Lending Club Case Study

Group Facilitator

Kuljeet Babber
Email ID: kuljeetbabber@gmail.com

Team Member

Anik Chowdhury
Email ID: dashinganik@gmail.com

Batch

ML C50

upGrad

Contents

Problem Statement

Lending Club is the leading online loan marketplace, offering a wide range of financial products including personal loans, business loans, and financing for medical procedures. One of the biggest challenges for lenders is managing credit risk, which is the financial loss that occurs when borrowers default on their loans.

To address this challenge, Lending Club is focused on understanding the key drivers behind loan defaults, including the factors that strongly indicate when a borrower is likely to default. By identifying these key variables, Lending Club can better manage its portfolio and assess risks to ensure the best possible outcomes for both borrowers and investors.

Problem solving approach
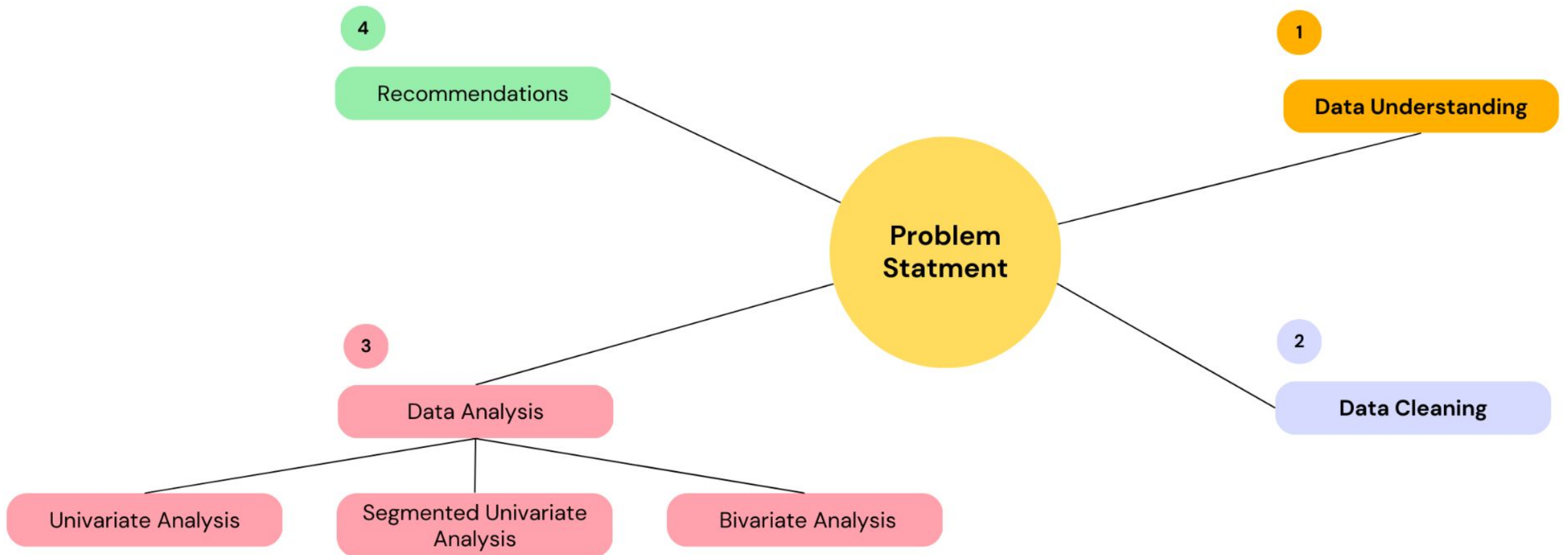
Data Cleaning

❖ We removed columns with Null % greater than 30 %
❖ We removed trailing spaces from the categorical data
❖ We dropped columns which had 1 unique values
❖ We dropped columns which had no significant impact on loan default prediction
❖ We dropped records for which loan payment is in progress
❖ loan_amnt , funded_amnt , int_rate & revol_util casted to float as they are amount & rate of interest respectively
❖ issue_d converted to date format and month & year derived as additional columns
❖ loan_status_num created to identify loan status numerically

Data Analysis & Observation

Interest Rates

loan_status vs Interest Rate
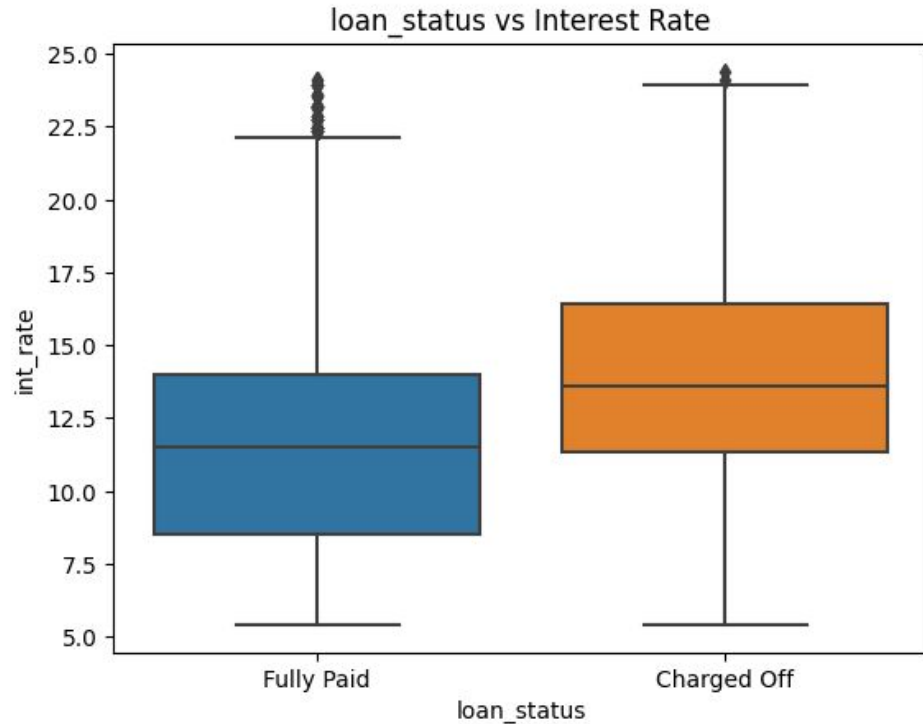
1. Interest rates less than 10% have very low chances of resulting in charge-offs. The minimum interest rate for charge-offs is 5.4%.
2. According to the quartile distribution, interest rates above 14% have a higher probability of resulting in default.
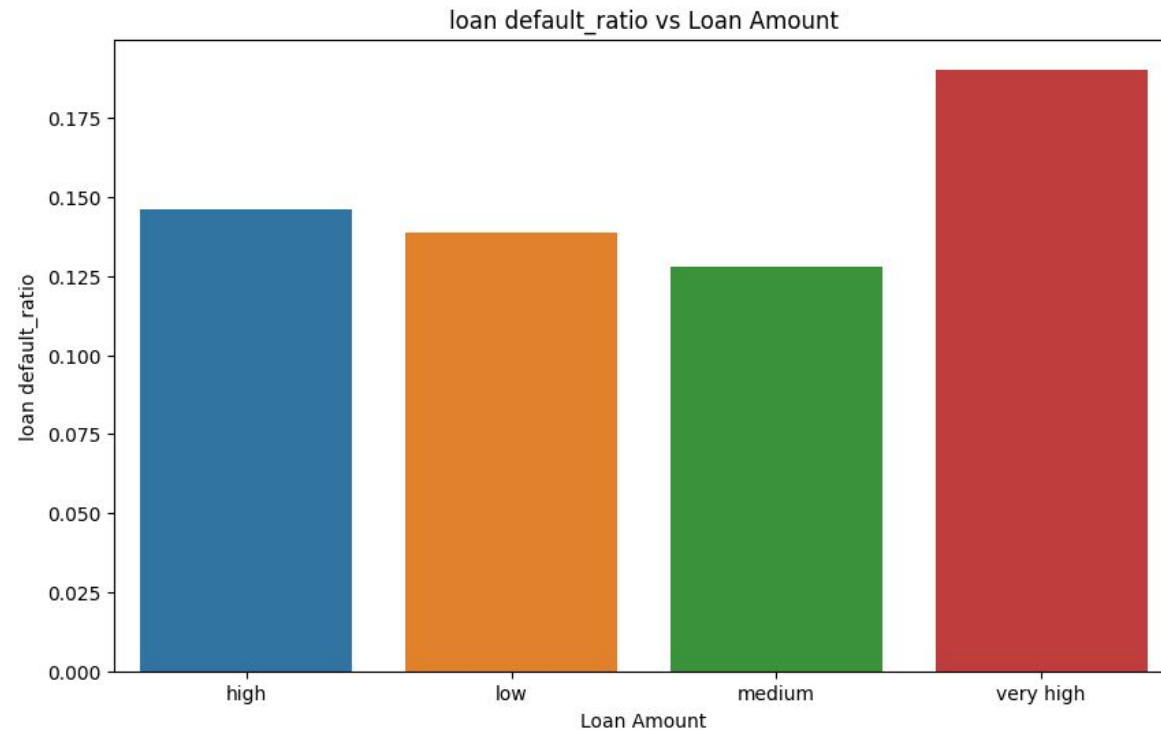3. The proportion of charge-offs increases with higher interest rates.

