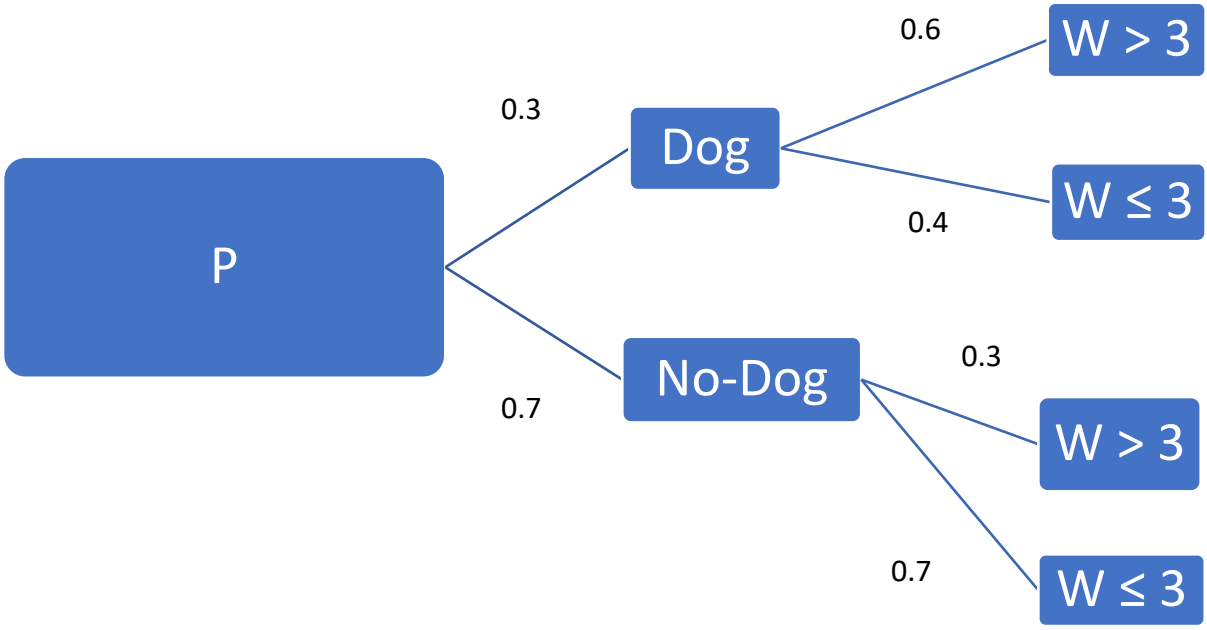
1.We are benchmarking two landing pages in our web client. Version A got 1000 new visitors of which 34 signed up and version B got 2000 new visitors of which 84 signed up. Which version performs better?

R: The version B performs better; it got a success rate of 4.2 and attract double of visitors than the version A.

2. Given the dataset (feature1, feature2, label) (-1, -1, 0), (-1, 1, 1), (1, -1, 1), (1, 1, 0), would you prefer to train a Logistic Regression or a Decision Tree Classifier for label prediction? Why?

R: At first sight I would say Logistic Regression, it's a binary classification task, but I think the best way to approach the problem will be applying a Classification Tree, the way they define the decision boundaries will adapt better to the features, I can't see this problem being separate with just a line.

3. The probability of a pet being a dog is 30%. The probability that a pet dog weighs more than 3 kg is
60%. If a pet is not a dog, the probability of it weighing less than or equal to 3 kg is 70%. What is theProbability of a pet being a dog if it weighs more than 3 kg?
R:



P(W>3) = P(D)*P(W>3|D) + P(ND)*P(W>3|ND)

P(W>3) = 0.6*0.3 + 0.3*0.7 = 0.39

P(D|W>3) = P(W>3|D)/P(W>3) = 0.21/0.39 = 0.54

4. We are solving a classification problem using a Decision Tree. When increasing the depth of the tree the accuracy improves until it stagnates at 90%. Why is this an unexpected behaviour? Can you offer a plausible explanation?

R:
I suppose that the 90% accuracy is in a test set, when we increase the depth of a tree, it becomes more complex, it will get more information but it is prone to overfitting. So usually, it will perform great on the train set and poorly on the test/validation set

5. What would you change in this python line?

operationResult: str = 193;

Remove the semicolon