

# ÉTUDE DE MARCHÉ — DATA INTERNATIONAL

Jérémy GUINAULT  
La Poule Qui Chante

# Sommaire

- *Présentation du contexte,*
- *Données initiales,*
- *Préparation et nettoyages des données,*
- *Présentation des variables,*
- *Analyse et exploration des données,*
- *Conclusion et recommandations.*

# Présentation du contexte

- **La Poule Qui Chante**, est une entreprise française d'agroalimentaire.
- Notre activité principale est l'élevage et la vente de poulets sous le label « **Poulet Agriculture Biologique** ».
- A l'heure actuelle nous avons une activité uniquement franco-française.
- L'objectif de l'étude est d'évaluer la possibilité de se développer à l'international.
- Aucun pays n'est ciblé en particulier pour le moment, nous ne fermons aucune porte.

# Données initiales

- 2 fichiers de bases :
  - Fichier de disponibilité alimentaire, regroupant les types d'aliment par pays, et leurs utilisations,
  - Fichier de population, présentant la population pour chaque pays sur plusieurs années.
- D'autres fichiers ont été rajouté pour compléter l'étude, comme un fichier avec le PIB par habitant, la stabilité politique par pays, grâce à des données en open data.

# Préparation et nettoyage des données

- Les fichiers de l'étude sont basés sur l'année 2017.
- Analyse des différents dataframes ajoutés :
  - Analyse de la taille des dataframes, et du type de données de chacun,
  - Vérification si présence de valeurs nulles, et traitement de celles-ci,
  - Suppression des colonnes superflues ou redondantes, pour ne conserver que les variables,
  - Modification du nom des colonnes pour que la variable soit claire,
  - Harmonisation du nom des pays entre le français et l'anglais,
  - Jointure des différents fichiers pour créer un dataframe final,
  - Création de nouvelles variables calculées.

# Présentation des variables

- 145 pays composent l'étude, avec chacun 12 variables, dont 3 colonnes calculées :
  - Population,
  - Volailles alimentaires en tonnes,
  - Volailles produites en tonnes,
  - Volailles importées en tonnes,
  - Volailles exportées en tonnes,
  - **Balance commerciale,**
  - Pib par habitant en \$,
  - Restaurants KFC,
  - Stabilité politique,
  - **Kg de volailles par habitant,**
  - **Déficit en volailles par habitant,**

# Analyse exploratoire des données

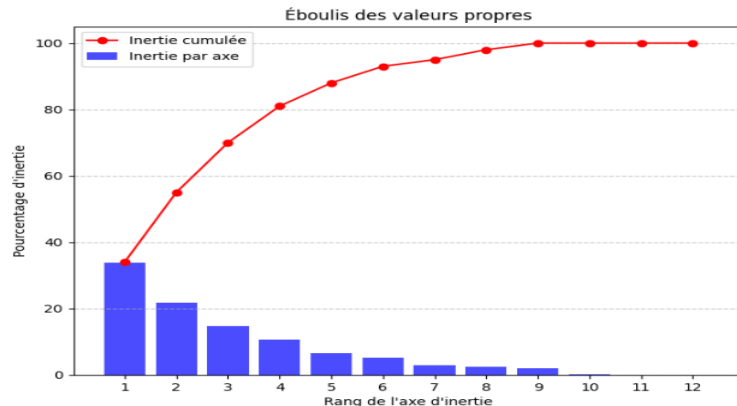
- Dataframe de 145 pays, qui couvre 88% de la population mondiale en 2017.
- L'analyse du dataframe est faite en deux temps : on va isoler dans un dataframe distinct les 4 pays que sont : les USA, la Chine, l'Inde et le Brésil, car pour certaines variables ils comportent des outliers.
- Pour le dataframe « principal », composé ,au final, de 141 pays, la méthode sera la suivante :
  - Standardiser les données,
  - Réaliser une ACP et déterminer le nombre de composantes principales nécessaires,
  - Faire une projection via ces composantes,
  - Création de clusters regroupant les pays ayant des similitudes, via CAH et K-Means,
  - Choisir le ou les clusters pertinents.

# Analyse exploratoire des données

- **Standardisation des données** : transformer les valeurs d'une variable pour qu'elles aient une distribution avec une moyenne de 0 et un écart-type de 1, avec *StandardScaler*.

