



Lending Club Case Study

SUBMISSION

Name:

Gayathri **C**

CHANDRIKA **PATURI**





Overview:

Background:

- Lending Club 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

Business Objective:

• Company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators to identify the risky loan applicants at the time of loan application

Strategy of Analysis:

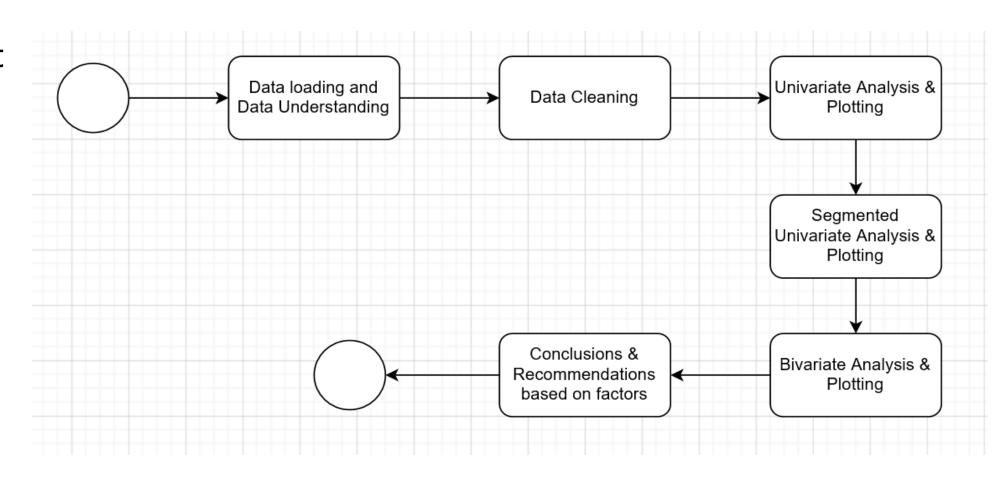
- Understanding data
- Cleaning the data
- Analysing the data
- · Plotting the data
- Concluding the analysis





Problem solving methodology

Flow Chart







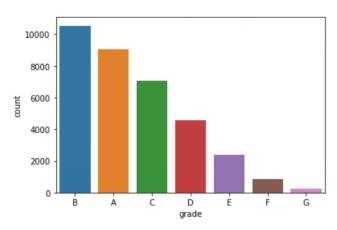
- Loading the csv
- Understand the data
- Data cleaning removing missing entries, converting the datatypes, data extraction and conversion, finding out the outliers and ignoring them, removing duplicate columns / rows, removing single valued columns etc





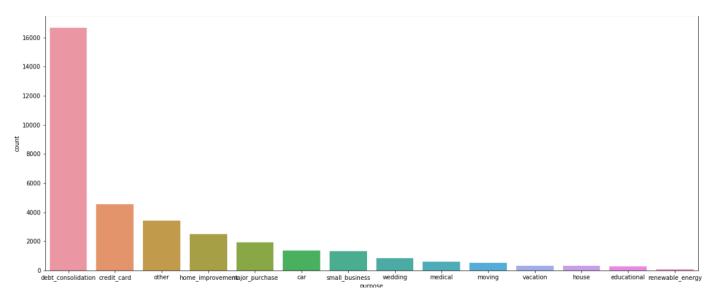
Univariate Analysis

Number of loans against grade



Higher number of loans are taken by higher grade borrowers

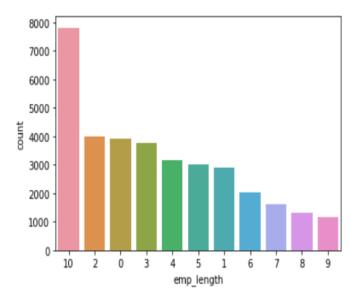
Number of loans against purpose



Higher number of loans are taken for debt consolidation. Based on it, a bank or finance lender can decide more offers to attract customers and give competition to competitors.



