



Lending Club Case Study: Assignment

Course : Machine Learning

Lecture On : Lending Club Case Study

Facilitator : Vignesh G

Agenda

- Problem Statement
- Charts and Explanations

What is Lending Club?

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

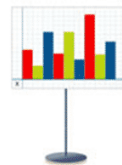
How Lending Club Works



Borrowers apply for loans.
Investors open an account.



Borrowers get funded.
Investors build a portfolio.



Borrowers repay automatically.
Investors earn & reinvest.

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:

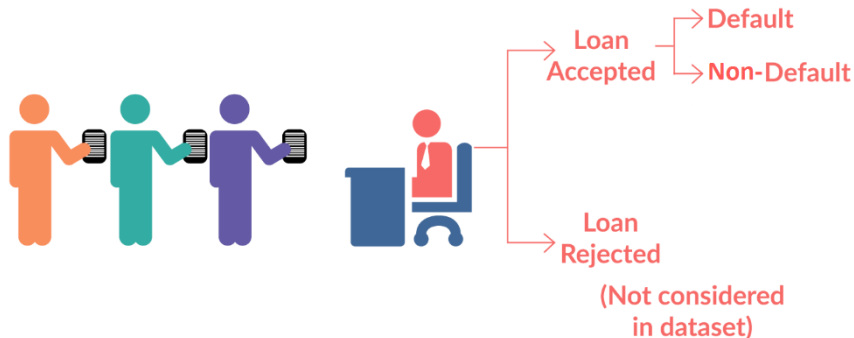
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, the 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 utilize this knowledge for its portfolio and risk assessment.



LOAN DATASET



Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

There are four major parts that are needed to be done for this case study:

1. Data understanding
2. Data cleaning (cleaning missing values, removing redundant columns etc.)
3. Data Analysis

Data Cleaning

1. Check the percentage of missing values
2. Remove all those with very high missing percentage
3. For columns with less missing percentage: perform Imputations
 - You don't need to impute the data, you can just identify the correct metric to impute the column.
4. You can drop rows where the missing percentage is quite high

Steps Followed for Lending Club Case Study

- Importing all necessary libraries
- Reading loan.csv file
- Finding out the data structure
- Looking at the split of loan_status
- Adding a New Column Count with 1 in every row for the sake of Using Pivot table
- Finding out outliers in the data

- Using quantile to differentiate annual_inc. After 99 %ile annual income doesn't look to be following a even pattern.
- Filtering out very high income group.
- Splitting Issue date into Month and Year.
- Adding bins for int_rate, annual_inc,funded_amt_inv,loan_amt,dti, revol_util and open_acc.

PEOPLE WITH PURPOSE OF DEBT_CONSOLIDATION MORE LIKELY TO DEFAULT.

loan_status purpose	Charged Off	Current	Fully Paid
car	160.0	50.0	1339.0
credit_card	542.0	103.0	4485.0
debt_consolidation	2767.0	586.0	15288.0
educational	56.0	NaN	269.0
home_improvement	347.0	101.0	2528.0
house	59.0	14.0	308.0
major_purchase	222.0	37.0	1928.0
medical	106.0	12.0	575.0
moving	92.0	7.0	484.0
other	633.0	128.0	3232.0
renewable_energy	19.0	1.0	83.0
small_business	475.0	74.0	1279.0
vacation	53.0	6.0	322.0
wedding	96.0	21.0	830.0

CHARGED_OFF ON AVERAGE HAVE LOWER INCOME THAN FULLY_PAID APPLICANTS

loan_status emp_length	Charged Off	Current	Fully Paid
1 year	53592.422593	65844.550725	60681.941733
10+ years	68831.621915	81631.058394	77162.507199
2 years	53326.943173	71128.679381	60880.642978
3 years	57852.321996	69227.915854	64110.563193
4 years	59334.793224	65441.585106	65156.208822
5 years	62239.914825	72236.620230	65552.779097
6 years	63372.017705	64752.724138	66201.516506
7 years	61668.907939	67533.018710	66966.160195
8 years	67778.180099	66729.209302	70908.400806
9 years	65846.839236	73355.881250	71877.011323
< 1 year	52581.045906	70987.216216	59637.499069

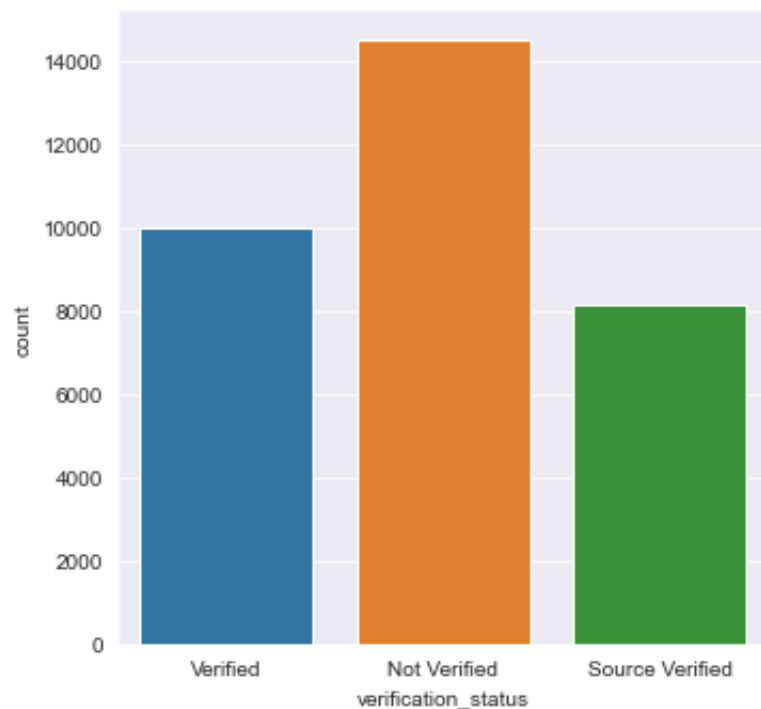
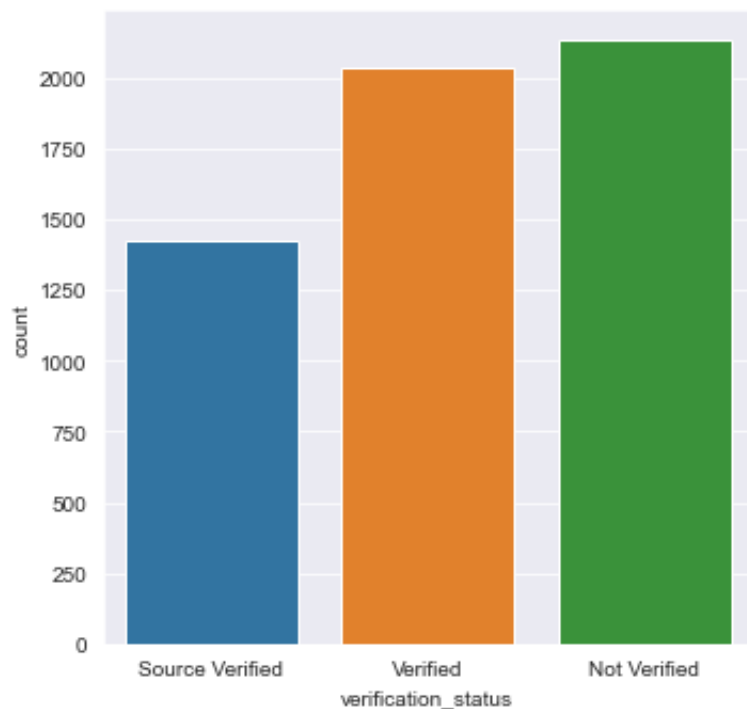
INTEREST_RATE FOR CHARGED_OFF ARE HIGHER ON AVERAGE

