**Mini Project – Python**

**Topic - Loan Dataset**

**Introduction -** Loans dataset aims to identify variables which indicate if a

person is likely to default, which can be used for identifying the risky loan

applicants to avoid any financial loss to the company.

**Dataset Description -** It contains the complete loan data for all loans

issued through the time period 2007 to 2011.

**Data Dictionary -**

1.annual\_inc - The self-reported annual income provided by the borrower

during registration.

2.dti - A ratio calculated using the borrower’s total monthly debt payments on

the total debt obligations, excluding mortgage and the requested LC loan,

divided by the borrower’s self-reported monthly income.

3.emp\_length -Employment length in years. Possible values are between 0 and

10 where 0 means less than one year and 10 means ten or more years.

4.funded\_amnt - The total amount committed to that loan at that point in

time.

5.funded\_amnt\_inv -The total amount committed by investors for that loan at

that point in time.

6.grade - LC assigned loan grade

7.id - A unique LC assigned ID for the loan listing.

8.installment - The monthly payment owed by the borrower if the loan

originates.

9.int\_rate - Interest Rate on the loan

10.last\_pymnt\_amnt-Last total payment amount received

11.last\_pymnt\_d -Last month payment was received

12.loan\_amnt -The listed amount of the loan applied for by the borrower. If at

some point in time, the credit department reduces the loan amount, then it

will be reflected in this value.

13.loan\_status - Current status of the loan

14.member\_id -A unique LC assigned Id for the borrower member.

15.purpose - A category provided by the borrower for the loan request.

16.term -The number of payments on the loan. Values are in months and can

be either 36 or 60.

17.total\_acc -The total number of credit lines currently in the borrower's credit

file

18.total\_pymnt -Payments received to date for total amount funded

19.total\_pymnt\_inv -Payments received to date for portion of total amount

funded by investors

20.total\_rec\_int -Interest received to date

**Questions -**

1. Import the dataset and understand it.

2. List down the number of rows and columns.

3. ‘Int\_rate’ column is character type. With the help of lambda function

convert into float type.

4. Check the datatype of each column.

5. Cleaning the dataset- Remove the columns having complete NaN value in

the entire dataset.

6. Write the code to find the value counts of the ‘loan\_status’ category

column and filter only the ‘fully paid’ and ‘charged off’ categories.

7. Filter the ‘Emp\_Len’ column to extract the numerical value from the

string.

Hint - Emp\_len : < 1year, 2 years , 3 years as 1 , 2, 3 so on.

8. Using the Lambda function, remove the month from the ‘term’ column

such that ‘36 months’, ‘60 months’ appear as 36 and 60 respectively.

9. Create a new column as risky\_loan\_applicant by comparing loan\_amnt

and funded\_amnt with the following criteria -

If loan\_amnt is less than equals to funded\_amnt set it as ‘0’ else set it as

‘1’.

10. Using the bar plot visualize the loan\_status column against categorical

column grade, term, verification\_status . Write the observation from

each graph.

11.Using a user defined function convert the ‘emp\_len’ column into

categorical column as follows -

If emp\_len is less than equals to 1 then recode as ‘fresher’.

If emp\_len is greater than 1 and less than 3 then recode as ‘junior’.

If emp\_len is greater than 3 and less than 7 then recode as ‘senior’

If emp\_len is greater than 7 then recode as ‘expert’.

12.Find the sum of ‘loan\_amnt’ for each grade and display the distribution

of ‘loan\_amnt’ using a pie plot.