Lending Club Case Study

SUBMISSION

Team: Sukanya Roy Sikdar ; Uday Kalakala

Lending Club Case Study

Business Objectives:

- Identification of the largest source of risky applicants
- Identification of risky applicants
- Identifying the driving factors behind loan default, i.e. the variables which are strong indicators of default (Top 5 variables)

Problem solving methodology

Data understanding - We observed the data and drilled down which variables we can run the analysis through to find the top 5 variables which would be significant indicators of default.

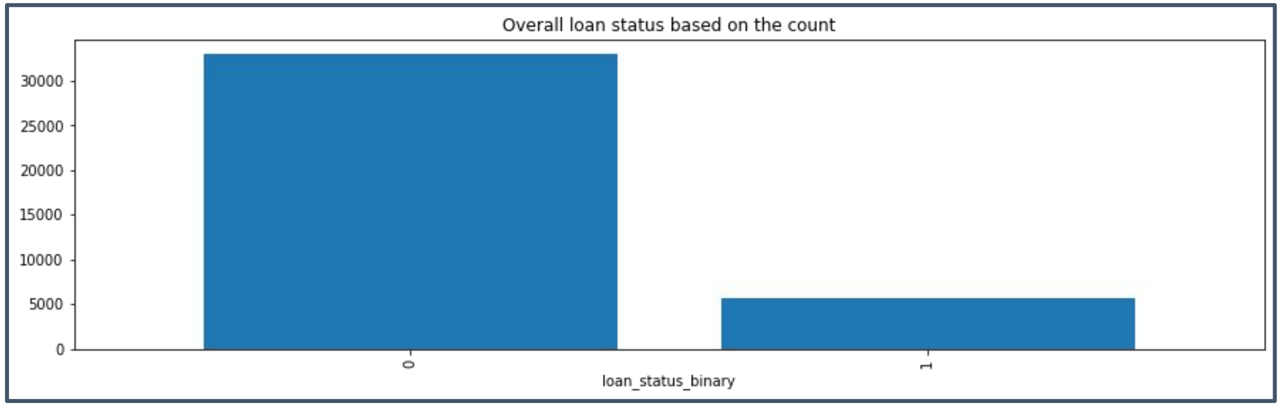
Data Cleaning and Manipulation - We identified columns with missing values and removed them from the analysis. We also removed "Current" from loan status since the payment is still ongoing and we cannot determine whether it is a loan that is going to be fully paid or charged off. We have also replaced few null values with the mode / mean wherever applicable and prepared the data for analysis. Considered "Fully Paid" as 0 and "Charged Off" as 1

Data analysis - Once we cleaned the data. We began with our data analysis, using the EDA techniques, which involved univariate, bivariate analysis across the variables also correlation between the variables. We have used charts along the way to get a clear picture on different variables and identify the most significant indicators of default.

Data Cleaning

- We identified columns with missing values and removed them from the analysis.
- We also removed "Current" from loan status since the payment is still ongoing and we cannot determine whether it is a loan that is going to be fully paid or charged off.
- We have also removed columns which are customer behavioural variables as they are not available while the loan application.
- We have also replaced few null values with the mode / mean wherever applicable and prepared the data for analysis.
- **Please note : Considered "Fully Paid" as 0 and "Charged Off" as 1**

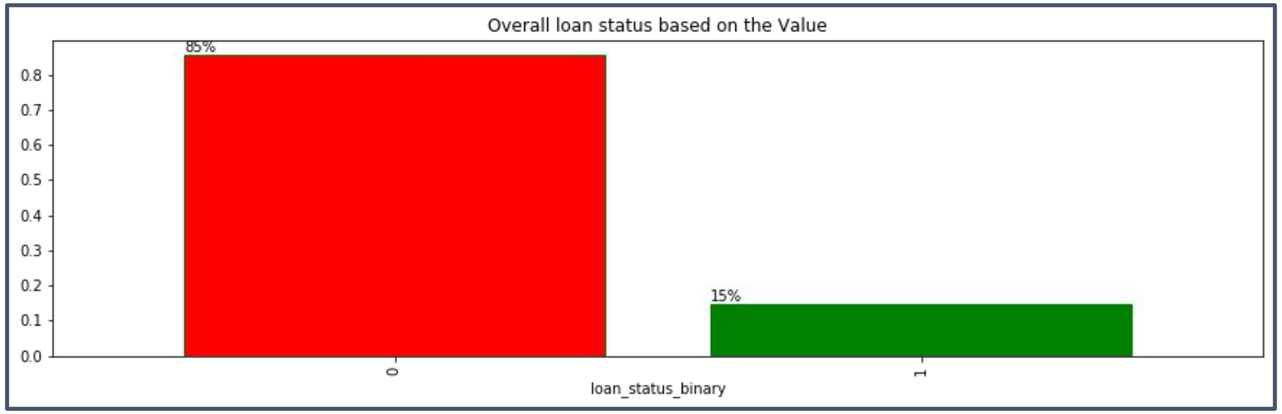
Data Analysis Observations



We assumed key variable as Loan Status

We see that among the loans provided over the online marketplace by the consumer finance company, most of its loan has been Fully Paid.

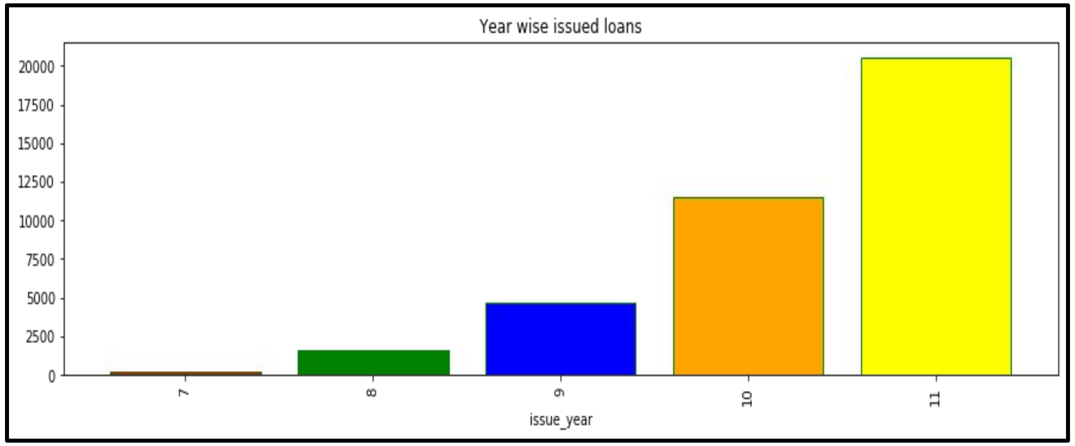
Overall Default rate is 15 %



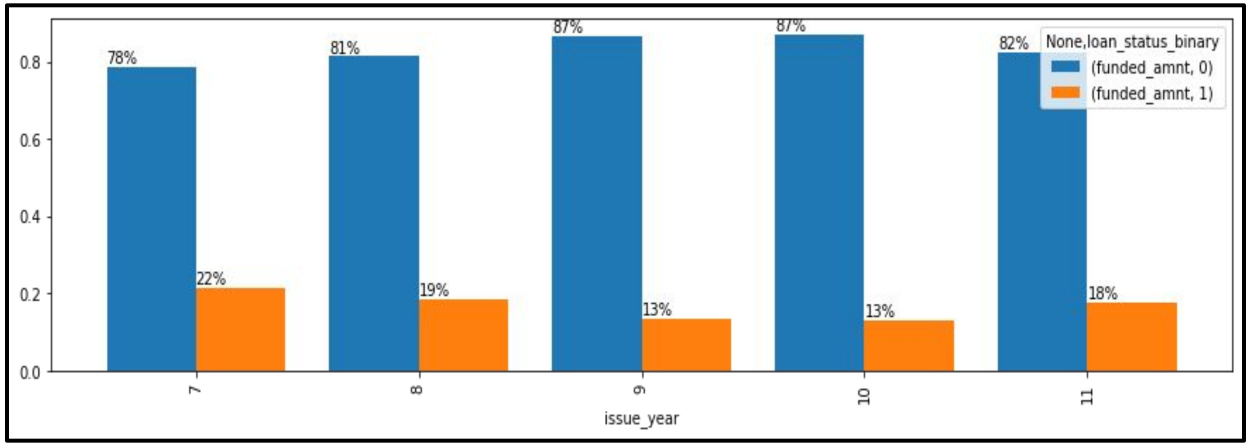
Data Analysis Observations

	loan_amnt	funded_amnt	funded_amnt_inv
loan_amnt	1.000000	0.981790	0.939091
funded_amnt	0.981790	1.000000	0.957145
funded_amnt_inv	0.939091	0.957145	1.000000

Based on the correlation we can assume that applicants are getting almost the same amount they have requested for, i.e (loan amnt ~ funded amnt by investors)



Year wise default rate



We see that 2011 had the highest number of loan issued, and 2007 had the highest default rate (22 %) followed by 2008 (19%). We assume this was because of the Great Economic Recession during those years.

