



La poule qui chante

Etude de marché Implantation à l'international

Melvin Derouck - Data Analyst
Décembre 2023

SOMMAIRE

- Contexte & Objectifs
- Méthodologie
- Réduction des dimensions (ACP)
- K-Means
- CAH
- Caractérisation des clusters
(Heatmap)
- Scoring et résultats





Contexte & Objectif

Après une croissance significative en France, l'entreprise “La Poule qui chante” souhaite se développer à l'international. Mais aucun pays ni continent n'est pour le moment choisi.

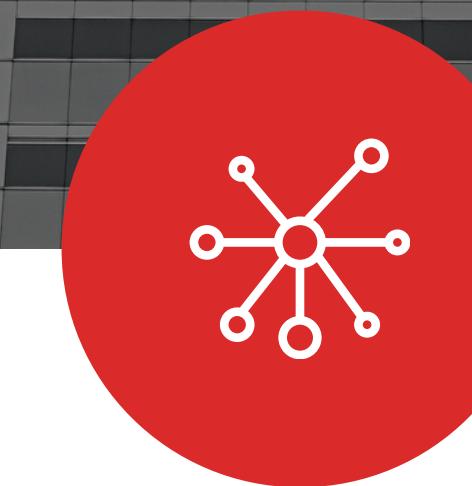
L'objectif principal de ce projet de data analyse est d'établir une liste de 3 pays les plus pertinents, et cela à l'aide de plusieurs méthodes de clustering.

Méthodologie



Réduction des dimensions via l'ACP

Transformation des variables originales en un nouveau jeu de variables non corrélées, afin de faciliter la visualisation et l'interprétation des tendances.



Clustering avec K-Means et CAH

Regroupement des pays en clusters basés sur des similarités dans leurs caractéristiques de marché. Les résultats des 2 méthodes seront ensuite comparés.



Scoring et sélection des pays

Attribution de scores aux pays dans nos clusters, selon 2 scénarios différents, afin d'identifier et mettre en évidence les trois meilleurs pays cibles.

Nettoyage et traitement des données

4 exports de données utilisés dans le cadre du projet :

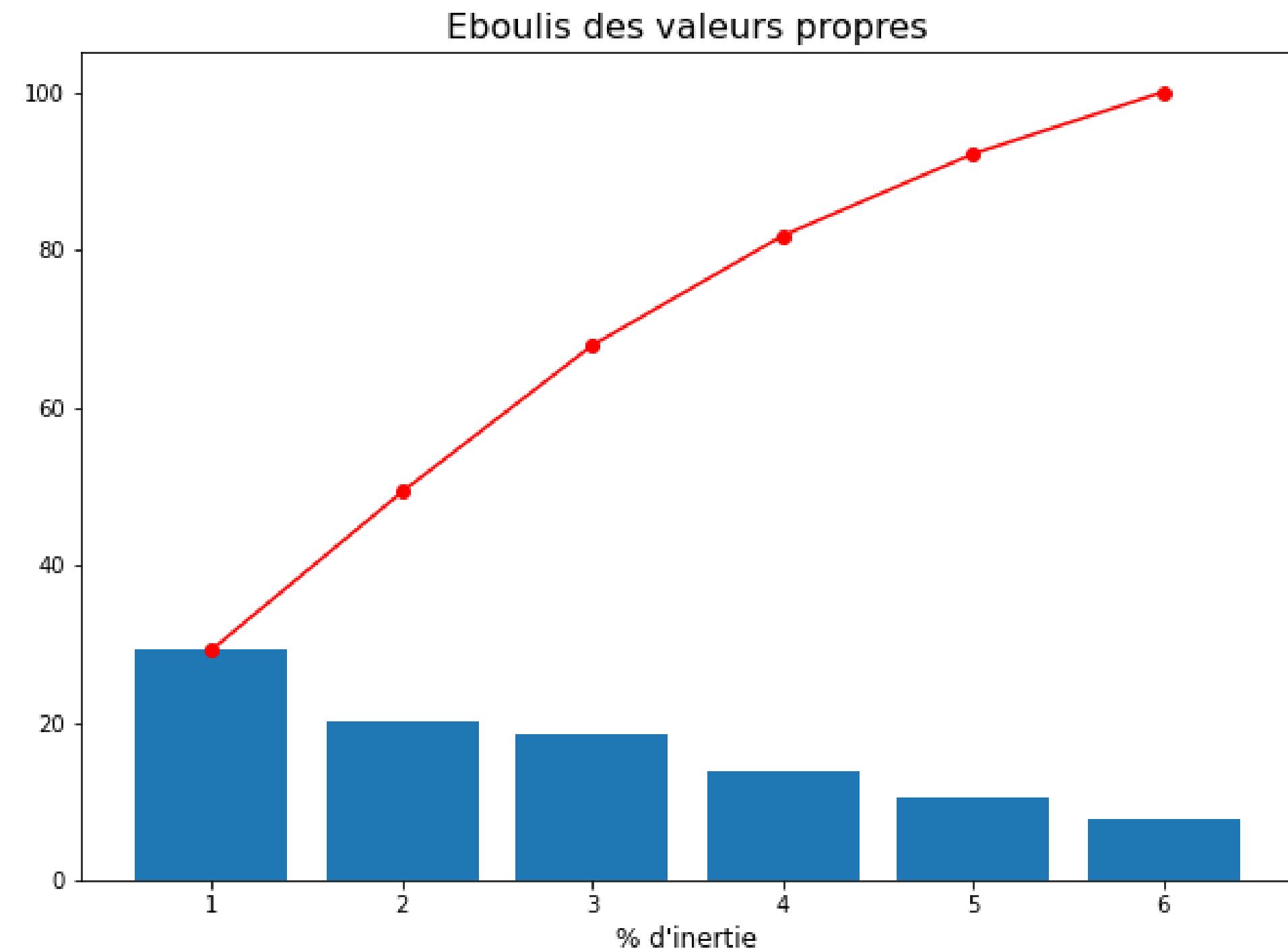
- Fichier “Disp_alim”
- Fichier “Population”
- Fichier “PIB”
- Fichier “Distance”



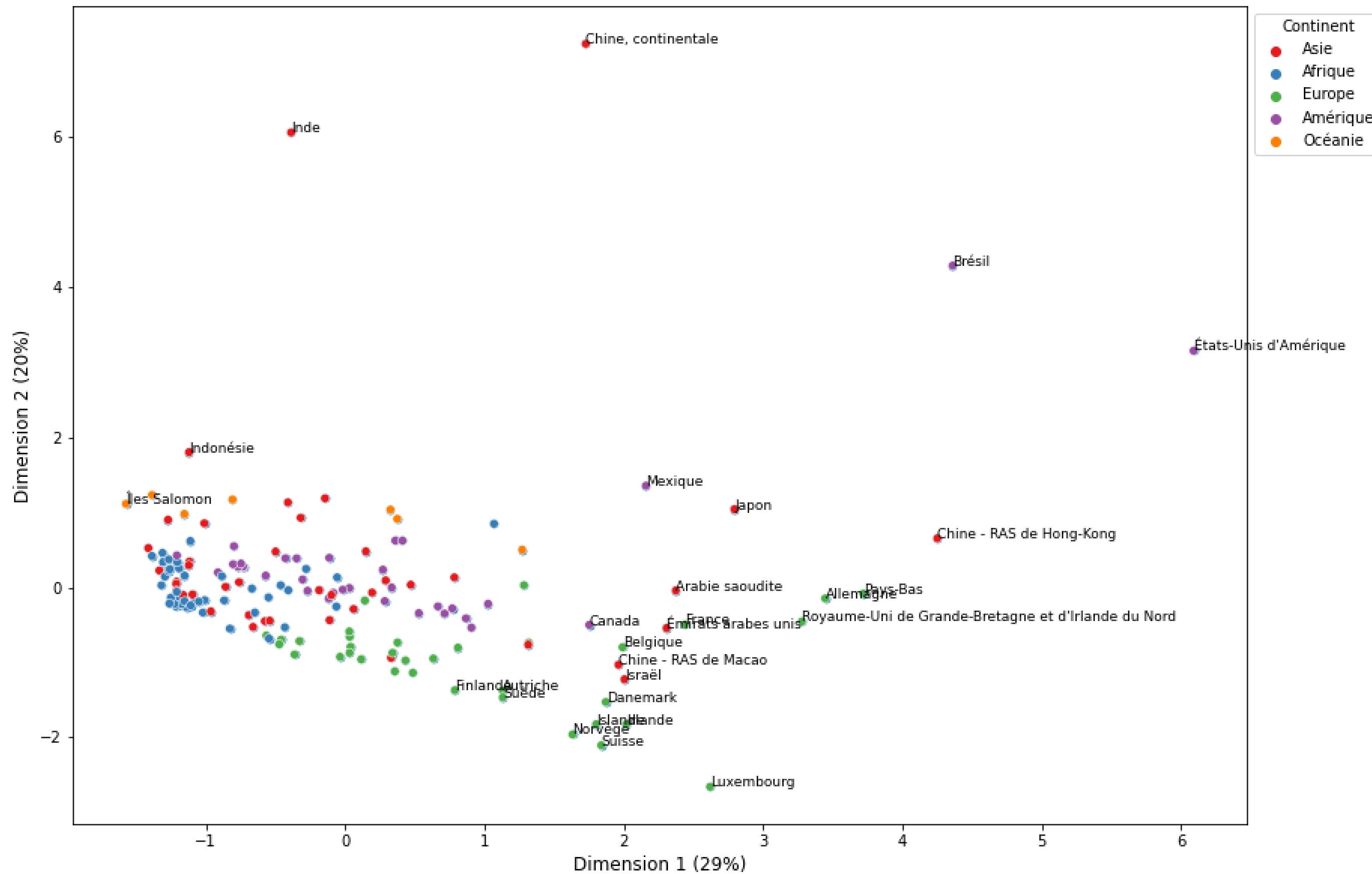
kaggle

Réduction des dimensions avec l'ACP

- Transformer des données complexes en structures plus simples à comprendre et à visualiser
- Réduction du nombre de variables sans perte d'informations essentielles.
- On applique le modèle sur 2 composantes principales
- Elles couvrent 49% de la variance expliquée

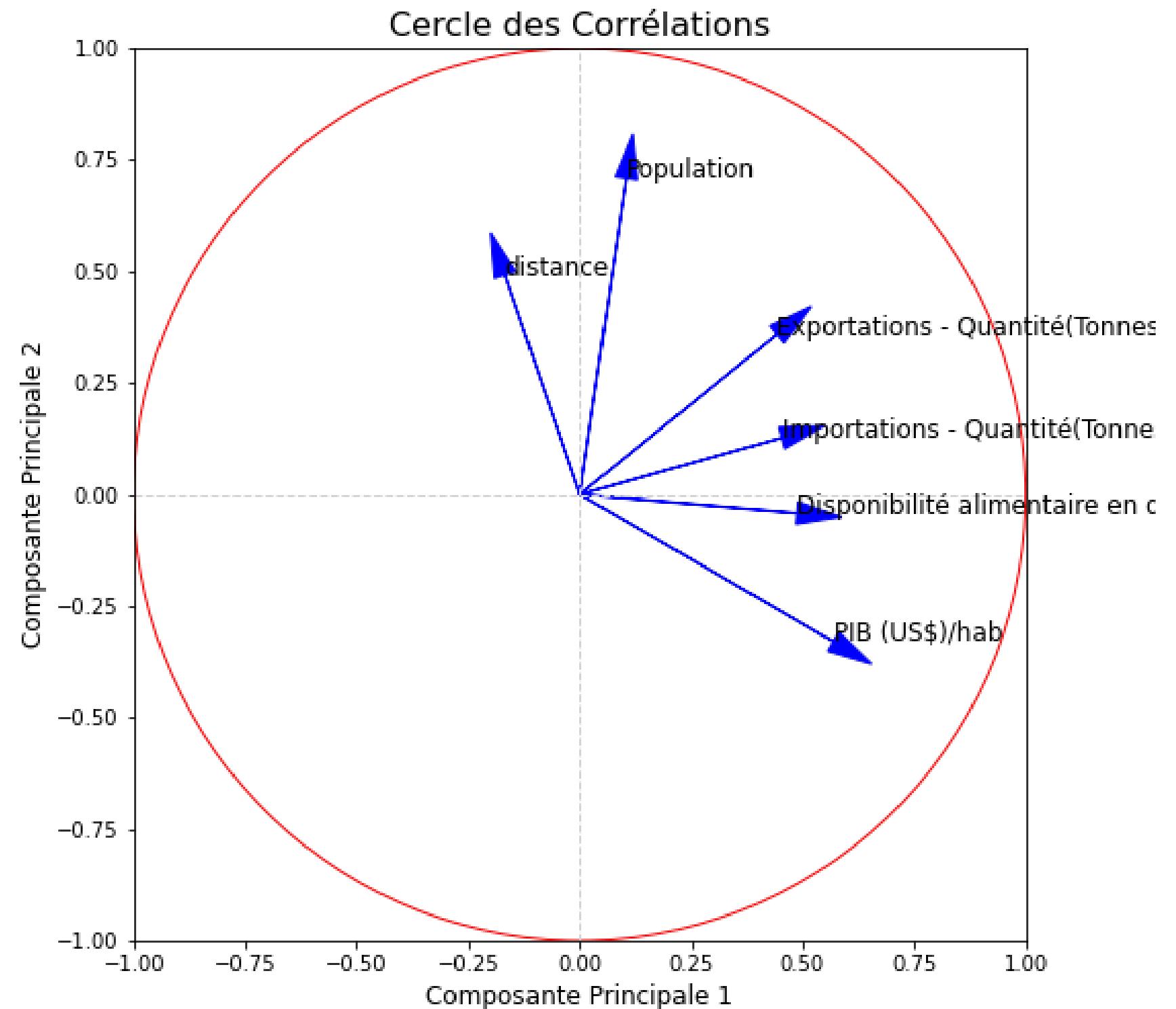


Premier plan factoriel (49%)



Cercle des corrélations

- Permet de simplifier la complexité des différentes dimensions de l'ACP
- Axe horizontal = 1ère composante principale
Axe vertical = 2ème composante principale
- Chaque flèche représente une variable
- Il permet aussi de voir les corrélations au sein des données



Détermination des clusters

K-MEANS
&
CAH

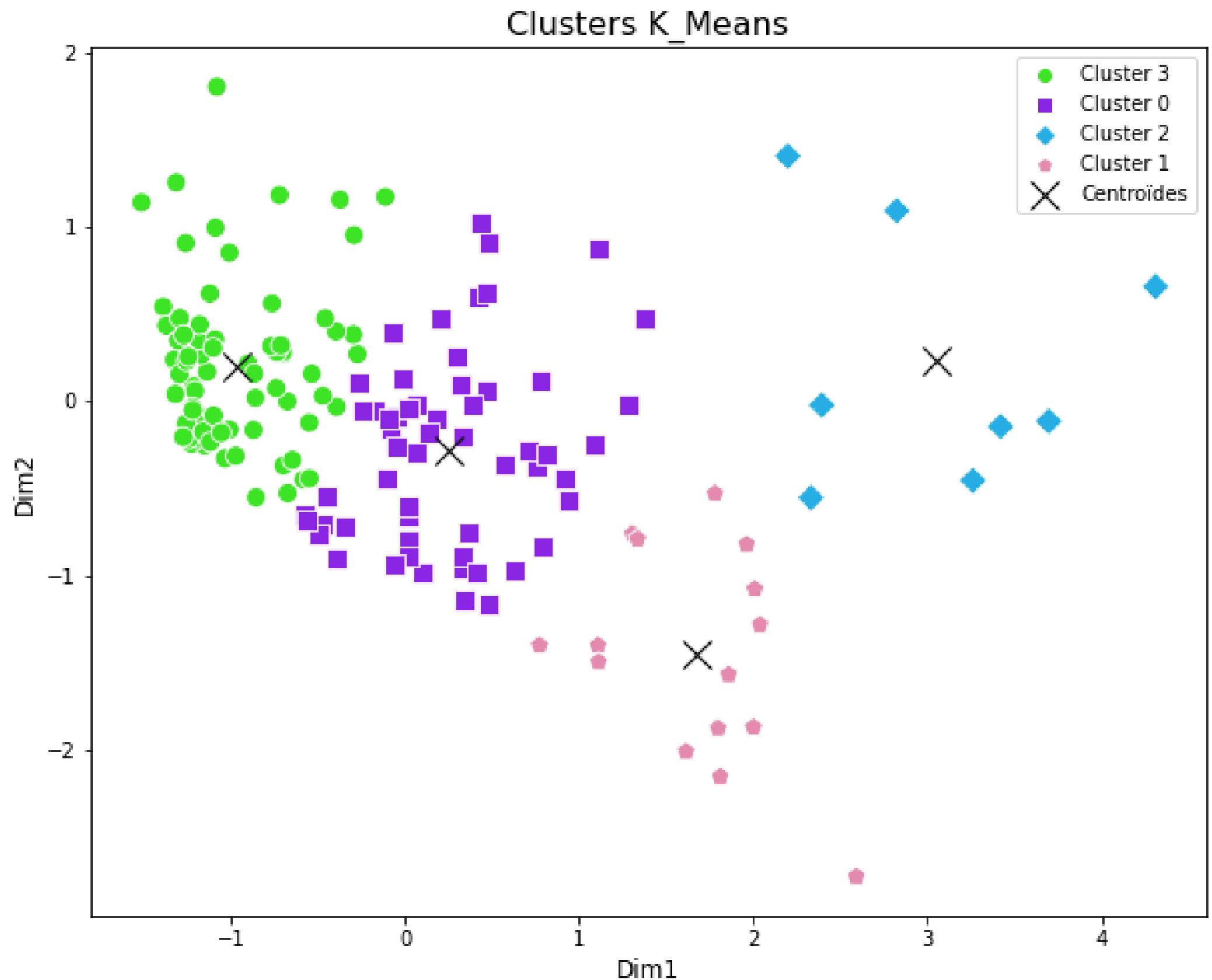
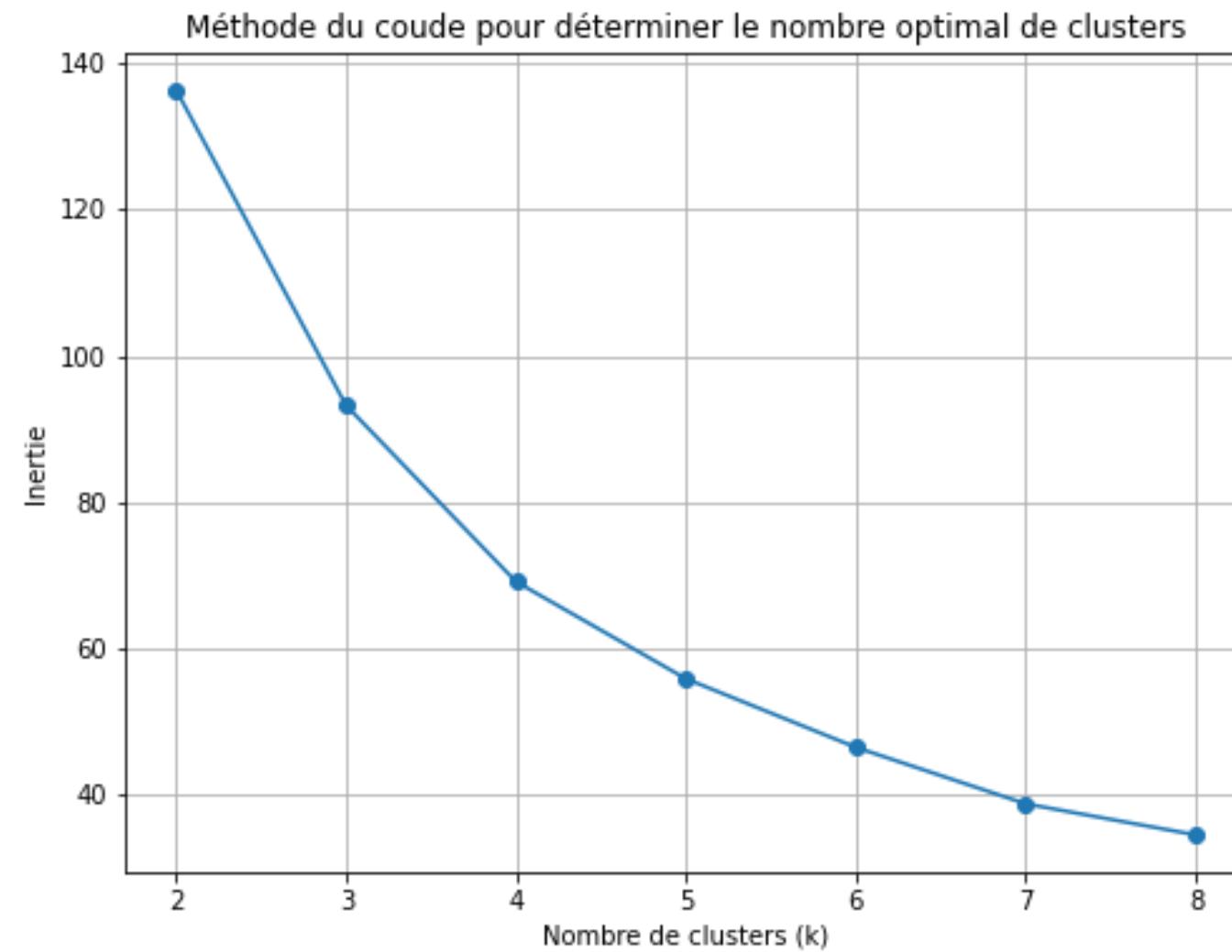
Méthode des K-Means

2 variables : PIB et Importations

Standardisation des données

Méthode du coude pour définir le nombre optimal de clusters

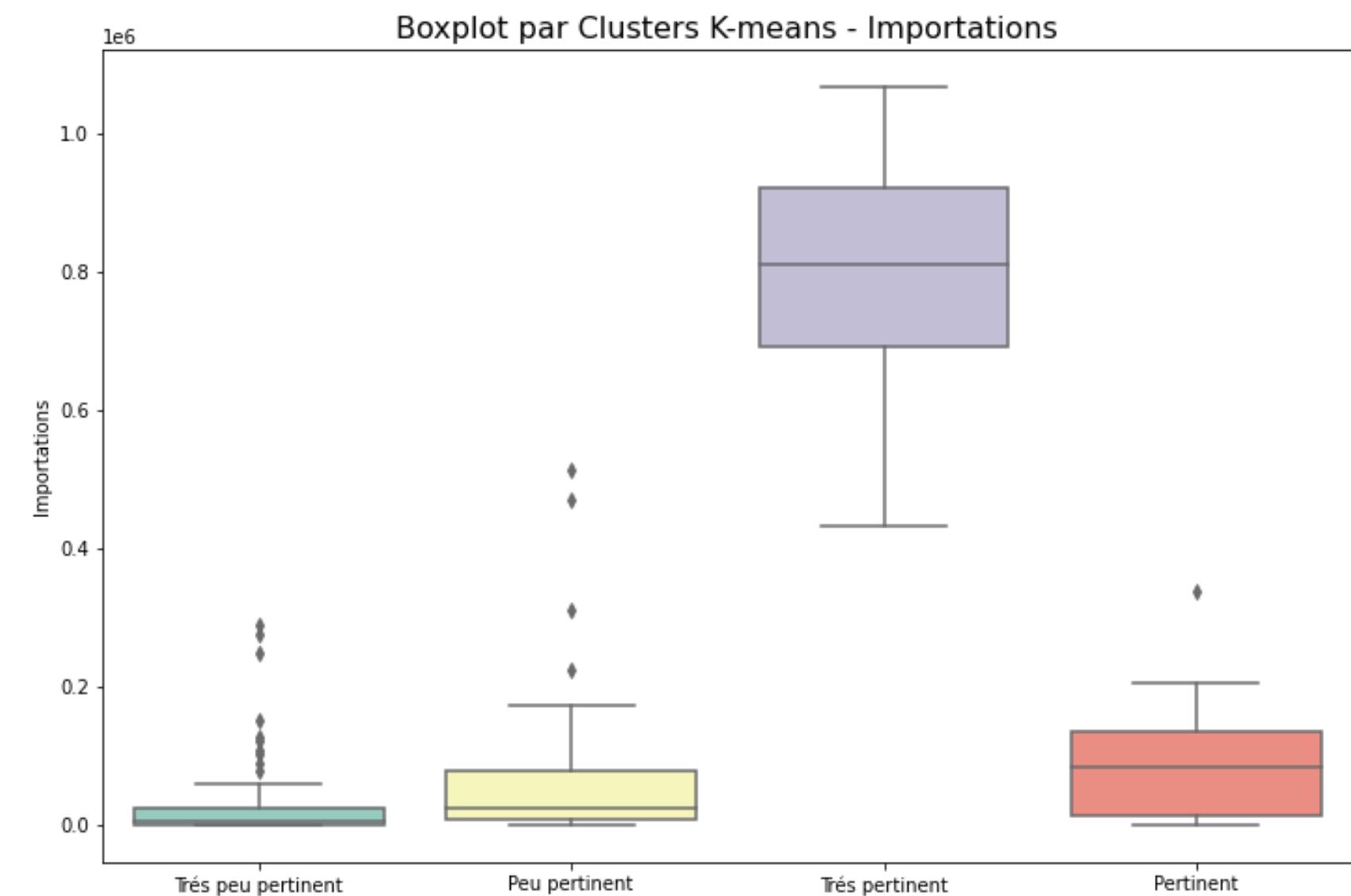
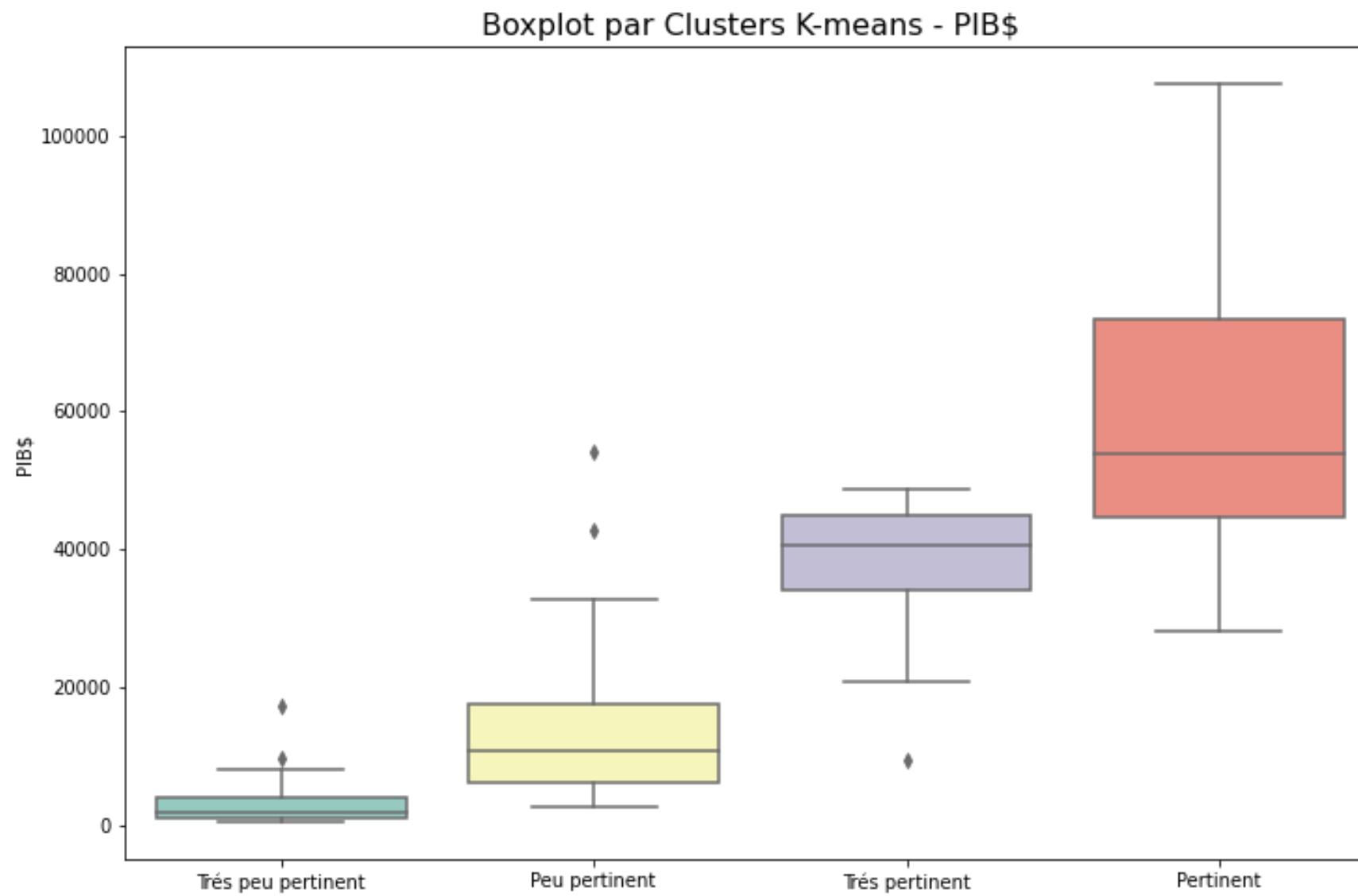
=> ici 4



