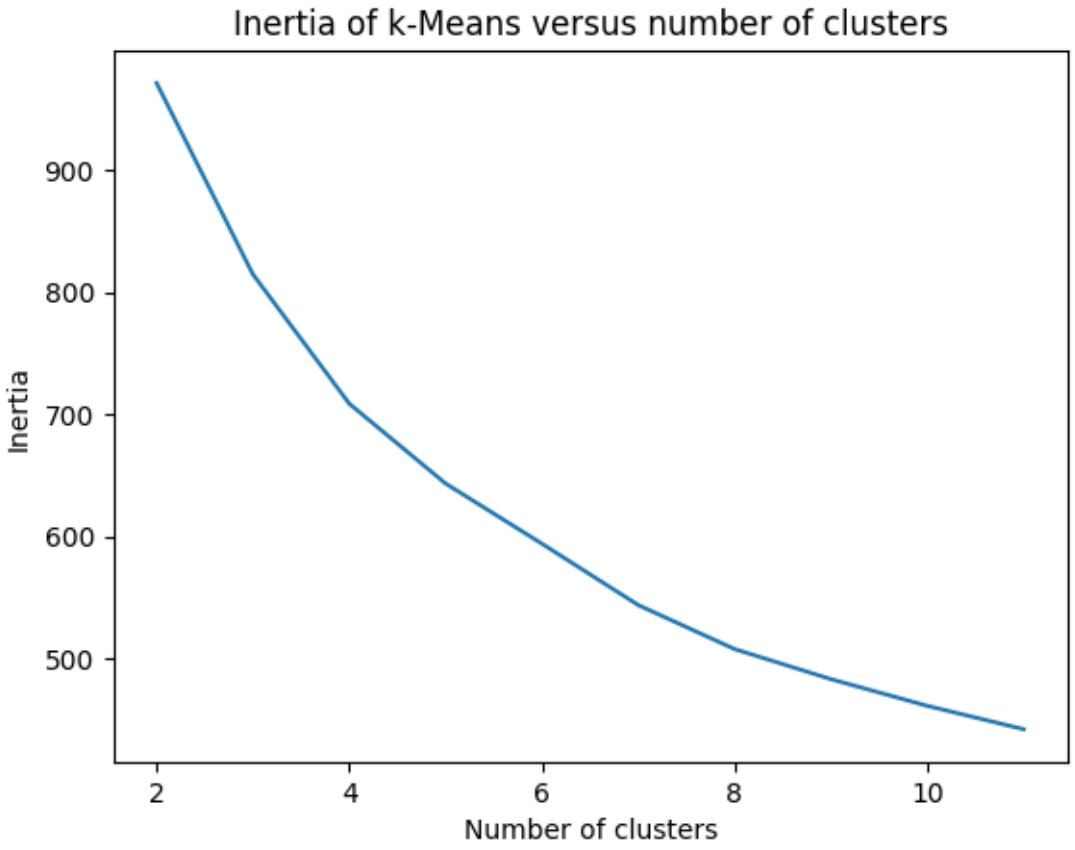
Introduction

Three clustering techniques on performance raquets. N = 148 raquets and clustering done on 8 variables. Note that the variable mass_unstrung is multiplied by 2 after scaling so as to give mort importance to it. 8 techniques were tested but the K-means, Gaussian Mixture Modeling and Hierarchical Clustering looked the best in the aim of our study. The clusters are with the csv file raquetclusters attached.

1 K-means clustering

1.1 Elbow criteria

We choose a K = 5 for this criteria.



1.2 Cluster centers and distinction

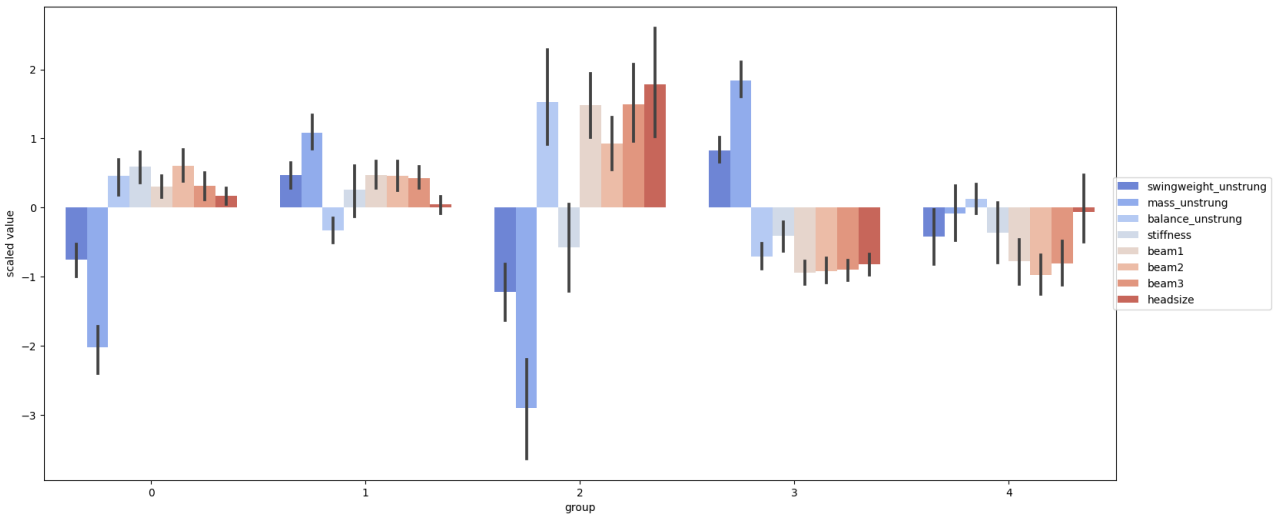


Figure 1: Cluster centers of the K-means technique

1.3 PCA plot

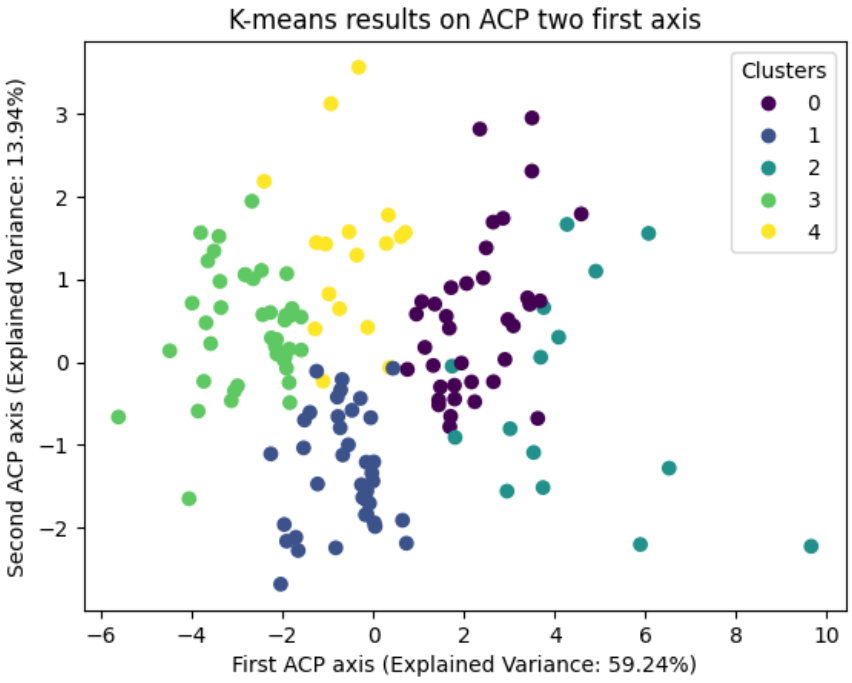


Figure 2: PCA plot with the K-means clusters

Clusters look distinct on the PCA (explaining 74% of the variance). Moreover the number of observations in each cluster is homogeneous (respectively 36, 39, 15, 41 and 17).

2 Gaussian Mixture Model

2.1 Parameter selection

Best parameter choice using the GridSearchCV library on Python and the criteria of the Bayesian Information Criterion (BIC) score.

Number of components	Type of covariance	BIC score
6	diag	706.098725
4	diag	754.021459
5	diag	769.322472
3	diag	795.194757
6	tied	814.499381

Figure 3: Best parameters for the Gaussian Mixture model

We will choose next 6 components (clusters) for our Gaussian Mixture model and a diagonal covariance type.

2.2 Cluster centers and distinction

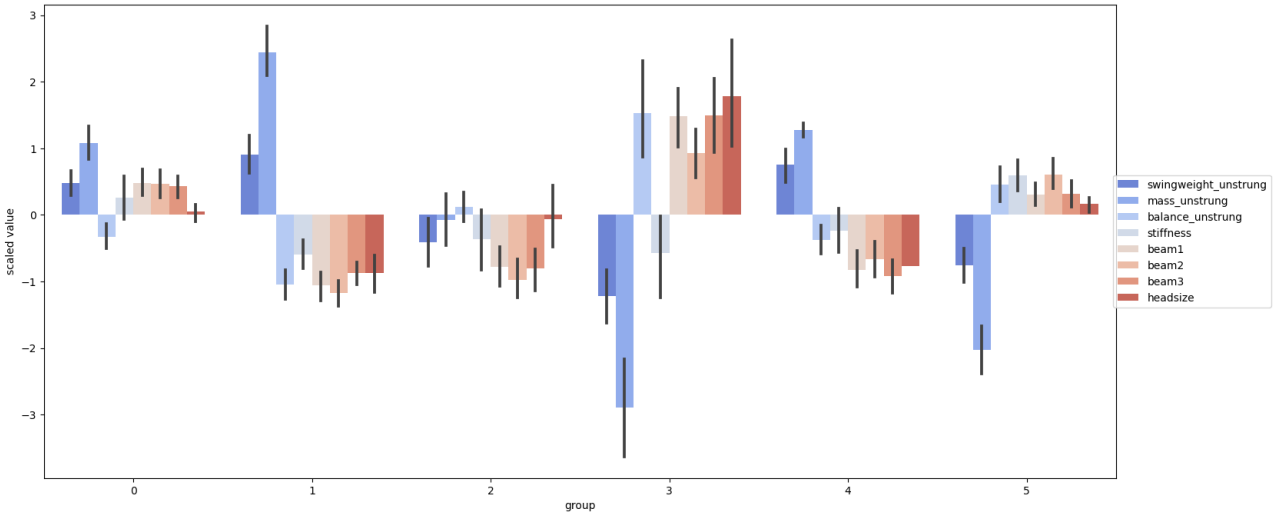


Figure 4: Cluster centers of the Gaussian Mixture technique

2.3 PCA plot

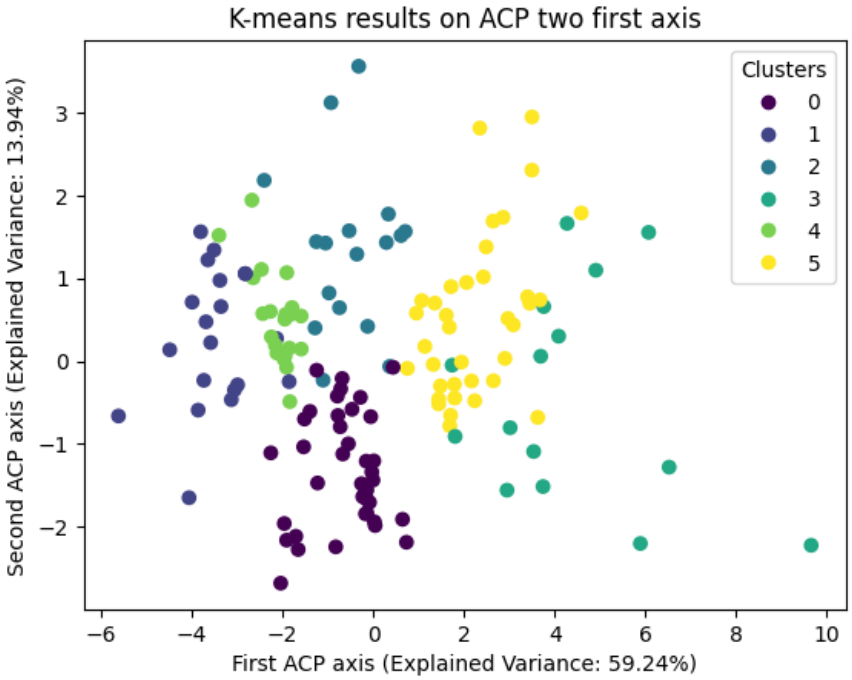


Figure 5: PCA plot with the Gaussian Mixture clusters

Here there is also quite separated groups, and homogeneous number of observations per cluster.

3 Agglomerative clustering

3.1 Dendrogram

Looking at this graph, we can make the choice of k = 4 clusters.

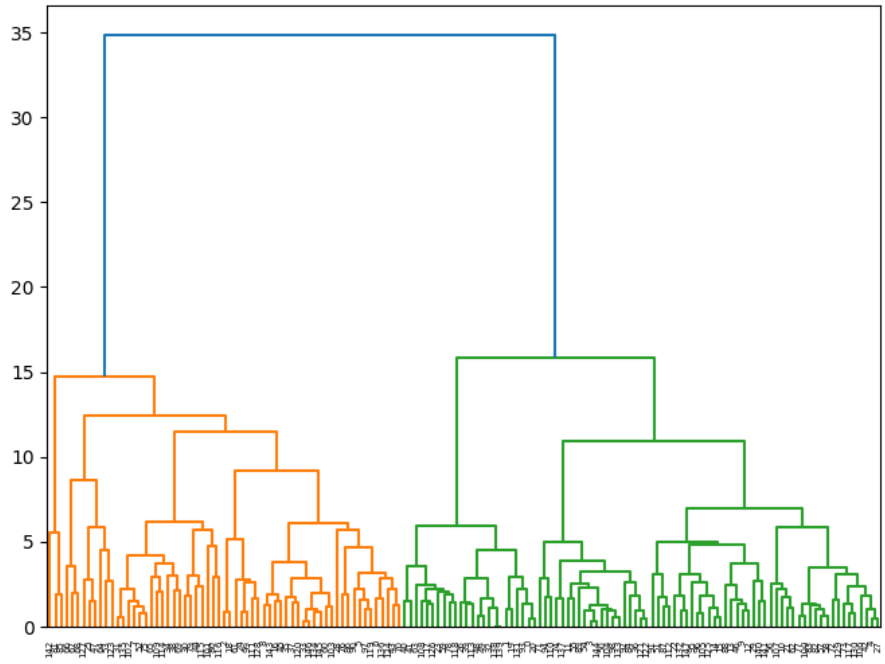


Figure 6: Dendrogram

3.2 Cluster centers

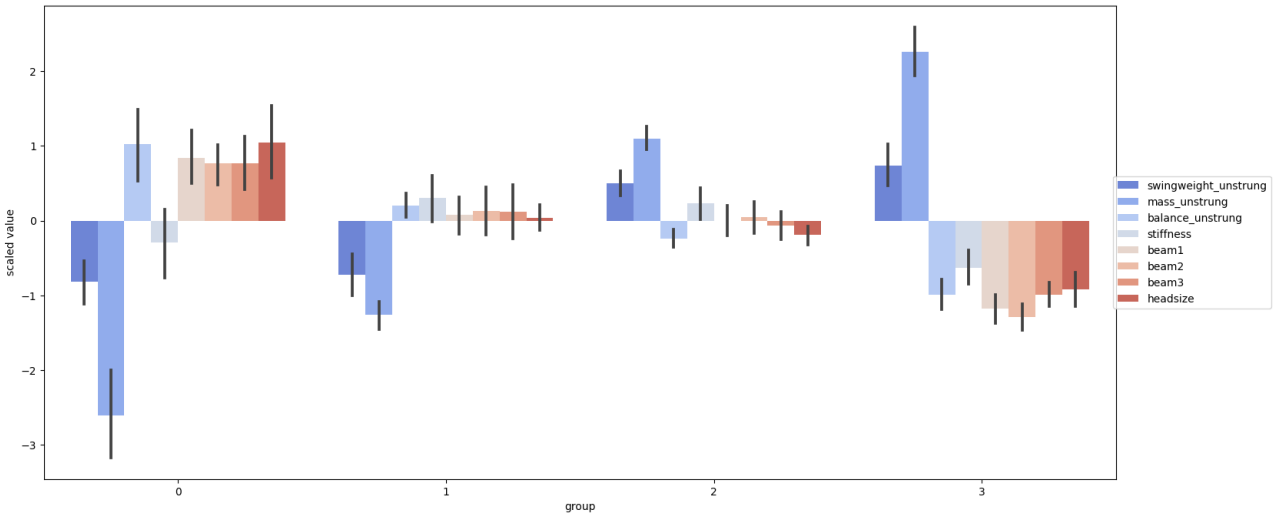


Figure 7: Cluster centers of