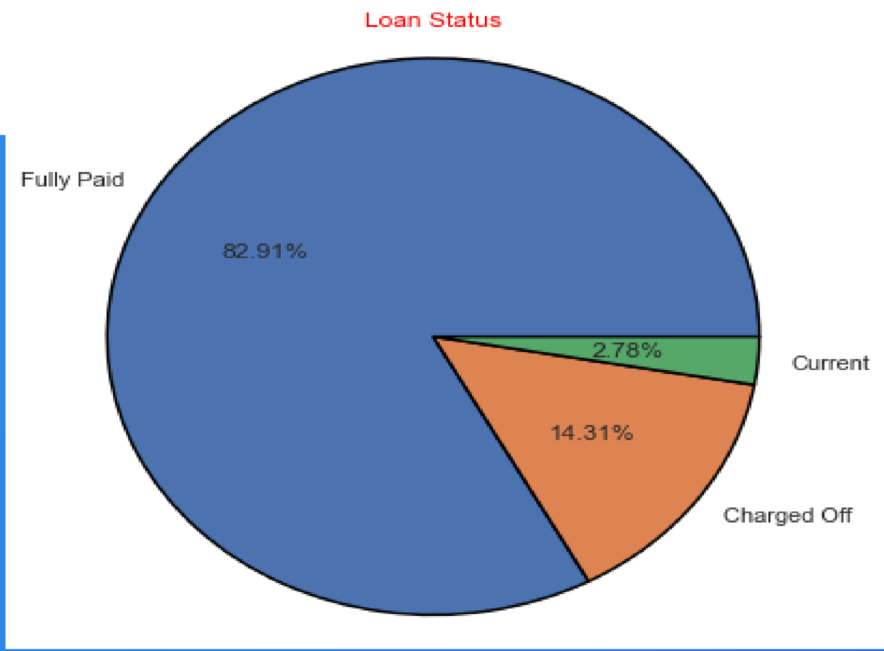
LENDING CASE STUDY

NISHITH B RAO

BALABHASKAR

/01 Problem Statement

- The Business Case was to identify potential indicators for Loan Defaulters from the Lending Club Case Study Data.
- Data Set Information
  - RangeIndex: 39717 entries
  - Columns: 111 entries
- Our target column was “loan_status”



/02 DATA CLEANING

01 Fixing data series and filling Missing values

- Dropping columns that have only “NA” values
- Dropping columns which have high percentage of missing values
- Dropping column that has description

02 Standardizing values

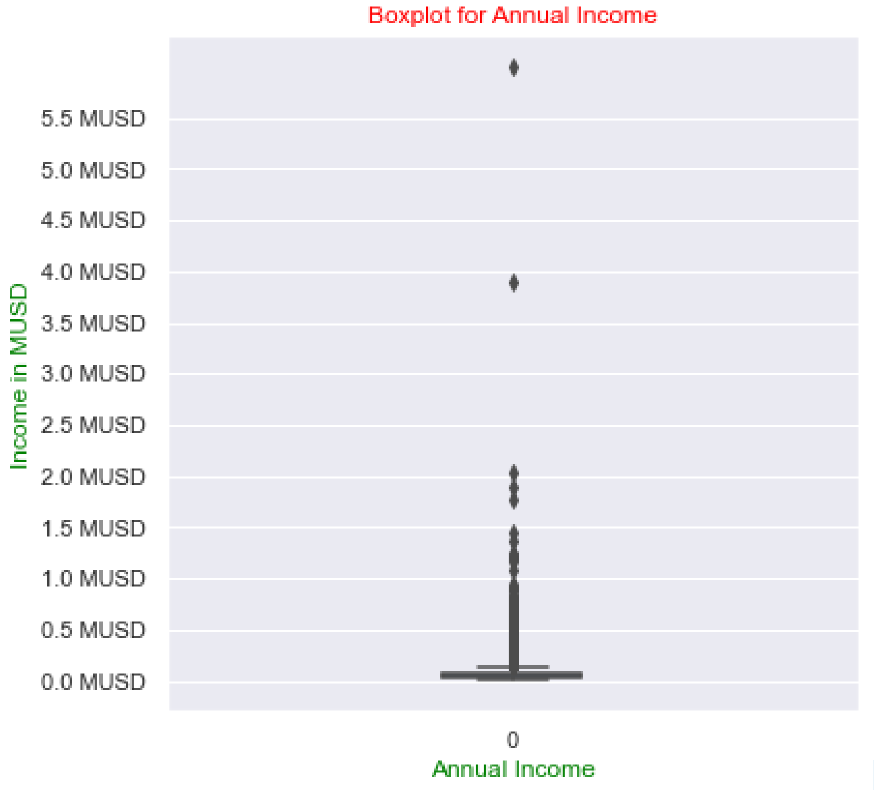
- Column emp_lenght missing values were filled with NA not to exaggerate the data
- Column mnt_since_last_del was filled with median value
- Column revol_util was filled with median value

