LENDING CLUB

CASE STUDY

1. AMIRTHA RAJAN PKS
2. SUDARSAAN G

LOAN DEFAULTERS ANALYSIS:

OBJECTIVE:

A Consumer finance company , wants to identify patterns which can indicate if a person is likely to default, and that can be used by the company for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate.

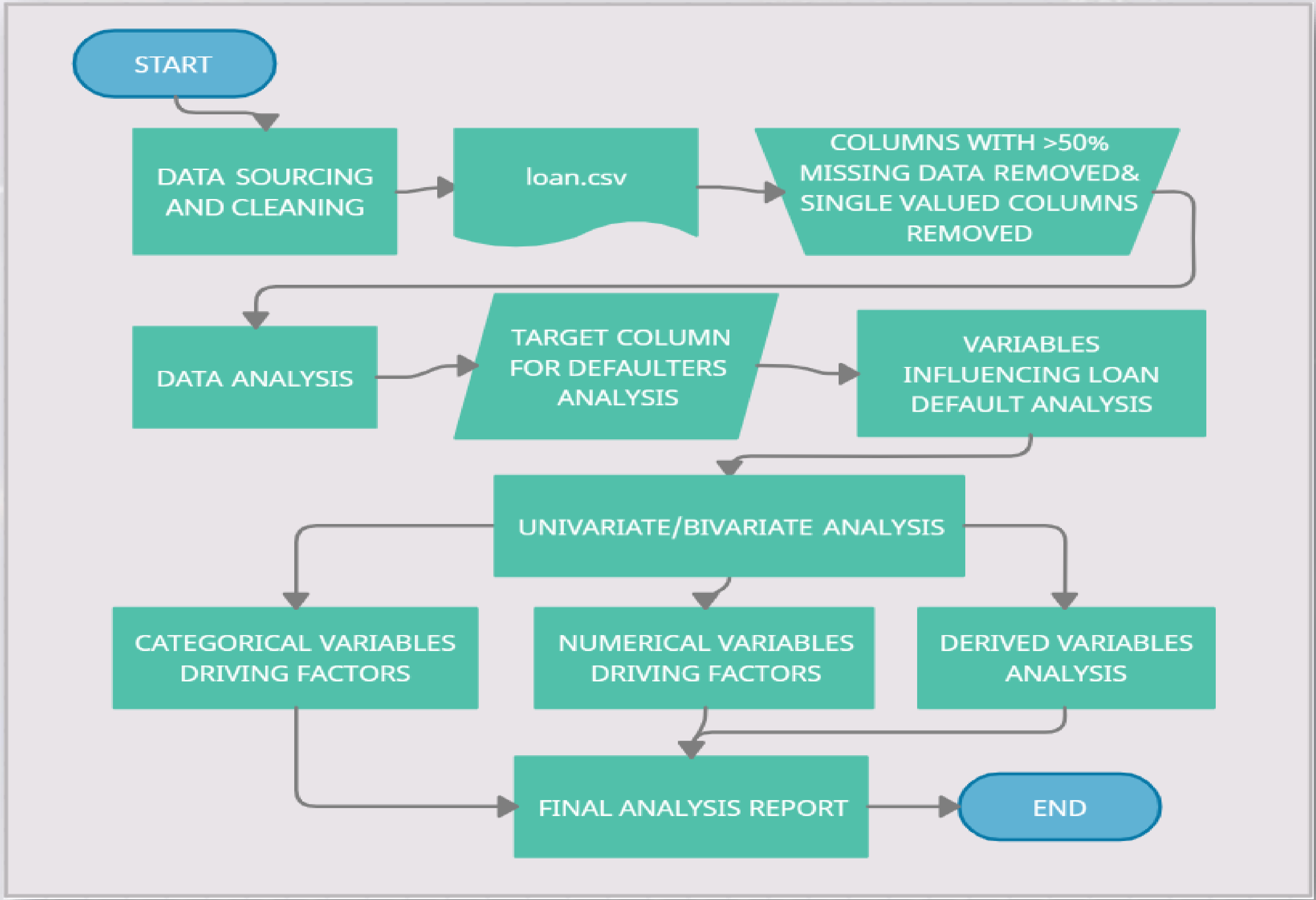
GOALS:

- Identify the suitable Data column that distinguishes the defaulters from Fully paid borrowers.
- Identify Demographic and Loan Characteristics variables/columns that helps predict the default.
  - Identify the driver variables behind loan defaults,
    - Categorical driving factors
    - Numerical driving factors
    - Derived data factors

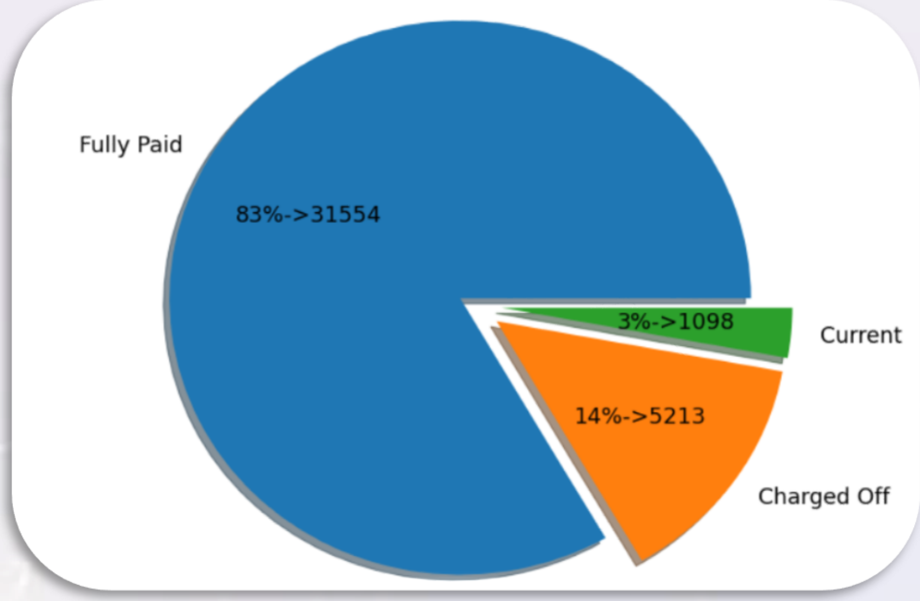
STRATEGY:

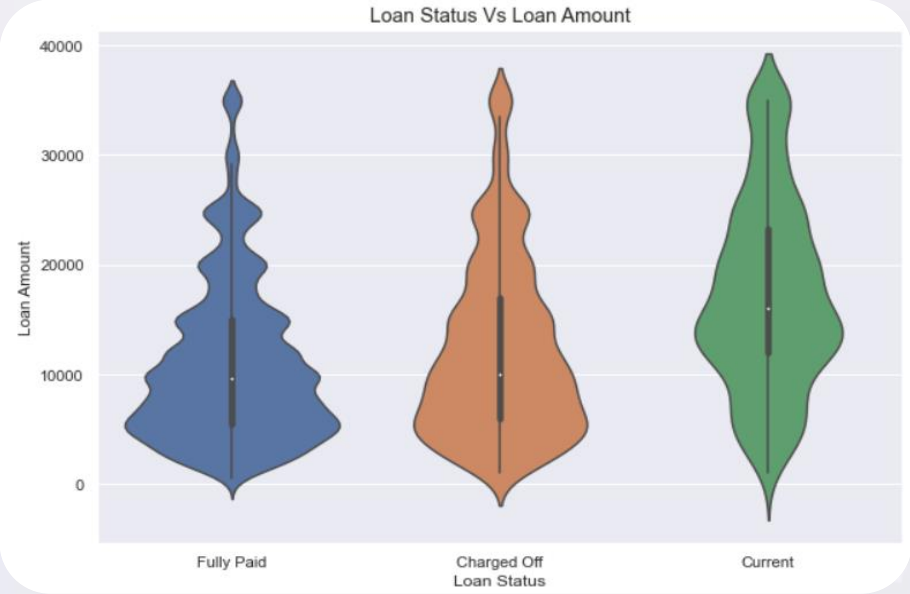
- Higher the Charged Off percent compared to Fully paid in each category of Column, will indicate how likely a person could go default.

FLOW CHART:


START
DATA SOURCING AND CLEANING
loan.csv
COLUMNS WITH >50% MISSING DATA REMOVED& SINGLE VALUED COLUMNS REMOVED
DATA ANALYSIS
TARGET COLUMN FOR DEFAULTERS ANALYSIS
VARIABLES INFLUENCING LOAN DEFAULT ANALYSIS
UNIVARIATE/BIVARIATE ANALYSIS
CATEGORICAL VARIABLES DRIVING FACTORS
NUMERICAL VARIABLES DRIVING FACTORS
DERIVED VARIABLES ANALYSIS
FINAL ANALYSIS REPORT
END

DEFAULTERS COLUMN ANALYSIS:





<u>OBJECTIVE:</u>

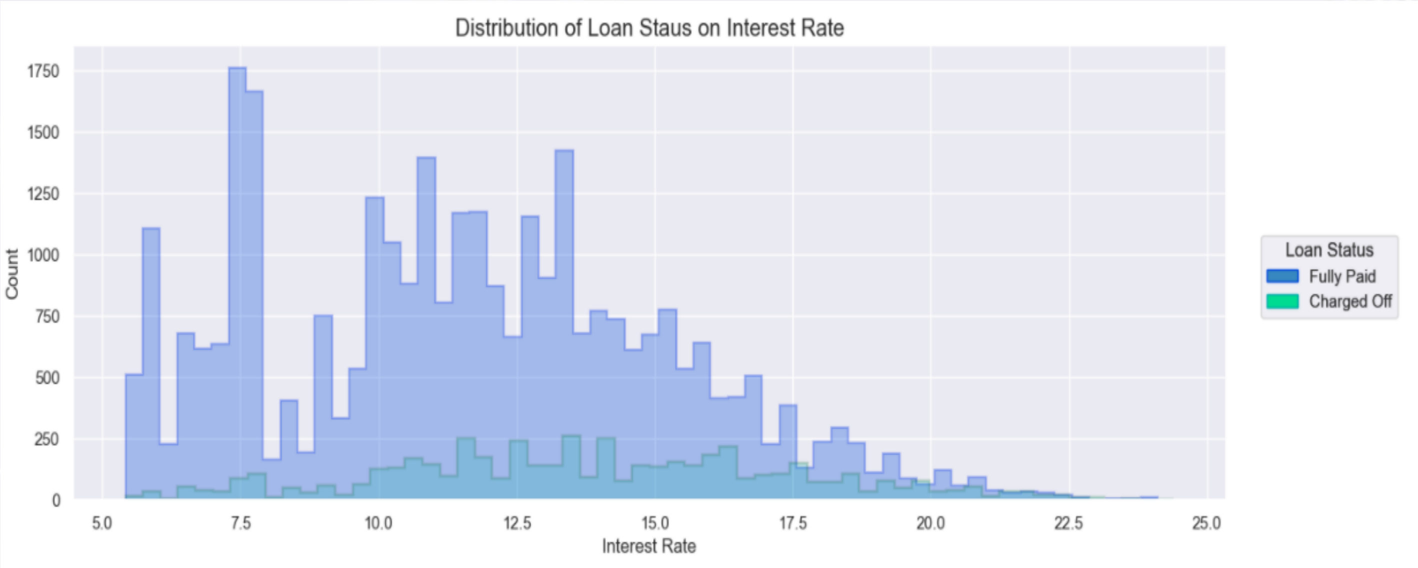
Identify most suitable data column that distinguishes defaulters from fully paid borrowers.

