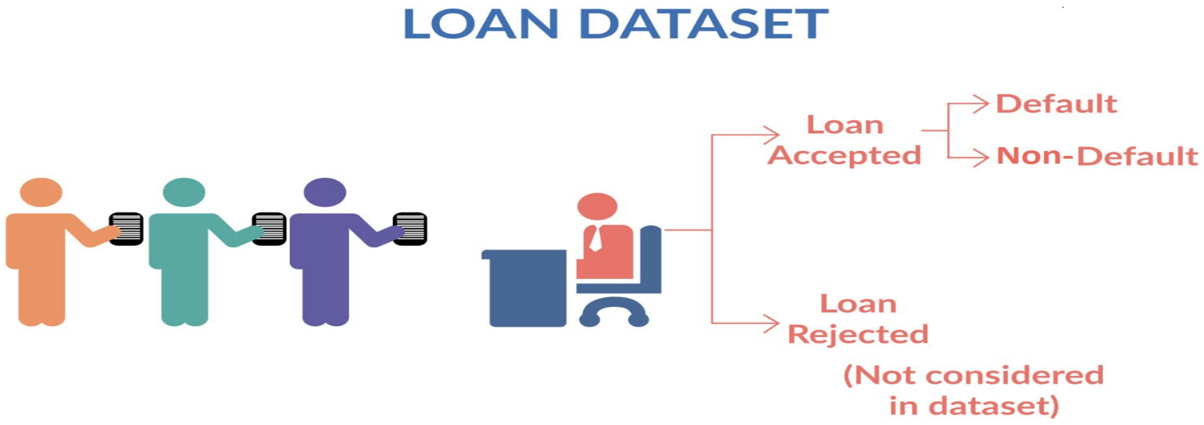
Lending Club Analysis

By Aahuya Rakshakahari & Pragyanshree

Problem Statement

- Analyze the data of a consumer finance company which specialises in lending various types of loans to urban customers.
- Default in a loan results in the loss to the bank, in this study, we analyze some of the factors involved in defaults.

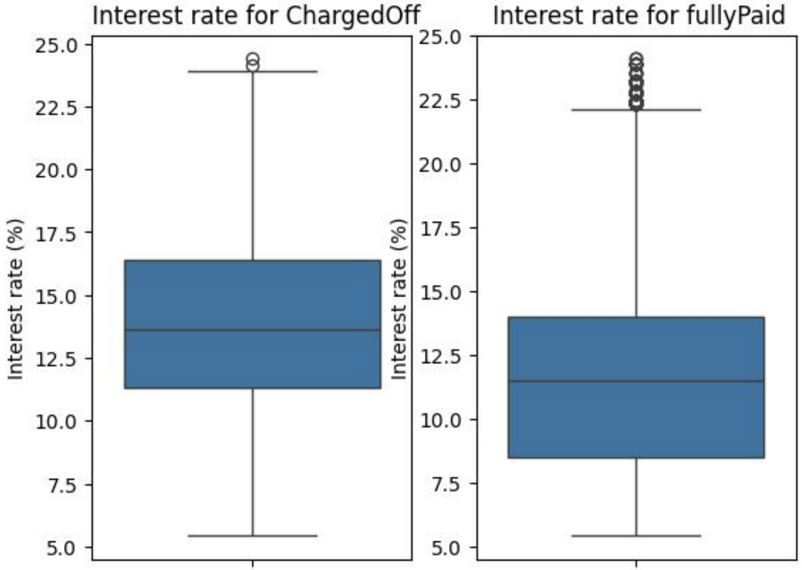


Data Sourcing / Cleansing

- The data is readily provided with the assignment, hence no effort done to source the data.
- Cleansing :
    - Analysed and removed the columns that had all null values.
    - Analyzed the variance in columns and found some columns had the same value for all the rows, which are not useful for our study hence removed those as well.
    - Some columns were not in the right format, for example interest rates had % in it and we converted by trimming and removing %. Duration had text but since the duration had only two values, we didn’t convert it into a numeric field.

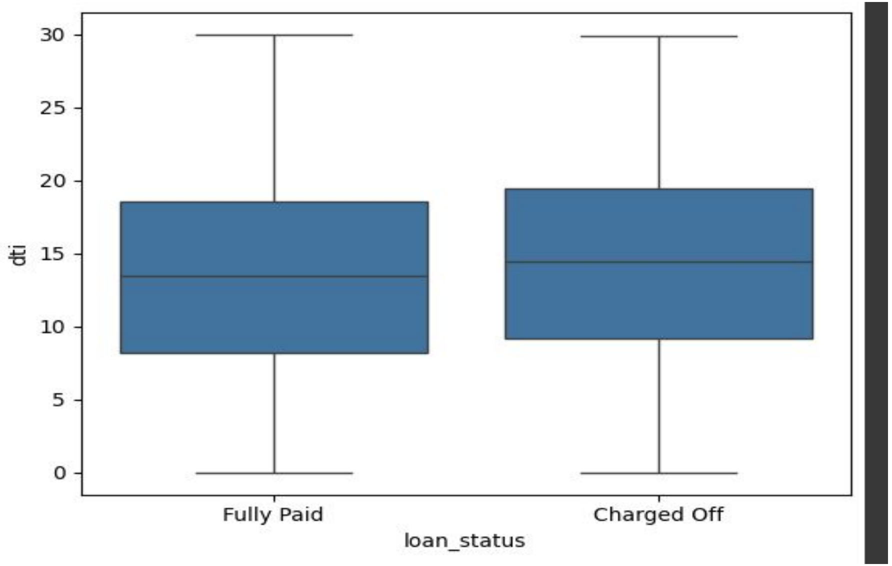
Default - Interest rate

We observe in general, the median interest rate for charged off loan was higher than the fully paid loans.



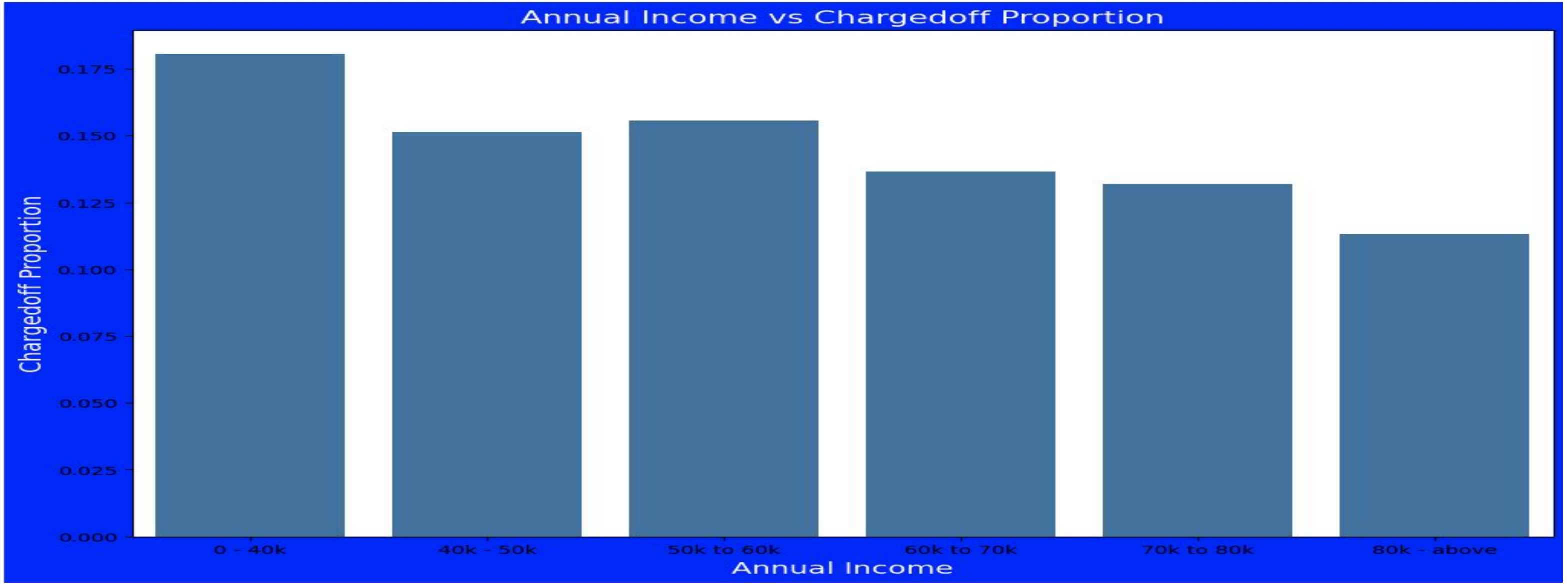
Default - DTI

Dti was also higher for the defaulting customers.



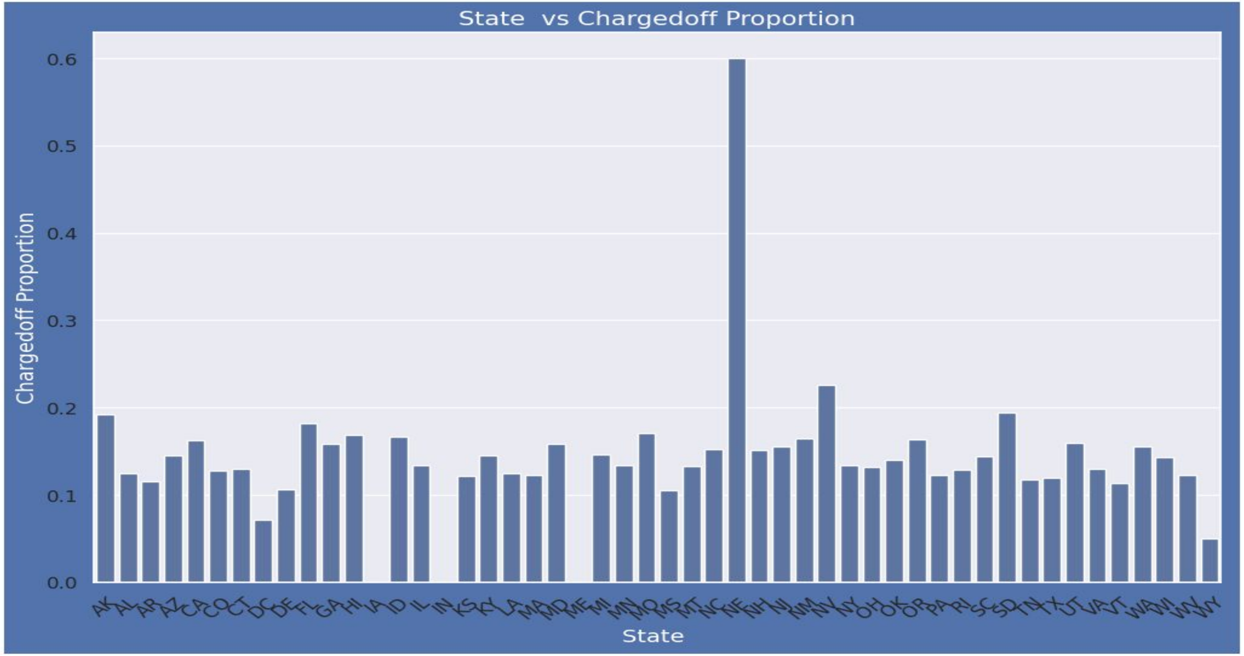
Default vs Annual Income

- Proportionally more defaults are in 0-40k annual income range.



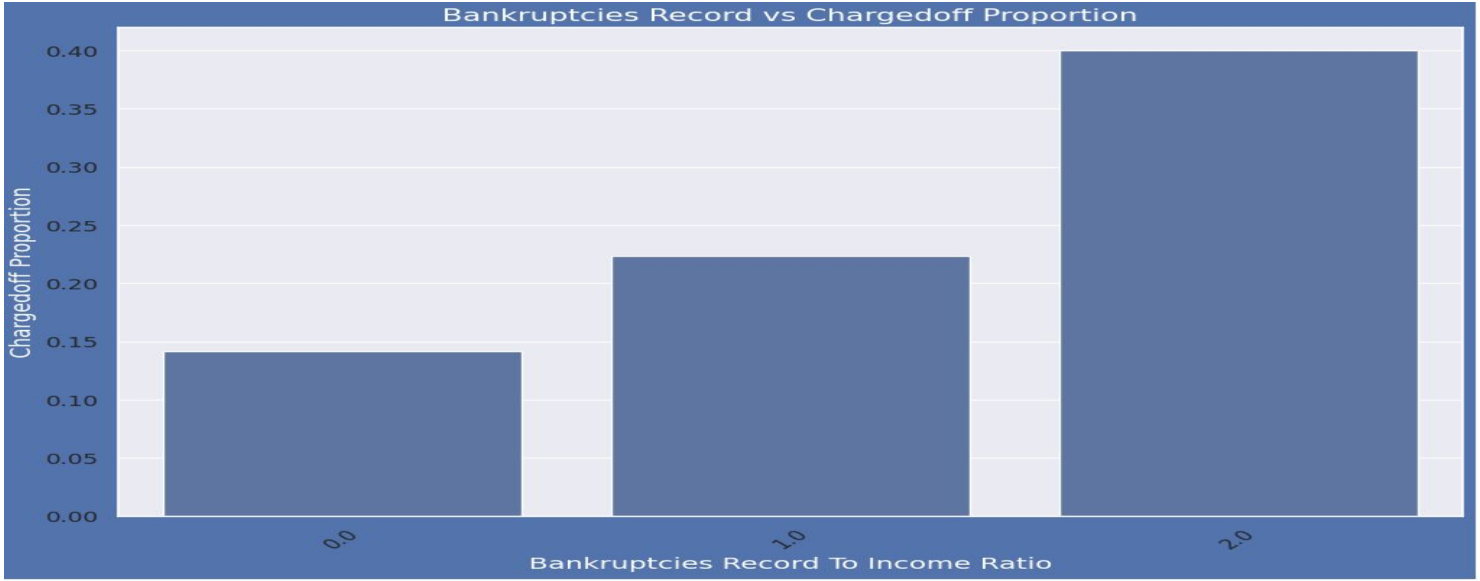
Default based on states

- Nebraska (NE) is a having higher number of defaults compared to other states.



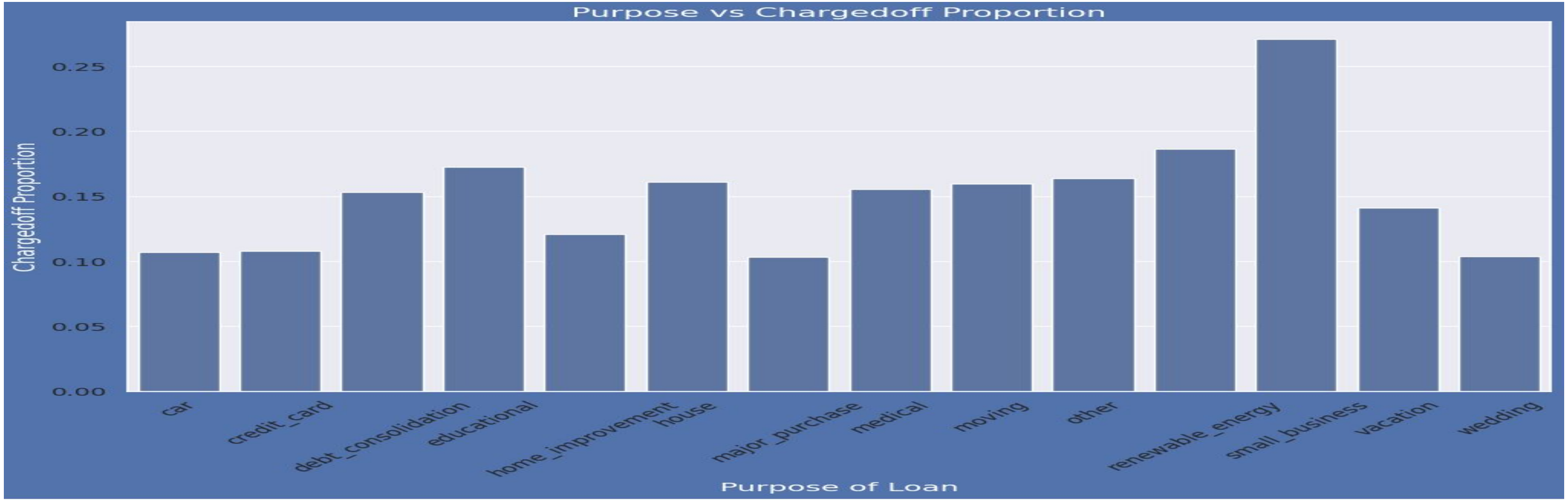
Bankruptcies

- Higher the Bankruptcies to income ratio, higher is the chance of default



Purpose

- The loans taken by small_business is more prone for default.



Summary

- To summarize, the defaults are correlated to following variables
  - Interest rate
  - DTI
  - Annual Income
  - Some states have higher default rate
  - Previous bankruptcies
  - Purpose of the loan
- The bank can diversify by reducing the loan granted by considering above parameters to reduce defaults.