

GRAMENER CASE STUDY

Submission

Group Name:

1. Richa Goel
2. Indranil Nanda
3. Ankur Johari
4. Ravi Ranjan Kumar

Business Analysis Objectives

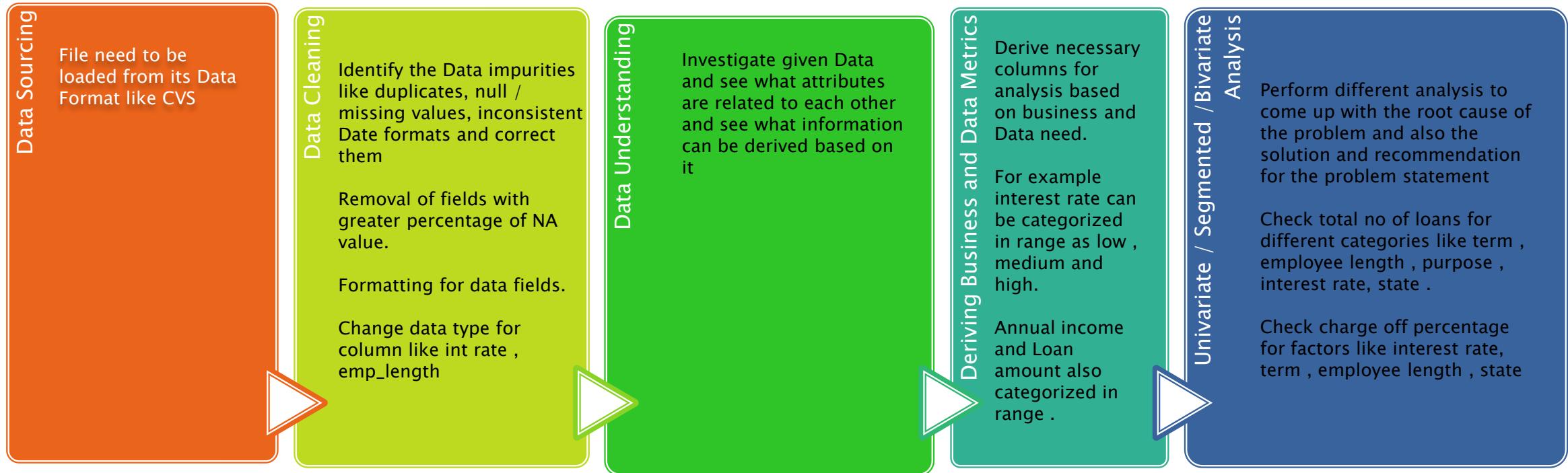
Problem Statement:

When a consumer finance company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. There could be two types of risks associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Identify the driving factors to identify new borrowers that would likely default on their loans.

Problem solving methodology



Understanding & Cleaning the Data

- As per given data ,There are total 39717 loan request approved in 2018.
- We have below types of data types in our given data
 - Float64 – 74 variables
 - Int64 – 13 variables
 - Object – 24 variables
- There are 111 column in provided data set . Below are some important column which are used for Analysis:-

Field Details	Description	Additional details
Loan status	Loan status of a loan , it can have three value :-Fully paid , Charged off and Current	Fully paid: Applicant has fully paid the loan (the principal and the interest rate) Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan
addr_state	The state provided by the borrower in the loan application	
annual_inc	The self-reported annual income provided by the borrower during registration.	
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.	
grade	LC assigned loan grade	
home_ownership	The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.	
int_rate	Interest Rate on the loan	

Cleaning the data

- ▶ We have 54 columns that have 100 % missing value like bc_open_to_buy , acc_open_past_24mths , bc_util .
- ▶ Removed all the columns which has missing value for more than 90% of data like tot_coll_amt, open_acc_6m.
- ▶ Removed all the columns where we have identical value for all rows for ex:- application_type
- ▶ Removed Column that had unique value for all observations as url , Desc , Member id
- ▶ Removed column that was masked , for example zipcode
- ▶ Updated Employee length column to maintain data consistency.

Derived Metrics : Business and Data Understanding

Below additional column can be derived based upon Business and Data understanding :

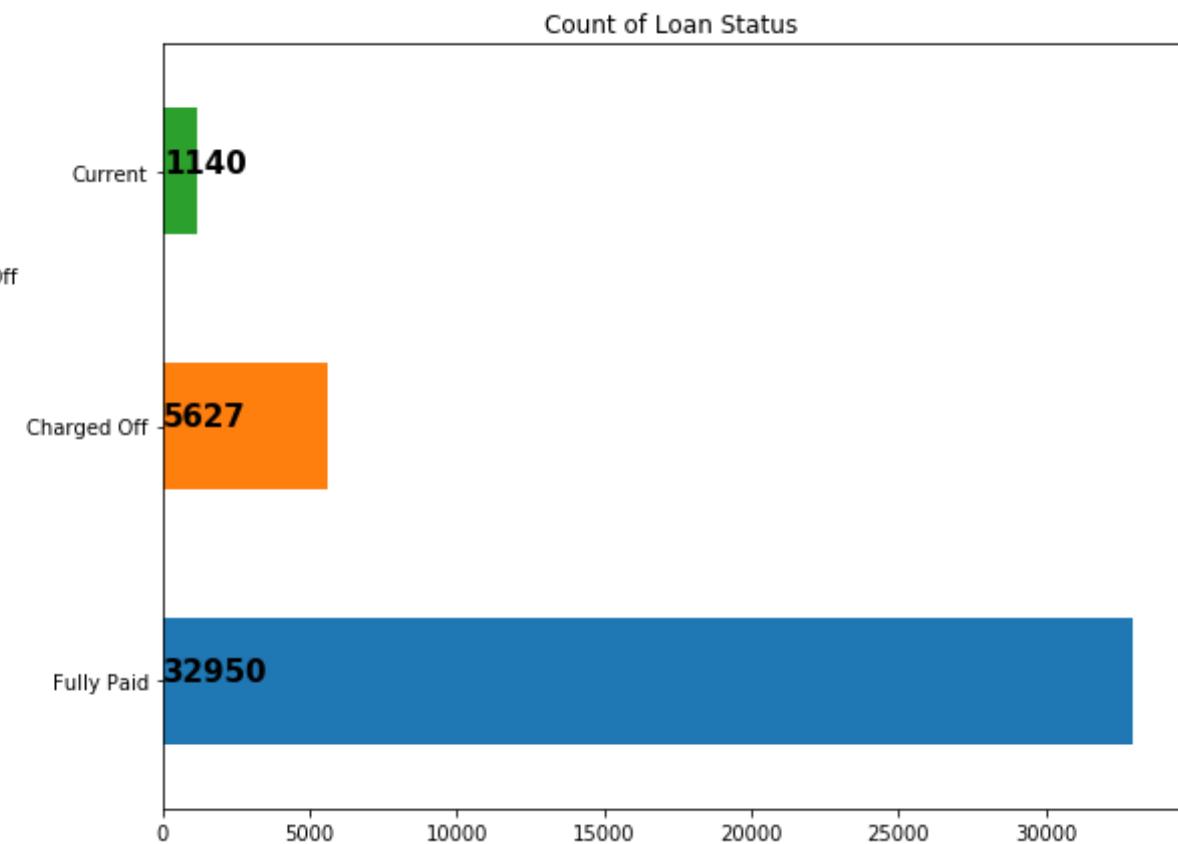
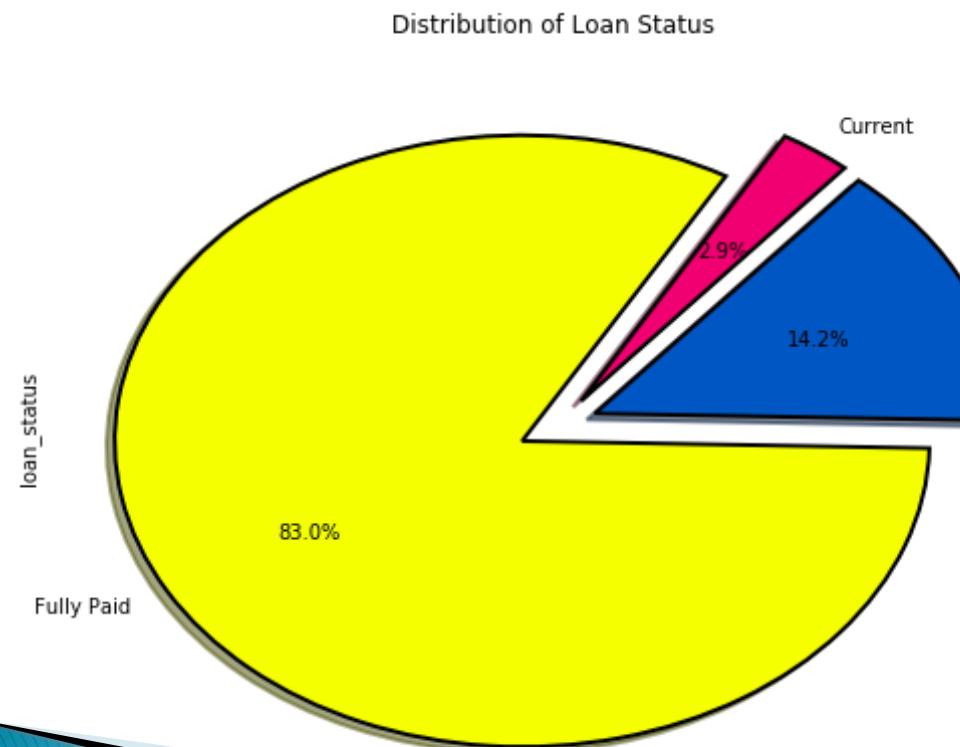
- ▶ **Interest rate range** :-Interest rate can be divided in three category , Low -1-9 , Medium :-9 to 15 and High ;-> 15
- ▶ **Credit_age_year** ;-It is derived from earliest_cr_line.
- ▶ **Annual income Range** ;-Annual income is divided in ten interval of 12K .
- ▶ **Loan Amount Range** :-Loan amount is divided in ten different slot of 2800 K.
- ▶ **Inquiry Range** :- Categorized the inquiry into , '0-2','2-4','4-6','6-8'.
- ▶ **Revol_UTL_Range**;-revol_util is divided in ten different slot of 10%.
- ▶ **DTI_Range**;-DTI_Range is divided in 6 different slot of 5% each.

