

LENDING CLUB CASE STUDY SUBMISSION

Group Members:
Manoj Kumar Sharma
Adharsh Belagapu

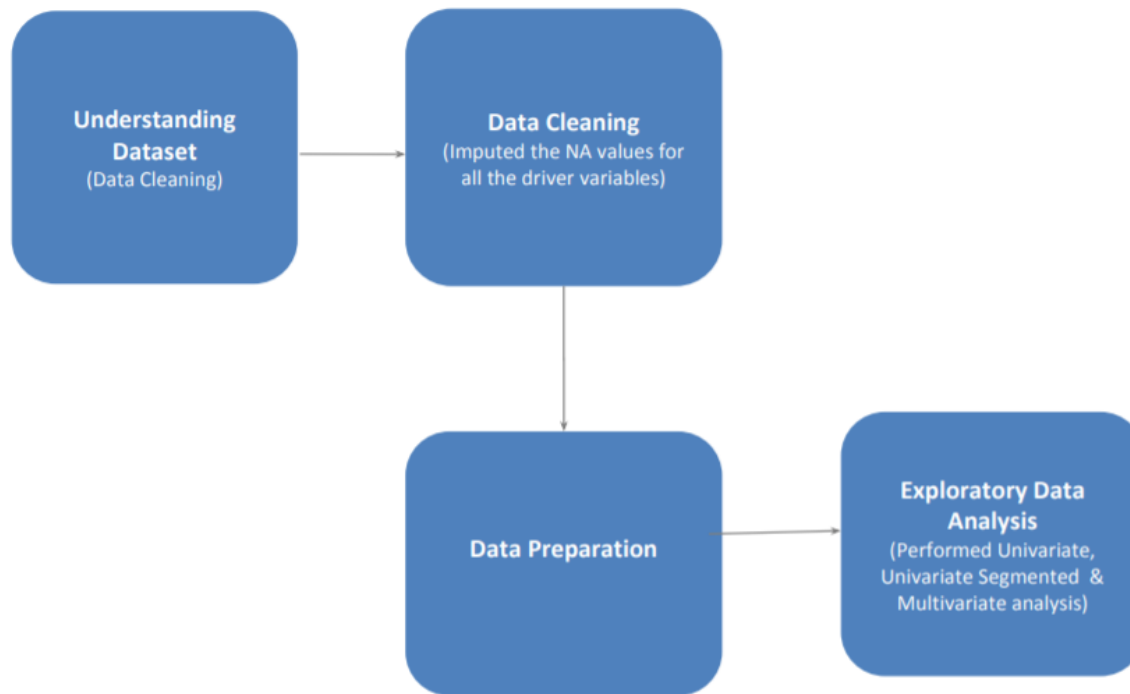
Objectives

Understand the 'Driving Factors' or
'Driver Variables' behind Loan Default

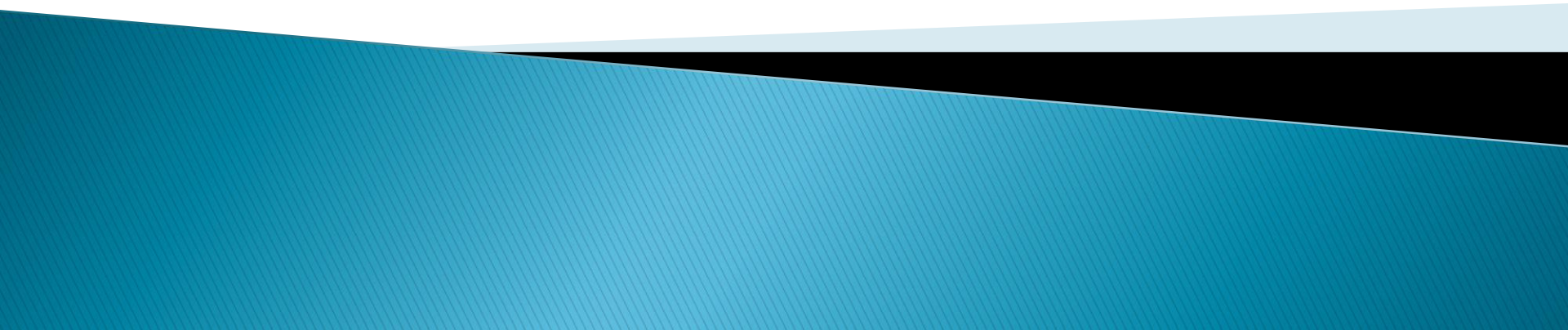
Loan Lending Organizations may choose to utilize
this knowledge for its risk assessment of new loan
applicants

