

# Lending Club Case Study

## EDA

---

Manish Bindal – ML C36

9<sup>th</sup> February 2022

# Problem Statement

The largest online loan marketplace company wants to understand the driver variables which are strong indicators of loan default. They want to utilize this knowledge for their portfolio and risk assessment.

Data for all the issued loans from 2007 to 2011 is shared for this company. Objective is to analyze this data set using EDA to identify these driver variables.

# Analysis Approach

---



Standard EDA steps and techniques like “Data Cleaning”, “Data Manipulation”, and “Analysis” are used to produce the observations and recommendations.



Data cleaning and manipulation includes removal of columns having high null/missing values, having same value across rows, and having unique values across all rows, removing outliers, standardizing values, and splitting columns like obtaining issue month and issue year from issue date.



Data analysis is accomplished by analyzing one column at a time and then analyzing two columns together to study the effect on loan status.

# Observations

Five strong indicators of loan default are:

- ✓ Grade (lower the grade, higher the chance of defaulting the loan)
- ✓ Interest Rate (defaults are more with higher interest rate)
- ✓ Debt to Income Ratio (if debt to income ratio is higher than more chances of defaulting the loan)
- ✓ Public Record of Bankruptcies (with public record of bankruptcies, there is a higher chance of defaulting the loan)
- ✓ Term (high chances of defaulting the loan if payment term is 60 months)

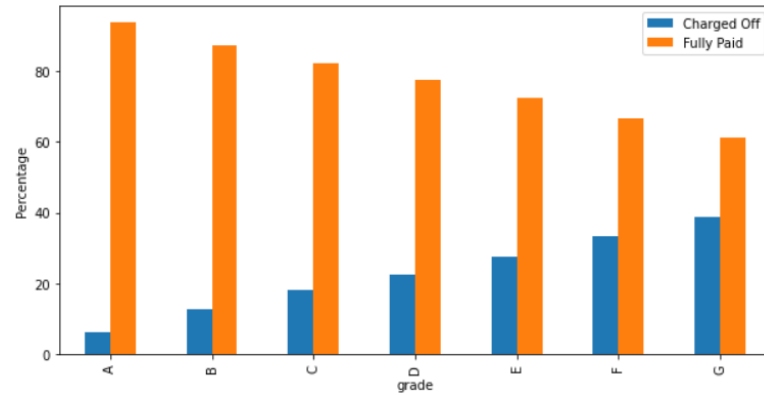
These indicators will be discussed in detail in further slides along with recommendations.

# Recommendations

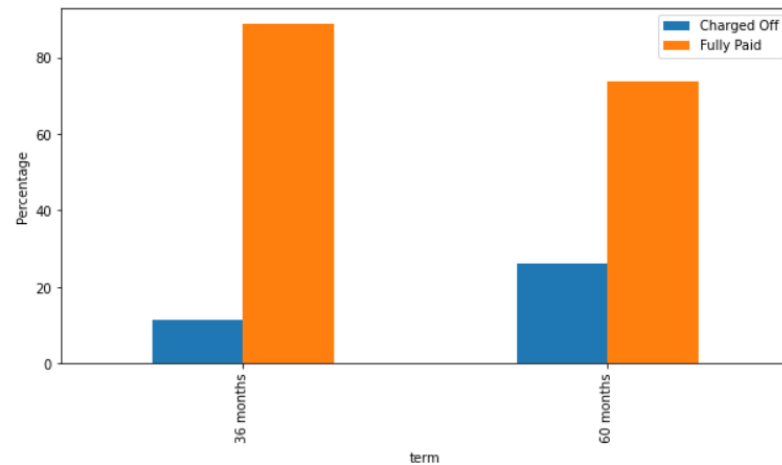
Following are the recommendations from the analysis:

1. Implement business rules based on the observations around following five variables before approval of the loan applications. ○
  - ✓ Grade
  - ✓ Interest Rate
  - ✓ Debt to Income Ratio
  - ✓ Public Record of Bankruptcies
  - ✓ Term
2. Extra level of scrutiny is required if borrowers are from Nebraska or loan is taken for small business.
3. Loan defaults are high if funded amount by investors is high or annual income is low. Further analysis (multi variable) should be conducted to find more insights and design better loan approval process.

# Details on Recommendation# 1

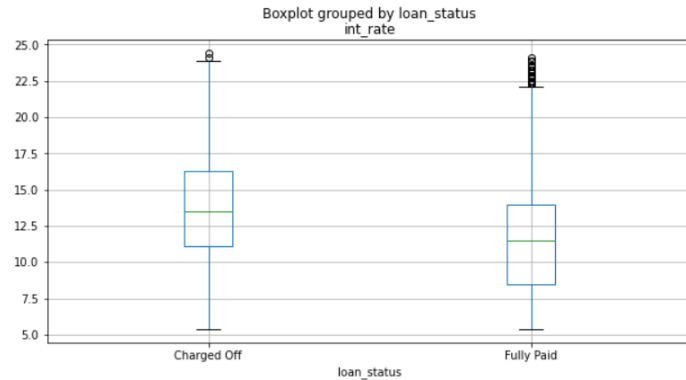


Percentage of loan issued among different grades are plotted and grouped by loan status in the left side graph. This clearly shows that as the grade gets lower, loan default percentage becomes higher.

