Lending Club Data Analysis

Prabir Mallick
Sreejith Rajashekaran

01 Introduction
About company and problem statement

02 Analysis
Explorer data Analysis for lending club and analyze result

03 Recommendation
Recommendation and suggestion to investors for the business

About Lending Club

This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss)

39718 loan request successfully processed

17% revenue

Fully Paid Loan

57% money was recovered for charged off loan

Analysis

❖ Problem

As a data scientist for lending club
analyze the past loan information and
Try to understand the different features,
Attributes to influence the Defaulter

❖ Context

Company wants to understand the **driving
factors (or driver variables)** behind loan
default,
i.e. the variables which are strong
indicators of default. The company can
utilise this knowledge for its portfolio and
risk assessment.

❖ Approaches

- Analyze raw data and understand the overall attributes
- Clean data (Drop columns with null values or single category value)
  Convert column values to integer format

- Check the distribution and frequency of each attribute and derived attributes

- Analyze the variables against target variable

- Analyze correlation between features
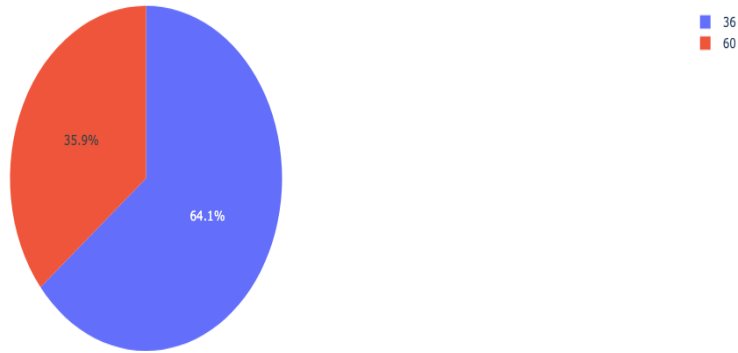
❖ Solutions

- Summarize results

- Recommendation

Analysis of Overall Loan
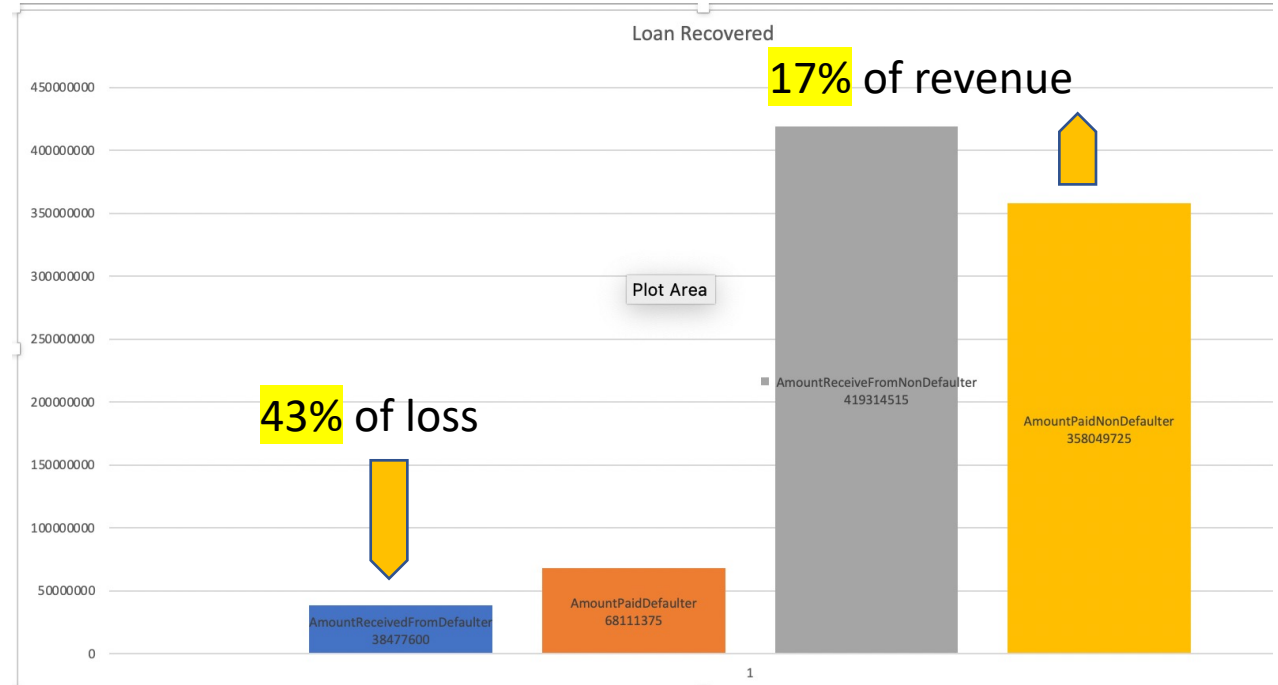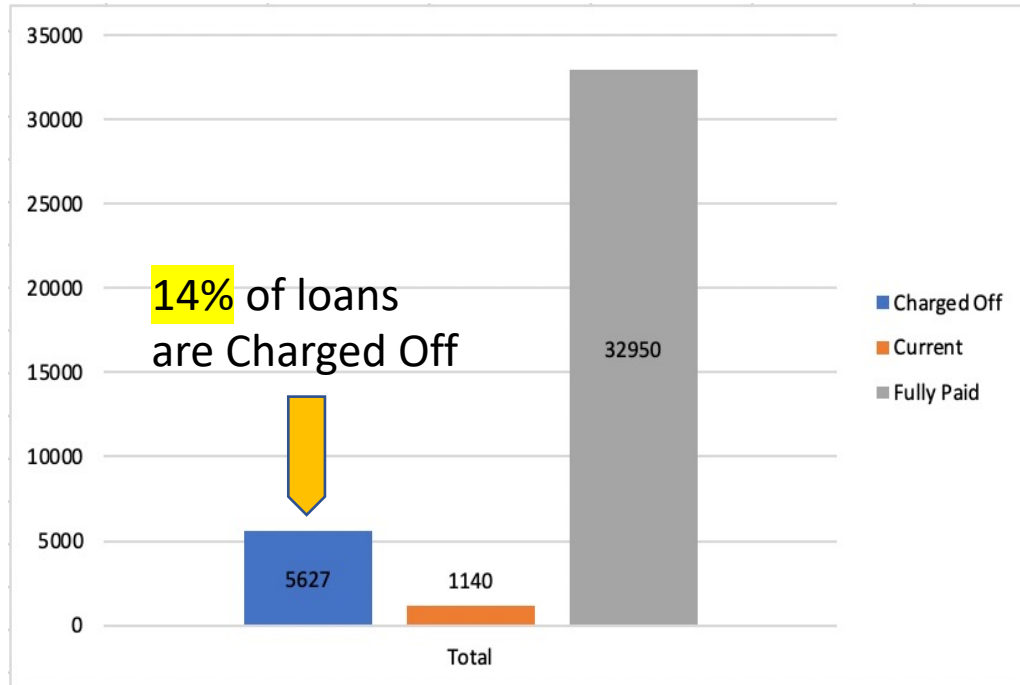
- 14% loans are charged off
- 43% loan amount is loss due to Defaulter
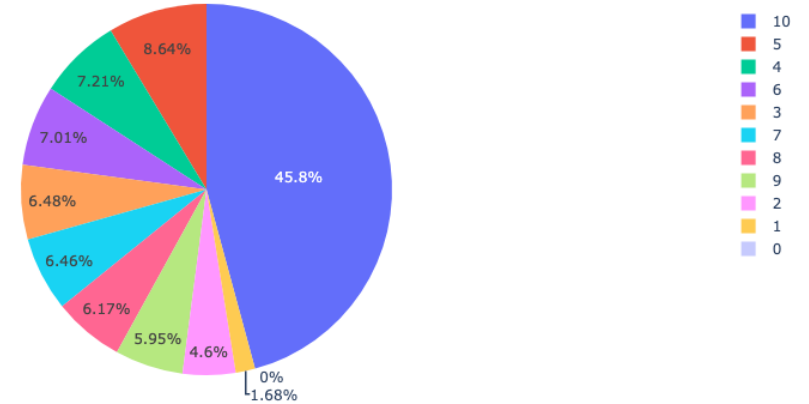- 17% revenue for Fully Paid loan

Term Loan

64% are of 60 months
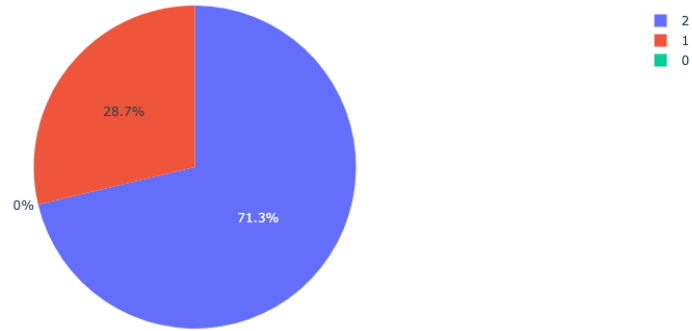
Observation: Preference of loan term is maximum

Emp Length

46% of loan applicants are from 10+ years
Then 5 years of experience

Observation: As exp is more, many responsibility will come and to fulfil in short time, customer will look for loan
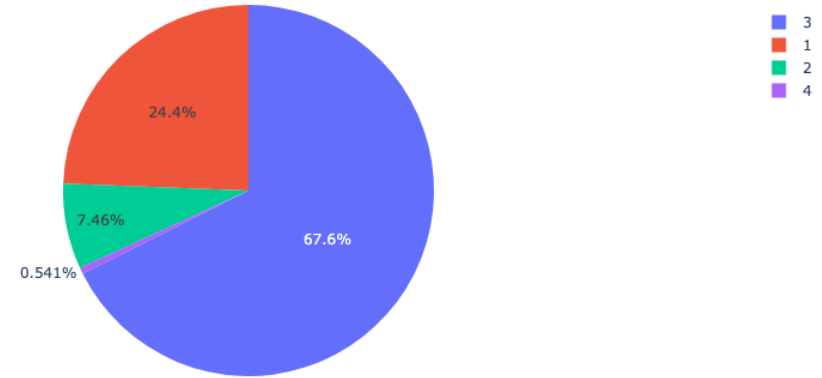
Verification

Home Ownership

Observation:

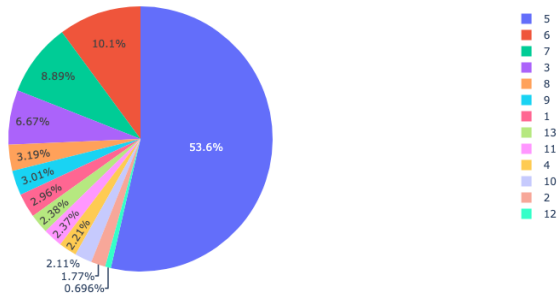- 71% loans are source verified.
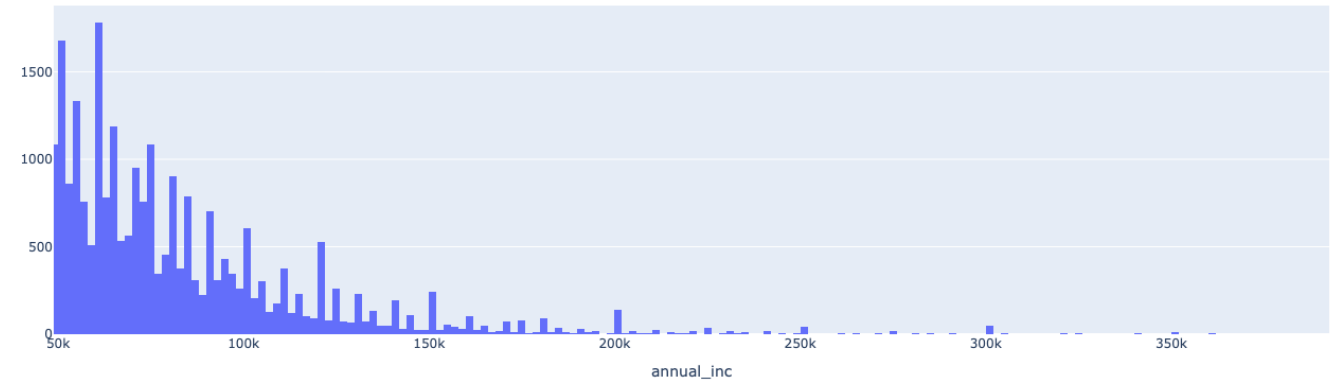- 68% applicants have mortgage home ownership.

