GRAMENER CASE STUDY

SUBMISSION

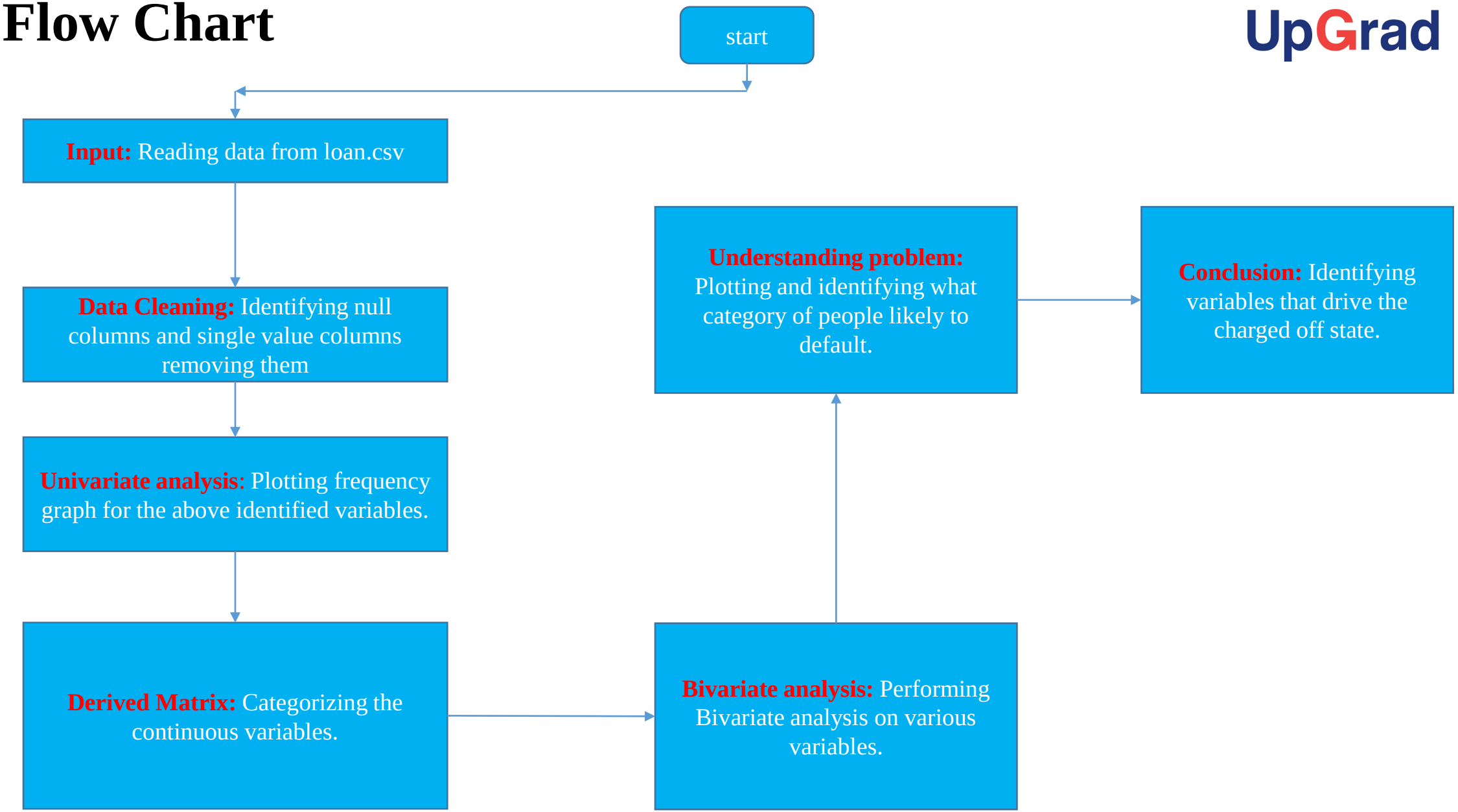
Group Name:

1. Ravi Biradar
2. Tanmaya Ray
3. Ashish Hota
4. SivaRamaKrishnaReddy. V

Objective

- The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default.
- Identify If the applicant is **likely to repay the loan.**
- Identify If the applicant is **likely to not repay the loan.**

Flow Chart

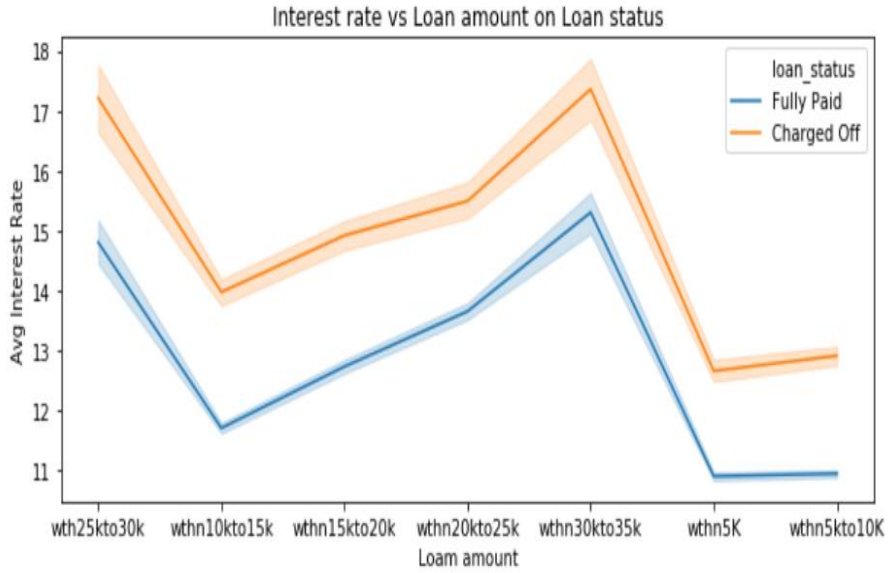

start
Input: Reading data from loan.csv
Data Cleaning: Identifying null columns and single value columns removing them
Univariate analysis: Plotting frequency graph for the above identified variables.
Derived Matrix: Categorizing the continuous variables.
Bivariate analysis: Performing Bivariate analysis on various variables.
Understanding problem: Plotting and identifying what category of people likely to default.
Conclusion: Identifying variables that drive the charged off state.

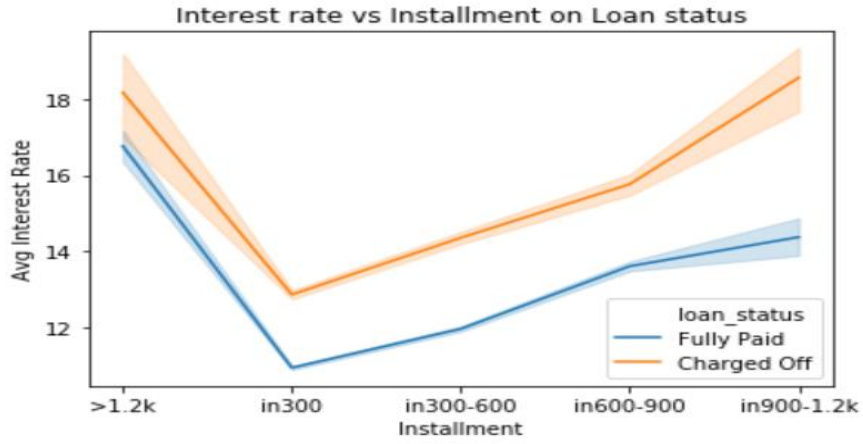
Assumptions made for Case Study

- Columns with all the null values have no impact .So we remove such columns
- Columns with only one value also has no impact as if we try to combine with any other parameter to do bivariate analysis as one of the column has single value , it would be like doing univariate analysis on other column . So we remove such columns.
- Columns with date value like Issue date , Next payment date , last payment date and last credit pull date have no impact to loan status.

Bivariate Analysis

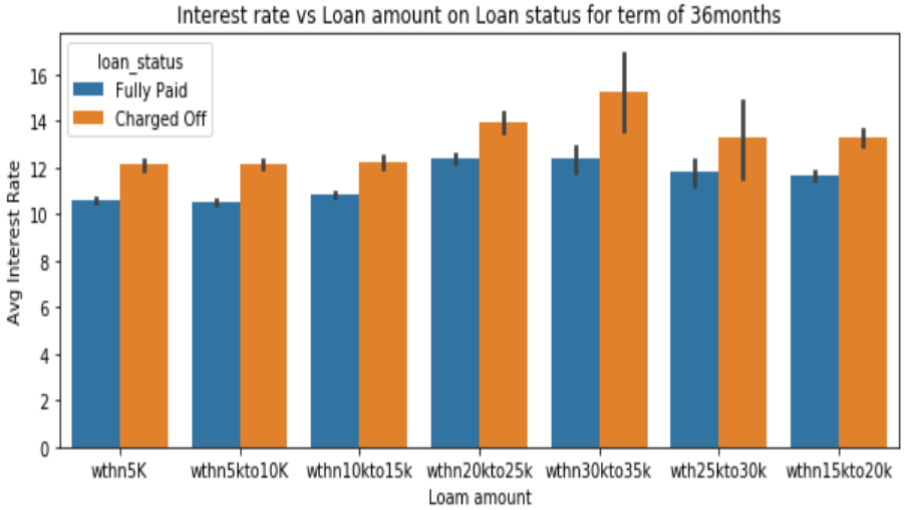
- From the graph 1 , If the avg interest rate high for the loan amount, the installment is also high which increases the chances of Charged off.
- From graph 2,With smaller interest rate say for loan within 5k if the avg interest rate within 11 then the people likely to pay .

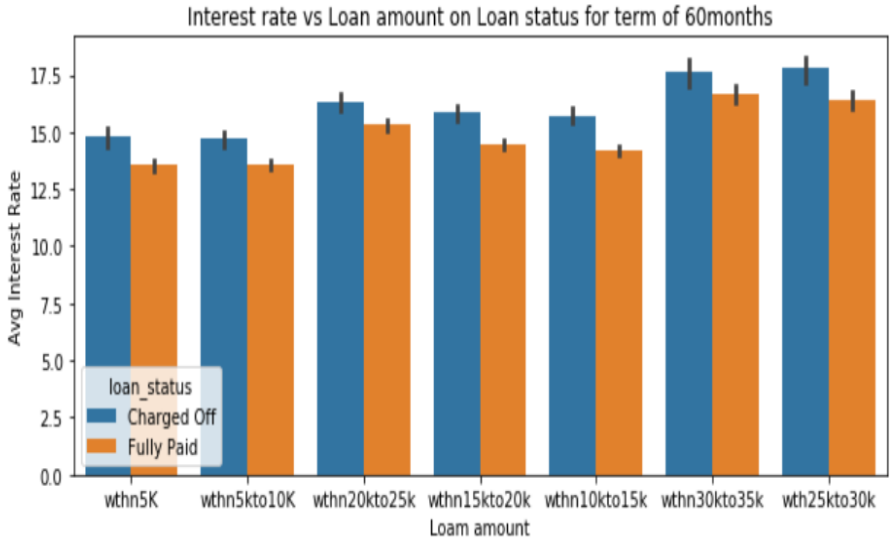




Bivariate Analysis

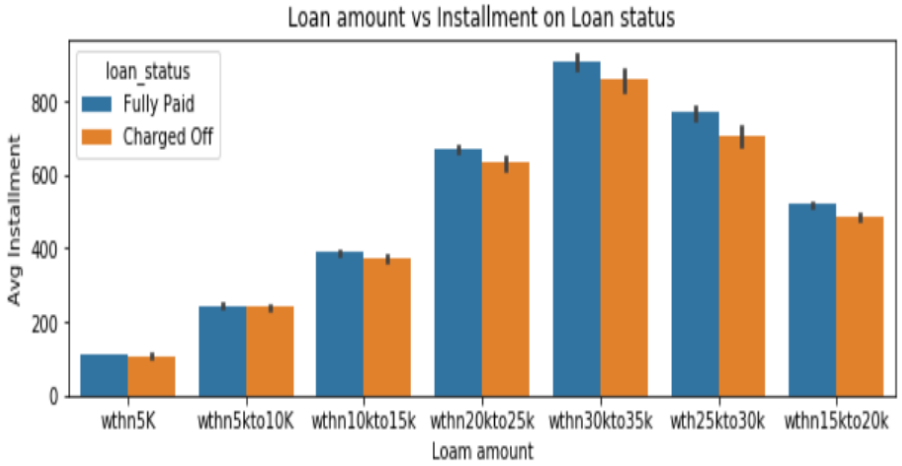
- From the graph 1 , If the avg interest rate for 36 months starts from 10% as the loan amount increases it becomes had to pay.
- From graph 2,For 60months term avg int rate is from 12.5 and goes high with the loan amount.

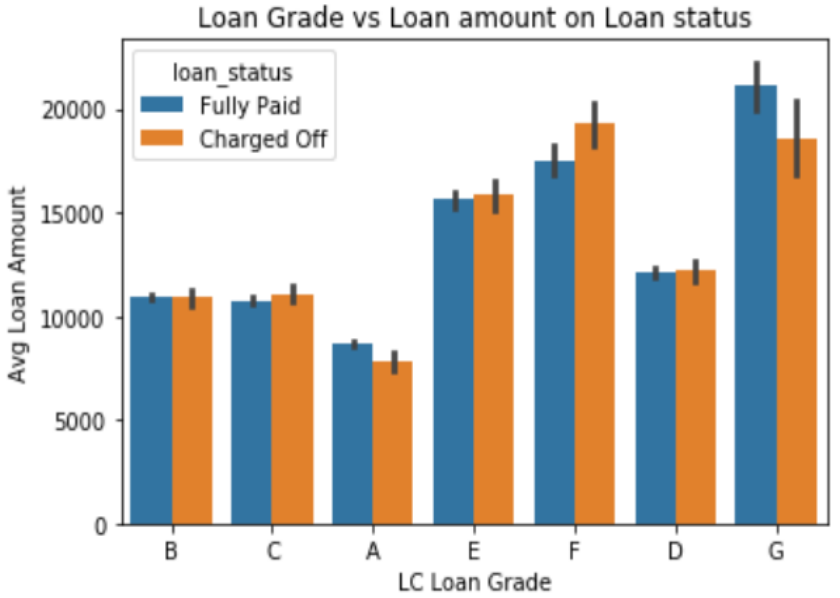




Bivariate Analysis

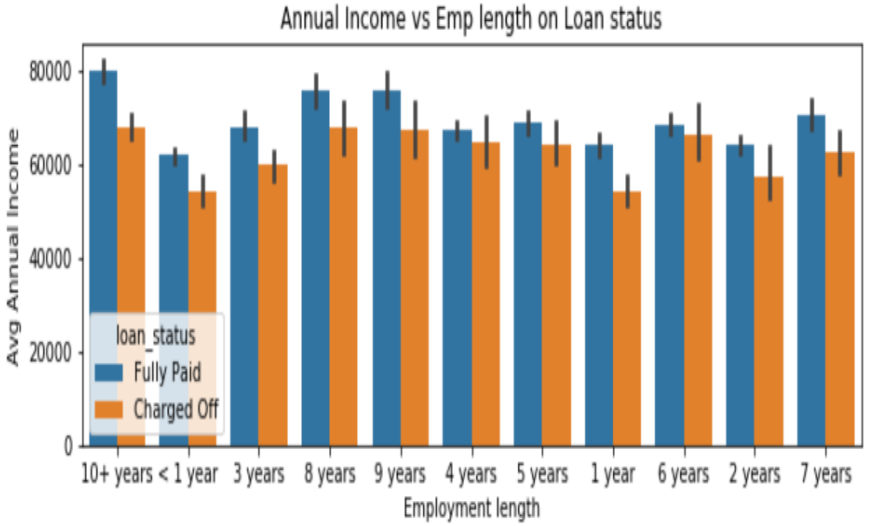
- From the graph1 ,The avg installment on loan has less impact compared to installment with interest rate.
- From graph 2,Despite the more no of loans by Grade A it has the least charge off.

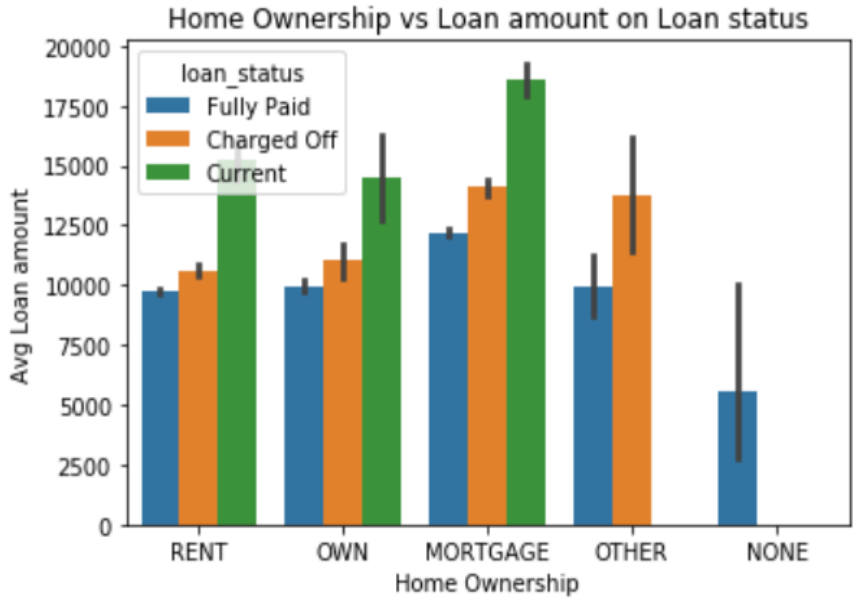




Bivariate Analysis

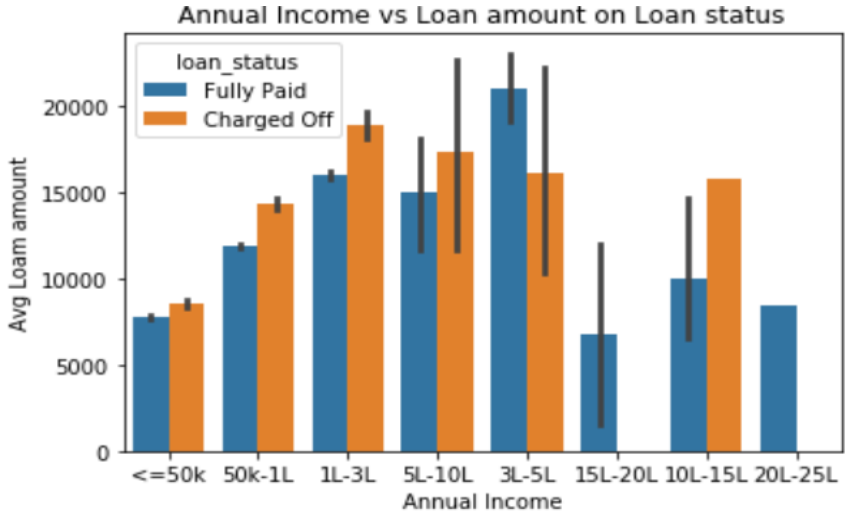
- From the graph1,If the annual income is more with the experience they tend to pay the loan else the chances of default are high.
- From graph 2,Despite the more no of loans taken by rent and mortgage home owners,the chances of charged off are high for own and mortgage

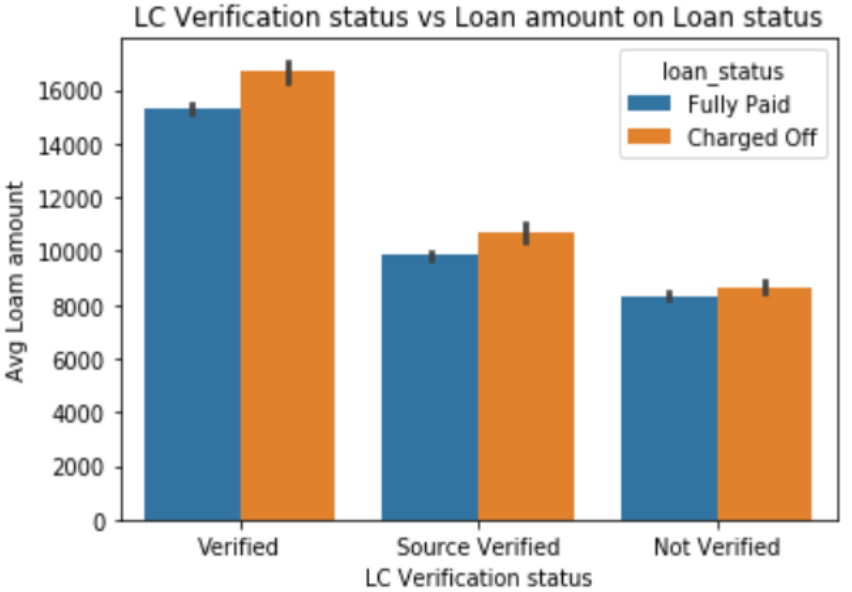




Bivariate Analysis

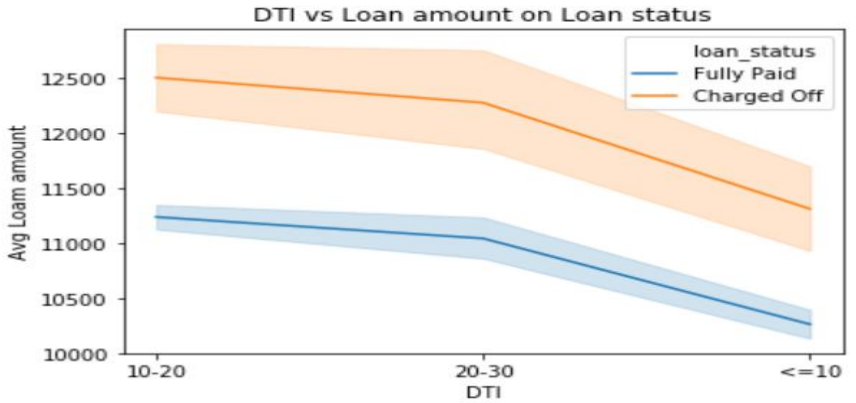
- From the graph1,If the annual income is less and take higher loan the chances of default are high.
- From graph 2,Verified have higher chances of default,compared to other categories.

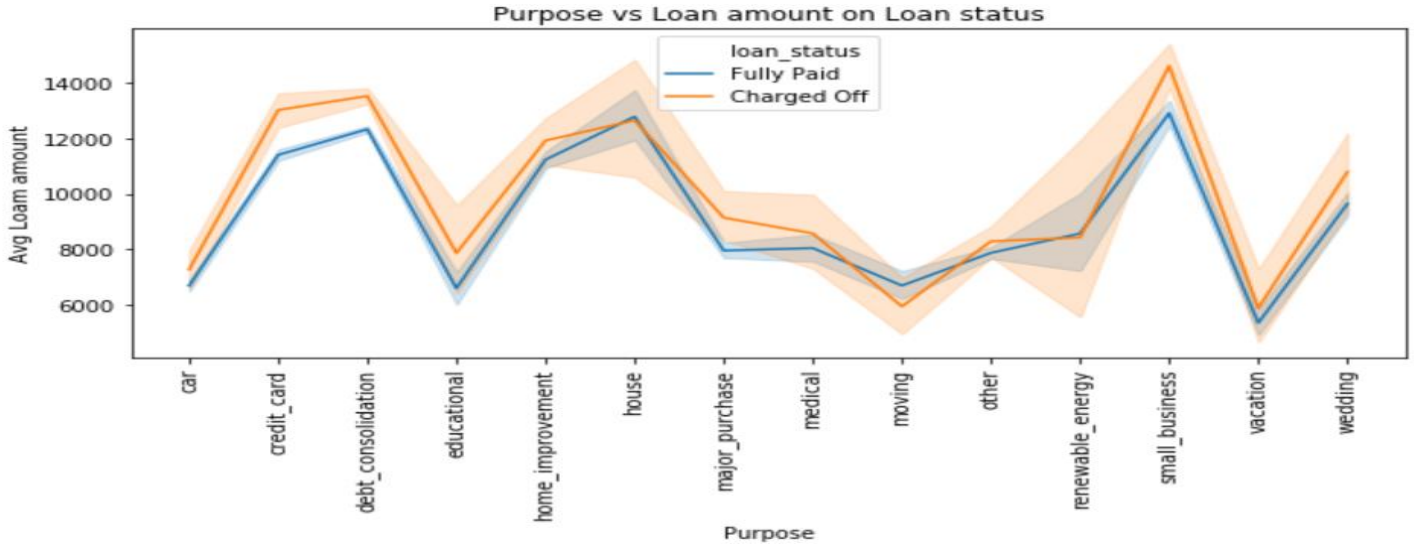




Bivariate Analysis

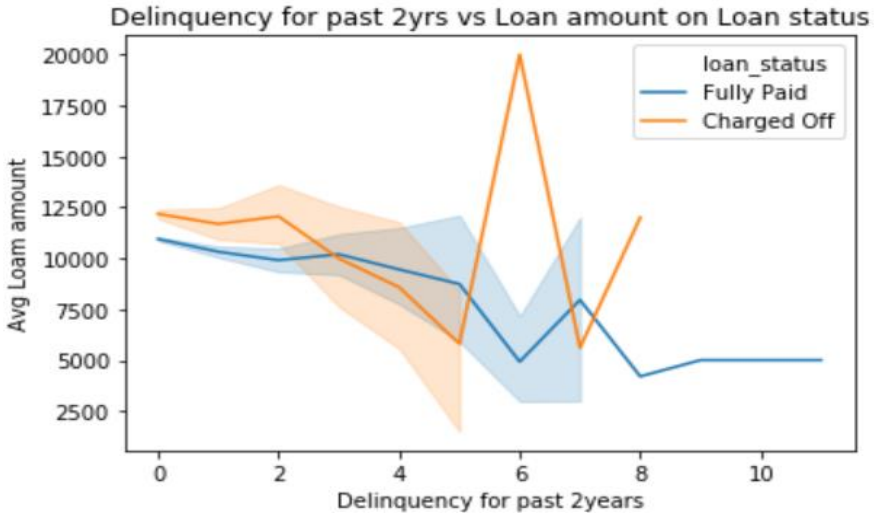
- From the graph1 for DTI,DTI score of 10-20 have more chances of default.
- From graph 2 for purpose,If the loan amount is within the limit,debt consolidation and credit card default can be avoided as those purpse serve the most for loan takers.

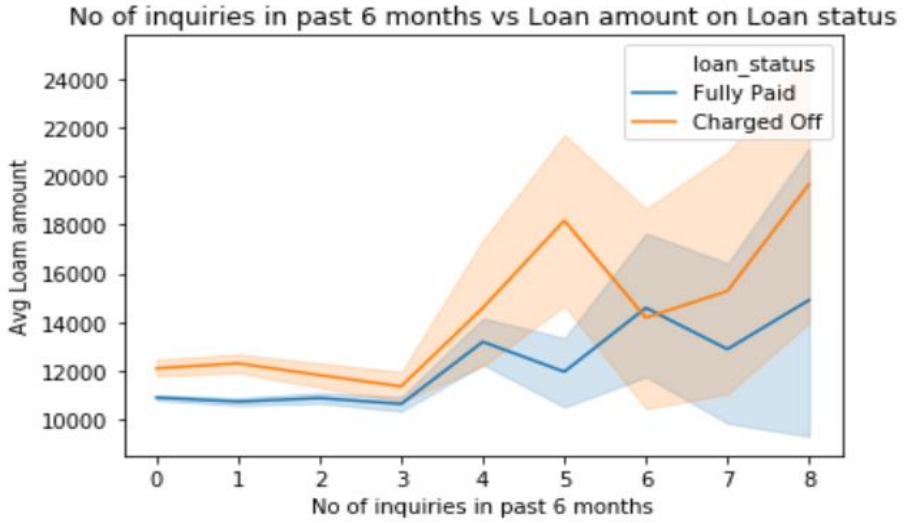




Bivariate Analysis

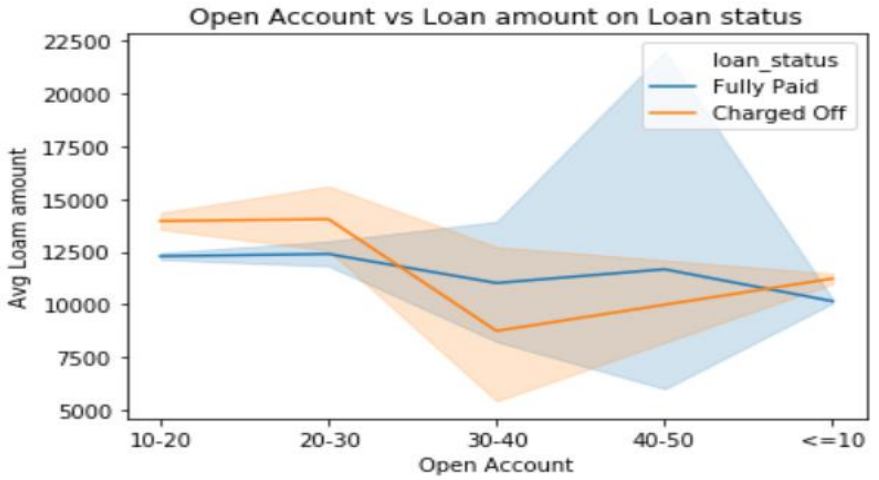
- From the graph1,Higher the delinquency higher the default chances.
- From graph 2,Higher the inquiry higher the charged off status.

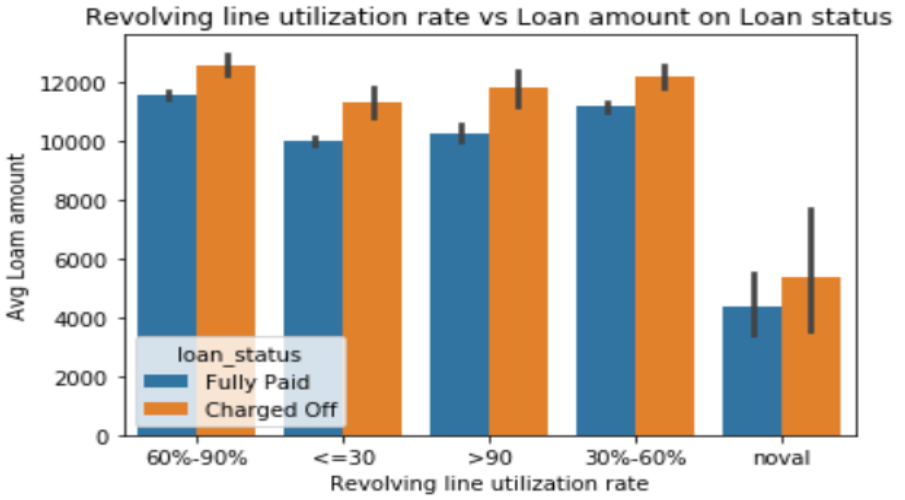




Bivariate Analysis

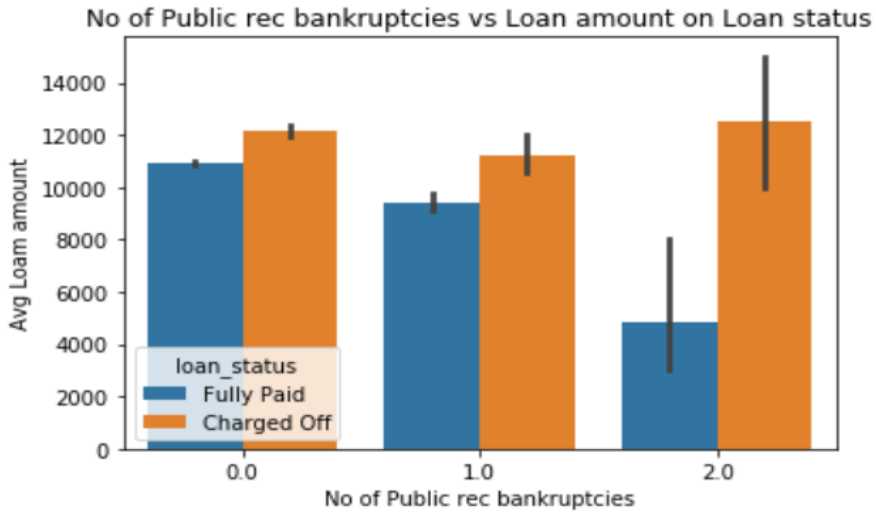
- From the graph1,People with open account <=20 take more loans but if the loan amount exceeds certain value the chances of default is high.
- From graph 2,accepted revol util is <=30% but with higher values the chances of default are higher.

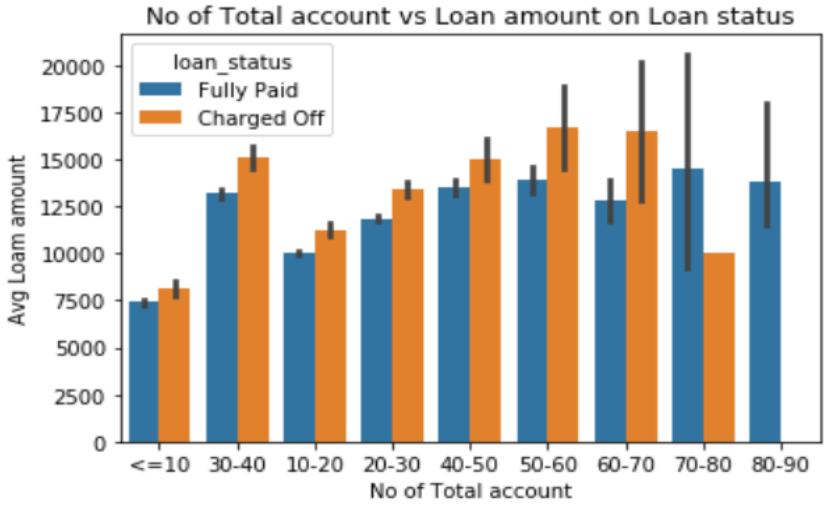




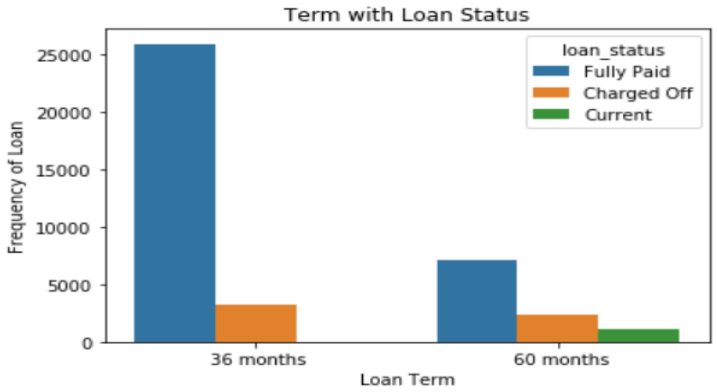
Bivariate Analysis

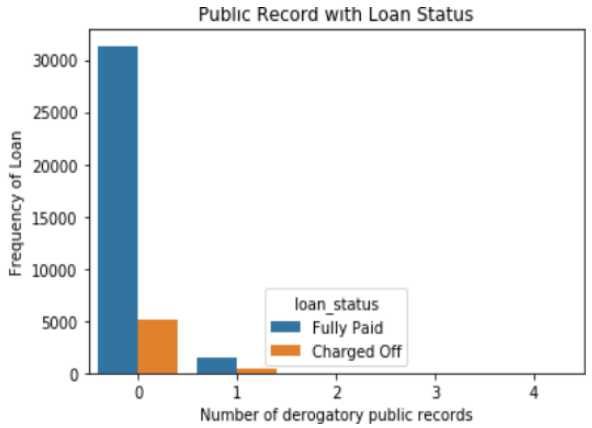
- From the graph1,People with 0 and more than 1 bankruptcies tend to default rather than 1.
- From graph 2,Higher the total no of accounts higher the chances of default.

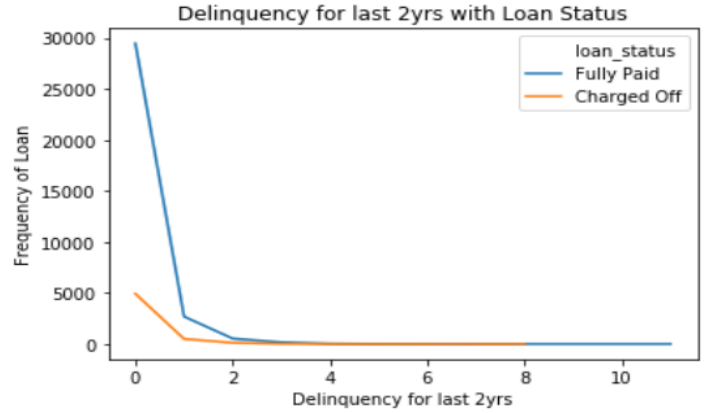


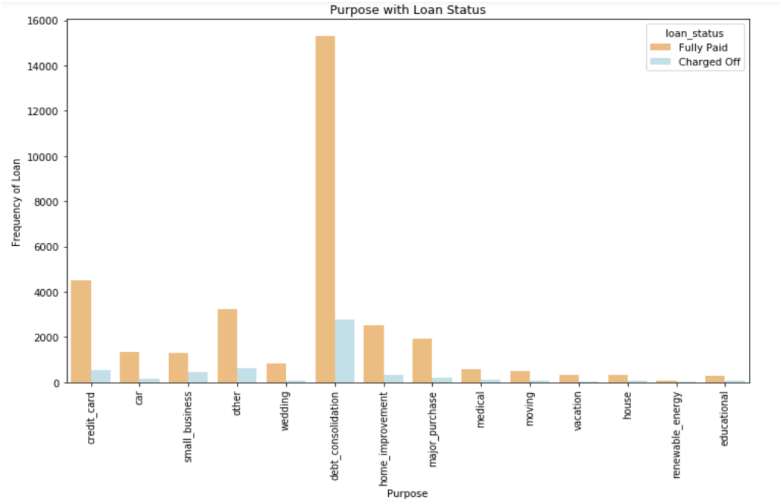


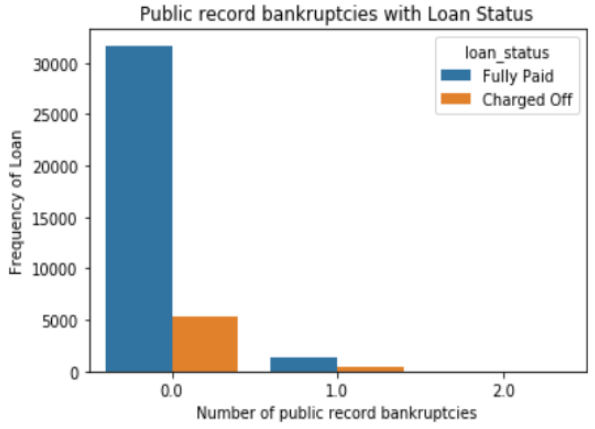
Results of Univariate Analysis

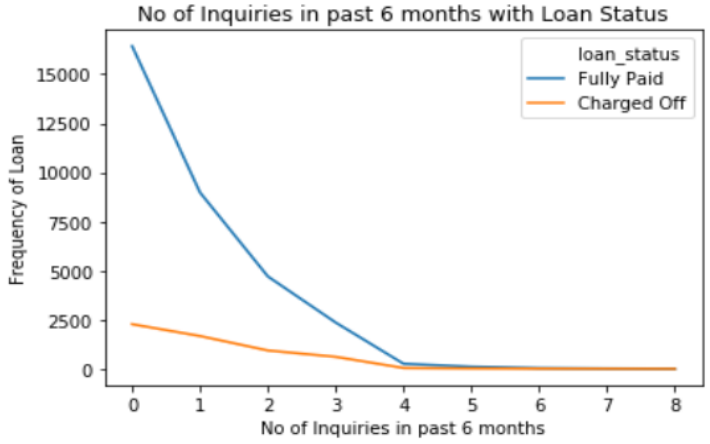




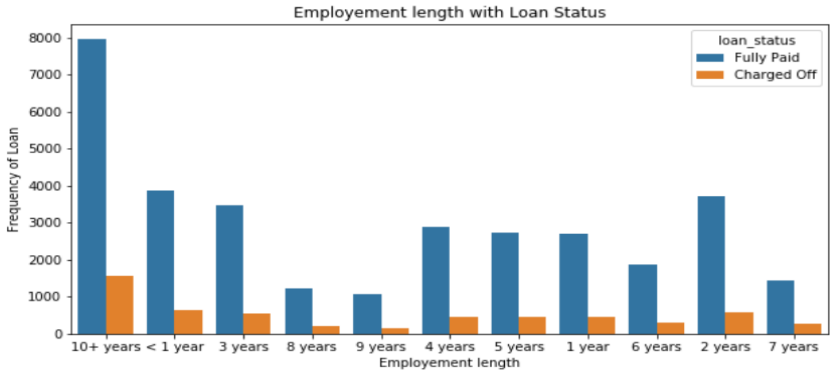


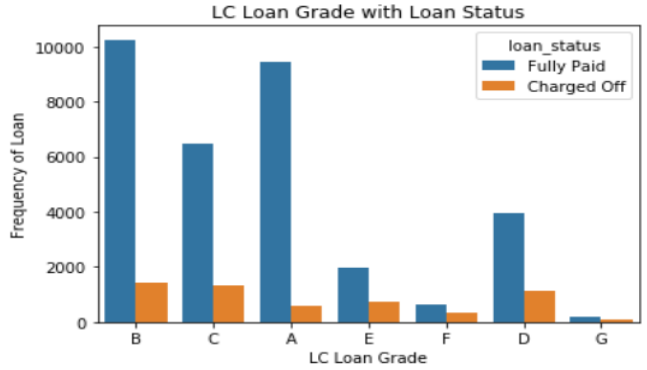


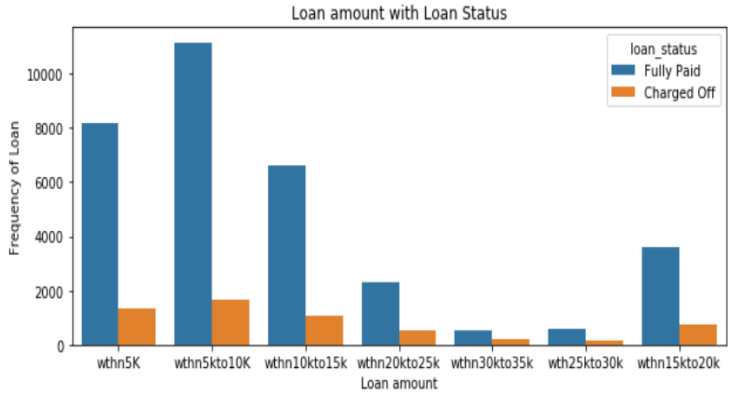


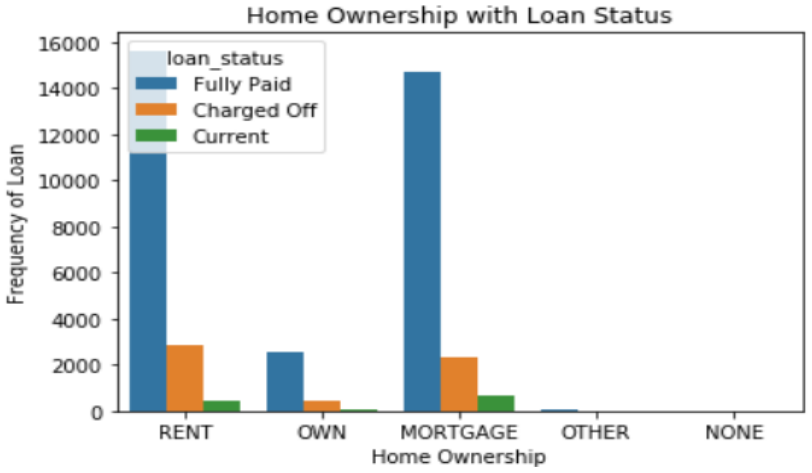


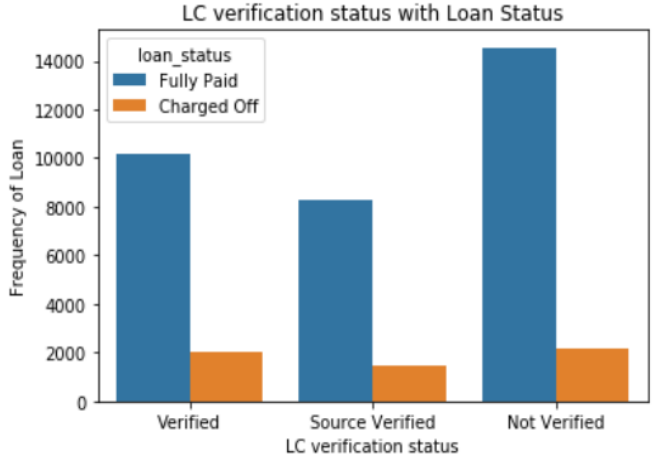
Results of Univariate Analysis

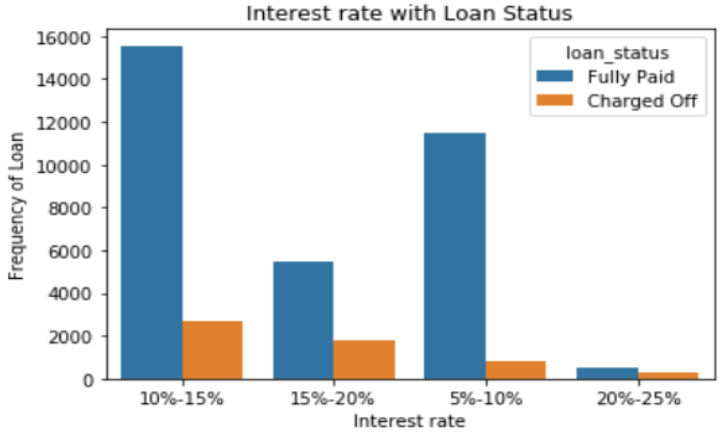




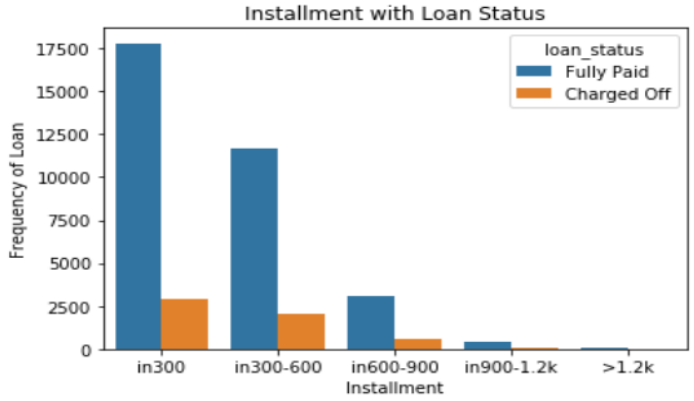


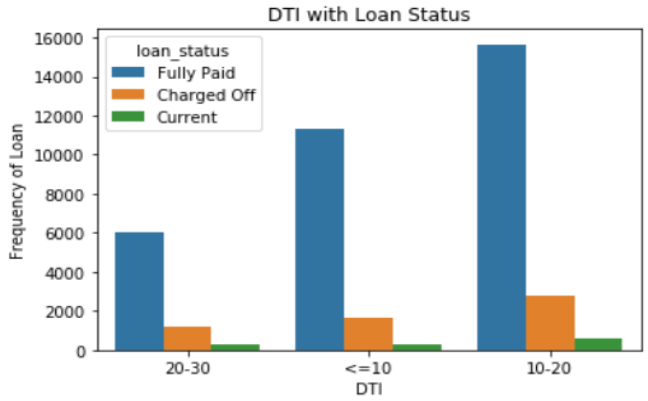


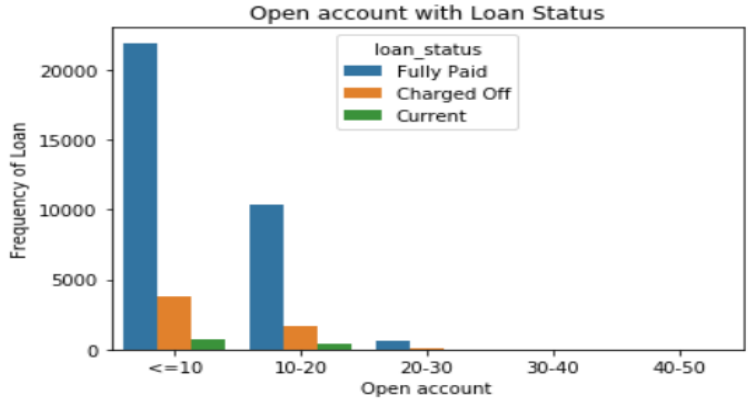


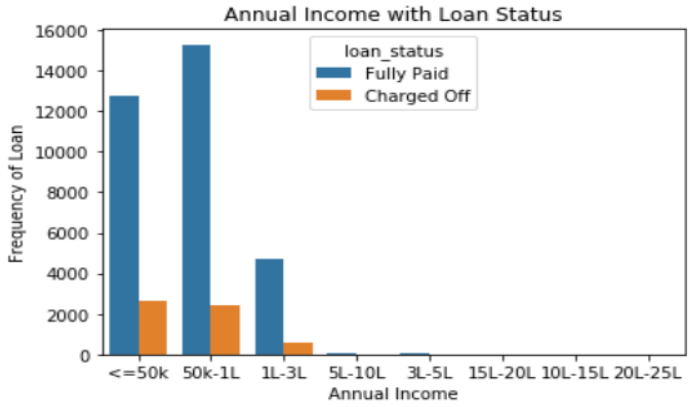


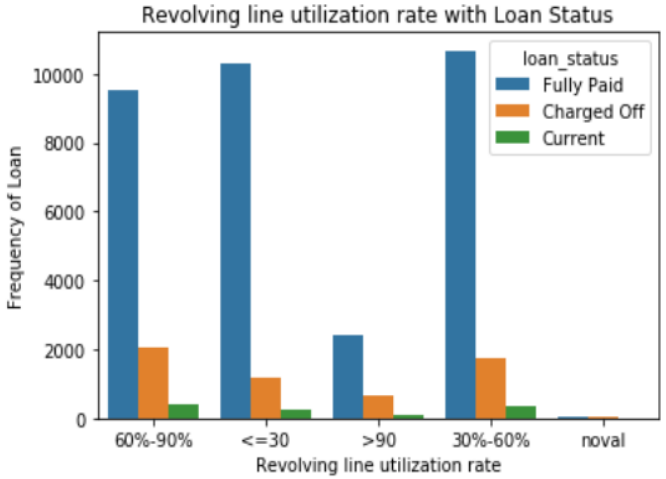
Results of Univariate Analysis

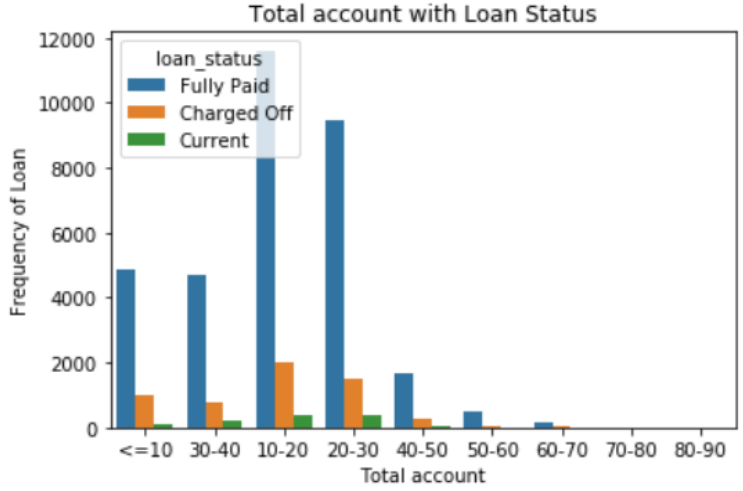












Conclusion

We have identified the driving factors which effect the loan status. From the previous graphs we are able to see clearly what category of people are more likely to pay the loan and who are likely to go default.

We even see that people with extreme lower and higher values for the listed variables in the graph tend to go default.