EDA CASE STUDY

SUBMISSION

Group Name:
1. Jang Bahadur Singh Umath
2. Jai Shankar Bhagat
3. Satyam Satyajeet
4. Kesab Kumar Nayak

Abstract- For Lending Risk Management and Decision Support System

Why did we start?

•To Minimize Risk of Lending

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

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

Objective:

*   Identification of Loan Applicant traits that tend to 'default' paying back

*   Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena

*   Gramener may choose to utilize this knowledge for its portfolio and risk assessment of new loan applicants

Data Exploration

- There are 111 variables in the loan dataset.

- Most of the variables contains a single value or more number of Nas

- There are total 39717 records in the dataset.

ANALYSIS OF LOAN STATUS

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

Data Cleaning and Manipulation

Dealing with Missing values :

- In this dataset 54 variables contain all the observations as NA so, they are removed.

- 14 columns have more than 70% of 0 so they are removed.

- If some column contain missing values, then we replaced it with the median value.

- 4 columns from remaining has same data in column hence they were removed.

- Date is converted into standard format and % is removed from columns wherever required.

- Removed the variables having too many levels like title.

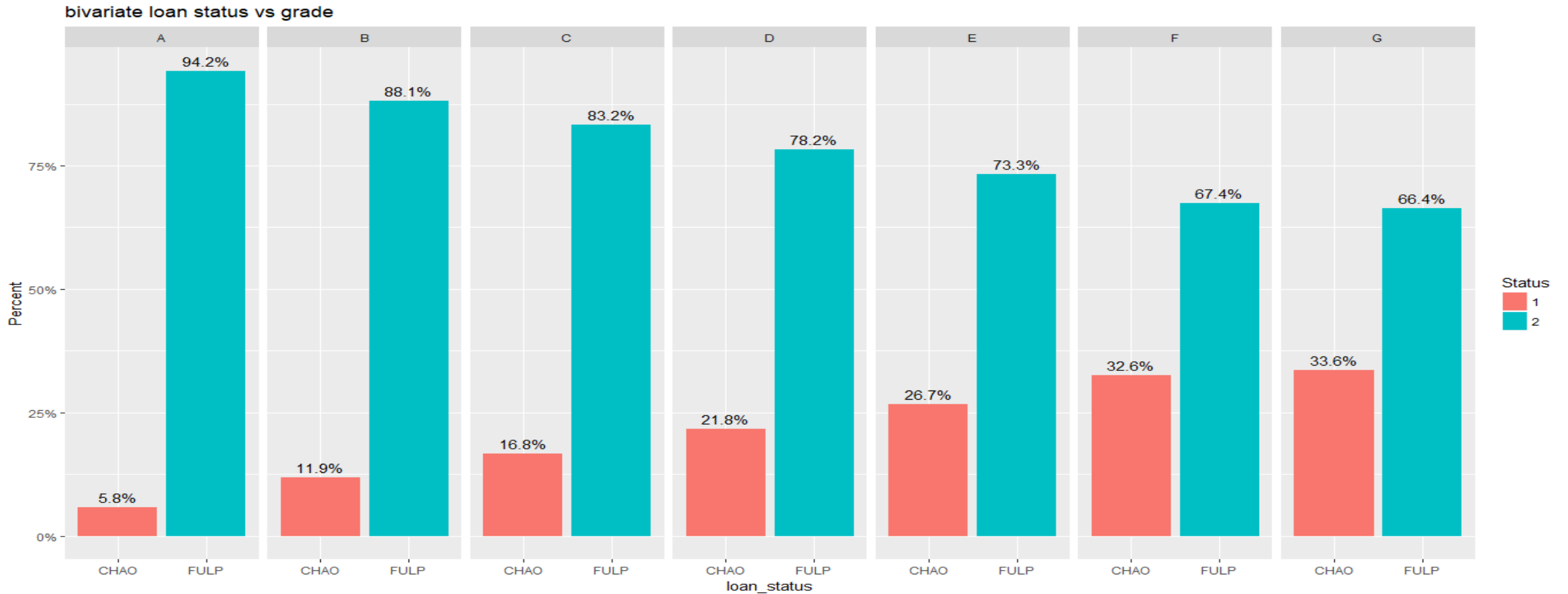- Removed the columns having more than 50% NAs.

Top 6 Deciding Parameters for Loan Defaulter

1. Loan Amount

2. Grade

3. Purpose of Loan

4. State

5. Home Ownership

6. Verification Status

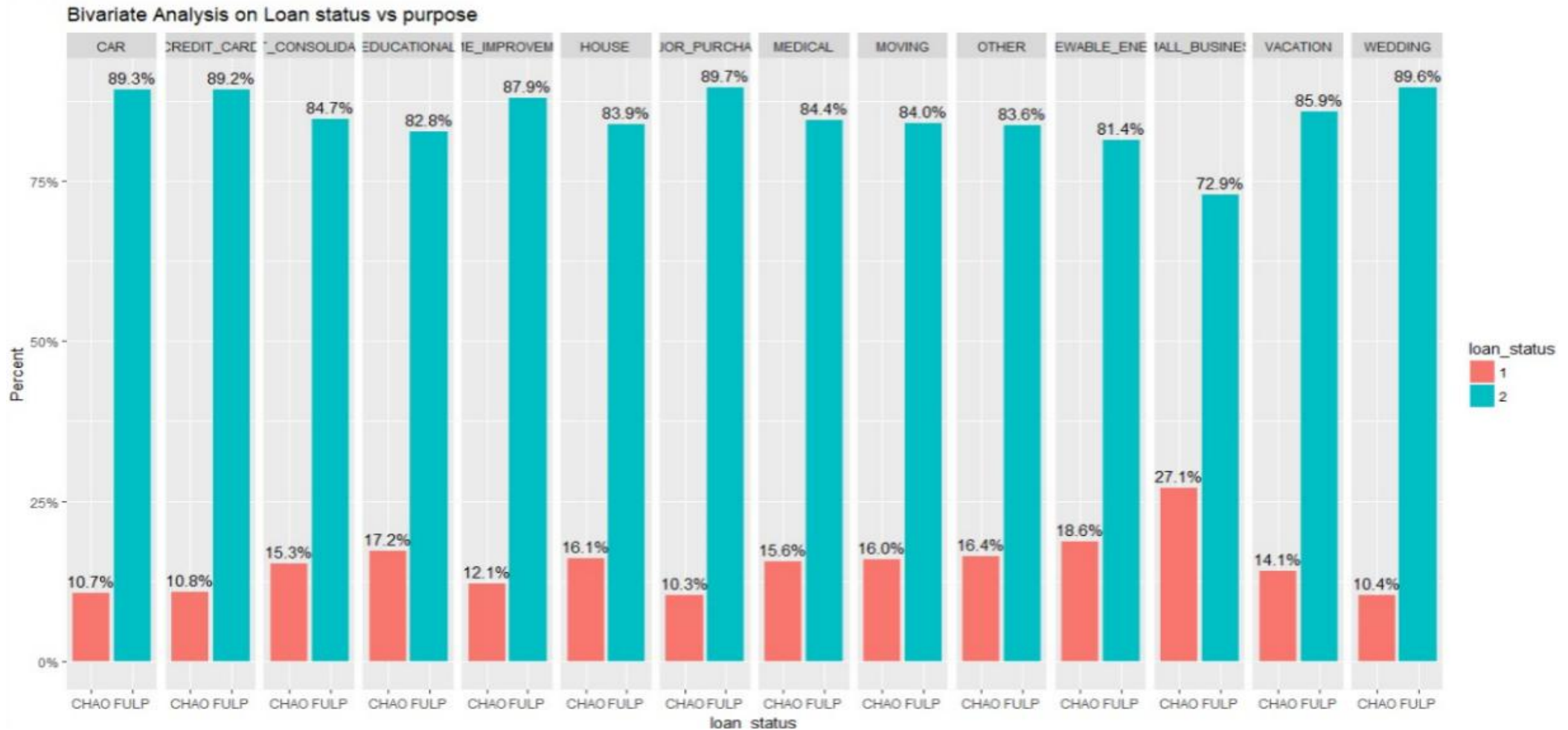Following slides are the visualization of above variables.