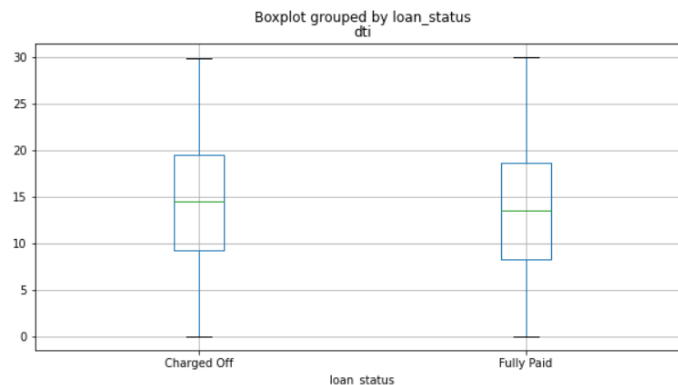


Percentage of loan issued between two terms i.e. 36 months and 60 months are plotted and grouped by loan status in the left side graph. It is evident from it that extra amount of time to pay back the loan contributes towards higher percentage of loan default.

# Details on Recommendation# 1 continued....



Box plots for loan status values (Charged Off, Fully Paid) are plotted against interest rate (y-axis). Lower edge of the box represents 25<sup>th</sup> percentile of total population and upper edge of the box represents 75<sup>th</sup> percentile. It is clear that defaults are higher with higher interest rate.



As explained above, same box plots for loan status values (Charged Off, Fully Paid) are plotted against debt to income ratio this time. It is not as evident as interest rate but still it can be observed that with higher debt to income ratio, chances for loan defaults are higher.

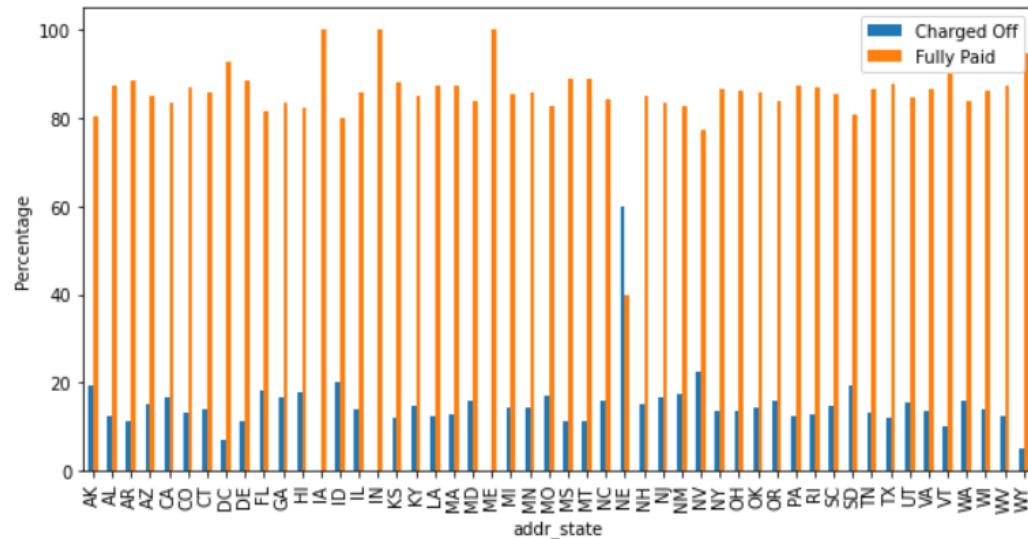
Mean of total public record of bankruptcies : 0.05

Mean of public record of bankruptcies based on loan status:

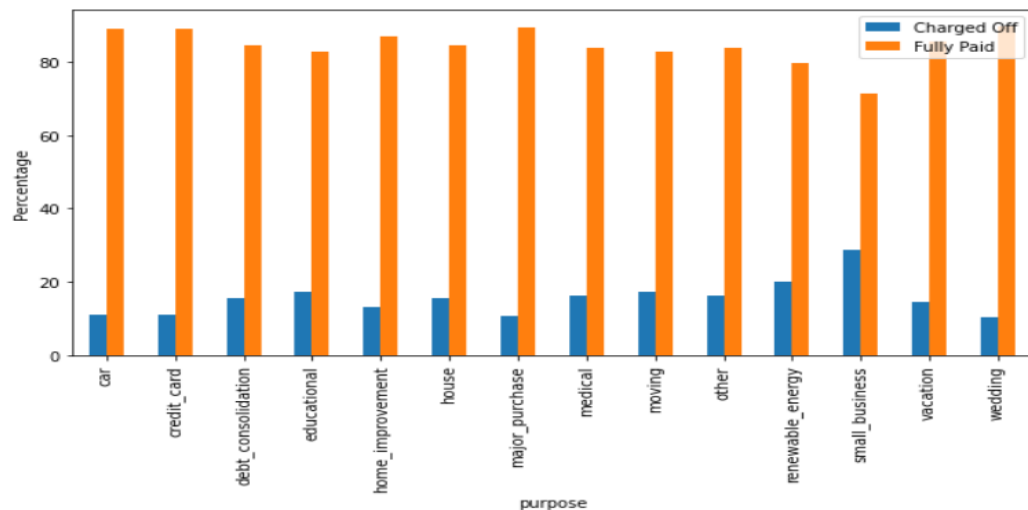
-----  
loan\_status  
Charged Off    0.07  
Fully Paid     0.04

Mean value of the public record of bankruptcies for entire data set is 0.5, whereas it goes up to 0.7 for “charged off” loan status which means borrowers having bankruptcies record tends to have higher chance of defaulting the loan.

# Details on Recommendation# 2



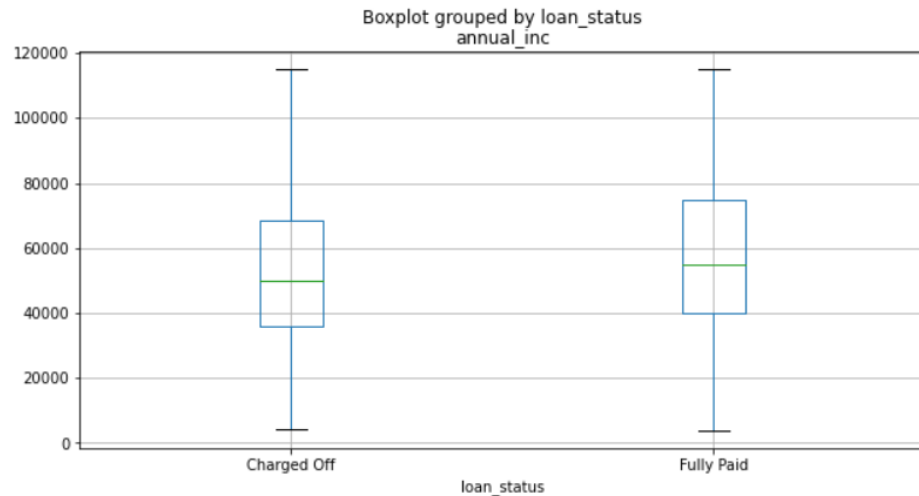
Percentage of loan issued among different states are plotted and grouped by loan status in the left side graph. This clearly shows that borrowers from Nebraska have the highest percentage of loan default.



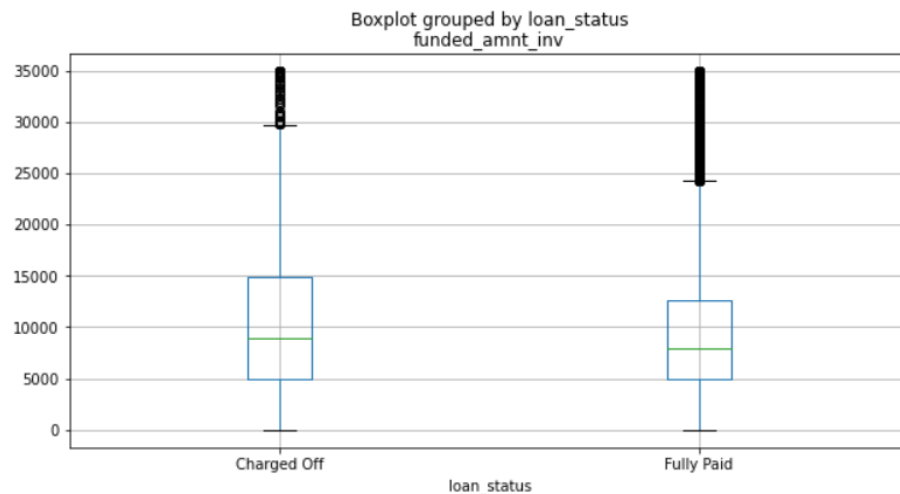
Percentage of loan issued among purpose values are plotted and grouped by loan status in the left side graph. This clearly shows that loan taken for small business have the highest percentage of loan default.



# Details on Recommendation# 3



As explained earlier, box plots for loan status values (Charged Off, Fully Paid) are plotted against annual income. It seems that lower income group tend to have a greater number of loan defaults.



This time the box plots for loan status values (Charged Off, Fully Paid) are plotted against funded amount by investors. Though at the lower level of fund amount given by investors, loan defaults are almost equal to loan paid but at the higher amounts, loan defaults are high.



Thank You !

