LENDING CLUB ASSIGNMENT

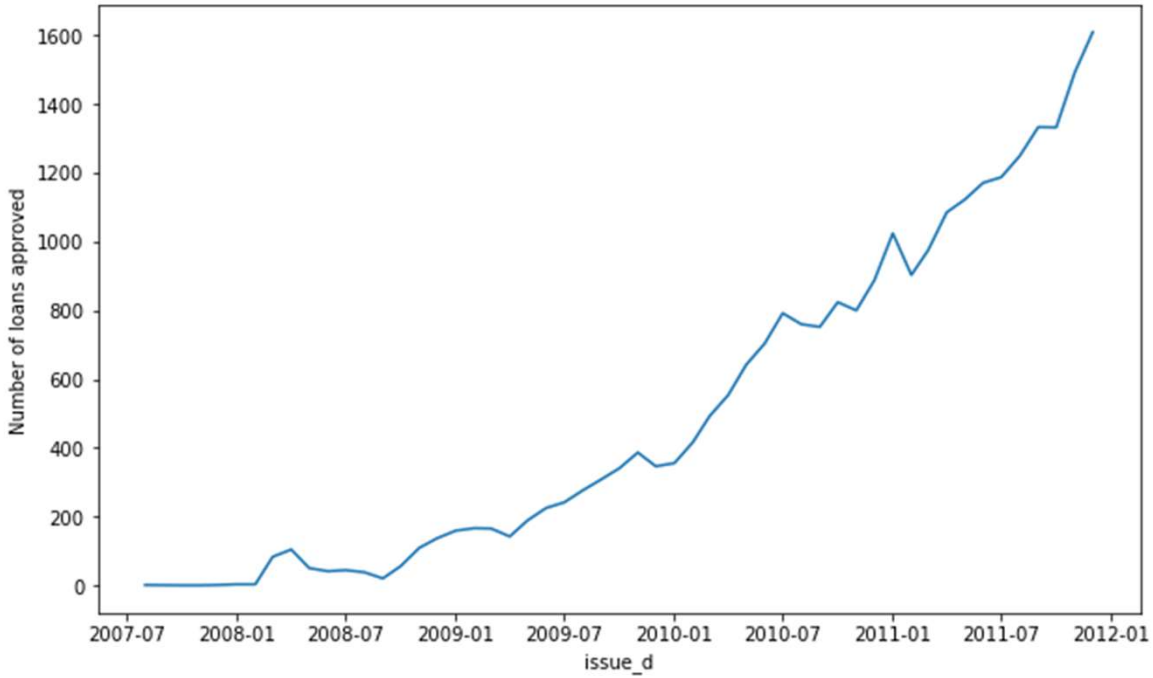
SUBMISSION

Name: Sahil Bajaj, Rajkumar Lakkimsetty

Abstract

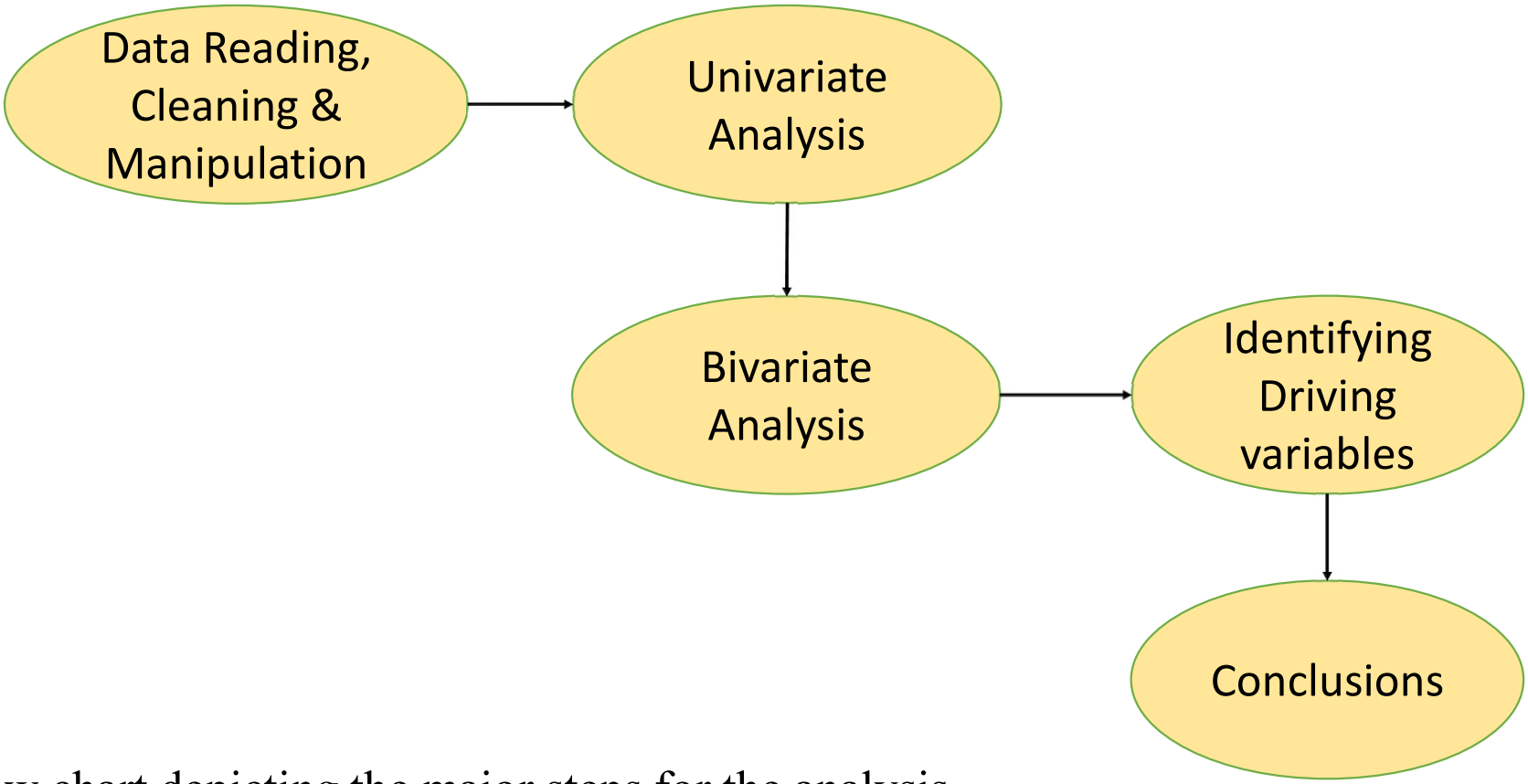
- Analytics has a crucial role in all major sectors today, and Banking sector is one of the sectors where the use of Data analysis can help in Investment, loan approval and other decisions.
- Our company ‘Lending Club’, has been growing at an impressive rate over the last few years, the number of loan approvals has grown considerably as well, but with an increase in loan approvals the risk associated Increases as well.
