



La poule qui chante

Etude de marché



Hello !

Je suis Lokman

Je suis data analyst pour La poule qui chante...



Le contexte

La poule qui chante est une entreprise française qui commercialise de la viande de volaille, elle souhaite s'étendre à l'internationale.

Ma mission

Trouver le ou les pays les plus propices à cette expansion en considérant plusieurs facteurs

Ma méthode

Je vais faire le tour du monde !



01

Tour du monde

1. Population

La Chine :

- 1,4 Milliard d'habitant en 2017



1. Population

2. Croissance démographique

L'Inde :

- 1,35 Milliards d'habitants en 2017
- Augmentation de la population de 9%
entre 2010 et 2017



1. Population

2. Croissance démographique

3. Disponibilité alimentaire

Le Brésil :

- 208 Millions d'habitants en 2017
- Augmentation de la population de 7%
entre 2010 et 2017
- 55kg par personne par an



1. Population
2. Croissance démographique
3. Disponibilité alimentaire

4. PIB par habitant

Les Etats Unis :

- 329 Millions d'habitants en 2017
- Augmentation de la population de 6% entre 2010 et 2017
 - 50kg par personne par an
 - 60 000\$ de PIB par habitant



1. Population
2. Croissance démographique
3. Disponibilité alimentaire
4. PIB par habitant

5. Importation

Le Mexique:

- 122 Millions d'habitants en 2017
- Augmentation de la population de 8% entre 2010 et 2017
 - 33kg par personne par an
- 20 000\$ de PIB par habitant
- 972 000 tonnes de viande de volaille importées par an



1. Population
2. Croissance démographique
3. Disponibilité alimentaire
4. PIB par habitant
5. Importation

6. Dépendance à l'importation

L'Arabie Saoudite :

- 34 Millions d'habitants en 2017
- Augmentation de la population de 14% entre 2010 et 2017
 - 38kg par personne par an
 - 47 000\$ de PIB par habitant
- 744 000 tonnes de viande de volaille importées par an
- 1,37 points d'indice de dépendance à l'importation



1. Population
2. Croissance démographique
3. Disponibilité alimentaire
4. PIB par habitant
5. Importation
6. Dépendance à l'importation

7. Stabilité

Le Japon :

- 126 Millions d'habitants en 2017
- Stagnation de la population entre 2010 et 2017
 - 18kg par personne par an
 - 41 000\$ de PIB par habitant
- 1 000 000 de tonnes de viande de volaille importées par an
- 0,48 points d'indice de dépendance à l'importation
- 1,11 points d'indice de stabilité politique



Population
Croissance démographique
Disponibilité alimentaire
PIB par habitant
Importation
Dépendance à l'importation
Stabilité





02

Clustering

Un peu de théorie :

Les méthodes de clustering sont un type d'analyse de données qui vise à regrouper des données en groupes ou « clusters ». Ces groupes partagent des caractéristiques communes et se distinguent des autres.

Classification Ascendante Hiérarchique

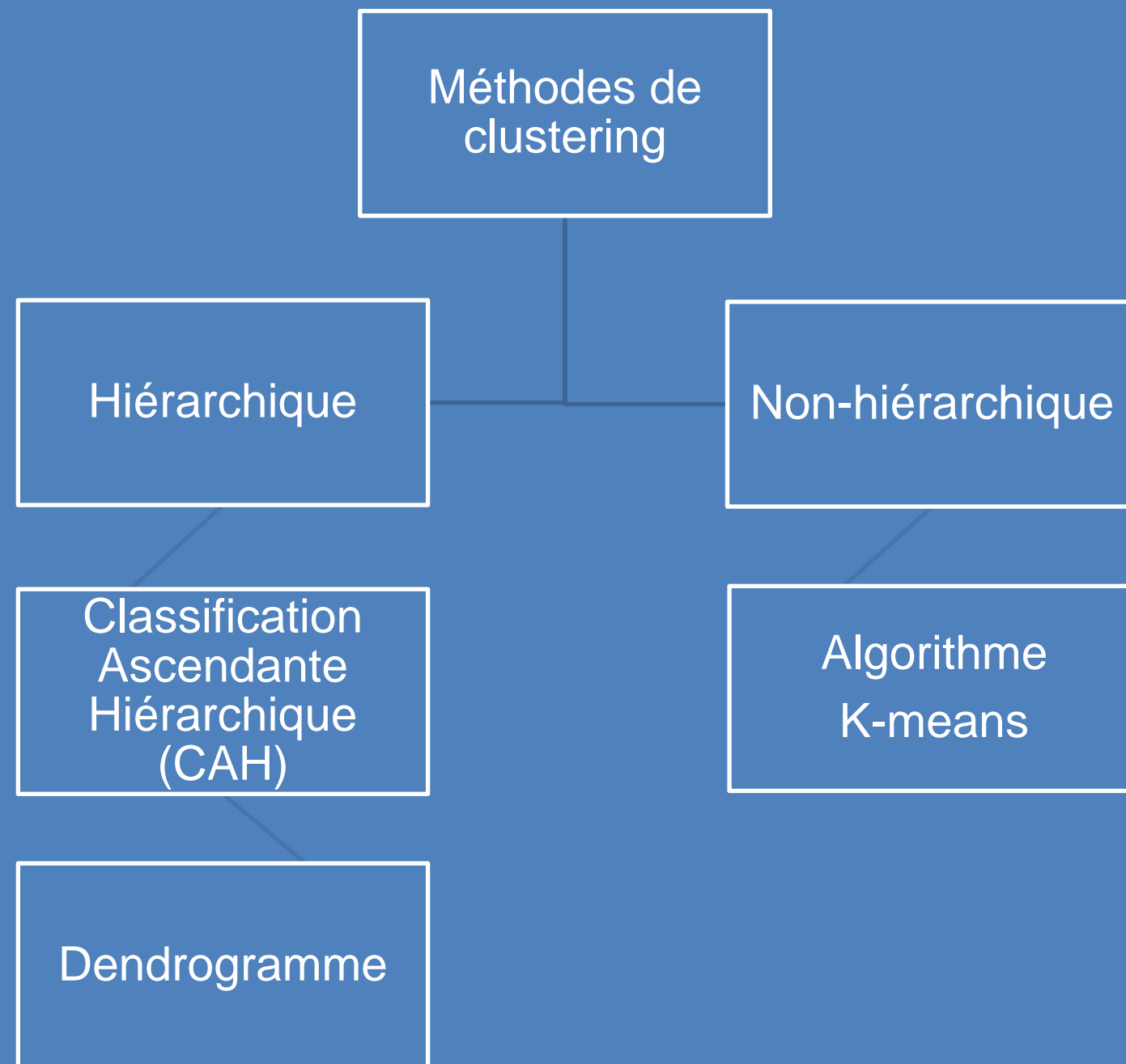
La Classification Ascendante Hiérarchique (CAH) consiste à regrouper des données en un certain nombre de groupes hiérarchiques selon des critères prédéfinis.

Algorithme K-means

La méthode K-Means est une méthode couramment utilisée de clustering, qui se concentre sur les groupes de données qui partagent le même centre.

Analyse en Composante Principale

Enfin, l'Analyse en Composantes Principales (ACP) est une méthode utilisée pour extraire des informations intéressantes à partir de grandes quantités de données. Elle réduit la dimension des données afin d'identifier les variables corrélées.



