Lending Club Case Study

Submission

Team:
Vikram Pawar

Date : 5th July 2023

Index

What is this case study about?

In this case study, we are working on applying the knowledge of Exploratory Data Analysis gained to understand how consumer and loan attributes influence the tendency of default for the Lending Club using dataset which includes complete loan data for all loans issued by Lending Club through the time period 2007 t0 2011.

Finally we would have to recommend driving factors (or driver variables) behind loan default which company can for its portfolio and risk assessment.

Problem Statement

LendingClub is an online peer-to-peer lending platform that connects borrowers and investors.The platform operates as an intermediary, facilitating loans between individual borrowers and investors who are seeking to earn interest on their funds.

Borrowers who are seeking personal loans can apply for loans through LendingClub's website. The platform evaluates the creditworthiness of the applicants using various factors, including credit history, income, and debt-to-income ratio. Approved borrowers are then assigned an interest rate based on their credit profile.

Two **types of risks** associated with the Lending Club'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

Like most other lending companies, **lending loans to 'risky'** applicants is the largest source of financial loss **(called credit loss)**. The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, **borrowers who default** cause the largest amount of **loss to the lenders**. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

Business Objective

The goal is to **identify these risky loan applicants**, then risky 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.

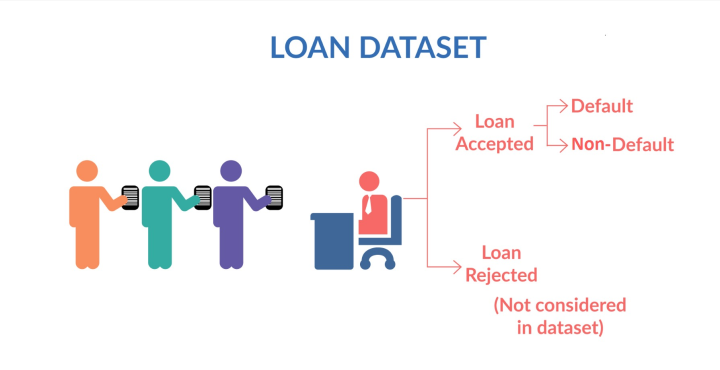
In other words, **the company wants to understand the driving factors** (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Business Understanding

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

1.Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

1. **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
2. **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'.
3. **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



2.Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Analysis Approach

01

Data Cleaning

- Rows Analysis
- Column Analysis
- Column removals
- Data conversion
- Outlier removal
- Create Derived column

02

Univariate Analysis

Analyse single variable as below

- Quantitative Variables
- Ordered Categorical Variables
- Un-Ordered Categorical Variables
- Derived Variables

03

Bivariate Analysis

Analyse influence of one variable on other as below

- Categorical Variables
- Quantitative Variables

04

Collect Insights

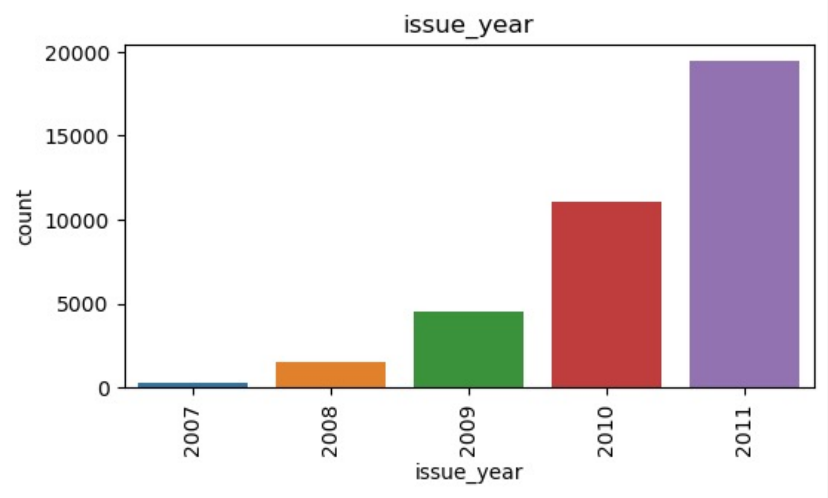
- Conclusion
- Recommendation on driving factors

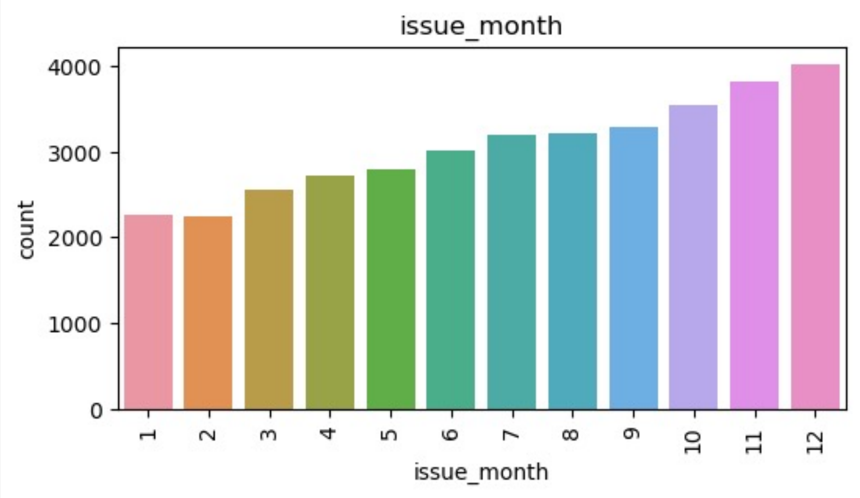
Analysis

After loading, analysing, cleaning and performing univariate, bivariate analysis on the dataset provided we were able to get interesting insights and able to recommend some of stronger driving factors (variables) leading to defaults.

In this section we will see columns/variables one by one and its univariate, bivariate analysis and if possibly it is stronger variable for the loan default.

Let's first see loan application trend for LC





- We can see growth in loans sanctioned by Lending Club year on year.
- Loans applications are seen more for last 3 months of year and lesser for first 3 months. May be by year ends people are looking for more cash as festival is around same time or they are consolidating debt by year end

upGrad

SINCE 1998

SILVER JUBILEE YEAR

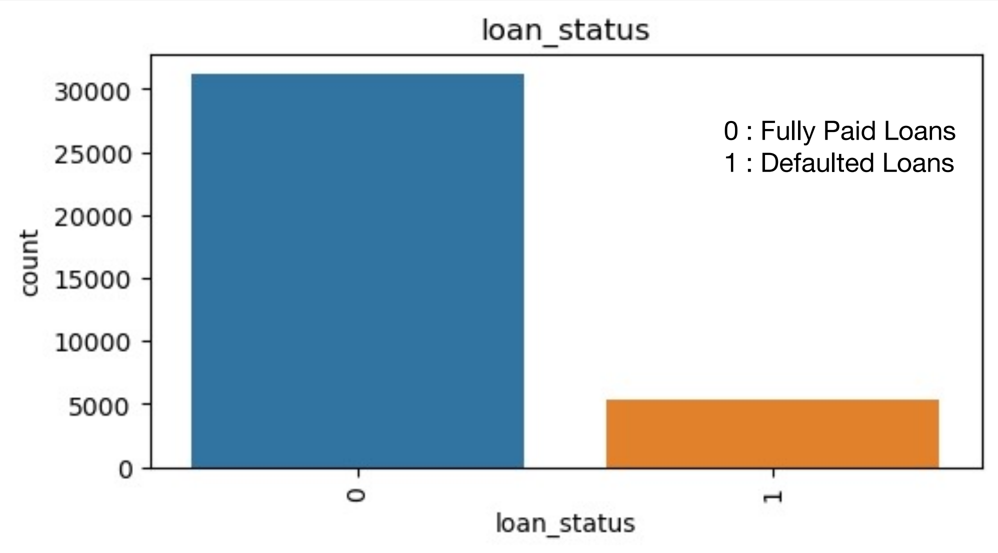
Pioneering Excellence in Education, Research & Innovation

01

Analysis of Loan Attributes

upGrad

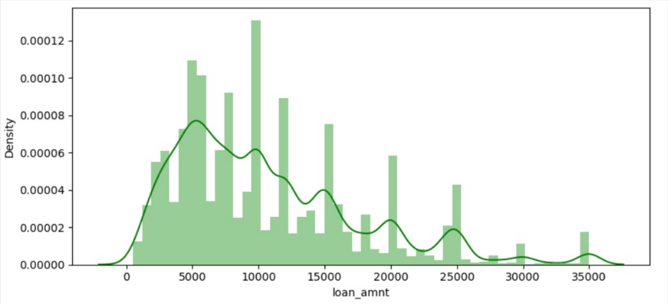
loan_status

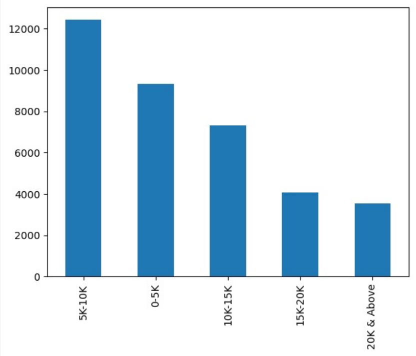


- Lending Club have approx. 14% default rate as per the analysis

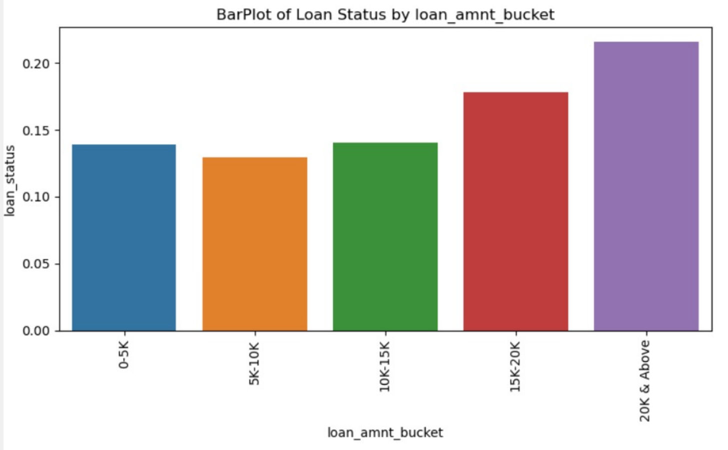
Loan Amount (loan_amount)

