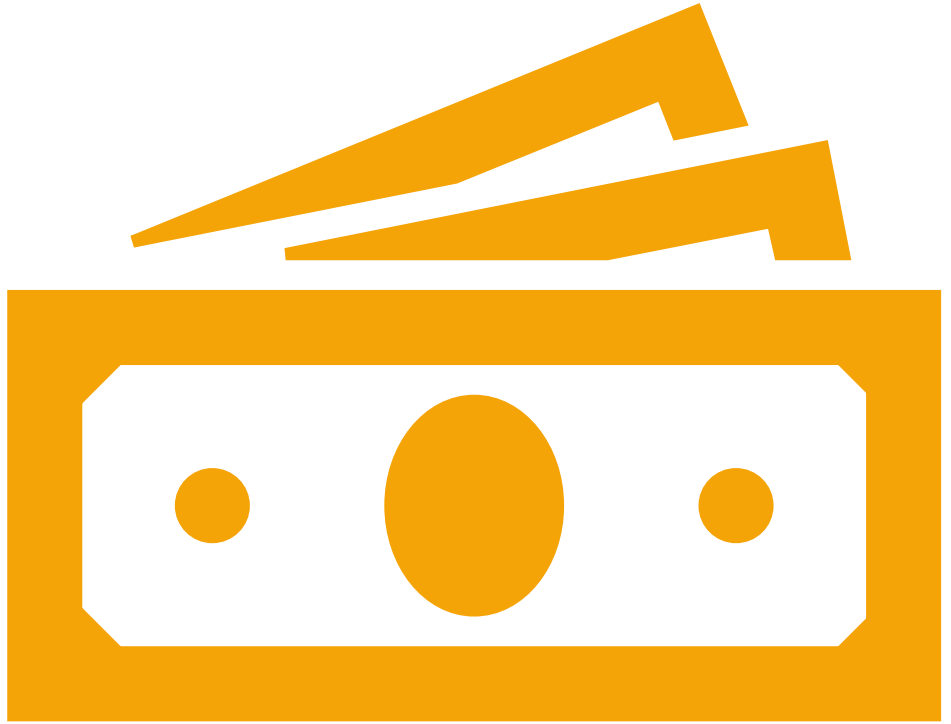
Lending Club Case Study

- Prabhat Sharma
- Brij Bhushan Paliwal



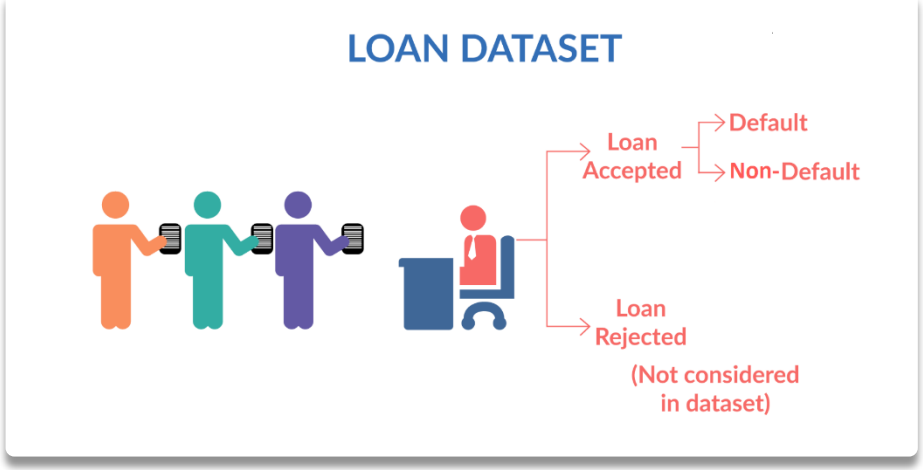
Objective

- The objective of the study is to provide the company with factors that help in analyzing loan applications and minimizing business loss and/or financial loss.
- At the end of the study, driving factors/variables are identified which help analyze a loan application.
- Below decisions can be taken based on the analysis:
  1. Rejecting the loan
  2. Reducing the amount of loan
  3. Lending at a higher rate of interest

Data Set

The dataset provided by the company to perform the analysis is:

- **loan.csv** – Loan Data from 2007 to 2011 (39717 rows * 111 columns)
- **Data_Dictionary.xlsx** – Meaning of the variables in loan.csv



