LENDING CLUB CASE STUDY

SUBMITTED BY:
KHURAIJAM GUNINDRO

CONTENTS

PROBLEM STATEMENT

Lending Club, a finance company serving city folks, faces a big problem: figuring out who will pay back their loans and who won't. If people don't pay, it's bad news for the company, especially when they label accounts as "Charged-Off."

The goal is to help Lending Club avoid losing money. There are two main challenges:

1. **Picking Good Borrowers:** They need to spot people who will likely pay back their loans. Turning down these folks means losing out on potential business.

2. **Avoiding Bad Loans:** They also want to steer clear of lending money to people who might not pay back. Approving loans for these folks could lead to big losses.

This case study aims to use exploratory data analysis to spot those likely to default on loans. By doing this, Lending Club hopes to understand what factors make people more likely to default, helping them make better decisions about who to lend to.

DATA DESCRIPTION

Lending Club gave us access to a large dataset containing detailed information about borrowers' credit histories and their interactions with Lending Club. With over 39,717 records and 111 columns, there was plenty of data for our team to work with.

We focused on selecting specific variables from the dataset that were most relevant to understanding whether a borrower might default on their loan. By narrowing down the data to these key variables, we aimed to uncover the factors that could directly or indirectly influence a borrower's likelihood of defaulting. This approach allowed us to concentrate on the most impactful aspects of a borrower's profile when assessing their risk level.

1	LoanStatNew	Description
2	acc_now_delinq	The number of accounts on which the borrower is now delinquent.
3	acc_open_past_24mths	Number of trades opened in past 24 months.
4	addr_state	The state provided by the borrower in the loan application
5	all_util	Balance to credit limit on all trades
6	annual_inc	The self-reported annual income provided by the borrower during registration.
7	annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
8	application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
9	avg_cur_bal	Average current balance of all accounts
10	bc_open_to_buy	Total open to buy on revolving bankcards.
11	bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts.
12	chargeoff_within_12_mths	Number of charge-offs within 12 months
13	collection_recovery_fee	post charge off collection fee
14	collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections
15	delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
16	delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
17	desc	Loan description provided by the borrower
18	dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested L
19	dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LO
20	earliest_cr_line	The month the borrower's earliest reported credit line was opened
21	emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
22	emp_title	The job title supplied by the Borrower when applying for the loan.*
23	fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
24	fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to.

BUSINESS UNDERSTANDING

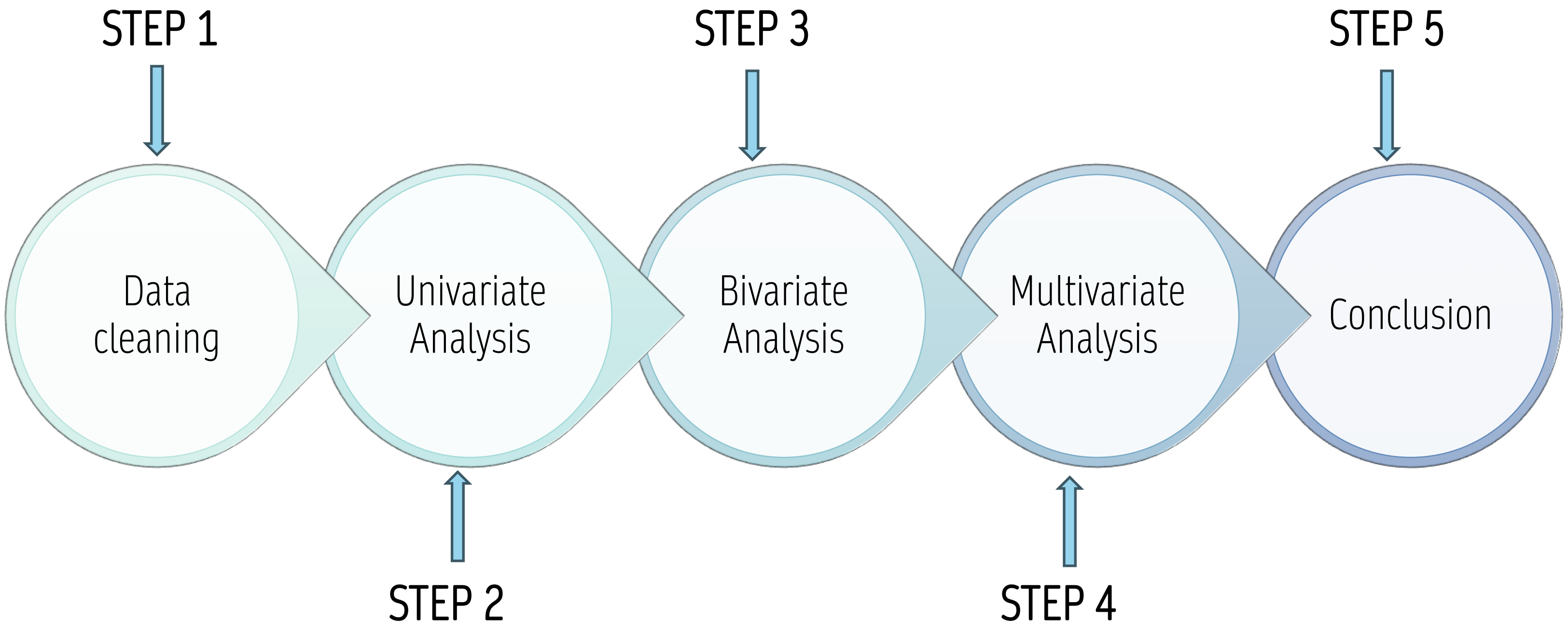
When a person applies for a loan, there are two types of decisions that could be taken by the company:

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

- 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

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 STEP



DATA CLEANING & PREPROCESSING

Data cleaning is the process of identifying and correcting errors, inconsistencies, and inaccuracies in datasets to ensure their reliability and usability for analysis, decision-making, and insights generation, thereby enhancing the quality and credibility of the results obtained from data-driven approaches.

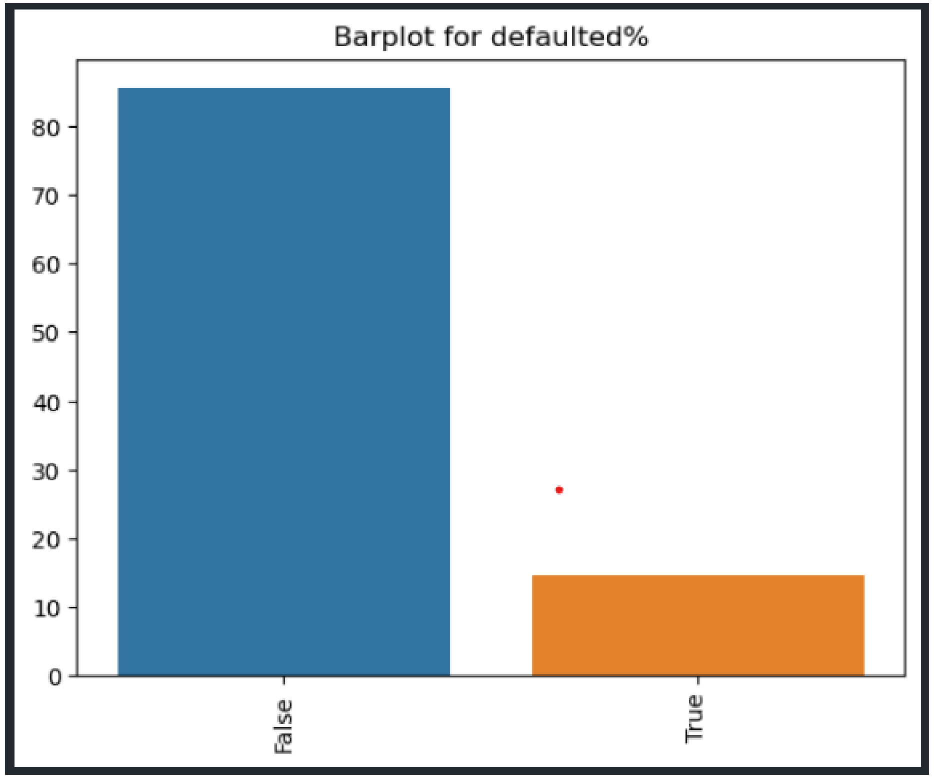
1. Firstly we checked for the columns which had NaN(NULL) values for all rows and dropped them with the help of pandas.
2. Then we dropped the columns which had 30% more than any other values as NaN.
3. Dropped the columns with only one unique value as it has no effect on the analysis.
4. Member id and url is of no use to for the analysis so it's of no use to the dataset so we dropped it.
5. Removing emp_title and title because they contain string values that are more unique values.
6. Removing the features which doesn't have anything to do with loan defaulting.
7. We cannot tell or predict anything for the current loans so we also dropped it.
8. We also need to standardize some values such as rounding the funded amount invested, splitting the no. of months such as 36 and 60 from the string, removing the percentage symbol from the interest rate, creating new columns with months & year differently from issue date and creating a defaulted column derived from loan status(defaulted column has two values i.e. True and False).

ANALYSIS ON LOAN STATUS

From this we can see that :

