CASE STUDY

SUBMISSION

Business objective and problem statement

- Background:
  - The company under consideration is a **consumer finance company** operating in **online loan marketplace**
  - It facilitating personal loans, business loans, and financing of medical procedures
  - There are risks associated with the company's decision on loan approval:
    - Not approving the loan for an applicant, who is likely to repay the loan
    - Approving loan for an applicant, who is not likely to repay the loan

- Problem statement:
  - This **risk of financial loss (credit loss)** is largely due to borrowers who default ("**charged-off**")

Business Objective:

- To understand the driving factors which are **strong indicators of loan default** & minimize credit loss, which can be utilized for portfolio and risk assessment

- Goal of Data analysis:
  - To analyze the complete loan data for all loan,clean the data for any quality issues, create new variables, conduct exploratory data analysis to identify variables which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Step-by-step approach:

1. Data understanding
2. Data cleaning
3. Categorisation of Data
4. Data Visualisation
5. Analysing the Trends
6. Summary

Data understanding and key assumptions

- **Original data set** contains information about **39,717 applicants** who have been issued loans by the company

- Various **consumer attributes** and **loan attributes** are spread across 111 columns

- Loans are broadly classified into the 3 categories based on their status: **Fully paid**, **Current**, **Charged-off**
  - **83 columns dropped** for analysis:
    - **54** columns with 100% null values dropped
    - **5** columns with > 5% null values dropped and not relevant for analysis basis understanding from metadata
    - 24 columns that are of no use in our analysis (Ex:url,address,etc) are removed
  - **Changed the data type** to suit our analysis:
    - date related columns converted into datetime format,
    - Similarly, "%" suffix removed from 'int_rate', 'dti', etc.

- **Revised data set has 38577 rows and 28 columns**

ANALYSIS OF LOAN STATUS

Loan Status	Number of Loans
Current	1140
Charge Off	5627
Fully Paid	32950

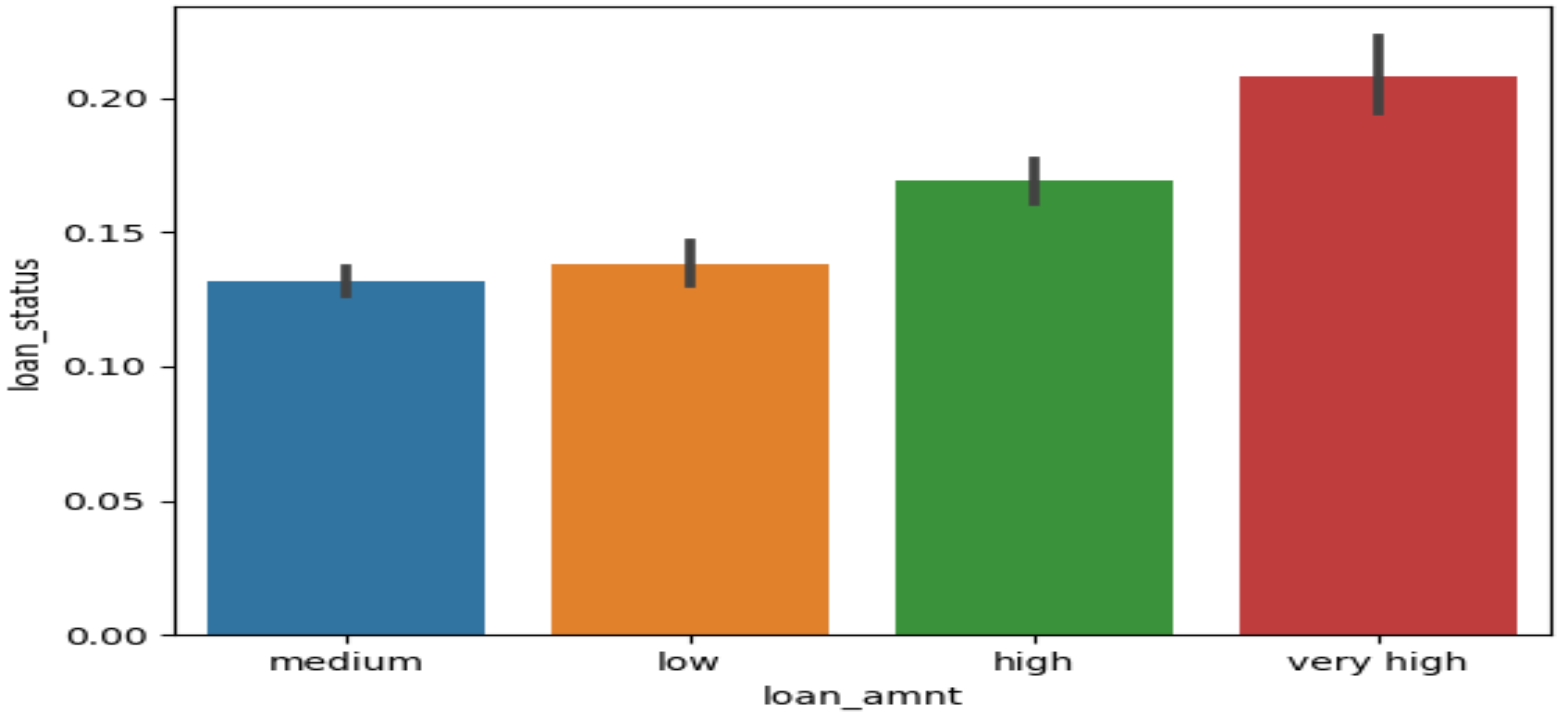
We cannot analyis Current status data so first we remove them and convert other two in 0/1 form

```
x=x[x['loan_status']!='Current']
x['loan_status'] = x['loan_status'].apply(lambda i: 0 if i=='Fully Paid' else 1)
x['loan_status']=x['loan_status'].astype('int')
x['loan_status'].value_counts()
```

