

# LENDING CLUB CASE STUDY

## SUBMISSION

**Raghuannndan Kavi**  
**(ML/Ai -C51)**

## Lending Club Case Study

### Background:

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans and financing of medical procedures. It's largest source of credit loss is lending loans to 'risky' applicants.

### Problem Statement:

Lending Club wants to identify the risky loan applicants, so that such loans can be reduced thereby cutting down the amount of credit loss.

Our aim is to find the driving factors behind loan getting default.

# Problem Solving Methodology

## Data Understanding:

1. Loan Data Set contains the complete loan data for all loans issued through the time period 2007 to 2011.
2. Data Dictionary describes the meaning of each variable used in Loan Data Set.

## Data Cleaning:

1. We dropped the columns having more than 60% missing values, where number of unique value is only one and columns which are not influencing our target variable i.e **loan status**.
2. We are assuming here that the columns representing customer behavior (e.g. Delinquency, revolving balance/utility etc.) are not relevant w.r.t our analysis.
3. We have removed all data where loan status is "Current" as this is not relevant to our analysis.

## Categorical Variables vs Percentage of Charged Off:

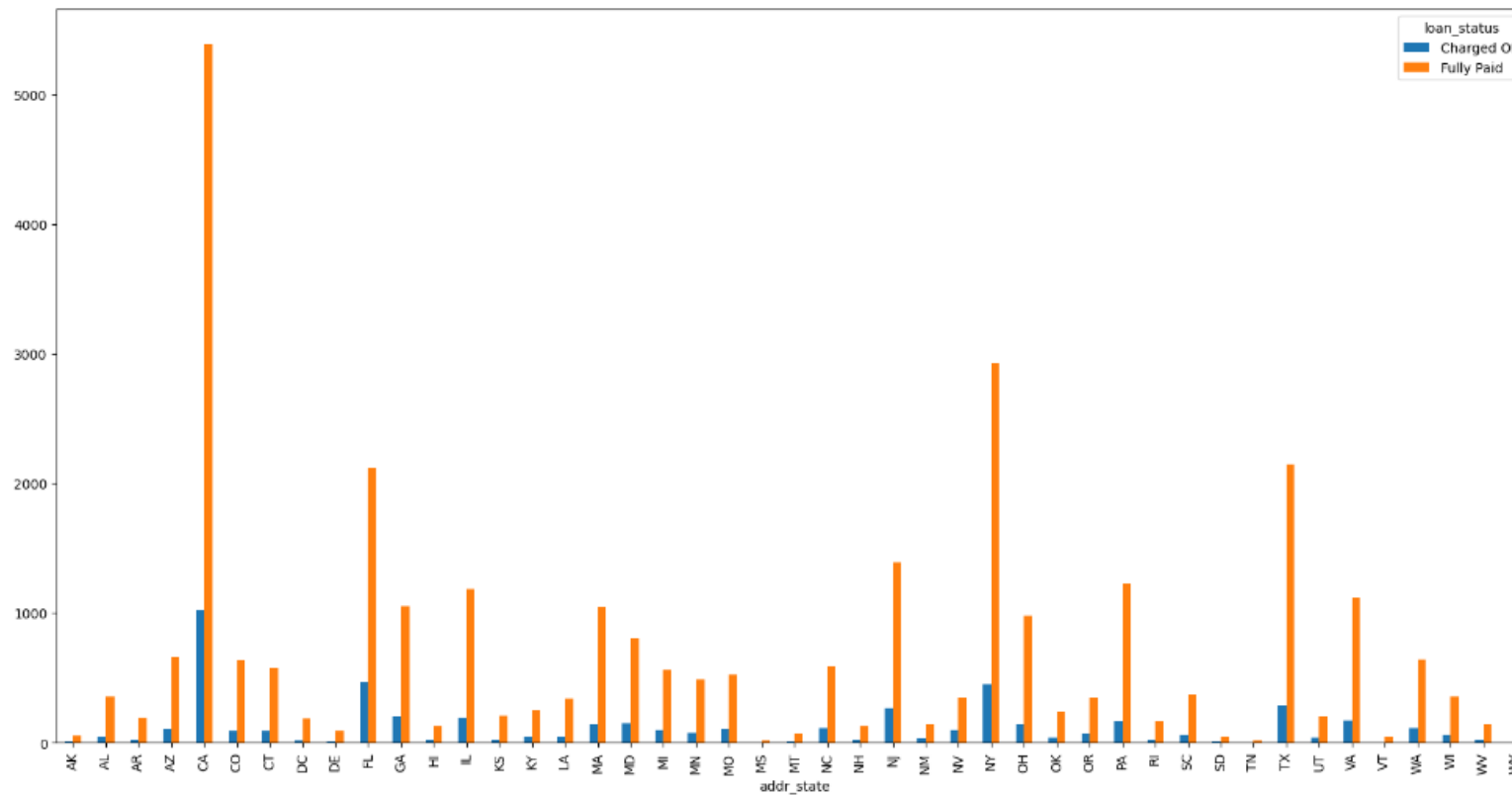
The main motive of this use case is to find what parameters are impacting Loan Status the most i.e. if an applicant will successfully complete the loan term or default. So, we will be using a new term **Percentage of Charged Off** that will be equal to:

$$\text{Percentage of Charged Off} = (\text{Number of Applicants Charged Off} / \text{Total No. of Applicants}) * 100$$

We will calculate this percentage w.r.t each column in bivariate analysis & will see how the Percentage of Charged Off changes with these columns.

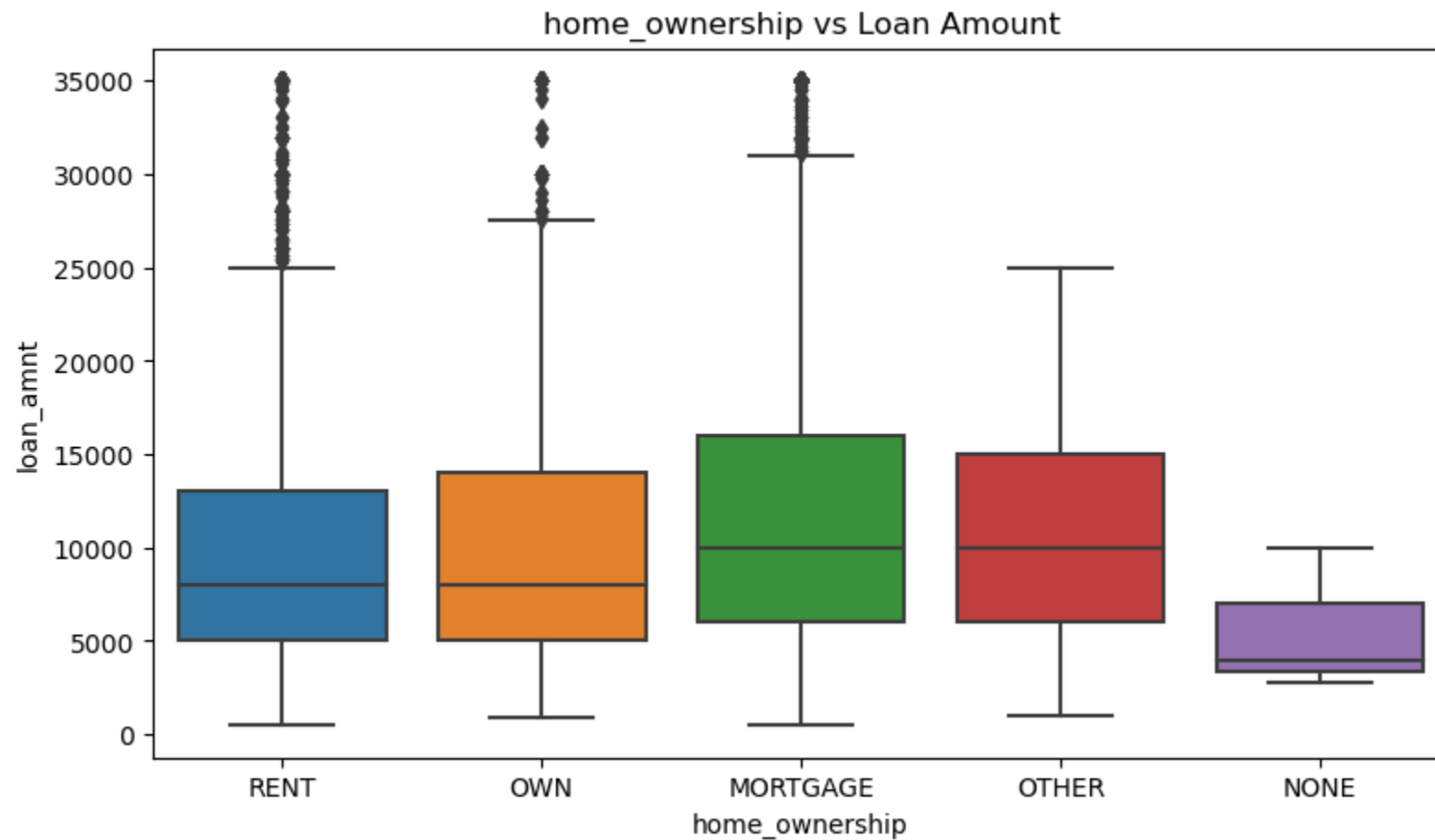
# Loan Analysis

- 1.Maximum no. of loans are applied by people from CA.
- 2.Maximum no. of loans are charged off for people of NV.



