



# LENDING CLUB CASE STUDY

SUBMITTED BY:

ASWANTH KUMAR V

MARY DIVYA ANUSHA YERUVA



# AGENDA



- Problem Statement
- Data Understanding
- Data Cleaning and Manipulation
- Data Analysis
- Plotting the Graphs
- Observations and conclusion
- Recommendations



# Problem Statement



- Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.
- 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.



# Data Understanding



- Go through the problem statement clearly and understand the business requirement
- Understand the columns with given dictionary data and the usage of the columns
- Compare the columns with the relevant columns to get more clarity
- From the data set explore the data by looking at sample data by filtering few columns
- Ignore the columns having none values which makes understanding the data is easier
- Finding out the data quality issues in the data set
- Filtering out the special characters and spaces which blocks your data transforming
- Finding out the columns having yes unique item or categories makes filtering out the other columns easier to understand the data much better



# Data Cleaning and Manipulation



- Checking the percentage of Nan values in dataset.
- Dropping the Columns having greater or equal to 75% Nan values.
- Dropping the columns which are having unique data.
  - Finding the columns having unique data
  - Sorted unique values based on the columns.
  - Dropping the unnecessary columns as per the column's data uniqueness.
- Dropping the columns which are not required for the data analysis (Customer behavior)
- Dropping the rows home\_ownership is having NONE values
- Dropping the rows home\_ownership is having Other values
- Dropping the rows loan\_status is having Current Value

## Contd..

- Removing the % from int\_rate column(eg: 13% as 13)
- Converting the term column to integers ( eg: 36 months as 36)
- Formatting issue\_d
- converting installment column to int
- Approved ratio of loan amount  $(\text{funded\_amnt\_inv}/\text{loan\_amnt}) * 100$
- Manipulating the subgrading to numeric values for better visualization
- emp\_length column modification
  - emp\_length column modification to number of years ( eg: 6 years as 6)
  - Filling the empty cells of emp\_length with 0
  - If emp length is less than 1 year round off to 0
  - If emp length is greater than 10 years round off to 10
- revol\_util column data modification
  - Removing the % from revol\_util column
  - revol\_util column is customer behavior column can be removed if not required