03 Outlies Treatment

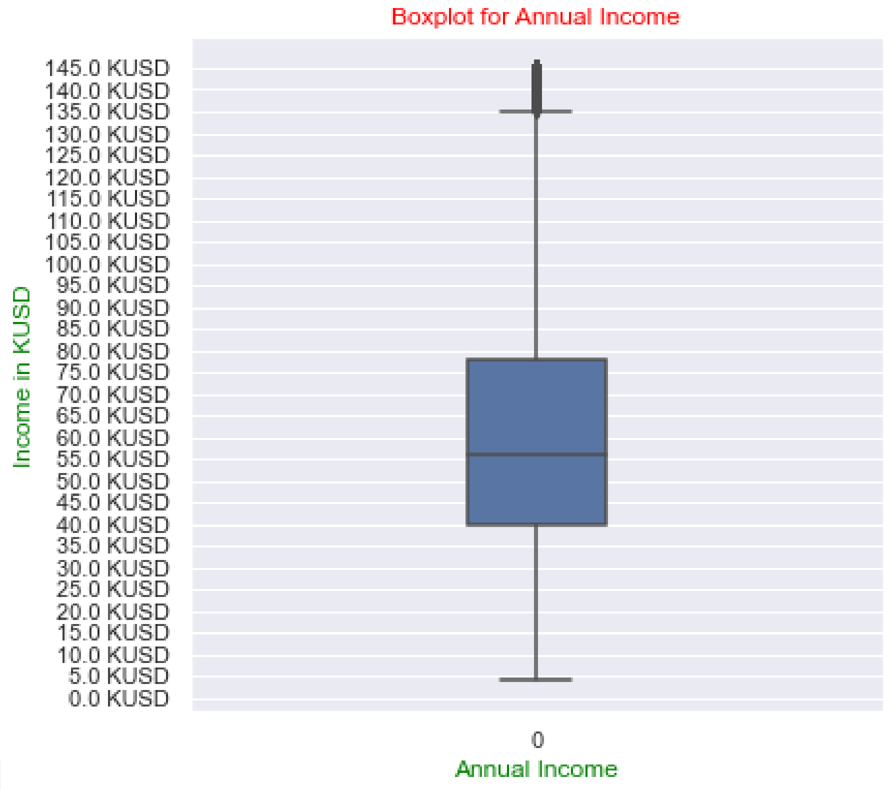
Outlier Treatment – Using 1.5 * IQR Method

- Outlier treatment on Annual_inc, Open_account,total_pymnt
- Removing extra charecters from columns revol_uti,int_rate that had %
- Emp_lenght had values like <1year and 10+year which was formated to 0 and 10 years

Before



After



/03 EXPLORATORY DATA ANALYSIS

01 Categorical Univariate Analysis

- We found significant relationship for PURPOSE, GRADE, HOMEOWNERSHIP, LOAN ISSUE MONTH, LOCATION
- We were also able to conclude that TERM, SUBGRADE, VERIFICATION STATUS had minimal/null effect on loans that are charged off

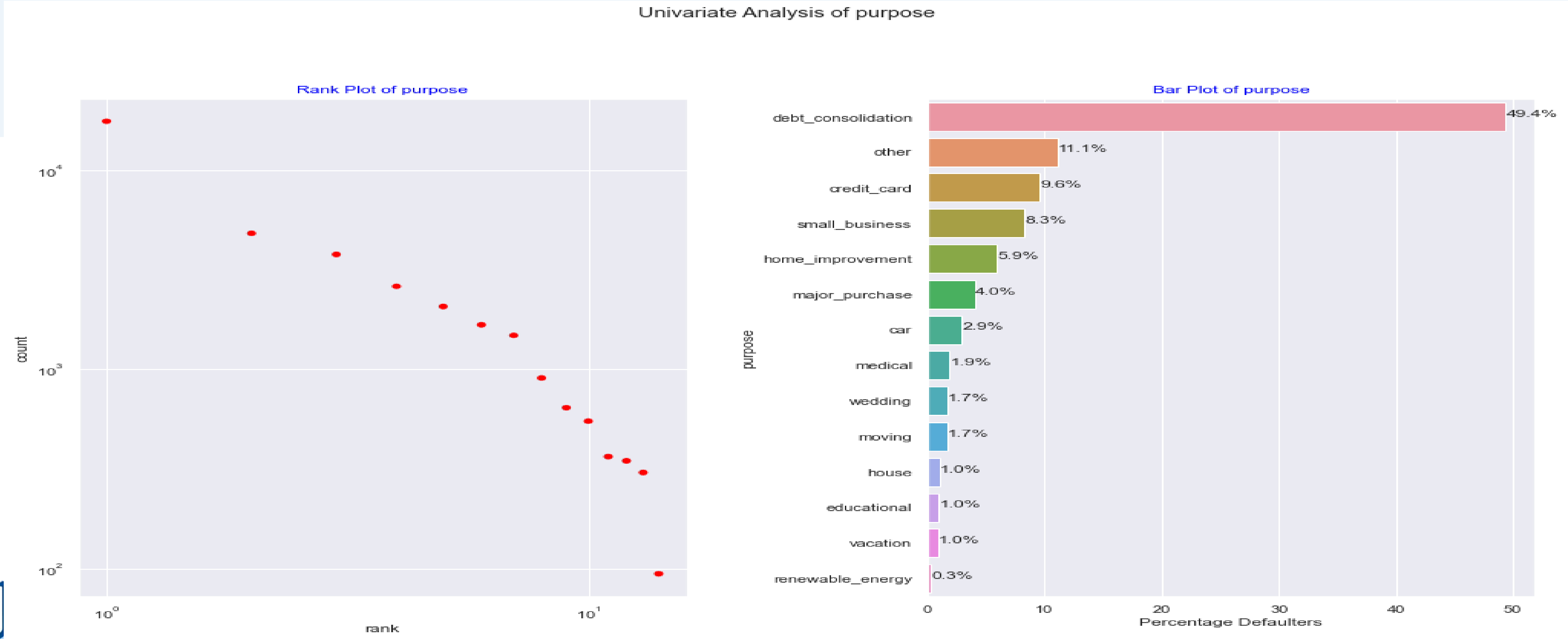
02 Continuous Univariate Analysis

- We were able to plot violin plots for Interest Rate, Employment Length and DTI

03 Bivariate Analysis

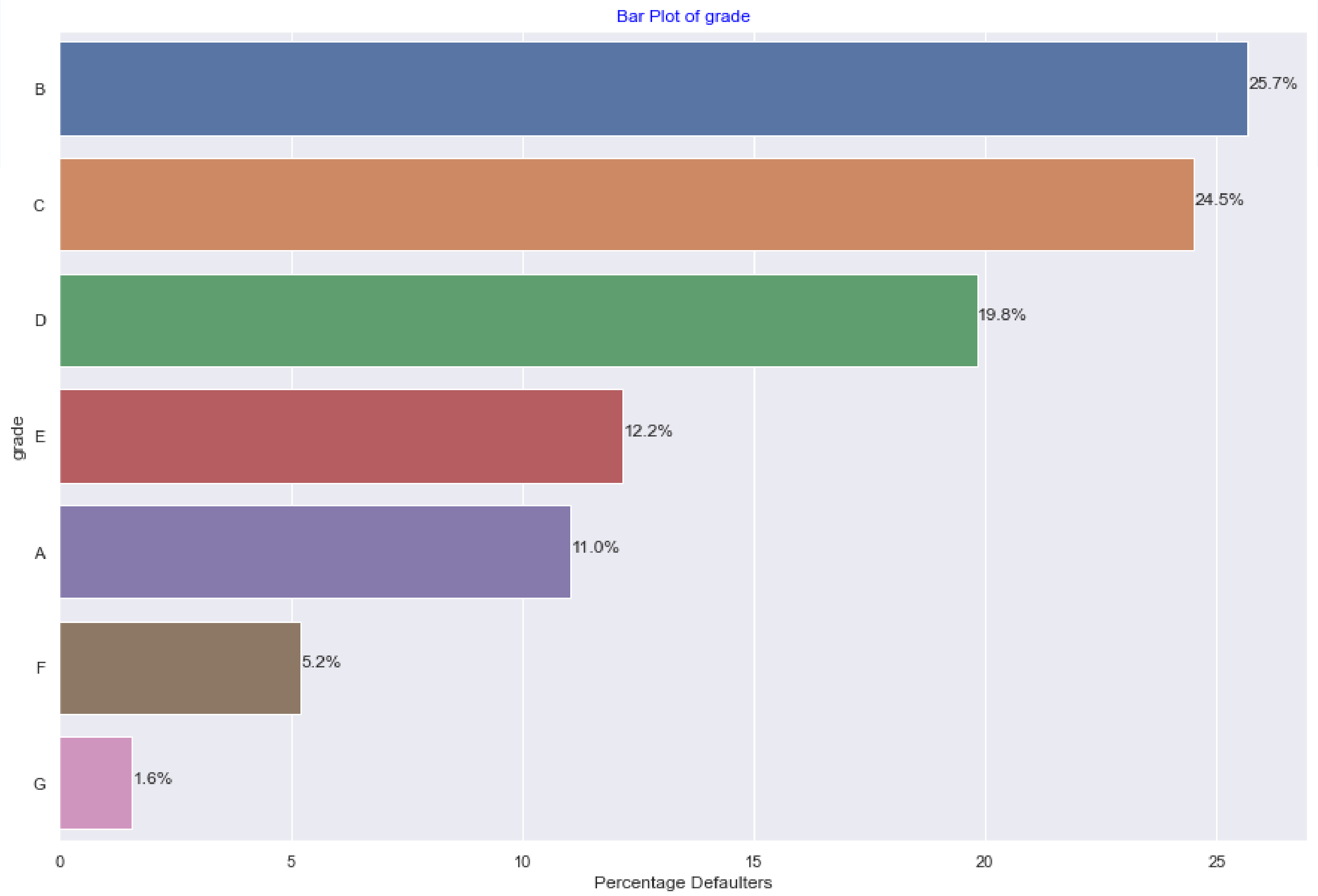
- We were able to analyse the influence of Interest Rate on different types of Loan grades.
- We were also able to segment annual income group and analyse it against different purpose

- The log - log rank v/s frequency plot of Purpose follows a Zipf Distribution. Additionaly it could possibly serve as an independent variable



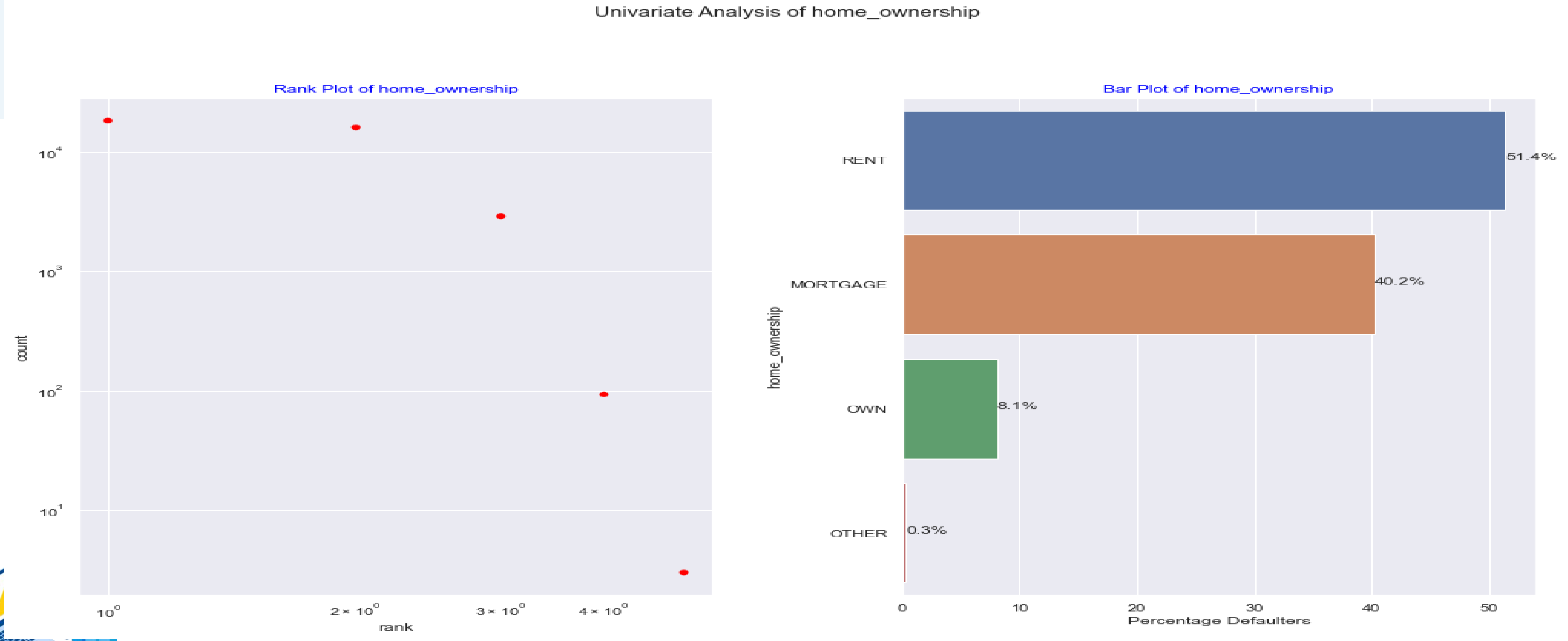
PURPOSE

- We can conclude that loan types that are categorized under B, C, D are most likely to default compared to rest of others. Grade G seems to default the leas



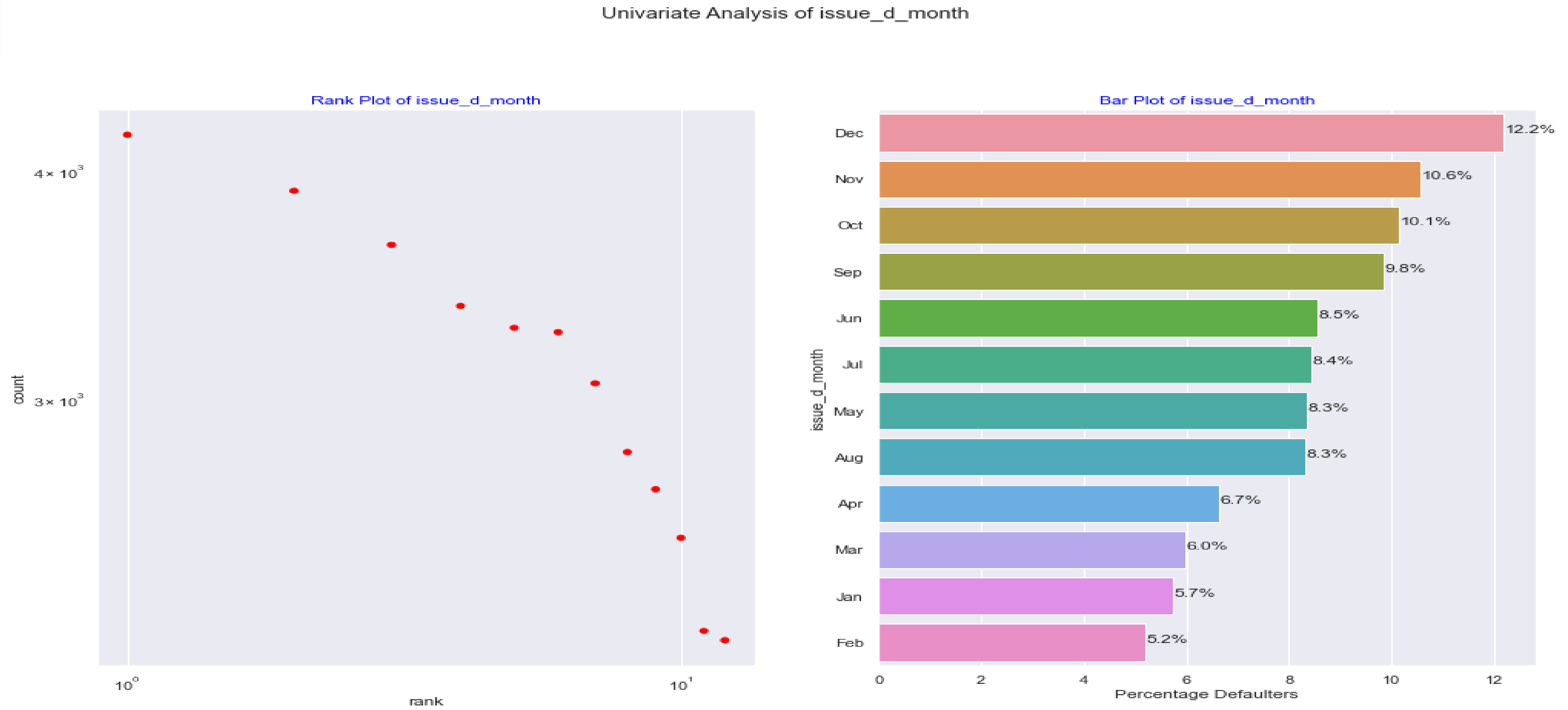
GRADE

- People who lives on Rent or has their house Mortaged will default more frequently (~ approx 5 times) more than people who own a house.

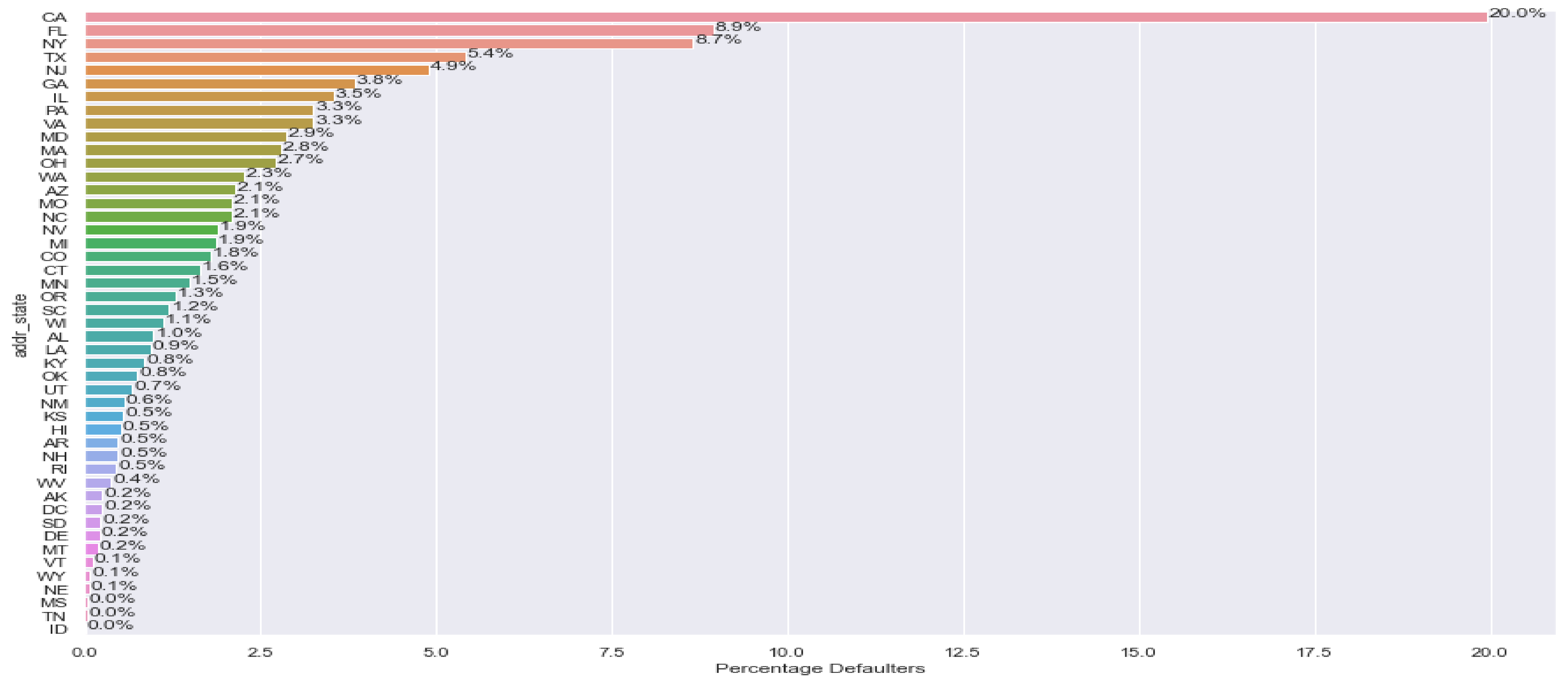


HOME OWNERSHIP

- Loans issued at the end of the year tend to default more than loans issued at the beginning of the year. The rank plot follows Zipf Distribution which implies majority of the loans that tend to go default are issued during year ends.
  - 1. A possible explanation could be that year end time is when you track your expenses or it could probably be the time when we have to settle up our credit card expenses and plan for holiday expenses. Inorder to do this we mostly would consolidate our debts.
  - 2. Another possible explantion for this could be that during end of year people could apply for loans for vacations, holiday expenses. This is more likely to default compared to a loan that is taken during year start which could be most probably for non holiday related activity

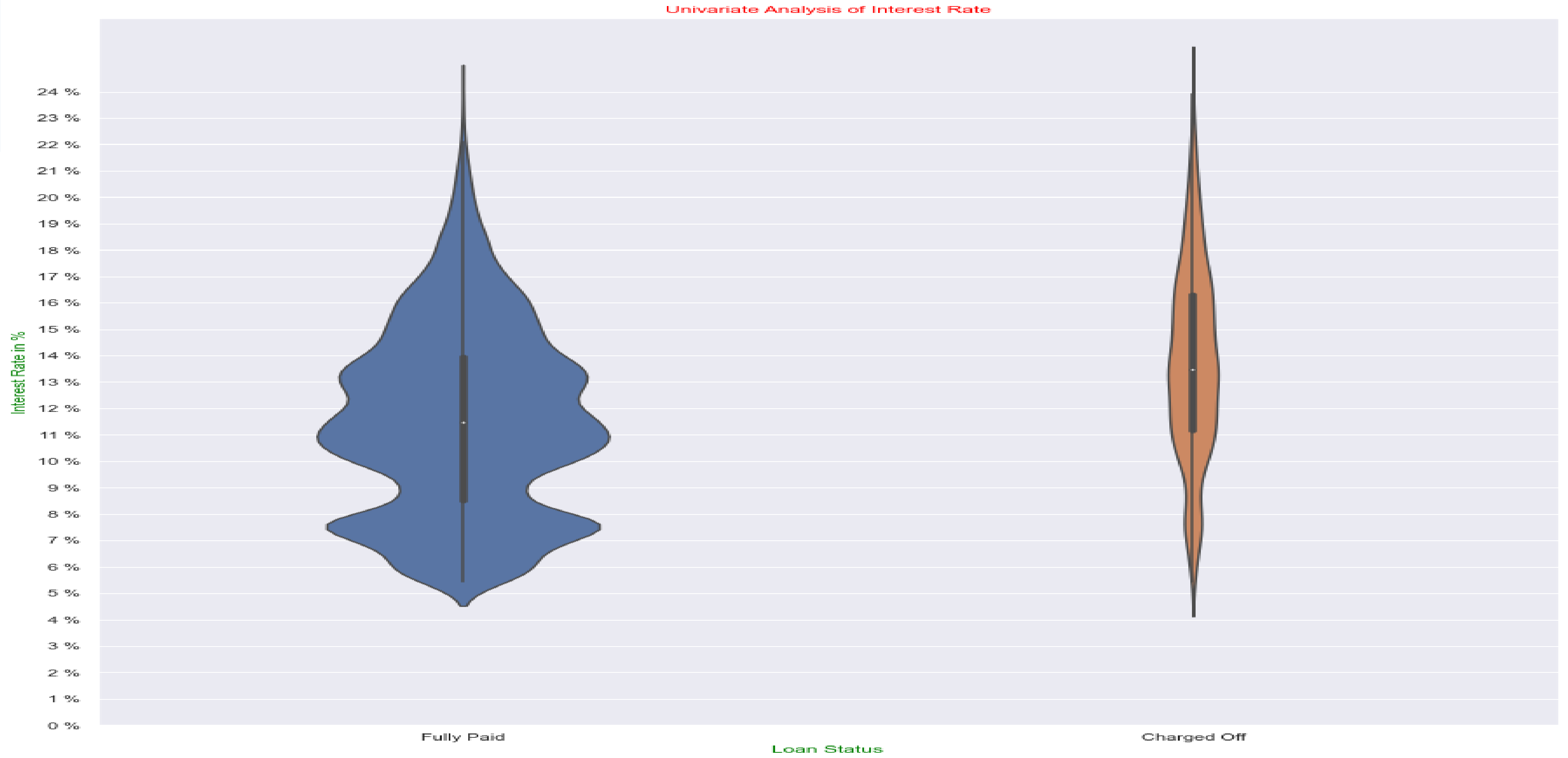


- Large number of defaulters hail from the state of California, followed by Florida and New York



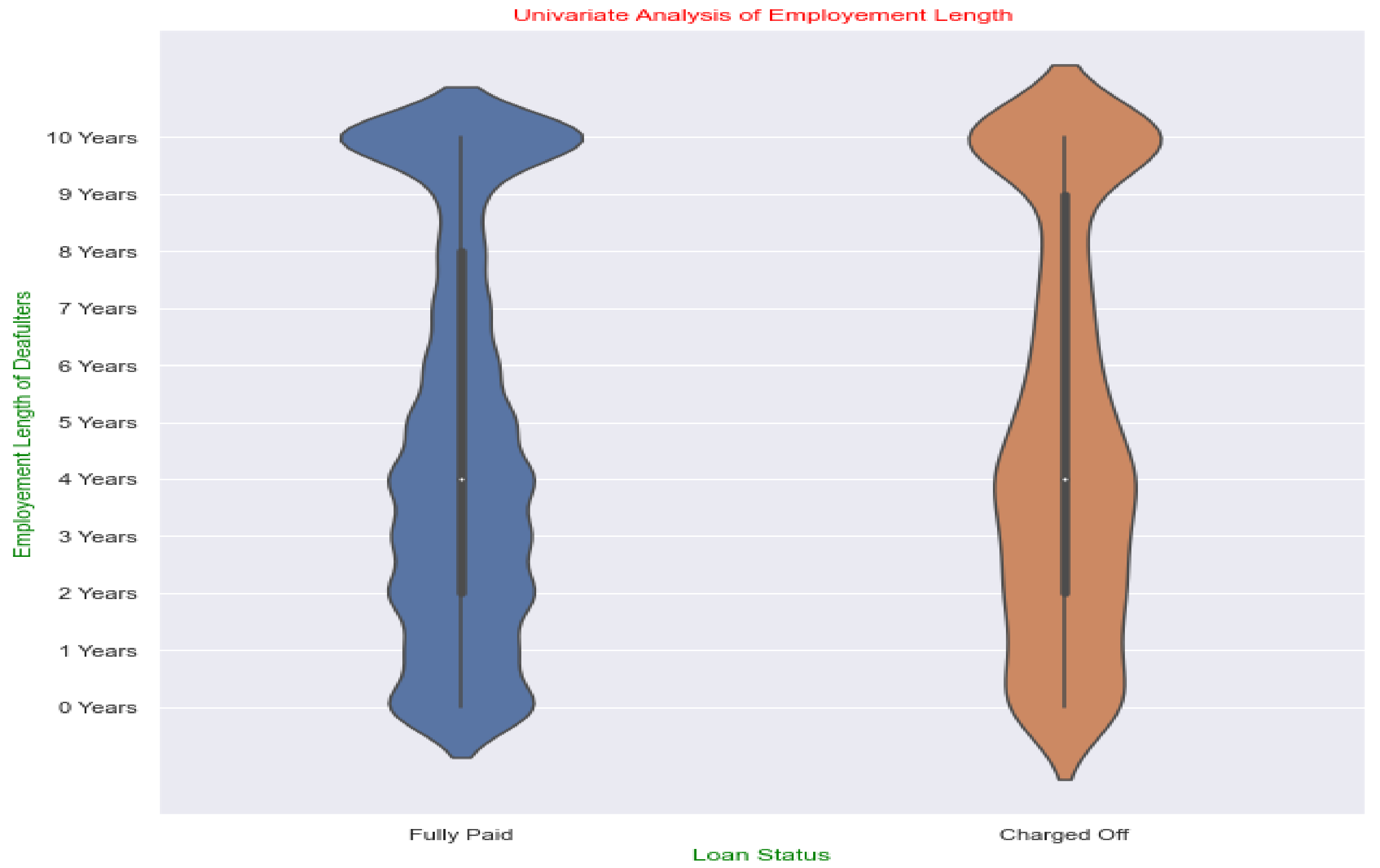
ADDR_STATE

- The median interest rate of Charged Off Loan is 13.5% whereas for loans that are fully paid are 11.5%. Since the First Quartile of default loans are approximately at 11% we can conclude that interest rates that go beyond 11% has more chance to default. Additionally the chances to default are highest in the interest rate range of 11-16%.



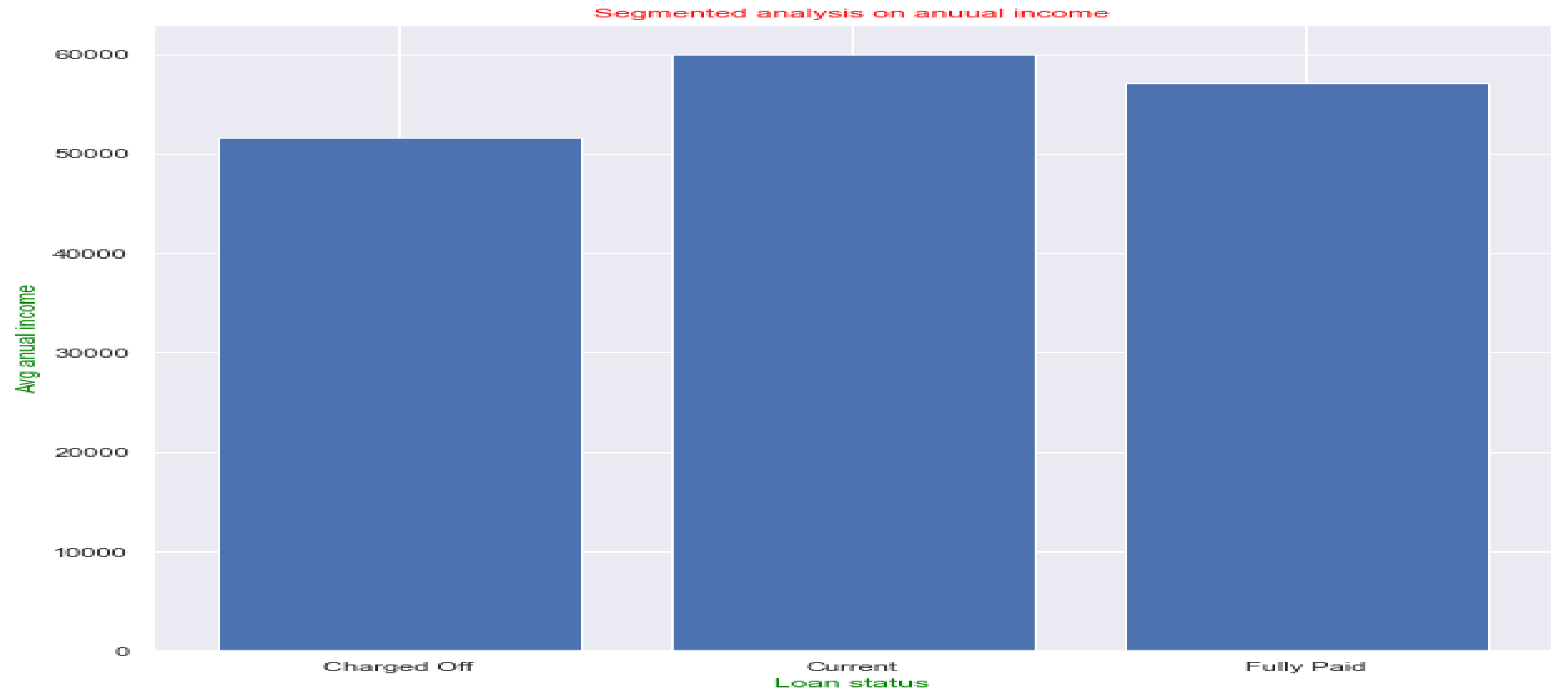
Int_rate

- The key learning from the visualization is that the 50% of defaulters are employed for 4 or less than 4 years. An interesting trend that we could observe here is the tendency to default reduces once the employement period moves through 5 to 9 years the tendency decreases and shoots up at 10+ years.

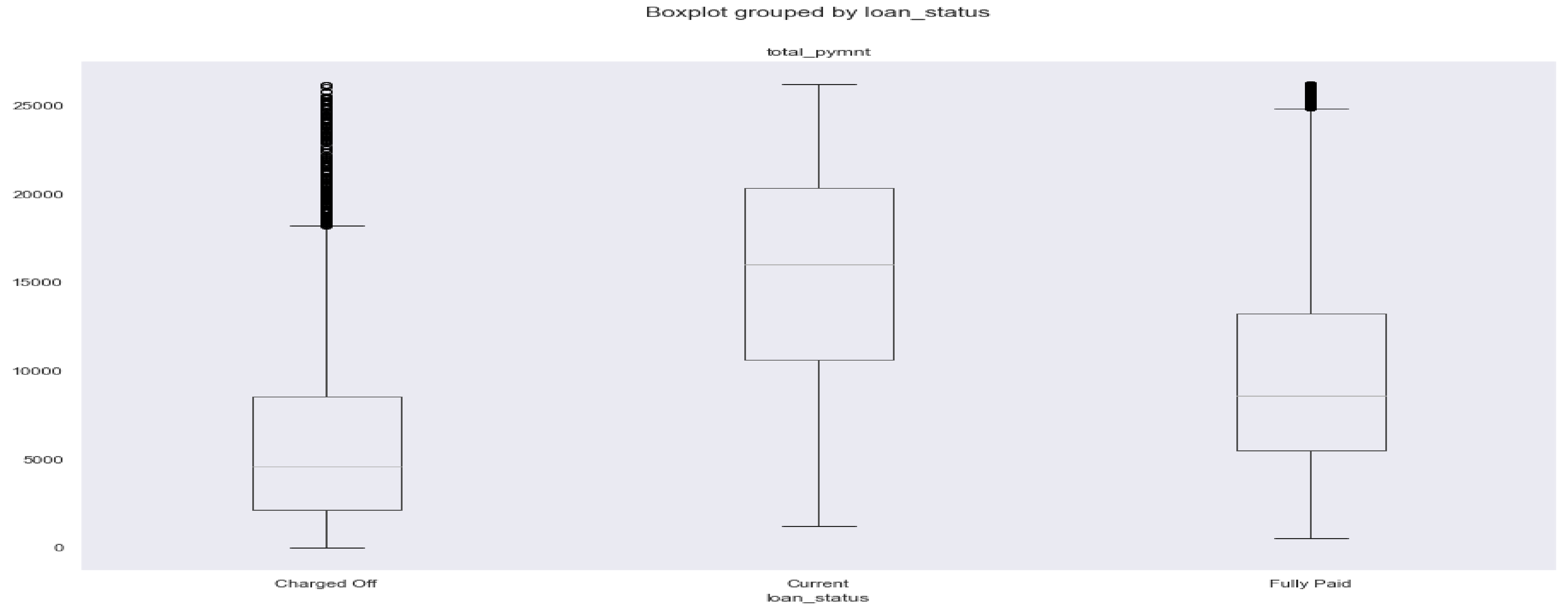


Emp_len_
mod

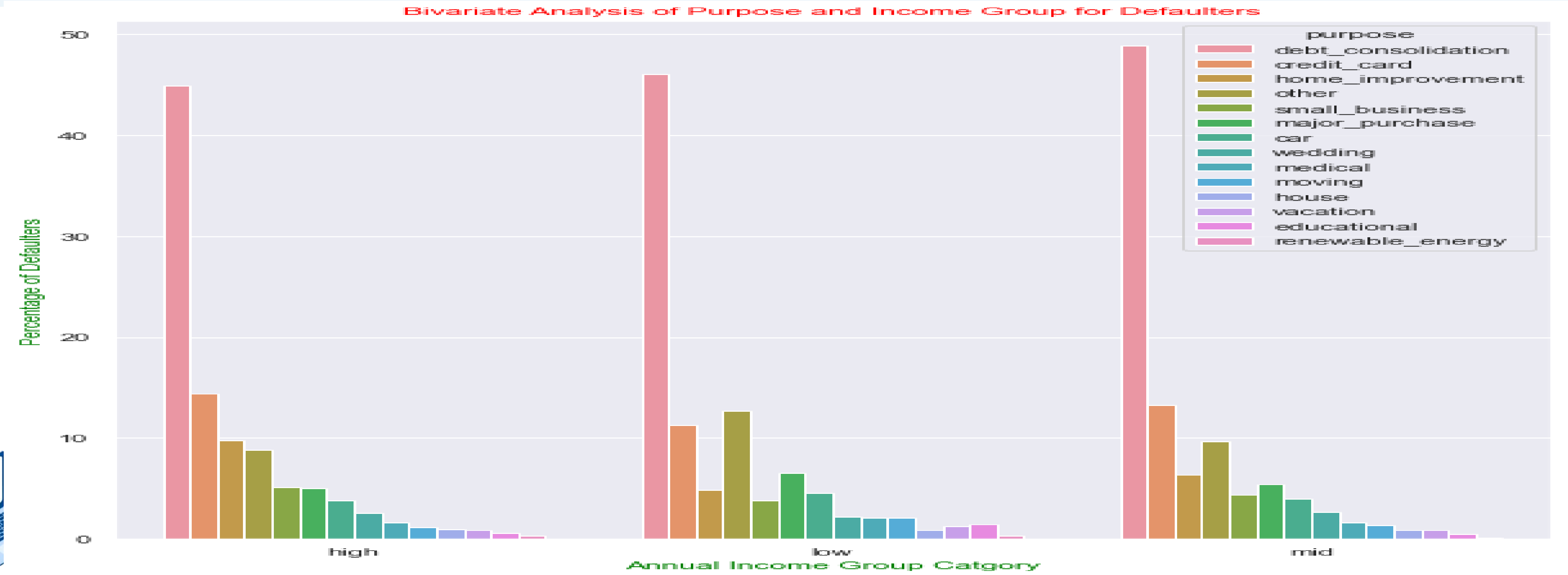
- People who have defaulted have comparitively less average income



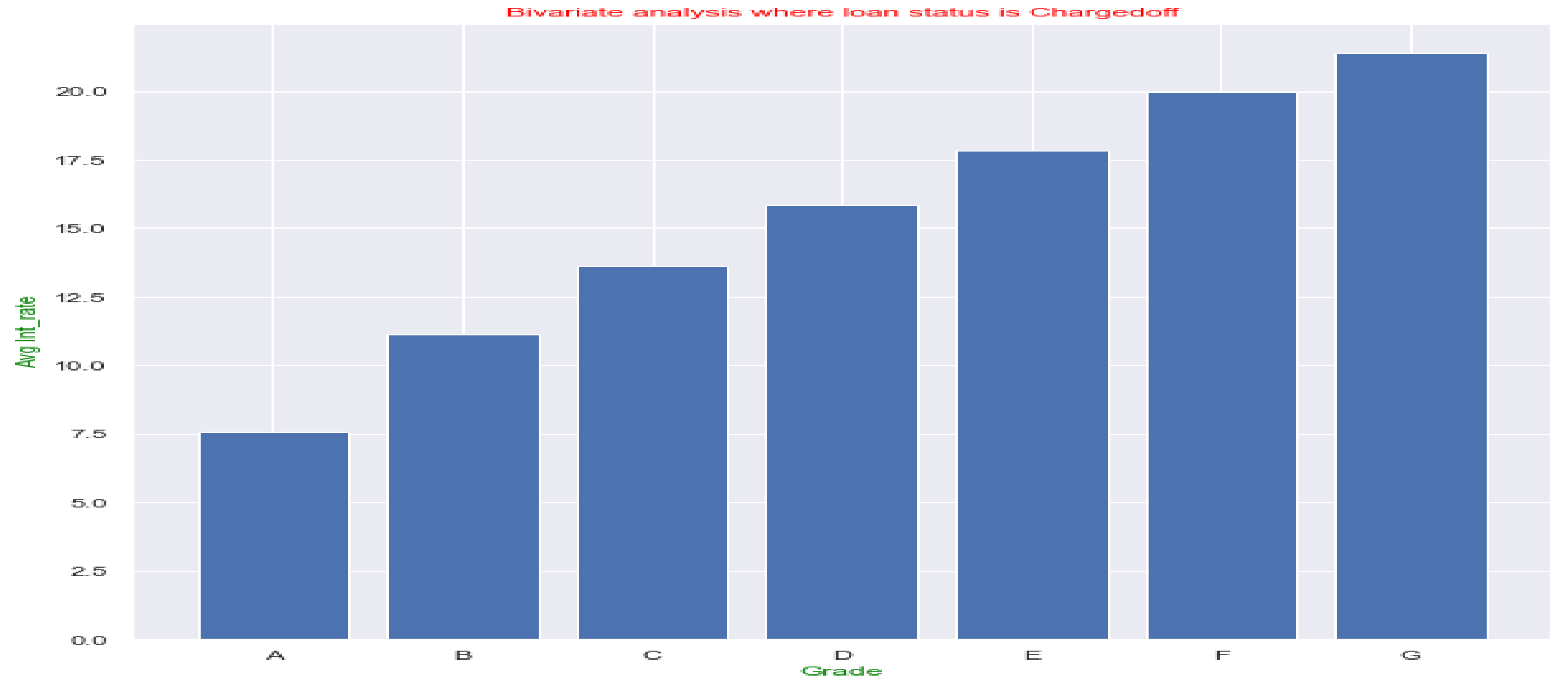
- The total Amount paid by defaulters is less



- Everywhere debt consolidation is highest
- -High income group follow the earlier mentioned patter where debt consolidation is high.
- Low income group has purpose other rising up. They take small and short loans -Medium income has large rise in credit card defaulters



- From this we can dram one conclusion is that even though intrest rate is high for G grate there are less defaulters compared to grade B where average intrest rate is low



Summary

1. From our analysis we were able to identify that the loan that is most likely to default is for Debt Consolidation.
2. Loan which has interest rate above 11% is also a good flag for defaulters.
3. Loan Grades B> C > D are more likely to default owing to their lower interest rates. Grade G Loans are the best with minimal default but however has highest interest rate.
4. Loans applied during year end has more chance to default compared to loans that are applied at the beginning of the year.
5. Employment Period is critical. 50% of the defaulters are employed for less than 4 years. Additionally requestors from California state to default a lot more than rest of the states.