



La poule qui chante

Etude de marché sous Python

Etude sur l'exportation de Poulets sous le label "Poulet Agriculture Biologique"

Jean-Robert Vives

Contexte et objectif de la mission

DATA ANALYST

pour la société “La poule qui chante”

Produit : Vente de poulets BIO en France



OBJECTIF

Développement à l'international



Etude de marché

TOP 5 des pays potentiels





MÉTHODOLOGIE

1. Création de notre fichier de données (FAO+PESTEL)
2. Nettoyage et chargement des données (Colaboratory)
3. Analyse générale du marché
 - 3.1 Etat général de la production de poulet
 - 3.2 Distribution des importations
 - 3.3 Distribution des exportations
 - 3.3 Distribution de la production de poulet
 - 3.4 Matrice de corrélation
4. Analyse en Composantes Principales ACP
5. Classification Ascendante Hiérarchique CAH
6. K-means
7. Conclusion

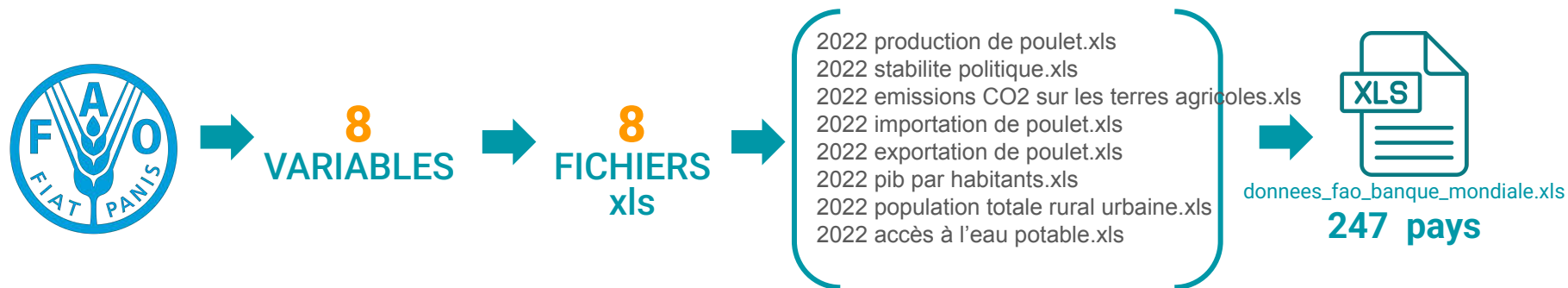
1. Création de notre fichier de données (FAO+PESTEL)



« Aucune variable technologique quantifiable n'était disponible dans les bases de données utilisées (FAO, Banque Mondiale).

Cependant, il serait intéressant, dans une étude complémentaire, d'intégrer des données sur la logistique, la traçabilité ou l'innovation dans la filière. »

1. Création de notre fichier de données (FAO+PESTEL)



2. Nettoyage et chargement des données (Colaboratory)

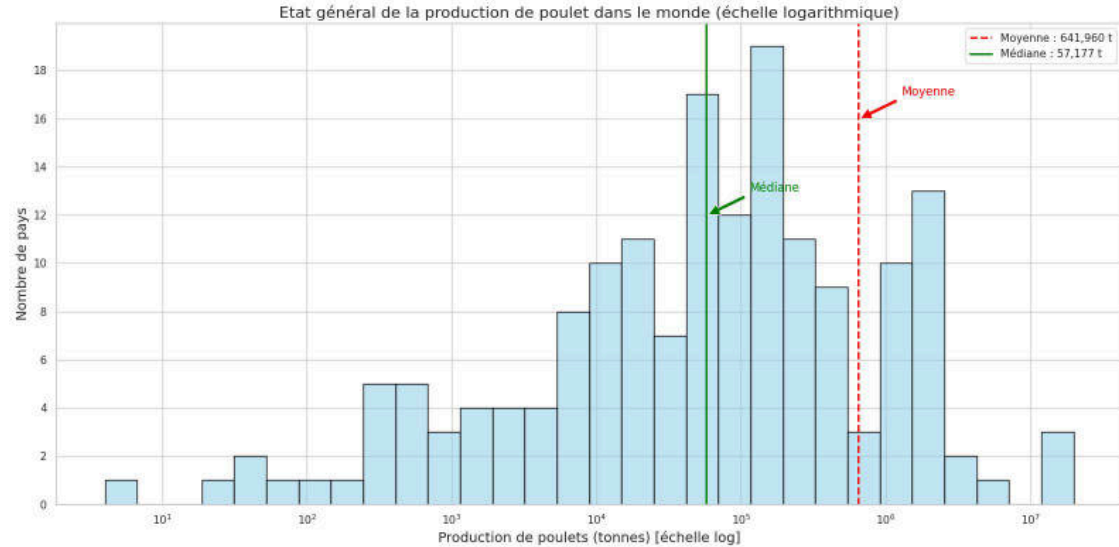


3. Analyse générale du marché

3.1 Etat général de la production de poulet dans le monde

Ce graphique confirme la forte concentration de la production mondiale de poulet

La moyenne (638 393 t) placée à l'extrême droite du graphique est bien supérieure à la médiane (57 049 t), ce qui renforce l'idée que quelques pays produisent des volumes massifs, tandis que la majorité des pays ont une production limitée.

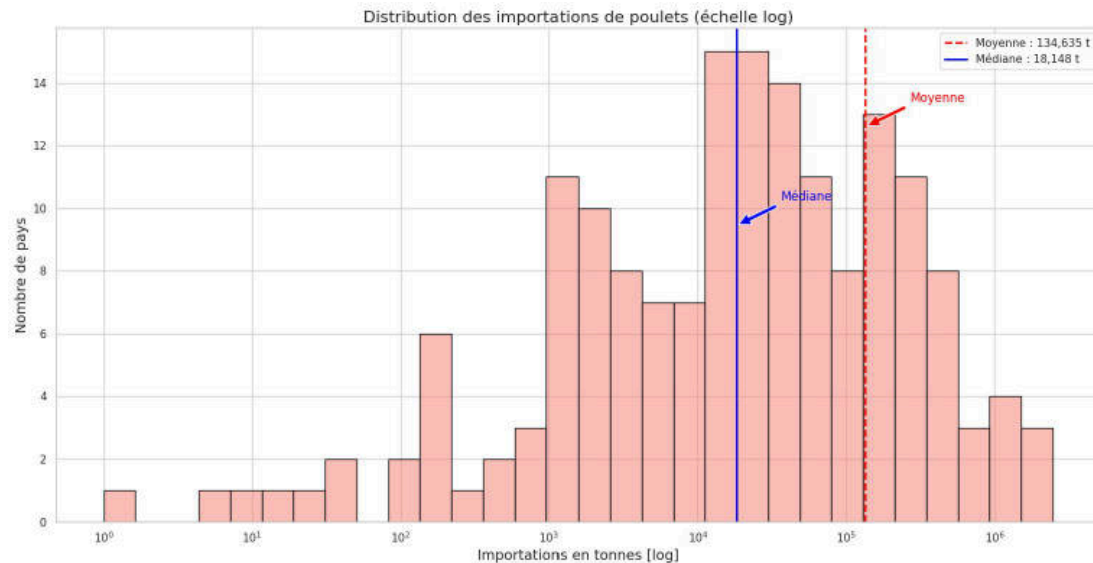


3. Analyse générale du marché

3.2 Distribution des importations

La valeur médiane est de 18 148 tonnes
ce qui explique qu'une grande majorité de pays
exportent des volumes très faibles.

La moyenne s'élève à 134 635 tonnes ? Positionnée à
l'extrême gauche du graphique et indiquant **qu'un nombre
restreint de pays est responsable de la majeure partie des
importations mondiales de poulet.**

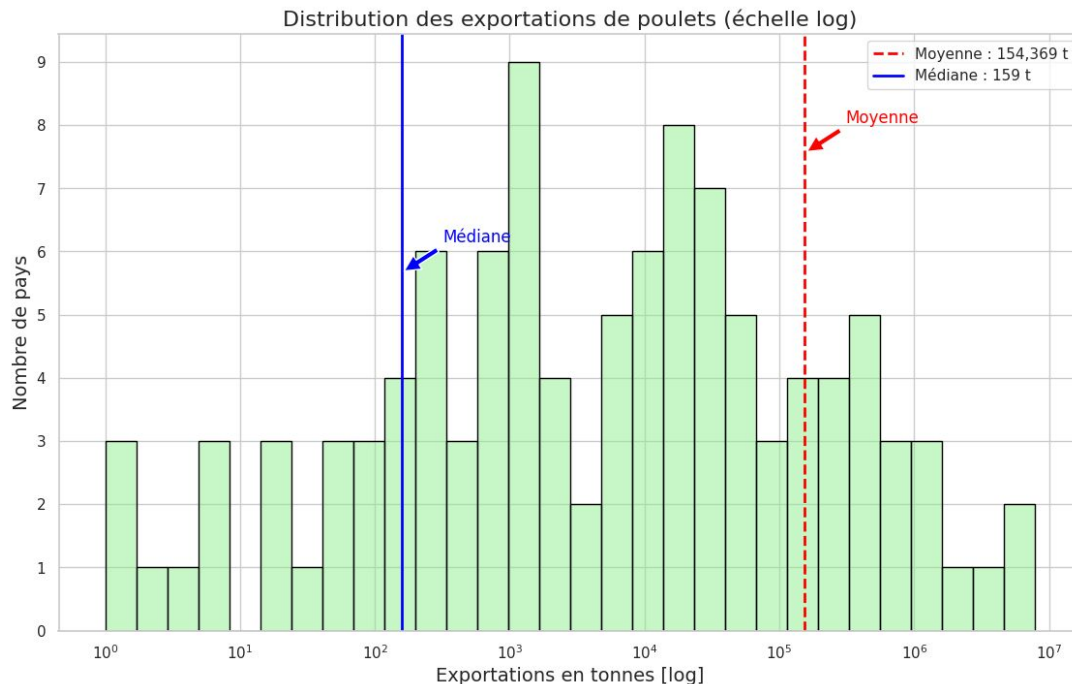


3. Analyse générale du marché

3.3 Distribution des exportations

La valeur médiane est très faible 159 tonnes
ce qui explique qu'une grande majorité de pays
exportent des volumes très faibles.

La moyenne s'élève à un niveau considérablement plus élevé,
atteignant **154 369 tonnes** ? Positionnée à l'extrême gauche du
graphique et indiquant qu'**une minorité de pays dominant**
massivement le marché de l'exportation de poulet



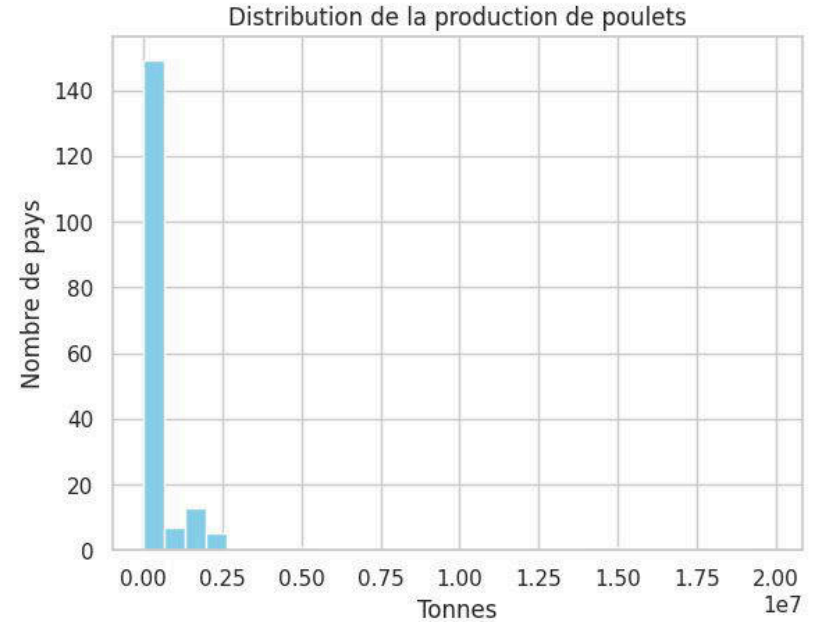
3. Analyse générale du marché

3.3 Distribution de la production de poulet

Ce graphique montre que **la production mondiale de poulet est fortement concentrée.**

Une grande majorité de pays produisent des volumes très faibles, tandis qu'un très petit nombre de nations sont responsables de la majeure partie de la production globale.

