CSE514, Fall 2022, HW 5   Name: Student ID:

Note: This homework is worth a total of 15 points

Please refer to the
        Color preferences.csv
file posted on Canvas to find the probability values needed for this first page of questions.

Q1 (9pts): Create a stump decision tree by splitting the data based on purple preference, where the left child holds all the samples that answered Yes, and the right child holds all the samples that answered No.
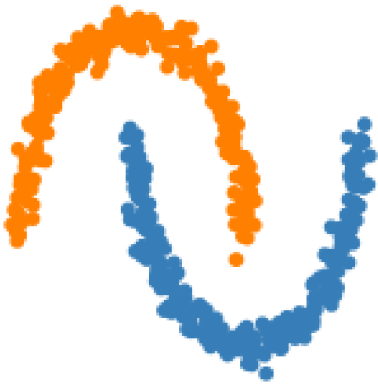
> **Q1a:** Given the task of predicting blue preference, what's the Total Weighted Gini Impurity of this split?

> **Q1b:** What would be the predicted labels of the left and right leaves?
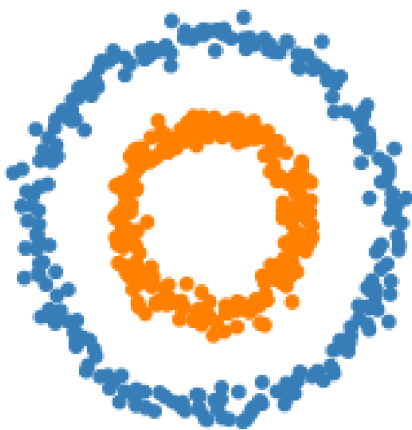
> **Q1c:** Given the task of predicting percentage of white t-shirts, what's the Total Weighted Variance of this split?

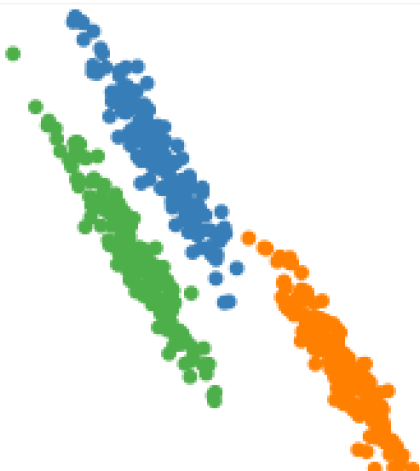> **Q1d:** What would be the predicted values of the left and right leaves?

Q2 (6pts): For each of the following clusters, give a clustering method you would NOT recommend, and a reason why you think it would fail.
Don't repeat clustering methods

Q2a:

Q2b:

Q2c: