Name: Md Shakil Hossain
Id: 213002042

Solve (b):

- 13 instances with 5 attributes: Age, Graduated, Income, Gender, Smokes
- Target variable: Smokes (Yes/No)
- 6 'Yes', 7 'No'

Gini Index of Entire Dataset

Total: 13 samples

Smokes	Count
Yes	6
No	7

- Gini(S) = 1 – $(P_{yes}^2 + P_{no}^2)$
- Gini(S) = 1 – $((\frac{6}{13})^2 + (\frac{7}{13})^2) \approx 0.497$

Gini Index for each Attributes

A) **Graduated**

Graduated	Count	Yes	No
Yes	6	2	4
No	7	4	3

Gini(Yes) = $1 - \left(\left(\frac{1}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right) = 0.444$

Gini(No) = $1 - \left(\left(\frac{4}{7}\right)^2 + \left(\frac{3}{7}\right)^2\right) = 0.490$

Weighted Gini= $1 - \left(\left(\frac{6}{13}\right)^2 * 0.444 + \left(\frac{7}{13}\right)^2 * 0.490\right) = 0.468$

Gini Index for each Attributes

B) **Gender**

Gender	Count	Yes	No
Yes	7	4	3
No	6	2	4

Gini(Yes) = $1 - \left(\left(\frac{4}{7}\right)^2 + \left(\frac{3}{7}\right)^2\right) = 0.490$

Gini(No) = $1 - \left(\left(\frac{2}{6}\right)^2 + \left(\frac{4}{6}\right)^2\right) = 0.444$

Weighted Gini= $1 - \left(\left(\frac{7}{13}\right)^2 * 0.490 + \left(\frac{6}{13}\right)^2 * 0.444\right) = 0.468$

Gini Index for each Attributes

C) Age (splitting on **Age < 30)**

Age < 30?	Count	Yes	No
Yes	6	4	2
No	7	2	5

Gini(Yes) = $1 - \left(\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right) = 0.444$

Gini(No) = $1 - \left(\left(\frac{2}{7}\right)^2 + \left(\frac{5}{7}\right)^2\right) = 0.408$

Weighted Gini= $1 - \left(\left(\frac{6}{13}\right)^2 * 0.444 + \left(\frac{7}{13}\right)^2 * 0.408\right) \approx 0.425$

Gini Index for each Attributes

D) **Income (split at 50K)**

Income <50K?	Count	Yes	No
Yes	6	4	2
No	7	2	5

Gini(Yes) = $1 - \left(\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right) = 0.444$

Gini(No) = $1 - \left(\left(\frac{2}{7}\right)^2 + \left(\frac{5}{7}\right)^2\right) = 0.408$

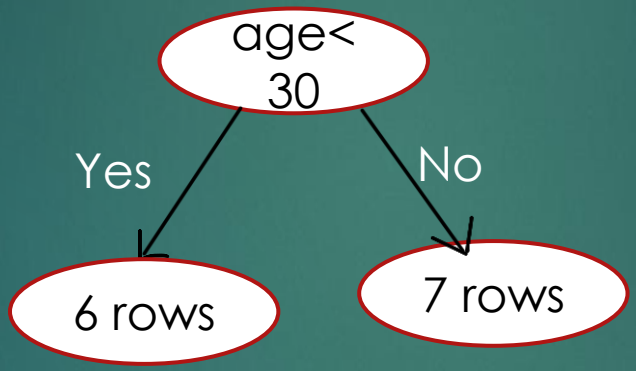
Weighted Gini= $1 - \left(\left(\frac{6}{13}\right)^2 * 0.444 + \left(\frac{7}{13}\right)^2 * 0.408\right) \approx 0.425$

Best Attribute to Split On

- Income < 50K → Gini = **0.425**
- Age < 30 → Gini = **0.425**
- Gender → 0.468
- Graduated → 0.468

Root Node: Age < 30

- Best Gini = 0.425 (tie between Age and Income)
- Tie broken by attribute order → Chose: Age < 30



- Left Subtree (Age < 30): 6 samples (4 Yes, 2 No)
- Best split: Income < 25K
- Right Subtree (Age ≥ 30): 7 samples (2 Yes, 5 No)

Splitting Age < 30

- Left Subtree (Age < 30): 6 samples (4 Yes, 2 No)
- Best split: Income < 25K
- Right Subtree (Age ≥ 30): 7 samples (2 Yes, 5 No)

Final Tree Structure