LENDING CLUB ASSIGNMENT

SUBMISSION

Group Members:

Medha Pundir

Saumya Bisht

INTRODUCTION

LendingClub , the leading marketplace lender in the US, reported first-quarter results today. According to the online lender, applications grew 31% with originations hitting $2.7 billion. Importantly, 73% of customers were approved within the first 24 hours – a necessary metric in the digital finance era.

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

We are trying to understand the most dominant factors contributing to future credit risk default by it's customers.

PROBLEM SOLVING APPROACH

Data Understanding

- Firstly we will try to understand the data provided
- Understand the significance of each column,datatype,columns for numerical data.

Data Cleaning

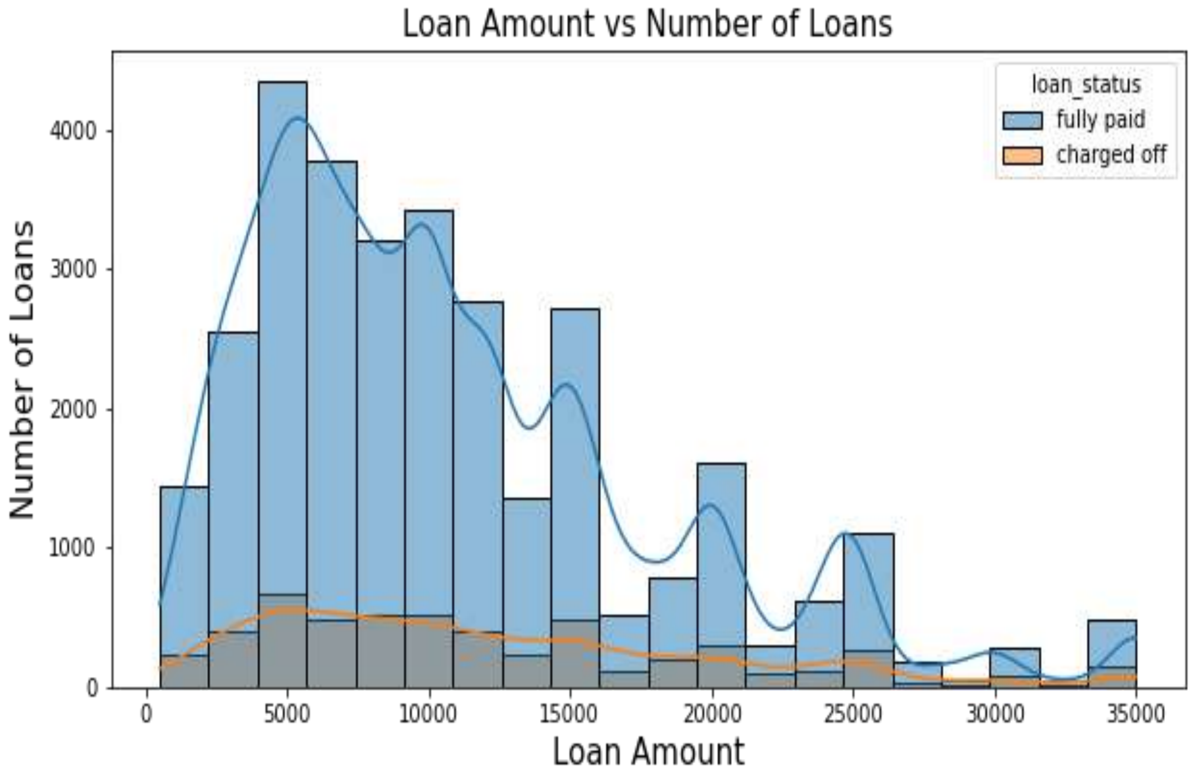
- Identify the missing values
- Identify the unique values
- Remove the high missing percentage.
- Removed the null values.
- Stripped the % and years from columns
- Drop the rows where the missing percentage is quite high .
- Create new columns if required for identification of new metric .
- Derived two columns issue_year,issue_month from issue_d.
- Derived default_rate column from loan_status i.e assuming 0 for fully paid and 1 for charged off.

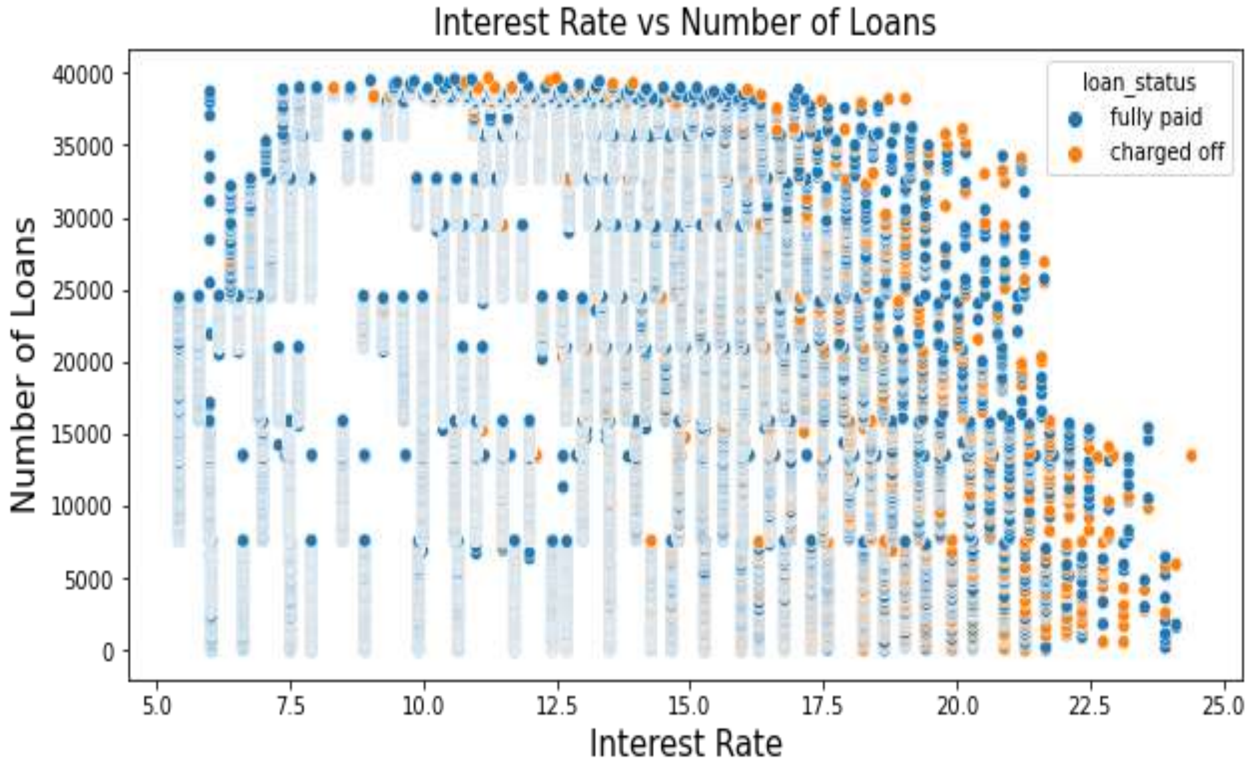
Data Analysis

- Discover the target variables leading to high percentage default s .
- Determine the combination of factors contributing to higher credit risk.

