* Draw a decision tree diagram to predict number of rows to play based on weather conditions like outlook, temperature, humidity, windy, consider databas shown below

Outlook	Temperature	Humidity	Windy	Hours to Play
Rainy	Hot	High	false	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	46
sunny	Mild	High	False	45
Sunny	cool	Normal	False	52
Overcast	cool	Normal	True	43
Rainy	Mild	High	False	35
Rainy	cool	Normal	False	38
Sunny	mild	Normal	False	46
Rainy	mild	Normal	True	48
Overcast	mild	High	Tyue	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30
Sunny	Cool	Normal	True	23

Termination criteria: CV <= 10% of minimum number of Sample

Calculating mean, standard deviation (SD), co-efficient of variation(v)

$$\text{mean} = \frac{\Sigma x}{n} = \frac{557}{14} = 39.78$$

$$SD = \sqrt{\frac{\Sigma(x-\text{mean})^2}{n}} = 9.67$$

$$CV = \frac{SD}{\text{mean}} \times 100 = \frac{9.67}{39.78} \times 100 = 24.80$$

Now, dataset is split into different attributes, The SD of each branch is calculated.

$$SD(\text{altr}) = \Sigma W(\text{branch}) . SD(\text{branch})$$

and the result SDR (standard deviation reduction) is calculated, SDR = SD - SD(attr)

$$\boxed{\therefore SD = 9.67}$$

Outlook:-

Outlook	mean	SD	CV	n	W(v)
Rainy	35.2	8.7	24.7	5	5/14
Overcast	46.25	4.03	8.72	4	4/14
Sunny	39.2	12.2	81.0	5	5/14

$$\therefore SD(\text{Outlook}) = \frac{5}{14} * 8.7 + \frac{4}{14} * 4.03 + \frac{5}{15} * 12.2 = 8.59$$

$$SDR(\text{Outlook}) = SD - SD(\text{Outlook}) = 9.67 - 8.59 = 1.08$$

Temperature:-

Temperature	mean	SD	CV	n	W(v)
Hot	36.25	10.34	30.6	4	4/14
cool	39	12.14	31.1	4	4/14
mild	42.6	8.38	19.65	6	6/14

$\therefore SD(temperature) = \frac{4}{14} * 10.34 + \frac{4}{14} * 12.14 + (\frac{6}{14}) * 8.38 = 10.01$

$SDR(temperature) = SD - SD(temperature) = 9.67 - 10.01 = -0.34$

Humidity:-

Humidity	mean	SD	CV	n	W(HV)
High	37.5	10.11	26.92	7	7/14
Normal	4.2	9.4	22.4	7	7/14

$\therefore SD(humidity) = \frac{7}{14} \times 10.11 + \frac{7}{14} \times 9.4 = 9.77$

$SDR(humidity) = SD - SD(humidity)$

$= 9.67 - 9.77 = -0.1$

Windy:-

Windy	mean	S.D	CV	n	W(V
True	37.6	11.6	30.8	6	6/14
False	41.3	8.41	20.3	8	8/14

$\therefore SD(windy) = \frac{6}{14} * 11.6 + \frac{8}{14} * 8.41 = 9.77$

$\therefore SDR(windy) = SD - SD(windy) = 9.67 - 9.77 = (-0.1)$

SDR(outlook) = 1.08

SDR(windy) = -0.1

SDR(Temperature) = -0.34

SDR(humidity) = -0.1

The value that has highest SDR is consider as root node (i.e decision node)

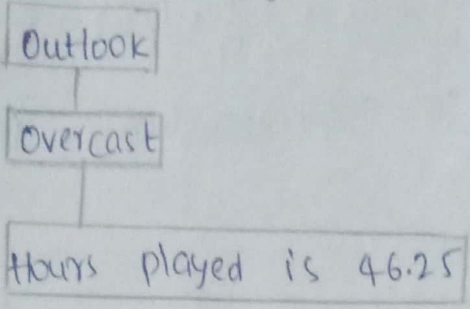
Considering termination criteria

CV is 10% or CV is (n ≤ 4)

Outlook

Over cast has CV of 8%, which is less than threshold.

value therefore, we need not go for further spliting.



We need to split sunny and rainy columns

Outlook	Temperature	Humidity	Windy	Hours played
Sunny	mild	High	False	45
Sunny	cool	Normal	False	52
Sunny	cool	Normal	True	23
Sunny	mild	Normal	False	46
Sunny	mild	High	True	30

∴ mean = 39.2; SD = 12.2; CV = 31.0

Temperature:-

Temperature	mean	SD	CV	n	w(v)
Mild	40.3	8.96	22.23	3	3/5
Cold	37.5	20.50	54.66	2	2/5

SD(temperature) = $\frac{3}{5}*8.96 + \frac{3}{5}*20.5 = 13.576$

SDR (temperature) = SD - SD(temperature) = 12.2 - 13.576 = -1.37

Humidity:-

Humidity	mean	SD	CV	n	w(v)
High	37.5	10.6	28.26	2	2/5
Normal	40.3	15.30	37.96	3	3/5

SD(humidity) = $\frac{2}{5} * 10.6 + 3/5 * 15.30 = 13.44$

SDR(humidity) = SD - SD(humidity) = 12.2 - 13.42 = -1.22

Windy:-

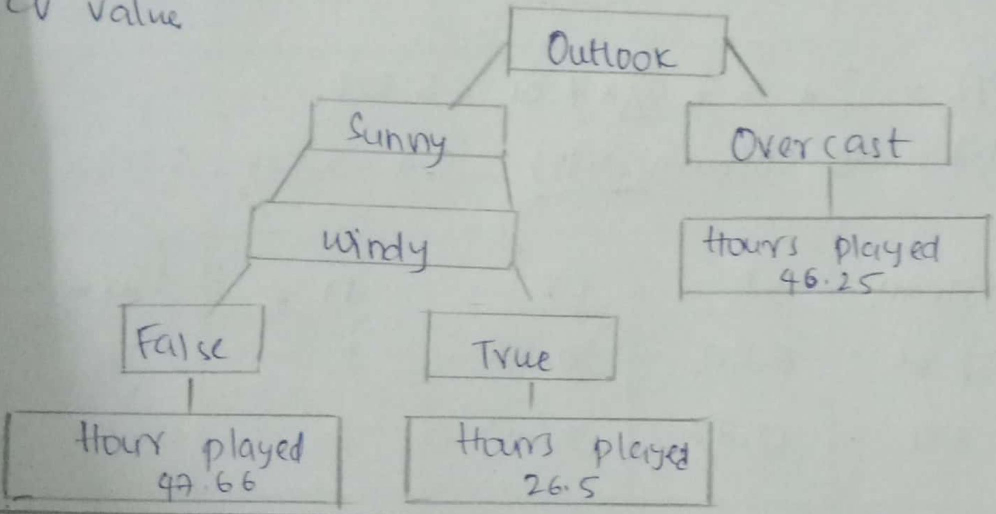
Windy	mean	SD	CV	n	w(v)
False	47.66	3.78	7.94	3	3/5
True	26.5	4.94	18.65	2	2/5

$SD(windy) = \frac{3}{5} * 3.78 + \frac{2}{5} * 4.94 = 4.23$

$SDR(windy) = SD - SD(windy = 12.2 - 4.23 = 7.97$

In outlook, among temperature, humidity and windy SDR value is high for windy SDR = 7.97

Then, check for CV value both true and false satisfy the CV value



Rainy:-

Outlook	Temperature	Humidity	Windy	Hours played
Rainy	hot	High	False	25
Rainy	hot	High	True	30
Rainy	mild	High	False	35
Rainy	cool	Normal	False	38
Rainy	mild	Normal	True	48

mean = 35.2, SD = 8.7, CV = 24.7

Temperature:-

Temperature	mean	SD	CV	n	w(v)
Hot	27.5	3.53	12.83	2	2/5
mild	41.5	9.19	22.144	2	2/5
cool	38	0	0	1	1/5

