LendingClub - EXPLORATORY DATA ANALYSIS

RESULTS AND OBSERVATIONS BY VIKAS GARG & CHACKO MULACKEL

AGENDA

- OVERVIEW
- EDA ROADMAP
- ANALYSIS
- RECOMMENDATIONS

OVERVIEW

- Lending Club is a peer to peer lending company based in the United States, in which investors provide funds for potential borrowers and investors earn a profit depending on the risk they take (the borrowers credit score). Lending Club provides the "bridge" between investors and borrowers.

LENDING CLUB

Wants	Fears	Needs
Be Profitable	Taking Unwanted Risks	EDA On Existing Dataset
Understand Customers	Losing Investment	To Find Bad Loans
Give Good Loan		Critical Variables To Help Give Good Loans

OBJECTIVE

The company wants to understand the **driving factors** 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.

As a Data Scientist analyze the past dataset by using EDA to understand customer and loan patterns and their influence of not repaying loan on time or simply putting 'Defaulting'

EDA ROADMAP

Data Cleaning

1. Null Value

2 Removing unwanted Data

Univariate Analysis

1. Frequency

2. Distributions

(Categorical and Numerical)

Segmented Analysis

1. Derive New Variables

2. Create Plots

Bivariate Analysis

1. Understanding how combination of variables estimates good loan's

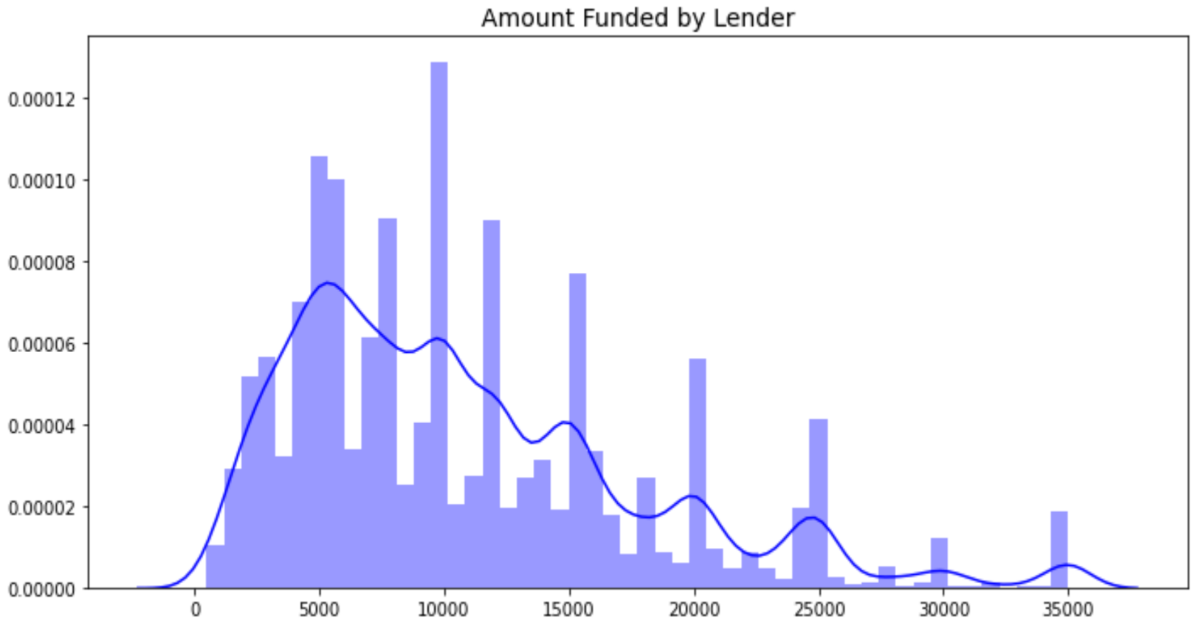
Finalize Variables/ SUMMARY

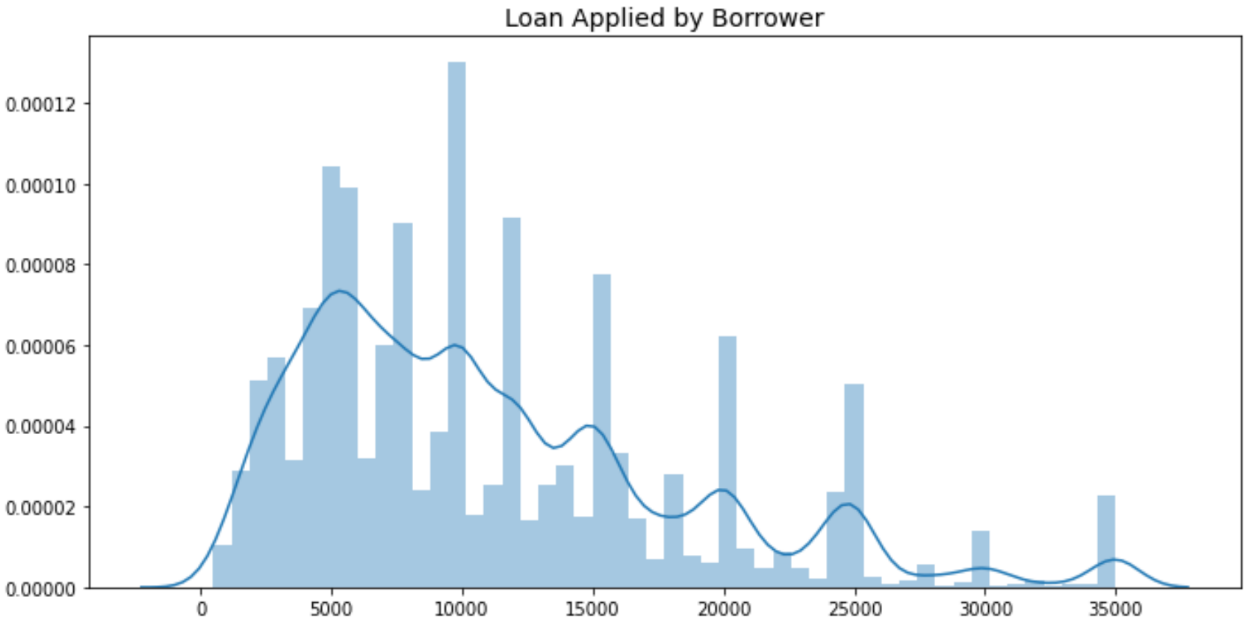
ANALYSIS

DATA CLEANING PROCEDURE

- Column Reduction – Removal of columns with more than 65% missing values
- Removing unimportant/ insignificant columns by intuition
- Data Imputation – Converting NA/ NAN to suitable values
- Data Correction – Removing bad data, e.g. Date columns
- Correction in data type
- Addition of newly derived columns