Term

Loan Status vs Term
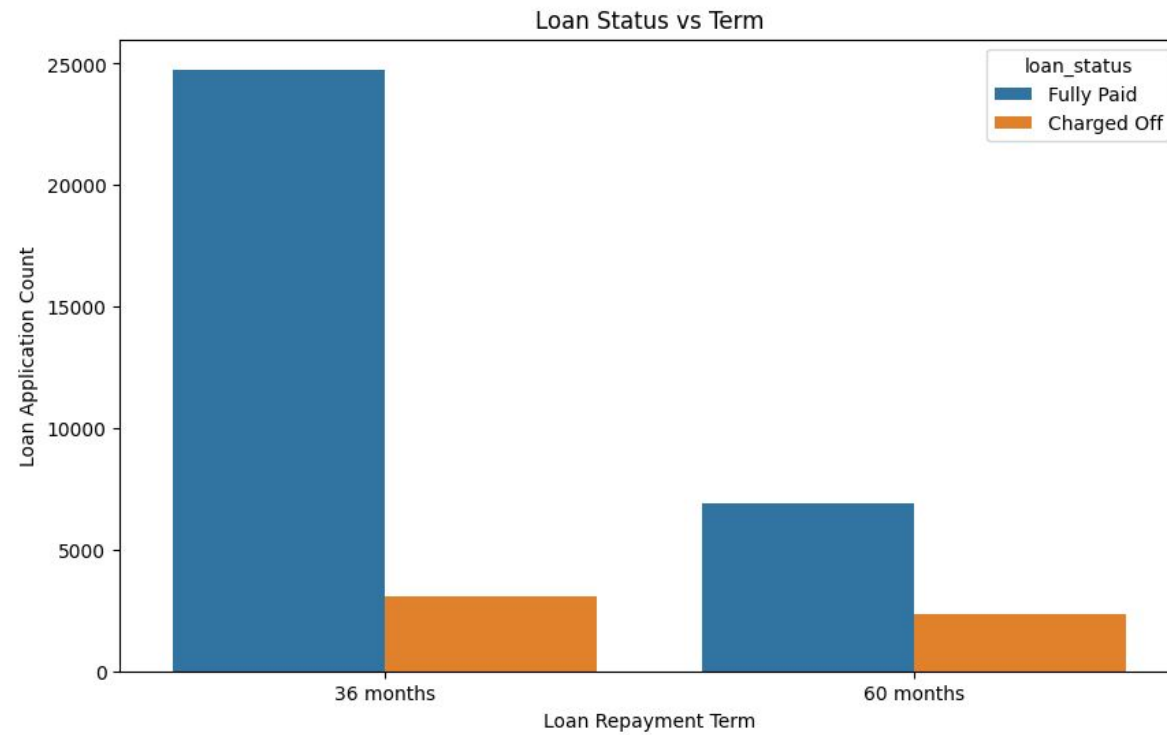
People who have taken loan for a tenure of 60 months have higher chances of defaulting
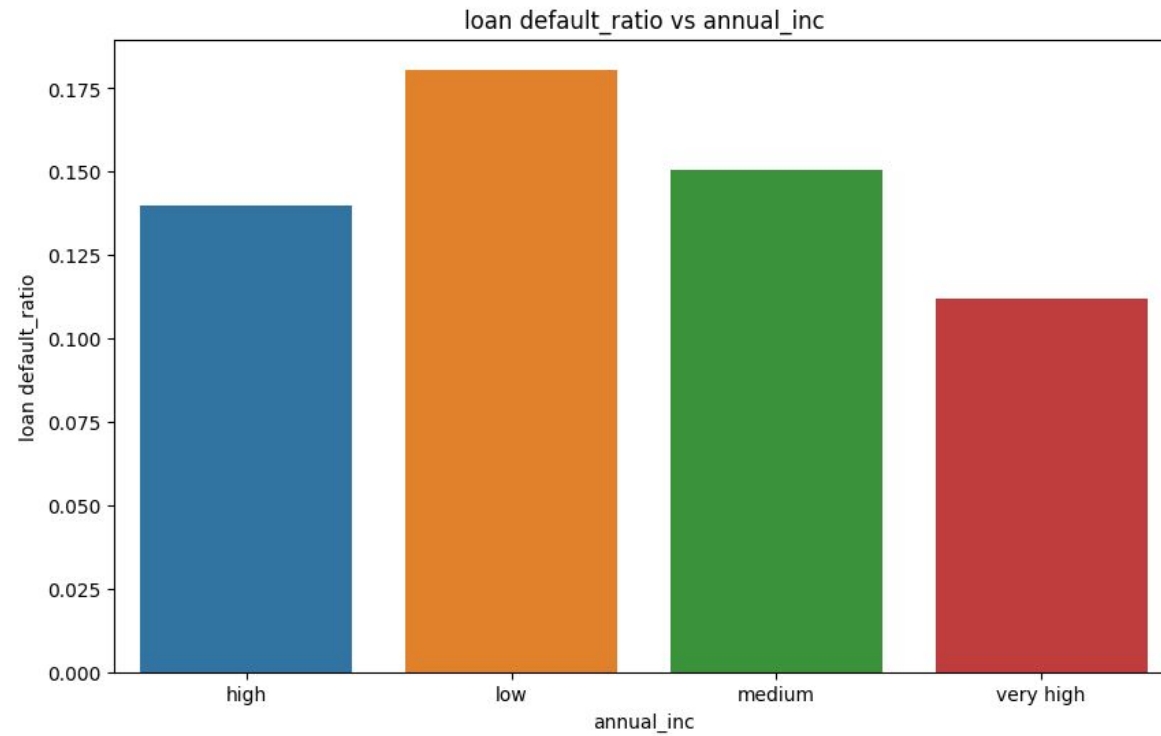
Loan Amount

