



PRODUIRE UNE ÉTUDE DE MARCHE AVEC PYTHON

Saliou Ndiaye

Septembre 2023

Étant consultant Data Analyst chez La poule qui chante, une entreprise française d'agroalimentaire, mon mission consiste à effectuer une analyse d'un groupement de pays cibles pour une exportation des poulets à l'internationale.

- 1. PREPARATION ET NETTOYAGE DES DONNEES**
- 2. MÉTHODE DE CLASSIFICATION ASCENDANTE HIÉRARCHIQUE (CAH)**
- 3. MÉTHODE K-MEANS**
- 4. PROJECTION DU NUAGE DES INDIVIDUS SUR F1 F2 AVEC LES CLUSTERS CAH ET KMEANS POUR COMPARAISON**
- 5. ANALYSE DU CLUSTER SÉLECTIONNÉ**
- 6. ANALYSE DES COMPOSANTES PRINCIPALES (ACP)**
- 7. CONCLUSION**

☐ Les données de l'entreprise :

- Données Population (2000-2018)
- Données Disponibilité alimentaire (année 2017)

☐ Les données du site FAO rajoutées pour une analyse PESTEL

- PIB (année 2017)
- Stabilité politique

☐ Nouvelles variables à partir des données Disponibilité alimentaire (année 2017) :

- ***ratio proteine_animal/proteine_total*** : en calculant la part de protéine animale par rapport à la disponibilité de protéines en quantité (g/personne/jour) total par zone.
- ***Taux importations*** : en calculant le taux en importation de volaille par zone .
- ***Taux exportations*** : en calculant le taux d'autosuffisance en volaille par zone .

☐ Nouvelles variables à partir des données Population (2000-2018) :

- ***% évolution population 5 ans*** : en calculant le pourcentage de l'évolution de la population de chaque pays de 2012 à 2017 .

☐ Renommer certaines variables des données du PIB et Stabilité politique :

- ***PIB/Habitant \$***
- ***Indice stabilité***

❑ Jointure des dataframes

Après jointure des données , la dataframe complet a été nettoyé en :

- Insérant les valeurs manquantes des variables de certains pays après la jointure.
- Supprimer les pays sous dépendance de pays puissants .
- Supprimer les provinces de chine.

❑ Le dataframe complet contient 8 variables et 167 pays.

