Lending Club Study

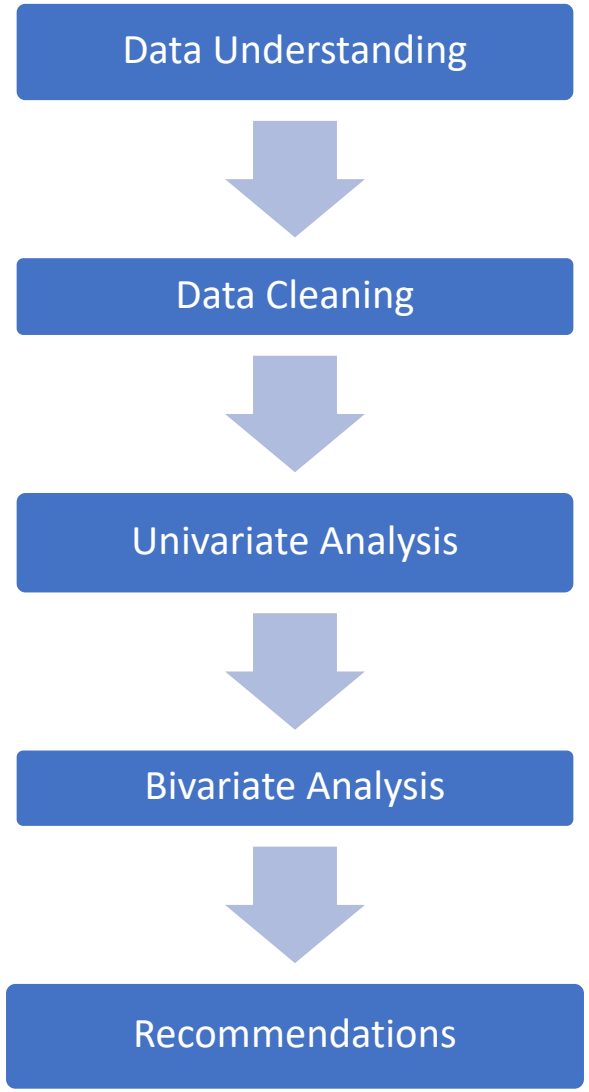
Group Member:

Nishit Goyal
Sonigara Chetan

Problem Statement

- Consumer finance company faces a critical challenge in making informed loan approval decisions, as these decisions are fraught with two distinct risks: denying a loan to a creditworthy applicant, resulting in a loss of potential business, or approving a loan to a high-risk applicant, potentially leading to financial losses due to default.
- To mitigate these risks, we need to leverage Exploratory Data Analysis (EDA) techniques on historical loan applicant data. This analysis aims to uncover meaningful patterns and insights within the dataset, specifically focusing on consumer attributes and loan attributes. By doing so, we seek to identify key indicators that correlate with the likelihood of loan default.
- Ultimately, the objective is to develop a robust understanding of how various factors influence the propensity for default, which will enable us to make data-driven decisions such as denying loans to high-risk applicants, adjusting loan amounts, or offering loans at higher interest rates when necessary.

Flowchart for Mitigating Loan Default Risk through EDA



Data Understanding

Data Understanding and Observations

- There are 39717 rows, 111 columns present in loan.csv
- Null or NA values are present in all rows for 54 columns in the dataset.
- The dataset contains mixed data types, primarily consisting of numeric and object types. To facilitate our analysis, we will convert certain data types to numeric.
- We will prioritize the analysis of key columns, such as loan_amnt, dti, loan_status, annual_inc, and others with significant relevance.
- Some key columns exhibit outliers, and our objective is to address and remove these outliers for more robust analysis
- Some columns like int_rate needs to be transform for further analysis

Data Cleaning

Removing Unnecessary Columns

- Following columns have either all null or NA values.We can drop these columns

- 'mths_since_last_major_derog', 'annual_inc_joint', 'dti_joint', 'verification_status_joint', 'tot_coll_amt', 'tot_cur_bal', 'open_acc_6m', 'open_il_6m', 'open_il_12m', 'open_il_24m', 'mths_since_rcnt_il', 'total_bal_il', 'il_util', 'open_rv_12m', 'open_rv_24m', 'max_bal_bc', 'all_util', 'total_rev_hi_lim', 'inq_fi', 'total_cu_tl', 'inq_last_12m', 'acc_open_past_24mths', 'avg_cur_bal', 'bc_open_to_buy', 'bc_util', 'mo_sin_old_il_acct', 'mo_sin_old_rev_tl_op', 'mo_sin_rcnt_rev_tl_op', 'mo_sin_rcnt_tl', 'mort_acc', 'mths_since_recent_bc', 'mths_since_recent_bc_dlq', 'mths_since_recent_inq', 'mths_since_recent_revol_delinq', 'num_accts_ever_120_pd', 'num_actv_bc_tl', 'num_actv_rev_tl', 'num_bc_sats', 'num_bc_tl', 'num_il_tl', 'num_op_rev_tl', 'num_rev_accts', 'num_rev_tl_bal_gt_0', 'num_sats', 'num_tl_120dpd_2m', 'num_tl_30dpd', 'num_tl_90g_dpd_24m', 'num_tl_op_past_12m', 'pct_tl_nvr_dlq', 'percent_bc_gt_75', 'tot_hi_cred_lim', 'total_bal_ex_mort', 'total_bc_limit', 'total_il_high_credit_limit,acc_now_delinq', 'delinq_amnt'

- Some Columns contains value like 0.0 or NAN after analyzing them. Our next step is to remove these columns as they do not contribute

	collections_12_mths_ex_med	policy_code	application_type	\
0	0.00	1	INDIVIDUAL	
1	0.00	1	INDIVIDUAL	
2	0.00	1	INDIVIDUAL	
3	0.00	1	INDIVIDUAL	
4	0.00	1	INDIVIDUAL	
...	...	...	...	
39712	NaN	1	INDIVIDUAL	
39713	NaN	1	INDIVIDUAL	
39714	NaN	1	INDIVIDUAL	
39715	NaN	1	INDIVIDUAL	
39716	NaN	1	INDIVIDUAL	