Caractérisation des clusters

Les 4 clusters sont représentés indépendamment par un boxplot.

Ces représentations nous aide à comprendre les particularités de chaque cluster.

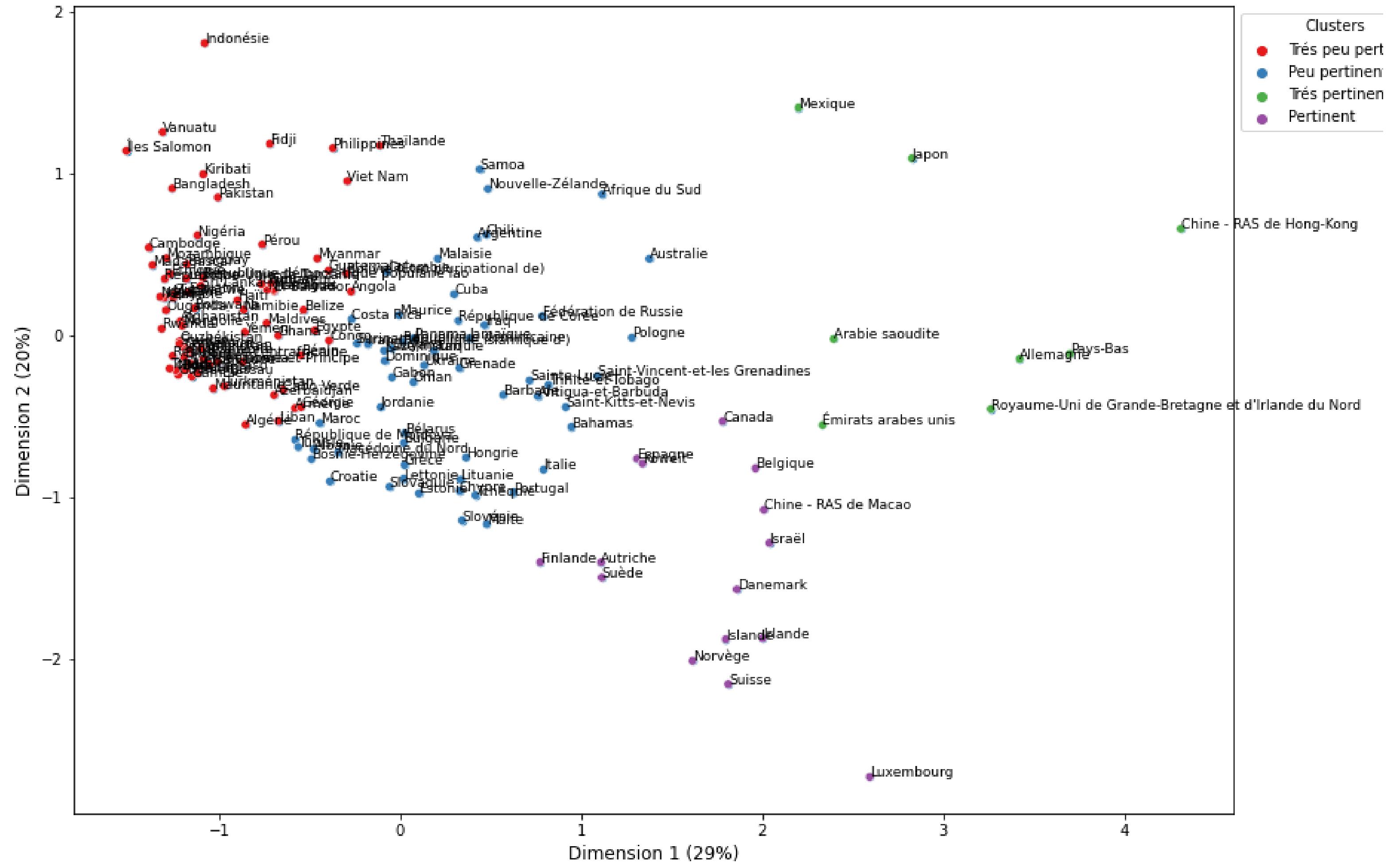


Résultats - K Means

8 pays ont été identifiés comme “Très pertinent”