Driving factors / variables behind loan default

1. Term

Fig 1.1

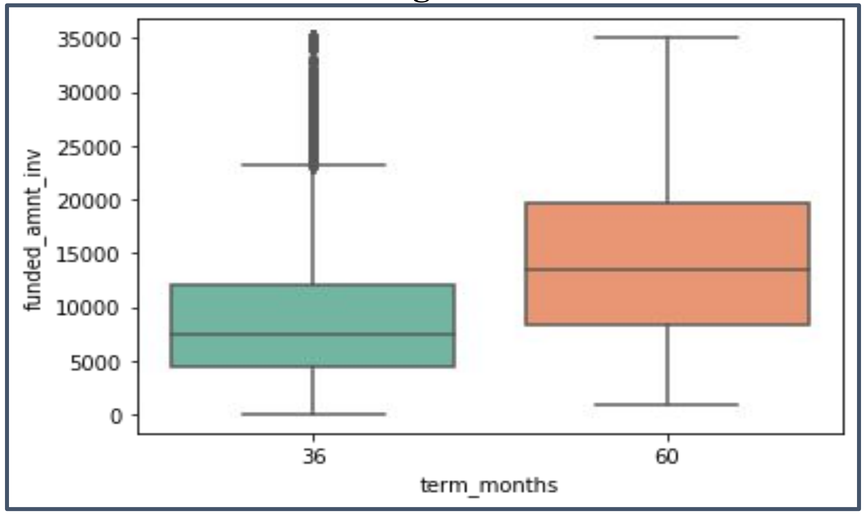


Fig 1.2 Terms wise default rate

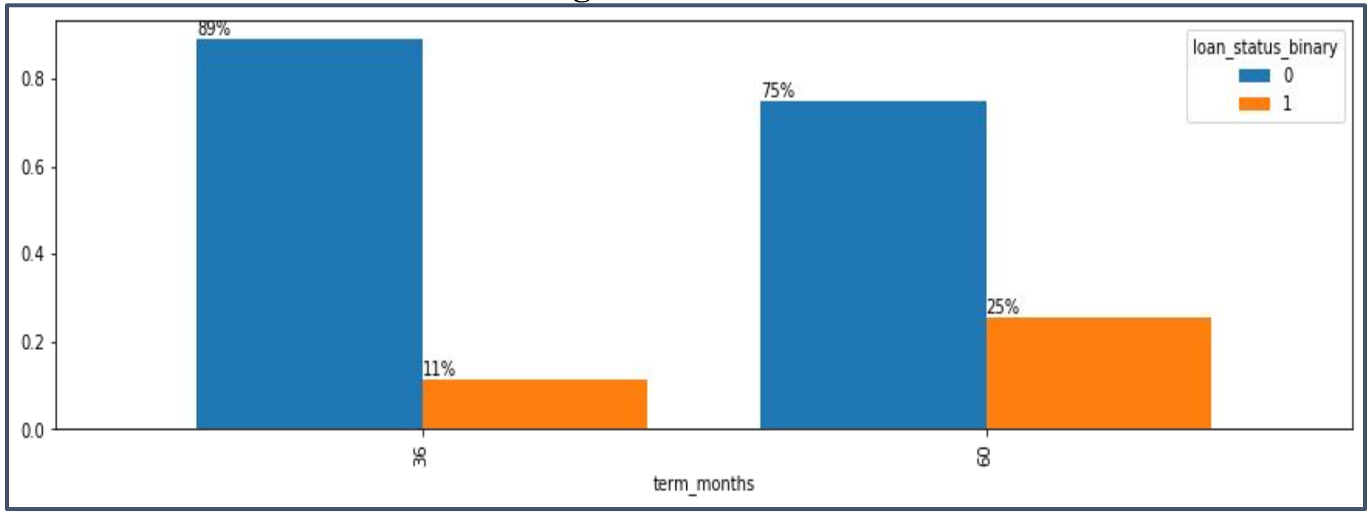
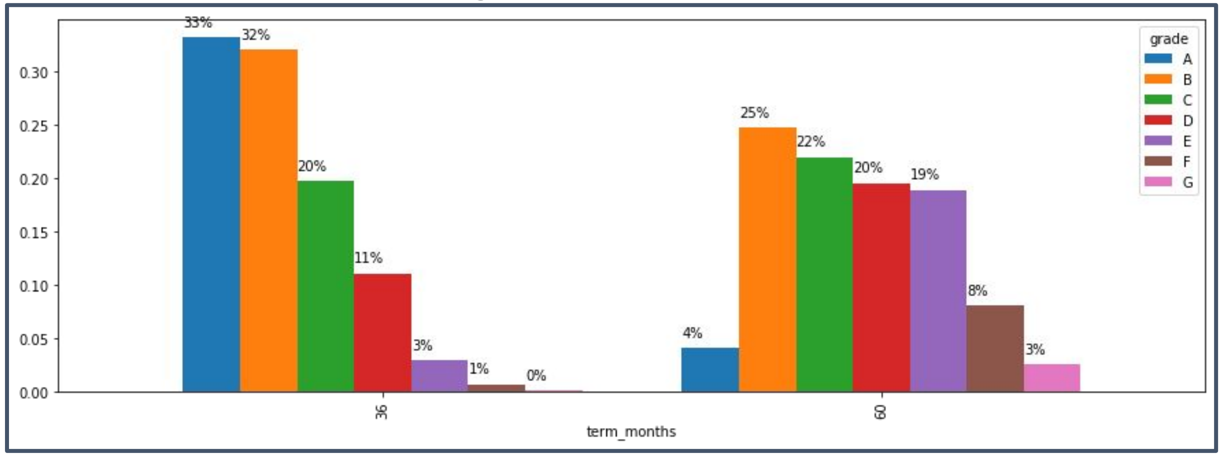


Fig 1.3 Term across Grade



- **60 months** term loans have the highest funding amount with majority of its loans lying between 10000 and 15000 USD (Fig 1.1)
- **60 months** term loan also has a higher **default rate of 25%** (Fig 1.2)
- The distribution of lower quality loans (based on grades) is more in case of **60 months** term loans (Fig 1.3)

Driving factors / variables behind loan default

2. Grade

Fig 2.1

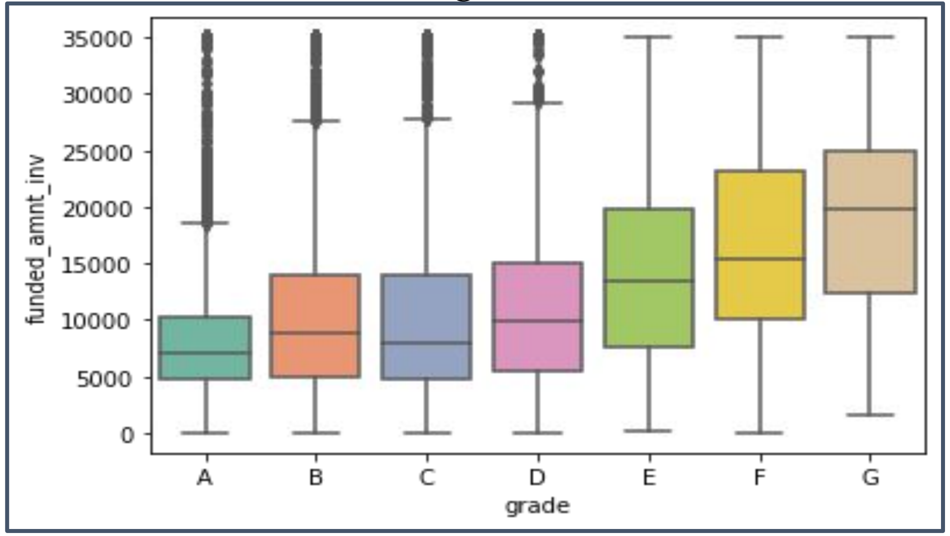


Fig 2.2 Interest rate across Grade

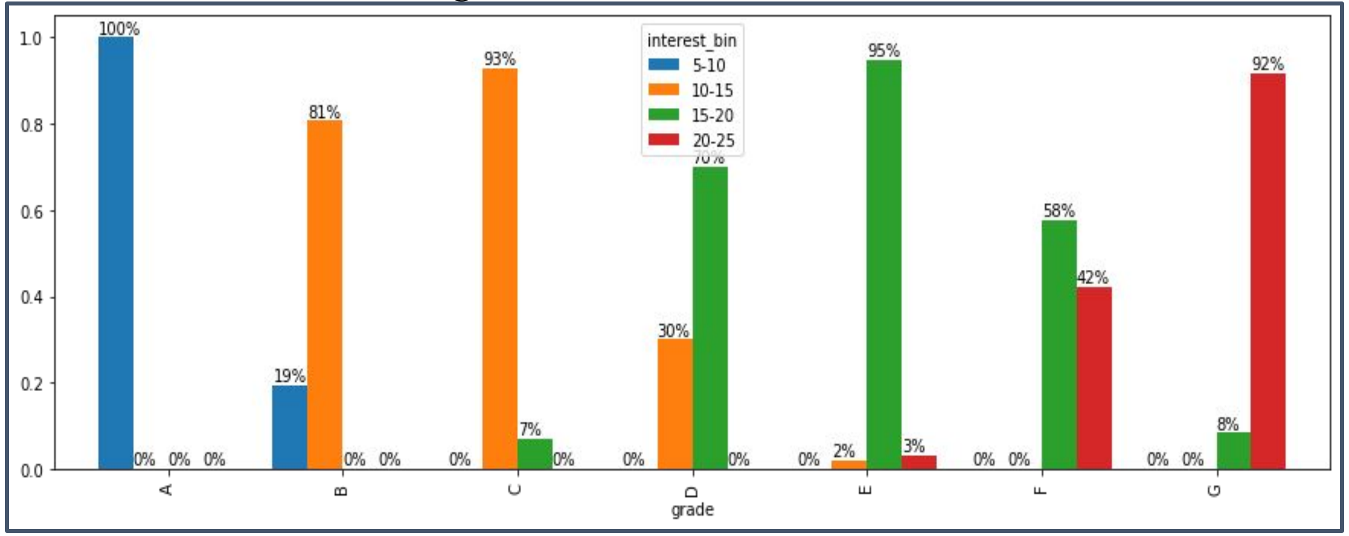
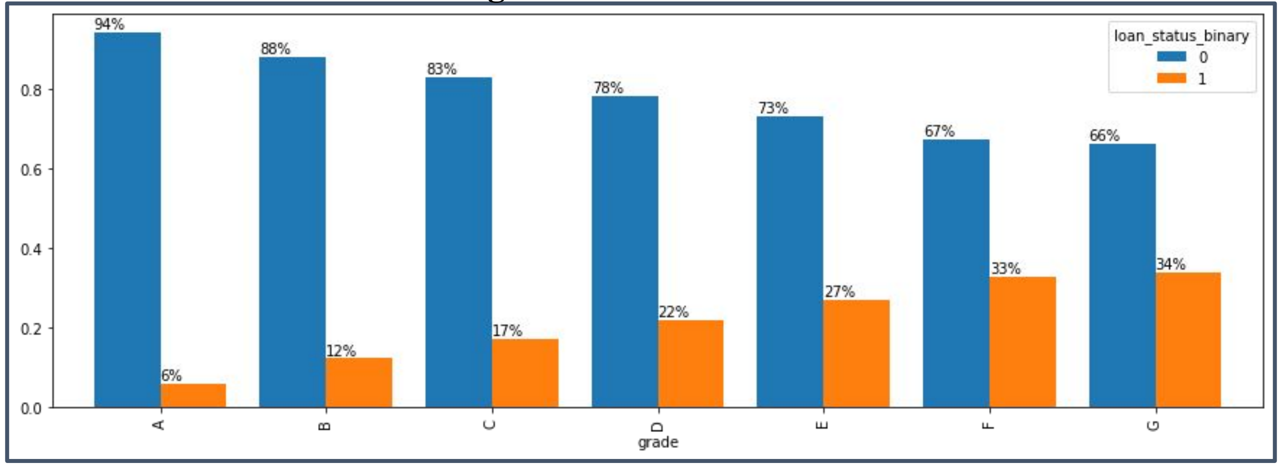


