Lending Club Case Study

- By
Sai Josree
Nilesh

Business Objective

Lending club is a finance company which specializes in providing analysis that helps lenders, to decide on processing the loan applications.

Objective is to understand the driving factors or variables behind loan default i.e., the variables which are strong indicators of default and utilize this knowledge for portfolio and risk assessment.

Two types of risks are associated with the lender's decision:

- *If the applicant is likely to repay the loan, then not approving the loan results in 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*

Data Cleaning and Correction

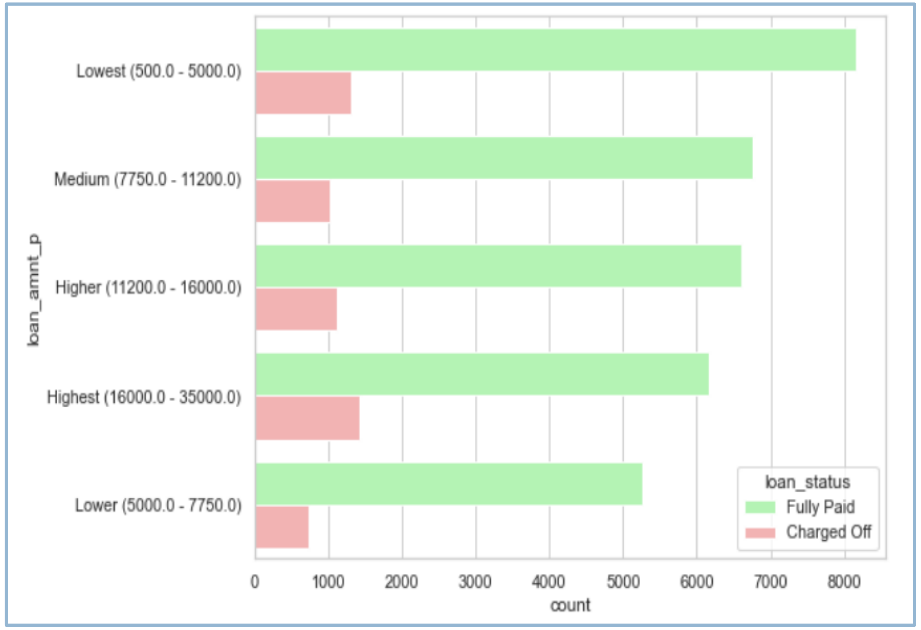
Initial data structure provided for analysis is ***(39717, 111)***

Columns with most null values are removed.

Fields created after loan approval are not helpful towards business objective and hence these are removed.

New columns derived are added to dataset from existing columns.

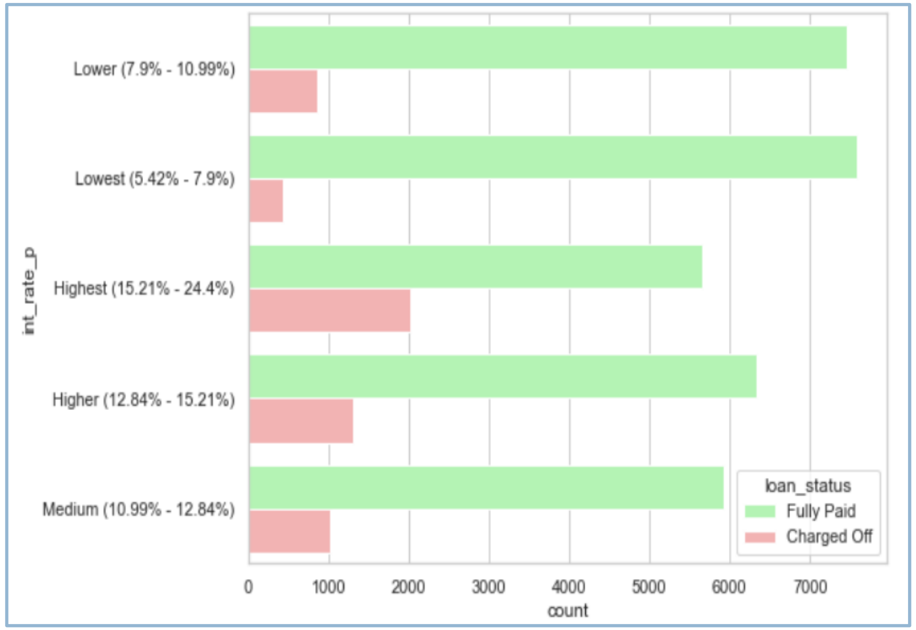
The resulted data structure after the process is ***(38577, 28)***



loan_amnt_p	Charged off %	Record count
Highest (16000.0 - 35000.0)	0.187624	7579
Higher (11200.0 - 16000.0)	0.145368	7739
Lowest (500.0 - 5000.0)	0.138725	9472
Medium (7750.0 - 11200.0)	0.131613	7788
Lower (5000.0 - 7750.0)	0.123521	5999

Data Analysis

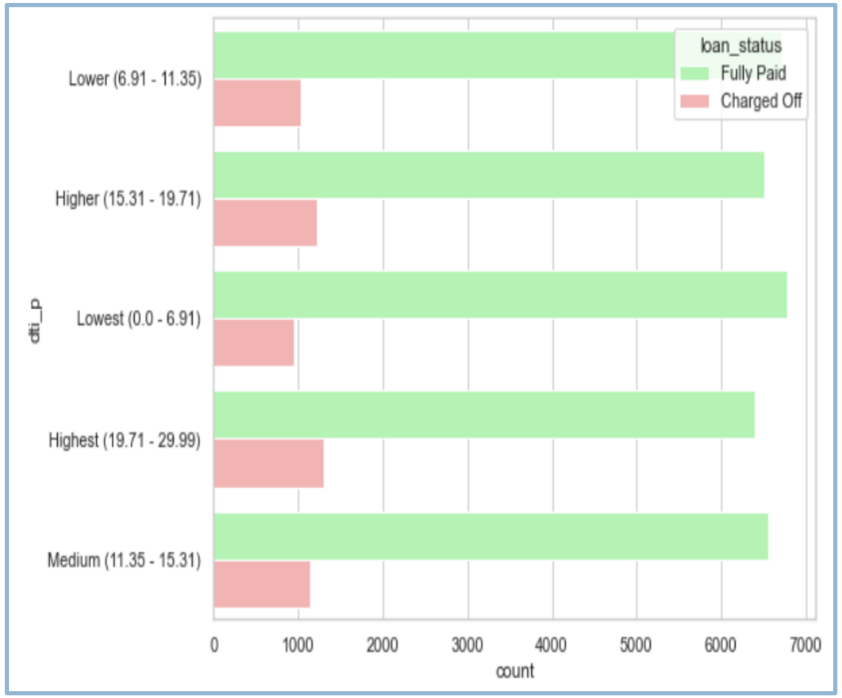
loan amount, interest rate trend on loan status



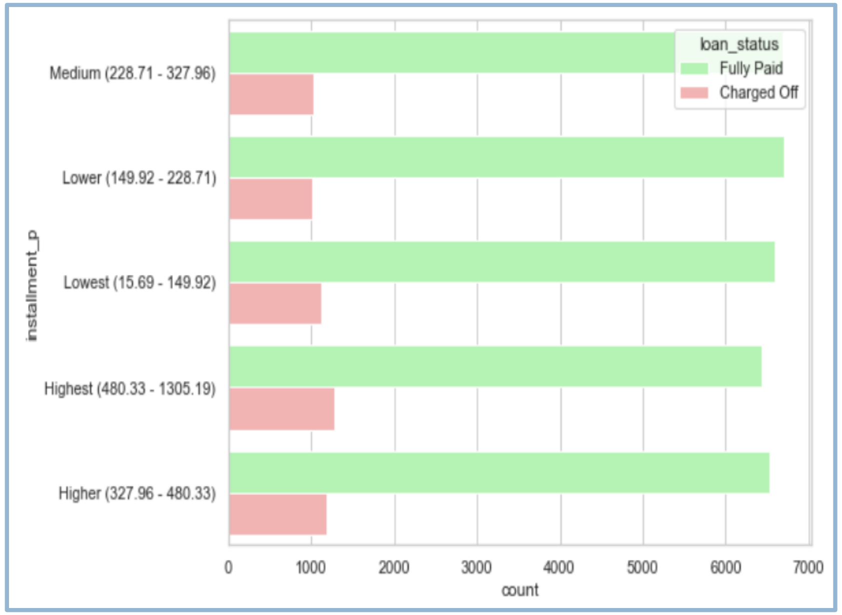
int_rate_p	Charged off %	Record count
Highest (15.21% - 24.4%)	0.262150	7675
Higher (12.84% - 15.21%)	0.171619	7639
Medium (10.99% - 12.84%)	0.146995	6939
Lower (7.9% - 10.99%)	0.102876	8311
Lowest (5.42% - 7.9%)	0.053538	8013

Both loan amount and interest rates have impact on loan defaulting.

Higher the loan amount or interest rate, higher is the chance of loan defaulting



dti_p	Charged off %	Record count
Highest (19.71 - 29.99)	0.168853	7699
Higher (15.31 - 19.71)	0.157908	7726
Medium (11.35 - 15.31)	0.147609	7696
Lower (6.91 - 11.35)	0.132627	7736
Lowest (0.0 - 6.91)	0.122409	7720



installment_p	Charged off %	Record count
Highest (480.33 - 1305.19)	0.166321	7714
Higher (327.96 - 480.33)	0.152988	7713
Lowest (15.69 - 149.92)	0.145153	7716
Medium (228.71 - 327.96)	0.132936	7718
Lower (149.92 - 228.71)	0.131934	7716

Data Analysis

debt-to-income, installment trend on loan status

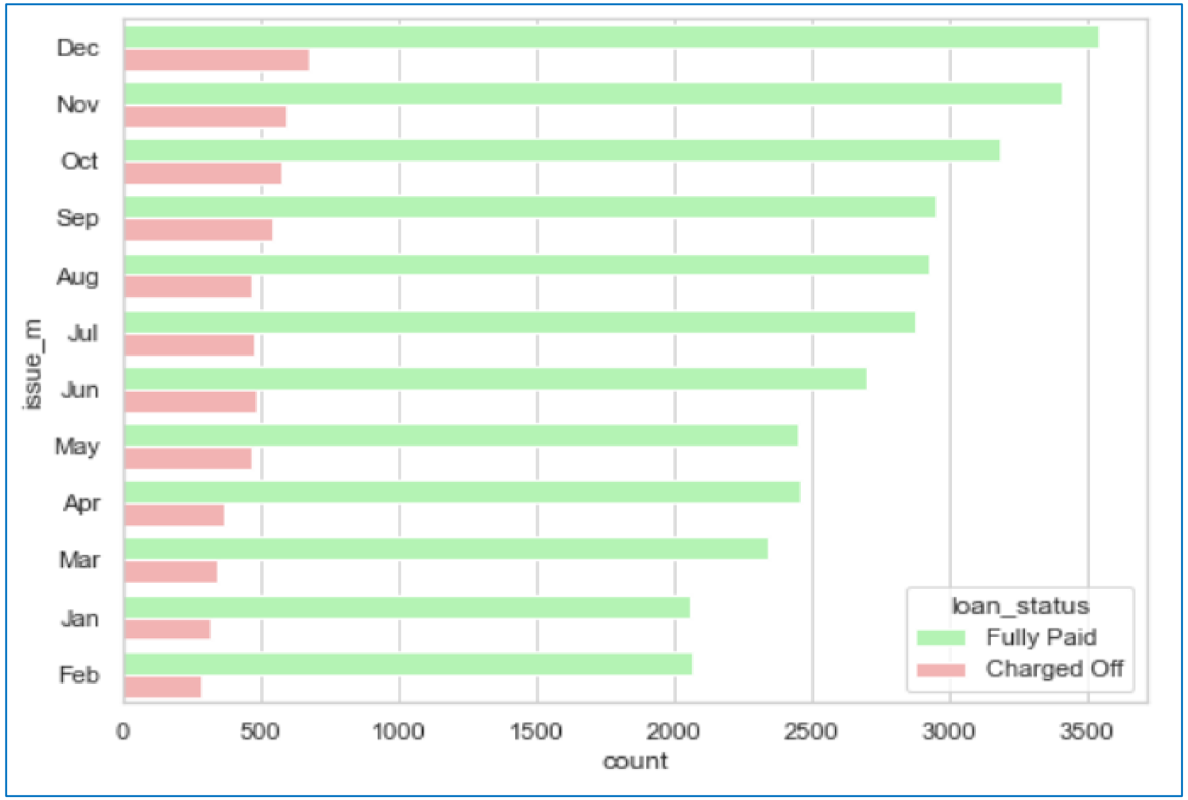
Installment amount and debt to income ratio also show positive impact on loan defaulting.

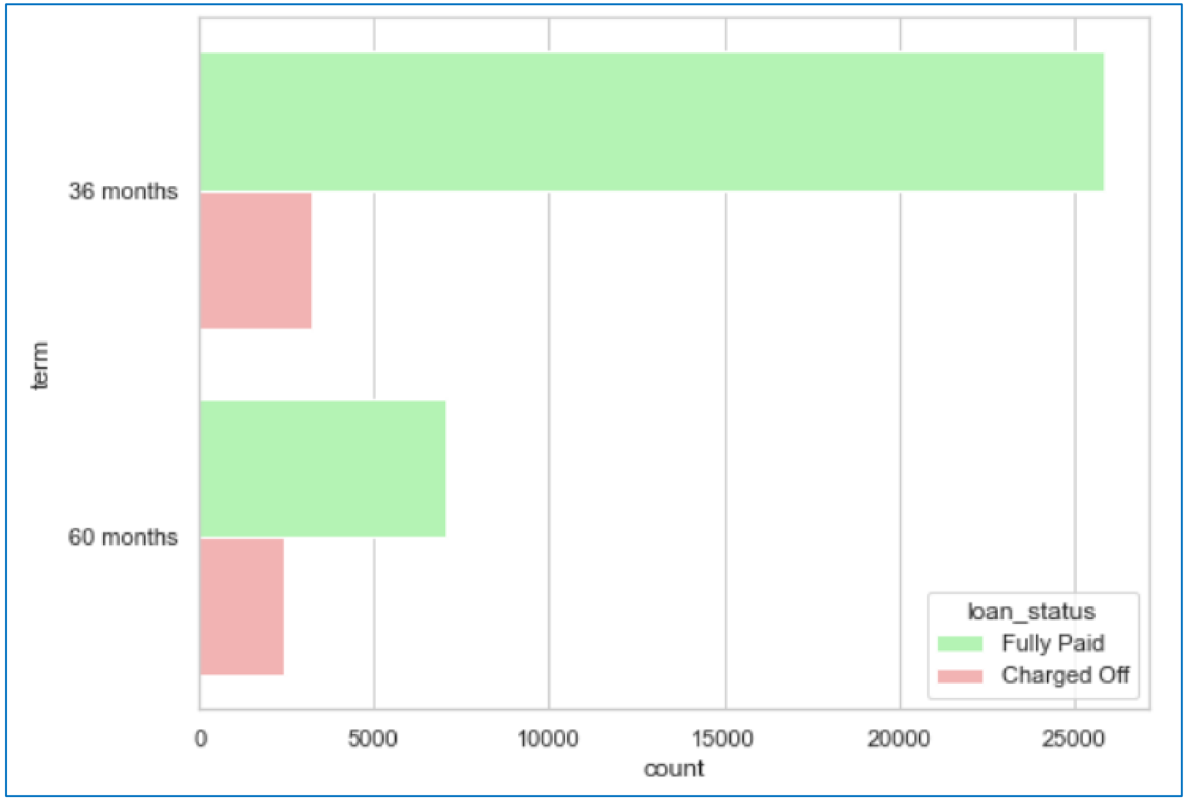
Higher the installment or debt-to-income ratio, higher is the chance of loan defaulting

Impact of term and issue month on loan repayment

- December being Holiday season for Christmas, it is the month which has the highest number of loan applications and have the biggest default ratio.
- Month of May is also another one, which is during the summer break in US where people love to travel.

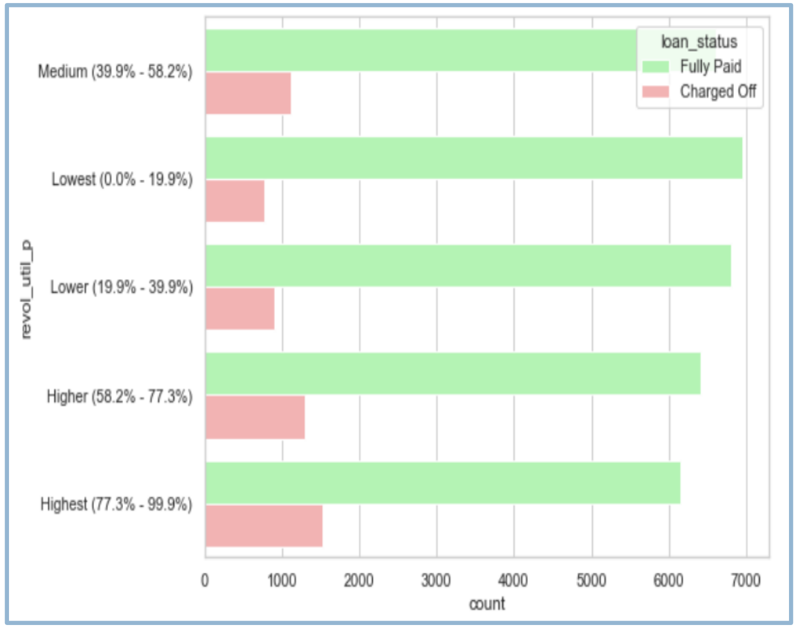
- For loans with 5-year repayment term, the default percent is 25%. And for 3-year loan repayment term, the default is only for 11% of the cases.
- Therefore, loan repayment term plays a major factor in judging the default rate.



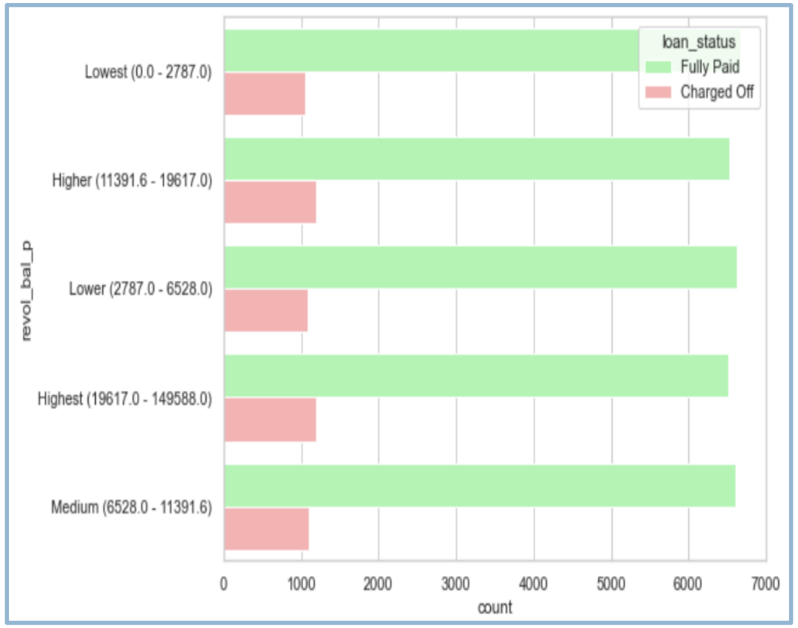


Data Analysis

revolving utilization rate, total credit revolving balance trend on loan status



revol_util_p	Charged off %	Record count
Highest (77.3% - 99.9%)	0.198069	7664
Higher (58.2% - 77.3%)	0.167727	7703
Medium (39.9% - 58.2%)	0.145001	7731
Lower (19.9% - 39.9%)	0.118356	7714
Lowest (0.0% - 19.9%)	0.099417	7715



revol_bal_p	Charged off %	Record count
Highest (19617.0 - 149588.0)	0.154913	7714
Higher (11391.6 - 19617.0)	0.154335	7717
Medium (6528.0 - 11391.6)	0.142894	7712
Lower (2787.0 - 6528.0)	0.141283	7715
Lowest (0.0 - 2787.0)	0.135898	7719

High revolving utilization rate has positive impact on loan defaulting.