- For a sustainable growth which is beneficial for the borrower and the Investors a proper approach of risk assessment is quite essential.

Number of loan approvals over the years:



- The number of loan approvals has increased has approximately doubled every year.

Methodology:



The Flow chart depicting the major steps for the analysis.

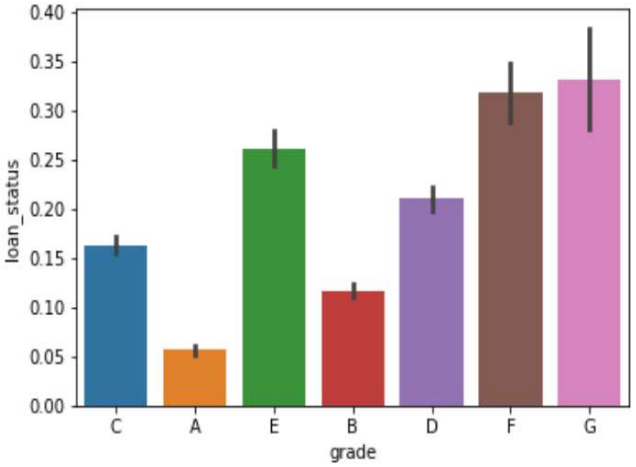
Analysis:

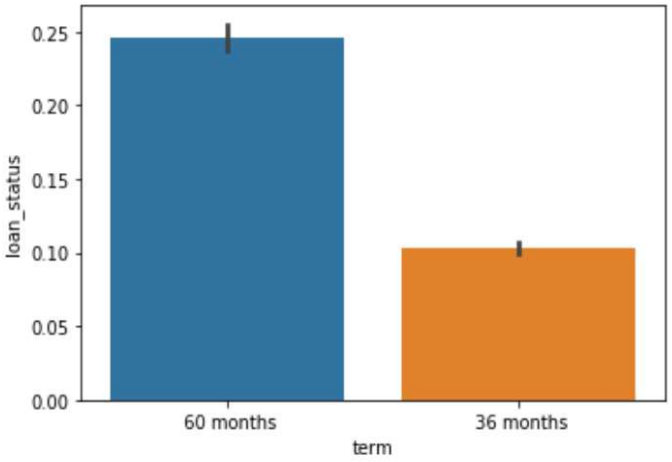
Data Reading, Cleaning & Manipulation:

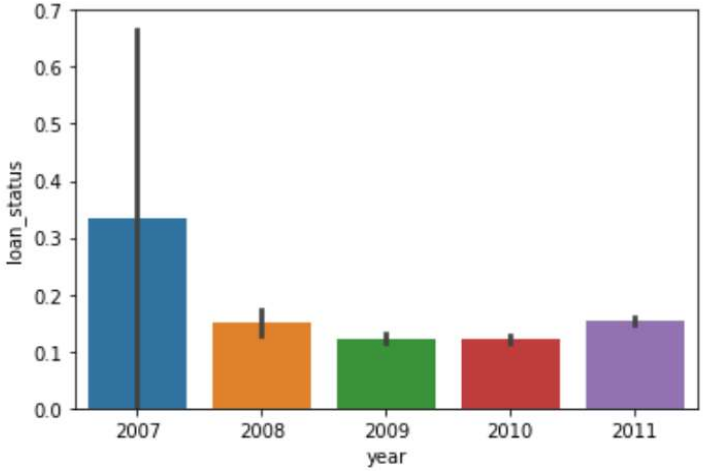
- First and foremost the number empty rows for each column of the given dataframe is calculated. After which the columns where empty value is more than 50% is dropped.
- The missing value rows are omitted for each column.
- Several redundant columns are identified where there is only one unique value, such rows are junk in nature and hence are dropped.
- Many unnecessary columns are dropped which play no crucial role in our analysis, e.g. Unique ids, Customer behaviour variables, etc.

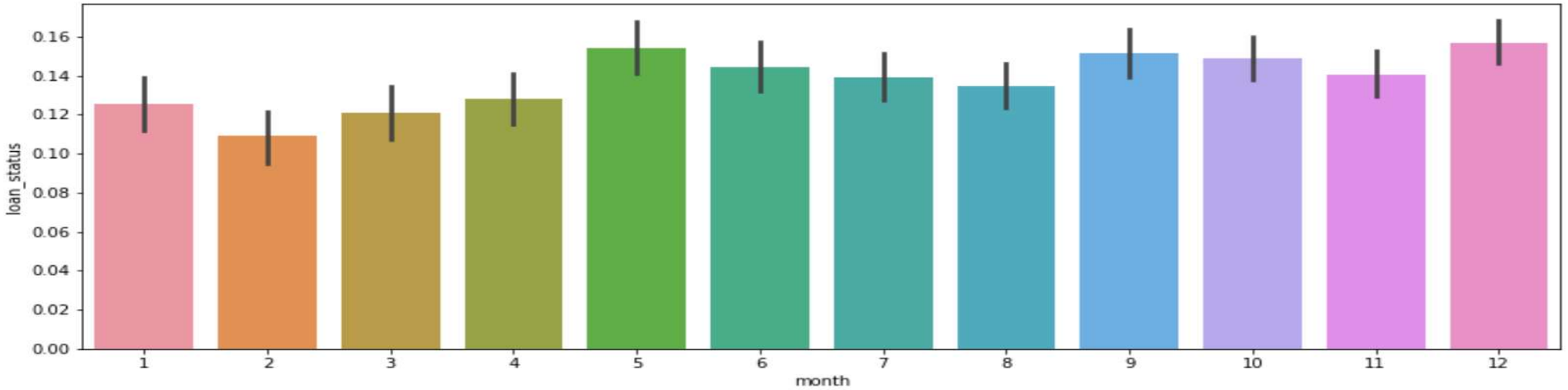
Univariate Analysis:

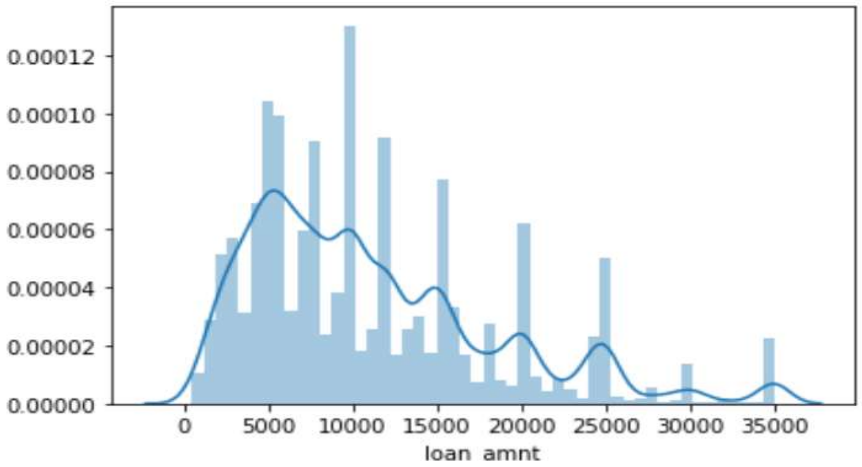
- Default rate for various variables is evaluated and plotted in bar-plot form.
- Most of the default occur in the month of December.
- Approximately, 14% of loans are defaulted

