upGrad | International Institute of Information Technology Bangalore

Name of the case study: Lending Club Case Study

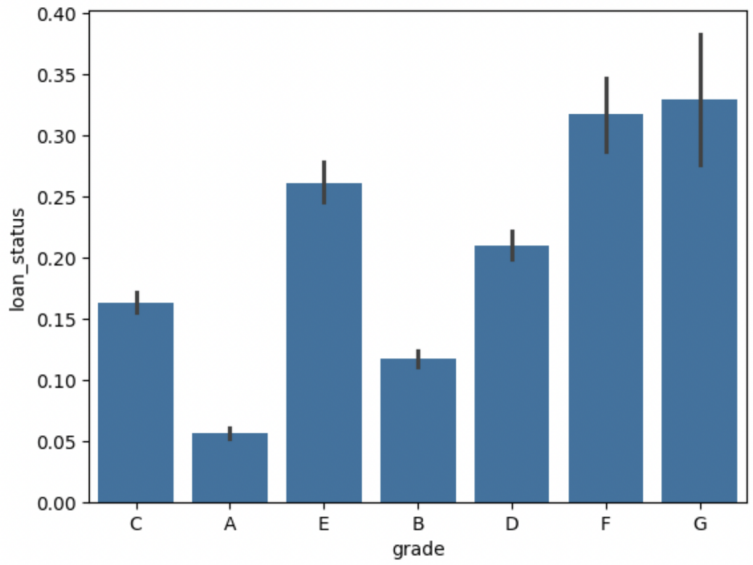
Assignment Brief: Dive in the data analytics of a financial company who is the largest online loan marketplace facilitating business loans, personal loans and financing of medical procedures. Objective is to perform data analytics to identify various contributing factors for a financial company to identify what are the attributes of consumer and loan which results in the default or losses for the company. To analyse the above scenario, we will perform exploratory data analysis and apply univariate & bivariate analysis and identify driving variables or factors:-

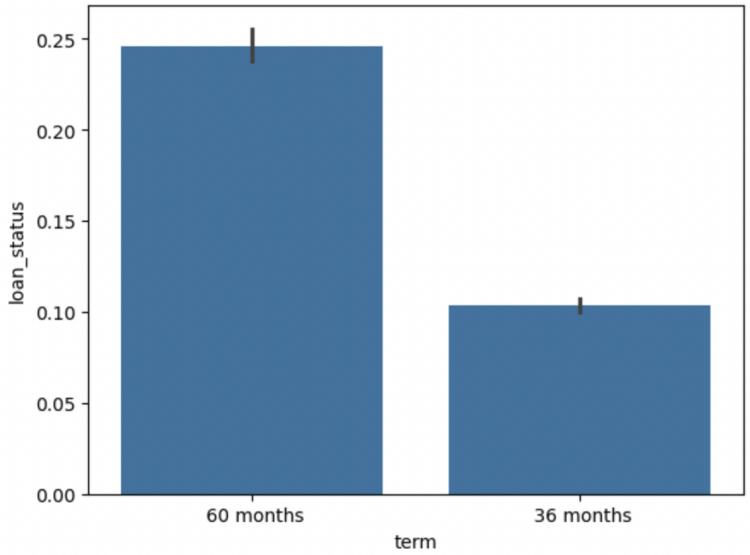
Data Cleaning:

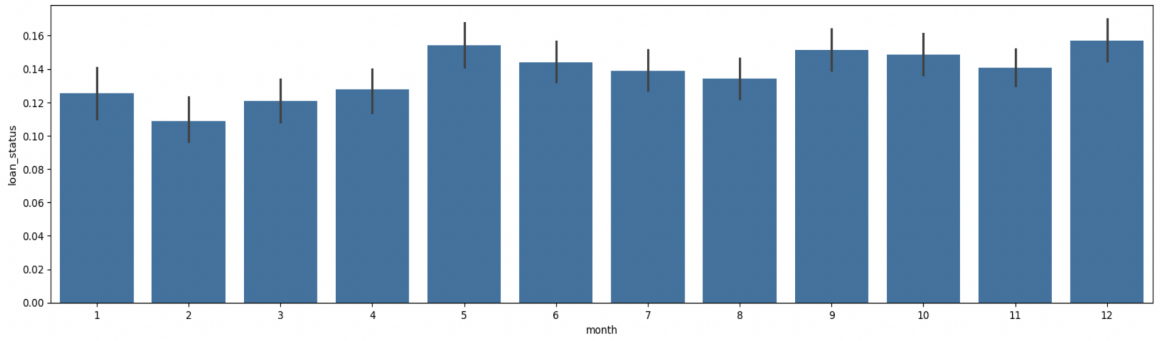
- Identify the empty rows for each column from the given data frame and drop the columns wherein the empty value is more than 50%>. This is followed across data industry standard as common practise and also instructed by the instructor in the previous classes
- Omitted the missing value rows for each column
- Identified columns are identified wherein, only one unique value is available hence, dropped those rows
- Few more columns were dropped to retain the relevant or needed data for the analysis like unique ids, customer behaviour variables

