**Project: Bank loan of customers**

Cleaning steps followed –

* Converted finance\_1 csv file into .xlsx form.
* From Finance\_1 - I removed unnecessary columns - payment plan (just had one value – n).
* Change the format of column funded\_amnt\_inv, installment, dti, total\_pymnt to numerical (decimal up to 2 places) from general.
* Trimmed the term column(space present before).
* Checked for duplicate records, no duplicate record found.
* Checked for errors in entire table, no error found
* changed format of columns - issue\_d to date(yyyy-mm-dd) from custom(ddd-yy-mm).
* Replaced empty values with NULL.
* From Finance\_2 removed unnecessary columns –
* next\_pymnt\_d – 38577(out of 39717 records) had value – 0
* out\_prncp & out\_prncp\_inv – 39717 records had value – 0
* initial\_list\_status - just had one value – f
* delinq\_2yrs – 35405(out of 39717 records) had value – 0
* changed format of columns - earliest\_cr\_line, last\_pymnt\_d, last\_credit\_pull\_d to date(yyyy-mm-dd) from custom(ddd-yy-mm).
* Change the format of all columns with decimal values to numerical (decimal up to 2 places) from general.
* Checked for duplicate records, no duplicate record found.
* Checked for errors in entire table, no error found.
* Replaced empty values with NULL.