Lending club case study

Summary of findings

Rahul Sinha

PURPOSE

- Using EDA to evaluate risk in lending loans to the potential customers.
- Figure out potential customers for profitability.

Business understanding

You work for a **consumer finance company** which specialises in lending various types of loans to urban customers. 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:

- 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.

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Steps taken in analysis

- Data understanding
- Data cleansing
  - ✓NaNs removal
  - ✓Single value columns removal
  - ✓Removal of columns adding no value due to be distinctive nature
  - ✓Final data for analysis
- Outliers treatment
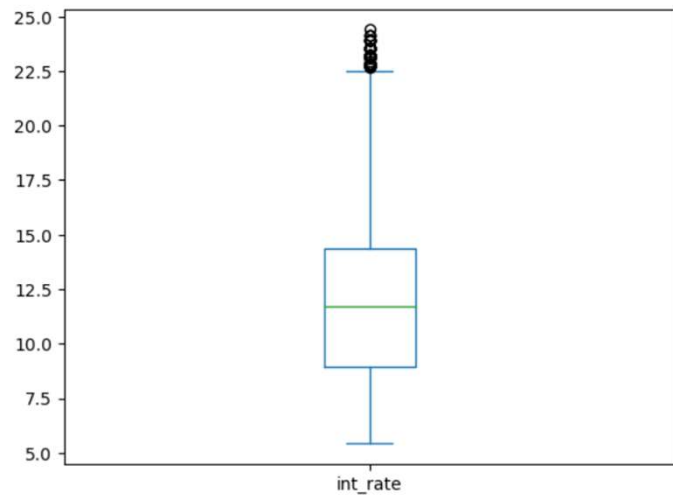- Univariate analysis
- Bivariate analysis
- Recommendations

Data understanding

Available data was checked for nulls, single values, duplicate values and outliers.
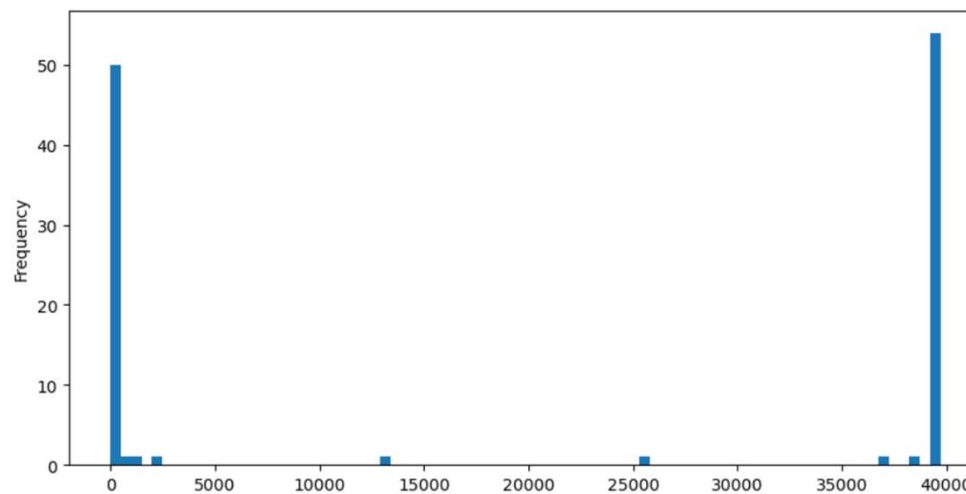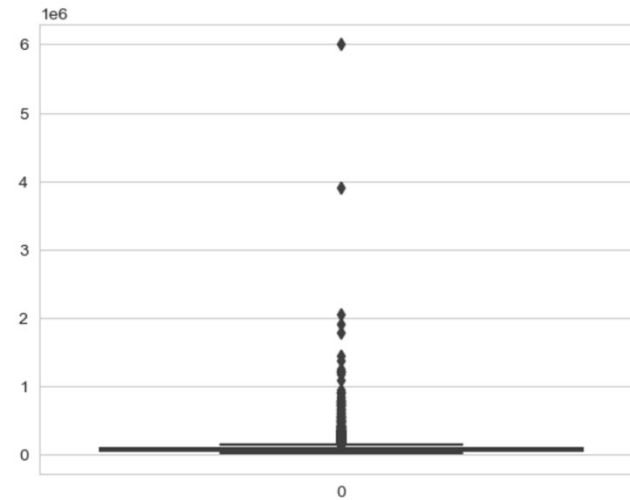

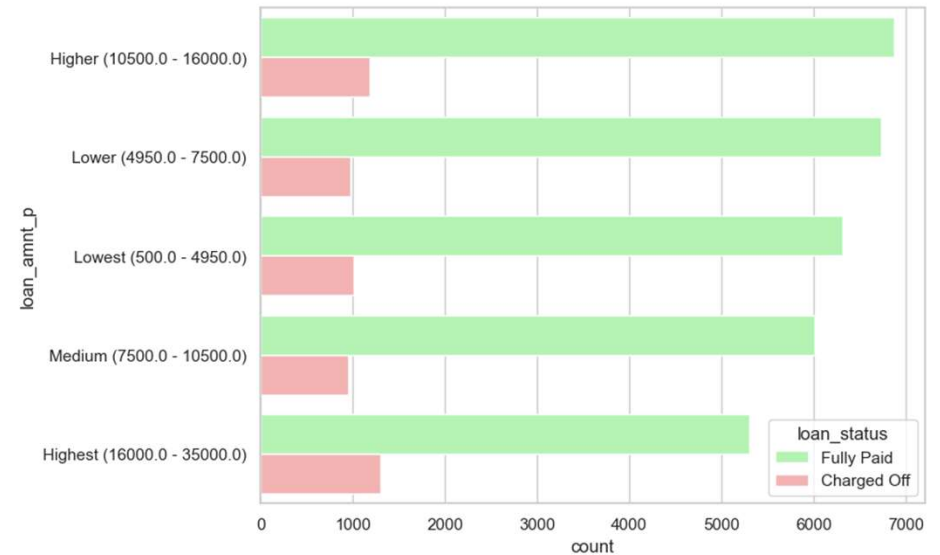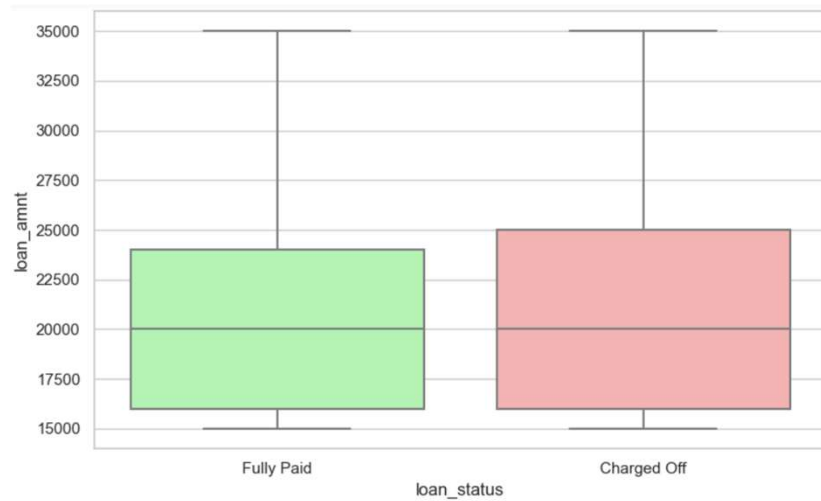Showing null columns in histogram

Outliers

Outliers in interest rate

Outliers in annual income

Loan amount vs loan status

loan_amnt_p	Charged off %	Record count
Highest (16000.0 - 35000.0)	0.196485	6601
Higher (10500.0 - 16000.0)	0.146602	8049
Lowest (500.0 - 4950.0)	0.138298	7332
Medium (7500.0 - 10500.0)	0.137114	6965
Lower (4950.0 - 7500.0)	0.126119	7707

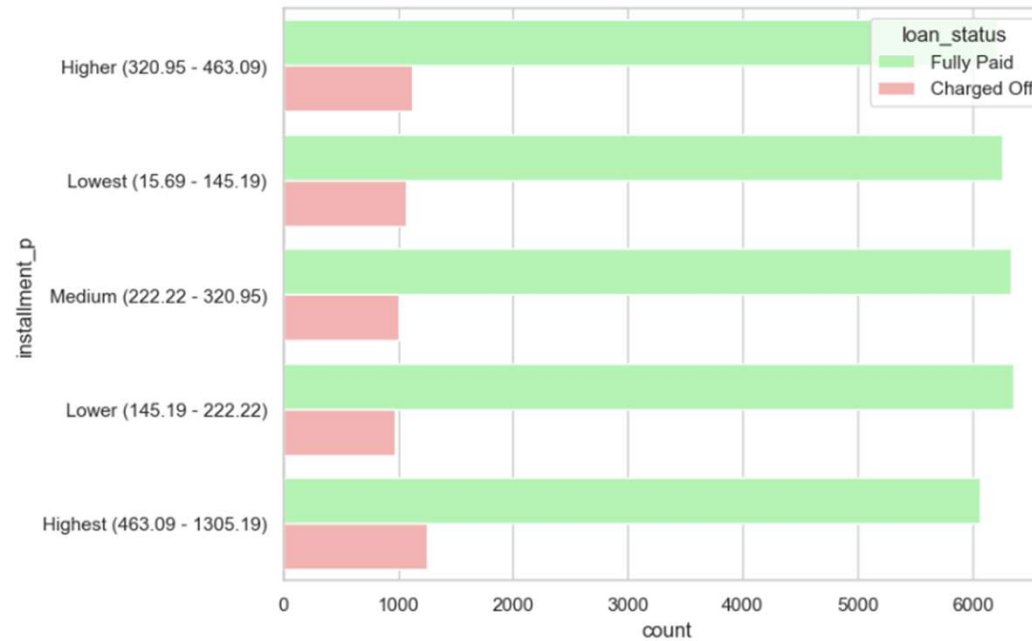Higher the loan amount more likely to be charged off

Interest rates
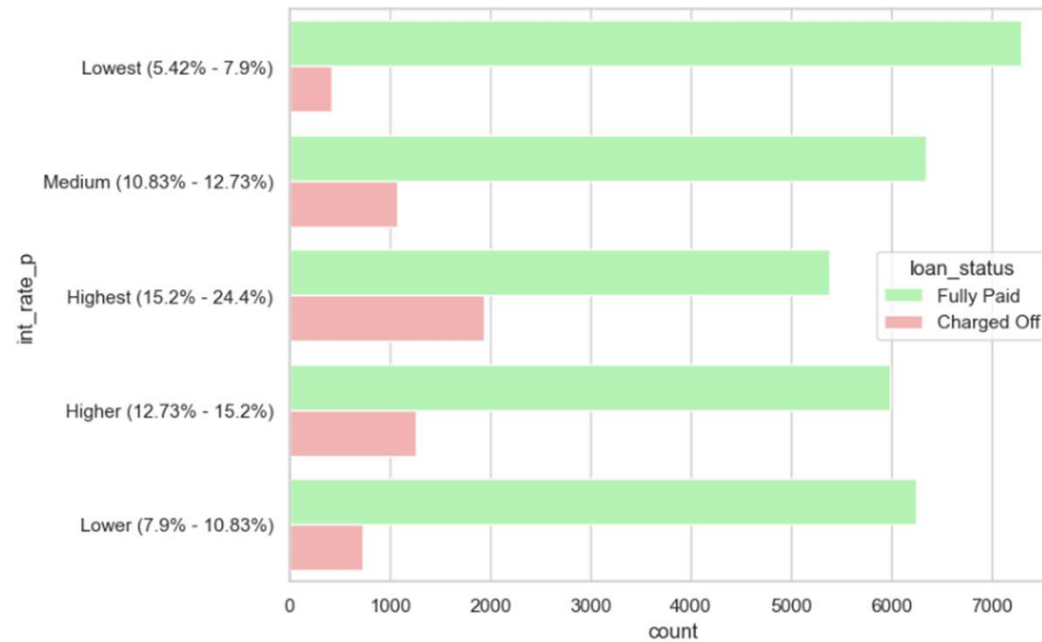
int_rate_p	Charged off %	Record count
Highest (15.2% - 24.4%)	0.265083	7326
Higher (12.73% - 15.2%)	0.173457	7241
Medium (10.83% - 12.73%)	0.144687	7416
Lower (7.9% - 10.83%)	0.103993	6962
Lowest (5.42% - 7.9%)	0.054871	7709

