Bank Marketing Analysis

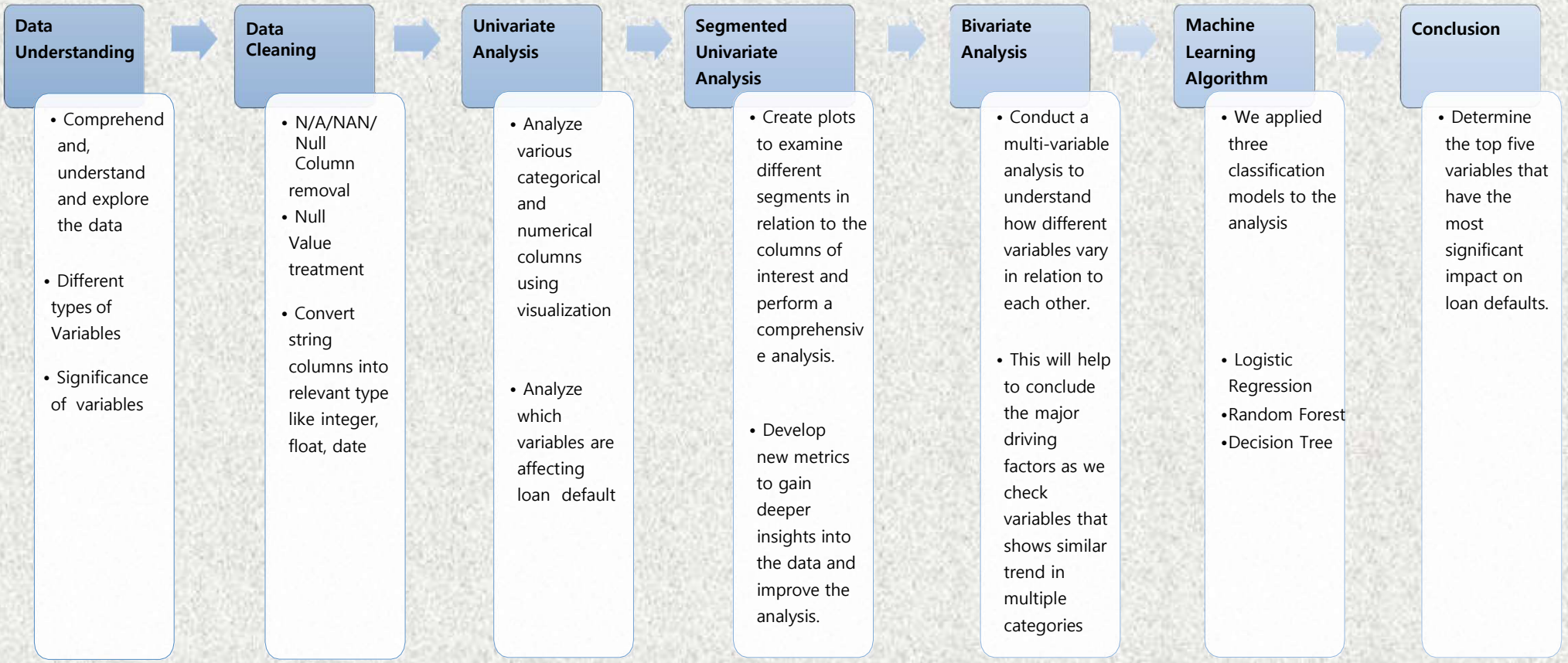
Group 11 – Parag Garg, Yashagra Sharma, Saugat Nayak, Saksham Chauhan

I. Problem Statement / Business Objective

- BRICS is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Through a quick internet interface, borrowers can readily get loans with cheaper interest rates.
- To conduct a risk analysis of the company, exploratory data analysis (EDA) and machine learning algorithms were used which helped in understanding the potential risks associated with the company's operations and devise suitable strategies to mitigate them.

Business Loss	• The high incidence of credit loss, which refers to the financial loss incurred by lenders, can be attributed to the loans extended to applicants who are considered to be 'risky'. • Borrowers who default cause the largest amount of loss
Risk Factors	• Approving a loan application that can result in default • Disapproving a loan application that can be fully paid
Analysis Goals	• Finding the factors or variables that derive the loan default

II. Analysis Approach



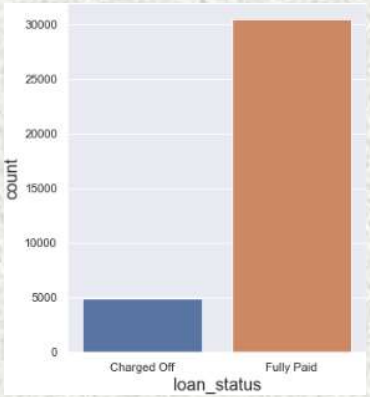
III. Data Understanding

We have loan dataset, containing details of accepted loans.

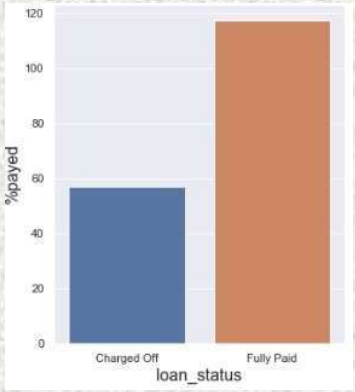
- These loans are divided into these major segments based on their status : -
  - **Fully paid:** The applicant has fully repaid the loan, including the interest and principle.
  - **Current:** The applicant is currently paying the installments, hence the loan's term has not yet ended. They are not listed as "defaulted" candidates.
  - **Charged-off:** The borrower has defaulted on the loan if the installment payments have not been made on schedule for an extended period of time.
- We are only analyzing Fully Paid and Charged-Off loans, as current loans can end up in either of the other two states. Analyzing them will bring ambiguity to our finding.
- There are 110 columns/variables in the data set covering various details about the loan applications and repayment
- The Variables can be broadly classified into 3 categories:
  - **Applicant Details :** Information about the applicant, including "Annual Income," "Employment Length," "Address," "Home Owned Status," etc.
  - **Loan Details:** Information from the loan application, such as the "Loan Amount," "Interest Rate," "Grade," and "Term."
  - **Customer Behavior Details:** variables like "Recoveries" and others that describe how customers behave while making payments after the loan has been approved.
- We will be concentrating more on "Applicant Details" and "Loan Details" variables in order to uncover characteristics that can help us detect loans that may default before sanction.

IV. Univariate Analysis

Loan Overall

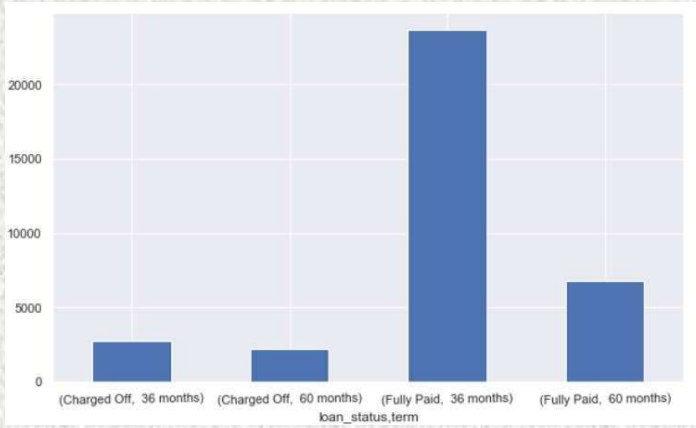


~14 % loans are charged off in the dataset.



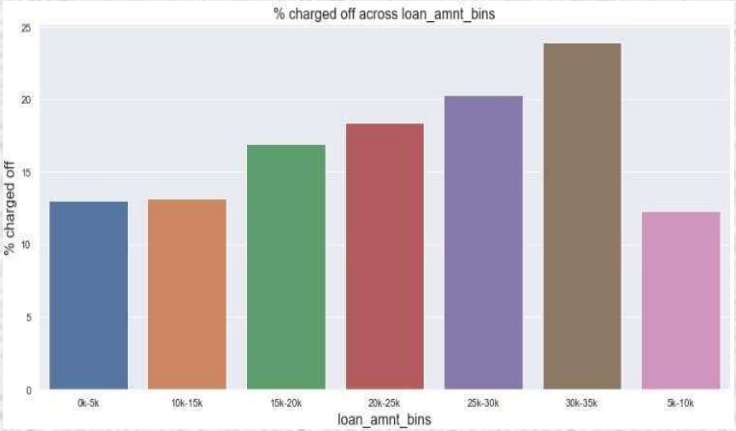
~57.12 % of total loan amount for charged off cases has been payed.
LC is gaining some profit from the fully paid ones (17.24 %)

Term

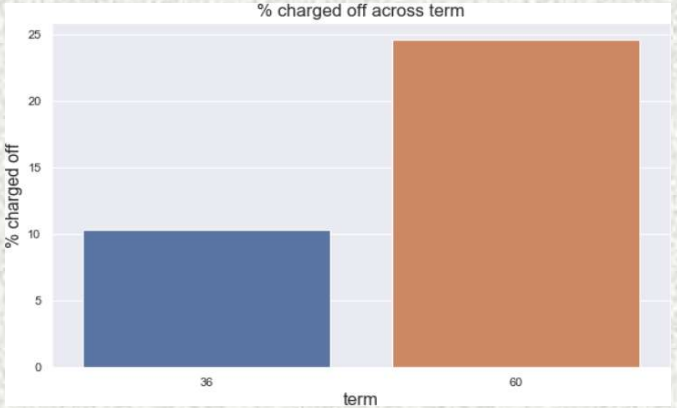


Number of short term loans is ~3 times more than long term loans.

Still, the total number of charged off cases is very slightly different.



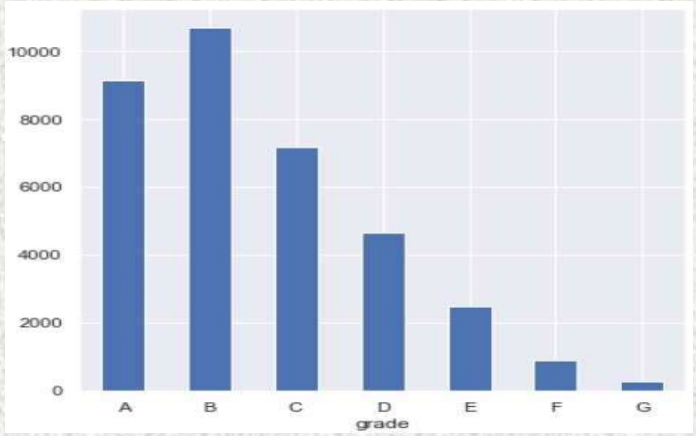
As the loan amount increases the % of charged off cases also increases.



Long term loans are more likely to get charged off compared to short term.

IV. Univariate Analysis

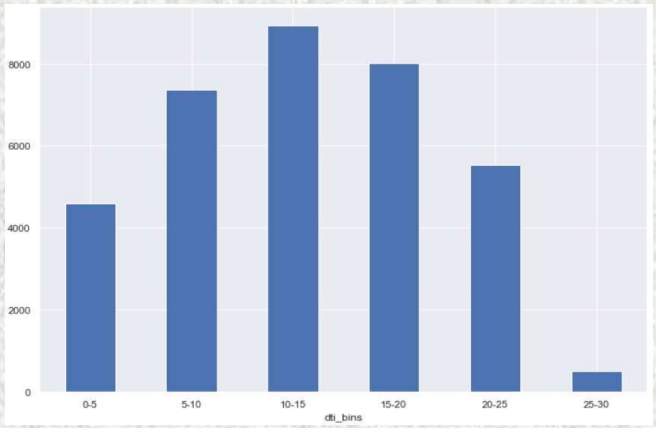
Grade



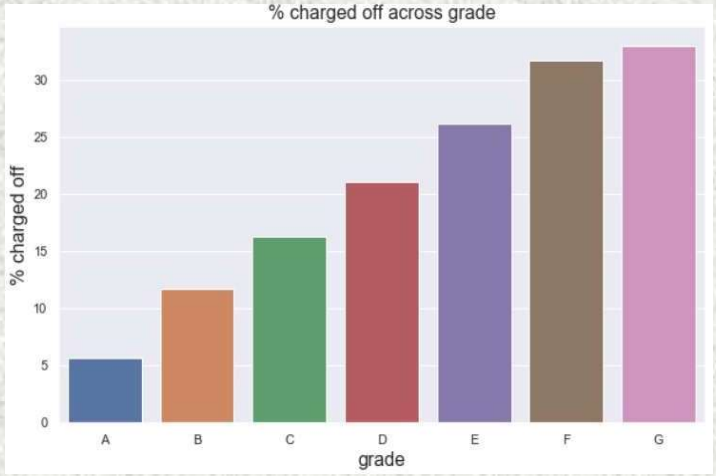
Loans of grades A, B, C have the highest count.

A similar trend can be seen for sub grade.

DTI

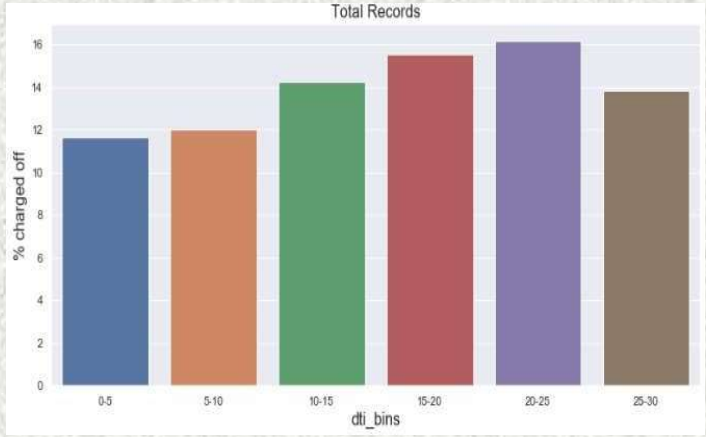


~ 40 % loans are given to borrowers with DTI greater than 15



# of charged off cases increase as the grade of the loan increases.
With the least being in A,B,C

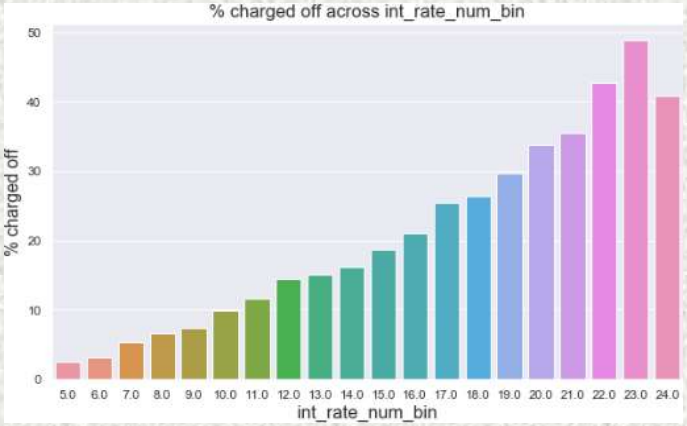
A similar trend can be seen for sub grade.



For DTI's greater than 15%, the chances of default gets of close to 16 %

IV. Univariate Analysis

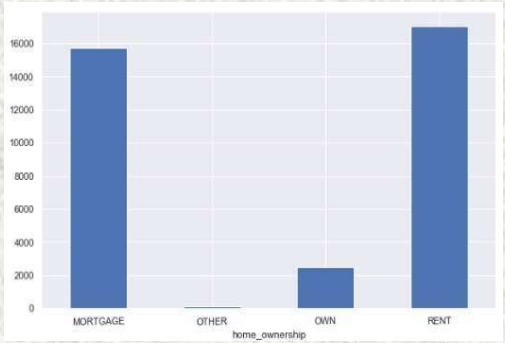
int_rate



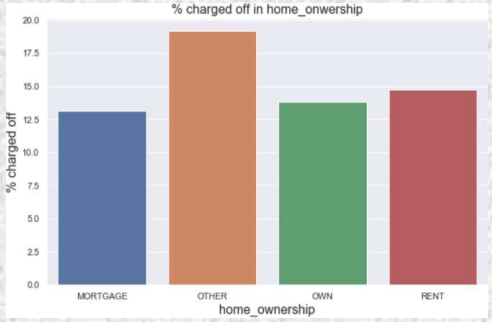
Interest rate is ranging between 5 – 25 %

And as the interest rate increases the percentage of charged off also increases.

Home Ownership

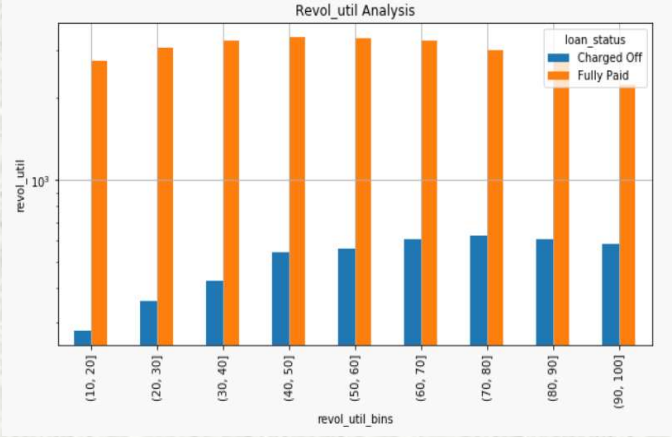


Most loans are taken by borrowers who have their home on mortgage and rent



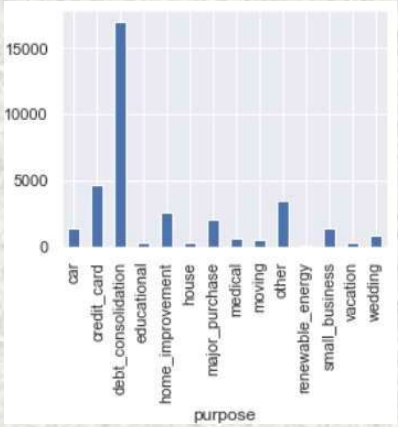
% charged off is higher than 12.5 % for all, but mortgage and rent are critical as they have the highest count.

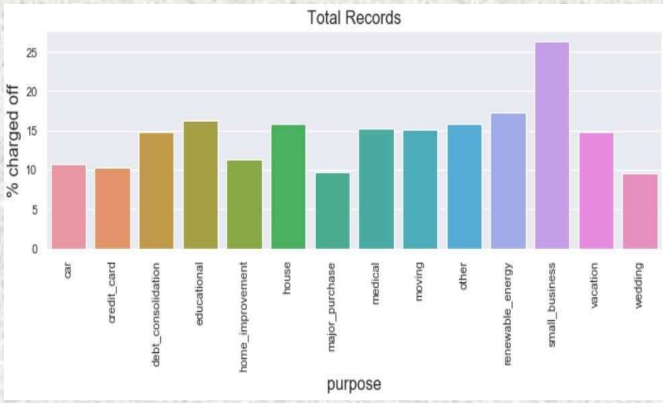
Revol util



Charged off loans are affected by revol_util as the revol % increases risk of loan being charged off increases with highest risk lying when revol_util is under the range of 60 - 90 %

Purpose



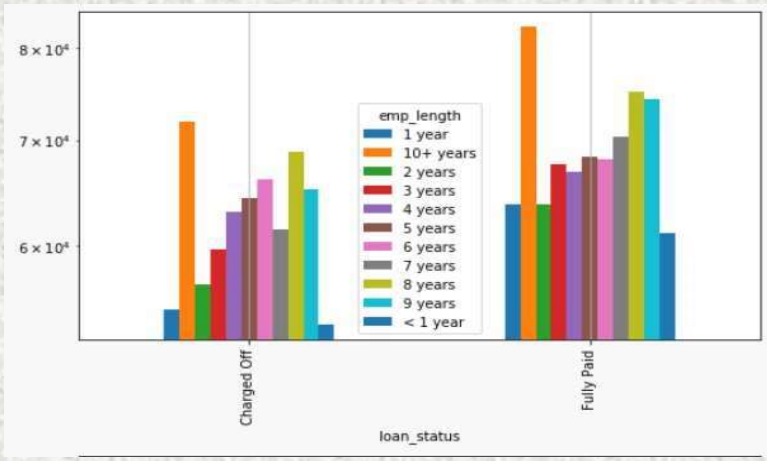


Most loans (close to 50 %) are for the purpose of debt_consolidation. Small business have highest percentage of charged-off loans. All other purposes on average have ~12 % charged off loans but debt_consolidation due to its high frequency is always important.

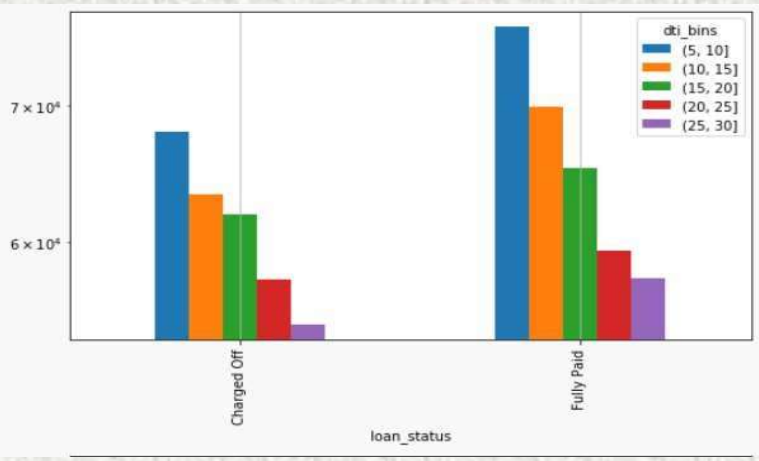
V. Bivariate Analysis

Annual Income VS

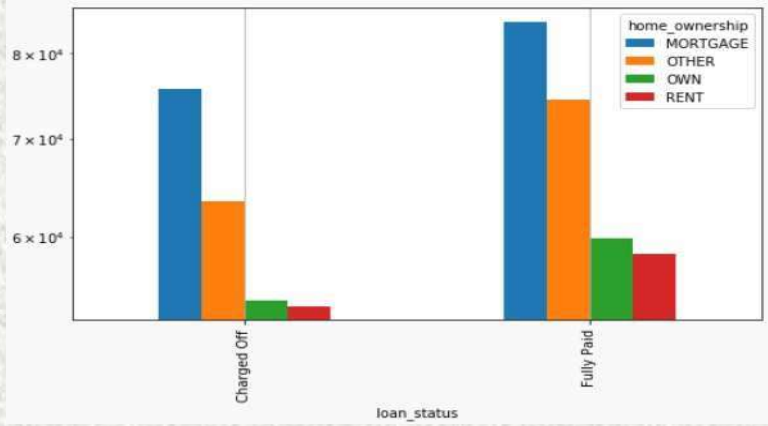
Emp_length



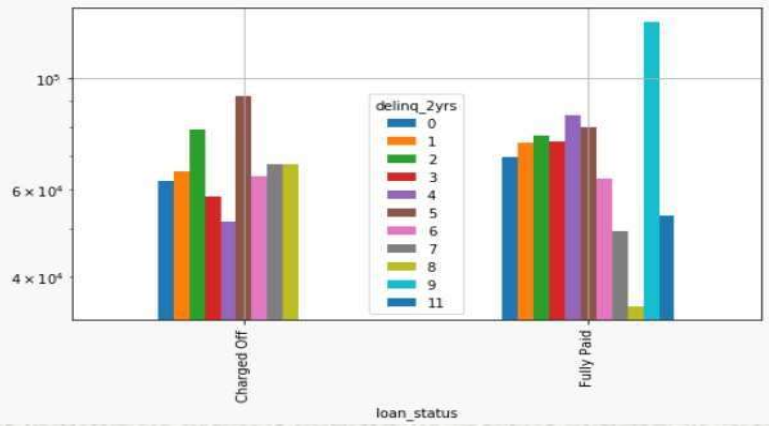
DTI bins



Home ownership



Delinq_2yrs



Annual Income for charged off cases is lower than Fully paid, whereas the annual income varies in a different pattern against different variables as seen in plots.

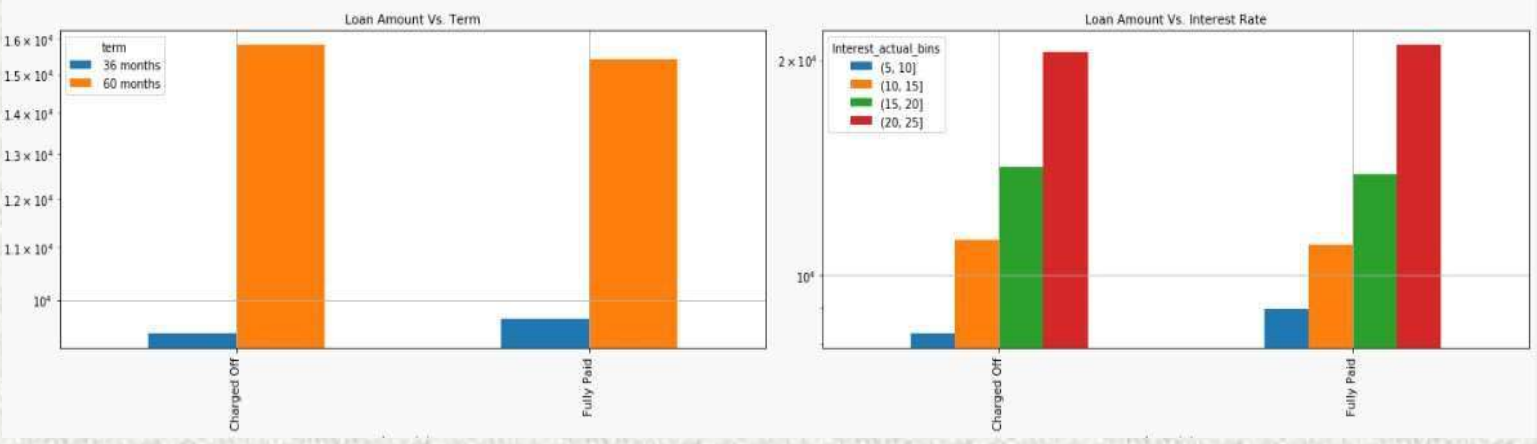
Annual income is highest for mortgage home ownership & for emp_length greater then 10 yrs Univariate analysis showed that Highest defaulters are also from mortgage & over 10years employed

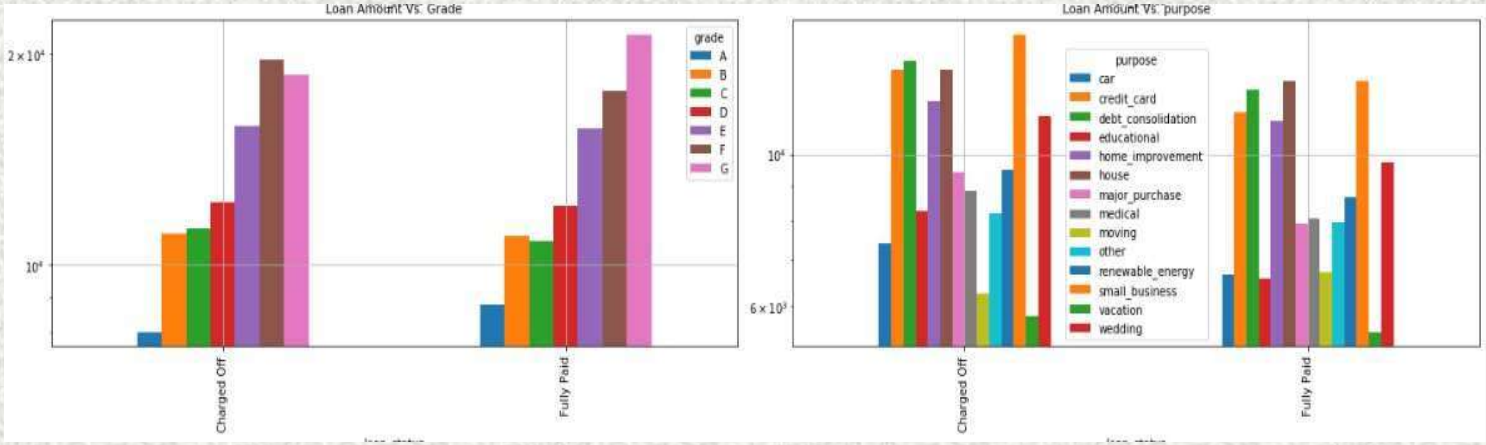
DTI however stands true showing that people having high DTI have lower annual income, and as we know it is higher for charged off cases

Thus it seems to be a strong indicator

V. Bivariate Analysis

Loan Amount
Vs



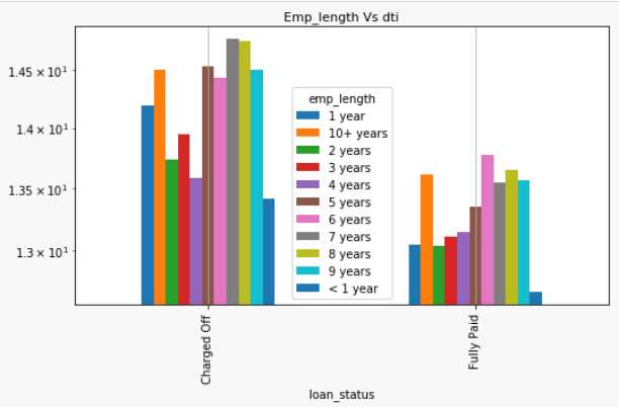


Loan Amount, Term, Grade & Interest rate all are directly proportional to each other
-Higher Loan amount tends to have higher term
-Higher loan amount & higher term tends to have higher grade
-Higher grade will imply higher interest rate
Loans with higher amount, term, grade & interest rate tends to het charged off more often

Purpose also holds true in this analysis as most of the cases that are charged off are from debt_consolidation & small business Higher amount loans are also tend to be given in these two categories

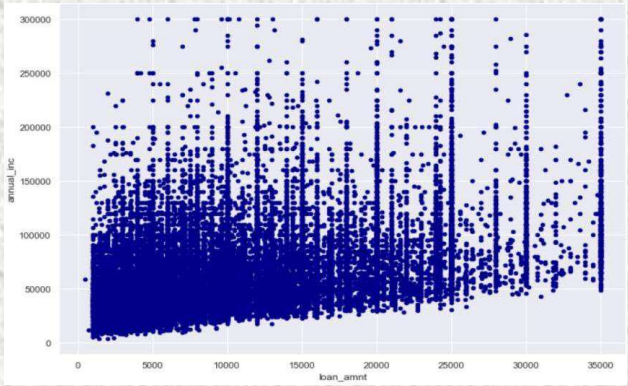
V. Bivariate Analysis

Emp length vs DTI



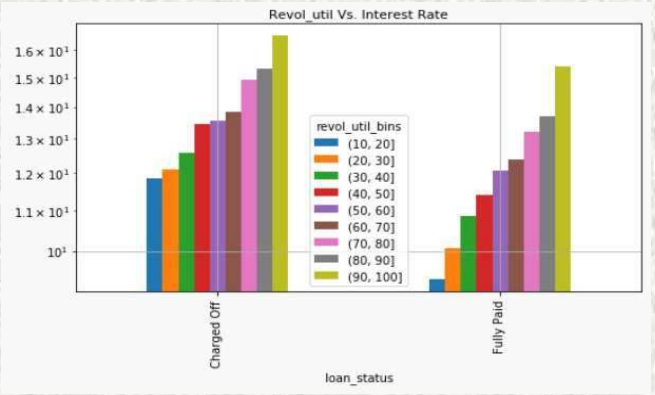
DTI holds true as it increases with emp length and average dti is greater for charged off cases

Loan amount vs annual income



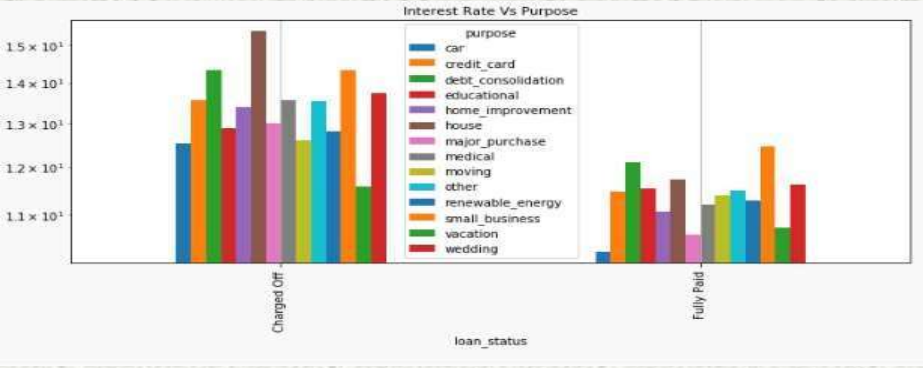
Its clear as the annual income is increasing the maximum loan amount is also increasing.
But there are a lot of cases where the annual income is less than 50 k and they have loans of amount greater than 20k, up to the max amount of 35k.

Revolving util vs interest rate and DTI

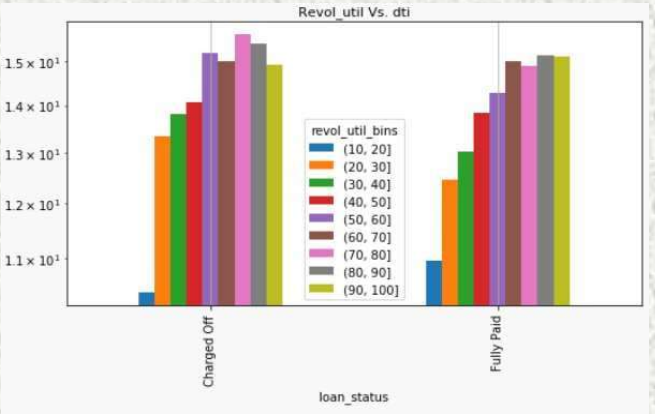


Interest rate is higher for higher Revol_Util & also DTI is higher for higher Revol_Util peaking between 70 ~ 80%. Thus we can observe Revol_Util is one of the stronger indicators of loan default, as it increases risk of default increases.

