Lending Club Case Study

Problem Statement and Objective

- This project is for a large financial company that provides various types of loans to applicants.
- The purpose of this project is to minimize credit loss to the company due to loan default by borrowers.
- A data set is available that contains historic data with details (both loan related and borrower related data) for each loan provided by the company.
- The objective is to use data analytics to analyze this loan data set and determine the factors(variables) that influence loan default.
- These factors will help the company when issuing new loans to predict the risk and minimize the risk by deciding whether to accept or reject the new loan application.

Analysis Approach

- Data Exploration
- Data Cleaning
  - Handling Nulls
  - Removing unwanted columns
  - Converting to right datatypes
- Data Enriching
  - Data Imputation
  - Derived Metrics
- Data Analysis
  - Univariate
  - Bivariate Analysis
- Observations and Recommendations

Data Cleaning

- Dropped columns that contain all missing(null/NA) values.
  - Eg: annual_inc_joint, dti_joint
- Dropped columns with large percentage of missing values.
  - Eg: next_pymnt_d(97%), mths_since_last_record(92%)
- Verified rows with all missing values – found none.
- Dropped columns that seem not relevant for any analysis.
  - Eg: id, url, title, out_prncp_inv
- Dropped columns with single value only.
  - Eg: policy_code, application_type, pymnt_plan

Data Enriching

- Converting columns to right data types
  - Eg: int_rate(from object to float), revol_util
- Data Imputation for certain columns with very less % of missing values
  - Eg: emp_length
- Creating derived metrics
  - Eg: issue_year, issue_month from issue_d

Observations and Recommendations