Case Study
Exploratory Data Analysis

Lending Club Dataset

Dataset Column Selection for Analysis

Un Ordered Categorical (Borrower/Customer Attribute)
Purpose of Loan
Home Ownership Status
Loan Verification Status
Residing State of the Loan Applicant

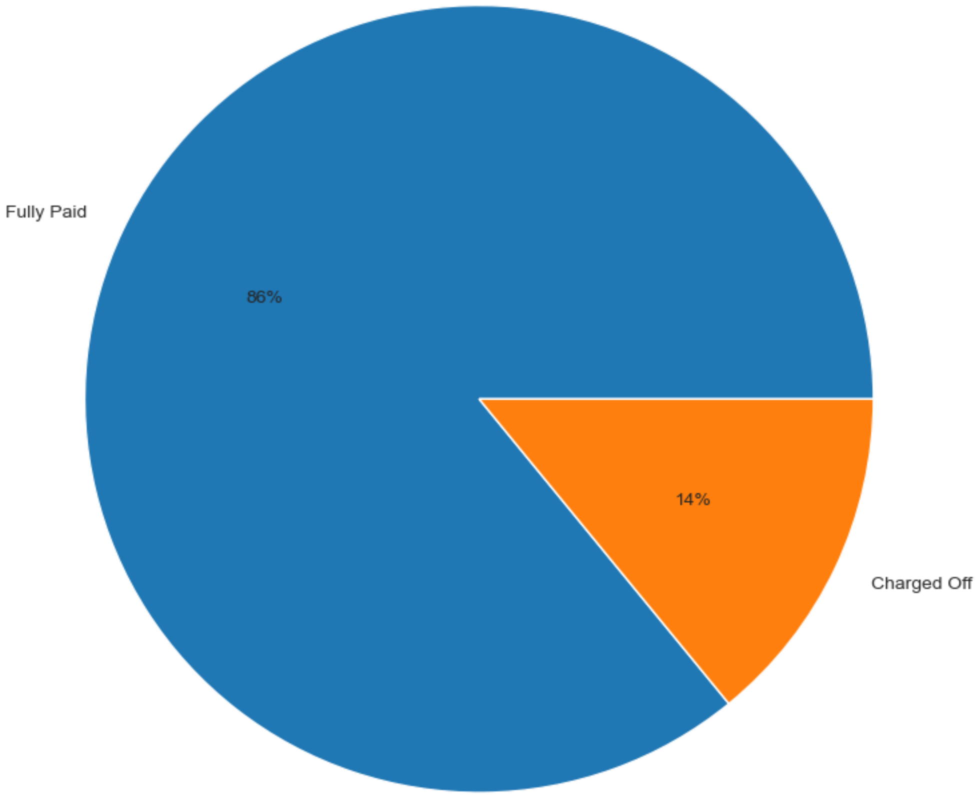
Ordered Categorical (Borrower/Customer Attribute)
Annual Income Range
Debt to Income Ratio Range
Employment Tenure of the Loan Applicant
Public Recorded Bankruptcies
Credit Utilization Range

Ordered Categorical (Loan/Product Attribute)
Loan Amount Range
Interest Rate Range
Total Repayment
Total Received Principal
Total Received Interest
Term/Tenure of Loan
Grade of Loan
Sub Grade of Loan
Loan Issue Year
Loan Issue Month

Derived Attributes

- **Principal Recovery Ratio** as the measure of severity of charge-off. The lesser the ratio indicates the more severe is the damage caused by the charge-off.
  - `prncp_recovery_ratio_range = total_rec_prncp / loan_amt`
- **Charge Off Percentage** as a measure of percentage of loans charged off across all the dibursals.
  - `charged_off_loan_percentage = round(loan_raw[loan_raw.loan_status == 'Charged Off'].shape[0] / loan_raw.shape[0]`
  - This percentage becomes the benchmark to analyse the performance of Charged-Off loans across different variables/dimensions.

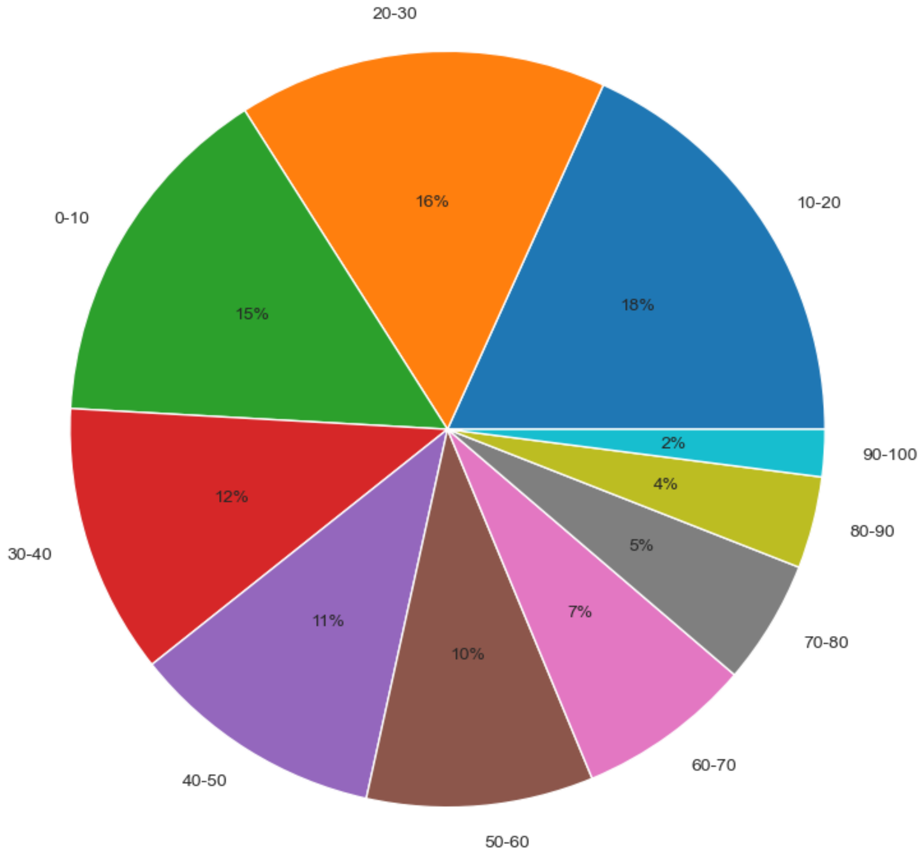
Proportion of Charged-Off loans



Distribution of Loan Status across Loans

- Analyzing ~34k records indicates the Charged-Off loans constitutes around ~14% of all the loans disbursed.

- Univariate Analysis of all the records to be performed taking this proportion as the benchmark.

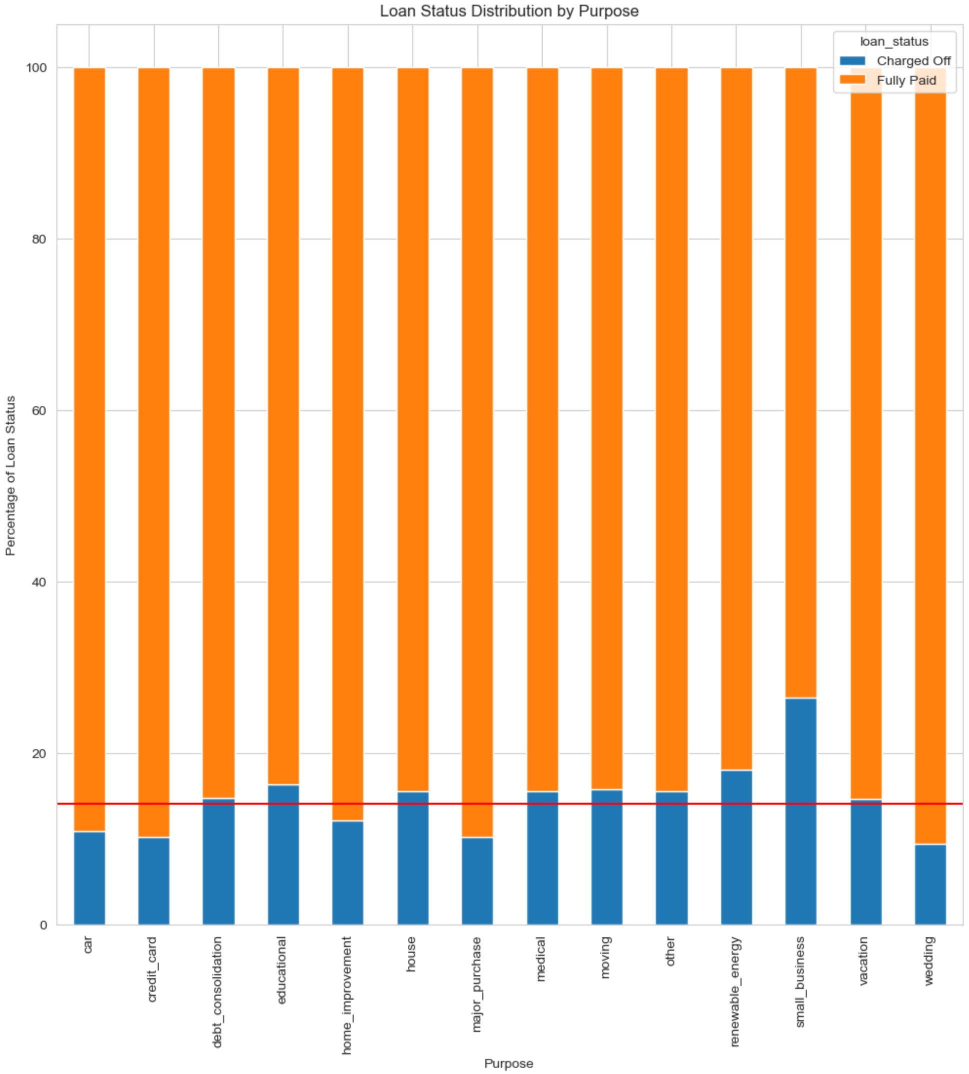


Distribution of Principal Recovery Ratio across Charged Off Loans

- Around **~70%** of the charged-off loans ends up paying **< 50%** of the principal

- Analyzing the customer and loan attributes that affects the payment of loan based on Principal Recovery Ratio

Univariate Analysis of Charged Off Loans

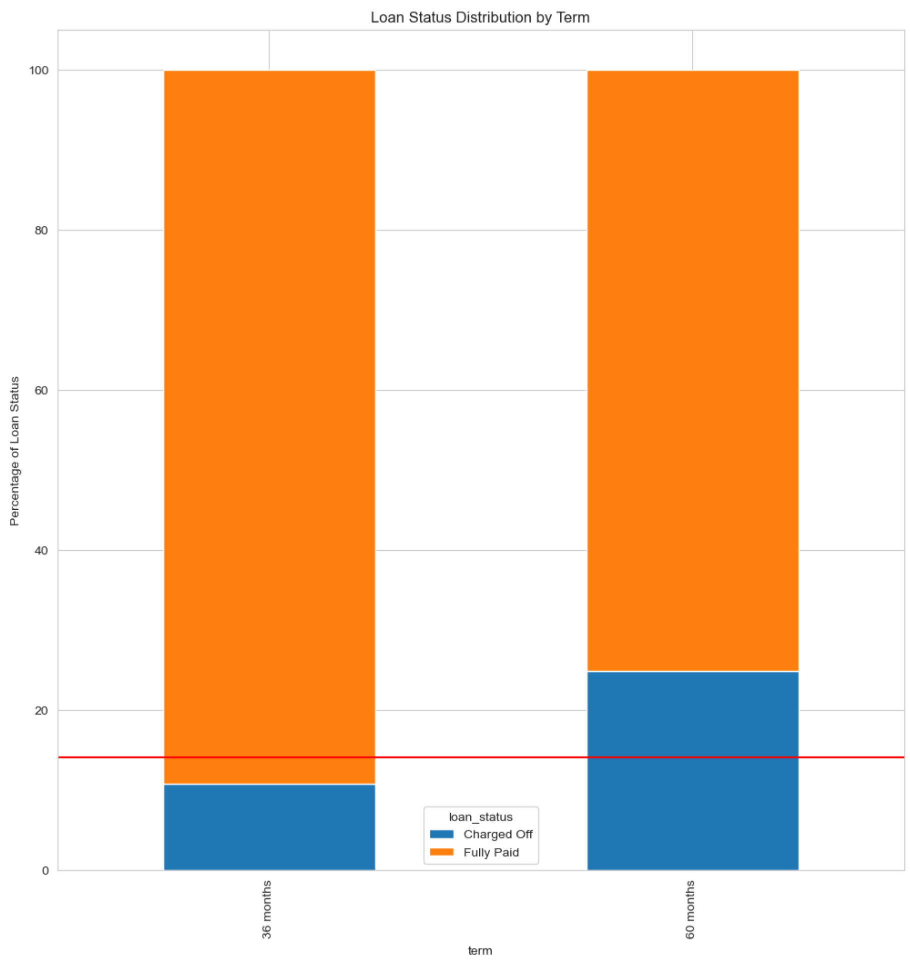


In the **small_business** and **renewable** catgory, the ratio of borrowers with "Charged Off" loan status exceeds the overall ratio of "Charged Off" borrowers.

- small_business charged-off loans ~**30%**
- renewable charged-off loans ~**20%**

Hence, the probability of defaulting is more when the Purpose is **small_business** or **renewable**.

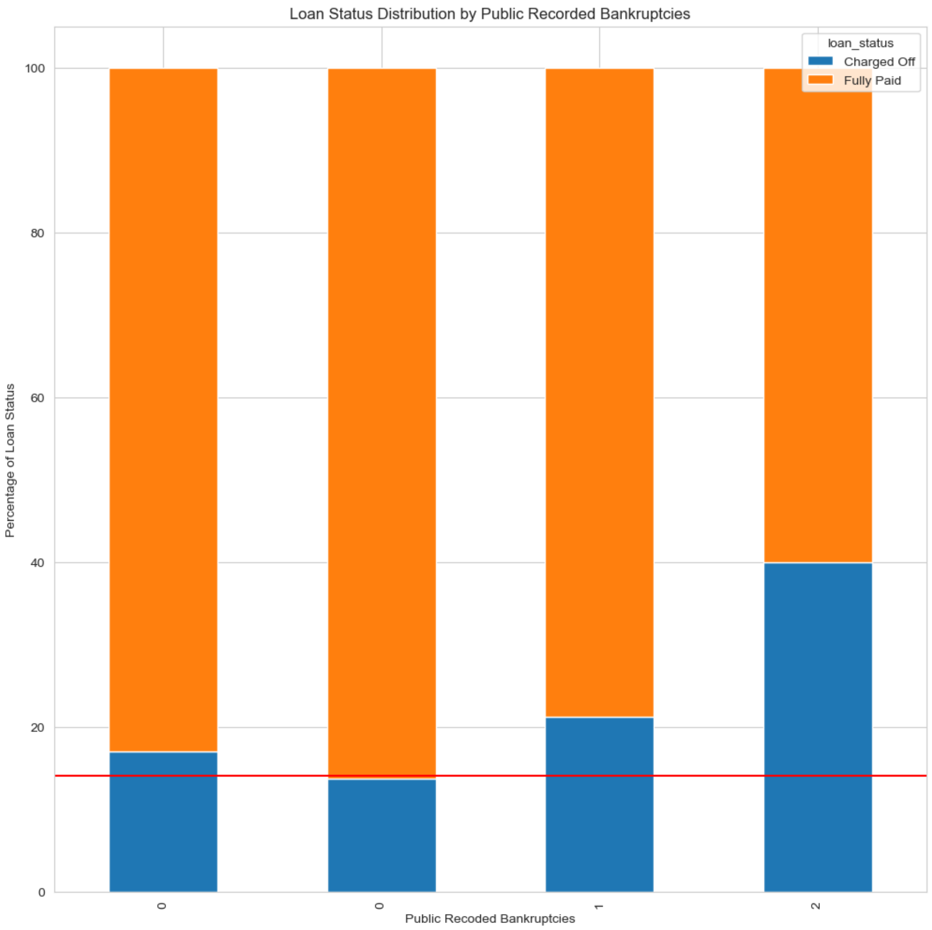
Univariate Analysis of Charged Off Loans



The number or charged-off loans are **~25%** for loans with tenure = **60 months** which is greater than the benchmark.

Hence, loans with **higher tenure** are more vulnerable.

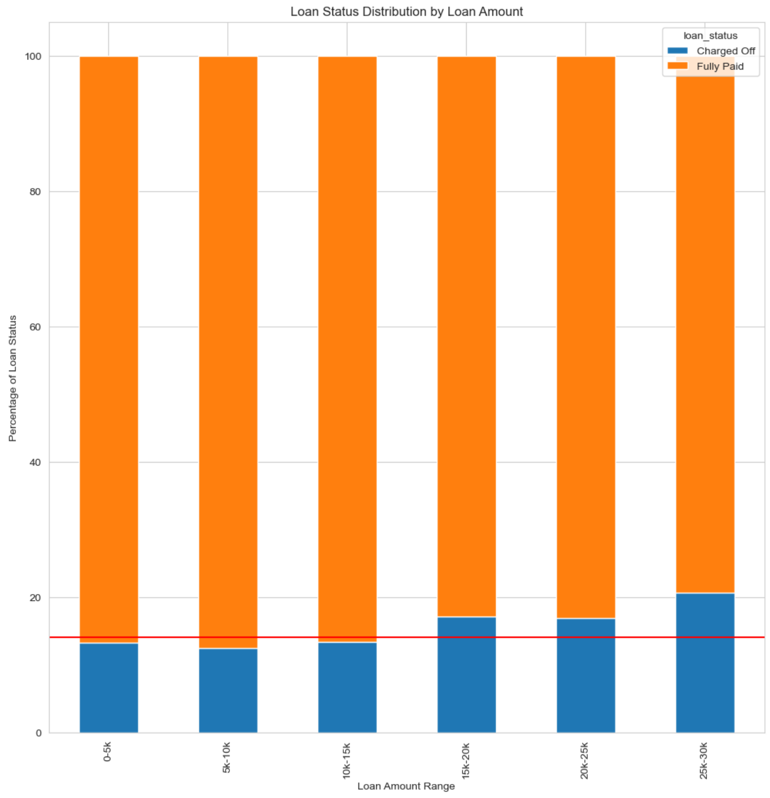
Univariate Analysis of Charged Off Loans

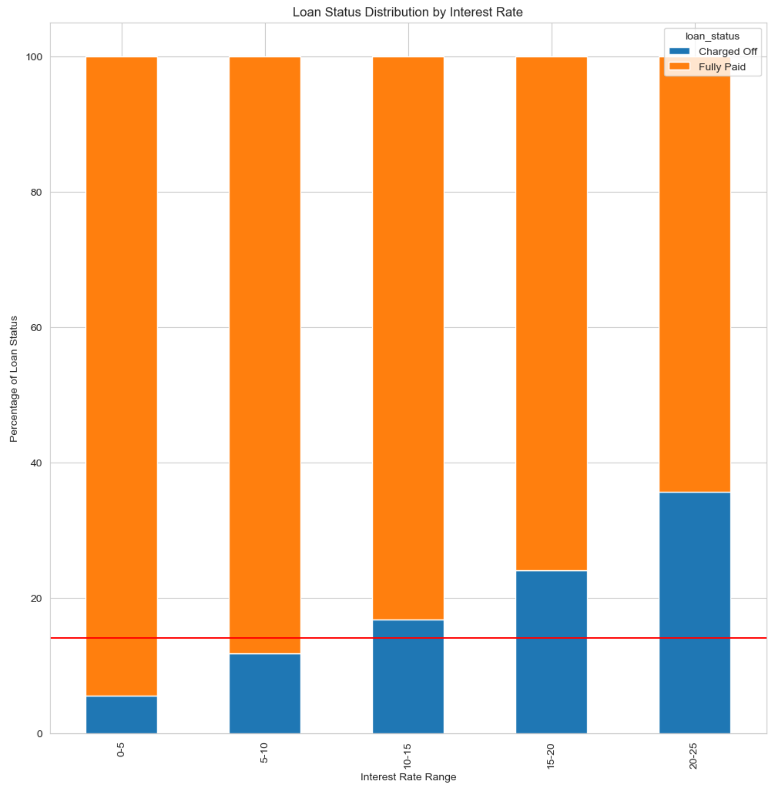


Analysis of loans where applicants have declared **>2** public recorded bankruptcies shows that **40%** of such loans ends up being Charged-Off.

A clear indicator that such disbursals should be avoided.

Univariate Analysis of Charged Off Loans



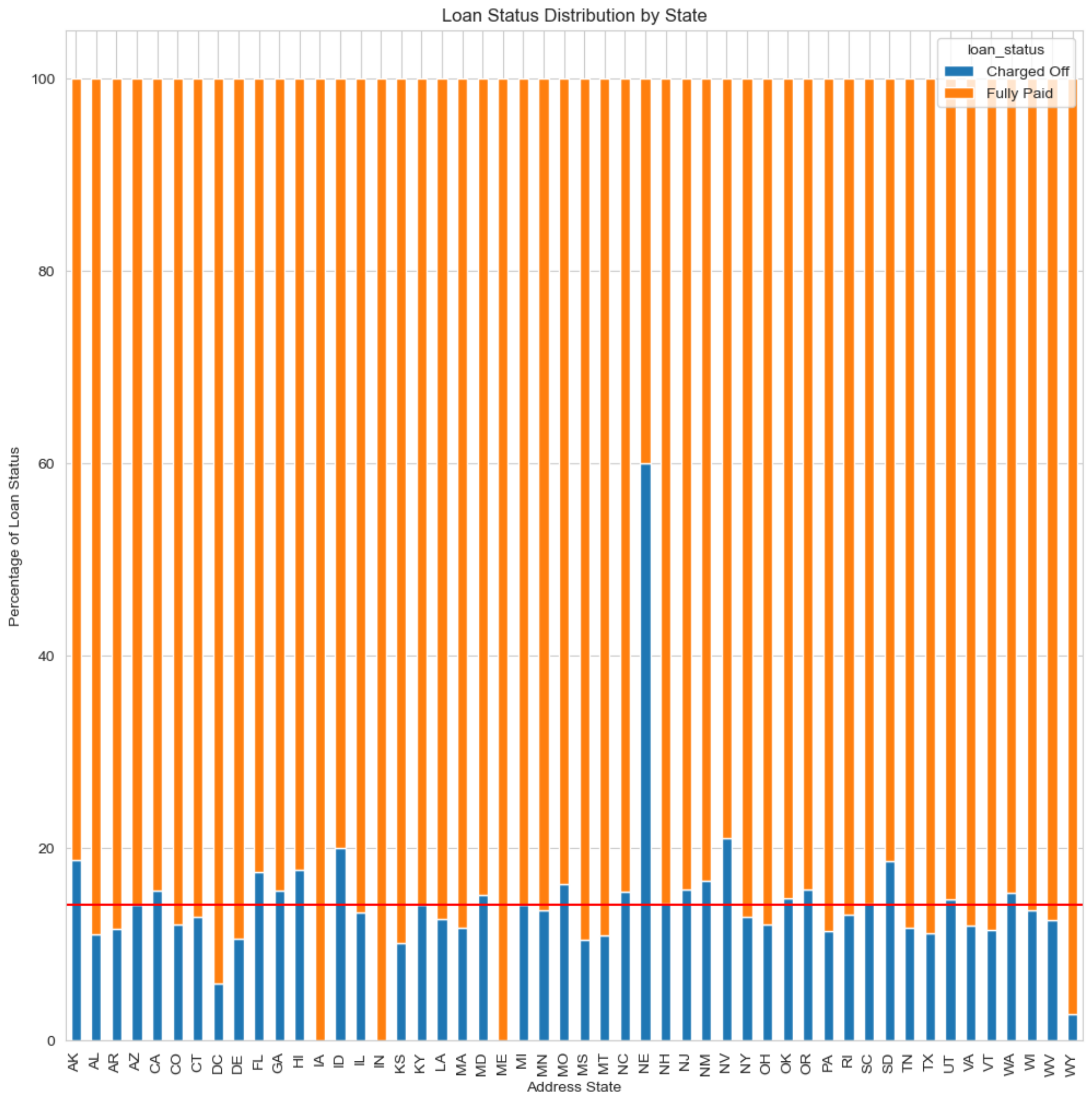


The percentage of "Charged-Off" loans increases with increasing Loan Amount and Interest Rate.

Interest Rate > 20% tends to have a high impact on the loans being defaulted with **~40%** loans being "Charged Off"

Hence, loans offering at higher interest rate should be avoided.

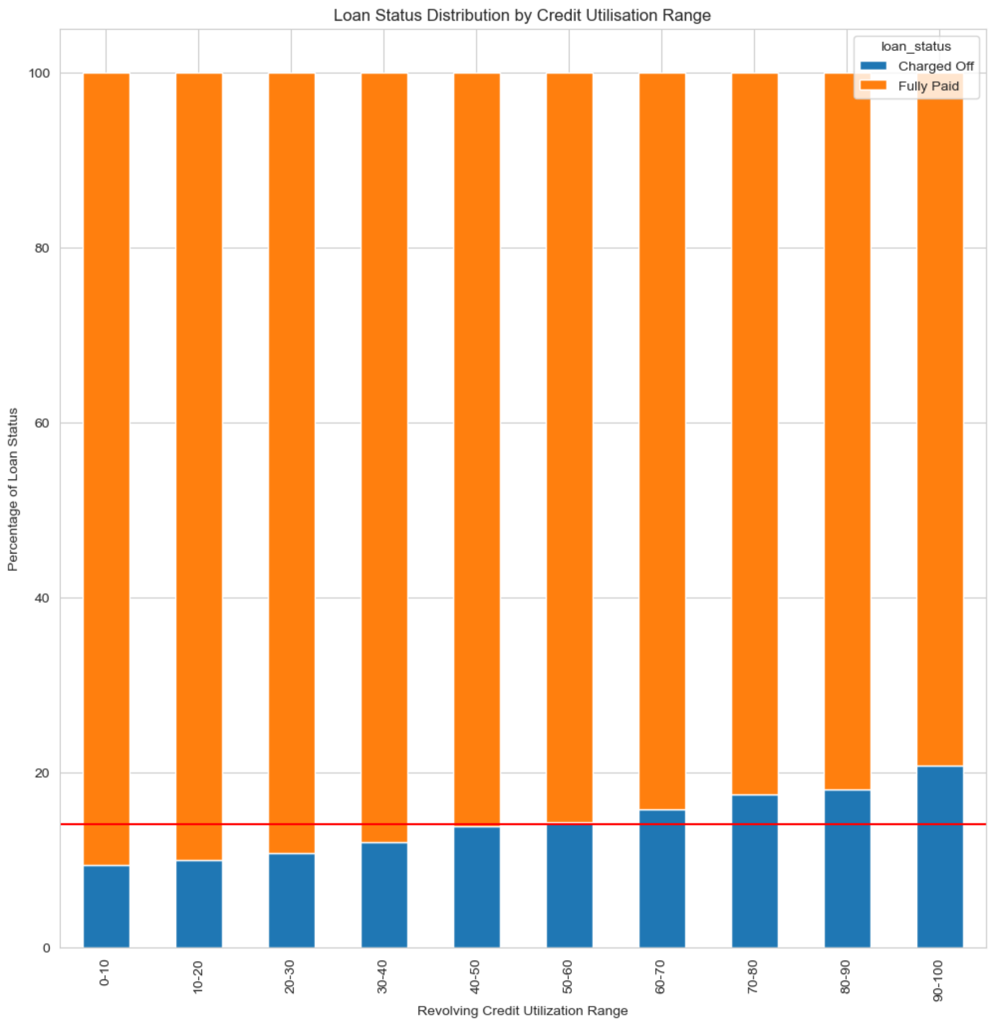
Univariate Analysis of Charged Off Loans



- **NV, SD, ID, AK** states have higher percentage of “Charged Off” loans.

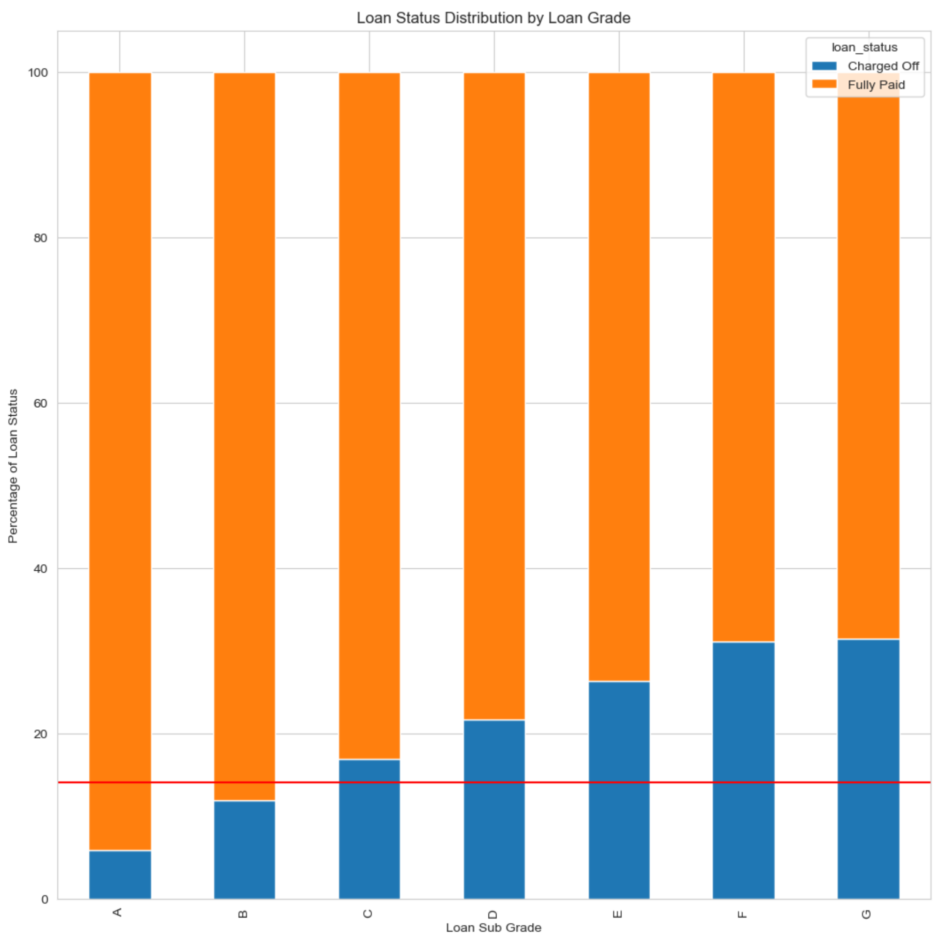
- **NE** state seems to have exceptionally high number **(~60%)** of loans being “Charged Off” which requires further analysis. This could also indicate a lapse in the application verification or recovery processes.

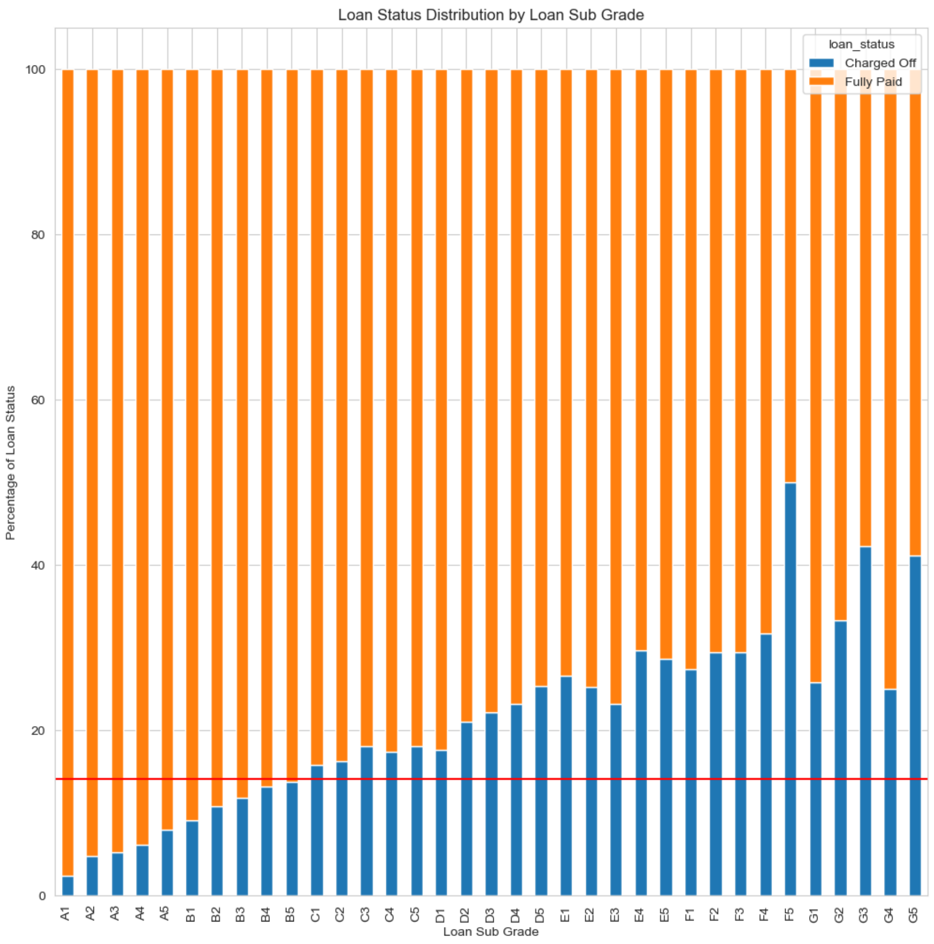
Univariate Analysis of Charged Off Loans



Loan applicants with Higher Credit Utilization Ratio tends to default more.

Univariate Analysis of Charged Off Loans



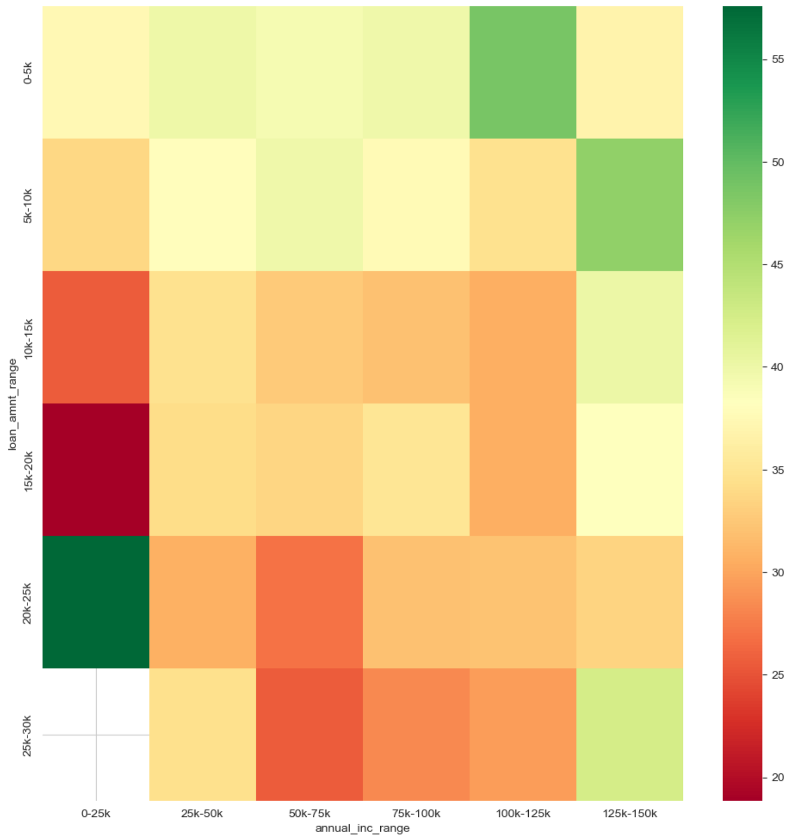


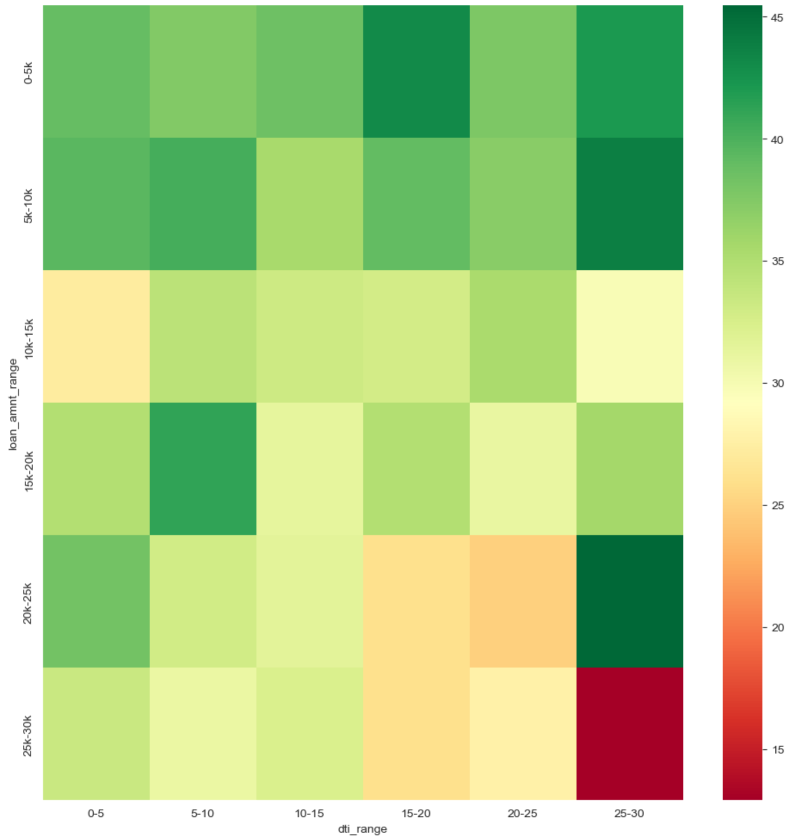
Higher Grade loans **E, F, G** tends to result into defaults

Similar trend is observed in the subgrades of loan grade

Hence, higher grade loans should be avoided.

Bivariate Analysis of Charged Off Loans

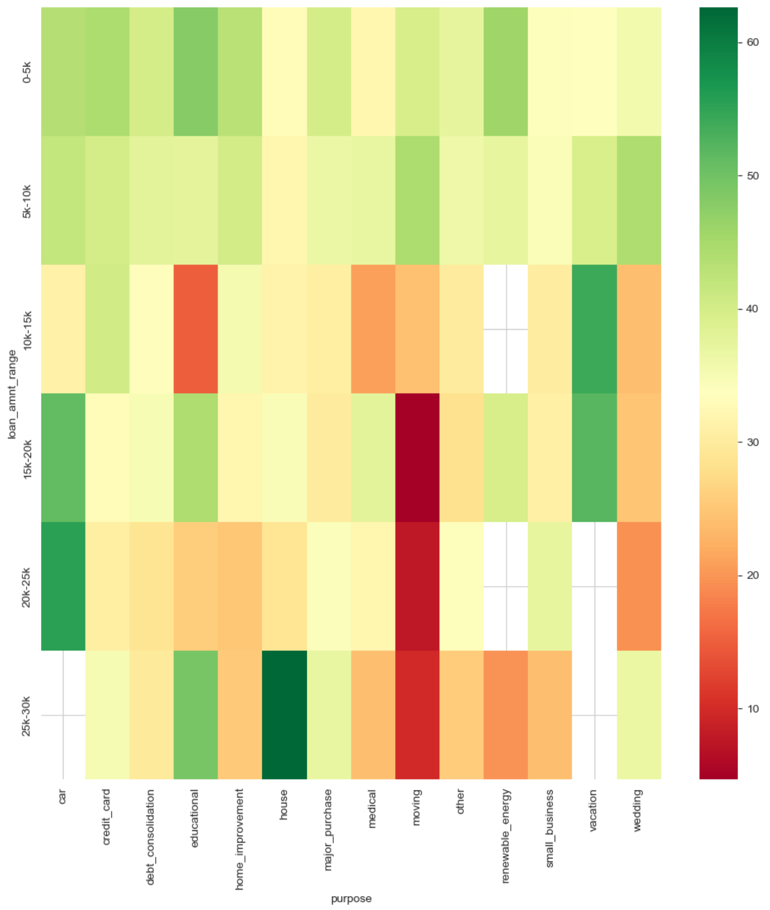


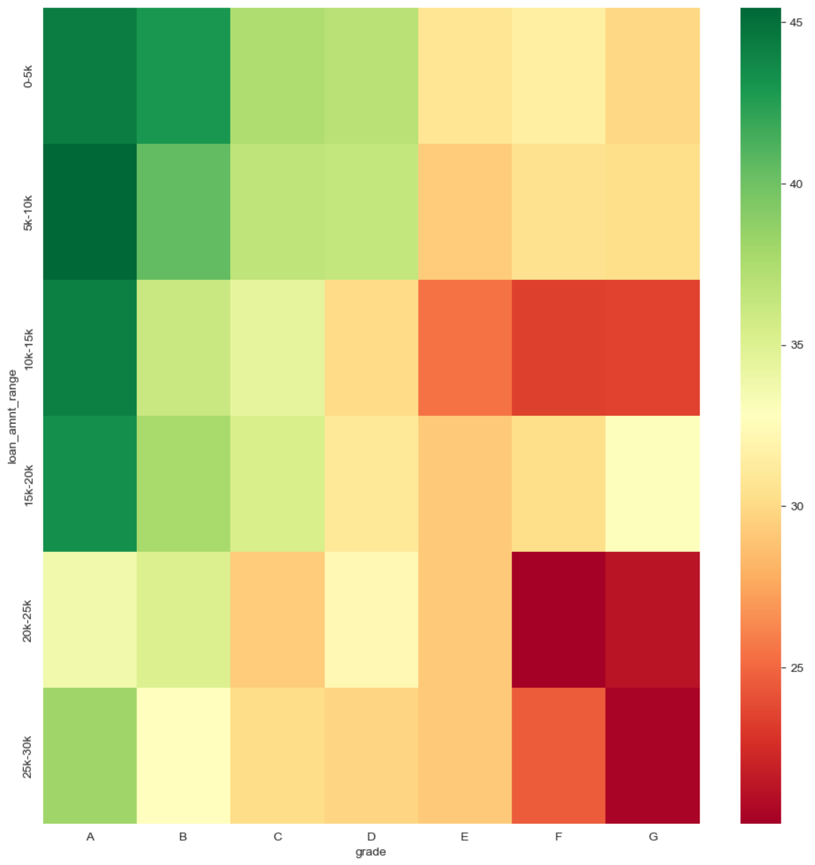


- In general, for all the income groups, the PRR shows a decrease with increasing loan amount. However, for the lowest income group **Annual Income < 25000**, the PRR decreases with the increase in Loan Amount, the least being for **Loan Amount ~ 15-20k** range.
- This indicates that loan amounts > 15k should be avoided for borrowers with annual income in the range 0-25k
- The PRR seems to be dropping below **15%** for loans with **Loan Amount ~ 25k - 30k** and **DTI ~ 25 – 30**
- This indicates that high DTI affects severely the repayment of higher loan amounts

- The exceptionally high PRR for **Loan Amount ~ 20-25k** seems to be an outlier.

