Lending Club Study

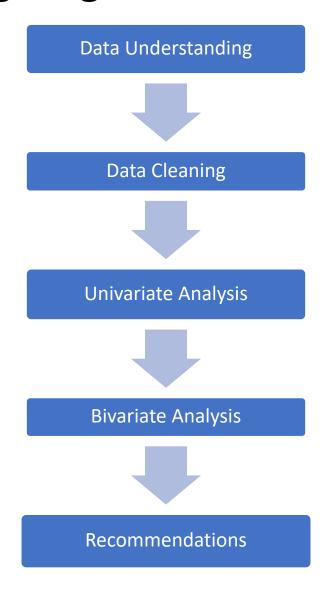
Group Member:

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
                                                          INDIVIDUAL
1
                               0.00
                                                 1
                                                          INDIVIDUAL
2
                               0.00
                                                          INDIVIDUAL
3
                                0.00
                                                          INDIVIDUAL
4
                                                 1
                                0.00
                                                          INDIVIDUAL
39712
                                 NaN
                                                 1
                                                          INDIVIDUAL
                                                 1
39713
                                 NaN
                                                          INDIVIDUAL
39714
                                 NaN
                                                          INDIVIDUAL
39715
                                                 1
                                 NaN
                                                          INDIVIDUAL
39716
                                                 1
                                 NaN
                                                          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
```

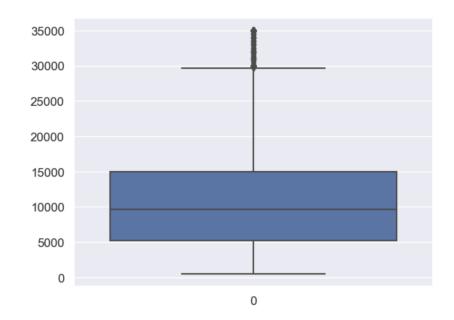
 Some columns like emp_length and int_rate needs to be transform to proper format for further analysis

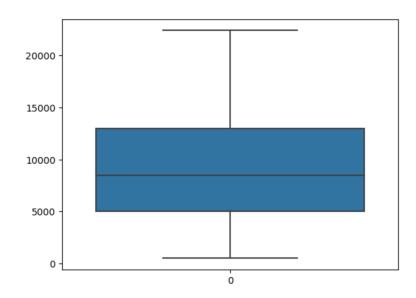
• Example: emp_length

Before	After
<1 month	0
36 month	36

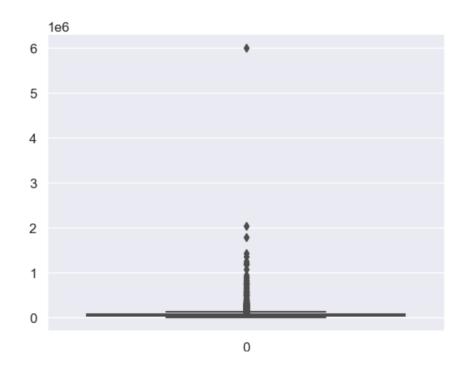
Removing Outliers

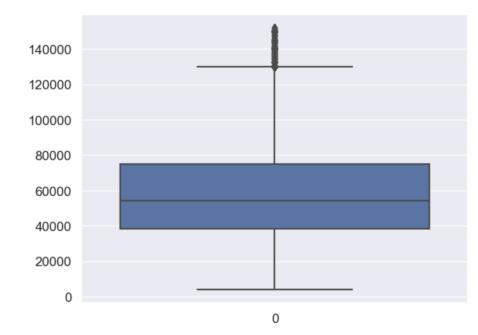
- Loan_amnt
- Box plot of loan_amnt before and after removing outliers





Annual_inc before and after removing outlier

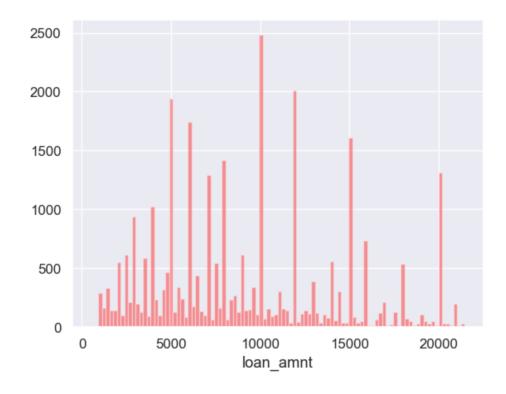




Univariate Analysis

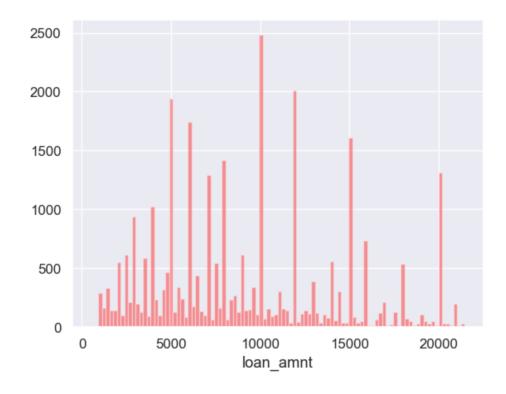
Loan Amount

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



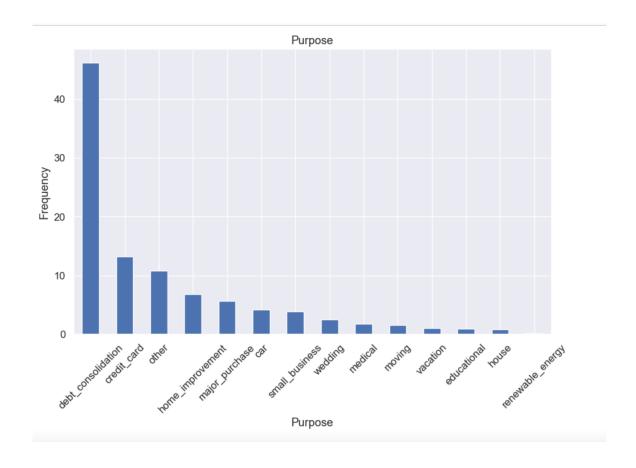
Loan Status

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



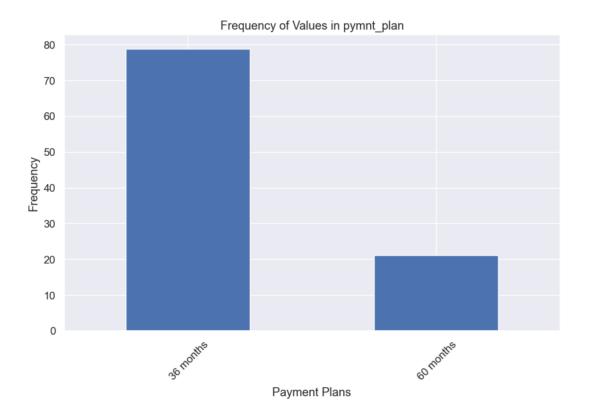
Purpose

Most people are taking loan for debt consolidation



Term Of Loan

Most Loans are for short term



Bivariate Analysis

Correlation Matrix

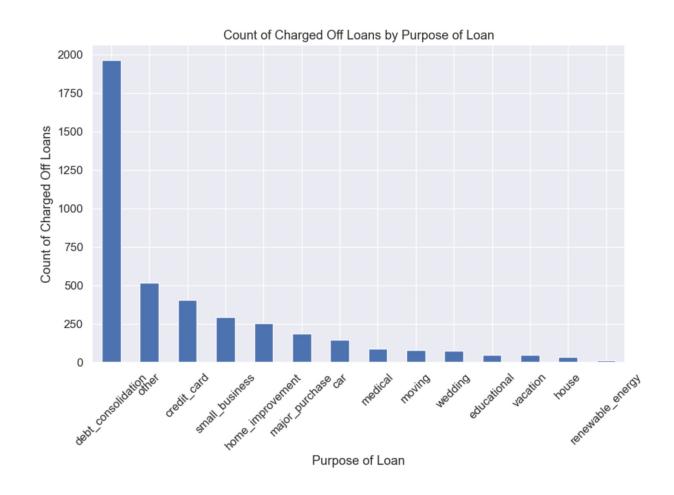
		Correlation Matrix																												
id	1.00	0.99	0.06	0.07	0.20	0.02	0.01	0.09	0.00	0.11	-0.00	-0.04	0.12		0.01	-0.01	-0.01	0.02	0.17	0.17	0.06	0.18	0.04	0.10	-0.07	0.01	-0.03	0.08		-0.00
member_id	0.99	1.00	0.05	0.07	0.22	0.02	-0.00	0.09	0.00	0.11	-0.00	-0.04	0.13		0.00	-0.01	-0.01	0.02	0.16	0.16	0.06	0.19	0.04	0.10	-0.08	0.01	-0.03	0.09		-0.00
loan_amnt	0.06	0.05	1.00	0.99		0.13	0.93	0.13	0.31	0.08	-0.06	-0.01	0.05	-0.02	0.17	-0.04	0.25	0.21	0.13	0.13					0.01	0.07	0.04	0.39		-0.03
funded_amnt	0.07	0.07	0.99	1.00	0.94	0.14	0.95	0.13	0.31	0.08	-0.06	-0.01	0.05	-0.02	0.17	-0.04	0.24	0.21	0.14	0.14					0.01	0.08	0.04	0.39		-0.03
funded_amnt_inv	0.20	0.22	0.92	0.94	1.00	0.14	0.87	0.14	0.29	0.09	-0.06	-0.03	0.12	0.47	0.15	-0.05	0.22	0.20	0.15	0.15					-0.02	0.06	0.02	0.39		-0.03
int_rate	0.02	0.02	0.13	0.14	0.14	1.00	0.13	-0.01	-0.02	0.09	0.14	0.12	-0.04	0.31	-0.04	0.10	0.02	-0.10	0.09	0.09	0.15	0.15	0.06	0.43	0.07	0.07	0.03	0.05		0.08
installment	0.01	-0.00	0.93	0.95		0.13	1.00	0.10	0.31	0.07	-0.04	-0.01	0.02	-0.07	0.16	-0.04	0.24	0.18	0.06	0.06				0.58	0.02	0.06	0.04	0.34		-0.02
emp_length	0.09	0.09	0.13	0.13	0.14	-0.01	0.10	1.00	0.20	0.06	0.02	0.01	0.06	0.18	0.10	0.06	0.15	0.21	0.04	0.04	0.12	0.13	0.11	0.10	-0.02	0.01	-0.01	0.06		0.05
annual_inc	0.00	0.00	0.31	0.31	0.29	-0.02	0.31	0.20	1.00	-0.09	0.03	0.03	0.00	0.03	0.26	-0.00	0.39	0.38	0.02	0.02	0.30	0.28	0.31	0.18	-0.02	0.00	-0.00	0.16		-0.00
dti	0.11	0.11	0.08	0.08	0.09	0.09	0.07	0.06	-0.09	1.00	-0.04	0.00	0.07	0.19	0.30	-0.01	0.26	0.24	0.04	0.04	0.07	0.09	0.05	0.12	-0.01	0.02	0.01	0.01		0.00
delinq_2yrs	-0.00	-0.00	-0.06	-0.06	-0.06	0.14	-0.04	0.02	0.03	-0.04	1.00	-0.00	-0.56	-0.05	0.02	-0.00	-0.06	0.07	-0.01	-0.01	-0.05	-0.05	-0.05	-0.01	0.02	0.00	0.00	-0.02		-0.01