Cela indique une structure de **marché dominée par quelques acteurs majeurs.**



3. Analyse générale du marché

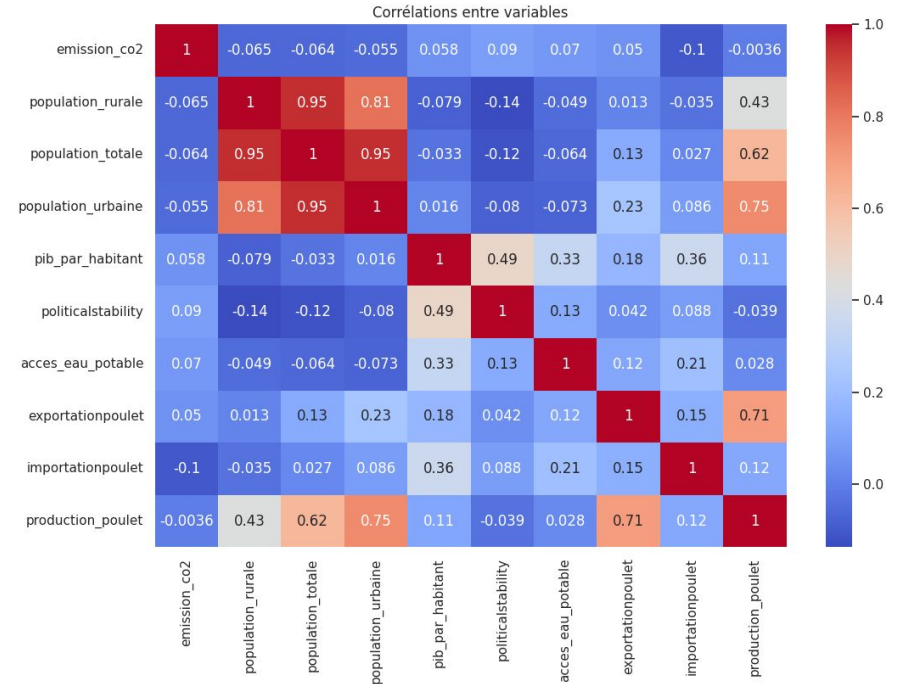
3.4 Matrice de corrélation

Cette matrice met en évidence les relations clés suivantes :

La production de poulet est fortement corrélée à la taille de la population (0,75), indiquant une **production axée sur la demande intérieure**.

Les exportations sont fortement liées à la production (0,71), confirmant que **les grands producteurs sont les grands exportateurs**.

Les importations ont une corrélation modérée avec le PIB par habitant (0,36), suggérant que **les marchés plus riches sont plus susceptibles d'importer**, y compris des produits à valeur ajoutée.



4. Analyse en Composantes Principales ACP

Pour effectuer cette opération de standardisation nous allons utiliser la bibliothèque "SCIKIT-LEARN" et l'objet "StandardScaler" qui va automatiser le calcul pour l'ensemble des données de mon dataframe

Etape 1 Standardisation des données

1 - On soustrait la valeur moyenne à chaque variable

Exemple :

Population_totale de l'Afghanistan = 40578,842
la moyenne est de 42131,8
J'obtiens $40578,842 - 42131,8 = -1552,958$

2 - on divise ce chiffre par l'écart-type de cette variable

Exemple : 'Afghanistan' = 1552,958

l'écart type est de 15679,44

J'obtiens $-1552,958 / 15679,44 = -0,0099$

Résumé statistique complet des variables sélectionnées (X.describe()) :

	pib_par_habitant	population_totale	exportationpoulet	importationpoulet	acces_eau_potable	production_poulet	politicalstability	emission_co2
count	180.000000	1.800000e+02	1.800000e+02	1.800000e+02	180.000000	1.800000e+02	180.000000	180.000000
mean	22765.074444	4.213180e+04	1.543685e+05	1.346346e+05	78.622222	6.383935e+05	-0.105611	2.100111
std	26042.388482	1.567944e+05	7.461188e+05	3.338995e+05	31.483286	2.224580e+06	0.953741	3.245590
min	0.000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.000000e+00	-2.550000	0.000000
25%	3608.350000	1.831061e+03	0.000000e+00	1.683250e+03	70.750000	6.299055e+03	-0.617500	0.540000
50%	13512.950000	8.534495e+03	1.590000e+02	1.814850e+04	95.000000	5.704919e+04	0.000000	1.000000
75%	33895.525000	3.040557e+04	1.470600e+04	1.075765e+05	99.000000	2.394528e+05	0.655000	2.232500
max	137059.000000	1.456770e+06	7.869511e+06	2.464589e+06	99.000000	1.984124e+07	1.590000	19.850000

4. Analyse en Composantes Principales ACP

Etape 2 Obtention du tableau standardisé

J'ai donc un dataframe de 8 dimensions.
que l'on nommera Composantes Principales.

Aperçu des 5 premières lignes du DataFrame standardisé (df_standardized) :

	pib_par_habitant	population_totale	exportationpoulet	importationpoulet	acces_eau_potable	production_poulet	politicalstability	emission_co2
0	-0.325098	0.129489	-0.050103	0.667958	0.489805	0.591693	-0.645985	-0.392426
1	-0.876593	-0.009932	-0.207473	-0.199578	0.107587	-0.274924	-2.570097	-0.528373
2	-0.211940	-0.251373	-0.207473	-0.296866	0.521657	-0.280409	0.226699	-0.355350
3	-0.372514	0.021397	-0.206529	-0.404341	0.521657	-0.103971	-0.667014	-0.546911
4	1.534097	0.268322	1.494561	5.578311	0.649062	0.197263	0.752413	-0.302825

4. Analyse en Composantes Principales ACP

Etape 3 Réduire la dimensionnalité

Moins de dimensions, c'est plus simple à comprendre
je décide donc d'en calculer 5 qui est une valeur de base (et inférieure)
UTILISATION DE LA BIBLIOTHEQUES “**SCIKIT-LEARN**”

```
📊 Variance expliquée par composante : [26.8 22.5 13.7 11.4  9.9]  
100 Variance cumulée : [26.8 49.3 63.  74.4 84.3]
```

La variance cumulée m'indique je suis arrivé en créant ses 5 composantes principales à capturer 84,3 % de l'information.

