



Lending Club Case Study

GROUP MEMBERS

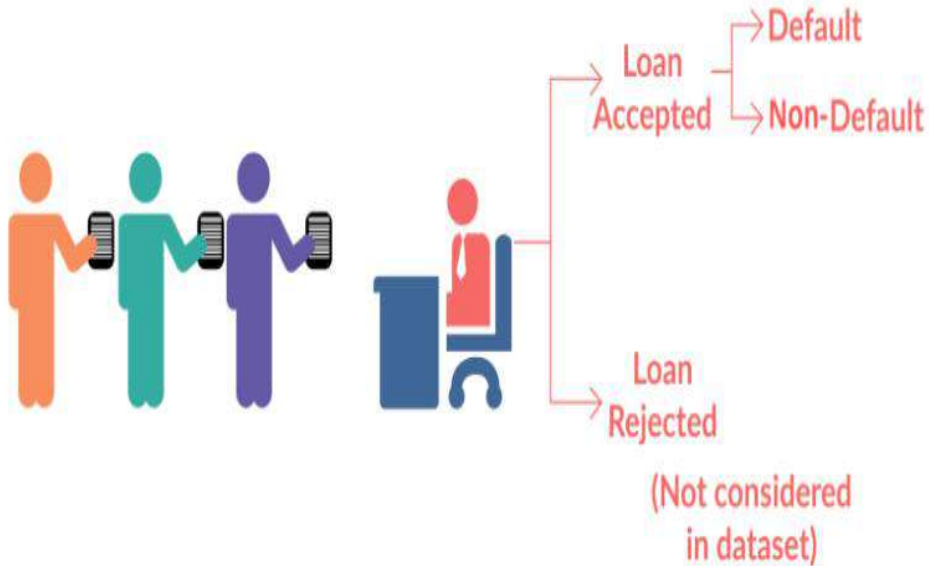
- AAKANKSHA
- GUNEET SINGH GANDHI

PROBLEM STATEMENT

As a data scientist working for Lending Club analyze the dataset containing information about past loan applicants using EDA to understand how *consumer attributes* and *loan attributes* influence the tendency of default

DATA SET UNDERSTANDING

LOAN DATASET



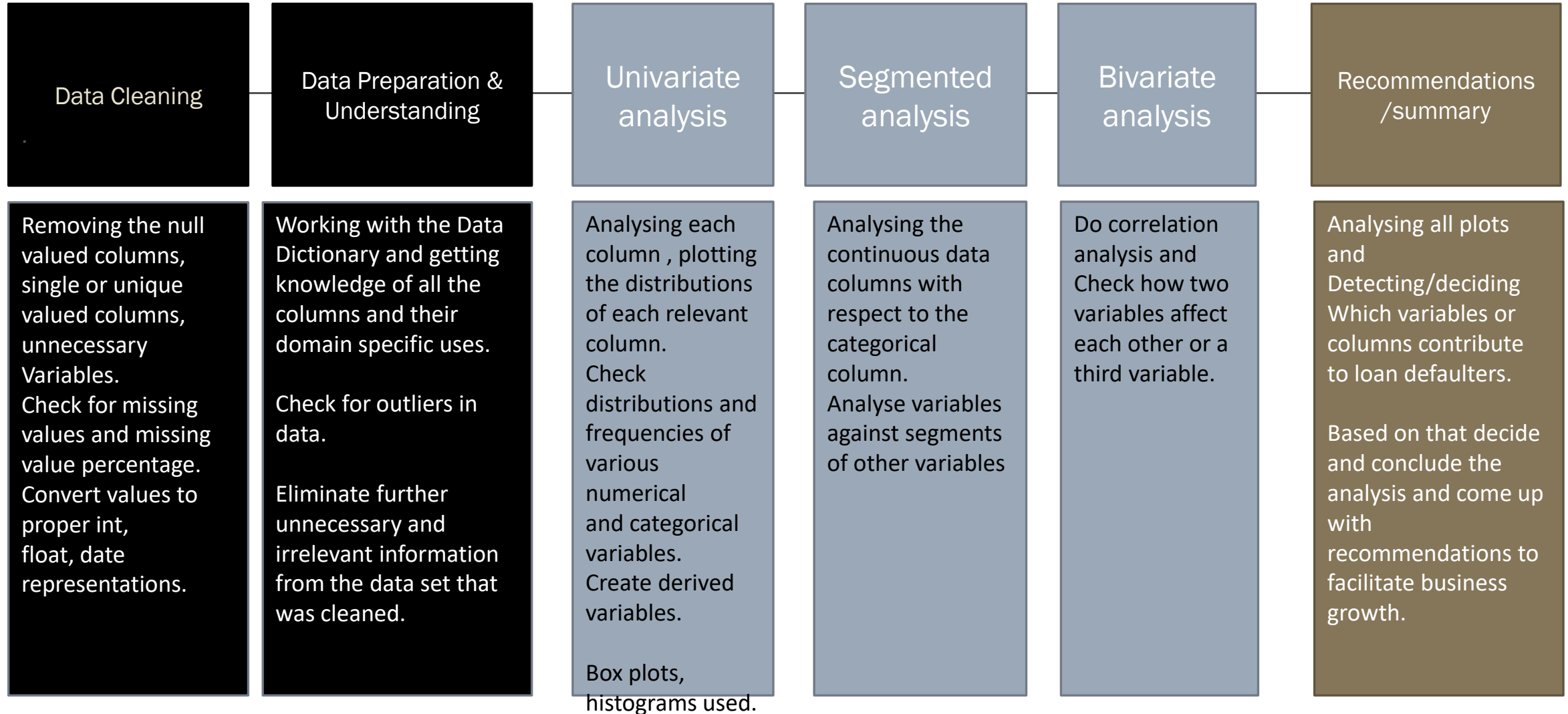
When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

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

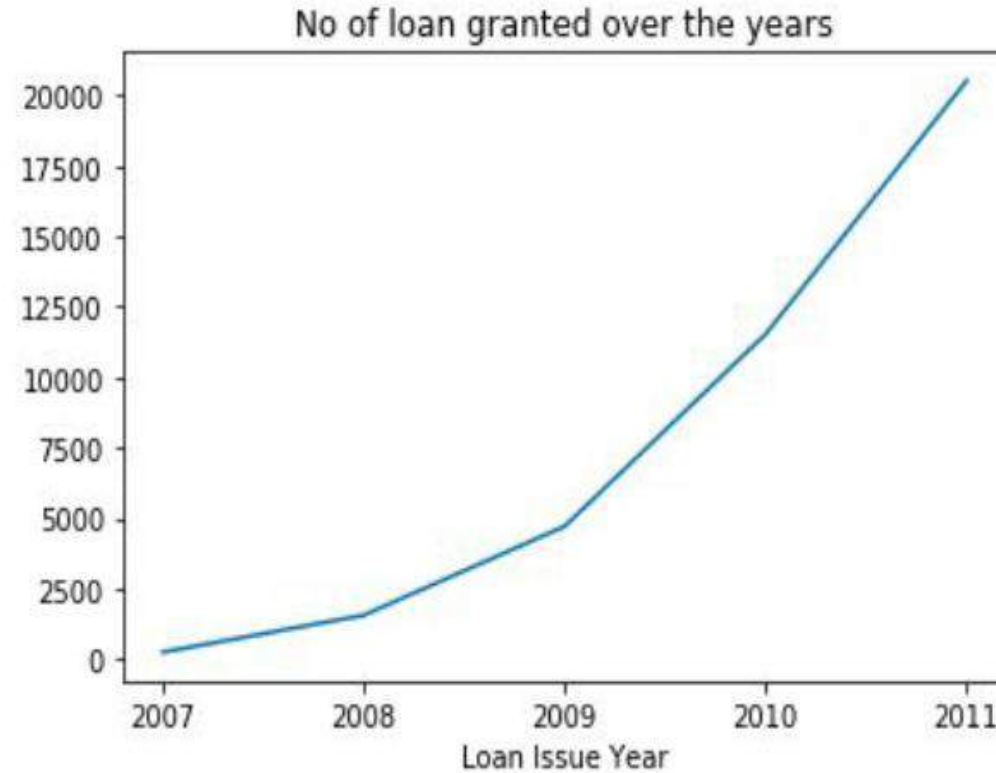
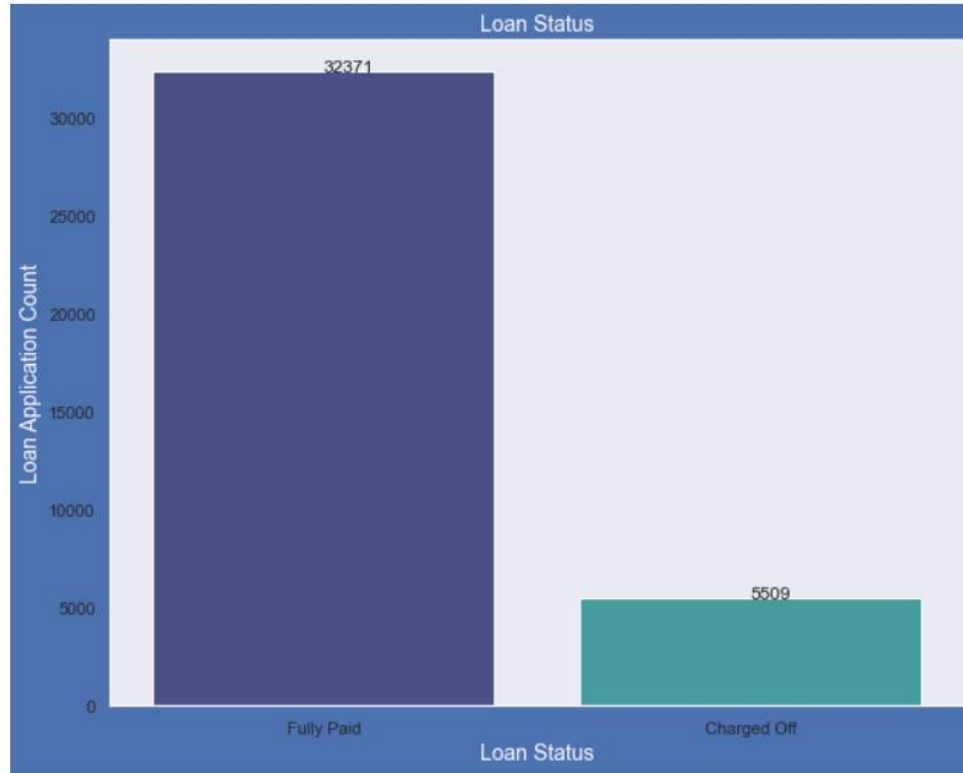
- 1. Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
- 2. Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- 3. Charged-off:** Applicant has not paid 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)

PROBLEM SOLVING APPROACH



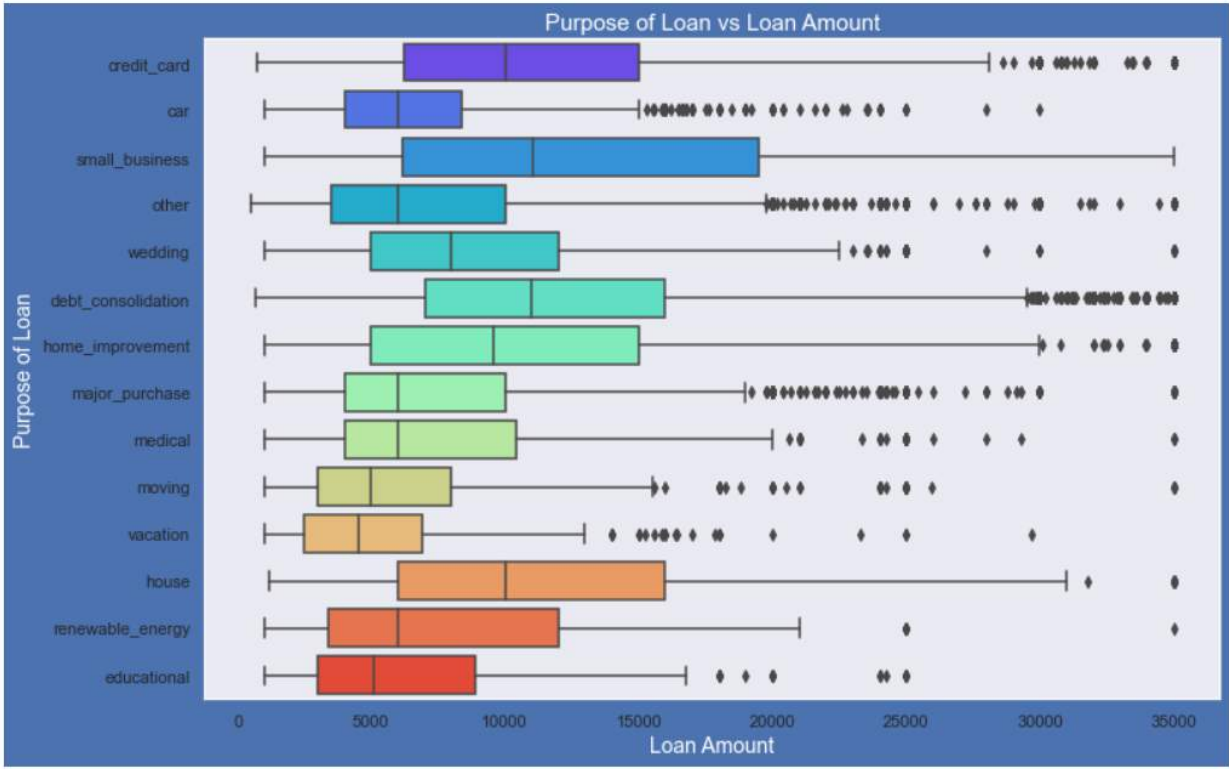
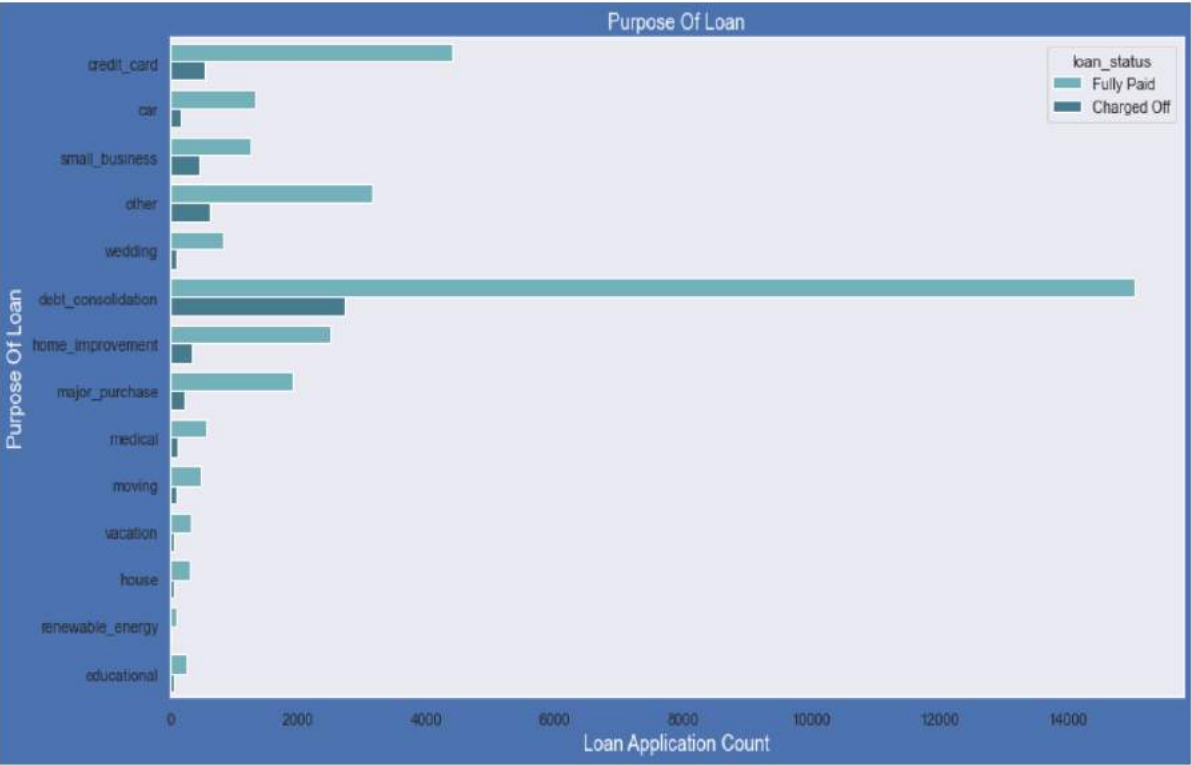
LOAN STATUS ANALYSIS