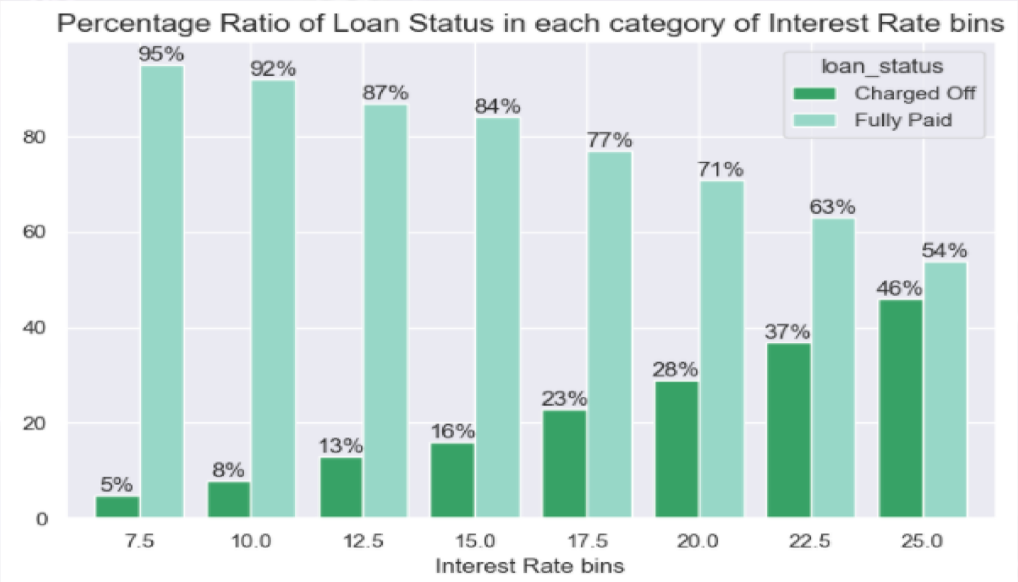
<u>ANALYSIS:</u>

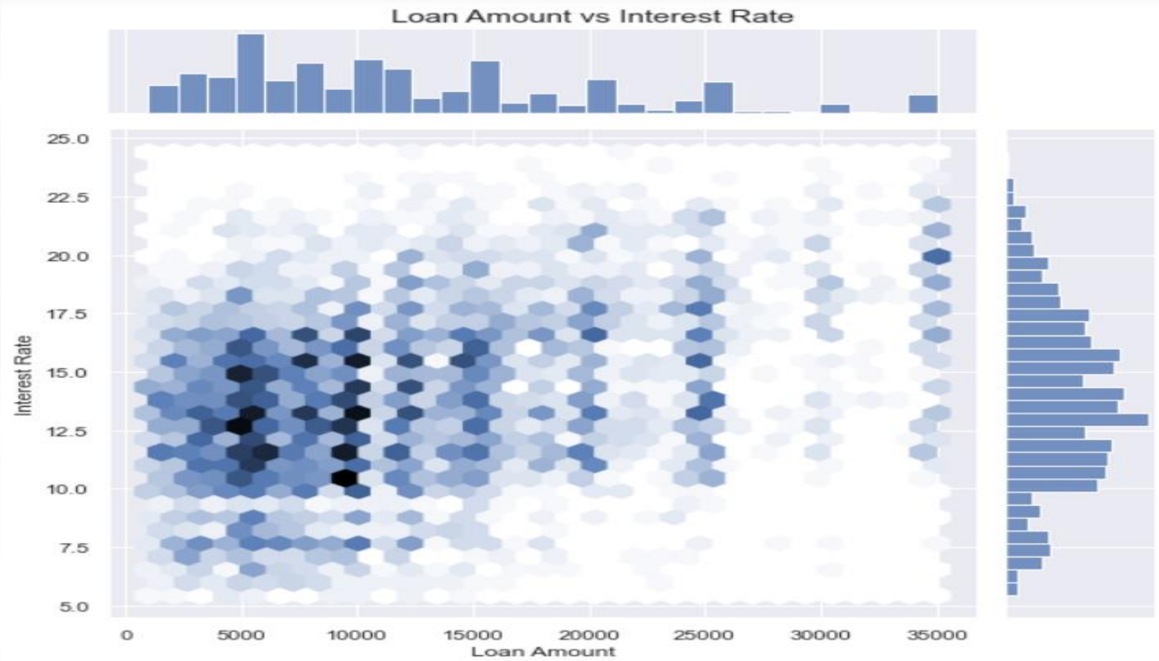
The Column **Loan status** clearly distinguishes Defaulters from Fully paid and Current borrowers and is the most suitable column to be considered to be the **Target Column** for further Analysis.

With ambiguity that **Current Loan data** can either fall under Fully-paid or Charged-off in Future, **Current Loan data** is **not considered for analysis**.

NUMERICAL CONTINUOUS VARIABLE ANALYSIS:



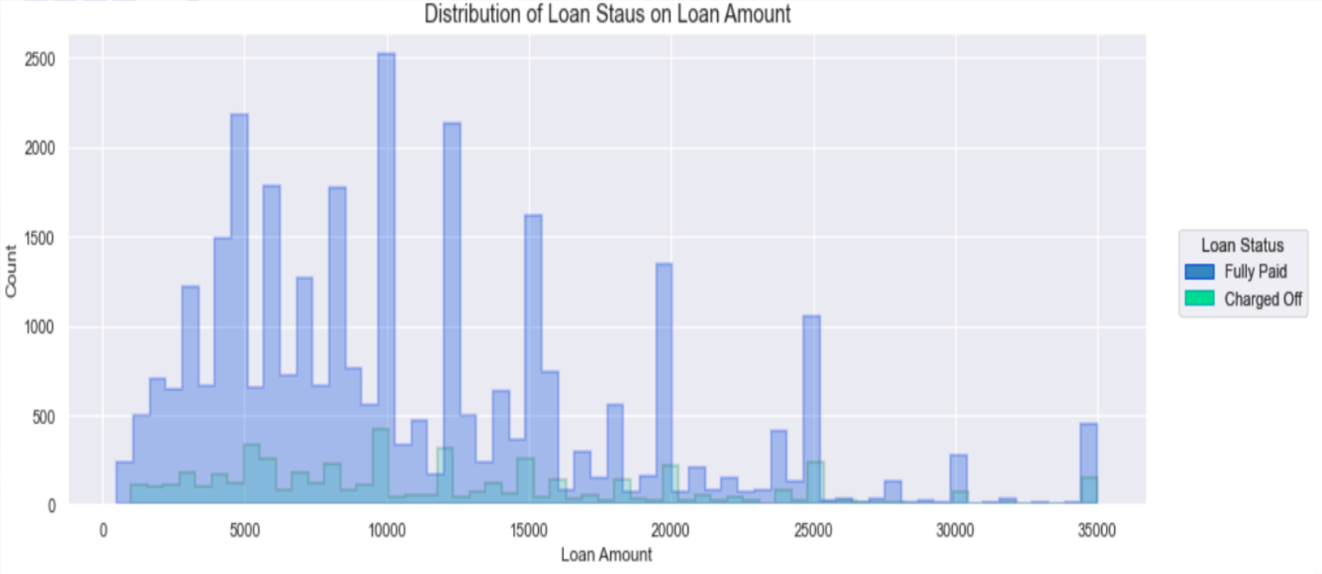


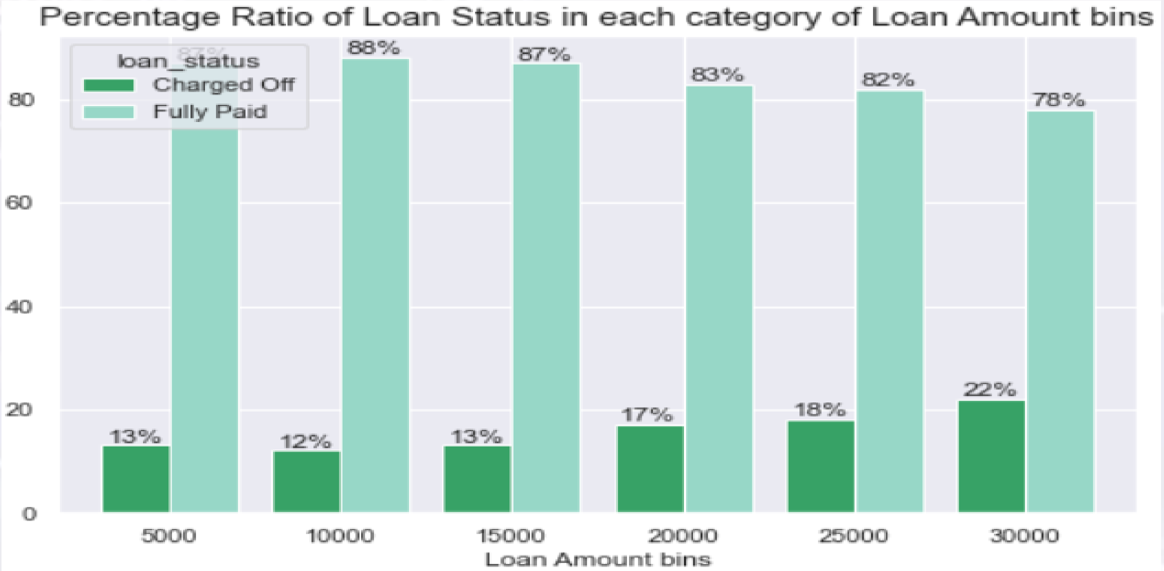


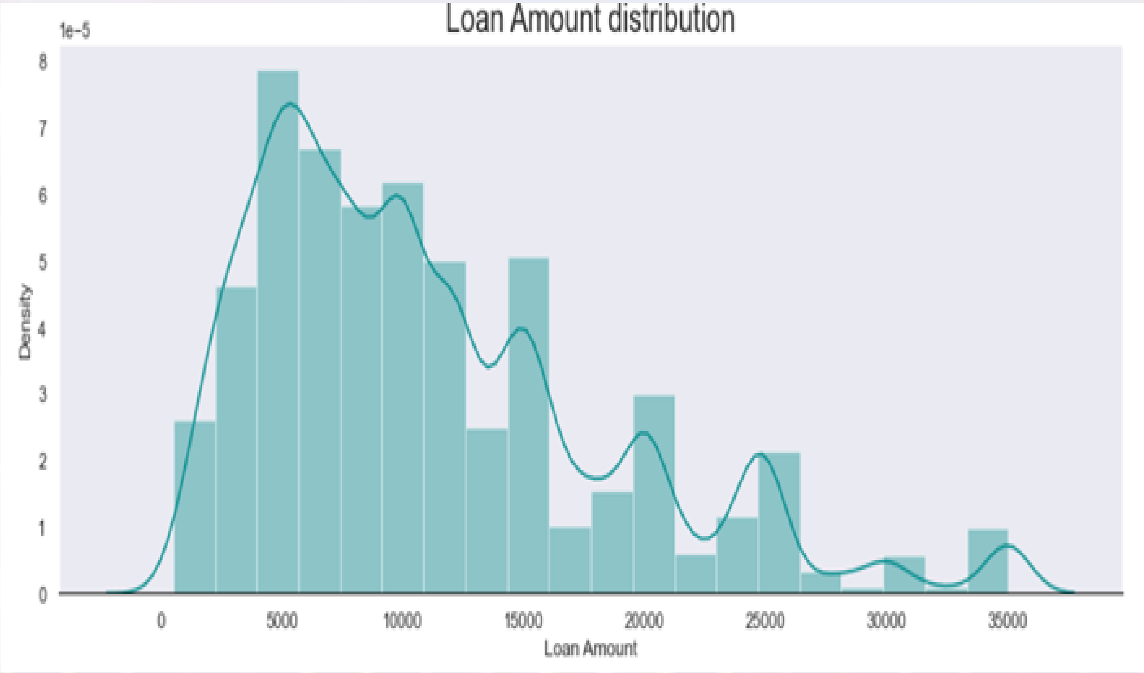
<u>INTEREST RATE:</u>

- Hex plot shows higher density of defaulters between 10 and 15% interest rate.
- The plot also signifies that most defaulters fall between 10-15% interest rate for Loan amount less than 10,000.
- Categorising Interest rate to bins (with 2.5 difference), We could see **anyone having more than 20% interest are likely to go default.**

NUMERICAL CONTINUOUS VARIABLE ANALYSIS:



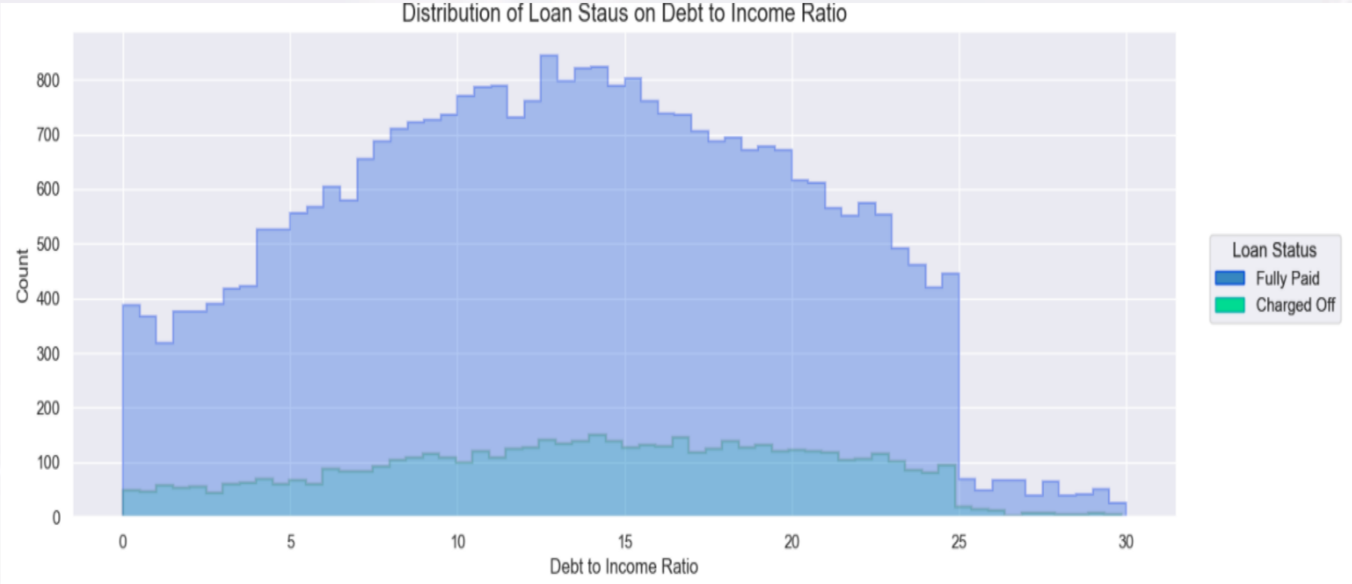


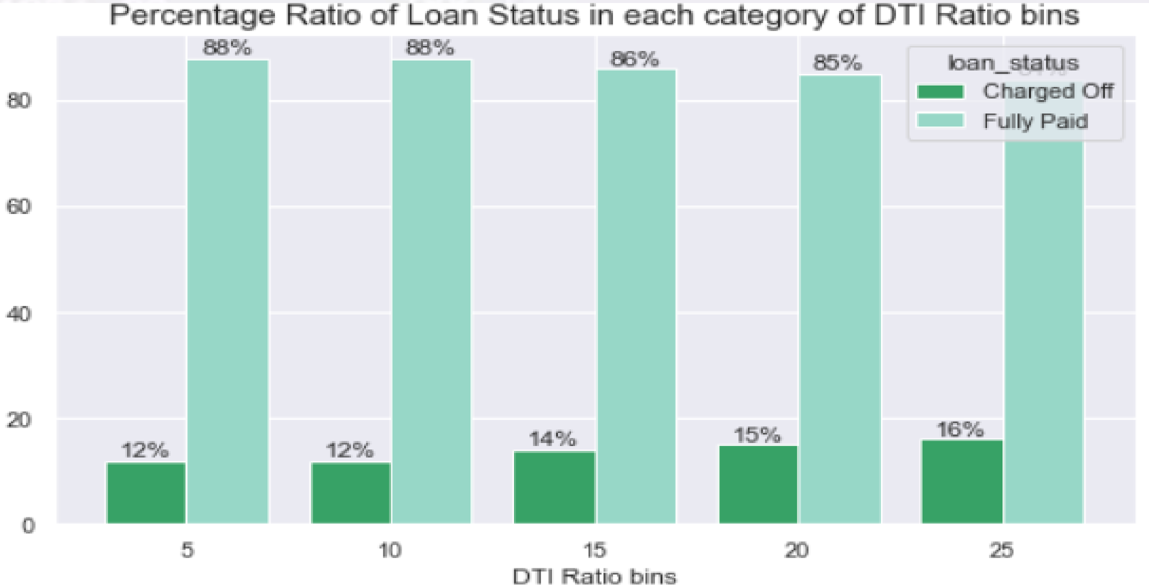


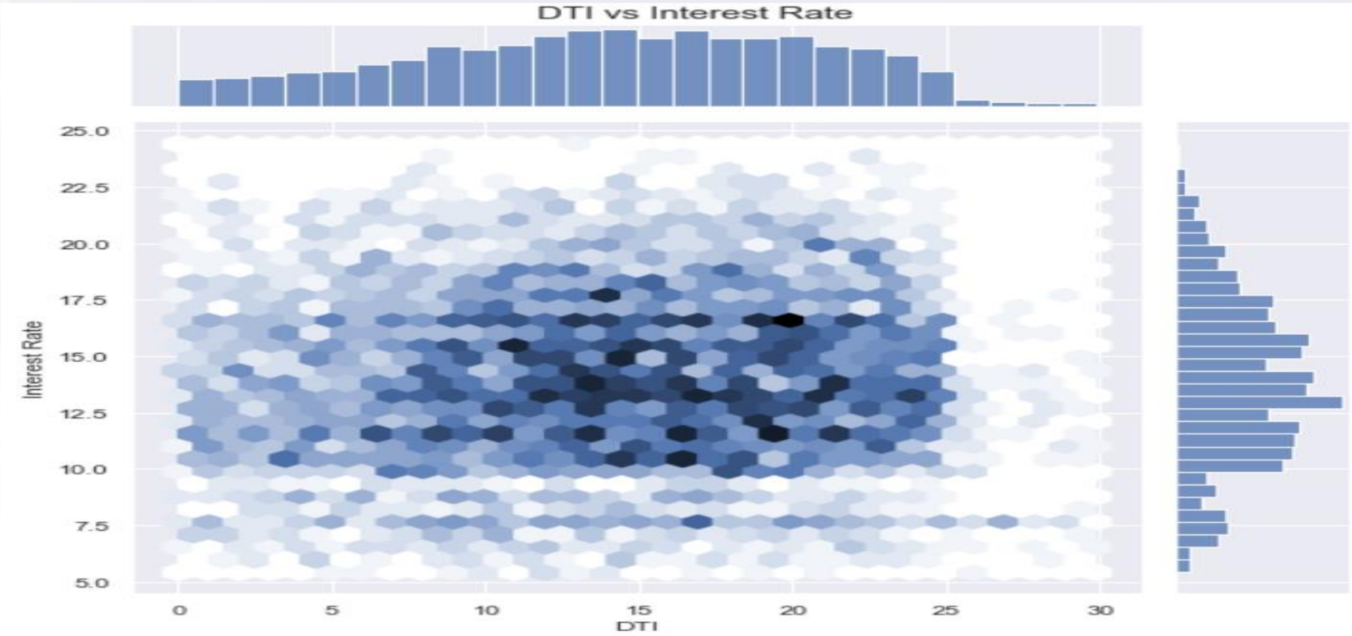
LOAN AMOUNT:

- In general, more loan amount gets raised between 5000 and 10000 dollars.
- Loan Amount is binned categorically at 5000$ interval and anything beyond 25000 will be considered as 30000.
- The distribution increases at faster rate from 0 to 5000 and gradually decreases as the Loan amount Increases making it a right skewed distribution.
- **Higher the loan amount, higher the chance of borrower going default.**

NUMERICAL CONTINUOUS VARIABLE ANALYSIS:



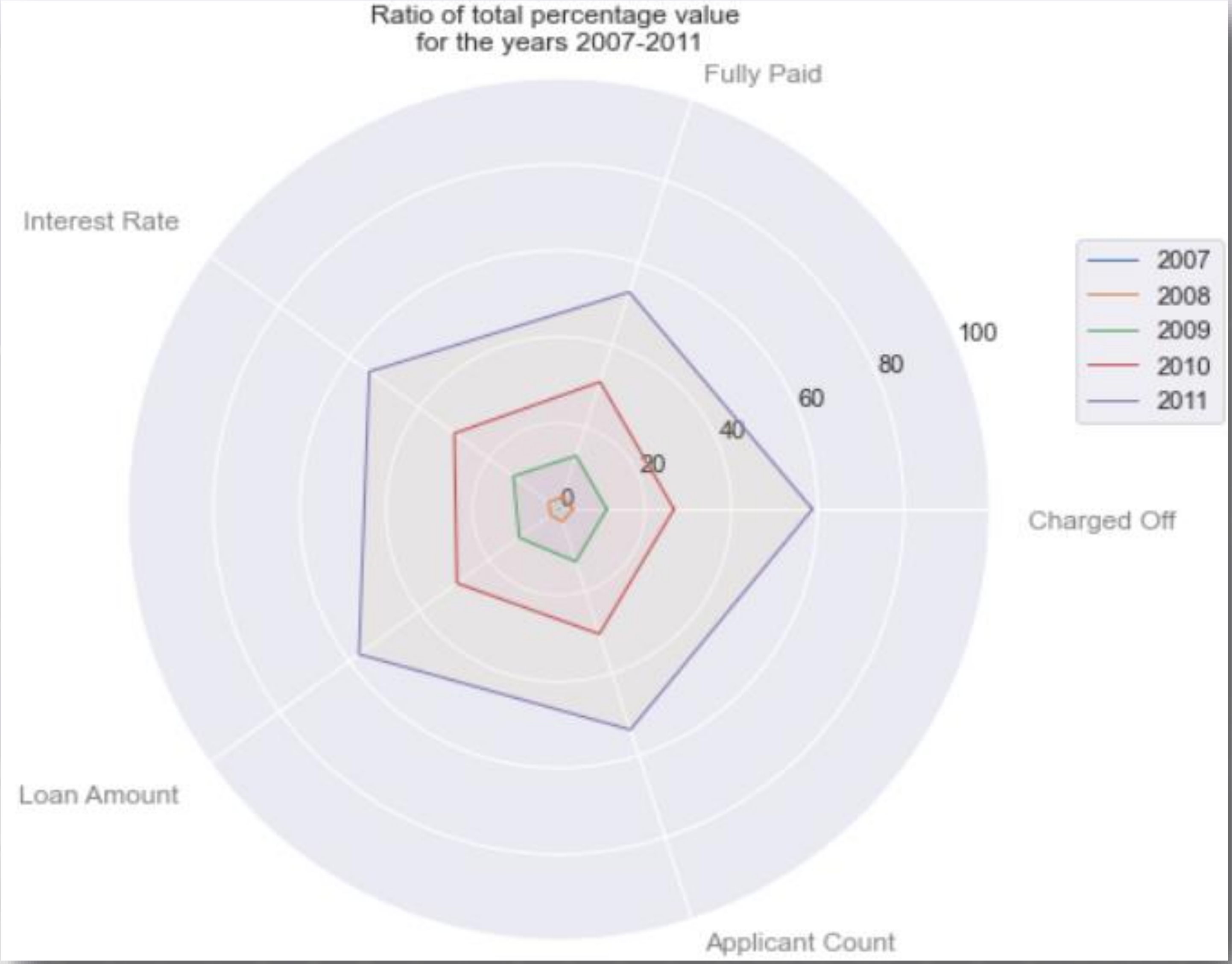




DTI: (Ratio of Total monthly debts to Total monthly Income)

- On Comparing DTI with Interest rate, we could see **higher volume of defaulters lying between 10 to 17.5 %**.
- DTI is categorically binned at 5% interval and anything beyond 20 will be considered 25.
- **Defaulters percent plot is more likely to be normally distributed and chance of going default is equally spread across DTI**.

