Submission sheet
Assignment III

TASK 1: k-Means

Corresponds (more or less) to the three expected species? ☒YES ☐ NO

Number of records in each cluster: 1) 85 2) 112 3) 103

TASK 2: preprocessing

Is it better to rescale before or after detecting and filtering out the outliers? Normalice before

Corresponds (more or less) to the three expected species? ☒YES ☐ NO

Number of records in each cluster: 1) 83 2) 112 3) 103

Coordinates of the three centroids:

	PW	PL	SW	SL
1)	5.77	3.04	3.53	1.18
2)	5.76	2.94	3.74	1.23
3)	5.75	3.05	3.70	1.27

TASK 3: choice of k

Which K corresponds to the best clustering? (using the Davies-Boulding index).
2

TASK 4: Hierarchical clustering

Using SingleLink, how many records are included in each of the two top clusters?

Cluster 1: 286

Cluster 2: 14

Which approaches produce a (more or less) correct clustering corresponding to the three species, if any?

SingleLink: does not make correct clustering

CompleteLink: make correct clustering

AverageLink: make correct clustering

TASK 5: DB-Scan

How many clusters does DB-SCAN find with eps=1, min_samples=5? 1

Can you give a value for epsilon leading to two clusters (plus noise)? 10000

K-DISTANCES:

Which K did you use? 7

According to the k-distances plot, what value(s) of epsilon would you consider as a parameter to DB-Scan and why?

The point where it changes drasticlly the k-distance plot since it indicates that transicion from the points that are close together to points that are far away that represents the points in a different cluster