



La poule qui chante

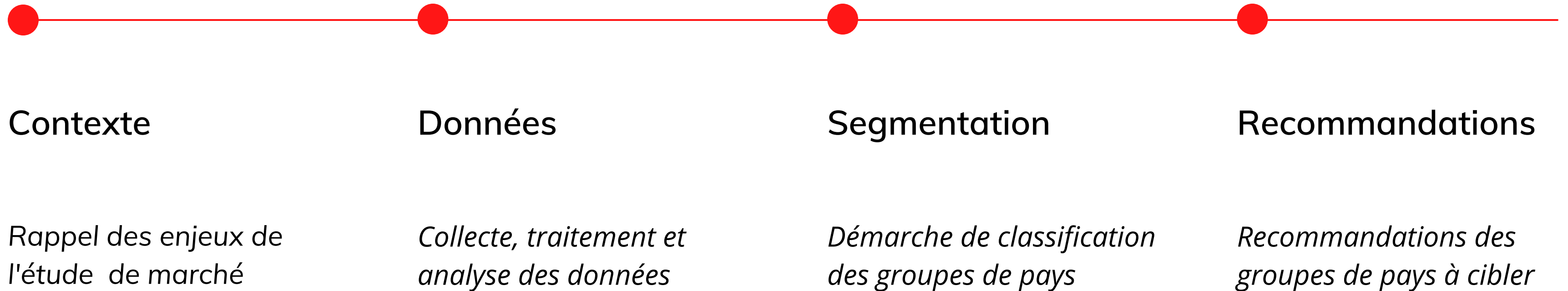
Etude de marché

Commercialisation de poulets à
l'échelle mondiale

Novembre 2022



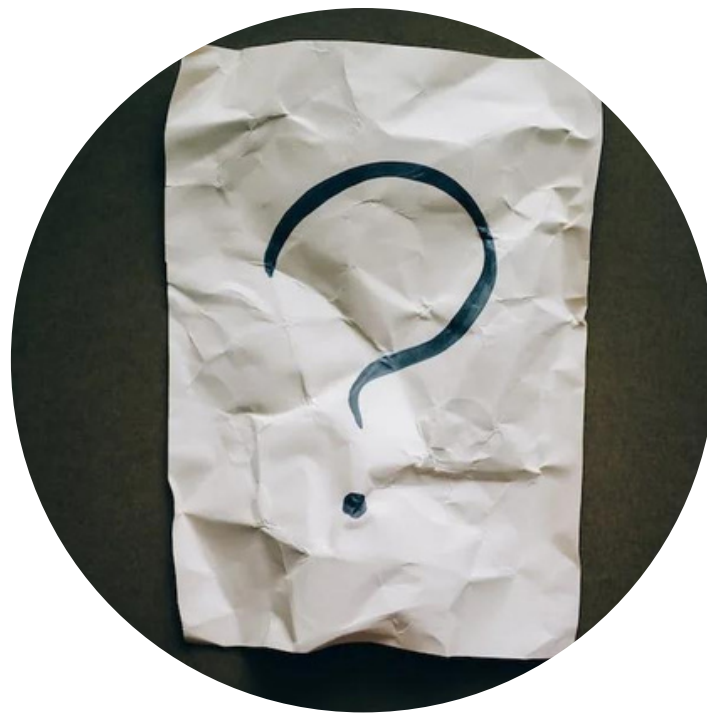
Plan de la présentation



Contexte



La poule qui chante



Qui ?

Entreprise
agroalimentaire
spécialisée dans la vente
de poulets.



Quoi ?

Volonté de lancer la
commercialisation à
l'international.



Comment?

Critères de sélection
établis à partir des
données de la FAO
(Food and Agriculture
Organization).



Où?

Sur la base de ces
critères, déterminer les
pays les plus profitables.

Données : analyse exploratoire



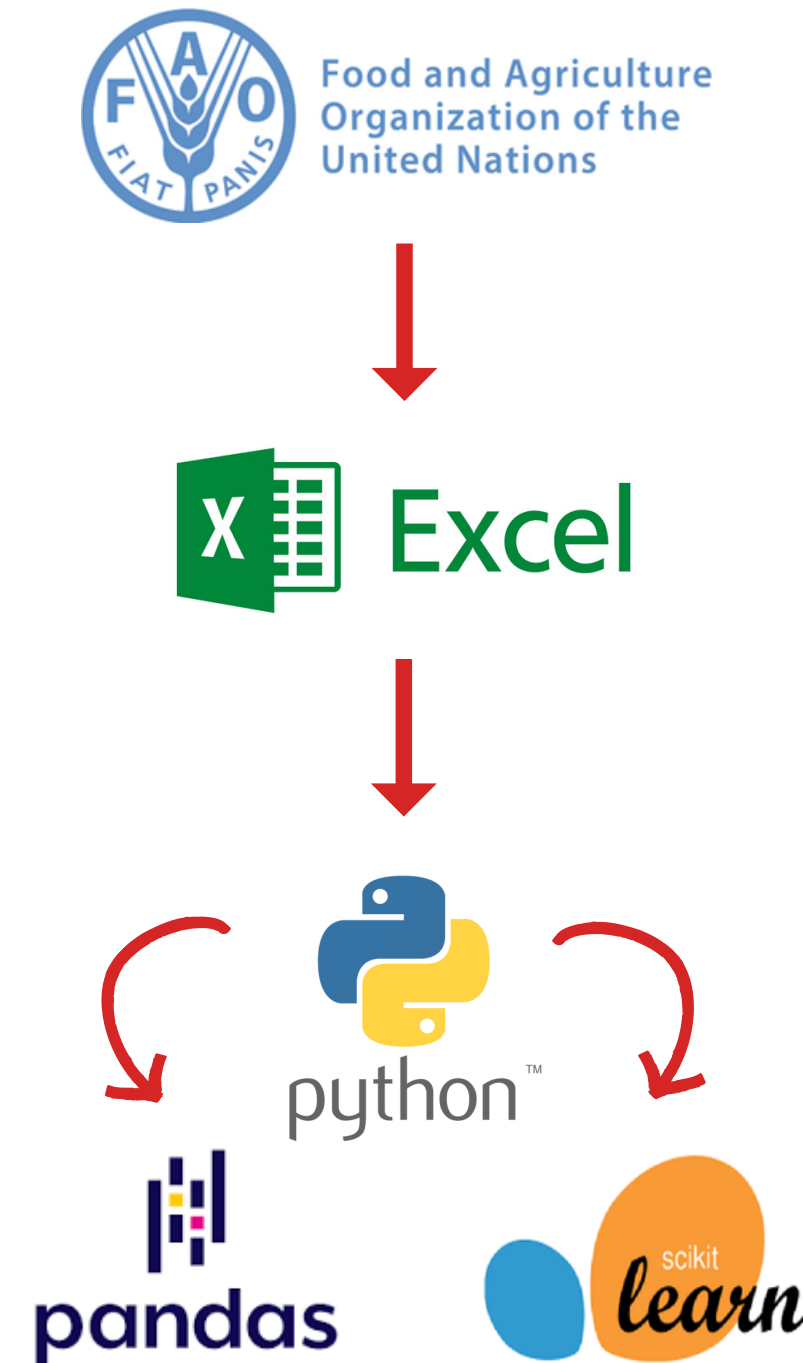
La poule qui chante

1 Collecte

Toutes les données de l'année 2017 ont été collectées sur le site de la FAO :

- Focus sur les données "volailles"
- Disponibilité alimentaire (disponibilité individuelle, consommation, production, importations...)
- Population
- PIB
- Insécurité alimentaire
- Stabilité politique...

Type de fichiers : .csv



Principe de collecte et traitement des données

2 Nettoyage

Les différentes tables sont ensuite jointes pour n'en faire qu'une seule :

- Renommage des variables (colonnes)
- *Suppression des variables jugées "inutiles"* (semences, pertes, exportations...)
- Gestion des valeurs nulles :
 - Suppression
 - Imputation de valeurs précises (ex : PIB en 2020 de Taïwan)

Nombre de pays étudiés : 170

	valeurs_nulles	pourcentage
NPSA_m	61	35.88
NPSIAG_m	55	32.35
SAEAM_%	10	5.88
Stabilite	2	1.18
PIB_habitant_usd	1	0.59
DA_Kcal_personne_jour	0	0.00
DA_kg_personne_an	0	0.00
DMG_g_personne_jour	0	0.00
DP_g_personne_jour	0	0.00
DI_kt	0	0.00
Importations_kt	0	0.00
Nourriture_kt	0	0.00
Production_kt	0	0.00
Population_m	0	0.00

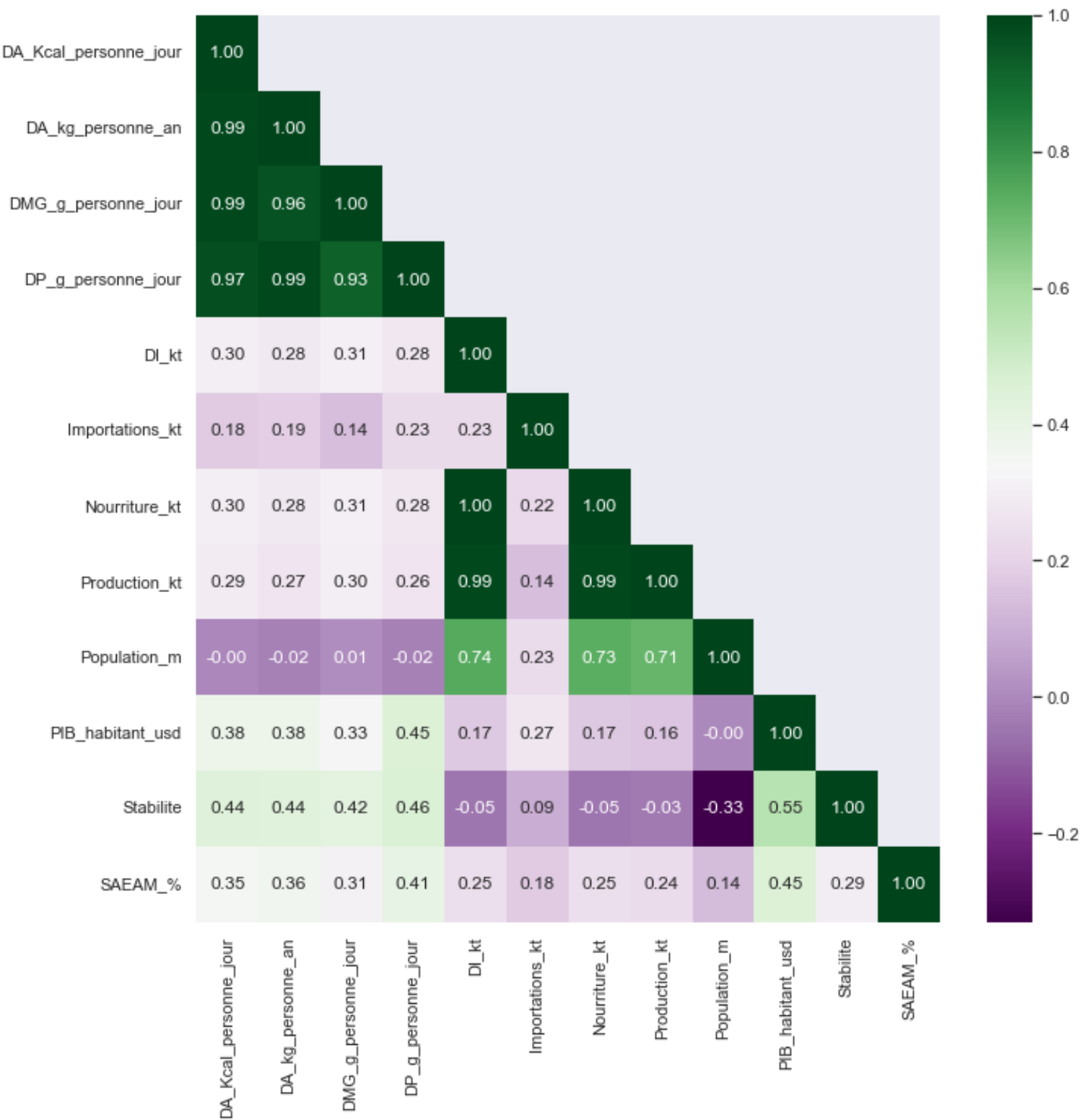
Pourcentage de valeurs nulles

3 Analyse

Une fois la table nettoyée, elle est encore traitée pour améliorer l'analyse :

- *Suppression de certains outliers (Inde et Chine continentale)*
- Pré-traitement des données pour la réduction de dimensionnalité (réduction des variables)

Table finale : 170 lignes x 12 colonnes

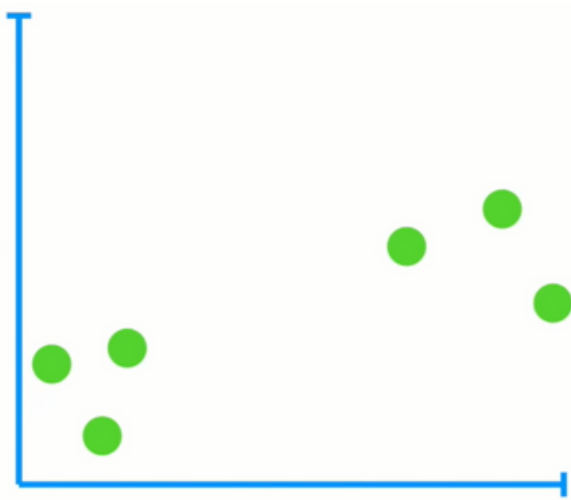


Matrice de corrélation entre les variables

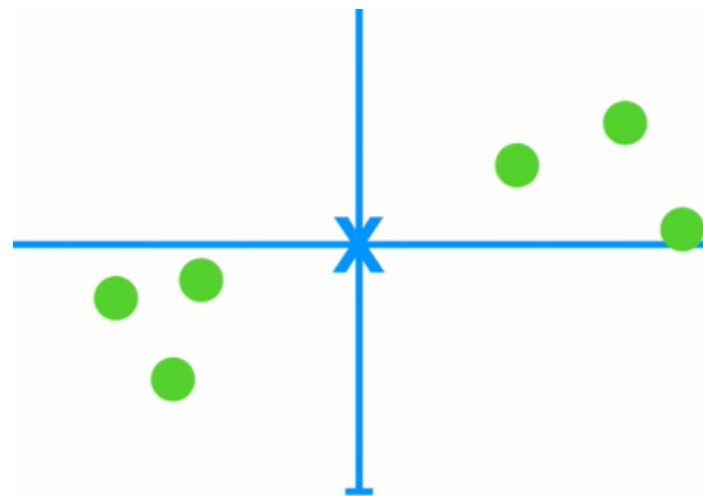


Données : réduction de dimensionnalité (ACP)

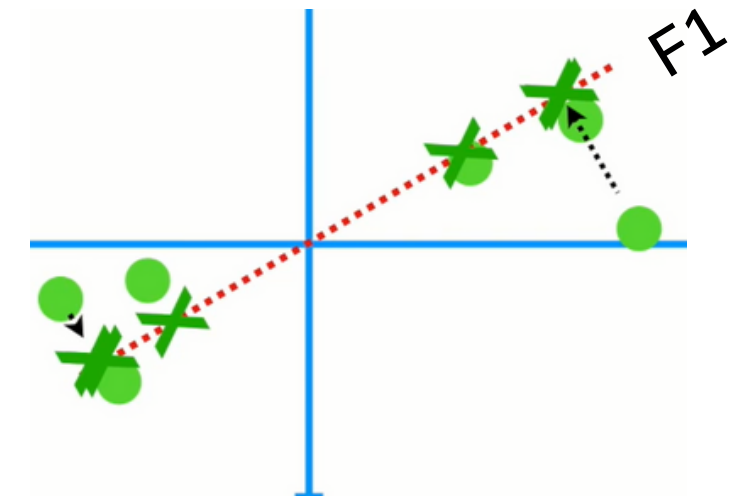
L'Analyse en Composantes Principales (ACP) est une technique linéaire non supervisée permettant de réduire les variables (colonnes) d'un jeu de données. Ci-dessous une explication du principe :