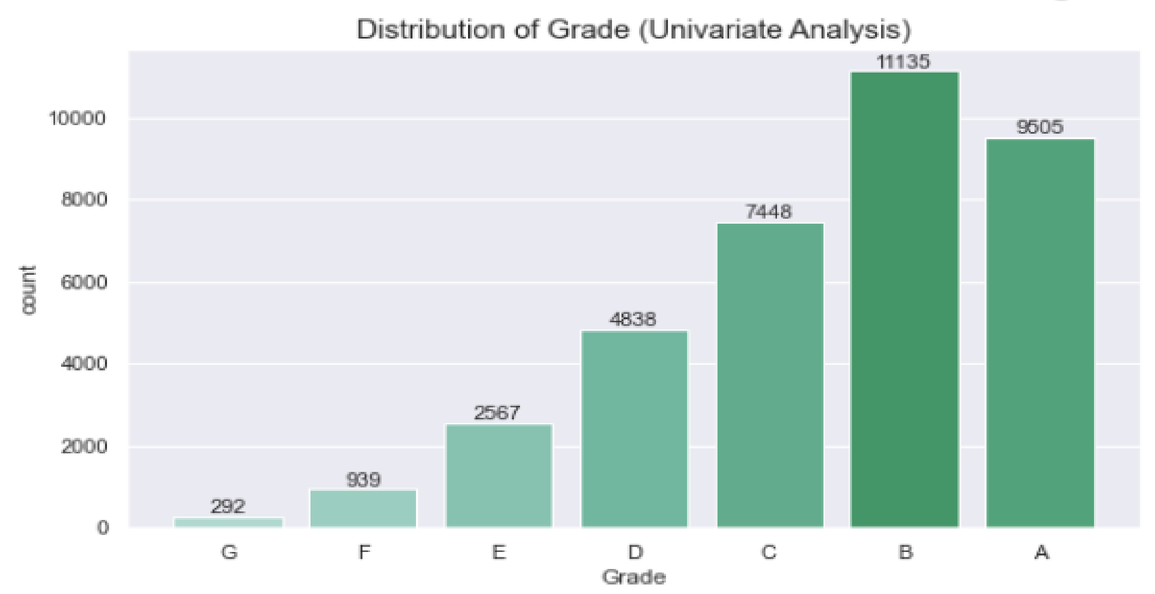
CATEGORICAL DRIVING VARIABLES:

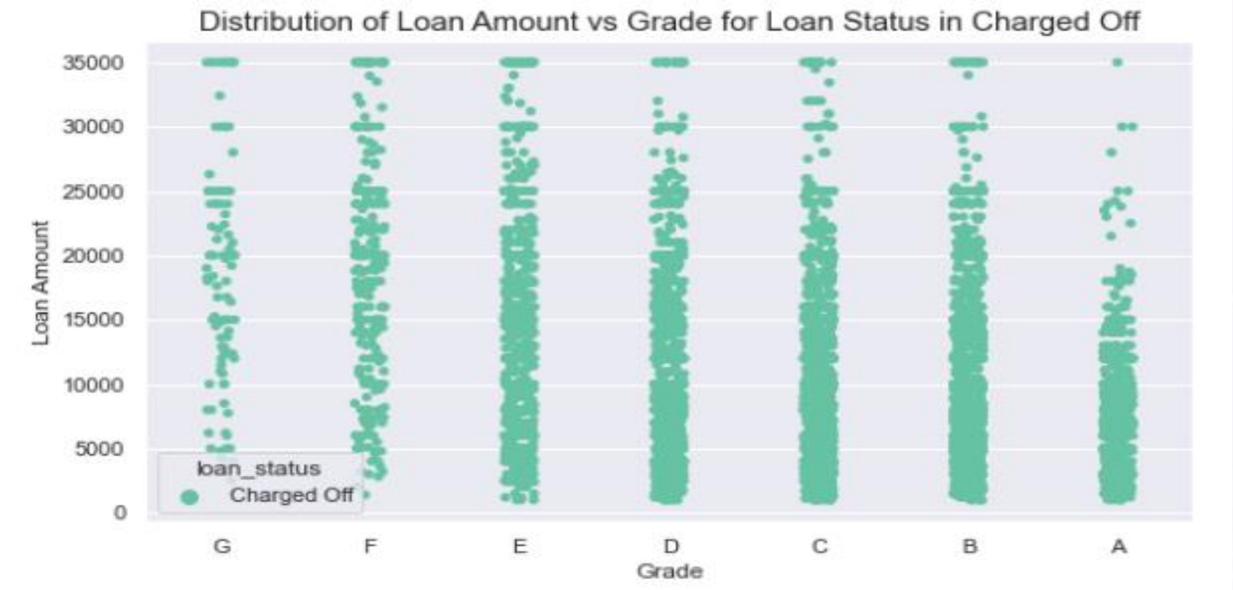


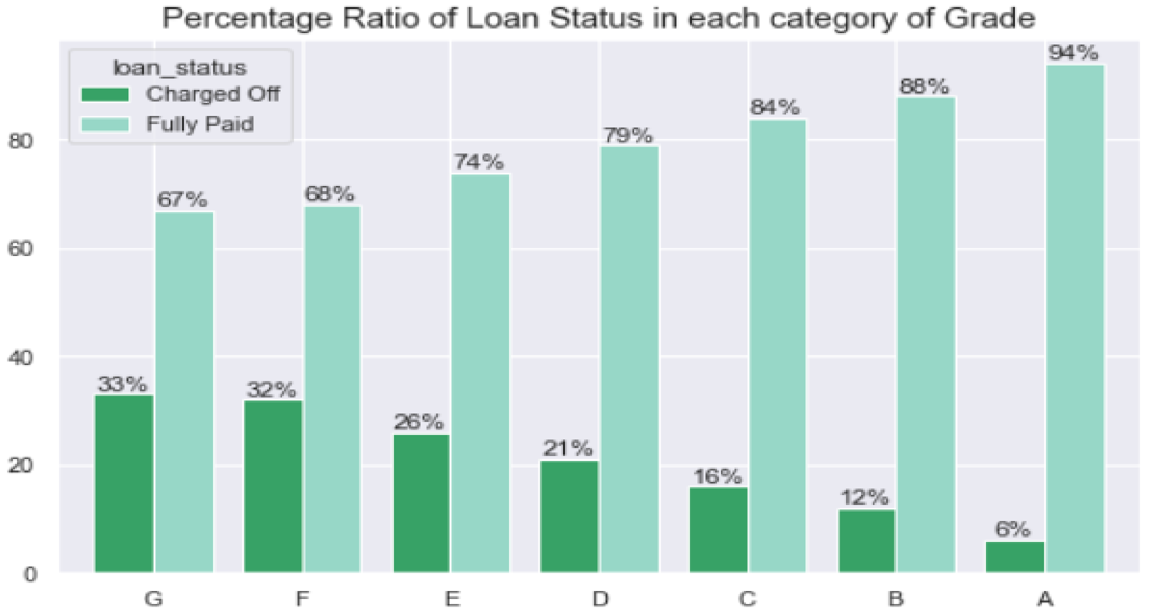
Radar Chart showing the ratio of total value for the features distributed between the years 2007-2011:

- On an Overview we could clearly see that the ration increases significantly as the each year.
- For the Applicant Count(loan issued) the chart depicts that in the year 2011 the applicants are about 50% of the total count.
- The Charged off ratio in years has been increasing a lot. Comparing 2010 which has ratio close to 27% with the year 2011 of ratio nearly 59% which is of about 32% increase.
- On calculating the Total Loan amount for these years 2011 has funded nearly 57% of the overall loan amount while 2010 funded only 29% of the overall sum.

CATEGORICAL VARIABLE ANALYSIS:



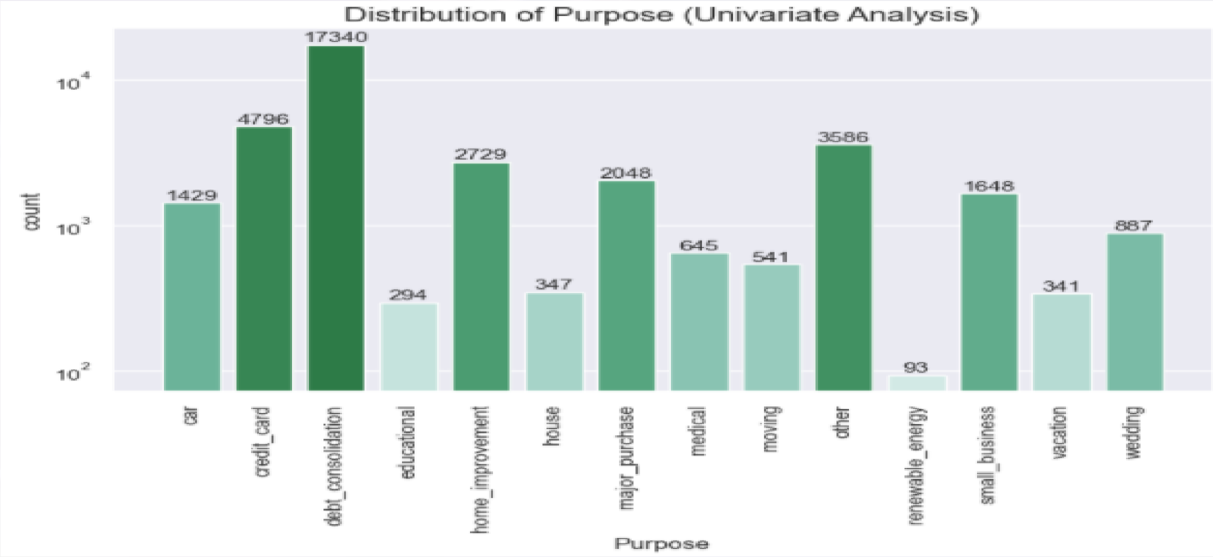


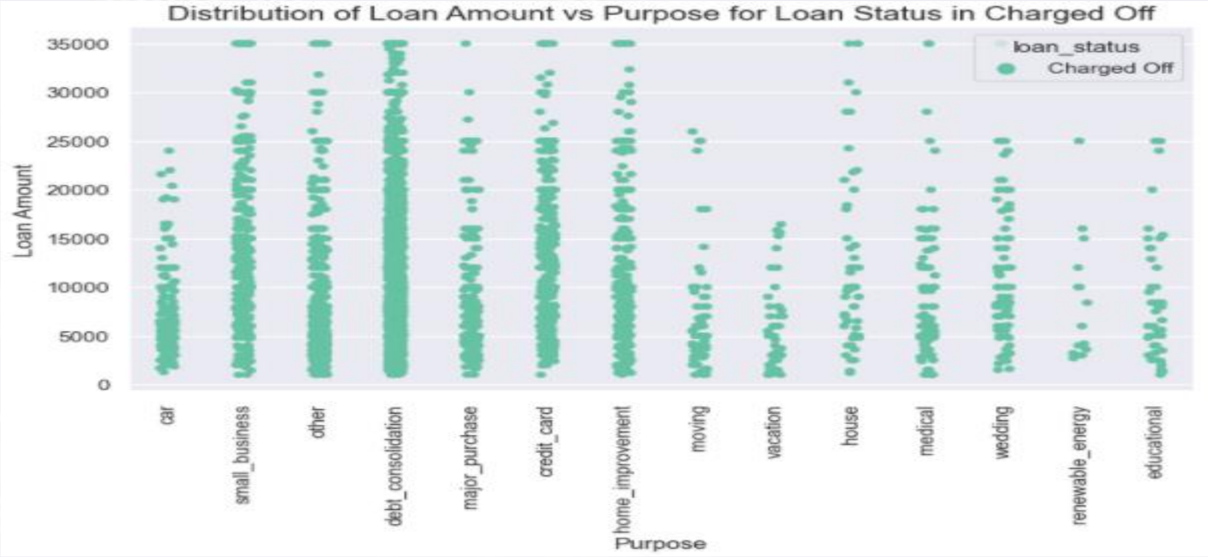


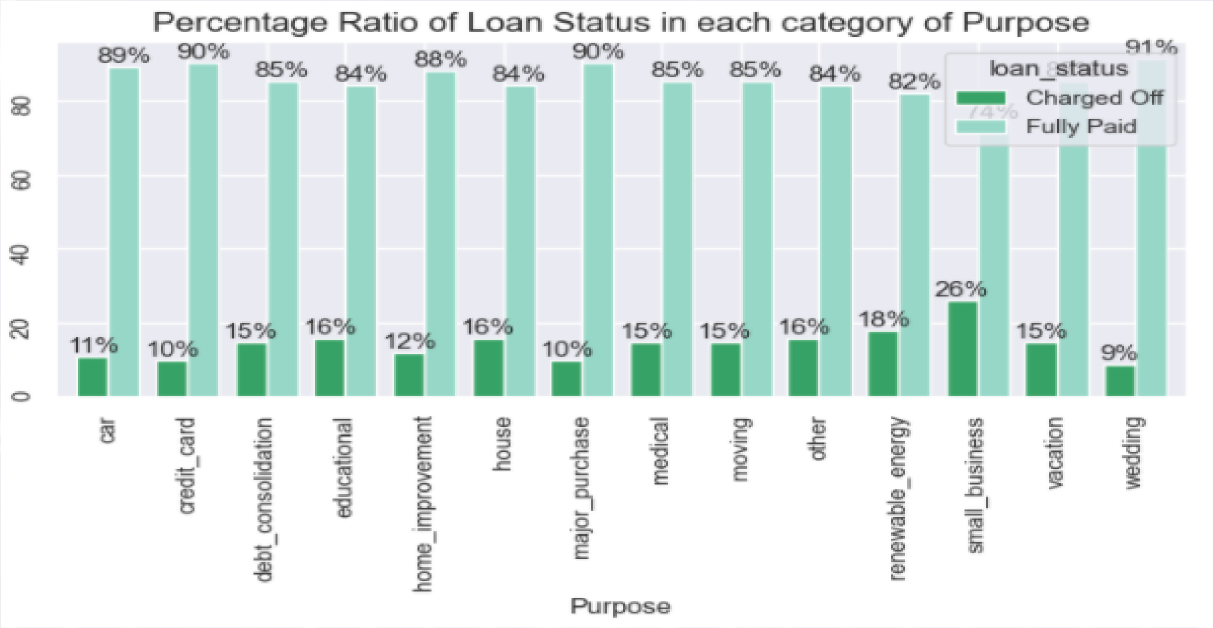
GRADE:

- We could see more loans raised by borrowers with low interest rate and volume of loans decreases over the Grades (from A to G)
- Grade G has higher defaulters due to higher interest rate and the defaulters rate decreases from G to A.
- **Anyone in Grade G and F are most likely to be defaulters.**

CATEGORICAL VARIABLE ANALYSIS:



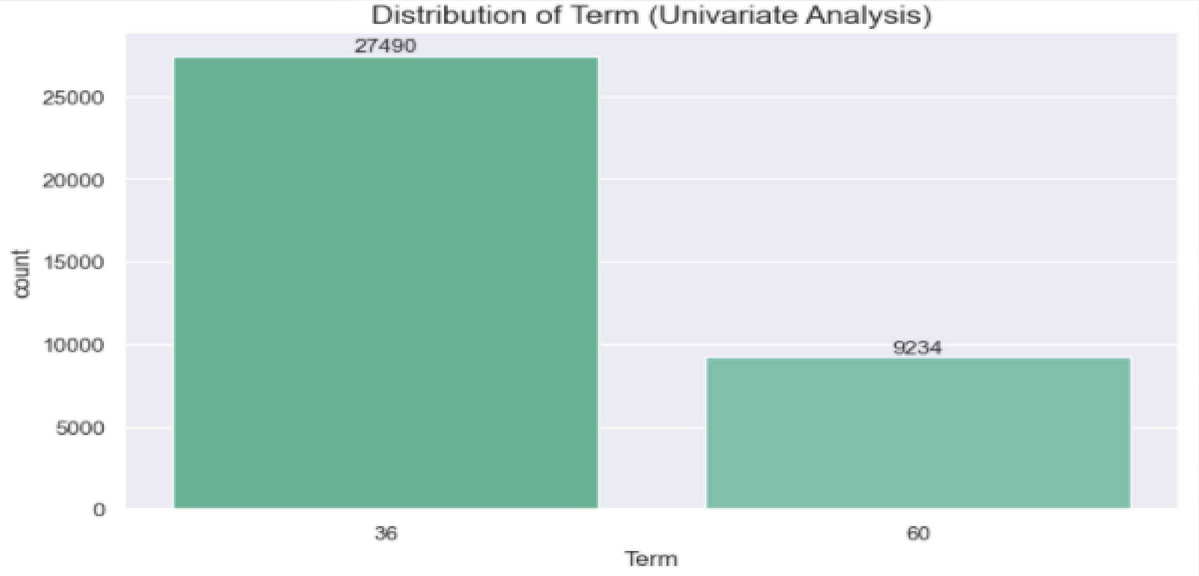


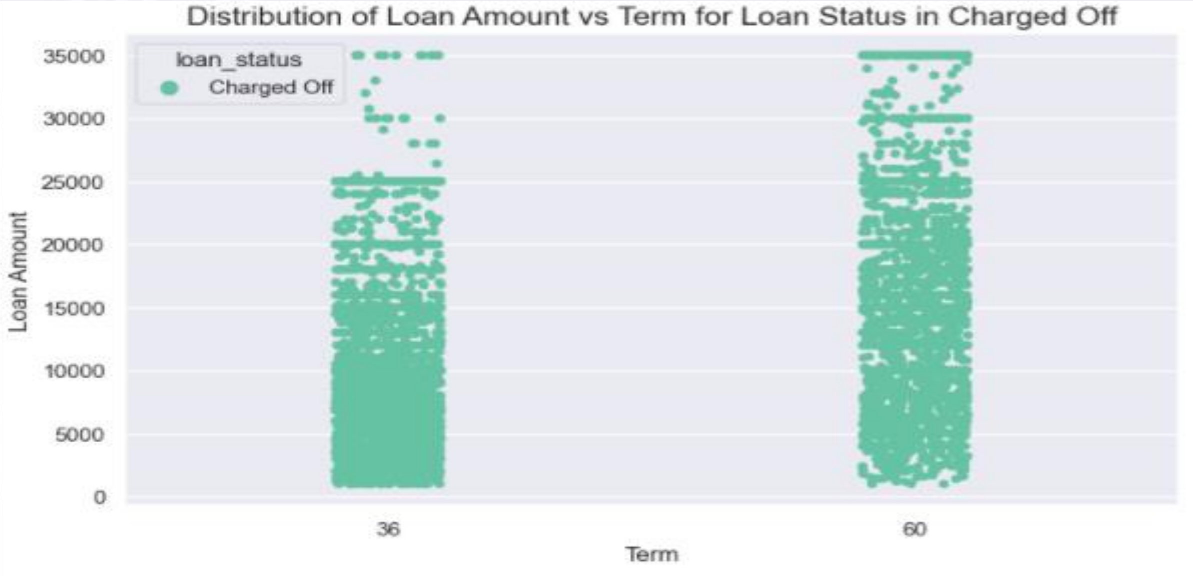


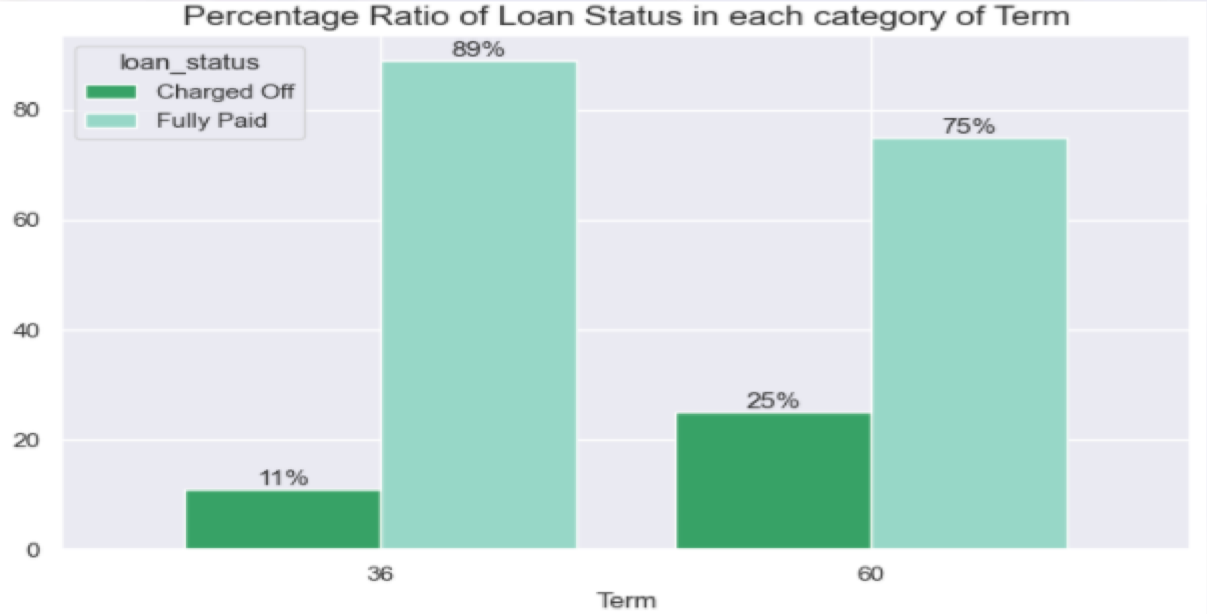
PURPOSE:

- More loans have been borrowed for Debt consolidation followed by Credit card and Others.
- **And Most defaulters are most likely to be in Small business and renewable energy as there is more financial risk involved.**

CATEGORICAL VARIABLE ANALYSIS:



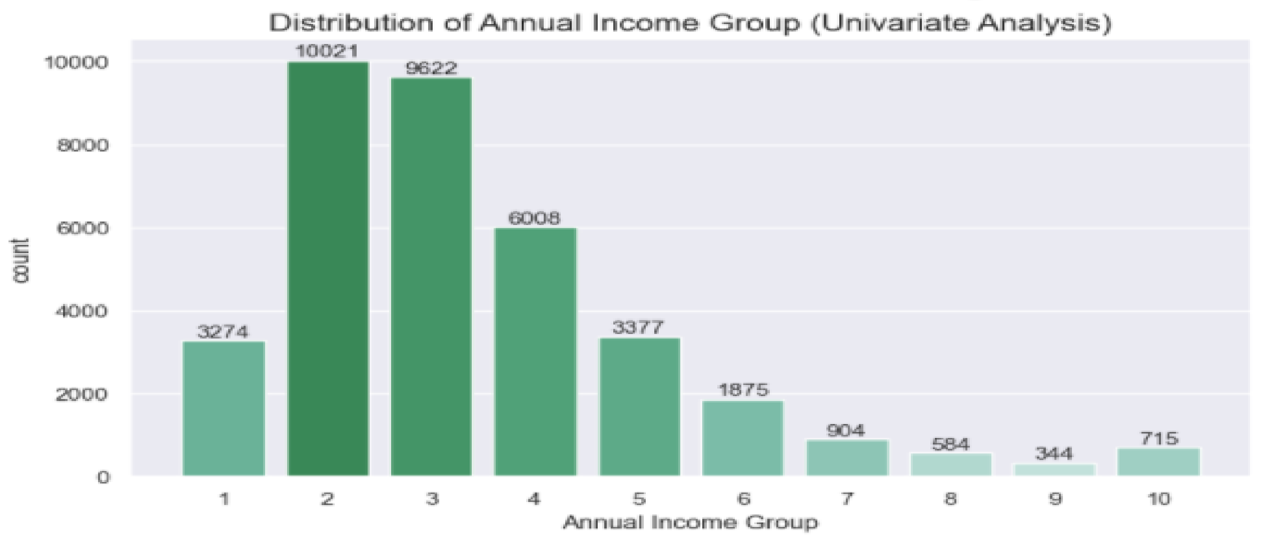


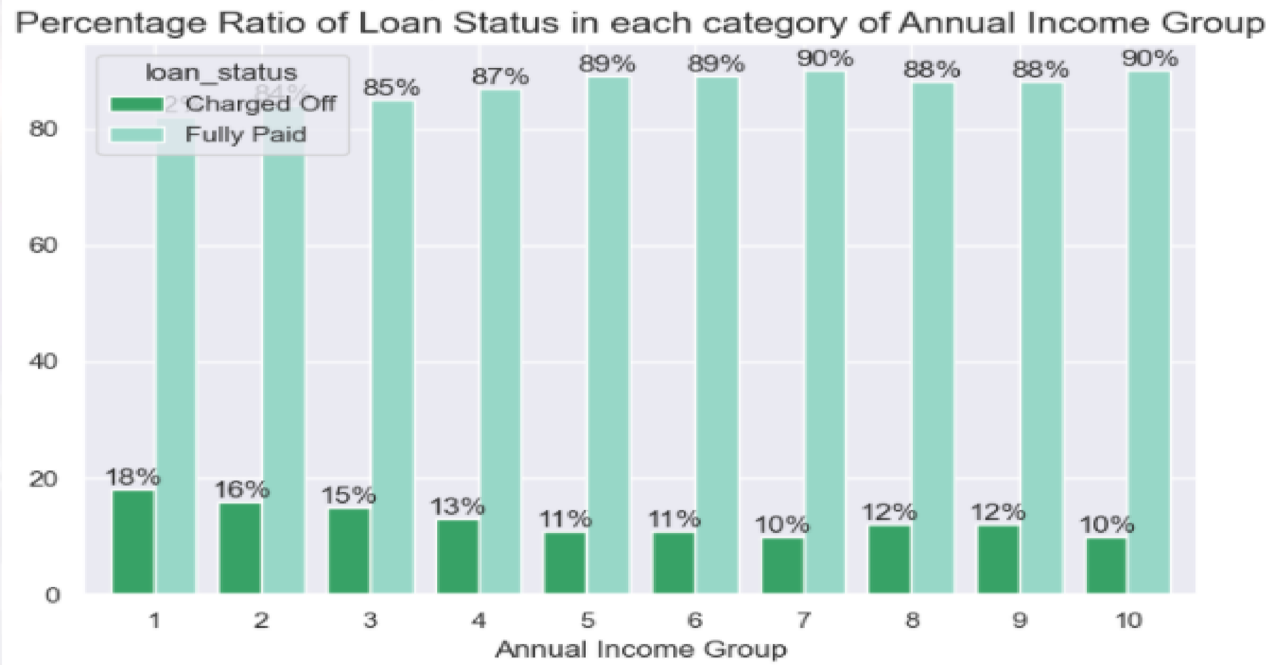


<u>TERM:</u>

- Higher the volume of loans under 36 months
- Interestingly with higher volume of loans under 36 month period, there are comparatively less defaulters.
- **It is likely that borrowers having 60 months tenure are more likely to default.**
- This can be an effect of time, as longer the loan gets, the borrowers commitment to repay could eventually go off.

DERIVED DATA ANALYSIS:

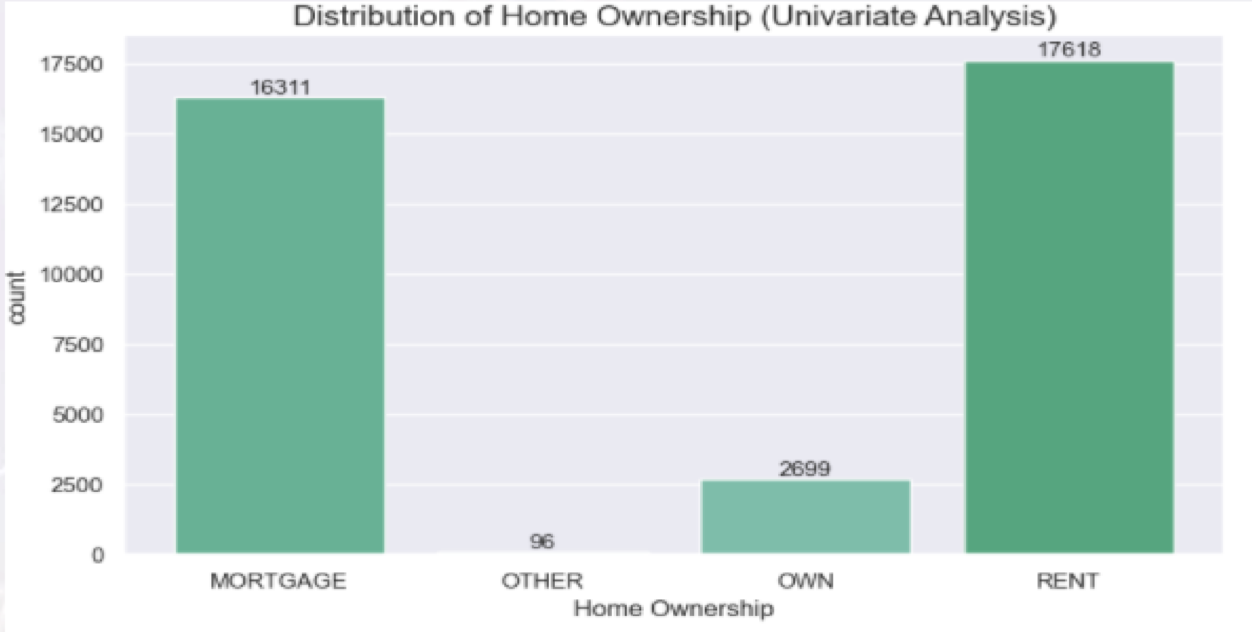


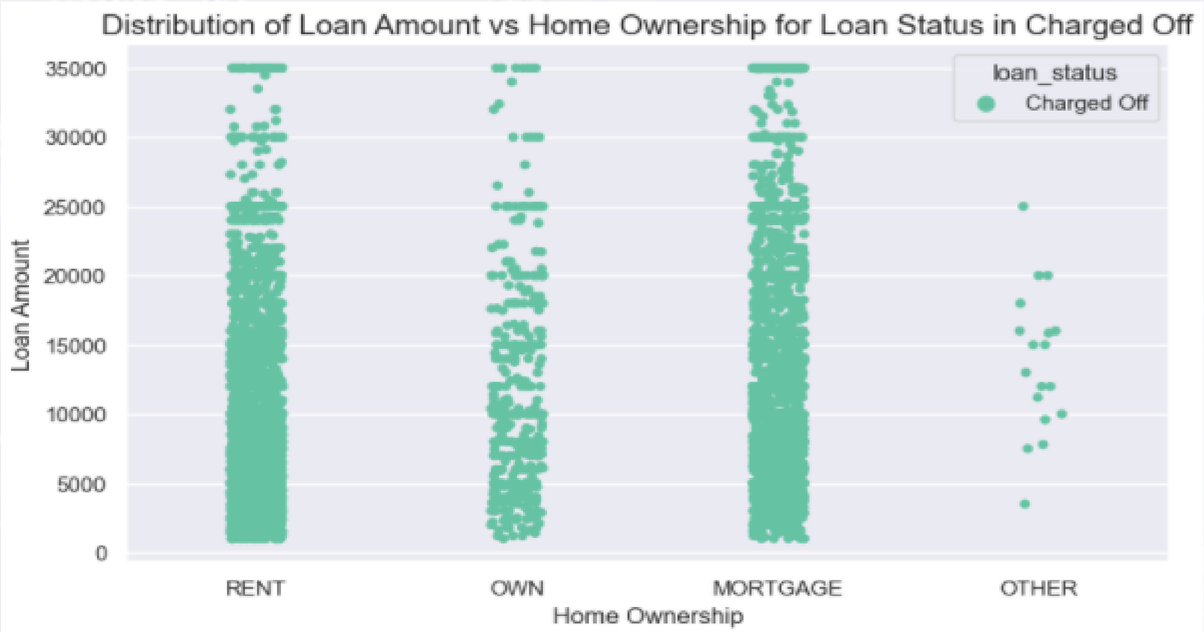


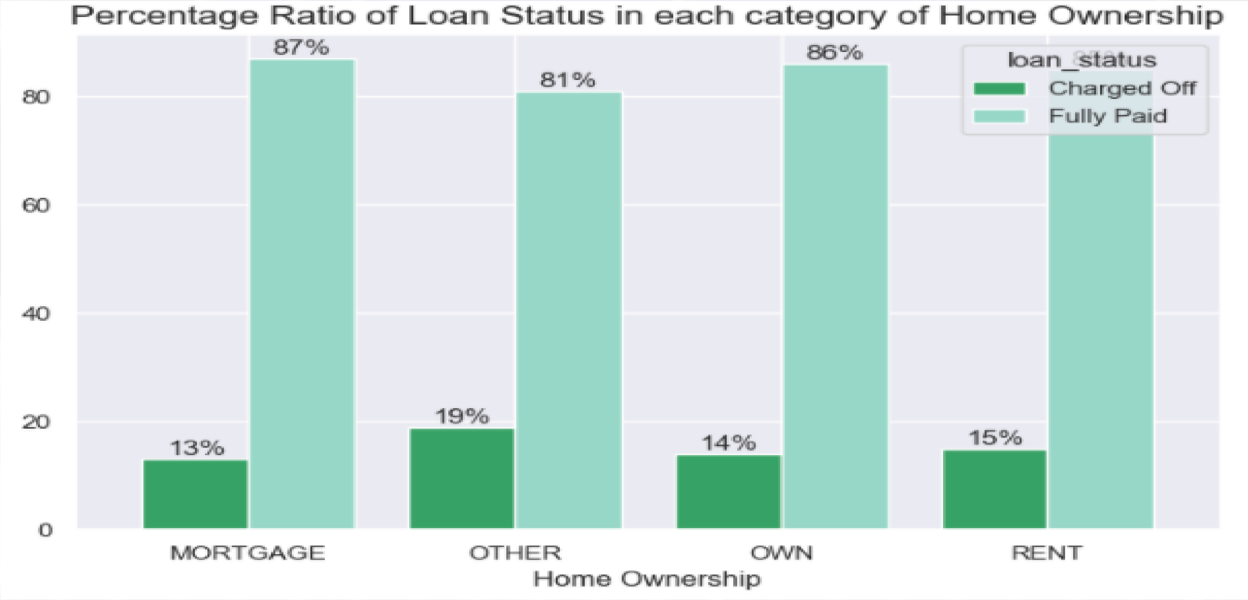
ANNUAL_INCOME_DERIVED:

- Annual income has been binned with 20K dollars interval with any value above 180K to be considered as 10.
- With borrowers holding less than 60K dollars income are more likely to get loans within 25,000$ and in this people below 60K dollars are likely to be defaulters.
- Higher goes the annual income, lesser gets the Defaulters rate.
- **Borrowers having less than 40K dollars as annual income are more likely to go default.**

CATEGORICAL VARIABLE ANALYSIS:



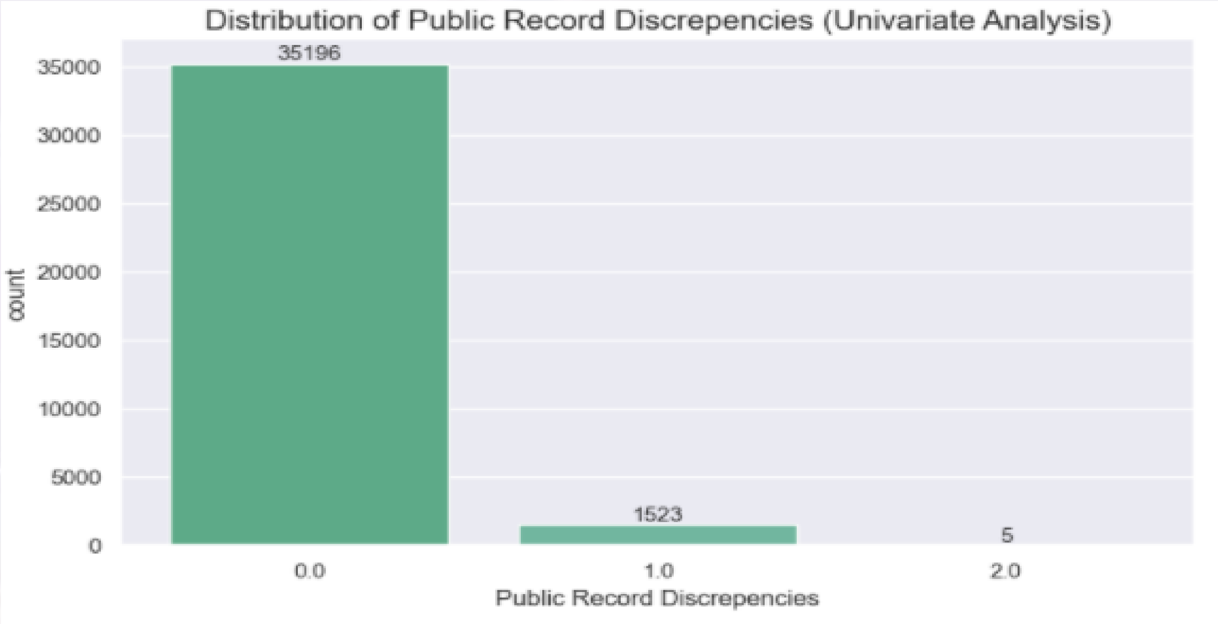


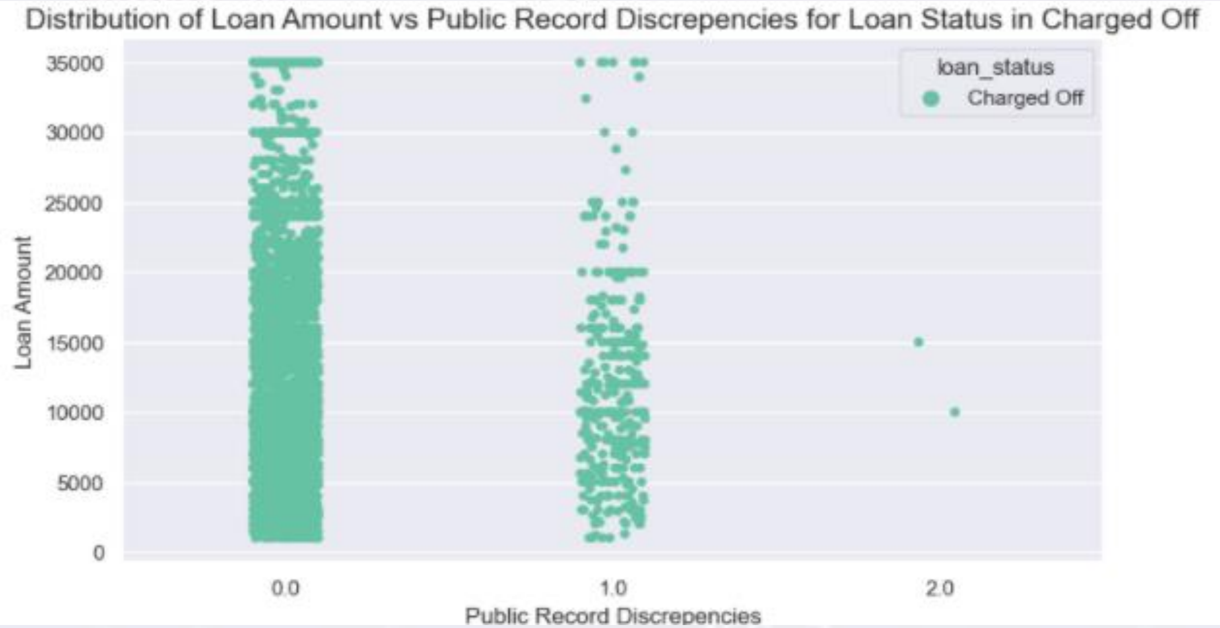


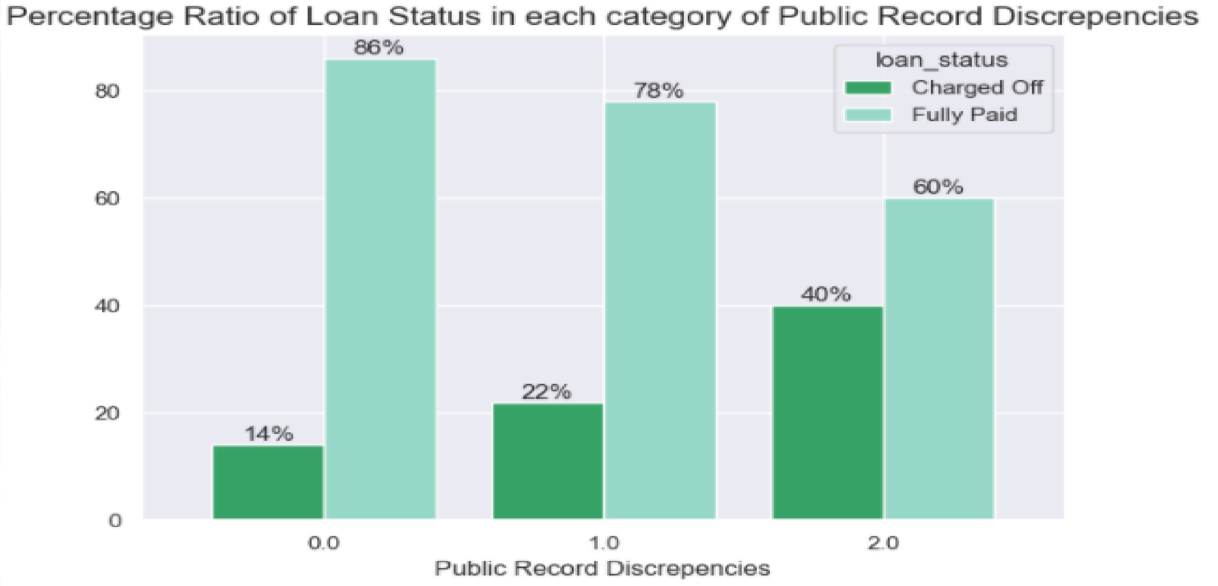
HOME OWNERSHIP:

- Higher volume of loans have been raised for Rent and Mortgage.
- Getting deeper with each section, It clearly depicts that **people who are in Other have higher chance of being Charged Off**.

CATEGORICAL VARIABLE ANALYSIS:



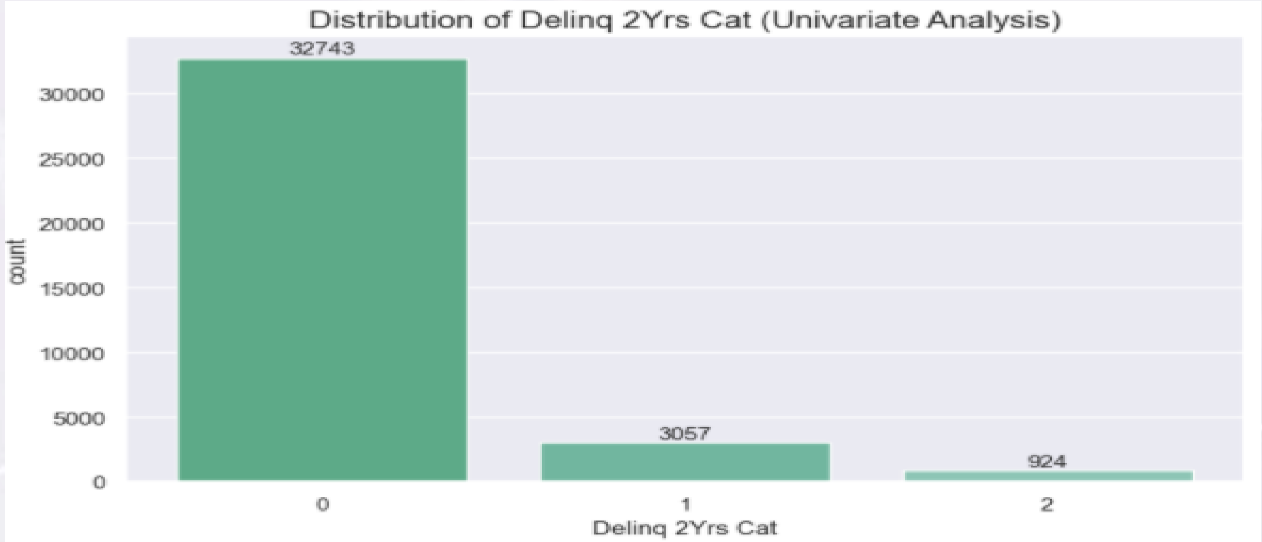


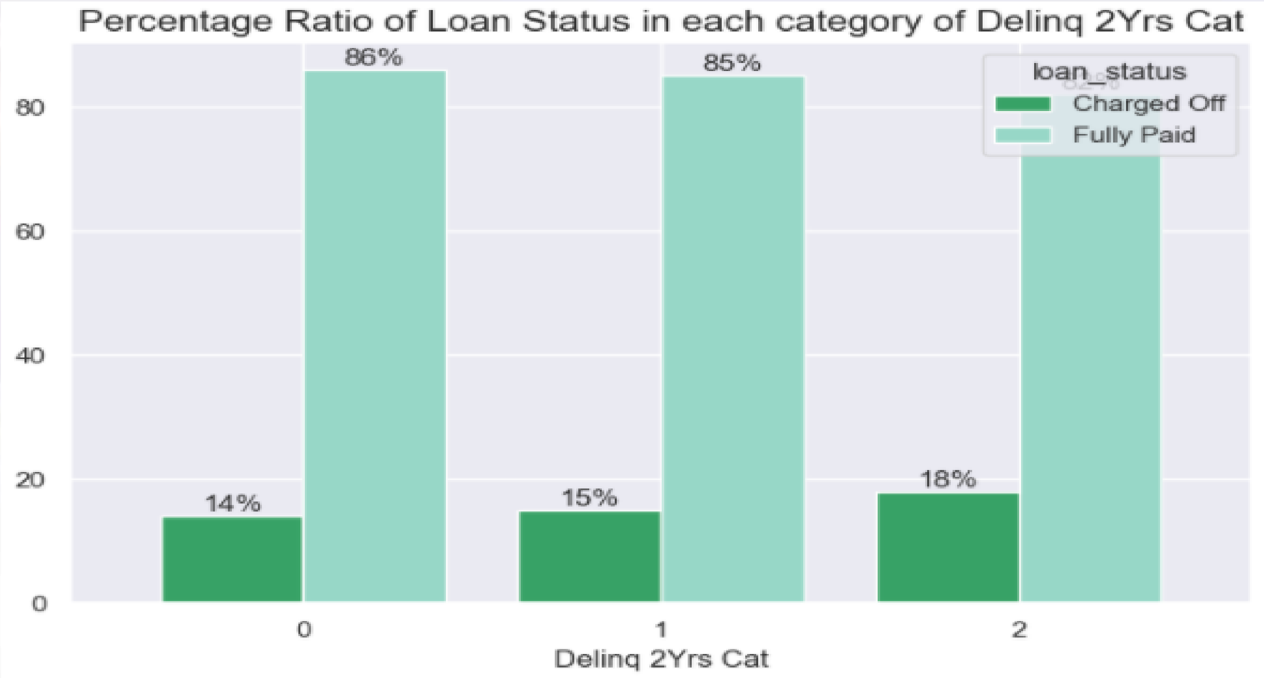


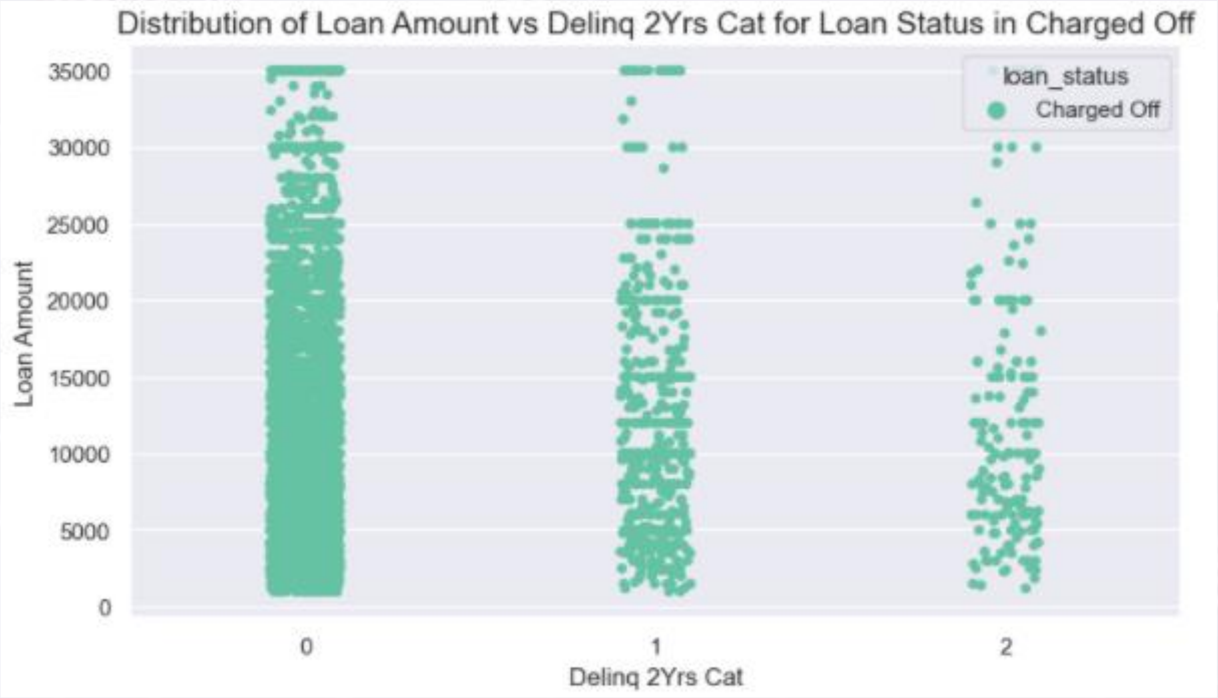
PUBLIC RECORD BANKRUPTCIES:

- We could see steep decrease in the number of loans approved for public record bankruptcies, Which implies the Business has already restricted the No of loans for users based on above mentioned parameters.
- As aforementioned restriction is believed to be in place, **borrowers with more than 0 bankruptcies are likely to be defaulters**.

CATEGORICAL VARIABLE ANALYSIS:







DELINQUIENT ON PAST 2 YEARS:

- Higher the delinq number of a borrower, higher is the chance of borrower running the path of a defaulter.
- **Anyone having more than 1 delinq in past 2 years are more likely to go default**.

CONCLUSION: DEFAULTERS DRIVING VARIABLES

- INTEREST RATE:

  With higher interest rate, the amount to be paid off also gets higher. Anyone above 20% Interest rate are can mostly go default.

- LOAN AMOUNT:

  Higher the Loan amount, higher will be the instalments and anyone borrowing more than 25000$ are most likely to go default. Same pattern follows for funded amount.

- GRADE:

  Higher the Grade(A-G), higher the defaulters percent. F and G are most likely to be defaulters here.

- PURPOSE:

  With high risk involved in small business and renewable energy, These 2 categories of Loan are most likely to take the path of Defaulters.

- TERM:

  As an effect of time, It is likely that borrowers having 60 months tenure are more likely to default compared to ones under 36 months.

- PUBLIC RECORD BANKRUPTCIES:

  With already being bankrupt, Anyone with 1 or more bankruptcies are likely to be defaulters.