Univariate Analysis:

- Evaluated various variables and plotted in bar form
- As per data, most of them occur in the month of December
- To be precise, 13.98% loans are defaulted

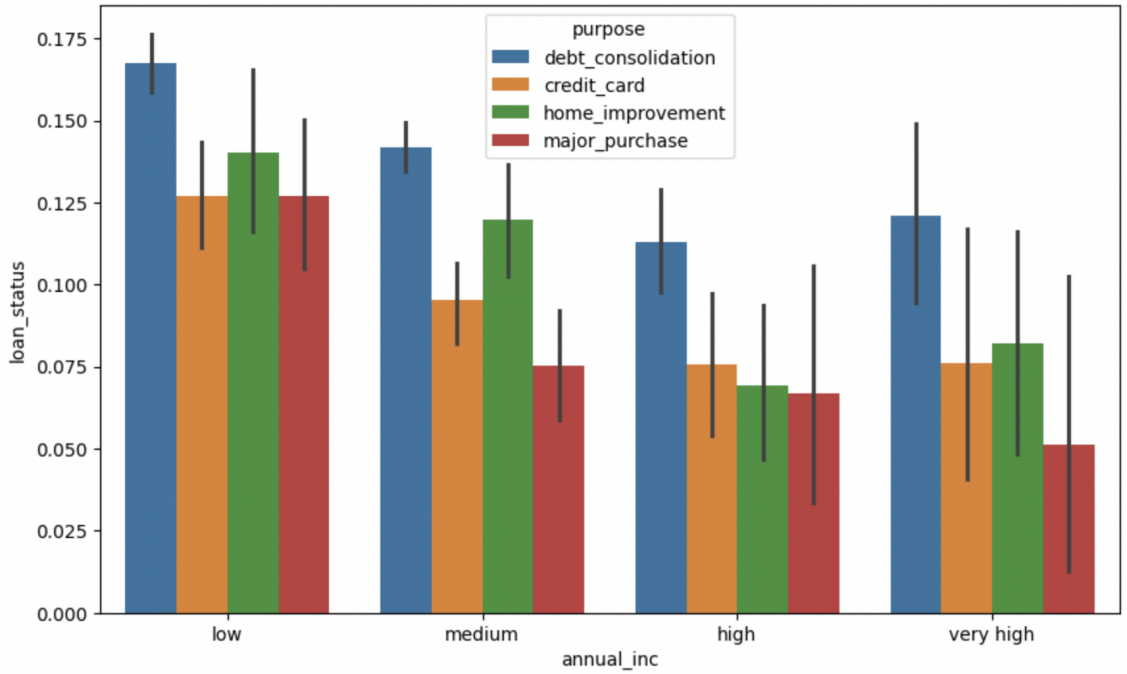


