

Loan Prediction

Group 1:

48	Prabhav Pandya	11910000
49	Omkar Prabhune	11910471
50	Pranav Tambaku	11910000
53	Pritesh Pawar	11910435



Problem Statement

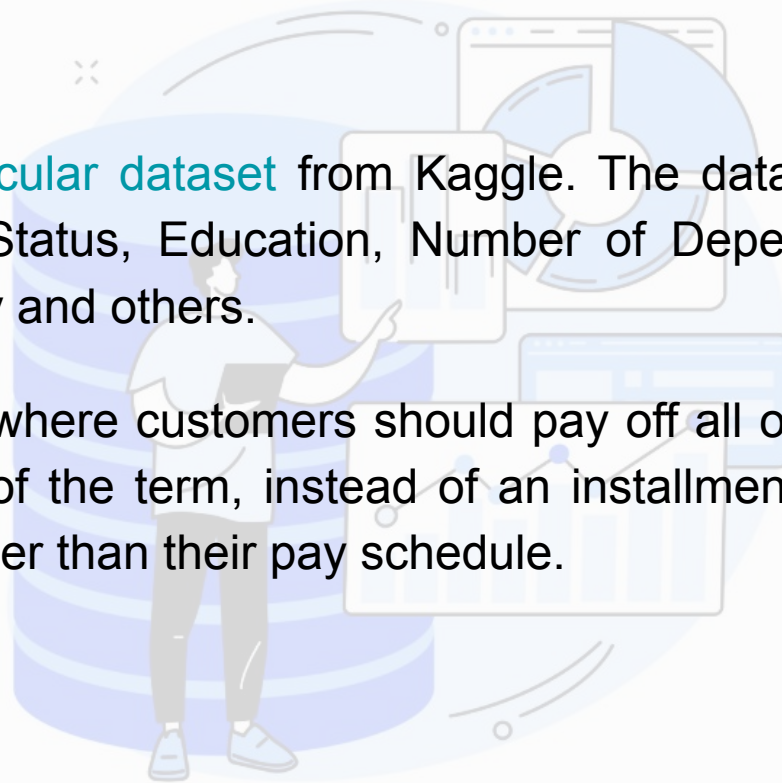
To automate the loan eligibility process based on customer detail provided while filling online application form. These details are Gender, Education, Income, Marital Status, Loan Amount and others. And to identify the customer segments, which are eligible for loan amount



Dataset

We sourced [this particular dataset](#) from Kaggle. The dataset includes variables like Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others.

The loan here is one where customers should pay off all of their loan debt in just one time by the end of the term, instead of an installment schedule. Of course, they could pay off earlier than their pay schedule.



Data Content

Categorical features(Nominal): These variables are categorical without order or ranking. Below are the list of such variables in this data:

- Gender,
- Married,
- Self_Employed,
- Credit_History

Categorical features(Ordinal): These variables in categorical in nature having some order involved.

- Dependents,
- Education,
- Property_Area



Data Content

Numerical features(Ratio): These features have numerical values.

- ApplicantIncome
- CoapplicantIncome
- LoanAmount
- Loan_Amount_Term

Analysis

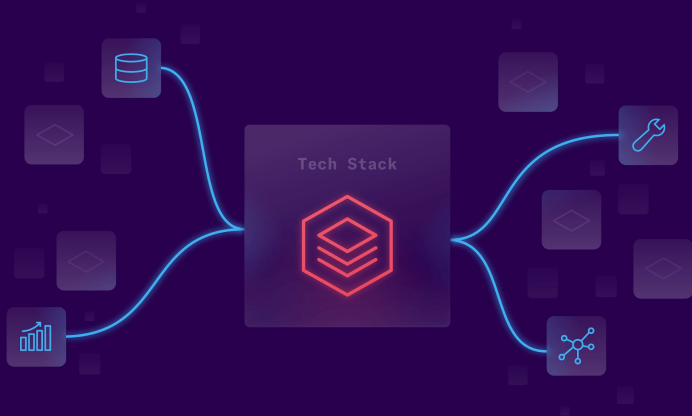
We plan on making two main types of analyses in this project.

First to explore the dataset by finding correlations between data (Exploratory Data Analysis)

Second we plan to use Regression techniques to predict whether a loan will be paid or not by the applicant



Techstack



Language Used: R

Dataset Format: csv (Comma Separated Values)

Libraries needed:

1. ggplot2
2. caret

Thank You