



Lending Club

Case Study

Sujay N and Syed Towhid

iiitB and UpGrad Executive PG Programme
Machine Learning & Artificial Intelligence

Background – Lending Club

Lending club is the largest peer-to-peer marketplace connecting borrowers with lenders. Borrowers apply through an online platform where they are assigned an internal score. Lenders participating in the club are the primary source of finance and they decide the following:

- 1) Whether to lend money?
- 2) The terms of loan such as
 - 1) Interest Rate
 - 2) Monthly Instalment
 - 3) Tenure
 - 4) Any Other Term

Some popular products of lending club are:

- Credit card loans
- Debt consolidation loans
- House loans
- Car loans

Case Study Objective:

To identify variables which are strong indicators of applicants to be defaulters and potentially use the insights in approval / rejection decision making.

Case Study - EDA Overview

THE ANALYSIS FRAMEWORK AND APPROACH:

1. Data understanding

- Import the dataset
- Review the data --dataset head
- Review the columns
- Shortlist variables of interest

2. Data cleaning

- cleaning missing values
- removing redundant columns

3. Data Analysis

4. Recommendations

Data Understanding

Types of variables

The objective is to identify predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan. Now, there are broadly three types of variables –

1. Applicant associated

variables [demographic variables such as age, occupation, employment details etc.]

2. Loan characteristics

variables [amount of loan, interest rate, purpose of loan etc.]

3. Customer Behaviour*

variables [generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.]

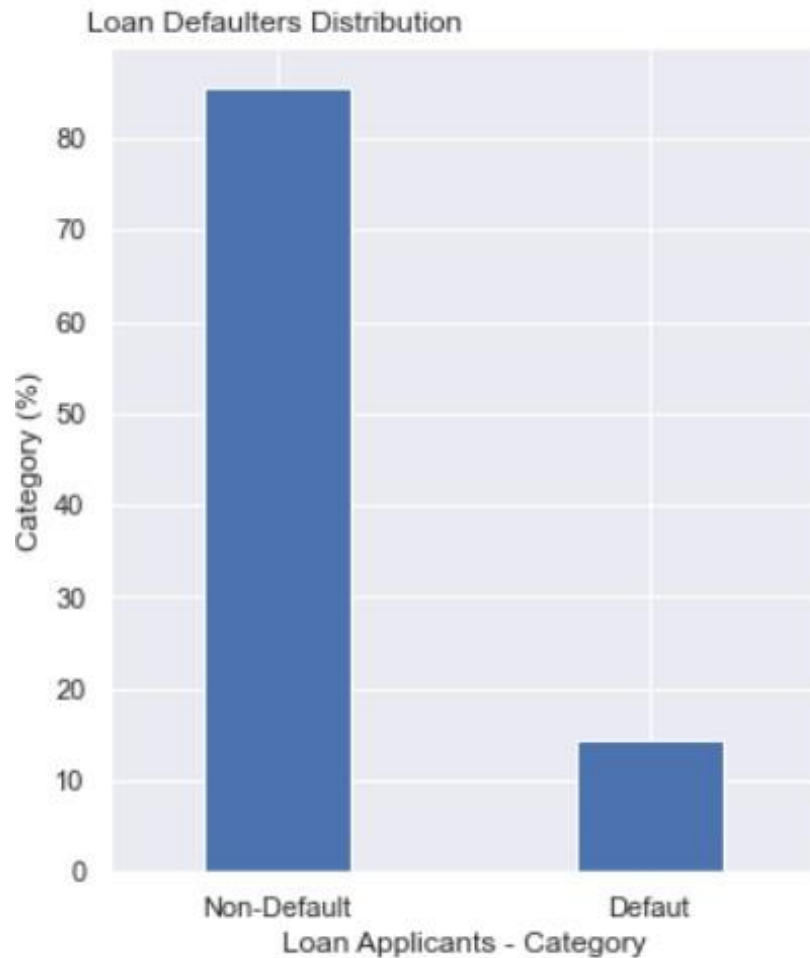
* The customer behaviour variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval. Thus, going forward, we will use only the other two types of variables.

Loan Information and Characteristics
Loan Amount
Funded Amount
Interest Rate
Loan Status
Loan Grade and Sub-Grade
Funded Amount Investment

Customer Behavior Variables
Application Type
Loan Purpose
Revolving Balance
Delinquency Year-2
Earliest Credit line

Applicant Associate Variables
Annual Income
Emploment Length
Description
Zip Code
Employment Title
Address State

Default Rate Assessment

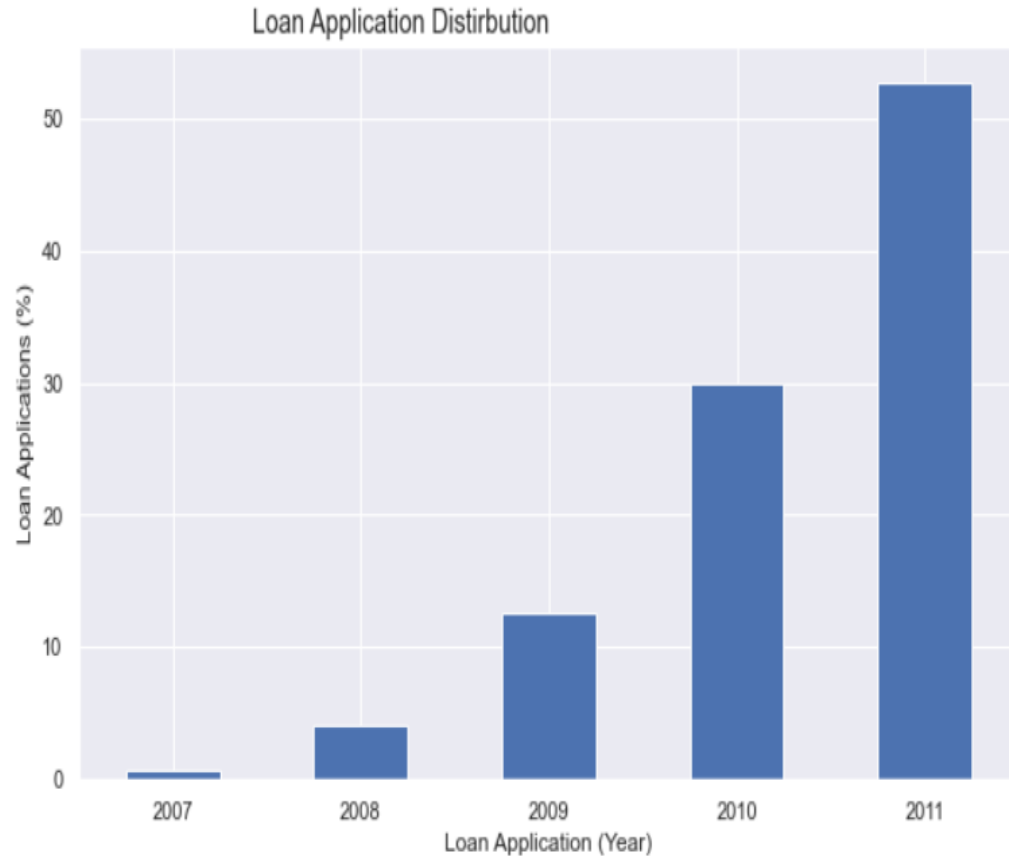


14%

Among the loan status which are Fully Paid or Charged-off, 14% of the applicant are defaulters.

Population: 5399 of 37544 records

Time Period (2007-2011) Assessment



Key Observations

78%

Increment in applicants in
2011 compared to 2010

145%

Increment in applicants in
2010 compared to 2009

Lending Club Portfolio

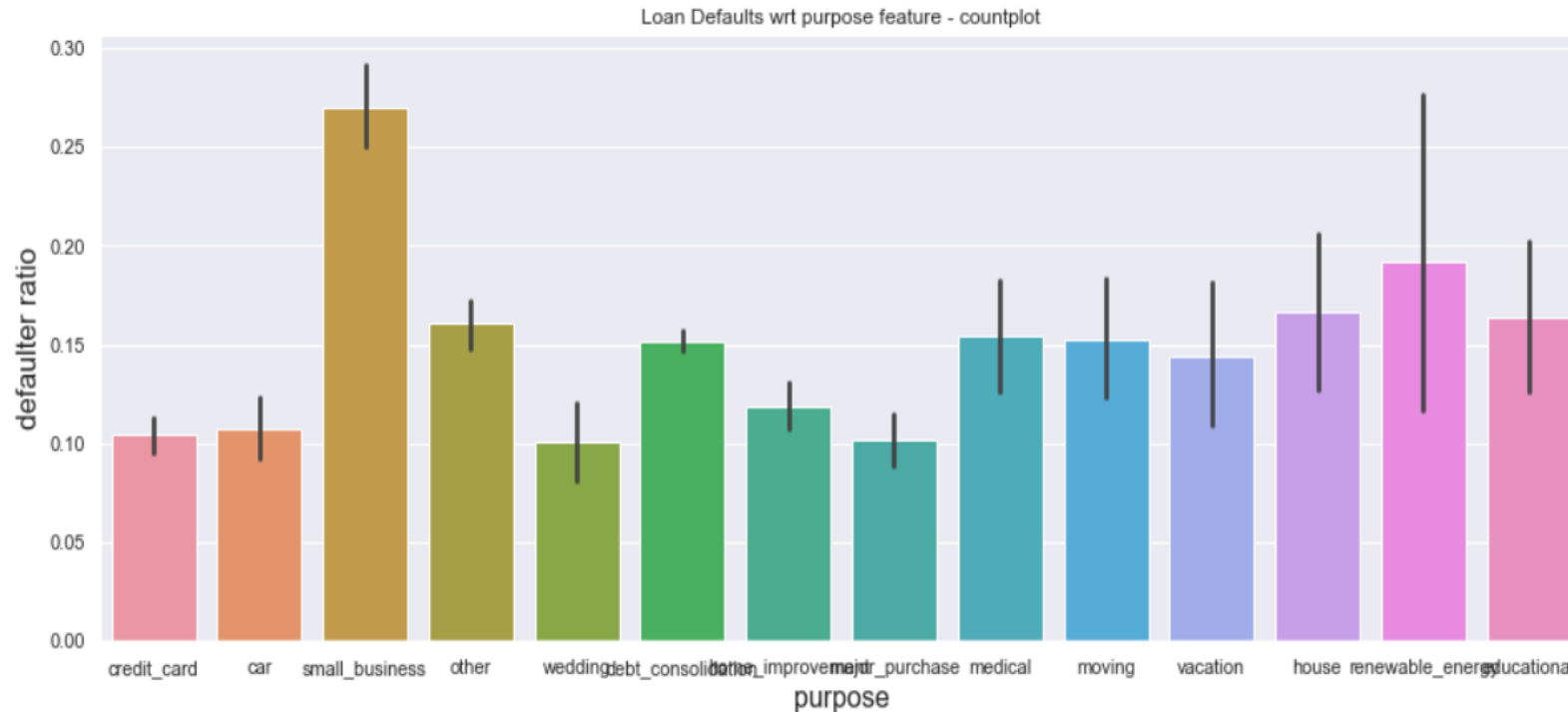


Key Observations

- 46.8% debt consolidation loans
- 13% credit card loans
- 7.5% home improvement loans
- 5.6% major purchase loans