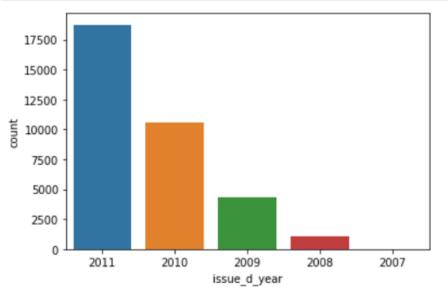
Number of loans based against emp length



Higher number of loans are taken by borrowers with emp length 10 or above.

With #years of experience, the salary will get increased and the person can afford to take loan compared with a fresher or medium experienced person.

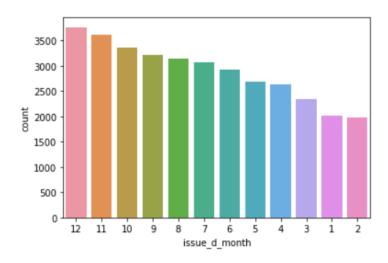
Number of loans based on issue year



Higher number of loans are taken in 2011. Which determines the bank is running in profit with more #customers year on year.

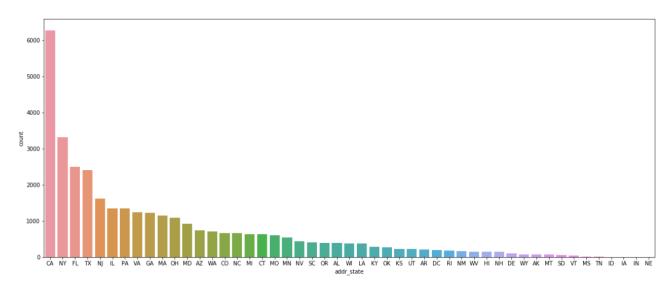


Number of loans based on issue month



Higher number of loans are taken in Dec. Due to higher offers during festival time.

Number of loans by each state

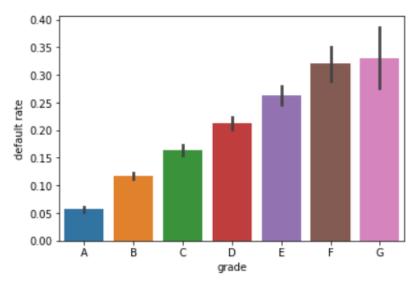


Higher number of loans are taken by borrowers in state CA. by looking at the above, lender can decide what to be done so that the performance will be go high in other states too.



Segmented Univariate Analysis

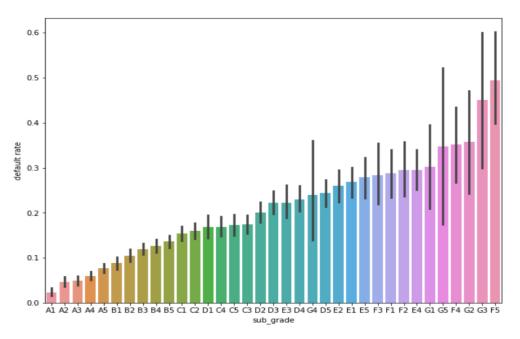
default rates across all grades.



As the grade of loan (risky of loan) increases the default rate increases



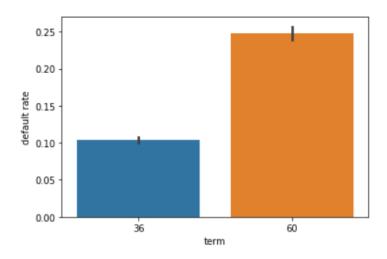
default rates across subgrades



In-depth analysis which gives idea which sub grade is riskier than other grade. For example: A1 is less riskier than A2. A2 is less riskier than A3 and so on...