Fig 2.3 Grade wise default rate

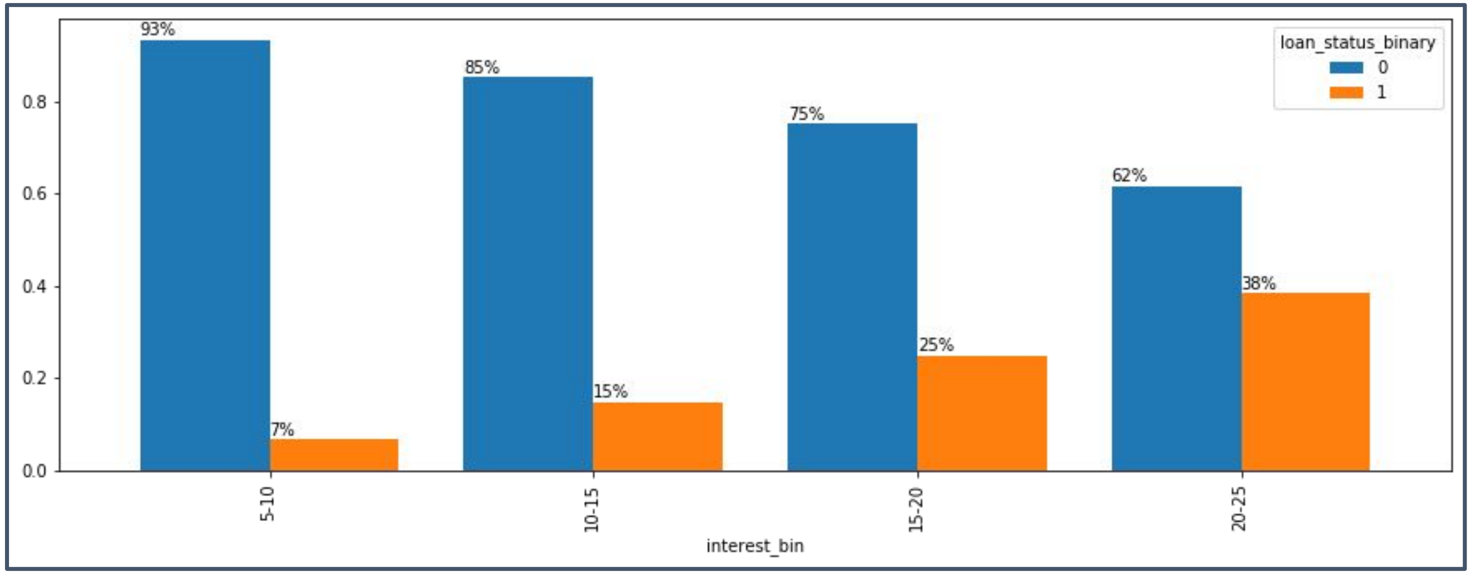


- The lowest quality of loan i.e **Grade G** has the higher funded amount (Fig 2.1)
- We also observed that **Grade G** has the highest **default rate of 34%** (Fig 2.3)
- As the quality / grade of loan moves from A to G (high to low), the rate of interest increases and hence chances of default also increases (Fig 2.2)
- Subgrade also has a similar effect on default rate (check Appendix**)