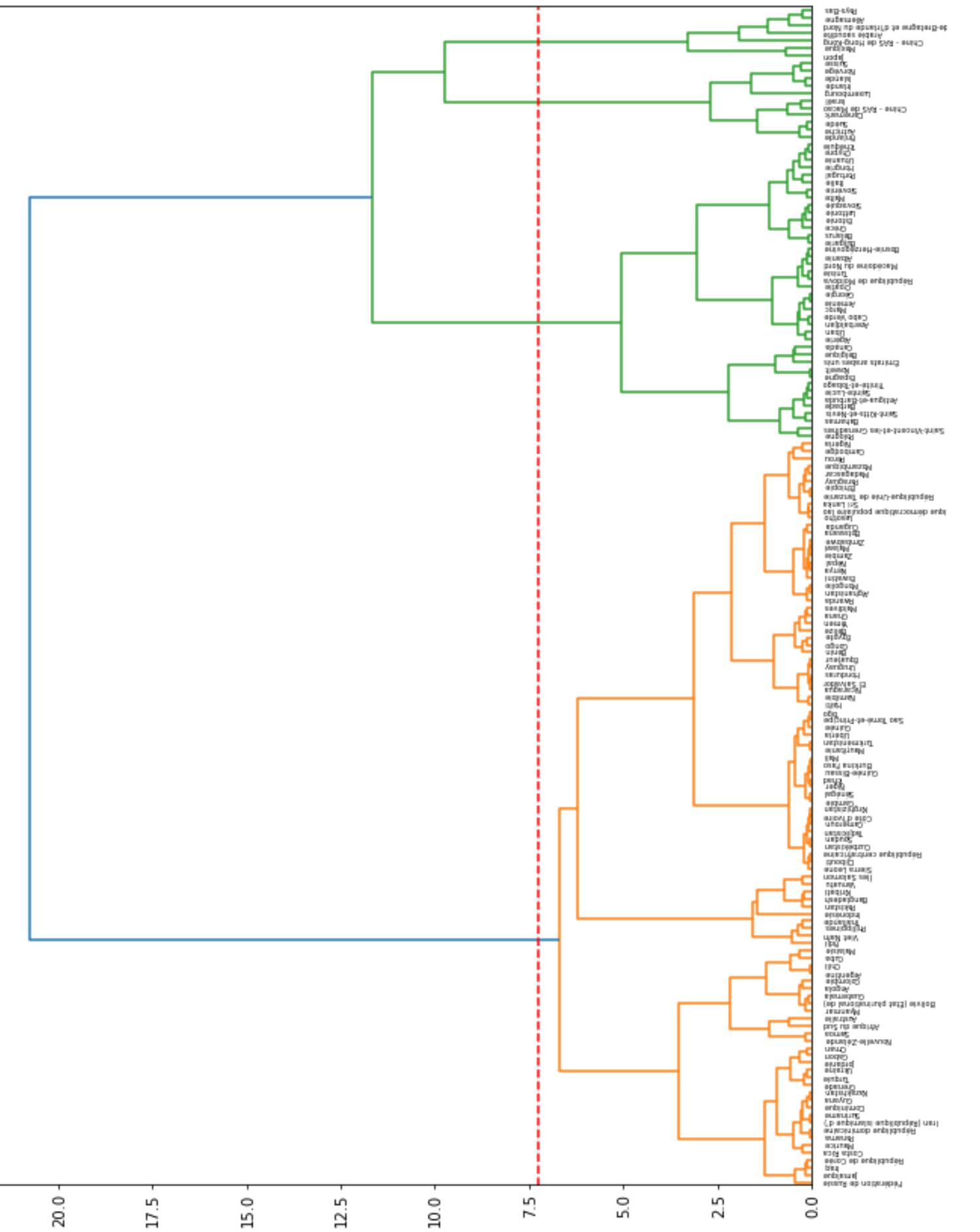
Silhouette score = 0.68

	iso3	Pays	Continent	Population	PIB (US\$)/hab	Disponibilité alimentaire en quantité (kg/personne/an)	Exportations - Quantité(Tonnes)	Importations - Quantité(Tonnes)	distance	Cluster_kmeans
4	DEU	Allemagne	Europe	82658409.0	44552.819372	19.47	646000.0	842000.0	439.8984	Très pertinent
7	SAU	Arabie saoudite	Asie	33101179.0	20803.745918	43.36	10000.0	722000.0	4694.5000	Très pertinent
30	HKG	Chine - RAS de Hong-Kong	Asie	7306322.0	46165.856512	53.51	663000.0	907000.0	9639.4760	Très pertinent
70	JPN	Japon	Asie	127502725.0	38386.511146	18.50	10000.0	1069000.0	9725.6290	Très pertinent
93	MEX	Mexique	Amérique	124777324.0	9287.849736	32.52	9000.0	972000.0	9206.8130	Très pertinent
110	NLD	Pays-Bas	Europe	17021347.0	48675.222335	20.33	1418000.0	608000.0	427.9169	Très pertinent
115	GBR	Royaume-Uni de Grande-Bretagne et d'Irlande du...	Europe	66727461.0	40361.417383	31.94	359000.0	779000.0	342.9475	Très pertinent
154	ARE	Émirats arabes unis	Asie	9487203.0	40644.804043	43.47	94000.0	433000.0	5249.5350	Très pertinent



Méthode de la CAH

- Méthode de clustering qui construit une hiérarchie de clusters
- Le Dendrogramme représente la manière dont les clusters sont formés en hiérarchie
- Chaque fusion de clusters est représentée par une ligne, et la hauteur de la ligne indique la distance entre les clusters fusionnés
- 4 clusters sont sélectionnés