	0	1	2	3	4	5	6	7	8	9	10	11
mean	-0.0	-0.0	-0.0	0.0	0.0	0.0	-0.0	0.0	0.0	0.0	-0.0	0.0
std	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

- Réalisation de l'ACP et détermination du nombre de composantes nécessaires :

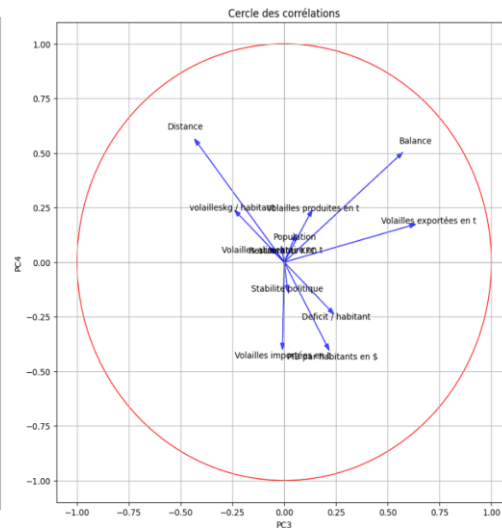
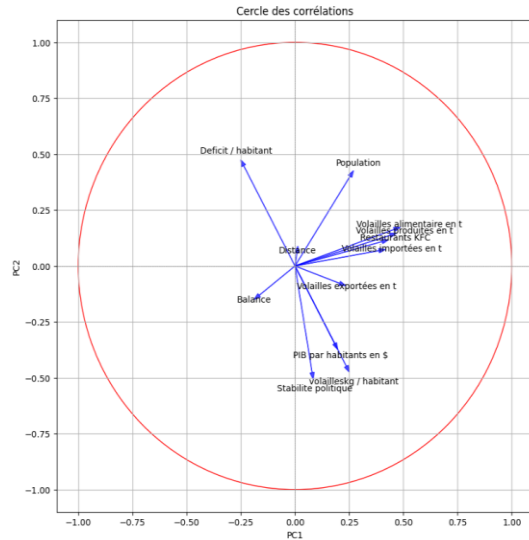


**4 composantes retenues pour expliquer 80% de la variance**



# Analyse exploratoire des données

- Cercle des corrélations pour les 4 composantes :

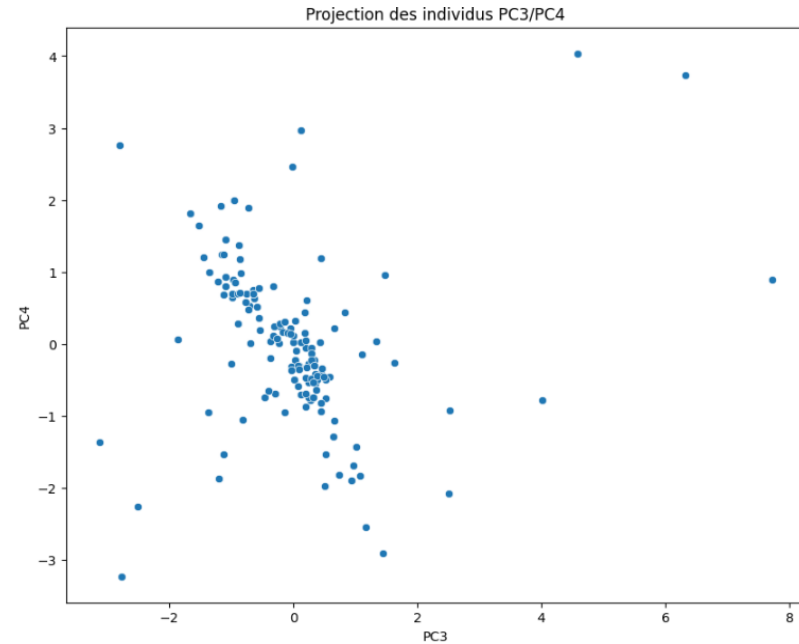
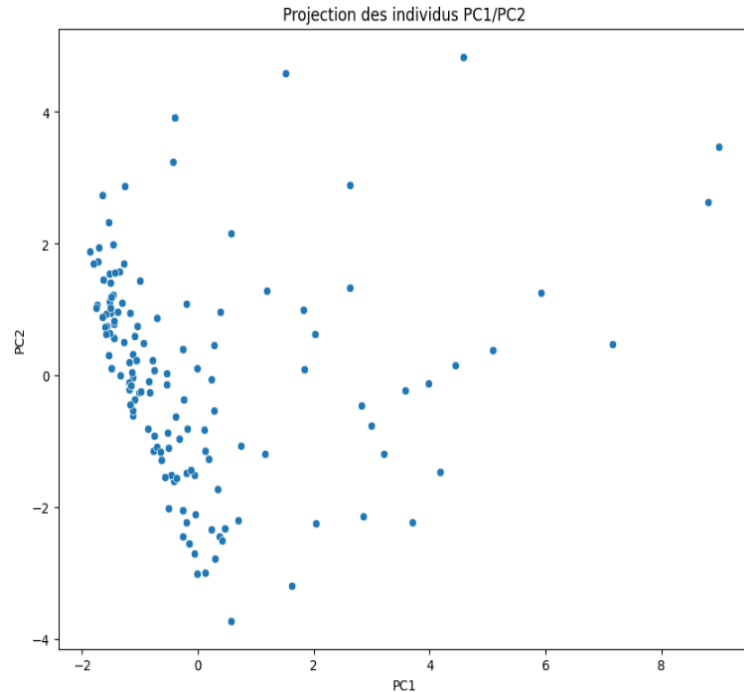