Suggestion

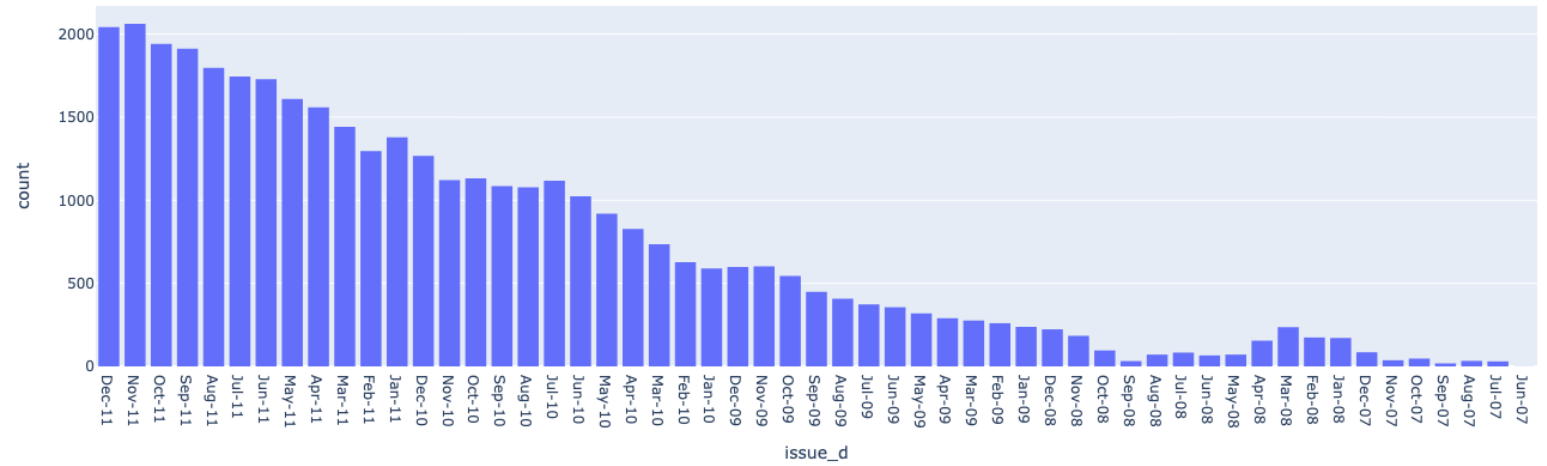Before lending money , company should continue to verify

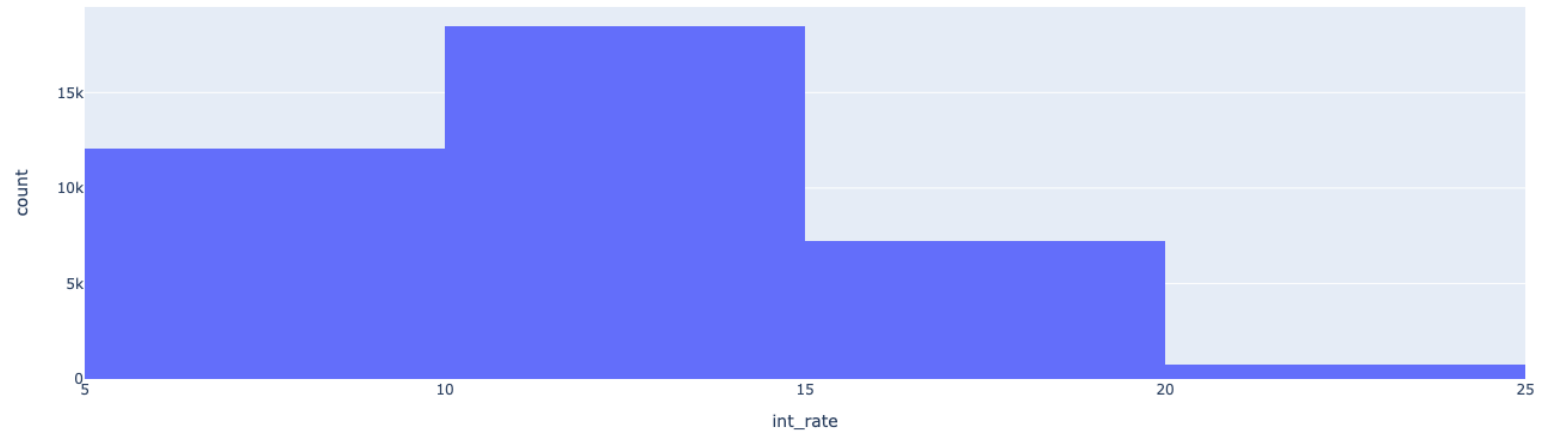# Purpose

Income

Observation:
- Mainly applicants are taking loan for the debt consolidation
- Annual income varies from 52K - 74K
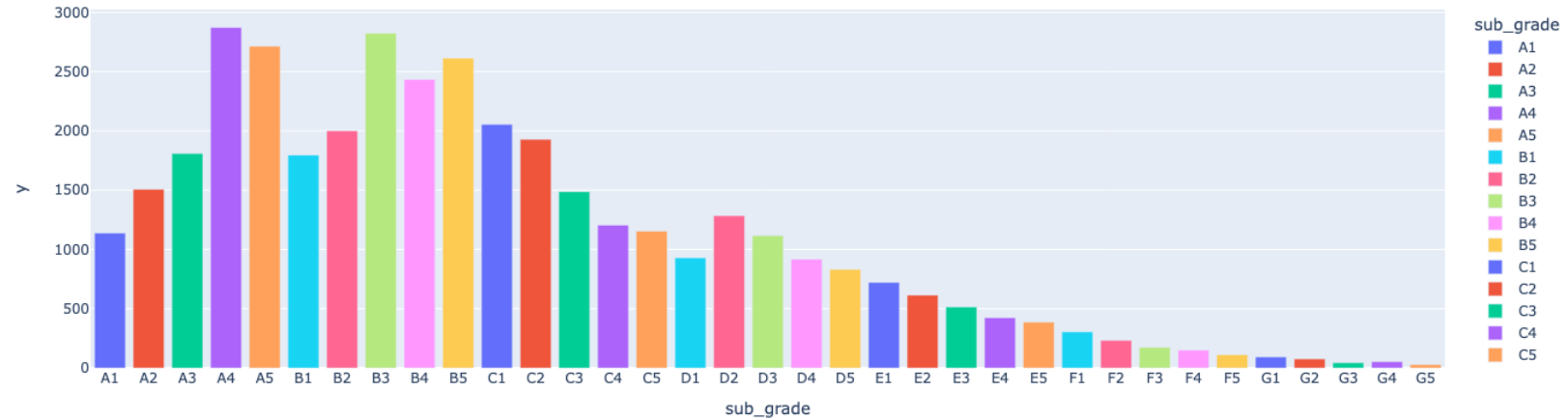
Issue Date

Interest Rate

Observation:
- maximum loan processed between June 2010 and Dec 2011
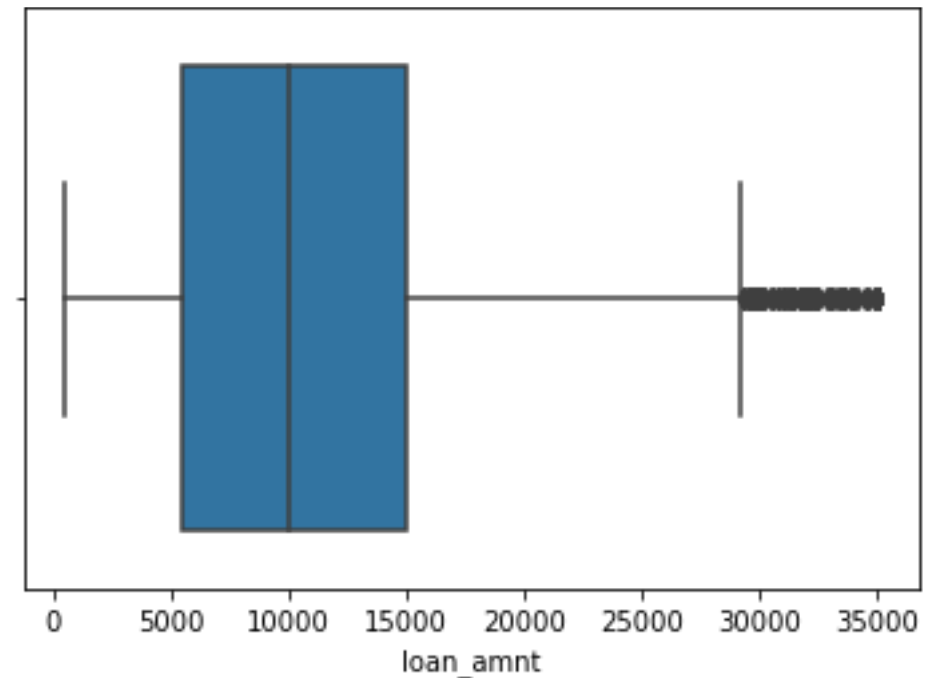- maxium applicants interest rate varies between 10 to 14.99
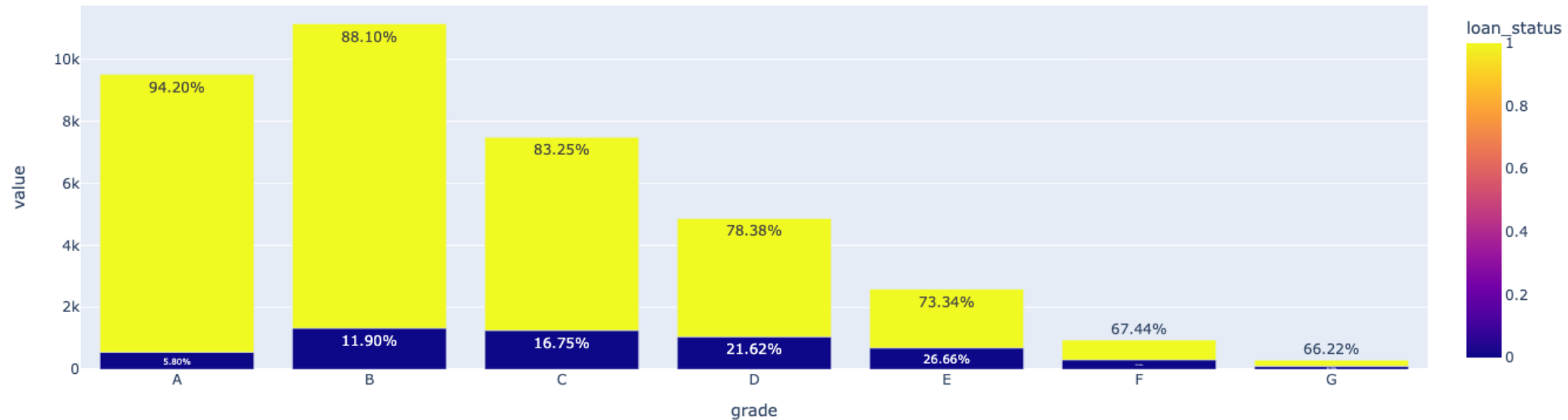
Subgrade

Loan Amount

Observation:
- Most Loan applicants are with grade A4,A5,B3,B4,B5
- Maximum loan amount can be funded till 30K

Suggestion:
- Be careful while approving higher amount loan, likely to be maximum defaulter.
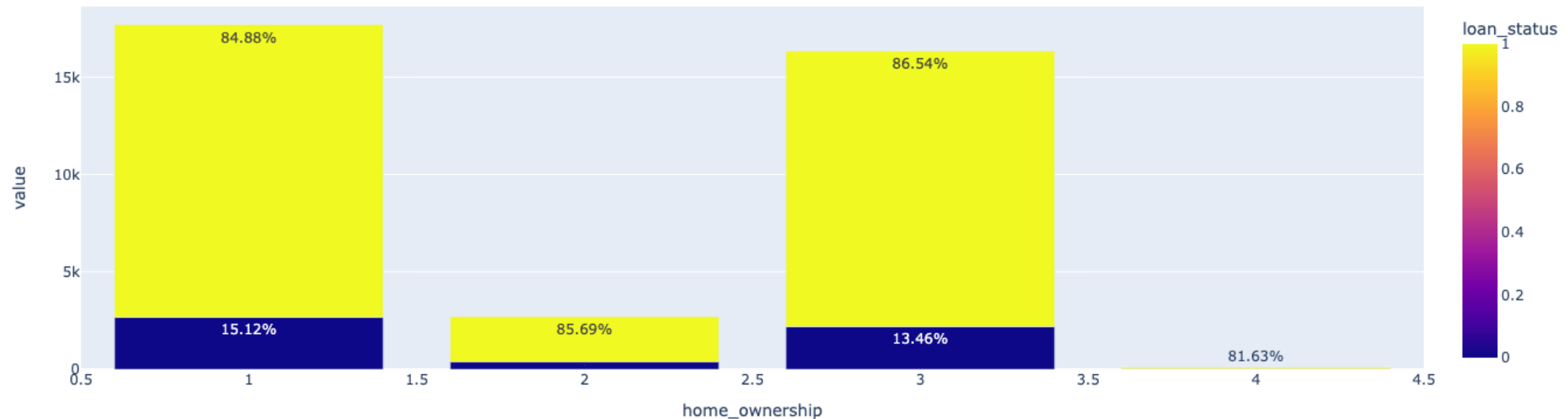
Grade is influencing the defaulter

