upGrad

Lending Club Case Study

Group Members
Rekha Kailas
Raghuvaran Gadhar
Komala M
Rajasekhar Challa

upGrad

Objective

Identification of risky applicants using EDA is the objective of this case study.

upGrad

Lending Club Analysis Overview

- Understanding the data
- Data Cleaning/ Processing
- Outlier Detection and Removal
- Univariate Analysis
- Bivariate Analysis

upGrad

Understanding the data

- Among the loan dataset, we had 111 columns and 39717 rows.
- More than 50% of column has 100% null values.
- We found columns with single unique value.
- Data included both numeric and non-numeric

upGrad

Data Cleaning / Processing

We'll clean the dataset and handle the missing data, data conversion and categorical variables.

- Data Removal

✓ Removed columns with null values having more than 30%

✓ Removed Rows with null values having more than 30%

- Data Conversion

✓ Converted all date column object to datetime type

✓ Casted all continuous variables to numeric so that we can find a correlation between them

upGrad

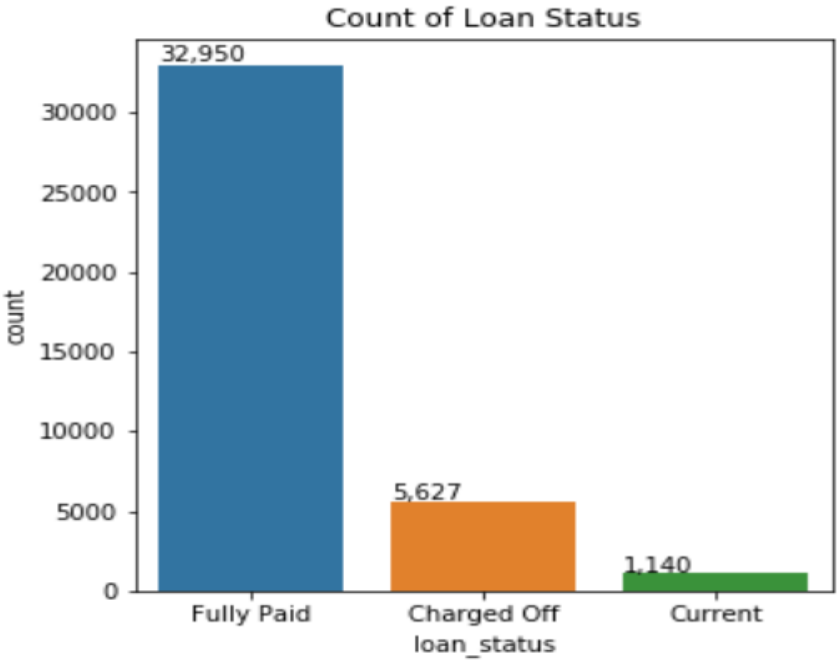
Outlier Detection and Removal

- Here, any value which out of range of 5th and 95th percentile are considered as outlier.
- For handling such outliers, we have **removed the values containing Outliers**.
- After removing the outliers, we have filled the missing values with the estimated one.
- In this case, we have gone with **Generalized Imputation** i.e., we calculate the median for all non missing values of that variable and replace missing values with median.