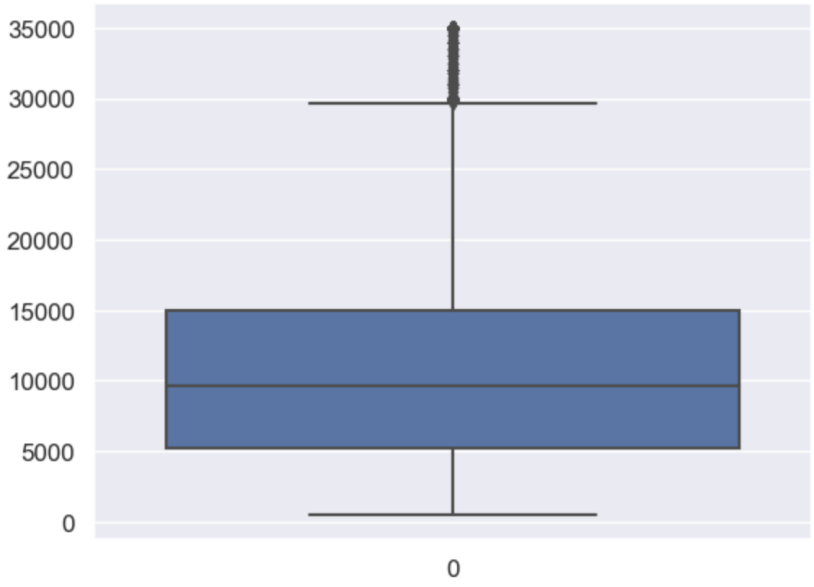
	chargeoff_within_12_mths	pub_rec_bankruptcies	tax_liens
0	0.00	0.00	0.00
1	0.00	0.00	0.00
2	0.00	0.00	0.00
3	0.00	0.00	0.00
4	0.00	0.00	0.00
...	...	...	...
39712	NaN	NaN	NaN
39713	NaN	NaN	NaN
39714	NaN	NaN	NaN
39715	NaN	NaN	NaN
39716	NaN	NaN	NaN

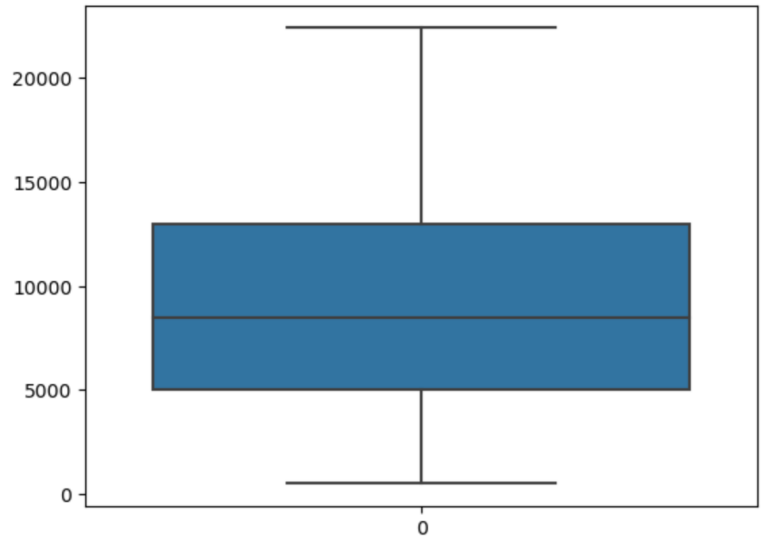
- Some columns like emp_length and int_rate needs to be transform to proper format for further analysis
- Example: emp_length

