

Lending Club: Case Study

Team:

Chaithra.R

C.Monika Raj

Introduction

- **Problem Statement:**

To predict if in case a default is likely to occur based on the characteristic trends on combination of features.

- **Problem Interpretation:**

Our aim is to understand the risky loan applicants and identify the driving factors that is able to predict the default and control before it happens.

- **Data Description:**

Dataset consists of behavioral factors, Loan amount and other factors which are collected during application of loan and the status of their loan which consists of Fully paid(non-defaulter),current and charged off(defaulter).

Data Cleaning

Total No of rows given: 39717

Total No of columns given: 111

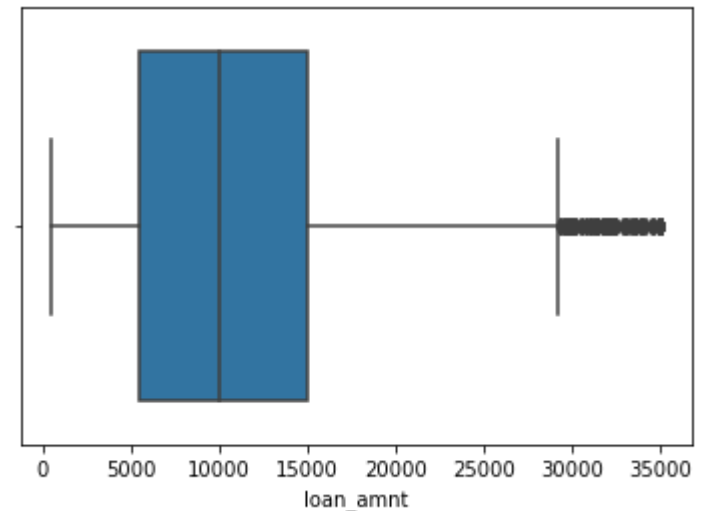
- Out of 111 columns in the given data, columns with respect to Missing values more than 85%, Behavioural columns, columns with textual data, columns with only 2 values which are 0 and nan that cant be imputed are dropped.
- Hence, only 18 columns are retained for the further analysis after cleaning the data wrt to missing value, imputation and removing rest of the columns based on the values they contain to make the analysis easier.
- Rows which contain 'Current' as their loan status are dropped as the analysis is being done for defaulters and non defaulters.
- Rows after dropping current status is 38577

Outlier Detection

- Outliers are present in most of the numerical columns.
- And it is handled by:

IQR Technique

- Concepts of “Flooring” and “Capping” is used since do not want to drop any values from the data.

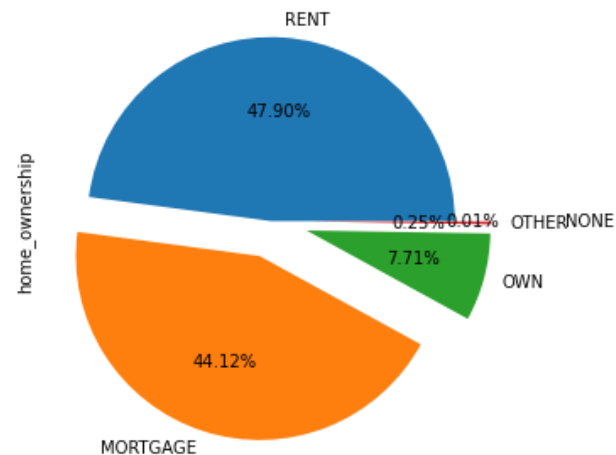
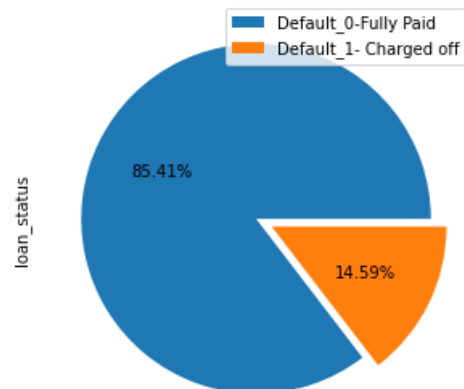