upGrad

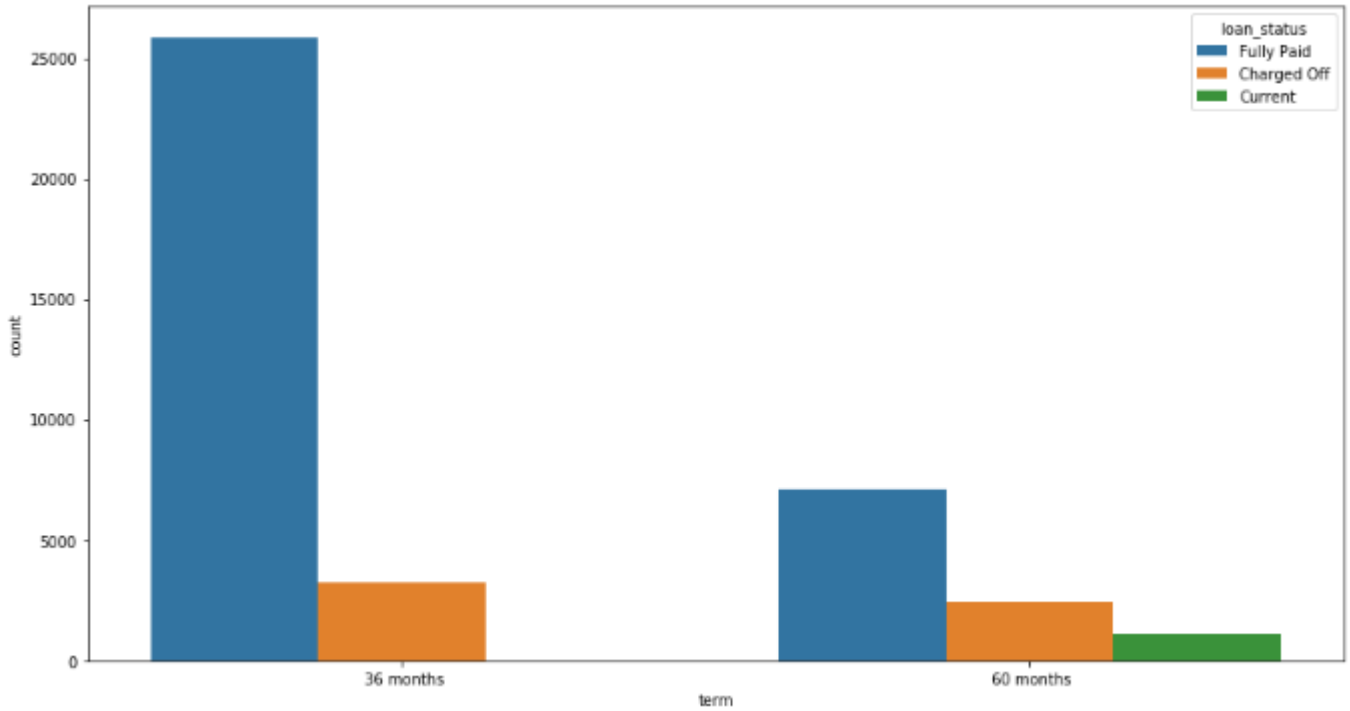
Univariate Analysis

Loan Status Count



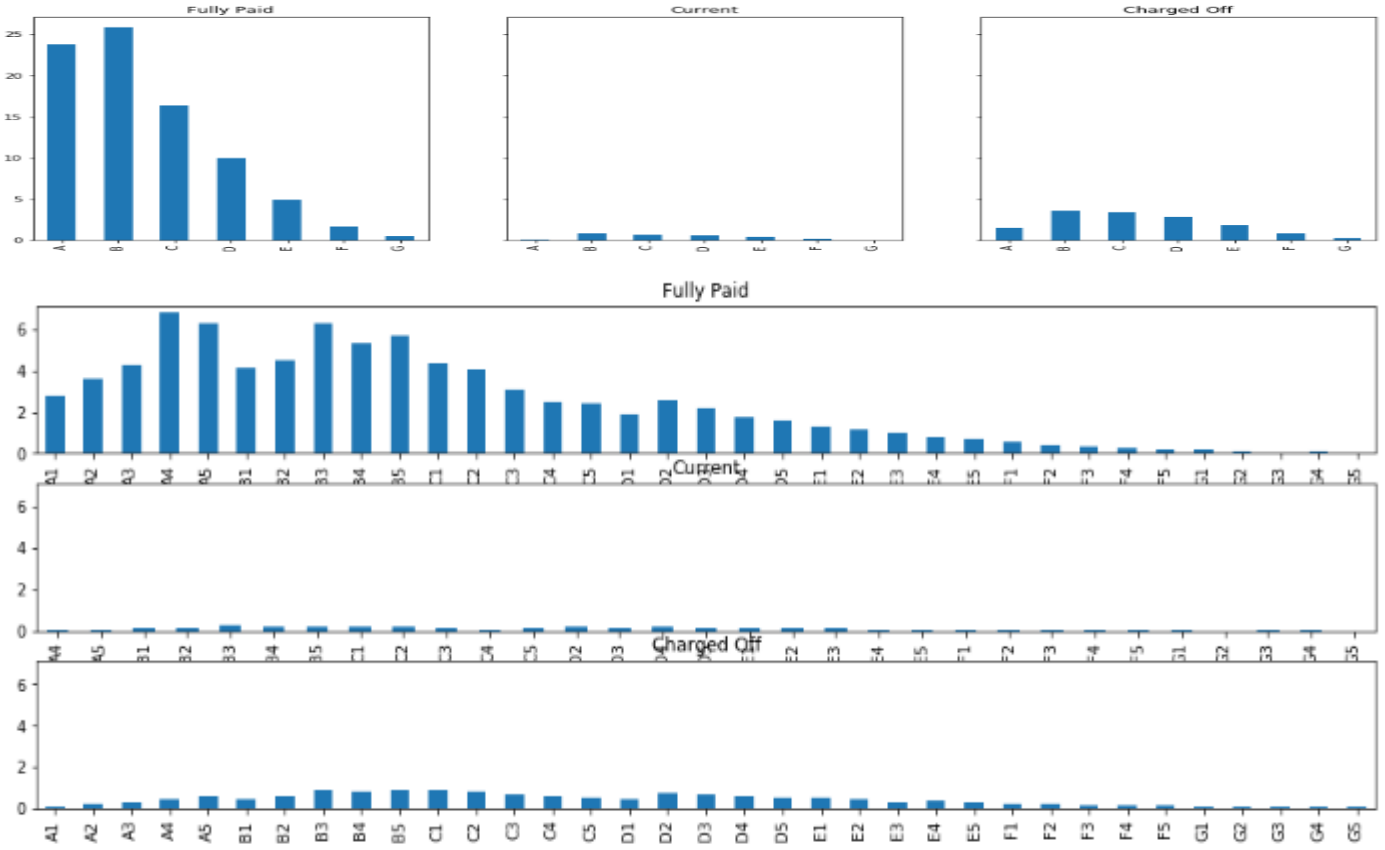
upGrad

Term(Tenure)