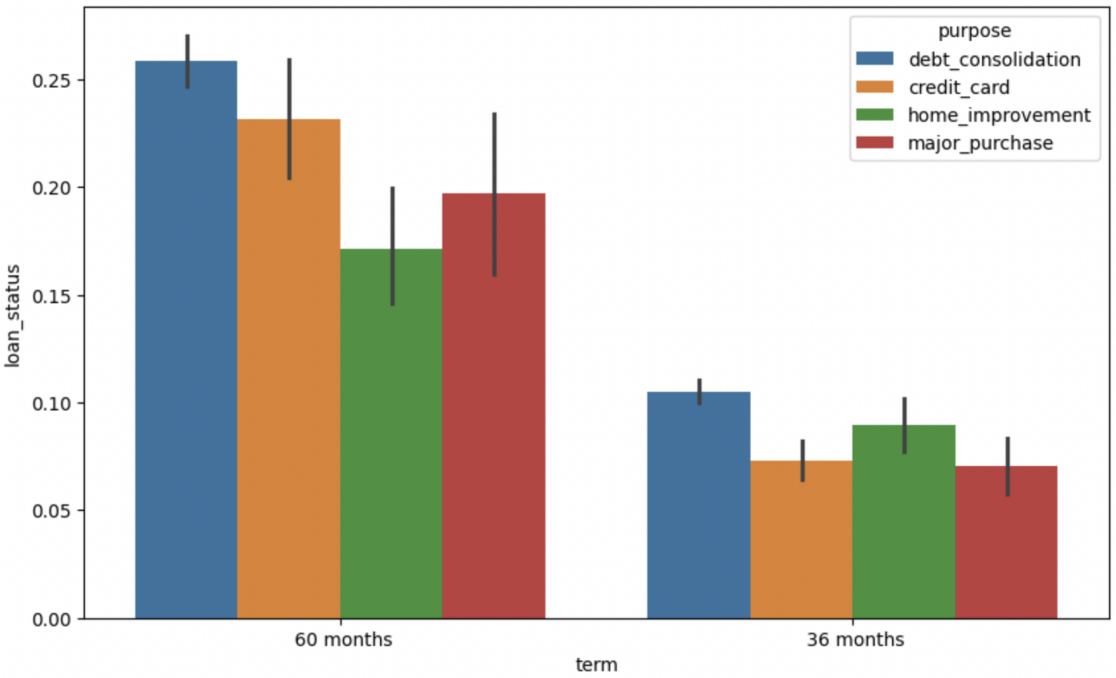


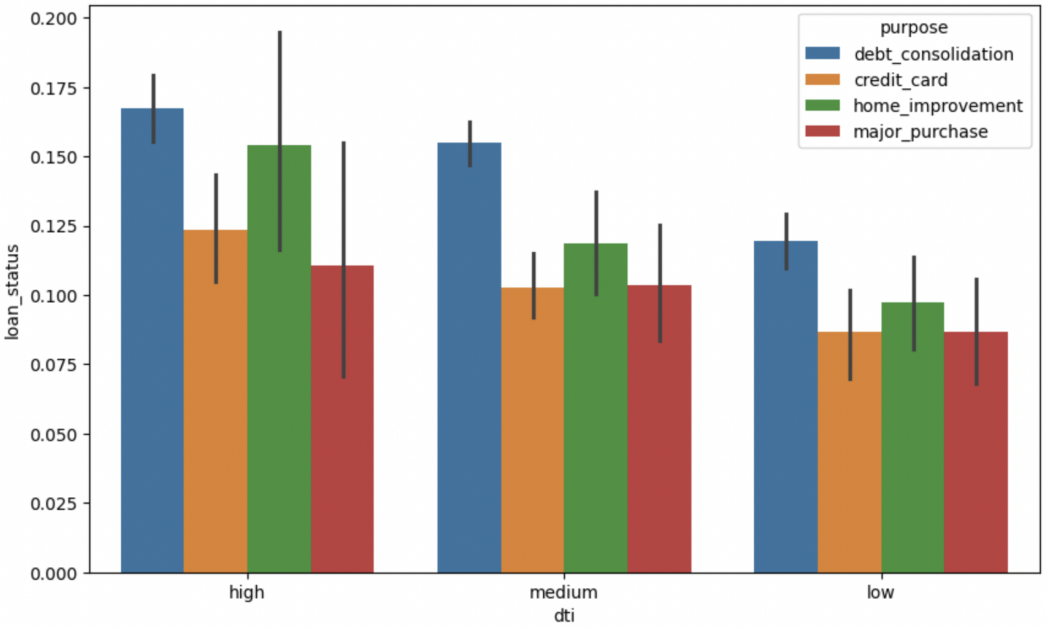


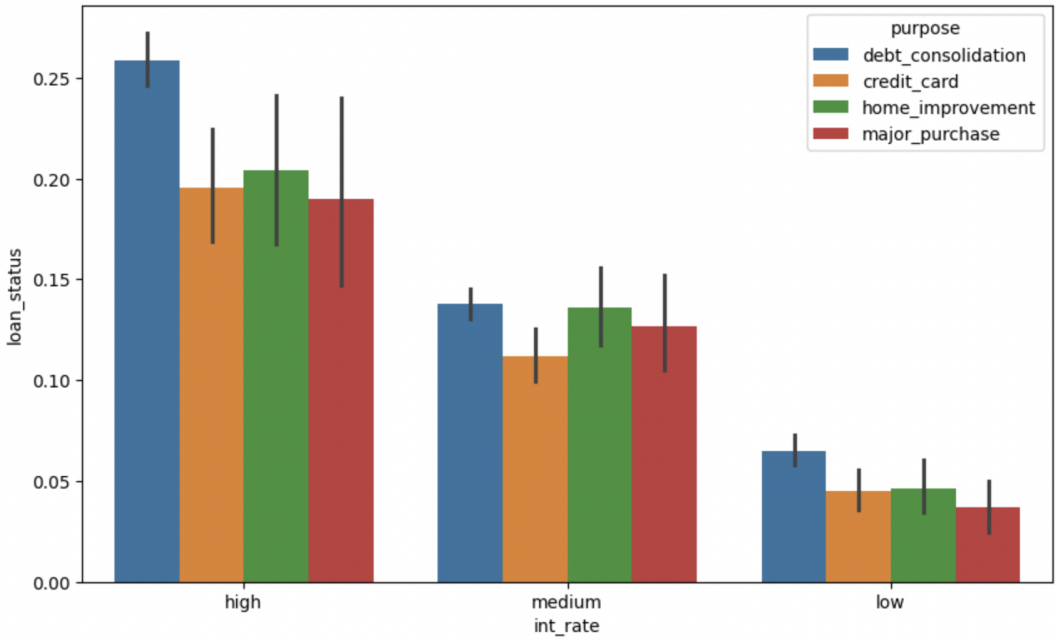
Bivariate Analysis:

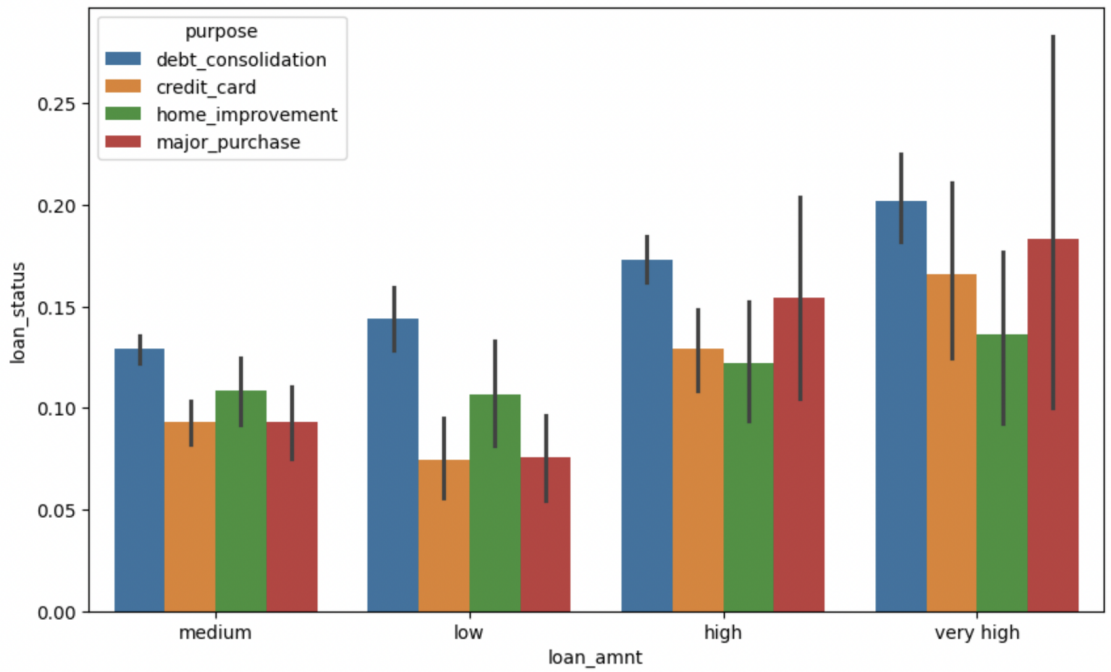
- As part of analysis, identified two variables are considered to see how defaulter behaviour changes
- Heatmap is used to analyse which US states have more defaulters in comparison
- Identifying default driving factors which have more influence on the defaulting behaviour of the borrower

