Lending Club Case
▸ # Study

Puneet Aggarwal, Gagan Dua

Problem Statement

- You work for a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- 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

- The data given below contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Objective

▶ The Objective of this case study is to implement EDA technique on a real world problem.

▶ Understand the insights and present in a business first manner via presentation.

▶ Benefits of the case study:

  ➢ Gives an idea about how EDA is used in real life business problems.

  ➢ Develops a basic understanding of risk analytics in banking and financial services.

  ➢ Helps to minimize the loss of money while lending it to the clients.

Business Understanding

▶ The business objective is to take a decision whenever they receive a loan application whether to reject or approve based on certain variables.

▶ **Dataset Details**:

▶ The data given below contains information about past loan applicants and whether they 'defaulted' or not.

▶ Data has details regarding approved loan not the rejected ones. It has 3 status of loan which is Fully Paid, Current and Charged-Off. But for our analysis, we only need Fully Paid and charge-off loan status to get insight on Loan defaulter's case.

Followed EDA process steps list:

▶ Importing the data

▶ Removing unnecessary column

▶ Removing Null values

▶ Removing duplicate row and columns

▶ Fixing Null and missed values

▶ Datatype updation as per variable

▶ Creation of Derived columns from formal column

▶ Remove Outlier

▶ Visualize the data through Charts to get better insights

Loan Distribution By state

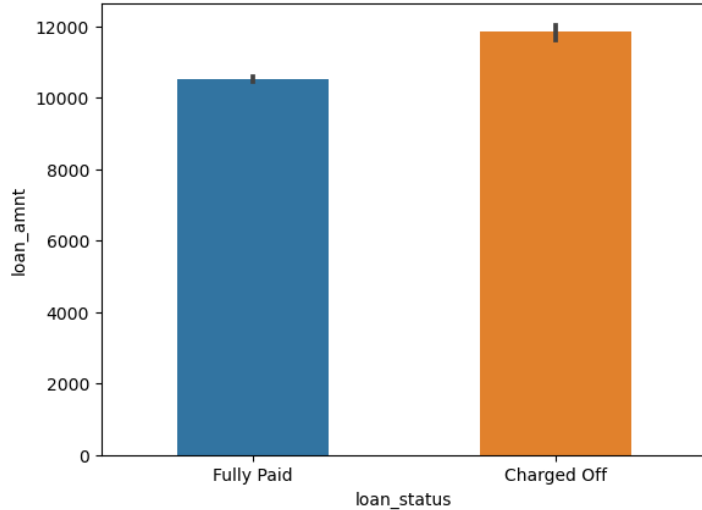California has the most number of defaulted one's.

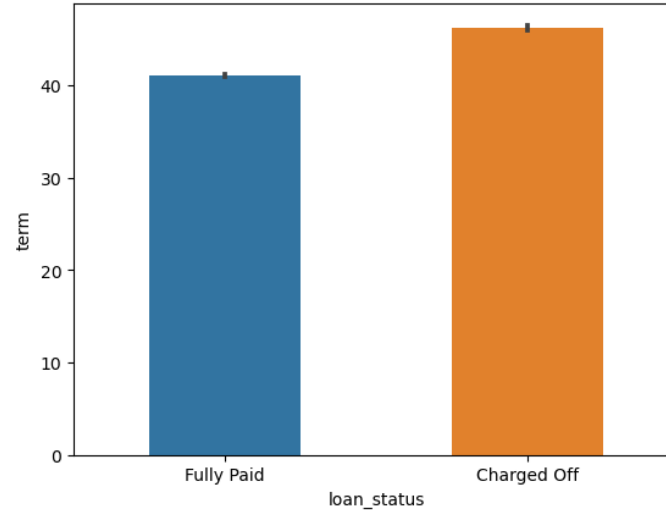Loan distribution by State
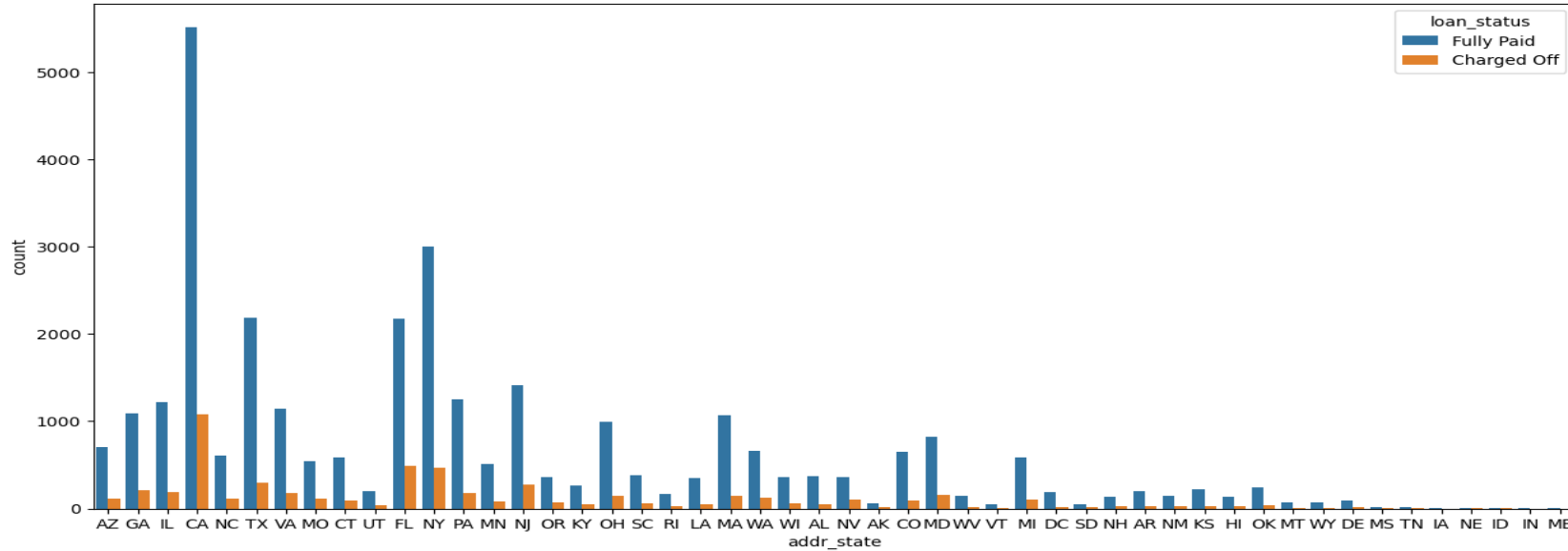
Loan Status, Loan Amount And Loan term