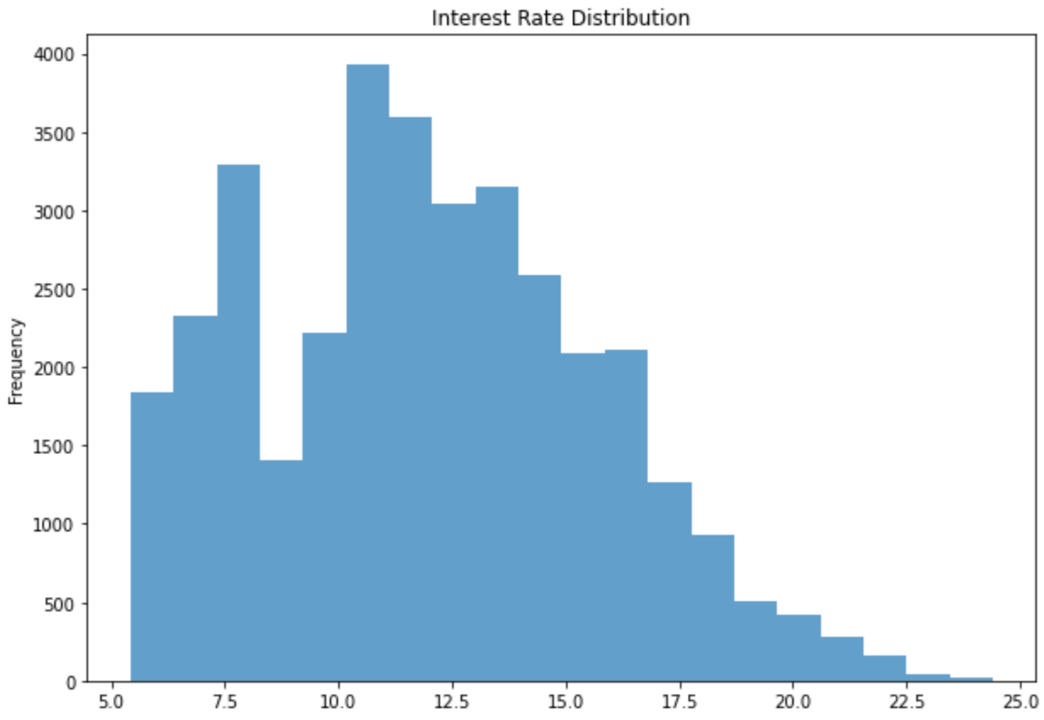
LOAN DISTRIBUTION FREQUENCY

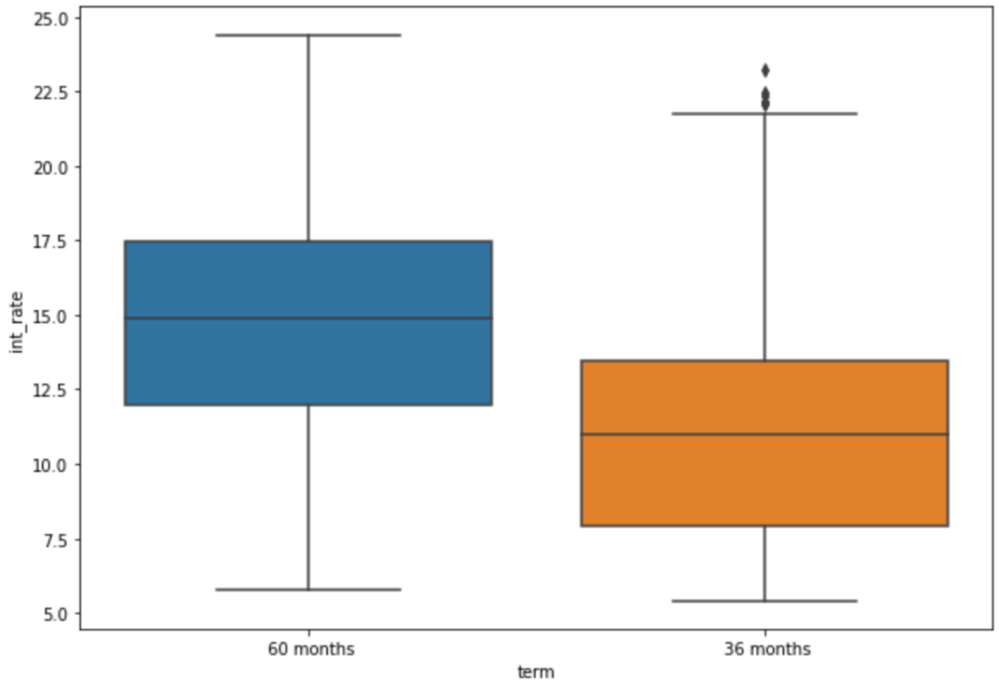




- There are big difference while comparing the applied amount and the funded amount.
- The year 2011 was having the highest amount of loan that got distributed.