Revolving credit balances also shows some impact on loan defaulting, though the difference is minimal

However, high utilization has high chance of loan defaulting

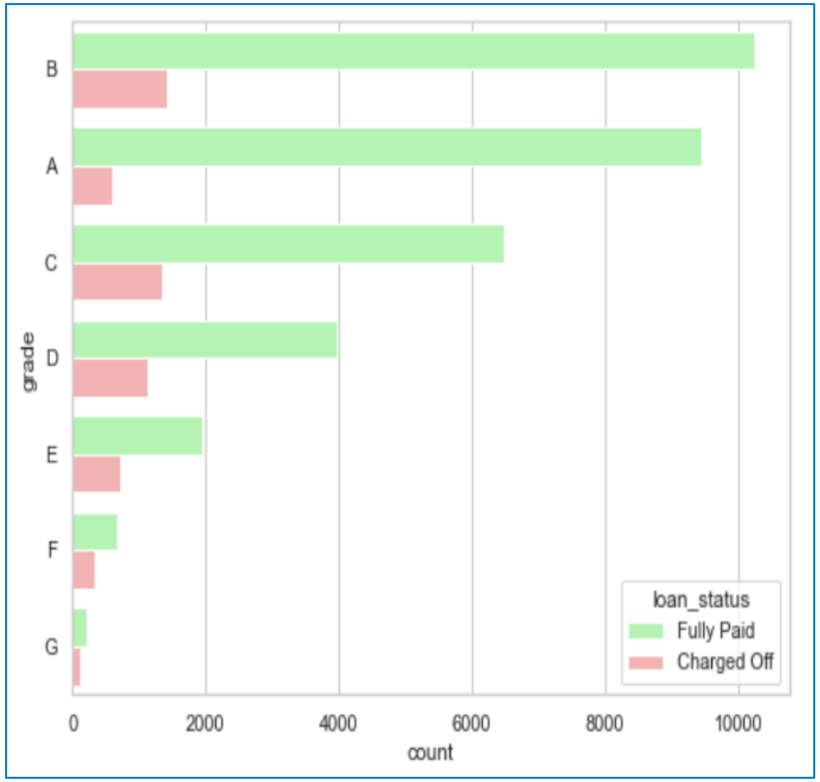
Inferences:

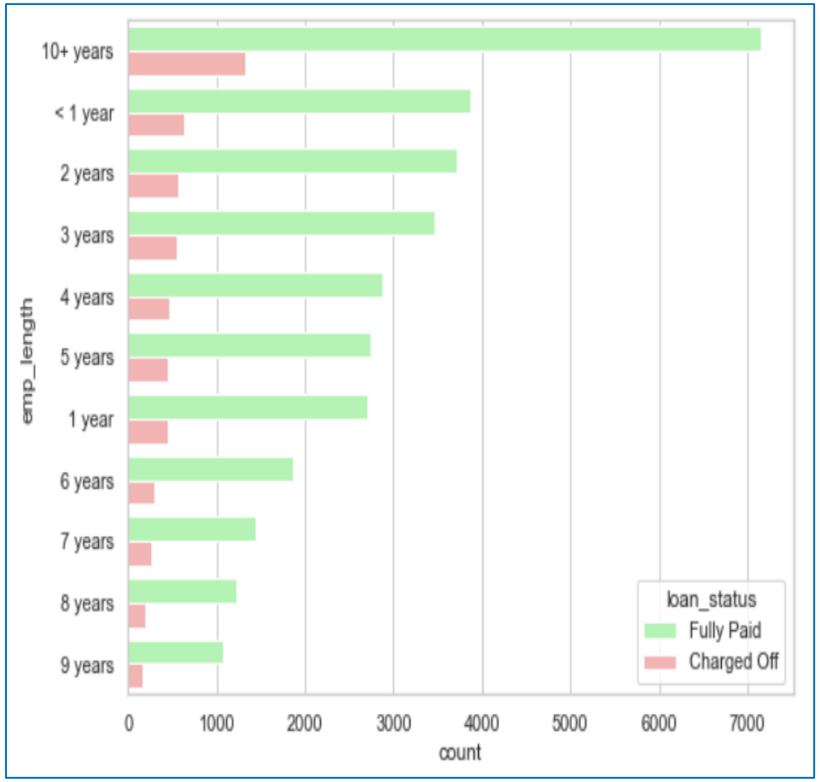
Verified borrowers have higher ratio of loan charged off compared to Not verified users.

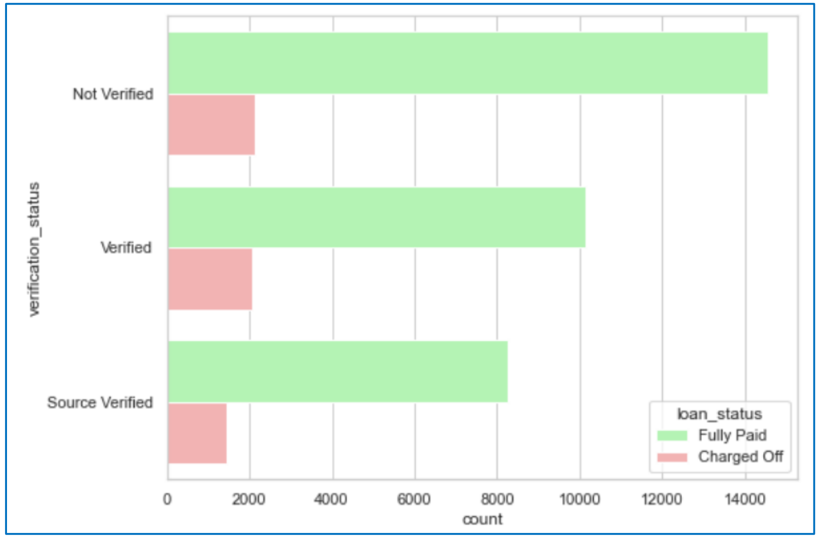
This raises questions on the verification process followed by the company

