



# La poule qui chante

**Identifier les marchés internationaux privilégiés pour l'exportation de volailles**

**Projet9-Produisez une étude de marché avec R ou Python**

# **SOMMAIRE**

**TENDANCES ET PERSPECTIVES DU MARCHÉ GLOBAL DE LA VOLAILLE**

**MISSION ET DÉMARCHE DE L'ANALYSE**

**CHOIX DES VARIABLES UTILISÉES**

**ANALYSE EN COMPOSANTES PRINCIPALES**

**CLUSTERING : CAH ET MÉTHODE DES K-MEANS**

**QUELLE MÉTHODE DE CLUSTERING CHOISIR ?**

**INTERPRÉTATION ET DÉCISION**

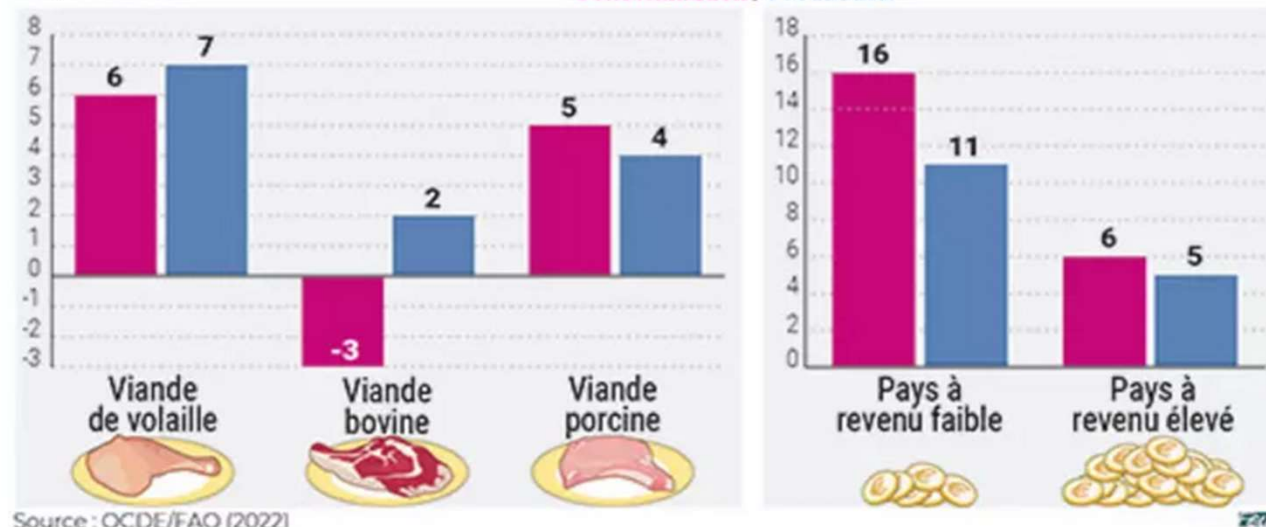
## TENDANCES ET PERSPECTIVES DU MARCHÉ GLOBAL DE LA VOLAILLE

### LA VOLAILLE, MOTEUR DE LA HAUSSE DE CONSOMMATION DE VIANDE

Variation prévue en % entre 2019-2021 et 2031

Évolution de la production et de la consommation de viande par habitant, exprimée en protéines

Consommation / Production



Source : OCDE/FAO (2022)

© Les Marchés

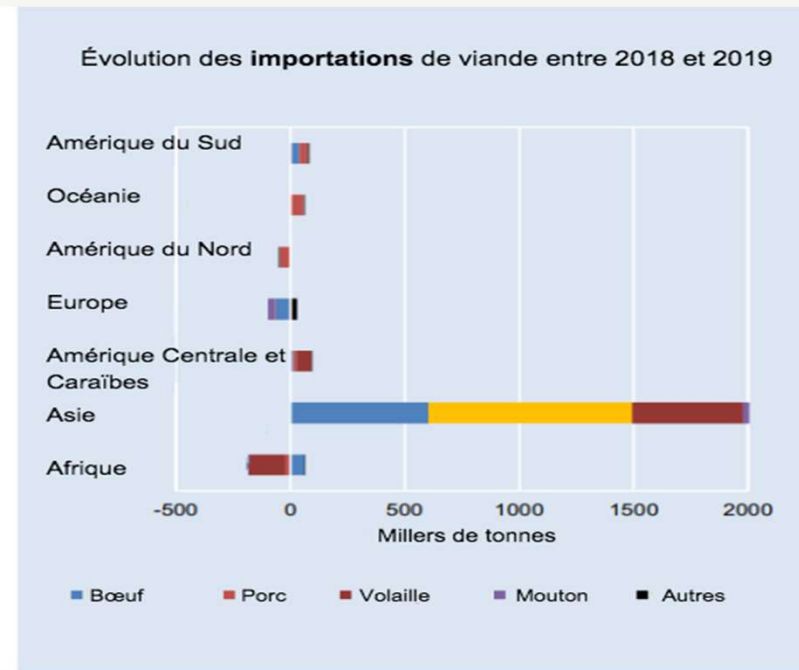
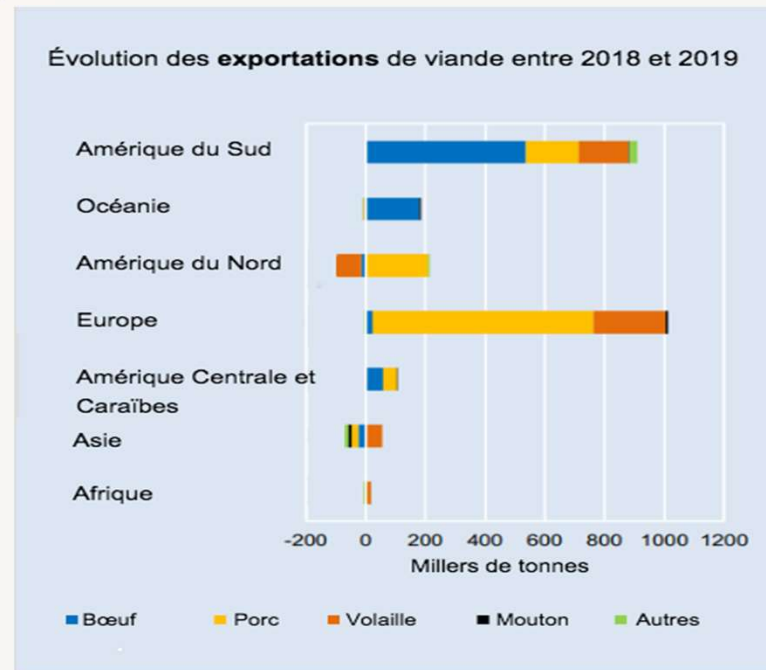
N°1 DES VIANDES CONSOMMÉES  
DANS LE MONDE

#### PERSPECTIVE DU MARCHÉ

- DEFIS SANITAIRES EN DEVELOPPEMENT
- RISQUE DE GRIPPE AVIAIRE

#### TENDANCE ACTUELLE

- MARCHÉ DU POULET EN PLEINE CROISSANCE
- IMPORTATIONS MONDIALES DE POULET EN HAUSSE



Le marché est prometteur, un plan de développement d'exportation en croissance continue , importation en hausse : Depuis 2018, l'importation mondiale de poulet a connu une croissance de 30%.

## MISSION ET DÉMARCHE DE L'ANALYSE

### CONTEXTE:

La Poule qui Chante cherche à étendre ses activités à l'international.

Nous allons utiliser l'analyse de données pour identifier les marchés les plus prometteurs.

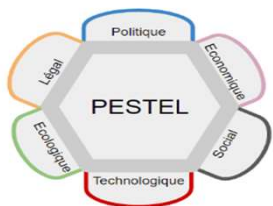
### MISSION ET OBJECTIFS

L'objectif est de regrouper les pays en fonction de plusieurs critères pour cibler notre expansion

### DÉMARCHE À SUIVRE:

Nous allons sélectionner, préparer et nettoyer les données, suivies d'une analyse exploratoire, d'une analyse en composantes principales et de méthodes de clustering.

Le traitement a été réalisé en langage Python.

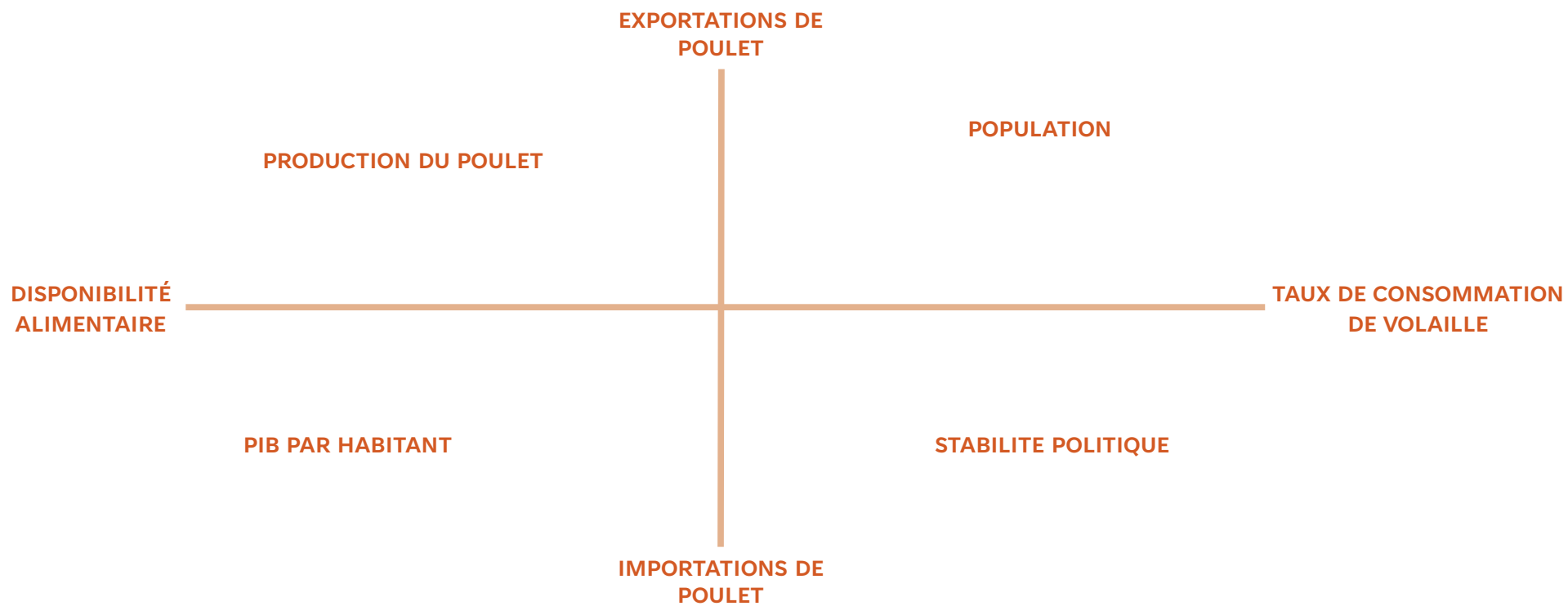


## VARIABLES UTILISÉES

*Données extraites du site fao.org sous fichiers .csv*



**Variables issues des  
bilans alimentaires**



NETTOYAGE DES DONNÉES ET ANALYSE  
EXPLORATOIRE (HORS ACP)

#J'affiche Les pays qui ont une exportation supérieur à un million de tonnes de poulet  
poulet\_2017[poulet\_2017['Exportations poulet']>1000].sort\_values('Exportations poulet',ascending=False)

Élément	pays	Exportations poulet	Importations poulet	Nourriture poulet	Production poulet	poulet_al
2073	Brésil	4223	3	9982	14201	204
15860	États-Unis d'Amérique	3692	123	18100	21914	219
11036	Pays-Bas	1418	608	346	1100	70
11227	Pologne	1025	55	1150	2351	107

#J'affiche La liste des pays qui présentent une prospérité économique très élevée  
pib\_2017.nlargest(n=10, columns=['pib par habitant'])

Année	pays	pib par habitant
114	Monaco	173611.814788
99	Liechtenstein	170875.739925
22	Bermudes	112877.376531

#J'affiche Les pays qui ont une importation supérieur à 400 mille tonnes de poulet  
poulet\_2017[poulet\_2017['Importations poulet']>400].sort\_values('Importations poulet',ascending=False)

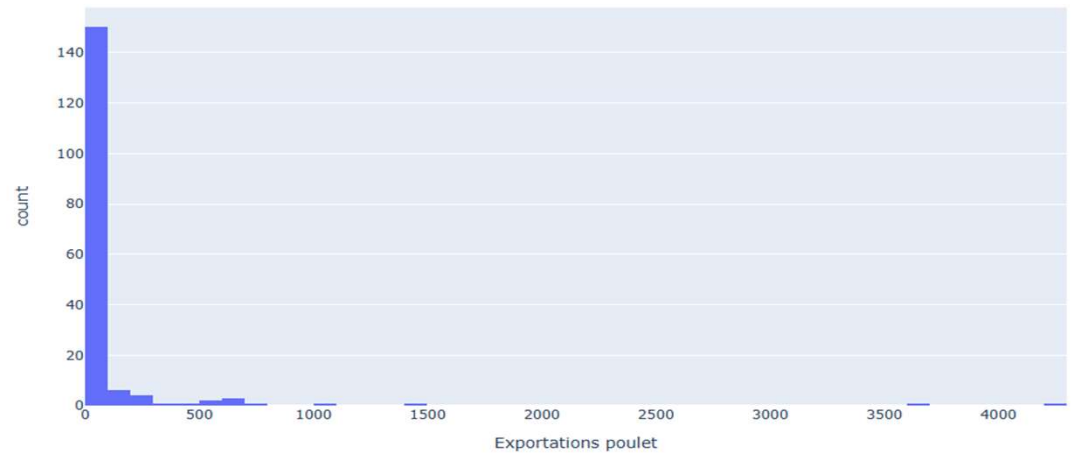
Élément	pays	Exportations poulet	Importations poulet	Nourriture poulet	Production poulet	poulet_al
7143	Japon	10	1069	2359	2215	67
9259	Mexique	9	972	4058	3249	123
3014	Chine - RAS de Hong-Kong	663	907	391	24	210
452	Allemagne	646	842	1609	1514	71

## ANALYSE UNIVARIÉE DES VARIABLES

### « ÉVALUATION DE LA DISTRIBUTION AVEC LES COEFFICIENTS D'ASSYMETRIE ET DE KURTOSIS »

• **Skewness**: Mesure de l'asymétrie d'une distribution. Une valeur négative indique une queue à gauche, une valeur positive une queue à droite.

• **Kurtosis**: Indicateur de l'aplatissement d'une distribution. Une valeur positive signifie une distribution plus "pointue" que la normale, une valeur négative une distribution plus "aplatie"



Évaluation de la distribution avec les coefficients d'asymétrie et de Kurtosis.  
Skewness: 7.3912924275999625  
Kurtosis: 59.95692289814235

```
df_final.describe()
```

	Exportations poulet	Importations poulet	Nourriture poulet	Production poulet	poulet_al	dispo_al	taux consommation volaille	Population en 2017	stabilité politique	pib par habitant
count	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	167.000000	1.670000e+02	167.000000	167.000000
mean	94.868263	83.682635	557.035928	609.832335	73.449102	2853.610778	2.492278	3.536833e+04	-0.075629	13644.677213
std	451.809342	181.723645	1710.019978	2107.109471	60.715777	459.365524	2.079863	1.122848e+05	0.900249	19236.369604
min	0.000000	0.000000	0.000000	0.000000	0.000000	1754.000000	0.000000	5.204500e+01	-2.940000	414.740323
25%	0.000000	3.000000	28.000000	12.000000	21.000000	2508.500000	0.830089	2.902508e+03	-0.635000	1939.899238
50%	0.000000	16.000000	91.000000	65.000000	62.000000	2864.000000	1.994942	9.729823e+03	-0.040000	5825.224294
75%	11.500000	78.000000	339.000000	339.500000	104.000000	3254.000000	3.397296	3.046071e+04	0.640000	16102.611534
max	4223.000000	1069.000000	18100.000000	21914.000000	243.000000	3770.000000	9.592425	1.338677e+06	1.600000	110203.008202



## TRANSFORMATION EFFECTUEE SUR LES DONNEES

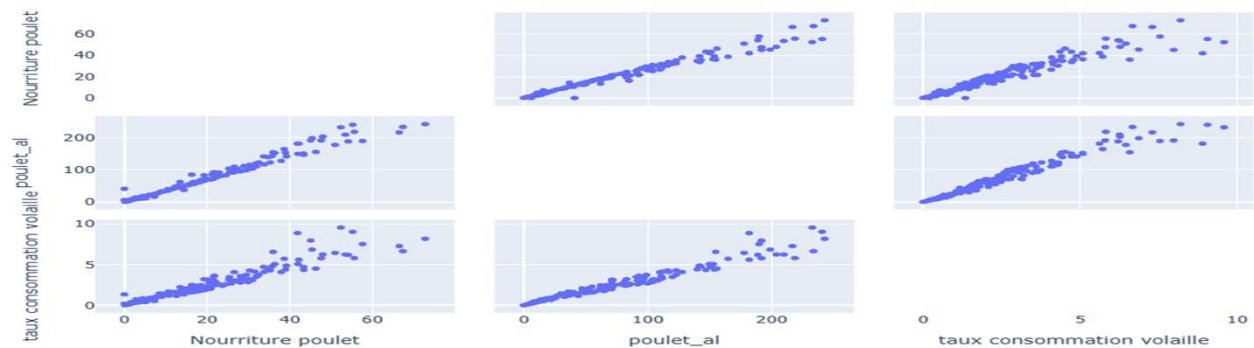
Conversion des unités : passant de milliers de tonnes à kg par habitant en fonction de la population en 2017 afin de permettre une comparaison plus équilibrée entre différentes régions ou pays.

```
col=['Exportations poulet','Importations poulet','Nourriture poulet','Production poulet']
for colonne in col:
    df_final[colonne]= df_final[colonne]*1000/ df_final['Population en 2017']
df_final
```

	pays	Exportations poulet	Importations poulet	Nourriture poulet	Production poulet	poulet_al	dispo_al	taux consommation volaille	Population en 2017	stabilité politique	piB par habitant
0	Afghanistan	0.000000	0.798984	1.515314	0.771432	5	1997	0.250376	36296.113	-2.80	530.149831
1	Afrique du Sud	1.105074	9.016001	35.695645	29.240609	143	2987	4.787412	57009.756	-0.28	6723.928582
2	Albanie	0.000000	13.175372	16.295855	4.507364	85	3400	2.500000	2884.169	0.38	4521.752219
3	Algérie	0.000000	0.048322	6.378477	6.644247	22	3345	0.657698	41389.189	-0.92	4134.936054

```
# Observation graphiques des variables très fortement corrélées
selection =['Nourriture poulet',
            'poulet_al',
            'taux consommation volaille' ]

import plotly.express as px
fig = px.scatter_matrix(df_final, dimensions=selection)
fig.update_traces(diagonal_visible=False)
fig.show()
```



# TRANSFORMATIONS EFFECTUEES SUR LES DONNEES

Transformer la variable "Population en 2017" en son logarithme permet de:

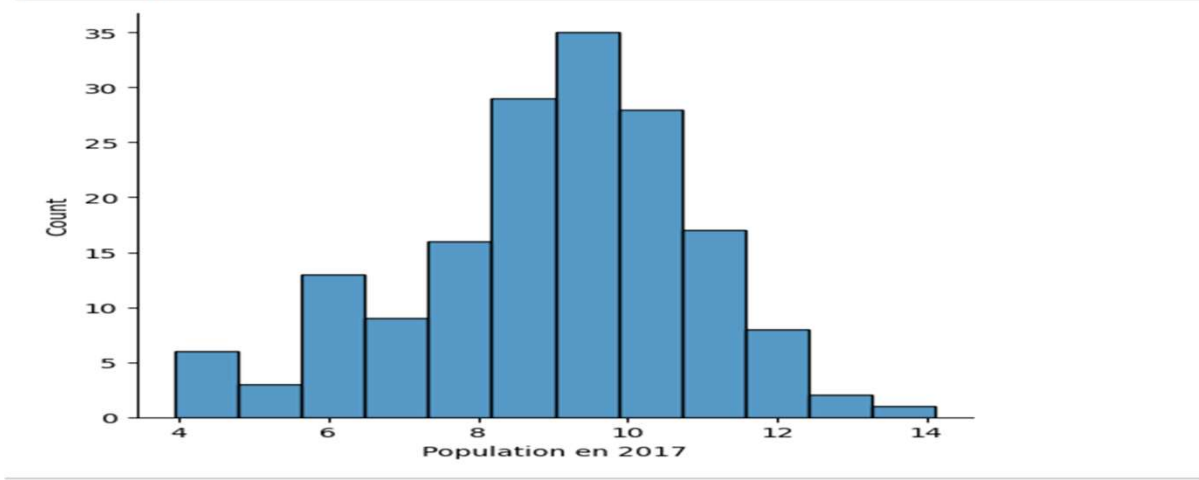
- Normaliser sa distribution
- faciliter l'interprétation et la comparaison de cette variable avec d'autres indicateurs.
- Diviser d'autres variables par la population permet de décorréler ces mesures de la taille de la population elle-même.

Cette étape est essentielle pour des comparaisons internationales justes, en particulier lorsque les tailles de population entre les pays varient grandement.

```
#transformer pop en log afin de normaliser la distribution de Population en 2017
df_final['Population en 2017']=np.log(df_final['Population en 2017'])
df_final['Population en 2017']
```

0	10.499466
1	10.950978
2	7.966992
3	10.630775
4	11.322472
...	
162	9.157699
163	9.728262
164	12.691841
165	11.574960
166	6.455260
Name: Population en 2017, Length: 167, dtype: float64	

```
sns.displot(data=df_final,x='Population en 2017')
plt.show()
```

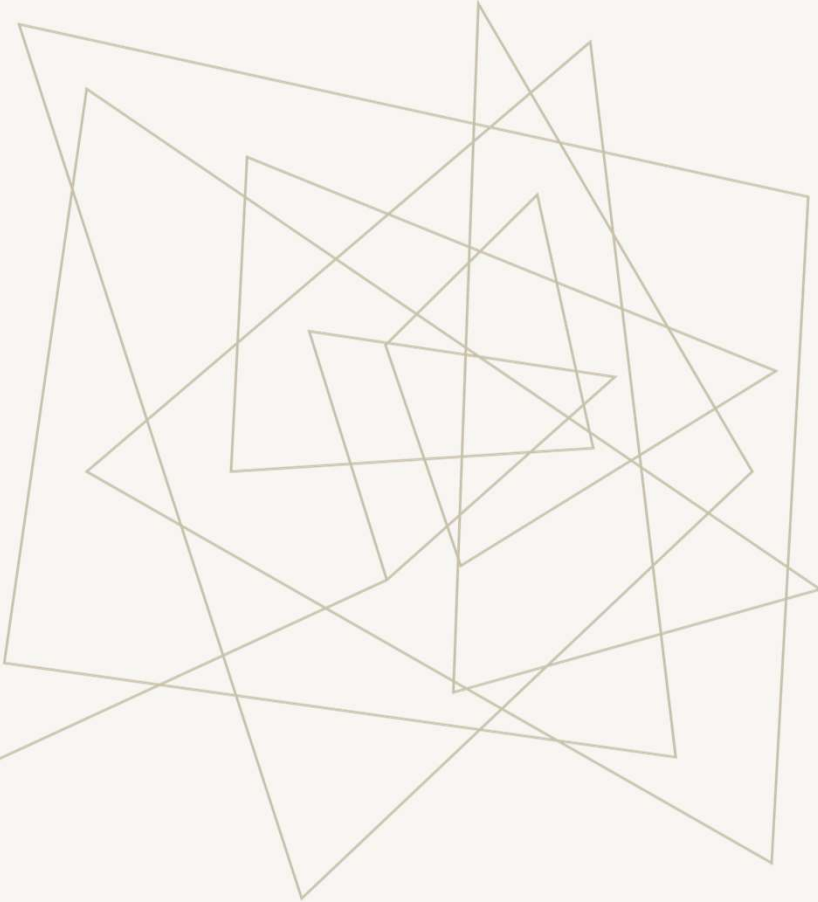


## DATAFRAME FINAL PRÊT A L'ANALYSE

```
#Dataframe final prêt à l'analyse  
df_final
```

	pays	Exportations poulet	Importations poulet	Production poulet	dispo_al	taux consommation volaille	Population en 2017	stabilité politique	pib par habitant
0	Afghanistan	0.000000	0.798984	0.771432	1997	0.250376	10.499466	-2.80	530.149831
1	Afrique du Sud	1.105074	9.016001	29.240609	2987	4.787412	10.950978	-0.28	6723.928582
2	Albanie	0.000000	13.175372	4.507364	3400	2.500000	7.966992	0.38	4521.752219
3	Algérie	0.000000	0.048322	6.644247	3345	0.657698	10.630775	-0.92	4134.936054
4	Allemagne	7.815297	10.186501	18.316346	3559	1.994942	11.322472	0.59	44670.222288

## ANALYSE EN COMPOSANTES PRINCIPALES



### OBJECTIF :

- RÉDUCTION DE DIMENSION
- FACILITER LA VISUALISATION DES DONNÉES
- D'ÉTUDIER LA VARIABILITÉ ENTRE LES INDIVIDUS
- D'ÉTUDIER LES LIAISONS ENTRE LES VARIABLES

### PRINCIPE :

- VISUALISATION DES CLUSTERS PAYS SUR LE PREMIER PLAN FACTORIEL

### • AVANTAGES :

- VISUALISATION DES DONNÉES, GAIN EN STOCKAGE ET TEMPS DE TRAITEMENT PAR LA SUITE

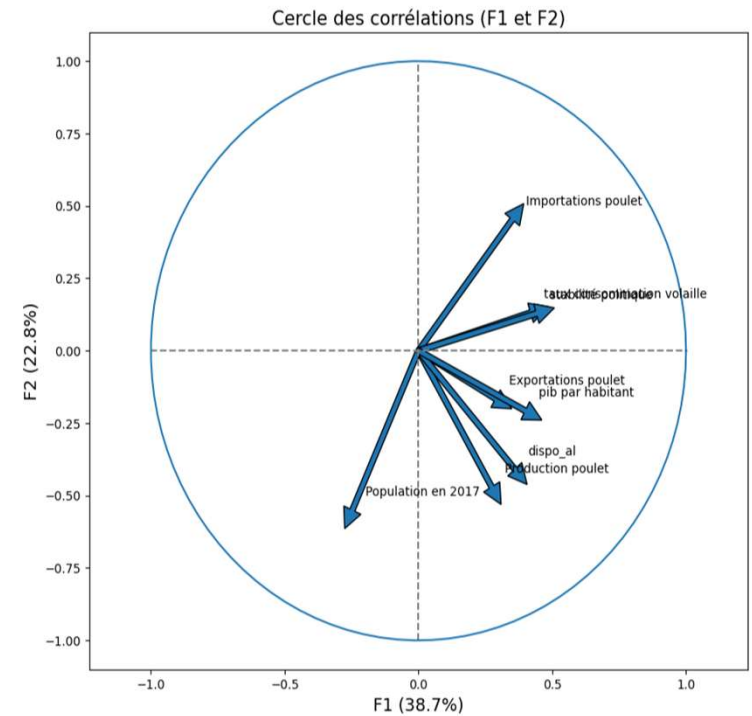
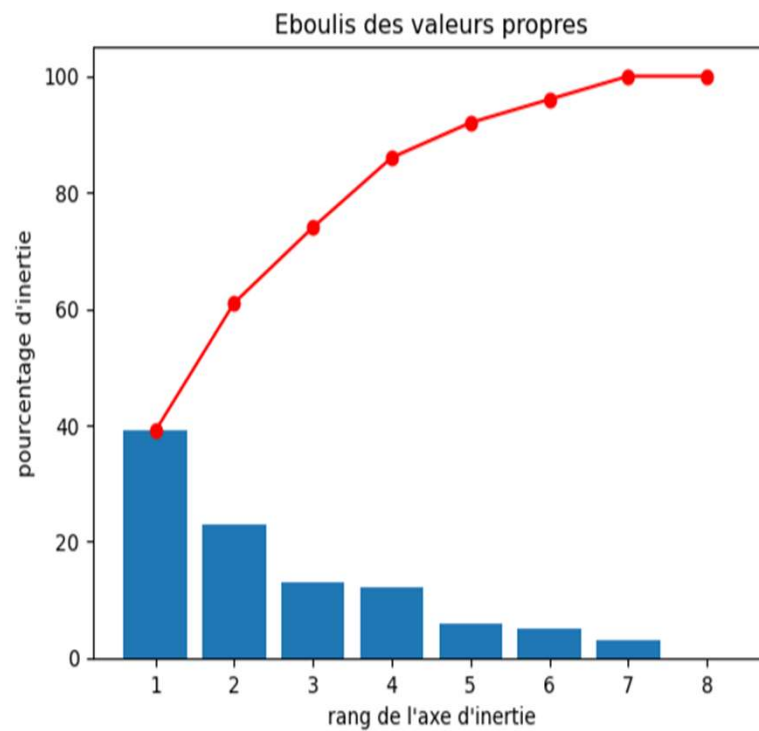
### • INCONVENIENTS :

- PERTE D'INFORMATIONS

## ANALYSE EN COMPOSANTES PRINCIPALES

*Plus la projection du nuage de points est étalée;  
Plus faible est la perte d'informations.*

*Pays étalés à droite présentent un meilleur PIB/habitant  
avec un potentiel en exportation du poulet élevé.*





## CLASSIFICATION ASCENDANTE HIÉRARCHIQUE (CAH)

La Classification Hiérarchique Ascendante (CAH) utilise la méthode de Ward et commence par traiter chaque point de données comme son propre cluster : 1 individu = 1 cluster

Ensuite, en suivant une approche itérative, les clusters sont fusionnés deux à deux de manière à minimiser la réduction de l'inertie Intra-cluster.

Le processus continue jusqu'à ce que toutes les données soient regroupées en un seul cluster. Les étapes de cette agglomération sont souvent visualisées à l'aide d'un *dendrogramme*.

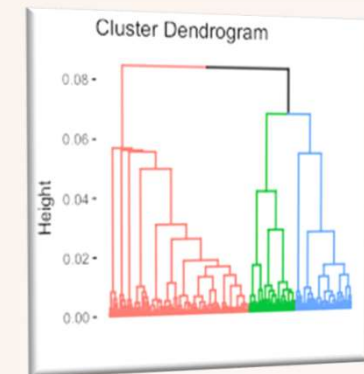
**DENDROGRAMME** : est une représentation graphique en forme d'arbre

Les individus = feuilles

Les branches = liaisons entre les individus

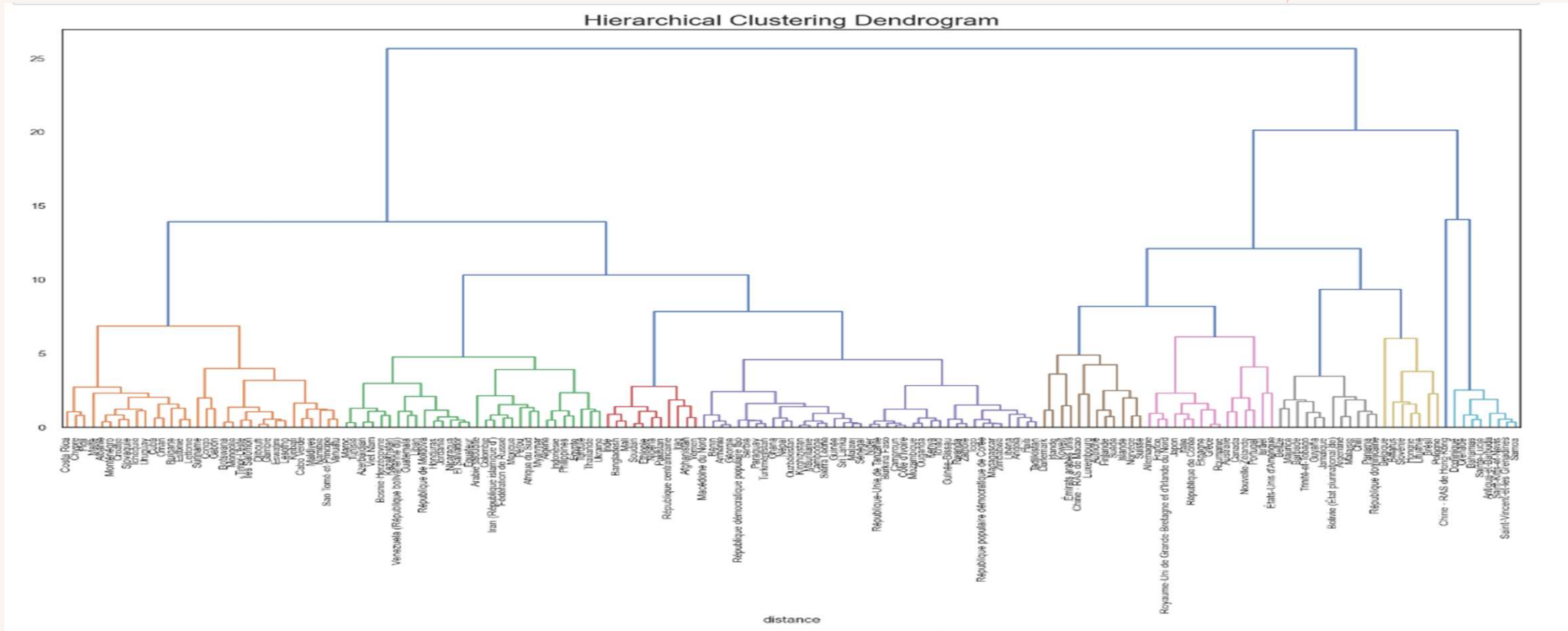
La distance entre les individus = la longueur de la branche

Les branches plus courtes indiquent une plus grande similitude entre les éléments qu'elles relient.



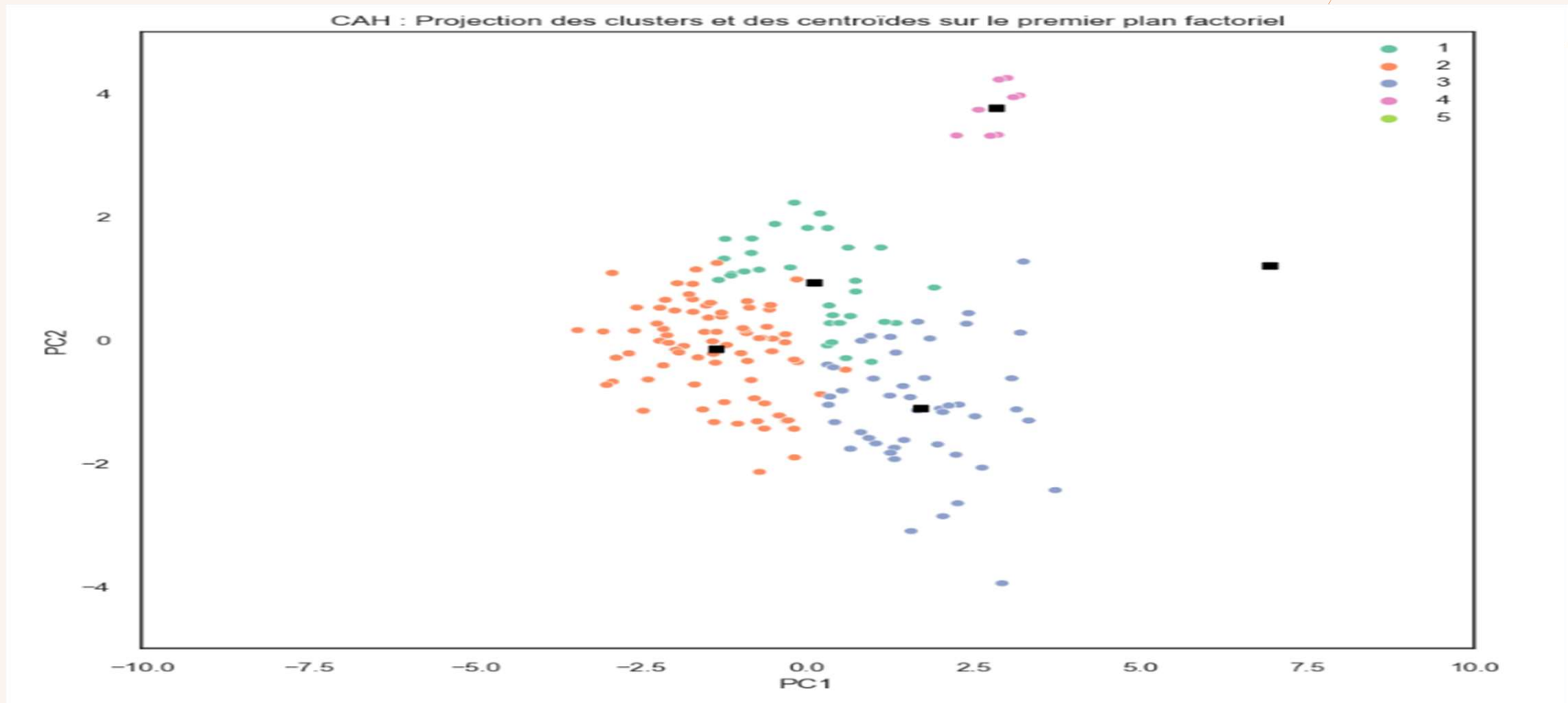
## PARTITIONNEMENT - CLASSIFICATION HIÉRARCHIQUE

Premier découpage en 5 groupes, un dendrogramme a été fait pour pouvoir identifier les pays les plus similaires entre eux.

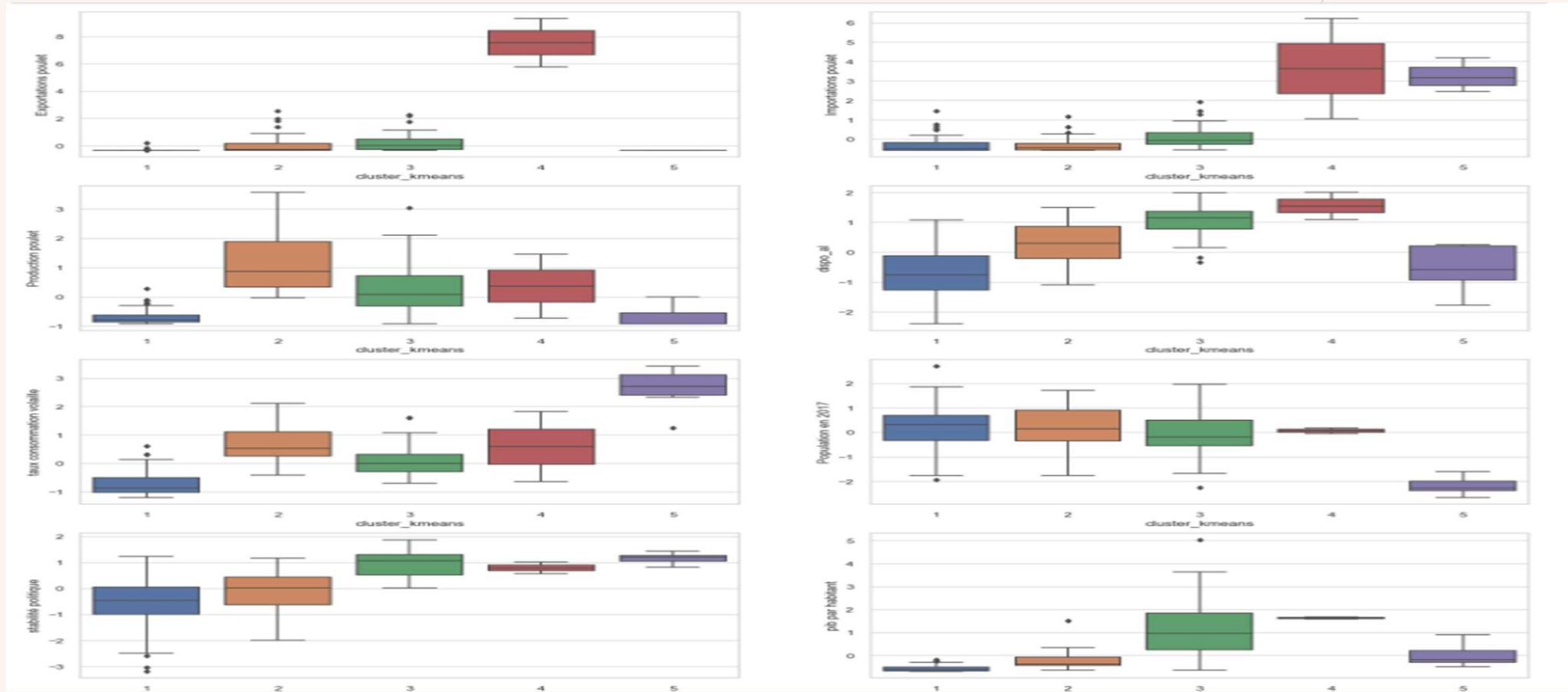




*Projection des 5 groupes issus du CAH.*



*BOXPLOT : Visualisation des groupes selon les variables.  
Ici un aperçu de toutes les variables*



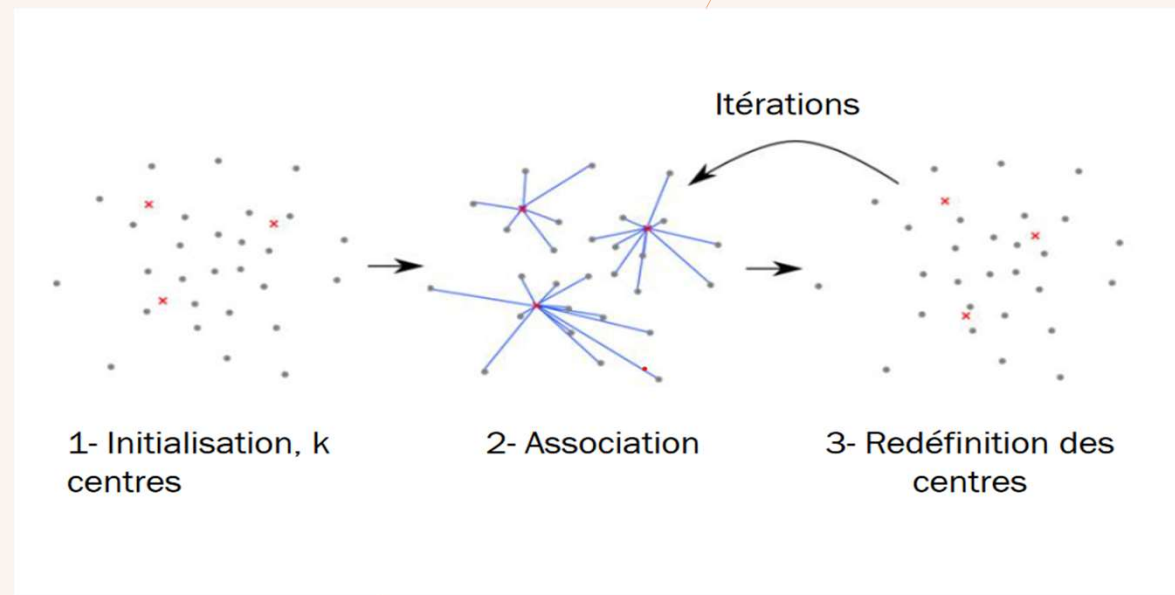
## Méthode des k-means

**k-means**: est une méthode de clustering non supervisée qui partitionne un ensemble de points en  $K$  groupes distincts en minimisant la somme des carrés des distances entre chaque point et le centre de son cluster attribué.

Le choix de  $K$ , le nombre de clusters, est souvent crucial et peut être évalué avec diverses méthodes comme la « méthode du coude » ou « le coefficient de silhouette ».

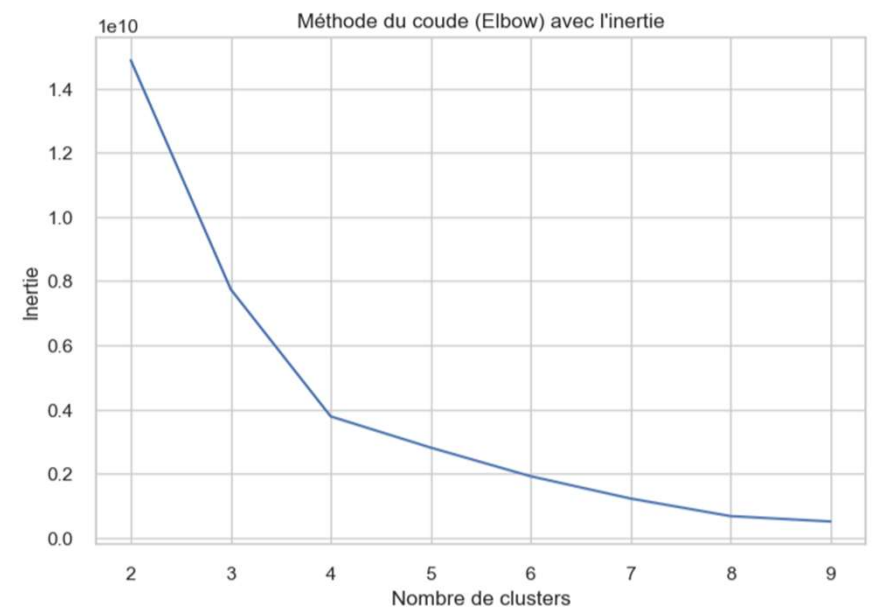
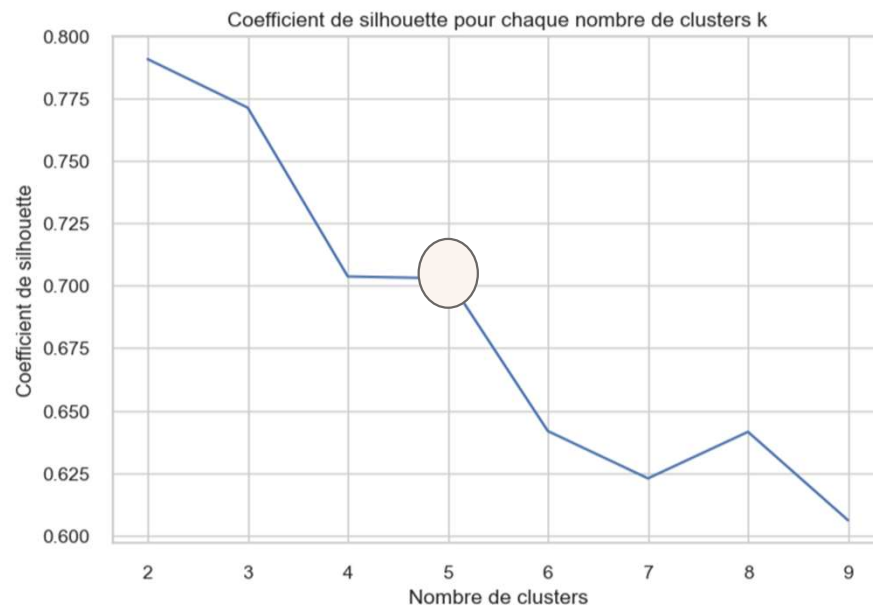
### Itérations :

- Association des points au centre le plus proche
- Calcul des nouveaux centres de gravité
- Association des points aux nouveaux centres
- Etc... jusqu'à ce que les centres ne bougent plus (= convergence).



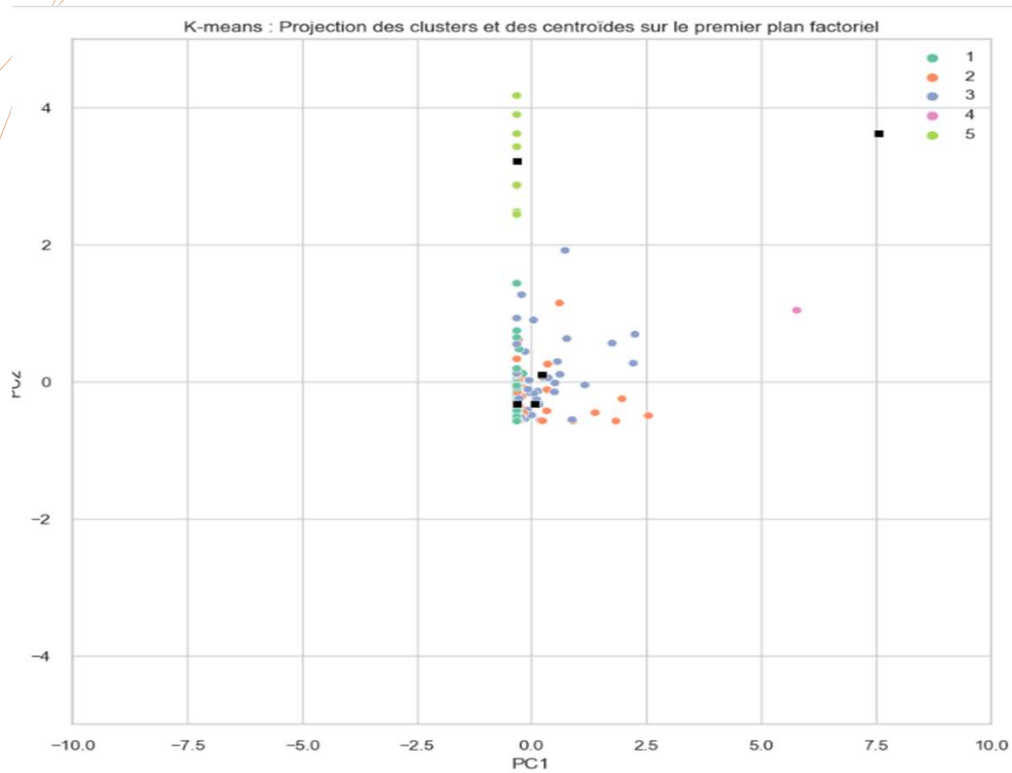
## Méthode des k-means

### CHOIX DE K=5 CLUSTERS

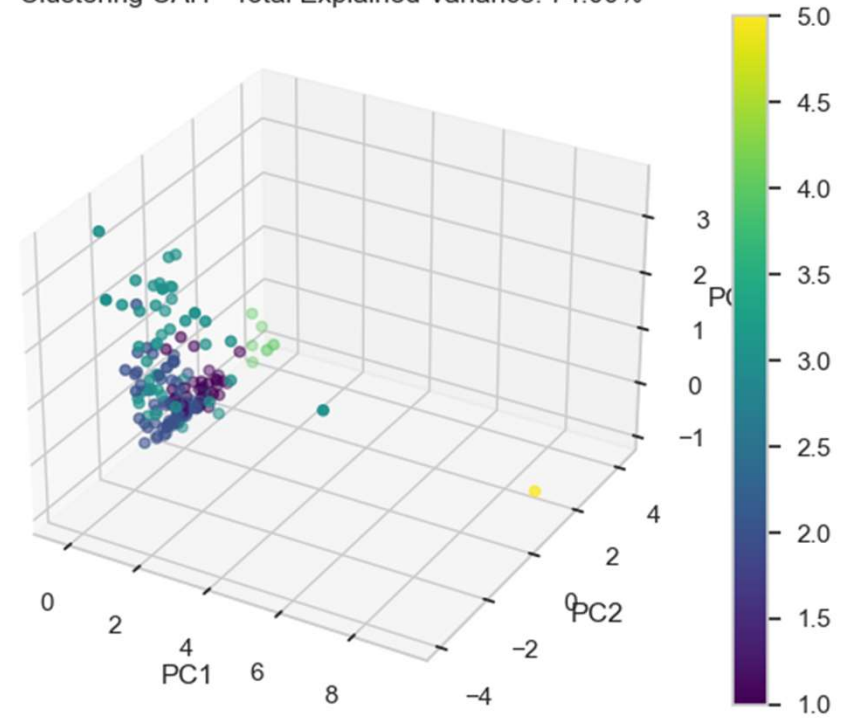


## Méthode des k-means

*Projection des 5 groupes issus De k-means sur le premier plan factoriel*



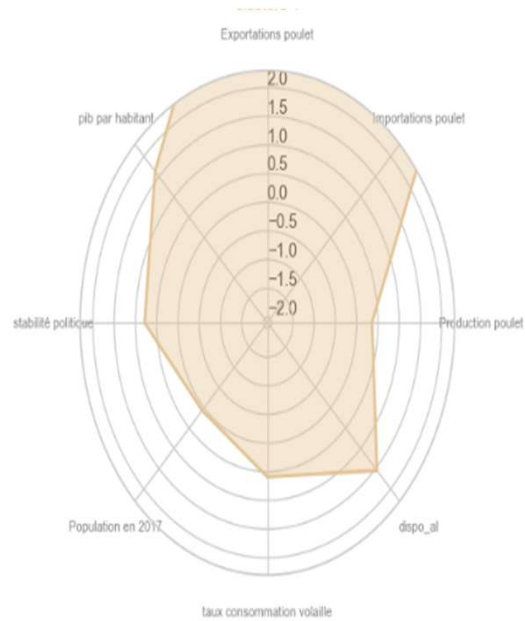
Clustering CAH - Total Explained Variance: 74.00%



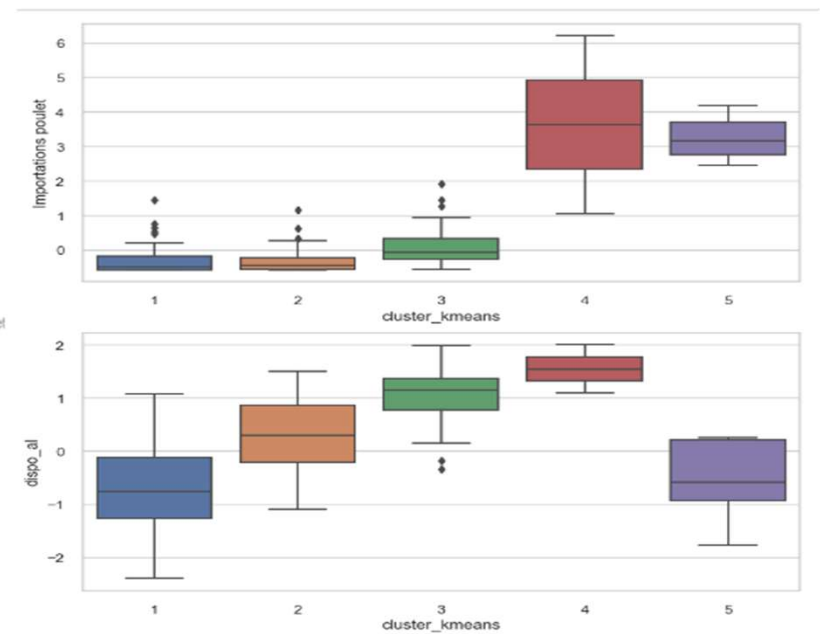
## Méthode des k-means

Visualisation des groupes selon les variables.  
Ici un aperçu selon 2 variables.

### RADAR CHARTS



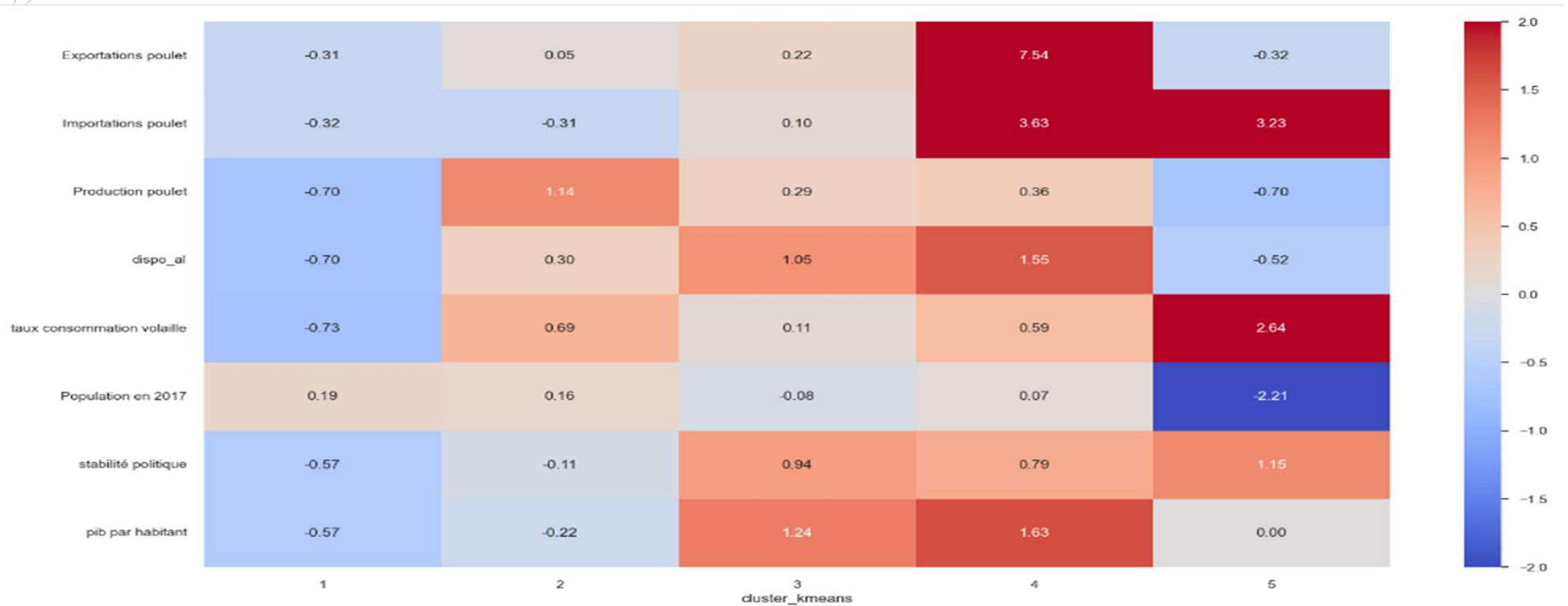
### BOXPLOTS



## Méthode des k-means

Visualisation des groupes selon les variables.

HEATMAP



## QUELLE MÉTHODE DE CLUSTERING CHOISIR ?

NOUS AVONS UTILISÉ LES INDICATEURS DE QUALITÉ **DAVIES-BOULDIN** ET LE **COEFFICIENT DE SILHOUETTE** POUR DECIDER QUEL SET DE CLUSTERS CHOISIR.

```
# Calcul de l'indice de Davies-Bouldin
print("\033[1mIndice de Davies-Bouldin:\033[0m")
print("Classification hiérarchique:", round(metrics.davies_bouldin_score(X_scaled,clusters), 3))
print("Kmeans:", round(metrics.davies_bouldin_score(X_scaled,labels+1), 3))

print("-----")

# Calcul du coefficient de silhouette
print("\033[1mCoefficient de silhouette:\033[0m")
print("Classification hiérarchique:", round(metrics.silhouette_score(X_scaled,clusters), 3))
print("Kmeans:", round(metrics.silhouette_score(X_scaled,labels+1), 3))
```

**\*\*Indice de Davies-Bouldin :** Plus cette valeur est faible, mieux c'est. Elle mesure la moyenne de la «similitude» entre chaque cluster et son cluster le plus similaire. La similitude est le rapport de la distance intra-cluster à la distance inter-cluster.\*\*

Classification hiérarchique: 1.086

Kmeans: 1.13

Interprétation : La Classification hiérarchique est presque similaire selon cet indice.

**\*\*Coefficient de Silhouette:** Les valeurs vont de -1 à 1. Une valeur élevée indique que l'objet est bien assorti à sa propre grappe et mal assorti aux grappes voisines.\*\*

Classification hiérarchique: 0.226

Kmeans: 0.316

Interprétation: Kmeans est nettement meilleur selon cette métrique.

**\*\*À partir de ces informations, on décide de poursuivre l'analyse avec les clusters du K-means.**



## Interprétation et décision

À l'aide du nuage de point, des radar charts, heatmap et des boxplot ci-dessus nous pouvons choisir **les pays cibles** :

**Le cluster 3** est recommandé pour l'exportation des volailles ainsi que Hong-Kong qui est inclus dans **le cluster 4**

les pays de ces deux clusters présentent:

- des stabilités politiques très élevées
- Leurs habitants sont des gros consommateurs de volailles (fortes disponibilités alimentaires).
- Ils ont des taux de dépendance aux importations acceptables relativement à d'autres groupes de pays.

```
cluster 3: 40 pays
Index(['Albanie', 'Allemagne', 'Australie', 'Autriche', 'Canada',
      'Chine - RAS de Macao', 'Croatie', 'Cuba', 'Danemark', 'Espagne',
      'Estonie', 'Finlande', 'France', 'Grèce', 'Irlande', 'Islande',
      'Italie', 'Japon', 'Kiribati', 'Koweït', 'Lettonie', 'Lituanie',
      'Luxembourg', 'Malte', 'Monténégro', 'Norvège', 'Nouvelle-Zélande',
      'Oman', 'Portugal', 'Roumanie',
      'Royaume-Uni de Grande-Bretagne et d'Irlande du Nord',
      'République de Corée', 'Slovaquie', 'Slovénie', 'Suisse', 'Suède',
      'Tchéquie', 'Uruguay', 'Émirats arabes unis', 'États-Unis d'Amérique'],
      dtype='object', name='pays')
```

```
cluster 4: 2 pays
Index(['Belgique', 'Chine - RAS de Hong-Kong'], dtype='object', name='pays')
```

## SYNTHÈSE

### Prochaines étapes :

- Identification d'autres pays cibles : Une analyse plus poussée pourrait être réalisée sur les clusters 3 et 4 pour identifier d'autres pays cibles (classification réalisée sur ces clusters uniquement)

Dans le cadre de l'analyse de marché :

- La proximité géographique des pays cibles à la France sera évaluée pour son influence sur les frais de transport, la durée de conservation des produits frais et l'empreinte écologique.
- L'existence ou l'absence d'accords de libre-échange sera également un facteur à considérer dans notre évaluation.
- Pour la visualisation des données et le suivi de l'étude, nous envisageons d'utiliser Tableau Software.



# MERCI

Samira MAHJOUB