

# Lending Club Case Study

## Problem Statement:

- Given the past data of a lending company, apply the basics of data analysis and come up with meaningful conclusions about its defaulters

## Approach:

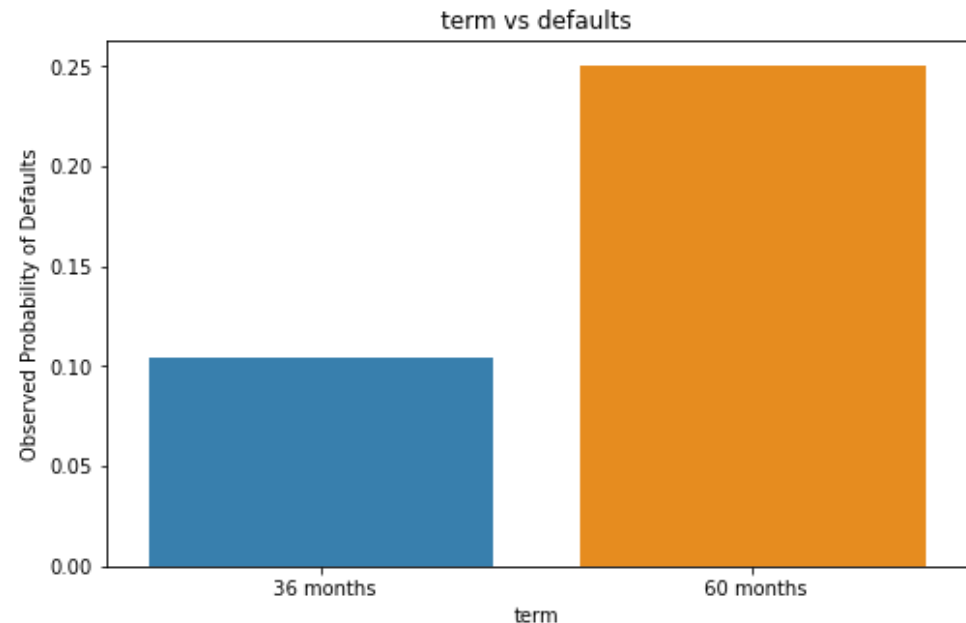
- We started by Understanding the problem domain and the variables in the data. We moved onto to cleaning up the data, removing outliers, if any and formatting them.
- For the data analysis itself, we began with univariate analysis and segmented univariate analysis to identify the driver variables. We then extended this with Bivariate analysis.

## Note:

- If the presentation is viewed via the GitHub browser, please click on “More Pages” at the end of page 5, to view all the pages & conclusions

