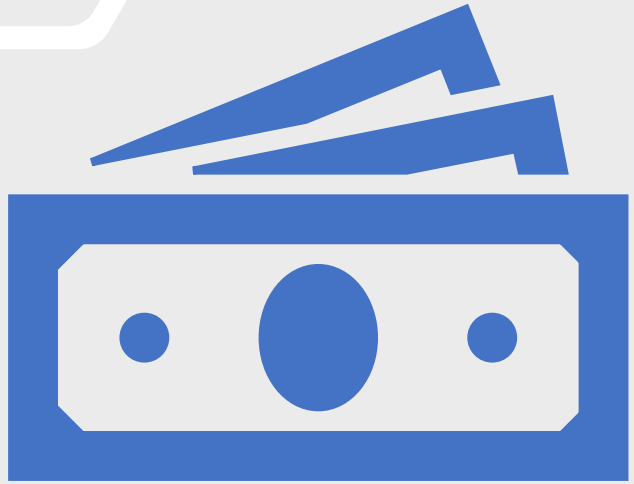
Lending Club
Case Study



By:
Aditya Repekar & Alok Agarwal

Company Profile

- A consumer finance company, specializes in lending various types of loans to urban customers

Problem Statement & Action

Problem

- When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:
  - If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
  - If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company.

Action

- We will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default. What are the driving factors which are strong indicators of getting default which can be utilized for risk assessment.
- We have collected a dataset that contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Case Study

Data Understanding

What we did

Download the Loan dataset & load to pandas dataframe.

Gone through all the different variables.

For better understanding, data dictionary sheet provides detailed description of these variables.

Outcome

Better understanding of different variables in dataset

Understanding of Business terminologies.

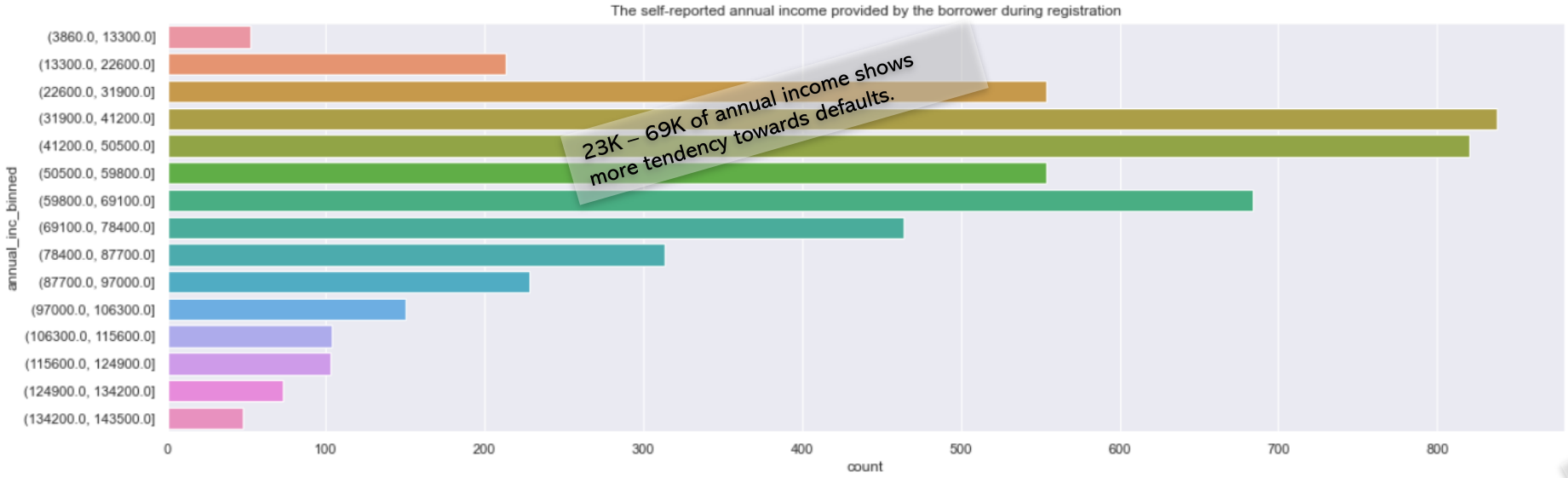
Data Cleaning & Manipulation

- In this process, we identify & get rid of the columns which have:
  - Only NA values
  - Only one constant value, as no meaningful can be derived out of that column.
  - More than 30% is either blank or NULL, as that will impact the analysis.
  - Random strings, IDs, urls etc which has no significance in analysis
  - Other redundant variables (Behavioural attributes)
  - Removing the outliers for e.g from column ‘annual income’ while checking the IQR values.
  - Some columns like interest rate(int_rate) is manipulated to strip off the “%” character.
  - Manipulating column issue_d to represent month & year in separate derived columns.

Univariate Analysis

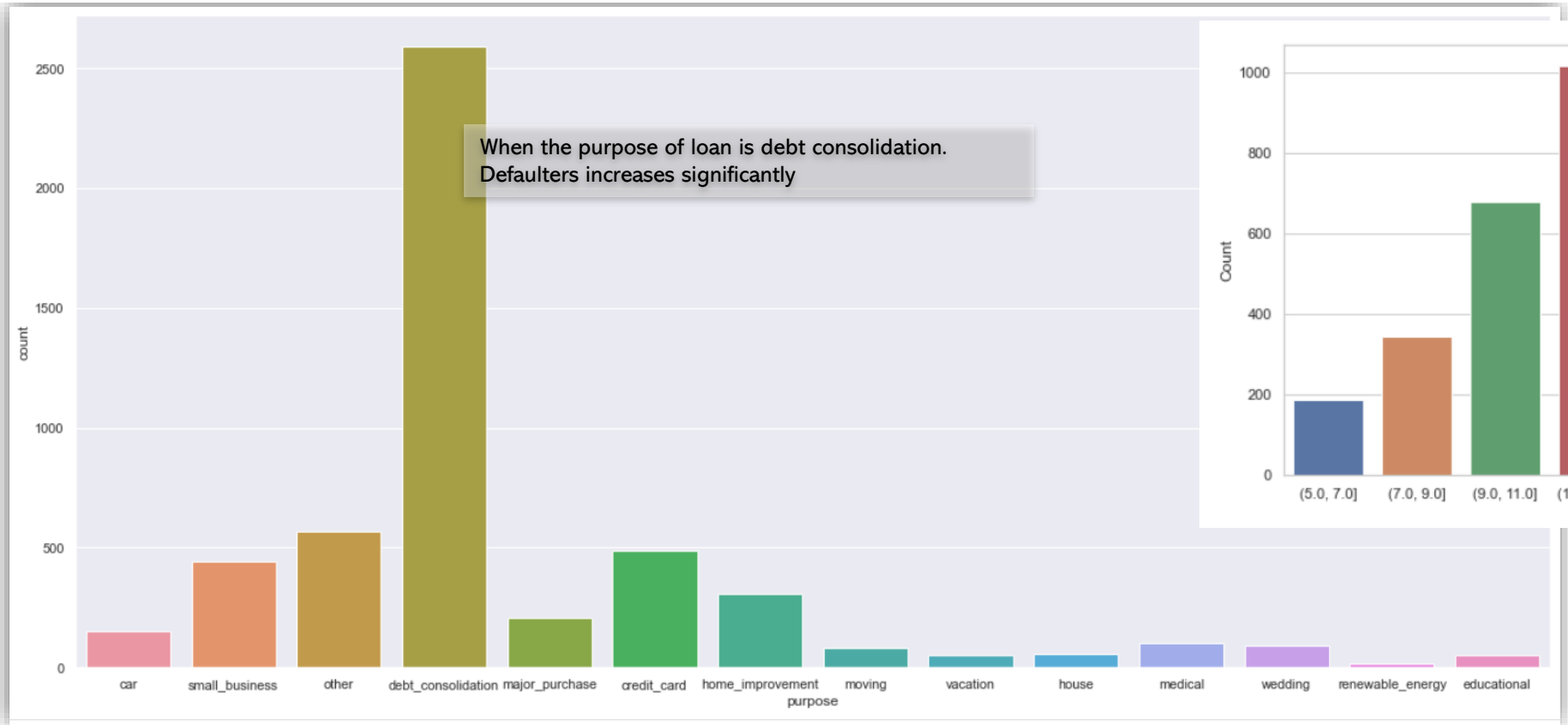
Univariate Analysis

- Univariate analysis covers examination of distribution of default cases against one indicator (variable) at a time.
- During the data analysis, we figured out some strong indicators of default.
  - Purpose of the loan
  - Interest Rate
  - Home Ownership
  - Employment length
  - Instalment amount
  - Sub grade etc

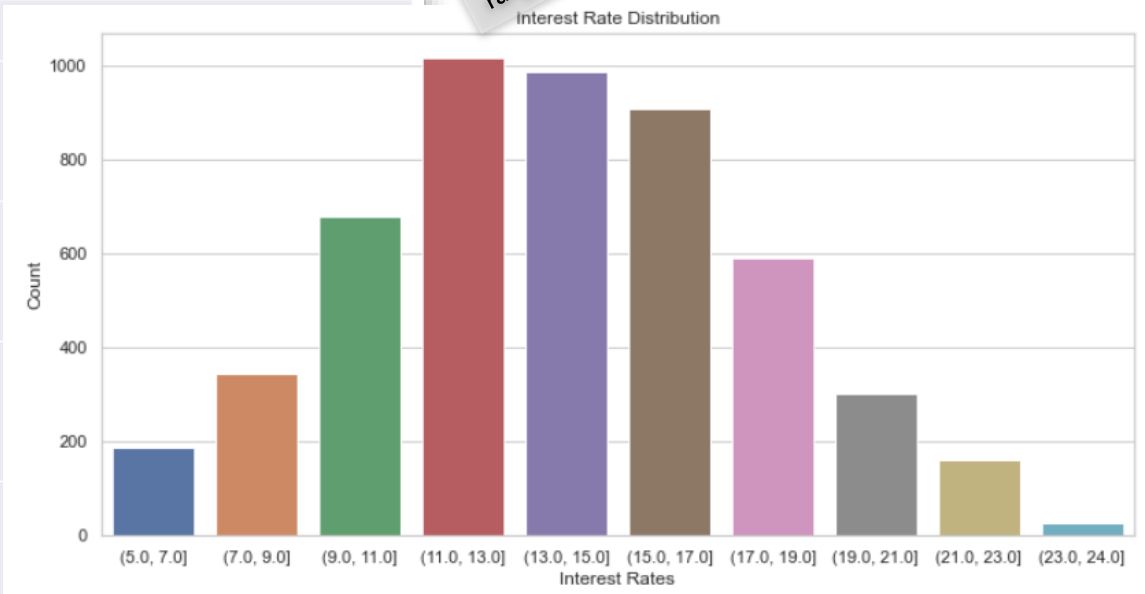


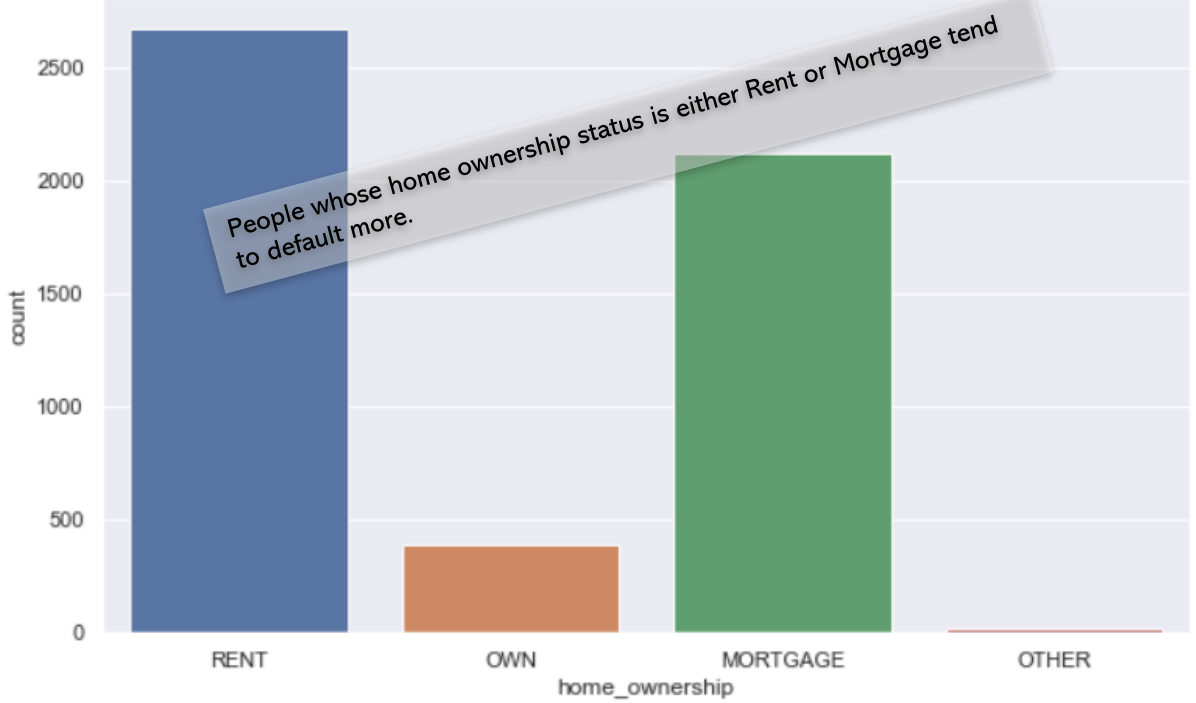
23K – 69K of annual income shows more tendency towards defaults.

