Step 1:

Question 1: • Entropy of the Parent node

(Cal the entropy of target variable (Risk level) before split. Entropy measure the impurity of the dataset

$$Entropy = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

p_i is the proportion of each class

- $P_{low} = 4/8 = 1/2$
- $P_{high} = 4/8 = 1/2$

$$Entropy(parent) = -(0.5\log_2(0.5) + 0.5\log_2(0.5))$$
$$= +\underline{1}$$

• Step 2: Split on "CreditScore at 650"

+ Left child (≤ 650) ID 2, 4, 6, 8
+ Right child (> 650) ID 3, 1, 5, 7.

• Step 3 Cal Entropy of each child node

• Left child (≤ 650)

+ $P_{low} = 0$
+ $P_{high} = \underline{1}$

$$Entropy(left) = -(0\log_2(0) + 1\log_2(1))$$
$$= 0$$

• Right child (> 650)

$+p_{low} = 1$

$+p_{high} = 0$

Entropy (right) = 0

Step 4 Cal the weighted Avg Entropy after the Split

$$\text{Weighted Entropy} = \left(\frac{\text{Size of left}}{\text{total size}}\right) \times \text{Entropy (left)} + \left(\frac{\text{size of right}}{\text{total size}}\right) \times \text{Entropy (right)}$$

$$= 0$$

Step 5 Cal Information Gain

$$\text{Information Gain} = \text{Entropy (parent)} - \text{weighted Entropy}$$

$$= \underline{1}$$

$\Rightarrow$ I would choose this as the root node because it results in a perfect separation of the ~~data~~ classes, achieving the maximum possible information gain of $\underline{1}$.