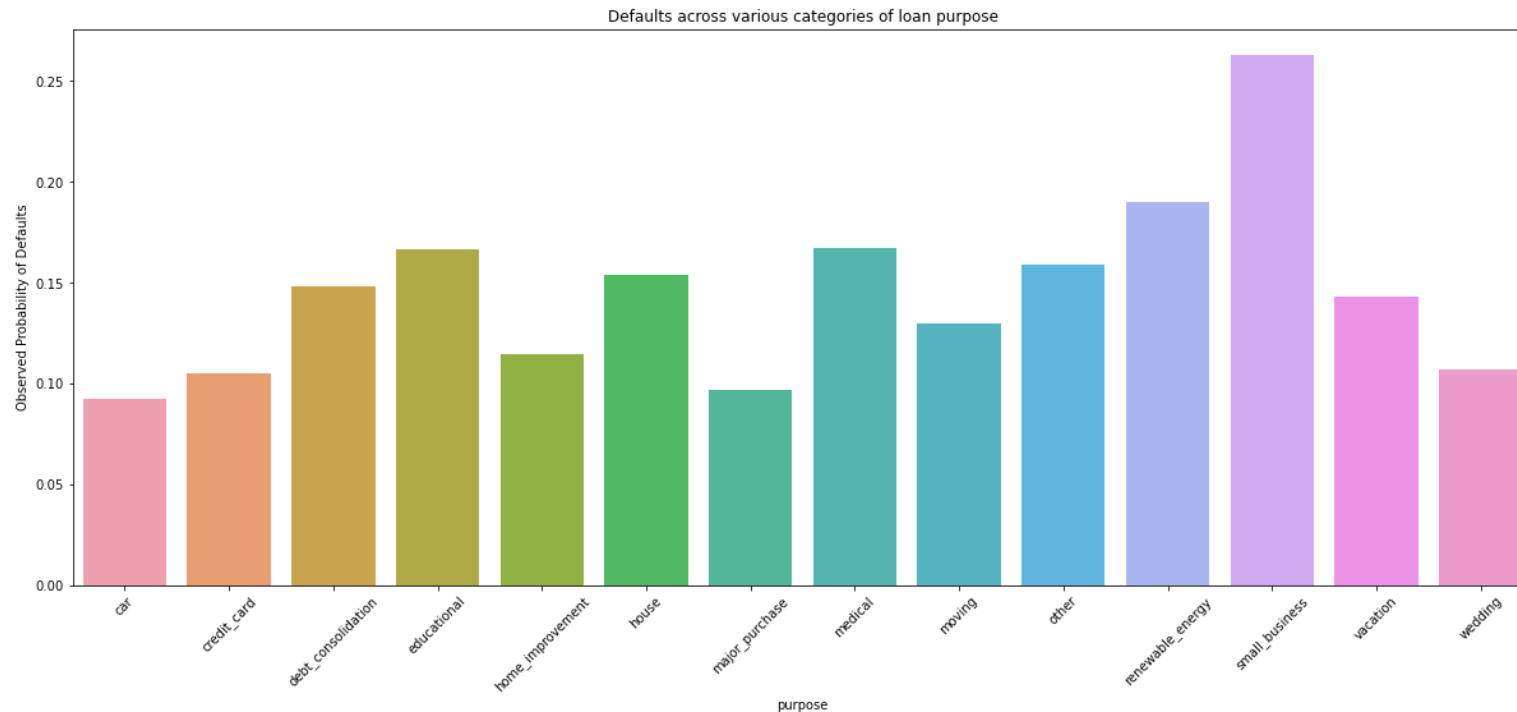
# Conclusion 1

- “**term**” is a strong driver variable of default
- **At least 1 out of 4 loans (25%)** in “60 months” end up defaulting
- This helps the company to manage their portfolio by **not allocating a high number of loans** in the high risk “**60 month**” category