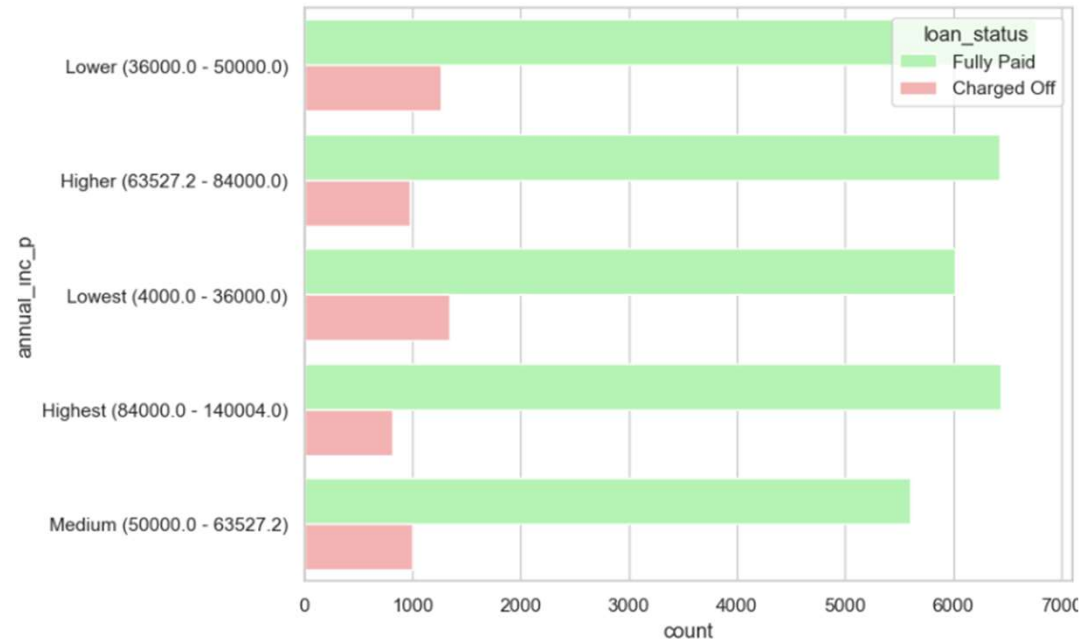Higher interest rates more likely to be charged off