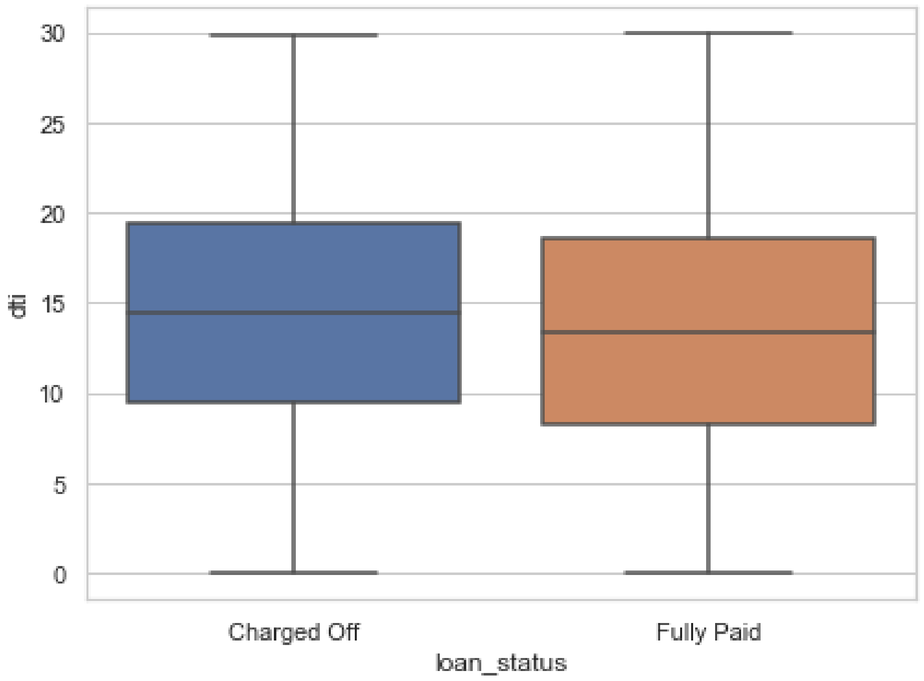
INTEREST RATE DISTRIBUTION

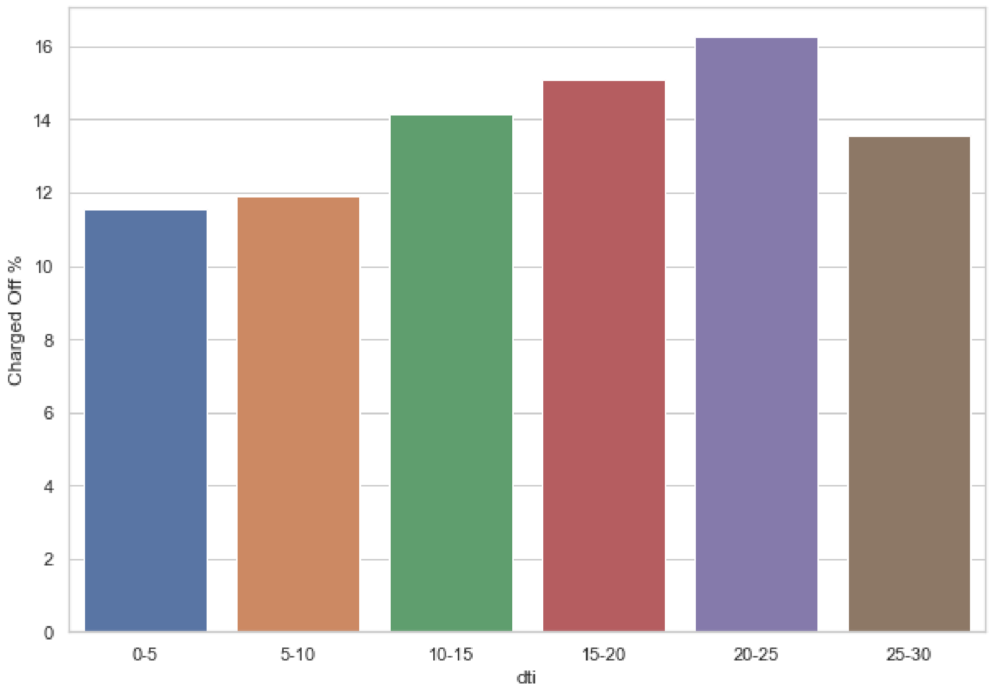




- Interest rate vary from 5.42% to 24.59%
- High interest rates are associated with longer period loans