Sachant qu'un seul de 80 à 90 % est considéré comme très bon, je reste à 5

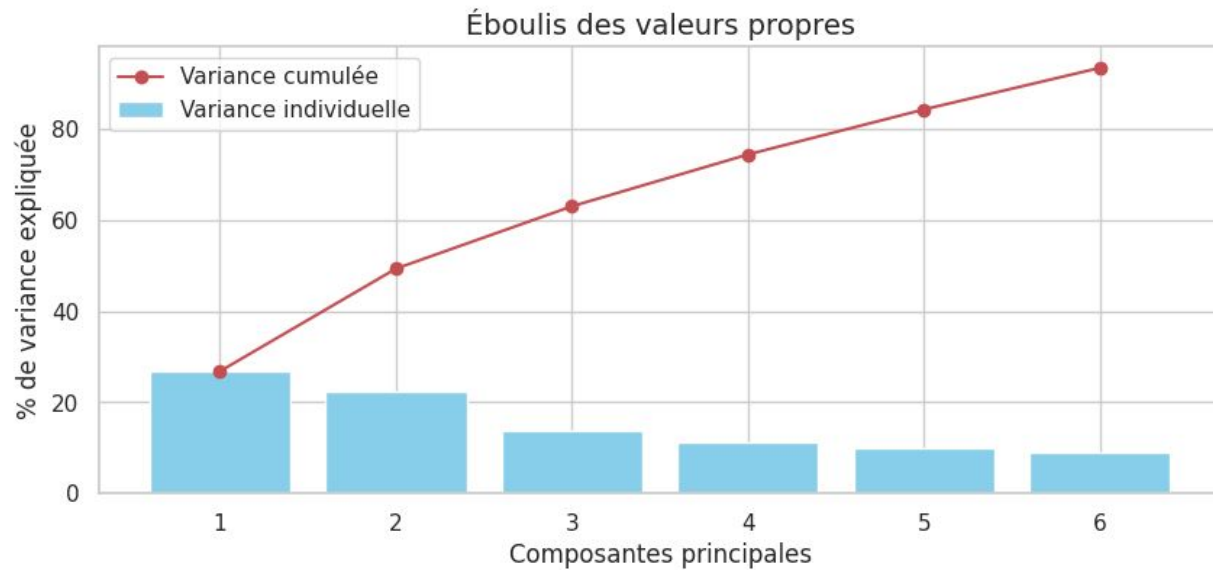


4. Analyse en Composantes Principales ACP

Etape 4 Visualisation de l'éboulis

Volontairement je choisis de visualiser l'éboulis pour 6 composantes principales.
Ce graphique va permettre de m'assurer dans le fait que j'ai bien fait de ne choisir que 5 niveaux pour le calcul de mes CP

Concernant la courbe de variance cumulée, j'observe un point d'inflexion léger au CP6 et surtout un gain qui n'est plus notable, **ce qui me rassure sur le fait d'avoir choisi de m'arrêter au CP5.**



4. Analyse en Composantes Principales ACP

Etape 5 Visualisation du cercle de corrélation

Quadrant en bas à droite :

Les flèches sont proches : ces variables vont de pair donc les pays très producteurs de poulet sont souvent aussi de gros exportateurs et ont une grande population.

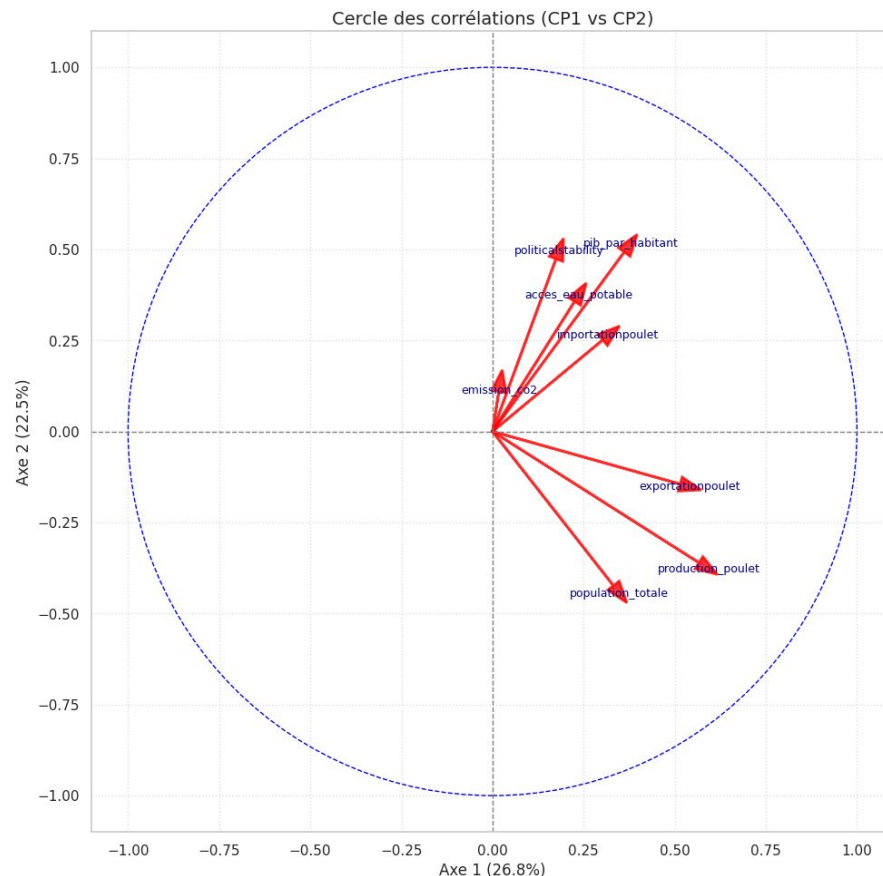
Conclusion : Ce quadrant nous renseigne sur l'existence de "poids lourds" du marché mondial du poulet

Quadrant en haut à droite

importationpoulet :

Le vecteur importationpoulet n'est PAS confondu avec population ou exportation : il forme un angle avec eux (environ 60 à 90°, soit une corrélation faible à nulle). Il est relié modérément à : PIB/habitant et à stabilité politique.

Conclusion : Les gros importateurs ne sont pas les mêmes pays que les très gros producteurs/exportateurs. Les importateurs sont plus proches des pays riches, stables et à bonne infrastructure.



