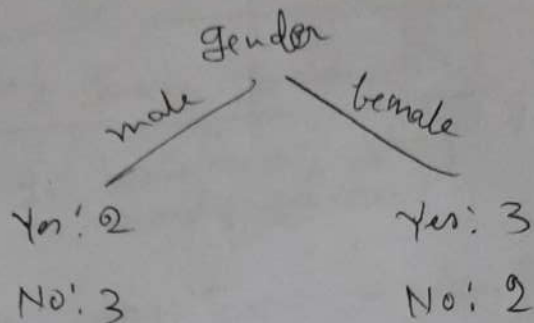


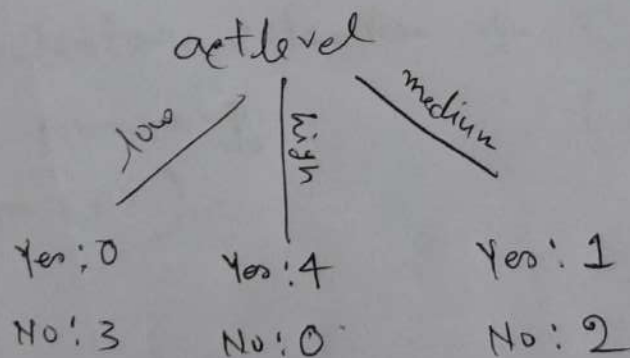
## calculating Gini Impurity of Gender

G<sub>I</sub>(Gender)



$$\begin{aligned}
 G_I(\text{gender}) &= \left[ 1 - \left( \frac{2}{2+3} \right)^2 - \left( \frac{3}{3+2} \right)^2 \right] \times \frac{5}{10} \\
 &+ \left[ 1 - \left( \frac{3}{3+2} \right)^2 - \left( \frac{2}{3+2} \right)^2 \right] \times \frac{5}{10} = 0.29 + 0.29 \\
 &= 0.48
 \end{aligned}$$

G<sub>I</sub>(activity)



$$\begin{aligned}
 G_I(\text{activity}) &= \left[ 1 - \left( \frac{0}{3} \right)^2 - \left( \frac{3}{3} \right)^2 \right] \times \frac{3}{10} \\
 &+ \left[ 1 - \left( \frac{4}{4} \right)^2 - \left( \frac{0}{4} \right)^2 \right] \times \frac{4}{10} \\
 &+ \left[ 1 - \left( \frac{1}{1+2} \right)^2 - \left( \frac{2}{1+2} \right)^2 \right] \times \frac{3}{10} \\
 &= 0 + 0 + 0.13 = 0.13
 \end{aligned}$$

GrI (monthly Inc)

sent monthly income and get midpoints

rent(monthlyInc) = [22, 25, 30, 35, 40, 48, 55, 60, 65, 70]

Income < 23.5	
<u>True</u>	<u>False</u>
Yes: 0	Yes: 5
No: 1	No: 4

$$\begin{aligned} G I(\text{inc} < 23.5) &= \\ &[1 - (\frac{0}{1})^2 - (\frac{1}{1})^2] \times \frac{1}{10} \\ &+ [1 - (\frac{5}{9})^2 - (\frac{4}{9})^2] \times \frac{9}{10} \\ &= 0.44 \end{aligned}$$

Income  $< 27.5$

True False

Yes: 0 Yes: 5

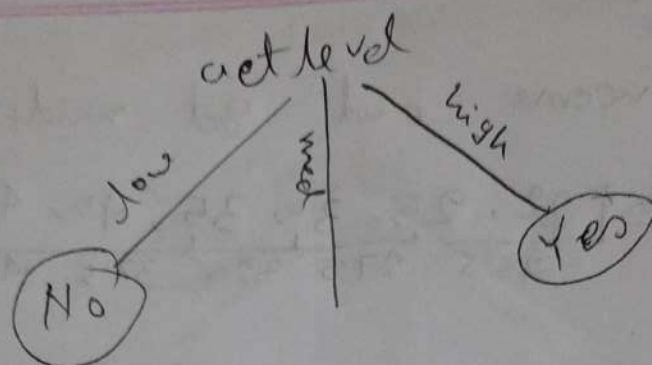
No: 2 No: 3

$$\begin{aligned} & \{ \text{GI}(\text{income} < 27.5) = \\ & [1 - 0 - 1] \times \frac{2}{10} \\ & + [1 - \left(\frac{5}{8}\right)^2 - \left(\frac{3}{8}\right)^2] \times \frac{8}{10} \\ & = 0.375 \end{aligned}$$

Income < 32.5

True	False
Yes: 0	Yes: 5
No: 3	No: 2

$$GI(\text{income} < 32.5)$$
$$[1 - 0 - 1] \times \frac{3}{10} +$$
$$[1 - (\frac{5}{7})^2 - (\frac{2}{7})^2] \times \frac{7}{10}$$
$$= 0.285$$

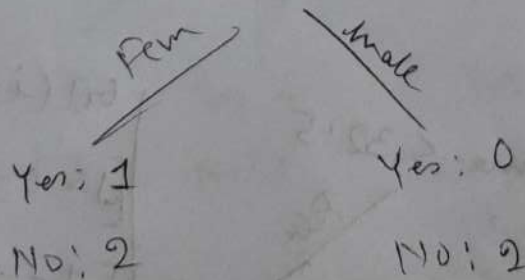


New set

<u>ID</u>	<u>Gender</u>	<u>Activ</u>	<u>MonInc</u>	<u>buy</u>
3	fem	med	48	Yes
4	male	med	35	No
8	male	med	40	No

G<sub>I</sub>(Gender)

Gender



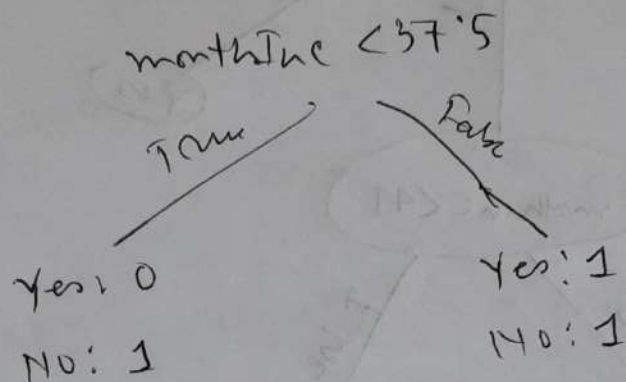
$$G_I(\text{Gender}) = \left[ 1 - \left( \frac{1}{3} \right)^2 - \left( \frac{2}{3} \right)^2 \right] \times \frac{3}{5}$$

$$+ 0 = 0.267$$

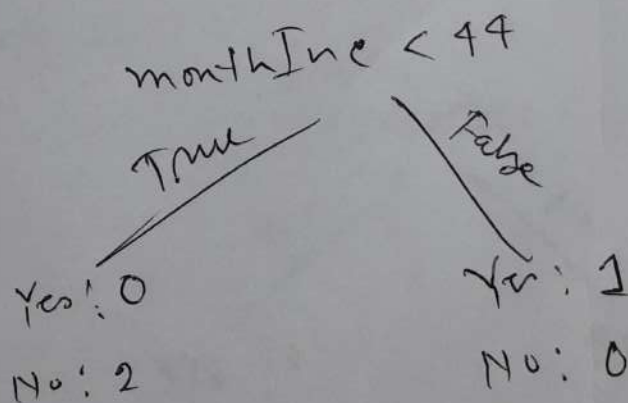


G<sub>I</sub>(monthInc)

sorted =  $\frac{35, 40, 48}{37.5 \quad 44}$

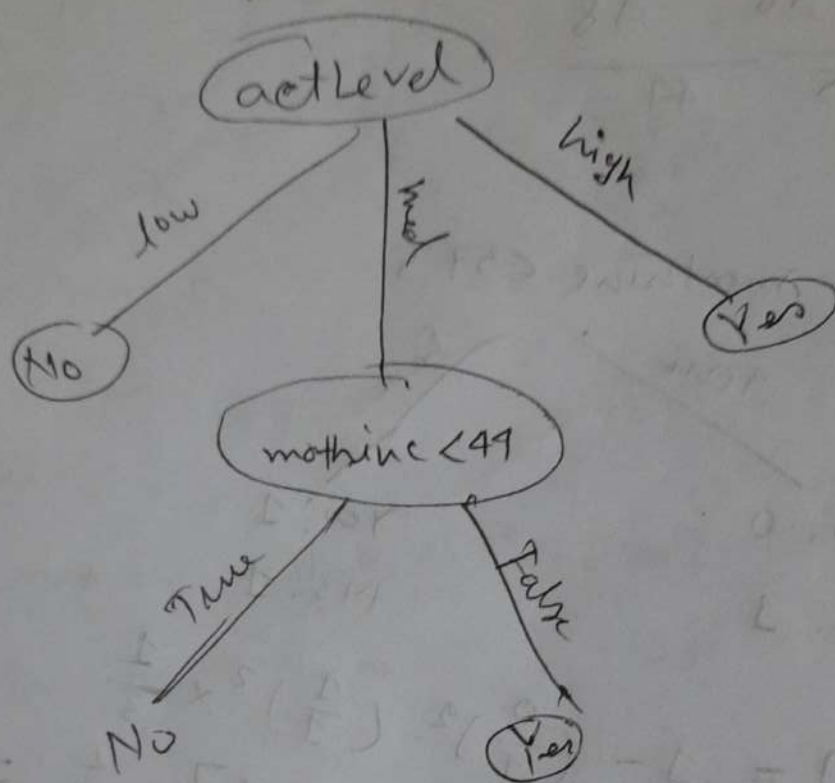


$$G_I(m < 37.5) = 1 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2 \times \frac{1}{3} + \left[1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right] \times \frac{2}{3} = 0.33$$

