LENDING CLUB ASSIGNMENT

SUBMISSION

Names: Aruna Sri Kurapati

Vidya Latha Pyla

Lending Club Problem Statement

Lending Club specializes in lending various types of loans to urban customers.

From the given loan data set we need to do Exploratory Data Analysis(EDA) and identify parameters of a customer which plays vital role in deciding to approve or reject a loan.

So the parameters identified after EDA should help the lending club to take decision for the below major risks

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Given Data set has a loan status parameter is having values like “Charged Off”, ”Fully Paid” and “Current which is considered as the target variable to decide whether a customer pays the loan or not.

Lending Club EDA Strategy

Data Cleaning Steps

Step 1: Initially we started analysing size of the data, type of various columns, missing data, description of the numeric columns.

Step 2: We have removed all the columns which has more than 90% of null values

Step 3: We have observed only one unique value for the column like payment plan, initial list status etc.. Removed all the columns having single unique which will not have any impact to decide whether the customer pays the loan or not

Step 4: Removing the unnecessary characters in integer columns and converting the columns to proper data type.

Lending Club EDA Strategy

Univariate Analysis – We have done the univariate analysis to understand the spread of different continuous and categorical variables.

Step 1: Analysed various categorical columns with the count plots.

Step 2: Analysed various continuous columns by describing them and using distribution plots.

Step 3: Segmented analysis of annual income, interest rate and debt to interest ratio(dti).

Bivariate Analysis – We have done Bivariate analysis with the target variable loan status to decide the important parameters for identifying charged off loans.

Step 1: Created a Correlation plot to understand the correlation between numeric columns

Step 2: Analysis of different continuous & categorical columns with the loan status.

Step 3: Calculated the percentage of Charged off loans based on data driven columns.

Lending Club – Conclusions

Below are the list of major parameters identified by us after EDA which helps in deciding if the customer can pay loan or not.

- **Interest Rate** – Higher the interest rates the chances are more for the customer not paying loan
- **Purpose** – 25% of loans in "Small Business" and 15% of loans in "Debt Consolidation" are not paid by customers
- **Grade** – Customer who belong to grade "E,F,G" has high chances of not paying loan compared to other grades.
- **DTI(Debt to Income Ratio)** – We have observed that above 15 DTI has more risk in paying loan.
- **Annual Income** – Customers having annual income below 50000 has chances of not paying the loan

Contd...

Lending Club – Conclusions

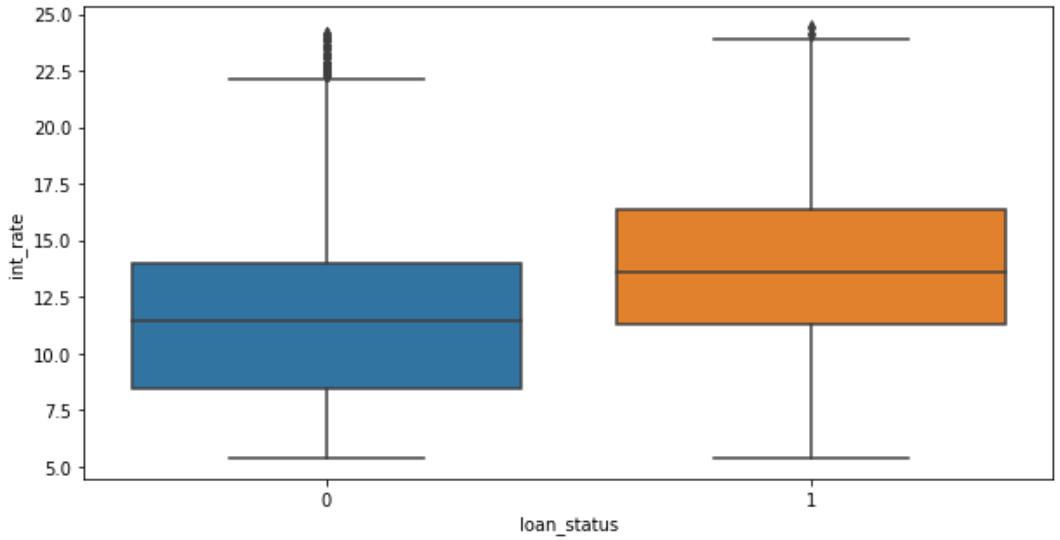
Below are the list of major parameters identified by us after EDA which helps in deciding if the customer can pay loan or not.