Data Cleaning

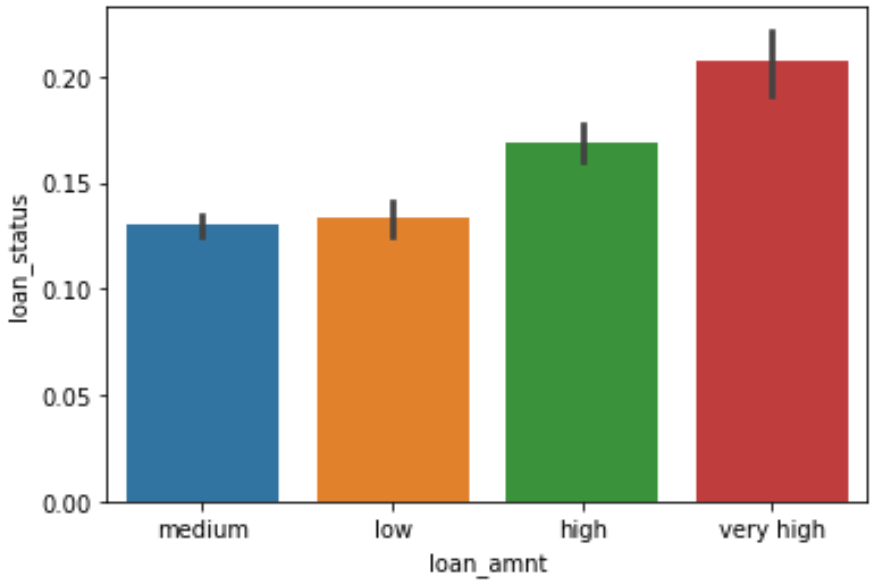
- We checked how many columns had missing values and the percentage of missing values in them.
- We then removed all the columns with more than 90% of missing values – 56 columns.
- We then got rid of behavior related and other columns of no use to the analysis.
- We then removed the current ongoing loans as they are not useful to the analysis.

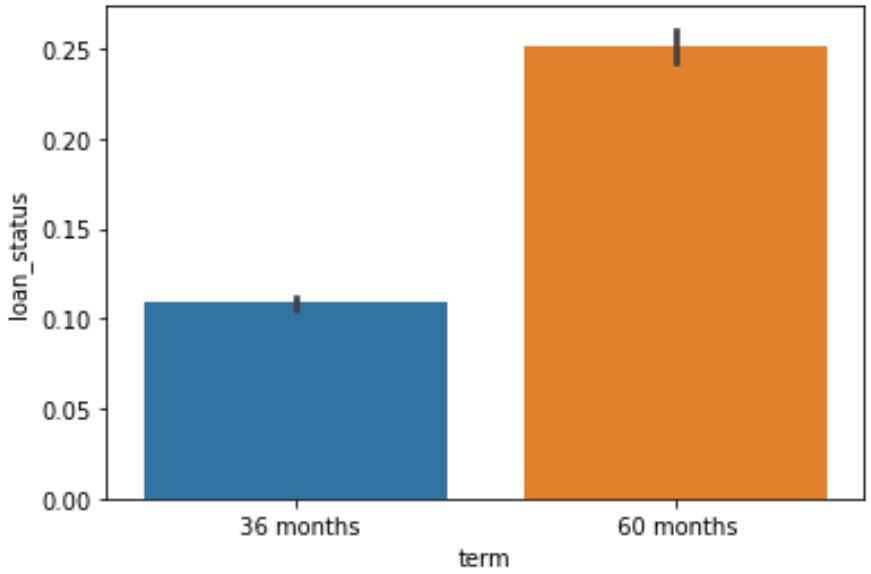
Univariate Analysis

- We plotted bar plots for default rates with different categorical variables such as :
  a. Grade
  b. Subgrade
  c. Term
  d. Home ownership
  e. Verification status
  f. Purpose
  g. Loan Amount(Low, Medium, High, Very High)
  h. Date(Year, Month)

  and so on…

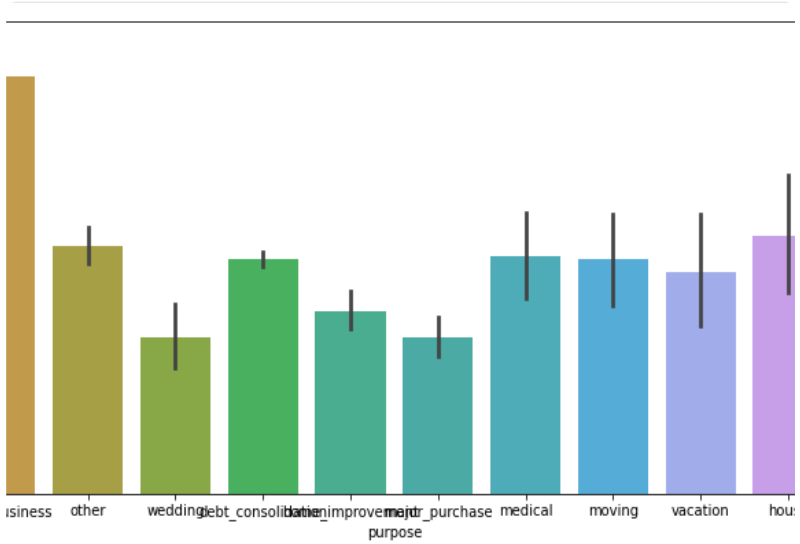
Univariate Analysis

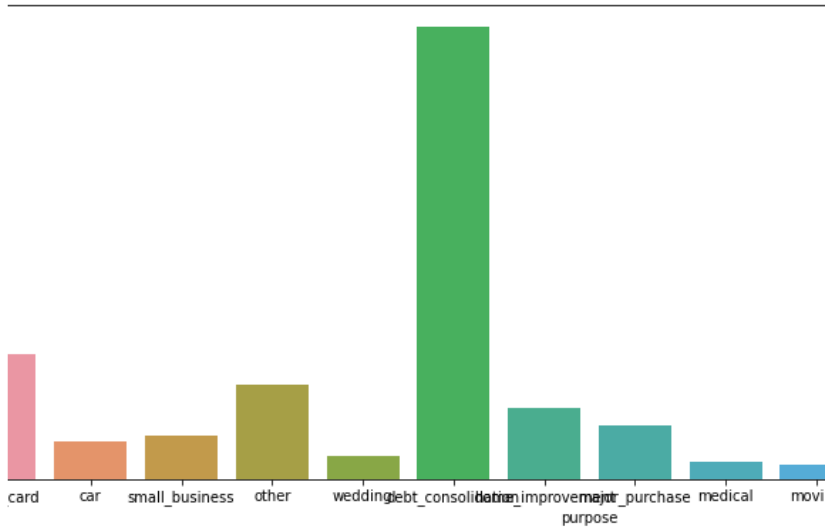




Univariate Analysis

- Purpose of the loan seems like an important factor affecting the default rate. We will now divide the purpose variable into segments and analyse further.

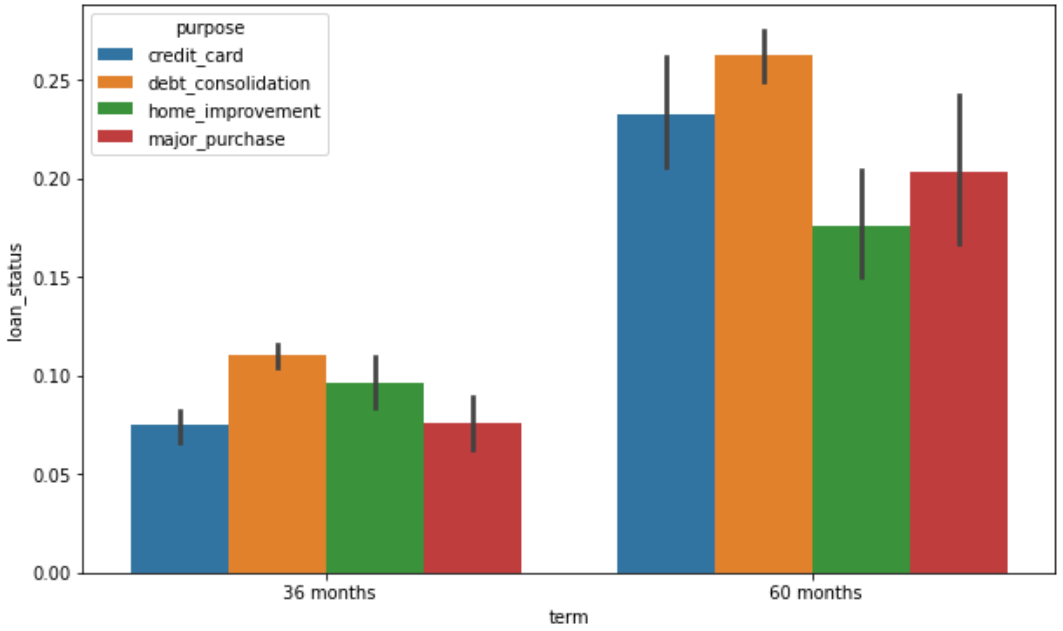


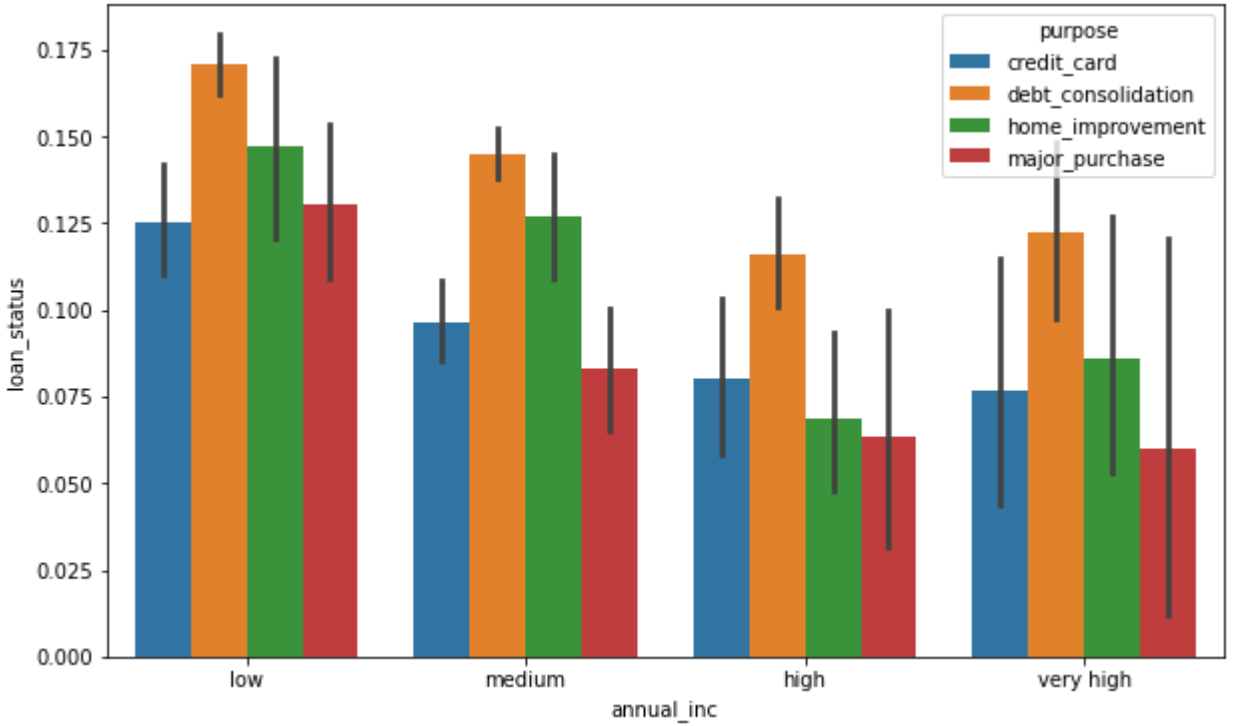


Segmented Univariate Analysis

- After performing univariate analysis, we identified few important variables such as purpose of the loan, grade, interest rate, etc. that impact the default rates the most.
- We then divided **purpose of the loan** into **4 major segments** to analyze further:
  a. Consolidation
  b. Credit Card
  c. Home Improvement
  d. Major Purchase
- Now, we plotted graphs with default rate, the above segments and other important variables

Segmented Univariate Analysis





Inference

As per the exploratory data analysis performed on the data set, below are the **driving factors/variables/columns** that will help the company anlayze each new loan application:

['id', 'member_id', 'loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'term', 'int_rate', 'installment', 'grade', 'sub_grade', 'emp_title', 'emp_length', 'home_ownership', 'annual_inc', 'verification_status', 'issue_d', 'loan_status', 'pymnt_plan', 'desc', 'purpose', 'dti', 'mths_since_last_delinq', 'initial_list_status', 'collections_12_mths_ex_med', 'policy_code', 'acc_now_delinq', 'chargeoff_within_12_mths', 'delinq_amnt', 'pub_rec_bankruptcies', 'tax_liens', 'month', 'year']

Based on the analysis, a decision can be made whether a loan should be accepted or rejected and if accepted, what should be the interest rate and loan amount, etc.