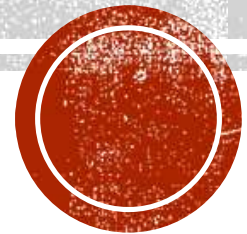


LENDING CLUB

CASE STUDY

Contributor: Sukanya Karri



Problem statement:

- You work for a consumer finance company that specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. The data given contains information about past loan applicants and whether they 'defaulted' or not.

Objective:

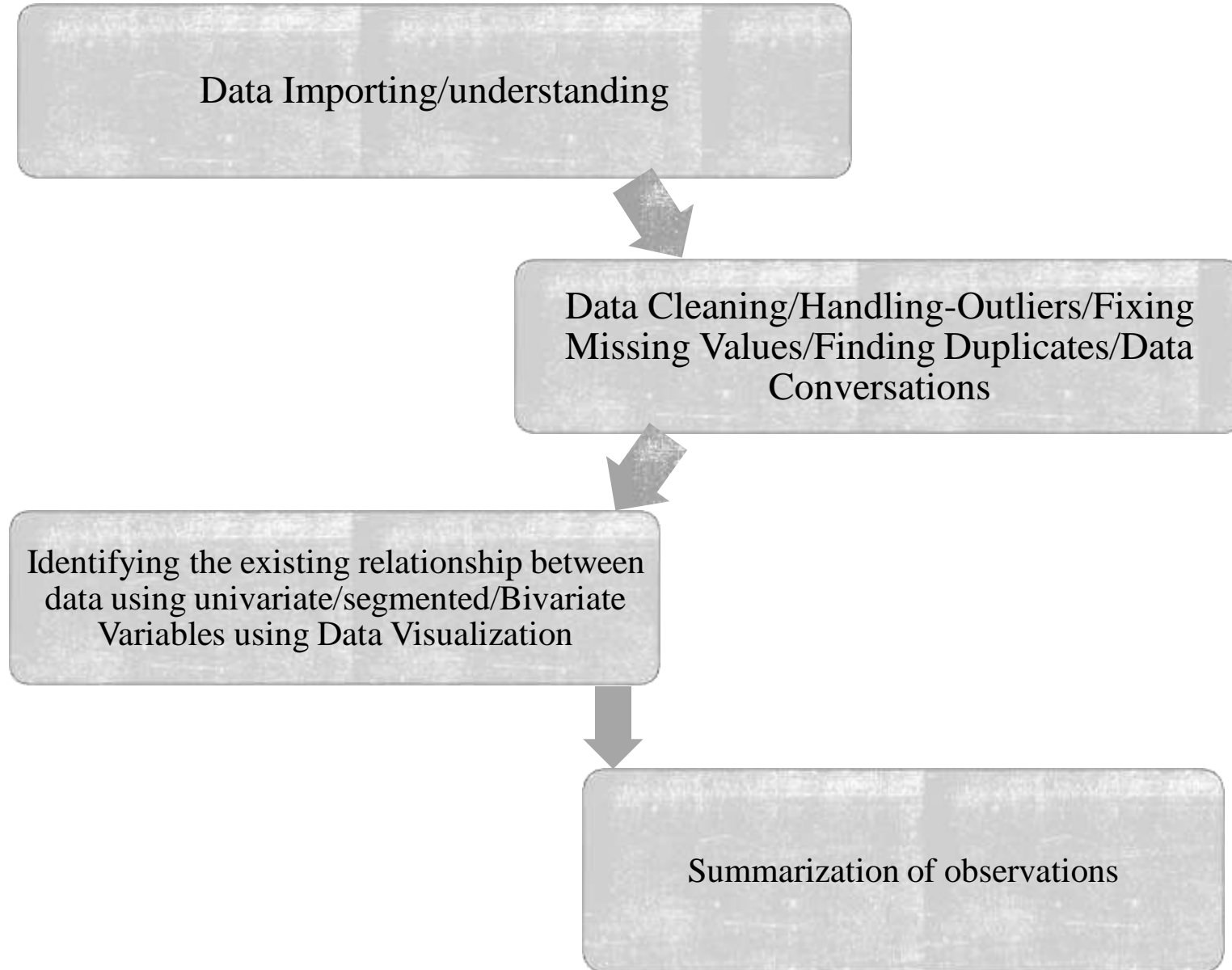
The aim is to understand the driving factors (or driver variables) behind loan default, i.e. the variables that are strong indicators of default, so that the client can utilize this knowledge for its portfolio and risk assessment.

Benefits of the case study:

- Will understand the techniques involved in exploratory data analysis and the steps involved like cleaning, wrangling, and finding missing values.
- Domain knowledge- Based on the data provided helps us to get an insight into its attributes, this EDA case study focuses on loans and defaulters, and risk analytics is a helpful concept to learn here
- Understanding how data performs using various visualization charts



STEPS INVOLVED: DATA ANALYSIS



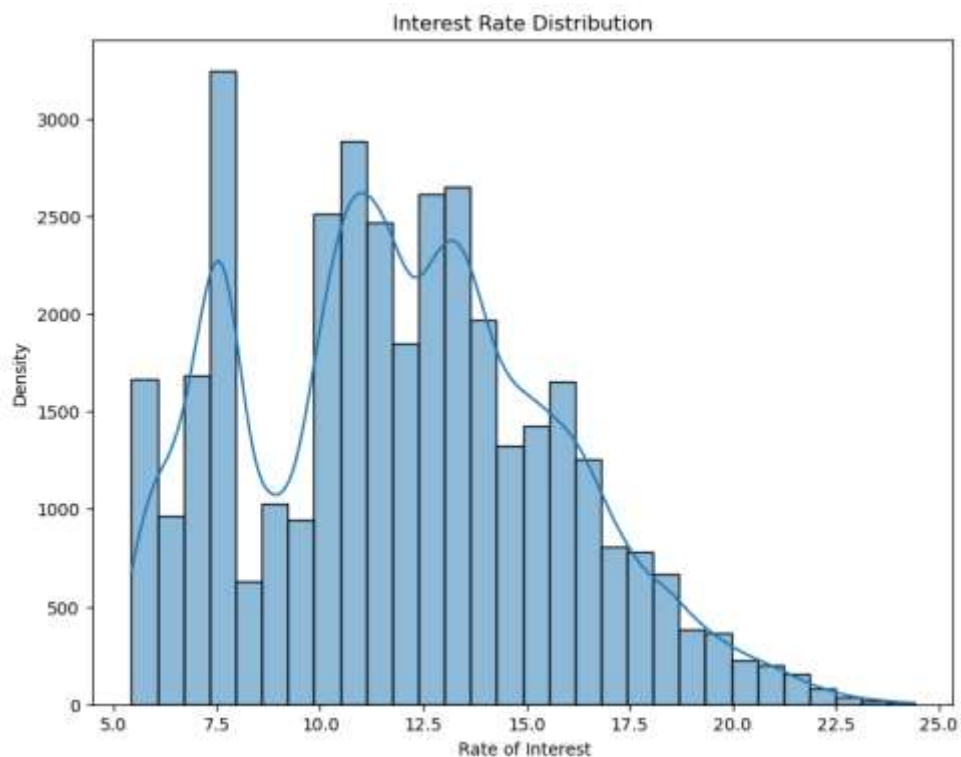
DATA ANALYSIS PROCEDURE IN BRIEF

- ❖ Import the given loan data set
- ❖ Read the file in the CSV format in jupyter notebook
- ❖ Understand the data and the various columns present
- ❖ Search for the duplicated values
- ❖ Identify the missing/Null values
- ❖ Post the search and identification drop the duplicates and fix/remove the missing values
- ❖ Search for the unique values and drop them if found not useful for the business objective
- ❖ Do a little background research and figure out the attributes which will be useful for data analysis
- ❖ Drop the columns that are not required for the analysis based on the project requirement and result expectation. For example: Zip Code, URL, etc..
- ❖ Perform the data conversion and extract the necessary columns say month and year as and when required
- ❖ Check for the empty spaces in categorical objects using the string function and remove them same.
- ❖ Check for the outliers and remove the same
- ❖ Perform the Univariate/Bivariate/Segmented univariate with different attributes using Data visualization techniques to understand the distribution of data and its performance compared to other elements.
- ❖ Summarize the observations

