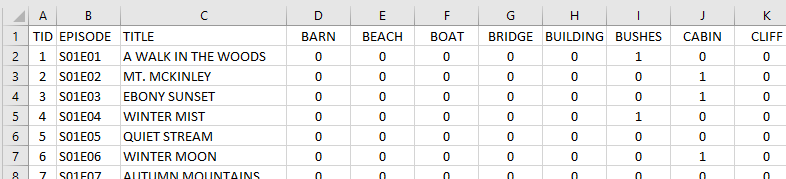
HW: Association Rules

|  |  |
| --- | --- |
| For this homework assignment, we will consider the Bob Ross Paintings dataset. Bob Ross starred in the “Joy of Painting” television show which, according to Bob Ross website, is the most recognized, most watched TV art show in history.  This dataset contains   * Episode number for which painting was created * Title of painting * 37 items commonly used by Bob Ross in his paintings (1: Item exists in painting, 0: Item does not exist). | Image result for bob ross images |

A snip-it of the BobRossPainting dataset.



Complete the following.

1. For this problem, we will consider all length-2 association rules that imply that a will be included in the painting. Length-2 in this case implies one item on left-hand side and on right hand side of the rule.

Find the support, confidence, and lift for each of these rules. Sort this list from largest lift to smallest. (9 pts)

<Paste list of rules with support, confidence, and lift computations here>

1. Which rule has the largest lift? (1 pt)

Rule with largest lift: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Provide the value and a brief explanation as to what each of the following terms measure for the rule identified in Problem #2. (5 pts)

|  |  |  |
| --- | --- | --- |
| Term | Value | Description of its meaning in the context of this problem |
| Support |  |  |
| Confidence |  |  |
| Lift |  |  |

1. Identify some of the items whose lift is substantially lower than 1. What can be said about these items and their association with a CABIN being present in the painting? Discuss. (2 pts)
2. Repeat problem #1, but this time build rules for being present. That is, replace with on the right-hand side of the rule. (6 pts)
   1. Which items are most indicative of TREES being present in the painting?
   2. Which items are least indicative of TREES being present in the painting?
3. Which of the following length-3 association rules are more indicative of a being present in the painting? Use Lift to make your determination. (4 pts)

* Rule #1:
* Rule #2:

1. Complex association rules are often difficult to build, see the rule below. Other than just having messy formulas, why might these rules be difficult to create? Discuss. (3 pts)

* A complex rule: