Clustering demographic data using a classification tree:

- I created random data that resembles the original data and a column target with class as 1.
- Named the target as 0 and randomly permuted the features in the dataset.
- Built a decision tree model.

```
> summary(model)
Call:
rpart(formula = target ~ ., data = acombined_data)
  n= 18818

    CP nsplit rel error xerror xstd
1 0.01      0         1      0    0

Node number 1: 18818 observations
  predicted class=0  expected loss=0.5  P(node) =1
    class counts:  9409  9409
   probabilities: 0.500 0.500

> predicted = predict(model, acombined_data[,-c(15)])
> predicted
           0   1
    [1,] 0.5 0.5
    [2,] 0.5 0.5
    [3,] 0.5 0.5
    [4,] 0.5 0.5
    [5,] 0.5 0.5
    [6,] 0.5 0.5
    [7,] 0.5 0.5
    [8,] 0.5 0.5
    [9,] 0.5 0.5
   [10,] 0.5 0.5
```

- From the above model, we can observe that features do not have any predictive power to do a classification as it has only single root indicating that.
- To cross verify this, we can predict the model on the training set itself and observe that the probability for every row is one-half for both the classes.
- Therefore, we can conclude that it doesn't have any predictive power.