

Lending Club Case Study

SUBMISSION

Group members: Jennis Vaishnav
Pranay Bhavaraju

<Abstract>

We are developing a number of data-driven investment strategies that show how to use exploratory data analysis for lending club case study. We'll take a closer look at the process from getting data from a CSV file to developing and evaluating investment strategies based on different approaches. We'll also take a look at how a variable affects another variable.

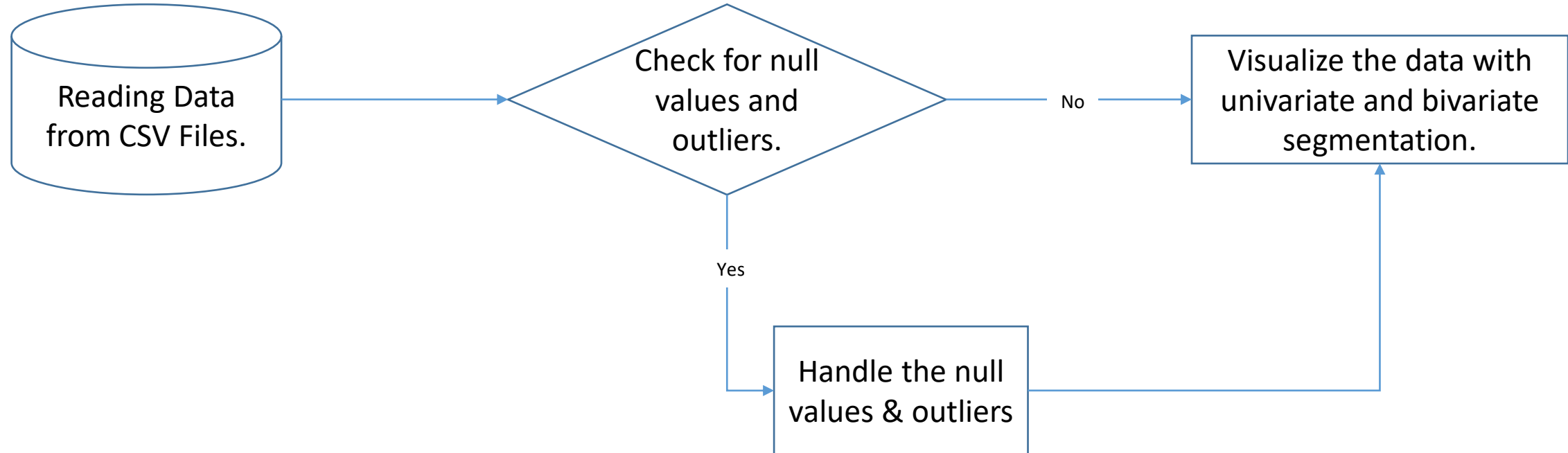
We focus on how to apply and evaluate data science techniques in a real-world business environment and the resulting strategies. The analysis presented in this presentation can suggest to the lenders what parameters to look at, before lending a certain amount to the borrower. What is important is that we've gone beyond the analysis of the data, but there is a thorough and critical observation.

Our treatments are comprehensive and range from univariate to bivariate, strictly speaking, from a business perspective. The visualization process includes the following: data cleaning and truncation were ever needed, provided we don't remove important/critical data, classification/probability estimation based on the plots, Python performing exploratory data analysis for lending club case study.

Here we have removed the data which is no longer needed for the part of analysis or which would just deviate our results from the actual analysis.

<Problem solving methodology>

Data Gathering and Cleaning



<Analysis>

1. First we imported the files and libraries necessary for performing the analysis operation.
2. There were numerous columns which were totally empty which were removed as they were not fit for analysis.
3. From the data dictionary the parameters were identified which were likely to have an impact on rather the borrower would turn out to be defaulter or fully pay the loan
4. Several analysis were made by using boxplots, bar charts where ever necessary.
5. In certain cases where there was not enough significance generated, then in such cases they were double checked by doing same analysis and visualizing them in different way.
6. It was found that employee experience, rate of interest, tenure, monthly income, verification, played a crucial role in the analysis.