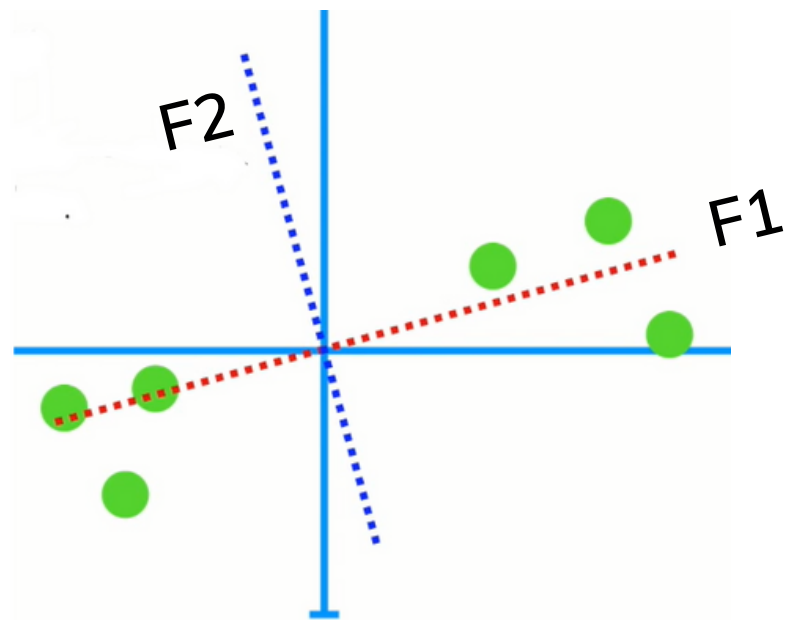
1 Données initiales



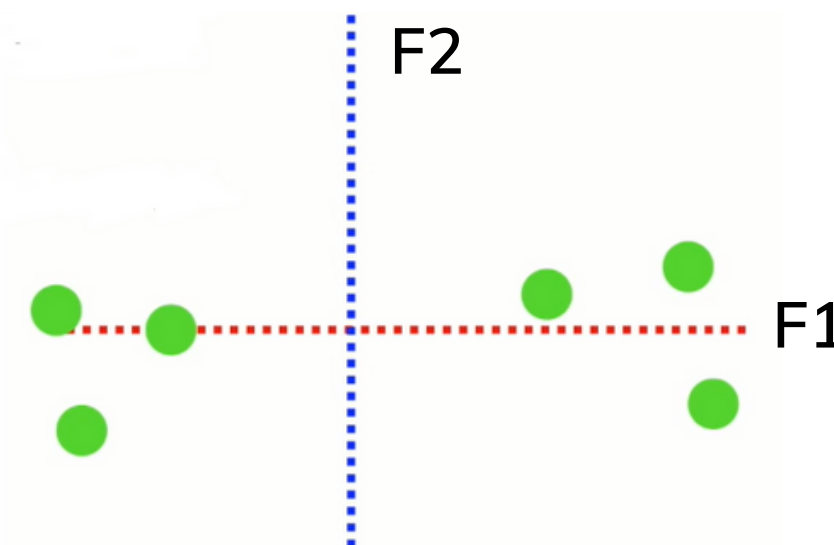
2 Données centrées-réduites



3 F1 : axe se rapprochant le plus de tous les points



4 F2 : axe orthogonal à F1



5 Projection des points suivant F1 et F2

6 Etapes 1 à 5 répétées jusqu'à un certain point

Données : réduction de dimensionnalité (ACP)



Ainsi, toutes les variables ont pu être résumées en 3 "composantes" :

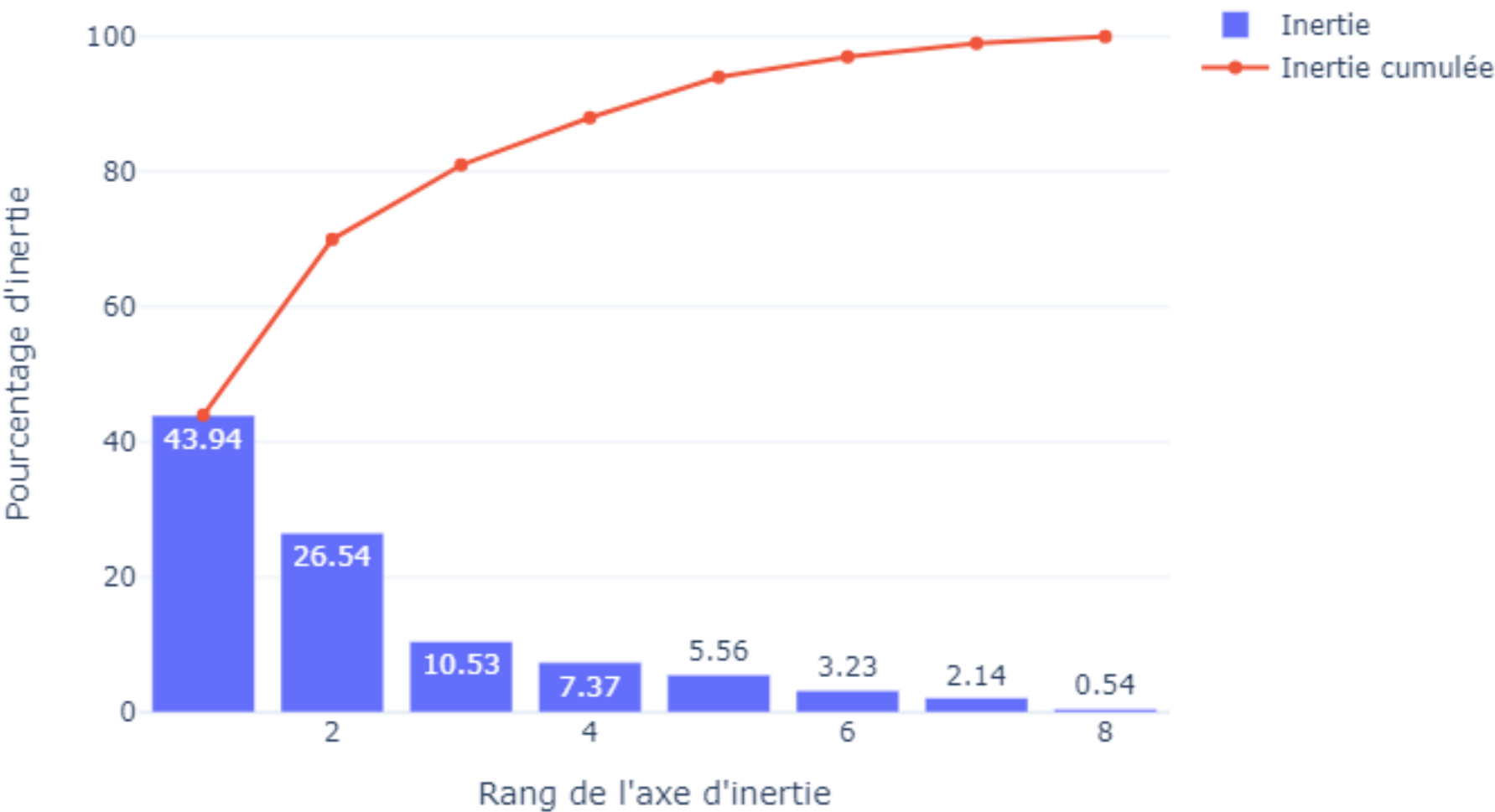
- *F1 : disponibilité alimentaire individuelle*
- *F2 : population-consommation de poulets*
- *F3 : PIB/habitant*

