Lending Club Case Study

Case Study Submission by:

Aishee Ghosh

Chanchal Narang

8th June 2022

Index

- Understanding the Problem Statement
- Understanding the given dataset
- Data Cleaning and Transformation
- Data Analysis
- Conclusions Drawn

Understanding the Problem Statement

- Lending Club is a consumer finance company which specializes in lending various types of loans to urban customers.

- The objective is to identify who are the risky loan applicants and what are the factors weighs in for them to be the risky ones.

- How can we save the lender from money loss and do the risk assessment for the same.

- Now we have the loan dataset which contains the complete loan data for all loans issued through the time period 2007 t0 2011.

Understanding the given Dataset provided with Problem Statement

Below are the columns which can be important indicators for the Analysis,

- annual_inc =>this will decide eligible for loan or not.

- Application_type(individual)=>Individual loan borrowers

- avg_cur_bal=> important factor how much balance he has whether he will be able to pay EMI or not.

- chargeoff_within_12_mths=> how many times he/she is defaulter in last 12 months, this will tell us about his/her behavior.

- delinq_2yrs=>Defaulter for how much time. Again tells about borrower behavior.

- delinq_amnt=> for how much amount he /she is delinquent ,will help us to decide it a risky one or not .

- emp_length=> he/she has continuous source of income or not.

# Data Cleaning and Transformation

- Checking the percentage of null values in each column. We got 100 percentage of null values in few column.

- There are 3 types of loan status & we are removing the "loan_status=current"the "current" ones are already approved.

- Checked and Changed the datatype of columns as per analysis's requirement.

- Removed the null values from the data set.

Data Analysis: The categorical and continuous

- Now from the clean dataset we need to segregate Categorical variables/columns which are finite & in an order and Continuous variables/columns which are infinites .

- Categorical Variables are:
  annual_inc,application_type,avg_cur_bal,delinq_2yrs,delinq_amnt,loan_status,

  term,int_rate,grade,issue_d,sub_grade,dti,home_ownership,verification_status

- Continuous Variables are: emp_length, loan_amnt, annual_inc

- Now we need to run Univariate , Segmented Univariate , Bi Variate analysis, Multivariate analysis on these columns to analyze the data further.
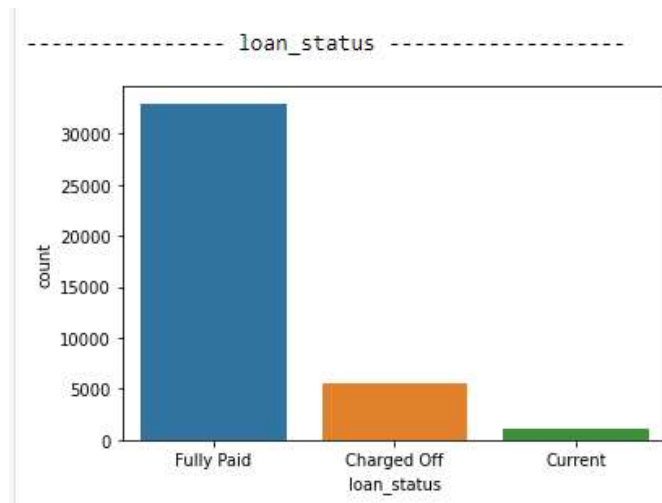
Data Analysis: Univariate Analysis

- Fetch the Statistical data from the Clean Data Set using describe().
- Identified the Outliers as shown in the below box plots. Few loan amounts are going beyond USD29000. Also very few people has huge income.
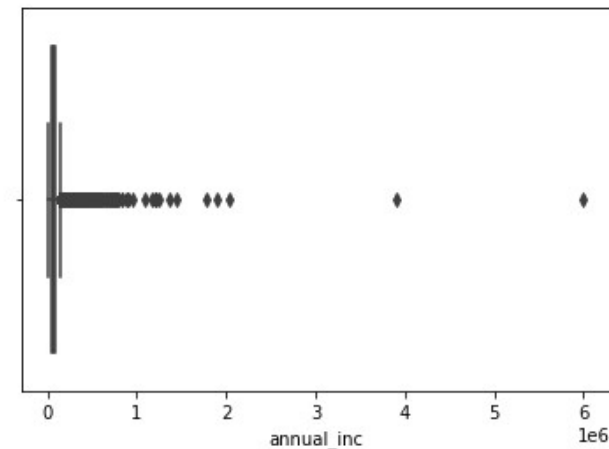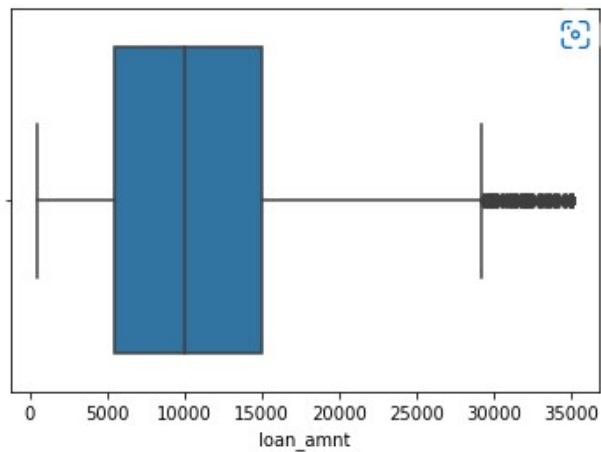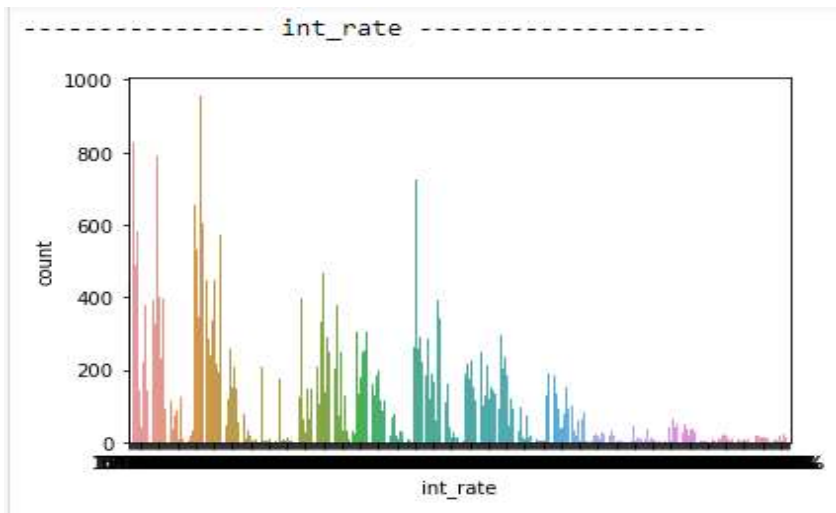
Data Analysis:

- Maximum number of loan takers has already paid their loans
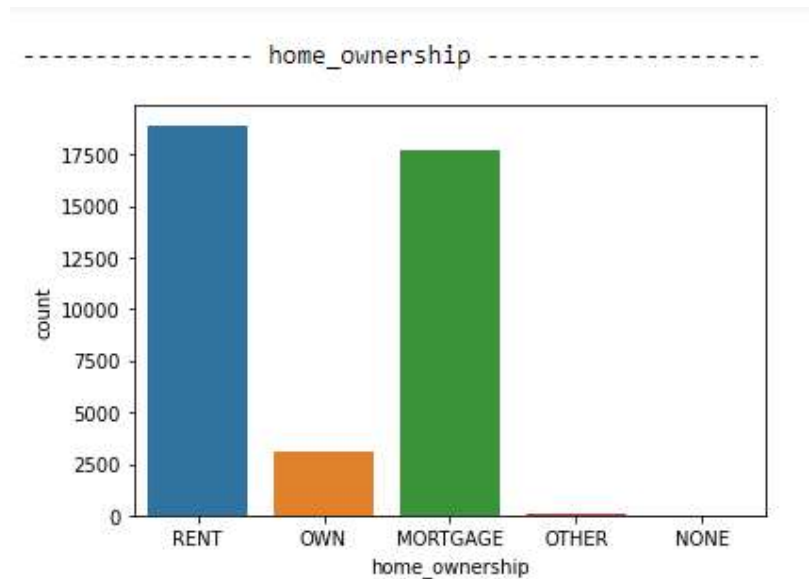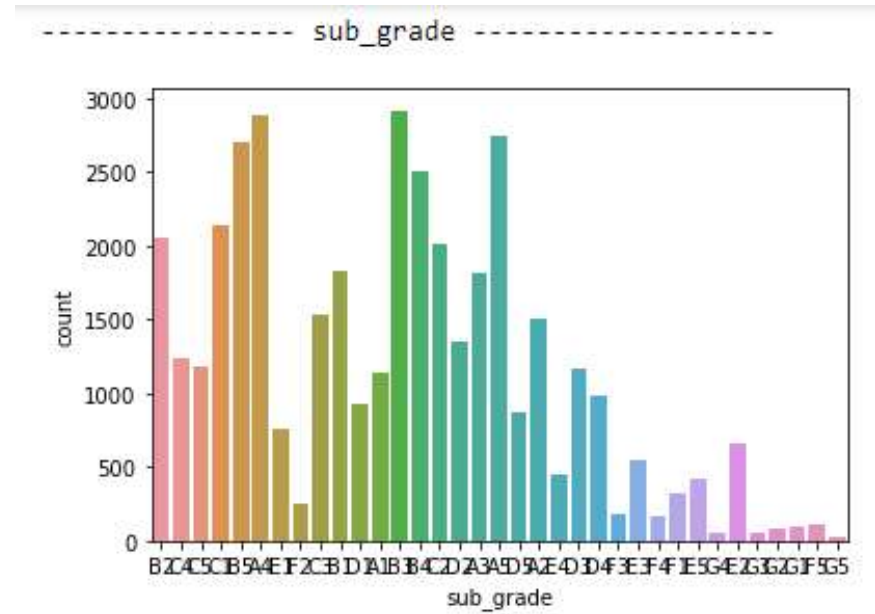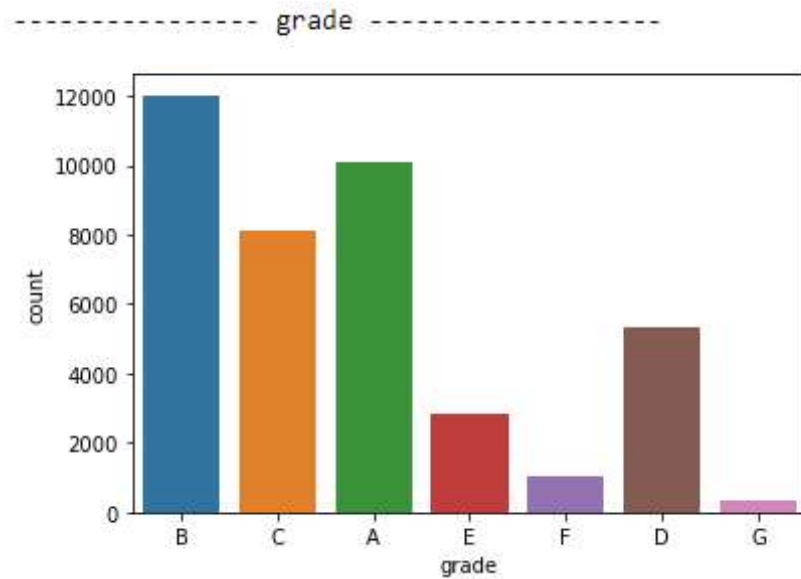- More loan takers are opting for 36 months term.

Data Analysis

There are majority of the loans which are of Low Interests.
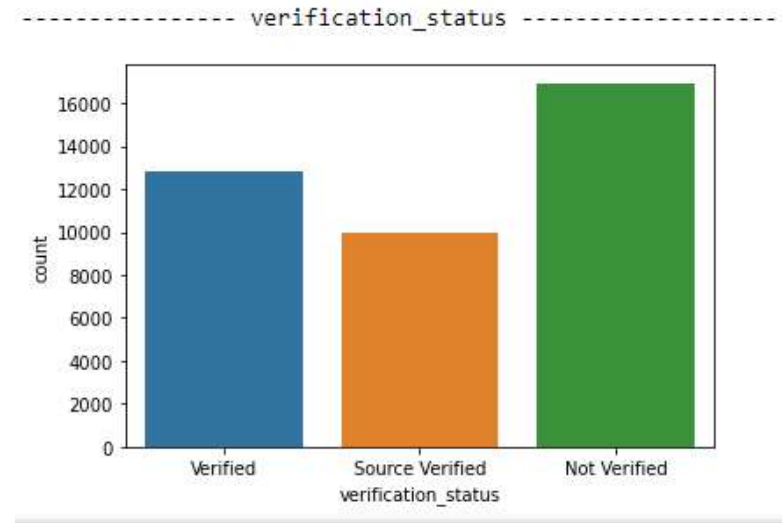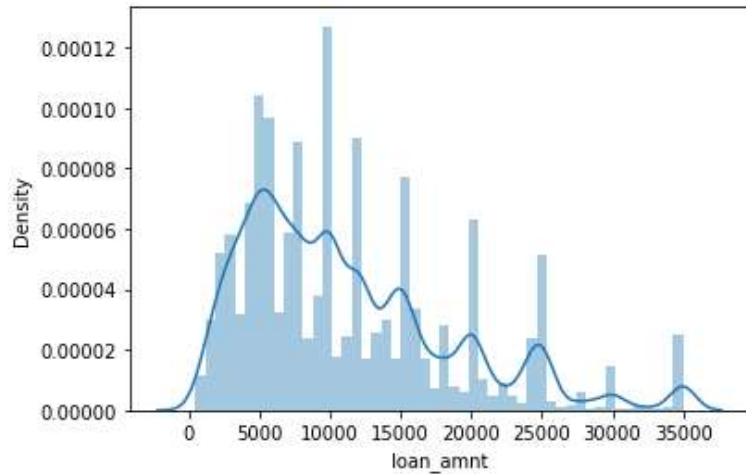

Majority of the Loan Takers lives under Rent.

Data Analysis

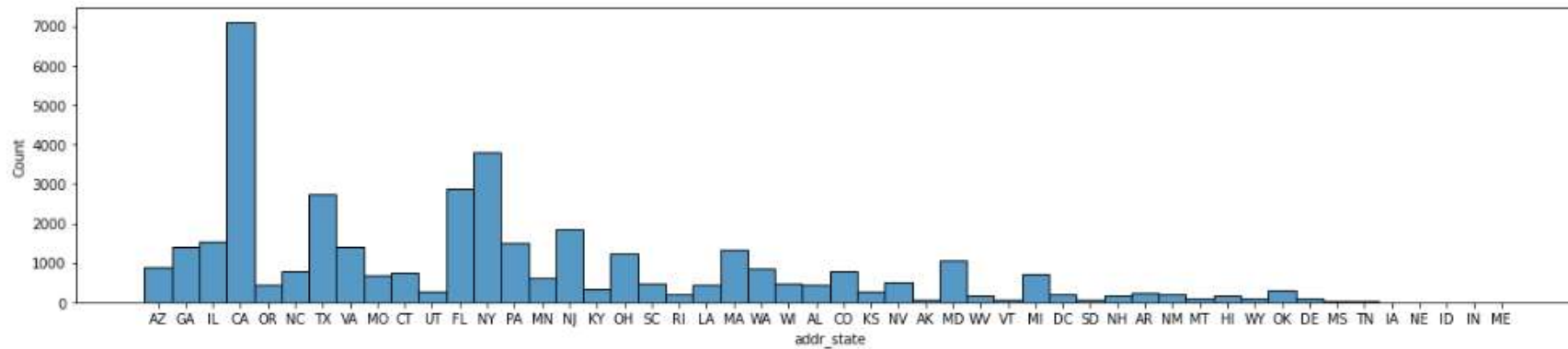**B grade loans are more popular among loan takers! Sub_grade B3 type of loan is topping the charts!**
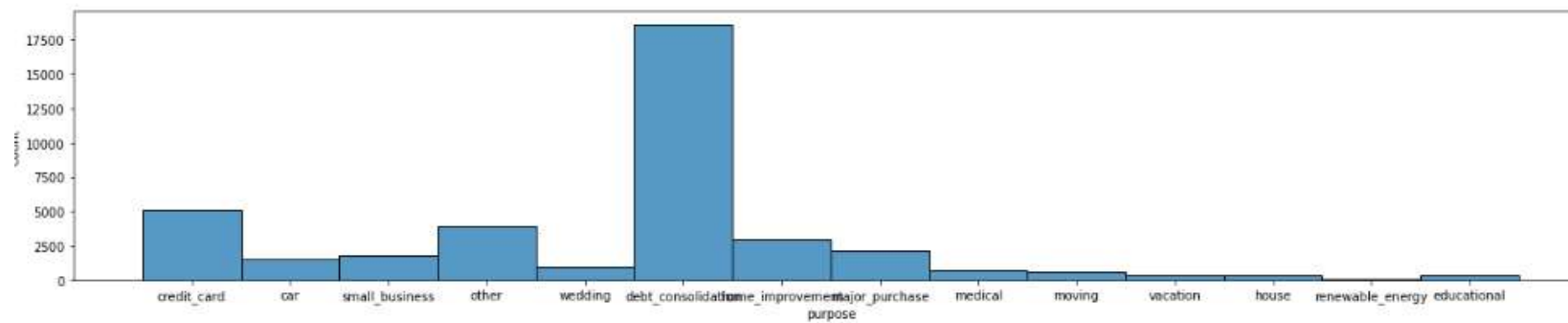
Data Analysis