the Hierarchical Clustering technique

3.3 PCA plot

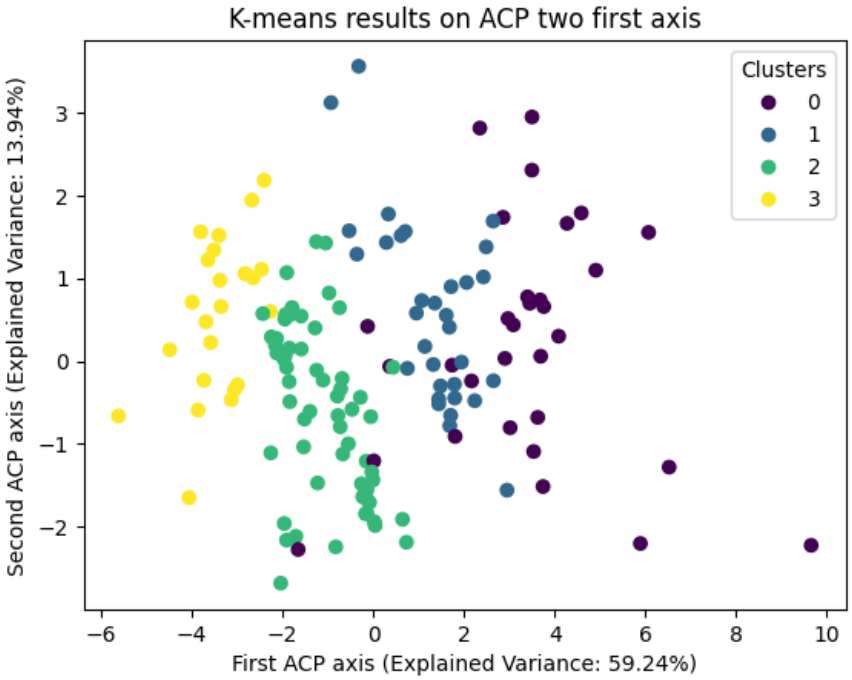


Figure 8: PCA plot with the Agglomerative clustering clusters