Assignment No.2

1). A dataset contains 300 samples classified into two classes: 220 positive and 80 negative.

a. Compute the Gini index before splitting.

b. If a split results in subsets: Left: (90 positive, 10 negative) Right: (100 positive, 100 negative) Compute the weighted Gini index and determine whether the split improves purity.

2) Consider the given dataset with two independent variables (x ,x) and one dependent variable (y):

a. Use the sum of squared errors (SSE) to determine the best splitting point for x .

b. Construct the first split of a regression tree using SSE as the impurity measure

X1	X2	Y
1	5	10
2	6	12
3	8	15
4	10	18
5	12	21
6	15	25
7	18	28
8	20	30