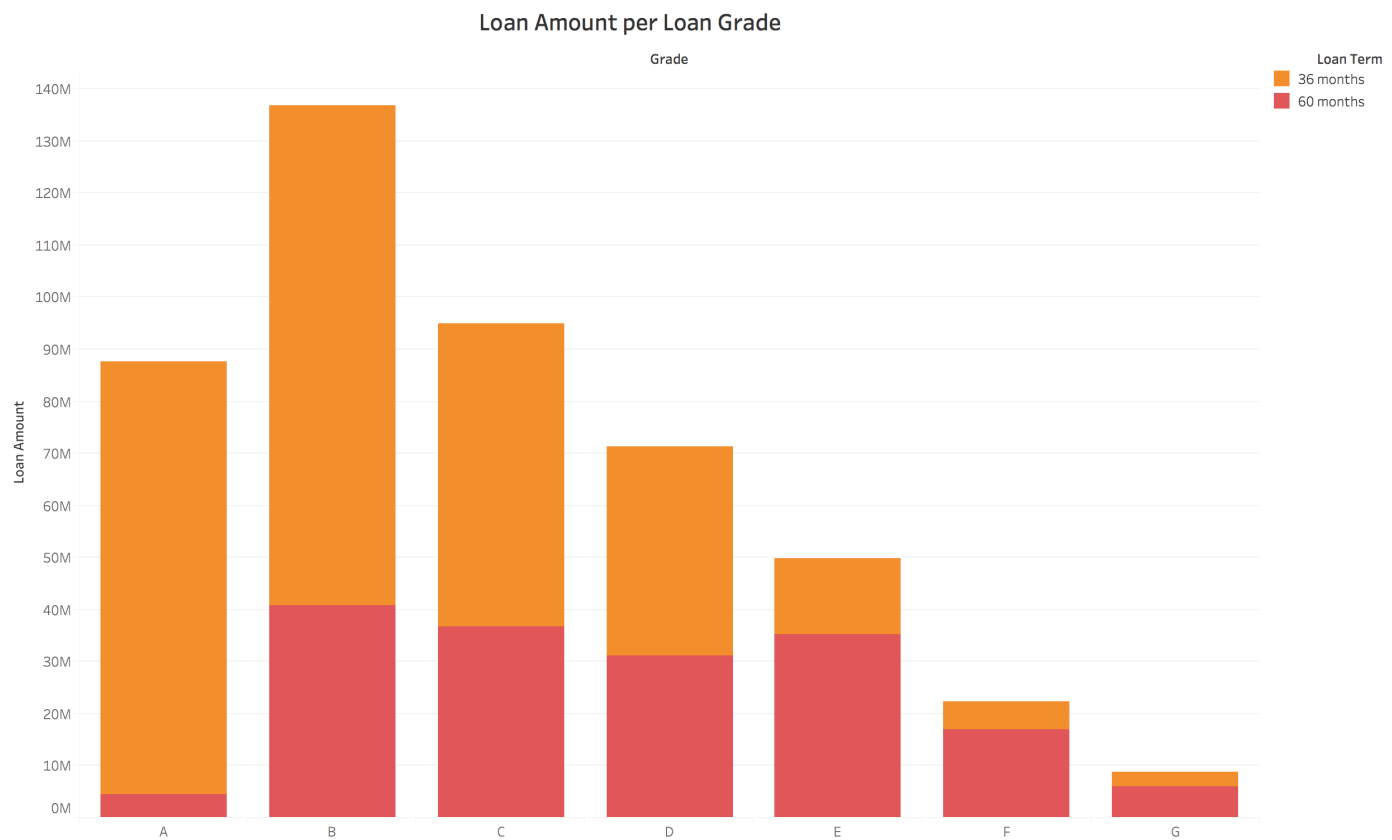
Something to know about the data to understand the plots better:

- **Grade:** This categorical variable goes from A to G and it indicates the risk level assigned to the loan. Beginning at A, which means very low risk, and ending at G, which means very high risk. This risk level is calculated by Lending Club and it is related to the borrower credit health. This is the variable I will be focusing on the most because it acts like a summary of the client health. I will be using this variable in most of my visualizations.
- **SubGrade:** It is a sublevel of Grade, it will open up the Grades in smaller levels by adding a number to the Grade, these numbers go from 1 to 7. So A1 means super low risk, and G7 means super high risk. This sublevel is also calculated by Lending Club to help investors decided in which loans to invest.
- **Investors:** Lending Club allows regular people (lenders) to invest in loans approved by Lending Club. They can invest in several loans at the time, and they can invest in the entire loan amount or just a percent of it. So, knowing this, a loan can be funded by several investors. And, the possibility exists that not the entire loan is funded by investors, there could be some percent that hasn't been covered by them.

## ANALYSIS

As I mentioned before, given the data that I have, I wanted to perform a simple analysis of risk appetite of the company, how much risk are they willing to take when granting loans? And the best variable to take a look at is Grades given that is a good summary of the risk level of the client requesting the loan.

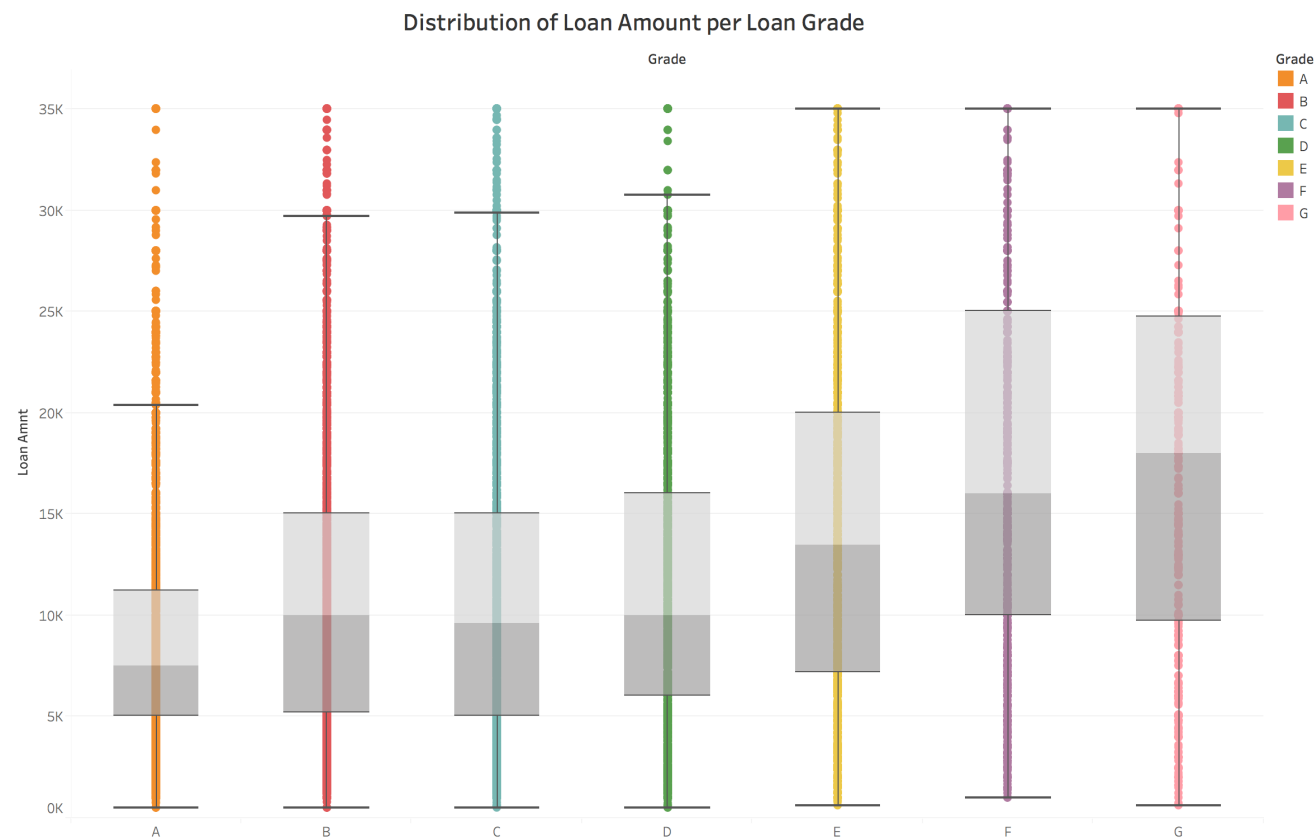
So, I started by doing a simple bar plot to show how much of the total loans granted where allocated to one grade or the other.



