Landing Club Case Study

Company Information:

- 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.
- 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.

Risk Definition:

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

Expectation From This Exercise:

Loan Rejected: The company ahs 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. Hence a data set is provided to analytics team to analyze and identify the key area where applicant is not likely to repay the loan.

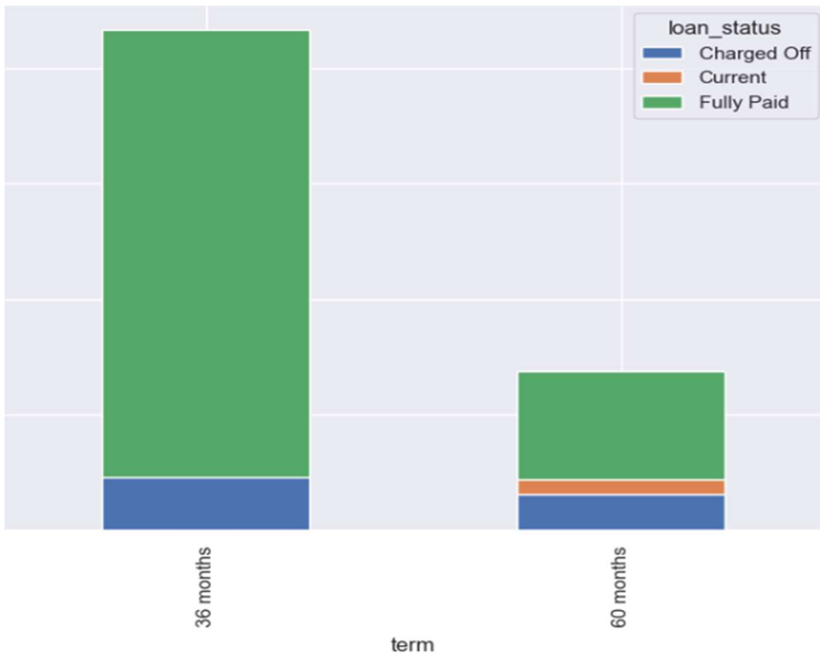
This analysis will help company to reduce financial or business loss. However after analysis detail discussion with business will be required to discuss those key areas where applicant is not likely to repay the loan. This discussion will help to reach end factor of default cases.

Technology:

- I Python - version 3.11
- pandas - version 2.0.3
- numpy - version 1.24.3
- matplotlib - version 3.7.2
- seaborn - version 0.12.2
- warnings

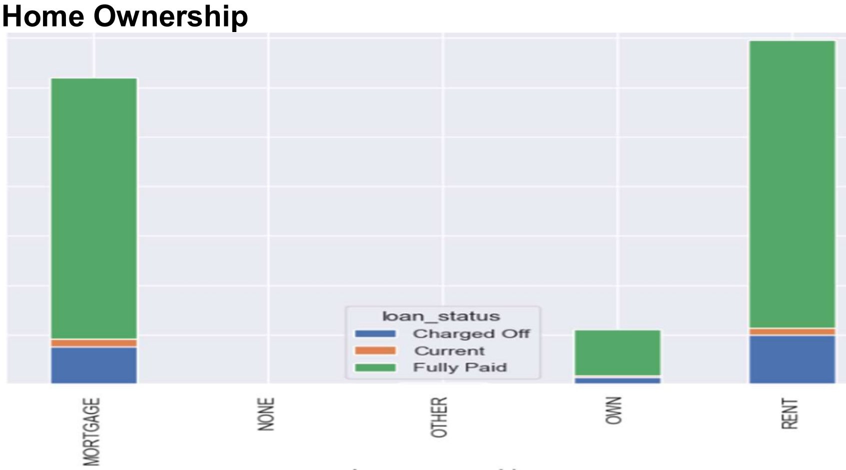
Analysis on Loan Data based on different variables

1. Term



By looking at above chart we can draw the conclusion that percentage of 'Charged Off' cases is much higher when term is 60 Months. Hence avoid long term loan if there is any suspicion.