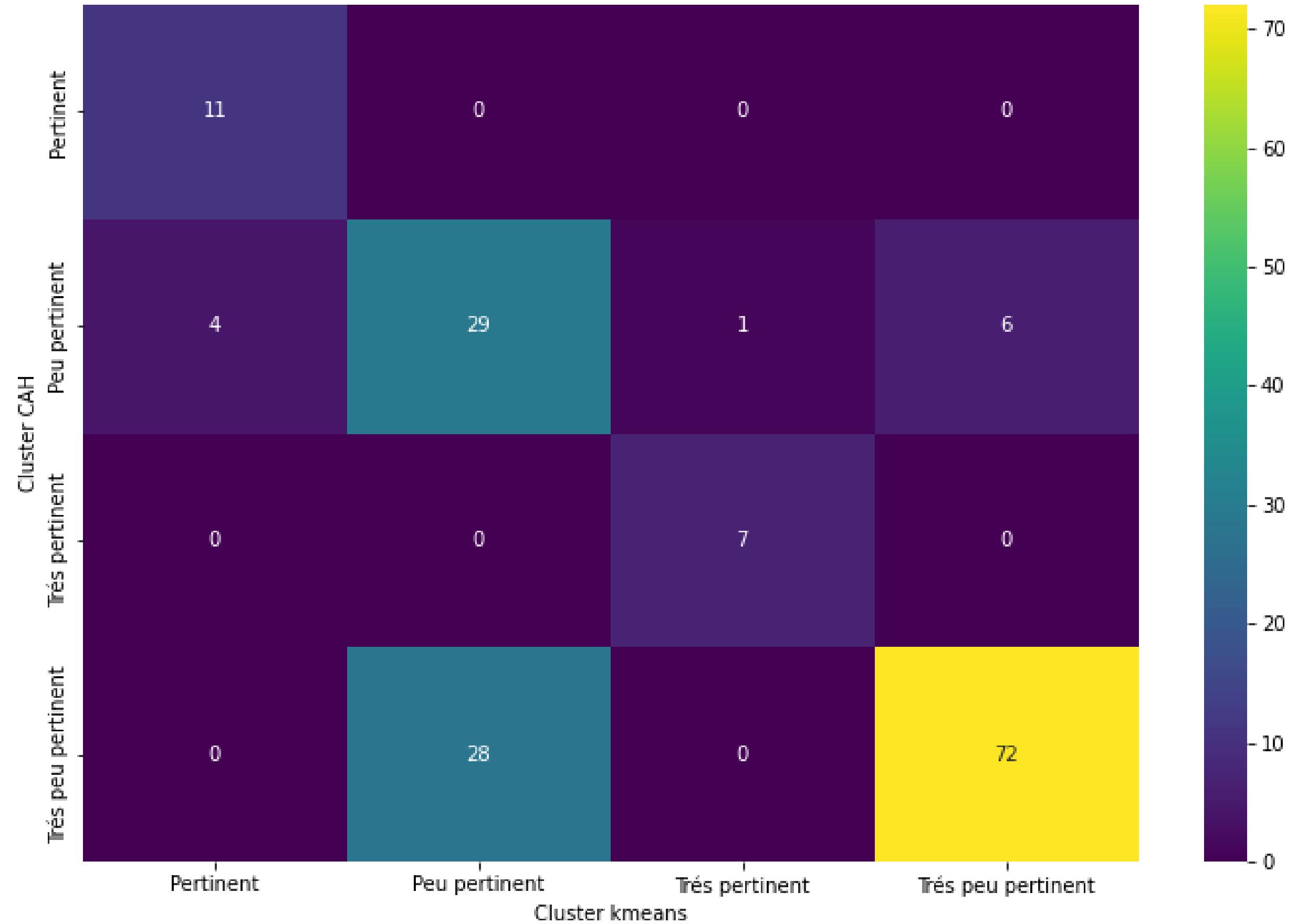
Résultats - CAH

7 pays ont été identifiés comme “Très pertinent”

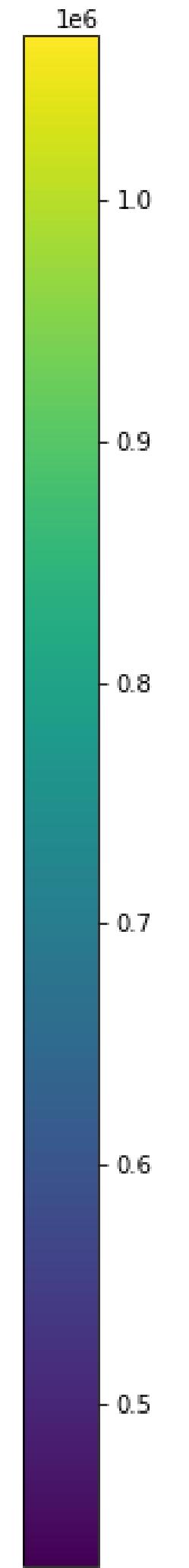
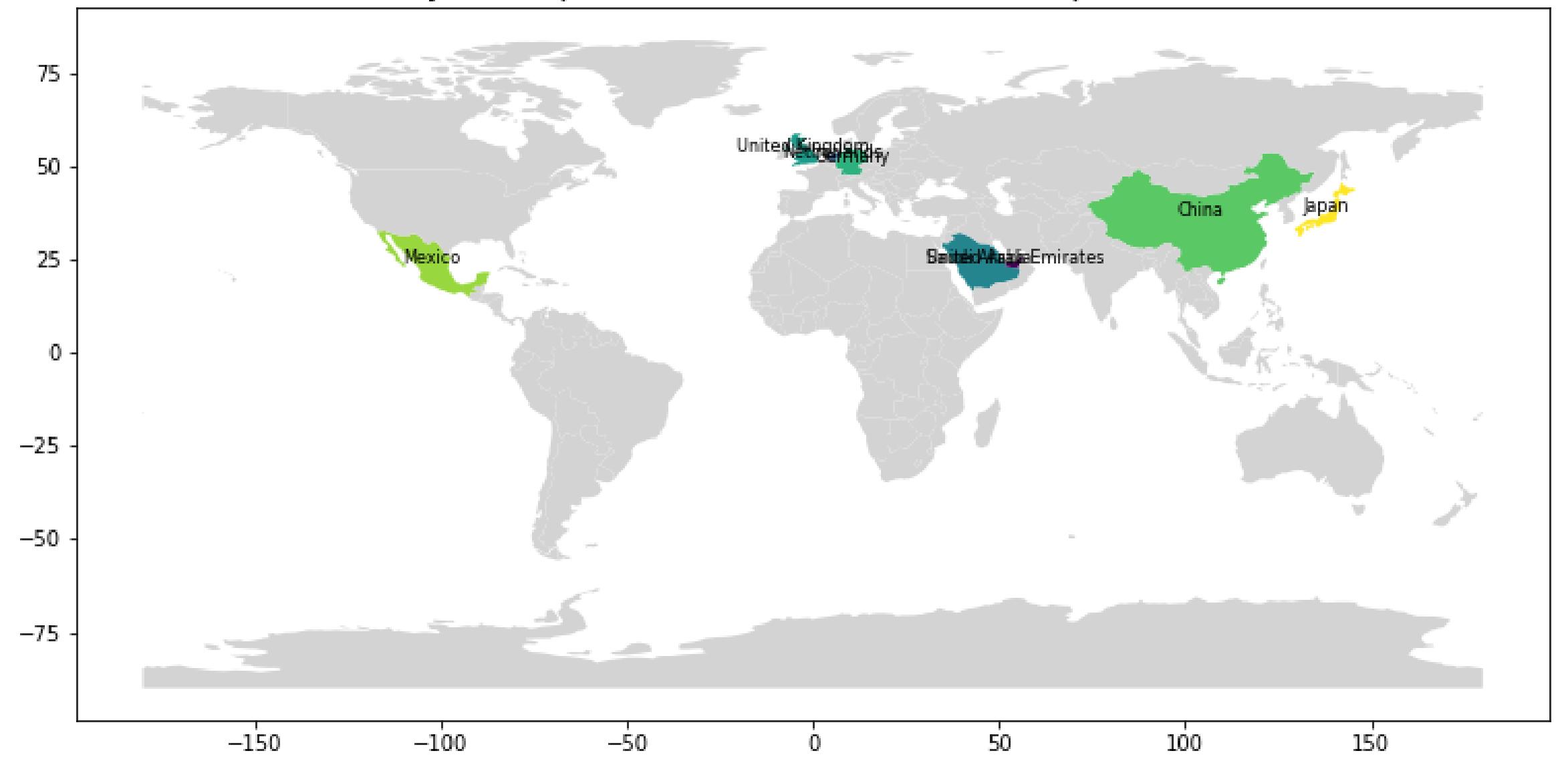
Indice de Davies-Bouldin = 0.81