Analysis on Loan Amount

Bivariate analysis on Loan Amount
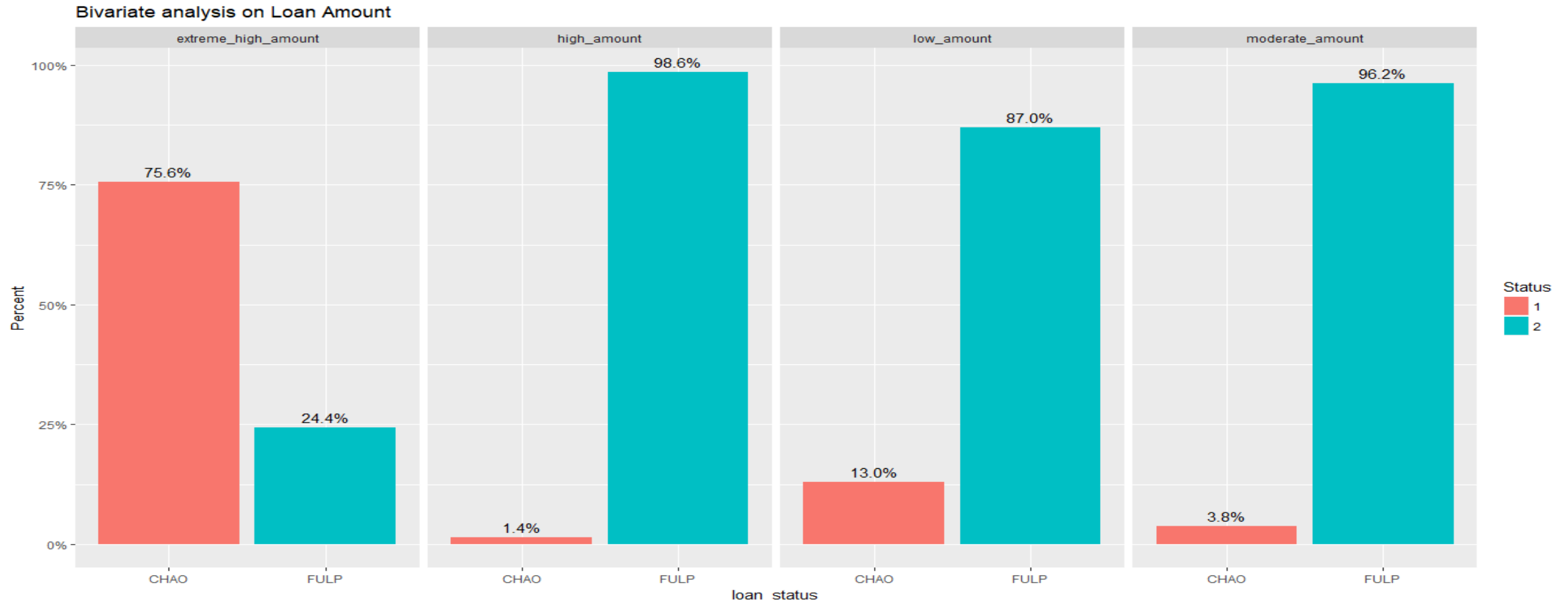
Insight 1: Most defaulting borrowers have very high loan amount

Analysis on Grade of Borrower

bivariate loan status vs grade

Insight 2: Loans with Grade greater than C are most likely to be defaulted

Analysis on Purpose of Loan

Bivariate Analysis on Loan status vs purpose

Insight 3: Most defaulting borrows mention purpose as SMALL BUSINESS

Analysis based on State

Status of Loan on basis of State
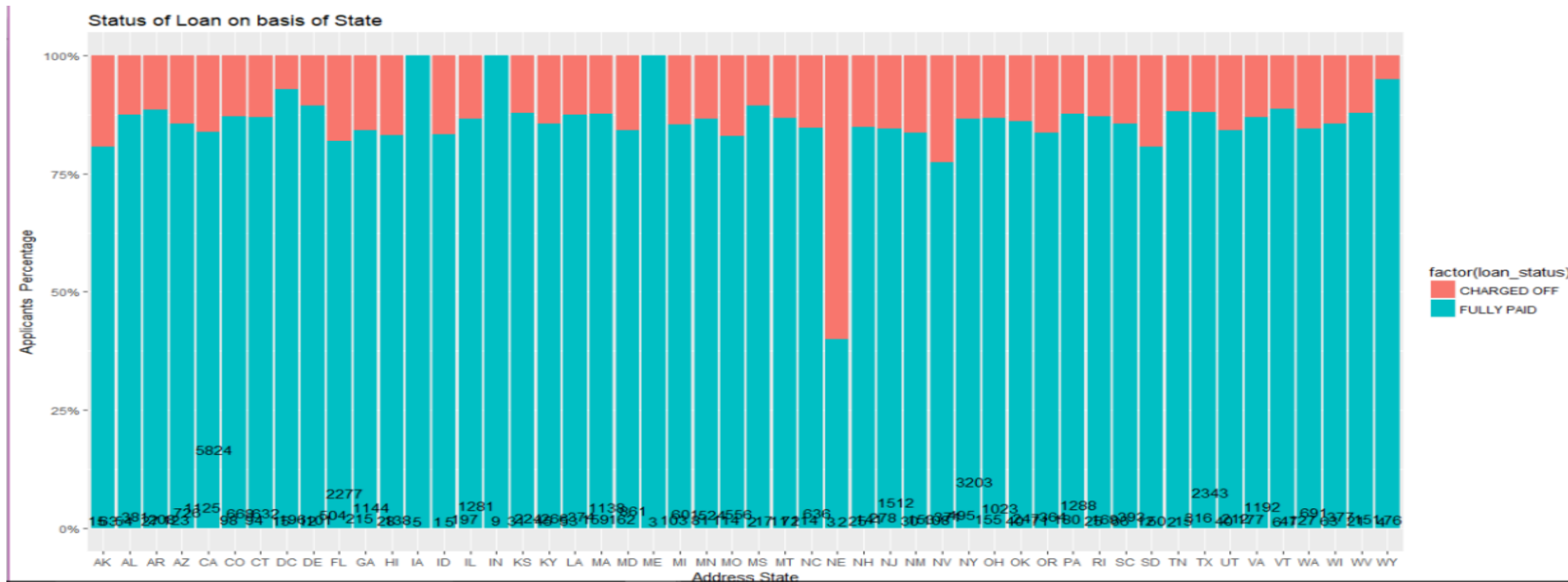
Insight 4: Most defaulting borrowers' are from NEVADA