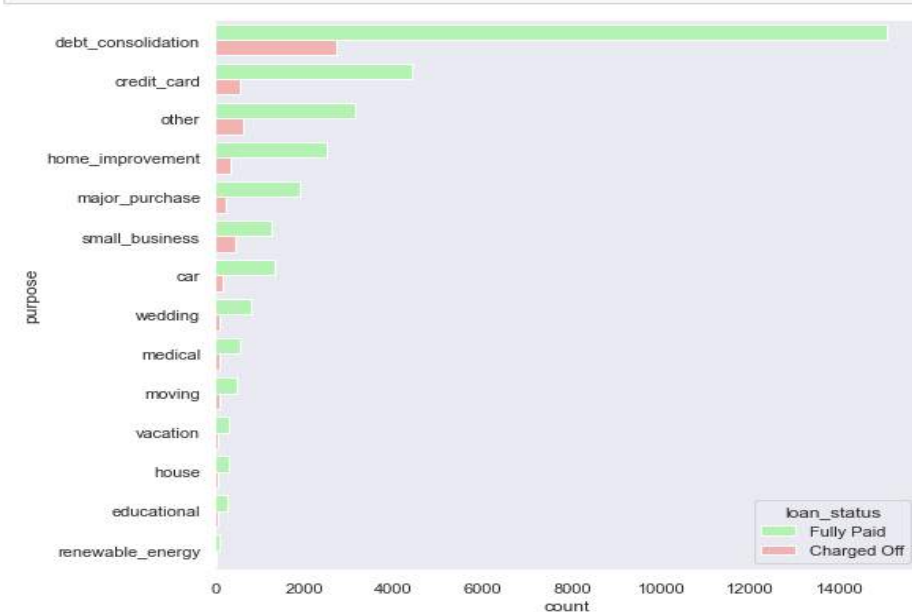
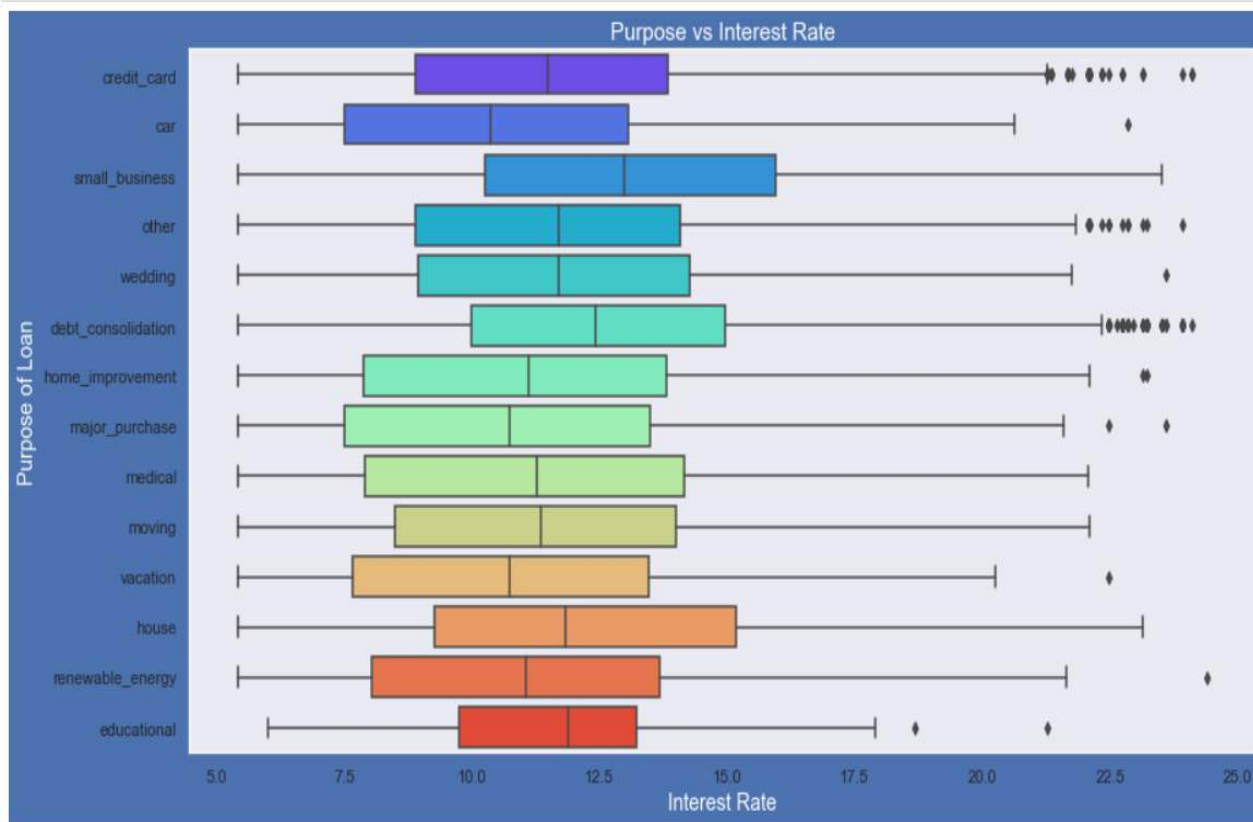


Observations –

1. About 14% of loan are having status as defaulters.
2. The loan has been increasing exponentially over the years.

LOAN PURPOSE



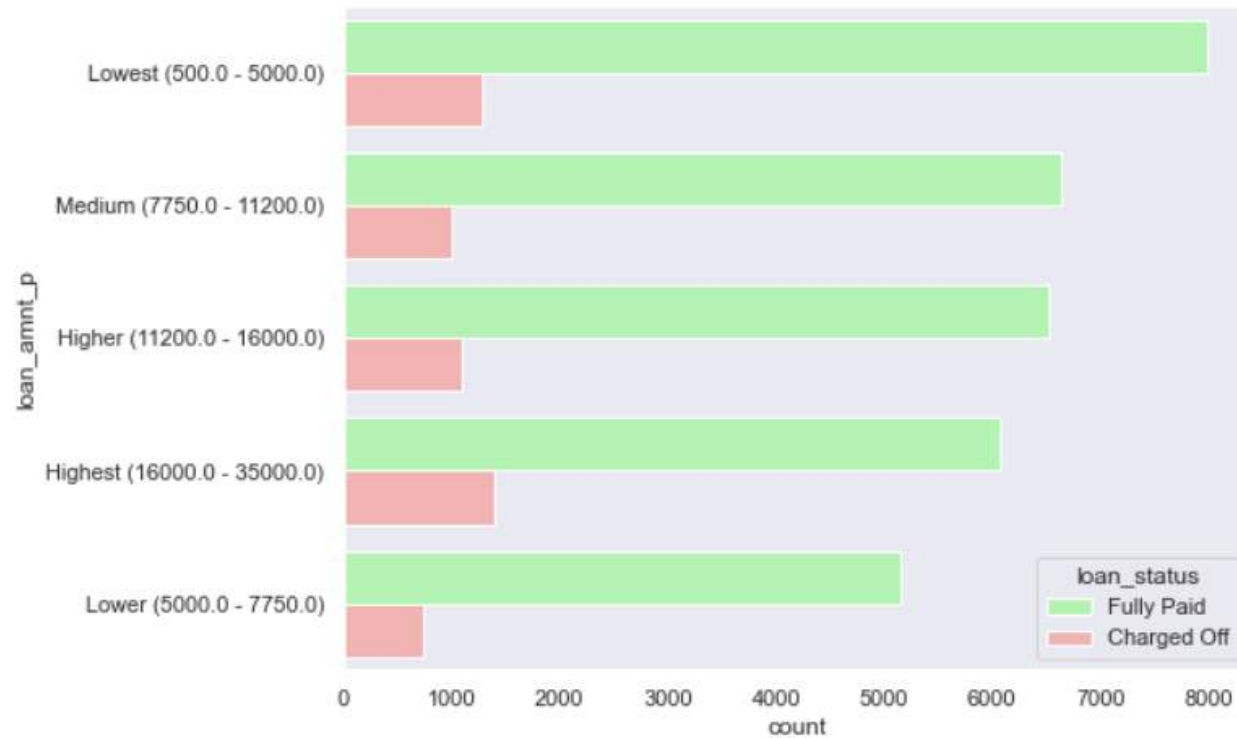


purpose	Charged off %	Record count
small_business	0.268665	1701
renewable_energy	0.186275	102
educational	0.172638	307
other	0.164628	3760
house	0.160221	362
moving	0.156966	567
medical	0.156250	672
debt_consolidation	0.152893	17764
vacation	0.143631	369
home_improvement	0.119561	2827
credit_card	0.108338	4929
car	0.105938	1482
wedding	0.102536	907
major_purchase	0.102299	2131

Observations :

1. Most of the loans were taken for the purpose of debt consolidation & paying credit card bill.
2. Number of charged off count also high too for these loans.
3. Loan amount is highest for loan taken for small business purpose among all purposes. Debt consolidation is second and Credit card comes 3rd.
4. Loans taken for small business are the risky ones followed by renewable energy.

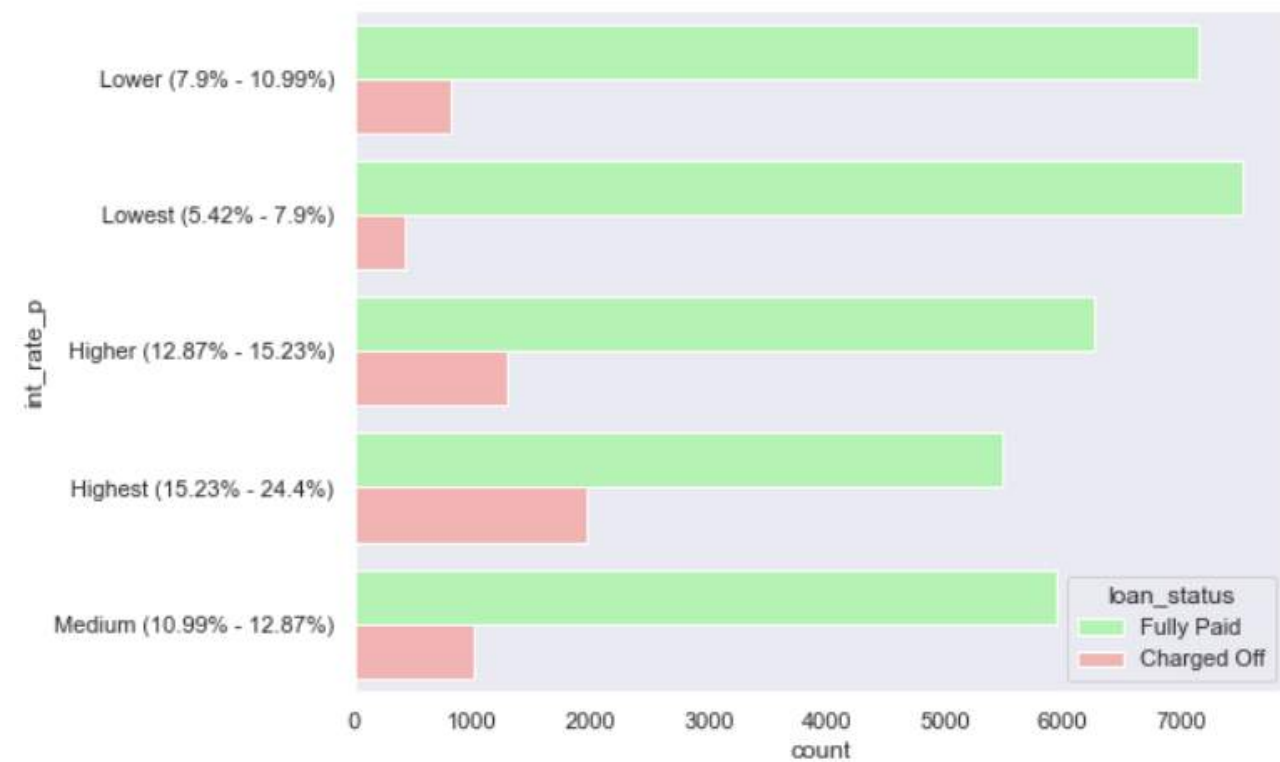
UNDERSTANDING LOAN AMOUNT



loan_amnt_p	Charged off %	Record count
Highest (16000.0 - 35000.0)	0.186009	7462
Higher (11200.0 - 16000.0)	0.144656	7625
Lowest (500.0 - 5000.0)	0.138898	9273
Medium (7750.0 - 11200.0)	0.130589	7627
Lower (5000.0 - 7750.0)	0.124555	5893

Higher the loan amount , greater the chance of loan getting defaulted.

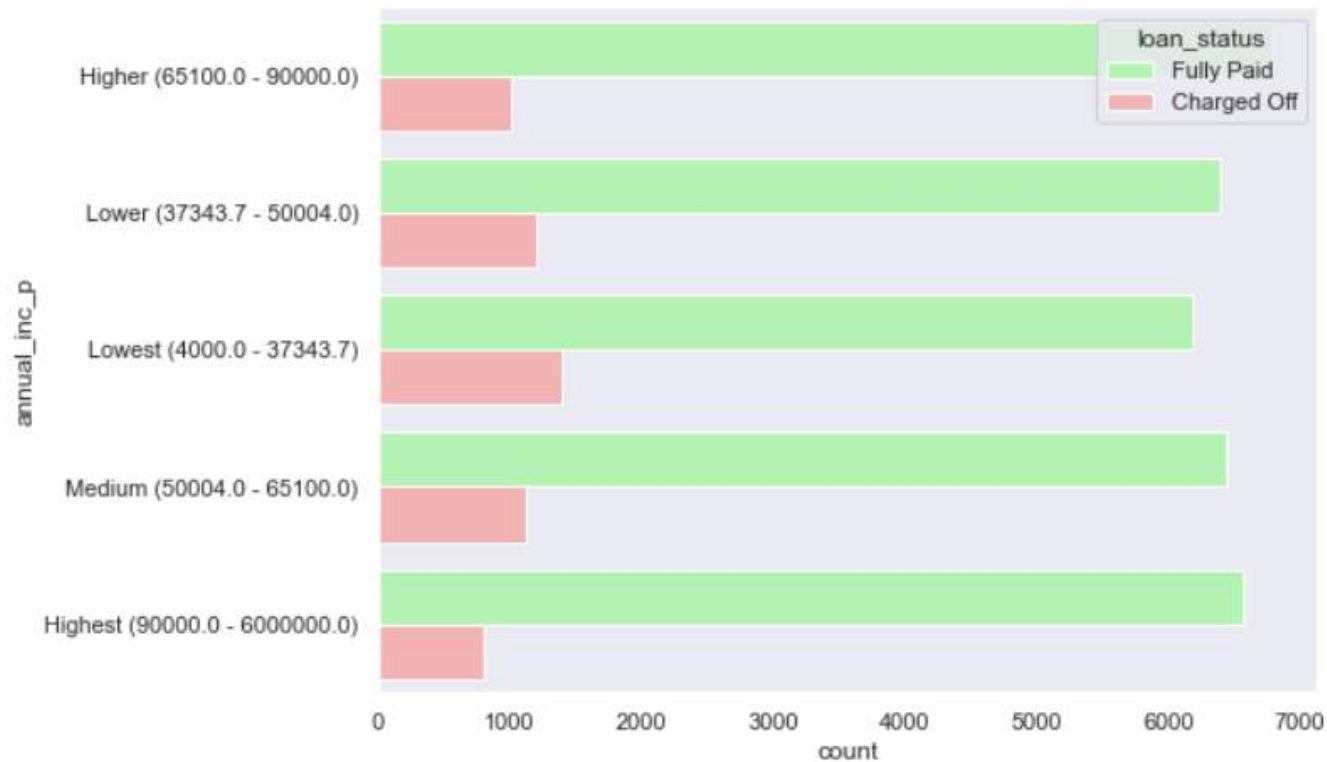
UNDERSTANDING INTEREST RATE



int_rate_p	Charged off %	Record count
Highest (15.23% - 24.4%)	0.263574	7459
Higher (12.87% - 15.23%)	0.172086	7566
Medium (10.99% - 12.87%)	0.144420	6945
Lower (7.9% - 10.99%)	0.102110	7962
Lowest (5.42% - 7.9%)	0.053473	7948