170 lignes x
12 colonnes

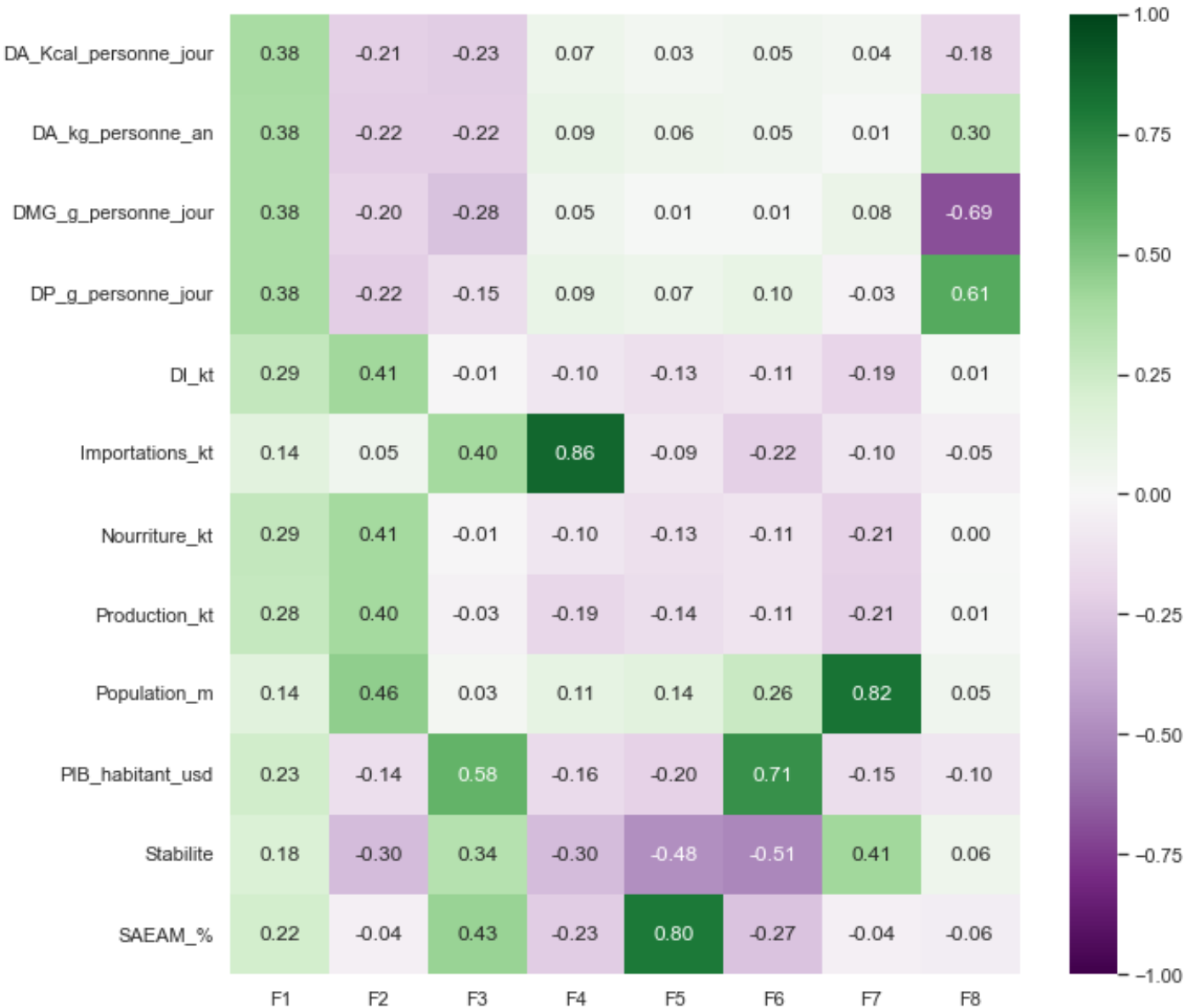


170 lignes x
3 colonnes

Eboulis des valeurs propres



Pourcentage de variance des composantes



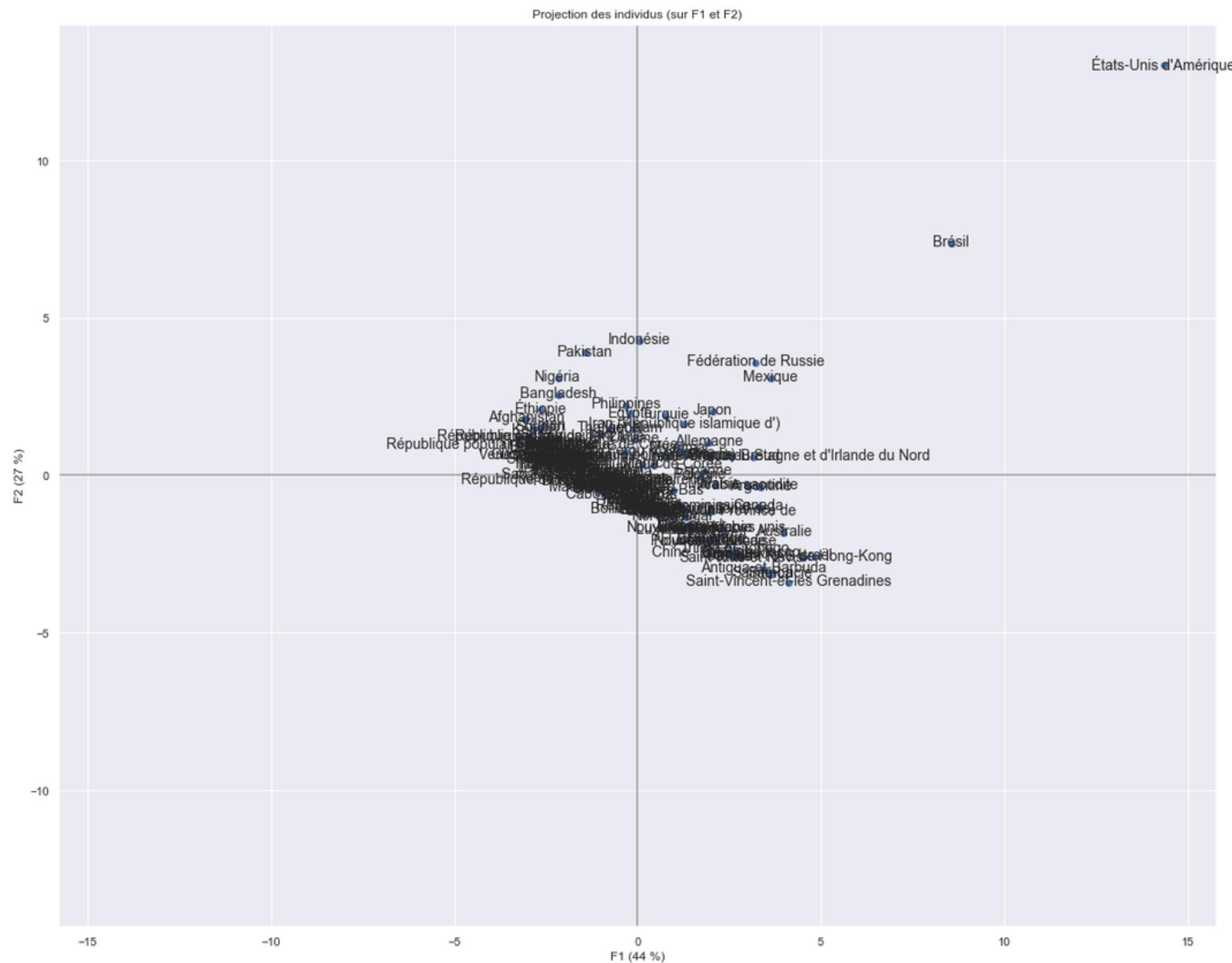
Proportion des variables vis-à-vis des composantes



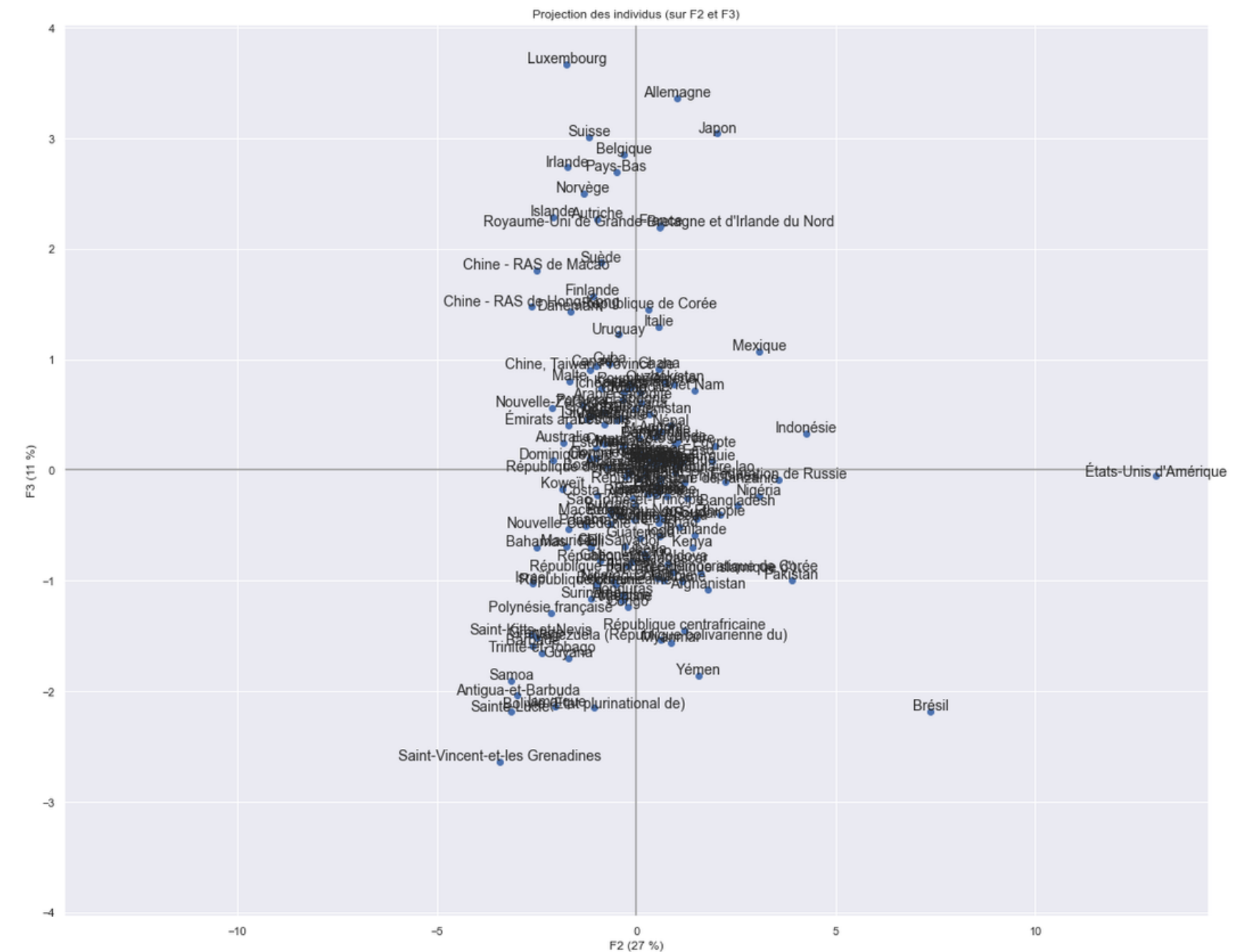
Données : réduction de dimensionnalité (ACP)

Les pays sont ensuite "projetés" suivant les 3 composantes retenues. Exemples :

- Etats-Unis et Brésil : pays peuplés grands consommateurs de poulets
- Luxembourg : pays à faible population ayant un revenu par habitant élevé



Projection des pays suivant F1 et F2



Projection des pays suivant F2 et F3

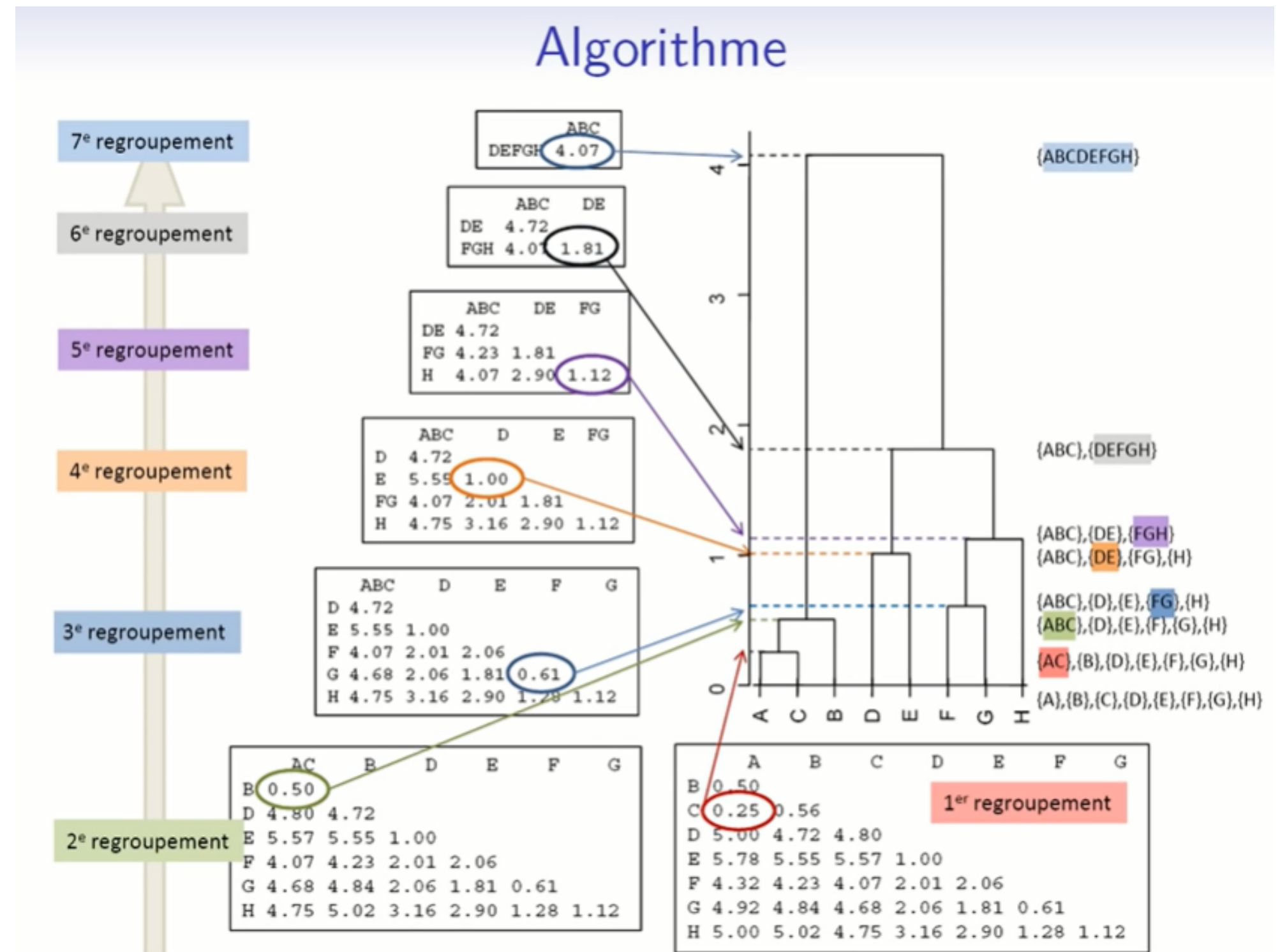


Segmentation : CAH

La Classification Ascendante Hiérarchique Classification (CAH) est un algorithme de classification non supervisée permettant de segmenter des données.

Principe

- Calculs des distances minimales entre points
- Minimiser ces distances
- Grouper les points les plus proches
- "Remonter l'arbre" jusqu'à ce qu'il n'y ait plus de points à grouper



