Lending Club Case Study

Prepared by: Sameer Raghavendrarao Jere & Prakash parthiban

Internal

Objective

- **The 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.**

Internal

Data Cleaning

- 1) member_id is unique and of no use for data analyis. So need to drop these.
- 2) loan_amnt & funded_amnt are same so one of them can be dropped. We will drop loan_amnt.
- 3) Subgrade contains the information about grade. Hence Subgrade can also be dropped as it is redundant.
- 4) We need to remove the NAN or missing values

Internal

Standardizing the data

- 1) The data contains special charcters like %, >.
- 2) we need to standradize this data by removing % and data binning
- 3) emp_length can be divided in different buckets:
  - a) < 1 year can be considered as 0 yrs experience
  - b) > 10years can be considered as 10yrs experience

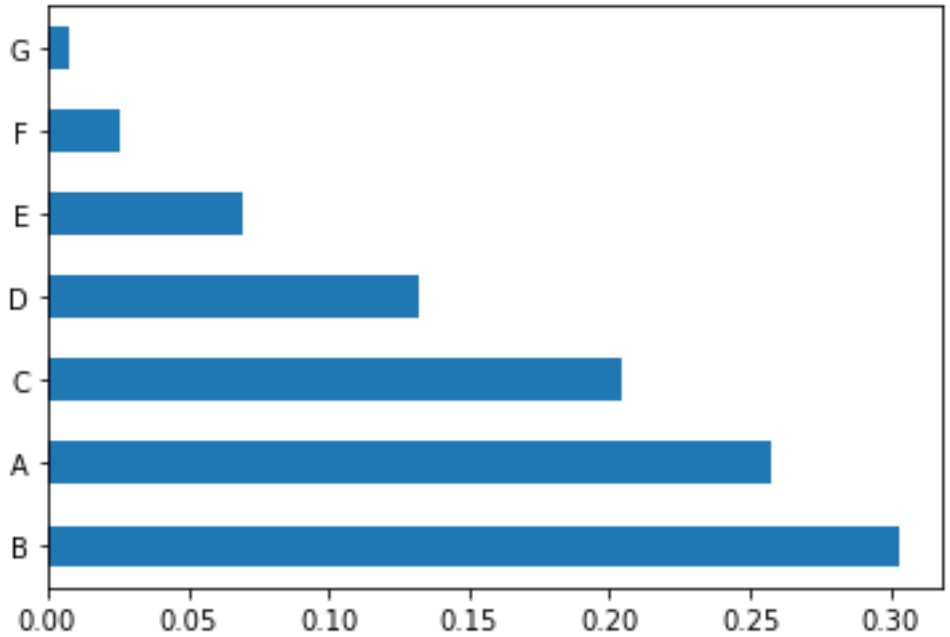
Internal

Handling Outliers

- 1) Univariate outliers - Univariate outliers are the once whose data points are far beyond the range of expected values with respect to one variable.
- 2) Multivariant outliers - Identifing the outliers which are outside the expected range when you plot the data along with some other variable.

Internal

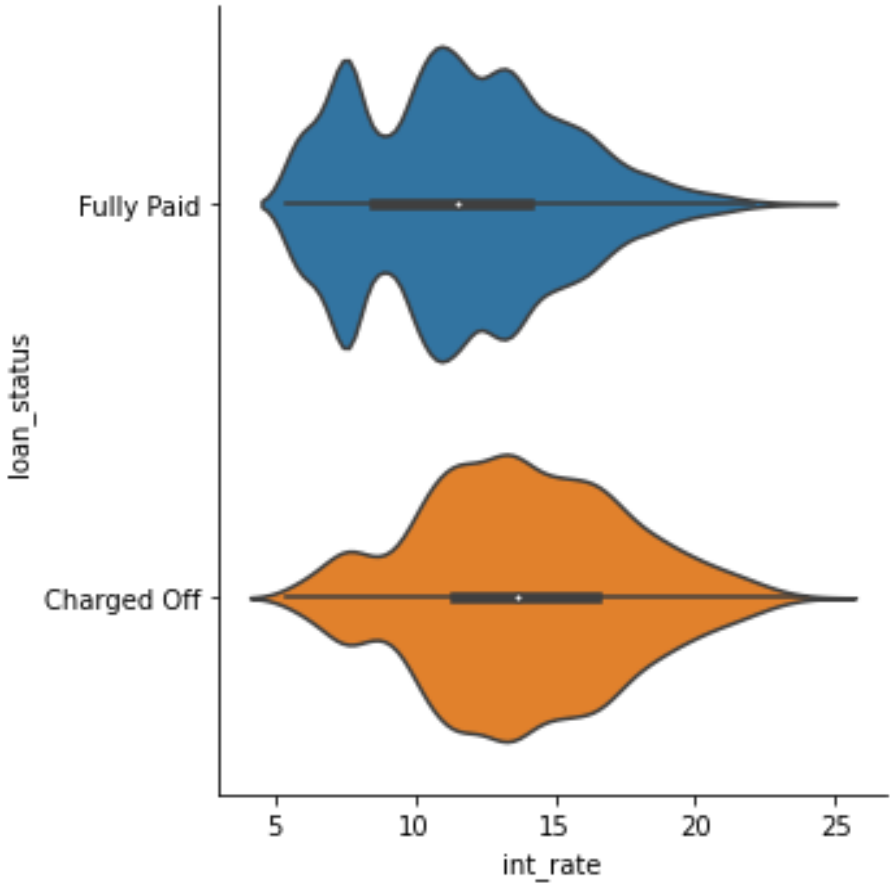
Univariant Analysis



The above bar plot shows more than 75% of the customers have credit score A, B & C. This is a good indicator as majority customers will be willing to pay the loan back inorder to maintain the good credit score or get upgraded to a better credit score by paying the loan on time.¶