Univariate Analysis





Bivariate Analysis

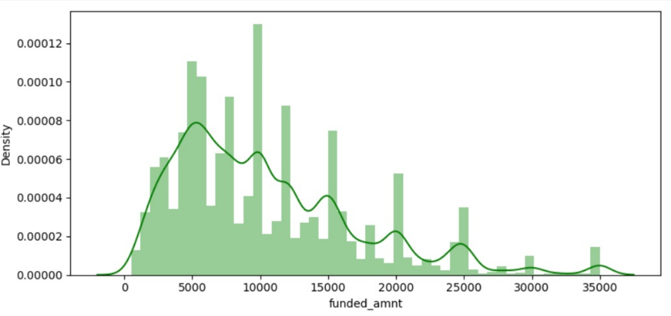


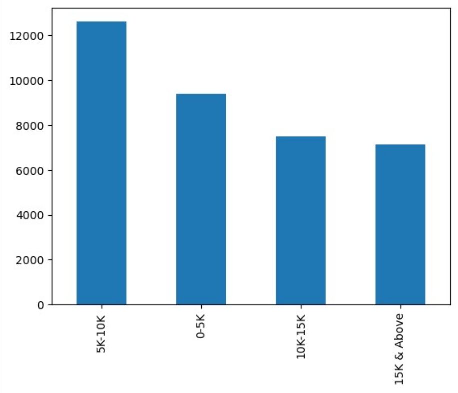
- Major loans are sanctioned between 5k to 10k
- Loans in range of 5K-10K are more but default by customer on this loans is lesser
- Whereas loans taken for amount 15K and above is comparatively lesser but have higher chance of default by customer

Is strong attribute for loan default? : Yes

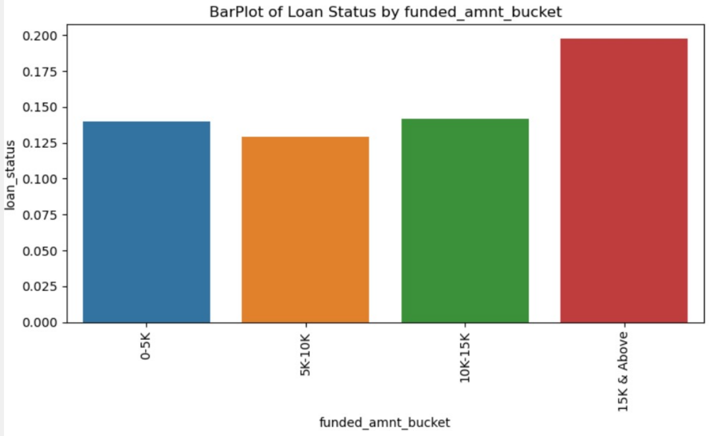
Loan Funded Amount (funded_amount)

Univariate Analysis





Bivariate Analysis

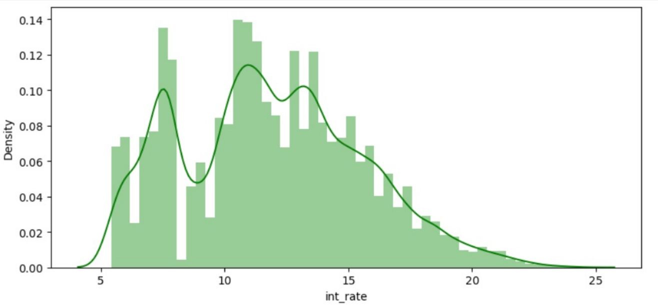


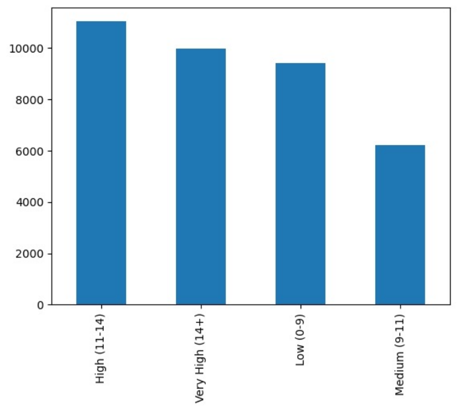
- Most of the funded_amnt is in the range of 5K to 15K which is obvious considering the loan amount sanctioned are higher in same range
- Similar to loan_amount, we are seeing more loans are funded in range of 5K-10K range but default by customer on this loans is lesser compare to 15K & above funded loans
- 15K and above funded loans are having higher chance of default by customer

Is strong attribute for loan default? : Yes

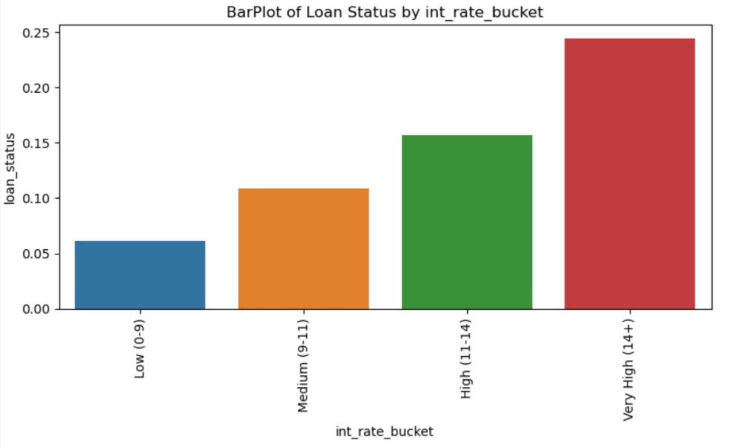
Interest Rate of loan (int_rate)

Univariate Analysis





Bivariate Analysis

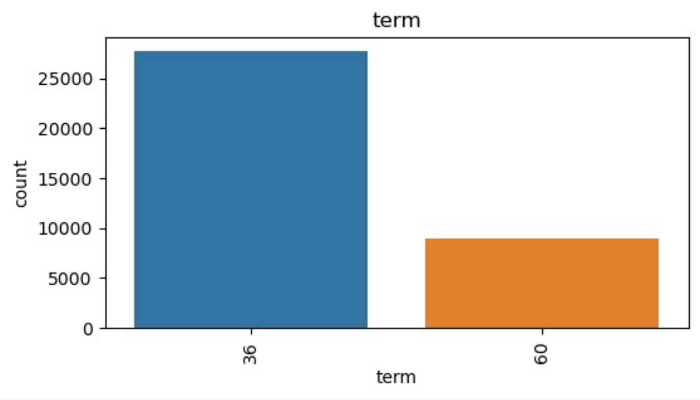


- Lending Club seems to be sanctioning loans at higher interest rate
- Majority of the interest rate is in the range of 5% to 14% going at the max to 24%
- As the inerest rate increases the the default percent increases steeply
- Loans with Very High interest rate of 14% and above have possibility of chargeoff

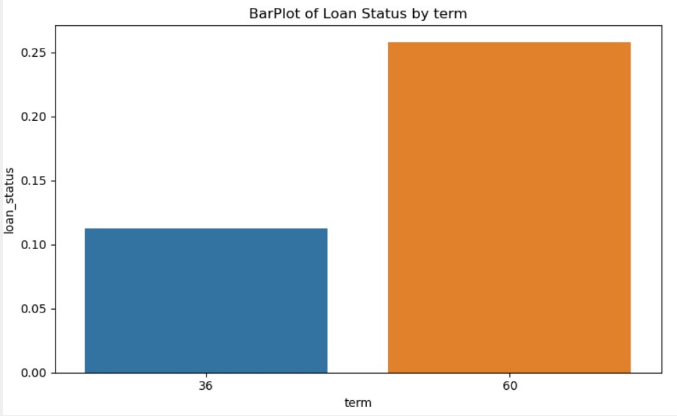
Is strong attribute for loan default? : Yes

Term of loan (term)

Univariate Analysis



Bivariate Analysis



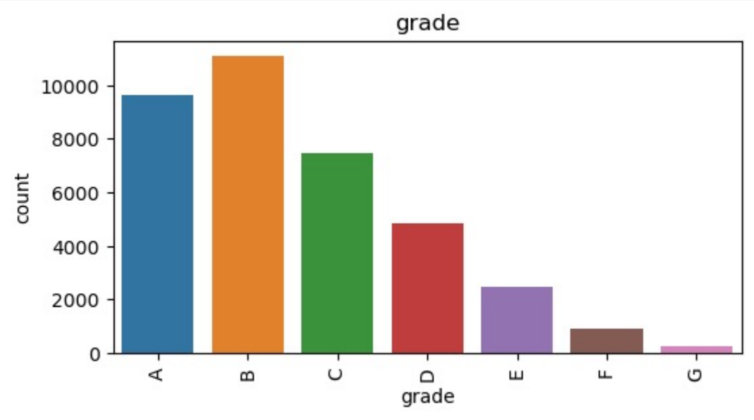
- Majority of the loans are in with term of 36 months
- Applications with term=60 potentially will have high Charge Offs

Is strong attribute for loan default? : Yes

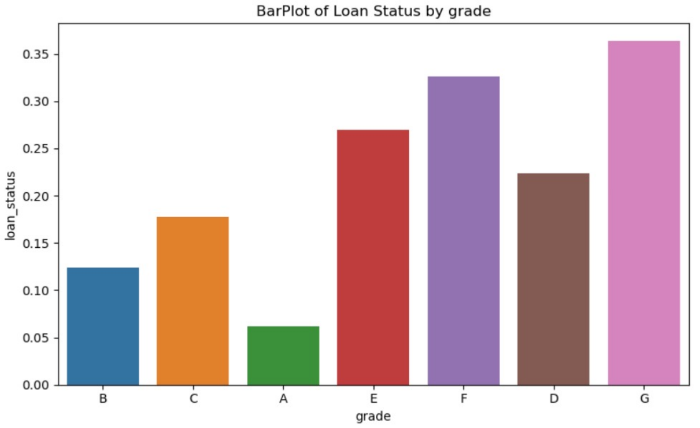
upGrad

Grade of loan (grade)