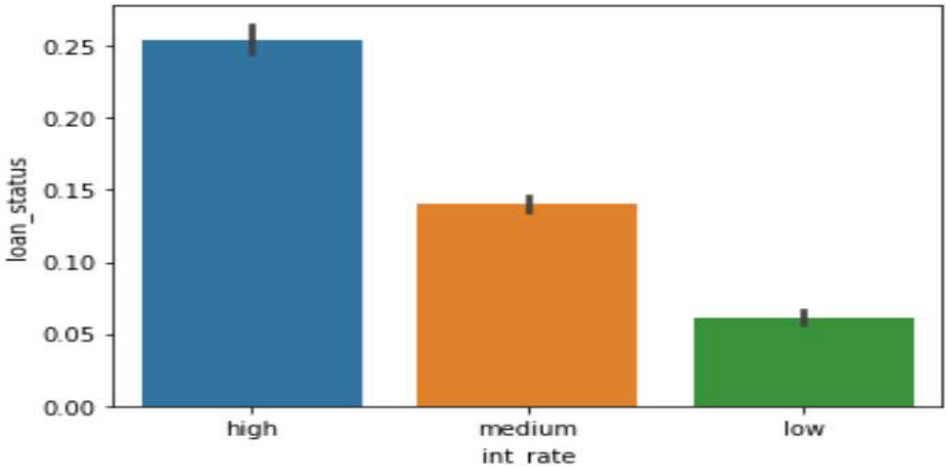

loan_status
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
C
A
E
B
D
F
G
grade

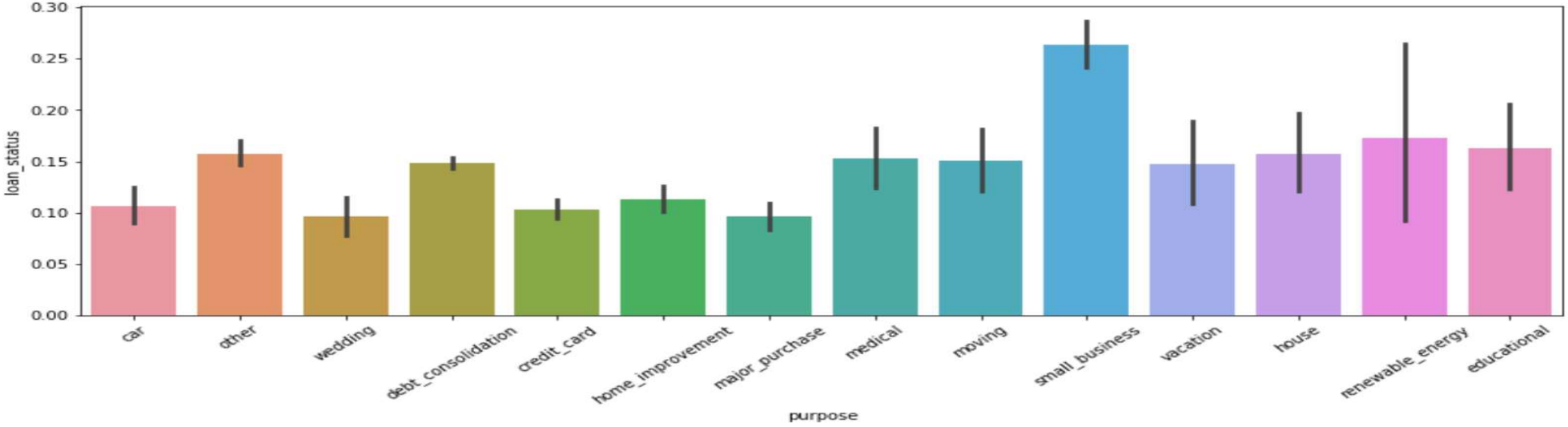

loan_status
0.25
0.20
0.15
0.10
0.05
0.00
60 months
36 months
term


loan_status
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
2007
2008
2009
2010
2011
year


loan_status
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1
2
3
4
5
6
7
8
9
10
11
12
month


0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
5000
10000
15000
20000
25000
30000
35000
loan_amnt


0.25
0.20
0.15
0.10
0.05
0.00
loan_status
high
medium
low
int_rate


0.30
0.25
0.20
0.15
0.10
0.05
0.00
loan_status
car
other
wedding
debt_consolidation
credit_card
home_improvement
major_purchase
medical
moving
small_business
vacation
house
renewable_energy
educational
purpose

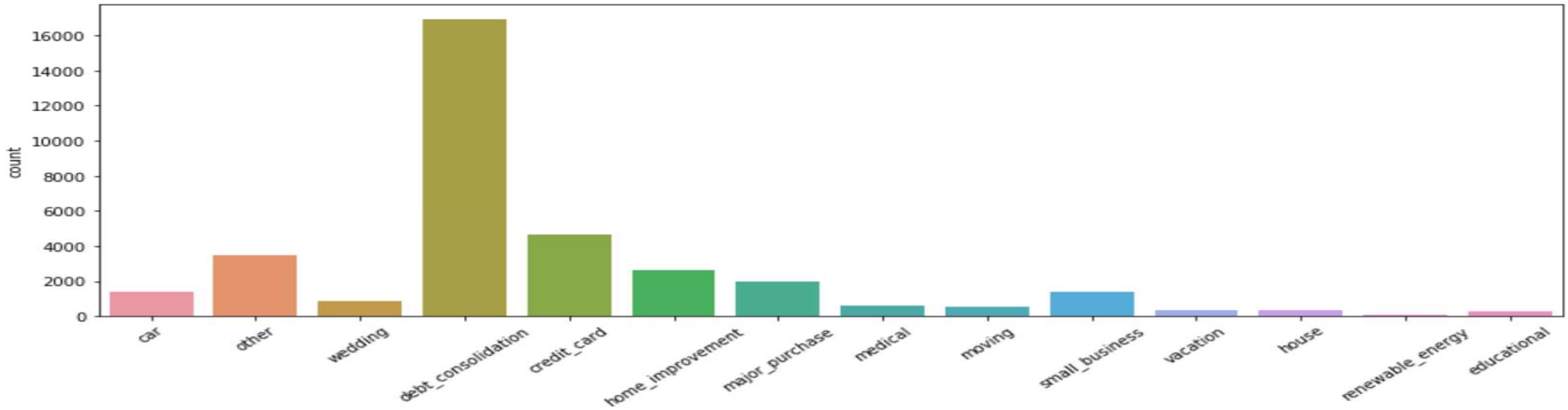
Analysis

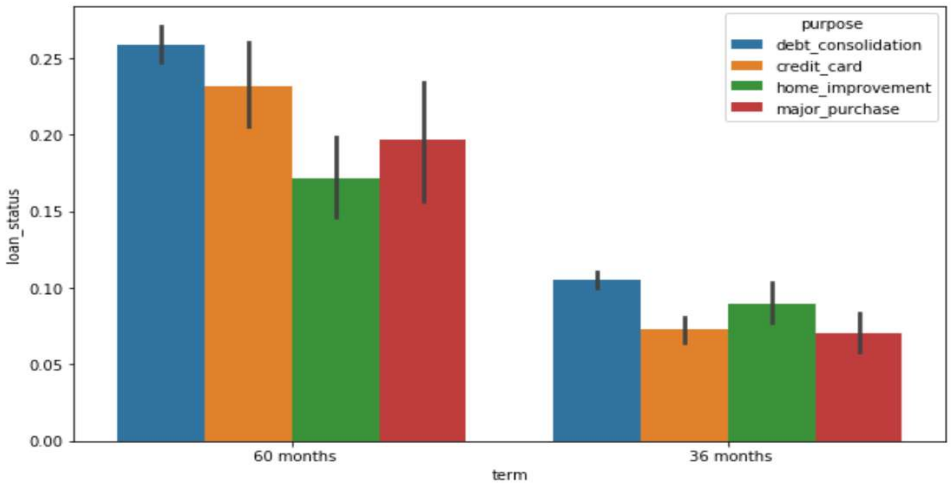
Bivariate Analysis:

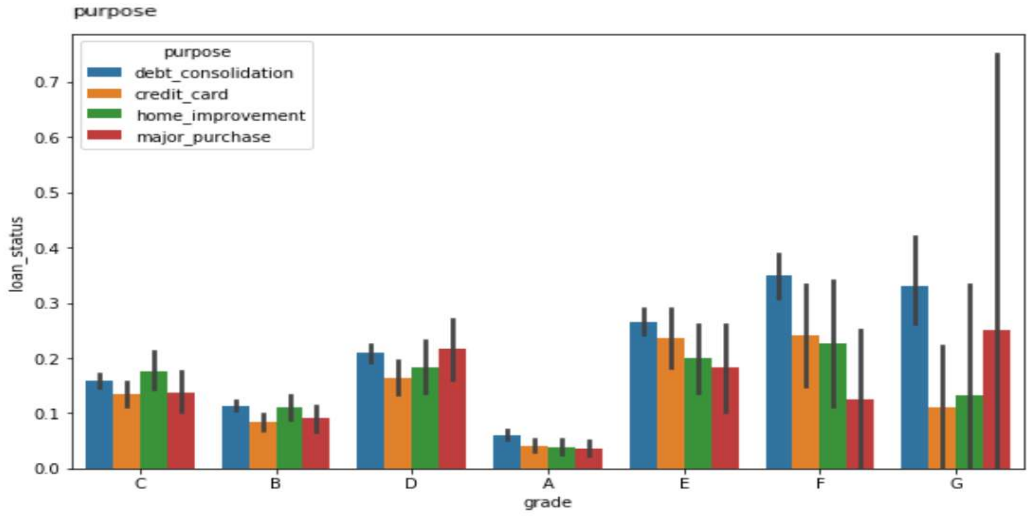
- For this part of the analysis two column variables are considered to see how the defaulter behaviour changes as the variables.
- Heatmap is used to analyse which US states have more defaulters in comparision.

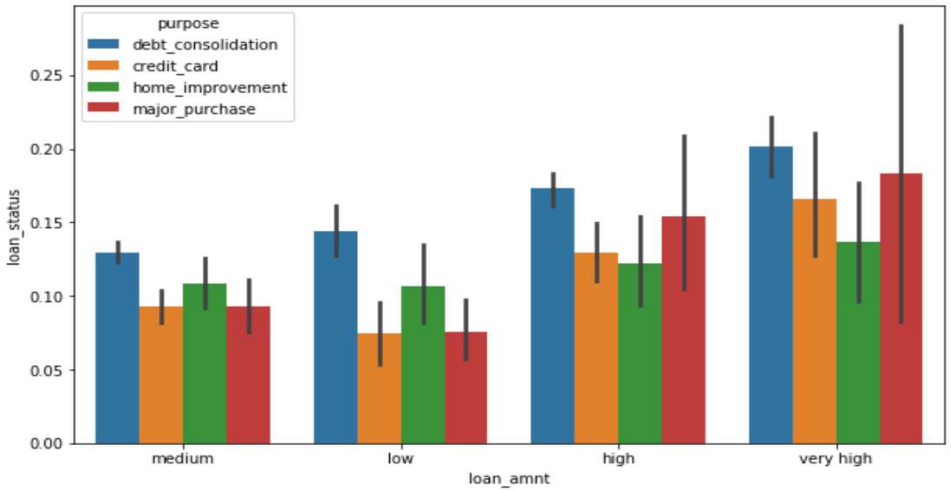
Identifying Driving variables:

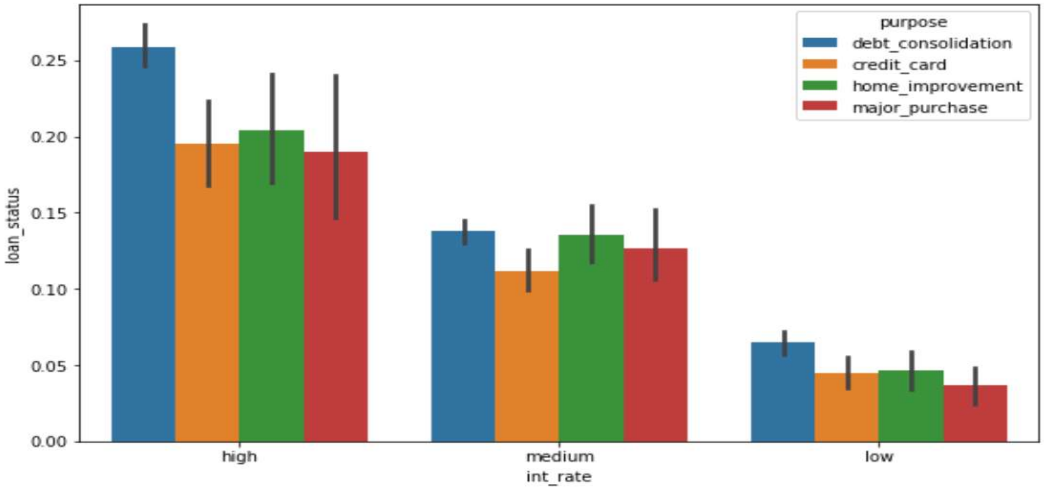
- Driving variables are those which have more influence on the defaulting behaviour of the borrowers applying for loan.
- Top five driving variables are Identified and listed out in conclusion