Sum of Loan Amnt for each Grade. Color shows details about Term. The view is filtered on sum of Loan Amnt, which keeps non-Null values only.

In the plot above, we can see that most of the loans are granted to grades A and B, which are the good grades, the ones that are assigned to clients that represent a low risk and that are very likely to pay back the loan. We can notice that for riskier Grades (C to G) there is decreasing trend in the total amount granted. That means that the company, while having a diverse portfolio, is focusing on less risky loans. I decided to open it up by Loan Term simply to get an idea of which term is preferred depending on the risk level, we can notice that as the risk increases, the preferred term for those loans is 60 months (grades from C to G); and for loans A and B, we notice that they are mostly short term loans (36 months).

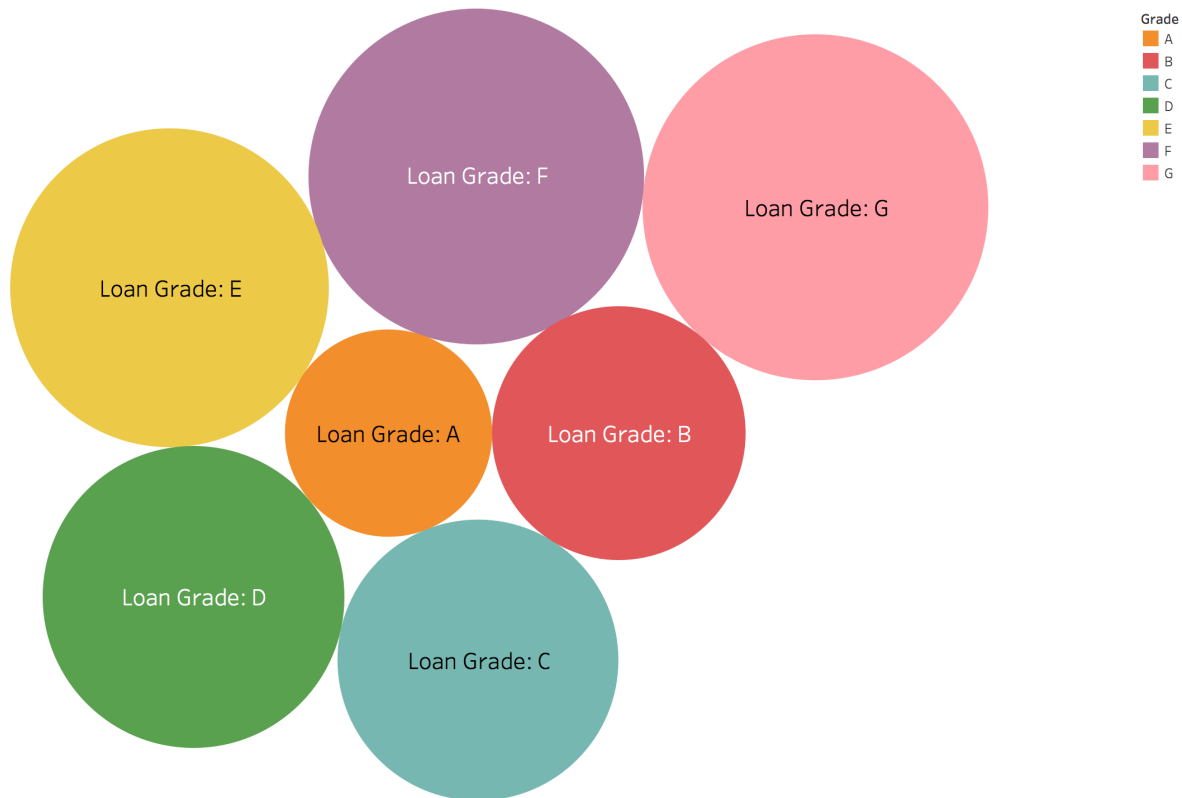
Now that we know that and feel a little better about the risk appetite of the company, let's take a look at the amount of the loans per grade but let's take at the average amount granted to each grade.



This is interesting. In *Figure 1* we noticed that there was a lot of allocation for A and B grades and not too much for grades from C to G. But in this visualization, we can see that the average loan amount granted for grades A and B is much less than the amount granted for grades C to G. There actually seems to be an inverse correlation between the grade and the average loan amount approved: as the grade gets worse (riskier borrowers), the loan amount granted gets larger. This is interesting because we can say that even if we don't allocate much to risky grades, they are large allocations, large in terms of money, not quantity.

Ok, so not a lot of risky loans are being granted but the few that there are, are large amounts of money. I would like to know now if the company is compensating taking the risk for those loans by assigning them a higher interest rate than to less risky loans. In order to review that, I will use a Bubble Map where each bubble represent a Grade and the size of the bubble is dictated by the interest rate for that grade.

Loan Interest Rate per Loan Grade from June 2007 to December 2011

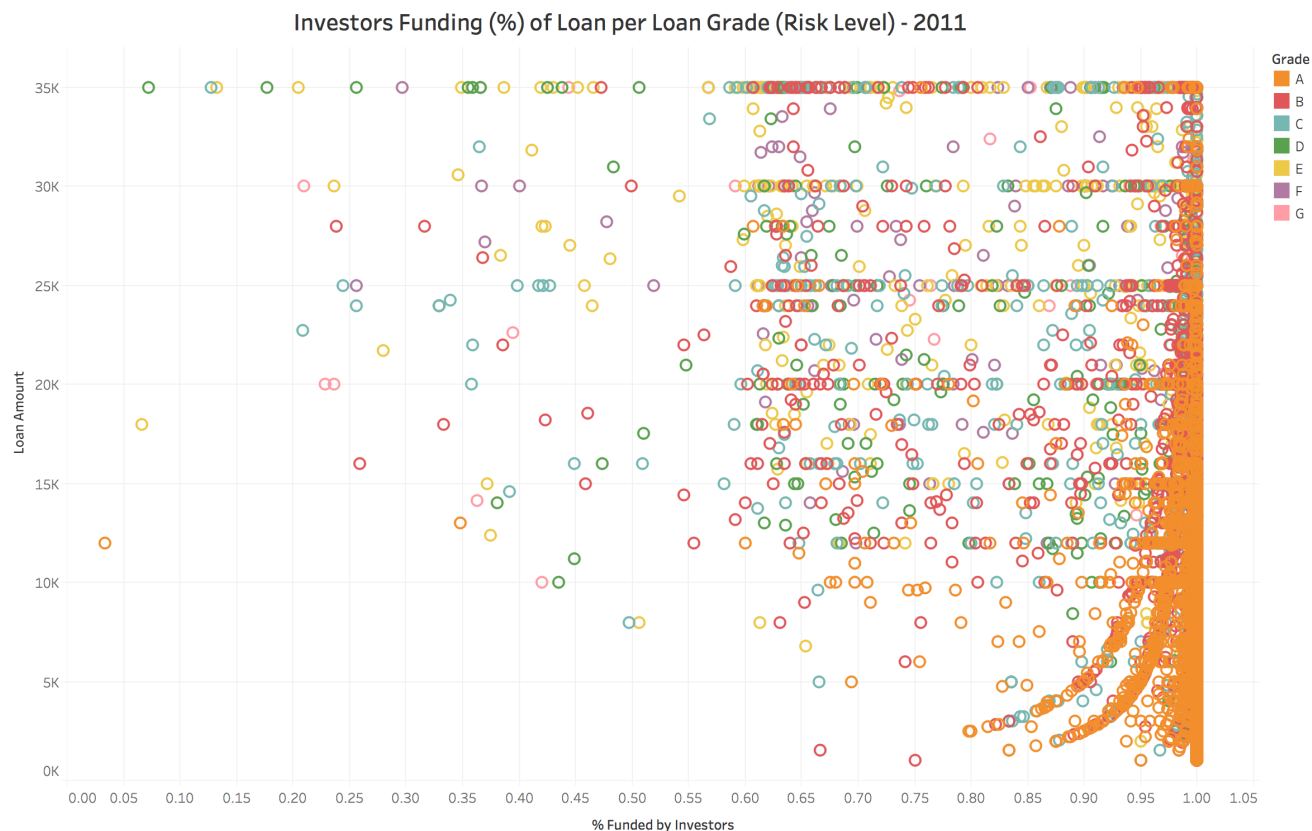


Grade. Color shows details about Grade. Size shows average of Int Rate. The marks are labeled by Grade. The view is filtered on average of Int Rate, which keeps non-Null values only.

We can observe in the visualization above that the size of the bubbles increases as the risk level (grade) gets worse. This is good to know, because it means that the company is compensating the risk they are accepting by giving the borrowers a higher interest rate than to less risky borrowers.

Ok, so up until this point, I think I have managed to get enough information to answer my question, as I mentioned before, given that I only have information from one company, it is not possible to extrapolate this and make a general conclusion about the overall US risk appetite during 2007 - 2011.

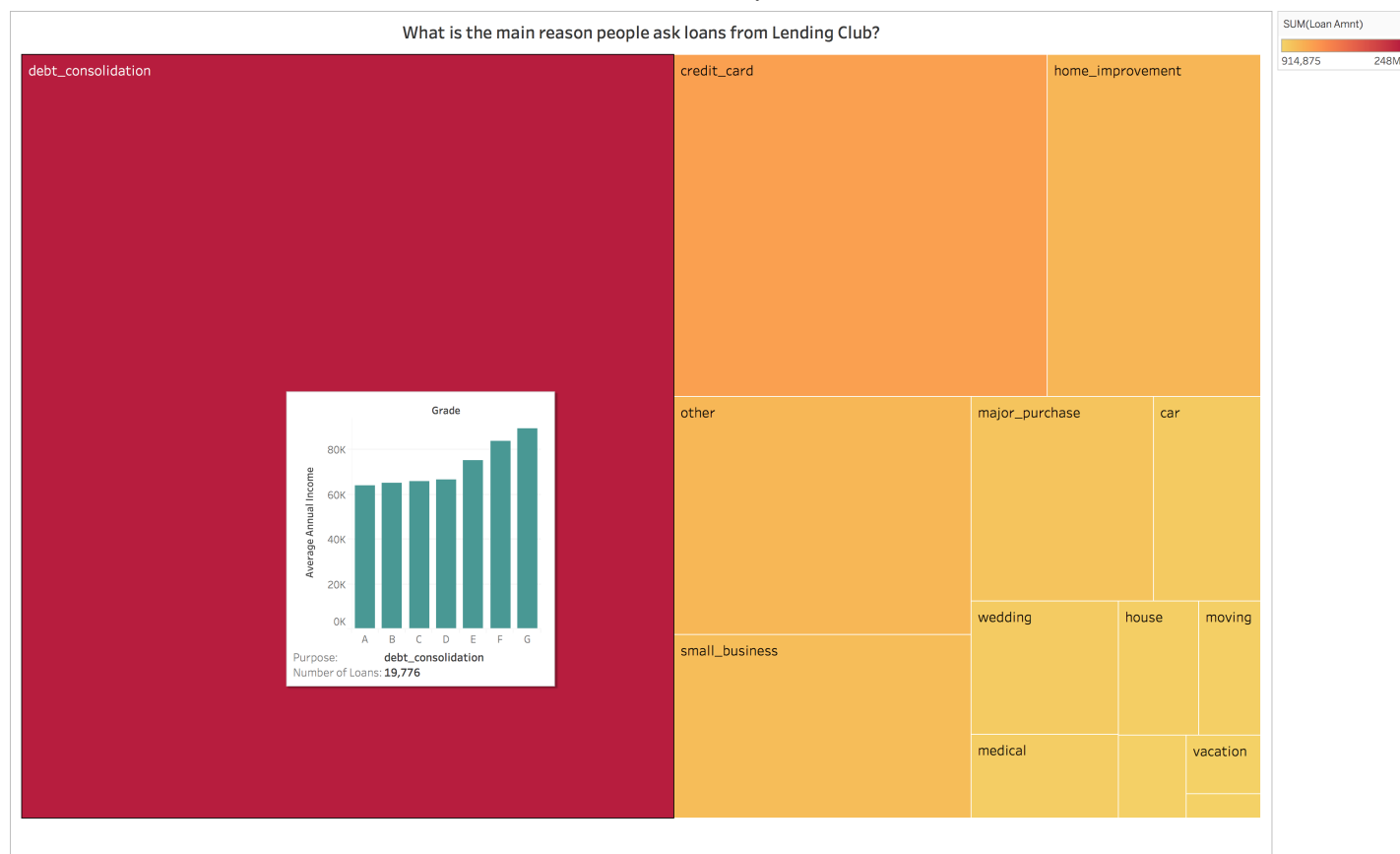
So far I've checked everything from the company's side, knowing about the company's risk appetite, but what about the investors? Can we say something about their risk appetite? I had mentioned that, even though the company approves the loans, investors are the ones that give the money to fund the loans, and not always all of it, something only part of the loan is funded by investors. So, in order to get an idea of the investors risk appetite, I generated the following visualization:



% Funded by Investors vs. Loan Amnt. Color shows details about Grade. The data is filtered on Issue D, which ranges from 1/1/2011 to 12/1/2011. The view is filtered on Grade, which has multiple members selected.

In this scatter plot, each point is colored based on its grade, the x-axis represents the percentage of the loan amount that was funded by investors, and the y-axis represents the loan amount. A fast conclusion we can reach by looking at the graph, is that loans graded A (orange) are funded almost completely, the majority falls between 90% and 100% of funding. And for the other loans, there is a lot of variance about how much of it is funded by investors. Knowing this, I think we can make an inference about the company's investors: even though we seem to have some risk taking investors, the majority of the investors prefer to invest in safe loans.

And to finish this analysis, I was curious about what is the main reason people go to the company to request a loan. It is not related to the my main question, I was just curious about this. So, I created a Tree Map where each block represents a reason for the loan, and the size of the block is determined by what percentage of the total amount of loans is due to that reason.



From the graph above, we can see that the main reason is debt consolidation. I also wanted to see what is the grade distribution for that reason. That is why when you hover on a block, you will see a barplot showing the distribution of the loan amounts per Grade. We can see that even though most of the portfolio is made up by debt consolidations, they seem to be very risky investments, most of them are graded from C to G.

## CONCLUSIONS

The question I set out to answer at the beginning was if, right after the financial crisis of 2007, the company was more careful when deciding to whom they granted a credit to. Ideally, I would've liked to answer this question not only for this company but also for the entire US financial system. I need more data to make a conclusion like that, but with what I have, I can make a conclusion for this company. This company's clients are both the borrowers and the investors, for the borrowers, it is in their best convenience that the company is lenient when assigning a grade; and for investors, it is in their best convenience that the company is harsh when assigning a grade.

The company seems to be very interested in having a diverse portfolio, based on risk level, but their focus are not risky loans. So, it seems like yes, the lesson has been learned.