Top 6 Deciding Parameters for Loan Defaulter

1. Loan Amount
2. Grade
3. Purpose of Loan
4. State
5. Home Ownership
6. Verification Status

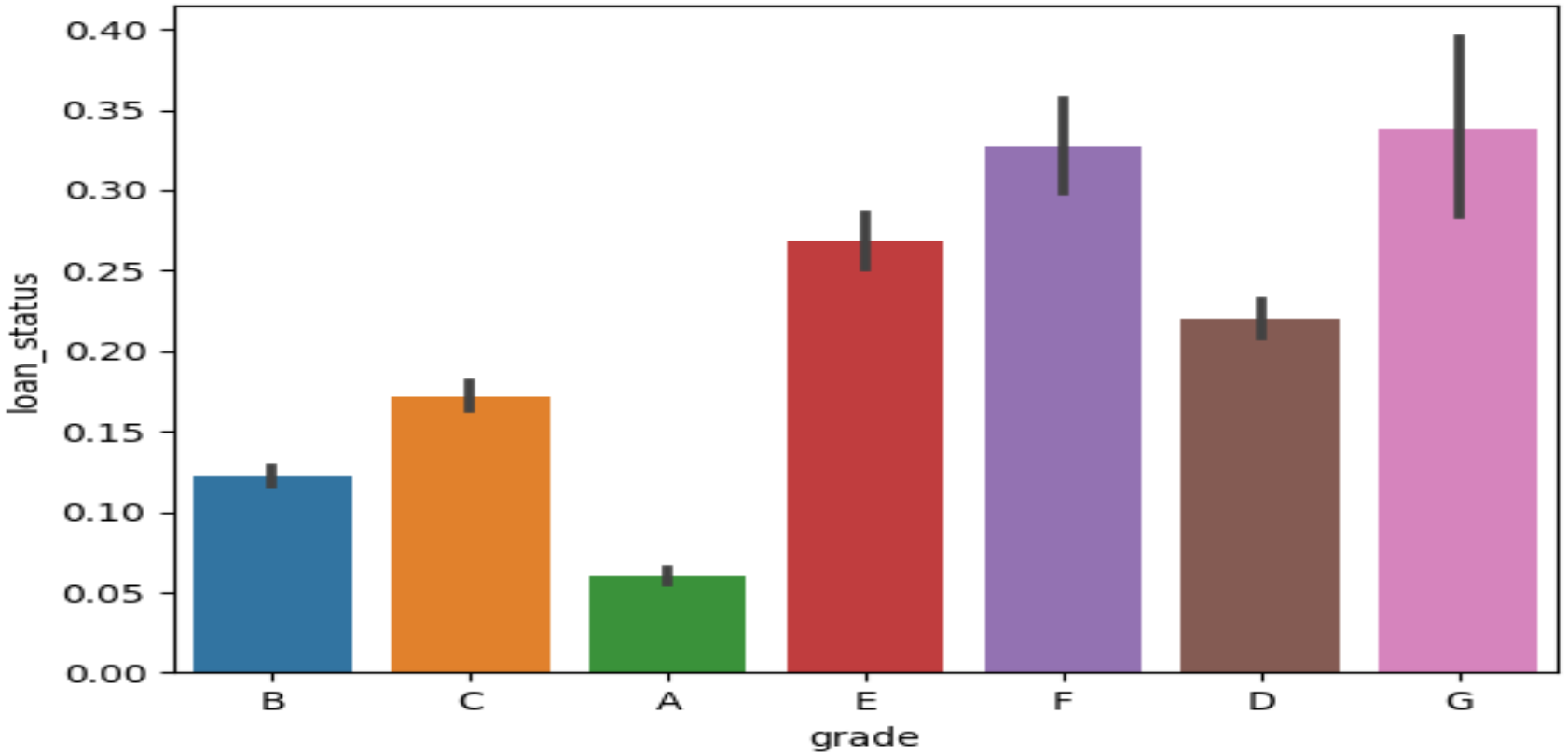
Analysis on Loan Amount



Trend that Loan Amount vs loan_status shows :-

Most defaulting borrowers have very high loan amount

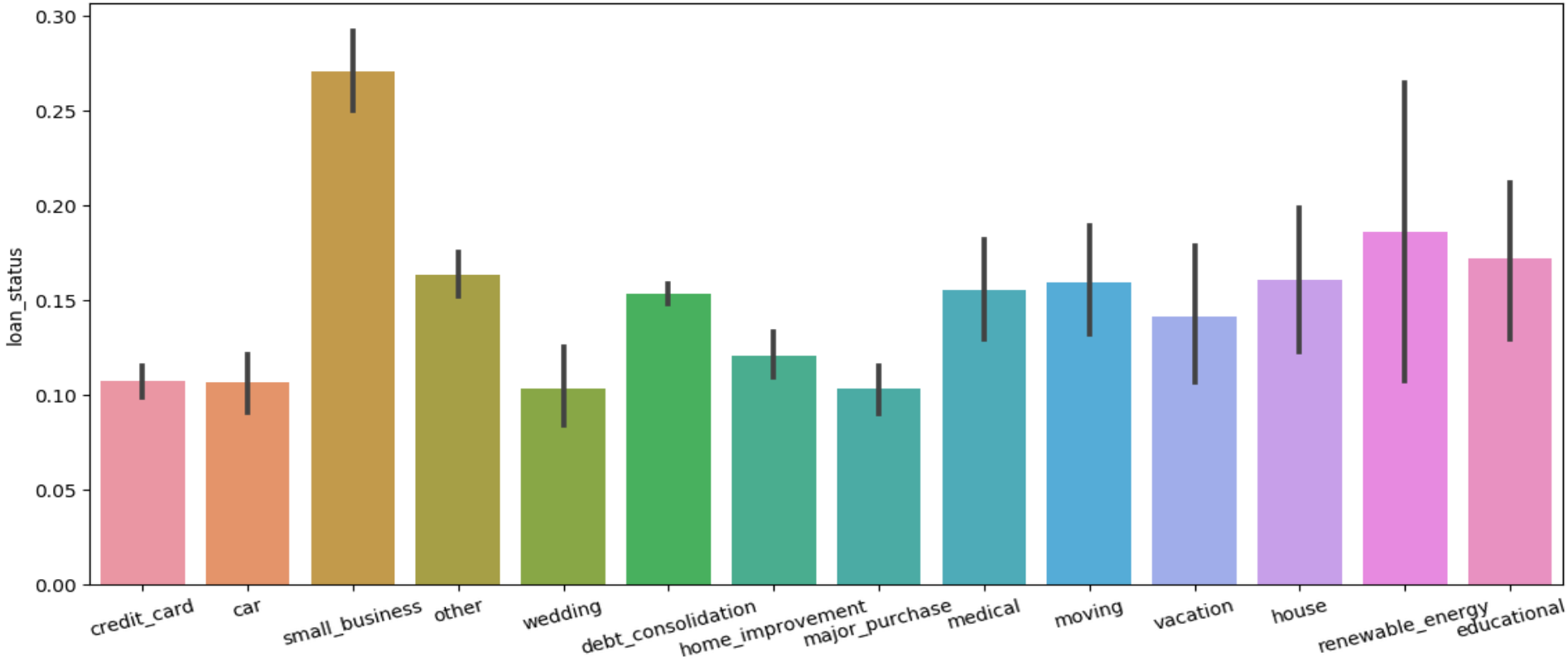
Analysis on Grade



Trend that Grade vs loan_status shows :-

Loans with Grade greater than C are most likely to be defaulted

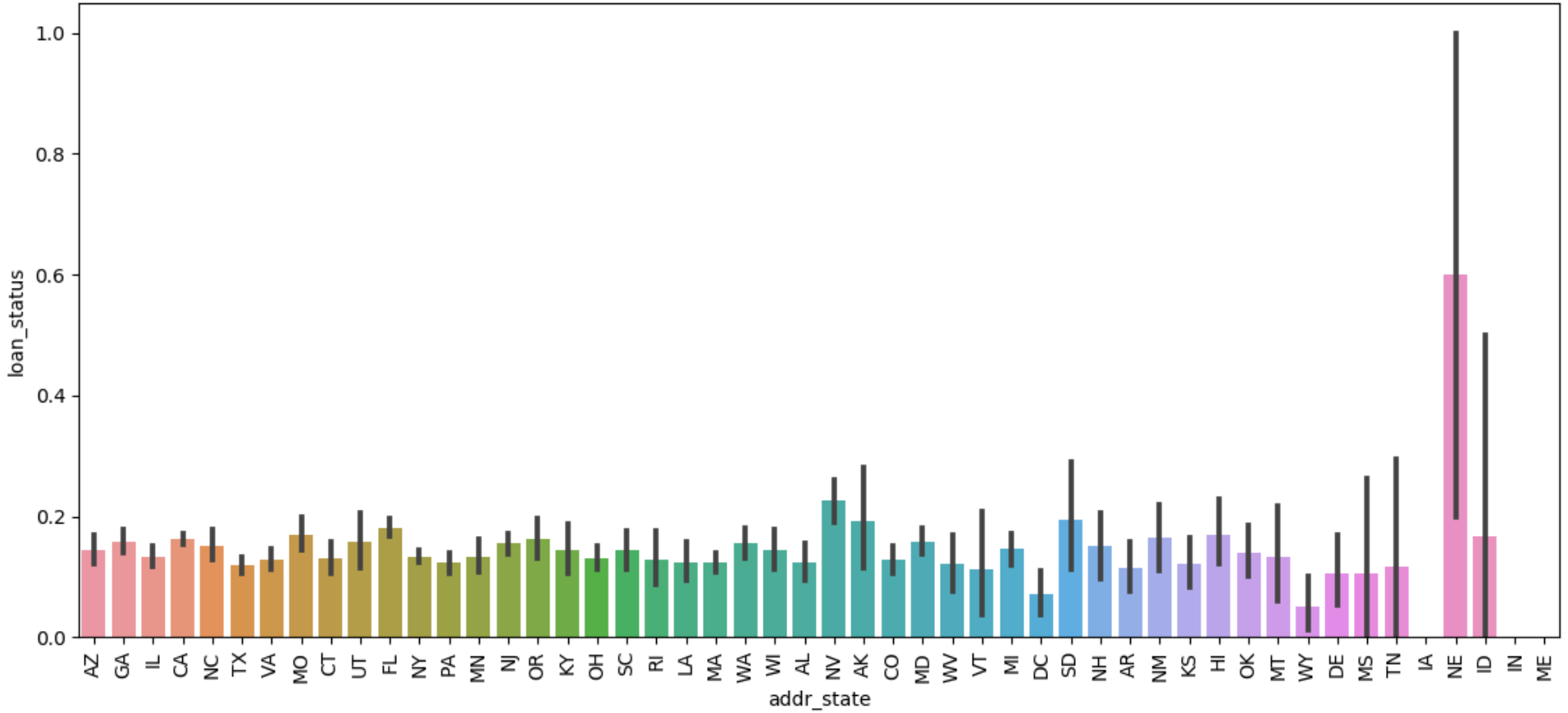
Analysis on Purpose of loan



Trend that Purpose vs loan_status shows :-

Most defaulting borrows mention purpose as SMALL BUSINESS

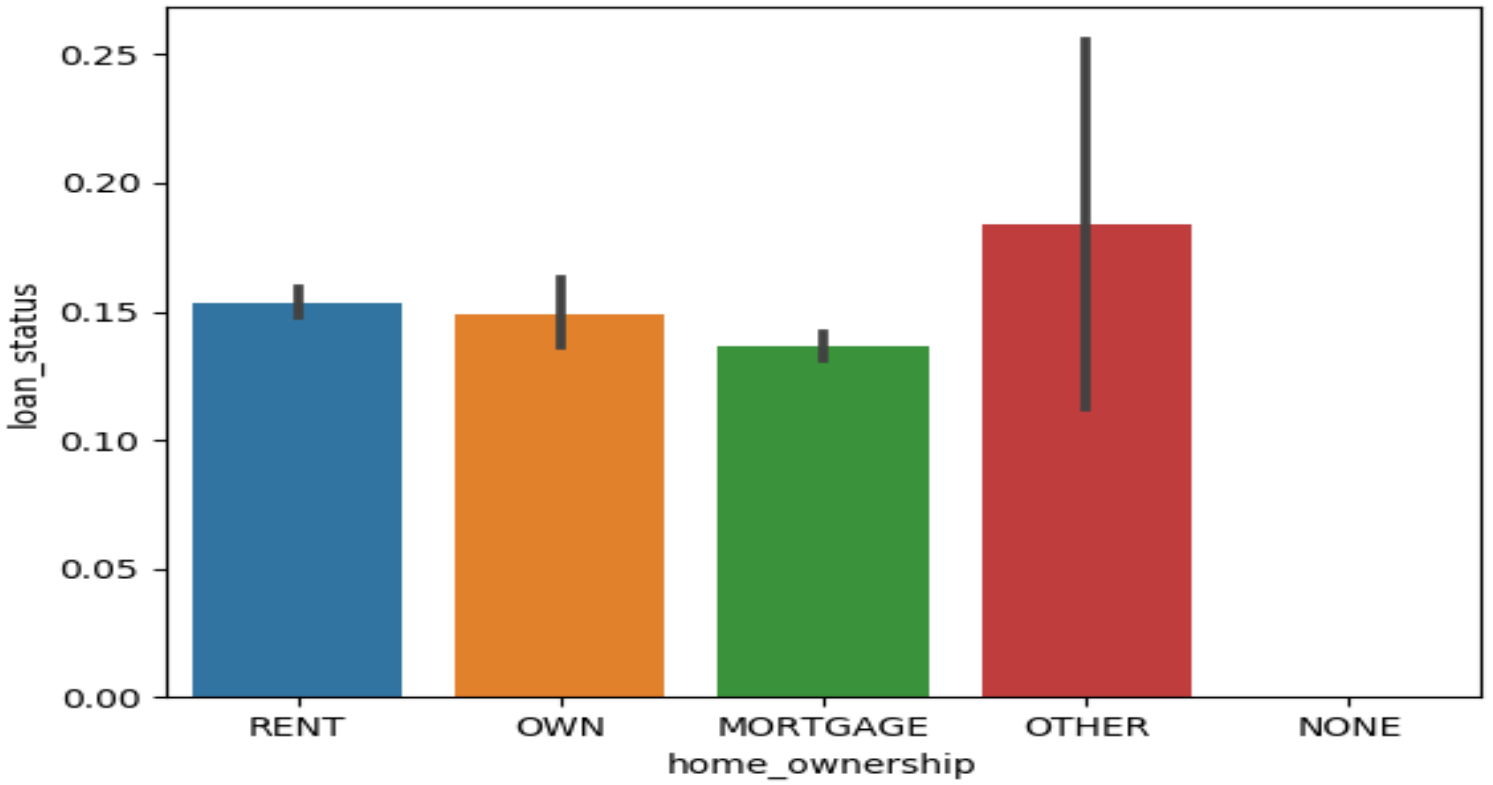
