LENDING CLUB CASE STUDY

SUBMISSION

Group Name:

1. Soura Khandual
2. Mallikarjuna Das
3. Abhishek Prusty
4. Megha Goyal

Problem Statement

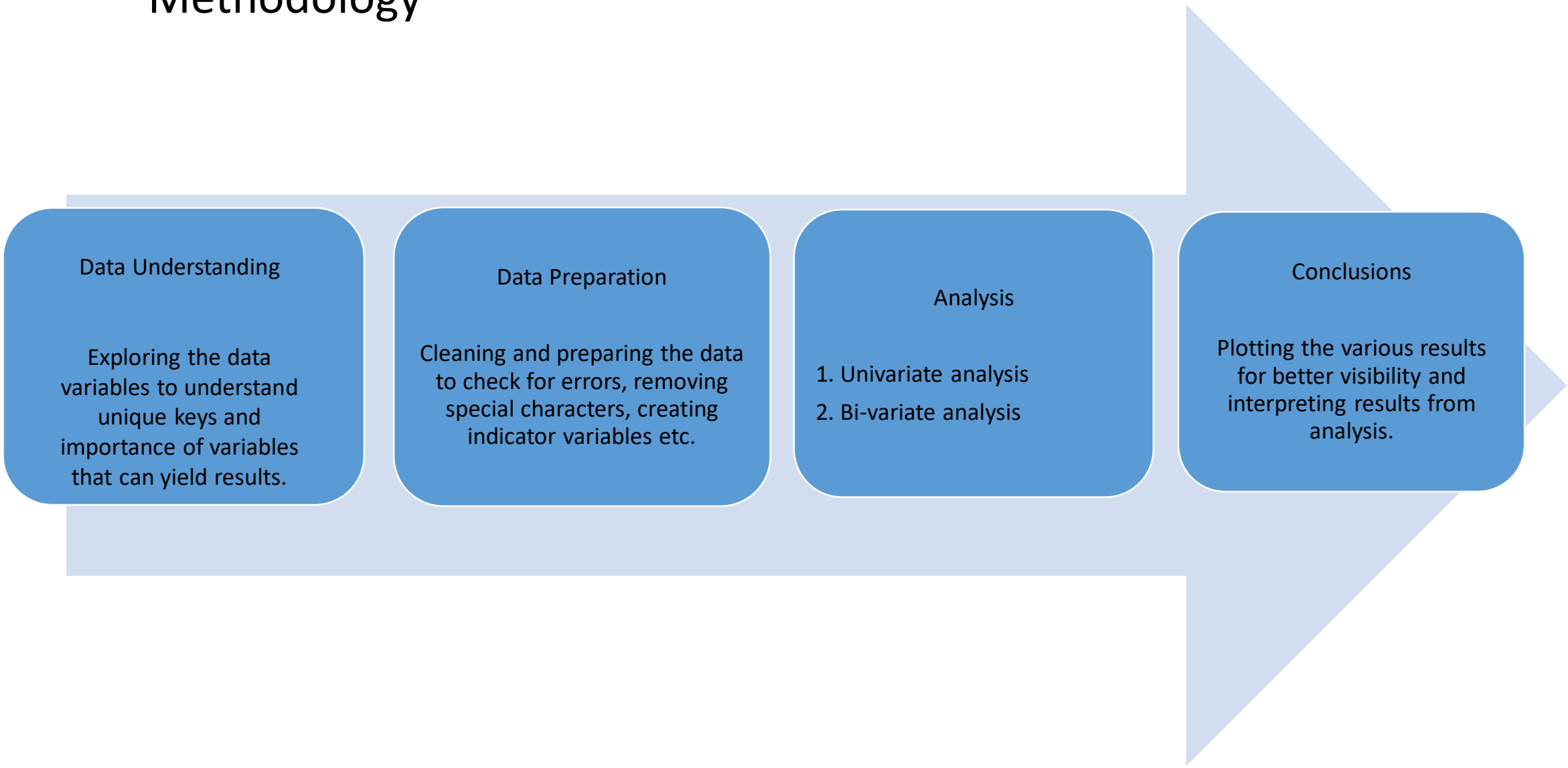
•A consumer finance company which specializes in lending various types of loans to urban customers wants to decide for loan approval based on the applicant's profile.

Business Objective:

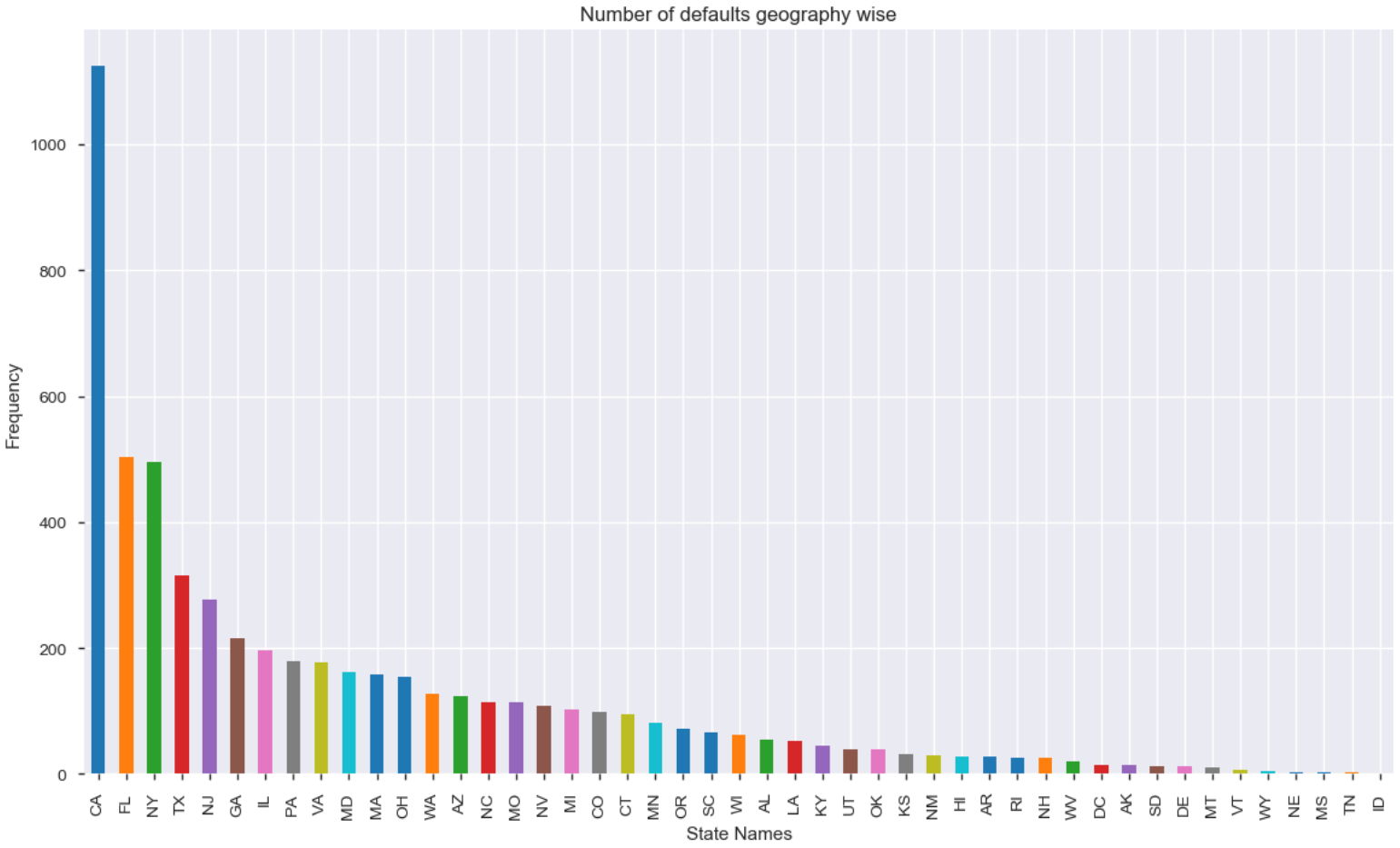
•The company wants to understand the **driving factors (or driver variables)** behind loan default. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

•The company can utilise this knowledge for its portfolio and risk assessment. If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.

Methodology



Univariate Analysis – State wise default cases

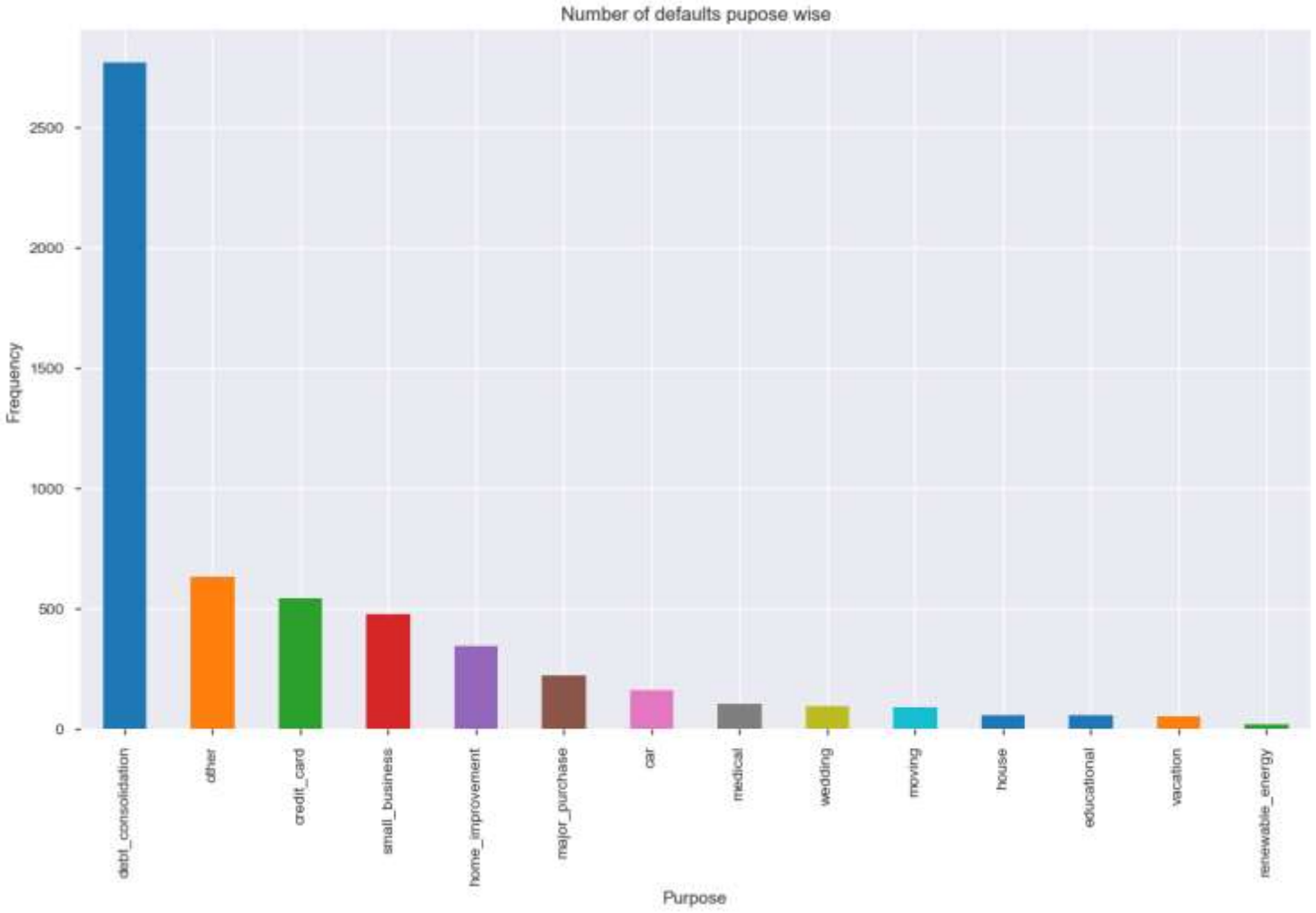


The graph of the left shows the number of default cases in every state.