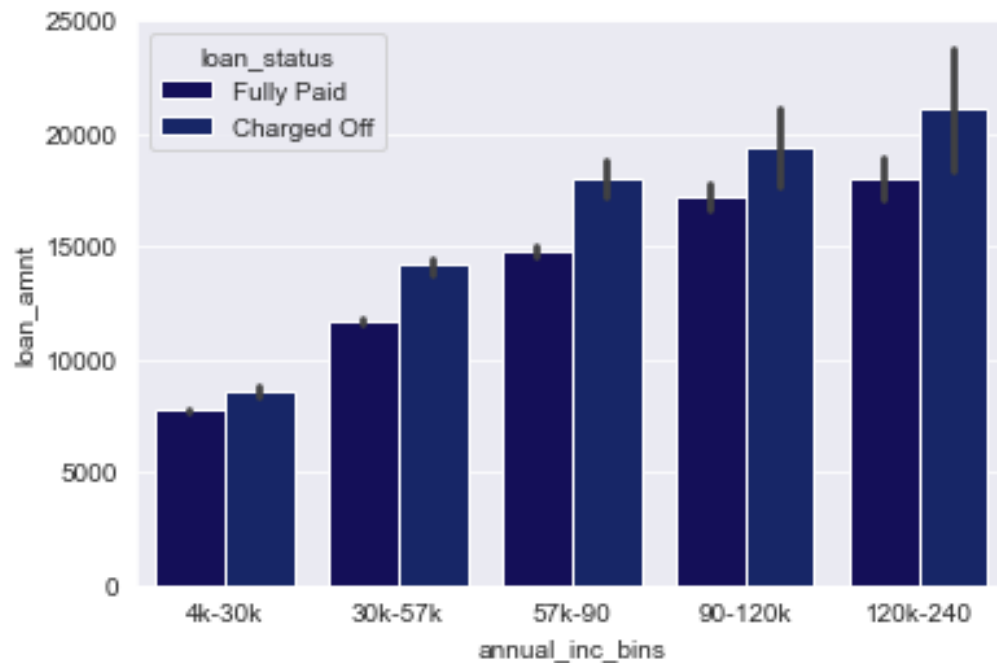
loan_status emp_length	Charged Off	Current	Fully Paid
1 year	13.652813	14.862899	11.707453
10+ years	14.063032	14.957150	11.546196
2 years	13.821283	15.419897	11.693531
3 years	13.921924	15.016951	11.625476
4 years	13.950329	15.287128	11.772850
5 years	13.850044	14.503103	11.645074
6 years	13.978033	15.805517	11.607914
7 years	14.120191	14.592742	11.718146
8 years	13.753793	14.857674	11.469276
9 years	13.765541	14.943750	11.548299
< 1 year	13.484614	15.820676	11.570010

PERCENTAGE OF LONG TERM LOAN ARE HIGHER FOR CHARGED OFF APPLICANT

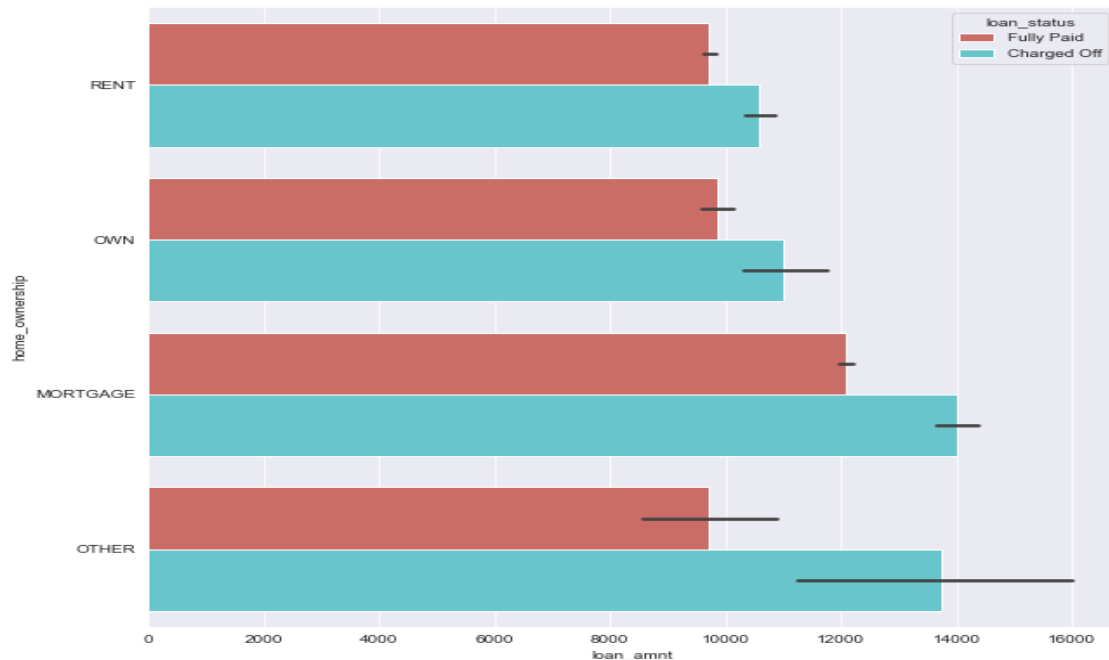
loan_status term	Charged Off	Current	Fully Paid
36 months	3227.0	NaN	25869.0
60 months	2400.0	1140.0	7081.0

