

LENDER GROUP CASE STUDY

By – Silapada Hansda



Problem Statement

- Developing a basic understanding of risk analytics in banking and financial services and understand how data from existing history of customers can be used to minimize the risk of losing money while lending to customers.
- Identifying driving factors using EDA to help business identify the red-flags while processing the loan applications to reduce losses due to defaulters

Analysis Approach

- We have the loan data set of all the loans issued.
- We will only be looking for the data of loans which are either Fully Paid or Charged Off
- Our target variable is “Loan Status”.
- We will be cleaning the data i.e. removing data which have missing values/null values, or which doesn't have any effect on the target variable , correcting data to make it more meaningful
- We will be doing univariate and bivariate analysis on the numerical and categorical variables to identify the factors which directly affects the Loan Status.
- We will identify variables which are common among the defaulters , so that those applicants can be identified during the loan submission process hence saving the loss due to bad loans
- **Driving Factors will be marked in Red** and non driving in yellow for better viewing.

Dataset Information



There are 39717 records and 111 columns(variables) in the raw dataset

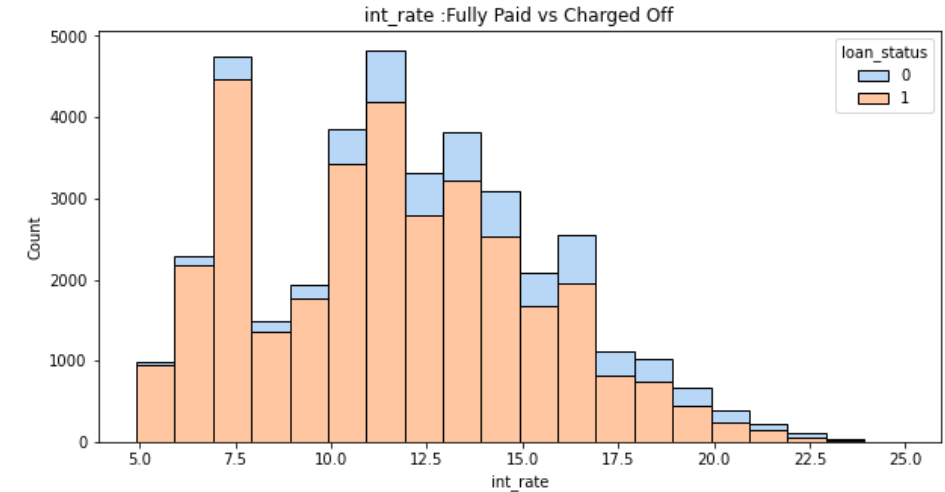
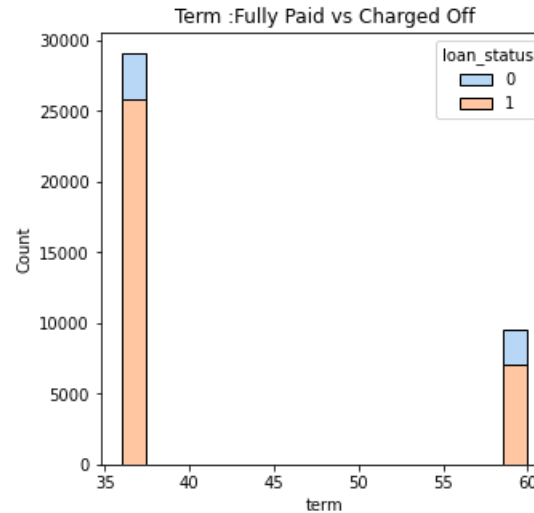
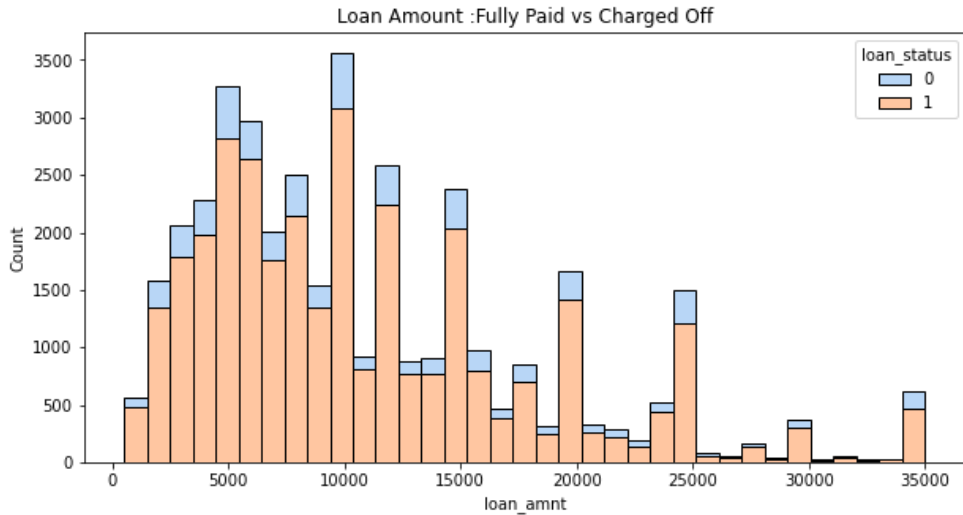


There are some extra information in the dataset like variable which will not be available during loan submission , hence won't be affecting our analysis .After removing such columns, our new dataset's shape is (39717, 38)



Final shape of dataset after removing columns which have majority of values missing or are having same value throughout(so not affecting the analysis) : 39716 rows and 25 columns

Analysis on Loan Amount , Term and Interest Rates



Loan Amount:

- People prefer round amounts while taking loans like 5000, 10000, 15000, 20000 etc. and hence no of defaulters are also more in those brackets.
- Not much relation between loan amount and defaulter can be derived

Term:

- Most of the loans are short term loans.
- Defaulters doesn't have a direct relationship with terms

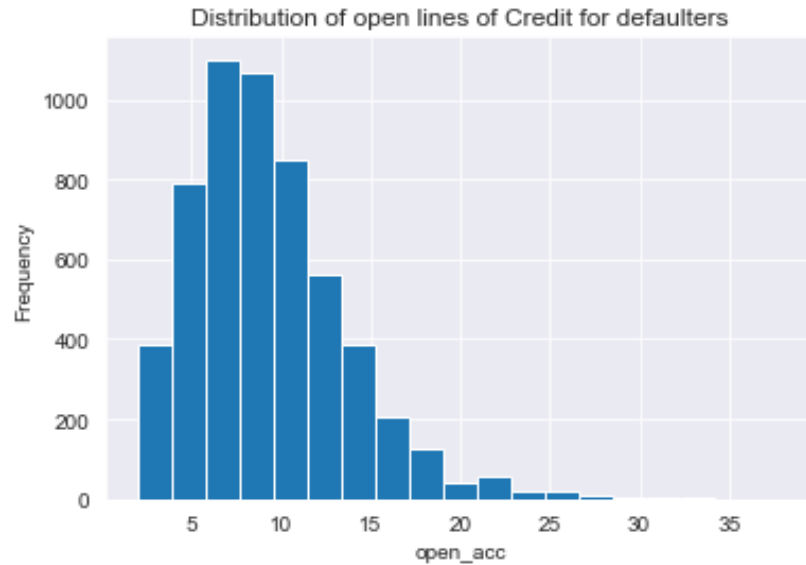
Interest Rates:

- Almost all the interest rates(low, medium and high) have defaulters
- not much relation between defaulters and Rate of Interest

Legend:

Loan Status 1 : Fully Paid
Loan Status 0 : Charge Off

Analysis on open_acc , pub_rec , revol_bal



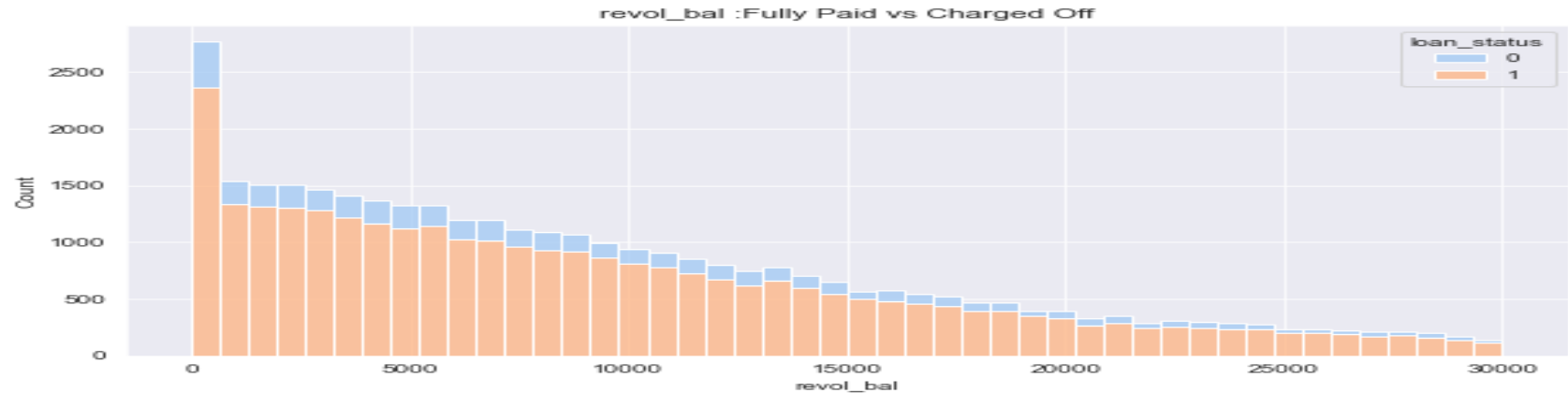
open_acc

- Most of the defaulters are people having 6-10 open credit lines in there account.
- Most of the loan appliers are also from this range only.
- very less applications from people who already have high open credit lines(more than 20)



pub_rec

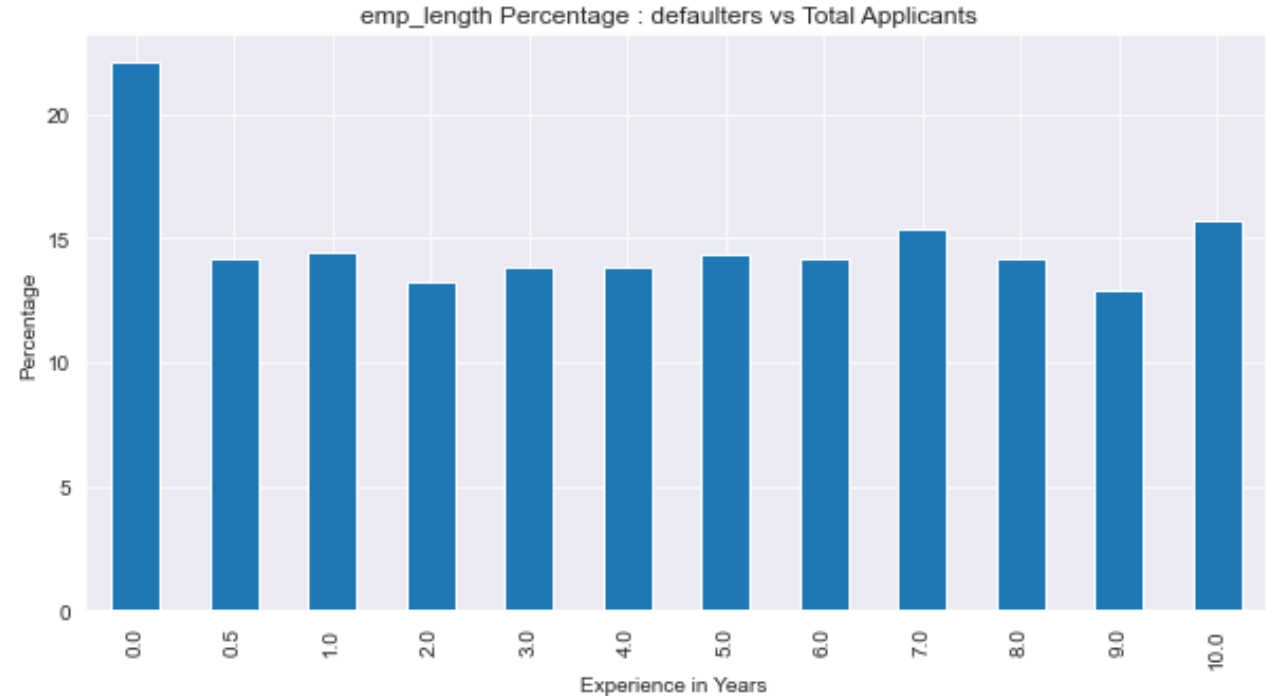
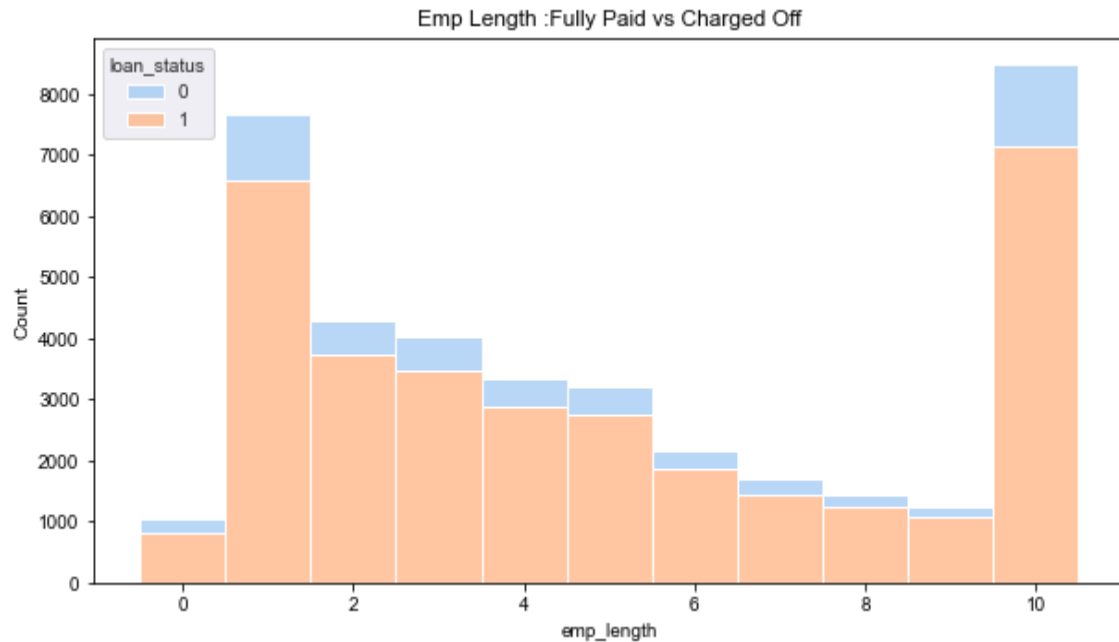
- No direct relation between No of defaulters and No of Derogatory records of the applicant



revol_bal

- A lot of people who don't have any revolving balance("0") are applying for the loan, hence defaulters are also more
- Almost all values of revolving balance have defaulters.
- No direct relation between revolving balance and defaulters

Analysis on Employment Length

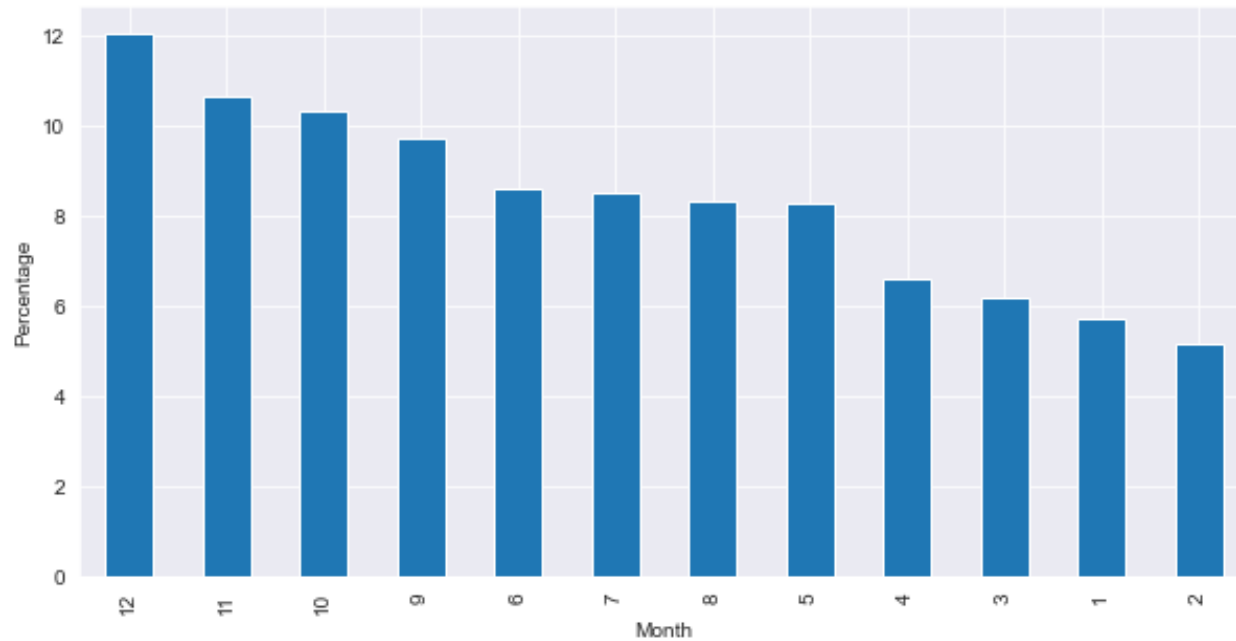


- most people who took loans have either employment length of "1 and <1" years or "10 and 10+"
- defaulters are spread out across all the employment length , **but more at 1 year and "10 and 10+"**

- 13-15% defaulters for almost all the years of experience
- people having 0 years of work exp have the most percentage of defaulters among defaulters belonging to that years of exp **which is around 22%**

Analysis on Loan Issue Month , Annual Income

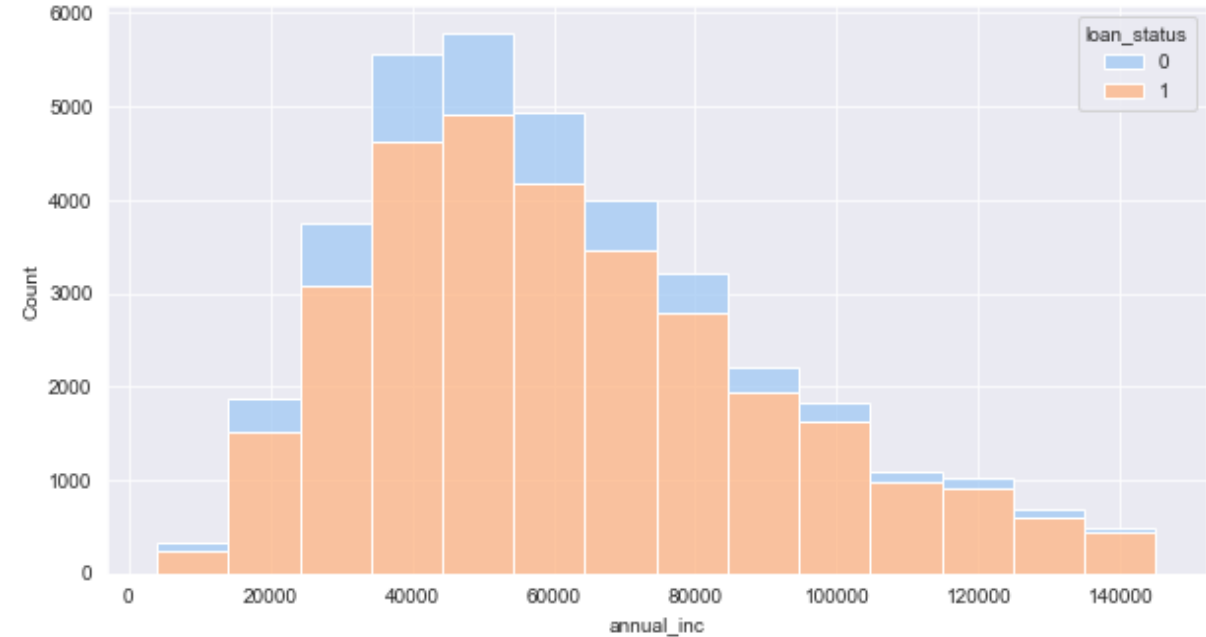
Issue date contribution to Defaulters Percentage



Loan Issue Month

- Number of loans starts increasing when the financial year ends in September and goes on till December
- As a result, **the number of defaulters are also more in those months (September to December)**
- percentage of defaulters is in the range of 12-16% across all the months.

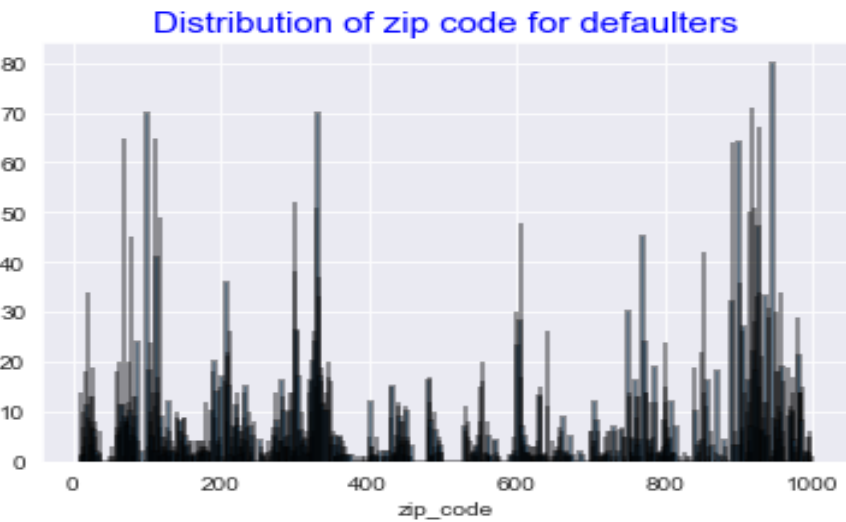
Annual Income :Fully Paid vs Charged Off



Annual Income

- defaulters are more in the lower income section specially between **30000-70000 annual income** range

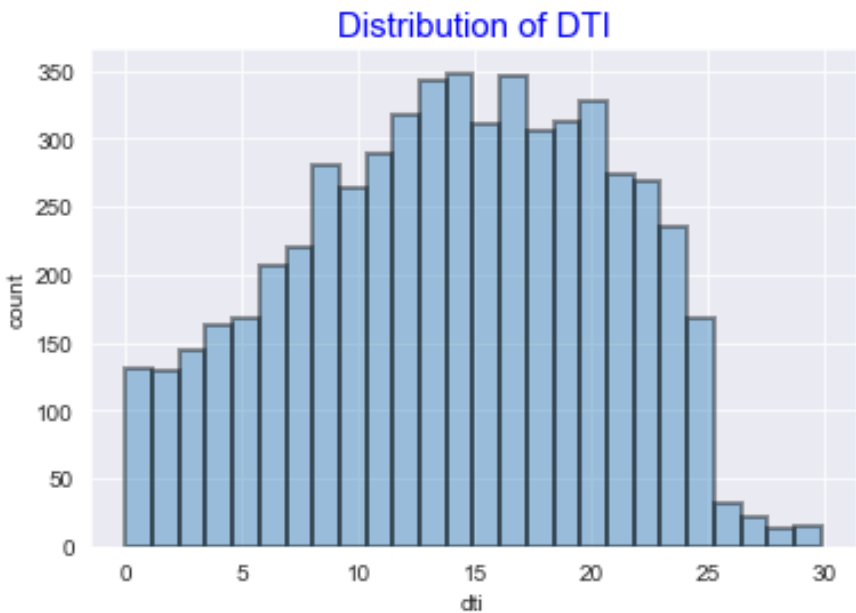
Analysis on Zip Code , DTI



Zip Code	%age
945	1.421717
917	1.261774
100	1.244002
331	1.244002
926	1.190688
70	1.155145
112	1.155145
900	1.137373
891	1.137373

Zip Code

- There are peaks at some places which means more no of defaulters
- **Following Zip code contribute more defaulters** among the 1000 zip codes

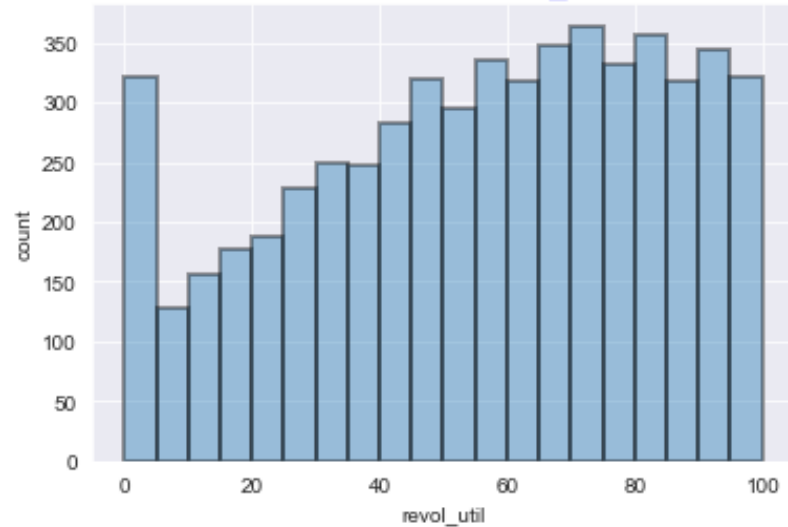


DTI

- People with DTI till 25% are applying for loan, higher than this ratio there are very few apliers
- There are considerable no of defaulters from the **range 10% to 25 % Debt to Income ratio**

