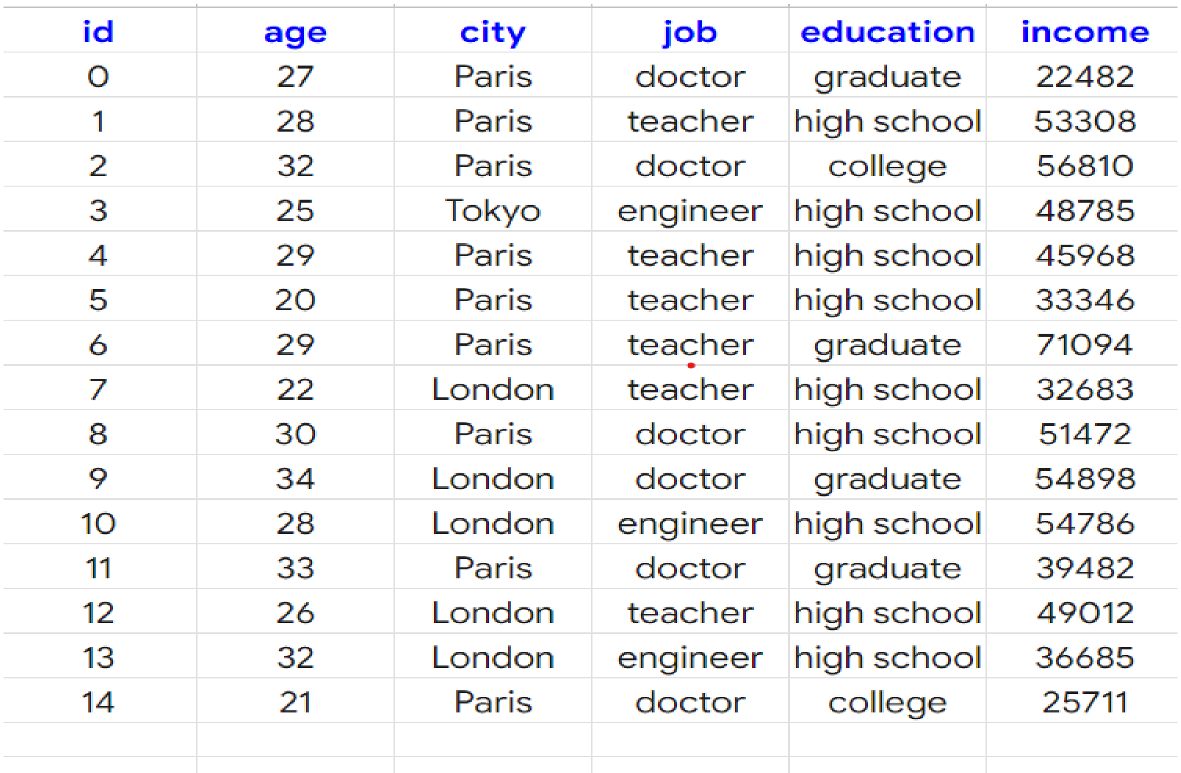
1. **Create a Decision Tree of this dataset.**

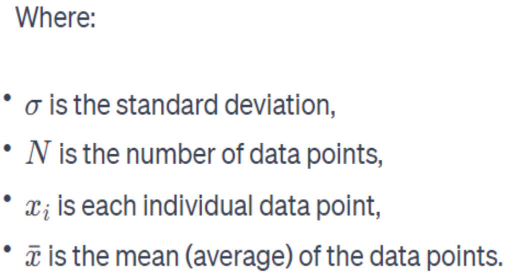


id	age	city	job	education	income
0	27	Paris	doctor	graduate	22482
1	28	Paris	teacher	high school	53308
2	32	Paris	doctor	college	56810
3	25	Tokyo	engineer	high school	48785
4	29	Paris	teacher	high school	45968
5	20	Paris	teacher	high school	33346
6	29	Paris	teacher	graduate	71094
7	22	London	teacher	high school	32683
8	30	Paris	doctor	high school	51472
9	34	London	doctor	graduate	54898
10	28	London	engineer	high school	54786
11	33	Paris	doctor	graduate	39482
12	26	London	teacher	high school	49012
13	32	London	engineer	high school	36685
14	21	Paris	doctor	college	25711