Univariate Analysis



Bivariate Analysis

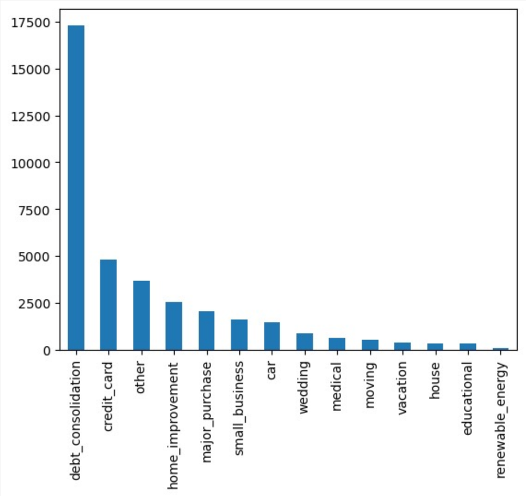


- Majority of loan application fall under Grade B
- G grade loans are having possibility of default followed by F & E

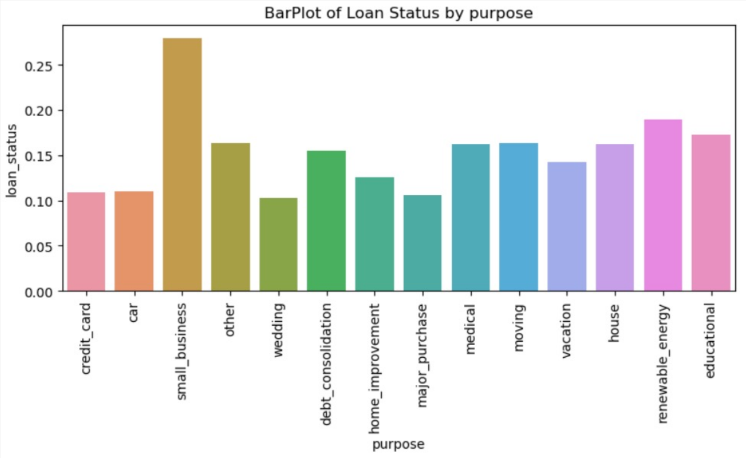
Is strong attribute for loan default? : Yes

Purpose of loan (purpose)

Univariate Analysis



Bivariate Analysis



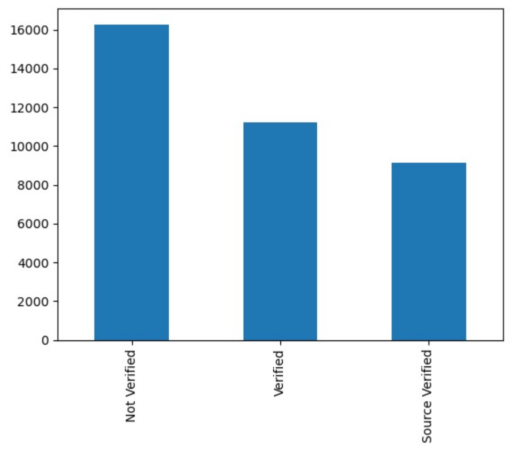
- Approx. 47% of loans are applied for debt_consolidation

- Highest loans are taken for debt_consolidation but charge off is comparatively lesser than other purposes
- Highest probability of Charge Offs is for small_business but the volume of loans is comparatively low
- Renewable energy have lowest loans but have pretty high charge off potential

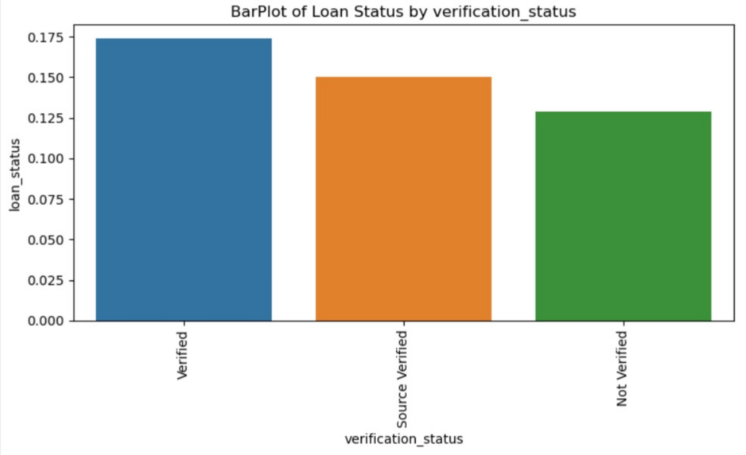
Is strong attribute for loan default? : Yes

Income verification status of loan (verification_status)

Univariate Analysis



Bivariate Analysis



- Approx. 44% of loans are sanctioned without approving income or source of income
- If defaults for not verified source is higher than Lending Club may have to revisit strategy around income source verification

- We can see loans with not verified income source are more but the default on them is lesser than that of verified income source loans
- All 3 values in verification status seems to having charge off similar so conclusive decisions is difficult