DEFAULTED PERCENTAGE – 14.5%

NON DEFAULTED PERCENTAGE – 85.4%

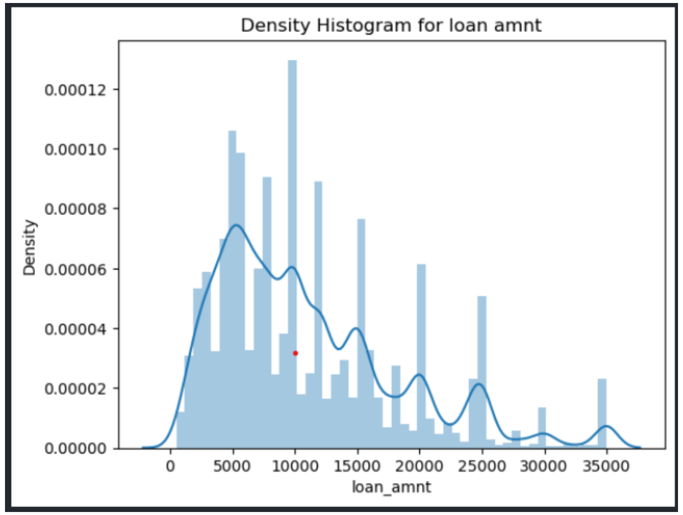


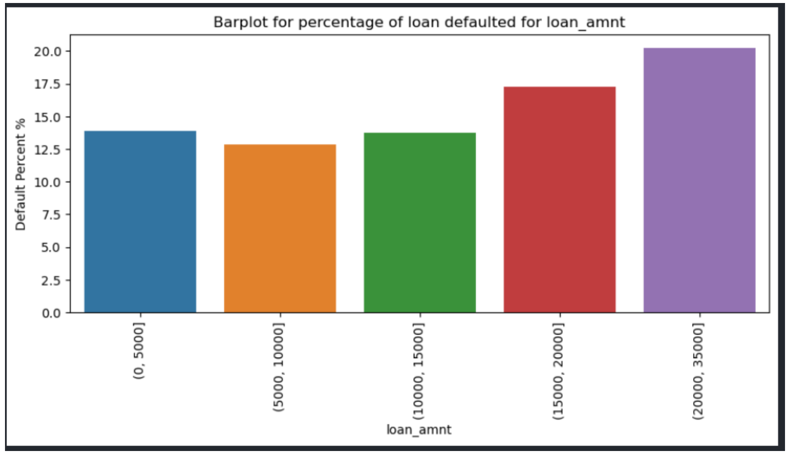
ANALYSIS ON LOAN AMOUNT

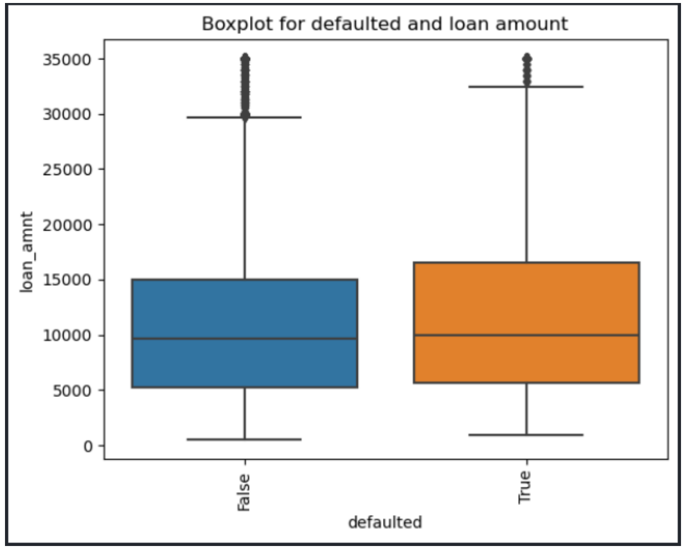
We've observed some unexpected surges in loan requests for specific amounts: $10k, $15k, $20k, $25k, $30k, and $35k.

Interestingly, the default rate for loans ranging from $15k to $35k is notably higher compared to loans below $15k and surpasses the overall default rate of 14.58%. This trend suggests that borrowers requesting larger loan amounts are more prone to defaulting on their loans.

Similarly, we've noticed similar patterns in the funded amounts and invested funded amounts, indicating that higher funding levels are associated with increased default risks.

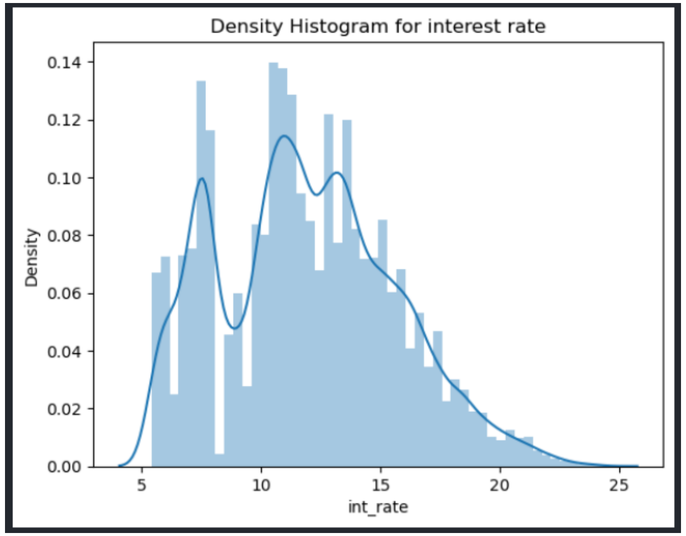


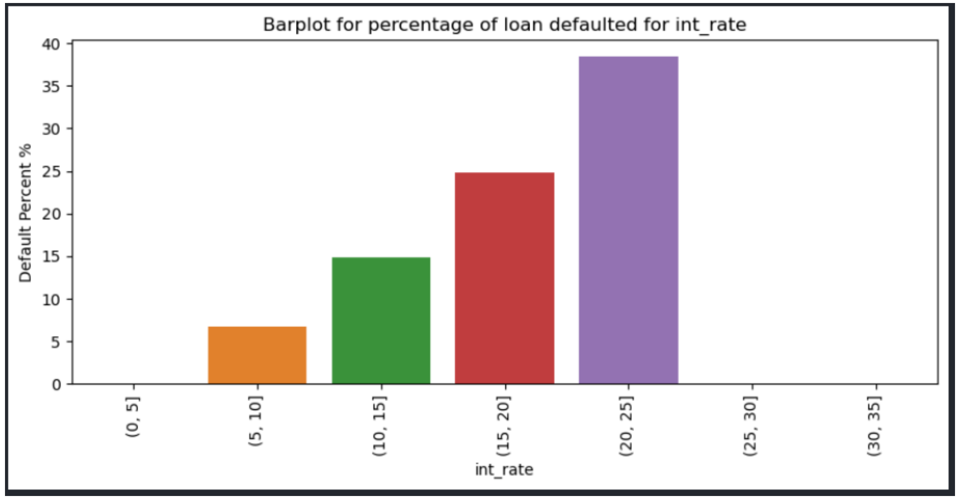


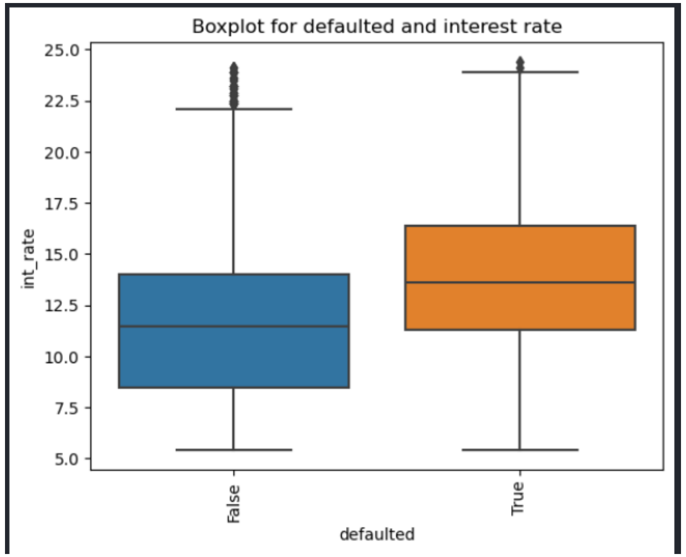


ANALYSIS ON INTEREST RATE

The loan market is experiencing an unusual trend: there's a significant drop in loans with interest rates ranging from 8% to 10%. However, as interest rates rise, the percentage of defaulted loans sharply increases. Notably, loans with interest rates between 15% and 25% have a default rate higher than the overall average of 14.58%.