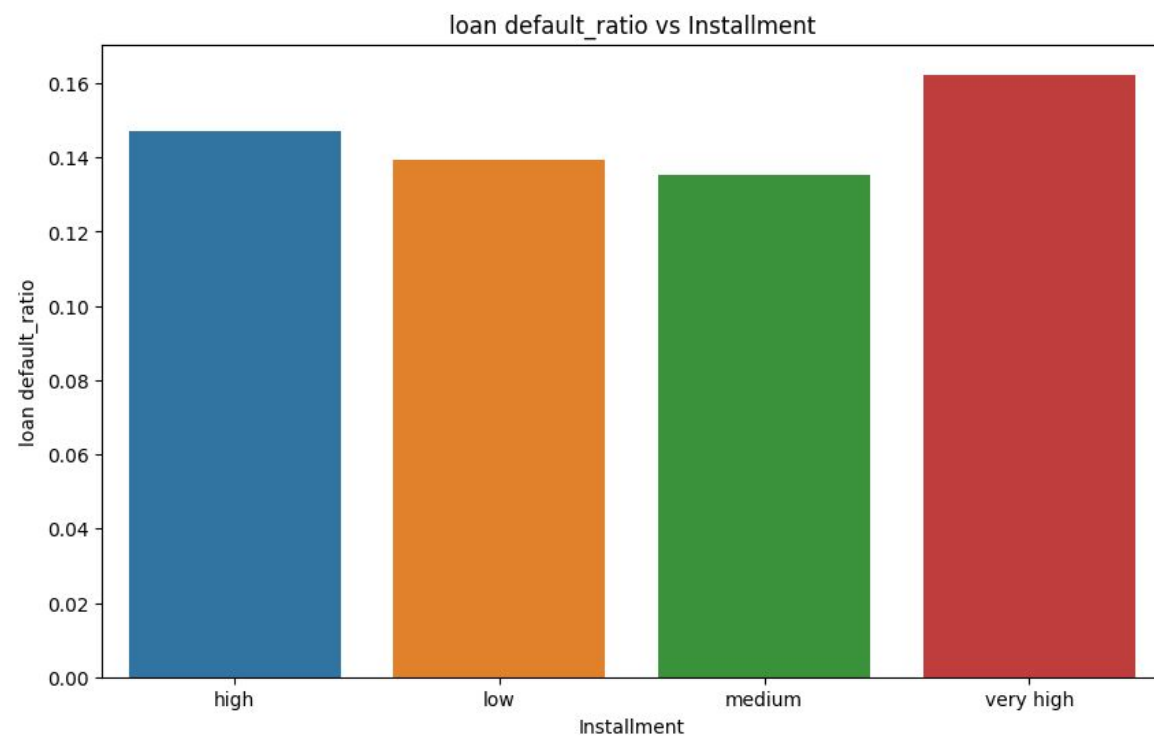loan default_ratio vs Loan Amount

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

Annual Income

Low Annual Income have slight chances to default

Installment

loan default_ratio vs Installment

Higher installment have around 16 % chances of default

Grade

Loan Status vs Grade
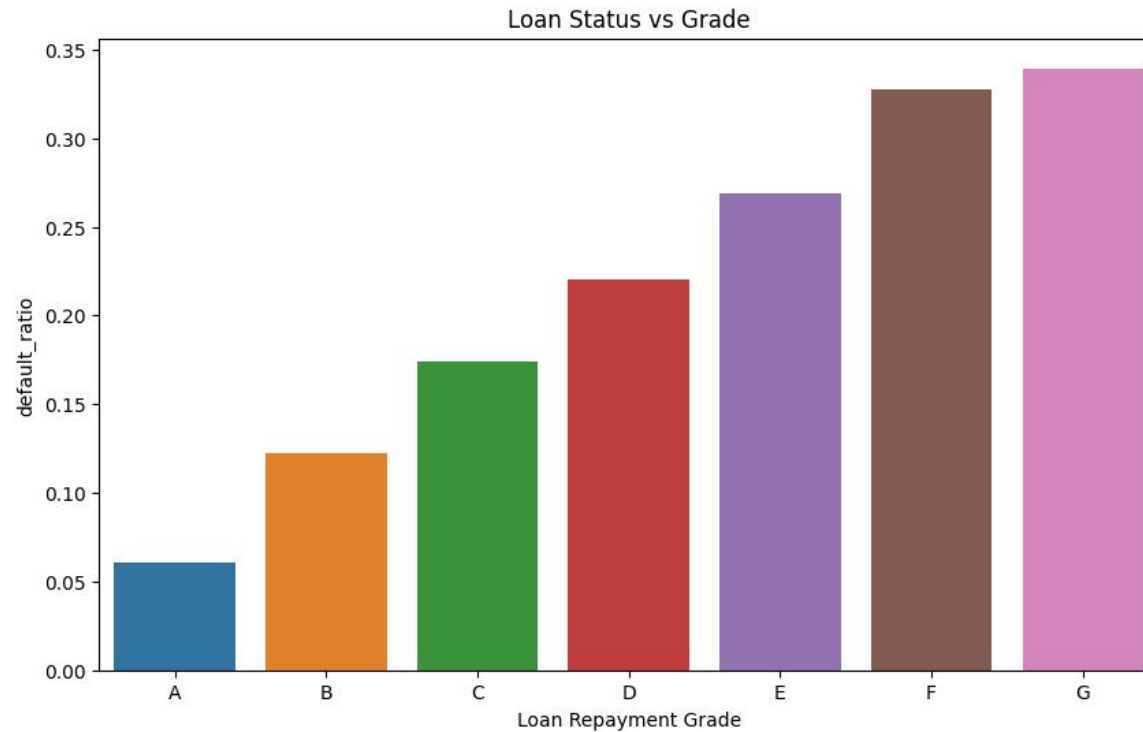
1. Grade F & G have highest chances of default
2. Grade A has the least chance of default
3. As we saw earlier Grade and sub grade have same trend hence this behaviour would be same in subgrade also , where in subgrades of F & G would be high risk to default
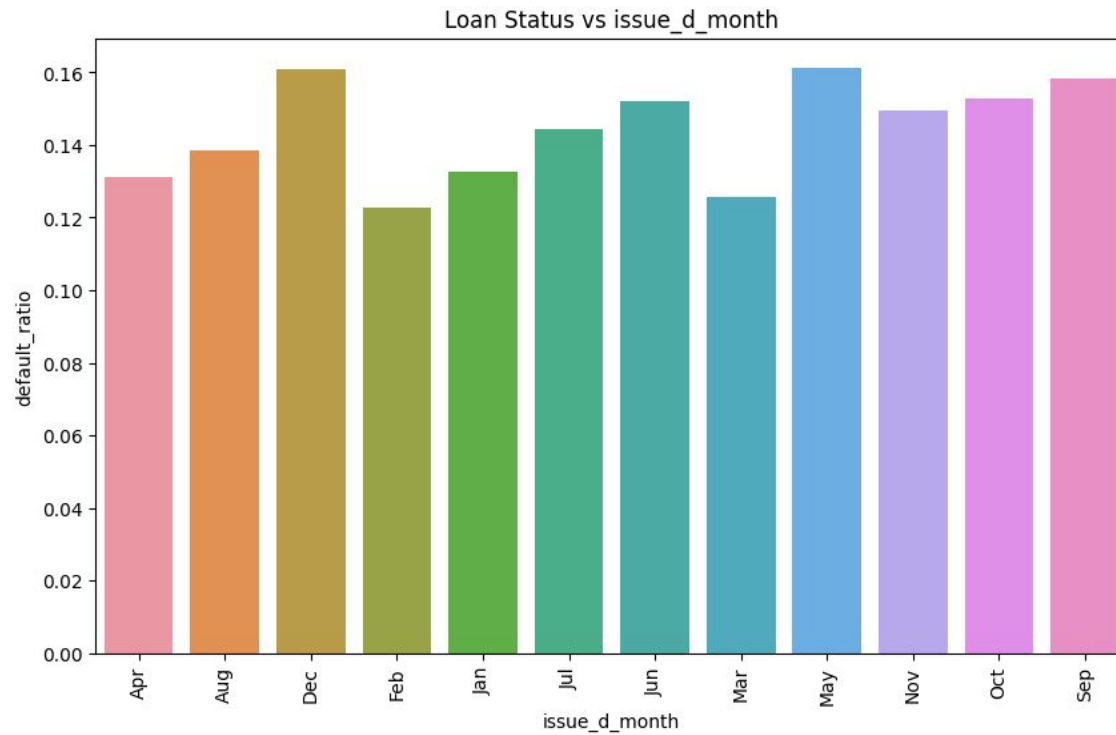
Home Ownership
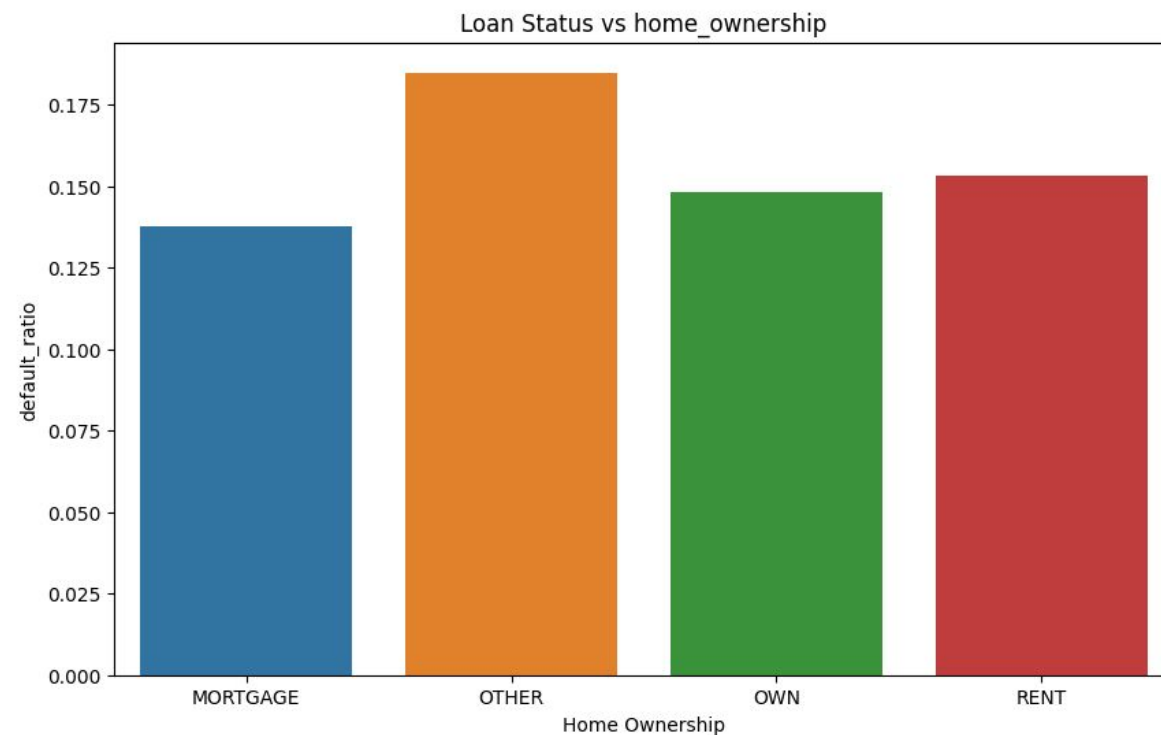
There is slight impact of other category in home_ownership to default