inq_last_6mths	-0.04	-0.04	-0.01	-0.01	-0.03	0.12	-0.01	0.01	0.03	0.00	-0.00	1.00	0.01	-0.01	0.09	0.02	-0.02	0.11	-0.01	-0.01	-0.03	-0.05	-0.04	0.00	0.03	0.01	0.01	0.02		0.01
mths_since_last_deling	0.12	0.13	0.05	0.05	0.12	-0.04	0.02	0.06	0.00	0.07	-0.56	0.01	1.00	0.49	0.03	0.06	0.03	0.02	0.02	0.02	0.04	0.10	0.04	0.04	-0.03	0.00	-0.01	0.03		0.03
mths_since_last_record	0.70	0.72	-0.02	-0.02	0.47	0.31	-0.07	0.18	0.03	0.19	-0.05	-0.01	0.49	1.00	0.04	0.79	-0.07	0.12	0.04	0.04	-0.02	0.41	-0.05	0.10	-0.11	-0.02	-0.04	0.09		0.82
open_acc	0.01	0.00	0.17	0.17	0.15	-0.04	0.16	0.10	0.26	0.30	0.02	0.09	0.03	0.04	1.00	0.01	0.29	0.68	0.01	0.01	0.15	0.14	0.15	0.11	-0.02	0.00	0.00	0.07		0.02
pub_rec	-0.01	-0.01	-0.04	-0.04	-0.05	0.10	-0.04	0.06	-0.00	-0.01	-0.00	0.02	0.06		0.01	1.00	-0.06	-0.01	-0.01	-0.01	-0.05	-0.05	-0.06	-0.00	-0.00	0.00	-0.00	-0.02		0.85
revol_bal	-0.01	-0.01	0.25	0.24	0.22	0.02	0.24	0.15	0.39	0.26	-0.06	-0.02	0.03	-0.07	0.29	-0.06	1.00	0.31	0.03	0.03	0.22	0.20	0.22	0.17	-0.01	0.01	0.00	0.08		-0.04
total_acc	0.02	0.02	0.21	0.21	0.20	-0.10	0.18	0.21	0.38	0.24	0.07	0.11	0.02	0.12		-0.01	0.31	1.00	0.01	0.01	0.18	0.18	0.19	0.10	-0.03	0.01	0.00	0.15		0.01
out_prncp	0.17	0.16	0.13	0.14	0.15	0.09	0.06	0.04	0.02	0.04	-0.01	-0.01	0.02	0.04	0.01	-0.01	0.03	0.01	1.00	1.00	0.17	0.18	0.11	0.33	-0.01	-0.02	-0.01	-0.06		-0.01
