

# Lending Club Case Study

Group ID: **74994**

Group members:

**Nikhil Kumar Patel**

**Shruthi N**

# Business Objectives



## Understanding

When the company receives a loan application, the company must decide for loan approval based on the applicant's profile.



## Analysis

To identify these risky loan applicants using Exploratory data Analysis.



## Recommendations

The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e., the variables which are strong indicators of default. .

# Analysis Approach

## ❖ Data understanding

Understanding the available dataset and how each variable would play a role in our analysis.

## ❖ Data Cleaning

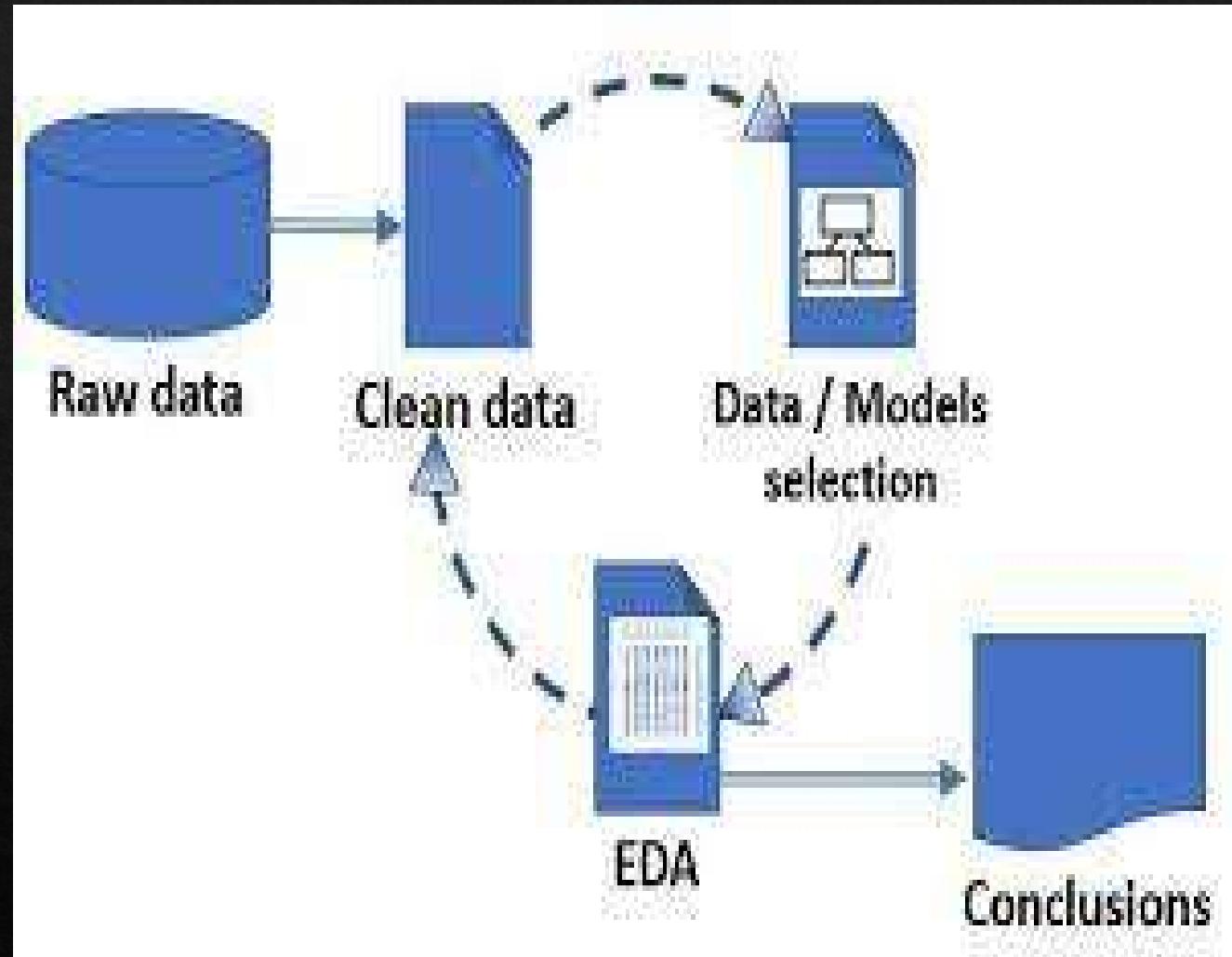
Remove N/A columns/rows

Check for duplicate data

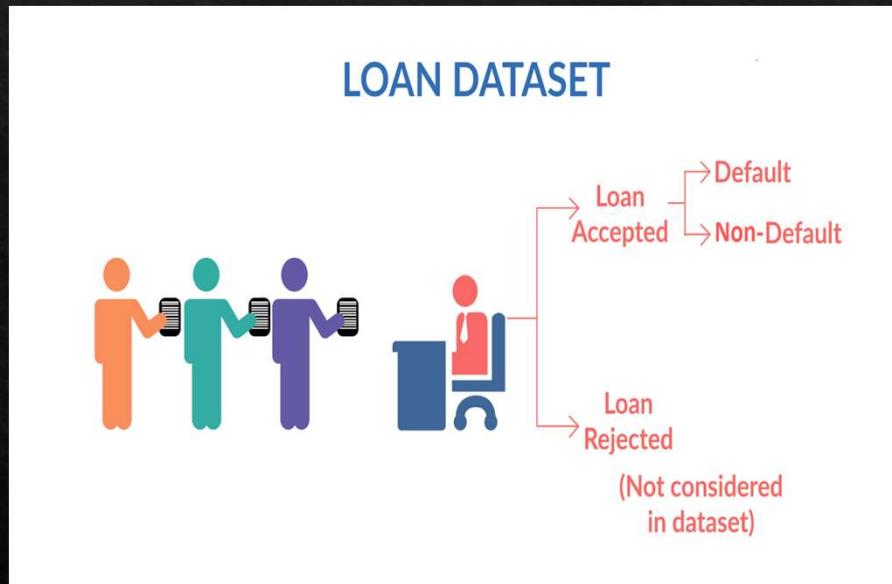
Imputation

## ❖ Data Analysis

Performing EDA using univariate, bivariate and correlation maps to identify the relationship between various variables



# Data Understanding



- ❖ Our dataset contains the complete loan data for all loans issued through the time period 2007 to 2011.
- ❖ It consists of 39717 rows and 111 columns.
- ❖ Consists of categorical and numerical variables.
- ❖ We need to identify the variables that would help us determine the risky loan requests.

# Data Cleaning

- ❖ Remove duplicate data
- ❖ Removed columns having more than 50% of null values as imputing them would exaggerate the data
- ❖ Removed customer behavioral columns as these are not available at the time of loan applications, hence they can't be used as a predictor for loan approval
- ❖ Removed columns having all unique values as it won't serve much purpose in the analysis
  - ❖ Example: member\_id ,url ,zip\_code ,title
- ❖ Removed outliers
- ❖ Removed rows having 'Current' loan status
- ❖ Standardized variables such as
  - ❖ int\_rate (removed % symbol)
  - ❖ Created new columns using the issue\_d variable (issue\_d\_year and issue\_d\_month)
- ❖ Used binning technique to derive new variables that were required for the analysis

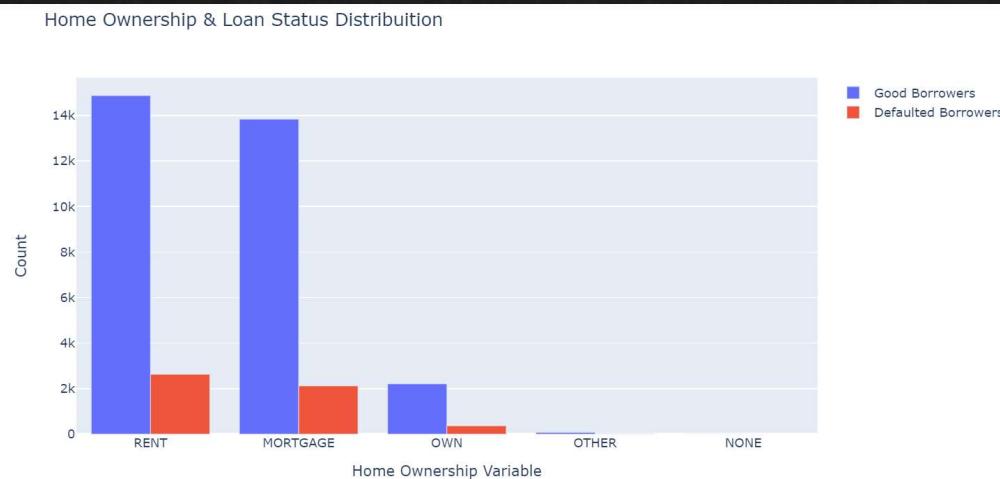
# Data Analysis

- ❖ Analysis of Loan status and finding relationship between the employee experience Vs Loan status.
- ❖ 15% of the total applications are defaulters
- ❖ Most of the loans are fully paid
- ❖ The number of applications is high for 10+year of employee experience.
- ❖ Also, upon checking the default rates, we see that employees with 10+year of experience are high.

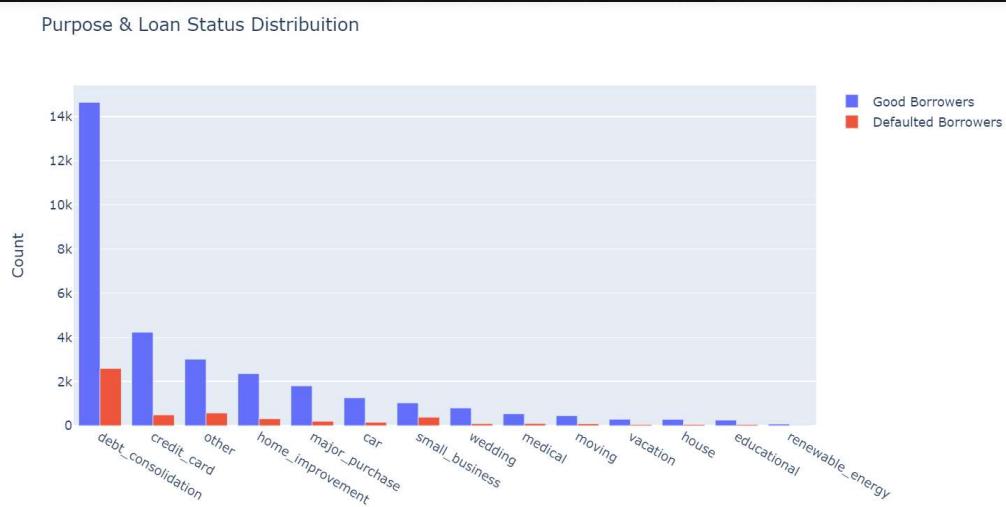


# Distribution of Loan status Vs Home ownership and Purpose

Home Ownership & Loan Status Distribution



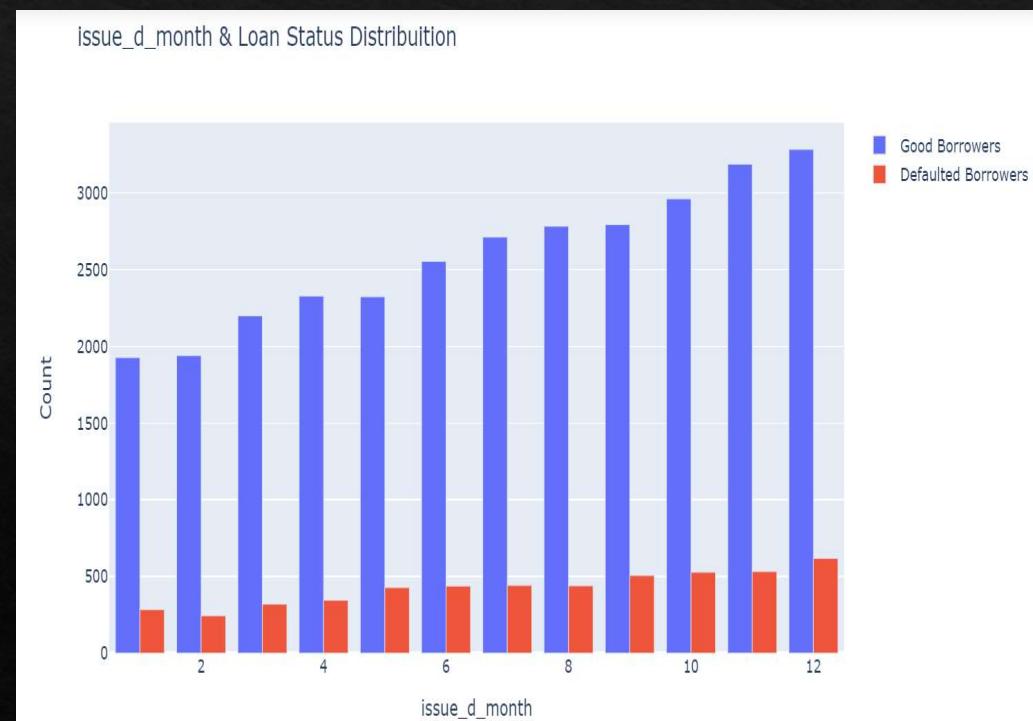
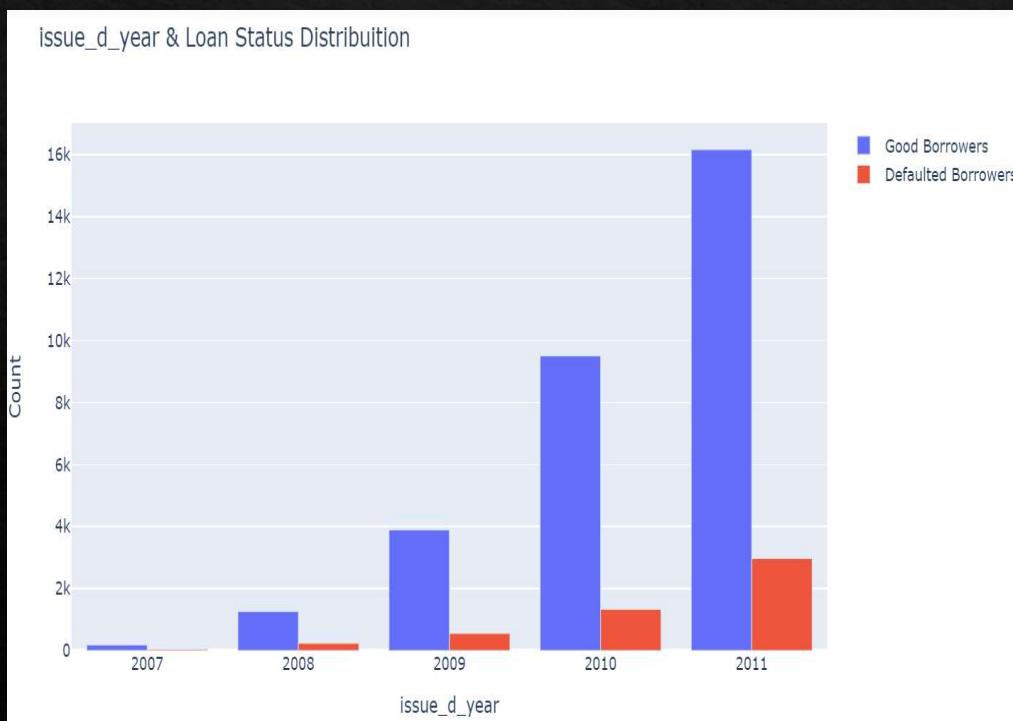
Purpose & Loan Status Distribution