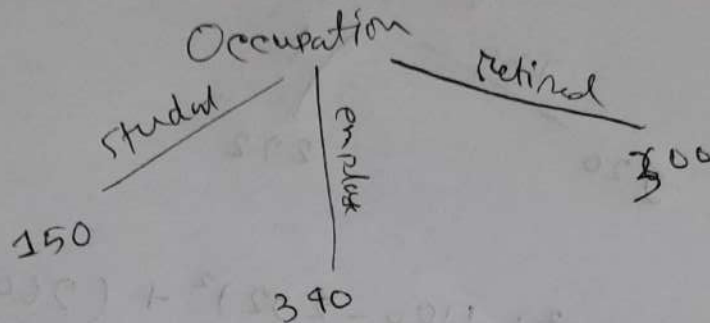


$$G_I(m < 44) = 0$$

Final Tree

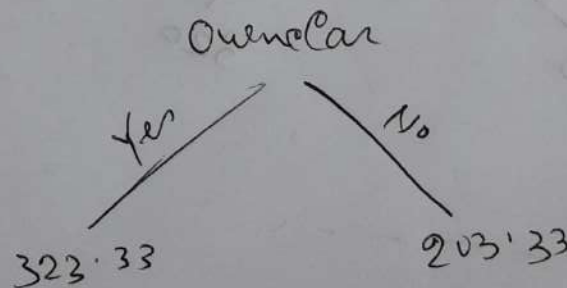


SSR (Occupation)



$$\begin{aligned} SSR(\text{Occupation}) &= (120-150)^2 + (180-150)^2 + (260-340)^2 \\ &\quad + (420-340)^2 + (230-300)^2 + (370-300)^2 \\ &= 8200 + 16200 = 24400 \end{aligned}$$

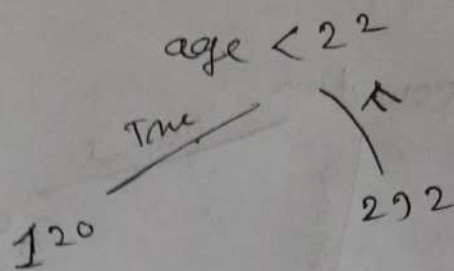
SSR (OwnCar)



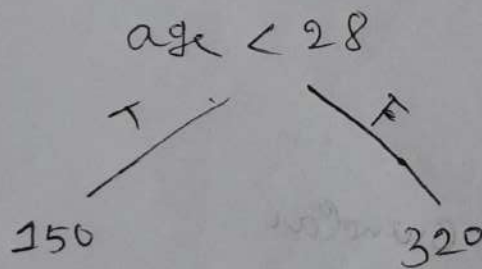
$$\begin{aligned} SSR(\text{OwnCar}) &= (120-203.33)^2 + (180-323.33)^2 \\ &\quad + (260-203.33)^2 + (420-323.33)^2 + (230-203.33)^2 \\ &\quad + (370-323.33)^2 = 42964.24 \end{aligned}$$



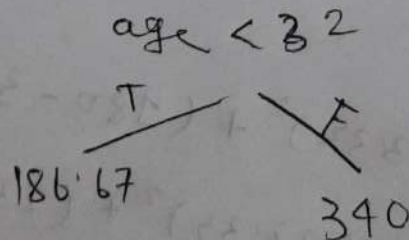
SSR(age)



$$\begin{aligned}
 SSR(\text{age}) &= (120 - 120)^2 + (180 - 292)^2 + (260 - 292)^2 \\
 &+ (420 - 292)^2 + (230 - 292)^2 + (370 - 292)^2 \\
 &= \dots
 \end{aligned}$$



SSR(age < 28)



$$\begin{aligned}
 SSR(\text{age} < 28) &= (120 - 186.67)^2 + (180 - 186.67)^2 + (260 - 186.67)^2 \\
 &+ (420 - 340)^2 + (230 - 340)^2 + (370 - 340)^2 \\
 &= 9866.67 + 19400 = 29266.67
 \end{aligned}$$