- Extrait du dataframe avec les données standardiser et les 4 composantes

	PC1	PC2	PC3	PC4	Zone
0	0.320426	-0.479314	-0.526388	-0.310917	Afghanistan
1	0.840410	2.779533	2.089916	2.287723	Afrique du Sud
2	-0.518329	-0.492481	-0.550332	-0.262695	Albanie
3	0.448280	-0.135324	-0.132106	-0.455584	Algérie
4	1.484279	2.078387	1.845685	4.045153	Allemagne

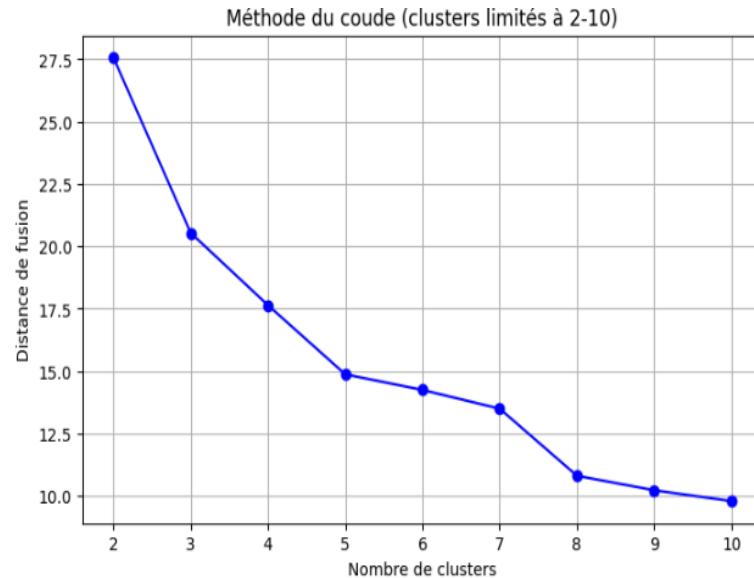
# Analyse exploratoire des données

- Projection des individus avec les 4 composantes

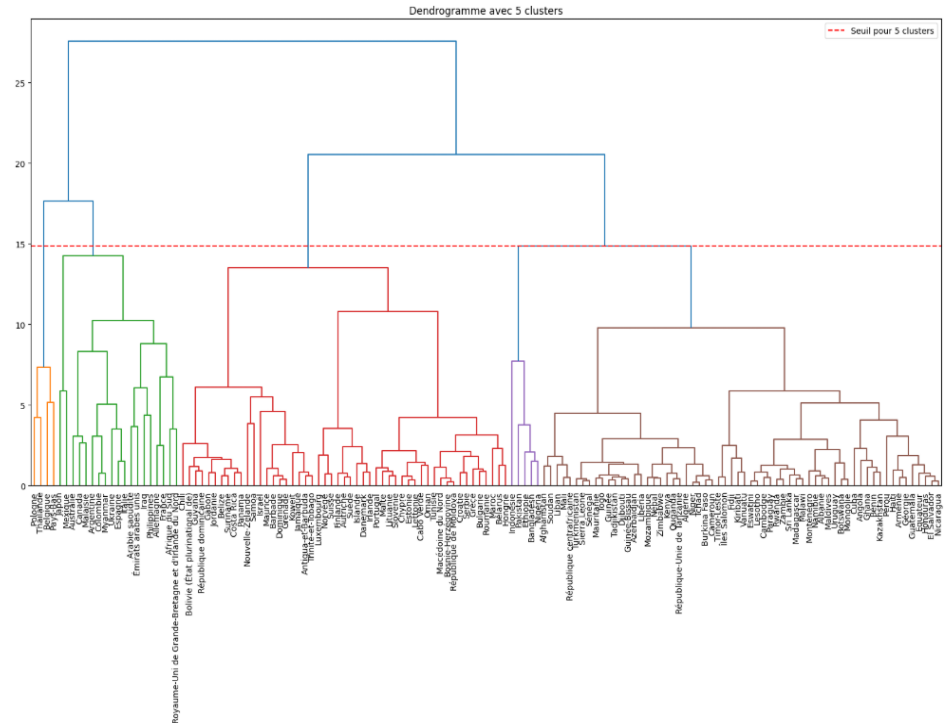


# Analyse exploratoire des données : CAH

## Détermination du nombre de clusters

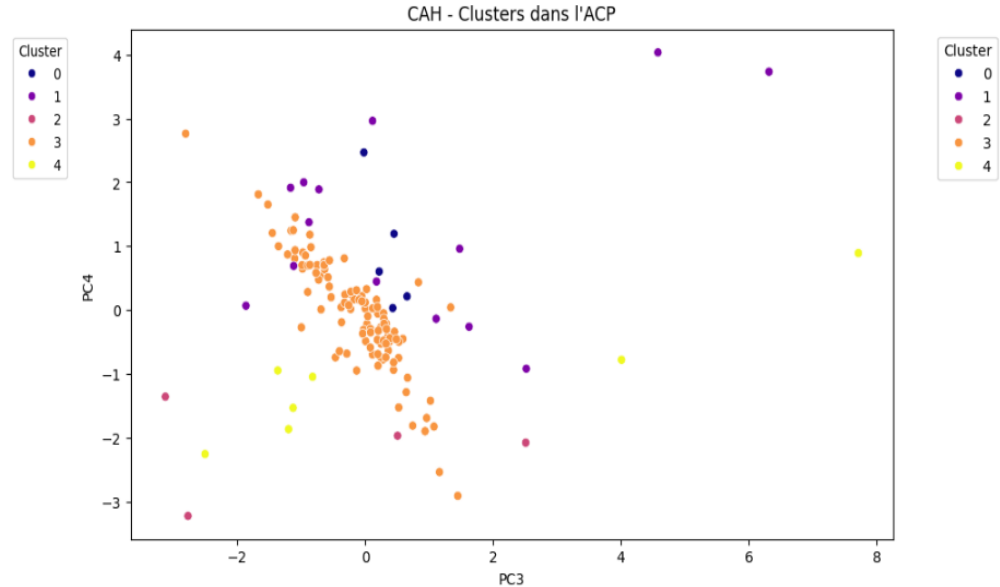
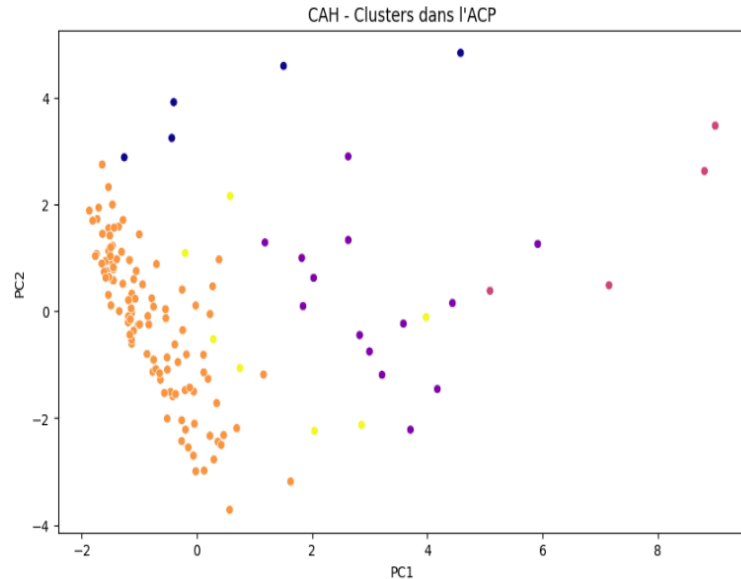