- ❖ Applicants who live in rent or mortgages homes tend to default more.
- ❖ Out of the total count of the defaulters nearly 50% of applicants do not own a house and are rented.
- ❖ And approx., 45% of applicants are having mortgage.
- ❖ Similarly, we notice the maximum number of loan requests are for debt\_consolidation, credit\_card, other and small\_business
- ❖ Among which the ratio between charged off and fully paid is high for small\_business.
- ❖ Also, many applications do not have annual incomes verified.

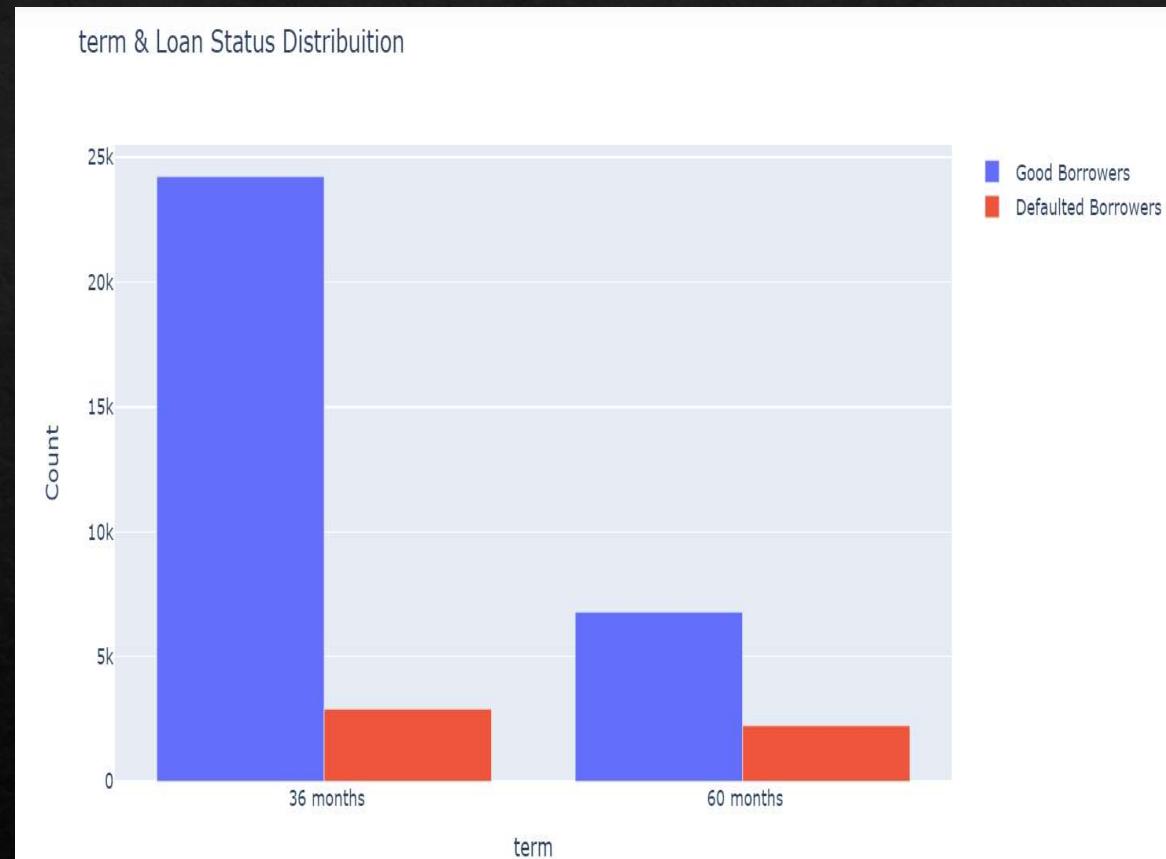
# Distribution of Loan status Vs Issue\_d

- ❖ The loan application has an exponentially increasing trend year on year.
- ❖ Also, a higher number of applications are funded during the year ends.

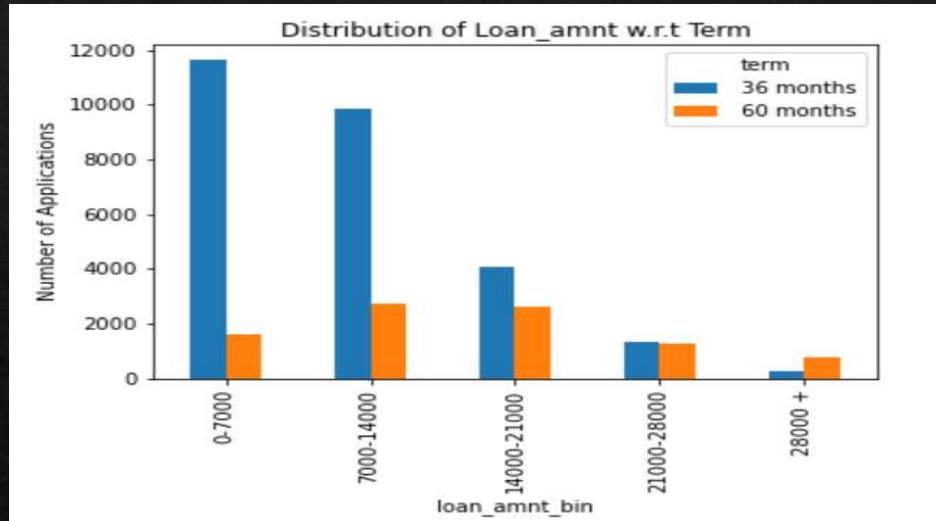


# Distribution of Loan status Vs Term

- ❖ Many applications are for 36-month term
- ❖ We see that only ~10% of the total applicants under 36-month tenure fall under the defaulter category.
- ❖ However close to 25% of the total number of borrowers who requested 60-month tenure are defaulters.

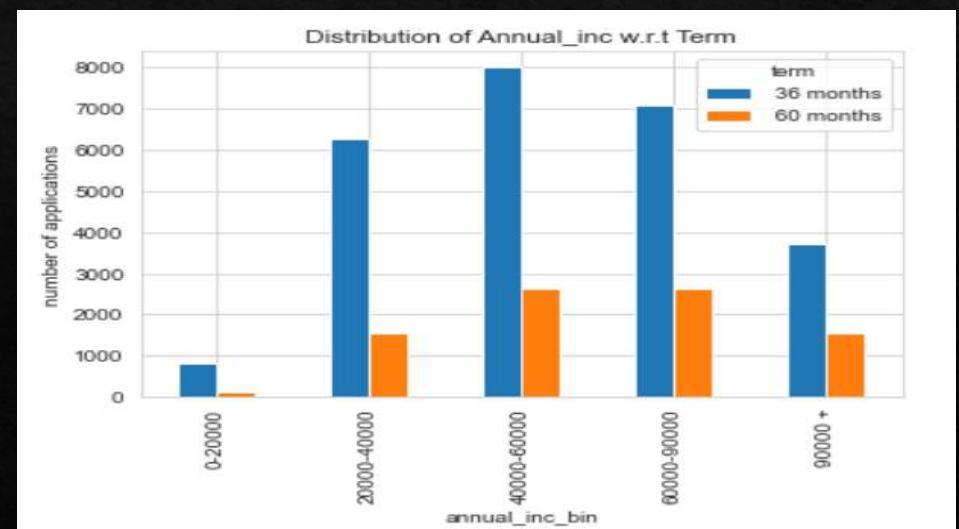


# Checking Term Vs different variables



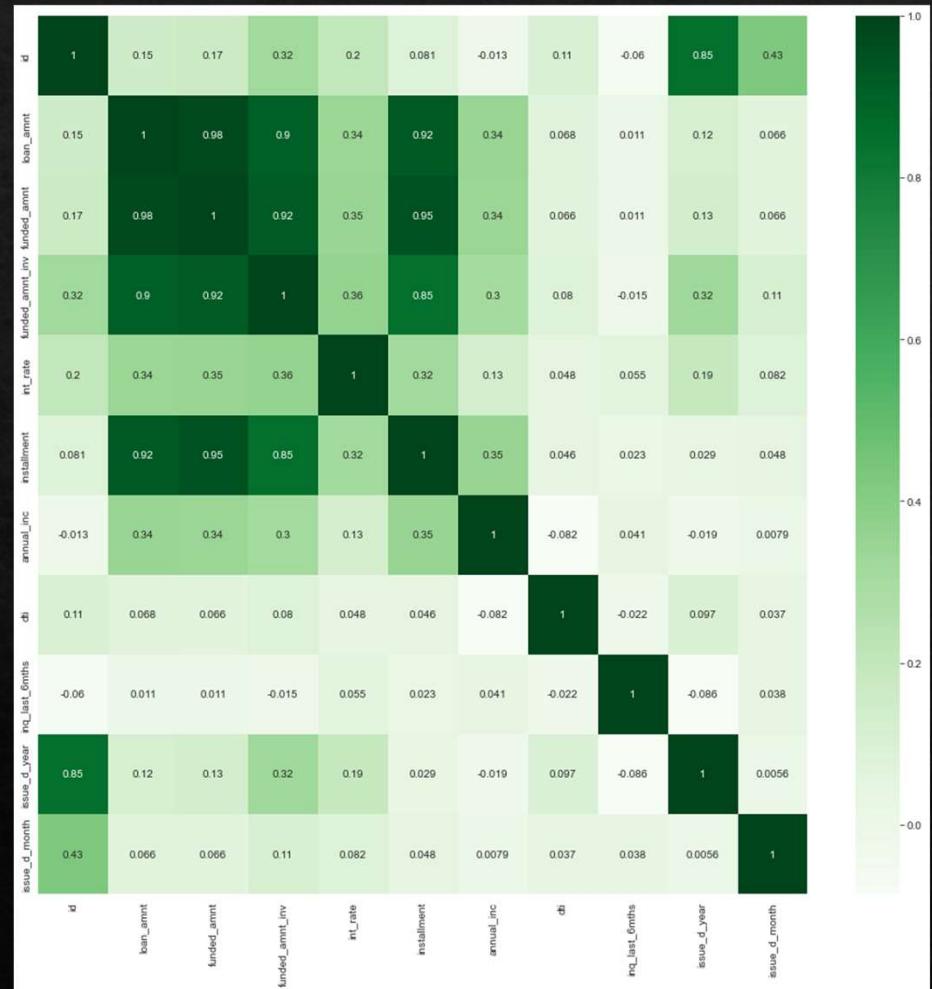
- ❖ Lending club to be cautious when applicants request for 36 months plan with
  - ❖ **Loan amounts** ranging between 0 and 14000
  - ❖ **Annual incomes** in ranges 40000-60000

- ❖ Lending club to be cautious when applicants request for 60 months plan with
  - ❖ **Loan amounts** ranging between 14000 and 28000
  - ❖ **Annual incomes** in ranges 40000-60000, 60000-90000



# Correlation map

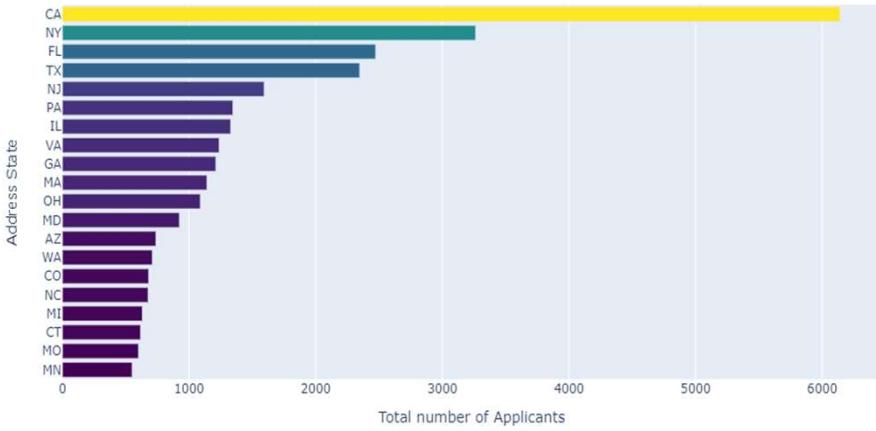
- ❖ There is high correlation between loan\_amnt, funded\_amnt and funded\_amnt\_inv
  - ❖ Inverse correlation between dti and annual income, this relationship will help us determine if the applicant can pay the loan and manage existing debts.
  - ❖ We can infer the following from the heat map: lower the dti indicates higher annual income implying the applicant would pay the loan. On the other hand, when the dti is high, it indicates low income which may lead to applicant becoming a defaulter.
  - ❖ dti and inq\_last\_6mnths have negative correlation, indicating that the higher the dti, the lesser inquires and we need to be cautious when we come across such combinations.