CHARGED_OFF APPLICANTS HAVE HIGHER PERCENTAGE OF NOT_VERIFIED STATUS

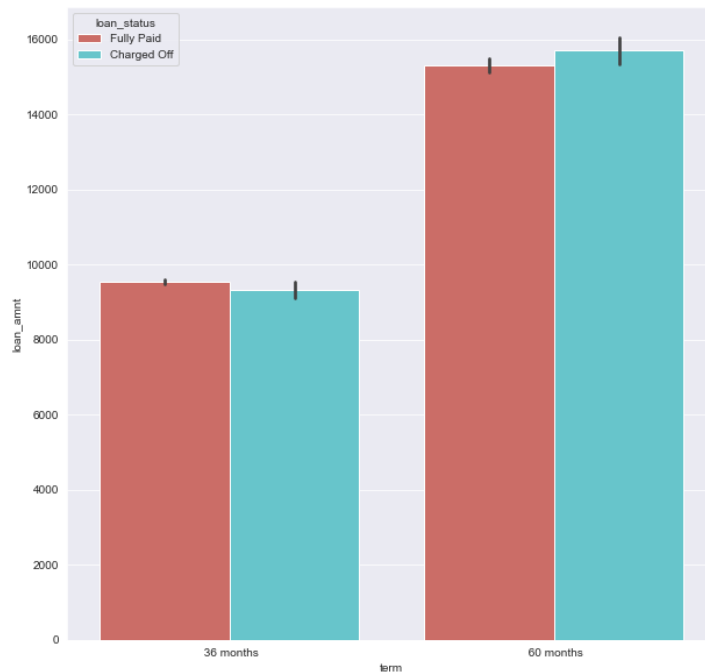




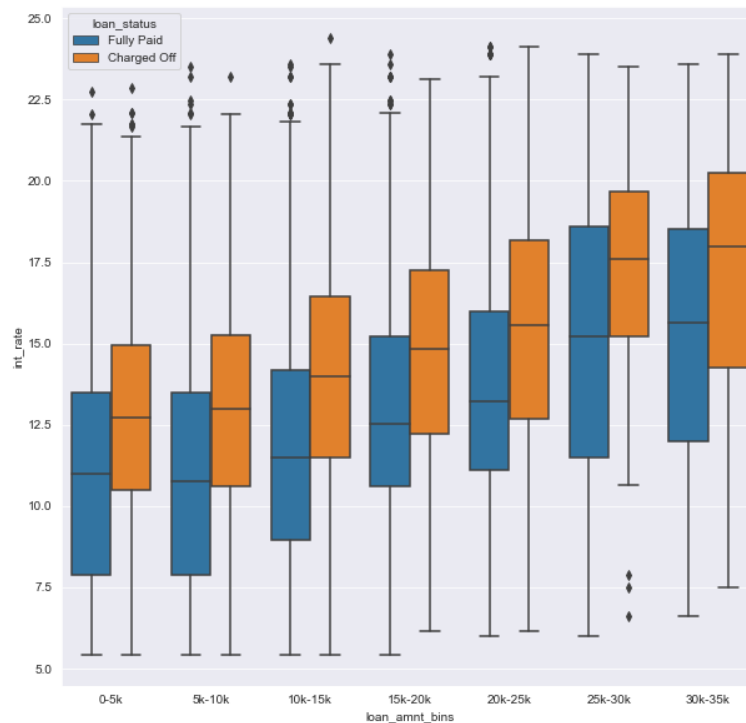
APPLICANTS WITH MORTGAGE AND INCOME OF 12-14000



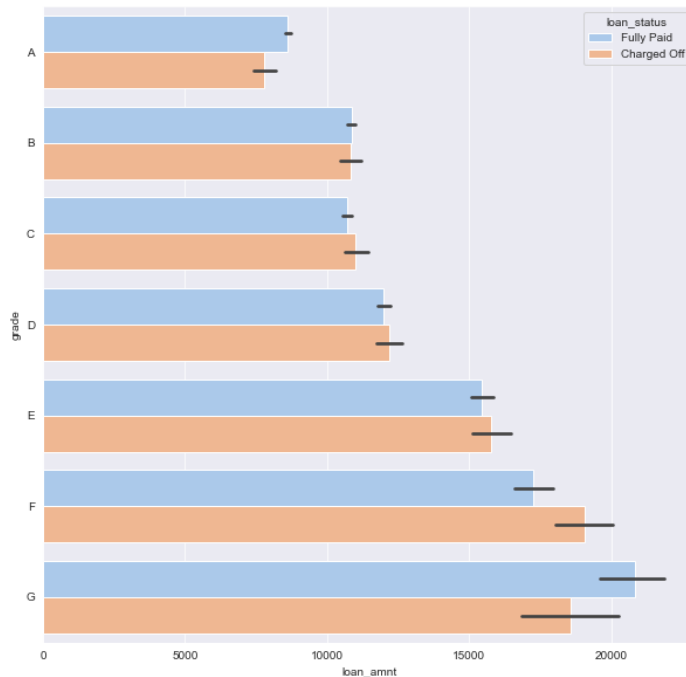
TERM AND LOAN AMOUNT ARE THE SAME FOR BOTH APPLICANTS



CHARGED_OFF APPLICANTS HAVE HIGHER LOAN AMOUNT AND HIGHER INTEREST RATES



CHARGED OFF APPLICANT WITH GRADE F ARE LIKELY



- Applicants who defaulted on average has lower annual_inc than fully paid.
- Applicants who defaulted on average has higher interest rates than fully paid.
- Top three reasons for loan were debt consolidation, other and small business.
- Charged off has the highest percentage of Not verified category.
- Applicants with Charged_off status vs Fully Paid status doesn't have a significant difference in loan amounts and related to terms.
- Person with longest term takes larger loans.
- Higher the income, higher the loan amount.
- For grade G and interest rate above 20% for fully paid and charged_off applicants.
- When employment length is high the interest rate is high for charged_off applicants.
- Applicants taking loan for 'home improvement' have annual incomes of 60k -70k
- Charged_off applicants whose home ownership is 'MORTGAGE' have loan amount of 12k-14k.
- Charged_off applicants who receive interest at the rate of 21-24% and have an annual income of 70k-80k
- Charged_off applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %