

Lending Club Case Study

DR RANJEET SINGH MAHLA
MSC IN AI AND ML
CASE STUDY I

SUBMITTER: RANJEET SINGH MAHLA
CONTRIBUTOR: RANJEET SINGH MAHLA

OBJECTIVES

Using exploratory data analysis techniques and understand the loan data sets and present a business model for avoiding loan defaults.

- (1) Implication of exploratory data analysis on real world scenario (here the example is loan data) and provide a solution to the client.
- (2) To emphasize on data driven decision making process for avoiding bad loans.
- (3) To present business centric recommendation for lending loans to clients while avoiding bad loans.

Problem Statement

PROBLEM:

The problem is to infer business driven decision making from the existing loan data set and present a recommendation for lending money to clients while avoiding bad loans.

DATASET

The loan data set contains information about past loan applicant's whether they defaulted the loans or paid on time and if paid on time and defaulted what the variates which might have driven the loan outcomes to a bad loan or a good loan.

APPRAOCH

The approach is understanding the data sets, data cleaning, data wrangling, data visualization

Univariate

Bivariate

Corelation

And segmented univariate and sub plotting

EDA visualization guided inference can be used for business centric decision making

Summary

KEY MESSAGE

WHEN GIVING LOAN CONSIDER FOLLOWING KEY POINTS TO AVOID DEFAULTS

(1) GRADES (2) VERIFICATION STATUS (3) ANNUAL INCOMES (4) PUBLIC RECORD BANKRUPTCIES

SHORT SUMMARY

MORE DERAULTERS WHEN

(1) WHEN HOUSE OWNERSHIP IS RENT (2) WHEN LOAN FOR CLEARING OTHER DEBT (3) WHEN APPLICANT RECEIVING INTEREST AT 13-17% (4) WHEN INCOME RANGE 13201-58402 (5) WHEN APPLICANT HAVE 20-37 OPEN ACCOUNTS (6) WHEN EMPLOYMENT LENGTH 10 (7) WHEN LOAN AMOUNT RANGING FROM 5429-10357 (8) WHEN MONTHLY INSTALLMENT BETWEEN 145-274 (9) FUNDED AMOUNT BY INVESTOR IS BETWEEN 5,000-10,000 (10) WHEN LOAN TERM IS 36 MONTHS OR SO (11) WHEN LOAN STATUS IS NOT VERIFIED (12) WHEN NUMBER OF INQUIRES IN LAST SIX MONTH IS ZERO (13) WHEN DEROGATORY RECORDS ARE 0 (14) WHEN GRADE IS GIVEN AS B (15) WHEN TOTAL GRADE IS OF B5 AND ABOVE LEVEL

UNIVARIATE BIVARIATE ANALYSIS

TAKE AWAY FROM UNIVARIATE ANALYSIS

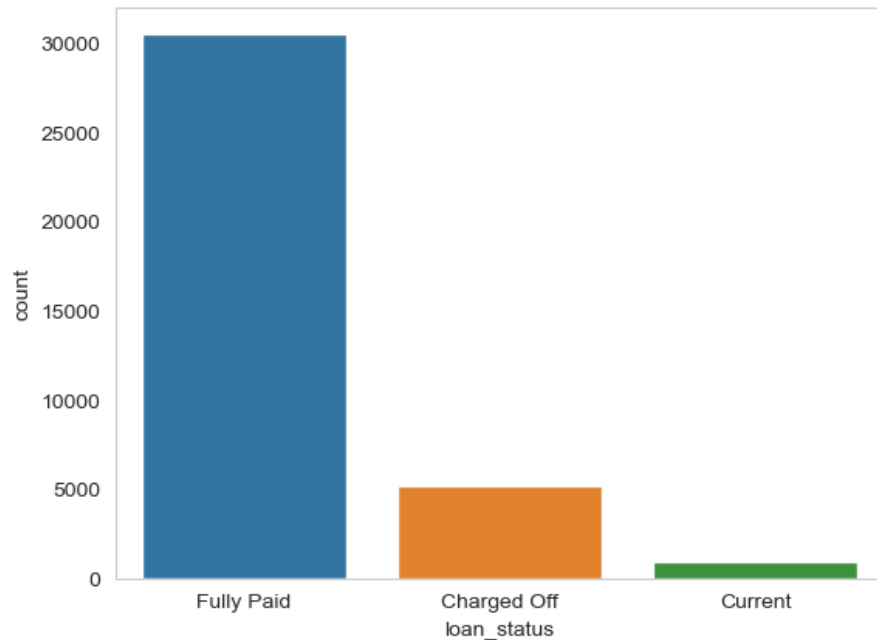
(1) IN TERM OF DISPERON MAJORITY OF LOANS ARE GIVEN FOR 36 MONTHS THAN THOSE OF 60 MONTHS WHERE MAXIMUM DEFAULTERS EXISTS (2) MAJORITY OF LOANS ARE HIGH GRADE LOANS ARE ARE FALLING IN THE TIME FRAME OF 36 MONTHS THAN LOWER GRADE LOANS FALLING IN 60 MONTHS (3) MOST PEOPLE WHO ARE DEFAULTERS ARE THOSE WHO LIVE ON RENT RATHER THAN HAVING OWN HOUSE (4) INCOME HAS A LARGE IMPACT ON DEFAULTING, PEOPLE WHO DEFAULTED HAVE LOW INCOME (5) PAYING OTHER DEBS IS THE MAJOR PURPOSE OF TAKING LOAN AND THAT IS WHERE MORE LOAN HAS THE HIGH PROBABILITY OF DEFAULTING.