- **Home Ownership** – Customers who stay in Rented house has high risk in paying loans compared to others
- **Term** – Long term loans has high risk than short term loans
- **State** – Customers from Florida and California has high risk in paying loans compared to others
- **Inquiries** – Customers who make high number of inquiries has high risk in paying loans
- **Public Bankruptcies** – Customers having high number of bankruptcies has high risk in paying loans.

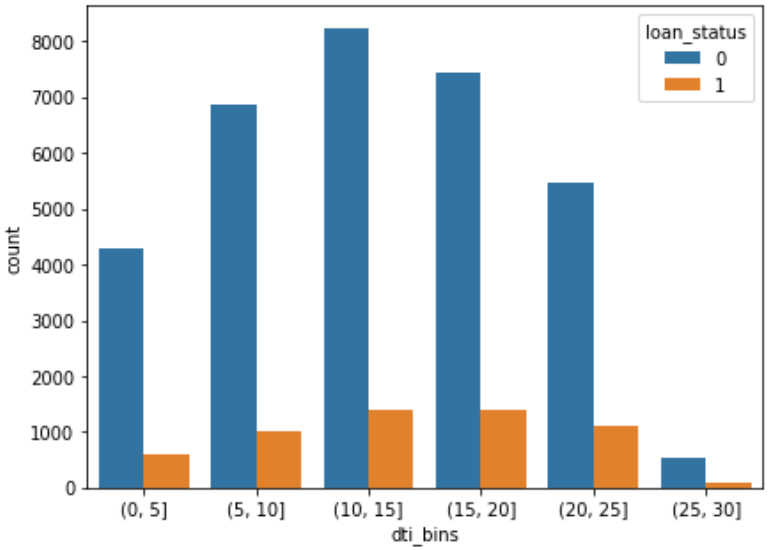
Contd...