Debt Consolidation Loan is the most popular product in Lending Club portfolio

Loan Purpose Vs Defaulter Ratio

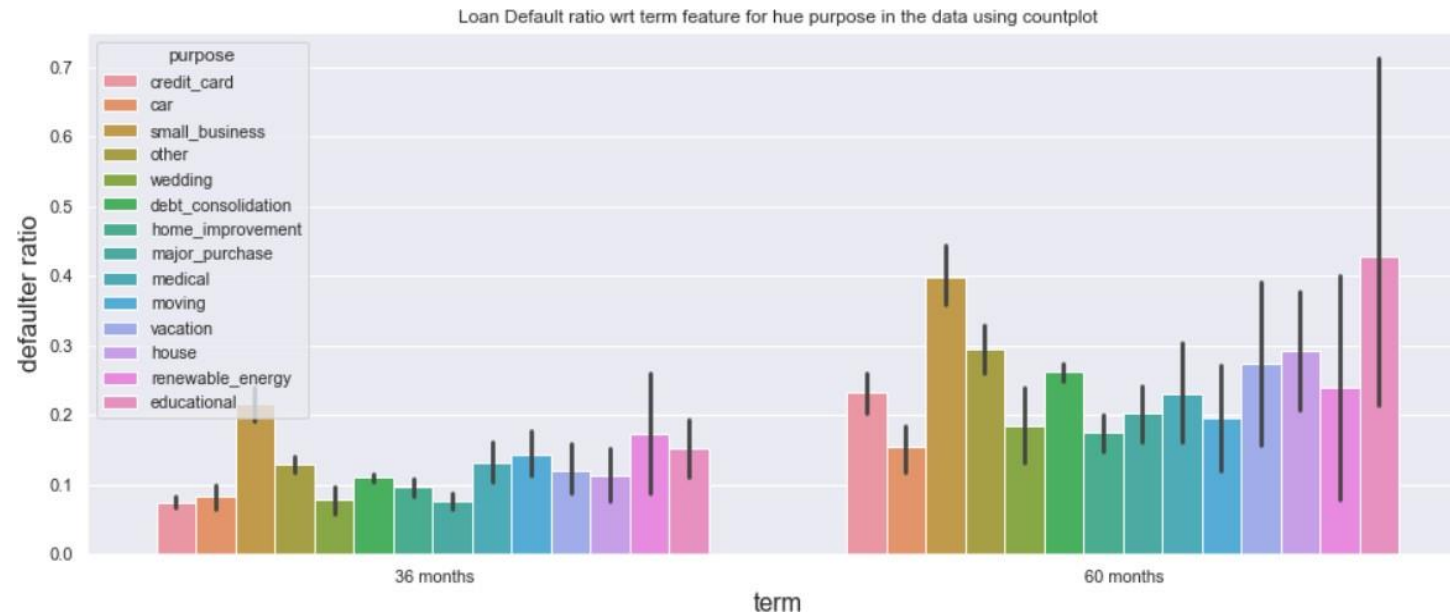


Key Observations

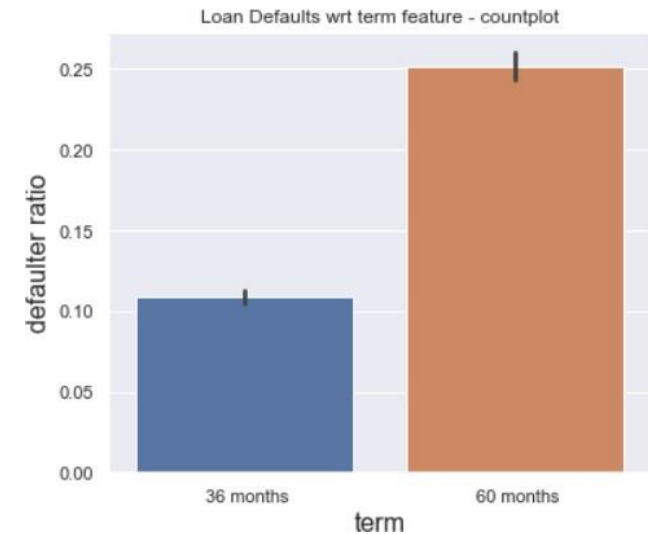
- Small business: 27%
- House: 17%
- Educational: 16%
- Debt consolidation: 15%
- Credit card: 10%
- Major purchase: 10%

Small Business are the highest defaulters.

Product Summary – Tenure

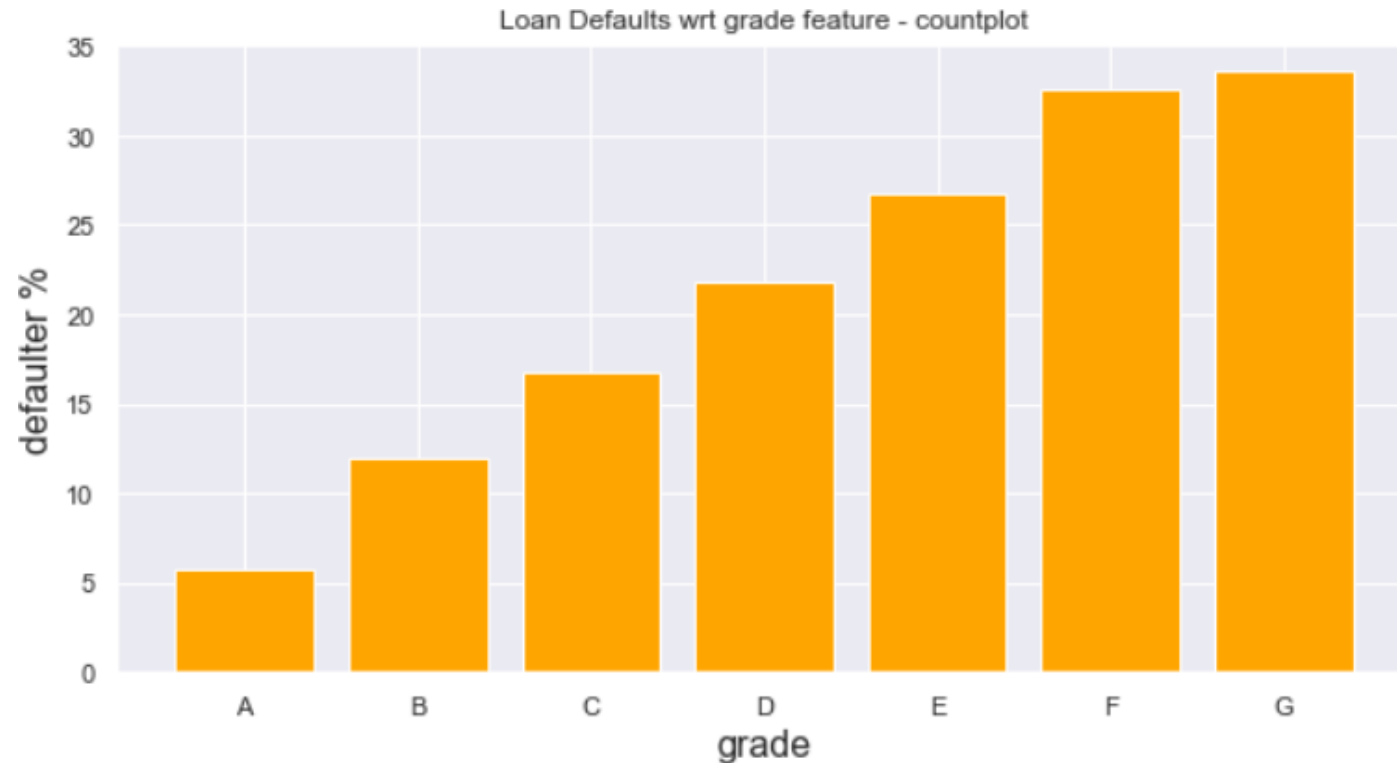


Clearly, we have more concentration on 60 months compared to 30 months. The Defaulter ratio is directly proportional to loan tenure in this segment where, 60-months have high default rate.