TAKE AWAY FROM BIVARIATE ANALYSIS

(1) GRADE OF LOAN IS A POTENTIAL RISK FACTOR; INTERST RATE INCREASE WHEN LOAN GRADES IS INCREAING FROM A TO F OR SO. (2) OVERALL LOAN APPLICANT HAVE SPARSE RECORD OF BANK RUPTCY AND CONSIDERED AS GOOD LOAN APPLICANT EXCEPT THOSE WHO ARE DEFAULTERS WITH LOW GRADE.

FROM CORRELATION HEATMAP IT IS CLEAR THAT ONLY FEW FACTORS ARE CRITICAL (1) LOAN AMOUNT (3) INTEREST RATE (4) ANNUAL INCOME

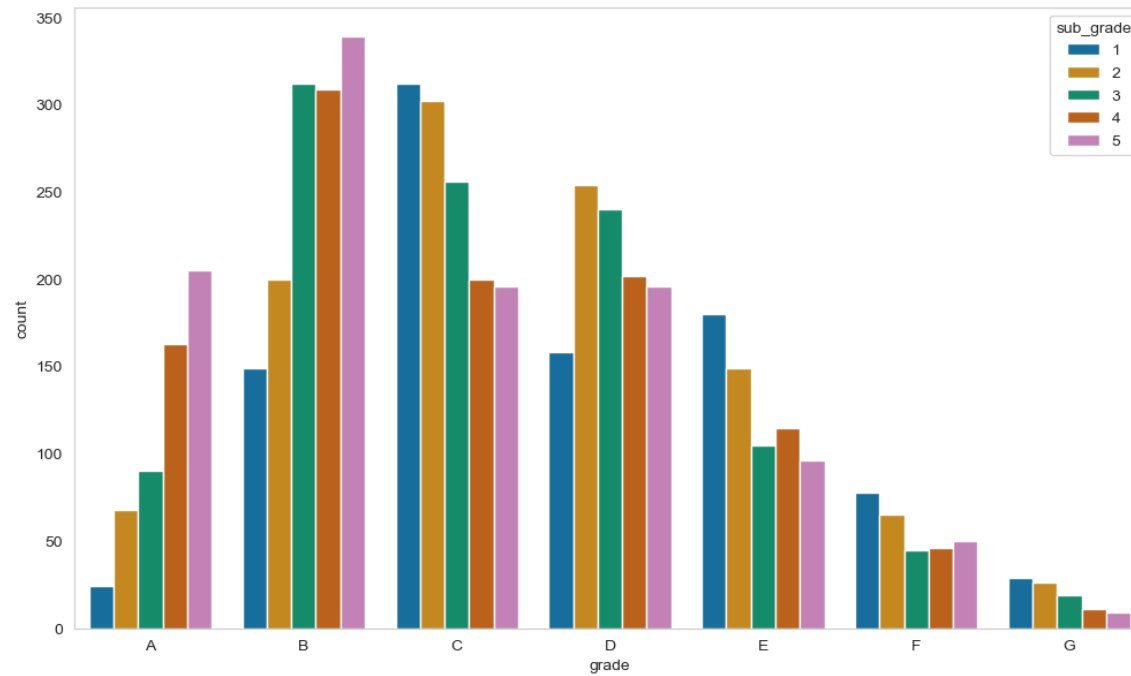
LOAN STATUS



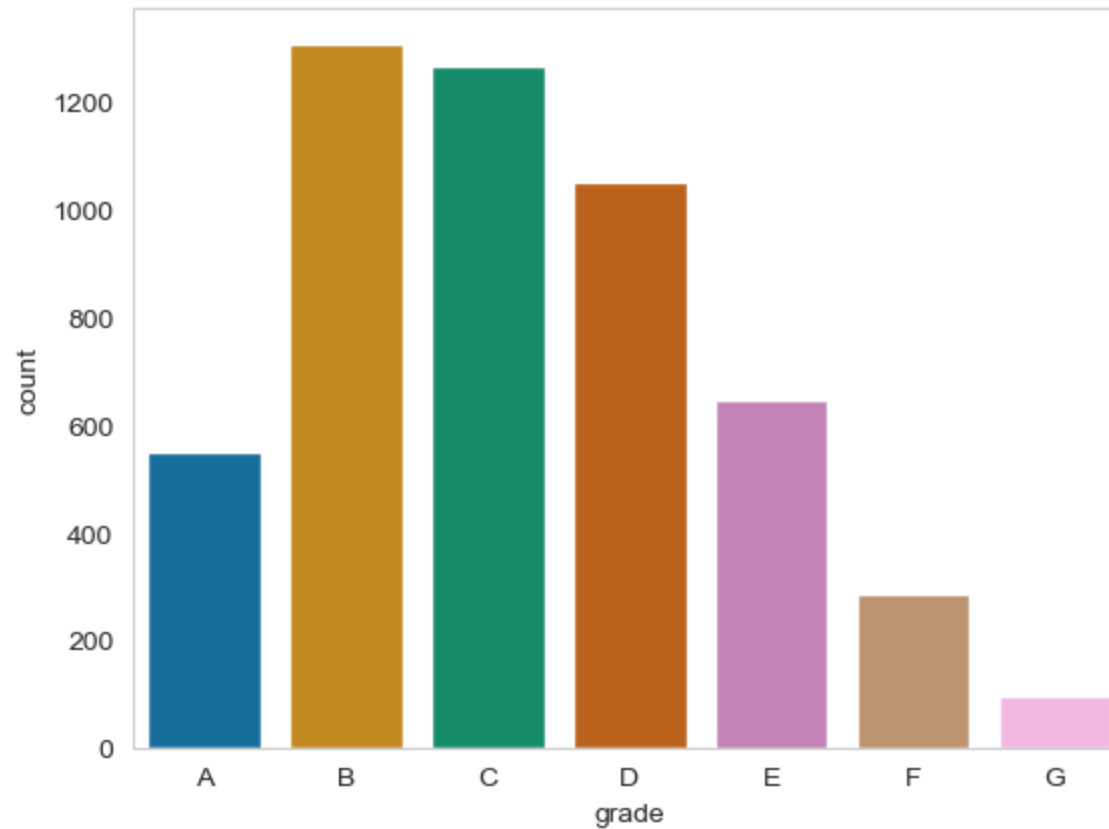
THERE ARE MORE FULLY PAID LOAN THEN DEFAULT AND CURRENT LOAN

A MINOR FRACTION OF LOAN ARE DEFAULT AND WITH EDA WE CAN PREDICT THE MEASURES OF DEFAULT

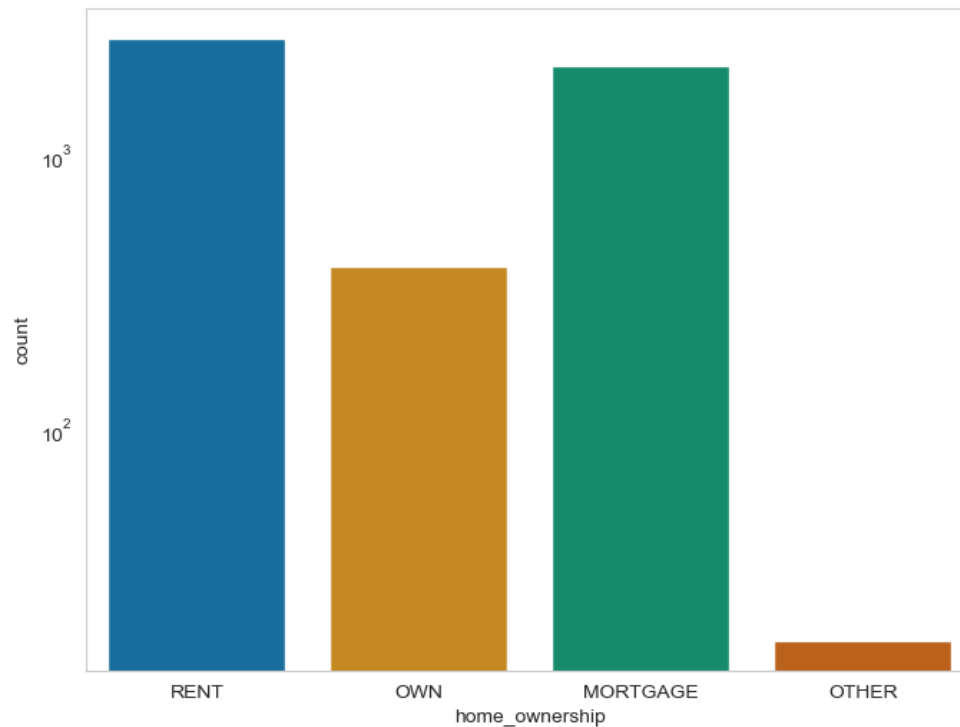
GRADE WIDE LOAN



CHARGED OFF LOANS AND GRADES

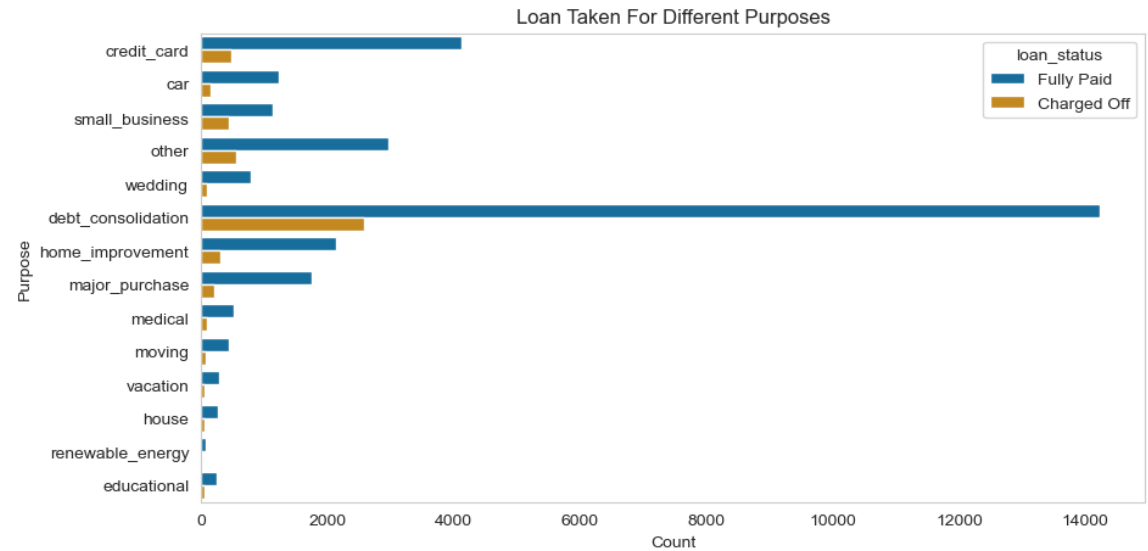
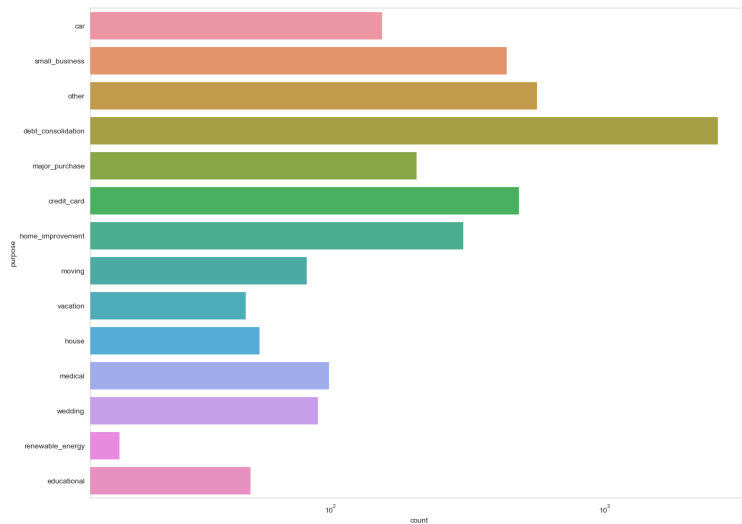


HOME OWNERSHIP AND LOAN



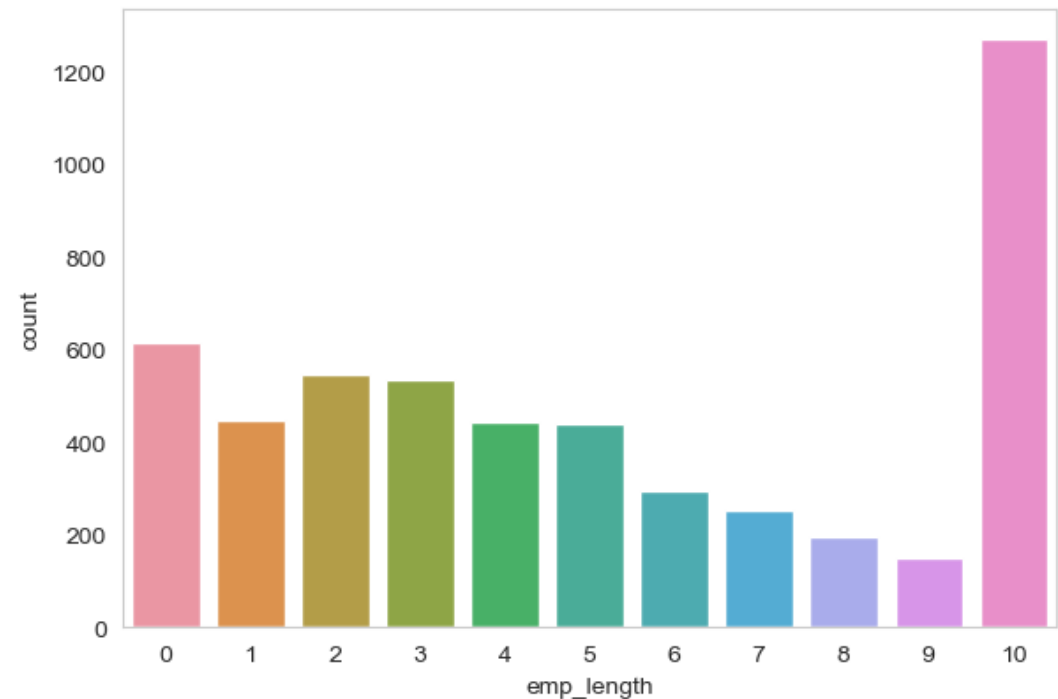
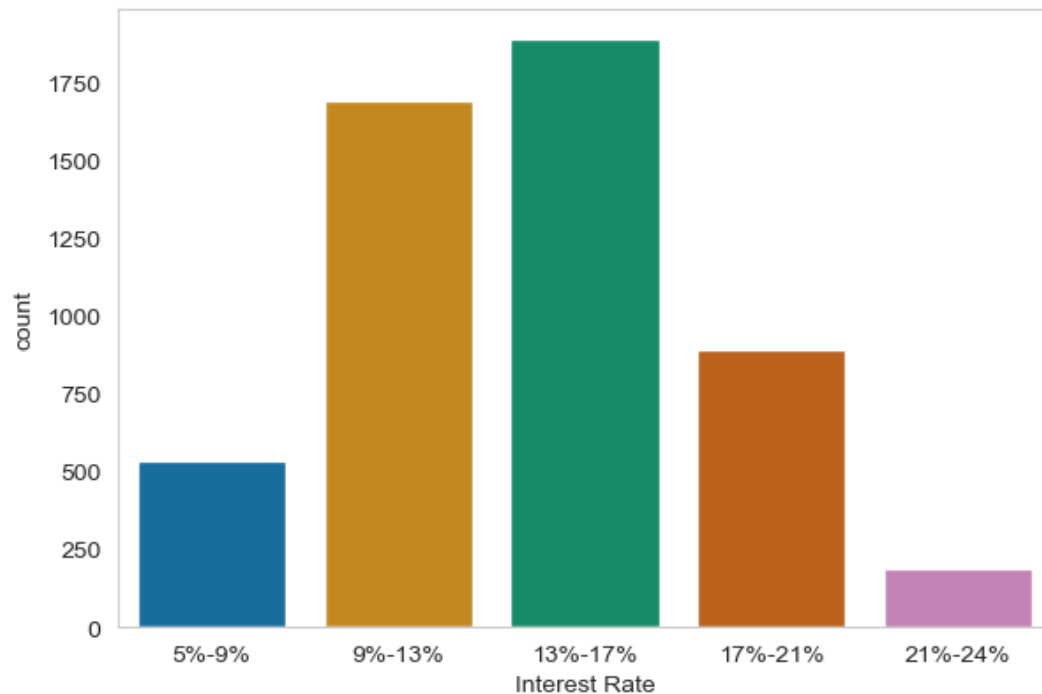
PEOPLE LIVING IN RENTED HOUSE ARE MORE LIKELY TO DEFAULT THE LOAN

