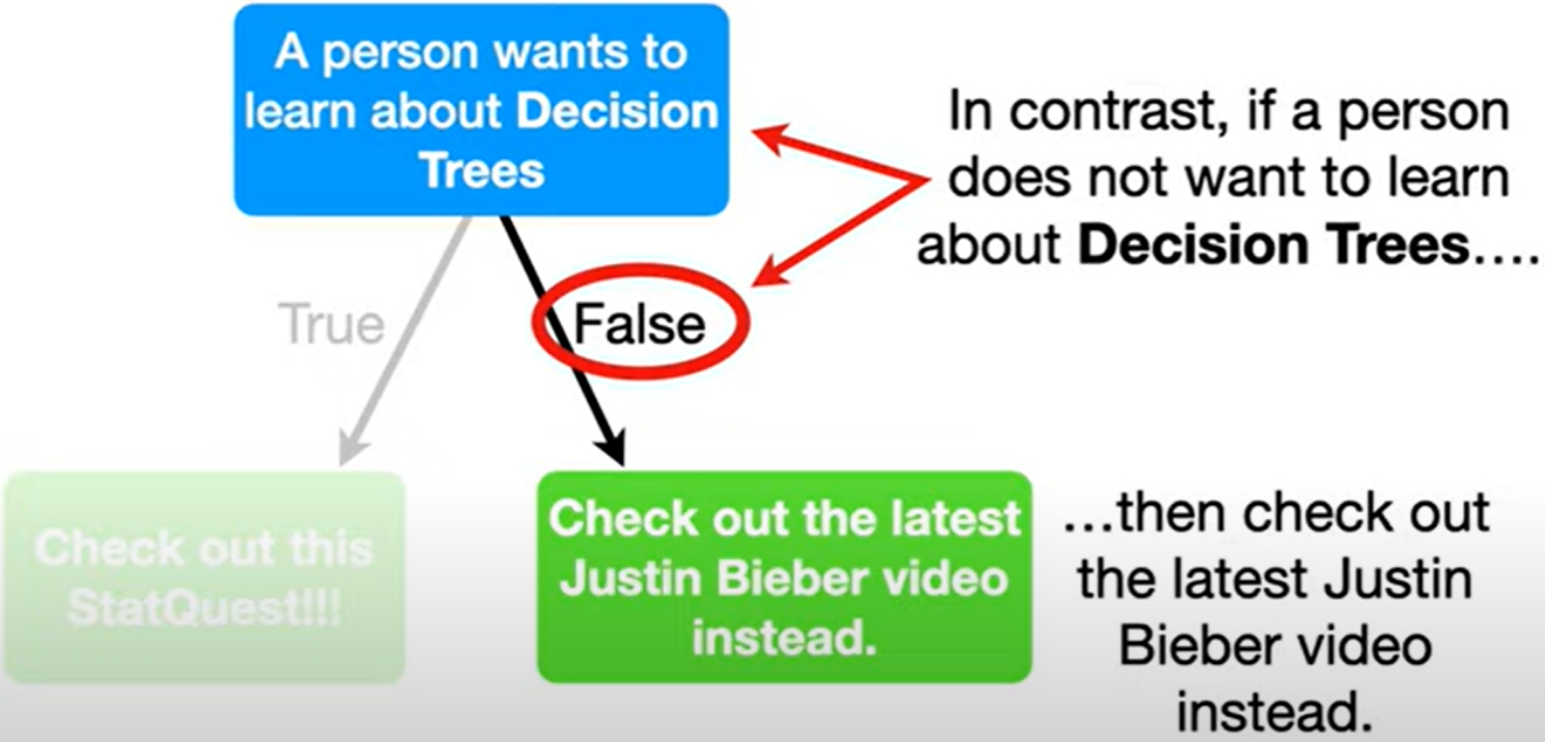
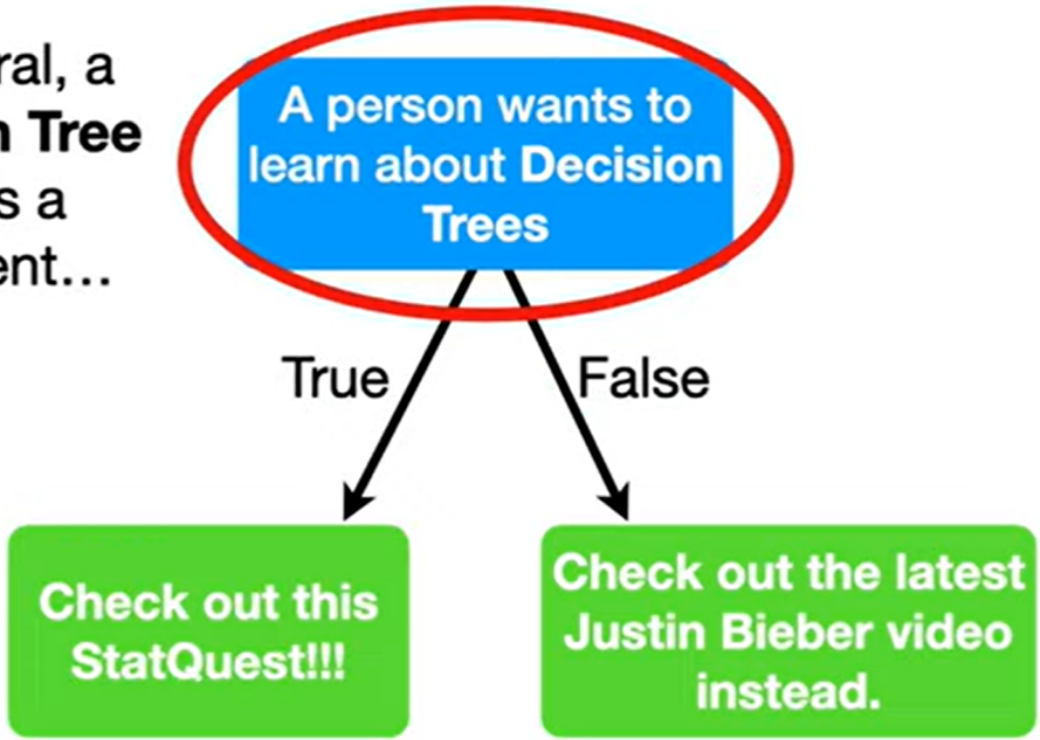
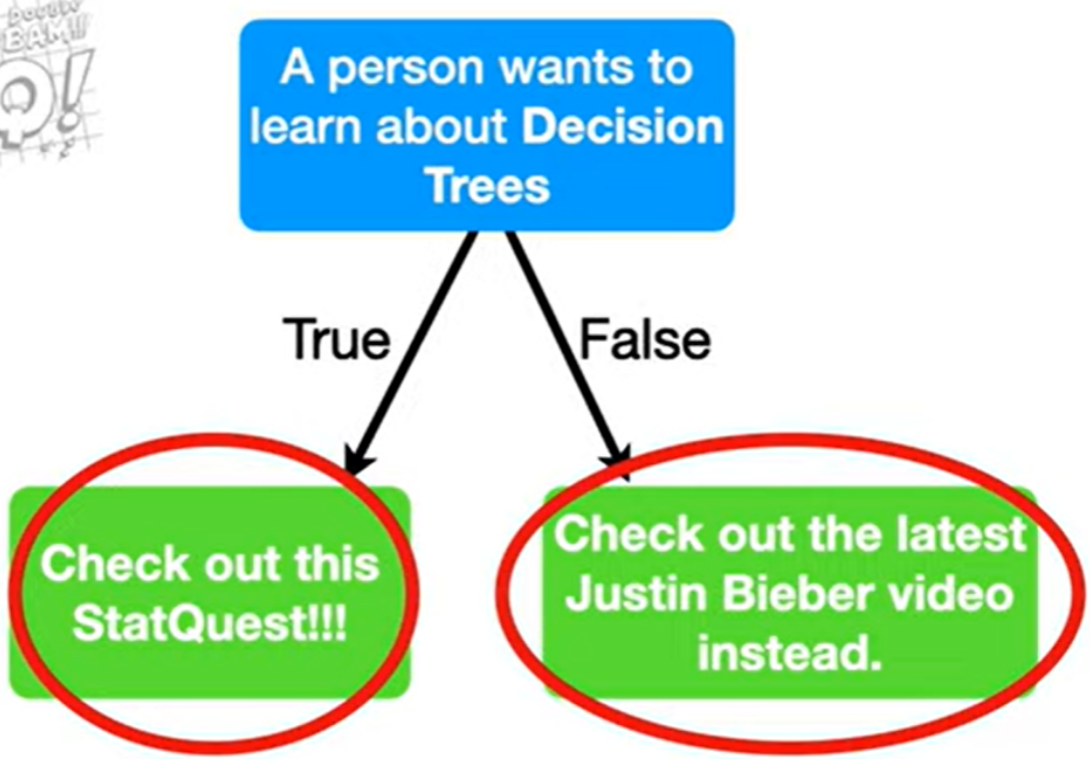

A person wants to learn about Decision Trees
True
False
Check out this StatQuest!!!
Check out the latest Justin Bieber video instead.
In contrast, if a person does not want to learn about Decision Trees....
...then check out the latest Justin Bieber video instead.

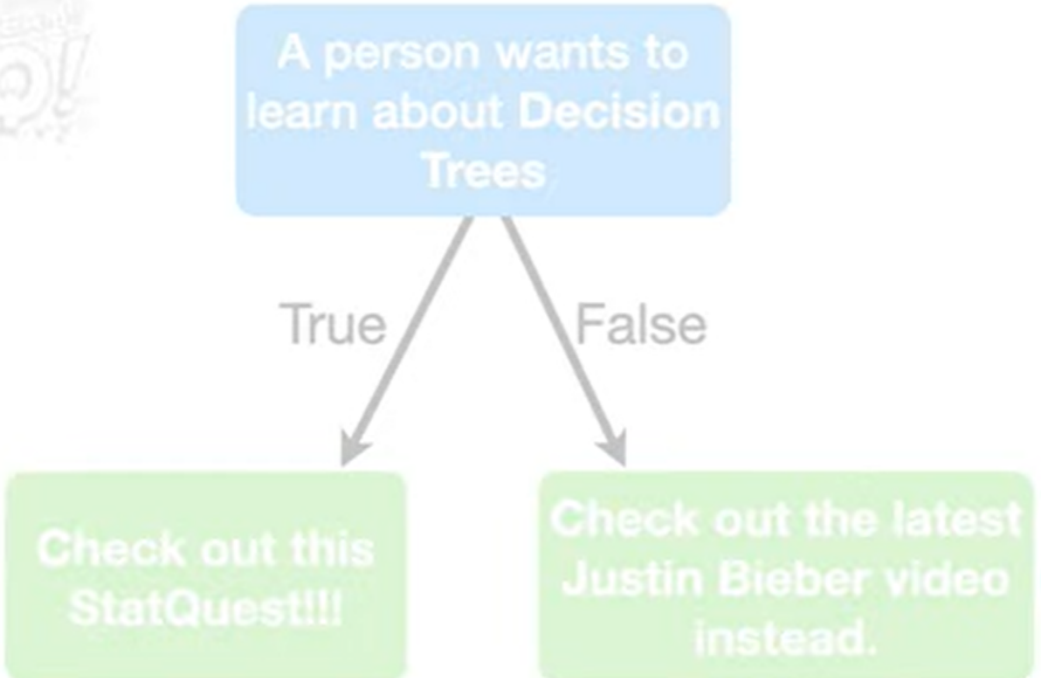
In general, a **Decision Tree** makes a statement…





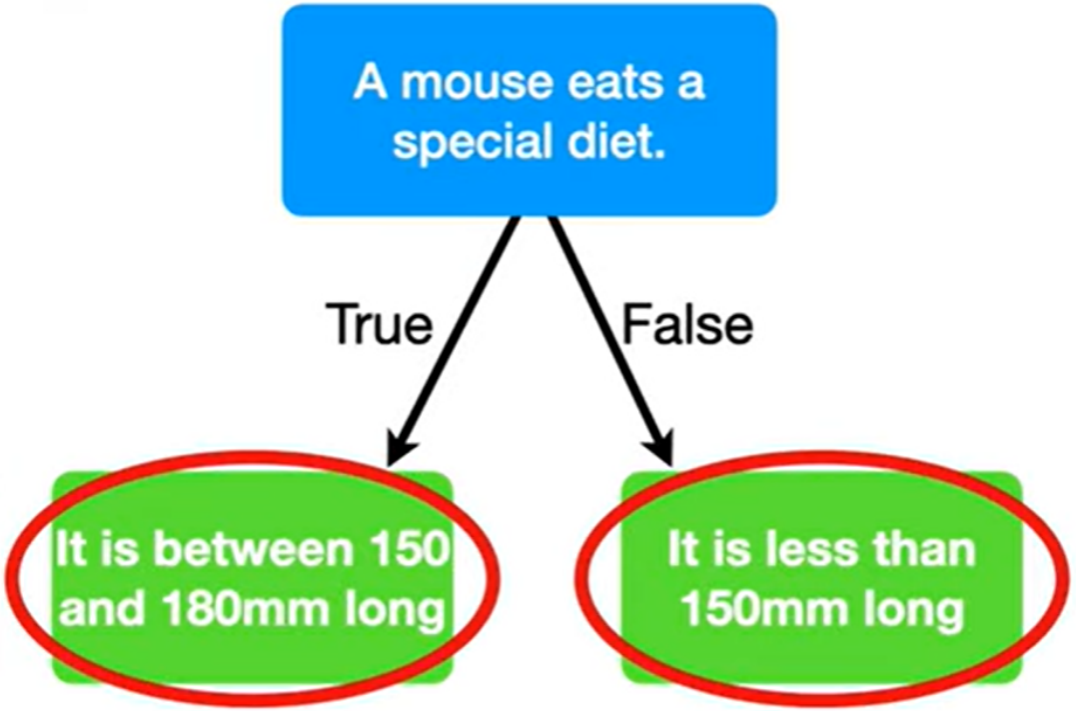
When a **Decision Tree** classifies things into categories…

…it's called a **Classification Tree**.