Analysis on Home Ownership

Bivariate Analysis on Loan status vs Home Ownership
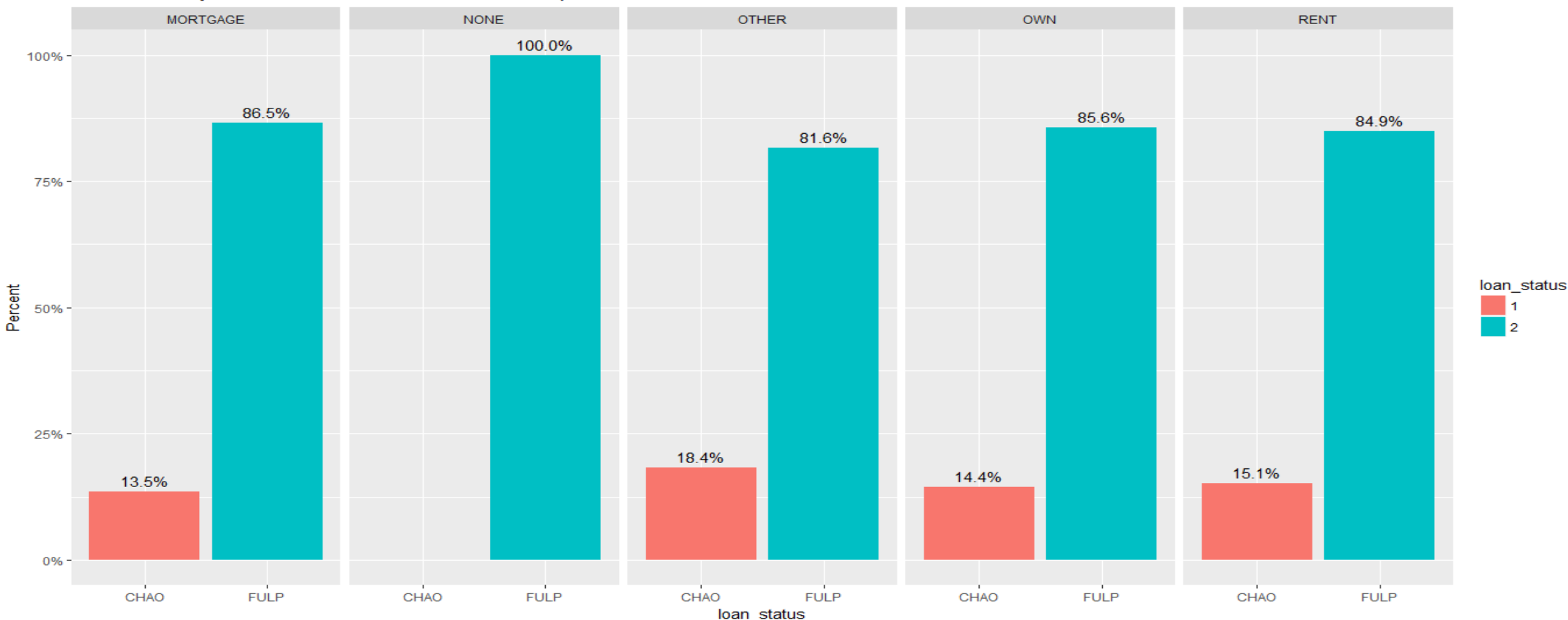
Insight 5: Most defaulting borrowers have "OTHERS" as ownership status