Is strong attribute for loan default? : No

Pioneering Excellence in Education, Research & Innovation

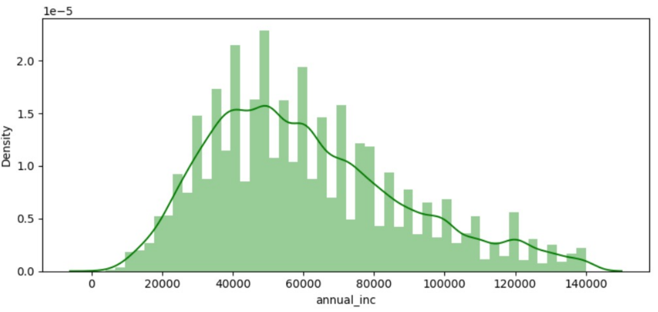
02

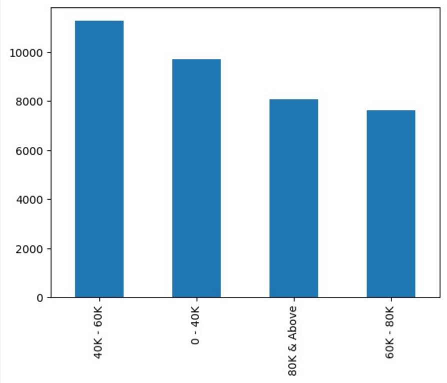
Analysis of Customer Attributes

upGrad

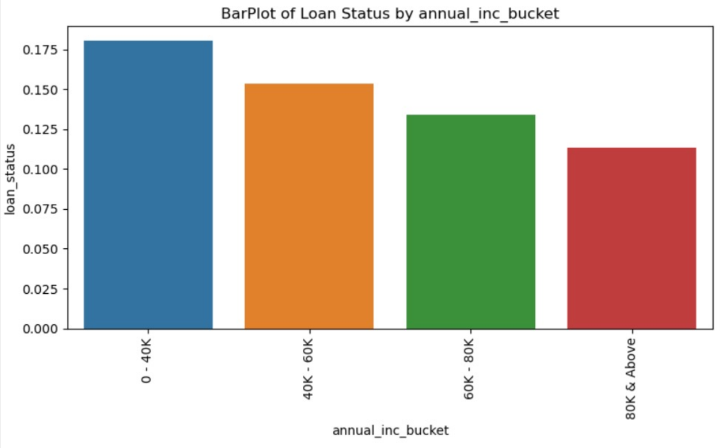
Annual Income of Customer (annual_inc)

Univariate Analysis





Bivariate Analysis

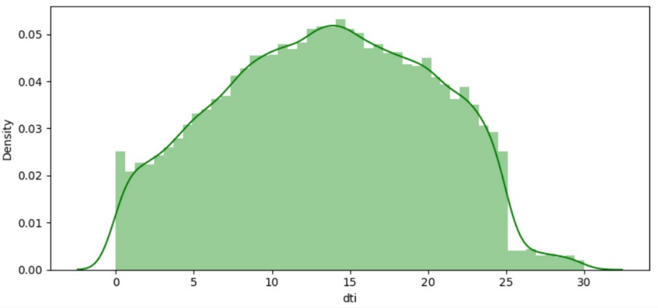


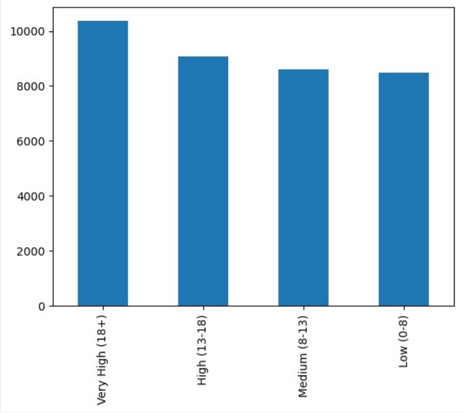
- Majorly customer annual income is between 40000 to 78000
- Loan taken by customer in annual income range of 0-40K has the highest charge offs
- Customers with Annual income range of 80K & above have lesser possibility of charge off
- Increase in annual income decreases the possibility of charge off

Is strong attribute for loan default? : Yes

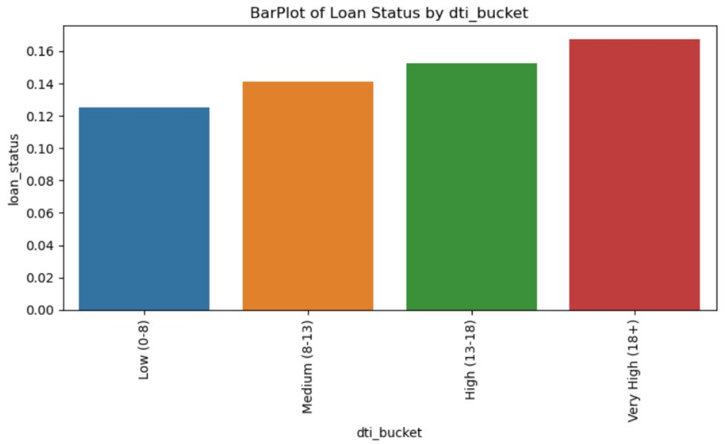
Debt To Income Ratio of Customer (dti)