<Analysis>

7. Here a quick visualization was done for those who have paid the loan fully and those who have defaulted the loan.

8. It can be seen that there are more number of defaulters than the number of borrowers who actually completed the loan on time.

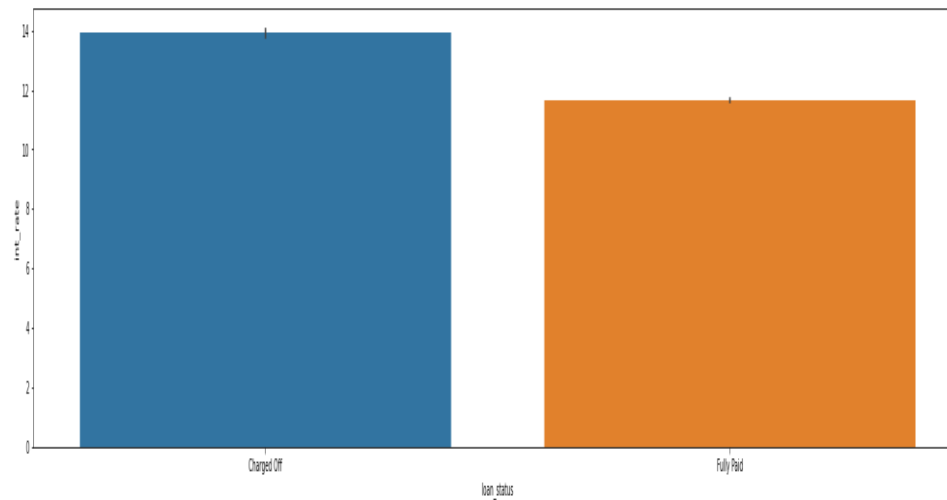
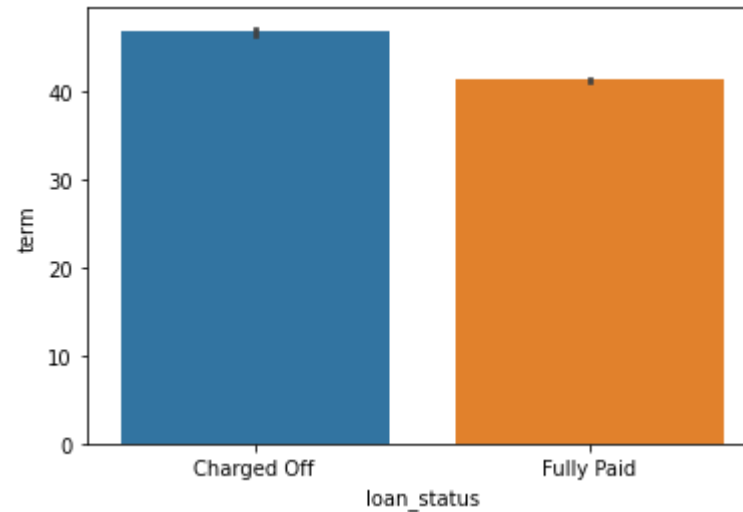