Borrowers having higher loan amount has higher chances of being a defaulter

Borrowers applying for 60 months loan have higher chances of being a defaulter

Loan Amount vs. Loan Status
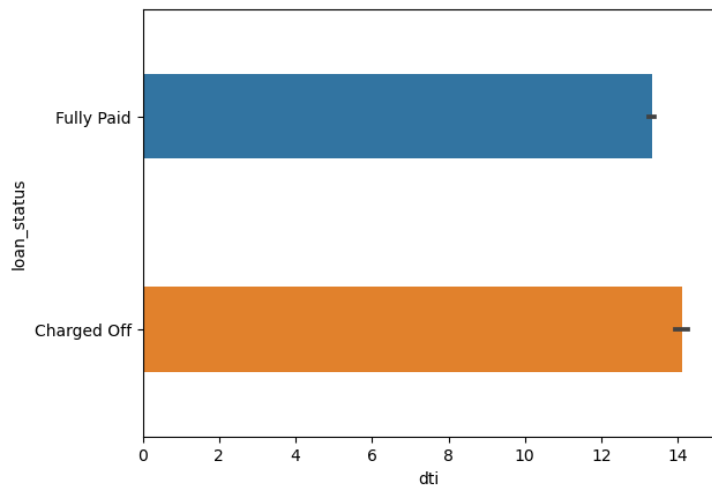

Loan Status vs. Loan term

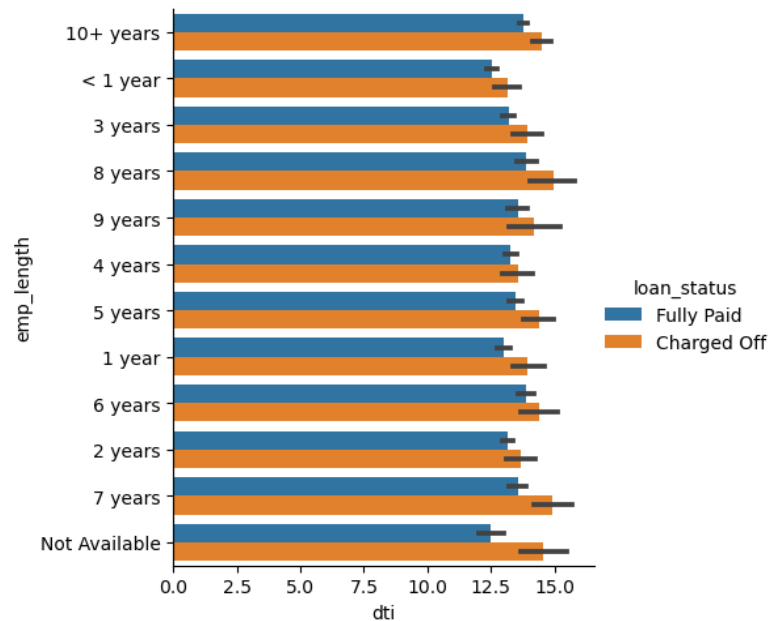# DTI vs Loan Status vs emp_length

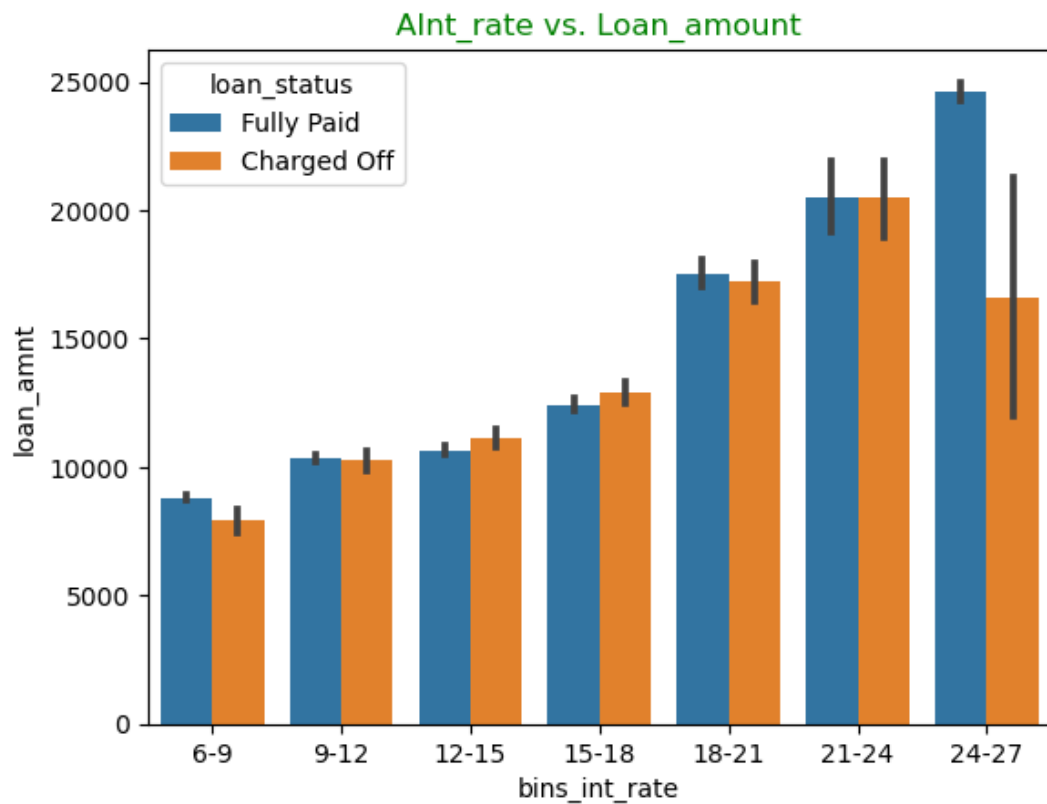The chances of being a default increase with the higher dti