Before	After
<1 month	0
36 month	36

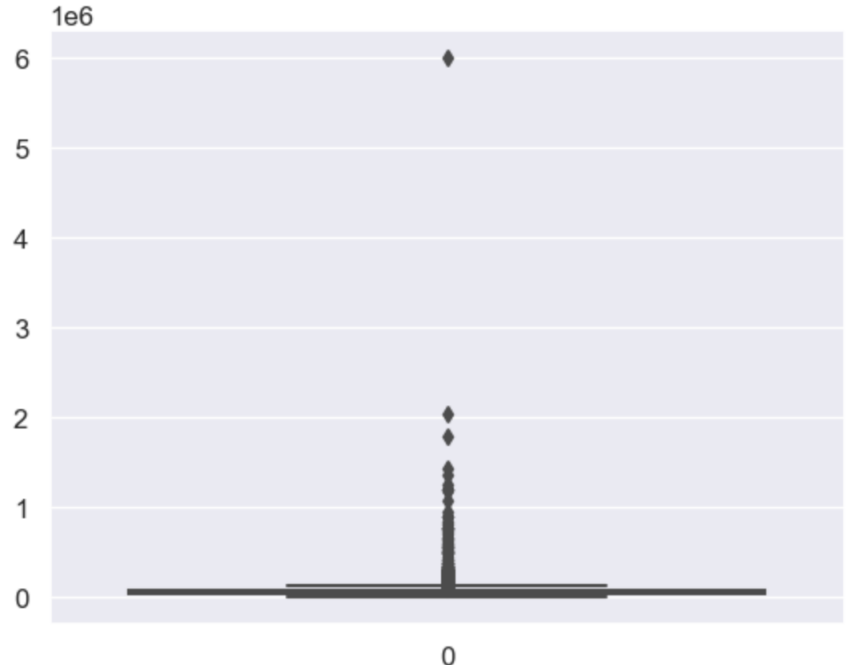
Removing Outliers

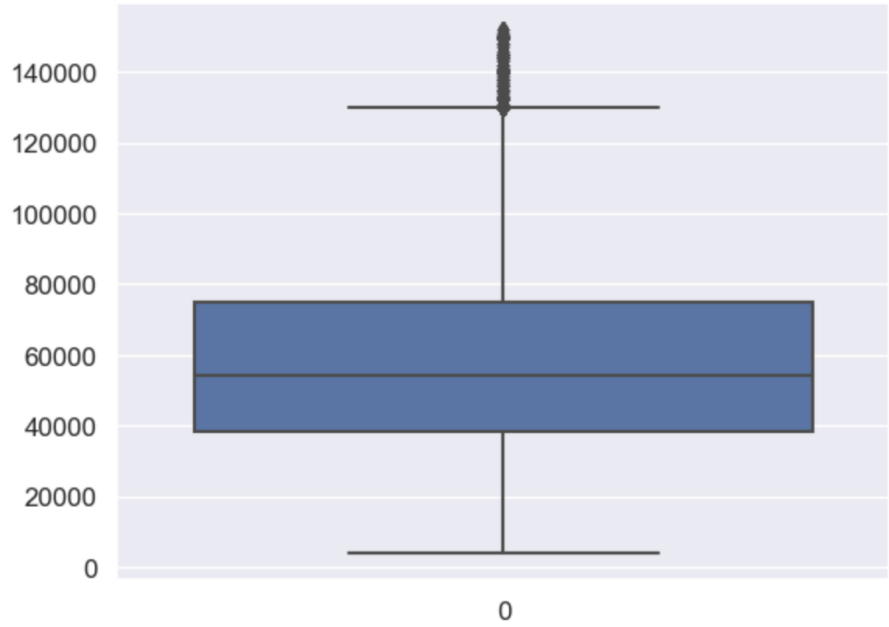
- Loan_amnt
- Box plot of loan_amnt before and after removing outliers





- Annual_inc before and after removing outlier

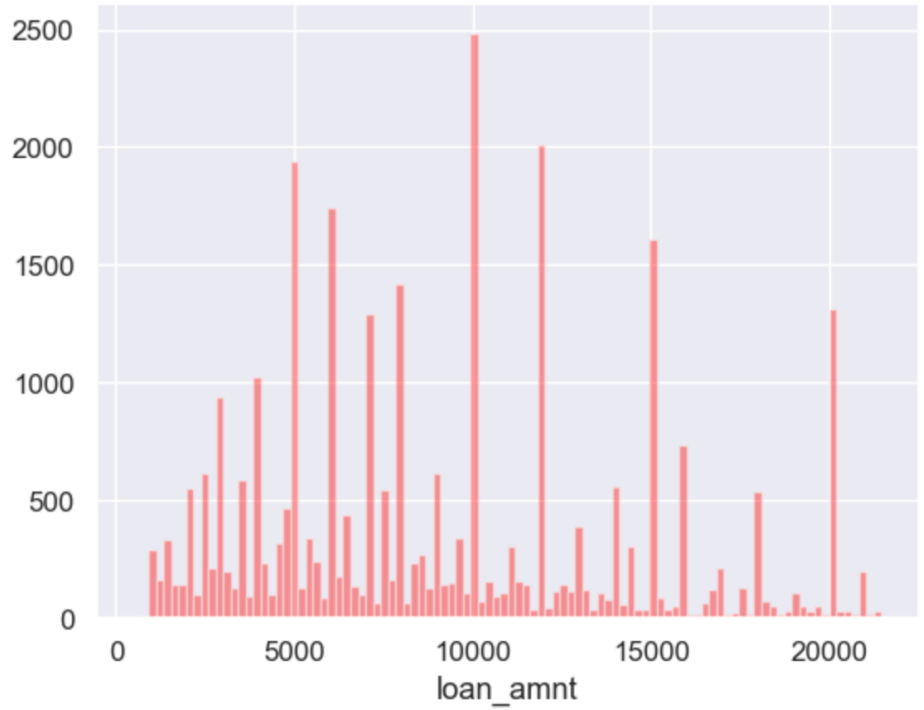




Univariate Analysis

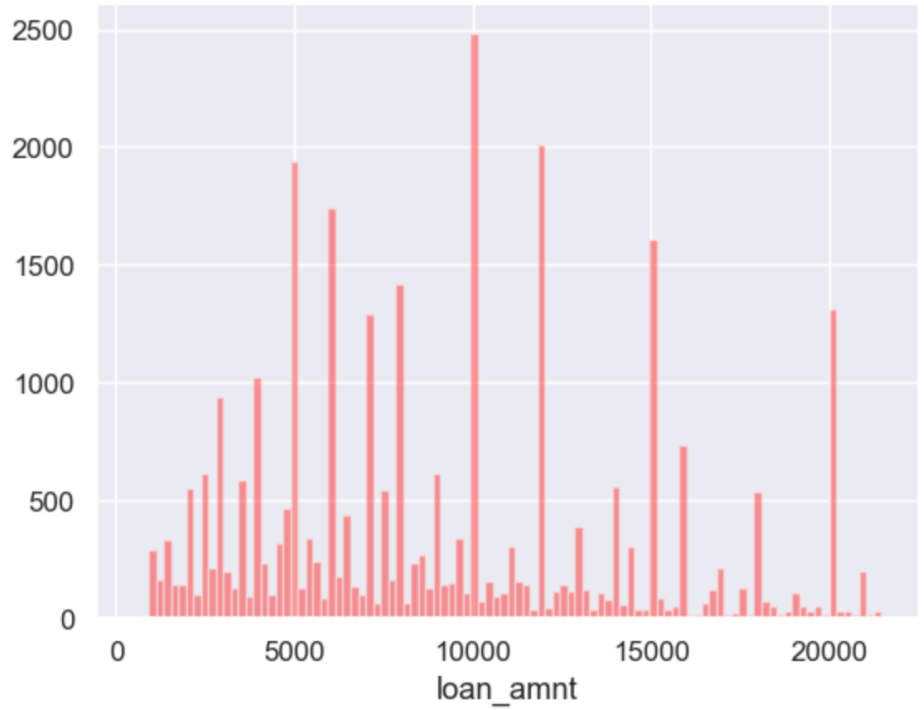
Loan Amount

- Mostly people applying for loan has annual income between 25k-80k



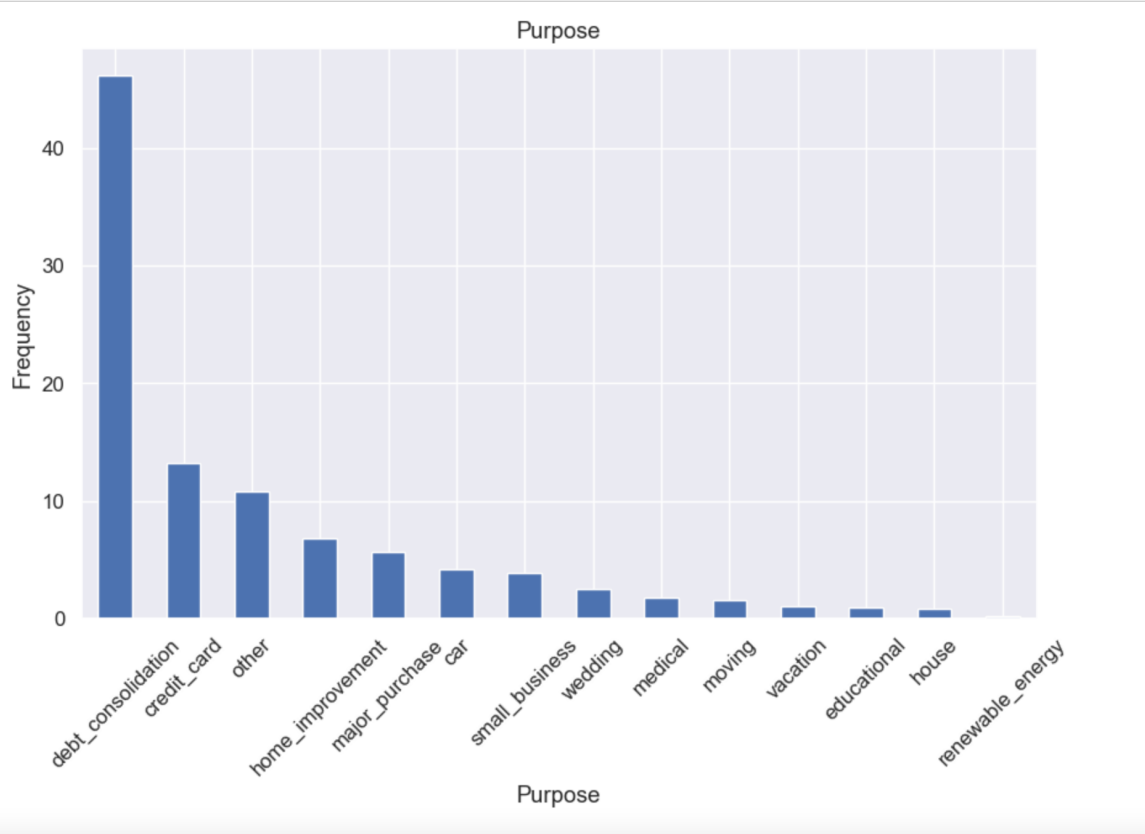
Loan Status

- Around 15-20% of total loans are getting charged off