Defaulters increases between interest rate range of 11% to 17%

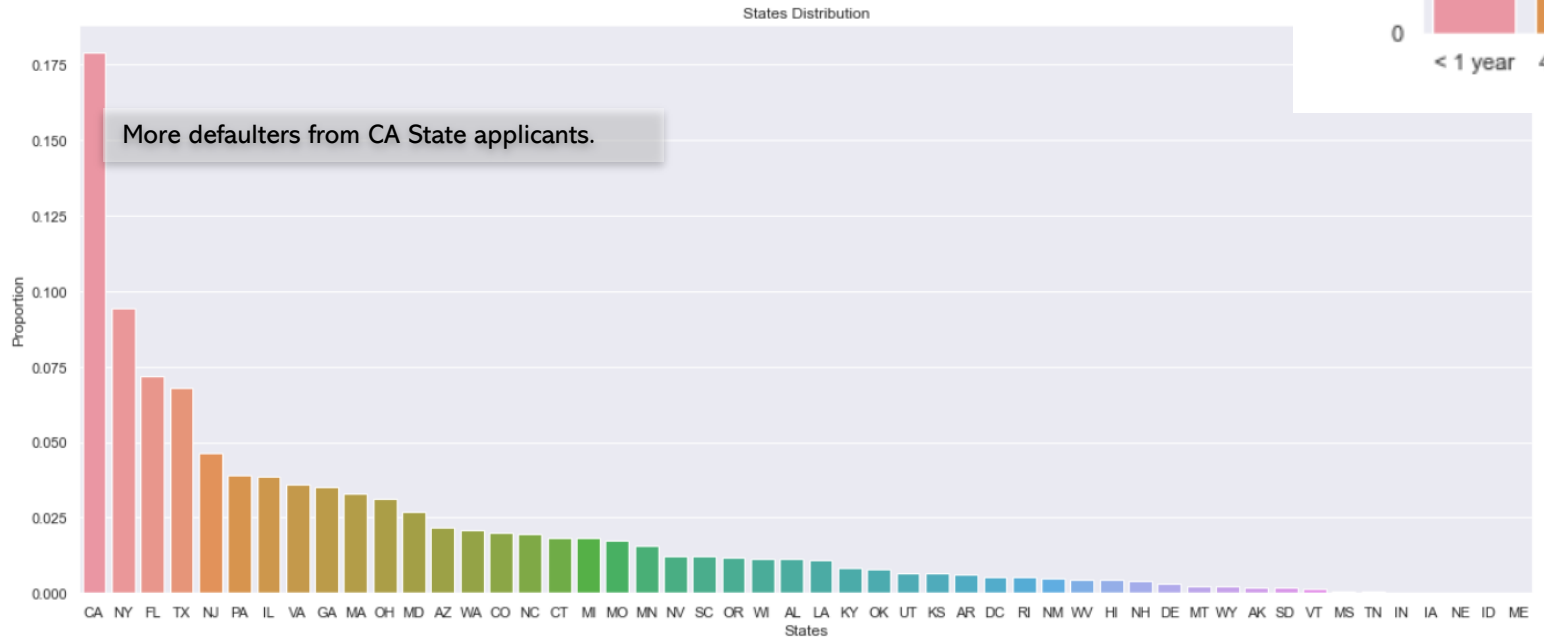


When the purpose of loan is debt consolidation. Defaulters increases significantly

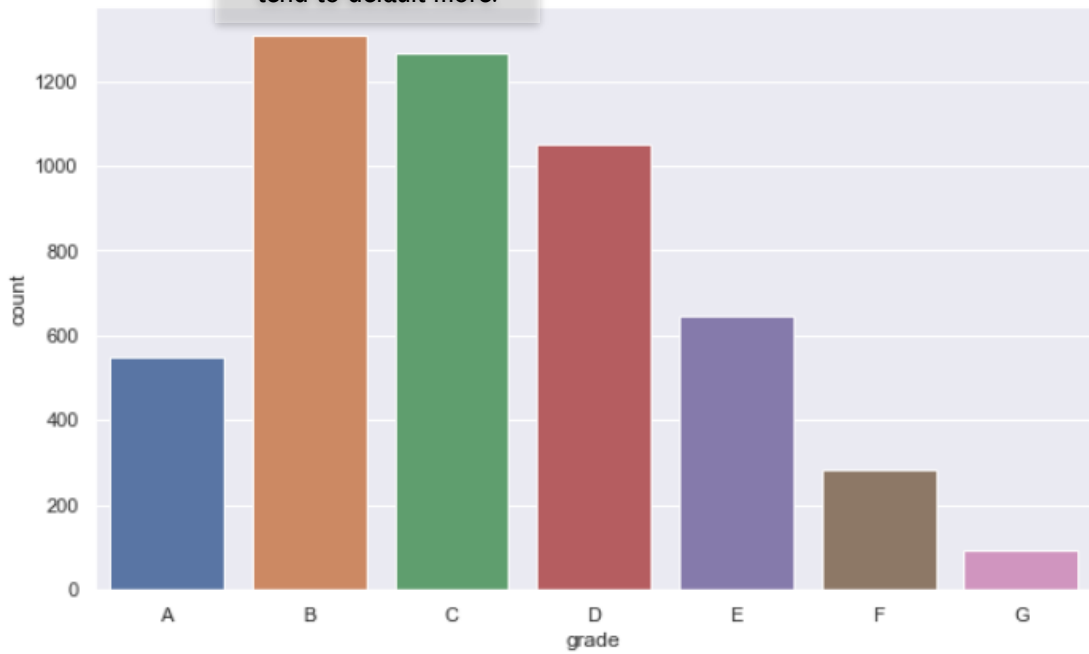



People whose home ownership status is either Rent or Mortgage tend to default more.
count
2500
2000
1500
1000
500
0
RENT
OWN
MORTGAGE
OTHER
home_ownership


Employment Length Distribution
People employed for > 10 years default more.
Count
1200
1000
800
600
400
200
0
< 1 year
4 years
3 years
10+ years
1 year
9 years
2 years
8 years
7 years
5 years
6 years
Employment Length

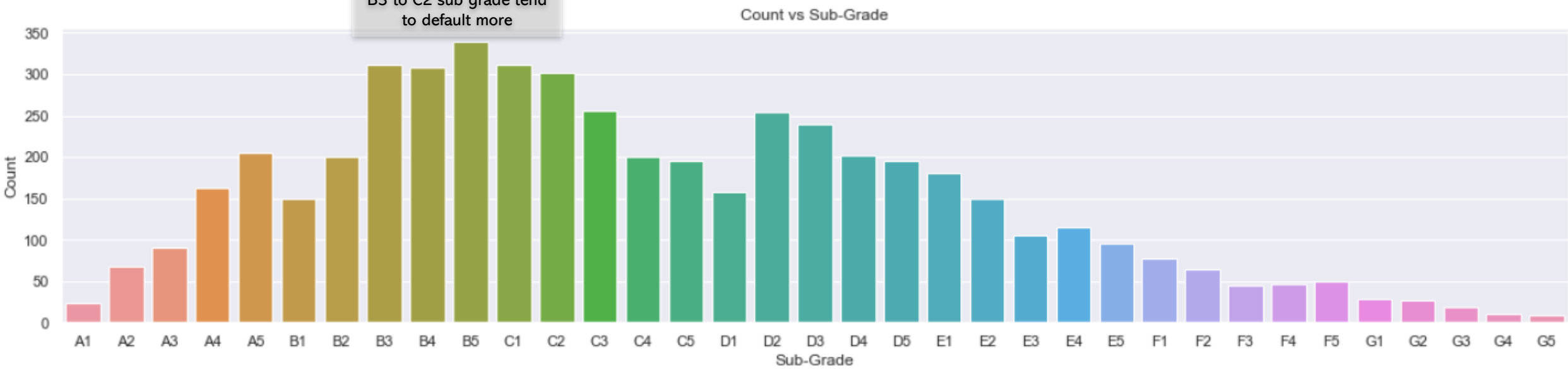

States Distribution
More defaulters from CA State applicants.
Proportion
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
CA NY FL TX NJ PA IL VA GA MA OH MD AZ WA CO NC CT MI MO MN NV SC OR WI AL LA KY OK UT KS AR DC RI NM WV HI NH DE MT WY AK SD VT MS TN IN IA NE ID ME
States

B & C Grade
tend to default more.



In B & C grade,

B3 to C2 sub grade tend
to default more



Bivariate Analysis

Bivariate Analysis

- Bivariate analysis covers examination of distribution of default cases against two different indicator (variable) at a time. i.e how one variable is influencing or providing more insights to results of another variable.
- During the data analysis, we captured degree of strength of correlation of different indicators that represents defaults.

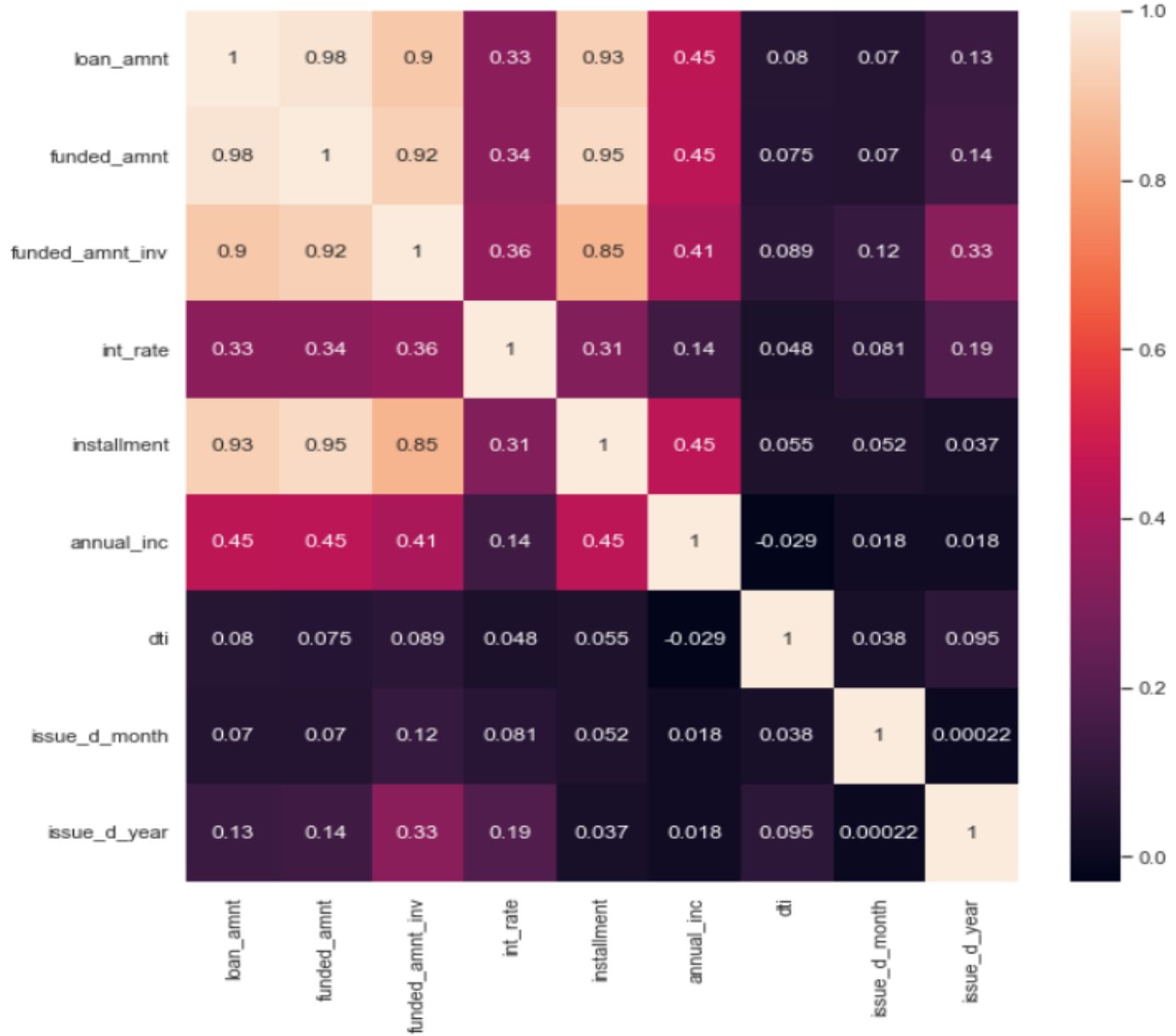
Correlation matrix

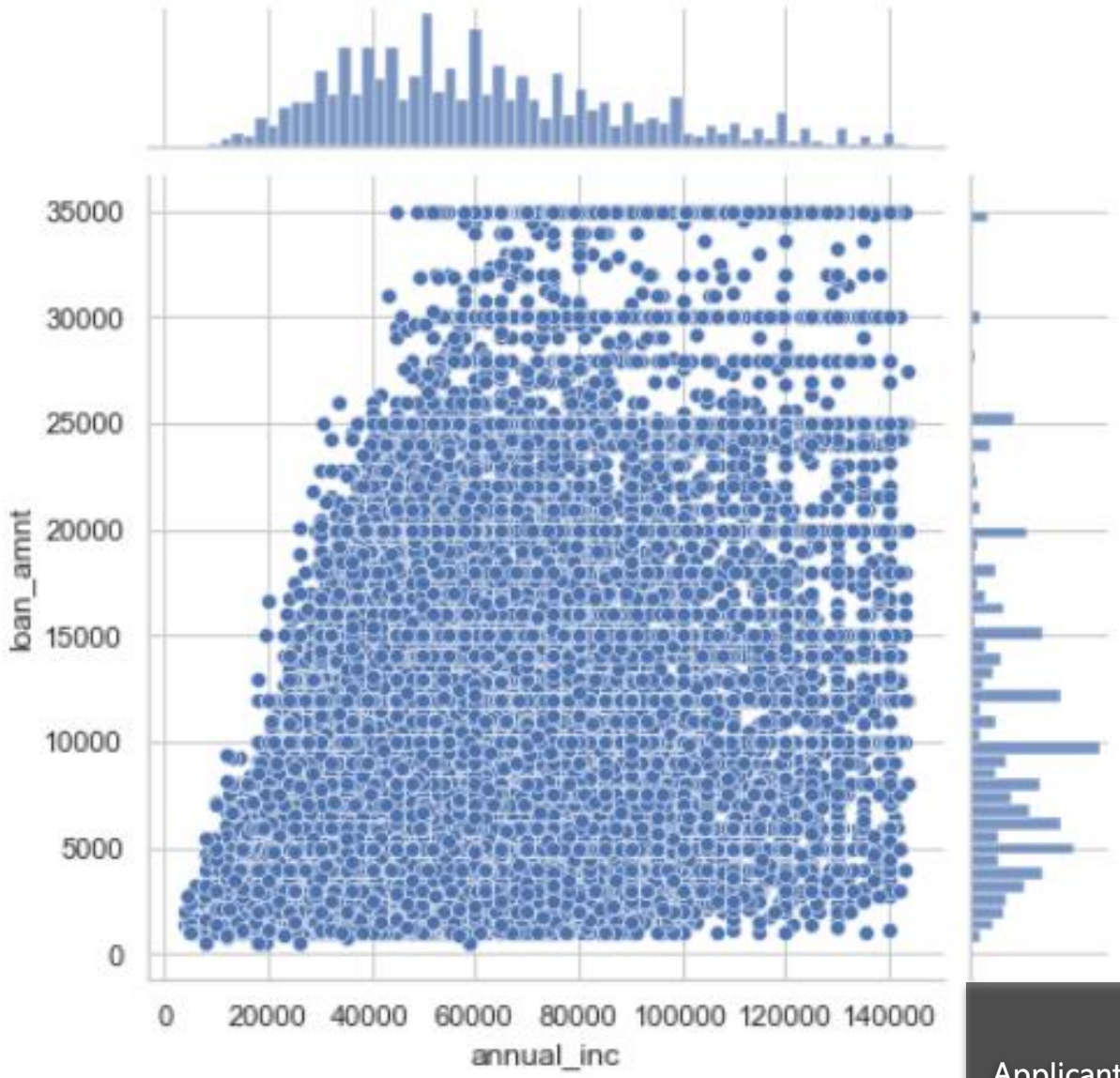
reflects that

loan_amnt, funded_amnt, funded_amnt_inv are highly correlated

and

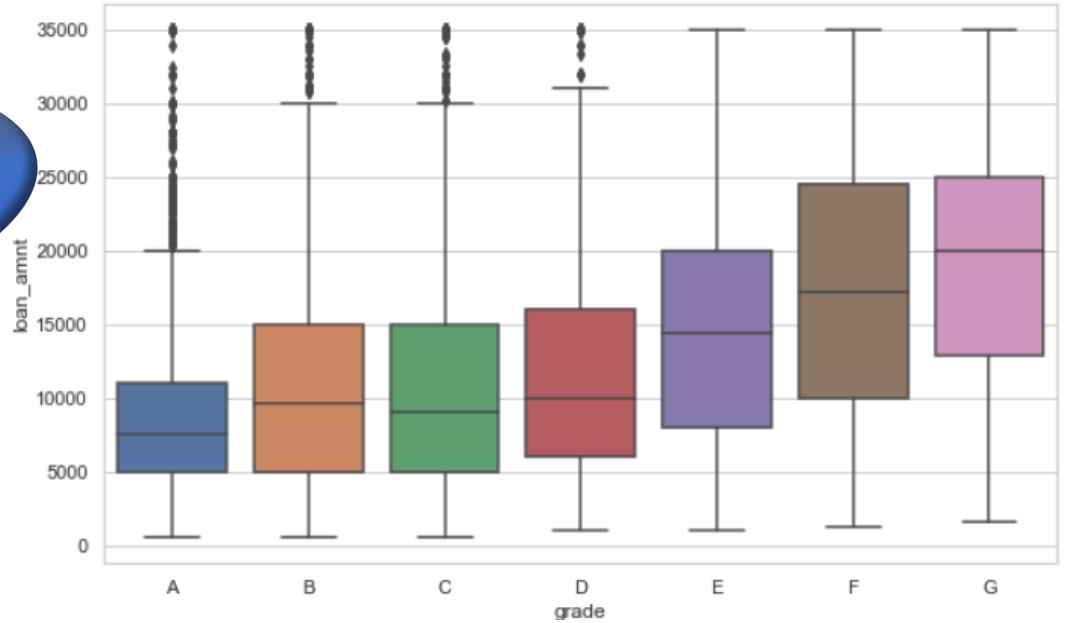
annual income is negatively correlated with debt-to-income ratio