Analysis based on Verification Status

Bivariate analysis on verification status
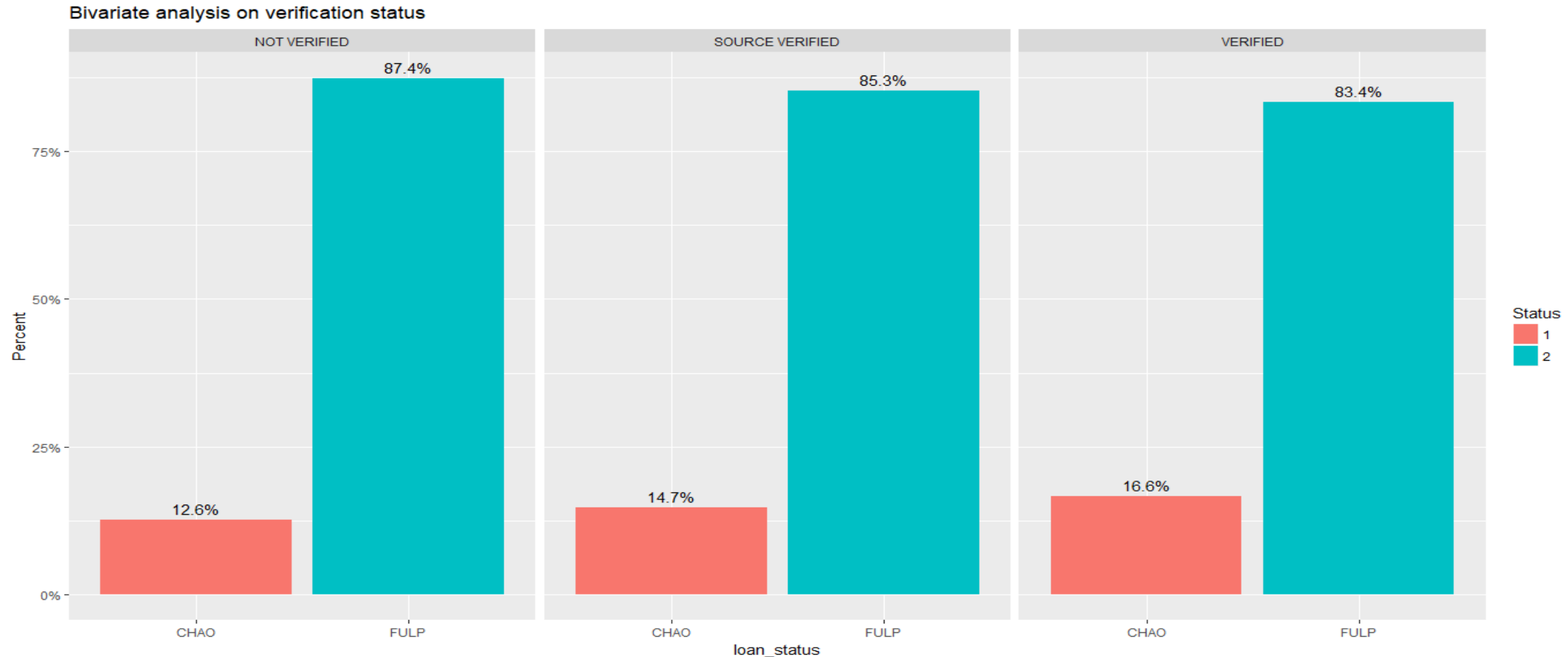
Insight 6: Even though, LC has verified borrowers and borrowers sources, the percentage of defaulters is higher than "Not Verified" cases.

Leading Loan Parameters Correlation
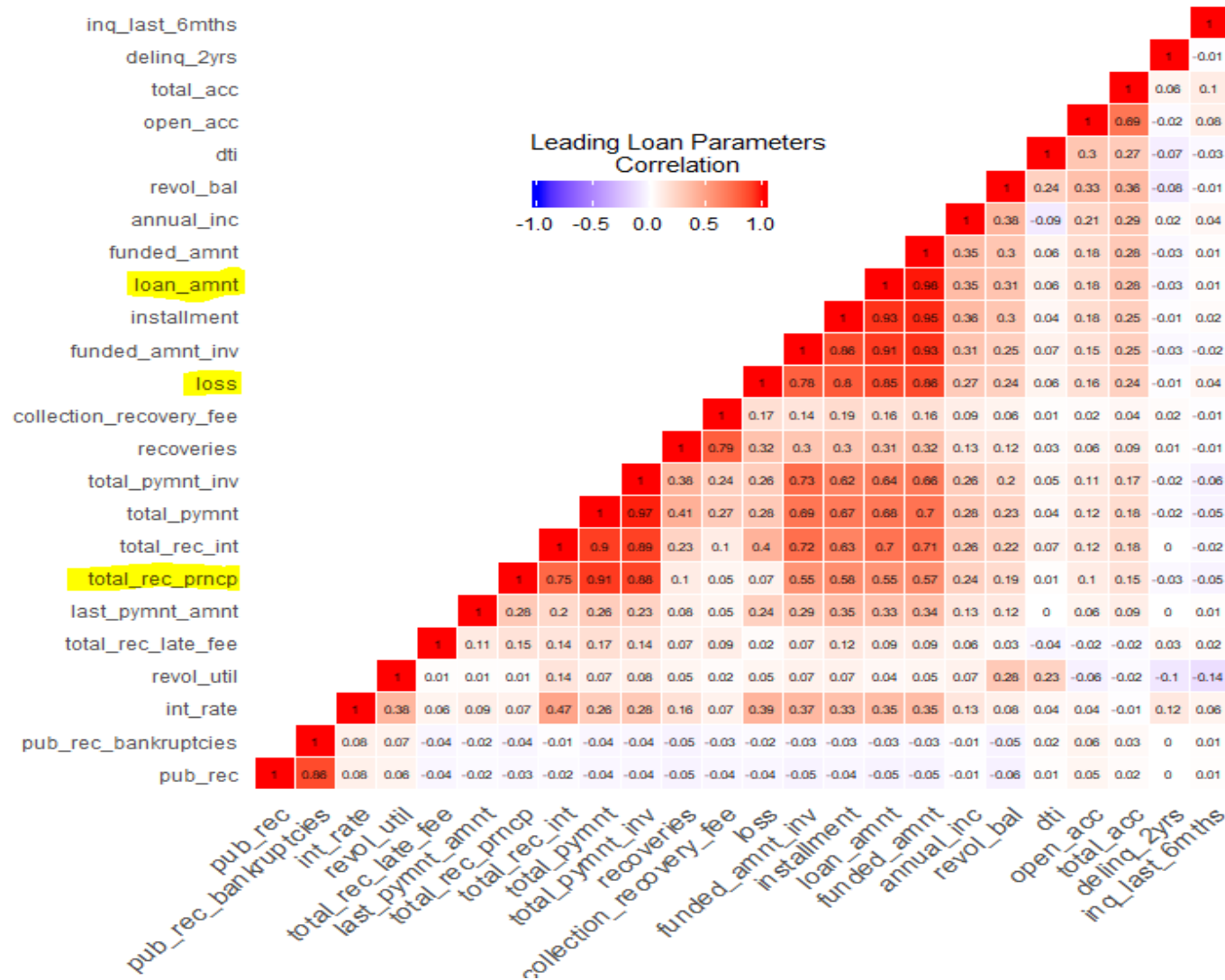
- As we can see that for less Total_rec_prncp and high interest rate, percentage of defaulters are most.
- Less total_rec_prncp and moderate interest rate, **defaulters are high**.
- But we cant use total payment received because it is known only after once the loan is sanctioned
- Co-relation plot that 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**