Higher the interest rate leads to higher charged off%

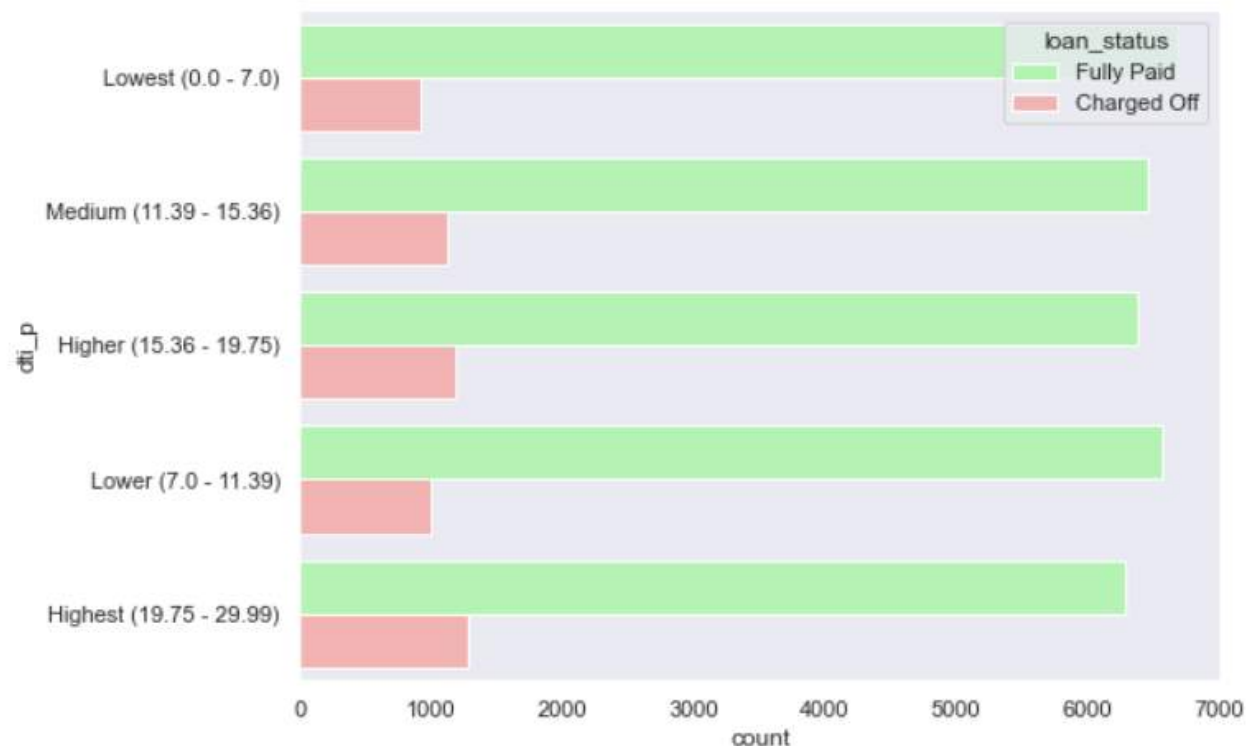
ANNUAL INCOME



annual_inc_p	Charged off %	Record count
Lowest (4000.0 - 37343.7)	0.183474	7576
Lower (37343.7 - 50004.0)	0.157597	7589
Medium (50004.0 - 65100.0)	0.148995	7564
Higher (65100.0 - 90000.0)	0.128767	7797
Highest (90000.0 - 6000000.0)	0.107696	7354

Higher the annual income higher the repayment percentage.

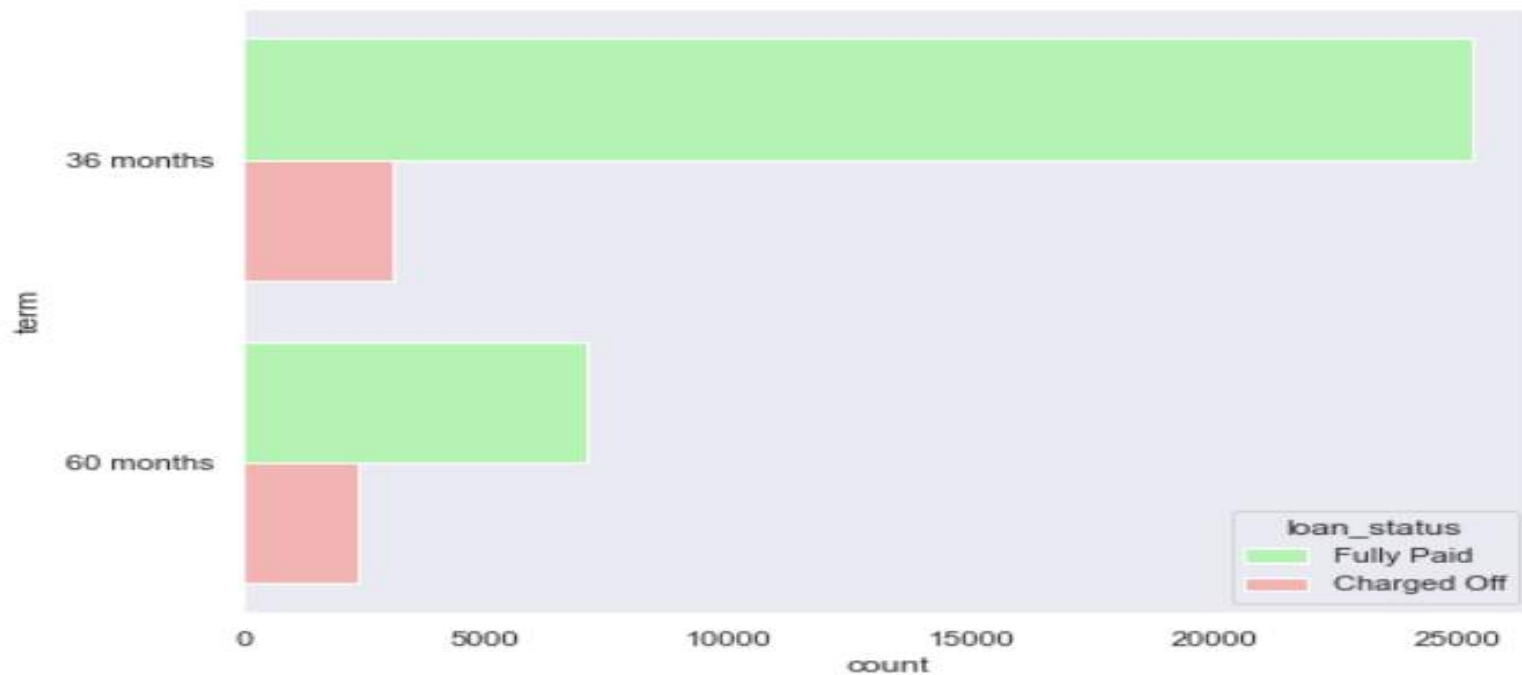
DTI (Debt to Income Ratio)



dti_p	Charged off %	Record count
Highest (19.75 - 29.99)	0.168585	7557
Higher (15.36 - 19.75)	0.156852	7574
Medium (11.39 - 15.36)	0.148524	7588
Lower (7.0 - 11.39)	0.131711	7562
Lowest (0.0 - 7.0)	0.121595	7599

DTI- A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
Higher DTI will lead to higher charged off.