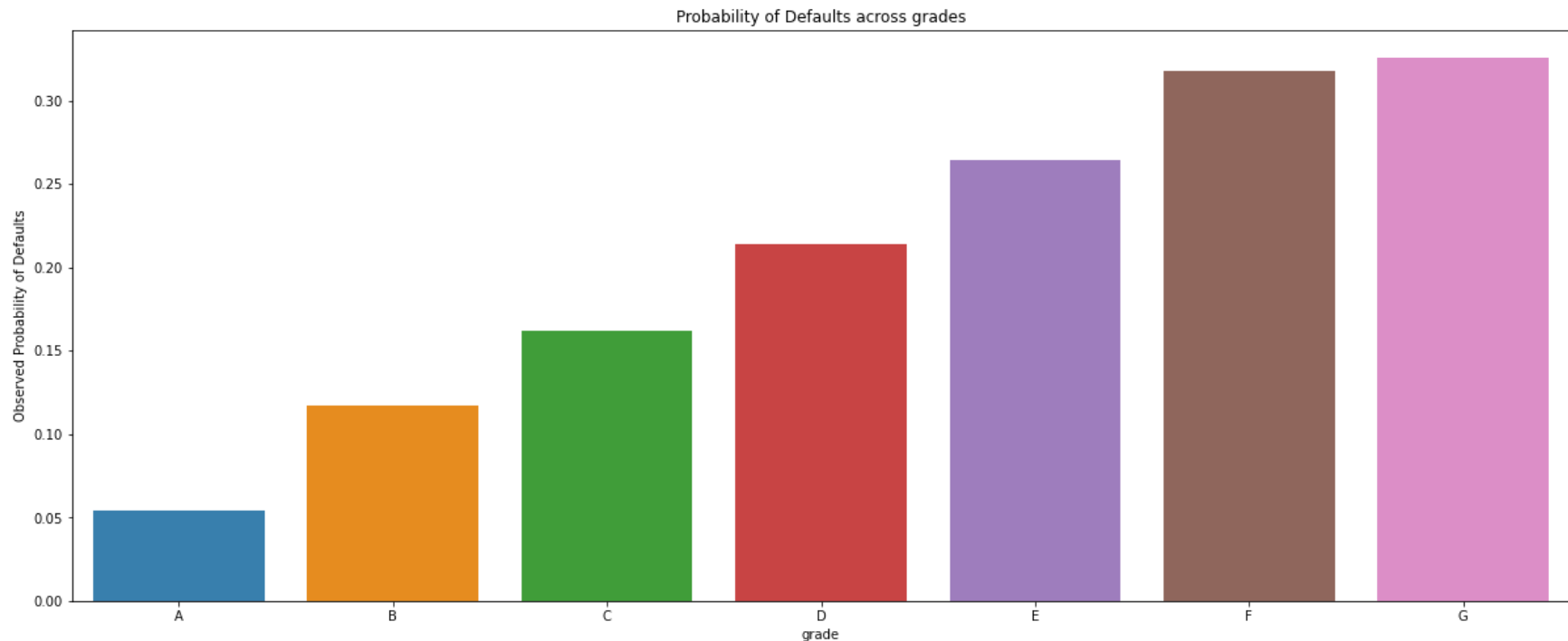
# Conclusion 2

- “**purpose**” is a strong driver variable of default
- Ex: at least **1 out of 4 loans** (>25%) in “**small business**” end up defaulting
- This helps the company to manage their portfolio by **not allocating a high number of loans in the high risk “purpose” categories like “small business”, “renewable energy”**.



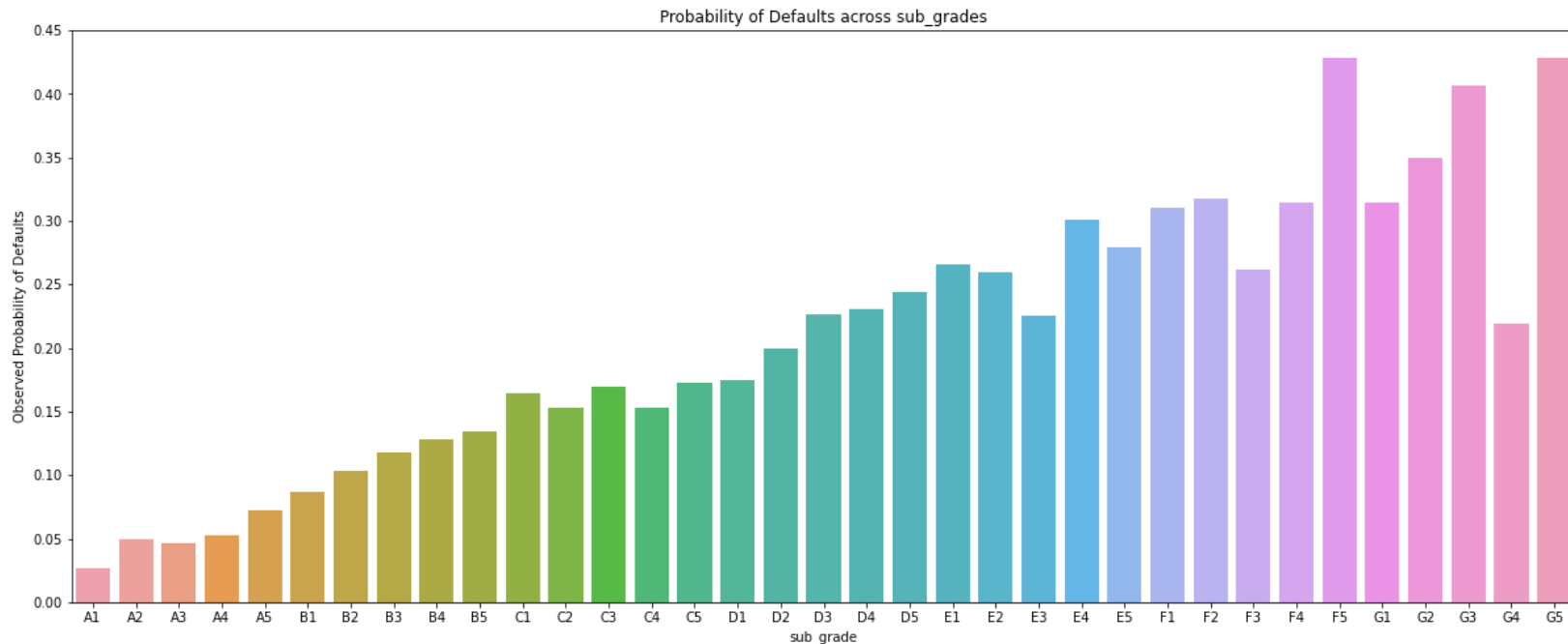
# Conclusion 3

- “**grade**” is a strong driver variable of default
- Ex: **more than 30% of loans in grade G & F** end up defaulting
- This helps the company to manage their portfolio by **not allocating a high number of loans in the high risk “grade” categories like “G”, “F”**