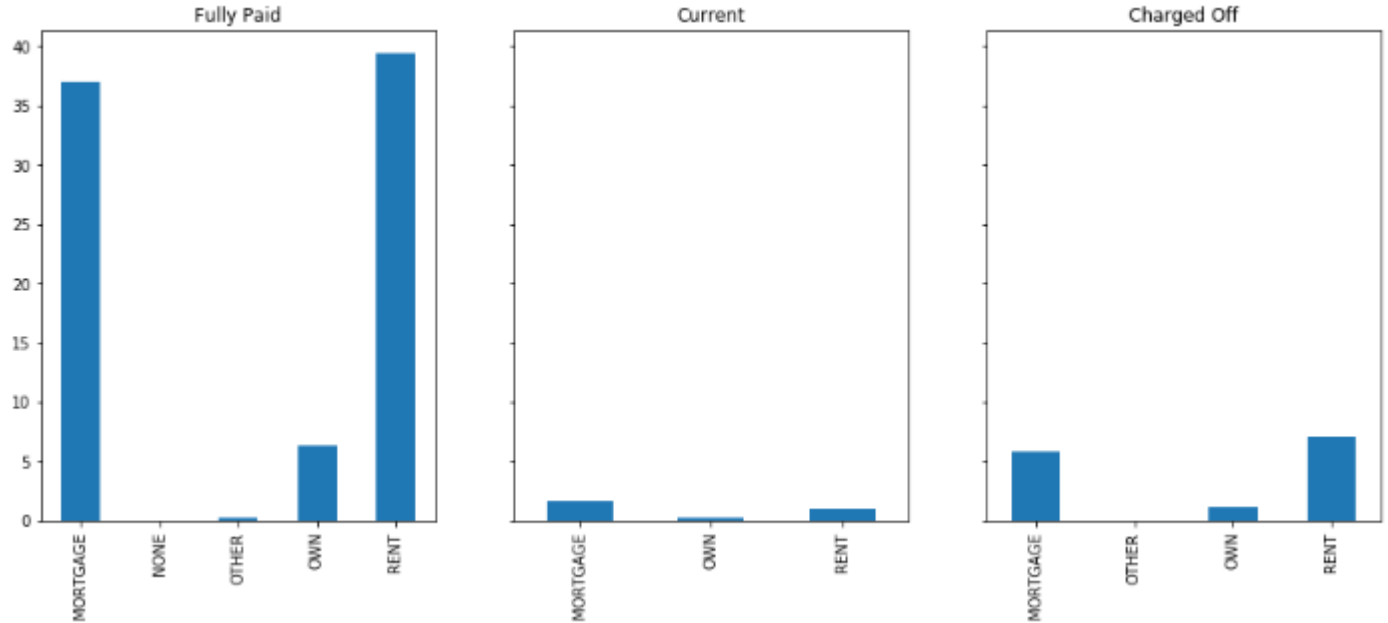
upGrad

Grade and its Sub Grade



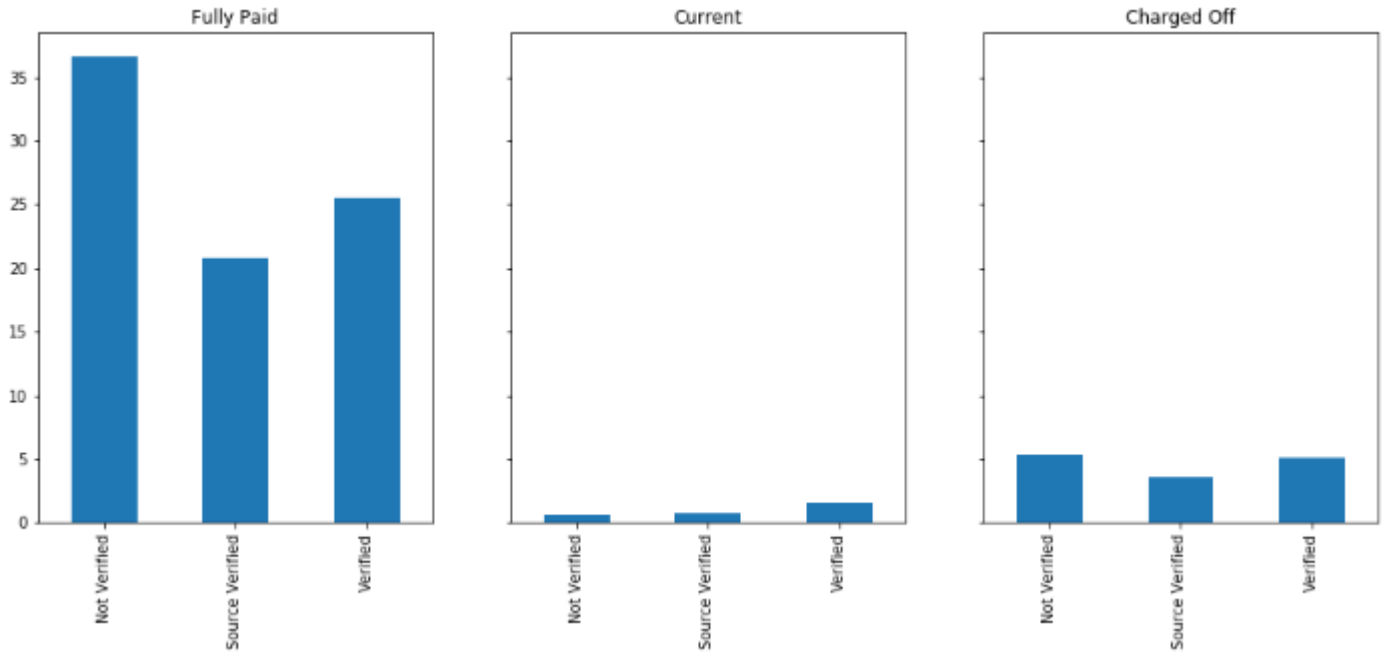
upGrad

Home Ownership

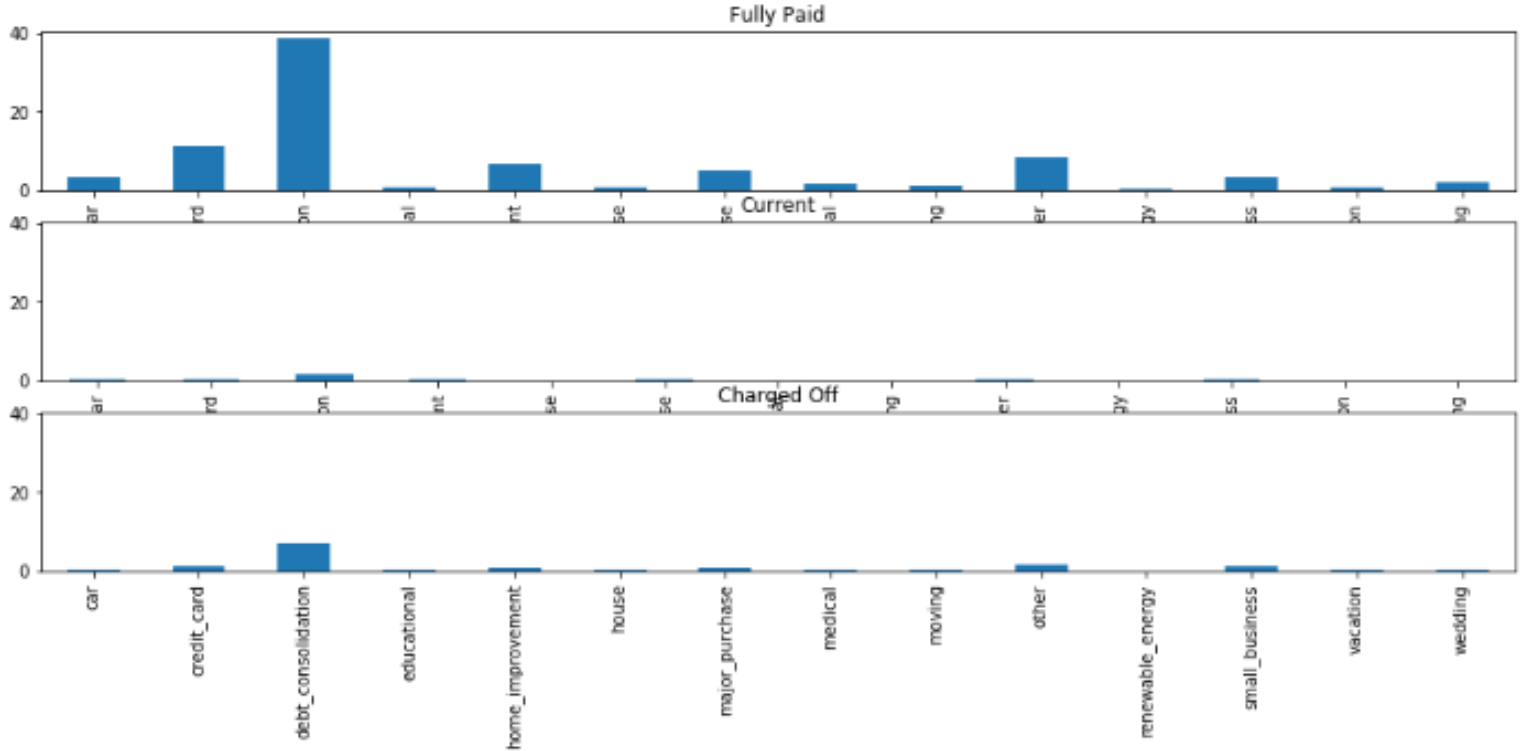


upGrad

Verification Status

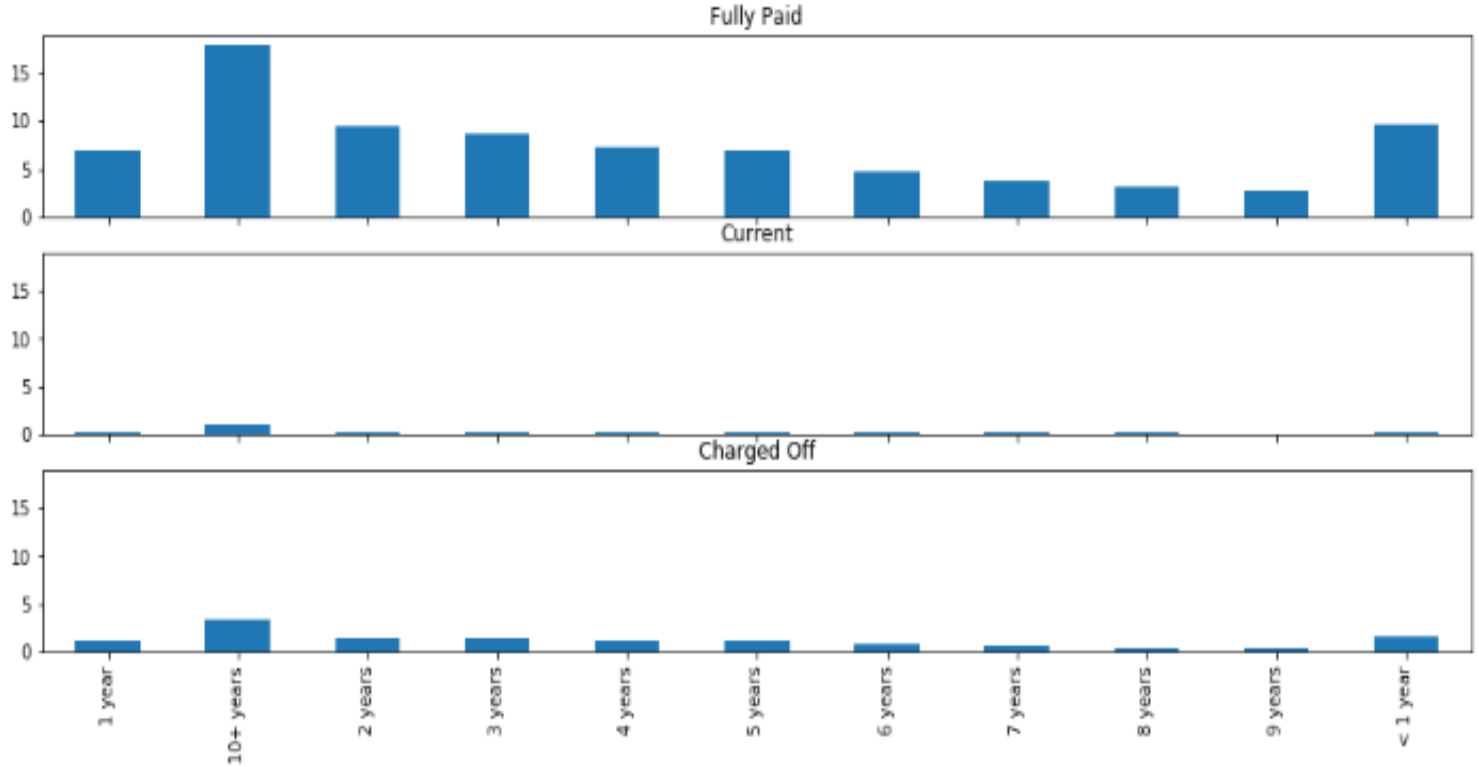


Loan Purpose



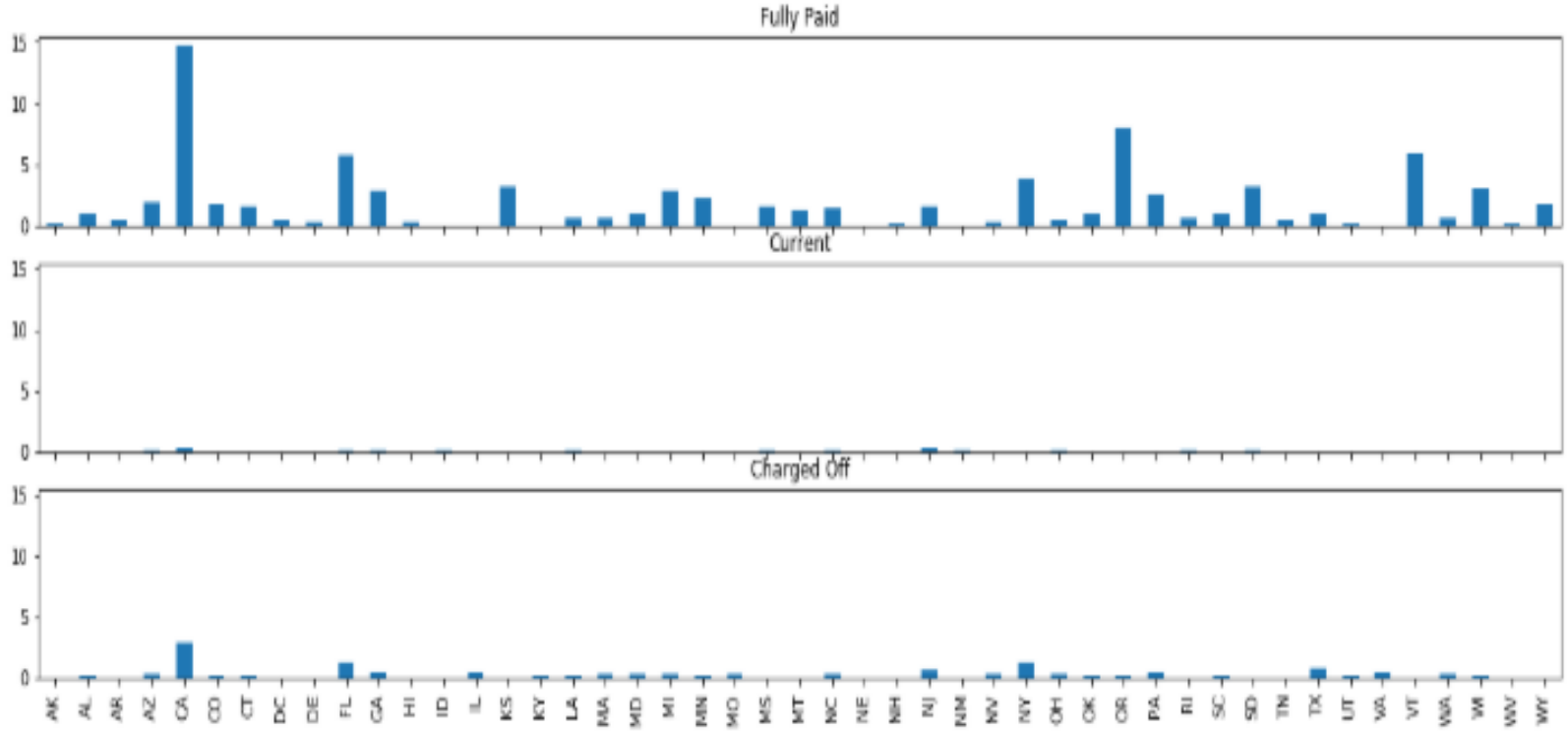
upGrad

Employment Length



