

# La poule qui chante

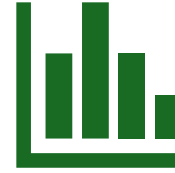
Etude de marché 2024

Plan de développement à l'international

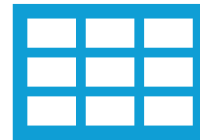
# Sommaire



Contexte: Le marché de la  
volaille et notre objectif de  
développement



Analyse en composantes  
principales



Clustering: CAH et K Means



Résultats et  
recommandations



# Le marché mondial de la volaille



**Production mondiale** : En 2023, la production mondiale de volaille a atteint environ **135 millions de tonnes**. Les États-Unis, le Brésil et la Chine représentent ensemble près de **50 %** de cette production.

**Exportations** : Le Brésil est le premier exportateur mondial de viande de volaille, représentant environ **30 %** des exportations mondiales, suivi des États-Unis avec environ **15 %**.

**Consommation** : La consommation mondiale de volaille a fortement augmenté ces dernières années. Elle est estimée à environ **16 kg par habitant** en moyenne mondiale, avec des pics dans les pays comme les États-Unis où la consommation est supérieure à **45 kg par habitant** par an.

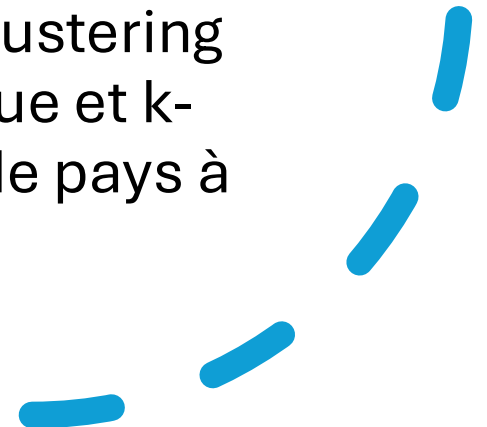
**Demande croissante en Asie et en Afrique** : Ces deux continents enregistrent la plus forte augmentation de la consommation de volaille. En Chine, par exemple, la consommation de volaille a augmenté de **20 %** au cours des dix dernières années.

# La mission

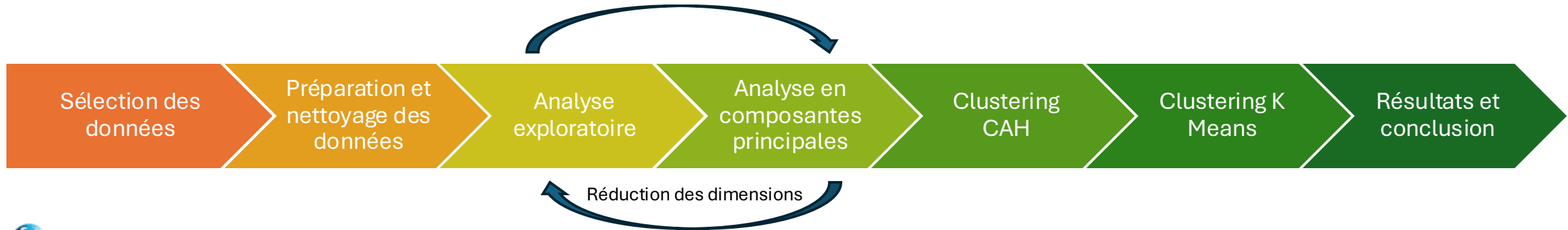
**Objectif** : Proposer une analyse des pays pour cibler les marchés potentiels d'exportation de nos poulets issue de l'agriculture biologique.

**Analyse des données** : Utiliser les données de la FAO et autres sources ouvertes (Banque Mondiale, etc.) pour construire un jeu de données propre, incluant au moins 8 variables pour environ 100 pays.

**Approche analytique** : Réaliser une ACP (Analyse en Composantes Principales) et un clustering (classification ascendante hiérarchique et k-means) pour identifier les segments de pays à cibler.



# Les différentes étapes de l'analyse



Création de variables:  
**tx\_prod\_import** = Taux de production par rapport à l'importation  
**tx\_import\_dispo** = Taux d'importation par rapport à la disponibilité intérieure  
**tx\_prod\_dispo** = Taux de production par rapport à la disponibilité intérieure  
**Evol\_pop\_sur\_5ans\_(%)**

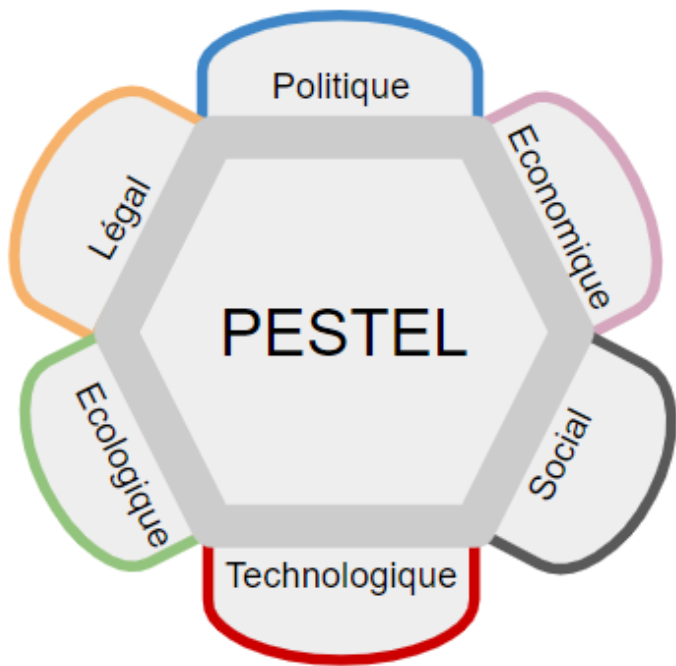
Régression linéaire sur 3 variables pour déterminer les valeurs manquantes de 'pop\_ss\_moyen\_alim\_saine\_(%)'

Suppression des outliers : 15 pays supprimés

**Distance\_from\_france\_km :**  
Calcul de la distance entre les capitales des pays du dataframe

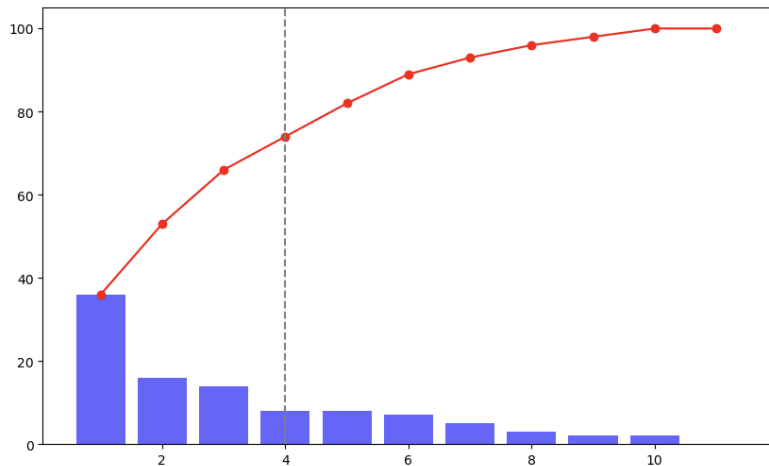
# Dataframe final

- 146 pays
- 11 variables

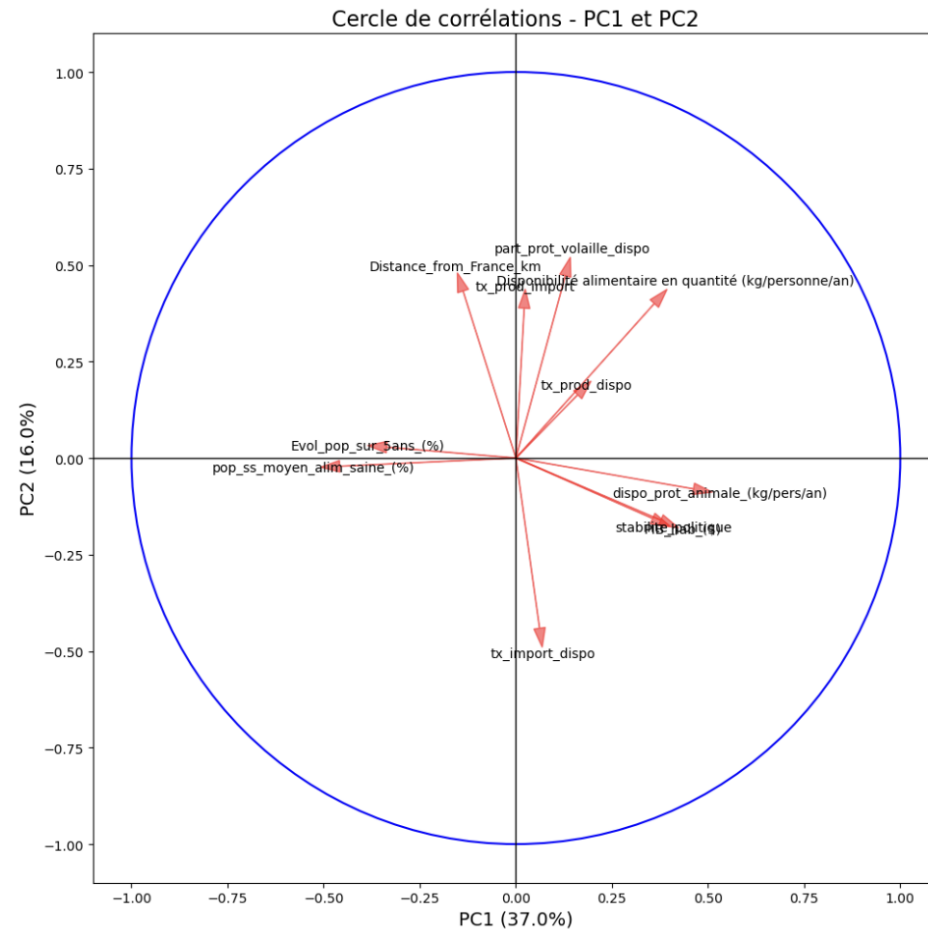


stabilite_politique	Indice de stabilité politique du pays en 2017
pop_ss_moyen_alim_saine_ (%)	Part de la population du pays n'ayant pas les moyens d'acheter de la nourriture saine en 2017
Evol_pop_sur_5ans_ (%)	Evolution de la population du pays entre 2012 et 2017
Distance_from_France_km	Distance du pays depuis la France (distance calculée entre les capitales)
Disponibilité alimentaire en quantité (kg/personne/an)	Disponibilité alimentaire en viande de volaille en kg/personne/an en 2017
PIB_hab_ (\$)	PIB par habitant en 2017
dispo_prot_animale_ (kg/pers/an)	Disponibilité en protéines animales (tous animaux confondus) en kg/pers/an
tx_prod_dispo	Taux de production par rapport à la disponibilité intérieure
tx_import_dispo	Taux d'importation par rapport à la disponibilité intérieure
tx_prod_import	Taux de production par rapport à l'importation
part_volaille_dispo_prot	Part de la viande de volaille dans la disponibilité totale en protéine animale

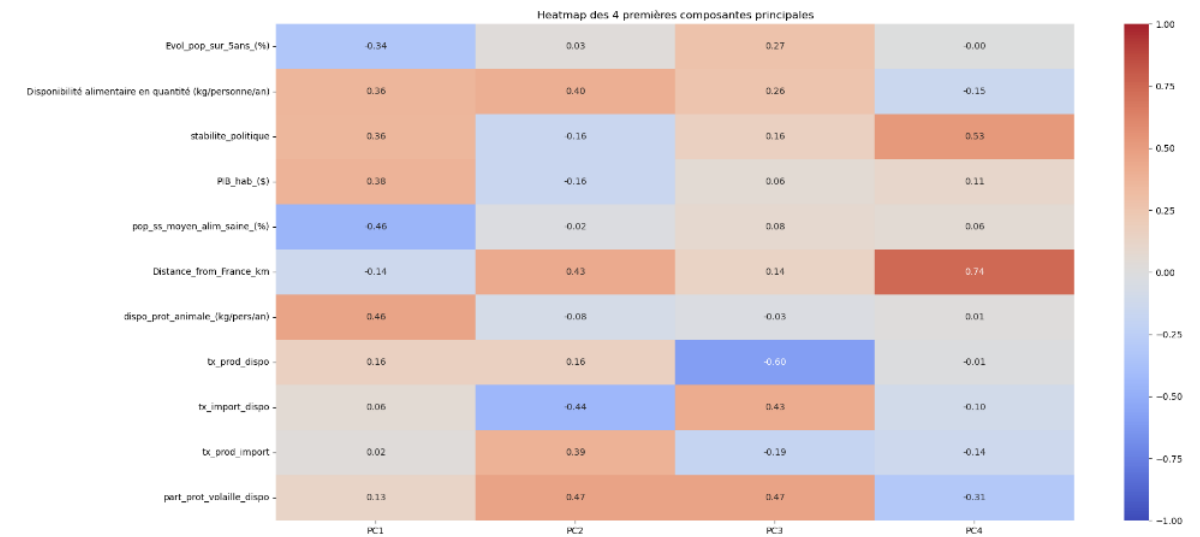
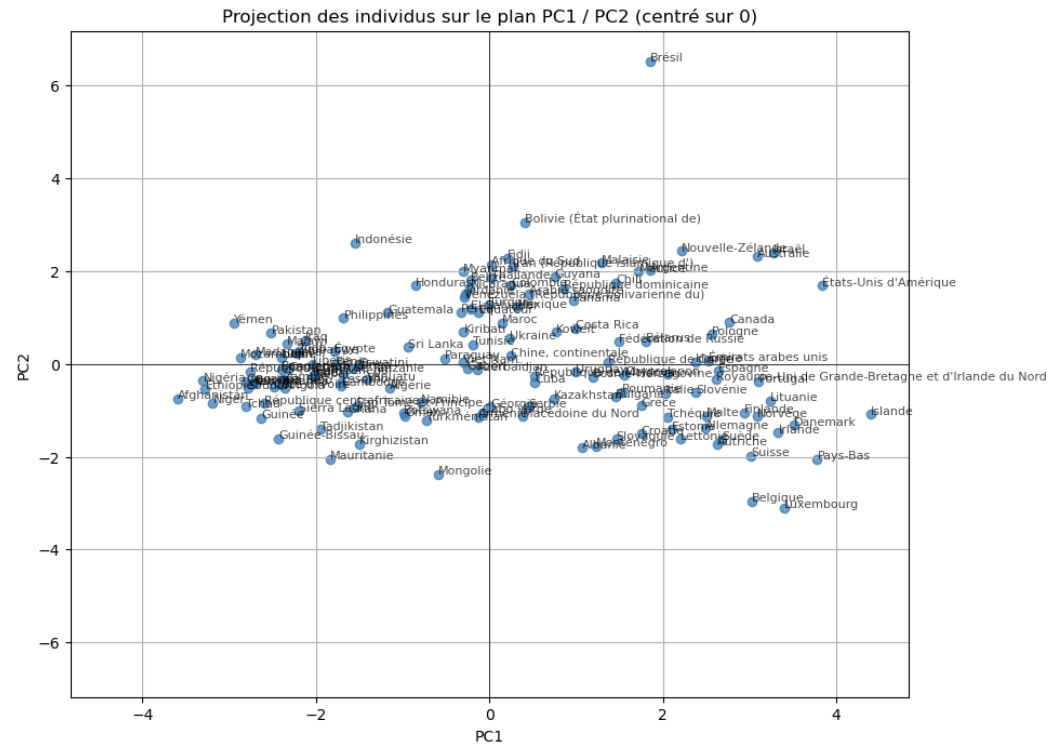
# Analyse en composantes principales