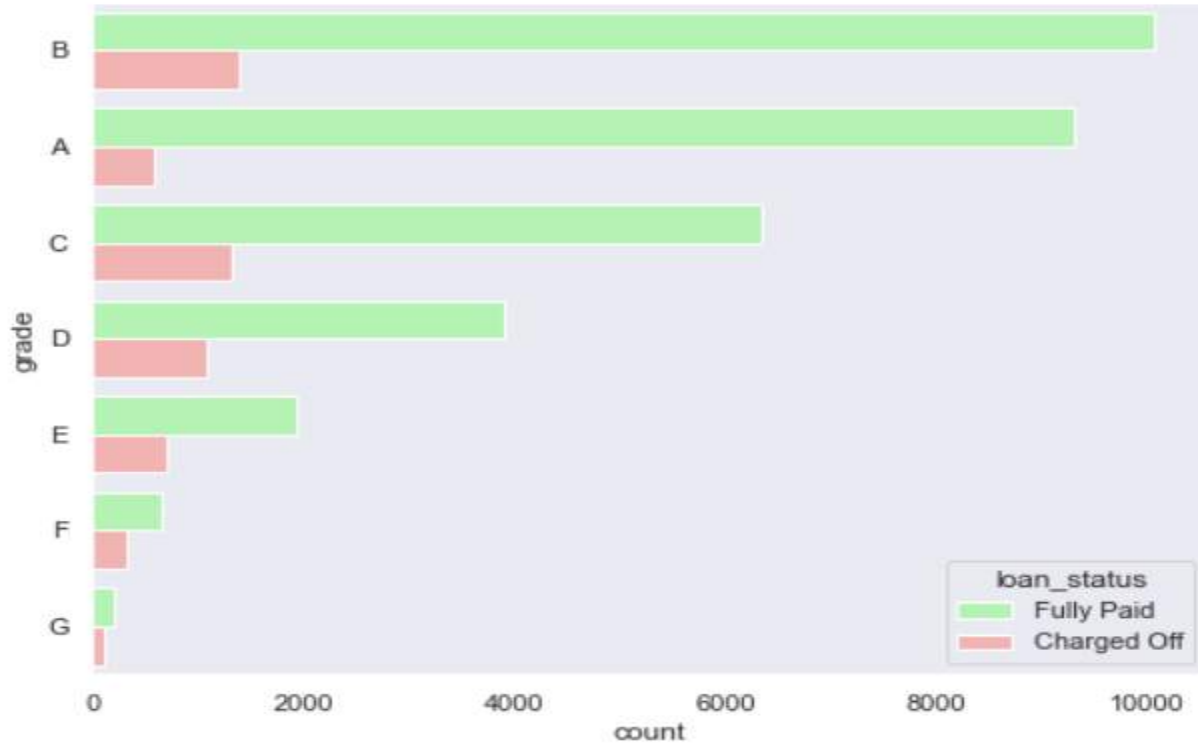
LOAN TERM



term	Charged off %	Record count
60 months	0.253138	9481
36 months	0.109476	28399

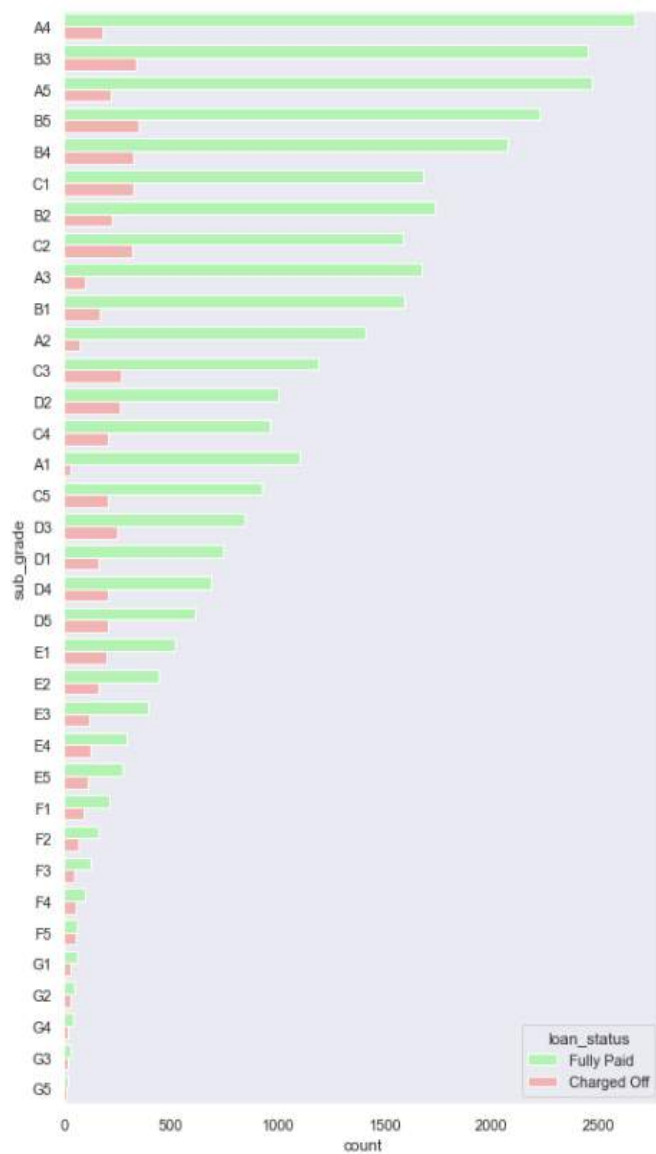
For loans with 60 months repayment term, the default percent is 25%.
For Loans with 36 months loan repayment term, the default is only for 11% of the cases.

LOAN GRADES



grade	Charged off %	Record count
G	0.340067	297
F	0.325702	961
E	0.268265	2628
D	0.218134	4974
C	0.171612	7651
B	0.121751	11466
A	0.060184	9903

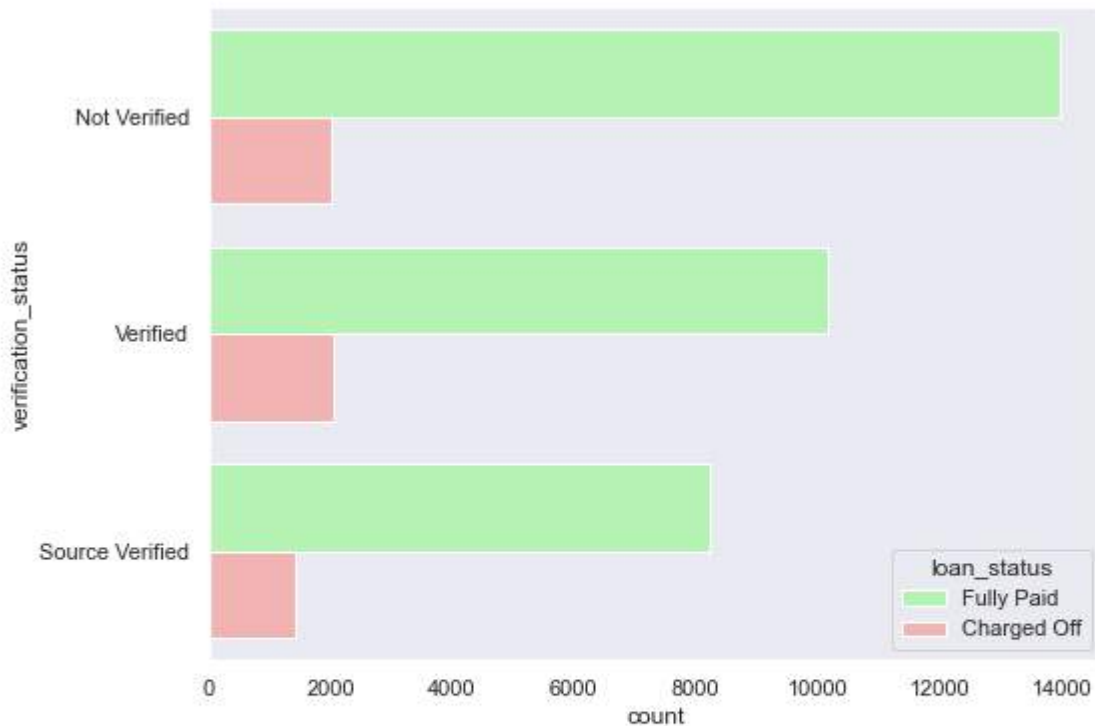
Loan grades having highest default percentages - G, F, E and D form grades where default rate is much higher than others. Lower grades have higher incidence of defaults on loans.
The grading system is working!



sub_grade	Charged off %	Record count
F5	0.473214	112
G3	0.422222	45
G2	0.363636	77
F4	0.358106	146
G5	0.357143	26
G1	0.333333	93
F2	0.302632	228
E4	0.298795	415
F1	0.294702	302
F3	0.286550	171
E5	0.282723	382
E1	0.273492	713
E2	0.264463	605
D5	0.251220	820
G4	0.240741	54
D4	0.231265	895
E3	0.230019	513
D3	0.226311	1095
D2	0.208979	1261
C3	0.182381	1453
C5	0.179556	1125
D1	0.178295	903
C4	0.175918	1171
C2	0.165964	1896
C1	0.162176	2004
B5	0.135639	2573
B4	0.134864	2395
B3	0.121190	2789
B2	0.112532	1955
B1	0.094641	1754
A5	0.080507	2683
A4	0.062256	2843
A3	0.056465	1771
A2	0.049391	1478
A1	0.026596	1128

Loan subgrades G3 and F5 sub categories that have above 40% default rate.

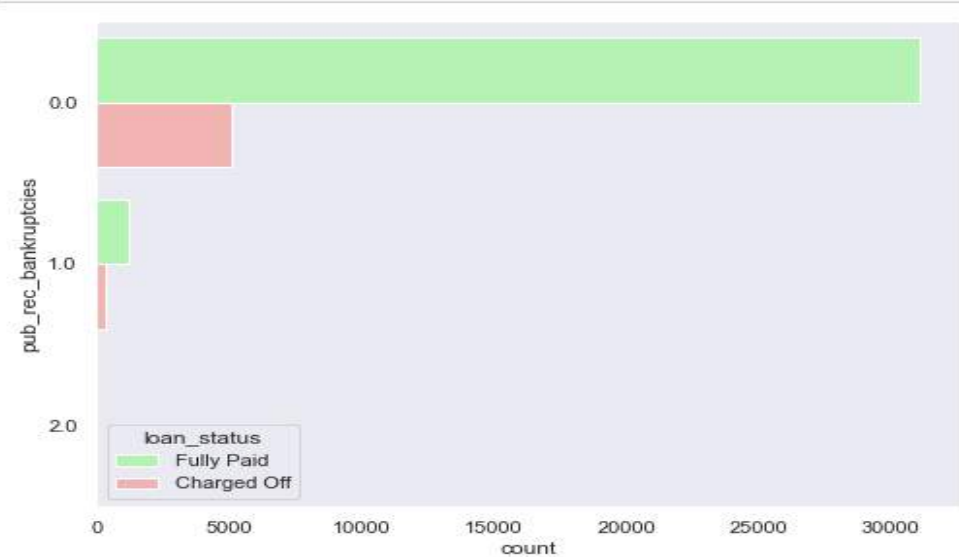
VERIFICATION STATUS



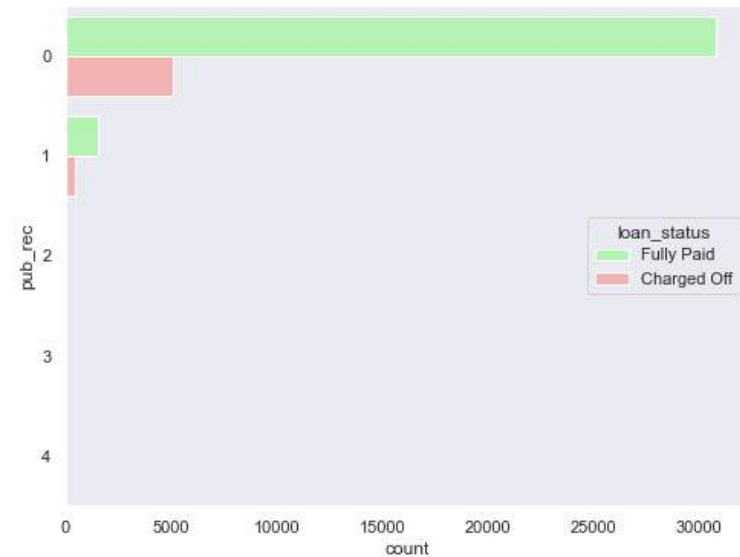
verification_status	Charged off %	Record count
Verified	0.168046	12205
Source Verified	0.148186	9677
Not Verified	0.126516	15998

Verified applicants have a higher charged off percentage .
The verification process needs to be validated/ checked or
changed.