<u>Step 1</u>:

standard deviation of target : -

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



Where:

- σ is the standard deviation,
- N is the number of data points,
- x_i is each individual data point,
- $\bar{x}$ is the mean (average) of the data points.

$$\text{Mean} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Mean=

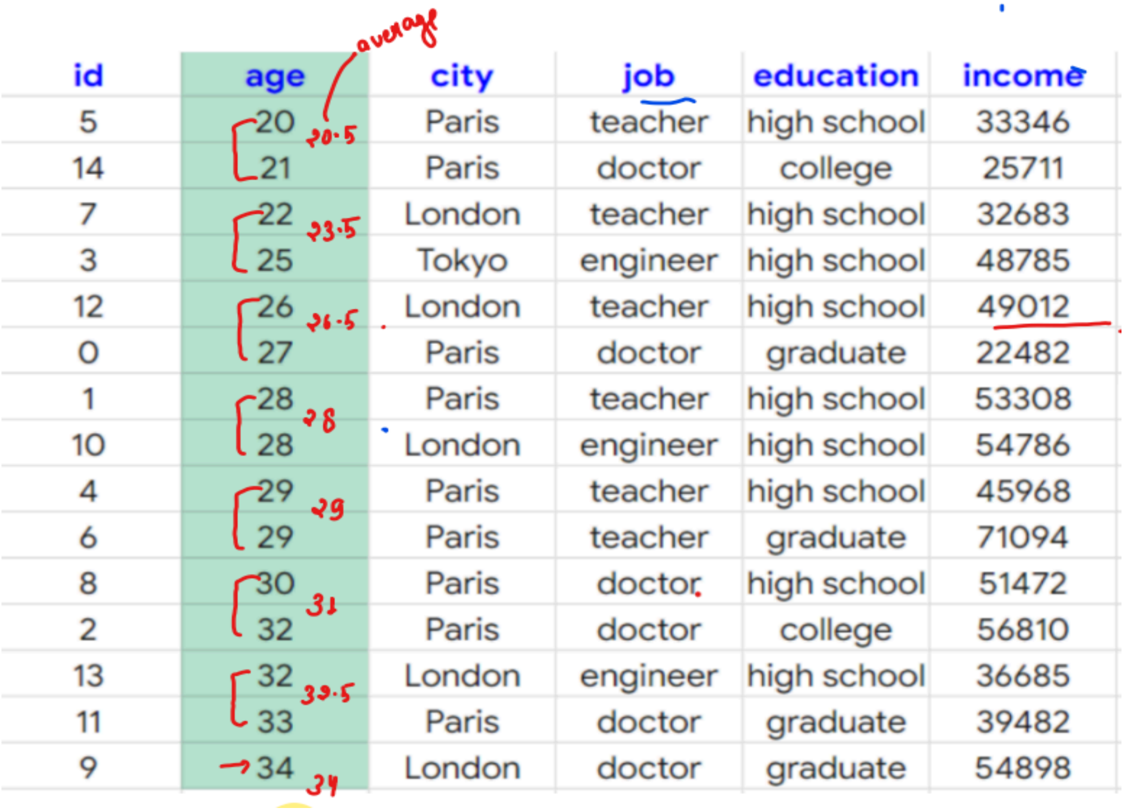
$$(33346+25711+32683+48785+49012+22482+53308+54786+45968+71094+51472+56810+36685+39482+54898)/15$$

$$\text{Mean} \approx 41919.33$$

standard deviation = 12708.16363

Step 2 :

Now find the weighted standard deviation of each attribute . Age is a continuous feature so firstly we have to make it discreet. Sort table on the basis of Age.



id	age	city	job	education	income
5	20	Paris	teacher	high school	33346
14	21	Paris	doctor	college	25711
7	22	London	teacher	high school	32683
3	25	Tokyo	engineer	high school	48785
12	26	London	teacher	high school	49012
0	27	Paris	doctor	graduate	22482
1	28	Paris	teacher	high school	53308
10	28	London	engineer	high school	54786
4	29	Paris	teacher	high school	45968
6	29	Paris	teacher	graduate	71094
8	30	Paris	doctor	high school	51472
2	32	Paris	doctor	college	56810
13	32	London	engineer	high school	36685
11	33	Paris	doctor	graduate	39482
9	34	London	doctor	graduate	54898