DEBT COLSOLIDATION



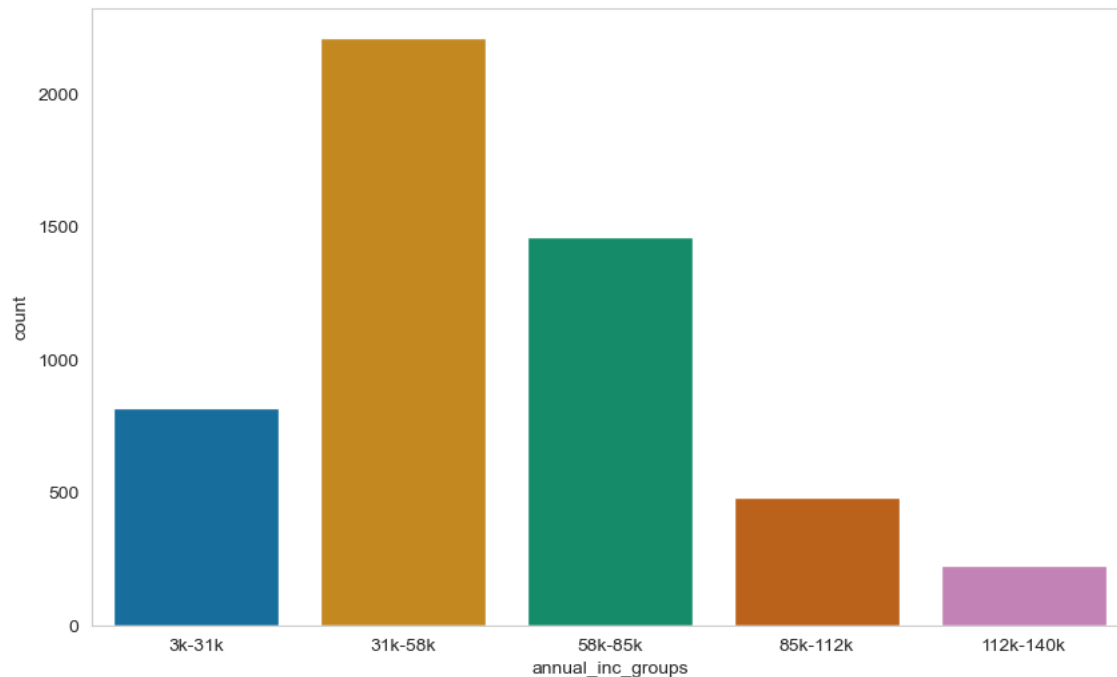
PEOPLE WHO ARE TAKING LOAN TO PAY FOR OTHER LOAN ARE MORE LIKELY TO DEFAULT THE LOAN

EMPLOYMENT AND INTEREST RATE



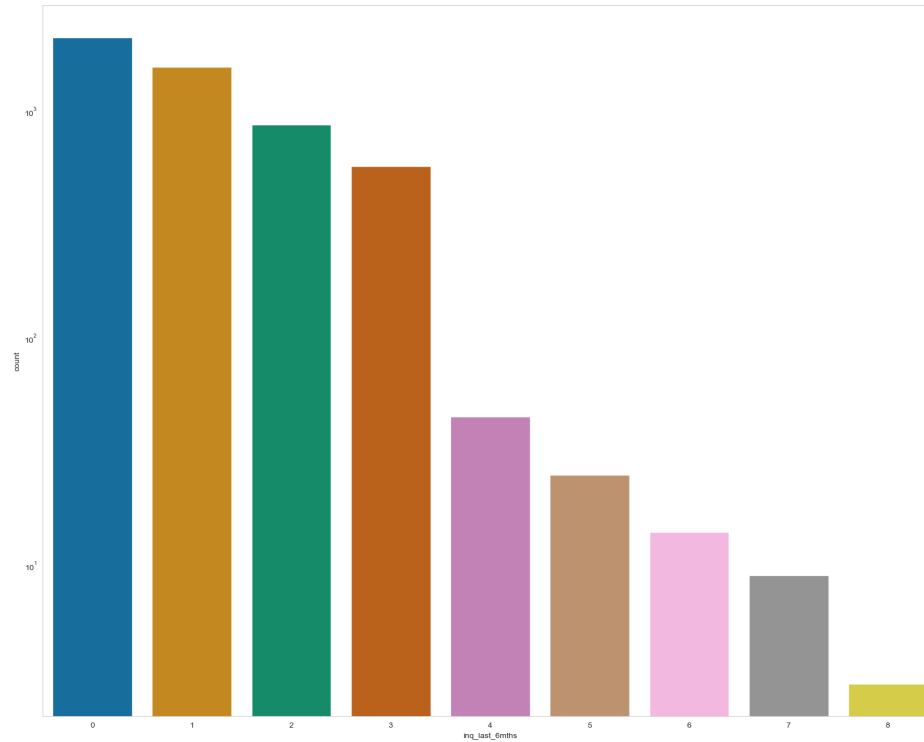
WHEN INTEREST RATE IS 13-17% AND WHEN PEOPLE WHO ARE MORE THEN 10 YEARS OF EMPLOYMENT ARE MORE DEFAULTERS

ANNUAL INCOME AND LOAN STATUS



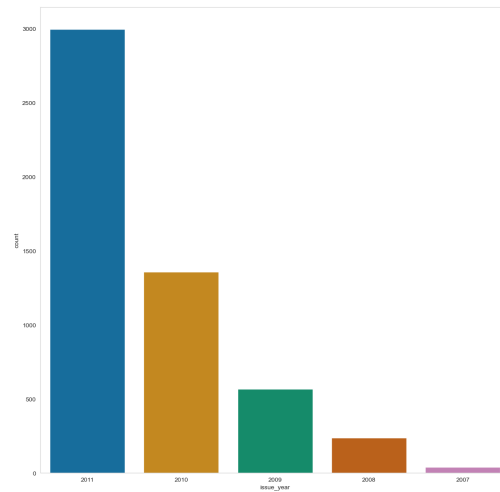
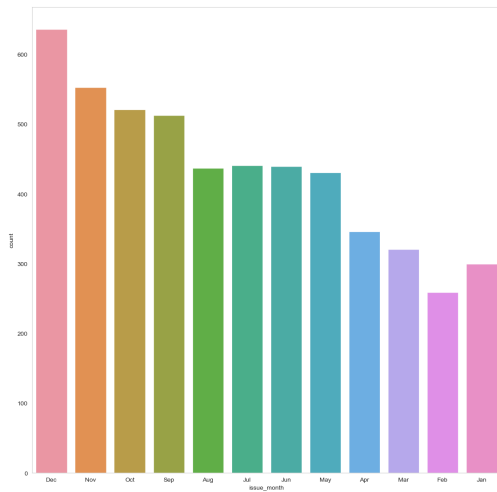
PEOPLE WITH LOW INCOME ARE MORE LIKELY TO DEFAULT THE LOAN

NUMBER OF INQUIRIES AND DEFAULTS



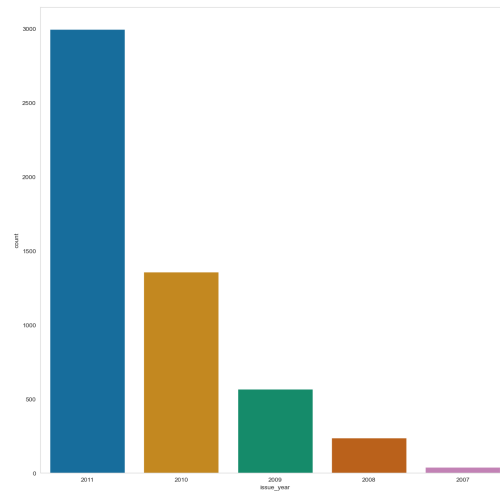
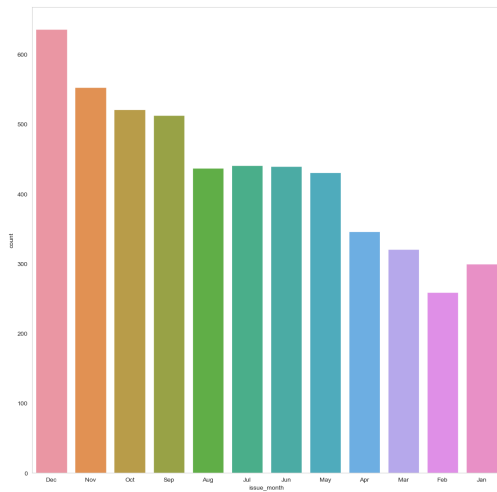
WHEN IN LAST 6 MONTHS LESS INQUIRES MORE DEFAULTS CAN OCCUR

DEFAULTERS IN TIME FRAME



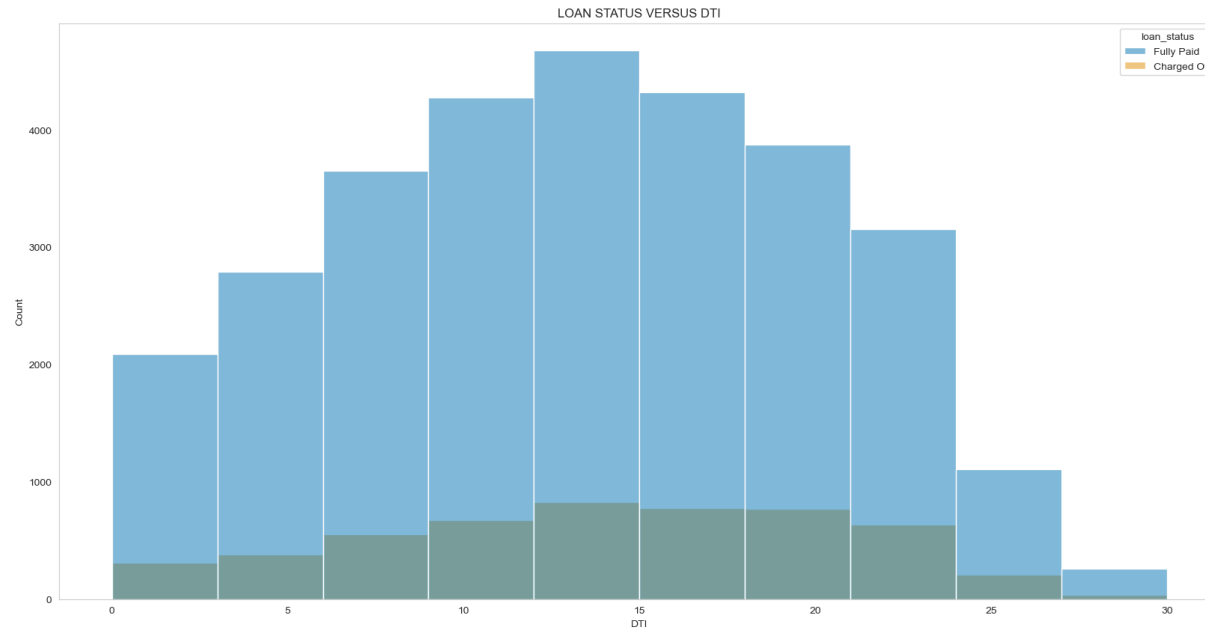
MORE DEFAULTS IN THE YEAR 2011 AND ALSO THE MAXIMUM LOAN

DEFAULTERS IN TIME FRAME



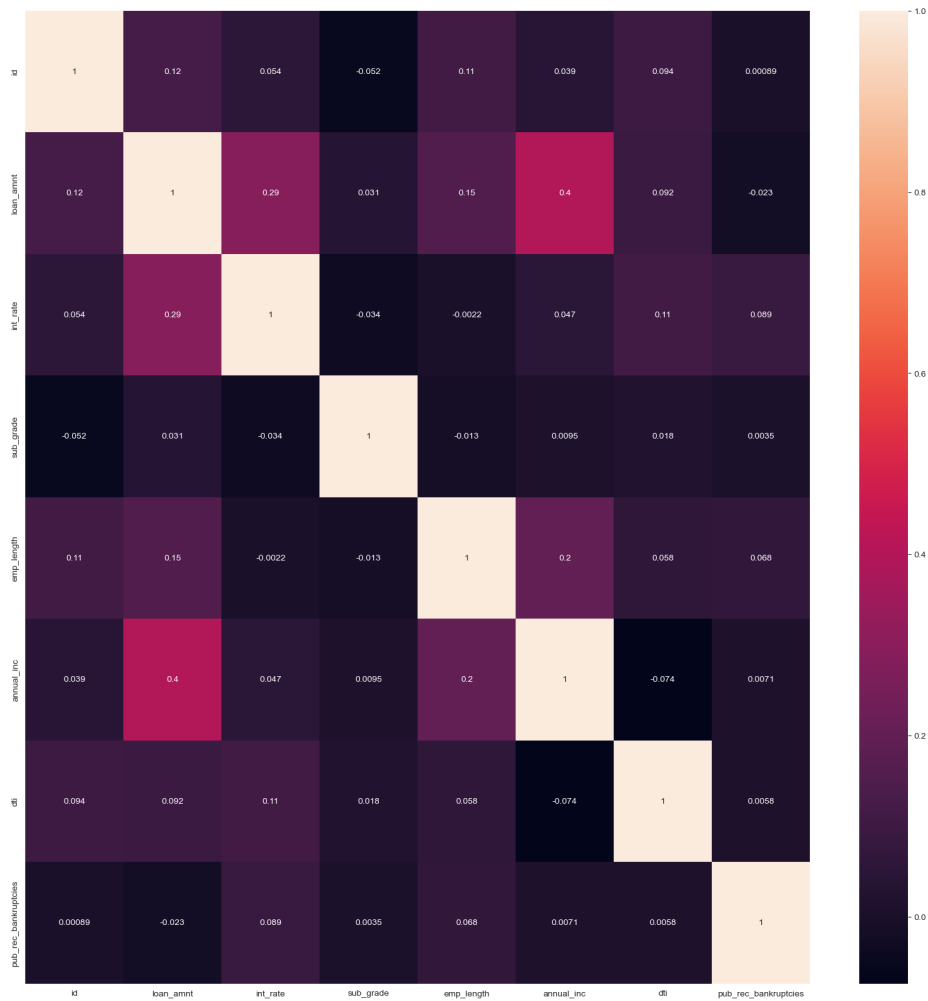
MORE DEFAULTS IN THE YEAR 2011 AND ALSO THE MAXIMUM LOAN

DTI AND LOAN DEFAULTS

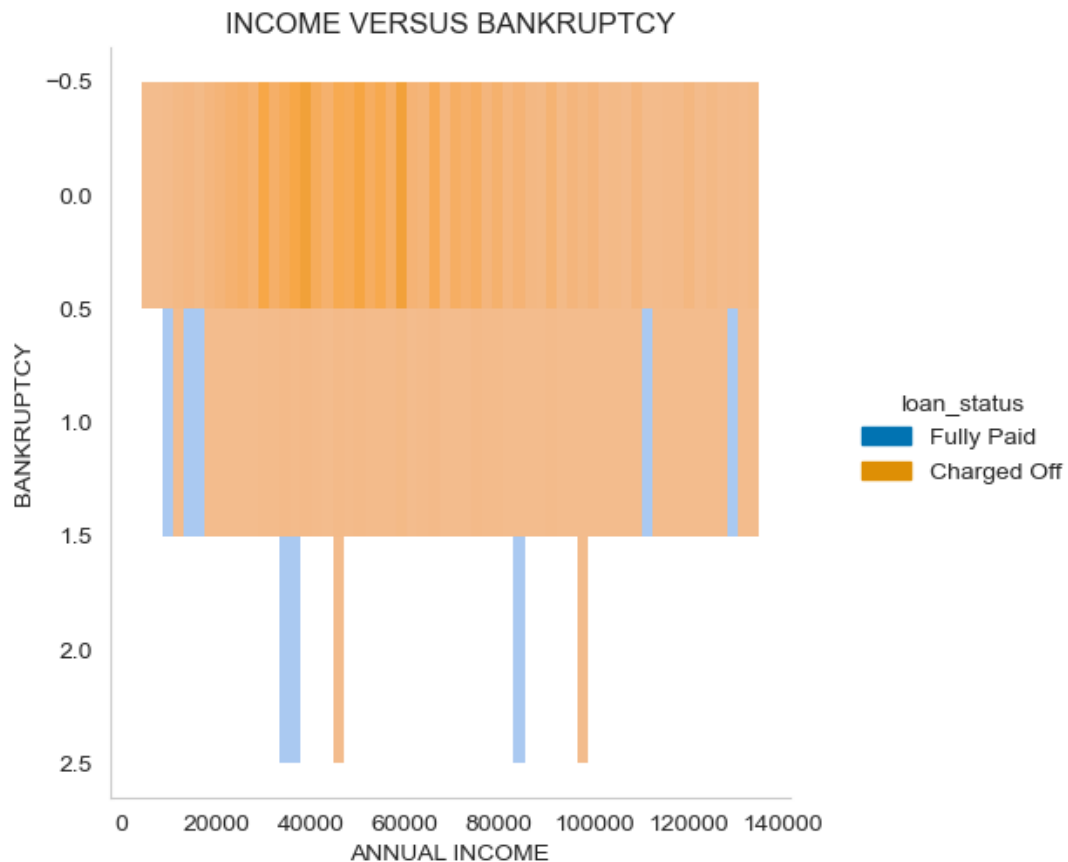


MORE DEFAULTS WHEN **DTI** RATIO PROPORTION IS 15

CORRELATION MATRIX



ANNUAL INCOME VS BANKRUPTCY



OVERALL RECOMMENDATION

MAJOR FACTOR OF LOAN DEFAULT

- (1) LOW LOAN GRADES (E/G/F)
- (2) LOAN VERIFICATION STATUS
- (3) ANNUAL INCOME
- (4) PUBLIC RECORD OF BANKRUPT
- (5) HIGH LOAN DTI (DEBT TO INCOME RATIO)
- (6) EMPLOYMENT > 10YEARS