

DATA ANALYSIS REPORT:

LOAN DATASET

DATE: 30 JUNE 2019

BY

ANIKET GAIKWAD

• OVERVIEW OF THE PROBLEM STATEMENT:

- A data set was provided comprising data of bank loan details of the customers for a loan lending company.

AIM:

- In this case study, we used Exploratory Data Analysis (EDA) to understand how consumer attributes and loan attributes influence the tendency of default.

➤ THE FOLLOWING FACTORS WERE KEPT IN MIND:

When a person applies for a loan, there are two types of decisions that were considered:

- 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. These candidates are not labelled as 'defaulted'.
- 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

- **DATA SOURCE AND DESCRIPTION:**

The data was provided by Mr. Arjun Joshi Sir for the data analytics internship to evaluate my skills and providing valuable experience and exposure that will help me a lot to improve.

- **DATA SET CONTENT DESCRIPTION:**

- ✓ The given consist of the following:

NUMBER OF ROWS: 39717

NUMBER OF COLUMNS: 111

- ✓ Many of the columns present in the dataset were containing more than half of empty values, hence dropping those columns was preferable for better analysis.
- ✓ After data cleansing was performed, the dataset was reduced to:

NEW NUMBER OF ROWS: 38577

NEW NUMBER OF COLUMNS: 20

• IMPORTANT FEATURES IN THE DATASET:

The following features or attributes were evaluated as the most important features for finding loan defaulters:

- ✓ `loan_amnt`: The total loan amount.
- ✓ `term`: The no. of months for which the loan is issued.
- ✓ `int_rate`: The interest rate with which the loan was issued.
- ✓ `Instalment`: The monthly payment which needs to be paid.
- ✓ `Grade`: The grade of the customer. The grade can be between A to G.
- ✓ `emp_length`: The number of years of experience of the customer while working.
- ✓ `Purpose`: The reason behind the issue of the loan.
- ✓ `home_ownership`: The type of home ownership of the customer. It can have types like RENT, MORTGAGE, OWNED, OTHERS.
- ✓ `annual_inc`: The annual income of the customer.
- ✓ `addr_state`: The state in the United States where the customer belongs to.
- ✓ `verification_status`: The verification status of the customer. It can have the values: Verified, Source Verified, Not Verified.
- ✓ `delinq_2yrs`: Number of delinquent accounts a customer has linked with him.
- ✓ `inq_last_6mths`: Number of enquiries in the last 6 months.

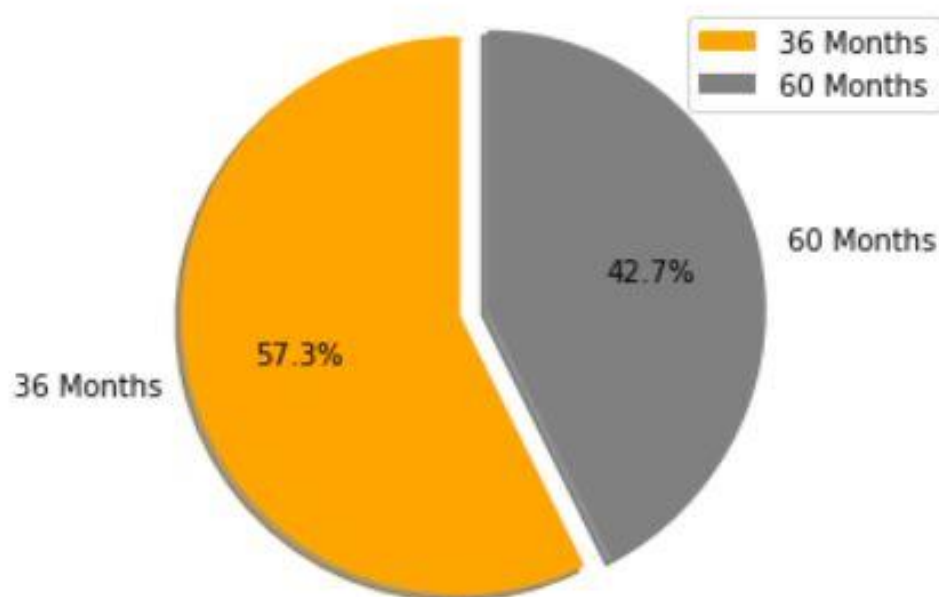
- ✓ `open_acc`: Number of open credit lines in the customer's account.
- ✓ `pub_rec`: Number of derogatory public records.
- ✓ `revol_bal`: Total credit revolving balance.
- ✓ `revol_util`: Amount of credit the borrower is using relative to all available revolving credit.
- ✓ `total_acc`: The total number of credit lines currently in the borrower's credit file.
- ✓ `pub_rec_bankruptcies`: Number of bankruptcies occur in a customer's account.
- ✓ `loan_status`: The current loan status of the customer.

- The features mentioned above had relevance to our aim.
- EXPLORATORY DATA ANALYSIS(EDA) was performed with considering the features mentioned above and the following insights were found out which may be essential for our aim.

- **INSIGHTS FOUND WITH RELEVANCE TO THE AIM:**

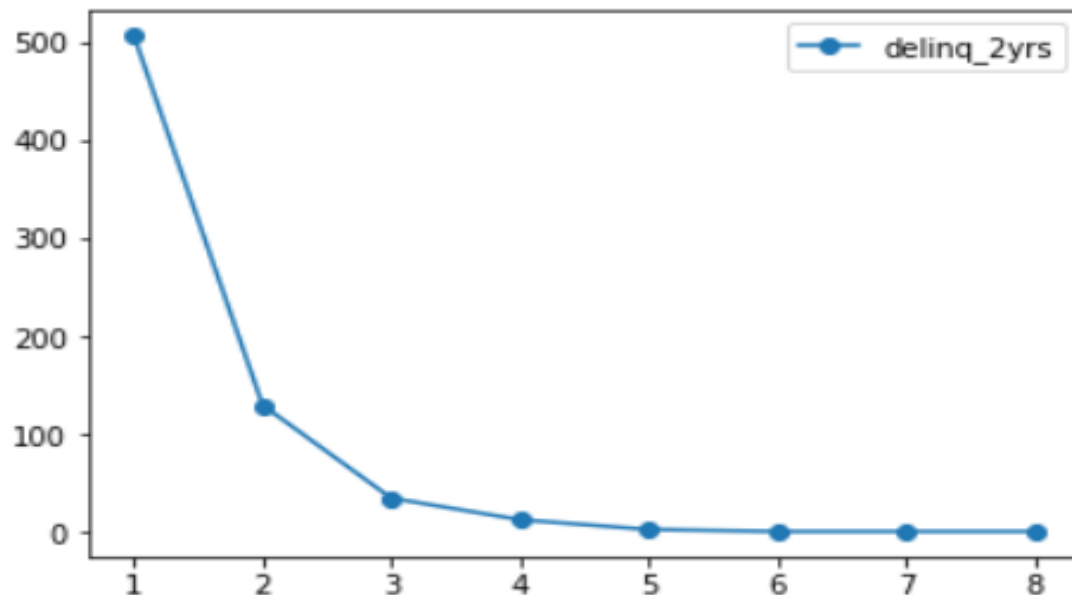
UNIVARIATE ANALYSIS:

1. PERCENTAGE OF DEFAULTER WITH RESPECT TO TERM:



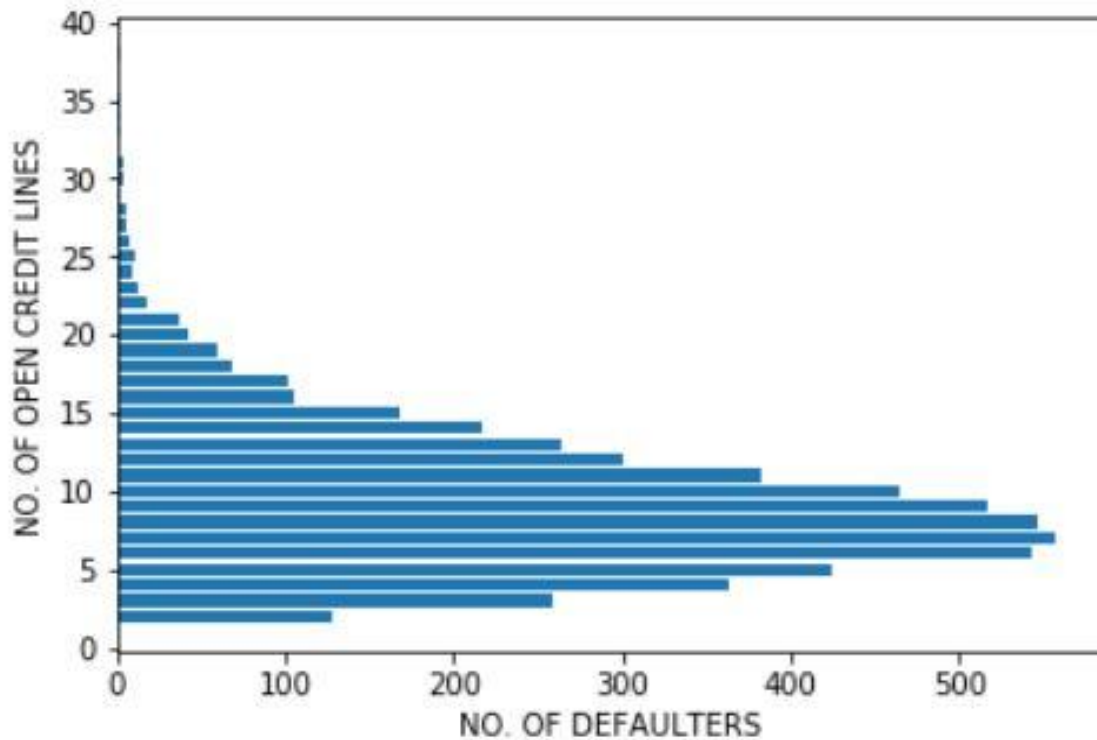
- ✓ The above pie chart shows the percentage of defaulters with respect to the term duration which can be either 36 Months or 60 Months.
- ✓ Clearly, the customers with term duration of 36 months default more than the others.
- ✓ The company should check the term duration applied by the new costumer for reducing losses.

