

LENDING CLUB CASE STUDY

Suma T A & Shashank Singh

PROBLEM STATEMENT

- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- Losses associated with problem:

Business loss	Financial loss
If the applicant is likely to repay the loan, then not approving loan is a loss of business (rejecting loans for non - default).	If the applicant is not likely to repay the loan, then approving loan may lead to financial loss (approving loans for default).

OBJECTIVE AND TARGET

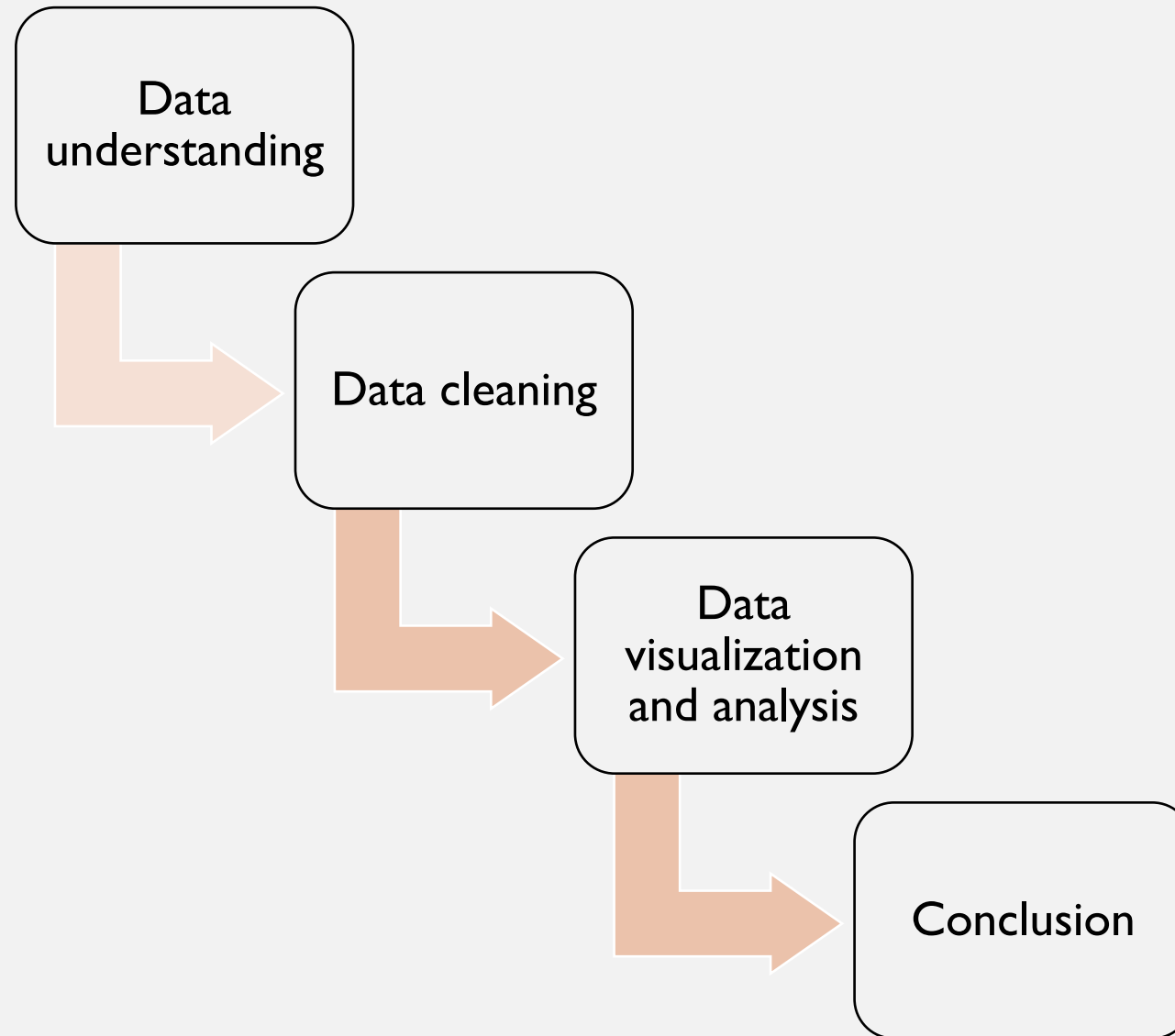
Objective

- How data can be used to minimize the risk of losing money while lending to customers

Target

- Reduce the Credit Loss by using EDA techniques to find driving factors for the defaulting of loan

STEPS IN USED IN DATA ANALYSIS



DATA CLEANING

- Initial shape of dataset-111 columns
- Remove columns with only null values-54 columns
- Removed columns with high number of missing values(>50%)-3 columns
- Remove columns with only one unique value-9 columns removed
- Removing irrelevant columns- removed 28 columns
- Final shape of data set after cleaning: 17 columns

CORRECTING DATA TYPES AND DERIVED COLUMNS

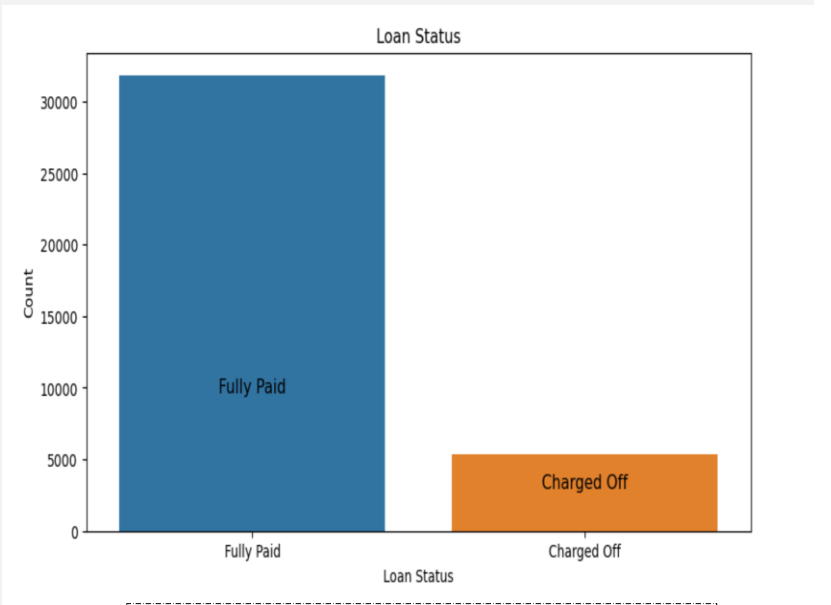
- The datatype of the following columns were corrected