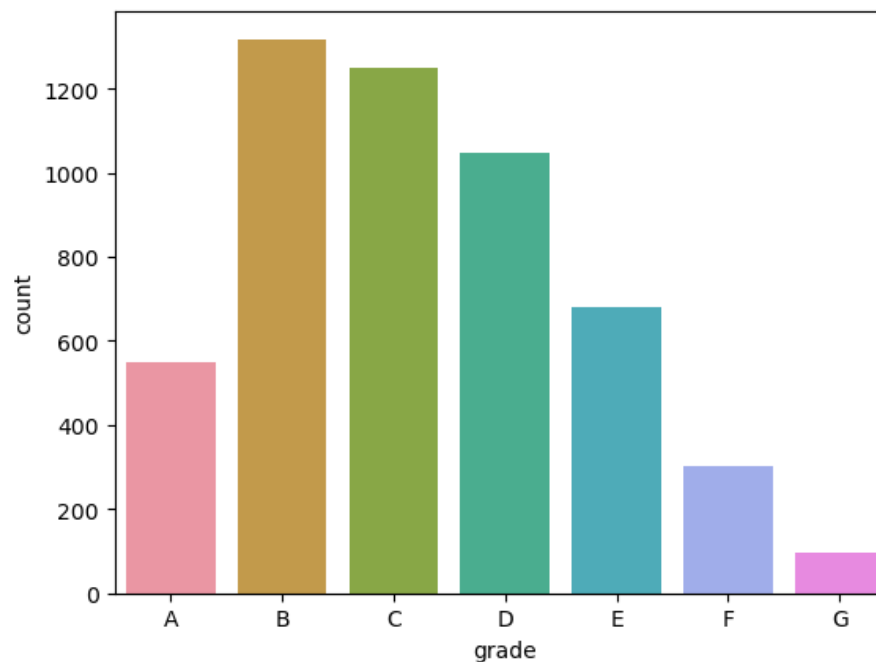
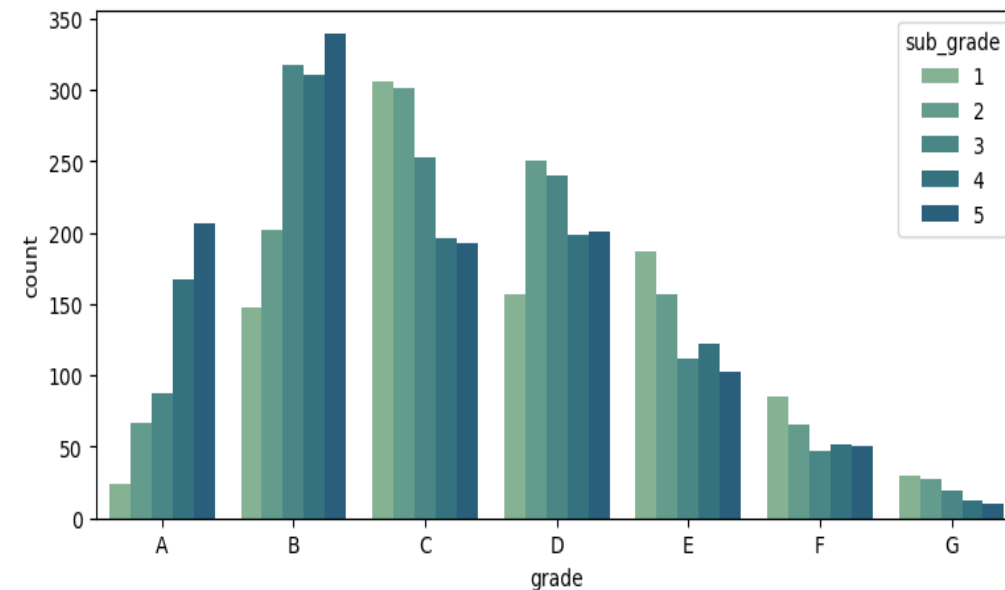


DISTRIBUTION OF ATTRIBUTES AT A GLANCE-1:

GRADE/SUB GRADE, INTEREST RATE



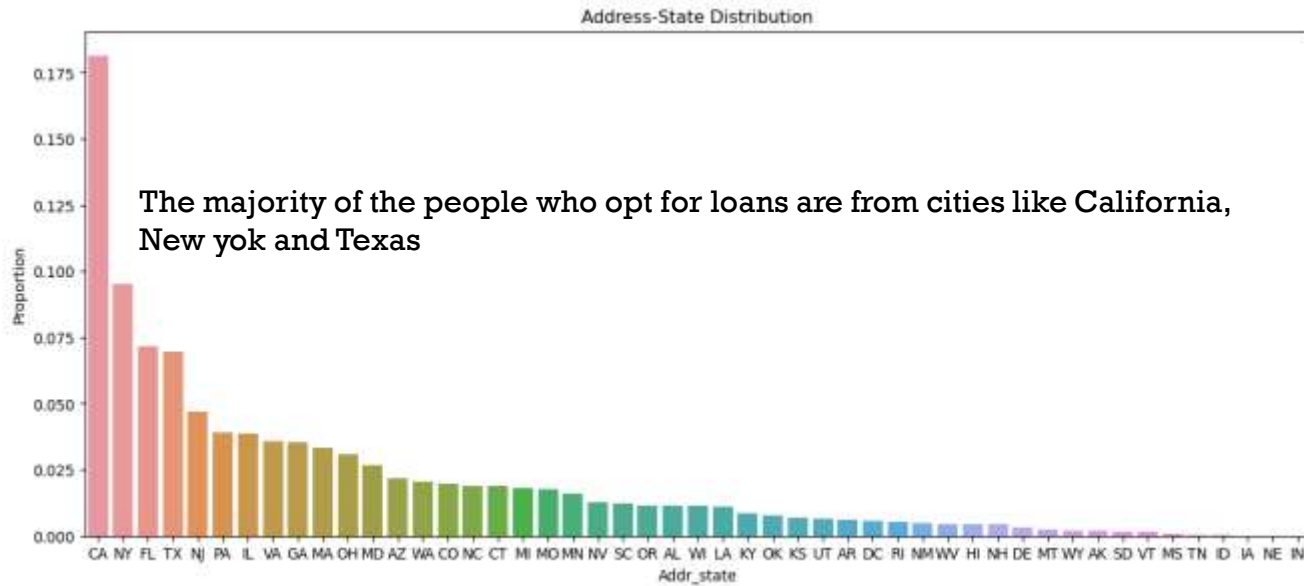
Interest Rate is high at 7.5 and has a sharp decline at 10 and most of the people opting for ROI falls in 10 to 14



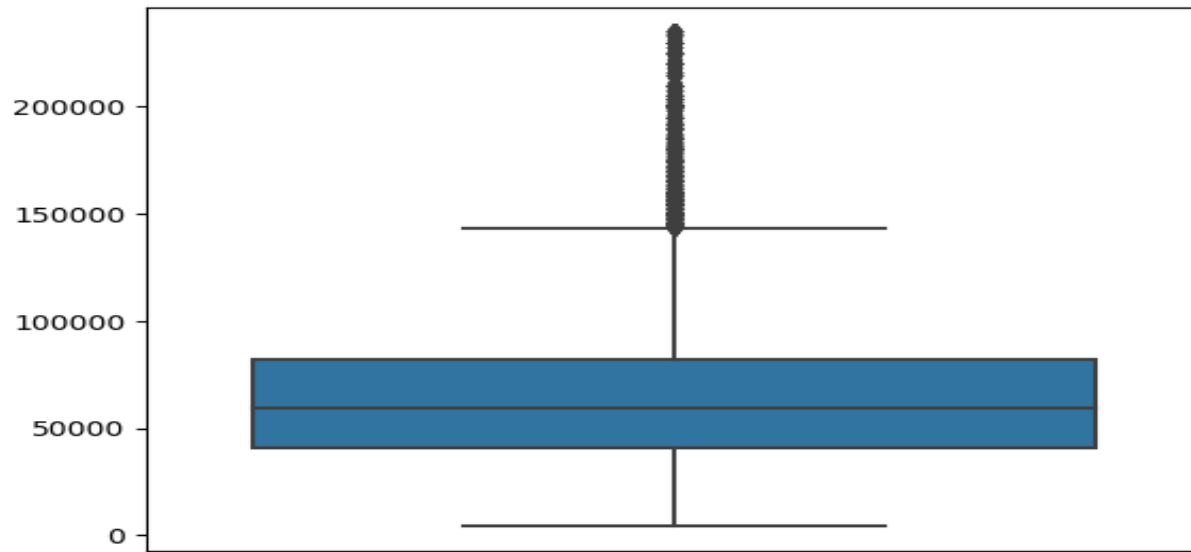
Most borrowers fall in the B and C categories of grades/subgrades