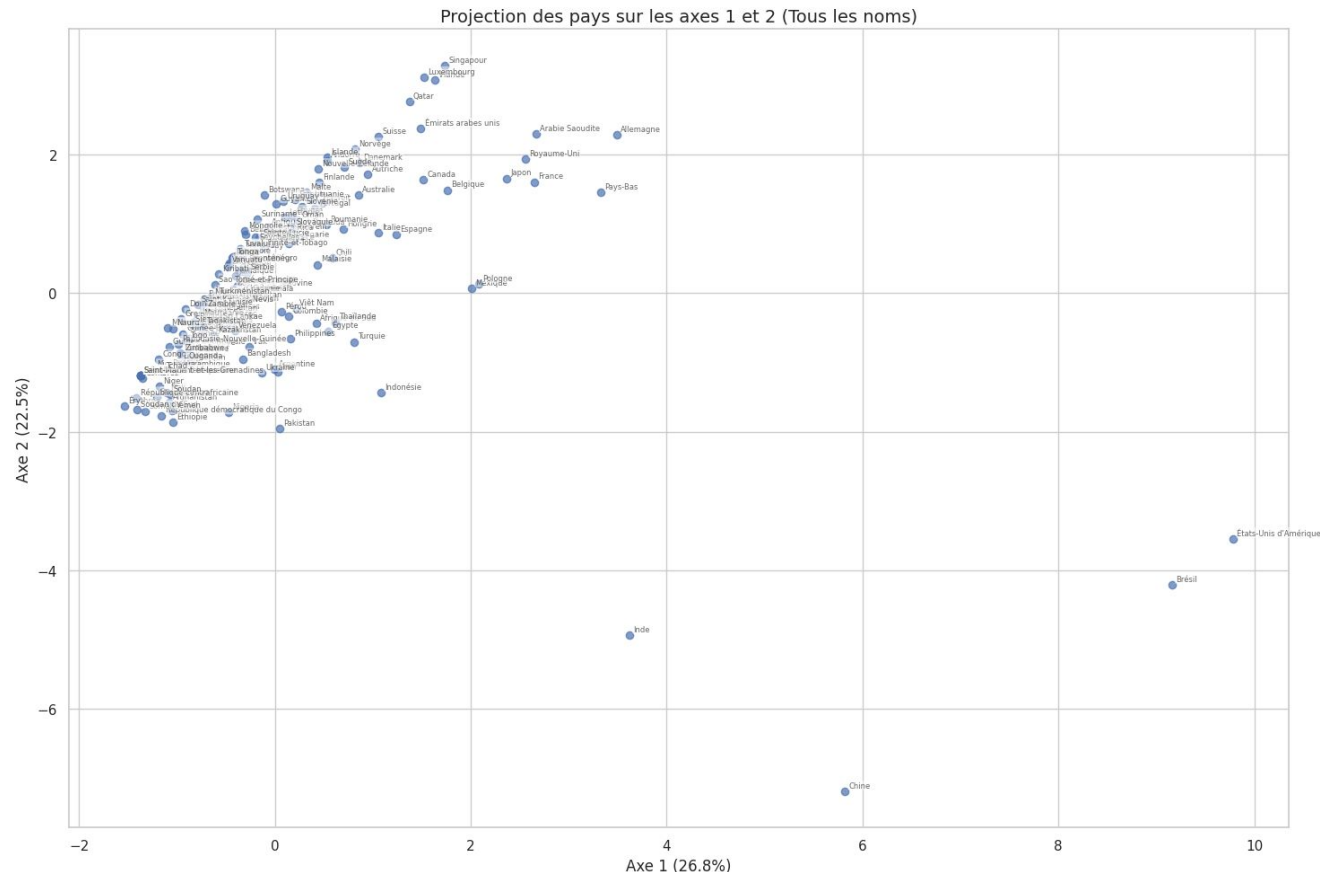
4. Analyse en Composantes Principales ACP

Etape 6 Projection des pays

La projection des pays montre
que **les pays les plus intéressants**
sont en effet situés dans le
quadrant Haut/ Droit

De même, on retrouve nos poids lourds du marché mondial de la production de poulet :

INDE
CHINE
ETATS-UNIS
BRÉSIL



4. Analyse en Composantes Principales ACP

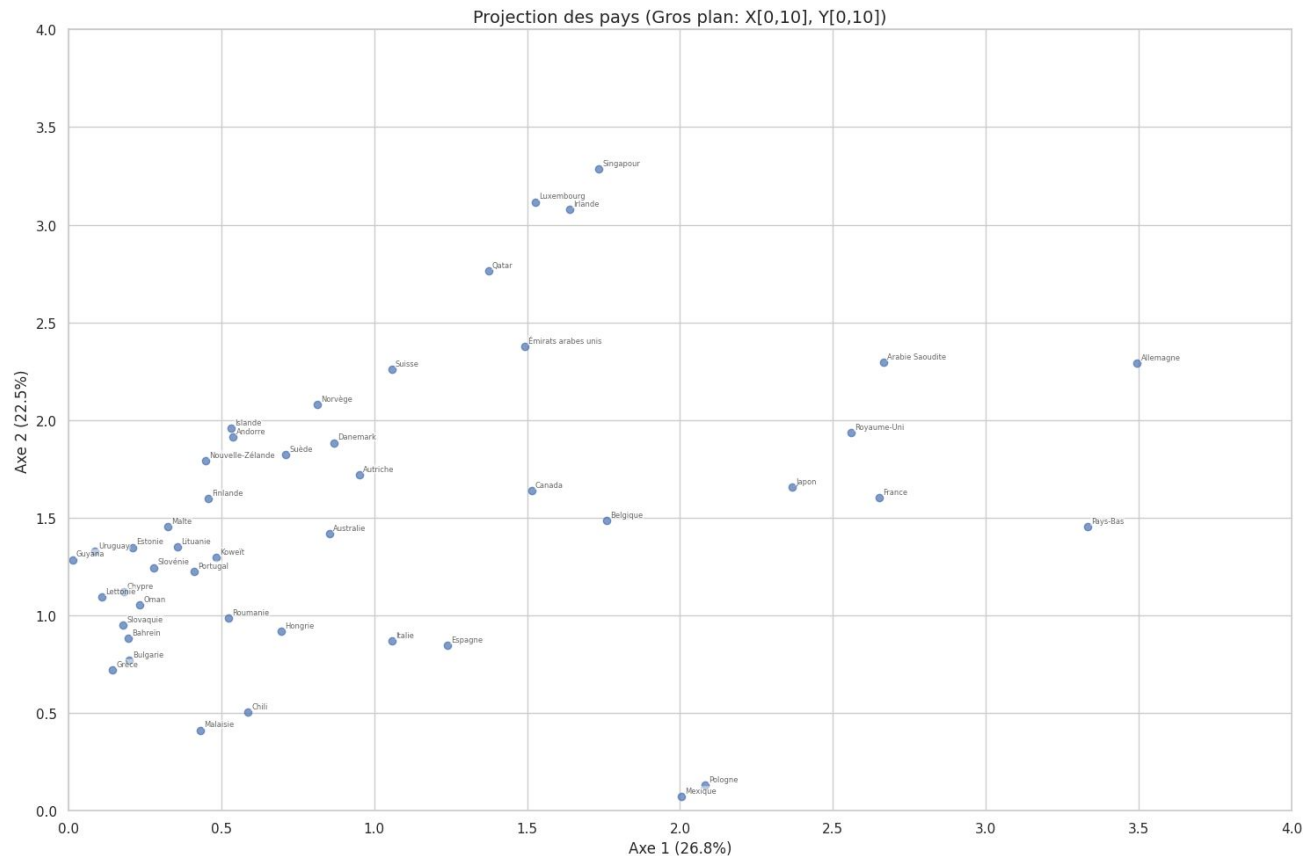
Etape 6 Projection des pays Quadrant haut droit

Plus un pays est en haut à droite (grandes valeurs sur Axe1 ET Axe2), plus il est censé être un excellent candidat (selon la logique du cercle et du nuage des pays).

Allemagne, Royaume-Uni, Arabie Saoudite, Pays-Bas, France, Japon, Suisse, Luxembourg, Belgique, Qatar, Singapour...

Ce sont les cibles idéales

Profil : ils se caractérisent par une forte stabilité politique, un haut niveau de vie, d'importantes importations de poulet et des exigences de qualité (BIO, sécurité sanitaire, traçabilité).



4. Analyse en Composantes Principales ACP

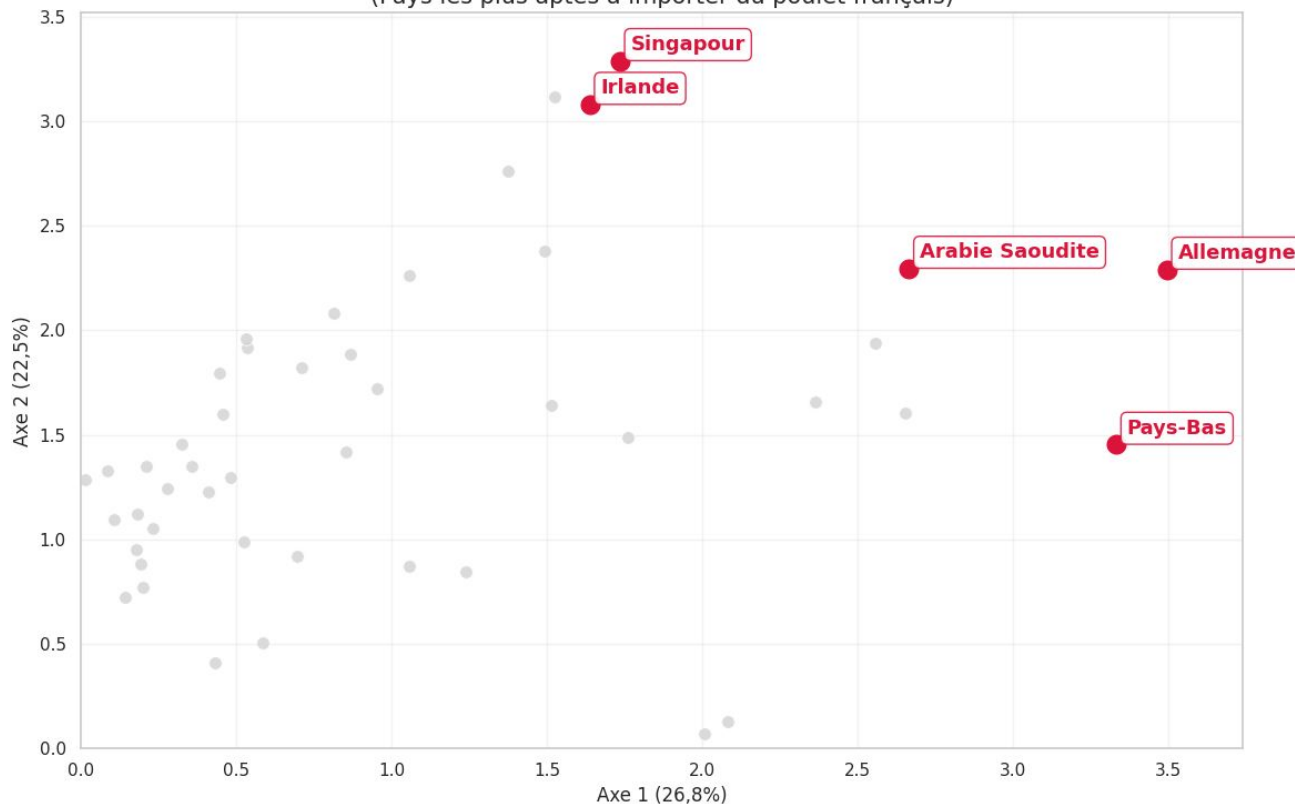
Etape 7 LE TOP 5

Les 5 pays les “plus haut à droite”,
sont les plus intéressants
commerciallement pour la
problématique d’importation de poulet
français.

