[Rajesh Bisht]
[Partner – Bhavesh bihani]

Lending Club case study

Introduction

By this case study we need to understand the real time problem scenario and how we can analyse the problem by the help of EDA . By EDA analysis we need to minimize the risk in banking industry

Understanding

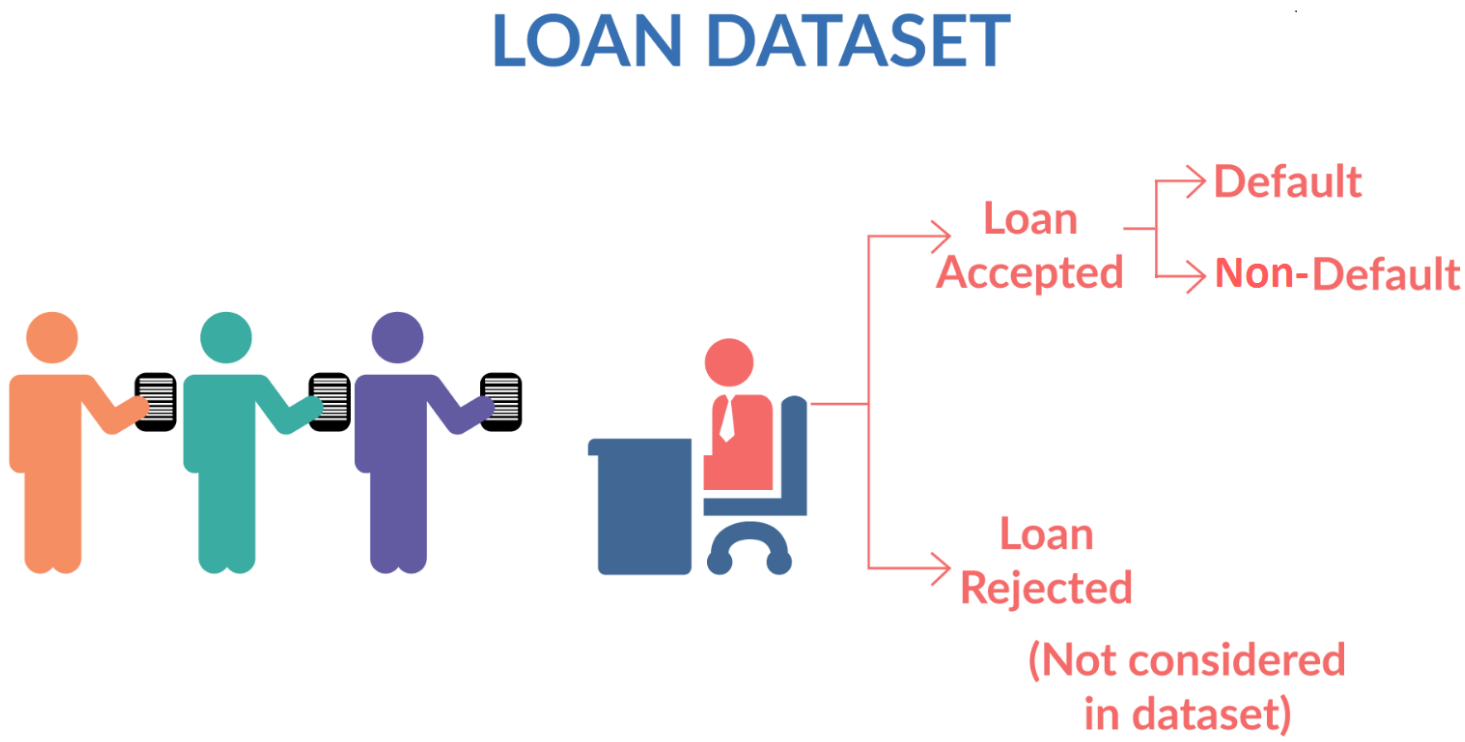
We work for a **consumer finance company** which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's 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

Data provided

Data given contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Data Set


LOAN DATASET
Loan Accepted
Default
Non-Default
Loan Rejected
(Not considered in dataset)

Decision making

When a person applies for a loan, there are **two types of decisions** that could be taken by the company:

- **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
  - **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
  - **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
  - **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan
- **Loan rejected**: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Objective

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.