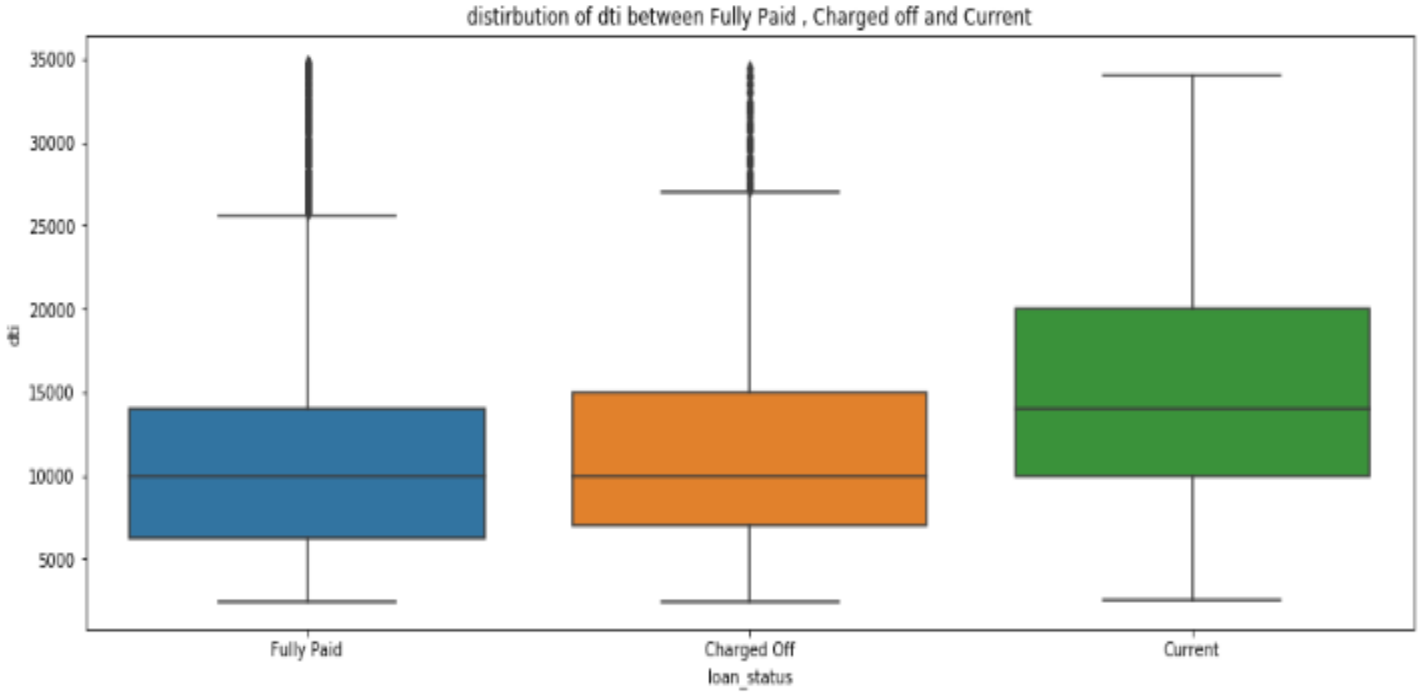
upGrad

Address State



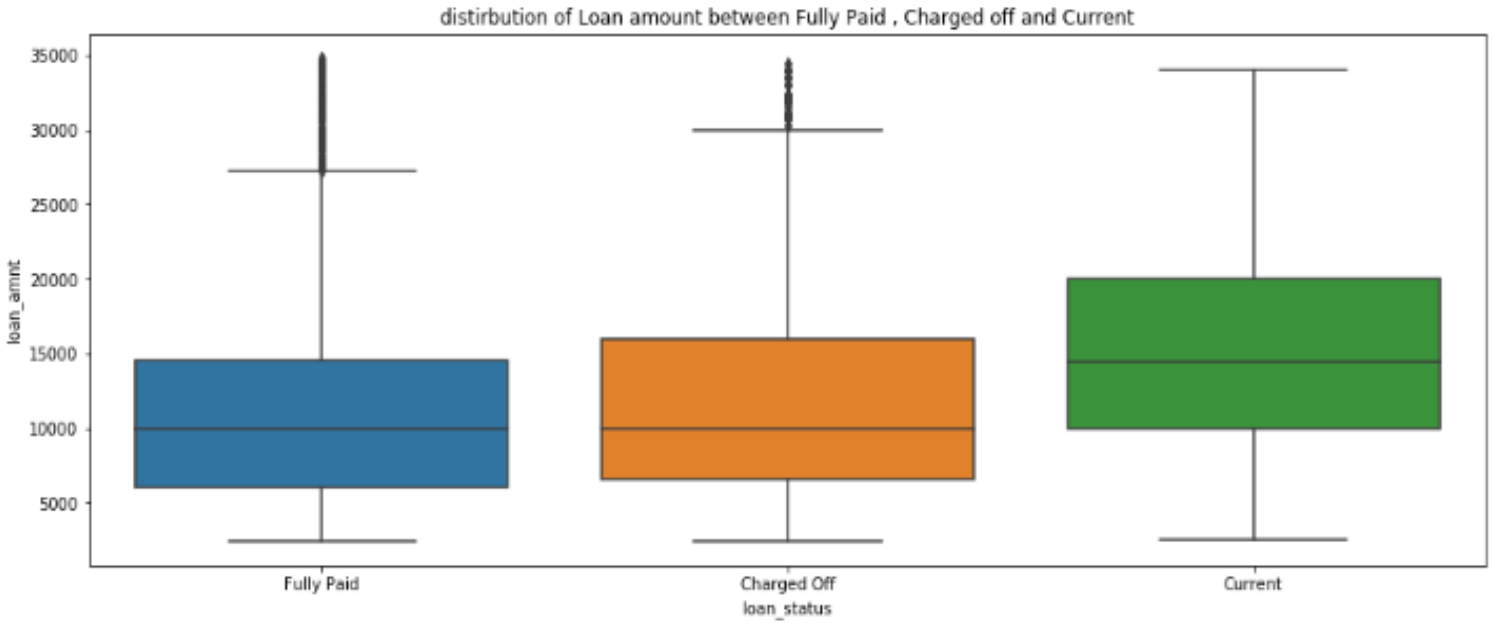
upGrad

Distribution of DTI



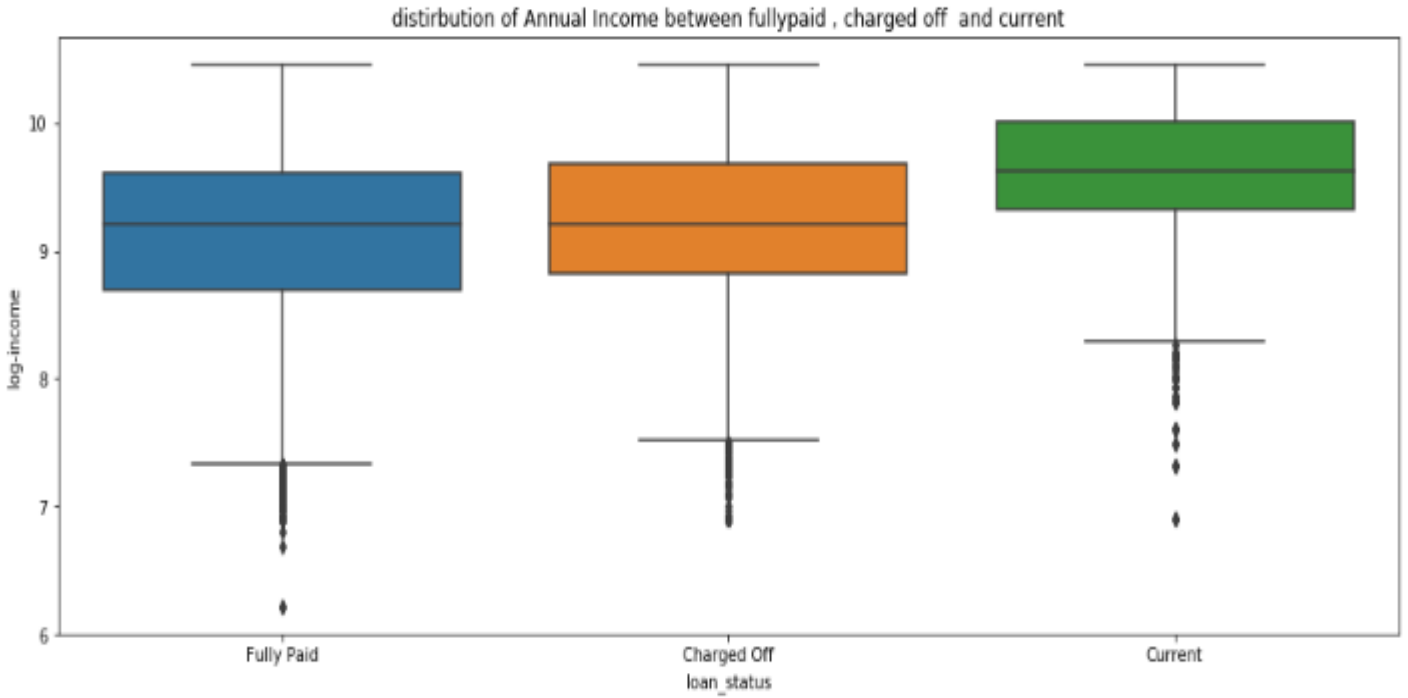
upGrad

Distribution of Loan Amount



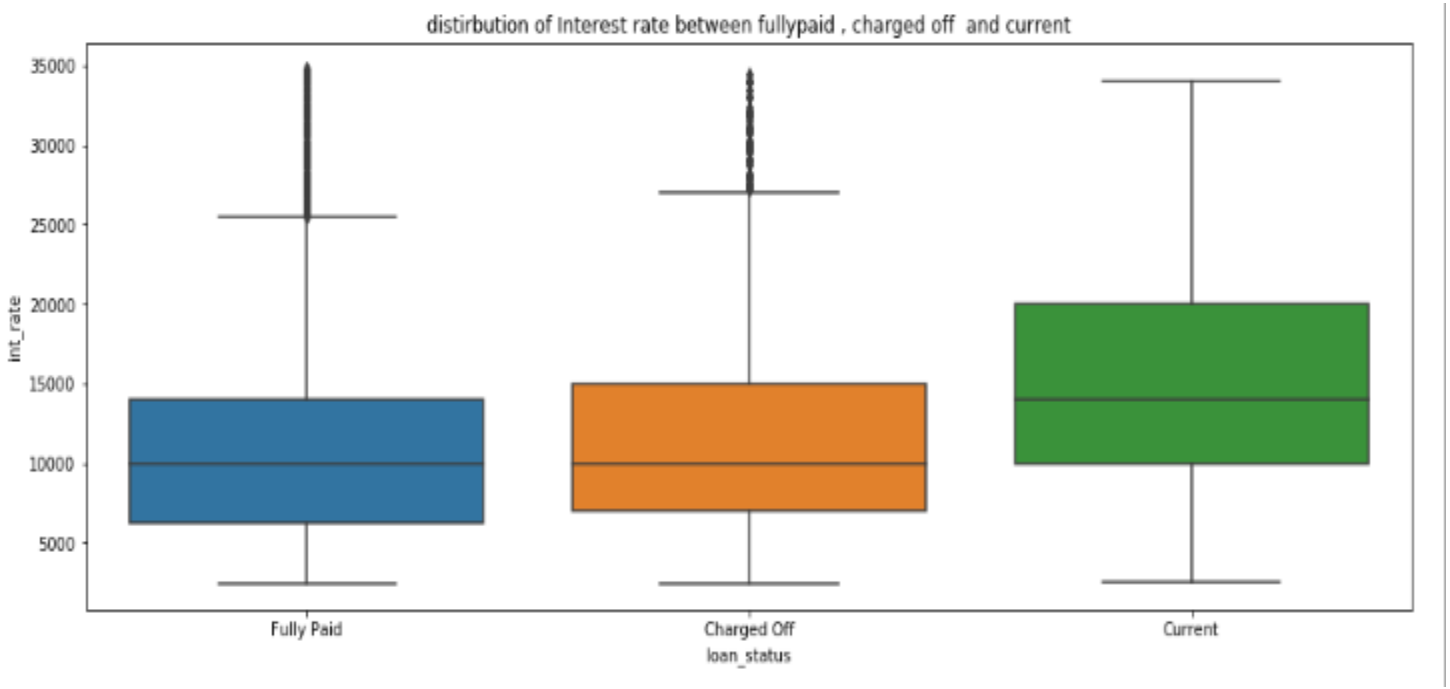
upGrad

Distribution of Annual Income



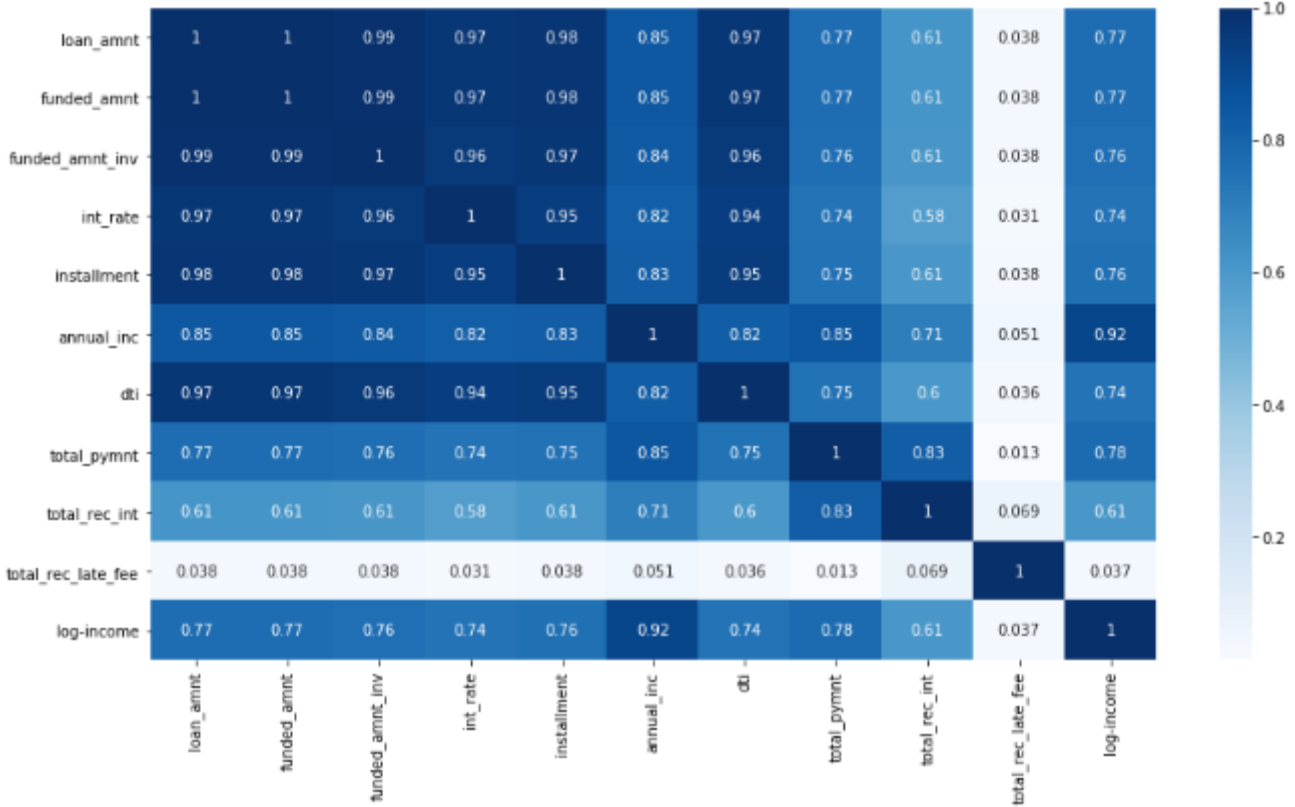
upGrad

Distribution of Interest Rate