Variable	Change(s)
Term	Trailing % removed. dtype changed to float
Grade	dtype changed to category
Sub grade	dtype changed to category
Emp_length	dtype changed to float
Home_ownership	dtype changed to category
Verifivation_status	dtype changed to category
Issue_d	dtype changed to datetime
Purpose	dtype changed to category
Addr_state	dtype changed to category

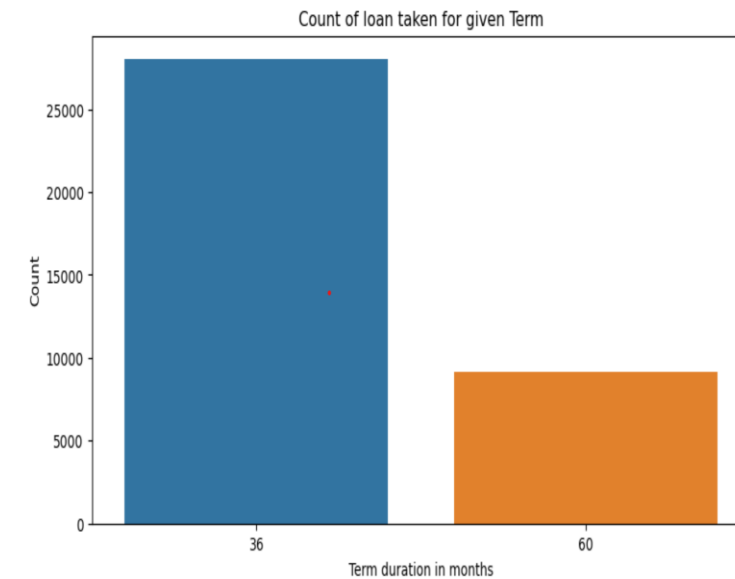
Two new derived columns

Two new columns 'issue_year' and 'issue_month' were derived using the issue.d column using 'to date_time'

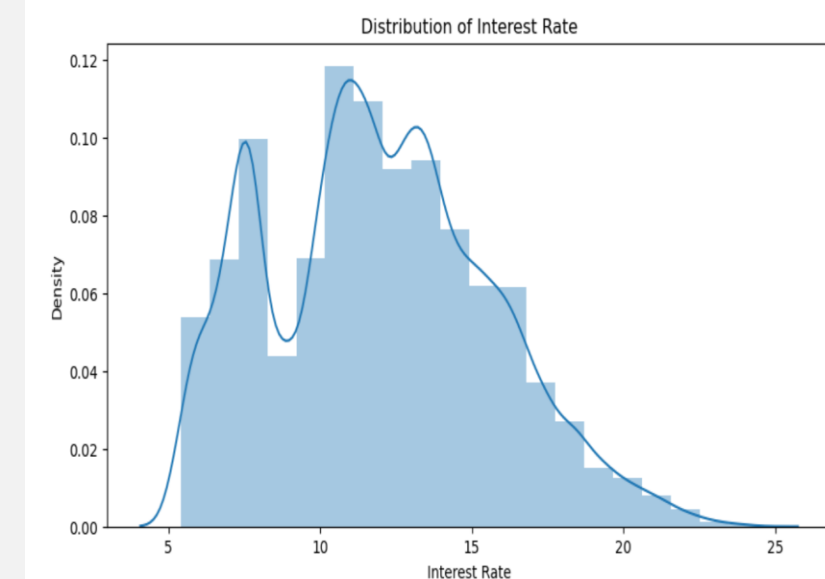
UNIVARIATE ANALYSIS



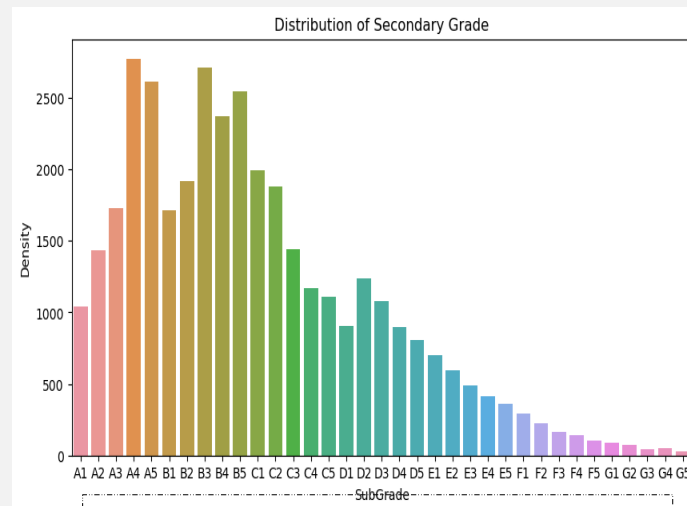
Defaulted loan are low in numbers compared to Fully Paid.



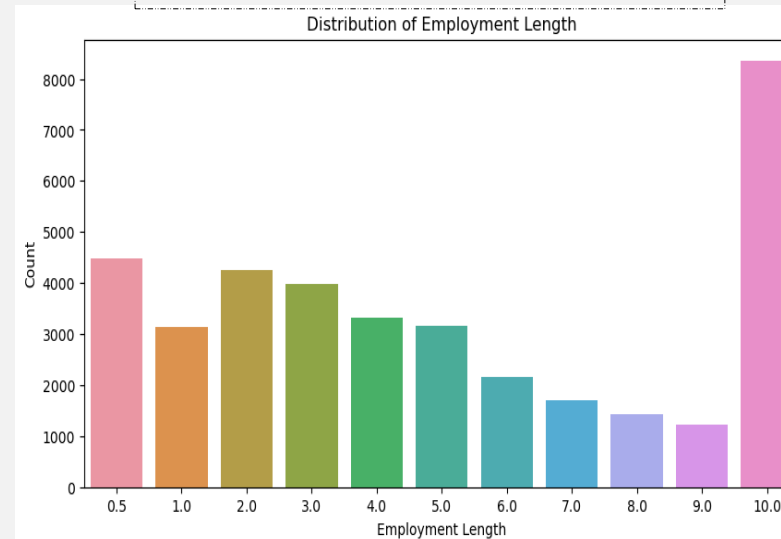
More than half of the loan taken has term of 36 months compared to 60 months