Age (average)	Standard deviation	weighted SD
20.5	$\leq$ 0 $>$ 13227.01071	$= 0 + \frac{15}{15}(13227.01071)$ $= 13227.01$
23.5	$\leq$ 3454.46 $>$ 8011.40	7100.012
26.5	$\leq$ 4451.98 $>$ 6950.04	6117.35 lowest
28	$\leq$ 14859.64 $>$ 17079.18	15895.34
29	$\leq$ 12254.34 $>$ 5941.99	10180.223
31	$\leq$ 10490.94 $>$ 18849.27	12719.31
32.5	$\leq$ 10309.23 $>$ 1131.73	9085.56
34	$\leq$ 10490.98 $>$ 0	10490.98

weighted standard deviation

$$S(23.5) = \frac{3}{15}(3454.46) + \frac{12}{15}(8011.40)$$

$$= \frac{1}{5}(3454.46) + \frac{4}{5}(8011.40)$$

$$= 690.892 + 6409.12$$

$$= 7100.012$$

$$S(26.5) = \frac{5}{15}(4451.98) + \frac{10}{15}(6950.04)$$

$$= \frac{4451.98}{3} + \frac{2}{3}(6950.04)$$

$$= 1483.99 + 4633.36 = 6117.35$$

(26.5) — best split for Age

PAGE No
DATE

weight SD(Age) = 6117.35

Target SD = 12708.16363

gain = 12708.16363 − 6117.35

= 6590.81363

Next feature (City)

City:

City

City	no. of rows	S.D.	weighted SD	gain
Paris	9	7713.33		
london	5	11216.08	8366.66	4341.50
Tokyo	1	0		

SD(London) = 11216.0

SD(Paris) = 7713.33

$$= \frac{9}{15}\left(7713.33\right) + \frac{5}{15}\left(11216.08\right) + \frac{1}{15}(0)$$

= 8366.664

gain = 12708.16363 − 8366.66

= 4341.50

Next feature Job:-

no. of rows	Job	SD	weighted SD	gain
6	Teacher	8699.96		
6	Doctor	12528.36	9609.81	6590.81
3	Engineering	5590.91		

weighted SD = $\frac{6}{15}(8699.96) + \frac{6}{15}(12528.36) + \frac{3}{15}(55909$

$= \frac{2}{5}(8699.96) + \frac{2}{5}(12528.36) + \frac{1}{5}(5590.91)$

$= \underline{9609.51}$

Gain = Target SD − 9609.51

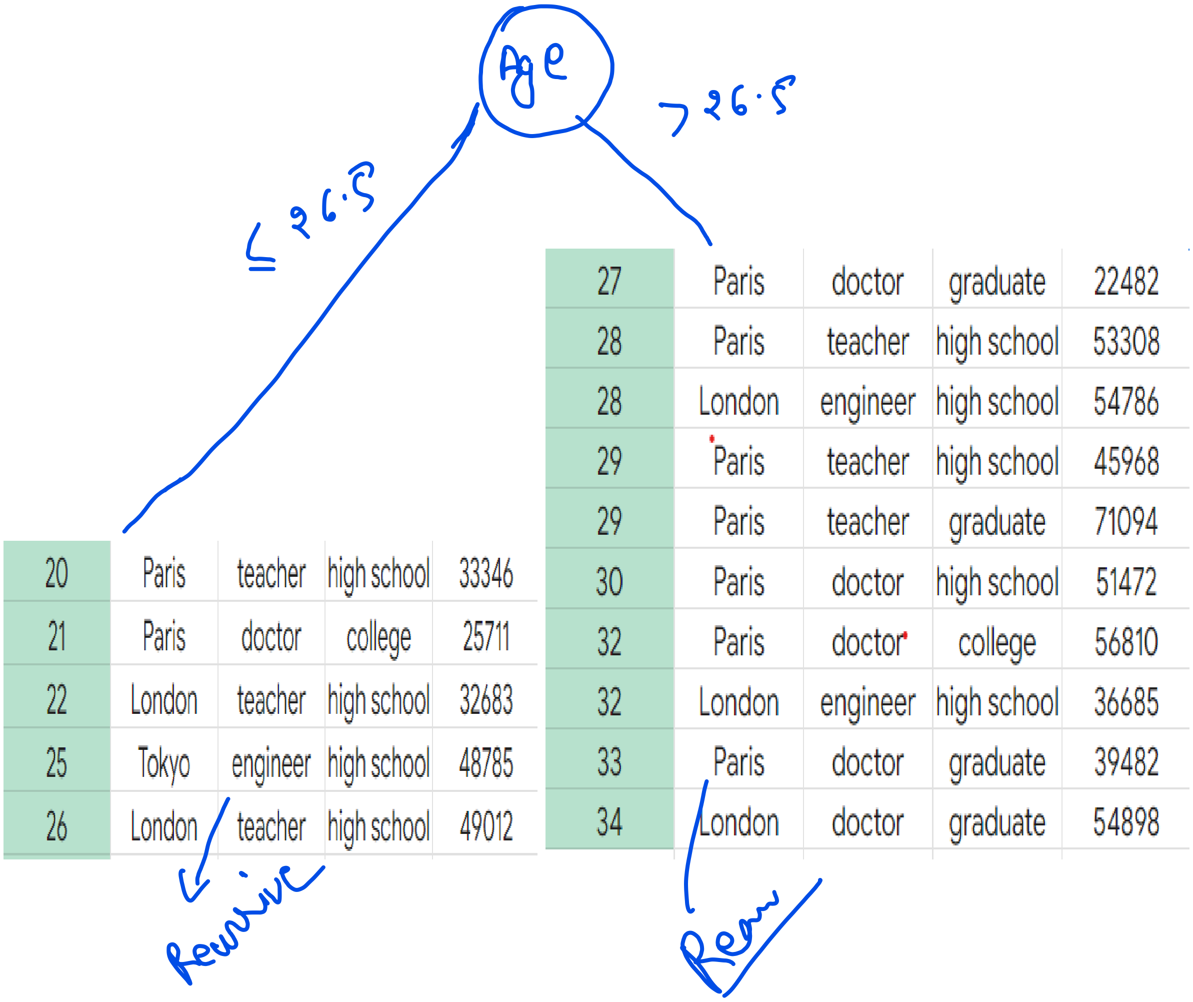
= 6590.81

Next feature :- Education

no. of rows	Education	S.D.	weighted SD	gain
9	High school	13979.03		
2	college	19341.86	15109.26	−2401.09
4	graduate	15535.99		

$= \frac{9}{15}(13979.03) + \frac{2}{15}(19341.86) + \frac{4}{15}(15535.99)$

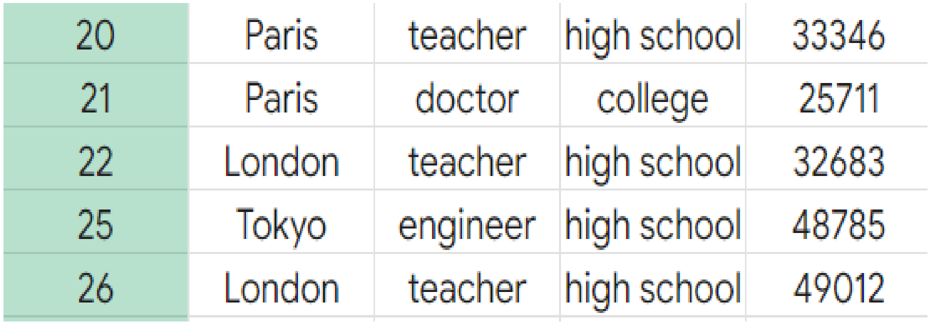
Now highest gain is of Age , so Root node should be Age.



20	Paris	teacher	high school	33346
21	Paris	doctor	college	25711
22	London	teacher	high school	32683
25	Tokyo	engineer	high school	48785
26	London	teacher	high school	49012

27	Paris	doctor	graduate	22482
28	Paris	teacher	high school	53308
28	London	engineer	high school	54786
29	Paris	teacher	high school	45968
29	Paris	teacher	graduate	71094
30	Paris	doctor	high school	51472
32	Paris	doctor	college	56810
32	London	engineer	high school	36685
33	Paris	doctor	graduate	39482
34	London	doctor	graduate	54898

Level 2:



20	Paris	teacher	high school	33346
21	Paris	doctor	college	25711
22	London	teacher	high school	32683
25	Tokyo	engineer	high school	48785
26	London	teacher	high school	49012