PUBLIC RECORDS



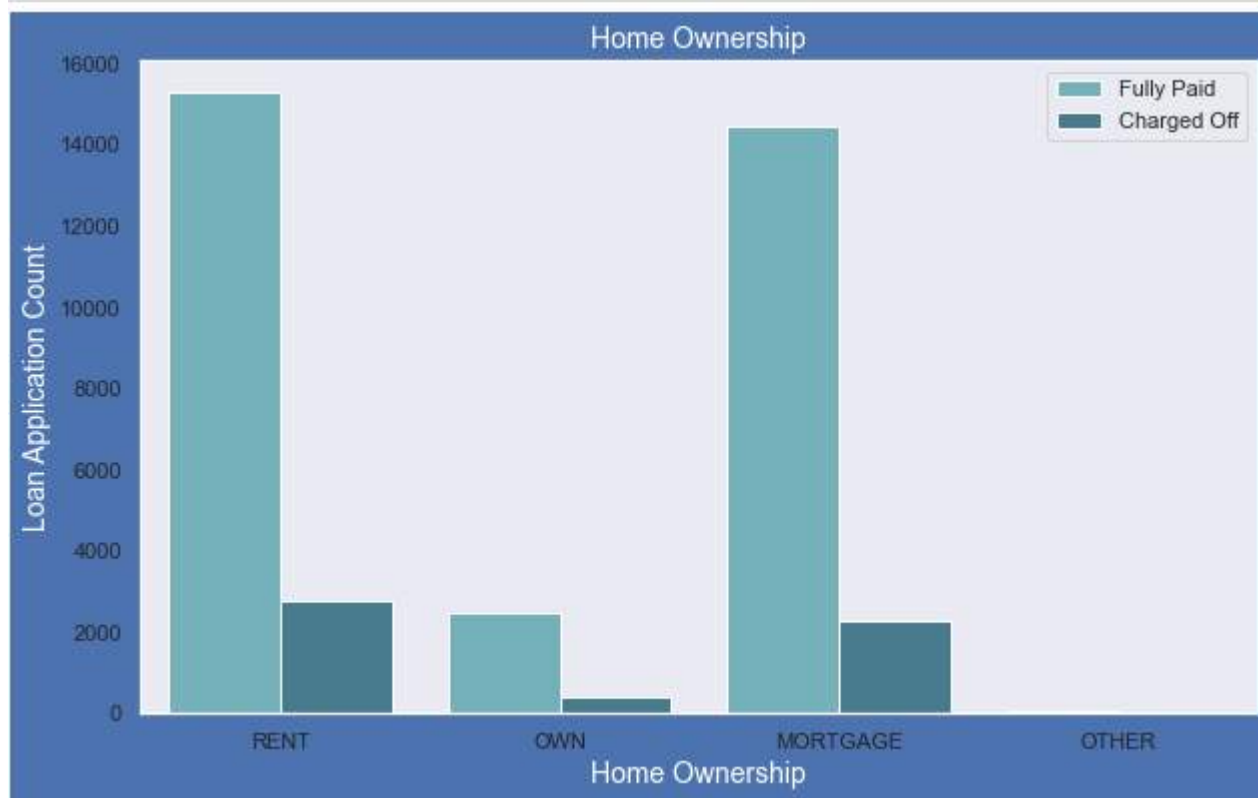
pub_rec_bankruptcies	Charged off %	Record count
2.0	0.400000	5
1.0	0.223580	1637
0.0	0.141868	36238



pub_rec	Charged off %	Record count
1	0.226179	1994
2	0.212766	47
0	0.140888	35830

Higher the number of public bankruptcy records, bigger the chance of defaulting the loan. Having even 1 derogatory record increases the chances of Charge Off significantly. These two factors are correlated and we can use any one to conclude.

HOME OWNERSHIP



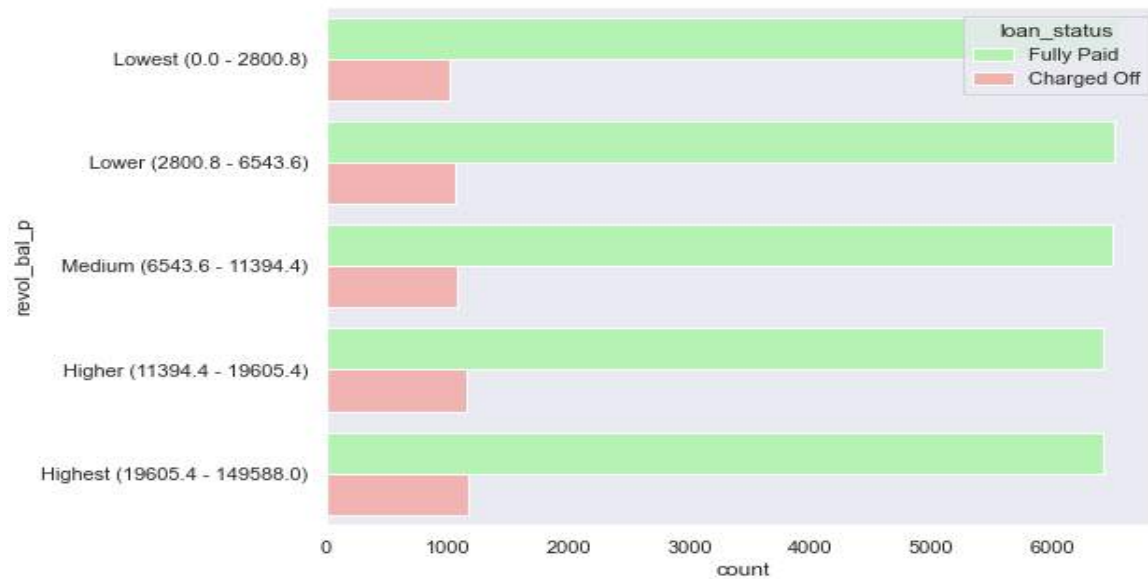
People who live in a rented home or a mortgaged one are most likely to default.

LOAN PAYING TERM



Those who had taken loan to repay in 60 months had more % of number of applicants getting charged off as compared to applicants who had taken loan for 36 months.