Assumption

- ❑ Considered missing value and <1 years as 0(years) and > 10 years as 10 for emp_length.
- ❑ Annual income :-Removed annual income more than 1200000 to remove outlier
- ❑ Loan amount :- Removed observation for loan amount > 30K to remove outliers.
- ❑ Considered interest rate range as Low -1-9 , Medium :- 9 to 15 and High ;-> 15

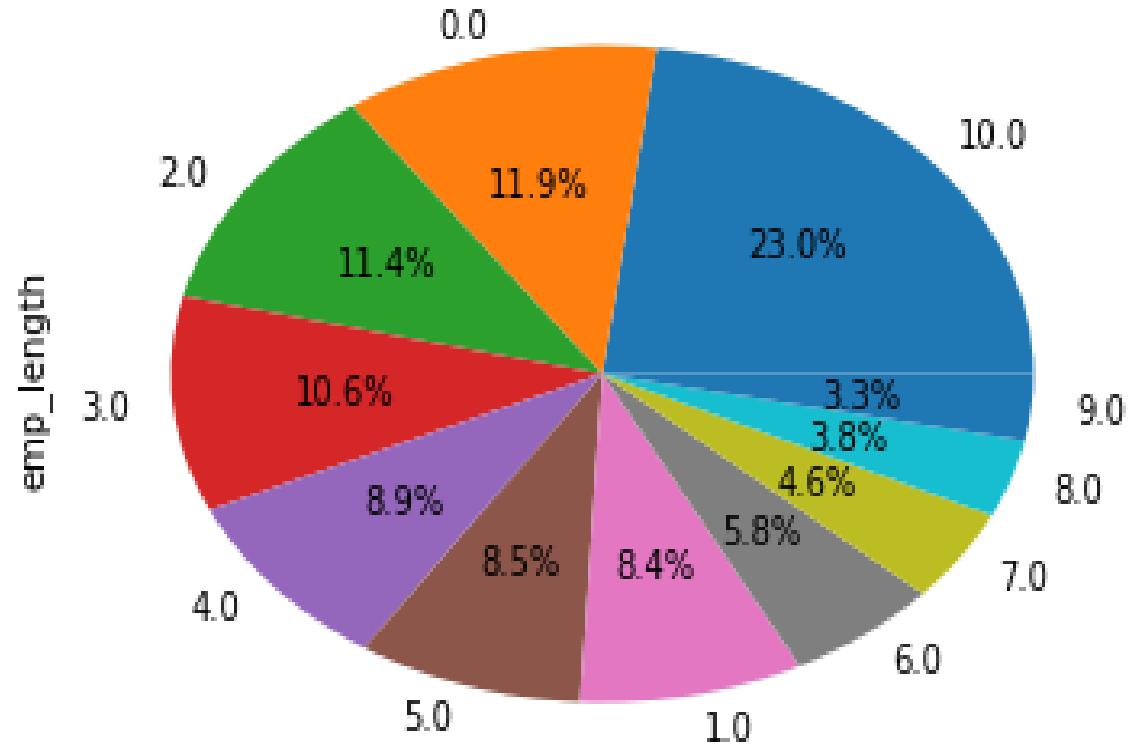
Data Visualization

Loan Status distribution which shows that approx 14.2 % loans are getting defaulted which is a pain area of our organization which we need to analyzed.



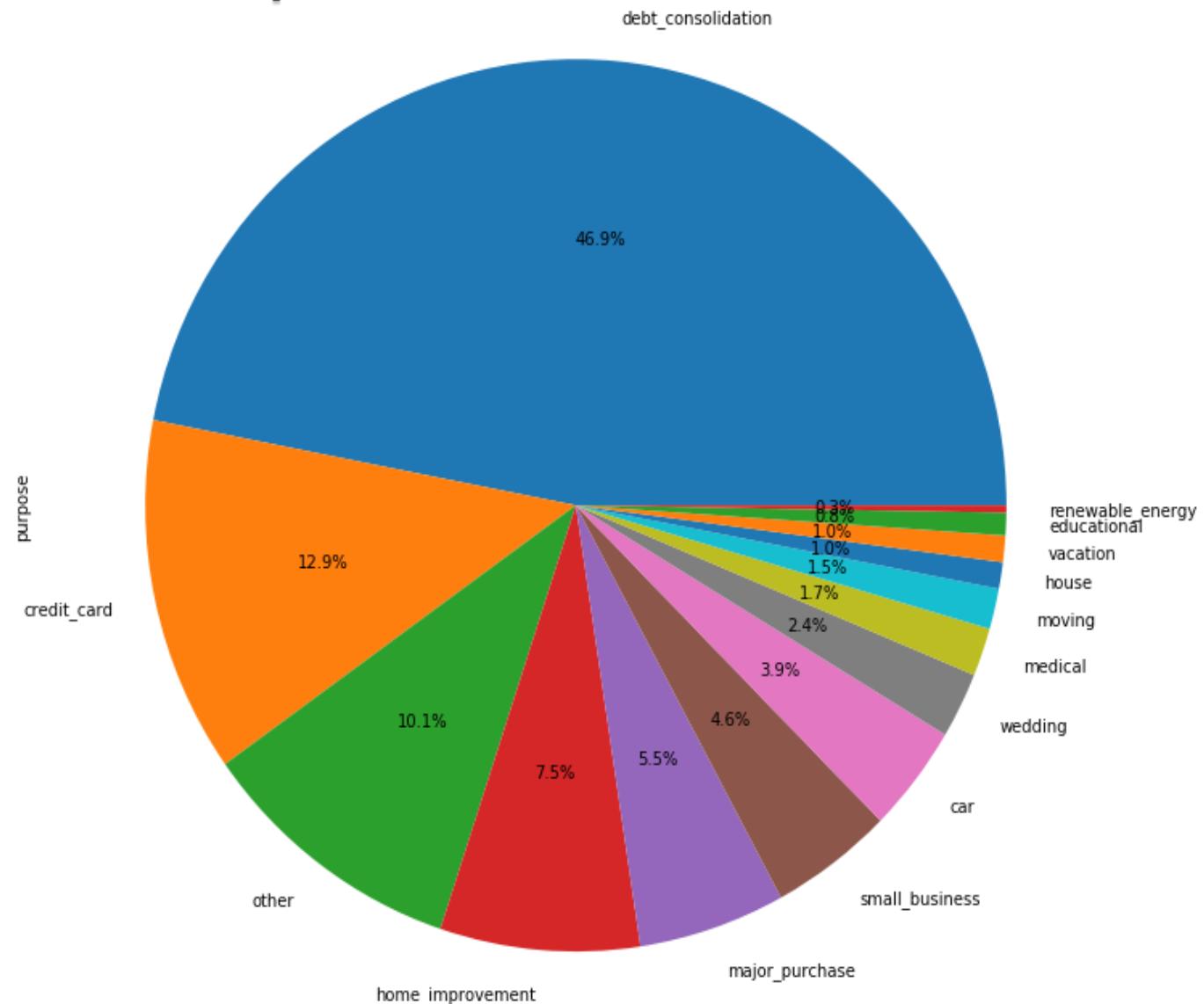
Univariate Analysis – Employment Length

Approx 23 % loans are given to Employees having served for 10 or more than 10 Years.



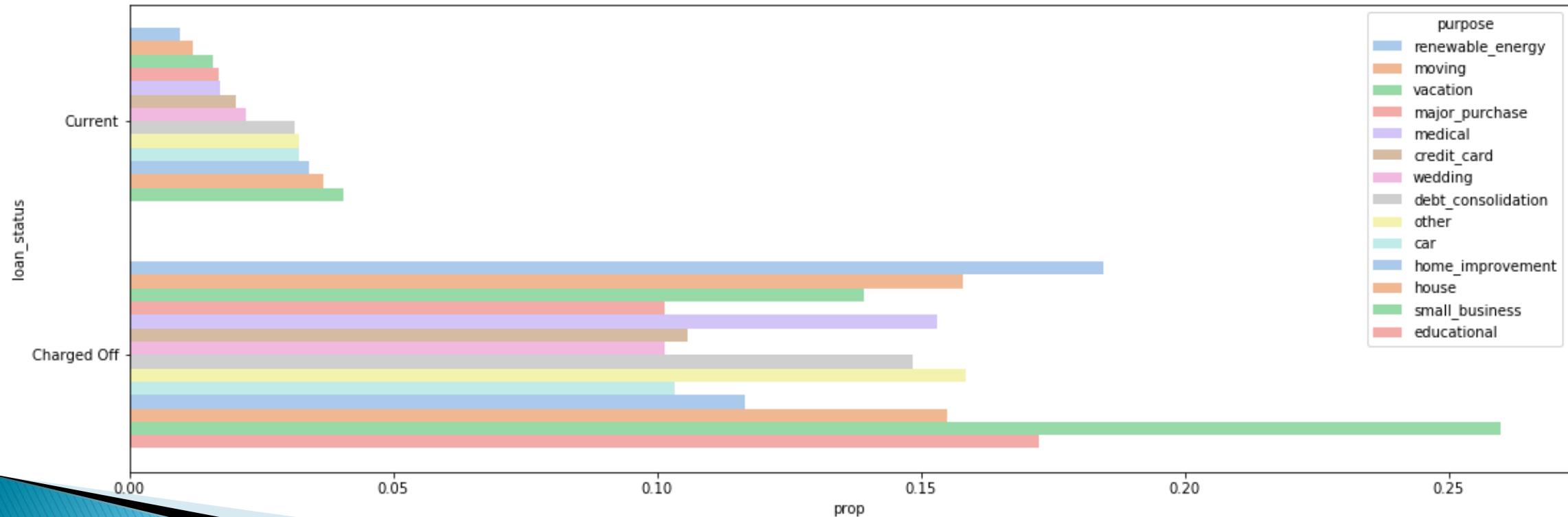
Univariate Analysis – Loan Purpose

Univariate Analysis on Loan Purpose shows that majority of loans requested for Debt Consolidation ,followed by Credit Card.