Data Analysis

Univariate Analysis

- Loan Status -
 - Fully Charged (non defaulter) – 85.41%
 - Charged off (defaulter)-34.59%
- Home Ownership –
 - Rent – 47.90%
 - Mortgage – 44.12%
 - Own - 7.71%
 - Other - 0.25%
 - None - 0.01%

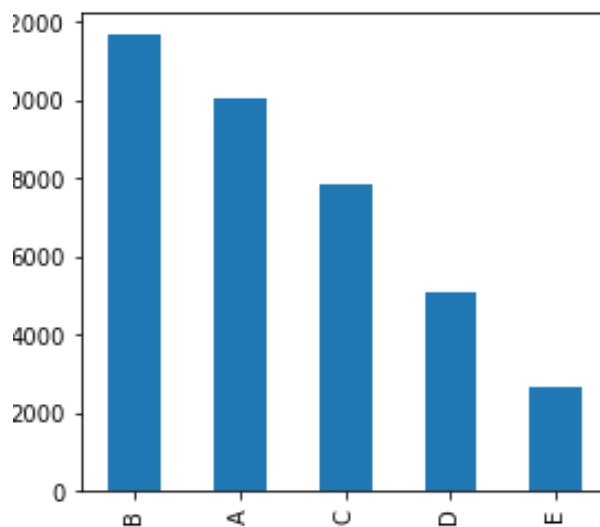
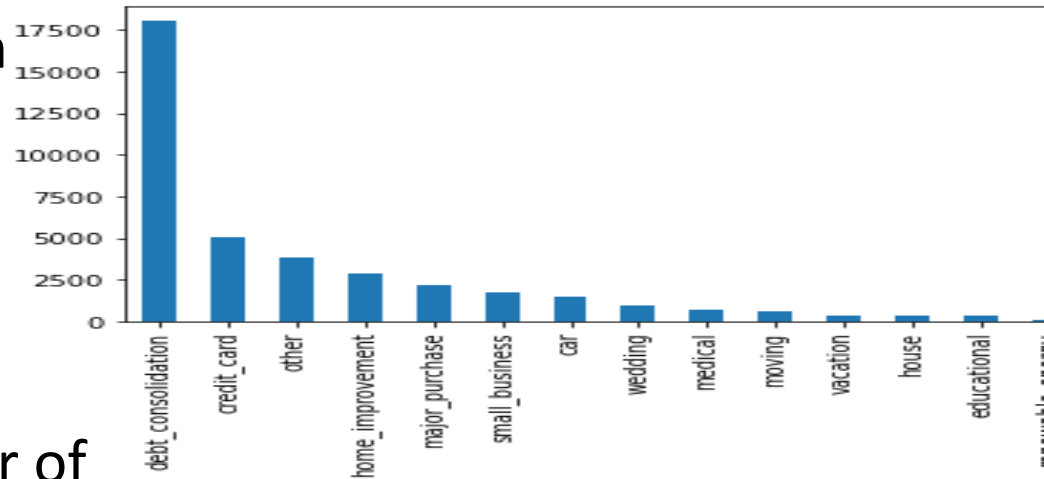
Percentage of Defaulters



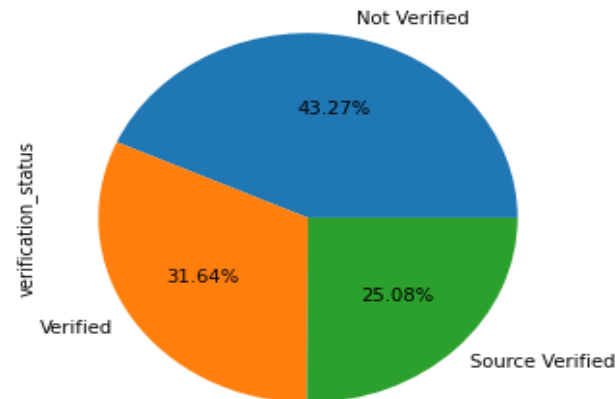
Verification Status shows that percentage of applicants with Not verified are higher.

Purpose of Applying loan is higher for Debt consolidation.

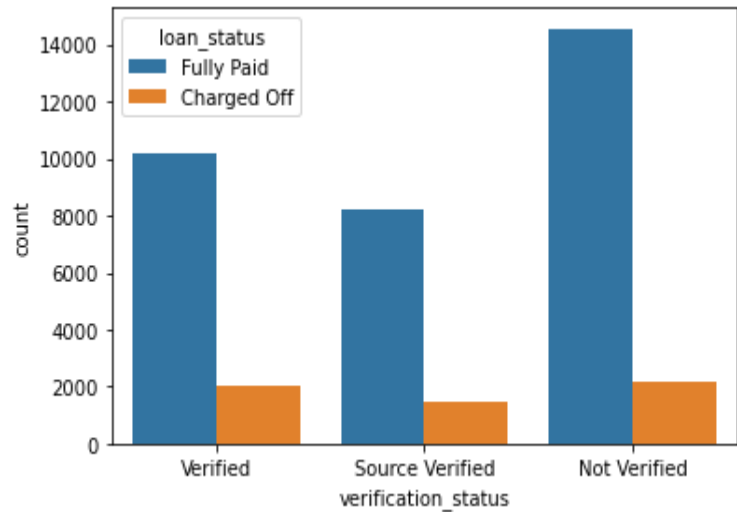
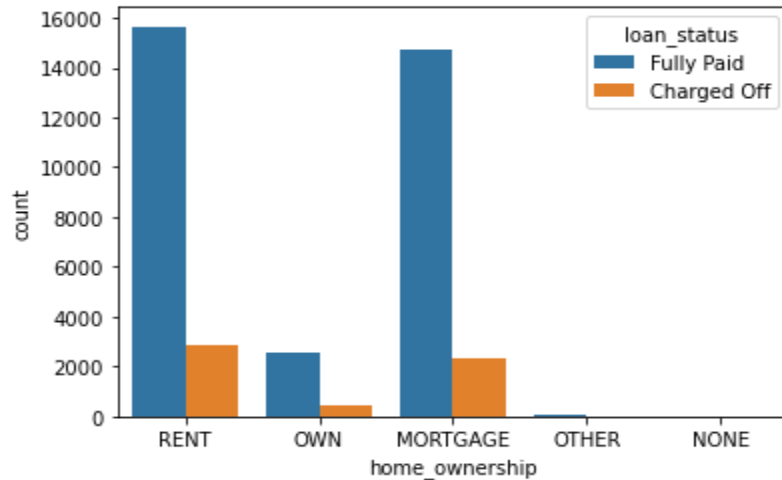
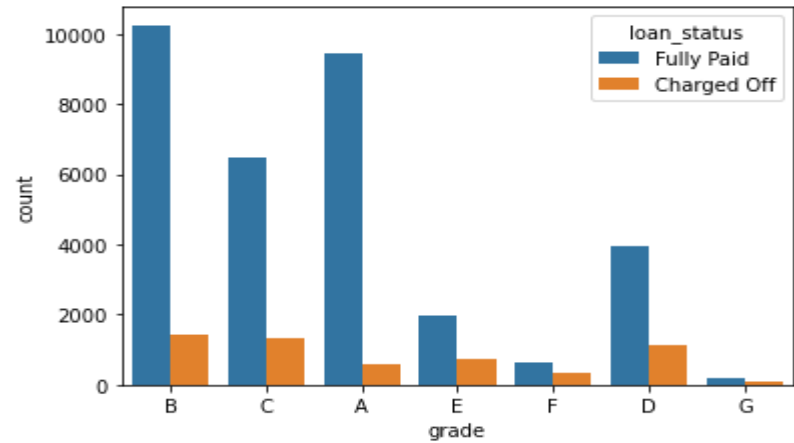
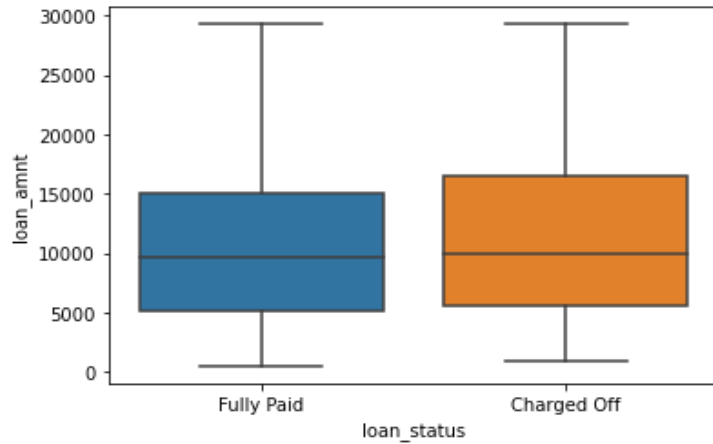
Grade B has a greater number of applicants.

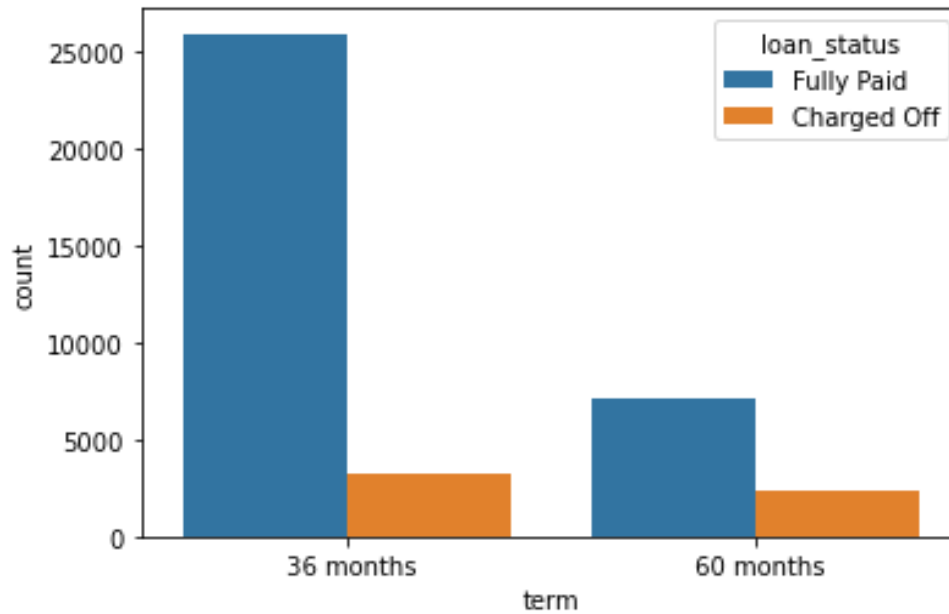


percentage of income status being verified

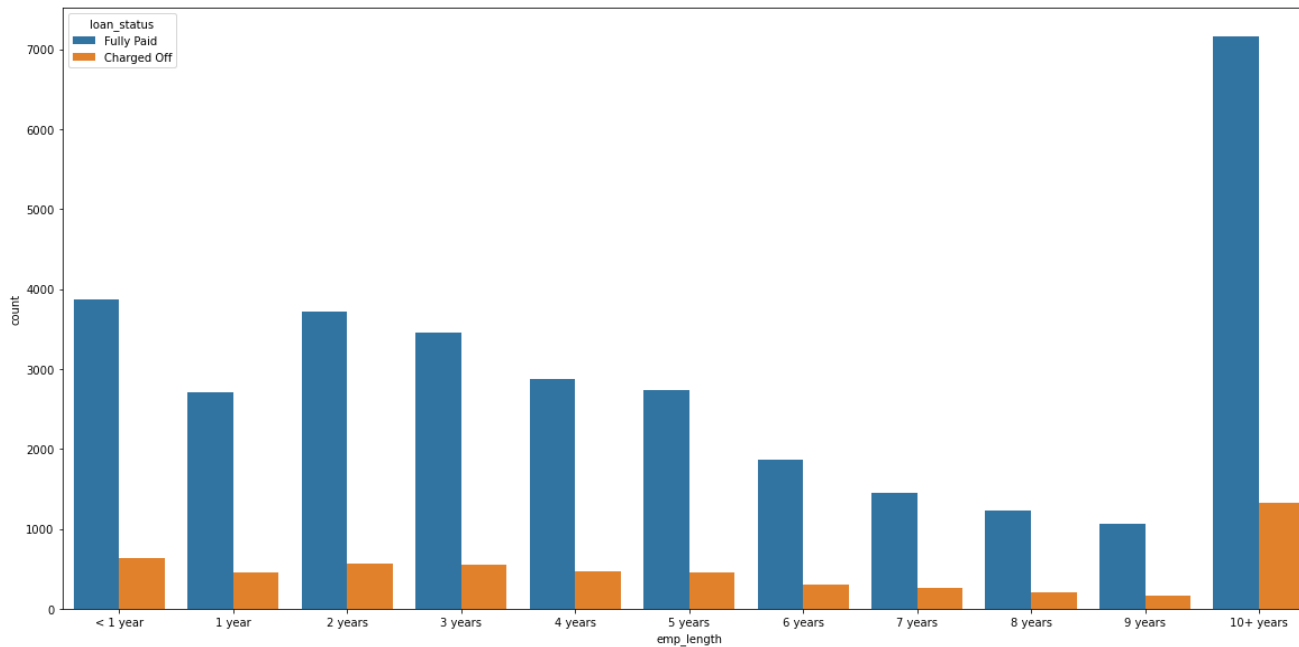


Bivariate Analysis

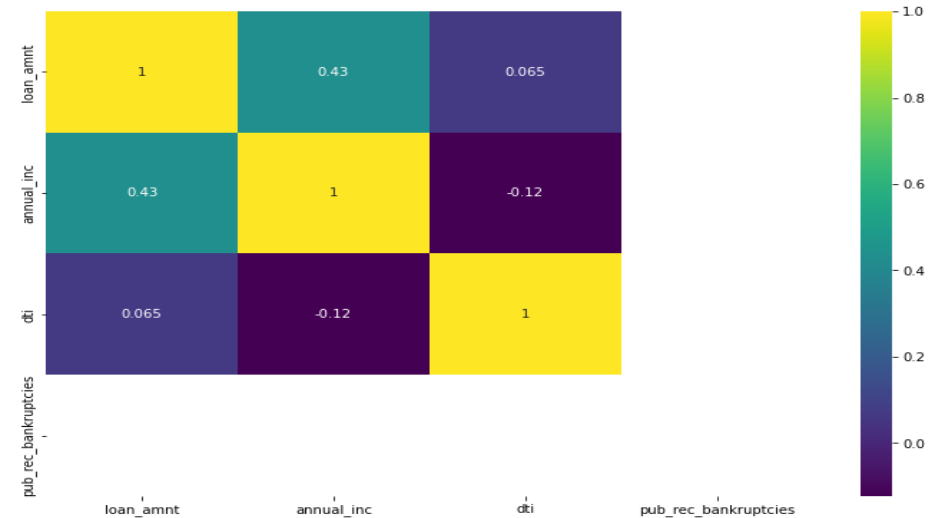
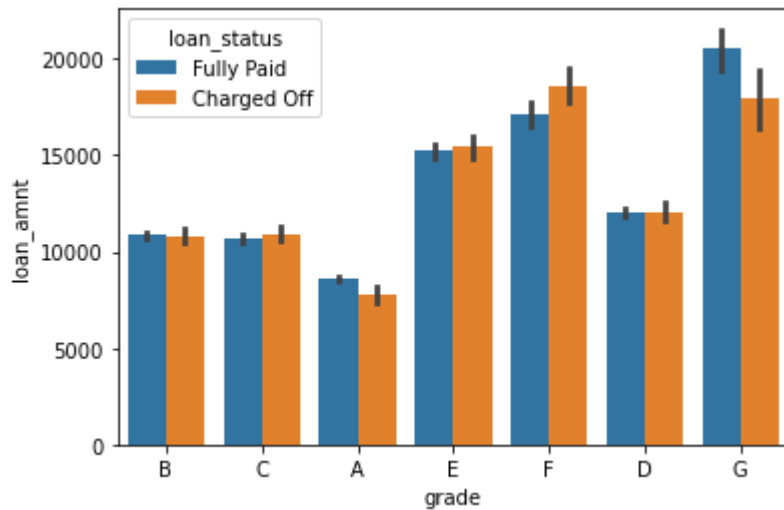
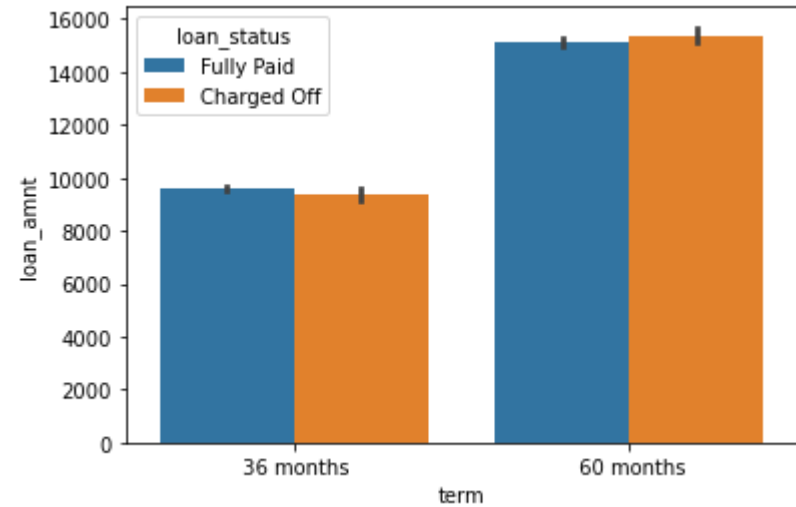
