

# Pandas Assignment - Loan Data Cleaning & Insight

## Dataset Link

1. Find out the dataset dimensions
2. check the data type of each column
3. Check for the null values in the Dataset
4. fill null value of emp\_title column as "None" and count the value of each emp\_title
5. Fill the null value of emp\_length column as "Unknown" and count the value of each emp\_length.
6. Fill the null value of the title column as "Unknown" and print the number of unique values.
7. Clean the revol\_util column and fill in appropriate null values.
8. Drop null value rows of last\_pymnt\_d, issue\_d and last\_credit\_pull\_d then change the data type using to\_datetime function (NOTE : give a try )
9. Handle the null values & Noise in the Dataset for each column. Drop all those columns whose nulls are equal to or greater than 70%
10. Find out that columns which have constant values and drop them after identification
11. Check and delete the duplicate rows in the data set
12. Take appropriate action for the following columns - desc, pub\_rec\_bankruptcies .
13. Find out the statistical summary of the Dataset
14. Calculate annual\_inc column's following aggregation values - mean, std, min, 50%
15. Create a new column income\_range and divide the annual\_inc into three groups - if annual\_inc is less than 40000 then assign 'Low' and if annual\_inc is greater than 40000 and less than 80400 then assign medium and if above 80400 then 'High'
16. Count how many persons are in each income\_range.

17. For each loan\_status find the number of installment for different home\_ownership