Lending Club – Graphs

1. Interest Rate versus Loan Status



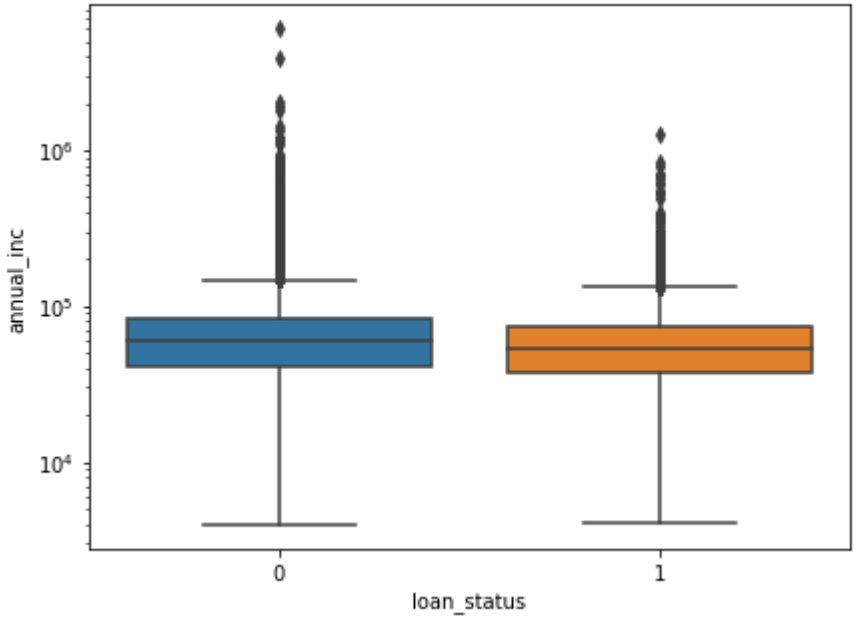
2. DTI versus Loan Status



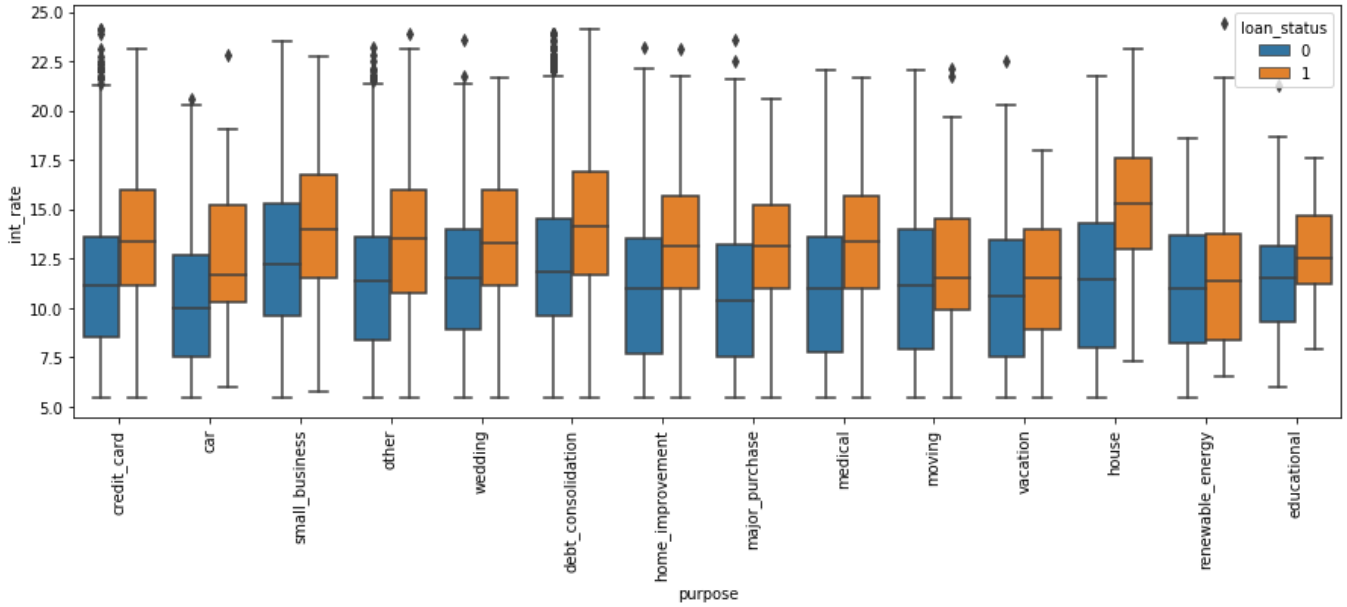
Contd…

Lending Club – Graphs

1. Annual Income versus Loan Status



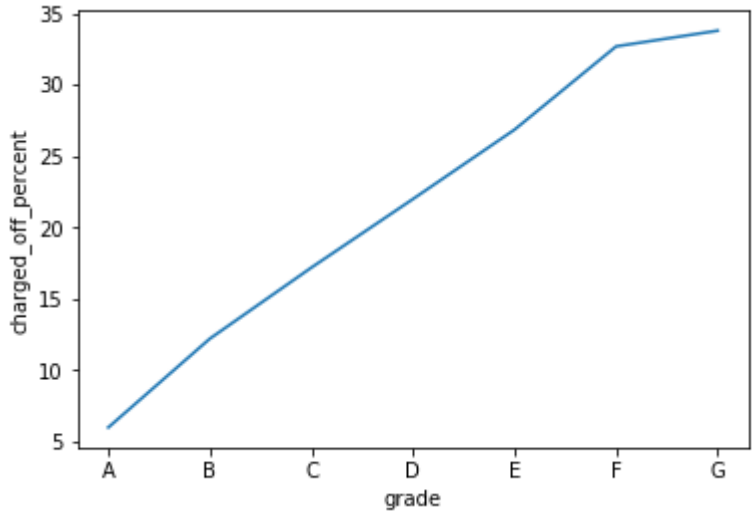
2. Purpose Versus Loan Status