Installment amount

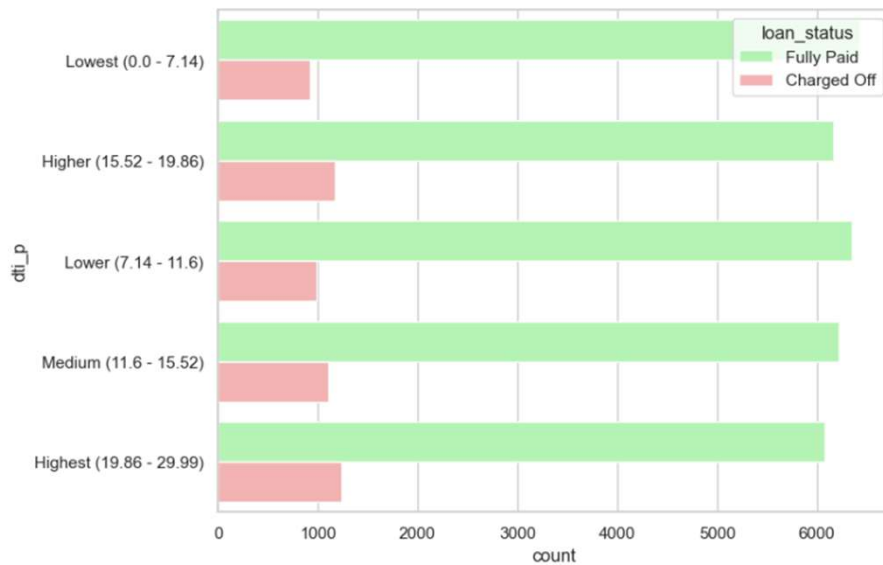| installment_p	Charged off %	Record count
Highest (463.09 - 1305.19)	0.171311	7320
Higher (320.95 - 463.09)	0.152568	7341
Lowest (15.69 - 145.19)	0.146052	7333
Medium (222.22 - 320.95)	0.136543	7331
Lower (145.19 - 222.22)	0.132624	7329

Higher installment amount more likely to be charged off

Annual income

annual_inc_p	Charged off %	Record count
Lowest (4000.0 - 36000.0)	0.182967	7362
Lower (36000.0 - 50000.0)	0.157718	8027
Medium (50000.0 - 63527.2)	0.151749	6603
Higher (63527.2 - 84000.0)	0.132730	7406
Highest (84000.0 - 140004.0)	0.113010	7256

Higher annual income insures higher chances of loan repayment

DTI & Revolving balance

dti_p	Charged off %	Record count
Highest (19.86 - 29.99)	0.168831	7315
Higher (15.52 - 19.86)	0.159602	7337
Medium (11.6 - 15.52)	0.150601	7324
Lower (7.14 - 11.6)	0.134461	7333
Lowest (0.0 - 7.14)	0.125664	7345
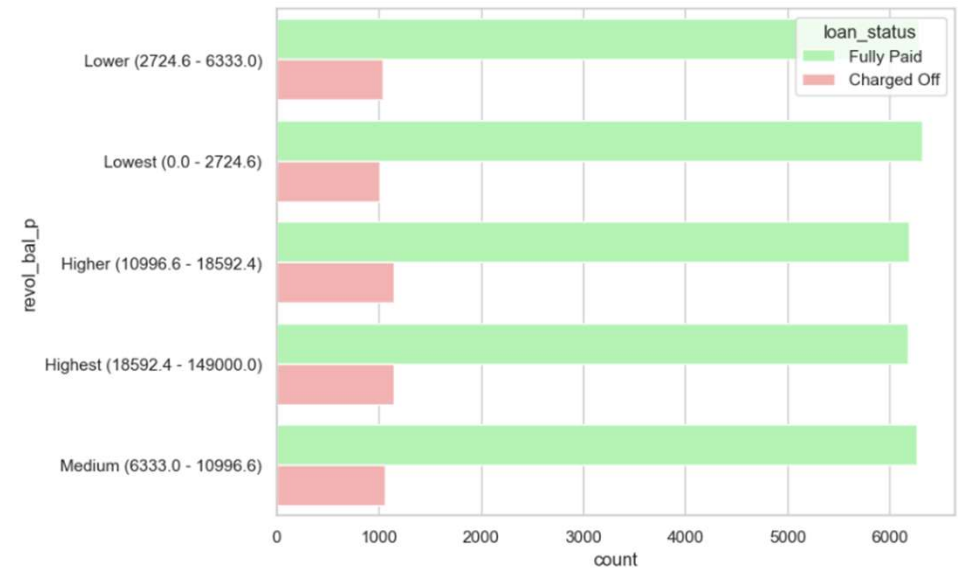
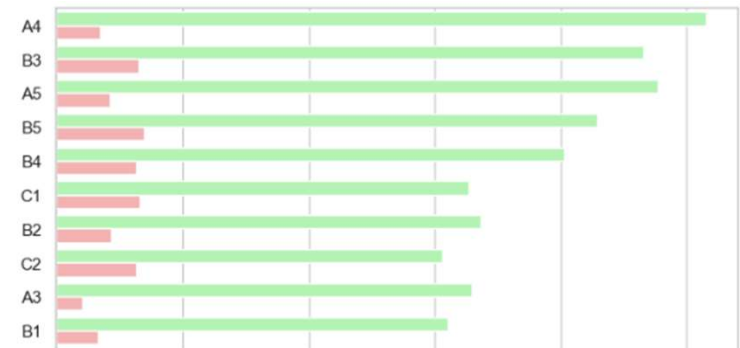| revol_bal_p	Charged off %	Record count
Highest (18592.4 - 149000.0)	0.156732	7331
Higher (10996.6 - 18592.4)	0.156322	7331
Medium (6333.0 - 10996.6)	0.145313	7329
Lower (2724.6 - 6333.0)	0.142390	7332
Lowest (0.0 - 2724.6)	0.138317	7331

Higher DTI & revolving balance higher chances of charged-off
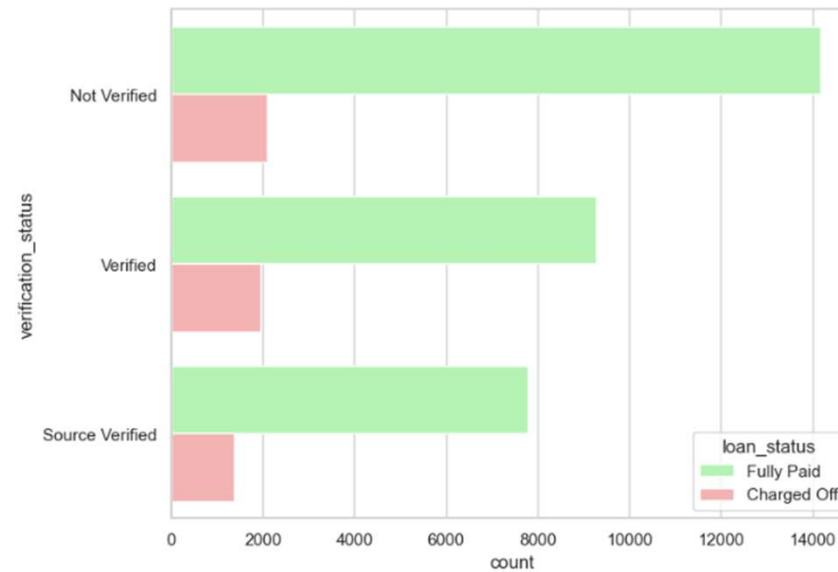
Grade and Sub-grade
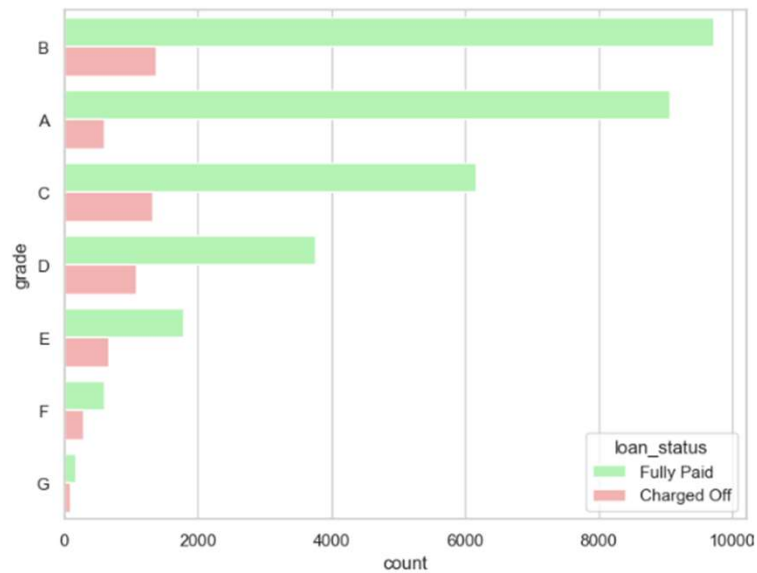
grade	Charged off %	Record count
G	0.363985	261
F	0.326185	886
E	0.269608	2448
D	0.223371	4835
C	0.177115	7481
B	0.123997	11097
A	0.061476	9646