As depicted, the highest number of defaults are in CA, almost twice more than that of FL and NY. This is nearly 20% of the total default cases in these states.

The lowest default cases arising from WY, NE,MS,TN,ID.

Univariate Analysis – Purpose wise default cases

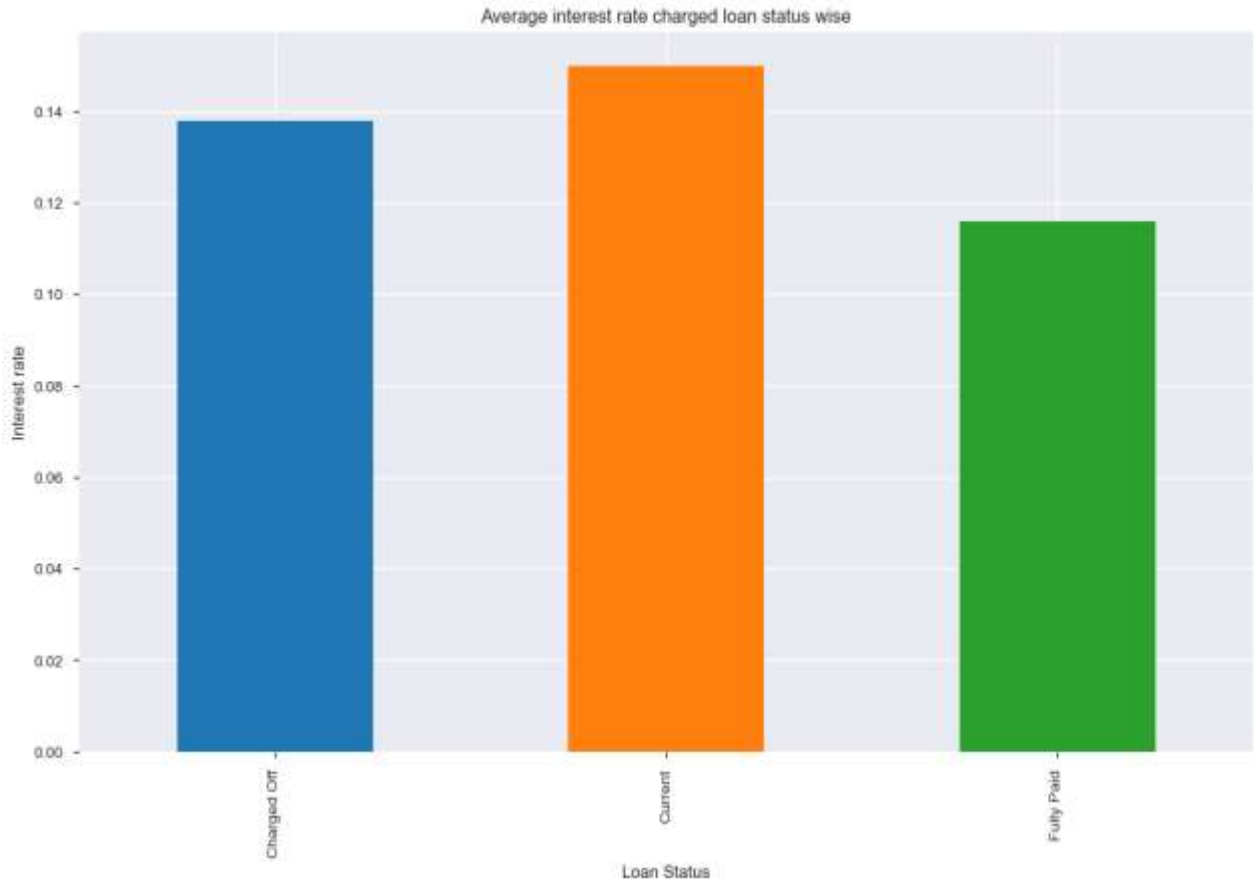


The graph on the left shows the number of defaults with respect to the purpose of the loan.

As depicted, the highest number of defaults are with the debt_consolidation purpose with nearly 47% of the total default cases. The borrower takes loan in order to pay several other loans and ends up defaulting.

The lowest default case accounts to renewable energy with just 0.34% of the total default cases .

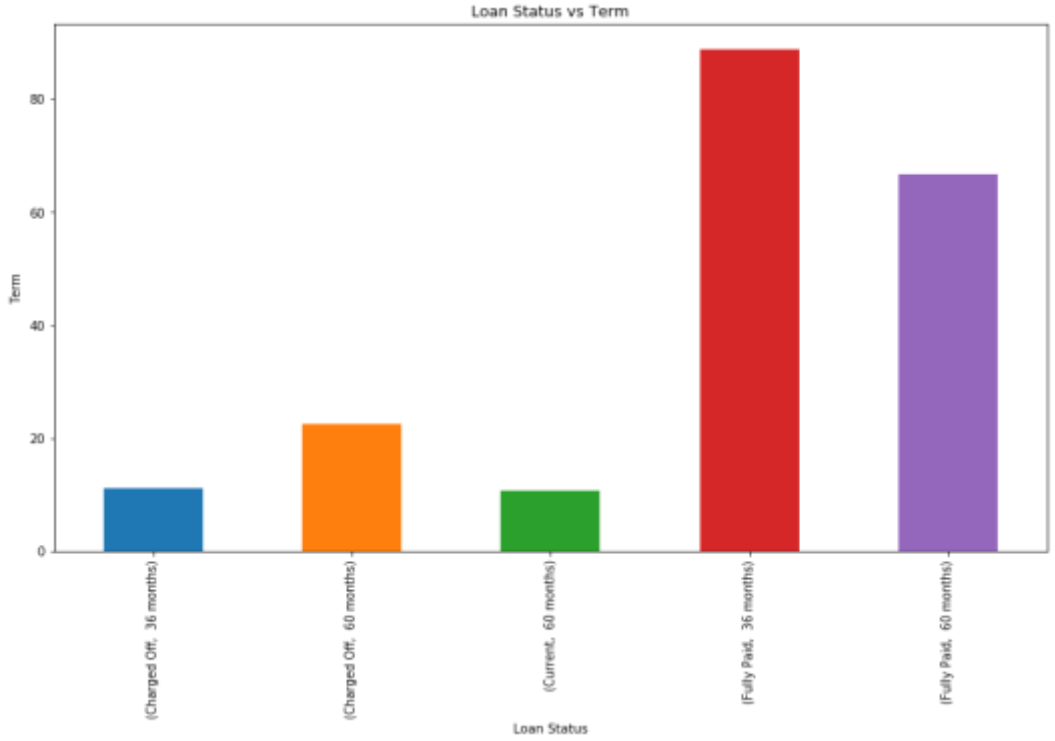
Univariate Analysis – Average interest rate charged based on loan status



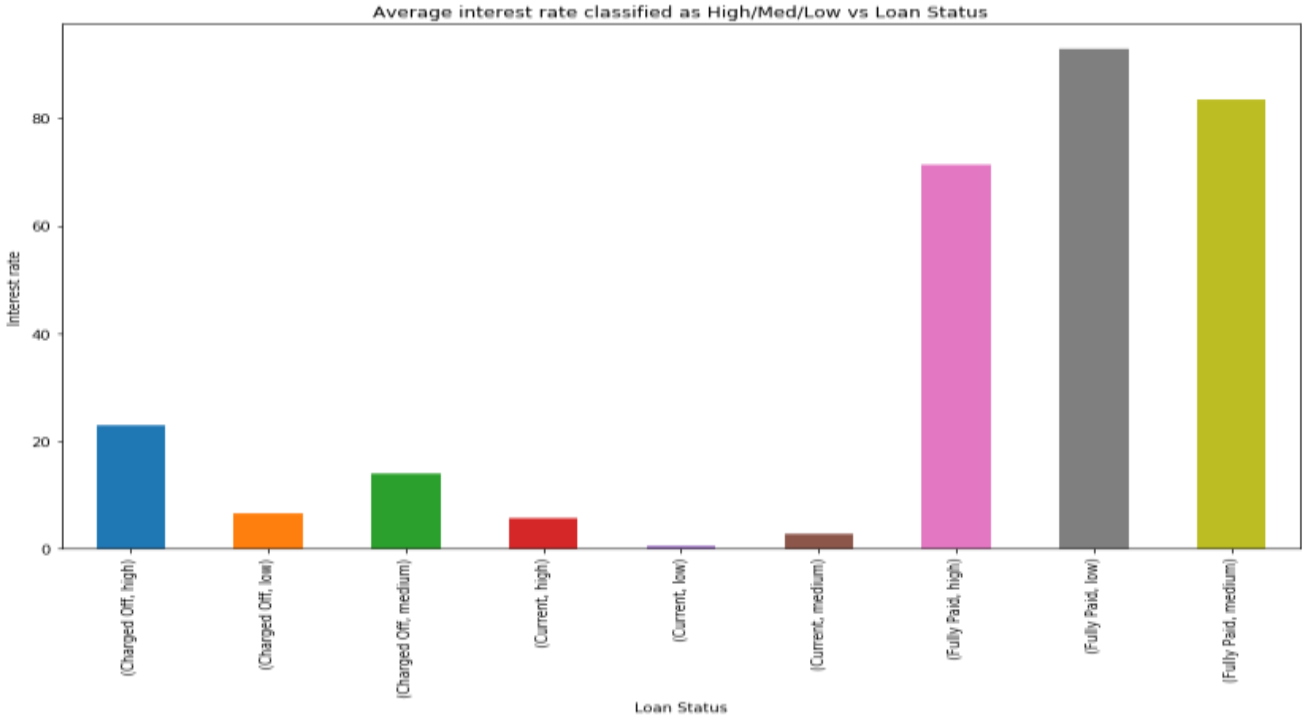
The graph on the left shows the average interest rate charged basis loan status of the borrower.

The average interest rate charged is highest for current loan status which is approx 15%.
The charged off cases have an interest rate of 14% on average which is competent with current loans.
The lowest interest rates are seen for fully paid loan cases accounting to 11.6%.

