DAT 520 Problem Set Five

Michael Surdek

Southern New Hampshire University

At first glance, I think the decision tree is predicting someone’s age based on any combination of selected categories such as native country, gender, job, and highest level of education from a sample of 965,173 people.

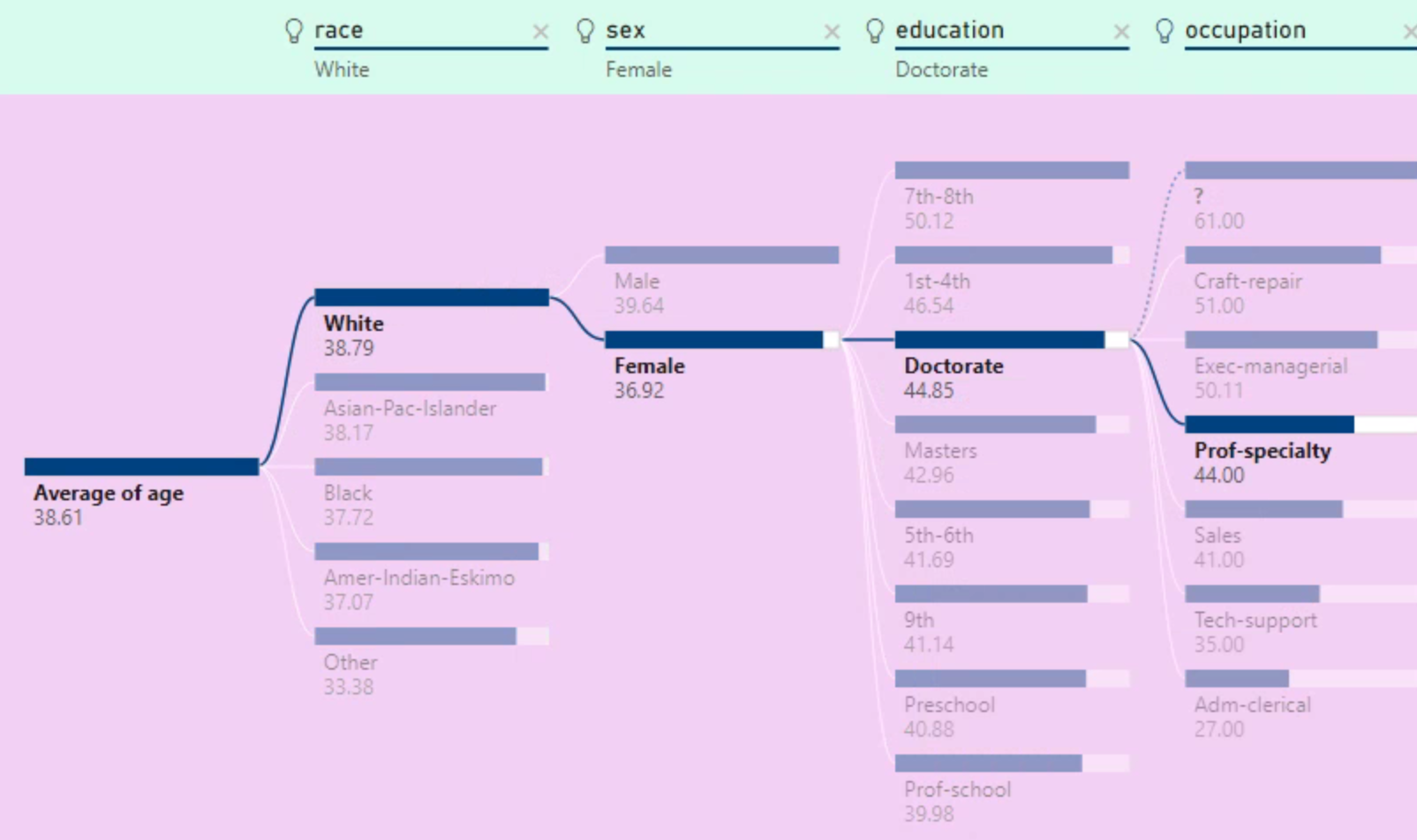
# All but age, capital gain, capital loss, education fields (2), hours-per, occupation, race, and sex:

# 

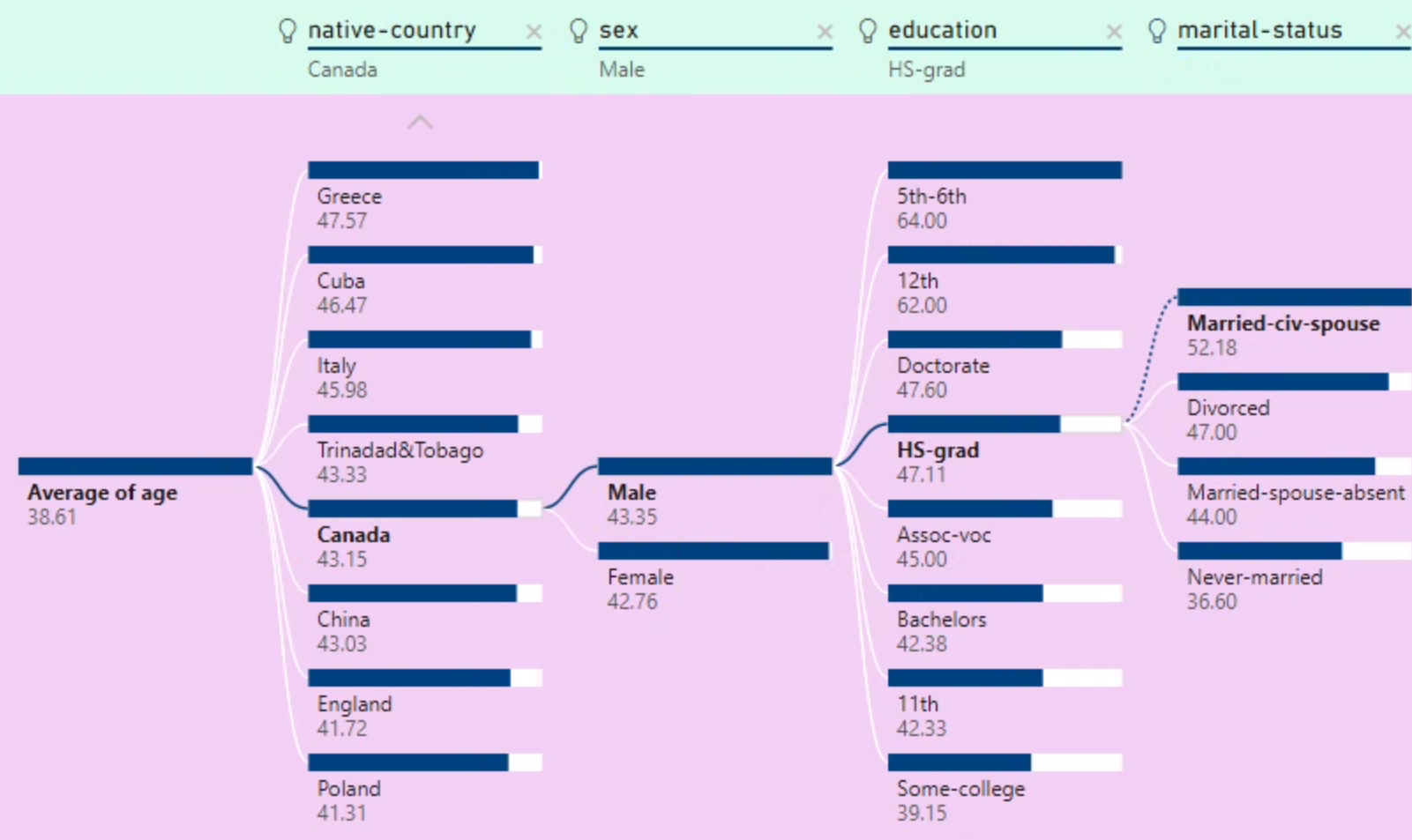
# What is the average age with these new selections?