applicants having 5+ years have more dti hence resulting into the possibility of a defaulter
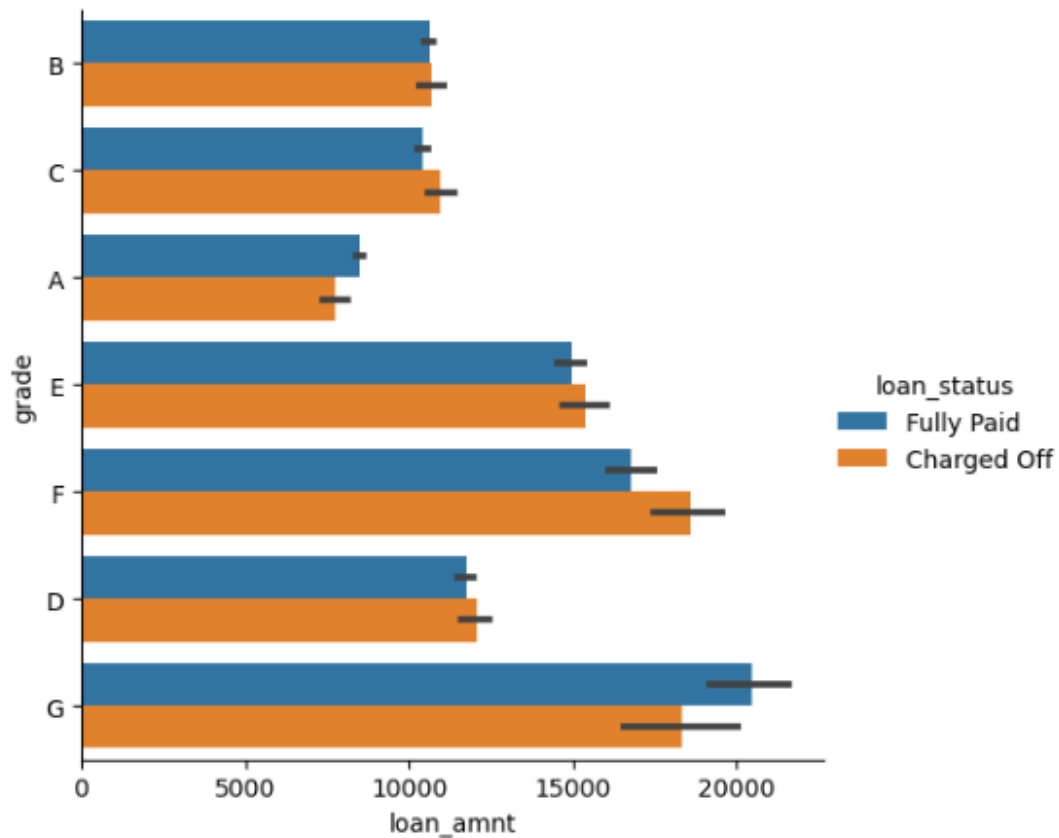
Int_Rate and Loan_amount

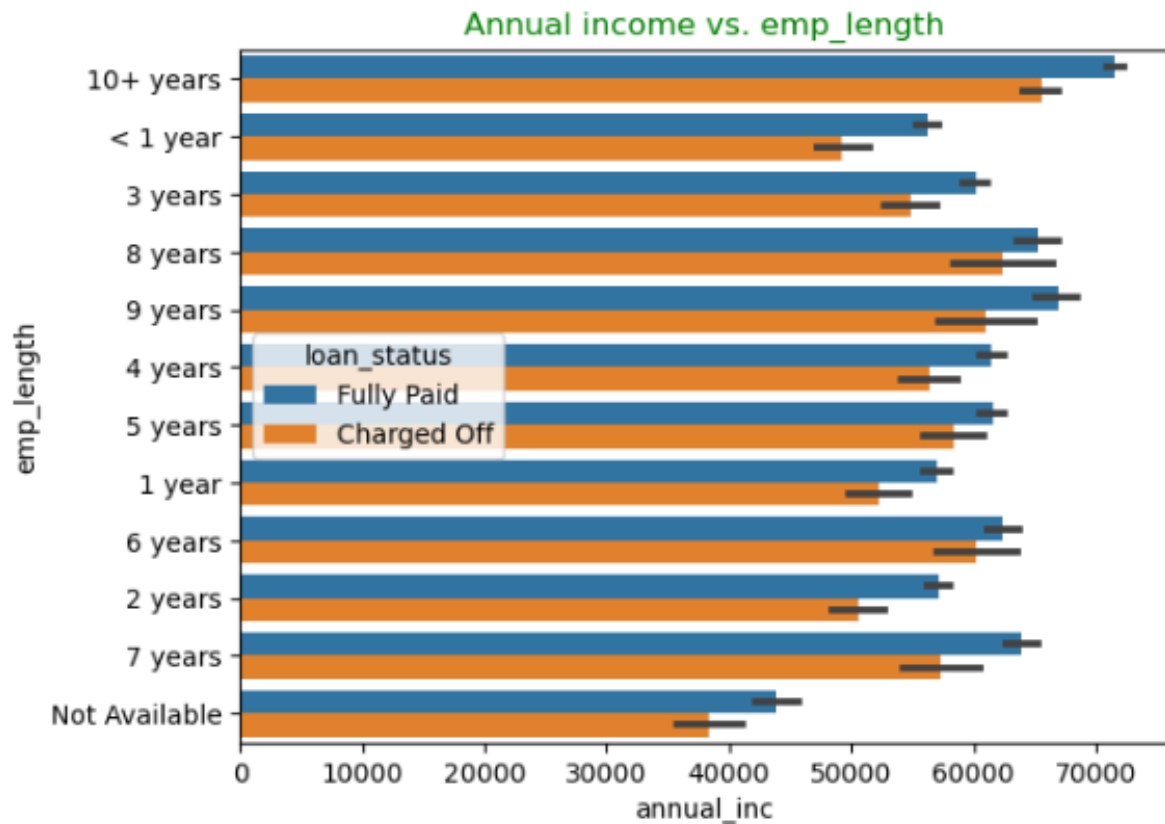▶ Higher interest rate results in higher number of defaulter