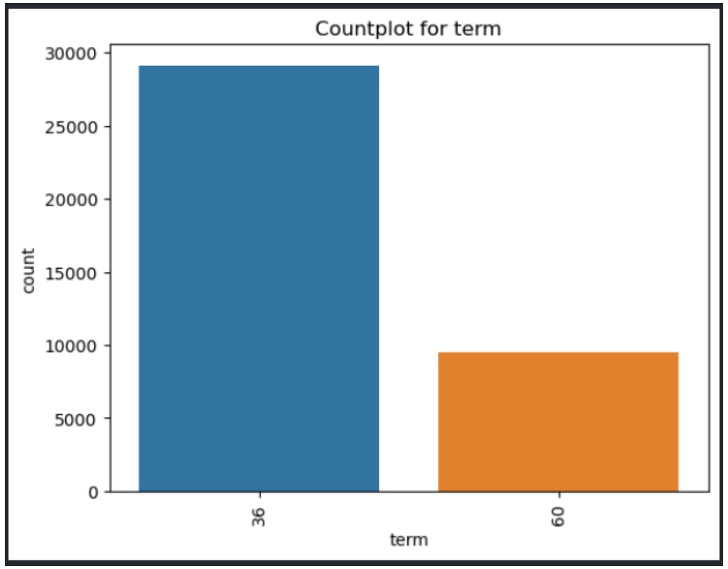


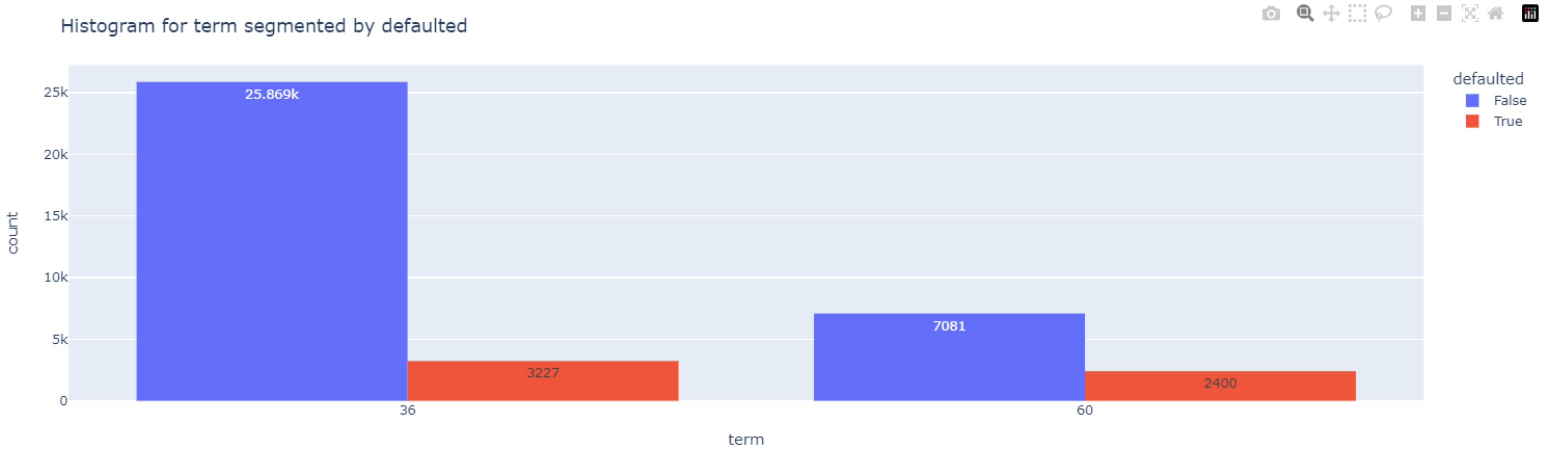




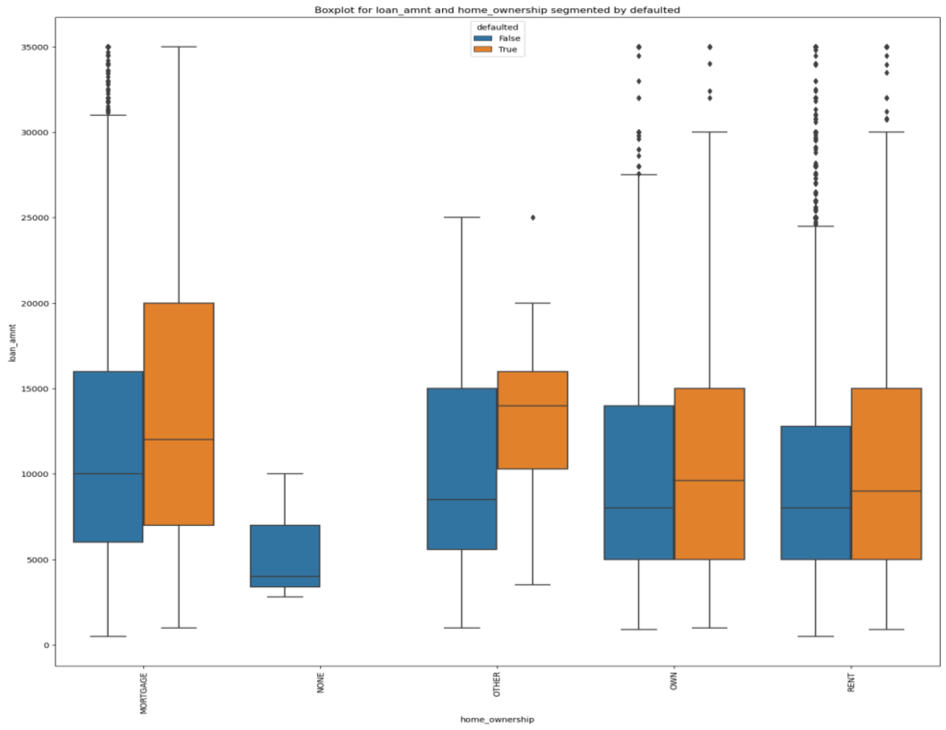
ANALYSIS ON TERM

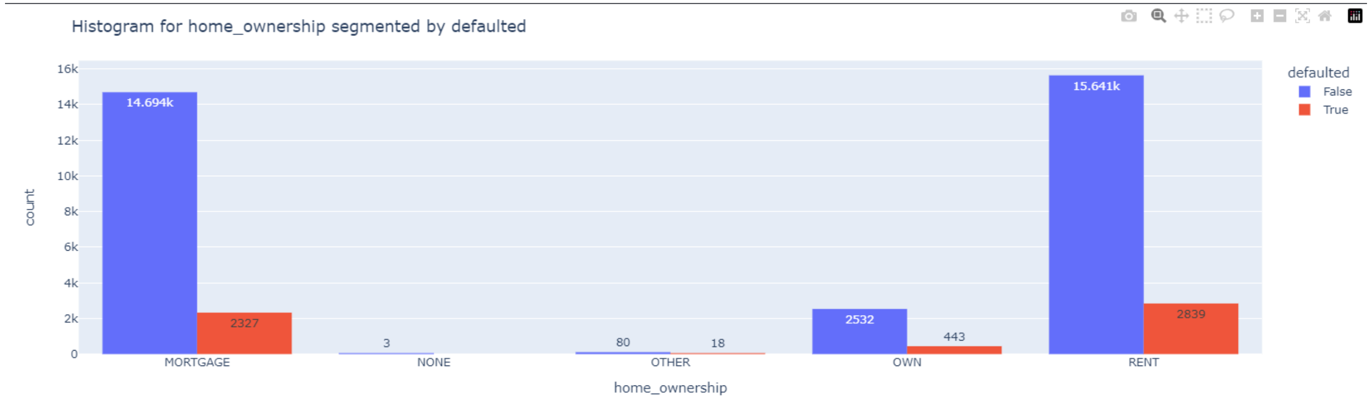
Borrowers are favouring loans with a 36-month term over those with a 60-month term.