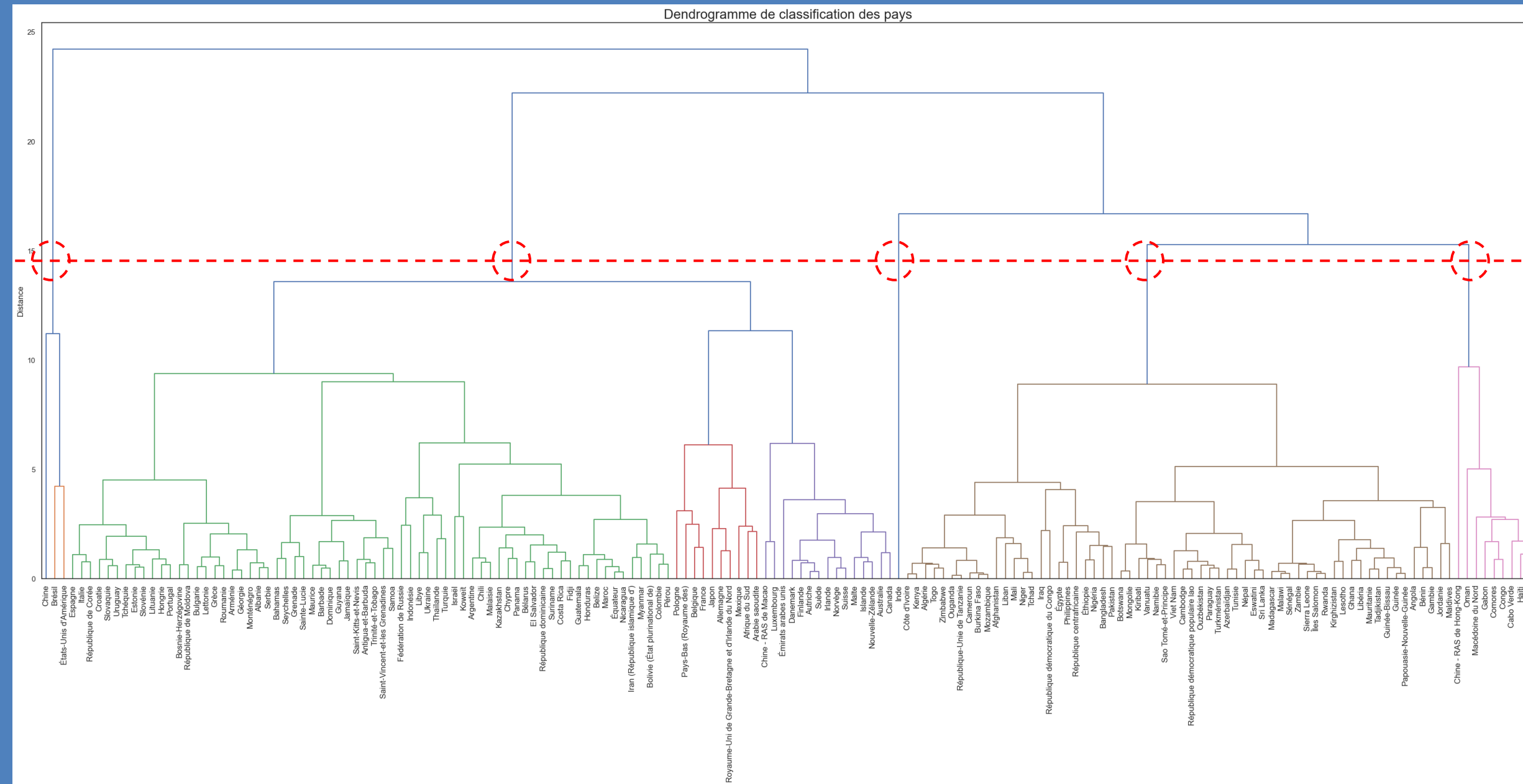


Classification Ascendante Hiérarchique

Classification Ascendante Hiérarchique

Dendrogramme

Choix de découper nos individus en 5 classes distinctes (repère rouge)

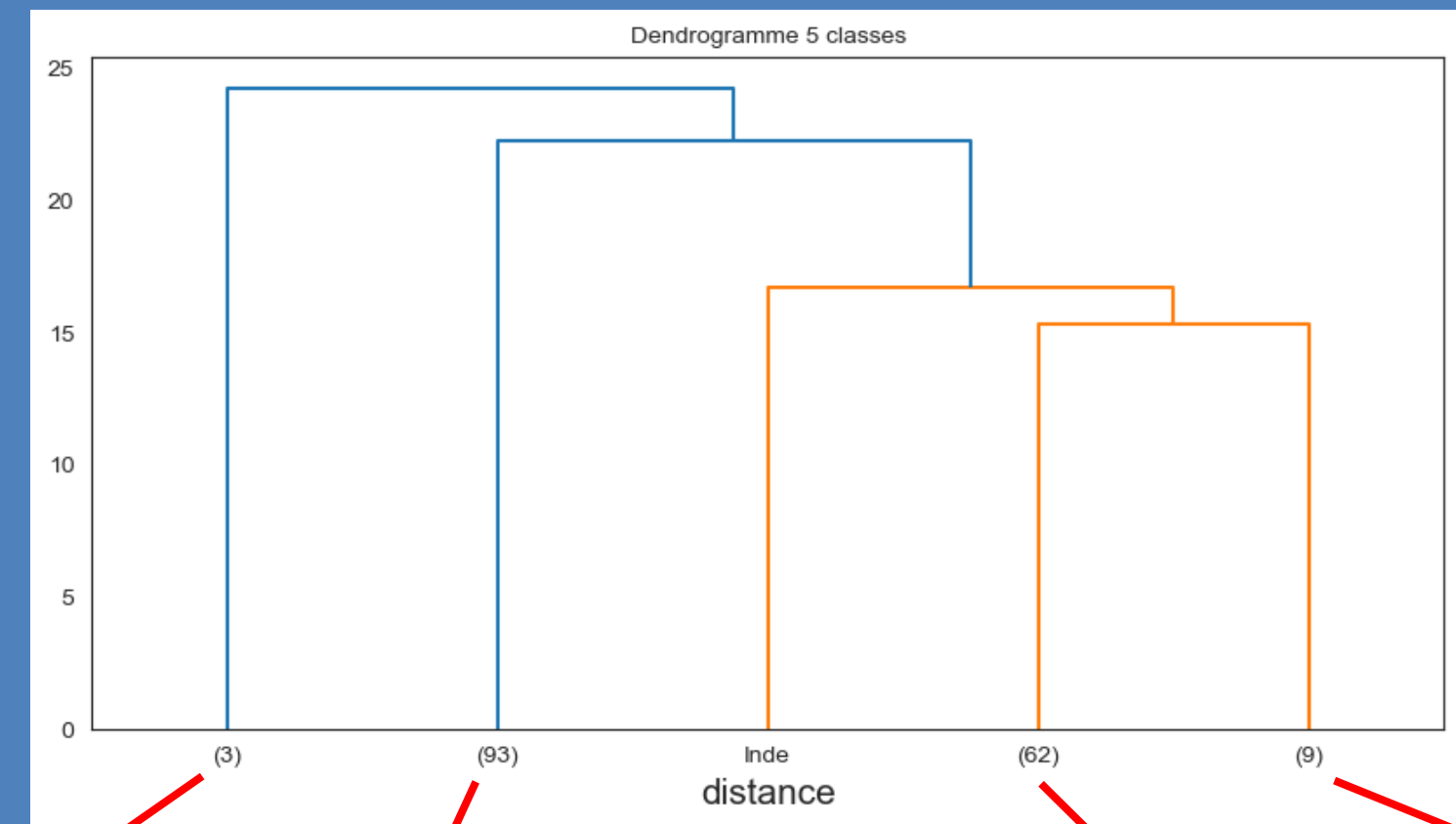


Classification Ascendante Hiérarchique

Dendrogramme 5 classes

Affichage des 5 clusters obtenus

- Clusters déséquilibrés
 - Entre 1 et 93 pays
 - L'Inde est un cluster



classe1

pays	
23	Brésil
49	États-Unis d'Amérique
167	Chine

classe2

pays	
1	Afrique du Sud
2	Albanie
4	Allemagne
6	Antigua-et-Barbuda
7	Arabie saoudite
...	...
154	Thaïlande
157	Trinité-et-Tobago
160	Turquie
161	Ukraine
162	Uruguay

classe5

pays	
69	Inde

classe3

pays	
26	Cabo Verde
31	Chine - RAS de Hong-Kong
35	Comores
36	Congo
55	Gabon
65	Haïti
92	Macédoine du Nord
114	Oman
155	Timor-Leste

classe4

pays	
0	Afghanistan
3	Algérie
5	Angola
12	Azerbaïdjan
14	Bangladesh
...	...
159	Turkménistan
163	Vanuatu
164	Viet Nam
165	Zambie
166	Zimbabwe

Analyse des classes

Visualisation des distributions en boxplot de chaque variable par classe

Classe 1

Pays très gros producteurs et exportateurs de viande de volaille

Classe 2

Pays riches stables qui consomment et importent de la viande de volaille en quantité

Classe 3

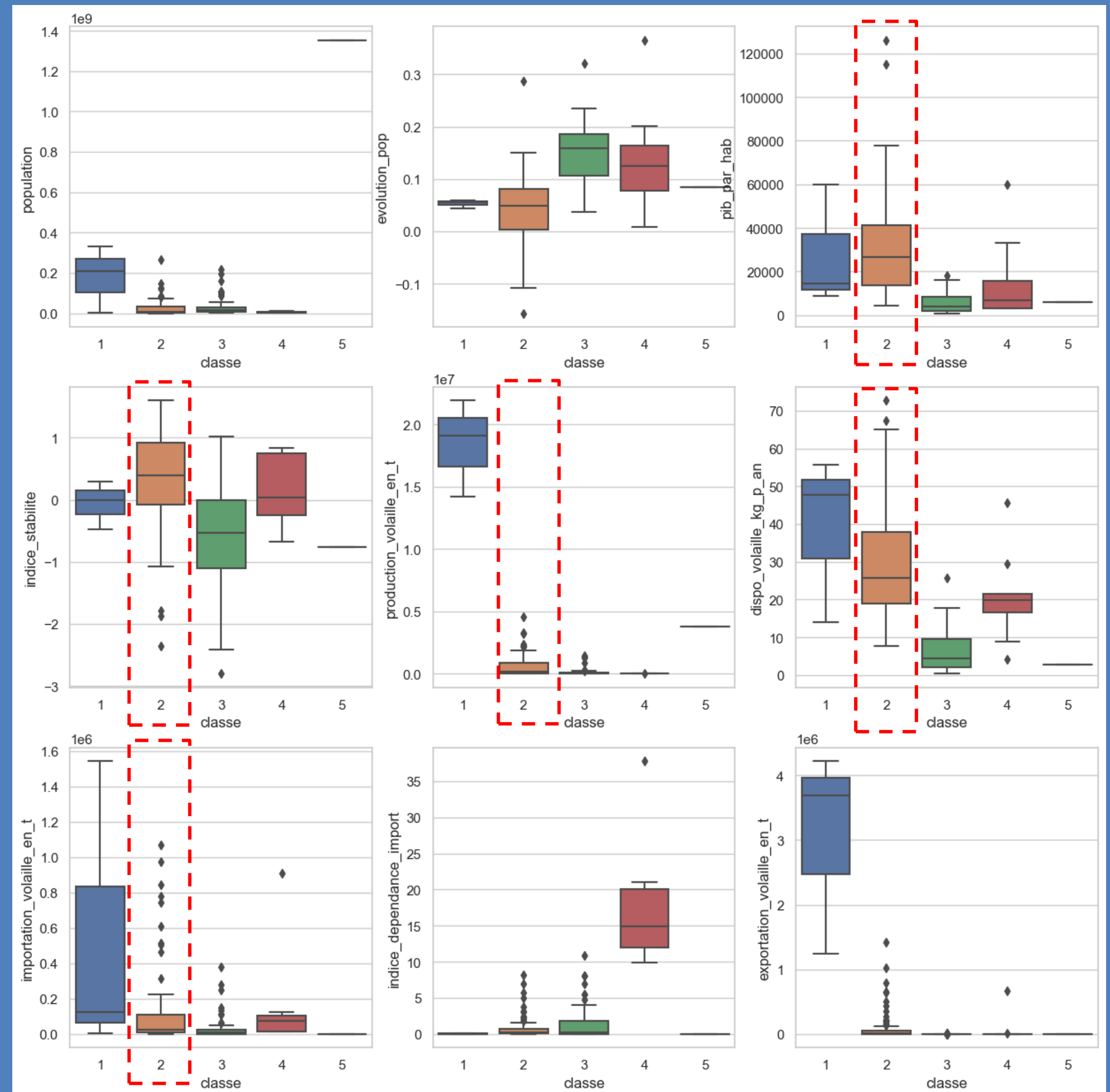
Pays qui consomment mais n'exportent et n'importent pas

Classe 4

Pays pauvres qui consomment peu mais dépendent beaucoup des importations

Classe 5

Inde, pays très peuplé producteur mais pas importateur



Classification Ascendante Hiérarchique

Classement des pays

Création d'un score à partir des valeurs pondérées
de chaque variable

	pays	score
0	Japon	59.705019
1	Mexique	54.530161
2	Allemagne	48.238815
3	Royaume-Uni de Grande-Bretagne et d'Irlande du...	45.718630
4	Arabie saoudite	42.088397
5	Indonésie	39.959683
6	Émirats arabes unis	38.605685
7	France	35.672071
8	Pays-Bas (Royaume des)	34.765680
9	Fédération de Russie	34.746038
10	Afrique du Sud	34.682957
11	Chine - RAS de Macao	29.737881
12	Canada	27.351340
13	Koweït	26.904297
14	Espagne	24.918073
15	Belgique	24.110028
16	République de Corée	23.561802
17	Grenade	22.981591
18	Australie	22.507534
19	Italie	22.329417