Allemagne
Singapour
Arabie Saoudite
Pays-Bas
Irlande

Avec une préférence pour
l’Allemagne

Top 5 pays du quadrant haut droit
(Pays les plus aptes à importer du poulet français)

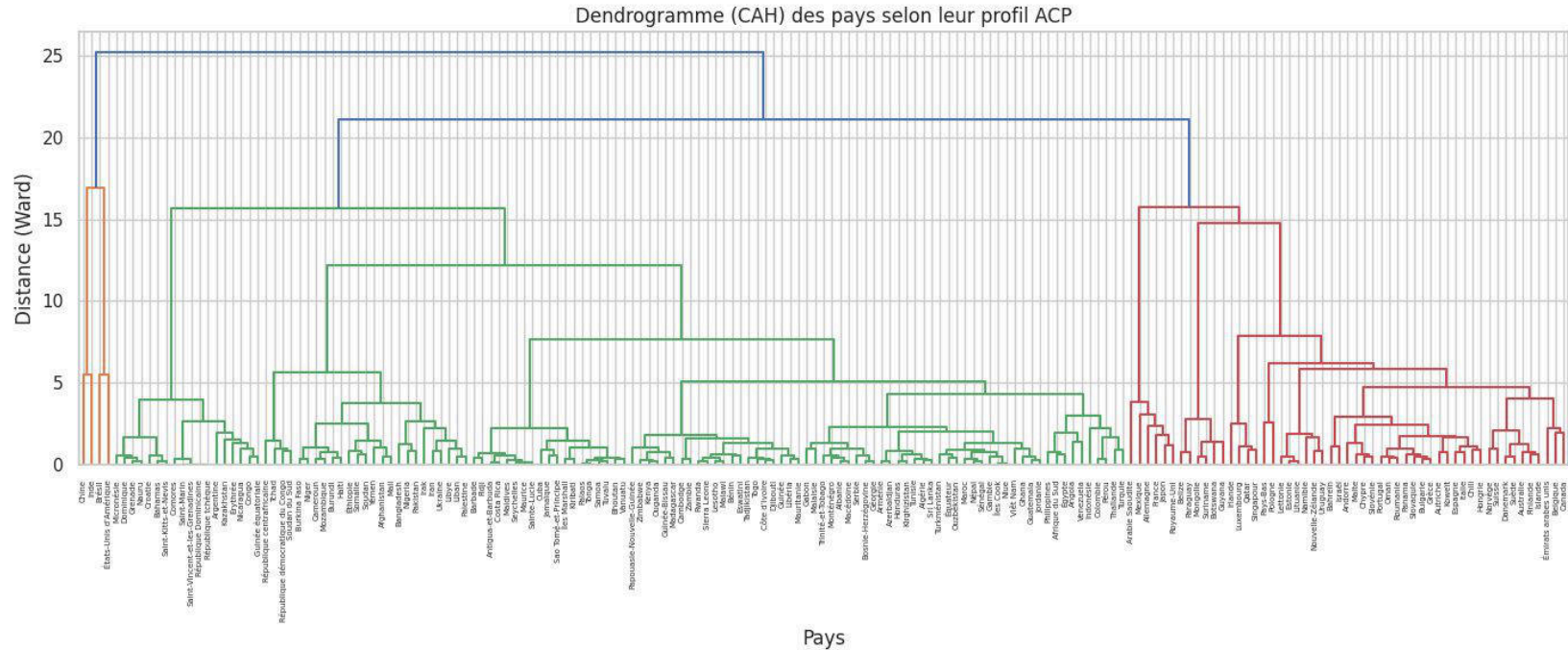


Etape 1 LE DENDROGRAMME

Principe : on rassemble chaque individus deux par deux selon leur ressemblance.

Visualisation : on obtient une sorte d'arbre généalogique "dendrogramme" qui montre qui ressemble à qui et qui forme des groupes.

Calcul : pour définir un nombre de groupe logique, on repère le “coude” dans l’arbre.



Etape 2 LE COUDE

Si on coupe en 19 : 3 clusters majeures (orange vert et rouge) d'où émergent les 3 branches colorées

Choix de 3 CLUSTERS



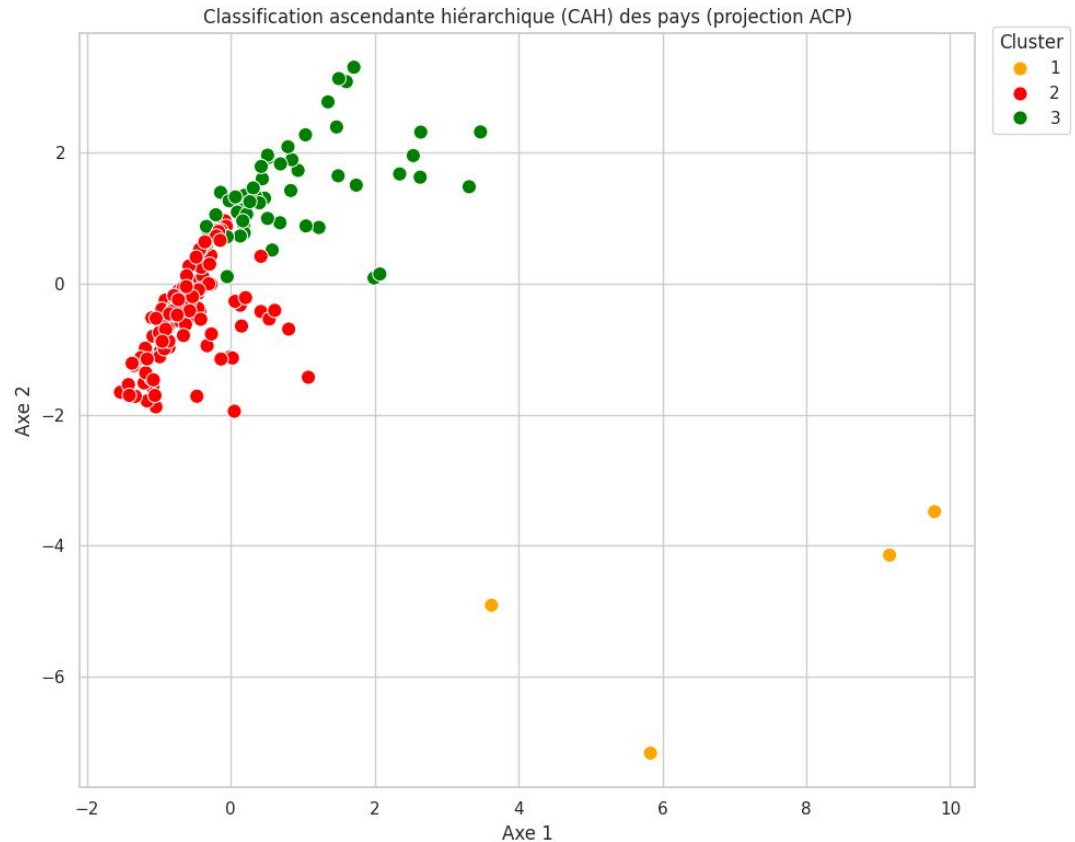
5. Classification Ascendante Hiérarchique CAH

Etape 3 Visualisation ACP

Vérification :

- CLUSTERS bien séparés et sans chevauchement
- On visualise facilement le cluster situé dans le quadrant haut droit
- Visualisation sans appel des 4 mastodontes du marché de l'export

Nos pays cibles se situent en vert dans le cluster 3



6. K-means

Etape 1 Explication

Principe : une méthode automatique qui crée des groupes d'individus selon leurs ressemblances. cette méthode calcule des "centres" et regroupe les individus les plus proches de ces centres. Le nombre de centre que l'on calcule est la résultante du calcul du dendrogramme

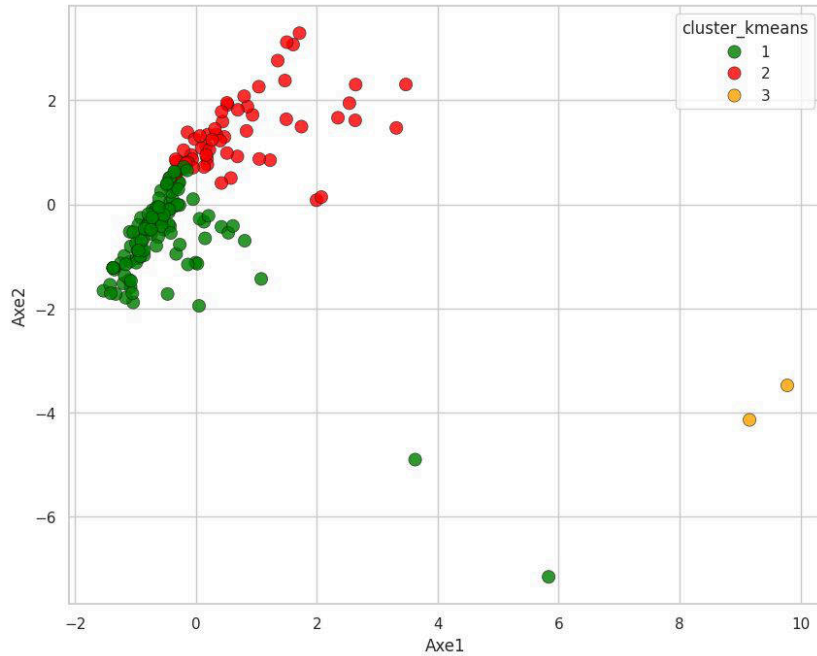
Fonctionnement : L'algorithme va placer 3 centres de groupe nommés centroïdes au hasard. Ensuite il va répartir chaque pays dans le groupe dont il est le plus proche. Ensuite les centres vont être remplacés selon la moyenne des points de chaque groupe

Résultat : chaque pays appartient à un et un seul groupe, des groupes homogènes sont mis en évidence.

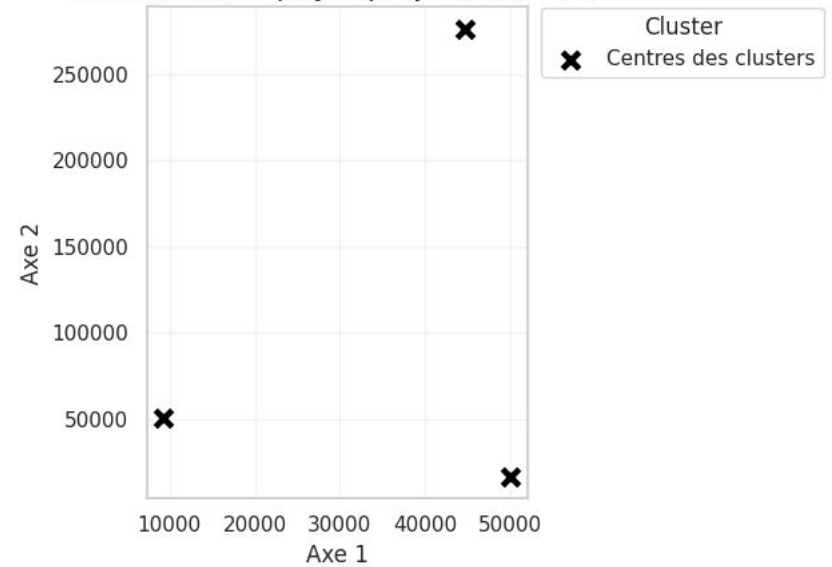


6. K-means

Etape 2
3 CLUSTERS



K-means des pays (projection ACP)



6. K-means

Etape 3 MOYENNE ET TOP 5

Après le calcul des moyennes, par rapport au point d'origine on trouve que le cluster le plus intéressant est le cluster 2

```
cluster_kmeans      Axe1      Axe2
0      -0.611473  -0.473748
1       0.829182   1.403632
2       7.095806  -4.969070
Cluster le plus haut/droit : 1
```

Calcul du Top 5 avec la méthode K-means

```
Top 5 pays les plus favorables de ce cluster :
country      Axe1      Axe2      score
4      Allemagne 3.495460  2.289647  5.785108
152     Singapour 1.734932  3.286740  5.021672
8      Arabie Saoudite 2.665320  2.294039  4.959360
129     Pays-Bas 3.332256  1.452704  4.784960
77      Irlande 1.637913  3.077621  4.715534
```

Allemagne
Singapour
Arabie Saoudite
Pays-Bas
Irlande

7. Conclusion

TOP 5 : Allemagne, Singapour, Arabie Saoudite, Pays-Bas, Irlande

ACP + CAH + K-means

3 méthodes d'analyse du marché différentes
mais que l'on utilise souvent ensemble car complémentaires

ACP

Analyse en Composantes Principales

**REDUCTION DE LA DIMENSION DES
DONNÉES**

Sert à projeter et visualiser
les individus (pays) dans l'espace

CAH

Classification Ascendante Hiérarchique

**REGROUPER LES PAYS EN FAMILLES DE
PROFILS SIMILAIRES**

Montre comment structurer les groupes
"naturels", décide d'un nombre de familles
"logiques"

K-means

**SEPARER LES PAYS EN GROUPES
"SIMILAIRES"**

Permet de partitionner automatiquement en
assignant chaque pays à un groupe
opérationnel.