Bivariate Analysis-1

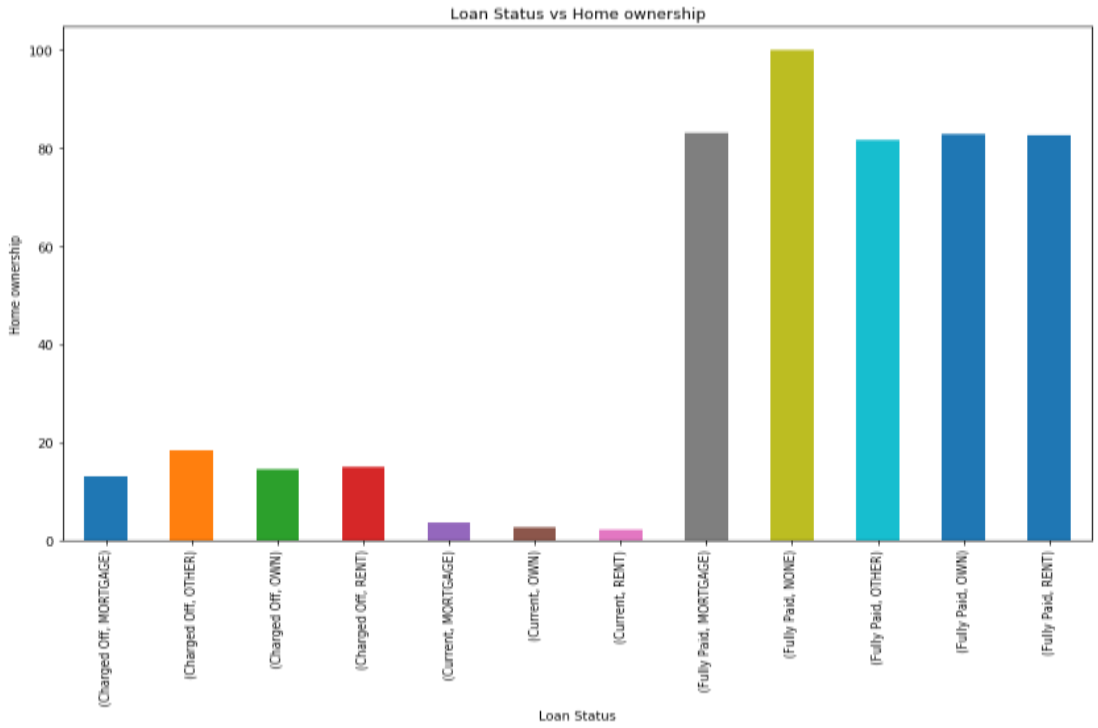


Plot 1:
Hypothesis: Term impacts the defaults.
Conclusion: The term period the loan clearly impacts defaults. 23% of those who took a loan with a 60 month term defaulted as against 11% of those with 36 months.
Clearly, one should be giving lesser 60 month loans



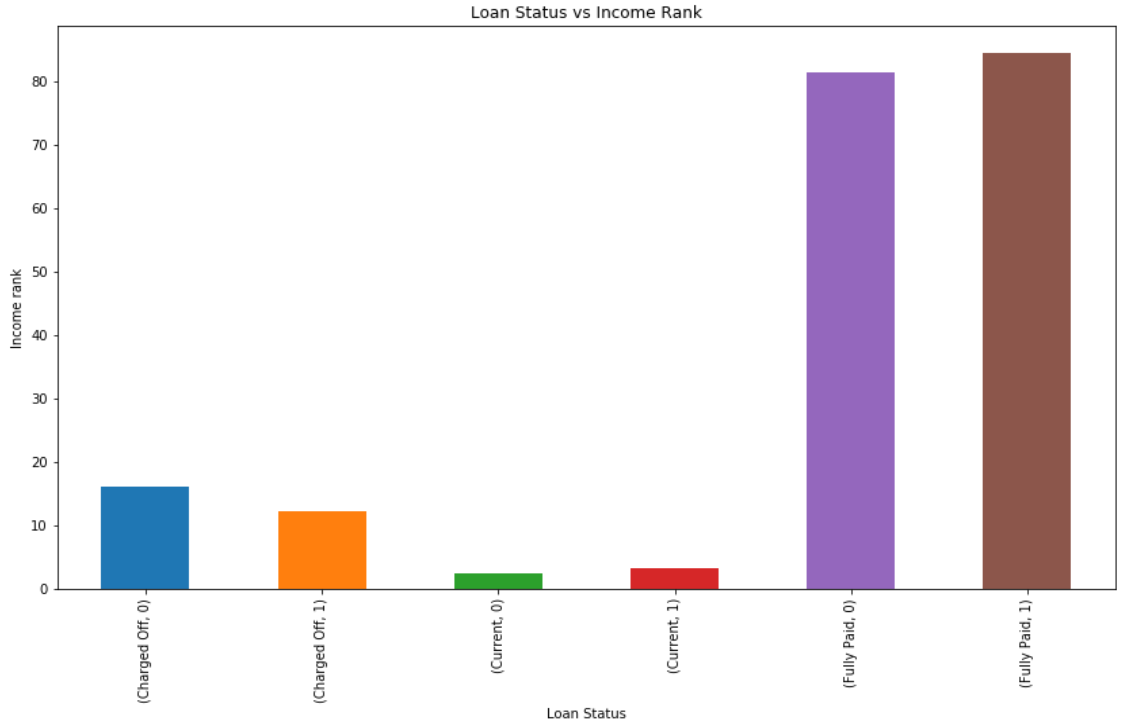
Plot 2:
Hypothesis: Interest rate and defaults have a positive corelation..
Conclusion: Clearly, a higher interest rate leads to more defaults. Those with a very high interest rate (in my classification, an interest rate > 14.4% is classified as VH), default at 23.37% as against an average of 14.17%. Low interest rates leads to lesser defaults. Only 6.6% of loans in this bucket defaulted. This is something for the management to think about. Are they risking their loan by charging too high an interest rate?