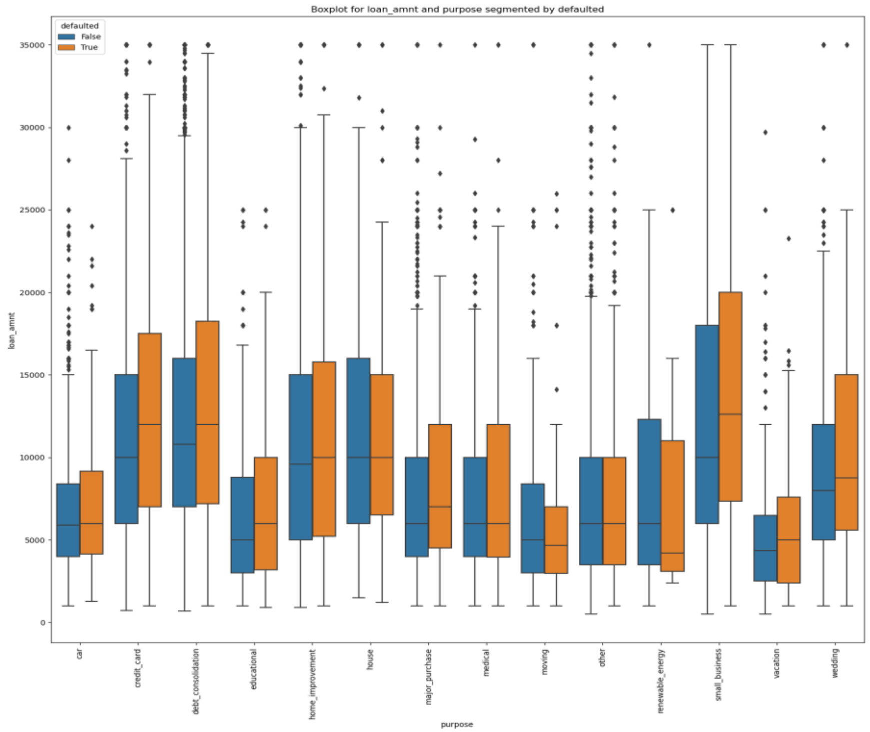
ANALYSIS ON HOME OWNERSHIP

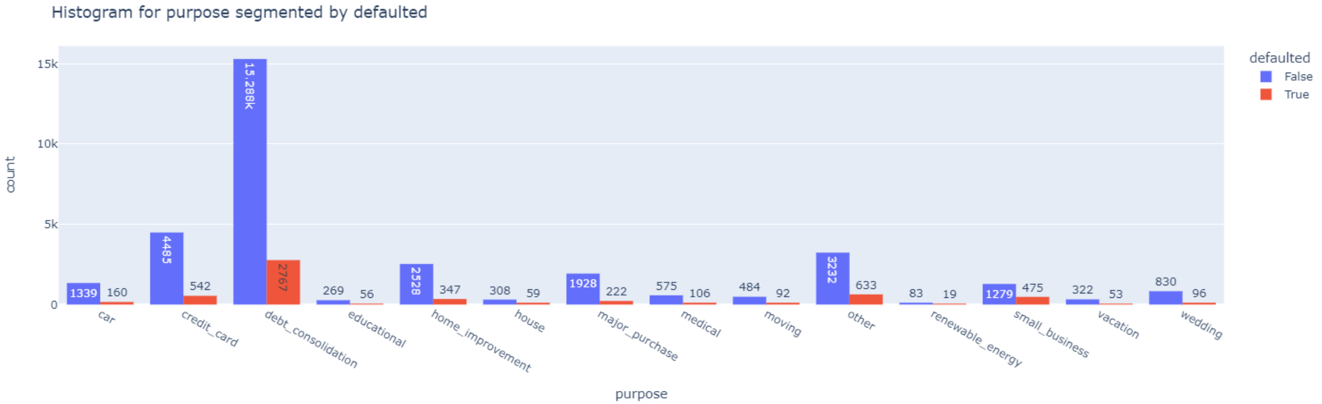




Borrowers who are renting or paying a mortgage have the highest number of loans. Interestingly, the percentage of defaulted loans remains consistent across mortgage holders, homeowners, and renters. However, when it comes to larger loan amounts, those with mortgages show a wider spread in loan default rates.