	Dimension	Variance expliquée	% variance expliquée	% cum. var. expliquée
0	Dim1	4.037231	36.0	36.0
1	Dim2	1.797142	16.0	53.0
2	Dim3	1.512328	14.0	66.0
3	Dim4	0.904072	8.0	74.0
4	Dim5	0.855357	8.0	82.0
5	Dim6	0.721155	7.0	89.0
6	Dim7	0.526432	5.0	93.0
7	Dim8	0.281607	3.0	96.0
8	Dim9	0.221119	2.0	98.0
9	Dim10	0.167528	2.0	100.0
10	Dim11	0.051892	0.0	100.0

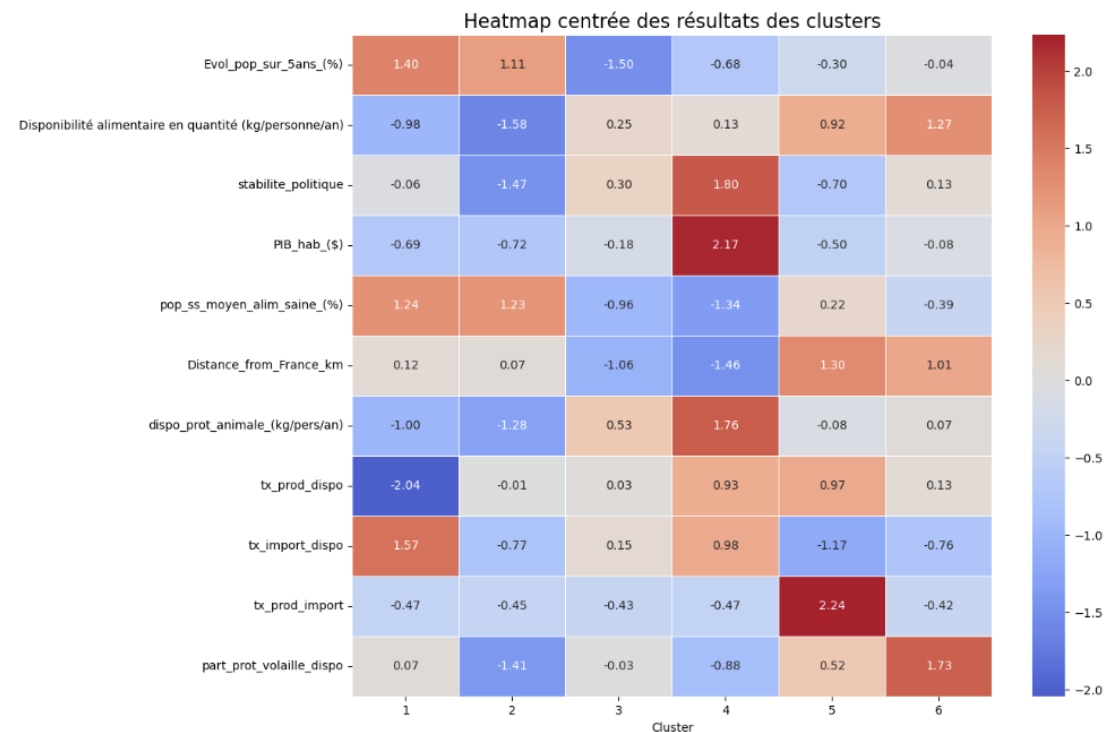
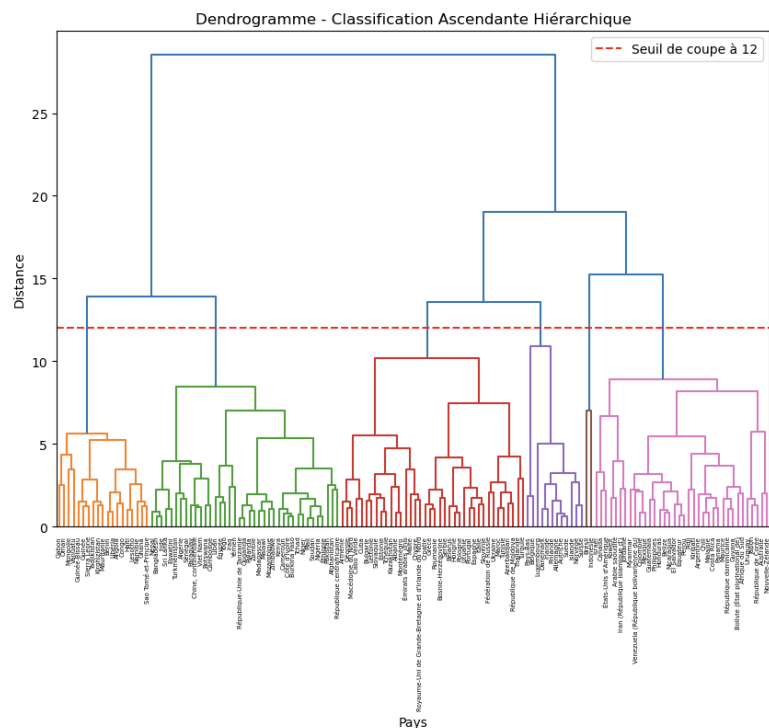


