Task 1a)
i. What would be the root node in this scenario?
Total sales

ii. 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
By following this hierarchy, we could clearly identify geographical trends.

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.?
i.   Total Sales
ii. sales EUR
iii. Total Sales→ years→ Country→ City→ supplier

Task 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.

Task 2b)

i. From the size of the rectangle, we could estimate it is Berlin, but it is similar to Hamburg so
we cannot be sure. "City" should be level 3 to answer this question properly. We can clearly
identify the country, Germany.

ii. Samsung, with the most sales in Hamburg. As in the previous question, it would be more
appropriate if "city" was level 3.

iii. Germany has the highest cumulative sales, corresponding to around 75% of the total
sales.

iv. Yes: Winterthur (Simons), Solingen (Samsung) and Osnabrück (LG)

v.   Several suppliers operate only in Germany: Micromax, HP, Ulefone, Spice, Toshiba,
Yezz, Icemobile, Panasonic and vivo.
   - Plum operates mainly in Switzerland and just has a very low sales volume in Austria
     and in Hannover, Germany.

vi. Six, the highest sales are generated by Nokia ($91826)