Lending Club case study

Submitted by
Rakesh C

CONTENTS

- About the company
- Study objectives
- Methodology
- Outcome (Conclusion)
- Technology used
- Acknowledgement

LENDING CLUB

Company profile

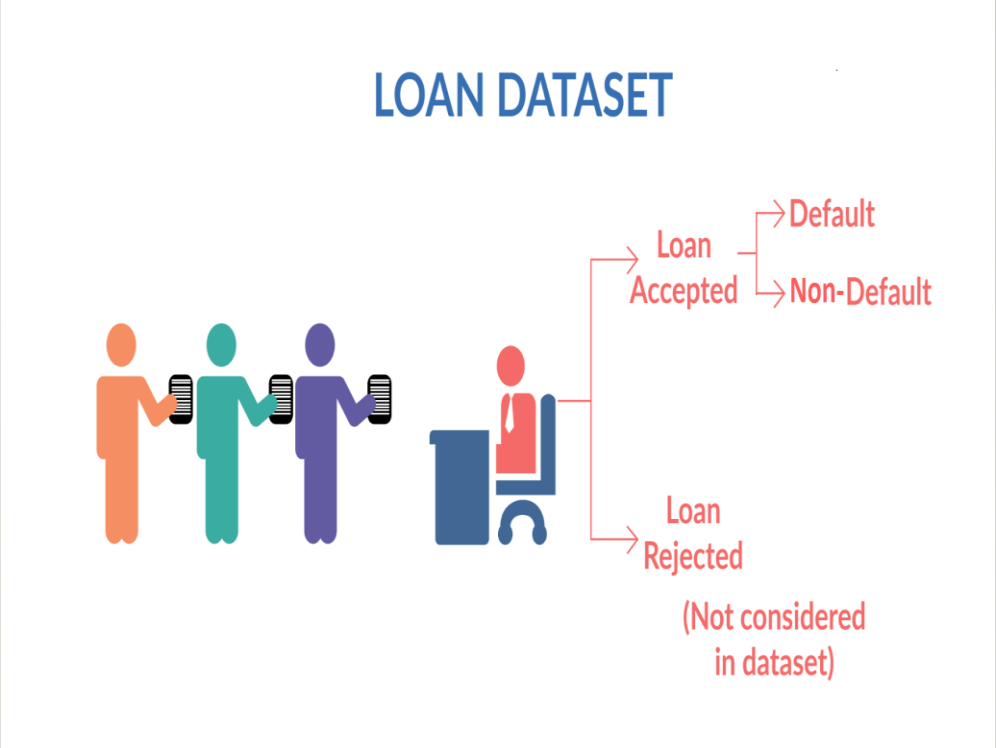
This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower-interest-rate loans through a fast online interface.

STUDY OBJECTIVES

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:

1) If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company.

2) 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.



METHODOLOGY

1)The data in the .csv format was imported into jupyter notebook, for further analysis
2)Data cleaning- null data and inadequate data were filtered and eliminated
3)Data visualization- the correlations and EDA analysis performed on the data to optimize for the desired objective
4)Data validation was conducted through visualization and priorities from EDA.

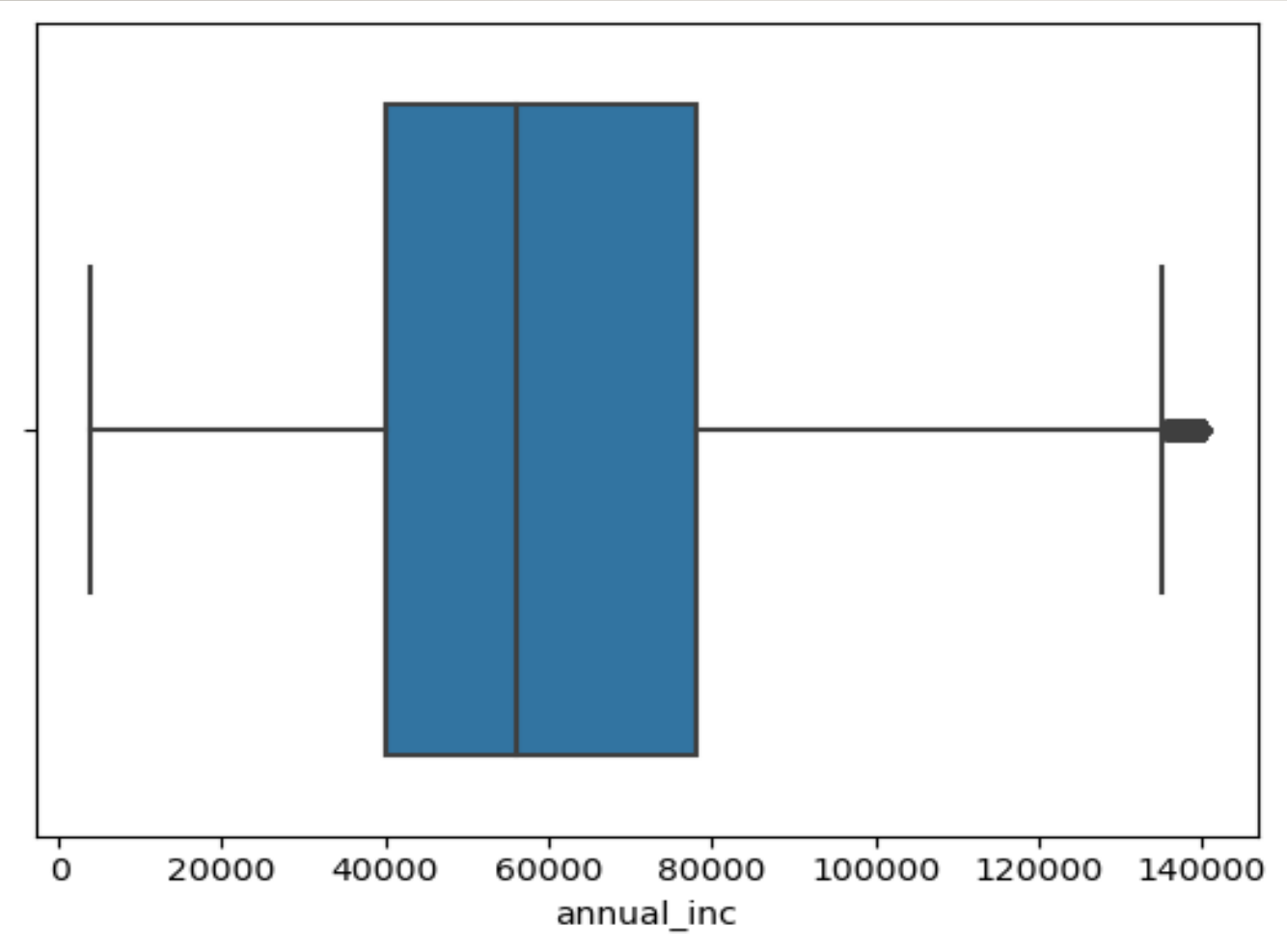
Data cleaning

	loan_amnt	term	int_rate	installment	grade	sub_grade	emp_length	home_ownership	annual_inc	issue_d	...	dti	ea
0	5000	36 months	10.65	162.87	B	2	10	RENT	24000.0	Dec-11	...	27.65	
1	2500	60 months	15.27	59.83	C	4	0	RENT	30000.0	Dec-11	...	1.00	
2	2400	36 months	15.96	84.33	C	5	10	RENT	12252.0	Dec-11	...	8.72	
3	10000	36 months	13.49	339.31	C	1	10	RENT	49200.0	Dec-11	...	20.00	
5	5000	36 months	7.90	156.46	A	4	3	RENT	36000.0	Dec-11	...	11.20	