# Analyse en composantes principales





# Clustering méthode CAH



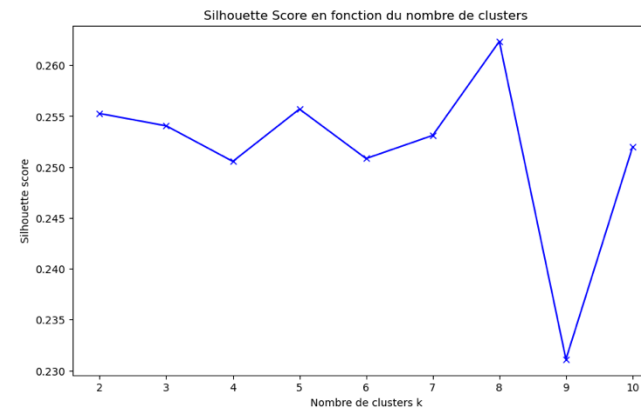
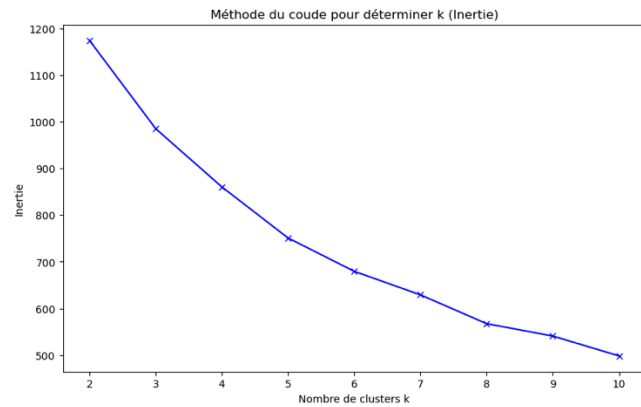
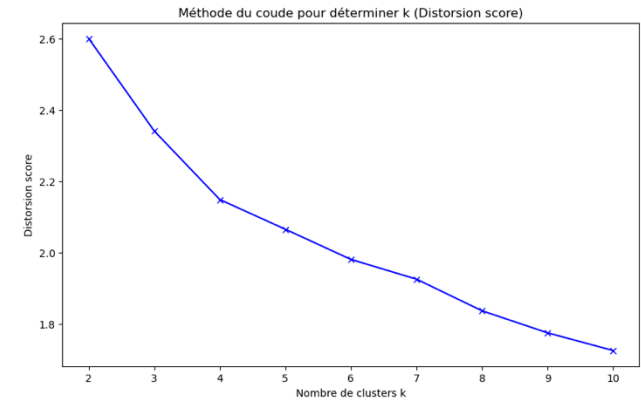
Cluster	Evol_pop_sur_5ans_(%)	Disponibilité alimentaire en quantité (kg/personne/an)	stabilite_politique	PIB_hab_(\$)	pop_ss_moyen_alim_saine (%)	Distance_from_France km	dispo_prot_animale_(kg/pers/an)	tx_prod_dispo	tx_import_dispo	tx_prod_import	part_prot_volaille_dispo
1	11,67	10,03	-0,10	2991,61	73,48	6235,16	8,02	23,75	92,80	28,44	17,77
2	10,56	4,52	-0,94	2313,64	73,12	6071,38	6,29	88,92	13,36	2786,88	9,39
3	0,53	21,41	0,11	13515,50	11,01	2356,80	17,39	90,35	44,83	5259,80	17,17
4	3,67	20,32	1,00	62859,43	0,29	1018,75	24,95	119,17	72,87	632,92	12,39
5	5,12	27,61	-0,49	6868,25	44,50	10110,15	13,68	120,66	0,04	351733,33	20,31
6	6,12	30,92	0,01	15707,69	27,12	9162,57	14,56	93,53	13,72	6806,28	27,16