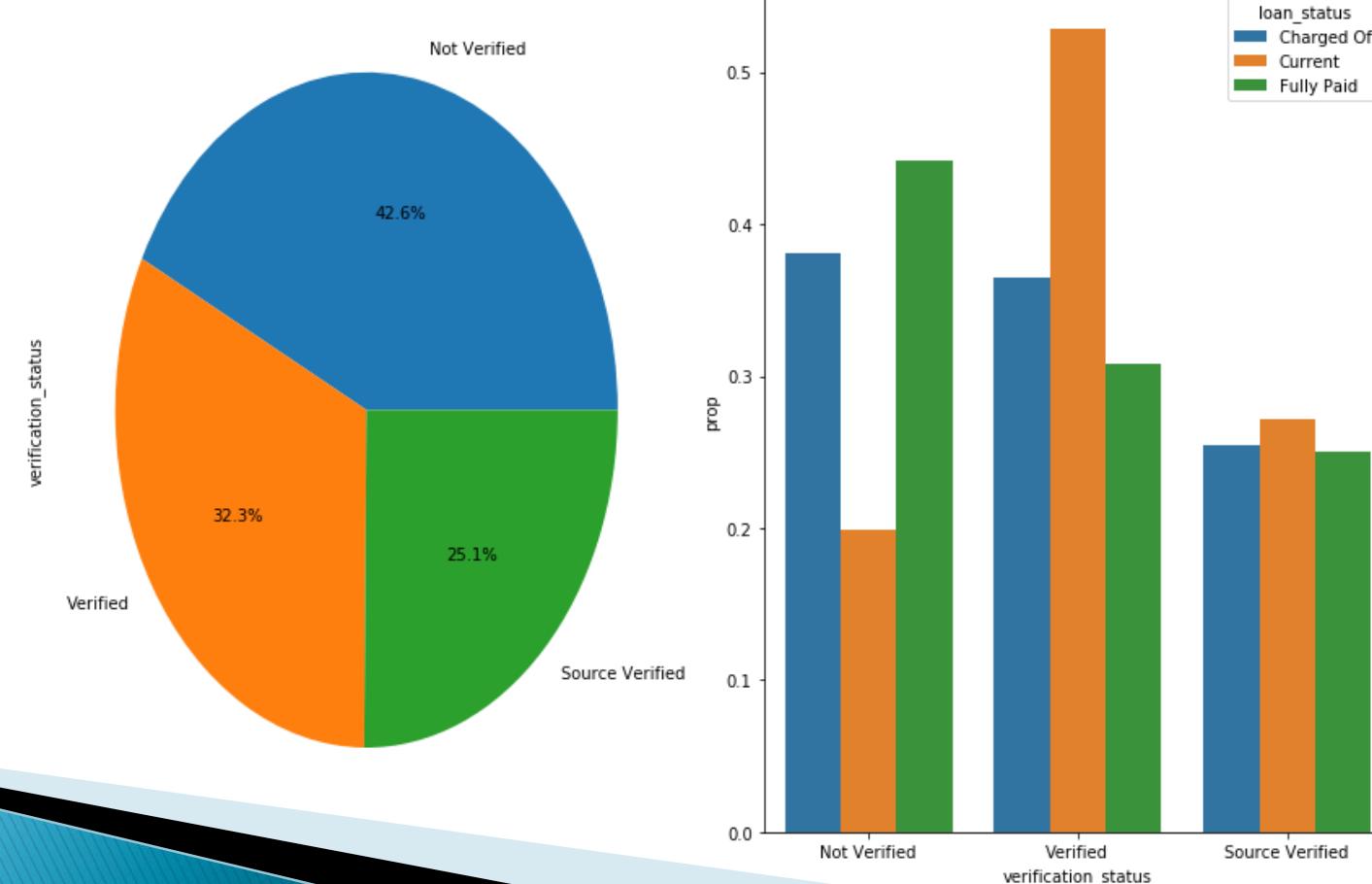
Segmented Univariate Analysis – Loan Status vs Purpose

Charged Off % is significantly higher for Small Business where as Debt Consolidation volume is high but Charged Off % is low.



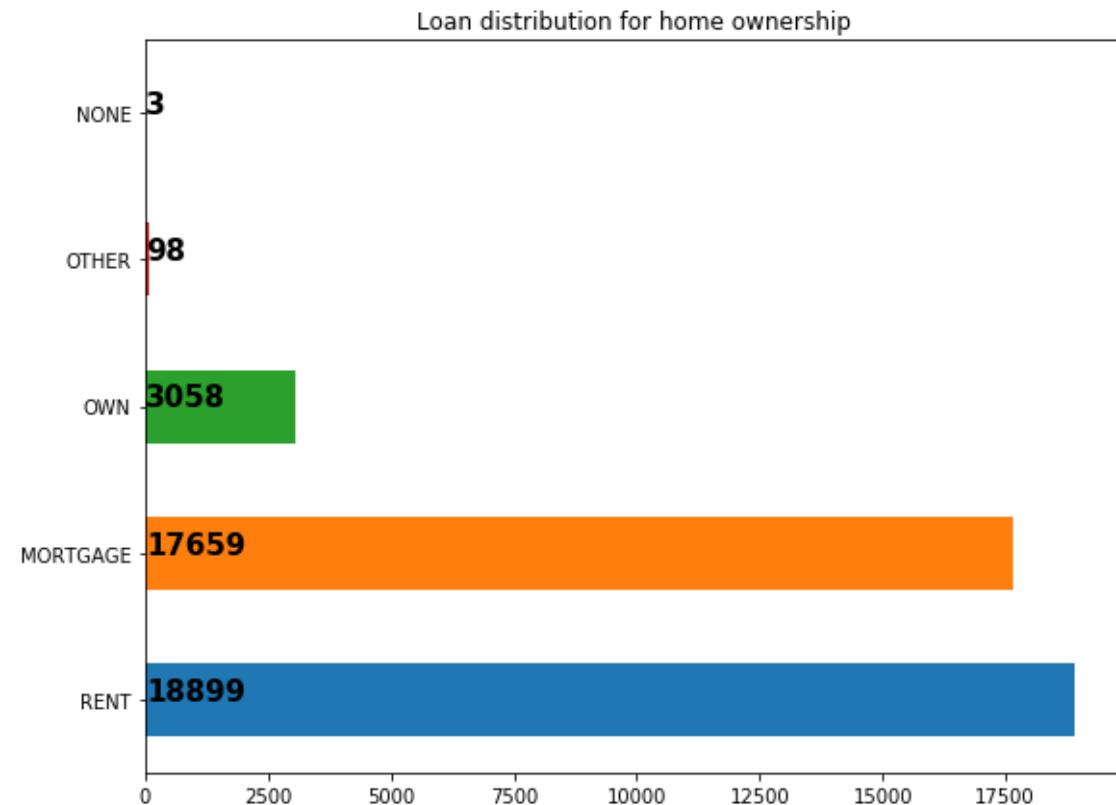
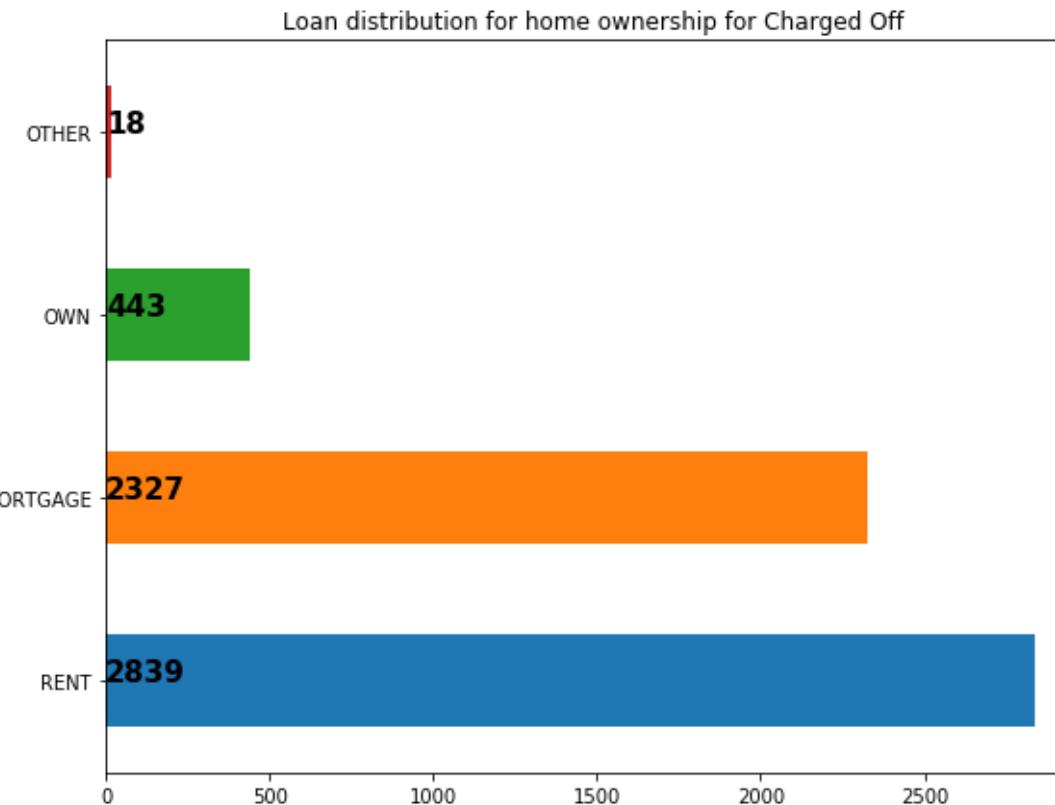
Univariate Analysis – Income Source Verification Status

Univariate Analysis on Income Source Verification Status indicates that, around 42.6 % of loan requests were issued even though income source was not verified.



Univariate Analysis – Home Ownership

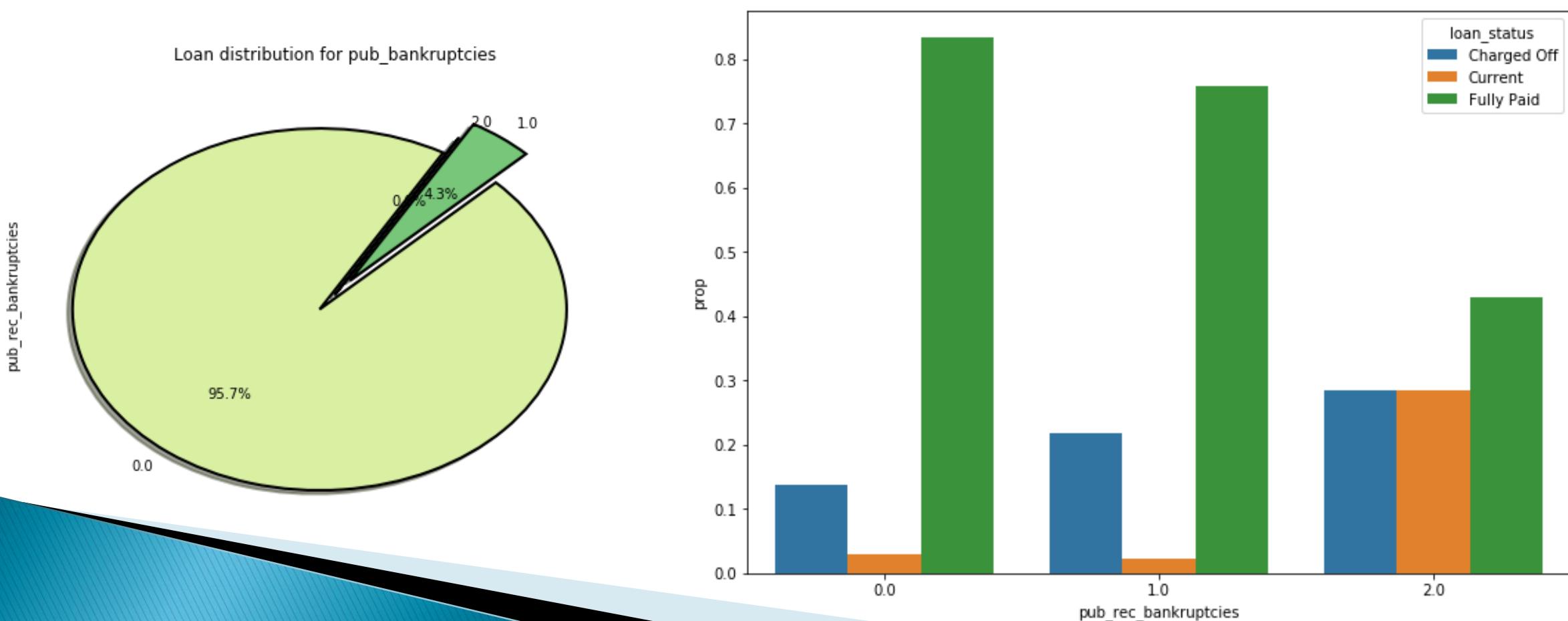
Univariate Analysis on Home Ownership indicates that, people with own home took significantly lesser number of loans compare to people with mortgage and rent.



Univariate Analysis – Loan distribution for pub_bankruptcies

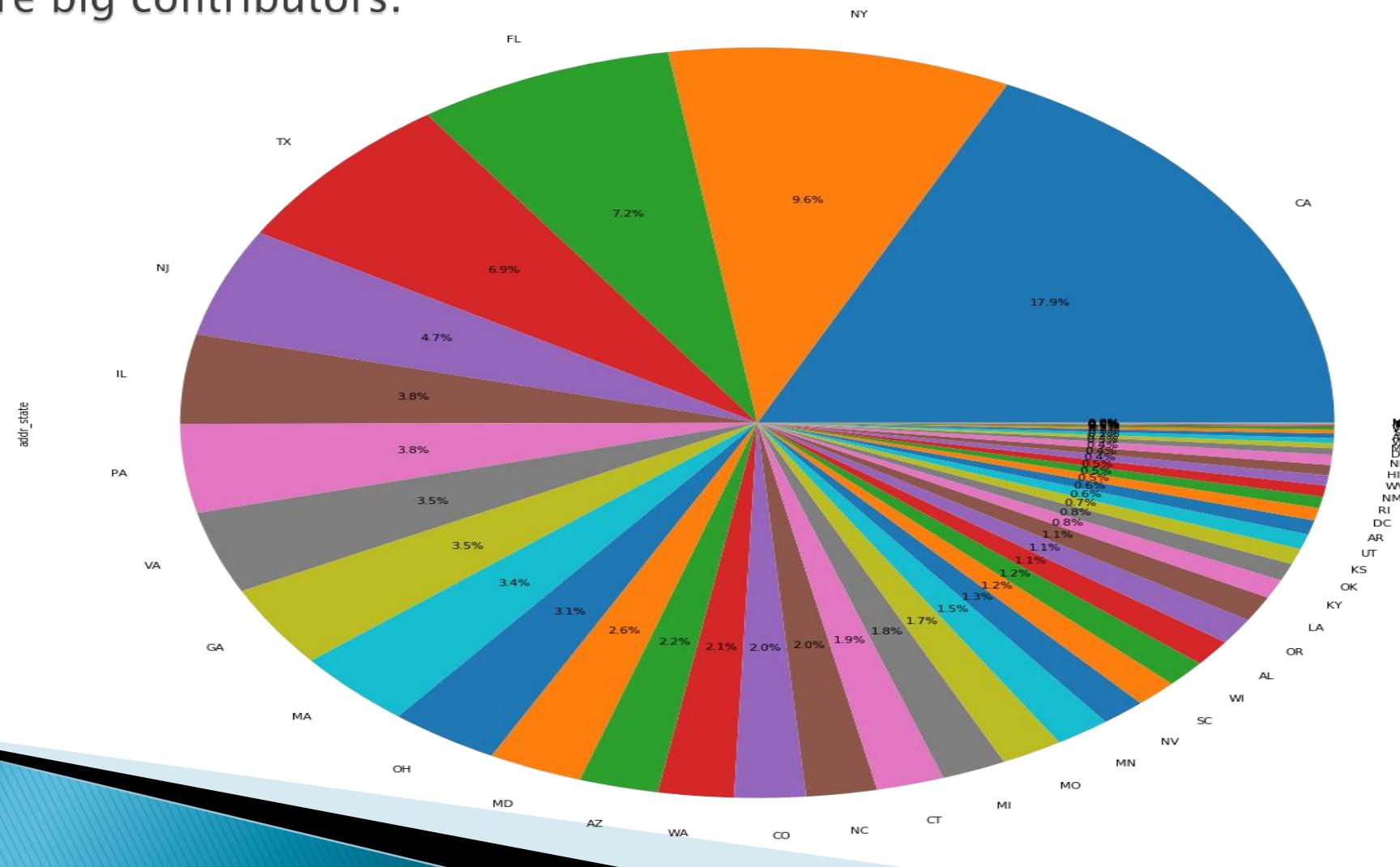
Univariate Analysis on Loan distribution for pub_bankruptcies indicates that maximum loans are approved for people who has pub_rec_bankruptcies is 0 which is less risky.

Also ,charge percentage are high for public bankruptics is 2.



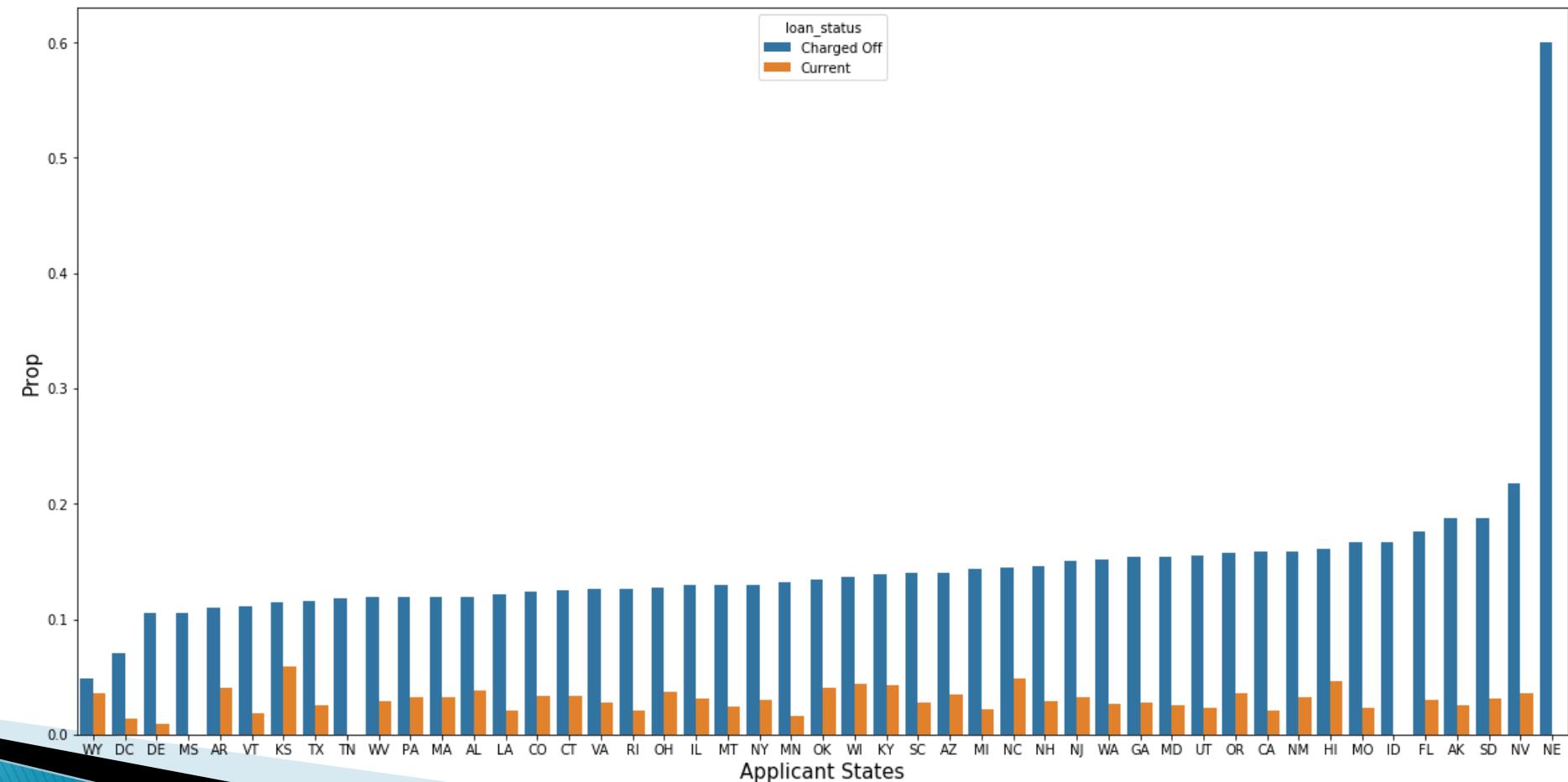
Univariate Analysis – State

Univariate Analysis distribution of States of the Applicants which shows California , New York and Florida are big contributors.



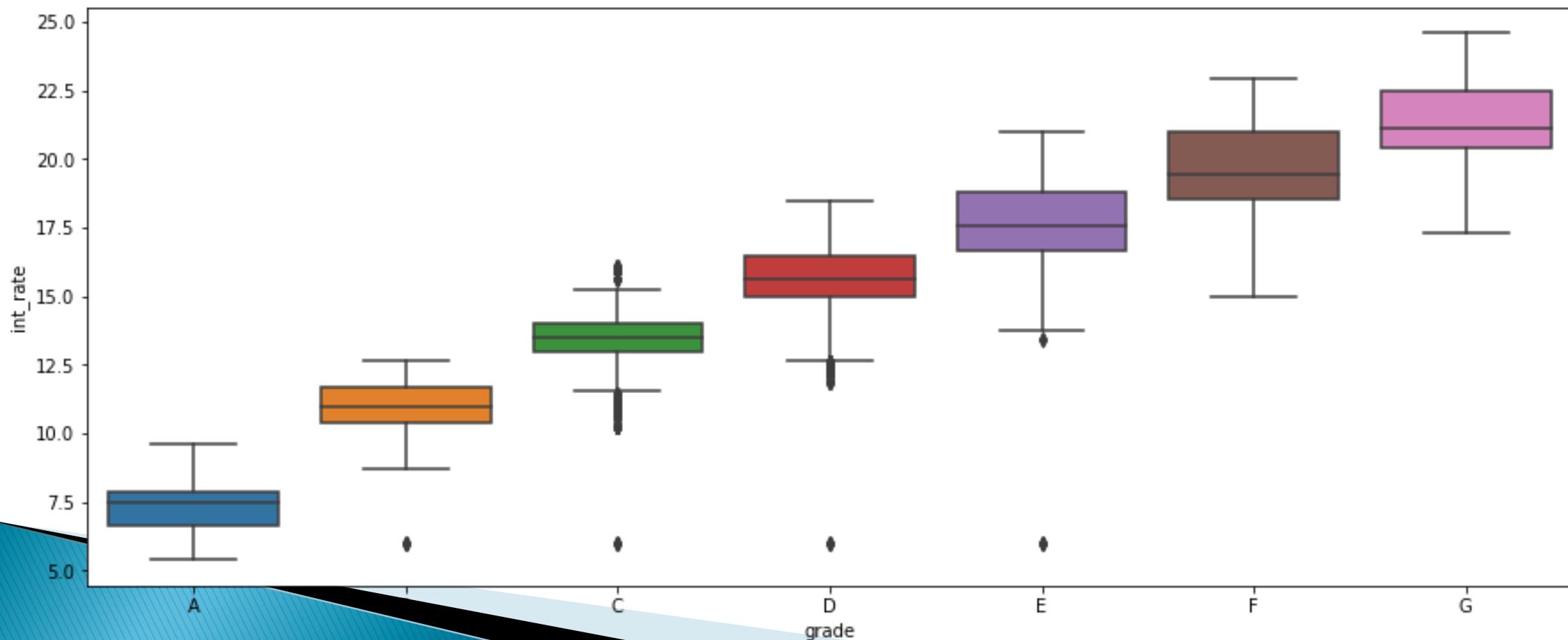
Segmented Univariate Analysis – Loan Status vs State

It is observed that charged off percentage is very high for Nebraska (60%) but loan for this state is very less which is good , another state which has 1.3% of loan and higher charged off percentage is Nevada.



Bi-variate Analysis – Interest Rate vs Grade

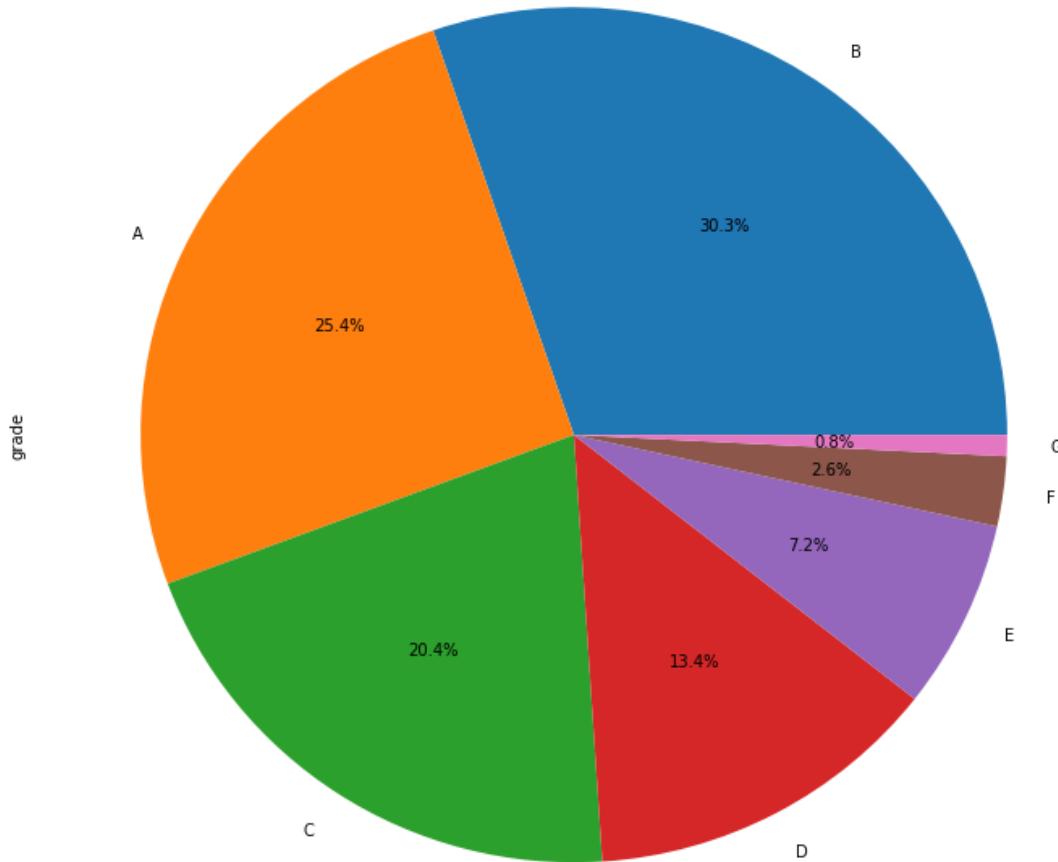
In this Bi-Variate Analysis we can infer that Interest rate is higher as we moved the Grades from A to G .



Univariate Analysis – Loan Grade

Univariate Analysis on Loan Grade which shows that maximum loan is approved for grade status 'B' which is a good factor and reduce percentage of risk.

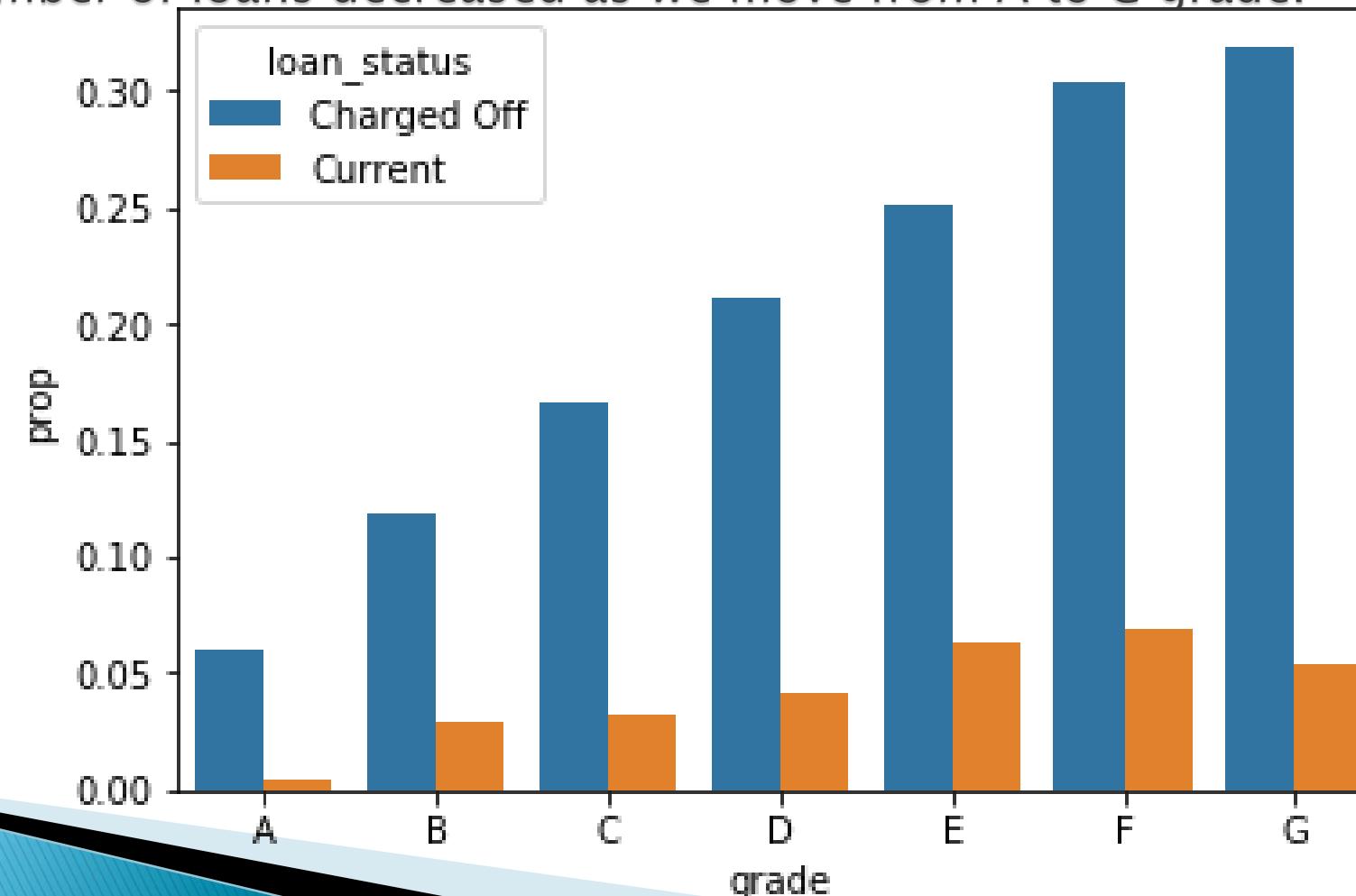
As we moved from Grade A to G , number of approved loans dropped significantly.



Segmented Univariate Analysis – Loan Status vs Grade

The Chance of being Charged off is higher for Higher Grades.

But this makes sense by seeing the previous plots where we can see the interest rate increases and number of loans decreased as we move from A to G grade.

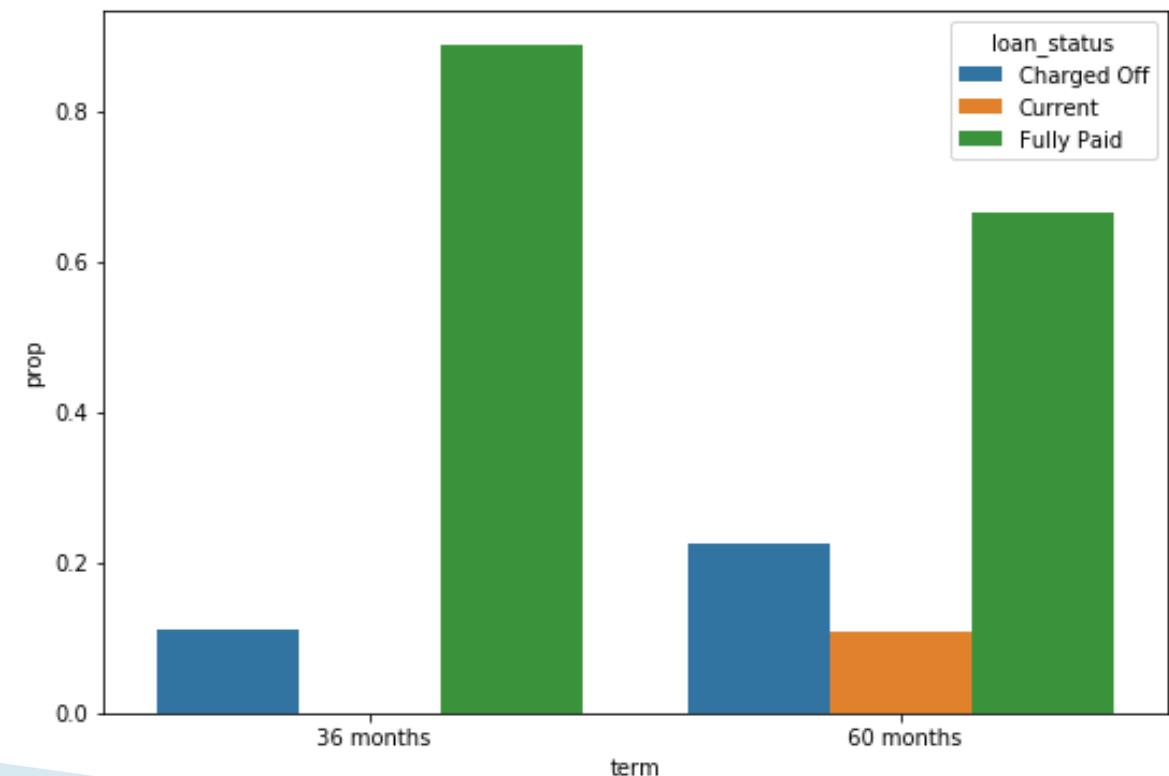
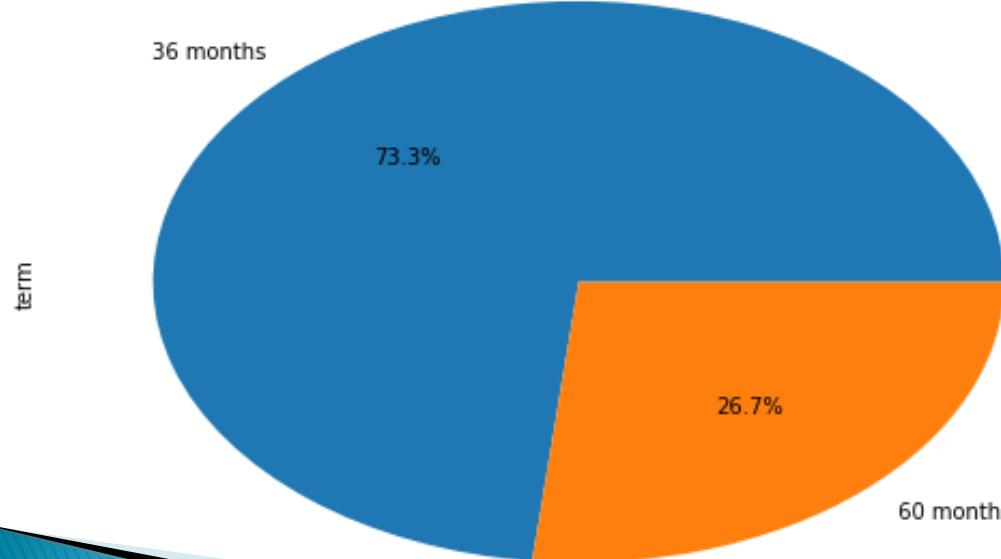


Segmented Univariate Analysis – Loan Status vs Term

We can see the number of approved short term(3 yr) loan is higher than long term(5 yr) loan.

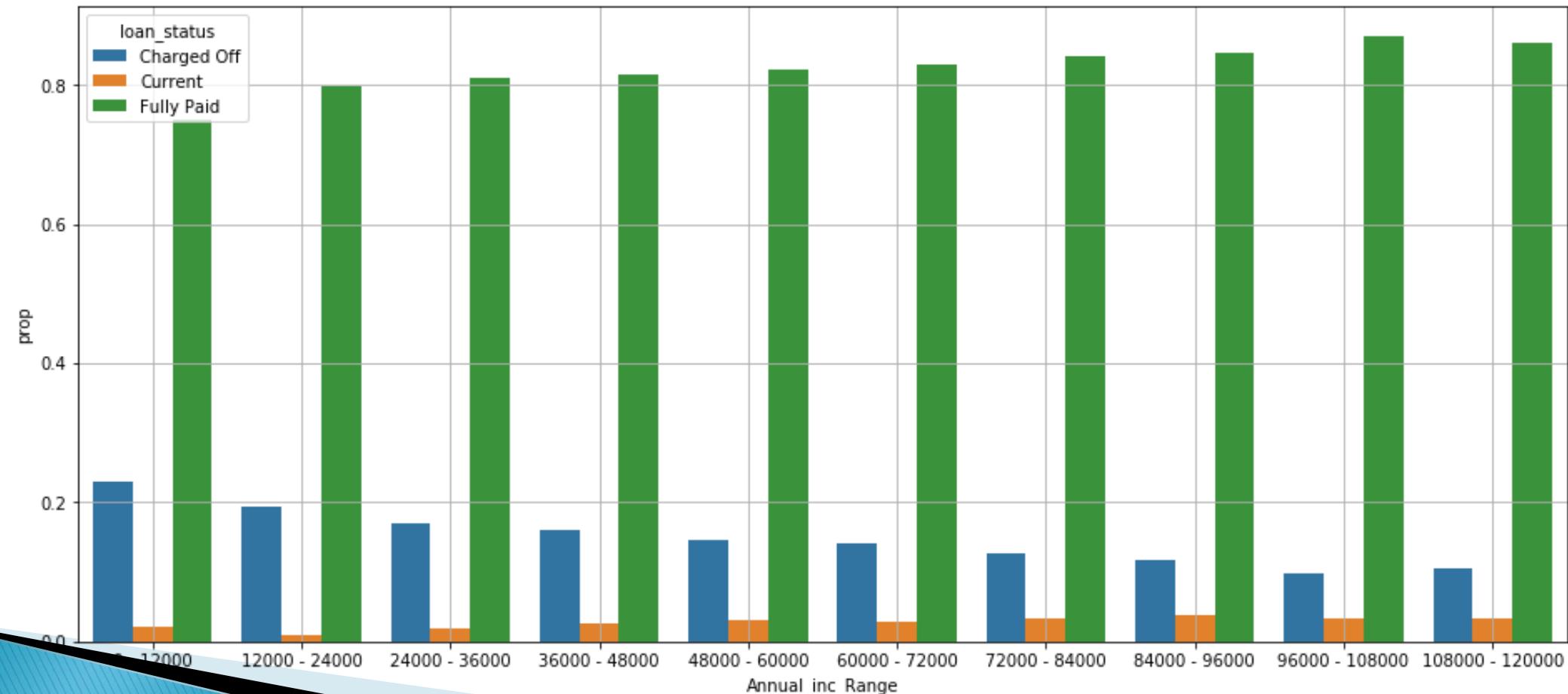
But for 36 months charged off rate is 11% where as for 60 months it is 22% which is almost double.

Thus 36 month time frame is more preferable .



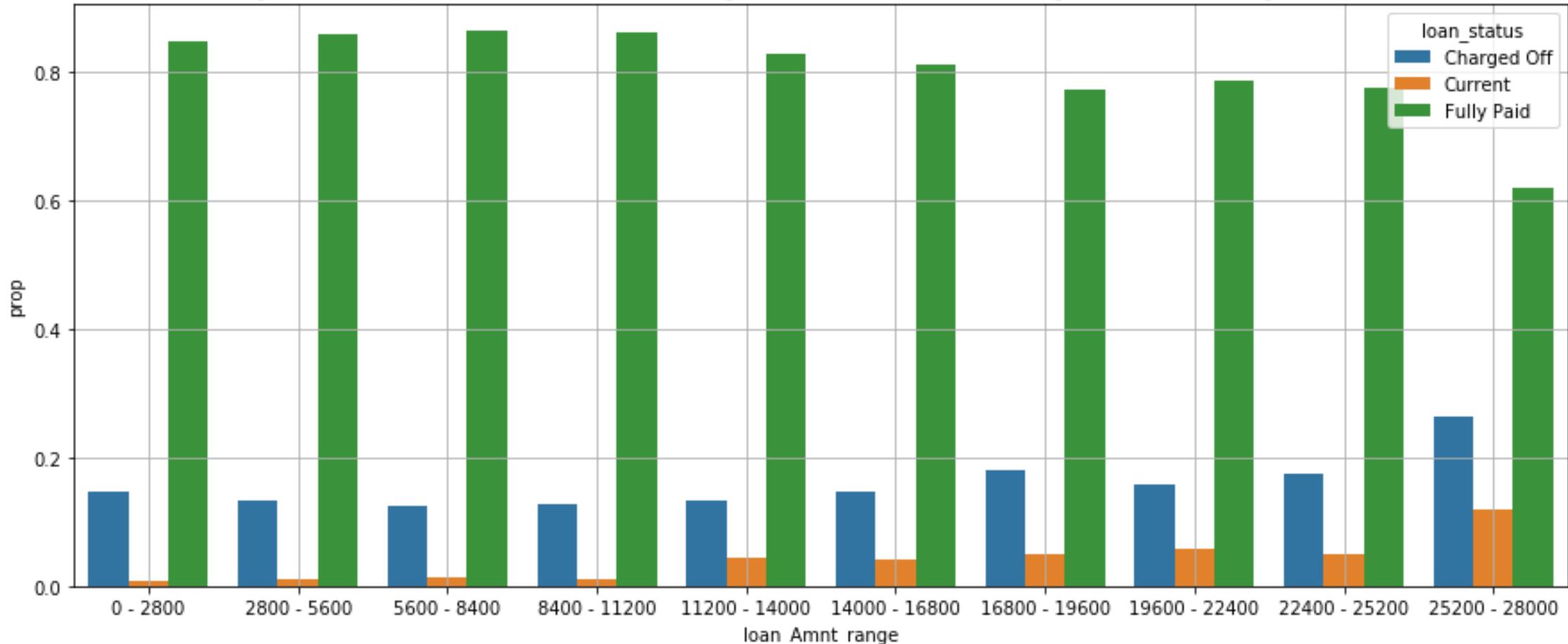
Segmented Univariate Analysis – Annual Income Range

We can infer % of Charged Off Applicants are decreasing as Income Range is Increasing .



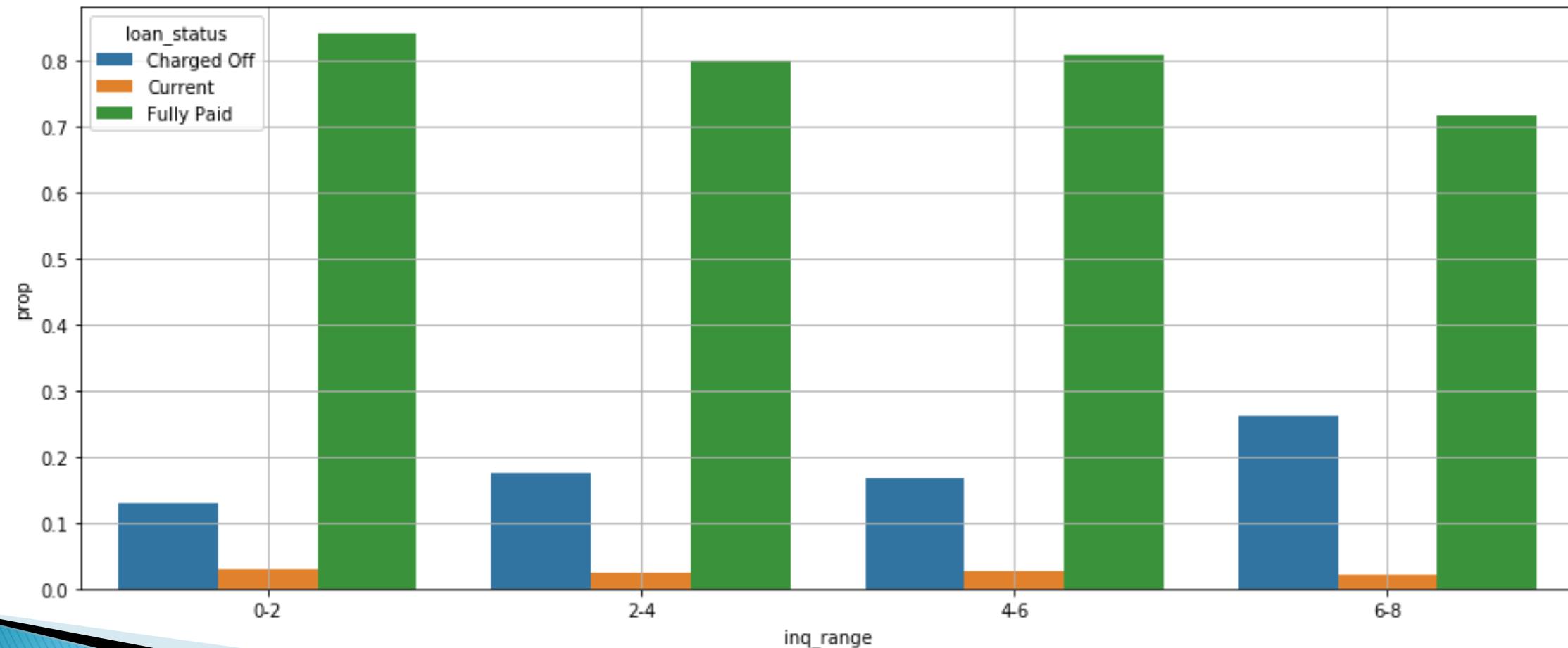
Segmented Univariate Analysis – Loan Amount Range

We can infer % of Charged Off Applicants are increasing as Loan Amount Range is Increasing.



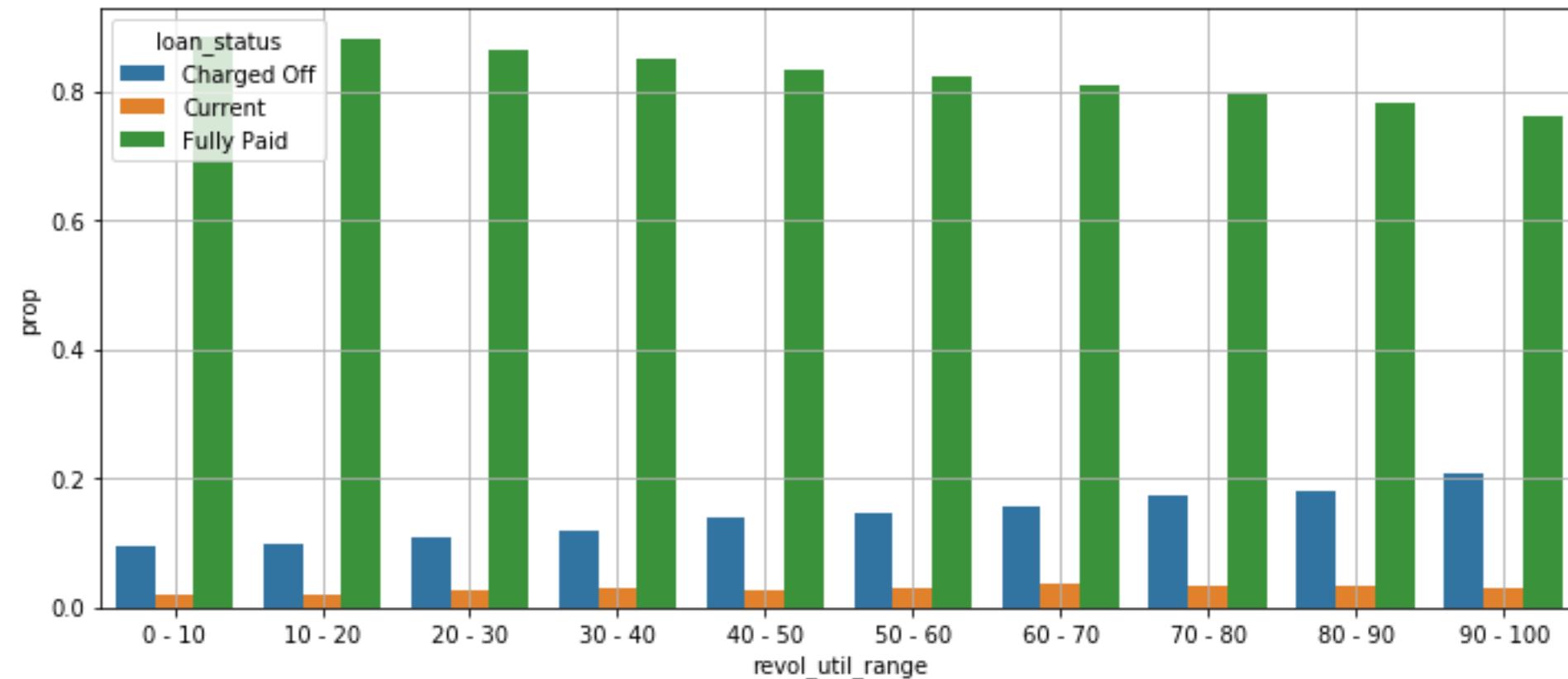
Segmented Univariate Analysis – Inquiry Range

We can infer % of Charged Off Applicants are increasing as Inquiry Range is Increasing.



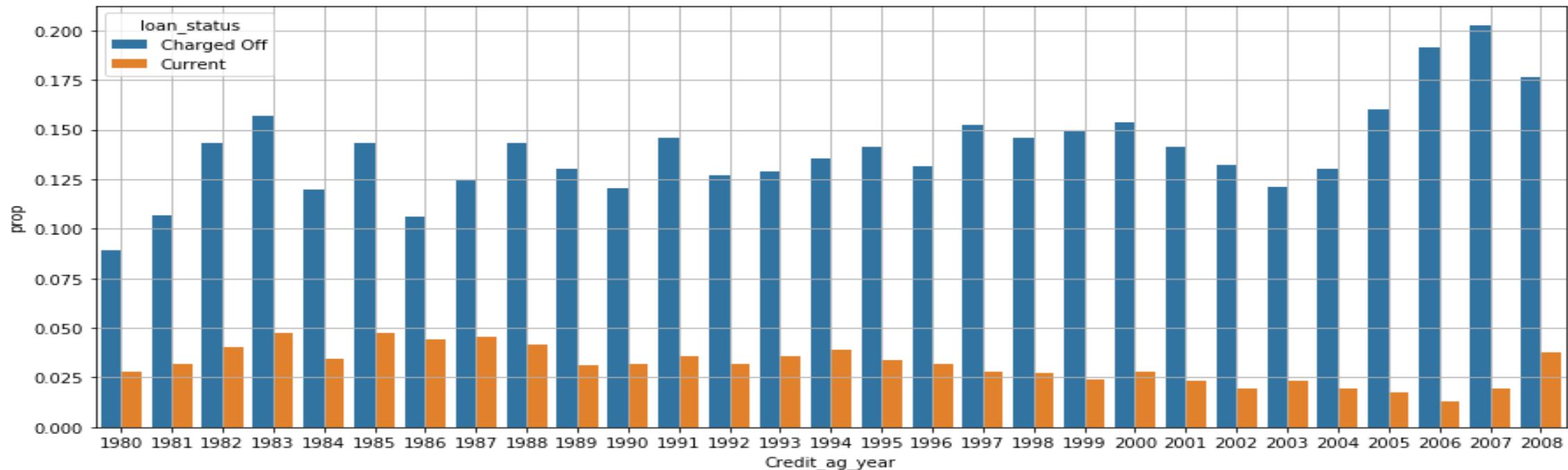
Segmented Univariate Analysis – Revolving Utilization

If we have higher % of Applicant with higher revolving Utilization then Charged off % is also high which is a risk for Organization.



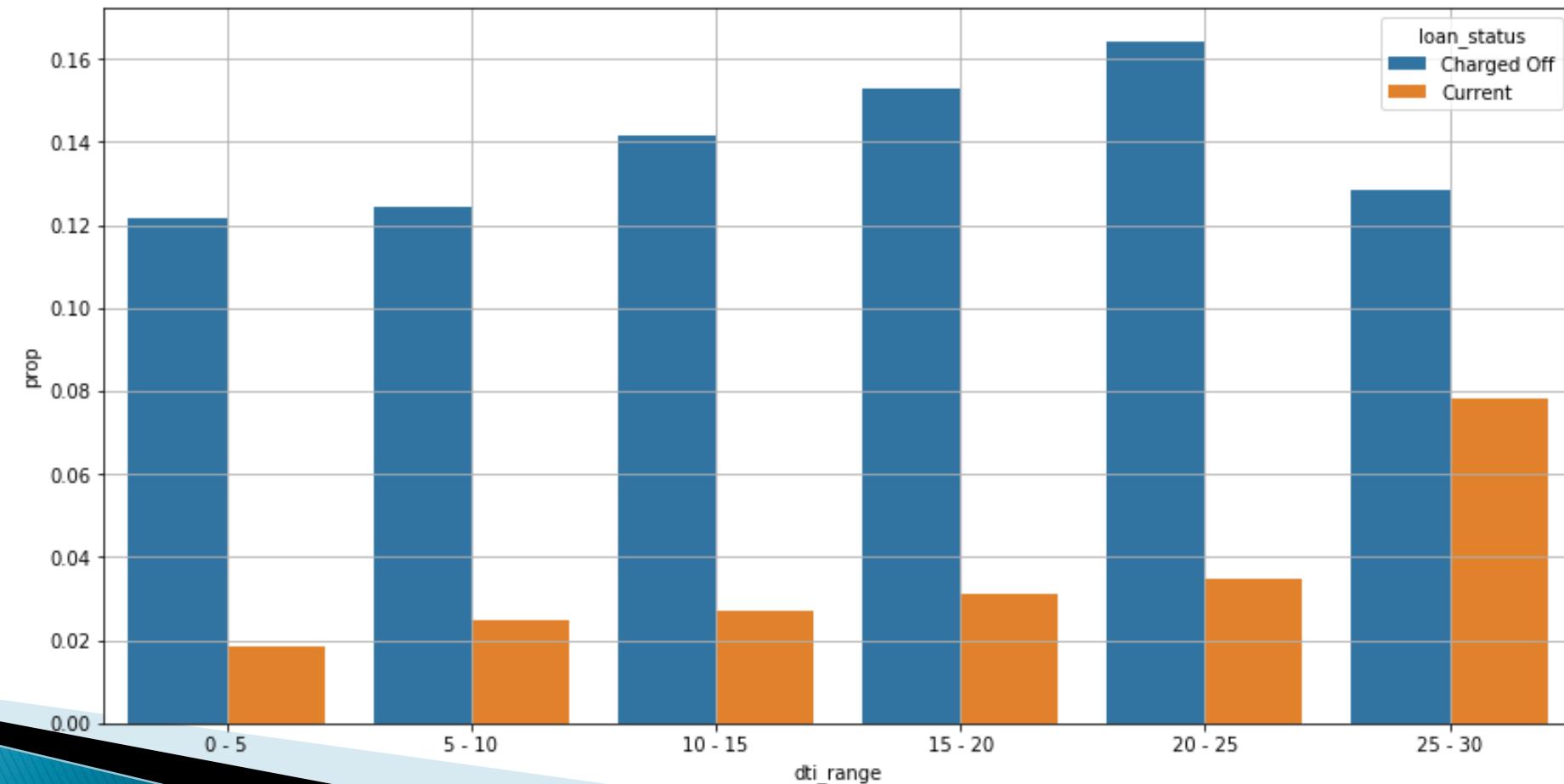
It seems in post Recession period of US economy , Charged Off % is increasing . For example after 1981– 1982 , we can see % Charged off is increased .

Similarly , In Early 2001 to 2003 during contraction in the North American economy which lead to Higher % Charged Off .



Bi-variate Analysis – Loan Status vs Debt to Income Ratio

Based on Bivariate Analysis on Debt To Income(DTI) and Loan Status, we observed a steady increase in the charged off loans ratio as DTI increases compare to fully paid loans.



Conclusion

Borrower's with following factors are highly prone to be defaulters

- Loan Purpose
 - Small Business
 - Renewable Energy
 - Educational
- State
 - Nebraska
 - Nevada
- Home Ownership
 - Mortgage
 - Rent
- Loan Term as 5 year (60 months)
- High Public Bankruptcies
- Higher Grade / High Interest Rate
- Debt to Income Ratio
- Income Range
- Higher Revolving Utilization Factor
- Credit Age
- Inquiry Range