Analysis Based on Interest Rate and Loan Amount



Analysis based on interest rate vs Loan Amount
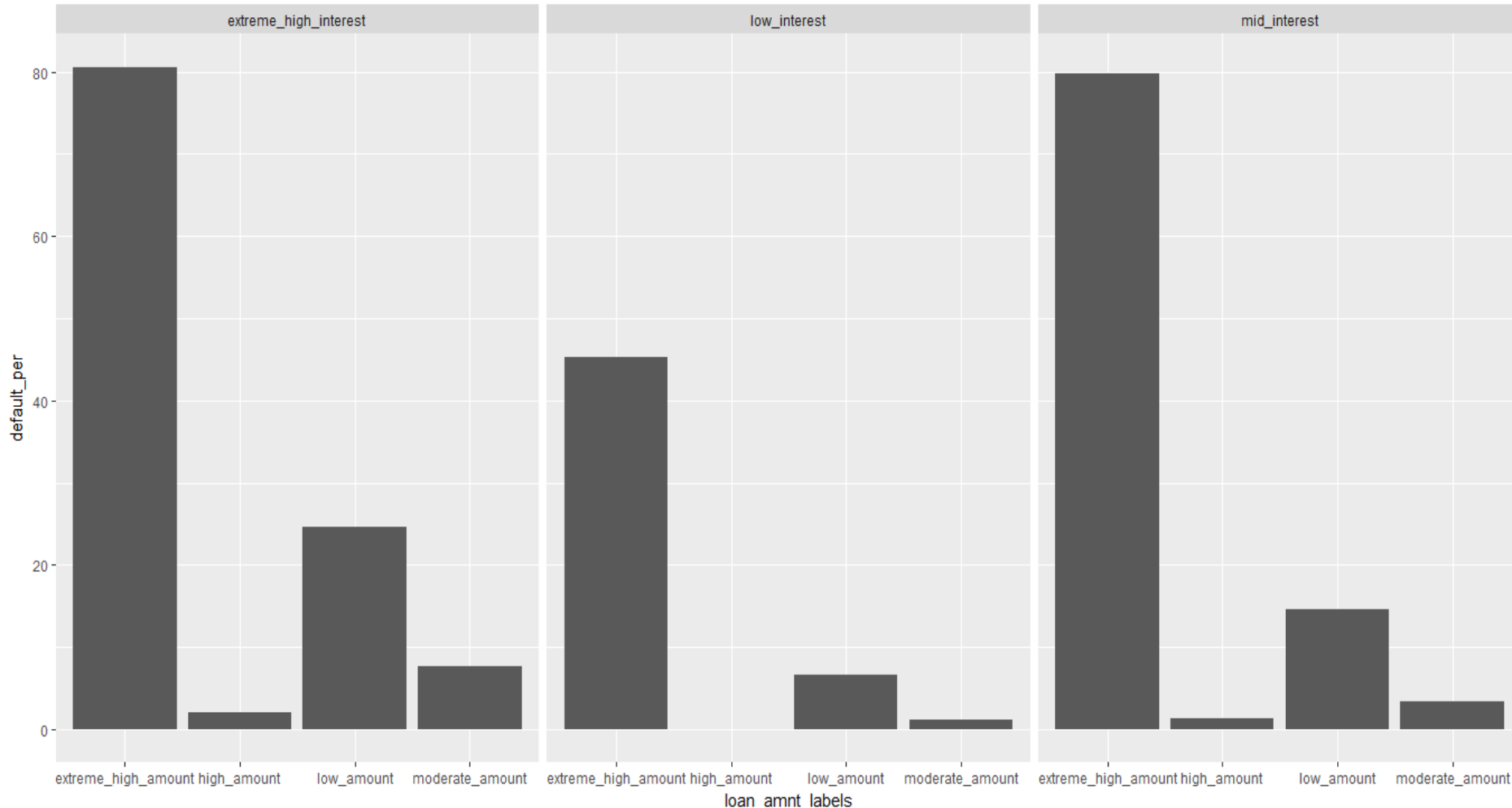
- low interest and extremely high amount given can lead to 45% default
- In case of extremely high amount and high interest and mid interest are contributing to 80% of defaults
- In case of low amount extremely high interest is significant