Observation:
22% and 27%grade D and E category is mostly defaulter.

Suggestion:
- Avoid to lend money to these category
- Loan can be given in higher interest rate.
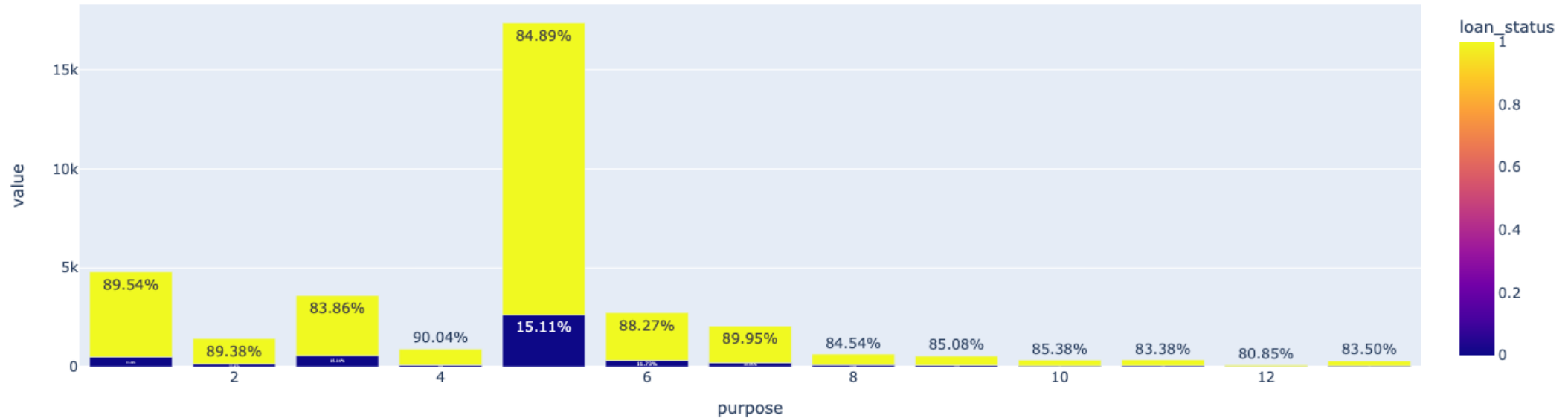
Home Ownership

Observation:
- 18% of loan applications are defaulter for other type of home ownership.

Suggestion:
- Lender should add process to check home ownership.
- so lender should not provide loan to them or provide higher rate interest.
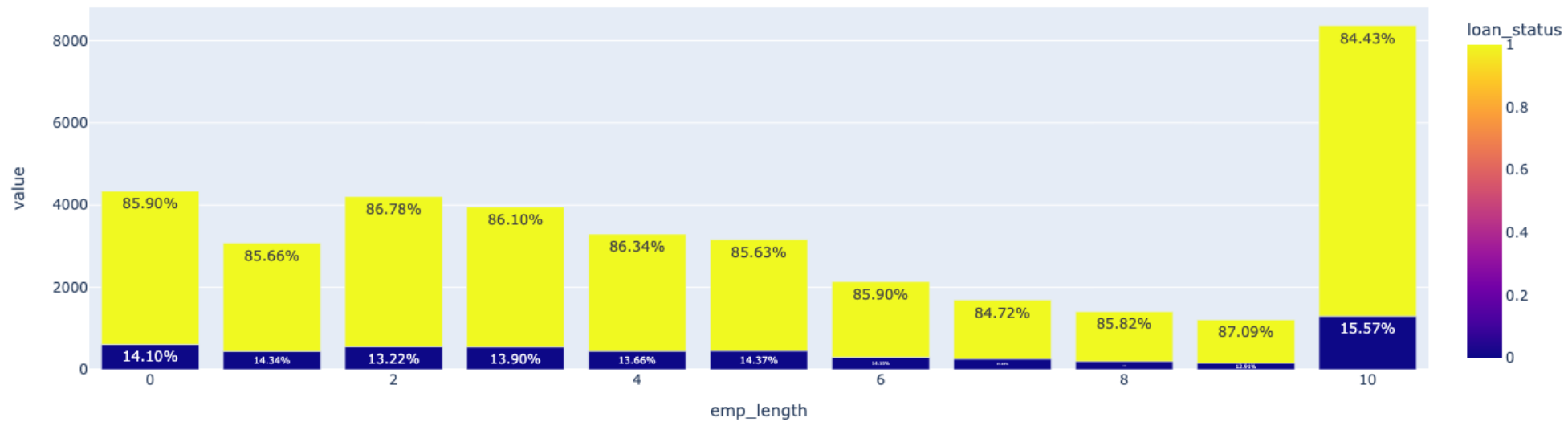
Purpose

