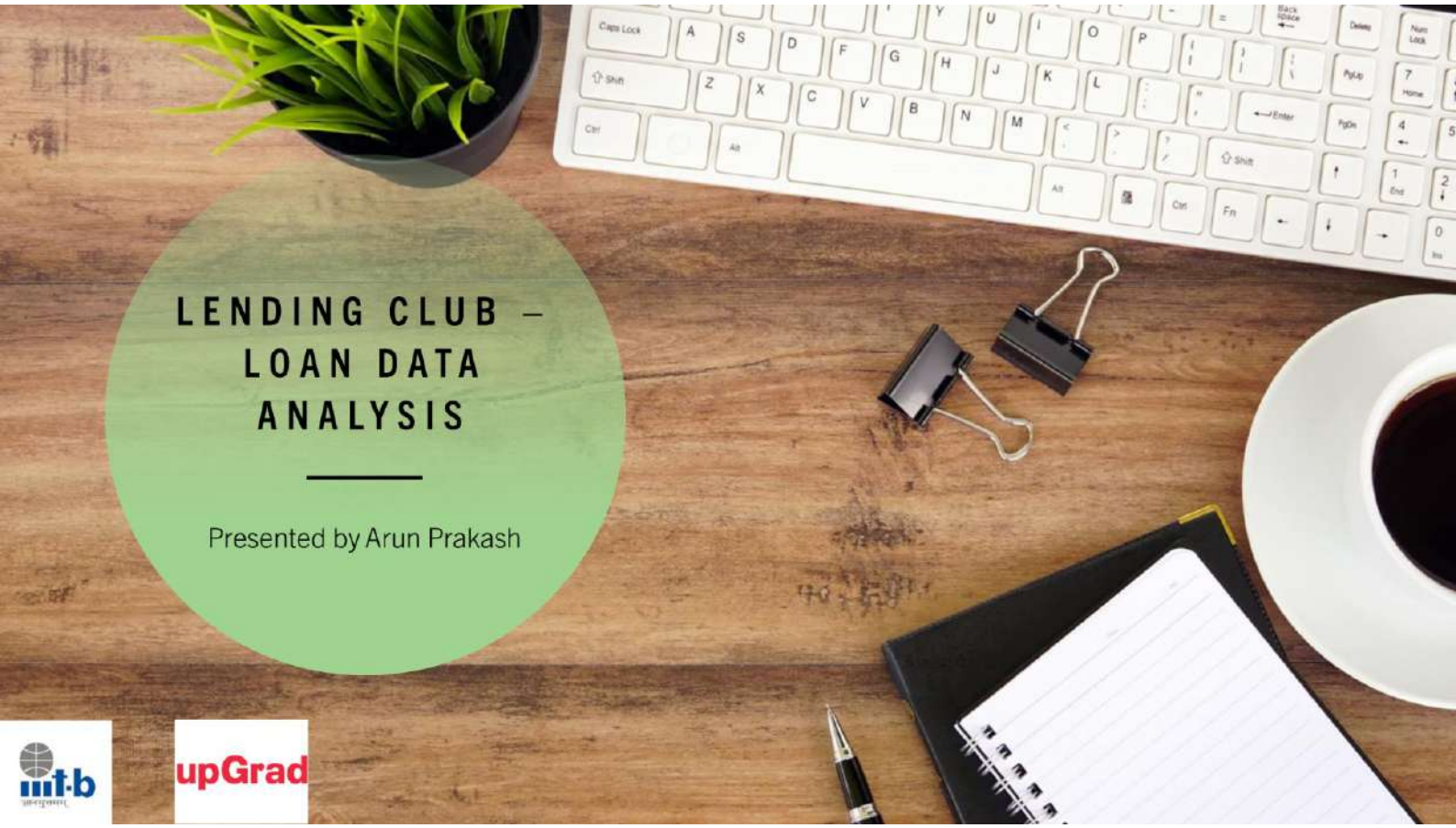

LENDING CLUB – LOAN DATA ANALYSIS
Presented by Arun Prakash
upGrad

AGENDA

- Objective and Problem Statement
- Approach
  1. Problem Solving
  2. Data analysis
- Business review of Statistical analysis
- Summary

OBJECTIVE & PROBLEM STATEMENT

- Objective of the case study was to identify risk applicants, what factors contribute to defaulting proportions and how can company leverage this information for portfolio and Risk assessment.

- Problem Statement: In the last three financial year (2008-11), on an average at 10% default rate at Lending company has increased . Over same period of time, the median annual income has remained same by increasing by ~5%

LOAN DEFAULT: POSSIBLE HYPOTHESES

- Fraud : Over last 3 years the possible cases of fraud over consumer purchases has increased
- Interest rate : The interest rate over the loan amount and tenure is high.
- Verification : The client selection, monitoring of application has typically dropped over 3 years
- Ticket size of Loan : With increased annual income, the possible increase in loan amount in last 3 years

APPROACH- UNDERSTANDING BUSINESS

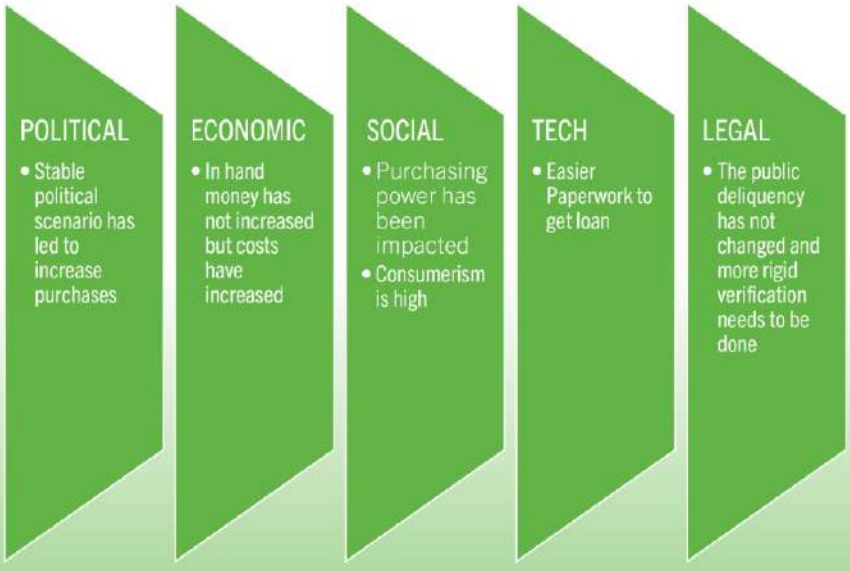
The typical consumer finance company is looking at building business model relying on past purchases to make a decision on whether to reject or select a loan application. For building this model ,they are looking to identify the attributes which impact the payment . Attributes which indicate tendency to default.

Why is the model Important?

1.Because it impacts the revenue bottom line of the company.

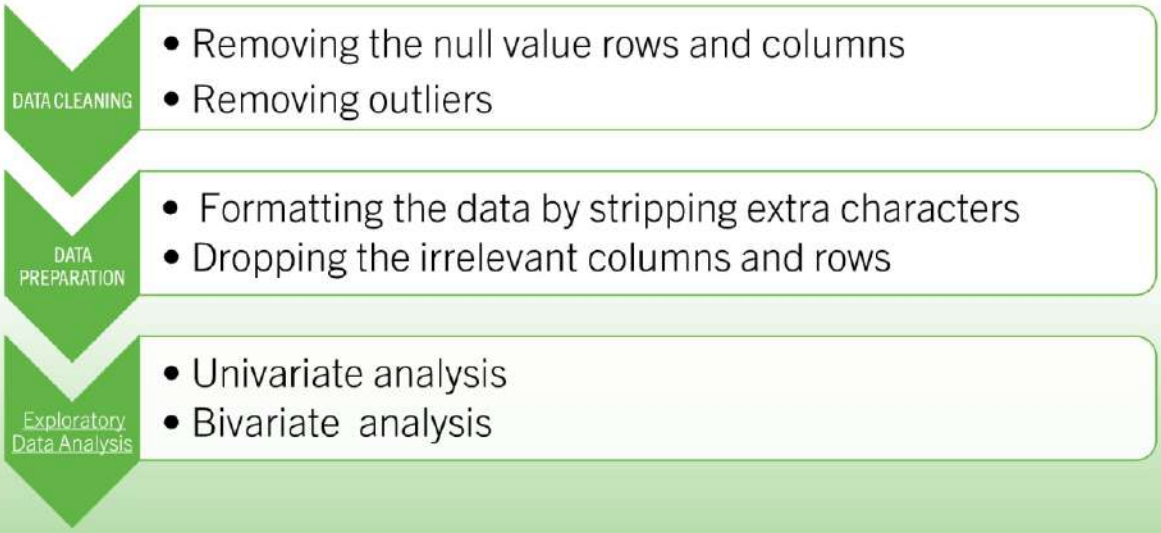
2. Because no finance company wants NPA on their accounts.

APPROACH- BUSINESS ANALYSIS



PESTEL Frame work Analysis –
Consumer Finance Lending Scenario

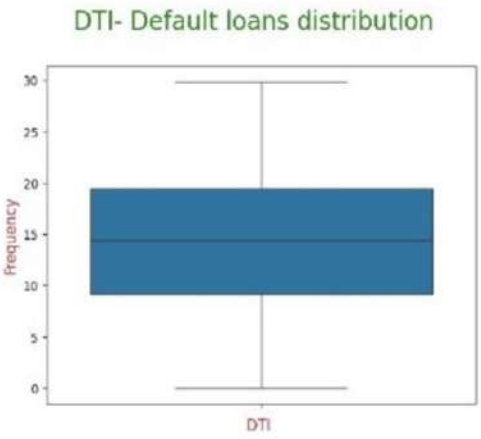
APPROACH- DATA ANALYSIS



upGrad

EXPLORATORY DATA ANALYSIS- BUSINESS REVIEW

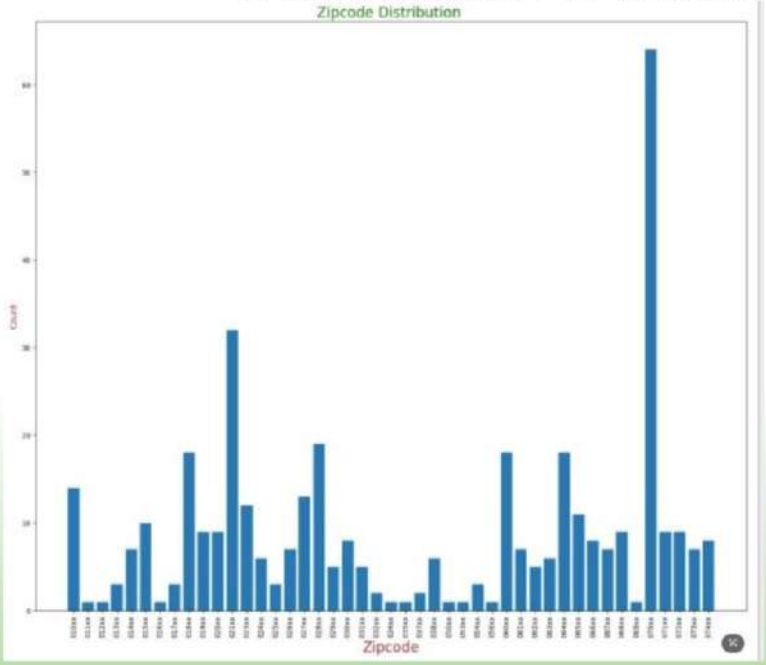
UNIVARIATE ANALYSIS-



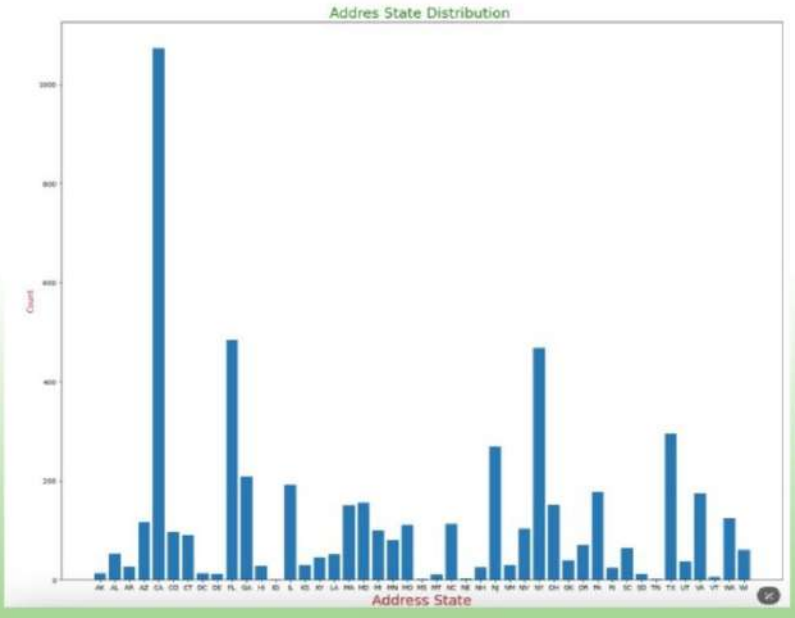
DTI

#As per industry standard good DTI should be less than 36% ,since the Loans have been dispersed with max dti of 30% it is good.

