

Answer the following questions based on the table below:

	Sailboat	Company	Sail
1	Big	Small	No
3	Big	Small	No
2	Big	Big	Yes
4	Big	Big	Yes
5	Small	Small	Yes
7	Small	Small	No
6	Small	Big	Yes
8	Small	Big	No

$$-\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{3}{4}\log_2\left(\frac{3}{4}\right) = .811$$

- What's the entropy of the entire table?
- What's the entropy of Sailboat(Big) if our first split is by 'Sailboat'?
- What's the entropy of Sailboat(Small) if our first split is by 'Sailboat'?
- What's the expected/weighted entropy of Sailboat?
- What's the entropy of Company(Big) if our first split is by 'Company'?
- What's the entropy of Company(Small) if our first split is by 'Company'?
- What's the expected/weighted entropy of Company?
- Based on your answers above, what's the first feature that you should split by?

If your answer to 'Q40h' is 'Sailboat', then answer the next 6 questions

- 1) Are you going to split subtable Sailboat(Big) by another feature or not?
- 2) If the answer to 'Q-Sailboat-a' question is 'Yes', then what's the feature that you will split subtable Sailboat(Big) by?
- 3) If the answer to 'Q-Sailboat-a' question is 'Yes', then what's the information gain after making that split on subtable Sailboat(Big)?
- 4) Are you going to split subtable Sailboat(Small) by another feature or not?
- 5) If the answer to 'Q-Sailboat-d' is 'Yes', then what's the feature that you will split subtable Sailboat(Small) by?
- 6) If the answer to 'Q-Sailboat-d' question is 'Yes', then what's the information gain after making that split on subtable Sailboat(Small)?

If your answer to Q40h is 'Company', then answer the next 6 questions

- 1) Are you going to split subtable Company(Big) by another feature or not?
- 2) If the answer to 'Q-Company-a' question is 'Yes', then what's the feature that you will split subtable Company(Big) by?
- 3) If the answer to 'Q-Company-a' question is 'Yes', then what's the information gain after making that split on subtable Company(Big)?
- 4) Are you going to split subtable Company(Small) by another feature or not?
- 5) If the answer to 'Q-Company-d' is 'Yes', then what's the feature that you will split subtable Company(Small) by?
- 6) If the answer to 'Q-Company-d' question is 'Yes', then what's the information gain after making that split on subtable Company(Small)?

Draw the final table here