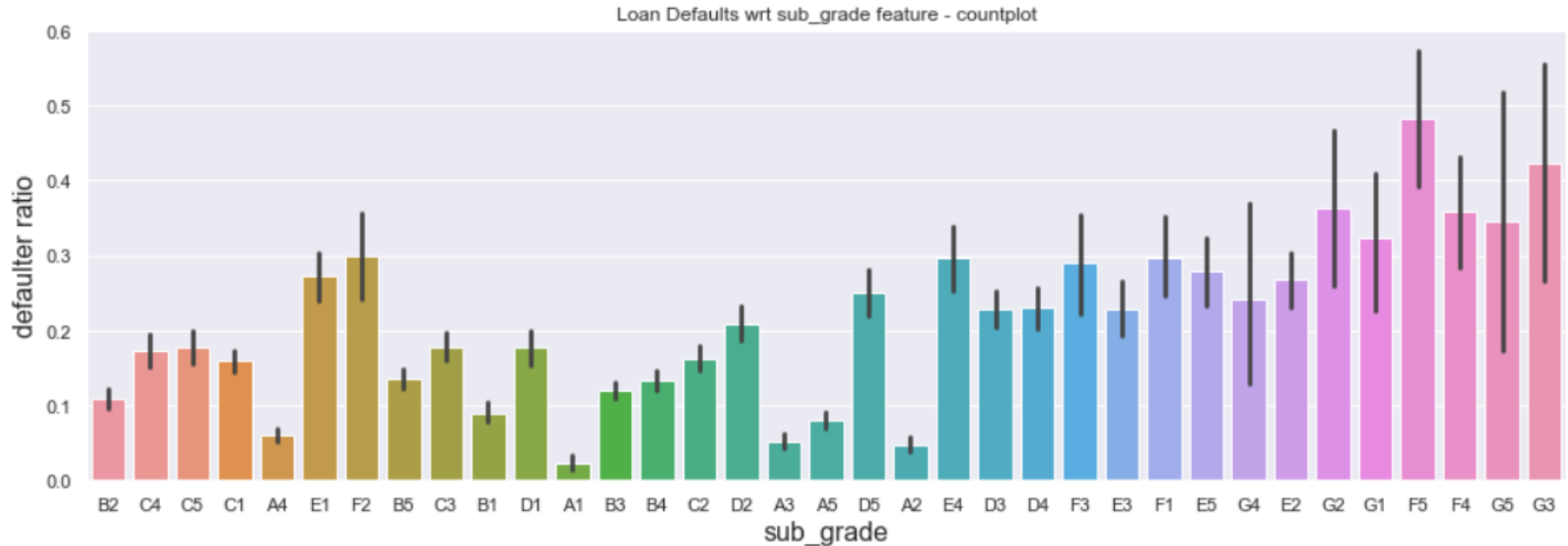
Lending club has given out 36 month term period loan for major purpose and home improvement.

Defaulters Summary – Grade



From the plot for 'grade' we can infer that the defaulters rate is increasing with respect to grade, hence the chances of loan getting defaulted increases with the grade from A moving towards G.

Product Summary – Sub-Grade