upGrad

Heat Map with the continuous variables



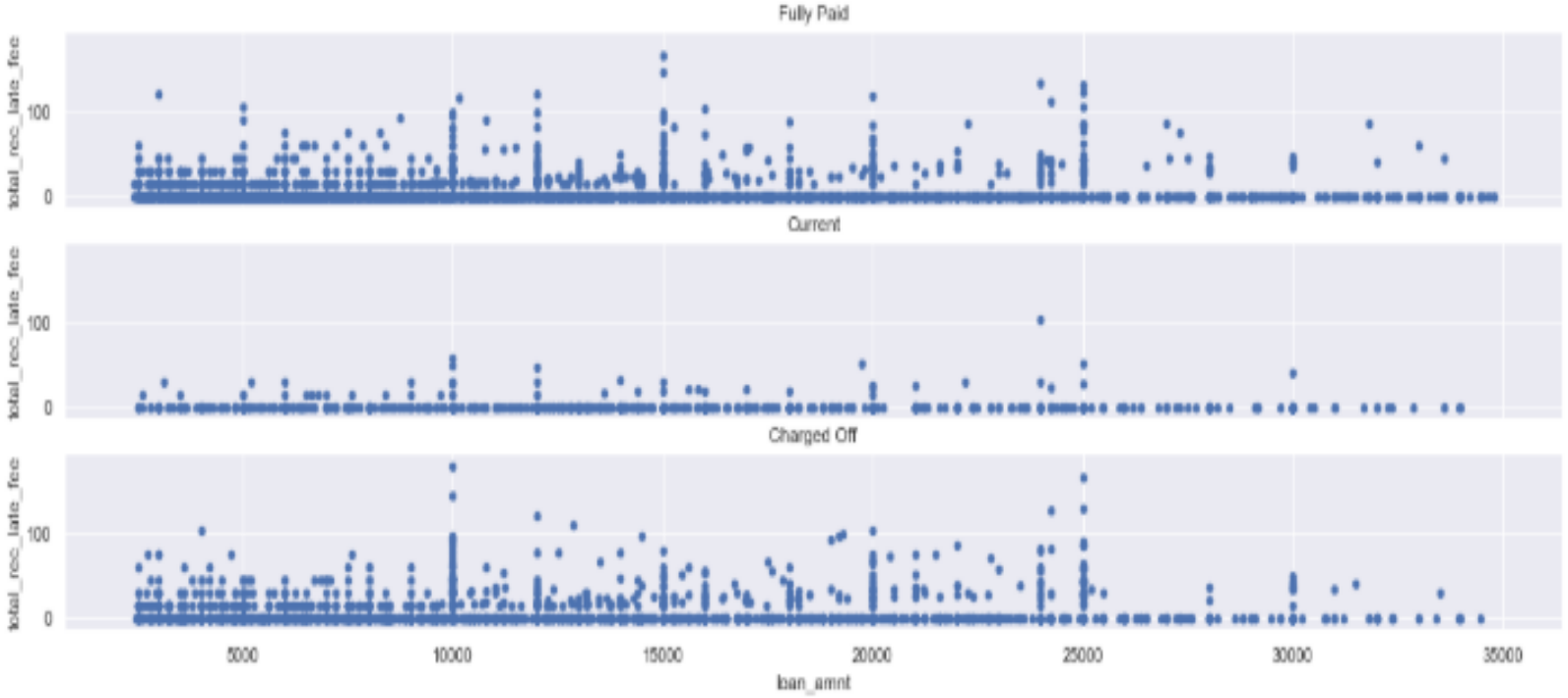
upGrad

Bivariate Analysis

Distribution of Interest Rate v/s Term

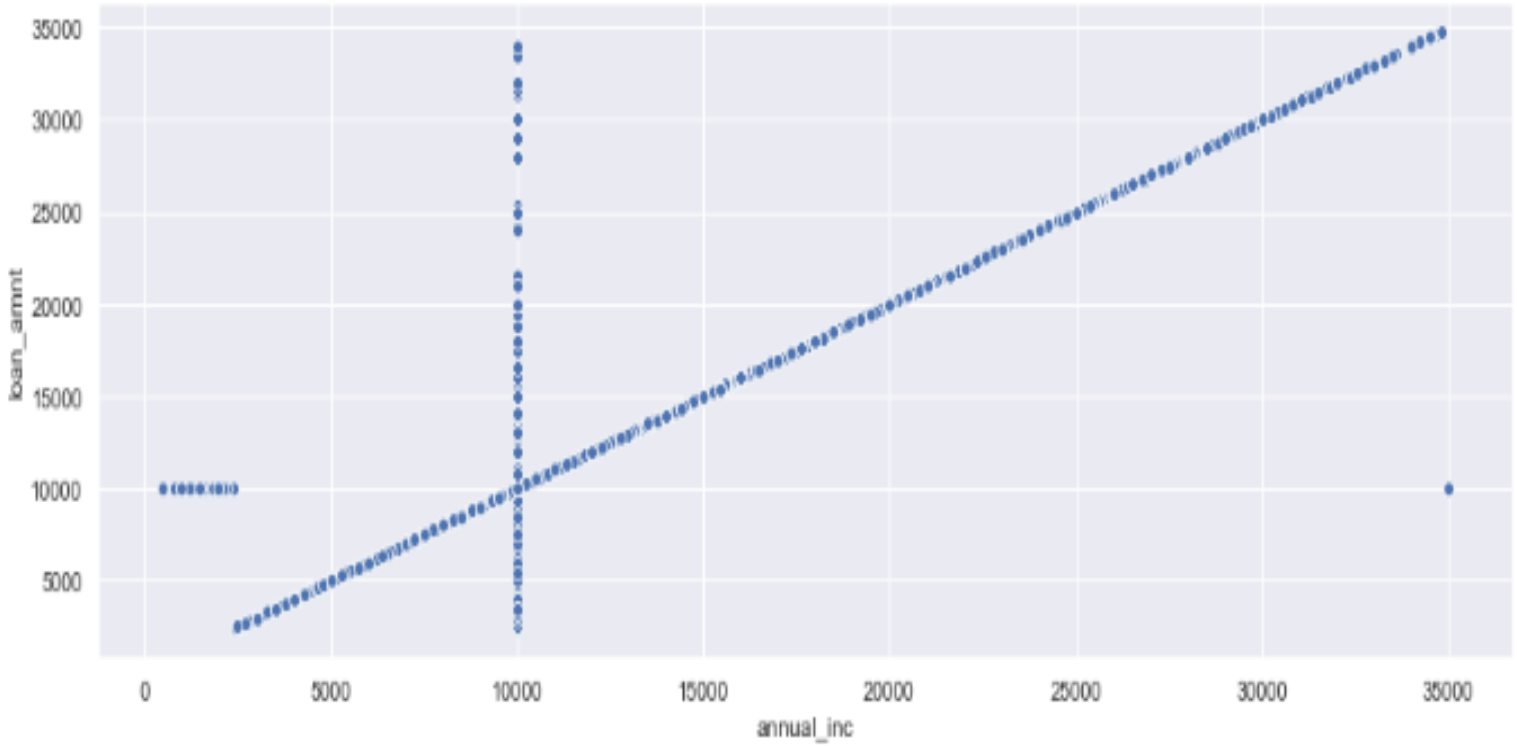
upGrad

Distribution of loan amount v/s total recovery late fee



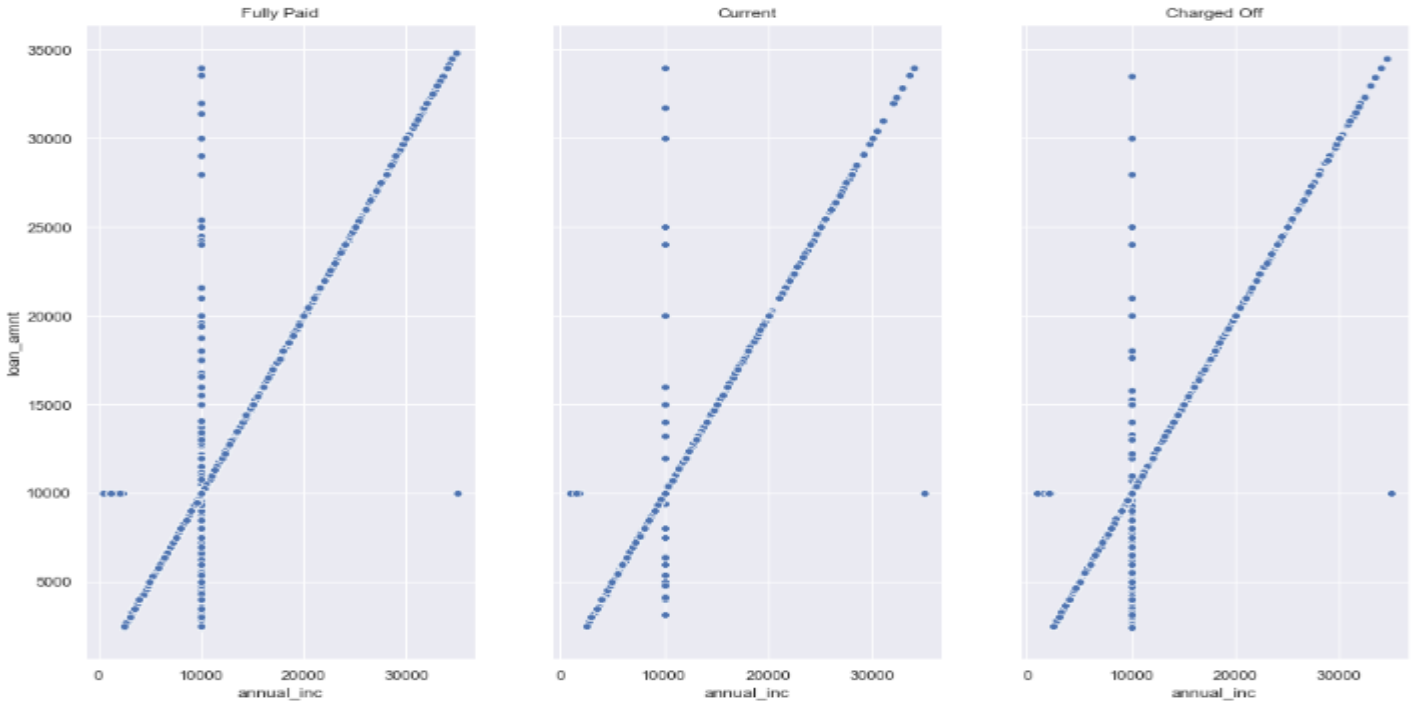
upGrad

Distribution of Annual Income v/s Loan Amount



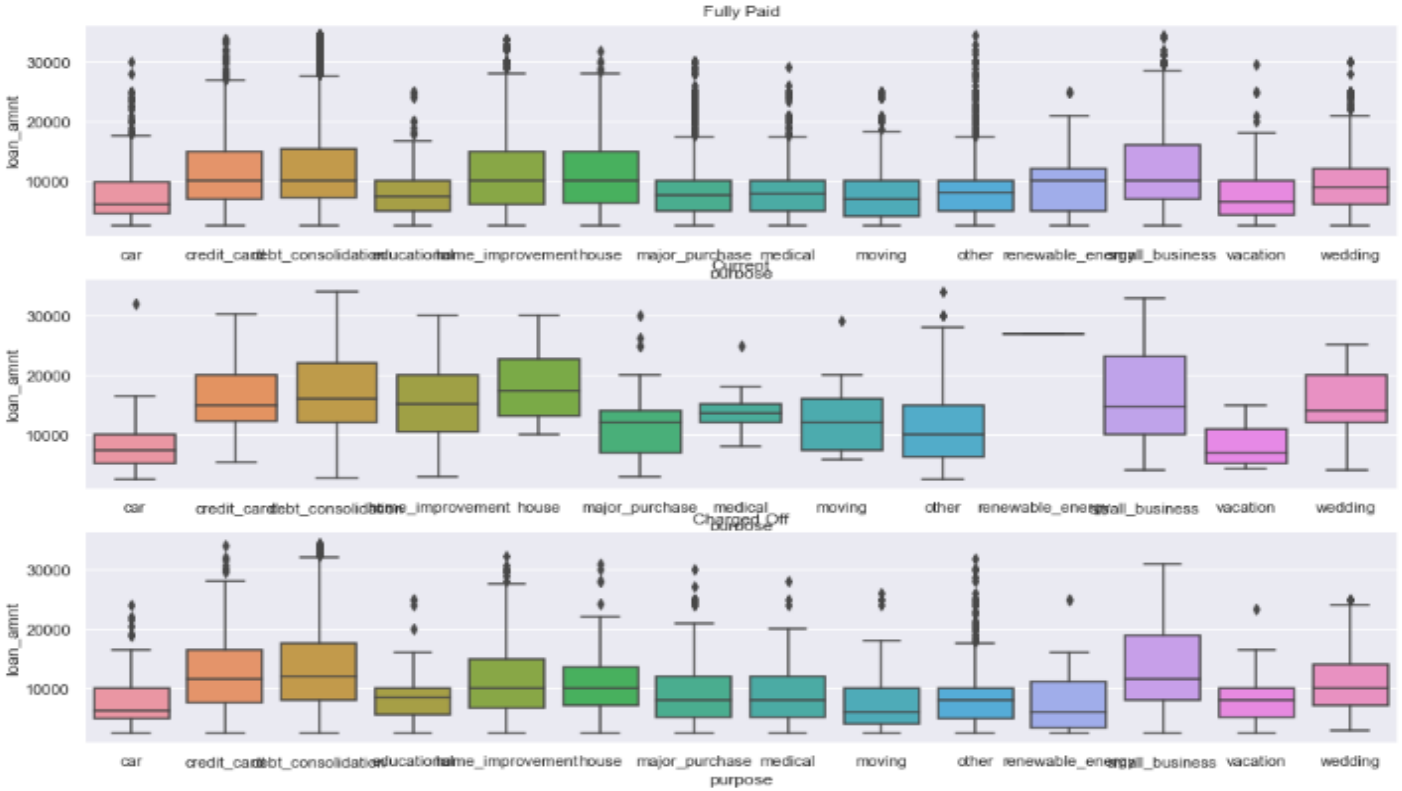
upGrad

Distribution of Annual Income v/s Loan Amount based on loan status



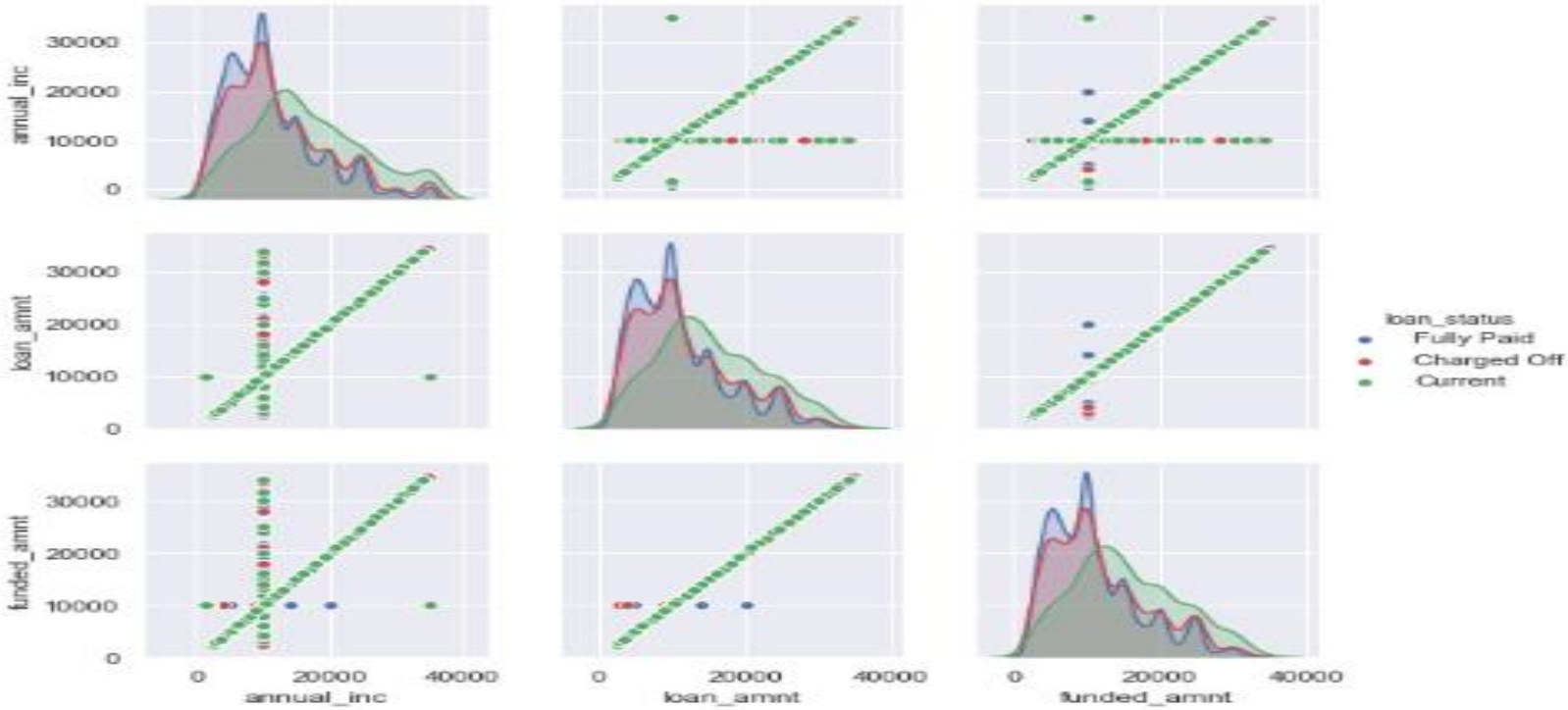
upGrad

Distribution of Loan Amount v/s Purpose



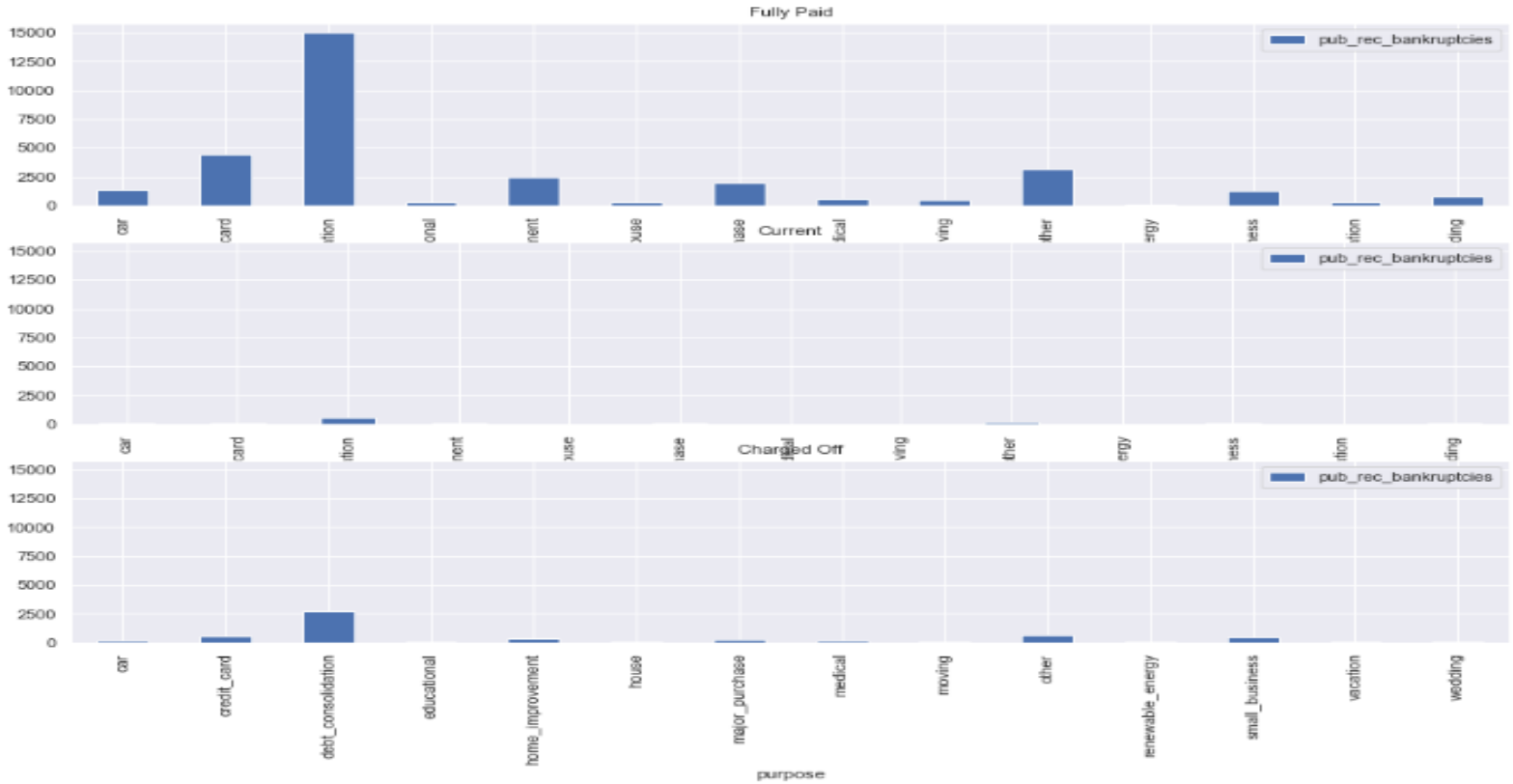
upGrad

Pair plot for Annual income, loan amount, funded amount and purpose based on loan status



upGrad

Pivot table for purpose against bankruptcies based on loan status



upGrad

Summary

After exploratory analysis of lending club case, following variables need to be verified to prevent from approving the loan for borrowers who likely to default

1. Address state
- a. states with CA,FL,IL,NJ,NY,PA and TX are defaulted more.

2. Purpose
- a. Credit_card, debt_consolidation, small_business and others are defaulted more

3. Employment length
- a. employers with > 10 years and > 1 year are defaulted more

4. Loan amount
- a. Defaulters take higher range of loan amount

5. Interest rate
- a. Defaulters goes for higher range of interest rate

6. Term (tenure)
- a. Defaulters goes for higher tenure(term)

upGrad

Thank You