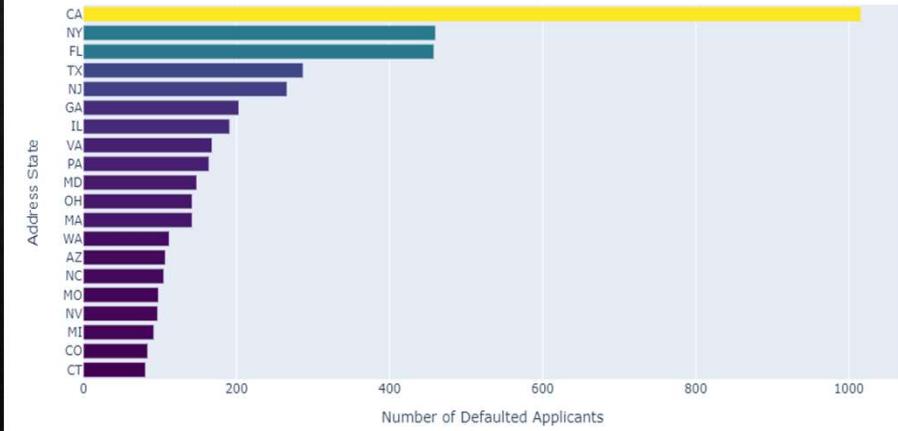
# Distribution of Loan status Vs State address

- ❖ We notice that maximum Loan applications are from CA, NY and FL.
- ❖ Also, these top 3 states are having high default rates.