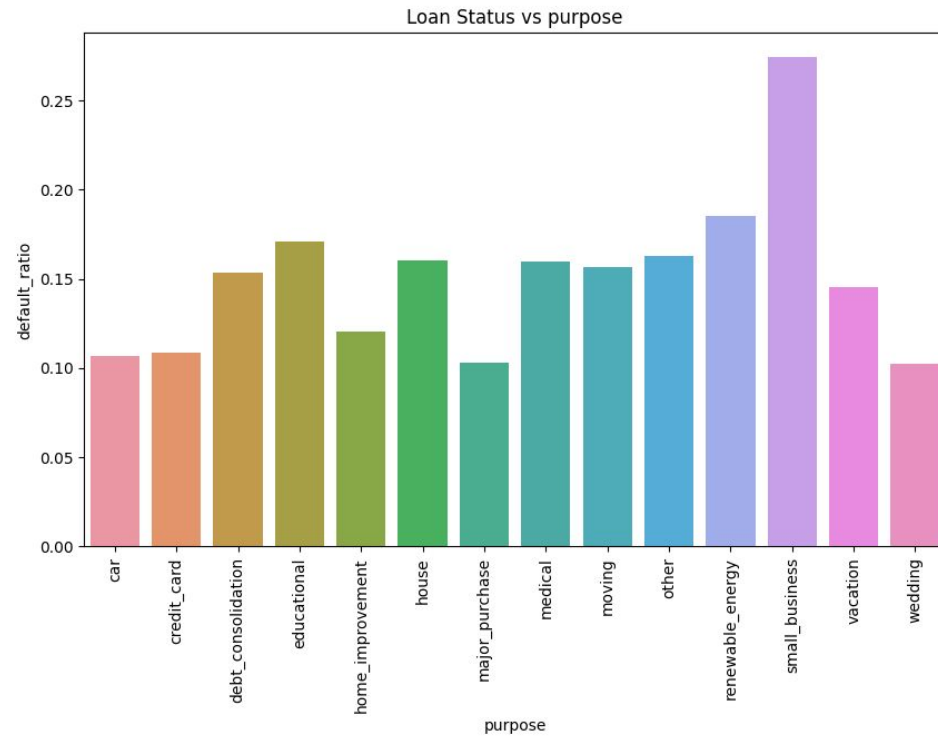# Address State

Loan Status vs addr_state
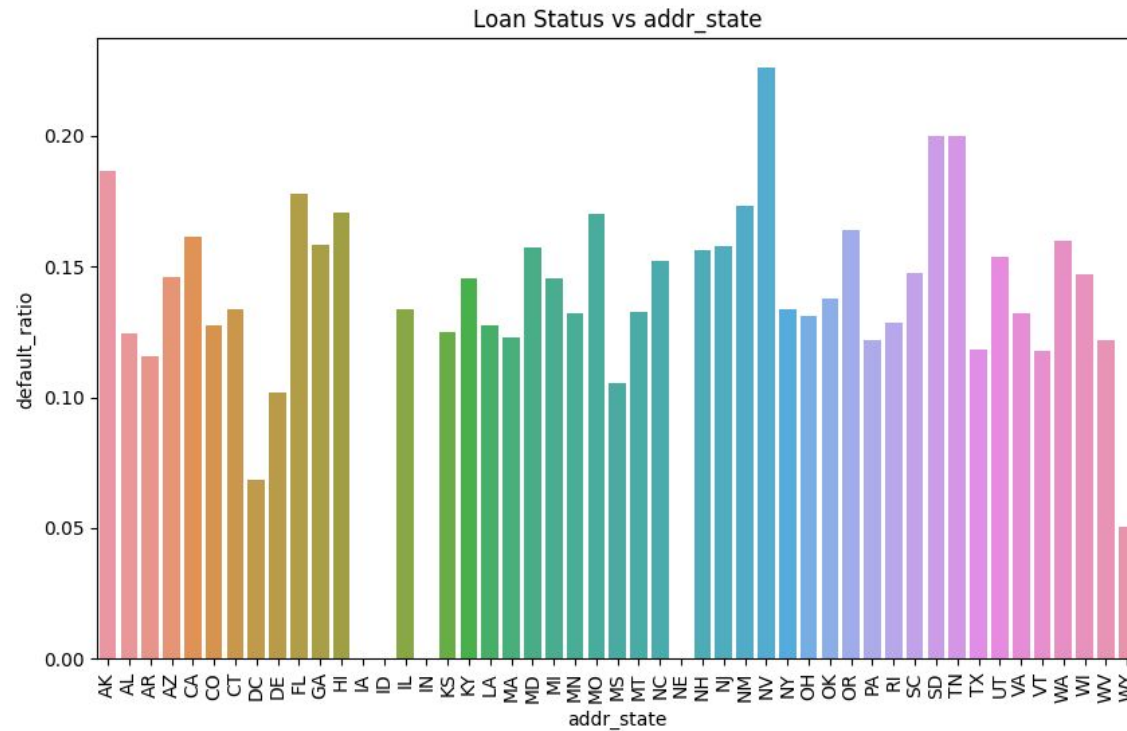
High chances of default are in NV,SD,TN,NM,MO & the least are in IA,ID,IN,NE

Issue Month

The issuance of loans in May, September, October, November, and December has minimal impact on loan defaults.
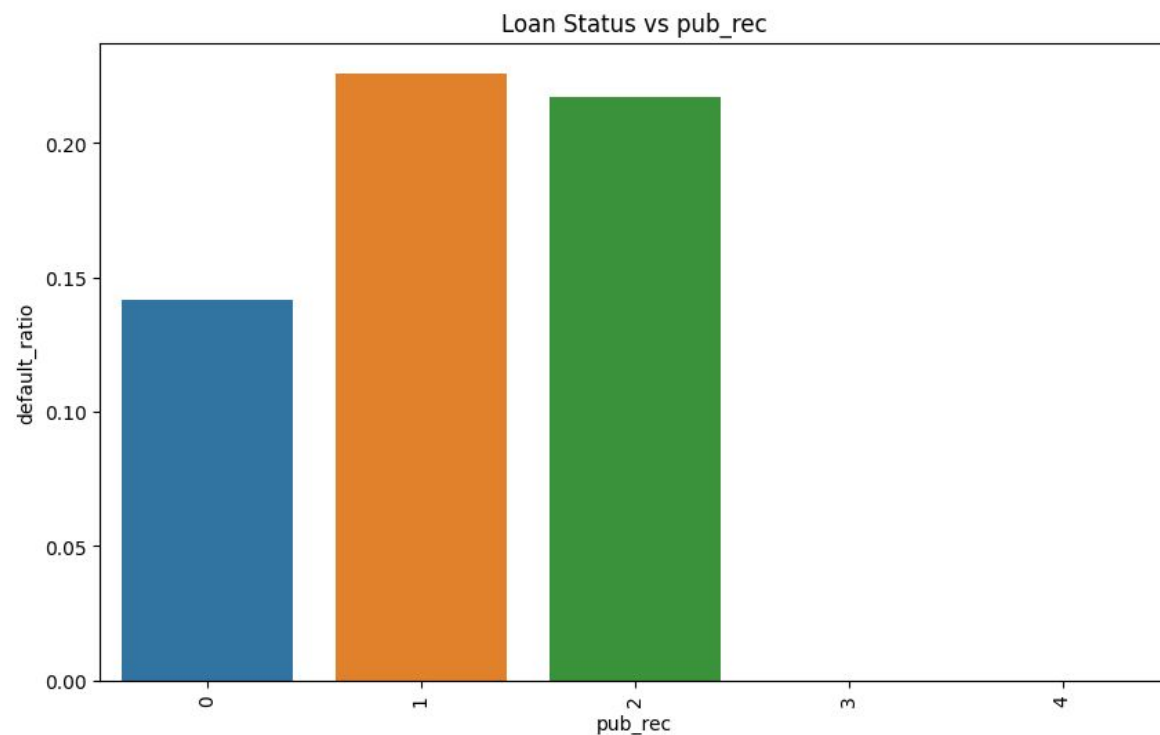
Purpose

Loan Status vs purpose

Small businesses, renewable energy, and educational purposes have a higher probability of defaulting on their loans.