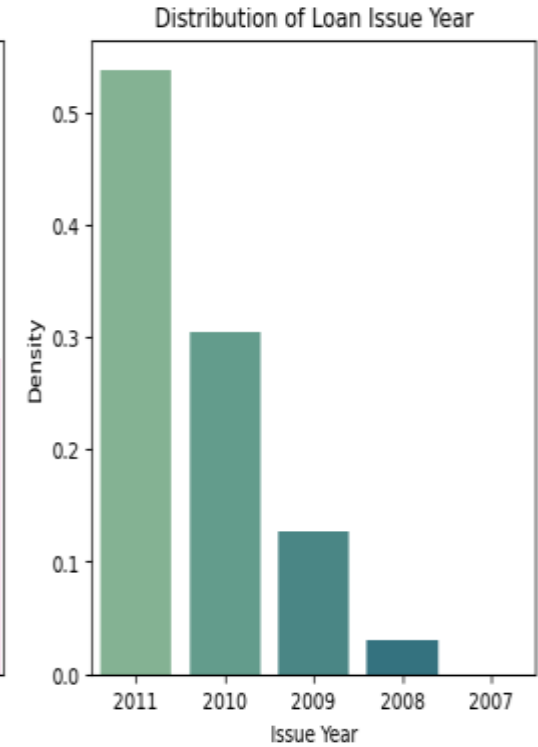
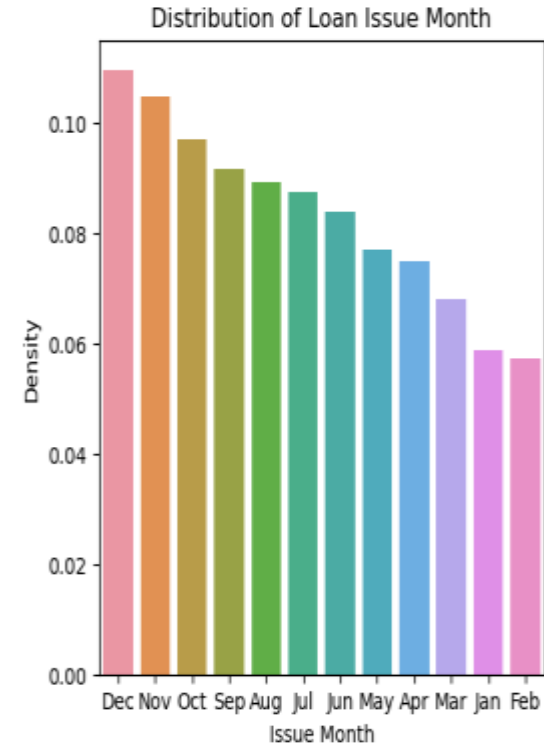
DISTRIBUTION OF ATTRIBUTES AT A GLANCE-2: ADDRESS STATE,ISSUE MONTH/YEAR, ANNUAL INC



Annual income distribution



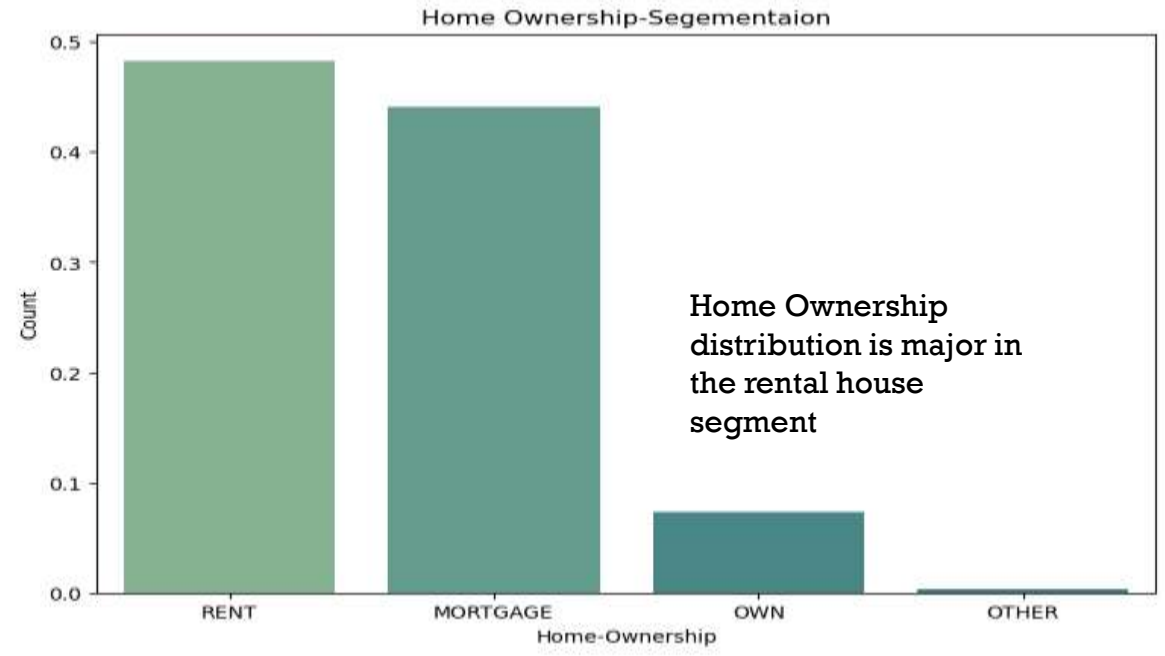
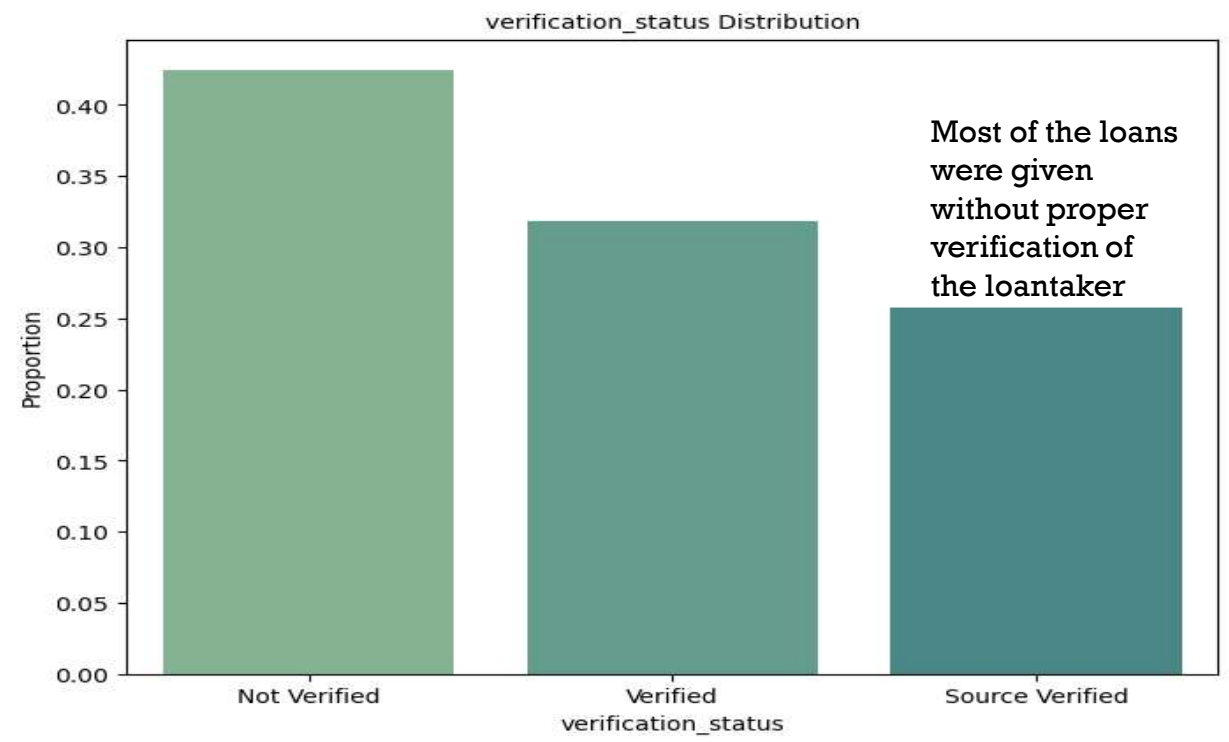
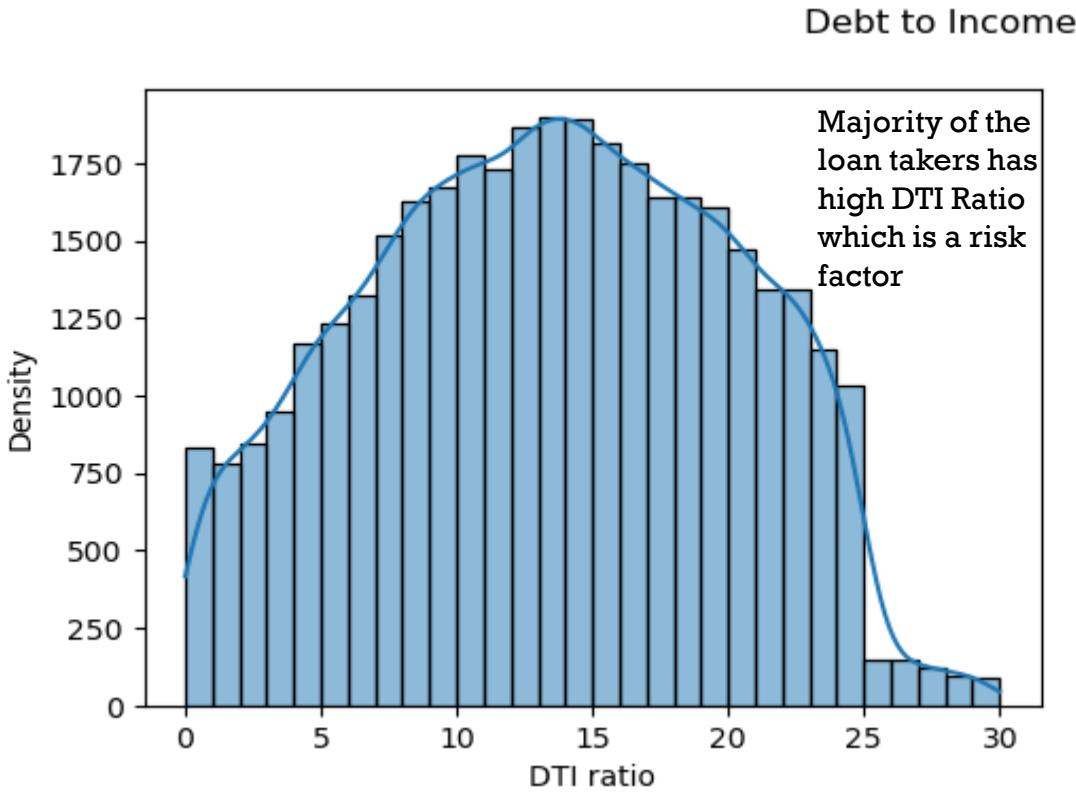
Most of the people opting for loan has low annual income



Most of people took loans in December and in 2011 the highest records of loans for Dec month were given the most and the loans taken by people had a huge rise in 2011 compared to the previous years



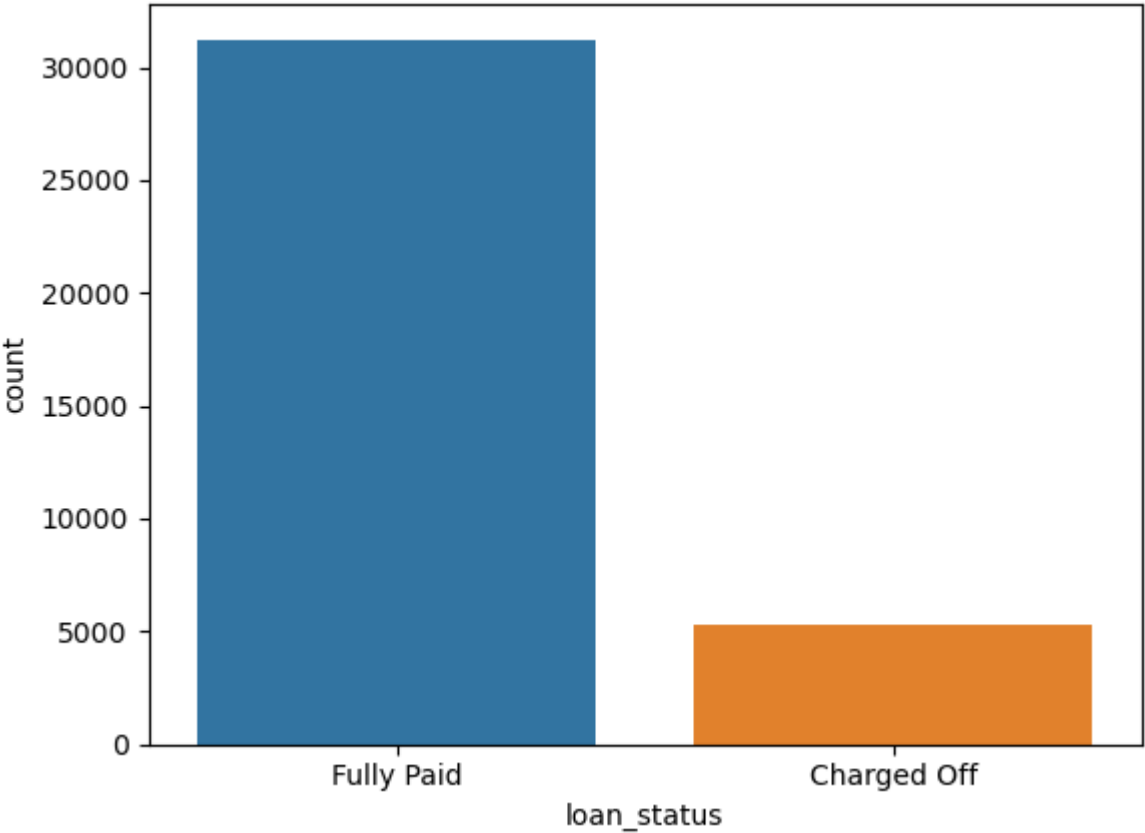
DISTRIBUTION OF ATTRIBUTES AT A GLANCE-3: DTI, HOME OWNERSHIP & VERIFICATION STATUS



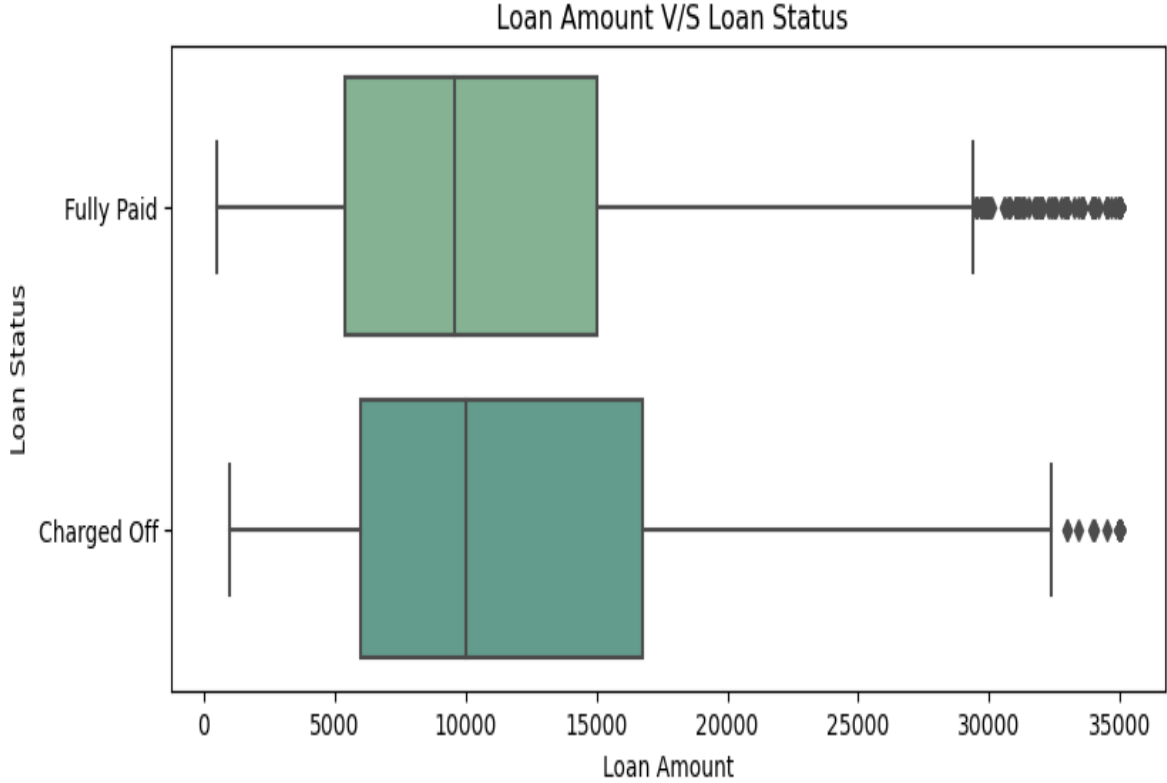
COMPARISONS (V/S)



LOAN STATUS VS LOAN AMOUNT:



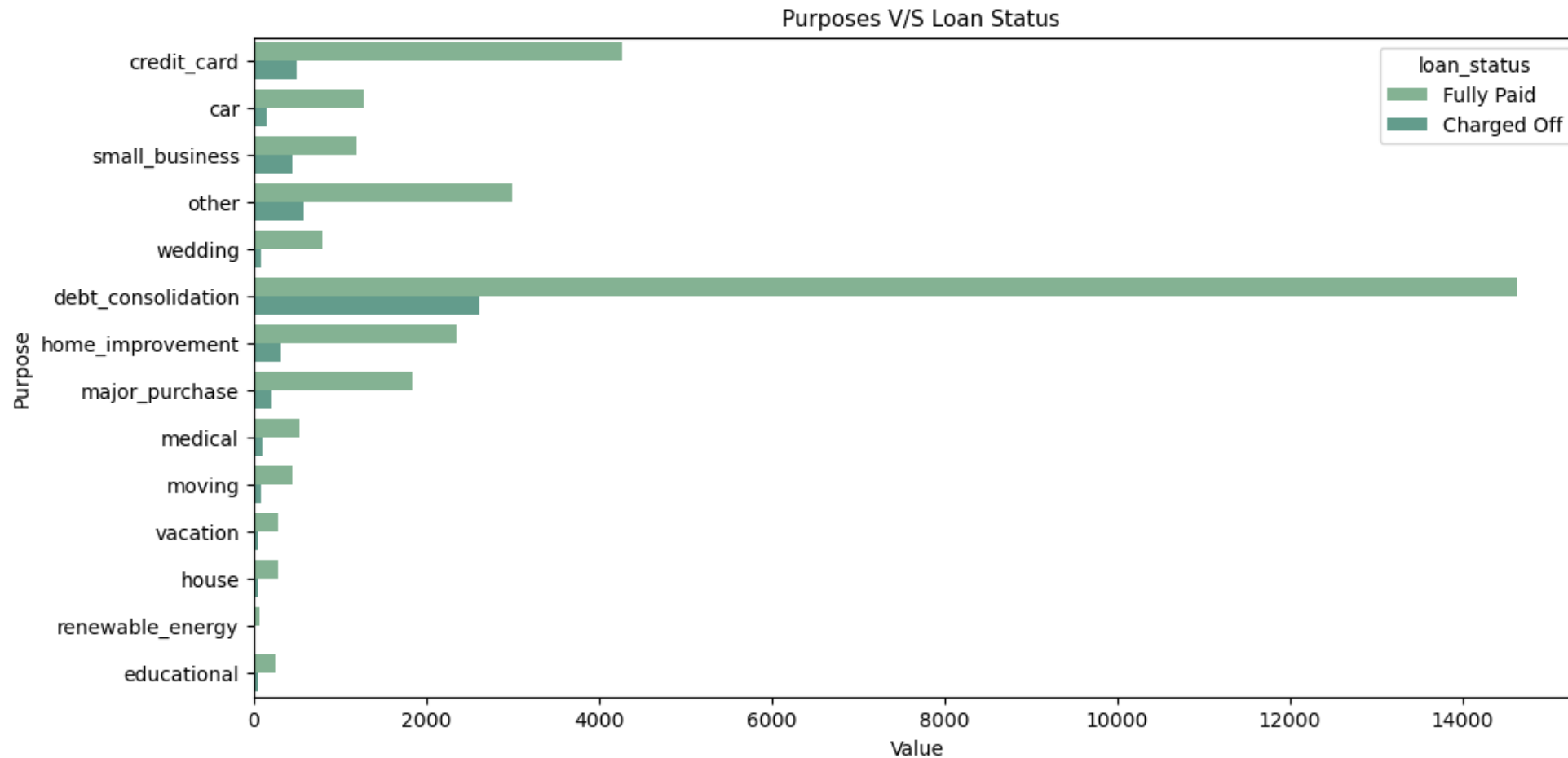
Value of Charged off loans is less compared to Fully Paid loans



Median and 1st quartile are almost the same but the majority of the people falling in the 3rd quartile which is 15k+ seem to be the largest defaulters



PURPOSE: DATA CONSOLIDATION

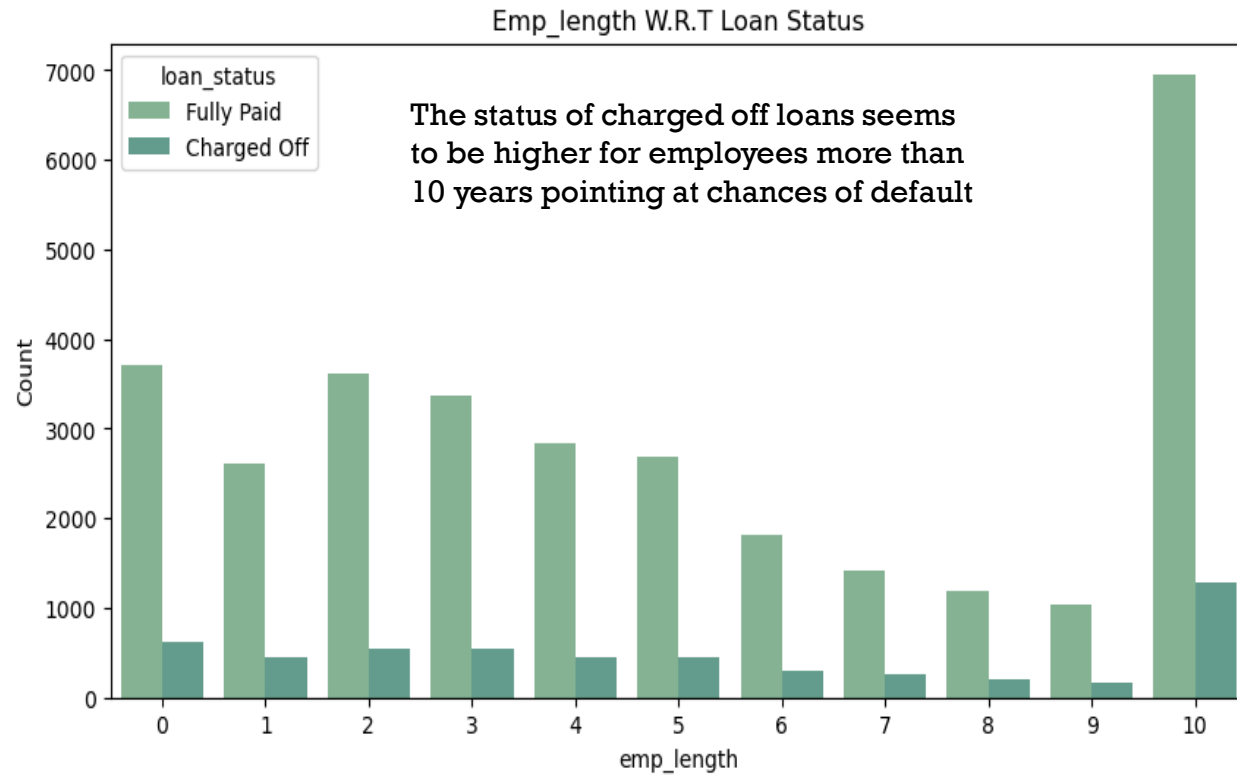


Debt consolidation stands as the major purpose for loan defaulters

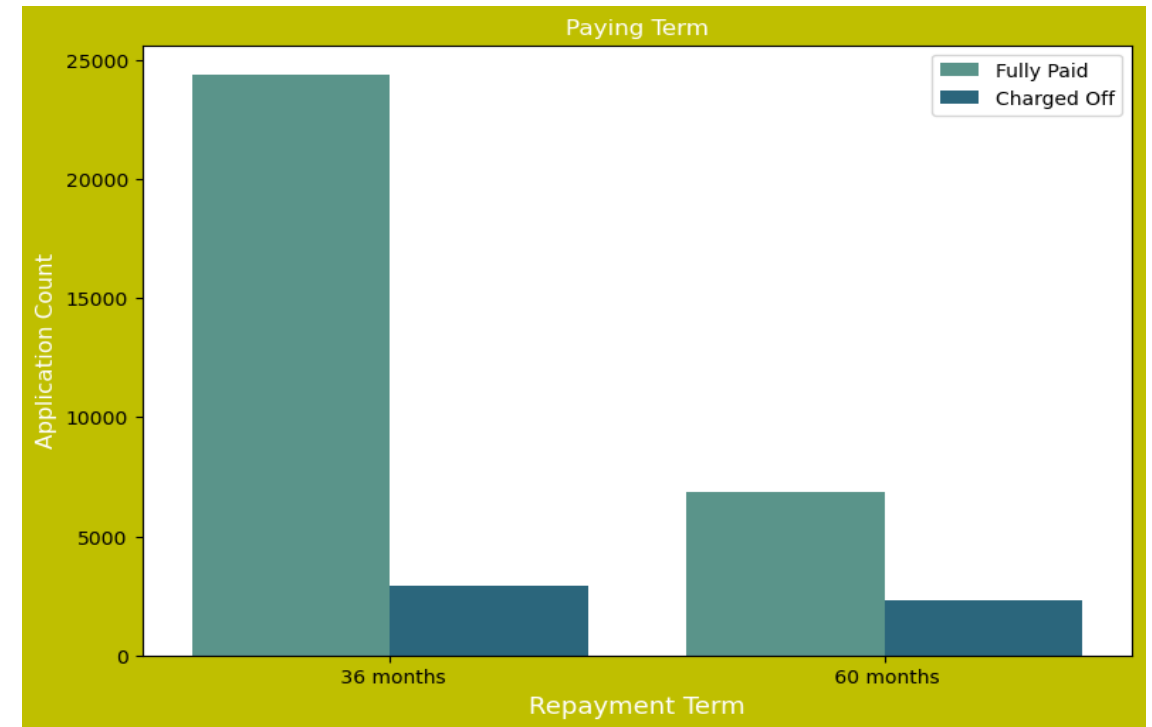


LOAN STATUS

Employee Length



Term

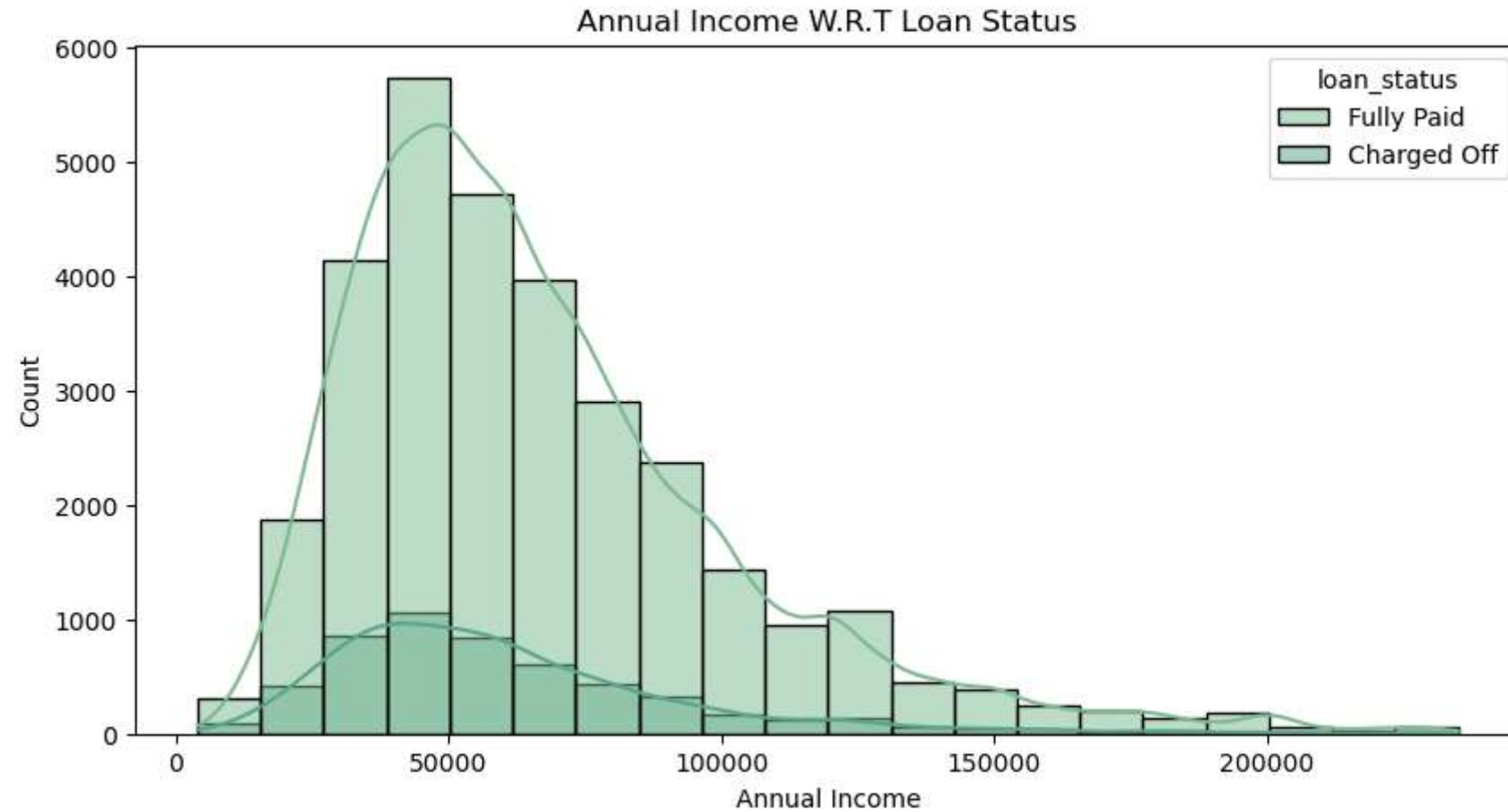


36 Months Charged off percentage is high compared to 60 Months with respect to fully paid loans



LOAN STATUS

Annual Income

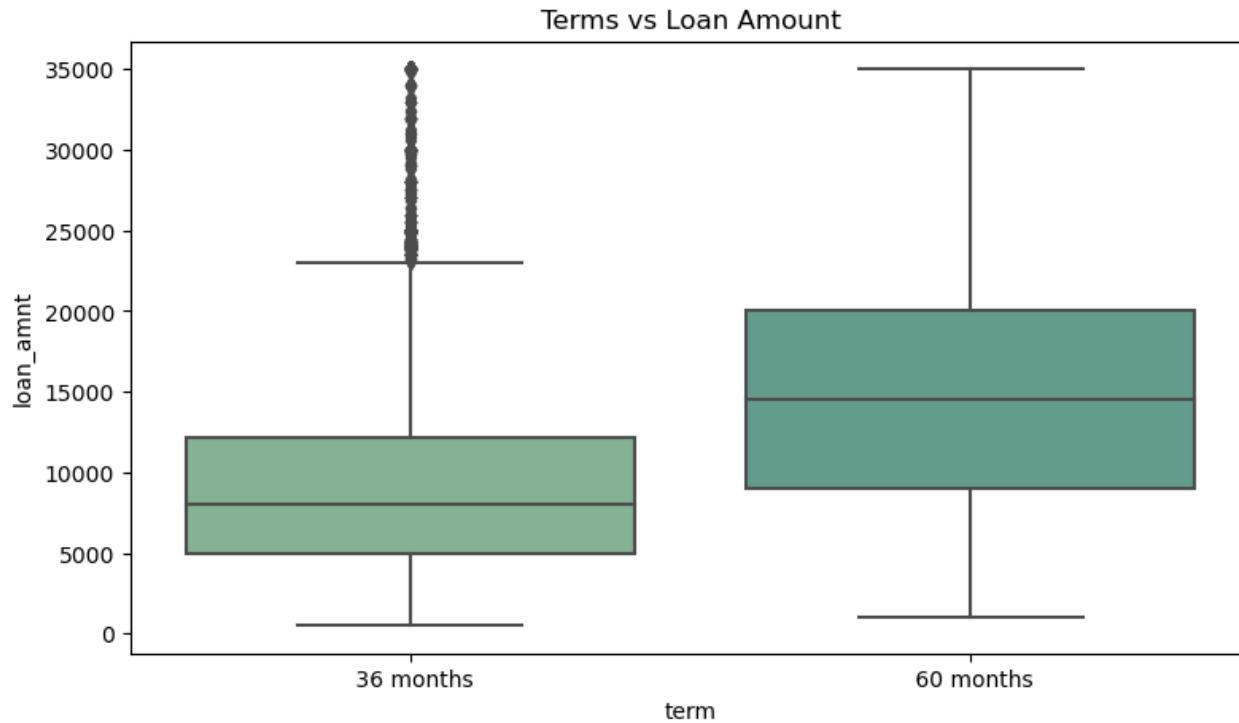


Majority of the people who may default fall in the range of 50000 to 70000

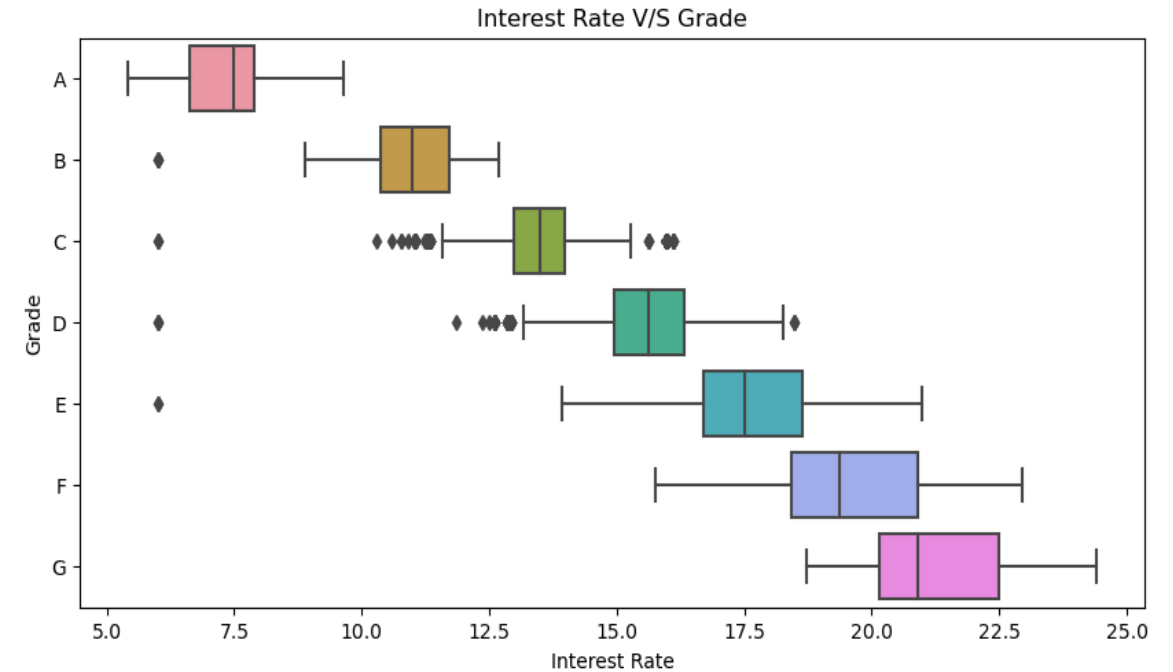


LOAN AMOUNT

Term



Grade



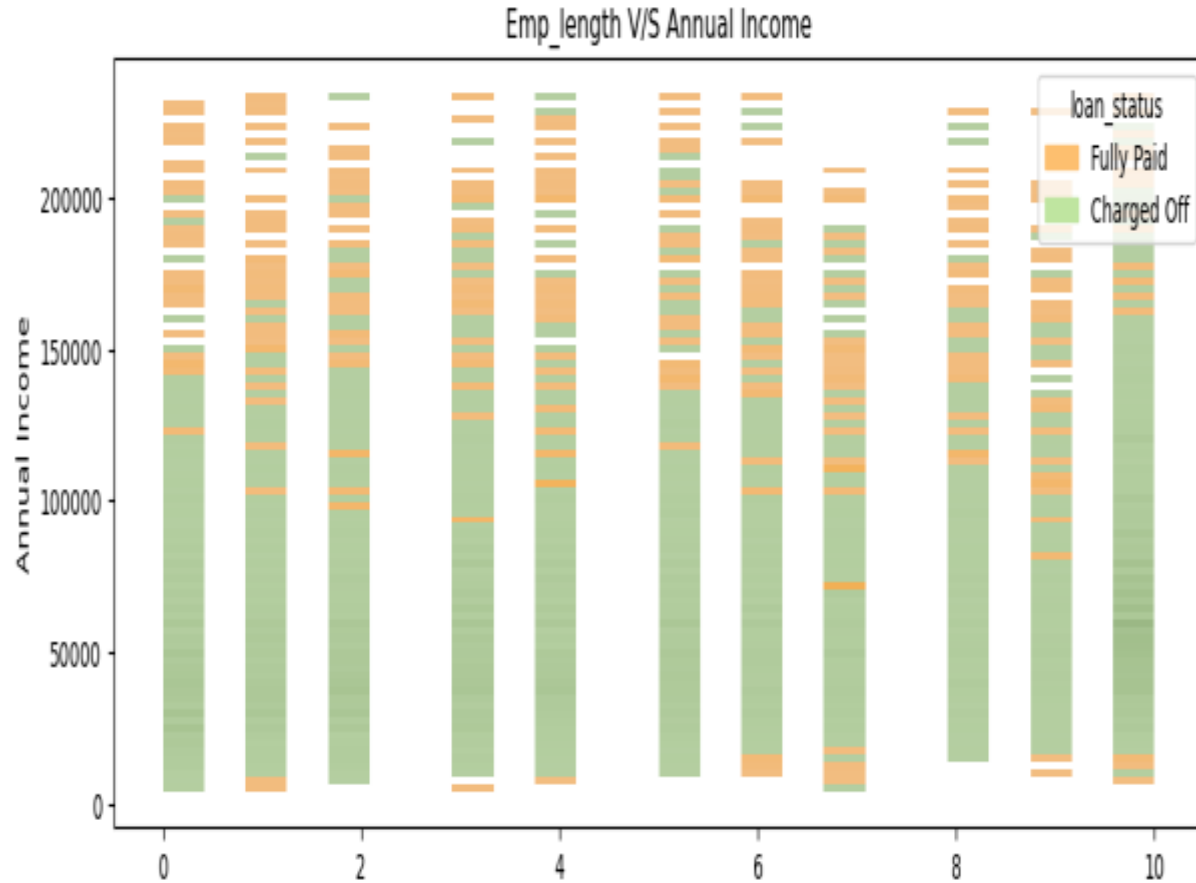
Plot 1: Defaulters seem to be more for the 3rd quartile of 36 months and 1st quartile of 60 months and the loan amount that can defaulted in both scenarios falls in the range of 10k-16k

Plot 2: The higher the rate of interest the lesser the grade



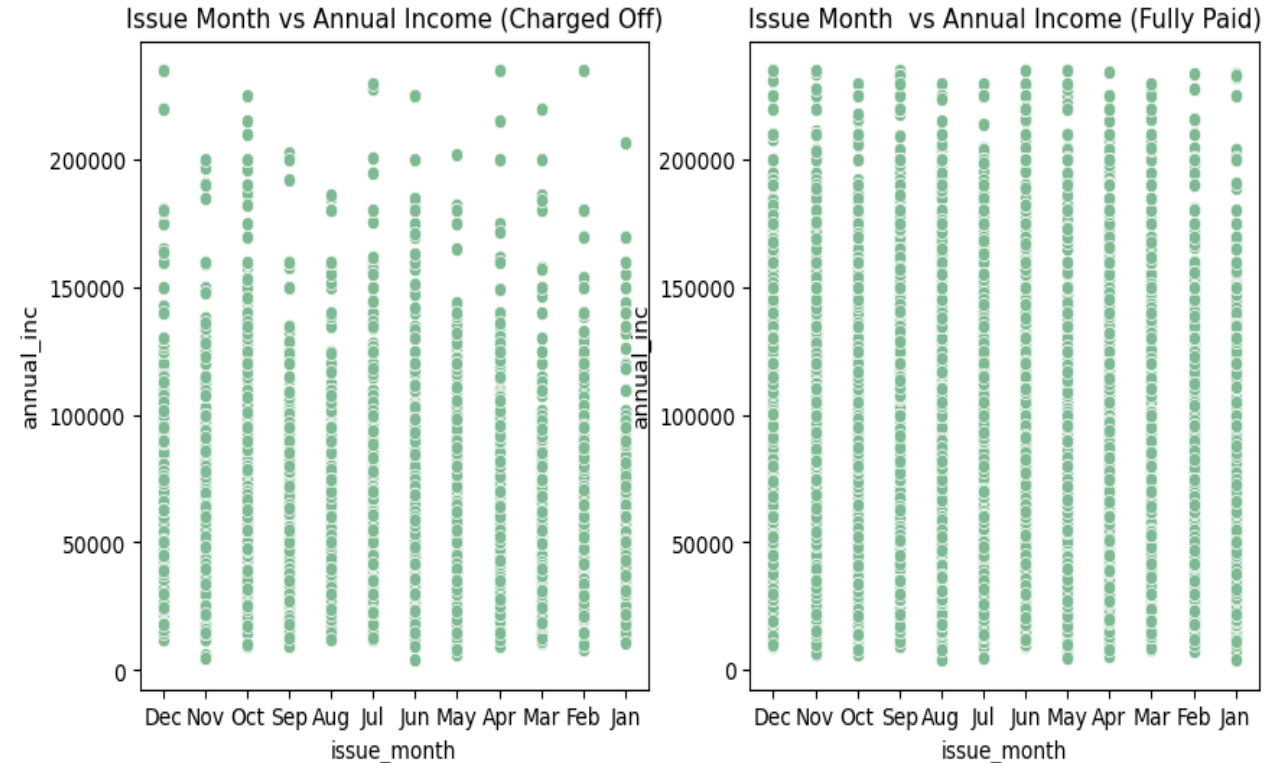
ANNUAL INCOME

Employee Length



There are chances of more defaulters in the 10+ year range, however when considering the annual income range of defaulters is more in the range of 10k to 15k irrespective of length of employment.

Issue Month

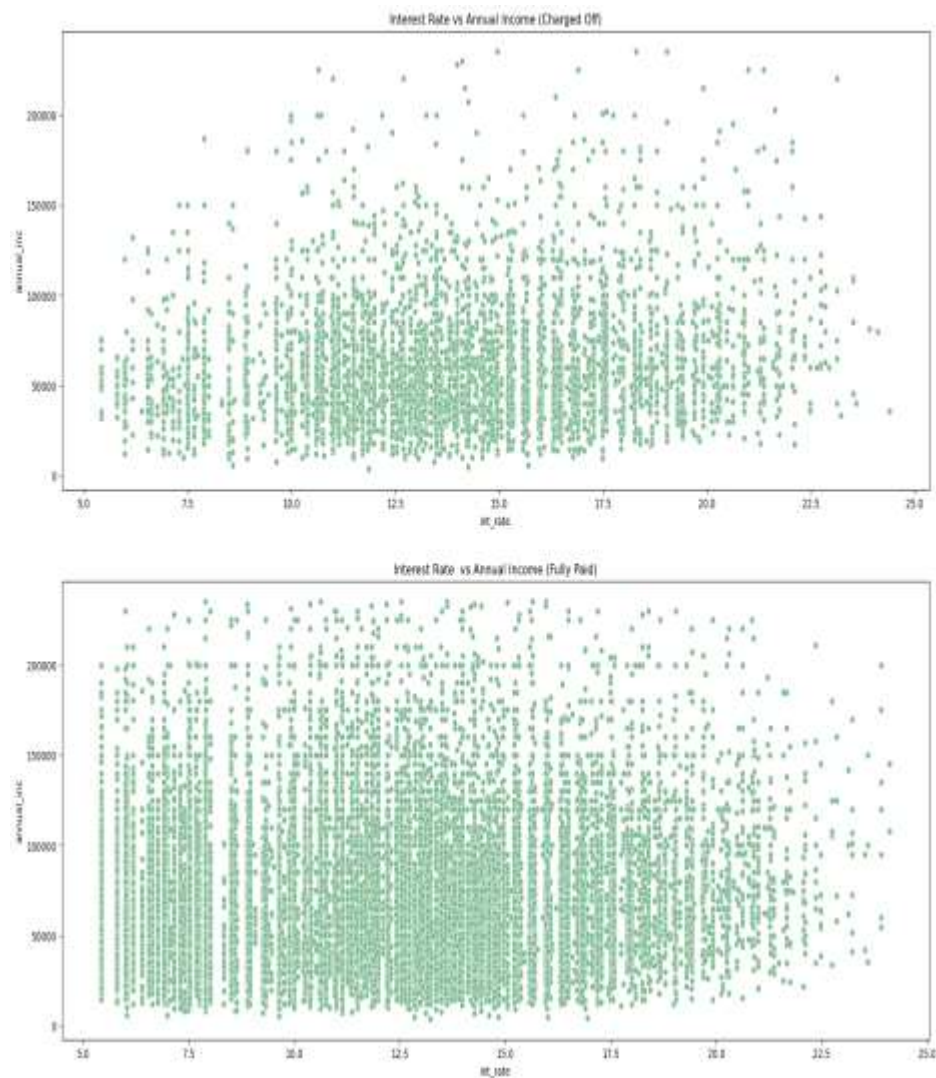


October month irrespective of annual income ranges indicates a large chances defaulting

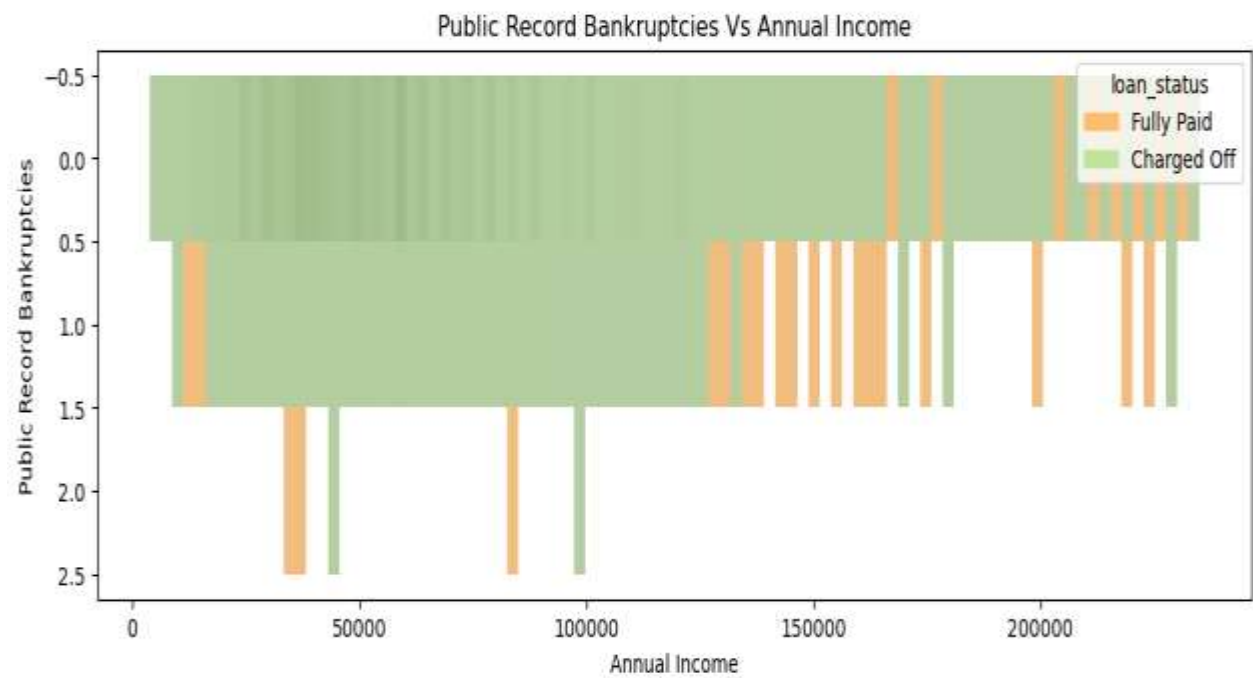


ANNUAL INCOME

Interest Rate



Public record Bankruptcies



Bankruptcies seem to be higher for charged-off loan status inclining more in the income range of 10-13k

From the charged-off plot we can say that maximum defaulters fall in the annual income range <10k with an interest rate opted for 11 to 16 percent Not much info can me made from fully paid



Summary:

Observations:

Major driving factors (or driver variables) behind loan default from the observations made are as follows:

- ❖ **Employment Length** – observed that consumers who have more than 10 years of experience in loans are charged off most.
- ❖ **Grades/Sub-Grades-** The analysis tells that increasing rates increase the default rate and persons with lower grades seem to default most.
- ❖ **Verification Status-**Most of the loans were provided without verification
- ❖ **Annual income-** Range of 10k-16k Income has high chances of defaulters
- ❖ **Term-** Preference of loan term plays a role in loan defaulting
- ❖ **Purpose-** From observation, it was found that purpose plays an important role in loan defaulting