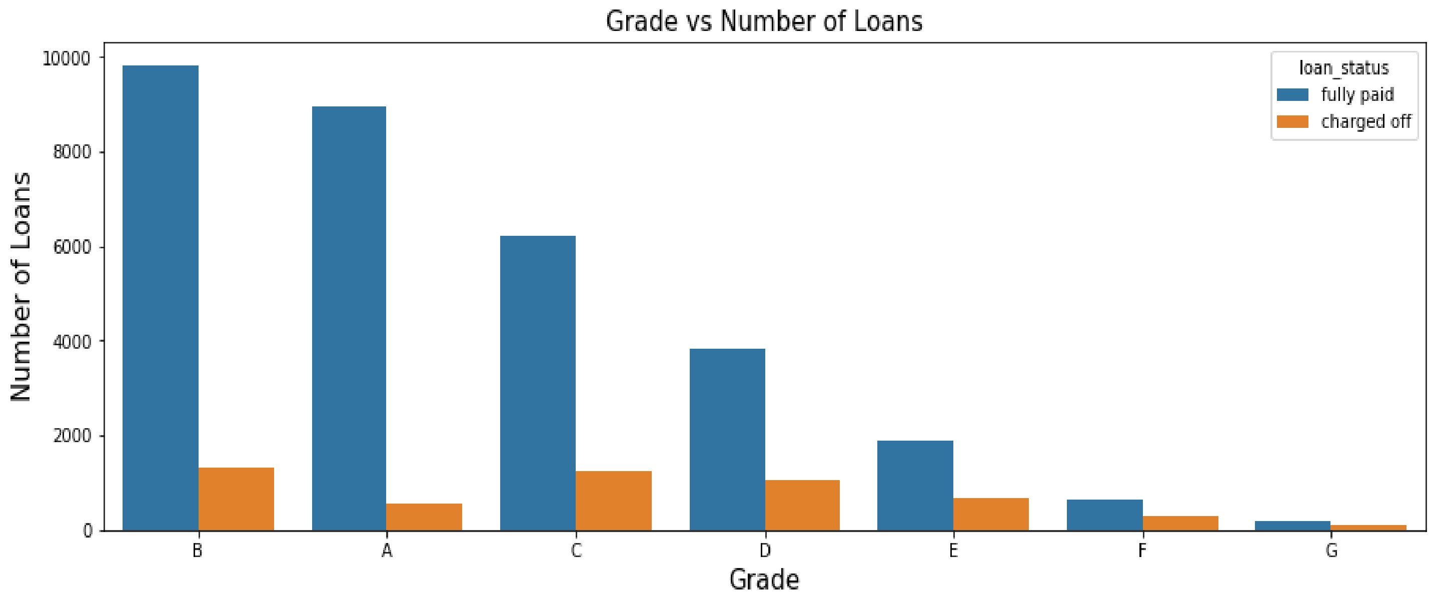
UNIVARIATE ANALYSIS

- Here we are trying to understand how loan status varies with Loan Amount and Interest rate charged.
- No of defaults increase ,when loan amount increases and the same observation holds for interest rate to default rate.