Principe de calcul de la CAH

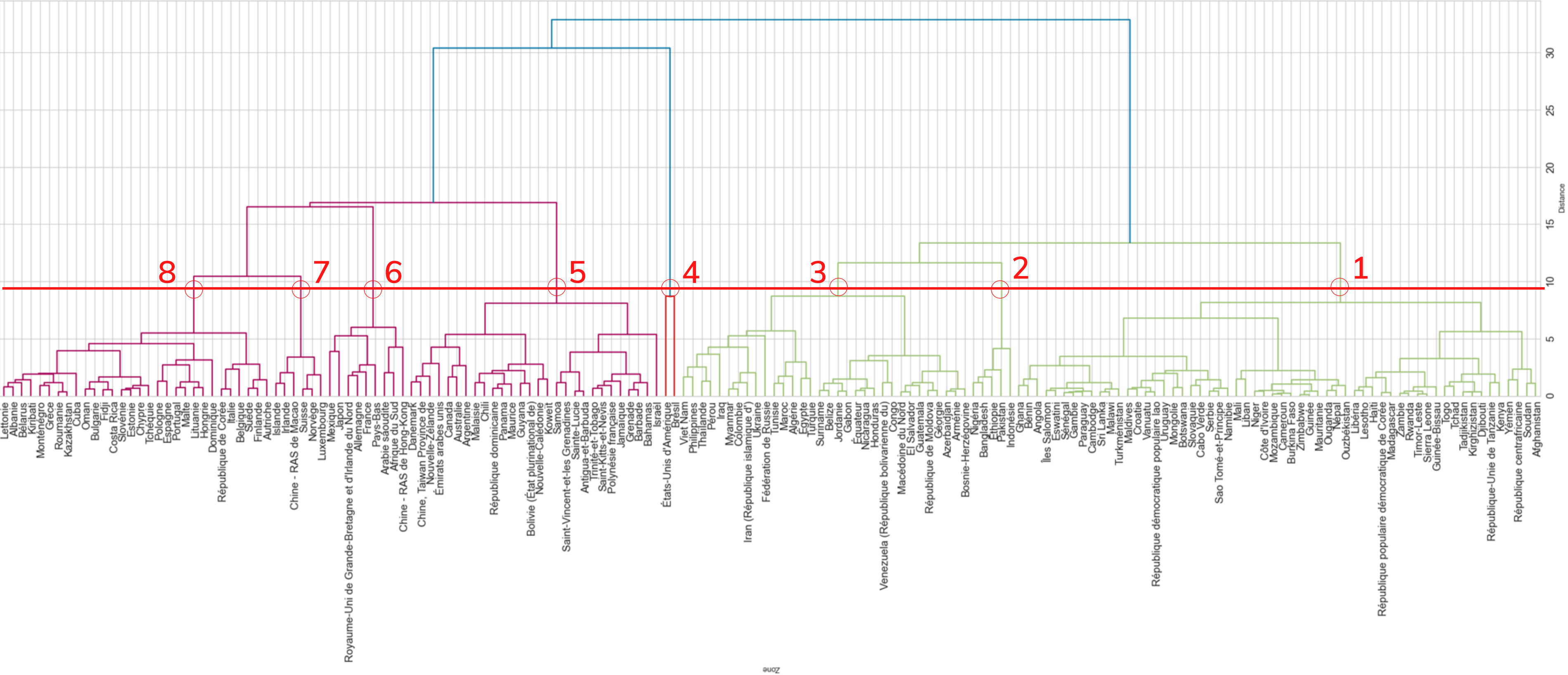
Segmentation : CAH



La poule qui chante

Arbre de classification : dendrogramme

Total obtenu : 8 clusters

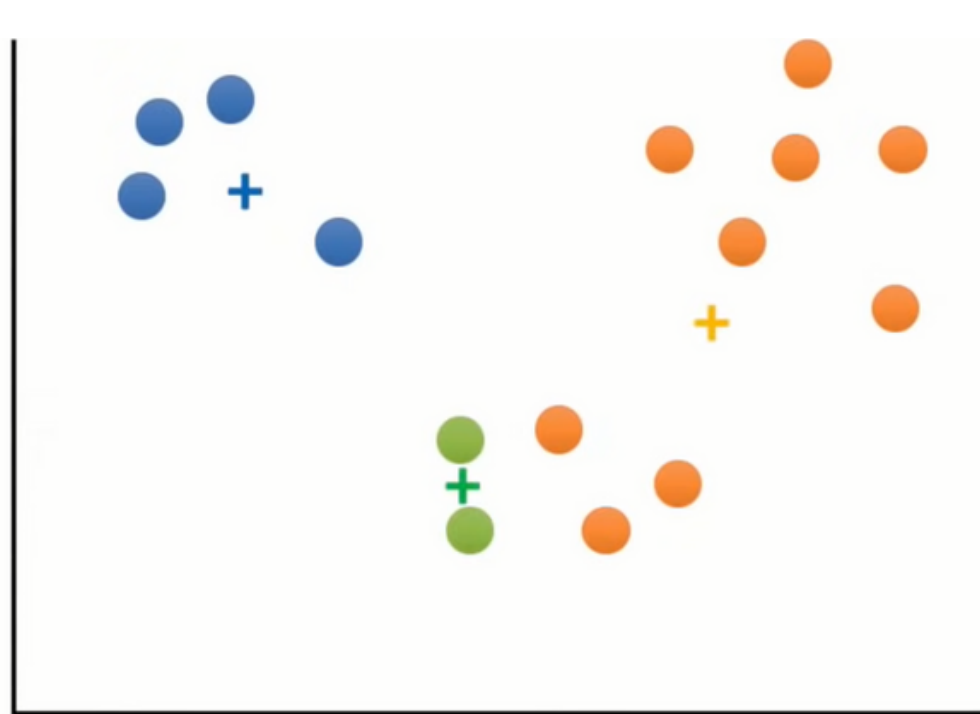


Segmentation : K-means

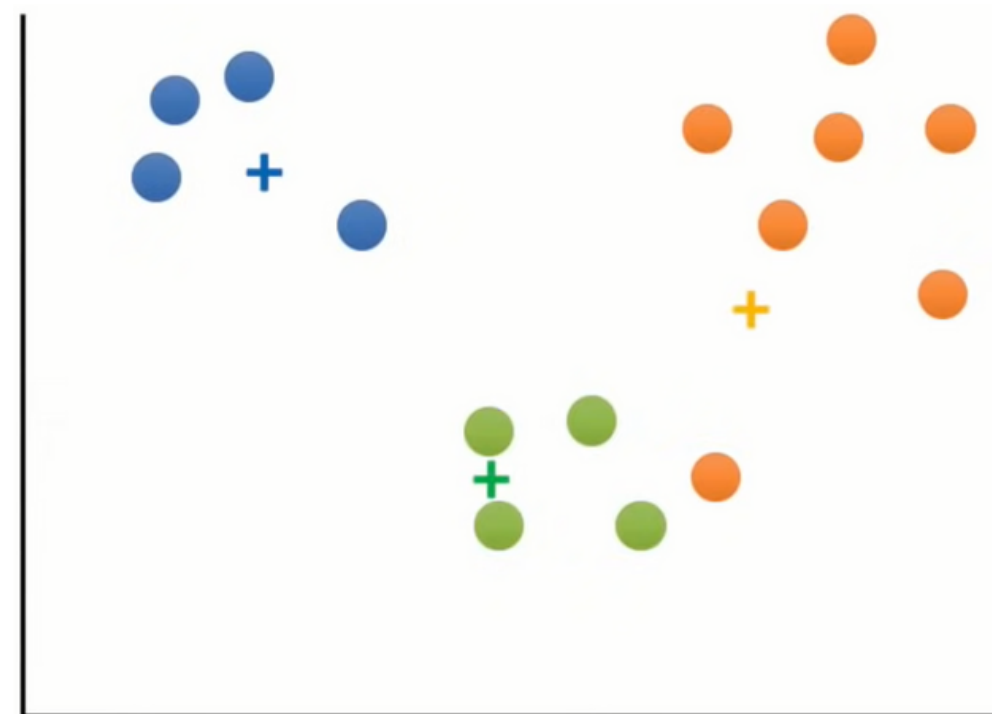


La poule qui chante

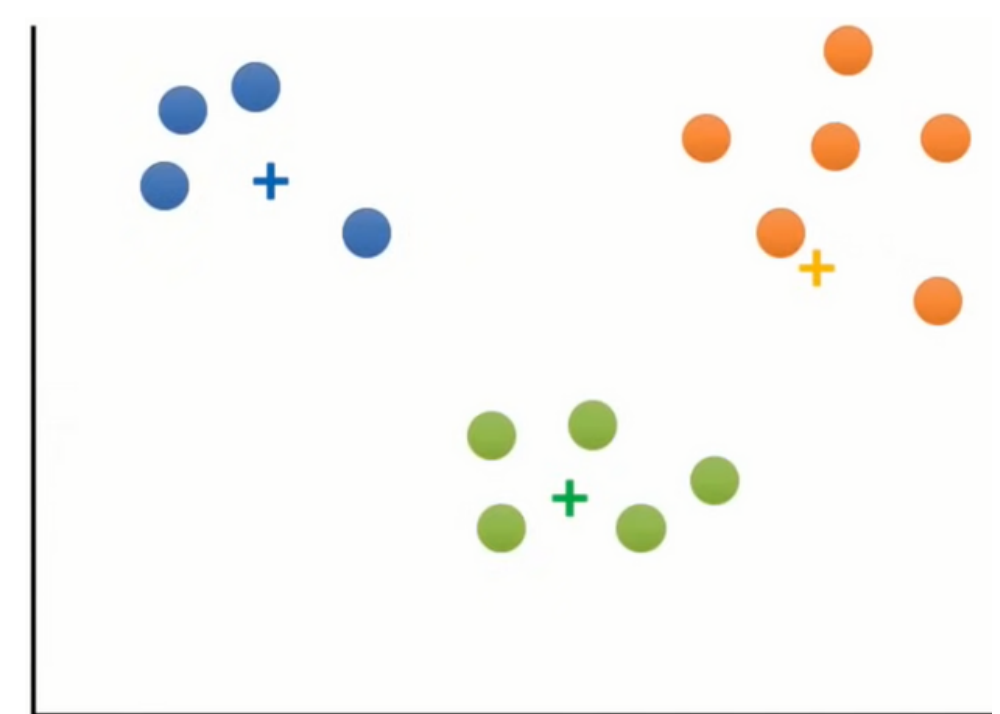
Le K-means est un algorithme de classification non supervisée permettant de segmenter des données.
Il fonctionne par calculs itératifs des centroïdes (centre de gravité d'un groupe).



