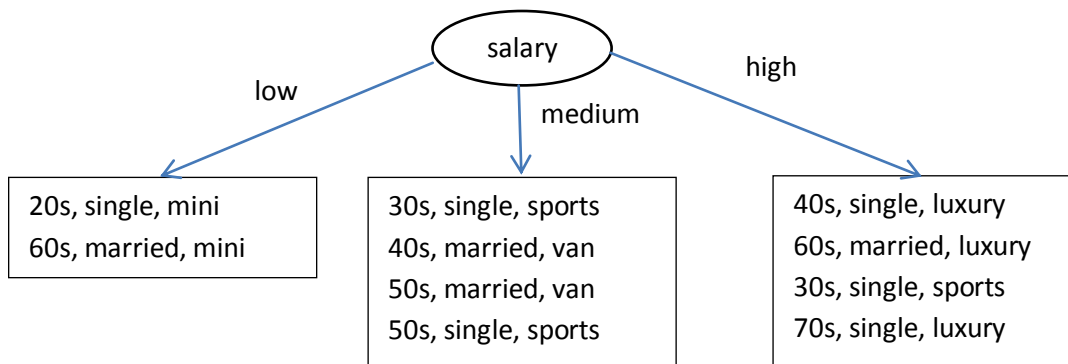


Quiz 2. Decision tree. Solution

The following dataset contains the demographic data about people who buy cars. The goal is to predict the car type which the particular person will probably buy if targeted with the specialized promotion.

age	salary	marital	car
30s	medium	single	sports
20s	low	single	mini
40s	medium	married	van
40s	high	single	luxury
60s	high	married	luxury
30s	high	single	sports
50s	medium	married	van
50s	medium	single	sports
60s	low	married	mini
70s	high	single	luxury

Select the best attribute to put at the top of the decision tree. Use the entropy to justify your answer.

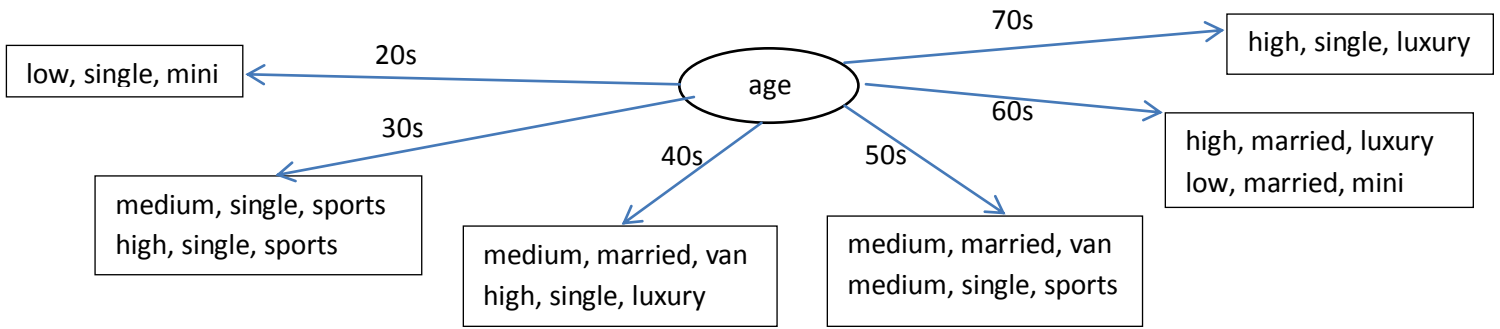


$$\text{Entropy}(2/2,0)=0$$

$$\text{Entropy}(2/4,2/4)=1$$

$$\text{Entropy}(3/4,1/4)=-3/4 \log 3/4 - 1/4 \log 1/4=0.81$$

$$\text{Average entropy: } 0.2*0+0.4*1+0.4*0.81=0.72$$



Entropy(1,0)=0

Entropy(2/2,0)=0

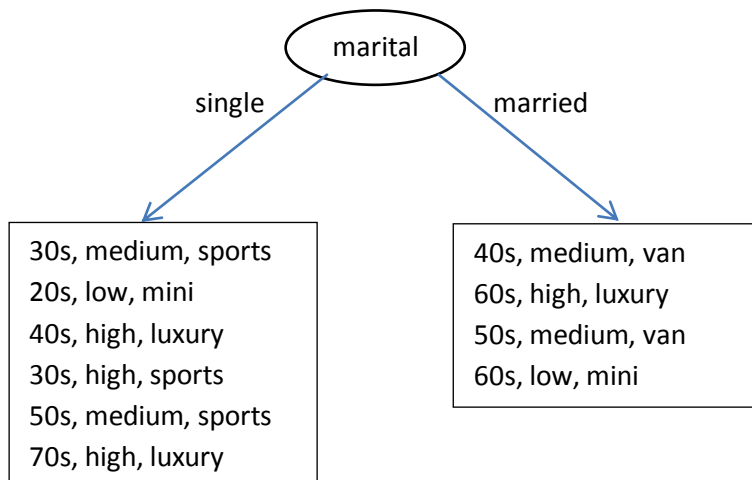
Entropy(1/2,1/2)=1

Entropy(1/2,1/2)=1

Entropy(1/2,1/2)=1

Entropy(1,0)=0

Average entropy: $0.2*1+0.2*1+0.2*1=0.60$



Entropy(1/6,2/6,3/6)=1.45

Entropy(2/4,1/4,1/4)=1.5

Average entropy: $0.6*1.45+0.4*1.5=1.475$

Answer: choose age