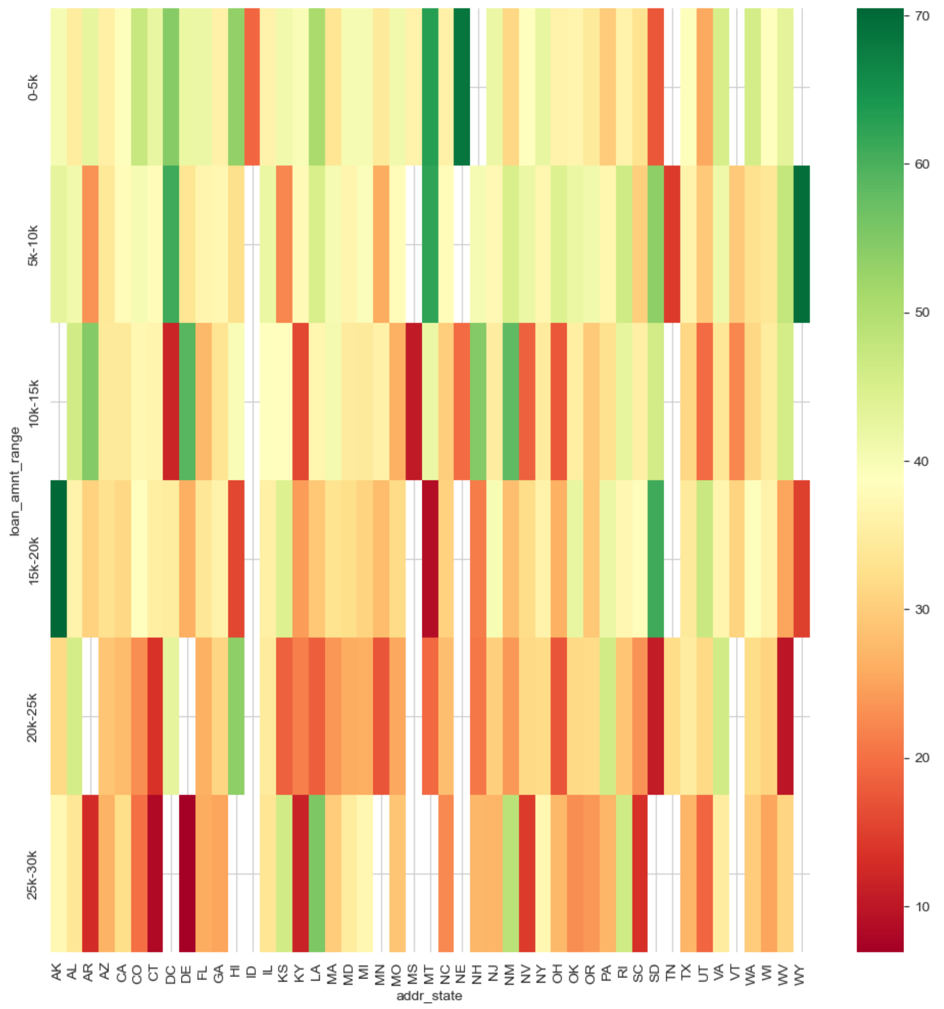
Bivariate Analysis of Charged Off Loans





- Amongst all the categories, **moving** category seems to show exceptionally low PRR's for **Loan Amount > 15k** and should be avoided or not supported.
- As also pointed out earlier Loan Grade **F** and **G** tends to have effect on the loan principal recovery. These grade loans should be discontinued.

Bivariate Analysis of Charged Off Loans



- General pattern is that the PRR decreases with the increase in the Loan Amount. However, few states are exceptional w.r.t the PRR and should be scrutinized. See the section below:

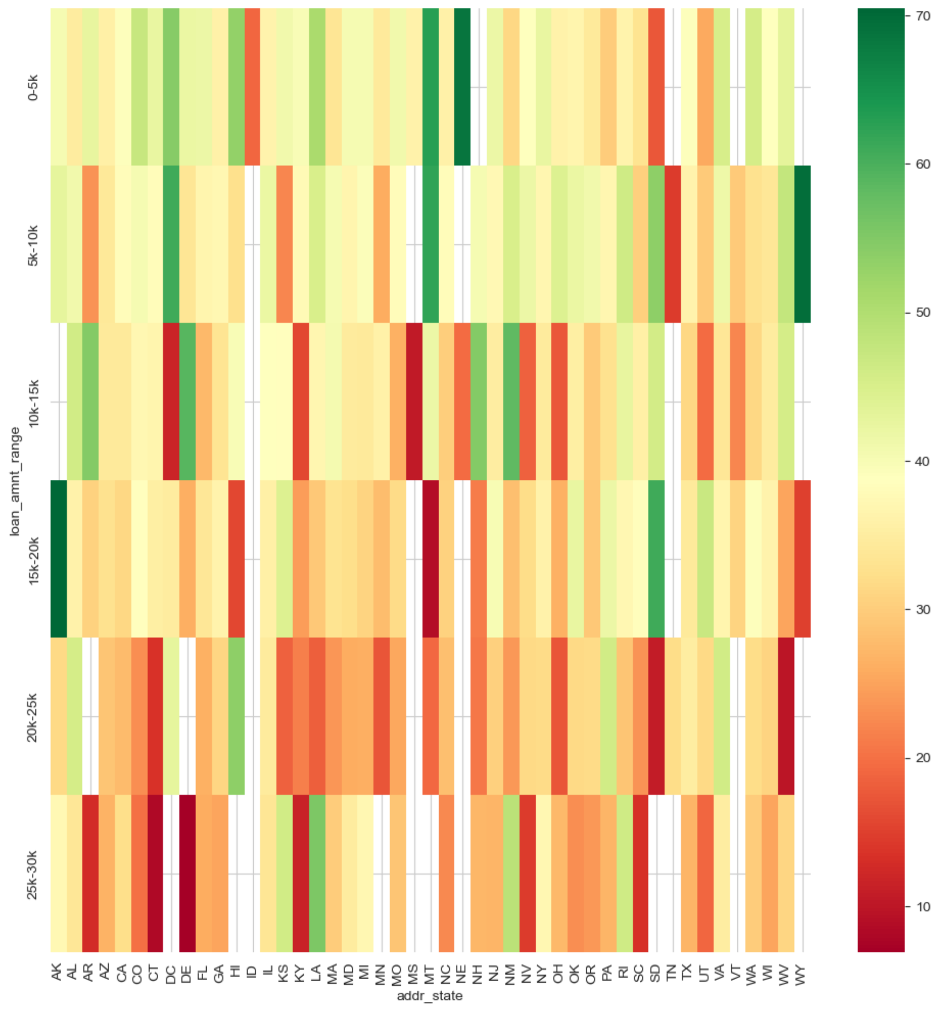
- For Loan Amount **25k-30k**:
  - states with least PRR: **CT, DE**

- For Loan Amount **20k-25k:**
  - states with least PRR: **WV**

- For Loan Amount **15k-20k:**
  - states with least PRR: **MT**

- For Loan Amount **10k-15k:**
  - states with least PRR: **MS, DC**

Bivariate Analysis of Charged Off Loans



- General pattern is that the PRR decreases with the increase in the Loan Amount. However, few states are exceptional w.r.t the PRR and should be scrutinized. See the section below:

- Annual Income **125k-150k:**
  - states with least PRR: **UT**

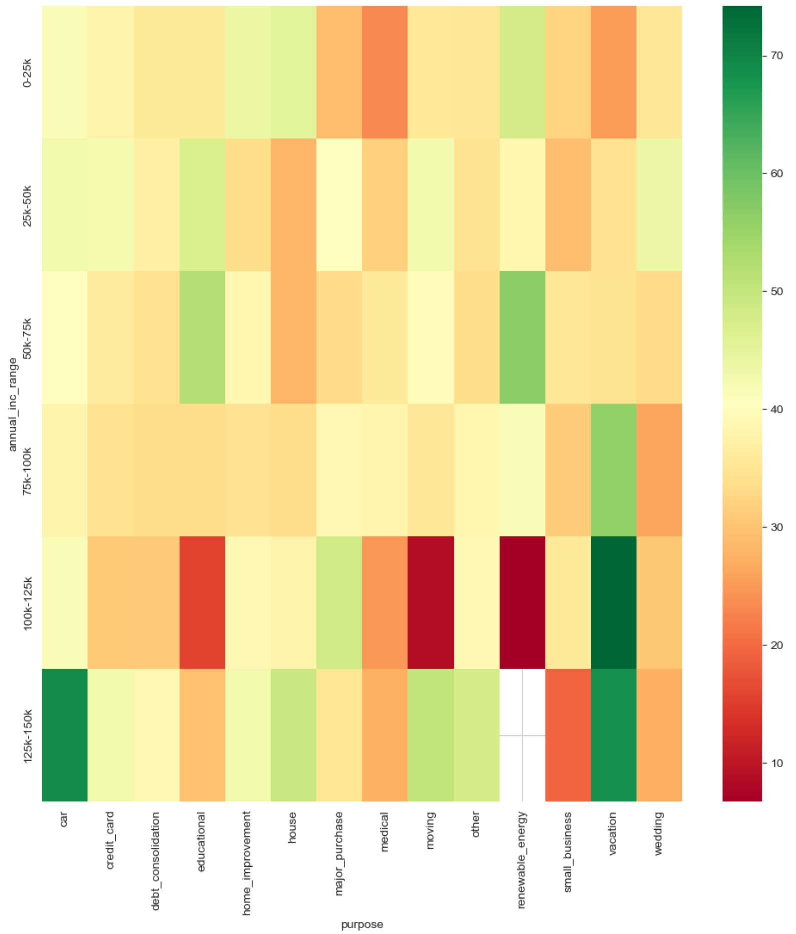
- Annual Income **100-125k:**
  - states with least PRR: **HI,NV**

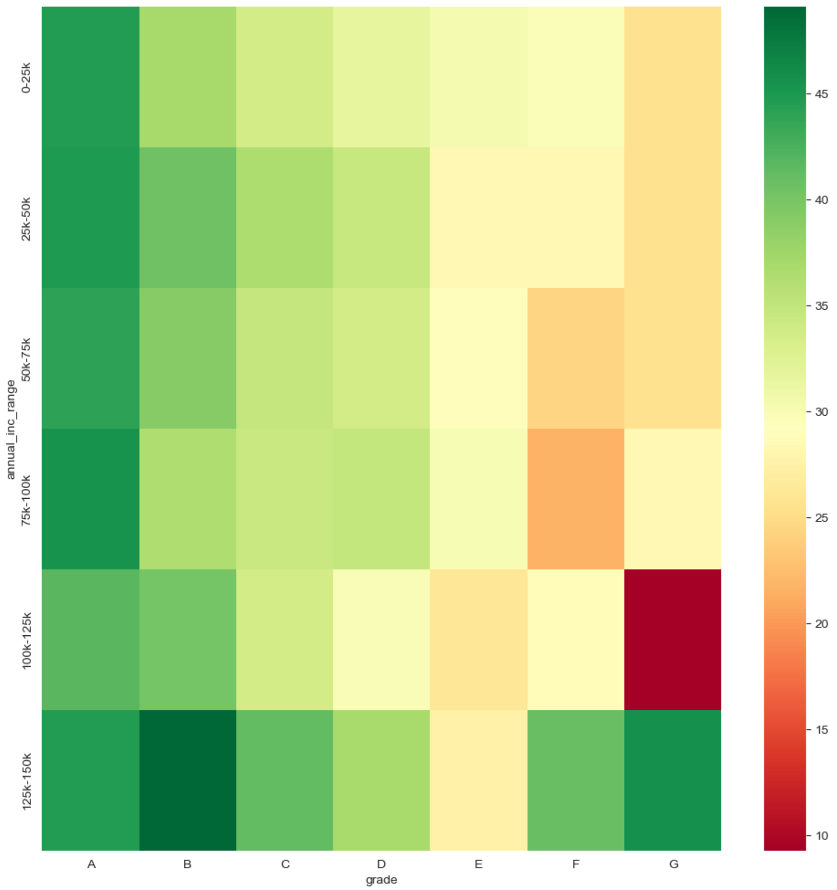
- Annual Income **50k-75k:**
  - states with least PRR: **DE, SD**

- Annual Income **25k-50k:**
  - states with least PRR: **MS**

- Annual Income **0k-25k:**
  - states with least PRR: **VT**

Bivariate Analysis of Charged Off Loans





- Below category loans shows exceptionally lower PRR and should be avoided:
  - moving
  - renewable_energy
  - small_business
  - wedding

- Loan Grades E, F, G tends to affects severely the repayment of loan irrespective of the annual income

- The exceptionally high PRR for **Loan Amount ~ 20-25k** seems to be an outlier.

Conclusion

Minor Impact

- Higher loan amount
  - (above 15k)
- Applicant's address state
  - (NV, SD, AK, ID, NE)
- Higher debt to income ratio
  - (above 15%)
- Credit utilization rate
  - (above 50%)

Major impact

- Loan purpose
  - (small business, renewable energy, educational)
- Loan Tenure
  - (5 year loans tends to cause more defaults)
- Public bankruptcy records
  - (1 or 2)
- Higher interest rate
  - (above 20%)
- Loan grade & sub-grade
  - (E, F, G)
- Higher DTI's affect the payment of Higher Loan Amounts