EDA Data Cleaning and identifying Variant for Analysis

Steps followed during Data Cleaning:

❖ Columns with all NAN or Same values are dropped

❖ Most of the data are Categorical variables, therefore only the rows corresponding to defaulter "Charged Off" in "loan_status" column is considered

❖ From the derived sub dataset, column show in the below table are considered for analysis

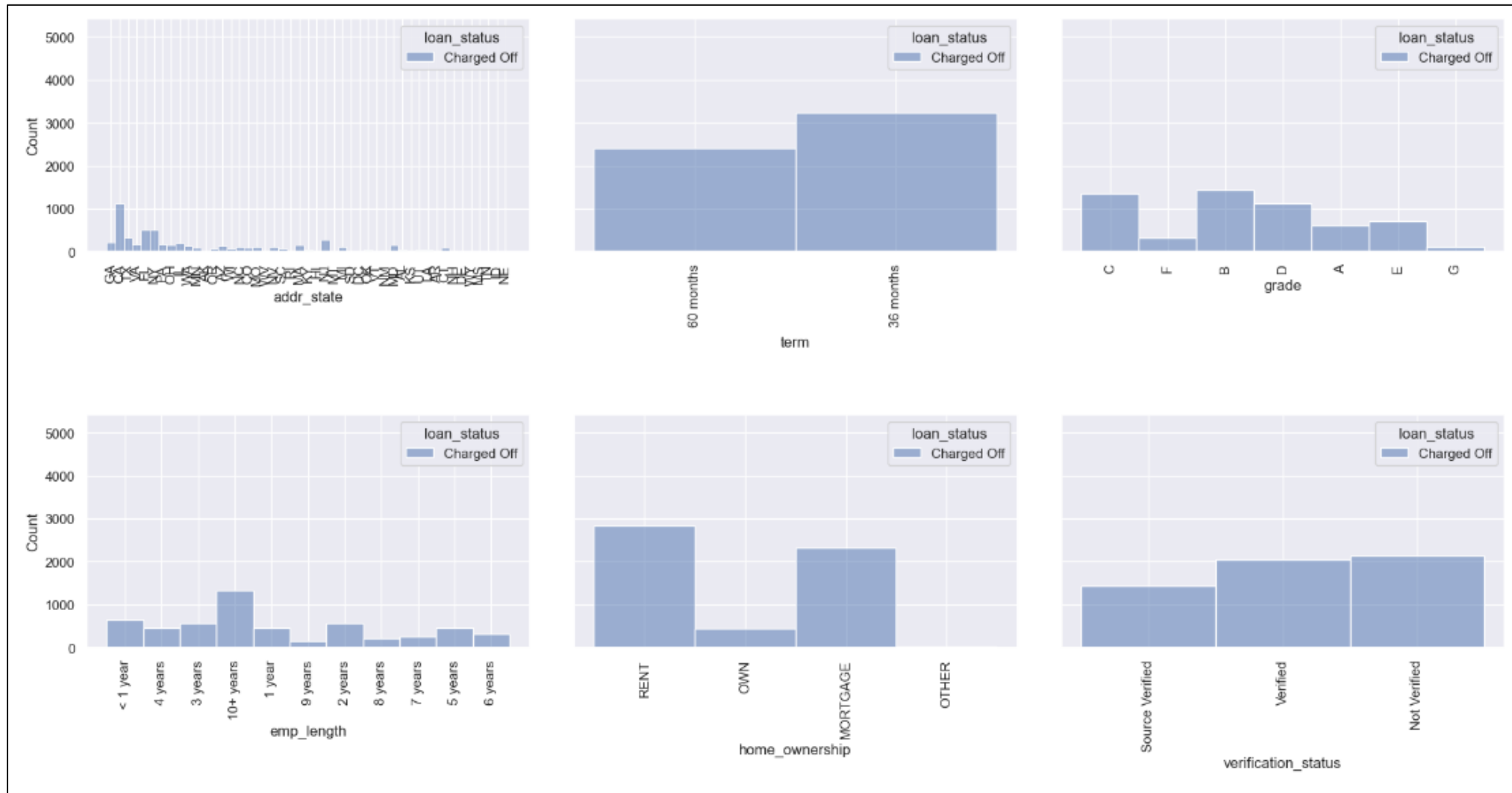Note: unique count values less or equal to 14 is considered for selecting the variables.

column	unique count
term	2
grade	7
emp_length	11
home_ownership	5
verification_status	3
loan_status	2
purpose	14
delinq_2yrs	11
inq_last_6mths	9
pub_rec	5
pub_rec_bankruptcies	3

Univariant Analysis

Analysis Result:
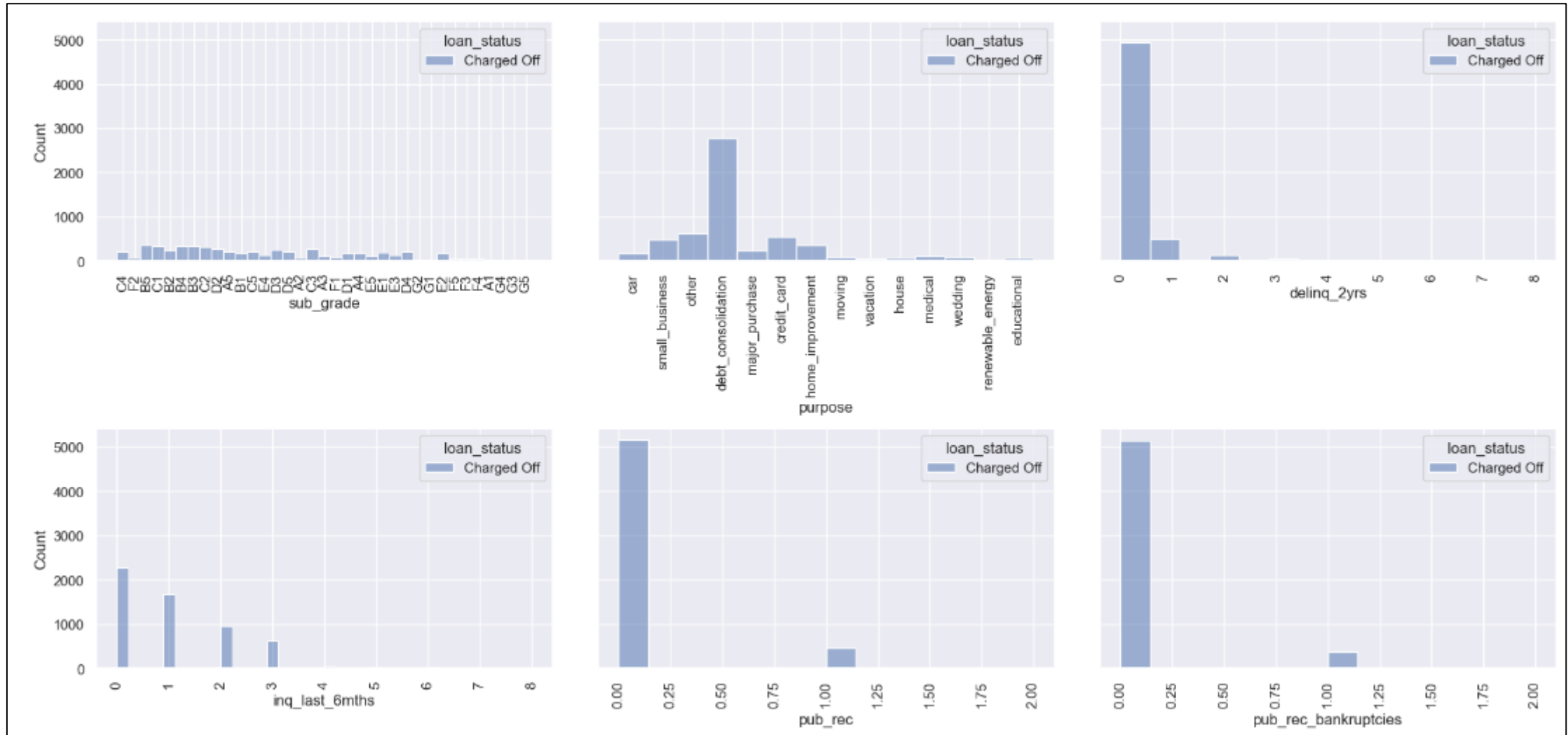- ❖ Resulting plot is shown below:
- ❖ **driver variables are:** "addr_state", "term", "grade", "emp_length", "home_ownership", "verification_status"
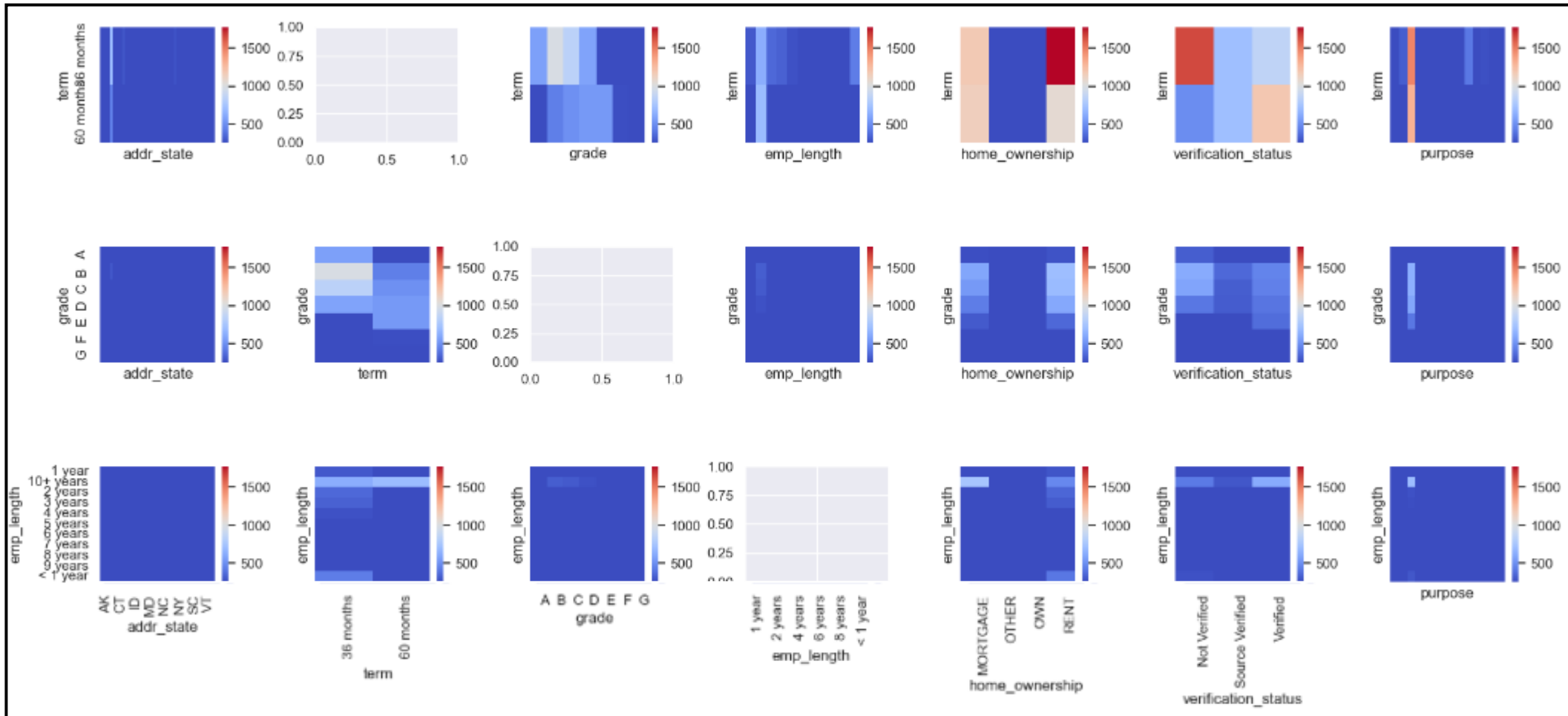
Univariant Analysis – Cont's

