A wine producer has collected chemical measurements from 178 wines produced from three different grape cultivars. The dataset contains 13 features describing various chemical properties of each wine.

Analyze the dataset using hierarchical clustering, identify patterns, and interpret the results.

N.B: The true cultivars of the wines are stored in the target variable *y*. *y* is not used for clustering ( it is used to evaluate the results : unsupervised learning) .

---

**Dataset**

- **Features (X)**: 13 continuous chemical measurements (Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315, Proline)
- **Target (y)**: true wine cultivar (1, 2, 3) — for evaluation only

To load the dataset use the following code :

```
from ucimlrepo import fetch_ucirepo

# fetch dataset

wine = fetch_ucirepo(id=109)

# data (as pandas dataframes)

X = wine.data.features

y = wine.data.targets

# metadata

print(wine.metadata)

# variable information

print(wine.variables)

print(X.shape)
```

1. ***Data exploration***
    - *Check for missing values and describe the range of the features.*
    - *Plot histograms or boxplots for at least 3 features to visualize distributions.*

2. ***Data preprocessing***
    - *Standardize the features using z-score normalization (mean = 0, std = 1).*
    - *Explain why standardization is necessary in hierarchical clustering.*
3. ***Hierarchical clustering***
    - *Compute the **linkage matrix** using the Ward method.*
    - *Plot the **dendrogram** for all observations.*
    - *From the dendrogram, choose a reasonable number of clusters and assign cluster labels to the wines.*
4. ***Cluster analysis***
    - *For each cluster, compute the **mean and standard deviation** of the features.*
    - *Interpret the clusters: are there clear differences in chemical composition?*
5. ***Evaluation***
    - *Compare your cluster assignments with the true labels y.*
    - *Compute the **confusion matrix***
    - *Interpret the confusion matrix*