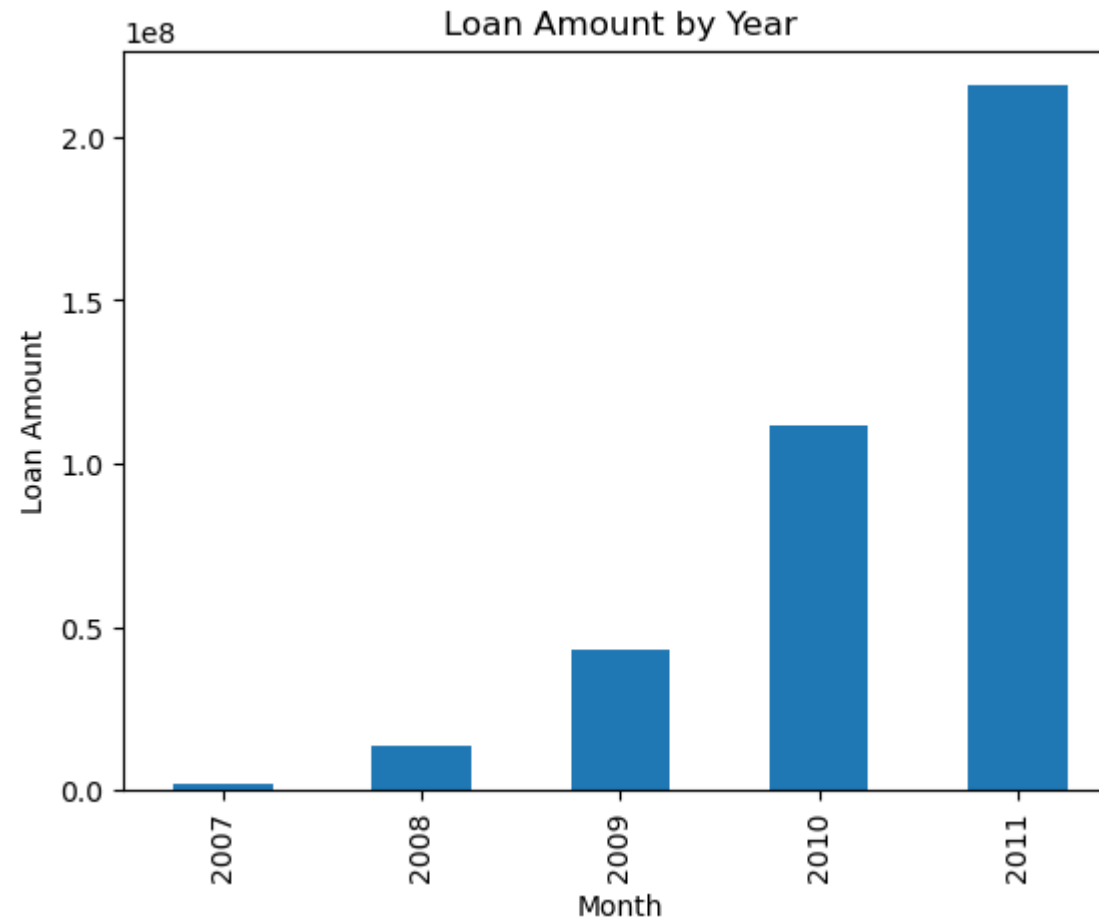
# Home Ownership and loan amount

1. Mortgage home ownership have taken maximum loans. This shows that having collateral is a safe for banki in providing loans