D, E, F and G have very high default rate

The G3 and F5 sub categories have above 40% default rate. This field is a clear indicator of the default percent.
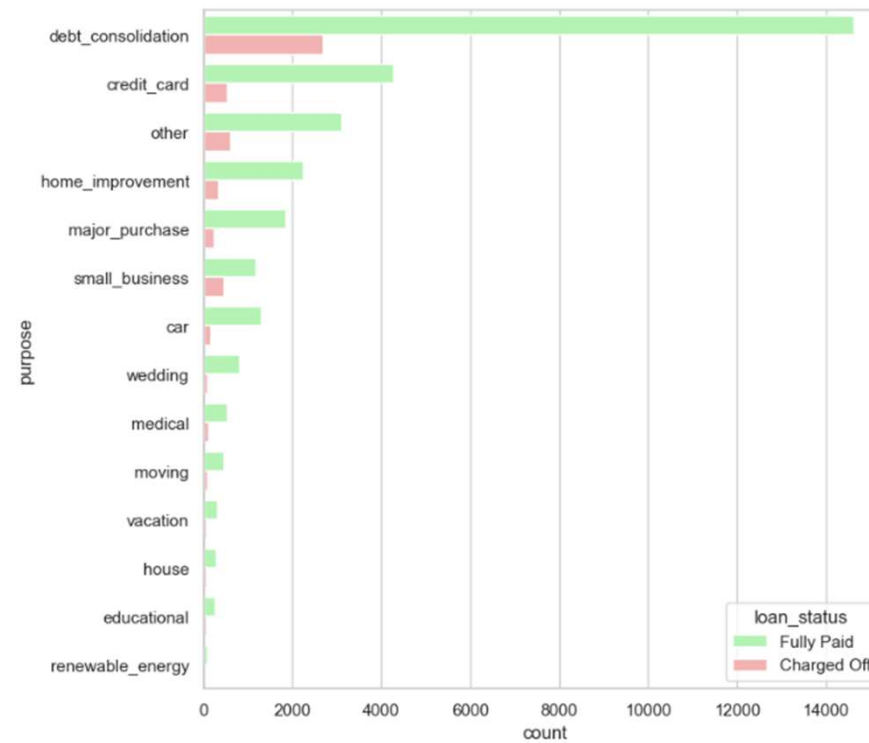
VERIFICATION

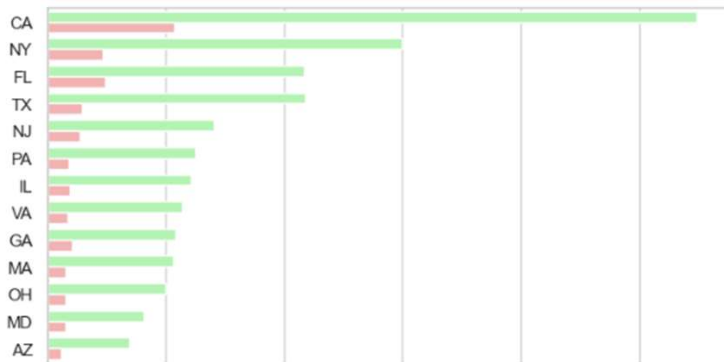| verification_status	Charged off %	Record count
Verified	0.173723	11219
Source Verified	0.149973	9155
Not Verified	0.128747	16280