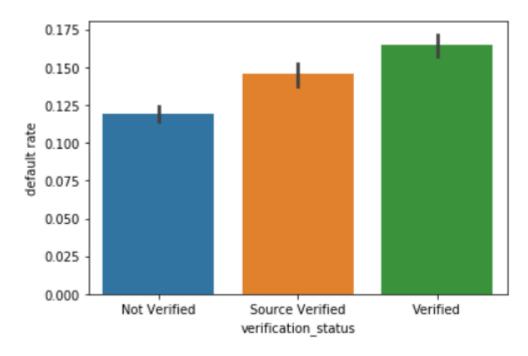


default rates across monthly terms



The risk is more, i.e, defaulters are more when the term to repay is 60.

default rates across verified loans

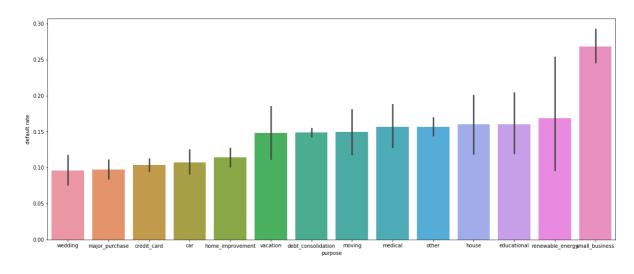


verified users have more defaulters.

That means the verification process need to be corrected. Fraud during approvals of loans (could be bribing to get loan, could be relative / close knit to approve loan / submitting the fake proofs etc)

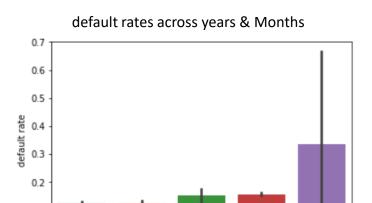


default rates based on purpose of taking loan



small business loans have more defaulters. Because,

- 1. in the world most of the people are middle class people
- 2. so, small business will be run by them
- 3. if no profitable business, may not have the financial condition to repay



0.1

0.0

2010

2009

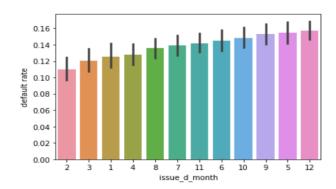
In 2007 the defaulter's rate is more the lending club has been established in 2006. So, it is new experience and difficult to manage how to recover the loans from customers by lender.

2011

2007

2008

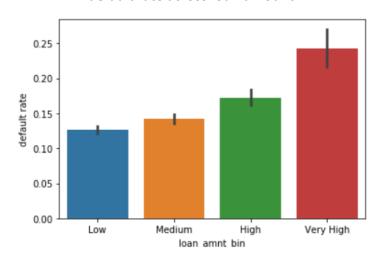
issue d year



In December month the defaulter are more

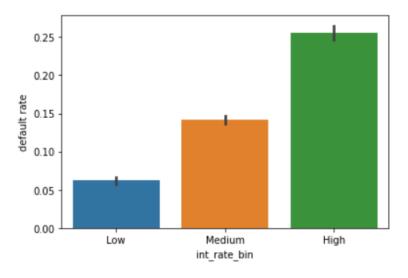


default rate across loan amount.



greater the loan amount, greater the risk with defaulters. because they may pay or may not pay huge amount.

default rate across interest rate.



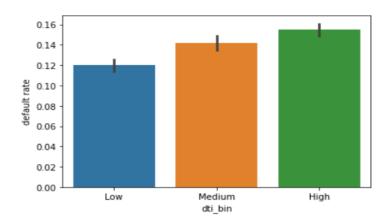
as the interest rate increases, the chance of having defaulters in the list.

As they may or may not pay the interest (if no profit in their business)



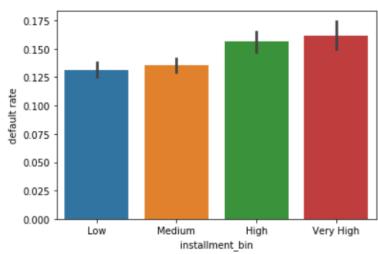


default rate across debt to income ratio.



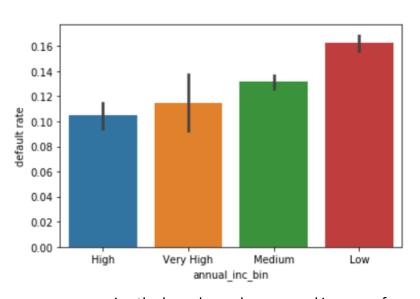
higher the DTI, more defaulters.

default rate across instalments



higher instalments , then more defaulters.

default rate across income

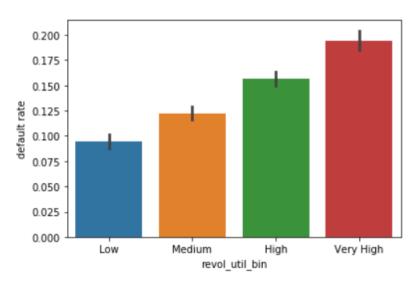


repaying the loan depends on annual income of a person / organization. So, lower the annual income, greater the risk by having more defaulters.



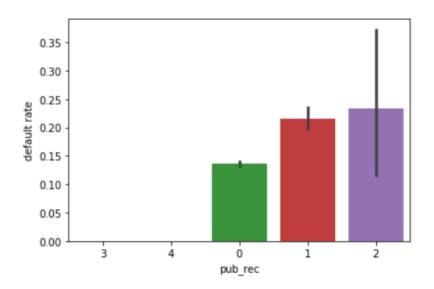


default rate across revolving utilization



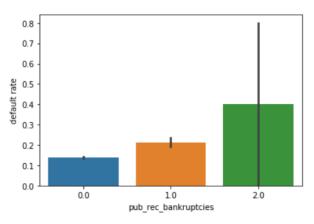
higher the revolving utilization, more risk to receive the amounts from customers.

default rate across public records



default rate increases with increase in number of derogatory public records

default rate across public records bankruptcies

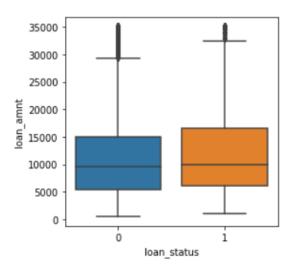


default rate increases with increase in number of pub_rec_bankruptcies



Bivariate Analysis

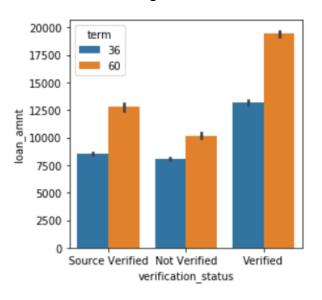
loan amount for each loan status



higher average loan amount gives more chances of defaulters.

UpGrad

loan amount against verification status

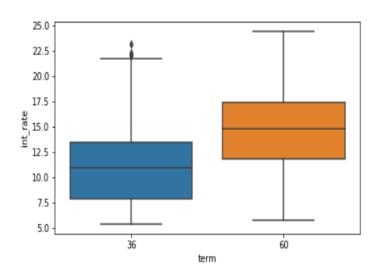


Higher loan amounts are more verified



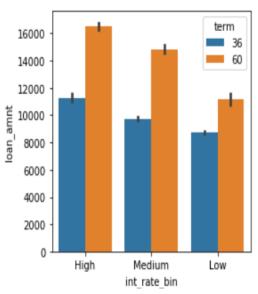


Interest rate against term



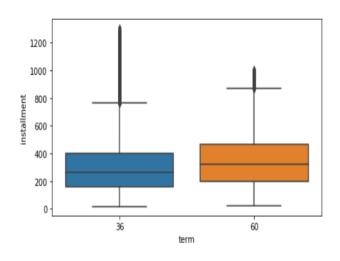
60 months term has higher interest rate compared to 36 months

