

### Lending Club Case Study

By- Vishakha Singh

Tuesday, May 3, 2022

Developed as part of the Exploratory Data Analysis Module required for Executive PG Programme in Machine Learning and AI - IIIT, Bangalore.

#### **Problem Statement**

A consumer finance company which specializes in lending various types of loans to urban customers receives loan applications, 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

We need to identify indicators which can be used for accepting or rejecting the Loan Application



#### **Problem Analysis**

- •We have historical data for applicants who have successfully paid the loan, defaulted and are currently in process of paying loan
- •Need to find indicators for detecting loan defaulters

**Current State** 

### To Fill gap between current and future state

- As a data scientist we need to do exploratory data analysis on the available data
- •Study the pattern of defaulters
- Based on EDA result Create a framework to identify future defaulters from the applicants

- •Indicators are Provided based on EDA
- •Defaulters are detected based on the indicators

**Future State** 

### **EDA Approach**

- Distribution of features is plotted on histogram
- Skewness is identified
- Create derived variable

- Correlation matrix fro numerical variables
- Analysed Effect of two variables

Data cleaning and understanding

Univariate Analysis Segmented univariate analysis

Bivariate Analysis

Insights and recommendations

- Data is expanded and studied for available columns
- Datatypes, missing values are corrected
- Outliers are detected
- Standardization of data

 Analysed aggregated metric across segments  Strong indicators of defaulter results published

#### Data Understanding

### Categorization of variables happened based on two categories-

- Pre and post loan acceptance variables
- We considered pre loan accepted variables for our analysis as this information would be present at the time of loan application

### Pre Loan acceptance variables

```
loan amnt
funded amnt
funded amnt inv
term
int rate
installment
grade
sub grade
emp length
home ownership
annual inc
verification status
issue d
purpose
addr state
dti
pub rec
pub rec bankruptcies
```

### Post loan acceptance variables

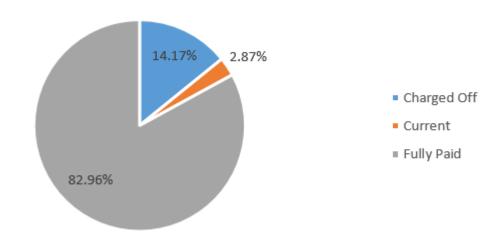
```
deling 2yrs
earliest cr line
inq last 6mths
open acc
pub rec
revol bal
revol util
total acc
out prncp
out prncp inv
total pymnt
total pymnt inv
total rec prncp
total rec int
total rec late fee
recoveries
collection recovery fee
last pymnt d
last pymnt_amnt
last credit pull d
application type
```

# Univariate and segmented Univariate analysis

#### Loan status distribution

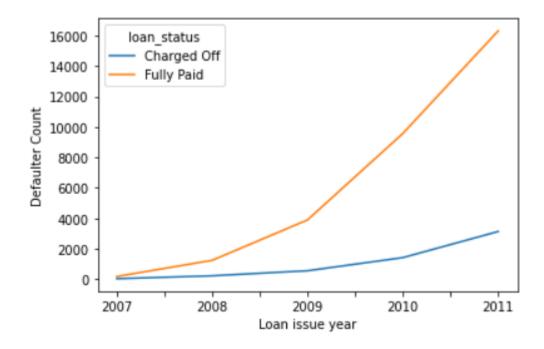
There are ~14% defaulters in the data set