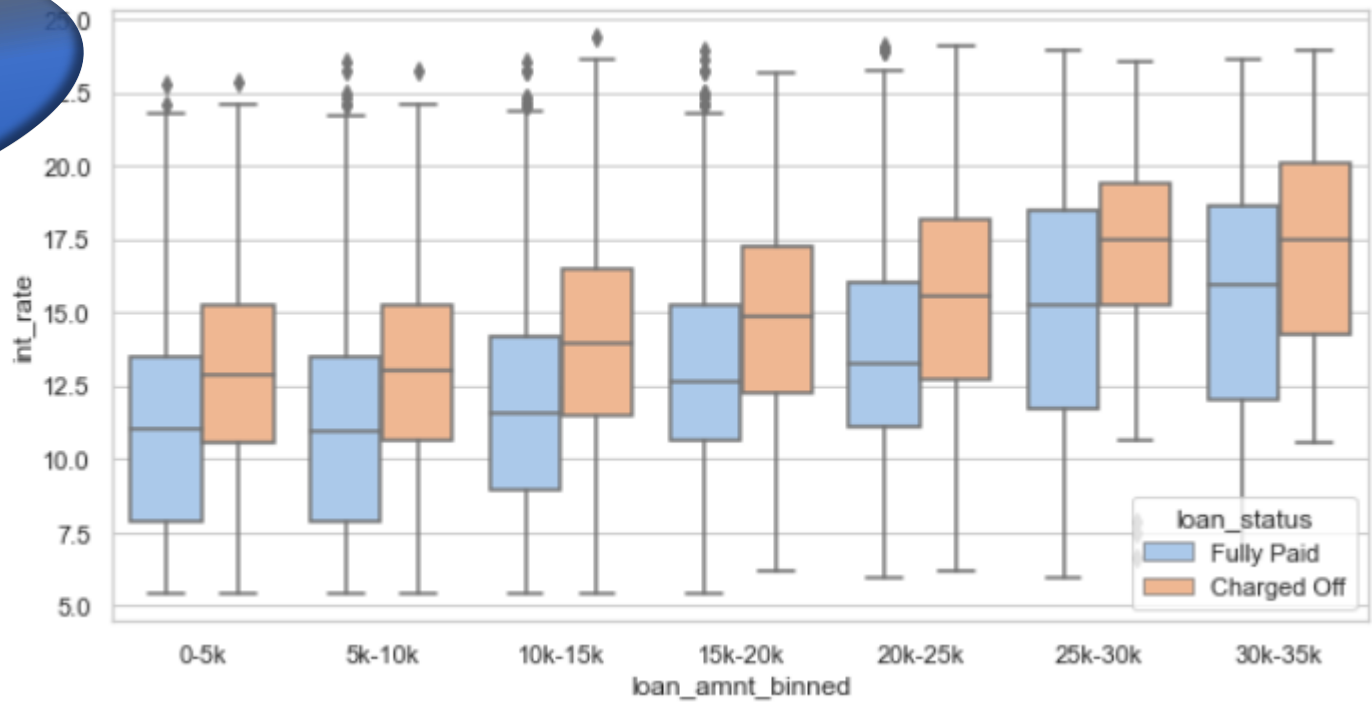


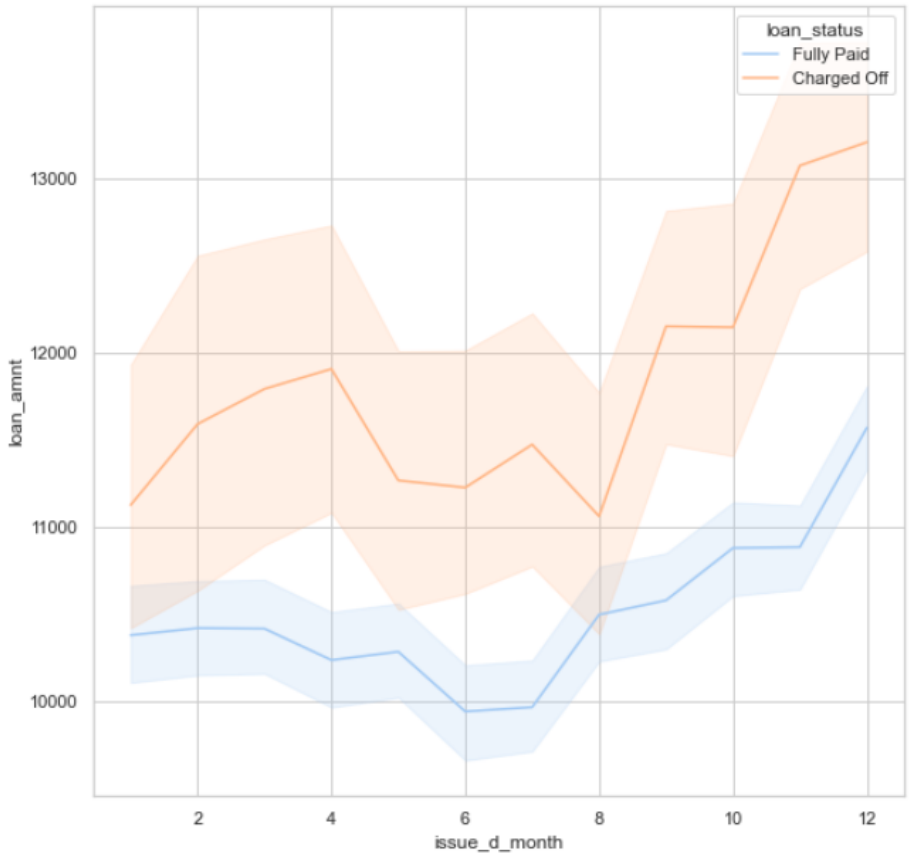
Applicants whose annual income is between 40K-60K & taking loans amounting above 25K are more likely to default

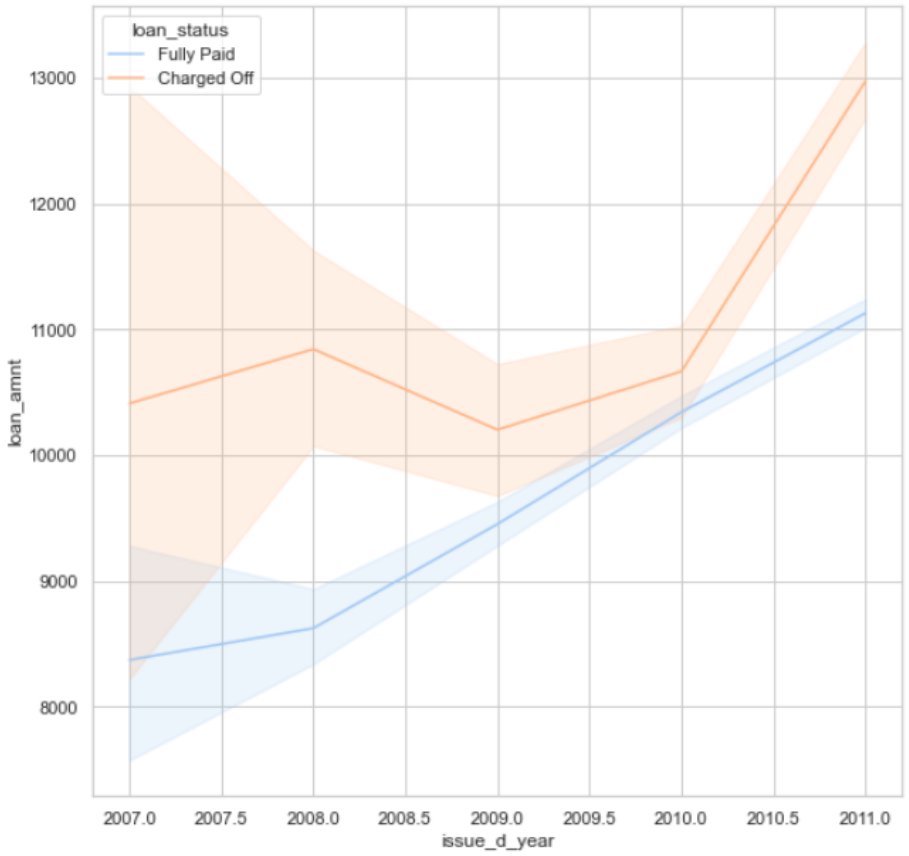
The plot illustrates a trend that shows loan amount increases inversely with grade.



