

# Lending Club Case Study

Group Members:

Shubhada Prabhakar Babhulgaonkar  
Shravan Kumar Yadav

# Problem Statement

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

# Steps to solve the problem



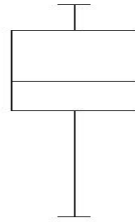
**Data  
Cleaning**



**Data  
Understanding**



**Data  
Filtering**



**Univariate  
Analysis**



**Bivariate  
Analysis**



**Recommendations**



## Data Cleaning

We have cleaned the data on the basis of below observations.

- 1) Remove columns having all NA values
- 2) Remove columns having single unique values as they do not have any impact on analysis.
- 3) Removing columns having values NA and blank.
- 4) Remove columns that are unique to each column. Example id, url
- 5) Removing the columns which are not know to bank at the time of loan application filled. These are created after loan is approved and running. For example `revol_bal`, `out_prncp` etc.
- 6) Removing columns in which we have null values more 90% of rows. example `mths_since_last_record`
- 7) Removing “%” character from interest rate column



## Data Understanding

- No. of rows and columns in data
- Name of columns
- Not Null values for each column
- Data types of columns
- Number of numeric and string columns
- Check on mean, median,min,max for numeric columns
- View on sample of rows to see values stored in each column

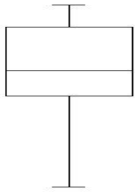


## Data Filtering

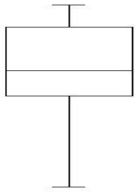
We observed that the data contains 3 different loan status.

- 1) Fully Paid
- 2) Charged Off
- 3) Current

Fully paid are loans which are completely paid by borrowers. Charged off is where the borrower is unable to pay whole loan. Current is the current running loan which is neither complete nor charged off. For analysis the impact of attributes on loan status , we need to drop the loan with status “Current” because these are running loan. We are not including it in analysis because our main aim was to analyse the key factors which will impact the loan being fully paid or charged off. Therefore minimising the company loses.



# Univariate Analysis



# Univariate Analysis





# Bivariate Analysis



# Bivariate Analysis



# Bivariate Analysis



# Bivariate Analysis



# Bivariate Analysis



## Recommendations

1. LC should reduce the rate of interest for 60 months tenure to borrowers as the chance for them to default is more.
2. Grade is a good attribute in detecting the defaulters. LC should be more careful for borrowers having grade G to A.
3. LC should avoid giving loans to states ID, NV and NE.
4. Loan for the purpose of small businesses are likely to default the loan. LC should reduce approving loan to them.
5. LC should look into the annual income of the borrower. Borrowers with less annual income are likely to default the loan.