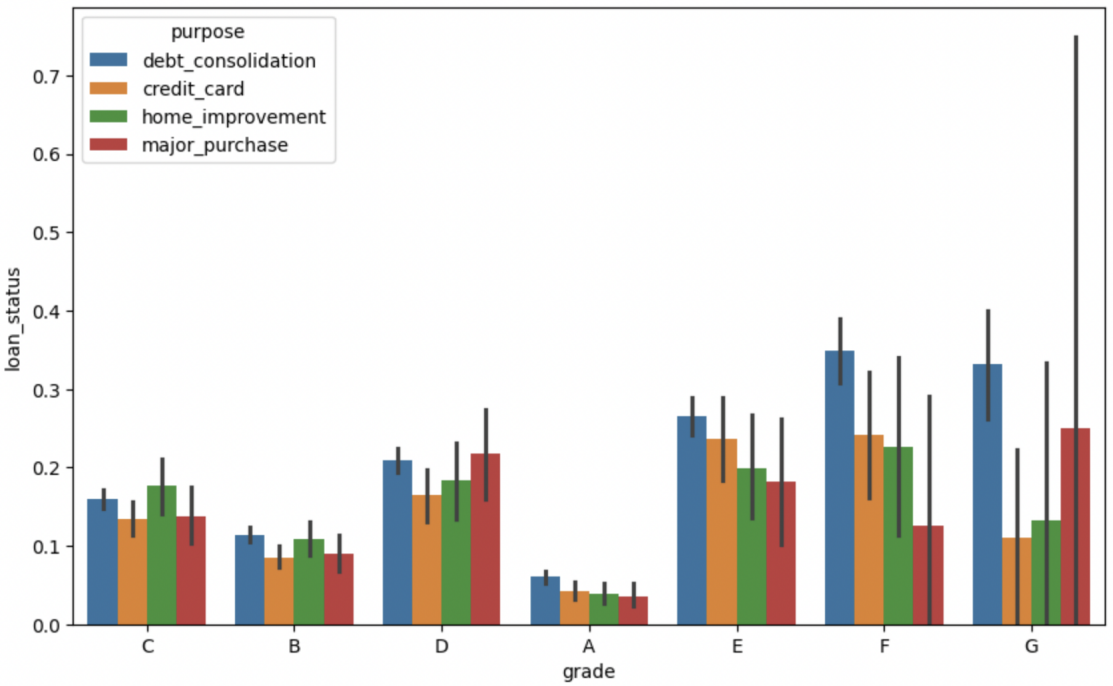

purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
low
medium
high
very high
annual_inc

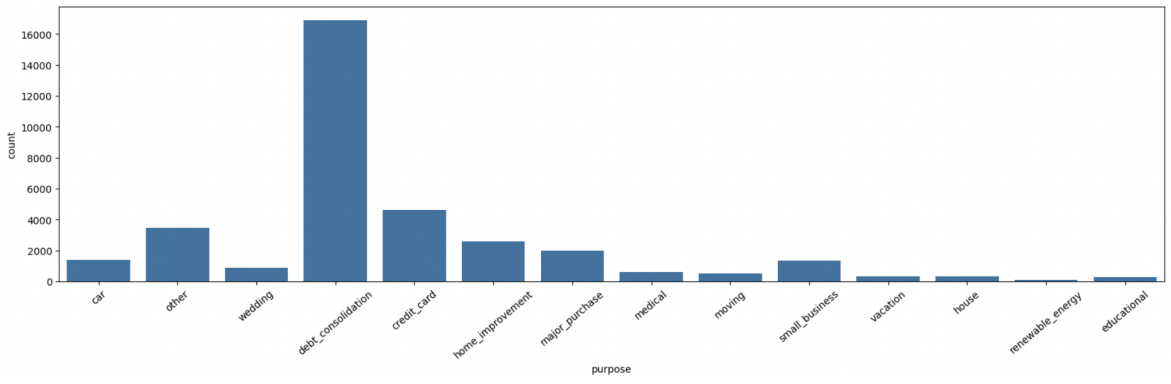

purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.25
0.20
0.15
0.10
0.05
0.00
60 months
36 months
term