Majority of the loan takers income is not verified by Lending Club. This is major flaw which needs to be rectified if LC wants to avoid risky loan takers.

Probability Distribution of the Loan Amount. Looks like majority has opted to apply a loan for USD $10000
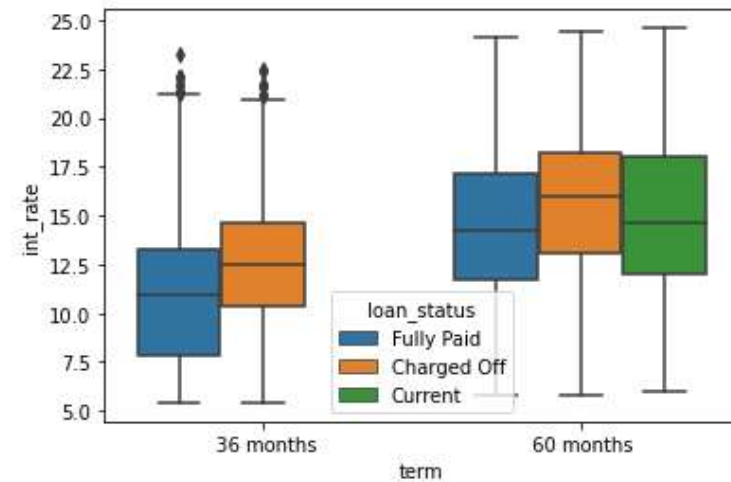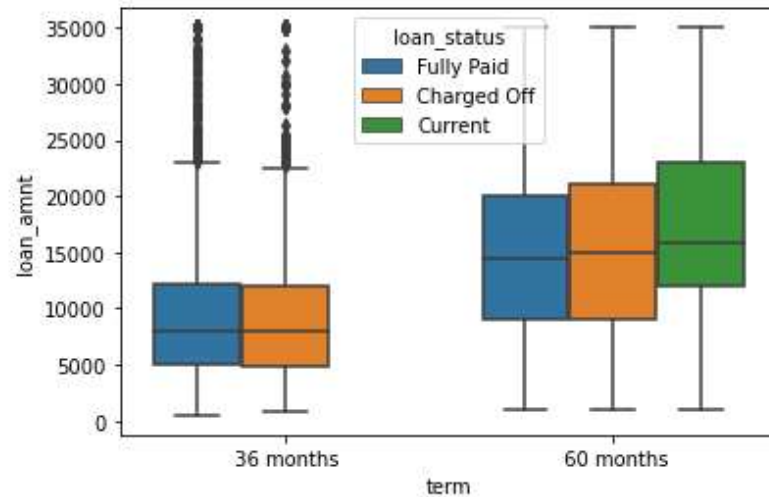
This has been observed that the most of the borrowers are from CA and NY states.