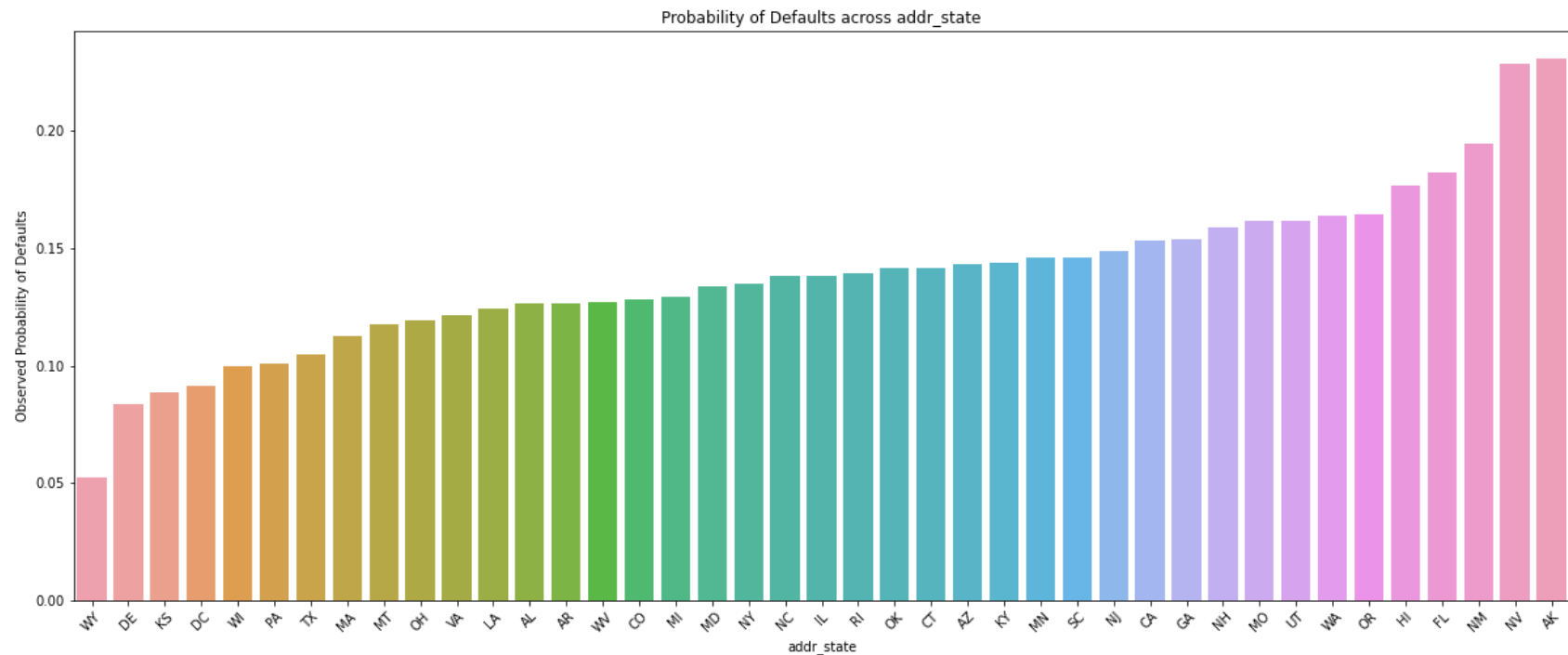
# Conclusion 4

- “**sub\_grade**” is a strong variable of default
- Ex: **more than 40% of loans** in “F5” & “G5” end up **defaulting**
- This helps the company to manage their portfolio by **not allocating a high number of loans in such high risk “sub\_grade” categories**



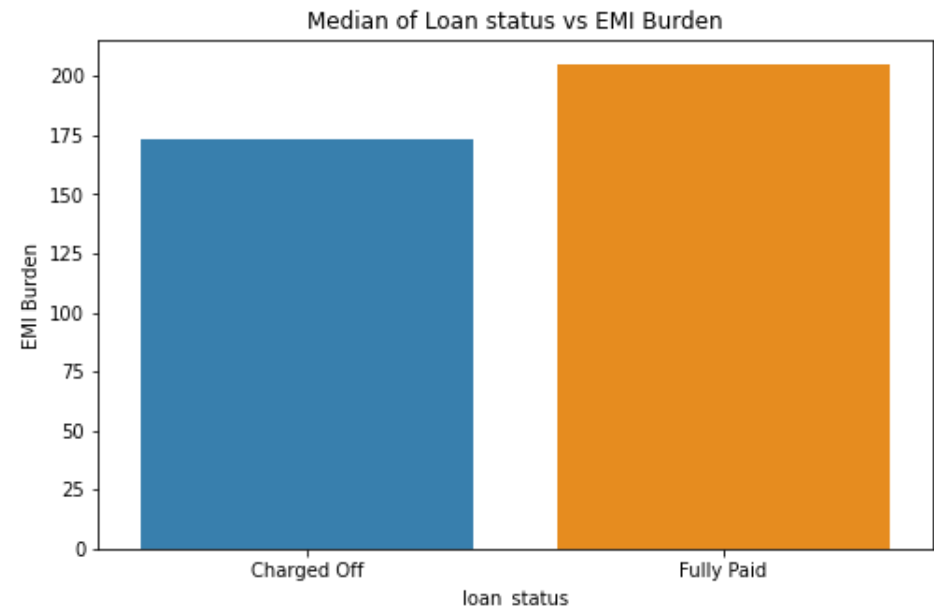
# Conclusion 5

- “**addr\_state**” is a strong variable of default
- Ex: **more than 20% of loans** in “NV” & “AK” end up **defaulting**
- This helps the company to manage their portfolio by **not allocating a high number of loans in such high risk “addr\_state” categories**