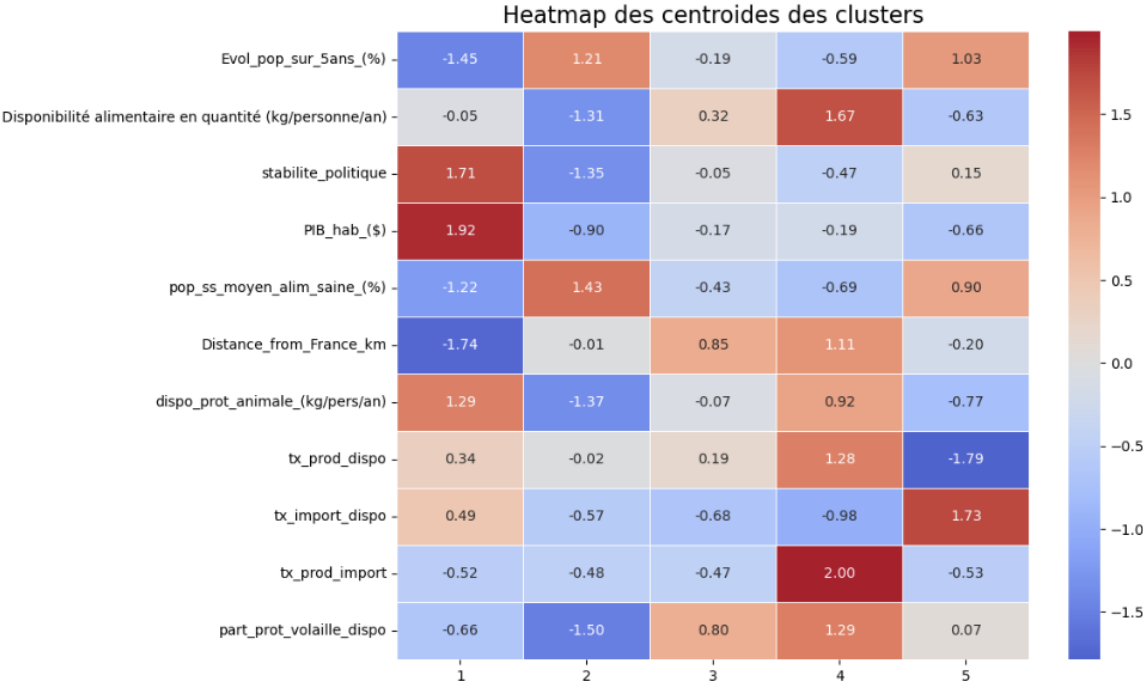
# Clustering méthode K Means

## Déterminer le nombre de clusters

- **Méthode du coude avec le distorsion score** = 4 Clusters
- **Méthode du coude avec l'inertie** = 5 Clusters
- Le **silhouette score** est relativement constant pour  $k=2$  à  $k=5$ , avec un score autour de **0.255**. Il y a une augmentation importante pour  $k=8$ , qui montre un score maximal à environ **0.260** puis le score chute brutalement à partir de  **$k=9$** .
- La **cohérence entre les trois analyses (distorsion, inertie, silhouette score)** suggère que  **$k=4$  ou  $k=5$**  reste le choix optimal, avec peut-être une préférence pour  **$k=5$**  si l'on souhaite un compromis entre compacité et qualité des clusters.

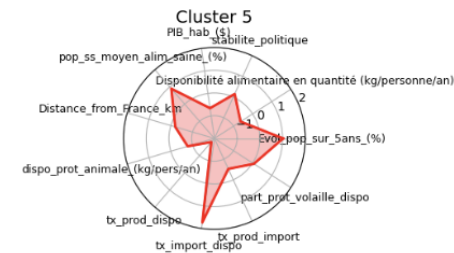
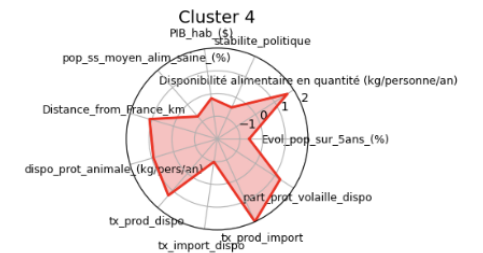
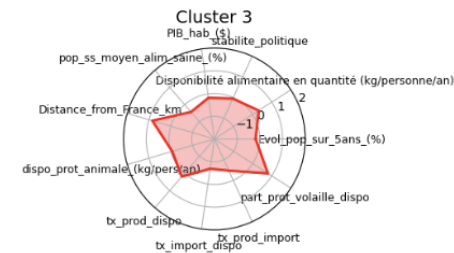
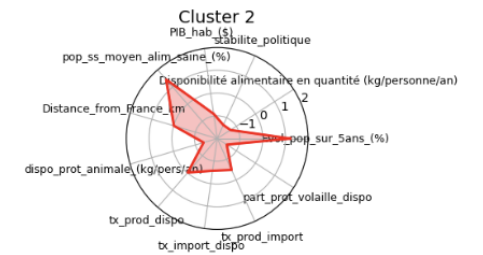
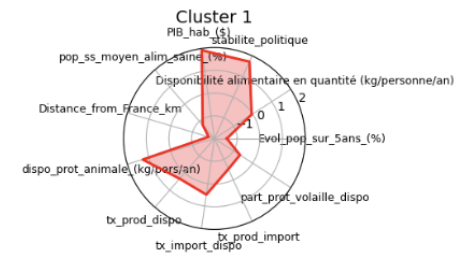
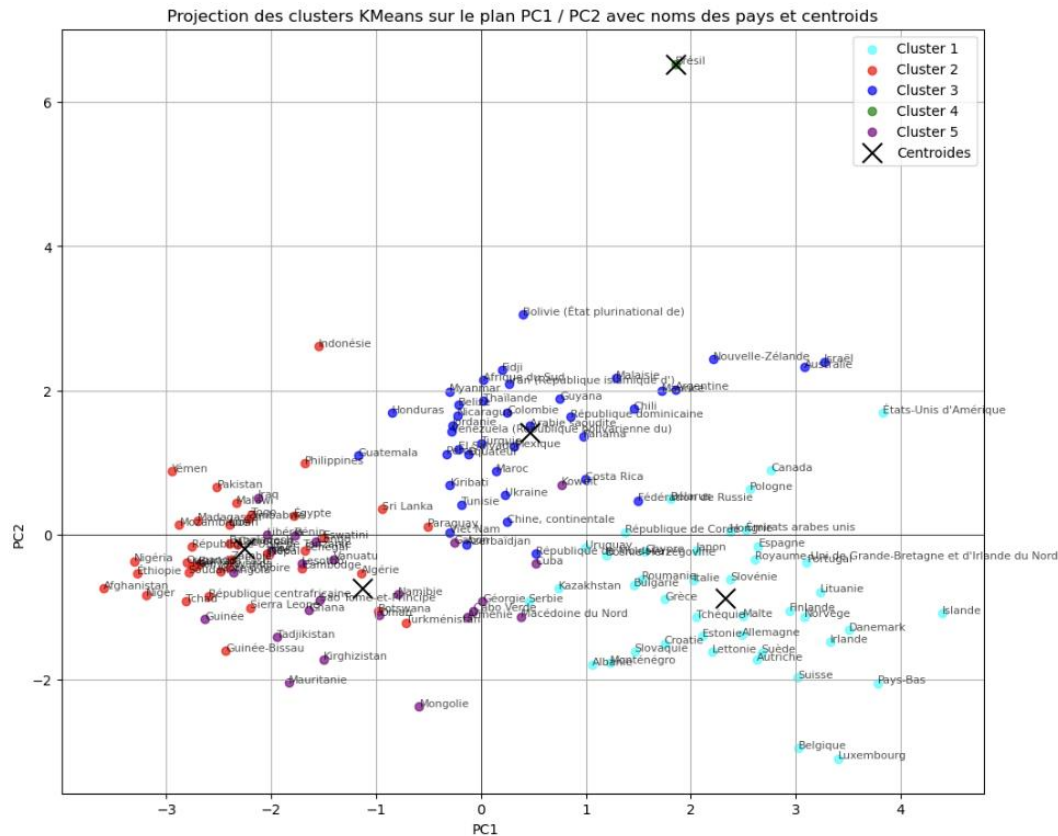