2. Home Ownership

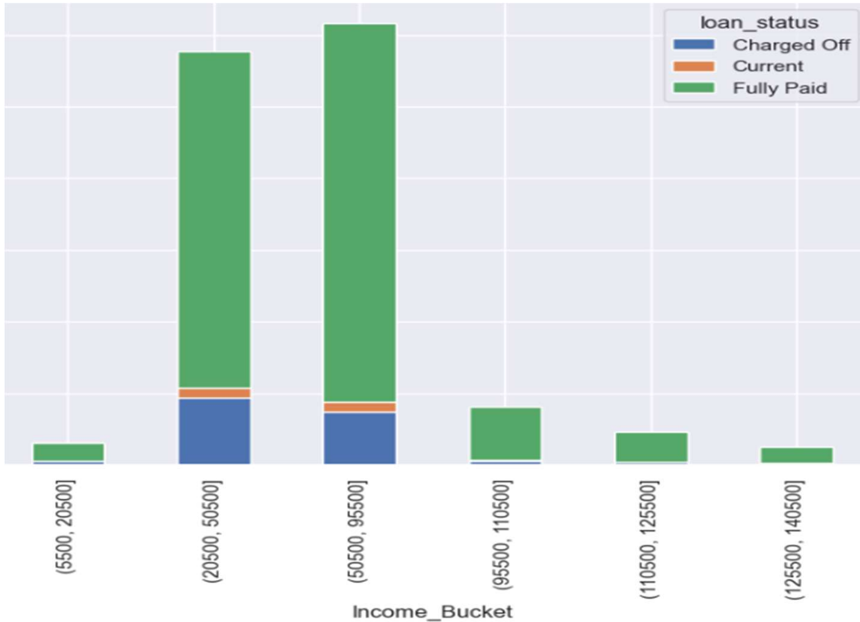


	home ownership
	Charged Off
home_ownership	
MORTGAGE	0.122881
NONE	NaN
OTHER	0.177215
OWN	0.135233
RENT	0.144162

Charged off cases in % -->

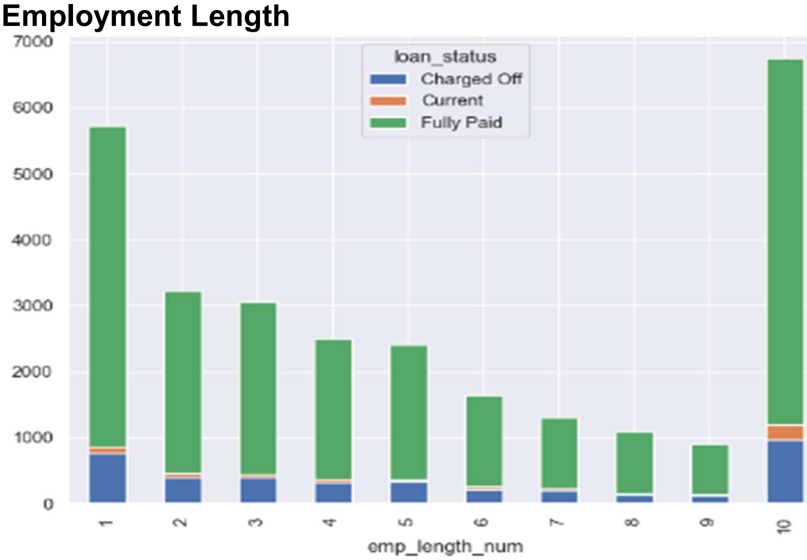
persons who are on Rent, are having high percentage of charged-off case. Hence this category need to be verified more accurately.

3. **Income**



Above graph shows that income bucket of 20500-50500 has more cases of Charged Off cases.