verified customers showing more defaults so
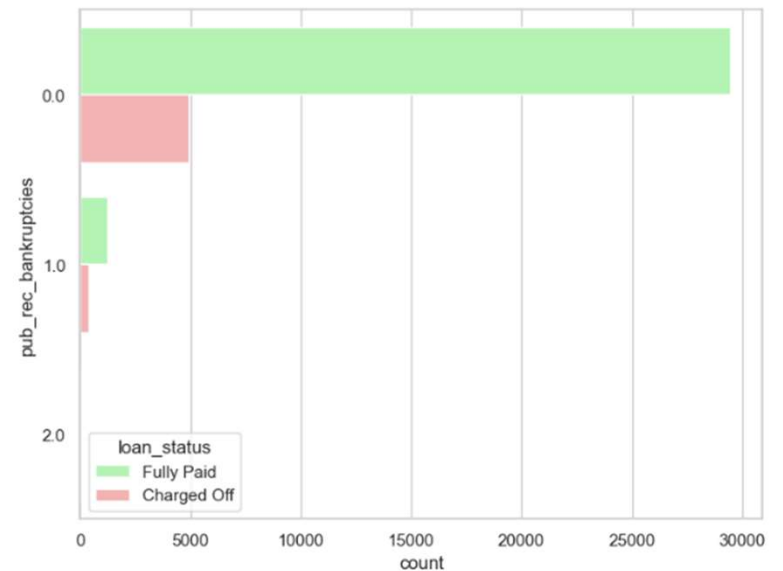verification has a flaw

Purpose

small business purpose, Renewable and Education loan has maximum defaults
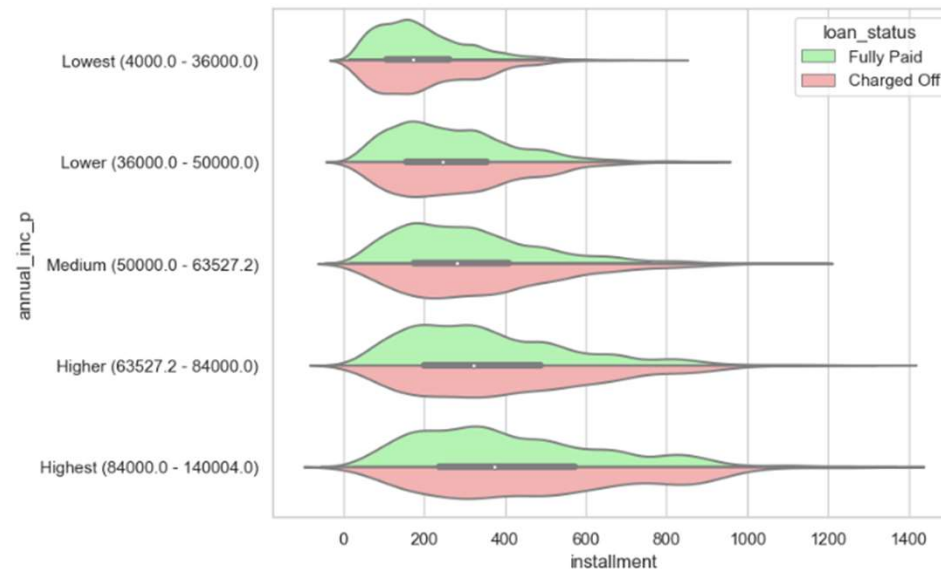
Address state & Bankruptcy

From the above table we can see that the some of the risky states are NE, NV, SD, AK, FL, MO

Higher the number of public bankruptcy records, bigger the chance of defaulting the loan

Bivariate analysis

Above figure shows that for higher installments for any income group have more number of defaults

Recommendations

- Applicant with loan amount more than 10k are more likely to be a risk.
- Loans with Interest rates more than 12% are under risk.
- Installment amount more than 320/- is likely to be a risk.
- Annual income more than 36k mostly ensures good repayment chances.
- DTI more than 15 and revolving balance more than 13k is under risk.
- Grades D, E, F, G and G3, F5 have default percent more than 40. It is a clear risk.
- Verified customer showing higher defaulters which is a sign of ill verification process.
- Loans for education, small business and renewable is under risk.
- Risky states are NE, NV, SD, AK, FL, MO from default point of view.