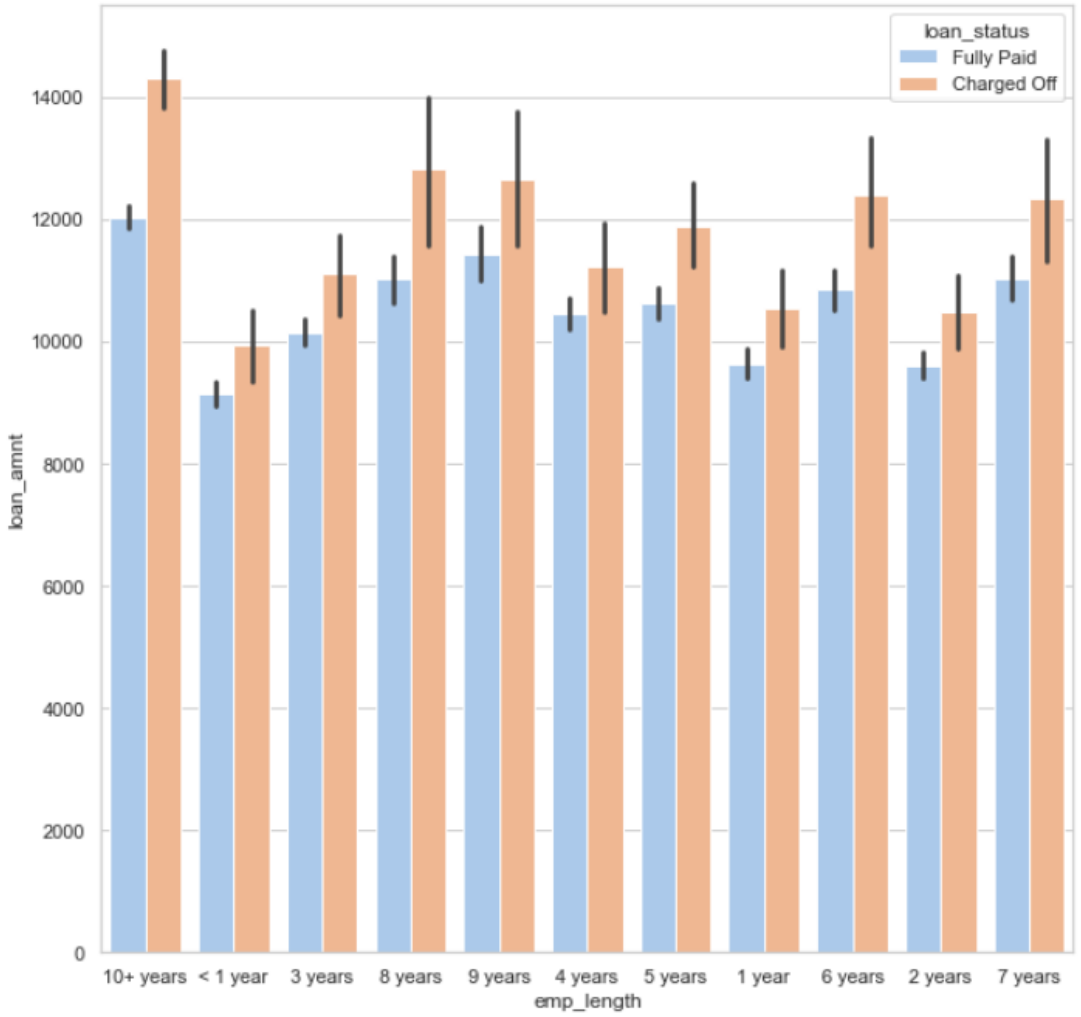
Higher Interest rate at higher loan amounts tend results in loan defaults.



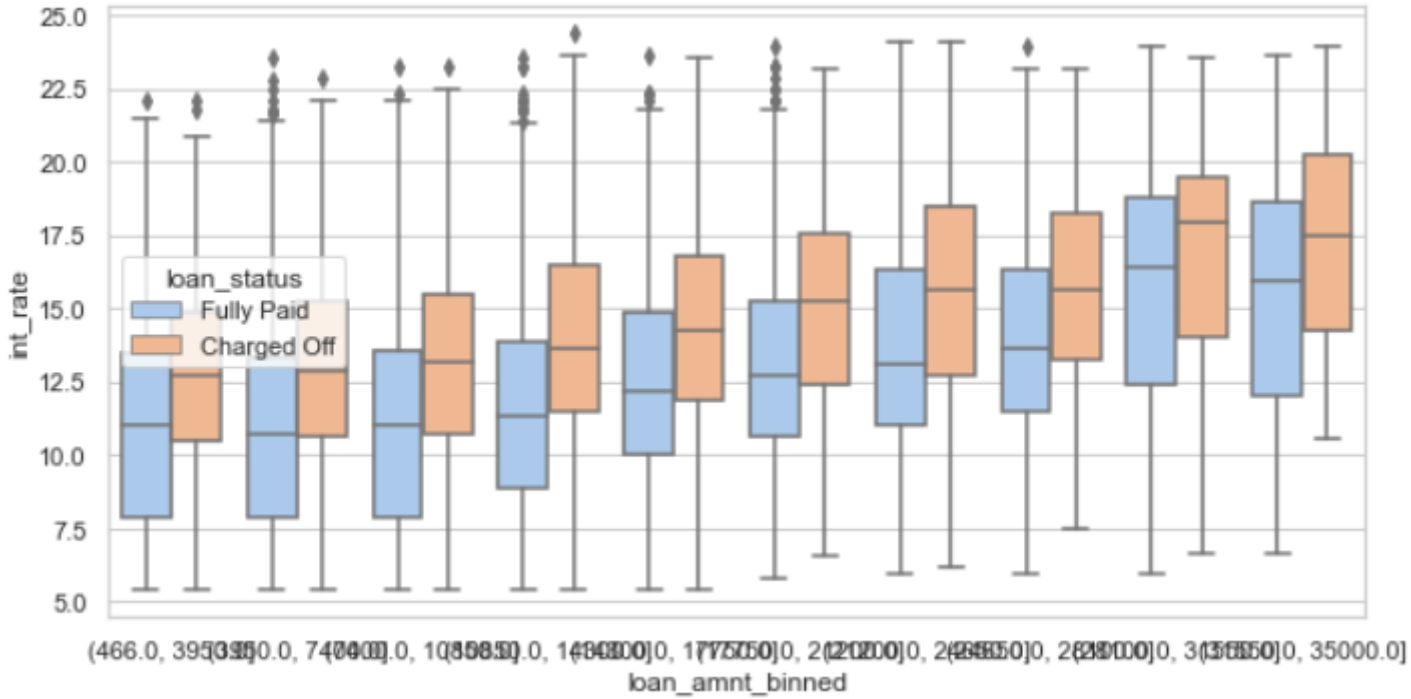




Loan amount requested during and after sub-prime mortgage crisis has shown an upward trend of being charged off