Number of derogatory public records

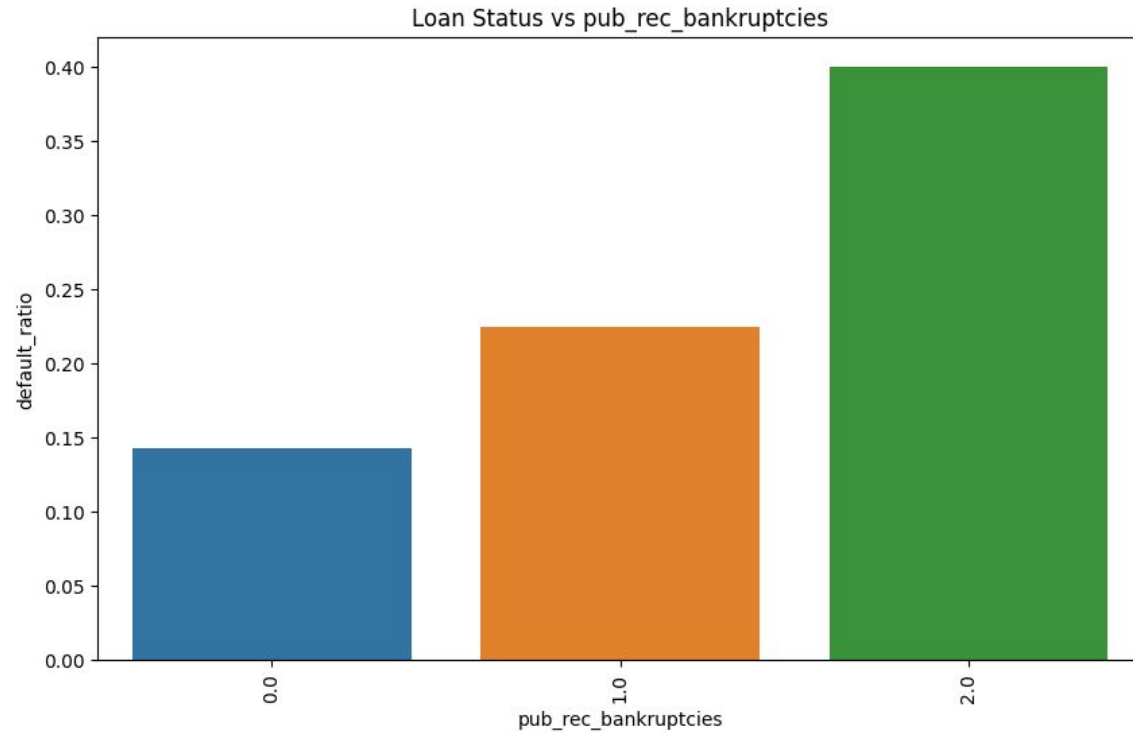Loan Status vs pub_rec

Borrowers with derogatory public records, specifically those with public records marked as 1 or 2, have a higher likelihood of defaulting on their loans.

Number of public record bankruptcies

Loan Status vs pub_rec_bankruptcies

Borrowers with public bankruptcy records (1 & 2) are more likely to default on their loans.

Correlation between addr_state & purpose wrt Defaulting chances
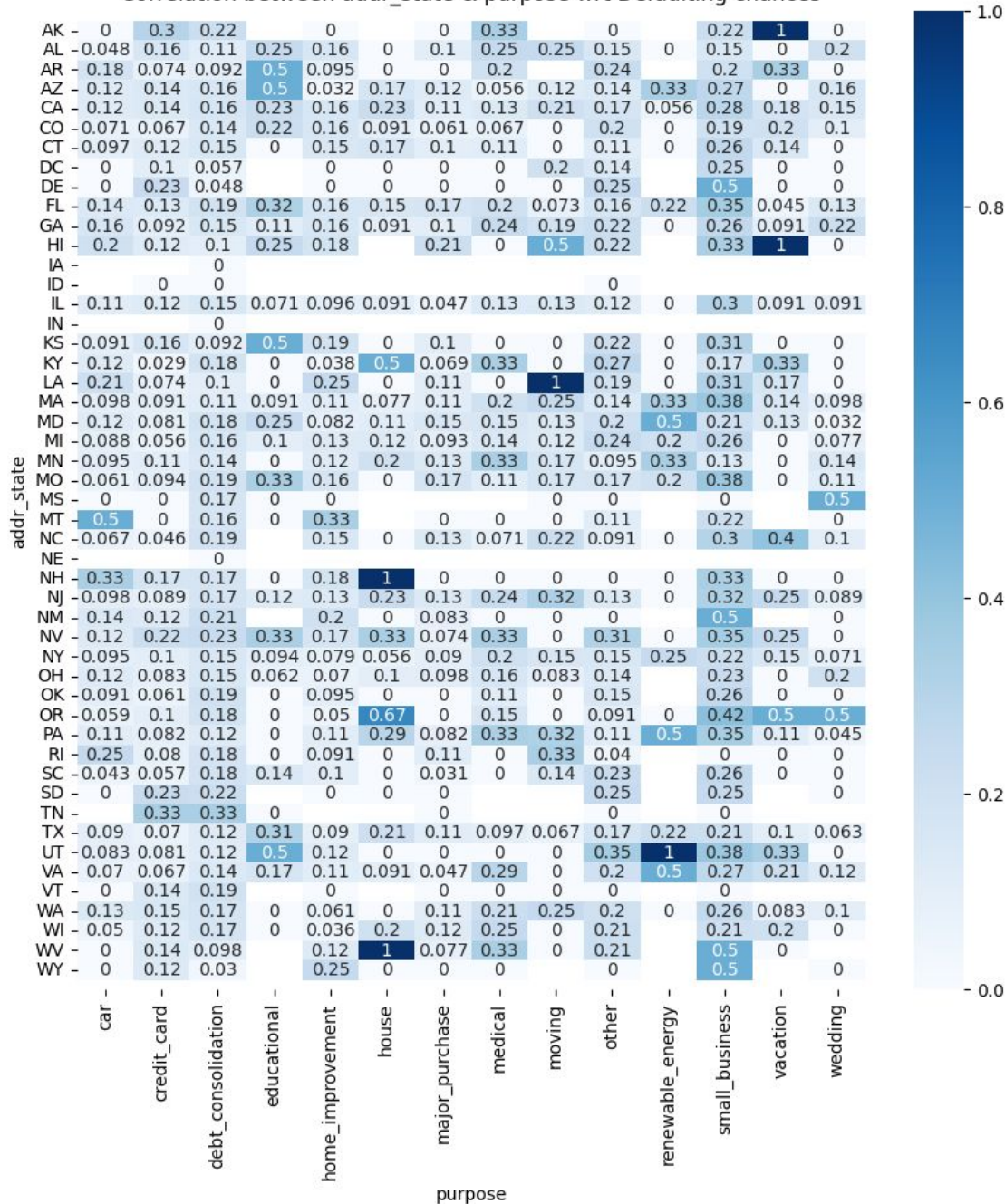
addr_state & purpose wrt Defaulting chances