From the plot above for 'sub-grade' we can infer that the defaulters rate is increasing with respect to grade, hence the chances of loan getting defaulted increases with the grade from A moving towards G. This further confirms the associated with main grades such as A-G

Lending Club Portfolio – Top Products vs Average Interest Rate

Credit Card

11.62%

Debt Consolidation

12.04%

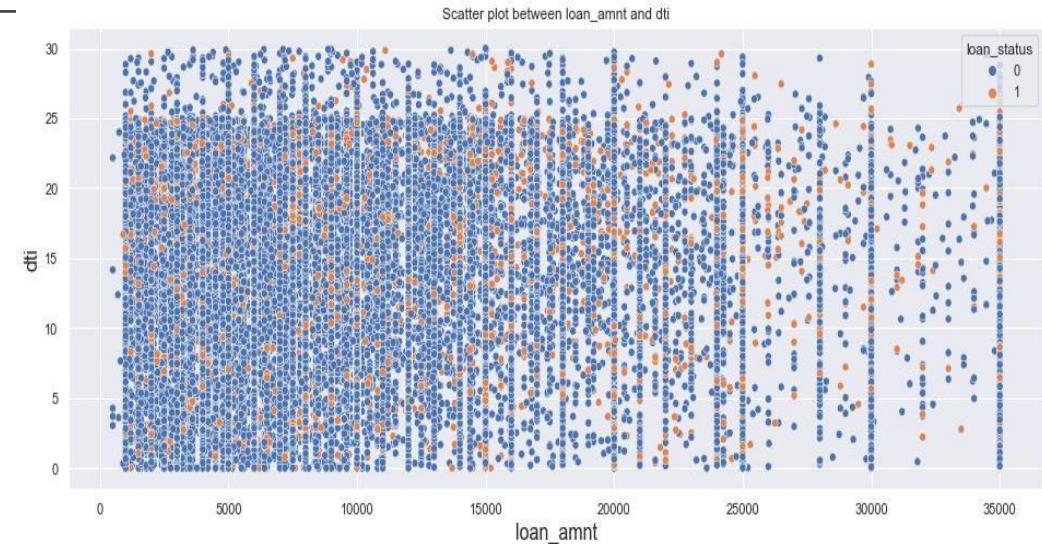
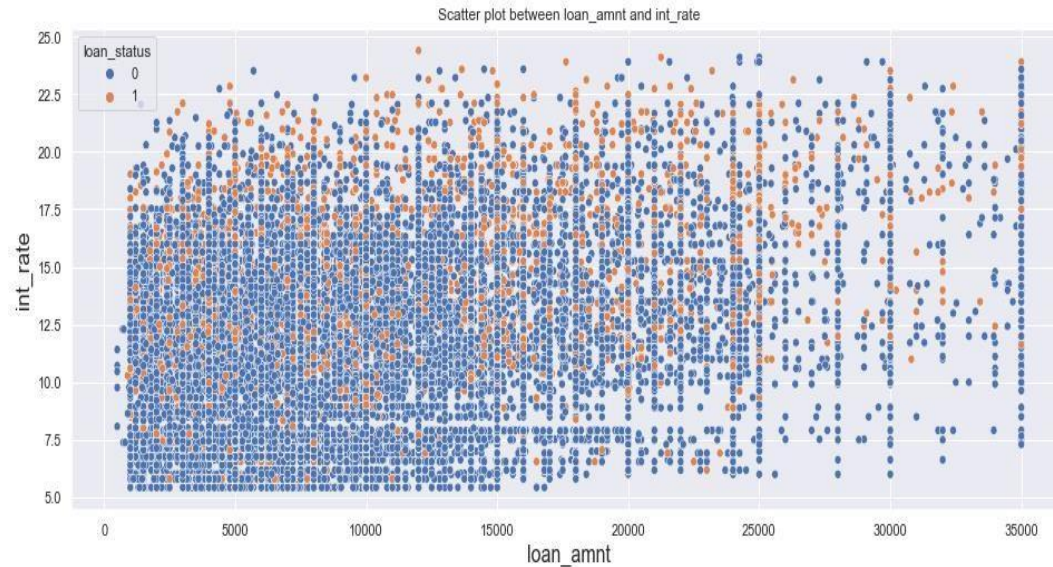
Home Improvement