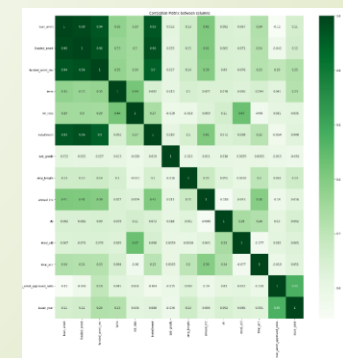
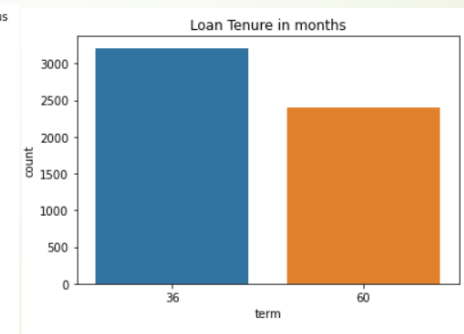
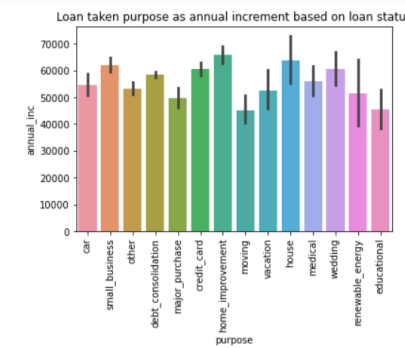
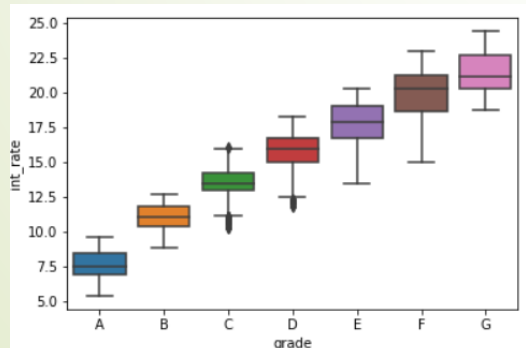
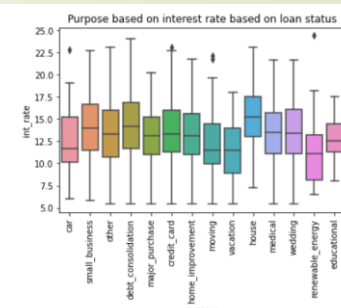
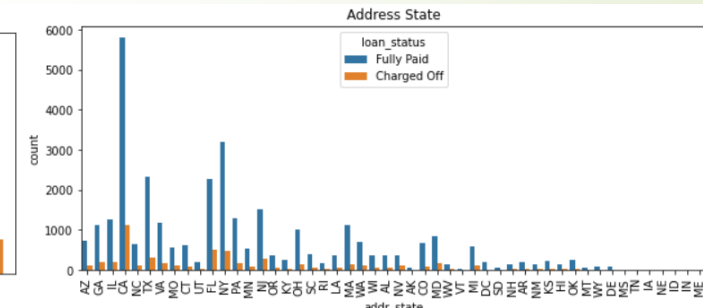
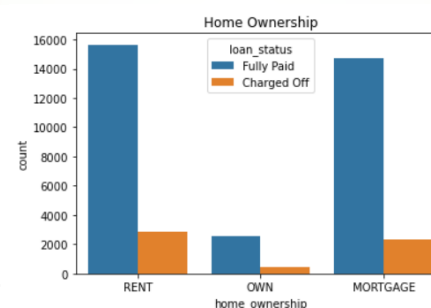
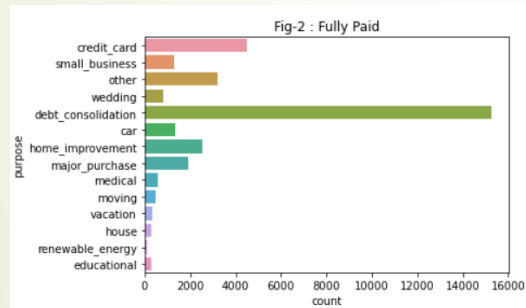
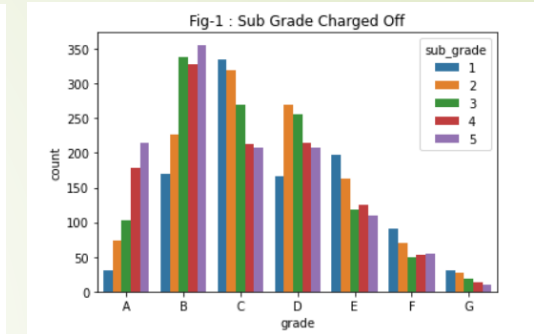
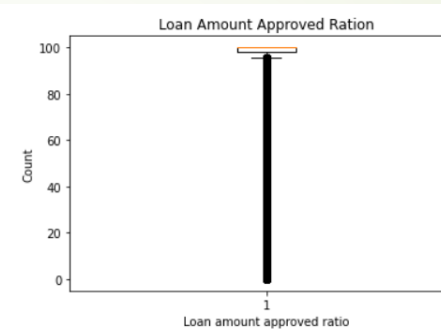
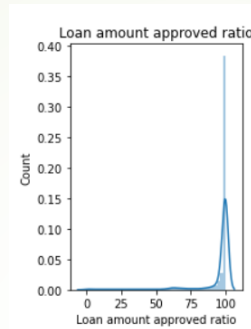
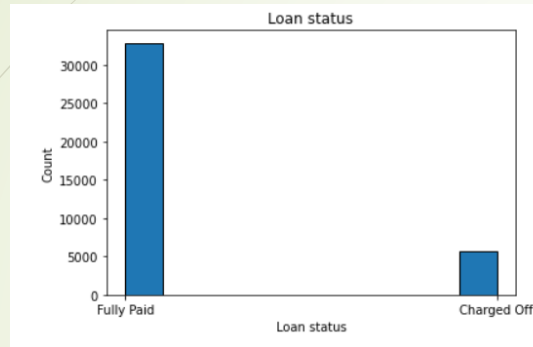