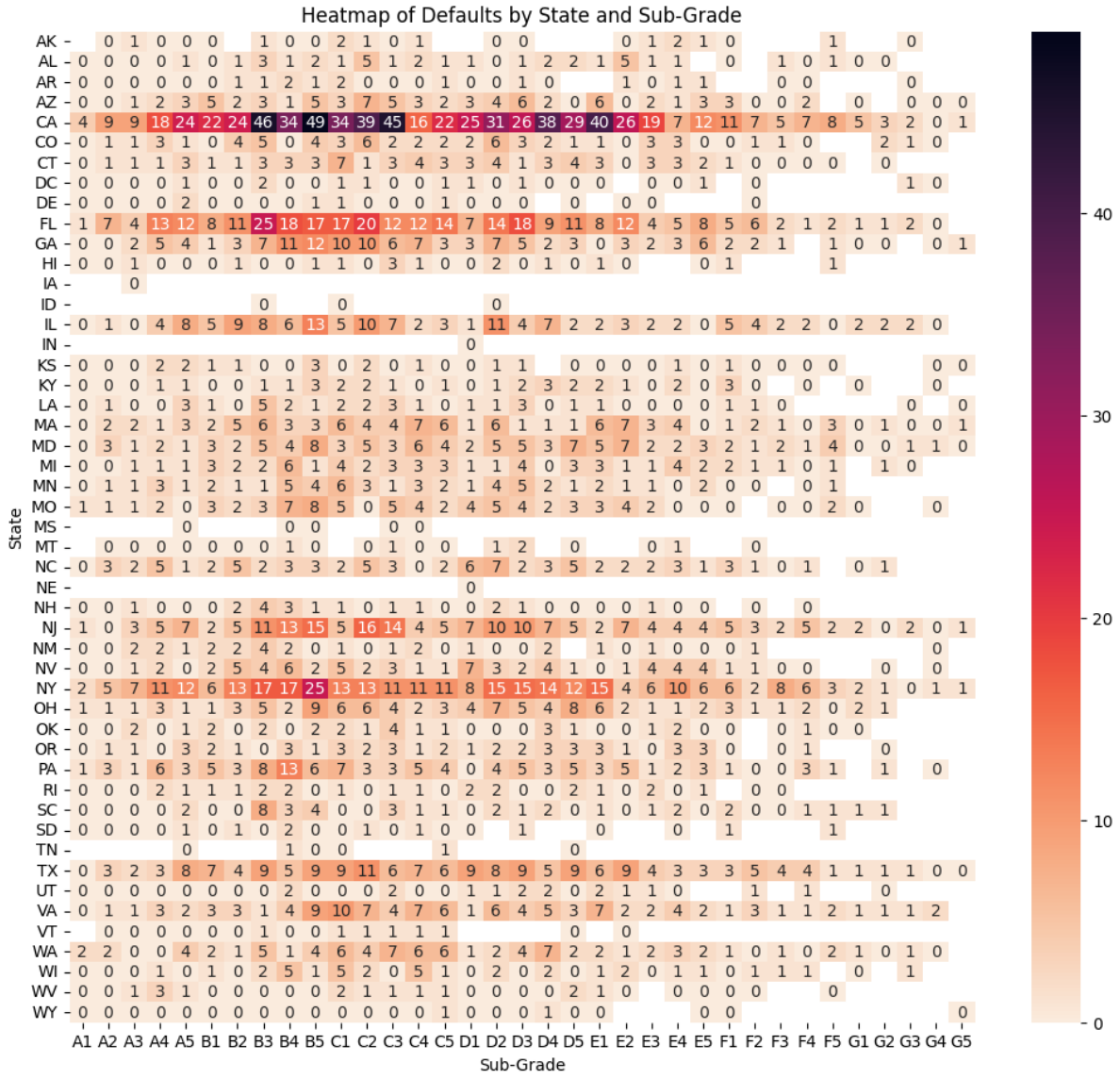

purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
high
medium
low
dti


purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.25
0.20
0.15
0.10
0.05
0.00
high
medium
low
int_rate


purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.25
0.20
0.15
0.10
0.05
0.00
medium
low
high
very high
loan_amnt


purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
C
B
D
A
E
F
G
grade


count
16000
14000
12000
10000
8000
6000
4000
2000
0
car
other
wedding
debt_consolidation
credit_card
home_improvement
major_purchase
medical
moving
small_business
vacation
house
renewable_energy
educational
purpose

Recommendations and conclusion as per analysis and Heatmap:

- Borrowers with previous history of default is bigger indicator of default or loss
- Avoid approving higher value loans with previous default or history of single default
- Small business loans have been regulated and reduced
- Interest rates to spiked for loans with dti greater than 20
- Enable eligibility criteria for loan approval to be <30% of the borrower's income
- As per heatmap, Location of the borrower definitely plays important role in driving default behaviour
- Also, default driving factors include 'sub_grade', 'grade', 'int_rate', and 'addr_state'