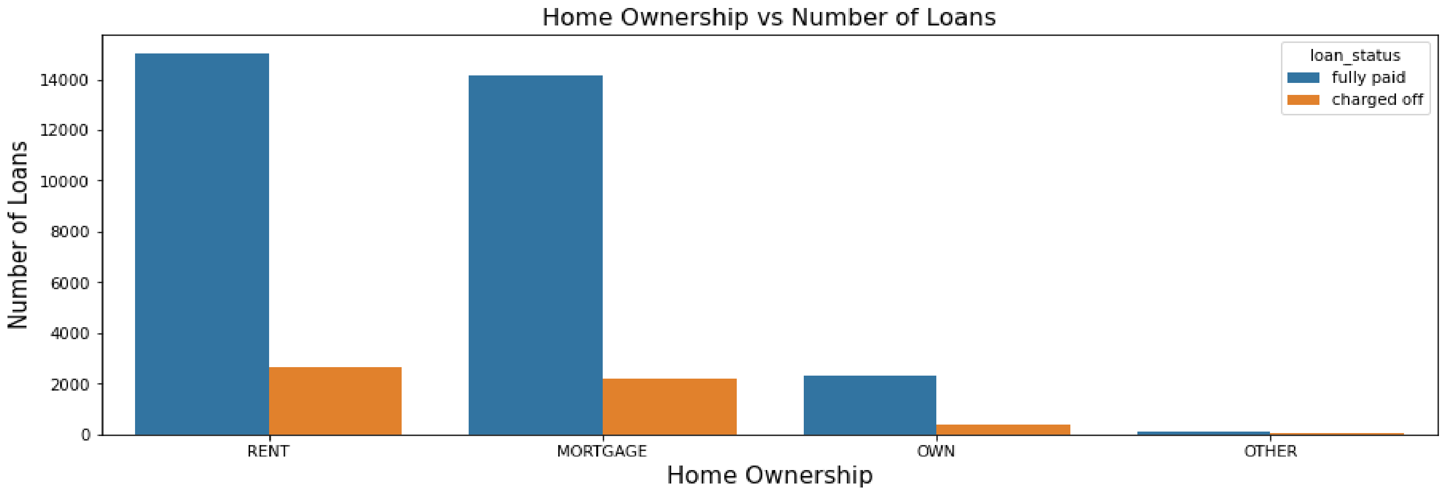




UNIVARIATE ANALYSIS

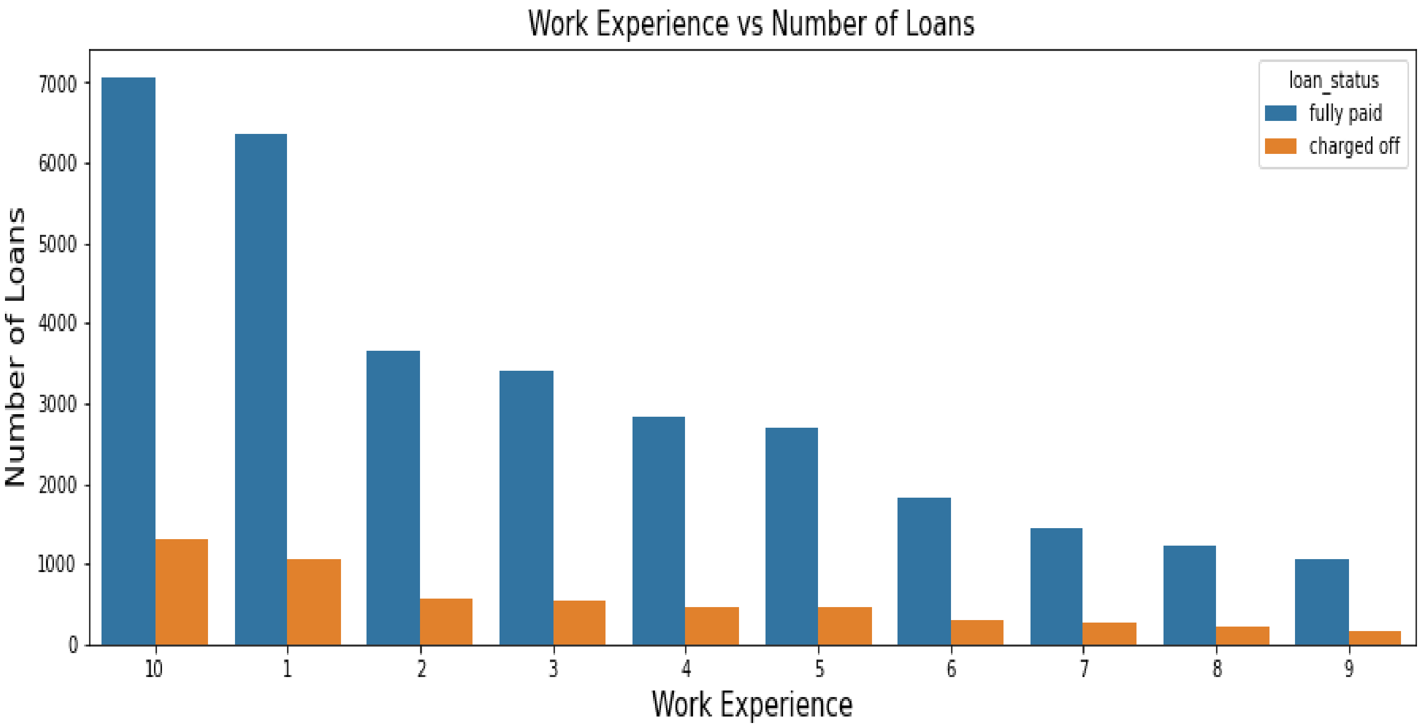


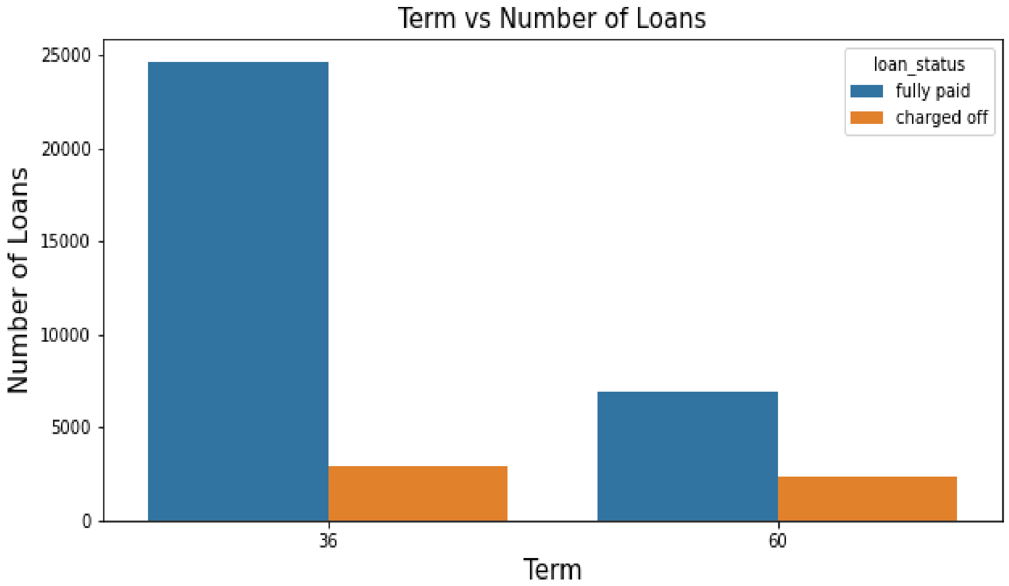
Here we can see that no of loan taken by people.
We can most people having lower grading tends to default on their loan



Here we can see default rate to home ownership.
Most Loans are taken by people who are on rent.People who don't have their own home ,have higher default rate

UNIVARIATE ANALYSIS

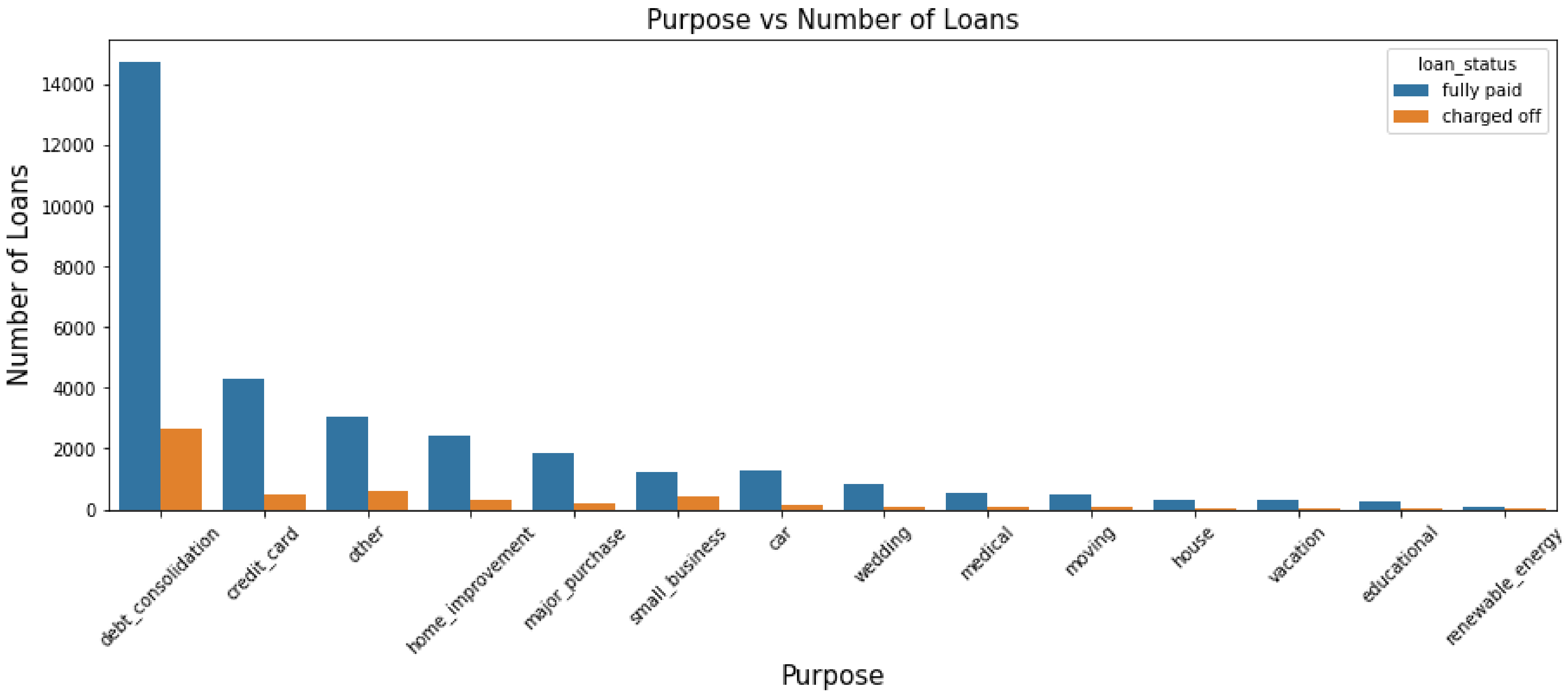




Most loans paid and defaults are mostly prevalent for 10 years experience and starters .

