Lending club case study

Initial table data:

- 111 columns
- 39717 Rows of data

- Columns post-filtration of data by removing the columns which has values as NA:

user_attributes = ["id", "member_id", "desc", "zip_code", "addr_state", "annual_inc", "dti", "inq_last_6mths", "open_acc", "open_acc_6m", "pub_rec", "recoveries", "collection_recovery_fee", "mths_since_last_major_derog", "emp_title", "emp_length", "home_ownership", "verification_status", "last_credit_pull_d"]

loan_attrinutes = ["loan_amnt", "funded_amnt", "funded_amnt_inv", "term", "int_rate", "installment", "last_pymnt_amnt", "grade", "sub_grade", "issue_d", "loan_status", "pymnt_plan", "purpose", "title", "delinq_2yrs", "mths_since_last_record", "application_type", "policy_code", "last_pymnt_d", "next_pymnt_d"]

Univariate analysis:

➢ Removal of columns "pymnt_plan", "policy_code", "application_type", "open_acc_6m".
  These columns have single values and their existing will not change the analysis
   to reduce the size of the data frame memory:

➢ Analyzed various attributes which might be contributing to the loans getting defaulted.

➢ Observed the different values of the attributes.

Bivariate Analysis:

- Analyzed the relationship between the different entities. Such as :
  - term_months and loan_status
  - emp_length and loan_status
  - addr_state and loan_status
  - sub_grade and loan_status
  - verification_status and loan_status