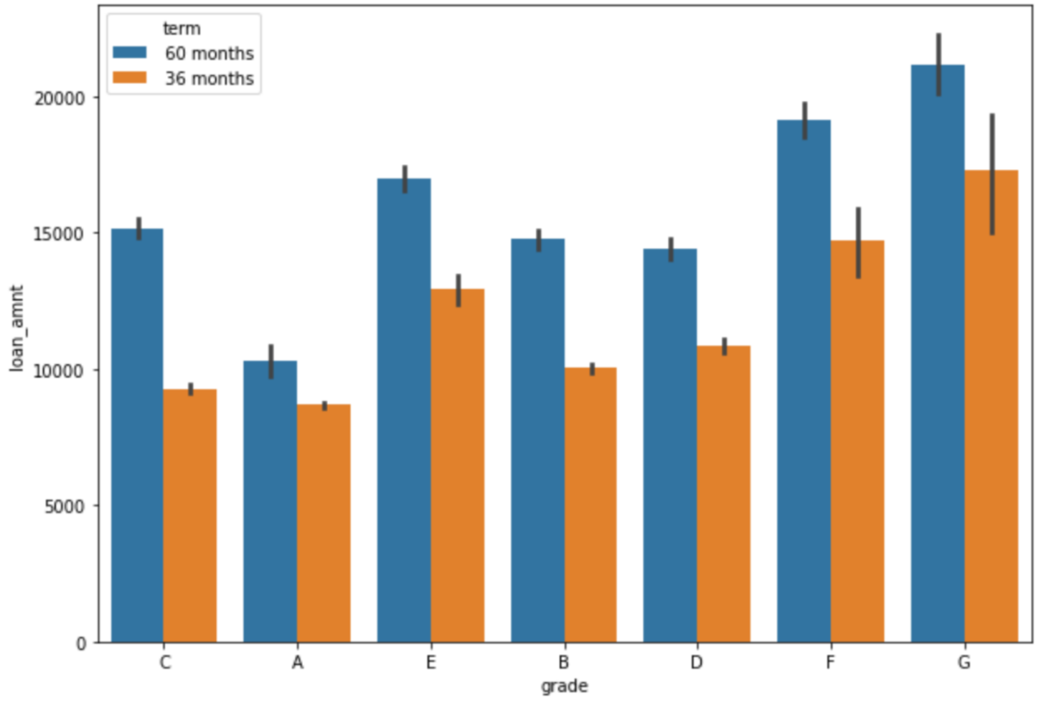
CHARGED OFF CUSTOMERS VS DEBT TO INCOME RATIO

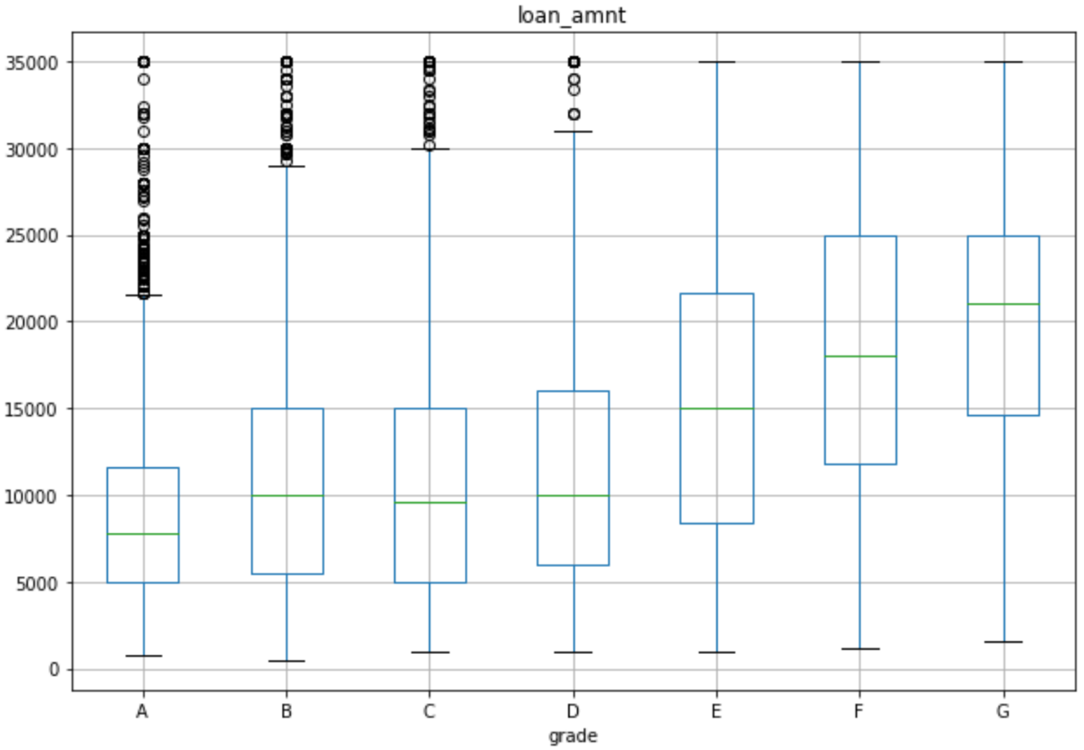




- Charged off customers have comparatively high debt to income ratio.
- Charged off customers are increasing slightly with respect to dti ratio so giving loan to customers with high dti will be more risky

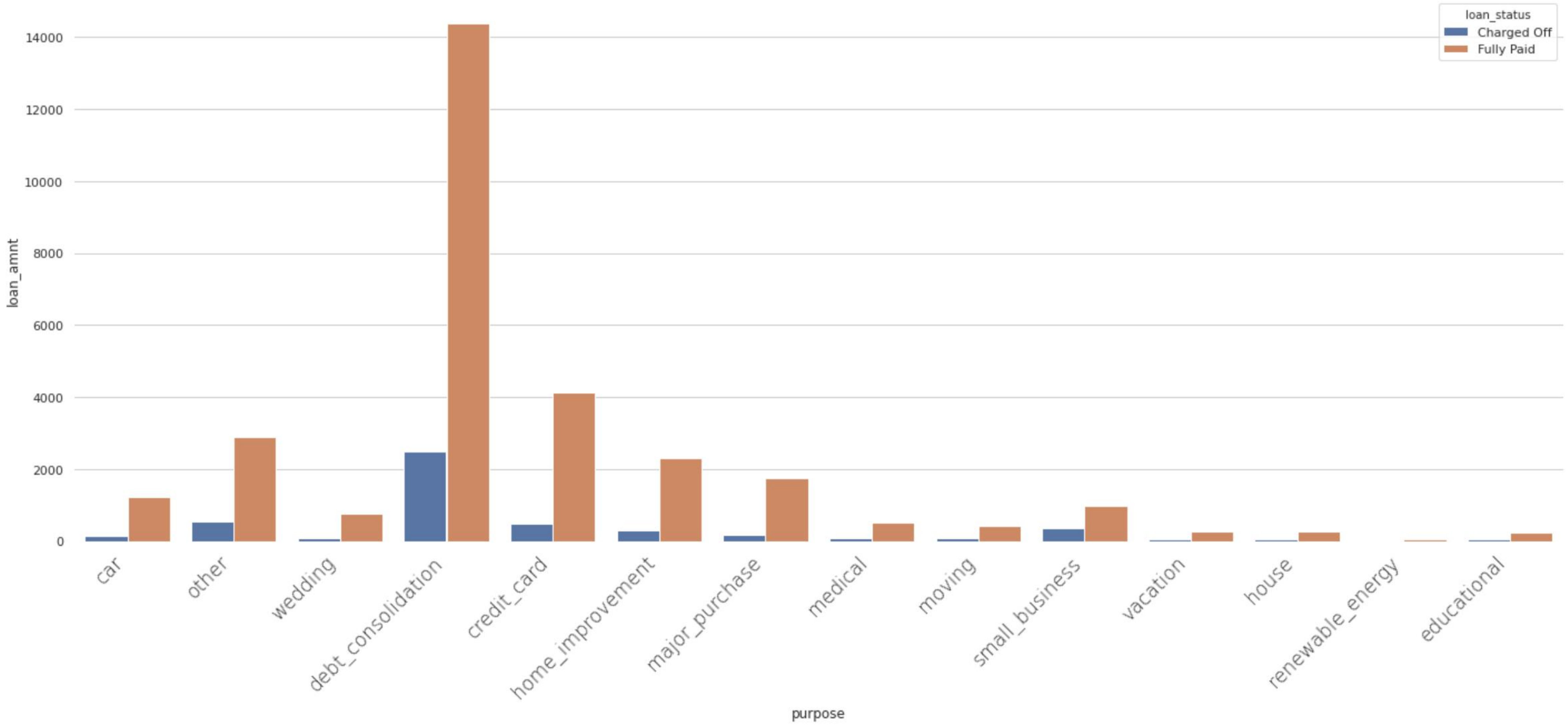
RELATION OF INTEREST RATE VS GRADES





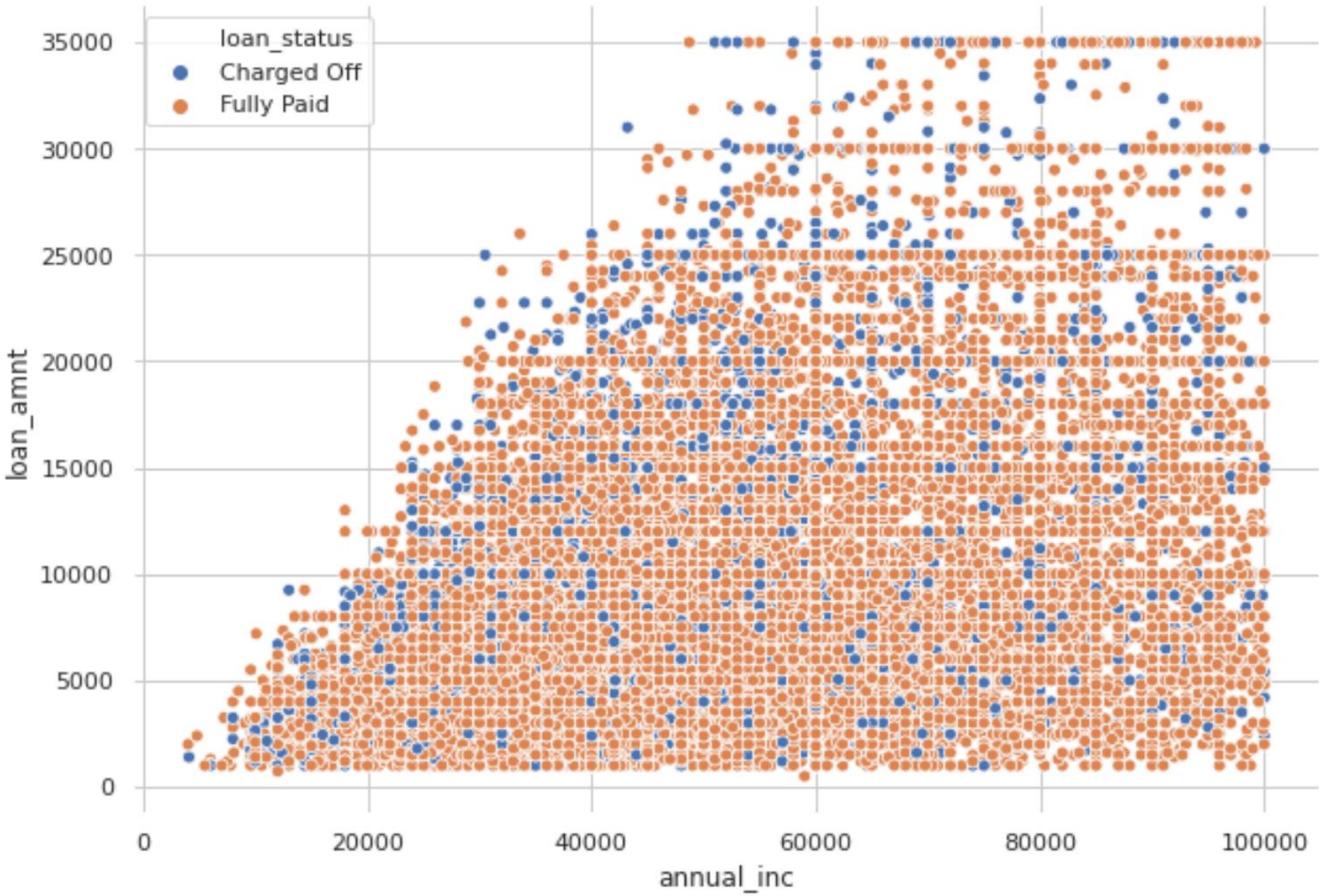
- Interest rate vary from 5.42% to 24.59%
- The median loan amount for a grade G is higher than that of A,B or C grade types
- Lower the grades higher the interest rate.

LOAN AMOUNT VS PURPOSE



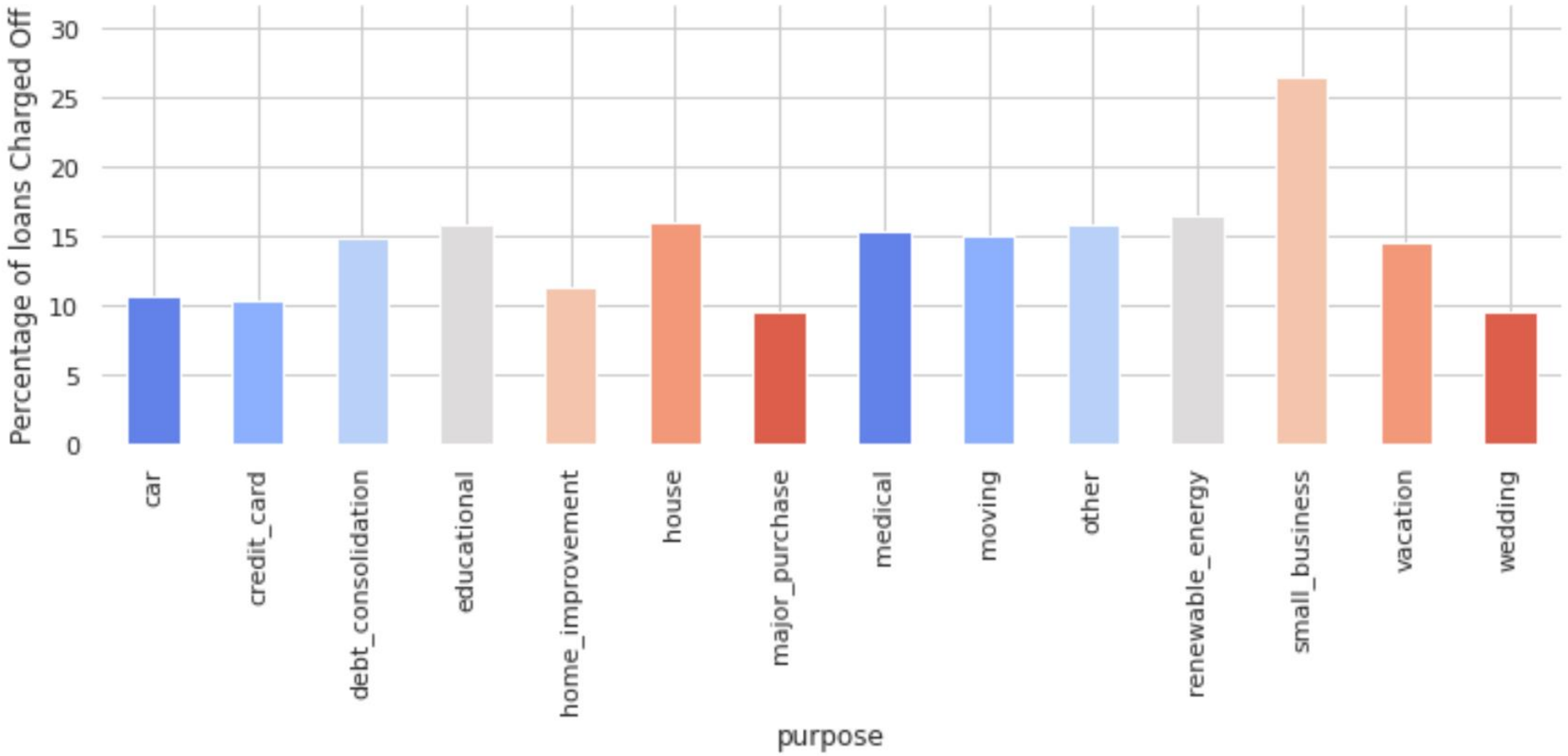
- Majority of the loans where taken for debt consolidation

ANNUAL INCOME OF DEFAULTERS AND GOOD LOANS



- A large number of low income candidates where given bigger loan

ANNUAL INCOME OF DEFAULTERS AND GOOD LOANS



- As you can see Small Business shows trends of defaulting

RECOMMENDATION BASED ON FINDINGS

DTI

Approving loans where debt to income ratio higher than 30% will be risky and has to be diligent on such applicants.

SMALL BUSINESS

Small business shows the highest number of defaulters. The company have to hand pick applicants in this category

PRIOR RECORD

Giving loans to prior defaulters will be risky. Stringent background checks has to be done for such candidates

INTEREST RATES

Higher Interest rates should be imposed for candidates from small business and higher DTI

THANK YOU

VIKAS GARG & CHACKO MULACKEL