Driving factors / variables behind loan default

3. Interest rate

Interest rate bin wise default rate



- We have noticed that as the quality of loan reduces, the rate of interest increases
- This typically showcases the fact that the riskiest loans have the highest amount of interest rate
- Hence, from the chart we can see that higher the interest rate, higher is the default rate.
- At **20-25% rate of interest** the **default rate** is highest i.e **38%**

Driving factors / variables behind loan default

4. Purpose

Fig 4.1 Purpose wise default rate (by number of applicants)

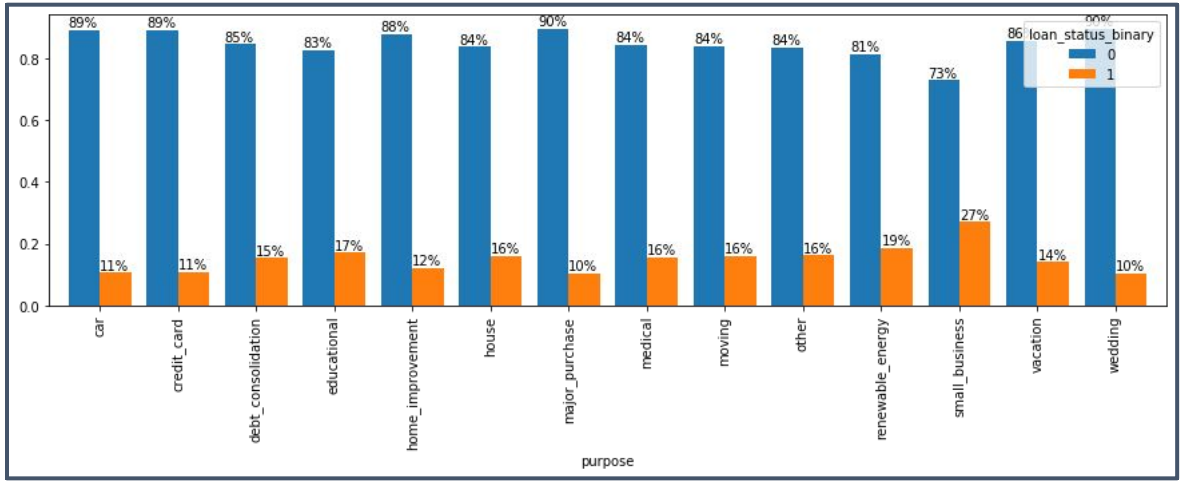


Fig 4.2 Purpose wise default rate (by total funded amount)

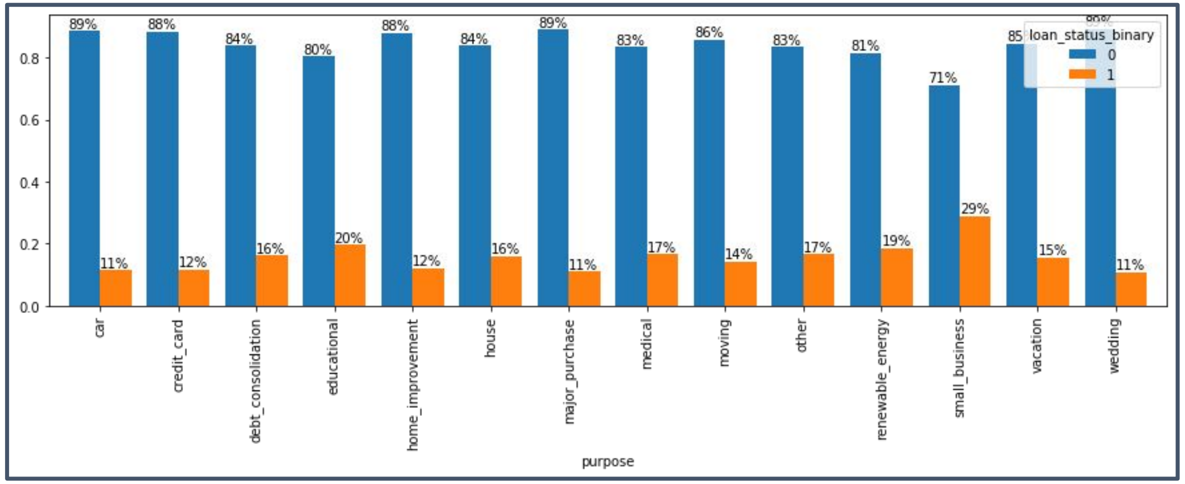
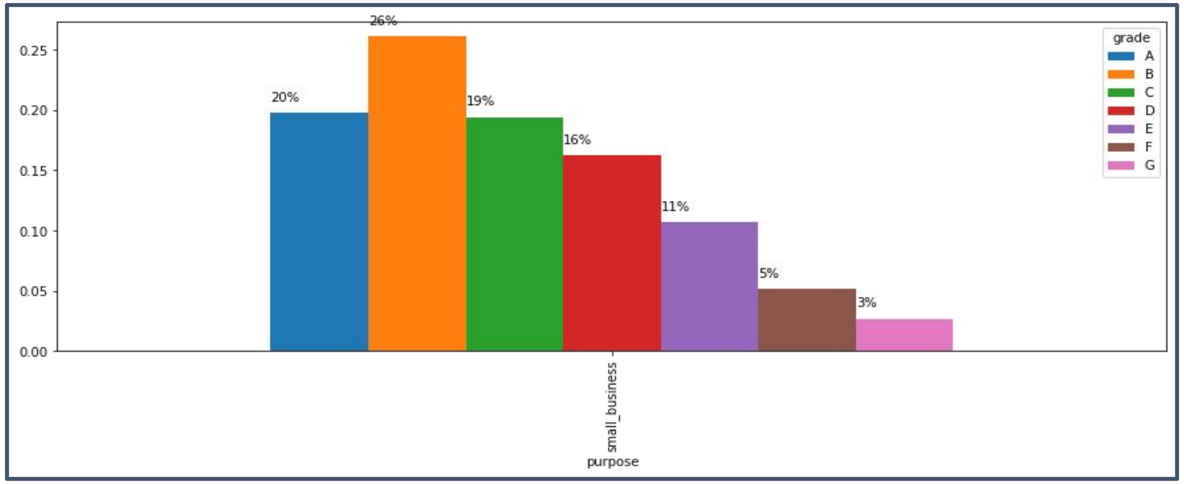