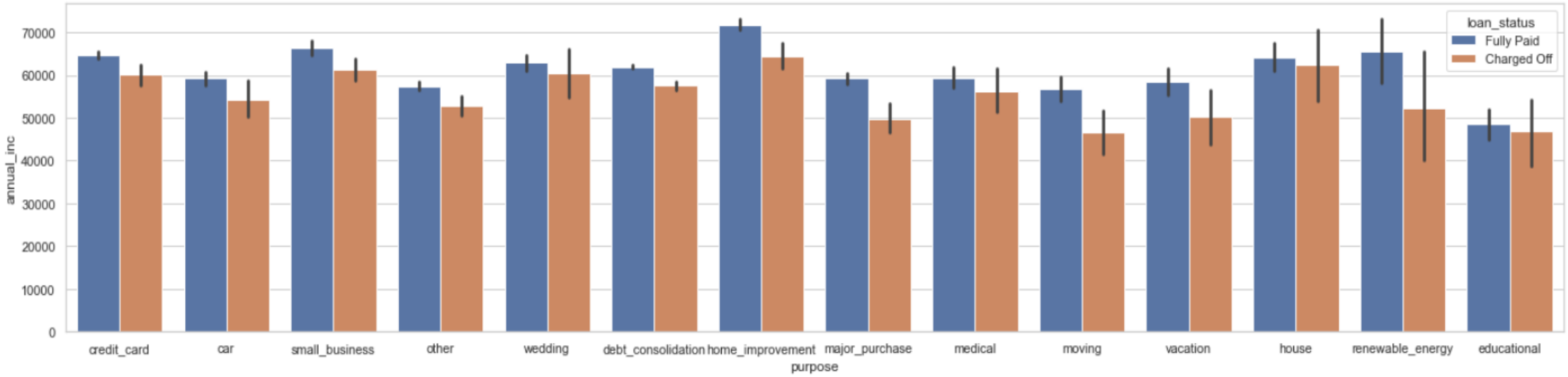


People with 10+ years of experience got the loan approved for higher amount, but eventually resulting in most defaulters

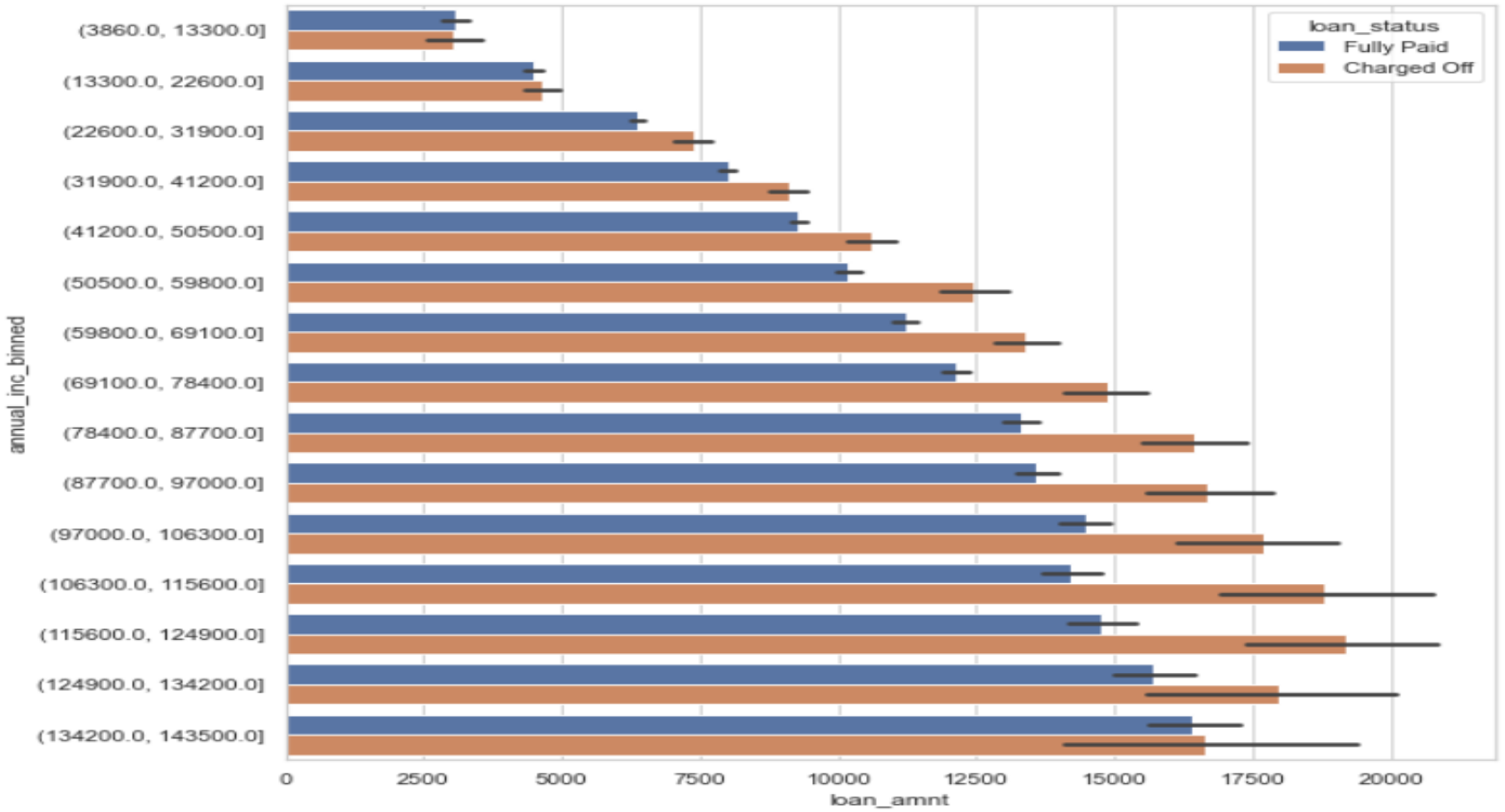


Interest rate is higher for high loan amounts which can strongly result in loan default

High annual income applicants when take loan for home improvement, house, credit card, small business, wedding tend to default more



As annual income of applicants increase, they take higher loan amount & tend to default more



Prescriptive Insights

- Based on the univariate & bivariate data analysis and understanding the behaviour of different plots, we have the following outcomes

Risk Considerations (Defaulters trend)

Univariate Analysis

- Annual Income range ➔ 23k – 69k
- Employment length ➔ 10+ years
- Interest rate ➔ 11-17%
- House Ownership ➔ RENT
- Loan amount ➔ 4k-10k
- Purpose ➔ Debt consolidation
- Debt-to-income ratios ➔ 12-18
- Monthly Instalments ➔ 145-274
- Grade ➔ B & C
- Sub Grade ➔ B3 to C2
- State ➔ CA

Bivariate Analysis

- Annual Income between 40k-60k are taking loan of 25000 and more
- Status is verified and loan amount above 15000
- Home ownership is MORTGAGE and loan amount is between 12k-14k
- Employment length is 10+ years and loan amount is 12k-14k
- Loan in the range 30k - 35k and are charged interest rate of approx. 15-20 %
- Loans taken for maximum term for higher loan amount in the range of 14k-16k and with high interest rates between 12.5-17.5%
- Loan for home improvement, house, credit card, small business, wedding and has annual income of 60k -70k
- Home ownership is MORTGAGE and have income of 60-70k
- Annual income is between 112k-140k and loan amount above 17k
- Interest Rate of 21-24% and have an income of 60k-70k
- Home ownership is MORTGAGE and have loan of 14-16k

Content Repository

GitHub Repository Of Source Code & Presentation

https://github.com/adityarepekar/Lending_Club_Case_Study_EDA

Thankyou