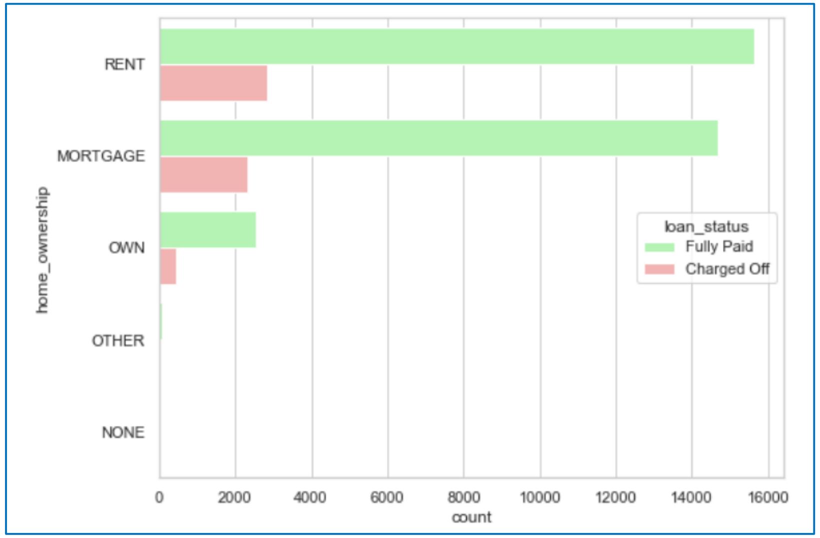
Employment length of borrower and home ownership doesn't give much information regarding loan default

It is evident that loan grades having highest default percentages. G, F, E and D form grades where default rate is much higher than others.





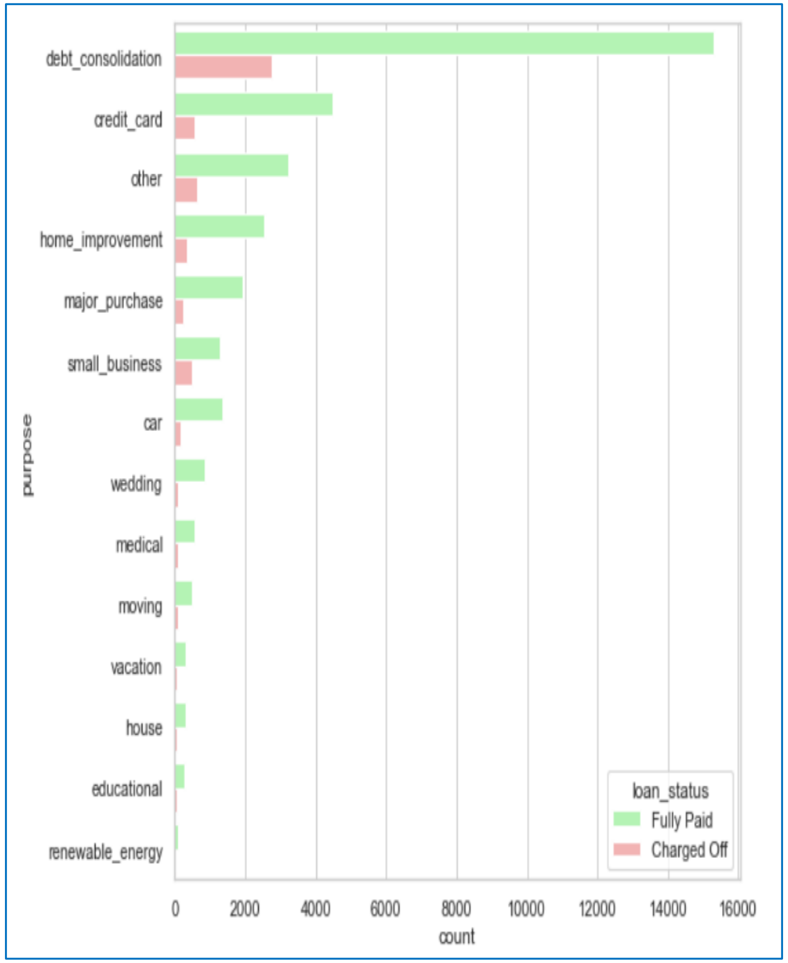


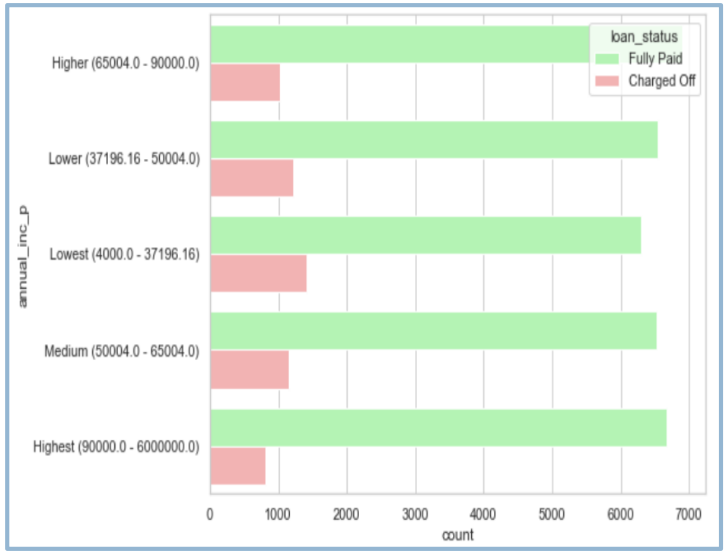


Data Analysis

annual income, purpose of loan trend on loan status

annual_inc_p	Charged off %	Record count
Lowest (4000.0 - 37196.16)	0.183126	7716
Lower (37196.16 - 50004.0)	0.156926	7768
Medium (50004.0 - 65004.0)	0.149629	7679
Higher (65004.0 - 90000.0)	0.129651	7929
Highest (90000.0 - 6000000.0)	0.109285	7485





purpose	Charged off %	Record count
small_business	0.270810	1754
renewable_energy	0.186275	102
educational	0.172308	325
other	0.163777	3865
house	0.160763	367
moving	0.159722	576
medical	0.155653	681
debt_consolidation	0.153254	18055
vacation	0.141333	375
home_improvement	0.120696	2875
credit_card	0.107818	5027
car	0.106738	1499
wedding	0.103672	926
major_purchase	0.103256	2150

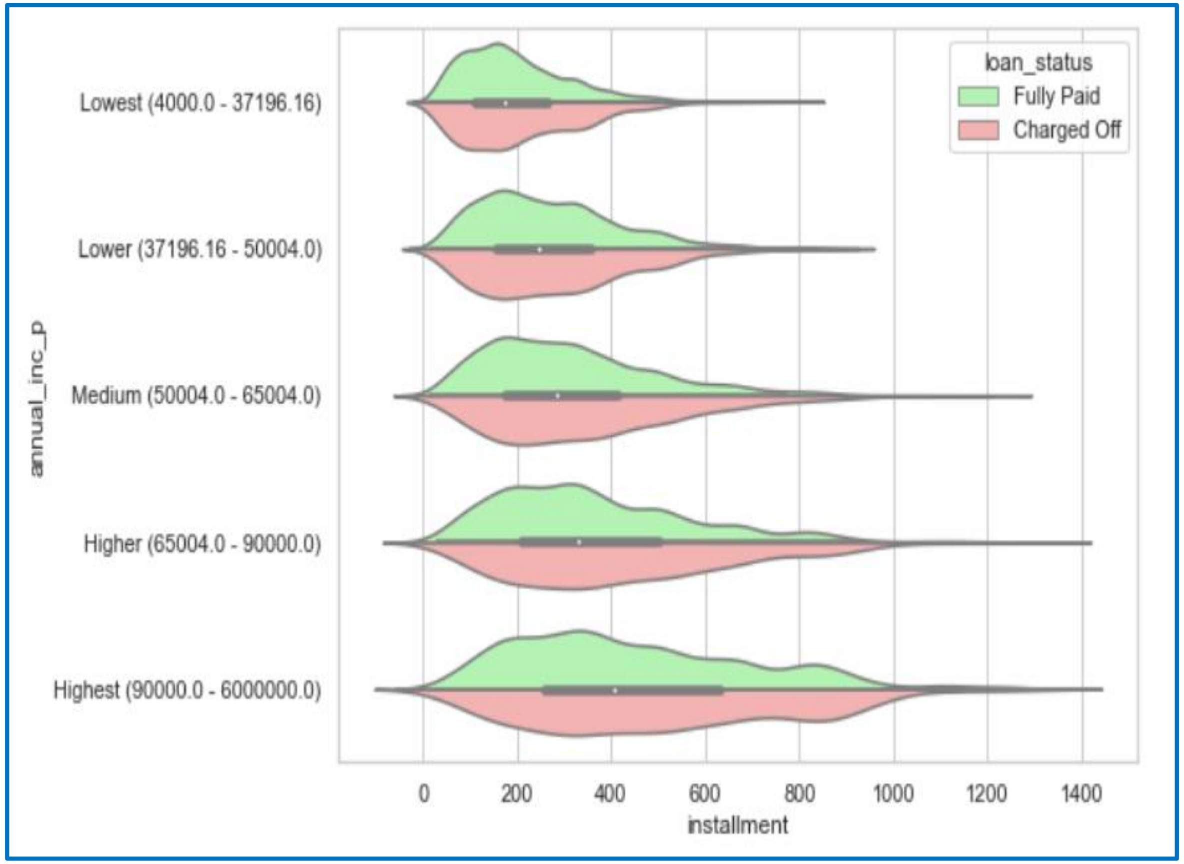
Higher the annual income of the borrower higher is the repayment percent

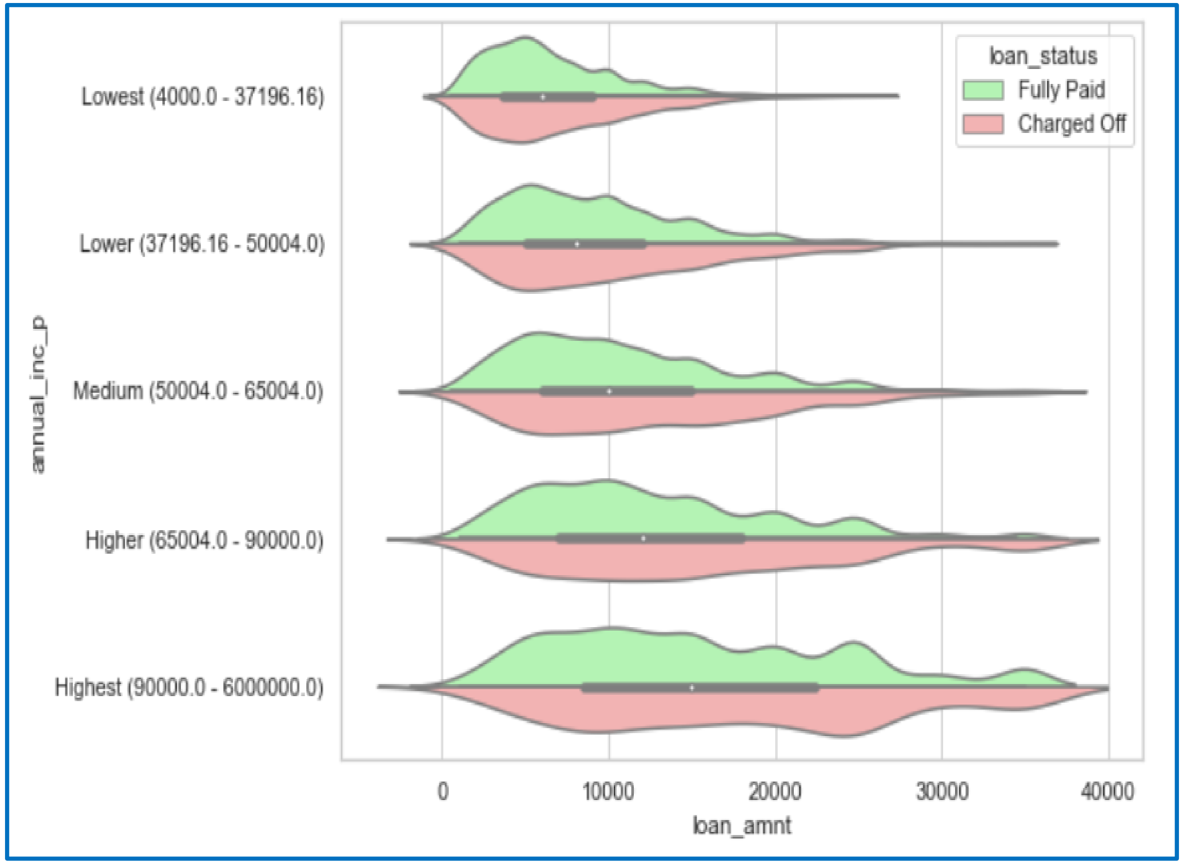
Lesser income will increase the risk of loan default