## Dendrogramme



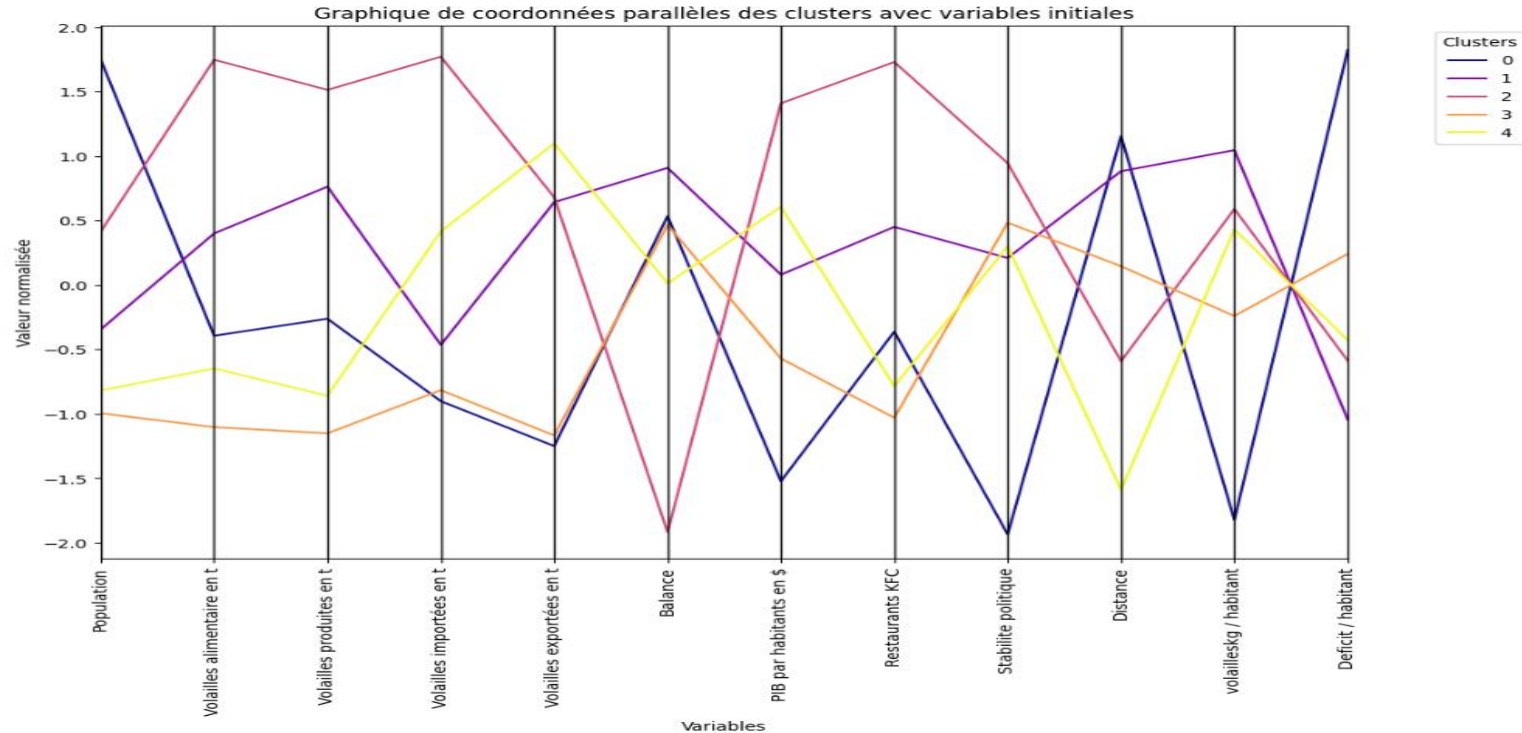
# Analyse exploratoire des données : CAH

- Cela nous a permis de générer 5 clusters avec la répartition suivante :



Cluster 0 : 5 pays, Cluster 1 : 14 pays, Cluster 2 : 4 pays, Cluster 3 : 111 pays, Cluster 4 : 7 pays.

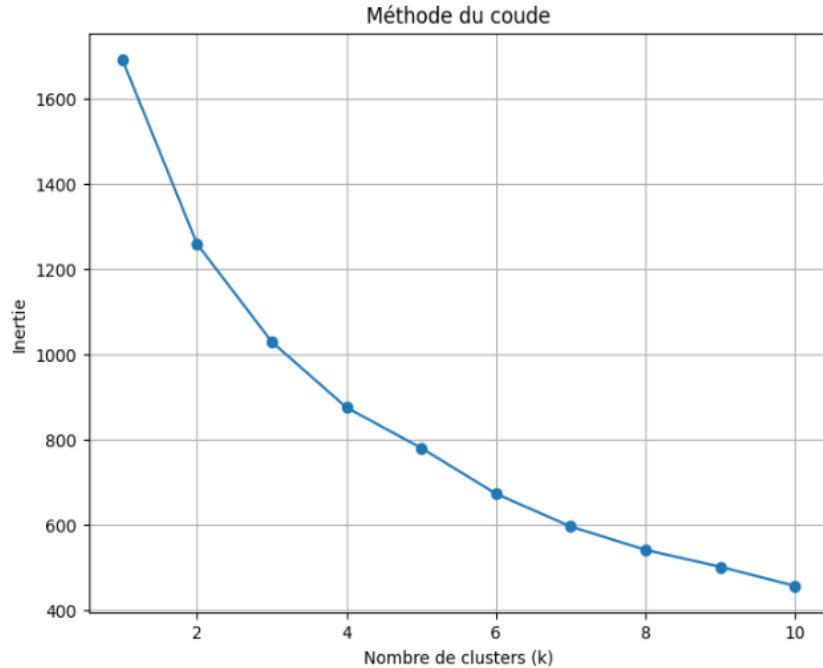
# Analyse exploratoire des données : CAH



Le cluster 2 se dégage favorablement par rapport aux variables choisies.