4. **Employment Length**

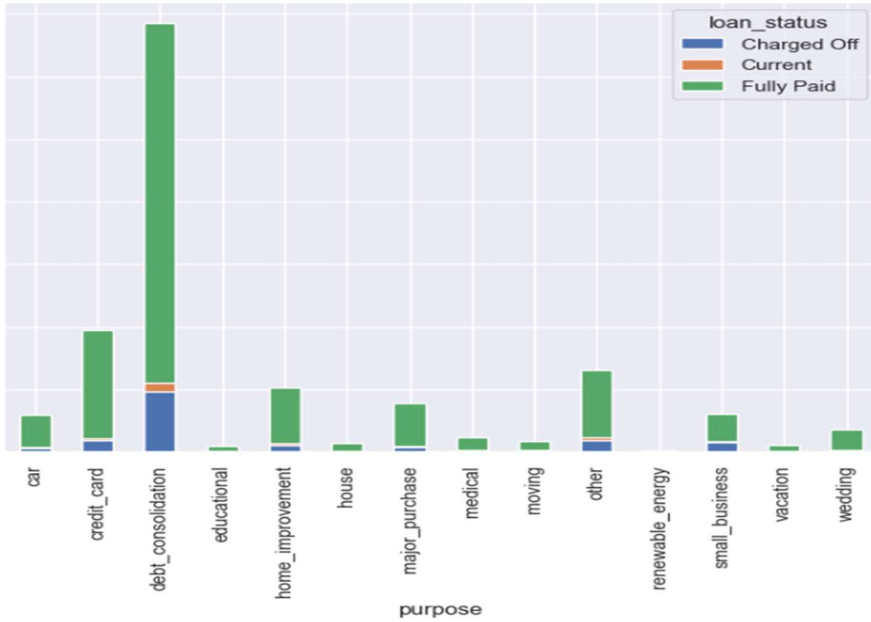


Charged off cases in % -->

emp_length_num	Charged Off
7	0.148773
10	0.143429
5	0.137328
1	0.133893
6	0.132603

above graph and table clearly shows that candidates with emp_length of 7,10+ have higher percentage of chargedoff cases.

5. **Purpose**

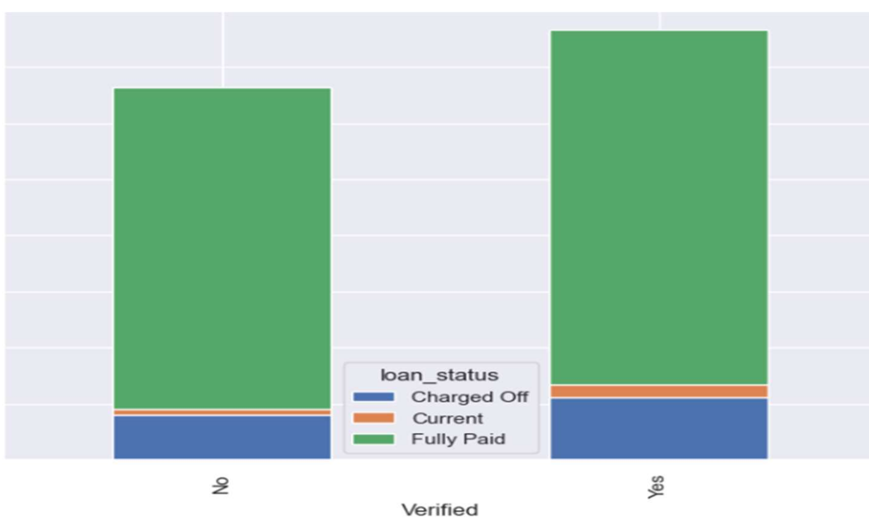


Charged off cases in % -->

purpose	Charged Off
small_business	0.252260
educational	0.179612
renewable_energy	0.158730
moving	0.150000
medical	0.149897

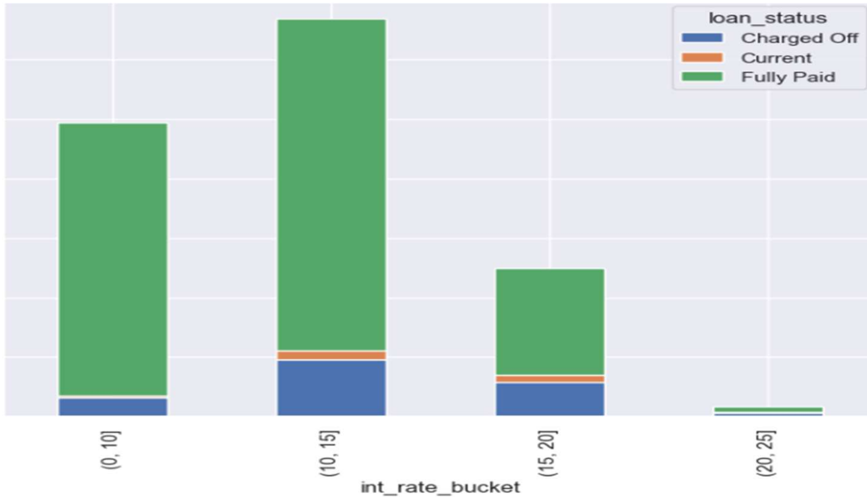
Above table and graph shows that category small_business have highest percentage of chargedoff cases.

6. **Verification Status**



After combining verified values we can see that verified cases have higher percentage of chargedoff cases, hence we can draw the conclusion that verification department need to improve its practices

7. **Interest Rate**

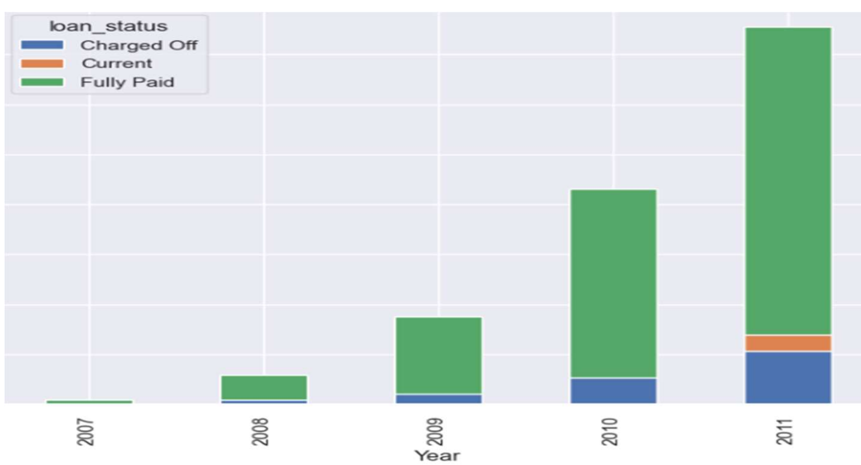


int_rate_bucket	Charged Off
(0, 10]	0.063912
(10, 15]	0.143819
(15, 20]	0.232461
(20, 25]	0.362832

Charged off cases in % -->

By looking into above tables and graph we can conclude that applications where interest rate is 0 to 10 has lowest chargedoff cases.

8. **Year**

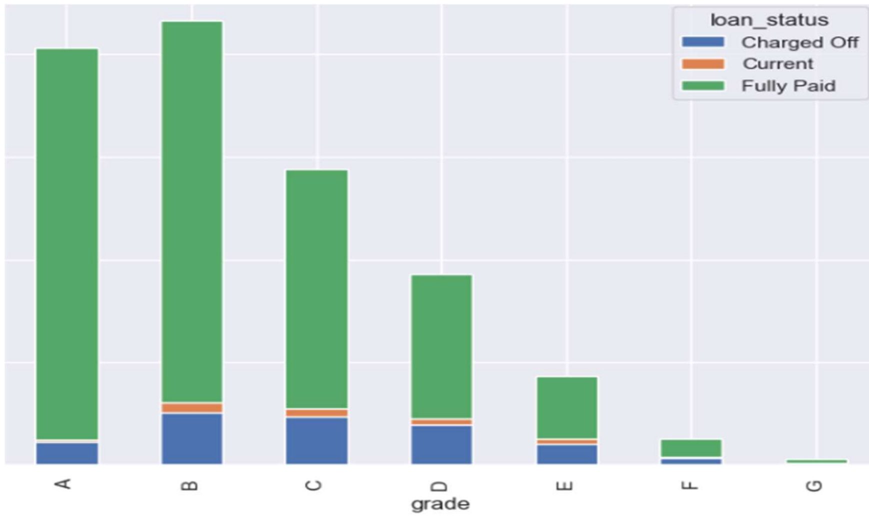


Year	Charged Off
2007	0.133721
2008	0.156277
2009	0.121903
2010	0.124275
2011	0.141268

Charged off cases in % -->

Above analysis shows that from 2010 to 2011, charged off percentage has increased from 12 to 14

9. **Grade**

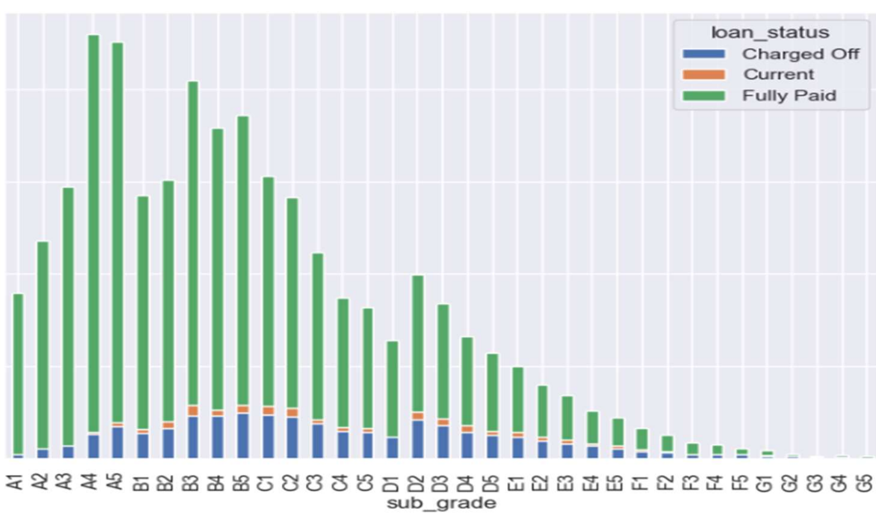


grade	Charged Off
G	0.401709
F	0.284600
E	0.246528
D	0.210753
C	0.166377
B	0.117994
A	0.056692

Charged off cases in % →

Above chart and table shows that grade A has lowest percentage of chargedoff cases. Hence A is safest category of grade.

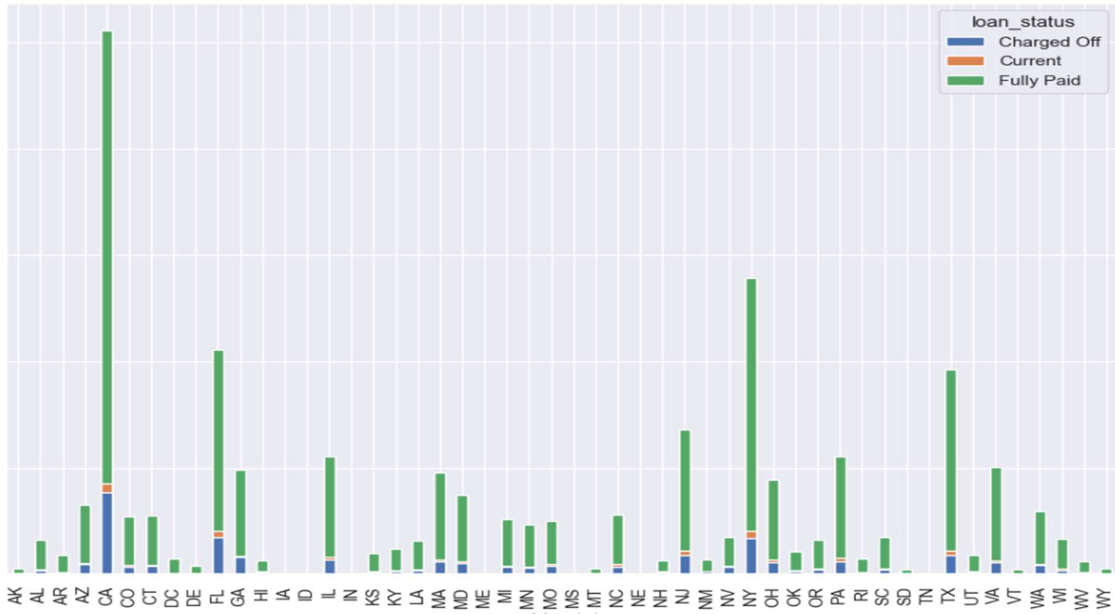
10. **Sub Grade**



By looking at above chart we can conclude that D2 ,D3 have highest percentage of chargedoff cases.

Also B3,B4,B5,C1 and C2 has highest cases of chargedoff cases.

11. Add State



Above chart shows higer number of chargeoff cases in CA,FL,NY and TX

Charged off cases in % →

addr_state	Charged Off	
NE	0.333333	
SD	0.219512	0
NV	0.201149	0
TN	0.181818	
NM	0.176471	0
FL	0.167142	0
HI	0.166667	0
GA	0.162743	0
MO	0.161677	0
AK	0.160714	
NH	0.150794	0
CA	0.149844	0
WA	0.146959	0
VT	0.146341	0
AZ	0.141757	0

However when we look by percentage of chargedoff vs total count, we can conclude that NY,FL,GA,MO,CA,WA,AZ has highest percentage of charegedoff cases.

12. Correlation Matrix

	dti	funded_amnt	open_acc	pub_rec	total_rec_int	loan_status_num	i
dti	1.00	0.08	0.29	-0.00	0.12	0.04	
funded_amnt	0.08	1.00	0.12	-0.02	0.61	0.03	
open_acc	0.29	0.12	1.00	0.01	0.06	-0.01	
pub_rec	-0.00	-0.02	0.01	1.00	0.03	0.06	
total_rec_int	0.12	0.61	0.06	0.03	1.00	-0.03	
loan_status_num	0.04	0.03	-0.01	0.06	-0.03	1.00	
int_rate	0.11	0.18	-0.02	0.12	0.55	0.20	
annual_inc	-0.12	0.12	0.12	-0.01	0.06	-0.05	
revol_bal	0.23	0.19	0.27	-0.05	0.12	0.00	
total_acc	0.22	0.17	0.68	-0.01	0.04	-0.03	
revol_util	0.27	0.07	-0.11	0.06	0.24	0.10	
term_num	0.07	0.23	0.02	0.03	0.54	0.15	
recoveries	0.02	0.07	-0.00	0.00	0.02	0.35	
collection_recovery_fee	0.02	0.04	-0.00	0.00	-0.00	0.19	
last_pymnt_amnt	-0.00	0.32	0.05	-0.02	0.01	-0.24	
out_prncp	0.04	0.12	0.01	-0.01	0.33	-0.05	
delinq_2yrs	-0.03	-0.03	0.02	0.01	0.04	0.02	
emp_length_num	0.04	0.09	0.08	0.07	0.06	0.01	

By looking at above matrix we can conclude below points.

a. recoveries and chargeoff are positively correlated.

b. chargedoff is negatively correlated with 'Last Payment amount'.

c. chargedoff is positively correlated with interest rate.