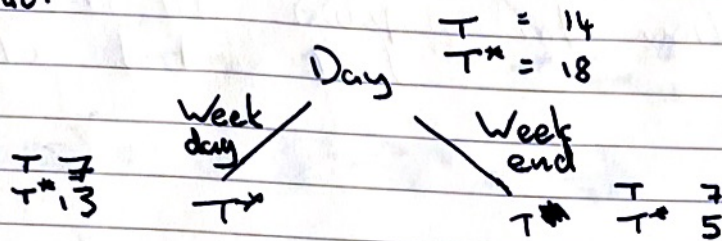


Required Activity 9.3

1) $G(N) = 1 - \frac{1}{2} \left(\frac{14}{32} \right)^2 - \left(\frac{18}{32} \right)^2 = \text{answer } \frac{63}{128} \approx 0.4922$

2) Day Split:



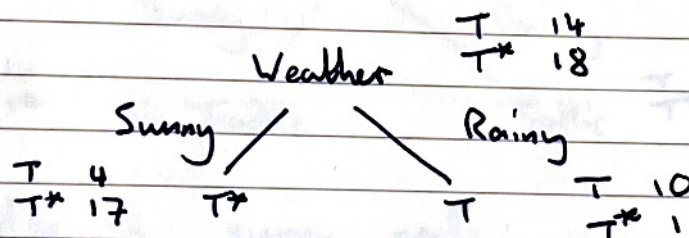
Let $T = F_q$ of Traffic
 $T^* = F_q$ of no Traffic

$$G(\text{Week day}) = 1 - \frac{7}{20}^2 - \frac{13}{20}^2 = \frac{91}{200}$$

$$G(\text{Week end}) = 1 - \frac{7}{12}^2 - \frac{5}{12}^2 = \frac{35}{72}$$

$$\Rightarrow \Delta G = \frac{63}{128} - \frac{20}{32} \cdot \frac{91}{200} - \frac{12}{32} \cdot \frac{35}{72} = \frac{49}{1920} \approx 0.02552$$

Weather Split:

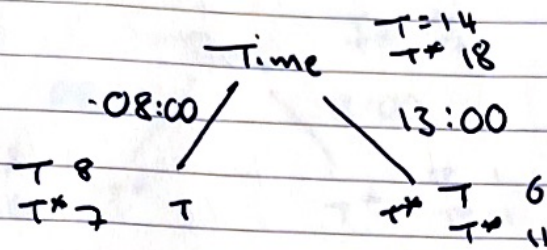


$$G(\text{Sunny}) = 1 - \frac{4}{21}^2 - \frac{17}{21}^2 = \frac{136}{441}$$

$$G(\text{Rainy}) = 1 - \frac{10}{11}^2 - \frac{1}{11}^2 = \frac{20}{121}$$

$$\Rightarrow \Delta G = \frac{63}{128} - \frac{21}{32} \cdot \frac{136}{441} - \frac{21}{32} \cdot \frac{20}{121} = \frac{6889}{24968} \approx 0.2758$$

Time Split



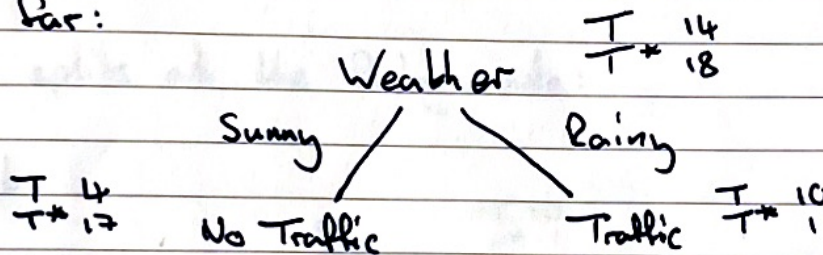
$$G(08:00) = 1 - \frac{8}{15}^2 - \frac{7}{15}^2 = \frac{112}{225}$$

$$G(13:00) = 1 - \frac{6}{17}^2 - \frac{11}{17}^2 = \frac{132}{289}$$

$$\Rightarrow \Delta G = \frac{63}{128} - \frac{15}{32} \cdot \frac{112}{225} - \frac{17}{32} \cdot \frac{132}{289} = \frac{579}{32640} \approx 0.01621$$

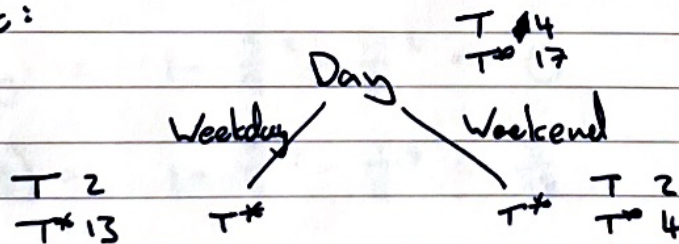
Best split = Weather split with an information gain of 0.233

Tree so far:



34) Evaluate splits at sunny node:

Day split:

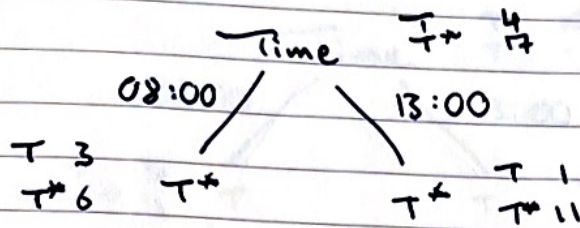


$$G(\text{Weekday}) = 1 - \frac{2}{15}^2 - \frac{13}{15}^2 = \frac{52}{225}$$

$$G(\text{Weekend}) = 1 - \frac{2}{6}^2 - \frac{4}{6}^2 = \frac{4}{9}$$

$$\Rightarrow \Delta G = \frac{136}{1441} - \frac{15}{21} \cdot \frac{52}{225} - \frac{6}{21} \cdot \frac{4}{9} = \frac{4}{245} \approx 0.01633$$

Time Split



$$G(08:00) = 1 - \frac{3}{9}^2 - \frac{6}{9}^2 = \frac{4}{9}$$

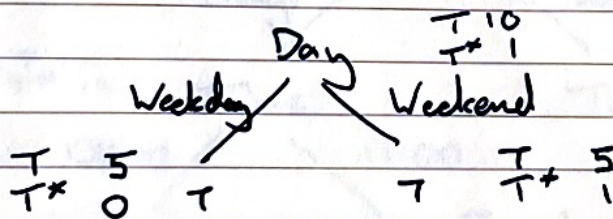
$$G(13:00) = 1 - \frac{1}{12}^2 - \frac{11}{12}^2 = \frac{11}{12}$$

$$\Rightarrow \Delta G = \frac{136}{121} - \frac{9}{21} \cdot \frac{4}{9} - \frac{12}{21} \cdot \frac{11}{12} = \frac{3}{98} \approx 0.03061$$

\therefore Splitting on time gives the highest information gain.
 $0.03061 > 0.02 \Rightarrow$ continue.

2) Evaluate splits at the Rainy node:

Day split

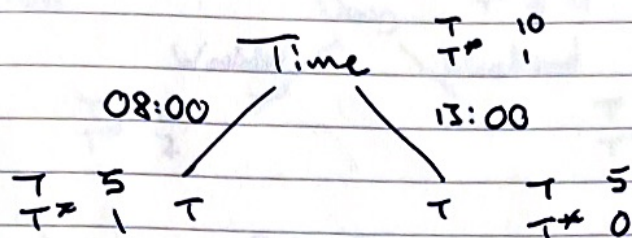


$$G(\text{Weekday}) = 1 - \frac{5}{5}^2 - \frac{0}{5}^2 = 0$$

$$G(\text{Weekend}) = 1 - \frac{5}{6}^2 - \frac{1}{6}^2 = \frac{5}{18}$$

$$\Rightarrow \Delta G = \frac{20}{121} - \frac{5}{11} \cdot 0 - \frac{5}{11} \cdot \frac{5}{18} = \frac{5}{363} \approx 0.01377$$

Time Split



$$G(08:00) = 1 - \frac{5}{6}^2 - \frac{1}{6}^2 = \frac{3}{18}$$

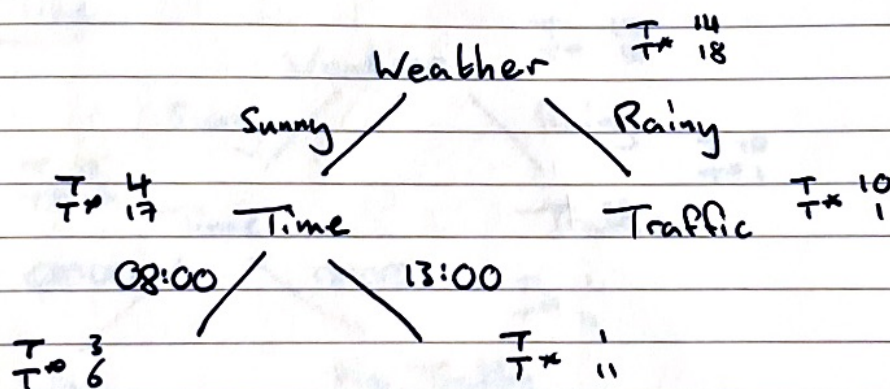
$$G(13:00) = 1 - \frac{5}{5}^2 - \frac{0}{5}^2 = 0$$

$$\Rightarrow \Delta G = \frac{20}{121} - \frac{6}{11} \cdot \frac{3}{18} - \frac{5}{11} \cdot 0 = \frac{5}{163} \approx 0.01877$$

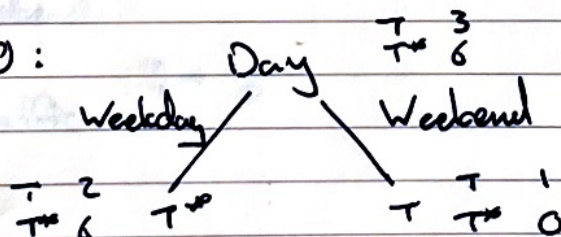
\therefore Both splits have a ΔG of < 0.02

\therefore Do not split.

\therefore The Decision Tree now looks like:



a) Evaluate 08:00:

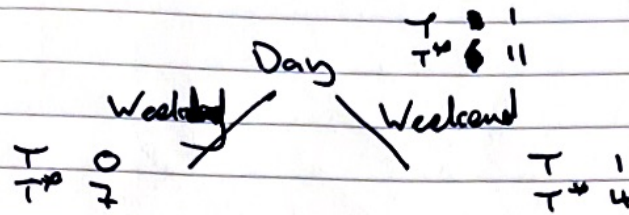


$$G(\text{Weekday}) = 1 - \frac{2}{8}^2 - \frac{6}{8}^2 = \frac{3}{8}$$

$$G(\text{Weekend}) = 1 - \frac{1}{1}^2 - \frac{0}{1}^2 = 0$$

$$\Rightarrow \Delta G = \frac{4}{9} - \frac{8}{9} \cdot \frac{3}{8} - \frac{1}{9} \cdot 0 = \frac{1}{9} \approx 0.1111$$

2) Evaluate 13:00



$$G(\text{Weekday}) = 1 - \frac{0}{7}^2 - \frac{7}{7}^2 = 0$$

$$G(\text{Weekend}) = 1 - \frac{1}{5}^2 - \frac{4}{5}^2 = \frac{8}{25}$$

$$\Rightarrow \Delta G = \frac{11}{72} - \frac{7}{12} \cdot 0 - \frac{5}{12} \cdot \frac{8}{25} = \frac{7}{360} \approx 0.01944$$

\therefore Should split at 08:00 but not at 13:00 as:

$$\frac{1}{5} > 0.02 \text{ and } \frac{7}{360} \leq 0.02.$$

5) Final Tree