Top 20 Address State By Number of Total Borrowers

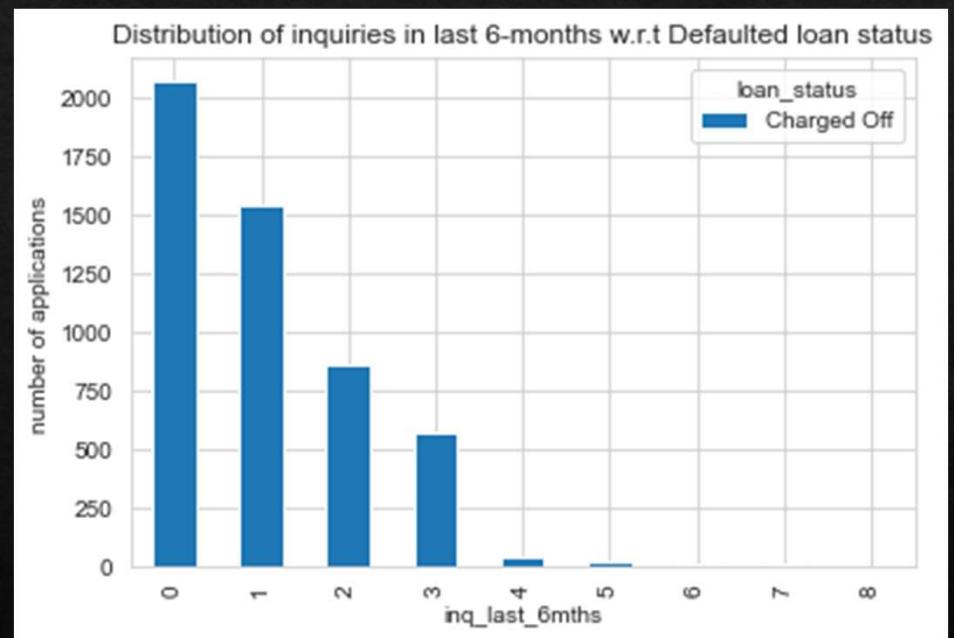


Top 20 Address State By Number of Defaulted Borrowers



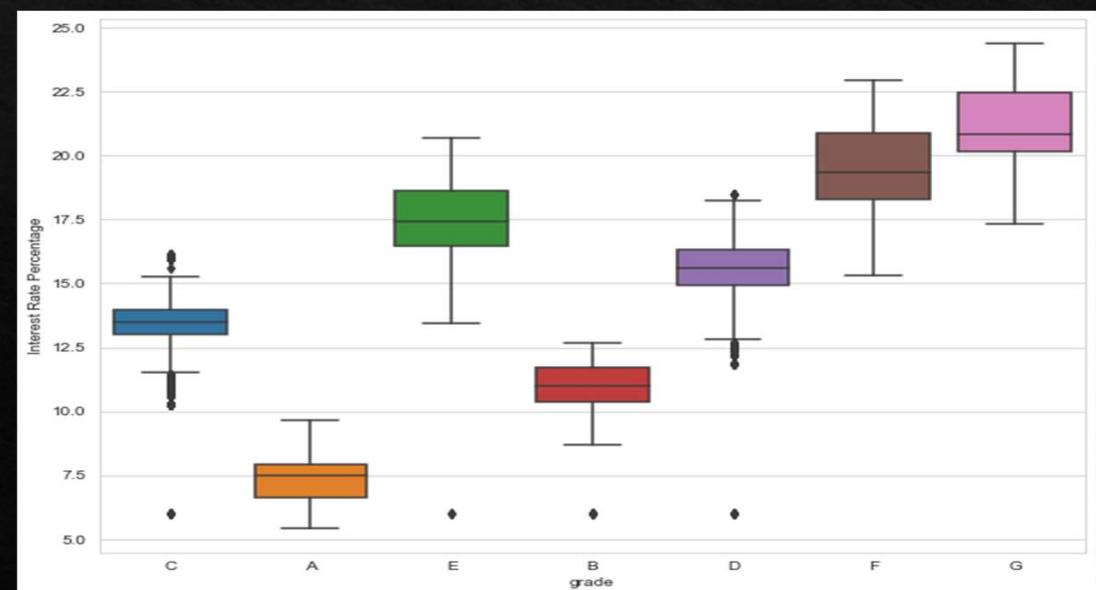
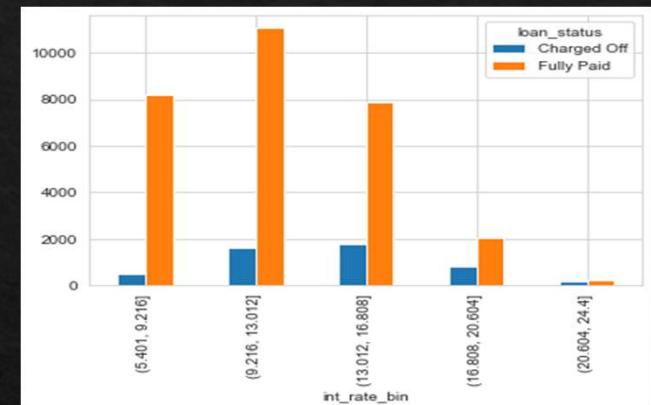
# Distribution on Defaulted Loans Vs inquiry made in last 6 months

- ◊ There is a decreasing trend on the number of defaulters based on the number on inquires, as the inquiries increase the defaulter rate decreases.



# Distribution of Interest Rate with Loan Status & Grade

- ◆ Default rates are high for int\_rates ranging between 9 and 16(approx.,)
- ◆ Grades A and B have the highest loan application requests.
- ◆ Higher loan amounts were requested with high quality loans. [grades A and B]
- ◆ Higher grades have lower interest rate <12% (approximately)



# Recommendations

- ❖ Close to 15% of the total applicants are defaulters.
- ❖ Lending club needs to be more cautious on the below factors:
  - ❖ Purpose
    - ❖ Debt consolidation and Small business
  - ❖ Home ownership
    - ❖ Rented and Mortgage with loan amounts as below
      - ❖ loan amount = 5000 to 13000 under RENT category
      - ❖ loan amount = 10000 to 25000 under MORTGAGE category
  - ❖ Employment length
    - ❖ Employees with 10+years of experience
  - ❖ Verification Status
    - ❖ All applications need to have annual incomes verified.

# Continuation..

- ◊ DTI rates and Annual incomes.
  - ◊ Applicants with higher dti and low annual incomes tend to default.
- ◊ Inquiries in last 6 months
  - ◊ Less number of inquires indicate more dti rates which further indicate lesser annual incomes, implying that the borrower could default.
- ◊ Term
  - ◊ Borrowers who request for loan amounts between 0 and 14000 with 36 months tenure tend to default more.
  - ◊ Also, borrowers having annual incomes in ranges, 20000-40000, 40000-60000, 60000-90000 tend to default more.
- ◊ State
  - ◊ Applicants from CA tend to default more compared to other states.
- ◊ Grades and int\_rate
  - ◊ High quality grades are funded more. High grades imply low interest rates, hence grading system is working fine.
  - ◊ Higher interest rate have higher tendency to default the loan.