Analysis based on State



Trend that State vs loan_status shows :-

Most defaulting borrowers' are from NEVADA

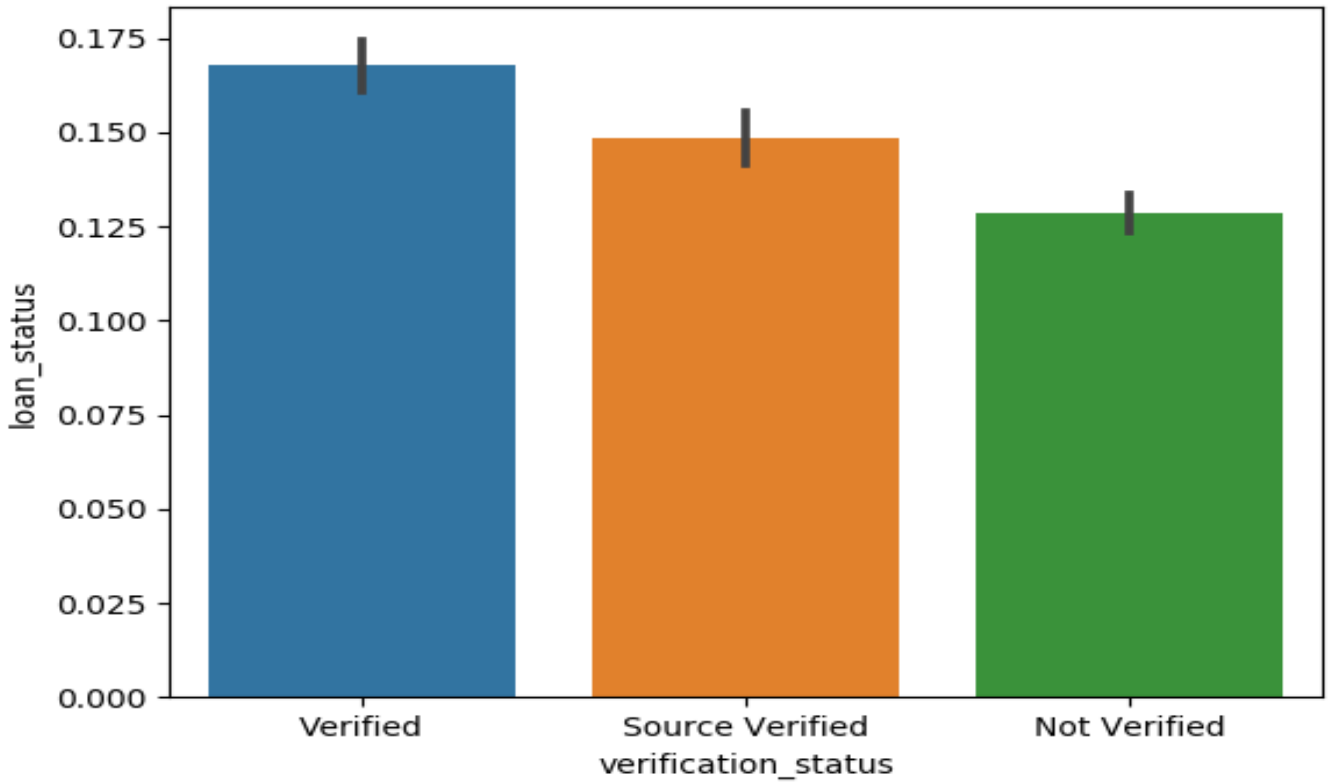
Analysis based on Home Ownership



Trend that Home Ownership vs loan_status shows :-

Most defaulting borrowers have "OTHERS" as ownership status

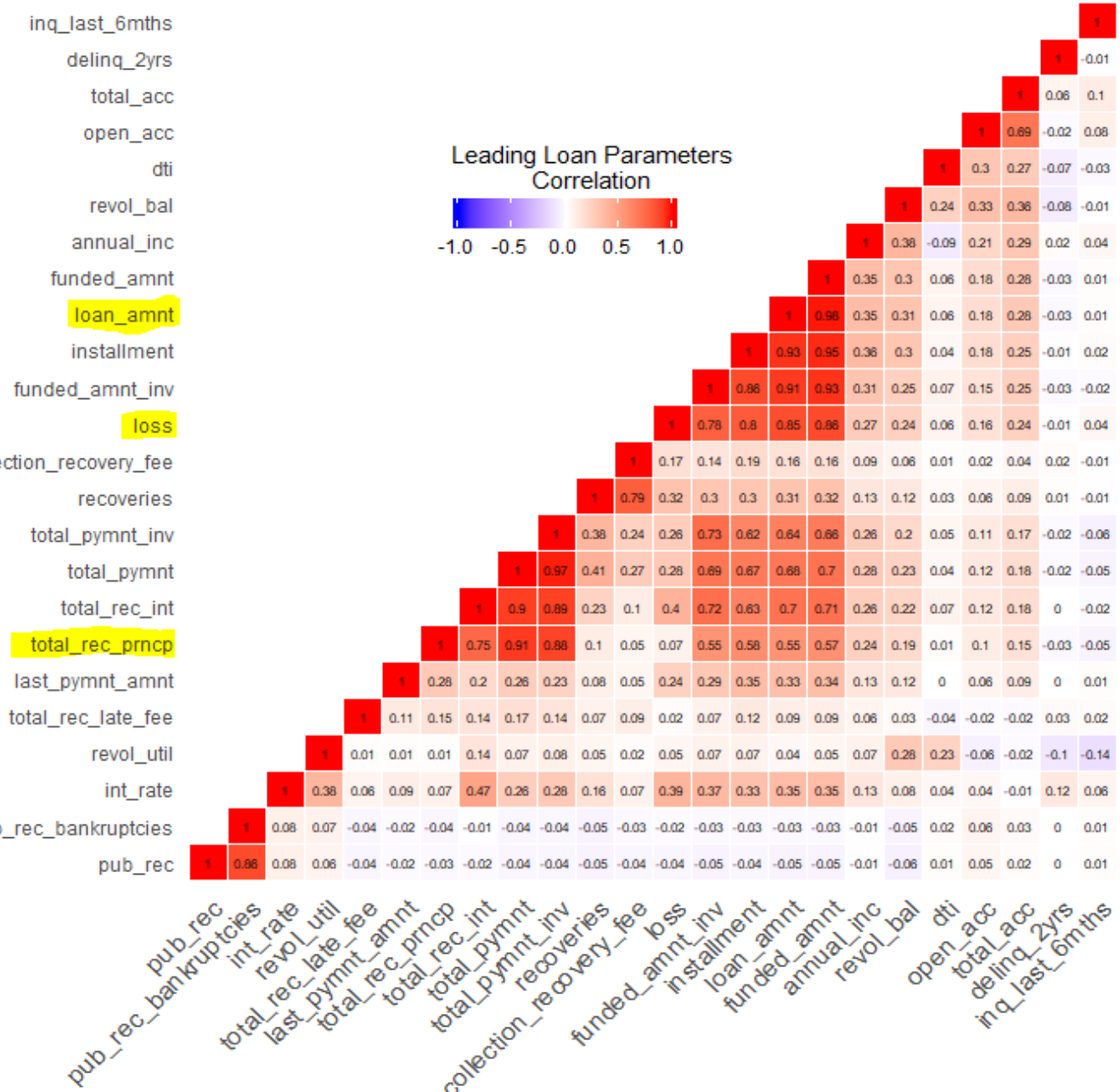
Analysis based on Verification Status



Trend that Verifiaction Status vs loan_status shows :-

Even though, LC has verified borrowers and borrowers sources, the percentage of defaulters is higher than "Not Verified" cases.

Correlation Matrix



- As we can see that for less Total_rec_prncp and high interest rate, percentage of defaulters are most.
- Less total_rec_prncp and moderate interest rate, defaulters are high.
- But we cant use total payment received because it is known only after once the loan is sanctioned
- Co-relation plot that Loan Amount is highly correlated(85%) with total received principal
- Credit Loss = Funded Amount – Principle returned
- Credit loss is highly correlated with loan amount, so we can take loan amount as one of the significant parameter while deciding defaulters

Conclusion

- **As per as the analysis, we found few deciding factors which ends up in determining the loan defaulter applicant. So, the bank should consider the deciding factors before sanctioning the loan to avoid the credit loss.**