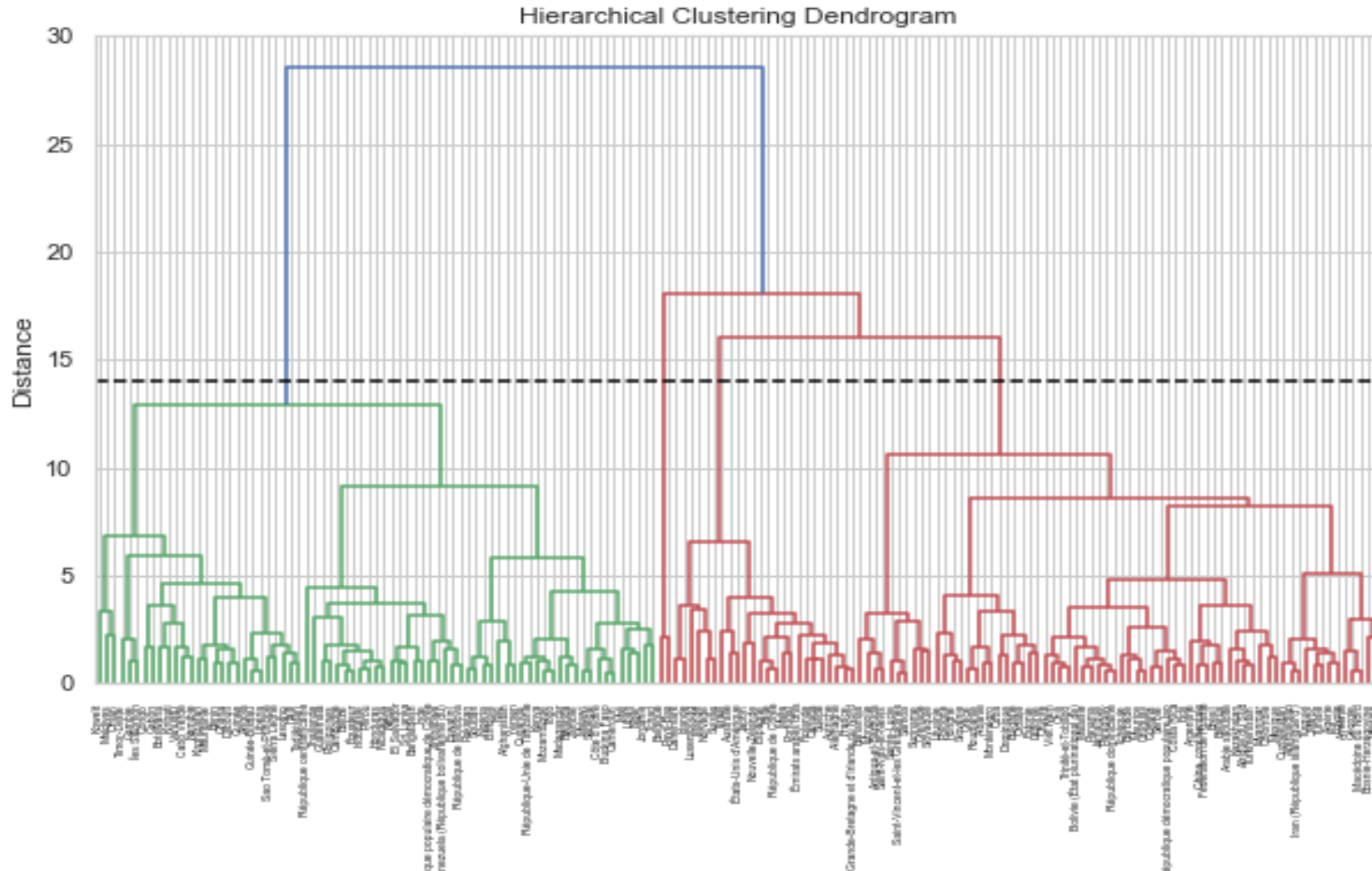
```
df.shape
```

```
(167, 8)
```

Zone	Taux importations	Taux autosuffisance	Disponibilité de protéines en quantité (g/personne/jour)	ratio proteine_animal/proteine_total	Disponibilité alimentaire (Kcal/personne/jour)	% evolution population 5 ans	PIB/Habitant \$	Indice stabilité
Afghanistan	50.88	0.00	54.09	8.39	1997.0	16.48	530.149831	-2.80
Afrique du Sud	24.27	2.97	83.36	33.84	2987.0	7.91	6723.928582	-0.28
Albanie	80.85	0.00	119.50	21.77	3400.0	-1.03	4521.752219	0.38
Algérie	0.72	0.00	92.85	10.48	3345.0	10.71	4134.936054	-0.92
Allemagne	48.42	37.15	104.07	33.33	3559.0	2.08	44670.222288	0.59

2. MÉTHODE DE CLASSIFICATION ASCENDANTE HIÉRARCHIQUE (CAH)

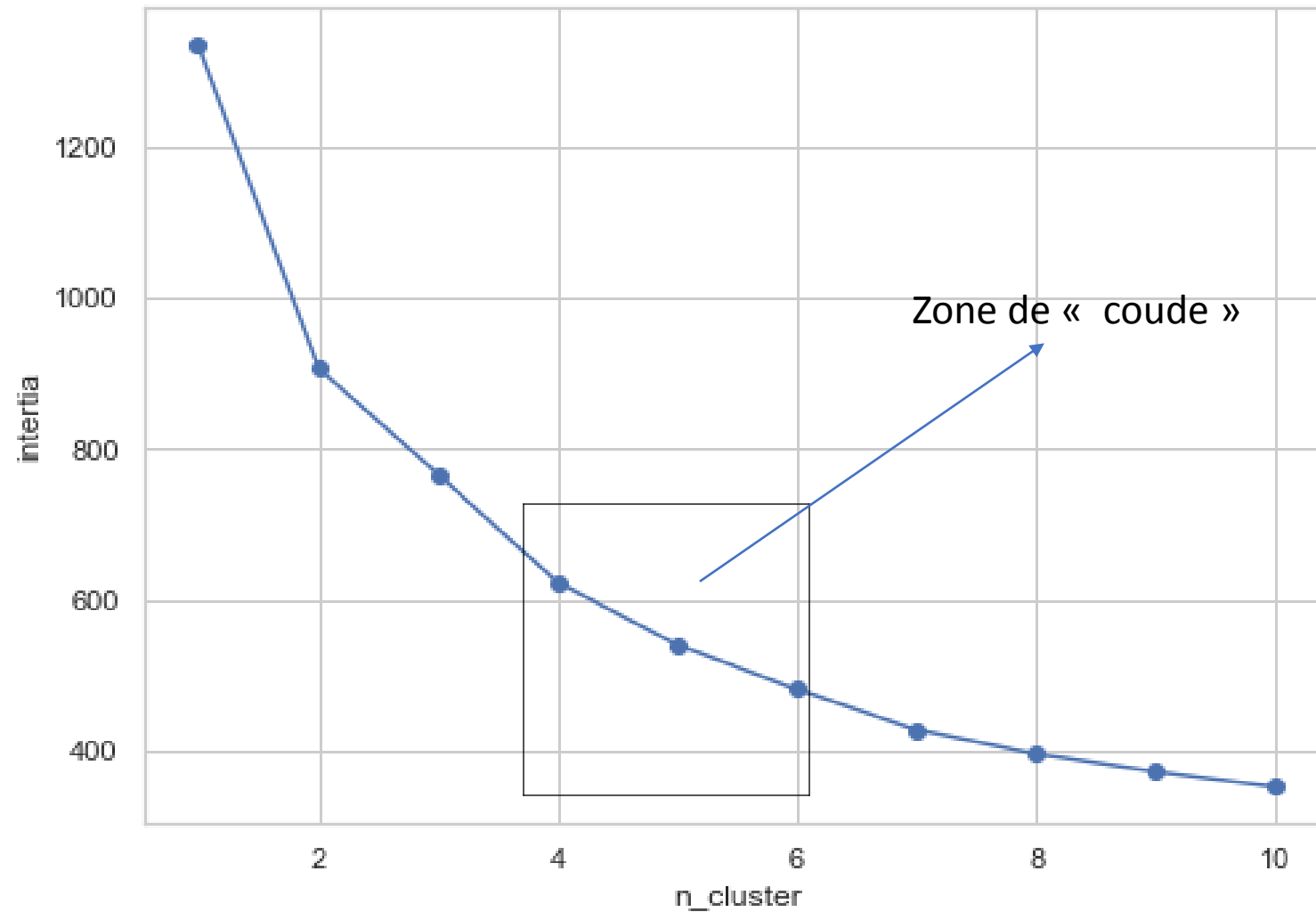
❑ Création du dendrogramme avec la méthode de ward



❑ Le dendrogramme montre que, on pourrait diviser notre arbre en **deux clusters** comme indiqué avec un cluster en vert et l'autre en rouge. On pourrait couper l'arbre pour obtenir **4 clusters**.

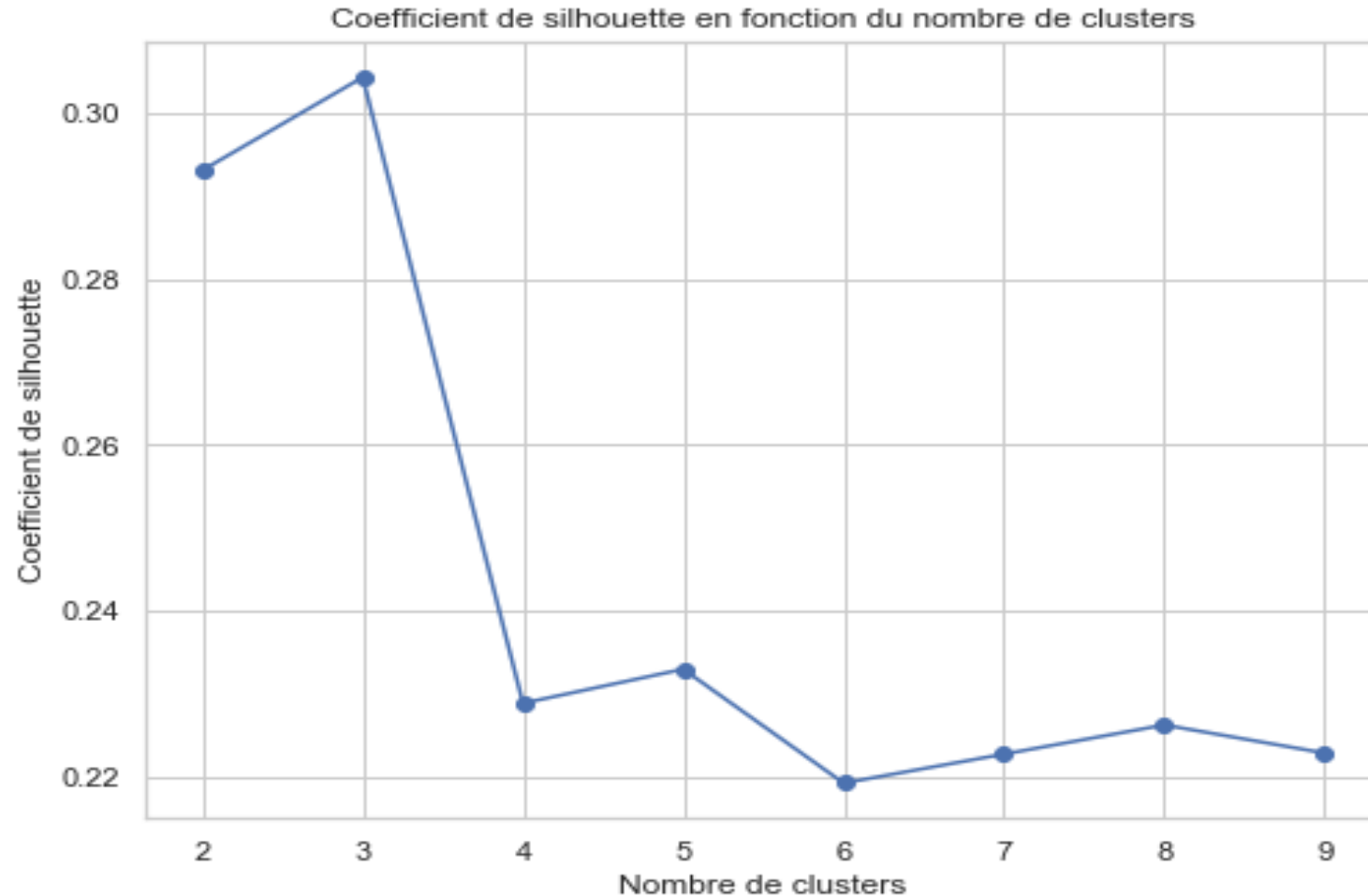
❑ Cependant, pour déterminer la “meilleure” valeur du nombre de cluster, on utilise la **méthode de coude et de silhouette**.

❑ la méthode du coude



- ❑ la décroissance de l'inertie intra-classe indique une zone de "coude" pour un nombre de clusters de **4 à 6**.

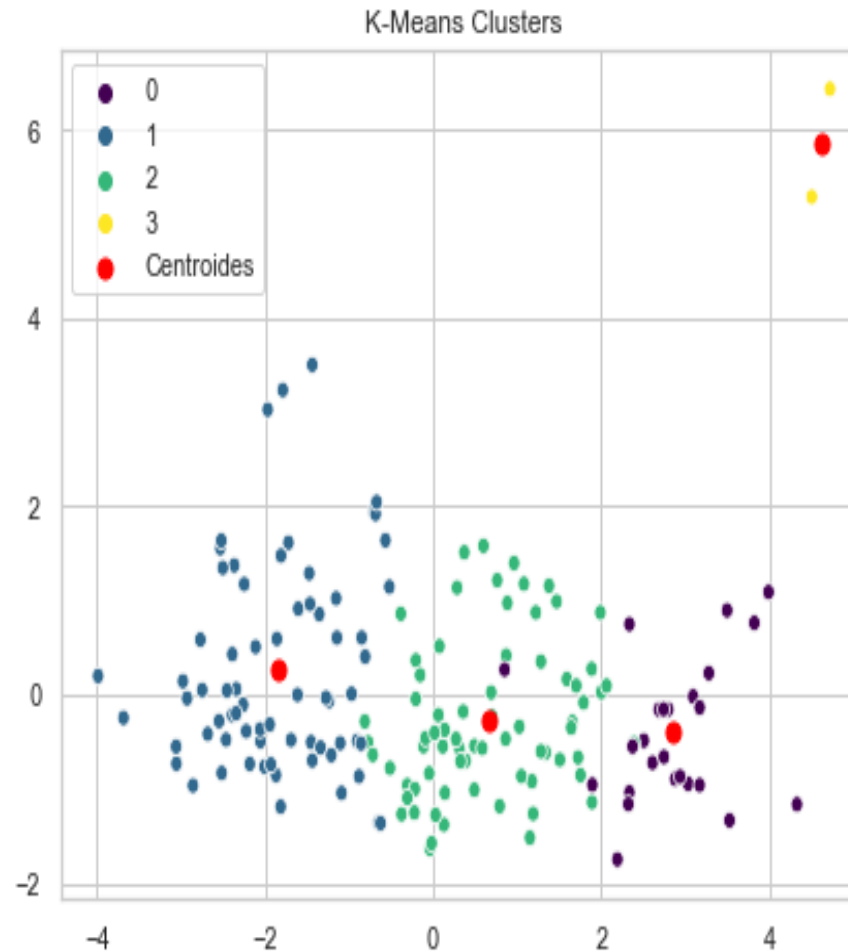
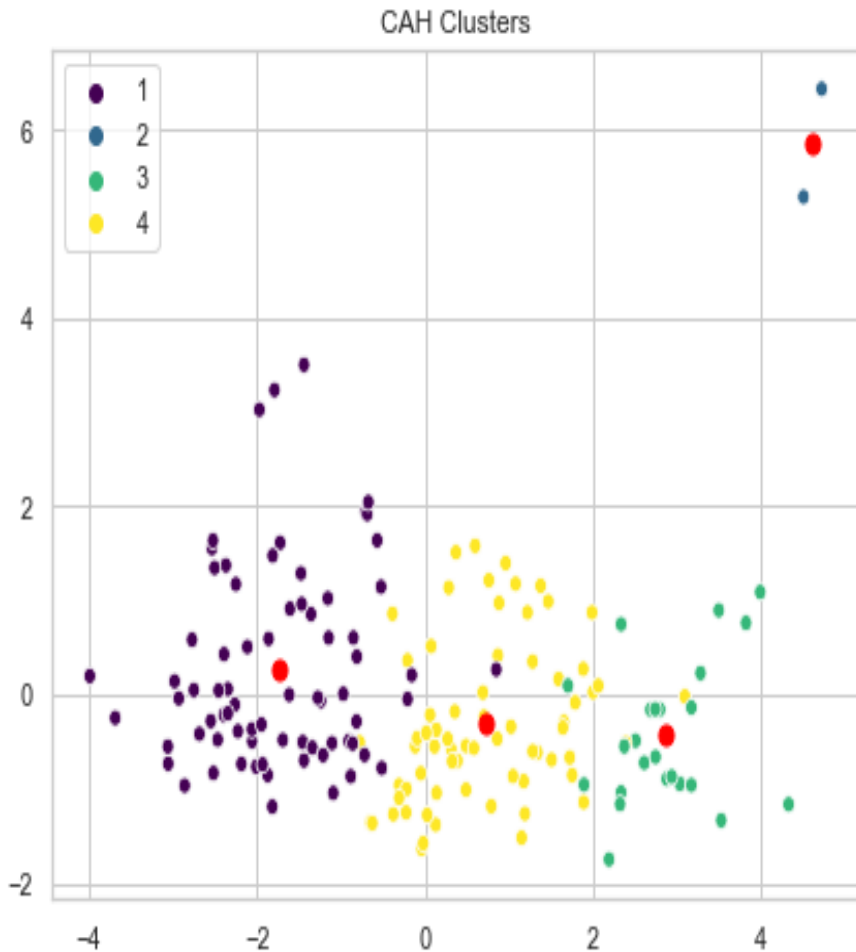
☐ Méthode de silhouette



☐ le plus haut coefficient de silhouette qui détermine le nombre idéal de cluster indique ici **3 cluster**.

4. Projection du nuage des individus sur F1 F2 avec les clusters CAH et Kmeans pour comparaison

□ Projeter sur les axes du PCA les CAH et Kmeans partition avec 4 clusters

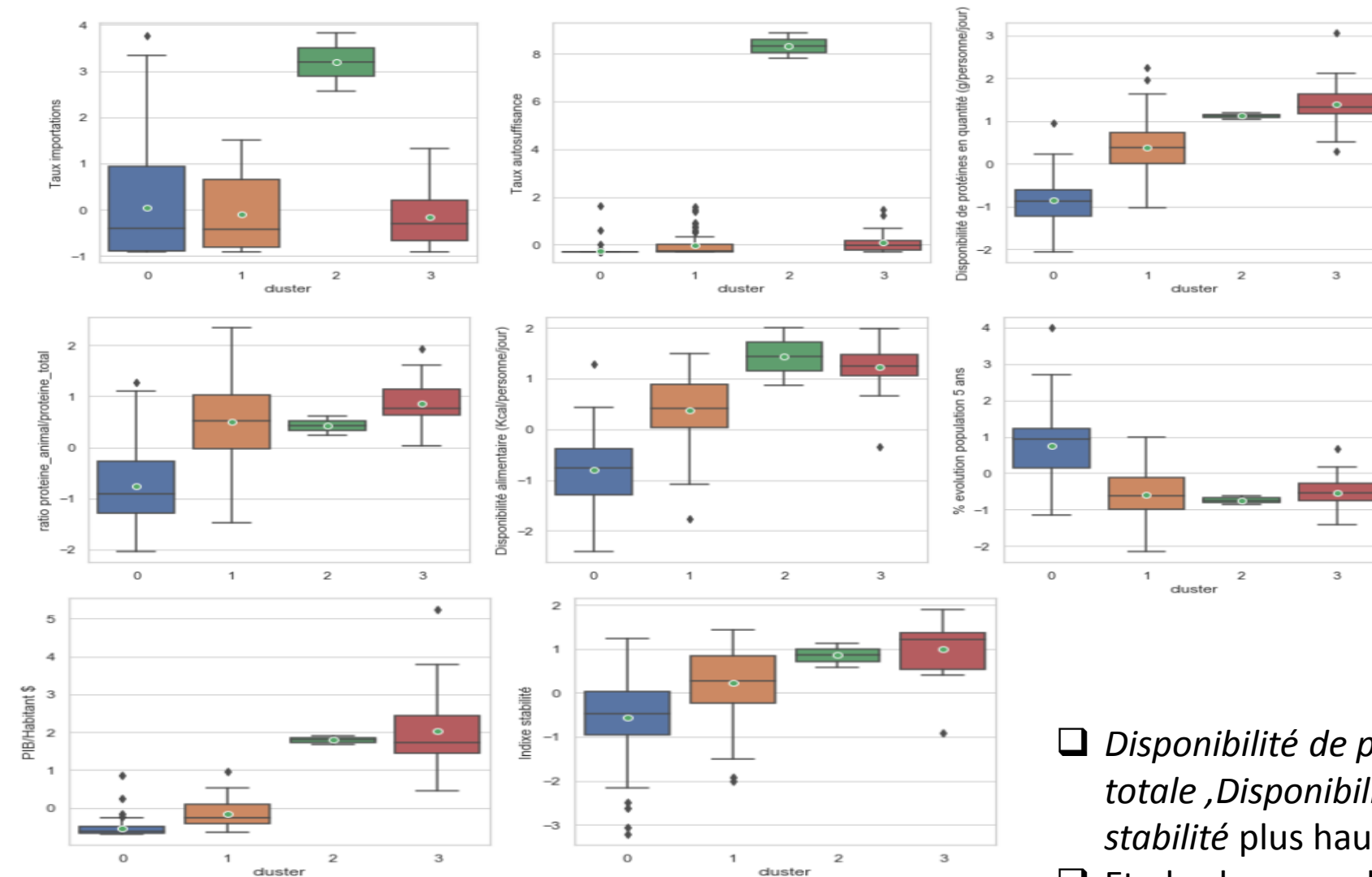


□ Les figures du nuage de individus sur les deux axes montrent :

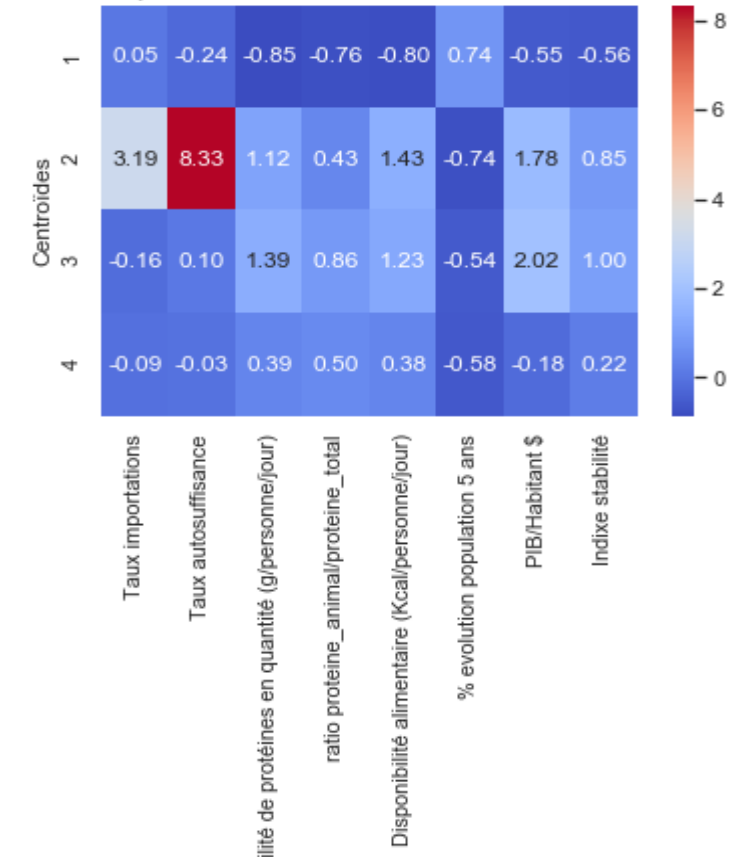
- le clustering semble assez similaire d'une méthode à l'autre
- ne se chevauchent pas sur les axes F1 et F2.
- Des centroïdes bien distants

5.Analyse des clusters

Les boxplot des différentes variables de chaque cluster avec CAH.



Heatmap entre centroïdes et variables initiales



- ❑ Disponibilité de protéine, ration protéine animal/protéine totale ,Disponibilité alimentaire, PIB par habitant et indice stabilité plus haut pour le **cluster 3**.
- ❑ Et plus bas pour les **clusters 1 et 2**.

5.Analyse des clusters (suite)

Cluster sélectionné

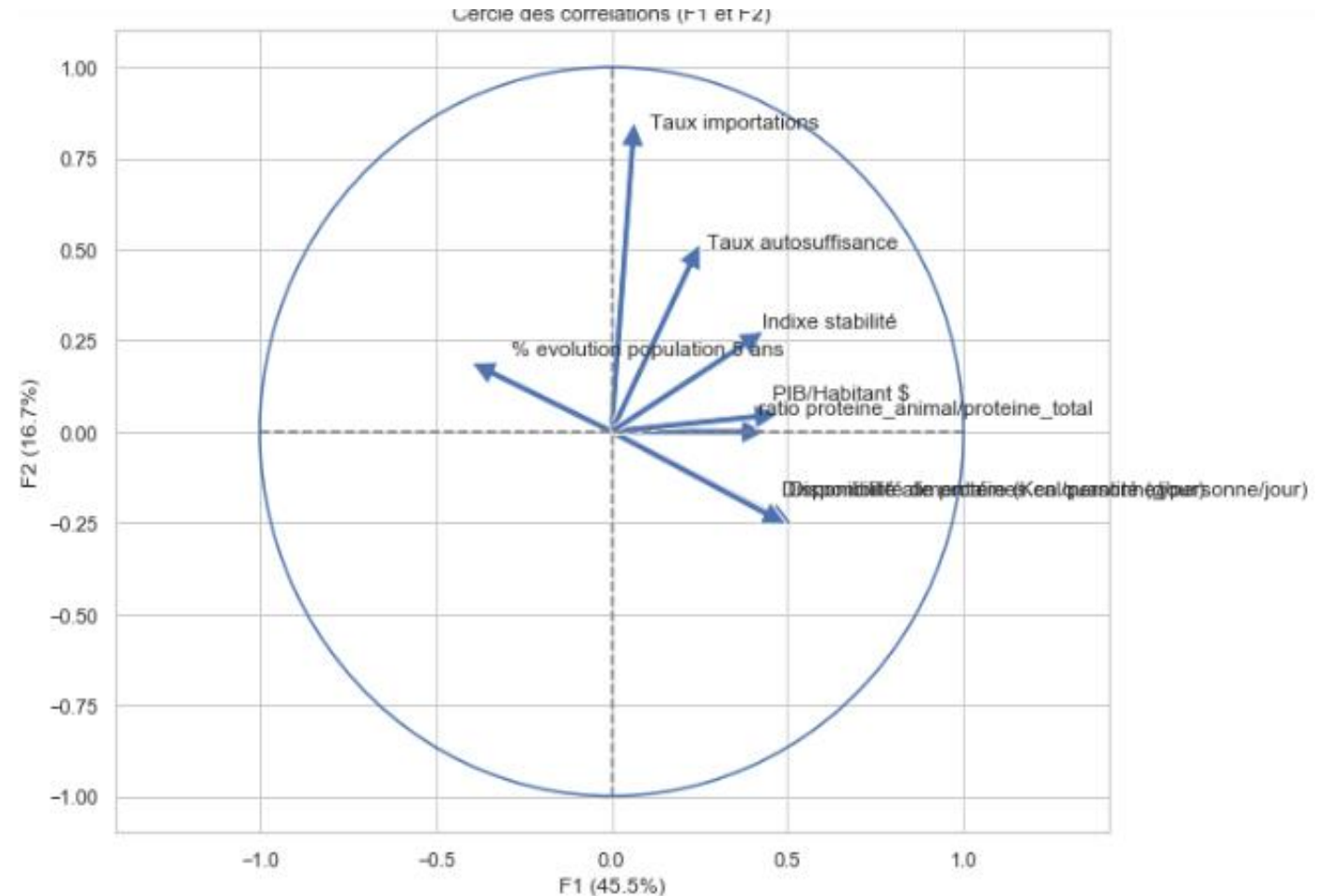
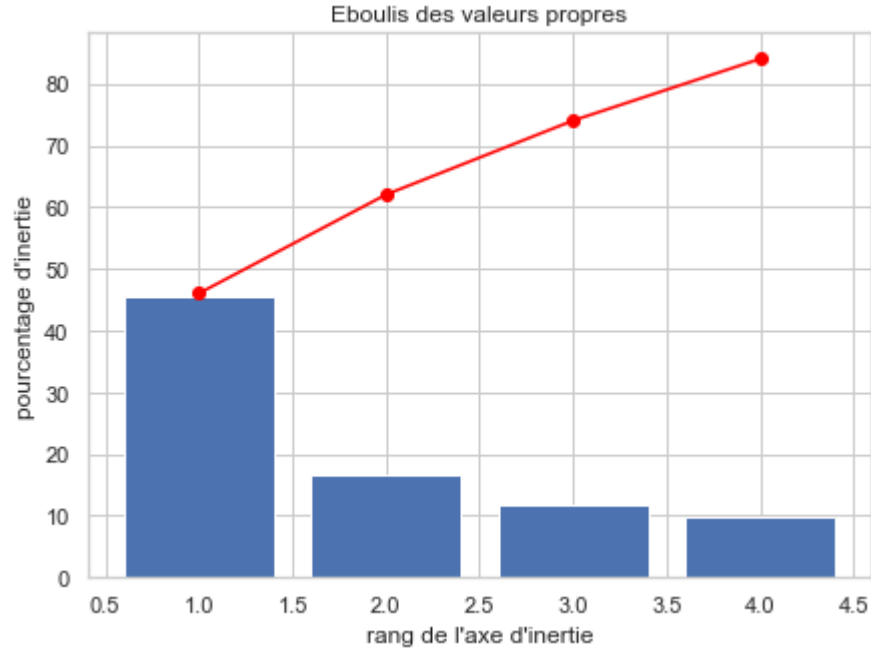
Zone	Taux importations	Taux autosuffisance	Disponibilité de protéines en quantité (g/personne/jour)	ratio proteine_animal/proteine_total	Disponibilité alimentaire (Kcal/personne/jour)	% evolution population 5 ans	PIB/Habitant \$	Indice stabilité
Allemagne	48.42	37.15	104.07	33.33	3559.0	2.08	44670.222288	0.59
Australie	1.37	3.59	108.01	46.66	3307.0	7.34	57430.803787	0.90
Autriche	63.58	45.09	108.11	34.38	3694.0	3.74	47429.535836	1.05
Canada	12.70	11.37	101.33	34.42	3492.0	5.18	45125.109795	1.10
Danemark	79.64	83.23	112.92	33.51	3383.0	2.16	57888.203512	0.87
Espagne	13.69	14.16	106.99	40.19	3270.0	-0.88	28190.806662	0.28
Finlande	14.41	10.81	117.77	30.02	3337.0	1.78	46412.760006	1.08
France	32.17	31.85	112.08	34.26	3556.0	2.01	40458.149875	0.28
Irlande	77.34	72.66	105.81	31.89	3714.0	3.15	70491.993301	1.00
Islande	20.00	0.00	141.34	31.52	3629.0	2.68	71961.532046	1.35
Israël	0.00	0.47	122.62	40.08	3507.0	8.26	42756.825069	-0.89
Italie	8.27	15.60	106.69	33.15	3522.0	1.33	32694.406219	0.31
Japon	44.27	0.41	86.52	35.18	2699.0	-0.72	38928.952602	1.11
Luxembourg	100.00	9.09	105.79	38.17	3334.0	11.50	110203.008202	1.33
Malte	64.29	0.00	116.32	34.16	3460.0	3.82	28122.287168	1.25
Norvège	1.96	0.00	109.98	30.43	3384.0	5.64	75490.688450	1.17
Nouvelle-Zélande	0.60	13.69	90.80	42.81	3153.0	5.23	43534.089383	1.60
Portugal	25.23	13.54	114.41	38.23	3494.0	-2.26	21475.355851	1.12
Royaume-Uni de Grande-Bretagne et d'Irlande du Nord	34.87	16.07	103.87	35.09	3428.0	3.41	40617.679075	0.39
République de Corée	16.04	0.70	96.97	33.93	3370.0	2.07	31524.942664	0.39
Suisse	38.35	3.01	95.46	31.32	3414.0	5.59	83353.597572	1.26
Suède	44.92	12.30	107.03	34.73	3214.0	3.79	53788.877502	0.98
Émirats arabes unis	105.10	22.82	107.26	26.42	3370.0	3.78	42522.377622	0.62
États-Unis d'Amérique	0.67	20.21	113.61	43.24	3762.0	3.52	59059.596403	0.29

❑ Le cluster choisi contient 24 pays.

❑ La majorité de ces pays ont :

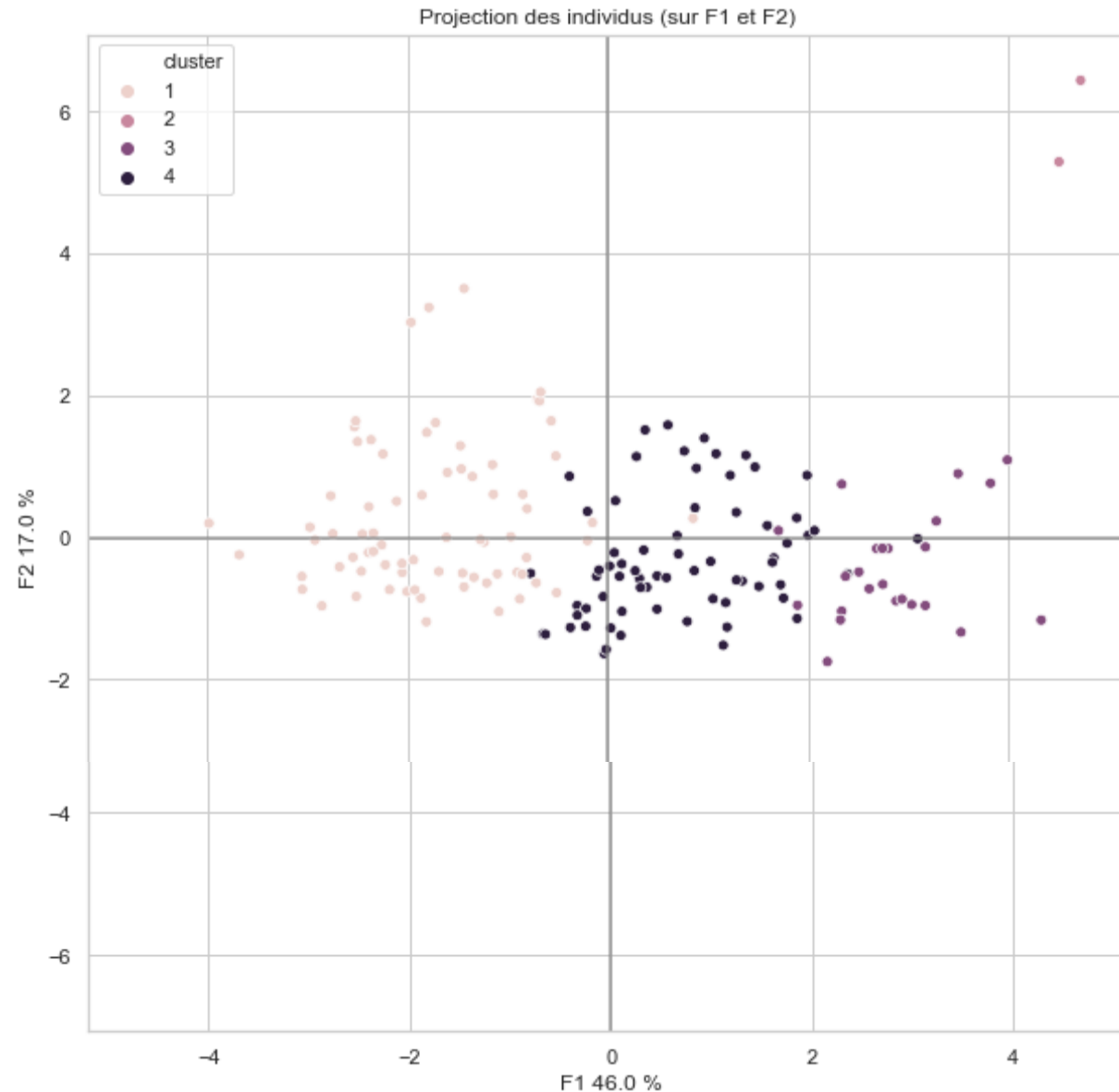
- *PIB*(pays riches) fort.
- *Taux d'autosuffisance* en volailles faible .
- *Indice stabilité* fort(pays stables).
- *ratio proteine_animal/protéine_total* élevée.

6. ANALYSE DES COMPOSANTES PRINCIPALES (ACP)



- ❑ Les 2 premières dimensions captent 63% de l'information.
- ❑ Les variables *Disponibilité protéine*, *ratio proteine_animal/proteine_total*, *PIB/habitant*, *Ratio_prot_alim*, *indice stabilité* sont corrélées positivement à la première dimension (F1).
- ❑ Les variables *Taux importations* et *Taux autosuffisance* sont fortement corrélées positivement à la deuxième dimension (F2).

6. ANALYSE DES COMPOSANTES PRINCIPALES (ACP) (suite)



- ☐ les individus(pays) qui sont situés sur la droite du graphique sont les pays avec ***PIB, stabilité politique et ratio proteine_animal/proteine total*** fort et sur le haut du graphique les pays importateur de volaille fort.
- ☐ Cela confirme la classification (CAH)

7. Conclusion

❑ Classement selon le pays qui a le plus de taux d'importation en volaille.

	Zone	Taux importations	Taux autosuffisance	Disponibilité de protéines en quantité (g/personne/jour)	ratio proteine_animal/proteine_total	Disponibilité alimentaire (Kcal/personne/jour)	% evolution population 5 ans	PIB/Habitant \$	Indice stabilité
13	Luxembourg	100.00	9.09	105.79	38.17	3334.0	11.50	110203.008202	1.33
4	Danemark	79.64	83.23	112.92	33.51	3383.0	2.16	57888.203512	0.87
8	Irlande	77.34	72.66	105.81	31.89	3714.0	3.15	70491.993301	1.00
14	Malte	64.29	0.00	116.32	34.16	3460.0	3.82	28122.287168	1.25
2	Autriche	63.58	45.09	108.11	34.38	3694.0	3.74	47429.535836	1.05
0	Allemagne	48.42	37.15	104.07	33.33	3559.0	2.08	44670.222288	0.59
21	Suède	44.92	12.30	107.03	34.73	3214.0	3.79	53788.877502	0.98
20	Suisse	38.35	3.01	95.46	31.32	3414.0	5.59	83353.597572	1.26
18	Royaume-Uni de Grande-Bretagne et d'Irlande du...	34.87	16.07	103.87	35.09	3428.0	3.41	40617.679075	0.39
7	France	32.17	31.85	112.08	34.26	3556.0	2.01	40458.149875	0.28
17	Portugal	25.23	13.54	114.41	38.23	3494.0	-2.26	21475.355851	1.12
9	Islande	20.00	0.00	141.34	31.52	3629.0	2.68	71961.532046	1.35
6	Finlande	14.41	10.81	117.77	30.02	3337.0	1.78	46412.760006	1.08
5	Espagne	13.69	14.16	106.99	40.19	3270.0	-0.88	28190.806662	0.28
11	Italie	8.27	15.60	106.69	33.15	3522.0	1.33	32694.406219	0.31
15	Norvège	1.96	0.00	109.98	30.43	3384.0	5.64	75490.688450	1.17
10	Israël	0.00	0.47	122.62	40.08	3507.0	8.26	42756.825069	-0.89

- ❑ Les pays européens à privilégier.
- ❑ On peut sélectionner les pays ayant plus de **Taux d'importation** en volaille.

MERCI