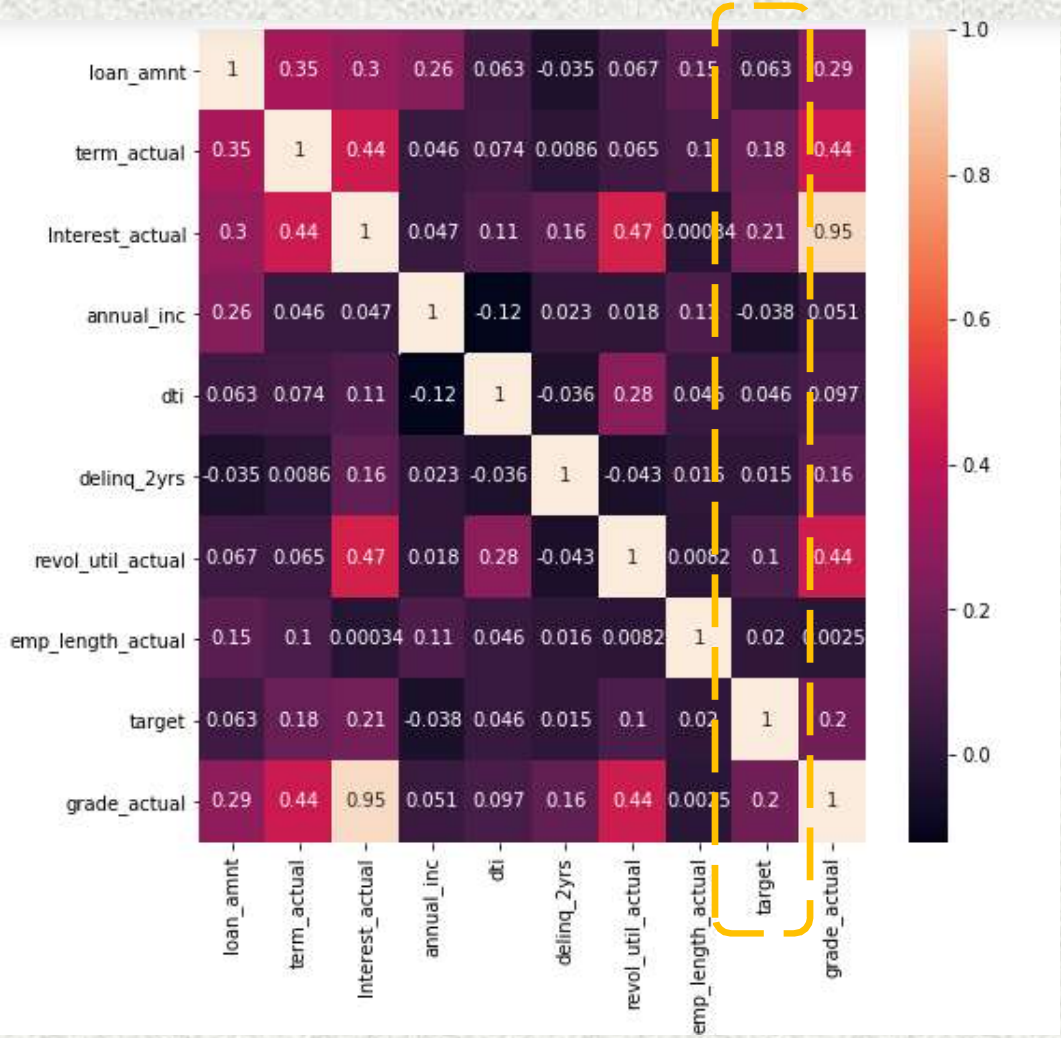
Interest Rate vs Purpose



Small business , debt consolidation , house and credit card have the highest interest rates.
Small business is risky as it has the highest % of default (univariate)



VI. Correlation Matrix



•**Correlation matrix** is drawn with continuous variables to check the correlation with target column

•As per the correlation matrix the variables which are the most correlated with the target column **("Charged Off loans")** are :

1) Interest Rate
2) Grade
3) Term
4) revol_util
5) Dti

Note: As interest rate is derived using the term & grade thus we will exclude interest rate from driving factor and take term +grade instead

VII. Machine Learning Algorithm

	Logistic Regression	Random Forest	Decision Tree
Accuracy	0.862501	0.788616	0.505796
Misclassification Rate	0.137499	0.211384	0.160871
True Positive Rate	0.999672	0.885515	0.845082
False Positive Rate	0.999314	0.820178	0.782341
Specificity	0.000686	0.179822	0.217659
Precision	0.862732	0.871519	0.871219
Prevalence	0.999623	0.876543	0.557629

The results suggest that the logistic regression model performed better than both the random forest and decision tree models in terms of accuracy, misclassification rate, and specificity. However, the decision tree model had the highest true positive rate and precision.

VII. Conclusion

- **After all the analysis the 5 driving factors that indicate the loan can be defaulted for a given loan amount are:**

Driving Factors	Description	Trend
Term	The number of payments on the loan. Values are in months and can be either 36 or 60.	• Term increases as the loan amount increases implying higher interest rate thus increasing risk of loan default
Grade	Assigned loan grade	•As grade increases interest rate increases increasing the risk of loan default • A & B are the safest grades
DTI	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.	• Risk of loan default increases as DTI increases • DTI is higher for charged off cases •Applicants having high DTI tend to have low annual income
Revol_Util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.	• Revol util is directly proportional to risk of loan default • Revol_util ranging from 60~90% has highest default cases
Interest Rate	The amount a lender charges a borrower and is a percentage of the principal—the amount loaned	• Loan Purpose also indicates the risk of loan default •Small business & debt_consolidation has highest default cases