2. NUMBER OF DEFAULTERS WITH RESPECT TO NUMBER OF DELINQUENT ACCOUNTS:



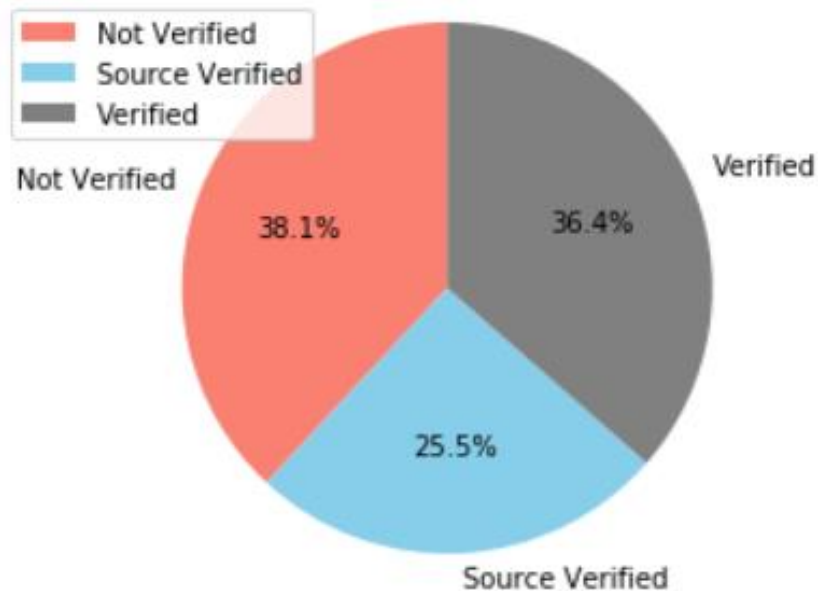
- ✓ The graph above shows the number of defaulters having the number of delinquent accounts associated with them.
- ✓ From the graph above, we can observe that there are more customers having number of delinquent accounts as 1 than the customers having number of delinquent accounts as 2.
- ✓ The curve reduces as the number of delinquent accounts increases, hence we can make an observation that lesser the number of delinquent accounts, more chances of the customers being a defaulter.