Most people took loan for Debt Consolidation and very less people took loan for renewable energy.
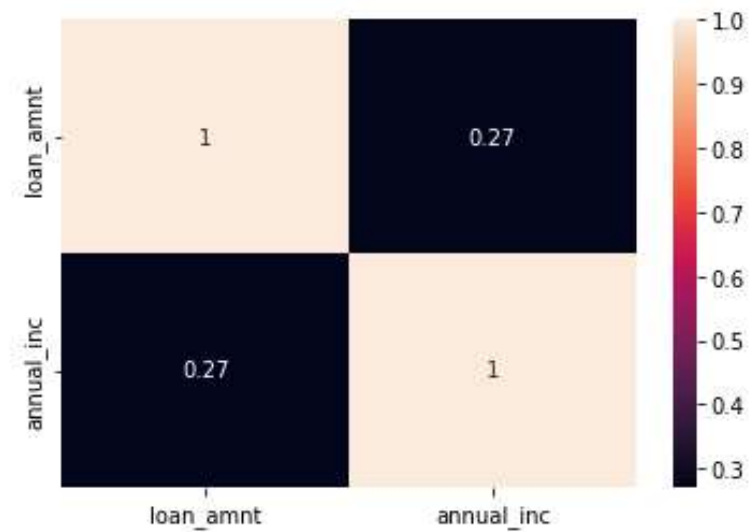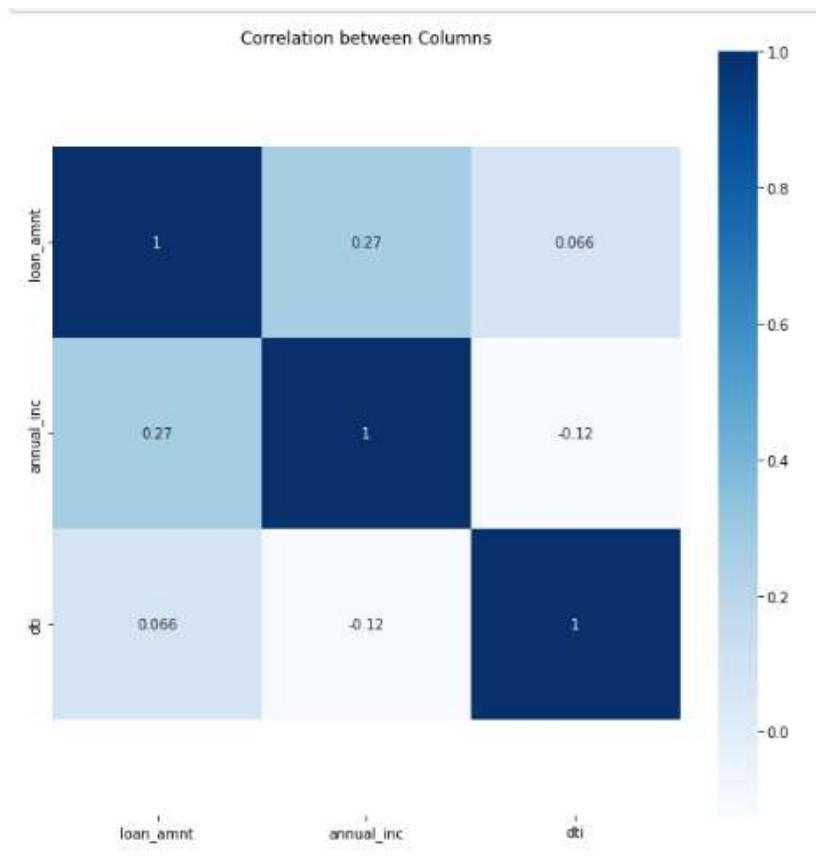
Data Analysis

The defaulters rate is high in 60 months tenure. Looks like most people who took high loan amount in high interest rate have become defaulters and facing it hard to return back the loan.

Data Analysis

From this heatmap we observe there is a overall positive correlation between loan_amount & annual_inc.

Correlation between Columns

A heatmap to understand the Correlation between the columns of the data set
##The correlation between the loan_amnt & annual_inc is positive
##The correlation between the loan_amount &dti is positive
##The correlation between the annual_inc & dti is negative.

Conclusion & Recommendations:

So Based on all the previous graph we can conclude these following. If Lending Club Loan Company can make below necessary changes , it might help them to reduce defaulters and avoid financial loss to the company.

1. Lending Club should perform a thorough KYC about their Borrowers income, Identity validation before granting them Loan. This will reduce the risk of Company loosing on Money.

2. Apart from NY, NJ, FL, CL , in other states very few people are interested in opting for a loan. Need to research more on those states and take action accordingly to increase business across USA.
3. The Offers & the Policies for F grade and G grade loans needs to be reviewed and make it more lucrative to the borrowers.. These loan types are the least ones to be opted.
4. Introduce more options with respect to loan terms and a floating interest rate should definitely attract more genuine borrowers.