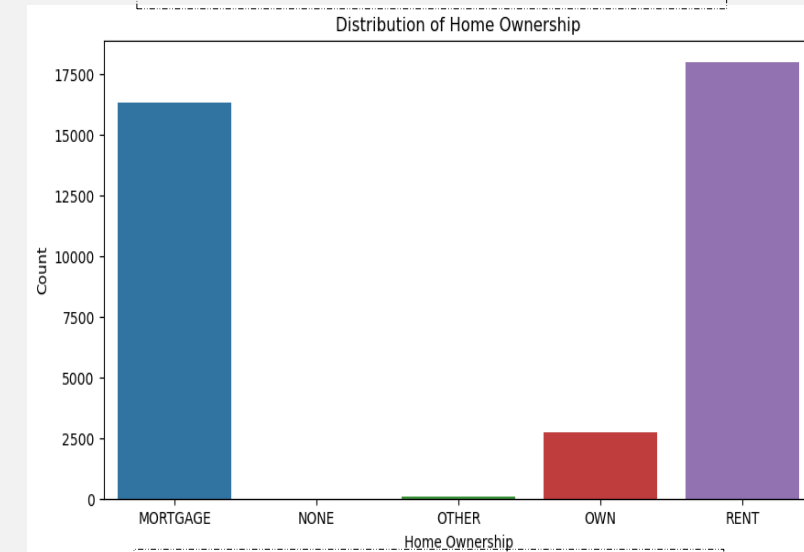
The interest rate is more crowded around 5-10 and 10-15 with a drop near 10.



A large amount of loans are with lower subgrade in grade 'A' and 'B' compared to rest showing most loans are high grade loans

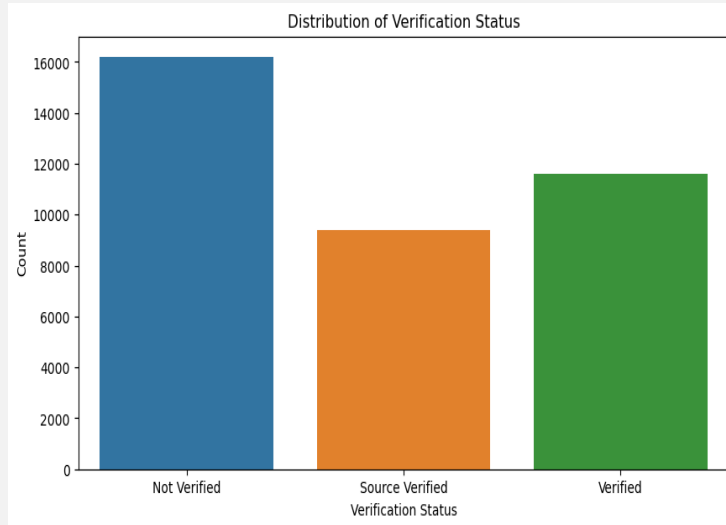


Majority of borrowers have working experience greater than 10 years

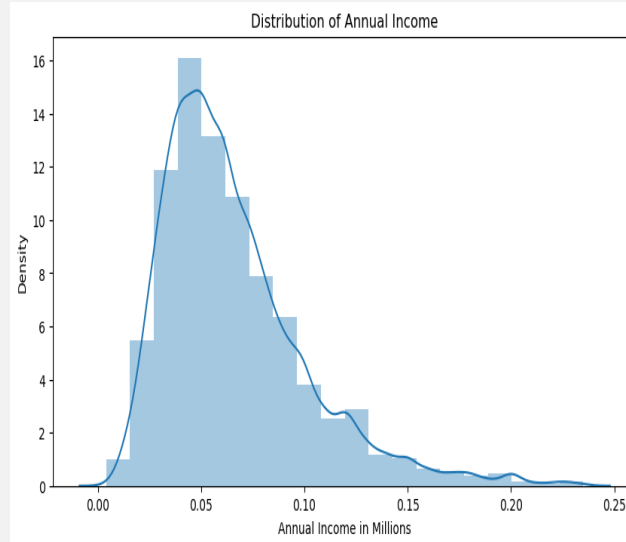


Majority of borrowers don't possess property and are on mortgage or rent.

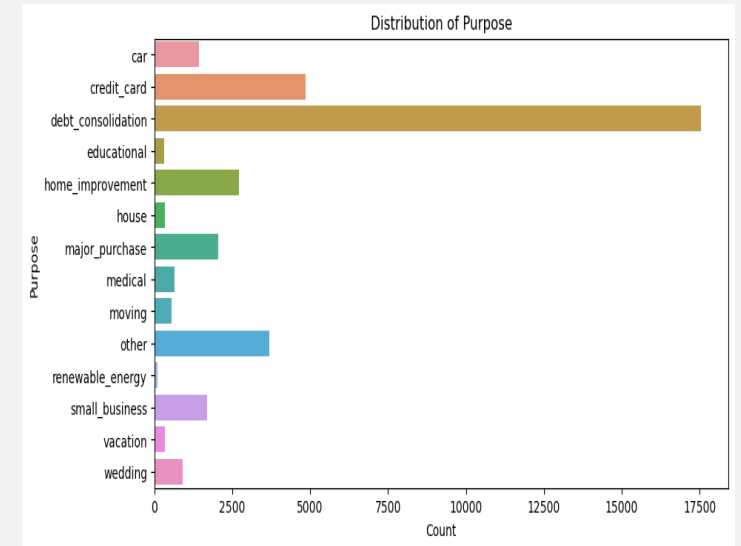
UNIVARIATE ANALYSIS



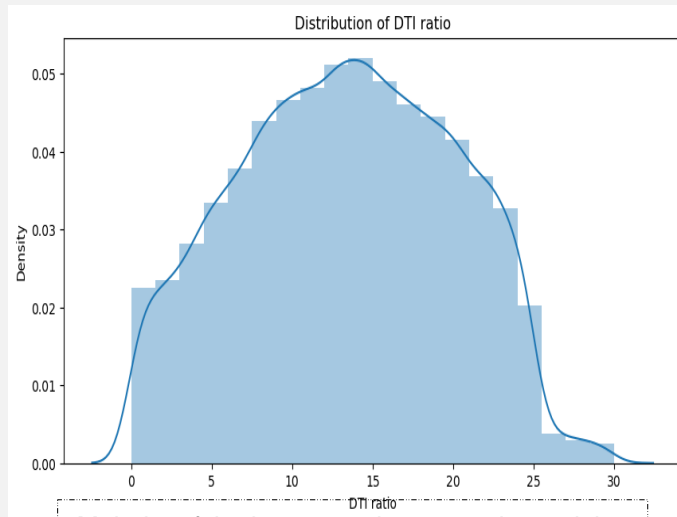
About 50% of the borrowers are verified by the company or have source verified.



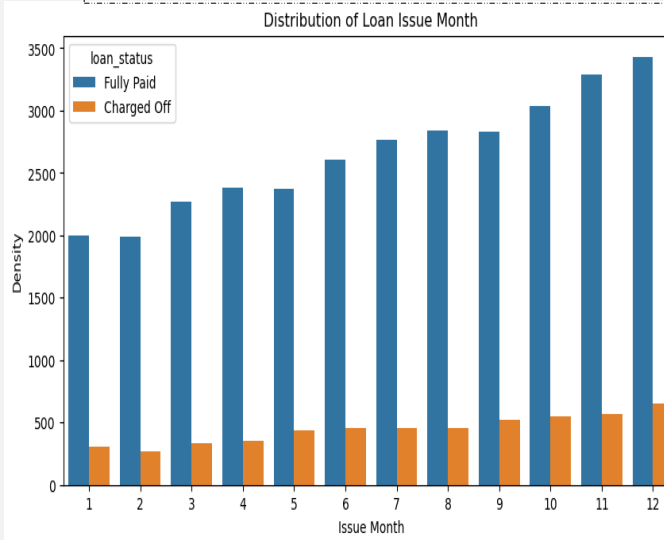
Annual Income shows left skewed normal distribution thus we can say that the majority of borrowers have very low annual income



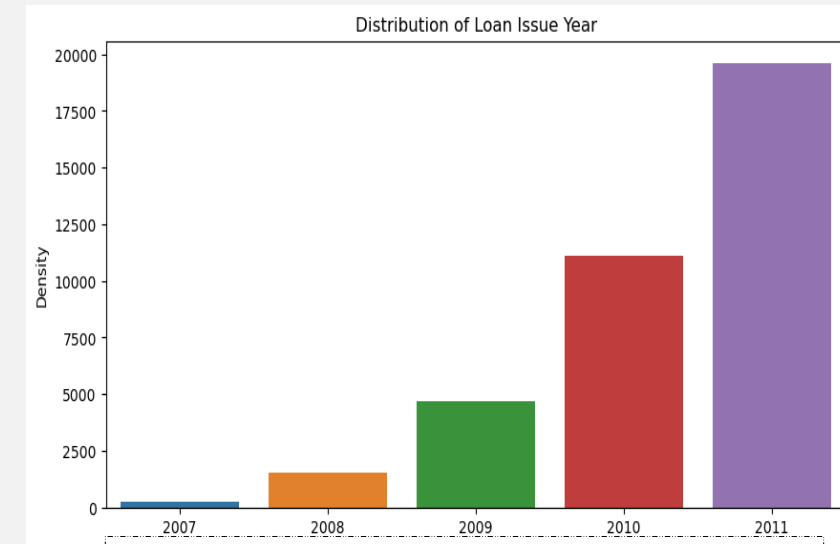
A large percentage of loans are taken for debt consolidation followed by credit card.



Majority of the borrowers have very large debt compared to the income registered, concentrated in the 10-15 DTI ratio

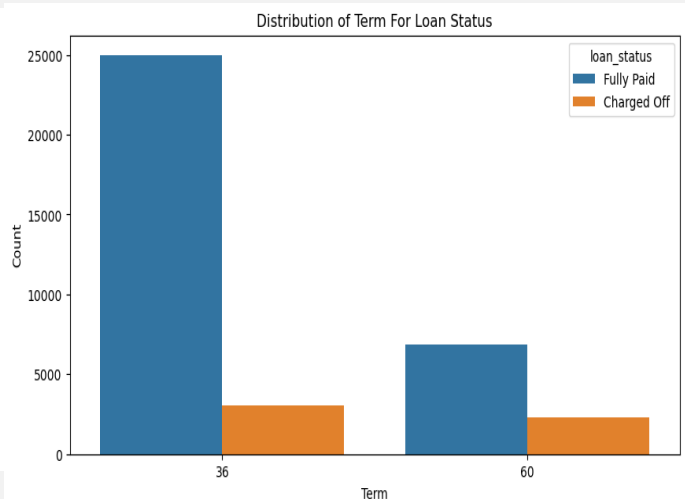
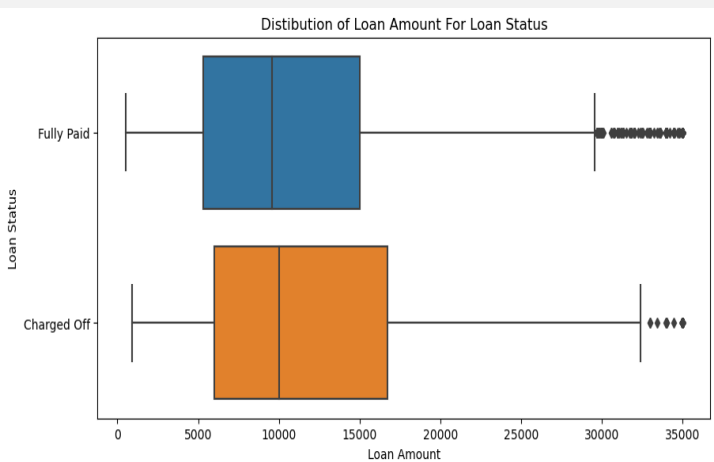


Majority of the loans are given in last quarter of the year.

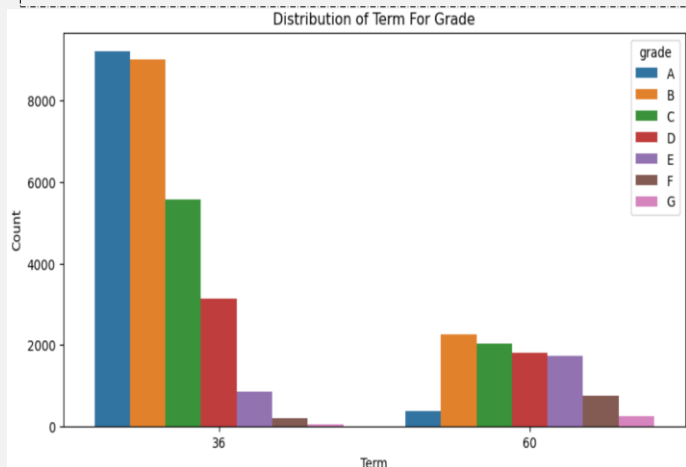


The number of loans approved increases with the time at exponential rate, thus we can say that the loan approval rate is increasing with the time

SEGMENTED UNIVARIATE ANALYSIS

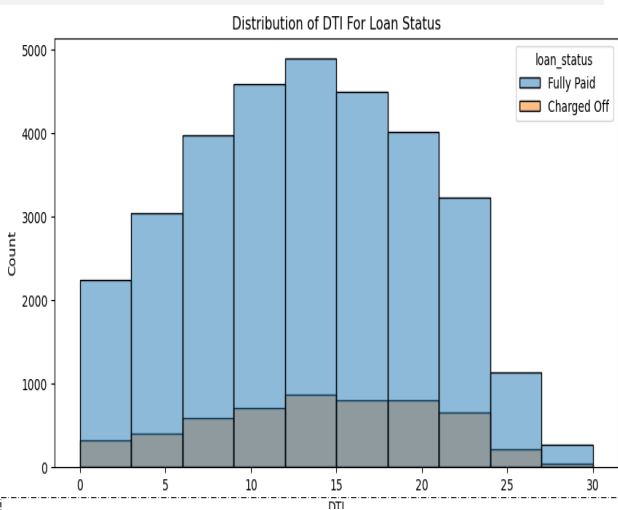


The mean and 25% percentile are same for both but we see larger 75% percentile in the defaulted loan which indicate large amount of loan has higher chance of defaulting.

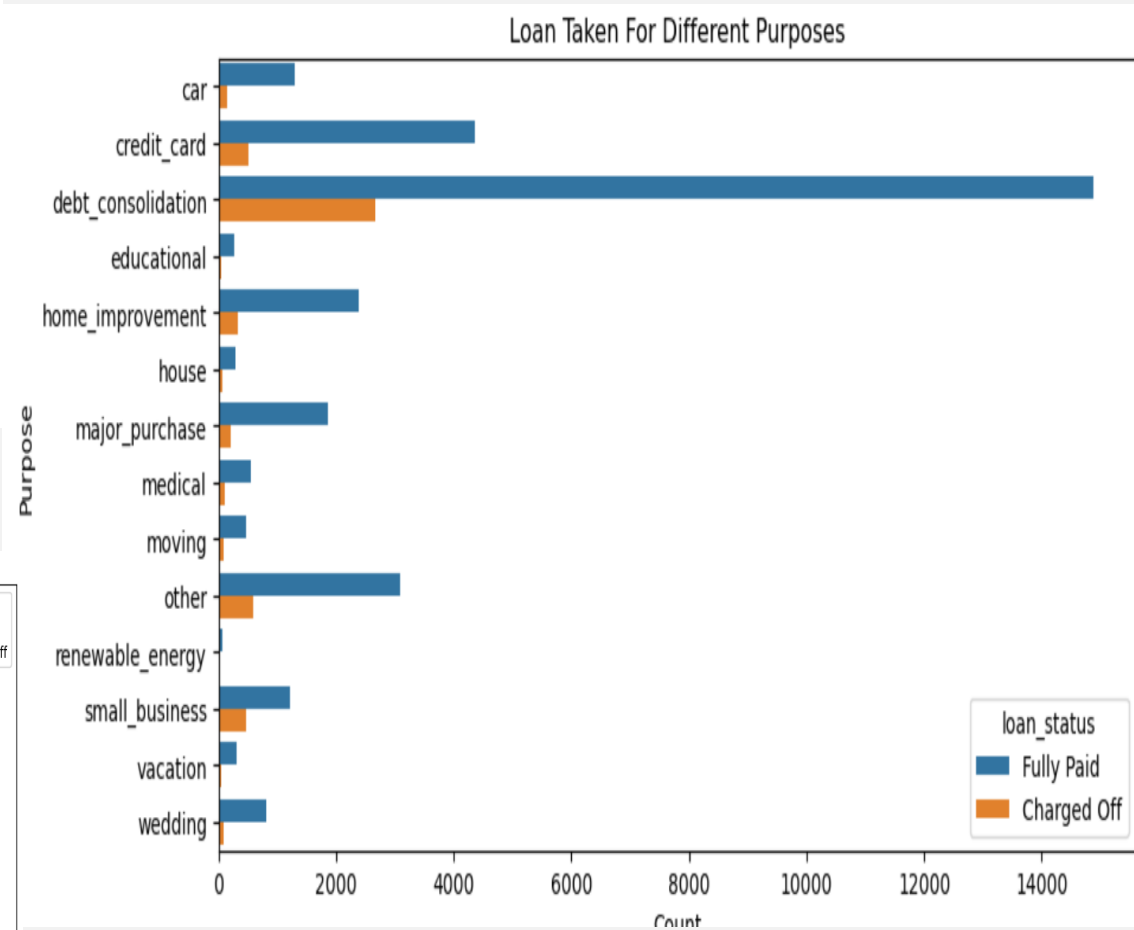


The loans in 36 month term majorly consist of grade A and B loans whereas the loans in 60 month term mostly consist of grade B, C and D loans

The 60 month term has higher chance of defaulting than 36 month term whereas the 36 month term has higher chance of fully paid loan