Internal

Bivariate analysis

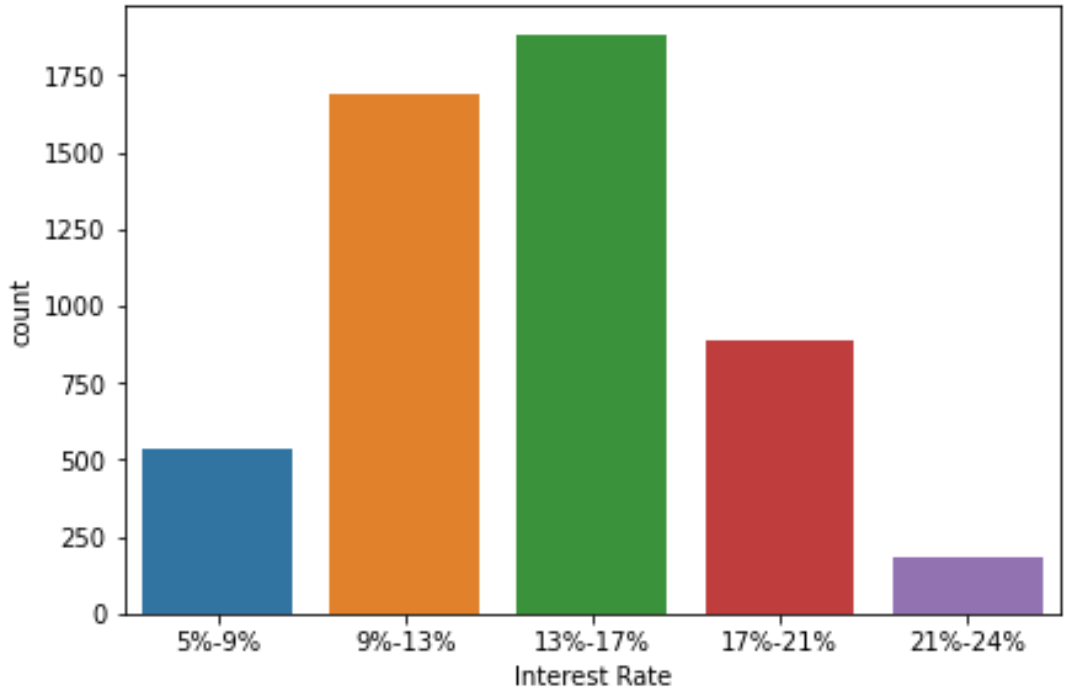


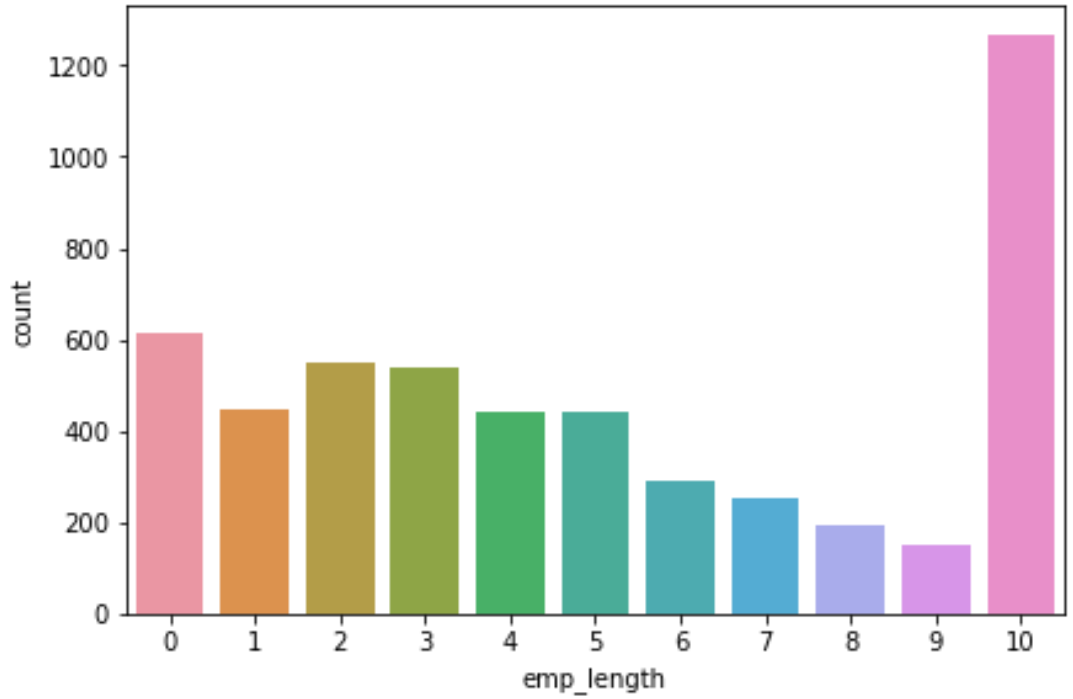
The above category plot shows we are checking the relation between loan_status and interest rate

- the customers who are offered lower interest rates tend to pay the loan on time
- also the customers having interest rate between 12% to 14% are likely to default

Internal

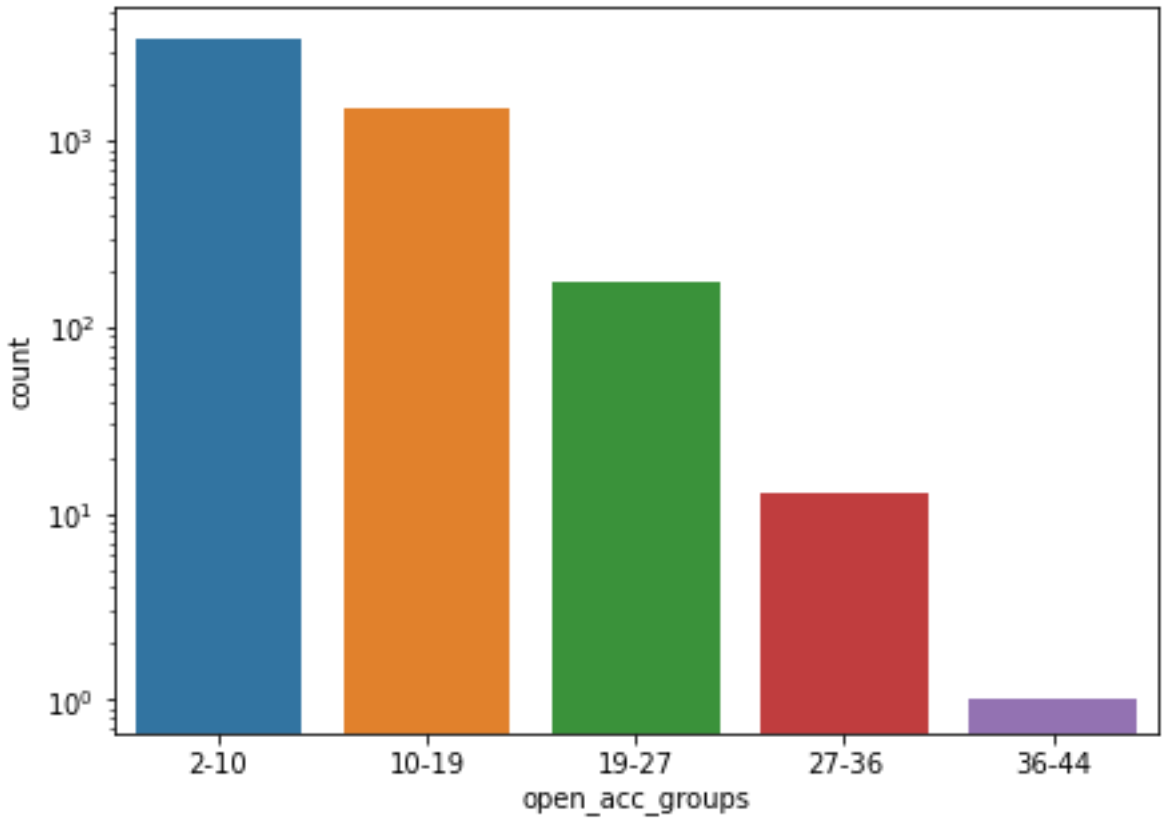
Create bins to convert numerical variables to categorical variables.

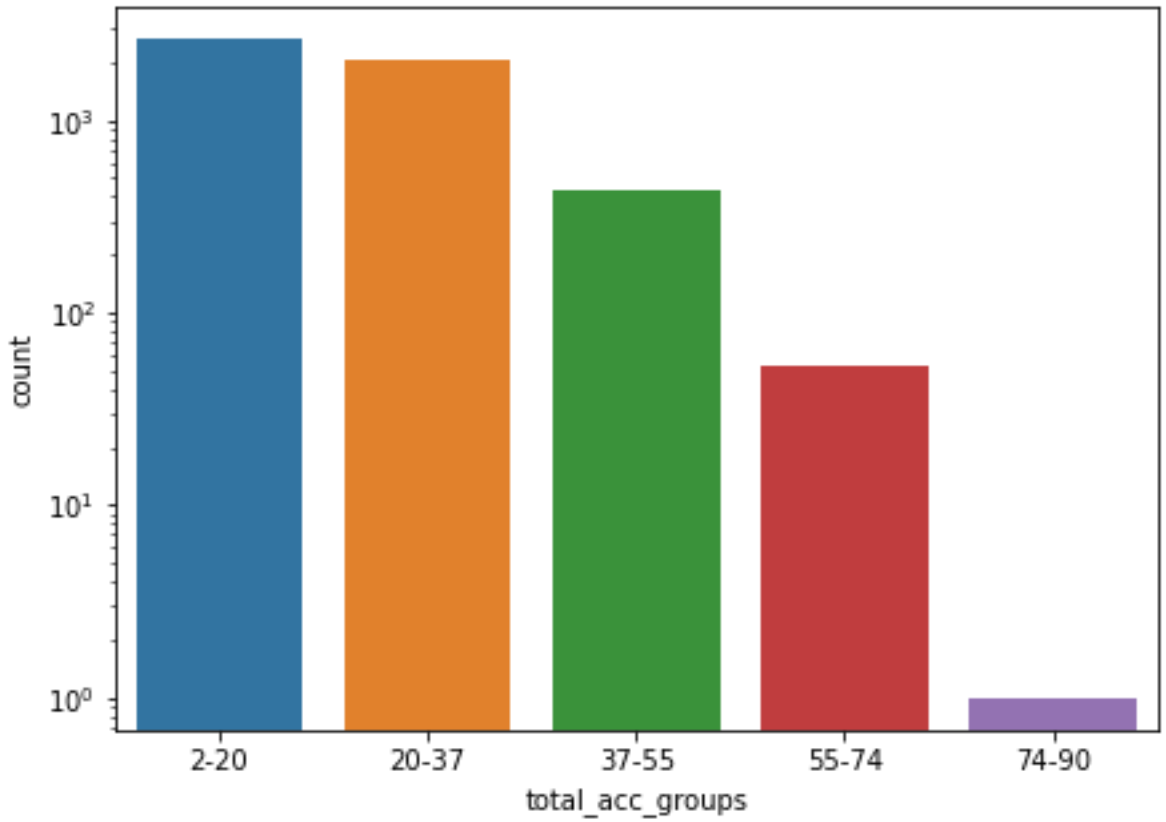




Internal

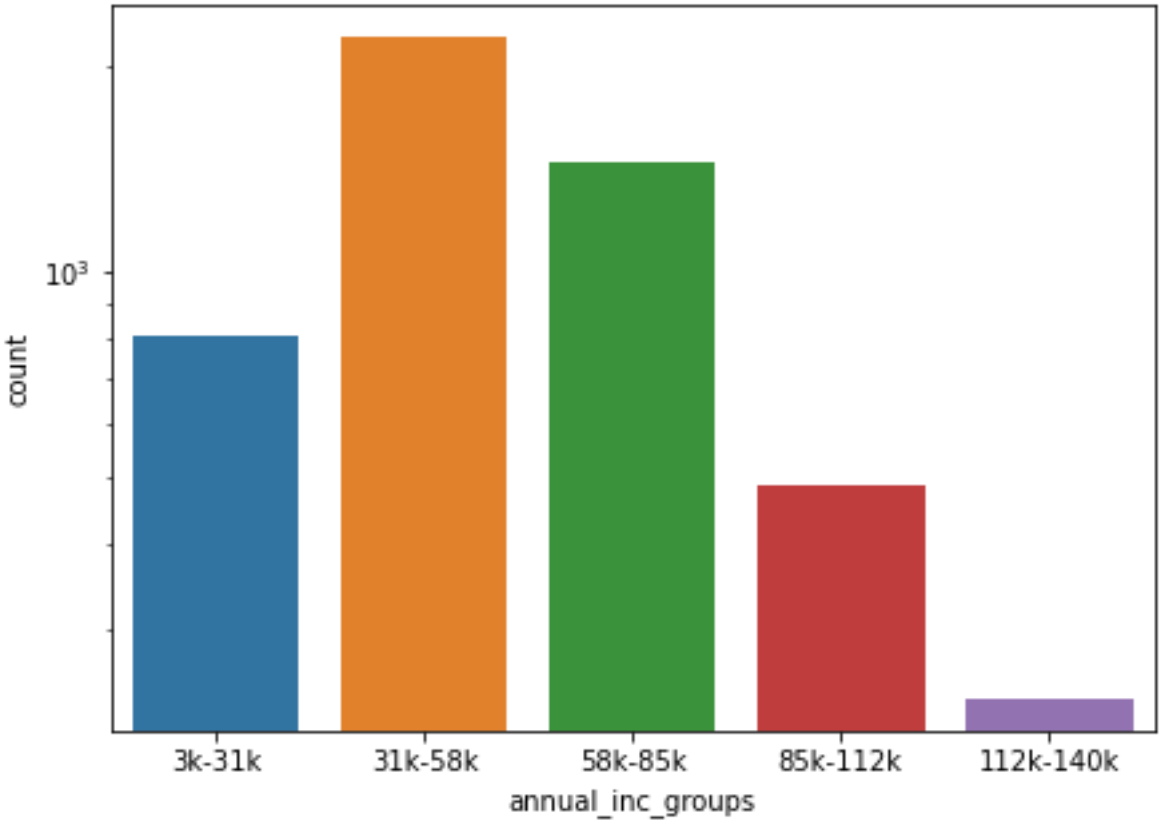
Analysis with bins created.

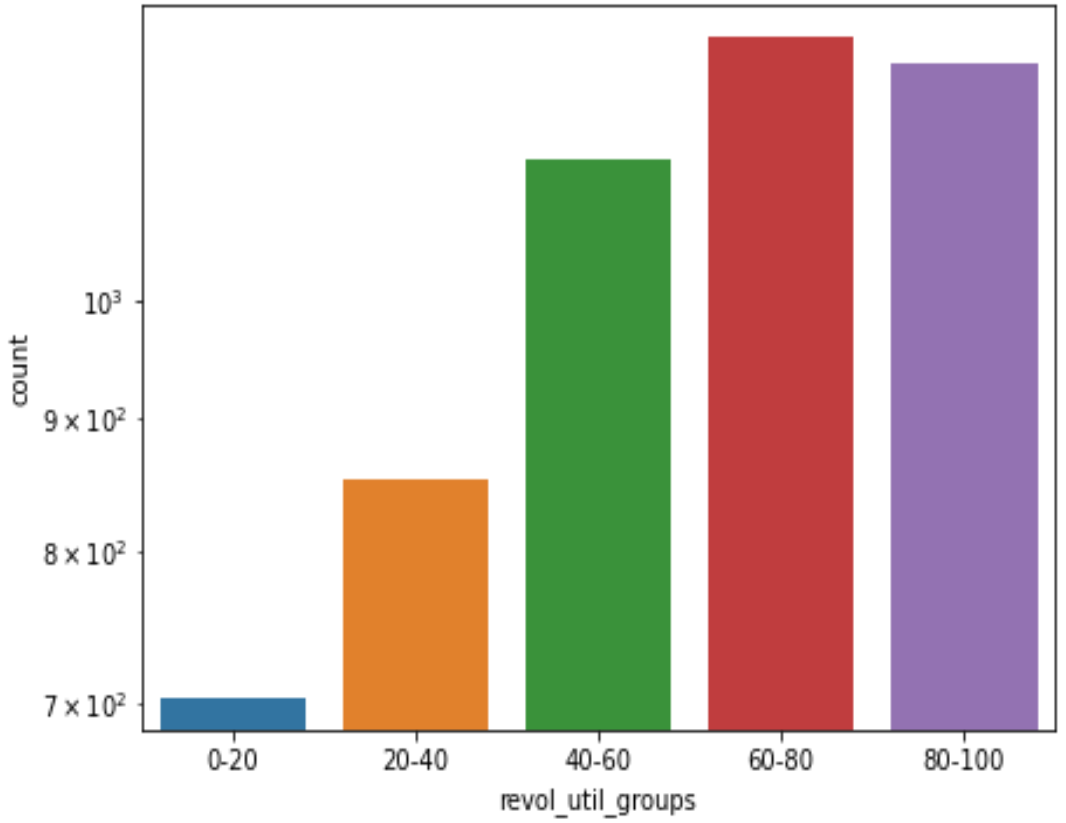




Internal

Analysis with bins created.





Internal

Final Observations

1) Customers with interest rates between 13% to 17% are likely to default

2) Customers with lowest interest rates are the once who likely to payback the loan amount on time.

3) Customers who are offered the highest interest rates (above 20%) are likely to payback the loan amount on time.

4) Customers with more credit lines are paying the loan on time.

a) This cannot be generalized because there might be a possiblity they take more loans to pay the existing loans.

b) So the bank should avoid giving credits to the customers who already have lot of credits as they might be scammers.

c) we have examples of Nirav Modi and Vijay Mallya who used such practices of taking loan from one bank to pay other loan and do money laundering.

5) The total credit lines also drives the same observation which we made in point no 4) above.

6) The customers with annual income of 31k to 58k are likely to become default customers.

7) The final graph on revolving credit lines also proves that higher the revolving credits higher the chances of the customer defaulting the loans.

Internal

Thank you!

Internal