LENDING CLUB CASE STUDY

By Elayaraja Muthuraj
EPGP ML and AI, UpGrad

PROBLEM STATEMENT

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Lending loans to ‘risky’ applicants is the largest source of financial loss (called 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.

In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

Aim of this case study is Identify the risky loan applicants using Exploratory Data Analysis, then such loans can be reduced thereby cutting down the amount of credit loss.

Lending Club wants to understand the driving factors behind loan default and identify the variables which are strong indicators of default.

CASE STUDY APPROACH

Data Cleansing

Find Columns to apply to Exploratory data analysis methods

Data Analysis

- Univariate Analysis
- ByVariate Analysis
- Plot and charts

Observations

DATA CLEANSING

Pandas, numpy, seaborn and matplotlib.pyplot libraries used for the analysis

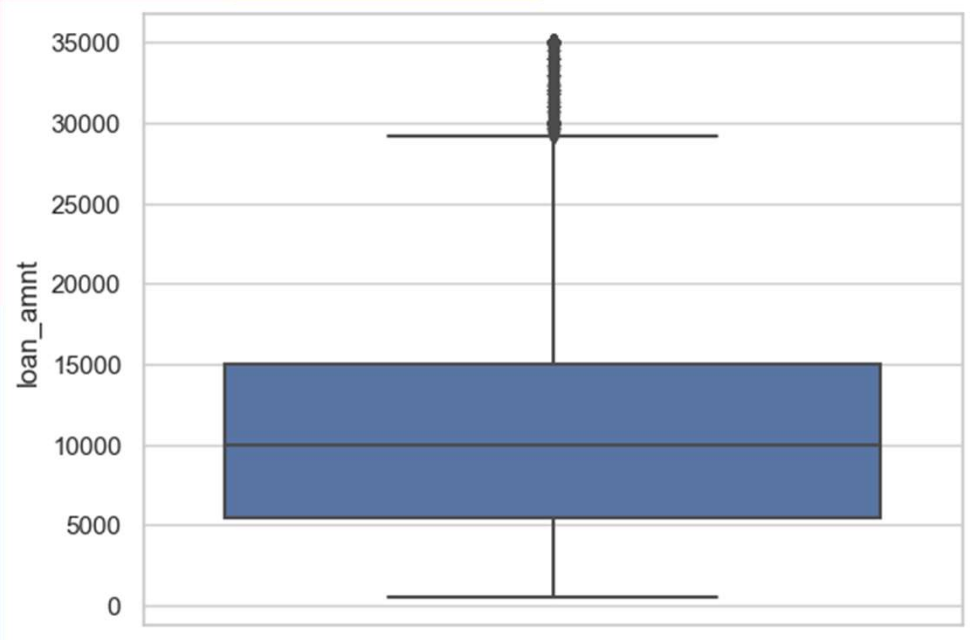
Almost 50+ columns identified with null values are dropped from the data

Derived variables created from Date and year columns of loan issued year

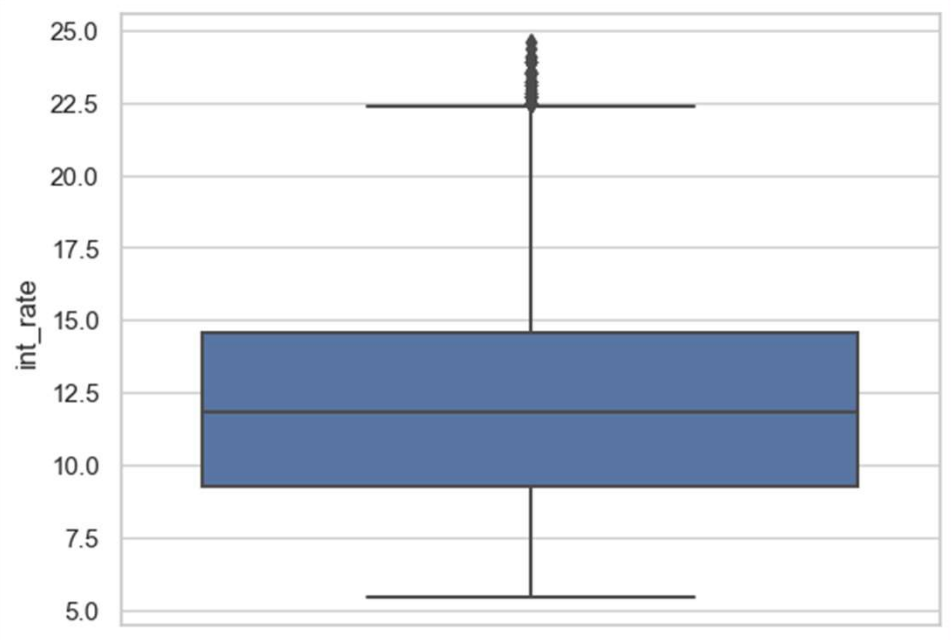
% to be stripped from intrest rate column and convert emp_length column to have only digit values

converted 'loan_amnt', 'funded_amnt','int_rate', 'funded_amnt_inv', 'installment', 'annual_inc','dti', 'emp_length', 'total_pymnt' columns into numeric data for considering to calculations

UNIVARAITE ANALYSIS

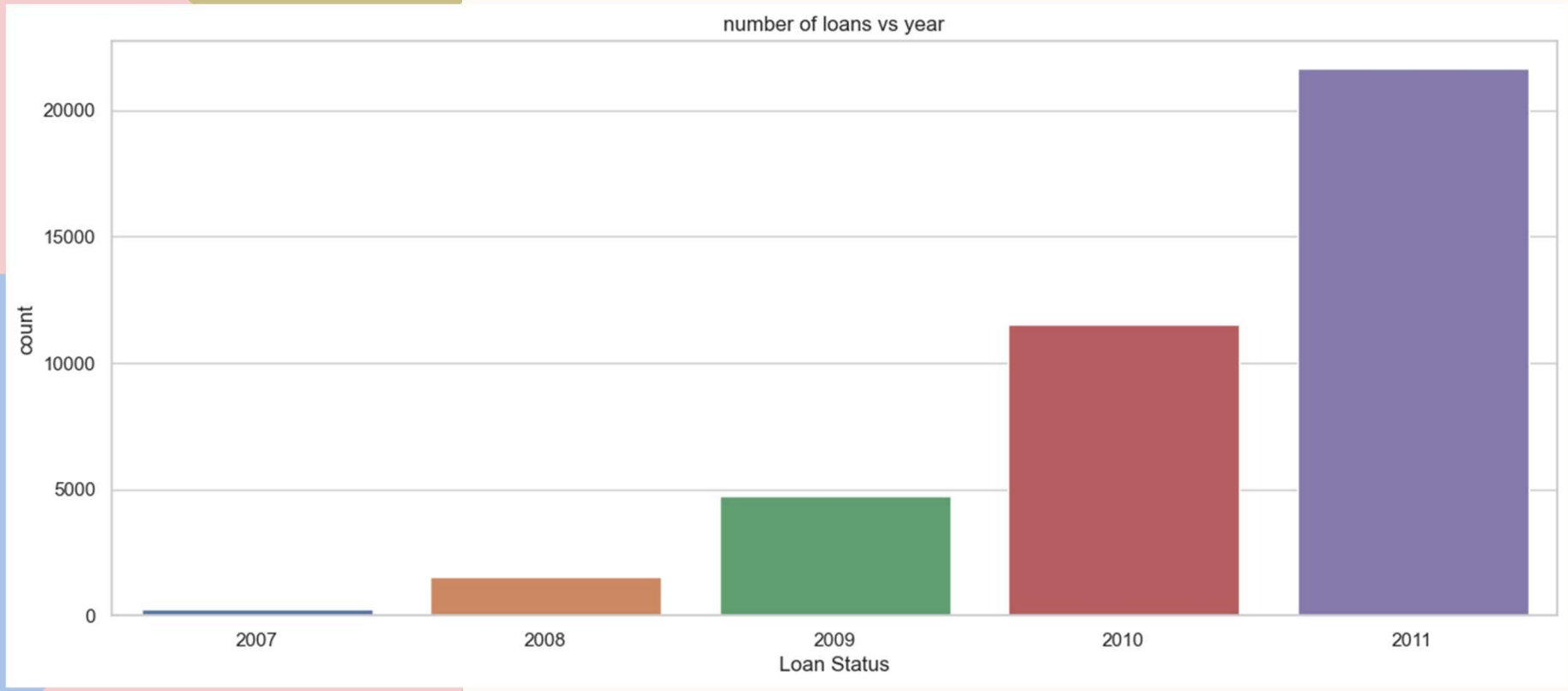


Most loan amounts within the range of 5000 to 15000

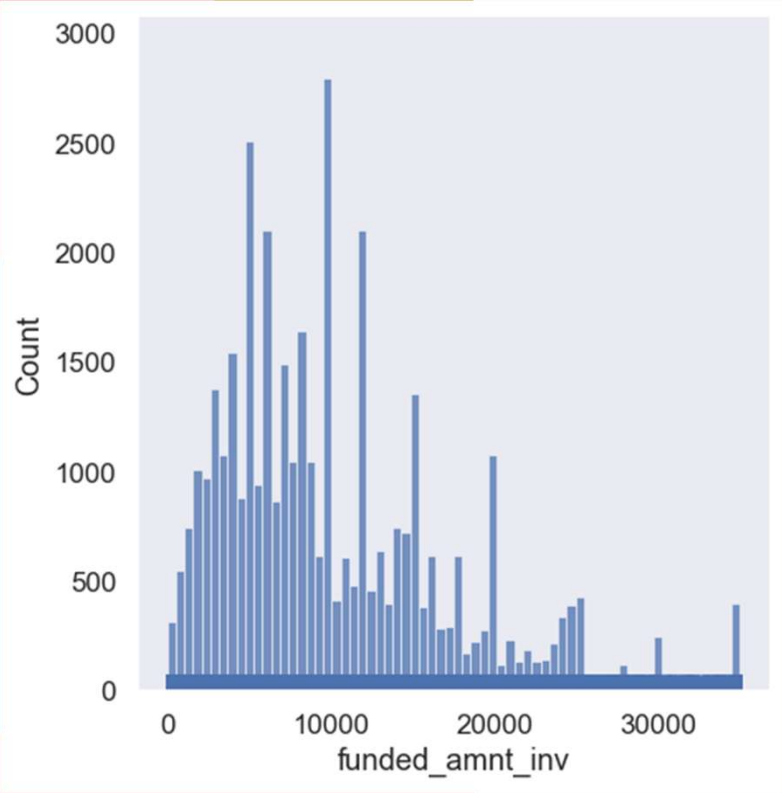


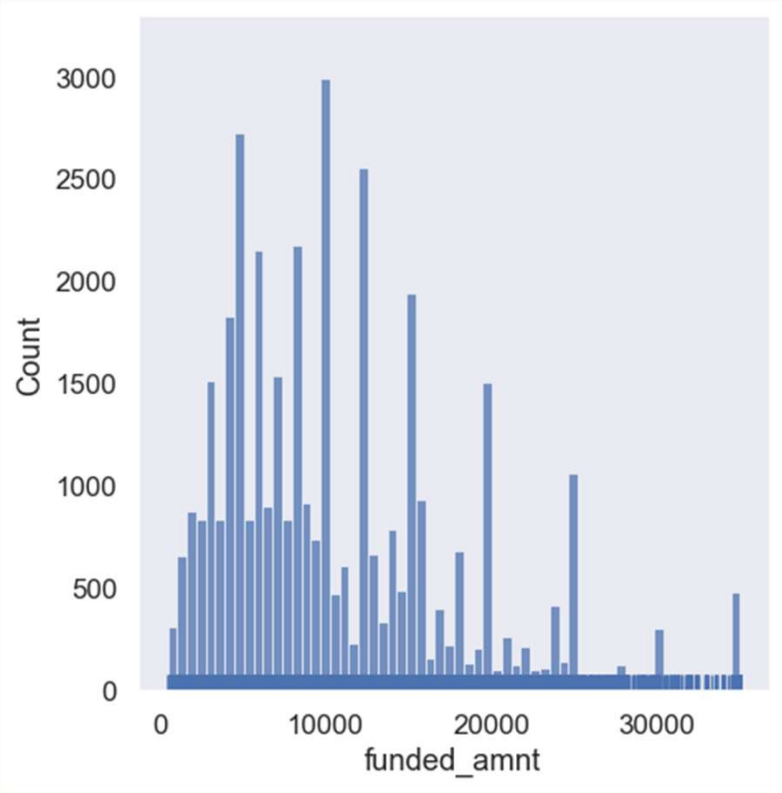
Most loan interest are within the range of 8% to 15%