# Analyse exploratoire des données : K-Means

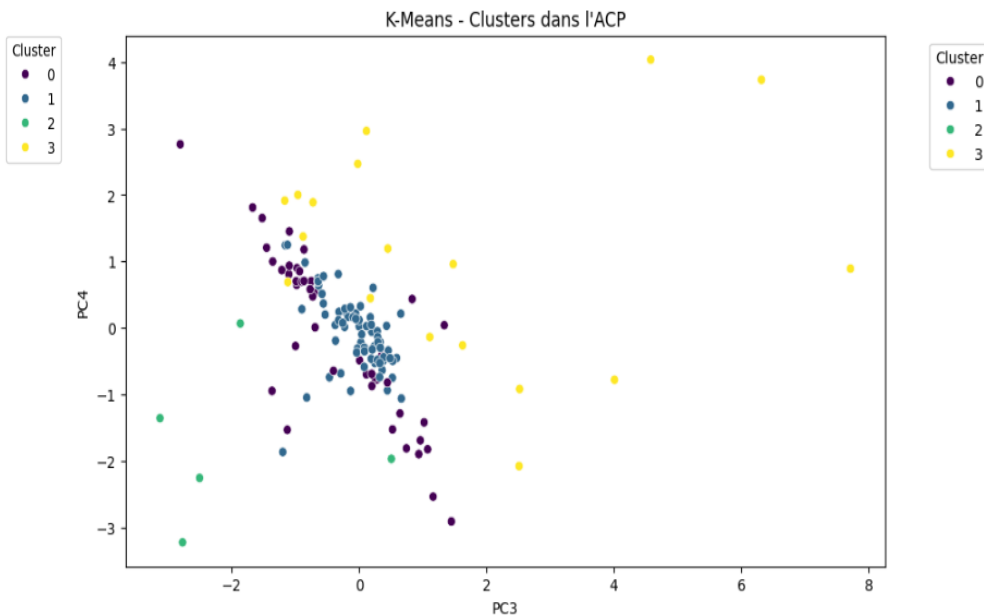
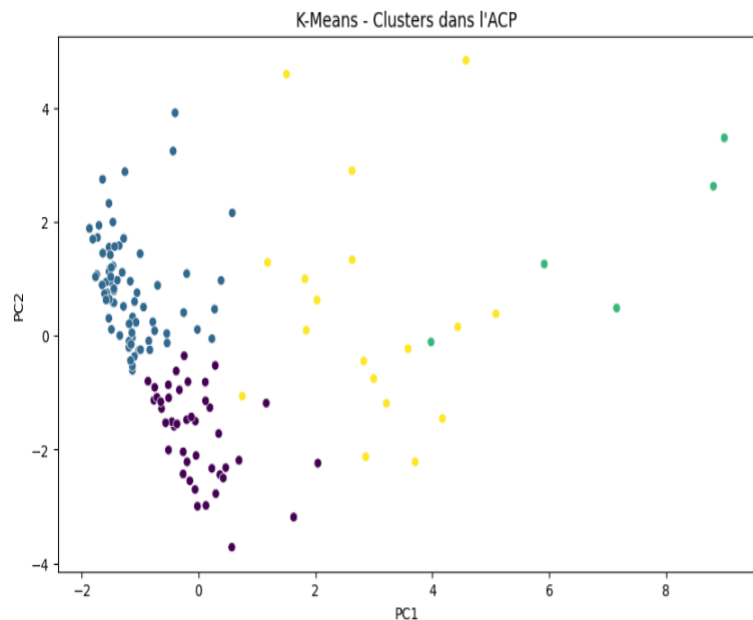
- Détermination du nombre de clusters :



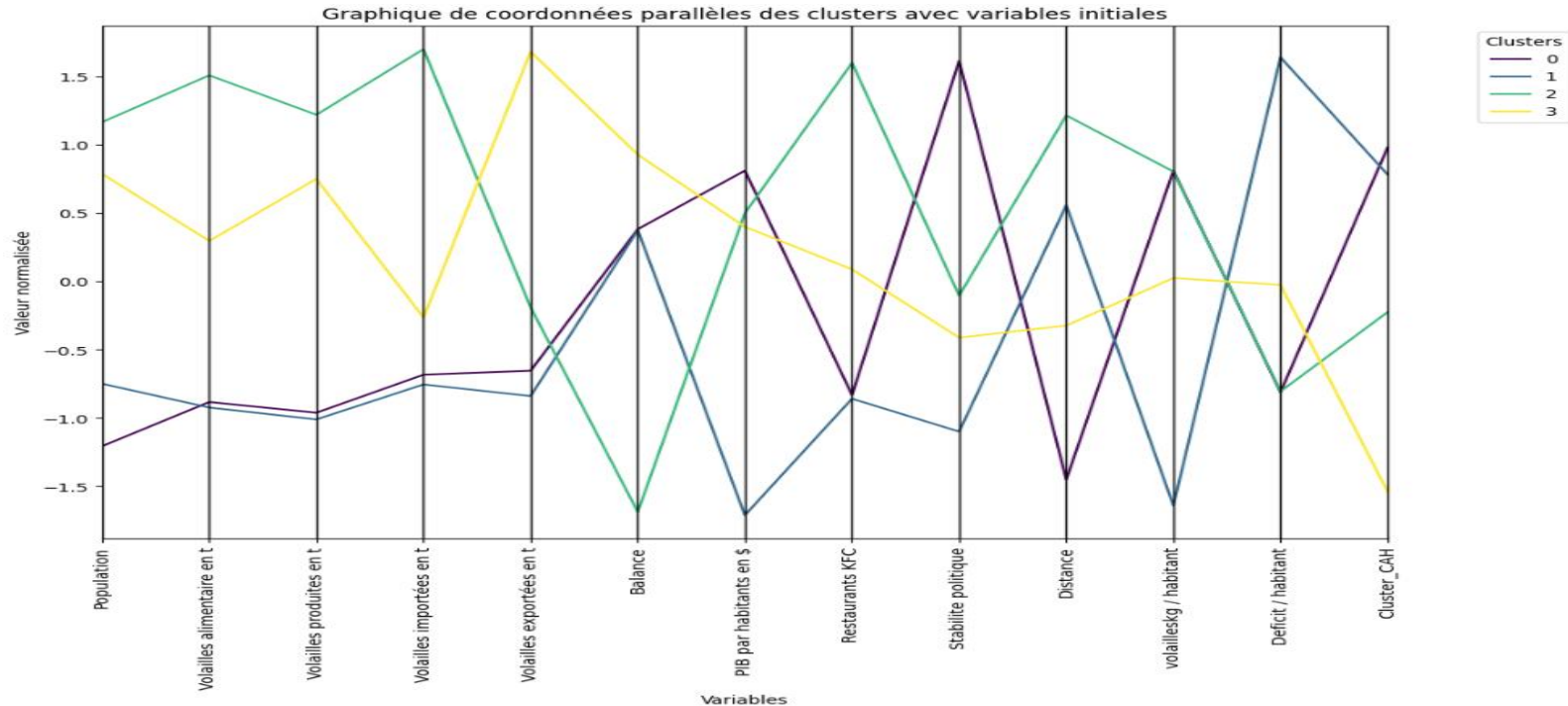
- 4 clusters choisis,
- Application de l'algorithme K-Means,
- Répartition des clusters :
  - Cluster 0 : 44 pays,
  - Cluster 1 : 74 pays,
  - Cluster 2 : 5 pays,
  - Cluster 3 : 18 pays.

# Analyse exploratoire des données : K-Means

Répartition des individus par clusters :