out_prncp_inv	0.17	0.16	0.13	0.14	0.15	0.09	0.06	0.04	0.02	0.04	-0.01	-0.01	0.02	0.04	0.01	-0.01	0.03	0.01	1.00	1.00	0.17	0.18	0.11	0.33	-0.01	-0.02	-0.01	-0.06		-0.01
total_pymnt	0.06	0.06	0.89	0.90	0.86	0.15	0.85	0.12	0.30	0.07	-0.05	-0.03	0.04	-0.02	0.15	-0.05	0.22	0.18	0.17	0.17	1.00	0.95	0.98	0.80	-0.02	-0.02	0.01	0.40		-0.03
total_pymnt_inv	0.18	0.19	0.83			0.15	0.78	0.13	0.28	0.09	-0.05	-0.05	0.10	0.41	0.14	-0.05	0.20	0.18	0.18	0.18	0.95	1.00		0.77	-0.05	-0.03	-0.01	0.39		-0.04
total_rec_prncp	0.04	0.04	0.86			0.06	0.85	0.11	0.31	0.05	-0.05	-0.04	0.04	-0.05	0.15	-0.06	0.22	0.19	0.11	0.11	0.98		1.00	0.67	-0.05	-0.12	-0.07	0.47		-0.04
total_rec_int	0.10	0.10	0.70			0.43	0.58	0.10	0.18	0.12	-0.01	0.00	0.04	0.10	0.11	-0.00	0.17	0.10	0.33	0.33	0.80	0.77	0.67	1.00	0.03	0.02	0.00	0.08		0.00
total_rec_late_fee	-0.07	-0.08	0.01	0.01	-0.02	0.07	0.02	-0.02	-0.02	-0.01	0.02	0.03	-0.03	-0.11	-0.02	-0.00	-0.01	-0.03	-0.01	-0.01	-0.02	-0.05	-0.05	0.03	1.00	0.10	0.09	-0.07		-0.01
recoveries	0.01	0.01	0.07	0.08	0.06	0.07	0.06	0.01	0.00	0.02	0.00	0.01	0.00	-0.02	0.00	0.00	0.01	0.01	-0.02	-0.02	-0.02	-0.03	-0.12	0.02	0.10	1.00	0.85	-0.07		0.00
collection_recovery_fee	-0.03	-0.03	0.04	0.04	0.02	0.03	0.04	-0.01	-0.00	0.01	0.00	0.01	-0.01	-0.04	0.00	-0.00	0.00	0.00	-0.01	-0.01	0.01	-0.01	-0.07	0.00	0.09	0.85	1.00	-0.04		-0.00