REVOLVING BALANCE

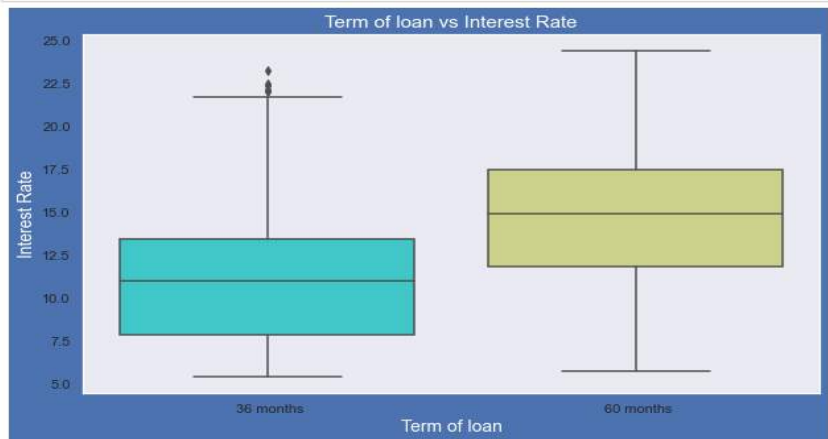
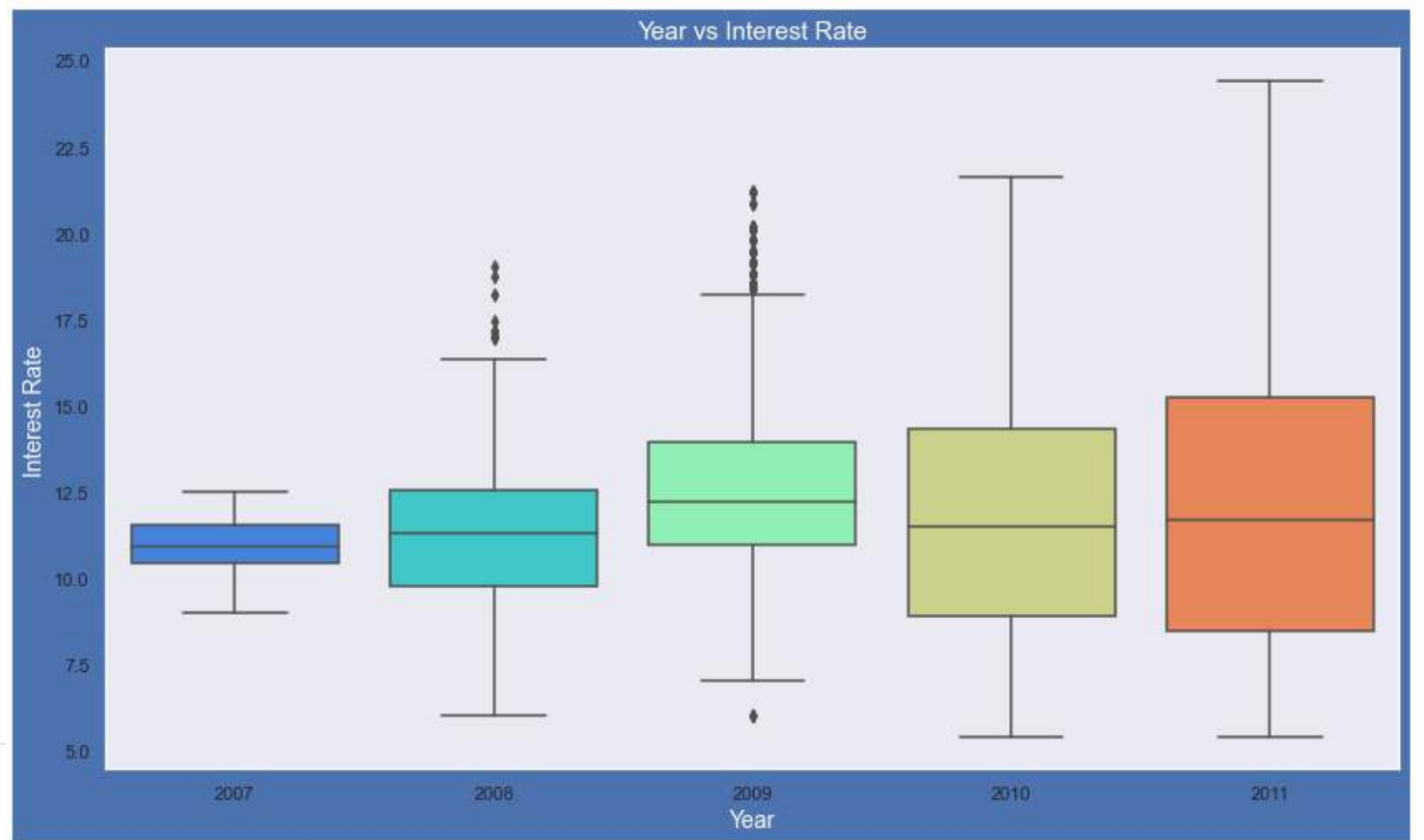


revol_bal_p	Charged off %	Record count
Highest (19605.4 - 149588.0)	0.154303	7576
Higher (11394.4 - 19605.4)	0.153775	7576
Medium (6543.6 - 11394.4)	0.143347	7576
Lower (2800.8 - 6543.6)	0.140576	7576
Lowest (0.0 - 2800.8)	0.135164	7576

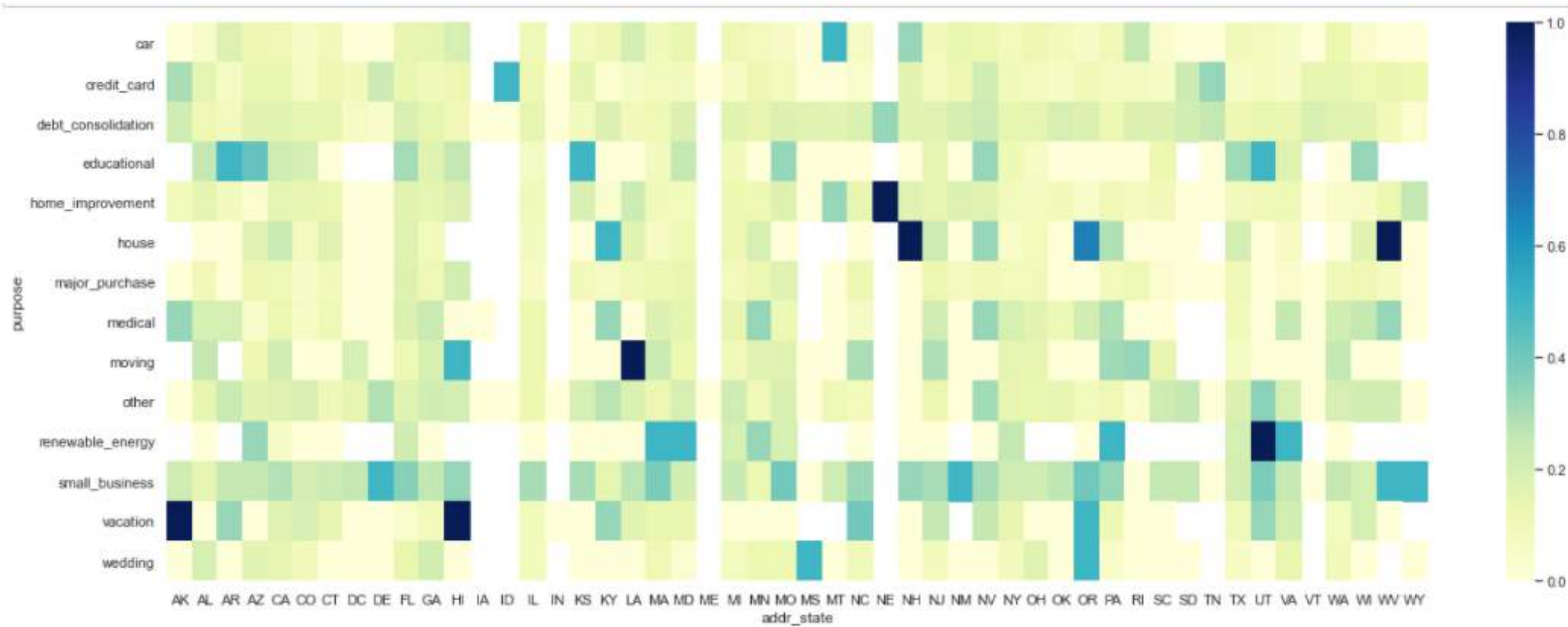
Total credit revolving balance - Higher the revolving balance, bigger the chance of the loan getting defaulted. People with high utilization of Revolving Line of Credit at the time of taking loan default more.

Loans with utilization > 75% are risky.

Plot shows interest rate is increasing slowly with increase in year. Average interest rate is higher for 60 months loan term. Most of the loans issued for longer term had higher interest rates for repayment.



ADDRESS STATE



As per the above plot, the darker the intersection of addr_state has with the purpose of the loan, the riskier the loan application is. Some of the examples are below: •vacation loans in AK, HI, OR •education loans in AR, KS, UT •small business loans in DE, NM, WV, wY •wedding loans in MS, OR

RECOMMENDATIONS

1. Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan default.
2. Grades are good metric for detecting defaulters. Low grade loans have high tendency to default. Grading system is working as expected. Lending club should examine more information from borrowers before issuing loans to Low grade (G to A).
3. Small business loans are defaulted more. Lending club should stop/reduce issuing the loans to them. Reduce the number of approvals where purpose is small business. A deep scrutiny can also be done when the purpose is Debt consolidation.
4. Borrowers with mortgage home ownership are taking higher loans and defaulting the approved loans.
5. People with more number of public derogatory records are having more chance of filing a bankruptcy. Lending club should make sure there are no public derogatory records for borrower. Stop approving loans to people with prior bad record.
6. Loans having higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.
7. When approving loans for a high amount, check the background of the applicant thoroughly as the default rate is directly proportional to amount.
8. Prefer approving loans for people with higher incomes for profit making.
9. Start charging higher interest rates for loans with DTI greater than 20
10. Verified applicants have a higher charged off percentage. The verification process needs to be validated/ checked or changed.

SUMMARY

VARIABLES HAVING A MINOR EFFECT

- Higher loan amount
- Higher installment amount
- Lower annual income
- Higher debt to income ratio
- Applicant's address state (NV, SD, AK, FL, etc.)
- Loan issue month

VARIABLES HAVING A MAJOR EFFECT

- Higher interest rate
- Higher revolving line utilization rate.
- Repayment term (5 years)
- Loan grade & sub-grade (D to G)
- Loan purpose (small business, renewable energy, educational)
- Derogatory public records (1 or 2)
- Public bankruptcy records (1 or 2)

COMBINED VARIABLES

- High loan amount & interest rate for lower income group
- High installment and longer repayment term
- Home ownership (other) and loan purpose (car, moving or small business)
- Income group and loan purpose