Univariate Analysis





Bivariate Analysis



- Majorly dti is ranging between 8 to 18 & we can see sharp drop in dti after 25
- We are seeing similar loan count across bands of DTI
- Obviously we are seeing people with higher DTI are applying for more loans and which is again obvious from purpose as major loans are applied for debt_consolidation

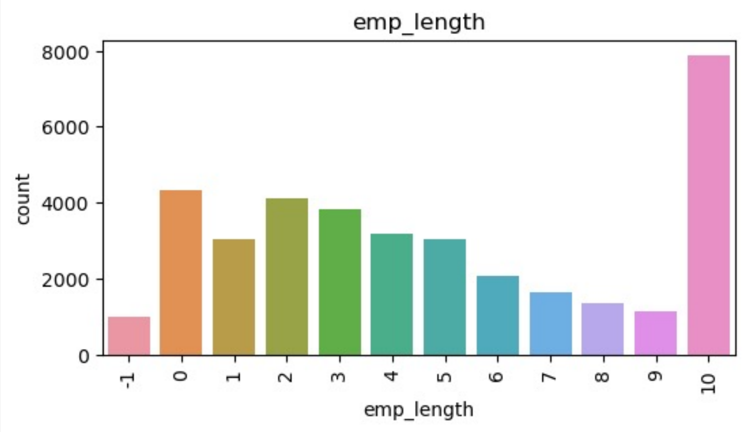
- As the dti increase the percentage of loan defaulted increase.

Is strong attribute for loan default? : Yes

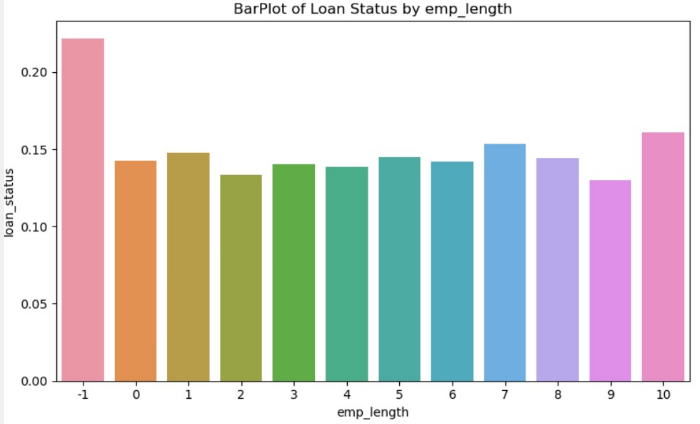
Employment Length of Customer (emp_length)

Pioneering Excellence in Education, Research & Innovation

Univariate Analysis



Bivariate Analysis



- Majority of the loan applicants are having employment length of 10+ years
- Also interesting to see the loan applications in experience range of 0-3 years is more than in experience range of 6-7 years

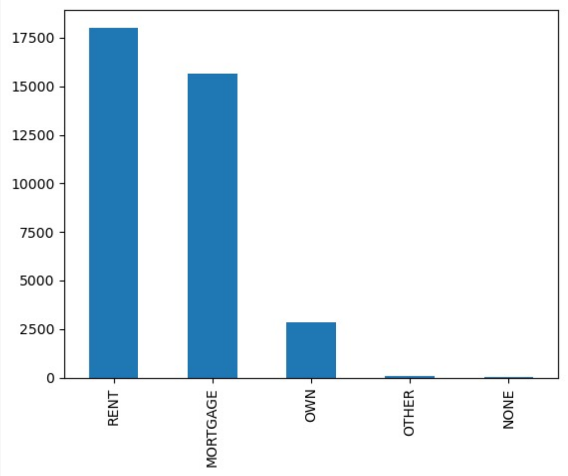
- Highest Charge Offs are in customer with emp_length of 10+ Years and above but we have maximum loans in same employment length
- Charge Off across other employment length is similar hence inconclusive

Is strong attribute for loan default? : No

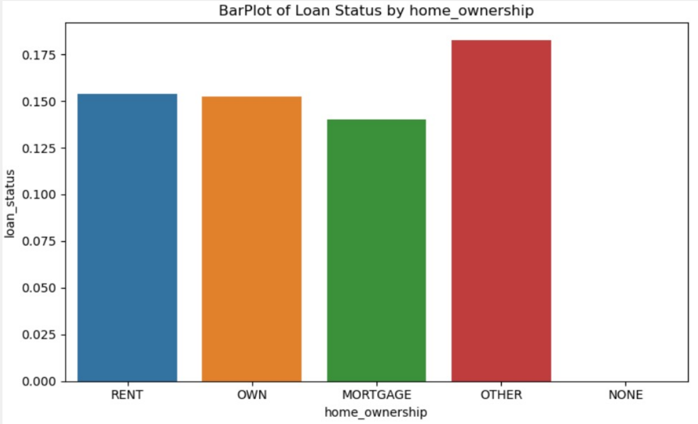
upGrad

Home Ownership of Customer (home_ownership)

Univariate Analysis



Bivariate Analysis



- Customers on rent and mortgage tend to opt for loans more than those owing house