Recommendation

Multivariate Analysis on loan amount, interst rate , purpose and loan status
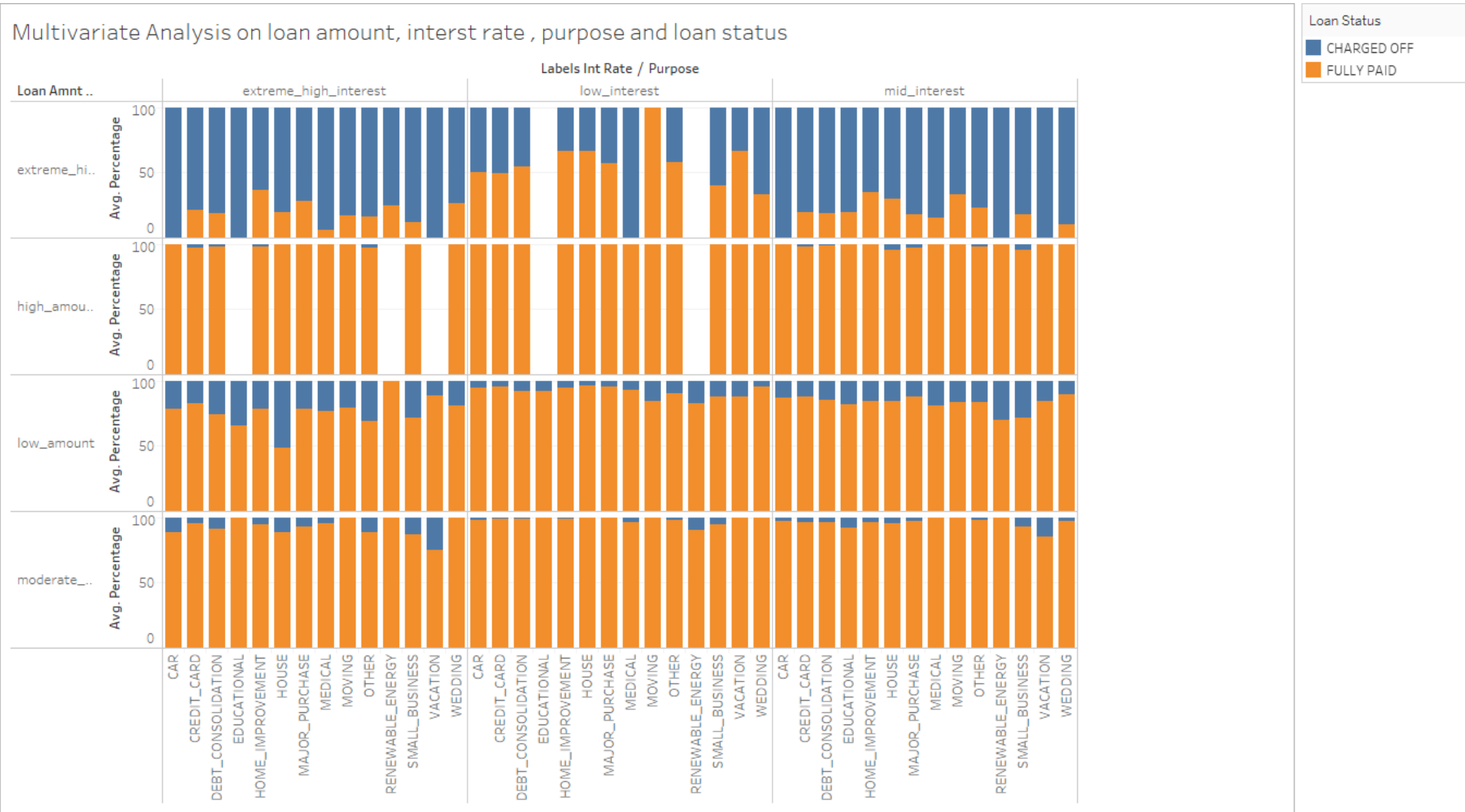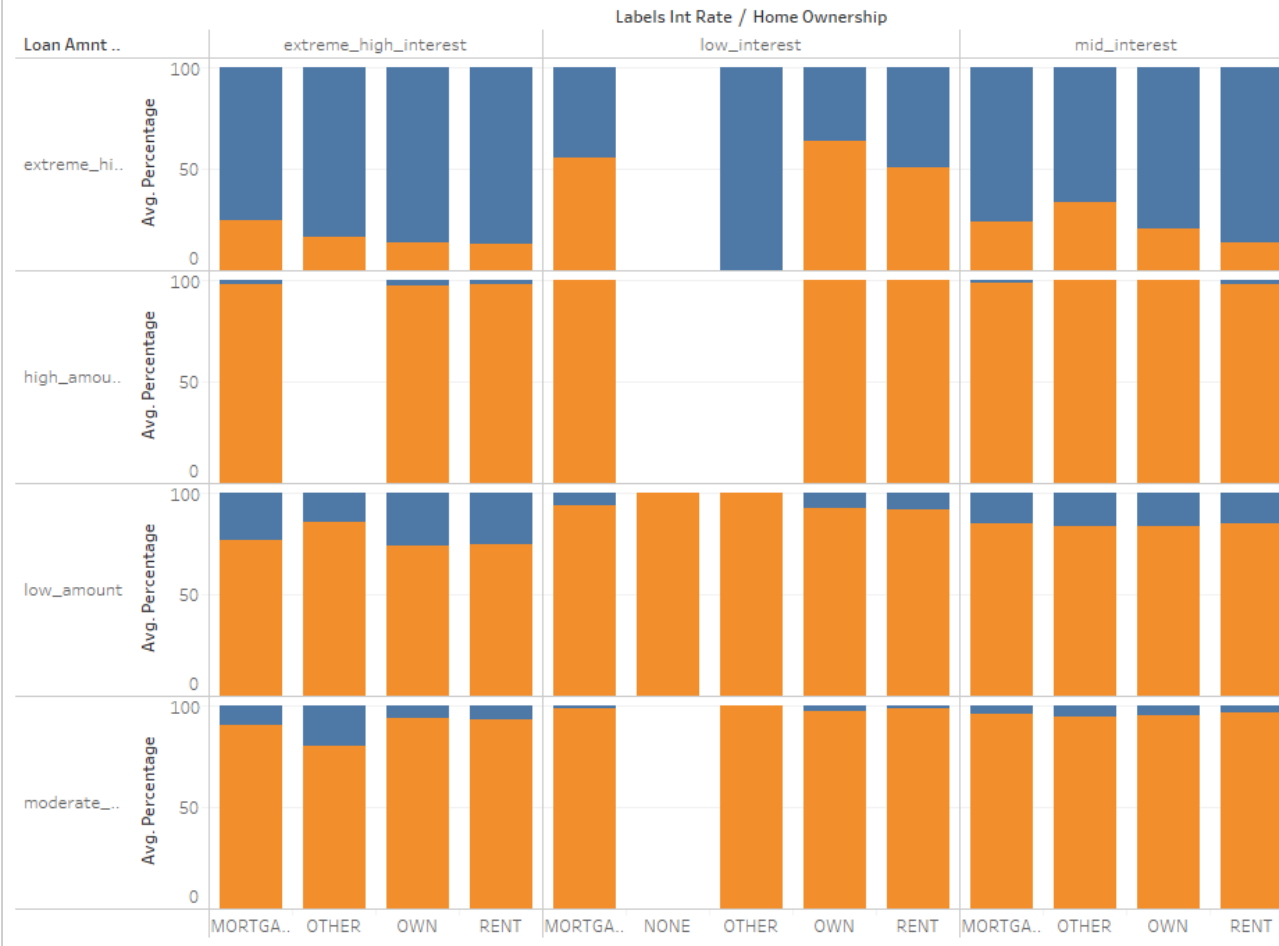
- For Extremely high loan amount and extremely high interest rate , Small Business and Debt Consolidation are leading into maximum defaulter.

- For Extremely high loan amount and mid interest rate , Small Business and Debt Consolidation are leading into maximum defaulter

Recommendation

- For Extremely high loan amount and extremely high interest rate , Home Ownership as OTHERS is leading into maximum defaulter.

- For Extremely high loan amount and mid interest rate , Home Ownership RENT is leading into maximum defaulter.

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.