Observation
- debt consolidation purpose is highest reason for loan followed by Car
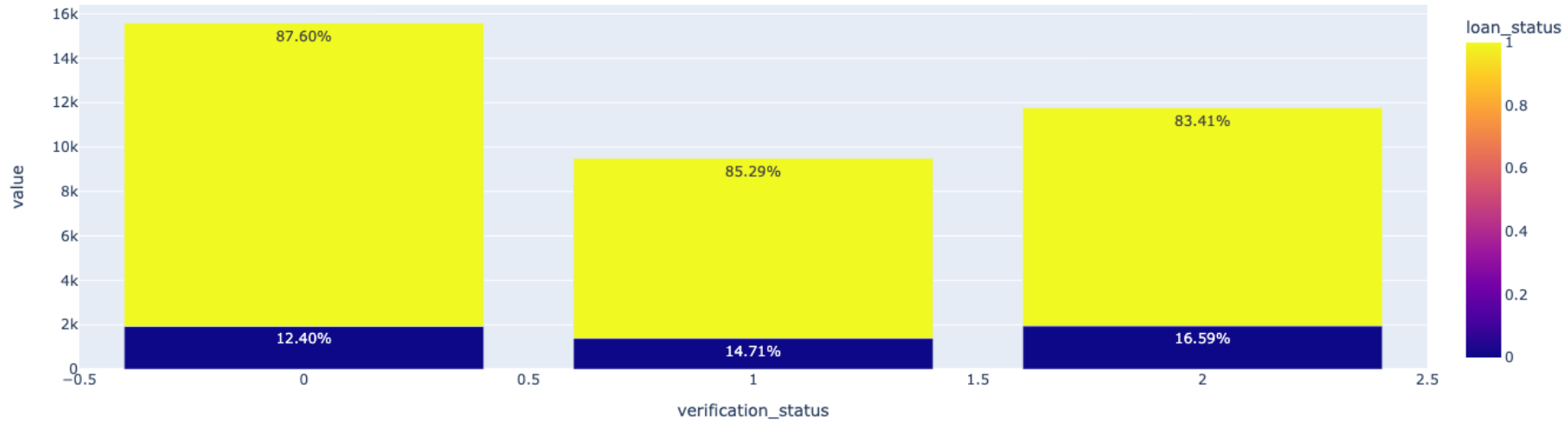- 18% renewable_energy type of application are defaulter

Employee length

Observation:
10+ years exp applicants more
Avg 14% defaulter from each employee group. This variable will not major impact of defaulter

Verification status
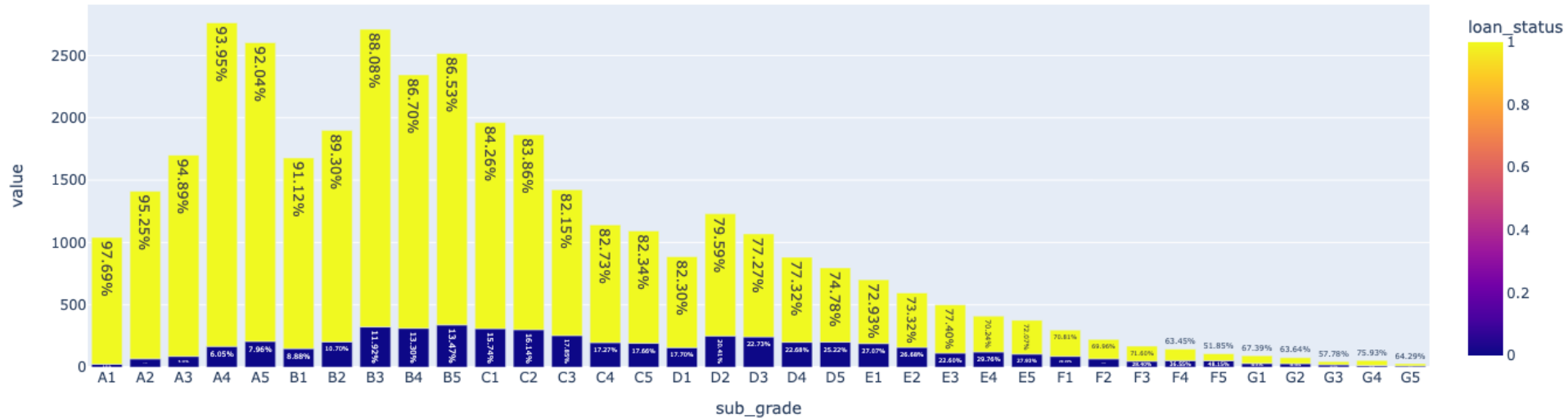

Observation
12% applications are not verified.
Though verified application has more defaulter

Suggestion
Check why 12% applications are not verified
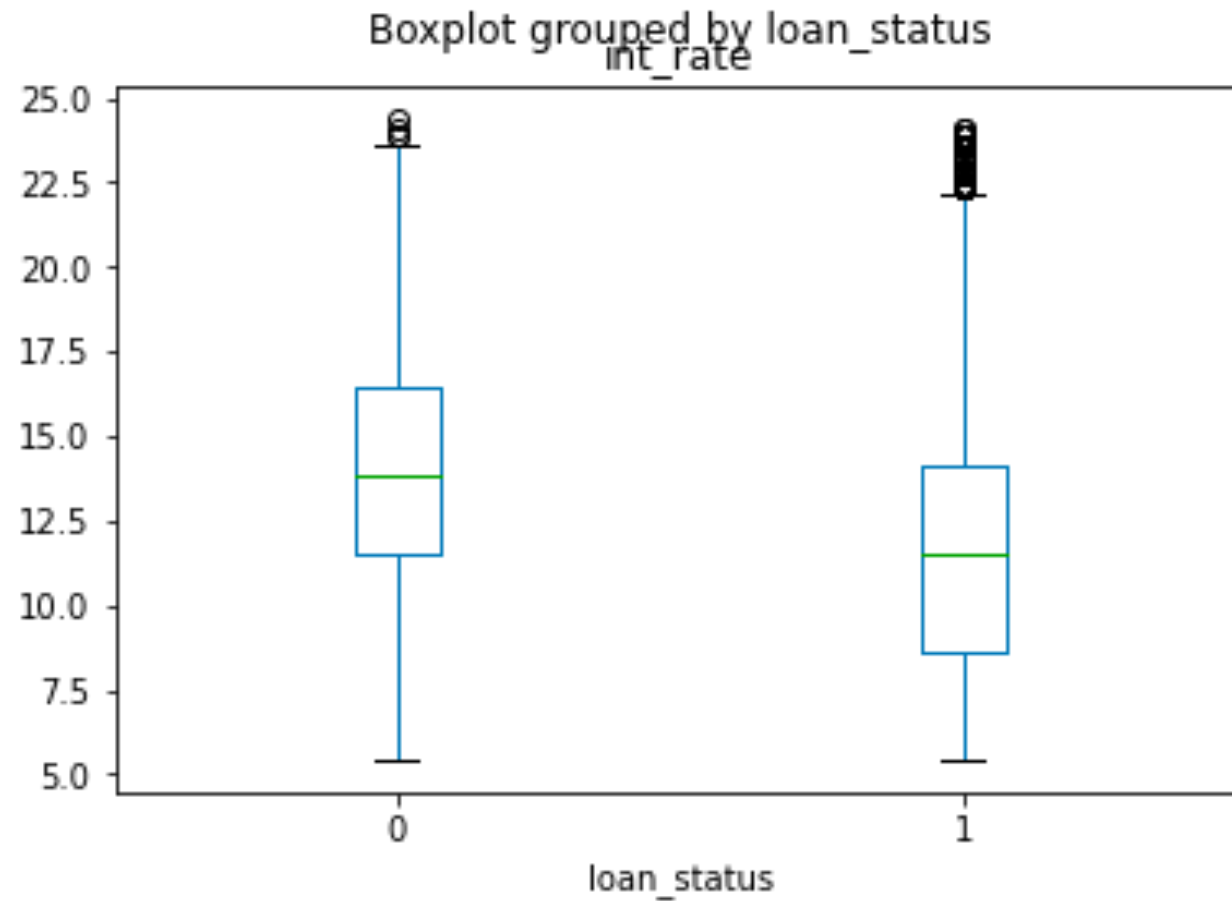Relook the verification process