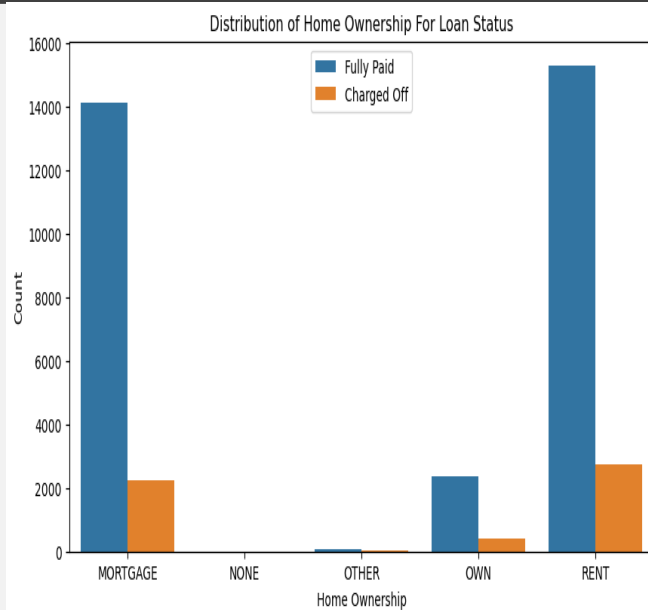


The Loan Status varies with DTI ratio, we can see that the loans in DTI ratio 10-15 have higher number of defaulted loan but higher dti has higher chance of defaulting.

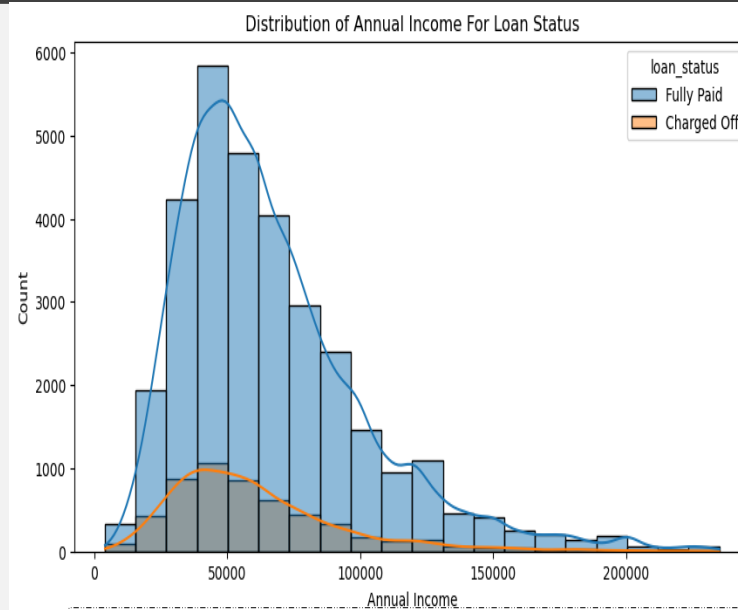


Debt Consolidation is the most popular loan purpose and has highest number of fully paid loan and defaulted loan.

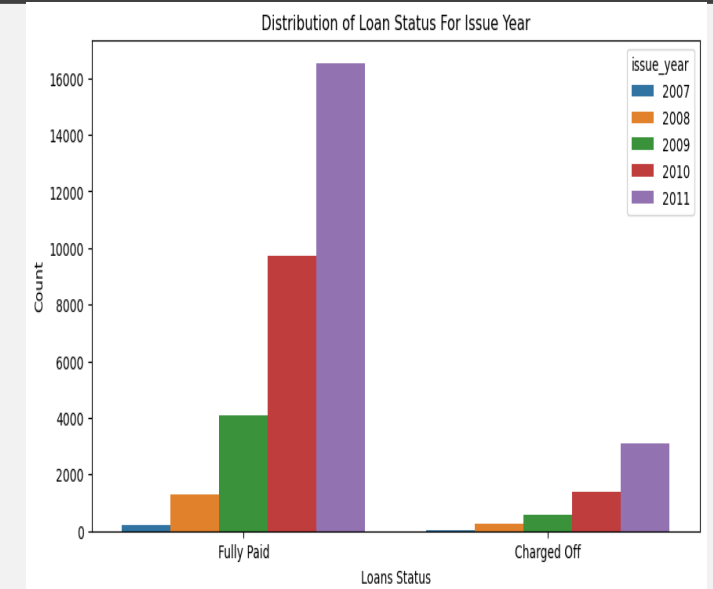
SEGMENTED UNIVARIATE ANALYSIS



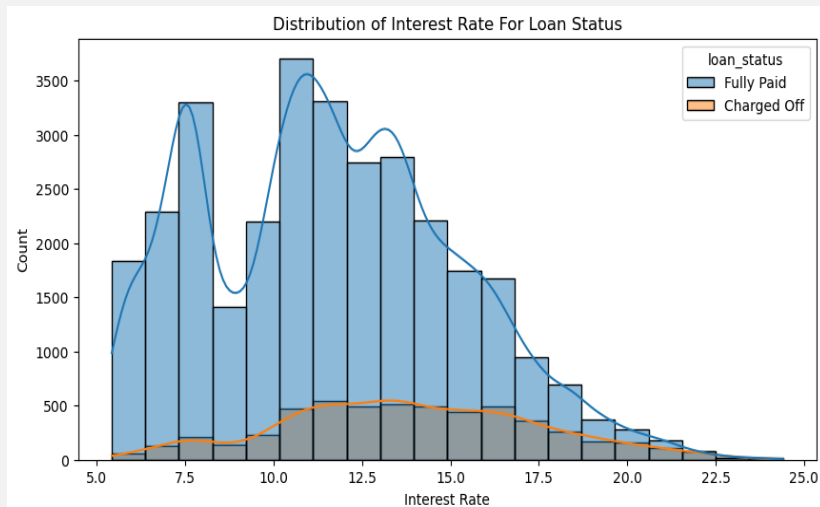
The Defaulted loan are lower for the borrowers which own their property compared to one who mortgage or rent