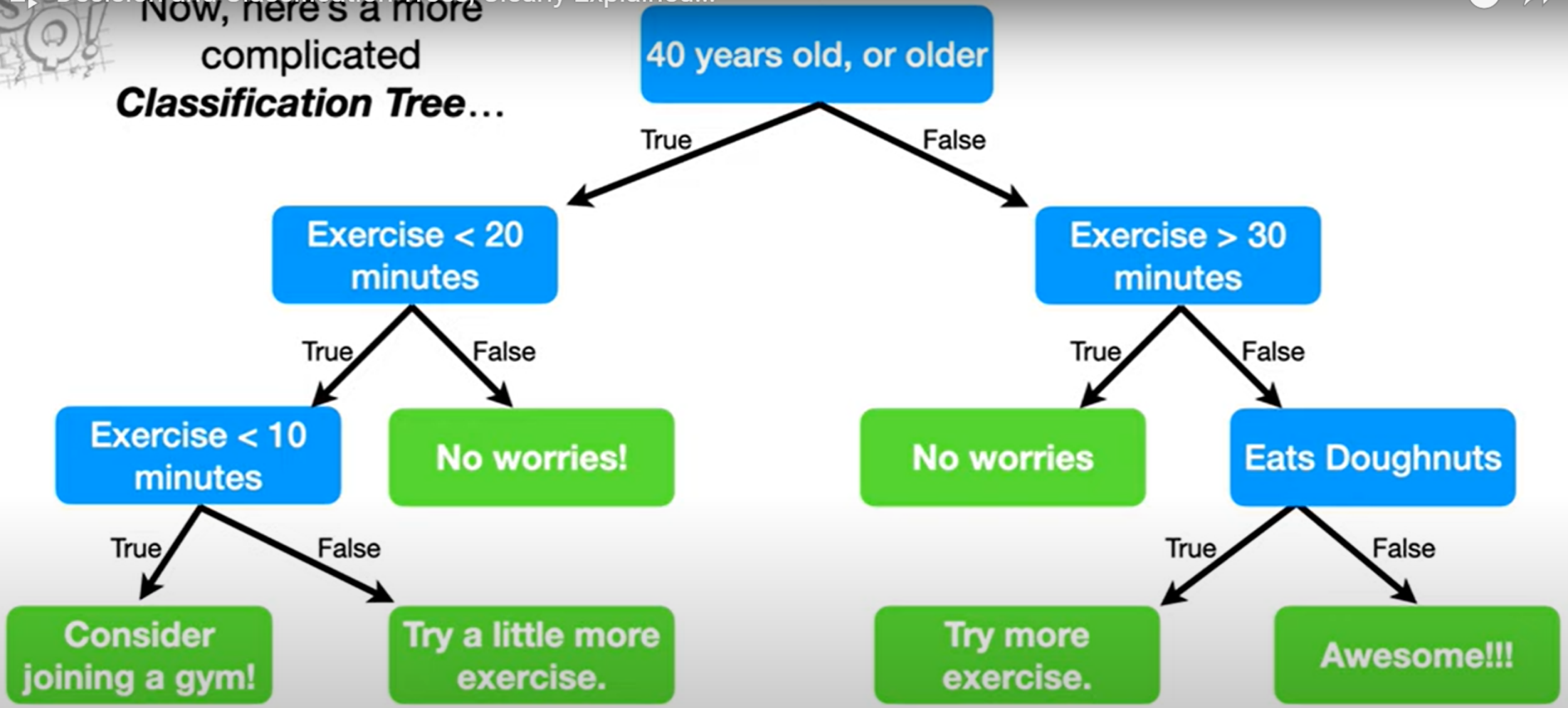
When a **Decision Tree** classifies things into categories…

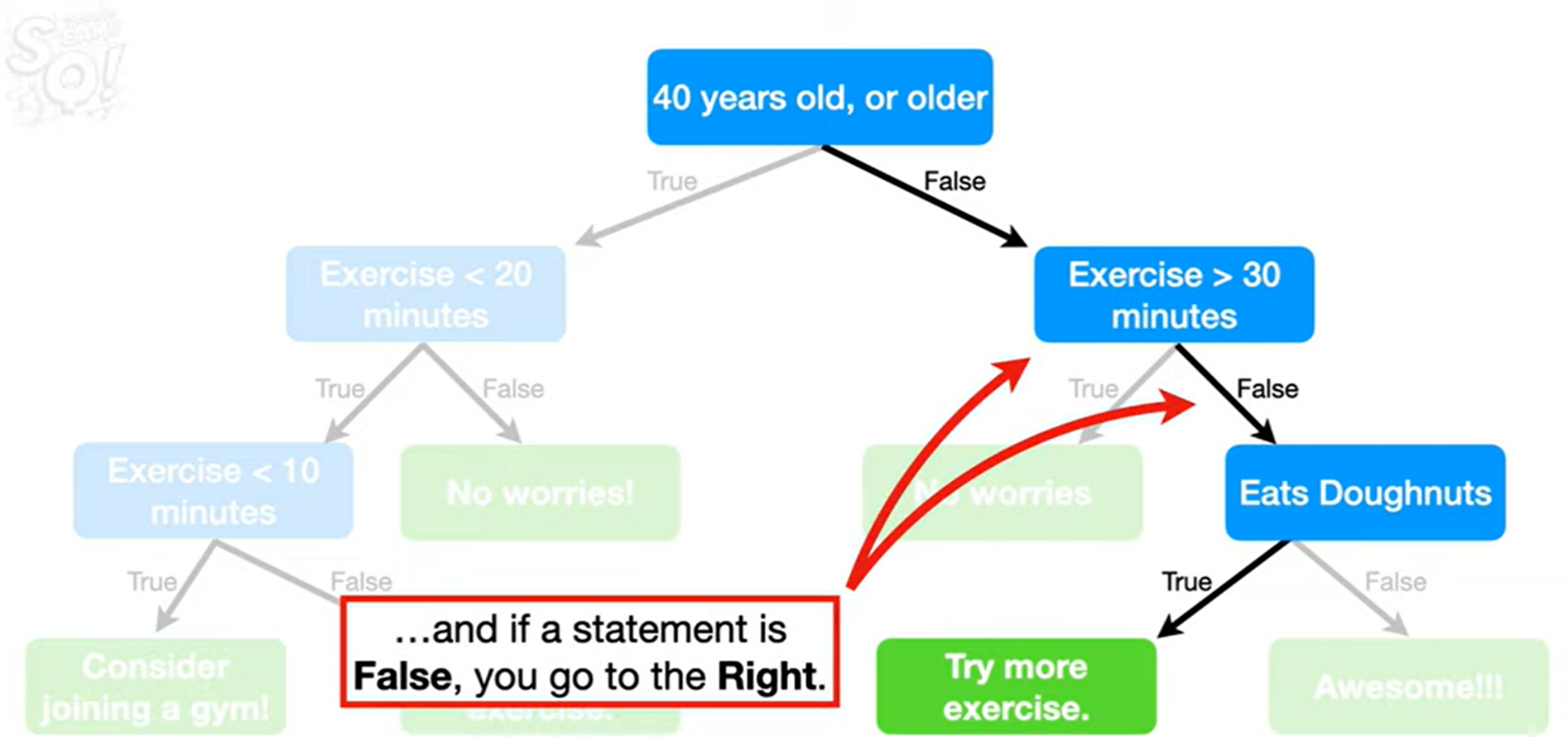
…it's called a **Classification Tree**.



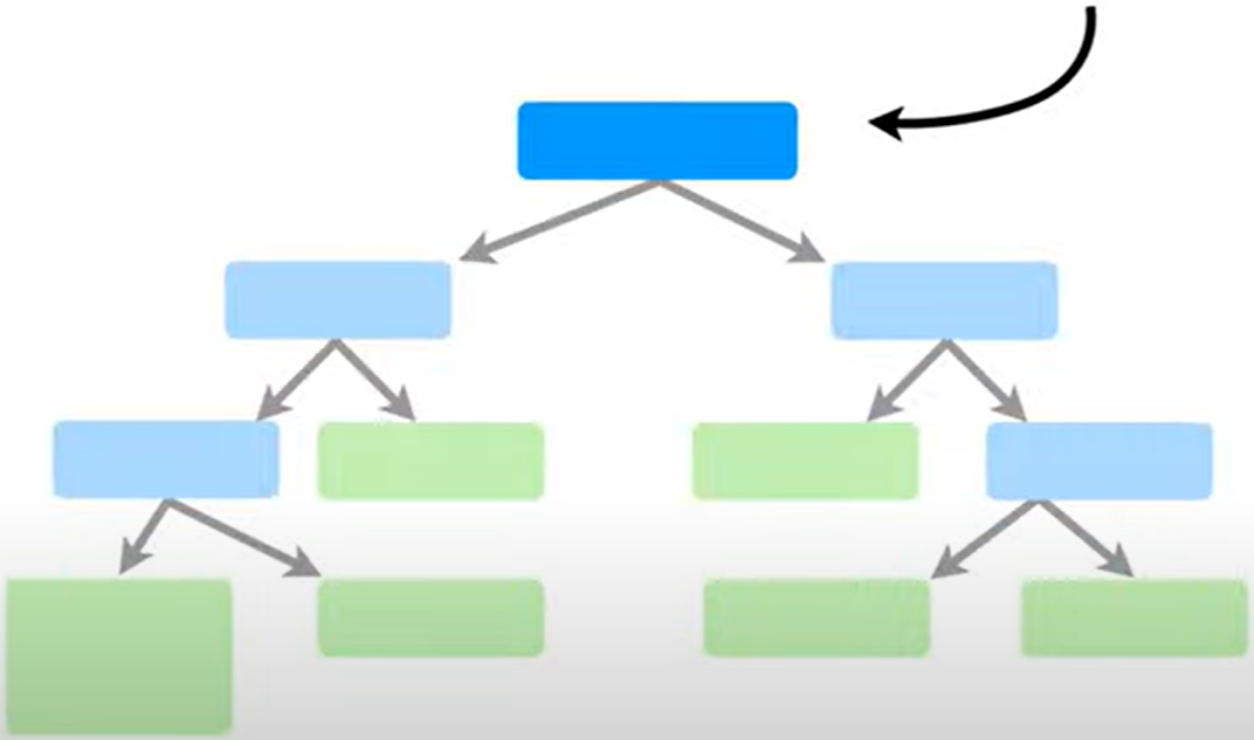
And when a **Decision Tree** predicts numeric values…

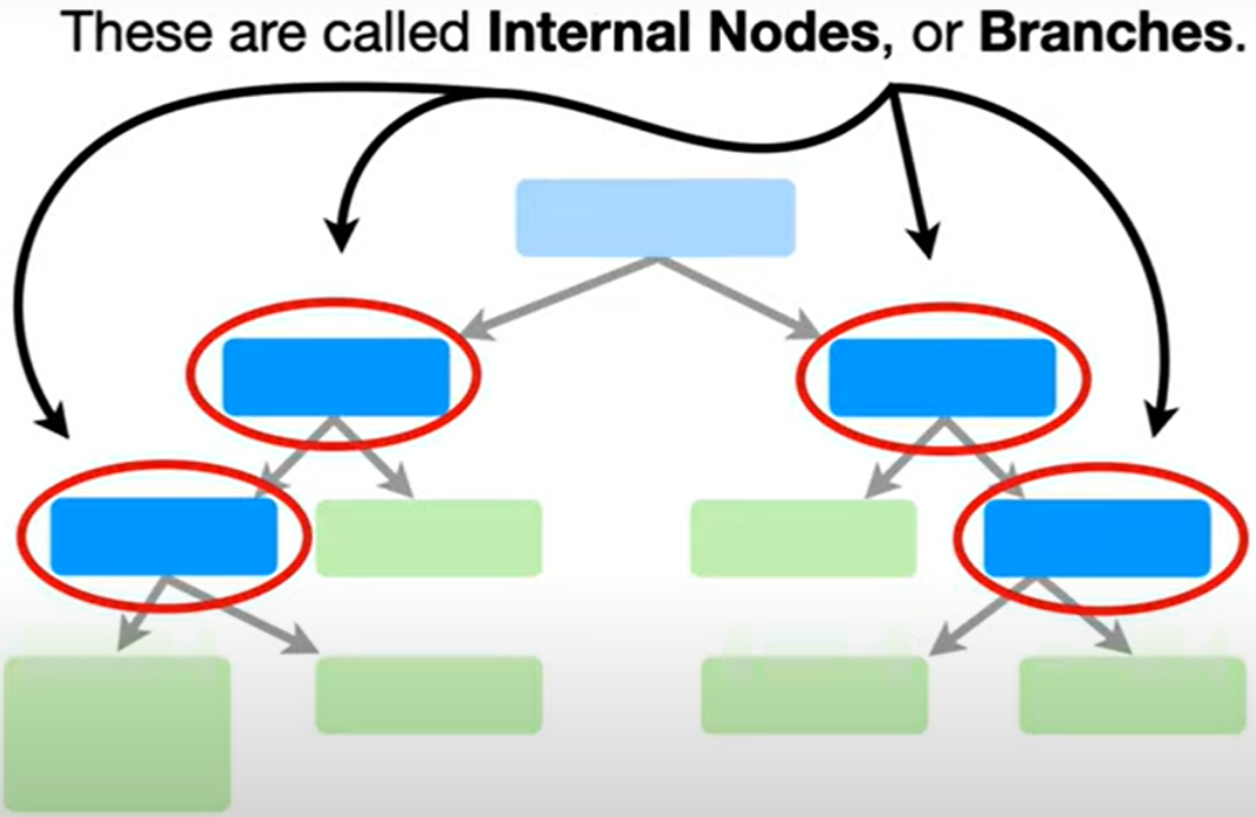
Now, here's a more complicated ***Classification Tree***…



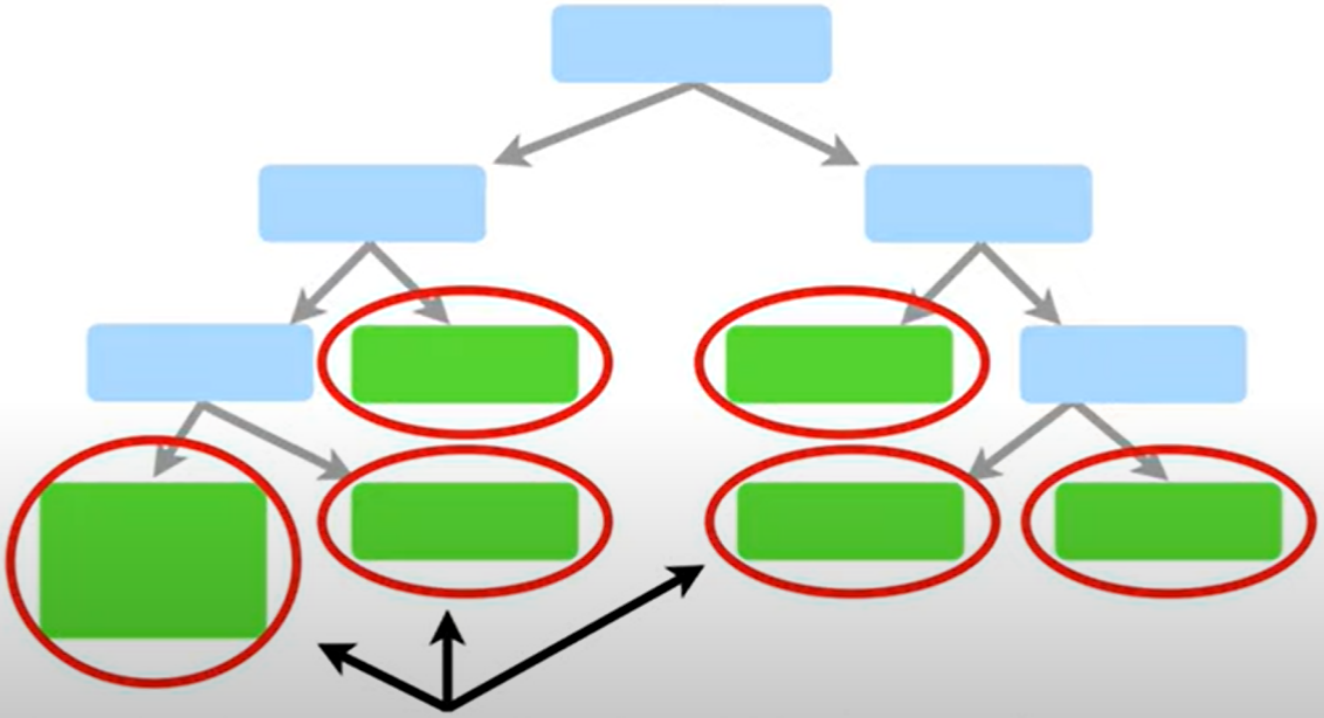

40 years old, or older
True
False
Exercise < 20 minutes
Exercise > 30 minutes
True
False
True
False
Exercise < 10 minutes
No worries!
No worries
Eats Doughnuts
True
False
True
False
Consider joining a gym!
…and if a statement is **False**, you go to the **Right**.
Try more exercise.
Awesome!!!

The very top of the tree is called the **Root Node** or just **The Root.**



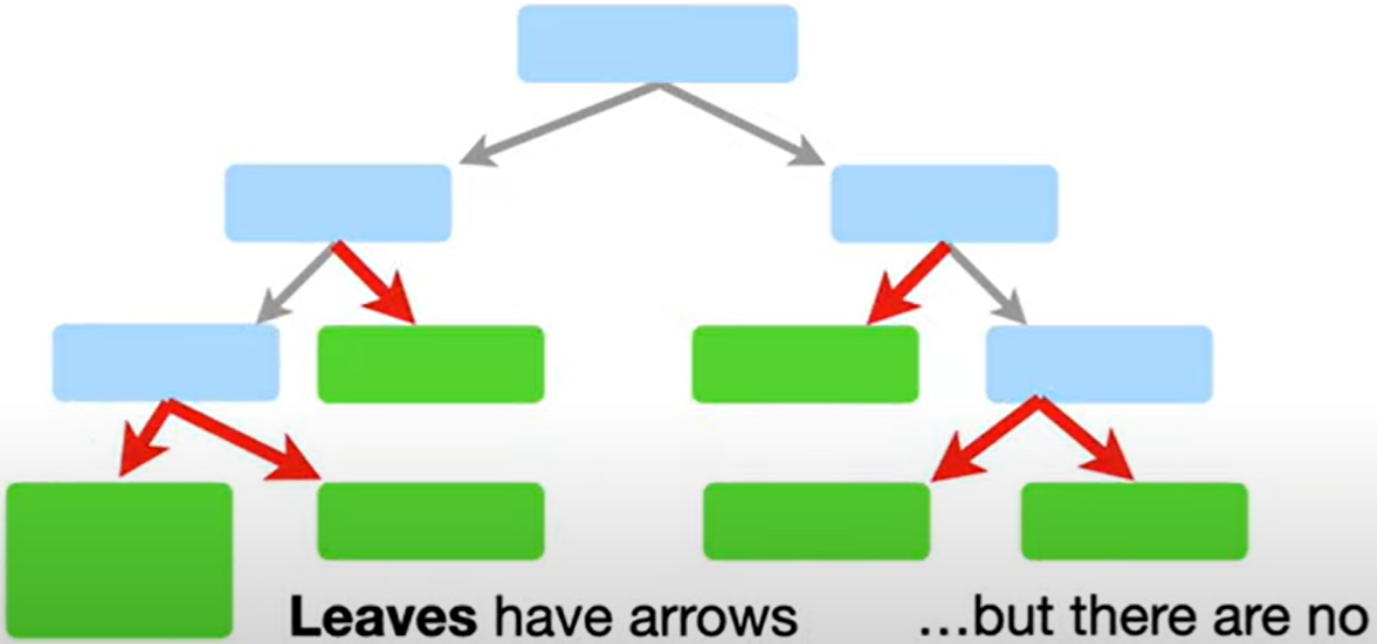

These are called **Internal Nodes**, or **Branches**.

Decision and Classification Trees, Clearly Explained!!!



Lastly, these are called **Leaf Nodes**, or just **Leaves**.

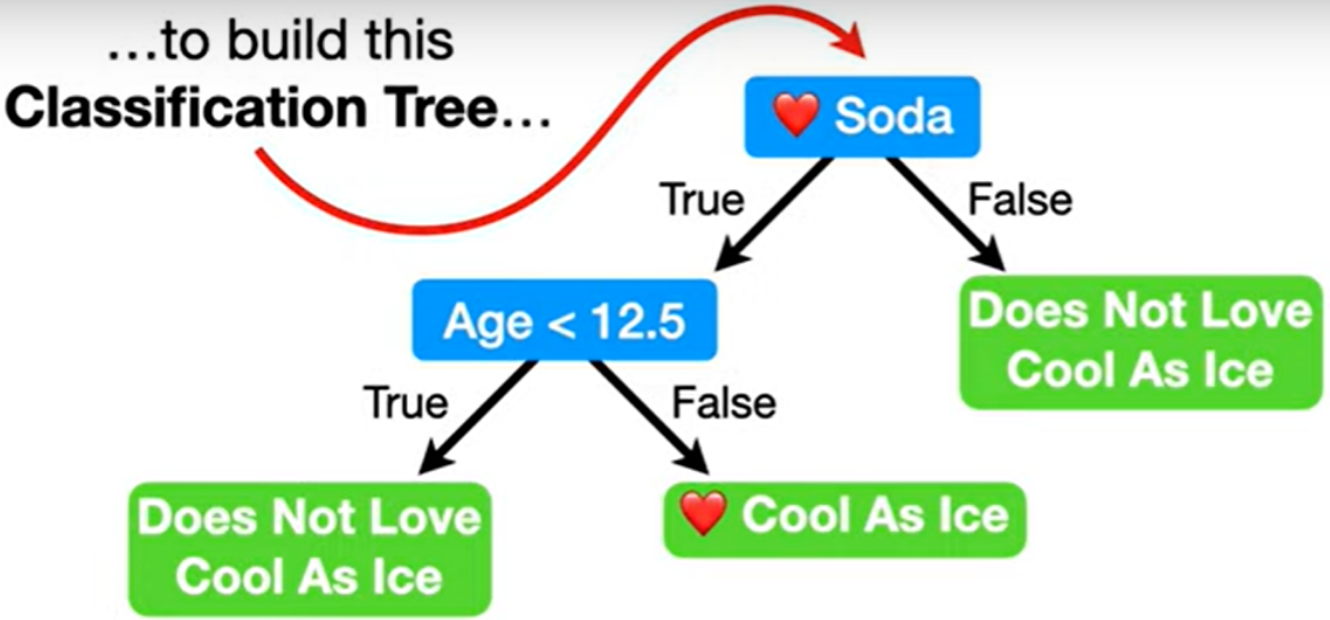
Decision and Classification Trees, Clearly Explained!!!

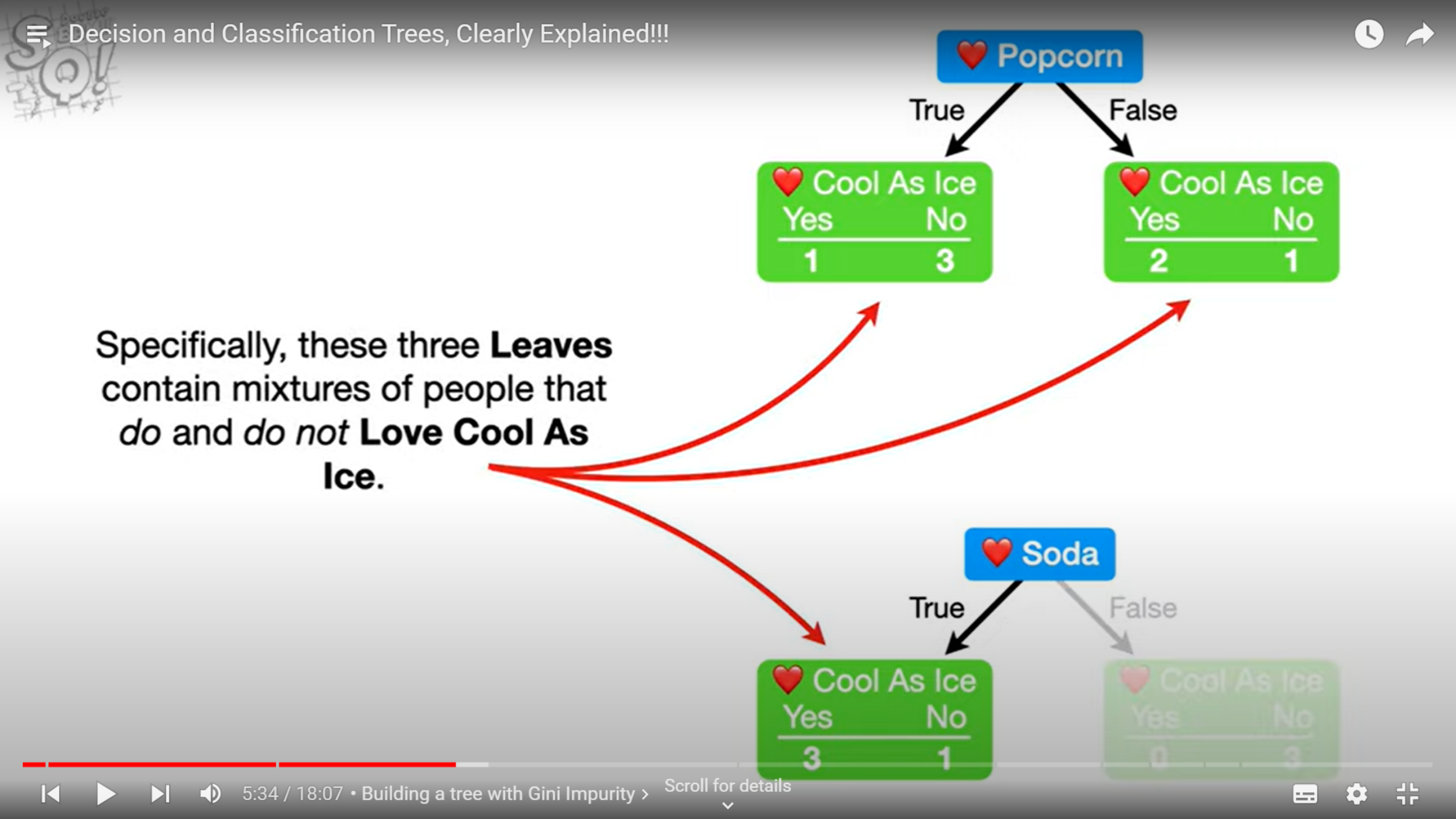


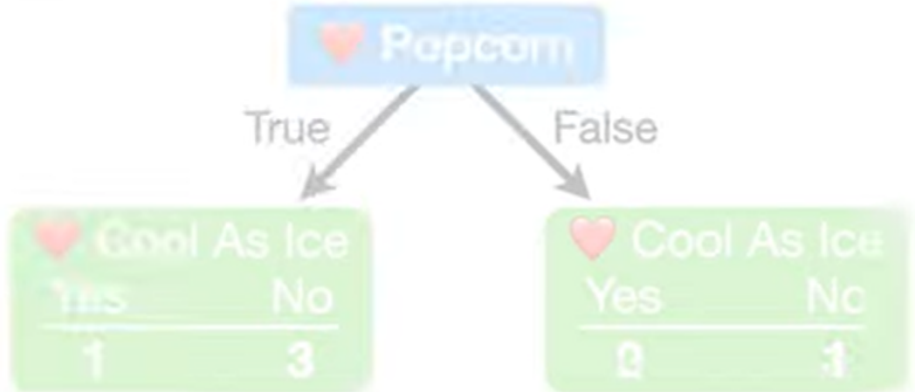
3:10 / 18:07 • Basic decision tree concepts ›

Scroll for details

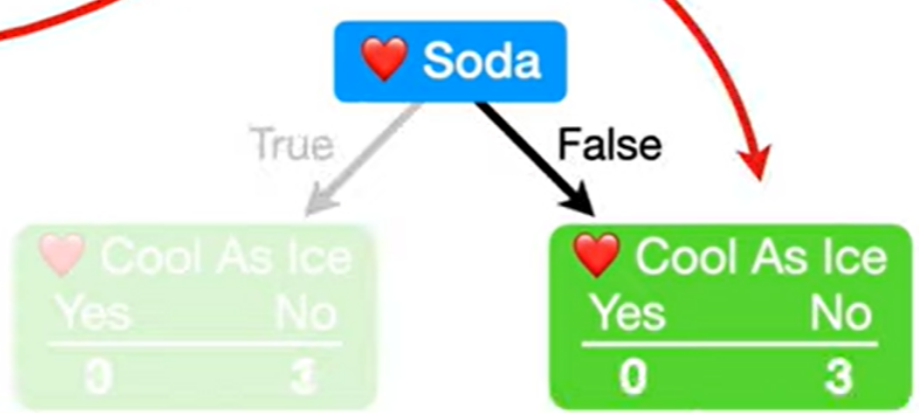
Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No