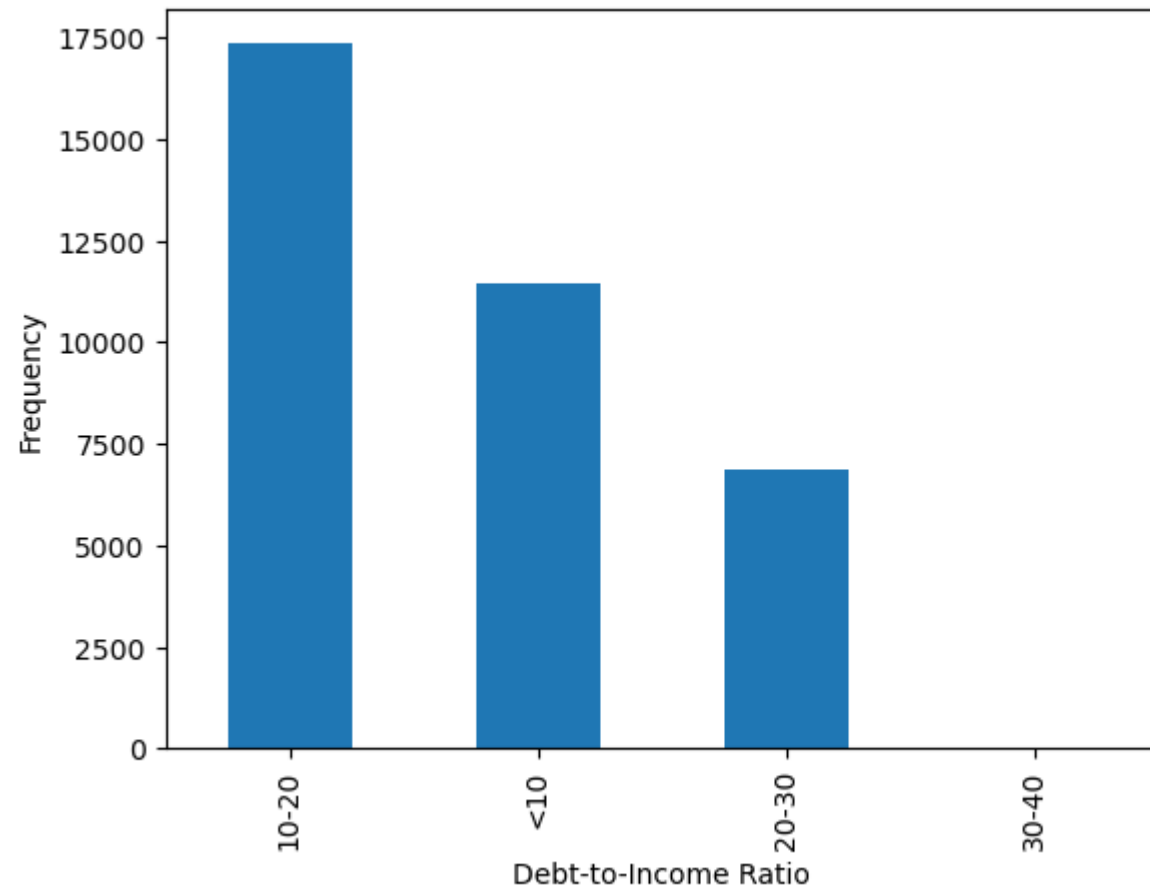
# Year & Month Wise trends of Loans

Number of loans provided each year is increasing



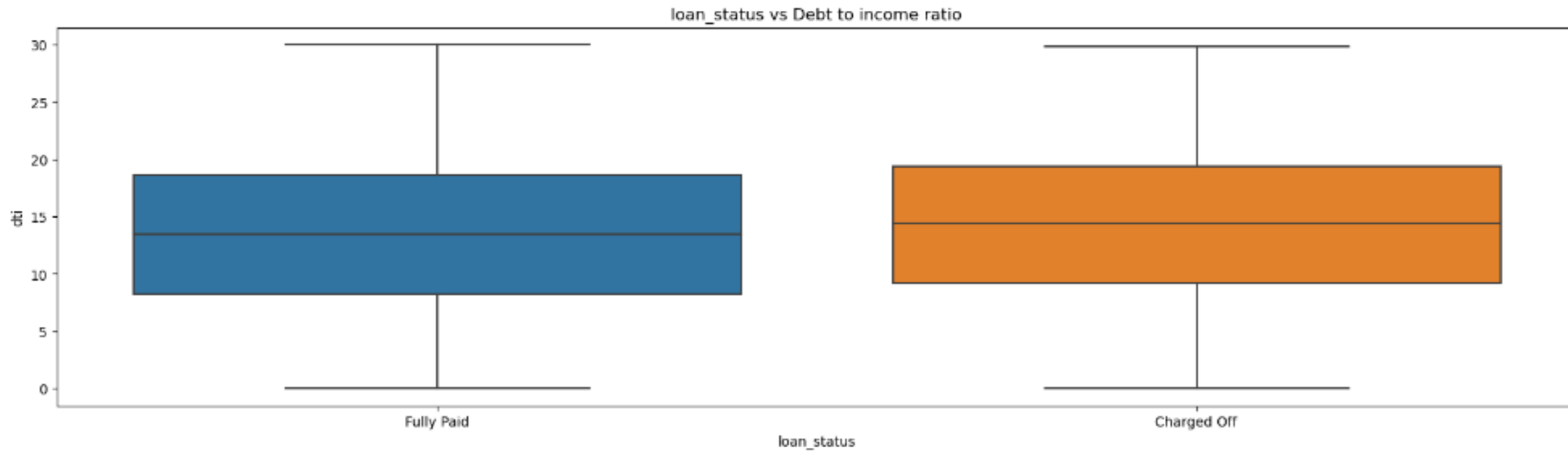


Debt-to-Income Ratio Distribution



If debt to income ratio is more it is more likely the person will default. This poses a higher risk for lender.





People with more debt to income ratio are more likely to default

## Correlation Heat-Map

1. Loan Amount, Funded Amount, Funded Amount invested and installment are highly correlated with each other.
2. Annual Income is negatively correlated with DTI.



## Recommendations to Lending Club

- **Higher debt to income ratio is likely to default. Its not safe to provide loans to these people**
- **Higher loan term has more number of defaulters. So the risk increases**
- **There is a co-relation between loan amount, funded amount and funded amount by investor.**
- **Number of loans increase each year.**
- **Higher the annual income, higher the loan amount given**
- **Longer employee length and verified source is a must check to provide loans as employee length indicates stability and verification decreases risks**
- **Loans provided to People with house mortgage is safe as this involves house as some guarantee for the lender.**