Looking at loan_status vs int_rate vs issued year by plotting

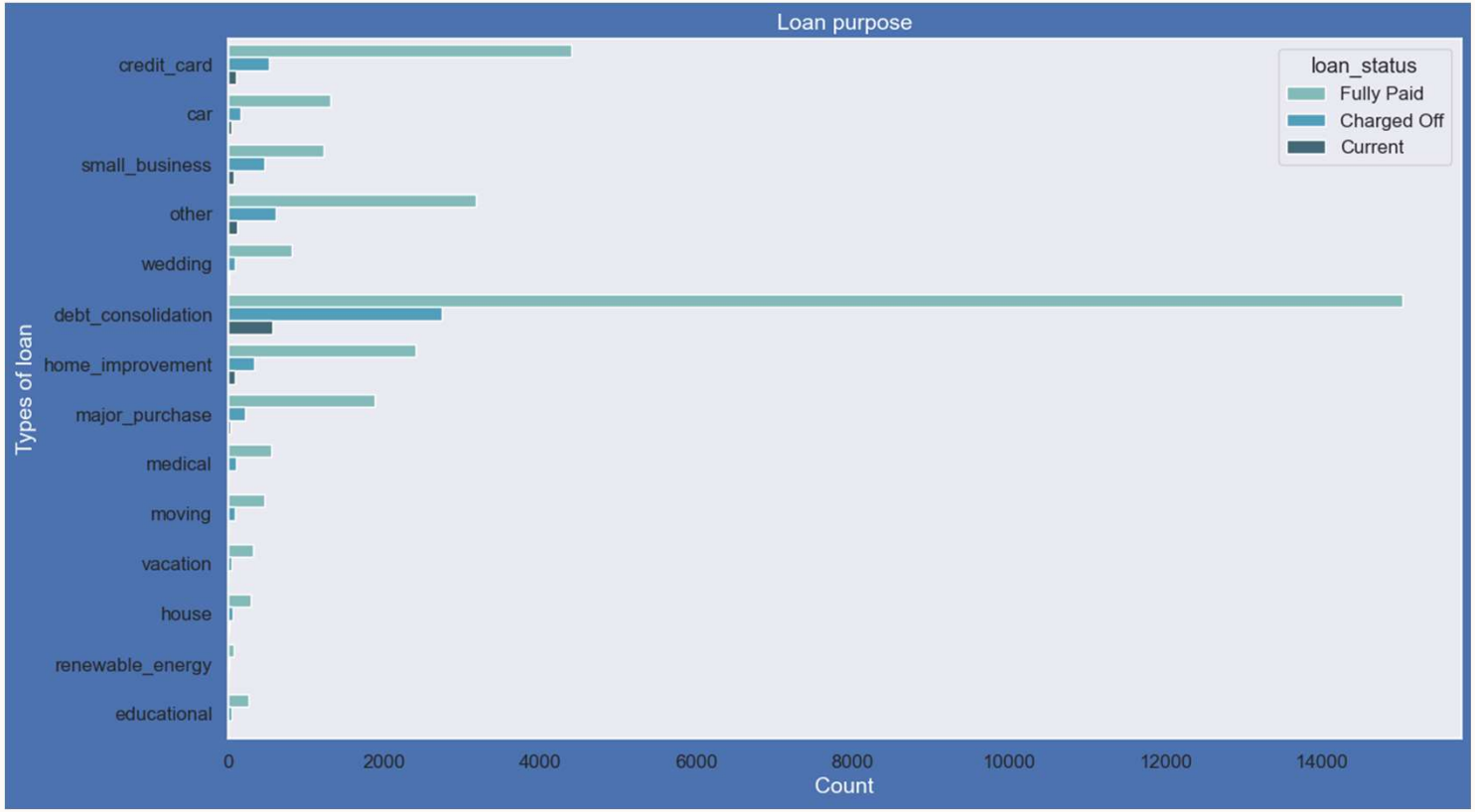


Most loan amounts within the range of 5000 to 15000

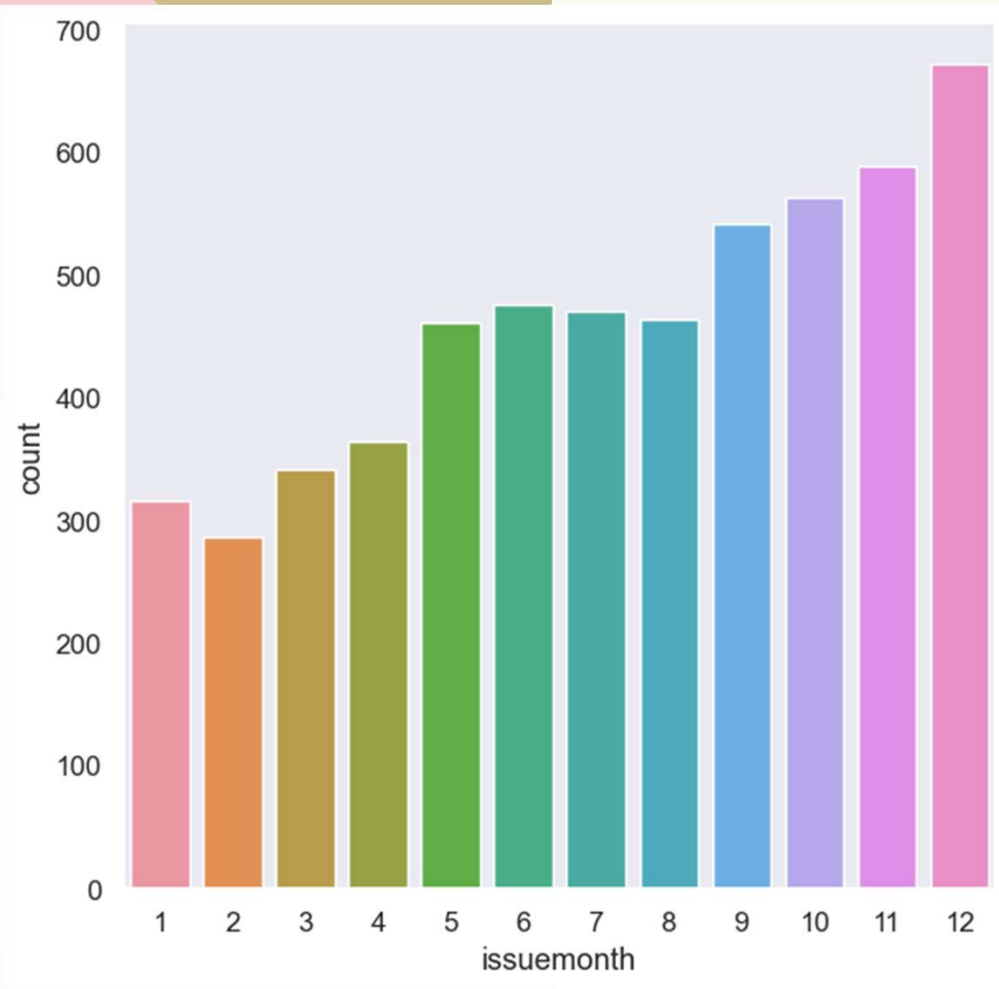