Algorithme K-means

Algorithme K-means

Détermination de K

La recherche de K détermine le nombre de clusters de notre analyse

Coefficient de silhouette

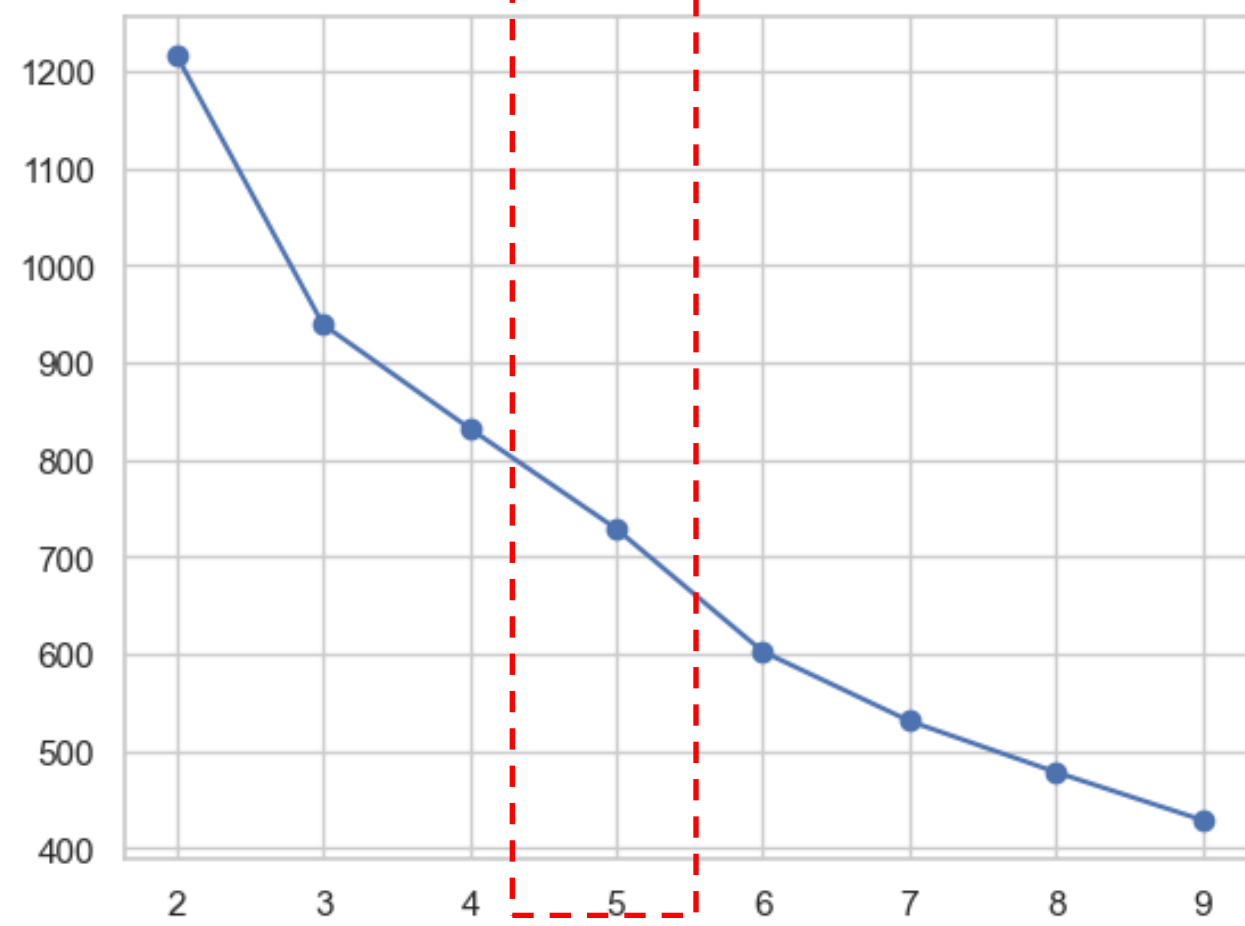
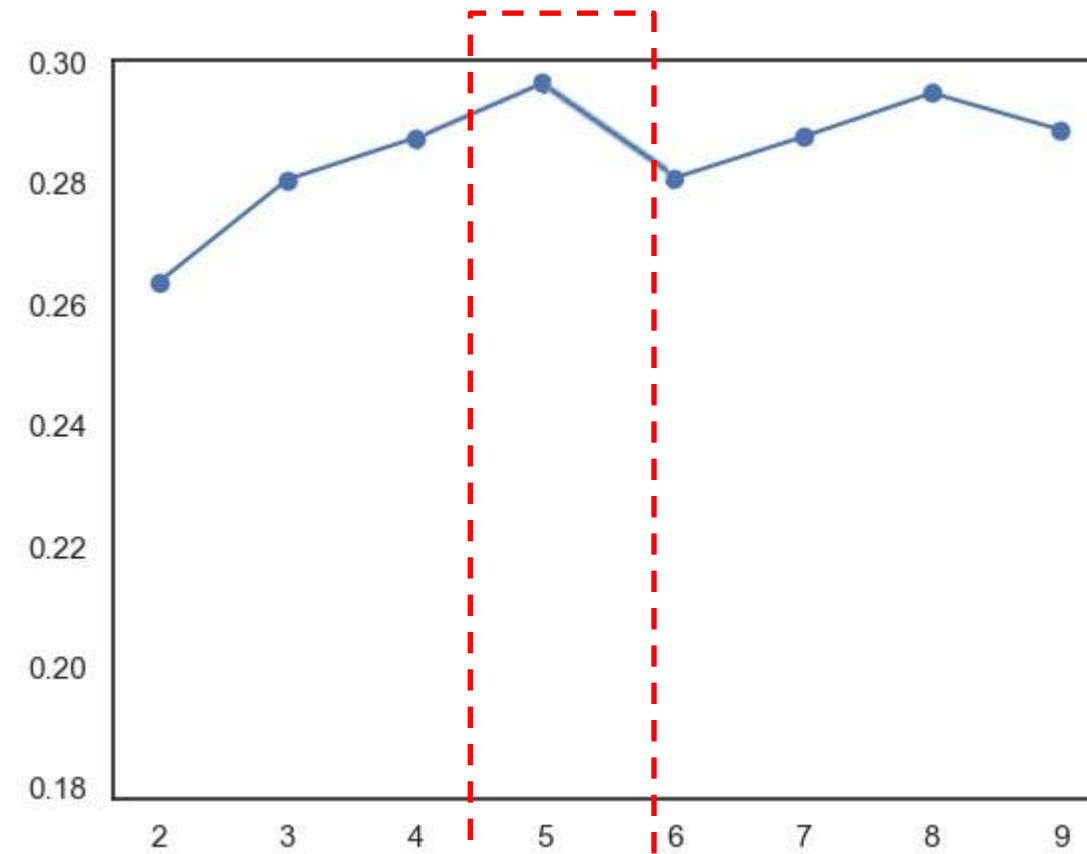
Le coefficient de silhouette mesure la similarité de chaque point de données d'un cluster par rapport aux autres clusters.

Nous retenons la valeur la plus élevée dans une sélection entre 2 et 10, ici $k=5$

Méthode du coude

En traçant la courbe de l'inertie intra-classe en fonction de k , nous cherchons alors à identifier les étapes où l'on observe une rupture dans la décroissance de cette courbe

Cette courbe ne nous renseigne pas car aucun coude se forme dans notre courbe



Algorithme K-means

Détermination de K

La recherche de K détermine le nombre de clusters de notre analyse

Coefficient de silhouette

Le coefficient de silhouette mesure la similarité de chaque point de données d'un cluster par rapport aux autres clusters.

Nous retenons la valeur la plus élevée dans une sélection entre 2 et 10, ici k=5

Méthode du coude

En traçant la courbe de l'inertie intra-classe en fonction de k, nous cherchons alors à identifier les étapes où l'on observe une rupture dans la décroissance de cette courbe

Cette courbe ne nous renseigne pas car aucun coude se forme dans notre courbe

Algorithme K-means

classek0

	pays
26	Cabo Verde
31	Chine - RAS de Hong-Kong
35	Comores
36	Congo
55	Gabon
92	Macédoine du Nord
114	Oman
155	Timor-Leste

classek1

	pays
0	Afghanistan
3	Algérie
5	Angola
9	Arménie
12	Azerbaïdjan
...	...
161	Ukraine
163	Vanuatu
164	Viet Nam
165	Zambie
166	Zimbabwe

classek2

	pays
23	Brésil
49	États-Unis d'Amérique
167	Chine

classek3

	pays
1	Afrique du Sud
2	Albanie
4	Allemagne
6	Antigua-et-Barbuda
7	Arabie saoudite
...	...
149	Suisse
150	Suriname
153	Tchéquie
157	Trinité-et-Tobago