last_pymnt_amnt	0.08	0.09	0.39	0.39	0.39	0.05	0.34	0.06	0.16	0.01	-0.02	0.02	0.03	0.09	0.07	-0.02	0.08	0.15	-0.06	-0.06	0.40	0.39	0.47	0.08	-0.07	-0.07	-0.04	1.00		-0.01
policy_code																														
pub_rec_bankruptcies	-0.00	-0.00	-0.03	-0.03	-0.03	0.08	-0.02	0.05	-0.00	0.00	-0.01	0.01	0.03	0.82	0.02	0.85	-0.04	0.01	-0.01	-0.01	-0.03	-0.04	-0.04	0.00	-0.01	0.00	-0.00	-0.01		1.00
	.12	member_id	ban_amnt	funded_amnt	funded arrort inv	int_rate	instalment	emp_length	annalinc	-6	deling_2yrs	ing_last_6mths	since last deling	since last record	oben_acc	pub_rec	levo] bal	pos acc	dound Ino	out pmop inv	btal_gymnt	total_pymnt_inv	total rec pmp	bbal rec_int	total_rec_late_fee	recoveries	cion_recovery_fee	lasi_pymnt_amnt	policy_code	rec bankruptoies

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

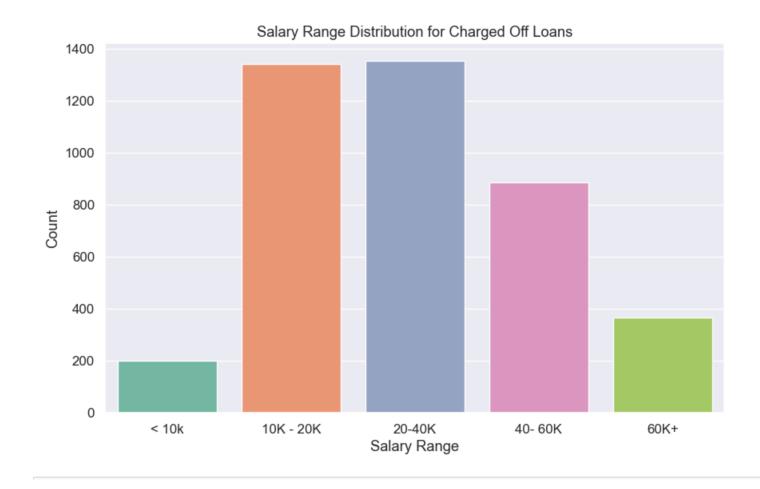
Charge off loan's vs Purpose of Loan

Most charged off loans are for debt consolidation purpose



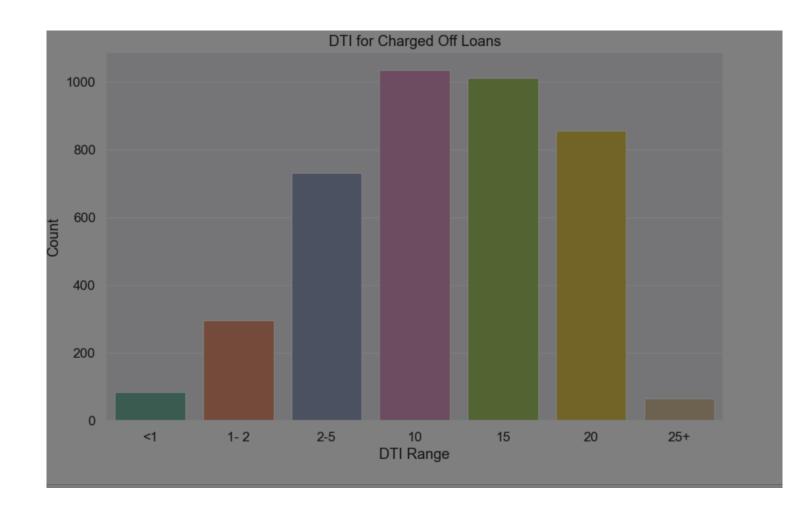
Salary Range for Charge off loans

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



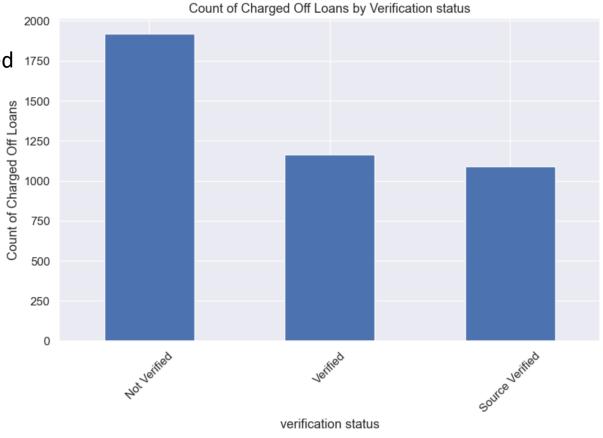
DTI vs Charge off loans

Higher the DTI, more are chances of charged Off loans



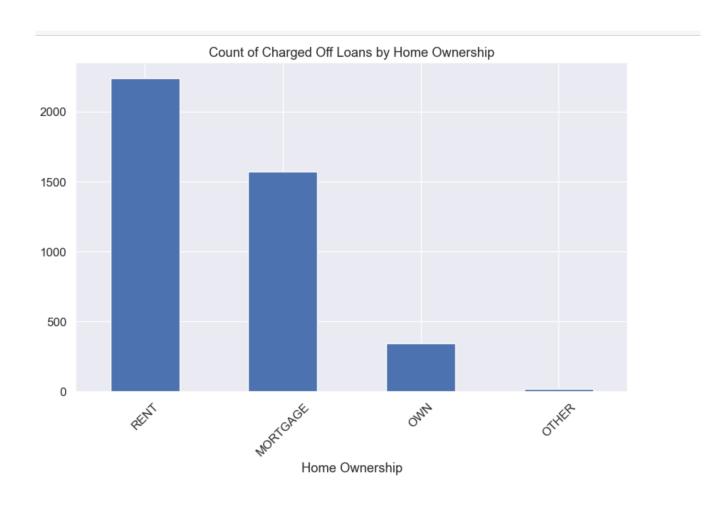
Verification vs Charge Off Loans

Loans which are not verified ,are getting charged 1750 Off more likely



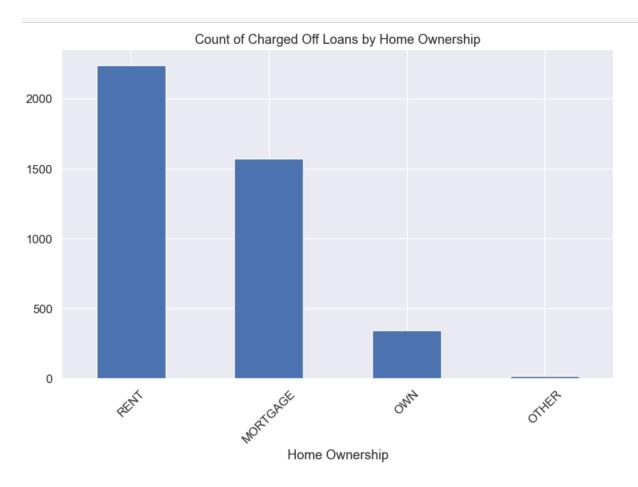
Home membership vs Charge Off Loans

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