loan amount against interest



Higher loan amount has higher interest rate depending on the tenure. by considering earlier factors, more tenure has more defaulters

instalment against term

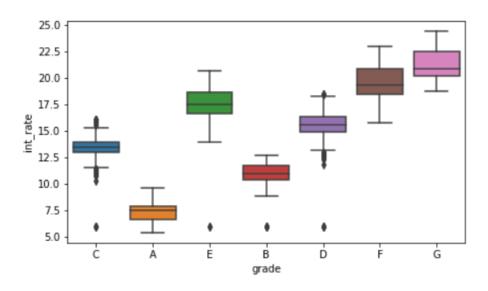


60 months term has higher average instalment compared to 36 months



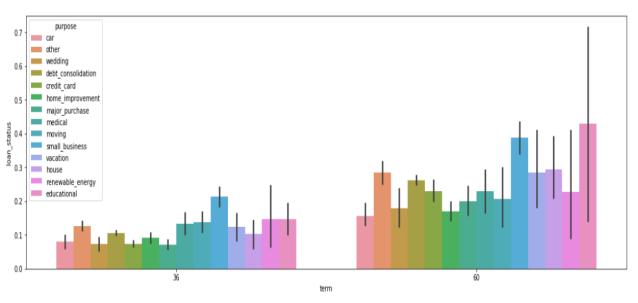


Interest rate against grade



lower grades have higher interest rate

default rate against term each purpose



Small business purpose loan are higher in 36 months term whereas renewable energy is higher for 60 months



Heat Map to Find the Correlation between different numeric factors

Few of the findings are,

interest rate and revolution-util is correlated by a factor of 0.47

loan amount and last payment amount is also correlated by 0.47

etc....