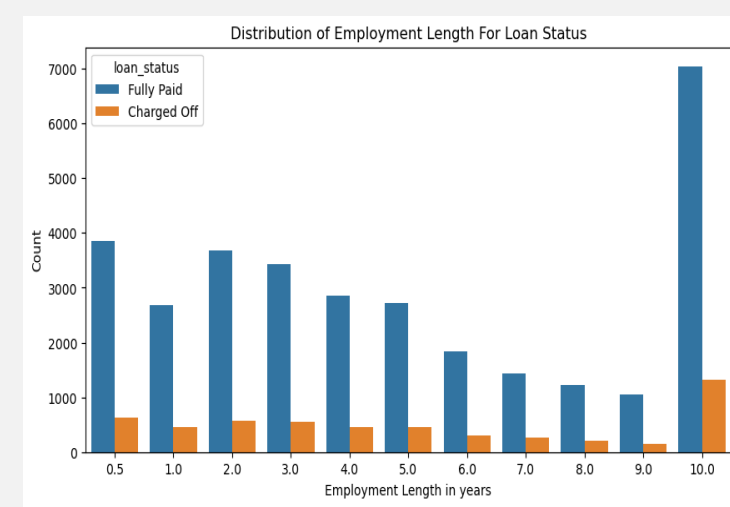
Borrowers with less 50000 annual income are more likely to default and higher annual income are less likely to default



The Fully paid loan are increasing exponentially with the time compared to defaulted loan

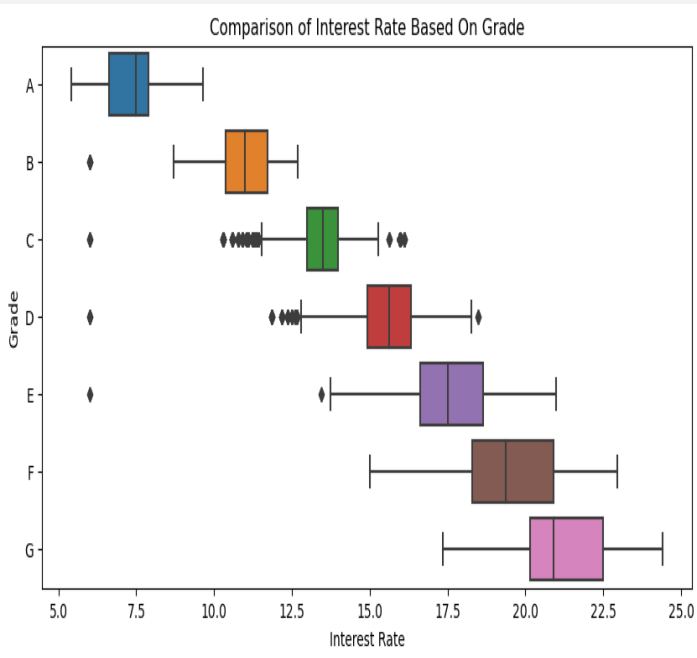


The default loan amount increases with interest rate and shows are decline aftr 17.5 % interest rate

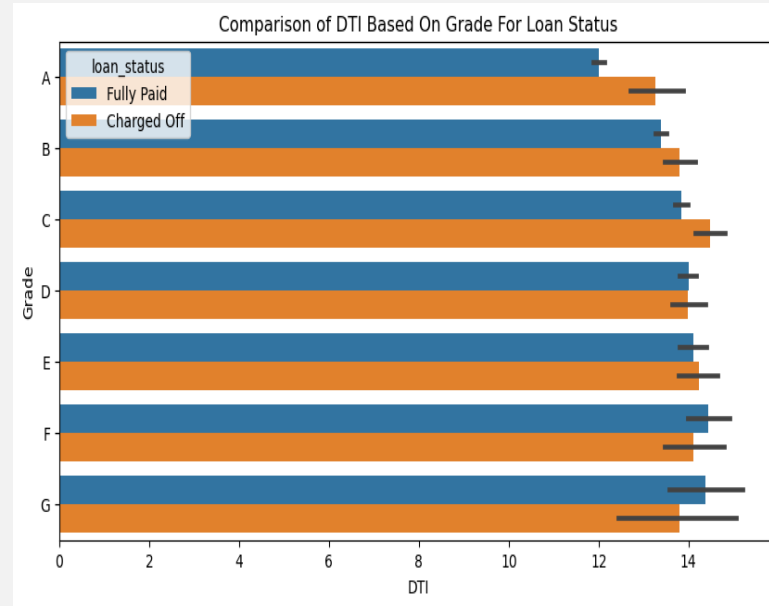


The Employees with 10+ years of experience are likely to default and have higher chance of fully paying the loan

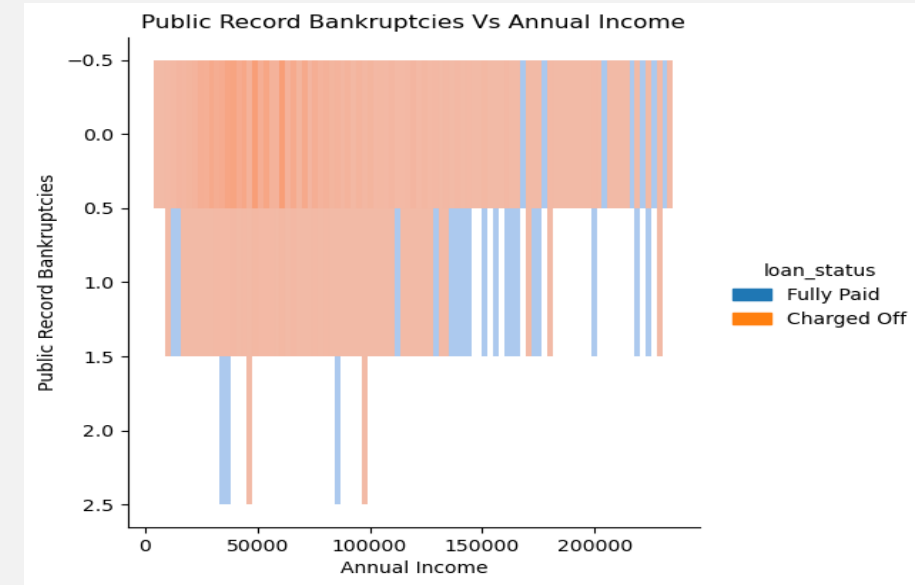
BIVARIATE ANALYSIS



The Grade represent risk factor thus we can say interest rate increases with the risk