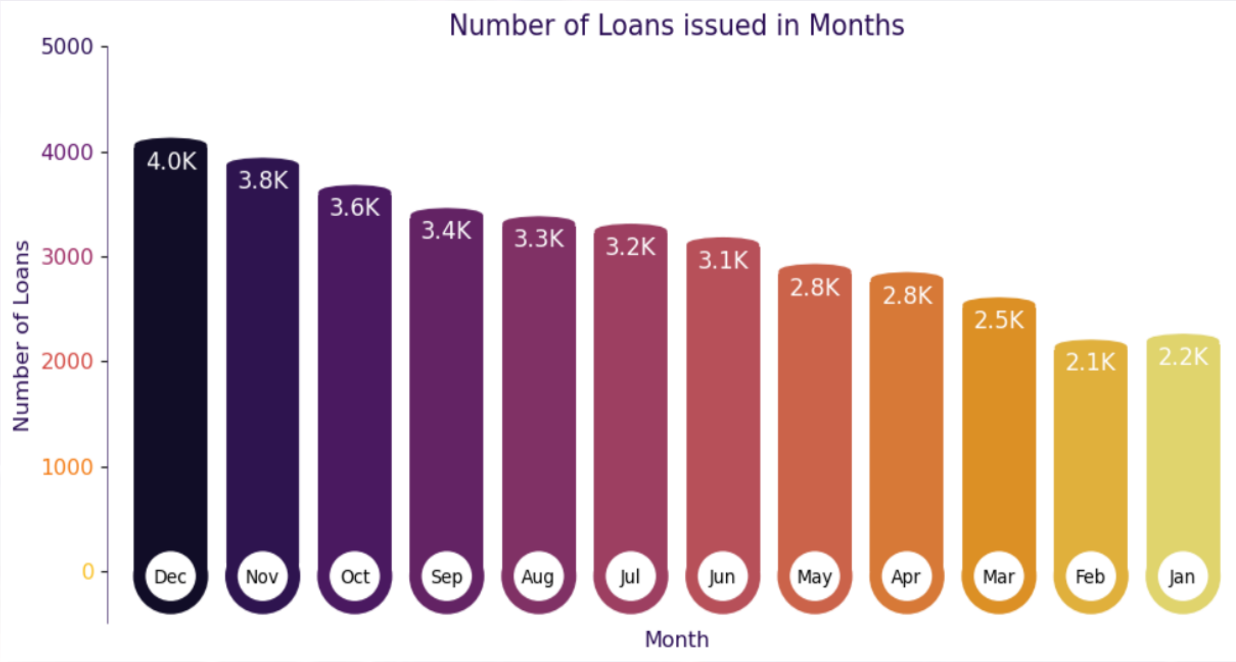
- ANNUAL INCOME:

  Borrowers holding less than 40K dollars as annual income are most likely to go default.

- DELINQ_2YRS:

  Anyone having more than 1 delinq in past 2 years are more likely to go default

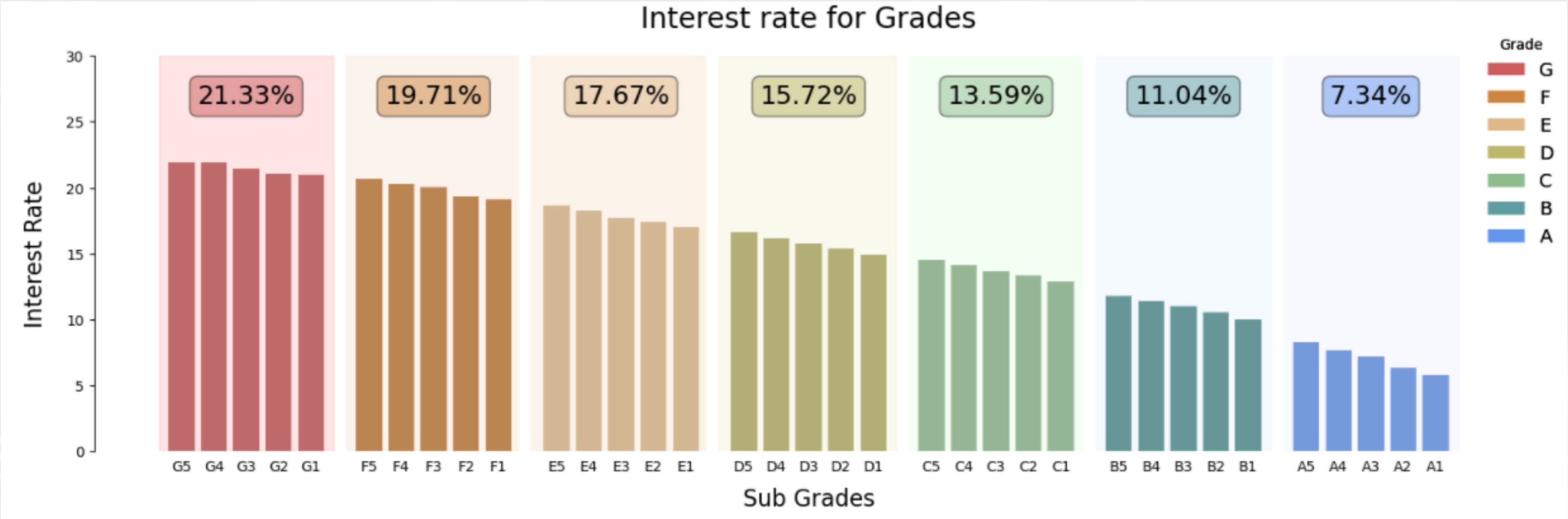
INTERESTING FACTS:



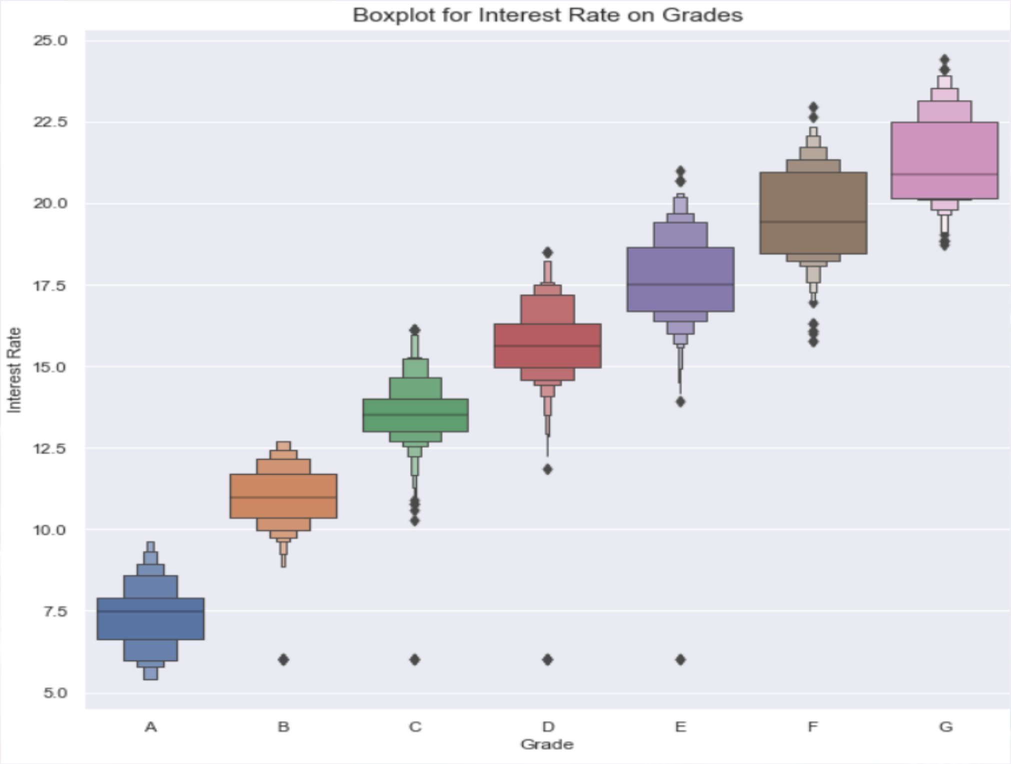
MONTH:

- We can see a trend where in **the no of loans being issued gets increased over every month of the year**.
- As one reason could be that US has many celebration and festivals towards the end of the year so people tend to take up a lot of loans over this period.

- SUB-GRADE Vs INTEREST RATE:
- Sub-Grade and Interest rate has a directly proportional relationship.
- **Higher the grade (from A to G), higher will be the interest rate.**

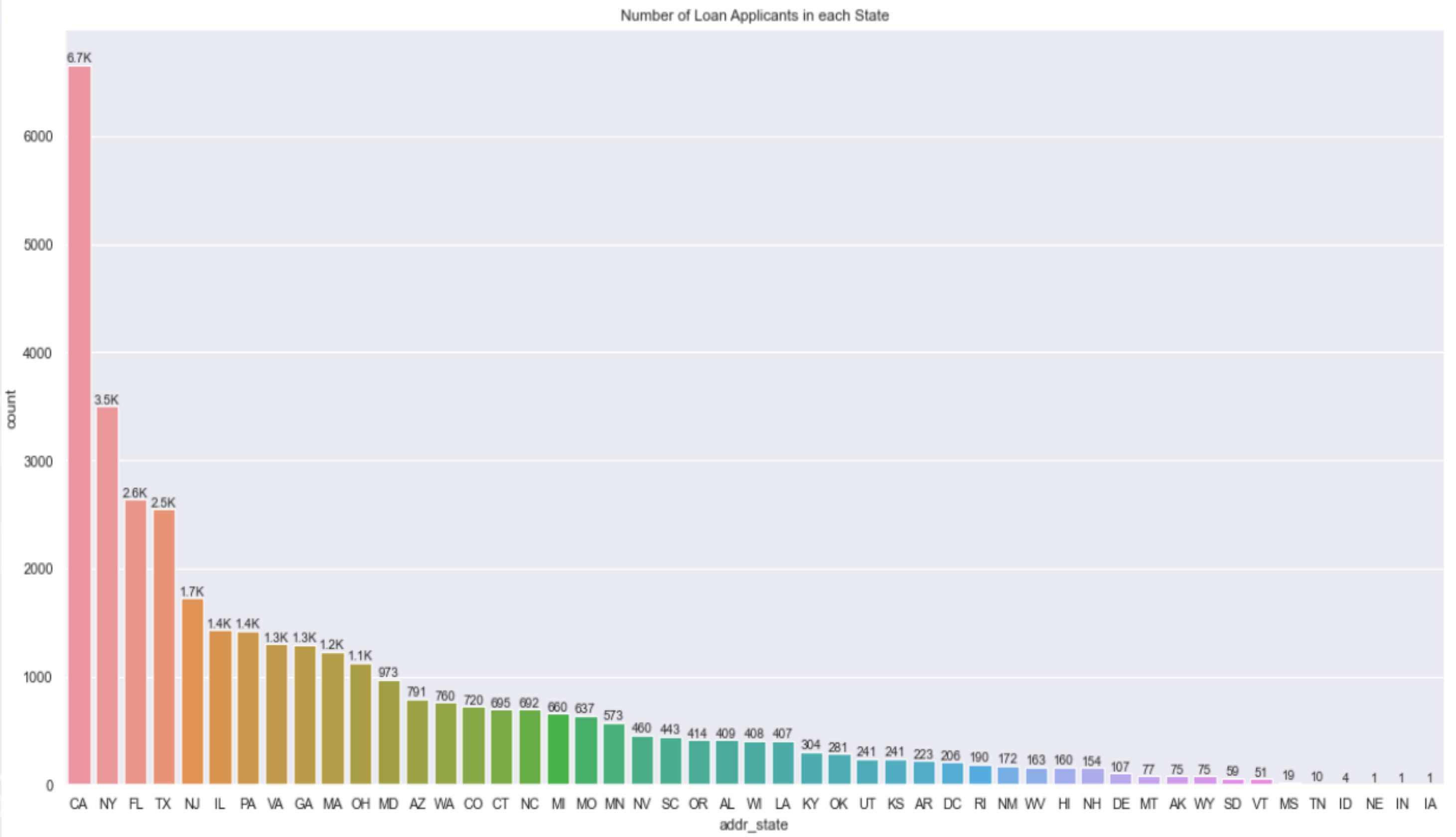


INTERESTING FACTS:



GRADE:

The Boxen plot clearly pictures that higher the grade (alphabetically), higher gets the interest rate.



STATE:

CA followed by FL and NJ are States with higher percent of Defaulters.