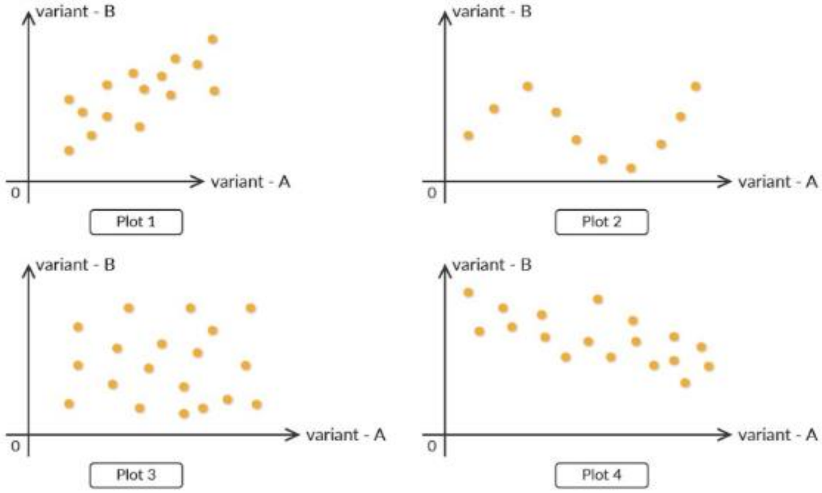
Dataset description:

Dataset contains the values of 39 persons (P1 to P39). It has a total of 7 columns. C1, C2, C3, C4, C5, C6, C7. Column C1 to C6 are independent columns or features and C7 is target column or dependent column containing three levels as ‘N’, ‘S’, ‘I’.

Questions:

1. Do the null value imputation if Column 'C6' is binary and has mean value of 0.040941.
2. Calculate the percentage of each level of column 'C7'.
3. Plot the mean and range of column 'C5' w.r.t each level of the target column.
4. Find the count of persons having 'C3' & 'C6' = 0 in common.
5. Which two levels of target column are overlapping in nature according to given independent features. Show the overlap on a plot to support your answer.
6. Create a data frame using ‘group by’ function showing counts of each level of target column and mean values of column ‘C4’.
7. Find the person code that got highest number of entries in the dataset.
8. Find the person code that got highest percentage of level 'I' w.r.t others.
9. Out of the independent features among C2 to C6, arrange these features in ascending or descending order of importance depending upon any feature selection method you wish to apply. Write the code and explain.
10. Consider the following four scatter plots of two variables A and B.



    a) Out of the four plots, which has least correlation coefficient?
    b) Which of the plots above has very a low correlation coefficient but has some relation between the variables?
    c) Which of the plots above has a negative correlation?

11. Consider column 'C2' as a time-series column.
    Null Hypothesis (H0): This is not stationary.
    Alternate Hypothesis (H1): This is stationary.
    Perform any stationary test which can show that column 'C2' is stationary or not.
12. Choose any algorithm of your choice and perform classification task for the dataset given.
    a) Print the confusion matrix.
    b) Calculate the F1 score and Recall. (Scores of 0.75+ will be appreciated).