Applicants distribution across loan status

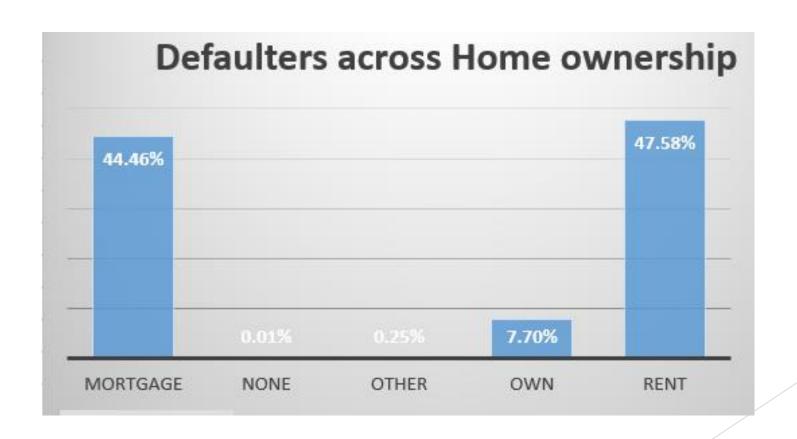


### Analysis across years

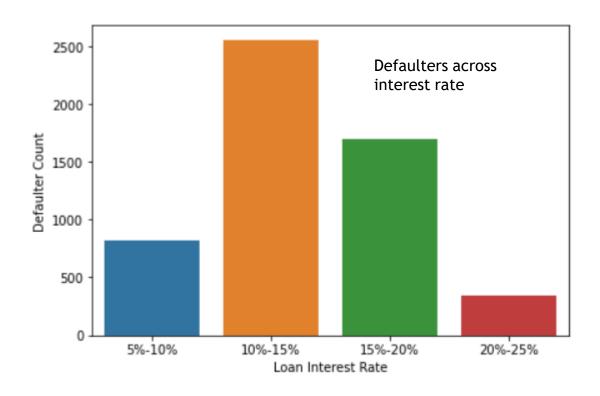
- Data is spread across the years 2007 to 2011
- Around 58% loans issued in year 2011 resulted in default



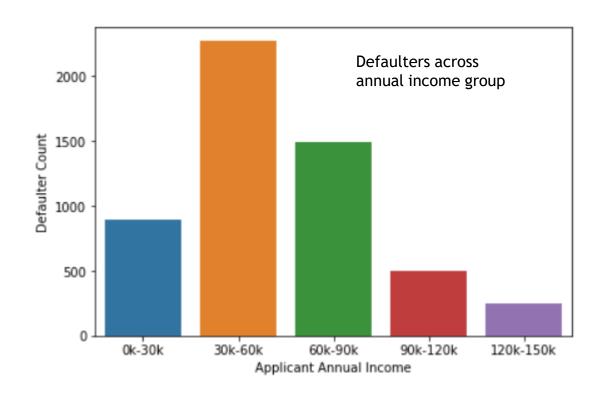
Around 92% applicants having home ownership status as rented or mortgage were defaulters



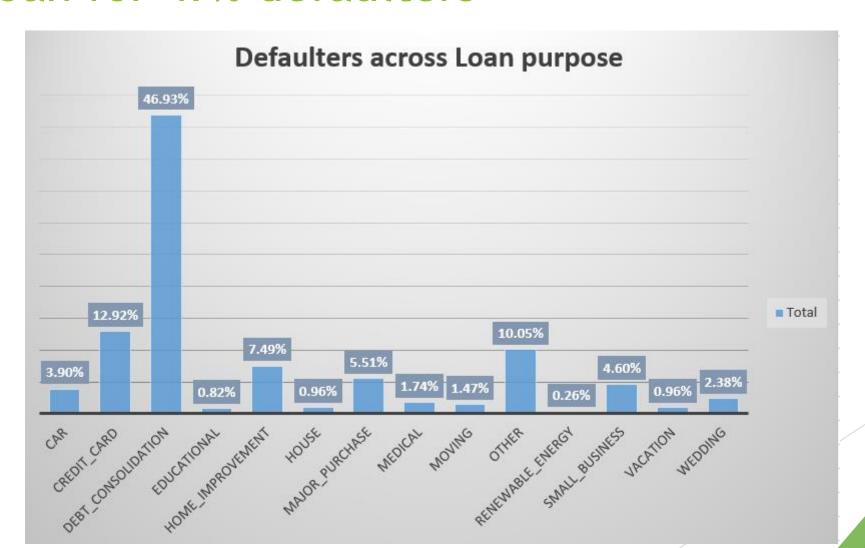
## Around 46% Applicants who had an interest rate in range 10%-15% defaulted



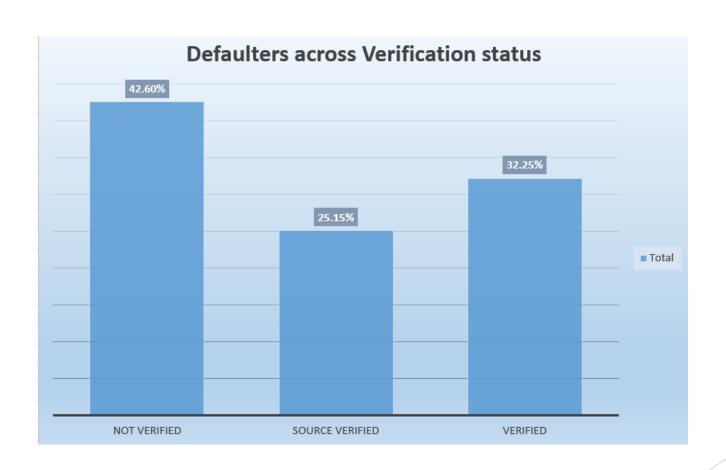
## 42% Defaulters had an annual income in range 30k-60k



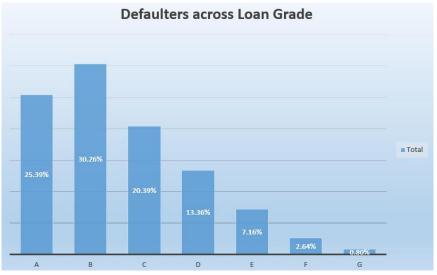
### Debt\_consolidation was the purpose of loan for 47% defaulters

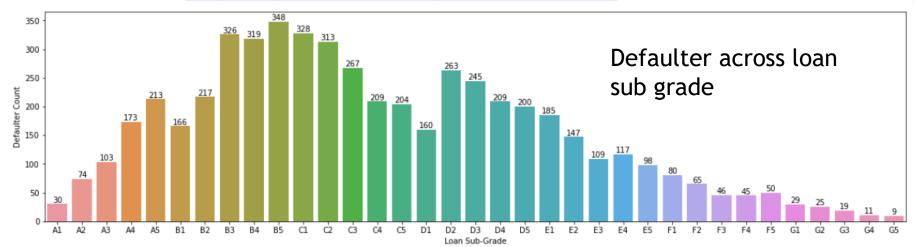


### 42% Defaulters had not have their stated income verified

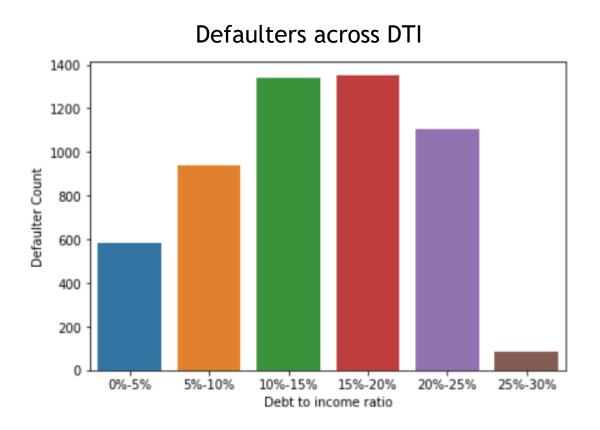


### Loan Grade B had ~30 % of the defaulters, under which B5 was the worst performer

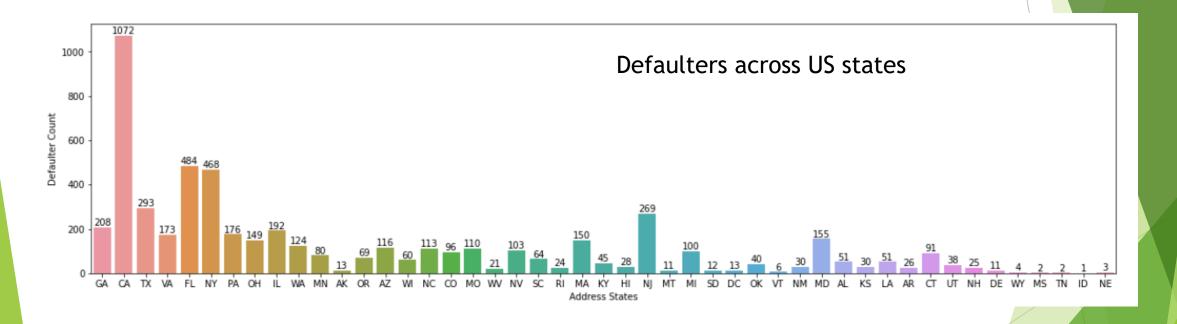




## Around half of the total defaulters had Debt to income ration in range 10%-20%

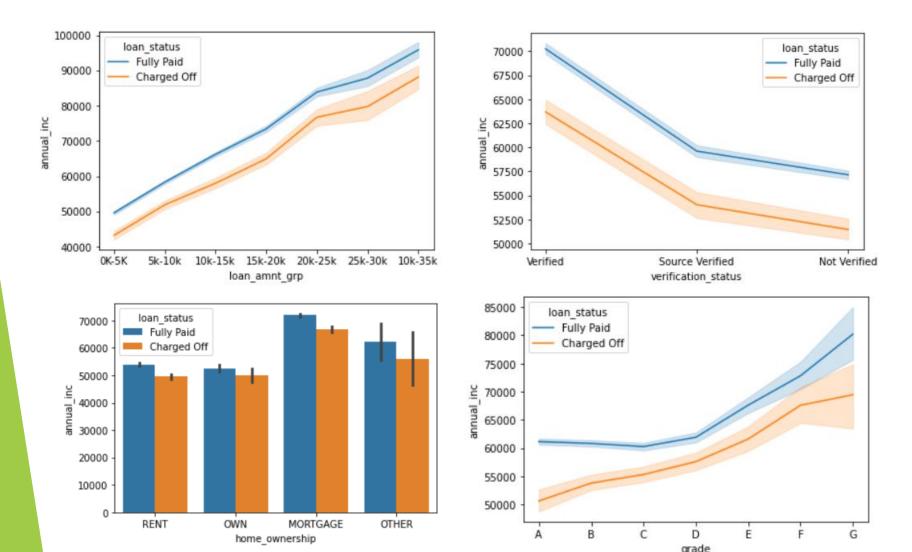


### ~20% defaulters are from California state



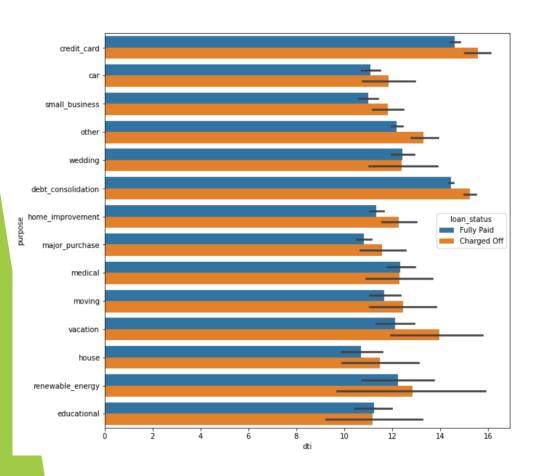
### **BIVARIATE ANALYSIS**

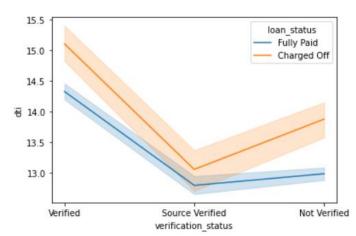
#### Mean Annual Income across various variables is less for defaulters compared to "fully paid" customers

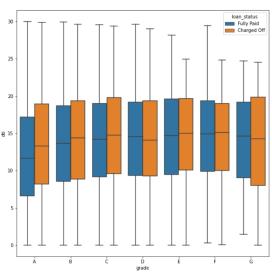