5 rows × 22 columns

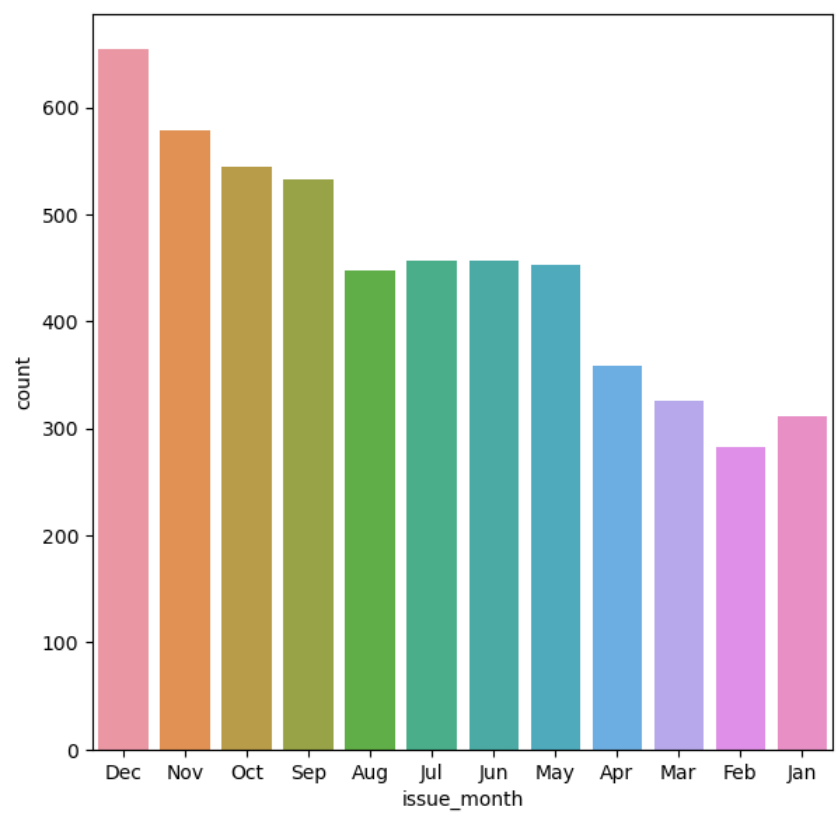
The goal of the analysis is to find the defaulters, in order to do that remove the inadequate columns and data which doesn't impact much and has the same weightage

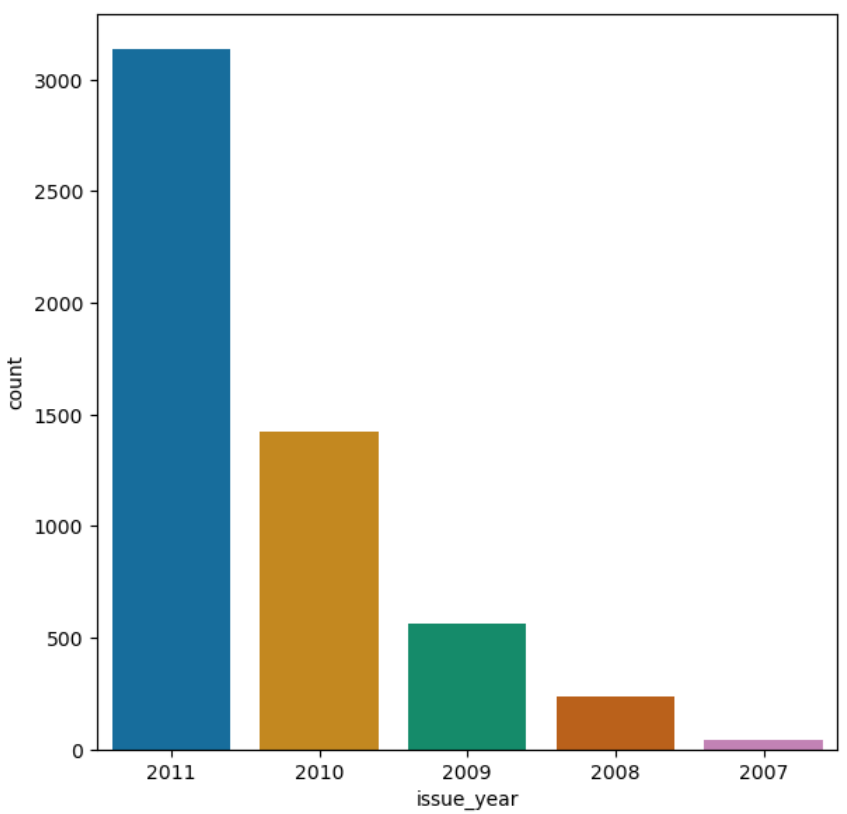
Data standardizing



Removing the outliners with constraints not less than 0.95 and analyze the data which has to be prioritized

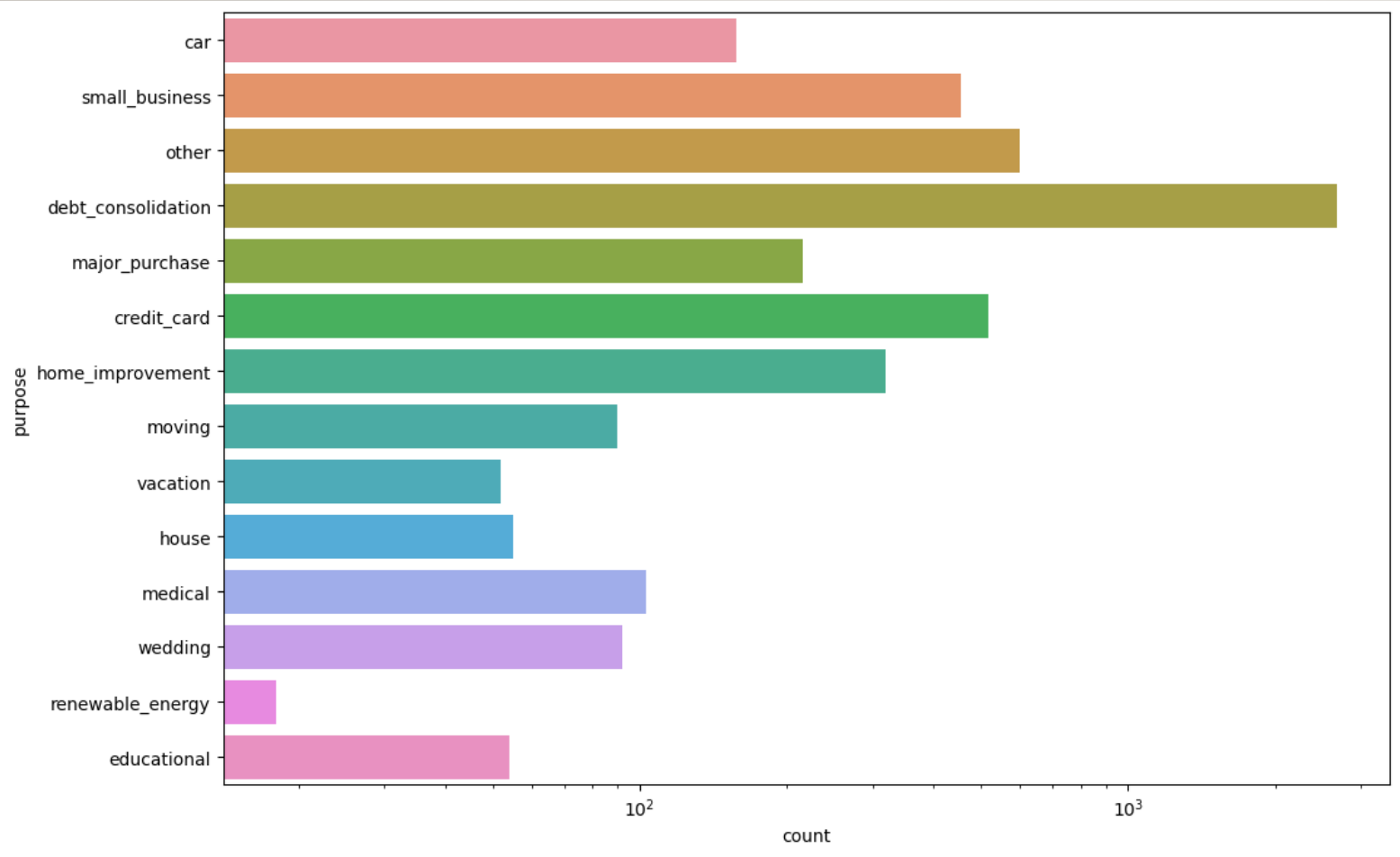
Data analysis and visualization -EDA





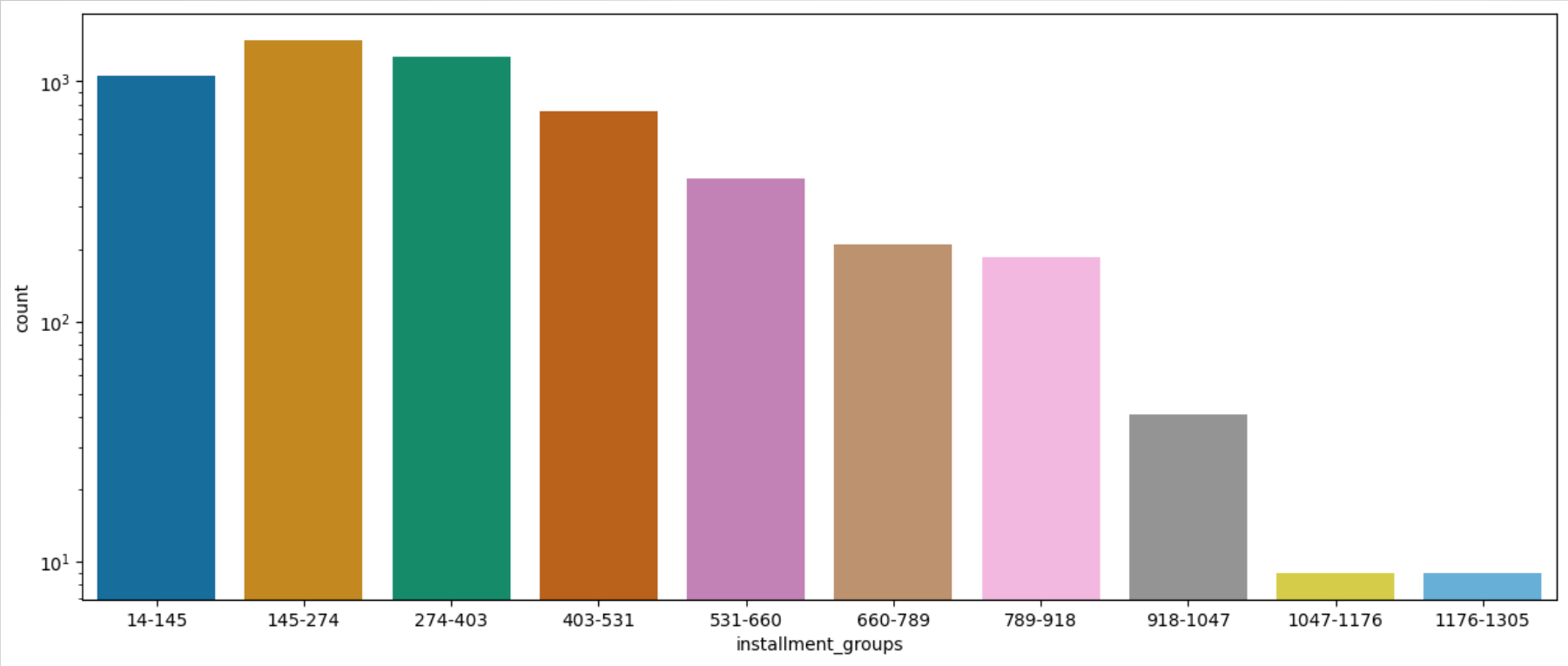
The loan application was maximum for the year 2011, therefore data related to 2011 were analyzed and the count of more than 2500 was prioritized

Data analysis and visualization -EDA



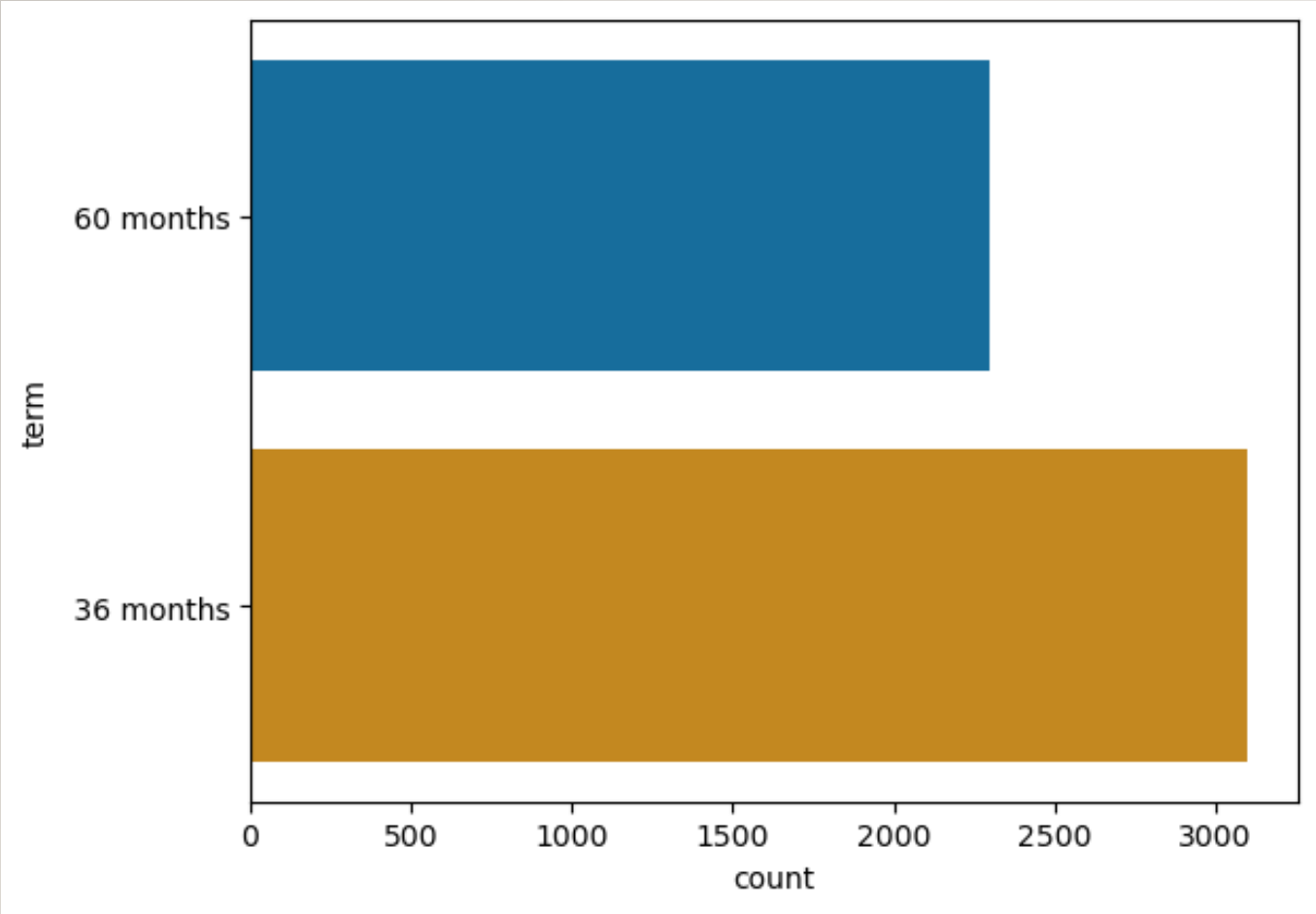
The maximum loan was taken for debt consolidation, compared with the other entities

