2 Approaches

1. Converting categorical variables to dummy variables
2. Converting categorical variables to label encoded variables

Steps

1. Importing the required packages
2. Changing the working directory
3. Reading Train and Test data
4. Preparation for removing columns with zero variance # Not Used
5. Function for removing columns with zero variance
6. Getting list of zero variance columns in train and test datasets
7. Remove zero variance columns in train data from test data

* *drops columns from test where variance is zero in test data as well as the columns for which variance is zero in train data*
* *train and test have different columns which have zero variance*
* *Hence dropping the same columns in train and test data. Dropping the columns with zero variance in train data from test data.*

1. Finding Unique, Total Count and NAs and writing it to a csv
2. Finding the list of categorical variables
3. Creating dummy variables from categorical variable
4. Deleting categorical variables from train and test
5. Plotting Scree plot to get the number of components which will explain 90% variance in data
6. Performing PCA on train and test data
7. Separating x and y variables to be passed to xgboost
8. Defining xgboost model
9. Predict from the xgboost model
10. Calculating Root Mean Square Error