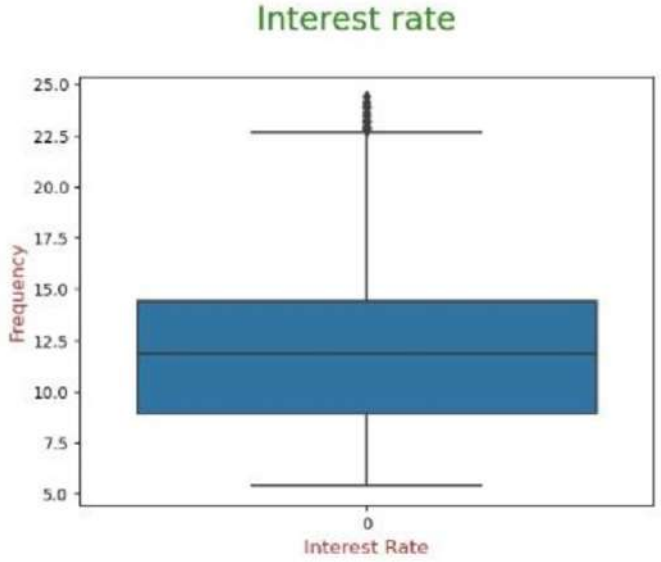
UNIVARIATE ANALYSIS-



Address State and Zipcode : Many of the defaulters are from CA and NY and from 070 zipcode



UNIVARIATE ANALYSIS-

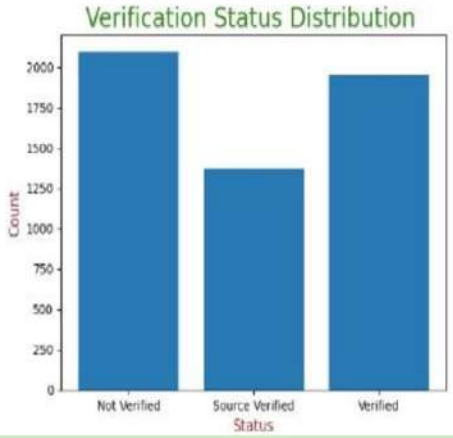


Interest Rate

#Most defaulters received interest at the rate of 11.5-17% against median interest of all borrowers of 10%

#Some defaulters received loans at more than 22.5%

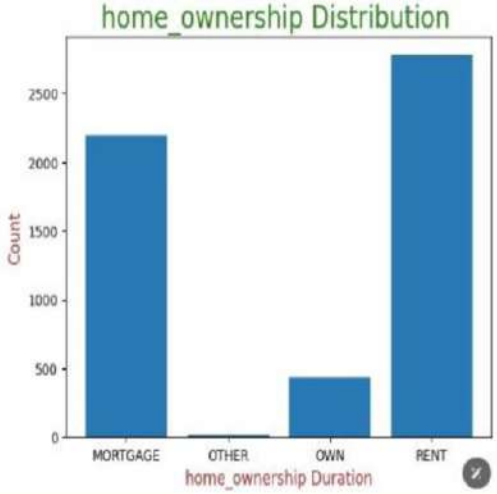
UNIVARIATE ANALYSIS-



Verification status

#Most defaulters were not verified for annual income.

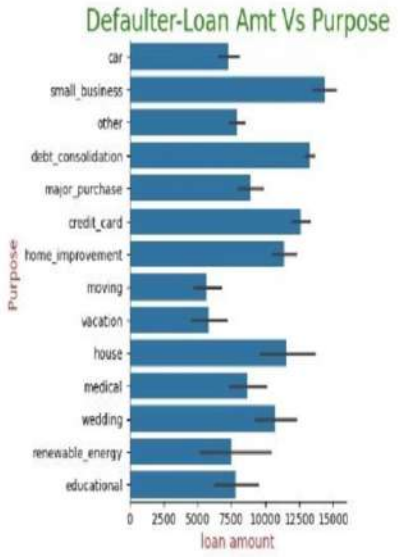
BIVARIATE ANALYSIS-



Home Ownership vs Loan Amount

In case of "Charged of Loans"- Most of the defaulters whose home ownership is 'MORTGAGE and RENT ".