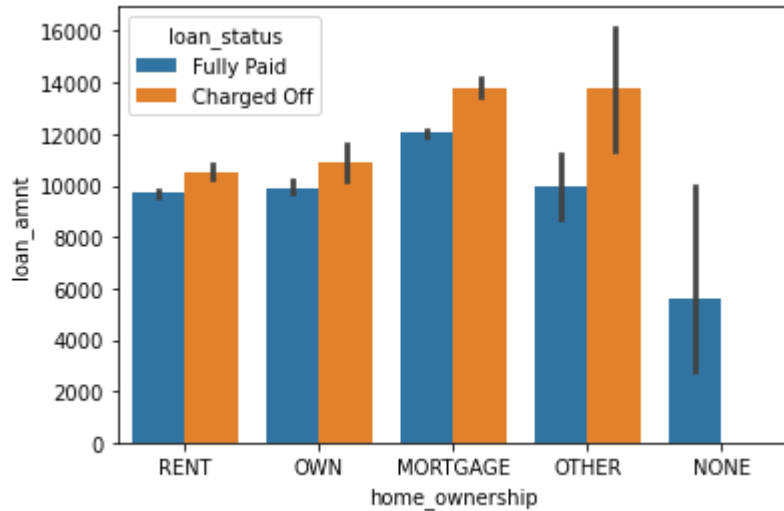




In Bivariate Analysis, loan status is considered for the comparison.



Multivariate Analysis



Recommendations:

During approval of loan keeping the points below regarding defaulters are helpful:

- Providing higher loan amount for the people with home ownership status rented and others are risky.
- Providing loan for verification status being not verified are more likely to default.
- Experience in the job should be analysed before the approval of the loan.

Thank You