Assignment 5.1
Student name: Chenxu Wang
ID: 10457625
Purpose: Construct a classification and regression tree to classify salary based on the other variables only one split level.

TABLE 1 ：Candidate splits for *t* = root node

Candidate Split	Left Child Node, t(L)	Right Child Node, t(R)
1	Occupation=Service	Occupation∈{Management, Sales, Staff}
2	Occupation=Management	Occupation∈{Service, Sales, Staff}
3	Occupation=Sales	Occupation∈{Service, Management, Staff}
4	Occupation=Staff	Occupation∈{Service, Management, Sales}
5	Gender=Male	Gender=Female
6	Age⩽30	Age⩾31
7	Age⩽40	Age⩾41

TABLE 2 : Values of the components of the optimality measure Φ(*s*/*t*) for each candidate split, for the root node

Split	PL	PR	Level	P(j | tL)	P(j | tR)	2PL*PR	Q(s | t)	Φ(s | t)
1	0.273	0.727	L1	0.333	0.125	0.397	0.583	0.231
			L2	0.333	0.25			
			L3	0.333	0.375			
			L4	0	0.25			
2	0.364	0.636	L1	0	0.286	0.463	1.429	0.662
			L2	0	0.429			
			L3	0.5	0.286			
			L4	0.5	0			
3	0.18	0.82	L1	0	0.22	0.295	0.89	0.263
			L2	0.5	0.22			
			L3	0.5	0.33			
			L4	0	0.22			
4	0.18	0.82	L1	0.5	0.11	0.295	1.33	0.392
			L2	0.5	0.22			
			L3	0	0.44			
			L4	0	0.22			
5	0.545	0.455	L1	0.33	0	0.496	0.93	0.461
			L2	0.33	0.2			
			L3	0.33	0.4			
			L4	0	0.4			
6	0.455	0.545	L1	0.4	0	0.496	0.933	0.463
			L2	0.2	0.333			
			L3	0.4	0.333			
			L4	0	0.333			
7	0.727	0.273	L1	0.25	0	0.397	0.58	0.23
			L2	0.25	0.33			
			L3	0.375	0.33			
			L4	0.125	0.33			

The maximum observed value for $\Phi(s|t)$ among the candidate splits is therefore attained by split 2, with $\Phi(s|t) = 0.662$. CART therefore chooses to make the initial partition of the data set using candidate split 2, Occupation=service, occupation $\in$ {Management, Sales, Staff}