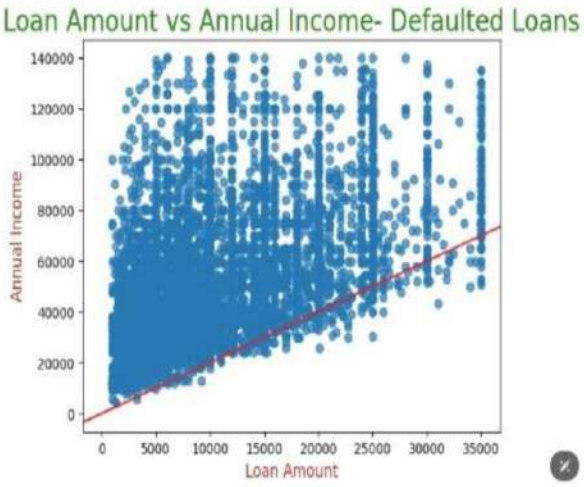
BIVARIATE ANALYSIS-



Purpose vs Loan Amount

IIn case of "Charged off Loans"- Most defaulters took loan had small business' or had to pay of Credit Card Debt.

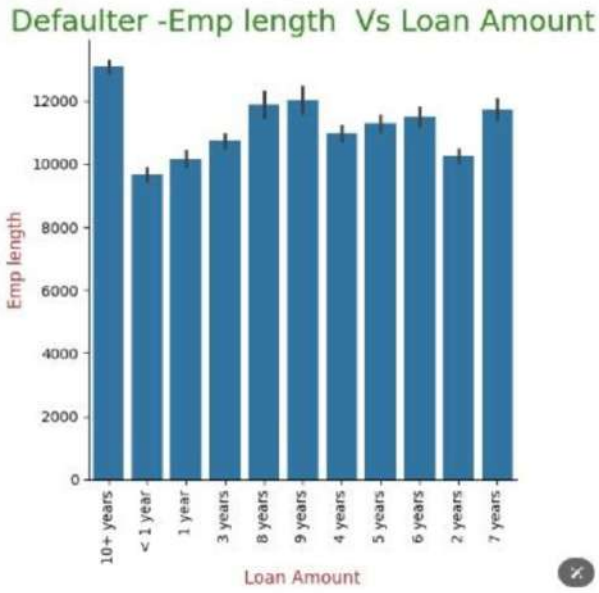
BIVARIATE ANALYSIS-



Annual Income vs Loan Amount

IIn case of "Charged off Loans"- Only high income people with salary higher than 23500$ took more than 30000$ loan amount.

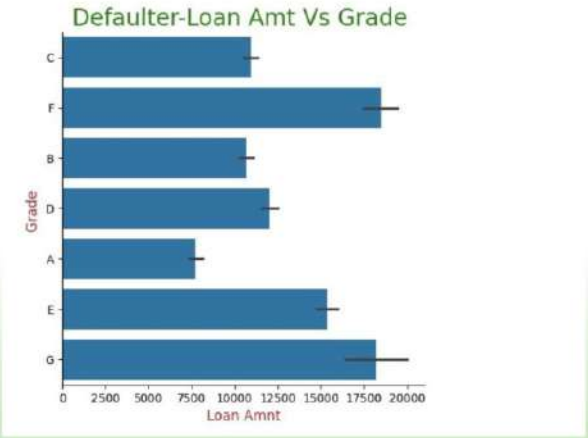
BIVARIATE ANALYSIS-



Employment Length vs loan Amount .

In case of Charged of Loans-Most defaulters have 10+yrs experience and loan amount is 14k+

BIVARIATE ANALYSIS-



Grade vs Loan Amount

In case of "Charged off Loans"- Most defaulters were grade F&G employees who took loan of upwards of 17500$

CONCLUSION

- Key driver variables behind loan default are

1. Loan Amount requested and Application date
2. Debt to Income Ratio
3. Purpose
4. Employment Length
5. Applicant residing State and Zipcode

CONCLUSION-

- Observations

1. Most defaulters took loan amount between 5000-15000$ at interest rate of 10-15%
2. Once median annual income crossed 57000$, less likely people are to go for loans exceeding 15000$
3. Most of borrowers were employees with 10+ years' experience based out of California 070 zipcode .
4. Lending Club should investigate borrowers who are renters or homeowners with a mortgage seeking loans for debt consolidation, as they have a higher chance of default, particularly in the fourth quarter of the year.
5. Most of loans were assigned without annual income verification

Thank
You!