loan_amnt	1	0.0049	0.076	0.14	-0.04	0.29	0.93	0.4	0.068	0.0049	0.17	-0.043	0.31	0.47	-0.027
inq_last_6mths	-0.0049	1	0.0059	0.011	0.0075	0.13	0.0026	0.036	-0.068	1	0.093	0.024	-0.025	0.028	0.015
đti	- 0.076	0.0059	1	0.051	-0.034	0.11	0.067	-0.1	0.28	0.0059	0.3	-0.0065	0.25	0.016	0.0057
emp_length	0.14	0.011	0.051	1	0.013	0.0011	0.12	0.18	0.008	0.011	0.1	0.067	0.16	0.08	0.064
delinq_2yrs	-0.04	0.0075	-0.034	0.013	1	0.16	-0.029	0.028	-0.043	0.0075	0.012	0.013	-0.06	-0.016	0.0066
int_rate	0.29	0.13	0.11	-0.0011	0.16	1	0.27	0.057	0.47	0.13	0.0026	0.1	0.089	0.17	0.086
installment	0.93	0.0026	0.067	0.12	-0.029	0.27	1	0.4	0.1	0.0026	0.16	-0.039	0.31	0.41	-0.024
annual_inc	- 0.4	0.036	-0.1	0.18	0.028	0.057	0.4	1	0.041	0.036	0.26	-0.012	0.42	0.22	-0.0086
revol_util	- 0.068	-0.068	0.28	0.008	-0.043	0.47	0.1	0.041	1	-0.068	-0.09	0.058	0.31	-0.011	0.06
inq_last_6mths	-0.0049	1	0.0059	0.011	0.0075	0.13	0.0026	0.036	-0.068	1	0.093	0.024	-0.025	0.028	0.015
open_acc	- 0.17	0.093	0.3	0.1	0.012	0.0026	0.16	0.26	-0.09	0.093	1	0.0058	0.28	0.079	0.012
pub_rec	0.043	0.024	-0.0065	0.067	0.013	0.1	-0.039	-0.012	0.058	0.024	0.0058	1	-0.058	-0.03	0.84
revol_bal	0.31	-0.025	0.25	0.16	-0.06	0.089	0.31	0.42	0.31	-0.025	0.28	-0.058	1	0.12	-0.044
last_pymnt_amnt	0.47	0.028	0.016	0.08	-0.016	0.17	0.41	0.22	-0.011	0.028	0.079	-0.03	0.12	1	-0.019
_rec_bankruptcies	0.027	0.015	0.0057	0.064	0.0066	0.086	-0.024	0.0086	0.06	0.015	0.012	0.84	-0.044	-0.019	1
	loan_amnt -	ing last_6mths -	offi -	emp_length -	delinq_2yrs -	int_rate -	installment -	- annual inc	revol_util -	ing_last_6mths -	oben_acc -	- bnp rec	revol_bal -	last_pymnt_amnt -	pub_rec_bankruptcies -
	inq_last_6mths dti emp_length delinq_2yrs int_rate installment annual_inc revol_util inq_last_6mths open_acc pub_rec revol_bal last_pymnt_amnt	inq_last_6mths -0.0049 dti - 0.076 emp_length - 0.14 delinq_2yrs0.04 int_rate - 0.29 installment - 0.93 annual_inc - 0.4 revol_util - 0.068 inq_last_6mths -0.0049 open_acc - 0.17 pub_rec0.043 revol_bal - 0.31 last_pymnt_amnt - 0.47 rec_bankruptcies0.027	inq_last_6mths -0.0049 1 0.0059 0.011 0.0075 0.13 0.0026 0.036 -0.068 1 0.093 0.024 dti - 0.076 0.0059 1 0.051 -0.034 0.11 0.067 -0.1 0.28 0.0059 0.3 -0.065 emp_length - 0.14 0.011 0.051 1 0.013 -0.0011 0.12 0.18 0.008 0.011 0.1 0.067 delinq_2yrs - 0.04 0.0075 -0.034 0.013 1 0.16 -0.029 0.028 -0.043 0.0075 0.012 0.013 int_rate - 0.29 0.13 0.11 -0.0011 0.16 1 0.27 0.057 0.47 0.13 0.0026 0.1 installment - 0.93 0.0026 0.067 0.12 -0.029 0.27 1 0.4 0.1 0.0026 0.16 -0.039 annual_inc - 0.4 0.036 -0.1 0.18 0.028 0.057 0.4 1 0.041 0.036 0.26 -0.012 revol_util - 0.068 -0.068 0.28 0.008 -0.043 0.47 0.1 0.041 1 0.068 -0.09 0.058 inq_last_6mths -0.0049 1 0.0059 0.011 0.0075 0.13 0.0026 0.036 -0.068 1 0.093 0.024 open_acc - 0.17 0.093 0.3 0.1 0.012 0.0026 0.16 0.26 -0.09 0.093 1 0.0058 pub_rec0.043 0.024 -0.0065 0.067 0.013 0.1 0.039 -0.012 0.058 0.024 0.0058 1 revol_bal - 0.31 -0.025 0.25 0.16 -0.06 0.089 0.31 0.42 0.31 0.025 0.28 0.058 last_pymnt_amnt - 0.47 0.028 0.016 0.08 -0.016 0.17 0.41 0.22 -0.011 0.028 0.079 -0.03	inq_last_6mths -0.0049	inq_last_6mths -0.0049										



- 0.8

-0.6

- 0.4

-0.2

-0.0





After our analysis, below are the Major Factors to identify the risky loan applicants at the time of loan application.

- term
- grade and subgrade
- purpose
- verification_status
- loan amount
- interest rate
- dti
- installments
- revolving util
- annual income
- public records





• Improve the verification process

Reason: we are seeing more defaulters in the approved verification status. Which is not normal

- Reduce approving loans to small business purposes as defaulters are more
- Increase the interest rate where DTI is more