Analysis Result:
  ❖ driver variables from :
  ❖ **driver variables are:** ""purpose""

Bivariate Analysis

➢ Driver variables from the Univariate analysis are consider for the analysis,

❖ "addr_state", "term", "grade", "emp_length", "home_ownership", "verification_status", "purpose"

➢ Analysis performed using Crosstab and heatmap, Result shown in below figure:

❖ Resulting plot is shown below:

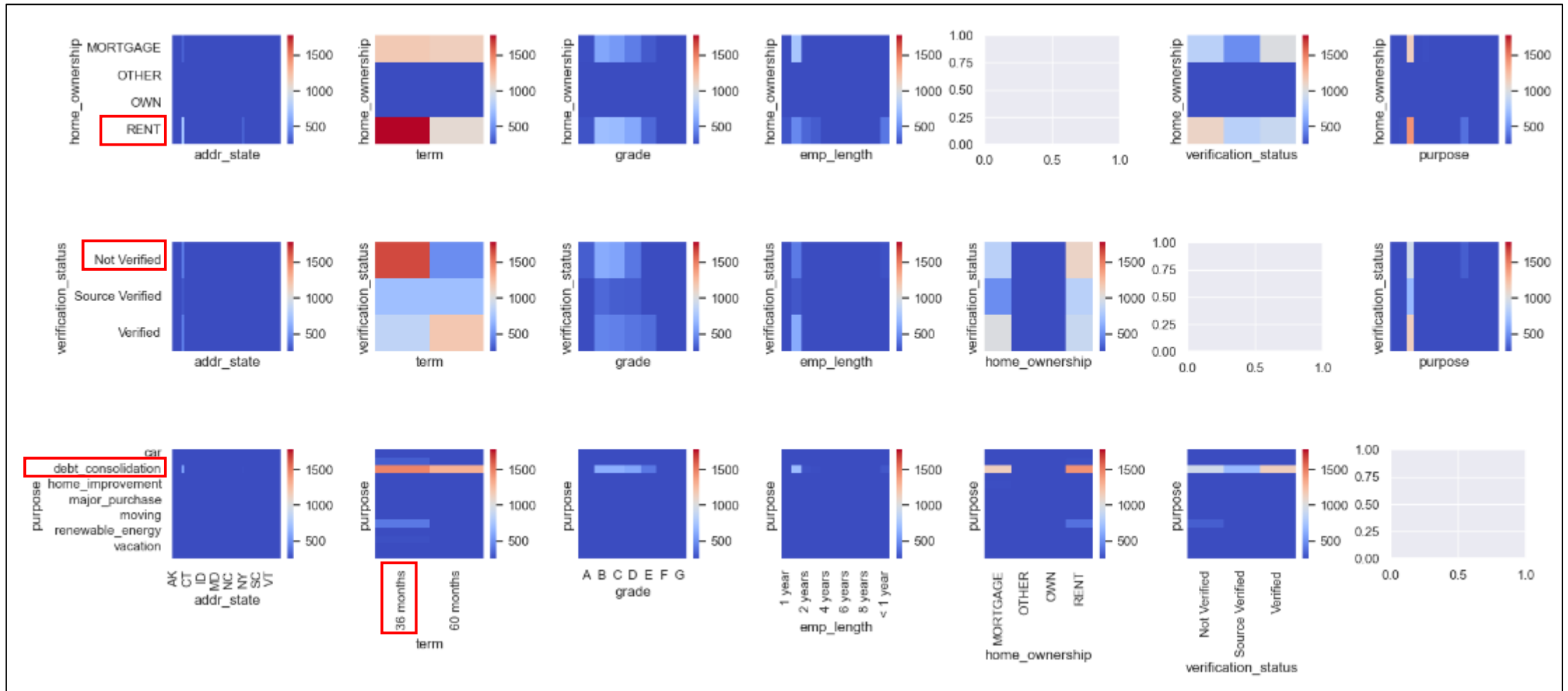❖ driver variables are: **"term", "home_ownership", "verification_status"**

Bivariate Analysis – Cont's

➢ Analysis Result – Cont's:

  ❖ Resulting plot is shown below:

  ❖ Driver variables are: **"purpose", "term", "home_ownership", "verification_status"**

Note: Driving values for each Driver variables are highlighted with Red box.

Conclusion

➢ From the variables considered for the analysis, the most driving factors (or driver variables) behind loan defaulter are:

    ❖ Resulting plot is shown below:

    ❖ driver variables are: "purpose", "term", "home_ownership", "verification_status"

    ❖ Below Tables shows the critical values for the driver variables.

Driver Variables	Critical Values
Purpose	debt_consolidation
Home_ownership	RENT
Verification_status	Not Verified
Term	36 months

    ❖ The most Critical case is the combination of above values, in these cases Loan should be rejected.