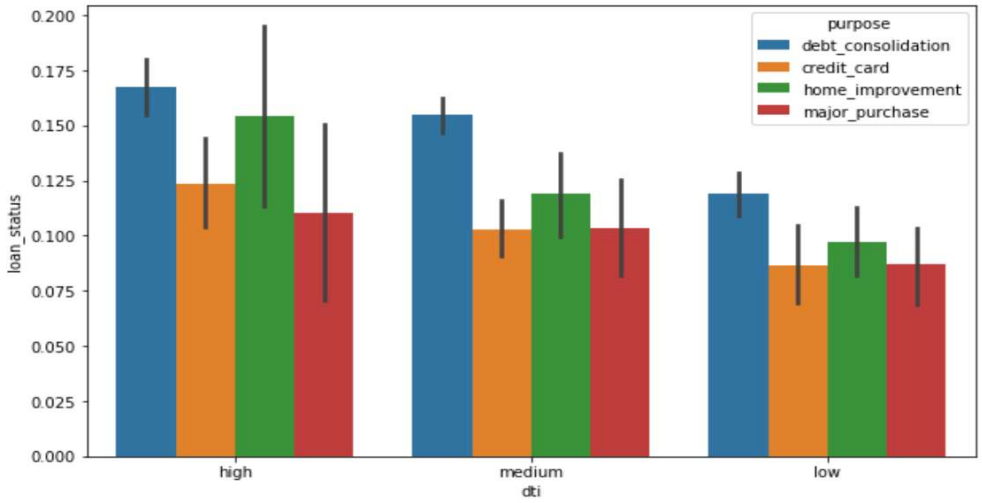

count
16000
14000
12000
10000
8000
6000
4000
2000
0
car
other
wedding
debt_consolidation
credit_card
home_improvement
major_purchase
medical
moving
small_business
vacation
house
renewable_energy
educational

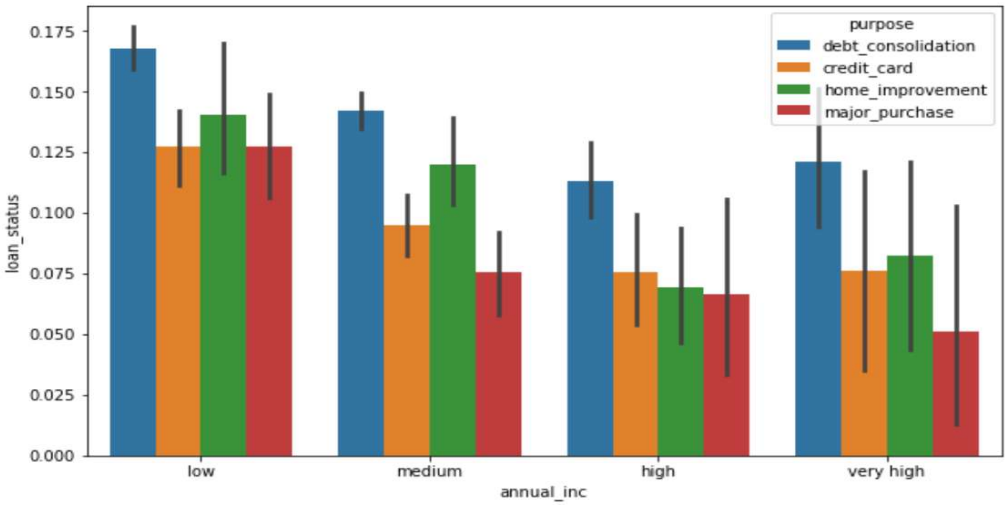

purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.25
0.20
0.15
0.10
0.05
0.00
60 months
36 months
term