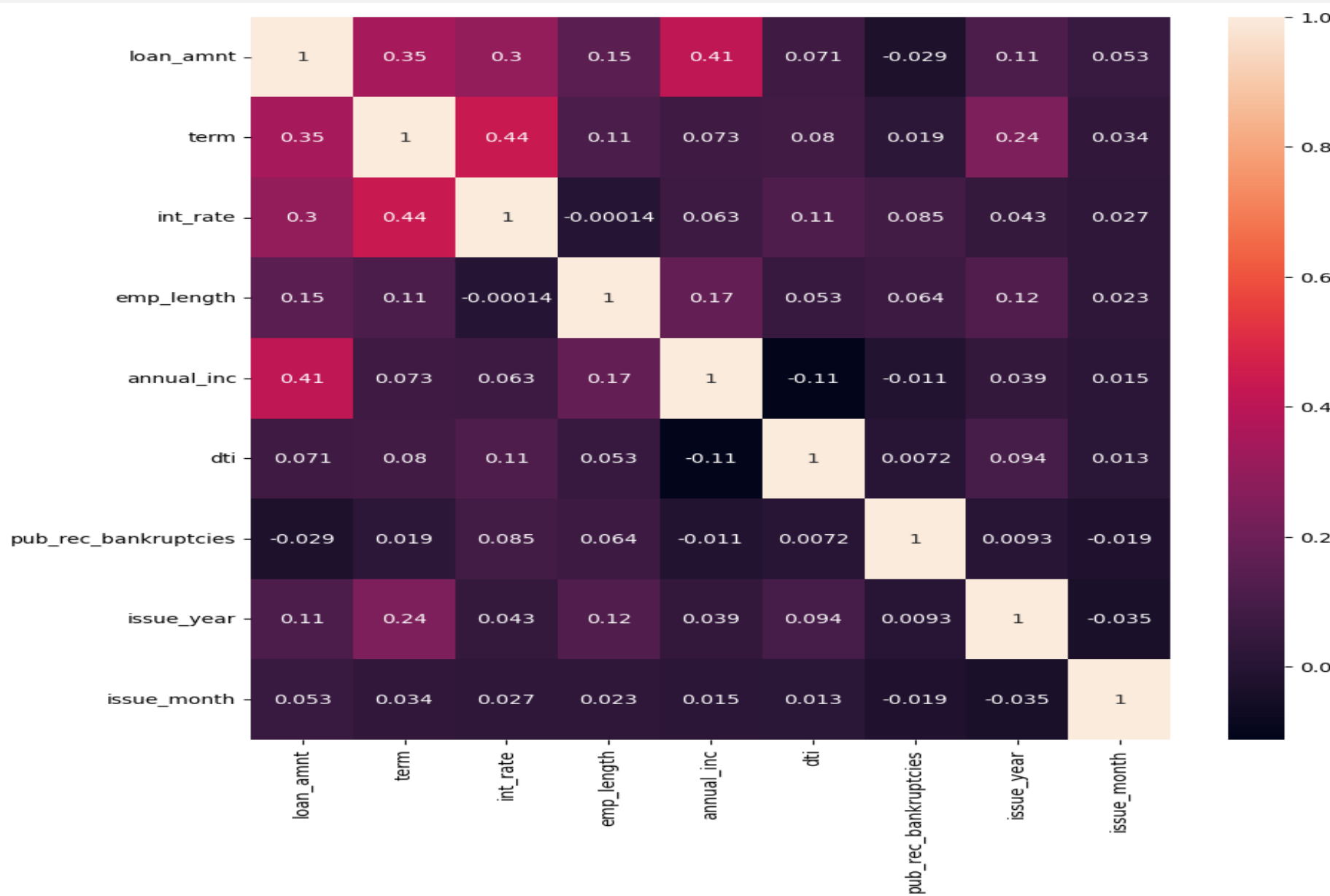


The Grade A which is lowest risk also has lowest DTI ratio which we can say that higher grade has lower rate of default.



The borrowers are mostly having no record of Public Recorded Bankruptcy and are safe choice for loan issue.

CORRELATION HEAT MAP



CONCLUSIONS

- **Univariate Analysis**

- ❖ The number of defaulted loan is 7 times less than the number of fully paid loan.
- ❖ The majority of loan has a term of 36 months compared to 60 months.
- ❖ The interest rate is more crowded around 5-10 and 10-15 with a drop near 10.
- ❖ A large amount of loans are with grade 'A' and 'B' compared to rest showing most loans are high grade loans.
- ❖ Majority of borrowers have working experience greater than 10 years.
- ❖ Majority of borrowers don't possess property and are on mortgage or rent.
- ❖ About 50% of the borrowers are verified by the company or have source verified.
- ❖ Annual Income shows left skewed normal distribution thus we can say that the majority of borrowers have very low annual income
- ❖ A large percentage of loans are taken for debt consolidation followed by credit card.
- ❖ Majority of the borrowers are from the large urban cities like California, New York, Texas, Florida etc.
- ❖ Majority of the borrowers have very large debt compared to the income registered, concentrated in the 10-15 DTI ratio.
- ❖ Majority of the borrowers have no record of Public Recorded Bankruptcy.
- ❖ Majority of the loans are given in last quarter of the year.
- ❖ The number of loans approved increases with the time at exponential rate, thus we can say that the loan approval rate is increasing with the time.

Bivariate Analysis

- The Grade represent risk factor thus we can say interest rate increases with the risk.
- The Grade A which is lowest risk also has lowest DTI ratio which we can say that higher grade has lower rate of default.
- The borrowers are mostly having no record of Public Recorded Bankruptcy and are safe choice for loan issue.

Recommendations

- Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:
 - DTI
 - Grades
 - Verification Status
 - Annual income
 - Pub_rec_bankruptcies
- Other considerations for 'defaults' :
 - Borrowers not from large urban cities like california, new york, texas, florida etc.
 - Borrowers having annual income in the range 50000-100000.
 - Borrowers having Public Recorded Bankruptcy.
 - Borrowers with least grades like E,F,G which indicates high risk.
 - Borrowers with very high Debt to Income value.
 - Borrowers with working experience 10+ years.