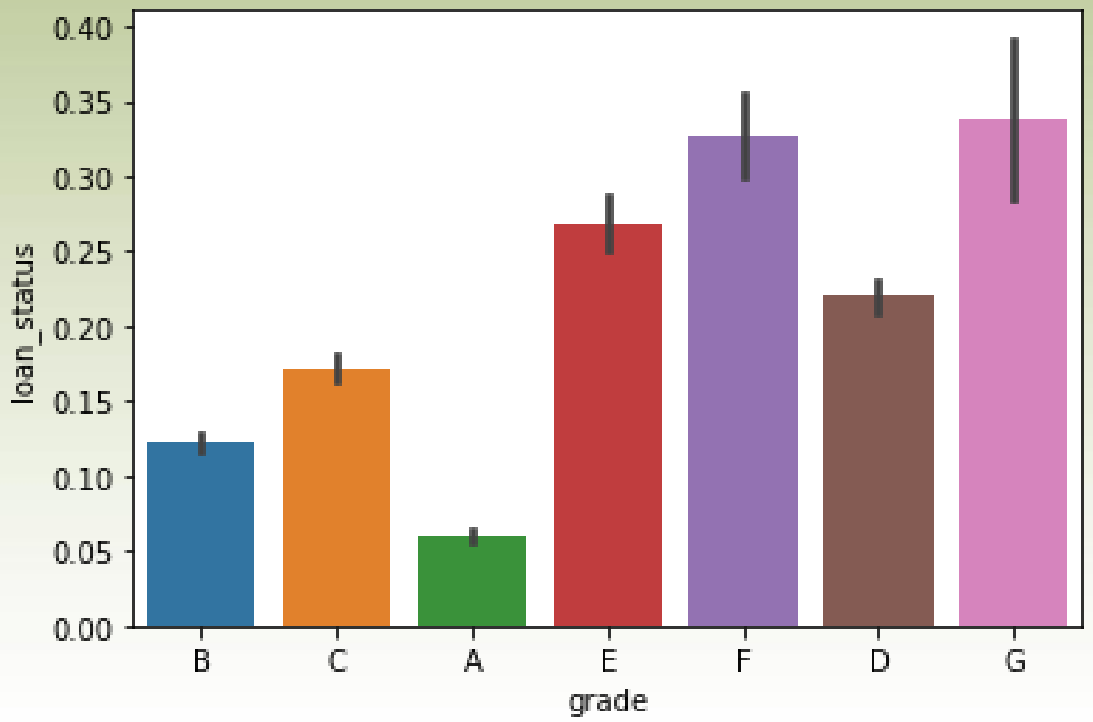
List some interesting facts about Black History Month

Data Cleaning

Input

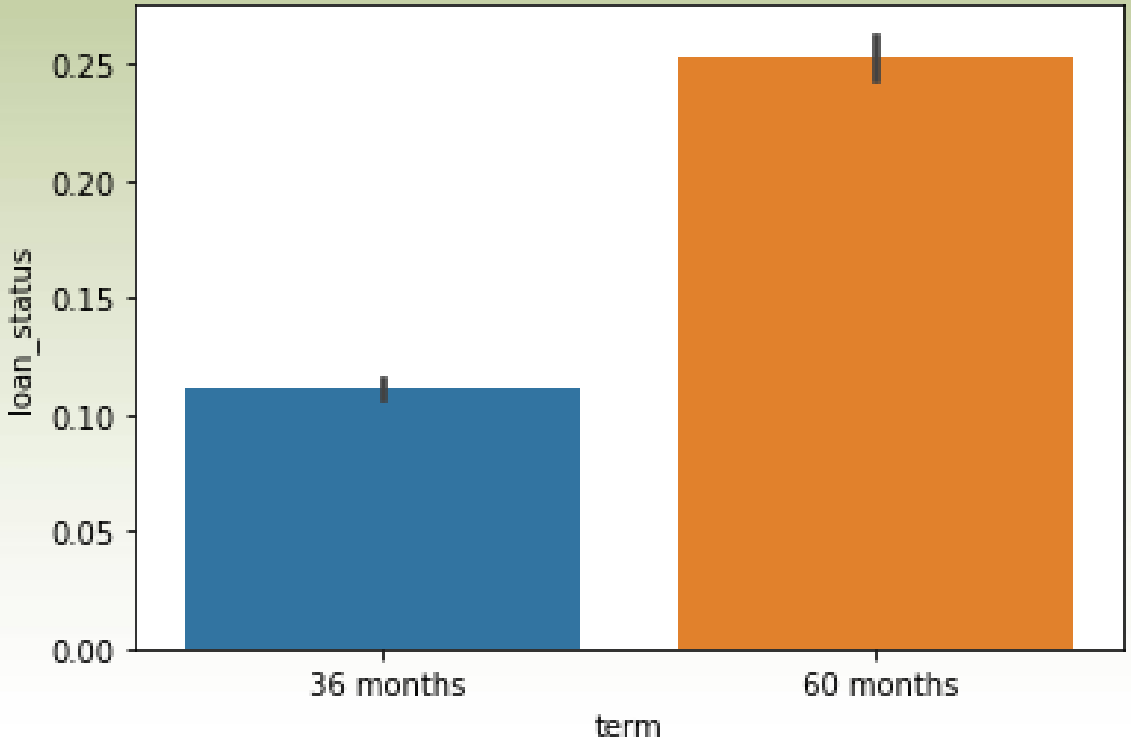
- Total provided inputs are in 39717 columns and 111 row .
- Import directories
- Check for missing values
- Data cleaning by removing missing values.
- After cleaning 39717 colums and 56 rows
- Highlight the values required for analysis

Analysis



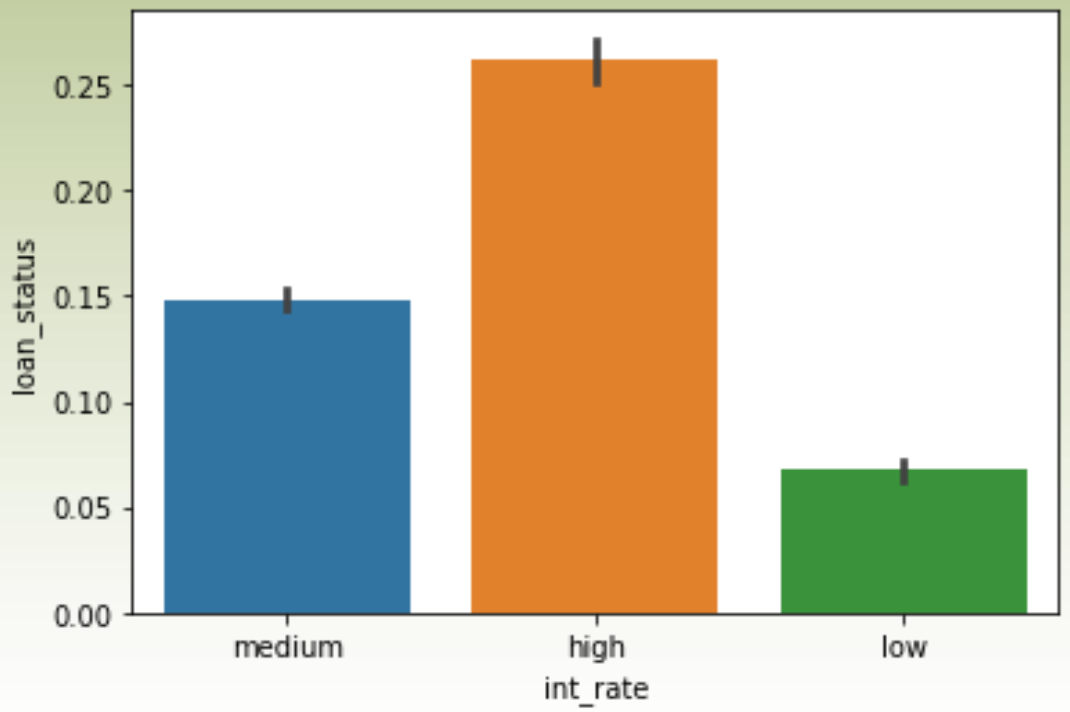
- Defaulter can be mapped with the grade of loan as A is having least defaulters and G is highest.

Analysis



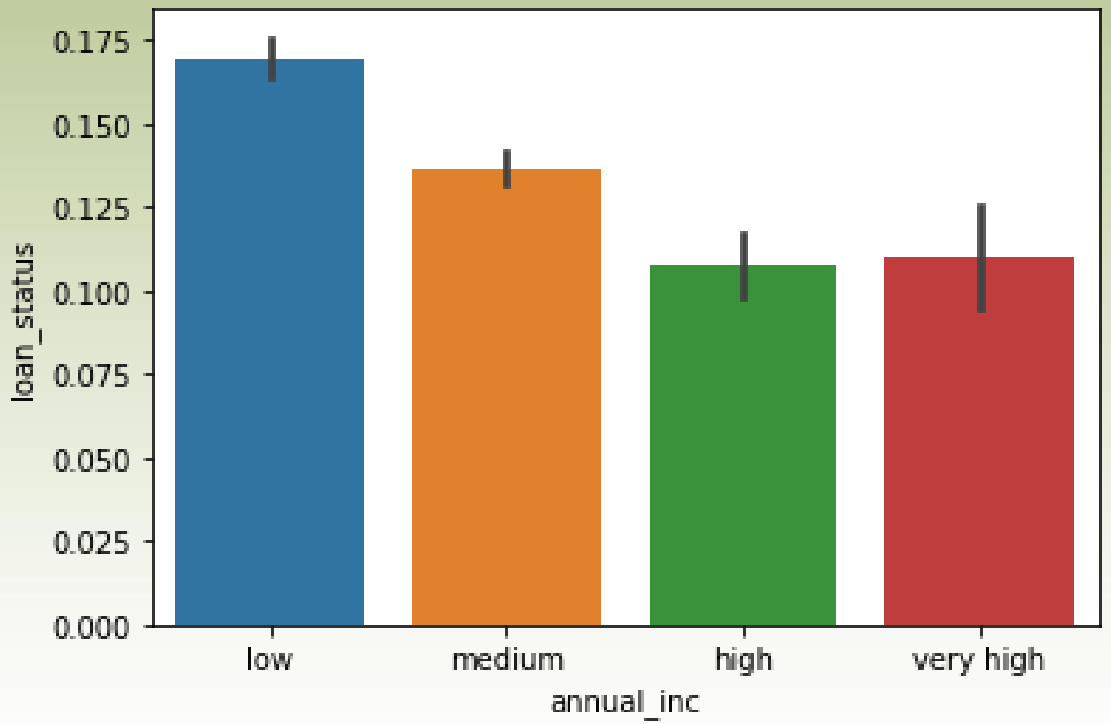
- Persons applying for more tenure are likely to default more .

Analysis



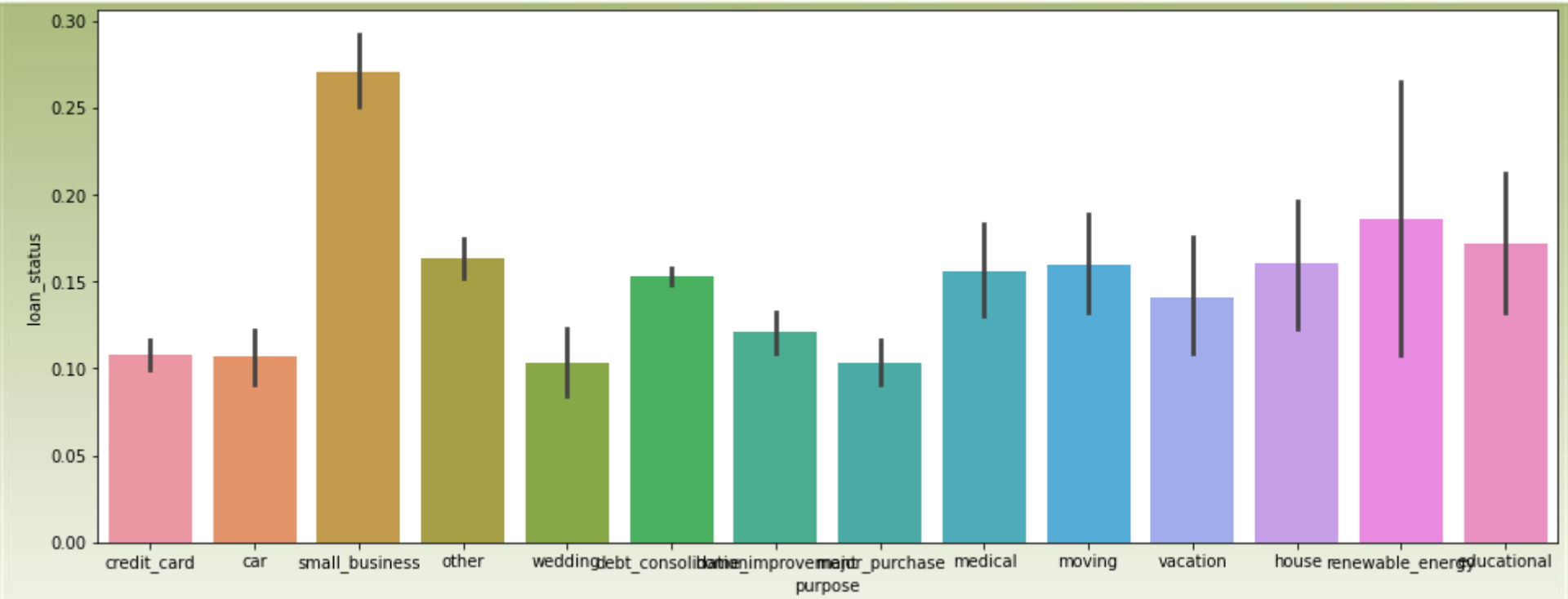
- Higher the intrest rate more risk of default

Analysis



- Less Annual income means more chances to default

Analysis



Above is analysis based on purpose of loan .Loan taken for small business is on high risk , however for wedding purpose on low risk

Thank You