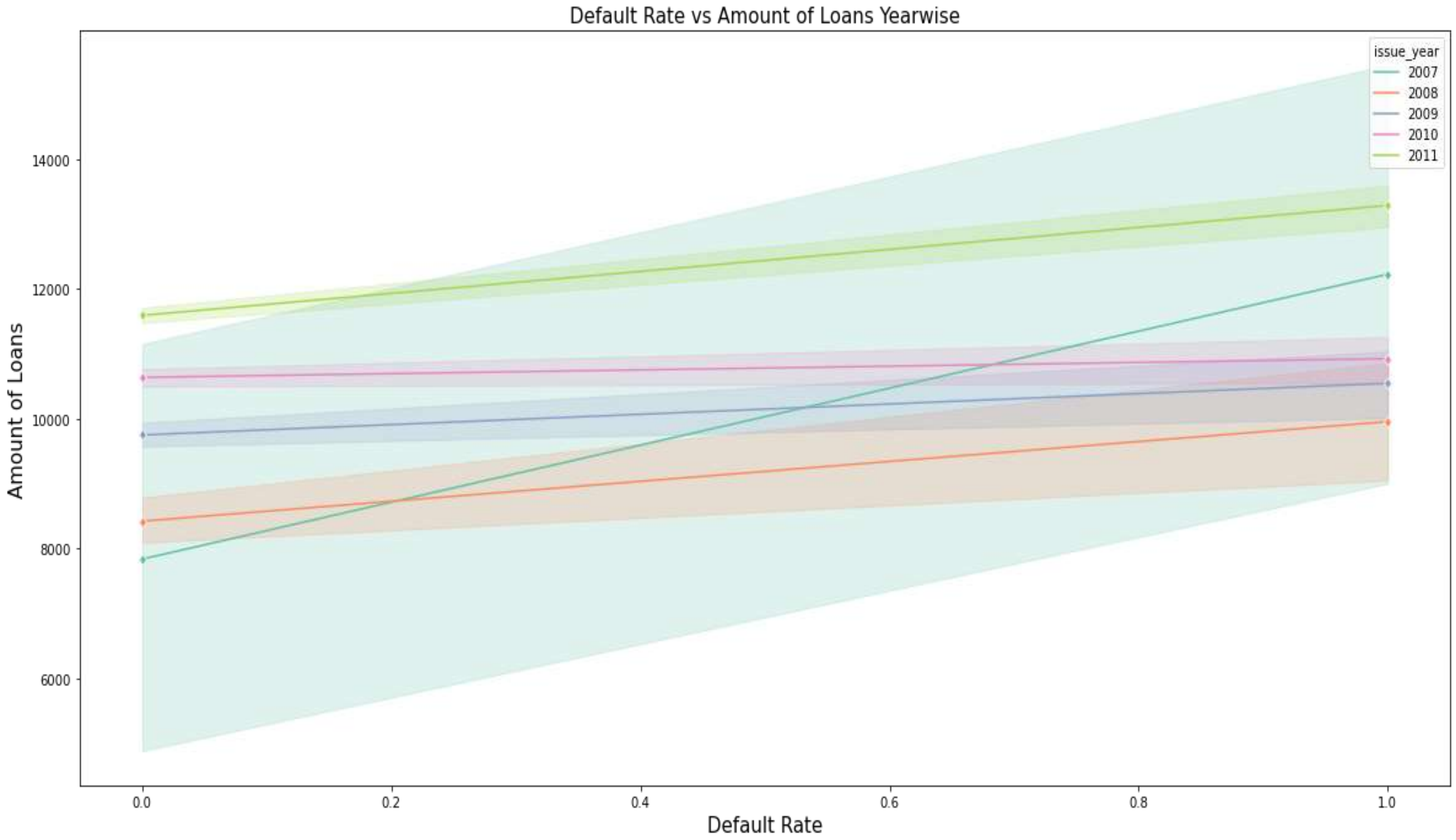
As we can see 36 months is the most preferred term for loan .

UNIVARIATE ANALYSIS



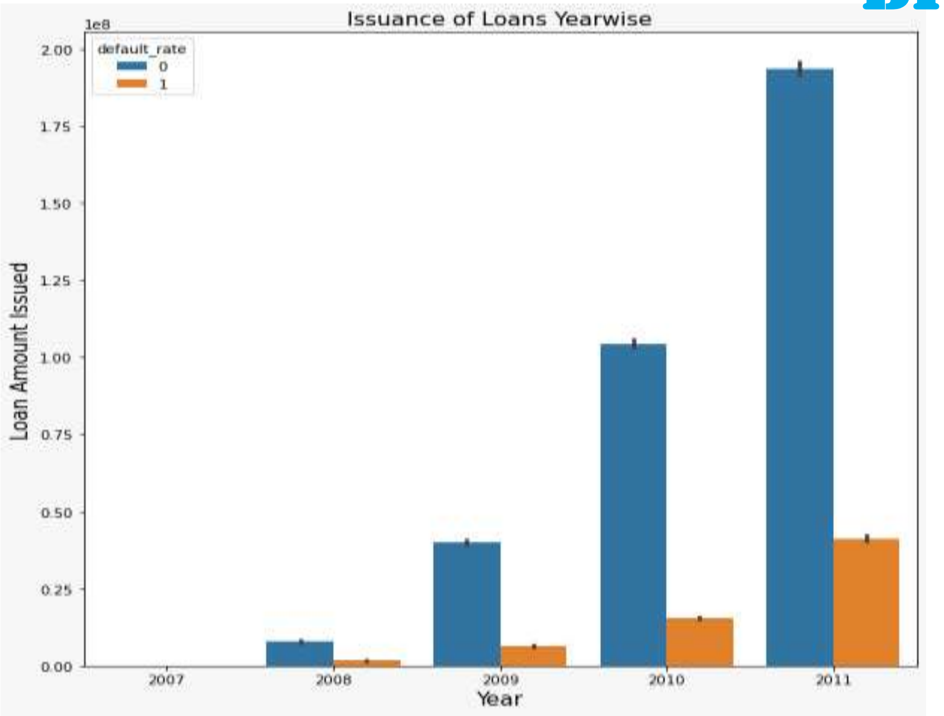
For most borrowers the purpose of loan was for debt consolidation and is clearly evident that most defaulters lie in the same category.