classek4

	pays
69	Inde

Analyse des classes

Visualisation des distributions en boxplot de chaque variable par classe

Classe k0

Pays qui consomment mais n'exportent et n'importent pas

Classe k1

Pays pauvres et instables qui n'importent et n'exportent pas

Classe k2

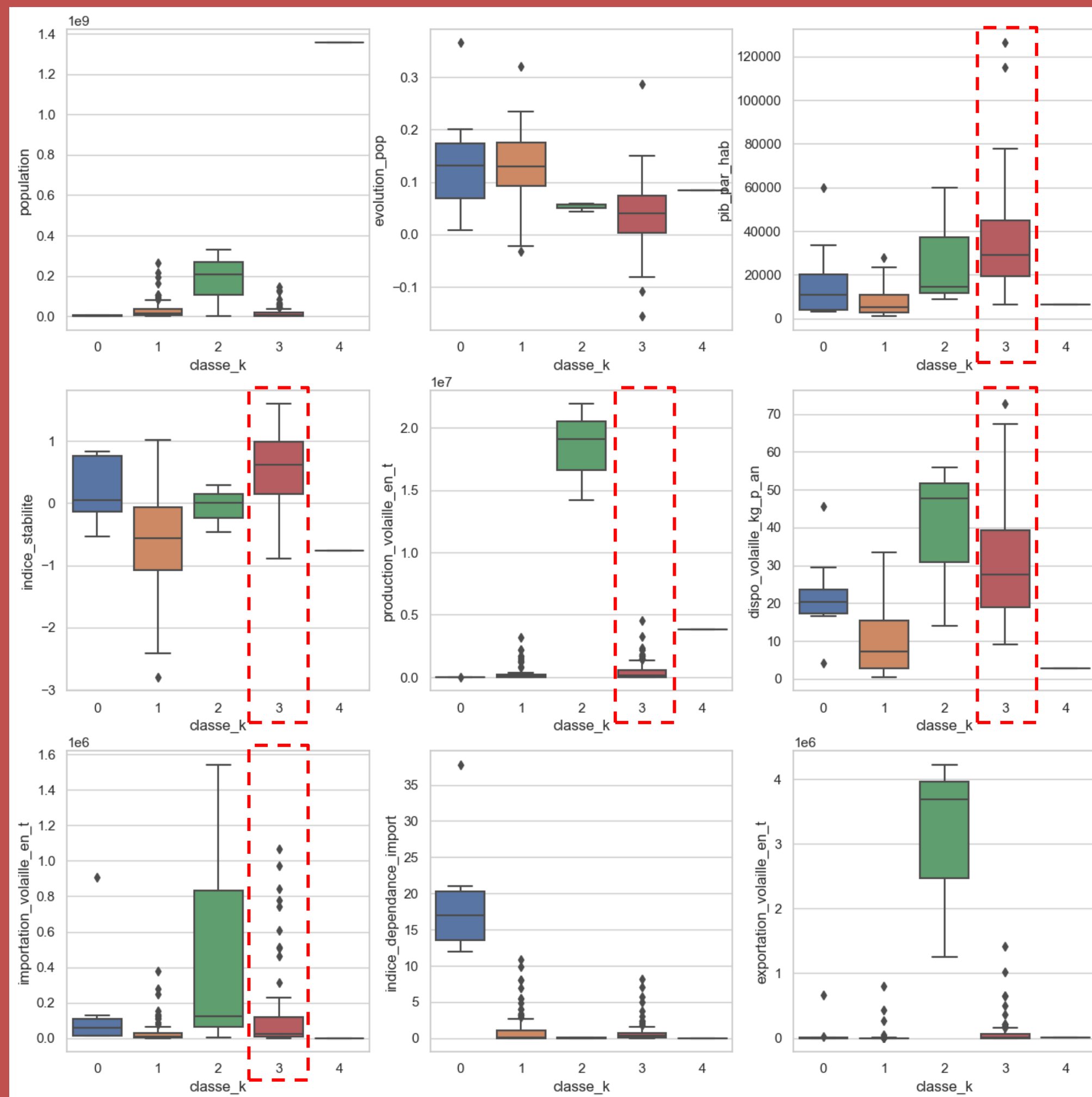
Pays très gros producteurs et exportateurs de viande de volaille

Classe k3

Pays riches et stables qui consomment et importent de la viande de volaille en quantité

Classe k4

Inde, pays très peuplé producteur mais pas importateur



Algorithmme K-means

Classement des pays

Création d'un score à partir des valeurs pondérées
de chaque variable

Comparaison avec le classement de la méthode
précédente

	pays	score
0	Japon	71.202769
1	Mexique	63.473552
2	Allemagne	54.947325
3	Royaume-Uni de Grande-Bretagne et d'Irlande du...	50.621130
4	Fédération de Russie	46.025948
5	Arabie saoudite	42.775178
6	France	40.265552
7	Émirats arabes unis	38.345245
8	Afrique du Sud	37.919885
9	Pays-Bas (Royaume des)	35.624187
10	Canada	30.258330
11	Chine - RAS de Macao	29.603731
12	Espagne	27.839655
13	République de Corée	27.078680
14	Italie	26.563122
15	Koweït	25.404849
16	Australie	24.046061
17	Belgique	23.852227
18	Argentine	22.874467
19	Grenade	22.316505

	pays	score
0	Japon	59.705019
1	Mexique	54.530161
2	Allemagne	48.238815
3	Royaume-Uni de Grande-Bretagne et d'Irlande du...	45.718630
4	Arabie saoudite	42.088397
5	Indonésie	39.959683
6	Émirats arabes unis	38.605685
7	France	35.672071
8	Pays-Bas (Royaume des)	34.765680
9	Fédération de Russie	34.746038
10	Afrique du Sud	34.682957
11	Chine - RAS de Macao	29.737881
12	Canada	27.351340
13	Koweït	26.904297
14	Espagne	24.918073
15	Belgique	24.110028
16	République de Corée	23.561802
17	Grenade	22.981591
18	Australie	22.507534
19	Italie	22.329417



Analyse en Composantes Principales

Analyse en Composantes Principales

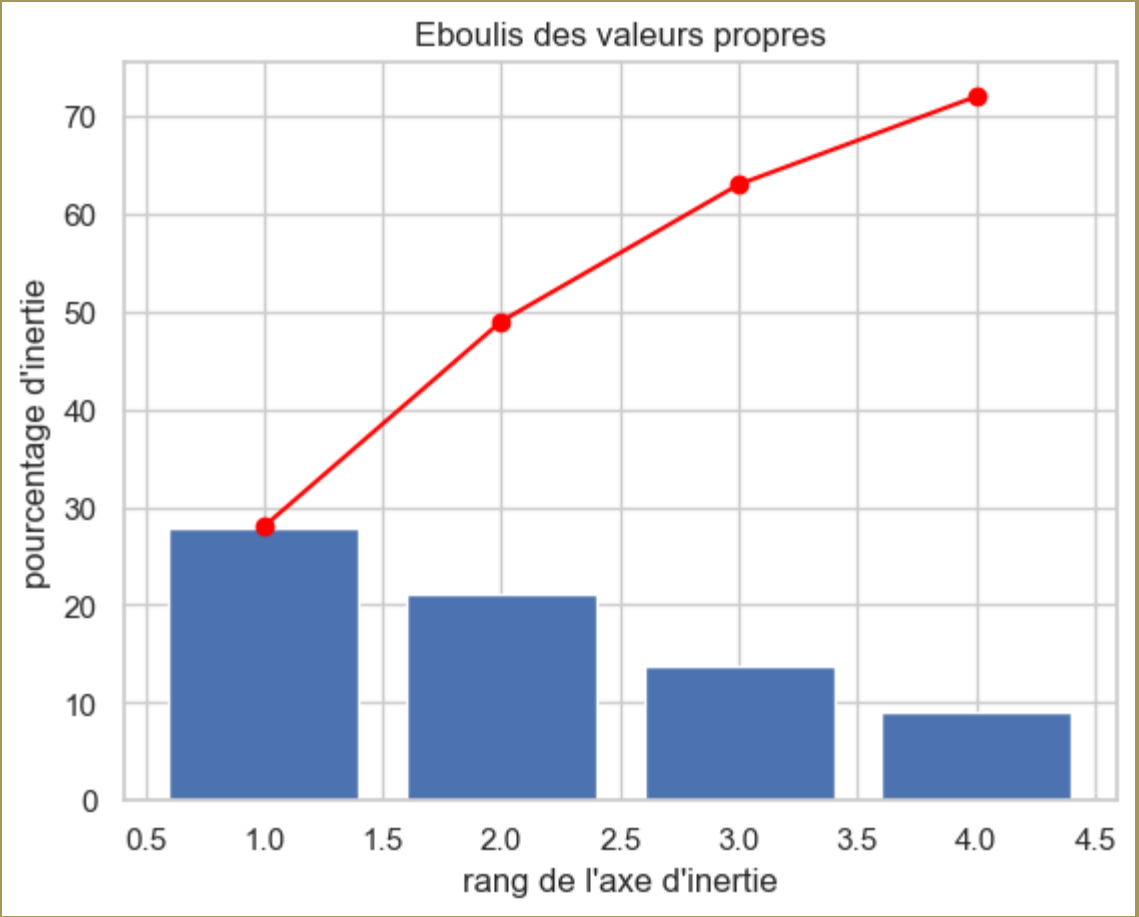
Le nombre de composantes à conserver est déterminé en analysant l'éboulis des valeurs propres obtenu à partir de l'analyse en composantes principales. Les valeurs les plus importantes sont celles qui expliquent le plus la variabilité.