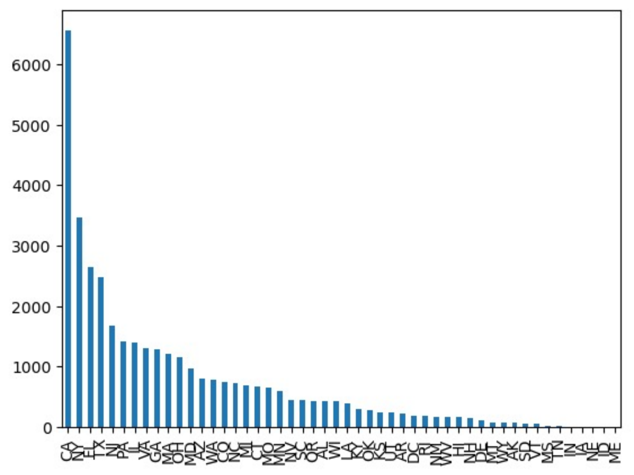
- Person having house on Rent & Mortgage have taken more loans and also default is high
- Person with owned house have taken lesser loans but default too is very similar to loan with rent or mortgaged home customers

Is strong attribute for loan default? : No

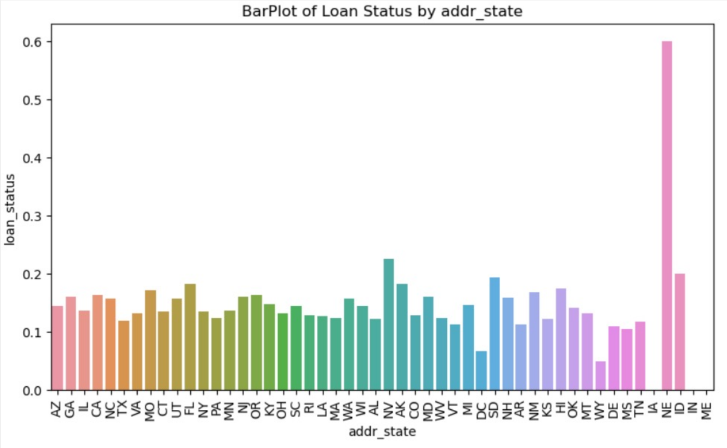
upGrad

Address Staet of Customer (addr_state)

Univariate Analysis



Bivariate Analysis



- Maximum loans/business is coming from CA, BY, FL & TX states
- Charge off for NE state is spiked but volume of loan for the NE state is lesser (only 5) it may some exceptional loan application from the state so we can ignore in analysis.
- For rest of states the default is very similar so cannot make conclusive decision

Is strong attribute for loan default? : No

Findings/Conclusion – Univariate Analysis

Considering Customer Attributes

- Customers below 60K annual income are majorly seeking loans
- Majorly customer with 10+ years working experience are majorly seeking loans followed by customers with working experience upto 2 years
- Customer in Rented or Mortgaged houses are majorly seeking loans compared to owned house
- Highest loan applications are seen for debt_consolidation purpose
- CA, BY, FL & TX states has the maximum amount of loan applications
- Majority of the debt to income is in the range of 0 to 20

Considering Loan Attributes

- Highest applications are seen in range of 5k to 10k of loan amount
- Majorly loans are sanctioned with interest rate is in the range of 5% to 14% some going upto 24%
- Majority of the loan applications are sanctioned for term of 36 months
- Most of the loan installment are in the range of 160 to 400
- Majority of loan application counts are sanctioned under Grade B
- Major loans of Lending Club seems to be sanctioned without income or income source verification

Considering Other Attributes

- For Lending Club, loan applications are increasing year by year
- Major loan applications are seen in Oct, Nov, Dec and lowest for Jan, Feb, Mar

Findings/Conclusion – Bivariate Analysis

- Percent of loan default is more for 60 months term than 36 months term
- Percent of loan default is more for lower grade of loans (G, E, F default possibility is more)
- Percent of loan default is higher for small business borrowers
- Percent of loan default is higher for loans of higher loan amount & funded amount majorly above 15K
- Percent of loan default increases as rate of interest increases
- Percent of loan default is higher if annual income is less than 40K for customer
- Percent of loan default increases as loans are sanctioned with higher dti customer
- Percent of loan default is similar across emp length
- Percent of loan default is similar across home ownership
- State is also of not much importance in loan default except NE state which had spike in loan defaults
- Income or Income Source verification is not of much influence for default as even after verified income borrowers seems to be defaulting more than non verified.

Recommendation on driving factors influencing tendency of default

Recommendation on driving factors influencing tendency of default

— Customer Attributes

dti
A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.

annual_inc
The self-reported annual income provided by the borrower during registration.

— Loan Attributes

term
The number of payments on the loan. Values are in months and can be either 36 or 60.

grade
LC assigned loan grade.

purpose
A category provided by the borrower for the loan request.

int_rate
Interest Rate on the loan

loan_amnt
The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.

funded_amnt
The total amount committed to that loan at that point in time.

Pioneering Excellence in Education, Research & Innovation

Thanks!

Have any questions?

Email: vikrampawar08@gmail.com
Phone: +91-9870731602
Linkedin: https://www.linkedin.com/in/vikrampawar88/

upGrad