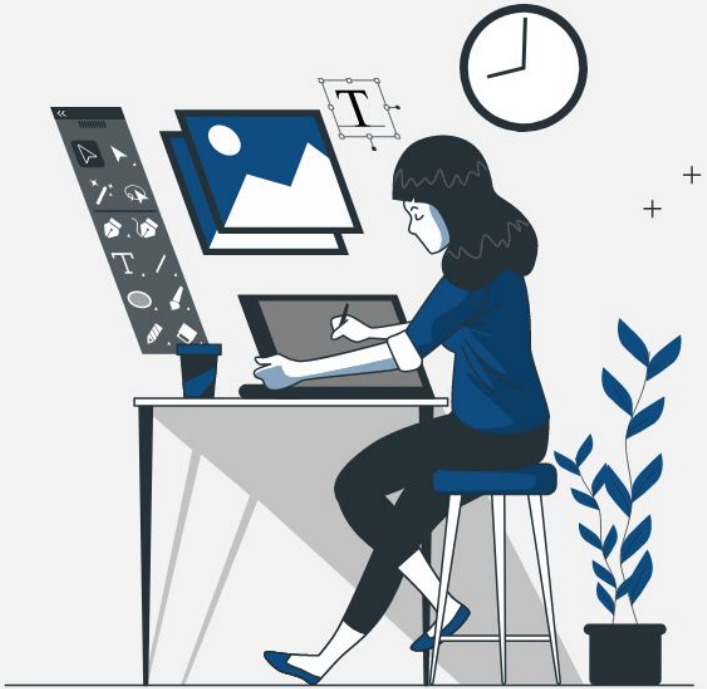
LENDING CLUB

CASE STUDY PRESENTATION

Understanding the meaning of data and Processing



01. ABOUT THE PROJECT

02. Data understanding

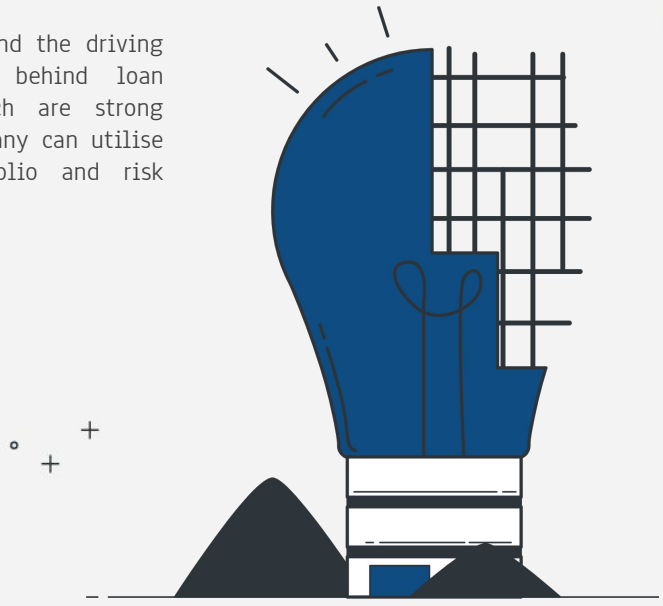
03. DATA CLEANING

04. Data Analysis

ABOUT THE PROJECT

PROJECT PROBLEM STATEMENT

The company wants to understand the driving factors (or driver variables) behind loan default,i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.



BUSINESS UNDERSTANDING

You work for a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, 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

WHAT WE ARE WORKING ON

Data understanding:

Understanding the meaning of data, Find out any quality issues in data

Data Analysis

Univariate and segmented univariate analysis.
Business-driven, type-driven and data-driven metrics.
Bivariate analysis.
Plots for the analysis.

Data Cleaning

Cleaning process for data such as adding missing values and outlier treatment.
Standardisation of data

02

DATA UNDERSTANDING

POINTS ANALYSED

Data Overview:

Analysed the dimensions of the dataset (number of rows and columns).
Examined the first few rows of the dataset to get a sense of the data structure.
Identified the types of variables (categorical, numerical, etc.).

Loan Attributes:

Investigated on loan-specific attributes such as loan amount, interest rate, term, etc.
Analyzed the distribution of loan amounts and interest rates.
Explored the relationship between loan attributes and the likelihood of default.

Target Variable Analysis:

Identified the target variable, which is likely to be 'default' in this case.
Analysed the distribution of the target variable (countplot/bar chart).
Understanding the percentage of defaults compared to non-defaults.

DATA QUALITY ISSUES

01

'funded_amnt_inv' some of the data in this column is not in standard format

02

'emp_title' there is empty data

03

There is lot of fields which is don't have any usefull value except NA so we can avoid these columns

04

Data in some fields are not in the standard form like most of the data is in number format but some of them are in decimal so we need to round of the decimal

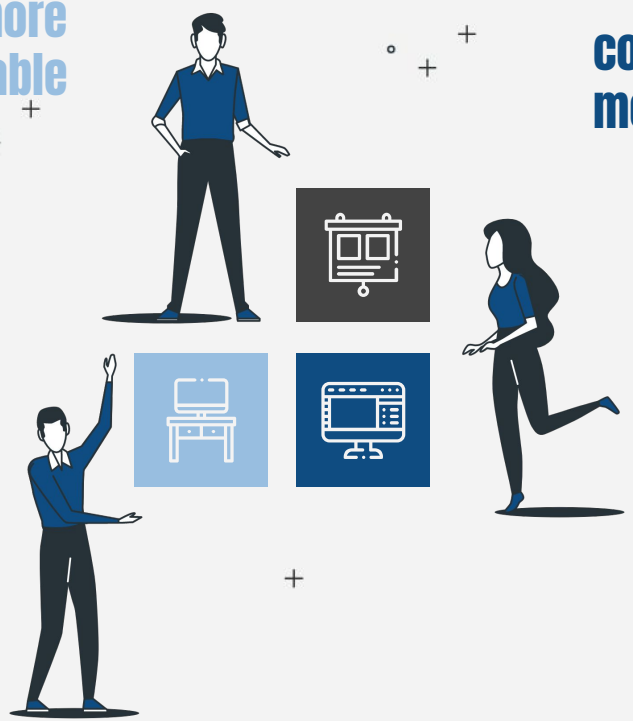
05

Finding Outliers in the Loan Amount column using IQR

DATA CLEANING PROCESS

Removing the columns with more values are not available

conversion of terms from month to number



standardising the funded_amnt_inv column

Finding Outliers in the Loan Amount column using IQR

04

DATA ANALYSIS

You can enter a subtitle here if you need it

TECHNIQUES USED

UNIVARIATE ANALYSIS

CATEGORICAL ANALYSIS

BUSINESS DRIVEN ANALYSIS

BIVARIATE ANALYSIS

CATEGORICAL AND NUMERICAL ANALYSIS

PREDICTED RESULTS

- Points derived Using Univariate analysis
  - people who owns home took less loans
  - Most of the loan terms are for 36 months and the most approved loan amount is 5,000
  - It is identified that people from the state with code CA have took more loans and also most of the loan applicants are people who have income of 40000 to 60000
- Points derived Using Bivariate analysis
  - it is analysed from the Interest Rate to Installment Ratio vs. Loan Amount' that the interest to installment ratio is less for higher loan amounts
  - From Relationship between Loan amount and terms (Filtered by Interest Rate) it is identified that for higher amount of loan the interest rate is less
  - From Scatter Plot: Loan amount and Verification ratio it is clear that people who have a verification badge got more loan amount
  - From Income vs Loan chart from this it is identified that people who have income between 30000 to 40000 took more loanand the most approved loan amount in between 0 to 10000
  - From Committed amount and Amount committed by investors it is identified that most part of the committed amount has been funded by the investors

THANKS!...