Analysis on revol_util , total_acc

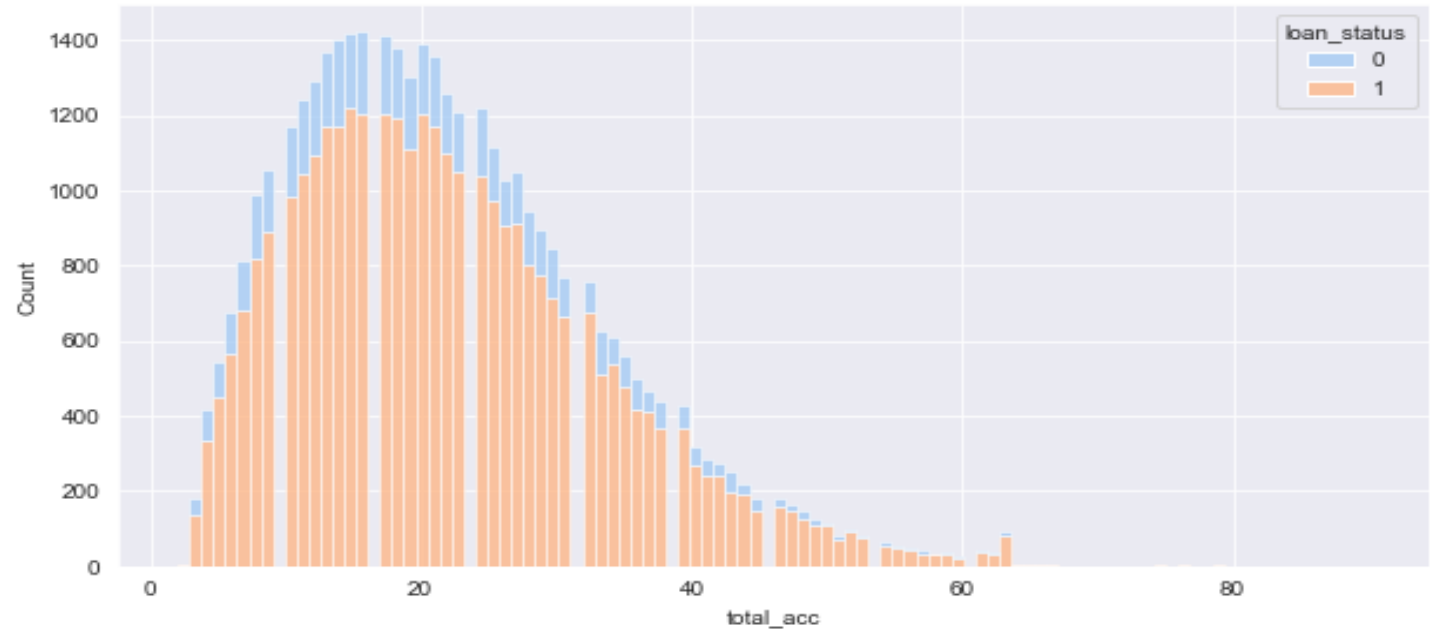
Distribution of revol_util



revol_util

- Higher number of **defaulters** have high Revolving line utilization rate

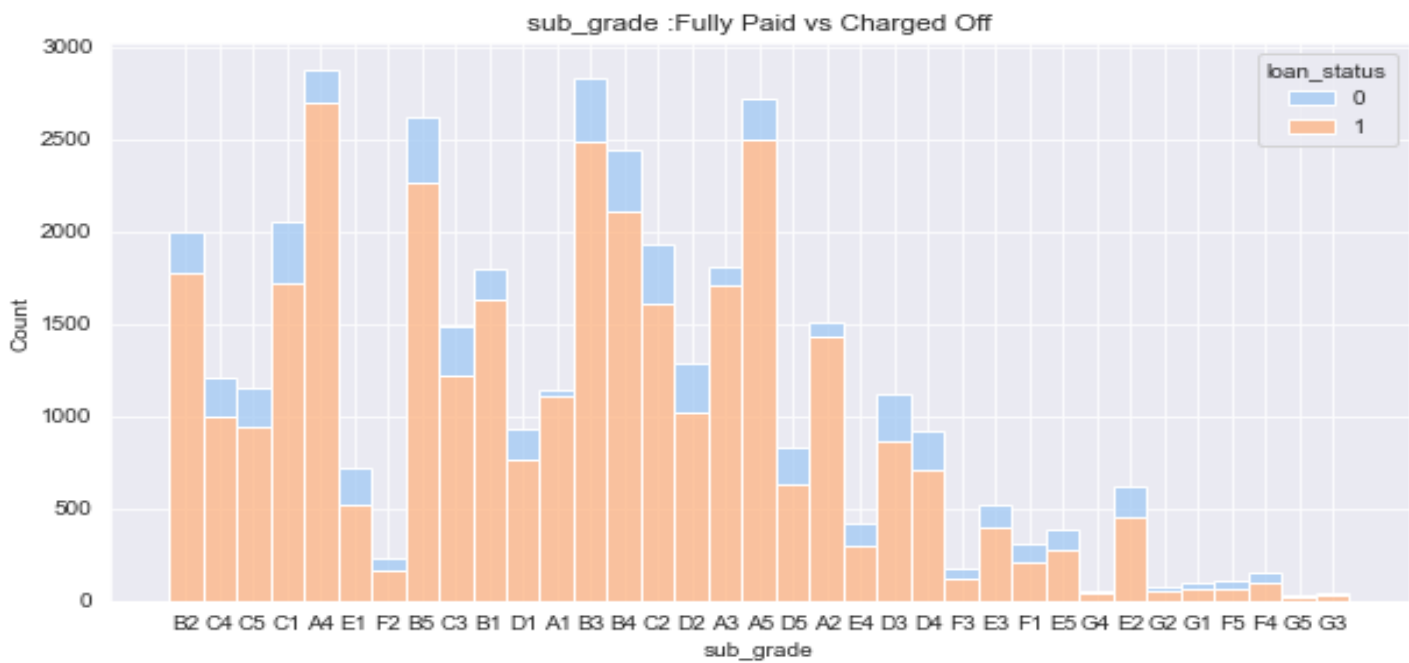
total_acc :Fully Paid vs Charged Off



total_acc

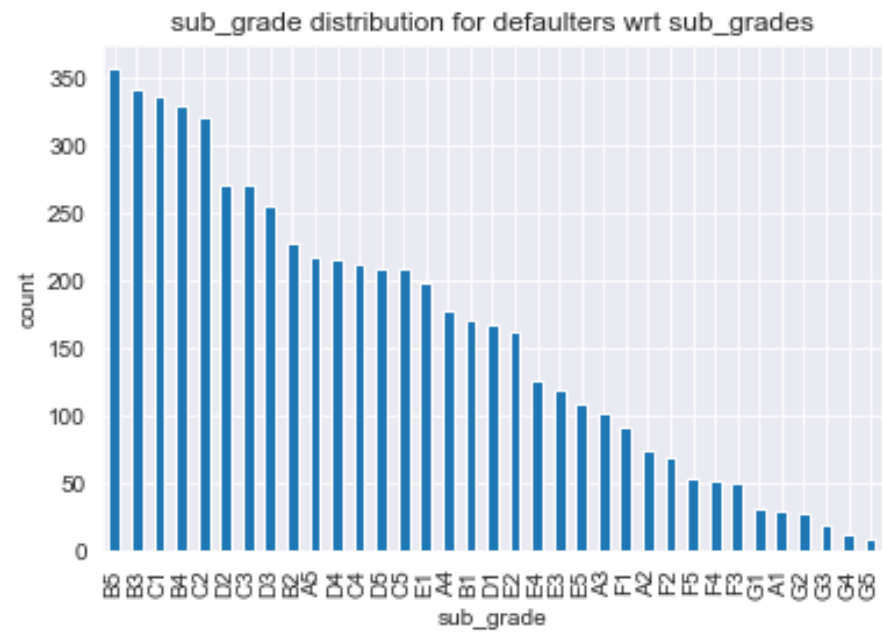
- Most of the people applying loans have already 10-30 open credit lines on there account.
- **Percentage of defaulters is also high** in those range

Analysis on grade , sub_grade



Grade

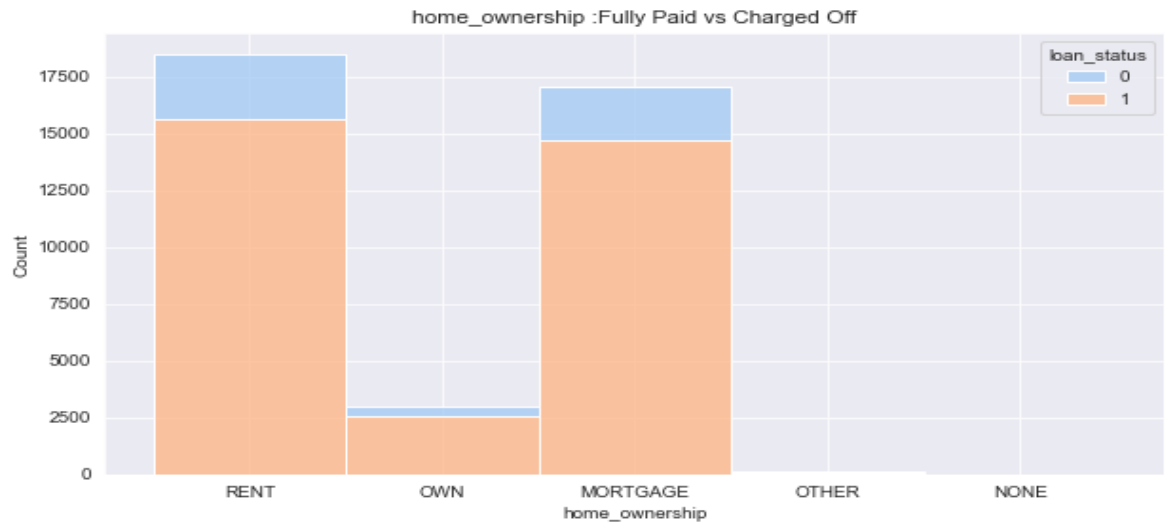
- Most of the applicants are from Grade A, B and C
- **Lower the grade ,higher the percentage of defaulters**



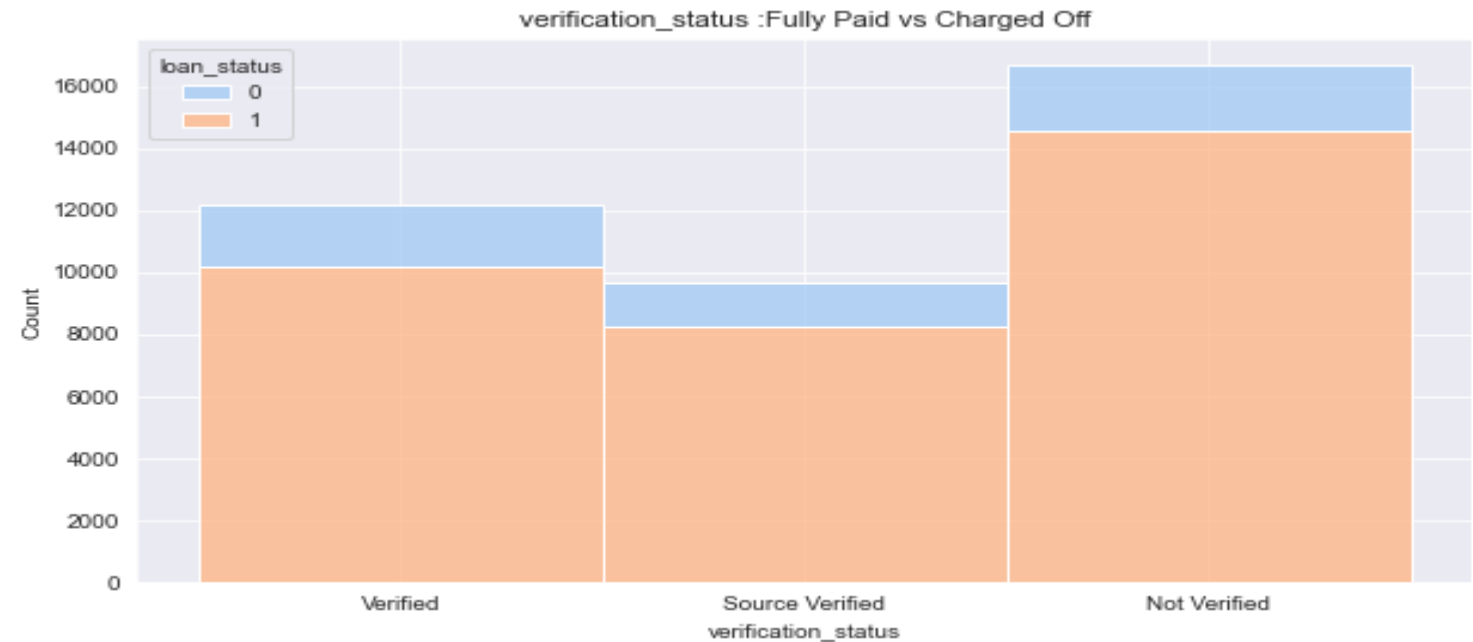
sub_grade

- There are defaulters in almost all the sub grades
- But of all the defaulters , **no of defaulters is relatively large in anyone having sub grade \geq B2 i.e B2,B3,...,C1,C2...D1 etc**

Analysis on home_ownership , verification_status



- home_ownership**
- Most of the defaulters are people who generally Rent or Mortgage

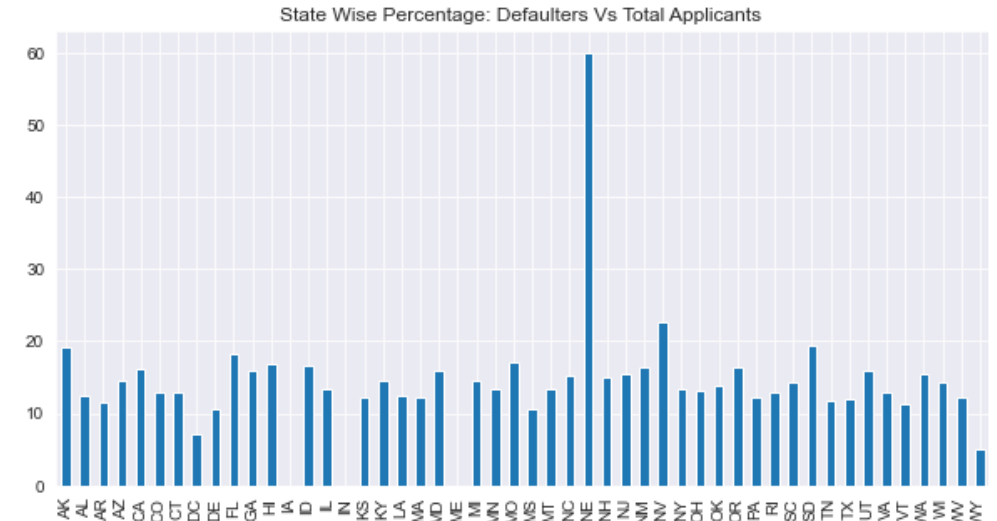
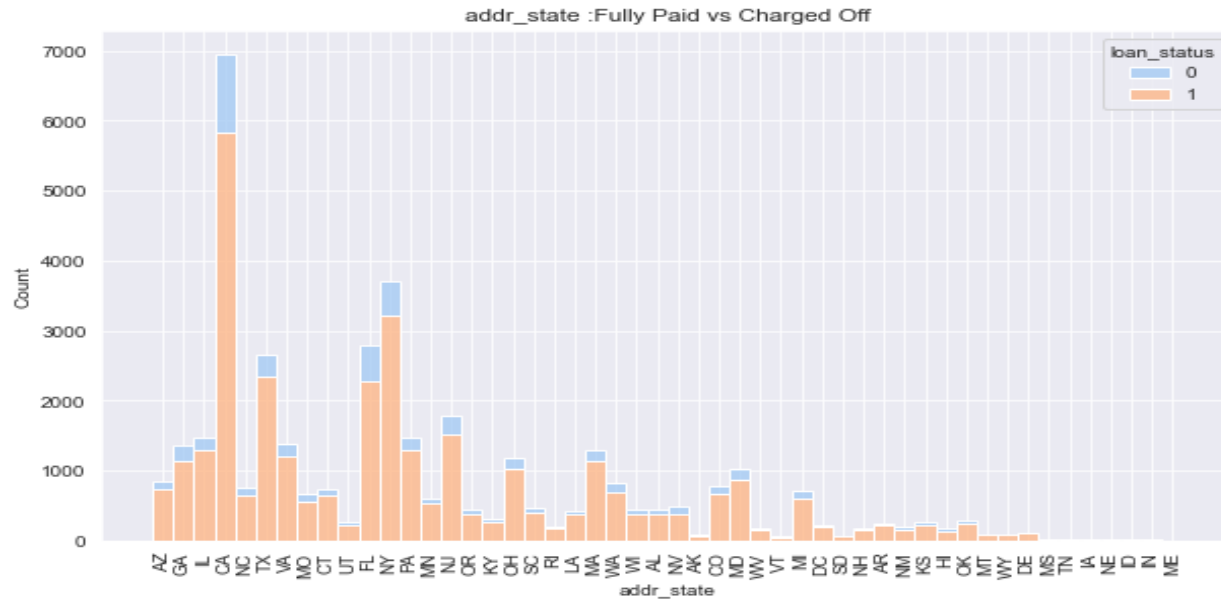
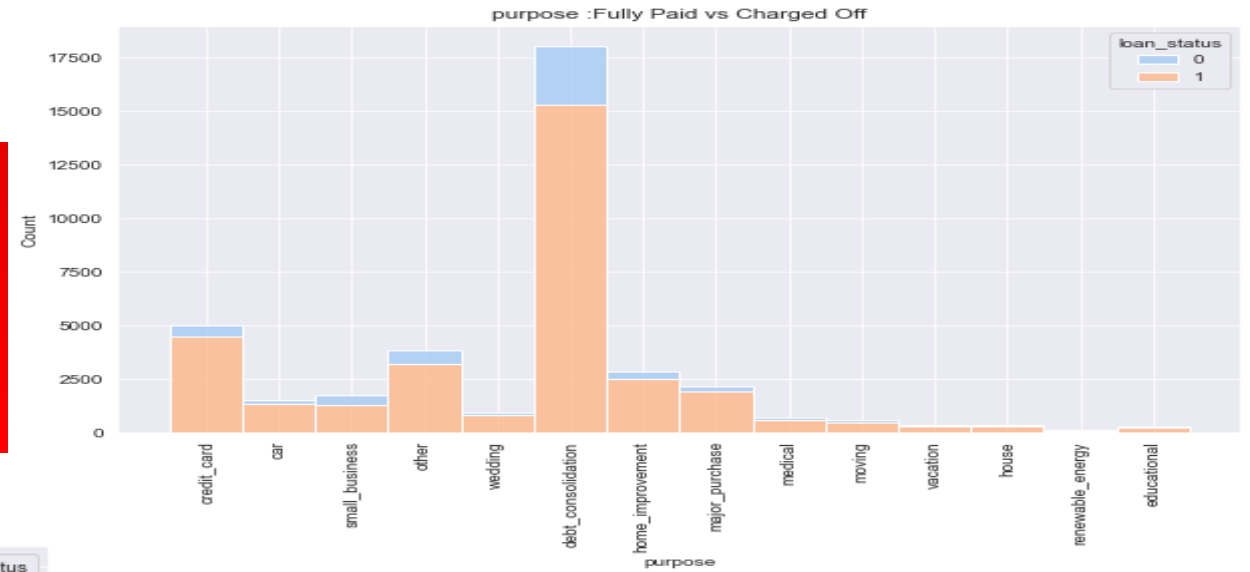


- verification_status**
- Very less Defaulters when the applicant's income is source verified.
 - **Not Verified and Verified applicants have a higher number of defaulters.**

Analysis on purpose , addr_state

Purpose

- Most of the loans taken are for **Debt Consolidation, resulting in more defaulters**
- People having "Other" as a purpose have high percentage of defaulters



addr_state

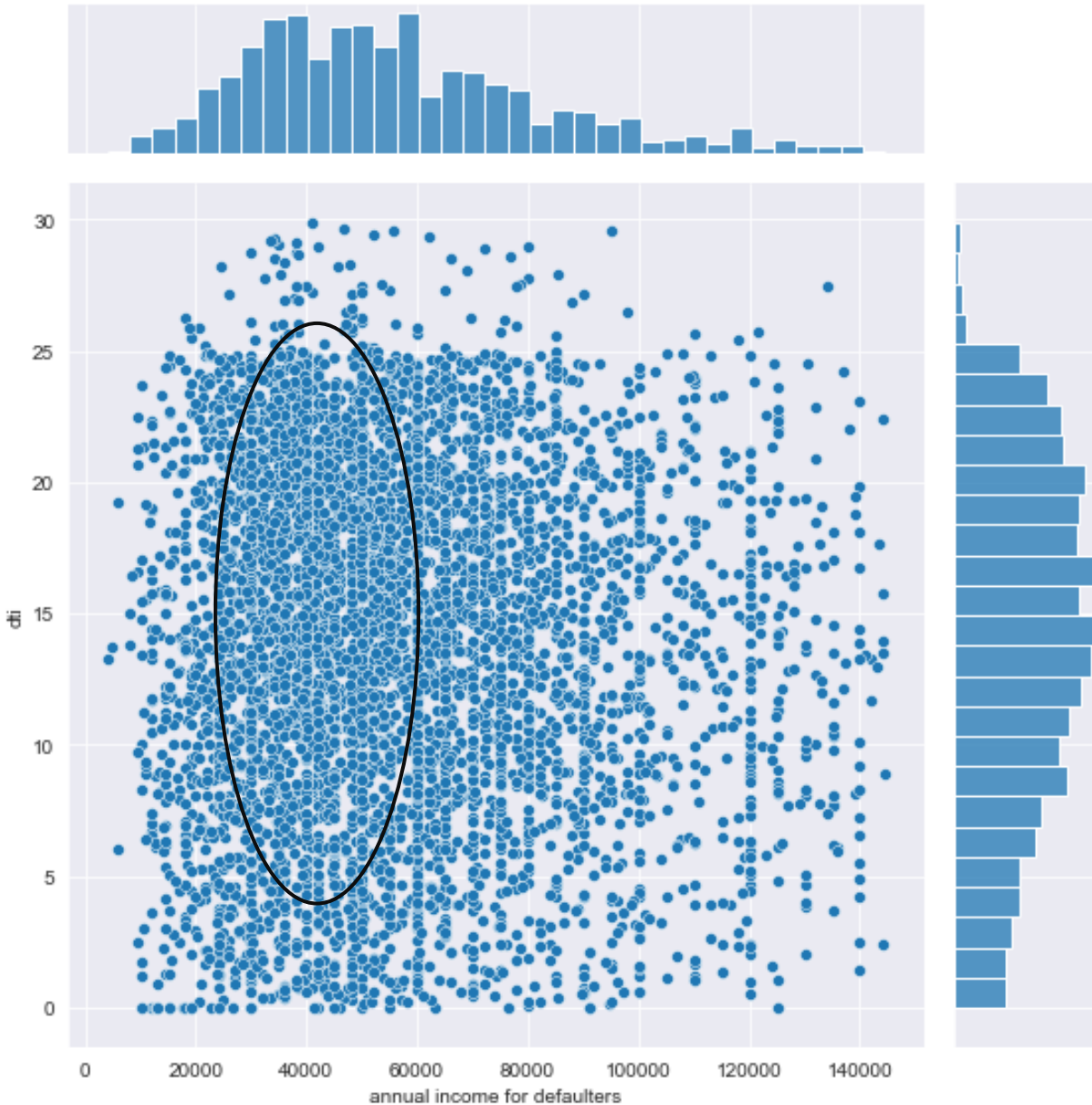
- States like **CA, NY, FL, TX** are having more no of applicants and hence more no of defaulters.
- Some states have relatively higher percentage of defaulters like **AK, FL, NV, SD** with around 20% of applicants defaulting.
- One peak of around 60% for a state **NE** , but the number of applicants were very less (3 defaulters out of 5)
- Few states having **0 defaulters** like **IA, IN, ME**

Bivariate Analysis on sub grade and Annual Income



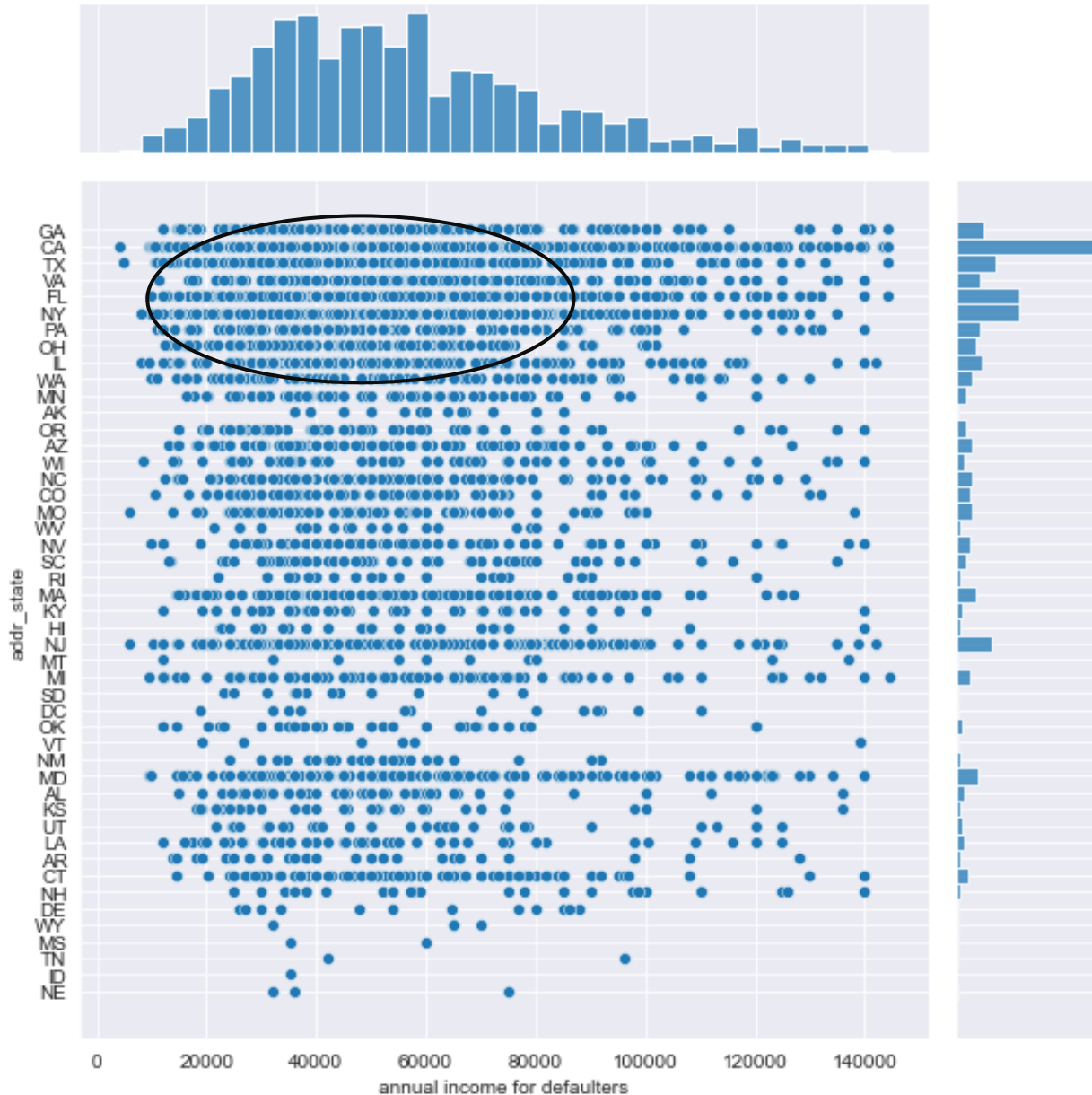
- We see a cluster of applicants who have defaulted belonging to a lower sub grade and have lower incomes
- **We can infer that there is a direct relation between income and sub grades, lower income is usually assigned lower grades among the defaulters**

Bivariate Analysis on dti and Annual Income



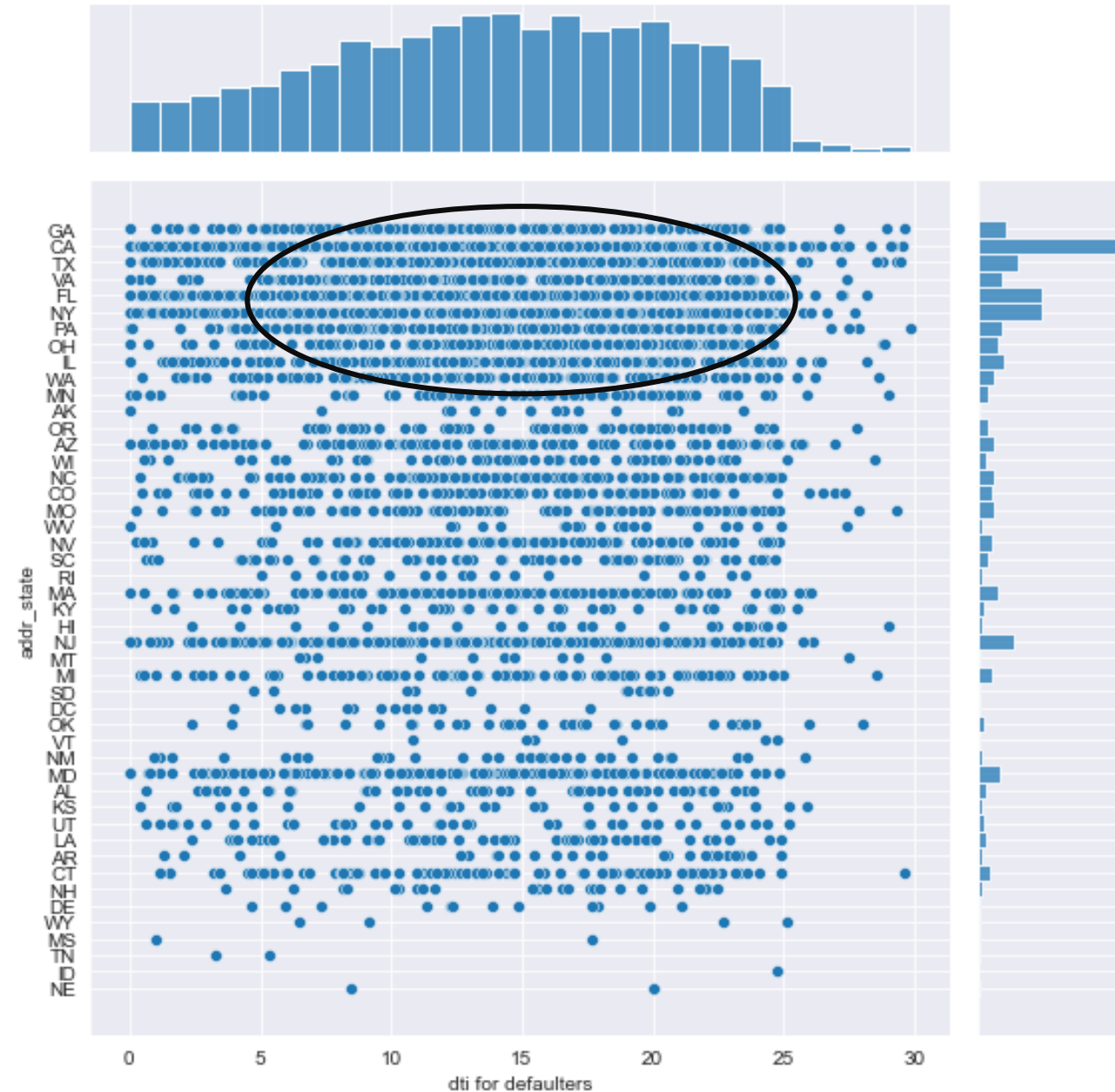
- We see that among the defaulters there are a lot of low-income group with considerable high dti rate 10-25%

Bivariate Analysis on addr_state and Annual Income



- we see a cluster of points for lower income groups and areas like (GA,CA,TX,VA,FL,NY,PA,OH,IL) defaulting

Bivariate Analysis on addr_state and dti



- we see a cluster of points for people in areas like (GA,CA,TX,VA,FL,NY,PA,OH,IL) and having a dti of 10-25% defaulting

Conclusion

Driving Factors		Action Item
emp_length :	Applicants with 0 years of experience tend to default more	
issue_d	The number of defaulters are more in the months of September to December, so need to verify them thoroughly.	
annual_inc	Lower income section specially between 30000-70000 needs to be verified thoroughly	
zip_code	Zip codes(945,917,100,331,926,70,112,900,891) contribute more defaulters among the 1000 zip codes, applicants from that area should be verified thoroughly	
dti	More defaulters have 10% to 25 % Debt to Income ratio	
revol_util	Defaulters have high Revolving line utilization rate	
total_acc	Defaulters usually have 10-30 open credit lines on there account	
grade	Lower the grade(<A) ,higher the percentage of defaulters	
sub_grade	no of defaulters is relatively large in anyone having sub grade >= B2 i.e B2,B3,...,C1,C2...D1	
home_ownership	Most of the defaulters are people who generally Rent or Mortgage, they need to be verified thoroughly	
verification_status	Not Verified and Verified applicants have a higher number of defaulters, needs to be verified thoroughly	
purpose	Applicants with Debt Consolidation as purpose needs to be verified thoroughly	
addr_state	Applicants from few states like CA, NY, FL,TX, AK, FL, NV, SD needs to be verified	

Thank you