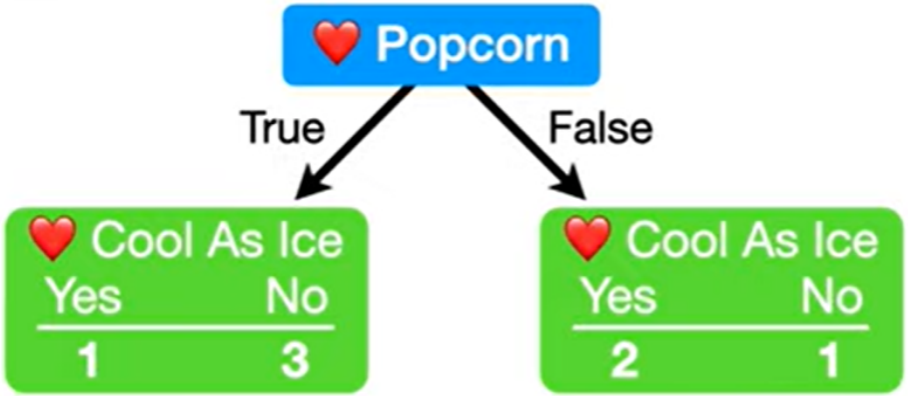



Popcorn
True
False
Cool As Ice
Yes No
1 3
Cool As Ice
Yes No
2 1
Specifically, these three **Leaves** contain mixtures of people that *do* and *do not* **Love Cool As Ice**.
Soda
True
False
Cool As Ice
Yes No
3 1
Cool As Ice
Yes No
0 3

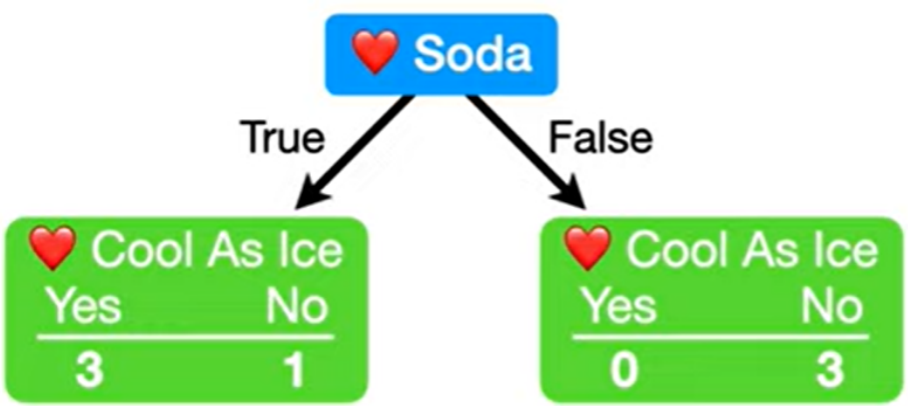


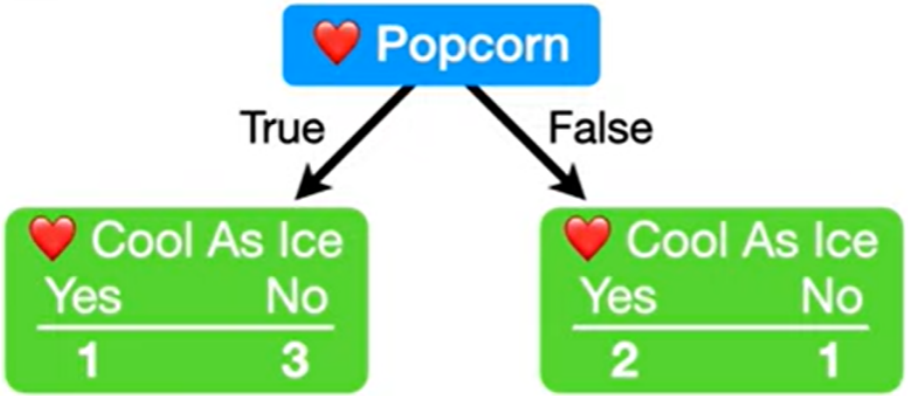
In contrast, this **Leaf** only contains people who *do not* **Love Cool As Ice**.



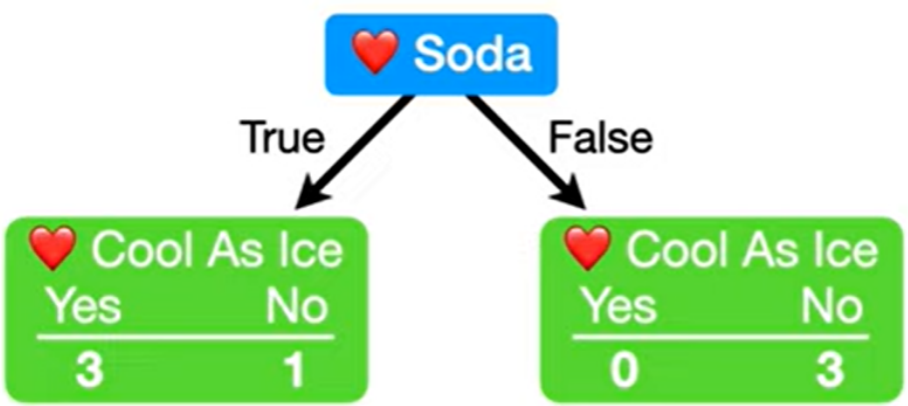


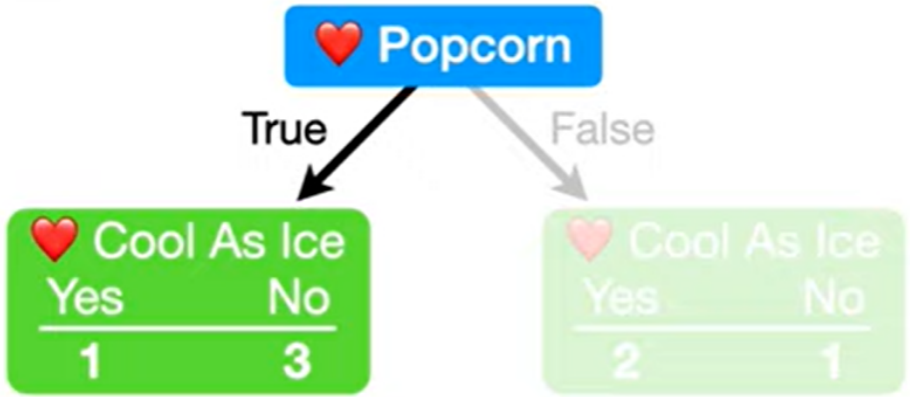
…and only one **Leaf** in the **Loves Soda** tree is **Impure**…





One of the most popular methods is called **Gini Impurity**, but there are also fancy sounding methods like **Entropy** and **Information Gain**.