Eboulis

Ici, la variance est distribuée de manière équitable.

Composantes

Nombre de composantes sélectionnées : 4



Cercle de corrélations

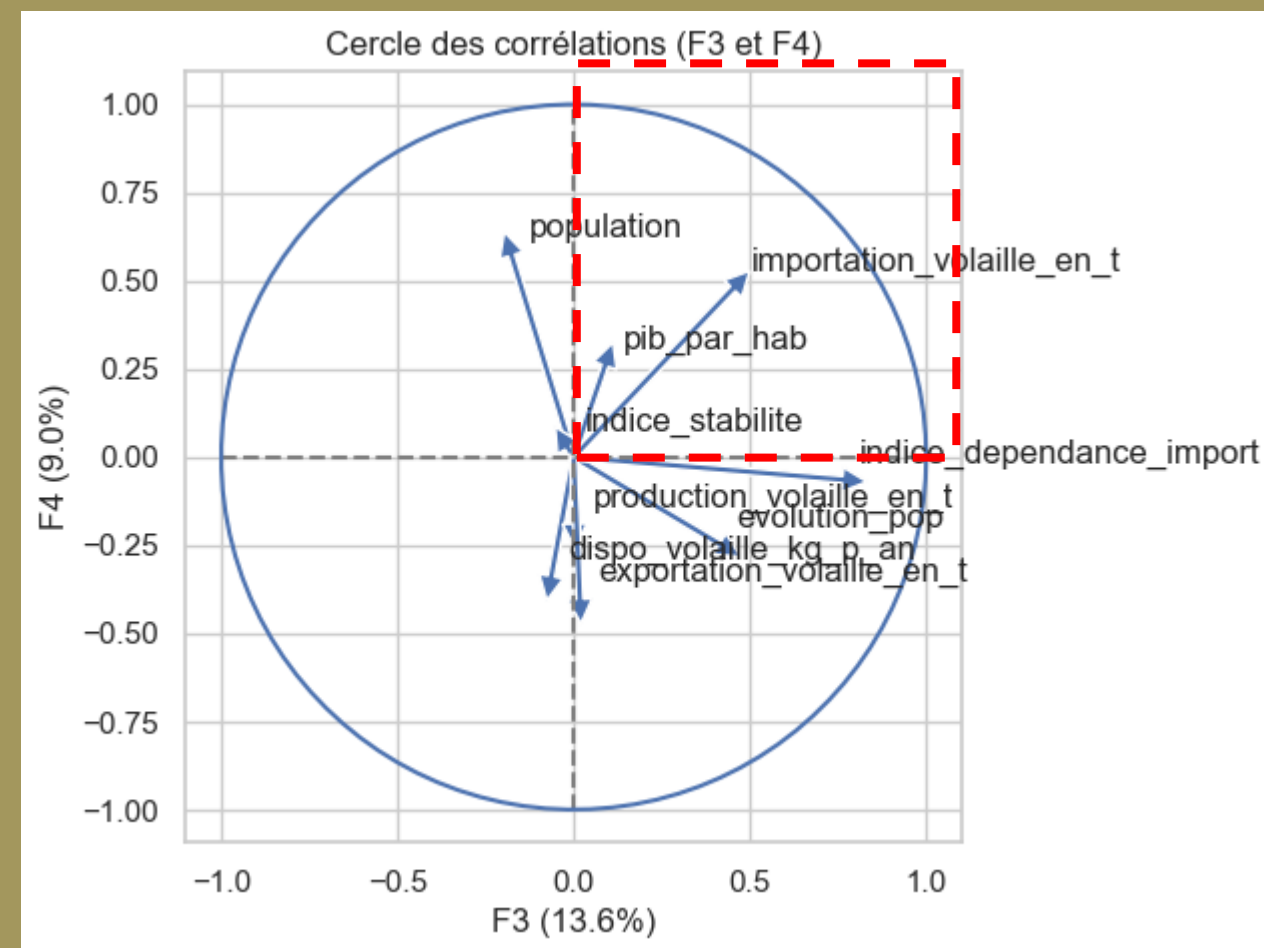
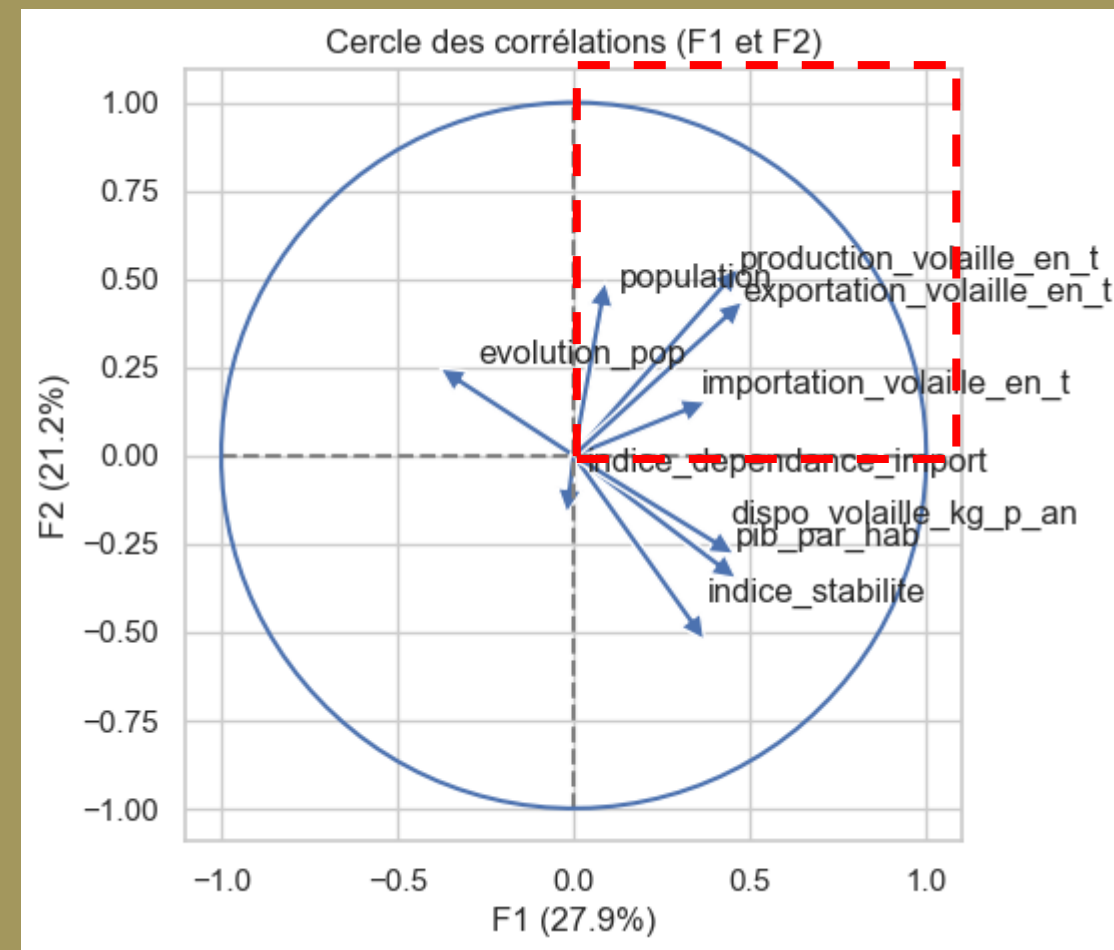
Le cercle de corrélations permet de visualiser les coefficients de corrélation entre les variables et les 4 composantes principales appelées F1 F2 F3 et F4. Il permet aussi de repérer les variables corrélées entres elles.

F1 et F2

La production et l'exportation sont fortement corrélées sur les 2 premières composantes. La population et l'importation aussi à moindre mesure.

F3 et F4

L'importation et le PIB par habitant sont fortement corrélés sur les 2 composantes suivantes.



Projection des individus par classe

Le but est de visualiser les individus dans un système à 2 dimensions pour mieux comprendre l'organisation des données. Il est possible d'évaluer la force des corrélations entre les individus et les variables à travers les composantes.

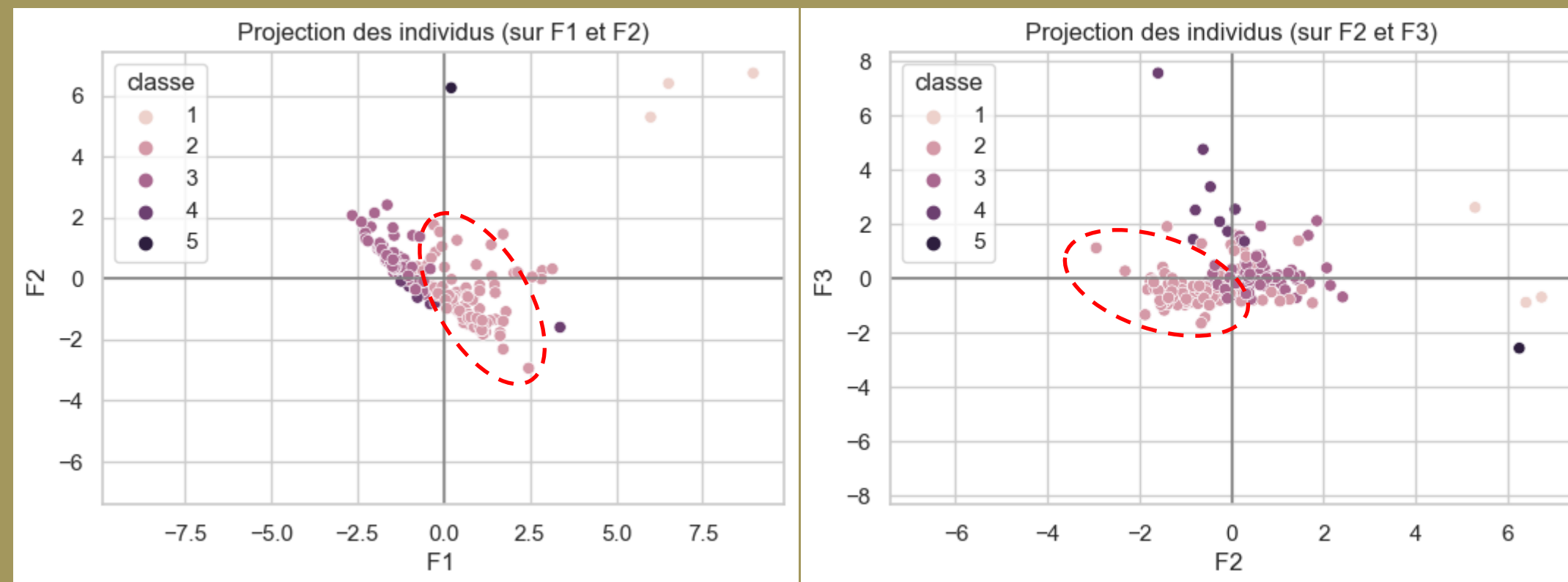
Classes CAH

Les classes 4 et 5 dépendent des composantes 0 et 1
& les classes 1 et 5 dépendent des composantes 2 et 3

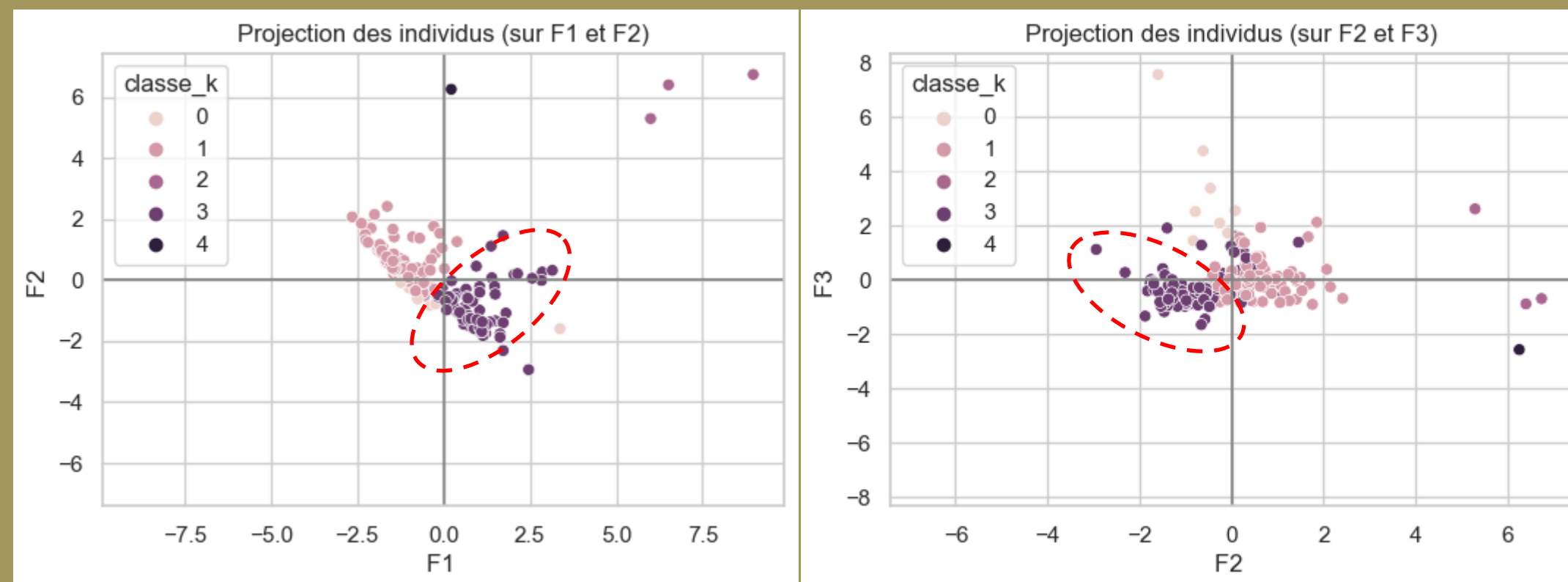
Classes K-means

Les classes 4 et 1 dépendent des composantes 0 et 1
& les classes 3 et 4 dépendent des composantes 2 et 3