Mean Debt to income ratio across various variables is high for defaulters compared to

"fully paid" customers

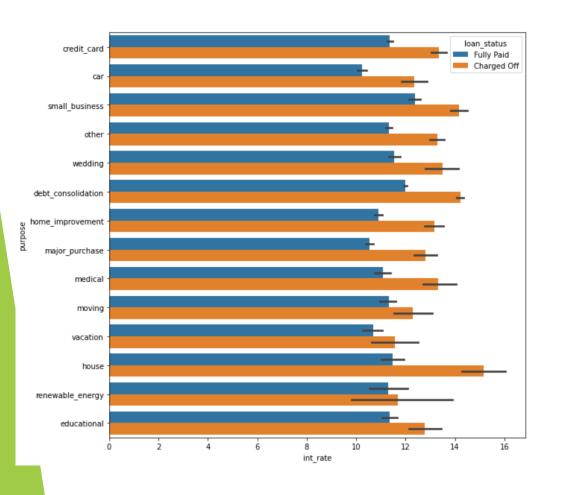


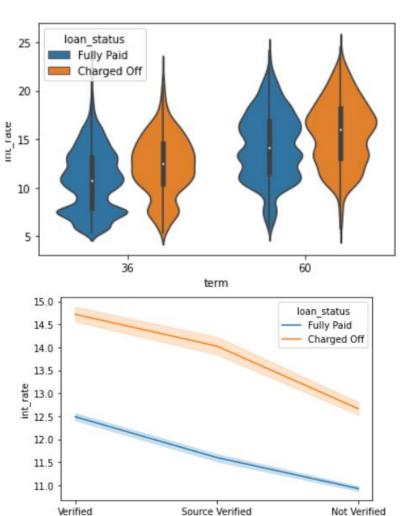




Mean interest rate across various variables is high for defaulters compared to "fully paid"

customers





verification status

### Recommendations

#### Insights and Recommendation

Look for the following indicators for new applicants who are likely to default

- income is not verified
- People with annual income in range 30k-60k
- Purpose of loan is debt\_consolidation
- ▶ Debt to income ratio 10%-20%
- Home ownership as RENT or MORTGAGE
- ► Loan grade assigned as B

Internal

Thank You!!!!!!!