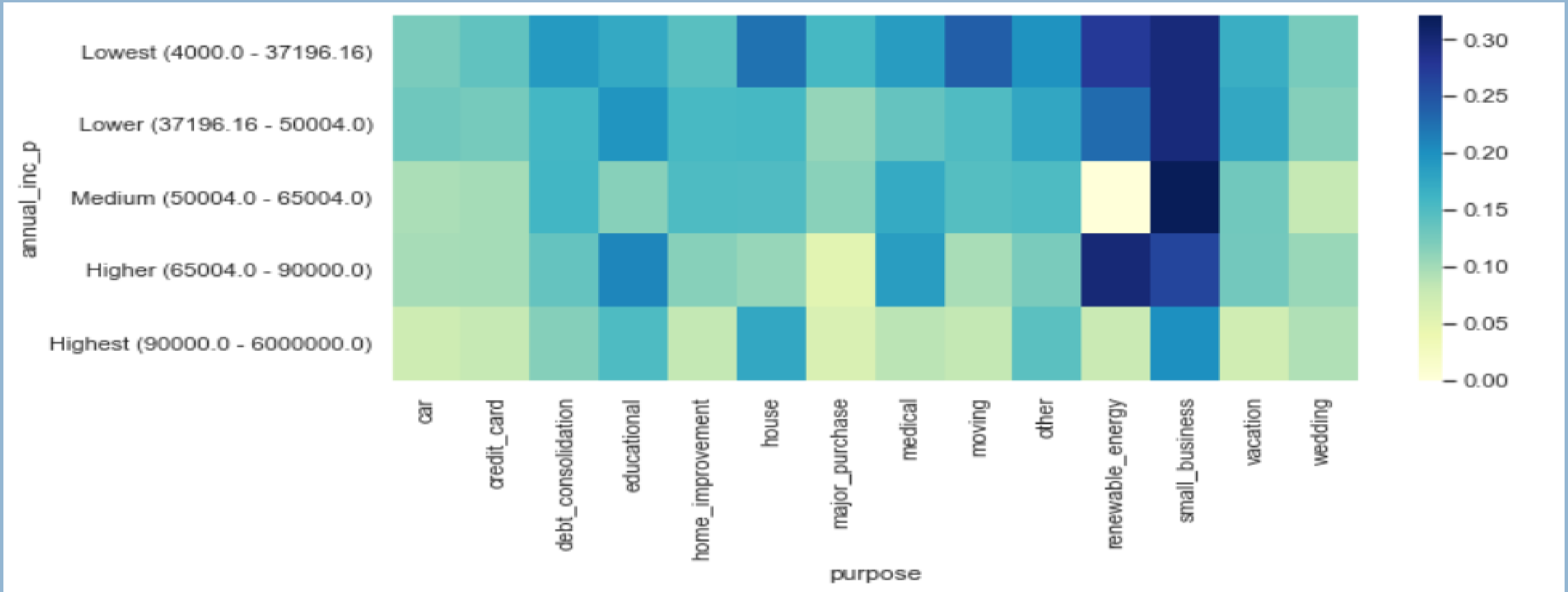
From the analysis it is evident that the loans taken for **small_business**, **renewable_energy** and **educational** are the riskier ones.

Data distribution analysis

- Plotted Graphs show that for any income group, higher installments and higher loan amount have a greater number of defaults.





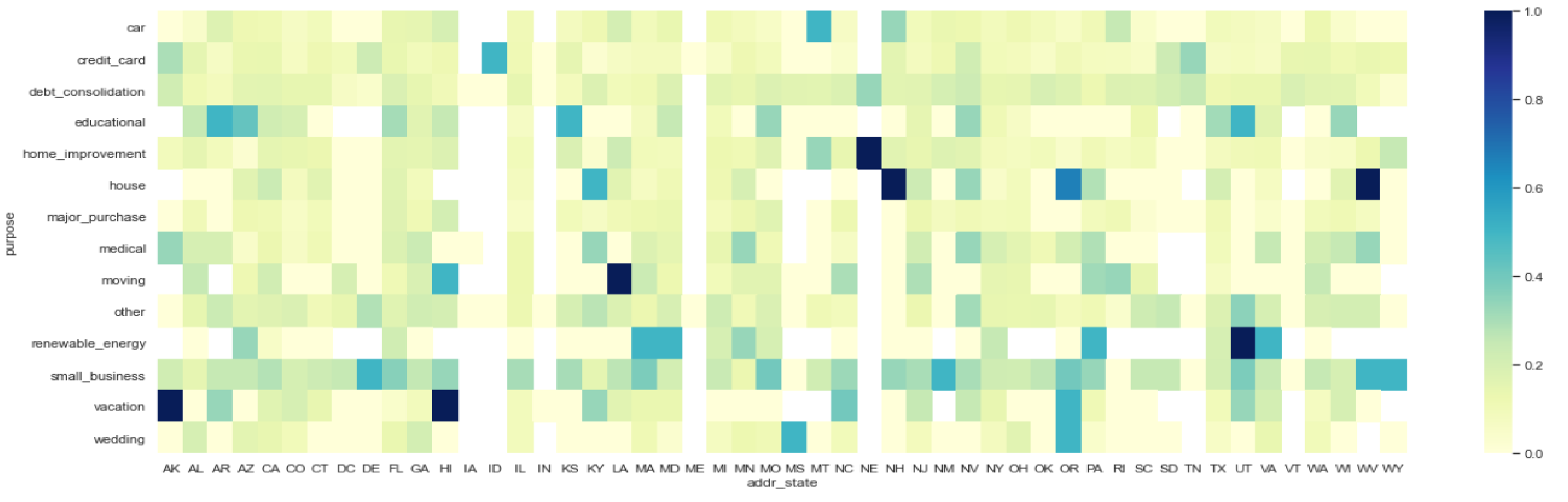


Plot showing the risks of income groups versus the purpose of loan.

- Small business loans for low-income and middle-income groups.
- Renewable energy loans for high income group and lowest income group