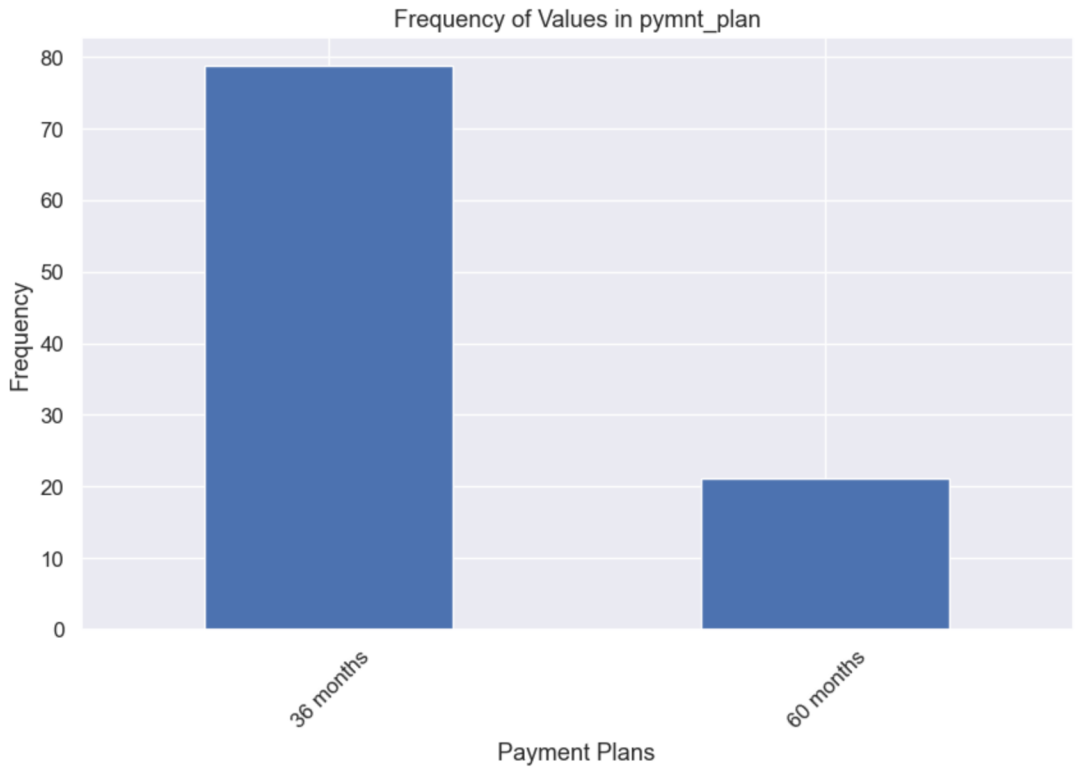
Purpose

- Most people are taking loan for debt consolidation



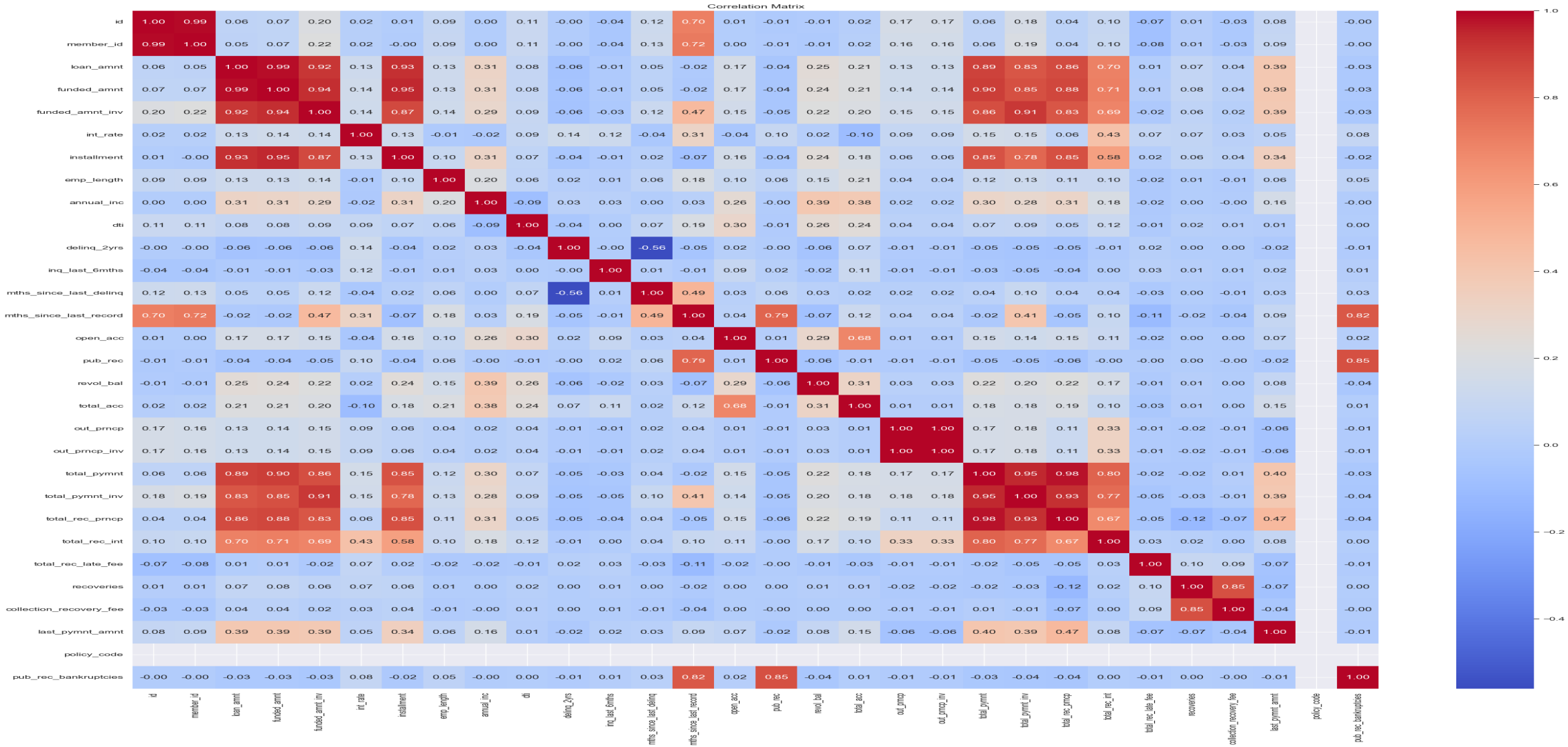
Term Of Loan

- Most Loans are for short term



Bivariate Analysis

Correlation Matrix



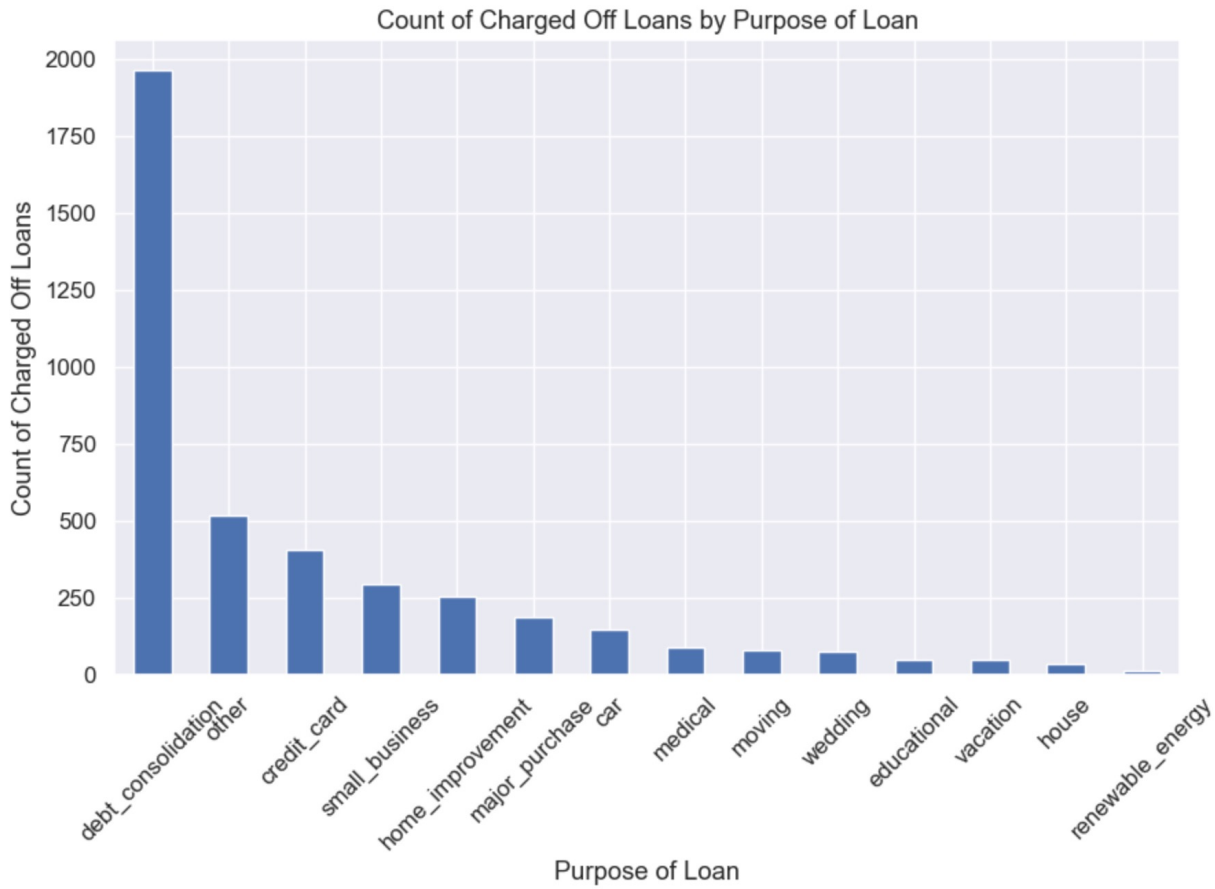
Observations from Correlation Matrix

- Correlation matrix for loan data
- loan amount,funded_amnt,funded_amnt_inv,installment -- strongly correlated with each other
- total_pymnt,total_pymnet_iv,total_rec_prcp are storngly correlated with each other
- loan_amnt is negatively correlated with delinq_2yrs and pub_rec

- int_rate is negatively correlated total_Acc
- delinq_2yrs is negatively correlated with mnths_since_last_deling

Charge off loan's vs Purpose of Loan

Most charged off loans are for debt consolidation purpose

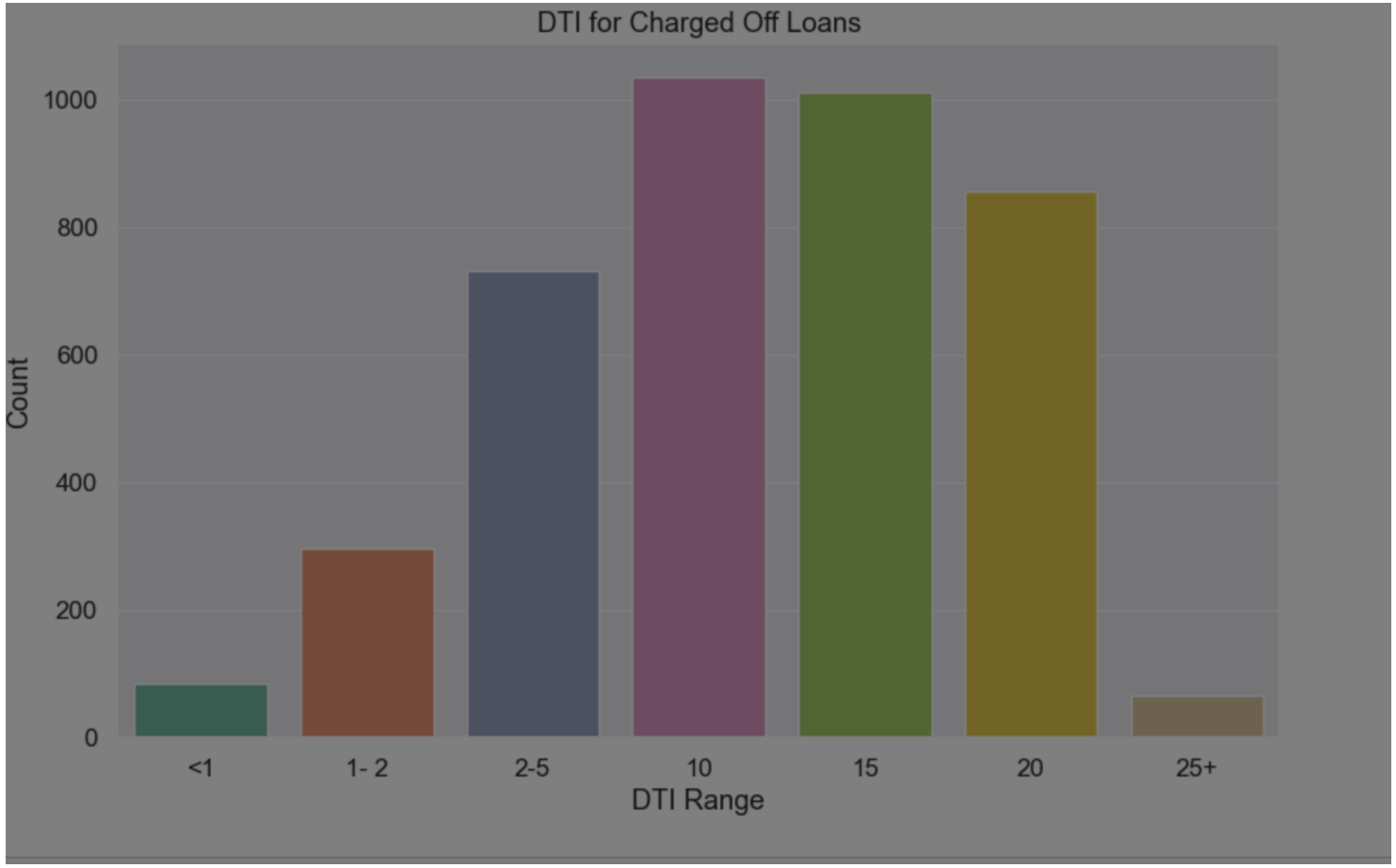


Salary Range for Charge off loans

People having salary less than 40k are having more charged off loans

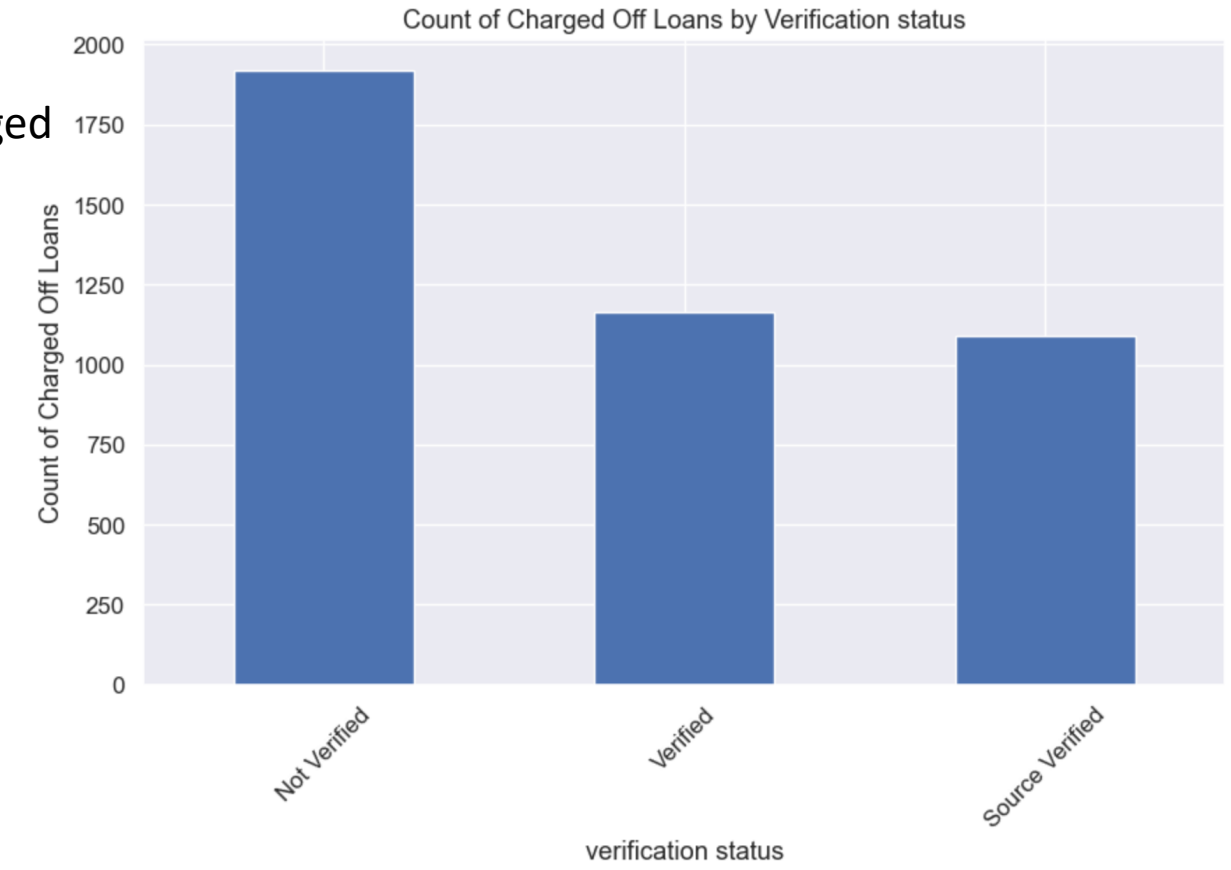
DTI vs Charge off loans

Higher the DTI,more are chances of charged Off loans



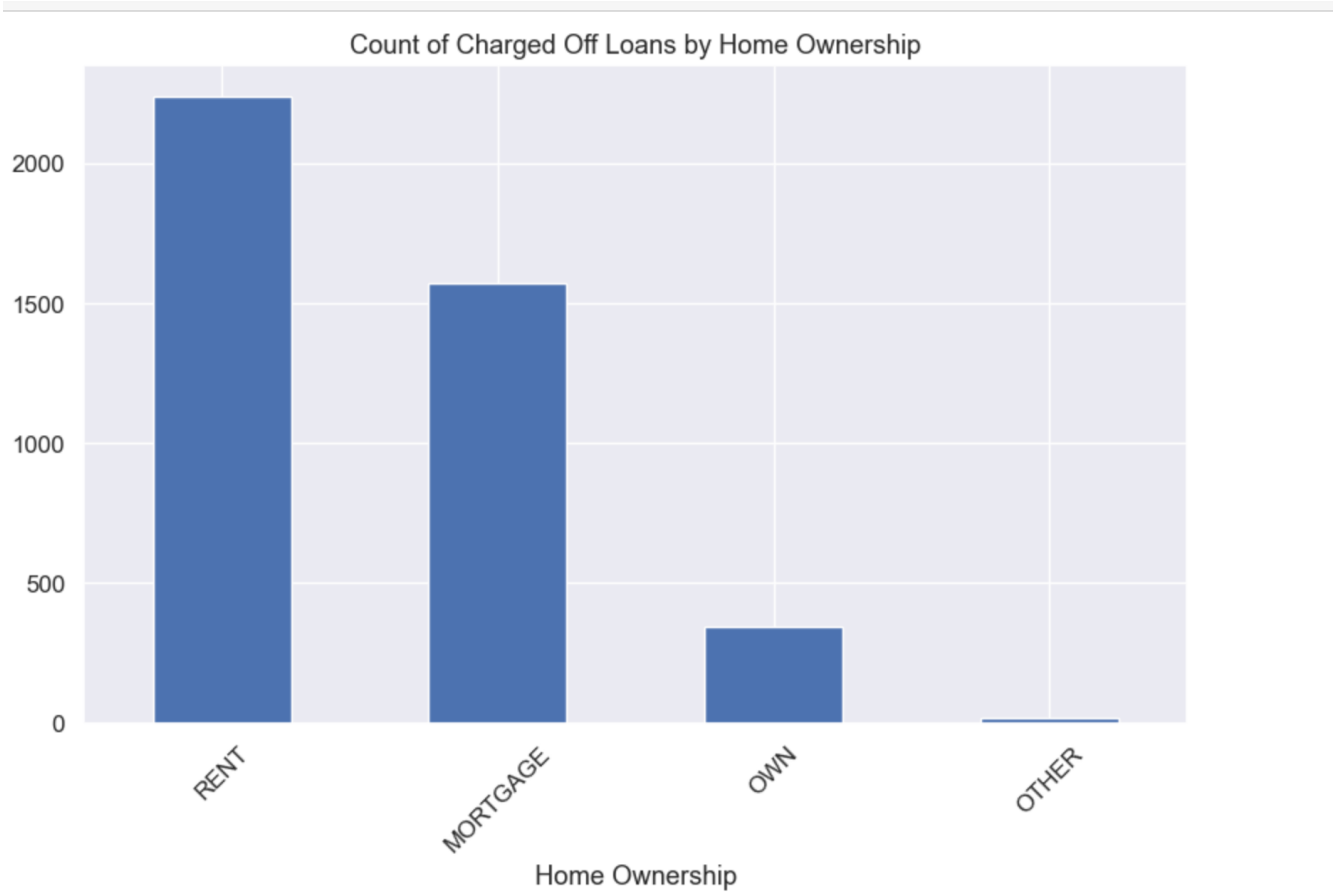
Verification vs Charge Off Loans

Loans which are not verified ,are getting charged Off more likely



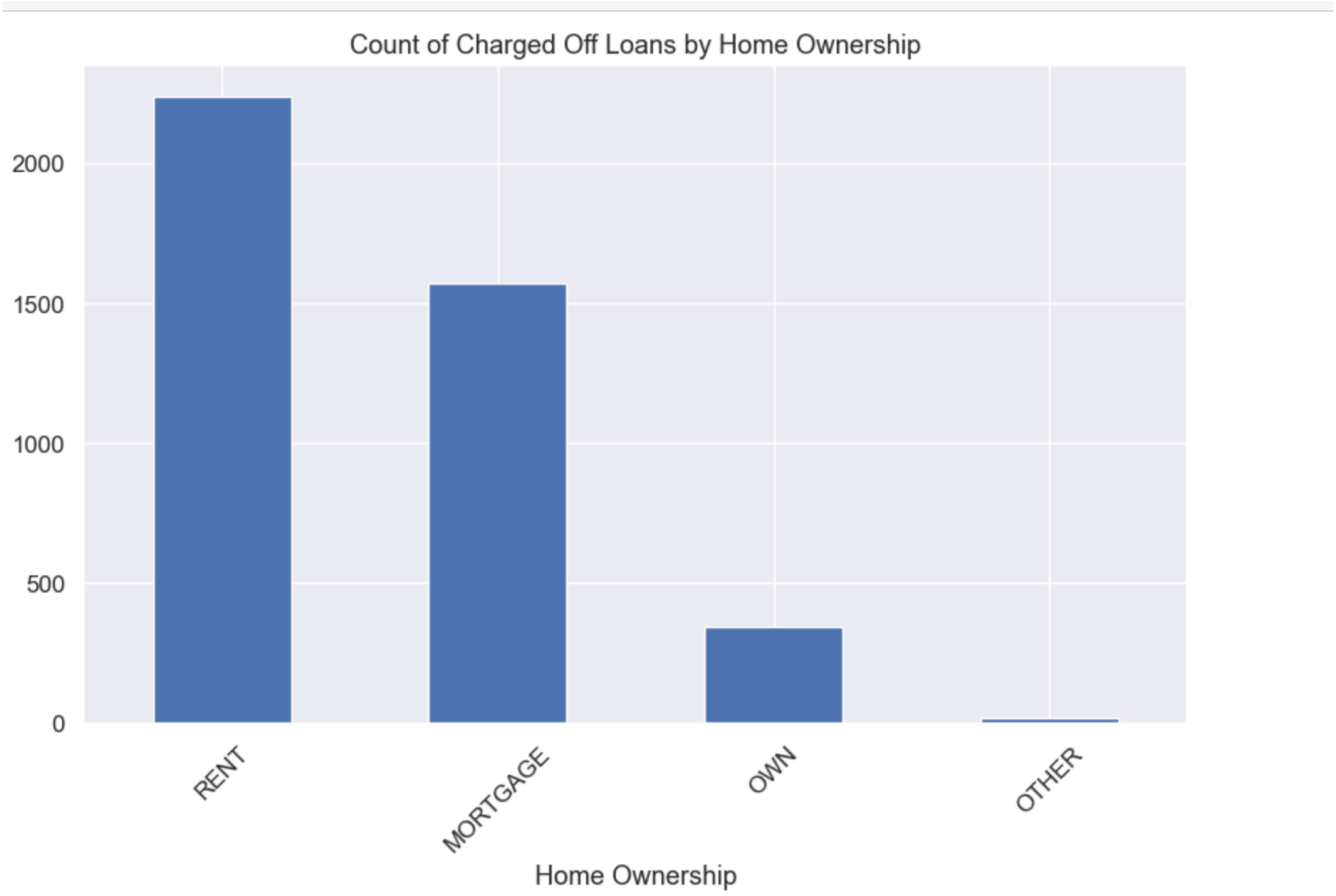
Home membership vs Charge Off Loans

People having their own house are having
Less charged off loans



Grade vs Charge Off Loans

Lower the grade, more the chances of loan to get Charged off



Recommendations

1)Lending Club can significantly reduce the risk of charge-offs by implementing thorough verification processes for all loan applications.

2)To mitigate risk, Lending Club should consider limiting the number of loans issued for the purpose of debt consolidation.

3)Implementing restrictions on loans to individuals who do not own a house can help minimize risk for Lending Club.

4)Applicants with a higher Debt-to-Income Ratio (DTI) appear to have a higher incidence of loan charge-offs, suggesting the importance of managing such loans more cautiously.

5)To reduce the likelihood of charge-offs, Lending Club should exercise caution when granting loans with poor credit grades, as lower grades are associated with a higher risk of default.

6)It's advisable for Lending Club to refrain from granting large loans to individuals with a public record (pub_rec) as this appears to be negatively correlated with loan performance