	iso3	Pays	Continent	Population	PIB (US\$)/hab	Disponibilité alimentaire en quantité (kg/personne/an)	Exportations - Quantité(Tonnes)	Importations - Quantité(Tonnes)	distance	Cluster_kmeans	Cluster_CAH
4	DEU	Allemagne	Europe	82658409.0	44552.819372	19.47	646000.0	842000.0	439.8984	Très pertinent	Très pertinent
7	SAU	Arabie saoudite	Asie	33101179.0	20803.745918	43.36	10000.0	722000.0	4694.5000	Très pertinent	Très pertinent
30	HKG	Chine - RAS de Hong-Kong	Asie	7306322.0	46165.856512	53.51	663000.0	907000.0	9639.4760	Très pertinent	Très pertinent
70	JPN	Japon	Asie	127502725.0	38386.511146	18.50	10000.0	1069000.0	9725.6290	Très pertinent	Très pertinent
93	MEX	Mexique	Amérique	124777324.0	9287.849736	32.52	9000.0	972000.0	9206.8130	Très pertinent	Très pertinent
110	NLD	Pays-Bas	Europe	17021347.0	48675.222335	20.33	1418000.0	608000.0	427.9169	Très pertinent	Très pertinent
115	GBR	Royaume-Uni de Grande-Bretagne et d'Irlande du...	Europe	66727461.0	40361.417383	31.94	359000.0	779000.0	342.9475	Très pertinent	Très pertinent

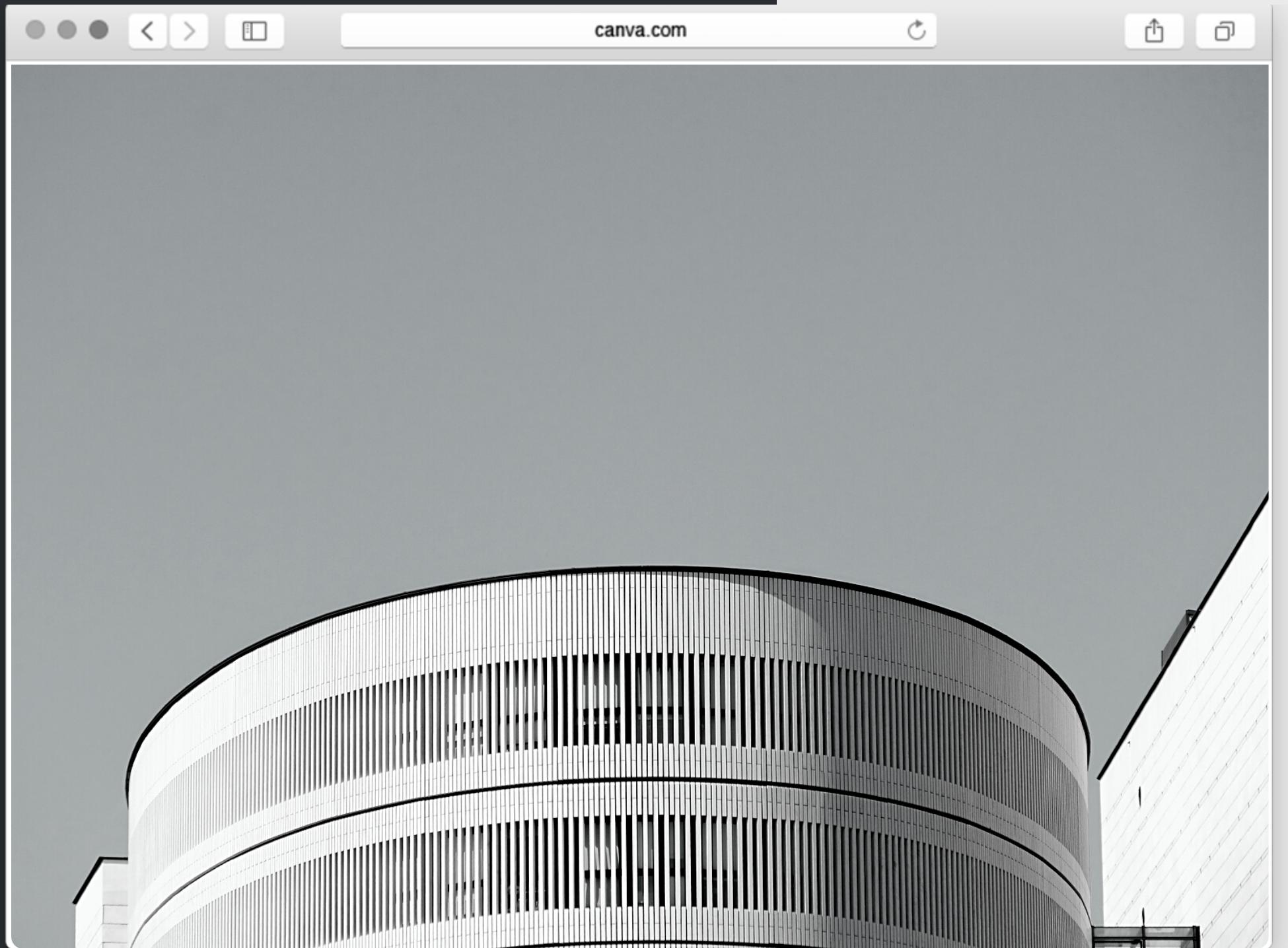
Heatmap des Clusters



Pays 'Très pertinent' avec indicateur d'importations



Scoring & Résultats



Scénario écologique

**Poids élevé sur variable “Distance” et
“Population”**

"PIB (US\$)/hab": 5,
"distance": 8,
"Population": 6,
"Exportations – Quantité(Tonnes)": 4,
"Importations – Quantité(Tonnes)": 4

	Pays	Score_Total
4	Allemagne	20.536171
110	Pays-Bas	18.513139
115	Royaume-Uni	18.080526

Scénario business

Poids élevé sur “PIB” et “Importation”

"PIB (US\$)/hab": 9,
"distance": 3,
"Population": 4,
"Exportations – Quantité(Tonnes)": 5,
"Importations – Quantité(Tonnes)": 8

	Pays	Score_Total
4	Allemagne	20.939788
110	Pays-Bas	19.497395
70	Japon	18.652582

**Merci pour votre
attention !**



La poule qui chante