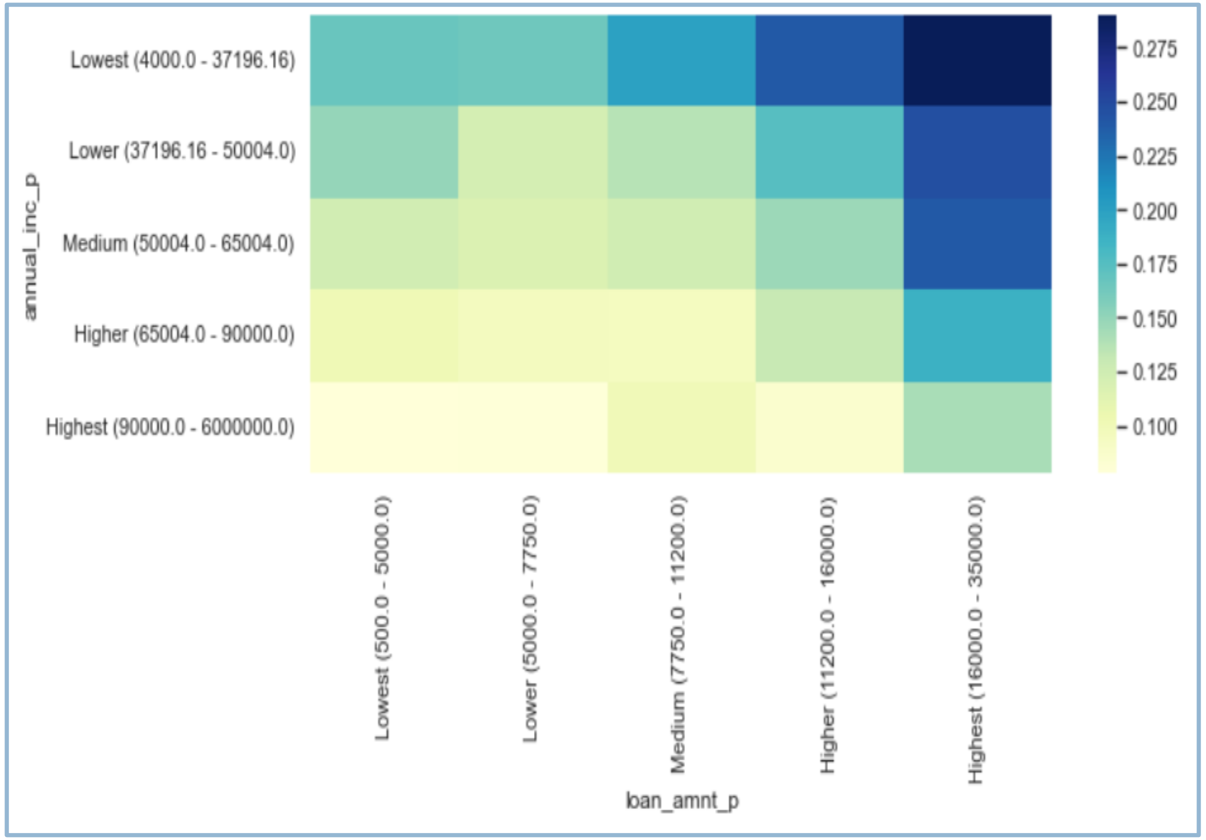
Purpose of loan and state of borrower correlation Analysis

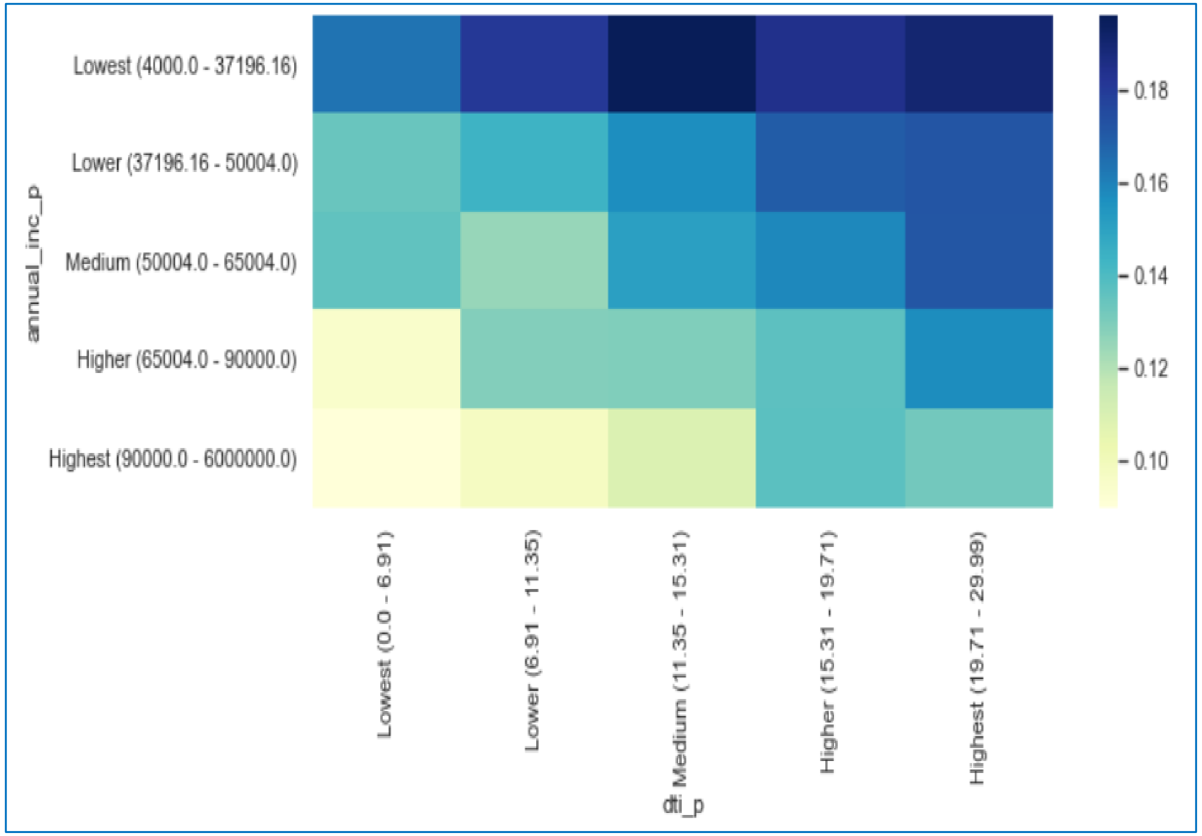
- As per the plot, the higher the correlation between addr_state has and purpose of the loan, the risker the loan application is. Some of the examples are below:
    - Vacation loans in AK, HI, OR
    - Education loans in AR, KS, UT
    - Small business loans in DE, NM, WV, WY
    - Wedding loans in MS, OR



Annual income variable impact analysis on loan repayment

- Medium or High debt-to-income group in the low-income range is the riskiest in terms of loan repayment.
- Approving higher loan amount for a low-income borrower is with the Highest risk.
- Therefor, lower income group with other high-risk factors, have least chance of full loan repayment, due to financial crisis





Inferences from EDA

Driving factors identified with high impact on loan default

- Higher interest rate
- Higher revolving line utilization rate
- Loan repayment term
- Missing employment record
- Derogatory public records
- Public bankruptcy records
- Loan purpose

Driving factors with Minor impact on loan default

- Higher loan amount
- Higher installment amount
- Lower annual income
- Applicant's address state
- Loan issue month

Combination of Factors that have high impact on loan default

- High loan amount for lower income group
- Low income with Debt-to-income ratio
- High interest rate for lower income group
- High installment and longer repayment term
- Home ownership (other) and loan purpose (car, moving or small business)
- Residential state and loan purpose
- Income group and loan purpose

Recommendations

- Loan requests for large sum by lower income groups with higher interest rates should be either denied or sum invested should be reduced
- Loan requests for longer term period with huge amount by any income groups should be properly analyzed and revisited before the loan approval
- Loan request of small capital by mid or high-income groups can be approved, as they have high chance of loan repayment
- Verification process for loan should be further analyzed and revised, as this factor is not helpful in loan default detection
- Home ownership factor should be segregated properly with definite options to further utilize this factor in loan default detection
- Any loan requests with Combined risk factors from the inferences list should be re analyzed before the approval

THANK YOU !