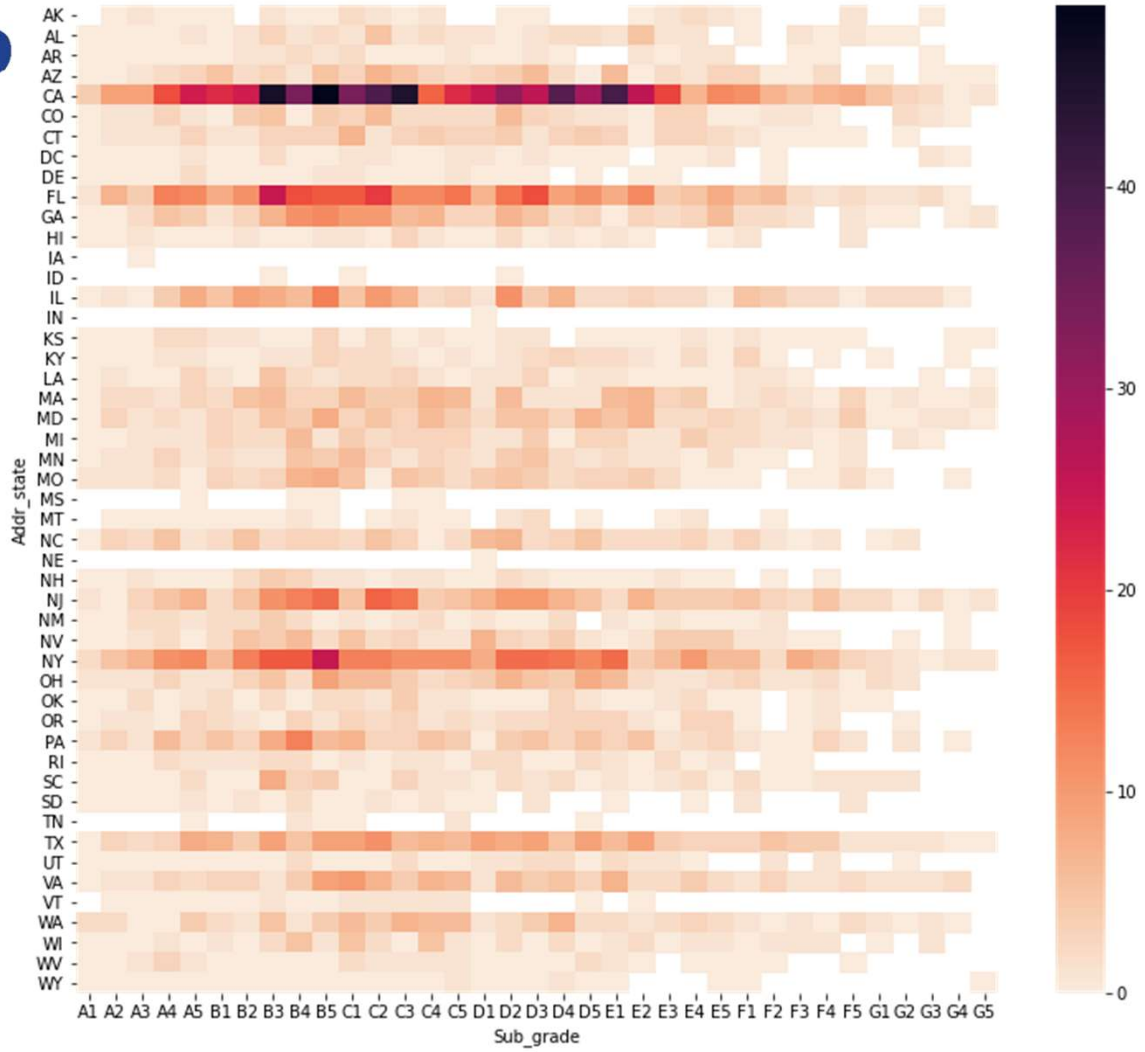

purpose
purpose
debt_consolidation
credit_card
home_improvement
major_purchase
loan_status
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
C
B
D
A
E
F
G
grade


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
0.25
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
0.200
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
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
low
medium
high
very high
annual_inc



Recommendations

- Stop approving loans to people with prior bad record. Or at least stop approving high-value loans.
- Reduce number of approvals where purpose is small business.
- Start charging higher interest rates for loans with dti greater than 20
- Stop approving loans where amount/income is higher than 30%

Conclusions

- As per the heatmap we can conclude location of the borrower has an important role in identifying the defaulting behaviours, but there are several other column variables which have quite and influence on the defaulters behaviour.
- The following columns are important driving factors, 'sub_grade', 'pub_rec_bankruptcies', 'addr_state', 'grade' and 'int_rate'. Since grade and subgrade are dependent variables we can add the 'term' column also as a driving factor for defaulting.