Identify type of Loan applicant

Methodology Used: CRISP-DM



Data Cleaning Steps

- Delete columns: Delete unnecessary columns.
 - Remove outliers: Remove high and low values that may affect the results of analysis.
 - Missing values: Treat missing values with appropriate approach.
 - Duplicate data: Remove identical rows, remove rows where some columns are identical.
 - Filter rows: Filter by segment, filter by date period to get only the rows relevant to the analysis.
- 

Analysis

AIM of project was to analyze how customer variables affect in loan being defaulted

NA values were imputed

Few NA values were deleted from dataset

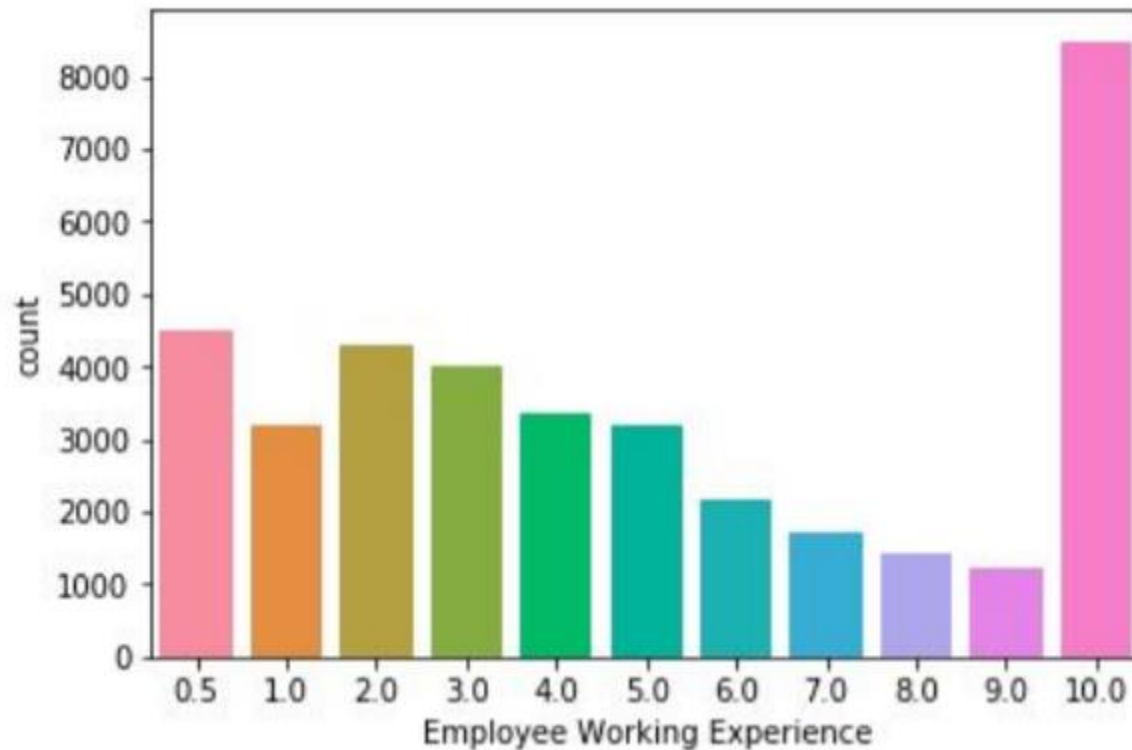
Some derived columns were prepared using existing columns

Histogram, Bar Charts and Box plots were used in EDA

Outliers were found using Box plots and were mentioned in the EDA

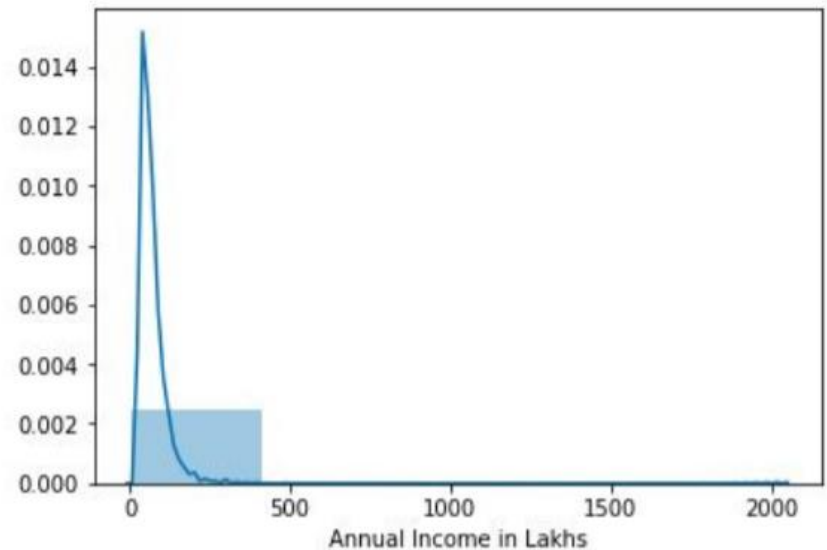
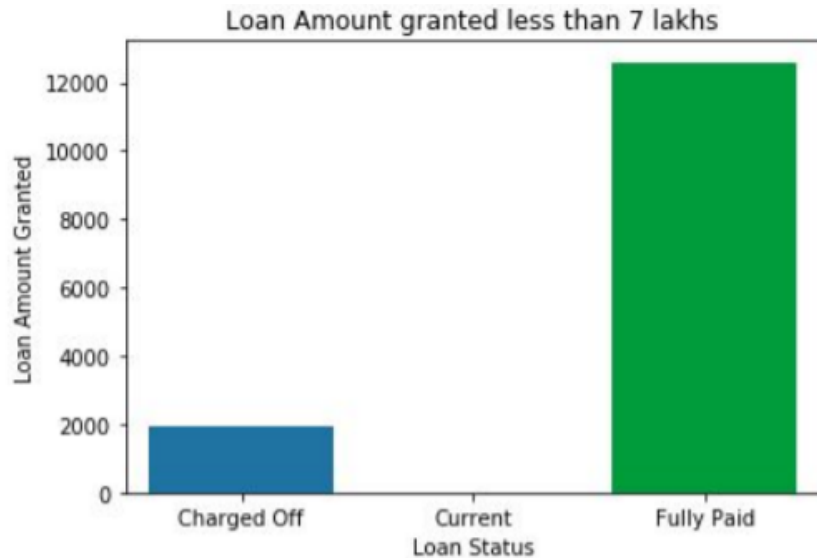
Bivariate and multi variate analysis was done of different columns

Work Experience analysis of Applicants



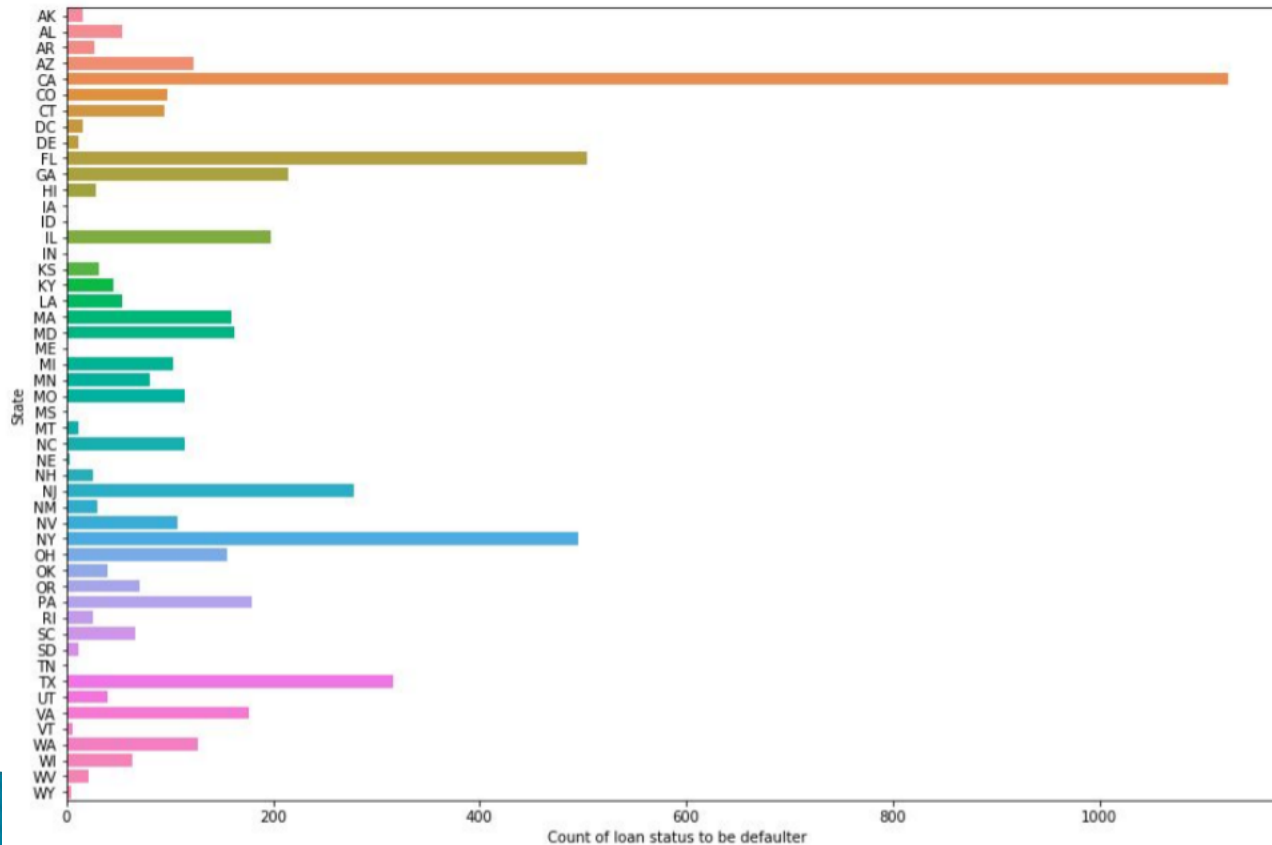
High percentage of employees having
10+ years of work experience

ANALYSIS: LOAN AMOUNT



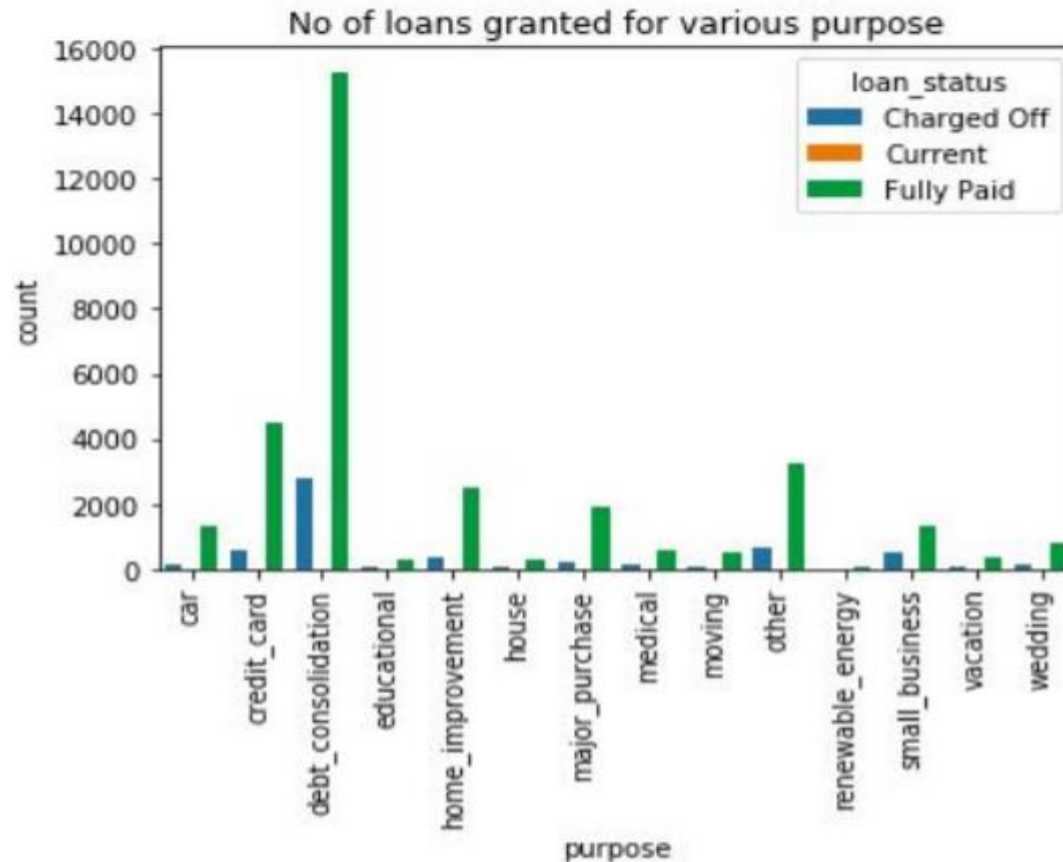
- Most of the loans are fully paid
- Majority of loan applicants have low annual income
- Very less loans are there which are charged off

Defaulters by state



➤ California, florida , Newyork and Texas have high probablity for default

Reason for loan



➤ Most loans were granted for Debt consolidation

Thank you