# Data Analysis



- The objective is to identify predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan.
- Data Analysis can be performed-
  - Univariate Analysis
  - Bivariate Analysis
  - Multivariate Analysis
- There are broadly three types of variables –
  - Those which are related to the applicant (demographic variables such as age, occupation, employment details etc.),
  - Loan characteristics (amount of loan, interest rate, purpose of loan etc.) and
  - Customer behavior variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.).

# Plotting the Graphs





# Observation and conclusion

Graph	Observation
loan_status	Around 5000 people who are defaulters in the loan status from the above graph
grade vs sub_grade , loan_status	Maximum number of charged off loans is happen in the grade B category from Fig-1
purpose, loan_status	debt_consolidation is having more Charged Off and Fully Paid
emp_length, loan_status	As per the above graph if emp lenght is more than 10 years charged off is more
home_ownership, loan_status	Home Ownership having RENT loan status is fully paid is in between 14000 and 16000
addr_state, loan_status	People take loan in CA Address State is Fully paid
verification_status, loan_status	As per the above graph if Verification Status is Verified then loan status is Fully paid
int_rate_cat, loan_status	Fully paid loan amount is happend if Interest Rate is between 10%-15%
revol_util_cat, loan_status	Revol util is in between 70-80 we are having more charge off loan status from the above graph
total_acc_cat, loan_status	If total number of accounts is in between 10-20 loan status is more in Fully paid and charged off from the above graph
term, loan_status	Charged Off is more if the term plan is 36 months from the above graph
loan_amnt_cat, loan_status	More charged off and fully paid is happened in 0 to 10000 range
dti_cat, loan_status	Dti between 12-18 having more charged off when compared with other ranges
emp_length, loan_status, funded_amnt_inv	Emp having 10 yrs having more charged off for funded loan amount investment 12000+
verification_status, loan_status, funded_amnt_inv	More charged off is in verified status in between 14000 to 16000 funded amount investment
grade, funded_amnt_inv, loan_status	17500 funded amount investment have more charged off in F grade
grade, loan_amnt, loan_status	18k - 20k loan amount have more charged off in F grade
annual_inc_cat, loan_status	More Charged off in 30k to 60k range
purpose, loan_status, annual_inc	Annual income with 50k to 60k with purpose home improvement have more charged off
int_rate_cat, annual_inc, loan_status	Interest rate 20 to 25 percentage with annual income 60000+ have more charged off
purpose, loan_status, int_rate	Loan taken for house purpose has more charged off
int_rate, grade, loan_status	Interest high with grade G has more defaulters
home_ownership, loan_status, annual_inc	Mortgage home ownership with 60000 annual income have more charged off
annual_inc_cat, loan_status, int_rate	Annual income with 90k-120k having high interest rate have more charged off
issue_year, loan_status	In 2011 Charged off loans are more as per the above graph



# Recommendations

- ▶ Verification of employment proof of loan applicant.
- ▶ Reduce loan approval for Rent category in home ownership. Because it has higher rate in charge off.
- ▶ Proper investigation and verification on the Rent category home ownership before giving loan approval.
- ▶ Avoid approving loans for grade B and Sub Grad B, C
- ▶ Do proper inquiry for Grade B,C,D and Sub Grad A, B, C
- ▶ Avoid approving loans for sub grade B.
- ▶ Applications who take loan on debit consolidation purpose are likely to be charge off
- ▶ The purpose of loan is lesser than annual income range have a proper verification
- ▶ Applicant annual income range greater than 60k request to recommendation highly for this category
- ▶ A mortgage home ownership required strict enquiry



THANK YOU