Int_Rate and Loan_amount

▶ Grade E, F and G has high number of defaulters.
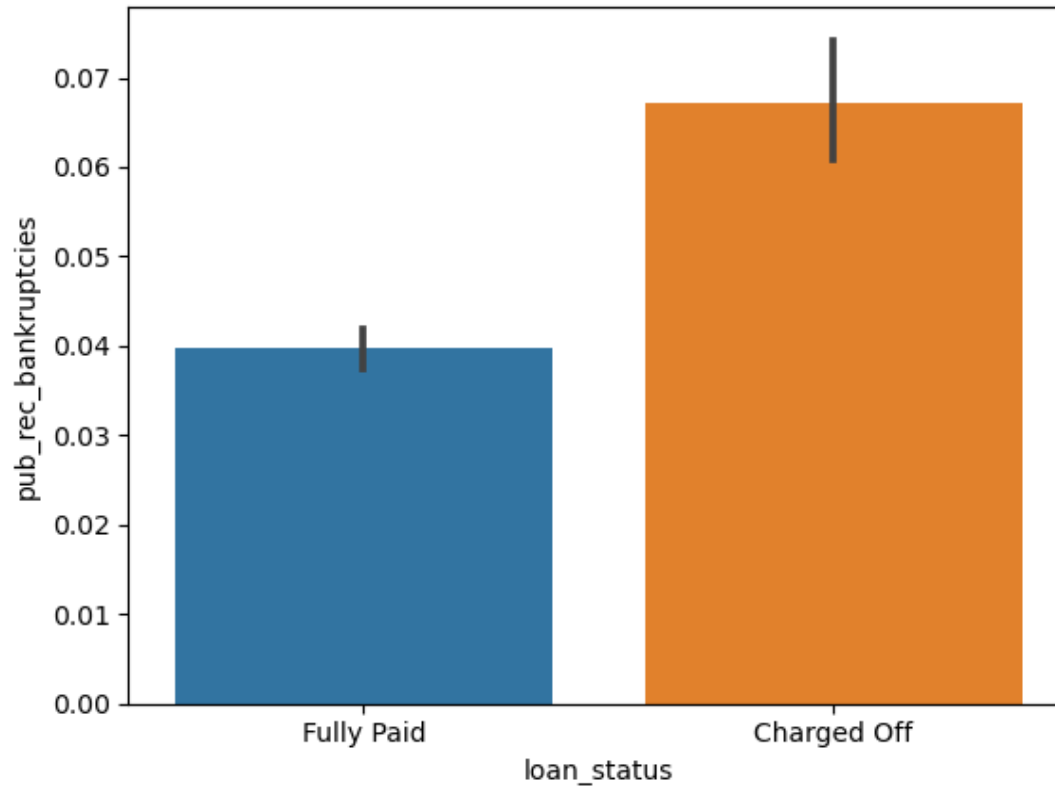
Int_Rate and Loan_amount

► Applicant with emp_length of 10+ years have a high
chance of being a defaulter.

Annual income vs. emp_length

Pub_rec_bankruptcies vs Loan_Status

Pub_rec_bankruptcies is affecting the number of loan defaulters.

Important Insights from Dataset Analysis

1. California has the most number of defaulted one's.
2. Three borrowers having higher loan amount has higher chances of   being a defaulter
3. Borrowers applying for 60 months loan have higher chances of being a   defaulter
4. Highly number of defaulter had annual income approx. from 35k to   55k. High salary leads to less chances of being a defaulter
5. The chances of being a default increase with the higher dti
6. Emp length having 10+ is offered more loan
7. Applicant with emp_length of 10+ years have a high chance of being a   defaulter.
8. Higher interest rate results in higher number of defaulter
9. Applicant taking loan For small business is likely a defaulter
10. A large number of applicant taking loan for mortgage are defaulters.
11. Applicants having 5+ years have more dti hence resulting into the   possibility of a defaulter.
12. Grade E, F and G has hugh number of defaulters.

Conclusion:

Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:

1. Pub_rec_bankruptcies
2. Grades
3. Annual income
4. Verification Status
5. DTI

The case study helped to identify the key insights i.e. variables including interest rate, loan issue year etc.