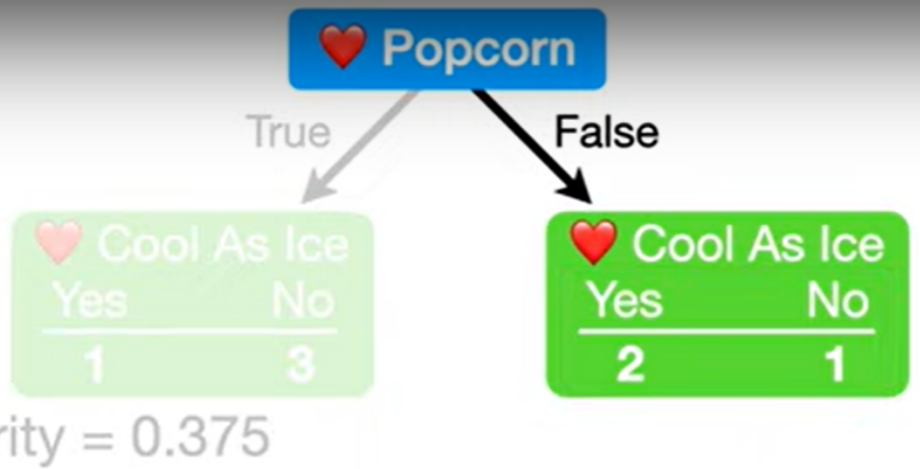




The **Gini Impurity** for the **Leaf** on the left is…

Gini Impurity for a Leaf = 1 - (the probability of “Yes”)^2 - (the probability of “No”)^2

Popcorn
True
False
Cool As Ice
Yes
No
1
3
Cool As Ice
Yes
No
2
1
Gini Impurity = 0.375
Gini Impurity for a Leaf = 1 - (the probability of "Yes")² - (the probability of "No")²
$= 1 - (\frac{1}{1+3})^2 - (\frac{3}{1+3})^2$
= 0.375
So let's put **0.375** under the **Leaf** on the left so we don't forget it.

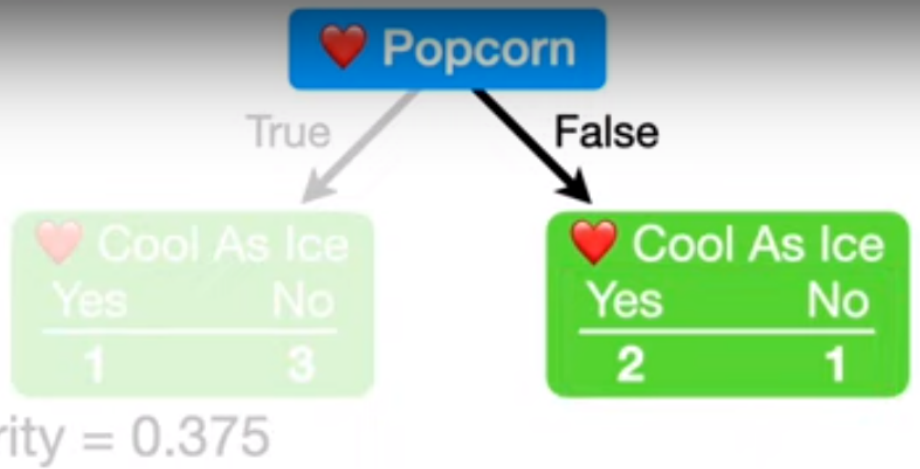


Gini Impurity for a Leaf = 1 - (the probability of "Yes")^2 - (the probability of "No")^2

$$= 1 - \left(\frac{2}{2+1}\right)^2 - \left(\frac{1}{2+1}\right)^2$$

$$= 0.444$$

And when we do the math we get **0.444**.

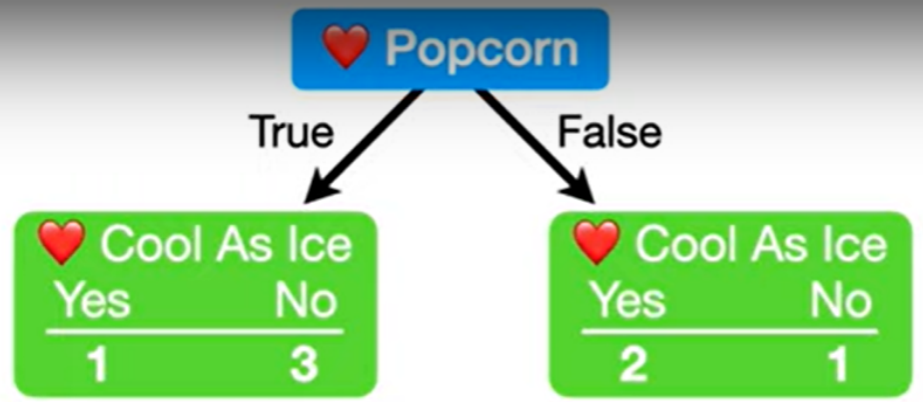


Gini Impurity for a Leaf = 1 - (the probability of "Yes")^2 - (the probability of "No")^2

$$= 1 - \left(\frac{2}{2+1}\right)^2 - \left(\frac{1}{2+1}\right)^2$$

$$= 0.444$$

And when we do the math we get **0.444**.



And when we do the math, we get **0.405**.

Gini Impurity = 0.375

0.444

Total **Gini Impurity** = weighted average of **Gini Impurities** for the **Leaves**

$$= \left(\frac{4}{4+3}\right) 0.375 + \left(\frac{3}{4+3}\right) 0.444$$

$$= 0.405$$

Lastly, we calculate the **Gini Impurity** values for each average age.

Age	Loves Cool As Ice
7	No
12	No
18	Yes
35	Yes
38	Yes
50	No
83	No

9.5 → Gini Impurity = 0.429

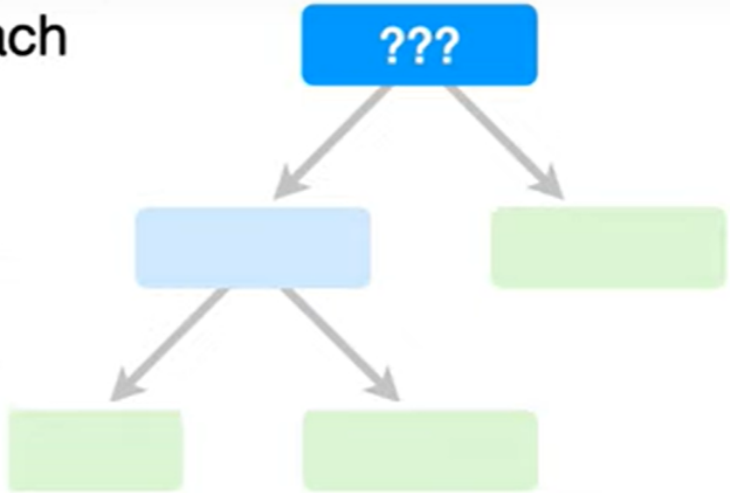
15 → Gini Impurity = 0.343

26.5 → Gini Impurity = 0.476

36.5 → Gini Impurity = 0.476

44 → Gini Impurity = 0.343

66.5 → Gini Impurity = 0.429



Lastly, we calculate the **Gini Impurity** values for each average age.

Age	Loves Cool As Ice
7	No
12	No
18	Yes
35	Yes
38	Yes
50	No
83	No

9.5 → Gini Impurity = 0.429

15 → Gini Impurity = 0.343

26.5 → Gini Impurity = 0.476

36.5 → Gini Impurity = 0.476

44 → Gini Impurity = 0.343

66.5 → Gini Impurity = 0.429