



LENDING CLUB CASE STUDY

Varun Juyal



CONTENTS

Introduction

Business Objective

Data preparation and cleaning

Analysis with Observations

Recommendations

INTRODUCTION

Lending Club - largest online marketplace for peer-to-peer lending.

Provides an online platform for borrowers and lenders where they can transact smoothly and quickly.

Loan types - personal loans, business loans, and financing of medical procedures

Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss - credit loss.

Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

BUSINESS OBJECTIVE

Company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment

DATA PREPARATION & CLEANING

- The data provided has a total of 39,717 rows and 111 columns
- Data type details - float (74), integer (13) and object (24)
- While inspecting the data set we found, there are lot of columns with missing values. Of the total fields in the data set, almost 51% of the data was missing.
- There were columns, which had all the rows as “NA”, we can completely get rid of these. The threshold for dropping column was any column with more than 40% NA values.

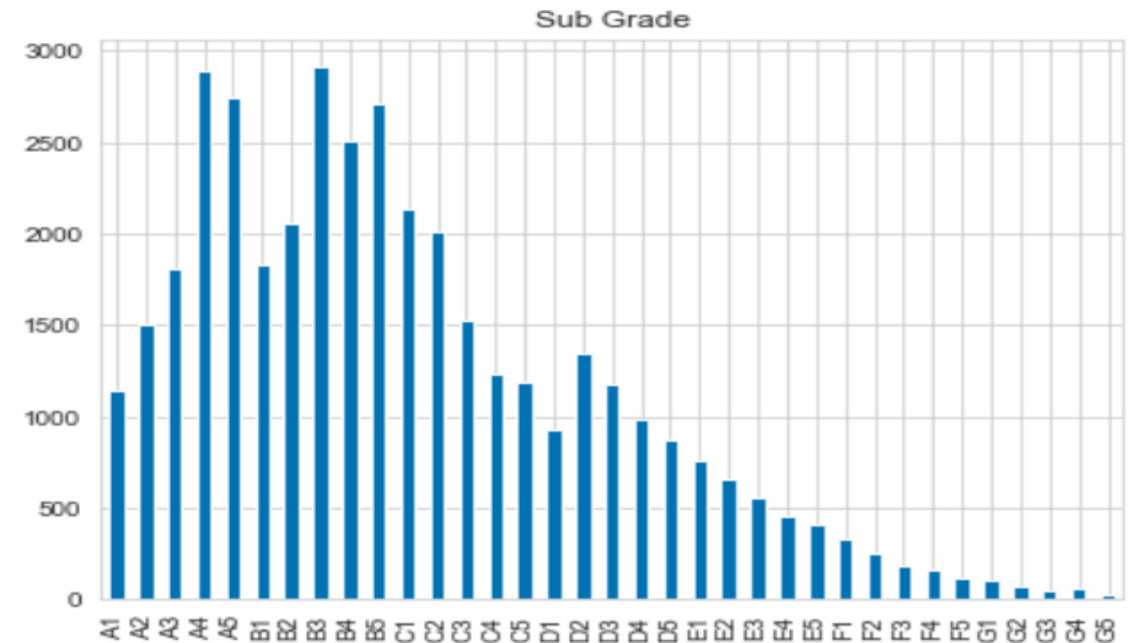
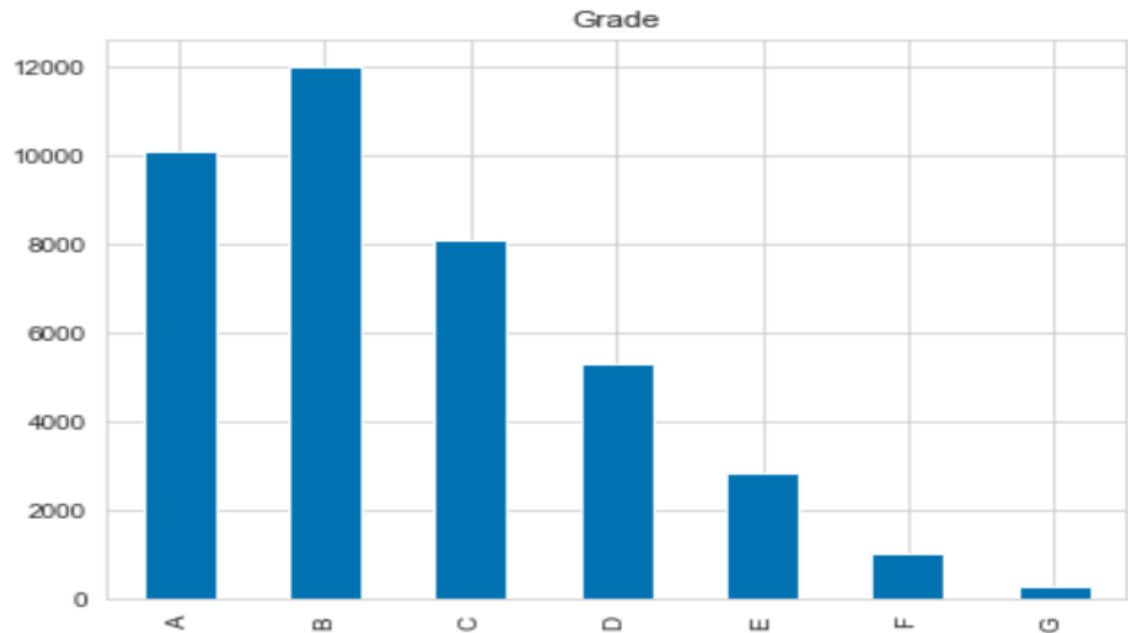
DATA PREPARATION & CLEANING

- On further analysis, we observe that there are columns with data on customer behavior after loan has been granted, these will be redundant for the analysis needed.

`'funded_amnt','funded_amnt_inv','emp_title','url','desc','title','pymnt_plan','zip_code','addr_state','delinq_2yrs','earliest_cr_line','open_acc','pub_rec','revol_bal','revol_util','total_acc','out_prncp','out_prncp_inv','total_pymnt','total_pymnt_inv','total_rec_prncp','total_rec_int','recoveries','collection_recovery_fee','last_pymnt_d','last_pymnt_amnt','last_credit_pull_d','application_type','tax_liens'.`

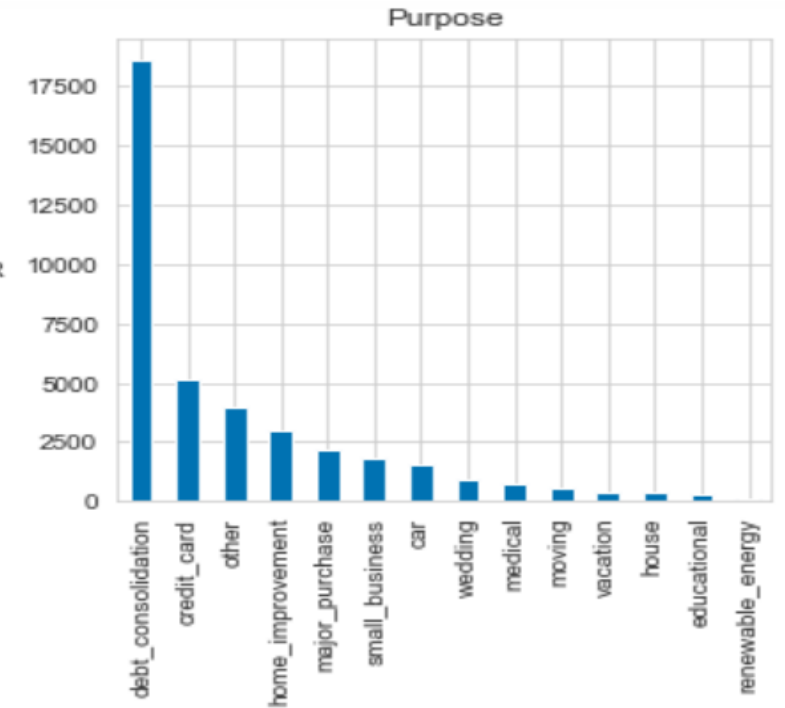
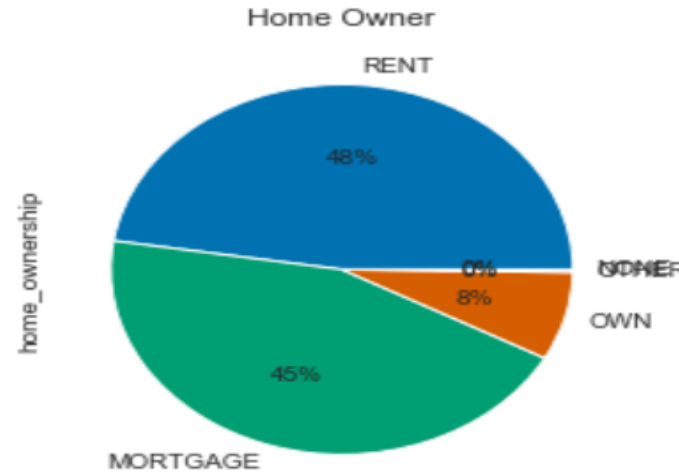
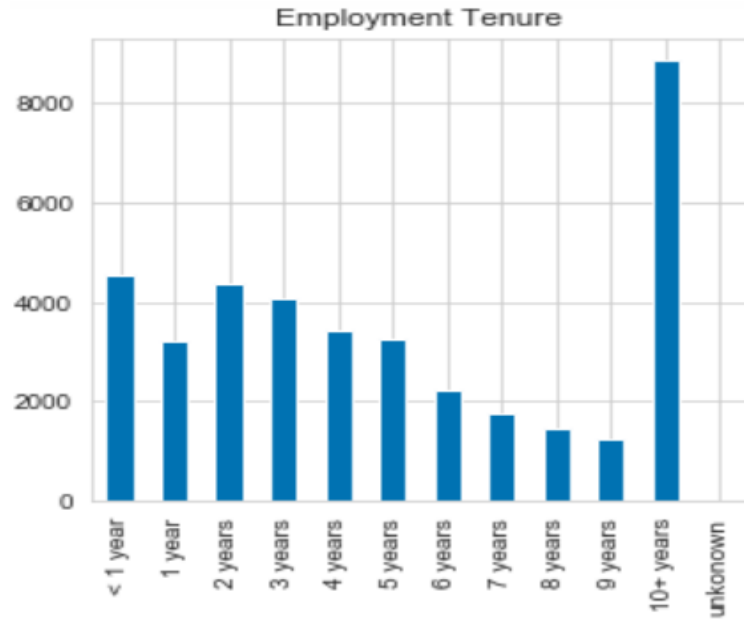
- In addition, upon removing the “NA” values from the columns, we still had some “NA” values, which we dropped by removing some of the rows.

ANALYSIS – UNIVARIATE



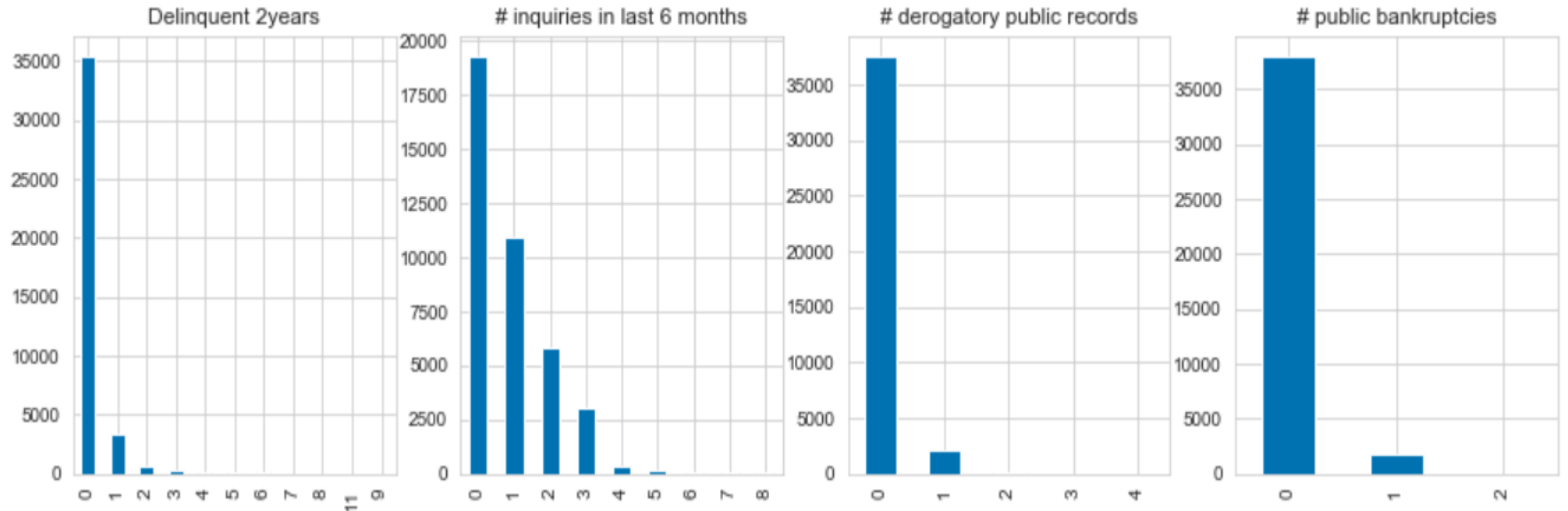
- Even though loan application acceptance decreases with decrease in grade, however grade B has a higher frequency than A. This could be because grade A are more affluent and fewer people need a loan.
- Within grades A & B, number of loan application acceptance increases with decrease in subgrades, while for C and lower, the acceptance decrease with decrease in sub-grades. A reason could be that for grades A & B, loans are accepted irrespective of subgrades and more applications are received with decreasing sub grades, whereas for C & below, sub grades have an impact on application acceptance.

ANALYSIS – UNIVARIATE



- Loan applications accepted generally decreases with increase in years if employment, indicating more need of loan in early years of employment.
- Customers with rented and mortgaged house apply and get accepted for loan the most
- Most loans are for debt consolidation

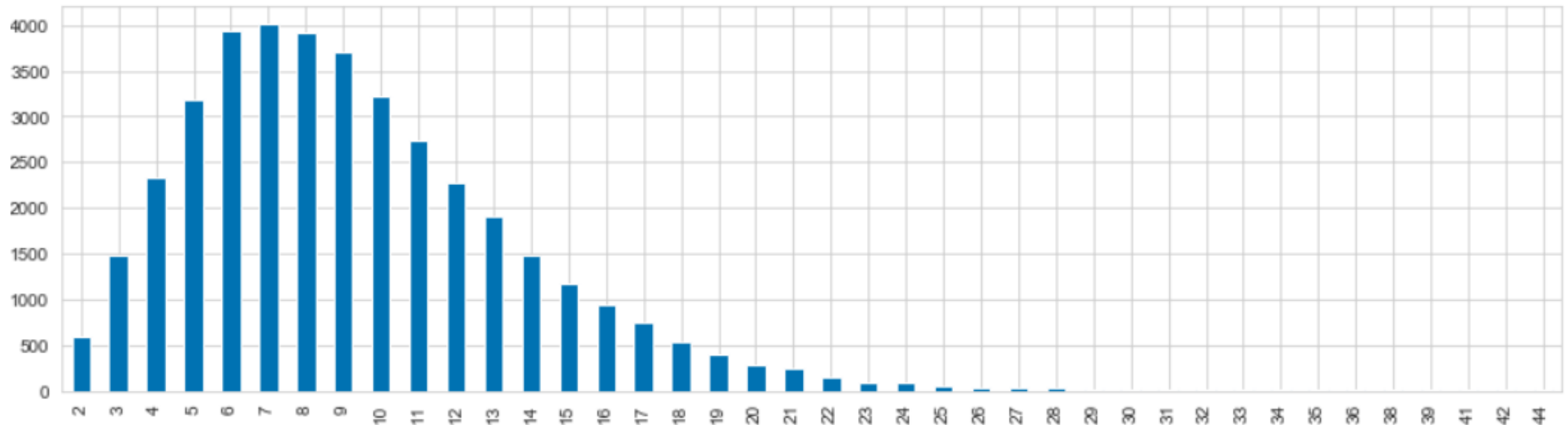
ANALYSIS — UNIVARIATE



Loan applications are almost always being accepted for customers with no derogatory* records.

A *derogatory item is an entry that may be considered negative by lenders because it indicates risk and hurts your ability to qualify for credit or other services

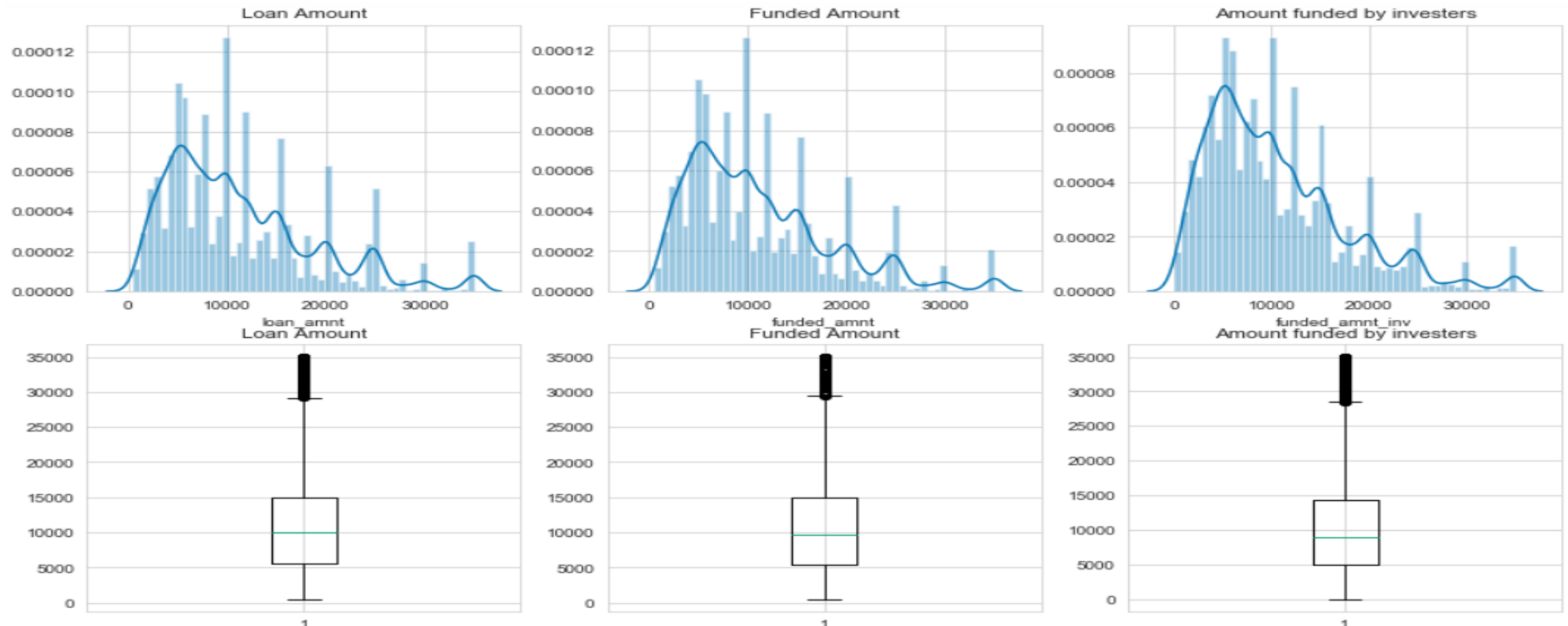
ANALYSIS – UNIVARIATE



Loan acceptance have a tendency towards 7 open accounts* on credit file

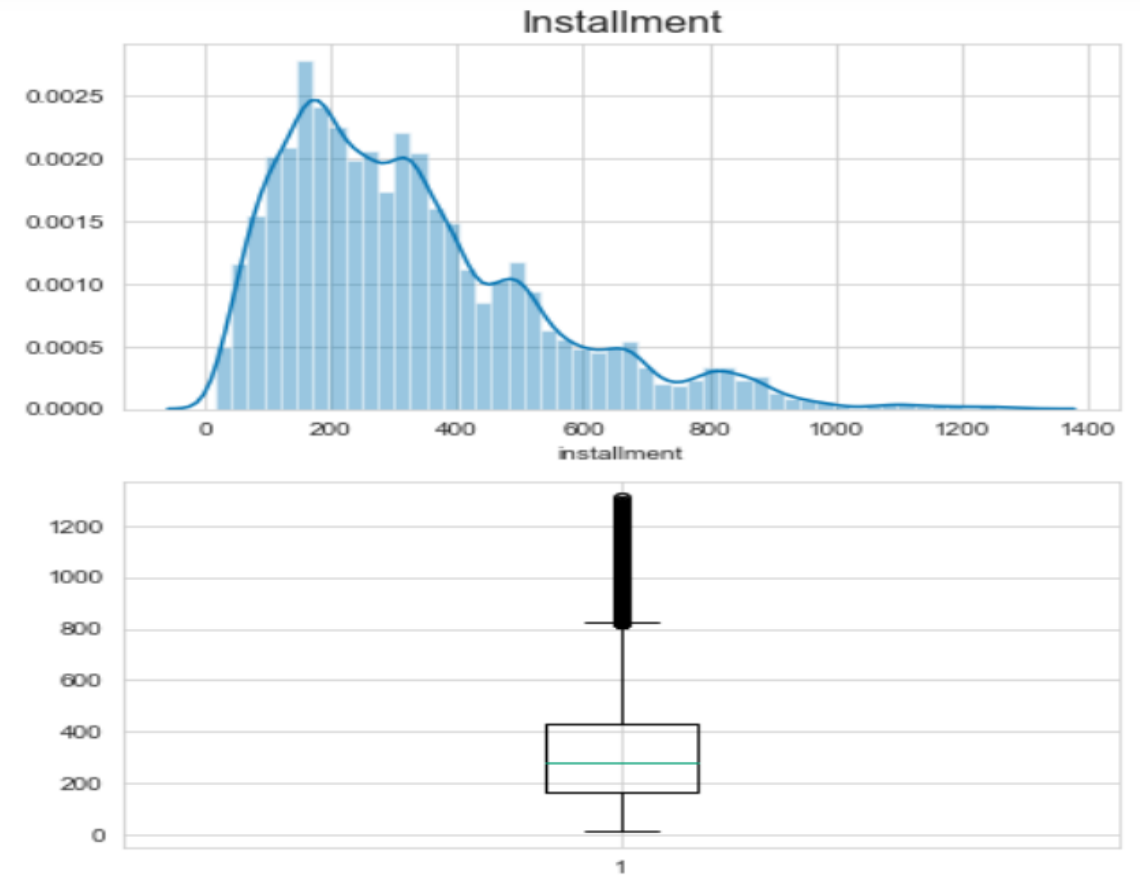
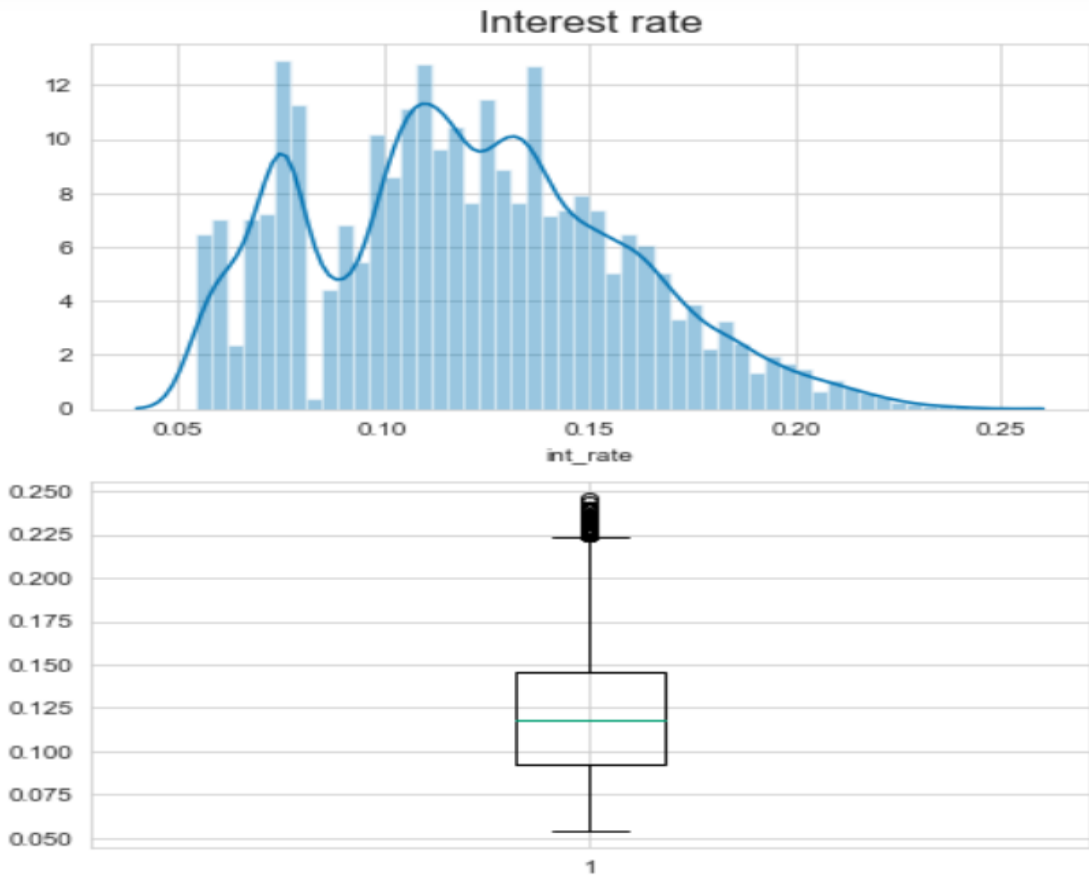
Also referred to as “open credit,” *open accounts are a hybrid of installment and revolving credit. The payment is not the same each month, and it’s usually due in full at the end of each billing cycle. The consumer satisfies financial responsibility for the account when the bill is paid in full each month

ANALYSIS – UNIVARIATE



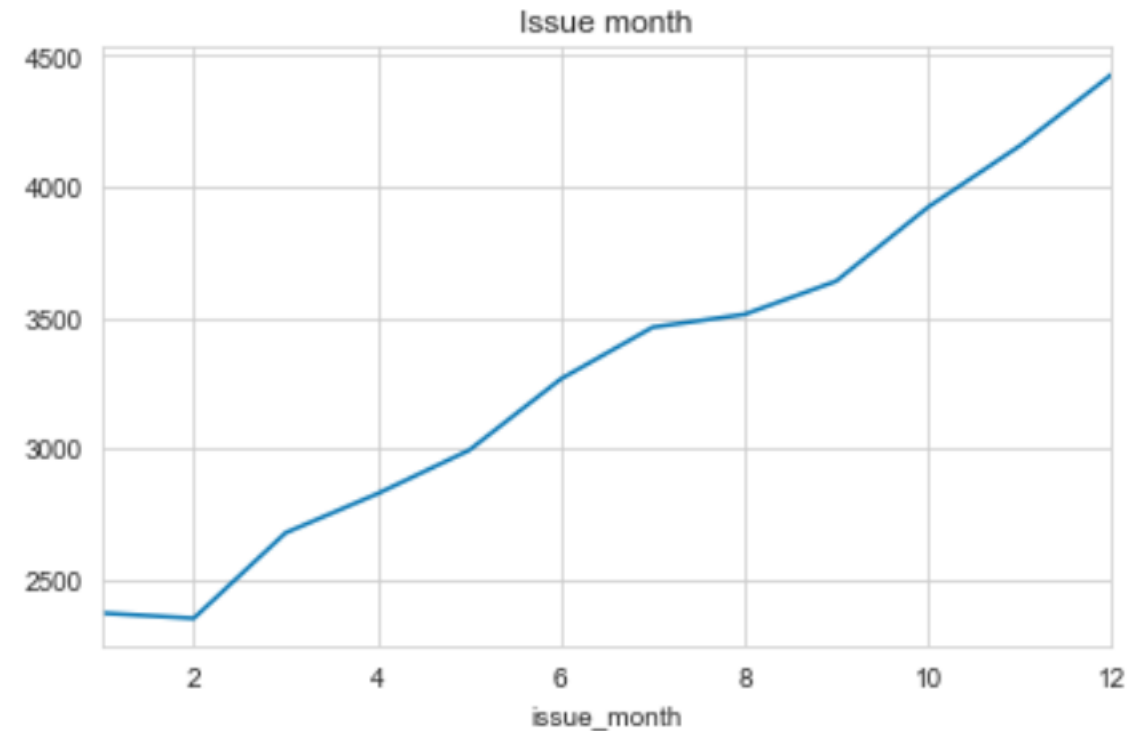
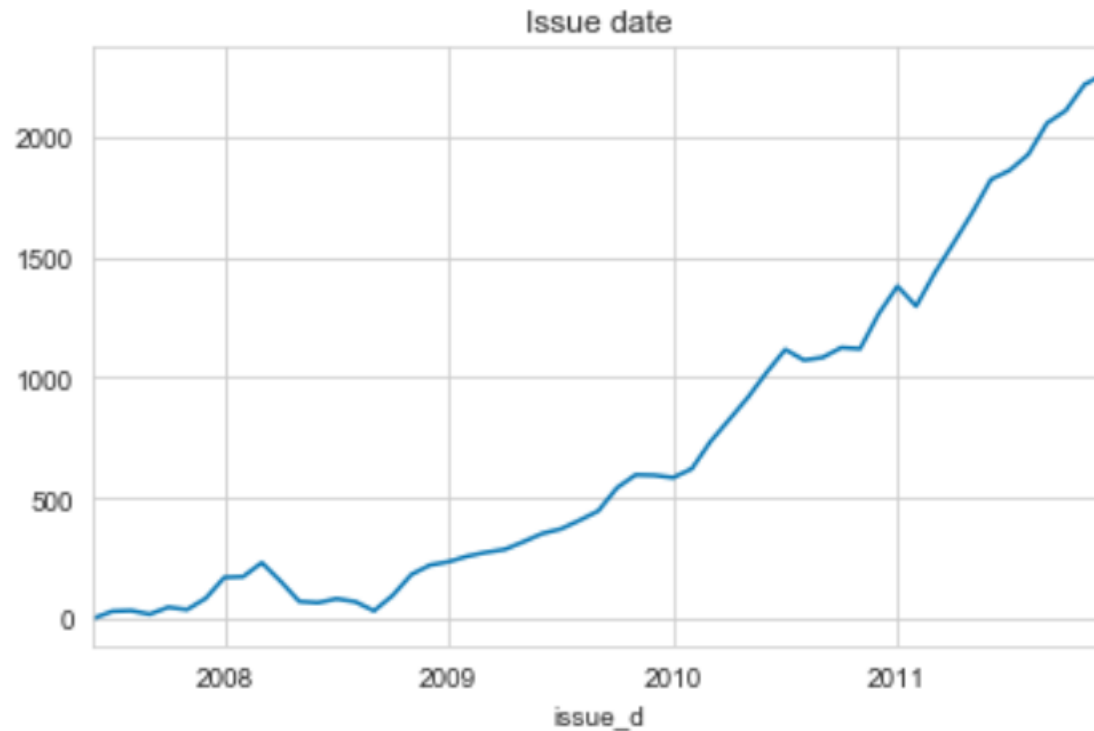
There are peaks at intervals possibly due to rounding off. Most loan amounts are in the range of IQR(5000-15000)

ANALYSIS – UNIVARIATE



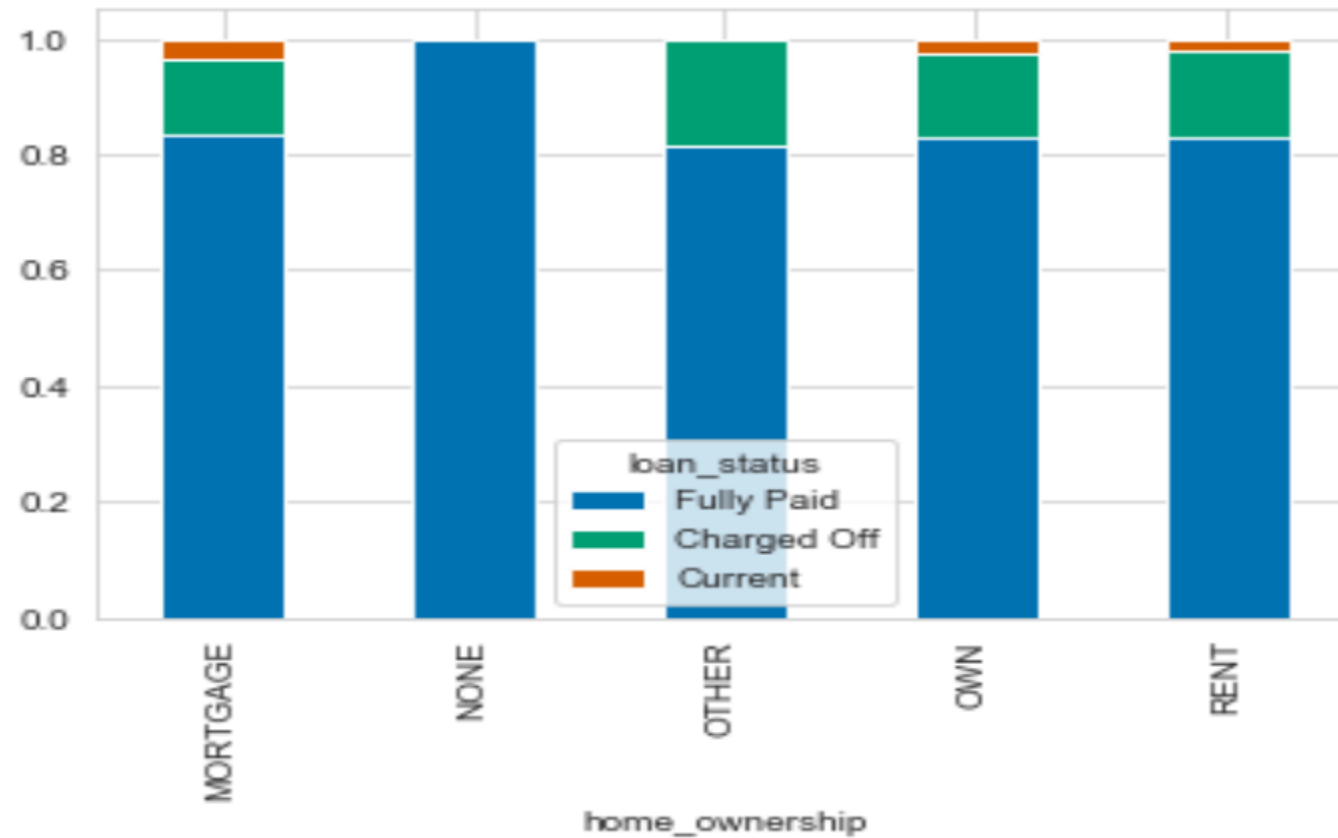
Very few loans given at 8% interest rate

ANALYSIS – UNIVARIATE



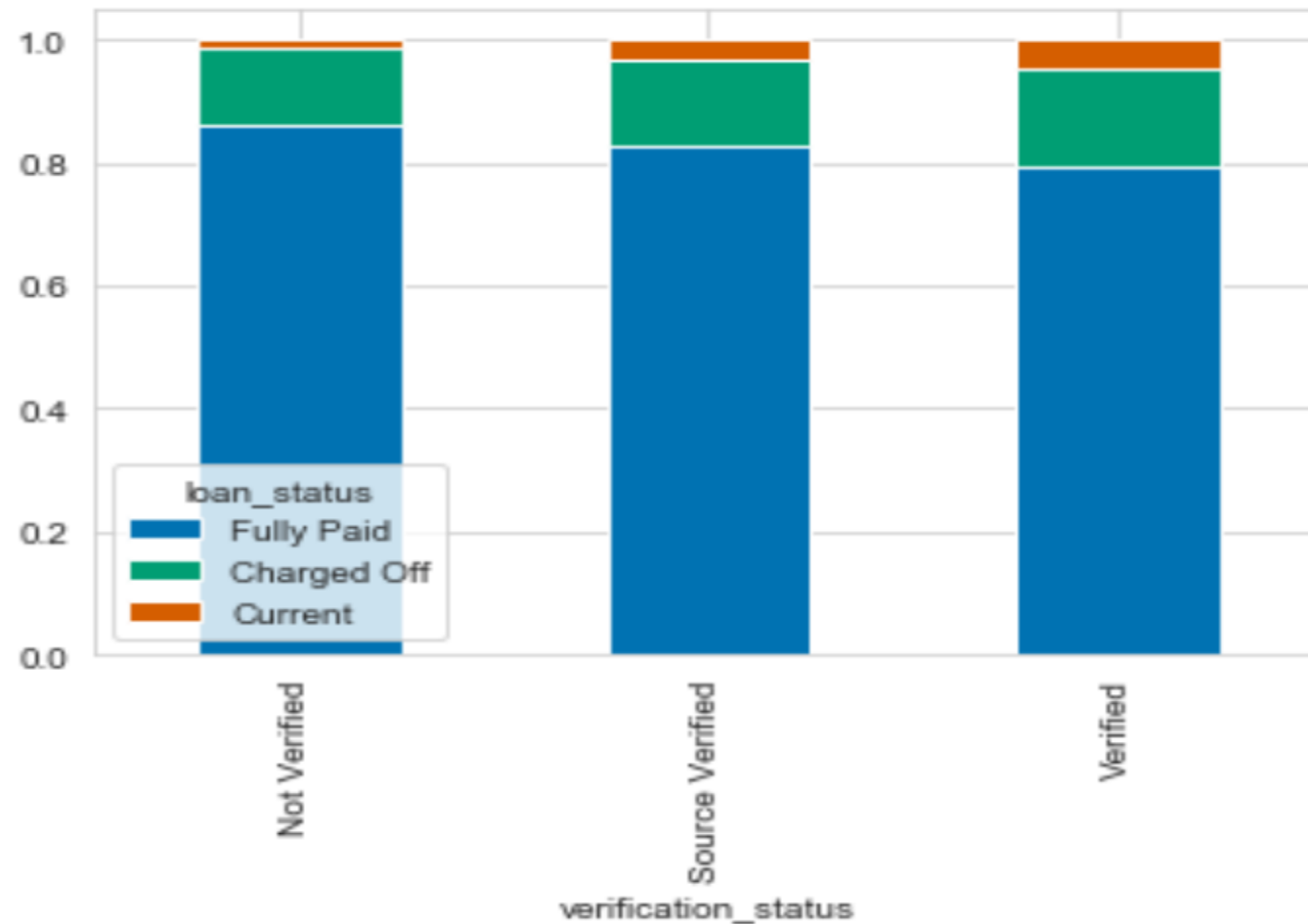
- Almost linear growth in loans since 2009
- Almost linear growth in loans starting March throughout Dec
- Slope: Jan-Feb least number of loans issued | March-May, Jul-Sep growth rates of loan issuance are slower than May-Jul, Sep-Dec

ANALYSIS — BIVARIATE



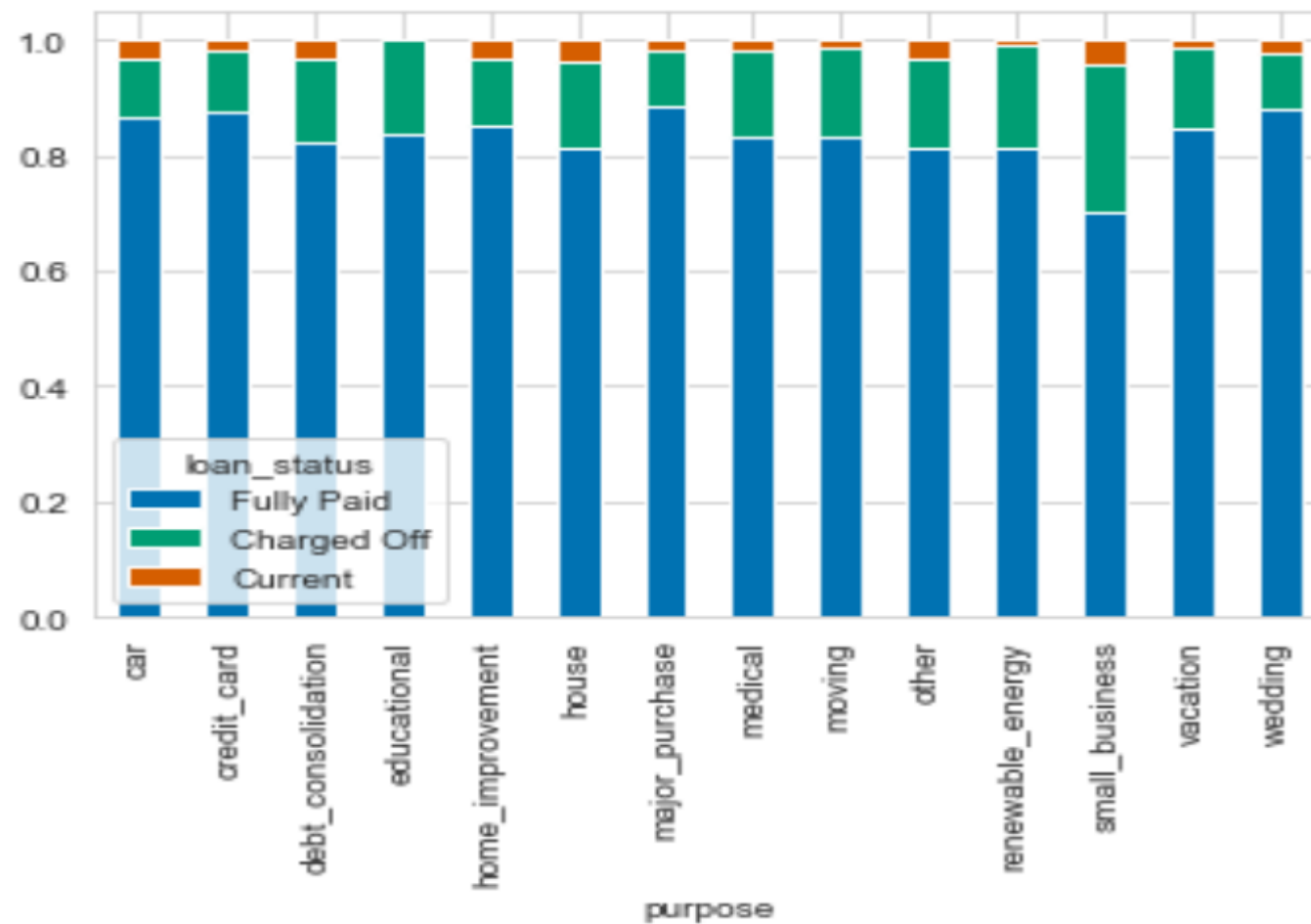
Defaults tend to increase as this grades drop – Bad performance leads to increase in defaults. The bad performance could possibly be linked to some other factor.

ANALYSIS — BIVARIATE



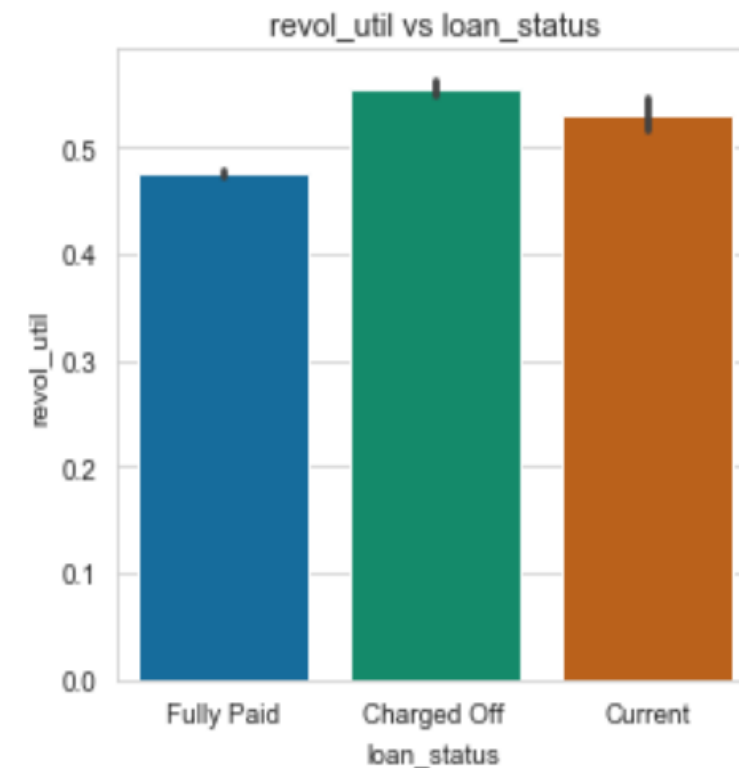
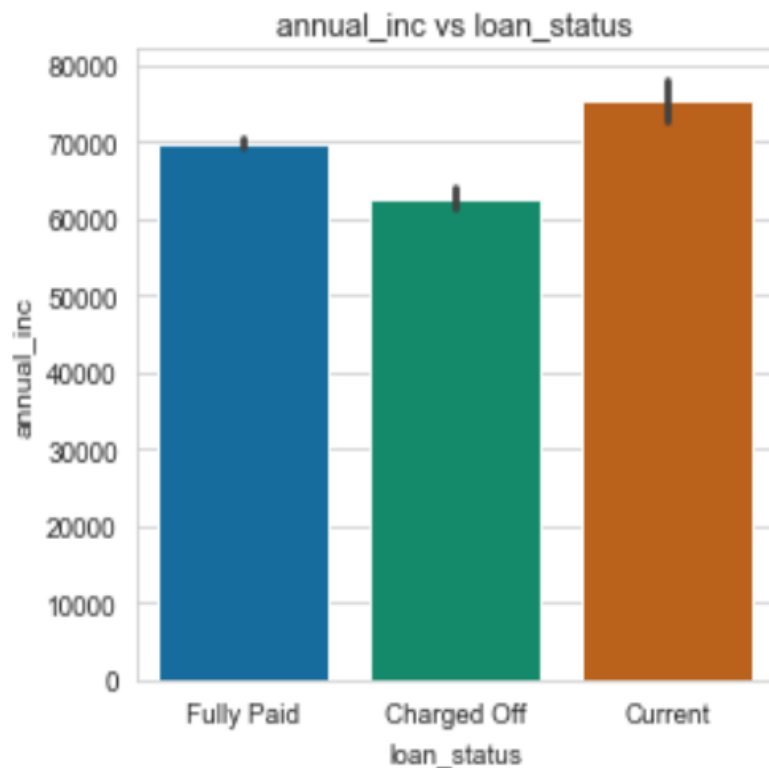
Non verified accounts seem to be doing better as their default rate is lower

ANALYSIS — BIVARIATE



Small business seem to have the highest proportion of defaults when compared with other purposes

ANALYSIS — BIVARIATE



- Delinquency is high for lower annual income.
- Higher DTI for Charged off customers opposed to Fully paid
- Charge off frequencies higher for high revol_util

RECOMMENDATIONS

- Small Business' have defaulted more. They are potential for higher interest rates and also proper due diligence
- Reduce the high interest loans for 60 months tenure as they have high propensity for default
- Higher interest rates should be charges to borrowers with debt-to-income ratio
- Focused effort required to tap low experience customers
- Focused effort required to tap customers with rented and mortgaged houses
- More discretion and due diligence required while sanctioning loans to some states - CA, FL and NY