

DSPL Lab Assignment 2

Refer the Loan dataset in the given file and perform the following queries.

1. Identify the most frequent values for all categorical features
2. Give descriptive statistics of numerical features in the dataset. Comment about the distribution of data from it.
3. Replace the missing values in categorical features using appropriate techniques.
4. Demonstrate various encoding techniques for categorical features.
5. for numerical features, replace missing values using
 - a. using simple imputer (mean, median)
 - b. using random sample imputation
6. Give descriptive statistics of numerical features in the dataset after handling missing values. Comment about the distribution of data from it.
7. Plot following graphs. Label X and Y axis, give appropriate title to the graph.
 - a. Plot histogram for Loan Amount and mention no. of observations
 - b. Plot histogram for Loan Amount and mention no. of observations
 - c. plot bar graph showing income for graduate and non-graduate applicant and mention no. of observations
 - d. Plot the boxplot for Loan amount. Give the five value summary from it.
 - e. Comment on the correlation between Applicant's income and Loan amount using appropriate graph.
 - f. Give descriptive statistics of numerical features in the dataset. Comment about the distribution of data from it.

Subject Incharge

Dr. Shachi Natu