You will find a file in Day 5 Folder named “nuclear_explosions.csv”. Perform the EDA on the data

1. Print the columns of the data
2. Rename the columns which is more relatable according to you
3. Check for missing values (presence and sum) (Some values are in string “NAN” format)
4. Treat the missing values using appropriate strategy
5. Check for duplicates and treat them
6. Look for unique values in categorical column
7. Check data types of each column
8. Plot for following analysis
    a. Country wise count of nuclear explosion
    b. Deployment location count
    c. Year wise count
    d. Country wise average explosion yield
    e. Surface Type count
    f. Surface type average explosion yield
    g. Data Purpose Distribution (Frequency)
    h. Correlation between yield and depth of nuclear test
    i. Box plot of all numeric variables

Task 2 : Try adding 10 more random observations to the data in supervised learning flow

Re perform training and testing

Play with random_state and train_size

Observe the accuracy of both the train and test