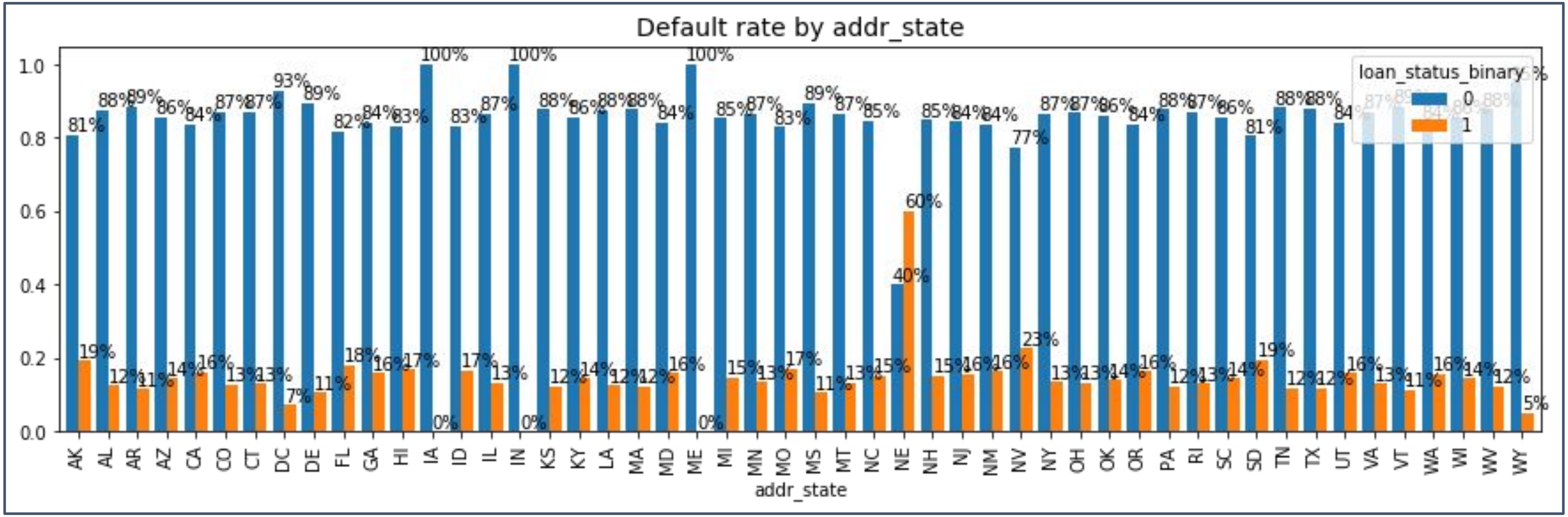
Fig 4.3 Purpose across grade (For Small Business)



- For purpose **Small business** has the highest default rate of **27%** (based on applicants) and **29% (**based on the total funded amount)
- Small business have loans skewed mostly towards Grade B and A, however there is significant amount of small business loans towards the lower grades. Hence, there is a higher default rate (Fig 4.3)
- Renewable energy also has a higher default rate (19%), followed by educational (17%) based on the number of applicants. (Fig 4.1)
- Educational also has a higher default rate (19%), followed by Renewable energy(17%) based on the total funded amount. (Fig 4.2)

Driving factors / variables behind loan default

5. Address State



The state with the highest default rate is **NE** at **60%**, which is higher than it's rate of fully paid. NE is the only state where the charged off borrowers are more than fully paid. NE is followed by NV with 23% of default rate.