3. RELATION OF NO. OF DEFAULTERS WITH NO. OF OPEN CREDIT LINES:



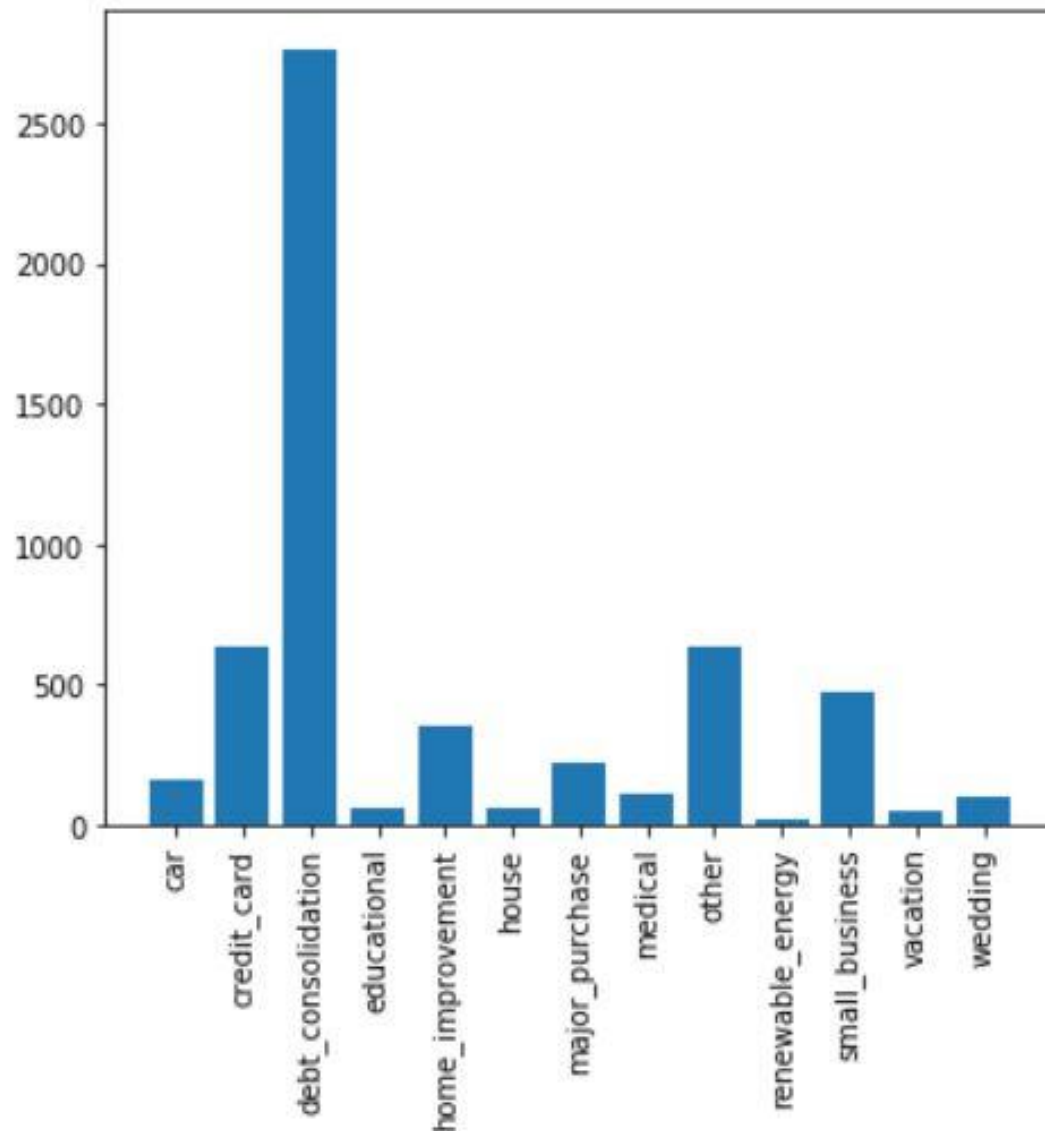
- ✓ The graph above shows the number of defaulters with the number of open credit lines present in the customer's account.
- ✓ From the above, we can observe that customers having 5 to 10 open credit lines in their account have the tendency to default the most.
- ✓ Hence, the company should look at the open credit lines present for better evaluation.

4. PERCENTAGE OF DEFAULTERS ACCORDING TO THE VERIFICATION STATUS:



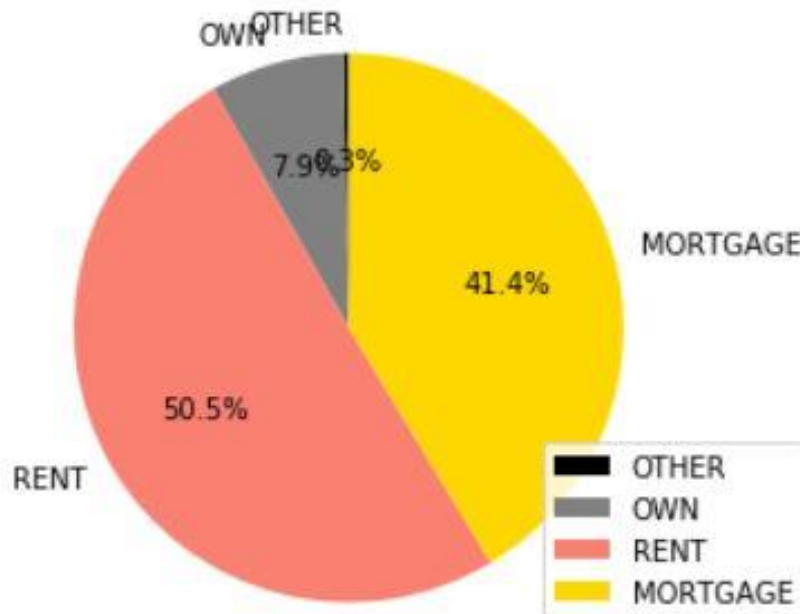
- ✓ The graph above shows the percentage of defaulters with respect to verification status.
- ✓ Looking at the pie chart, we can observe that customers having Verification status as 'Verified' and Not Verified have the tendency to default.

5. LOAN PURPOSE FOR THE PEOPLE HAVING LOAN STATUS AS CHARGED OFF(DEFAULTERS):



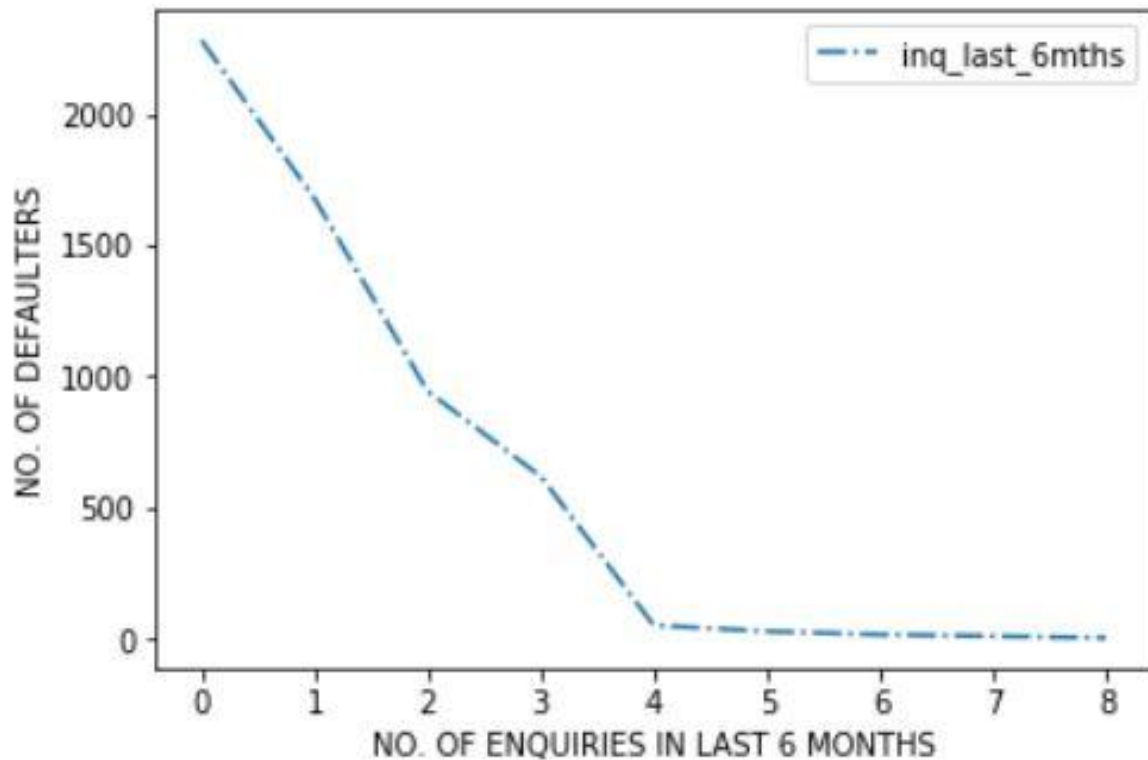
- ✓ The graph above shows the purpose or reasons behind the issuing of loan by the defaulter.
- ✓ We can observe that customers with purpose as 'Debt Consolidation' tend to default more prominently.

6. NO. OF DEFAULTERS ACCORDING TO THEIR HOUSE OWNERSHIP:



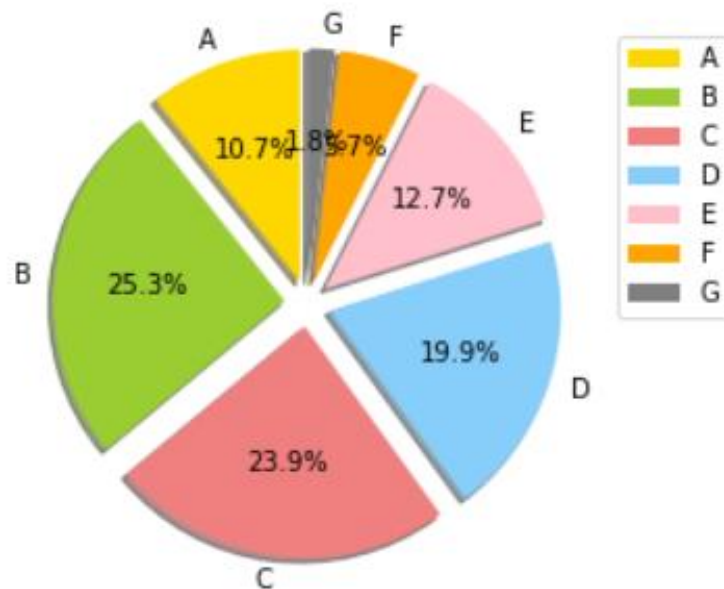
- ✓ Here, we can observe that customers having house ownership status as 'RENT' or 'MORTGAGE' have the tendency to default.
- ✓ The company should check the house ownership status of the new customers to evaluate the tendency of defaulting.

7. NO. OF DEFAULTERS WITH RESPECT TO NO. OF ENQUIRIES IN LAST 6 MONTHS:



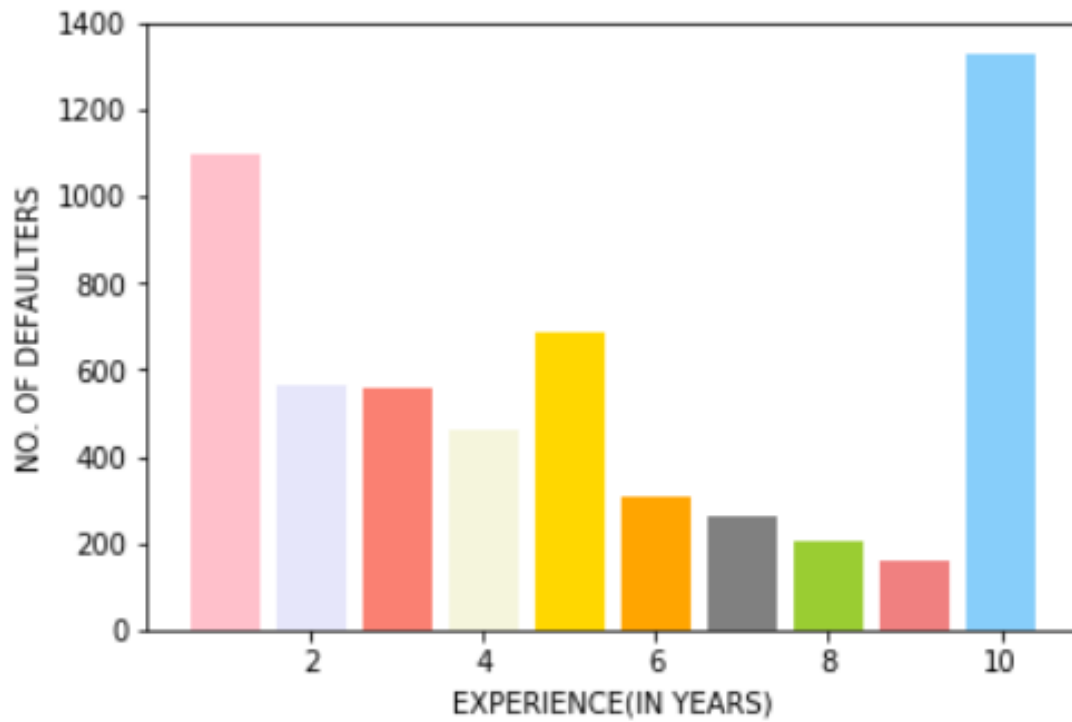
- ✓ The line plot above shows the behaviour of number of defaulters with respect to number of enquiries in last 6 months.
- ✓ We can observe that number of defaulters reduce as we increase the number of enquiries in last 6 months.
- ✓ We can make an outcome that as the number of enquiries in last 6 months increase, the number of defaulter decrease.

8. THE GRADE WITH NUMBER OF DEFAULTERS:



- ✓ The pie chart plotted above clearly shows the relation between the grade and number of defaulters.
- ✓ Customers having grade as 'B', 'C' or 'D' have a greater chance of defaulting.

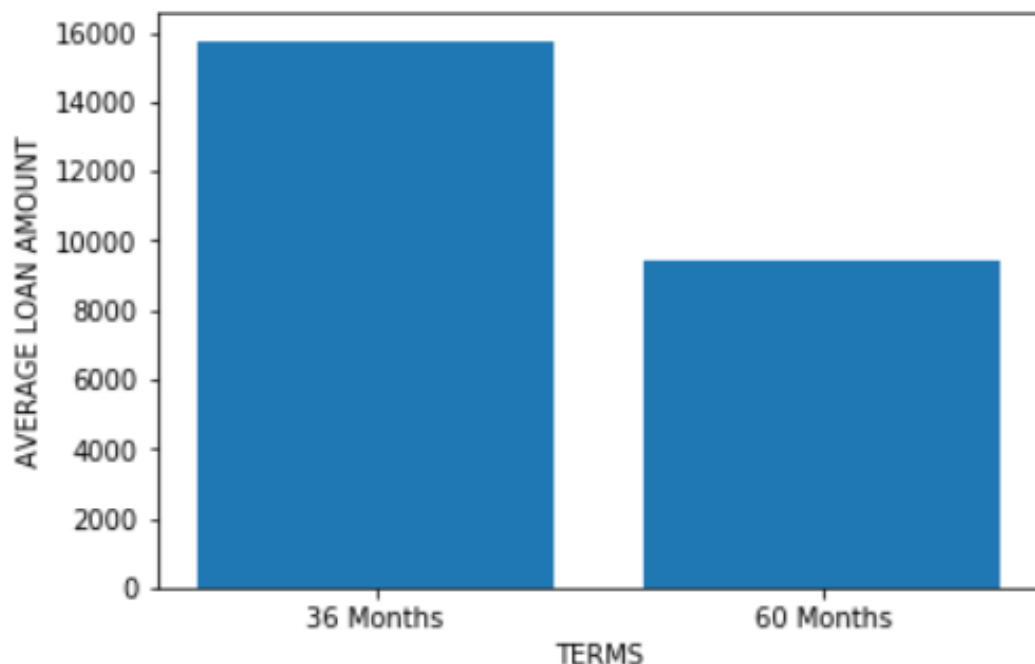
8. RELATION BETWEEN EXPERIENCE (IN YEARS) WITH NO. OF DEFAULTERS:



- ✓ The plot displays the variation in number of defaulters with respect to the number of years of experience.
- ✓ Clearly, the customer having years of experience as 1 or 10 default the most.
- ✓ Hence, the company should check the number of years of experience when checking for defaulters.

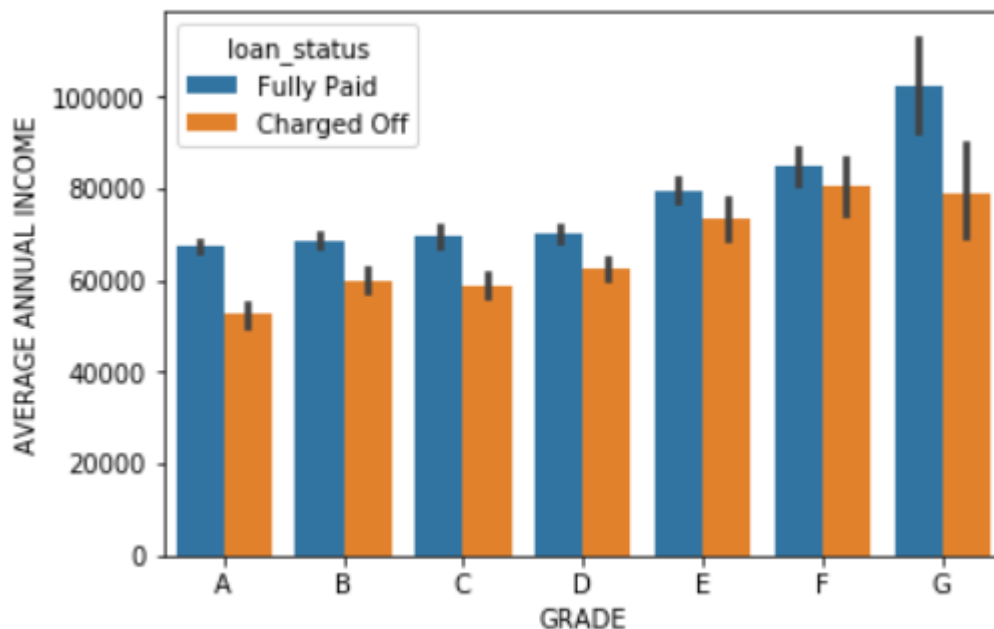
BIVARIATE ANALYSIS:

1. AVERAGE LOAN AMOUNT WITH RESPECT TO THE LOAN TERM FOR DEFAULTERS ONLY:



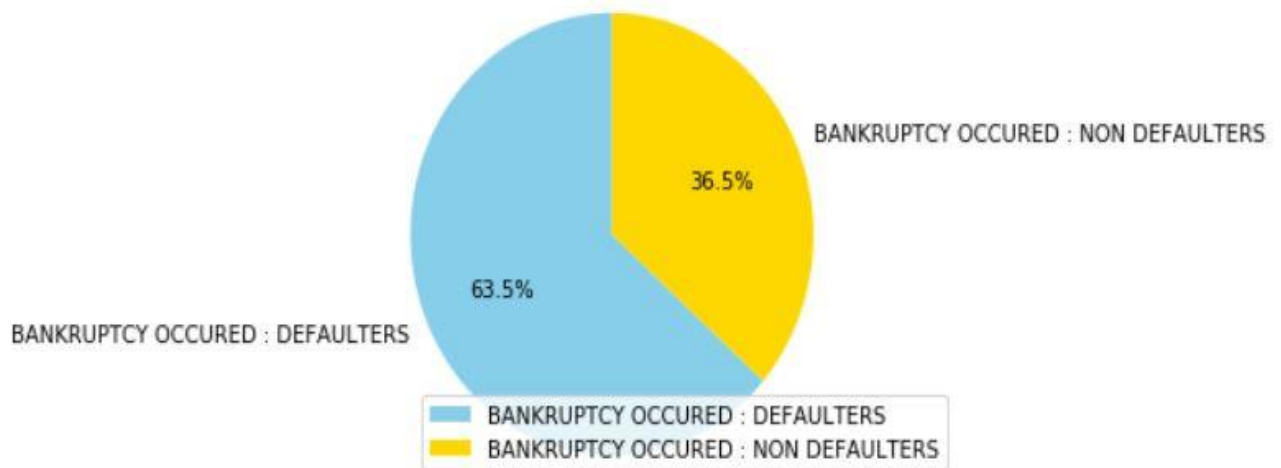
- ✓ The average loan amount was calculated with respect to the term of the defaulters.
- ✓ We can observe that defaulters having term as 36 months have a much higher loan amount than the defaulters having loan amount as 60 months.
- ✓ Hence the company should access the loan amount too before providing loan to the new customer.

2. RELATION BETWEEN ANNUAL INCOME AND GRADE ACCORDING TO THE LOAN STATUS:



- ✓ Plotted a bar graph for average annual income with respect to grade and differentiated the graph with different loan status.
- ✓ By analysing the bar plot, we came to know that the average annual income for the 'Fully Paid' customers is a little higher than that of the 'Charged Off' customers.
- ✓ The company should check the annual income of the customer and the grade as well to evaluate whether the new customer may or may not default.

3. NO. OF BANKRUPTCY OCCURED FOR DEFAULTERS AND NON-DEFAULTERS:



- ✓ The above pie chart shows the occurrence of bankruptcy for both, defaulters as well as non-defaulters.
- ✓ We can summarize that occurrence of bankruptcy in defaulters is way more than its occurrence in non-defaulters.
- ✓ Hence, the company should check the bankruptcy records of the new customers to evaluate the tendency to default or not.

CONCLUSION:

The following conclusion can be made from the report:

- The company should analyse or check the following features in a new loan application such as:
 - Term Duration of the loan.
 - Number of delinquent accounts linked with the customer.
 - Number of open credit lines in the customer's account.
 - The Verification Status of the customer.
 - Purpose of the loan.
 - House Ownership Status of the customer.
 - Number of Enquiries in the last 6 months.
 - Experience of the customer.
 - Loan Amount with respect to the term for new customer.
 - Annual Income with respect to the grade.
 - The occurrence of bankruptcy in new customers account.

The relations and relevance explored in the report if scrutinized and executed properly, will minimize losses and help the company improve on financial terms.