# Analyse exploratoire des données : K-Means

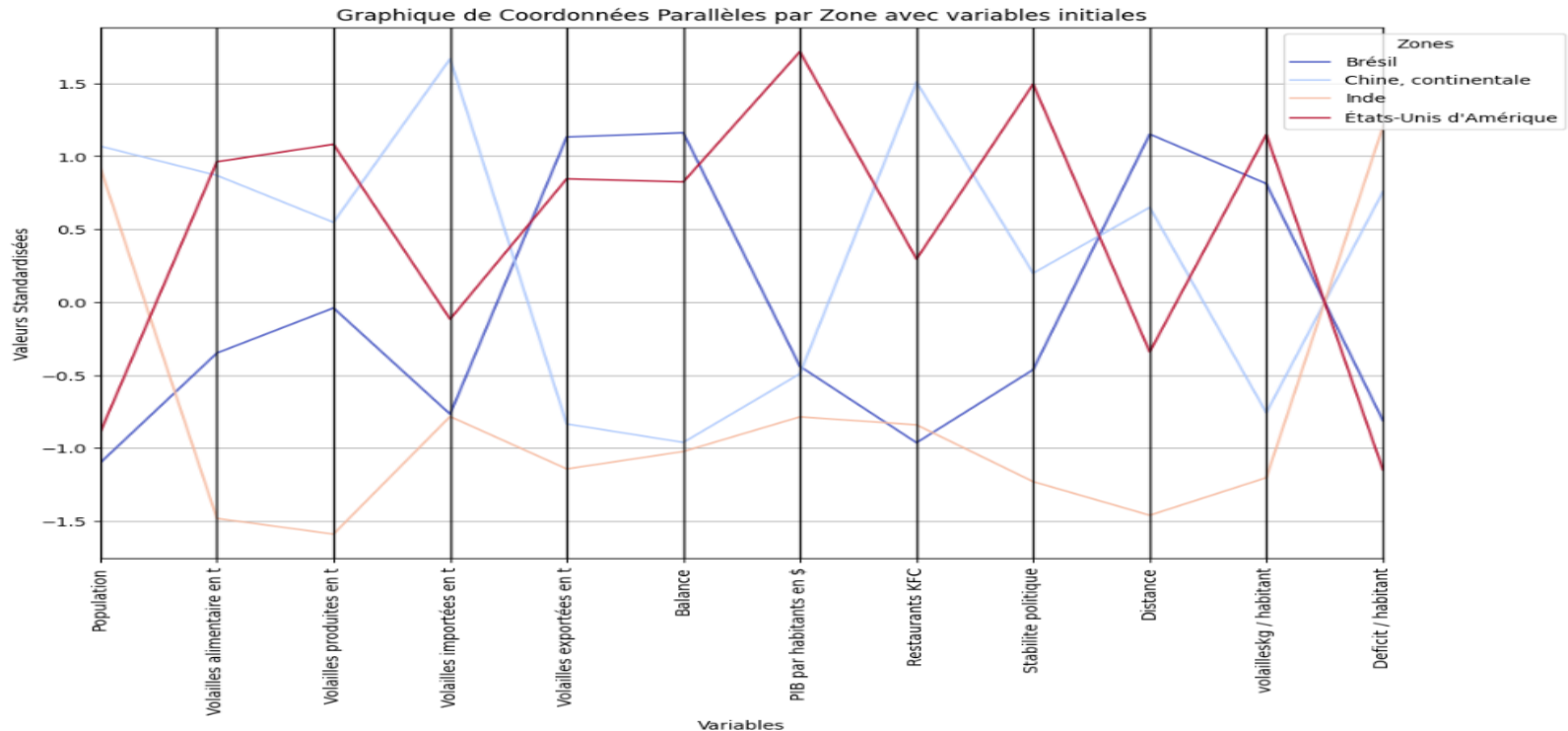


Le cluster 2 se dégage favorablement par rapport aux variables choisies.



# Analyse exploratoire des données

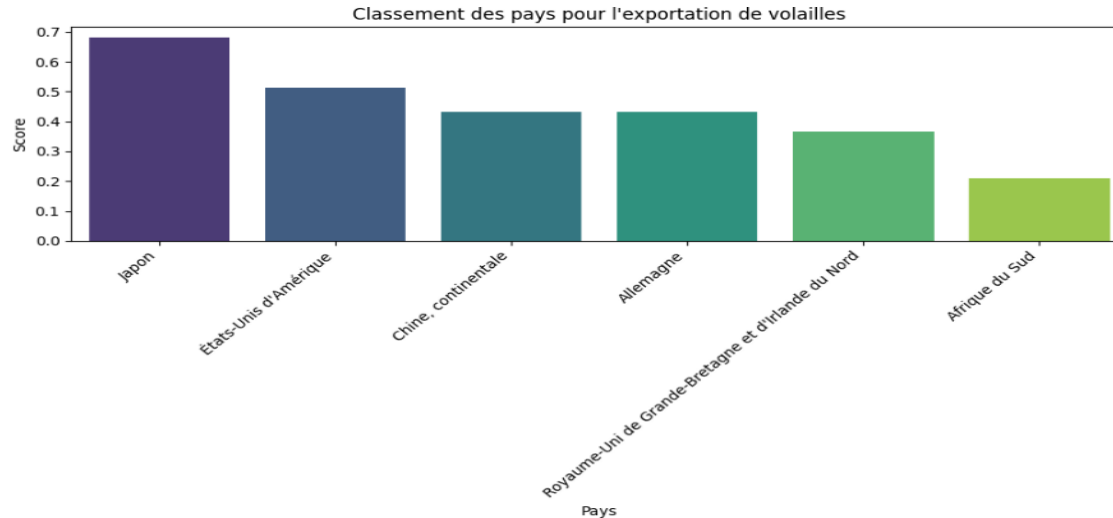
Pour les 4 pays isolés au début de l'analyse :



# Analyse exploratoire des données

Création d'un dataframe final avec le cluster 2 de la CAH et le cluster 2 du K-Means, et les 4 pays isolés. Puis filtre pour ne conserver que les pays qui tendent vers une stabilité politique ( $> 0,3$ ).

Résultat, un dataframe de 6 pays. Si l'on cherche à les classer avec un score en fonction de variables sélectionnées (population, importation, balance, distance...) :



# Conclusion et recommandations

- Par rapport à nos variables initiales, 6 pays ressortent de l'analyse : Japon, USA, Chine, Allemagne, Royaume-Uni et Afrique du Sud.
- En classant les pays via un score pondéré se basant sur les variables, le Japon ressort en tête. Si le Japon s'avère être une solution viable il serait pertinent de s'intéresser également à la Chine.
- Nous pouvons également nous diriger vers les pays les plus proches pour minimiser les risques et les coûts, dans ce cas l'Allemagne et le Royaume-Uni seraient une bonne opportunité.
- Les USA et l'Afrique du Sud peuvent être des solutions intéressantes dans un second temps, mais le marché est peut-être plus saturé, compte tenu du disponible par habitant.