SD(Temp) = $2/5 * 3.53 + 2/5 * 9.19 + \frac{1}{5} * 0 = 5.088$

SDR(temperature) = $8.7 - 5.088 = 3.612$

Humidity:-

Humidity	mean	SD	CV	n	w(v)
High	30	5	16.66	3	3/5
Normal	43	7.07	16.44	2	2/5

SD(humidity) = $\frac{3}{5} * 5 + 2/5 * 7.07 = 5.828$

SDR (humity) = SD - SD(humidity) = $8.7 - 5.828 = 2.872$

Windy:-

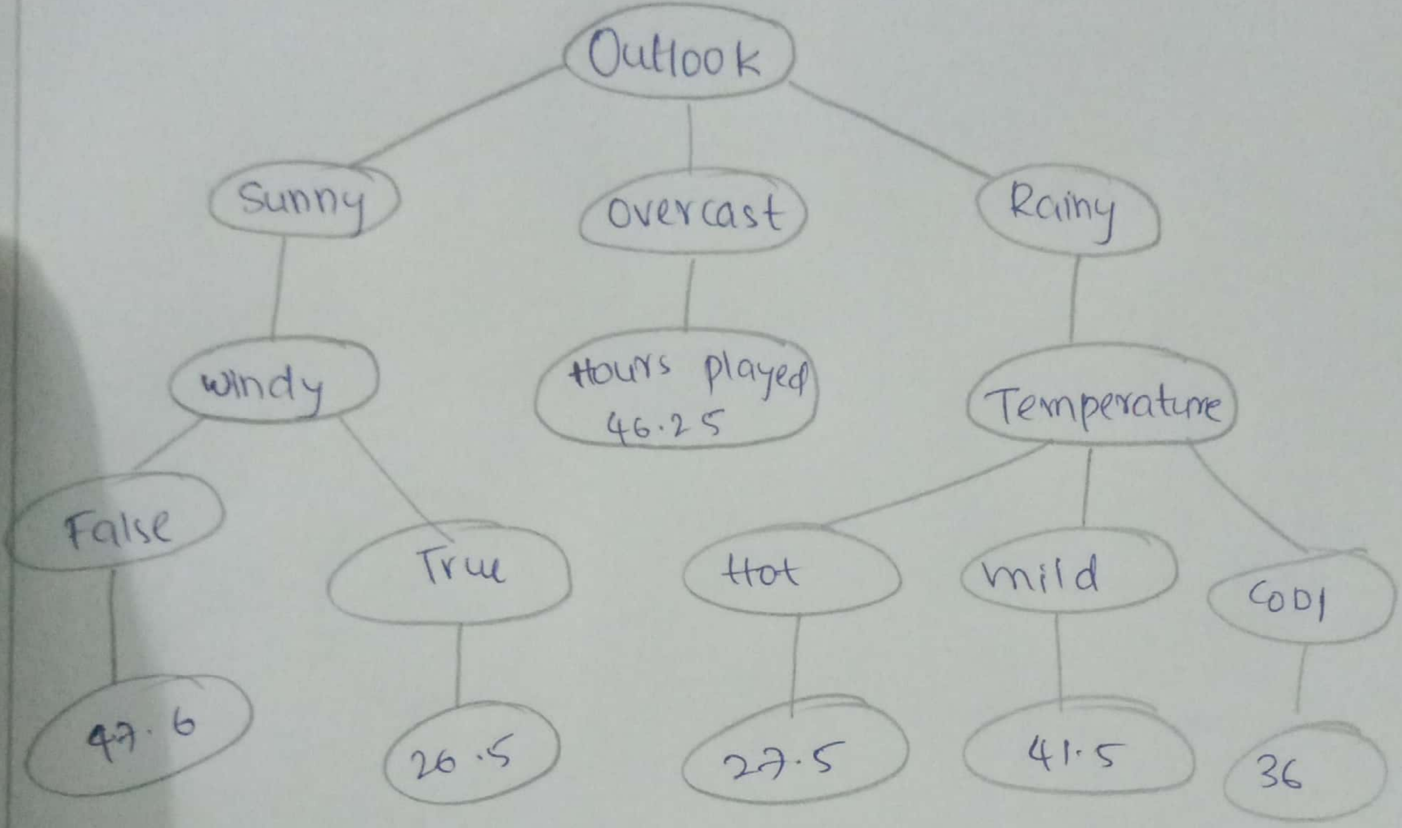
Windy	mean	SD	CV	n	w(v)
False	32.66	6.80	20.85	3	3/5
True	39	12.72	32.5	2	2/5

SD(windy) = $3/5 * 6.80 + \frac{2}{5} * 12.72 = 9.168$

SDR(windy) = $8.7 - 9.168 = -0.468$

The SDR value is high for temperature among Temperature, humidity & windy. Then check for CV value of hot, mild and cold satisfy the CV value

* Design tree diagram to predict number of hours to play based on weather conditions.


Outlook
Sunny
Overcast
Rainy
windy
Hours played 46.25
Temperature
False
True
Hot
mild
cool
47.6
26.5
27.5
41.5
36