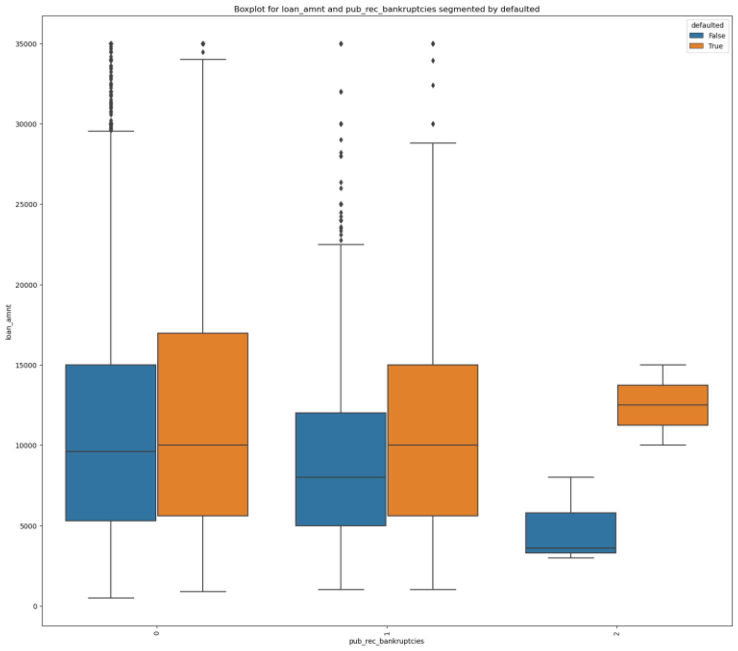
ANALYSIS ON PURPOSE

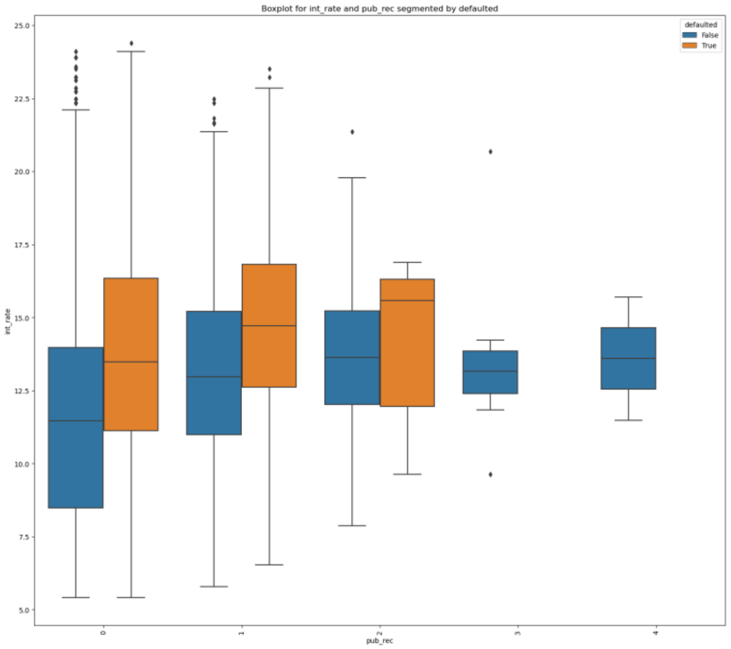


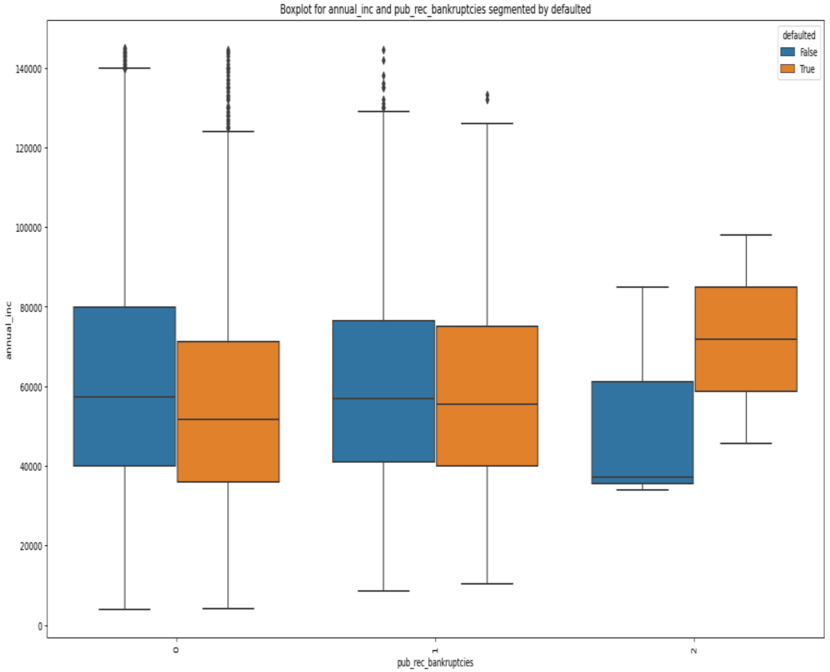


Debt consolidation and credit card borrowers are taking out more loans compared to other groups. Additionally, the default rate for loans taken by small business and renewable energy borrowers exceeds the overall default rate of 14.58%.

