1a)

1. What would be the root node in this scenario?

Total Sales

1. Which variable could we encode by node size?

sales EUR

iii. What could be an appropriate hierarchical structure with 3 levels (root→

level 1 → level 2 → level 3)? Explain your choice.

Total Sales→ Country→ City→ supplier

iv. The data contains sales data of one year. Imagine we have data over multiple

years – how would this change your answers from i. - iii.?

1. Total Sales
2. sales EUR
3. Total Sales→ years→ Country→ City→ supplier

1b)

Advantages

Clear relationships between different levels.

Scalability: can represent large datasets by nesting nodes.

Disadvantages

When used on large datasets, it has too complex a structure, making it not easy to understand.

Takes up large space and may get cluttered.