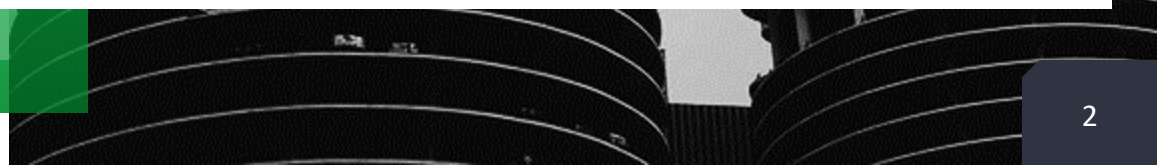
LENDING CLUB

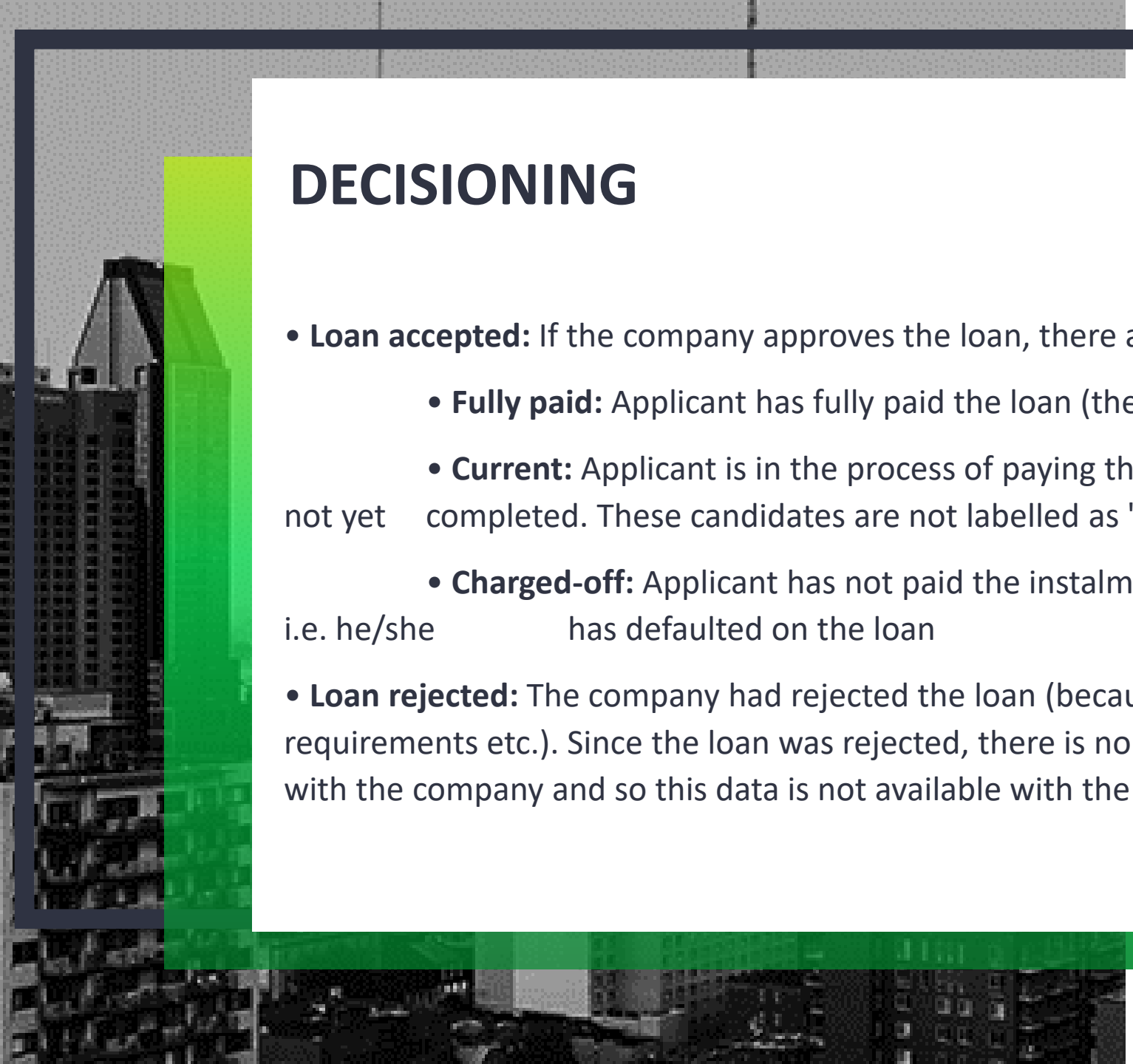
Risk Analysis

INTRODUCTION

Lending club is a **Consumer finance company** which specializes 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





DECISIONING

- **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)

BUSINESS OBJECTIVES

• Lending club lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).

• Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

• The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default using EDA is the aim of our case study .

TASKS PERFORMED

1. Data Understanding
2. 2. Data Cleaning and Manipulation
3. 3. Data Analysis
    - i. Univariate Analysis
    - ii. Standard Univariate Analysis
    - iii. Bivariate Analysis

DATA ANALYSIS AND CLEANING

1. The Entire data set contains n 39717 rows and 111 columns

2. The data is divided into three major categories :

- **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

3. The fields with 100% null values will not be useful.

4. Rows with Loan Status as ‘Current’ will not help to decide whether the customer is defaulter or not. So won’t be useful

DATA ANALYSIS AND CLEANING

5. Columns 'desc', 'mths_since_last_delinq‘ are not helping to decide defaulters ,so can be removed .

6. Columns 'title', 'url', 'zip_code', 'addr_state‘ are also not useful for checking the replyments , so can be removed .

7. Removed % from int_rate values and converted them into float values for analysis .

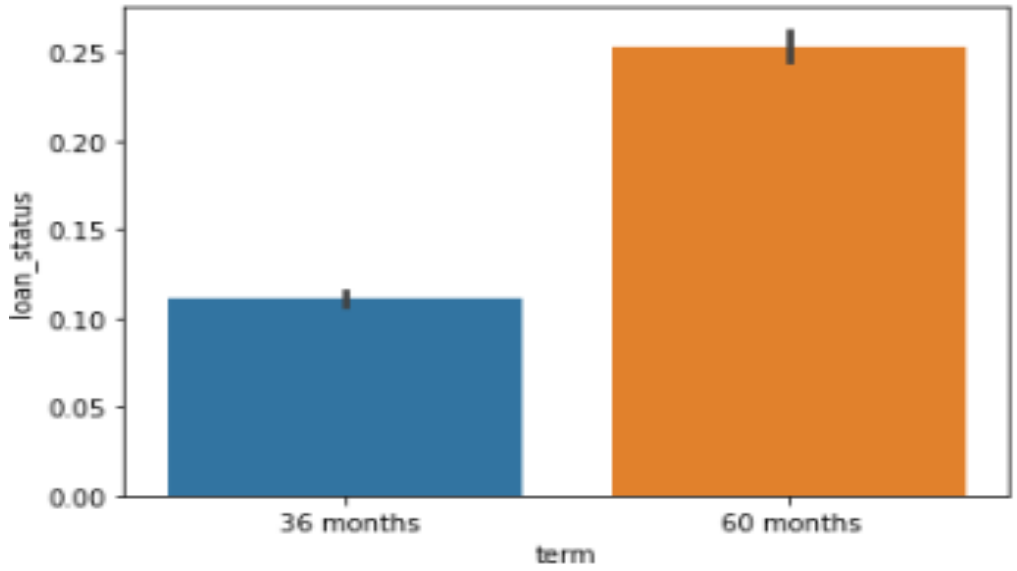
8. Removed xx from zip_code values and converted to integers

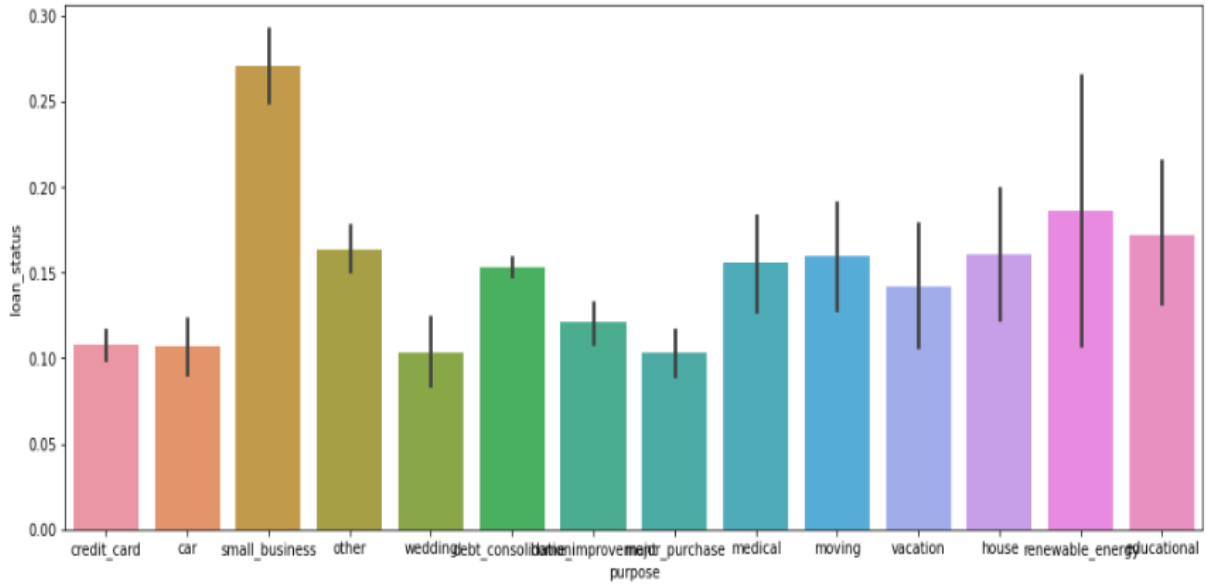
UNIVARIATE ANALYSIS:

1. Overall default rate is 15% .

2. With more term the Charge of rate increases .

3. Small Business loans defaults the most , then renewable energy and education





UNIVARIATE ANALYSIS:

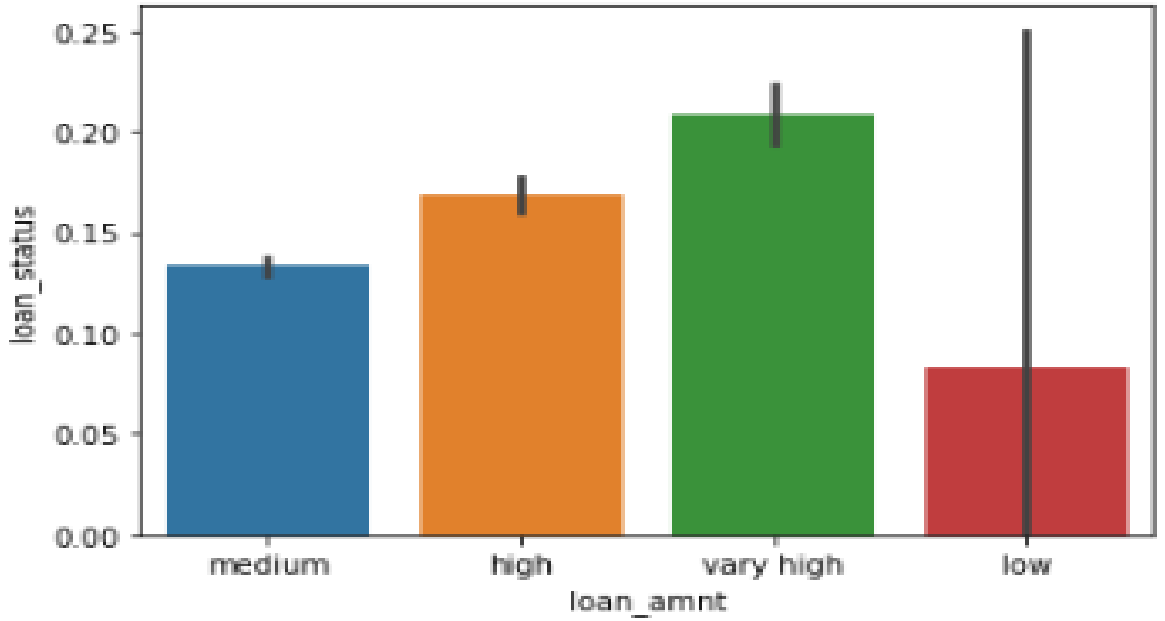
4. With more loan amount the charge off rate increases

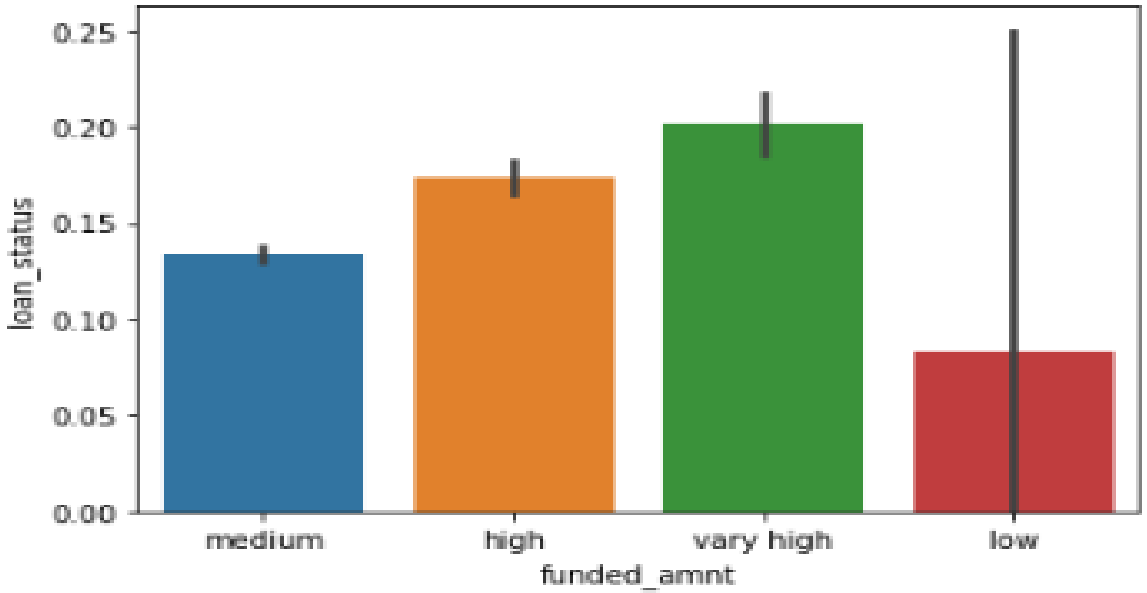
5. With more funded amount the charge off rate increases.

6. With more interest rate the charge off rate increases.

7. Higher interest rate higher rate of defaulters.

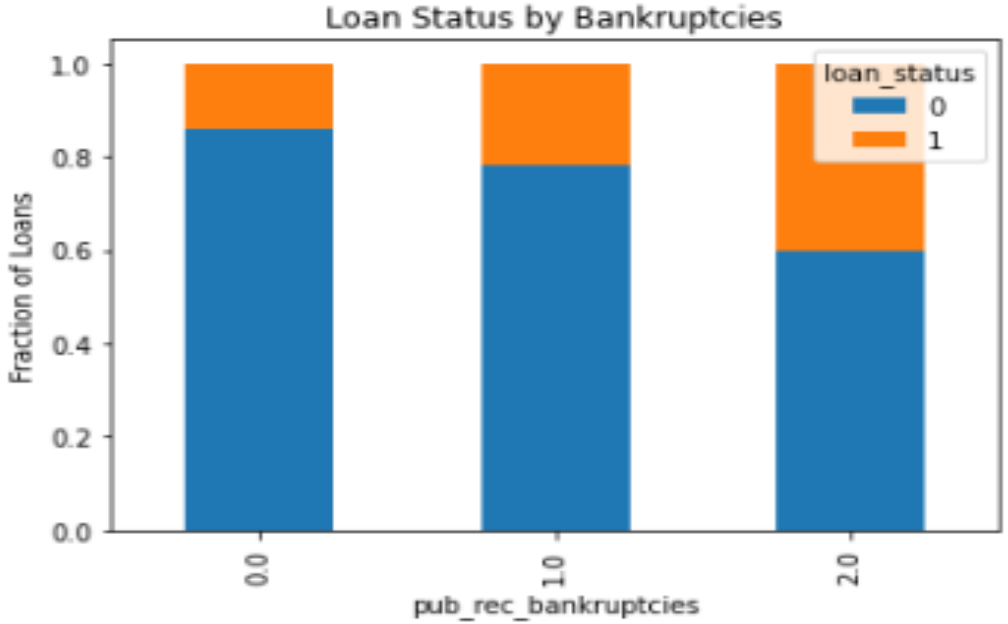
8. Lower the annual income higher the defaulter rate.

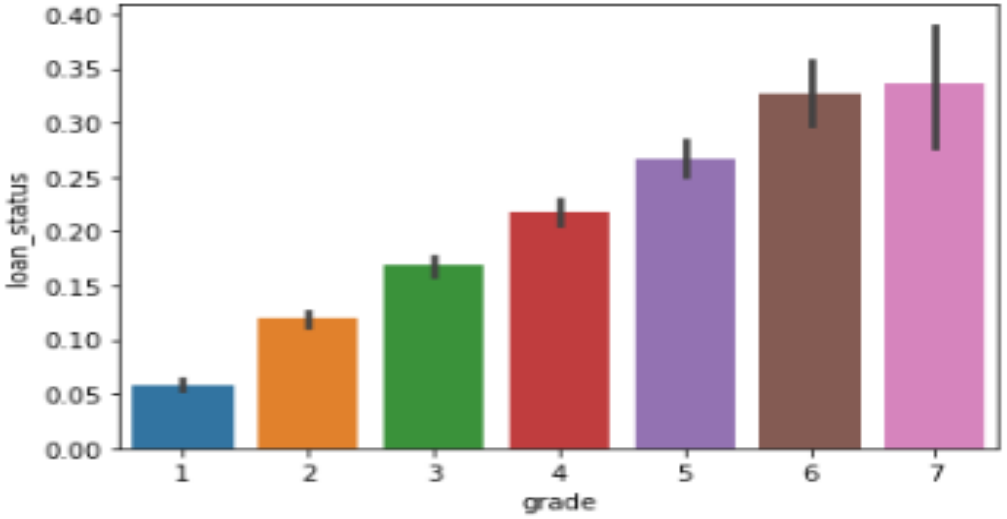




BIVARIATE ANALYSIS:

1. Recorded Bankruptcies increase in cases of Charged Off loan status
2. 'F' and 'G' grade customers are tends to be Defaulters
3. The Subgrades A1 to A5 have less chance to be defaulters





THANK YOU

Ankur Medhi