Conclusions

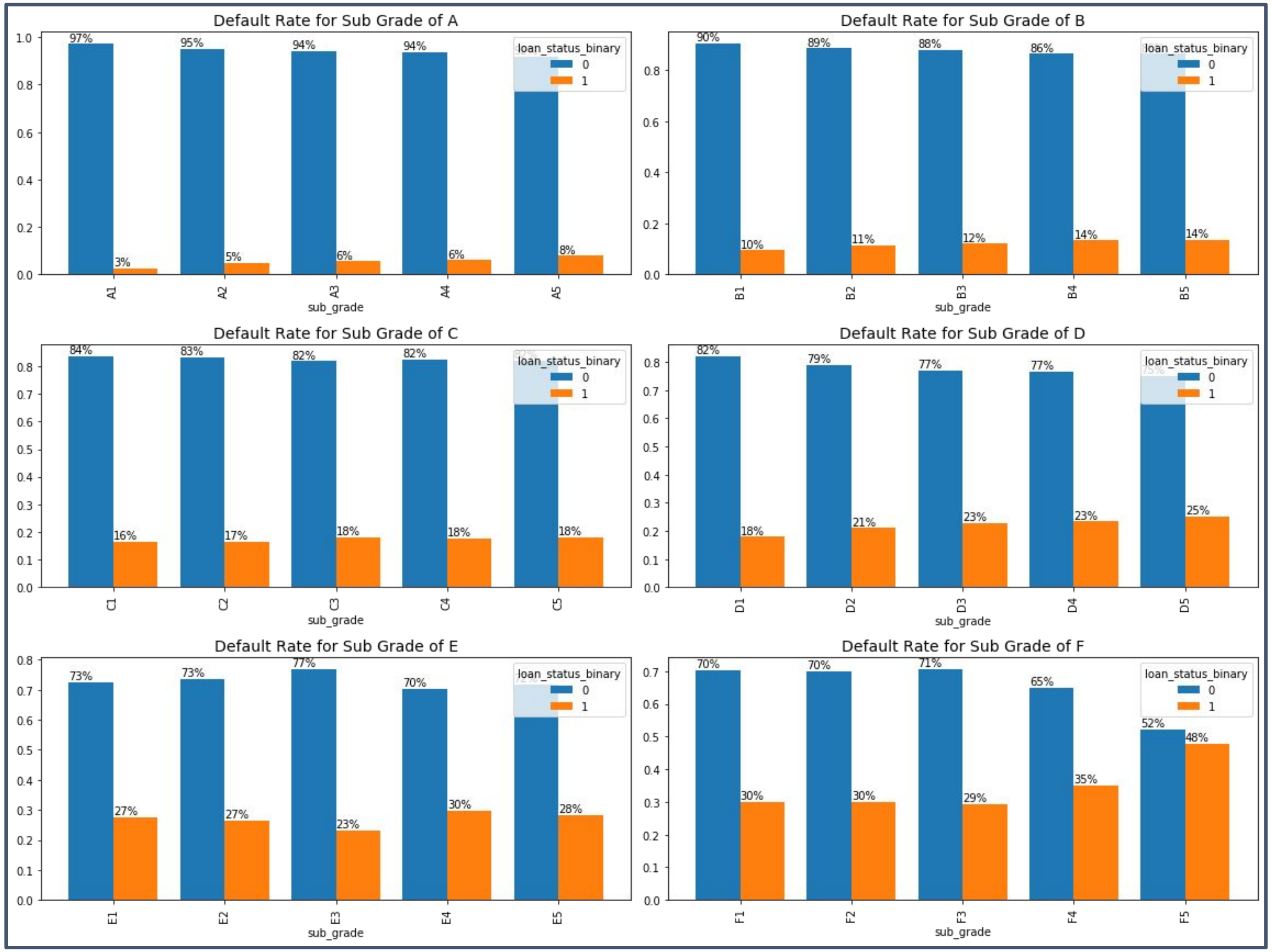
❖ Based on our observations we have identified **Term, Grade, Interest Rate, Purpose and Location (Address State)** as the major driving factors for defaulting.

❖ Additional Mentions -

- **Home ownership -** Other with default rate of 18%
- **Emp_Title**
  - ➢ LEVEL 3 COMMUNICATIONS with default rate of 80% (based on default rate)
  - ➢ US ARMY with default rate of 20% (based on the number of loans issued)
- **Annual Income -** USD 0-25000 bin with default rate of 19%, i.e, lower the income higher chances of default
- **Verification Status -** It's surprising that not verified has less default rate (13%) and verified as high default rate (17%)
- **DTI -** Default rate is high in the 20-30 bin of DTI (17%). Most of the application for charged off and fully paid is distributed between 10-20
- **Pub_rec_bankruptcies -** The applicant with higher public bankruptcies record is more likely to default (14%)

Appendix

2. Grade (**Sub-grade)



As the sub grade increase in each grade there is a increase in the default rate in all grades except in G