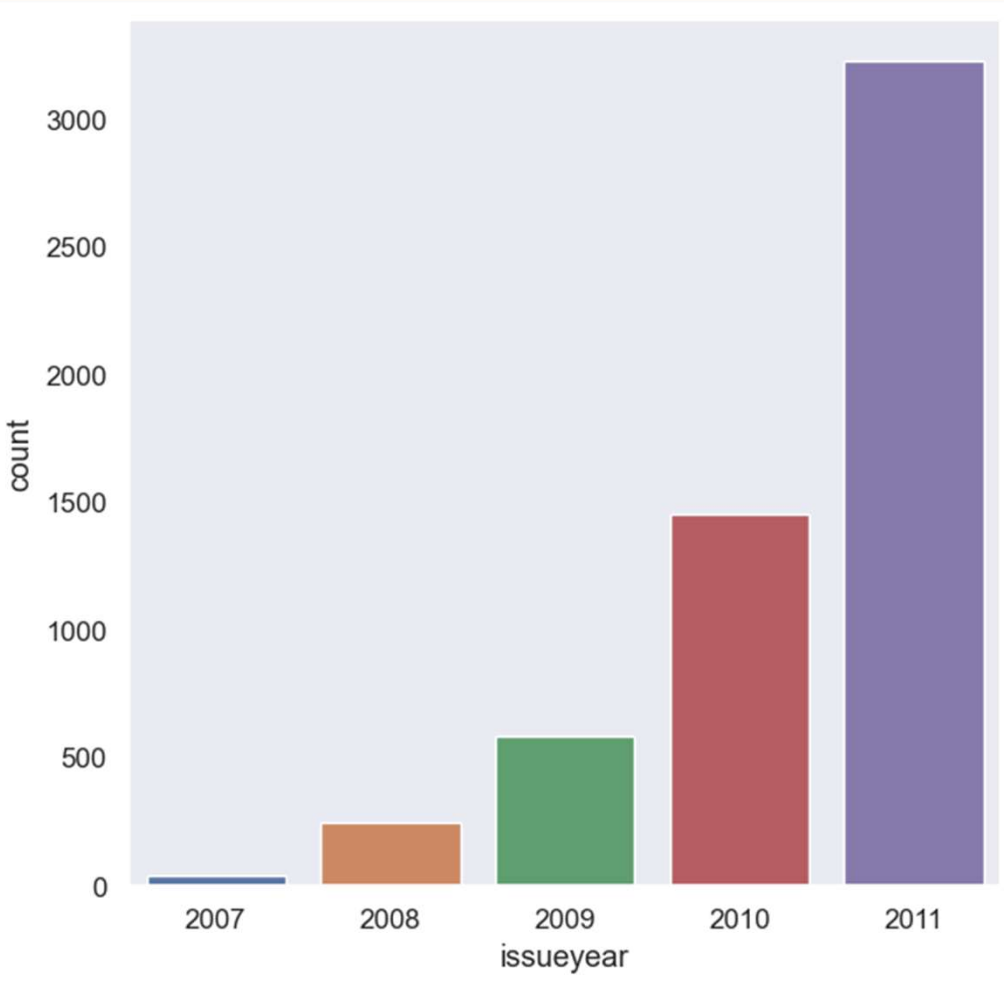


Loan purpose with status



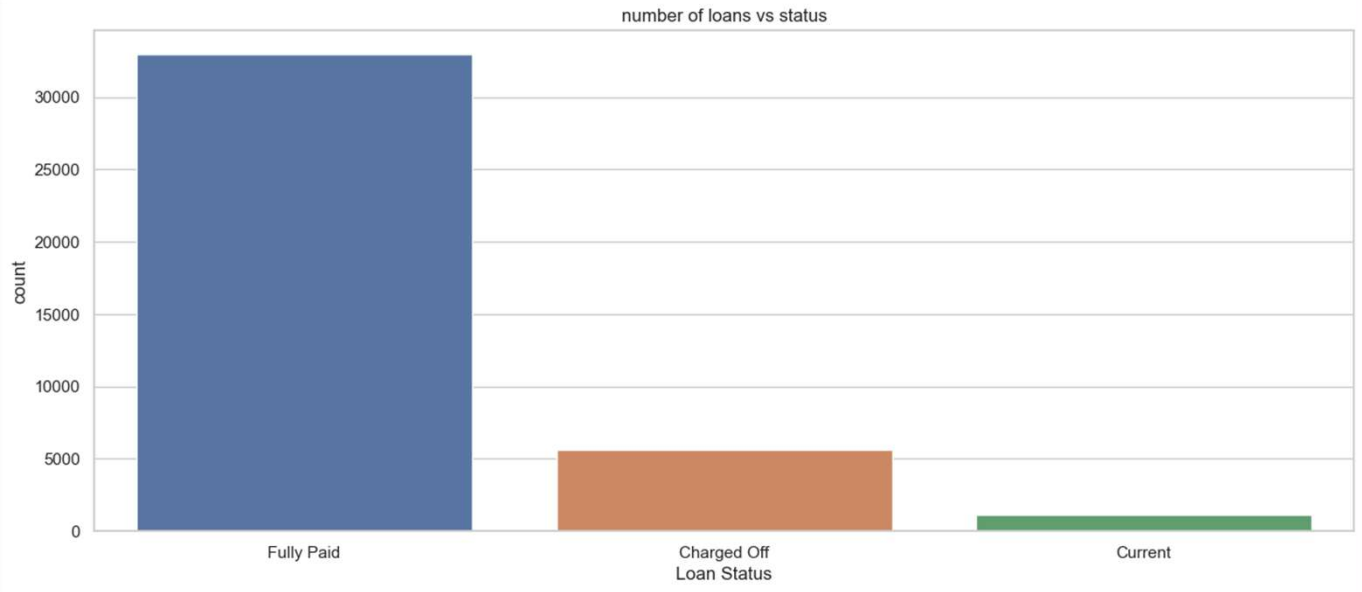
Charged off Loans vs Issued Month and Year





BIVARAITE ANALYSIS

Looking at loan_status vs int_rate vs issued year by plotting

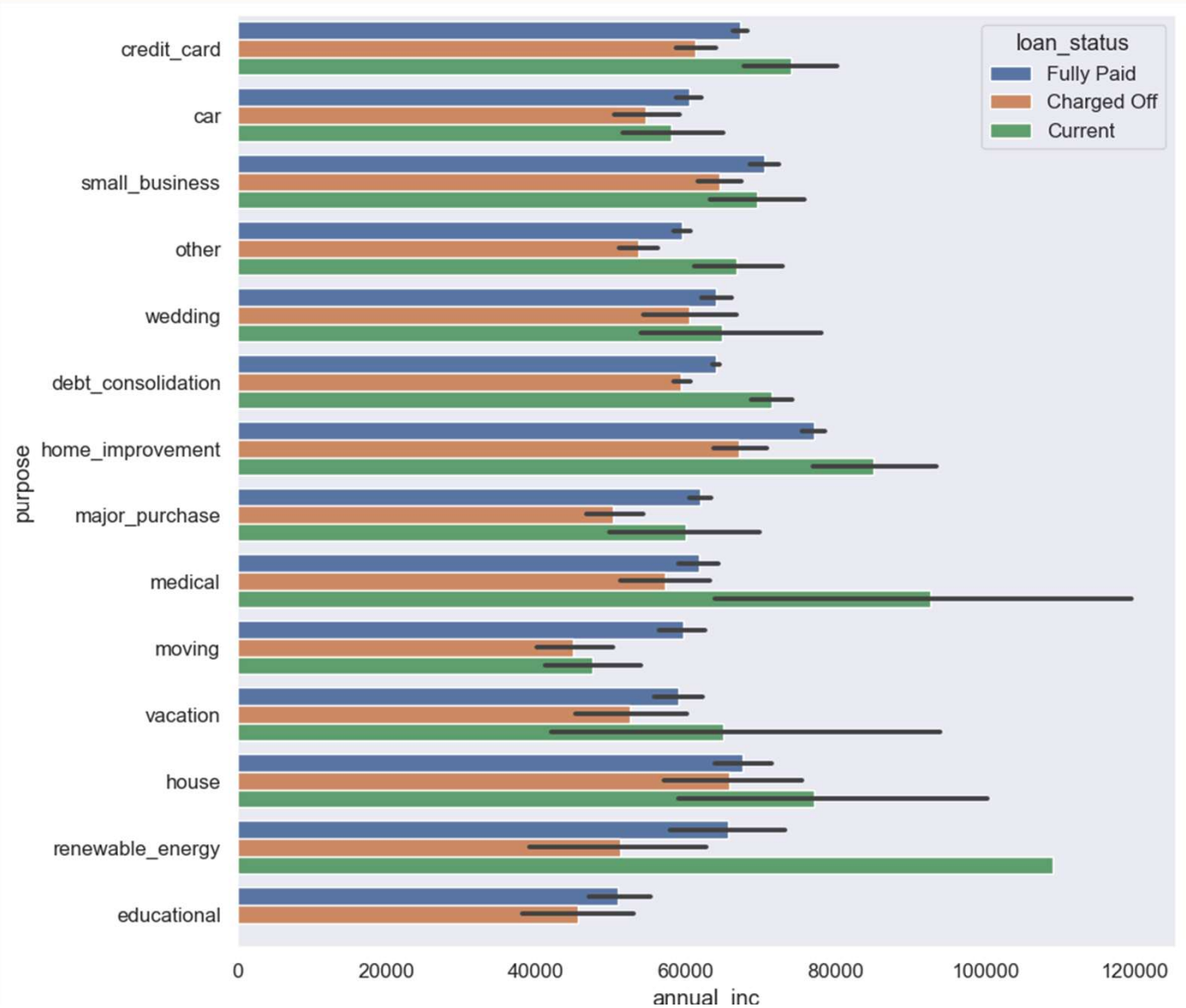


Bivariate Analysis of annual income versus default rate -

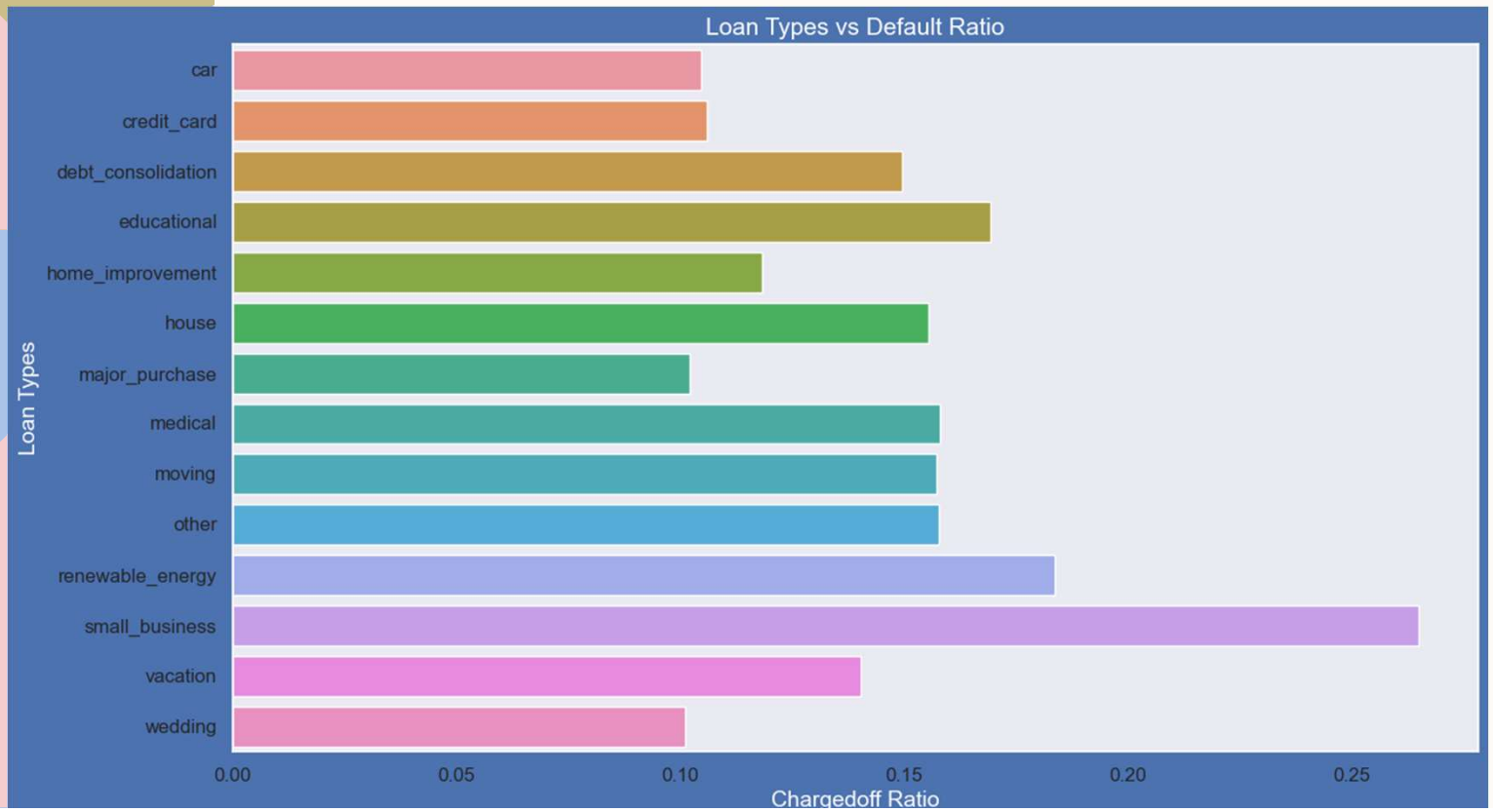
loan_status	annual_inc_grp	Charged Off	Current	Fully Paid	Total	Charged_off_ratio
0	0-20000	237	9	943	1189	0.199327
1	20000-40000	1514	170	7004	8688	0.174263
2	40000-60000	1729	345	9534	11608	0.148949
3	60000-80000	1024	240	6597	7861	0.130263
4	80000 +	1045	352	8181	9578	0.108292

Observations:
small Business applicants have high cha
charged off.
renewable_energy where chanrged off p
as compare to other categories.

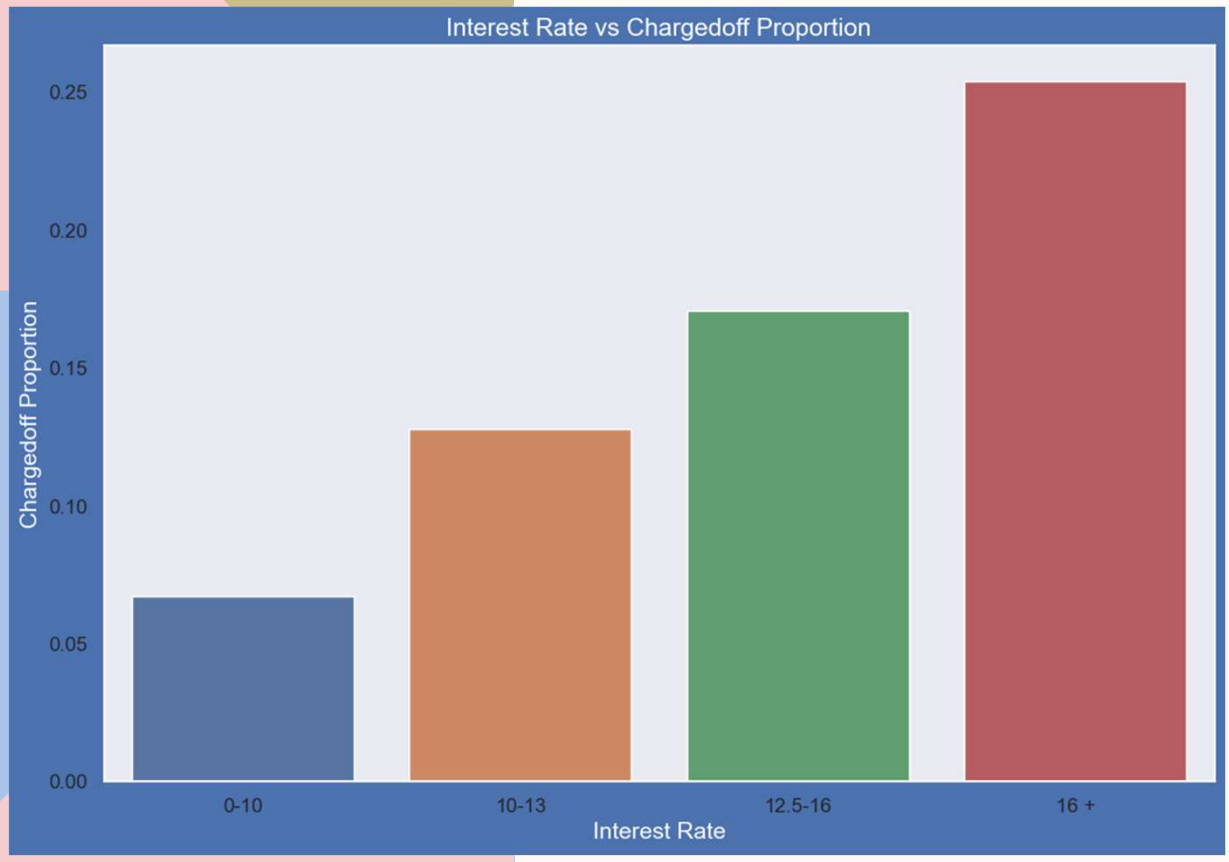
ANNUAL INCOME VS LOAN PURPOSE



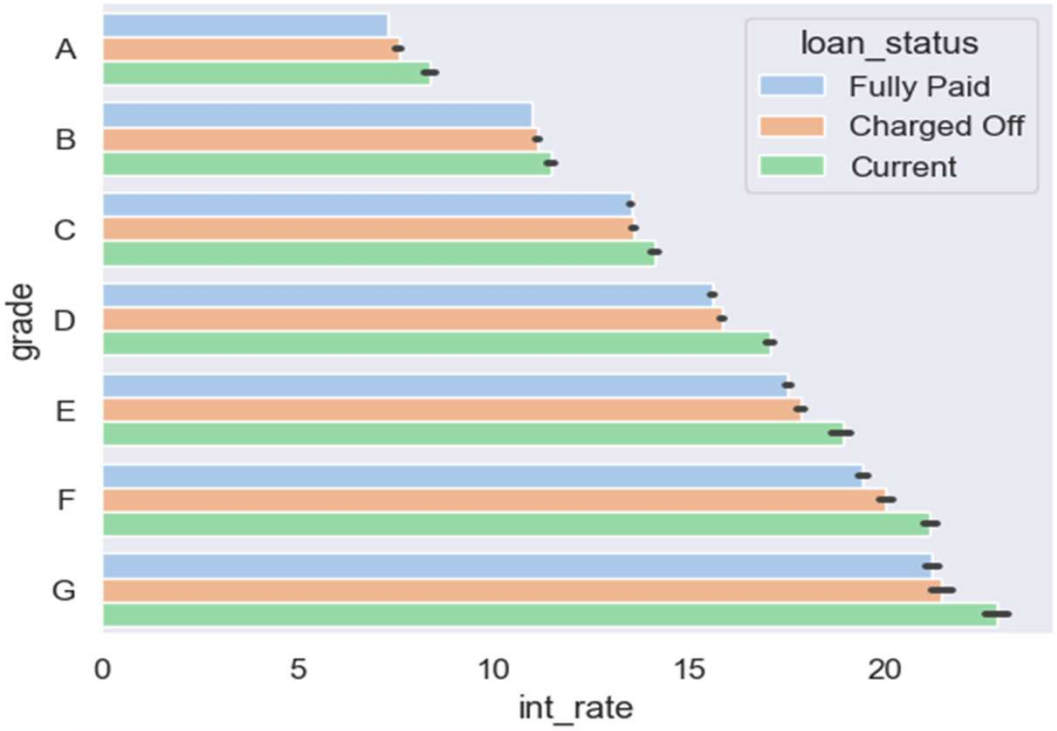
LOAN PURPOSE VS DEFAULT PROPORTION



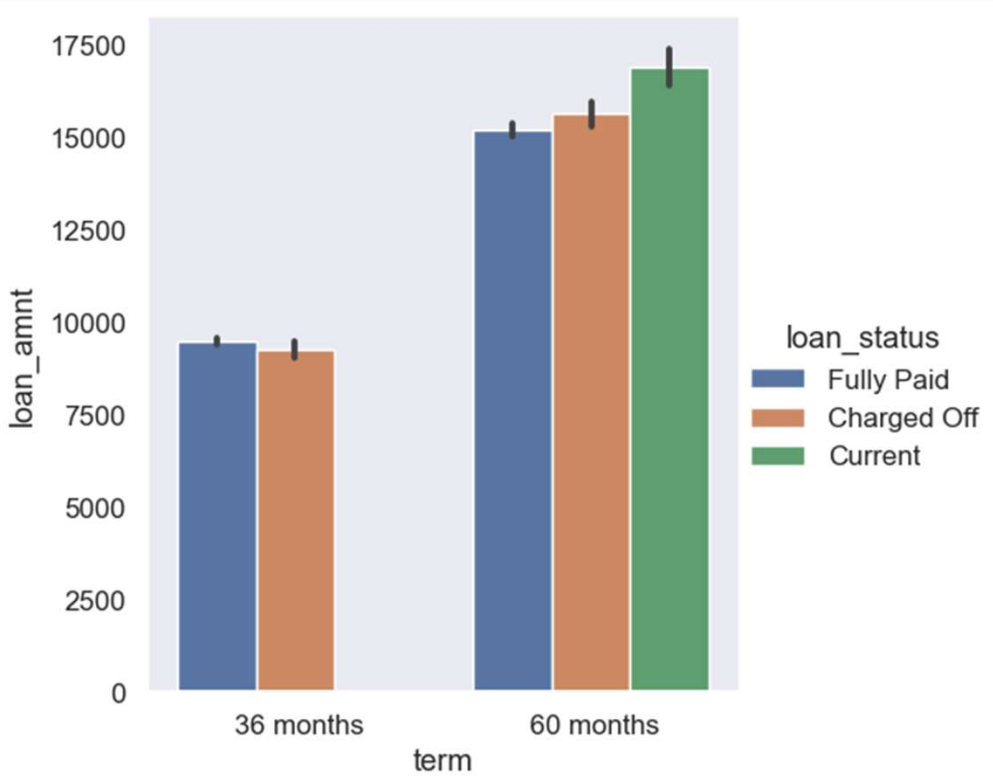
ANNUAL INCOME VS CHARGED OFF PROPORTION



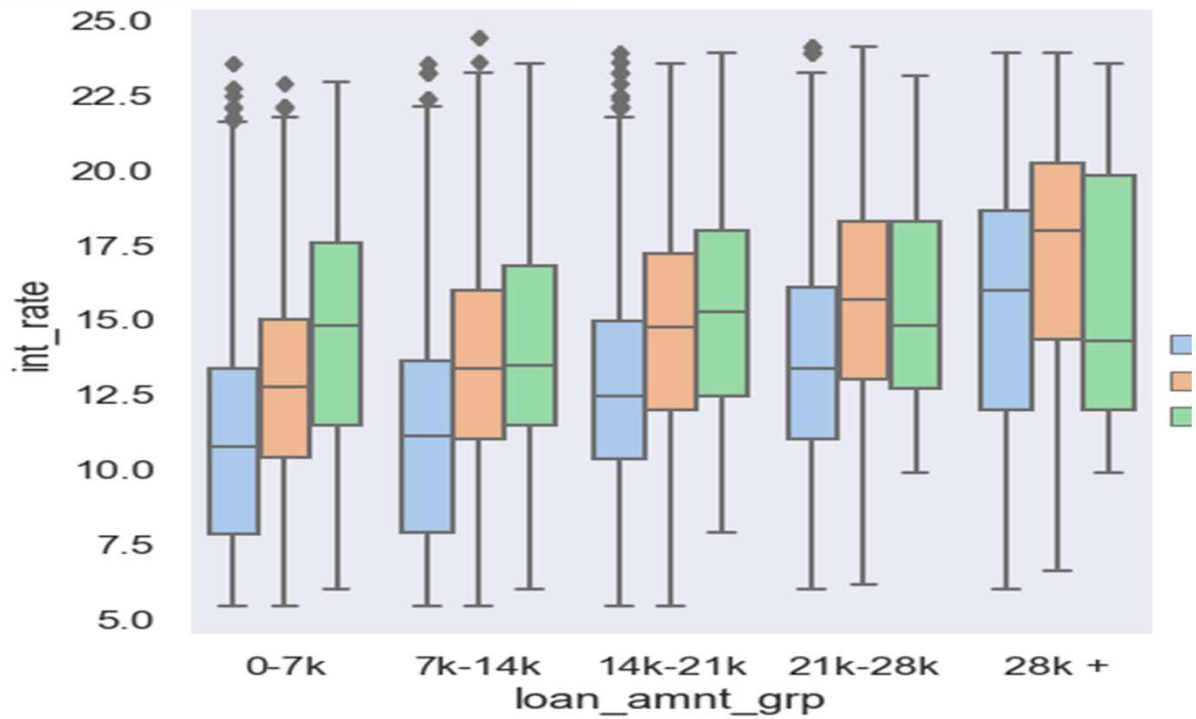
GRADE VS LOAN STATUS

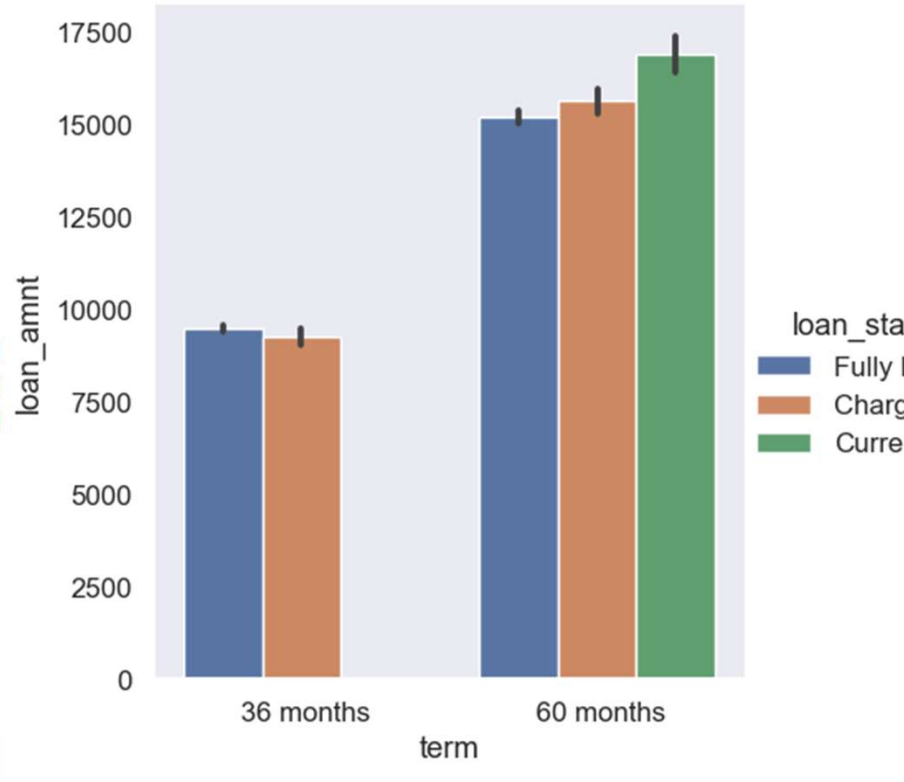


Duration vs Loan Status



Interest rate vs annual income vs status





OBSERVATIONS

The number of loans defaulted are the highest in the loan pupose"debt_consolation", But the annual income is not the highest.

The Loan is higher for all the income groups ,people who defaulted.

The interest rate for charged off loans is higher , compared to that of fully paid loans in all the groups. This is a driving factor for loan defaulting.

The loan was sanctioned/issued in December are maximum defaulted and Loans issued in the year 2011 are also higher comparing other years

Employees with highest tenure in an organization got higher Loan amount.

RECOMMENDATIONS

Modest interest for all types of loan would help in minimizing load default rate

Higher income group and lower income group , moderate loan amount would help in reducing defaulters

The number of loans defaulted are the highest in the loan pupose"debt_consolation", number of loans, loan amount can be reduced

December month loans seems seasonal with high default rate, thus can pushed to a month before or later , gives time to employees think and apply for loans

Employees with highest tenure likely to repay the loan successfully

THANK YOU

Elayaraja Muthuraj
Elaya.m@gmail.com
Elaya1984 (github.com)