Based on the above plot, the following observations can be made regarding risky loans:

1. Education loans in AR, KS, and UT
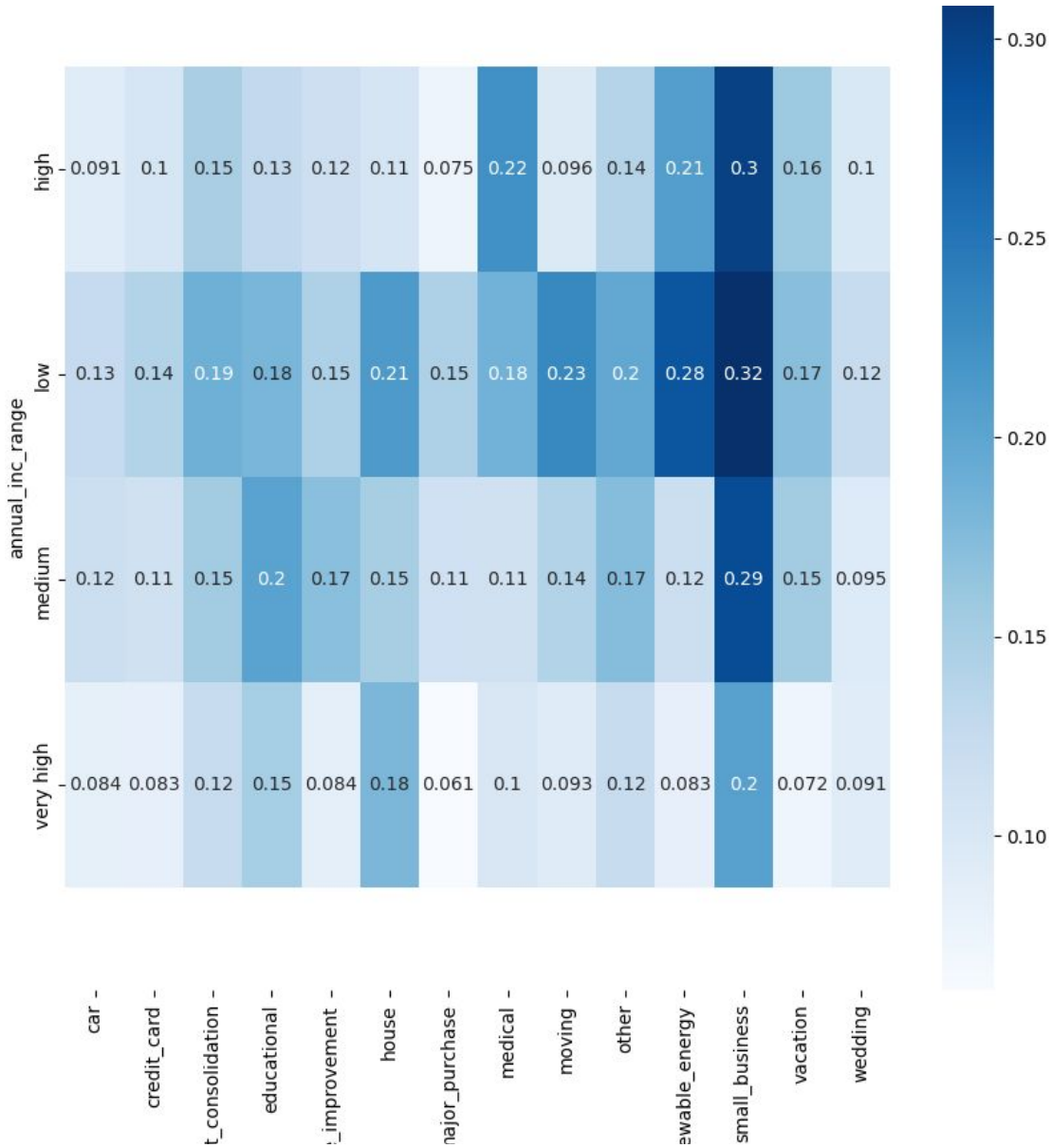
2. Small business loans in DE, NM, WV, and WY