Bivariate Analysis-2



Plot 3:

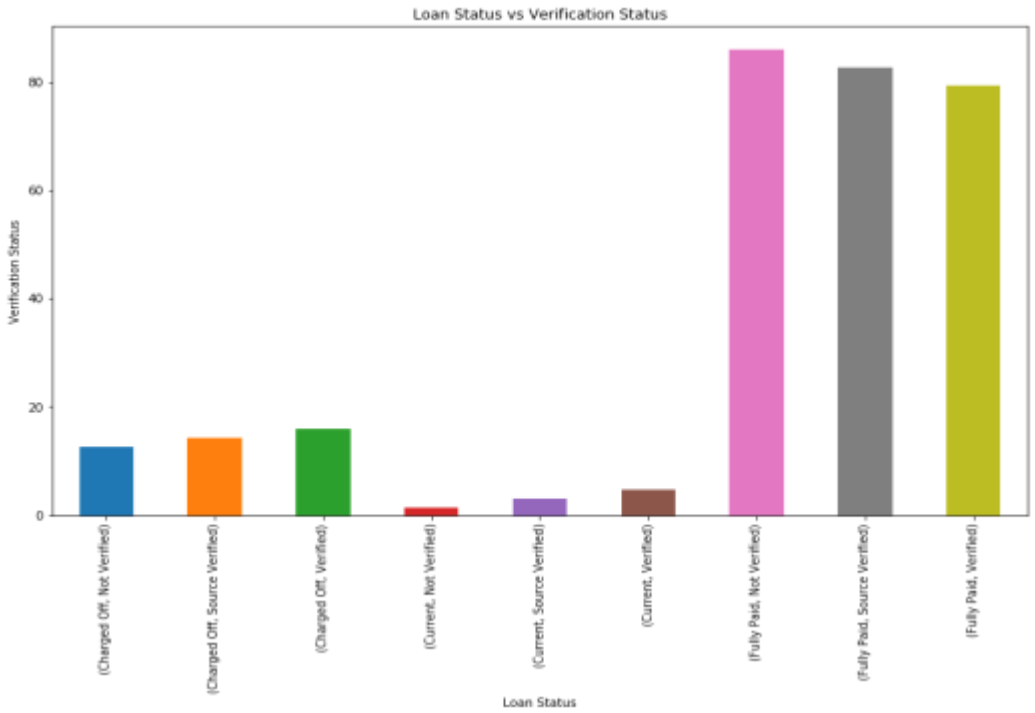
Conclusion : Those with mortgage homes have a slightly lesser default rate than average - approximately 7%. Though not very high it is still a significant one given the sample size. Possibly those with mortgages have an inculcated habit of paying loans on time and hence are better customers.



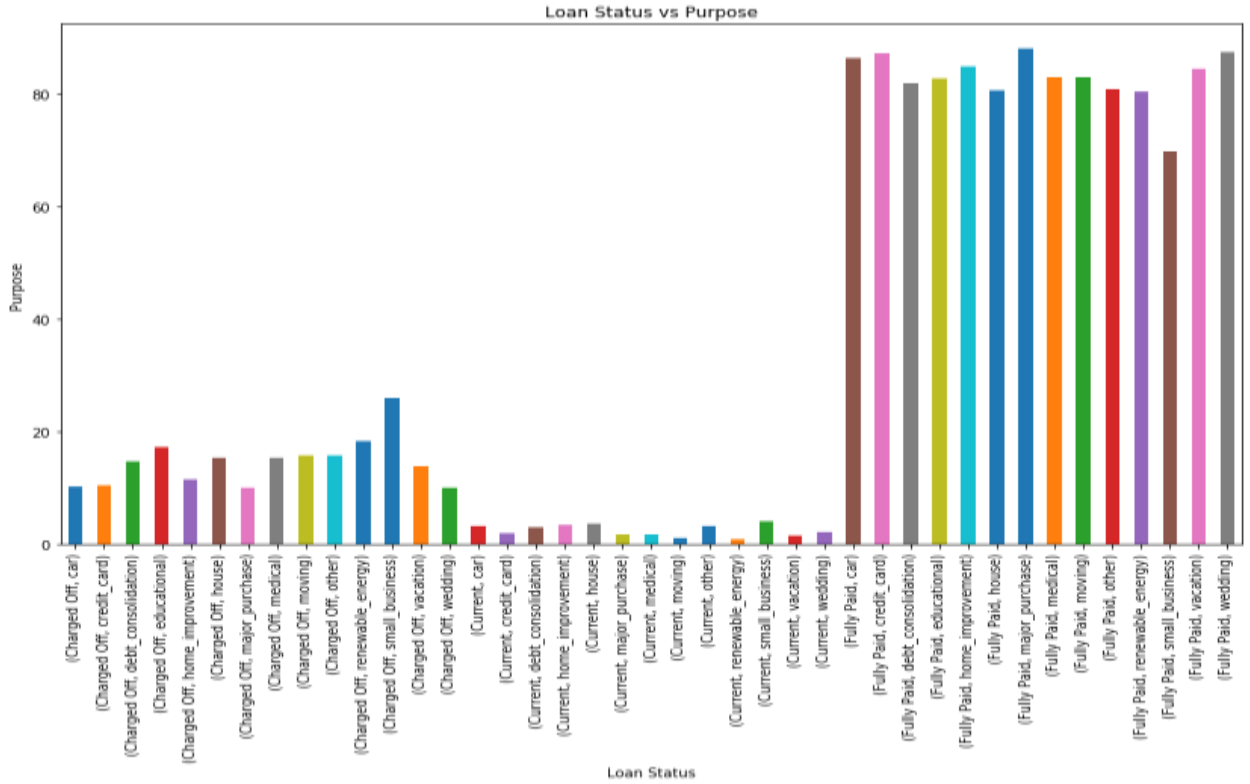
Plot 4:

Conclusion : Clearly, those with lower income have a higher default rate. They default approximately 14% higher than the average.

Bivariate Analysis-3

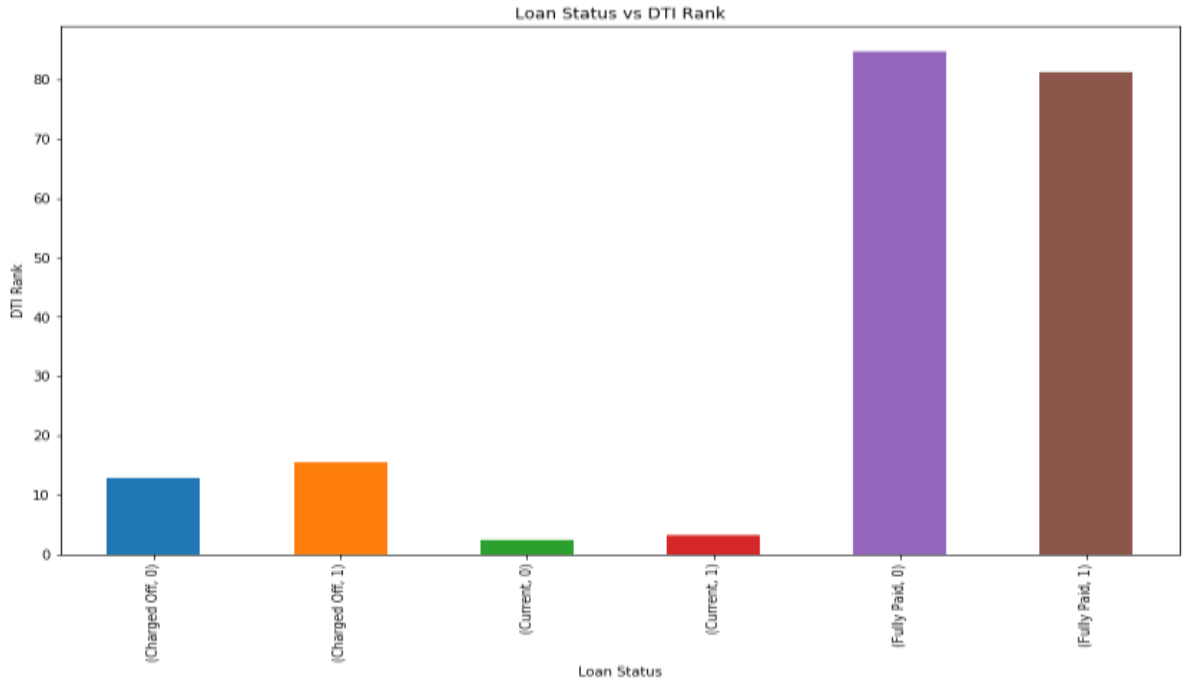


Plot 5:
Hypothesis: Verified loans will have lower defaults.
Conclusion : The data shows otherwise! Not verified loans have a 11% lower chance of default. The bank needs to check its sources of income verification and how accurate it is.

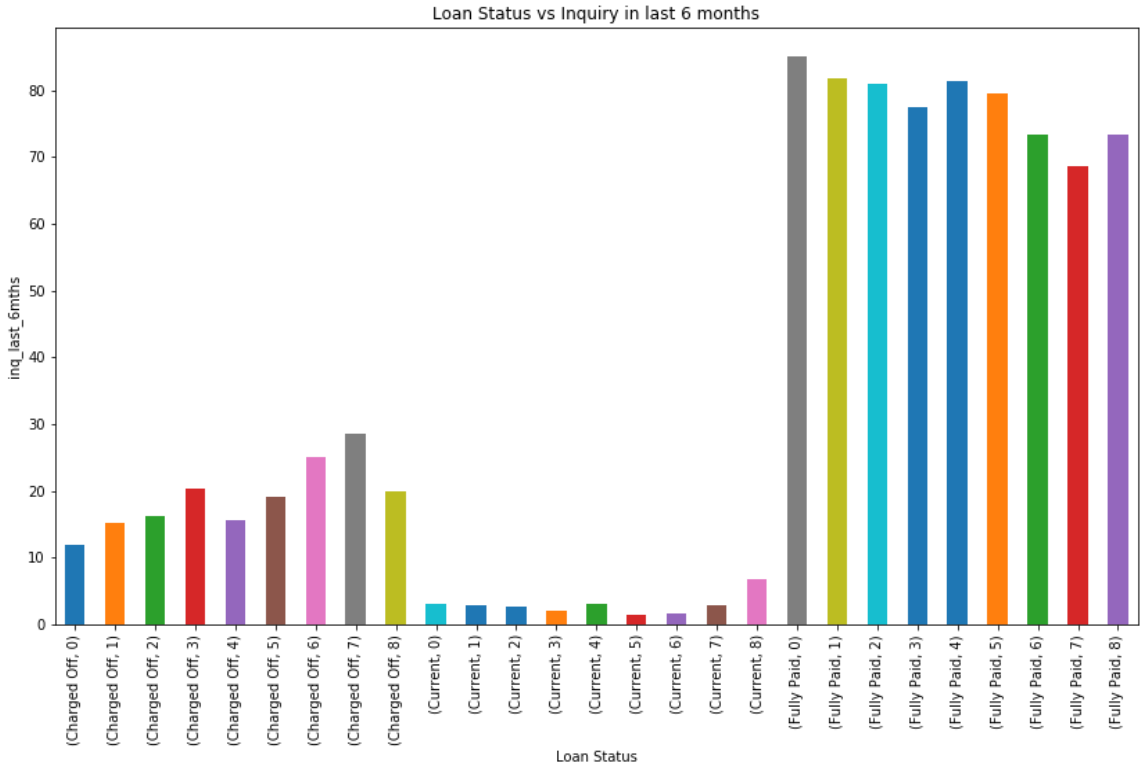


Plot 6:
Hypothesis: The purpose of applying loan should not impact defaults.
Conclusion : The purpose of taking the loan has an impact on defaults. Small business, renewable energy, educational are indexing significantly higher on defaults.