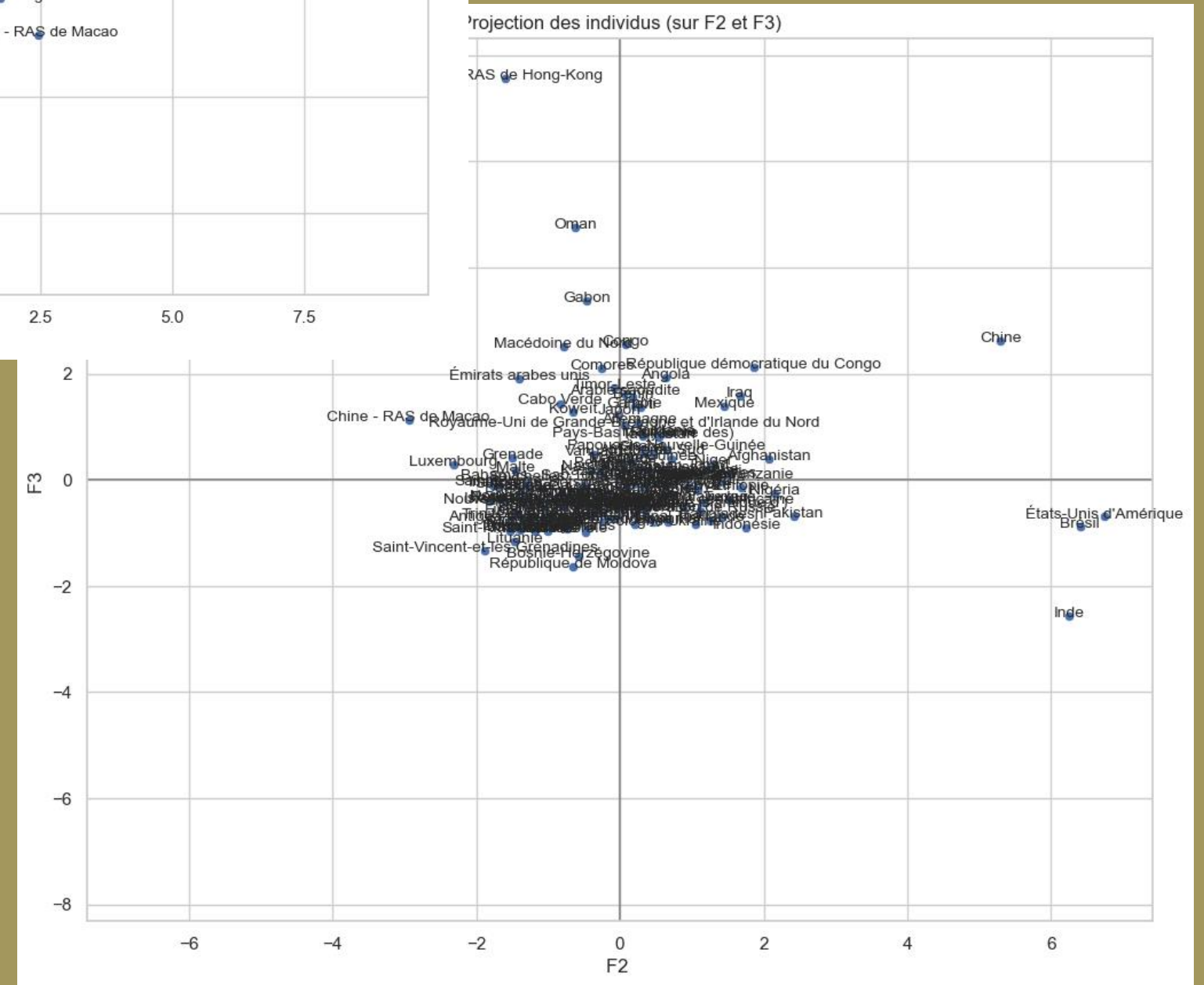
Projection des classes CAH



Projection des classes K-means



Les USA, le Brésil ou la Chine dépendent des
composantes F1 et F2
& la Chine, le Congo, ou le Mexique dépendent des
composantes F3 et F4





03

Conclusion

Plusieurs clusters géographiques ont été repérés

#1 Europe

le cluster le plus propice à une internationalisation est le cluster européen, avec en tête l'Allemagne et le Royaume Unis qui obtiennent un excellent score

#2 Moyen-Orient

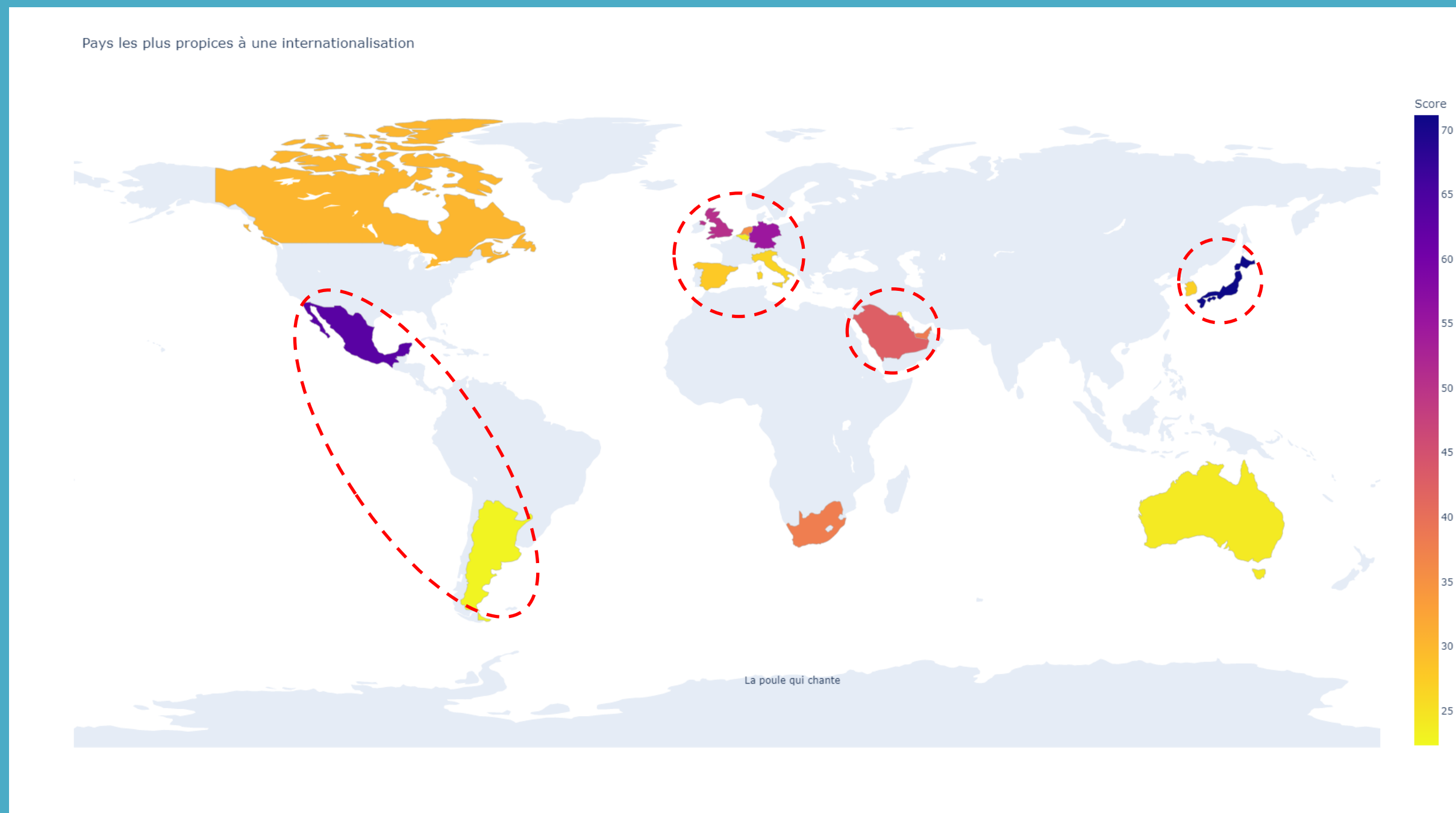
Le cluster est composé de l'Arabie Saoudite et des Emirats Arabes Unis, c'est un cluster intéressant de par la proximité

#3 Asie

En Extrême Orient, nous avons un cluster composé du Japon qui obtient le score le plus élevé, et de la Corée du Sud

#4 Amérique du Sud

Enfin L'Amérique du Sud composé du Mexique avec un excellent score et de l'Argentine





Merci !

C'était Lokman

Data analyst pour La poule qui chante...