Figure:1.0

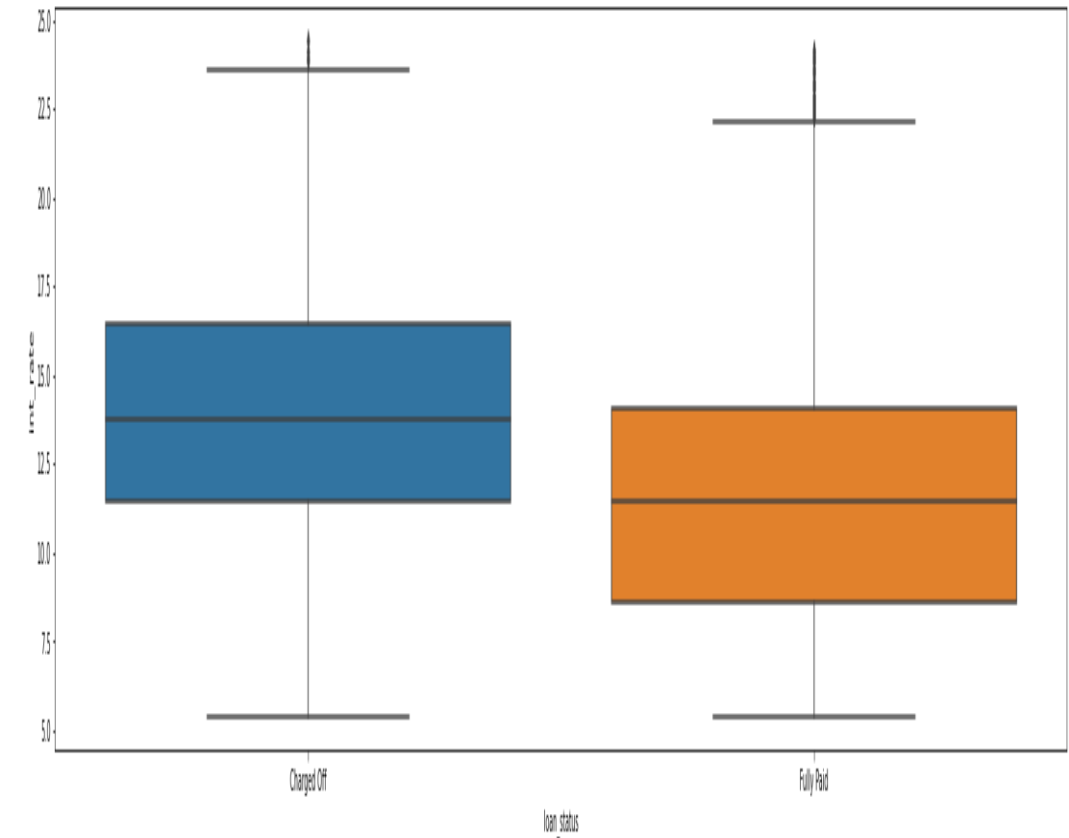
<Analysis>

9. Here we have taken the number of loans paid with respect to the term of the loan.
10. Here it turned out to be that higher the term the most likely the loan is to be defaulted.
11. The loans which were having less tenure are being paid.
12. So it would be ideal to lend a loan which has a tenure matching the below barplots.



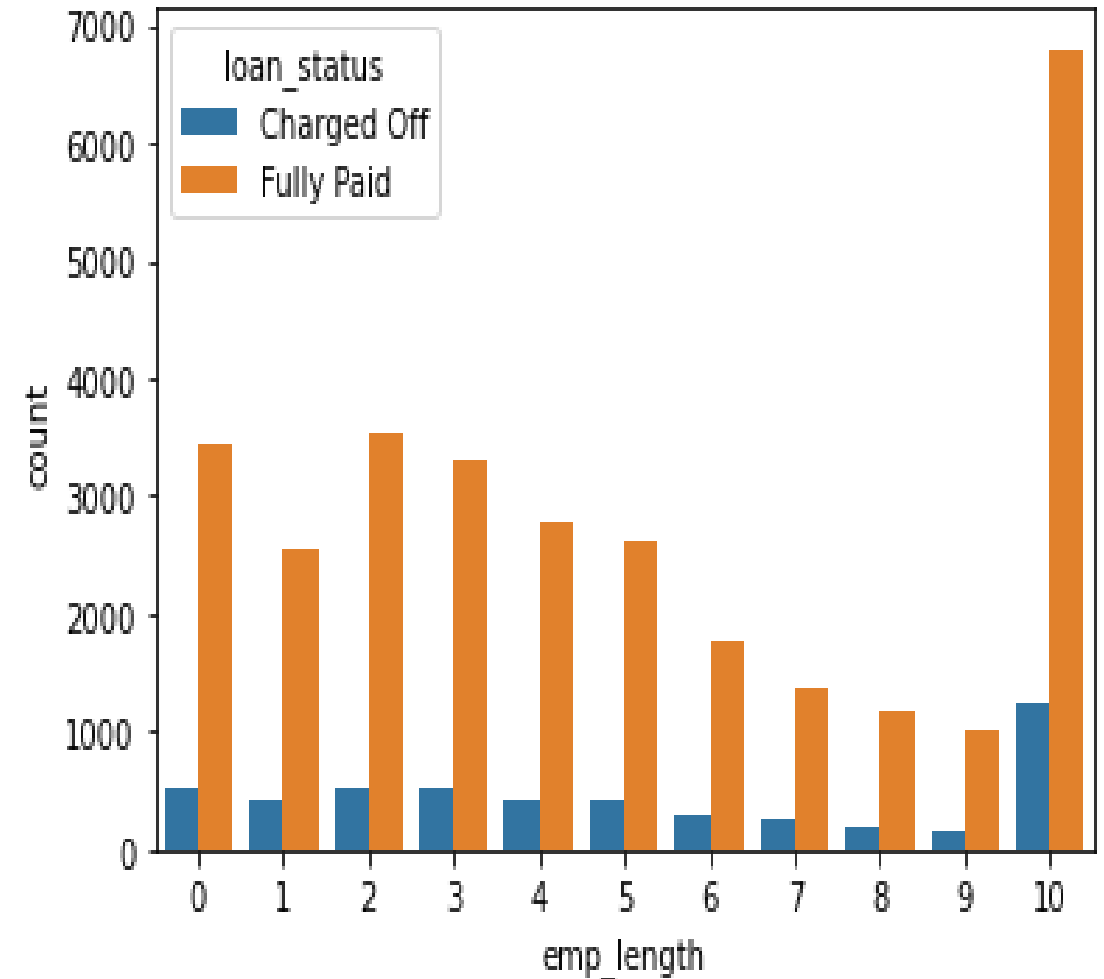
<Analysis>

- ❑ Here there are two boxplots loan status(X-axis), and rate of interest(Y-axis).
- ❑ Here a close thing to observe is that the median of defaulters(Charged off category) is quite higher than the ones who have paid the loan.
- ❑ This implies that the higher the rate of interest the higher the rate of defaulters would be. That is higher the possibility of borrower to default the loan increases with increased rate of interest.



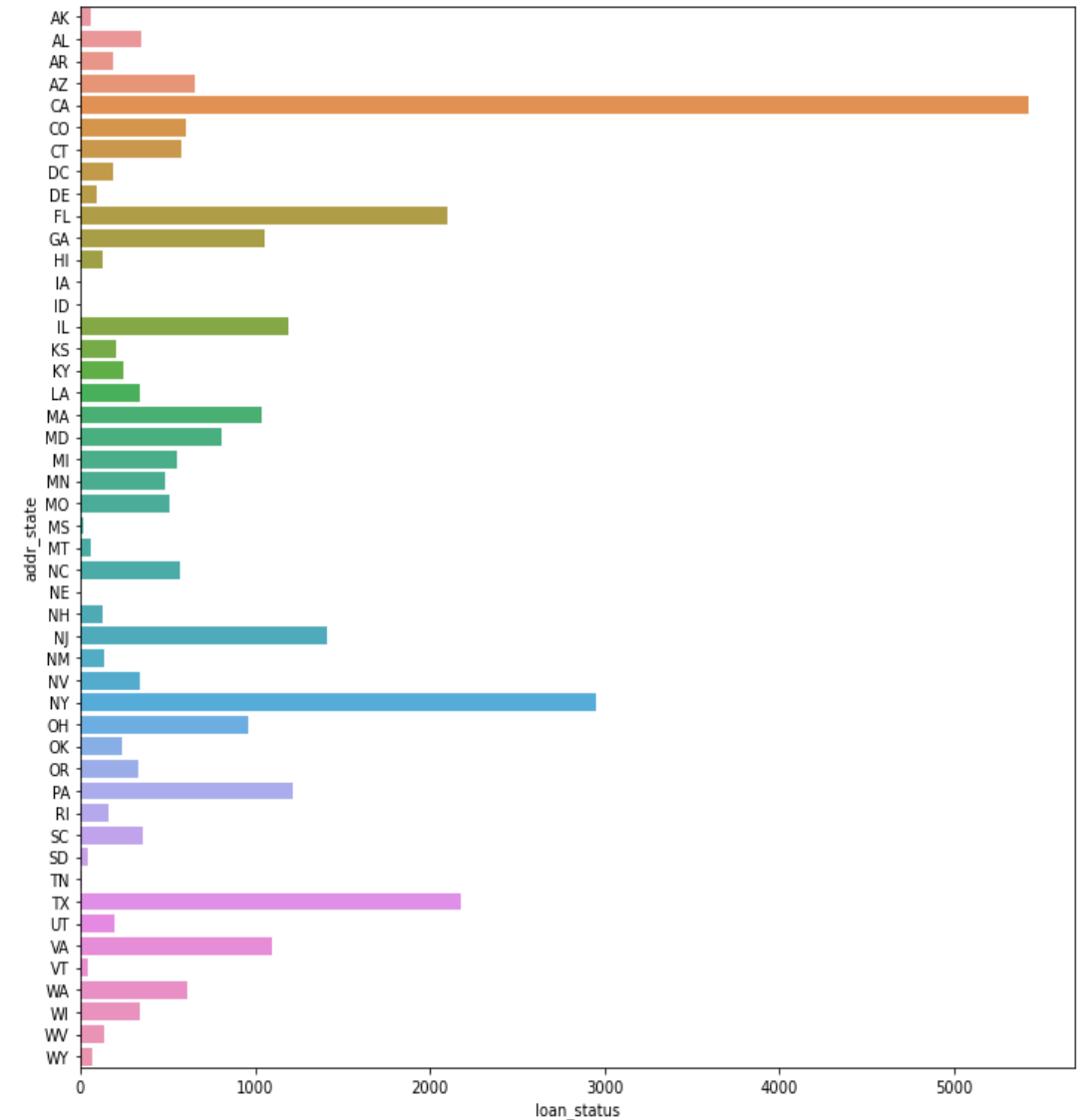
<Analysis>

- ❑ Here we have a grouped bar plot. Where we have employee length and their loan status.
- ❑ Here it was found that higher the employee is experienced, the more likely he would be able to pay off the loan.
- ❑ Employee length is the number of years of experience the borrower has as an employee.
- ❑ The employee who has higher experience are, that is, greater than 10 years are most suitable ones for our lending the loan.
- ❑ Consequently the rate of defaulters also increases with experience of the employee. The same customer who is paying loan with >10 years of service, that very same category is having the highest default rate to. So employee experience becomes an important parameter for analysis.



<Analysis>

- ☐ Here we have analysed state(X-axis), and defaulters on the Y-axis.
- ☐ Here we came to know that a state named CA has highest number of defaulters.
- ☐ It can used if a person is living in one of states where the rate of defaulters is higher, one can think twice before lending the loan.
- ☐ This plays a signific role because it shows us, what is the exact financial, economic situation in that particular state.
- ☐ If a state has poor financial situation then it implies that we should do a proper check on the borrower and then an then lend the loan to the borrower.



<Conclusions>

Based on the above slides we have identified the following :

- i) The rate of interest is an factor which effects the analysis.
- ii) Years of experience of an employee/borrower also plays an significant role
- iii) The amount to lended also plays an significant role.
- iv) The state of the borrower also plays an important role on the analysis.
- v) The tenure of the loan also plays a significant role. Higher the tenure the higher are the chances of a loan to get defaulted.

Thank you