3. Vacation loans in AK and HI

4. Housing loans in NH and WV

5. Car loans in MT

6. Moving loans in LA and HI

Correlation between home_ownership & loan_amnt_range wrt Defaulting chances
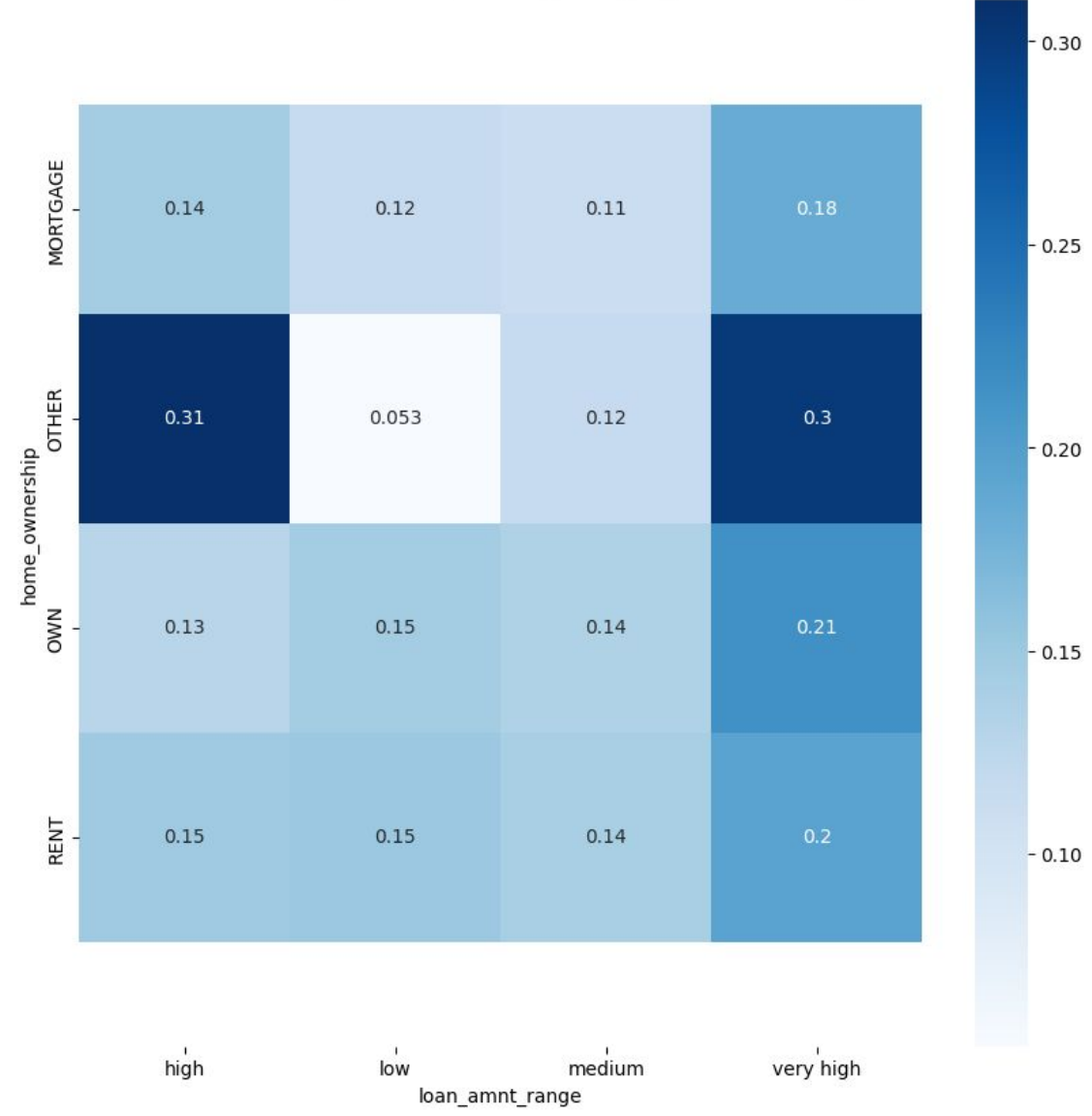
# Home_ownership & loan_amnt_range wrt Defaulting chances

People with High and Very High Income with Homeownership(other) tend to default on Loan

Annual income & purpose wrt Defaulting chances

People with both high and low income who are involved in small business tend to have a higher tendency to default on loans.

Annual income & addr_state wrt Defaulting chances

Correlation between annual income & addr_state wrt Defaulting chances
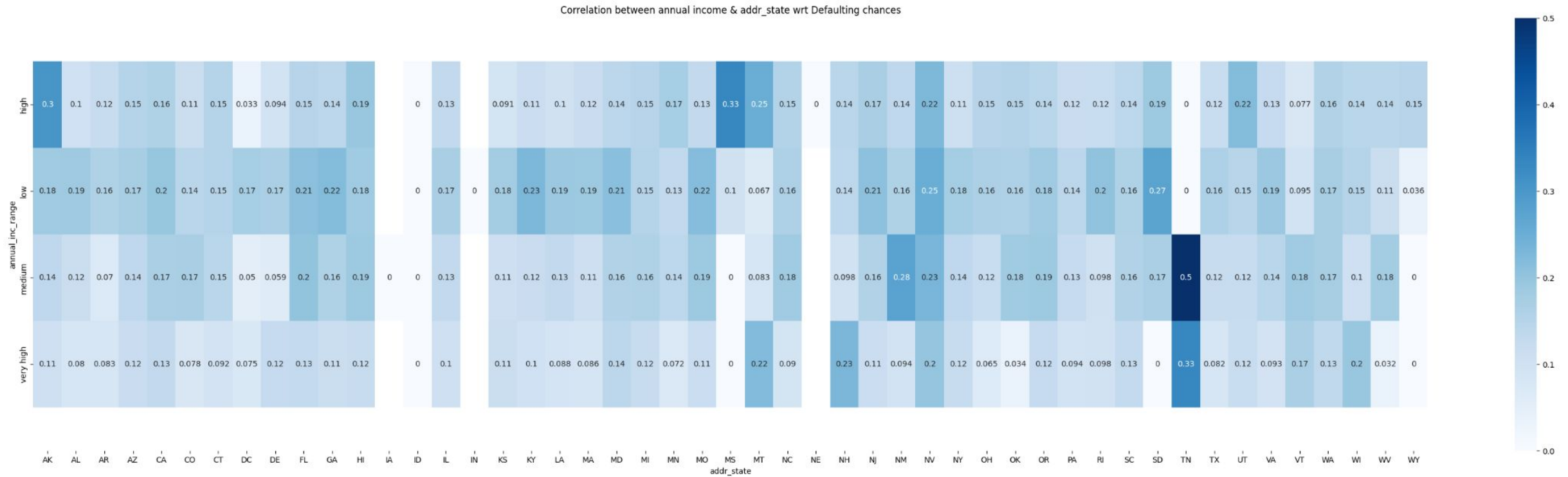
Based on the analysis, it was observed that people with high and low incomes in small business tend to default on loans. Additionally, individuals with high incomes in AK, MS, and MT have a medium chance of default, while those with medium incomes in TN and NM have a high chance of default. Lastly, very high-income earners in TN, NH, and MT have a medium chance of default."

Recommendations

1. Interest rate more than 14% have high probability to default

2. People who have taken loan for a tenure of 60 months have higher chances of defaulting

3. Higher the loan amount, greater the chance of the loan getting default.

4. Low Annual Income have slight chances to default

5. Higher installment have around 16 % chances of default(300 +)

6. Grade F & G along with its subgrade have highest chances of default

7. There is slight impact of other category in home_ownership to default

8. High chances of default are in NV,SD,TN,NM,MO

9. Small Business , renewable energy ,educational purpose have higher chances to default

10. Derogatory public records (1 & 2) have high default chances

Recommendations

11. Public bankruptcy records (1 & 2) have higher chances to default

12. Vacation loans in AK, HI

13. Housing loans in WV,NH

14. Small Business in DE, NM, WV, WY

15. Car Loan in MT

16. Moving Loan in LA , HI

17. People with High and Very High Income with Homeownership(other) tend to default on Loan

18. High Income People in AK,MS,MT have medium chances of default

19. Medium Income people in TN,NM have high chances of default

20. Very High Income people in TN,NH,MT have medium chances of default

Thankyou