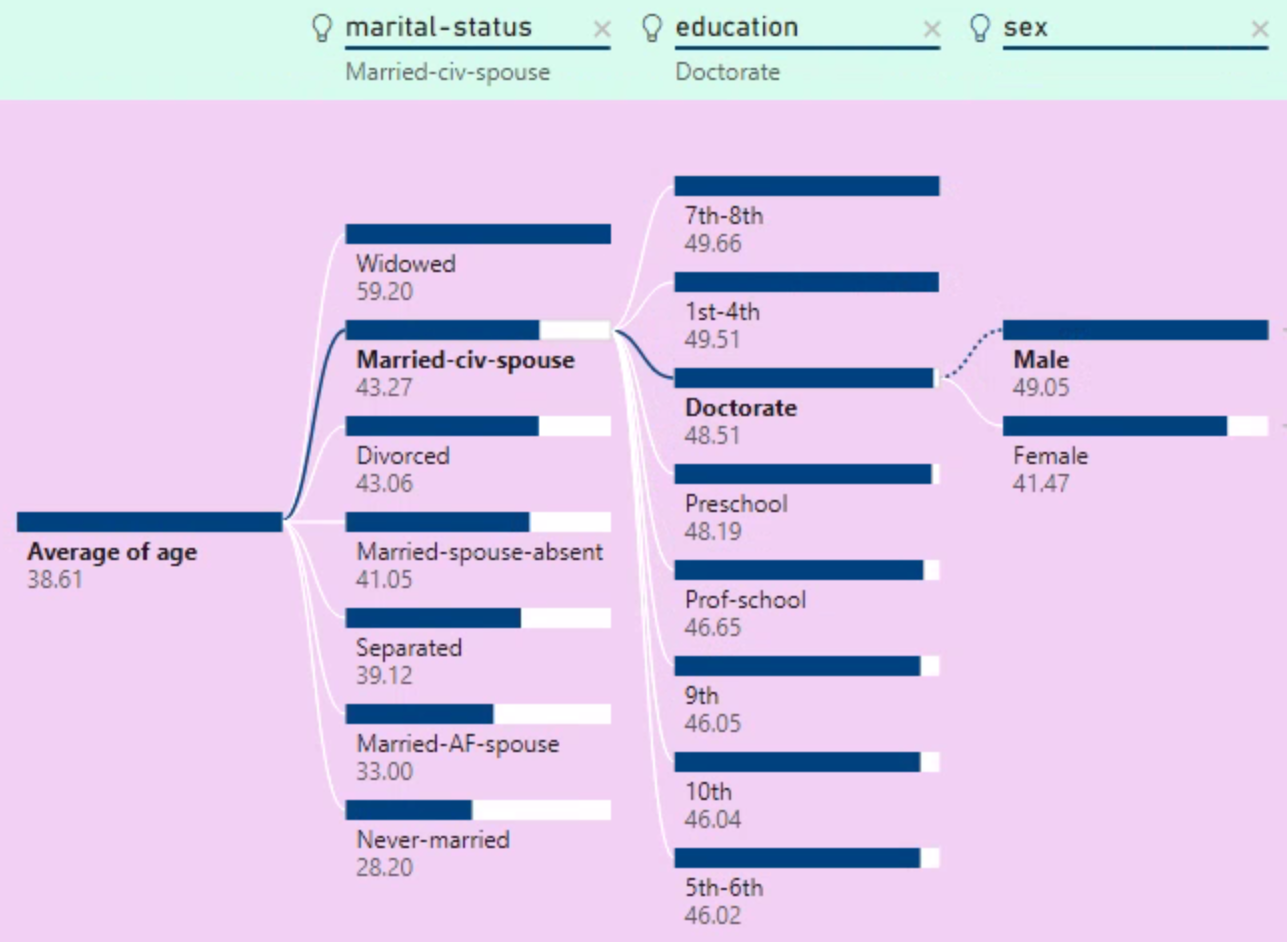
38.61

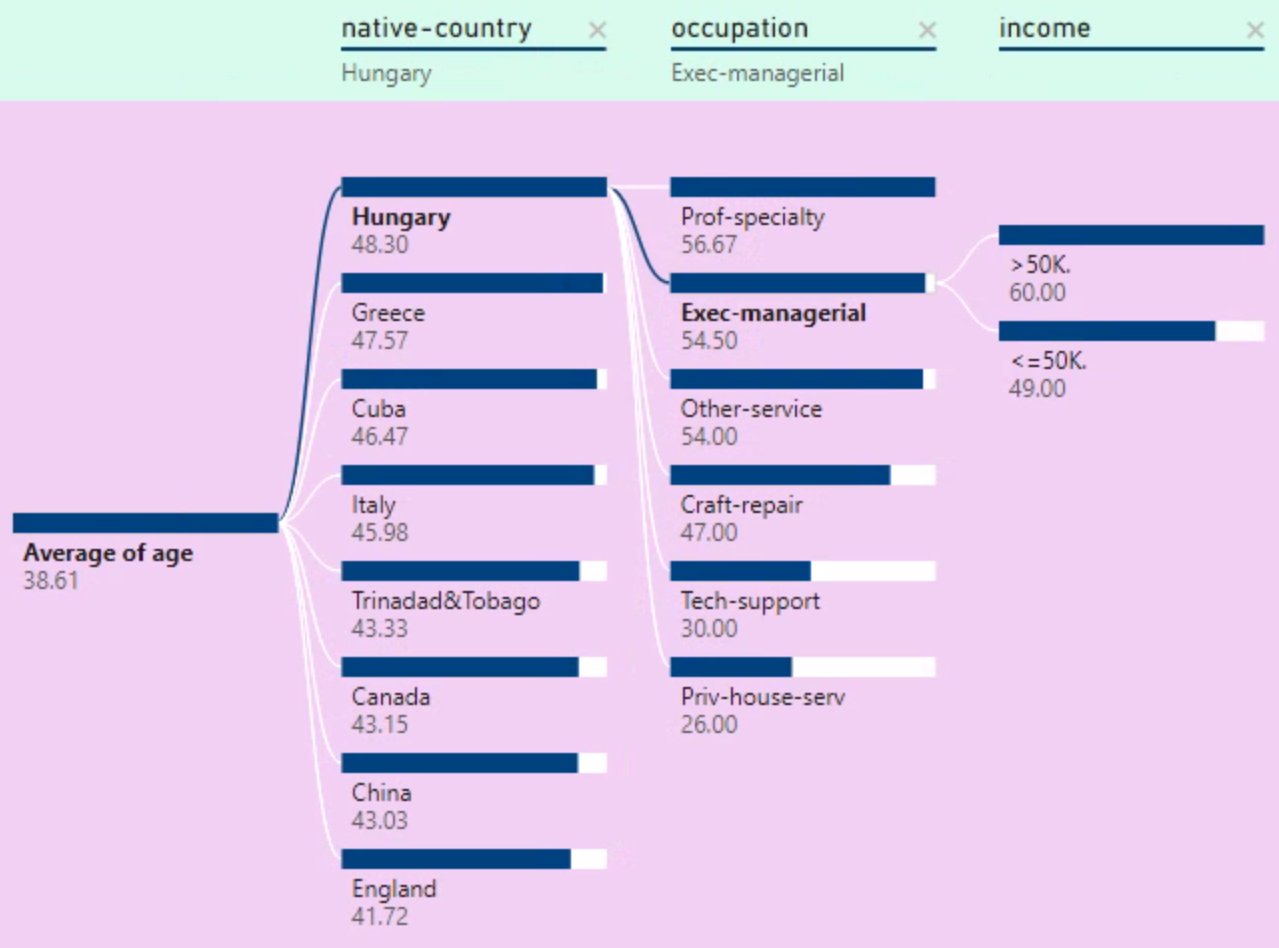
Weighted percentage of white, female, occupation of prof-specialty, with a doctorate degree:



The numbers change as you click on different inputs because you are selecting factors that change the sample set. As you move along the decision tree, the average age listed is based only on the people who fit that category and all of those to its left. For example, the 44.85 under Doctorate predicts the age of a person given that they are white, female, and a doctor. Additionally, the 38.79 under White predicts the age of a person given that they are white. Changing the inputs and the selections changes the combinations of variables that are used to calculate the age prediction.

 The calculations in the far-right column of this decision tree predicts a person’s age based on various martial statuses given the fact that the person is a Canadian male high school graduate. If the person is currently married, the decision tree predicts that their age is 52.18, and if the person is divorced, the decision tree predicts that their age is 47.00, given the selected values for native country, sex, and education.

 The calculations in the far-right column of this decision tree predicts a person’s age based on their sex given the fact that the person is married and has a Doctorate degree. If the person is male, the decision tree predicts that their age is 49.05, and if the person is female, the decision tree predicts that their age is 41.47, given the selected values for marital status and education.



The calculations in the far-right column of this decision tree predicts a person’s age based on their income given the fact that the person is Hungarian company executive. If the person makes over $50,000 per year, the decision tree predicts that their age is 60.00, and if the person makes under $50,000 per year, the decision tree predicts that their age is 49.00, given the selected values for native country and occupation.