# Clustering méthode K Means



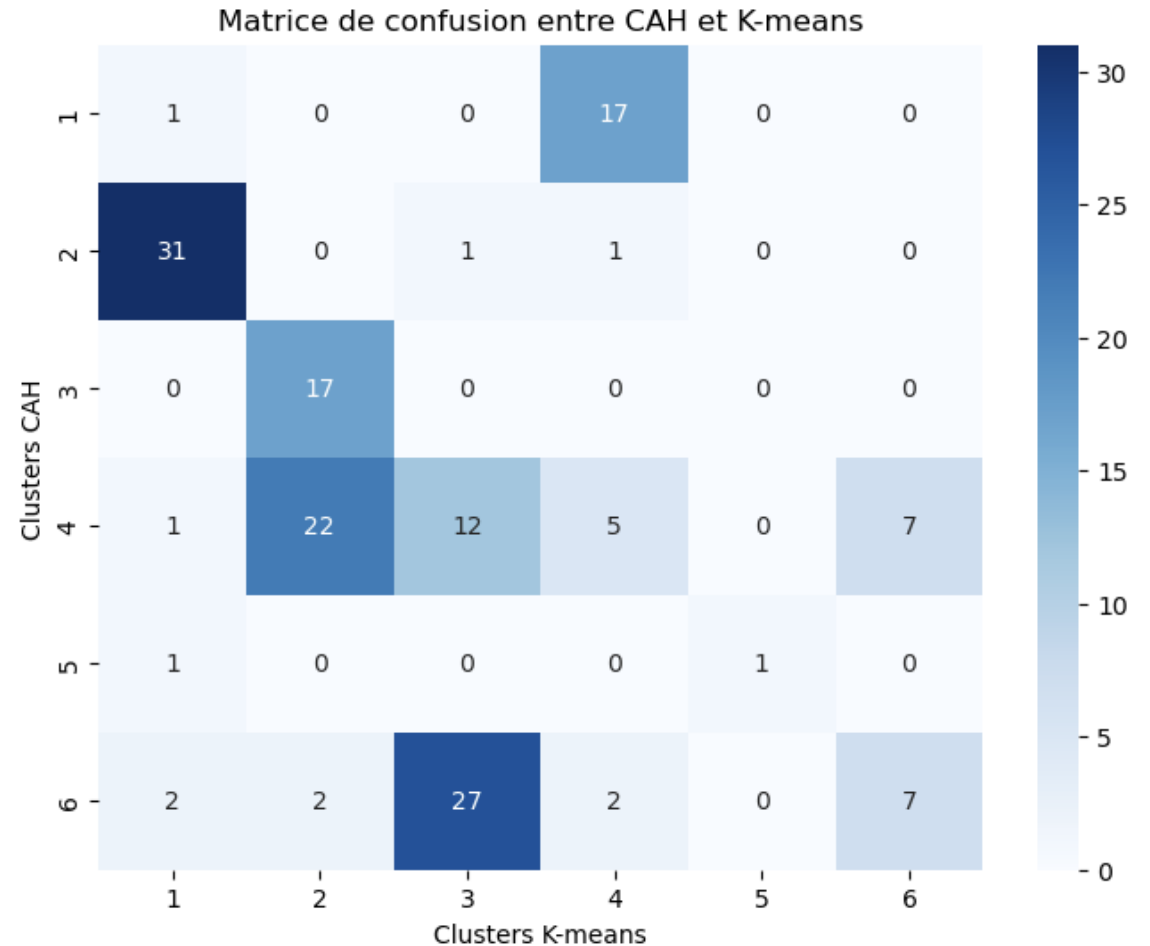
Cluster	Evol_pop_sur_5ans_(%)	Disponibilité alimentaire en quantité (kg/personne/an)	stabilite_politique	PIB_hab_(\$)	pop_ss_moyen_alim_saine_(%)	Distance_from_France_km	dispo_prot_animale_(kg/pers/an)	tx_prod_dispo	tx_import_dispo	tx_prod_import	part_prot_volaille_dispo
1	3,92	27,00	0,19	27519,61	12,62	4736,91	20,39	104,83	47,06	229353,24	17,98
2	8,03	12,16	-0,40	8485,44	61,97	6760,13	10,10	96,25	21,52	319085,80	15,30
3	5,88	31,35	-0,15	13427,19	27,30	7773,11	15,15	101,34	18,86	346372,04	22,67
4	5,25	47,39	-0,23	13284,14	22,41	8077,05	18,98	127,40	11,47	5972930,34	24,22
5	7,75	20,17	-0,11	10156,86	51,98	6536,48	12,42	54,15	76,93	217469,74	20,31

# Clustering méthode K Means



# Comparaison des 2 methodes

- **Correspondance élevée** : Le **Cluster 2 de CAH** correspond bien au **Cluster 2 de K-Means**, ce qui indique une similarité dans la détection de ce groupe entre les deux méthodes.
- **Fragmentation des clusters** : Les **Clusters 1, 3, 4, et 5 de CAH** montrent une fragmentation importante dans les résultats de K-Means, avec des individus répartis dans plusieurs clusters K-Means. Cela pourrait indiquer que la méthode K-Means propose une segmentation différente des données.
- **Différences dans la structure des clusters** : Globalement, la méthode K-Means semble regrouper différemment les individus par rapport à la CAH, ce qui peut être dû à la manière dont les distances et les regroupements sont calculés dans chaque algorithme.
- En résumé, il y a des **différences notables** entre les deux méthodes de clustering, notamment une meilleure correspondance pour certains clusters (ex : Cluster 2), tandis que d'autres clusters sont plus **fragmentés** ou **redistribués** dans K-Means.



# Analyse Comparative : CAH vs K-Means

---

## 1. Distance Intra-Cluster

**K-Means : 2.35**

**CAH : 2.95**

Ici, **K-Means** obtient une **meilleure performance** avec une distance intra-cluster plus faible (2.35) que CAH (2.95), ce qui signifie que K-Means forme des groupes plus cohérents.

## 2. Silhouette Score

**K-Means : 0.256**

**CAH : 0.226**

Un score plus élevé indique une meilleure qualité de clustering. Ici encore, **K-Means** obtient un **meilleur score** avec 0.256, contre 0.226 pour CAH, indiquant une meilleure séparation et cohésion des clusters.

## 3. Jaccard Score

**Jaccard Score : 0.159**

Avec un score de 0.159, il y a une **faible correspondance** entre les deux méthodes, ce qui signifie que les résultats des clusters sont assez différents.

## 4. Adjusted Rand Index (ARI)

**ARI : 0.529**

Avec un ARI de 0.529, il y a une **similitude modérée** entre les deux méthodes, mais pas suffisante pour conclure que l'une surpasse nettement l'autre.

## Conclusion : K-Means à privilégier

En se basant sur les performances de la distance intra-cluster et du Silhouette Score, **K-Means** est la méthode à privilégier pour segmenter les pays en clusters.

Les clusters formés par K-Means sont plus compacts et mieux séparés que ceux de CAH.

# Conclusion et recommandations

---

## **Cluster 1 est le plus favorable pour l'exportation:**

- Pouvoir d'achat élevé permettant aux consommateurs d'acheter des produits premium.
- Demande existante pour des aliments biologiques et de haute qualité.
- Proximité géographique (pour les pays européens) réduisant les coûts logistiques.
- Réglementations alimentaires souvent alignées avec les normes européennes, facilitant les procédures d'exportation.

42 pays du Cluster 1 :

Albanie, Autriche, Bulgarie, Canada, Chypre, Danemark, Bélarus, Estonie, Finlande, Allemagne, Bosnie-Herzégovine, Grèce, Hongrie, Croatie, Islande, Irlande, Italie, Kazakhstan, Japon, République de Corée, Lettonie, Lituanie, Malte, Pays-Bas, Norvège, Tchéquie, Pologne, Portugal, Roumanie, Slovénie, Slovaquie, Espagne, Suède, Suisse, Émirats arabes unis, Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, États-Unis d'Amérique, Uruguay, Belgique, Luxembourg, Serbie et Monténégro.

## **Cluster 3 présente également des opportunités :**

- Économies émergentes avec une classe moyenne en croissance.
- Intérêt croissant pour les produits de meilleure qualité et potentiellement biologiques.
- Nécessité d'évaluer les conditions spécifiques de chaque pays, y compris les barrières commerciales et les préférences culturelles.

39 pays du Cluster 3 :

Argentine, Australie, Bolivie (État plurinational de), Belize, Myanmar, Chili, Chine, continentale, Colombie, Costa Rica, Azerbaïdjan, République dominicaine, Équateur, El Salvador, Fidji, Kiribati, Guatemala, Guyana, Honduras, Iran (République islamique d'), Israël, Jordanie, Malaisie, Maurice, Mexique, Maroc, République de Moldova, Nouvelle-Zélande, Nicaragua, Panama, Pérou, Fédération de Russie, Arabie saoudite, Afrique du Sud, Thaïlande, Tunisie, Turquie, Ukraine, Venezuela (République bolivarienne du), Viet Nam