Initialisation
Centroïdes placés
aléatoirement



n-itérations
Calculs succesifs de
différents centroïdes

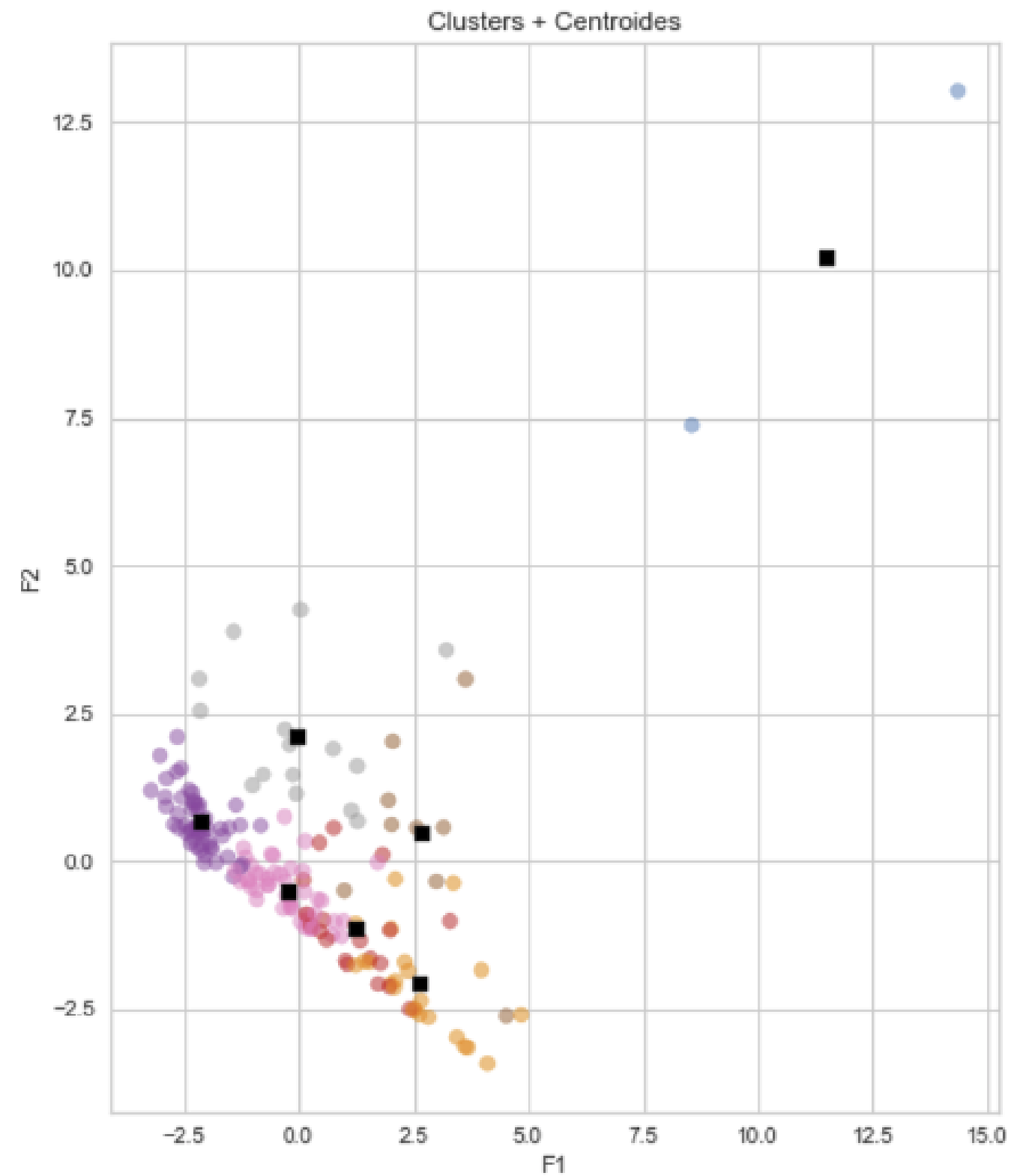
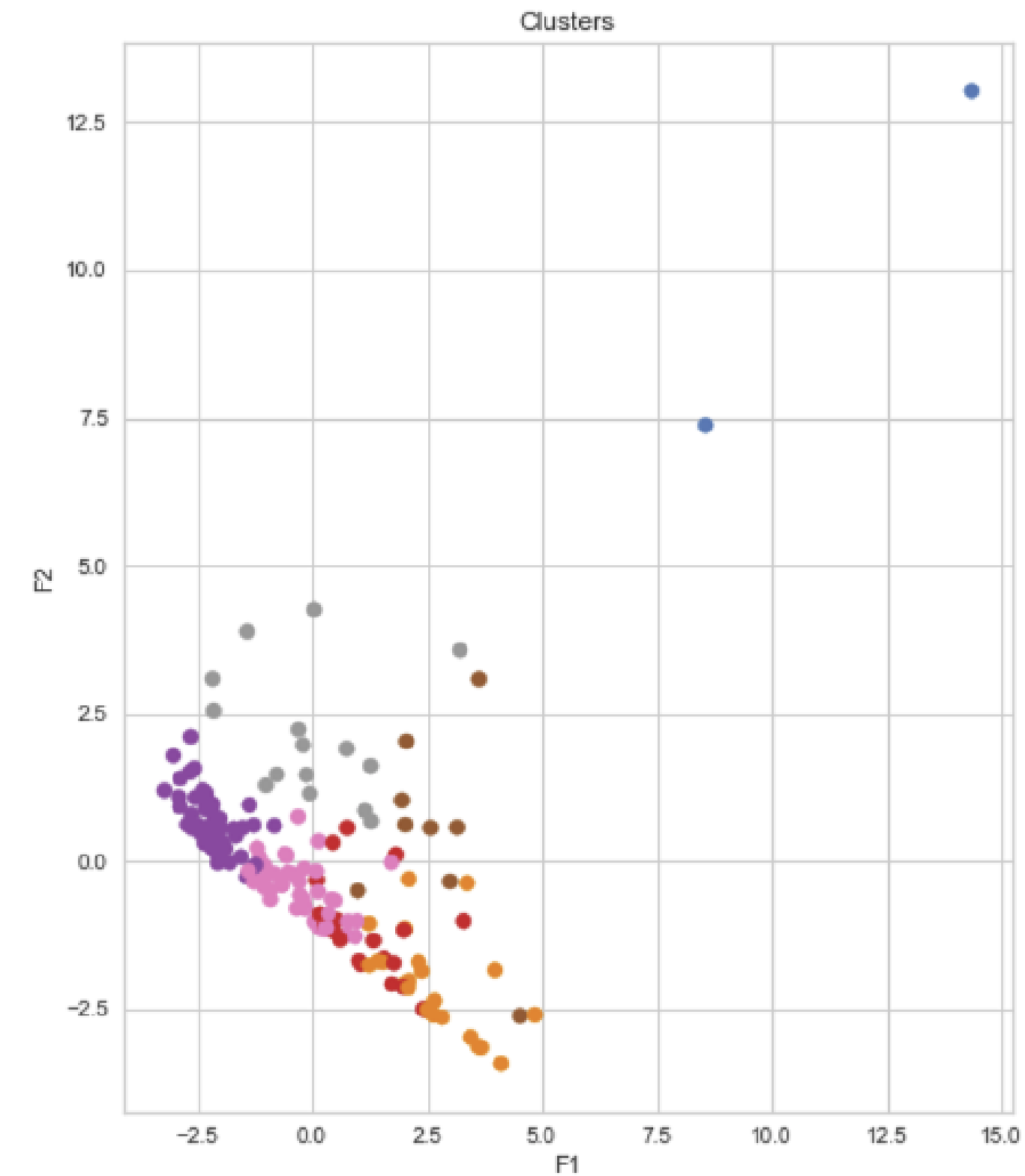


Convergence
Centroïdes placés à
leur position finale

Segmentation : K-means



La poule qui chante



Clusters + centòides obtenus

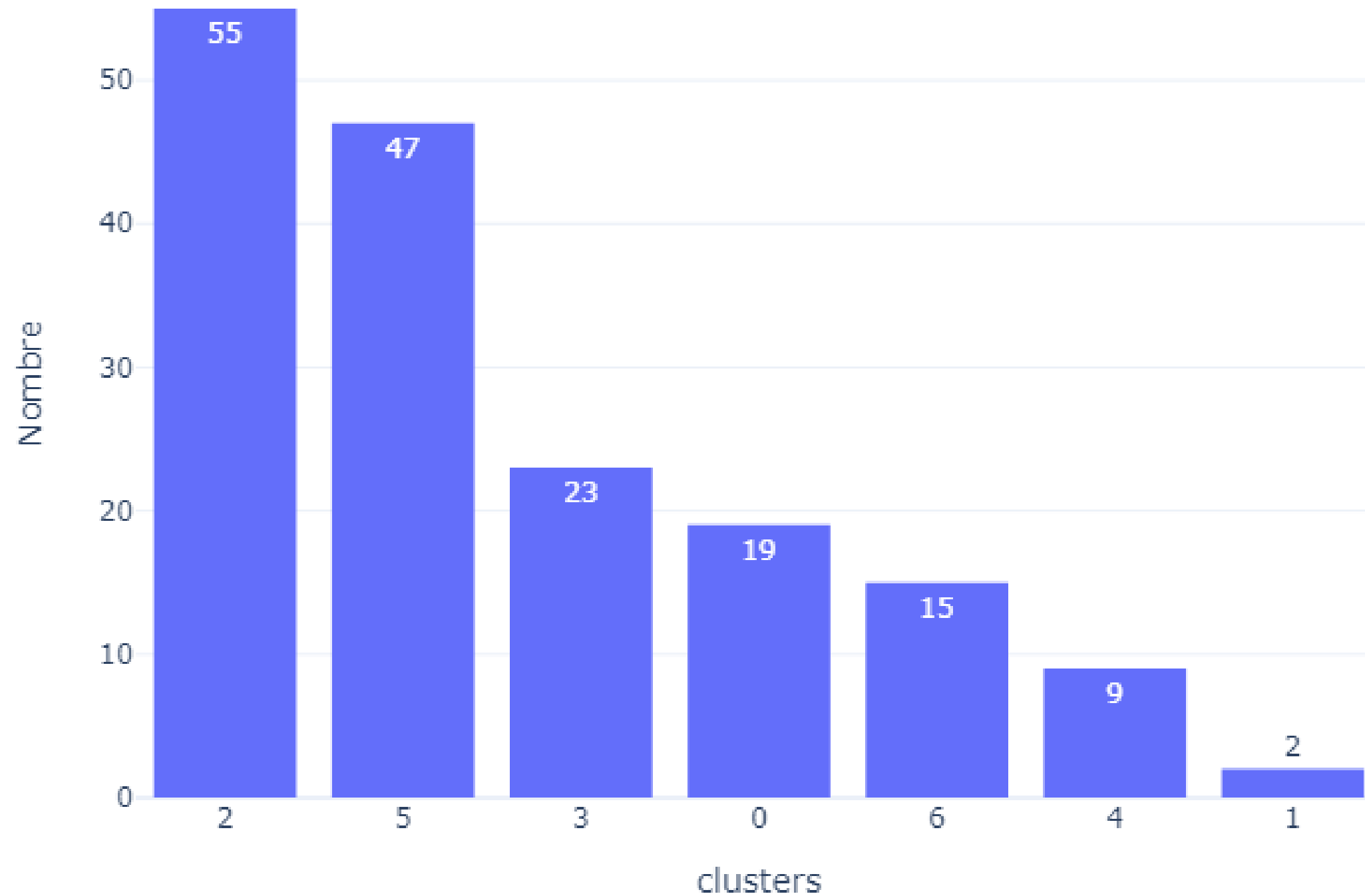
Analyse du clustering :