Bivariate Analysis-4

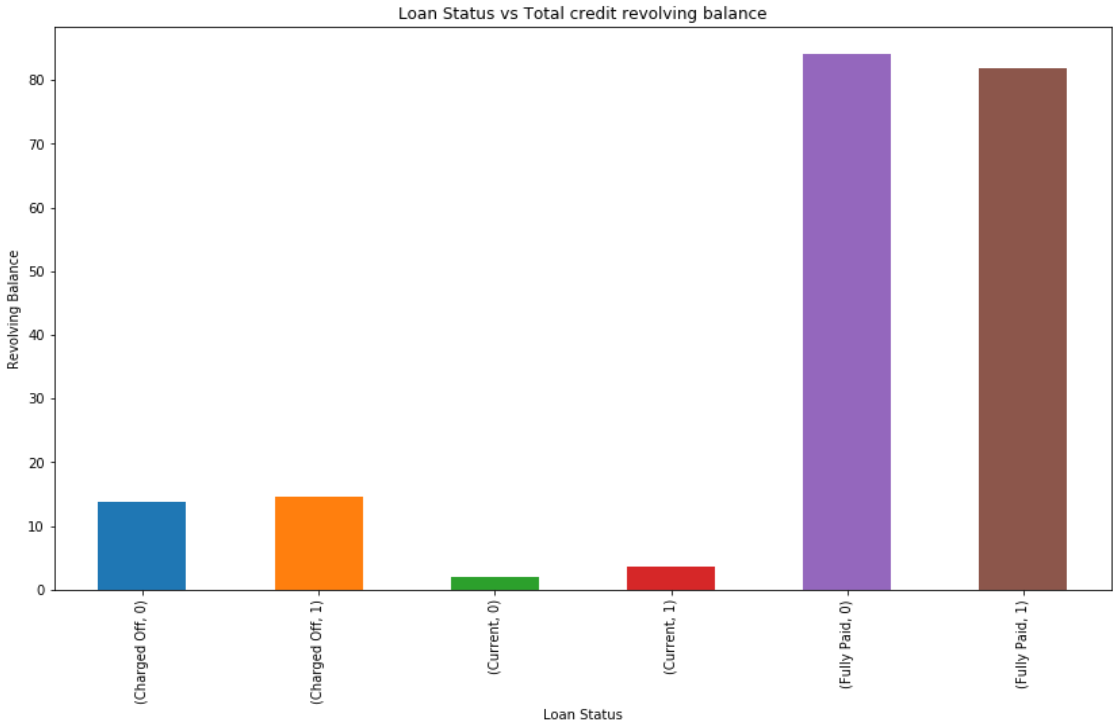


Plot 7:
Conclusion: Clearly, having a higher dti increases default by 10% vs the average of 14.17%.

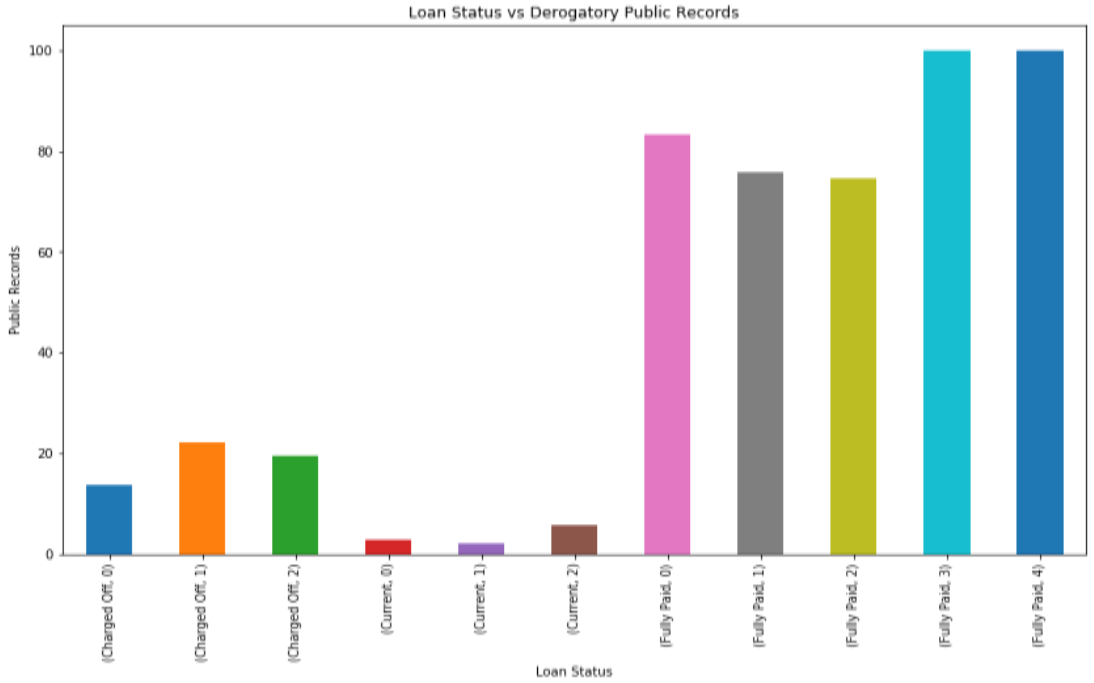


Plot 8:
Conclusion : In general, there seems to be a correlation between number of enquiries and defaults. More than 4 enquiries shows a significant increase vs. the average defaults of 14.17%. This is a red-flag for the bank's management. Whenever the credit enquiries go higher than 4, they need to tighten up processes as default risk is higher.

Bivariate Analysis-5



Plot 9:
Conclusion: There is only a marginal 3% difference vs. the average. This variable does not seem to have a strong influence on defaults.



Plot 10:
Conclusion : The fact is that if there is a derogatory public record, it significantly increases defaults. Both 1 and 2 records index much higher than the mean at 157 and 138% respectively.

Conclusion

A bivariate analysis of the customer attributes with other loan attributes was conducted to arrive at the below listed driver variables:

1. Term
2. int_rate
3. home_ownership
4.annual_inc
5.verification_status
6.purpose
7. dti
8. inq_last_6mths
9.pub_rec
10. addr_state