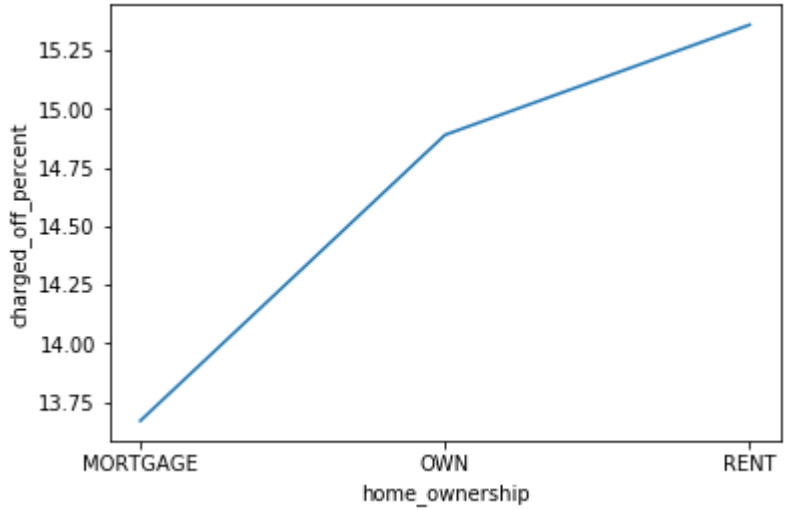
Contd…

Lending Club – Graphs

1. Grade versus Loan Status



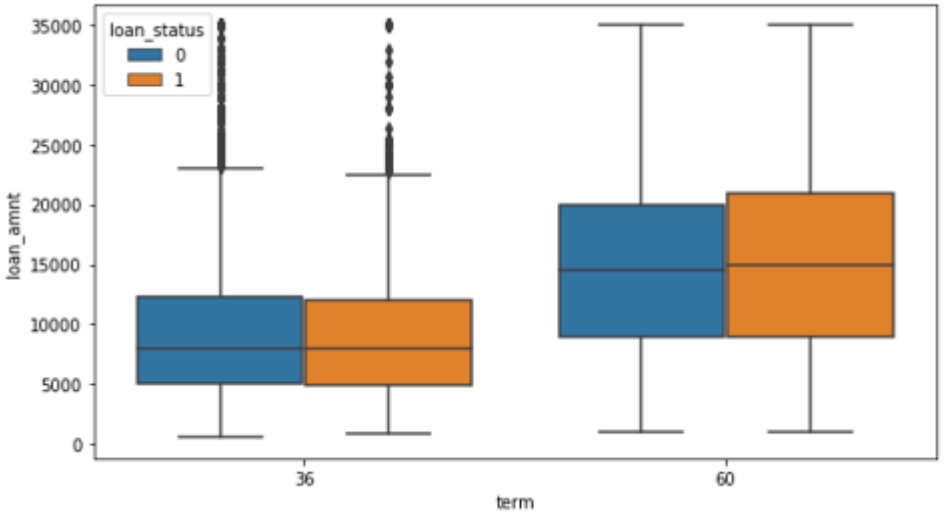
2. Home Ownership Versus Loan Status



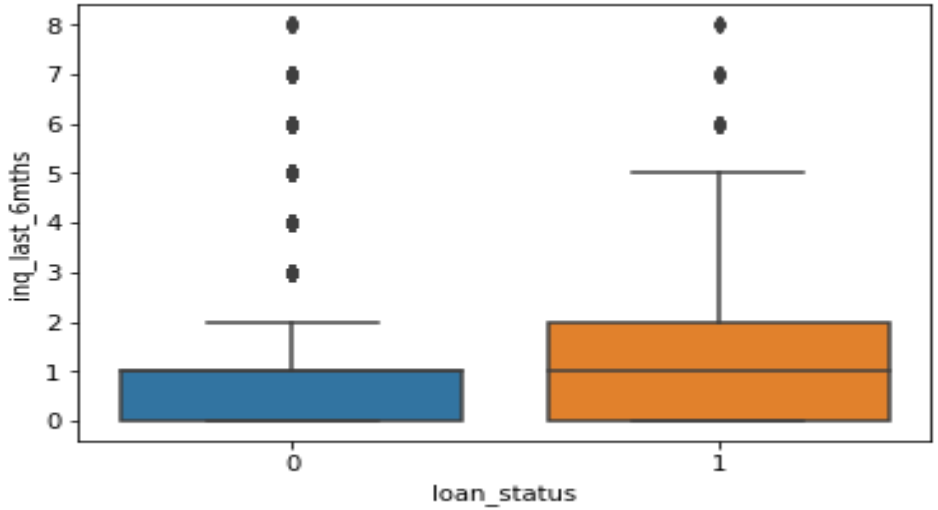
Contd…

Lending Club – Graphs

1. Term versus Loan Status



2. Enquiries Versus Loan Status



Contd…

Thank You