- *Outliers*
- *Centroïdes distincts*
- *Chevauchement entre points de clusters différents*



Segmentation : K-means

Répartition des zones par clusters - KMeans



Nombre de pays par clusters K-means

Analyse du clustering :

- 7 clusters
- Répartition hétérogène
- Outliers (2 pays)



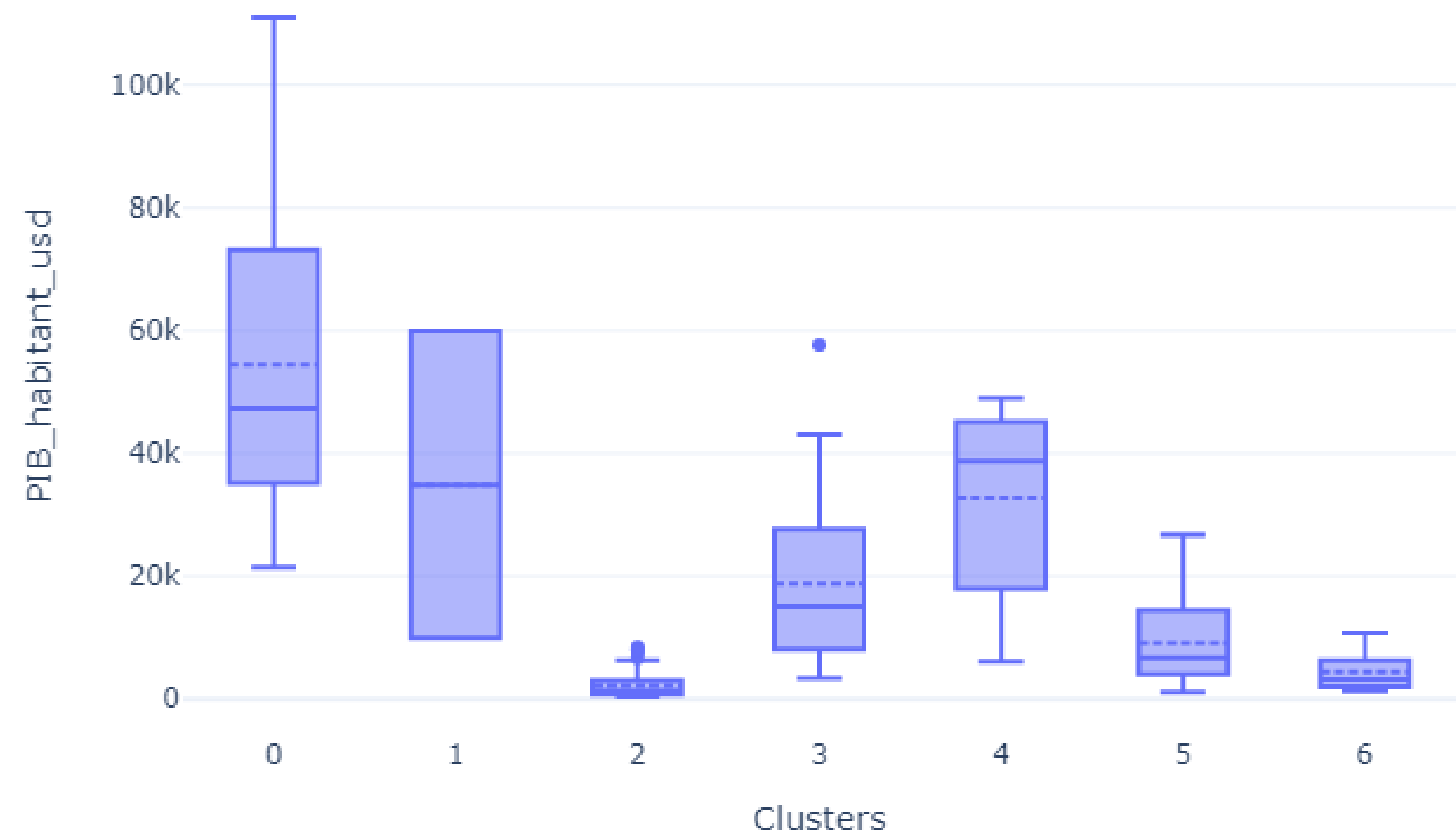
Recommandations

Choix pour les recommandations : Clustering K-means / Variables pertinentes et prépondérantes

Critères d'analyse de la segmentation des pays

- 1 Disponibilité alimentaire individuelle
- 2 Consommation de poulets
- 3 Population du pays
- 4 Revenus par habitant (PIB)
- 5 Stabilité politique du pays

Distribution des clusters K-means suivant : PIB_habitant_usd

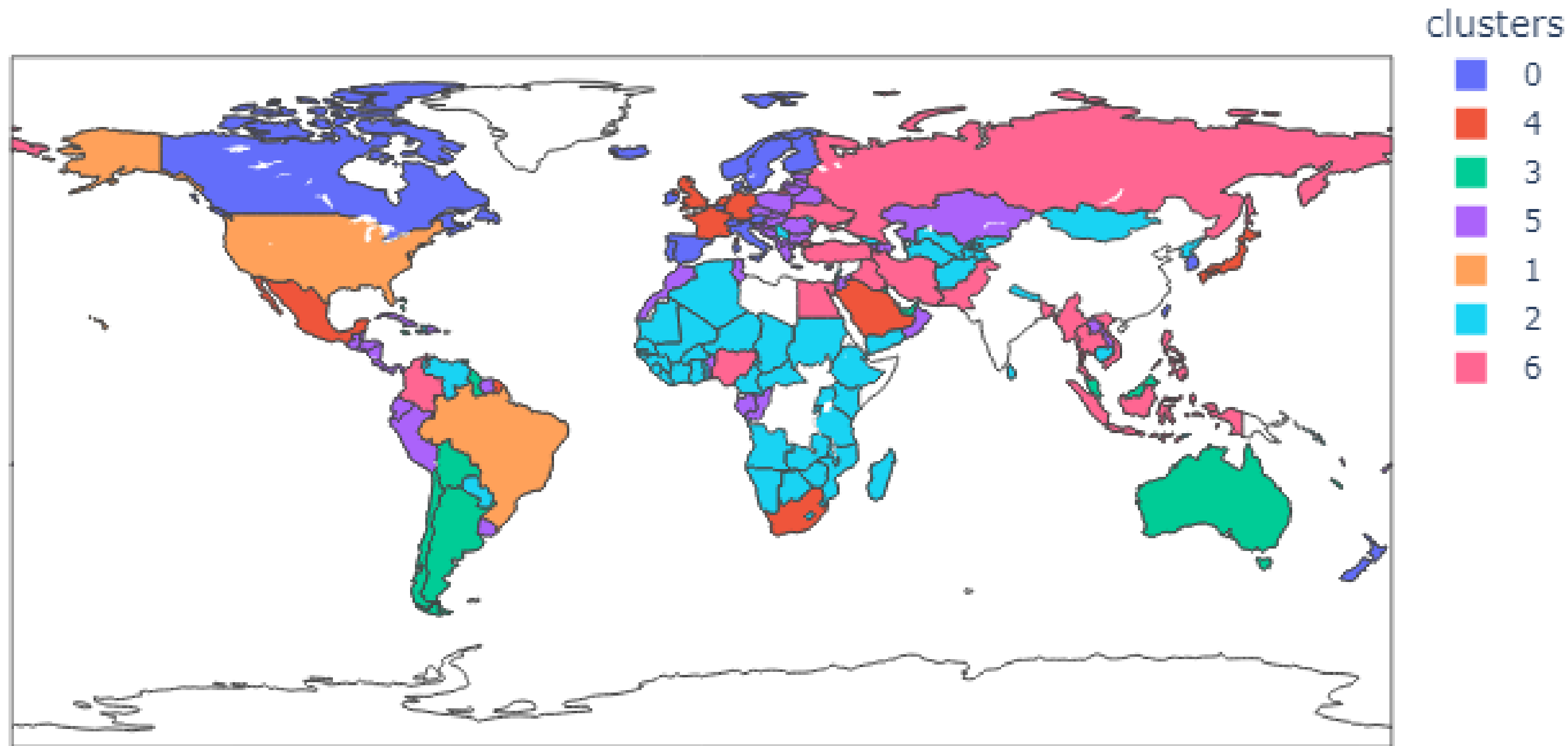


Analyse des clusters suivant le PIB/habitant



Recommandations

Repérage géographique des groupements de pays à cibler par ordre de potentiel.



Potentiel fort :

- Cluster 0
- Cluster 4

Potentiel moyen :

- Cluster 3
- Cluster 5
- Cluster 1

Potentiel faible :

- Cluster 2
- Cluster 6

Recommandations : cibler les pays stables avec une population élevée ayant des revenus conséquents, où la consommation et la disponibilité alimentaire liées au poulet sont faibles.