Loan Prediction

Steps:-

A. Data Cleaning: -
   a. Remove all Nan column and set a threshold of Nan value to 30% and then remaining rows is remove from the dataset.
   b. Drop duplicates values

B. Data Filtering:-
   a. Base on the columns details given in data dictionary data filter the most require columns only.
   b. Then analyzing the remaining column values from the dataset filter more columns from it.

C. Analyze the dataset:
   a. Univariate analysis:-
      i. Perform univariate analysis on different columns to understand their distribution, frequency and dependency on data set.
      ii. I use distplot, histplot and box plot to analyze the single variable.
      iii. Filter dataset in two categories
         A. person who fully paid the loan
         B. person who charged off.

         Then compare their respective graphs.

      iv. After Univariate, analysis there is conclusion that this data set is highly imbalance towards the fully paid category.
   b. Bivariate Analysis:-
      i. Perform bivariate analysis on the multiple values of data-frame to check the working of the variables.
      ii. For that, I replace the categorical values to numeric values and then understand the behavior of each values on different parameter.
      iii. After the analysis, I found that fully paid and charged off slightly same features than the current category.