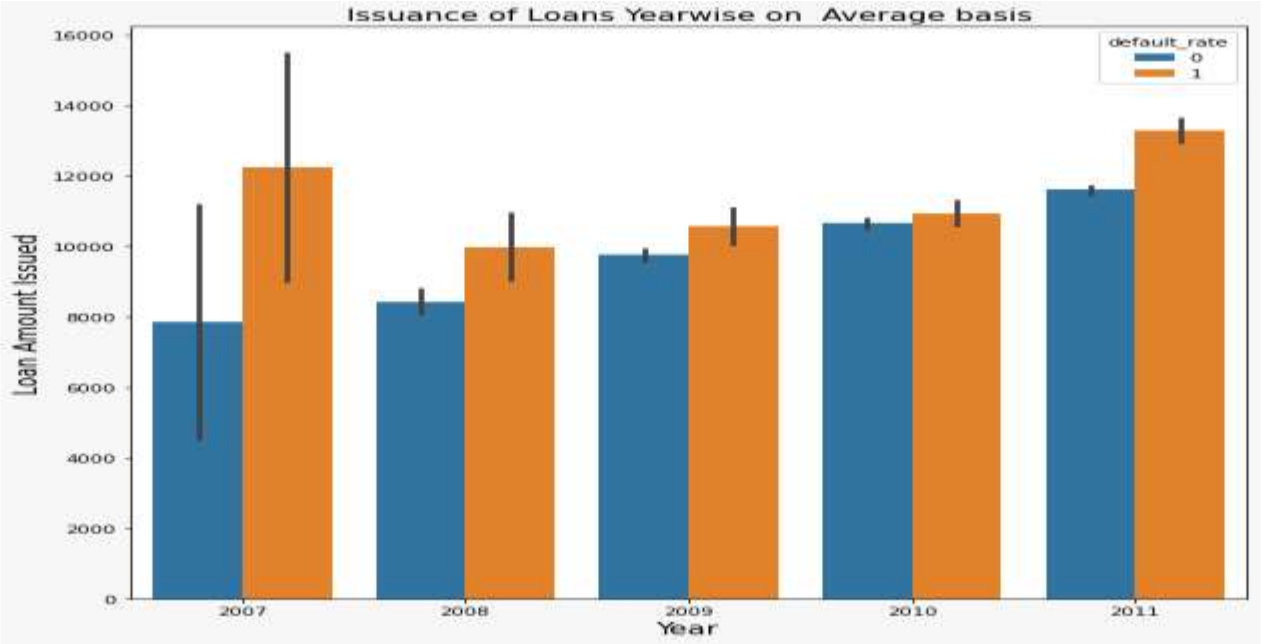
BIVARIATE ANALYSIS



Through this graph ,we can see a variation between defaulters and non defaulters for the year 2007 and this could be due to economic crisis of 2007. Same pattern can be observed for the year 2011 ,which could be a indictator of Financial crisis of 2011.

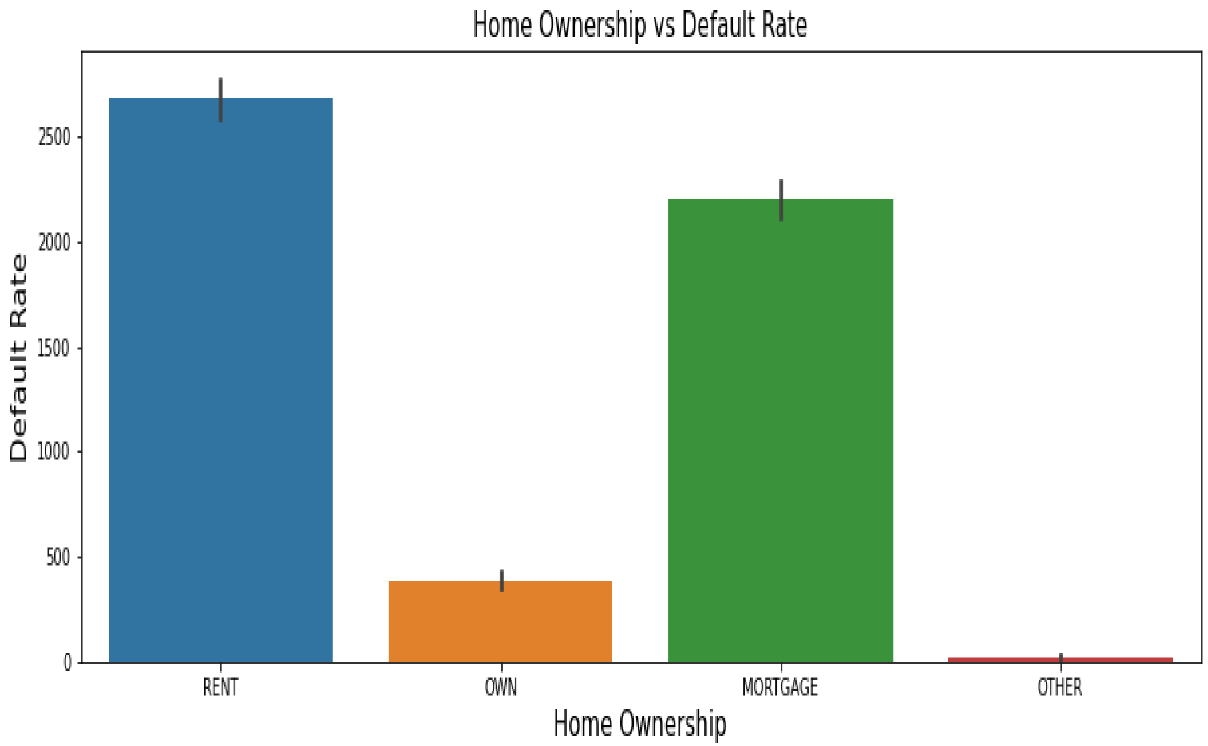
BIVARIATE ANALYSIS

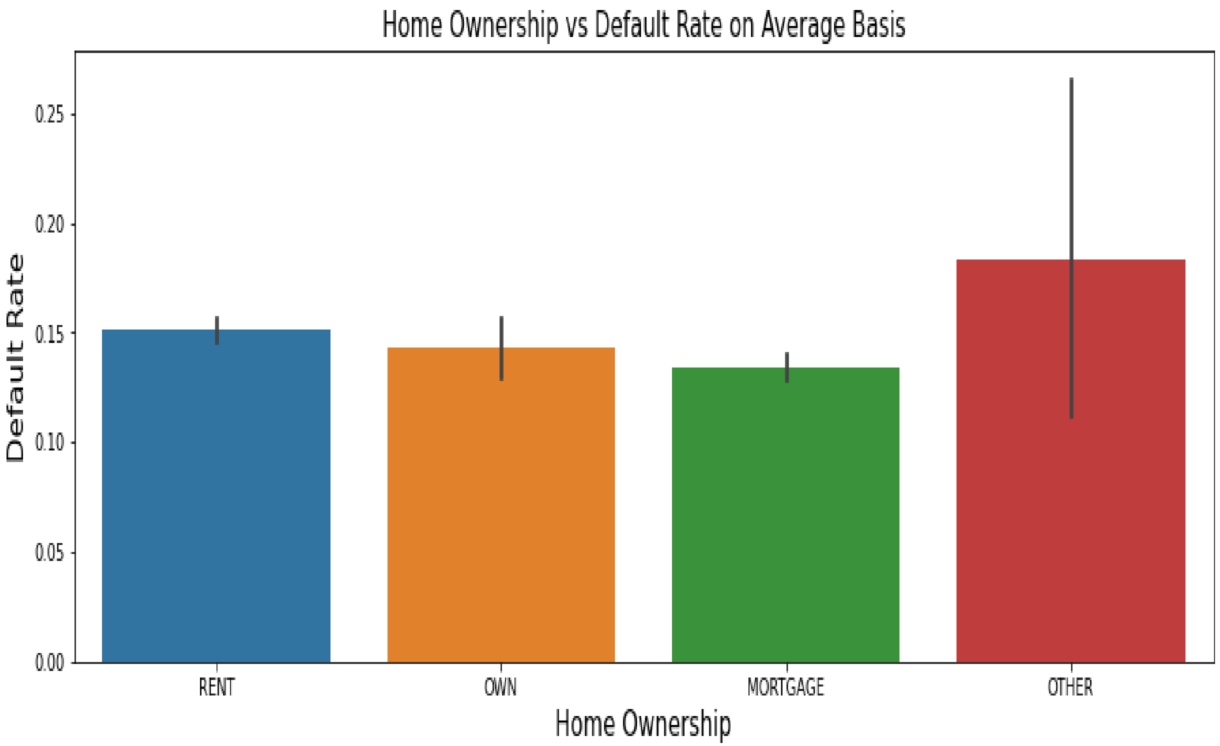




On the left we can see that loan amount is increasing yearly whereas in the right on an average default rate is high for the year 2007 ,then stable for the years 2008 to 2010 and again increased for the year 2011.This variation once again concludes that economic crisis could be a triggering factor for most defaults.

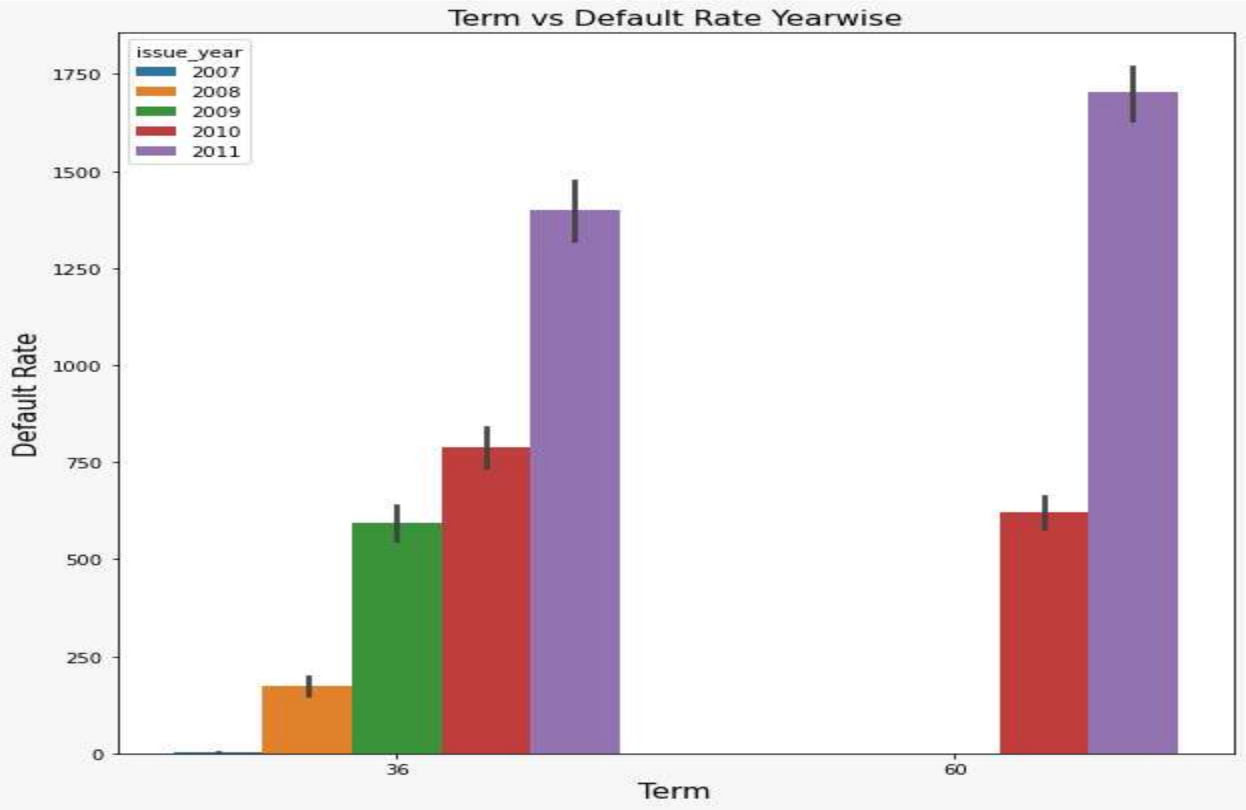
BIVARIATE ANALYSIS

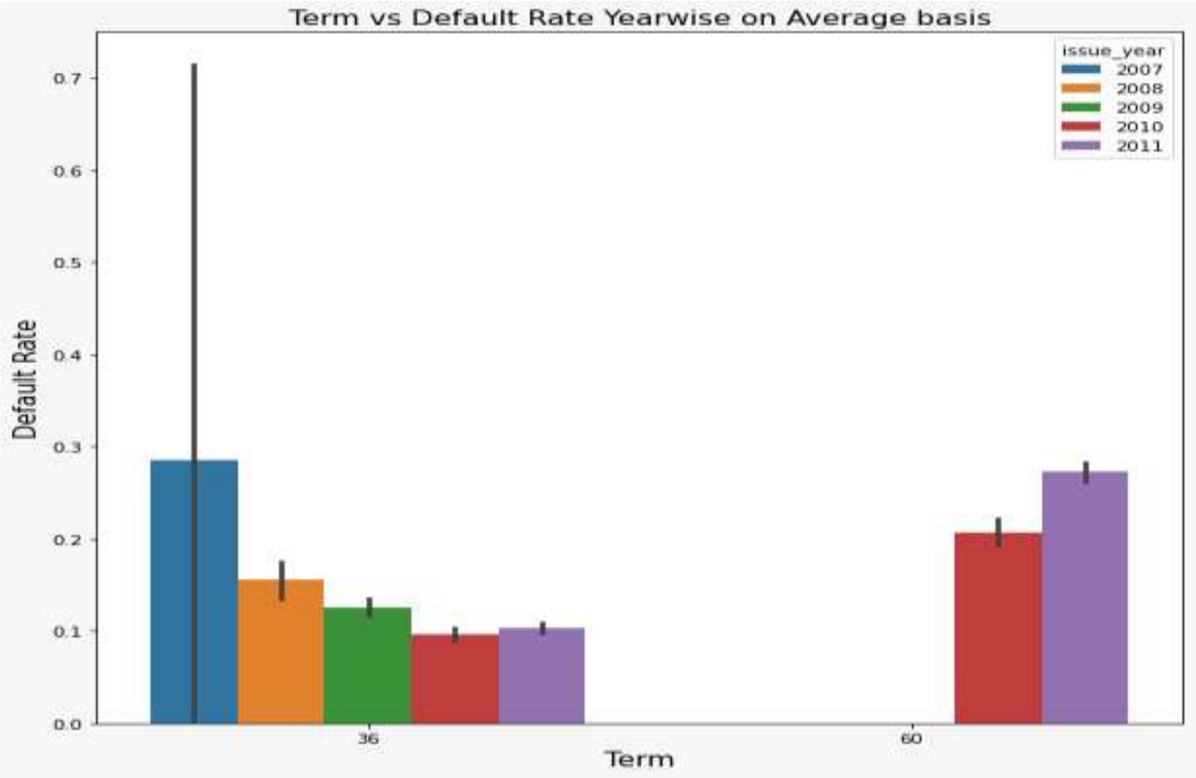




From the left we can see that overall rent and mortgage have the higher default rate but when we see the average picture ,Other is dominant .

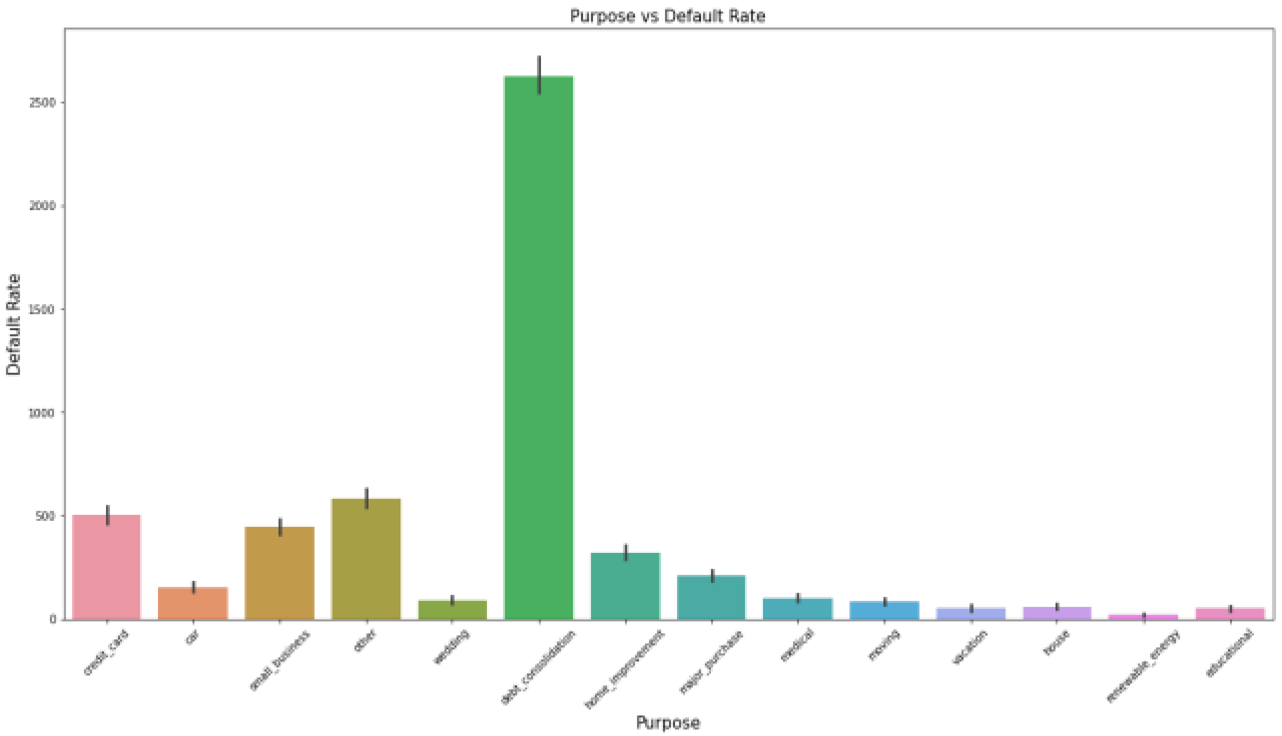
BIVARIATE ANALYSIS

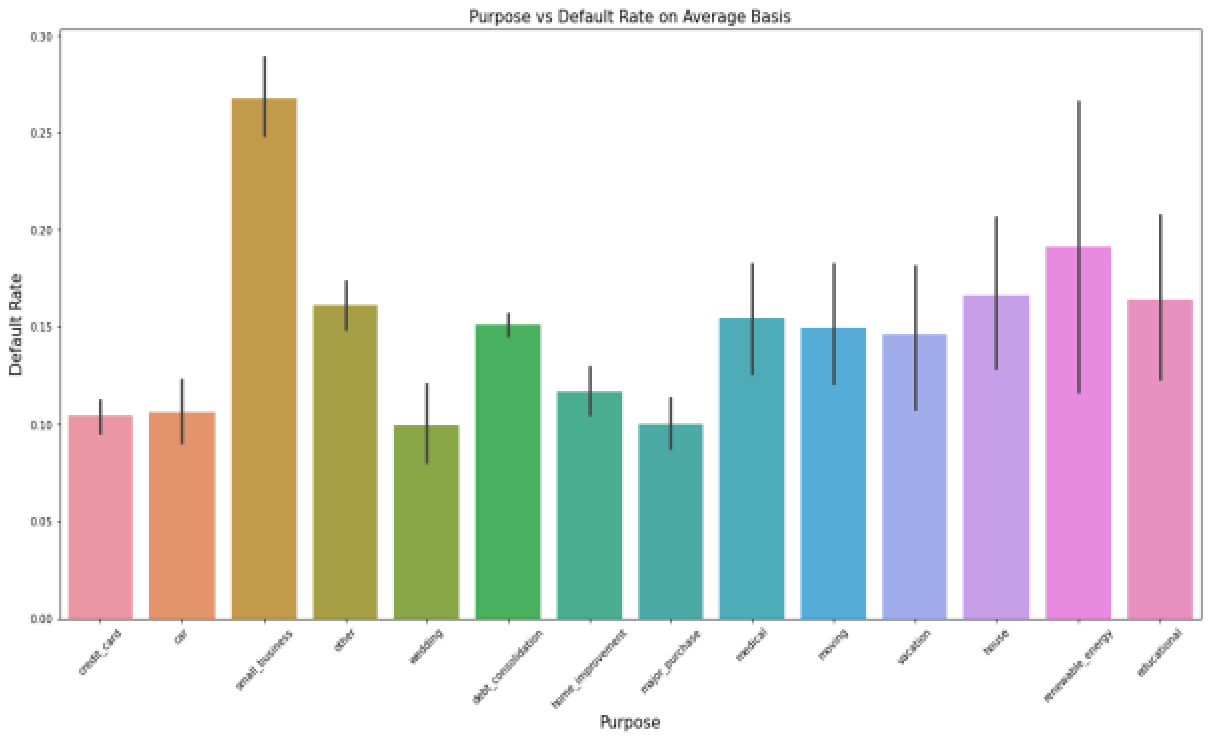




On the left we can for both terms default rate is increasing on yearly basis but On the right for the term of 36 months there is a anomaly against the year 2007,and decreasing till year 2011 and for the term of 60 months the default rate is on the higher side .

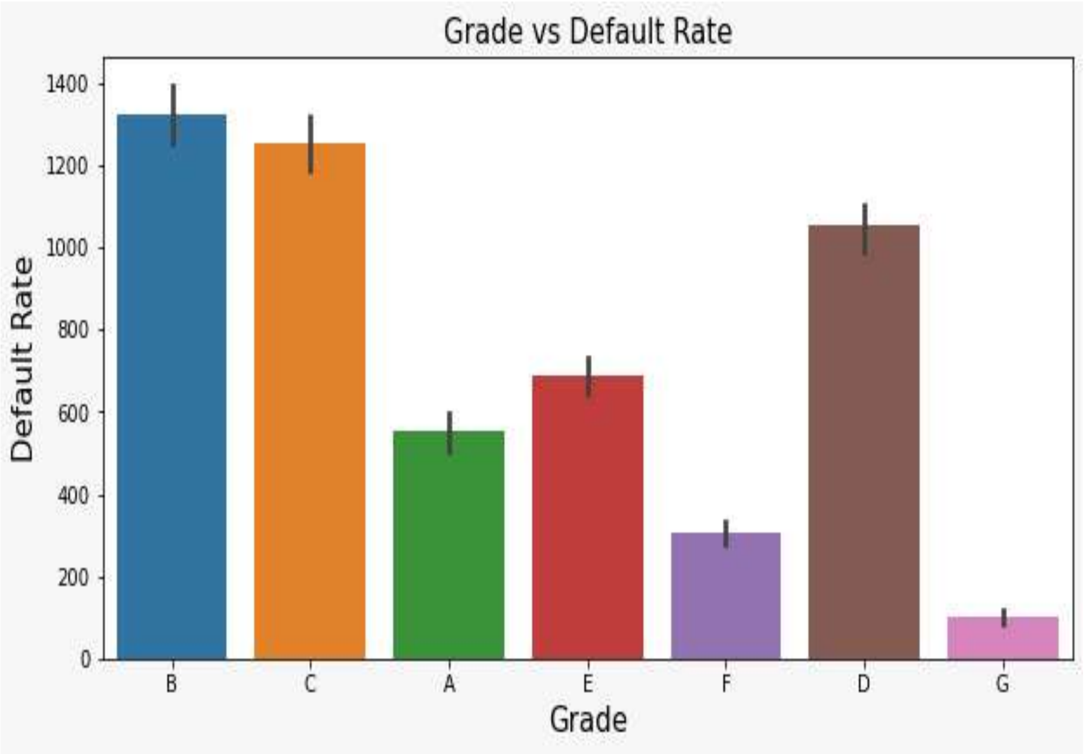
BIVARIATE ANALYSIS

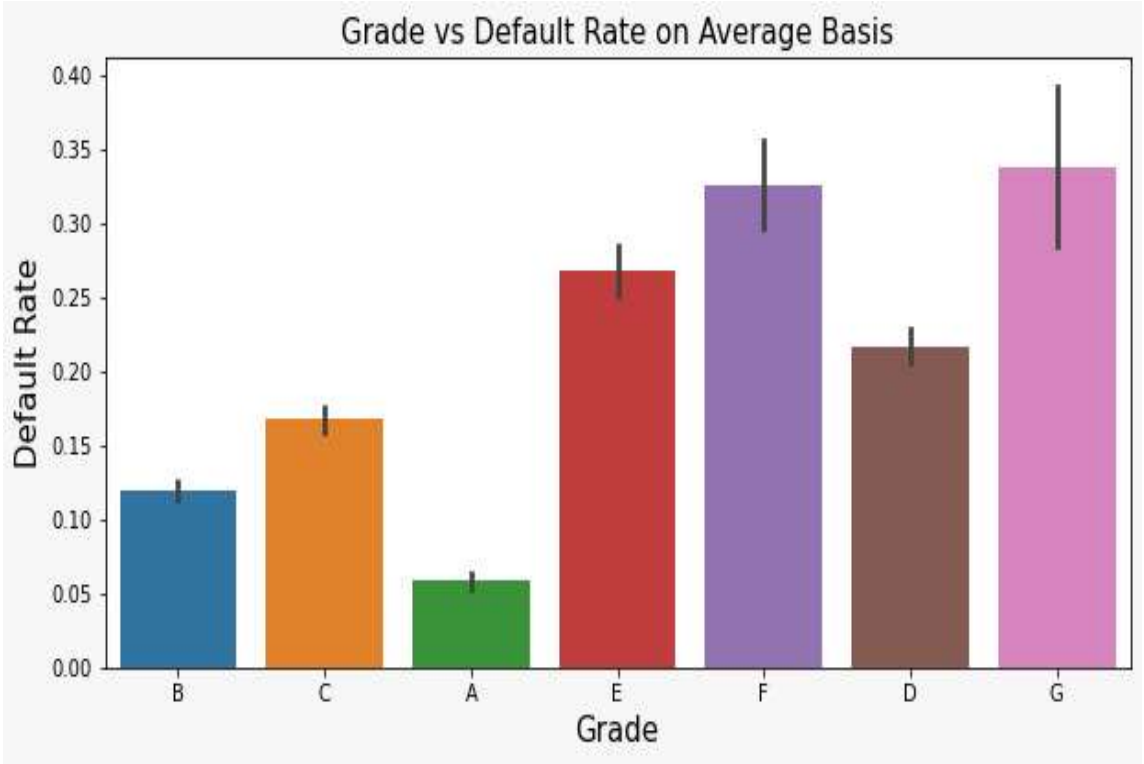




On the left we can see that debt consolidation have higher defaults but on and average basis small business have more default rate

BIVARIATE ANALYSIS





On the left we can see B and C grades are having the highest default rate .When we are looking at the average value the lower grades are having higher default rate

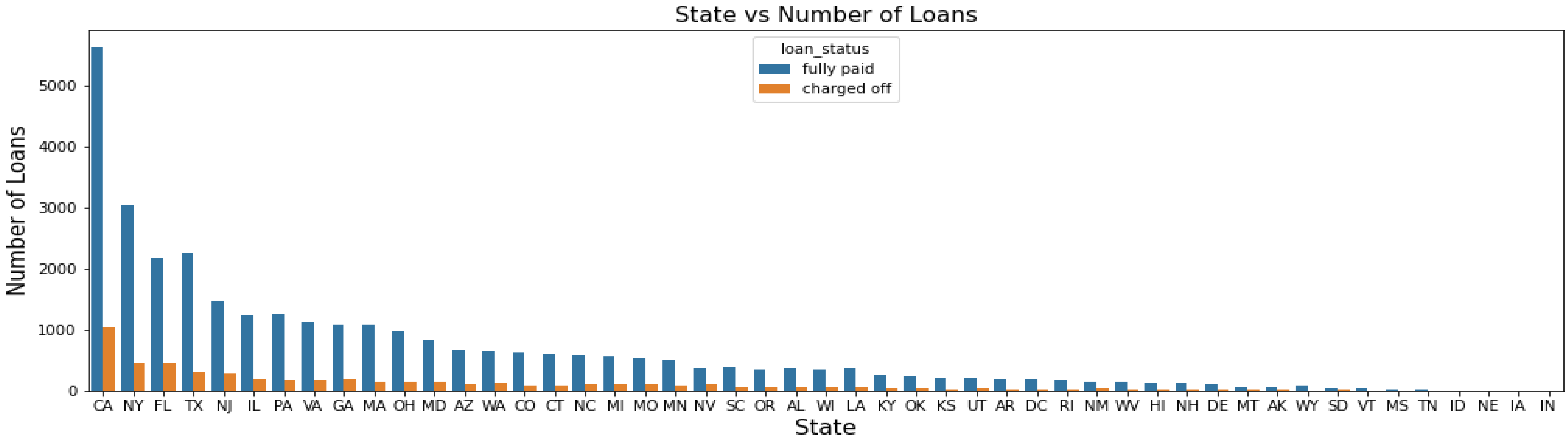
Conclusion

From the analysis we found that :-

We can conclude that most important indicators for default rate /loan status are :

- Interest rate .
- Term
- Loan Amount
- Grade
- Work Experience
- Also we have observed that defaults rate is higher for the year 2007 and 2011 but the amount of data for 2007 is very low and might be skewing observation. Loans issued and default rate is highest in the year 2011.

BIVARIATE ANALYSIS



The above figure shows the defaulters and non defaulters for each state .