Loan Sanction EDA

Dhrubajyoti Debnath

Problem Statement

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

1.Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

1. **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
2. **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'.
3. **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

2.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)

The data given for the case study 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.

Solution Approach

Out of all the attributes, we have identified some probable attributes which might have some effect on the loan status.

- Grade
- Employment Length
- Annual Income
- Home ownership
- Purpose of taking loan
- Dti
- inq_last_6mths
- open_acc
- revol_bal

- We have plotted the correlation of these individual attributes and loan status in the python notebook

EDA Observation

Below are the observations as per the plots in **Group_Facilitator_Name.ipynb** notebook

- **Grade** – In the plot, we can see that grade has not much effect on the Charged off status, however it can be seen that charged off is comparatively higher for grade B,C and D compared to E, F and G grade
- **Employment Length** - it seems people with employment length of about 5 is more susceptible to be charged off
- **Annual Income** - it seems customers with about 50000 annual income are more susceptible for charged off
- **Home ownership** - it seems there is not much difference between people with rented and owned home
- **Purpose of taking loan** - it is quite clear from the plot that people with a purpose of setting up small business are way more susceptible to be charged off
- **Dti** - it is seen that there is not much difference between median value of dti for various loan status
- **inq_last_6mths** - from the plot we can see median value of inq_last_6mths are same for charged off and fully paid, so no clear conclusion can be drawn
- **open_acc** - Not much difference in the number of open account among loan status
- **revol_bal** - This plot is also not giving any clear direction on loan status