Data analysis and visualization -EDA



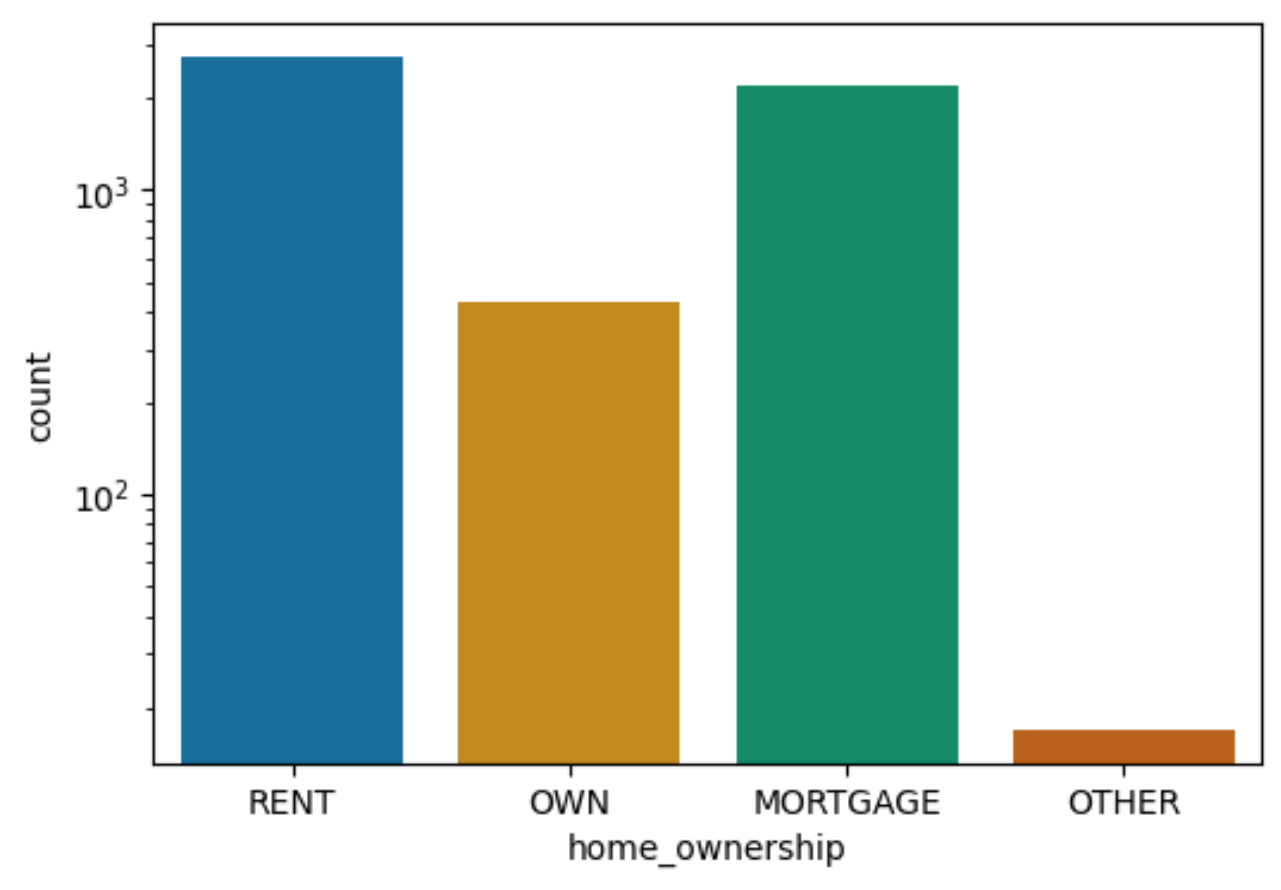
The maximum installment group is in the range of 145-274, this category has to be monitored

Data analysis and visualization -EDA



The 36 monthly term count are considerably higher compared to 60 term, hence this short duration count has to be noticed

Data analysis and visualization -EDA



The Rent and mortgage group are higher, hence they have to be noticed before the approvals of loan

Conclusion

analysis for the charge of loan is as per the following conditions:

The probability of defaulting occurs at:

- Applicants having house_ownership as 'RENT' and 'MORTGAGE’,
- Applicants use the loan to clear other debts,
- Applicants receive interest at the rate of 13-17%,
- Applicants have an income of range 31K to 58K,
- Applicants have 20-37 open_acc,
- Applicants employement length of 10, the Loan amount is between 5429 - 10357, Dti is between 12-18

When monthly installments are between 145-274, there are more defaulters in 36 months,(when no of enquires in the last six months is zero, When the number of derogatory public records is 0, and When the purpose is 'debt_consolidation'

Technology used

Library
Notebook
Matplotlib
Notebook_shim
NumPy
Openpyxl
Pandas
NumPy
Plotly
Git
GitHub

Acknowledgement

I would like to thank for the extended support provided by the following members

1. Lending club
2. M/s Upgrad
3. Faculty experts from IIIT-B
4. Support team