Question 2

Step 1 : Cal variance of the Parent node

$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$$\text{Variane (parent)} = \frac{1}{8}\sum(x_i - 660)^2$$

$$= 4200$$

Step 2 Cal Variance of each child node

- Left Child ($\leq 35$) : ID 1, 2, 4, 6, 8

$$\text{Variane (left)} = \frac{1}{5}\sum(x_i - 648)^2$$

$$= 1576$$

- Right child ($> 35$) : ID 3, 5, 7

$$\text{Variance (right)} = \frac{1}{3}\sum(x_i - 746.67)^2$$

$$\approx 822.22$$

Step 3 Cal Weighted Var after Split:

$$\text{Weighted Variance} = \frac{5}{8} \times 1576 + \frac{3}{8} \times 822.22$$

$$= 1293.33$$

Step 4 Cal Var Reduction

$$\text{Var Reduction} = \text{Var(parent)} - \text{Var}_{\text{weighted}}$$

$= 2906.67$

- How splitting criterion differs from Information gain in classification tree:
  + The key difference lie in the type of target variable (continuos vs categorical), the metric used (variance vs entropy), and the goal of the split.