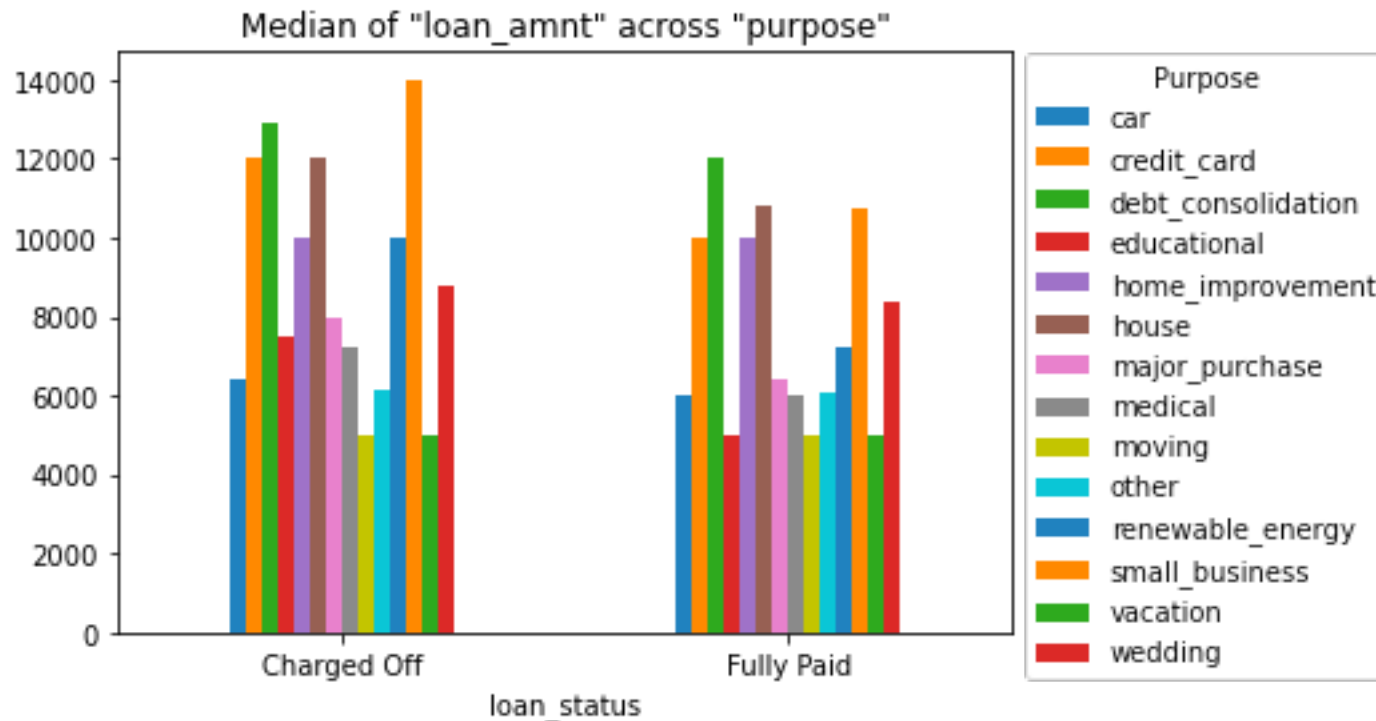
# Conclusion 6

- **Business Driven Metric – EMI\_BURDEN**
- Defined as the ratio of “annual\_inc” to “installment”. It tries to quantify the ability to handle the burden of installments.
- The medians of the calculated emi burden are “175” & “200” for “Charged Off” & “Fully Paid” respectively.
- Fully paid loans **have much higher emi\_burden** values than that of defaults
- This derived metric helps in predicting defaults at the time of approval
- **Note:** The metric must be used with caution. An increase/decrease in “annual\_inc” during the term can greatly affect its importance



# Conclusion 7

- Defaulters have significantly higher “**loan\_amnt**” for “purpose” – “small\_business”, “credit\_card”, “renewable\_energy” categories
- This helps the company to manage their portfolio by **not approving loans of higher “loan\_amnt” among such categories**





# Conclusion 8

- Among loans in grade “G”, “defaulters” have **significantly lesser “annual\_inc”** compared to the ones who have “Fully Paid”
- This potentially means that the **company must consider higher values of “annual\_inc”** for approval of loans among applicants of grade “G”