ANALYSIS ON PUBLIC BANKRUPTCIES



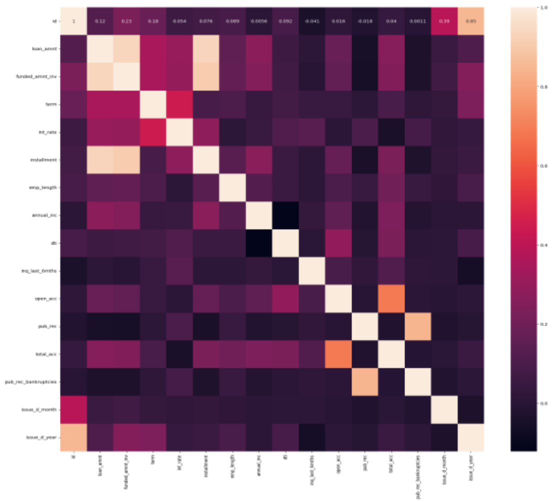




Borrowers with lower interest rates and a history of two public bankruptcies are defaulting, as are those with higher instalment amounts and the same bankruptcy record. Similarly, borrowers with higher incomes who have undergone two public bankruptcies are also defaulting on their loans.

CORRELATION MATRIX

	id	loan_amnt	funded_amnt_inv	term	int_rate	installment	emp_length	annual_inc	dti	inq_last_6mths	open_acc	pub_rec	total_acc	pub_rec_bankruptcies	issue_d_month	issue_d_year
id	1.000000	0.120614	0.231604	0.176405	0.053936	0.076088	0.089479	0.005572	0.091785	-0.041021	0.016256	-0.017683	0.039902	0.001093	0.386712	0.846859
loan_amnt	0.120614	1.000000	0.937922	0.346650	0.301265	0.932260	0.157440	0.268999	0.062436	0.012940	0.177200	-0.049997	0.256179	-0.035141	0.051902	0.109814
funded_amnt_inv	0.231604	0.937922	1.000000	0.343923	0.297473	0.905464	0.165543	0.251981	0.070663	-0.002800	0.162738	-0.051470	0.242715	-0.036327	0.068091	0.252332
term	0.176405	0.346650	0.343923	1.000000	0.440206	0.090456	0.102564	0.043866	0.076182	0.047708	0.046162	0.010266	0.096168	0.016690	0.033173	0.237704
int_rate	0.053936	0.301265	0.297473	0.440206	1.000000	0.277203	0.013762	0.048899	0.110913	0.133362	0.006022	0.098635	-0.046539	0.084016	0.025285	0.039417
installment	0.076088	0.932260	0.905464	0.090456	0.277203	1.000000	0.132876	0.267842	0.052038	0.011014	0.172893	-0.045706	0.229860	-0.033038	0.029439	0.053066
emp_length	0.089479	0.157440	0.165543	0.102564	0.013762	0.132876	1.000000	0.121792	0.050475	0.009980	0.100244	0.048763	0.201869	0.045861	0.019716	0.094719
annual_inc	0.005572	0.268999	0.251981	0.043866	0.048899	0.267842	0.121792	1.000000	-0.121530	0.035465	0.156927	-0.017864	0.234534	-0.015955	0.008980	0.008558
dti	0.091785	0.062436	0.070663	0.076182	0.110913	0.052038	0.050475	-0.121530	1.000000	0.002178	0.287849	-0.004742	0.229119	0.007315	0.014197	0.092857
inq_last_6mths	-0.041021	0.012940	-0.002800	0.047708	0.133362	0.011014	0.009980	0.035465	0.002178	1.000000	0.093434	0.023726	0.113516	0.014821	0.013356	-0.059495
open_acc	0.016256	0.177200	0.162738	0.046162	0.006022	0.172893	0.100244	0.156927	0.287849	0.093434	1.000000	0.000028	0.687260	0.005616	0.000371	0.011357
pub_rec	-0.017683	-0.049997	-0.051470	0.010266	0.098635	-0.045706	0.048763	-0.017864	-0.004742	0.023726	0.000028	1.000000	-0.023494	0.841571	-0.022398	-0.006026
total_acc	0.039902	0.256179	0.242715	0.096168	-0.046539	0.229860	0.201869	0.234534	0.229119	0.113516	0.687260	-0.023494	1.000000	-0.009998	0.002768	0.052972
pub_rec_bankruptcies	0.001093	-0.035141	-0.036327	0.016690	0.084016	-0.033038	0.045861	-0.015955	0.007315	0.014821	0.005616	0.841571	-0.009998	1.000000	-0.018413	0.011869
issue_d_month	0.386712	0.051902	0.068091	0.033173	0.025285	0.029439	0.019716	0.008980	0.014197	0.013356	0.000371	-0.022398	0.002768	-0.018413	1.000000	-0.035536
issue_d_year	0.846859	0.109814	0.252332	0.237704	0.039417	0.053066	0.094719	0.008558	0.092857	-0.059495	0.011357	-0.006026	0.052972	0.011869	-0.035536	1.000000



The term length and interest rate exhibit the strongest correlation with defaulted loans, with coefficients of 0.17 and 0.21 respectively.

RECOMMENDATIONS

- 1. Loan amount greater than 15000$ have higher default rate.
- 2. Funded amount invested greater than 15000$ have higher default rate.
- 3. With the increase of interest rate the default rate also increases.
- 4. Applicants with higher annual income have more chances not defaulting.
- 5. Default rate increases with DTI.
- 6. Longer term (60 months) have more default rate than lower term(36 months).
- 7. As grade decreases default rate increases (Same for sub grade).
- 8. Fun fact : Verified applicants have higher default rate.
- 9. Small business borrowers have higher default rate .

THANK YOU