11.29%

Major Purchase

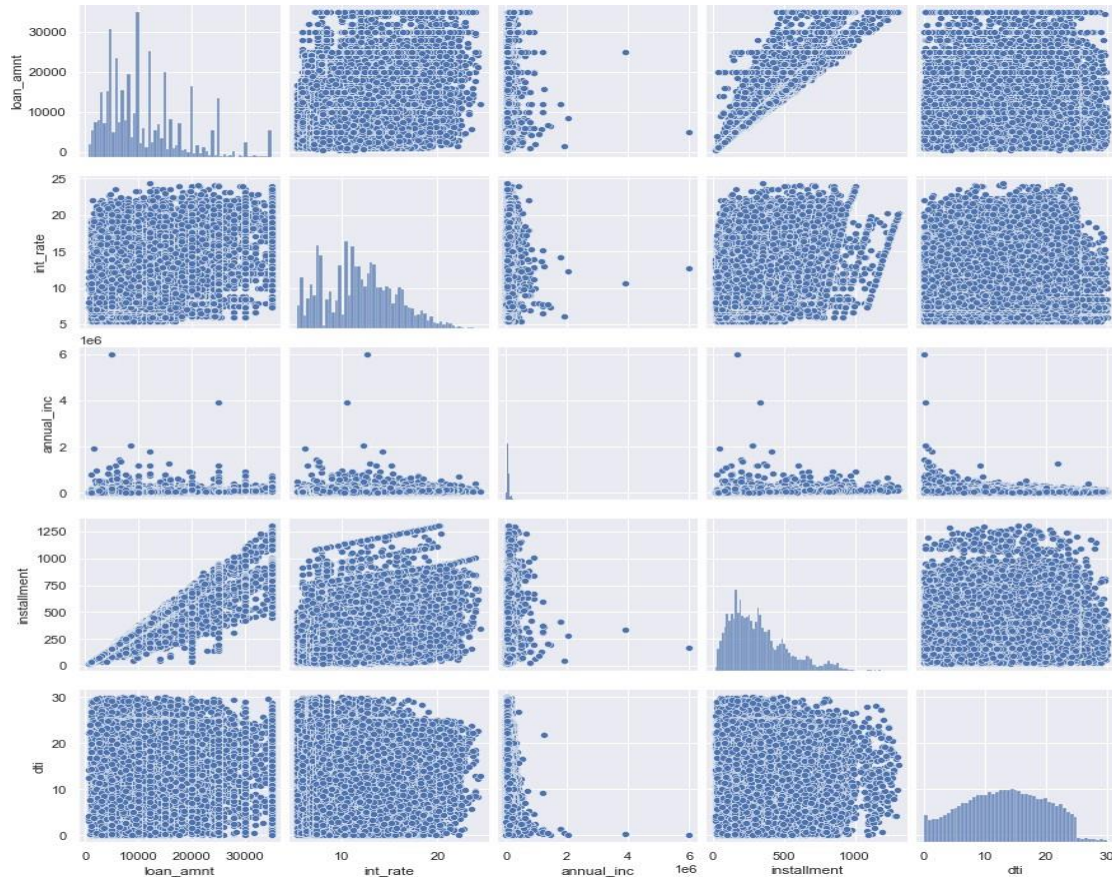
10.80%

Loan Amount in relation with Interest and DTI



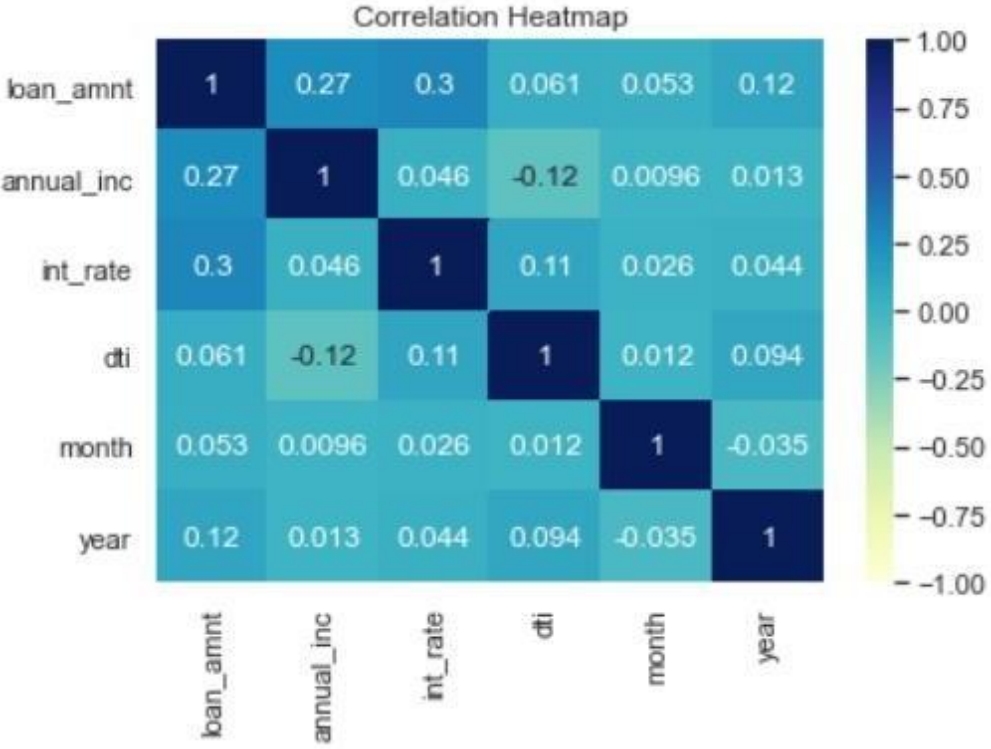
There a dense correlation between these factors. Lower loan amount tend to have a close affinity with Interest and Debt-To-Income (DTI) ration. As the Load amount increases the correlation is decreasing indicating the high loan rate is least opted.

Paired Analysis – key Indicators



We can see the correlation between these key indicator or influencing variables such as Loan_amount vs Instalment, Interest Rate Vs Instalment.

Key Indicators Correlation



Correlation Summary - Variables	
Term	
Grade	
Purpose	
Revolving Balance	
Interest Rate	
Installment	
Annual Income	
Loan Amount	

There is a strong correlation between the key indicators and thus, the factors influencing defaulters and decision making for loan approval are highly cohesive.

Conclusion

After analyzing all the related features available in the dataset, we have come to an end, deducing the main driving features for the Lending Club Loan Default analysis.

The best driving features for the Loan default analysis are: term, grade, purpose, revol_util, int_rate, installment, annual_inc, funded_amnt_inv