Sub Grade

Observation:
- Sub grade from D1 onwards more defaulter.

Suggestion:
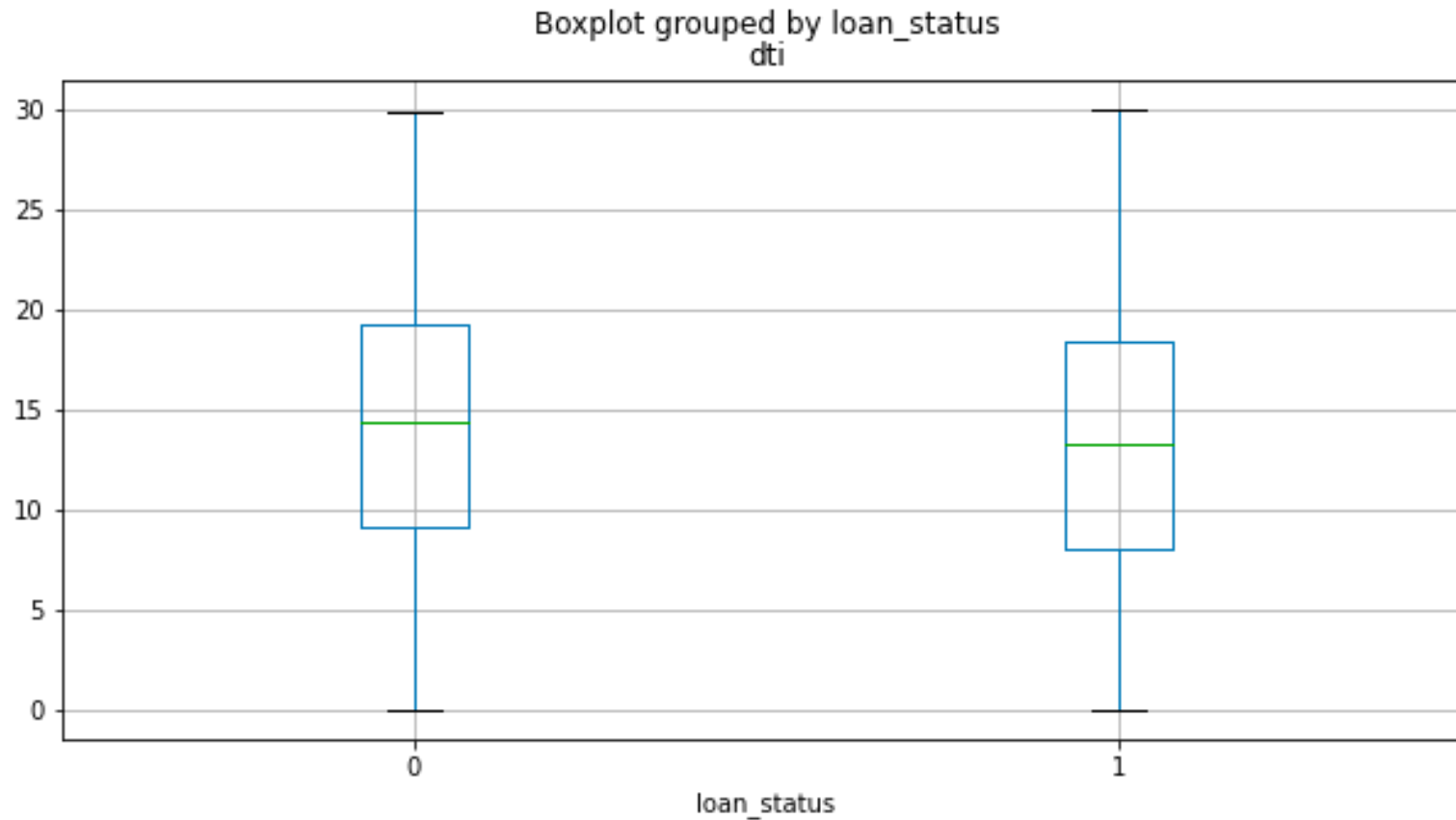- Avoid to approve sub grade after D1 onwards as it likely to defaulter

Loan interest rate

Boxplot grouped by loan_status
int_rate

Observation:
If rate of interest is more than 14%, there will be higher chance of defaulter
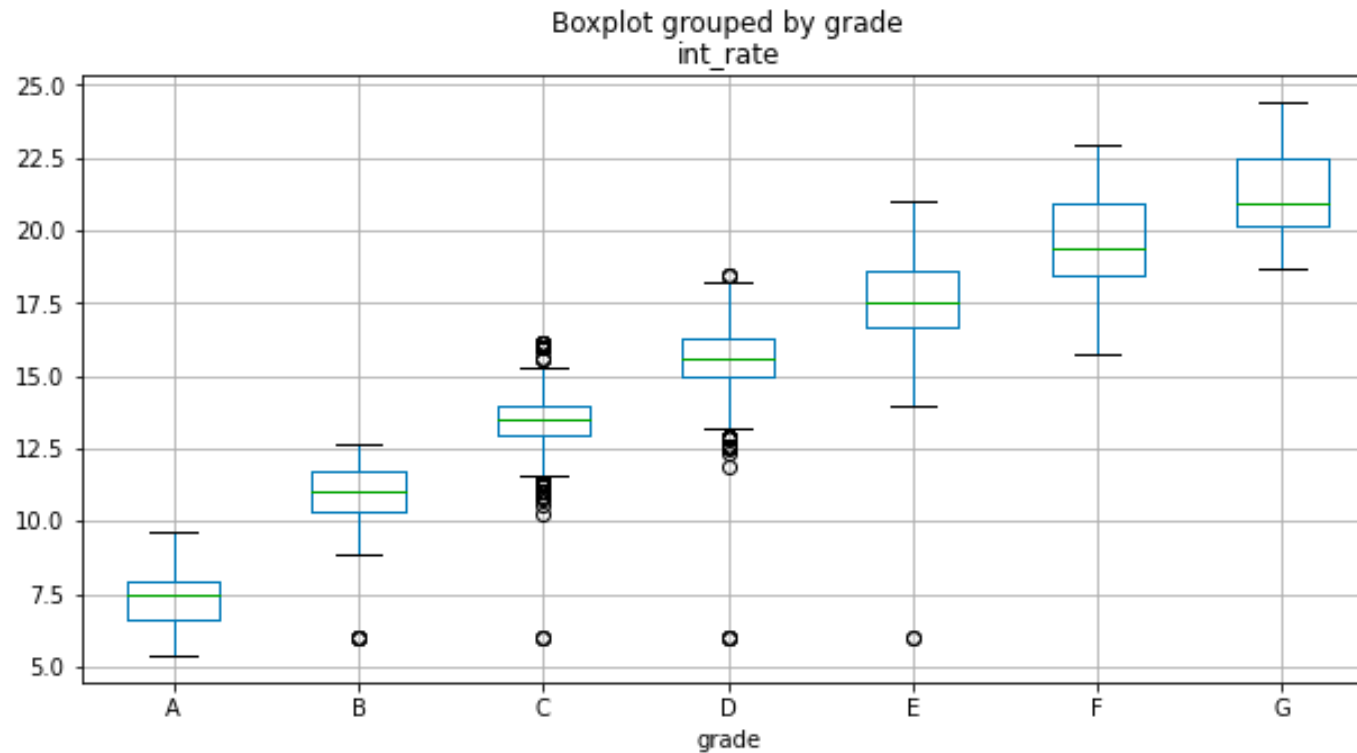
Debt to Income

Boxplot grouped by loan_status — dti

Observation:

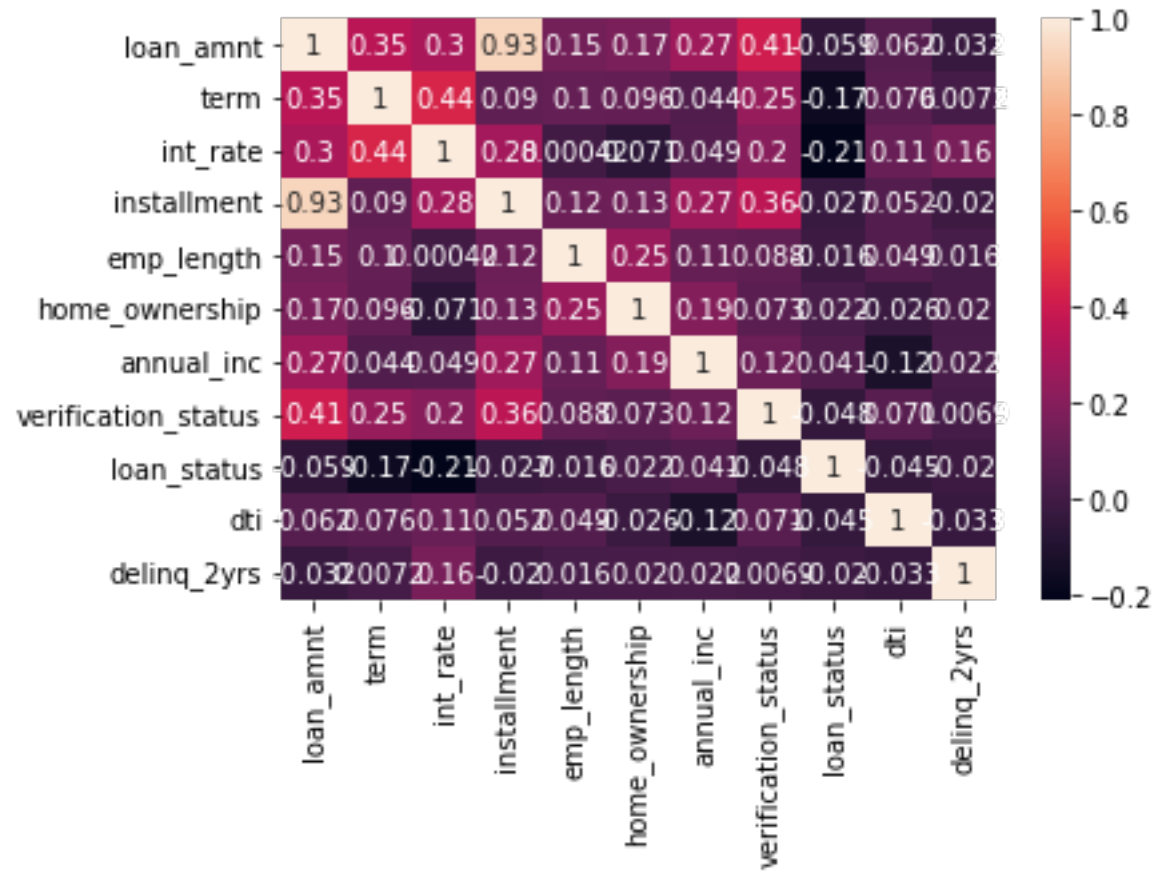If dti is more than 18% then there will be highly chance of defaulter

Interest rate with subgrade

Observation:
grade D onwards , there will be more higher rate of interest

Correlation

Observation:
Higher loan amount is higher chance of defaulter
Higher interest rate is also likely to be defaulter

Suggestion

Stop Approving higher loan amount > 35K

Start to retrospect validating process

Stop Approving loan higher dti >18

Avoid to approve sub grade after D1 onwards as it likely to defaulter

Emi amount is corelated with Amount funded

Stop Approving the loan without home ownership specified